REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0798

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPOF
	December, 1994	Final

4. TITLE AND SUBTITLE
USAF Summer Research Program - 1994 Summer Faculty Research Program Final Reports, Volume 5B, Wright Laboratory

5. FUNDING NUMBERS

6. AUTHORS
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research and Development Labs, Culver City, CA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
4040 Fairfax Dr, Suite 500
Arlington, VA 22203-1613

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Contract Number: F49620-93-C-0063

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Summer Faculty Research Program (USAF- SFRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES

15. NUMBER OF PAGES

16. PRICE CODE

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

RESEARCH & DEVELOPMENT LABORATORIES
AFOSR PROGRAM OFFICE

March 5, 1999

Lilla Mae Davis,

The following pages are unavailable because of bindery problems. In 1994 the publishing company we used had problems with the bindery.

Volume 5A page 4-5 missing.

Volume 5B page 34-6 missing.
Volume 5B page 58-14 side of page cut-off.

Volume 8 page 15-4 missing.

Johnetta Thompson
Program Administrator

RDL, AFOSR PROGRAM OFFICE, 5800 UPLANDER WAY, CULVER CITY, CALIFORNIA 90230-6608 (800)677-1363 & (310)410-1244
e-mail (INTERNET) ADDRESS: afosr@rdl.com FAX (800)677-1363 & (310)216-6940

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1994

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 5B

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

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.

December 1994

PREFACE

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

Due to its length, Volume 5 is bound in two parts, 5A and 5B. Volume 5A contains reports #1-33. Volume 5B contains reports #34-66. The Table of Contents for Volume 5 is included in both parts.

This document is one of a set of 16 volumes describing the 1994 AFOSR Summer Research Program. The following volumes comprise the set:

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

SFRP FINAL REPORT TABLE OF CONTENTS i-xxi

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

1. INTRODUCTION

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

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

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

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

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

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

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

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest; and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1994 participants expressed in end-of-tour SRP evaluations (Appendix B).

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

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

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

Table 1: SRP Participation, by Year

YEAR	Number of Participants			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years; in one case a laboratory did not fund any additional associates. However, the table shows that, overall, the number of participating associates increased this year because two laboratories funded more associates than they had in previous years.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 44-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous (over 600 annually) individual requesters.

Due to a delay in awarding the new contract, RDL was not able to place advertisements in any of the following publications in which the SRP is normally advertised: *Black Issues in Higher Education, Chemical & Engineering News, IEEE Spectrum* and *Physics Today*.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

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

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

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1994 are shown in this table.

Table 2: 1994 Applicants and Participants

PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	600	192	30
(HBCU/MI)	(90)	(16)	(7)
GSRP	322	117	11
(HBCU/MI)	(11)	(6)	(0)
HSAP	562	133	14
TOTAL	1484	442	55

4. SITE VISITS

During June and July of 1994, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

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

In previous years, an RDL program representative visited from seven to ten different HBCU/MIs to promote interest in the SRP among the faculty and graduate students. Due to the late contract award date (January 1994) no time was available to visit HBCU/MIs this past year.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. This year, seven HBCU/MI SFRPs declined after they were selected. The following table records HBCU/MI participation in this program.

Table 3: SRP HBCU/MI Participation by Year

YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6

6. SRP FUNDING SOURCES

Funding sources for the 1994 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1994 SRP selected participants are shown here.

Table 4: 1994 SRP Associate Funding

FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	150	98*1	121*2
USAF Laboratory Funds	37	19	12
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	5	0	0
TOTAL	192	117	133

*1 - 100 were selected, but two canceled too late to be replaced.
*2 - 125 were selected, but four canceled too late to be replaced.

7. COMPENSATION FOR PARTICIPANTS

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

Table 5: 1994 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994
Faculty Members	$690	$718	$740	$740
Graduate Student (Master's Degree)	$425	$442	$455	$455
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391
High School Student (First Year)	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $50/day for faculty and $37/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and GSRP associates, 58% (178 out of 309) claimed travel reimbursements at an average round-trip cost of $860.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-one percent (78 out of 192) of faculty associates took orientation trips at an average cost of $498. Many faculty associates noted on their evaluation forms that due to the late notice of acceptance into the 1994 SRP (caused by the late award in January 1994 of the contract) there wasn't enough time to attend an orientation visit prior to their tour start date. In 1993, 58 % of SFRP associates took orientation visits at an average cost of $685.

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

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

8. CONTENTS OF THE 1994 REPORT

The complete set of reports for the 1994 SRP includes this program management report augmented by fifteen volumes of final research reports by the 1994 associates as indicated below:

Table 6: 1994 SRP Final Report Volume Assignments

	VOLUME		
LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, FJSRL, WHMC	6	11	16

AEDC = Arnold Engineering Development Center
FJSRL = Frank J. Seiler Research Laboratory
WHMC = Wilford Hall Medical Center

APPENDIX A -- PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP and GSRP associates represent 158 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 46 states plus Washington D.C. and Puerto Rico. Selectees represent 40 states.

GSRP - Applicants came from 46 states and Puerto Rico. Selectees represent 34 states.

HSAP - Applicants came from fifteen states. Selectees represent ten states.

C. Academic Disciplines Represented

The academic disciplines of the combined 192 SFRP associates are as follows:

Electrical Engineering	22.4%
Mechanical Engineering	14.0%
Physics: General, Nuclear & Plasma	12.2%
Chemistry & Chemical Engineering	11.2%
Mathematics & Statistics	8.1%
Psychology	7.0%
Computer Science	6.4%
Aerospace & Aeronautical Engineering	4.8%
Engineering Science	2.7%
Biology & Inorganic Chemistry	2.2%
Physics: Electro-Optics & Photonics	2.2%
Communication	1.6%
Industrial & Civil Engineering	1.6%
Physiology	1.1%
Polymer Science	1.1%
Education	0.5%
Pharmaceutics	0.5%
Veterinary Medicine	0.5%
TOTAL	100%

Table A-1. Total Participants

Number of Participants	
SFRP	192
GSRP	117
HSAP	133
TOTAL	442

Table A-2. Degrees Represented

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	189	0	189
Master's	3	47	50
Bachelor's	0	70	70
TOTAL	192	117	309

Table A-3. SFRP Academic Titles

Academic Titles	
Assistant Professor	74
Associate Professor	63
Professor	44
Instructor	5
Chairman	1
Visiting Professor	1
Visiting Assoc. Prof.	1
Research Associate	3
TOTAL	192

Table A-4. Source of Learning About SRP

SOURCE	SFRP		GSRP	
	Applicants	Selectees	Applicants	Selectees
Applied/participated in prior years	26%	37%	10%	13%
Colleague familiar with SRP	19%	17%	12%	12%
Brochure mailed to institution	32%	18%	19%	12%
Contact with Air Force laboratory	15%	24%	9%	12%
Faculty Advisor (GSRPs Only)	--	--	39%	43%
Other source	8%	4%	11%	8%
TOTAL	100%	100%	100%	100%

Table A-5. Ethnic Background of Applicants and Selectees

	SFRP		GSRP		HSAP	
	Applicants	Selectees	Applicants	Selectees	Applicants	Selectees
American Indian or Native Alaskan	0.2%	0%	1%	0%	0.4%	0%
Asian/Pacific Islander	30%	20%	6%	8%	7%	10%
Black	4%	1.5%	3%	3%	7%	2%
Hispanic	3%	1.9%	4%	4.5%	11%	8%
Caucasian	51%	63%	77%	77%	70%	75%
Preferred not to answer	12%	14%	9%	7%	4%	5%
TOTAL	100%	100%	100%	100%	99%	100%

Table A-6. Percentages of Selectees receiving their 1st, 2nd, or 3rd Choices of Directorate

	1st Choice	2nd Choice	3rd Choice	Other Than Their Choice
SFRP	70%	7%	3%	20%
GSRP	76%	2%	2%	20%

APPENDIX B -- SRP EVALUATION RESPONSES

1. OVERVIEW

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

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	275
HSAPs	116
USAF Laboratory Focal Points	109
USAF Laboratory HSAP Mentors	54

All groups indicate near-unanimous enthusiasm for the SRP experience.

Typical comments from 1994 SRP associates are:

> "[The SRP was an] excellent opportunity to work in state-of-the-art facility with top-notch people."
>
> "[The SRP experience] enabled exposure to interesting scientific application problems; enhancement of knowledge and insight into 'real-world' problems."
>
> "[The SRP] was a great opportunity for resourceful and independent faculty [members] from small colleges to obtain research credentials."
>
> "The laboratory personnel I worked with are tremendous, both personally and scientifically. I cannot emphasize how wonderful they are."
>
> "The one-on-one relationship with my mentor and the hands on research experience improved [my] understanding of physics in addition to improving my library research skills. Very valuable for [both] college and career!"

Typical comments from laboratory focal points and mentors are:

"This program [AFOSR - SFRP] has been a 'God Send' for us. Ties established with summer faculty have proven invaluable."

"Program was excellent from our perspective. So much was accomplished that new options became viable "

"This program managed to get around most of the red tape and 'BS' associated with most Air Force programs. Good Job!"

'Great program for high school students to be introduced to the research environment. Highly educational for others [at laboratory]."

'This is an excellent program to introduce students to technology and give them a feel for [science/engineering] career fields. I view any return benefit to the government to be 'icing on the cake' and have usually benefitted."

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below (Note: basically the same as in previous years.)

A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B. Laboratory sponsor seminar presentations of work conducted by associates, and/or organized social functions for associates to collectively meet and share SRP experiences.

C. Laboratory focal points collectively suggest more AFOSR allocated associate positions, so that more people may share in the experience.

D. Associates collectively suggest higher stipends for SRP associates.

E. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same arguement was used to raise the faculty and graduate student participation time a few years ago.)

2. 1994 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 109 LFP evaluations received.

1. LFP evaluations received and associate preferences:

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

		How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	10	30	50	0	20	50	40	0	10	40	60	0	0
AL	44	34	50	6	9	54	34	12	0	56	31	12	0
FJSRL	3	33	33	33	0	67	33	0	0	33	67	0	0
PL	14	28	43	28	0	57	21	21	0	71	28	0	0
RL	3	33	67	0	0	67	0	33	0	100	0	0	0
WHMC	1	0	0	100	0	0	100	0	0	0	100	0	0
WL	46	15	61	24	0	56	30	13	0	76	17	6	0
Total	121	25%	43%	27%	4%	50%	37%	11%	1%	54%	43%	3%	0%

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

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

(Continued on next page)

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

Laboratory Focal Point Reponses to above questions

	AEDC	*AL*	*FJSRL*	*PL*	*RL*	*WHMC*	*WL*
= *Evals Recv'd*	10	32	3	14	3	1	46
Question =							
2	90 %	62 %	100 %	64 %	100 %	100 %	83 %
2a	**3.5**	**3.5**	4.7	4.4	4.0	4.0	**3.7**
2b	4.0	3.8	4.0	4.3	4.3	4.0	3.9
3	4.2	3.6	4.3	3.8	4.7	4.0	4.0
4	3.8	3.9	4.0	4.2	4.3	NO ENTRY	4.0
5a	4.1	4.4	4.7	4.9	4.3	3.0	4.6
5b	4.0	4.2	4.7	4.7	4.3	3.0	4.5
6a	**3.6**	4.1	**3.7**	4.5	4.3	**3.0**	4.1
6b	3.6	4.0	4.0	4.4	4.7	3.0	4.2
6c	3.3	4.2	4.0	4.5	4.5	3.0	4.2
7a	3.9	4.3	4.0	4.6	4.0	3.0	4.2
7b	4.1	4.3	4.3	4.6	4.7	3.0	4.3
7c	3.3	4.1	4.5	4.5	4.5	5.0	4.3
8	4.2	4.3	5.0	4.9	4.3	5.0	4.7
9	3.8	4.1	4.7	5.0	4.7	5.0	4.6
10a	4.6	4.5	5.0	4.9	4.7	5.0	4.7
10b	4.3	4.2	5.0	4.3	5.0	5.0	4.5
11	4.1	4.5	4.3	4.9	4.7	4.0	4.4
12	4.1	3.9	4.0	4.4	4.7	3.0	4.1
13a	**3.8**	**2.9**	4.0	4.0	4.7	**3.0**	**3.6**
13b	**3.8**	**2.9**	4.0	4.3	4.7	**3.0**	**3.8**
14	4.5	4.4	5.0	4.9	4.7	4.0	4.5

3. 1994 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 275 SFRP/GSRP evaluations received.

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

1. The match between the laboratories research and your field:		4.6
2. Your working relationship with your LFP:		4.8
3. Enhancement of your academic qualifications:		4.4
4. Enhancement of your research qualifications:		4.5
5. Lab readiness for you: LFP, task, plan:		4.3
6. Lab readiness for you: equipment, supplies, facilities:		4.1
7. Lab resources:		4.3
8. Lab research and administrative support:		4.5
9. Adequacy of brochure and associate handbook:		4.3
10. RDL communications with you:		4.3
11. Overall payment procedures:		**3.8**
12. Overall assessment of the SRP:		4.7
13. a. Would you apply again?	Yes:	85%
b. Will you continue this or related research?	Yes:	95%
14. Was length of your tour satisfactory?	Yes:	86%
15. Percentage of associates who engaged in:		
a. Seminar presentation:		52%
b. Technical meetings:		32%
c. Social functions:		03%
d. Other		01%

16. Percentage of associates who experienced difficulties in:

a. Finding housing: 12%
b. Check Cashing: 03%

17. Where did you stay during your SRP tour?

a. At Home: 20%
b. With Friend: 06%
c. On Local Economy: 47%
d. Base Quarters: 10%

THIS SECTION FACULTY ONLY:

18. Were graduate students working with you? Yes: 23%

19. Would you bring graduate students next year? Yes: 56%

20. Value of orientation visit:

Essential: 29%
Convenient: 20%
Not Worth Cost: 01%
Not Used: 34%

THIS SECTION GRADUATE STUDENTS ONLY:

21. Who did you work with:

University Professor: 18%
Laboratory Scientist: 54%

4. 1994 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

The summarized results listed below are from the 54 mentor evaluations received.

1. Mentor apprentice preferences:

Table B-3. Air Force Mentor Responses

		How Many Apprentices Would You Prefer To Get ?			
		HSAP Apprentices Preferred			
Laboratory	*# Evals Recv'd*	*0*	1	2	3+
AEDC	6	0	100	0	0
AL	17	29	47	6	18
PL	9	22	78	0	0
RL	4	25	75	0	0
WL	18	22	55	17	6
Total	**54**	**20%**	**71%**	**5%**	**5%**

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

2. Mentors involved in SRP apprentice application evaluation process:
   a. Time available for evaluation of applications:
   b. Adequacy of applications for selection process:
3. Laboratory's preparation for apprentice:
4. Mentor's preparation for apprentice:
5. Length of research tour:
6. Benefits of apprentice's work to U.S. Air force:
7. Enhancement of academic qualifications for apprentice:
8. Enhancement of research skills for apprentice:
9. Value of U.S. Air Force/high school links:
10. Mentor's working relationship with apprentice:
11. Expenditure of mentor's time worthwhile:
12. Quality of program literature for apprentice:
13. a. Quality of RDL's communications with mentors:
    b. Quality of RDL's communication with apprentices:
14. Overall assessment of SRP:

	AEDC	*AL*	*PL*	*RL*	*WL*
# Evals Recv'd	**6**	**17**	**9**	**4**	**18**
Question #					
2	100 %	76 %	56 %	75 %	61 %
2a	4.2	4.0	**3.1**	**3.7**	**3.5**
2b	4.0	4.5	4.0	4.0	3.8
3	4.3	**3.8**	**3.9**	**3.8**	**3.8**
4	4.5	**3.7**	**3.4**	4.2	**3.9**
5	**3.5**	4.1	**3.1**	**3.7**	**3.6**
6	4.3	3.9	4.0	4.0	4.2
7	4.0	4.4	4.3	4.2	3.9
8	4.7	4.4	4.4	4.2	4.0
9	4.7	4.2	3.7	4.5	4.0
10	4.7	4.5	4.4	4.5	4.2
11	4.8	4.3	4.0	4.5	4.1
12	4.2	4.1	4.1	4.8	3.4
13a	**3.5**	**3.9**	**3.7**	4.0	**3.1**
13b	4.0	4.1	3.4	4.0	3.5
14	4.3	4.5	3.8	4.5	4.1

5. 1994 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 116 HSAP evaluations received.

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

1. Match of lab research to you interest:		3.9
2. Apprentices working relationship with their mentor and other lab scientists:		4.6
3. Enhancement of your academic qualifications:		4.4
4. Enhancement of your research qualifications:		4.1
5. Lab readiness for you: mentor, task, work plan		**3.7**
6. Lab readiness for you: equipment supplies facilities		4.3
7. Lab resources: availability		4.3
8. Lab research and administrative support:		4.4
9. Adequacy of RDL's apprentice handbook and administrative materials:		4.0
10. Responsiveness of RDL's communications:		**3.5**
11. Overall payment procedures:		**3.3**
12. Overall assessment of SRP value to you:		4.5
13. Would you apply again next year?	Yes:	88%
14. Was length of SRP tour satisfactory?	Yes:	78%
15. Percentages of apprentices who engaged in:		
a. Seminar presentation:		48%
b. Technical meetings:		23%
c. Social functions:		18%

MODELING AND MITIGATION OF TERRAIN SCATTERED INTERFERENCE

Ismail I. Jouny
Assistant Professor
Department of Electrical Engineering

Lafayette College
Easton, PA, 18042

Final Report for:
Summer Faculty Research Program
Wright Labs

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

and

Wright Patterson Laboratory

August 1994

34-1

MODELING AND MITIGATION OF TERRAIN SCATTERED INTERFERENCE

Ismail I. Jouny
Assistant Professor
Department of Electrical Engineering
Lafayette College

Abstract

A model for terrain scattered interference (TSI) or hot clutter is developed. This model includes all relevant TSI parameters such as spatial delay, doppler, and carrier phase angle. The effects of number of array elements, number of tap delays, TSI spatial delay, doppler, tap spacing, jamming angle of arrival, number of multipath signals, doppler spread, weight fixing time, relative bandwidth, and subbanding are examined. Recommendations concerning tap spacing, number of array elements, subbanding, weight lifetime, and the use of tap delay elements are presented.

MODELING AND MITIGATION OF TERRAIN SCATTERED INTERFERENCE

Ismail I. Jouny

I. Introduction

This report addresses issues concerning terrain scattered interference (TSI) or hot clutter. TSI poses a unique adaptive signal processing problem because of mainbeam interference which cannot be cancelled by adding more angular degrees of freedom, and also because of doppler which result in a time dependent covariance matrix which leads to time varying weights and time dependent signal-to-interference plus noise ratio (SINR). Aircraft and jammer altitudes and the distance separating them are also important parameters that determine the nature of TSI It is also understood that terrain structure is another significant contributor to TSI, where rough terrain results in diffuse low power TSI and smooth terrain results in specular high power TSI. Although, the existence of diffuse and specular TSI is dependent on the type of terrain enclosed within the glistening surface, several studies have described TSI as a statistical phenomenon which can be modeled as Weibul, Rayleigh, or (the most recent) K distributed random process (see the work of Raghavan [?]). Regardless of the statistical nature of TSI and whether TSI is deterministic or random, it is known that TSI enters the mainbeam of an airborne radar in a multipath scenario which is characterized by doppler effects, disproportionate time delay, and variable intensity. What makes TSI a serious threat is that it can be modeled as a large number of correlated jammers with significant doppler disparity.

Currently, there are three sources of TSI data that are known to the author of this report; 1) TSI data generated by Westinghouse (code originally provided by MIT/LL),

2) TSI real data generated by NRL, and 3) Mountaintop data provided by MIT/LL and Rome lab. Although the collection of this data has been costly and required an enormous effort on behalf of the investigators, there are subtle differences between all three types of TSI data which imply that the statistical nature of TSI as well as its underlying bistatic reflectivity rules are not fully understood yet. Predicting the severity of TSI depends on the type of terrain, jammer and radar altitudes, and jammer elevation and azimuth view angles. A recent gathering at MIT/LL has confirmed the lack of accurate TSI modeling techniques that would help with the design of future array based radars. There are also many similarities between the three types of TSI data. All three data sources yield TSI with similar power spectrum and similar doppler effects as well as similar dependence on jammer elevation and azimuth look angles. This study uses similar TSI parameters as those implied by these data sources. Other TSI parameters that are not common to all data sources are simply chosen according to Mountaintop database which is described in some detail in the MIT/LL workshop on Adaptive Processing digest.

A background on TSI mitigation is presented in the following section. In Section III we formulate the TSI problem where we include all TSI relevant parameters. Section IV describes the results obtained, and Section V presents some conclusions and recommendations concerning TSI mitigation.

II. Background

TSI mitigation work summarized in this section describes avenues of research followed by several array processing clusters in the U.S. and is an indication of the enormous recent research efforts that have focused on TSI cancellation. These activities are being funded by all sectors of the United States Armed Forces. The state of

the art TSI mitigation technology requires hundreds and even thousands of degrees of freedom (DOF) particularly when doppler compensation is being employed. Therefore, a TSI mitigation system will be extremely costly, highly complex, and awaiting the next generation of semiconductor devices. It is however, extremely important to understand the significance of each of the factors that contribute to TSI and how array design can be optimized with respect of each of these parameters. For example, there are studies that suggest that a faster and more frequent weight updating algorithm (assuming the SMI algorithm is used) will reduce the number of DOF sharply. In the following we begin by describing some of the earlier TSI mitigation algorithms and end with the latest that was presented at MIT/LL in March 1994.

Earlier efforts on TSI mitigation such as the work of Brennan et al [9] have described TSI as a linear combination of weighted and delayed replicas of a jammer waveform. The obvious solution is to reverse TSI formation process by employing a tap delay line filter (a Finite Impulse Response Filter) whose coefficients are chosen according to an adaptive algorithm. The number of delay elements needed has to match the number of multipath components and their perspective delays. Clearly, such a model does not account for doppler effects and assumes a moderate number of multipath components. In [16] it is shown that the spectral shape of the radar receiver may influence TSI mitigation and that the most desirable receiver spectrum is shaped as raised cosine. The study in [16] included Gaussian and rectangular spectral shaped receivers. Monzingo and Miller [17] provide an interesting TSI example and show that the cancellation ratio of a tap delay line canceller is very dependent on the bandwidth tap delay product ($B\Delta$) as well as the number of taps.

The idea of using tap delay elements was further exploited in a study by Mayhan et al [8] where the choice of uniform tap spacing was described as reasonable and where it was determined that the shape of the antenna array has some impact on the number of taps needed. Mayhan's study addresses the problem of wideband

and examined the effect of eclipsing on the performance of the adaptive processor. All of the above studies do not consider the effect of doppler modulation on the performance of the TSI cancellation system.

Jao from MIT/LL [31] devised a TSI model based on Geometric Optics and Rice's perturbation rule. Preliminary results in [31] indicate a reasonable level of TSI predictability under certain frequency, altitude and view angle scenarios. Coutts (also from MIT/LL) [22] used the Mountaintop database to demonstrate 20 db cancellation of TSI using tap delay filters. The study in [22] uses both pulsed CW data and pseudo-random noise. With the pulsed CW data a 20 db cancellation using 200 tap delay elements was achieved. Coutts also demonstrated the effect of doppler on the weights lifetime where it was shown that fixed weights yield optimal results for three ms. The results in [22] also show that a cancellation ratio of 30 db would require up to 800 taps and that the spacing between taps must about half the reciprocal of the jammer bandwidth. The PN approach produced higher cancellation ratio but required more frequent weight updating than the pulsed CW approach. Miller and Ortiz from Hughes [23] used several array scenarios with doppler compensation for TSI mitigation. Miller's study used the conventional sidelobe canceller, a canceller with high gain beam towards jammer, a beam space adaptive array, and an element space adaptive array, all with tap delay line filters and doppler compensation. Although these methods differ in the number of tap delays needed, what is surprising is the amount of information about TSI needed beforehand, and the number of degrees of freedom required particularly when doppler compensation is employed. More importantly, the number of compensation channels needed depends on the RF frequency of the radar, the velocity , and the weight update frequency. Miller's study did not consider subbanding but did demonstrate that desert terrain TSI can be cancelled effectively using any of the four techniques mentioned above. The numerical examples provided in [23] excluded any doppler shift. Basham from SAIC [24] de-

vised an approach that uses multiple adjacent beams to sample TSI. Basham used a 3-Dimensional weight vector (range/doppler/time) in a space time adaptive processing framework. Although Basham's approach minimizes the eclipsing problem, but his technique is a computationally demanding. Brovko et al [25] from Hughes also studied TSI mitigation capabilities of various array shapes using tap delay filters. Harrison from Lockheed [20] recommended that the Principal Component Inverse (PCI) method proposed by Tufts is better suited to mainbeam jammer nulling than the SMI algorithm and it requires less number of samples. Finally, Compton in a study involving an array with one auxiliary and no tap delays [26] examined the problem of weight aging because of doppler. Compton's results agree with Coutts's 3 ms weight lifetime and indicate that maintaining the signal to interference plus noise ratio within 3 db of its maximum value is insensitive to jammer angular spread, the number of multipath components, and the number of auxiliary elements. Weight lifetime is, however, sensitive to doppler spread of multipath jamming [26].

III. Problem Formulation

A TSI model that includes all pertinent parameters of multipath interference and doppler effects is developed in this section. An adaptive array with M-1 auxiliaries and K tap delay elements with or without subbanding is used to cancel TSI. The signal-to- interference plus noise ratio of the array and the cancellation ratio are used interchangeably as performance measures of the array nulling capability. Let the vector

$$X = [X_1|X_2|\ldots|X_M]^T \tag{1}$$

denote the signal received by the array where

$$X_m = [X_{m1}, X_{m2}, \ldots, X_{MK}]^T . \tag{2}$$

"T" denotes the transpose and $X_{mk} = x_{mk}(t) = x_m(t-(k-1)\Delta)$ which is the signal received by the m^{th} array element and delayed by $k\Delta$, where Δ is the tap spacing in seconds. Clearly,

$$x_{mk}(t) = d_{mk}(t) + i_{mk}(t) + n_{mk}(t) \tag{3}$$

where d, i and n denote the desired signal, the interference, and additive thermal noise. Assuming that the desired and interference signals arrive at angles θ_d and θ_i respectively then

$$\begin{aligned} d_{mk}(t) &= d(t-(k-1)\Delta-(m-1)T_d) \\ i_{mk}(t) &= i(t-(k-1)\Delta-(m-1)T_i) \\ n_{mk}(t) &= n_{m1}(t-(k-1)\Delta) \end{aligned} \tag{4}$$

where

$$T_d = \frac{L}{c}\sin(\theta_d) \tag{5}$$

$$T_i = \frac{L}{c}\sin\theta_i \tag{6}$$

"L" being the separation between array elements chosen as $L = \frac{\lambda}{2}$ and c is the speed of light. The weight vector needed is

$$W = [W_1|W_2|\ldots|W_M] \tag{7}$$

where $W_m = [w_{m1}, w_{m2}, \ldots, w_{mK}]$ represents the weight vector for each array element. It is well known that

$$W = \Phi^{-1}S \tag{8}$$

where $\Phi = E\{X^*X^T\}$ is the covariance matrix and $S = E\{X^*d(t)\}$ is the steering vector. The covariance matrix Φ can be expressed as

$$\Phi = \Phi_d + \Phi_i + \Phi_n \tag{9}$$

which represents the sum of the covariance matrices of desired signal, interference, and noise components. It is assumed that the interference, desired signal, and noise all have rectangular spectral densities centered at the radar operating frequency ω_0 and with spread $\Delta\omega$ so that $\Delta\omega/\omega_0$ represent the relative bandwidth B of all three signals. Some of the useful parameters needed in this study are $\Phi_d = \omega_0 T_d$ which represents the interelement phase shift, and $\Phi_i = \omega_0 T_i$. Let Δ_{90} be the time delay required to produce a 90 degrees phase shift $\Delta_{90} = \frac{\pi}{2\omega_0}$, and r is chosen such that

$$r = \frac{\Delta}{\Delta_{90}} \tag{10}$$

Therefore increasing r corresponds to increasing the spacing between array elements. The structure of the covariance matrix is

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \dots & \Phi_{1M} \\ \Phi_{21} & \Phi_{22} & \dots & \Phi_{2M} \\ \dots & \dots & \dots & \dots \\ \Phi_{M1} & \dots & \dots & \Phi_{MM} \end{bmatrix} \tag{11}$$

where $\Phi_{mn} = E\{X_m^* X_n^T\}$ and $\Phi_{mn}(j,k) = E\{X_{mj}^* X_{nk}\} = \rho((j-k)\Delta + (m-n)T_i)$ where $\rho(\tau)$ is the autocorrelation of each of the three signals (desired, interference, and noise) defined as $\rho(\tau) = P sinc(\Delta\omega\tau/2)\exp j\omega_0\tau$ where P is the power of the associated signal. Therefore, the covariance matrix of the desired signal can be constructed as

$$\Phi_{mn}(p,q) = P_d sinc\left[\Delta\omega_0\left((p-q)\Delta + (m-n)T_d\right)/2\right]\exp j\omega_0\left((p-q)\Delta + (m-n)T_d\right) \tag{12}$$

The covariance matrix for the noise component is constructed as

$$\Phi_{mn}(p,q) = P_n sinc\left[\Delta\omega_0(p-q)/2\right]\exp j\omega_0(p-q) \tag{13}$$

Assuming that the desired signal and the interference are uncorrelated, the steering vector reduces to $S = E\{X_d d(t)\}$,

$$S = [S_{11}, \ldots, S_{1K}, \ldots, S_{MK}]^T \tag{14}$$

where

$$S_{mk} = \sqrt{P_d} sinc\left[\Delta\omega_0 \left((k-1)\Delta + (m-1)T_d\right)\right] \exp j\omega_0 \left((k-1)\Delta + (m-1)T_d\right) \tag{15}$$

The output of the array can be expressed as $s(t) = s_d(t) + s_i(t) + s_n(t)$ thus representing all three signal components, where $s(t) = W^T X$. Therefore the power of the desired output signal is $\frac{1}{2} W^{*T} \Phi_d W$ (see [10]). Therefore the signal to interference plus noise ratio of the array is

$$SINR = \frac{W^{*T} \Phi_d W}{W^{*T} (\Phi_i + \Phi_n) W} \tag{16}$$

Another measure of performance for the above adaptive architecture is the cancellation ratio. The cancellation ratio CR is defined as

$$CR = \frac{\sum_{ks=0}^{KS} E\{|r_{ks}|^2\}}{\sum_{ks=0}^{KS} E\{|X_1^{ks}|^2\}} \tag{17}$$

where r_{ks} is the residue in the ks subband assuming a total of KS subbands. The ks subband of the main channel signal is denoted by X_1^{ks}. The residue power needed to compute CR is

$$E\{|r_{ks}|^2\} = E\{|X_1^{ks}|^2\} - Z_{ks}^{*T} R_{ks}^{-1} Z_{ks} \tag{18}$$

where $Z_{ks} = E\{A_{ks} X_1^{ks}\}$ and $A_{ks} = \left[X_{21}^{ks}, X_{22}^{ks}, \ldots, X_{2K}^{ks}, X_{31}^{ks}, \ldots, X_{MK}^{ks}\right]$, and $R_{ks} = E\{A_{ks}^* A_{ks}^T\}$ which can be generated from Φ_i by deleting the first row and the first column. If R is used to compute the weights in the auxiliaries then

$$W_{ks} = R_{ks}^{-1} Z_{ks} \tag{19}$$

The weights in the main channel in this case are assumed $W = [1, 0, 0, \ldots, 0]$.

The multipath scenario considered in this study is defined such that the desired signal arrives at the broadside $\theta_d = 0$, and a correlated group of interference arrives at angles θ_l, $l = 1 : LM$ where LM is the total number of multipath components. The interference signal arriving at m^{th} element of the array and the k^{th} tap delay is

$$i_{mk}(t) = \sum_{l=0}^{LM} \beta_l i(t - mT_{il} - (k-1)\Delta - T_{sl}) \exp j\left[\omega_l\left(t - mT_{il} - (k-1)\Delta\right) + \psi_l\right] \quad (20)$$

where β_l denotes the amplitude of the l^{th} multipath component with $\beta_0 = 1$ which is associated with the direct interference signal. For specular multipath $LM = 1$ and $\beta_1 \neq 0$ and $\beta_l = 0$, $l = 2, \ldots, LM$. For diffuse multipath $\beta_l \neq 0$ for $l = 2, \ldots, LM$. Statistically, β_l can be chosen according to a Rayleigh or K distribution, but it may also be chosen as depicted by Beckmann and Spizzichino [30] according to the geometry and standard deviation of the roughness of the glistening surface. Beckmann's model was used in [13] as a tool for generating TSI. According to such a TSI model, diffuse multipath jamming arrives from near the radar or the jammer, while the specular multipath is reflected off the center. Signals reflected off other areas are described as moderately diffuse. The altitudes of the radar and the jammer are important factors that contribute to the size of the moderately diffuse region. The parameter ψ_l represents the carrier phase angle of the l^{th} multipath component with $\psi_0 = 0$. The values of ψ_l must be carefully chosen because they may effect the structure of the covariance matrix and may significantly reduce the effect of doppler shift. For example in [13], ψ_l, $l = 1, \ldots, LM$ are assumed uniformly distributed and uncorrelated random variables over $[0, 2\pi]$. In [26], ψ_l are assumed zero (deterministic). To avoid diluting the doppler problem, we have assumed that all values of ψ_l are identically zero. The parameter T_{il} denotes the interelement phase shift where

$$T_{il} = \frac{L}{c} \sin \theta_l \quad (21)$$

with $T_{i0} = 0$. T_{sl} is the spatial delay of the l^{th} multipath. The value of T_{sl} depends on the geometry of the glistening surface and may range from 1 to 1.6 times the spatial delay of the specular interference component. The doppler shift of the l^{th} multipath component is denoted by ω_l

$$\omega_l = \frac{-2\nu}{c}\omega_0 \tag{22}$$

where ν is the relative velocity of the jammer with respect to the radar and ω_0 is the radar RF frequency. The values of ω_l (except $\omega_0 = 0$) may range from about 50 Hz to few KHz. The doppler spectrum of real data indicates that high powered TSI has a doppler frequency in a 100 Hz range. The above model is consistent with the models used in [26, 13, 11, 23] although none of these referenced models incorporate all pertinent TSI parameters such as spatial delay, doppler frequencies, and tap delays. The covariance matrix of the interference in the ks subband can now be expressed as

$$\Phi_{mn}^{ks}(p,q) = E\{I^{ks}(m, t - p\Delta)I^{ks}(n, t - q\Delta)\} \tag{23}$$

where $I^{ks}(m, t - p\delta)$ is a sample of interference at the m^{th} array element and behind the $(p-1)^{th}$ tap so that $m, n = 1, \ldots, M$ and $p, q = 1, \ldots, K$. It can be shown that the above covariance matrix element is

$$\begin{aligned}
&\Phi_{mn}^{ks}(p,q) = \\
&\sum_{l=0}^{LM}\sum_{r=0}^{LM} \beta_l\beta_r\rho\left[(m-1)T_{il} - (n-1)T_{ir} + (p-q)\Delta + T_{sl} - T_{sr}\right] \\
&expj\left[(\omega_l - \omega_r)t - \omega_r\left((n-1)T_{ir} + (q-1)\Delta\right) + \omega_l\left((m-1)T_{il} + (p-1)\Delta\right)\right] \\
&\exp j\left[2\pi(ks)\Delta\omega_{ks}\left((m-1)T_{il} - (n-1)T_{ir} + (p-q)\Delta\right)\right]
\end{aligned} \tag{24}$$

where $\Delta\omega_{ks}$ is the bandwidth of the ks subband. The last exponential term in the above expression is an adjustment that is needed because of subbanding and may be applied to the expressions of Φ_d and Φ_n. The function $\rho(x)$ is simply

$\sin(\Delta\omega_{ks}x/2)/(x/2)$. Similarly the vector Z_{ks} in the cancellation ratio expression can be found as

$$Z_{ks}^{m}(p) = E\{I_{m}^{*ks}(t,(p-1)\Delta)\, I_{1}^{ks}(t,0)\} \tag{25}$$

which corresponds to the m^{th} array element and the p^{th} tap delay with $m = 2,\ldots,M$ and $p = 1,\ldots,K$. It can be shown that

$$\begin{aligned} &Z_{ks}^{m}(p) = \\ &\sum_{r=0}^{LM}\sum_{l=0}^{LM} \beta_l\beta_r\rho\left[(m-1)T_{il} + (p-1)\Delta + T_{sl} - T_{sr}\right] \\ &\exp j\left[\omega_l(t-(m-1)T_{il}-(p-1)\Delta) + 2\pi(ks)\Delta\omega_{ks}\left((m-1)T_{il}-(p-1)\Delta\right)\right] \end{aligned} \tag{26}$$

The above expressions can be written in various forms depending on the programming approach used. For example the above expressions can be rewritten in terms of the interelement phase shift $\phi_l = \omega_0 T_{il}$ and so forth. The above expressions of Φ_i and Z can be used to determine SINR and cancellation ratio CR. The covariance matrix Φ is such that $\Phi = \Phi_d + \Phi_i + \Phi_n$. In determining the cancellation ratio, the desired signal is often ignored because it bears little significance on the covariance matrix in a high power jamming environments.

Two theoretical remarks are appropriate at this point:

A. Impulse Response Model

The transfer function of the m^{th} element of the adaptive array used to mitigate TSI is

$$H_m(\omega) = \sum_{k=1}^{K} w_{mk}\exp\left(-j\left[\omega(k-1)\Delta\right]\right) \tag{27}$$

Therefore, the transfer function of the overall array is

$$H(\omega) \;=\; \sum_{m=1}^{M} H_m(\omega)\exp\left(-j\left[\omega(m-1)T_i\right]\right) \tag{28}$$

$$= \sum_{m=1}^{M}\sum_{k=1}^{K} w_{mk} \exp\left(-j\left[\omega((k-1)\Delta + (m-1)T_i\right]\right)$$

Hence, the impulse response of the adaptive antenna array is

$$h(t) = \sum_{m=1}^{M}\sum_{k=1}^{K} w_{mk}\delta(t-(m-1)T_i-(k-1)\Delta)) \tag{29}$$

Assuming that the jamming waveforms incident on the array can be modeled as

$$I(t) = \sum_{l=0}^{LM} \beta_l I_0(t-T_{sl}) \exp\left(j\left[\omega_l t + \psi_l\right]\right) \tag{30}$$

where $I_0(t)$ is the actual signal transmitted by the jammer. The output component of the array due to jamming only is the convolution of $I(t)$ with $h(t)$ which is

$$\begin{aligned} s_i(t) \;=\; & \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{l=0}^{LM} \beta_l w_{mk} I_0\left[t-T_{sl}-(k-1)\Delta-(m-1)T_{il}\right] \\ & \exp\left(j\omega_l\left[t-(k-1)\Delta-(m-1)T_{il}\right]\right) \end{aligned} \tag{31}$$

The above equation can be solved for the optimal weights w_{mk} (set $s_i(t)=0$) or using an adaptive processor. The significance of the above equation is that the number of degrees of freedom needed to mitigate TSI (which is equal to $K \times M$) depends on the number of multipath components LM, and the doppler frequencies ω_l particularly the doppler spread. Clearly, some of the weights may be insignificant depending on whether the coefficients β_l are relatively large. Notice that the number of array elements M plays a limited role in the above equation since the parameters ω_l and T_{sl} are controllable by the number of taps used K.

B. Doppler Spread

By further examining the above equation which, if solved, yields the optimal set of weights needed to completely mitigate TSI, we recognize that the contribution of the doppler frequency ω_l to the overall array response is due to the expression

$\sum_{l=0}^{LM} \exp(-j\omega_l t)$. If we assume that all multipath components have doppler frequencies centered around ω_{l0} and increase by an increment of $\Delta\omega_{l0}$, it can be shown that the doppler contribution is equivalent to $\exp(j\omega_{l0}t/2)\frac{\sin(LM)\Delta\omega_l t/2}{sin\Delta\omega_{l0}t/2}$. Thus, the contribution of doppler is equivalent to a sinc function whose first zero occurs about $\Delta\omega_{l0}(LM)t/2 = \pi$ which corresponds to $t = 2\pi/(\Delta\omega_{l0}(LM)) = 1/f_s$ where f_s is the total doppler spread ($f_s = (LM)\Delta\omega/(2\pi)$). Therefore, a proper weight fixing time that minimizes the effect of doppler modulation is $t \approx 1/f_s$. In [26], it was found experimentally that SINR remains within 3 db of its optimal value for about $2.3/f_s$ seconds (i.e. about $1.15/f_s$ seconds before or after the time optimal SINR was achieved). Our approximation above agrees with the results reported in [26, 22]. We therefore recommend, based on doppler features extracted from Mountaintop Data base, a weight fixing time of about 1 to 3 ms.

IV. Conclusions

A TSI model that incorporates all pertinent scattering parameters was developed in this study. The model used incorporates multipath spatial delay, doppler shift, tap delays, and diffuse and specular multipath. Extensive simulation was carried to examine the significance of each of these parameters in an adaptive architecture framework. The number of auxiliaries, the relative bandwidth, doppler spread, jammer arrival angle, and multipath spatial delay were all variables that affected the signal to interference plus noise ratio as well as the cancellation ratio. The results of this study can be summarized as follows:

- An array with three auxiliary elements (a total of 4 channels) provides sufficient TSI cancellation. Increasing the number of auxiliaries beyond 3 is of limited reward unless the jammer bandwidth increases above 5%.

- In this study, it is assumed that all multipath components arrive at an angle $\theta_i \pm n\Delta\theta_i,\ n = 0, 1, \ldots$. All simulations indicate that neither θ_i nor $\Delta\theta_i$ have a significant impact on the performance of the array given that the arrival angle of the desired signal is different from that of the closest multipath component by about 5 degrees or more.

- In this study, it is assumed that all multipath components have a doppler frequency $\omega_{l0} \pm n\Delta\omega_{l0}, n = 0, 1, \ldots$. The central frequency ω_{l0} has little effect on the performance of the array for a weight lifetime of 10 ms. The doppler spread $(LM)\Delta\omega_{l0}$ is a significant factor that determines the weights lifetime. However, if the relative bandwidth of the jammer increases above 5% the doppler spread is an insignificant factor compared to the impact of wideband jamming, and doppler compensation would be of little advantage in this case.

- It is recommended that all array weights be updated at a rate of $\approx 1/f_s$ where f_s is the doppler spread. If the doppler parameters of a real TSI data base are used, the array weights age in about 3 to $4ms$.

- Antenna arrays with nonuniform tap spacing outperform those that are constructed of uniformly spaced delay elements. Nonuniform tap spacing has no effect on weight aging. If uniform tap spacing is used, a time delay $\Delta \approx 1/(2\Delta\omega)$ is recommended where $\Delta\omega$ is the jammer bandwidth.

- The number of taps used is the most important factor in TSI mitigation particularly when the jammer relative bandwidth is above 5%. Increasing the number of taps beyond a certain level is not rewarding. This level was 16 taps in our study where the number of multipath components was limited to 9.

- The spatial delay of multipath components has little effect on either SINR or CR.

- As the power and bandwidth of a jammer increases, TSI becomes a more serious problem where both the number of auxiliaries and tap delay filters must be increased.

- The performance of an adaptive array decays slowly as the number of multipath components increases. In our study any increase beyond 7 was of little significance.

- Subbanding and tap delay filtering in a hybrid structure is recommended. Unlike the number of tap delay elements, The number of subbands has to be increased significantly for any noticeable improvement in the array performance. Nonetheless, subbanding followed by tap delay filtering yields near optimal performance particularly when TSI is due to a wideband jammer. Nonuniform subbanding using wavelet decomposition did not outperform subbanding via a DFT, but did reduce the rate of fluctuation of CR as a function of time.

An antenna array with three auxiliary elements, uniform subbanding, and a sufficient number of nonuniformly spaced tap delays is recommended for TSI mitigation. It is the author's conclusion that wideband jamming with or without TSI presents a more serious threat than narrow band jamming with TSI. Wideband jamming has been traditionally mitigated using subbanding and tap delay filtering. Doppler compensation would not be needed, thus reducing the number of degrees of freedom needed, if the array weights are updated every 3 to 4 ms.

Acknowledgement

The author wishes to thank RDL and Wright Patterson AFB for their support and for introducing him to the TSI problem. The author particularly acknowledges

the support of WL/AARM3. An expanded copy of this report which contains more details can be obtained from Wright Patterson AFB WL/AARM3, Dayton, OH.

References

[1] W. D. White, "Wideband Interference Cancellation in Adaptive Sidelobe Cancellers," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 19, No. 6, pp 915-924, November, 1993.

[2] F. W. Vook, R. T. Compton, "Bandwidth Performance of Linear Adaptive Arrays with Tapped Delay-Line Processing," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 3, pp 901-908, July 1992.

[3] W. F. Gabriel, "Adaptive digital processing investigation of DFT subbanding vs transversal filter canceler," *Naval Research Laboratory technical report*, NRL report 8981, July 1986.

[4] W. F. Gabriel, "Adaptive Processing Array Systems," *Proceedings of The IEEE*, Vol. 80, No. 1, pp 152-162, January 1992.

[5] J. T. Mayhan, A. J. Simmons, and W. C. Cummings, "Wide-Band Adaptive Nulling Using Tapped Delay Lines," *IEEE Transactions on Antennas and Propagation*, Vol. 29, No. 6, pp 923-936, November 1981.

[6] L. E. Brennan and I. S. Reed, "Adaptive Cancellation of Scattered Interference," *Adaptive Sensors, Inc.*, final report, December 1982.

[7] R. T. Compton, "The bandwidth performance of a two-element adaptive array tapped delay-line processing," *IEEE Transactions on Antennas and Propagation*, Vol. 36, No. 1, pp 5-13, January 1988.

[8] D. R. Morgan and A. Aridgides, "Adaptive Sidelobe Cancellation of Wide-Band Multipath Interference," *IEEE transactions on Antennas and Propagation*, Vol. 33, No. 8, pp 908-917, August 1985.

[9] R. T. Compton, "The Relationship Between Tapped Delay-Line and FFT Processing in Adaptive Arrays," *IEEE Transactions on Antennas and Propagation*, Vol. 36, No. 1, pp 15-26, January 1988.

[10] R. L. Fante, "Cancellation of Specular and Diffuse Jammer Multipath Using a Hybrid Array," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 823-836, Vol. 27, No. 5, September, 1991.

[11] D. R. Morgan and A. Aridgides, "Comments on; Cancellation of Specular and Diffuse Jammer Multipath Using a Hybrid Adaptive Array," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 30, No. 3, pp. 932-933, July, 1994.

[12] S. Applebaum, "Adaptive Arrays", *IEEE Transactions on Antennas and Propagation*, pp. 585-598, Vol. 585, No. 5, September 1976.

[13] L. E. Brennan, J. D. Mallett and I. S. Reed, "Adaptive Arrays in Airborne MTI Radar," *IEEE Transactions on Antennas and Propagation*, pp. 607-615, Vol. 24, No. 5, September 1976.

[14] R. A. Monzingo and T. W. Miller, "Introduction to Adaptive Arrays," New York: Wiley, 1980.

[15] R. T. Compton, "Adaptive Antennas," Englewood Cliffs, NJ: Prentice Hall, 1988.

[16] K. Gerlach, "A Numerically Efficient Band-Partitioned Noise Canceller," NRL Report #9050, 1987.

[17] W. Harrison and D. Tufts, "Rapidly Adaptive Mainbeam Jamming Nulling," *Proceedings of ASAP Workshop*, MIT/LL, March 1994.

[18] J. K. Jao, "Application of Two-Scale Surface Scattering to Bistatic Clutter Modeling", *Proceedings of ASAP Workshop*, MIT/LL, pp. 427-447, March 1994.

[19] S. D. Coutts, "Mountaintop Jammer Multipath Mitigation Experiment," *Proceedings of ASAP Workshop*, MIT/LL, pp. 595-625, March 1994.

[20] T. W. Miller and J. Ortiz, "Jammer Multipath Phenomenology and Mitigation," *Proceedings of ASAP Workshop*, MIT/LL, pp. 665-683, March 1994.

[21] C. L. Basham, "Mitigation of Multipath Jamming by Space- Time Adaptive Processing," *Proceedings of ASAP Workshop*, MIT/LL, pp. 665-683, March 1994.

[22] O. Brovko, T. T. Nguyen, and B. J. Heiman, "High PRF TSI Mitigation in Fighter Radars," *Proceeding of ASAP Workshop*, MIT/LL, pp. 683-714, March 1994.

[23] R. T. Compton, "The Effects of Multipath Jamming on an Adaptive Array," *Proceedings of ASAP Workshop*, MIT/LL, pp. 833-861, March 1994.

[24] G. Beylkin, R. Coifman, and V. Rokhlin, *Communications on Pure and Applied Mathematics*, Vol. 44, pp. 141-183.

[25] B. Z. Steinberg and Y. Leviatan, "On the Use of Wavelet Expansion in the Method of Moments," *IEEE Transactions on Antennas and Propagation*, Vol. 41, pp. 610-619, May 1993.

[26] "Adaptive Processing Radar Study", Final Report, Westinghouse Electric Corporation, Electronic Systems Group, Baltimore, MD, 1994.

[27] P. Beckmann and A. Spizzichino, "The Scattering of Electromagnetic Waves From Rough Surfaces," Dedham, MA: Artech House, 1987.

[28] R. S. Raghavan, "A Model for Spatially Correlated Radar Clutter," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 38, No. 7, pp. 268-275, March 1991.

LAMB WAVE SCANNING OF A MULTILAYERED COMPOSITE PLATE SPECIMEN

Tribikram Kundu
Professor
Department of Civil Engineering
and Engineering Mechanics

University of Arizona
Tucson, AZ 85721

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

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

and

Wright-Patterson Laboratory

July 1994

LAMB WAVE SCANNING OF A MULTILAYERED COMPOSITE PLATE SPECIMEN

Tribikram Kundu
Professor
Department of Civil Engineering
and Engineering Mechanics
University of Arizona

Abstract

In this research a new scanning technique using leaky Lamb waves is developed and applied to detect internal defects in a multilayered composite plate specimen. Images generated by this new Lamb wave scanning technique (L-scan) are compared with the conventional C-scan images. This comparison shows that the L-scan technique is more effective for detecting some internal defects such as fiber breakage and matrix burned out zones in a multilayered specimen than the C-scan technique.

LAMB WAVE SCANNING OF A MULTILAYERED COMPOSITE PLATE SPECIMEN

Tribikram Kundu

Introduction

Ultrasonic technique has become one of the most popular nondestructive testing technique because of its versatility and ease of operation. It can detect internal cracks and inclusion type defects in homogeneous or layered materials most of the times without much difficulty. However, it has its own shortcomings. It is not very effective in detecting cracks which are vertical to the plate surface. This is because the ultrasonic signal is not reflected by the crack when the signal propagation direction is parallel to the crack surface. The back scattering technique and acoustic microscopy are used for detecting such vertical defects. However, when these defects are not located very close to the surface these techniques also encounter difficulties. Use of Lamb waves to detect such defects may be a viable alternative to the currently practiced methods. Theoretical studies by Kundu and Blodgett {1}, Yang and Kundu {2,3} and Yang {4} have shown that different Lamb wave modes produce different levels of excitation in various layers in a multilayered solid plate. This phenomenon is exploited here to experimentally generate C-scan images of different layers of a composite plate by propagating Lamb waves of different modes through the plate. The C-scan image generated by receiving the leaky Lamb waves is denoted as the "L-scan" image in this paper.

Previous efforts of using leaky waves to inspect defects in composite and metal plates include the works of Chimenti and Nayfeh {5}, Nagy et al. {6}, Pearson and Murri {7}, Rose et al. {8}, Nayfeh {9}, Bar-Cohen and Chimenti {10}, Chimenti and Bar-Cohen {11}, Martin and Chimenti {12}, Mal and Bar-Cohen {13}, and Chimenti and Martin {14} among others. Most of these works involve relating the material defects such as the porosity and delamination to the change in the Lamb wave propagation characteristics, the dispersion curves, phase velocity and attenuation. Not many investigators attempted to scan the specimen using Lamb waves to generate the L-scan image. Chimenti and Martin {14} did it with some success. They positioned the transmitter and the reflector in a pitch-catch arrangement over the specimen to generate leaky Lamb waves. They placed the receiver in the null zone and scanned the specimen.

The null zone position changes in presence of an internal defect. Hence, when a defect is encountered the receiver voltage amplitude is altered and the image of the defect is generated. The major problem with this arrangement is that the null zone position is very sensitive to the plate thickness. Hence, a few percent change in the plate thickness alters the receiver voltage amplitude significantly. To avoid this problem one needs to filter the L-scan generated data through a special filter. This signal processing helps to minimize the effect of the plate thickness variation on the null zone but retains the sensitivity to defects of interest.

In this research this problem is avoided by placing the receiver beyond the null zone as well as the specularly reflected zone. Thus only propagating leaky Lamb waves are received by the receiver; its amplitude is comparatively less sensitive to the plate thickness and more sensitive to the defects inside the plate.

Methodology

The first step of generating an L-scan image is to produce propagating Lamb waves in the specimen. For this purpose two transducers are positioned in a pitch-catch arrangement over the plate specimen as shown in Figure 1. The transmitter is excited by the tone burst excitation. The excitation frequency is then varied continuously from a minimum value to a maximum value within the band width of the transducers. The reflected signal is received by the receiver, the signal amplitude is then displayed on an oscilloscope screen as a function of the frequency. When the transmitters and the reflector are positioned such that the reflected energy is maximum then the reflected amplitude spectrum can have shapes as shown in Figures 1a or 1b. If no Lamb waves are generated then the spectrum looks like the plot shown in Figure 1a. However, if leaky Lamb waves are generated at some frequency then at that frequency a dip is observed, as shown in Figure 1b, due to the generation of leaky Lamb waves. If the transducers are moved further down, i.e. the specimen-transducer distance is reduced without altering the distance between the two transducers then, because of defocusing, the reflected amplitude spectrum changes its shape and magnitude. In this defocused position peaks are observed at the frequencies producing the Lamb waves as shown in Figure 1c. This is because when the defocusing is high the specularly reflected beam

cannot reach the receiver but the leaky waves can.

The frequency corresponding to the Lamb wave modes can be obtained from the dips (Figure 1b) or peaks (Figure 1c) of the reflected signal spectrum. The Lamb wave speed or the phase velocity can be obtained from the Snell's Law,

$$C_L = \frac{\alpha_o}{\sin(\theta)} \qquad (1)$$

where C_L is the Lamb wave phase velocity, α_o is the longitudinal wave speed in the coupling fluid (for water it is equal to 1.49 km/sec) and θ is the angle of inclination of the transducers, i.e. the angle between the vertical axis and the transducer axis.

If the transducer angle is changed, corresponding Lamb wave speed is changed, hence the dip positions along the frequency axis vary. Thus the Lamb wave dispersion curves can be experimentally generated by monitoring the transducer angles and dips of the reflected signal spectra. After selecting a specific Lamb wave mode for scanning, that mode is generated by first setting the transducer angle (θ), then proper defocusing is done by vertically moving the transducers and observing the Lamb wave peaks and finally the frequency is set at a value corresponding to a Lamb wave peak of interest. The specimen is then scanned with this transmitter-receiver arrangement.

Experimental Results

A twelve layer ceramic composite plate of dimension 8.5 inch (215.9 mm) x 3 inch (76.2 mm) x 0.094 inch (2.4 mm) is used as the specimen for the Lamb wave scanning. The NICALON fibers in the MAS-5 (magnesium-aluminum-silicate glass) matrix are oriented in 0° and 90° directions in alternate layers and is symmetric about its central plane. The specimen was made by the CORNING company and it has some internal flaws as shown in Figure 2. Top two layers do not contain any flaw. The third layer has a fold in the fibers at a distance of 2.5 inch from the left edge. This fold may cause some fibers to break at this location. This layer also has a fiber concentrated zone or "die plug" in between the central line and the bottom edge. The fourth layer is flaw less. The fifth layer has a "die plug" in the same position as the third layer and two fiber folds at 1 inch and 2.5 inch distances from the left edge. The sixth layer has a 2 inch x 3 inch zone along the right edge where the

matrix has been burned out at 550°C. The specimen is symmetric about its central plane which is the interface between the sixth and seventh layers. Fiber orientations and defect locations in each of the twelve layers of the composite plate are shown in Figure 2.

Figure 3 shows the experimentally obtained dispersion curves of the specimen in three different directions. Square (□), cross (+) and diamond (◇) symbols correspond to 0°, 90° and 45° directions. Top and bottom figures show the same curves in two different frequency ranges - 0 to 5 MHz for the top figure and 0 to 15 MHz for the bottom figure. From these figures one can clearly see that, as expected, at low frequencies (less than 2 MHz) there is no difference in the dispersion curves in three directions. However, as the frequency increases the anisotropy becomes more prominent and differences in the dispersion curves become more noticeable.

The ultrasonic images, generated by scanning the specimen by first and second Lamb wave modes, are shown in Figure 4. The edge to edge distance (W in Figure 2) between the two transducers was kept constant at 20 mm for all scans. The transducer-reflector distance was varied until strong peaks were observed at the critical frequencies in the reflected signal spectrum as shown in Figure 1c. The transducer angle was set at 17.5° (it corresponds to a phase velocity of 4.96 km/sec, from eqn.1) and frequency values for the two modes of scanning are set at 1.92 MHz (mode 1) and 3.50 MHz (mode 2). Two black squares in the dispersion curve plot of Figure 4 indicate the frequency - phase velocity combinations that produced the two images shown in Figure 4. It is interesting to note that in the top image, that corresponds to mode 1 scanning, the matrix burned out zone can be clearly seen on the right side but other defects such as fiber folds and die plugs are invisible. In the bottom image (mode 2 image) on the other hand the matrix burned zone is not clearly visible but several other defects in the central region of the plate can be clearly seen. The defect marked by the arrow corresponds to the fiber fold defect in the third layer. Hence, the energy associated with the first mode propagated mostly through sixth and seventh layers and that for the second mode propagated through the top three or four layers.

A slight change in the frequency of excitation of mode 1 from 1.92 MHz to 1.72 MHz brings the black square right on the dispersion curve as shown in Figure 5. Then fiber folds in the fifth layer become also visible as shown in

the top image of Figure 5. The image of the fold at one inch distance from the left end is marked by the arrow. The second fold can be seen in this image also, however, it is not as prominent as the image at the same location produced by the mode 2 excitation, shown in the bottom image. From these two images one can conclude that the first mode imaged the fiber folds of the fifth layer and the second mode imaged the fiber fold of the third layer.

L-scan Images of the plate produced by the second (bottom image) and third (top image) modes are shown in Figure 6. For these images the transducer angle is again set at 17.5° and the frequency is set at 3.5 MHz for the second mode and 5.8 MHz for the third mode. The mode 3 image shows the matrix burned out zone better than the mode 2 image; however, fiber folds and other types of defects in the central region of the plate are clearer in the mode 2 image.

Since at high frequencies the signal has poor penetration property not much information about the defects in internal layers of the composite plate can be obtained by scanning the specimen with higher Lamb modes. Figure 7 confirms this expectation. The acoustic image shown in this figure is generated with a frequency of 6.71 MHz and the transducer angle of 12.5° - it corresponds to the fifth mode. Hardly any defect can be seen in this image.

L-scan images are then compared with the conventional C-scan images. Figure 8 shows back surface reflected C-scan images generated by 2.2 MHz(top), 5.0 MHz(middle) and 10.0 MHz(bottom) transducers used in the pulse-echo mode - the transducer axis is positioned normal to the plate specimen. Matrix burned out zone has been poorly imaged by all these three transducers. However, the 10 MHz transducer could image the die plug, marked by the arrow. Figure 9 shows two more images generated by the C-scan technique. The top image is generated by a 15 MHz transducer in the vertical pulse-echo mode and the bottom image is generated by the back scattered signals received by a 10 MHz transducer inclined at an angle of 16°, as shown in Figure 9. This orientation is particularly useful in detecting cracks which are vertical to the plane of the plate. No defect can be seen in either images. Further increase in the transducer frequency doesn't produce any better quality image of this specimen.

Finally, the L-scan images produced in this research are compared with the L-scan images that are produced when the receiver is placed in the null zone as done in reference {14} in stead of the leaky Lamb wave zone beyond the specularly reflected region as done here. Figure 10 shows the image generated

by modes 1 and 2 - the transducer angle is equal to 17.5°, and the signal frequency is equal to 1.72 MHz for the top image and 3.5 MHz for the bottom image. Hence, transducer orientation and signal frequency for these images are identical to those shown in Figure 4. The only difference is that in this case the receiver is located in the null zone unlike the previous case. Hardly any known defect can be seen in the two images of Figure 10.

Concluding Remarks

It is shown in this research that the Lamb wave scanning technique is superior to the conventional C-scan technique for imaging defects in internal layers of a multilayered composite plate. The Lamb wave scanning image can show fiber folds and burned out matrix type defects that cannot be seen in images generated by the conventional C-scan in the pulse-echo arrangement, whether the transducer is positioned vertically, i.e. normal to the specimen or at an inclination to receive the back scattered signal. It is also shown that if the receiver is placed in the null zone the defect detection capability goes down significantly because of the other factors, such as the plate thickness, that can affect the null zone position.

Acknowledgment

This research was supported by the summer faculty research grant from the Air Force Office of Scientific Research. I would like to express my gratitude to Dr. Tom Moran for giving me the opportunity to carry out this research. Helps of Mark Ruddell, Mark Blodgett, Prasanna Karpur, Teo Matikas, Jeff Fox and Laura Mann are also acknowledged.

References

1. Kundu, T., and Blodgett, M., 1993, "Detection of Material Defects in Layered Solids Using Lamb Waves", Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, Vol.13, p.1343-1350.

2. Yang, W., and Kundu, T., 1994, "Lamb Wave Propagation in Multilayered Anisotropic Solids and Its Applications Towards Imaging Material Defects", Proceedings of the 21st International Symposium on Acoustical Imaging, Laguna Beach, California.

3. Yang, W., and Kundu, T., 1994, "Efficient Use of Lamb Waves to Characterize Multilayered Anisotropic Plates", ASME Winter Annual Meeting, Chicago, November, 1994.

4. Yang, W., 1994, "Elastic Wave Propagation in Multilayered Anisotropic Plates and Its Potential Application in Non Destructive Testings", Ph.D. Dissertation, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, AZ, USA.

5. Chimenti, D.E., and Nayfeh, A. H., 1985, "Leaky Lamb Waves in Fibrous Composite Laminates", Journal of Applied Physics, Vol.58, p.4531-4538.

6. Nagy, P. B., Rose, W. R., and Adler, L., 1986, "A Single Transducer Broadband Technique for Leaky Lamb Wave Detection", Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, Vol.5, p.483-490.

7. Pearson, L. H., and Murri, W. J., 1986, Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, Vol.5, p.1093-1101.

8. Rose, W. R., Rokhlin, S. I., and Adler, L., 1986, Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, Vol.5, p.1111-1117.

9. Nayfeh, A. H., 1986, "Acoustic Wave Reflection from Water-Laminated Composite Interfaces", Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, Vol.5, p.1119-1128.

10. Bar-Cohen, Y., and Chimenti, D.E., 1986, "Nondestructive Evaluation of Composites by Leaky Lamb Waves", Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, Vol.5, p.1199-1206.

11. Chimenti, D. E. and Bar-Cohen, Y., 1986, "Signal Analysis of Leaky Lamb Wave Spectra for NDE of Composites", Proceedings of 1985 IEEE Ultrasonic Symposium, , Ed. McAvoy, B.R., Pub. IEEE, New York, USA, p.1028-1031.

12. Martin, R.W., and Chimenti, D.E., 1987, "Signal Processing of Leaky Lamb Wave Data for Defect Imaging in Composite Laminates", Review of Progress in Quantitative NDE, Eds. Thompson, D. O., and Chimenti, D. E., Pub. Plenum Press, New York, USA, Vol.6, p.815-824

13. Mal, A. K., and Bar-Cohen, Y., 1988, "Ultrasonic Characterization of Composite Laminates", Wave Propagation in Structural Composites, Eds. Mal, A. K., amd Ting, T. C. T., AMD Vol.90, Pub. ASME, New York, p.1-16.

14. Chimenti, D. E., and Martin, R. W., 1991, "Nondestructive Evaluation of Composite Laminates by Leaky Lamb Waves", Ultrasonics, Vol.29, p.13-21.

Figure Captions

Figure 1: Left column - Transmitter-receiver-specimen positions in a pitch-catch arrangement. Right column - Expected reflected signal spectra when defocussing is low (top row) and high (bottom row). (a) No Lamb waves are generated, (b) & (c) Lamb waves are generated.

Figure 2: Flaws in different layers in the twelve-layer plate specimen.

Figure 3: Experimentally generated dispersion curves in 0°, 90° and 45° directions denoted by cross (+), square (□) and diamond (◊) symbols respectively. Frequency range - 0-5 MHz (top figure) and 0-15 MHz (bottom figure).

Figure 4: Mode 1 (θ = 17.5°, f = 1.92 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Arrow in the mode 2 image shows the fiber fold of the third layer.

Figure 5: Mode 1 (θ = 17.5°, f = 1.72 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Arrow in the mode 1 image shows the fiber fold of the fifth layer.

Figure 6: Mode 3 (θ = 17.5°, f = 5.8 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Arrow in the mode 2 image shows the fiber fold of the third layer.

Figure 7: Mode 5 (θ = 12.5°, f = 6.71 MHz) image (top), and dispersion curves (bottom) with the black square showing the phase velocity - frequency combinations used for scanning.

Figure 8: Conventional C-scan images with 2.2 MHz (top), 5.0 (MHz) and 10.0 (MHz) transducers at normal incidence. Arrow in the bottom image shows the die plug location.

Figure 9: C-scan images with 15 MHz transducer at normal incidence (top) and 10 MHz transducer at 16 degree incidence (bottom).

Figure 10: Mode 1 (θ = 17.5°, f = 1.72 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Receiver is in the null zone.

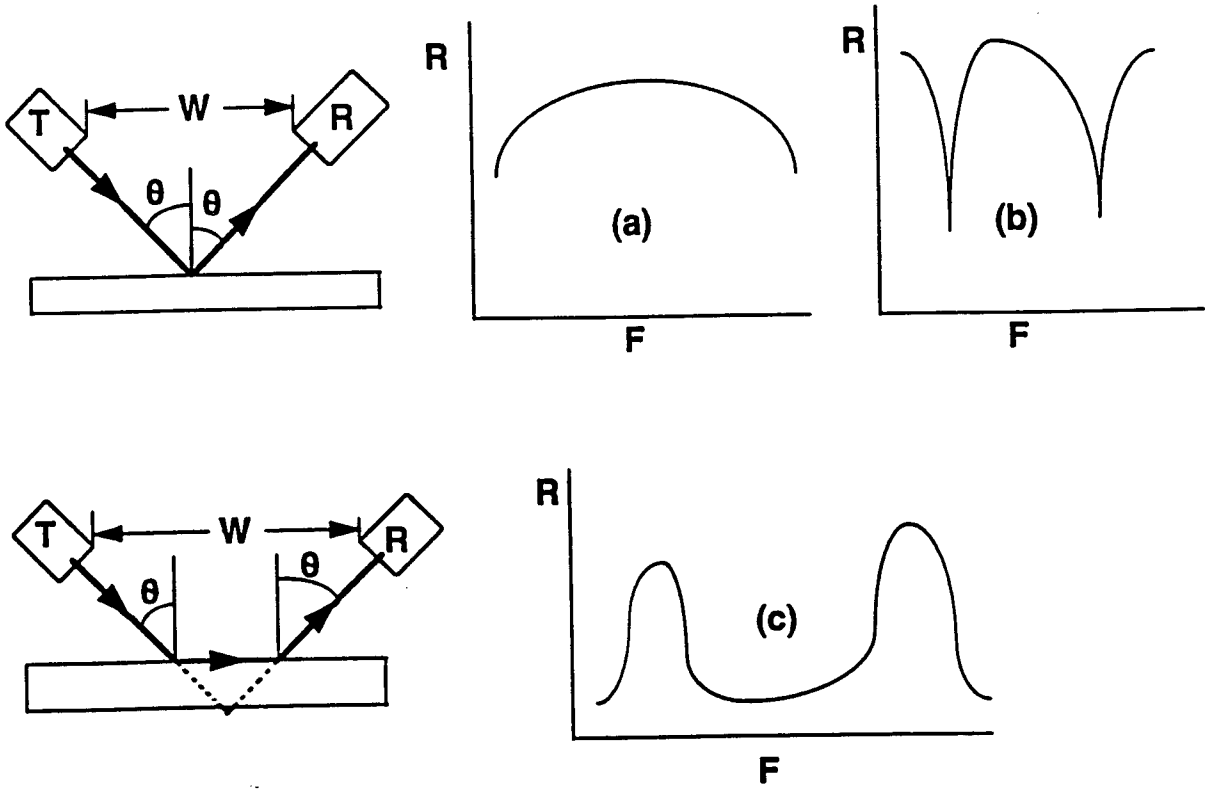


Figure 1: Left column - Transmitter-receiver-specimen positions in a pitch-catch arrangement. Right column - Expected reflected signal spectra when defocussing is low (top row) and high (bottom row). (a) No Lamb waves are generated, (b) & (c) Lamb waves are generated.

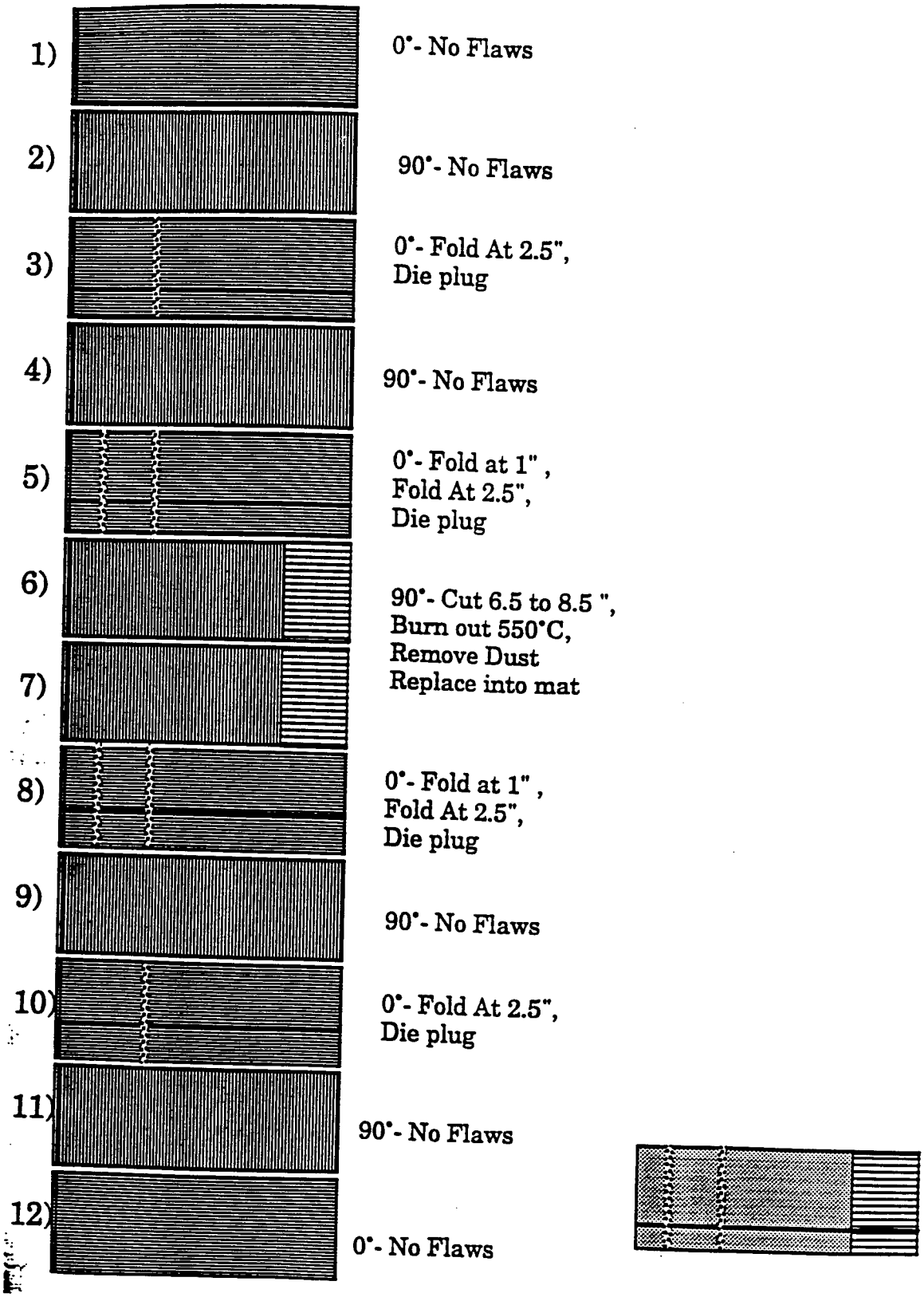


Figure 2: Flaws in different layers in the twelve-layer plate specimen.

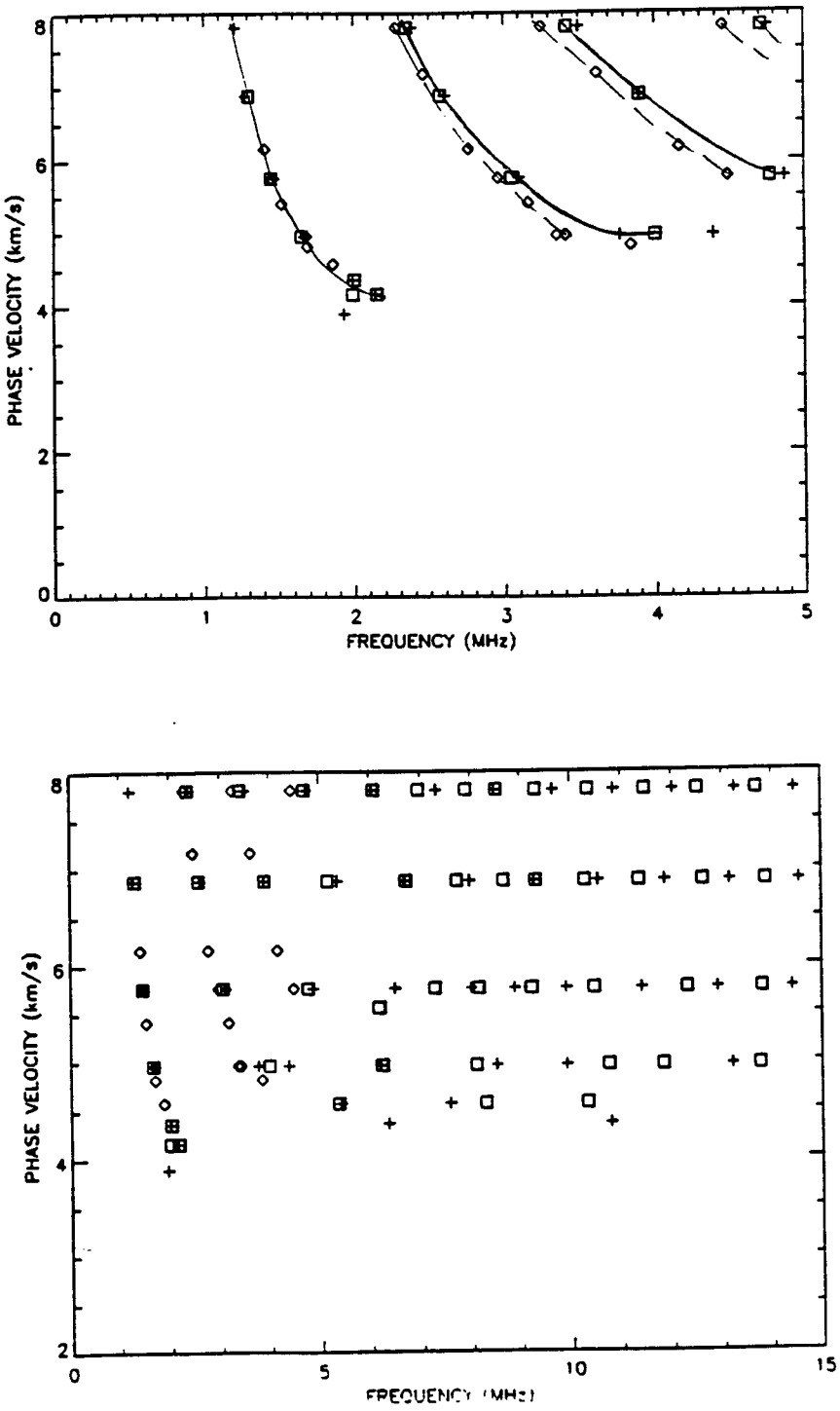


Figure 3: Experimentally generated dispersion curves in 0°, 90° and 45° directions denoted by cross (+), square (□) and diamond (◊) symbols respectively. Frequency range - 0-5 MHz (top figure) and 0-15 MHz (bottom figure).

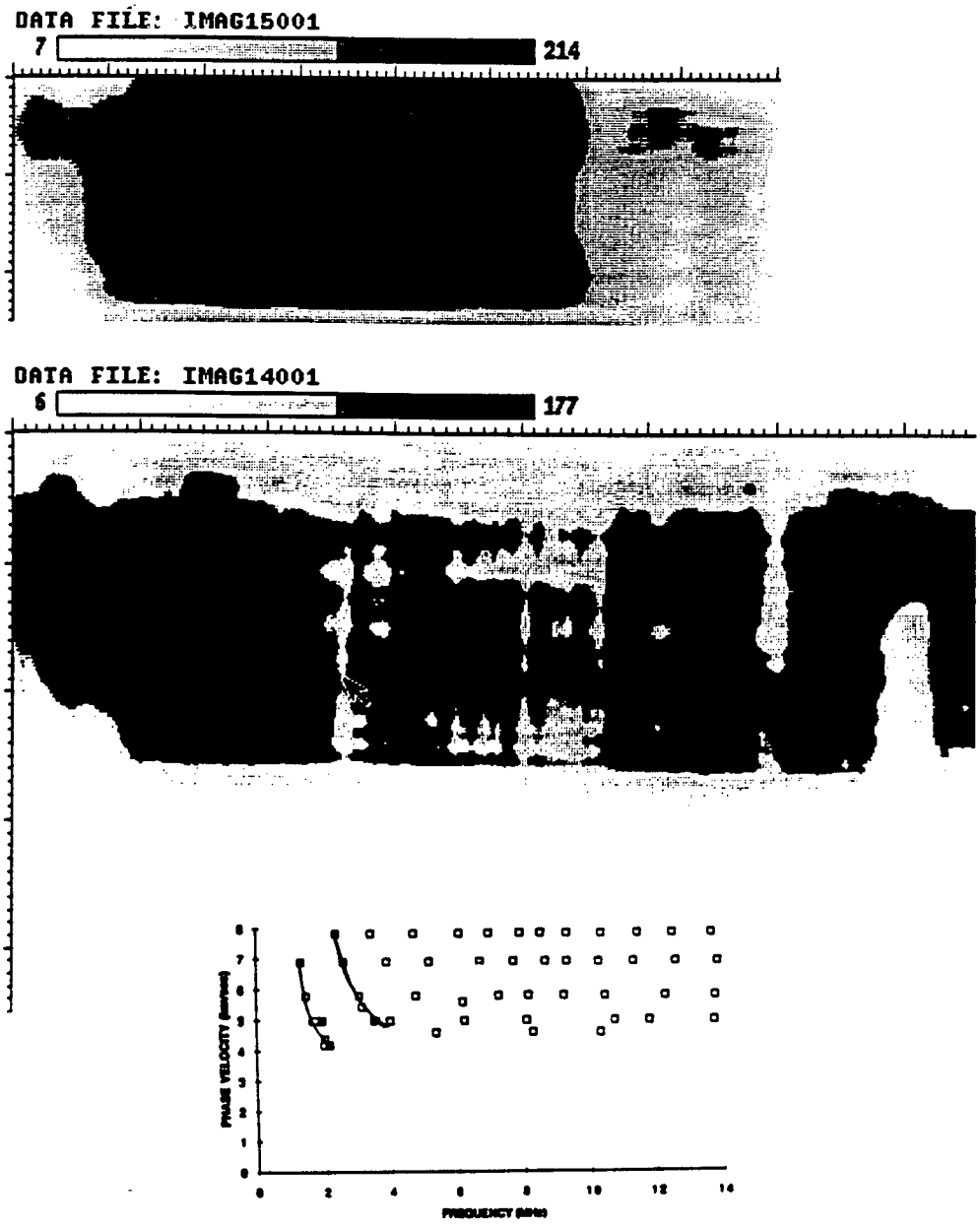


Figure 4: Mode 1 (θ = 17.5°, f = 1.92 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Arrow in the mode 2 image shows the fiber fold of the third layer.

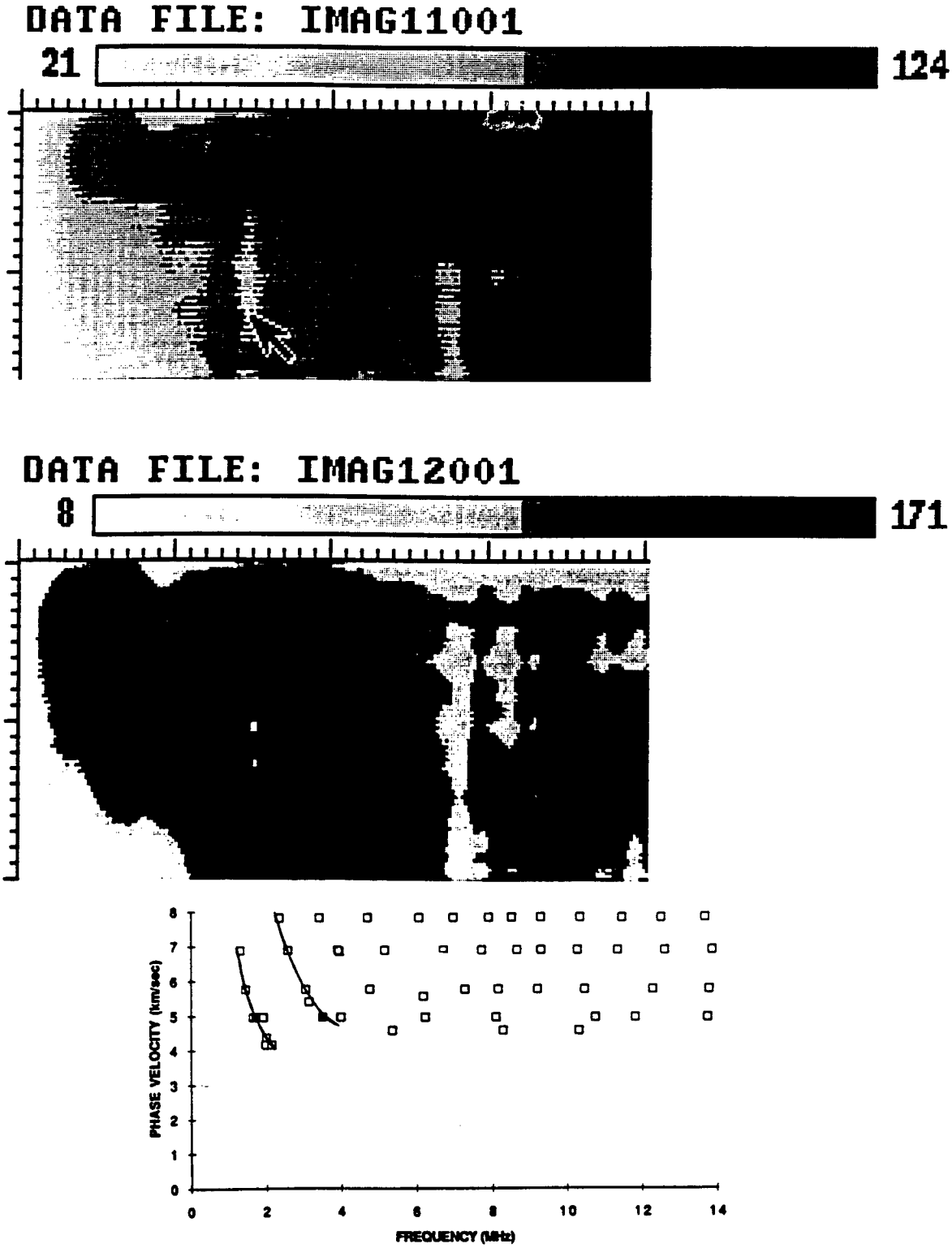


Figure 5: Mode 1 (θ = 17.5°, f = 1.72 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Arrow in the mode 1 image shows the fiber fold of the fifth layer.

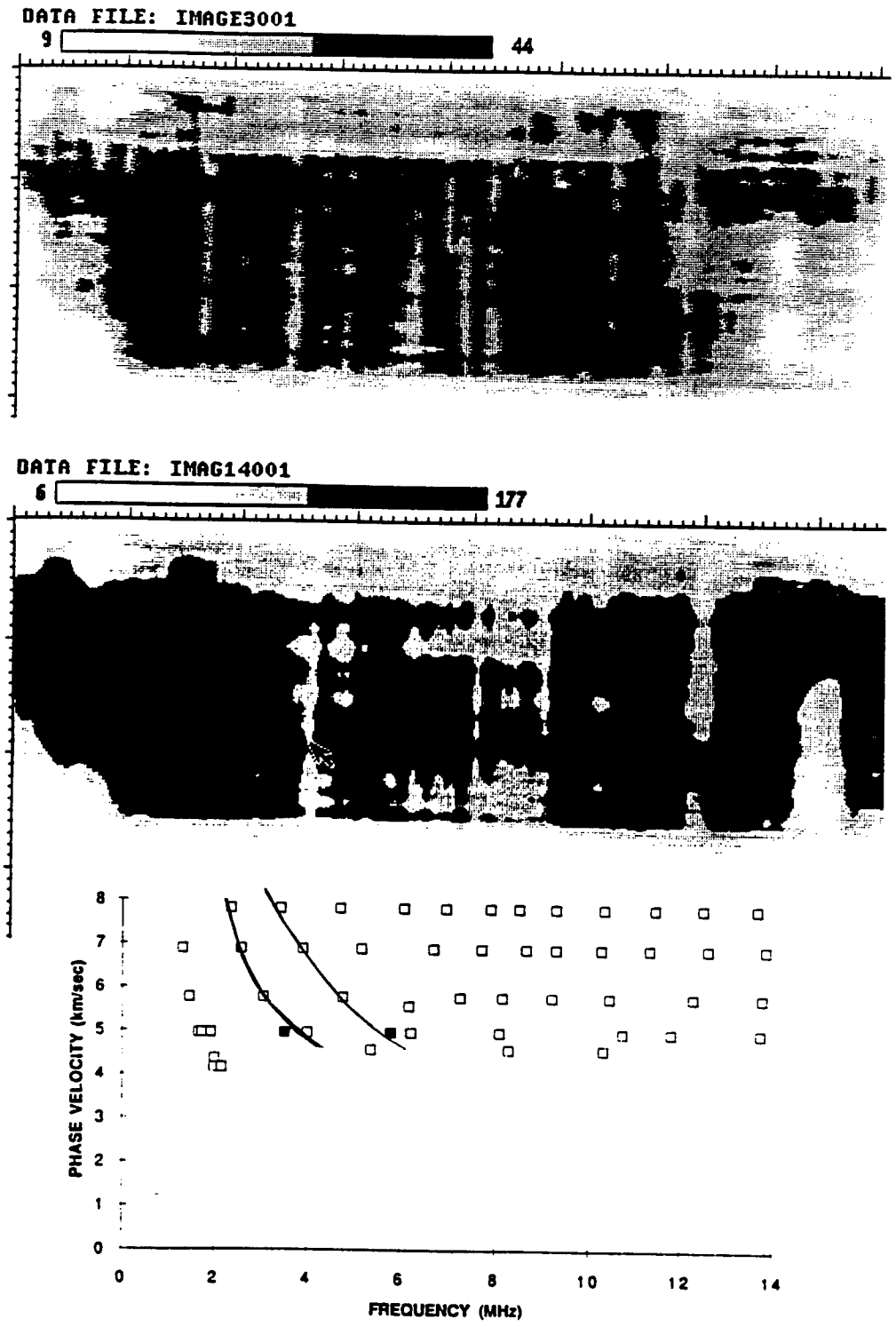


Figure 6: Mode 3 (θ = 17.5°, f = 5.8 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Arrow in the mode 2 image shows the fiber fold of the third layer.

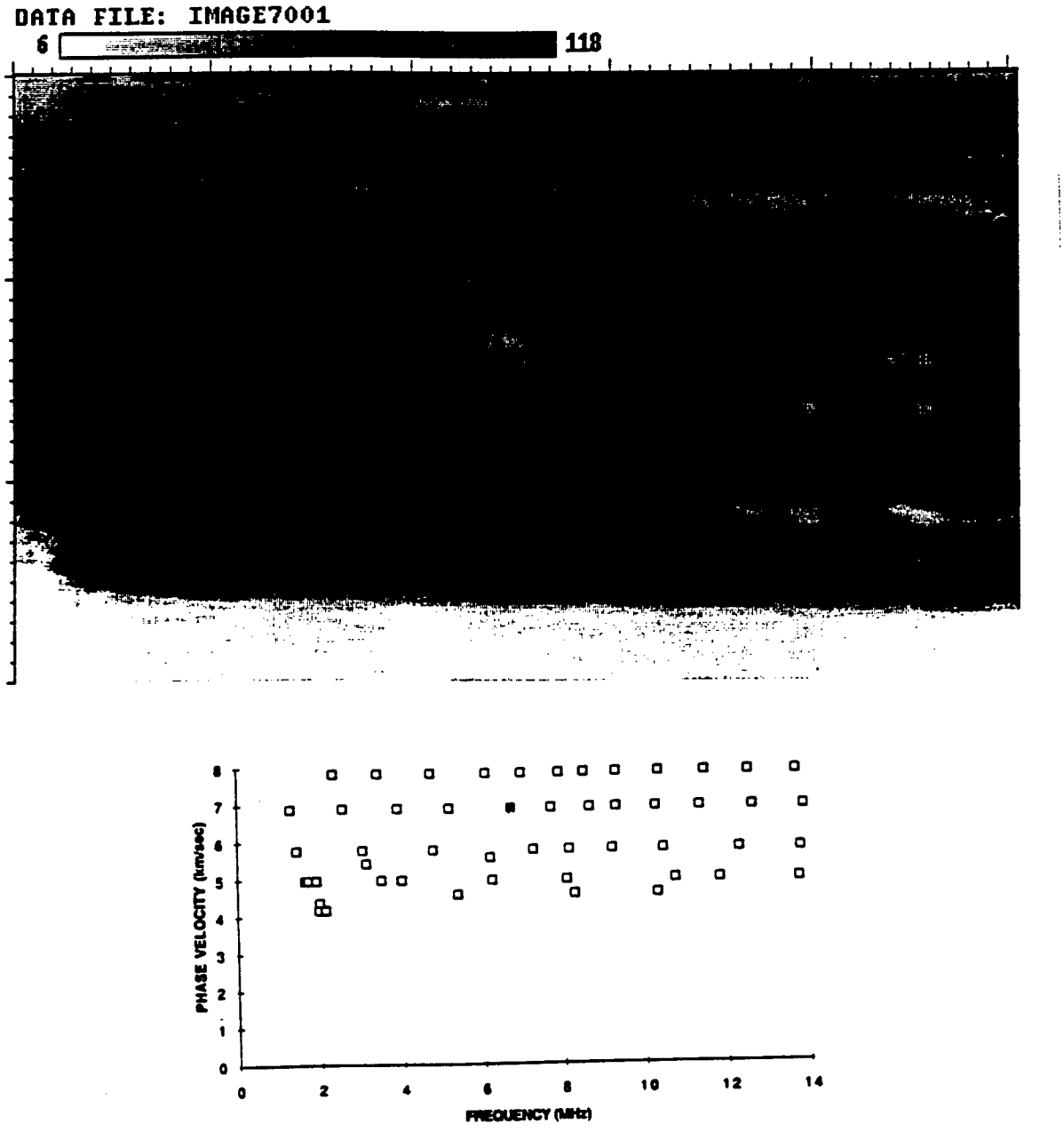


Figure 7: Mode 5 (θ = 12.5°, f = 6.71 MHz) image (top), and dispersion curves (bottom) with the black square showing the phase velocity - frequency combinations used for scanning.

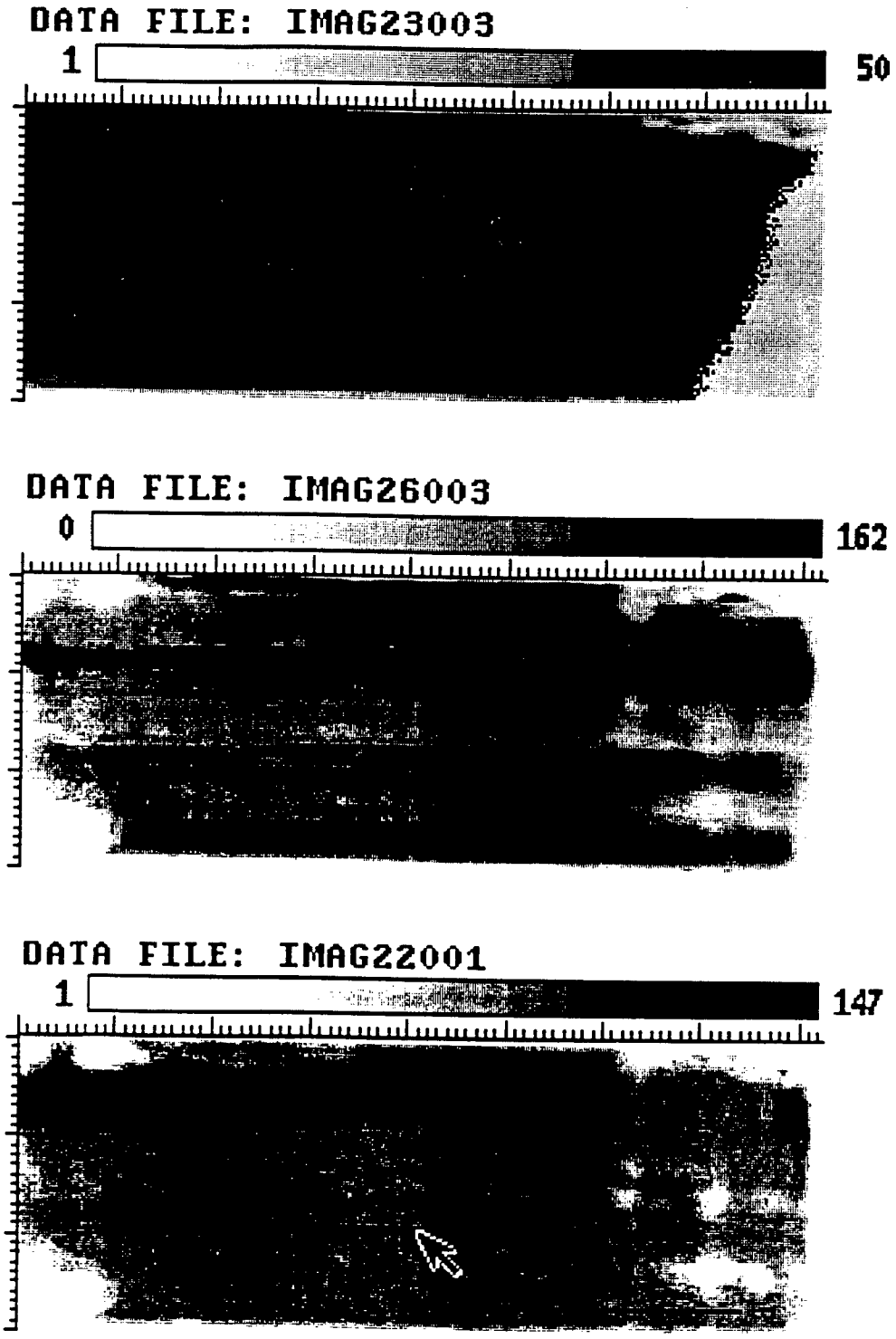


Figure 8: Conventional C-scan images with 2.2 MHz (top), 5.0 (MHz) and 10.0 (MHz) transducers at normal incidence. Arrow in the bottom image shows the die plug location.

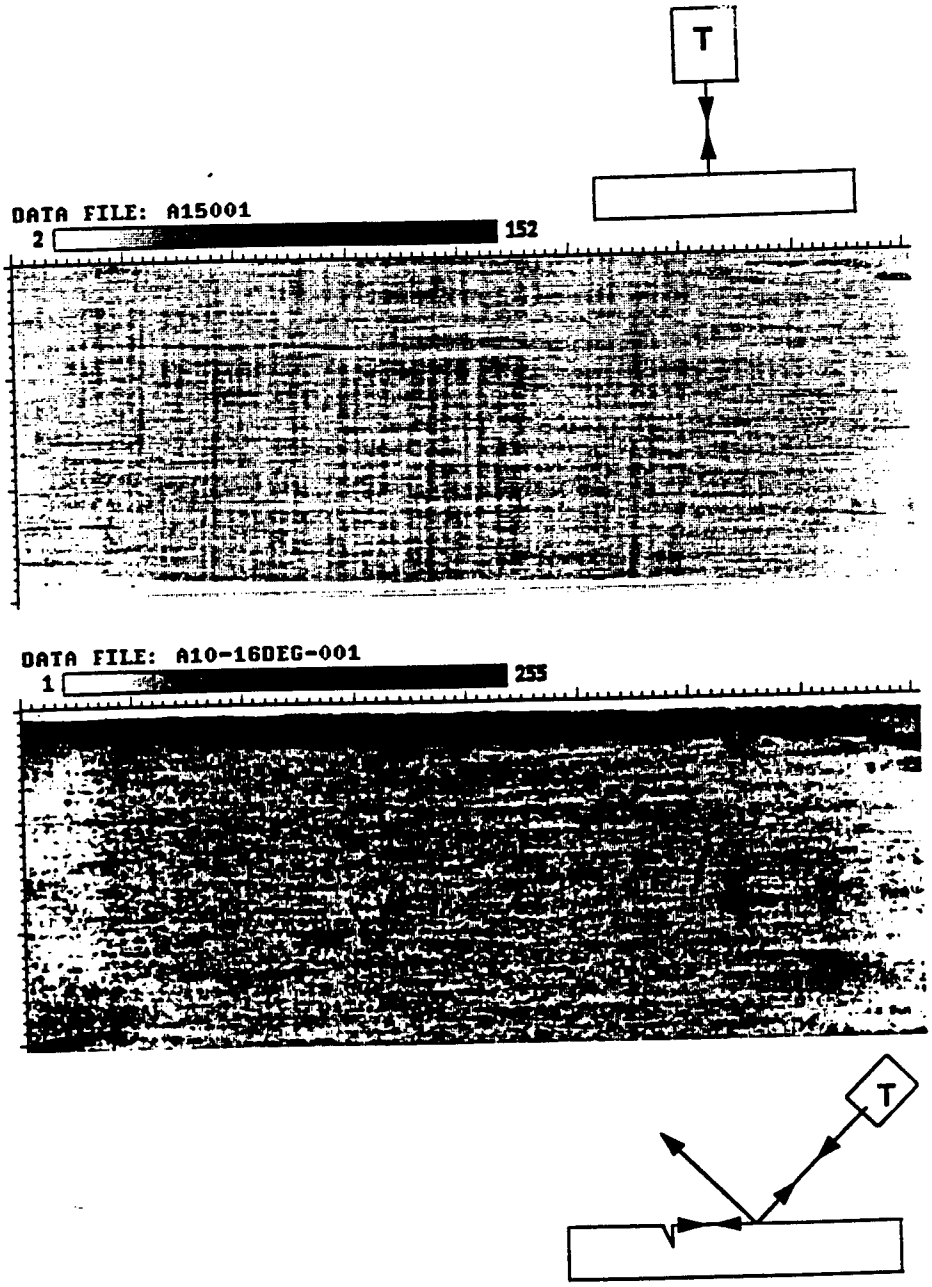


Figure 9: C-scan images with 15 MHz transducer at normal incidence (top) and 10 MHz transducer at 16 degree incidence (bottom).

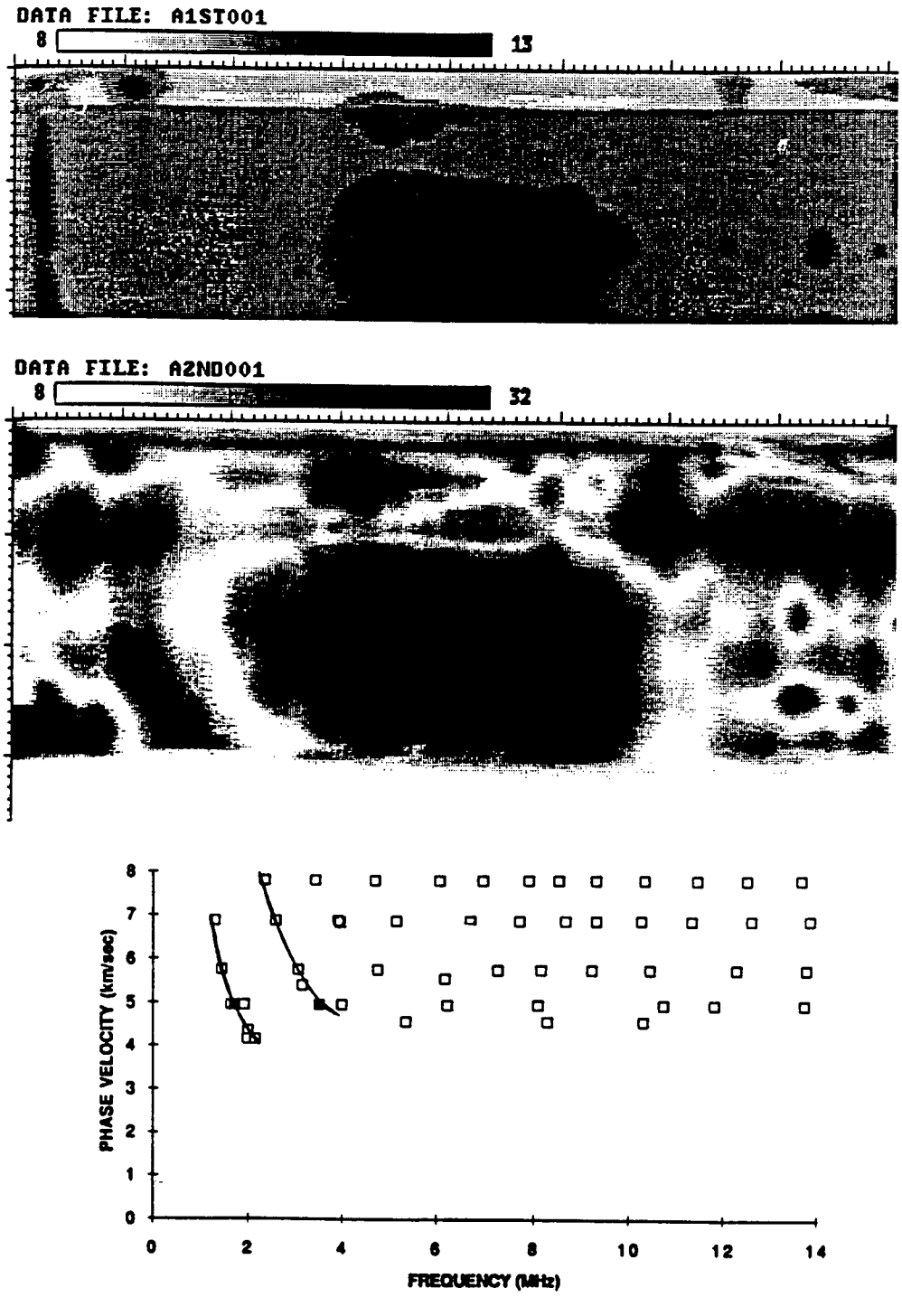


Figure 10: Mode 1 (θ = 17.5°, f = 1.72 MHz) (top), mode 2 (θ = 17.5°, f = 3.50 MHz) (middle) images, and dispersion curves (bottom) with black squares showing the phase velocity - frequency combinations used for scanning. Receiver is in the null zone.

HIGH RESOLUTION RANGE SIGNATURE ESTIMATION

Jian Li
Assistant Professor
Department of Electrical Engineering

405 CSE, Bldg. 42
University of Florida
Gainesville, FL 32611

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

August 1994

HIGH RESOLUTION RANGE SIGNATURE ESTIMATION

Jian Li

Assistant Professor

Department of Electrical Engineering

University of Florida

405 CSE, Bldg. 42

Gainesville, FL 32611

Abstract

This report presents a robust parametric data model for estimating high resolution range signatures of radar targets. This paper also presents an estimation algorithm for the data model. The algorithm is referred to as the APES (Amplitude and Phase Estimation of a Sinusoid in unknown colored noise) algorithm. We shall describe how the APES algorithm can be used to estimate range signatures. We shall show, with both numerical and experimental examples, that our modeling and estimation approach yields better resolution and lower sidelobes than the conventional nonparametric FFT (fast Fourier transform) method. We shall also show that our approach is more robust than modeling the radar data as a certain number of complex sinusoids in noise and estimating the frequencies, amplitudes, and phases of the sinusoids with one of the best sinusoidal parameter estimation methods.

HIGH RESOLUTION RANGE SIGNATURE ESTIMATION

Jian Li

I. Introduction

This paper proposes a robust parametric data model for estimating high resolution range signatures of radar targets. For the range signature estimation, estimating the radar cross section (RCS) of the scattering center of a radar target at a certain range is modeled as estimating the amplitude and phase of a complex sinusoid with known frequency in unknown colored Gaussian noise. This new modeling approach models the thermal noise and the interferences from other scattering centers of a radar target as unknown colored noise.

This paper also presents a parameter estimation algorithm for the data model. The algorithm is referred to as the APES (Amplitude and Phase Estimation of a Sinusoid in unknown colored noise) algorithm. We shall describe how the APES algorithm can be used to estimate the range signatures. The APES algorithm avoids the search over the parameter space and requires only simple matrix multiplications and matrix inverses.

We will show that our modeling and estimation approach yields better resolution and lower sidelobes than the conventional nonparametric FFT (fast Fourier transform) method. We will also show that our approach is more robust than modeling the radar data as a certain number of complex sinusoids in noise and estimating the frequencies, amplitudes, and phases of the sinusoids with one of the best sinusoidal parameter estimation methods. We will present both numerical and experimental examples comparing the performance of our approach with the two afore-mentioned approaches.

II. High Resolution Range Signature Estimation

In this section, we propose a parametric data model and present an estimation algorithm for estimating high resolution range signatures of a radar target. High resolution range signatures are useful for many applications including the non-cooperative target identification (NCTI).

A. Problem Formulation

The range resolution of a radar is determined by the radar bandwidth. To achieve high resolution in range, the radar must transmit wideband pulses, which are often linear frequency modulated (chirp) pulses. Upon receiving each pulse returned by a radar target, the radar demodulates the pulse by mixing the pulse with a reference chirp signal and low-pass filter the mixed signal. As a result, the scattering centers of a radar target at different ranges correspond to different frequencies of the output of this demodulation operation.

Let $\mathbf{z}$ denote an $M \times 1$ vector containing the samples of the output of this operation. We assume that the radar bandwidth is moderate so that the radar cross sections (RCSs) of the scattering centers of the radar target do not change with frequency. Then $\mathbf{z}$ may be written as

$$\mathbf{z} = \sum_{k=1}^{K} \gamma_k \mathbf{f}_k + \mathbf{n}_0, \tag{1}$$

where γ_k, $k = 1, 2, \cdots, K$, denotes the RCS of the kth scattering center of the radar target, $\mathbf{n}_0$ denotes the additive noise vector, and

$$\mathbf{f}_k = \begin{bmatrix} 1 & e^{-jt_k} & \cdots & e^{-j(M-1)t_k} \end{bmatrix}^T, \tag{2}$$

with $(\cdot)^T$ denoting the transpose and t_k denoting the time delay proportional to the range of the kth scattering center.

A simple nonparametric method of estimating γ_k and t_k is to use FFT (Fast Fourier Transform), which is both computationally efficient and robust to model errors. However, FFT is known for its high sidelobes and poor resolution. Many different types of windows may be applied to $\mathbf{z}$ to reduce the sidelobes. Yet using windows with FFT further reduces the already poor resolution.

When the number of scattering centers K of a radar target is small, one may use many existing fre-

quency estimation methods, such as the computationally and asymptotically (for high signal-to-noise ratio) statistically efficient MODE method [1, 2, 3], to estimate the t_k. Once t_k are estimated, the straightforward least-squares fitting method may be used to estimate the corresponding γ_k.

However, there are two cases where the above parametric modeling approach may result in poor performance. First, for a complicated radar target such as an airplane, the number of scattering centers K may be very large and may even be larger than the number of samples M in $\mathbf{z}$. For this case, the frequency estimation methods cannot be used since to uniquely determine the time delays t_k, even in the absence of the additive noise $\mathbf{n}_0$, the K must be less than $(M-1)/2$ [4, 5]. Second, when two or more scattering centers are very closely spaced, the frequency estimation methods may not be able to resolve all of the scattering centers. When this result occurs, the least-squares fitting method yields poor RCS estimates. As a result, the estimated range signature, i.e., γ_k versus t_k, may be distorted.

We consider below a more robust parametric data model. To estimate the RCS $\gamma(t)$ of a scattering center whose range corresponds to a time delay t, we model the received data $\mathbf{z}$ as

$$\mathbf{z} = \gamma(t)\mathbf{f}(t) + \mathbf{n}(t), \tag{3}$$

where $\mathbf{f}(t)$ is defined in ((2)) with t_k replaced by t and $\mathbf{n}(t)$ denotes the additive noise. The $\mathbf{n}(t)$ is assumed to be circularly symmetric complex Gaussian random vector with zero-mean and unknown covariance matrix $\mathbf{Q}(t)$. The unknown $\mathbf{Q}(t)$ models both the thermal noise and the interferences from other scattering centers. The $\mathbf{Q}(t)$ is assumed to be a positive definite and Toeplitz matrix.

The problem of interest herein is to determine the range signature $\gamma(t)$ from $\mathbf{z}$. We remark that to compute a discrete range signature with M_s samples, we set $r = 2\pi(m-1)/M_s$ for the mth sample. We shall show below that the M_s RCS estimates in the signature can be computed in parallel.

B. The APES Algorithm

We present below an algorithm that may be used to estimate the unknown $\gamma(t)$ in model ((3)). We note that the amplitude and phase of $\gamma(t)$ may be considered as the amplitude and phase of the complex

sinusoid with frequency t described by $\mathbf{f}(t)$. Thus we refer to our estimator below as the *Amplitude and Phase Estimation of a Sinusoid in unknown colored noise*, or simply the *APES*, method.

The APES algorithm is obtained by mimicking the computationally and asymptotically statistically efficient large sample maximum likelihood estimator obtained in Appendix A for estimating the complex gain of a signal with known waveform and known steering vector in unknown colored noise. To obtain the APES algorithm, we first divide the measurement vector $\mathbf{z}$ into $\tilde{N}$ overlapping subvectors $\tilde{\mathbf{z}}(n)$ of dimension $\tilde{M} \times 1$, where $\tilde{N} = M - \tilde{M} + 1$. The nth subvector $\tilde{\mathbf{z}}(n)$ contains the nth to the $(n + \tilde{M} - 1)$th element of $\mathbf{z}$. Let $\tilde{\mathbf{n}}(t, n)$ be formed from $\mathbf{n}(t)$ in the same way $\tilde{\mathbf{z}}(n)$ are formed from $\mathbf{z}$. The $\tilde{\mathbf{n}}(t, n)$ are assumed to be circularly symmetric complex Gaussian random vectors with zero-mean and the same unknown covariance matrix $\mathbf{Q}_1(t)$, which is a submatrix of $\mathbf{Q}(t)$. Then $\tilde{\mathbf{z}}(n)$, $n = 1, 2, \cdots, \tilde{N}$, may be written as

$$\tilde{\mathbf{z}}(n) = \gamma(t)\mathbf{a}_0(t)y(t, n) + \tilde{\mathbf{n}}(t, n), \tag{4}$$

where $\mathbf{a}_0(t)$ is referred to as the steering vector and has the form

$$\mathbf{a}_0(t) = \begin{bmatrix} 1 & e^{jt} & \cdots & e^{j(\tilde{M}-1)t} \end{bmatrix}^T, \tag{5}$$

and

$$y(t, n) = e^{j(n-1)t}, \quad n = 1, 2, \cdots, \tilde{N}. \tag{6}$$

Let $\mathbf{J}$ be the $\tilde{M} \times \tilde{M}$ exchange matrix (with ones on the antidiagonal and zeros elsewhere). Let

$$\bar{\mathbf{z}}(n) = \mathbf{J}\tilde{\mathbf{z}}^*(\tilde{N} - n + 1), \tag{7}$$

where $(\cdot)^*$ denotes the complex conjugate. Let

$$\bar{\mathbf{n}}(t, n) = \mathbf{J}\tilde{\mathbf{n}}^*(t, \tilde{N} - n + 1), \tag{8}$$

where $\bar{\mathbf{n}}(t, n)$ are assumed to have the same statistics as $\tilde{\mathbf{n}}(t, n)$. Then $\bar{\mathbf{z}}(n)$ may be written as

$$\bar{\mathbf{z}}(n) = \gamma^*(t)\beta(t)\mathbf{a}_0(t)y(t, n) + \bar{\mathbf{n}}(t, n), \tag{9}$$

where

$$\beta(t) = e^{-j(N-1)t}. \tag{10}$$

Let

$$\hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t) = \frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} \tilde{\mathbf{z}}(n) y(t,n). \tag{11}$$

Let $\hat{r}_{yy}$, $\hat{\mathbf{R}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$, $\hat{\mathbf{R}}_{\bar{\mathbf{z}}\bar{\mathbf{z}}}$, and $\hat{\mathbf{r}}_{\bar{\mathbf{z}}y}(t)$ be defined similarly as $\hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t)$. Note from ((6)) that $\hat{r}_{yy} = 1$. Note also that as may be seen from ((11)), the mth element of $\hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t)$ is the discrete-time Fourier transform of the mth sequence $\{\tilde{z}_m(1), \cdots, \tilde{z}_m(\tilde{N})\}$ divided by $\tilde{N}$, where $\tilde{z}_m(n)$ denotes the mth element of $\tilde{\mathbf{z}}(n)$. Thus to compute a discrete range signature with M_s samples, where M_s is a power of 2, the $\hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t)$, and similarly the $\hat{\mathbf{r}}_{\bar{\mathbf{z}}y}(t)$, may be computed with the FFT method and zero-padding. The amount of computations required for each sequence is $O[M_s \log_2(M_s)]$. Since the 2nd through the $\tilde{N}$th elements of the mth sequence are the 1st through the $(\tilde{N}-1)$th elements of the $(m+1)$th sequence, the FFT of the $(m+1)$th sequence may be updated from the FFT of the mth sequence with $O(M_s)$ operations.

Let $\tilde{\mathbf{b}}(t) = \gamma(t)\mathbf{a}_0(t)$. Applying the results in Appendix A to $\tilde{\mathbf{z}}(n)$, we obtain the estimate of $\tilde{\mathbf{b}}(t)$ as

$$\hat{\tilde{\mathbf{b}}}(t) = \hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t). \tag{12}$$

Similarly, let $\bar{\mathbf{b}}(t) = \gamma^*(t)\beta(t)\mathbf{a}_0(t)$. Applying the results in Appendix A to $\bar{\mathbf{z}}(n)$, we have

$$\hat{\bar{\mathbf{b}}}(t) = \hat{\mathbf{r}}_{\bar{\mathbf{z}}y}(t). \tag{13}$$

By applying the results in Appendix A to both $\tilde{\mathbf{z}}(n)$ and $\bar{\mathbf{z}}(n)$, the covariance matrix $\mathbf{Q}_1(t)$ may be estimated as

$$\begin{aligned}
\hat{\mathbf{Q}}_1(t) &= \frac{1}{2}\left[\hat{\mathbf{R}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} - \hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t)\hat{\mathbf{r}}^H_{\tilde{\mathbf{z}}y}(t) + \hat{\mathbf{R}}_{\bar{\mathbf{z}}\bar{\mathbf{z}}} - \hat{\mathbf{R}}_{\bar{\mathbf{z}}y}(t)\hat{\mathbf{r}}^H_{\bar{\mathbf{z}}y}(t)\right] \\
&= \frac{1}{2}\left\{\hat{\mathbf{R}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} + \hat{\mathbf{R}}_{\bar{\mathbf{z}}\bar{\mathbf{z}}} - \left[\begin{array}{cc} \hat{\tilde{\mathbf{b}}}(t) & \hat{\bar{\mathbf{b}}}(t) \end{array}\right]\left[\begin{array}{cc} \hat{\tilde{\mathbf{b}}}(t) & \hat{\bar{\mathbf{b}}}(t) \end{array}\right]^H\right\} \\
&\triangleq \frac{1}{2}\left[\hat{\mathbf{R}} + \hat{\mathbf{B}}(t)\hat{\mathbf{B}}^H(t)\right],
\end{aligned} \tag{14}$$

where $(\cdot)^H$ denotes the complex conjugate transpose. Then the estimate of $\gamma(t)$ may be computed as (see

Appendix A)

$$\hat{\gamma}(t) = \frac{1}{2}\left[\frac{\mathbf{a}_0^H(t)\hat{\mathbf{Q}}_1^{-1}(t)\hat{\tilde{\mathbf{b}}}(t) + \beta(t)\hat{\tilde{\mathbf{b}}}^H(t)\hat{\mathbf{Q}}_1^{-1}(t)\mathbf{a}_0(t)}{\mathbf{a}_0^H(t)\hat{\mathbf{Q}}_1^{-1}(t)\mathbf{a}_0(t)}\right]. \tag{15}$$

Note that since only $\hat{\mathbf{B}}(t)$ is a function of the range related time delay t, the $\hat{\mathbf{Q}}_1^{-1}(t)$ can be computed more efficiently with

$$\hat{\mathbf{Q}}_1^{-1}(t) = 2\hat{\mathbf{R}}^{-1} - \hat{\mathbf{R}}^{-1}\hat{\mathbf{B}}(t)\left[\hat{\mathbf{B}}^H(t)\hat{\mathbf{R}}^{-1}\hat{\mathbf{B}}(t) - \mathbf{I}\right]^{-1}\hat{\mathbf{B}}^H(t)\hat{\mathbf{R}}^{-1}. \tag{16}$$

We note that computing $\hat{\mathbf{R}}^{-1}$ requires $O(\tilde{M}^3)$ computations. Once $\hat{\mathbf{R}}^{-1}$ is obtained, computing $\hat{\mathbf{Q}}_1^{-1}(t)$ for each t requires $O(\tilde{M}^2)$ computations. Computing $\hat{\gamma}(t)$ for each t also requires $O(\tilde{M}^2)$ computations.

The the APES estimator for estimating $\gamma(t)$ may be summarized as follows:

Step 1: Compute $\hat{\tilde{\mathbf{b}}}(t)$, $\hat{\bar{\mathbf{b}}}(t)$, and $\hat{\mathbf{Q}}_1(t)$ with ((12)), ((13)), and ((16)), respectively.

Step 2: Determine $\hat{\gamma}(t)$ with ((15)).

We remark that the vectors $\tilde{\mathbf{n}}(t,n)$ and $\bar{\mathbf{n}}(t,n)$ are not independent of each other since they are formed as the overlapping subvectors of $\mathbf{n}(t)$. Yet since these vectors are not completely correlated with each other, it can be shown that as $\tilde{N}$ goes to infinity, $\hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t)$, $\hat{\mathbf{R}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$, $\hat{\mathbf{R}}_{\bar{\mathbf{z}}\bar{\mathbf{z}}}$, and $\hat{\mathbf{r}}_{\bar{\mathbf{z}}y}(t)$ are all consistent estimates. As a result, the $\hat{\gamma}(t)$, for all possible t, is also a consistent estimate of $\gamma(t)$.

We also remark that the APES algorithm only requires simple matrix multiplications and matrix inverses. Moreover, many computations of the APES algorithm can be done in parallel. It is thus possible to achieve real-time range signature estimation by implementing the APES algorithm with parallel processors and/or specially designed hardwares.

We note that for the special case of $\tilde{M} = 1$, the $\hat{\gamma}(t)$ in ((15)) can be shown to be the Fourier transform of $\mathbf{z}$. Thus the Fourier transform method is a special case of the APES method.

Finally, the parameter $\tilde{M}$ has the following effects on the APES performance. Note first that the larger the $\tilde{M}$, the larger the dimension of $\mathbf{Q}_1(t)$, and thus the better the modeling of the interferences. We shall show in the following subsection that the larger the $\tilde{M}$, the better the resolution of the APES method. On the other hand, the larger the $\tilde{M}$, the smaller the $\tilde{N} = M - \tilde{M} + 1$. Thus increasing $\tilde{M}$ increases the variance of $\hat{\mathbf{Q}}_1(t)$ since $\hat{\mathbf{r}}_{\tilde{\mathbf{z}}y}(t)$, $\hat{\mathbf{R}}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$, $\hat{\mathbf{R}}_{\bar{\mathbf{z}}\bar{\mathbf{z}}}$, and $\hat{\mathbf{r}}_{\bar{\mathbf{z}}y}(t)$ are poorer estimates for larger $\tilde{M}$. We shall show

in the following subsection that for very large $\tilde{M}$, the variance of $\hat{\gamma}(t)$ may increase. Also, increasing $\tilde{M}$ increases the amount of computations needed by the APES method.

C. *Numerical and Experimental Results*

We present below both numerical and experimental examples showing the performance of the APES algorithm. In the simulated numerical examples, the root-mean-squared errors (RMSEs) of the APES estimator are obtained with 100 independent Monte-Carlo trials and are compared with the corresponding Cramer-Rao bounds (CRBs).

We first use simple numerical examples to illustrate the performance of the APES estimator. Consider first an example where the true range signature is shown in Figure 1(a). We assume that there are 66 scattering centers in the signature. The number of data samples in the received data vector $\mathbf{z}$ in ((3)) is assumed to be $M = 128$. The additive thermal noise in $\mathbf{z}$ is assumed to be a zero-mean white Gaussian random process with variance 1. Figure 1(b) shows the estimated range signature (amplitude only) with the FFT method, which is also equivalent to the APES method with $\tilde{M} = 1$. Note that due to the large sidelobes of the FFT method, the fourth scattering center is not discernible in Figure 1(b). Figure 1(c) shows the estimated range signature with the APES method with $\tilde{M} = 48$. Comparing Figures 1(b) and (c), we note that using APES method with $\tilde{M} = 48$ yields a much better range signature estimate, which has much reduced sidelobes and higher resolution for the large scattering centers, than using the FFT method. For the 62 very small and very closely spaced scattering centers between 0.97 and 2.47, the APES method with $\tilde{M} = 48$ tends to suppress them. Figure 1(d) shows the estimated range signature obtained when using the MODE with the least-squares fitting (MODE-LSF) method by assuming that there are 48 complex sinusoids plus white noise in the received data vector $\mathbf{z}$. We note that since the data model is incorrect for the MODE-LSF method and some of the estimated scattering centers are very close to each other, using the MODE-LSF method may yield very large false peaks. For this case, they occur near $t = 2.5$. (Note that this type of large false peaks do not occur in every Monte-Carlo simulation. They occur in about 20% of the Monte-Carlo simulations.) The presence of these large false peaks makes the MODE-LSF method

less preferable even than the FFT method. Figure 1(e) shows the estimated range signature obtained by using FFT with Kaiser window and shape parameter 4. Comparing Figures 1(b), (c), and (e), we note that using FFT with Kaiser window can reduce the FFT sidelobes, but the already poor resolution of the FFT method is made poorer due to the windowing.

We now consider the effects of $\tilde{M}$ on the performance of the APES estimator. Figure 2 shows the RMSEs of $\hat{\gamma}(t = 0.2332)$ (the first scatterer in Figure 1(a)) obtained with the APES estimator as a function of $\tilde{M}$. We note from Figure 2 that the APES estimator with a proper $\tilde{M}$ can give much more accurate estimates of $\gamma(t = 0.2332)$ than the FFT method, which is equivalent to the APES method with $\tilde{M} = 1$. With a proper $\tilde{M}$, the performance of the APES estimator can be close the corresponding CRB, which is also the best unbiased performance that can be achieved by an estimator. Note also that for very large $\tilde{M}$, the RMSE of $\hat{\gamma}(t = 0.2332)$ may increase due to the poor estimate of $\mathbf{Q}_1(t)$. To achieve the best estimate of $\hat{\gamma}(t = 0.2332)$, the $\tilde{M}$ should be within $28 \leq \tilde{M} \leq 64$.

Figures 3(a) and (b) show the range signature estimates obtained with the APES method when $\tilde{M} = 28$ and $\tilde{M} = 64$, respectively. We note from Figure 2(a) that using the APES method with any $\tilde{M}$ within $28 \leq \tilde{M} \leq 64$ yields similar RMSEs for $\hat{\gamma}(t = 0.2332)$. Yet Figure 3 shows that the larger the $\tilde{M}$ within $28 \leq \tilde{M} \leq 64$, the better the resolution of the estimated range signature. The larger the $\tilde{M}$, however, the larger the amount of computations needed by the APES method.

Consider next an example where the true range signature is shown in Figure 4(a). We assume that there are 75 scattering centers in the signature. The number of data samples in the received data vector $\mathbf{z}$ in ((3)) is assumed to be $M = 128$. The additive thermal noise in $\mathbf{z}$ is assumed to be a zero-mean white Gaussian random process with variance 10. Figure 4(b) shows the estimated range signature with the FFT method. Note that due to the large sidelobes of the FFT method, the two small scattering centers that are to left and right of the first group of the large scattering centers are not discernible in Figure 4(b). Figure 4(c) shows the estimated range signature with the APES method with $\tilde{M} = 48$. We note that the two small scattering centers that are to left and right of the first group of the large scattering centers are resolved in Figure 4(c) since the sidelobes are very small for the APES method with $\tilde{M} = 48$. Comparing Figures 4(b) and (c), we

note that neither the FFT nor the APES method with $\tilde{M} = 48$ can resolve the three groups of very closely spaced scattering centers. Both methods tend to combine RCSs of very closely spaced scattering centers together to yield a large peak in the range signature since these scattering centers are assumed to have the same phase. Figure 4(d) shows the estimated range signature with the MODE-LSF method by assuming that there are 48 complex sinusoids plus white noise in the received data vector $\mathbf{z}$. We note that since the data model is incorrect for the MODE-LSF method and some of the estimated scatterer locations are very close to each other, using the MODE-LSF method may yield large false peaks. For this example, they occur near $t = 1.7$. Figure 4(e) shows the estimated range signature obtained by using FFT with Kaiser window and shape parameter 4. Comparing Figures 4(b) and (e), we note that using FFT with Kaiser window further widens the large peaks in 4(b).

Finally, we apply below the APES method to the experimental data measured by a ground-to-air radar. (All other information about the radar and the radar target is not available for public release.) The measured data consists of $M = 128$ samples and is also degraded with the zero-mean white Gaussian noise with variance 0.1. Figure 5(a) shows the target range signature of an aircraft obtained with the FFT method. Figure 5(b) shows the target range signature of the aircraft obtained by using the FFT method with the Kaiser window and shape parameter 4. Figure 5(c) shows the target range signature obtained with the APES method with $\tilde{M} = 48$. Comparing Figures 5(a) and (c), we note that using the APES method with $\tilde{M} = 48$ yields much lower sidelobes than using the FFT method. Comparing Figures 5(b) and (c), we note that using the APES method with $\tilde{M} = 48$ yields better resolution and sharper peaks than using the windowed FFT. Figure 5(d) shows the target range signature obtained when using the MODE-LSF method by assuming that there are 48 complex sinusoids plus white noise in the measured data. We note that using the MODE-LSF method may yield poor target range signatures due to inaccurate data models.

III. Conclusions

We have presented a robust parametric data model for estimating high resolution range signatures of radar targets. We have described how the APES (Amplitude and Phase Estimation of a Sinusoid in unknown colored noise) algorithm can be used to estimate the range signatures. With both numerical and experimental examples, we have shown that our modeling and estimation approach yields better resolution and lower sidelobes than the conventional nonparametric FFT (fast Fourier transform) method. We have also shown that our approach is more robust than modeling the radar data as a certain number of complex sinusoids in noise and estimating the frequencies, amplitudes, and phases of the sinusoids with one of the best sinusoidal parameter estimation methods.

Appendix A – Large Sample Maximum Likelihood Estimator and Cramer-Rao Bound

We consider below the estimation of the complex gain of a signal with known waveform and known steering vector in unknown colored Gaussian noise. Let $\mathbf{x}(n)$, $n = 1, 2, \cdots, N$, denote some received data vectors of dimensions $M \times 1$. Let $y(n)$, $n = 1, 2, \cdots, N$, denote a known waveform. Assume that the received data vectors $\mathbf{x}(n)$ can be written as

$$\mathbf{x}(n) = \gamma \mathbf{a}_0 y(n) + \mathbf{n}(n), \tag{17}$$

where $\mathbf{n}$ is the additive noise vector of dimensions $M \times 1$ and $\mathbf{a}_0$ is referred to as the *steering vector* and is known. The noise vectors $\mathbf{n}(n)$ are assumed to be circularly symmetric complex Gaussian random vectors with zero-mean and arbitrary covariance matrix $\mathbf{Q}$ and are temporally white, i.e.,

$$\mathrm{E}[\mathbf{n}(i)\mathbf{n}^H(j)] = \mathbf{Q}\delta_{i,j}, \tag{18}$$

where $(\cdot)^H$ denotes the complex conjugate transpose and $\delta_{i,j}$ is the Kronecker delta.

We consider below a large sample ($N \gg 1$) maximum likelihood (ML) estimator for γ from $\mathbf{x}(n)$, $n = 1, 2, \cdots, N$. It is easy to show that an exact ML estimator requires a multidimensional search over the parameter space and is computationally burdensome. We shall present below a large sample ML estimator that is both computationally and asymptotically (for large N) statistically efficient. The approach we use to derive the large sample ML estimator is similar to the one in [6, 7], which is devised for a different problem.

The log-likelihood function of the received vectors $\mathbf{x}(n)$, $n = 1, 2, \cdots, N$, is proportional to (within an additive constant):

$$-\ln|\mathbf{Q}| - \mathrm{tr}\left\{\mathbf{Q}^{-1}\frac{1}{N}\sum_{n=1}^{N}[\mathbf{x}(n) - \mathbf{b}y(n)][\mathbf{x}(n) - \mathbf{b}y(n)]^H\right\},$$

where $|\cdot|$ denotes the determinant of a matrix and

$$\mathbf{b} = \gamma\mathbf{a}_0. \tag{19}$$

Consider first the estimate of $\mathbf{Q}$ and the unstructured estimate of $\mathbf{b}$. It is easy to show that

$$\hat{\mathbf{Q}} = \frac{1}{N}\sum_{n=1}^{N}[\mathbf{x}(n) - \hat{\mathbf{b}}y(n)][\mathbf{x}(n) - \hat{\mathbf{b}}y(n)]^H, \tag{20}$$

and $\hat{\mathbf{b}}$ may be obtained by minimizing the following cost function

$$F = \left|\frac{1}{N}\sum_{n=1}^{N}[\mathbf{x}(n) - \mathbf{b}y(n)][\mathbf{x}(n) - \mathbf{b}y(n)]^H\right|. \tag{21}$$

Let

$$\hat{\mathbf{r}}_{y\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}y(n)\mathbf{x}^H(n), \tag{22}$$

and

$$\hat{r}_{yy} = \frac{1}{N}\sum_{n=1}^{N}y(n)y^H(n). \tag{23}$$

Let $\hat{\mathbf{R}}_{\mathbf{xx}}$ be defined similarly as $\hat{r}_{yy}$. Then let [8]

$$\begin{aligned}
\mathbf{G} &= \frac{1}{N}\sum_{n=1}^{N}[\mathbf{x}(n) - \mathbf{b}y(n)][\mathbf{x}(n) - \mathbf{b}y(n)]^H && (24)\\
&= \hat{\mathbf{R}}_{\mathbf{xx}} - \mathbf{b}\hat{\mathbf{r}}_{y\mathbf{x}} - \hat{\mathbf{r}}_{y\mathbf{x}}^H\mathbf{b}^H + \mathbf{b}\hat{r}_{yy}\mathbf{b}^H && (25)\\
&= \left[\mathbf{b} - \hat{\mathbf{r}}_{y\mathbf{x}}^H\hat{r}_{yy}^{-1}\right]\hat{r}_{yy}\left[\mathbf{b} - \hat{\mathbf{r}}_{y\mathbf{x}}^H\hat{r}_{yy}^{-1}\right]^H + \hat{\mathbf{R}}_{\mathbf{xx}} - \hat{\mathbf{r}}_{y\mathbf{x}}^H\hat{r}_{yy}^{-1}\hat{\mathbf{r}}_{y\mathbf{x}}. && (26)
\end{aligned}$$

Since $\hat{r}_{yy}$ is a positive number and the second and third terms in ((26)) do not depend on $\mathbf{b}$, it follows that

$$\mathbf{G} \geq \mathbf{G}\,|_{\,\mathbf{b}=\hat{\mathbf{b}}}, \tag{27}$$

where

$$\hat{\mathbf{b}} = \hat{\mathbf{r}}_{y\mathbf{x}}^H\hat{r}_{yy}^{-1}. \tag{28}$$

Since the whole sample covariance matrix $\mathbf{G}$ is minimized, the unstructured estimate $\hat{\mathbf{b}}$ of $\mathbf{b}$ in ((28)) will minimize any nondecreasing function of $\mathbf{G}$ including the determinant of $\mathbf{G}$, which is F in ((21)). It is easy to see that $\hat{\mathbf{b}}$ is a consistent estimate of $\mathbf{b}$.

By using ((28)) with ((20)), the $\hat{\mathbf{Q}}$ may be rewritten as

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}}_{\mathbf{xx}} - \hat{\mathbf{r}}_{y\mathbf{x}}^H \hat{r}_{yy}^{-1} \hat{\mathbf{r}}_{y\mathbf{x}}. \tag{29}$$

It is easy to see that $\hat{\mathbf{Q}}$ is a consistent estimate of $\mathbf{Q}$.

Let us now consider the structure of $\mathbf{b}$. The cost function in ((21)) may be rewritten as

$$\begin{aligned} F &= \left| \hat{\mathbf{R}}_{\mathbf{xx}} - \mathbf{b}\hat{\mathbf{r}}_{y\mathbf{x}} - \hat{\mathbf{r}}_{y\mathbf{x}}^H \mathbf{b}^H + \mathbf{b}\hat{r}_{yy}\mathbf{b}^H \right| && (30) \\ &= \left| \hat{\mathbf{R}}_{\mathbf{xx}} - \hat{\mathbf{b}}\hat{r}_{yy}\hat{\mathbf{b}}^H + (\mathbf{b}-\hat{\mathbf{b}})\hat{r}_{yy}(\mathbf{b}-\hat{\mathbf{b}})^H \right| && (31) \\ &= \left| \hat{\mathbf{Q}} \right| \left| \mathbf{I} + \hat{\mathbf{Q}}^{-1}(\mathbf{b}-\hat{\mathbf{b}})\hat{r}_{yy}(\mathbf{b}-\hat{\mathbf{b}})^H \right|. && (32) \end{aligned}$$

The ML estimate of γ may be obtained by minimizing F in ((32)) or equivalently $\ln F$. In [6, 7], we have shown that minimizing $\ln F$, with F defined in ((32)), is asymptotically (for large N) equivalent to minimizing

$$F_1 = \mathrm{tr}\left[\hat{r}_{yy}(\mathbf{b}-\hat{\mathbf{b}})^H \hat{\mathbf{Q}}^{-1}(\mathbf{b}-\hat{\mathbf{b}}) \right]. \tag{33}$$

Minimizing F_1 in ((33)) yields

$$\hat{\gamma} = \frac{\mathbf{a}_0^H \hat{\mathbf{Q}}^{-1} \hat{\mathbf{b}}}{\mathbf{a}_0^H \hat{\mathbf{Q}}^{-1} \mathbf{a}_0}. \tag{34}$$

The large sample ML estimator for estimating γ may be summarized as follows:

Step 1: Compute $\hat{\mathbf{b}}$ and $\hat{\mathbf{Q}}$ with ((28)) and ((29)), respectively.

Step 2: Determine $\hat{\gamma}$ with ((34)).

We remark that since $\hat{\gamma}$ is a consistent and large sample realization of the ML estimate, it follows that $\hat{\gamma}$ is asymptotically (for large N) statistically efficient according to the general properties of ML estimators [9]. Using the results in [6, 7], we can show that the asymptotic (for large N) distribution of $\hat{\gamma}$ is complex

Gaussian with mean γ and variance

$$\mathrm{Var}[\hat{\gamma}] = \frac{1}{\sum_{n=1}^{N} \mathbf{y}^H(n)\mathbf{a}_0^H \mathbf{Q}^{-1}\mathbf{a}_0\mathbf{y}(n)}, \tag{35}$$

which is also the CRB.

References

[1] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, pp. 1132–1143, July 1990.

[2] P. Stoica and K. C. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proceedings, Pt. F*, vol. 137, pp. 19–26, February 1990.

[3] M. Viberg, "Sensitivity of parametric direction finding to colored noise fields and undermodeling," *Signal Processing*, vol. 34. pp. 207–222, 1993.

[4] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 996–1000, July 1989.

[5] A. Nehorai, D. Starer, and P. Stoica, "Direction-of-arrival estimation in applications with multipath and few snapshots," *Circuits, Systems, and Signal Processing*, vol. 10, pp. 327–342, 1991.

[6] J. Li, B. Halder, P. Stoica. M. Viberg, and T. Kailath, "Decoupled maximum likelihood angle estimation for signals with known waveforms," Technical Report No. CTH-TE-8, Chalmers University of Technology, Gothenburg, Sweden, February 1994.

[7] J. Li, B. Halder, P. Stoica, and M. Viberg, "Decoupled maximum likelihood angle estimation for signals with known waveforms," *submitted to IEEE Transactions on Signal Processing.*

[8] T. Söderström and P. Stoica, *System Identification.* London, U.K.: Prentice-Hall International, 1989.

[9] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York, NY: John Wiley & Sons Inc., 1968.

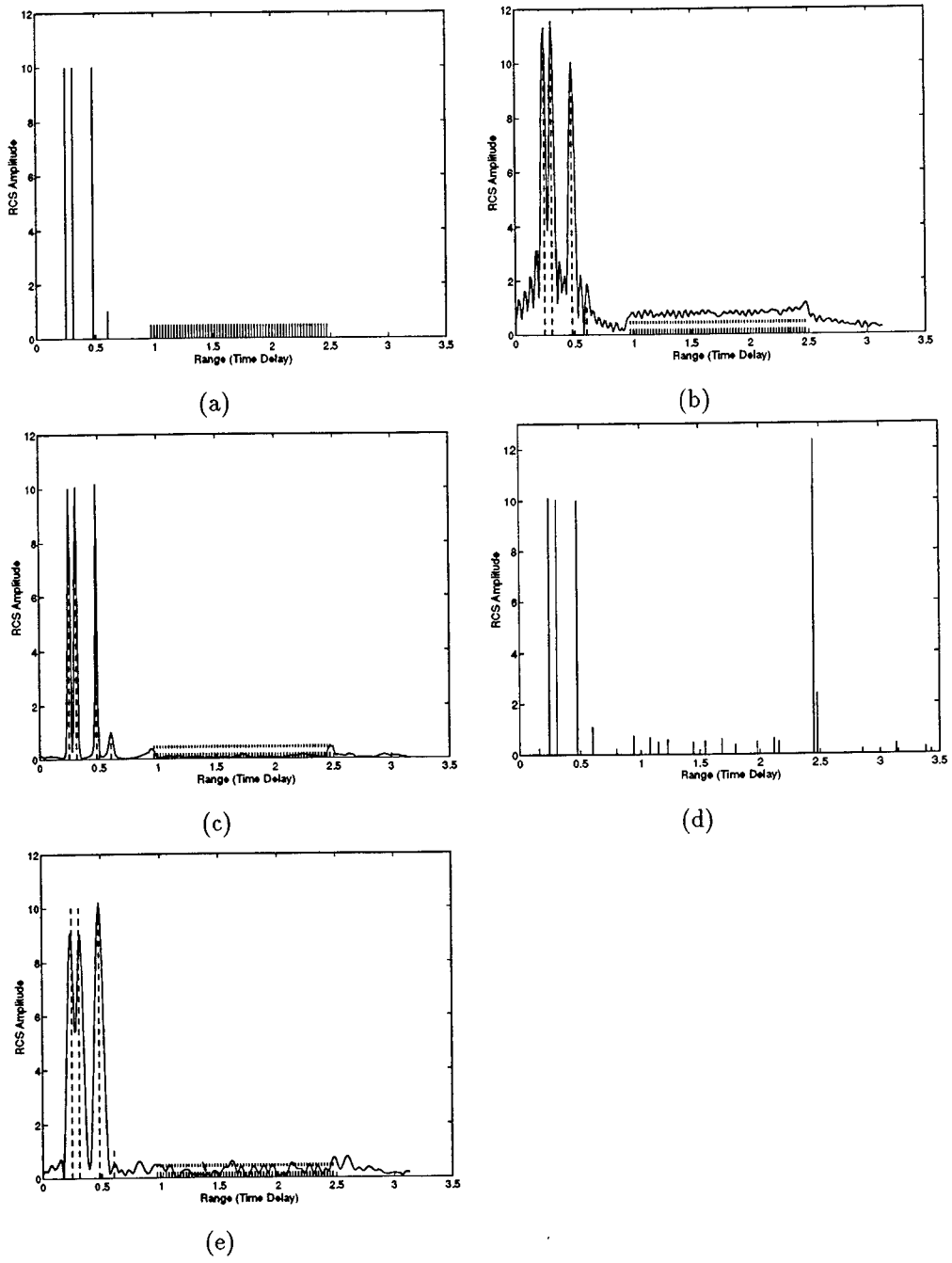


Figure 1: Range signature estimates (solid lines in (b) – (e)) compared to the true range signature (dashed lines in (b) – (e)) when the white noise variance is 1 and $M = 128$. (a) True range signature. (b) FFT or APES with $\tilde{M} = 1$. (c) APES with $\tilde{M} = 48$. (d) MODE-LSF by assuming 48 complex sinusoids. (e) FFT with Kaiser window and shape parameter 4.

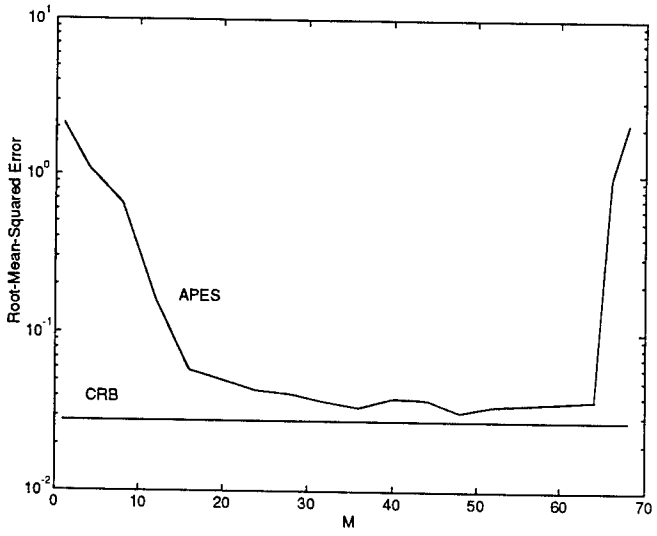
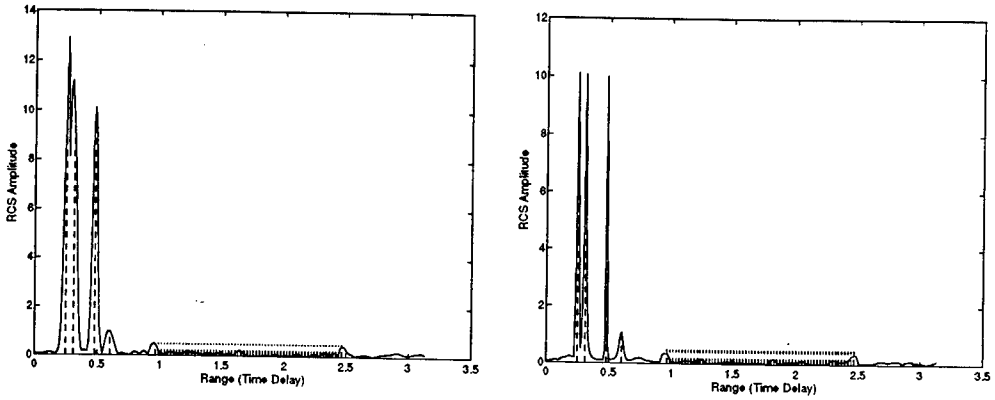


Figure 2: Root-Mean-squared errors (MSEs) of $\hat{\gamma}(r = 0.2332)$ obtained with the APES estimator as a function of $\tilde{M}$ for the example in Figure 1.



(a) (b)

Figure 3: Range signature estimates (solid lines) compared to the true range signature (dashed lines) when the white noise variance is 1 and $M = 128$. (a) APES with $\tilde{M} = 28$. (b) APES with $\tilde{M} = 64$.

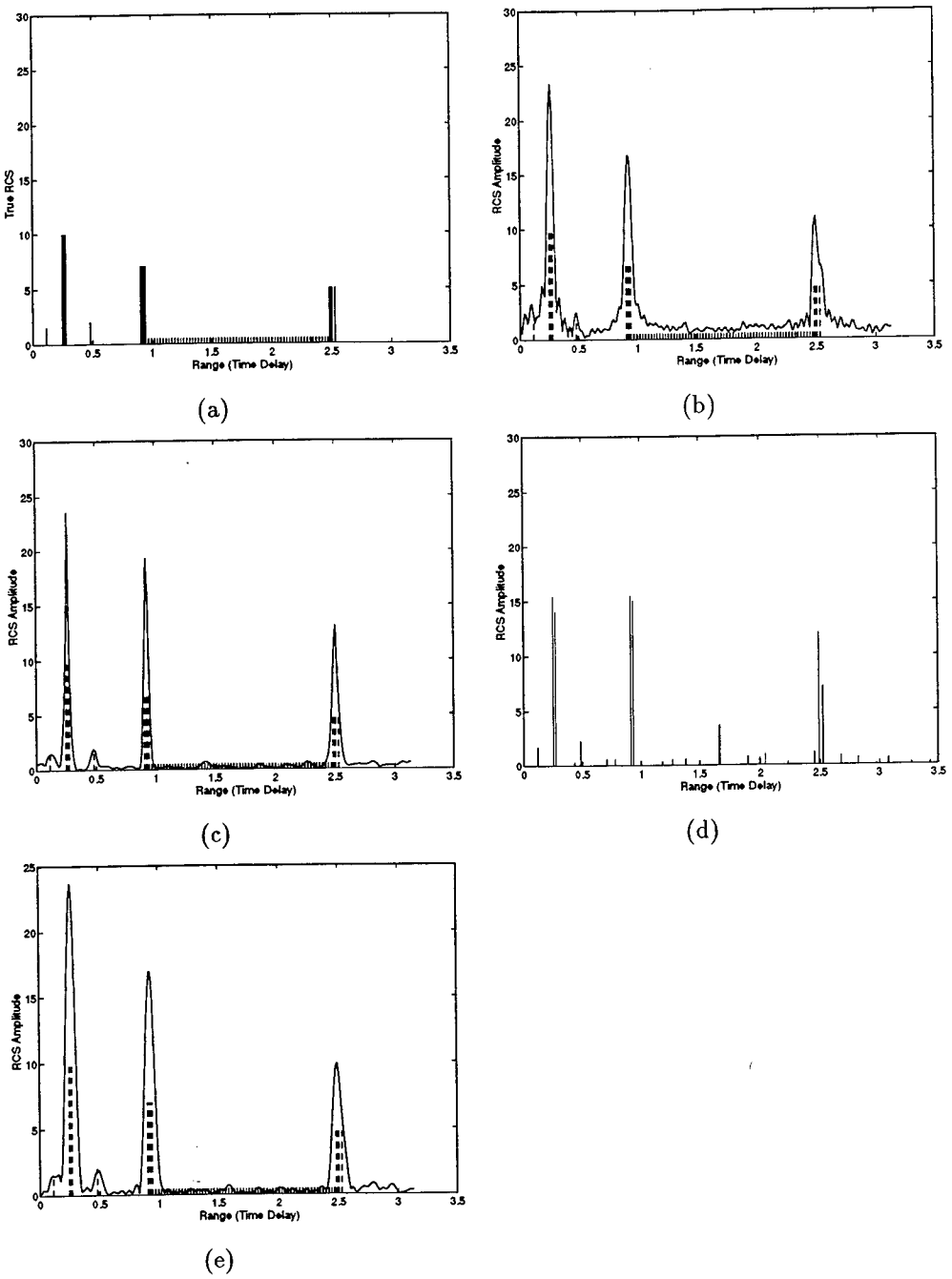


Figure 4: Range signature estimates (solid lines in (b) – (e)) compared to the true range signature (dashed lines in (b) – (e)) when the white noise variance is 10 and $M = 128$. (a) True range signature. (b) FFT or APES with $\tilde{M} = 1$. (c) APES with $\tilde{M} = 48$. (d) MODE-LSF by assuming 48 complex sinusoids. (e) FFT with Kaiser window and shape parameter 4.

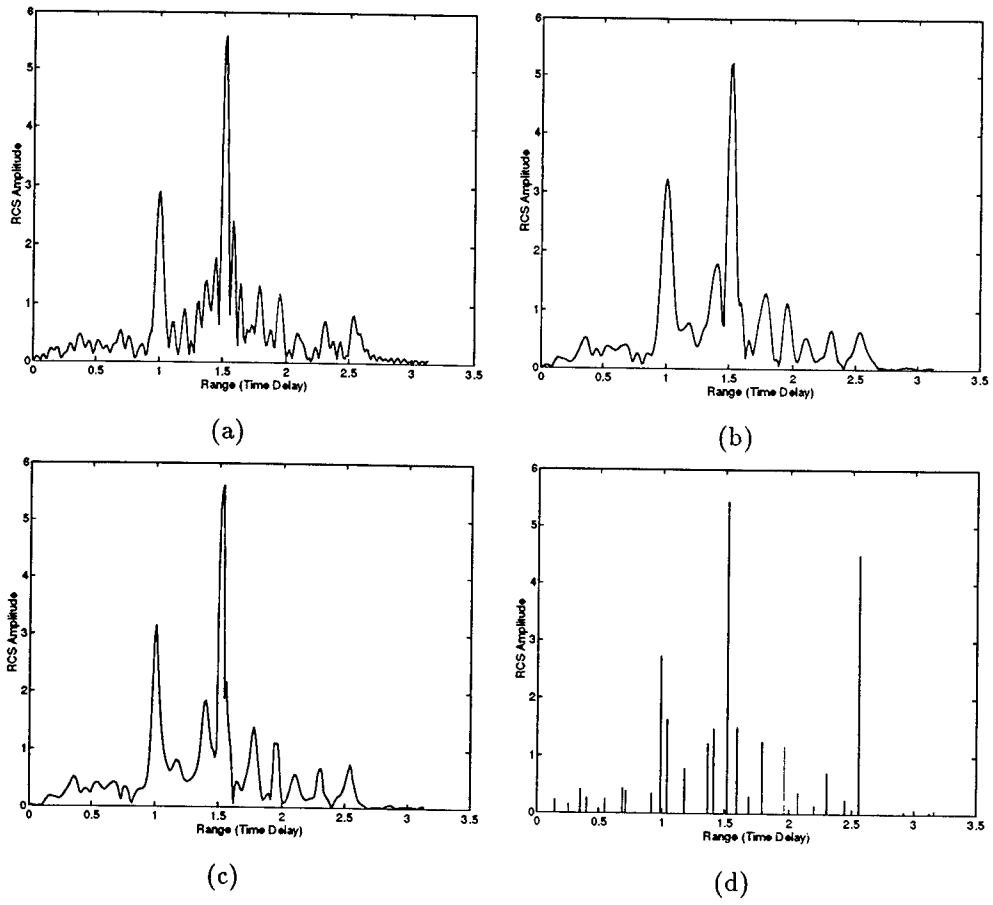


Figure 5: Range signature estimates of an airplane measured with a ground-to-air radar. (a) FFT or APES with $\tilde{M} = 1$. (b) FFT with Kaiser window and shape parameter 4. (c) APES with $\tilde{M} = 48$. (d) MODE-LSF by assuming 48 complex sinusoids.

PREDICTION OF MISSILE TRAJECTORY

Chun-Shin Lin
Department of Electrical and Computer Engineering

University of Missouri-Columbia
Columbia, MO 65211

Final Report for
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory
Pilot/Vehicle Interface Technology Branch
Wright-Patterson AFB, Dayton OH

August 1994

PREDICTION OF MISSILE TRAJECTORY

Chun-Shin Lin
Department of Electrical and Computer Engineering
University of Missouri-Columbia

Abstract

This study concerns the survivability of an airship that is under attack by a missile. The pilot may employ an appropriate maneuver or countermeasure at the right moment to escape from the attack. Real-time prediction of the missile trajectory will help prepare the pilot to take action at the right moment to survive. The missiles considered have a short boosting period followed by a non-powered period. The proposed technique takes some missile trajectory data at the early stage of the non-powered period to estimate two of the most important missile parameters for predicting the future trajectory. The complete flying course of the target airship must be known for the prediction. Nine sets of trajectory data, three sets from each of three different types of missiles, are available for this study. One of these nine is not used because the impact happens during the boosting period. The developed technique generated good prediction results on these missile flyout data. One merit of the proposed technique is that no priori knowledge about the missile is needed. Indeed, throughout this study, no information about these missiles has been made available to this researcher. Prediction for the earlier trajectory of the missile including the boosting period has not been studied but relevant issues have been addressed and discussed in the introduction and discussion sections.

SUMMARY OF NOTATIONS

t:	current time
$\mathbf{V}_M(t)$:	missile velocity vector at t
$V_M(t)$:	missile speed at t
$\mathbf{V}_T(t)$:	target velocity vector at t
$V_T(t)$:	target speed at t
$\mathbf{P}_M(t)$:	missile position vector at t
$\mathbf{P}_T(t)$:	target position vector at t
$\mathbf{R}_{TM}(t)$:	line-of-sight vector at t; it equals $\mathbf{P}_T(t)$ - $\mathbf{P}_M(t)$
$R_{TM}(t)$:	length of the vector $\mathbf{R}_{TM}(t)$
$\mathbf{V}_{TM}(t)$:	relative velocity between the target and the missile at t; it equals $\mathbf{V}_T(t)$ - $\mathbf{V}_M(t)$
$V_c(t)$:	closing rate ($= -\dot{R}_{TM}(t)$)
$A_c(t)$:	command acceleration for the missile at t
$\omega_s(t)$:	derivative of line-of-sight angle at t (called the line-of-sight rate)
M:	mass of the missile; the variable cancels out in our final equations
fric:	friction parameter of the missile; estimated from the missile trajectory data
α:	navigation constant; estimated from the missile trajectory data
g:	gravity (32.15ft/sec^2)
$h(t)$:	altitude of the missile at t; the third component of $\mathbf{P}_M(t)$.
Δt:	sampling period for missile and target trajectory data (50ms in this study)
$\mathbf{k}_N$:	the unit vector that is perpendicular to $\mathbf{R}_{TM}(t)$ and lies on the plane consisting of $\mathbf{R}_{TM}(t)$ and $\mathbf{R}_{TM}(t+\Delta t)$

PREDICTION OF MISSILE TRAJECTORY

Chun-Shin Lin

1. INTRODUCTION

This study concerns the survivability of an airship that is under attack by a missile. The pilot may employ an appropriate maneuver or countermeasure at the right moment to escape from the attack. Real-time prediction of the missile trajectory will help prepare the pilot to take action at the right moment to survive [1].

The missiles considered have a short boosting period followed by a non-powered period. Figure 1 shows the typical velocity curve for these two periods. In this study, we consider the prediction of the missile trajectory only for the non-powered period. Some sampled trajectory data after the motor burns out are used to estimate two of the most important missile parameters for predicting the future missile trajectory. The complete flying course of the target airship must be known for the prediction. Nine sets of data, three sets from each of three different types of missiles, are available for this study. One of these nine is not used because the impact happens before the motor burns out. Precisely predicting the complete missile trajectory using just a small set of trajectory data during the boosting period is not possible unless the missile type can be identified. The length of the boosting period and the acceleration curve must be known for reasonably good prediction. However, less accurate trajectory prediction is possible by estimating the acceleration and the navigation constant. If the trajectory data can be continuously obtained until a short period after the motor burns out, the prediction can be continuously updated to give satisfactory results.

In this study on the prediction for the non-powered period, the parameters identified are the friction parameter and the navigation constant in the proportional navigation guidance. Whether the missile has its motor burned out or not can be detected by checking whether it is accelerating or decelerating. Once the motor burns out, the friction, the control technique, and the flying course of the target determine how the missile will fly. With the friction parameter and the navigation constant identified, and the target trajectory given, the future missile trajectory can be predicted.

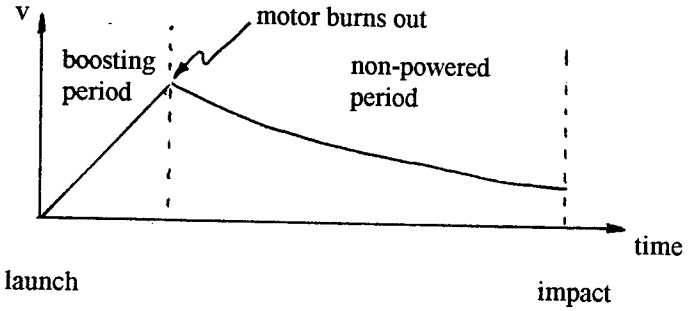


Figure 1. A typical missile velocity curve

No information about the studied missiles has been made available to this researcher. The developed technique has generated good prediction results on these missile flyout data without a priori information about the missiles.

2. IDENTIFICATION OF PARAMETERS

This section discusses how the friction effect and the navigation constant are estimated. We assume that n position vectors and n velocity vectors for both the missile and the target are available for the parameter estimations.

2.1 Friction

The friction coefficient is assumed to be a constant in this study. Thus the basic equation for the missile velocity $V_M(t)$ can be expressed as

$$V_M(t+\Delta t) = V_M(t) + fric * V_M(t) * \Delta t. \quad (1)$$

where "*fric*" is referred to as the friction parameter in this report and has a negative value.

The possible change of the missile altitude will affect the velocity. The change of the potential energy will be converted into the kinetic energy. Thus the velocity change from this factor can be solved from the following relationship:

$$1/2\ [(M\ V_M^2(t+\Delta t) - M\ V_M^2(t)] = - Mg[h(t+\Delta t) - h(t)] \quad (2)$$

where M is the mass of the missile, g equals to 32.15ft/sec^2, and $h(t)$ is the altitude of the missile at time t. With $V_M(t+\Delta t) \equiv V_M(t) + \Delta V_M(t)$, the following can be derived from (2):

$$\Delta V_M(t) \approx g[h(t+\Delta t) - h(t)]/V_M(t) \quad (3)$$

Note that the mass in (2) cancels out and does not appear in (3). With this factor added, (1) can be modified as

$$V_M(t+\Delta t) = V_M(t) + fric * V_M(t) * \Delta t + g[h(t+\Delta t) - h(t)]/V_M(t). \quad (4)$$

From the missile flyout data, we can also observe the correlation between the friction and the angular velocity of the missile. In this study, although such effect was not considered, our prediction results seem good. Whether the effect can be modelled and estimated, and whether the inclusion of this will improve the prediction accuracy and reliability require additional investigation.

With $V_M(t+\Delta t)$, $V_M(t)$, $h(t+\Delta t)$, and $h(t)$ given, *fric* can be computed from (4). Thus, theoretically, only the trajectory data at two time instants, t and $t+\Delta t$, are needed to evaluate this parameter. However, with possible measurement errors, the average value computed over a longer time period is preferred. If the altitude and velocity of the missile at n consecutive sampled time instants are available, one can compute the mean value of *fric* solved from (4) as

$$fric = (1/n)\sum_{k=1}^{n}\{V_M(t+k\Delta t)-V_M(t+(k-1)\Delta t)-g[h(t+k\Delta t)-h(t+(k-1)\Delta t)]/V_M(t+(k-1)\Delta t)\}/[V_M(t+(k-1)\Delta t) * \Delta t] \quad (5)$$

2.2 Navigation Constant

How a missile behaves relies on its control scheme. This study focuses on proportional navigation guidance because most missiles use the technique. The following statement regarding the popularity of this guidance technique is made in Chapter 2 of [2]:

"The Lark missile , which had its first successful test in December 1950, was the first missile to use proportional navigation. Since that time proportional navigation guidance has been used in virtually all of the world's tactical radar, infrared(IR), and television (TV) guided missiles."

To briefly introduce the proportional navigation guidance technique, we can first consider the two-dimensional case. Figure 2 shows a two-dimensional missile-target engagement geometry. The line from the missile to the target is called the line-of-sight (LOS). θ is the LOS angle and its time derivative is called the LOS rate. The LOS distance vector is denoted by $\mathbf{R}_{TM}$ and its length is indicated by R_{TM} (a bold-face character for a vector and an Italic one for a scalar). V_T and V_M are the speeds of the target and the missile, respectively. A_c is the acceleration applied perpendicular to the line-of-sight. The proportional navigation control intends to make the missile heading rate (angular velocity) linearly proportional to the LOS rate. The purpose is to approximate a constant bearing course, that keeps the LOS angle constant. One proportional navigation law [2] is to issue acceleration commands, perpendicular to the instantaneous LOS, which are proportional to the LOS rate and closing speed. The law can be mathematically expressed as

$$A_c = \alpha V_c \omega_s \quad (7)$$

where α is the control gain known as the navigation constant, ω_s is the line-of-sight rate, and V_c is the closing speed which equals to $-\dot{R}_{TM}(t)$.

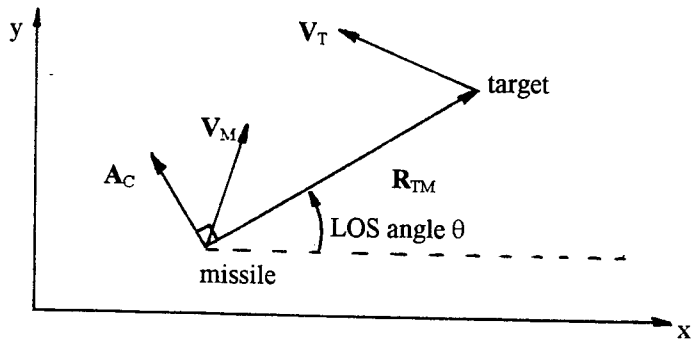


Figure 2 Two-dimensional missile-target engagement geometry

In control, (7) is used to determine the needed acceleration. But for estimation of α, A_c, V_c and ω_s are computed from the measured missile trajectory data, and α is computed from (7). This constant can be computed from $A_c(t)$, $V_c(t)$ and $\omega_s(t)$ at a single specific t. However, for the same reason mentioned in estimating "*fric*", to reduce the effects of possible measurement errors, it is preferred to have α evaluated as an average of $A_c/(V_c\omega_s)$ from the data over a longer time period.

The real missile-target engagement is in three dimensional space. We need to deal with vectors. In the following, we will focus on computation of ω_s, A_c and V_c. Again, bold-faced symbols will be used to denote vectors. Given some trajectory data, with the use of the law of cosines, the line-of-sight rate can be estimated as (see Figure 3)

$$\omega_s(t) = \cos^{-1}\{[(R_{TM}^2(t) + R_{TM}^2(t-\Delta t) - |\mathbf{R}_{TM}(t) - \mathbf{R}_{TM}(t-\Delta t)|^2]/[2R_{TM}(t)R_{TM}(t-\Delta t)]\}/\Delta t \quad (8)$$

where $|\cdot|$ indicates the vector length. Note that $R_{TM}(t)$, $R_{TM}(t-\Delta t)$ and $|\mathbf{R}_{TM}(t) - \mathbf{R}_{TM}(t-\Delta t)|$ are lengths of three sides of a triangle. The angular rate is about the axis perpendicular to both $\mathbf{R}_{TM}(t-\Delta t)$ and $\mathbf{R}_{TM}(t)$.

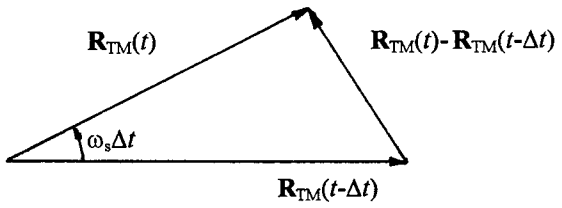


Figure 3. Relationship between ω_s and $\mathbf{R}_{TM}$ vectors

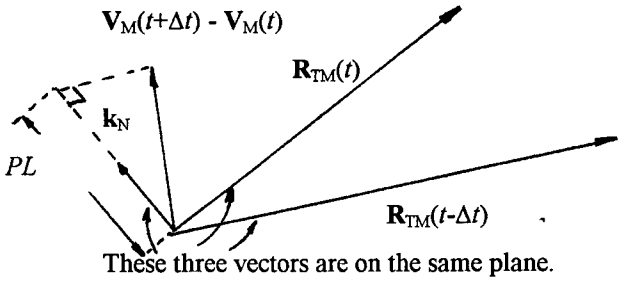


Figure 4. Illustration of vectors and the length used for computing $A_c(t)$.

Now lets turn our attention to A_c. The change of missile velocity direction is caused by the acceleration command A_c. The direction of $(\mathbf{V}_M(t+\Delta t) - \mathbf{V}_M(t))$, the same as that for $A_c(t)$, is supposed to be perpendicular to $\mathbf{R}_{TM}(t)$ and on the plane consisting of $\mathbf{R}_{TM}(t)$ and $\mathbf{R}_{TM}(t-\Delta t)$, *i.e.*, along the $\mathbf{k}_N$ axis as indicated in Figure 4. While the change of the measured velocity may not be on the direction of $\mathbf{k}_N$, for parameter estimation using this proposed technique, we project $\mathbf{V}_M(t+\Delta t) - \mathbf{V}_M(t)$ to the unit vector $\mathbf{k}_N$. The projection can be done by the following computations:

$$\mathbf{k}_N = \{[\mathbf{R}_{TM}(t-\Delta t) \times \mathbf{R}_{TM}(t)] \times \mathbf{R}_{TM}(t)\} / |\{[\mathbf{R}_{TM}(t-\Delta t) \times \mathbf{R}_{TM}(t)] \times \mathbf{R}_{TM}(t)\}| \quad (9)$$

$$\text{projected length } PL = \mathbf{k}_N \cdot (\mathbf{V}_M(t+\Delta t) - \mathbf{V}_M(t)) \quad (10)$$

The length is divided by Δt to give A_c for estimating the navigation constant, *i.e.*,

$$A_c(t) = PL/\Delta t \quad (11)$$

Another needed quantity for estimation is the closing velocity V_c, which is easy to obtain. With a small effort in derivation, V_c can be obtained as

$$\begin{aligned} V_c(t) &= -\partial |\mathbf{R}_{TM}(t)| / \partial t \\ &= [\mathbf{R}_{TM}(t) \cdot \mathbf{V}_{TM}(t)] / |\mathbf{R}_{TM}(t)|. \end{aligned} \quad (12)$$

Given n target and missile flying data samples, the navigation constant, α, can be estimated by the following steps:

1. Compute $\omega_s(t+k\Delta t)$, $A_c(t+k\Delta t)$ and $V_c(t+k\Delta t)$ for $k = 1,2 \ldots n$ using (8)-(12).

2. Compute α as $\alpha = \frac{1}{n}\sum_{k=1}^{n} [A_c(t+k\Delta t)/(V_c(t+k\Delta t)\omega_s(t+k\Delta t))]$

A summary showing the needed step-by-step computation for obtaining $A_c(t)$, $V_c(t)$ and $\omega_s(t)$ is given in Appendix A.

3. PREDICTION OF MISSILE TRAJECTORY AND THE OVERALL PROCEDURE

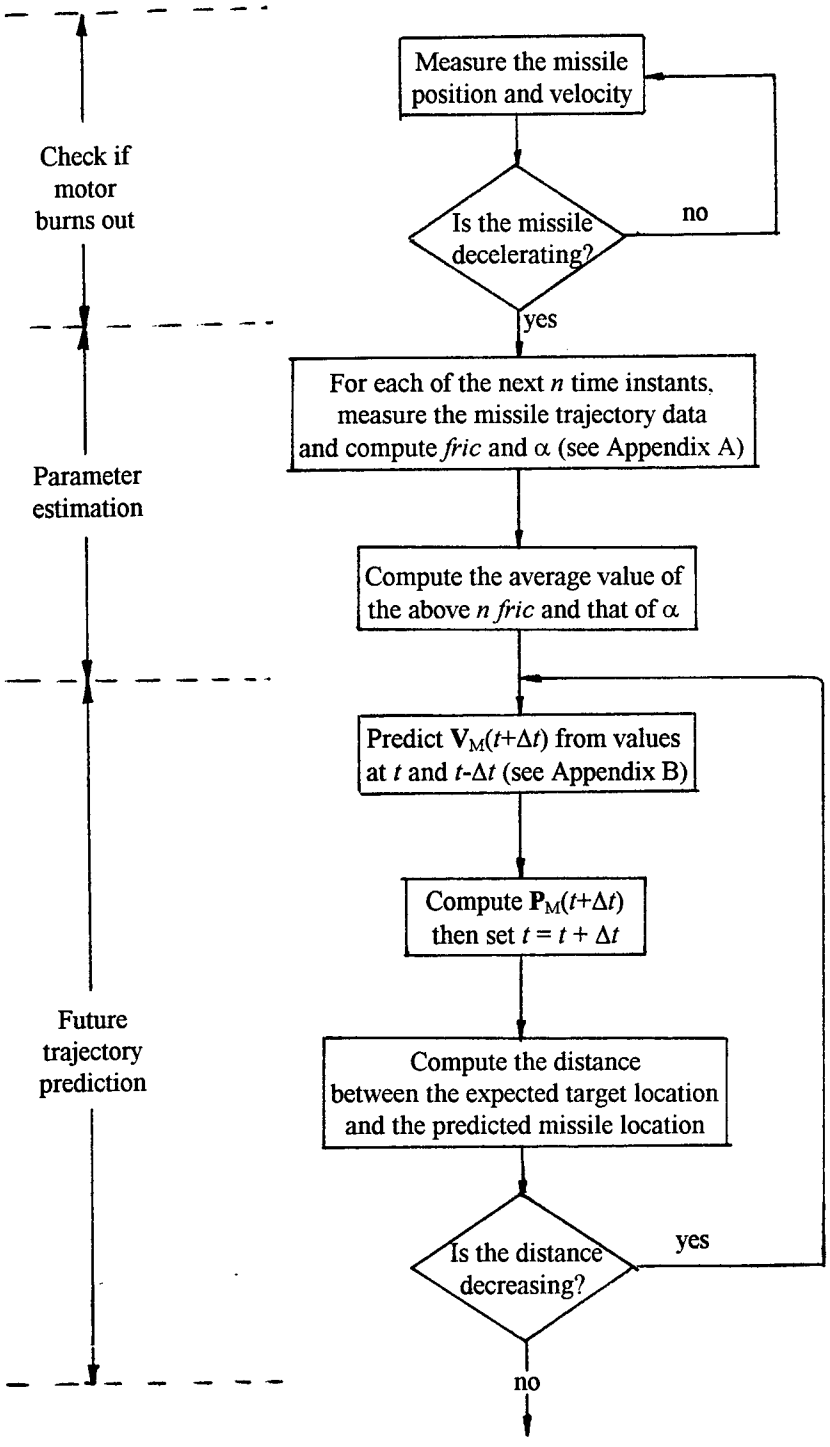
3.1 Prediction of Missile Trajectory

With the friction parameter "*fric*", the navigation constant α and the target trajectory known, the predicted missile trajectory can be computed. This section briefs the computation procedure.

The new missile speed after every Δt period can be estimated using (4). To determine the vector velocity, $A_c(t)$ must be first computed by (7). Equations (8) and (12) are used to compute $\omega_s(t)$ and $V_c(t)$. The acceleration A_c is along the vector $\mathbf{k}_N$, which is perpendicular to $\mathbf{R}_{TM}(t)$ and on the plane consisting of $\mathbf{R}_{TM}(t)$ and $\mathbf{R}_{TM}(t-\Delta t)$ as stated before. The new velocity vector can be computed as the sum of the previous velocity and the acceleration vector times by Δt. The new velocity vector is then adjusted to have the length equal to the speed computed using (4). A procedure can be easily developed to compute the velocity and position of the missile step by step along the time. The complete set of equations used for the computations are given in Appendix B to show the procedure.

3.2 Overall Estimation and Prediction Procedure

The following diagram provides an overall procedure for estimation and prediction:


Measure the missile position and velocity
Is the missile decelerating?
no
yes
For each of the next n time instants, measure the missile trajectory data and compute fric and α (see Appendix A)
Compute the average value of the above n fric and that of α
Predict $\mathbf{V}_M(t+\Delta t)$ from values at t and $t-\Delta t$ (see Appendix B)
Compute $\mathbf{P}_M(t+\Delta t)$ then set $t = t + \Delta t$
Compute the distance between the expected target location and the predicted missile location
Is the distance decreasing?
yes
no
Check if motor burns out
Parameter estimation
Future trajectory prediction
(having passed the closest position)

4. EXPERIMENTS AND RESULTS

The proposed technique has been implemented in Turbo C and applied to eight sets of missile flyout data provided by Wright Laboratory. This section presents the details regarding the data, experiments and results.

4.1 Missile Flyout Data

Nine sets of missile flyout data were provided by Wright Laboratory for usage in this study. Those data were obtained from simulations for three different types of missiles. The nine sets of data are named M1RUN1, M1RUN2, ... M3RUN3, of which each specifies the missile type and the simulation run. Among these nine, M1RUN1 has the impact happen before the motor burns out. This is not the situation considered in this study. Thus that set of data was discarded and only the other eight were used.

Each data file consists of the velocities and positions of the target and the missile for the complete period from launching to impact. Velocities and positions are represented in three coordinates: North, East and Down. Figures 5(a) through 12(a) show the plots of missile and target trajectories for the eight engagements.

Observations from the data indicate that the motor burning periods for the three types of missiles are about 5.4 seconds, 10 seconds and 7.5 seconds respectively. The maximum velocities are 37,000, 44,000 and 38,000 ft/sec. Information about these missiles and their control techniques are not available to this researcher. Trajectories for the first two types are pretty normal while those for the third one look trickier. Figure 12(a) shows that the missile tries to turn into east away from the target at the early stage. There may be an offset added on purpose to the actual target position in control. This may be done just to create the confusion on which object is the real target. In this example, the offset could be about 4,000ft in the east direction and 10,000 ft up in the altitude. However, in the last 12-15 seconds, the offset seems removed and the missile turns back toward the target.

4.2 Experiment Setup

The motor burning periods are 5.4, 10 and 7.5 seconds. Thus our experiments simply discarded the first 10 seconds of data. The data of the next 5 seconds were used to estimate the friction parameter and the navigation constant. The identified parameter values were used to calculate the predicted missile trajectories for the rest of the time period. Note that the parameter estimation was done only once using the data from a 5-second period.

4.3 Experimental Results

Trajectory Curves

All the predicted results for the eight cases are provided in Figures 5(b) to 12(b). In each plot, there are three curves showing the target trajectory, the real missile trajectory and the predicted missile trajectory. In most plots, the real and predicted missile trajectories are too close to be distinguishable. The difference between the predicted one and the real one in Figure 6(b) seems noticeable but is actually insignificant because the unit used in that figure is "foot" (thousand feet in most of other plots). All others in Figures 5(b) to 8(b) show indistinguishable curves. The largest errors happen in Figure 12(b) for the case M3RUN3. This may be caused by the target

position offset mentioned previously. The parameters were estimated at a time when the missile was not really heading toward the target. However, one may be surprised that the predicted intercept time is very close to the real one.

As mentioned earlier that the estimation of the parameters was done only once using the data between the period of 10-15 seconds. If we can continuously obtain new estimates using the most current trajectory data, the accuracy will be improved and the problem in M3RUN3 will be eventually removed.

Estimated Parameters

One will expect that the estimated friction parameters and the navigation constants for the same type of missiles should be close to each other. The following table shows the estimated results from all runs:

run name	M1RUN2	M1RUN3	M2RUN1	M2RUN2	M2RUN3	M3RUN1	M3RUN2	M3RUN3
fric	-.0616	-.0608	-.0342	-.0394	-.0354	-.0184	-.0174	-.0308
α	7.33	5.78	9.17	9.46	7.59	2.79	2.92	1.77

Data for the same type of missiles are close to each other. The deviations are within 20-25% except the ones for M3RUN3. The larger deviations for the M3RUN3 case are believed to be caused by the same reason stated earlier.

Time of Impact

Knowing when the impact will occur can be important. An airplane can sharply turn away from the missile-target plane or employ a countermeasure shortly before the impact to efficiently escape from the attack [3].

Our plots show that the time of impact for the real one and the predicted one are close. We estimate the time at which the predicted distance from the missile to the target is smallest. The difference between the actual time of impact and the predicted one for each run is listed in the following table:

run name	M1RUN2	M1RUN3	M2RUN1	M2RUN2	M2RUN3	M3RUN1	M3RUN2	M3RUN3
time error (sec)	0	0.05 (lag*)	0	0.05 (lag)	1.0 (lag)	.25 (lag)	.15 (lag)	.15 (lead*)

*lag: predicted one occurs later; lead: predicted one occurs earlier.

5. DISCUSSIONS AND REMARKS

About the technique

This study shows that the trajectory can be well predicted. Estimation of the friction parameter and the navigation constant can be done from a small set of data points. This computational load for parameter estimation is light. Calculation of the predicted trajectory can be implemented into a procedure in which Δt determines the number of compuational loops needed. Using a larger Δt reduces the computational burden but, on the other hand, degrades the precision of prediction. For practical implementation, the best value needs to be decided.

Computational efficiency is one important issue but has not been investigated yet. Whether the computation speed is a problem in practical application or not must be checked. If it is, then whether the problem should be solved in software or hardware needs to be answered. Some computations in parameter estimation and trajectory prediction involve vector operations. Devices with special vector operation capabilities may help provide faster computation.

The prediction results are less reliable if the total time period predicted is long. This is reasonable and expected. In this study, we have assumed that only a small set of data are observed. The prediction is then done all the way until the time of impact. The scheme gives the prediction based on a small piece of information available just following motor burnout. Prediction can be updated if new real data are continuously provided. This will for sure make the prediction more reliable and precise.

Extension and Relevant Issues

While the technique provides pretty good results, there is still plenty of space for improvement. Only the constant friction and navigation constant are estimated and used in this study. Further improvement is very likely by having some more details. The precision of the predicted trajectory and time of intercept may be improved. Further improvement may be achieved by having additional parameters and updating the prediction by using the most recent trajectory information.

Providing predicted information starting from the time of launching can be done. During the boosting period, the acceleration can be measured and the navigation constant can be estimated. Some assumptions on the missile such as the maximum velocity, the length of boosting period and the friction parameter must be made. The prediction can be good for the boosting period and will start to deviate from the actual one after the motor burns out. To obtain the reliable prediction, trajectory data must be continuously measured and provided until some data from the non-powered period are obtained.

A relevant problem is to guess which airship is the most likely target of a missile if there are many airships around. Finding out the target earlier will help in creating the prompt optimum reaction. This may be done by carefully evaluating the trajectories of the potential target and the missiles. Some ideas used in this study can be extended for developing that kind of technique.

6. CONCLUSION

This study investigated the possibility of predicting the missile trajectory and the time of impact. The purpose is to obtain information in order to give the pilot enough time to well prepare for the best action. Since the missile trajectory is determined by the flying course of the target, the planned target path must be known to this prediction system. The proposed prediction technique only predicts the trajectory for the non-powered period. During that period, the trajectory is mainly determined by the friction, the navigation constant used in proportional navigation guidance as well as the target trajectory. Thus, procedures were developed to estimate the friction

parameter and the navigation constant at the early stage of the non-powered period. Tests on eight sets of data from three different missiles show promising results. Future extension should on the real implementation, development of technology that utilizes the predicted information for optimal defense, continuing improvement of the prediction accuracy, etc.. A relevant problem is to earlier find out the target from a group of airships based on the missile trajectory. There must be correlation between the missile trajectory and the target trajectory. Some ideas and techniques can be extended for that usage.

REFERENCES

1. Halski, D. J., R. J. Landy and R. P. Meyer, "Low Risk Evaluation and Medium Risk Development," WL-TR-94-3026, Wright Laboratory, Wright-Patterson AFB. (An ICAAS report).

2. Zarchan, Paul, *Tactical and Strategic Missile Guidance* (Volume 124 of Progress in Astronautics and Aeronautics). American Institute of Aeronautics and Astronautics, Inc. 1990.

3. Shaw, Robert L., *Fighter Combat*, Naval Institute Press, Annapolis, Maryland. 1985

ACKNOWLEDGEMENTS

The study is supported by the Air Force Summer Faculty Research Program. This project was initialized by Major Peter Raeth. His help and valuable discussion throughout the project period are greatly appreciated. I would also like to thank Mr. Robert Whitmoyer for his valuable suggestions and discussion, and Mr. Terry Christian and Mr. James Zeh for providing the missile flyout data used in this study.

APPENDIX A. Step-by-step computation for estimating the navigation constant α

Given

the target positions $\mathbf{P}_T(t)$, $\mathbf{P}_T(t\text{-}50ms)$

the missile positions $\mathbf{P}_M(t)$, $\mathbf{P}_M(t\text{-}50ms)$

the target velocity $\mathbf{V}_T(t)$

the missile velocities $\mathbf{V}_M(t)$, $\mathbf{V}_M(t+50ms)$

To compute $\omega_s(t)$, $A_c(t)$, and $V_c(t)$, and then α.

Summary of step by step computation from the above given values:

Compute

$\mathbf{R}_{TM}(t) = \mathbf{P}_T(t) - \mathbf{P}_M(t)$

$\mathbf{R}_{TM}(t\text{-}50ms) = \mathbf{P}_T(t\text{-}50ms) - \mathbf{P}_M(t\text{-}50ms)$

$\mathbf{V}_{TM}(t) = \mathbf{V}_T(t) - \mathbf{V}_M(t)$

compute $\mathbf{R}_{TM}(t) - \mathbf{R}_{TM}(t\text{-}50ms)$

* $\omega_s(t) \approx \cos^{-1}\{[(R_{TM}^2(t) + R_{TM}^2(t\text{-}50ms) - |\mathbf{R}_{TM}(t) - \mathbf{R}_{TM}(t\text{-}50ms)|^2$

$/[2R_{TM}(t)R_{TM}(t\text{-}50ms)]\}/50ms$ (8)

compute $\mathbf{V}_M(t+50ms) - \mathbf{V}_M(t)$

$\mathbf{k}_N = \{[\mathbf{R}_{TM}(t\text{-}50ms) \times \mathbf{R}_{TM}(t)] \times \mathbf{R}_{TM}(t)\} /$

$|\{[\mathbf{R}_{TM}(t\text{-}50ms) \times \mathbf{R}_{TM}(t)] \times \mathbf{R}_{TM}(t)\}|$ (9)

projected length $PL = \mathbf{k}_N \cdot (\mathbf{V}_M(t+50ms) - \mathbf{V}_M(t))$ (10)

* $A_c(t) = PL/50ms$ (11)

* $V_c(t) = [\mathbf{R}_{TM}(t) \cdot \mathbf{V}_{TM}(t)]/|\mathbf{R}_{TM}(t)|$

* compute $A_c(t)/[V_c(t)\omega_s(t)]$ for computing α.

APPENDIX B. A step-by-step example showing the computation for predicted missile trajectory

Given

the target positions $\mathbf{P}_T(t)$, $\mathbf{P}_T(t\text{-}50ms)$

the missile positions $\mathbf{P}_M(t)$, $\mathbf{P}_M(t\text{-}50ms)$

the target velocity $\mathbf{V}_T(t)$

the missile velocities $\mathbf{V}_M(t)$

two parameters *fric* and α.

To compute the predicted value of $\mathbf{V}_M(t+50ms)$ and $\mathbf{P}_M(t+50ms)$.

Step by step computation from the above given values:

Compute

$\mathbf{R}_{TM}(t) = \mathbf{P}_T(t) - \mathbf{P}_M(t)$

$\mathbf{R}_{TM}(t\text{-}50ms) = \mathbf{P}_T(t\text{-}50ms) - \mathbf{P}_M(t\text{-}50ms)$

$\mathbf{V}_{TM}(t) = \mathbf{V}_T(t) - \mathbf{V}_M(t)$

compute $\mathbf{R}_{TM}(t) - \mathbf{R}_{TM}(t\text{-}50ms)$

* $\omega_s(t) \approx \cos^{-1}\{[(R_{TM}^2(t) + R_{TM}^2(t\text{-}50ms) - |\mathbf{R}_{TM}(t) - \mathbf{R}_{TM}(t\text{-}50ms)|^2$

$/[2R_{TM}^2(t)R_{TM}^2(t\text{-}50ms)]\}/50ms$ (8)

$\mathbf{k}_N = \{[\mathbf{R}_{TM}(t) \times \mathbf{R}_{TM}(t\text{-}50ms)] \times \mathbf{R}_{TM}(t)\} /$

$|\{[\mathbf{R}_{TM}(t) \times \mathbf{R}_{TM}(t\text{-}50ms)] \times \mathbf{R}_{TM}(t)\}|$ (9)

* $V_c(t) = [\mathbf{R}_{TM}(t)\cdot\mathbf{V}_{TM}(t)]/|\mathbf{R}_{TM}(t)|$

$A_c(t) = \alpha V_c(t)\omega_s(t)$

$\mathbf{V}_M(t+50ms) = \mathbf{V}_M(t) + A_c(t)\mathbf{k}_N$ (11)

$V_M(t+50ms) = V_M(t) + fric * V_M(t) * 50ms + g[h(t+50ms) - h(t)]/V_M(t).$ (4)

* $\mathbf{V}_M(t+50ms) = \{\mathbf{V}_M(t+50ms)/|\mathbf{V}_M(t+50ms)|\}V_M(t+50ms)$

* $\mathbf{P}_M(t+50ms) = \mathbf{P}_M(t) + \mathbf{V}_M(t+50ms)*50ms$

The newly predicted $\mathbf{V}_M$ and $\mathbf{P}_M$ are then used to compute the next values.

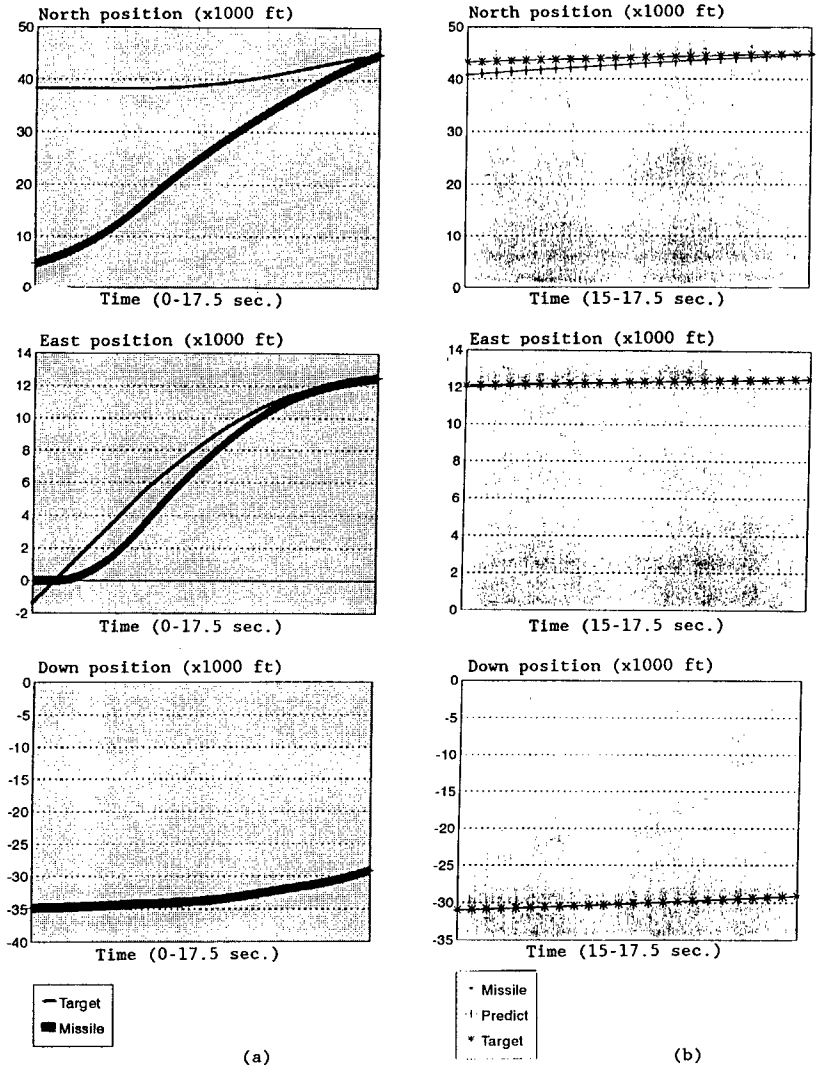


Figure 5. For M1RUN2: (a) Complete trajectories (b) Predicted results

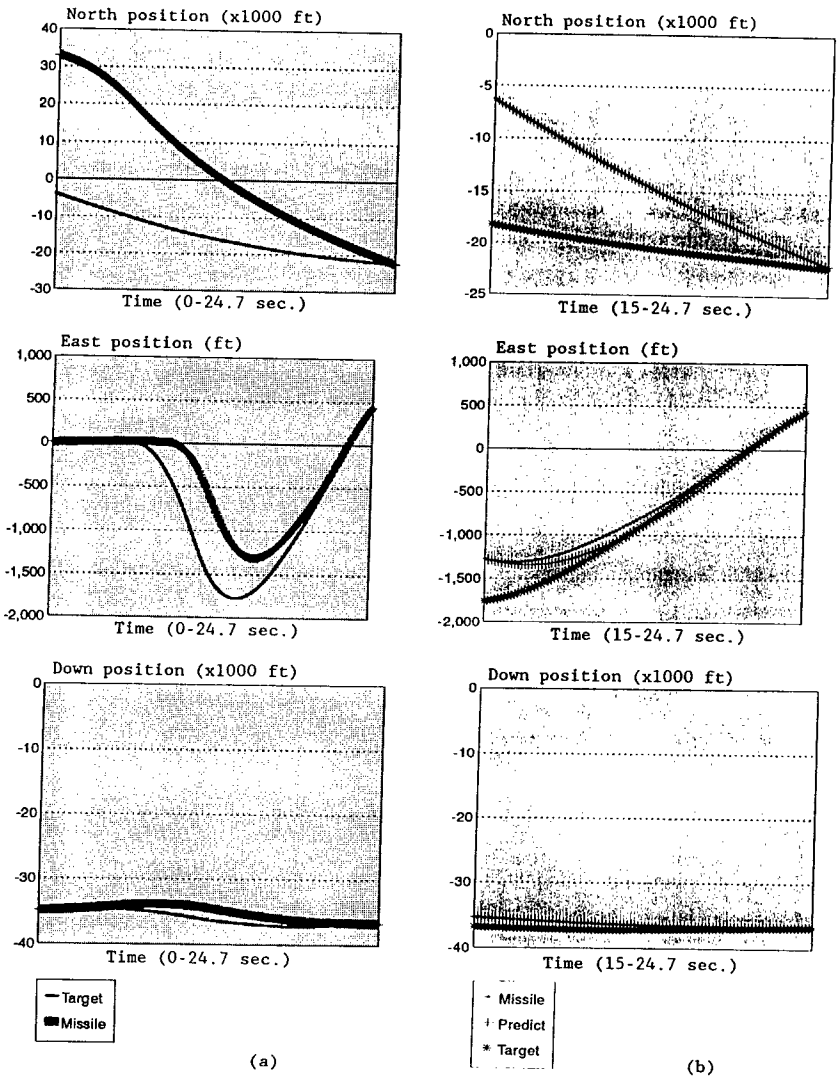


Figure 6. For M1RUN3: (a) Complete trajectories (b) Predicted results

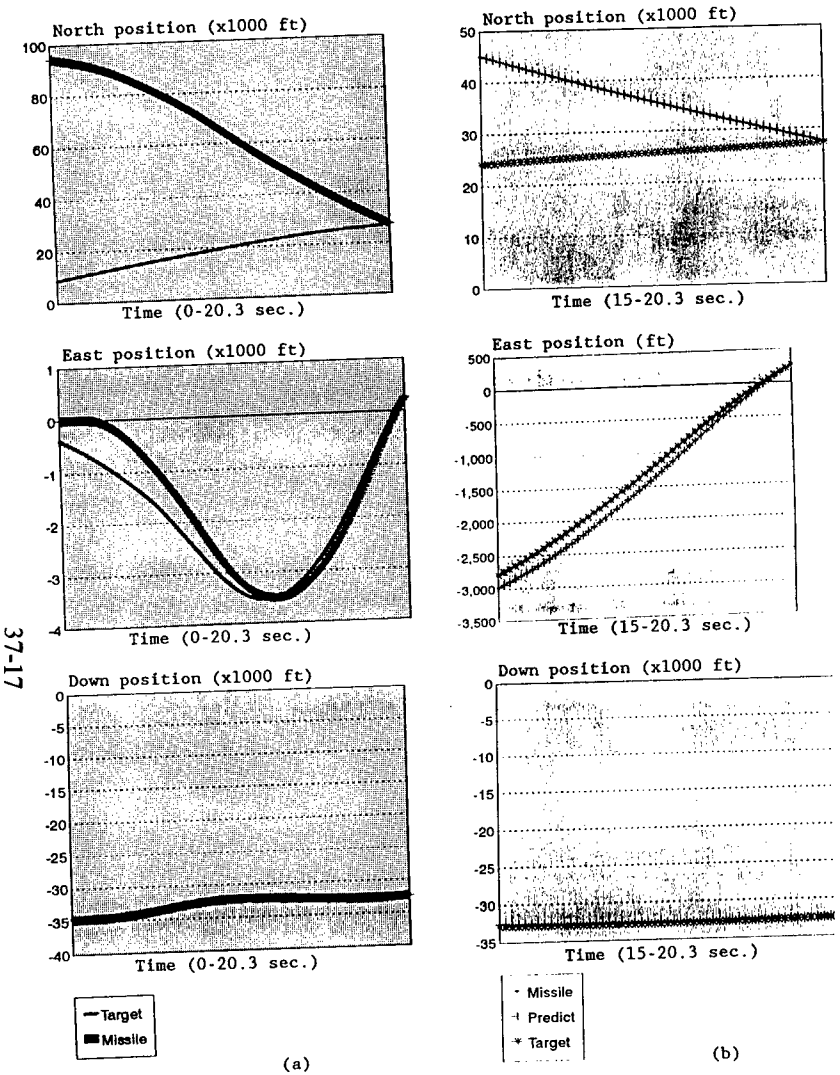


Figure 7. For M2RUN1: (a) Complete trajectories (b) Predicted results

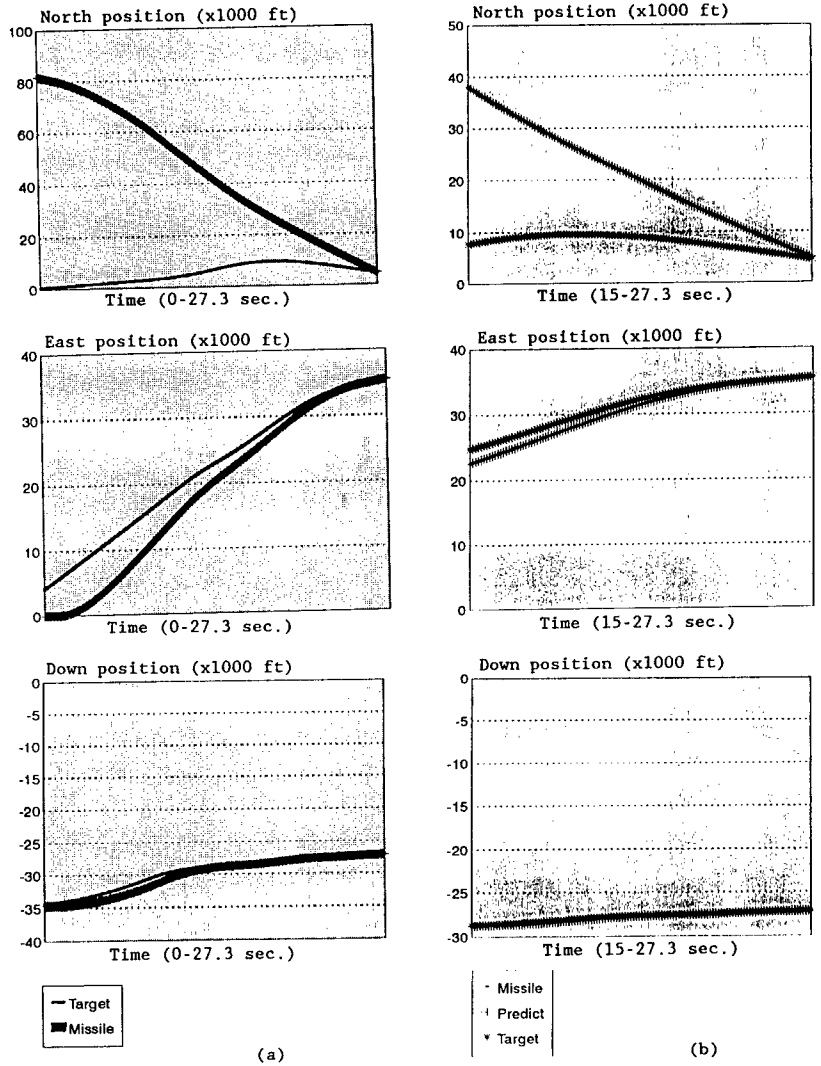


Figure 8. For M2RUN2: (a) Complete trajectories (b) Predicted results

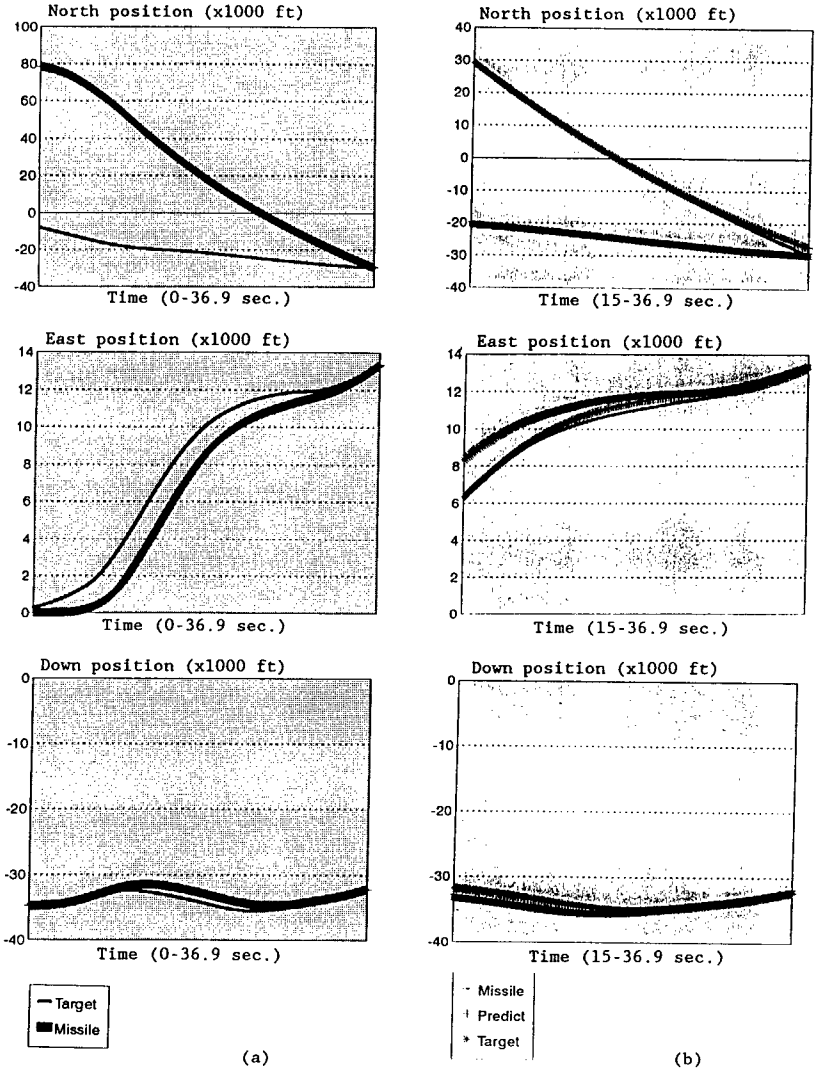


Figure 9. For M2RUN3: (a) Complete trajectories (b) Predicted results

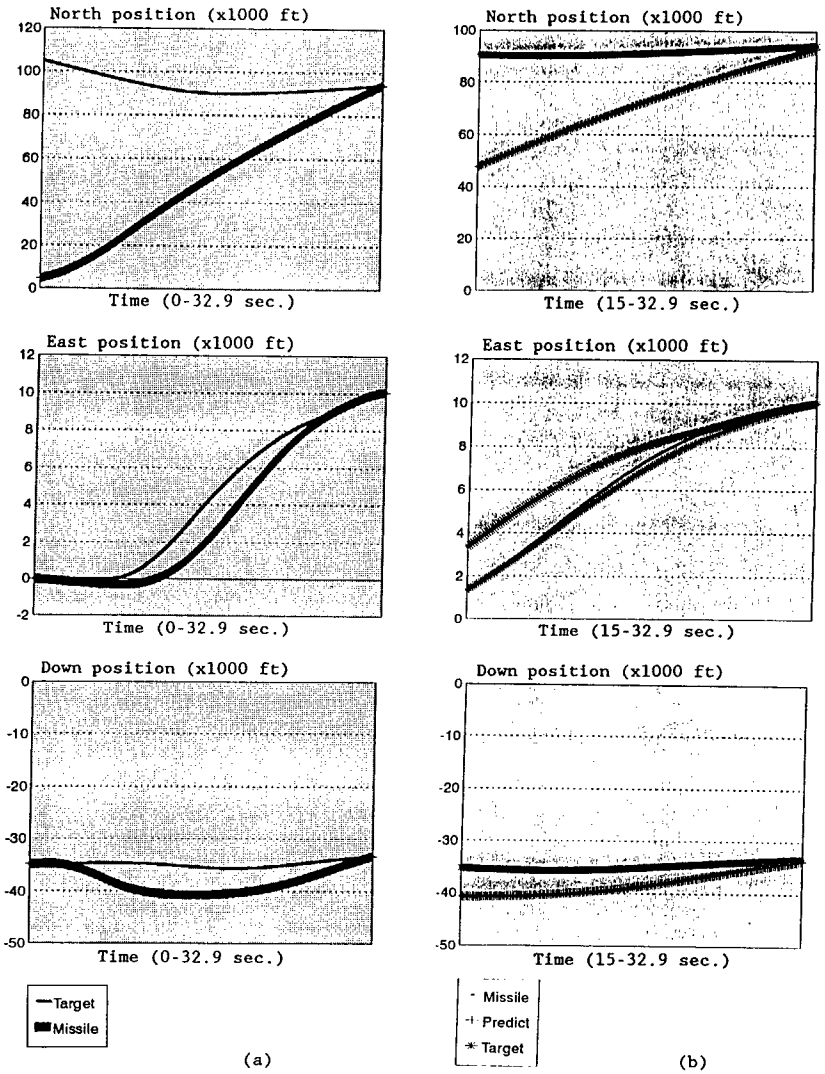


Figure 10. For M3RUN1: (a) Complete trajectories (b) Predicted results

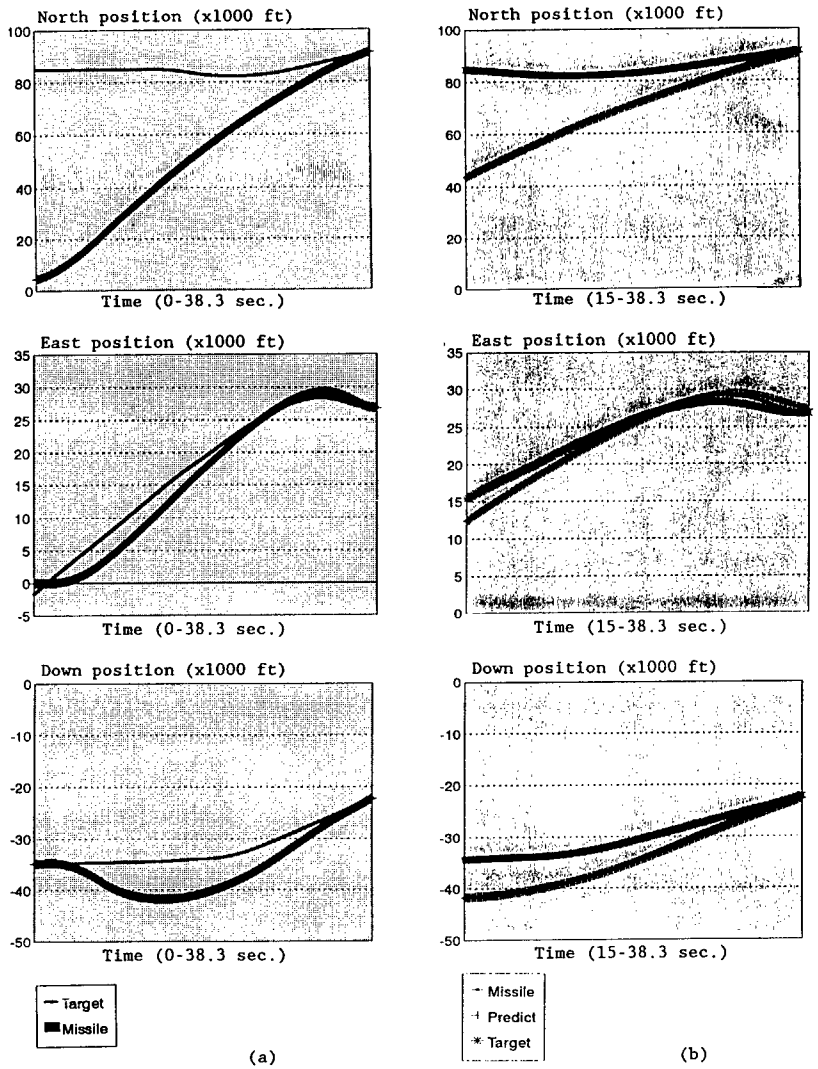


Figure 11. For M3RUN2: (a) Complete trajectories (b) Predicted results

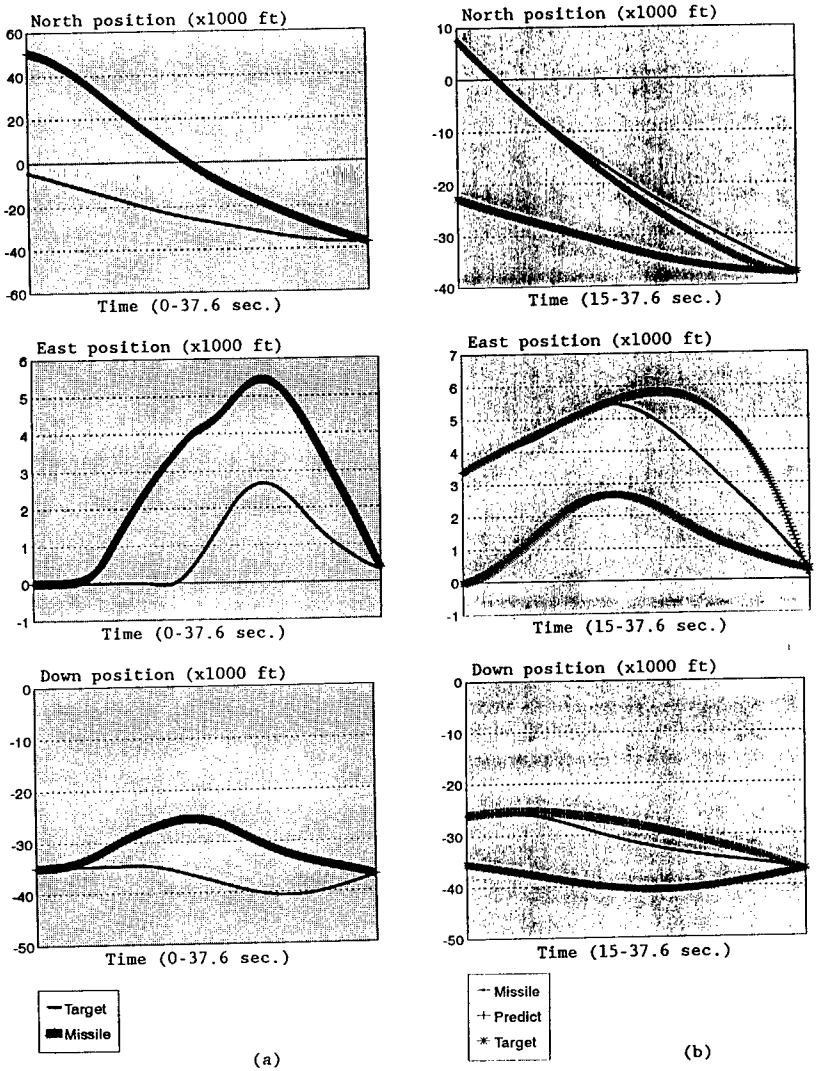


Figure 12. For M3RUN3: (a) Complete trajectories (b) Predicted results

THREE DIMENSIONAL GEOMETRY MEASUREMENT OF TIRE DEFORMATION USING AN OPTICAL TECHNIQUE

Paul P. Lin
Associate Professor
Mechanical Engineering Department
Cleveland State University
Cleveland, OH 44115

Final Report for:
Summer Research Program
Wright Laboratory

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

and

Wright Laboratory

August, 1994

THREE DIMENSIONAL GEOMETRY MEASUREMENT OF TIRE DEFORMATION BY MEANS OF AN OPTICAL TECHNIQUE

Paul P. Lin
Associate Professor
Mechanical Engineering Department
Cleveland State University

Abstract

The main objective of this research was to apply an optical technique called fringe projection to quantifying aircraft tire deformation and strains. Unlike Moire Fringes, the proposed technique using a single light source and only one grating requires no image superposition. As a result, the measurement is not as sensitive to vibration as the Moire method does. The other objective was to compare the magnitudes of three dimensional deformation between two types of F16 aircraft tires made of distinct tire cords, namely Bias and Radial, subjected to loading conditions such as with flat plate and flywheel, different percentages of tire deflection and different yaw angles. Since the optical technique used is based on the geometrical change relative to the selected reference point within the tire, a close-range fiber optic displacement sensor was installed to accurately detect the point's height change in the direction parallel to the wheel axle. The experimental results indicates that with the measuring system and the proposed technique, the tire deformation can be quantified and the three dimensional geometry of a deformed tire can be reconstructed.

THREE DIMENSIONAL GEOMETRY MEASUREMENT OF TIRE DEFORMATION BY MEANS OF AN OPTICAL TECHNIQUE

Paul P. Lin

Introduction

In non-contact measurement, several optical techniques are available. Laser ranging can yield a dense set of depth values with which surface structure can be obtained through surface fitting or approximation (Vemuri and Aggarwal, 1984). This technique, however, is usually slow and expensive. Stereo vision utilizes the disparity between the projected positions of a point in two images to infer the depth of this point (Marr and Poggio, 1976). But the correspondence between points in the stereo images is difficult to establish, and the computation is sensitive to errors introduced in digitization and camera calibration. The well known Projection Moire technique uses a white light or laser light source and two gratings of the same pitch (one in front of light and the other one in front of camera) to generate Moire interference patterns. The image is recorded in a single CCD camera, in lieu of two cameras used in stereo vision. Another very similar technique, Shadow Moire, uses only one grating near the object to generate the Moire patterns. The Moire contours thus obtained, however, do not make a difference between peaks and pits unless prior information or additional algorithms are applied. Furthermore, this technique is very sensitive to vibration due to the necessity of superimposing two images. The most accurate optical technique available today is phase-shifting interferometry (PSI). It takes time to generate three or four consecutive phase shifts to form interferograms, and thereby rendering PSI not useful for measurement of dynamic motion. This technique, by nature is also very sensitive to vibration. Another consideration is that Moire technique requires the use of two very fine pitch gratings (usually over 250 lines per inch), which makes it very difficult to visualize the generated Moire pattern on a low reflectivity aircraft tire. As a result, it is often necessary to paint the tire surface in light color.

The proposed fringe projection technique (Lin and Parvin, 1990; Lin, et.

al.,1991) uses a single light source and only one grating in front of light projector to generate optical fringes. No image superposition is required. In comparison with the Moire or phase-shifting technique, the fringe projection technique is less sensitive to vibration and much more computationally efficient. Lin, et. al., 1994 used an optical technique to measure the tire deformations of different aircraft bias tires: F-16, F-111 and KC-135. This research extends the scope of the previous work to 3-D deformations considering the effect of yaw angle. In this report, the measuring system and methodology for tire deformation and strains are described, and the results of three dimensional tire deformation analysis are presented. The difference of deformation between bias and radial tires is compared. Finally, the conclusions of this research and the recommendations for the future work are made.

METHODOLOGY

The measuring system consists of

(1) Optical equipment: White light projector, grating, optical rails, CCD camera and close-range fiber-optic displacement sensor.
(2) Image acquisition equipment: Frame grabber, image acquisition and processing software.
(3) Data acquisition equipment: Dual-channel digital data storage oscilloscope.
(4) High Speed Flash: Microseconds flash duration with a few nanoseconds response time.
(5) Synchronization Device: Synchronize the tire rotation with the flash.
(6) Recording equipment: Super VHS video recorder (VCR).
(7) Computing equipment: 486-based micro-computer and RGB monitor.

The light produced by the white light projector passes through the grating (Ronchi ruling) and illuminate the tire surface (see Figure 1). The image captured by the CCD camera is recorded in the VCR, and then transmitted to the frame grabber where the image data are digitized and processed. The digital

image is then displayed in a high resolution RGB monitor. The frame grabber and CCD camera both have the same resolution of 512 by 480 pixels. The highest shutter speed available in this camera is 100 micro seconds. The camera, frame grabber and VCR's frame rate is 1/30 seconds.

The principle of the 3-D optical measurement used here is based on the curvature change of projected fringes and the spacing between two adjacent fringes. The first step in image analysis is to accurately detect the locations of fringe centers, line by line. Usually the fringe center locations are limited to the so-called one-pixel resolution which is provided by the CCD camera. In this research, the so-called sub-pixel resolution was employed in order to improve the fringe center detection accuracy. It should be noted that three dimensional geometry determination greatly depends upon the accuracy of fringe centers. When performing image analysis, fringe centers are scanned from top to bottom and left to right (see Figure 2).

It is necessary to specify the location of the reference point within the tire and near the wheel flange. In addition, a reference plane (xy plane) passing through this point and perpendicular to the viewing direction has to be established. It is worthwhile to note that when a tire is loaded, not only the location (x and y components) of the reference point changes, the height (z component) of the point (i.e. perpendicular to the sidewall) changes as well. In this study, a close-range fiber-optic sensor was installed to accurately measure the reference point's height change when a tire is loaded. During the tire loading process, this sensor moved with the tire in order to keep the detecting position constant.

A wobbling motion occurs when a tire rotates on a flywheel. The wobble becomes very noticeable particularly when the tire is loaded. To overcome this problem, a rotational motion synchronization device was designed to turn on the high speed flash exactly once per revolution so that same points of interest in the tire can be captured by the CCD camera. A reflective tape is placed on the reference point and the probe of the fiber-optic sensor is then focused on the tape. Initially, the height of this point is detectable. As the tire rotates the tape goes away from the fiber-optic light beam that the height cannot be

detected because the distance from the probe to the target goes out of range. In this case, the measuring voltage is set to zero. As the tire comes back exactly to the initial angular position, the voltage will be back to the initial highest value. Since the voltage generated by the fiber-optic sensor changes so fast that it is necessary to use a dual-channel digital data storage oscilloscope to store the digital data for later examination. Any voltage near the peak value can be used to trigger the flash. Thus, the flash is fired only when necessary, once per revolution. The flash takes a few nanoseconds of response time and 9 microseconds of flash duration operating at the power of 60 Watts. This flash, due to an unexpected delay from the manufacturer, was not available at the time of testing. Nevertheless, the synchronization device worked well when tested using the digital oscilloscope.

The acquired images were filtered and analyzed with and without the use of grating. Without grating, two points (one inch apart in the radial direction) were marked with white dots every 15° (Fig. 3). The displacements of these points were traced when the tire was subjected to different loads and yaw angles. In addition, strains can be calculated. These displacements, however, are only two dimensional. With grating, optical fringes were generated and from which the fringe centers were accurately detected (Fig. 4). To determine the three dimensional geometry of tire deformation, the in-house developed computing algorithm was employed.

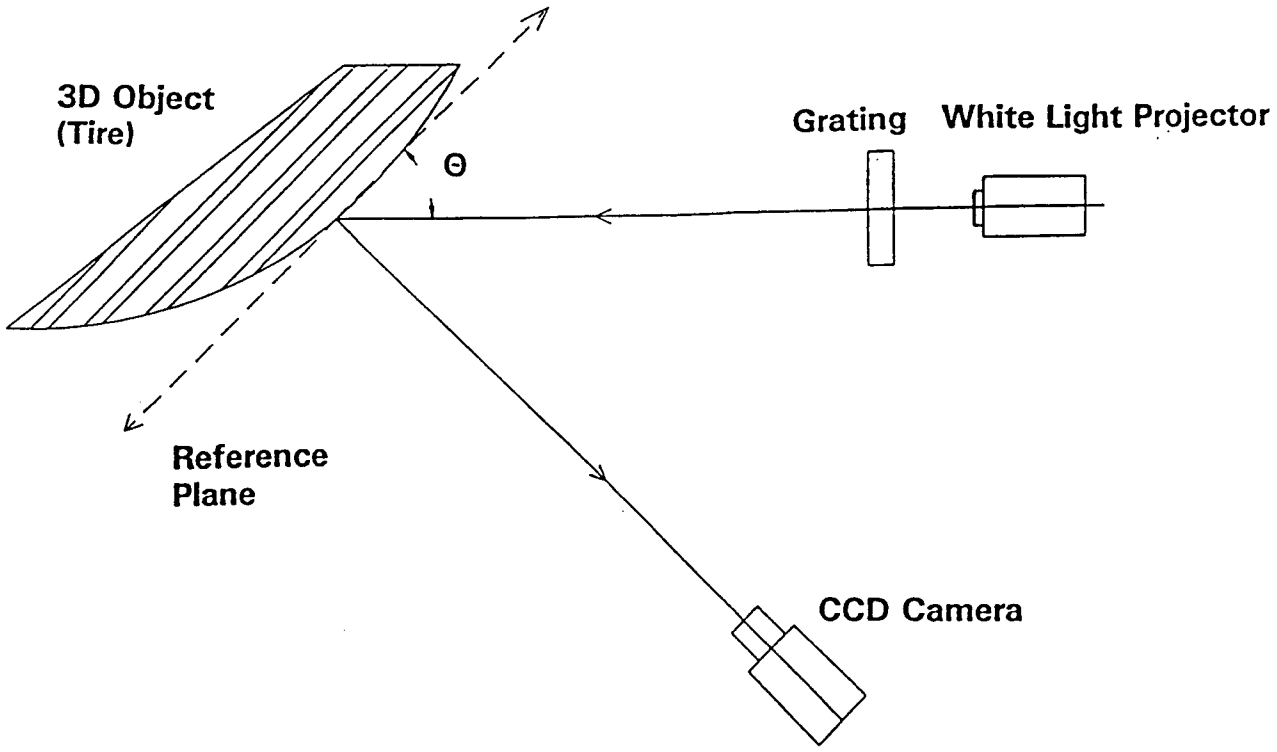
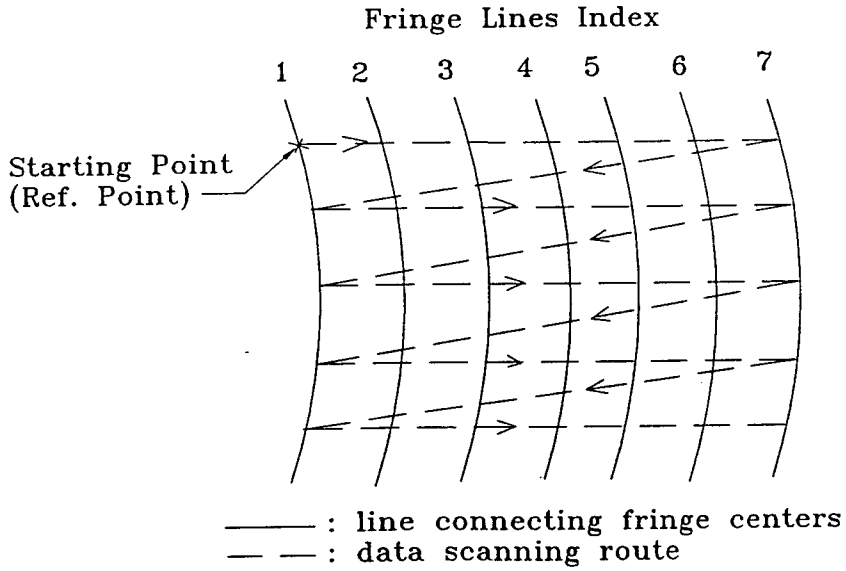


Fig.1 Experimental Arrangement for 3D Geometry Measurement



Note: Fringe 1 is scanned first (from top to bottom) and the remaining fringes are then scanned horizontally line by line as shown in this figure.

Fig. 2 Image Data Scanning Process

Fig. 3

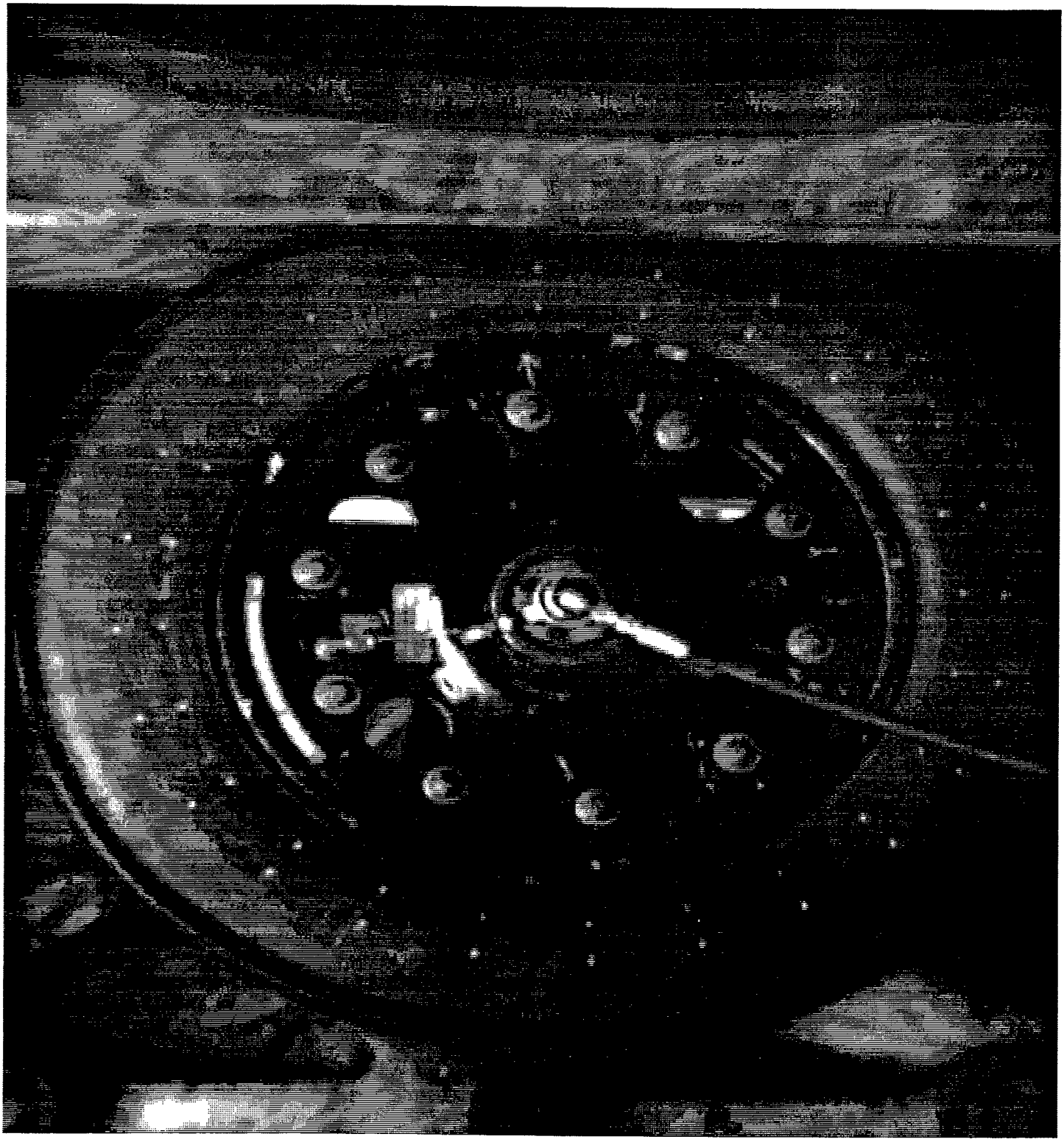
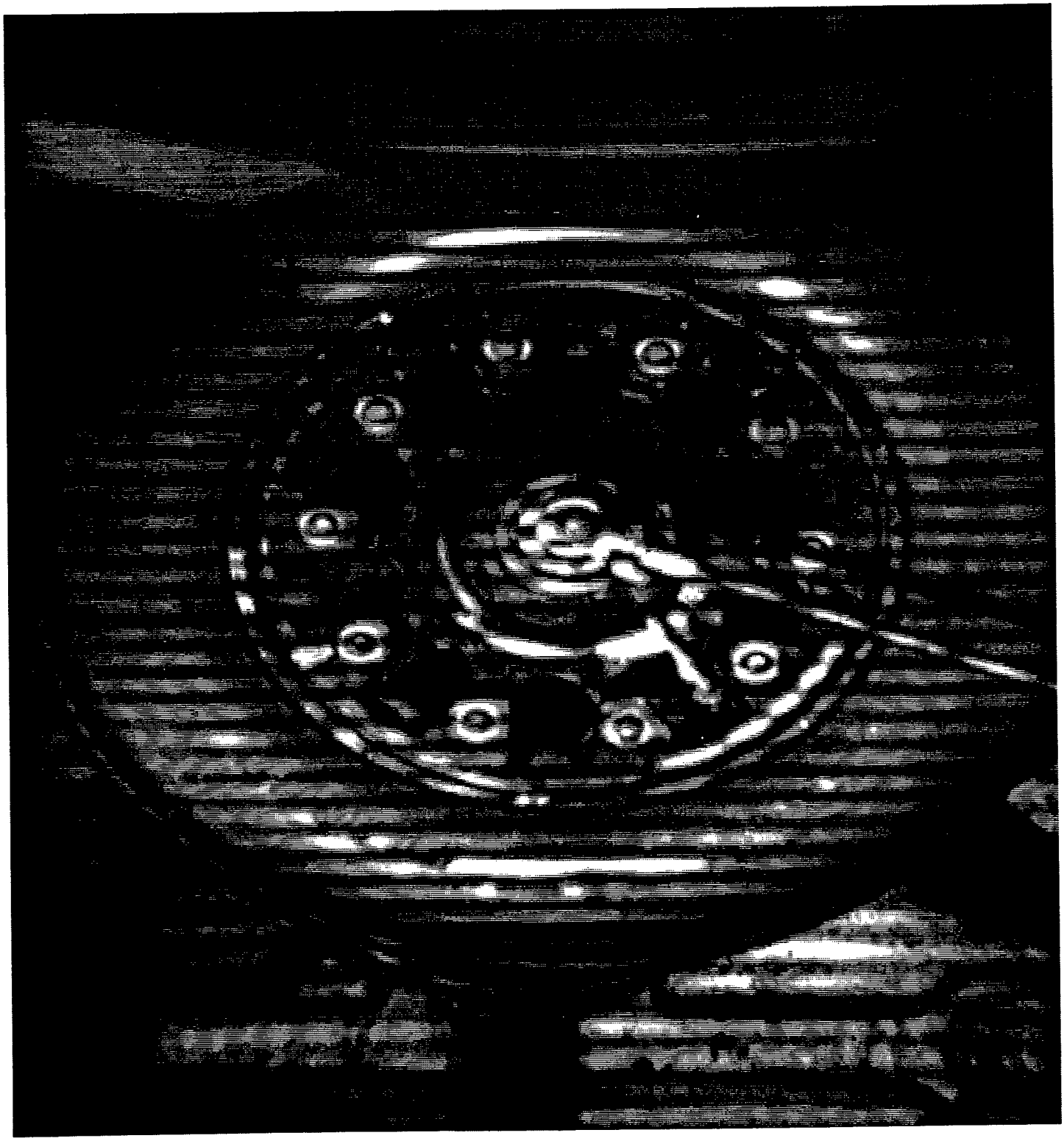


Fig. 4



Results and Discussion

In this research, only F-16 bias and radial tires were used. Each type of tire was loaded against flat plate and flywheel, at 30% and 40% tire deflections, and at 0°, ±2°, ±4° yaw angles. Both 2-D and 3-D deformations were examined and analyzed.

TABLE I

Loading Comparison between F16 Bias and F16 Radial Tires

Loading Condition: Rated Pressure (310 psi initially)
Corrected Load (as shown below)

(A) Subjected to Flat Plate Loading

Deflection	Bias	Radial
30 %	15800 lb	13300 lb
40 %	22600 lb	20300 lb

(B) Subjected to Flywheel Loading

Deflection	Bias	Radial
30 %	14000 lb	12100 lb
40 %	20600 lb	18500 lb

It can be seen from Table I that the radial tire requires about 16% less load than the bias tire for producing 30% deflection, and likewise the former requires about 10% less load than the latter for producing 40 % deflection. In other words, when subjected to the same load, it is easier to deflect a radial tire. The 2-D displacement data in polar coordinates also indicates the same finding (see Appendix). The 3-D graphs included in the Appendix show only the 3-D reconstructed geometry of a small but significant portion of a deformed tire (a rectangular region near the contact between tire and flat plate or flywheel). Furthermore, due to deformation symmetry only the upper half of the portion is shown. Thus, the bottom horizontal line of each 3-D reconstructed surface corresponds to the line connecting the centers of tire and flywheel. It should be noted that this appendix includes only the test results with zero yaw angle.

The complete analysis of yaw angle effect on tire 3-D deformation is not available yet. However, the reference point's height changes subjected to

different yaw angles and percentages of tire deflection were measured directly with the fiber-optic sensor. The reference point is located on the tire sidewall, near the wheel flange and approximately 35° counter clockwise from the line connecting the centers of tire/wheel and flywheel. The height change is perpendicular to the tire sidewall, in the direction parallel to the wheel axle. TABLE II shows the measured height change data comparison.

TABLE II

Comparison of the Reference Point's Height Changes

Unit: mm

(A) F16 Bias Tire

With Flat Plate

Deflection	Yaw angle	0°	2°	4°	-2°	-4°
30 %		1.55	0.85	1.06	1.70	2.15
40 %		1.63	1.17	1.36	1.81	2.22

With Flywheel

Deflection	Yaw angle	0°	2°	4°	-2°	-4°
30 %		1.22	0.68	0.75	2.15	2.37
40 %		1.38	0.86	0.95	2.75	3.12

(B) F16 Radial Tire

With Flat Plate

Deflection	Yaw angle	0°	2°	4°	-2°	-4°
30 %		0.87	0.40	0.46	0.47	0.73
40 %		1.00	0.65	0.88	0.61	0.96

With Flywheel

Deflection	Yaw angle	0°	2°	4°	-2°	-4°
30 %		0.80	0.35	0.42	0.54	0.76
40 %		0.90	0.47	0.56	0.68	0.99

Table II indicates that for the radial tire there is not much difference in height change between the flat plate and the flywheel readings, whereas for

the bias tire the difference is significant. In general, the height changes are more with negative yaw angle than with positive ones. The difference becomes more obvious for the bias tire.

Conclusions and Future Work

The 2-D and 3-D data analyses thus far indicate that when subjected to the same loading a radial tire has higher 2-D displacement magnitudes (i.e. in the plane of sidewall) than a bias tire. Because of this phenomenon, a radial tire seems to exhibit less out-of-plane deformation (i.e. height change in z-direction, perpendicular to the sidewall). This preliminary finding in terms of out-of-plane deformation comparison needs to be confirmed when the complete image analysis becomes available in the near future.

The measuring system and the optical technique used worked very well, except that the high speed flash was not available. Future work will include testing of the integrated measuring system, and more detailed image analyses on the tire deformation comparison between F16 bias and radial tires. A new CCD camera with resolution better than 512 by 480 with the imager size larger than 1/2 inch is desired. A larger imager allows for a shorter distance between the camera and the viewing object, which might be an important factor particularly when the space in the test facility is limited. Future tests might include the deformation measurement of a tire at its critical speed and/or in braking condition.

References

Lin, P. P. and Parvin, F., "Edge Detection with Subpixel Resolution and its Application to Radius Measurement via Fringe Projection Technique," SME Technical Paper, MS90-576, 1990, pp. 4-13 - 4-27.

Lin, P. P., Parvin, F., and Schoenig, Jr., F. C., "Optical Gaging of Very Short-term Surface Waviness," Transactions of NAMAR/SME, 1991, pp. 327-322.

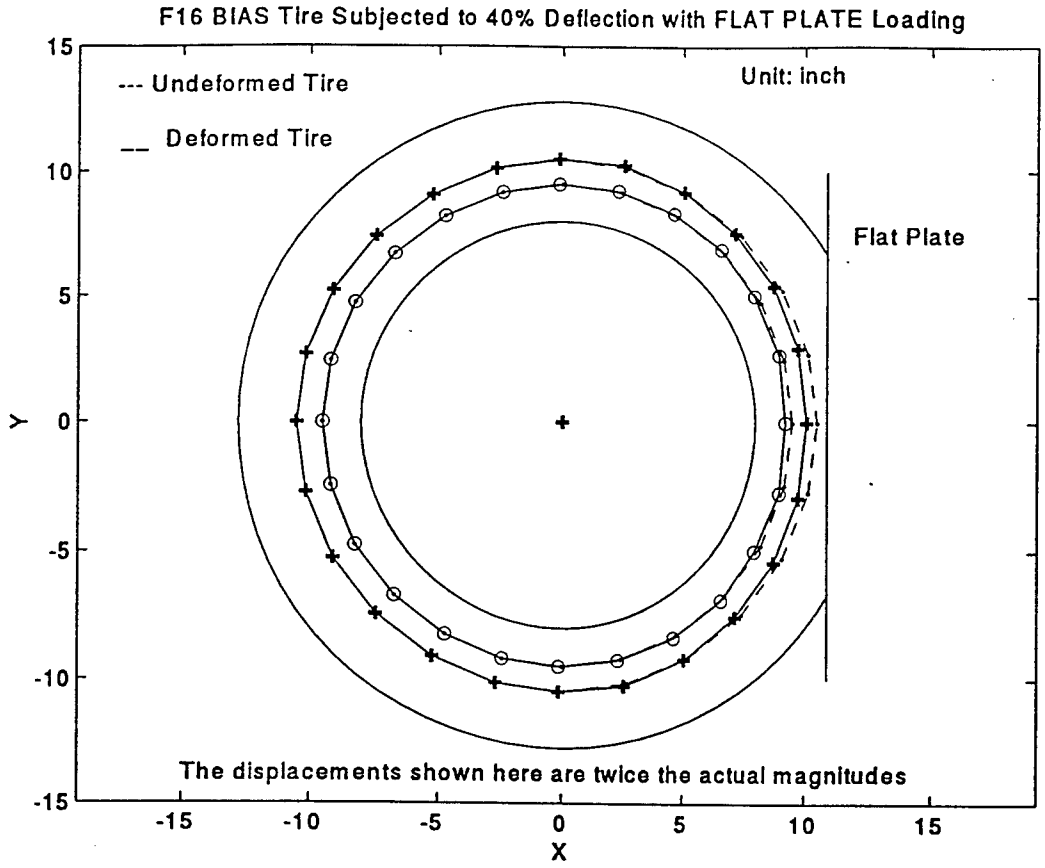
Lin, P. P., Chawla, M. D., and Ulrich, P. C., "Optical Technique for Measuring Tire Deformation and Strains - Preliminary Results," SAE Technical Paper No. 941178, Aerospace Atlantic Conference and Exposition, Dayton, Ohio, April, 1994.

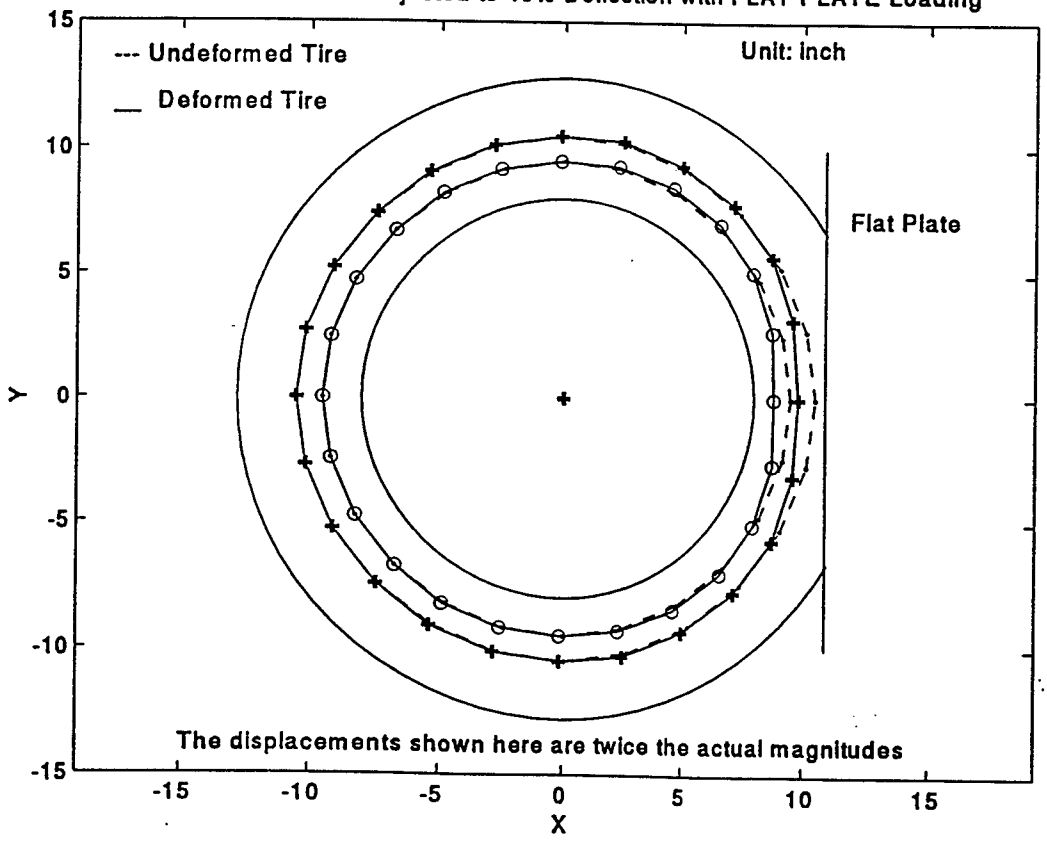
Marr, D. and Poggio, T., "Cooperative Computation of Stereo Disparity, "Science, V. 194, 1976, pp. 283-287.

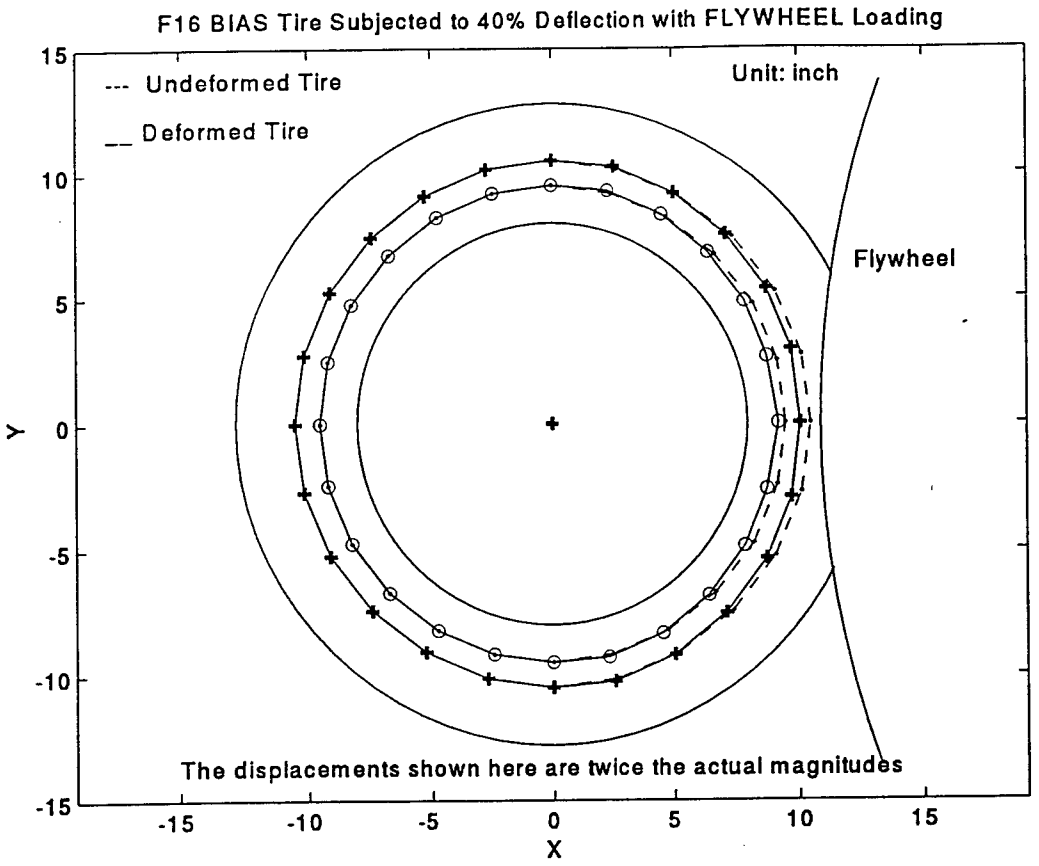
Vemuri, B. C. and Aggarwal, J. K., "3-Dimensional Reconstruction of Objects from Range Data," Proc. of 7th Int. Conf. on Pattern Recognition, V1, 1984, pp. 752-755.

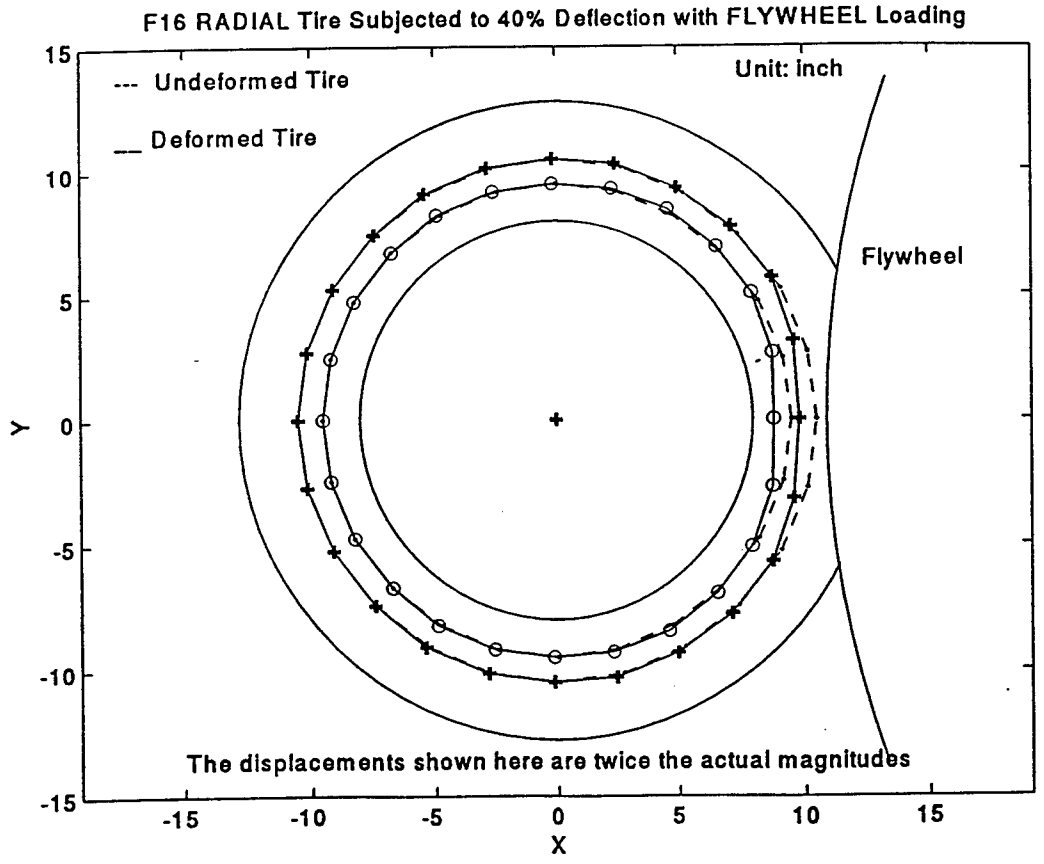
APPENDIX

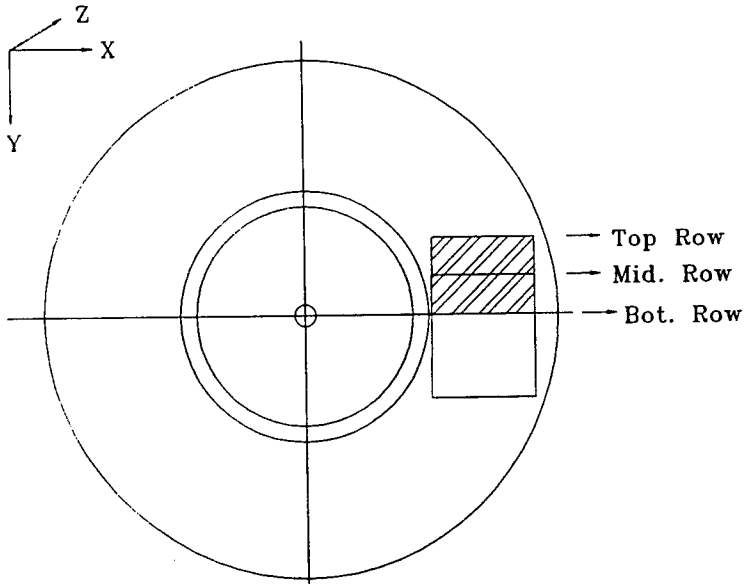
Selected Results of 2-D and 3-D Tire Deformations


F16 BIAS Tire Subjected to 40% Deflection with FLAT PLATE Loading
--- Undeformed Tire
__ Deformed Tire
Unit: inch
Flat Plate
The displacements shown here are twice the actual magnitudes
Y
X
15
10
5
0
-5
-10
-15

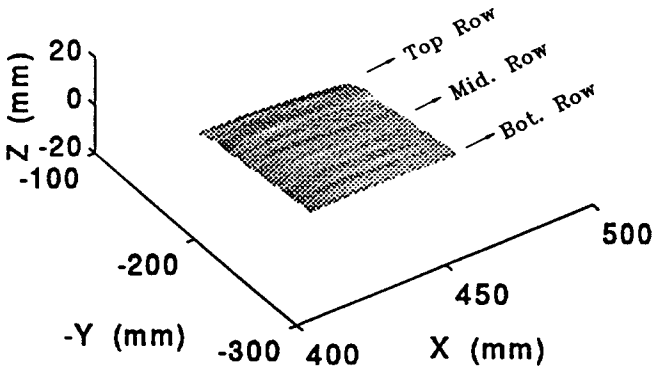

F16 RADIAL Tire Subjected to 40% Deflection with FLAT PLATE Loading
--- Undeformed Tire
__ Deformed Tire
Unit: inch
Flat Plate
The displacements shown here are twice the actual magnitudes
Y
X
15
10
5
0
-5
-10
-15


F16 BIAS Tire Subjected to 40% Deflection with FLYWHEEL Loading
--- Undeformed Tire
__ Deformed Tire
Unit: inch
Flywheel
The displacements shown here are twice the actual magnitudes
Y
X
15
10
5
0
-5
-10
-15

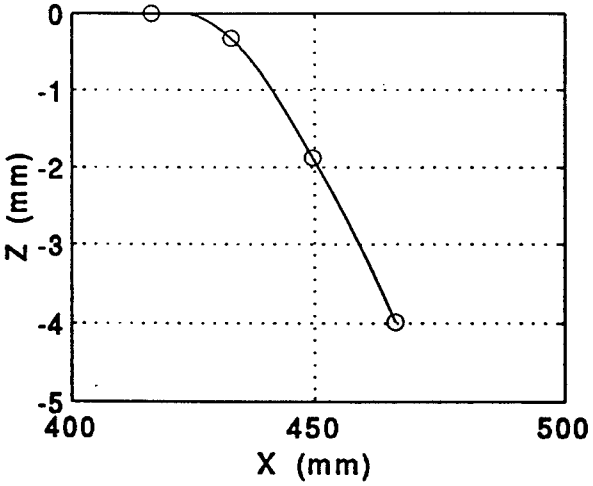

F16 RADIAL Tire Subjected to 40% Deflection with FLYWHEEL Loading
--- Undeformed Tire
__ Deformed Tire
Unit: inch
Flywheel
The displacements shown here are twice the actual magnitudes
Y
X
15
10
5
0
-5
-10
-15



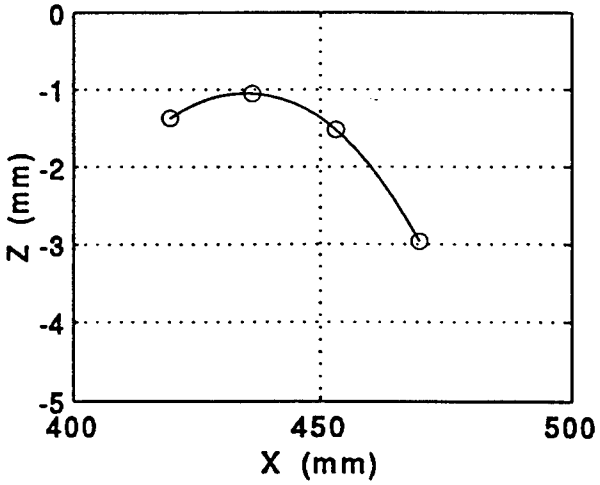
F16 Bias Tire:0% Deflection,Static (fwr00g)



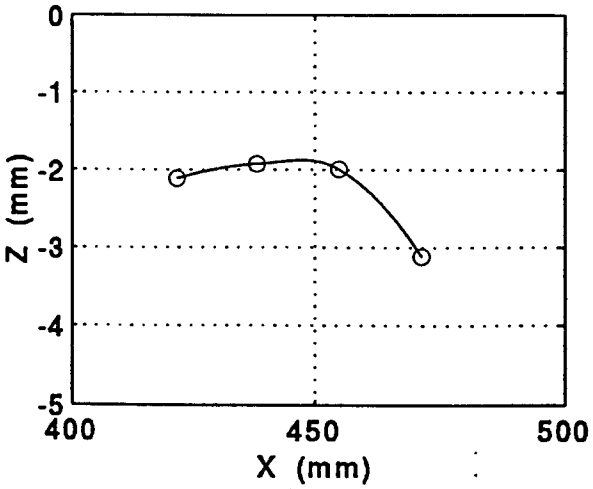
Deformation at Top Row



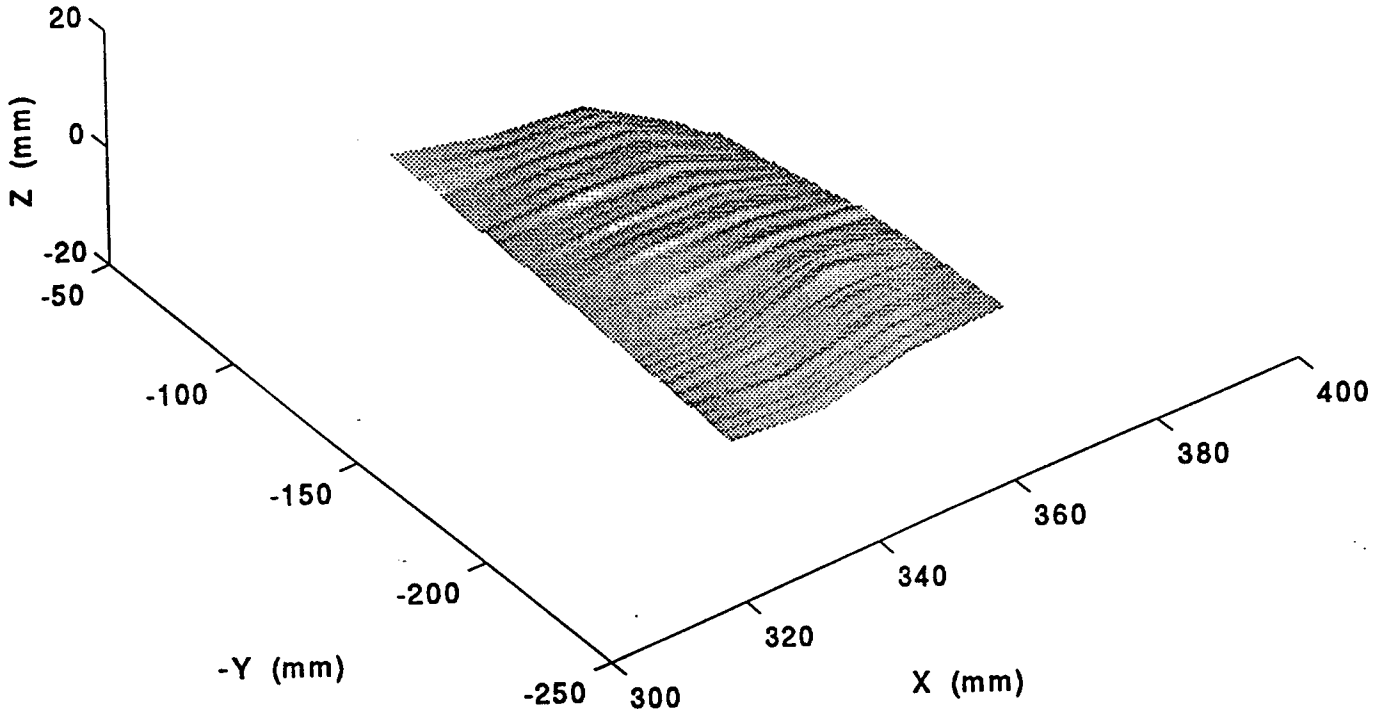
Deformation at Middle Row



Deformation at Bottom Row

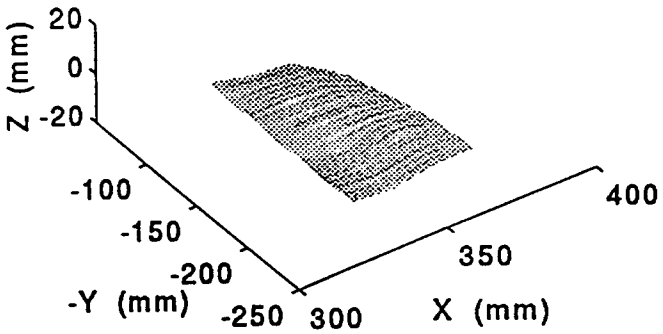


F-16 Bias Tire:Flatplate,R. Pres. C. Load, 30% Defl.,Static (fpra300)

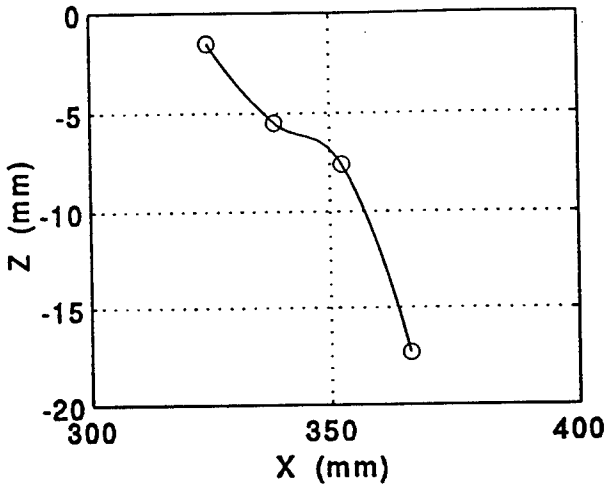


F16 Bias Tire:Flatplate, 30% Defl.,Static(fpra300)

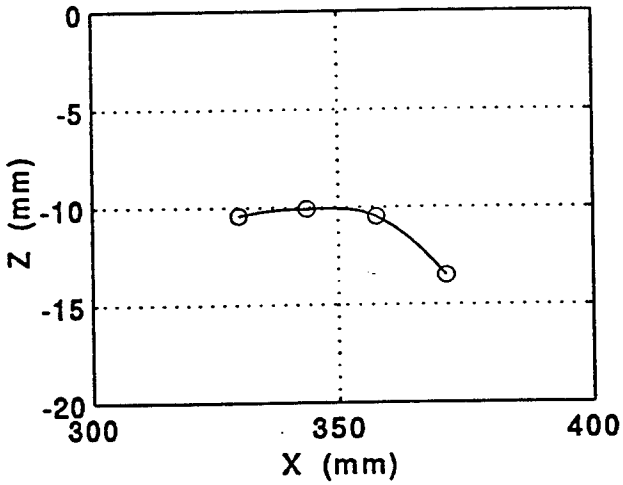
Initial Pressure:310 psi ; Load:15800 lb



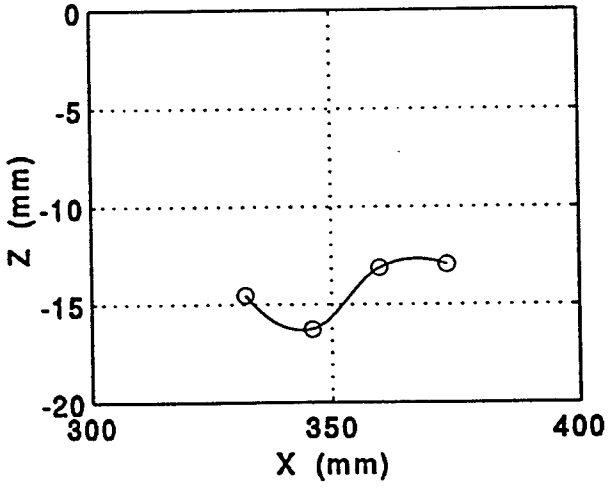
Deformation at Top Row



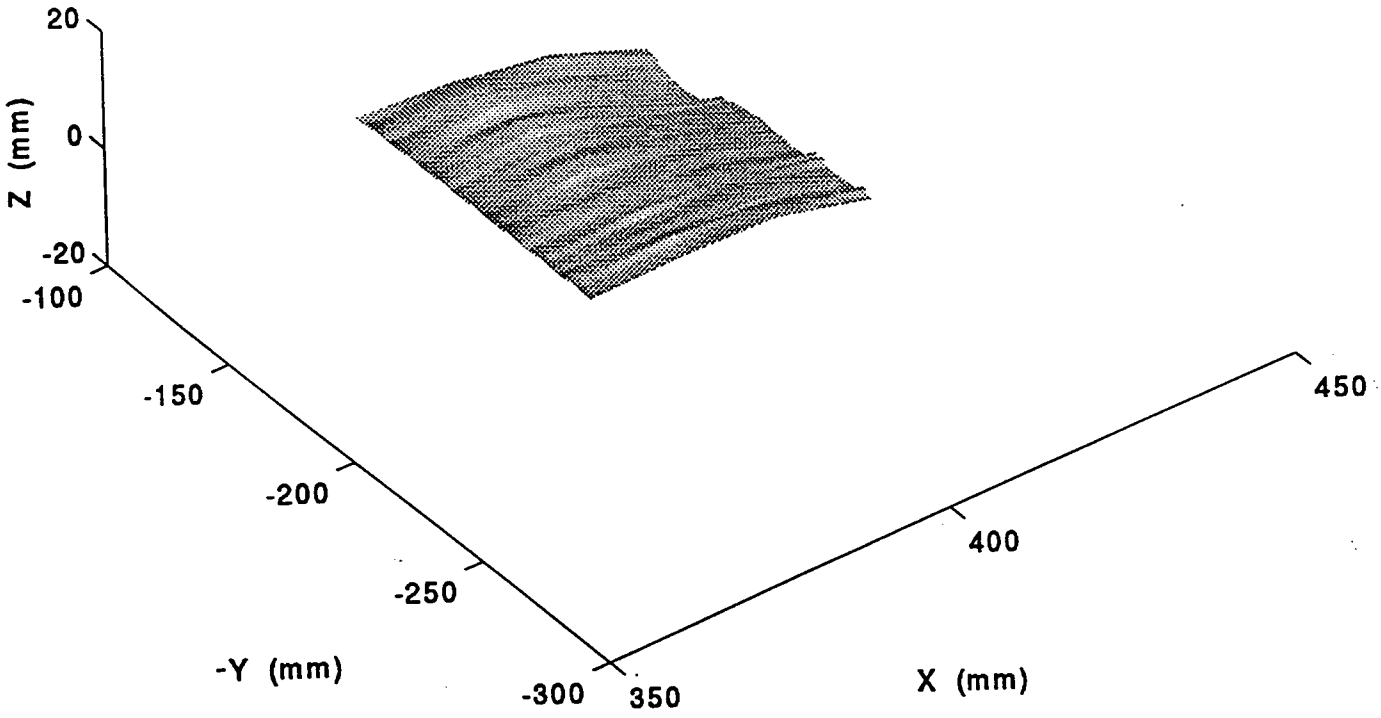
Deformation at Middle Row



Deformation at Bottom Row

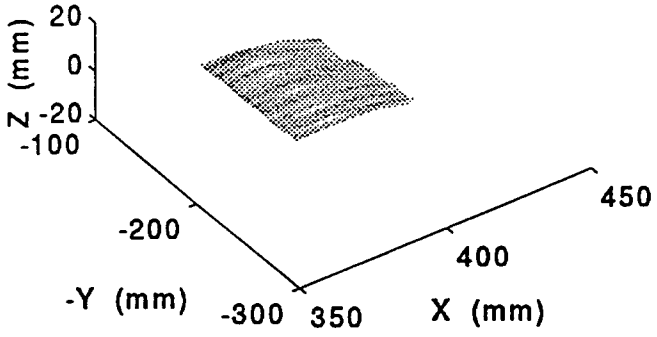


F-16 Bias Tire:Flywheel,R. Pres. C. Load, 30% Defl.,Static (fwrc300g)

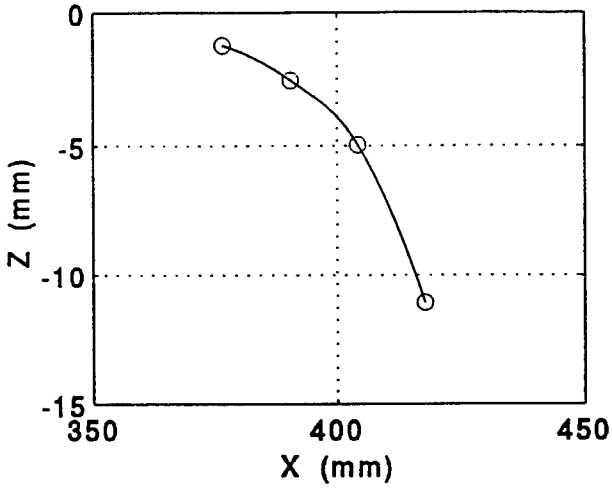


F16 Bias Tire:Flywheel, 30% Defl.,Static(fwrc300g)

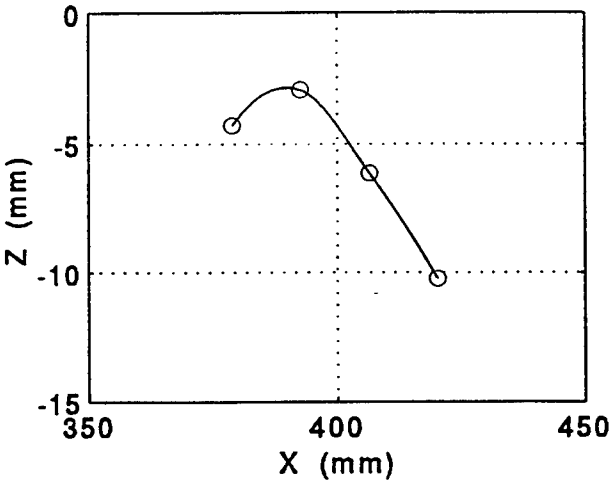
Initial Pressure:310 psi ; Load:14000 lb



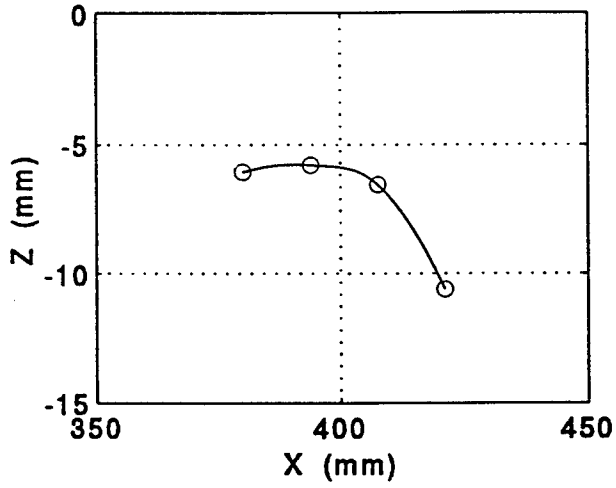
Deformation at Top Row



Deformation at Middle Row



Deformation at Bottom Row



A MODEL TO MONITOR THE CURRENT GAIN LONG-TERM INSTABILITY IN AlGaAs/GaAs HBTs BASED ON NOISE AND LEAKAGE CURRENT CHARACTERISTICS

Juin J. Liou
Associate Professor
Department of Electrical and Computer Engineering

University of Central Florida
Orlando, Florida, 32816

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

July 1994

A MODEL TO MONITOR THE CURRENT GAIN LONG-TERM INSTABILITY IN AlGaAs/GaAs HBTs BASED ON NOISE AND LEAKAGE CURRENT CHARACTERISTICS

Juin J. Liou
Associate Professor
Department of Electrical and Computer Engineering
University of Central Florida

Abstract

A simple model is developed to monitor the dc current gain long-term instability in the AlGaAs/GaAs heterojunction bipolar transistor (HBT). It is derived from the theory that the recombination current at the extrinsic base surface increases with time due to the surface degradation process. Furthermore, the initial 1/f noise and base leakage current characteristics have been used to provide the needed model parameters for the HBT surface recombination mechanism and surface quality, respectively. The current gain long-term variations calculated from the model for four HBTs compare favorably with those obtained from measurements. The model developed can be used to screen unreliable HBTs without having to carry out the long-hour stress test.

A MODEL TO MONITOR THE CURRENT GAIN LONG-TERM INSTABILITY IN AlGaAs/GaAs HBTs BASED ON NOISE AND LEAKAGE CURRENT CHARACTERISTICS

Juin J. Liou

1. INTRODUCTION

Current gain long-term instability in AlGaAs/GaAs heterojunction bipolar transistors (HBTs) is an important problem, since it sets a limit to the useable lifetime of such devices. Many HBT post-burn-in measurements have shown that while the collector current is affected only slightly by the stress test, the base current can increase considerably due to the degradation of surface property [1-2]. Thus the HBT dc current gain long-term instability is caused mainly by the increase in the recombination current at the extrinsic base surface.

Low-frequency noise has been widely used as a means to evaluate the extrinsic base surface property [3-4]. For example, most researchers suggested that the 1/f noise is originated from fluctuations in the occupancy of surface traps, which in turn perturbs the extrinsic base surface recombination current I_{bs} [5-6]. As a result, the magnitude of 1/f noise serves as an indicator for the recombination mechanism at the surface. The quality of the base surface, or the interface between dielectric layer and GaAs, on the other hand, is directly related to the base leakage current generated at the emitter-base periphery [7-8]. Normally, the larger the leakage current, the poorer the surface quality, and the more likely the surface will degrade as time is increased. Such a surface degradation then gives rise to the current gain long-term instability [9].

Previous work has been limited to the determination of the noise sources in the HBT and their origins, and a framework explicitly correlating the initial low-frequency noise and leakage current behavior with the HBT long-term performance has largely been overlooked in the literature.

This paper seeks to correlate the 1/f noise and base leakage current with the current gain long-term instability in the AlGaAs/GaAs HBT under room temperature. An analytical model is also developed which, when given the initial noise and leakage current characteristics, can be used to monitor the current gain long-term behavior of HBTs fabricated under the same process condition. Thus, the model has useful applications on screening unreliable HBTs without having to conduct the long-hour stress test.

The following sections describe the modeling of time-dependent base surface current $I_{bs}(t)$, which gives rise to the HBT dc current gain drift, the correlation between $I_{bs}(t)$ and the initial 1/f noise, the correlation between $I_{bs}(t)$ and the initial base leakage current, and finally some results calculated from the model and obtained from measurements for four HBTs having different emitter patterns and fabricated with different processes.

2. MODEL DEVELOPMENT

A ratio $\Delta h_{FE}/h_{FE0}$ (in %) is often used to characterize the dc current gain long-term instability [9], where h_{FE0} is the initial current gain and Δh_{FE} is the accumulated current gain drift over a particular time period t (in hour):

$$\Delta h_{FE} = \int_0^t dh_{FE}(t) \tag{1}$$

Note that $h_{FE} = I_C/I_B$, where I_C and I_B are the collector and base currents, respectively. The base current of the HBT consists of the intrinsic component I_{bi} and surface component I_{bs}:

$$I_B = I_{bi} + I_{bs} \tag{2}$$

Since Δh_{FE} is caused mainly by the surface degradation, I_{bi} can be assumed constant with respect to time, whereas I_{bs} is a function of time. Thus, for a constant I_C flowing through the device,

$$\Delta h_{FE}/h_{FE0} = \Delta h_{FE}/(I_C/I_{B0}) = \Delta h_{FE}/[I_C/(I_{bi} + I_{bs0})] =$$

$$-(I_{bi} + I_{bs0}) \int_0^t [I_{bi} + I_{bs}(t)]^{-2} dI_{bs}(t) \tag{3}$$

where $I_{B0} = I_B(t = 0)$ and $I_{bs0} = I_{bs}(t = 0)$.

The remaining task in developing the $\Delta h_{FE}/h_{FE0}$ model in (3) is to derive the expressions for I_{bi} and $I_{bs}(t)$.

2.1 Base Intrinsic Current I_{bi}

I_{bi} can be expressed as [8]

$$I_{bi} = qA_E D_p p(0)/W_E + I_C(1- \alpha_B) \tag{4}$$

where A_E is the emitter area, D_p is the hole diffusion coefficient, p(0) is the hole concentration at the emitter edge of the space-charge region, W_E is the emitter thickness, and α_B is the base transport factor. The first term on the right-hand side of (4) is the hole current injected from the base to emitter, and the second term is the recombination current in the base. For a typical HBT with a graded layer which removes the spike at the heterointerface, p(0) can be related to I_C, and (4) becomes

$$I_{bi} = (N_B/N_E)(D_p/D_n)(W_B/W_E)I_C \exp(-\Delta E_V/kT) \tag{5}$$

where N_B and N_E are the base and emitter doping concentrations, D_n is the electron diffusion coefficient, W_B is the base thickness, and ΔE_V is the effective valence band discontinuity including the graded layer effect [8].

2.2 Base Surface Current I_{bs}

The base surface recombination current I_{bs}, on the other hand, depends on the surface property and thus is related to the noise and leakage current. Summarizing the various experimental results, we find that the time-dependent I_{bs} follows the relation

$$I_{bs}(t) = I_{bs0} + I_{bs0}t^{\lambda} = I_{bs0}(1 + t^{\lambda}) \tag{6}$$

where λ is a parameter associated with the base surface quality. The first term on the right hand side of (6) describes the initial condition for I_{bs} and the second term accounts for the time dependence of I_{bs}. As will be shown below, I_{bs0} and λ can be determined from the measured initial 1/f noise and base leakage current, respectively.

2.2.1 Correlation Between I_{bs0} and 1/f Noise

The HBT noise measurements [6] show the existence of three distinct regions in the noise spectra: a 1/f shape (Flicker or 1/f noise) at lower frequencies, a Lorenztian spectrum (bump or burst noise) at intermediate frequencies, and a constant noise (white or shot noise) at higher frequencies. Based on the theories summarized in [6] for the three different noise regions, we have constructed a noise model for the HBT for a wide frequency range. Fig. 1 shows the model predictions, which agree well with the equivalent input base noise current spectral density (A^2/Hz) observed in measurements [6]. Our emphasis here will be the 1/f noise at low frequencies because it is related to the base surface recombination mechanism and thus the h_{FE} drift under study.

Assuming the initial surface recombination velocity S ($S \approx 5x10^5$ cm/sec without an AlGaAs ledge structure on the base surface and $S \approx 10^4$ cm/sec with a ledge) is independent of position and excess free-carrier density at the surface, Fonger [10] showed that the initial base current 1/f noise spectral density S_{Ib} can be expressed as

$$S_{Ib}(f) = (I^2_{bs0}/S^2)S_s(f) \tag{7}$$

where $S_s(f)$ is the spectral density characterizes the noise contribution at the surface. If we further assume that

$$I_{bs0} \approx qSn(0)L_dP_E \tag{8}$$

and that n(0) can be expressed in turn of I_C because

$$I_C \approx qA_ED_nn(0)/W_B \tag{9}$$

Here L_D is the electron lateral diffusion length in the base, n(0) is the injected minority electron carrier at the base edge of the space-charge region, and P_E is the emitter perimeter length. Putting (8)-(9) into (7) yields

$$S_{Ib}(f) \approx I_C^2(W_b L_d/D_n)^2(P_E/A_E)^2 S_s(f) \quad (10)$$

Since $S_s(f)$ is inversely proportional to the frequency f and is proportional to S, it can be given by [3]

$$S_s(f) = CS/f \quad (11)$$

where C is a constant. Thus, if $S_{Ib}(f)$ is measured at a particular I_C and for an HBT with a known device make-up and geometry, then C can be determined from (10). Consequently, from (7),

$$I_{bs0} = [S_{Ib}(f)fS/C]^{0.5} \quad (12)$$

It should be pointed out that $S_{Ib}(f)f$ is a constant and thus an arbitrary f can be selected as long as the 1/f noise dominates.

In this paper, four AlGaAs/GaAs HBTs (HBT-1, HBT-2, HBT-3, and HBT-4) are studied. HBT-1 and HBT-2 are fabricated in our laboratory, and both devices have 10 emitter fingers with a 3-μm diameter for each finger, have the same device make-up and dielectric layer (nitride layer), but have different finger patterns (HBT-1 has two 5-finger rows whereas HBT-2 has one 10-finger row) and are fabricated from different wafers. Their initial noise characteristics have very similar trends as those shown in Fig. 1, only the magnitude and Lorenztian spectrum are somewhat different [11]. At f = 100 Hz and collector current density $J_C = 5\times10^4$ A/cm^2, the base current noise spectral density for the two HBTs is about 8×10^{-19} A^2/Hz.

The second group of HBTs (HBT-3 and HBT-4) is supplied from an industry laboratory and is fabricated with a process different from that used in our laboratory. It has 3 rectangular-shape emitter fingers with a 2x20 μm^2 for each finger. The noise levels are 1.2×10^{-19} and 8×10^{-20} A^2/Hz for HBT-3 and HBT-4, respectively, at f = 100 Hz and $J_C = 1.5\times10^4$ A/cm^2. These noise spectral densities are higher than those in HBT-1 and HBT-2, as they correspond to 1.3×10^{-18} and 9×10^{-19} A^2/Hz at $J_C = 5\times10^4$ A/cm^2 ($S_{Ib}(f)$ is proportional to I_C^2).

2.2.2 Correlation Between λ and Leakage Current

Base and collector leakage currents are the dominant current components in AlGaAs/GaAs HBTs operated at relatively small bias voltages. Such currents have been widely observed in experimental measurements. According to a recent study [7-8], the base and collector leakage currents can take place at the emitter-base perimeter (circle 1 in Fig. 2), base-collector perimeter (circle 2), collector-subcollector perimeter (circle 3), and subcollector-substrate interface (circle 4). The magnitude of such currents depends strongly on the etching process, the quality of the emitter-base and base-collector peripheries which is covered by the dielectric (e.g., polyimide, nitride, etc.) layer, and the quality of n$^+$-GaAs/semi-insulating-GaAs interface. For discussion, we assume the HBT surface is protected with a polyimide layer. An inferior quality emitter and base peripheries can thus increase the possibility

for the free carriers to diffuse to the opposite side of the junction through the GaAs-polyimide and AlGaAs-polyimide interface and subsequently increase the leakage currents.

Since I_{bs} is of interest, we place the emphasis on the base leakage current. The total base current I_B consists of the normal current (I_{BN}), leakage currents at the emitter-base periphery ($I_{BL,E}$), base-collector periphery ($I_{BL,B}$), and n^+-GaAs/SI-GaAs interface ($I_{BL,SI}$) [7-8]:

$$I_B = I_{BN} + I_{BL,E} - I_{BL,B} - I_{BL,SI} \tag{13}$$

The base leakage current originated at the emitter perimeter ($I_{BL,E}$) is affected by both the base surface quality and the applied base-emitter voltage V_{BE}. As V_{BE} is increased, the potential barrier at the GaAs-polyimide interface is lowered, thus increasing the number of holes in the p-type GaAs to surmount the barrier and reach the emitter region. An analytical model for the initial base and collector leakage currents in AlGaAs/GaAs HBTs has been proposed in [7-8], which was derived analogous to the physical degradation mechanism described in the Arrhenius relationship [12] as

$$I_{BL,E} = J_{BL,E}P_E[1 - \exp(-V_{BE}F_L/V_T)] \tag{14}$$

where $J_{BL,E}$ (A/cm) is the initial leakage hole current density from the base to emitter, F_L is an empirical parameter, and $V_T = kT/q$ is the thermal voltage. The values of $J_{BL,E}$ and F_L can be obtained by fitting the model with measurement data, and the magnitude of $J_{BL,E}$ indicates the initial quality of the base surface. Equation (13) suggests that the base leakage current taking place at the emitter periphery increases very rapidly and becomes near constant as V_{BE} is increased. It is important to point out that the magnitude of $J_{BL,E}$ is of direct relevance to the base surface quality and thus, for the purpose of determining λ, is the key parameter to be extract from the total base current.

The leakage current $I_{BL,B}$ occurs at the base perimeter can be expressed in the same form as (14):

$$I_{BL,B} = J_{BL,B}P_B[1 - \exp(-V_{BC}F_L/V_T)] \tag{15}$$

where $J_{BL,B}$ is the leakage hole current density from the base to collector, P_B is the length of the base perimeter, and V_{BC} is the applied base-collector voltage. Since $I_{BL,B}$ flows in the opposite direction as the normal base current [7-8], a negative sign is associated with such a current in the total base current model (Eq. (13)). Note that this leakage current is absent if $V_{BC} = 0$ is used in measurements.

Another leakage current $I_{BL,SI}$ results from the leakage of free carriers through the n^+-GaAs/SI-GaAs (SI denotes semi-insulating) interface. Such an interface is "leaky" due to the very high defect density at the SI GaAs surface, and kinetic factors such as the chemical barrier and diffusion rate often prevent the system from reaching the equilibrium state. Since the SI substrate has a lower electron density and higher hole density than the n^+ sub-

collector, electrons and holes can leak, or diffuse, through the *nonequilibrium* interface and enter the SI substrate and the subcollector, respectively. The effects of the n^+-GaAs/SI-GaAs leakage on the base current can be treated by focusing on the hole transport. Once entering the n^+ region, the holes, which are minority carriers, will diffuse across the n^+ and n regions. Depending on the recombination process in these regions, a percentage of the holes will reach the base-collector depletion region and be swept into the base region by the large electric field at the base-collector junction. Since the number of holes allowed to be injected into the emitter is fixed by the emitter-base voltage, these extra holes are forced to flow out of the base terminal, which constitutes a current flow *opposite* to the normal base current flow. Thus, a negative base current component ($-I_{BL,SI}$) also needs to be included in the base current model to account for the n^+-GaAs/SI-GaAs interface leakage. Obviously, to maintain quasi neutrality in the SI substrate, holes need to be supplied from a current path through the grounded substrate.

Unlike the leakage process at the emitter-base and base-collector peripheries which require the free carriers to surmount the potential barrier associated with the dielectric layer, the leakage at the subcollector/substrate interface is caused by the diffusion of electrons and holes through the leaky interface. As a result, $I_{BL,SI}$ and $I_{CL,SI}$ are independent of the applied voltage.

Fig. 3 shows the total base currents, including both the leakage and normal components, calculated from the model and obtained from measurements for HBT-1 and HBT-2. In measurements and calculations, $V_{BC} = 0$ is used to eliminate the leakage component $I_{BL,B}$ at the base-collector perimeter, since it is not relevant to the base surface under study. Note that the leakage current dominates the base current for $V_{BE} < 0.8$ V. At higher voltages, the base current consists mainly of the normal components such as the hole injection current from the base to emitter and recombination current in the base. In the model calculations, $J_{BL,E} = 5.3\times10^{-6}$ A/cm and 3.2×10^{-6} A/cm and $I_{BL,SI} = 0$ have been used for HBT-1 and HBT-2, respectively, suggesting that HBT-1 has a poorer base surface initially than HBT-2 and therefore should have a larger surface degradation and a larger current gain drift as time is increased.

The behavior of HBT-3 and HBT-4 base leakage currents differs considerable from that for HBT-1 and HBT-2, as evidenced by the results shown in Fig. 4. For small voltages ($V_{BE} < 0.2$ V for HBT-3 and $V_{BE} < 0.4$ V for HBT-4), the base currents of these devices are negative. This is caused by the substantial leakage current ($I_{BL,SI} = 5\times10^{-8}$ A for HBT-3 and $I_{BL,SI} = 8\times10^{-8}$ A for HBT-4) through the subcollector/substrate interface in these devices. The leakage current densities at the emitter perimeter are also higher than HBT-1 and HBT-2 ($J_{BL,E} = 5\times10^{-5}$ A/cm^2 for HBT-3 and $J_{BL,E} = 4.5\times10^{-5}$ A/cm^2 for HBT-4), indicating that a large long-term current gain drift is likely in HBT-3 and HBT-4.

We found that λ can be correlated with $J_{BL,E}$ by the empirical relation

$$\lambda = B\cdot J_{BL,E}^{0.5} \qquad (16)$$

where B is an empirical parameter. For HBTs fabricated with the same process, B should be constant and can be determined by fitting the above model to the measured long-term current gain of one device. Once the value for B is established, λ for any HBT fabricated from the same process can be found from its initial base leakage current using (16). Such a λ, together with I_{bs0}, is then used to estimate the long term performance of all other HBTs under the same process condition.

According to the data in Fig. 3 and the relationship in (16), B = 1.5×10^{2} $(cm/A)^{0.5}$ for both HBT-1 and HBT-2, and λ = 0.35 and 0.27 for HBT-1 and HBT-2, respectively. For HBT-3 and HBT-4 fabricated from a different process, B needs to be re-determined and is found to be 0.75×10^{2} $(cm/A)^{0.5}$, which leads to λ = 0.53 for HBT-3 and λ = 0.45 for HBT-4.

3. RESULTS AND DISCUSSIONS

Fig. 5 shows the current gain variation ($\Delta h_{FE}/h_{FE}$) versus time at room temperature calculated from the model and obtained from measurements for HBT-1 and HBT-2. Two different collector current densities (2.5×10^{4} and 5×10^{4} A/cm^2) are considered. Clearly, the device (HBT-1) having a higher initial leakage current density at the base surface has a larger current gain drift. Also, it is shown that the gain variation increases with the constant collector current flowing through the device.

The current gain variations of HBT-3 and HBT-4 at two different stress current levels are shown in Fig. 6. Comparing to HBT-1 and HBT-2, these devices have a larger current gain degradation, which is expected because HBT-3 and HBT-4 possess higher base leakage current densities and therefore poorer base surface qualities.

All four HBTs considered do not have a passivated base surface. If an AlGaAs ledge structure were incorporated on the base surface of these devices, the initial recombination velocity at the base surface will decrease, which subsequently reduces the initial base current noise spectra [13] and thus the current gain long-term instability. This is evidenced by the results shown in Fig. 7 which are calculated from the model assuming HBT-1, HBT-2, HBT-3, and HBT-4 had an AlGaAs ledge on the extrinsic base surface ($S = 10^{4}$ cm/sec has been used). Comparing to the results in Figs. 5 and 6, we found that the ledge has reduced the current gain degradation of HBT-1 and HBT-2 by a factor of 10 and HBT-3 and HBT-4 by a factor of 5.

The above results have demonstrated that, when combined with the initial noise measurement and Gummel plot, the model developed can be used to estimate the AlGaAs/GaAs HBT long-term performance and to screen unreliable HBTs without having to resort to the time consuming stress test.

4. CONCLUSION

A simple model has been developed to explicitly correlate the AlGaAs/GaAs HBT current gain long-term instability with the initial 1/f noise and leakage current characteristics. It was found that the 1/f noise provides the information on the initial surface recombination mechanism and thus the initial base surface recombination current, whereas the leakage current serves as an indicator for the initial base surface quality and thus the surface long-term

degradation process. Two groups of HBTs, two in each group, fabricated from different processes were investigated, and their current gain long-term variations subjected to different current stresses have been calculated from the model and obtained from measurements. Very good agreement is found. The approach presented can be extended to the long-term instability of the HBT under both high current and temperature stress conditions.

References

[1] M. E. Haizi, L. M. Pawlowicz, L. T. Trn, D. K. Umemoto, D. C. Streit, A. K. Oki, M. E. Kim, and K. H. Yen, "Reliability analysis of GaAs/AlGaAs HBT's under forward current/temperature stress," Dig. IEEE GaAs IC Symp., p. 329, 1990.

[2] J. J. Liou and C. I. Huang, "The base and collector currents of pre- and post-burn-in AlGaAs/GaAs heterojunction bipolar transistors," Solid-St. Electron., vol. 37, p. 1349, 1994.

[3] A. van der Ziel, Noise in Solid State Devices and Circuits, New York: Wiley, 1986.

[4] O. Jantsch, "A theory of 1/f noise at semiconductor surfaces," Solid-St. Electron., vol. 11, p. 267, 1968.

[5] R. C. Jaeger and A. J. Brodersen, "Low-frequency noise source in bipolar junction transistors," IEEE Trans. Electron Devices, vol. ED-17, p. 128, 1970.

[6] D. Costa and J. S. Harris, Jr., "Low-frequency noise properties of N-p-n AlGaAs/GaAs heterojunction bipolar transistors," IEEE Trans. Electron Devices, vol. 39, p. 2383, 1992.

[7] J. J. Liou, C. I. Huang, B. Bayraktaroglu, and D. C. Williamson, "Base and collector leakage currents of AlGaAs/GaAs heterojunction bipolar transistors," J. Appl. Phys., to appear, Aug. 1994.

[8] J. J. Liou, Advanced Semiconductor Device Physics and Modeling, Boston: Artech House, 1994, Ch. 7.

[9] Y. Zhuang and Q. Sun, "Correlation between 1/f noise and h_{FE} long-term instability in silicon bipolar devices," IEEE Trans. Electron Devices, vol. 38, p. 2540, 1991.

[10] W. Fonger, "A determination of 1/f noise sources in semiconductor diodes and transistors," in *Transistor I*, Princeton, NJ: RCA Labs, 1956.

[11] Private communication with Prof. D. Pavlidis, Dept. of Electrical & Comp. Science, University of Michigan, Ann Arbor, Michigan.

[12] M. S. Ash and H. C. Gorton, "A practical end-of-life model for semiconductor devices," IEEE Trans. Reliability, vol. 38, p. 485, 1989.

[13] D. Costa and A. Khatibzadeh, "Use of surface passivation ledges and local negative feedback to reduce amplitude modulation noise in AlGaAs/GaAs heterojunction bipolar transistors," IEEE Microwave and Guided Wave Lett., vol. 4, p. 45, 1994.

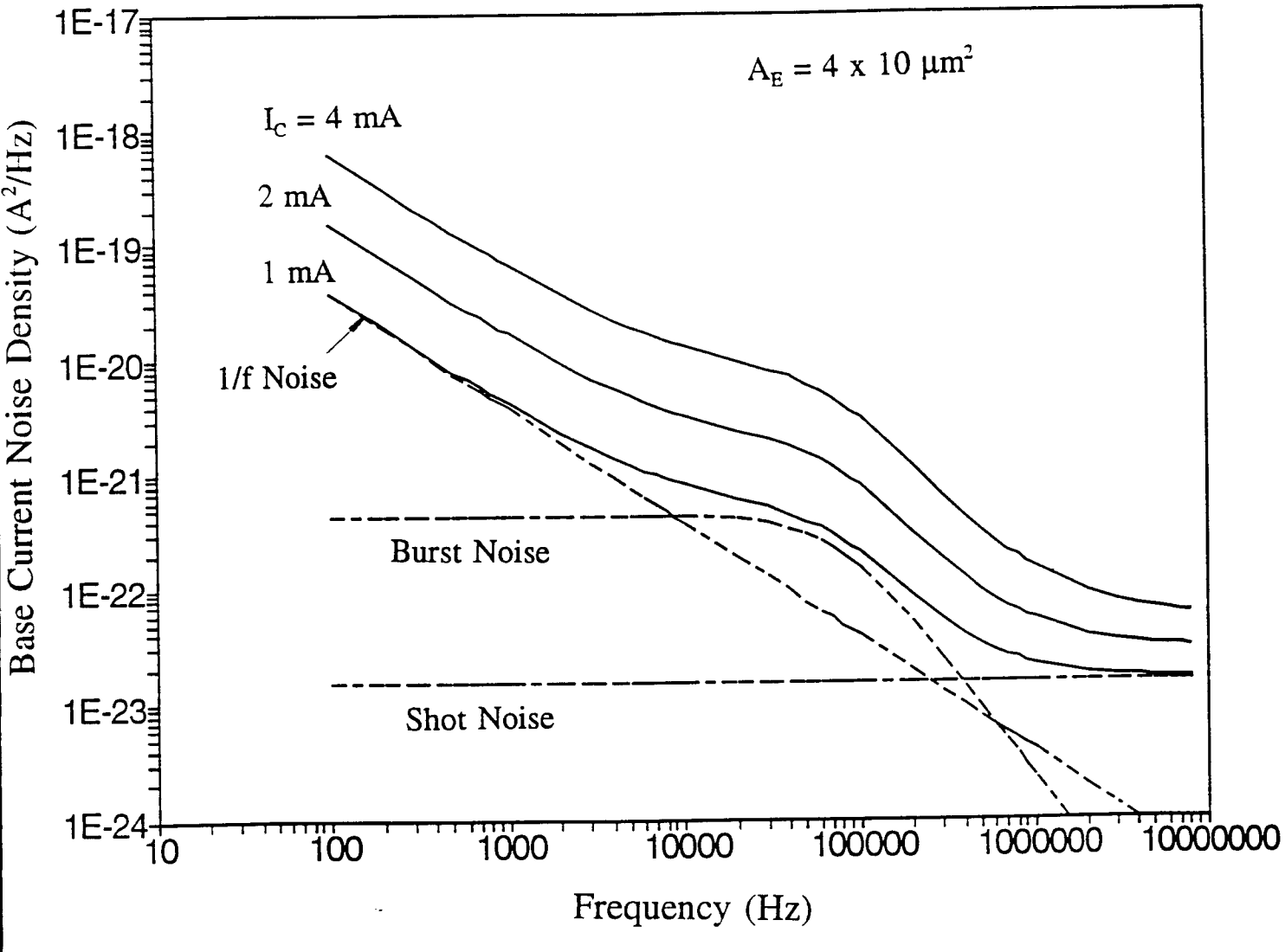


Fig. 1 Equivalent input base noise current spectral density calculated for an AlGaAs/GaAs HBT with an emitter area of 4x10 μm² under three different collector current levels.

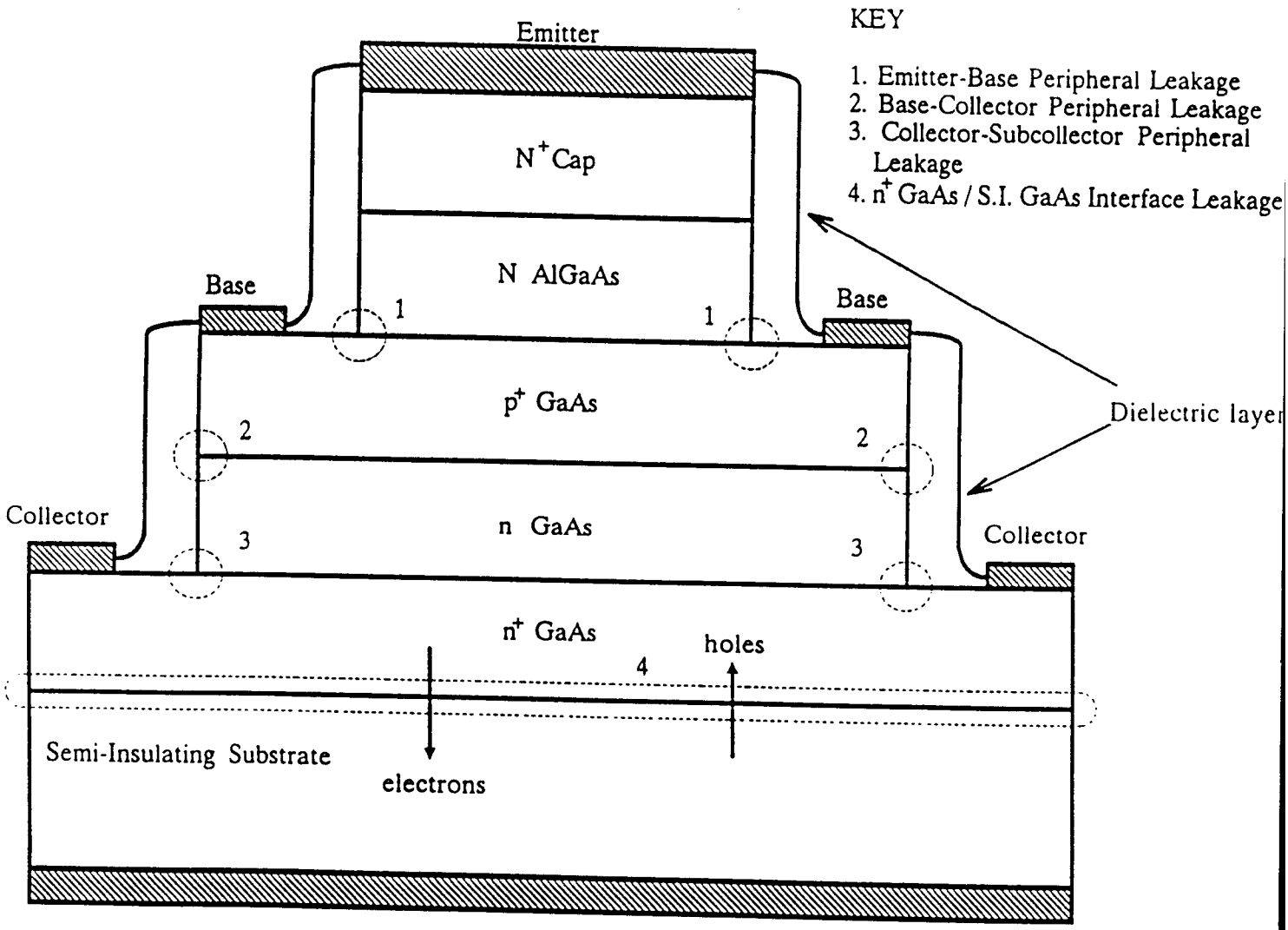


Fig. 2 HBT structure showing the four peripheries where the base and collector leakage currents can take place.

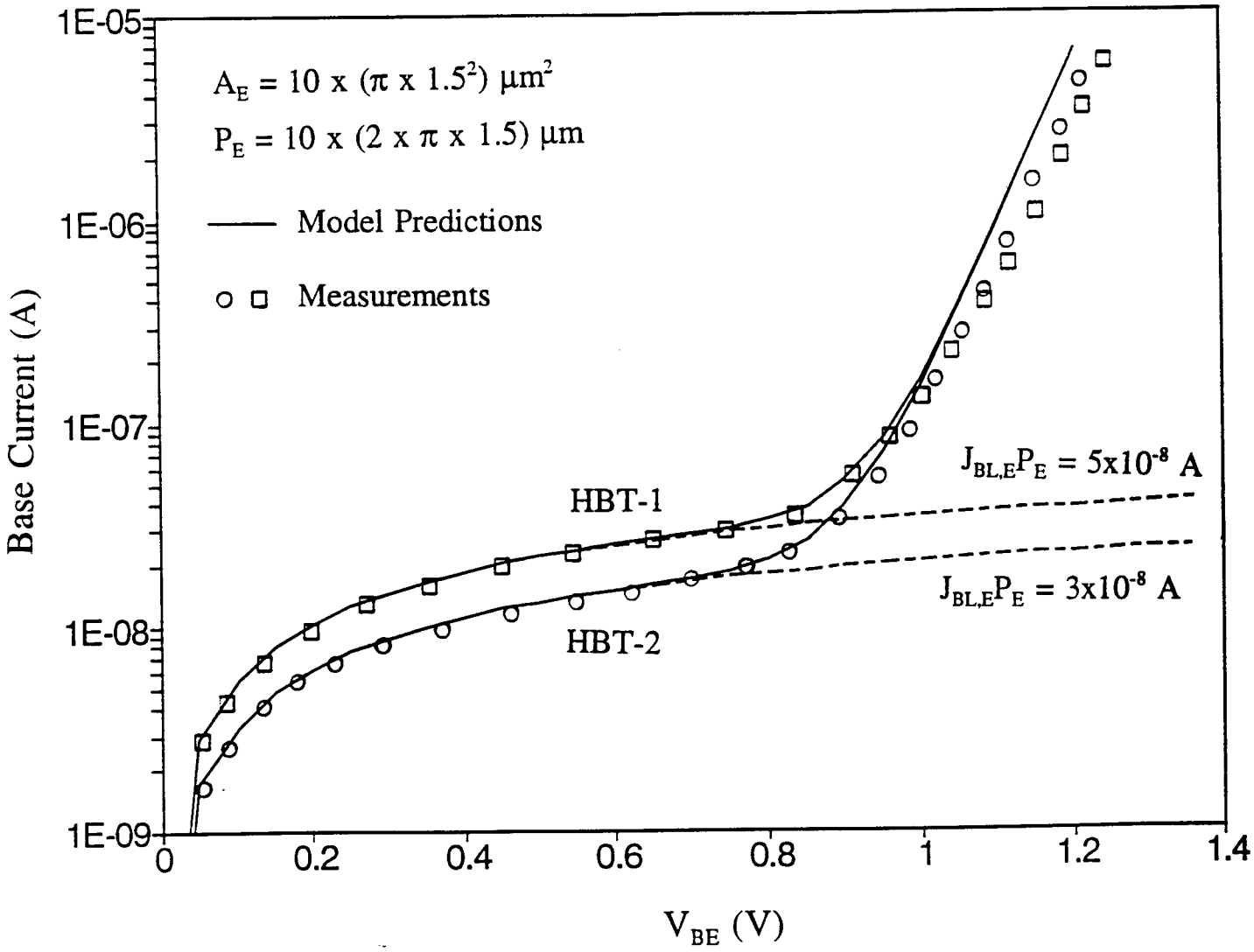


Fig. 3 HBT-1 and HBT-2 base currents, including both the leakage and normal currents, calculated from the model and obtained from measurements.

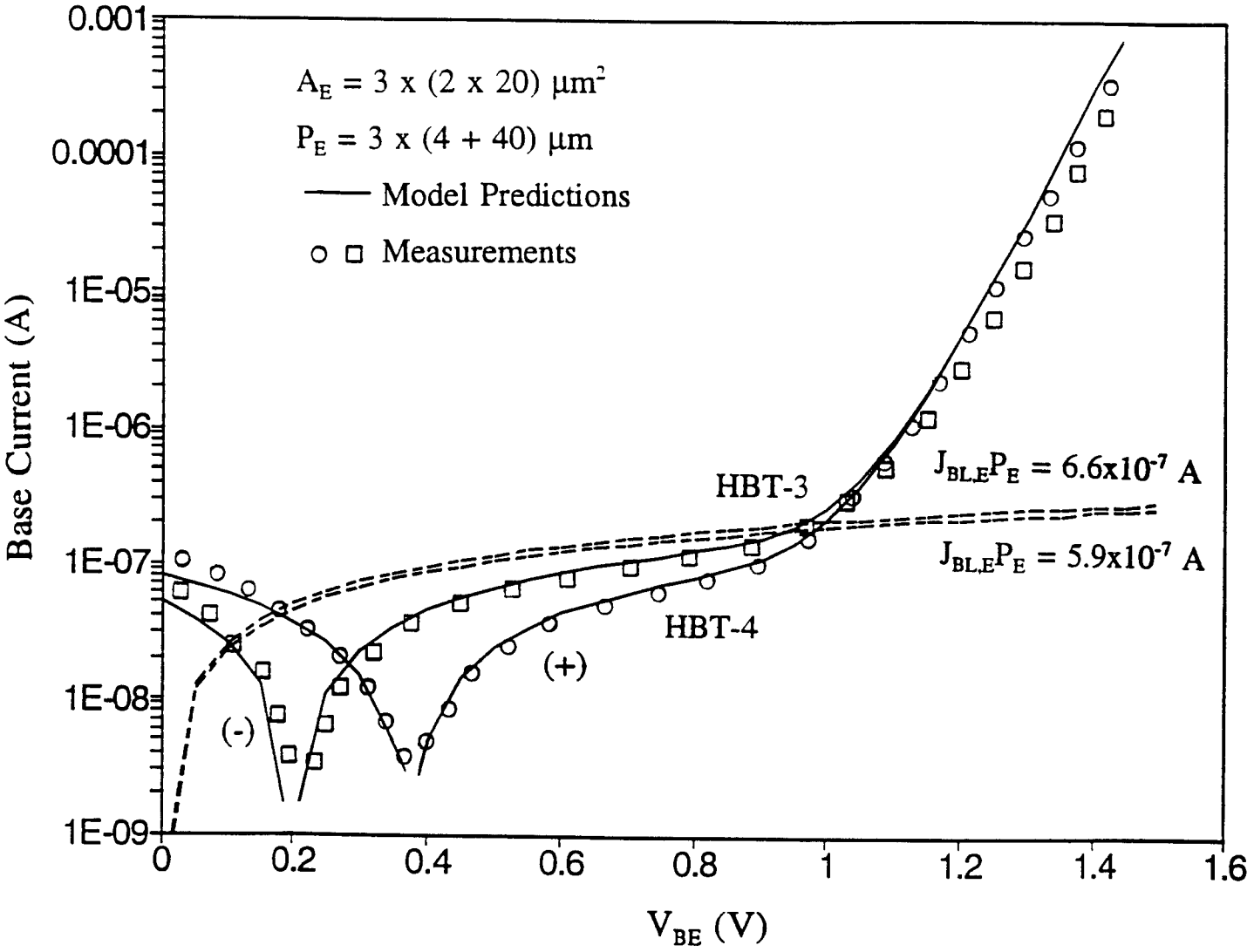


Fig. 4 HBT-3 and HBT-4 base currents, including both the leakage and normal currents, calculated from the model and obtained from measurements.

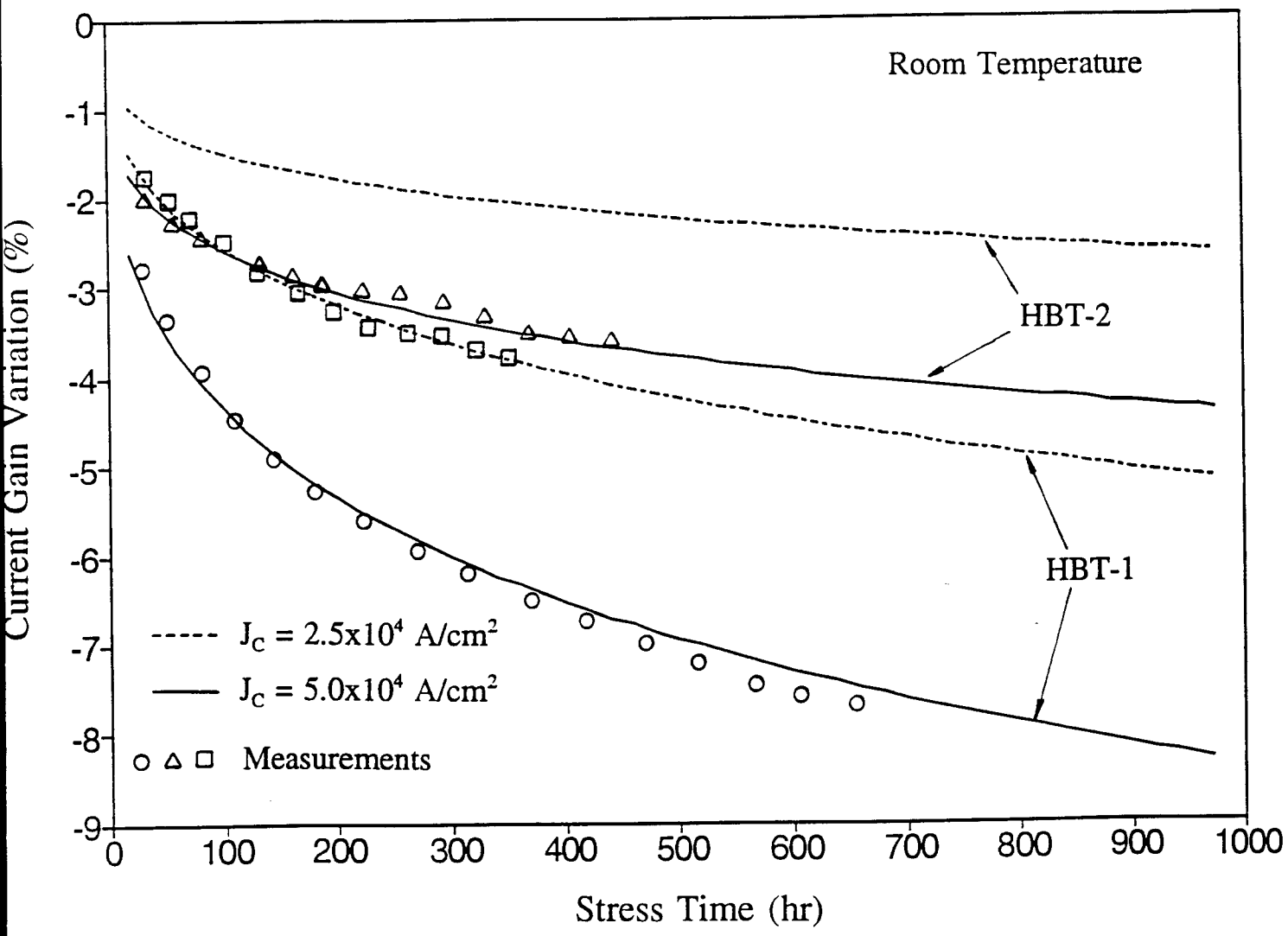


Fig. 5 Current gain long-term instability of HBT-1 and HBT-2 calculated from the model and obtained from measurements.

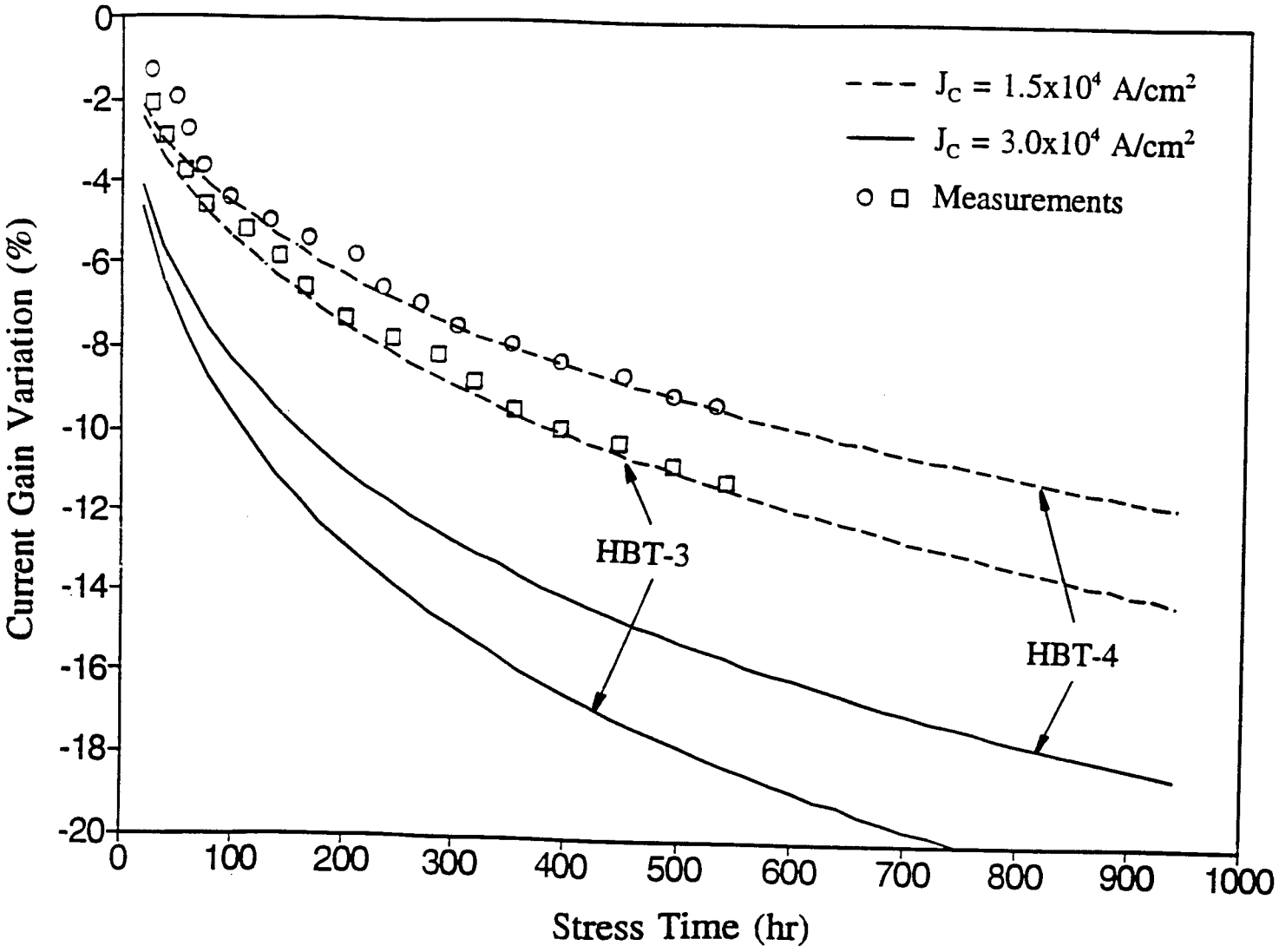


Fig. 6 Current gain long-term instability of HBT-3 and HBT-4 calculated from the model and obtained from measurements.

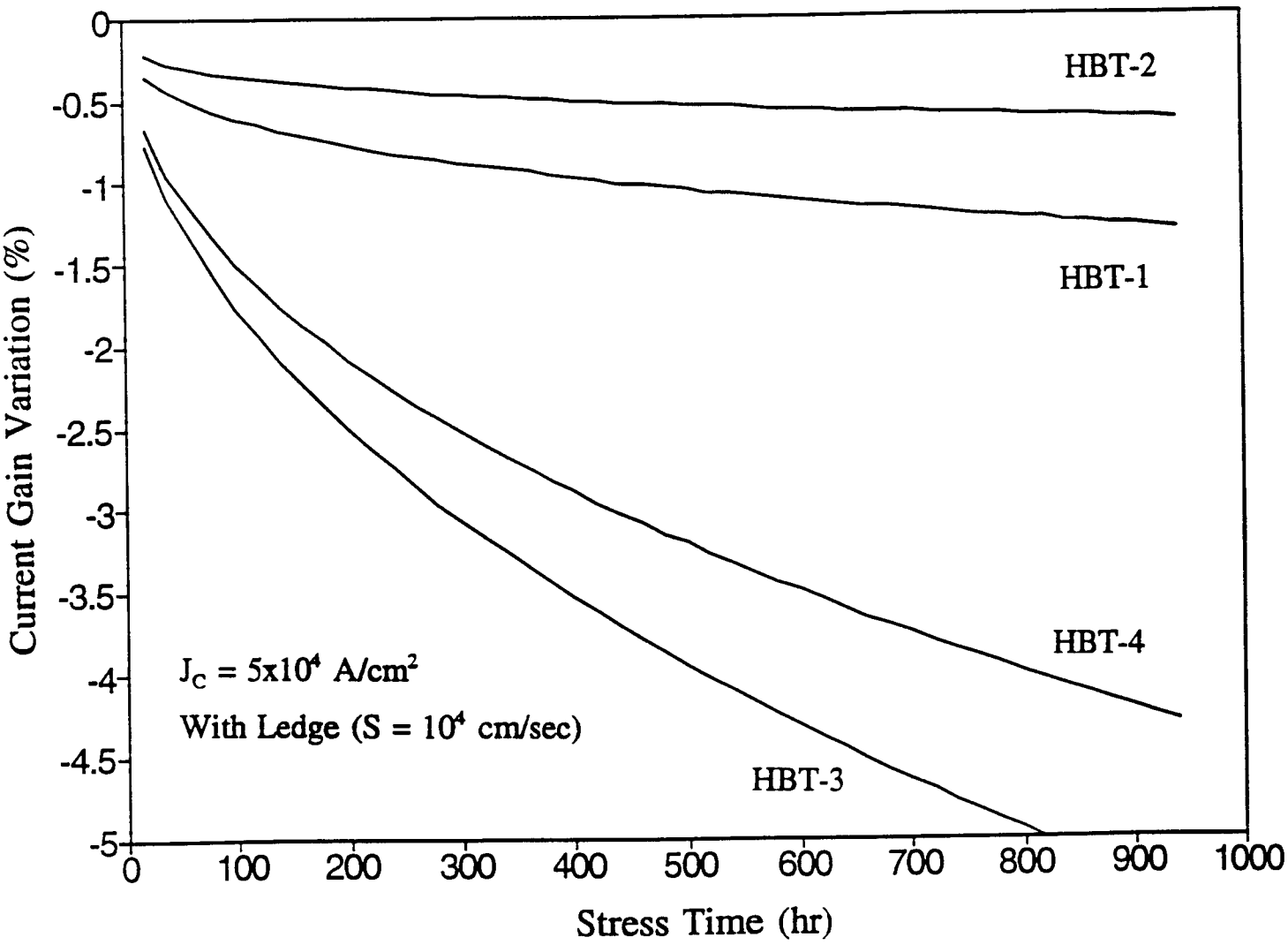


Fig. 7 Current gain variation versus time calculated assuming HBT-1, HBT-2, HBT-3, and HBT-4 have an AlGaAs ledge on the base surface.

NUMERICAL RECONSTRUCTION OF HOLOGRAMS IN ADVANCED BALLISTIC HOLOGRAPHY

James S. Marsh

Professor

Department of Physics

The University of West Florida

Pensacola, FL 32514

Final Report for:

Summer Faculty Research Program

Wright Laboratory

Armament Directorate

WL/MNSI, Eglin AFB

Sponsored by:

Airforce Office of Scientific Research

Bolling Air Force Base, Washington, D. C.

August 1994

NUMERICAL RECONSTRUCTION OF HOLOGRAMS IN ADVANCED BALLISTIC HOLOGRAPHY

James S. Marsh
Professor
Department of Physics
The University of West Florida

Abstract

The feasibility of numerical reconstruction of real holograms was studied. The fringes of real holograms are resolvable in good quality microscopes operating with visible light. Photographs of the hologram through the microscope can be made and scanned into a computer. A formalism based on the Fresnel approximation to Huygen's principle is developed leading to a mathematical simulation of the physical process of hologram reconstruction. The numerical intensive part is reduced to a two dimensional Fourier transform, to which the FFT can be applied. The formalism is successfully applied to numerically generated holograms.

NUMERICAL RECONSTRUCTION OF HOLOGRAMS IN ADVANCED BALLISTIC HOLOGRAPHY

James S. Marsh

Introduction

In advanced ballistic holography (Brooks, et. al; Hough and Gustafson), a cylindrical hologram is made of a projectile fired down the axis of the cylinder. The holographic exposure is made with light from a ruby laser (694 nm) with an exposure time of ~ 30 nsec, single or double pulsed. The idea is to capture, in three dimensions, an image of the projectile and accompanying fragments. The positions of the fragments are measured in the holographic reconstruction; if a double pulse has been used in the exposure, the velocities of the fragments may also be measured.

Data processing of the holographic image is obviously a crucial step in extracting the required information from the hologram. The principle process is to perform the reconstruction optically, in the traditional way. The reconstruction beam will be as nearly a replica of the original reference beam as possible, in the same wavelength, in order to preserve dimensional integrity and minimize aberrations of the holographic image. A video camera scans the image into a computer at various orientations to capture views from different angles. Software completes the data reduction.

As an alternate process supplementary to the principle process, there is a suggestion by Rodney Powell of WL/MNSI to scan the hologram itself, rather than the holographic image, and then to numerically, rather than optically, reconstruct the image in the hologram. In this report I describe efforts to produce a workable numerical algorithm to numerically reconstruct the hologram, to produce a scan of the actual hologram, and to assess the usefulness of this procedure.

Scanning the hologram

The hologram is a record of the interference pattern between the reference beam and the object beam formed during the exposure of the hologram. It may be regarded as a complex diffraction grating. The individual fringes in the hologram represent the information that must be collected in the scan. The

fringes in the hologram typically have a separation of 1 to 10 μ or more. Evidently simply putting the hologram down on a flat bed scanner will not capture the information.

However, 1 to 10 μ is larger than wavelengths of visible light and so the fringes will be resolvable in a high quality microscope working in the visible, with oil immersion if necessary. It will then be possible to photograph the fringes through the microscope and then to scan the photograph. An amplitude transmission hologram, produced in this lab, was viewed under a good quality microscope. This was a hologram of a real object, not a holographic diffraction grating. The fringes were clearly visible under the highest magnification available, though this was not an oil immersion microscope. It was apparent that a photograph of the microscopic view could have furnished an object quite suitable for scanning. However, no equipment for photographing the image was available. A phase transmission hologram could be captured in the same way with a phase contrast microscope.

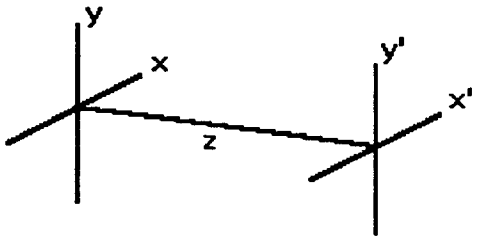
The field of view was estimated at some hundreds of fringes; this would correspond to a field of view on the order of magnitude of a millimeter. Thus using this process would be equivalent to reconstructing the holographic image through a 1 mm aperture. With objects at a nominal distance of 12 cm from the hologram, this gives a numerical aperture, $2\sin\vartheta/2$, of 1mm/120 mm. This gives an ideal resolution of $d=\lambda/2\sin\vartheta/2 \approx 100\mu$ using the light from a ruby laser, 694 nm.

This would give a resolution of the holographic objects an order of magnitude worse than will be achieved by viewing the image with the video camera, where the aperture will be centimeters wide. However, the scanned hologram can be numerically cleaned up and smoothed before image production, improving the signal to noise ratio, and perhaps some of the aberrations, of the image. This feature enhances the attractiveness of the scanned hologram. Where the balance can be struck between lower resolving power and better signal to noise ratio remains to be seen.

Numerical reconstruction

The idea for numerical reconstruction goes back to Rolleston and George, and before that, to Perlmuttern, Goodman, and Macovski. Rolleston and George numerically simulate image formation by a

lens using the Fresnel approximation. Their trick is to formulate this in a way that turns the numerically intensive part into a two dimensional Fourier transform. It is this idea which is modified here to produce a numerical simulation of the physical process of holographic image formation.



Assume a monochromatic wave field propagating from left to right, generally along the z axis. If the distribution of the field is E(x,y) in the x,y plane, then the distribution of the field E(x',y') in the x',y' plane a distance z from the x,y plane will be given, up to some unimportant factors, by

$$E(x',y') = \int dx\, dy\, E(x,y) \exp[i(k/2z)\{(x-x')^2 + (y-y')^2\}]$$

in the Fresnel approximation. k is the wave number, $k=2\pi/\lambda$, where λ is the wavelength of the radiation. Expand the quadratic factors in the exponent and rewrite this as

$$E(x',y') = \int dx\, dy\, E(x,y) \exp\{(ik/2z)(x^2+y^2)\} \exp\{(-ik/z)\,(xx'+yy')\}.$$

We have ignored the factor $\exp\{(ik/2z)\,(x'^2+y'^2)\}$ since the quantity of interest is $|E(x',y')|^2$.

This reveals that, up to unimportant factors, E(x',y') is just the two dimensional Fourier transform of $E(x,y) \exp\{(ik/2z)\,(x^2+y^2)\}$. This means that the fast Fourier transform can be use to generate the field in the x',y' plane from the field in the x,y plane. This makes the process numerically feasible on small computers.

To form E(x,y) in the integral we use the scanned density function of the hologram, T(x,y). Then we imagine the hologram to be illuminated by a reference beam, $E_0(x,y)$. The form that this takes is

generated in the computer. Then E(x,y) will be given by

$$E(x,y) = E_0(x,y)\, T(x,y)$$

or

$$E(x,y) = E_0(x,y) \exp\{imT(x,y)\}$$

depending on whether we want to generate the image as from an amplitude hologram or a phase hologram. the factor m in the second form is a phase modulation constant. As in physical holography, it can be chosen to eliminate the undiffracted light.

This process mathematically models the physical process used in reconstructing the actual hologram.

Numerical Simulation.

In order to test the feasibility of the numerical reconstruction process, an artificial hologram of a simple object was generated in the computer and then the process described in the preceding section was used to generate the object in the hologram. The object was 1 or more point objects. A point object at position x_0, y_0, z_0 would be represented by a spherical wave $\varphi_0 = \exp\{ik\,(z_0^2 + (x-x_0)^2 + (y-y_0)^2)^{1/2}\}$. To form a hologram of a collection of point objects we compute and plot

$$T(x,y) = |\, A + \Sigma\, \varphi_j \,|^2 .$$

φ_j is the wave representing the j^{th} point object and A is the reference wave. A has the form $B\,\varphi_0$, if we want the reference wave to be a spherical wave emanating from the point x_0,y_0,z_0, or $B \exp(i\alpha x)$ if we prefer the reference wave to be a plane wave coming in from the side. With $\alpha = 0$, the plane wave hits the plane of the hologram normally. The only constraint is that the amplitude of the reference wave, B, must be large enough that the reference wave is everywhere larger than the sum of the object waves. If this is not the

case then you get into the non-linear region in the reconstruction and can generate false images, etc. This forms a convenient way to study non-linear effects if that is your interest.

To reconstruct the hologram, we form $E(x,y) = E_0(x,y)\, T(x,y)$ as described above. To achieve the most faithful reconstruction, $E_0(x,y)$ should have the same form as the reference wave A. If E_0 departs from A then various aberrations will show up in the images, as described at length in the holographic literature. This numerical method also forms a convenient way to study those aberrations.

We successfully reconstructed images from computer generated holograms, with good image quality when the reconstructing beam, E_0 was the same as the reference wave A. We also saw multiple images from holograms in the non-linear regime, and aberrations that arise from E_0 not being identical to A. This convinces us that numerical reconstruction of holograms is mathematically and computationally feasible.

BIBLIOGRAPHY

R. E. Brooks, L. O. Heflinger, and R. F. Wuerker, "Pulsed Laser Holograms," IEEE J. Quantum Electron. **QE-2**, 275 (1966)

G. R. Hough and D. M. Gustafson, "Ballistic Applications of Lasers," SPIE Volume 1155, *Ultra Speed and High Speed Photography, Photonics, and Videography '89*, pp. 181-188 (1989)

R. J. Perlmutter, J. W. Goodman, and A. Macovski, "Digital holographic display of medical CT images", SPIE volume 367, *Processing andDisplay of Three-Dimensional Data*, pp 109-116 (1982)

R. Rolleston and N. George, "Image reconstruction from partial Fresnel zone information", Appl. Opt., **25**, 178-183 (1986)

INTEGRATED INFORMATION MANAGEMENT FOR ATR RESEARCH AND DEVELOPMENT

Rajiv Mehrotra
Associate Professor
Department of Mathematics and Computer Science

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

Final Report for
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

INTEGRATED INFORMATION MANAGEMENT FOR ATR RESEARCH AND DEVELOPMENT

Rajiv Mehrotra
Associate Professor
Department of Mathematics and Computer Science
University of Missouri-St. Louis

Abstract

A wide variety of alphanumeric, spatial, signature, modeling and image data is collected to support ATR research and development. ATR algorithm design, development and evaluation completely rely upon this data. Since the cost of both data collection and algorithm development is quite high, an efficient management of the necessary data is a minimum requirement for a successful ATR development program. With respect to the management and usage, the diversity and huge volume of this information pose some major challenges. To store, manipulate, and manage this diverse set of information, an integrated information management system that is capable of efficiently storing, retrieving, and manipulating alphanumeric, spatial, and multidimensional image information and pertinent algorithms need to be designed and implemented. This report discusses the central features and requirements of such an integrated ATR information management system (IAIMS), reviews the existing data management structures used by WL/AARA and its contractors, and outlines the key design and implementation issues.

INTEGRATED INFORMATION MANAGEMENT FOR ATR RESEARCH AND DEVELOPMENT

Rajiv Mehrotra

1. Introduction

A wide variety of signature and imagery data (see table 1) and ancillary data (see table 2) is collected on a regular basis by various DoD agencies and contractors. These data collection activities take place periodically in both controlled situations as well as operational (exercise) environments. For each type of imagery or signature data, a number of additional associated parameters (see table 3) are collected and the number and types of parameters varies with the image or sensor type. The total cost of planning and implementing a data collection activity is usually quite high. In addition to such measured data sets, a number of synthetic data collections are regularly generated via computer algorithms and computer target models. Air Force laboratories, universities, and other organizations involved in various aspects of ATR research and development activities utilize information from these collections for algorithm development, testing and performance evaluation. Again, a number of such efforts take place simultaneously and usually at a high cost. Extensive evaluation of algorithms and their relative performance analyses using some standard criteria is essential. The high cost and complexity of both the data collection and the algorithm development and evaluation tasks demand an efficient information environment which facilitates effective and integrated management of all imagery, signatures, ancillary data, associated parameters and other information necessary to perform ATR algorithm development.

SENSOR	Signature Type
Radar	RCS UHRR SAR, ISAR, IFSAR JEM
Electro-optical	FLIR (1st and 2nd gen.)
Lasar	LADAR
Acoustic	
Seismic	

Table 1. Types of signature data used in ATR research.

Ancillary ATR Data
Maps
Algorithms
Models
Meteorological Data
Performance Results

Table 2. List of typical ancillary data used in ATR research.

Associated ATR Parameters
Ground Truth
Target Articulation
Sensor Parameters
Security Information

Table 3. List of typical associated parameters used in ATR research.

WL/AARA has several large collections of data from a variety of sensors. Currently, each data set has a separate data storage and management system. These systems do not treat alphanumeric and imagery data equally. In other words, current approaches to the ATR data and related information management lack complete integration. Furthermore, these systems do not in any way directly aid in tracking the progress and management of data collection, data distribution, and algorithm development and evaluation tasks. The types of data retrieval are limited and in some cases retrieval of a desired set of data is not a straight forward process. There is no uniform software environment to support and manage data.

Recently, there has been a great deal of interest at WL/AARA and other DoD agencies in developing standards for data management and building an integrated information management environment. As a step towards this goal, WL/AARA awarded a contract to ERIM under PEATRS program to develop a baseline integrated data management system called ATR Data Management Environment (ADME). Also, a contract has been awarded to Veda to create an ORACLE-based relational database management system for storage and management of NCTI (Non Cooperative Target Identification) data. WL/AARA is also participating in several Air Force and DoD wide programs, such as ATRWG, Argus, JESEBEL and Header Consolidate, to move towards unified databasing practices.

This report outlines the requirements and objectives of an integrated information management system capable of effectively and efficiently storing and managing all pertinent ATR information. Key existing ATR data management systems are reviewed. Major design and implementation issues posed by the diversity of information and its usage are also discussed.

2. Objectives

Information or data management systems are used to store and manage organizational (or application domain) data in a way that gives users the illusion that their view of the data corresponds to the way that the data is actually represented in the system. The basic idea behind management of information is to enable sharing of the common stored data for a variety of applications or to assist in management and control of several tasks or operations in the application domain. The main objective of the ATR information management system is to manage information necessary for the effective management of all activities related to the ATR research and development. This objective translates into management of information to support four major activities: data collection, data storage, data distribution, and algorithm development and evaluation.

Typically, a **data collection** activity involves a collection plan which includes applicable information relative to the target(s), sensor(s), and local conditions that could effect the measurements. In order to implement, track progress , and perform post-collection evaluation, all information associated with a data collection plan must be stored and managed. This requires integrated management of spatial alphanumeric, and imagery data. The ability to efficiently retrieve, manipulate and display all three types of data through a user friendly interface is also desirable. A few examples of how a user may interface with the database for the purpose of data collection are as follows.

- Ability to interactively edit spatial data (i.e., map of a region or a collection plan to include the flight-lines and target locations to create a plan map or an actual collection map).
- Ability to display the map(s) associated with a given collection plan with flight-lines and marked target locations.
- Ability to retrieve information regarding target/object identification and its setup parameters, ground truth, or the collected imagery and other related data by clicking on the target location in a map.
- Ability to highlight, in a displayed map, locations of targets with a given identification (and/or setup parameters).
- Ability to locate and display plan (or collection) maps containing a specified set of targets.

Effective **data storage and distribution** is the most important objective of many ATR data management systems. Collected imagery, signature data and their associated parameters must be stored in a data management system. In many cases, the collected imagery (or signature) needs to be preprocessed (e.g., FFT or image enhancement) to create the proper input for users. The preprocessing needed varies for different ATR algorithms as well as for the sensor used to collect the data. The preprocessed data is also usually stored in the database. This duplication of data compounds to the data storage problem.

Distributing data on a regular basis or on demand is a major task associated with the database. Typically, data distribution involves selection of data based on some constrained parameter values specified by the users. Typical parameters include target aspect, image quality, or image contents. Data sets to be distributed may also involve general browsing and selection of stored data. However, due to security or data sequestering, some users can not be given access to the entire database. They are allowed access to selected data sets based on their needs. This requires the definition of multiple logical views of the stored data, where a view meets the needs of a user. With each view a list of authorized users needs to be associated in order to maintain data security. To effectively control data distribution task, information pertinent to customers, their data requests, responses to requests, etc. must also be stored and managed. Some other desirable characteristics of an IAIMS from the ATR data storage and distribution tasks viewpoint are:

- Ability to browse stored data (with proper display) to quickly identify signature (image) types.
- Ability to suitably display selected information (alphanumeric, signature plots or images).
- Ability of interactively select a set of images and associated parameters, that satisfy specified constraints on parameter values.
- Ability to retrieve (or select) other images of the same target, that is contained in the given (or selected) image.
- Ability to create, modify, and delete external logical views of the stored data and associated access authorizations.
- Ability to select a set of images that satisfy certain constraints, automatically apply some specified (pre)processing to selected set, and store the results temporarily or permanently.
- Ability to display (or select) all preprocessed signatures/images of a specified type that were obtained (as a result of processing) using a specified algorithm.
- Given a preprocessed image, ability to display the associated original image(s).

• Ability to share and modify existing set of preprocessing or synthetic imagery (or signature) generation algorithms.
• Ability to add new preprocessing or synthetic data generation algorithms to the information system.

Algorithm development and evaluation tasks involve coding new techniques utilizing existing ATR tools and algorithms, selecting desired collected data, running one or more selected algorithms on the selected data, and storing test and performance evaluation results. Algorithm evaluation techniques utilize signatures, imagery, ancillary data, associated target parameters, and ATR algorithms to generate performance analysis results in standard formats. Usually, a number of contractors, academia as well as DoD scientists are engaged in this activity.

From the algorithm development point of view, the information management system should have the ability to store and permit sharing of existing image processing algorithms and ATR tools in a user friendly environment for existing and newly developed code. From the algorithm evaluation point of view, the information system should have the ability to apply a selected set of algorithms to a selected set of imagery data and store the results (permanently or temporarily). It should enable classification and retrieval of stored data on the basis performance results of one or more ATR algorithms. To manage ATR algorithm development contracts, information regarding contractors, contracts, progress reports, periodic evaluation, deliverables, and contract status, etc. must also be maintained.

3. Integrated ATR Information Management

It is clear from the above discussions, that the key classes of information that must be managed by an IAIMS are

- signature and image information (imagery)
- signature or image related content independent information (alphanumeric)
- signature or image related content dependent information (alphanumeric)
- spatial information
- model information
- algorithmic information
- other alphanumeric data

The **first** type of information, signature and image information, includes the originally collected signatures and images as well as processed signature or images. As mentioned earlier,

a variety of imagery data, ranging from 1-dimensional to multidimensional, is used in ATR research. The **second** type of information, signature or image content independent information, includes image resolution, date, location, and time of collection, sensor setup parameters, atmospheric conditions, etc. which are found in the file header. The **third** type of information, signature or image related content dependent information, is the ground truth data. It provides the target identification, target location and orientation, etc. This information may also be found in the file header. All of the first three types of information come with the collected or generated data The **fourth** type of information, spatial information, includes maps of the collection plans and actual data collection activities. The **fifth** type of information, model information, includes the CAD model used for synthetic data generation and the models used by target recognition algorithms. The **sixth** type of information, algorithmic information, include all the preprocessing algorithms, signal or image processing and analysis algorithms, ATR algorithms, image quality assessment algorithms, algorithm evaluation and scoring algorithms. Finally, the **seventh** type of information, other alphanumeric information, includes (i) customers, data requests, and projects information needed to keep track of customers' data requests, data distribution activities, algorithm development contracts, and contract progress and evaluation, etc., (ii) algorithms' performance results to manage the algorithm evaluation task, and (iii) other model, algorithm, collection plan related information related.

An IAIMS must facilitate the storage and management of each of these seven classes of information. The advantages of data independence, data integrity, data sharing, controlled redundancy, and security offered by the conventional database management systems for alphanumeric data are required here for alphanumeric, signature, image, model, and map data.

A generic architecture for an IAIMS is shown in Fig. 1. There are four repositories of information: an alphanumeric database, a signal and imagery database, a model database, and a spatial database. These four repositories are shown separately since they are logically independent of each other. Any implementation of this generic architecture, however may choose to place these repositories in a single database or further decompose them (e.g., the signal/imagery database can be decomposed into 1-d signature database, 2-d image database, 3-d image database, etc.). The **alphanumeric database** holds all the information in the header, additional information about target, algorithm evaluation results, image quality assessment results, alphanumeric information pertinent to data collection plans and activities, etc. The **model database** contains all the CAD models used for generating synthetic signatures and images, as well as models user by ATR algorithms. The **signal and image database** contains the collected signatures and imagery data, processed signatures and imagery data, and generated

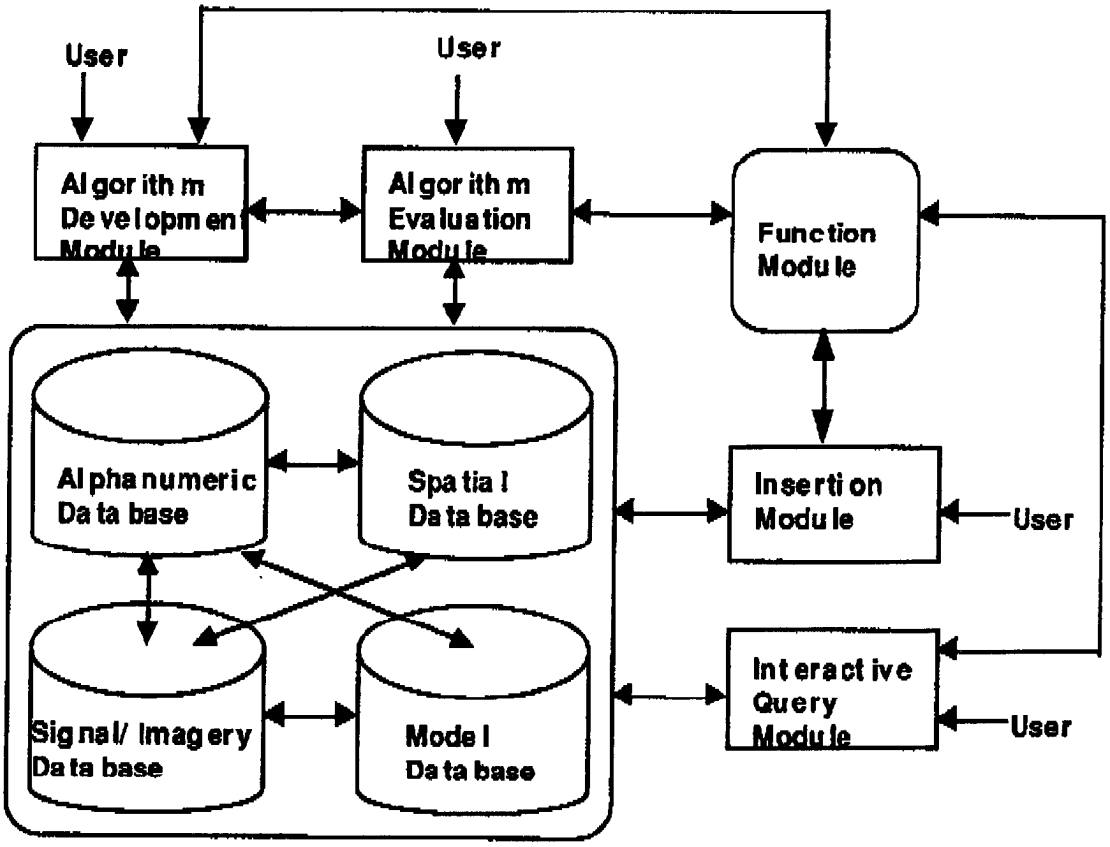


Fig. 1: A Generic Architecture for an IAIMS

synthetic signatures and imagery data. The **spatial database** holds region maps, collection plan maps, actual collection maps, georeferenced maps, etc. The **function module** is the collection of all algorithms needed for ATR research and development activities. It includes all the ATR algorithms, signal/image processing tools, spatial data processing tools, georeferencing algorithms, synthetic data generation algorithms, algorithm evaluation methods, etc. Note that the ATR information management system is expected to be insertion and retrieval (query) intensive and not update intensive.

4. Existing Data Management Systems

The current data management schemes are primarily designed for data storage and distribution. A variety of data management structures are used to store and manage different classes of signatures and imagery data. The largest database effort in WL/AARA is for the NCTI (non cooperative target identification program). A system, called RSRN (Radar Signature Retrieval Network) was developed for storage and distribution of radar signature data for the

development, training, and testing of UHRR (Ultra-High Range Resolution) classification algorithms. The RSRN contains raw video UHRR data only. A new Oracle based RDBMS system is being developed to house the RSRN data as well as processed and synthetically produced UHRR data.

Veda developed the RSRN database system for Automatic Radar Target Identification (ARTI) data, while other contractors and users developed their own in-house systems to manage their local data sets. All these data management structures are based on flat files and utilize a set of programs for management of data (i.e., do not have any database management system). Currently, the data files for all ARTI data types consists of records composed of four classes of information:

1. **Generic Header Information** – This information is stored in the first 256 bytes of a data record and provides general information and description of the signature type (in terms of signature data type, data source, data size, aircraft type and model, collection time and date, sensor parameters, etc.) in an ASCII readable form. Thus, only a selected first few bytes of each record need to be examined to determine the signature type. The same set of parameters are included in the generic header of all types of data.

2. **Summary Information** – Following generic header, the summary information is stored in a total of 80 bytes. This information is a set of floating point numbers and it varies for different types of data. Summary information is some additional information about the signature which includes description of (pre)processing and results (e.g., rejection test results).

3. **Native Header** – After summary information, the native header information is provided. This is simply the information in the header of the originally provided data (collected or synthetic). The native header for ARTI data is of size 561 bytes, whereas the size of the native header for synthetic data is 361.

4. **Data Samples (Signatures)** – The data samples are stored after the native header. These are pairs (number of pairs vary for different ARTI and synthetic data) of floating point numbers representing the complex I and Q values for the particular range, polarization, and Doppler bin.

This is a flat file-based data management system. The data file can be searched via one or more

of the predefined indexable parameters. A set of library routines are available for file manipulation, data retrieval, and data interpretation. This system is supporting only the data storage task. Since different users (contractors) use their own data management structures, sharing original and processed data, and results among contractors and Air Force personnel is not an easy task. Also, data distribution as per users request is not quite efficient.

Recently, a project was initiated to convert flat file-based NCTI data management system to an ORACLE (relational database) based database management system. In a relational database system, the world associated with the application domain is modeled in terms of entities and relationships among entities. These entities and relationships generally have some meaningful counterpart in the application domain. The initial proposal for the table definitions for processed ARTI Phase III data contains two tables: *track table* and *signature table.* The track table hold information that remains constant for each signature within a track/subtrack of GTAMS ARTI data. The signature table contains information which changes on a signature by signature basis. The model underlying this design doesn't represent (or abstract) the ATR problem (data) domain in a meaningful way. This design models the ATR problem (data) domain in terms of entities constant and variable, which do not correspond to anything in the ATR problem domain. A model that represents the ATR problem domain in terms of entities and relationships (or tables), like sensor, operator, target, target_setup, model, signature, etc., which correspond to real entities in the ATR data domain, represent the ATR world more meaningfully. Implementation of such a model in ORACLE (or any other database management system) will not only be modular, flexible, and easy to use, it will also reduce redundancy in the stored data (i.e., repetition of the same fields in several records) and improve data consistency.

The ADME project is primarily concerned with the management of data collection and algorithm evaluation activities. Though the overall objectives and functionality of the ADME are being formulated at this point, it is clear that it will permit storage and management of alphanumeric, imagery, and map data. It will offer a some what integrated management of collected data and related information as well as performance evaluation results. It will also offer a much improved system for data selection (for distribution). It is expected to be a tool set that will assist in planning and management of data collection and algorithm evaluation tasks. Though the ADME will offer a much improved baseline information management environment than the existing ones, it will not be a fully integrated (i.e., alphanumeric, map, and imagery data will be not be treated equally) system and only a limited types of imagery and map data-based retrievals and manipulation will be permitted. These limitations may be primarily due to the

scope of the project.

5. Design and Implementation of an IAIMS

One possible model to implement an IAIMS that meets all the aforementioned objectives is shown in Fig. 2. This is a modular approach to implementation since the entire system is implemented through a number of independent, logical modules. The most important and complicated component of this implementation model is the IAIMS controller, which is the driver of the entire ATR information management system. Some or all of the logical modules can be implemented using available (commercially or public domain) information management systems and programming environments. For example, GRASS can be used as the spatial database management system, ORACLE can be used as the alphanumeric database management system, and any program library management system can be as the algorithm management system. This approach also offers the advantage of integrating the existing ATR database system (e.g., ORACLE-based NCTI database and ADME) with other database systems by designing and implementing a proper IAIMS controller.

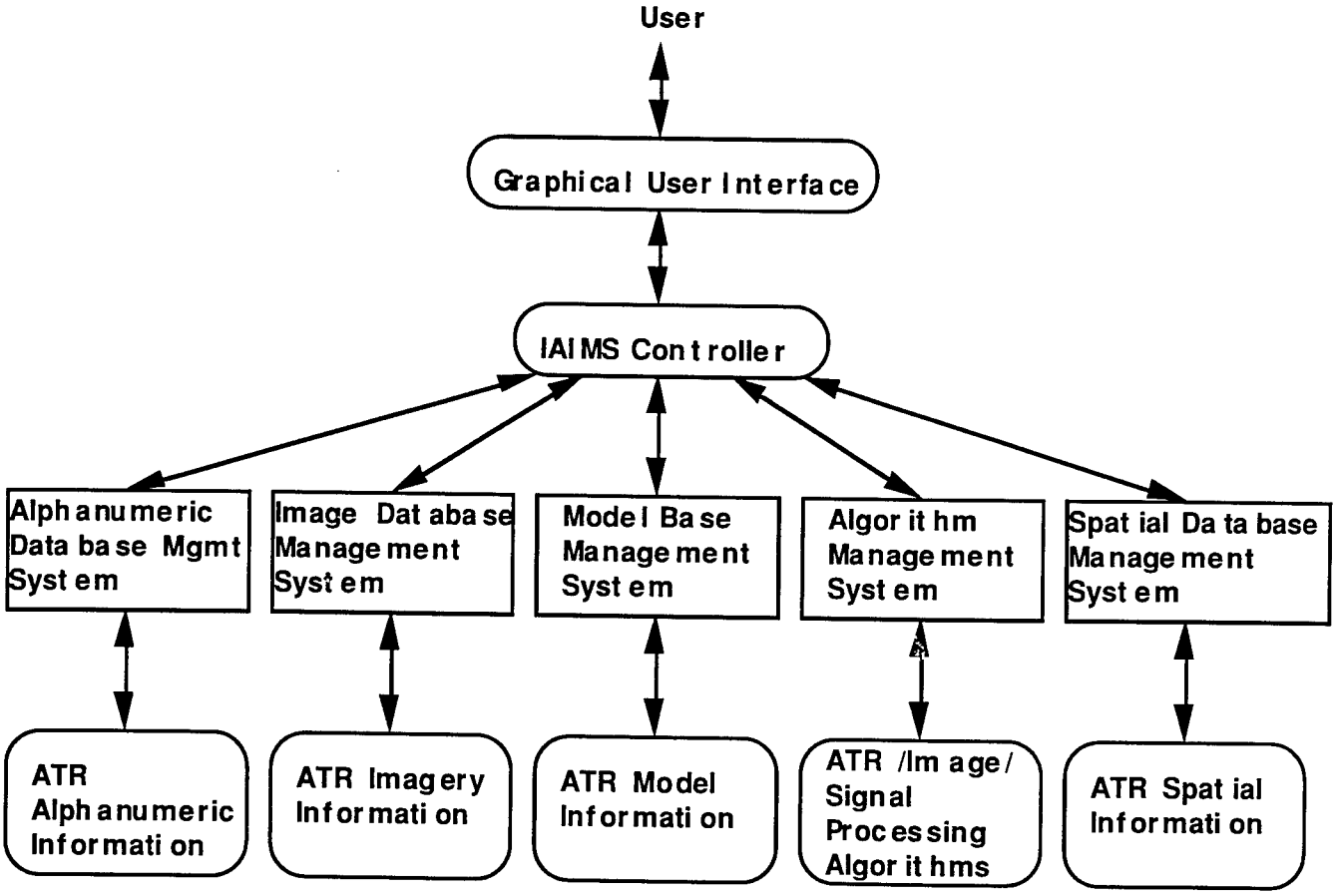


Fig. 2. Implementation Model for an IAIMS

As far as available database management systems are concerned, alphanumeric data management is a highly developed technology and several public-domain and commercial database management systems are available. Relational database management systems are the most popular these days for handling alphanumeric data. However, due to the advantages and flexibility offered by object-oriented technology, object-oriented database systems are rapidly gaining popularity. Several relational database management systems have been or are being extended to offer object-oriented features. Also, a few object-oriented database management and programming environments are now available.

The object-oriented technology has also made a great deal of impact and improvement on spatial database management systems but image database management technology is relatively new. Limitations of hierarchical, network, and relational data models are the primary reasons for the slow growth of image data management technology. With the advances in the object-oriented technology and knowledge-base system, several proposals and product announcements for elementary object-oriented image database management systems have recently appeared.

The complexity (or simplicity) of the overall IAIMS and the IAIMS controller module depends on the implementation of individual database management types. If all database management modules use a single database management system, the design and implementation of both the overall system and the IAIMS controller will be highly simplified compared to the case of separate database management systems for each of the database module. Only an object-oriented database management system can be used to manage all types of information. Thus, using the object oriented technology can simplify the design and implementation of an IAIMS. Advantages of object-oriented technology, potential concerns, and benefits over other technologies[1] are as follows:

Advantages:

1. Faster development
2. Higher quality
3. Easier maintenance
4. Reduced cost
5. Increased scalability
6. Better information structures
7. Increased adaptability

Potential Concerns:

1. Maturity of technology
2. Needs for standards
3. Needs for better tools
4. Speed of execution
5. Availability of qualified people
6. Cost of conversion
7. Support for large-scale modularity

[1] David A. Taylor, *Object-Oriented Technology,* Addison-Wesley, 1992, pp. 103-107.

Potential Benefits:

1. Reusability - New classes can built on existing classes
2. More reliable system
3. More rapid development
4. More flexible development
5. Models that more closely reflect reality
6. Better handling of nontraditional data types (e.g., imagery, spatial, CAD models)
7. Better technology independence and interoperability
8. Better basis for client-server and distributed computing
9. Availability of commercial class libraries.

Since two major WL/AARA funded ATR database efforts (ORACLE-based NCTI and ADME) have already started it would be cost effective to design and implement the IAIMS in such a way that the end products of those two efforts could be integrated into the systems without any major modifications. Furthermore, realizing that WL/AARA is a ORACLE house, the best approach to quickly implement a system that meets most of the WL/AARA data management needs, is to only implement only those additional logical components (i.e., in addition to the NCTI and ADME), that could be implemented under ORACLE. This implies that the requirement of managing images in an integrated manner with other types of data so that image-content-based retrieval is possible will have to be forfeited initially. Since image data type (or even vector type) and associated valid operations are not available in ORACLE, symbolic pointers (identifiers that point to image locations) will have to be used for image data. Images may have to be described in terms of alphanumeric descriptors (simple data types). Also, application programs will have to be written (i) to extract image representation and load them into corresponding tables, (ii) to interpret image information, and (iii) to retrieve images (using symbolic pointers). Access restrictions may have to imposed to ensure that the users don't update the image representation and image interpretation related table entries generated by the application programs. In other words, with an increase in the image information to be managed, the complexity and rigidity of the ORACLE-based information management system will also increase. In fact, the entire ATR information management system can be designed and implemented in ORACLE, but that would require a large number of complex application programs, which, in some sense, would simulate a desirable object-oriented environment.

The design and implementation of the algorithm management system and the algorithm database can also be postponed. The most important role of the algorithm management component, in an

integrated information management environment, is to manage a library of functions that can be used to defines complex object classes (i.e., data structures and associated valid operations). Feature of complex object definition is particularly desirable for image data and associated semantics. This feature is essentially needed for integrated management of image information. In absence of integrated image information management, the algorithm management component will also not be urgently needed initially.

Thus, major components (in addition to the NCTI database and the ADME, which may need to incorporate some additional features) that need to be implemented initially include those that are responsible for the management of data storage and distribution tasks, the model-base management (ORACLE-based model management will have limitations similar to ORACLE-based image management), a suitable GUI, and the ATR information management system controller, which controls all the underlying systems. If this approach is adopted, all the underlying system will be based on the ORACLE database management system and hence the overall ATR information management system, the GUI, and the IAIMS controller will be simplified. The other components can be implemented later when acceptable standards evolve and associated technologies mature.

The integrated management of imagery, the associated processing tools, and ATR algorithms pose the most challenges to the design and implementation of the complete IAIMS. The key reasons for the difficulty are:

- the lack of understanding the nature of imagery data and a suitable data model
- the lack of standards
- the need for image content-based or image semantic-based information retrieval methods

Actually, the whole area of integrated image information management is in its infancy and is currently a major focus of research in both the image processing as well as the database communities. Several issues including image information modeling to handle imprecision and multiple semantics, memory management, automatic semantic extraction, content-based retrieval, integration of multiple feature-based indexing, efficient query processing, and image integrity are topics of active research. In the following sections, two key problems in integrated management of image information, the image information modeling and the image information retrieval are briefly discussed. I propose to pursue research on both these issues in context of generalized image information management as well as the ATR image/signature information management.

6. Image Information Modeling

The purpose of a database or information management system is to manage an abstracted representation (in terms of the available structures) of the real world pertinent to the problem at hand. Before, considering any specific database management system for the implementation of the abstracted world, an implementation independent framework, called **data model**, is used to represent the application specific logical view of the information to be managed. Some important data models are entity-relationship data model, network data model, relational data model, functional data model, object-oriented data model. A data model is used to represent the logical view (or build a model) of the subject database in terms of entities, entity types, attributes of entities, relationships among entities, relationship attributes, and operations on entities, entity types, and relationships. This logical model must then be mapped into the given database management system's structures. Traditional data models are not capable of fully representing all aspect of an image database (ATR database in specific) because of the complexities associated with the representation of images, associated semantics, and operations. If an ATR database is only required to store images as attributes of traditional types of entities or relationships, and no operations need to be associated with the images (as is the case in the NCTI database), then any of the traditional models can be employed to model the entire ATR database. But as mentioned earlier, the existing view of the data is not sufficient to meet all the objectives of the ATR information management system. In an IAIMS, the image information that are required to be modeled can be categorized into the following classes:

1. **Iconic information** – This includes all imagery and signature data. There are two types of iconic information:

(a) Input imagery data – These are the collected or synthetic imagery data.

(b) Processed imagery data – These are images or signatures corresponding to imagery data obtained at different stages of processing.

2. **Symbolic information** – This includes all the header information, information needed to process images/signatures, and symbolic information extracted by processing imagery data. Specifically, it includes:

(a) Image/signature registration data – This includes information such as image resolution and

size, the date and time of image acquisition, atmospheric conditions, sensor types and parameters, etc.

(b) Model information – This includes symbolic descriptions of models used by synthetic data generation algorithms as well as target recognition algorithms.

(c) Extracted symbolic information – This includes symbolic information obtained by processing stored imagery/signature data by image/signal processing or ATR techniques. Specifically, it include symbolic description of detected images features, identified objects/targets and their characteristics, and information pertinent to performance of an algorithm with respect one or a set of images.

(d) Ground truth – This includes information pertinent to the identity, location, orientation, etc. associated with each of the stored images/signatures.

3. **Processing tasks** – This includes codes for all the image/signal processing and interpretation tasks needed to develop specific ATR algorithms. These tasks need to be grouped into generic categories for the ease of representation and flexibility. For example, all discontinuity detectors for a task category, which could contain several subcategories such as optimal detectors, model-based detectors, statistical, energy-based, etc. These task categories may or may not be mutually exclusive.

4. **Image/signature-to-domain transformations** – This includes transformation processes, defined in terms of a set of selected processing tasks, required to convert images/signatures to the domain dependent semantic descriptions (views). In case of an IAIMS, this includes all the ATR algorithms (defined in terms of processing tasks) that are used for target identification from imagery/signature data.

Though most of the traditional data models have been extended to handle image data, it is now widely accepted that for image information modeling an object-oriented data model with appropriate extensions would be most suitable. Research in this area is being actively pursued by both the multimedia information management and the visual information management scientific community. I am currently involved in the development of an extended object-oriented model capable of capturing all this image information as well as the required operations. I propose to continue this work and develop and implement the corresponding data definition language.

7. Image Information Retrieval

Data retrieval is a central problem in information management. In simple terms, the data retrieval problem can be stated as follows:

Given a large set of records (or objects), retrieve (or select) a subset that satisfies certain specified criteria.

A number of efficient index structures have been proposed in the literature to handle data retrieval in case of alphanumeric data. In image database systems, we often want to retrieve images whose contents satisfy certain conditions specified in an iconic query (i.e., queries that involve input images and conditions on them). In other words, an image database management system must support the retrieval of image data by content. As far as image data retrieval is concerned, most of the work has focused on devising index structures to retrieve geometric objects that lie within a specified spatial range. These indexing techniques are suitable for spatial data management, but are not suitable for similarity-based (exact or imprecise) retrieval that are often desired in an image database management system. There are three important types of similarity-based image retrieval in case of IAIMS:

1. **Shape similarity-based retrieval** - Given a definition of shape or object similarity, retrieve images having one or more objects similar (or dissimilar) to the given query object(s) or object(s) present in the input query image(s). An example is: "Find all images that contain one or more objects present in the input image."

2. **Relationship similarity-based retrieval** - Given a definition of shape and relationship (spatial, temporal, or spatio-temporal) similarity, retrieve images where certain specified objects (or query objects) that satisfy the specified relationship constraints. The relationship constraints can be specified by one or more input images or using some relationship description scheme. For example, "Find all images containing the object in image_1 to the left of the object in the image_2" or "Find all images having the same objects and same relationships among them as in the input image."

3. **Feature similarity-based** – Given a definition of image similarity based on image characteristics (e.g., histogram, signal in Fourier domain in a specified bandwidth, number of edges, number of corner, number of line segments, etc.), retrieve images that satisfy the specified constraints on image features.

To process such an image retrieval request, the query image data must be analyzed to identify its contents or features. In other words, image representation and interpretation should be components of a viable data retrieval or query processing strategy. The traditional index structures are not directly applicable for these types of image retrieval. The key issues to be handled in the design of an image information retrieval system are:

Shape and Image Representation: How can an image be described in terms of its features or how can the content of an image be described in terms of the properties of the shape of the objects and/or their spatial relationships? Of course, an important point is that these representations should be extracted automatically by processing each of the input images. Thus, we need to be concerned with how the representation can be automatically extracted from input images. The types of features used would determine the robustness and flexibility (with respect to noise, partial view, missing objects, transformations, etc.) of the overall similarity matching technique. For the first type of retrieval, an image is required to be represented as a set of shapes or regions. Each shape in turn may be represented in terms of its properties or primitive structural features. For the second type of image information retrieval, an image or an image sequence needs to be represented in terms of its regions (or objects) and their interrelationship. For the third type of retrieval an image can be represented in terms of a set of scalar and vector (one, two or multidimensional) measures.

Similarity Measure: What measures or criteria should be employed to automatically determine the similarity or dissimilarity of two shapes or the spatial relationships among objects? The similarity measure used by a system depends on the type of features or properties used to represent images, shapes and/or spatial relationships.

Index Structures: How should the image, shape, and relationship representation be organized so as enable an efficient search for similar image, shapes, and/or relationships based on some predefined similarity measure? Since a large set of images have to be searched to select a subset of models or images that satisfy certain specified conditions, the image data and associated representations must be organized in some index structures to facilitate efficient search.

Much work has to be done to develop image information retrieval techniques with desirable characteristics. One possible approach could be to develop representations and similarity measures which permit utilization of existing multidimensional point access methods (PAM), spatial access methods (SAM), or their modified version. Alternatively, retrieval systems based on new specialized index structures can be developed. These index structures should be

dynamic in nature, i.e., they must permit insertion and deletion of images and associated representations. I have been actively involved in research on the image retrieval problem and have proposed several shape similarity-based image retrieval schemes. I propose to investigate issues specific to the ATR image retrieval and develop techniques most suited to meet the ATR image information requirements.

8. Conclusions and Future Work

ATR information is multimedia in nature. Existing data management structures are not suitable to meet all the requirements with respect to the management of data collection, data distribution, algorithm development, project management, and algorithm evaluation tasks. An IAIMS capable of managing image/signature data, associated symbolic data, and image processing/ATR algorithms is essentially required. The central features of such an IAIMS are described and an a generic architecture for the IAIMS is proposed. Currently available database management systems are not suitable for the implementation of an IAIMS. It is my conclusion that the object-oriented technology needs to be employed for integrated management of image information and associated operations (algorithms). Since object-oriented technology is in a evolutionary state and there are no standards, committing to an object-oriented system at this point for a project of this magnitude is not advisable. Existing information management technologies are not capable of handling complex image representations and content-based image retrieval in flexible and user-friendly manner. An information management system capable of meeting all the requirements of the data storage and data distribution tasks management can be implemented using existing technologies. Several integrated image information management issues need further research to develop technologies required for the implementation of an IAIMS. Image information modeling and content-based image retrieval are identified as two key problems that need immediate attention. A brief description of these problems and pertinent issues is provided.

I plan to continue this work. My research and development work on this project will include the following major activities: 1. Development of an object-oriented image information modeling approach, 2. Development and implementation a data definition language and/or a GUI for the developed image information modeling approach, 3. Design and implementation of feature-based indexing mechanisms for a variety of content-based image retrieval, 4. Design and implementation of the input module and the query module for the ATR image information module, and 5. Development of a simple prototype ATR image management system.

A REVIEW OF NONFILLED INTRINSICALLY CONDUCTIVE ELASTIC MATERIALS

Douglas J. Miller, PhD
Associate Professor of Chemistry
Science & Mathematics Department

Cedarville College
PO Box 601
Cedarville, OH 45314-0601

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

A REVIEW OF NONFILLED INTRINSICALLY CONDUCTIVE ELASTIC MATERIALS

Douglas J. Miller, PhD
Associate Professor of Chemistry
Science & Mathematics Department
Cedarville College

ABSTRACT

The literature relating to nonfilled, intrinsically, electronically conductive elastic materials was reviewed. Metal or carbon filled elastomers and materials with ionic conductivity were excluded. Two major groups of conductive, elastic materials were found to exist. One group consisted of polymeric quaternary ammonium salts complexed with 7,7,8,8-tetracyanoquinodimethane. The other group was composed of graft copolymers and blends of conductive polymers with an elastomeric material.

A REVIEW OF NONFILLED INTRINSICALLY CONDUCTIVE ELASTIC MATERIALS

Douglas J. Miller, PhD

BACKGROUND

Research into the field of intrinsically conductive polymers has been intense over the last several decades, and yet the related field of intrinsically conductive elastic or elastomeric conductors has developed only fitfully. This is quite possibly due to the fact that commercially available metal or carbon filled elastomers have proven to be satisfactory for the vast majority of applications. Silicone rubbers, a popular choice for creating filled elastomers, have conductivities in the range of 10^1 to 10^4 S/cm when metal filled and in the range of 10^{-4} to 10^0 S/cm when carbon filled. Filled rubbers, however, may be undesirable due to certain of their mechanical or electrical properties or due to properties of the filler material itself. Alternatives to these filled systems should be studied for possible use in selected applications.

It should be noted here that the term "elastomer" is commonly used to refer to a wide range of substances with elastic or rubber-like properties. Definitions of rubbers or elastomers found in texts usually include a specification relating to their ability to retract to near original dimensions after substantial (100% to 1000%) elongation. This review does not restrict itself to these definitions, but instead surveys any conductive material that was described as elastomeric by its own makers. This also correctly implies that some elastic conductors, not described as elastomeric by their makers, may not directly appear in this review (i.e., several

elastic or flexible conductive plastics are not included).

INTRODUCTION

This article focuses on two groups of electronically conducting elastomeric substances, certain polymeric quaternary ammonium salts complexed with 7,7,8,8,-tetracyanoquinodimethane (TCNQ, see figure 1a) and elastomeric blends or graft copolymers which incorporate conducting polymers. Conducting elastomeric systems involving the use of conductive fillers, such as carbon black or metallic particles, will not be covered except for selected comparison purposes. Ionic conductors are also not covered. Mechanical and conductive properties are emphasized while the theory behind the conductivity is not elaborated upon. Methodologies for the preparation of the elastic conductors included in this review are closely related to those for certain nonelastic conductive polymers; therefore, some background on these latter conductive polymers is discussed.

DISCUSSION: POLYMERIC TCNQ SALTS

Polymeric quaternary ammonium salts complexed with the radical anion of tetracyanoquinodimethane ($TCNQ^{\cdot-}$, see Figure 1b) become electronically conductive. Several of these polymeric salts have been included in reviews of conductive polymers.[1-5] Ionic polymers of the type shown in Figure 2 were referred to as ionenes (for ionic polymers), and this name was subsequently extended to several types of polymeric quaternary ammonium salts.[6] The quaternary nitrogens of such polymers could be complexed with $TCNQ^{\cdot-}$ by reaction with LiTCNQ. Such materials with one $TCNQ^{\cdot-}$ molecule complexed to one quaternary nitrogen were termed simple salts and were

electronically conductive. Pure simple salts were best obtained by washing the initial product with solvent in order to extract away any extraneous neutral TCNQ which would alter the electrical properties. Neutral tetracyanoquinodimethane, TCNQ°, could be added to these simple salts to form what was called complex salts which usually possessed much higher conductivity. The ratio of TCNQ° to $TCNQ^{-}$ was normally in the range of zero to one. Maximum conductivity frequently, but not always, occurred at ratios of less than 1.0. The complex polymeric TCNQ salts had conductivities up into the range of 10^{-2}S/cm. Conductivity of these salts often followed the Arrenhius relationship of conductivity = $\text{conductivity}_o\ e^{-E/kt}$ (or $R = R_o\ e^{E/kt}$ for resistivity).

These early polymeric TCNQ salts were not labeled elastomeric. They were soluble in organic solvents and could form films that varied in consistency from brittle to flexible. They also tended to be hygroscopic and would decompose over a period of time. Most early papers on conductive ionene salts did not elaborate on their physical properties.

Articles appearing in the early 1970's described the synthesis of elastomeric ionenes synthesized starting from commercially available polymeric materials.[7-11] These elastomeric substances were complexed with TCNQ to form conductive simple and complex salts. One major category was based upon polypropylene glycols (PPG)[9,10]. Another was based upon commercial polyurethane polymers.[8,10]

Several ionenes were synthesized by reacting dihydroxypolypropylene glycol of various molecular weights with tolylene diisocyanate and chain extending with dimethylaminoethanol (DMAE) and trans 1,4-dichloro-2-butene

(TDCB) (see Figure 3).[9,10] These ionenes are identified with the initials, PPG, and then the reported molecular weight of the PPG used as starting material. The TCNQ salts are identified by adding the suffix, /TCNQ. Parameters for some simple PPG/TCNQ salts are given in Table 1. PPG150/TCNQ and and PPG450/TCNQ were reported to be brittle solids while ionenes based on higher molecular weight PPGs were reported to be elastomeric solids. Stability, as checked by UV/VIS, of the conducting ionenes listed in Table 1 was said to be good over a period of several weeks.[9] Conduction was electronic in nature and obeyed the Arrenhius relationship. Polymers PPG150/TCNQ and PPG450/TCNQ had a negative pressure coefficient of resistance, a feature common to electronically conducting materials.

Other workers, in Japan, followed a scheme similar to the one above except diiodo- or dibromopentane was used in place of TDCB, and the resulting polymer was reacted with either 4,4-bipyridine (BP) or 1,2-bis(4-pyridal)ethylene (BPE) (see Figure 4) to give crosslinked polymers.[12-15] Some properties of the TCNQ salts of these and related ionenes are included in Table 2.[12-14,16] The plain ionenes (with no added TCNQ), PPG420-BP, PPG1160-BP, PPG2020-BP, PPG2880-BP were extensively studied.[15] The approximate T_gs (by thermal methods) of these latter four ionenes were 0°C, -25°C, -45°C, and -55°C, respectively. Two mechanical relaxations were observed when the storage modulus was studied as a function of temperature for polymers PPG1160-BP and PPG2880-BP. The first was between -60 and -40°C. The second was between 20 and 70°C. The polymers gradually became fluid after the second temperature. Polymer PPG420-BP was a hard, brittle material. Polymers PPG1160-BP, PPG2020-BP, and PPG2880-BP were elastomeric

materials that broke at about 80% elongation. Polymers PPG2020-BP and PPG2880-BP were thermoplastic elastomers. The elastomeric ionenes were described as block copolymers with flexible PPG segments and hard ionene segments. The hard ionene segments were said to have a crosslinking effect which decreased as the weight fraction of the PPG fraction decreased and the mixing of the PPG fraction into the ionene portion increased. The storage modulus increased with increasing ionene content as the PPG segments became smaller. Water content of these polymers increased to a constant level after about 15 hours at room temperature and at 50% humidity. This water content was between 5.60 and 8.35% by weight for the various ionenes (higher for higher ionene content) and corresponded to between 1.8 and 2.8 water molecules per N^+ atom. The water swells the hard ionene segment and decreases the tensile modulus and tensile strength of the materials.

The temperature dependence of the storage modulus, E', and loss modulus, E", of polymers based on PPG2020-BP and PPG2880-BP was examined.[13,14] The studied polymers included the plain ionenes with no TCNQ, the simple salts, and the complex salts with a $TCNQ^0/TCNQ^-$ ratio of 0.5. Relaxations were observed for the TCNQ salts, both simple and complex, of these ionenes at around -46 and -54°C respectively. These relaxations were attributed to the polypropylene oxide segments of the structures. The TCNQ salts had a higher rubbery plateau in the E' data as compared to the plain ionene.

Polymers PPG1160-BP, PPG1290-BP, and PPG1290-BPE were examined by x-ray diffraction.[12,13] The plain ionenes and their simple TCNQ salts were amorphous. The complex salts showed definite diffraction peaks which were

assigned to the regularity in the arrangement of the TCNQ molecules. It was concluded that a microheterogeneous structure of crystalline TCNQ groupings existed along with amorphous polypropylene oxide segments. It was suggested that continuous conduction columns of $TCNQ^{-}$ and TCNQ° were responsible for the high conductivity.

The change of conductivity with time of the simple and complex TCNQ salts of PPG1160-BP and PPG1290-BPE was also studied.[12] Under ambient conditions, the conductivity slowly decreased with time over a study period of 300 days. Both the simple and complex salts had about half of their initial conductivity at the end of the study period.

PPG2880-BP/TCNQ based polymers were studied as drawn films under mechanical stress.[14] Both the simple and complex salt developed anisotropic electrical conductivity. These changes were ascribed to breaks in the crystalline conduction pathways. The anisotropic electrical conductivity of the simple salt decreased over a period of time (200 days) whereas the anisotropy persisted in the complex salt ($TCNQ°/TCNQ^{-}$ ratio of 0.5). All of the anisotropic electrical conductivities decreased with time except for that of the parallel dimension of the drawn simple salt which slowly and slightly increased over the test period.

Poly(tetramethylene oxide) (PTO) and poly(ethylene oxide) units were similarly incorporated into ionene polymers.[17,18] Only the PTO ionenes were labeled elastomeric.[17] These ionene was synthesized by reaction of commercially available isocyanate end-capped polymers (urethane polymers, designated Coronate 4080 and 4095) with DMAE and then p-bis(chloromethyl) benzene (BCMB). Simple and complex TCNQ salts were then made. (See

Table 3) The complex salts, as usual, had much higher conductivities than the simple salts. Both the simple and complex salts were flexible, but the complex salts were of higher stiffness.

Combinations of other polyurethane substances and TCNQ were also investigated.[7,8,10] Solithane 1,13 urethane resin was reacted with bromopropanol and then tetramethylamino poly(propylene glycol) to give a crosslinked elastomer (see Figure 5). The product, labeled PPGS, had a tensile strength of 954 psi, an elongation at break of 270%, and T_g (thermal method) of -20°C. This product was reacted with LiTCNQ to give a rubbery polymer containing about 5% by weight of bound TCNQ. Two specimens of differing thickness, PPGS 1 (0.5cm thick) and PPGS 3 (1.17cm thick), were studied. Also, a Goodrich polyurethane (tradename Estane, 5710F) was mixed with LiTCNQ to yield films with 5% LiTCNQ content by weight. Comparison of the above conducting TCNQ formulations to carbon filled polymers was made. It was reported that the conductivity of all of the polymer TCNQ complexes with 5% by weight of TCNQ was on the order of 10^{-8} S/cm. Stretching of these polymers between 0 and 150% resulted in essentially unchanged or somewhat increased conductivity. This behavior contrasted to that of carbon filled polymers which, as of 1970, were reported to need at least 15% by weight carbon black for reasonable conductivity changes and for which conductivity decreases upon elongation.[19]

Solithane can also be reacted with 1,3-bis(dimethylamino)2-propanol and then a dihalide such as dibromopropane, dibromopolybutadiene, or dibromohexane to give elastomeric materials.[7,10] The dibromohexane derivative was reacted with 10% by weight LiTCNQ to give a conductive

film.[10] Torsion measurement gave a T_g of 38°C for this Solithane-TCNQ complex which was only 8°C higher than for pure Solithane. Conductivity/temperature measurements yielded a T_g of 38°C, below which E_a was 0.5 eV and above which E_a was 1.2 eV. Conductivity was on the order of 10^{-9} S/cm near room temperature. Extensive dielectric measurements were made on this material.[10]

DISCUSSION: GRAFT COPOLYMERS AND BLENDS

The other major approach for making a conductive elastomer entails combining a conductive plastic with an elastomeric material. This is in actually an extension of attempts to produce environmentally stable blends of intrinsically conductive polymers. The vast majority of these attempts involved the blending of a conductive polymer with a fairly plastic matrix material, but a few systems did involve elastomeric matrices. One major approach to making such conductive combinations utilized an electrochemical polymerization process. Another major approach involved the incorporation of conductive polyacetylene (PA) into elastomeric matrices.

Recent work employed the electrochemical polymerization of aniline or 3-methyl-thiophene on a nitrilic rubber-coated platinum working electrode.[20-23] The insulating nitrilic rubber [variously referred to as a copolymer of butadiene and acrylonitrile, poly(butadiene-co-acrylonitrile), or NBR] was coated on the electrode and then swollen by a solvent. Dissolved monomer is then subsequently electrochemically polymerized into the nitrilic rubber. This process used with aniline produced a graft copolymer of polyaniline (PAn) and nitrilic rubber (NBR).[20,21] NBR rubber with 75% butadiene content and 29% acrylonitrile content produced the most

homogeneous product, a light green rubbery material. Infrared spectra indicated essentially a superposition of spectra for PAn and NBR with NBR spectral bands predominating. There was, however, a reduction of intensity of the band at 1640cm^{-1} assignable to the NBR carbon-carbon double bonds. This fact and TGA/DSC studies indicated that a graft copolymer of PAn and NBR had been formed. The graft copolymer was insoluble in solvents that would normally dissolve the NBR component. The electroactivity, as studied by cyclic voltammetry, was very similar to that of pure PAn. This electroactivity was reportedly retained even after several months of atmospheric exposure. No other mechanical or electrical properties were published.

A composite, not a graft copolymer, of poly(3-methyl thiophene), P3MeT, with nitrilic rubber was made using the same electrochemical methodology.[22,23] Dark, self-supported, flexible, homogeneous films could be produced under the appropriate conditions. The infrared spectrum of the P3MeT/NBR composite was essentially a superposition of spectra of the two components with those bands from NBR predominating. TGA/DSC work supported the conclusion that the material was a composite, not a graft copolymer. The T_g of the NBR component of the composite varied only slightly from that of pure NBR. This was said to indicate the low miscibility of the NBR with the P3MeT component. The NBR component could be removed by solvent extraction. The electroactivity, as studied by cyclic voltammetry, was very similar to that of pure P3MeT. The small composite films had a conductivity of 0.2 S/cm on the face originally adhering to the electrode surface while the opposite side of the film had no measurable conductivity. (Pressed

pellets of pure P3MeT had a conductivity of 1.2 S/cm.) Stress-strain testing revealed that the composite was a hard, tough material as compared to pure NBR which was a soft elastomeric substance. The composite had an average stress at break of 0.43 MPa and a strain at break of 74%. Pure NBR gave corresponding values of 0.18 MPa and 580%. The hardening was ascribed to the P3MeT component acting as a reinforcing fiber. Aging studies of the conductive or mechanical properties were not reported.

The disadvantages of the electrochemical method are the dimensional limitations imposed by the methodology and electrode shape and also possible anisotropy in the electrical conductivity of the film produced. These disadvantages lead the same group involved in the electrochemical studies to investigate a mechanical process for the creation of elastomeric blends.[24] Blends of polypyrrole and ethylene-propylene-diene terpolymer (EPDM rubber) were made by the absorption of pyrrole vapors into the EPDM matrix containing oxidant particles of $CuCl_2$. Copper chloride particles of various sizes were incorporated at different concentrations into the EPDM rubber by the process of calendering (rolling between heated rolls under high pressure). The most conductive blend contained 50phr (parts per hundred of the pure resin) of $CuCl_2$ particles smaller than 106 micrometers in diameter. This blend had 21.6% polypyrrole by weight, had a Young's Modulus of 3.9 MPa, and had a conductivity of 2×10^{-7} S/cm. IR, TGA, and DSC evidence indicated that the components do not interact and are immiscible. Both Young's Modulus and the conductivity for blends produced with 50phr of $CuCl_2$ increased when oxidant particle size decreased. An advantage of this method is the possibility of creating larger amounts of conductive, elastomeric

material. A potential disadvantage is the utilization of metallic compounds.

A different group of researchers investigated the formation of polyacetylene (PA) within either an elastomer or a thermoplastic elastomer.[25-32] This work was preceded by the efforts of Wnek, who through a process of in situ polymerization, succeeding in incorporating polyacetylene into low density polyethylene films.[33-36] This technique was applied to various elastomeric materials, including cis-1,4-polybutadiene,[25,26] cis-1,4-polyisoprene,[26] EPDM rubber,[26-28,32] and several thermoplastic elastomers.[29-32] The conductive materials are made by exposing a catalyst impregnated film to acetylene gas or by bubbling acetylene gas through a solution containing elastomer and catalyst and then casting a film. All of the resulting materials tended to have several traits in common. Higher PA content in any blend tended to increase its stiffness. All materials became conductive when doped, usually with I_2. This conductivity decayed with time, but sometimes slower than it did for pure PA. The PA was present in these blends as polycrystalline material within an amorphous matrix. Stretching the materials sometimes improved the conductivity of the materials due to an increase in the order of the PA phase, probably an increase in crystallinity.

The first reported study focused on the synthesis of a conductive blend of PA and cis-1,4-polybutadiene (PB).[25,26] A film of PB impregnated with a Ziegler-Natta type catalyst was exposed to acetylene gas, thus forming a PA/PB blend. This blend was then doped usually by solution techniques and usually with iodine. X-ray and infrared data supported the concept that

polycrystalline PA existed within amorphous PB. The PA in the blend, depending on the blend's PA content, formed continuous or discontinuous phases. The changeover occurred at about 40% PA content. The T_g, as measured by DSC, of the PB component in the blends was not significantly different from that of pure PB. This was interpreted as evidence of a lack of chemical reaction between the components of the blends. Increasing the PA content of the blend increased the tensile strength and decreased the maximum extension ratio (see Table 4). Blends with less than 30% PA content were elastomeric materials with mechanical stability of several weeks, whereas blends with higher PA content were more plastic-like. Pure PA turns brittle within a few hours after air exposure. Doped blends usually exhibited electrical conductivities in the range of 10 to 100 S/cm (see Table 5). Stretching before doping of blends containing 40 to 60% PA resulted in higher conductivities, up to 575 S/cm (see Table 5), probably due to increased crystallinity in the PA phase. The stability of the electrical conductivities reportedly did not differ substantially from pure PA. The conductivity of a 60/40 PA/PB blend dropped from 25 S/cm to 10^{-1} S/cm after 4 months of exposure to ambient conditions. PB and similar elastomers were concluded to be poor hosts for PA due to their poor ability to inhibit the diffusion of oxygen and moisture and due to the propensity of their plentiful double bonds for further reaction.

Polyacetylene was similarly introduced into a matrix film of an ethylene-propylene-diene terpolymer (EPDM rubber) in which ethylidene norbornene was the diene unit.[26-28,32] The EPDM rubber had low levels of unsaturation, less than 2%. Low levels of unsaturation were thought to

provide sufficient reactivity for crosslinking while limiting harmful further reactivity. Copolymers of isobutylene and isoprene (butyl rubber) or of isobutylene and butadiene were also suggested as possible blending candidates with low levels of unsaturation. The PA/EPDM blends were flexible, tough, homogeneous films whose elasticities were dependent upon the PA content. The IR spectrum of the blends had absorption bands assignable to either the PA or EPDM component. The PA was present primarily in the trans-configuration as evidenced by the infrared band at $1015cm^{-1}$. Electron micrographs of a PA/EPDM blend (5% PA by weight), before and after doping, gave evidence of a discontinuous PA phase. Essentially all of the EPDM rubber could be removed at this stage by solvent extraction. Blends doped with I_2 vapor (24 hour exposure) exhibited conductivities in the range of 10-90 S/cm. Conductivity depended on the PA content. Blends, apparently while being stretched, had higher levels of conductivity (in one case, changing from 10 S/cm to 700 S/cm at 600% elongation) due to increased crystallinity. Gamma radiation was normally used to crosslink the blends since a chemical crosslinking method utilizing S_2Cl_2 caused difficulties. IR data indicated that only the double bonds of the EPDM rubber were involved in the crosslinking process. Crosslinked PA/EPDM blends had conductivities as high as 100 S/cm and were insoluble. The conductive stability of the PA/EPDM blends, both before and after crosslinking, was better than that reported for pure PA. The initial conductivity of pure PA dropped to below 1% of its initial value after 35 days of air exposure. Conductivities of the EPDM/PA blends fell, during the 35 day test period, to between 1 and 10% of their initial values. The enhanced conductive

stability of the PA/EPDM blends was thought to be due to the protective effect of the EPDM component. The crosslinked blend consistently had higher conductivity values than the uncrosslinked blend at any point during this test period. Detailed mechanical properties, particularly for the crosslinked films, were not reported.

In situ polymerization could also place PA into the central rubbery portions of various triblock thermoplastic elastomers.[29-32] Kraton 1107 (styrene-isoprene-styrene triblock copolymer), Kraton 1101 (styrene-butadiene-styrene), and Kraton 4609 (styrene-ethylene butylene-styrene) from Shell Chemical Company were used as hosts for PA. One advantage that these commercial thermoplastic elastomers possessed was that the glassy polystyrene blocks acted as physical crosslinks thus precluding the necessity for post-curing processes. The PA phase forms within the rubbery central blocks and not within the glassy polystyrene blocks. The PA was present primarily in the trans-form and was highly crystalline. The PA/triblock blends shared several of the characteristics of the PA/EPDM blends. The blends with higher PA content were, as expected, stiffer than those with lower PA content. Most of the reported data was for blends with the styrene-butadiene-styrene (SBS) copolymer. A PA/SBS blend with 4% PA content had a stress-strain curve similar to the pure SBS elastomer with an elongation-to-break value of around 1100%. A PA/SBS blend with 20% PA was much stiffer with an elongation-to-break value of around 200%. Doping with iodine vapor or with ferric chloride in solution made the blends conductive, but also reduced the elongation-to-break values. One particular blend, probably a PA/SBS blend with 4% PA content, had an elongation-to-break value

before doping of 1127%. Doping with iodine for 15 minutes reduced this value to 335% while the conductivity was now 0.95 S/cm. Continued doping for a total of 180 minutes resulted in an elongation-to-break value of 167% and a conductivity of 3.00 S/cm (see Table 6). A reported range of conductivities values for PA/triblock blends was 60 to 100 S/cm. Decay of electrical conductivity in the blends was slower than in pure PA. One PA/SBS blend had, after 35 days of air exposure, between 1 and 10% of its original conductivity.

REFERENCES

(1) Rembaum, A. In *Encyclopedia of Polymer Science and Technology*, First ed.; Mark, H. F.; Gaylord, N. G.; Bikales, N. M., Eds.; Interscience Publishers, a division of John Wiley & Sons, Inc.: New York, 1969; Vol. 11, Plastics, Resins, Rubbers, Fibers, pp. 318-337.

(2) Rembaum, A. *J. Polym. Sci., Part C* **1970**, *29*, 157.

(3) Eisenberg, A.; King, M. *ion-containing polymers*, First ed.; Polymer Physics; Academic Press: New York, 1977; Vol. 2.

(4) Block, H. In *Advances in Polymer Science*; Cantow, H. J. et al., Eds.; Springer-Verlag: New York, 1979; Vol. 33, pp. 93-167.

(5) Duke, C. B.; Gibson, H. W. In *Encyclopedia of Chemical Technology*, 3rd ed.; Grayson, M.; Eckroth, D.; Mark, H. F.; Othmer, D. F.; Overberger, C. G.; Seaborg, G. T., Eds.; John Wiley & Sons, Inc.: New York, 1982; Vol. 18, pp. 755-793.

(6) Rembaum, A.; Baumgartner, W.; Eisenberg, A. *J. Polym. Sci., Part B* **1968**, *6*, 159-171; Polymer Letters.

(7) Rembaum, A.; Yen, S. P. S.; Landel, R. F.; Shen, M. *J. Macrom. Sci. Chem.* **1970**, *A4*(3), 715-738.

(8) Hermann, A. M.; Yen, S. P. S.; Rembaum, A.; Landel, R. F. *J. Polym. Sci., Part B* **1971**, *9*(8), 627-633.

(9) Somoano, R.; Yen, S. P. S.; Rembaum, A. *J. Polym. Sci., Part B* **1970**, *8*(7), 467-479.

(10) Hermann, A. M. In *Elec. Prop. Polym*; Frisch, K. C., Ed.; Technomic: Westport, Conn., 1972; pp. 120-138.

(11) Rembaum, A. *J. Elastoplastics* **1972**, *4*, 280-293.

(12) Watanabe, M.; Toneaki, N.; Shinohara, I. *Polym. J.* **1982**, *14*(3), 189-195.

(13) Watanabe, M.; Toneaki, N.; Takizawa, Y.; Shinohara, I. *J. Polym. Sci., Polym. Chem. Ed.* **1982**, *20*, 2669-2680.

(14) Watanabe, M.; Takizawa, Y.; Shinohara, I. *J. Polym. Sci., Polym. Chem. Ed.* **1983**, *21*(8), 2397-2404.

(15) Watanabe, M.; Takizawa, Y.; Shinohara, I. *Polymer* **1983**, *24*(4), 491-497.

(16) Watanabe, M.; Tsuchikura, A.; Kamiya, T.; Shinohara, I. *J. Polym. Sci., Polym. Lett. Ed.* **1981**, *19*(7), 331-334.

(17) Ikeno, S.; Yokoyama, M.; Mikawa, H. *J. Polym. Sci., Polym. Phys. Ed.* **1978**, *16*(4), 717-723.

(18) Takizawa, Y.; Aiga, H.; Watanabe, M.; Shinohara, I. *J. Polym. Sci., Polym. Chem. Ed.* **1983**, *21*(11), 3145-3153.

(19) Norman, R. H. *Conductive Rubbers and Plastics*; Elsevier Publishing Company Limited: New York, 1970.

(20) Tassi, E. L.; De Paoli, M. A. *J. Chem. Soc., Chem. Commun.* **1990**(2), 155-156.

(21) Tassi, E. L.; De Paoli, M. A. *Polymer* **1992**, *33*(11), 2427-2430.

(22) Zoppi, R. A.; De Paoli, M. A. *J. Electroanal. Chem.* **1990**, *290*, 275-282.

(23) Zoppi, R. A.; De Paoli, M. A. *Polymer* **1992**, *33*(21), 4611-4616.

(24) Zoppi, R. A.; Felisberti, I.; De Paoli, M. A. *J. Polym. Sci., Part A: Polym. Chem.* **1994**, *32*(6), 1001-1008.

(25) Rubner, M.; Tripathy, S.; Georger, J., Jr.; Cholewa, P. *Macromol.* **1983**, *16*, 870-875.

(26) Rubner, M.; Tripathy, S.; Lee, K.; Georger, J., Jr.; Jopson, H.; Cholewa, P. *Org. Coat. Appl. Polym. Sci. Proc.* **1983**, *48*, 544-547.

(27) Lee, K. I.; Jopson, H. *Makromol. Chem., Rapid Commun.* **1983**, *4*, 375-378.

(28) Lee, K. I.; Jopson, H. U.S. Patent US 4510075, 1985.

(29) Lee, K. I.; Jopson, H.; Tripathy, S.; Rubner, M. *Org. Coat. Appl. Polym. Sci. Proc.* **1983**, *48*, 548-551.

(30) Lee, K. I.; Jopson, H. *Polym. Bull. (Berlin)* **1983**, *10*(1-2), 105-108.

(31) Lee, K. I.; Jopson, H. U.S. Patent US 4510076, 1985.

(32) Lee, K. I.; Jopson, H. *ACS Symp. Ser.* **1984**, *242*(Polym. Electron.), 497-506.

(33) Galvin, M. E.; Wnek, G. E. *Polymer* **1982**, *23*, 795-797.

(34) Wnek, G. E. U.S. Patent US 4394304, 1983.

(35) Galvin, M. E.; Wnek, G. E. *J. Phys., Colloq.* **1983**, *C3*, 151-153; Conf. Int. Phys. Chim. Polym. Conduct., 1982.

(36) Wnek, G. E. In *Handbook of Conducting Polymers*; Skotheim, T. A., Ed.; Marcel Dekker, Inc.: New York, 1986; Vol. 1, pp. 205-212.

The tables and figures referred to in this version of the final report sent to the Air Force Office of Scientific Research have been omitted due to the page limit constraints. These tables and figures may be obtained by contacting:

WL/MLBP
2941 P STREET -- SUITE 1
WPAFB, OH 45433-7750

THERMOPHYSICAL AFFINE INVARIANTS FROM IR IMAGERY FOR OBJECT RECOGNITION

N. Nandhakumar
Assistant Professor
Department of Electrical Engineering,

University of Virginia
SEAS, Thornton Hall
Charlottesville, VA 22903

Final Report For:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

THERMOPHYSICAL AFFINE INVARIANTS FROM IR IMAGERY FOR OBJECT RECOGNITION

N. Nandhakumar
Assistant Professor
Department of Electrical Engineering,
University of Virginia

Abstract

An important issue in developing a Model-Based Vision approach is the specification of features that are - (a) invariant to viewing and scene conditions, and also - (b) specific, i.e., the feature must have different values for different classes of objects. We formulate a new approach for establishing invariant features. Our approach is unique in the field since it considers not just surface reflection and surface geometry in the specification of invariant features, but it also takes into account internal object composition and state which affect images sensed in the non-visible spectrum. A new type of invariance called Thermophysical Invariance is defined. Features are defined such that they are functions of only the thermophysical properties of the imaged objects. The approach is based on a physics-based model that is derived from the principle of the conservation of energy applied at the surface of the imaged object.

THERMOPHYSICAL AFFINE INVARIANTS FROM IR IMAGERY FOR OBJECT RECOGNITION

N. Nandhakumar

1. Introduction

Object recognition requires robust and stable features that are separable in feature space. An important characteristic of these features is that they be invariant to scene conditions, such as illumination, and changes in viewpoint/object pose. The formulation of invariant features, and the quantitative analysis of feature variance is currently being addressed by a number of researchers in the computer vision community, and has led to the establishment of a currenty incomplete but growing theory of feature invariance. Such efforts have primarily considered reflected-light imagery – formed by sensing visible wavelength energy. This investigation has resulted in a number of distinct (yet related) approaches that may be loosely grouped into three categories: (1) Geometric Invariance (GI), (2) Quasi-Invariance (QI), and (3) Intensity-based Invariance (II).

Geometric invariants have been investigated since the inception of the field of image analysis in the early 1960s (actually, such investigation can be traced back to the onset of photogrammetry in the 19th century). The main issue here is the investigation of features that are invariant with respect to changes in viewpoint. Geometric invariants come in two basic "flavors", algebraic and differential. Algebraic invariants are based on the global configuration of features extracted from an object, and involve the notion of algebraic shapes, e.g., shapes are analytically expressed as 2D conics, and invariant relationships are identified between conics belonging to an object. Differential invariants are polynomial expressions involving the local curvature properties of 2D and 3D curves and are computed for each point on the curve. Thus differential invariants are actually parameter space "signatures" (e.g., a locus of point in an abstract parameter space) that are unique to the object, so that differences between two parameter space signatures define different objects. A close relationship (an equivalence in some cases) has been shown between some formulations of differential invariants and algebraic invariants [Forsyth et al., 1991]. There are several examples of geometric (actually algebraic) invariants of planar configurations under projective transformation, such as the cross ratio using 4 collinear points, 5 coplanar points with no three being collinear, a conic and two non-tangent lines, and a pair of conics.

Quasi-Invariance (QI) can be thought of as a relaxation of the central notion of GI [Binford et al., 1989, Zerroug and Nevatia, 1993]. A Quasi-Invariant is a property of a geometric configuration that is almost invariant under a class of imaging transformations. Formally, a geometric configuration is a QI if the linear term(s) in the Taylor series expansion of the configuration with respect to the parameters of the imaging transformation vanish. This has the effect of making the QI measure nearly constant over a large region of the viewing hemisphere, and rapidly diverging only as the angle between the view direction and the surface normal approaches large values. This behavior lends itself to probabilistic modeling and the use of reasoning schemes such as Bayesian evidence accrual. A detailed study of the variation of "invariant" features with viewpoint has been undertaken [Burns et al., 1993].

Intensity-based Invariants (II's) are functions of image intensities that yield values which are invariant to scene illumination and viewpoint. To date, some investigation of II's has been conducted for visible imagery, but practically none has been reported for non-visible imagery. Examples of II's in computer vision include color features for object recognition [Klinker et al., 1988, Healey, 1991, Healey and Slater, 1994] and polarization cues for material identification [Wolf, 1990]. A more direct example of this approach computes ratios of albedos of homogeneous image intensity patches within objects in visible imagery [Nayar and Bolle, 1993].

Non-visible modalities of sensing have been shown to greatly increase the amount of information that can be used for object recognition. A very popular and increasingly affordable sensor modality is thermal imaging - where non-visible radiation is sensed in the long-wave infrared (LWIR) spectrum of $8\mu m$ to $14\mu m$. The current generation of LWIR sensors produce images of contrast and resolution that compare favorably with broadcast television quality visible light imagery. However, the images are no longer functions of only surface reflectance. As the wavelength of the sensor transducer passband increases, emissive effects begin to emerge as the dominant mode of electromagnetic energy exitance from object surfaces. The (primarily) emitted radiosity of LWIR energy has a strong dependence on internal composition, properties, and state of the object such as specific heat, density, volume, heat generation rate of internal sources, etc. This dependence may be exploited by specifying image-derived invariants that vary only if these parameters of the physical properties vary.

In this paper we describe the use of the principle of conservation of energy at the surface of

the imaged object to specify a functional relationship between the object's thermophysical properties (e.g., thermal conductivity, thermal capacitance, emissivity, etc.), scene parameters (e.g., wind temperature, wind speed, solar insolation), and the sensed LWIR image gray level. We use this functional form to derive invariant features that remain constant despite changes in scene parameters/driving conditions. In this formulation the internal thermophysical properties play a role that is analogous to the role of parameters of the conics, lines and/or points that are used for specifying geometric invariants when analyzing visible wavelength imagery. Thus, in addition to the currently available techniques of formulating features that depend only on external shape and surface reflectance discontinuities, the phenomenology of LWIR image generation can be used to establish new features that "uncover" the composition and thermal state of the object, and which do not depend on surface reflectance characteristics.

The derivation of thermophysical invariants (TI's) from non-visible wavelength imagery, the evaluation of the performance of these invariants, and their use in object recognition systems poses several advantages. The main advantage of this approach is the potential availability of a number of new (functionally independent) invariants that depend on internal compositional properties of the imaged objects. Note that it is possible to evaluate the behavior of thermophysical invariants using ground truth data consisting of images of objects of known composition and internal state. This additional information can be used to augment/complement the behavior of GI's. One way in which GI's can be integrated with TI's for object recognition is as follows: (1) Parametric curves and/or lines are extracted from an LWIR image. (2) The curves are used to compute GI's which are in turn used to hypothesize object identity and pose, and (3) TI's are computed for this hypothesis and compared to a stored model library for verification. Some details of this approach are presented later.

The ideas presented in this paper are continuations/extensions of previous and ongoing research in thermophysical model-based interpretation of LWIR imagery. A brief description of this thermophysical approach is presented in section 2, followed by the formulation of a method to derive themophysical invariants in section 3. Preliminary experimental results of applying this new approach to real imagery are presented in section 4, which is followed by a disussion of the behavior of the new method, issues to be considered in using this method for object recognition, and issues that remain to be explored.

2. A Thermophysical Approach to IR Image Analysis

An intuitive approach to thermo-physical interpretation of LWIR imagery is given in [Gauder, et al, 1993]. This approach rests upon the following observation, termed the "Thermal History Consistentcy Constraint" and analogous to Lowe's well known Viewpoint Consistentcy Constraint [Lowe, 1987]: "The temperature of all target features for a passive target must be consistent with the heat flux transfer resulting from exposure to the same thermal history." In [Gauder, et al, 1993] this constraint is exploited by analyizing objects to locate components that are similar in terms of thermo-physical properties and then examining a temporal sequence of calibrated LWIR data to experimentally assess the degree to which such thermo-physically similar components exhibit similar temperature state temporal behavior. Such analysis was shown to lead to formulation of simple intensity ratio features exhibiting a strong degree of temporal stability that could be exploited provided: (1) thermally homogenous regions in the LWIR image corresponding to the thermo-physically similar object components could be reliably segmented, and (2) a target-specific geometric refrence frame is available in order to correctly associate extracted regions with candidate object components.

To avoid the difficulties inherent in assumptions (1) and (2) above an alternative technique applicable to overall object signatures was suggested in [Gauder, et al, 1993]. An analysis of typical LWIR 'lumped parameter' object temperature modeling approaches suggests that over small time scales object temperature can be crudely modeled by a small dimensional linear system with algebraically seperable spatial and temporal components. Ratios of spatial integrals of temperature with a simple set of orthronormal 2D polynomials (obtained from applying Gramm-Schmidt to $1, x, y, xy, x^2y, xy^2, x^3$ and y^3) were tried. Some of the resulting functions were nearly constant with time when measured against 24 hours of LWIR imagery of a complex object (a tank) but no experimentation was done with multiple objects to examine between and within-class seperation, so little can be drawn in the way of a substantive conclusion with respect to utility as an object identification technique.

A physics-based approach that attempts to establish invariant features which depend only on thermophysical object properties is reported in [Nandhakumar and Aggarwal, 1988, Nandhakumar, 1990]. A thermophysical model is formulated to allow integrated analysis of thermal and visual imagery of outdoor scenes. An overview of this approach, it's advantages and limitations is

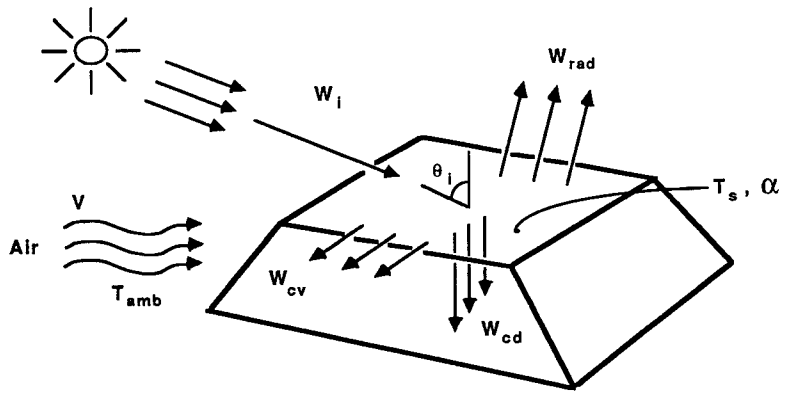


Figure 1: Energy exchange at the surface of the imaged object. Incident energy is primarily in the visible spectrum. Surfaces loses energy by convection to air, via radiation to the amosphere, and via conduction to the interior of the object.

presented below, and an improvement to this thermophysical approach using algebraic invariance theory is described in section 3.

At the surface of the imaged object (figure 1) energy absorbed by the surface equals the energy lost to the environment.

$$W_{abs} = W_{lost} \tag{1}$$

Energy absorbed by the surface is is given by

$$W_{abs} = W_I \, cos\theta_I \, \alpha_s \, , \tag{2}$$

where, W_I is the incident solar irradiation on a horizontal surface and is given by available empirical models (based on time, date and latitude of the scene) or by measurement with a pyranometer, θ_i is the angle between the direction of irradiation and the surface normal, and α_s is the surface absorptivity which is related to the visual reflectance ρ_s by $\alpha_s = 1 - \rho_s$. Note that it is reasonable to use the visual reflectance to estimate the energy absorbed by the surface since approximately 90% of the energy in solar irradiation lies in the visible wavelengths [Incropera and DeWitt, 1981].

A simplified shape-from-shading approach was used to compute $cos\theta_I$ and α_s from the visual image. The surface temperature was estimated from the thermal image based on an appropriate model of radiation energy exchange between the surface and the infrared camera.

The energy lost by the surface to the environment was given by

$$W_{lost} = W_{cd} + W_{cv} + W_{rad}, \tag{3}$$

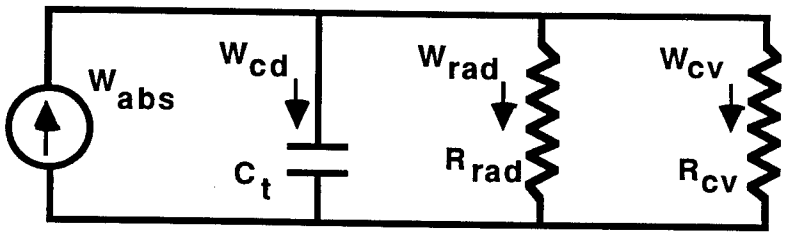


Figure 2: Equivalent thermal circuit of the imaged object. Here, internal heat sources are assumed to be non-existent. The conducted energy is used to charge/discharge the lumped thermal capacitance of the object.

where, W_{cv} denotes the heat convected from the surface to the air which has temperature T_{amb} and velocity V, W_{rad} is the heat lost by the surface to the environment via radiation and W_{cd} denotes the heat conducted from the surface into the interior of the object. The radiation heat loss is computed from:

$$W_{rad} = \epsilon \sigma (T_s^4 - T_{amb}^4), \tag{4}$$

where, σ denotes the Stefan-Boltzman constant, T_s is the surface temperature of the imaged object, and T_{amb} is the ambient temperature.

The convected heat transfer is given by

$$W_{cv} = h(T_s - T_{amb}) \tag{5}$$

where, h is the average convected heat transfer coefficient for the imaged surface, which depends on the wind speed, thermophysical properties of the air, and surface geometry [Incropera and DeWitt, 1981].

Considering a unit area on the surface of the imaged object, the equivalent thermal circuit for the surface is shown in figure 2. C_T is the lumped thermal capacitance of the object and is given by

$$C_T = DVc$$

where, D is the density of the object, V is the volume, and c is the specific heat. The resistances are given by:

$$R_{cv} = \frac{1}{h} \quad \text{and} \quad R_{rad} = \frac{1}{\epsilon\sigma(T_s^2 + T_{amb}^2)(T_s + T_{amb})}$$

It is important to note that implicit in the above formulation is that a low Biot number was assumed, viz., the surface was considered to be a thin plate.

2.1. Analyzing A Single Data Set

It is clear from figure 2 that the conduction heat flux W_{cd} depends on the lumped thermal capacitance C_T of the object. A relatively high value for C_T implies that the object is able to sink or source relatively large amounts of heat. An estimate of W_{cd}, therefore, provides us with a relative estimate of the thermal capacitance of the object, albeit a very approximate one. W_{cd} is useful in estimating the object's ability to sink/source heat radiation. a feature shown to be useful in discriminating between different classes of objects.

In order to minimize the feature's dependence on differences in absorbed heat flux, a normalized feature was defined to be the ratio $R = W_{cd}/W_{abs}$. The values are lowest for vehicles, highest for vegetation and in between for buildings and pavements. Classification of objects using this property value is discussed in [Nandhakumar and Aggarwal, 1988b].

2.2. Analyzing A Temporal Data Sequence

The availability of temporal sequences of registered thermal and visual images makes possible a direct estimation of C_T and hence a more reliable estimate of the imaged object's relative ability to sink/source heat radiation. Since

$$W_{cd} = C_T \frac{dT_s}{dt} \tag{6}$$

a finite (backward) difference approximation to this equation may be used for estimating C_T as

$$C_T = W_{cd} \frac{(t_2 - t_1)}{(T_s(t_2) - T_s(t_1))} \tag{7}$$

where, t_1 and t_2 are the time instants at which the data were acquired, $T_s(t_1)$ and $T_s(t_2)$ are the corresponding surface temperatures, and W_{cd} is the conducted heat flux which is assumed to be constant during the time interval.

2.3. Limitations of The Previous Approach

The above approach is powerful in that it makes available features that are completely defined by internal object properties, and hence may be considered to be deterministic. Hence, feature values may be compared with accurate *ground truth* values computed from known physical properties of test objects – one of the major advantages of using physics-based/phenomenological models as compared to statistical models.

There are several factors that limit the performance of the above approach. These are listed

below and a method to minimize these limitations is addressed in the next section.

Systemic Errors in Sensor Fusion.
The thermal and visual image pairs may not be perfectly registered. Also, segmentation errors typically cause a large portion of an object to be included with small portions of a different object in one region. These errors give rise to meaningless values of the surface energy estimates at/near the region boundaries. However, stable estimates are available in the interior of each of the regions. Also, shape from shading techniques perform poorly unless images are relatively noise-free and have high resolution [Zhang et al., 1994].

These errors result in a significant number of inaccurate estimates of the surface energy exchange components. The histogram of values of the ratio R tends to be heavy-tailed and skewed. A statistically robust scheme for computing the thermophysical feature R was proposed to minimize this drawback [Nandhakumar, 1994]. However, the computational complexity for such a technique is very high, and the following drawbacks below were not adequately overcome.

Unknown Surface/Scene Parameters.
Direct application of the energy exchange model as described above requires *a priori* knowledge of several surface and scene parameters such as emissivity, wind speed, solar insolation, etc. Values for these were either assumed, hypothesized, or empirically obtained. It would be beneficial to formulate an approach where knowledge of these parameters is not imperative.

Variation in R.
The value of the thermophysical feature, R, was found to be only weakly invariant. While separation between classes was preserved, the range of values of R for each class was observed to vary with time of day and season of year. Also, the feature R was able to only separate very broad categories of objects, such as automobiles, buildings, and vegetation - but lacked the specificity to differentiate between different models of vehicles.

An improved formulation for establishing thermophysical features is described in section 3, wherein the feature is constrained to be invariant to affine transformation of the driving (scene) conditions.

3. Formulation of Thermophysical Affine Invariants (TPI's)

An improved approach for computing invariant features that depend on the internal, ther-

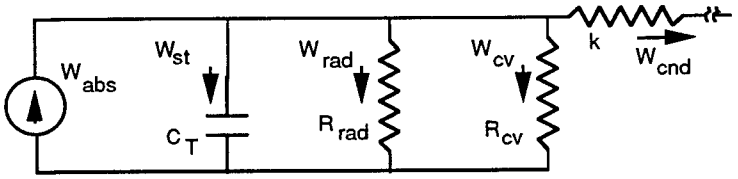


Figure 3: The equivalent thermal circuit for the extended model that separates the stored energy component and the conduction component to the interior of the object.

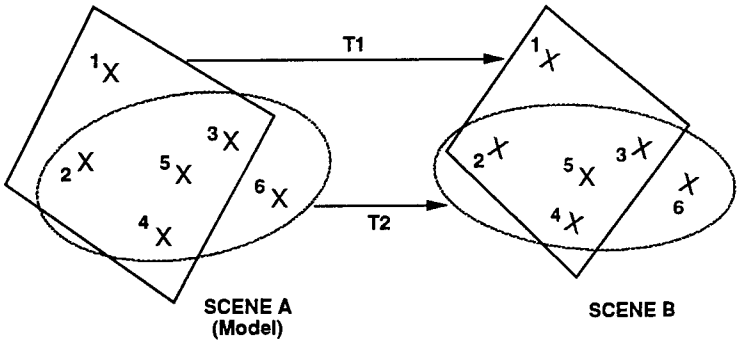


Figure 4: We need a collection of five point in 5D measurement space to compute the affine transformation, T_i, between two scenes. Two different sets of 5 points each can be used to define an absolute invariant provided $\det(T_1) = \det(T_2)$.

mophysical properties of the imaged object, and that are invariant to driving (scene and surface) parameters is based on a reformulation of the model outlined in section 2. First, to account for reasonable values of Biot numbers, viz. for objects other than thin plates, the conducted energy component W_{cd} is decomposed into two terms - W_{st} which raises the stored, internal energy of an elemental volume at the surface, and W_{cnd} which is conducted to the interior of the object. Note that W_{st} is given by eqn (6), while $W_{cnd} = k\ dT/dx$, where k is the thermal conductivity of the material, and x is distance below the surface. Here, we assume that lateral energy conduction is insignificant compared to conduction along the direction normal to the surface. Figure 3 shows the equivalent thermal circuit for this extended model.

The energy balance equation, $W_{abs} = W_{rad} + W_{cv} + W_{st} + W_{cnd}$ may be rewritten in the following linear form:

$$a^1 x_1 + a^2 x_2 + a^3 x_3 + a^4 x_4 + a^5 x_5 = 0. \tag{8}$$

Using the expressions for the various energy components as presented in section 2 we can express each term in the above expression as:

$$\begin{aligned} a^1 &= cos\theta_I & x_1 &= W_I \alpha_s \\ a^2 &= -\sigma\ (T_s^4 - T_{amb}^4) & x_2 &= \epsilon \\ a^3 &= -(T_s - T_{amb}) & x_3 &= h \\ a^4 &= C_T & x_4 &= -\frac{dT_s}{dt} \\ a^5 &= kT_s & x_5 &= 1 - \frac{T_{int}}{T_s} \end{aligned} \tag{9}$$

The term $a^5 x_5$ denotes W_{cnd} expressed in a finite difference form.

Note that a calibrated LWIR image provides radiometric temperature. Hence a^2 and a^3 can be computed from the LWIR image alone (and knowledge of the ambient temperature), while a^1 and a^4 are known when the identity and pose of the object is hypothesized, and a^5 is computed using the image and the hypothesis. The "driving conditions", or unknown scene parameters[1] that change from scene to scene are given by the x_i. For each pixel in the thermal image eqn (9) defines a hyperplane in 5-D space.

Consider two different LWIR images of a scene obtained under different scene conditions and from different viewpoints. Consider N points on the object that (a) are visible in both views, and (b) have been selected to lie on different components of the object which differ in material composition and/or surface normal direction. Assume (for the nonce) that the object pose for each view, and point correspondence between the two views are available (or hypothesized). A point in each view yields a contravariant tensor a^i as defined by eqn (9). Let the collection of these tensors be denoted by $a_k^i, k = 1, 2, \ldots, N$ for the first scene/image and $b_k^i, k = 1, 2, \ldots, N$ for the second scene. For the k-th point we denote the measurement tensor as $\mathbf{a_k}$ for the first view, and as $\mathbf{b_k}$ for the second view, and the driving conditions tensor as $\mathbf{x^k}$.

We assume that the scene/driving conditions, $\mathbf{x^k}$, in the two scenes are related by an affine transformation. The justification of this assumption is discussed below. We have found, empirically, that this assumption holds when the points are selected using the method discussed later in this paper. Since the $\mathbf{x^k}$ are transformed affinely, then it follows that the $\mathbf{a_k}$ are also transformed affinely. Note that an affine transformation from one scene to another is trivial to obtain if we have

[1] Here, we use the Einstein notation to denote the image-based measurement vector as a contravariant tensor, a^i, while the changing scene conditions form a covariant tensor, x_i. Therefore, $a^i x_i = 0$.

only five points that generate five non-coplanar tensors in our 5-D measurement space. Consider one such subset of 5 of the N points, and denote them as i, j, l, m, and n.

The determinant

$$d(\mathbf{a_i a_j a_l a_m a_n}) = \begin{vmatrix} a_i^1 & a_i^2 & a_i^3 & a_i^4 & a_i^5 \\ a_j^1 & a_j^2 & a_j^3 & a_j^4 & a_j^5 \\ a_l^1 & a_l^2 & a_l^3 & a_l^4 & a_l^5 \\ a_m^1 & a_m^2 & a_m^3 & a_m^4 & a_m^5 \\ a_n^1 & a_n^2 & a_n^3 & a_n^4 & a_n^5 \end{vmatrix} \tag{10}$$

defines the volume of the oriented parallelopiped formed by the pencil of the five contravariant tensors $\mathbf{a_i}, \mathbf{a_j}, \mathbf{a_l}, \mathbf{a_m}, \mathbf{a_n}$. The above determinant is a relative invariant to the affine transformation [Gurevich, 1964], i.e.,

$$d(\mathbf{a_i a_j a_l a_m a_n}) = \delta_{ijlmn} \times d(\mathbf{b_i b_j b_l b_m b_n})$$

where δ_{ijlmn} is the determinant of the affine transformation, T_{ijlmn}, which relates the measurement tensors, i.e., $\mathbf{a_k} = \mathbf{b_k} T_{ijlmn}$, $k \in \{i, j, l, m, n\}$.

Consider another set of five points in which at least one point is different from the previous set (see figure 4). Denote this second set as $\{p, q, r, s, t\}$. Again, assume that the measurement tensors for this collection of points undergo an affine transformation from the first scene to the second, and denote this transformation by T_{pqrst}.

$$d(\mathbf{a_p a_q a_r a_s a_t}) = \delta_{pqrst} \times d(\mathbf{b_p b_q b_r b_s b_t})$$

where $\delta_{pqrst} = \det(T_{pqrst})$. Hence, if $\delta_{ijlmn} = \delta_{pqrst}$, then we can define an absolute invariant as

$$I = \frac{d(\mathbf{a_i a_j a_l a_m a_n})}{d(\mathbf{a_p a_q a_r a_s a_t})} \tag{11}$$

Note that the existence of affine transformations T_{ijlmn} and T_{pqrst} is easy to ensure by selecting the surface points appropriately (discussed in section 5). The selection of two sets of five points such that $\delta_{ijlmn} = \delta_{pqrst}$ holds can be attempted as a data-driven training task as follows.

Calibrated LWIR imagery from different object classes are obtained at different times of day and different seasons of the year. N points are picked on an object – on distinctive components that differ in material composition and/or surface normal. Consider the image from time t_u and the image from t_v, $u \neq v$. The measurements at t_u along with the hypothesis of the identity of the

object form the tensors $\mathbf{a_k}$. Similarly, image information at time t_v is used to form the measurement tensors $\mathbf{b_k}$.

All combinations of two sets of five points each, $\{i, j, l, m, n\}$ and $\{p, q, r, s, t\}$, are examined. The measurement matrices $(\mathbf{a_i a_j a_l a_m a_n})$, $(\mathbf{b_i b_j b_l b_m b_n})$, $(\mathbf{a_p a_q a_r a_s a_t})$, and $(\mathbf{b_p b_q b_r b_s b_t})$ are constructed The transformations T_{ijlmn} and T_{pqrst}, if they exist, and the their determinants δ_{ijlmn}, and δ_{pqrst} are computed. The two sets that best satisfy $\delta_{ijlmn} = \delta_{pqrst}$ for different choices of pairs of scenes/images, i.e. different choices of t_u and t_v, are selected. With N points, the number of possible choices of the pair of sets of points is given by

$$n_N = \frac{1}{2}\left(\frac{N!}{(N-5)!5!}\right)\left(\frac{N!}{(N-5)!5!} - 1\right)$$

For example, $n_6 = 15$, $n_8 = 1,540$, and $n_{11} = 106,491$.

In order for the invariant feature to be useful for object recognition the value of I must be different if the measurement vector is obtained from a scene that does not contained the hypothesized object and/or the hypothesized pose is incorrect. Hence, the search for the optimal sets of points conducted during training phase should also take into consideration class separability, in addition to intra-class variation.

The linear form, eqn (9), must be slightly modified for interpreting LWIR imagery acquired at night. Since solar insolation is nonexistent, W_{abs} as defined above is zero, and the energy balance model has only four terms. Hence, the measurement tensor is four dimensional, and we consider sets of four points in evaluating the absolute invariant. A separate training phase is required, resulting in the specification of a different choice of points for night-time use.

4. Experimental Results

The method of computing thermophysical affine invariants discussed above was applied to real LWIR imagery acquired at different times of the day. Two types of vehicles were imaged: Type-A (figure 5) and Type-B (figure 6). Several points were selected (as indicated in the figures) on the surfaces of different materials and/or orientation. The measurement tensor given by eqn (9) was computed for each point, for each image/scene.

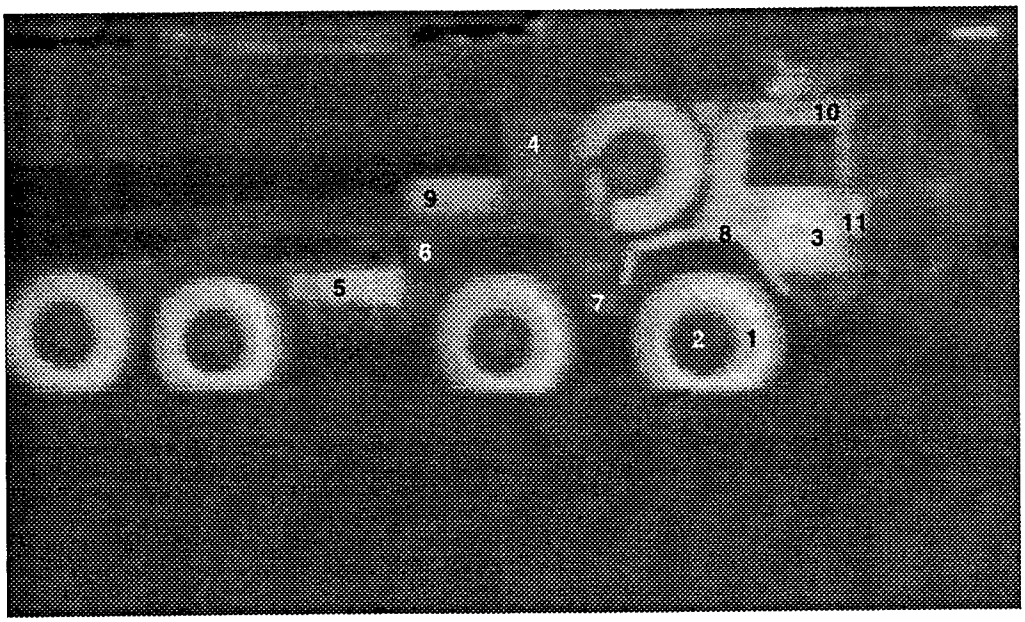


Figure 5: Type-A vehicle with points selected on the surface of different components with different material properties and/or surface normals.

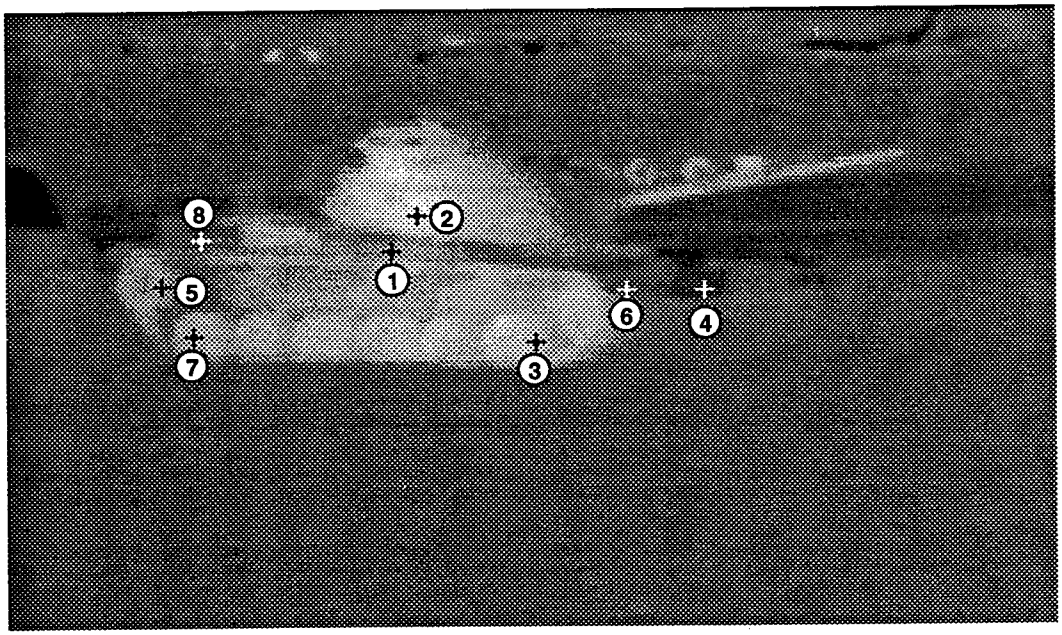


Figure 6: Type-B vehicle with points selected on the surface of different components with different material properties and/or surface normals.

Time	Set A-1	Set A-2
11 am	0.154	-10.9
12 n	0.152	-10.8
1 pm	0.153	-11.0
2 pm	0.154	-10.9
3 pm	0.154	-10.9

Table 1: Variation of feature values for two different choices of points. A-1 consists of point sets $\{1,3,10,6,7\}$ and $\{3,4,10,6,7\}$. A-2 consists of point sets $\{1,3,4,6,7\}$ and $\{2,3,6,7,8\}$. The numbers correspond to the labels in the image of the Type-A vehicle.

Time	Set B-1	Set B-2
11 am	-1.5	0.4
12 n	-1.6	0.7
1 pm	-1.5	0.2
2 pm	-1.6	0.4
3 pm	-1.5	0.4

Table 2: Variation of feature values for two different choices of points. B-1 consists of point sets $\{1,2,3,6,8\}$ and $\{1,2,5,6,8\}$. B-2 consists of point sets $\{1,2,4,5,7\}$ and $\{2,3,5,6,7\}$. The numbers correspond to the labels in the image of the Type-B vehicle.

Time of Day	Hypothesis: Type-A Data from: Type-B	Hypothesis: Type-A Data from: Type-A
11 am	.003	-.03
12 n	.01	-.01
1 pm	.01	-.02
2 pm	.01	-.01
3 pm	.001	-.01

Table 3: Inter-class variation vs. intra-class variation for set A-3 consisting of point sets $\{2,3,4,6,7\}$ and $\{3,4,10,7,8\}$ for Type-A vehicle. Thermophsyical properties are chosen for a Type-A vehicle hypothesis. The middle column shows feature values computed when the measurement tensor is obtained from a Type-B vehicle.

Time of Day	Hypothesis: Type-B Data from: Type-A	Hypothesis: Type-B Data from: Type-B
11 am	-.001	0.4
12 n	-.01	0.7
1 pm	-.03	0.2
2 pm	-.002	0.4
3 pm	.05	0.4

Table 4: Inter-class variation vs. intra-class variation for point set B-2. Points and thermophsyical properties are chosen for a Type-B vehicle hypothesis. The middle column shows feature values computed when the measurement tensor is obtained from a Type-A vehicle.

The method used to select optimal sets of points $\{i, j, l, m, n\}$ and $\{p, q, r, s, t\}$ was similar to that described in section 3 – however, instead of using the equivalence of the determinants of the two affine transformations as the selection criterion, we used the variance in the values of the feature computed for different scenes (i.e., images obtained at different times of day) containing the object. Many different pairs of five-point-sets yielded features with low variance from scene to scene. Table 1 shows values of the feature for two different choices of sets of points for the Type-A vehicle, and table 2 shows values for two different choices of points for the Type-B vehicle.

As mentioned in section 3 one must consider inter-class behavior as well as intra-class behavior. To investigate this we adopted the following procedure. Given an image of a vehicle, (1) assume the pose of the vehicle is known. (2) use the extremal forward and rear road-wheels to establish a reference frame. The coordinates of the selected points are expressed in terms of this frame. Thus when a type-A vehicle is hypothesized for an image actually obtained of a type-B vehicle, the material properties of the type-A vehicle are used, but image measurements are obtained from the image of the type-B vehicle at locations given by transforming the coordinates type-A object points to the type-B image frame.

Table 3 shows inter-class and intra-class variation when type-A vehicle is hypothesized, and for images obtained at five different times in the day. Table 4 shows inter-class and intra-class variation when type-B vehicle is hypothesized. Such investigation showed that the set of points A-1, A-2, and B-1 produced almost identical values irrespective of the source of the measurements. These sets have good (low) intra-class variation but poor inter-class separation, and they do not distinguish a type-A vehicle from a type-B vehicle. Sets A-3 and B-2, however, can be seen to demonstrate good inter-class separation and acceptable intra-class variation.

5. Discussion

The approach described above is promising in that it makes available features that are (1) invariant to scene conditions, (2) able to separate different classes of objects, and (3) based on physics based models of the many phenomena that affect LWIR image generation.

The specification of optimal sets of points for high inter-class separation and low intra-class variation is a crucial task in this approach. This is a complex search problem, and it is not clear that a solution will always exist for a collection of object classes. Note that different aspects of an object may be imaged – the set of visible points differ for each aspect. The complexity of the search

task is compounded by attempting to ensure inter-class separation in the presence of erroneous pose hypothesis.

Some criteria for the choice of point sets are obvious. Points should not be chosen such that the measurement matrix has less than full rank. This occurs, for example, when the five points lie on surfaces with identical surface normals and identical thermophysical properties. When the two sets of points have four points in common, and when the remaining two points are on different parts of the imaged object, but lie on materials that are identical/similar, then these points will form an invariant that has magnitude of one. Such invariants are useful only if there exists a unique set, or limited number of sets, of points that produces this value. Other thermophysical criteria for point set selection need to be investigated.

There remain many issues that are not well understood and require further research. An important assumption/observation that requires justification from a thermophysical viewpoint is that the tensor of driving conditions (for the two sets of points used to compute the invariant) undergoes an affine transformation from scene to scene, and that the affine transformations for the two sets have identical determinants for any two scenes. While an affine transformation can model a wide range of changes in the driving conditions, and hence is intuitively a good choice - a thermophysical interpretation of this assumption/observation would be helpful in understanding the behavior of invariants for different choice of points on the object. Another important issue is to choose a linear form that is not homogeneous - by replacing the zero in the right hand side of eqn (8) with terms from the left hand side. This would reduce the dimensionality of the measurement vector, and depending on the choice of terms – eliminate the requirements for hypothesis of pose and hypothesis of specific thermophysical parameter values, but at the risk of impaired invariance and separability.

The hypothesis of object pose and identity is best achieved by employing geometrical invariance techniques [Forsyth et al., 1991]. For example, conics may be fit to wheels which manifest high contrast in LWIR imagery, and their parameter values may be used to compute GI's. This may be employed to generate object identity and pose that may be verified by the thermophysical invariance scheme described above. Future effort will be devoted to: the integration of the above scheme with GI's to produce a complete system, the study of the nature of scene-to-scene transformation of driving conditions and justification of the affineness of this transformation, and

a detailed exploration of the performance of the scheme when applied to a significant collection of objects, aspects, and scene conditions.

References

[1] T. Binford, T.S. Levitt, and W.B. Mann, "Bayesian Inference in Model-Based Vision", *Uncertainty in AI, 3*, L.N. Kanal, T.S. Levitt, and J.F. Lemmer, (Ed's), Elsevier, 1989.

[2] J.B. Burns, R.S. Weiss, and E.M. Riseman, "View Variation of Point-Set and Line-Segment Features", *IEEE Trans PAMI*, vol 15, no 1, Jan 1993.

[3] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, C. Rothwell, "Invariant Descriptors for 3D Object Recognition and Pose", *IEEE Trans PAMI*, vol 13, no 12, Oct 1991

[4] M.J. Gauder, V.J. Velten, L.A. Westerkamp, J. Mundy, and D. Forsyth, "Thermal Invariants for Infrared Target Recognition", *ATR Systems and Technology Conf*, 1993.

[5] G.B. Gurevich, "Foundations of the Theory of Algebraic Invariants", (translated by J.R.M. Raddock and A.J.M. Spencer) P. Noordhoff Ltd - Groningen, The Netherlands, 1964

[6] G. Healey, "Using Color to Segment Images of 3-D Scenes", *Proc SPIE Conf Applications of AI*, vol. 1468, 1988, Orlando, FL, pp. 814-825.

[7] G. Healey and D. Slater, "Using Illumination Invariant Color Histogram Descriptors for Recognition", *Proc IEE Conf CVPR*, June 21-24, 1994, Seattle, WA, pp. 355-360.

[8] F.P. Incropera and D.P. DeWitt, *Fundamentals of Heat Transfer*, John Wiley and Sons, New York, 1981.

[9] G.J. Klinker, S.A. Shafer and T. Kanade, "Image Segmentation and Reflection Analysis through Color", *Proceedings of DARPA Image Understanding Workshop*, Cambridge, MA, 1988, pp 838 - 853.

[10] D.G. Lowe, "The Viewpoint Consistency Constraint," *International Journal of Computer Vision*, vol. 1, no. 1, 1987, pp. 57-72.

[11] J.D. Michel and N. Nandhakumar, "Unified Octree-Based Object Models for Multisensor Fusion", *Proc 2nd IEEE Workshop on CAD Model-Based Vision*, Champion, PA, 1994.

[12] N. Nandhakumar and J.K. Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, July 1988, pp. 469-481.

[13] N. Nandhakumar and J.K. Aggarwal, "Thermal and Visual Information Fusion for Outdoor Scene Perception", *Proc. of IEEE International Conference on Robotics and Automation*, Philadelphia, PA, April 1988, pp. 1306-1308.

[14] N. Nandhakumar, "A Phenomenological Approach to Multisource Data Integration: Analyzing Infrared and Visible Data", *Proc. IAPR TC7 Workshop on Multisource Data Integration in Remote Sensing,* College Park, MD, June 14-15, 1990.

[15] N. Nandhakumar, "Robust Physics-based Sensor Fusion", Journal of the Optical Society of America, JOSA-A, 1994, to appear.

[16] S.K. Nayar and R.D. Bolle, "Reflectance Ratio: A Photometric Invariant for Object Recognition", *Proc IEEE ICCV*, 1993.

[17] T.H. Reiss, "Recognizing Planar Objects Using Invariant Image Features", *Lecture Notes in Computer Science, 676,* G. Goos and J. Hartmanis (Eds), Springer-Verlag, Berlin, 1993.

[18] E. Rivlin and I. Weiss, "Semi-Local Invariants", *Proc IEEE CVPR* 1993, pp. 697-698

[19] D. Weinshall, "Direct Computation of Qualitative 3D Shape and Motion Invariants", *IEEE Trans PAMI*, vol 13, no 12, Dec 1991.

[20] D. Weinshall, "Model-based Invariants for 3D Vision", *Proc IEEE CVPR* 1993, pp. 695-696

[21] I. Weiss, "Noise-Resistant Invariants of Curves", *IEEE Trans PAMI*, vol 15, no 9, July 1993, pp. 943-948

[22] L.B. Wolff, "Polarization-based material classification from specular reflection", *IEEE Trans PAMI*, Nov 1990, pp 1059-1071.

[23] M. Zerroug and R. Nevatia, "Quasi-Invariant Properties and 3D Shape Recovery of Non-Straight, Non-Constant Generalized Cylinders", *Proc IEEE CVPR 1993*, pp 96-103

[24] R. Zhang, P.-S. Tsai, J.E. Cryer, M. Shah, "Analysis of Shape from Shading Techniques", *Proc IEEE CVPR*, Seattle, WA, 1994, pp. 377-384.

SURFACE OUTGROWTHS ON LASER-DEPOSITED $YBa_2Cu_3O_7$ THIN FILMS

M. Grant Norton
Assistant Professor
Department of Mechanical and Materials Engineering

Washington State University
Pullman WA 99164-2920

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

July 1994

SURFACE OUTGROWTHS ON LASER-DEPOSITED $YBa_2Cu_3O_7$ THIN FILMS

M. Grant Norton
Assistant Professor
Department of Mechanical and Materials Engineering
Washington State University

Abstract

The surface morphology of $YBa_2Cu_3O_7$ thin films formed on (001)-oriented $LaAlO_3$ substrates by pulsed-laser deposition has been examined using both scanning probe microscopy and transmission electron microscopy. Under certain deposition conditions a high density of surface outgrowths is observed. These outgrowths have been determined to be primarily $YBa_2Cu_3O_7$ grains oriented such that the c-axis of the unit cell is lying parallel to the film/substrate interface plane. The size and number density of these outgrowths do not vary significantly with deposition time, which suggests that they nucleate at the substrate surface at the same time as do the grains oriented with their c-axis perpendicular to the film/substrate interface plane.

SURFACE OUTGROWTHS ON LASER-DEPOSITED $YBa_2Cu_3O_7$ THIN FILMS

M. Grant Norton

Introduction

Several groups have reported the formation of $YBa_2Cu_3O_7$ (YBCO) thin films that consist of a mixture of grain orientations: where the c-axis is both perpendicular (c-perpendicular) and parallel (c-parallel) to the film/substrate interface plane [e.g., Refs. 1-4]. This type of microstructure, which can be formed in films grown by using various deposition techniques, typically has rectangular outgrowths (c-parallel grains) on the surface of an otherwise smooth (c-perpendicular grains) film. Two mechanisms have been proposed to account for the formation of the c-parallel grains. It is assumed that either compositionally different precipitates on, or else inclusions in, the c-perpendicular film act as nucleation sites for the c-parallel grains [e.g., Ref. 5]. These grains would then more rapidly grow outward in directions perpendicular to the c-axis of the YBCO unit cell, thus explaining their rectangular shape.

A second mechanism that has been proposed for the nucleation of the c-parallel grains is that step edges on c-perpendicular oriented spirals (associated with screw-dislocation-mediated growth) act as preferential sites for nucleation [e.g., Ref. 6]. Thus, in mixed c-perpendicular/c-parallel films the underlying film would be oriented in a c-perpendicular direction up to some arbitrary thickness, beyond which a significant number of c-parallel grains are then produced. In the results reported by Basu *et al.* [4], YBCO thin films deposited on (001)-oriented $LaAlO_3$ substrates by pulsed-laser deposition (PLD) were found to be exclusively aligned in the c-perpendicular orientation up to a thickness of 400 nm; then c-parallel grains were nucleated. In another study, transmission electron microscopy (TEM) examination of YBCO films grown on $SrTiO_3$ substrates by sputtering, under conditions that favored growth of c-parallel oriented films, showed that a 2 to 15 nm layer of c-perpendicular material was formed adjacent to the substrate surface. Again, beyond this thickness, c-parallel grains were nucleated [7]. On the other hand, the existence of completely c-parallel oriented films demonstrates that c-parallel grains can also be nucleated directly on the substrate surface [e.g., Ref. 8].

In this paper the surface morphology of a series of YBCO films formed on (001)-oriented $LaAlO_3$ substrates by PLD has been examined using TEM, atomic force microscopy (AFM) and scanning tunneling microscopy (STM).

Experimental

The method used to deposit YBCO films in our laboratory has been reported in detail elsewhere [9]. Briefly, the YBCO films were formed by PLD using a Lambda Physik LPX 305i excimer laser operating on KrF (λ = 248 nm). For film deposition, an average pulse energy of 440 mJ was used, the beam footprint was ~1 mm^2, the pulse duration was 25 ns, and the pulse repetition rate was 20 Hz. The stoichiometric YBCO target was rotated at 10 rpm and the laser beam simultaneously scanned radially across the target surface to avoid preferential ablation. The films were grown on single crystal (001)-oriented $LaAlO_3$ substrates (5 x 5 x 0.5 mm). The substrates were heated by a tungsten-filament lamp located below the stainless steel plate on which they were mounted. The substrate temperature during deposition, measured by means of a two-color infra-red pyrometer, was ~770 °C. Film deposition was performed in a flowing oxygen ambient (~160 mTorr). Following film deposition, the chamber was filled with oxygen and the samples allowed to cool to room temperature at a pressure of ~490 Torr of oxygen. Films were grown for the following times: 5, 15, 45, 125, and 250 seconds, and the reproducibility of the experimental conditions was determined by repeating several of the runs.

The surface morphology of the films was investigated via AFM using the Digital Instruments Nanoscope II system operating in contact mode with triangular SiN cantilevers. STM images were acquired using the Nanoscope II in constant current mode. The set point current was 1 nA and the bias voltage was 1000 mV. Good results were

obtained only with fresh Pt-Ir tips. No image filtering was used. Microstructural characterization of the films was performed using a Jeol 2000FXII TEM operating at an accelerating voltage of 200 kV.

Results

The AFM images shown here are representative of the surface morphology associated with YBCO films grown on (001)-oriented $LaAlO_3$ substrates using the deposition conditions described above. Images were recorded from many films and at least five areas in each film were examined.

Figure 1 shows an AFM image of an ~6-nm-thick film deposited during 5 seconds of growth. A number of surface features is visible in this image. These features are uniform in size (~200 x 300 nm) and oriented along a common direction with their long edges parallel. These outgrowths have a density of ~4 x 10^8 cm^{-2} and a height of ~30 nm above the film surface.

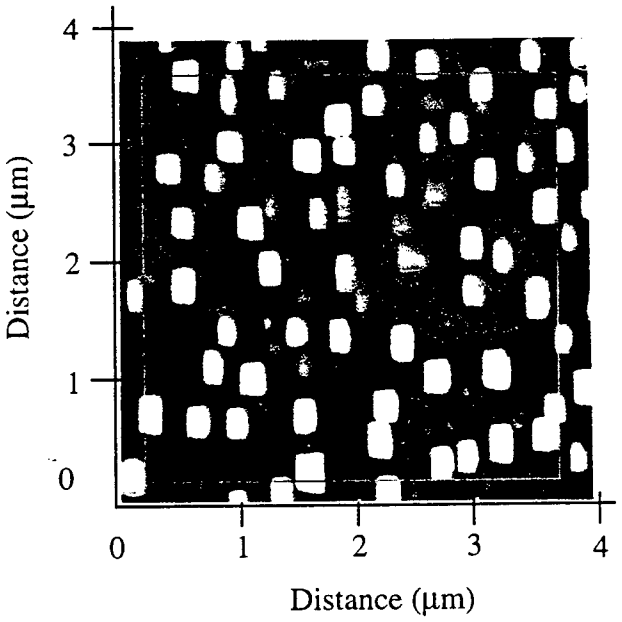


Figure 1. AFM image of a YBCO film deposited during 5 seconds. The z-range is 0 to 200 nm.

Figure 2 shows an AFM image of a 145-nm-thick film deposited during 125 seconds of growth. Several types of surface feature were found in these samples. These included elongated outgrowths (~200 x 300 nm and 30 to 50 nm high), small equiaxed features (100 to 300 nm square and 8 to 30 nm high), and irregular-shaped features of a similar size. All these features had their straight sides aligned either along or perpendicular to a common direction. The total outgrowth density remains ~4 x 10^8 cm^{-2}. A typical STM image of a region of the film, between the outgrowths, is shown in Figure 3.

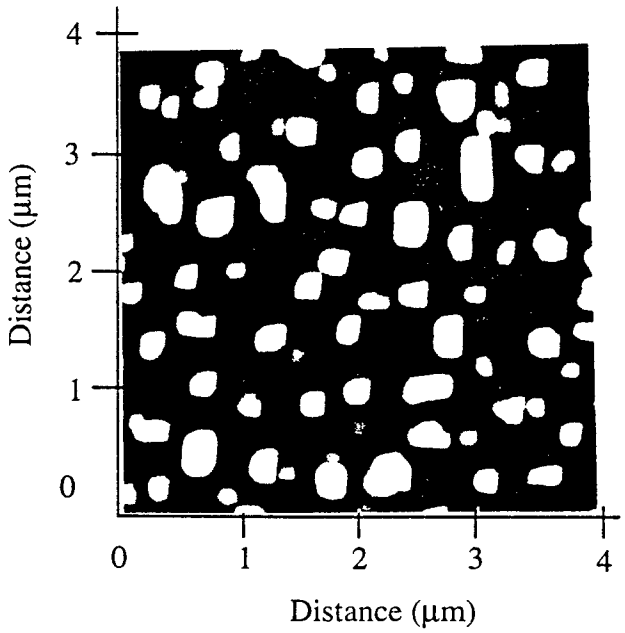


Figure 2. AFM image of a YBCO film deposited during 125 seconds. The z-range is 0 to 200 nm.

Films deposited during 250 seconds appeared similar to those formed during 125 seconds of growth. The only significant differences between the films were that, for the longer deposition times, some of the outgrowths had become more irregular in shape and in most cases their heights above the film surface had increased to between 75 to 100 nm.

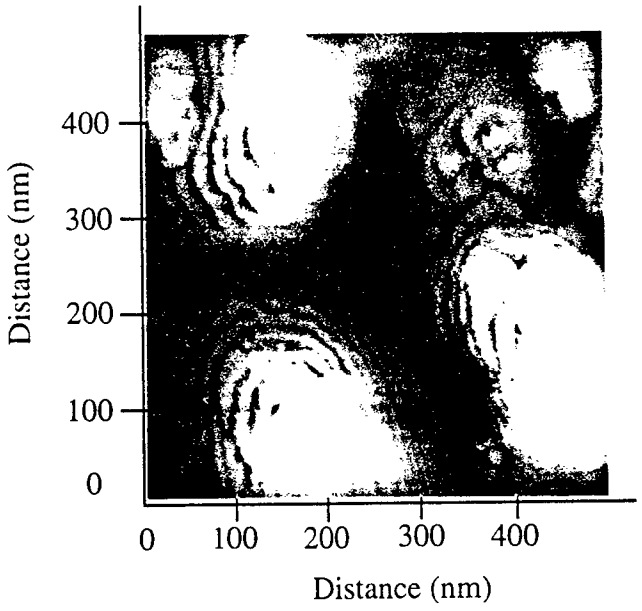


Figure 3. STM image of a YBCO film, recorded between the outgrowths, showing c-axis terraced growth. The z-range is 0 to 50 nm.

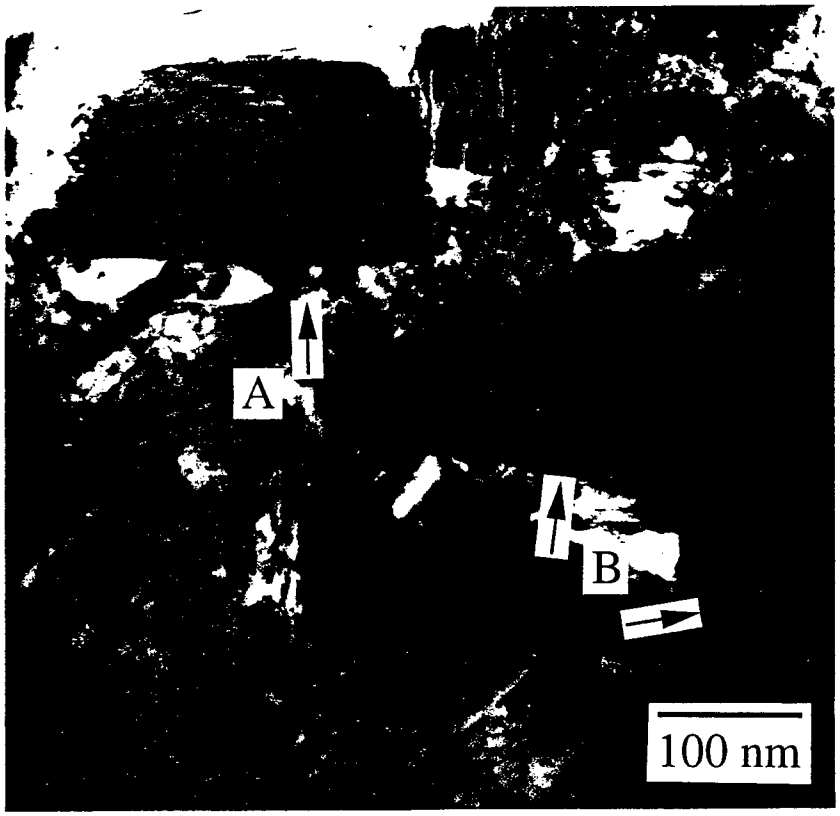


Figure 4. Bright-field plan view image of a YBCO film showing a typical rectangular outgrowth (A) and two merged outgrowths (B).

Rectangular outgrowths similar in size to those found by AFM were also observed by TEM. Figure 4 shows a typical bright-field TEM image of an area of a film deposited during 250 seconds where an example of a rectangular outgrowth has been indicated. The spacing of the lattice fringes visible in the images of the outgrowths is consistent with that of c-parallel oriented YBCO. Selected-area electron diffraction (SAED) patterns recorded from the same area of film shown in Fig 4, confirmed the identification of these outgrowths as c-parallel oriented grains.

Some of the larger, irregularly shaped features seen in the AFM images of the thicker films appear to be made up of two or more merged outgrowths. In some cases these merged outgrowths were c-parallel oriented YBCO grains aligned in orthogonal directions as illustrated by B in Fig. 4. In other cases, illustrated in Figure 5, the secondary outgrowths were precipitates having a composition different than the rest of the film. Energy dispersive spectroscopy (EDS) analysis in the TEM of several of these secondary outgrowths indicated that they consisted primarily of either copper- or yttrium-and-copper compounds. With the EDS system used in this study it was not possible to determine the presence of oxygen in these outgrowths, but it is presumed that the secondary outgrowths were either copper oxide or an yttrium-copper-oxide phase. Also found were some complex outgrowths that consisted of a combination of orthogonally oriented c-parallel YBCO grains and second-phase precipitates. An example of such an outgrowth is shown in Figure 6.

Discussion

As illustrated by Figs. 1 and 2, the density of outgrowths remains approximately constant over a wide range of deposition times (corresponding to average film thicknesses from 6 nm to 300 nm). This consistency suggests that these outgrowths (c-parallel grains) are nucleated at the film/substrate interface and continue to grow outward in concert with the c-perpendicular grains as the film is deposited. This growth mechanism therefore differs from that usually associated with the formation of mixed c-parallel/c-perpendicular YBCO films. The homogeneous size range of the outgrowths also suggests that they were nucleated at about the same time in the growth process.

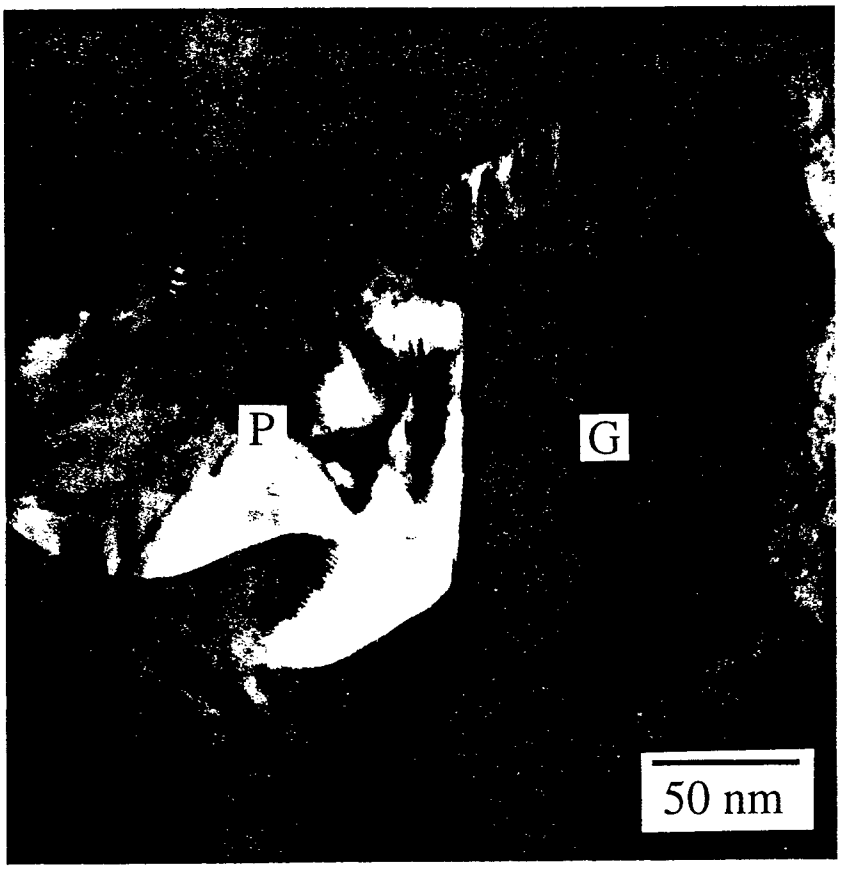


Figure 5. Bright-field image showing an example of precipitate (P) formation at the edge of a c-parallel YBCO grain (G).

STM images of the regions of the film between the outgrowths show typical c-axis terraced growth of up to 11 resolvable terraces (each one unit cell thick). No evidence of the spiral growth pattern associated with the presence of screw dislocations has been observed in films grown using the specific deposition conditions described above. Therefore, the outgrowths observed here are not related to step edges on c-perpendicular spirals, one type of nucleation site for c-parallel grains proposed earlier [6].

It is interesting that the density of the outgrowths reported here is very similar to the density of outgrowths reported for sputtered YBCO films deposited at similar temperatures on (001)-oriented $LaAlO_3$ substrates [10]. It should, however, be noted that the outgrowths observed by Han and co-workers [10] were determined to be Cu-rich particles, and not the type of YBCO outgrowths reported here. Furthermore, in this cited study the number of

outgrowths was found to saturate at an early stage of growth—the same process that appears to be occurring in this present study. The number density of c-parallel grains is also very similar to that reported by Nieh and co-workers in computer simulated studies of YBCO thin film growth on $LaAlO_3$ substrates [3]. In that study the c-parallel grains increased in size with increasing film thickness although the number density remained the same. Such behavior is similar to what is reported here but only for deposition times ranging from 125 to 250 seconds.

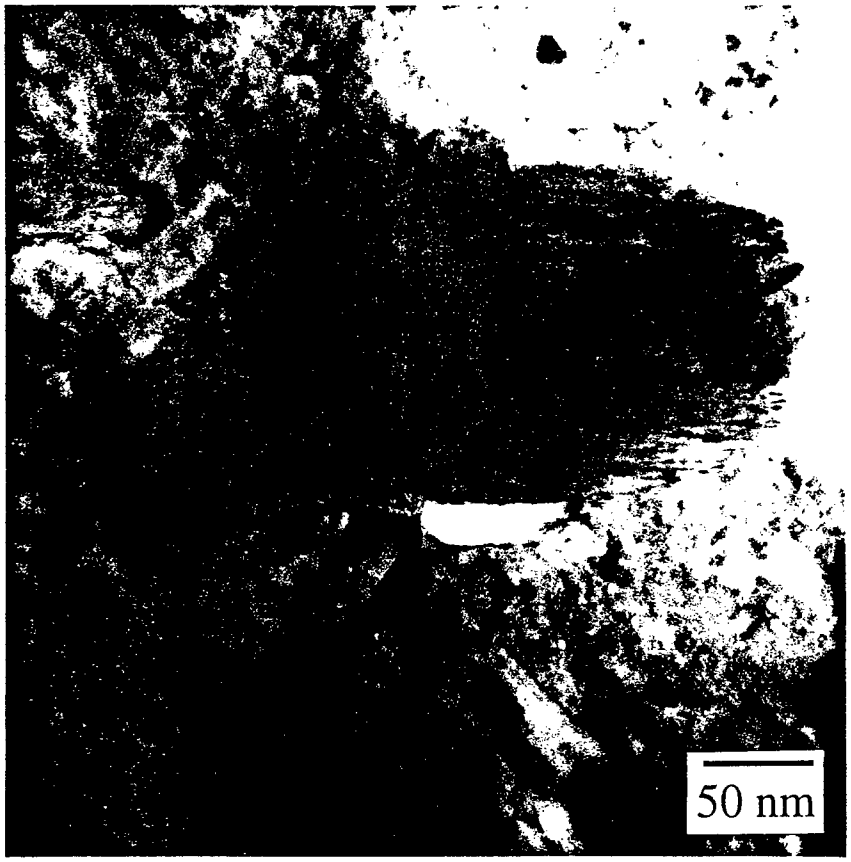


Figure 6. Bright-field image showing an outgrowth with a complex shape found on the surface of a YBCO film deposited during 250 seconds.

At the shorter deposition times most outgrowths show only a very slight change in size and shape with time. By about 125 seconds of growth, the ones that have merged with nearby primary outgrowths or significantly increased their size via the formation of secondary outgrowths continue to increase in size with increasing film

thickness. By 250 seconds of growth the outgrowths can extend from 75 to 100 nm above the film and have dimensions of 500 nm and upwards.

Preliminary studies in which the (001)-oriented $LaAlO_3$ substrates were annealed in air at temperatures up to 1100°C prior to film deposition indicate that, under the equivalent deposition conditions described above, the number of surface outgrowths was different. In general, the outgrowths appeared to be somewhat larger and more numerous. The high-temperature annealing process would be expected to remove some of the surface damage introduced by polishing.

The number density of outgrowths was also found to be strongly dependent upon the deposition parameters used. In general, decreasing the substrate temperature, lowering the laser energy, increasing the oxygen pressure, and increasing the laser pulse repetition rate all appear to increase the density of the outgrowths [9]. Several of these trends have been observed by other groups [e.g., Ref. 11].

For the longer deposition times some change in the outgrowth size and height occurs but as mentioned earlier this appears to be due to nucleation of secondary outgrowths and/or to outgrowth coalescence. The secondary outgrowths, which occur later in the growth process, are often associated with the c-parallel oriented YBCO grains and lead to the formation of the irregularly shaped features seen in the AFM images. For films deposited during shorter deposition times the outgrowths were all very regular in size and shape. Thus it is likely that the c-parallel YBCO grains act as favorable nucleation sites for precipitate formation. This correlation between c-parallel YBCO grains and the precipitation of other phases has also been reported for YBCO thin films deposited by sputtering [12]. In this present study most of the secondary outgrowths were found to be either copper oxide or yttrium copper oxide. Quantitative EDS analysis of these outgrowths was not possible, however $CuYO_2$ precipitates have been observed in sputter-deposited YBCO films [12] and both $Y_2Cu_2O_5$ and CuO precipitates in YBCO films deposited by PLD [13].

Conclusion

In conclusion, surface outgrowths produced on YBCO films deposited by PLD have been examined. The outgrowths were found to be predominantly c-parallel YBCO grains which under certain deposition conditions appear to nucleate on the substrate surface at the same time as do the c-perpendicular grains. During longer depositions some of the outgrowths become irregular in shape and significantly increase in size. This is due to the formation of precipitates at the edges of the c-parallel grains and/or the merging of two or more c-parallel grains.

MGN acknowledges numerous interactions with colleagues at Wright Laboratory, especially Dr. Rand Biggers, Dr. Iman Maartense, and Dr. Eric Moser. Technical support and assistance from David Dempsey, Jeff Brown, and John Woodhouse is greatly appreciated.

REFERENCES

1. A. Catana, D.G. Schlom, J. Mannhart, and J.G. Bednorz, Appl. Phys. Lett. 61 (1992) 720

2. R. Ramesh, A. Inam, D.M. Hwang, T.D. Sands, C.C. Chang, and D.L. Hart, Appl. Phys. Lett. 58 (1991) 1557

3. C.W. Nieh, L. Anthony, J.Y. Josefowicz, and F.G. Krajenbrink, Appl. Phys. Lett. 56 (1990) 2138

4. S.N. Basu, A.H. Carim, and T.E. Mitchell, J. Mater. Res. 6 (1991) 1823

5. C.C. Chang, X.D. Wu, R. Ramesh, X.X. Xi, T.S. Ravi, T. Venkatesan, D.M. Hwang, R.E. Muenchausen, S. Foltyn, and N.S. Nogar, Appl. Phys. Lett. 57 (1990) 1814

6. I.D. Raistrick and M. Hawley, in: Interfaces in High-Tc Superconducting Systems, edited by S.L. Shindé and D.A. Rudman, Springer-Verlag New York (1993) pp. 28-70.

7. S.K. Streiffer, B.M. Lairson, E.M. Zielinski, T. Umezawa, T.H. Geballe, and J.C. Bravman, Proc. 50th. Annual Meeting of EMSA, edited by G.W. Bailey, J. Bentley, and J.A. Small, San Francisco Press, San Francisco (1992) p. 240.

8. A.C. Westerheim, A.C. Anderson, D.E. Oates, S.N. Basu, D. Bhatt, and M.J. Cima, J. Appl. Phys. 75 (1994) 393

9. M.G. Norton, R.R. Biggers, I. Maartense, E.K. Moser, and J.L. Brown, in: Epitaxial Oxide Thin Films and Heterostructures, edited by D.K. Fork, R. Ramesh, J.M. Phillips, and R.M. Wolf, Mater. Res. Soc. Symp. Proc. 341 (1994 in press).

10. Z. Han, T.I. Selinder, and U. Helmersson, J. Appl. Phys. 75 (1994) 2020

11. R. Ramesh, A. Inam, T. Sands, and C.T. Rogers, Mater. Sci. and Eng, B14 (1992) 188

12. A. Catana, J.G. Bednorz, Ch. Gerber, J. Mannhart, and D.G. Schlom, Appl. Phys. Lett. 63 (1993) 553

13. O. Eibl and B. Roas, J. Mater. Res. 5 (1990) 2620

The Importance of Lower Orbital Relaxations in Polymer Band Gaps: Failure of the Frozen Orbital Approximation

James F. O'Brien
Professor Department of Chemistry

Southwest Missouri State University
Springfield, MO 65804

Final Report for:
Summer Faculty Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Wright Patterson Air Force Base, OH

and

Wright Laboratory

August 1994

The Importance of Lower Orbital Relaxations in Polymer Band Gaps: Failure of the Frozen Orbital Approximation

James F. O'Brien
Professor
Department of Chemistry
Southwest Missouri State University

Abstract

Theoretical calculations were performed on oligomers of all trans polyenes ranging from C_2H_4 to $C_{40}H_{42}$. All calculations used the GAMESS molecular orbital package. *Ab initio* calculations using the 6-31G* basis set were done on oligomers C_2H_4 to $C_{20}H_{22}$. The energies of the geometry optimized ground states of the oligomers were computed at the RHF level. The energies of the singlet and triplet excited states were calculated using the same basis set by a minimal configuration interaction (CI). In these CI calculations the ground state geometries were used; and all ground state orbitals but the HOMO and LUMO were frozen. In the limit of infinite chain length the energy difference between the HOMO and LUMO linearly approached hartrees 6.80 eV. The energy differences between the ground state and the CI excited states were also computed. The values of the coulomb and exchange integrals of the polyenes were computed.

Semi-empirical calculations were done on oligomers C_2H_4 to $C_{40}H_{42}$. Geometries were optimized at the AM1 level and the ground state energies computed. The energy of the triplet excited state was computed using the ground state geometry, but allowing full electronic relaxation in the ROHF approach. The HOMO-LUMO gap extrapolated to 6.17 eV. The energy difference between the ground state and the triplet excited state extrapolated to 1.67 eV. This result, in excellent agreement with the experimentally measured band gap in polyacetylene demonstrates that computation of accurate band gaps for polymers requires that electronic relaxation of excited state wavefunctions be allowed. Most previous workers have done no excited state calculations for polymer band gaps, but have merely extrapolated the value of the HOMO-LUMO gap as obtained from ground state calculations.

The Importance of Orbital Relaxations in Polymer Band Gaps:
Failure of the Frozen Orbital Approximation

James F. O'Brien

Introduction

Conjugated polymers are of considerable interest for a variety of electronic and optoelectronic applications [1-8]. An intrinsic fundamental property of such systems is the band gap. Accordingly much effort has been expended to compute the band gap, and a variety of approaches has been used. These are surveyed in a recent article [9]. Most of the previous efforts to calculate the band gap employed a variant of Koopman's theorem [10] in which the band gap is set equal to the difference in energy between the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO). When this is done for a series of oligomers, the value extrapolated to the infinite polymeric limit is taken to be the band gap. This approach holds that electronic relaxation vanishes in the long limit. Unfortunately, this grossly overestimates the band gap, typically by $\sim$4 eV. This overestimation has been attributed to the lack of post-HF correlation [11]. Recent work in this laboratory which included the effects of full electronic relaxation showed, for trans-polyacetylene (t-PA), that a band gap of $\sim$2.0 eV could be obtained by strictly Hartree-Fock Self Consistent Field (HF-SCF) procedures. This is within 0.5 eV of the experimental value. More importantly for the present case, this work shows that near experimental values for the band gap can be achieved from SCF calculations with no recourse to post Hartree Fock (dynamic) correlation.

The goals of this work are two-fold. The first goal is a numerical evaluation of the $1\ ^1A_g \longrightarrow 1\ ^1B_u$ and $1\ ^1A_g \longrightarrow 1\ ^3B_u$ transitions in all trans-polyenes, C_nH_{n+2} using the frozen orbital approximation. This allows an evaluation of the coulomb and exchange terms within this approximation, and, by comparison with previous results, a measure of the importance of relaxation effects as a function of oligomeric length. The second goal is to examine the utility of semi-empirical methods for the full relaxation approach, with emphasis on AM1 methods. As we will show, the results obtained by the "AM1 method" strongly depend on the particular implementation, and the differing results are directly attributable to how the

implementation considers relaxation effects. In practical terms, if an appropriate semi-empirical level can be found, rapid assessment of band gap sizes will become possible. This would then serve as a significant guide to polymer synthesis and interpretation of spectral and electronic properties.

Background

Ab Initio Calculations

The lowest energy optically allowed electronic transition in the all trans polyenes is from the $1\ ^1A_g$ ground state to the $1\ ^1B_u$ excited state. Extrapolation of this spin allowed electronic transition energy to the infinite chain length limit yields the optical bandgap. In this study the energies of the three states were computed: the $1\ ^1A_g$ ground state, the $1\ ^1B_u$ excited state, and the $1\ ^3B_u$ excited state. The expressions typically given for the energies of these states are shown in equations 1-3 [12,13].

$$1\ ^1A_g \longrightarrow 1\ ^1B_u \ = \ ^1E - E_0 \ = \ \epsilon_L - \epsilon_H \ - J_{H,L} \ + 2K_{H,L} \tag{1}$$

$$1\ ^1A_g \longrightarrow 1\ ^3B_u \ = \ ^3E - E_0 \ = \ \epsilon_L - \epsilon_H \ - J_{H,L} \tag{2}$$

Thus

$$^1E - \ ^3E = 2K_{H,L} \tag{3}$$

where: $1\ ^1A_g \longrightarrow 1\ ^1B_u$ is the energy difference between the singlet excited state and the ground state

$1\ ^1A_g \longrightarrow 1\ ^3B_u$ is the energy difference between the triplet excited state and the ground state

ϵ_L and ϵ_H are the orbital energies of the LUMO and HOMO respectively

^1E is the energy of the singlet excited state

^3E is the energy of the triplet excited state

E_0 is the energy of the ground state

J is the coulomb integral, $<\varphi_i(1)\varphi_j(2)|\,1/r_{12}\,|\varphi_i(1)\varphi_j(2)>$

K is the exchange integral. $<\varphi_i(1)\varphi_j(2)|\,1/r_{12}\,|\varphi_j(1)\varphi_i(2)>$

These equations implicitly neglect two effects. First, they assume that the self-consistent field of the $1\ ^1A_g$ state is identical to those of the $1\ ^1B_u$ and $1\ ^3B_u$ states. In fact, since they are of different symmetries for the polyenes, independent orthogonal HF-SCF solutions can be obtained for each state which contain the electronically relaxed orbitals for

each state. This relaxation is the static (or average) correlation which is contained within the HF-SCF procedure. The second component neglected in these equations is dynamic (instantaneous) or post-HF correlation. Neverless, it is not the absolute post-HF correlation which is important, but the differential correlation between the states which will impact the extrapolated gap. We consider only the SCF relaxation effects in this work.

Equation 1-3 pertain to a frozen orbital picture, - no SCF orbital relaxation of the $1\ ^1B_u$ and $1\ ^3B_u$ states. Electron - electron interactions for these states are evaluated through the J and K terms, but the orbitals remain those from the ground state, $1\ ^1A_g$ solution. It has been shown theoretically that in this scheme J and K vanish for an infinite system [13, 14], yet this approach leads to the problem of overestimated transition energies. Nevertheless, we know of no evaluation of J and K as a function of oligomer length (degree of polymerization). We report the calculation of the energies of the ground state and two excited state. Equations (1) and (2) were then used to calculate the values of the coulomb integral, J, and the exchange integral, K. Both J and K are positive quantities [15].

In two previous works [9, 16], we showed that the $1\ ^3B_u$ state can approximate the $1\ ^1B_u$ state for a long chain. In the conventional wisdom of the field, which follows the approach outlined in the previous paragraph and equations (1) and (2), this approximation is exact since the values of both $1\ ^1A_g \longrightarrow 1\ ^1B_u$ and $1\ ^1A_g \longrightarrow 1\ ^3B_u$ will approach $\epsilon_L - \epsilon_H$, the HOMO - LUMO gap as J and K vanish. Thus the detailed examination of the evaluation of J and K presented here serves two purposes. First, it numerically demonstrates the proof and, in conjunction with the previous work [9, 16], highlights the importance of full HF-SCF calculations for all of the states. We felt that since a strong sentiment exists that post-HF correlation is responsible for poor band gap calculations, but that our recent work casts doubt on this, a numerical confirmation of these terms was in order. Secondly, it provides an estimate of where the triplet approximation for the excited singlet energy should be good.

Thus the goal of the *ab initio* part of the project was to demonstrate 1) that the calculations were able to reproduce the theoretical behavior of the energies of the states and of the coulomb and exchange integreals; and 2) that such calculations, which do not allow electronic relaxation of the excited state wavefunctions, do not adequately compute polymer band gaps.

Semi-empirical Calculations

The semi-empirical AM1 calculations differed from the *ab initio* calculations in that the excited triplet states were computed at the ROHF-SCF level. We are aware of no formulation or implementation of the AM1 (or MNDO or PM3) methods truly suitable for the ^1B_u state [17]. The ^3B_u state can be treated by AM1 depending upon implementation (vide infra). Thus in the AM1 work full electronic relaxation of the excited state wavefunctions was permitted. This was not the case in the *ab initio* work. This is the crucial point of this work. These calculations permit one to explore the possibility that full electronic relaxation in the excited state will enable calculation of band gaps to satisfactory agreement with experiment.

This SCF procedure relies on the difference in energy between two states, the ground state and the lowest lying triplet excited state (assuming that the lowest triplet excited state is close in energy to the singlet excited state in this highly delocalized system). Such calculations are feasible using semi-empirical methods, particularly since the energy of the triplet state in question is not very sensitive to basis set quality [16]. Hence, the first goal is to address the viability of the semi-empirical method, AM1, to this type of problem. The secondary goal is to examine how the values of the coulomb and exchange integrals vary with chain length. We present a numerical demonstration that these terms vanish, if one assumes the appropriate conditions. This is important as it bears on the particular implementation of AM1 utilized.

Computational Details

The *ab initio* calculations, which used the 6-31G* basis set, employed the GAMESS (Generalized Atomic and Molecular Electronic Structure System) molecular orbital package maintained and distributed by Iowa State University [18]. All polyenes were constrained to C_{2h} symmetry, and their structures were previously optimized at the RHF/6-31G* level. Ground state energies were calculated at the RHF level. The excited state *ab initio* calculations used the configuration interaction , CI, approach. These were not CI's in the conventional sense. For both the 1 ^1B_u and the 1 ^3B_u states the orbitals used were the RHF orbitals obtained from the 1 ^1A_g ground state. Only one electron was promoted from the HOMO to the LUMO, with the spin appropriate to either the singlet or triplet state. No

further excitations were included. This allows no orbital relaxations, and corresponds to the scenario described in equations (1) to (3). We note that while the J and K terms are included for these two electron - that is the J and K interaction terms - no orbital relaxation occurs. These are the ground state orbitals.

The AM1 semi-empirical calculations were done using GAMESS. The RHF calculations on the ground state used the *ab initio* geometries as starting geometries; but then reoptimized them so that the AM1 energies were computed for geometries optimum for that level of calculation. The wavefunctions optimized in the ground state calculation were used as input for the triplet state calculations; but were reoptimized using the ROHF approach.

For oligomers larger than $C_{20}H_{22}$ there were no *ab initio* geometries to use as input. Therefore, for C_{n+4} systems, the C_n optimized Z-Matrix was used as the starting geometry with four additional CH units added to the interior of the chain. Angles and bond lengths in the interior of the chain were all quite similar. All AM1 ground state calculations converged readily. The triplet state calculations on the larger oligomers gave convergence problems. Convergence on the triplet state calculation on $C_{38}H_{40}$ required that a ground state calculation be done on the dianion, $C_{38}H_{40}^{2-}$. The final wavefunctions from this calculation were used as input wavefunctions for triplet state $C_{38}H_{40}$. In this way the input LUMO for the triplet state calculation was an optimized orbital

In the first part of the project *ab initio* calculations were done at the RHF level on the ground states of polyene oligomers C_nH_{n+2}, with n = 2 to 20. Next, *ab initio* level configuration interaction, (CI), calculations were done for the same species on both one electron excited states, the ^1B_u and the ^3B_u states. The energy differences between the ground state and these two excited states, $1\ ^1A_g \longrightarrow 1\ ^1B_u$ and $1\ ^1A_g \longrightarrow 1\ ^3B_u$, were computed and plotted versus 1/n. This was done, in accordance with the "reciprocal rule" for polymers [19], in order to examine the behavior of these energy differences under conditions allowing no geometry relaxation and only partial interaction of electrons, i.e. with all but two molecular orbitals frozen at the ground state result.

The second part of the project consisted of an examination of the capability of the semi-empirical AM1 method to compute the band gap. Initially the energy of the ground state was calculated while allowing both geometry and electronic relaxation. Next full

relaxation of the triplet excited state wavefunctions was permitted in these ROHF calculations; however, the AM1 ground state geometries were used. Semi-empirical calculations require much less computer time and memory. If a successful semi-empirical approach can be found, much larger systems will be able to be computationally studied. In particular, it was the goal of the project to examine the possibility that AM1 level calculations on the ground state and the excited triplet state would extrapolate to the same result as previously done *ab initio* calculations in which full wavefunction relaxation in the excited states was permitted. AM1 calculations are easier to perform on the triplet excited state than on the singlet excited state, and the two states become energetically degenerate in the polymer limit [11,12]. Thus, knowing the results from the comparison of the *ab initio* calculations to the experimentally determined value of the band gap, one could assess the utility of the AM1 approach.

Results

Plots of quantities calculated for oligomers were made as a function of 1/n, where n is the number of carbon atoms in the oligomer. These plots were extrapolated to the polymer limit, 1/n = 0. In the *ab initio* work this was done for energy differences between the ^1A_g ground state and the ^1B_u and ^3B_u excited states. The results were then extrapolated to obtain the value of the optical band gap for PA. Table 1 gives the *ab initio* energies for polyenes C_2H_4 to $C_{20}H_{22}$. Figure 1 is a plot of the three ΔE values from this data vs. 1/n. Least squares analysis of the linear part of the $\Delta E_{H,L}$ plot results in an intercept of 0.2475 hartrees or 6.74 eV (correlation coefficient = 0.9999). The triplet state energy difference has a minimum at $C_{10}H_{12}$. The singlet state energy difference has its minimum at a chain length longer than $C_{14}H_{16}$.

Table 2 shows the values of the coulomb and exchange integrals computed from the *ab initio* calculations and using equations (1) through (3). This data, plotted in Figure 2, shows that both J and K are approaching zero in the polymer limit. It is also of interest that the rates of change of both J and K with 1/n are positive; and both rates get more positive with increasing chain length (i.e. with smaller 1/n). We will refer to this variation of J and K in the discussion section.

Table 3 contains the results of the AM1 semi empirical calculations. Plots of $\Delta E_{H,L}$ and $1\ ^1A_g \longrightarrow 1\ ^3B_u$ versus the reciprocal of the chain lengthwere made and are shown in Figure 3. Least squares analysis of the linear part of the HOMO - LUMO plot gave an intercept in the polymer limit of 0.2267 hartrees or 6.17 eV (correlation coefficient = 0.9994). The ground state - triplet excited state energy difference leveled out at 0.0615 hartrees or 1.67 eV. This corresponds to the band gap, assuming that for an "infinite" chain the singlet and triplet excited states become energetically degenerate.

Discussion

Figure 1 suggests that the energies of the ^1B_u and the ^3B_u excited states both approach the same value as does the HOMO-LUMO gap, $\epsilon_L - \epsilon_H$. Equations 1 and 2 show that the condition for this to occur is that both J and K approach zero. Figure 2 shows that this is indeed happening. These results highlight the source of the success of the calculations mentioned above in which $1\ ^1A_g \longrightarrow 1\ ^3B_u$ decreases monotonically to the extrapolated value of 2.0 eV. Those earlier calculations allowed full electronic relaxation by using the ROHF approach to optimize the excited state wavefunctions. The CI calculations reported here were undertaken in order to evaluate J and K as a function of chain length; to highlight the divergence between results obtained from computations which allowed no relaxation compared with those that allowed electronic relaxation; and to understand the AM1 results from MOPAC calculations.

Our use of one electron CI calculations, which maintain frozen orbitals in both excited states and allow only the two unpaired electrons to interact, extrapolate to the same result as does the HOMO-LUMO gap and clearly are not adequate to reproduce experimental results. Any method that does not include electronic relaxation of the excited state wavefunctions will also be inadequate.

Figure 1 shows that the minimum in the plot of $1\ ^1A_g \longrightarrow 1\ ^3B_u$ versus 1/n occurs for shorter oligomers than does the minimum in the corresponding plot for $1\ ^1A_g \longrightarrow 1\ ^1B_u$ It is not clear if there is a physical reason for this earlier minimum; but there surely is a mathematical reason. If the derivatives with respect to 1/n of equations 1 and 2, the expressions for ΔE as a function of chain length, are set equal to zero in order to find the minima in the curves, the following expressions result.

$$d(^1E - E_0) / d(1/n) = d(\epsilon_L - \epsilon_H) / d(1/n) - dJ/ d(1/n) + 2[dK / d(1/n)] \quad (4)$$

$$d(^3E - E_0) / d(1/n) = d(\epsilon_L - \epsilon_H) / d(1/n) - dJ/ d(1/n) \quad (5)$$

When these derivatives are set equal to zero the result is:

$$dJ/ d(1/n) = d(\epsilon_L - \epsilon_H) / d(1/n) + 2[dK / d(1/n)] \quad (6)$$

$$dJ/ d(1/n) = d(\epsilon_L - \epsilon_H) / d(1/n) \quad (7)$$

Figure 2 shows that dJ/ d(1/n) gets more positive as the chain lengthens. When it reaches the magnitude of $d(\epsilon_L - \epsilon_H) / d(1/n)$, which, as seen in Figure 2, is relatively constant, the minimum occurs for the triplet state plot. In order for the minimum in the singlet state plot to occur, dJ/ d(1/n) must further increase until it equals the same $d(\epsilon_L - \epsilon_H) / d(1/n)$ term plus 2[dK / d(1/n)]. This will always occur at longer chain length.

The semiempirical AM1 data are plotted vs. 1/n in Figure 3. The significant features of the plot are: 1) $1\ ^1A_g \longrightarrow 1\ ^3B_u$ does not approach the same value as does the HOMO - LUMO gap. This stands in contrast with the frozen orbital calculations which did not employ ROHF. The conclusion is that electronic relaxation in the excited state must be considered if accurate band gaps are to be computed; 2) $1\ ^1A_g \longrightarrow 1\ ^3B_u$ approaches 1.69 eV for PA. This result is in excellent agreement with the experimental value.

Conclusion

The *ab initio* calculations done in the course of this work were quite lengthy and required considerable disk space. In fact, the C_{16} to C_{20} oligomers could not be done locally due to size considerations. It would be a great advantage if this type of calculation could be done semi-empirically. The length of the calculations and the size of the output files were not a problem for AM1 calculations on oligomers up to $C_{40}H_{42}$. Our results indicate that AM1 level calculations on excited states at the ROHF level do reproduce experimental results accurately. Polymers with repeat units more complex than the simple CH involved in the polyenes will benefit even more by the application of semi-empirical methods.

References

1. H. Shirakawa, E. J. Lewis, A. G. MacDiarmid, C. K. Chiang, and A. J. Heeger, *J. Chem. Soc. Chem. Commun., 1977* (1977) 578.
2. C. K. Chiang, C. R. Fincher, Jr., W. Park, A. J. Heeger, H. Shirakawa, E. J. Lewis, S. C. Gau, and A. G. MacDiarmid, *Phys. Rev. Lett., 39* (1977) 1098.
3. J. R. Ellis in *Handbook of Conducting Polymers*, Marcel Dekker, New York, 1986, p. 489-505.
4. A. G. MacDiarmid and R. B. Kaner in *Handbook of Conducting Polymers*, Marcel Dekker, New York, 1986, p. 689-727.
5. L. Alcacer, editor, *Conducting Polymers: Special Applications*, D. Reidel, Dordrecht, 1987.
6. A. G. MacDiarmid and A. Epstein, *J. Chem. Soc., Faraday Trans., 5* (1989) 1.
7. J. R. Reynolds, *Chemtech, 18* (1988) 440.
8. M. G. Kanatzidis, *Chemical and Engineering News*, December 3, 1990 36.
9. D. S. Dudis and A. T. Yeates, In preparation.
10. T. C. Koopmans, *Physika, 1* (1933) 104.
11. J. J. Ladik, *Quantum Theory of Polymers as Solids*, Plenum Press, New York, 1988, p. 211.
12. F. L. Pilar, *Elementary Quantum Chemistry*, McGraw-Hill, New York, 1968, p. 366.
13. Ref. 11, p. 271.
14. J. Avery, J. Packer, J. Ladik, and G. Biczo, J. Mol. Spec., *29 (1969)* 194.
15. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill, New York, 1989, p. 86.
16. D. S. Dudis, *Synth. Met., 49* (1992) 353.
17. Typically to get an excited singlet energy, the semi-empirical methods use the 'half electron' (or 'partial electron') SCF method followed by a CI calculation. Generally, there are very limited CI's and thus neglect much orbital relaxation. Furthermore, it has been shown that diffuse and polarization functions are very important for excited

states, at least fora the shorter mers, so semi-empirical methods are inherently at a disadvantage.

18. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, J. H. Jensen, S. Koseki, M. S. Gordon, K. A. Nguyen, T. L. Windus, and S. T. Elbert, GAMESS, *Quantum Chem. Program Exchange Bull., 10* (1990) 52.

19. P. M. Lahti, J. Obrzut, and F. E. Karasz, *Macromolecules, 20* (1987) 2023.

Table 1. Polyene Oligomer Energies From *ab initio* Calculations Using GAMESS with RHF/6-31G* Geometries

Oligomer	Orbitals	Electrons	E_L	E_H	$1\ ^1A_g$	$1\ ^1B_u$	$1\ ^3B_u$	$\Delta E_{H,L}$	$1\ ^1A_g \rightarrow 1\ ^1B_u$	$1^1Ag \rightarrow 1^3B_u$
C_2H_4	38	16	.1839	-.3744	-78.0317181	-77.6629858	-77.8738900	.5583	.368733	.157828
C_4H_6	72	30	.1335	-.3215	-154.9196540	-154.6281385	-154.7695770	.4550	.29152	.15008
C_6H_8	106	44	.1052	-.2936	-231.8082666	-231.5589055	-231.6664429	.3988	.24936	.14182
C_8H_{10}	140	58	.0879	-.2766	-308.6971128	-308.4728445	-308.5590113	.3645	.22427	.13810
$C_{10}H_{12}$	174	72	.0765	-.2654	-385.5860295	-385.3775813	-385.4486345	.3419	.20845	.13740
$C_{12}H_{14}$	208	86	.0685	-.2574	-462.4749689	-462.2767005	-462.3364977	.3259	.19826	.13847
$C_{14}H_{16}$	242	100	.0628	-.2519	-539.3639163	-539.1722318	-539.2233806	.3147	.19168	.14054
$C_{16}H_{18}$	276	114	.0584	-.2477	-616.2528666			.3061		
$C_{18}H_{20}$	310	128	.0551	-.2444	-693.1418176			.2995		
$C_{20}H_{22}$	344	142	.0525	-.2419	-770.0307688			.2944		

Table 2. Values of the Coulomb and Exchange Integrals From *ab initio* Calculations on Polyenes

Polyene	Coulomb Integral, J	Exchange Integral, K
C_2H_4	0.4005	0.1055
C_4H_6	0.3049	0.0707
C_6H_8	0.2570	0.0538
C_8H_{10}	0.2264	0.0431
$C_{10}H_{12}$	0.2045	0.0355
$C_{12}H_{14}$	0.1874	0.0299
$C_{14}H_{16}$	0.1742	0.0256
$C_{16}H_{18}$		

Table 3. Polyene Oligomer Energies From AM1 Calculations Using GAMESS

Oligomer	Electrons	E_L	E_H	$1\ ^1A_g$	$1\ ^3B_u$	$\Delta E_{H,L}$	$1\ ^1Ag \rightarrow 1\ ^3B_u$
C_2H_4	16	+.0528	-.3878	-11.4054449	-11.3056119	.4406	.0998330
C_4H_6	30	+.0165	-.3430	-21.8120466	-21.7299995	.3595	.0820471
C_6H_8	44	-.0031	-.3216	-32.2193636	-32.1461611	.3185	.0732025
C_8H_{10}	58	-.0149	-.3094	-42.6268274	-42.5615450	.2945	.0652824
$C_{10}H_{12}$	72	-.0226	-.3019	-53.03432252	-52.9683950	.2793	.0659275
$C_{12}H_{14}$	86	-.0280	-.2969	-63.4418244	-63.3794628	.2689	.0623616
$C_{14}H_{16}$	100	-.0318	-.2934	-73.8493269	-73.7854437	.2616	.0638832
$C_{16}H_{18}$	114	-.0346	-.2909	-84.2568289	-84.1950930	.2563	.0617359
$C_{18}H_{20}$	128	-.0367	-.2890	-94.6643315	-94.6008921	.2523	.0634394
$C_{20}H_{22}$	142	-.0384	-.2876	-105.0718336	-105.0102459	.2492	.0615877
$C_{22}H_{24}$	156	-.0397	-.2865	-115.4793334	-115.4159999	.2468	.0633335
$C_{24}H_{26}$	170	-.0407	-.2856	-125.8868236	-125.8252205	.2449	.0616031
$C_{26}H_{28}$	184	-.0417	-.2849	-136.2943361	-136.2310413	.2432	.0632948
$C_{28}H_{30}$	198	-.0424	-.2843	-146.7018380	-146.6402928	.2419	.0615452
$C_{30}H_{32}$	212	-.0430	-.2839	-157.1093389	-157.0460535	.2409	.0632854
$C_{32}H_{34}$	226	-.0435	-.2835	-167.5168405	-167.4552842	.2400	.0615563
$C_{34}H_{36}$	240	-.0440	-.2832	-177.9243426	-177.8610475	..2392	.0632951
$C_{36}H_{38}$	254	-.0444	-.2829	-188.3318429	-188.2703024	.2385	.0615405
$C_{38}H_{40}$	270	-.0447	-.2826	-198.7393441	-198.6760348	.2379	.0633093
$C_{40}H_{42}$	284	-.0450	-.2825	-209.1468449		.2375	

Figure 1. Ab initio Energy Differences vs. Chain Length

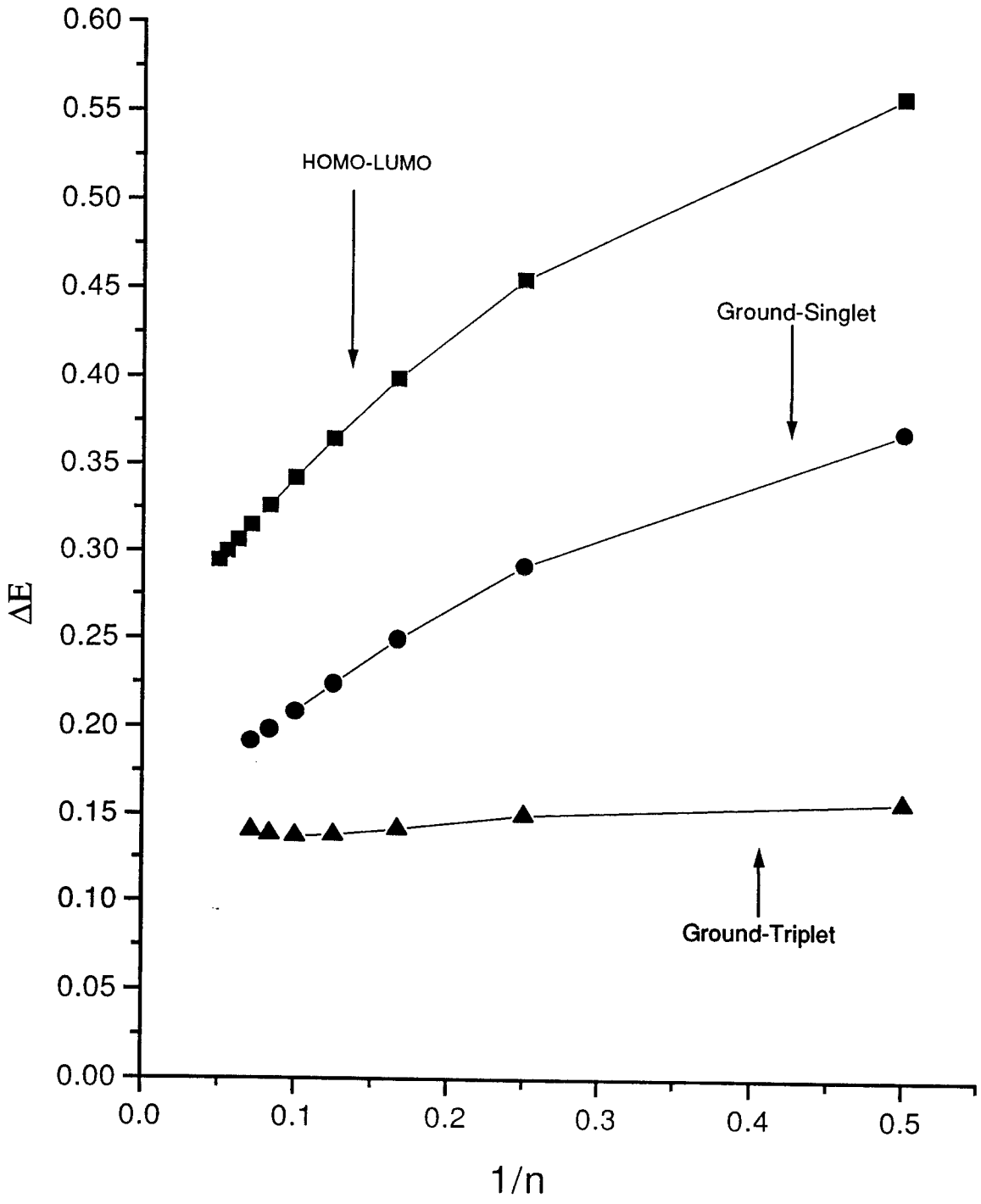


Figure 2. Coulomb and Exchange Integrals vs. Chain Length

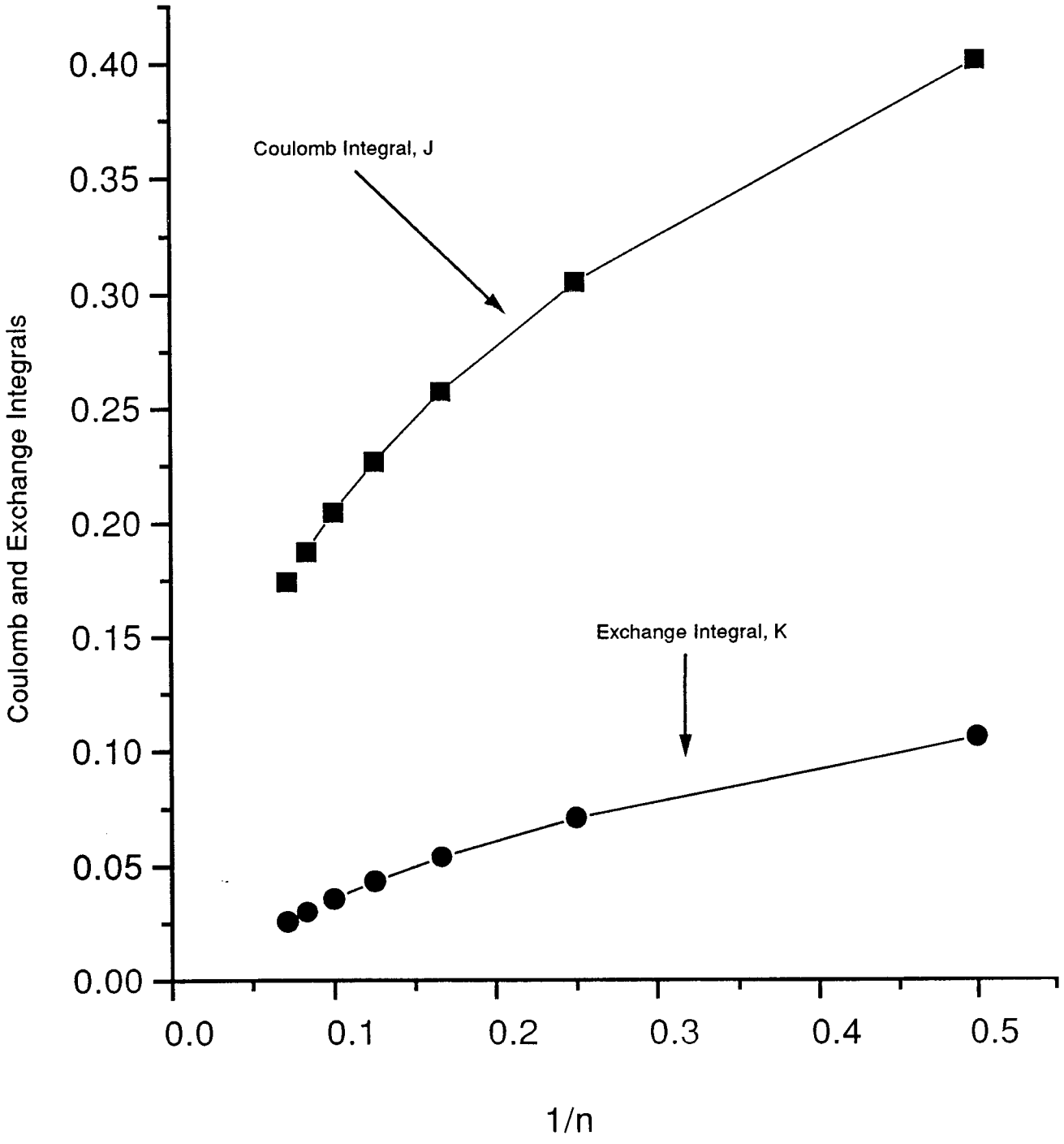
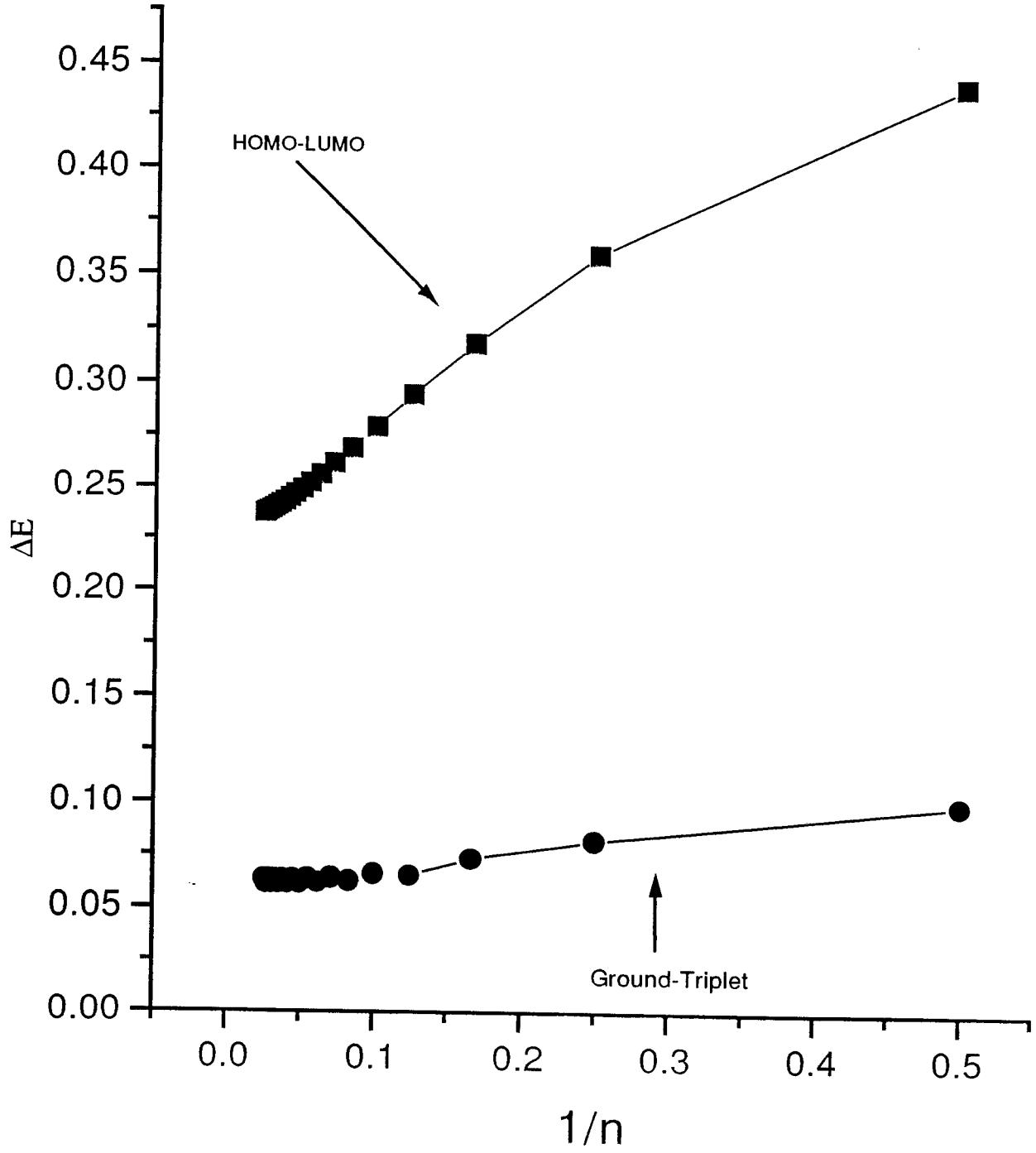


Figure 3. AM1 Energy Differences vs. Chain Length



PERFORMANCE OF MUSIC AND MONOPULSE ALGORITHMS IN THE PRESENCE OF RADOME REFLECTIONS AND MUTUAL COUPLING

Krishna M. Pasala
Associate Professor
Department of Electrical Engineering

University of Dayton
300 College Park
Dayton, Ohio 45419-0226

Final Report for:
Summer Faculty Research Program
WL/AARM - 3

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

and

WL/AARM - 3

August 1994

PERFORMANCE OF MUSIC AND MONOPULSE ALGORITHMS IN THE PRESENCE OF RADOME REFLECTIONS AND MUTUAL COUPLING

Krishna Murthy Pasala
Associate Professor
Department of Electrical Engineering
University of Dayton

Abstract

In this report we investigate the behavior of two direction finding algorithms under actual operating conditions rather than ideal conditions. These two algorithms are: 1) the Iterative Monopulse algorithm and 2) the MUSIC algorithm. The Iterative Monopulse algorithm is simple and efficient to use and does not demand excessive computational capacity but has a resolution dictated by the Raleigh criterion. When the spectrum is sparsely populated this algorithm gives accurate answers in an efficient fashion. The MUSIC algorithm is capable of super resolution beyond that dictated by the Raleigh criterion. The performance of these two algorithms is studied under realistic conditions wherein the effects of reflections from radome walls and the mutual coupling between elements of the array are included. The input data for this study is obtained from two sources. An electromagnetic code simulating the radome is used to generate the fields inside the radome. This data includes the effects of radome scattering on the wavefront distortion inside the radome but does not include the effects of mutual coupling. The second set of data is obtained from experiments carried out in a compact range on 16-horn array. Simulation results clearly indicate the degradation in the performance of both the algorithms. Also, the MUSIC algorithm is much more vulnerable to non-ideal conditions than is the Iterative Monopulse algorithm. It has been shown that the effects of mutual couplings can be compensated for using the mutual impedance matrix which is aspect insensitive. It remains a challenging problem to compensate for a near field scatterer in general and the radome in particular since these scattered fields are aspect dependent. Several lines of further research are suggested to overcome these problems.

PERFORMANCE OF MUSIC AND MONOPULSE ALGORITHMS IN THE PRESENCE OF RADOME AND MUTUAL COUPLINGS

Krishna Murthy Pasala

1. INTRODUCTION

The goal of the research presented here is to examine the performance of signal processing algorithms under actual operating conditions. In particular, we consider here various algorithms used for bearing estimation [1],[2]. All too often performance of these algorithms is assessed under ideal conditions. The sensors such as antennas are assumed to be perfect. The array of these sensors are assumed to operate with no mutual couplings present. In practice, antenna arrays are protected by a radome. The radome is assumed to be perfectly transparent but , of course, it is not. Also, the antenna array may be operating in the vicinity of a significant scatterer which may affect the performance of the algorithm.

Burks et al [3] have dealt with the problem of radar pointing error induced by radome. They used a ray based technique to account for the scattering from the radome and considered its effect on a monopulse radar. Temple and Pyati [4] have investigated the boresight error using a system concept and compared their results with experimental results. Friedlander and Weiss [5] have considered the problem of direction finding in the presence of mutual coupling. They describe an iterative technique which yields successively better estimates of the mutual coupling matrix and the direction of arrival. Yeh et al [6] and Litva and Zeytinoglu [7] also investigate the performance of the MUSIC algorithm [8] in conjunction with diplole arrays. Recently Pasala and Friel [9] have studied the performance of the MUSIC algorithm over a wide band of frequencies with array of dipoles, sleeve dipoles and spiral antennas. They also presented a way for correcting the antenna output voltages to compensate for the mutual coupling effects and restore the performance of the algorithm to near ideal.

In this report we consider the following problem: In the vicinity of the antenna array a scatterer is present which distorts the ideal wave front. In addition, the mutual coupling between the elements of the array also contribute to the distribution of the wave front. Under such " actual " conditions how is the

performance of a variety of spectrum estimation algorithms affected? We consider here three different spectrum estimation algorithms. These are: 1) the conventional beamformer [9] 2) an iterative monopulse method and 3) the MUSIC method. The data to be processed by these algorithms is obtained from two different sources. The first source is a radome simulation program which computes the fields inside a radome for any specified incident plane wave. This program accounts for the scatterer, which in this case is the radome, but does not account for the mutual coupling in the array. The second source is experimental data provided by WL/AARM - 3. This data is actually measured with a 16 horn array with and without the radome. This data accounts both for the scatterer, viz., the radome and the mutual coupling in the array. In a companion investigation, E. M. Friel [10] investigated the performance of the MUSIC algorithm in the presence of an edge type scatterer.

The investigations reported here clearly demonstrate the serious and deleterious effects of the scatterers in the vicinity of the array of sensors. We show that the boresight error may be reduced to virtually zero using the iterative monopulse algorithm in the ideal case. But the same algorithm results in significant error when using "actual data". We also show that the performance of a high resolution algorithm such as "MUSIC" degrades to such an extent, it is no better than the beamformer.

2. FORMULATION OF THE PROBLEM

Consider an array of antennas illuminated by a number of plane waves as shown in the figure - 2.1. such as i) an aircraft wing, ii) a building, iii) a radome etc. While the details of the scattering mechanisms could be different, the qualitative effect on any signal processing algorithm is essentially the same: the ideal signal model assumed by the algorithm ceases to be valid.

Desired information is extracted by processing the outputs of the antenna elements in the array. The antenna array may be modeled as a multiport network and the output currents may be obtained from

$$Z I = V. \qquad (1)$$

The (N x N) Z matrix is called the impedance matrix and represents the coupling between the elements of the array. In the ideal case, the coupling is deemed to be absent and the impedance matrix reduces to an identity matrix. That is, no element in the antenna is influenced by the presence of other elements in the

array. In any practical situation, the off-diagonal elements in the Z - matrix are significant. The excitation matrix V depends upon the incident fields at the antenna locations. If there were to be no mutual coupling,

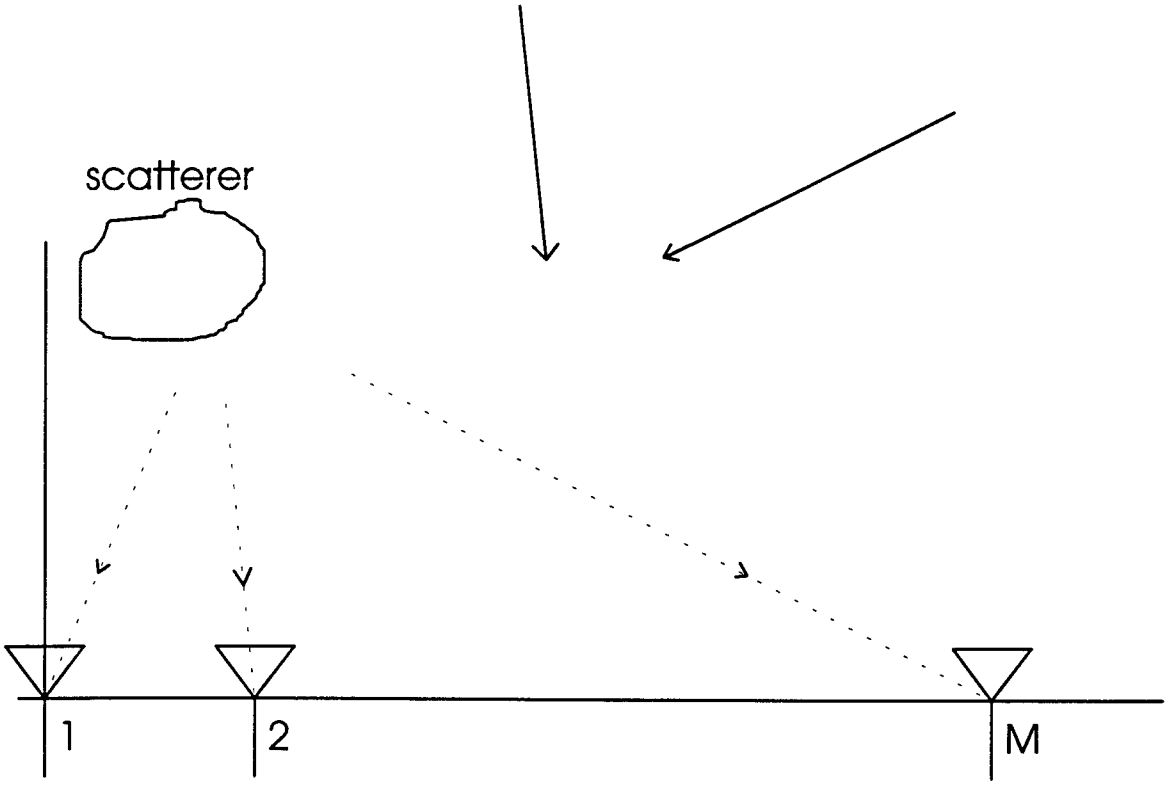


Figure 2.1. Geometry of the Array and the scatterer.

the output currents are directly proportional to the incident fields. In the presence of the scatterer, equation - 1 remains the same qualitatively but does become more complex as given below.

$$Z_1 I_1 = V_1 \tag{2}$$

where,

$$Z_1 = Z + Z_S, \tag{3}$$

$$V_1 = V + V_S, \tag{4}$$

and $$I_1 = I + I_S. \tag{5}$$

The N x N matrix Z_S accounts for the coupling between the antennas through the scatterer. The N x 1 vector V_S accounts for the excitation of the arrays resulting from the field at the antenna elements as a result of the scattering from the scatterer. This field is present in addition to the direct field. The N x 1 vector I_S is the change in antenna currents due to the presence of the scatterer. There is some evidence to indicate that the change in currents due to Z_S is negligible compare to the change in current due to V_S.

The currents I_1 obtained from solving equation - 2 are the "actual" currents and must be processed to obtain the desired information. For the purpose of this report, this "actual" data is obtained from two different sources. In the radome program, the total incident field (i.e. direct plus scattered field) is computed, but the antenna array is assumed to be ideal, i.e. Z_1 is assumed to be an identity matrix. The experimental data used here, however, accounts for both Z_S and V_S.

3. ALGORITHMS FOR BEARING ESTIMATION

We discuss here, briefly, the algorithms used for direction finding. The beamformer is the simplest and the monopulse method greatly improves the accuracy of the bearing estimation and the MUSIC algorithm is the most complicated and is capable of "super resolution".

3.1 Beamformer:

Let $w = [\, w_1, w_2, \ldots, w_N]^T$ be the weight vector and $s = [\, s_1, s_2, \ldots, s_N\,]^T$ be the signal vector. The weight vector W may be chosen to steer the beam in a given direction. The n^{th} weight is given by,

$$W_n = e^{[-j\beta(x_n \cos(\phi^k) + y_n \sin(\phi^k))]} \qquad (6)$$

where ϕ^k is the k^{th} steering angle. The signal at the n^{th} element due to a plane wave with an incident angle of ϕ_i is given by,

$$S_n = e^{[j\beta(x_n \cos(\phi_i) + y_n \sin(\phi_i))]} \qquad (7)$$

The array output is given by $W^T S$. The beamformer output is conveniently determined by forming the weight matrix [w], $[w] = [\, w^1\ w^2 \ \ldots\ w^K]$.

Each column of this matrix is the weight vector corresponding to the k^{th} steering angle. The beamformer output is then given by $[w]^T S$. The estimate of the direction of arrival is taken to correspond to the peak of the beamformer output. This is a simple and robust method but its resolution is limited to that dictated by Raleigh Criterion [2].

3.2 Iterative Monopulse:

Excellent accuracy and numerical efficiency may be achieved using the monopulse method. The direction of arrival falls between two beam peaks as shown (refer to figure 3.1).

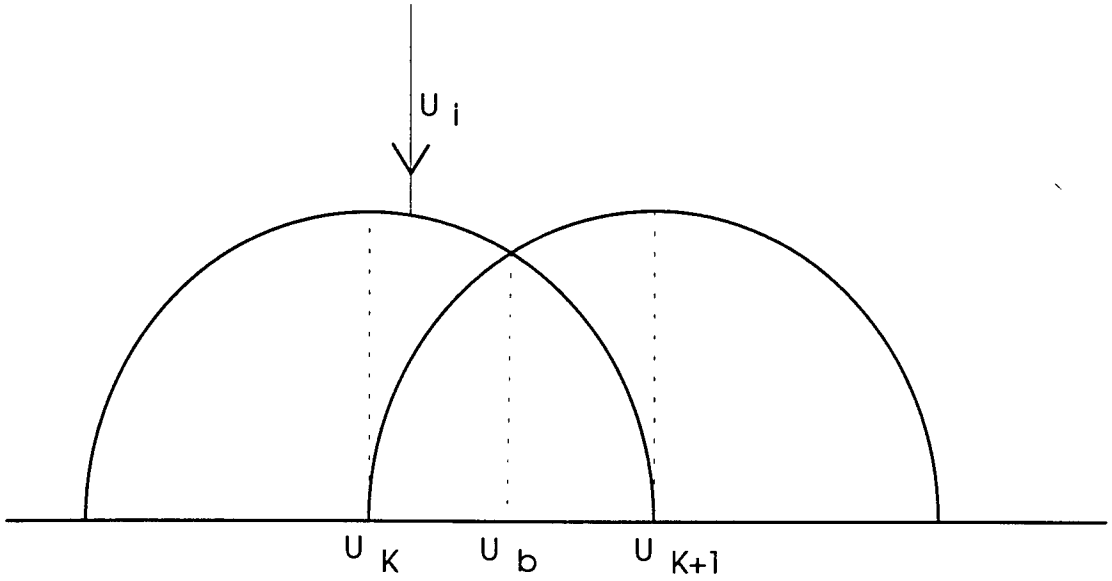


Figure 3.1. Geometry for Monopulse Processing.

The beams are separated in the U - space by $\frac{\pi}{N}$. The sum and difference beams are formed and the ratio of the outputs in difference and sum beams is denoted by r. Then, the estimate of the true angle in U - space, $\hat{U}_i$, is given by,

$$\hat{U}_i = U_b - \frac{\pi}{2N} r. \tag{8}$$

The closer U_i is to U_b, the more accurate this equation is. We can further improve the accuracy by creating two more beams such that the new boresight equals $\hat{U}_i$. That is,

$$U_b = \hat{U}_i \tag{9}$$

$$U_k = U_b - \frac{\pi}{2N} \quad \text{and} \quad U_{k+1} = U_b + \frac{\pi}{2N}. \tag{10}$$

Sum and difference beams are formed again and using equation - (8) above a better estimate is obtained and this iteration may be continued until there is no further improvement in the estimate. In practice, convergence is obtained in about three iterations.

3.3 MUSIC Algorithm:

The MUSIC algorithm is a sub-space technique based on eigendecomposition of the correlation matrix. It consists of the following steps:

i) Using the antenna outputs to estimate the correlation matrix, R_X.

ii) Carry out the eigendecomposition of R_X and separate the noise and signal eigenvalues and hence noise subspace.

iii) Then, determine the pseudo spectrum, $P_{music}(\theta)$ as follows:

$$P(\theta) = \frac{1}{\sum_{i=k+1}^{N} \left| e_i^H \, a(\theta) \right|^2} \tag{11}$$

where,

$$a(\theta) = [\, e^{-j\beta d_1 \sin(\theta)}, \ldots, e^{-j\beta d_N \sin(\theta)} \,]^T. \tag{12}$$

Since the search vector a(θ) corresponding to the direction of arrival is orthogonal to the noise sub-space, P(θ) exhibits a sharp peak when $\theta = \theta_i$.

4. RESULTS

In this section the algorithms described in section-3 are applied to simulated and experimental data. Before describing the various numerical simulations carried out, we first describe, briefly, the two sets of data.

1.Experimental data: The experimental data is obtained in a compact range with a 16-horn array. Several configurations of radomes are used, though only one set of data is used for simulations carried out in this report. For the case that is relevant here, the array is located in the middle of the radome where the reflections from the radome walls have the least effect. Also, the frequency used is 9.8 GHz, which corresponds to the center frequency of operation of the radome. The reflection coefficient is least for this frequency. Thus, the data used here does include the scattering from radome walls but this scattering could and would be a lot worse for frequencies at the edge of the band and for array elements located in the proximity of the wall. However, the data clearly includes the effect of mutual coupling.

2. Radome simulation data: This data is obtained by modeling a dielectric radome. This program computes the total field present at any specified location inside the radome for any specified angle of

incidence. Geometrical optics is the technique employed. The transmission coefficient is assumed to be unity and for reflections from the wall only one bounce is accounted for (as shown in figure 4.0).

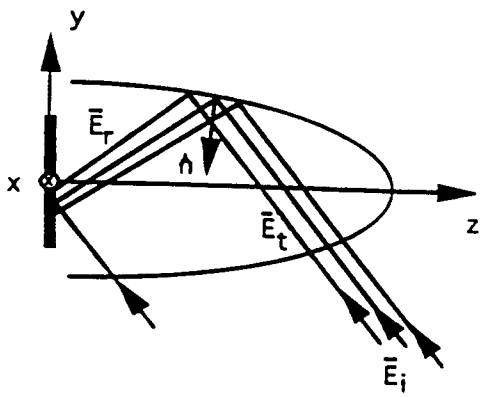


Figure 4.0 Geometry of the Reflections From the Radome.

The data generated by this program gives an estimate of the scattering from the radome but the antenna elements themselves are assumed to be ideal and the mutual coupling is ignored. The frequency of operation is varied at will as also the dielectric constant of the radome. The radome geometry is pre-set.

Simulation-1: Monopulse Processing.

In this simulation a 10-element array is used to determine the angle of arrival. Results were computed for both the ideal case and for the case where both the radome and mutual coupling are present. The actual array on which measurements were made is a 16-element array. We have chosen the middle 10 element sub-array from this array. Figures 4.1 a-f show the progression of results in the ideal case. Indeed, the iterative monopulse processing gives excellent results with zero error. That is, the estimated angle is the true angle.

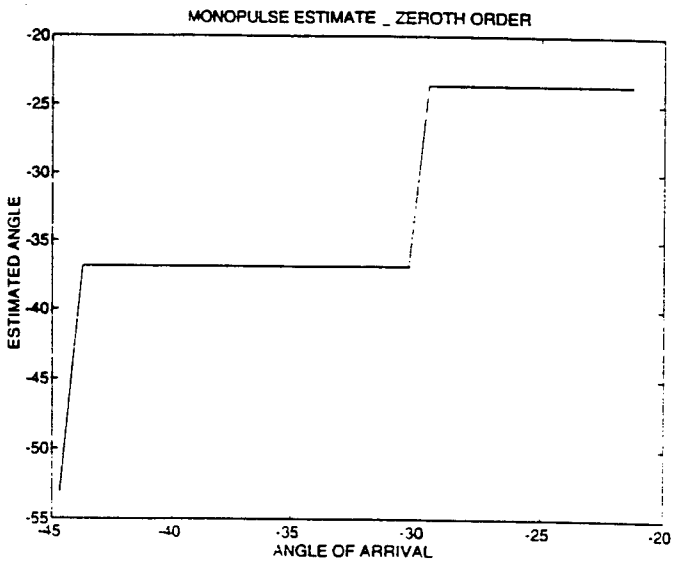


Figure 4.1 a

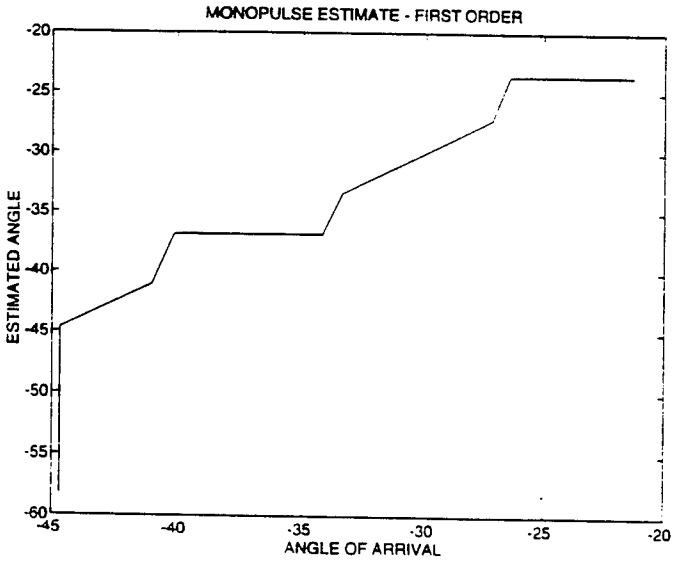


Figure 4.1b

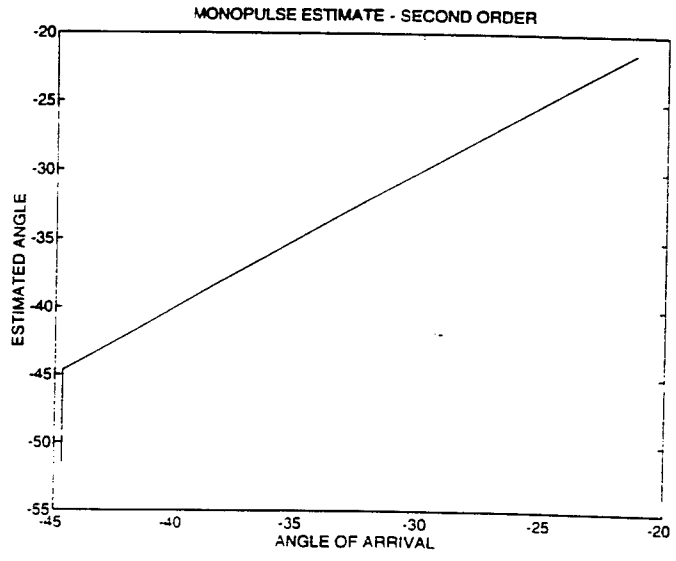


Figure 4.1c

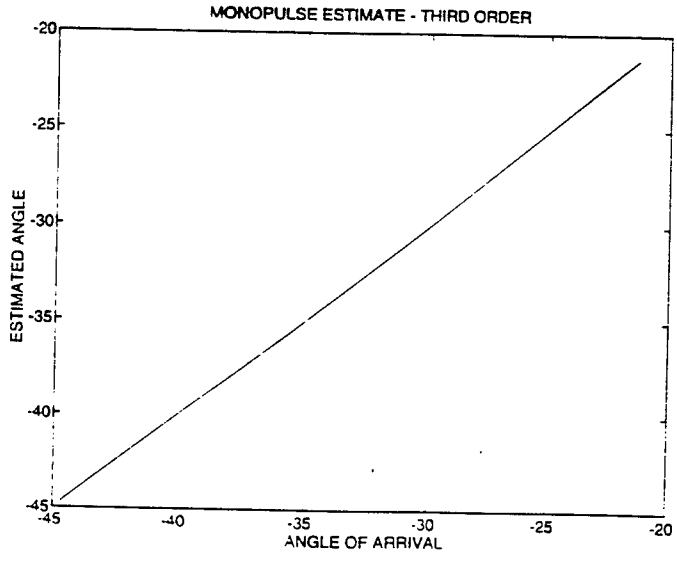


Figure 4.1d

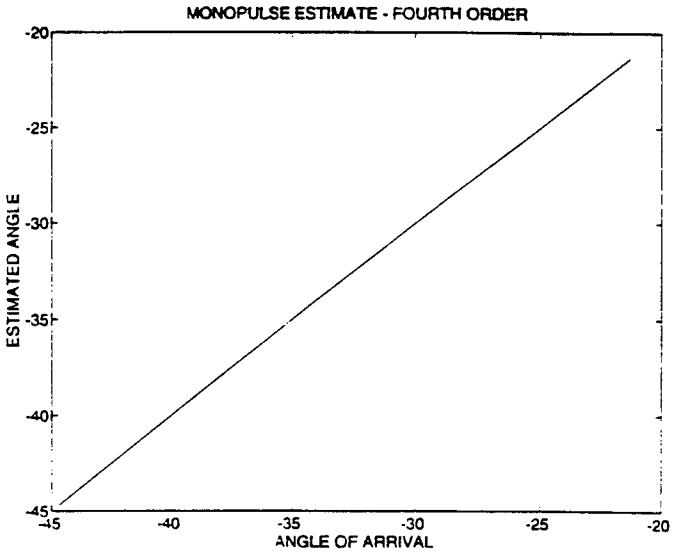


Figure 4.1e

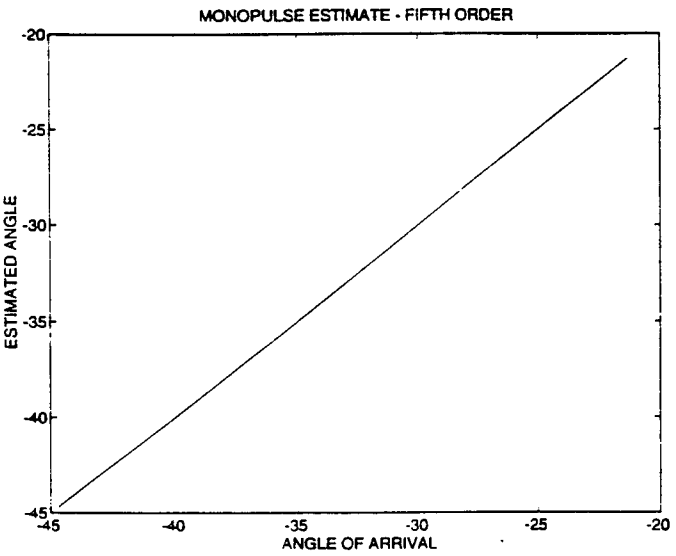


Figure 4.1f

The second part of the simulation is carried out using the actual measured data. Again, the progress of the iteration is shown figures 4.2 a-f. There is clear improvement in the bearing estimate as the order of iteration is increased. However, unlike in the ideal case, the error does not decrease to zero but settles to a small value. The error in bearing estimate for both ideal and actual cases is shown in figure 4.2 f

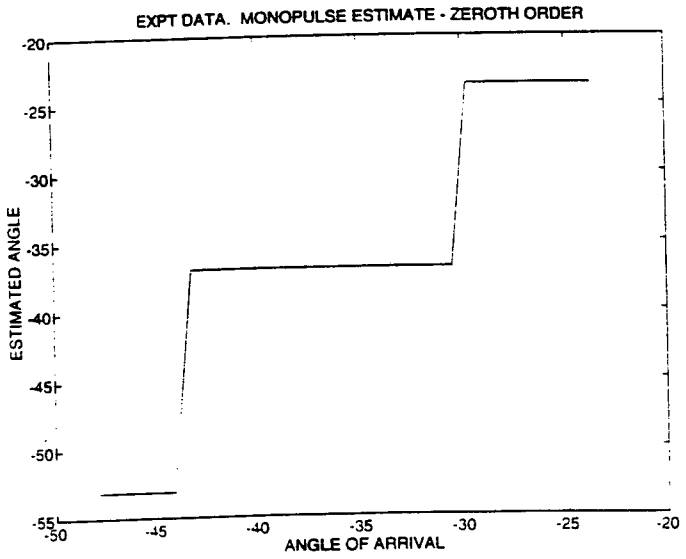


Figure 4.2a

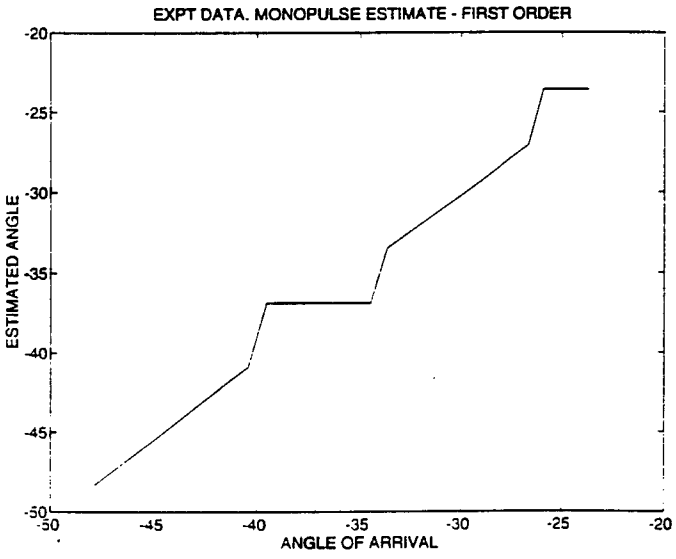


Figure 4.2b

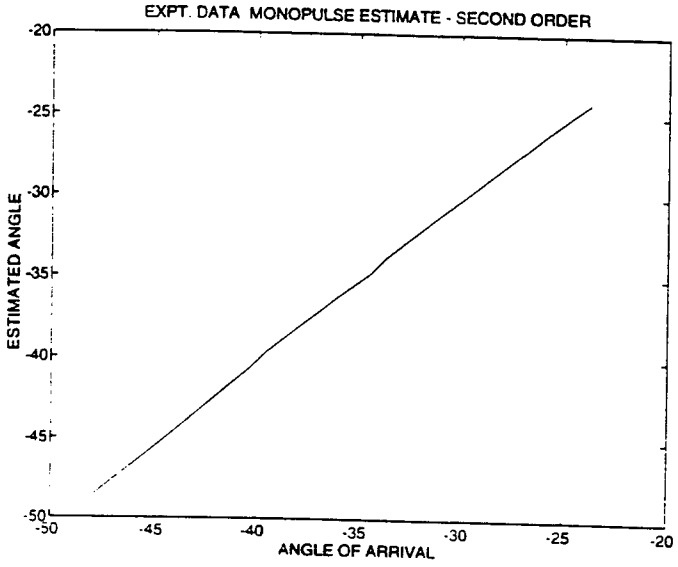


Figure 4.2c

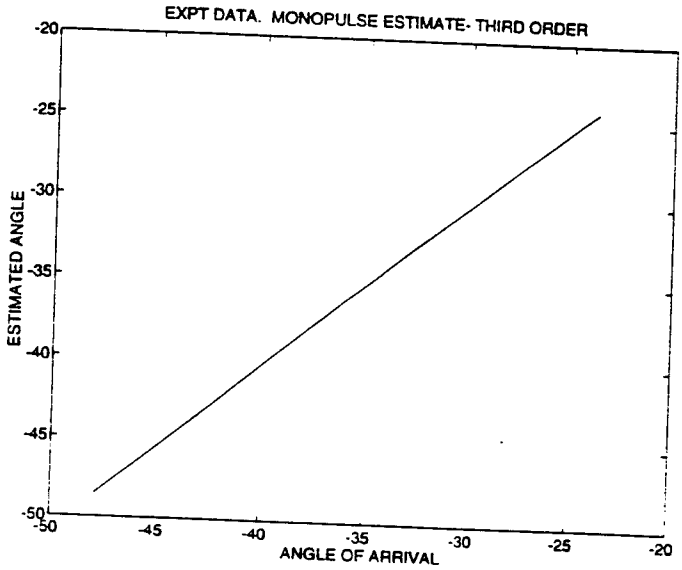


Figure 4.2d

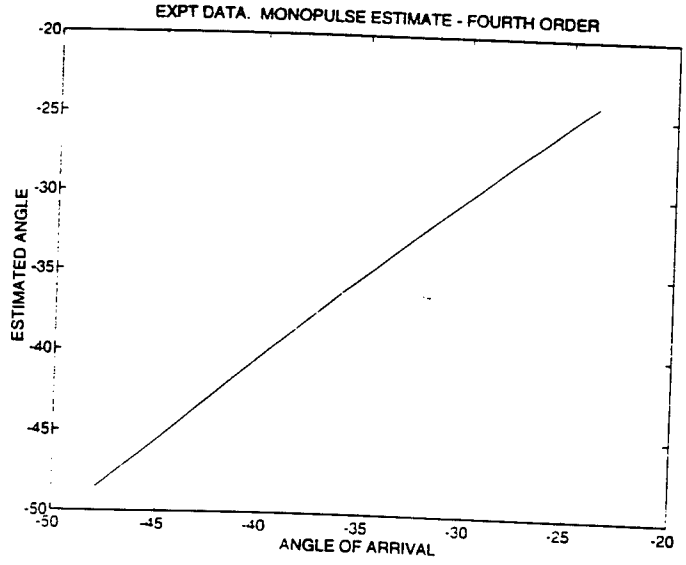


Figure 4.2e

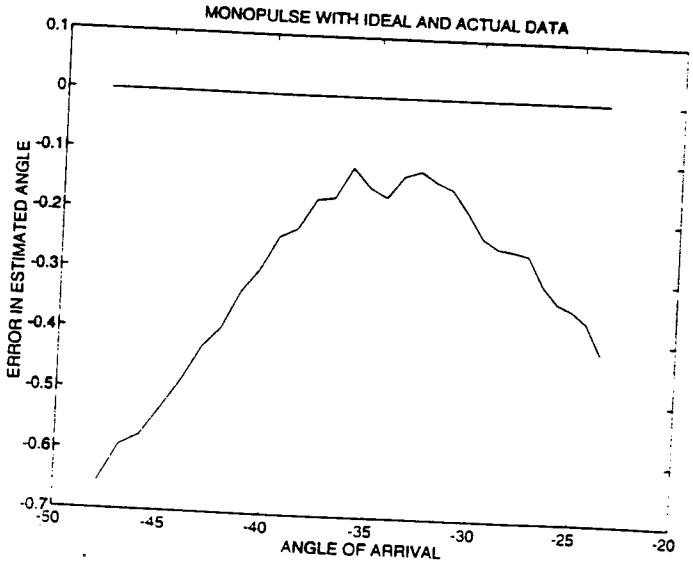


Figure 4.2f

Simulation-2: MUSIC with Radome simulation data.

In this simulation we conducted the following three numerical experiments:

1. The 10-element subarray is considered to be in the "hot spot". For angles of incidence in the range 0-70, the scattered field is strong in the locations occupied by array elements 1-10. The frequency of operation is taken to be 9795 and 9795.5 MHz. Figure 4.4 shows the spectra computed using the MUSIC and the beamformer algorithms. The severe degradation suffered due to the presence of the radome is quite evident. For purposes of comparison, the ideal spectrum is shown in figure 4.3. for the same sources.

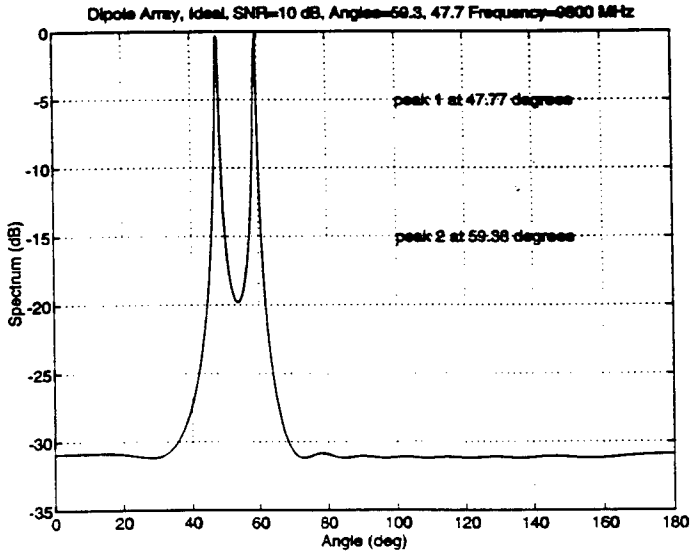


Figure 4.3 Ideal MUSIC Spectrum

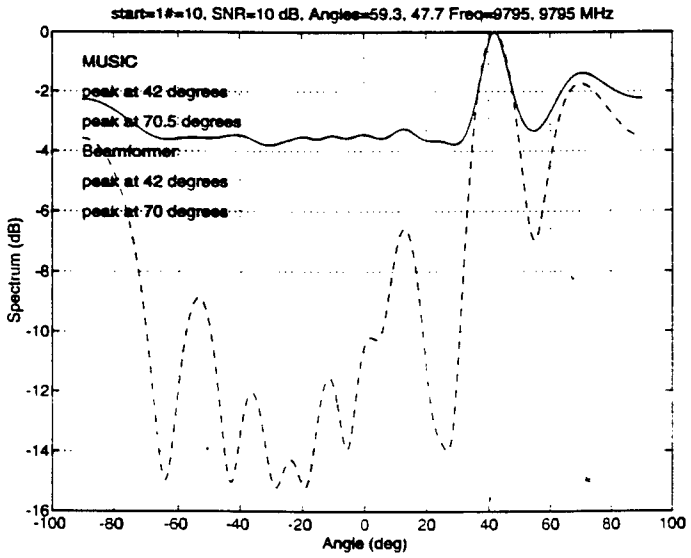


Figure 4.4 Actual MUSIC Spectrum

2. For this case the 10-element subarray is moved to the middle of the radome. The "hot spot" is not nearly as significant here. Nonetheless, the scattered field is not zero and the spectra are still significantly affected. Figure 4.5 shows the spectrum.

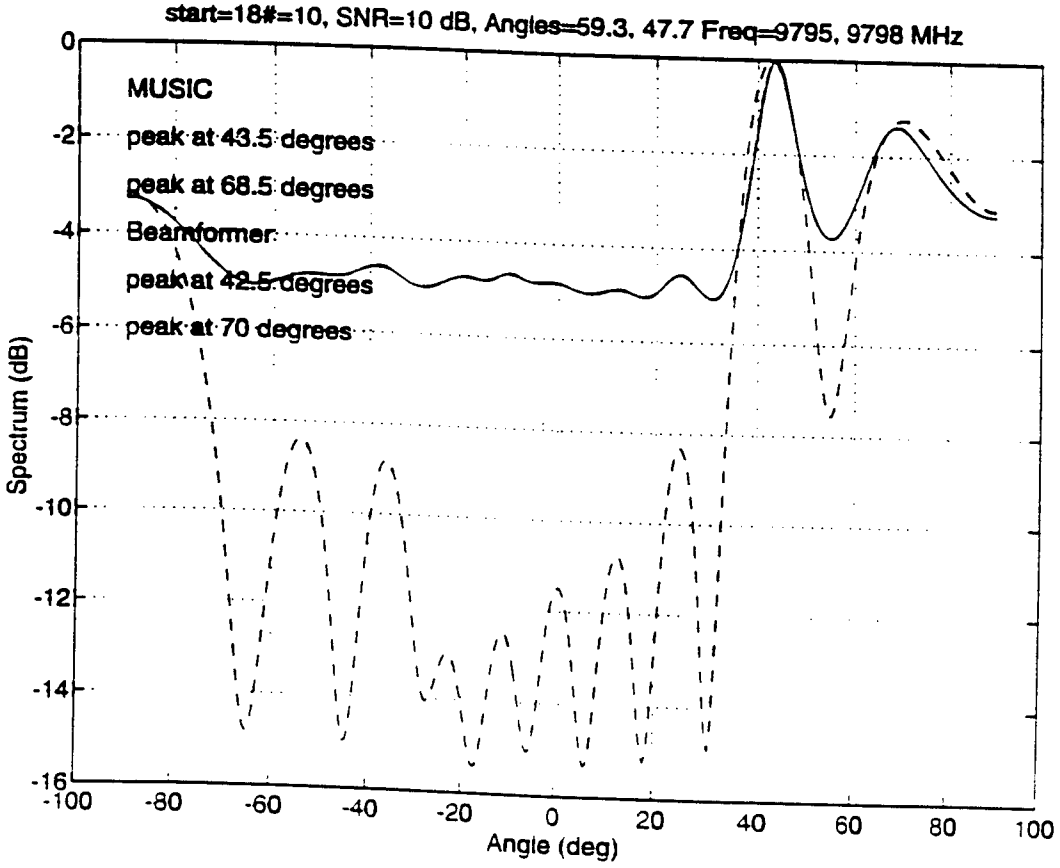


Figure 4.5 Actual Music Spectrum in the Presence of Radome.

3. All the spectra shown so far assume the antenna array to be ideal. Using the fields generated by the radome, we computed the outputs of a 10-element dipole array located in a hot spot as in case-1 above. Using these "actual" output signals we computed the MUSIC and the beamformer spectra. These spectra include the effects of the direct mutual coupling between the elements of the array but not the coupling through the walls of the radome. Figure 4.6 shows these spectra and as may be expected the spectra only get worse.

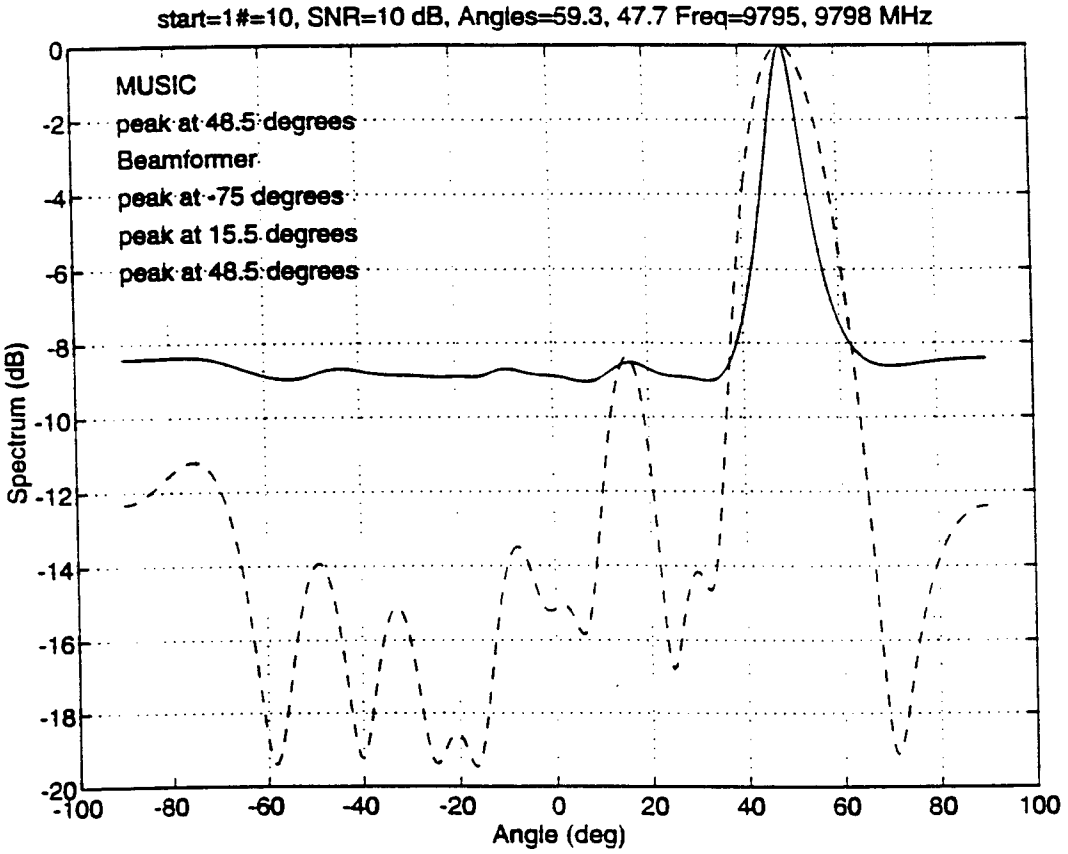


Figure 4.6 Effect of mutual coupling and radome.

Simulation-3: MUSIC spectra with 16-horn experimental data.

In this simulation MUSIC pseudo spectra are computed using the experimental data provided by WL/AARM - 3. These measurements consist of the output signals of a 16- horn array enclosed by a radome. The signal wavefront measured at the horn output, therefore, includes the effects of both the reflections from the radome walls and the mutual coupling between the horns. Figures 4.7-4.10 show the adverse effects of the wavefront distortions on the spectrum. By and large, the spectra fail to resolve the targets that would have been resolved under ideal conditions. These measurements are carried out in the middle of the passband of the radome. The spectra are likely to be worse at the for frequencies of operation at the edge of the passband.

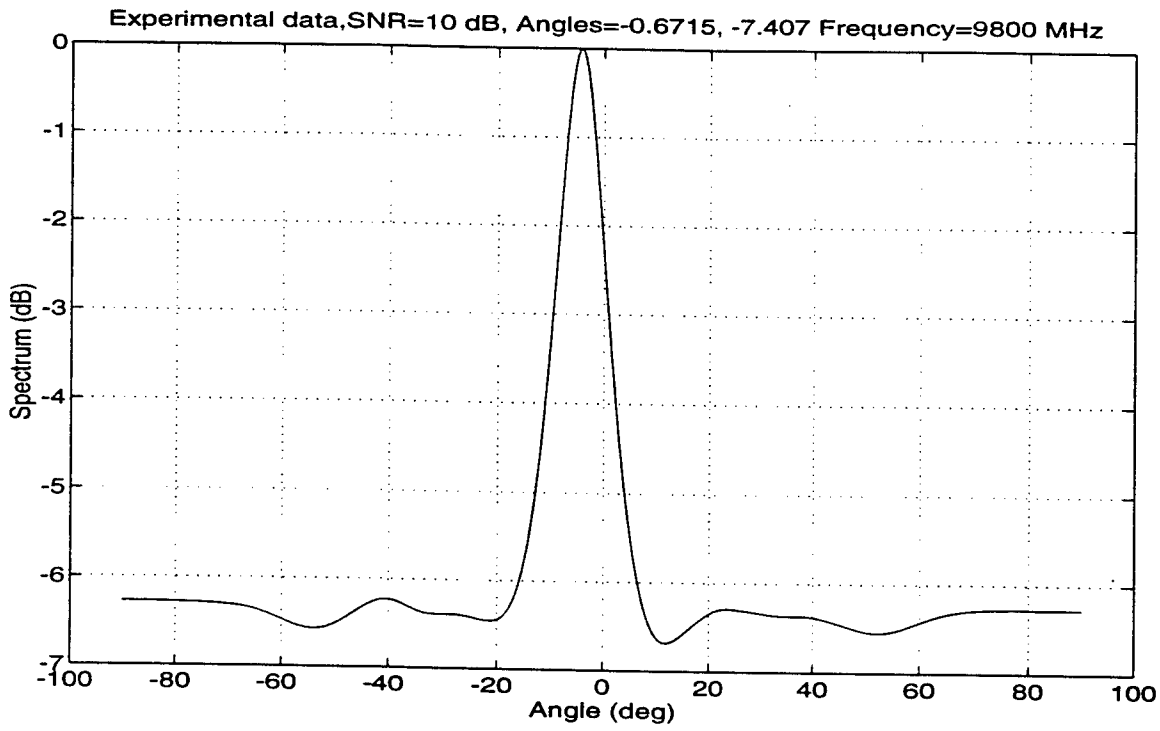


Figure 4.7 MUSIC Spectrum

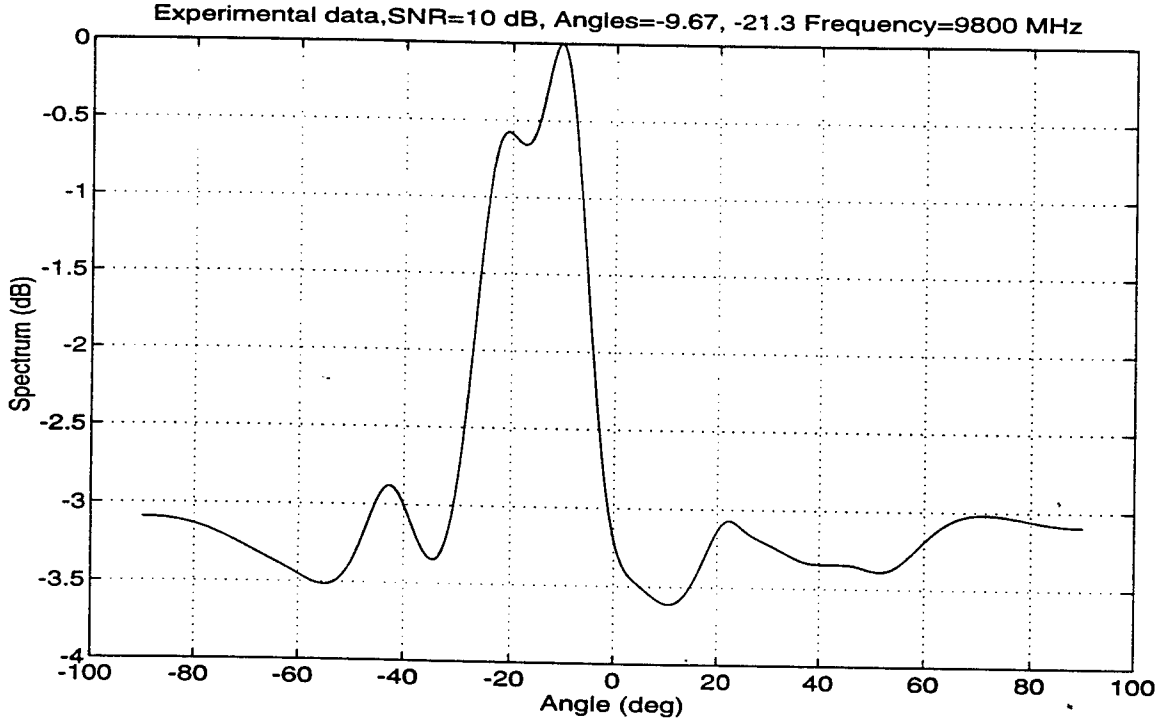


Figure 4.8 MUSIC Spectrum

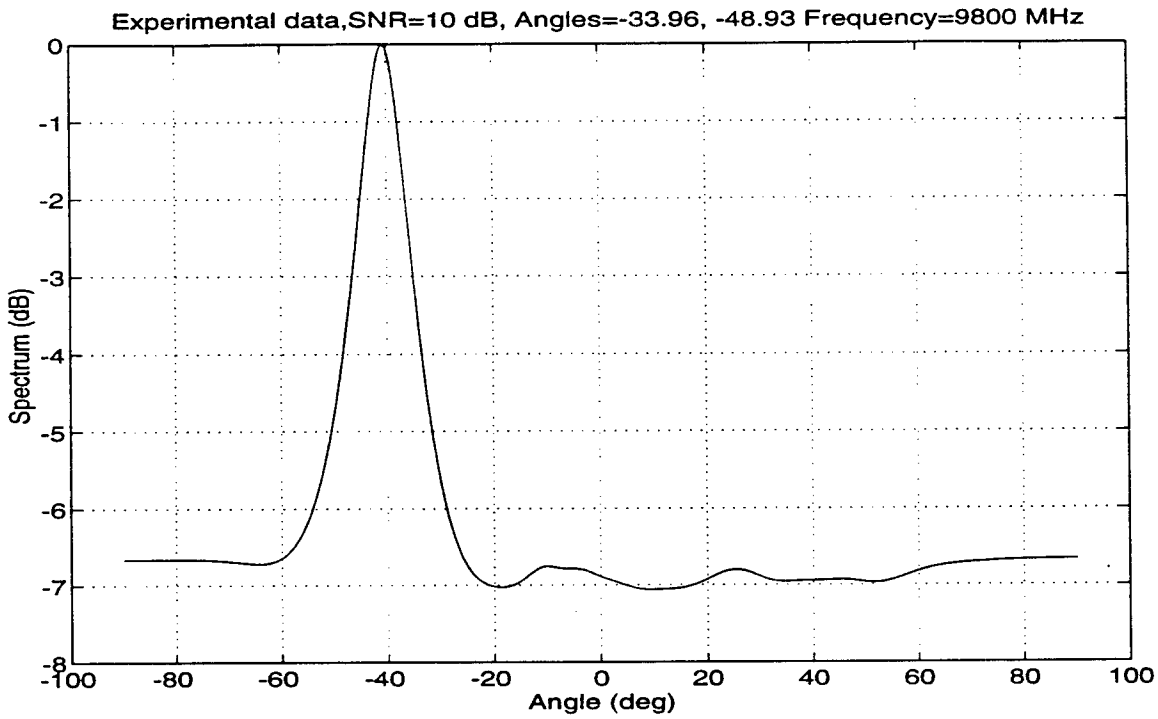


Figure 4.9 MUSIC Spectrum.

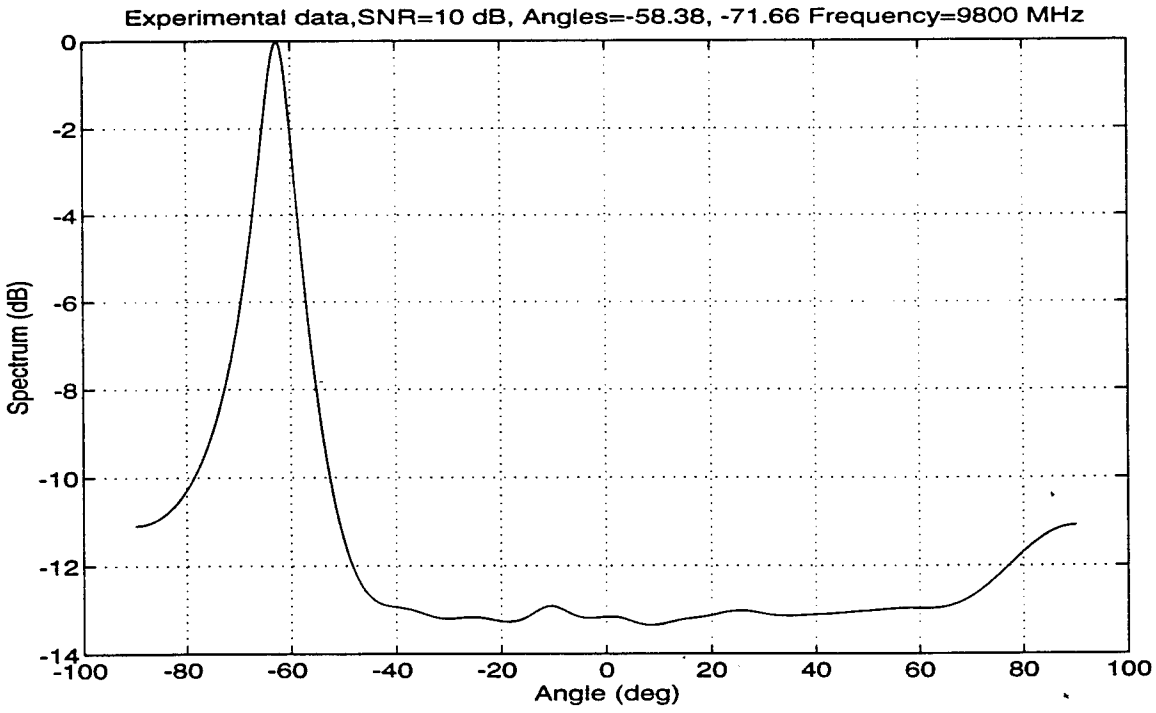


Figure 4.10 MUSIC Spectrum.

5. SUMMARY AND DISCUSSION

In the present effort we investigated the behavior of the Iterative Monopulse and the MUSIC algorithm used for bearing estimation under non-ideal conditions. The antennas in the array are not ideal and hence cause mutual coupling. The array may operate under a radome which is not perfectly transparent. The resulting reflections from the radome walls distort the signal wavefront. The results are quite clear: The performance of the bearing estimation algorithms suffer. The more sophisticated the algorithm is, the less robust it seems to be. MUSIC appears more susceptible than the monopulse algorithm. The study reported here points towards further study along a number of directions.

1) The iterative monopulse is an efficient and robust algorithm and is very useful for the case where the spectrum is sparsely populated. So long as two sources are separated by more than one beamwidth, their bearings are estimated quite accurately and simply using this algorithm. Computations with experimental data suggest a reduction in performance.

2) The MUSIC algorithm is capable of resolving two closely spaced targets operating on a short data record under ideal conditions. Investigations conducted here using both simulated and experimental data clearly establish the degradation in the performance of this algorithm. The degradation in MUSIC algorithm is more severe than that in the monopulse algorithm.

Based on these investigations, the following questions need to be addressed:

1) How much of the degradation in performance is due to the radome reflections and how much is due to the mutual coupling effects? This could be established by processing the 16 horn experimental data obtained with and without the radome.

2) Can the correction procedure developed in [9] to compensate for mutual coupling effects be extended to correct for unwanted radome reflections and in fact any near field scatterer?

3) How can one determine the mutual coupling matrix from the experimental data so that it could be used to correct the data for coupling effects?

6. REFERENCES

1. Pillai, S. U., **Array Signal Processing,** New York: Springer-Verlag, 1988.

2. Johnson, D. H. and Dudgeon, D. E., **Array Signal Processing,** Englewoodcliffs, NJ: Prentice Hall, 1993.

3. Burks, D. G., "A High Frequency Analysis of Radome -Induced Radar Pointing Error, " IEEE Trans.Antennas Propagat., Vol. AP-30, No. 5,September,1982.

4. Temple, A. and Pyati, V. P., "Radome Depolarization and Phase Front Distortion Effects on Boresight Error Prediction," IEEE national Radar Conference, Atlanta, Georgia, March 1994.

5. Friedlander, B. and Weiss, A.J., "Direction Finding in The Presence of Mutual Coupling," **IEEE Trans. Antennas Propagat.,**Vol. AP. 39, NO. 3, March,1991.

6. .Yeh, C. C., Leou, M. L., and Ucci, D. R., " Bearing Estimation with Mutual Coupling Present," **IEEE Trans. Antennas Propagat,** Vol. AP-37, NO. 10, October,1989.

7. Litva, J. and Zeytinoglu, M., "Application of High-Resolution Direction Finding Algorithms to Circular Arrays with Mutual Coupling Present," Final Report (McMaster University) Part II prepared for DREO, July, 1990.

8. Schmidt, R., " Multiple Emitter Location and Signal Parameter Estimation," In Proc. RADC Spectral Estimation Workshop, 1979.

9. Pasala, K. M., and Friel E. M., " Mutual Coupling Effects and their Reduction in Wideband direction of Arrival Estimation," To appear in October 1994 issue of **IEEE Transactions on Aerospace and Electronic Systems.**

10. Friel, E. M., "Direction Finding in the presence of a Near Field Scatterer," Final Report for Graduate student Research Program, WL/AArM-3, July 1994.

AN ASSESSMENT OF THE WL/AAAI-4 ANTENNA WAVEFRONT SIMULATOR

Robert P. Penno, Ph.D.
Assistant Professor
Department of Electrical Engineering

University of Dayton
300 College Park
Dayton, Ohio 45469-0226

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

August, 1994

AN ASSESSMENT OF THE WL/AAAI-4 ANTENNA WAVEFRONT SIMULATOR

Robert P. Penno, Ph.D.
Assistant Professor
Department of Electrical Engineering
University of Dayton

Abstract

Wright Laboratory has designed and constructed a one-of-a-kind Antenna Wavefront Simulator (AWFS) to produce signals that replicate those at the terminals of the elements of an antenna array. Such a unique research and development platform provides a compact, economic test environment upon which adaptive array electronics packages can be developed and tested. Furthermore, the AWFS could eliminate the need for testing in an anechoic chamber or open range until final test.

One of the most critical design issues raised in the development of the AWFS was the use of analog, linear phase shifters, instead of (tapped) delay lines, to produce the appropriate wavefront signals at respective elements in the array. Delay lines have been the traditional approach as they readily characterize a wavefront incident upon an array by delaying all frequency components in that wavefront by the same period of time. Phase shifters, on the other hand, are typically assumed to produce a constant phase shift, even over a band of frequencies. Consequently, phase shifters have never been used in wavefront simulation of broad band signals. Therefore, the purpose of this report is to describe in detail how phase shifters with a linear phase shift capability are the functional equivalent of delay lines, even for replicating wideband signals. Also, other aspects of the AWFS are discussed, and recommendations for future direction of its development are offered.

AN ASSESSMENT OF THE WL/AAAI-4 ANTENNA WAVEFRONT SIMULATOR

Robert P. Penno, Ph.D.

Introduction

The Antenna Wavefront Simulator(AWFS) being developed by WL/AAAI-4 is an innovative concept rapidly becoming a reality. The goal of the AWFS is to provide a means of testing antenna array electronics packages in a dynamic, controlled laboratory environment. Currently, such electronics packages are either simulated by computer model or evaluated via an anechoic chamber or open range field tests. The AWFS provides a real-time test platform that is inexpensive and compact. Moreover, it serves as an ideal supplement to the the current design process, by affording the opportunity to test and modify in real time, before final testing in an open range environment.

Operation

The function of the AWFS is to provide a radio frequency (rf) electromagnetic signal, equivalent to that provided by an antenna array, to the ports of an electronics package. Typically, this signal is comprised of several electromagnetic waves oscillating harmonically at different frequencies and incident upon the antenna array at different angles. These waves induce currents to flow upon the individual antenna elements comprising the array. These currents, then, are the signals delivered to the electronics package, and which the AWFS must simulate.

To perform this function, the AWFS utilizes an rf signal generator for each source to produce the initial harmonic output. As described in Figure 1, this signal is guided through a power divider. Each of the outputs of the power divider is then directed through a voltage-controlled phase shifter and a voltage-controlled attenuator. The phase shifter adds phase to the signal commensurate with a time delay representative of the time required for the wave to travel to a specific antenna element. The attenuator offers the capability of leveling the magnitude of the individual signals to account for nonuniform gains acquired elsewhere in the system. To account for multiple signals, which include message and jamming signals, a signal generator-phase

shifter-attenuator path is provided for each antenna element for each of the signals to be produced. For example, a message signal and two jamming signals incident upon an eight element antenna array would require 3 signal generators, 24 phase shifters and 24 attenuators to simulate. A power combiner sums the phase shifted outputs to produce the "antenna output" fed into the antenna electronics package. These signal modifications are orchestrated by a wavefront direction controller, which dictates the frequency of operation of the signal generators and the amount of phase shift and attenuation of each signal. Figure 1 shows a block diagram of this configuration.

Primary Goals

The primary goal of this research was to assess the AWFS in terms of its ability to provide a test environment for any antenna array electronics package. The first goal of this project, then, was to determine what aspects of the AWFS required scrutiny, and how that scrutiny would best be affected. This investigation determined that on a qualitative level, several questions required consideration:

(i) How can the AWFS better simulate a real time testing environment;

(ii) How can the effectiveness of the AWFS be evaluated; and

(iii) How effective are phase shifters in constructing the individual wave components?

The initial version of the AWFS is a somewhat idealized system. It models ideal signals incident upon an ideal antenna array. Recognizing this, more specific questions were asked regarding what modifications could be made to provide a more realistic simulation. Of particular interest were second order effects that distinguish the actual operation of an adaptive array from a laboratory test environment. Some of these second order effects are described by the following:

(a) Do linear phase shifters afford creation of accurate models of electromagnetic waves incident upon antenna array elements;

(b) What impact does the mutual impedance of the antenna elements have upon the accuracy of the model;

(c) As the adaptive weights of the electronics package vary, what impact do these changes have upon the antenna element impedance (moreover, how are these changes best characterized by the AWFS); and

(d) How does the time required by the AWFS to change test scenarios compare to the dynamic capability of the array electronics package ?

As work progressed, other concerns were added:

(e) To what extent should the effects of the radome be considered and, therefore, modelled;

(f) How can the effects of reflective obstacles and their associated contributions (i.e. multipath) best be included; and

(g) How can the AWFS be tested in a dynamic environment to know its dynamic limitations, (and not the dynamic limitations of the antenna electronics package) ?

Moreover, the incorporation of any and all second order effects might be accompanied by a degradation in the speed of the AWFS. So, an anciliary concern is a priorizing of these effects in terms of the impact of their implementation and means of implementation. Finally, a computer simulation of the AWFS, using MATLAB™ , was desired. as a supplementary tool.

The Use of Linear Phase Shifters

When an adaptive array establishes a null to remove an interfering, or jamming, signal, it is assumed that the jamming signal is narrowband. The subsequent null created is of the form, $|\sin(\omega-\omega_d)T/2|$, and is characterized by very sharp nulls. A wideband jamming signal is one, then, that has frequency components, ω_1, such that

$$|\sin(\omega_1-\omega_d)T/2|\neq 0, \qquad (\omega_d-\frac{B}{2})\leq\omega_1\leq(\omega_d+\frac{B}{2}), \qquad \omega_1\neq\omega_d$$

where B is the bandwidth of the wideband signal. Because of the sharp nulling performed by the adaptive array, components of the jamming signal not exactly at the nulling frequency, ω_d, have significant power passing through the processor. The presence of these components causes the adaptive processor to raise the weights so as to eliminate what the processor perceives as additional narrowband jamming signals. Since the adaptive processor is a

feedback loop, this process continues until the weights are high enough to null all of the jamming signal. Unfortunately, the sharpness of the null, in conjunction with the continuous nature of the spectrum of the wideband jamming signal and the finite number of nulls that the array can provide, cause the message signal to be appreciably attenuated. This phenomenon is often referred to as spatial dispersion, or bandwidth degradation due to channel mismatch, and is a well studied problem in adaptive array theory. The most practical solution to the spatial dispersion problem in adaptive arrays is the use of tapped delay lines, where the adaptive weights are effectively replaced by a discrete Fourier Series. The effect of this change is to provide a frequency domain shaping that broadens the null. In filter theory jargon, tapped delay lines can provide a wider stopband in the bandstop filtering that is being affected.

This problem assumes a slightly different form in the construction of the AWFS. The purpose of the AWFS is to produce rf signals that simulate a wavefront incident upon an antenna array. Thus, a single rf signal input to a reference port on the electronics package must be delayed by the amount of time it would physically take the wave to impinge upon the next antenna before it is input to the next port of the electronics package. Figure 2 shows a simple two element array and depicts the additional distance, d, that the wave must travel after impinging upon the reference element before it is incident upon the second element. This distance, d, shown in Figure 2, can be construed in terms of a time delay or a shift in phase of the wave, as per the property of the Fourier Transform:

$$\Im\{x(t-\frac{d}{c})\} \;=\; e^{j\omega\frac{d}{c}}X(\omega),$$

where c is the speed of light in free space. Thus, a phase shift comparable to the delay time is all that is required for a single frequency disturbance.

With a wideband jamming signal, however, each frequency component of the signal must be shifted by a different amount of phase, since each component completes a different number of cycles (or fractions thereof) over distance, d. A wideband signal requires, then, a <u>linear phase shift</u>, not a constant phase shift, the slope of which must be the ratio, d/c, over the band of

frequencies comprising the signal. Furthermore, this linear phase shift must be applied to each signal at each output port (simulating an antenna element). Thus, all frequency components of a wideband signal are given the appropriate, yet distinct, phase shift required to emulate the extra distance the wave travels relative to the reference element.

The linear phase shifter chosen to perform this task provides a linear phase shift in the range of $0.° \leq \phi \leq 380.°$. This device uses a backward wave coupler, a four-port microwave component, and two series inductor-capacitor (LC) circuits, the capacitor being a voltage-adjustable variac type. The LC circuits are connected to ports 3 and 4 of the coupler and provide total reflection of any wave incident upon that port, but with a phase shift proportional to the frequency of the incident wave. A signal input to port 1, is totally reflected, but shifted in phase, at both ports 3 and 4. The output at port 2 provides the appropriate phase shift for each frequency component (see Figure 3).

A computer simulation was performed based upon this model. As is shown in figure 4, the phase shift across a typical 20 Mhz band at 1.575 GHz showed extremely linear characteristics for a fixed series capacitance. Likewise, figure 5 displays the midband phase shift varying as a function of capacitance. By selecting a voltage, and hence a varactor capacitance, the slope of the phase shift was also selected. These are the essential characteristics required for operation of the linear phase shifter described above.

Using an HP network analyzer, all of the phase shifters were tested for these characteristics. Typical of these tests are the results shown in figures 6 and 7. Figure 6 displays the measured phase shift as a function of frequency over a 20 Mhz band centered at 1.2276 Ghz. The phase shift is quite linear, with a slope of approximately 5 nanoseconds. Figure 7 displays the group delay, as approximated by the network analyer using a forward difference algorithm. The time delay at the center frequency is noted as 5.0607 nsec, quite comparable to the slope of the previous figure. The apparent erratic nature of figure 7 can be explained by the unstable algorithm used to calculate it. The group delay is the derivative of the phase with respect to

frequency and numerical approximations of derivatives are notoriously unstable.

Moreover, the measurements and computer model support the statement that the phase shifters provide a linear phase shift across the desired band. As has been stated, such linear phase shift is equivalent to a constant time delay for all frequency components of a wideband signal. Thus, the AWFS can accurately simulate a wideband signal provided that (i) the phase shifters used provide a linear phase shift across the entire desired bandwidth of the signal, and (ii) one phase shifter per antenna element is used for each signal to be simulated.

Antenna Modelling-Mutual Coupling

A most frequent assumption in the assessment of any antenna array signal processing package is that the signal received by the individual antenna elements is unperturbed from the ideal signal. Most often, frequency dependent, interactive effects of the individual elements conspire to modify, or perturb, the received signal prior to admission into the signal processing electronics. The primary source of this perturbation is referred to as mutual coupling, and has its foundation in the very expression of Maxwell's equations, themselves. The essence of mutual coupling lies in the principle that an electromagnetic wave incident upon an antenna induces currents to flow on that antenna element. However, these induced currents serve as sources, themselves, and act to re-radiate some of this induced power. Adjacent antennas, then, receive not only the incident electromagnetic wave, but portions of that wave that have impinged upon other antennas and have been re-radiated. It is the power re-radiated from other adjacent antennas that perturbs the ideal signal. The degree to which this re-radiated power affects, or perturbs, the ideal is directly dependent upon the proximity of the antennas. More importantly, it is the electrical distance (i.e. distance as measured in wavelengths), that defines the magnitude of this perturbation. As the frequency of the incident wave decreases, its wavelength increases. So, while the physical separation between two antenna elements remains unchanged, its electrical distance decreases, which increases the amount of mutual coupling.

To succeed in its objective, the AWFS must be able to replicate mutual coupling to the extent that it is a factor in perturbing the signal. While mutual coupling is always present in an antenna array, it is a function of frequency, proximity, and antenna element type, and may be negligible in its effects. With respect to the AWFS (as it relates to the GPS, 7-element CRPA array), mutual coupling was deemed by some to be negligible in its effect upon the null-forming process. Nonetheless, the already impressive capabilities of the AWFS would be enhanced with the addition of a mutual coupling adjustment capability. Moreover, an analysis of the array involved in any AWFS test would seem to enhance confidence in that test. Such an analysis should include magnitude and phase measurements at each of the elements for illumination by a wave at a variety of angles of incidence. Additionally, measurements for individual elements would be desirable. Comparison of this data would provide a tabular listing of the effects of mutual coupling. These results, in turn, could be input to the controller of the AWFS, which would effectively "add" these effects to the the ideal wave generated by the AWFS. An alternative to actual measurement would be the theoretical or numerical analysis of the antenna elements to assess these effects. Here again, the goal would be to produce a table of corrections for the AWFS controller. On one hand, this approach is straightforward. On the other, it requires apriori knowledge of the antenna elements, and the use of the controller to search a table and add a correction factor. The time required for such a search could be prohibitive, requiring an alternative approach such as a neural network.

Antenna Modelling-Adaptive Weights

Similar to the notion of mutual coupling is the impact of loading of the antenna. As described above, an electromagnetic wave incident upon an antenna element ultimately produces a voltage across a set of terminals. Any load connecting those terminals experiences a current through it which is generated by the incident wave. As that load changes, a simplistic application of Kirchhoff's Law indicates that either voltage or current must change. In an adaptive array, the antenna elements are connected to an electronics package, which necessarily is characterized by an input impedance and acts as a load upon the antenna element. If this load changes, it will have an impact upon the currents flowing in the element, and hence upon the receiving properties

of the antenna element itself. The type of antenna element and the magnitude of the load change dictate the extent of this impact; such impact may again be negligible. There are two changes at work here: antenna loading other than a matched, or design, load, and a changing of the antenna loading.

In the case of the CRPA used in the GPS under test, the AE-1 antenna electronics package that it feeds includes a PIN diode limiter which feeds a LNA microwave amplifier. In practice, this combination seems to provide an impedance that is relatively constant over the frequencies involved, and as such, serves as a buffer between the antenna and the antenna electronics. Thus, changes in the adaptive weights of the electronics package have negligible effect upon the antenna. Since this is empirically derived, it would seem reasonable to perform either an analysis or measurements upon the package to determine the validity of the assertion that the input impedance of the electronics package is, indeed, constant. Moreover, this analysis should be extended to any package under evaluation by the AWFS, since, in general, a buffer between antenna element and electronics package cannot be expected. With respect to the CRPA/AE-1, it is suggested that such analysis be performed on a typical limiter/amplifier combination.

Dynamic Performance Measurement

The AWFS is intended to be a dynamic testing solution/platform. In the evaluation of the GPS AE-1 controller and the associated receiver, there is a need for simulating catch and lock of GPS signals by a highly maneuverable platform moving at very high speed. This high speed maneuverability is represented by the AWFS in terms of rapidly changing rf outputs from the AWFS. To create these outputs, the AWFS must generate and modify (i.e. add the appropriate phase shifts) the outputs from the rf signal generators.

There are two considerations here. First, how are the effects of high speed maneuvering(HSM) best introduced into the AWFS so as to simulate these effects? Second, having accounted for these effects, how is the effectiveness of these modification best evaluated. Ancilliary questions include whether the rf signal generators, phase shifters and other components can respond quickly enough to simulate the HSM. The use of look up tables to provide modifications that model these effects and others (such as antenna element attributes i.e. mutual coupling, multipath, etc.) take processing time that

limit the ability of the AWFS to accurately simulate these HSM characteristics. Presently, the AE-1 is being used to determine whether the AWFS is producing appropriate simulated signals. However, the current testing mode is static, i.e. the simulated signals are produced by the AWFS, submitted to the AE-1, the weights of the AE-1 are then frozen, and the nulling of the AE-1 is evaluated. If this process does not successfully produce the desired result, the question arises: Is the AWFS producing the proper signal, or is the AE-1 being challenged beyond its limitations.

A test controller/receiver is needed, then, that will afford the ability to check the received signal from the AWFS, and interpret whether that signal is correct (relative to the desired simulation). Such a device could be any commercially produced GPS receiver card, or even an AE-1 controller and the associated receiver. What is critical is the ability to access the interior workings of the unit to measure and subsequently asses the quality of that signal against the expected received signal in terms of HSM. Ultimately, development of this test device could evolve into a separate piece of test equipment whose measurement capabilities would not be limited to the evaluation of the AWFS. Moreover, such a device could be used to evaluate other pieces of equipment.

Computer Model (Matlab™)

To corroborate the analysis of portions of the AWFS, construction of an AWFS computer model was begun. The portion constructed have proven extremely useful. For example, in the understanding of the linear phase shifters, the Matlab™ model accurately predicted the linear phase across the band while describing the nonlinear relationship between phase shift and varacter capacitance. This was described earlier by Figures 4 and 5.

The model was expanded to include a graphic display of a multiple signal input. For a composite signal, consisting of several wavefronts incident at different azimuth and elevation angles upon an arbitrary planar array of isotropic antenna elements, the magnitude of the signals can be displayed as a function of azimuth ('phi') and elevation ('theta'). Figure 8 is an example of two signals incident upon a 100 x 100 element array at (phi, theta) of (45°, 75°) and (135°, 45°) respectively. Figure 9 displays a contoured outline

of Figure 8.

Suggested Paths for Future Investigation:

Of primary importance at this point would be considerations of the effects of mutual impedance of the individual antenna elements of the array. On a specific level, this is the most obvious source for signal perturbation from the ideal signal expected by the designer of an adaptive array electronics package. On a broader level, should mutual impedance prove to be an effect for which system compensation is required, the implementation of such compensation could prove challenging. One possible means of compensation consists of creating a look-up table of corrections for each antenna element at each angle. Such a table would reside in computer memory, but would require substantial memory and computer timme to implement. Other lines of approach include the use of self-learning techniques, such as neural networks, to provide compensation. The broader issue involves the implementation of a general scheme to implement any and all modifications to the AWFS that would provide a test scenario more closely related to the actual user environment.

Summary

The AWFS has the potential to provide a viable means of providing inexpensive testing of an adaptive array package. Its development would provide a tremendous alternative to the existing means of testing. What has been done here is threefold. First, critical areas in the development of the AWFS requiring focus and development have been identified. Secondly, the most immediate of these concerns, the use of linear phase shifters instead of delay lines, has been addressed and shown to be viable, if not advantageous. Finally, a computer simulation using Matlab™ has been initiated to model the AWFS. Such a model proved useful in understanding and validating the use of linear phase shifters. Further development of the computer model is expected as other salient issues of the AWFS' development are addressed. Finally, all of the comments cited herein should be viewed as having the potential to improve an already successful system. Moreover, the AWFS has the potential to be a primary tool in the testing and development of array electronics, at least the equal of a computer simulation. More likely, it will provide an economic alternative to some anechoic chamber or open range testing.

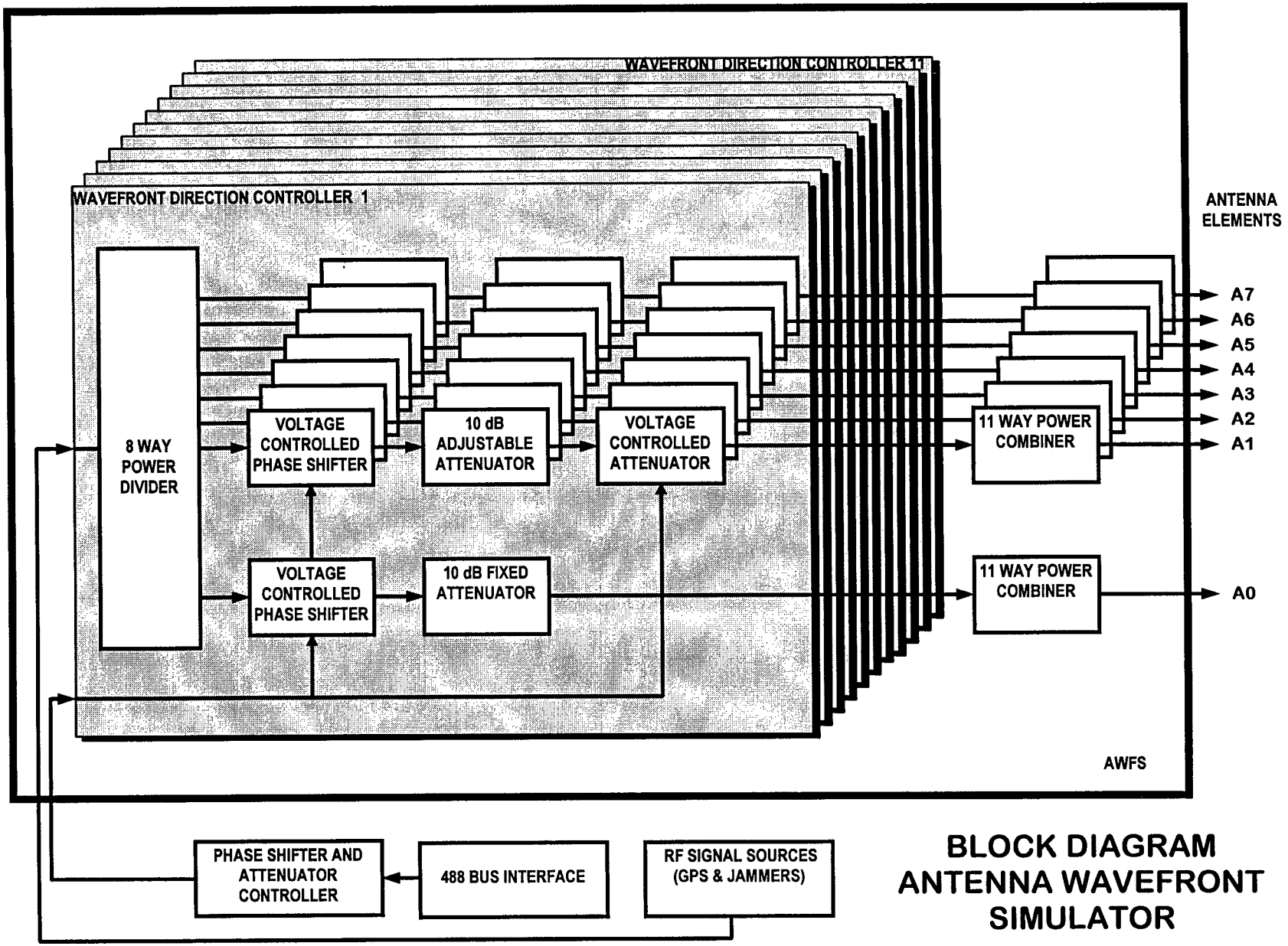

WAVEFRONT DIRECTION CONTROLLER 11
WAVEFRONT DIRECTION CONTROLLER 1
ANTENNA ELEMENTS
8 WAY POWER DIVIDER
VOLTAGE CONTROLLED PHASE SHIFTER
10 dB ADJUSTABLE ATTENUATOR
VOLTAGE CONTROLLED ATTENUATOR
11 WAY POWER COMBINER
A7
A6
A5
A4
A3
A2
A1
VOLTAGE CONTROLLED PHASE SHIFTER
10 dB FIXED ATTENUATOR
11 WAY POWER COMBINER
A0
AWFS
PHASE SHIFTER AND ATTENUATOR CONTROLLER
488 BUS INTERFACE
RF SIGNAL SOURCES (GPS & JAMMERS)
BLOCK DIAGRAM
ANTENNA WAVEFRONT
SIMULATOR

Figure 1

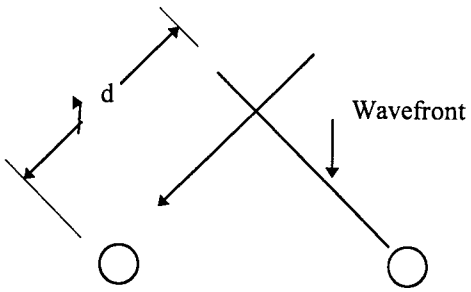


Figure 2: Phase Shift and Time Delay

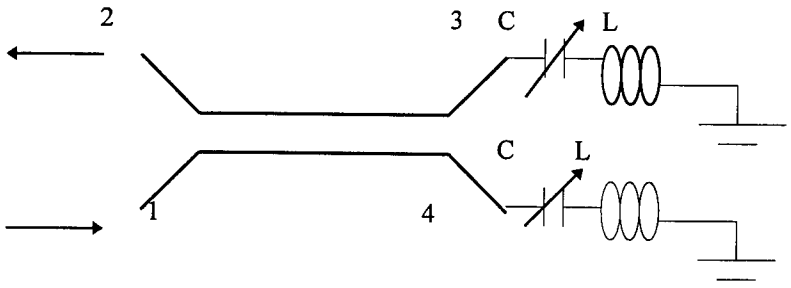


Figure 3: Schematic Diagram of Backward Coupled Phase Shifter

Figure 4. Phase(degrees) vs Frequency

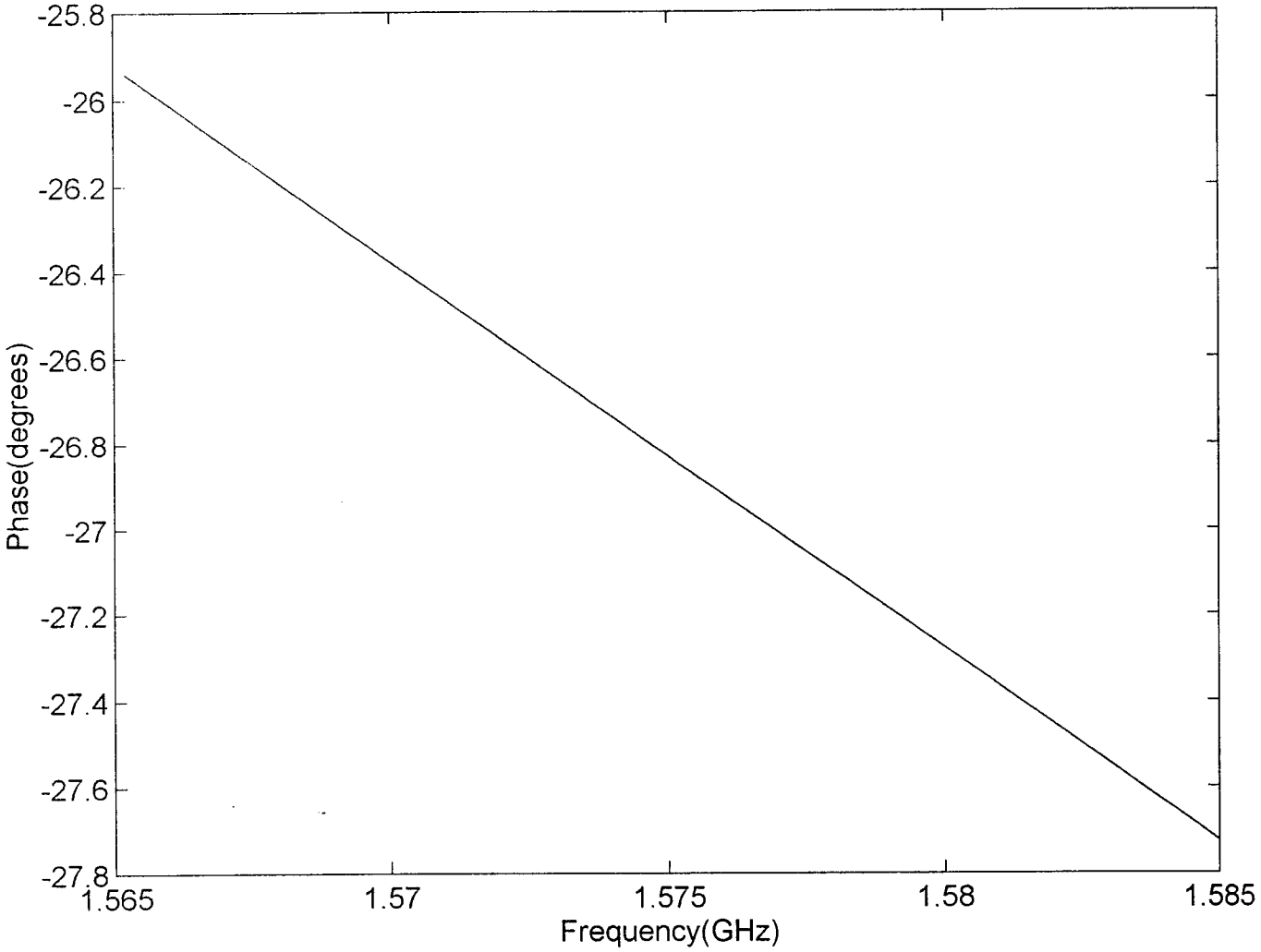
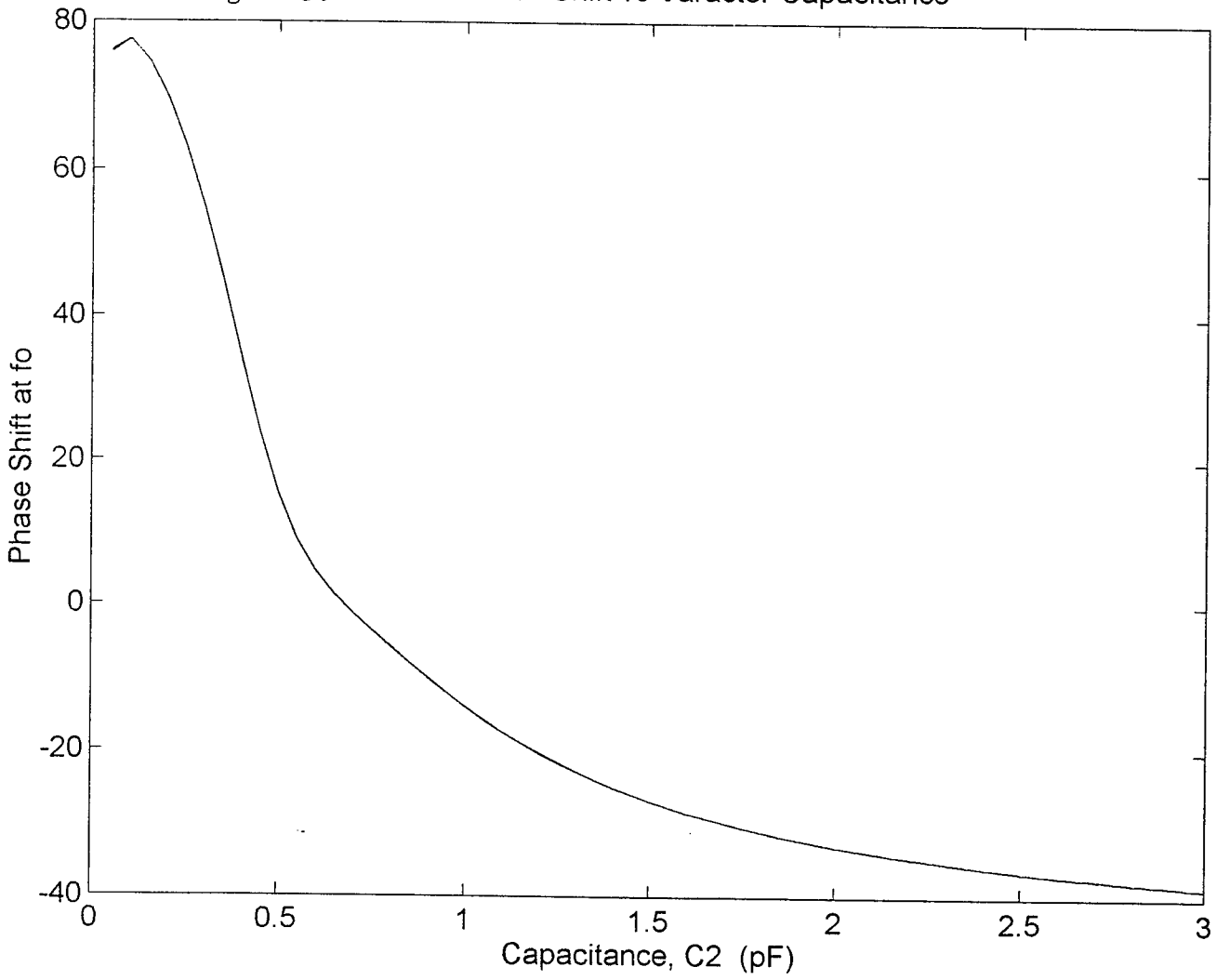


Figure 5. Center Phase Shift vs Varactor Capacitance



Phase Shift vs Frequency

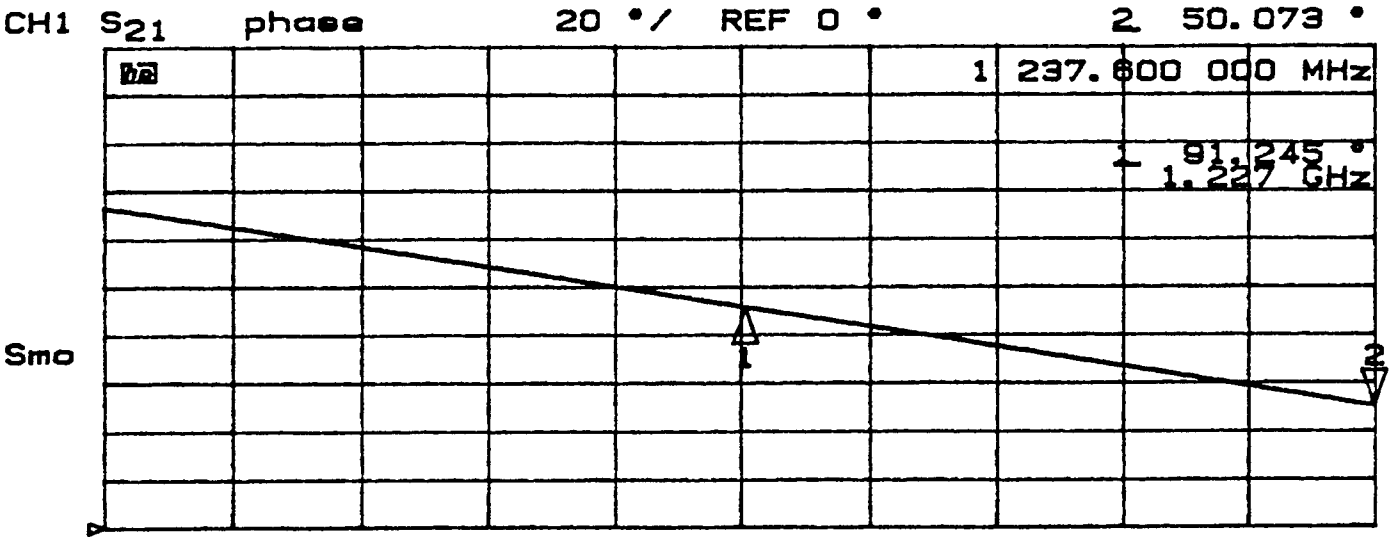


Figure 6

Delay Time vs Frequency

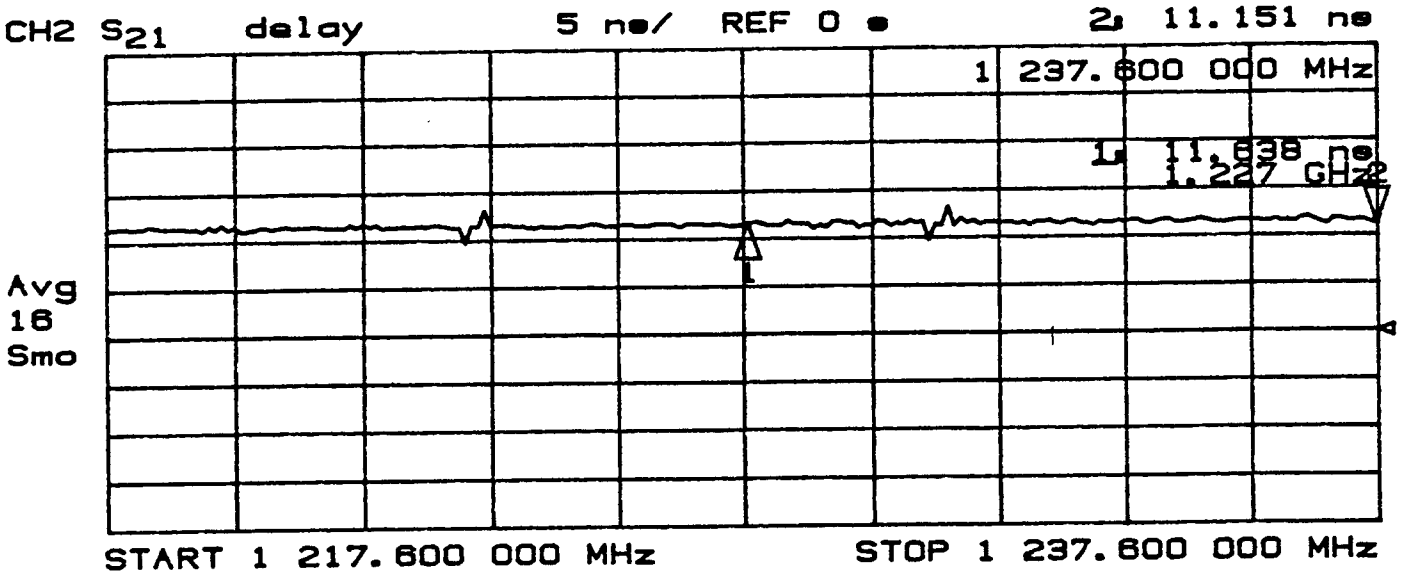


Figure 7

Figure 8. Field Pattern vs Spatial Angle

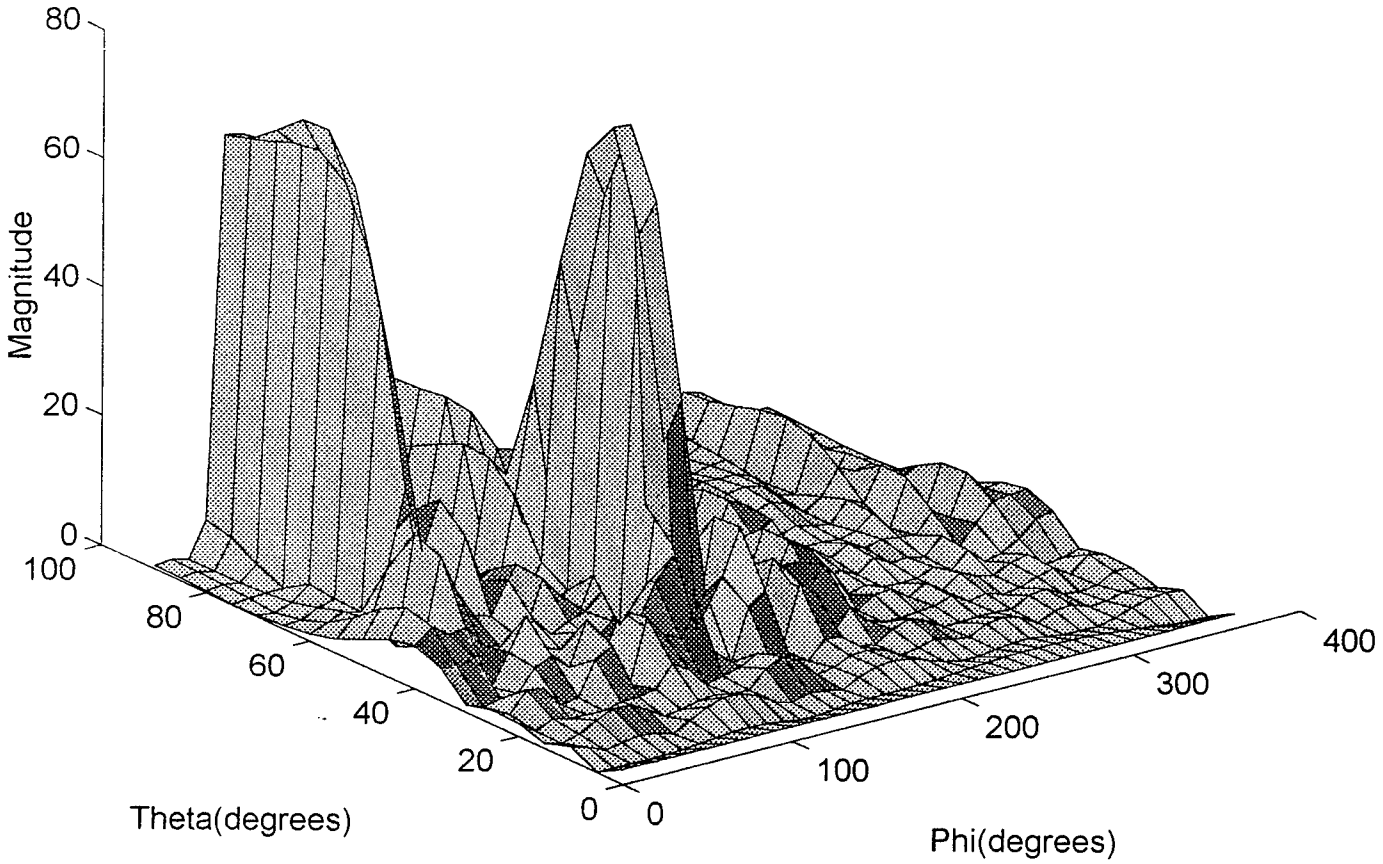
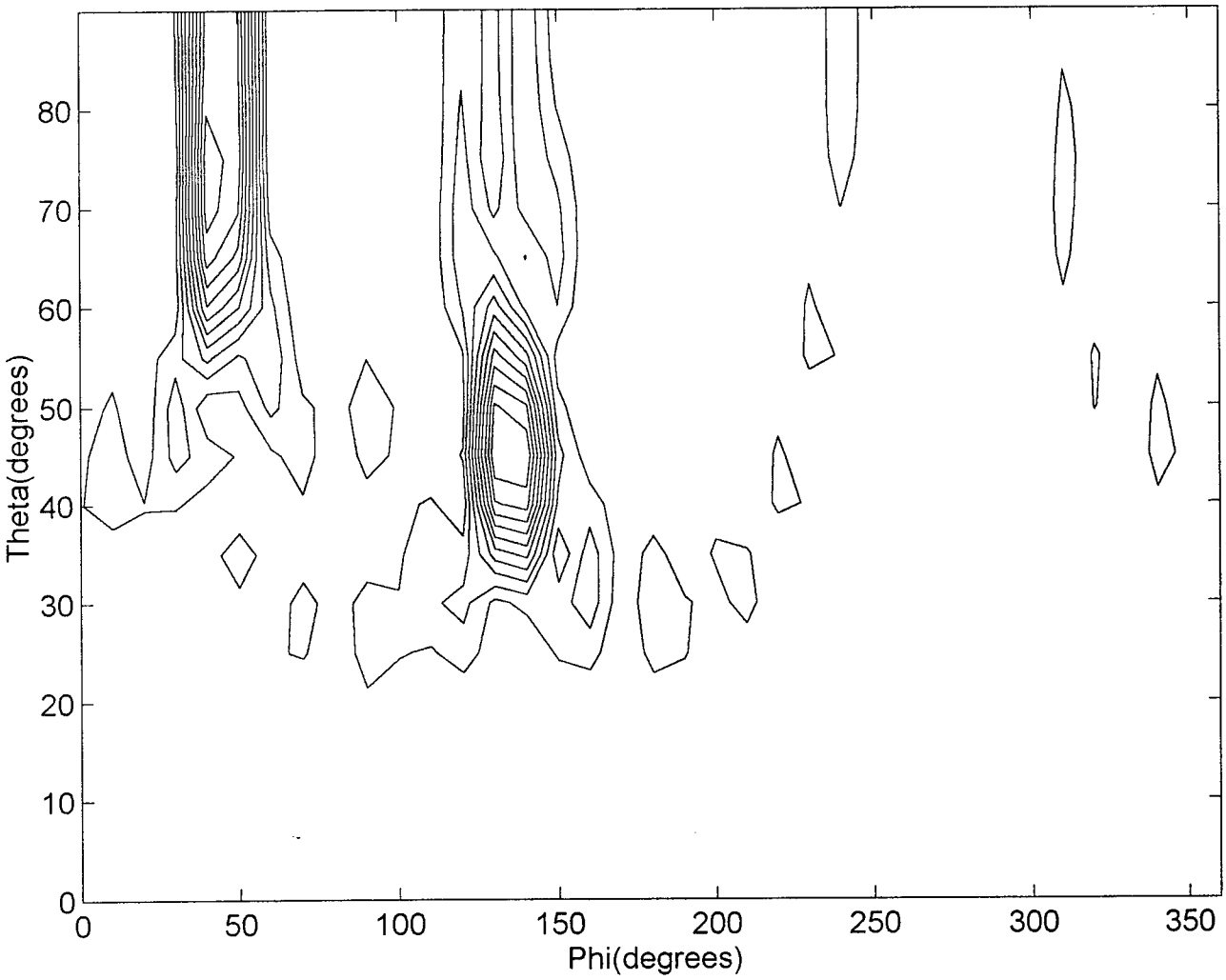


Figure 9. Field Pattern vs Spatial Angle



A SURVEY OF LITERATURE ON FUNCTION DECOMPOSITION

Marek A. Perkowski
Professor
Department of Electrical Engineering

Portland State University,
P.O. Box 751
Portland, Oregon, 97207-0751
mperkows@ee.pdx.edu

Final Report for:
Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

September 1994

A SURVEY OF LITERATURE ON FUNCTION DECOMPOSITION

Marek A. Perkowski
Professor
Department of Electrical Engineering,
Portland State University
Portland, OR, 97207-0751
mperkows@ee.pdx.edu

Abstract

This report surveys the literature on decomposition of binary, multiple-valued, fuzzy and continuous functions. It gives also references to relevant basic logic synthesis papers that concern topics important for decomposition, such as for instance representation of Boolean functions or symmetry of Boolean functions.

As a result of the analysis of the most successful decomposition programs for Ashenhurst-Curtis Decomposition, several conclusions are derived that should allow to create a new program that will be able to outperform all the existing approaches to decomposition. Creating such a superior program is necessary to make it practically useful for applications that are of interest to Pattern Theory group at Avionics Labs of Wright Laboratories.

In addition, the program will be also able to solve problems that have been never formulated before. It will be a test-bed to develop and compare several known and new partial ideas related to decomposition. Our emphasis is on the following topics:

1. representation of data and efficient algorithms for data manipulation,
2. variable ordering methods for variable partitioning to create bound and free sets; heuristic approaches and their comparison,
3. use of partial and total symmetries in data to decrease the search space,
4. methods of dealing with strongly unspecified functions,
5. special cases of decomposition, that can be efficiently handled (cascades, trees without variable repetition).

Acknowledgments.

I wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research for the sponsorship of this research. This has been one of the most exciting summers in my life - I learned that logic synthesis methods that I have been working for the past 26 years of my life have applications in Machine Learning, which opens up a lot of interesting research opportunities for me.

I would like to especially thank Dr. Timothy Ross, who not only provided me with an excellent guidance in machine learning, has been a creative "sparring partner" in discussions on logic problems, but also helped me in all kinds of other issues making my stay in Dayton a truly enjoyable one. I would like to thank the entire Pattern Theory team at WL/AART-2: Dave Gadd, Mike Noviskey, Jeff Goldman; and also Chris Vogt, Mark Axtel, John Carroll, and Craig Files, who all contributed to my better understanding of one or another aspect of the fascinating subjects that the Pattern Theory research is dealing with.

I would like to thank Julie Terri, Terri Willis and Martha Smith for their kindness and secretarial supports, and I would like to acknowledge Professors Robert Brayton, Malgorzata Marek-Sadowska, Tsutomu Sasao, Tadeusz Luba, Bernd Steinbach, Maciej Ciesielski, Bernd Becker, Radomir Stankovic, and Randall Bryant for their help in finding the necessary materials.

I would also like to thank the Air Force Office of Scientific Research for providing the opportunity and support to do this research.

A SURVEY OF LITERATURE ON FUNCTION DECOMPOSITION

Marek A. Perkowski

1 Introduction

Pattern Theory group at Avionics Laboratory of Wright Laboratories applies the so-called Ashenhurst-Curtis Decomposition of Boolean functions as a new machine learning paradigm [80]. Many decomposition ideas have been implemented in the programming system Flash developed by this group [85]. Although much research has been done recently world-wide on the decomposition of Boolean functions for logic circuit design applications, it seems to be a lack of recent monograph books or even survey papers in the area of Boolean and Multiple-Valued Logic Decomposition.

This survey is a preliminary attempt, and **it is only a part of a more complete document on the same topic and by the same author, that is available from the Pattern Theory group at Avionics Laboratory. The full document includes a description of all research issues in decomposition and respective computer programs, a complete bibliography with more than 500 positions, and three volumes of transparencies of more than 22 lectures delivered by the author to the group in the Summer of 1994.**

The functional decomposition is an extremely wide and multifaceted topic, it is in the very center of the area of logic synthesis. Therefore, in writing this survey, the author had to restrict himself to some point of view. He decided to look at the decomposition from the point of view of the research and development goals of the Pattern Theory group at Wright Labs.

The goals of the Pattern Theory work are the following:

- Creation of practical "machine learning" algorithms that will give useful results on military data. Which means, at the minimum, 30 binary input variables. Better, about 100 variables. The variables may be multiple-valued. The program must be robust across various classes of data. It must have better generalization capabilities than the C4.5 program or other decision-tree-based programs, or the decision-diagram approaches, as presented in 1994 by Ron Kohavi and Arlindo Oliveira. The program must be also faster than the current version of Flash.

- Contribute to the methods of "discovery". Upgrade Flash in such a way that it will be able to invent new algorithms and digital structures, for instance: it took about 100 years from the invention of discrete Fourier Transform to the discovery of Fast Fourier Transform (FFT). We would like the program to be able to find similarity patterns in few seconds and create the structure of the FFT automatically. Flash can already "invent" simple circuits such as an adder, but is unable to find more complex circuits, such as a combinational sorter, lattice realizations of symmetric functions, and butterflies. One of the reasons is, that at present Flash cannot deal with multi-output functions.

- Advance the state of the art of the **decomposition theory** with respect to many new aspects of decomposition that are specific to learning applications.

There are other points to be made about a special nature of the decomposition for Pattern Theory. The first very important observation is the following. In the past, the work in the area of Ashenhurst-Curtis Decomposition (AC decomposition) was conducted for application in circuit design, [14], [19]. In circuit design, for problems taken from engineering practice, when the number of input variables increases, the number of don't cares increases slightly with the number of minterms (true minterms, positive samples), and the number of zeros (false minterms, negative samples) increases dramatically. This kind of examples can be found in the well-known benchmark sets for logic synthesis, MCNC,

ISCAS and other, which have been created by industrial companies and universities, and that deal with control, DSP, computer, and telecommunication circuits, including large controllers, state machines, decoding logic, arithmetic, and interfaces.

This is in sharp contrast to machine learning, where with the increase in the number of input variables, there is a small increase in the number of both positive and negative samples, but dramatic increase in the number of don't cares. For instance, it is reasonable to expect that for a function of 100 variables there will be not more than 10,000 cares. This distinction makes completely different requirements on the decomposition programs in these two areas, a point that has been not yet sufficiently observed and appreciated. *This fact calls for the development of totally new approaches to synthesis.*

Another important point here is, that there is much confusion in the technical literature, graduate textbooks, and understanding of experts in the field, of what actually constitutes the Ashenhurst/Curtis Decomposition, and its many special cases and extensions. An attempt has been made in this research to find the original papers, give precize and uniforms definitions, and attribute the name of the decomposition always to its inventor. Until now, the main source of information on decomposition in USA is the book by Curtis [19] that presents early American research on what is now called the Ashenhurst-Curtis decomposition. Other monograph books include: Bochmann and Steinbach [76] (in German) that presents mostly German and European research, Bibilo and Yenin [62] (in Russian) that presents only the research from the former Soviet Union, and two monographs by Luba [58, 61] (in Polish) that concentrate mainly on the approach developed at the Warsaw University of Technology.

We found the following types of decomposition the most interesting for the above-mentioned requirements of Pattern Theory: Ashenhurst decompositions, defined in [14]; Curtis decompositions, defined in [19]; Roth/Karp decomposition [16, 18, 22, 23, 24]; Sasao multiple-valued PLA decomposition [64, 70, 96]; the decomposition of Perkowski, Hoang, Brown, and Wan [63, 86]; Yang/Ciesielski PLA decomposition [72, 81]; Luba's Partition-Based binary [58, 61], and multiple-valued decomposition [91, 107]; Bochmann/Posthoff/Dresig/Steinbach algebraic AN/OR/EXOR decomposition [76], so-called XBOOLE Decomposition; Zakrevskij/Bibilo decomposition [37, 62]; Orthogonal decompositions of Perkowski et al. [84, 92]; and Lai/Pedram/Vrudhula approach based on BDDs [89].

2 Goals.

The goals of this survey are the following:

1. Find what research has been already performed, not to repeat the same work again.
2. Find which computer programs did perform best from the practical point of view.
3. Find, as many as possible, small ideas, heuristics, and representations, that can be borrowed, enhanced, and integrated in the proposed new system.
4. Find what are the open problems, problems known to be not solved, or even not yet formulated.
5. Find how people are approaching the problems that we considered the most important ones based on our previous research and experience:
   - variable partitioning to free and bound sets, [14, 70, 67, 86, 101, 106].
   - coloring or covering approaches to find groups of compatible columns [67, 86, 62, 61].
   - encoding of compatible groups of columns [62, 86, 70].
   - subfunction reusing [86].
   - reduction of sets of vacuous variables [107, 58].

It is assumed that the reader is familiar with Ashenhurst [14] and Curtis [19] approaches to decomposition. Our goal here will be not to present all related ideas that can be found in the literature on the subject. We feel also, that there is a lot of repetitions and re-formulations, and some papers are strictly related to Kmap or cube representations, that we consider obsolete. Therefore, our survey concentrates on papers and ideas that seem to be the most fruitful for the implementation of decomposers; mostly approximate ones, for large numbers of input and output variables, and for strongly unspecified functions.

Therefore, we devote also a due amount of attention to some general logic synthesis problems that will be useful to design such decomposers, for instance:

- representation issues, such as describing functions by various forms of decision diagrams,
- search constraints, such as finding partial or total symmetries in input variables of the functions, that can limit the search space without sacrificing the decomposition quality.
- heuristic methods to reduce the search, such as the variable ordering techniques for Binary Decision Diagrams (BDDs) [60].

3 The Essence of Functional Decomposition and the Role of Bias.

The AC Decomposition is not an Algebraic nor a Boolean concept. It means, that it can be explained without using the concept of Boolean algebra and gates. What is necessary is only describing a function (i.e. a mapping) in terms of a composition of other mappings.

Analogous property is true for multiple-valued signals. The concept of decomposition can be explained without resorting to Post, Galois, or any other definite algebra to realize the functions. This is a very powerful property of decompositional methods, that places them at the very fundament of many research areas, and allows to generalize binary methods to other systems described by functions.

The above does not mean, however, that some kind of "algebraic" concepts cannot be used in the decomposition process. For instance, in binary decomposition one can use binary decision diagrams for data representation, and the Binary Decision Diagrams are a concept of Boolean algebra.

The main advantage of current Flash is that it is not using any particular gates or other "biases" that would restrict the space of solutions or would mimick human approaches. Flash finds then the "designs as they can be", and has thus a chance of a "discovery". In other synthesis methods, one assumes certain blocks to realize the "gates", for instance the AND, OR, EXOR, AND/EXOR or multiplexer, [98]. Such approaches are currently dominating in industry (standard cells, PLDs, sea of gates). Even the "universal logic module" approaches, such as the "lookup-table" approach to design with Xilinx devices [83], assume arbitrary tables, but only of a limited size (4, 5 or 6 inputs). This is in clear distinction to Flash approach, where no restriction on the number of block's inputs is given.

Such distinct and original approach of current Flash has the following consequences:

- The approach can be easily expanded to multiple-valued [66, 91, 107, 33], fuzzy [57, 45] and continuous [110] logic.
- It cannot be used directly to hardware realizations of circuits, so that some other program (like Espresso [59]) must be used to minimize and realize the non-decomposable blocks.
- There is a possibility that cost evaluation for large blocks is not accurate. The argument for Flash is that such blocks are rare.
- As mentioned, the current philosophy of Flash is that it is better to assume no any knowledge of types of gates, since assuming any kind of elements or structures would be considered a bias. I believe, however, that for efficiency Flash should be equipped also with user-controllable parametrizable biases towards symmetry [26, 35], restricted fan-in or fan-out of blocks, cascade design [21],

depth reduction [111], or other. The related problems of trees' construction, Maitra cascades [21], tandem network of Butler [47], partially symmetric functions [77], OR/AND/EXOR strong-weak decomposition of Steinbach [100, 111], and many other restricted cases of AC decomposition are also mentioned in this survey.

4 Research Areas Related to Functional Decomposition Applicable to Pattern Theory.

In this section we list the research topics and options that exist in functional decomposition.

1. **Type of variables.** On what kind of data the algorithm operates. We assume that in all cases the don't cares exist in data representation.
   (a) **Binary logic** (each variable is from set 0,1), [14, 19, 58, 80].
   (b) **Multiple-valued data** (each variable is from a finite set of integers), [66, 91, 107, 33].
   (c) **Fuzzy** (each variable is from a continuous interval [0,1]), [45, 57].
   (d) **Analog, or continuous** (each variable is an arbitrary real number), [110].
   (e) **Integer** (each variable is an arbitrary integer), introduced in [104], but not for AC decomposition.
   (f) **Mixed.** New research area.
2. **Representation of Functions.**
   (a) **Lists or Arrays of Samples.**, [80, 14, 19, 58].
   (b) **Binary Cubes and Multiple-Valued Cubes** [76, 100, 111, 62, 86].
   (c) **Binary Decision Diagram (BDD) [60] with extension for don't cares**, [89], the most successful computationally method published until 1994.
   (d) **A pair of a BDD for ON-set and a BDD for OFF-set.** A new approach, proposed below.
   (e) **Kronecker Decision Diagram (KDD) [105] with extension for don't cares, or a pair of KDD for ON-set and KDD for OFF-set.** These approaches have not yet been proposed.
   (f) **Other binary decision diagrams:** zero-supressed, pseudo-KFDD [95], mv KFDD, etc. These are new approaches.
   (g) **New classes of Decision Diagrams that expand data beyond binary**: Orthogonal [84, 92] and symmetric diagrams, Bryant's Binary Moment Diagrams [104], etc. These are all new approaches.
   (h) **Fuzzy Functions** and **Fuzzy Decision Diagrams**. These are both new approaches.
3. **Variable Partitioning to Free and Bound Sets.**
   (a) **Search** (simple genetic search [87], genetic classifiers, expert systems, heuristic search [101, 106], search in planning spaces.)
   (b) **Search-restricting theorems.** [69].
   (c) **Search-restricting heuristics.** [69].
   (d) **Applying learning to improve search performance.** [74].

4. **Column Minimization Problem.** [67, 86, 80, 103].
   (a) **Graph-coloring approach.** [67, 86].
   (b) **Set covering approach.** [58, 62, 22].
   (c) **"Set covering with weights of rows".** New approach.
   (d) **Covering/assignment approach of Bibilo et al.** [62].
   (e) **Reduced Covering approach of Luba et al.** [107].
5. **Encoding of groups of columns.**
   (a) **Fast ordering encoding of Wei Wan et al.** [86].
   (b) **Perkowski/LocNguyen fast rule-based heuristic encoder.** [75].
   (c) **Quadratic Assignment algorithms.**
   (d) **Sasao's method** (one hot code and binary code with maximum zeros - [64, 70, 96]).
   (e) **Yang and Ciesielski's approach.** [72, 81].
   (f) **Devadas' approach.** [65, 68].
   (g) **Spectral Approach** (Varma and Trachtenberg [71], Karpovsky [46]).
6. **Choice of Bound Partitions and Order of Creating Partitions.**
   In order to get some kind of restricted structure:
   - trees, tandem structures and other limited decompositions (Butler [47, 43, 38], Bender [55]).
   - partial and total symmetries [48, 26].
   - trees of non-repeated variables [16, 17, 20].
   - cascades [21, 54, 43].
   - iterative networks (such as a comparator of two numbers of arbitrary length).
   - combinational systolic structures, pipelines, FTT-like butterflies, or other computer structures.
7. **Realization of Nondecomposable blocks.**
   (a) **With SOP Minimizer such as Espresso [59].**
   (b) **With ESOP Minimizer such as Exorcism [102].**
   (c) **With a decision diagram, such as a BDD.**
   (d) **With a combination of programs, that would select the best solution, such as the current variant of Flash that mixes Exorcism and Espresso.**
   (e) **Transformations of non-decomposable blocks to decomposable blocks.** (Wei Wan's EXOR transformation [86]).
8. **Subfunction Reusing.**
   (a) **Reuse During Encoding**, this is the approach of TRADE [86].
   (b) **Reuse by Re-entering Subfunctions as Input Variables**, this general logic synthesis approach has been not yet used in the framework of AC decomposition. It is proposed below.
9. **Technological constraints** - Block packing (TRADE).
10. **Removal of largest sets of vacuous variables.**
11. **Dealing with multiple-output functions [61, 86, 89].**

Below we will briefly present the history of decomposition research and next the modern view at the issues that are most important for Pattern Theory.

5 Historical Overview of Research on Functional Decomposition.

In 1984 Boole published his fundamental book, [1]. (The literature below is in the chronological order). In this work Boole introduced the expansion, that was later on attributed incorrectly to Shannon. Thus Boole makes the beginning of the decomposition theory, both with respect to the Functional Decomposition, and the orthogonal expansions used in orthogonal decompositions [84, 92]. The so-called "Shannon Expansion" is **the** most fundamental formula of Boolean logic. Boole's book is referenced in all, both American and Russian, early research on decomposition.

In 1927 Zhegalkin in Soviet Union publishes his famous, but not known in the Western Hemisphere, paper [2], where he introduces what was next called the "Reed-Muller" transform and Reed-Muller Canonical Form. Not only was this transform successively used in efficient realizations of "Ashenhurst Decompositions", but it was a fundament in creation of all orthogonal transforms and decompositions.

In 1949 Shannon publishes his paper [3], which starts the history of engineering-oriented logic synthesis theory outside of USSR. This paper influences also research in USSR (Gavrilov, Povarov, Yablonskii), which started in the late 1930-ties but was not published in the literature available in the West until the end of the World War II.

Year 1952 brings the first papers on "Ashenhurst Decomposition" ever published [4, 5], which are the early versions of the most influential works that defined what we understand now by the "Ashenhurst Decomposition". In 1953 Ashenhurst introduces also the non-disjoint decomposition [6]. The Singer's paper [7] is also published. Although it is little known, the introduced by him representation of "Harvard Charts" (Decomposition Charts) became very influential.

In 1954 the paper by Povarov [8], formulates what is popularly known as the "fundamental Ashenhurst Theorem", as well as three other theorems useful in simple decompositions, found independently from Ashenhurst. Also this year, Reed [9] and Muller publish independently their fundamental papers, being not aware of the more or less the same ideas originated by Zhegalkin in 1927.

In years 1956 and 1957 R.L. Ashenhurst publishes his widely acclaimed paper "The decomposition of switching functions", in Proceedings of International Symposium of Theory of Switching [10]. In 1958 the first in the series of important papers by Curtis is published: [11]. Year 1959 brings a very important paper by S.B. Akers, "On a theory of Boolean functions" [12], where he introduced the concept of Boolean Difference, one of the most important concepts in switching theory, that has been next used much in conjunction with Boolean decomposition. This paper started one approach to decomposition that was next developed mostly in Europe: Belgium (Davio, Deschamps, Thayse), USSR (Gavrilov, Zakrevskij, Bibilo), and Germany (Bochmannn, Posthoff, Dresig, Baitinger, Steinbach).

New important development starts also in year 1960, when J.P. Roth from IBM publishes "Minimization over Boolean Trees," [16], a paper that started the sub-area of Ashenhurst decomposition, called "Roth-Karp Decomposition". This is the decomposition that is most extensively investigated in USA - all recent papers from U.C. Berkeley refer to Roth-Karp rather than other types of decomposition, since it was considered to be the most efficient one. They proposed the first Boolean method for extracting common subfunctions and an algorithm for identifying common subfunction between two functions based on the partitioning of compatible classes. Roth and Karp in their seminal paper [22] propose a new representation of the problem by using the cube covers of sets ON and OFF. Let us observe that this early representation, or its close derivatives, are still being used in some of the most successful programs for decomposition, since it is both compact and especially well suited for strongly unspecified functions of many variables [76, 100, 111, 62, 86].

In general, late 50s and early 60s bring many important papers that made the foundation of what became the AC decomposition: [17], [18], [22, 23], [24]. In 1963 H. A. Curtis published a book "A New Approach to the Design of Switching Circuits" [19], where he formulated a generalized theory. The book is still one of the basic references to what is called in Western Hemisphere an "Ashenhurst-Curtis Theory of Decomposition". In 1962 Maitra introduces an important concept of a cascade in [21], and presents a synthesis approach being a particular specialization of the AC decomposition. These ideas have been next applied and generalized by many authors.

In 1963 a book by Maley [25] includes early ideas on "NOR Decomposition" (a kind of "Gate Decomposition"), mainly intuitive graphical methods of "inhibiting" functions by other functions. The book was a formal explanation of ideas introduced by industrial designers. It put together many useful methods applied in practice, thus differing from the AC decomposition which was treated as a more "academic" approach, with respect of the lack of availability of computer programs for it, the difficult access to computers and their very limited capabilities at the time. These ideas of the Maley's book have been very important, since they laid a groundwork for further research on practical Boolean Decomposition programs of Dietmeyer, Darringer, Stoffers, Dietmeyer, Schneider, Duley and many other. Years 1964-1967 bring further developments in all three "schools" on decomposition: Ashenhurst-Curtis, Roth-Karp, and Maley. R.E. Prather publishes one of the first, if not the first, general logic synthesis textbooks, that mentions the Ashenhurst decomposition; "Introduction to Switching Theory: A Mathematical Approach" [27].

In 1969 V.Y. Shen defends a Ph.D. thesis in the Department of Electrical Engineering at Princeton, [30], where he uses for the first time the idea of Reed-Muller form for the representation of Boolean functions in the decomposition process, in order to test the bound sets more efficiently. He shows thus for the first time that the idea of decomposition is not related to tabular representations of Ashenhurst and Curtis. His thesis introduced also few other important ideas that all demonstrate how the better representation of the problem and its deeper mathematical analysis allow to reduce the space of variable partitions. Still, the computers of that time were not powerful enough to make his groundwork practical.

In 1970 Kellar and Shen publish [32]; their algorithm could provide a quick way to determine if the children decompose. The article by Walliuzzaman and Vranesic [33] deals for the first time with functions that have more than two possible values of both input and output variables, thus starting the sub-area of the *multiple-valued decomposition.* Since then, more papers have been published on various approaches to multiple-valued (mv) logic decomposition, that roughly follow the three lines of the AC, RK and Maley decompositions. However, the area of multiple-valued decomposition still remains much less developed than the binary decomposition, perhaps because mv decomposition continues to be linked by most researchers to circuit design, and not treated as a general "algorithm discovery" method, also for software applications. The same year, S.L. Height defends his Ph.D. dissertation which is the first systematic study of decomposition for incompletely specified Boolean functions (These results were next published in [39]). Two fundamental papers are also published, [29, 28], which influence the "gate" based approach to Boolean decomposition.

In 1971, Shen, McKellar, and Weiner publish an important paper "An fast algorithm for the disjunctive decomposition of switching functions," [36], which builds on the success of their earlier approach. Finally, the same year brings the book by Dietmeyer, [34], which introduces systematically two important ideas: cube calculus - an efficient representation of Boolean functions and operations on them, used next in many successful logic synthesis computer programs; and a new form of decomposition, that can be treated as a new "school" of thought in decomposition, "Dietmeyer Decomposition", an addition to all AC, RK and Maley approaches. The programs developed there became prototypes of many programs developed in the next years worldwide, and influenced also the state of the art in industry, originally in IBM. In 1971 Thayse published an early attempt at purely algebraic approach to decomposition, based on the concept of a Boolean derivative. The paper by Thayse and Davio from 1973, "Boolean differential calculus and its application to switching theory" [40], proved with time to be a very influential one. H.A. Curtis published in 1976 an important paper: "Simplified decomposition of Boolean functions," [44]. The new problem representation introduced in this paper influenced the very successful line of research represented recently by Luba et al. [58, 90, 107, 99]. The same year M.G. Karpovsky [46] publishes the first complete exposition of spectral approach to logic synthesis, "Finite orthogonal series in the design of digital devices". This book presents also a new approach to AC decomposition, and introduces the "Decomposition with EXOR Pre- and Post- Processors" for the first time. Finally, Kandel publishes in 1976 the first paper that related decomposition to Fuzzy Logic, [45].

In 1978 Butler [47] uses partition matrix to design "tandem networks", as a generalization to cascades. In another paper, "Analysis and design of fanout-free networks of positive symmetric gates" [48], he discusses use of symmetric functions. Thus, he discusses another two important special circuit architectures devised for realizing decomposed functions. In this way, Butler enhanced the idea of the decomposed structures of restricted form, which was next followed by several authors. In a very important book by Davio, Deschamps and Thayse, "Discrete and Switching Functions" [49], they further refine and extend the concepts of differential operators and their various uses, including the functional decomposition. Although very mathematical and difficult to read, this book is a source of many fundamental and deep ideas in logic synthesis, and should be recommended to any serious researcher in decomposition, Exor logic, spectral methods, or orthogonal transforms. In 1980 the first paper of Bibilo's school in Belorussia: [51], is published. Although their approach follows in principle the ideas of Roth-Karp and Zakrevskii, they created many ingenious combinatorial algorithms for the variable partitioning and encoding problems. Especially interesting is an algorithm that combines column compaction and encoding into a single process. This year marks also a beginning of an interest in applying spectral methods to various topics related to decomposition, [52, 53].

After increased interest in logic synthesis in the 80s, most research in multi-level circuits in the nineties was on algebraic factorization methods. However, starting in 1987, there is also an increased interest in Boolean decomposition, which has been reflected in the number of published papers. Several groups that have been active in future years, had published their first papers this year. The approach by Luba et al. introduces a totally different representation of the problem: for each input and output variable, the partition of cares to the different values of this variable is stored (k blocks of partition for k values of the variable). This representation allows for efficient variable partitioning, and has also several other advantages in various decompositions.

Paper [66] is a good reference to modular and multiple-valued decomposition. Perkowski and Brown [67], explain Ashenhurst/Curtis decomposition in a simple way, starting from synthesis with multiplexers. Next, a cube-based method for strongly unspecified method is given. The column minimization problem has been reduced here to graph coloring for the first time. In 1991 Varma and Trachtenberg [82] discuss the problem of evaluating complexity of a Boolean function, which is very important in decomposition for partition selection. As observed already by Yablonskyii [15], an accurate evaluation of complexity may be as expensive as minimization of the function with the respective method. There exist functional decompositions procedures Karpovsky [46] that use abstract complexity criteria based on functional approximations to logic complexity that arrive at optimal solutions by analytical methods.

An important paper by Yang and Ciesielski [81] is devoted to a Boolean decomposition of PLAs. Although Ashenhurst and Curtis work is never mentioned here, the paper introduces powerful methods to find variables' partitioning that are not exhaustive. The technique is applied there to PLA decomposition, but it can be adopted to general functional decomposition as well. Especially, the presented by them ideas may be useful to quickly find good-quality encodings during the decomposition.

A paper of Berkeley group, Murgai, Nishizaki, Shenoy, Brayton, and Sangiovanni-Vincentelli, "Logic Synthesis for Programmable Gate Arrays," [73] signifies an increased interest of this group, which was predominantly occupied with algebraic method in the past, with Boolean methods. This is the first of their several important papers, that are mostly devoted to Xilinx's Look-up Table model. Since 1991, Perkowski et al developed the so-called orthogonal decompositions [79, 84, 92] that have several similarities with the AC decomposition, but are based on the generalized spectral theory. It seems that such decompositions can be also incorporated into the framework of the proposed generalized decomposer. As evidenced by the amount of papers, in the last few years, especially since 1992, the increase of interest in Boolean Decomposition is significant. These papers will be presented in respective topical sections below.

6 Selected Approaches.

6.1 Varma and Trachtenberg Approach to AC Decomposition.

The main drawback of using the basic method of Disjoint Decomposition is to check $2^n - n - 2$ "maps" for n-input function. However, it is possible to devise algorithms that use necessary conditions for the existence of decompositions to prune out certain combinations of input variables that do not belong to any decomposition, effectively speeding up the procedure in all but pathological cases. Shen and McKellar devised such an algorithm that detected candidate partitions for disjoint decompositions of logic functions but required further testing of candidates, though the set of candidates contained far fewer partitions that the original set of all possible partitions. They also showed that the RM canonical form was easier to test for decompositions than the disjunctive normal form (DNF).

Varma and Trachtenberg presented an algorithm [71] that detects disjoint decompositions by inspecting the RM spectrum. The spectral method is more efficient than Boolean procedures in many cases, but requires the additional step of spectrum computation Their algorithm is only a variation of the one from Shen, but it incorporates both the necessary and the sufficient conditions for the existence of decompositions in one procedure. It also tests for different sizes of partitions successively.

6.2 PLA Decomposition.

Several authors, including Ciesielski and Yang [72, 81], Devadas [65, 68], and Sasao [64, 70, 96] proposed to apply Boolean decomposition to decompose a PLA into two or more cascaded PLAs. However, from the point of view of the mathematical apparatus used, some of these methods have little in common with classical functional decomposition problem formulation, and can be rather treated as multiple-valued minimization or input-encoding problems. Nevertheless, the techniques introduced by them are very efficient, and, because of certain similarities of decompositiona and multiple-valued logic, they should be considered by the proponents of the classical AC model as a possible source of additional model-enhancing ideas. Although their methods have been presented entirely in the PLA technology framework, they can be also modified to other types of "macro-blocks" or "universal cells".

Tsutomu Sasao [70] presented a method to decompose a PLA in a serial way. This resembles a classical Curtis decomposition with all predecessor blocks combined to a single multi-output block. Sasao divides the problem to two subproblems: *a partition problem* and *an encoding problem.* His first stage is very similar to all multiple-valued Curtis-like decomposition algorithms, with the only difference that the entries in the table are still binary (he discusses a multiple-valued input, binary output functions). Since this is a Curtis-like decomposition, he allows the column multiplicity index value to be higher than 2. In order to find a simple circuit he looks for the partition with the smallest multiplicity. For n variable function there are 2^n different partitions. When n is small ($\leq$16) he uses brute force method. For larger n he uses the following approach. Denote by n_1 the number of variables in bound set. He represents each column of the decomposition chart by a logical expression. The number of expressions in the chart is 2^{n_1}. Therefore when n_1 is small the decomposition chart can be represented with small memory storage. He calculates n_1 from value 2 to NE (a user parameter). An equivalence of two columns is checked as an equivalence of two logical expressions.

Sasao was one of the first authors who discussed the encoding problem in the framework of functional decomposition. He experimented with one-hot encoding and minimum-length encoding and calculated total sizes of predecessor and successors PLAs. He used a very simple encoding algorithm - "the more frequent the pattern occurs in the decomposition chart, the more the number of 0's in the code". He found that the one hot encoding worked better than the minimum-length encoding. He concludes also with the desire of finding a good heuristic method to find a good partition more efficiently and to find a better heuristic for the minimum-length encoding.

6.3 The Approach of Steinbach et al.

The approach of Steinbach et al. is the successor of many years of research of Bochmann, Posthoff and Dressing. Bochmann himself was influenced by the work of Zakrevskij's group, one of the three top logic synthesis groups in the former Soviet Union. Zakrevskij was the creator of the original algebraic ideas of this line of research; such as the definitions of the differential, maximum and minimum [37]. Some of them may go back perhaps to early papers of Davio, Thayse and Deschamps. All recent ideas of this German group are implemented in a programming system XBOOLE, therefore we will refer to the *XBOOLE approach.*

There are five properties that are interesting about the approach of XBOOLE:

- The totally algebraic formulation of the decomposition problem. Definition of Boolean operators that are powerful and universal, do not depend on the actually used representation of analytical functions used, and, being purely Boolean concepts, can be implemented in other representations. This can be taken advantage of by reformulating these operators for multiple-valued, fuzzy or other logic - the general decomposition method will remain the same.

- Efficient use of Ternary Vector Lists, a kind of Cube Calculus, in which all the cubes are disjoint. They are being made disjoint ("orthogonalized"), immediately after every operation that may produce non-disjoint cubes.

- The decomposition used by XBOOLE is actually not an AC decomposition, since each function is decomposable by it to two-input gates: AND, OR and EXOR, and possibly also Inverters. This approach allows to decompose large functions relatively quickly. Contrary to other "Gate Decomposition" methods, we discuss this system, however, here, because several ideas are still very close to AC decomposition (for instance, the general decomposition models or the variable partitioning), and because we believe that its methods are general and can be adapted for the purposes of Flash. We will call them "Generalized XBOOLE Decompositions". Such decompositions can be some of the "Special Bias" Decompositions in the future decomposer. XBOOLE Approach is much closer to AC decomposition than other "gate-oriented" Boolean decompositions (like those of Zaky, Stoffers, Dietmeyer, Duley, Darringer, and orthogonal decompositions by Perkowski). Its relationships to AC decompositions and especially Bibilo, Luba and Pedram researches must be studied in more detail.

- Every step in the decomposition can have a large amount of different results because of different algorithms, different basis elements, and the ambiguity of the decomposition. The program allows then to be controlled by some parameters in such a way that the result optimally meets user-requirements of some kind. Thus, it can be tuned to various technological parameters, such as area, number of gates, depth of the circuit, power consumption and testability.

- The method can handle efficiently don't cares, but not as efficiently as the approaches of Luba or Pedram.

6.4 FPGA synthesis.

A renewed interest in AC decomposition in recent years is caused by the introduction of Look-up Table Field Programmable Gate Arrays (FPGAs) by Xilinx in 1986 [83], and other companies in succeeding years (AT and T, Actel, Motorola, Algotronix). When it was found that the adaptation of earlier algebraic methods does not work properly for the Lookup Table model, researchers switched to Boolean decomposition. Currently, in most of the systems the decomposition is only an auxiliary process (MIS-PGA by Murgai et al [73, 78], HYDRA by Filo, Yang, Mailhot, and de Micheli 1991).

However, with the arrival of systems such as TRADE, as well as the Luba's system, that are primarily based on decomposition, the situation is becoming to change drastically. Even the U.C. Berkeley group

of Professor Brayton, who invented the algebraic factorization approach and were for long time its most stubborn adherents, are recently devoting more attention to Boolean decomposition. The approach presented in [73, 78] presents improved algorithms for Table Look Up Architectures which were included in the well-known program MIS-PGA(new). It uses several techniques, and Roth-Karp decomposition is only one of them. They use also other decomposition techniques such as: cofactoring and AND-OR decomposition. Infeasible node is one that cannot be realized by a single CLB. A feasible function can be realized by a single CLB. The decomposition phase starts from a possibly infeasible network and creates a feasible network. Kernel extraction and Roth-Karp decomposition are used. In kernel extraction, kernels are extracted from an infeasible node. The node and the kernel are then recursively decomposed. In the RK decomposition, a bound set X of cardinality m is chosen from the fanins of the infeasible node n.

6.5 Kim and Dietmeyer's Approach to Symmetric Decomposition.

Taking advantage of symmetry in multilevel synthesis requires detection of symmetry sets of multiple-output function F. Extensive work on the detection of symmetry sets using the array format was reported by [26]. Kim and Dietmeyer [77], observed that current systems for multilevel synthesis from PLA-like macros have a difficulty with symmetric functions. Designs of totally symmetric functions generated from descriptions that lack global network structure, have on average more than twice as many literals as the best designs, while the designs of nonsymmetric functions have on average 20% more literals than the best designs. Classic disjoint decomposition, which is generally considered for small networks, is particularly suitable for symmetric function synthesis.

The Boolean PLA decomposers of Devadas [65, 68] and Ciesielski [72, 81] overcome the two most important drawbacks of algebraic factorization methods: too much dependency on given expressions and the sequential generation of subfunctions. However, they have no special treatment for symmetric functions. Kim and Dietmeyer [77] developed a simple but effective heuristic method for synthesizing symmetric functions, that detects and takes advantages of symmetry and is based upon classical disjoint decomposition theory. Their programs use new general cube calculus operators developed by them especially for symmetry checking, and produce almost always the best designs known to the authors. It is the feeling of this author that with the new theory of symmetry recognition, that uses decision diagrams [108] the approach of Kim and Dietmeyer can be made even more efficient.

6.6 The Approach Based on Decision Diagrams.

Reduced, ordered Binary DDs (BDDs) provide a compact and canonical representation of Boolean functions. These data structures are used in the decomposition by Lai, Pedram and Vrudhula [89]. They use also their new data structure, called Edge-Valued Binary DDs (EVBDDs). EVBDDs provide a representation of Boolean functions over the integer domain and have been shown to be useful for the verification of arithmetic functions. Another similar and even more powerful decision diagrams have been recently created by Brayant and Chen, 1994, called Binary Moment Diagrams [104]. Computationally, the approach of Lai, Pedram and Vrudhula (Sastry) proposed in DAC '93 [89] is the most successful approach to AC decomposition. Their approach is applicable to both disjunctive and nondisjunctive decomposition, both completely and incompletely specified Boolean functions, single-output and multi-output. Their general theory uses a new algorithm, based on EVBDD representation, for generating the set of all bound variables that make the function decomposable. Their method for detecting simple disjunctive decomposition is based on the concept of a cut-set in BDD representation of Boolean functions. The algorithm is not using any new special properties, and it seems that the only reason that this algorithm is so successful, is the application of new representations - BDDs and EVBDDs.

7 Recommendation for the Future Decomposer for Pattern Theory Group.

During the course of this work, the following ideas have been created and integrated, some of which have its original sources in the surveyed above literature.

1. The most important aspect of a successful Functional Decomposer is the representation of data. From all representations in the literature, the most successful programs use Binary Decision Diagrams and Edge-Valued Decision Diagrams [89], lists of disjoint ON and OFF cubes [111], and variable-partitions on ON and OFF minterms [107]. The last representation has very interesting properties and allows to define, on incompletely specified functions, several operators, that have not been defined before, allowing thus for efficient variable partitioning and column compaction algorithms, as well as for formulating new kinds of decompositions, including a new method of "variable re-using". Additional ådvantage of this representation is its easy generalization for multiple-valued and continuous logic. On the other hand, the disadvantage of this approach is the use of big lists, thata are necessary to represent the ON and OFF sets of minterms. Hence comes our idea of encoding these large sets as two BDDs: the ON BDD, and the OFF BDD. All such sets for all input and output variables, together with the newly created sub-functions, can be represented as a "Shared" BDD. Such representation is totally new in the literature, and has all advantages of the previous ones, and none of their respective disadvantages. A ready public domain software package can be used to represent "reduced, ordered, shared BDDs with negated edges", since the only operations that we use are the standard ones: intersection, union and complement.

2. The representation allows to treat all newly created sub-functions as additional inputs variables, by calculating the partitions done by these functions on ON and OFF cares. Then, all new variables are treated on the same terms as the initial input variables - the analysis of vacuous variables and variable partitioning are applied to them. This allows for powerful function sharing and investigating a wider space of decompositions.

3. Another advantages of successful modern decomposers that are missing in Flash are: multi-output functions, and multiple-valued functions. A careful analysis of all operations used in the respective systems, suggests that the proposed above representation will allow easy generalization to multi-output and multiple-valued functions. In essence, it is based on approaches that have been already created or generalized for these cases. Moreover, our representation is well suited not only for AC decomposition, but for many other decomposition variants [111, 89, 107, 51, 62, 38, 43, 47, 48, 26, 34, 77]. Especially, it allows to take into account the symmetry, as a most natural design bias.

4. One more advantage of the representation is that it has been created to solve efficiently the two most important problem in decomposition: selecting best bound sets, and column compaction/coding problem. The sets of minterms are kept separated in the decomposition; for each input variable, output variable and sub-function. This representation allows for easy formulating of many partial problems associated with decomposition. Many heuristic or methodic criteria to select variables in order to achieve some separation goals can be easily defined. All information, that is lost and repeatedly found in classic decomposition algorithms, is retained here for further use.

5. All NP-hard combinatorial problems have been reduced here to **a single one**: the weighted covering problem. Excellent algorithms exist [69, 107] to solve this problems, and many publications are devoted to it.

6. Several search strategies have been used in the literature to find the best partitions for both the disjoint and non-disjoint cases. We propose to create a parametrized program that will allow to easily create and compare several of them: depth-first search; depth-first-search-with-one-successor; breadth-first search (a variant in Flash by Mark Axtell); best-bound (Haomin Wu/Perkowski [101]); limited best-bound search (Haomin Wu, Craig Files [106]); A* algorithm of Artificial Intelligence (Haomin Wu, Perkowski); searching subsets of a subset of variables (Shen/Kellar [30, 32, 36], Varma/Trachtenberg [71]); using predetermined order based on a heuristic of evaluating pairs of variables (cube based approach in presented by Wei Wan and Perkowski [86]); search by "add and subtract" variables from and to bound and free sets - Steinbach [100, 111]; search by selecting a single variable, based on complex

block separation criteria of partition calculus heuristics (Luba [58, 61, 90, 107]); other BDD variable selection and ordering methods [105, 98, 93]; and other decision tree like variable selection methods (such as in the C4.5 program by Quinlan).

References

[1] Boole, G., "An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities," *London*, 1854.

[2] I.I. Zhegalkin, "O Tekhnyke Vychyslenyi Predlozhenyi v Symbolytscheskoi Logykye," *Mat. Sb.*, Vol. 34, pp. 9-28, 1927 (in Russian).

[3] C.E. Shannon, "The Synthesis of Two-Terminal Switching Circuits," *The Bell System Technical Journal*, 28, pp. 59-98, 1949.

[4] R.L. Ashenhurst, "The Decomposition of Switching Functions," *Bell Laboratories Report*, No. BL-1(11), pp. 541-602, 1952.

[5] W.L. Semon, "Characteristic Numbers and Their Use in the Decomposition of Switching Functions," *Proc. ACM*, vol. 17, pp. 273-280, May 1952. (In Curtis'62).

[6] R.L. Ashenhurst, "Non-Disjoint Decomposition," *Bell Lab. Report*, No. 4., pp. 1V-1 - 1V-12, 1953.

[7] T. Singer, "The Decomposition Chart as a Theoretical Aid", *Harvard Comp. Lab*, Cambridge, Mass., Rept. BL-4, 1953.

[8] G.N. Povarov, "About Functional Separability of Boolean Functions", Doklady Akad. Nauk SSSR, Vol. 94, No. 5, pp. 801-803, (in Russian), May 1954.

[9] I.S. Reed, "A Class of Multiple-Error-Correcting Codes and the Decoding Scheme", *IRE Trans. on Inform. Th.*, Vol. IT-4, pp. 38-49, Sept.1954.

[10] R.L. Ashenhurst, "The Decomposition of Switching Functions," *Bell Telephone Labs Report*, 16, pp. III1-III72, 1956.

[11] H.A. Curtis, "Non-Disjunctive Decompositions," *Th. Swi. Bell Labs Rep.* No. 19, Sect. II, 1, 49, 1958.

[12] S.B. Akers, "On a Theory of Boolean Functions," *Journal of the Soc. for Ind. and Appl. Math.*, Vol. 7, No. 4, 1959.

[13] H.A. Curtis, "A Functional Canonical Form," *J. ACM*, 6, pp. 245-258, 1959.

[14] R.L. Ashenhurst, "The Decomposition of Switching Fuctions," *Proc. of an Intern. Symp. on the Th. Swi.* Apr. 2-5, 1957, Ann. Comp. Lab, Harv. Univ, Vol. 29, pp. 74-116, 1959.

[15] S.W. Yablonskii, "On Algorithmic Obstacles to the Synthesis of Minimal Contact Networks," *Problemy Kibernitiki*, No. 2, pp. 75-121, 1959, (in Russian).

[16] J.P. Roth, "Minimization over Boolean Trees," *IBM Journal*, Vol. 4, No. 5, pp. 543-555, 1960.

[17] H.A. Curtis, "A Generalized Tree Circuit", *J. ACM*, pp. 484-496, 1961.

[18] R.M. Karp, F.E. McFarlin, J.P. Roth, J.R. Wilts, "A Computer Program for the Synthesis of Combinational Switching Circuits," *In Proc. AIEE Ann. Symp. on Swi. Circ. Th.*, pp. 182-194, 1961.

[19] H.A. Curtis, "A New Approach to the Design of Switching Circuits", *Princeton,* N.J., Van Nostrand, 1962.

[20] H.A. Curtis, "Generalized Tree Circuit - The Basic Building Block of an Extended Decomposition Theory," *J.ACM,* 10, pp. 562-581, 1962.

[21] K.K. Maitra, "Cascaded Switching Networks of Two-Input Flexible Cells," IRE Trans. Comp. EC-11, pp. 136-146, 1962.

[22] J.P. Roth, R.M. Karp, "Minimization over Boolean Graphs," *IBM J. Res. Develop.* pp. 227-238, April 1962.

[23] J.P. Roth and E.G. Wagner, "Algebraic Topological Methods for the Synthesis of Switching systems. Part III: Minimization of Non-singular Boolean trees," *IBM J. Res. Develop.* Vol. 3, Oct. 1962.

[24] R.M. Karp, "Functional Decomposition and Switching Circuit Design," *J. Soc. Industr. Appl. Math.*, Vol. 11, No. 2, pp. 291-335, June 1963.

[25] G. Maley, J. Earle, "The Logic Design of Transistor Digital Computers," Englewood Cliffs, N.J. Prentice Hall, 1963.

[26] D.L. Dietmeyer, P.R. Schneider, "Identification of Symmetry, Redundancy and Equivalence of Boolean Functions," *IEEE Trans. Electron. Comp.* Vol. 16, pp. 804-817, Dec. 1967.

[27] R.E. Prather, "Introduction to Switching Theory: A Mathematical Approach," 1968.

[28] P.R. Schneider, D.L. Dietmeyer, "An Algorithm for Synthesis of Multiple-output Combinational Logic," *IEE Trans. on Comp.*, pp. 117-128, Febr. 1968.

[29] E.S. Davidson, "An Algorithm for NAND Decomposition Under Network Constraints," *IEEE Trans. on Comp.,* pp. 1098-1109, Dec.1969.

[30] V.Y. Shen, "On Simple Disjunctive Decompositions of Switching Functions," Ph.D. thesis, *Dept. Electr. Engng., Princeton University,* N.J., January 1969.

[31] S.L. Height, "Complex Disjunctive Decomposition of Incompletely Specified Boolean Functions," *Ph.D. dissertation,* Univ. New Mexico, Albuquerque, July 1970.

[32] V.Y. Shen, and A. C. McKellar, "An Algorithm for the Disjunctive Decomposition of Switching Functions," *IEEE Trans. on Computers,* Vol. C-19, pp. 239-248, March 1970.

[33] K.M. Walliuzzaman, Z.G. Vranesic, "On Decomposition of Multiple-valued Switching Functions," *Computer Journal,* Vol. 13, pp. 359-362, 1970.

[34] D.L. Dietmeyer, "Logic Design of Digital Systems," *Boston: Allyn and Bacon,* 1971.

[35] S.R. Das, C.L. Sheng, "On Detecting Total or Partial Symmetry of Switching Functions," *Trans. on Comp.*, pp. 352-355, March 1971.

[36] V.Y. Shen, A. C. McKellar, and P. Weiner, "A Fast Algorithm for the Disjunctive Decomposition of Switching Functions," *IEEE Trans.on Comp.,* Vol. C-20, No. 3, pp. 304-309, March 1971.

[37] A.D. Zakrevskij, "Algoritmy Sinteza Diskretnych Avtomatov, Moscow 1971. (in Russian)

[38] Butler, J., K.J. Breeding, "Some Characteristics of Universal Cell Nets," *IEEE Trans. on Comp.,* pp. 897-903, Oct. 1973.

[39] S.L. Height, "Complex Disjunctive Decomposition of Incompletely Specified Boolean Functions," *IEEE Trans. on Comp.,* Vol. C-22, No. 1, Jan. 1973.

[40] A. Thayse, M. Davio, "Boolean Differential Calculus and its Application to Switching Theory," *IEEE Trans. on Comp.*, C-22, No. 4, pp. 409-420, 1973.

[41] M. Elmasry, "Logic Partition for Multi-Emitter 2-Level Structures," *IEEE Trans. on Circ. and Syst.,* Vol. 21, No. 3, pp. 354-359, 1974.

[42] A. Thayse, "Boolean Differential Calculus and its Application to Switching Theory," *IEEE Trans. on Comp.*, Vol. C-22, April 1974.

[43] J.T. Butler, "On the Number of Functions Realized by Cascades and Disjunctive Networks," *IEEE Trans. on Comp.,* Vol. 24, No. 7, pp. 681-690, 1975.

[44] H.A. Curtis, "Simplified Decomposition of Boolean Functions," *IEEE Trans. on Comp.* Vol. 25, pp. 1033-1044, Oct.. 1976.

[45] A. Kandel, "Decomposition of Fuzzy Functions," *IEEE Trans. on Comp.* Vol. 25, No. 11, pp. 1124-1130, 1976.

[46] M.G. Karpovsky, "Finite Orthogonal Series in the Design of Digital Devices," *J. Wiley,* New York, 1976.

[47] J.T. Butler, "Tandem Networks of Universal Cells," *IEEE Trans. on Comp.,* Vol. 27, No. 9, pp. 785-799, 1978.

[48] J.T. Butler, "Analysis and Design of Fanout-free Networks of Positive Symmetric Gates," *J. of the Assoc. for Comp. Mach.,* Vol. 25, No. 3, pp. 481-498, 1978.

[49] M. Davio, J. P. Deschamps, and A. Thayse, "Discrete and Switching Functions," *Mc.Graw-Hill,* 1978.

[50] C.R. Edwards, S.L. Hurst, "Digital Synthesis Procedure Under Function Symmetries and Mapping Methods," *IEEE Trans. on Comp.,* Vol. 27, No. 11, pp. 985-997, 1978.

[51] P.N. Bibilo, S.V. Yenin, "Decomposition of a Boolean Function with Minimum Number of Significant Arguments of the Subfunctions," *Engn. Cybernetics,* Vol. 18, No. 3, pp. 75-81, 1980.

[52] A. M. Lloyd, "Design of Multiplexer Universal-Logic-Module Networks Using Spectral Techniques," *IEE Proc. Pt. E.,* Vol. 127, pp. 31-36, Jan. 1980.

[53] V.H. Tokmen, "Disjoint Decomposibility of Multi-valued Functions by Spectral Means," *Proc. IEEE 10th International Symp. on Multiple Valued Logic,* pp. 88-93, 1980.

[54] A.J. Tosser, D. Aoulad-Syad, "Cascade Networks of Logic Functions Built in Multiplexer Units," *IEE Proc. Pt. E,* Vol. 127, No. 2, pp. 64-68, March 1980.

[55] E.A. Bender, E.R. Canfield, "Fanout-free Functions," *American Math. Monthly,* Vol. 88, No.4., pp. 278-280, 1981.

[56] R.K. Brayton, and C. Mc. Mullen, "The Decomposition and Factorization of Boolean Expressions," In *Proc. Intern. Symp. on Circ. and Syst.,* pp. 49-54, May 1982.

[57] J.M. Francioni, A. Kandel, "Decomposable Fuzzy-valued Switching Functions," *Fuzzy Sets and Systems,* Vol. 9, No. 1, pp. 41-68., 1983.

[58] T. Luba, "Synthesis of Combinational Circuits Using Boolean Decomposition Method," *Pub. of Institute of Telecomm.*, Warsaw Techn. Univ., No. 108, 1983, (in Polish).

[59] R.K. Brayton, G.D. Hachtel, C. McMullen, A. Sangiovanni-Vincentelli, "Logic Minimization Algorithms for VLSI Synthesis," *Kluwer Academic Publishers,* Boston, 1984.

[60] R.E. Bryant, "Graph-Based Algorithms for Boolean Function Manipulation," *Trans. on Comp.,* Vol. C-35, No. 8, pp. 667-691, 1986.

[61] T. Luba, "A Uniform Method of Boolean Function Decomposition," *Rozprawy Elektrotechniczne,* No. 4, pp. 1041-1054, 1986.

[62] P.N. Bibilo, S.V. Yenin, "Synthesis of Combinational Networks Using the Method of Functional Decomposition," *Nauka i Tekhn.* Minsk, 1987, (in Russian).

[63] M. Perkowski, H. Uong, "Generalized Decomposition of Incompletely Specified Multioutput, Multi-Valued Boolean Functions," *Unp. man., Dept. Electr. Engn.* PSU 1987.

[64] T.Sasao, "Functional Decomposition of PLAs", *Proc. of the Intern. Workshop on Logic Synthesis,* Research Triangle Park, NC, May 12-15, 1987.

[65] S. Devadas, A.R. Wang, A.R. Newton, A. Sangiovanni-Vincentelli, "Boolean Decomposition in Multi-Level Logic Optimization," *Proc. IEEE ICCAD,* pp. 290-293, 1988.

[66] K.Y. Fang, A.S. Wojcik, "Modular Decomposition of Combinational Multiple-valued Circuits," *IEEE Trans. on Comp.,* Vol. 37, No. 10, pp. 1293-1301, 1988.

[67] M. Perkowski, J. Brown, "A Unified Approach to Designs with Multiplexers and to the Decomposition of Boolean Functions," *Proc.ASEE Annual Conf.,* pp. 1610-1619. 1988.

[68] S. Devadas, A. Wang, A.R. Newton, and A. Sangiovanni-Vincentelli, "Boolean Decomposition in Multilevel Logic Optimization," *IEEE J. of Solid-State Circuits,* Vol. 24, pp. 399-408, April 1989.

[69] M.A. Perkowski, J. Liu, J.E. Brown, "Rapid Software Prototyping: CAD Design of Digital CAD Algorithms," In G. W. Zobrist (ed), *Progress in Computer-Aided VLSI Design,* Vol. 1, pp. 353-401, 1989.

[70] T.Sasao, "Application of Multiple-Valued Logic to a Serial Decomposition of PLAs", *Proc. of the Intern. Symp. on Multiple-Valued Logic,* Zangzou, China, pp. 264-271, May 1989.

[71] D. Varma, E.A. Trachtenberg, "Design Automation Tools for Efficient Implementation of Logic Functions by Decomposition," *IEEE Trans. on CAD,* Vol. 8, No. 8, pp. 901-917, August 1989.

[72] S. Yang, M. Ciesielski, "A Generalized PLA Decomposition with Programmable Encoders," In the *Proc. of the Intern. Workshop on Logic Synth.,* May 1989.

[73] R. Murgai, Y. Nishizaki, N. Shenoy, R.K. Brayton, A. Sangiovanni - Vincentelli, "Logic Synthesis for Programmable Gate Arrays," *Proc. 27th DAC.,* pp. 620-625, 1990.

[74] M.A. Perkowski, P. Dysko, and B.J. Falkowski, "Two Learning Methods for a Tree-Search Combinatorial Optimizer," *Proc. of IEEE Int. Conf. on Comput. and Comm.,* Scottsdale, AZ, pp. 606-613, 1990.

[75] L. Nguen, M. Perkowski, *Proc. Midwest Symp. on Circuits and Systems,* 1990.

[76] D. Bochmann, B. Steinbach, "Logikentwurf mit XBOOLE," *Verlag Technik,* Berlin, 1991, (in German).

[77] B.G. Kim, D.L. Dietmeyer, "Multilevel Logic Synthesis of Symmetric Switching Functions," *IEEE Trans. on CAD,* Vol. 10, No. 4, pp. 436-446, Apr. 1991.

[78] R. Murgai, N. Shenoy, R.K. Brayton, A. Sangiovanni-Vincentelli, "Improved Logic Synthesis Algorithm for Table Look Up Architectures," *Proc. IEEE ICCAD,* pp. 564-567, 1991.

[79] M.A. Perkowski, and P. D. Johnson, "Canonical Multi-Valued Input Reed-Muller Trees and Forms", *Proc. 3rd NASA Symp. on VLSI Design,* Moscow, ID, pp. 11.3.1 - 11.3.13., Oct. 1991.

[80] T. Ross, M.J. Noviskey, T.N. Taylor, D. Gadd, "Pattern Theory: An Engineering Paradigm for Algorithm Design," *Final Technical Report,* WL-91-1060, Wright Laboratory, USAF, WL/AAART, WPAFB, OH 45453-6543, August 1991.

[81] S. Yang, M. Ciesielski, "Optimum and Suboptimum Algorithms for Input Encoding and its Relationship to Logic Minimization," *IEEE Trans. Comp. Aid. Des.* Vol. 10, No. 1, Jan. 1991.

[82] D. Varma, E.A. Trachtenberg, "On the Estimation of Logic Complexity for Design Automation Applications," *1990 Conference of IEEE,* pp. 368-371, 1991.

[83] Xilinx Inc., "The Programmable Gate Array Data Book", "Xilinx Programmable Gate Array User's Guide," 1991.

[84] M. A. Perkowski, "The Generalized Orthonormal Expansions of Functions with Multiple-Valued Inputs and Some of its Applications," *Proc. 22nd ISMVL,* pp. 442-450, Sendai, Japan, May 1992.

[85] T.D. Ross, "Function Decomposition Strategy for the Function Learning and Synthesis Hotbed," *Techn. Mem. WL-TM-92-110,* Wright Lab, USAF, WL/AART, WPAFB, Aug. 1992.

[86] W. Wan, M.A. Perkowski, "A New Approach to the Decomposition of Incompletely Specified Multi-Output Function Based on Graph Coloring and Local Transformations and Its Application to FPGA Mapping," *Proc. of Euro-DAC,* pp. 230-235, 1992.

[87] J.F. Frenzel, "Application of Genetic Algorithms to Pattern Theory," *Final Report,* Summer Faculty Research Program, WPAFB, July 1993.

[88] S. He and M. Torkelson, "Disjoint Decomposition with partial vertex chart," In *Notes of the Intern. Workshop on Logic Synthesis,* Lake Tahoe, CA, pp. P2a-1-P2a-5, May 1993.

[89] Y.T. Lai, M. Pedram, S. Sastry, "BDD-based Decomposition of Logic Functions with Application to FPGA Synthesis," *Proc. of 30th DAC,* pp. 642-647, 1993.

[90] T. Luba, H. Selvaraj, A. Krasniewski, "A New Approach to FPGA-based Logic Synthesis," *Workshop on Des. Method. for Microelectr. and Sign. Proc,* pp. 135-142, Cracow 1993.

[91] T. Luba, M. Mochocki, J. Rybnik, "Decomposition of Information Systems Using Decision Tables," *Bulletin of the Polish Acad. of Sci., Techn. Sci.,* Vol. 41, No.3, 1993.

[92] M. A. Perkowski, "A Fundamental Theorem for Exor Circuits," *Proc. of the IFIP WG 10.5 Work. Applic. Reed-Muller Exp. in Circ. Des.* Hamburg, Sept. 1993.

[93] R. Rudell, "Dynamic Variable Ordering for Ordered Binary Decision Diagrams," *Proc. ICCAD,* pp. 42-47, 1993.

[94] T.D. Ross, M.J. Noviskey, M.L. Axtell, D.A. Gadd, "Flash User's Guide," *Techn. Report, Wright Lab.,* USAF, WL/AART, WPAFB, OH 45433-6543, Dec. 1993.

[95] T. Sasao (ed.), "Logic Synthesis and Optimization," *Kluwer Acad. Publ.,* 1993.

[96] T. Sasao, "FPGA Design by Generalized Functional Decomposition," in *T. Sasao, "Logic Synthesis and Optimization,"* Kluwer Acad. Publ., 1993.

[97] I. Schaefer, M. A. Perkowski, and H. Wu, "Multilevel Logic Synthesis for Cellular FPGAs Based on Orthogonal Expansions," *Proc. IFIP WG 10.5 Workshop on Applications of the Reed-Muller Expansion in Circuit Design,* Hamburg, Germany, Sept. 1993.

[98] I. Schaefer, M.A. Perkowski, "Synthesis of Multi-Level Multiplexer Circuits for Incompletely Specified Multi-Output Boolean Functions with Mapping Multiplexer Based FPGAs", *IEEE Trans. on Comp. Aided Design,* Vol. 12, No. 11, pp. 1655 - 1664, Nov. 1993.

[99] H. Selvaraj, A. Czerczak, A. Krasniewski, T. Luba, "A Generalized Decomposition of Boolean Functions and its Application in FPGA-Based Synthesis," *IFIP Workshop on Logic and Architecture Synthesis,* pp. 147-166, Grenoble 1993.

[100] B. Steinbach, F. Schumann, M. Stoeckert, "Functional Decomposition of Speed Optimized Circuits," In *Auvergne, D., Hartenstein, R. (ed) "Power and Timing Modeling for Performance of Integrated Circuits,"* Proc. of the Third Intern, Workshop on Power and Timing Modeling and Optimization, IT Press Verlag, Bruchsal, 1993.

[101] H. Wu, M.A. Perkowski, "Synthesis for Reed-Muller Directed-Acyclic-Graph Networks with Applications to Binary Decision Diagrams and Fine Grain FPGA Mapping," *Proc. of IWLS '93,* Tahoe City, CA, May 1993.

[102] N. Song, M.A. Perkowski, "EXORCISM-MV-2: Minimization of Exclusive Sum of Products Expressions for Multiple-Valued Input Incompletely Specified Functions," *Proc. ISMVL'93,* pp. 132-137, CA, May 24-27, 1993.

[103] M. Breen, T.D. Ross, M.L. Axtell, "Computing Column Multiplicity in Function Decomposition," *report, WPAFB,* 1994.

[104] R. Bryant, "Moment Decision Diagrams", CMU Report, June 1994, draft.

[105] R. Drechsler, A. Sarabi, M. Theobald, B. Becker, M.A. Perkowski, "Efficient Representation and Manipulation of Switching Functions Based on Ordered Kronecker Functional Decision Diagrams," *Proceedings of DAC '94,* San Diego, CA, June 1994.

[106] C. Files, "Using a search heuristic in an NP-complete problem in Ashenhurst-Curtis Decomposition," Final Report, Wright Labs, Aug. 1994.

[107] T. Luba, R. Lasocki, "Decomposition of Multiple-valued Boolean Functions," *Appl. Math. and Comp. Sci.,* Vol. 4, No.1, pp. 125-138, 1994.

[108] M. Marek-Sadowska, "Detecting Symmetric Variables in Boolean Functions using Generalized Reed-Muller Forms," *Proc. ISCAS'94,* pp. 287-290, 1994.

[109] M. J. Noviskey, T.D. Ross, D.A. Gadd, M. Axtell, "Application of Genetic Algorithms to Function Decomposition in Pattern Theory," *Wright Labs, report WL-TR-94-1015,* 1994.

[110] T.D. Ross, J.A. Goldmann, D.A. Gadd, "On the Decomposition of Continuous Functions," *In Third Intern. Workshop on Post-Binary VLSI Systems,* 1994.

[111] B. Steinbach, M. Stoeckert, "Design of Fully Testable Circuits by Functional Decomposition and Implicit Test Pattern Generation," *Proc. 1994 IEEE Test Conference,* 1994.

A STUDY ON VIRTUAL MANUFACTURING

R. Radharamanan
Director of Advanced Manufacturing Center
College of Engineering

Marquette University
1515 West Wisconsin Avenue
Milwaukee, WI 53233

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

A STUDY ON VIRTUAL MANUFACTURING

R. Radharamanan
Director of Advanced Manufacturing Center
College of Engineering
Marquette University

Abstract

Emerging defense acquisition strategies require the capability to prove the manufacturability and affordability of new weapons systems prior to the commitment of large production resources and/or to shelving the system for restart in potential future threats. Loosing the manufacturing capability and experience in an era of "near zero" production is a major risk in the current defense environment. Maintaining the state-of-the-art manufacturing proficiency without actually building/manufacturing the weapons systems is a major challenge. Virtual Manufacturing (VM) meets the above challenges by providing the capability, in essence, to continue manufacturing in the virtual world of the computer. Through the use of distributed manufacturing modeling and simulation, VM enables the enterprises to evaluate the producibility and affordability of new product and/or process concepts with respect to risks, their impacts on manufacturing capabilities, production capacity, and cost. VM can also provide accurate and realistic means to predict schedule, cost, and quality; address affordability as an iterative solution; and bridge the gap between engineering (design) and manufacturing in an interactive fashion.

In this paper, a review of literature, the basic concepts, tools required, application areas, and the future of virtual manufacturing are presented. The actions necessary to implement and incorporate VM into the business processes are identified. Also, the benefits, costs, limitations, and risks associated with adopting VM are highlighted.

A STUDY ON VIRTUAL MANUFACTURING

R. Radharamanan

Introduction

The new acquisition strategies of the Department of Defense (DoD) place emphasis on the ability of manufacturing organizations to integrate military and civilian products (dual use) in low volume high variety production. This leads to reduction in one-time product development costs, capable of quick and easy change between successive products, and scheduling and control of many independent activities in a common production facility. Reduced budgets for new acquisitions require the ability to quickly identify proposed weapon system alternatives and reliably estimate the risks and costs associated with each alternative. Significant improvement in the manufacturing infrastructures are required to realize these concepts. Development of Virtual Manufacturing Systems (VMS) will significantly enhance these infrastructures. VM is based on the integrated application of modeling, simulation, tools, and technologies to enhance product/process design and production decisions with appropriate control at all process levels.

Virtual Manufacturing is one of the key technologies which allows to go beyond the assumptions driving the old acquisition strategies. It provides the following fundamental changes: VM can be used to "prove out" the production processes, resulting in "pre-production hardened systems" - i.e., the systems which are developed and verified but never actually undergo actual production runs; VM can support the generation of more reliable estimates of production schedule and costs since the models are based on actual processes and not just parametrics; through modeling and simulation, VM can significantly improve production flexibility, and hence, reduce the fixed costs; and VM can substantially improve the decision making process of acquisition managers by reliably predicting schedule, risks, and costs.

Air Force ManTech coordinates a program on "Virtual Manufacturing eXperimental Consortium" (VMXC), involving air force, army, navy, research centers, industries and academia to accelerate VM developments. Also, in coordination with the Joint Directors of the Laboratories (JDL) a new initiative in VM is being planned.

In July 1994, the First Virtual Manufacturing User Workshop {20} was organized in Dayton, Ohio to encourage the users from government and industry to discuss the issues relative to potential roles of modeling and simulation in manufacturing. The workshop established directions for a series of follow-on technical workshops in the near future. This paper is intended to provide an understanding of the basic concepts, benefits, and applications of Virtual Manufacturing.

Literature Review

McLean et al. {8} developed a VM cell as part of the control software for the Automated Manufacturing Research Facility (AMRF) at National Bureau of Standards. A shop based upon virtual manufacturing cells can provide greater flexibility than existing shop configurations through the time sharing of machining workstations. In this configuration, virtual cells are data files and processes in the control computer. The virtual cells are used to perform analysis, reporting, routing, scheduling, dispatching, and monitoring.

Mills et al. {9} addressed VM as a concept in which computer based tools help the designer analyze and visualize how his design will be built and tested on the shop floor or in a flexible robotic assembly cells. The thrust of this approach is to capture and use information generated (and usually discarded) during the design phase to provide - among other things - analysis of components, assembly sequence and task plans, and animation of the results of these plans where appropriate. This eventually will facilitate rapid prototyping of hardware.

Pentland and William {14} presented that VM concepts can fundamentally change and improve the design process by allowing iterative exploration of the space of the valid designs and permit real-time structural and dynamic analysis of the entire system as it is being designed.

Fujita et al. {1} described the development and functions of the 3 Ps (production planning partner) model-based real-time scheduling system. The discussions include the integration of CAD/CAM, virtual manufacturing and production models.

Mills et al. {10} presented a VM workbench by developing an integrated set of tools with which a team of engineers can design, program, test and debug complex manufacturing systems. The approach taken by them includes: developing a manufacturing system with individual tools, developing general requirements for an integrated system, defining an architecture that meets those requirements, and developing a prototype to test the architecture and test and refine the requirements.

Staffend {17} used the term "Virtual Factory" to identify an "Agile Manufacturing Enterprise". He also indicated several components of agile production which can be realized using products and tools already available and well proven. Some of them are: artificial intelligence, adaptive controls, and open system.

Hipwell {2} developed and implemented cooperative decision-making behaviors in an air combat simulation by using a knowledge-based system. He addressed the specific problem of generating autonomous forces for inclusion in the ARPA-DIS program. The system developed is based on phase control of a blackboard architecture. Cooperative behaviors are based on a leader-follower relationship. Agents share the workload in assessing threats. Leaders make the initial decision, but followers react independently if necessary. This relates to the virtual environment.

Irani et al. {4} introduced an approach for cell formation which integrates machine grouping and layout design, neglecting part family formation. They related the concept of a hybrid cellular layout and virtual manufacturing cells.

Kessler et al. {5} addressed the development of new tools and methodologies for the successful and timely transition of new product technologies to weapon systems. This depends heavily on the technical maturity, flexibility, and cost effectiveness of the critical manufacturing processes and systems required to turn these technologies into tangible products. They also addressed that virtual manufacturing and other tools will facilitate the early consideration of manufacturing in design.

Kimura {6} introduced a set of models pertinent for realizing Virtual Manufacturing Environment (VME). Among them product/process models are discussed in some detail as a kernel. Product models represent every antifact which appears during manufacturing, and process models are associated with them to derive their properties and behavior.

Mills et al. {11} addressed the VM issues in the integration of and the transformation (interfaces) among the various representations (geometry, material flow etc.); the concurrent access and subsequent merging of these representations into one coherent, distributed database; and the architecture and coordination required for their support.

Oliver et al. {12} worked on virtual environment for manufacturing systems and built a theoretical framework to support the development of a virtual environment model of a computer integrated manufacturing workcell to facilitate off-line programming, system simulation, and visual and analytical verification of actual system capabilities and performance. They emphasized the development and utilization of virtual environments for assembly sequence planning, material removal process simulation, and off-line programming.

Onosato and Iwata {13} developed a Virtual Manufacturing System (VMS) by integrating product models and factory models. They also introduced the concept of VMS, discussed the relationships with other concepts in manufacturing, and explained the architecture of VMS.

Varnado {18} analyzed the significant effects on producibility of weapon systems caused by suspending system development after prototype development. He provided a critical examination of the following affordability interface technologies: SIMNET, CAD/CAM, CIM, CAE, CAPP, CADFM, Virtual Prototyping, Virtual Manufacturing, Concurrent Engineering, Design for Manufacturability, Flexible Acquisition, Agile Manufacturing etc.

The Air Force ManTech perspective on VM {19} is the use of models for the integration of simulation and control for manufacturing, providing capabilities supportive of new Department of Defense acquisition strategies. Key to VM is the ability to interchange models between their use in a simulation environment and a control environment. VM supports the implementation of manufacturing systems composed of autonomous cooperative components, distributed in a decentralized manner throughout the enterprise. VM emphasizes formalized interacting product, process and resource models capable of supporting the capabilities required for each and every manufacturing component. The Air Force ManTech also believes that the effective VM development will require programs with broader scope supported by a number of major partners.

Hitchcock et al. {3} discussed the role of hybrid systems theory in VM. They also proposed that Virtual Manufacturing Environments (VME) will improve military preparedness, enable lean and agile manufacturing, improve product and process design, reduce manufacturing risks, improve manufacturing design and operation, support manufacturing system changes, enhance product service and repair, increase manufacturing understanding, and provide a vehicle for manufacturing training and research.

Kimura {7} proposed an open system framework for concurrent engineering based on VM. They also discussed about modeling of flexible configuration and execution of manufacturing activities, and quick adaptation of the system behavior to changing manufacturing requirements.

Radharamanan et al. {15} are working on the taxonomy of the technological underpinning and modeling approaches necessary to realize VM environment. Their work involves the identification of actions necessary to implement and incorporate VM into the business process and the description of benefits, risks, limitations, and costs associated with adopting VM.

Creation of Virtual Manufacturing Devices (VMD) by means of information systems models were presented by Rondeau et al. {16}. They developed an interface which allows one to generate the VMD of each manufacturing equipment automatically, by the information system study.

The objective of the Virtual Manufacturing User Workshop {20} was to encourage the "users" from government and industry to discuss the issues related to potential roles of modeling and simulation in manufacturing. One of the purposes of this workshop was to establish the directions for a series of follow-on technical workshops which will match user requirements generated at this workshop with technical capabilities. The workshop addressed the issues of VM through breakout sessions: the use of VM in production operations; the impact of VM in business culture; quantifying VM benefits; VM in design to reduce risk and cost; VM in education; and VM technologies. The results of the workshop are summarized in {20}.

Definition

According to the Air Force ManTech {3} "Virtual Manufacturing is an integrated, synthetic manufacturing environment exercised to enhance all levels of decision and control" in a manufacturing enterprise. Here, "synthetic" refers to a mixture of real and simulated objects, activities, and processes; "environment" supports the construction and use of distributed manufacturing simulations by synergistically providing a collection of tools (simulation, analysis, implementation, and control), models (product, process and resource), equipment, methodologies and organizational principles; "exercising" refers to the construction and execution of specific manufacturing simulations using the environment; "enhance" refers to increasing the value, accuracy and validity of decisions; "levels" indicate from product concept to disposal, from the shop floor to the executive suite, from factory equipment to the enterprise and beyond, and from material transformation to knowledge transformation; "decision" reflects how one can visualize, organize, and identify changes in the alternatives; "control" corresponds to predictions that effect actuality.

However, it is clear from the outcome of the "Virtual Manufacturing User Workshop" that a single definition of VM is inappropriate {20}. A definition of VM is proposed to capture design, production, and control aspects of manufacturing: Design-Centered VM adds Manufacturing information to Integrated Product Process Development (IPPD) with the intent of allowing simulation of many Manufacturing alternatives and the creation of many "soft" prototypes by "Manufacturing in the Computer"; Production-Centered VM addresses simulation capability to manufacturing process model with the purpose of allowing inexpensive, fast evaluation of many processing alternatives; and Control-Centered VM is the addition of simulation to control models and actual processes, allowing for seamless simulation for optimization during the actual production cycle.

As shown in Figure 1, the vision of VM is to provide the necessary capability to "Manufacture in the Computer". It is expected that VM will ultimately provide a powerful modeling and simulation environment so that design, manufacturing, and assembly of any product can be simulated in the computer.

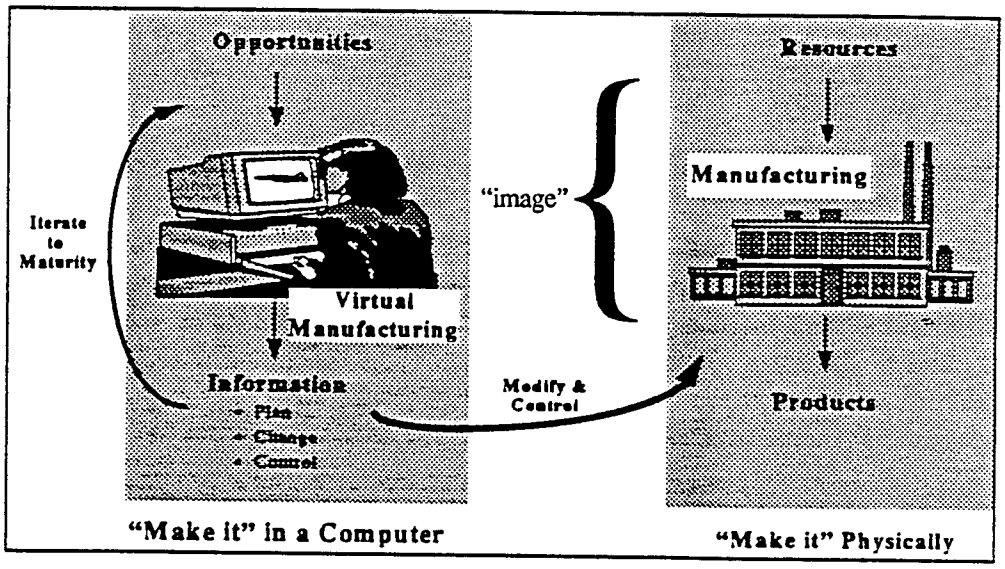


Figure 1: The Vision of Virtual Manufacturing

Concepts

VM has the ability to interchange models between their use in simulation and control environments. This provides confidence that VM simulations are robust, and will represent faithfully the real world situations. Throughout manufacturing, models of products, processes, and resources are used to represent the knowledge necessary to control various components of physical/intellectual operations. Rapid development of models and the component performance evaluation are well supported by the use of simulation. The use of VM concepts improves decision makin and quickly achieve products with high performance and quality at a low cost.

VM supports implementation of lean/agile manufacturing to achieve improvements in enterprise flexibility and economy. The use of simulation results in manufacturing systems that are less risky to change. Computer assisted model-based planning and control systems require less coordinating communications. The models provide a basis for sharing knowledge between organizations.

VM based systems are expected to enhance operations by providing timely answers to the questions: Can we make the product? What are the alternatives? What is the best way to produce the product? When can we deliver the product? How much will it cost?

As shown in Figure 2, in the defense environment VM bridges the gap between combat customer scenario and VM in operation scenario, considering product (product/process design, production system design, and production management) and process (concept development, dem/val, EMD, and production) life cycles, and process (unit processes, subsystem level, and system level) and enterprise (shop floor, factory, and industrial enterprise) hierarchies.

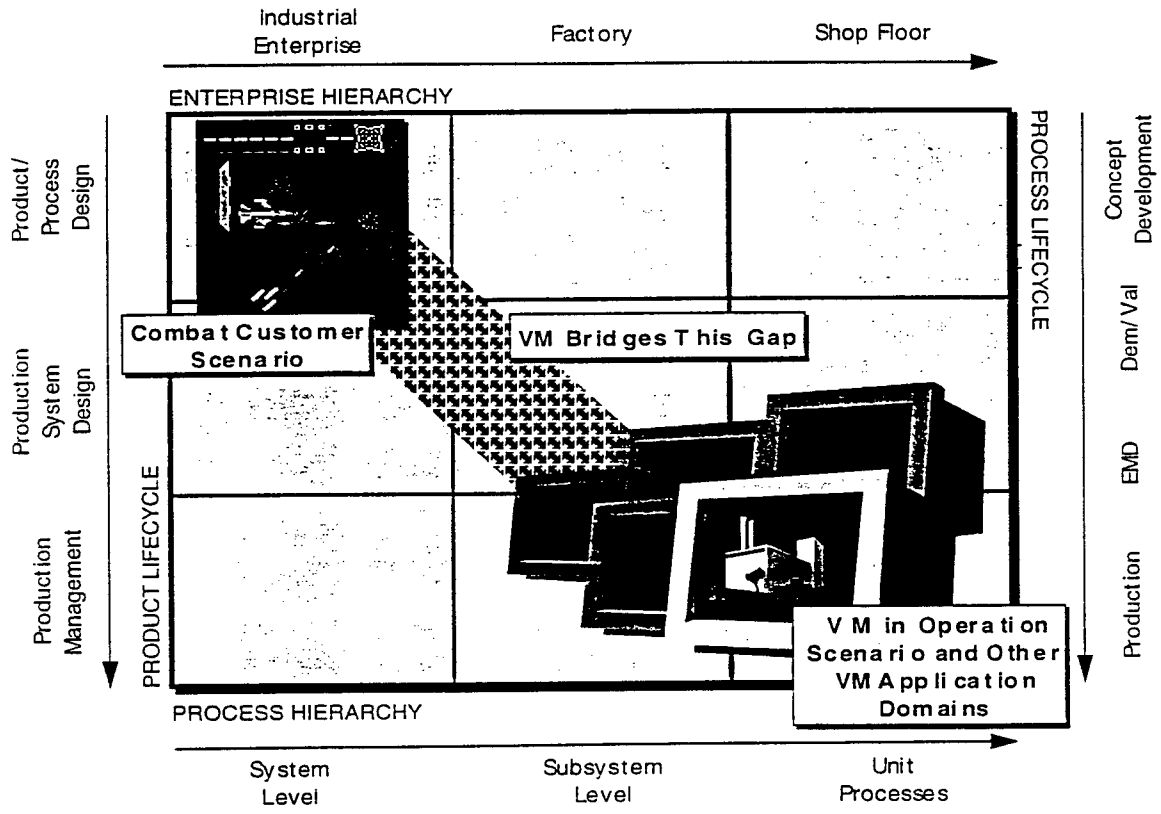


Figure 2: The Scope of Virtual Manufacturing

VM relies on modeling and simulation technology to simulate the production process and to enable us "make it virtually." It is an application of modeling and simulation, but extends that discipline beyond the conventional use. VM supplements the IPPD process since it provides a pathway for the manufacturing knowledge to be migrated to the early phases in the life cycle. VM also adds simulation to the Virtual Enterprise (VE) concept and Virtual Prototyping

The integration of VM with enterprise functions is shown in Figure 3. It represents an attempt to capture the idea that VM is not design or "...ilities," in order to accomplish the desired cross-functional trade-off analyses. In most cases, VM must be integrated with all the relevant enterprise functional areas via a trade-off mechanism (IPPD process).

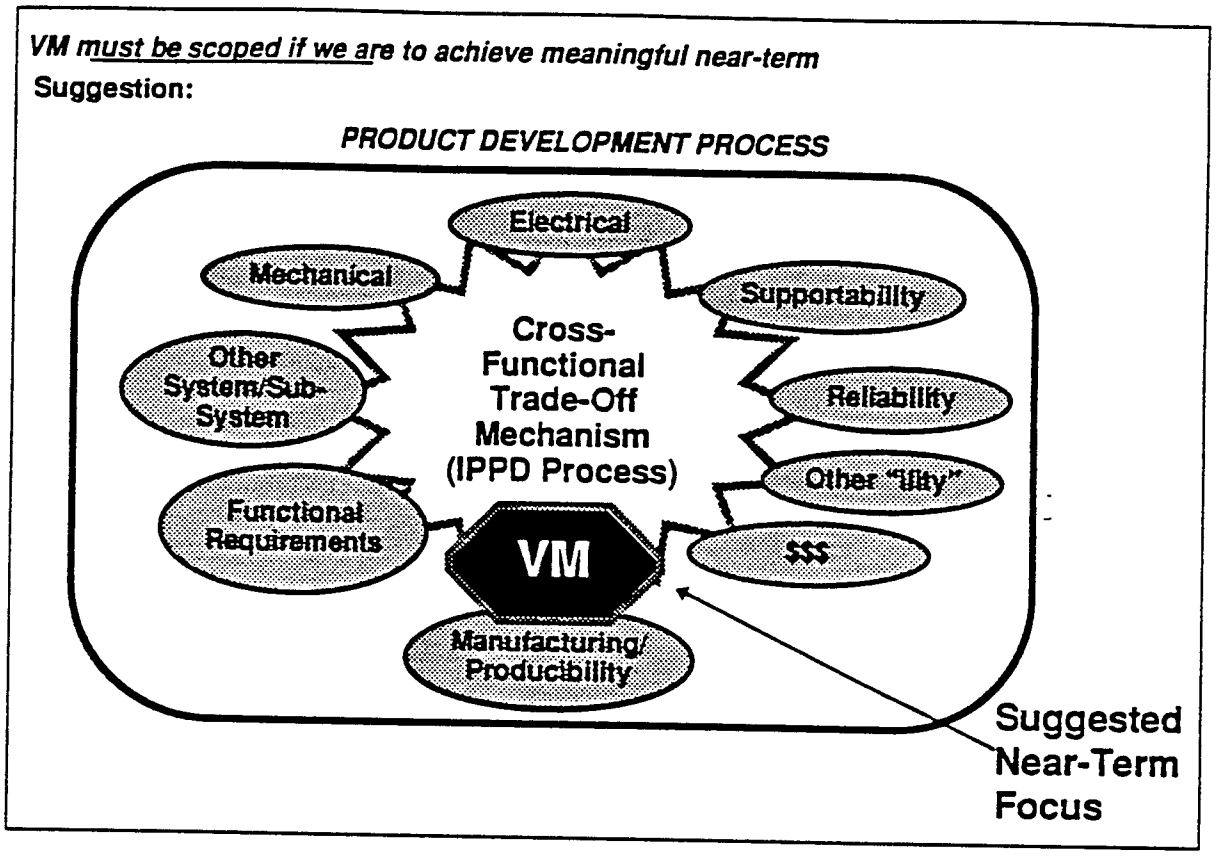


Figure 3: The Integration of VM with Enterprise Functions

Tools

There is a need for developing new tools for implementing the VM concepts in practice. Some of the existing design and production engineering tools that can be effectively incorporated are listed below:

Design Tools: Computer Aided Engineering (CAE), 3-Dimensional Computer Aided Design (CAD) Models, and Design for Manufacturability (DFM).

Production Tools: Computer Integrated Manufacturing (CIM), Advanced Modeling and Simulation, Distributed Interactive Simulation (DIS), Integrated Product/Process Development (IPPD), Just in Time (JIT), Materials Requirement Planning (MRP), Manufacturing Resources Planning (MRP II), Kanban, Virtual Reality, Visualization/Animation Tools, Hybrid System Theory, Complexity Theory, Distributed Computing, and Self-Directed Control.

Quality Tools: Total Quality Management (TQM), Quality Function Deployment (QFD), and Robust Design.

Artificial Intelligence (AI) Tools: Expert Systems, Neural Networks, Fuzzy Logic, Object Oriented Technologies, Autonomous Agents.

Computer Science Tools: Programming, National Information Highway, and Modern Telecommunications.

Management Tools: Quality Philosophies in Manufacturing (Deming, Juran, and Crosby), Manufacturing Strategies (Skinner, Wheelwright & Hayes, and Stalk Jr.), and Management Information Systems (MIS).

Mathematical Tools: Advanced Mathematics, Advanced Statistics, Optimization, and Stochastic Models.

Technology Issues

New integration technologies and philosophies are emerging. Visualization hardware and some software is becoming more affordable and widespread. New modeling and model abstraction techniques are appearing. The most important set of technologies center on modeling and simulation. The key areas that require attention in modeling and simulation are: model object selection (what to model); degree of abstraction; level of depth; flexibility and maintenance of models; integration of different models; and model validation.

Technical issues/barriers are of two types: tools are associated with the incremental, near-term of VM; and infrastructure, which is associated with thelong term vision. The components of technical architecture of VM include models, simulation, and environment. VM can contribute for the development of the following technologies:

- Modeling and simulation;
- Virtual reality and virtual environment;
- Integration and telecommunication infrastructures;
- Information processing;
- Hardware and Software;
- Implementation and control.

Benefits {20}

The expected benefits of VM are listed below:

Military Preparedness: VM can improve military preparedness by enabling the virtual production of weapons systems without the expense of actual production, storage, or disposal. Once production methods have been proven, these methods can be stored in the computer, ready

for deployment if needed. Manufacturing "know-how" can be preserved with near-zero production. The military complex would no longer need to heavily invest in technology which becomes obsolete without ever being used. It can prepare for war without actually investing in war-time production facilities or stockpiling excess, potentially environmentally dangerous munitions. By conserving resources until they are needed, and yet having the ability to produce a variety of needed weapons systems, the military can rapidly respond to changing mission requirements.

Affordability: A dramatic and pervasive benefit is expected to emerge in the area of affordability. Many risks and problems that have driven the costs of weapon systems in the past will be positively impacted as reliable cost and process capability information impacts key design and management decisions.

Shorter Cycle Times: The development time should be substantially shortened through the increased effectiveness of the IPPD process and due to the ability to go directly into production without false starts.

Producible Prototypes: The very first product (hardware) produced should be relatively trouble free if VM realizes its full potential since the design and the manufacturing processes will have been modeled, simulated, and refined in the computer prior to reaching the shop floor.

Quality: Product quality should be greatly enhanced through the more producible nature of the designs that will move to the shop floor and the high quality of the tools and work instructions available to support production.

Responsiveness: The ability to respond to the needs of customers for product capability should be significantly enhanced in both timeliness and cost. The ability to respond to customers about the impact of various funding profiles and delivery schedule should be markedly improved in both accuracy and timeliness.

Customer Relations: VM will be used as a marketing tool to improve customer relations even before actual parts are made in the shop floor.

Applications

VM environment enables a shortening and simplification of the life cycle, by improving the reliability of analyses and accelerating decisions through the use of modeling and simulation. VM helps to evaluate product making using simulation and supports operations to provide timely response to the Integrated Product and Process Design (IPPD) functions in the development of new products and/or processes. Collections of objects in a VM environment may also simulate the entire manufacturing enterprise to provide rapid response to customer requirements, e.g., proposed weapon systems developed as part of battlefield simulation. Customers with multiple VM-based supplier organizations can use models of their suppliers' enterprises to provide knowledge to an Enterprise Capabilities Expert. Vertical partners can contribute to capabilities models for use in the

Enterprise Capabilities Experts, knowledge-based computer programs, that will evaluate customer requirements and supplier capabilities to establish the organizations desirous of responding to specific customer needs. VM also will support rapid technology transfer by enabling the sharing of the advanced manufacturing capabilities between cooperating organizations. VM applications and tools of one organization may be shared by means of the National Information Highway to support the operations of manufacturing partners. VM may also be used in the design of systems and provide the tools necessary to continually and rapidly improve manufacturing system's capability. Using an evolutionary development approach the operation of individual components and the interactions between components are simulated.

Specific Areas of Application: {20}

Corporate Memory: Corporate memory will be enhanced in the near-term through the increase in development and use of expert systems to capture the knowledge of subject experts. Overlong term the impact will be much more significant. Using expert systems in conjunction with VM would be a significant improvement by providing process capability and cost information to guide the product design process as well as adding some viability to the concept of "shelf technology" where a product might go into production long after the initial design prototyping and testing are completed.

Capital Investment: Manufacturing models and simulations will and are having some influence on capital decisions currently, but this use is isolated to a few companies and not widespread within those companies. In thelong term, VM should be widely used in capital investment decisions since it should allow more credible comparisons of investment alternatives and should also provide history on the performance of the past investments which is frequently hard to obtain in the current environment.

Supplier Management: The current VM impact on suppliers is probably rare and the use of VM by suppliers themselves would probably be limited to the largest companies because of the anticipated large investment required to install VM. The future impact on supplier management, however, is expected to be very significant. Make/buy decisions will be enhanced through easy access to better quality and more detailed information on costs, capacity, process capability, and lead-times as part of the make/buy decision process. Cost control would also be enhanced because of the more accurate cost information available about suppliers.

Product Design: In the near-term, available and emerging simulation will enhance the effectiveness of systems integration in the design process and, as a result, the fit of components will be enhanced, inference between subsystems and structure will be minimized, and the dependence on hard-mockups will wane as hard-mockups are replaced by computer modeling and simulation. In thelong term, major improvements to the transition from design to production are envisioned because of much stronger and effective influence of process capacity and manufacturing

cost information on the product designer as well as the ability to do many more design iterations prior to committing to hardware.

Cost Estimation: VM necessitates more accurate cost information than can typically be provided by current cost accounting systems (and VM cannot succeed without this kind of information). This will, in turn, accelerate the current trend toward activity-based accounting systems and other accounting system changes that allow detailed and accurate product costing. Future VM systems will provide accurate cost data throughout the design, development, and production process. Cost estimating systems will become fully integrated with design and manufacturing databases and will have access to detailed process-level design feature related data.

Risk Management: In the near-term, VM is expected to see only isolated use in risk management because available models and simulations are exercised to identify risk areas for added management attention. In the future, the role of VM could evolve into having a major influence on management identification of risks and the merits of alternative courses of action at all levels of management. It is likely that the interfaces with VM would be different at each level of management or within each function. The net result would be to understand and manage risks better.

Customer Interface: The interest and enthusiasm of the customer for VM could potentially lead to a temptation for companies to exaggerate the use and impact of VM in their dealings with the customer. In spite of this risk, near-term impacts are more likely to favor the inclusion of the customer in the IPPD process, the inclusion of some requirements for VM in the customer statements of work, and better responses to customer "what-if" questions about changes to budget and delivery schedule. In the long term VM will enhance the credibility of responses to "what-if" queries significantly and this, in turn, will have an important impact on program stability by allowing decisions about program budgets and delivery schedules at all levels of the government to be based on accurate and credible information. The customers' ability to participate in the IPPD process should be greatly improved. Uncertainty remains about what changes might evolve in customer oversight as a result of the enhanced visibility available. The risk that extensive "how to" requirements for VM might be placed on future contracts might suboptimize the effectiveness of VM deployment and use.

Functional Interfaces: VM will potentially accelerate the current trend toward weaker functional distinctions within companies by promoting the widespread sharing of information and enhancing close inter-functional working relationships within the IPPD process. This trend, in the long term, should lead to influence the weakening of functional departments within the companies and their customers, as information sharing becomes even more widespread and effective, and as work efforts are more likely to be organized on product basis rather than being functionally oriented.

Shop Floor: In the near-term, shop floor people and concerns should have a greater influence on the design process, and the manufacturing approaches that have been modeled and

simulated above the shop floor will be brought out on the shop floor to validate the models and simulation. In the long term, significant improvements to work instructions will be seen through the ready availability of graphics. Much better tooling will be available on the shop floor with features that make it easier for the worker to succeed via access to better instructions and illustrations to promote error-free tool use. This will also make it easier to accommodate the envisioned drop in the average skill and education level of shop-floor workers. The proofing of design and manufacturing processes in the computer prior to commitment to hardware should sharply reduce the problems on the shop floor. Labor relations issues are anticipated to arise as the character and/or existence of some unionized positions such as process planning is impacted by the evolution of VM.

Measurement/Metrics

Defining a set of metrics which can adequately describe the benefits of VM is a critical path item for both development and implementation. While all may be derivable from an underlying set of metrics, the benefits must be reliable, believable and grounded in reality. Additionally, they must be mapped to the terminology and experience base of specific users and sponsors. This mapping will be markedly different depending upon the individual user viewpoint. Some examples are:

- The unit process/production view can measure benefits in terms of fewer engineering change notices, reduced MRB actions, reduced process variability, etc.
- A system level/concept development viewpoint would be much more interested in benefits measured in terms of less time to market, bigger market share, etc.

Infrastructure

Concerning the infrastructure, three primary components have been identified: 1. There is a need for a common, robust modeling method employed as the normal form of the "master, integrated model." The master model is not a central database, rather a capability to compose information at an arbitrary level of fidelity/granularit; 2. There is a need for an intelligent browser. This would be a new tool which leveraged the federating abstraction mechanism; 3. The infrastructure introduces a whole new level of issues concerning control of itself. Therefore, there are a class of configuration management and security issues which need to be addressed in the design of the infrastructure.

Conclusions

VM should be implemented incrementally. The potential scope of VM is very large. It is intended to improve the manufacturing process and the scope of VM products in the "big M" manufacturing domain. Disagreement remains on what VM is and its technological underpinnings: there is not a single, exclusive definition. For the Department of Defense (DoD), VM is a powerful tool to reduce the risks involved with new weapon systems acquisition. VM is a valuable tool for changing the way to do business, especially on cost. There is a need for mechanisms to integrate related programs. It must be determined how to establish the needed mechanisms and to tie them to governmental contracts.

Recommendations/Future Directions

It is recommended to conduct a series of technical workshops to investigate near-term and long term issues of VM. A study should be carried out to form an industry/government consortium to develop and validate VM infrastructure as a development pilot site, technology transfer vehicle, and agent for education/cultural change; validate VM process and metrics; define the current VM activities underway in the aerospace defense industry; identify the trends in research and development being pursued by other government agencies such as NSF, NIST and DoE; identify potential members of university/industry/government consortia wherein each group would be based on common VM interests.

There is a need for creating Virtual Manufacturing Experimental Environment (VMXE) to bring together Industry, University, Research Institute, and Government to perform cooperative research in creating prototypes and tools, methods and techniques towards the development of new products. Once the product is developed and available, it can be brought into "Technology Clearinghouse" and, then, the product and technology, through "best practices expos", can be made available to the down chain (Government - Research Institute - Industry - Academia) so that the life cycle of the product ends after exhaustive use through the chain. This continuous chain of technology transfer will aid to maintain the best and the competitive industrial base marketplace in the United States.

VMXE, a catalyst for manufacturing technology improvements for the Air Force and DoD Industrial base, will provide learning environment through constantly evolving example of manufacturing "best practices", will implement a technology distribution methodology into government and industrial base with chains through academia, and always will exceed customer expectations. The working methodology will constitute the following:

VMXE working partners: Will include participants from Government, Research Institute, Industry, and Academia who are producers and users of virtual manufacturing (VM) tools, techniques and equipment.

Advisory Committee: An advisory committee will be formed within and outside the working partners of VMXE consisting of experts in various aspects of VM related activities.

Technology Transfer: Technologies developed by ManTech and other supporting organizations will be made available at the VMXE. Technologies will be integrated and tested over the information highway to facilitate transfer of technology.

Cost Trust: The objective is to define the costs associated with the affordability assessment calculation which include costs of design, technology development, manufacturing, and all other related costs which would be incurred over the life of the product or weapon. To accomplish this, a survey on the manufacturing cost research that are currently being studied at universities and research institutes is a prerequisite.

Virtual Manufacturing Experimental Consortia (VMXC): This facility would help and deploy VM technologies as follows: it would take the basic results, mature them, and implement them in industry.

Showplace and Expos: VMXE will create an evolutionary technology turnover - "Technology Continuum". Due to high technical focus, emphasis on quality, available intellectual power, etc., vendors, suppliers, and users will compete for participation and demonstration of their products.

Defense Manufacturing Journal (DMJ): DMJ will be published periodically to update the state-of-the-art research and technical information on defense manufacturing related topics such as VM among the participants/partners.

Education and Training: Education and Training will be accomplished through technical workshops, short seminars and conferences on emerging technologies. The participants would include graduate students, government project engineers and planners, industry partners, and researchers.

Other Related Issues: Will address short and long term goals of VMXE, hardware and software issues, benefits and limitations, deliverables, budget details for short and long terms, and implications of VMXE among partners.

References

1. Fujita, S., Oonami, M., Watanabe, M., and Nakayama, Y., "Model Based Real-Time Scheduling System - 3 Ps (Production Planning Partner)," Systems, Control and Information, vol. 36, no. 6, p. 386-7, Japan, 1992.

2. Hipwell, D. P., "Developing Realistic Cooperative Behaviors for Autonomous Agents in Air Combat Simulation," M. S. Thesis, School of Engineering, Air Force Institute of Technology, WPAFB, OH, 169 p., December 1993.

3. Hitchcock, M. F., Baker, A. D., and Brink, J. R., "The Role of Hybrid Systems Theory in Virtual Manufacturing," Proc. IEEE Symposium on Computer-Aided Control Systems Design (CACSD), IEEE, New York, p. 345-50, 1994.

4. Irani, S. A., Cavalier, T. M., and Cohen, P. H., "Virtual Manufacturing Cells: Exploiting Layout Design and Intercell Flows for the Machine Sharing Problem," International Journal of Production Research, v. 31, n. 4, p. 781-810, 1993.

5. Kessler, W. C., Shumaker, G. C., and Hitchcock, M. F., "Early Manufacturing Consideration in Design," In AGARD, Integrated Airframe Technology, 7 p., December 1993.

6. Kimura, F., "Product and Process Modeling as a Kernel for Virtual Manufacturing Environment," CIRP Annals, v. 42, n.1, p.147-150, 1993.

7. Kimura, F., "Virtual Manufacturing as a Basis for Concurrent Engineering," IFIP Transactions B (Applications in Technology), vol. B-17, p. 103-17, 1994.

8. McLean, C. R., Bloom, H. M., and Hopp, T. H., "Virtual Manufacturing Cell," Proc. IFAC/IFIP Symposium on Information Control Problems in Manufacturing Technology 1982 (4th), Gaithersburg, MD, October 26-28, 1982, p. 207-215, 1983.

9. Mills, J. J., Furth, W., Sekine, Y., Wysocki, E., and Burzio, A., "Virtual Manufacturing : A New Concept in Automated Design Aids," CAD/CAM Robotics and Factories of the Future, Proc. 3rd Int. Conf. (CARS and FOF '88), Springer-Verlag, p. 115-18, vol. 1, 1989.

10. Mills, J. J., Huff, B., Criswell, T., and Graham, J., "Virtual Manufacturing Workstation," Flexible Assembly Systems - 1992 ASME Design Engineering Division (Publication) DE, v. 48, New York, p. 35-39, 1992.

11. Mills, J. J., Graham, J. K., Elmasri, R. A., and Weems, B. P., "The Virtual Manufacturing Workbench: Representation and Interface Issues," IFIP Transactions B (Applications in Technology), vol. B-10, p. 231-44, 1993.

12. Oliver, J. H., Vanderploeg, M. J., and Chen. L. L., "Virtual Environment for Manufacturing Systems," Annual Progress Report, Iowa State University, Ames, October 92 - September 93.

13. Onosato, M., and Iwata, K., "Development of a Virtual Manufacturing System by Integrating Product Models and Factory Models," CIRP Annals, v. 42, n. 1, p. 475-478, 1993.

14. Pentland, A., Williams, J., "Virtual Manufacturing," Preprints, NSF Engineering Design Research Conference, Univ. Massachusetts, Amherst, p. 301-16, 1989.

15. Radharamanan, R., Brink, J. R., Ashcom, J., and Hitchcock, M. F., "Modeling Approaches in Virtual Manufacturing Systems," To be Presented in the Int. AMSE Conf. on Information Processing: Methodologies and Applications, New Orleans, LA, Nov. 9 - 11, 1994.

16. Rondeau, E., Divoux, T., Lepage, F., and Veron, M., "Creation of Virtual Manufacturing Devices by Means of Information System Models," IFIP Transactions B (Applications in Technology) vol. B-17, p. 259-73, 1994.

17. Staffend, G. S., "Making the Virtual Factory a Reality," Conference Proceedings, AUTOFACT '92, SME Publication, p. 15-9-16, 1992.

18. Varnado, F., "Effect of Weapon Systems' Producibility of Suspending System Development After Advanced Technology Demonstration," M.S. Thesis, Naval Postgraduate School, Monterey, CA, 143 p., March 1993.

19. Virtual Manufacturing: A Methodology for "Manufacturing in a Computer" - An Air force ManTech Perspective, CNCPTFRN.DOC, 15p., October 1993.

20. "Virtual Manufacturing User Workshop," Report, Lawrence Associate Inc., Dayton, Ohio, August 1994.

ANNEALED PROTON EXCHANGED (APE) WAVEGUIDES in $LiTaO_3$ for DIFFERENCE-FREQUENCY GENERATION of INFRARED COHERENT RADIATION

Ramu V. Ramaswamy

Professor
Photonics Research Laboratory
Department of Electrical Engineering

University of Florida
135 Larsen Hall
Gainesville, FL 32608

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

September 1994

ANNEALED PROTON EXCHANGED (APE) WAVEGUIDES in $LiTaO_3$ for DIFFERENCE-FREQUENCY GENERATION of INFRARED COHERENT RADIATION

Ramu V. Ramaswamy

Professor
Photonics Research Laboratory
Department of Electrical Engineering
University of Florida

Abstract

Annealed proton-exchanged channel waveguides in lithium tantalate were investigated. Theoretical modeling was based on the results of experimental studies performed by using a white-light source technique. Parameters defining the reconstructed profiles were evaluated by matching computed and measured values of the channel widths defining the region of single-mode operation. The calculated modal fields are in agreement with those obtained from near-field measurements.

Annealed Proton Exchanged (APE) Waveguides in $LiTaO_3$ for Difference-Frequency Generation of Infrared Coherent Radiation

Ramu V. Ramaswamy and Robert F. Tavlykaev
Photonics Research Laboratory, Dept. of Electrical Engineering
University of Florida, Gainesville, FL 32611

I. Introduction

There is increasing interest in generating coherent infrared radiation in the wavelength region 2.5 - 4 μm due to its applications in spectroscopy, sensing, laser radar etc. Although there exist discrete laser sources emitting in this spectral region, nonlinear difference-frequency generation (DFG) offers an interesting and efficient alternative to the realization of compact, room-temperature operating milliwatt-power-level sources. DFG can be achieved almost at any wavelength where the nonlinear crystal is transparent by appropriate choice of the pump and idler wavelength. Unfortunately, DFG using bulk nonlinear crystals requires extremely high power levels. This problem can be overcome by using an integrated-optic waveguide geometry wherein the interacting beams can be confined to very small cross sectional areas (10 - 50 μm^2) over very long interaction lengths (10-20mm). Due to the above advantages of the waveguide geometry, efficient nonlinear interactions can be obtained even with moderately low power laser sources.

In order to achieve nonlinear interactions in waveguides, we must look for materials which are transparent in the wavelength region of interest, possessing a large nonlinear optical coefficient, and in which one can fabricate low-loss waveguides. In addition, it should be possible to satisfy the phase matching condition. Lithium niobate ($LiNbO_3$), lithium tantalate ($LiTaO_3$) and potassium titanyl phosphate (KTP) are materials which satisfy the above requirements. The transmission windows (1/e in 1 cm length) are 350 -5500 nm in $LiNbO_3$, 320 - 5200 nm in $LiTaO_3$,and 380- 4400 nm in KTP. These materials have been widely studied for second harmonic generation using the quasi-phase-matching (QPM) technique. At the present time, there are a number of ideal materials available for difference-frequency generation.

The proposed project aims at developing a coherent infrared source in the region 2.5 - 4.0 μm. In particular, we propose to develop an Optical Parametric Oscillator (OPO) via difference-frequency generation using QPM techniques. For achieving QPM, a domain-inverted grating will be used. For demonstration purposes, we will consider a source operating around 3.5μm wavelength region using QPM difference-frequency generation in $LiNbO_3$ and $LiTaO_3$. A preliminary investigation has been carried out in $LiNbO_3$ resulting in considerable expertise in the group. However, $LiTaO_3$ is superior to $LiNbO_3$ in several respects: higher damage threshold, lower losses, and more favorable domain-inversion profile. Based on the superior performance of $LiTaO_3$, we will attempt to fabricate several annealed proton exchanged (APE) waveguides in this material using a tantalum mask and pyrophosphoric acid as the proton source. In addition, we propose to demonstrate domain inversion using proton exchange and heat treatment near the Curie temperature (590^0C). This method had been demonstrated by several groups. It results in

semi-circular domain inverted regions which favor the frequency conversion due to a better overlap of the pump and OPO modes in the domain-inverted grating region.

This report summarizes the results on the fabrication, characterization, and modeling of APE waveguides in $LiTaO_3$.

II. Distinct features of APE waveguides in lithium tantalate

To date, several research groups have been active in the area of annealed proton exchanged (APE) waveguides [1-3]. However, most of the effort has been devoted to waveguides in lithium niobate. In the past few years, there has been increasing interest in analyzing APE structures in $LiTaO_3$ for the seasons stated previously. Although, some preliminary results have been obtained, these studies are still in their infant stage. In this connection it should be noted that despite similar crystallographic structure and electro-optical properties, these two crystals differ significantly with respect to specific characteristics of the waveguides due to the proton exchange process and subsequent annealing.

First, the refractive index increment that can be achieved in APE $LiTaO_3$ waveguides is almost an order of magnitude lower than that in $LiNbO_3$ [5]. Second, APE $LiTaO_3$ waveguides exhibit significant instabilities appearing as a temporal dependence of the waveguide index. Both short-term and long-term stability problems have been reported [5,6]. A third, and the most important, fact is the anomalous behavior of waveguides profiles upon annealing. Some studies seem to have shown that the refractive index in exchanged areas increases during the initial periods of the annealing process [5,7]. This has been attributed to a nonlinear dependence of the exchanged refractive index on the proton concentration.

A rigorous theoretical analysis of APE waveguides in lithium tantalate is hindered in large part since the phase diagram of this material is unknown. Some studies have suggested the possibility of obtaining buried profiles just in a single-step diffusion due to the aforementioned nonlinear dependence of the refractive index on concentration. As such, information on the shape of the waveguide profile is, in general, required in order to accurately predict the waveguide parameters for arbitrary fabrication conditions. However, the described peculiarity, namely, a significant increase of the refractive index during initial stages of the exchange process, has been shown to occur only at relatively high concentrations. For low concentrations, that correspond to well-annealed waveguides, the refractive index increment monotonically decreases with increasing exchange time. The resulting profile can, therefore, be approximated by the profiles conventionally used in modeling APE waveguides in lithium niobate [8].

III. Numerical routine for reconstructing APE profiles

To model APE waveguides in lithium tantalate, a numerical routine has been used that allows reconstruction of the waveguide profile using measured near-field pattern for the

fundamental mode of a waveguide with its channel width being in the middle of the range of single-mode operation defined by the known values. The necessary data has been obtained by using the white-light source technique [9] with an optical spectrum analyzer, as described in the report by David Mering [10] devoted to experimental studies.

The numerical technique is an iterative process that adjusts the parameters of the waveguide profile in order to obtain the best match between the computed and measured values to the channel width corresponding to the cut-off conditions for the fundamental and first asymmetric mode. An array of straight channel waveguides with monotonically increasing widths is considered as shown in Fig.1a. White-light source technique with an optical spectrum analyzer [10] is used to determine the channel widths W_1 and W_2 corresponding to cutoff of the fundamental and fist higher-order mode, respectively. Near-field measurements with the use of a conventional set-up (Fig.12) are performed, as described by David Maring [10], to obtain the modal distribution in a single-mode waveguide of a channel width between W_1 and W_2. This ensures a well-confined fundamental mode and the applicability of the effective index method [11].

The procedure is based on the following well-known assumptions:

1. The transverse refractive-index profile of a channel waveguide of width W can be expressed as [12]:

$$n(x, y, W) = n_s + \Delta n(x, y, W) = n_s + \Delta n_0 g(x, W) f(y) \tag{1}$$

where n_s is the substrate refractive index; Δn_0 represents the peak index change corresponding to an infinite channel width, i.e., to the planar waveguide. The index distributions in the depth and width direction, respectively, f(y) and g(x,W) are approximated by the following functions:

$$f(y) = \begin{cases} \dfrac{1-(y-b)^2/a^2}{1-b^2/a^2} & 0 \le y \le b+c \\ s \cdot \exp\left[-(y-b-c)/h\right] & b+c \le y \le \infty \end{cases} \tag{2}$$

$$g(x, W) = \frac{1}{2}\left\{\operatorname{erf}\left[(W/2-x)/d_x\right] + \operatorname{erf}\left[(W/2+x)/d_x\right]\right\} \tag{3}$$

where $h=a/2(a/c-c/a)$ and $s=(1-c^2/a^2)/(1-b^2/a^2)$; W is the mask opening width before exchange. The parameters a, b, c, and d_x are to be determined. The first three parameters describe the distribution in depth, Eq.(2), while d_x is the lateral penetration depth determined by the exchange temperature and time [13].

The advantage of representing the function f(y) in the form of Eq.(2), i.e., as a combination of a parabola and an exponential tail, is that it simplifies calculations and reduces CPU time, since there are analytic expressions available for the fields of waveguide eigen modes.

The shape of Eq.(2) is determined by the parameters (a, b, c) and depending on their values it can simulate a wide variety of refractive-index profile types. For example, for b=0, this profile can well approximate a Gaussian function widely used to model diffused waveguides, whereas for b<0 it has a non-zero derivative at the waveguide-air interface (y=0) and can model either buried (b>0) or shallow (b<0) profiles.

2. The transverse field of the fundamental mode is assumed to be separable, i.e., $\Psi(x,y)=\Psi_x(x)\,\Psi_y(y)$, where $\Psi_x(x)$ and $\Psi_y(y)$ are the field profiles along the width and depth direction, respectively.

3. Taking into account approximate values for the lateral diffusion depth that correspond to the parameters of the fabrication process we used, the one-dimensional lateral effective index distribution for a waveguide with width not exceeding 7 μm was approximated by a truncated parabola:

$$n_{eff}(x) = \begin{cases} n_s + \Delta n_{eff} \cdot (1 - x^2/u^2) & -u \le x \le u \\ n_s & |x| \ge u \end{cases} \qquad (4)$$

where u and Δn_{eff} are the width of the effective waveguide and its refractive index increment, respectively.

The reconstruction of a two-dimensional transverse profile consists of the three steps shown in the flowchart in Fig.1b. The procedure starts with the determination of the effective waveguide parameters which are then used to calculate the channel waveguide profile. Let us consider a single-mode waveguide of the width $W_1<W<W_2$ and first assume the propagation constant β_{in} of the fundamental mode to be known.

i) A two-dimensional area $(\Delta n_{eff},u)$ is used for computations. It should be chosen large enough to ensure that the region of monomode operation is enclosed. In our computations, we have used $10^{-5}<\Delta n_{eff}<10^{-2}$ and $0.5\mu m<u<10\mu m$. The area is divided into a mesh and the modal properties of the effective waveguide defined by Eq.(4) are calculated at each nodal point. The computed eigenfunction Ψ_x^c and eigenvalue β_c are compared to the measured field profile Ψ_x^m and the input value β_{in} of the propagation constant, respectively, in order to select a nodal point giving the minimum deviation. This point is used for the next iteration as the center of a smaller rectangle divided into a mesh of higher resolution. The iterations continue until the deviation is reduced below a prescribed level, determined by measurement errors, and finally, the optimum values $(\Delta n_{eff}^{opt},u^{opt})$ are found that define the effective waveguide profile.

ii). The reconstruction of the strip waveguide profile starts from the determination of the index distribution in depth at a position x=0. For this purpose, calculations similar to those used to reconstruct the effective waveguide are performed. The parameters to be adjusted are $\delta n_0=\Delta n_0$ g(0,W), a, b, and c. For each set of their values, the computed eigenfunction Ψ_y^c and propagation constant are compared to the measured field profile Ψ_y^c and $k_0(n_s+\Delta n_{eff}^{opt})$, respectively, where k_0 is the wavenumber in a vacuum. The best fitting between them corresponds to the optimum values $(\delta n_0)^{opt}$, a^{opt}, b^{opt}, and c^{opt}.

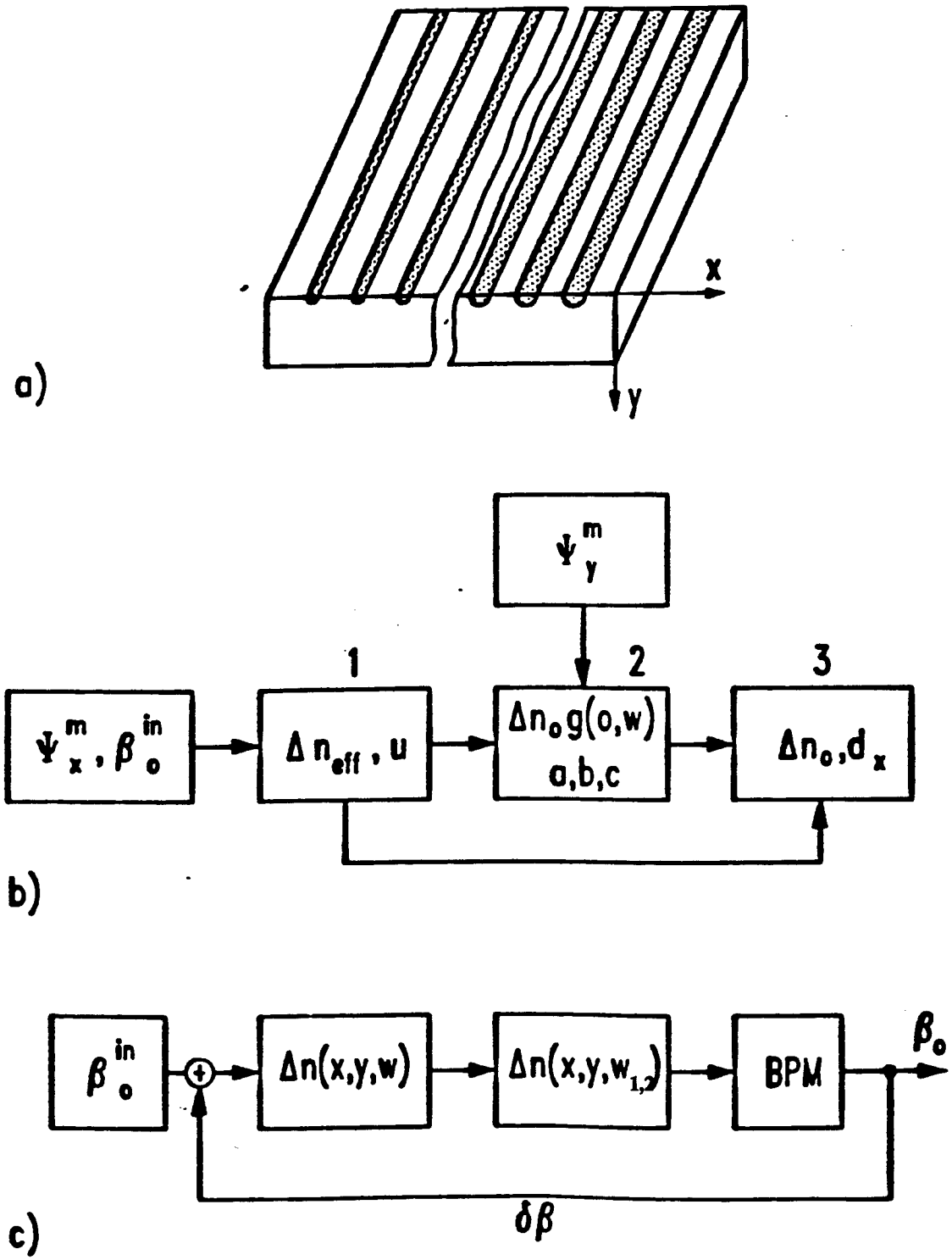


Fig.1 Refractive index reconstruction: (a) test structure with an array of channel waveguides; (b) flowchart of the numerical procedure for profile modeling; (c) evaluation of the propagation constant

iii). To get the full set of parameters characterizing the channel profile, the lateral diffusion length d_x and the peak index change Δn_0 must be determined. In fact, they are not independent, since there is an implicit relationship between them: $\Delta n_0\, g(0,W) = (\delta n_0)^{opt}$, where $(\delta n_0)^{opt}$ is the value determined at the second step. Therefore, if some value is assigned to d_x, we can calculate the strip waveguide profile as well as the corresponding effective profile. The value of d_x is adjusted, in order to obtain the best fitting of the computed effective profile to that defined by the values $(\Delta n_{eff}^{opt}, u^{opt})$ as determined at the first step. Thus, the procedure allows reconstruction of the refractive-index profile of a single-mode waveguide provided its mode near-field pattern and the propagation constant are known.

In order to avoid time-consuming measurements of propagation constants or, if they have somehow been obtained, to have an additional check of the accuracy of computations, we have taken use of the following computational process (see Fig.1c). At first a trial value for the fundamental mode propagation constant β_{in} is assumed to be lower than the expected one. If there is no estimate for the initial guess, it can be taken just near cutoff, i.e., $\beta_{in} \approx k_0 n_s$. The chosen value of β_{in} and the measured near-field pattern $\Psi_m(x,y)$ are used to determine the waveguide profile parameters, as described above (Fig.1b). Next, the profile of a strip waveguide of the width W_2 is calculated by substituting W_2 into (3) and its modal spectrum is analyzed by the conventional BPM [14]. If this waveguide is found to support only the fundamental mode, then one should make an increment $\delta\beta$ for the input value of the propagation constant. The cycles are repeated until the first higher-order mode in the waveguide of the width W_2 just appears above cutoff. Then, the profile of a strip waveguide of the width W_1 is analyzed in a similar manner with a reservation that the effective index profile (analyzed by one-dimensional BPM) is not applicable in this case since due to its symmetry there is no cutoff for the fundamental mode. After the parameters characterizing the index distribution in depth and the lateral diffusion length have been evaluated, the profile for any channel width can be calculated.

IV. Results

The profiles of APE waveguides have been modeled at λ=1.31μm. At this wavelength, the range of single-mode operation is bracketed by the values W_1=3.0μm and W_2 =8.4μm. The waveguide of width W=6μm was chosen as being sufficiently far from cut-off value for the fundamental mode so that the effective index method can be applied. On the other hand, this width is narrow enough to be comparable with the lateral penetration depth which can be roughly estimated to be $d_x \approx 1\mu m$ by using the results available in the literature [15]. As such, the effective index profile can be modeled by Eq.(4).

The iterative process started with a trial value of $0.4 \cdot 10^{-3}$ for the difference $N-n_s$ between the effective mode index and the bulk value (extraordinary index of lithium tantalate). The final value was obtained to be $N-n_s = 1.4 \cdot 10^{-3}$ that corresponded to W_1=2.8μm and W_2 =8.1μm matching reasonably well the values measured with the optical spectrum analyzer. The computed values were determined by propagating light through a waveguide with the modeled profile and computing loss associated with decreased optical field confinement as the mode approaches cut-off. Considering the weakly-guiding nature of APE waveguides, we used the standard BPM

algorithm that can be represented in a symbolical denotation as:

$$E(x, y, z + \Delta z) = \exp\left[-j\frac{\Delta z}{4n_s k_0}\nabla_\perp^2\right] \times \exp\left[-j\frac{n_s k_0}{2}\int_z^{z+\Delta z}\left(\left\{\frac{n_s + \Delta n(x, y, z')}{n_0}\right\}^2 - 1\right)dz'\right] \times \exp\left[-j\frac{\Delta z}{4n_s k_0}\nabla_\perp^2\right] E(x, y, z) + O((\Delta z)^3) \quad (5)$$

where $\nabla_\perp^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse laplacian; E represents the electric field of optical radiation; Δz denotes the step in the propagation direction.

The first and last term of the right side of Eq.(5) describe diffraction in a homogeneous medium with the refractive index n_s and can be computed by employing Fast Fourier Transform routines. The large drop of the refractive index at the boundary substrate-air was taken into account by substituting the substrate surface with a perfectly reflecting boundary. To accomplish that, the field distribution in depth was assigned a zero value at the surface and antisymmetrically flipped to the upper half-space (Fig.2). In performing computations, we used a computational mesh of the size D_x=128μm; D_y=64μm; and a step of Δz=10μm. The required CPU time is typically about 5min for an 10mm-long structure on a PC486.

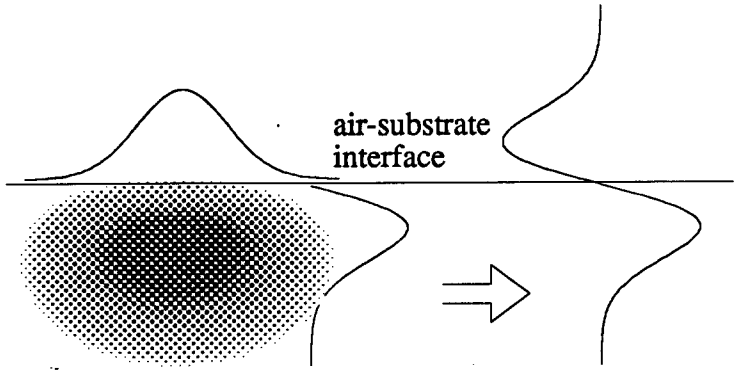


Fig.2 Modeling of light propagation along a waveguide by using Beam-Propagation Method

Shown in Figs.3 and 4, measured field distributions in depth and width are in fairly good agreement with the modal profiles computed for the optimum values of the waveguide parameters according to the algorithm described above. The measurement set-up used to arrive at the near-field profiles is described in the last section and illustrated in Fig.12.

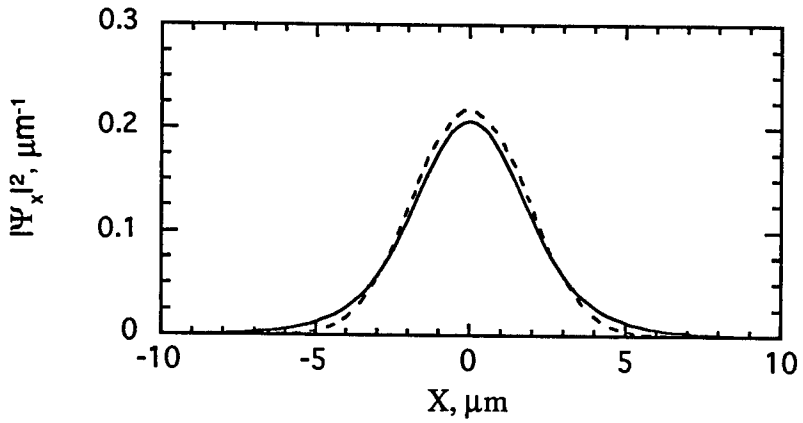


Fig.3 Measured (solid) and computed (dashed) horizontal modal field profile (W=6μm).

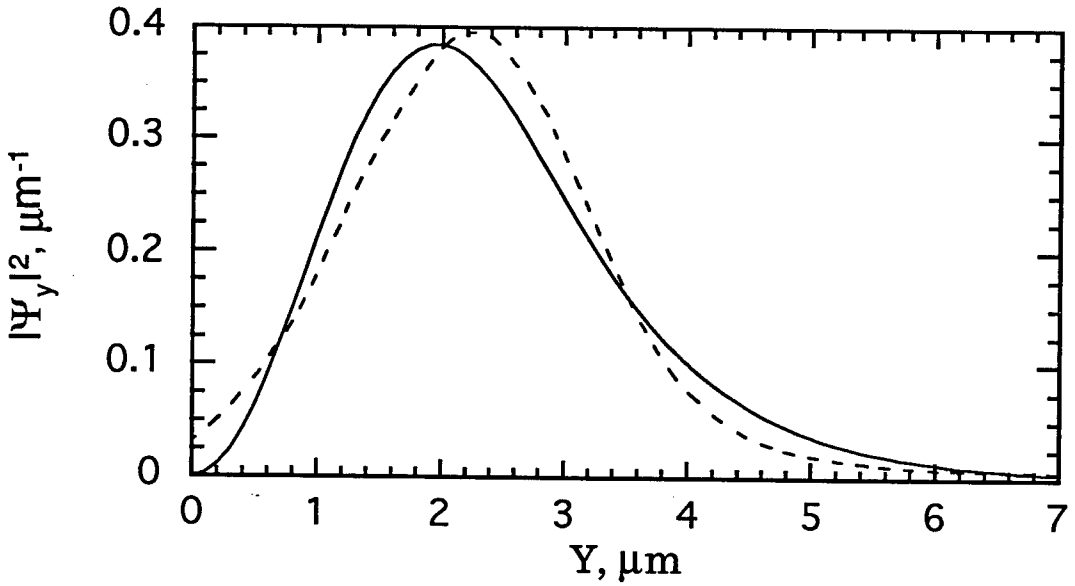


Fig.4 Measured (solid) and computed (dashed) vertical modal field profile (W=6μm).

The waveguide parameters corresponding to the waveguide with W=6μm are: $\Delta n_{eff}=3.03\cdot10^{-3}$, u=3.07μm, a=3.67μm, c=2.94μm, d_x=1.9μm, $\Delta n_0=1.18\cdot10^{-2}$. The index profile in depth and width are shown in Figs.5 and 6. The corresponding one-dimensional effective index profile is depicted in Fig.7 along with the truncated parabola profile used in the first step of reconstruction.

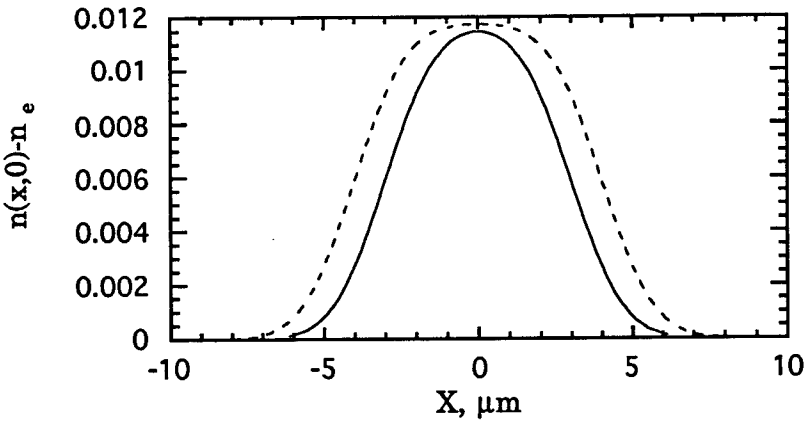


Fig.5 Waveguide index profile along the width direction for a strip width of 6 μm (solid) and 8 μm (dashed). Increasing channel width leads to an increase in the maximum increment, broader distribution, and a flattened top.

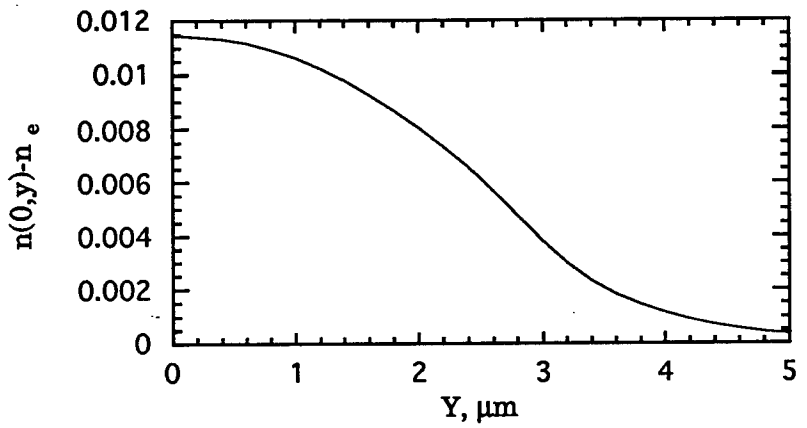


Fig.6 Reconstructed waveguide index profile along the depth direction for a strip width of 6 μm.

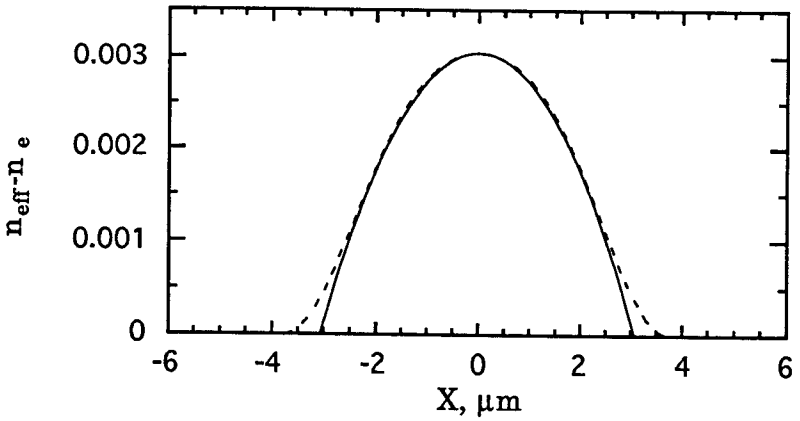


Fig.7 Effective index profile corresponding to the reconstructed 2D profile of a strip width of 6 μm (Fig.8 and 9): initial parabolic profile (solid) and approximating distribution (dashed)

Keeping in mind the remarks made at the beginning of Section II, it becomes desirable to extend the above technique to analyzing waveguide profiles fabricated under different conditions. In this connection, a further detailed study would be helpful to directly measure the shape of the waveguide profiles.

V. Experimental data.

As stated above, modeling of waveguides in lithium tantalate has been performed by comparing computed and measured modal fields as well as the channel widths confining the region of single-mode operation. In this section, only the graphs of the experimental data necessary for modeling are provided. A detailed description of the experimental set-up and techniques can be found in a report by David Maring [10].

The results of measurements with an optical spectrum analyzer are shown in Fig.8 for several channel widths. As seen, the transmission curves reveal peaks corresponding to the waveguide modes of different order. Note that higher-order modes reach cut-off at shorter wavelengths. The cut-off of the mode of the n-th order is assigned to the wavelength at which transmission rolls-off by 3dB below the corresponding peak (see Fig.10). To determine the values of W_1 and W_2, it is quite useful to summarize the results of measurements in a compact presentation and they are shown in Fig.11. These values can then be determined by simply finding intersections of the straight line $\lambda=\lambda_0$ (λ_0 is the wavelength of interest) with the curves for the fundamental and first antisymmetric mode.

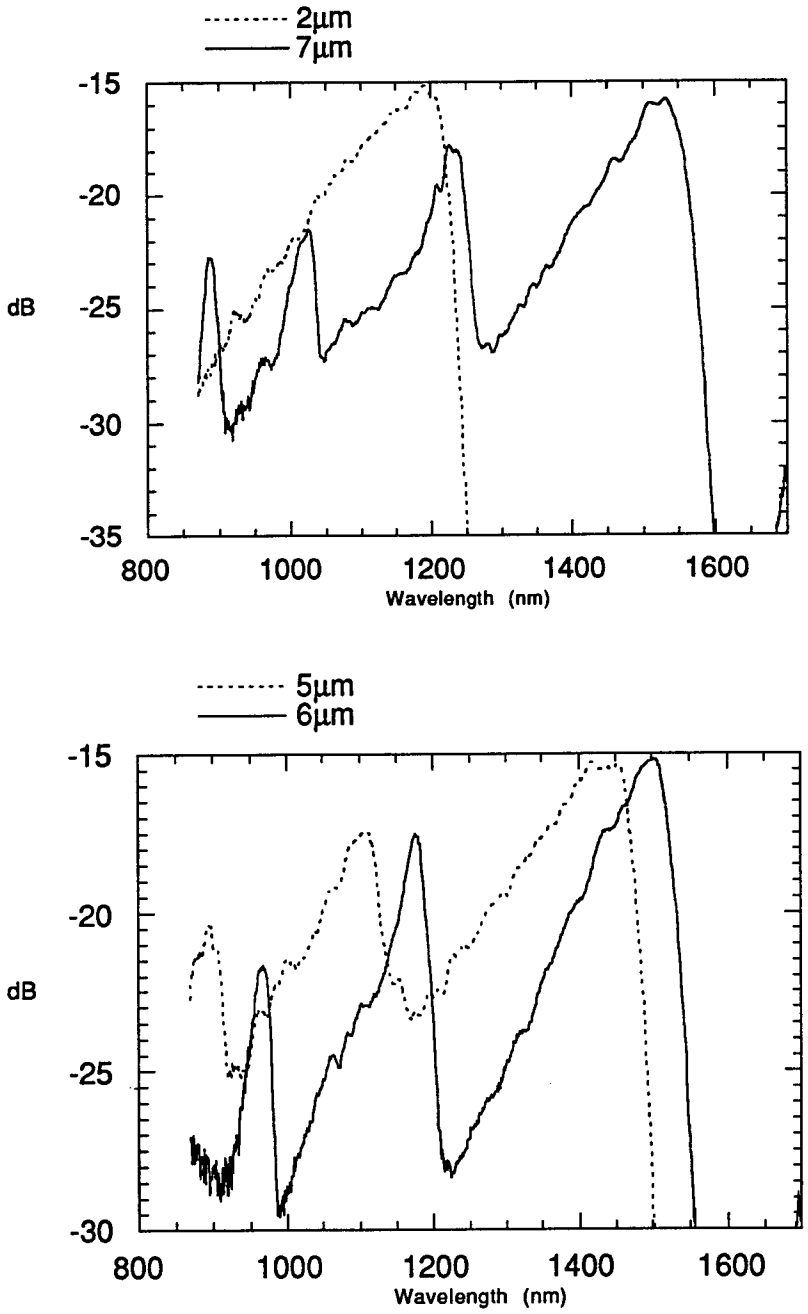


Fig.9. Scans of optical power transmission as a function of wavelength for various width guides excited by a white light source.

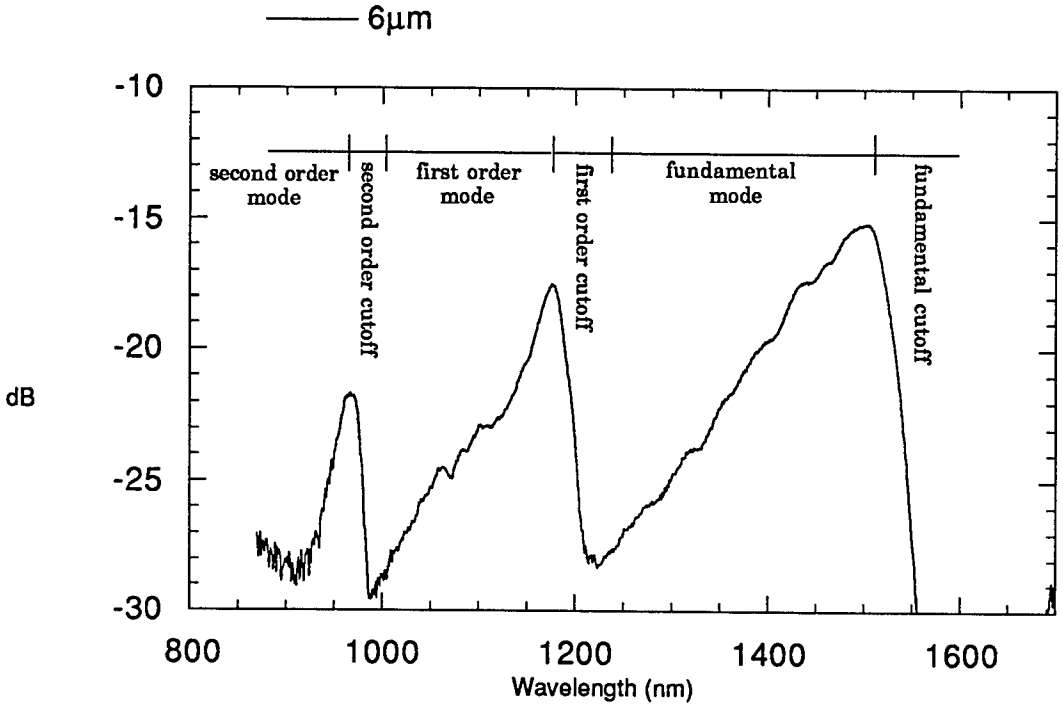


Fig.10. Scan of 6μm guide that indicates the regions of cutoff and modal operation.

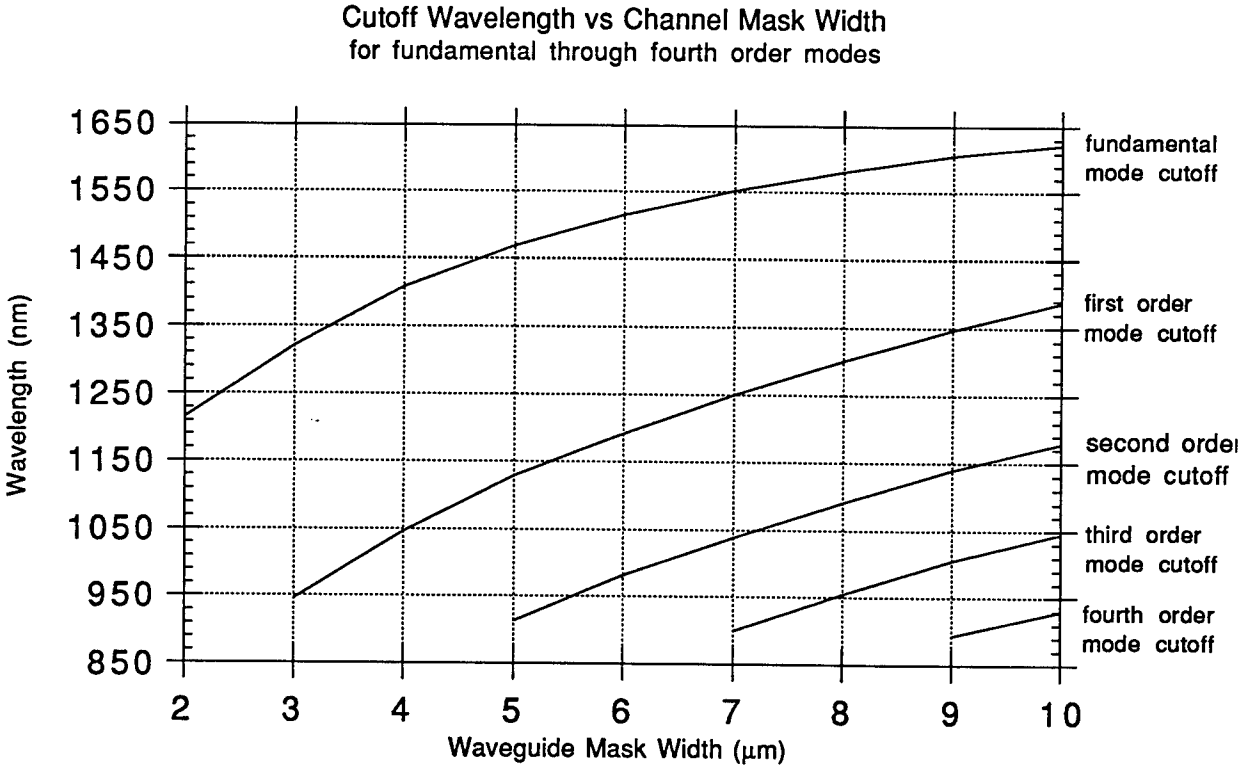


Fig.11. Cutoff wavelengths for each mode as a function of channel mask width.

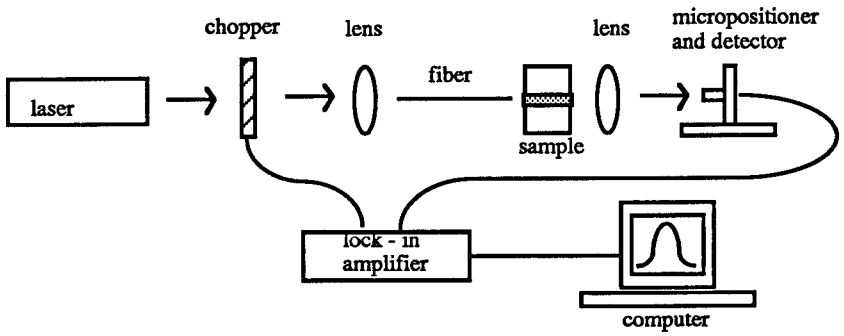


Fig.12. Experimental set-up used for near-field mode profile measurement.

VI. References

1. W. B. Spillman, Jr., N. A. Sanford, and R. A. Soref, "Optical waveguides in $LiTaO_3$ formed by proton exchange", *Opt. Lett.*, vol. 8, no. 9, p. 497, Sept. 1983.

2. J. L. Jackel, C. E. Rice, and J. J. Veselka, "Proton-exchange for high index waveguides in $LiNbO_3$,"*Appl. Phys. Lett.*, vol.41, p. 607, 1982.

3. X. F. Cao, R. Srivastava, R. V. Ramaswamy and J. Natour, "Recovery of Second-Order Optical Nonlinearity in Annealed Proton-Exchanged $LiNbO_3$," *IEEE Photon. Technol. Lett.*, vol.3, pp.25-27, 1991.

4. T. Findakly, P. Suchoski, and F. Leonberger, "High-quality $LiTaO_3$ integrated-optical waveguides and devices fabricated by the annealed-proton-exchange technique", *Opt. Lett.*, vol. 13, no. 9, p. 797, Sept. 1988.

5. H. Åhlfeldt, J. Webjörn, F. Laurell, and G. Arvidsson, "Postfabrication changes and dependence on hydrogen concentration of the refractive index of proton-exchanged lithium tantalate waveguides", *J. Appl. Phys.*, vol.75, No.2, p. 717, Jan. 1994.

6. P. Matthews and A. Mickelson, "Instabilities in annealed proton exchange waveguides in lithium tantalate," *J. Appl. Phys.*, vol.71, No.11, p.5310, June 1992.

7. T. Yuhara, K. Tada, and Y. S. Li, "Anomalous refractive index change and recovery of electro-optic coefficient r_{33} in proton-exchanged $LiTaO_3$ optical waveguides after annealing", *J. Appl. Phys.*, vol.71, No.8, p. 3966, April 1992.

8. X.F. Cao, R. V. Ramaswamy, and R. Srivastava, "Characterization of annealed proton exchanged $LiNbO_3$ waveguides for nonlinear frequency conversion", *J. Lightwave Technol.*, vol.10, No.9, p.1302, Sept. 1992.

9. J. E. Baran, Y. Silberberg, and P. Perlmutter, "White-light characterization of integrated-

optic waveguides and devices", *OFC/IOOC*, TUO2, 1987.

10. D. Maring, "Fabrication and white-light characterization of annealed proton exchanged channel waveguides in lithium tantalate," report submitted to Eglin AFB, October 1994.

11. G. B. Hocker and W. K. Burns, "Mode dispersion in diffused channel waveguides by the effective index method", *Appl. Opt.*, vol.16, pp. 113-118, 1977.

12. M. Fukuma and J. Noda, "Optical properties of titanium-diffused $LiNbO_3$ strip waveguides and their coupling-to-a-fiber characteristics," *Appl. Opt.*, vol.19, pp. 591-597, 1980.

13. K.Mizuuchi and K. Yamamoto, "Characteristics of periodically domain-inverted $LiTaO_3$,"*J. Appl. Phys.*, vol.72, No.11, p. 5061, Dec. 1992.

14. M. D. Feit and J. A. Fleck, Jr., "Light propagation in graded-index optical fibers," *Appl. Opt.*, vol.17, pp. 3990-3998, 1978.

15. P. Matthews and A. Mickelson, "Properties of proton exchange waveguides in lithium tantalate", *J. Appl. Phys.*, vol.72, No.7, p. 2562, Oct. 1992.

Superresolution of Passive Millimeter-Wave Imaging

Stanley J. Reeves
Assistant Professor
Department of Electrical Engineering

Auburn University
Auburn, AL 36849

Final Report for:
Summer Faculty Research Program

Wright Laboratory
Eglin Air Force Base

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

August 1994

Superresolution of Passive Millimeter-Wave Imaging

Stanley J. Reeves
Assistant Professor
Department of Electrical Engineering
Auburn University

Abstract

Passive millimeter-wave (PMMW) imagery has tremendous potential for imaging in adverse conditions; however, poor resolution poses a serious limitation to this potential. This project investigated the potential of various image acquisition and image processing strategies to superresolve PMMW images. Experimental results indicate that image processing alone cannot significantly superresolve general extended targets beyond the measured spatial frequency region. Furthermore, synthetic aperture acquisition techniques were found to be inappropriate in this setting.

A combination of acquisition strategies and image processing techniques holds the best promise for superresolution. We found that by designing the aperture weighting appropriately, one can make the acquired data more amenable to superresolution via image restoration techniques. In addition, we proposed a new acquisition method that allows one to increase the measured spatial frequency bandwidth in the data, which allows one to image beyond the diffraction limit. With these strategies, one should be able to improve resolution by a factor of three to four.

Superresolution of Passive Millimeter-Wave Imaging

Stanley J. Reeves

1 Introduction

Objects generally radiate across the entire frequency spectrum of electromagnetic energy. The radiation at visual and infrared wavelengths can be measured across a range of angles to form images. Images can also be acquired at millimeter wavelengths. Passive millimeter-wave (PMMW) images are particularly useful for navigation, guidance, and surveillance because millimeter waves penetrate fog and heavy rain and radiate at night as well as during daylight. Furthermore, millimeter waves penetrate smoke and small debris, making PMMW imagery especially advantageous under battlefield conditions [1]. Furthermore, the passive nature of this imaging modality makes stealth possible and reduces the opportunity for countermeasures. PMMW imaging also has a number of potential civilian applications, including aircraft landing, collision avoidance, and airport security screening.

A major drawback to PMMW images is poor angular resolution. Resolution is inversely proportional to wavelength and proportional to the size of the aperture. Because PMMW radiation has a comparatively long wavelength relative to visual radiation, the resolution is much lower. Furthermore, the use of very large apertures is not usually an option because of platform size constraints. For example, an aperture may have to fit within the diameter of a guided bomb or missile. To make PMMW useful for bomb guidance, the spatial resolution must be improved many times. In other applications, the necessary resolution improvement may not be as demanding, but higher resolution is still required to make these applications feasible.

Resolution improvement must be pursued via two strategies: image acquisition and image processing. It is our view that neither strategy is capable of solving the problem alone. The potential improvement of each strategy in isolation is quite limited, and some approaches may require both novel acquisition techniques and appropriate image processing of the resulting data to achieve their goal. Thus, this project has considered the two strategies in an integrated way. In the next section, we define the problem and discuss the assumptions required to do superresolution. In Section 3, we outline the most promising methods for superresolution and assess the potential of these methods in terms of the degree of resolution improvement. In Section 4, we explore the use of regularization to achieve superresolution and assess its performance. In Section 5, we propose a new method for superresolution that involves an innovative image acquisition strategy coupled with image processing to achieve the final superresolved image. In Section 6, we make recommendations on a number of different issues relating to the superresolution problem.

2 The Mathematics of Diffraction-Limited Imaging

To understand the nature of the superresolution problem as well as possible solutions, one needs a strong grasp of the mathematics that describe diffraction-limited imaging. Diffraction-limited imaging refers to the situation in which the aperture is small relative to the radiation wavelength. For simplicity, we consider the imaging equations in one dimension; extension to two dimensions is straightforward. We assume that the source is sufficiently distant from the aperture (in the Fraunhofer or far-field region) so that the wavefront from each radiator can be modeled as a plane wave. Furthermore, we assume that the radiation is essentially monochromatic with wavelength λ. Thus, the plane wave at time t from an angle θ from the aperture plane is $r(p)\exp\{j(\frac{2\pi}{\lambda}(ct+px)+\phi_p)\}$, where $p=\cos\theta$, x is the distance parallel to the aperture plane, and ϕ_p is a phase shift associated with a particular angle of arrival. Consider an aperture centered at $x=0$. If we assume that the elements of the aperture are insensitive to direction or that the scene exists only for a small angle from perpendicular to the aperture, then the waveform impinging upon the aperture element at a displacement x is a superposition of the waveforms from all angles[1]. Thus, we can express the waveform at location x as

$$F(x,t)=S(x)\int_p r(p)\exp\{j(\frac{2\pi}{\lambda}(ct+px)+\phi_p)\}dp, \tag{1}$$

where $S(x)$ is a weighting function that represents the gain of the aperture element at x. Note that this function is simply a weighted Fourier transform of $r(p)\exp\{j(\frac{2\pi}{\lambda}ct+\phi_p)\}$ with respect to p. For future reference, we denote the spatial response transfer function as $H(x)$.

The expression (1) defines the information about the source that is available at the aperture. Since the aperture is spatially limited, the spatial Fourier transform of the scene is only known over specified limits. This frequency-domain truncation is equivalent to a lowpass spatial filter of the image data, which limits the resolution of the image. The shape of the filter impulse response (the point-spread function or PSF) can be controlled by the choice of $S(x)$[2]. However, this only changes the weights on the Fourier components; those components that are zero remain zero regardless of the choice of $S(x)$.

As the radiation passes through the aperture, the information is transformed by the intervening lenses and stops. However, since higher spatial frequencies do not generate a response at the aperture, no system of lenses, optical materials, and stops can cause these frequencies to be measured. This is an important point that underscores the difficulty of the superresolution problem. In contrast to many other image deblurring problems for which higher spatial frequencies are attenuated in the acquired data, in this problem the higher spatial frequencies are eliminated altogether.

[1] Alternatively, we can replace $r(p)$ with $r(p)\sin\theta$ and then assume that the waveform measured at each aperture element is a superposition of the waveforms from all angles.

[2] The PSF is also referred to as the beam pattern when it is considered in terms of directional sensitivity at a particular point in the image plane

The intensity $r^2(p)$ rather than the complex value $r(p)\exp\{j\phi_p\}$ contains the information of interest in passive imaging, since ϕ_p is a random quantity unrelated to the emissivity or temperature of an object. If we denote the PSF of the imaging system as $h(p)$, we can express the intensity of the image at the focal plane as

$$\begin{aligned} |y(p,t)|^2 &= \int_u h(u)r(p-u)\exp\{j\frac{2\pi}{\lambda}ct + j\phi_{p-u}\}du \int_w h^*(w)r(p-w)\exp\{-j\frac{2\pi}{\lambda}ct - j\phi_{p-w}\}dw \\ &= \int_u\int_w h(u)h^*(w)r(p-u)r(p-w)\exp\{j(\phi_{p-u} - \phi_{p-w})\}du\,dw \end{aligned} \tag{2}$$

The processes that give rise to the radiation are essentially independent from one pixel to the next (noncoherent source), so that the phase ϕ_p is independent and uniformly distributed in $[0, 2\pi)$. Thus,

$$E[\exp\{j(\phi_u - \phi_w)\}] = \delta(u-w) \tag{3}$$

For wideband radiation with bandwidth $\Delta\omega$, the phase terms can be considered independent in time approximately every $\frac{2\pi}{\Delta\omega}$ seconds. Thus, one can estimate the expected value of $|y(p,t)|^2$ by integrating in time:

$$\begin{aligned} < |y(p,t)|^2 > &= \lim_{T\to\infty}\frac{1}{T}\int_0^T |y(p,t)|^2 dt \\ &= E[|y(p,t)|^2] \\ &= \int_u\int_w h(u)h^*(w)r(p-u)r(p-w)\delta(u-w)du\,dw \\ &= \int_u |h(u)|^2 r^2(p-u)du\,dw \\ &= |h(p)|^2 * r^2(p) \end{aligned} \tag{4}$$

The result is a convolution of the object intensity distribution with the square of the aperture PSF. The inverse Fourier transform of $|h(p)|^2$ is a convolution of $H(x)$ with itself (assuming a symmetric distribution). If $H(x)$ has a uniform distribution as shown in Figure 1, then the inverse transform will have a triangular shape with twice the width of the original $H(x)$ (Figure 2).

Note that the spatial frequencies are attenuated increasingly as x increases. In 1-D the half-power point is at $0.586\frac{2\pi}{\lambda}x_c$ as compared to $\frac{2\pi}{\lambda}x_c$ for the basic uniform aperture response. In this sense the bandwidth of the intensity imaging scheme is lower than that of the basic aperture spatial frequency response $H(x)$. However, there is an important advantage to the intensity response. Although some higher frequencies are attenuated more than for the uniform response, the intensity imaging system has a nonzero response to some spatial frequencies for which the uniform case has zero response. This means that some frequencies that are filtered out by the uniform response are still present for filtering in the intensity case. The presence of these

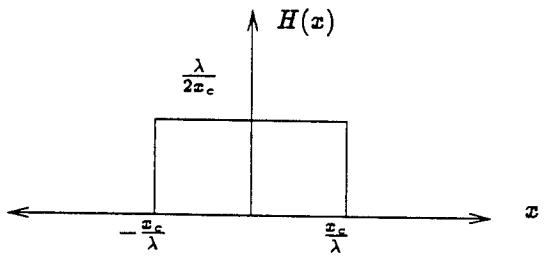


Figure 1: Angular frequency response of uniformly weighted aperture.

spatial frequencies provides a significant opportunity for image restoration techniques to recover most of the nonzero bandwidth $2\frac{2\pi}{\lambda}x_c$ of the scene.

Resolution improvement beyond this nonzero bandwidth $2\frac{2\pi}{\lambda}x_c$ is an extremely challenging task. In fact, it is impossible to go beyond this level of resolution with image processing apart from some a priori knowledge of the scene. The sampling theorem implies that the information available at the aperture can be uniquely represented by a set of samples in the spatial image domain with a spacing of $\frac{\lambda}{8\pi x_c}$. Thus, samples spaced more closely than this will be functions of adjacent samples and will not carry any more information about the signal [2].

The sampling theorem assumes that frequencies higher than $2\frac{2\pi}{\lambda}x_c$ are zero. However, if a priori information about the signal is available, one may be able to predict higher spatial frequencies from certain patterns formed by the lower frequencies. If these higher frequencies can be uniquely determined, this in turn specifies a more dense set of spatial-domain samples that are required to represent the larger bandwidth. The prediction of these higher spatial frequencies is the domain of a priori information.

3 A Priori Knowledge and Superresolution

With some a priori knowledge, it may be possible to determine a unique set of samples that are more closely spaced than the sampling theorem indicates [3]. A priori knowledge has the effect of eliminating solutions that are consistent with the data but inconsistent with knowledge of the scene under consideration. A number of assumptions are available for consideration for the superresolution problem.

One assumption that is sometimes used is that the image has a finite region of support. This assumption has been developed and implemented from a number of different perspectives [4, 5]. Region of support information is quite powerful in the context of astronomical imaging, in which a large part of the scene can be considered to be zero. However, the assumption that the image has a finite region of support is in general inappropriate for the problem considered here.

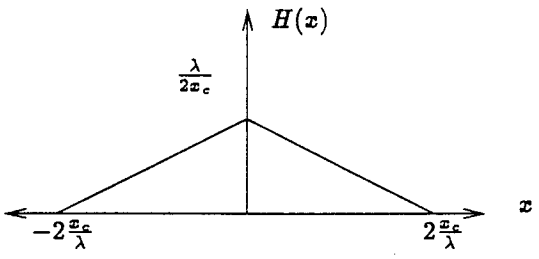


Figure 2: Angular frequency response for amplitude imaging with uniformly weighted aperture.

One assumption often used with more success in a general imaging context is that the image is nonnegative. This assumption holds for image intensity, which is normally the quantity of interest in PMMW imaging. The value of this assumption depends primarily on two factors: the amount of noise and the ratio of high-frequency intensity fluctuations to DC intensity. If the noise level is high, infeasible image solutions — including images with negative pixels — are more likely. In that case the nonnegativity constraint will be exercised and will rule out these infeasible solutions. Unfortunately, the nonnegativity constraint contributes nothing to superresolution when noise is the cause of the constraint being exercised. On the other hand, the absence of some high-frequency information may also cause the image solution to go negative. In this case, a nonnegativity constraint will force the missing high-frequency information to appear, which will accomplish a degree of superresolution in the process. Unfortunately, a complex scene's average intensity may be so high that missing high-frequency information does not cause the image to go negative. In this case, a nonegativity constraint is not exercised at all and thus does not accomplish any superresolution.

Edges are widely recognized as the basic "building blocks" of an image. Intensity edges generally reflect object boundaries in a scene, and these are extremely important in both human and computer vision [6]. In fact, some recent work has shown that certain classes of images can be completely reconstructed from knowledge of the edge locations and heights [7, 8]. The informational dominance of edges in real-life images can be used as a priori information in a superresolution scheme. Not surprisingly, a number of techniques have been developed to interpolate images by exploiting edge structure that can be detected in the low-resolution image [9–12]. In essence, each of these methods attempts to preserve the intensity discontinuities of oriented structures (edges). Using any of a number of different mechanisms, the image is interpolated along directional contours where they exist and is interpolated in a non-oriented fashion where no directionality is evident. These methods significantly improve the quality of interpolated images as compared to standard non-oriented interpolation schemes, such as splines or sinc-function interpolation. Blurring of edges is minimal since interpolation is done only along edges rather than across them.

Unfortunately, these methods have drawbacks as well. Although they reduce (and in some cases eliminate)

blurring of edges, they cannot determine the appropriate level of sharpness in a high-resolution edge from a low-resolution representation of the edge. Therefore, the edge may be either too smooth or too sharp in the interpolated image. More seriously, these techniques cannot distinguish edges that are spaced less than a pixel apart. Thus, no new details are created in the interpolation process. While the presence of edges could be an important a priori assumption, none of the current techniques appear to be capable of exploiting this assumption to do any significant superresolution.

Other structural assumptions, such as the existence of point targets, can also be of value when the assumptions are valid for the data under consideration. However, algorithms that rely on detecting these features in the low-resolution image risk being unable to distinguish overlapping features as well as detecting structure due to noise where no actual structure exists. This risk exists in the constrained iterative deconvolution (CID) algorithm of Richards *et al.* [13] as well as the differentiation-integration deconvolution (DID) algorithm of Ding [14]. Furthermore, these algorithms suffer from a philosophical problem for some applications: if the features of interest can be detected in the low-resolution image, then superresolution is unnecessary. Finally, the assumption of point targets, an integral part of the CID algorithm, is inappropriate for imaging of extended targets.

Another potentially powerful assumption is that the image can be modeled by a limited number of basis images. If these basis images contain frequencies both inside and outside the measured range, the measured frequency components can be used to determine the contribution of each basis image which together implicitly predict the unmeasured frequency components. This principle is the basis for the use of space-variant regularization, which we discuss in the next section.

4 Regularization

4.1 Background

Regularization is a method for converting an ill-posed and/or underdetermined problem into a problem with a unique, stable solution. An importance class of regularization techniques introduces a bias into the minimization criterion that defines the possible solutions. The bias is defined so that certain solutions are either disqualified or heavily penalized. Typically, the regularization term biases the result toward a smooth solution [15].

Regularization has been used successfully in a wide variety of related image processing problems — computer vision [16, 17], segmentation [18], restoration [19], and zooming [20]. The unifying feature of these problems is that the solution is known a priori to be smooth in some sense. By defining the problem so that rough solutions are penalized, one obtains a solution that more closely fits this a priori knowledge. The type of smoothing can be controlled in various ways to achieve a number of alternative results.

If we discretize the imaging equations, denoting an image as an ordered 1-D vector and the bandlimited imaging process as a matrix, we can write the imaging equations as follows:

$$g = Df + n, \tag{5}$$

where g is the low-resolution image, D is the filtering (and possibly downsampling) operation, f is the original high-resolution image, and n is noise. The simplest way to estimate the original image is to minimize the following with respect to f:

$$\sum_i (g_i - [Df]_i)^2. \tag{6}$$

Unfortunately, this criterion suffers from both noise sensitivity and nonuniqueness. However, by adding a regularization term to the criterion, we can reduce noise sensitivity and guarantee a unique solution:

$$\sum_i (g_i - [Df]_i)^2 + \alpha \sum_i [Lf]_i^2, \tag{7}$$

where L is a regularization operator that describes the type of smoothness imposed on the solution and α controls the degree of smoothness imposed [19]. The operator L is usually chosen to emphasize high frequencies so that these high frequencies tend to be deemphasized in the minimization of (7). The operator is often chosen to be a discrete Laplacian [21]. If L represents a linear shift-invariant (LSI) system (as in the case of the discrete Laplacian), no new spatial frequencies will appear in the solution. Thus, this type of regularized solution can only help restore frequency components that already exist in the measured data; it does not increase the bandwidth of the solution beyond that.

The criterion (7) can be modified to introduce new frequencies by making the degree of smoothing spatially dependent. This modification is justified by observing that images are not uniformly smooth. Instead, they generally show smooth regions separated by sharp transitions (edges). This observation motivates the use of spatially dependent smoothing, in which smoothness is enforced in some regions while sharp transitions are allowed between regions. The criterion is modified as follows:

$$\sum_i (g_i - [Df]_i)^2 + \alpha \sum_i s_{ii} [Lf]_i^2, \tag{8}$$

where the s_{ii} terms are weights on the filtered image pixels $[Lf]_i$. By making some weights relatively small, we reduce the smoothing enforced at those locations [22]. This has the effect of allowing sharp transitions at those locations, which can introduce new frequencies into the solution. We have shown in 1-D simulations that we can exactly reconstruct a signal composed of step edges simply by zeroing the smoothing at the locations of the discontinuities.

The primary difficulty in applying this method is that one must have a priori information about the locations of the reduced smoothing. In [22], the weights are estimated from the blurred (low-resolution) image. However, this is not likely to be very useful for the superresolution problem. If the weights that determine sharp transitions are themselves determined from a low-resolution image, then they are not likely to induce transitions that would be considered sharp at a higher resolution. For example, if two edges lie parallel to one another in close proximity, they may appear as a single edge in the low-resolution image. In this case, one cannot determine from the low-resolution image that the smoothness should be enforced between the edges. The edges may appear as a single edge in the solution, or the intensity may oscillate wildly in the region between the two edges.

A number of solutions have been proposed to intelligently estimate the appropriate weights from the data in the context of image restoration. Katsaggelos proposed the use of an adaptive iterative scheme in which the weights are estimated simultaneously with the restored image [23]. Reeves proposed an alternative scheme in which a sequence of restored images are computed and the weights estimated successively from each image in the sequence [24]. Neither of these techniques has been applied to the superresolution problem.

Other methods have been proposed that implicitly adapt the smoothing weights to the image as the solution is generated. Some of these techniques have been motivated in the context of a maximum a posteriori (MAP) estimation framework, although they can also be thought of as adaptive regularization schemes. Ground-breaking work in this area was done by Geman and Geman [18], who proposed the use of line processes that turn smoothing off during the estimation process if the local data indicate that the region is a boundary. The line process prevents smoothing across boundaries in the image. Unfortunately, the resulting criterion is highly nonconvex, which results in multiple local minima. As a consequence, the criterion must be minimized by simulated annealing, an extremely computationally intensive technique. Furthermore, this criterion is quite aggressive in finding edges and may build sharp edges where no edges or very subtle edges exist in the actual image.

A more appealing MAP estimation technique was proposed by Hebert and Gopal in the context of medical image reconstruction [25]. This criterion adopts a nonquadratic but convex penalty function for roughness, which enforces a smooth solution. However, at locations for which the data warrants a rough estimate, the penalty term does not continue to penalize the discrepancy quadratically. Thus, it allows for the possibility of outliers (edges) without giving up the smoothing property. In addition, because the criterion is convex, it can be optimized with gradient search techniques and converges to a unique solution. A related technique was proposed by Zervakis for the image restoration problem [26], although the problem was not formulated in a MAP estimation framework.

Another MAP estimation procedure was formulated by Schultz and Stevenson for the problem of image expansion [20]. This technique seeks to expand or interpolate an image so that edges are emphasized

in the expanded image. Their regularization criterion is essentially the same as the techniques described above [25, 26]; the criterion is convex and can be optimized by gradient search techniques. It can be shown that the results of these methods are approximately equivalent to the adaptive and multistage methods discussed previously.

4.2 Application to superresolution

The superresolution problem can be thought of as a combination of an image restoration and an image expansion (interpolation) problem. The attenuated spatial frequency data must be restored appropriately. In addition, the restoration must be done at a higher resolution than the recorded data, and new spatial frequencies must be predicted from the existing data in the process. Although regularization has never been applied successfully to the superresolution problem, it can be adapted from the applications discussed above in a very straightforward way. We define the superresolution criterion as follows:

$$\sum_i (g_i - [Df]_i)^2 + \alpha \sum_i \sum_j \rho([L_j f]_i), \tag{9}$$

where the L_j define regularization operators for different directions, and $\rho(\cdot)$ is a nonquadratic distance measure [20].

The Huber function has been used as $\rho(\cdot)$ by a number of researchers [25, 26, 20] to limit smoothing at edges in images. It is convex and nonquadratic. The Huber function is defined as follows:

$$\rho_T(x) = \begin{cases} x^2, & |x| \le T, \\ T^2 + 2T(|x| - T), & |x| > T. \end{cases} \tag{10}$$

This function is quadratic as long as the magnitude of the argument is less that T. However, when the argument magnitude increases beyond T, the function increases linearly instead of quadratically. The threshold T can be tuned to match the edge characteristics expected in a particular type of image. Four L_j operators are chosen to approximate second derivatives horizontally, vertically, and in each diagonal direction.

A number of experiments were performed using Huber regularization. We concentrated on the case in which the low-resolution image has no noise in it. Although this is unrealistic, it provides us with an estimate of the most optimistic potential for superresolution. The building image (Figure 3(a)) was blurred by the 2-D intensity response to a uniform circular aperture with a radius of $\frac{\pi}{4}$ radians per sample. The acquired image is shown in Figure 3(b). Finally, the restored image using Huber regularization with an optimized threshold T is shown in 3(c). The Huber image showed an improvement in mean square error (MSE) of a few percent as compared to the pseudoinverse restoration. However, the two images were visually indistinguishable.

Although a tiny amount of higher spatial-frequency information was reconstructed by the Huber regu-

larization, it is unclear how to quantify it. Ding [14] has proposed comparing the reconstructed image to a bandlimited image with a variable bandwidth corresponding to aperture size. The size whose image corresponds most closely to the reconstructed image yields an equivalent aperture achieved by the superresolution scheme. By this measure, the Huber reconstructed image is not superresolved beyond the psedudoinverse image. Both of these images correspond to a doubling of the aperture size. Thus, Huber regularization appears to be incapable of improving the resolution significantly beyond the measured frequency range. In the case of a noisy image, a gain of around 1.5 is probably more realistic. For the noisy case, the Huber regularization strategy (or any other adaptive regularization strategy) would perform significantly better than the pseudoinverse filter.

5 Image Acquisition

While image processing has the potential to improve the resolution of an image, the degree of potential improvement is quite limited. However, it may be possible to improve the resolution further by adopting new image acquisition strategies. We consider some options here.

5.1 Aperture weighting

If the image acquisition noise is a function of the overall signal level rather than at a fixed level, one can shape the aperture response to improve the superresolution potential. This technique, called apodization, is normally used to reduce tails in the PSF at the expense of increasing the mainlobe width. Although that strategy may produce a more subjectively pleasing image without any further processing, it actually reduces the resolution of the resulting image. If image acquisition is to be followed by superresolution processing, then the exact opposite procedure should be used. This has the effect of measuring higher spatial frequencies at a higher signal-to-noise ratio (SNR) relative to a uniform aperture response. Since these higher frequencies contain important information for superresolution, it is imperative that they be measured with an SNR as high as possible. Figure 2 shows the resulting spatial frequency when the aperture response is uniform. However, if the aperture response is shaped as in Figure 4, the resulting spatial frequency response is given by Figure 5. One can see that higher frequencies are not attenuated as drastically as in the uniform case. Thus, they can be more easily restored since the SNR is higher at those frequencies.

The optimization of the aperture response can be formulated rigorously as a problem of minimizing the expected mean-square error in the superresolved image as a function of aperture response. This should yield significantly better results than using a uniform aperture response.

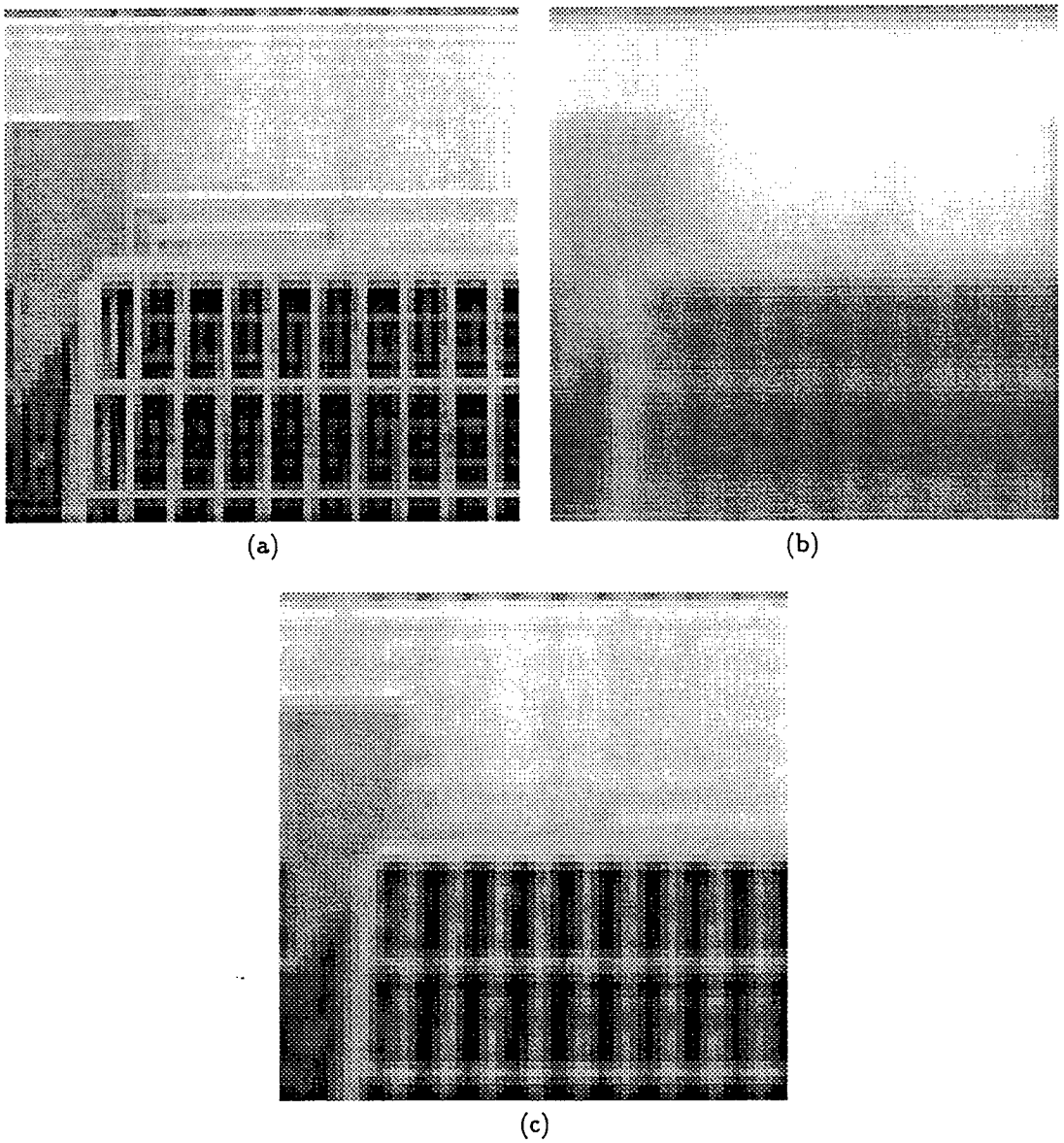


Figure 3: Regularization experiment: (a) original image, (b) intensity image acquired by uniform aperture, (c) Huber regularized reconstruction

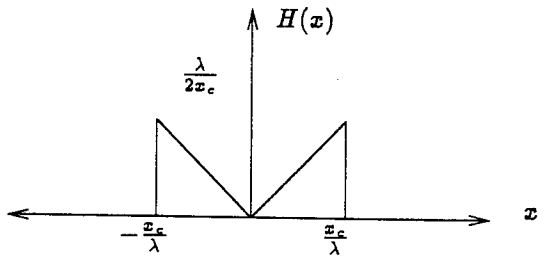


Figure 4: Angular frequency response of weighted aperture.

5.2 Synthetic aperture imaging

The concept of synthesizing a larger aperture by using a set of smaller apertures is used in a wide variety of settings. Two settings that are closely related to PMMW imaging are radio astronomy [27] and synthetic aperture radar (SAR) [28]. In radio astronomy, a distributed set of fixed apertures is used to improve the resolution of astronomical images. In SAR, the aperture moves along a path perpendicular to the direction to the object and sequentially traces out a distributed aperture. In both cases, coherent summation of the detected signals at each aperture gives the equivalent response of a single larger aperture.

Each case has its own similarities to PMMW imaging. Radio astronomy is similar in that the radiation arises from the scene rather than being transmitted by the observer. SAR is similar in that in most situations one must synthesize an aperture from a moving platform. Is it possible to do coherent summation for a noncoherent source? Clearly, that is exactly what is done in radio astronomy. As long as the phase shift from all angles of arrival is small across the entire aperture, one can synthesize a larger array. However, it is not clear that one can accomplish this with a single array that must move from one place to another, in spite of such assertions in the literature [29]. The problem is that the coherence time of the radiation is a fraction of $\frac{2\pi}{\Delta\omega}$. If the aperture is moving sideways relative to the scene at 1000 m/s, it will take 0.001 seconds to trace out a 1-meter aperture (in 1-D). If the bandwidth $\frac{\Delta\omega}{2\pi}$ is 10 GHz, the 1-meter aperture trace time is seven orders of magnitude longer than the coherence time of the radiation! Although it may be possible to reduce the bandwidth, a reduction of seven orders of magnitude is probably infeasible. Thus, it appears that SAR is not a viable image acquisition strategy for obtaining PMMW imagery at higher resolutions.

5.3 High-order imaging

Earlier we observed that by averaging intensity over time, the PSF becomes $|h(p)|^2$ instead of $h(p)$. The squared PSF has a Fourier transform that is twice as wide as the original PSF Fourier transform, which allows for the possibility of achieving twice the resolution obtainable with the original PSF. We have discovered

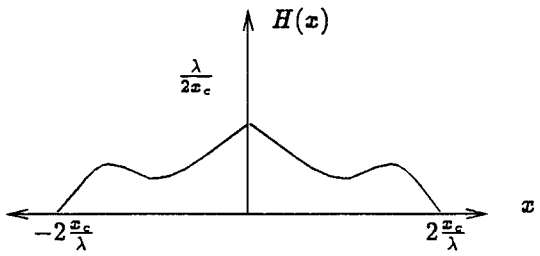


Figure 5: Angular frequency response for intensity imaging with weighted aperture.

that the concept of raising the amplitude to a higher power can be extended to powers greater than two. In fact, if we average the squared intensity over time, we get the following:

$$\begin{aligned}
< |y(p,t)|^4 > &= \lim_{T\to\infty} \frac{1}{T}\int_0^T |y(p,t)|^4 dt \\
&= E[|y(p,t)|^4] \\
&= \int_a\int_b\int_u\int_w h(a)h^*(b)h(u)h^*(w)r(p-a)r(p-b)r(p-u)r(p-w) \\
&\quad \times[\delta(a-b)\delta(u-w)+\delta(a-w)\delta(b-u)-\delta(a-b)\delta(u-w)\delta(a-u)]da\,db\,du\,dw \\
&= 2[\int_u |h(u)|^2 r^2(p-u)du]^2 - \int_u |h(u)|^4 r^4(p-u)du \\
&= 2[|h(p)|^2 * r^2(p)]^2 - |h(p)|^4 * r^4(p) \qquad (11)
\end{aligned}$$

The fourth-order image is then a function not only of the low-resolution intensity image but also a higher-resolution intensity-squared image. By subtracting a measured version of $2[|h(p)|^2 * r^2(p)]^2$, one can isolate the term $|h(p)|^4 * r^4(p)$. Thus, we can resolve beyond the diffraction limit with this scheme. The issues associated with the use of the intensity image also apply to this image. First, the convolution with $|h(p)|^4$ may attenuate higher frequencies even faster than $|h(p)|^2$. However, the Fourier transform is also twice as wide, meaning that the measured bandwidth is twice that of the intensity image. Thus, the measured fourth-order data has an increased potential for superresolution since these higher frequencies still exist in the data. Second, the expected value is only approximately achieved in a finite amount of time. Consequently, this model will have a noise variance that is inversely related to the time available for acquisition. Third, one can design the aperture weighting $S(x)$ to maximize SNR in the acquired image with the goal of superresolution in view.

To verify the validity of this method, we generated a 64-point square wave (a bar pattern in 2-D) and simulated an intensity image as well as an intensity-squared image. The signal was replicated 64 times to form a 64×64 image for display purposes (Figure 6(a)). The square wave was used as the amplitude with

the phase at each sample chosen randomly. The aperture size was chosen so that the spatial frequency of the square wave was higher than the highest measured frequency in the intensity image. Thus, one would expect a response only at DC in the intensity image. An intensity image and an intensity-squared image were formed by a Monte Carlo simulation taking 100,000 independent phase realizations and averaging the resulting signals. The resulting intensity signal is shown in 6(b) and the intensity-squared signal in 6(c). Note that the intensity signal has only a DC component as expected. However, the intensity-squared signal shows a frequency component at the same frequency as the square wave. Of course, this signal is blurred since even the intensity-squared signal does not pick up all frequencies of the square wave. This simulation confirms that higher frequencies can be imaged by the higher-order imaging scheme that are completely missed by a conventional intensity imaging scheme.

The concept behind the use of fourth-order images can be generalized to higher orders with a consequent increase in the measured bandwidth. Theoretically, one can continue this process to achieve any desired bandwidth (and resolution). Unfortunately, as the order is increased, the assumption that the expected value is achieved for a given amount of time becomes an increasingly poor assumption. Thus, noise becomes much more serious as the order is increased. Further analysis will shed light on the practical limitations of this approach.

In order to exploit higher-order imaging, one must determine how to combine the information in the second-order image (the intensity image) with the information in the fourth-order image. The relative weights given to each image will depend on a number of factors and may be functions of spatial frequency, spatial location, and the image itself. Furthermore, a clear understanding of the noise associated with the expected value assumption will be needed to appropriately process the data. Finally, an algorithm must be developed to incorporate the combined data and the noise model into a single image estimate.

6 Recommendations

- Image acquisition and image processing strategies should continue to be pursued in an integrated way.
- Optimal aperture weighting is an important tool for successful superresolution. The criterion for optimal weighting should be defined rigorously and an optimal aperture weighting derived.
- SAR should be abandoned as a viable superresolution strategy.
- A thorough study of the noise associated with the acquisition process should be undertaken. Noise arises from both thermal processes as well as the assumption that the acquired data is equal to its expected value. A comprehensive understanding of the noise processes is essential to intelligently address the image processing component of superresolution.

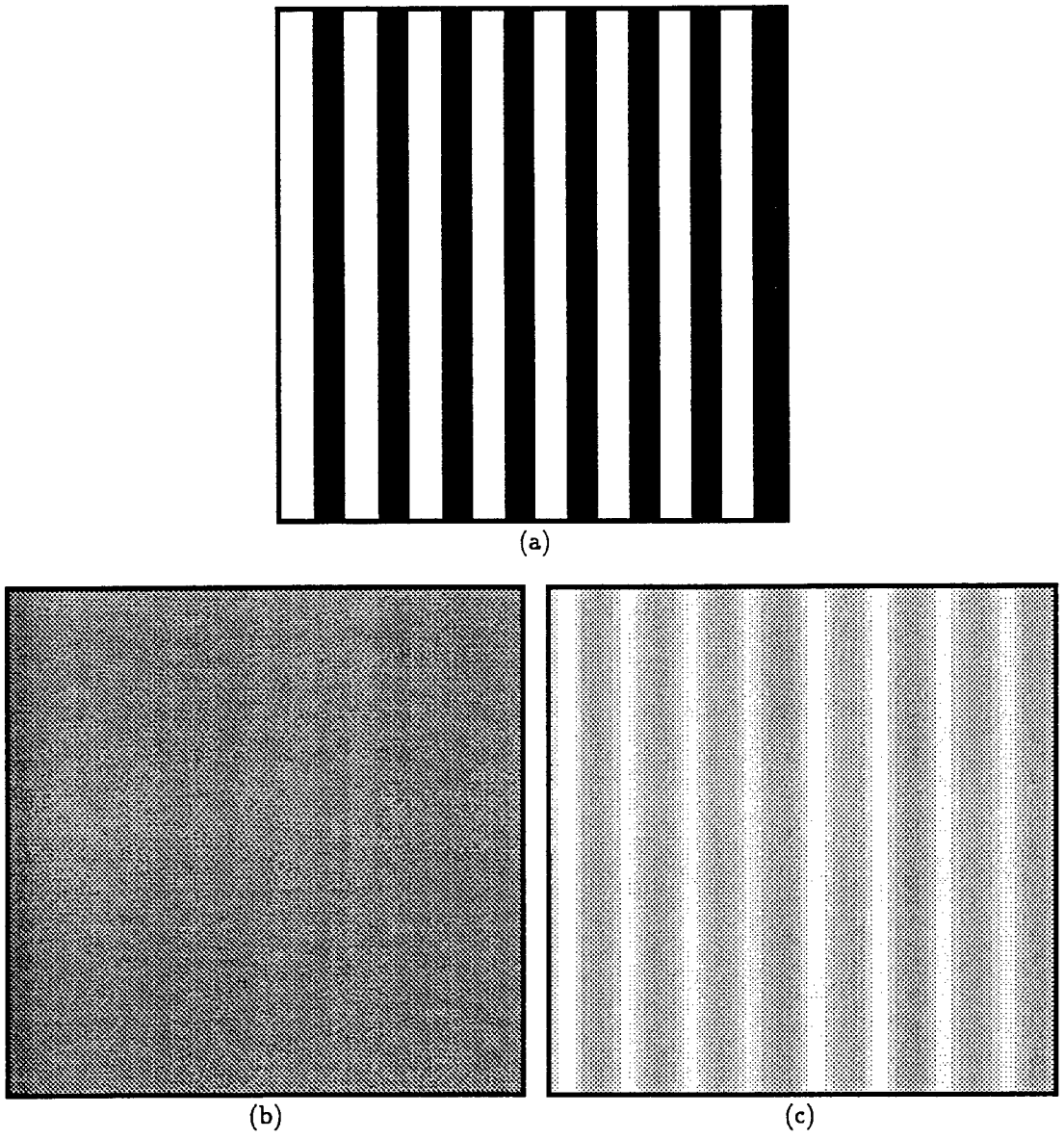


Figure 6: Higher-order imaging experiment: (a) original signal, (b) signal acquired by averaging intensity after passing through aperture, (c) signal acquired by averaging squared intensity after passing through aperture

- Imaging hardware must sample at least at the spatial density dictated by the sampling theorem. A higher sampling density is unnecessary, although there may be practical reasons for making a higher sampling density desirable.
- The optics should be carefully designed to minimize spatial variation in the PSF. Furthermore, this variation may need to be incorporated into a superresolution algorithm.
- The effect of the detector geometry in the image plane should be carefully modeled and optimized for superresolution.
- The effects of wideband radiation on the imaging model should be understood more thoroughly. For example, one must take the wideband characteristic into account when deriving an equivalent PSF for a given aperture.
- The atmospheric absorption curve and black-body radiation curve should be taken into account when characterizing the temporal frequency content of the incident radiation.
- A more complete understanding of higher-order imaging should be pursued in terms of noise, required integration time, and image processing strategies.
- The effect of a near-field scene on the imaging model should be studied so that the error can either be compensated or corrected during processing.
- A set of test images should be made available as standards for comparing proposed superresolution algorithms. Some of these images could be simulated from high-resolution images so that ground truth would be available. A series of images should be generated that are progressively more realistic (and more difficult).
- A MAP estimation technique that incorporates line processes should be investigated for creating sharper lines in images.

7 Conclusions

Superresolution is a very challenging problem. Our experiments suggest that image processing alone has an extremely limited potential for superresolving images — in the neighborhood of a 2:1 improvement. In light of this, how are we to understand claims of much larger superresolution ratios? If one assumes that salient features such as edges are highly structured and well-separated, then a number of methods can be used to predict the high-resolution version from the low-resolution image. However, these techniques are useless for

separating details that are separated by a fraction of the main lobe width of the PSF. Thus, they do not offer a general solution to the superresolution problem.

The most promising avenue for solving the superresolution problem is to continue to develop integrated acquisition and image processing strategies. With the further development of these strategies, one should be able to improve resolution by a factor of three to four.

Acknowledgment

I would like to thank Bryce Sundstrom, my contact point at Eglin, for his readiness to provide technical help and for his contagious enthusiasm for the topic of PMMW superresolution.

References

[1] R. Appleby, D. G. Gleed, and R. N. Anderton, "High-performance passive millimeter-wave imaging," *Optical Engineering*, vol. 32, pp. 1370–1373, June 1993.

[2] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* New Jersey: Prentice-Hall, 1989.

[3] S. P. Luttrell and C. J. Oliver, "Prior knowledge in synthetic-aperture radar processing," *Journal of Physics D: Applied Physics*, vol. 19, pp. 333–356, 1986.

[4] J. W. Goodman, *Introduction to Fourier Optics.* San Francisco: McGraw-Hill, 1968.

[5] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta*, vol. 14, no. 9, pp. 709–720, 1979.

[6] J. Aloimonos and D. Shulman, "Learning early-vision computations," *Journal of the Optical Society of America A: Optics and Image Science*, vol. 6, pp. 908–919, June 1989.

[7] R. Alter-Gartenberg, F. O. Huck, and R. Narayanswamy, "Compact image representation by edge primitives," *Journal of the Optical Society of America A*, vol. 7, pp. 898–911, May 1990.

[8] R. Alter-Gartenberg, F. O. Huck, and R. Narayanswamy, "Compact image representation by edge primitives," *CVGIP: Graphical Models and Image Processing*, vol. 56, pp. 1–7, January 1994.

[9] K. Jensen and D. Anastassiou, "Spatial resolution enhancement of images using nonlinear interpolation," in *Proceedings of the 1990 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2045–2048.

[10] V. R. Algazi, G. E. Ford, and R. Potharlanka, "Directional interpolation of images based on visual properties and rank order filtering," in *Proceedings of the 1991 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3005–3008.

[11] R. H. Bamberger, "A method for image interpolation based on a novel multirate filter bank structure and properties of the human visual system," in *SPIE Vol. 1657 — Image Processing Algorithms and Techniques III*, pp. 351–362, 1992.

[12] G. E. Ford, R. R. Estes, and H. Chen, "Space scale analysis for image sampling and interpolation," in *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. III–165–III–168.

[13] M. A. Richards, C. E. Morris, and M. H. Hayes, "Iterative enhancement of noncoherent radar data," in *Proceedings of the 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1929–1932.

[14] Z. Ding, "Resolution enhancement of passive millimeter-wave imaging," Final Report: Summer Faculty Research Program, AFOSR, 1993.

[15] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems.* Winston, 1977.

[16] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.

[17] D. Terzopoulos, "Regularization of inverse problems involving discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 413–424, July 1986.

[18] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, November 1984.

[19] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration," *Optical Engineering*, vol. 29, pp. 446–454, May 1990.

[20] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, pp. 233–242, May 1994.

[21] A. K. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "Non-stationary iterative image restoration," in *Proceedings of the 1985 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 696–699.

[22] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1874–1888, Dec. 1988.

[23] A. K. Katsaggelos, "Iterative image restoration algorithms," *Optical Engineering*, vol. 28, pp. 735–748, July 1989.

[24] S. J. Reeves, "Optimal space-varying regularization in iterative image restoration," *IEEE Transactions on Image Processing*, vol. 3, pp. 319–324, May 1994.

[25] T. J. Hebert and S. S. Gopal, "The GEM MAP algorithm with 3-D SPECT system response," *IEEE Transactions on Medical Imaging*, vol. 11, pp. 81–90, March 1992.

[26] M. E. Zervakis and T. M. Kwon, "Robust estimation techniques in regularized image restoration," *Optical Engineering*, vol. 31, pp. 2174–2190, October 1992.

[27] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy.* New York: John Wiley & Sons, 1986.

[28] D. L. Mensa, *High Resolution Radar Imaging.* Dedham, MA: Artech House, 1981.

[29] K. Komiyama, "High resolution imaging by supersynthesis radiometers (SSR) for the passive microwave remote sensing of the Earth," *Electronics Letters*, vol. 27, p. 389ff., February 1991.

EVALUATION OF NETWORK ROUTERS IN REAL–TIME PARALLEL COMPUTERS

Arindam Saha
Assistant Professor
Department of Electrical and Computer Engineering, and
NSF Engineering Research Center for Computational Field Simulations

Mississippi State University
P.O. Drawer EE
Mississippi State, MS 39762

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

July 1994

EVALUATION OF NETWORK ROUTERS IN REAL–TIME PARALLEL COMPUTERS

Arindam Saha
Assistant Professor
Department of Electrical and Computer Engineering, and
NSF/ERC for Computational Field Simulations
Mississippi State University

Abstract

Although general–purpose parallel processing research has been extremely vigorous over the last few years, real–time parallel processing has not received reciprocal attention. The two fields have some commonalities but the differences are many and significant. The unique requirements of real–time applications and the resultant ramifications on real–time parallel computers should be addressed in an adequate fashion. Like general–purpose parallel processing, a real–time parallel computer is as good as its communication capabilities. The crux of the communication potential lies in the underlying communication network and the router architecture.

Unlike the conventional parallel computers, little attention has been paid to the unique problem of efficient communication in real–time, application–specific, hard deadline parallel computers. These real–time applications contribute the added complications of *when* a task is completed and ultra–dependability in addition to the need of fast execution. Fast, predictable, schedulable, and dependable communication in real–time systems can be achieved by considering the various aspects of network characteristics, topology, switching methods, and router hardware *together* and study the interplay between them. The purpose of the research reported here is to evaluate a set of communication characteristics that are deemed desirable for real–time parallel processors with respect to performance and implementation cost.

A candidate network function is accurately defined and this characterization is detailed enough to be simulated. A concise, time–driven, flit–based, priority–driven, wormhole–routed, network simulator has been designed in C with a user–friendly Graphical User Interface (GUI). The network in general and the router in particular are simulated in great detail. Experimentation is performed by monitoring the latency and the throughput with variations in different parameters including the number and width of virtual channels, number of priority levels, maximum message length, load factor, and network size. Initially the destination address, message length and message priority are generated randomly with a uniform distribution. Then, various non–uniformities are introduced to mimic realistic applications. Results are plotted and analyzed.

EVALUATION OF NETWORK ROUTERS IN REAL–TIME PARALLEL COMPUTERS

Arindam Saha

1. Introduction

The performance of parallel computers depends considerably on the availability of fast, scalable interprocessor communication. In point–to–point networks, message exchange is limited by the lack of a shared communication medium. As a message makes its way through different nodes in the network, it incurs a delay at every hop which increases substantially as the communication changes from chip to backplane to box. Thus communication with processors just a few hops away can soon become prohibitively expensive. Considerable research in recent years has focused on reducing the message latency in general–purpose parallel computers. Unfortunately, little attention has been paid to the unique problem of efficient communication in real–time, application–specific, hard deadline parallel computers. These real–time applications contribute the added complications of *when* a task is completed and ultra–dependability in addition to the need of fast execution.

To tackle the problem of fast, predictable, schedulable, and dependable communication in real–time systems, one must consider the various aspects of network characteristics, topology, switching methods, and router hardware *together* and study the interplay between them. The purpose of the research reported here is to evaluate a set of communication characteristics that are deemed desirable for real–time parallel processors with respect to performance and implementation cost. This work will eventually lead to the design of an improved real–time network router in the future.

The research begins with the defining of an accurate candidate network function at a level such that it can be simulated extensively and with enough details such that the simulation results are realistic and meaningful. The simulator is then gradually developed to introduce more and more parameters that are relevant to the performance of a real–time communication network. The simulation results are finally used to measure the performance of the network which enables us to draw some conclusions that will lead to a future hardware implementation of the router.

In section 2 of this report we summarize the unique requirements of real–time systems and highlight some recent work in this area. The candidate network function is defined in section 3. The simulator is then described in section 4 and the simulation results are presented in section 5. In section 6 some conclusions are drawn.

2. Real–time Systems

Parallel processing is imperative for many real–time applications to achieve meaningful results in a timely manner. Unlike general–purpose parallel computers, real–time systems must meet stringent timing *and* dependability requirements because the consequences of a disruption of service caused by inadequate response time or physical failure can be catastrophic. An avionics system, the real–time system considered in this research, requires the following criteria among others.

- Predictability.
- Fast response time at least for tasks with higher priorities.
- Stability, even during overloading, to ensure that deadlines of critical tasks are met.
- High degree of schedulability.
- Dependability.

A great deal of work has been done by several researchers in real–time scheduling as summarized by Tilborg and Koob[1]. Rate Monotonic Scheduling (RMS) theory has emerged as a viable tool that enables the accomplishment of many of the stringent real–time systems requirements by managing system concurrency and timing constraints at tasking and message passing levels. In a nutshell, RMS theory provides a firm analytic basis for the development of real–time systems by ensuring that as long as the system utilization of all tasks is below a certain bound and appropriate scheduling algorithms are used, all tasks meet their deadlines. This theory has found a widespread use including the McDonnel Douglas Space Station on–board software [2] and IEEE Futurebus+ [3]. RMS theory calls for assignment of higher priorities to tasks with shorter request periods, independent of their run times. A set of *m independent* period-

ic tasks $\tau_1, \tau_2, ..., \tau_m$ is completely specified as $S(m) = \{(C_1, T_1), (C_2, T_2), ..., (C_m, T_m)\}$ where C_i and T_i denote the run time and the period of task τ_i respectively. Liu and Layland [4] show that the elements of the set S(m) scheduled by the RMS algorithm will always meet their deadlines for all task start times, if

$$\sum_{i=1}^{m} (C_i/T_i) \leq m(2^{1/m} - 1).$$

This inequality represents a theoretical bound and is rather conservative. In fact, one can find examples to show that the condition is sufficient but not necessary. In [5] a method is given to find if the idle processor time can be used to enlarge a given C_i, for instance to improve the accuracy of τ_i. To determine if tasks scheduled on a resource with utilization greater than the bound in (1) can meet their deadlines, one can use the exact schedulability test based on the following result [6]: for the set S(m) if τ_i meets its *first* deadline $D_i \leq T_i$, when all the higher priority tasks are started at the same time, then it meets all its future deadlines with any other task start times.

Priority inversion occurs when a high priority task is prevented from executing by a low priority task. Unbounded priority inversion is obviously catastrophic. Even when priority inversion is bounded we must ensure that it is minimized as much as possible. This problem can be controlled by the priority ceiling protocol [7] which satisfies the following two properties

- it prevents mutual deadlock between tasks, and
- under this protocol a task can be blocked by lower priority tasks at most once.

The priority ceiling of a binary semaphore S is defined to be the highest priority of all tasks that may lock S. When a task τ tries to execute one of its critical sections, it will be suspended unless its priority is higher than the priority ceilings of all semaphores currently locked by tasks other than τ. If task τ is unable to enter its critical section for this reason, the task that holds the lock on the semaphore with the highest priority ceiling is said to be blocking τ and hence inherits the priority of τ. As long as a task τ is not attempting to enter one of its critical sections, it will preempt every task that has a lower priority. In a simplified implementation of the priority ceiling protocol called the priority ceiling emulation [8], once a task locks a semaphore, its priority is immediately raised to the level of the priority ceiling of the semaphore. Deadlock avoidance and block–at–most–once results still hold provided one restricts tasks from suspending execution within their critical sections.

Dependability is achieved mainly through the placement and management of redundancy in various forms. The key to the success of using parallel systems with point–to–point interconnection networks for embedded real–time applications is the timely execution of several tasks which usually reside on different nodes and communicate with each other to accomplish a common goal. Deadline guarantees for these real–time tasks are not possible without a carefully designed communication network that supports the timely delivery of messages. The backbone of such a communication network is the router. This research deals with the design and evaluation of such routers and the associated network characteristics.

3. The Candidate Network Function

In this section we discuss the characteristics of the network as well as the communication paradigms and definitions. We closely follow the terminology outlined by Reed and Fujimoto [9].

Topology

Packaging constraints in real–time avionics make backplane and inter–box connections expensive. These parallel computers are often fielded with a number of empty slots for future expansion and thus the topology should be such that one can plug in additional processor boards without modifying the existing underlying hardware. The number of processors should range from a modest (say tens) to massive (say hundreds). The communication patterns of avionics applications to be run on these computers are mostly structured and uniform, and do not demand a high degree of connectivity or extremely low diameter from the parallel computer. Keeping these requirements in mind we have chosen the two–dimensional (2D) mesh without wraparound (henceforth referred to as mesh) topology. The dimension is low enough to have a high wire efficiency. It has been shown that the 2D mesh can have wider channels and faster communication rates than hypercubes, resulting in a higher bandwidth between nodes [10, 11]. Dally has shown that for uniformly distributed communication, 2D meshes will perform better than hypercubes even if we consider contention [11]. The wraparound links are omitted, at the cost of introducing asymmetry, to cater to the need of expandibility with minimal redesigning cost. Some of the other appealing features of the chosen topology include high bisection bandwidth, simple addressing scheme for shortest paths, bounded node degree, and simple layout.

Message

A message is a unit of information that an entity called the source node submits to the network to be transferred to a set of entities called destination nodes located elsewhere in the system. The set of destination nodes may range from simply one node to all nodes (known as broadcast). The message is partitioned into subunits called flits (flow control units). There will be the following three types of flits

- Header flit
- Body flit
- Tail flit.

Distinguishing between physical flits and logical flits may reduce the overall message latency, but will definitely increase the complexity of the router hardware. We feel that the cost of making this distinction outweighs the advantage and thus will not be made. The message generally has one header flit, followed by one or more body flits, and a tail flit at the end.

Switching Technique

Two switching techniques commonly used are *packet switching* [12] and *wormhole switching* [13]. Packet switching is a store–and–forward scheme where the entire routing information is embedded in every packet and each packet is self–sufficient. But the drawbacks of this method are the enormous buffering required in every node as well as the fact that the routing overhead is incurred by the entire packet. Wormhole switching, the one that is chosen eventually, is similar to the traditional circuit switching scheme. The message has multiple flits but the routing overhead is incurred solely by the header flit. The header flit establishes the entire route which the remaining flits simply follow. A relatively small amount of buffering is required per node to store a few of the flits and the method allows messages to be of any length. Once a flit is blocked, because either the outgoing link is busy or there is no space in the destination node, the preceding flits can still travel only if they are not blocked themselves. So once the header is blocked, eventually the remainder of the message is stopped, distributed over multiple nodes, holding the buffers along the already established path. The drawback of wormhole routing is that the corresponding resources (called virtual channels) in the route of the message are not relinquished even if there is blocking which apart from denying resources from other messages may also lead to deadlock. Hence we use deadlock–free algorithms [14] by multiplexing multiple virtual channels on the physical link.

Wormhole switching is ideally suited for hard deadline real–time systems. Given that the network has a large number of processors with a maximum degree of connectivity four, messages may have to go through numerous hops before they reach their destinations. Wormhole switching will minimize the message latency primarily because the routing overhead is limited to the header and the remaining flits will simply cut through as soon as resources are available along the established route. Wormhole switching can be viewed as the pipelining of the message transmission over a number of channels along the route. Packet switching method will have to inevitably break longer messages into smaller ones which will not only increase message latency but make the interfacing hardware complicated to enable reconstruction of the message at the destination.

Routing Technique

Routing decisions can be *explicit* or *implicit*. In an explicit scheme, the source node will compute the entire route *a priori* and the router will simply switch accordingly. This scheme is sometimes attractive in real–time systems for debugging and predictability reasons. But since most users will prefer to shift the burden of routing to the router we employ the implicit scheme whereby the router will determine the route based on local state information and minimal information provided by the source node.

The routing technique can be adaptive or fixed. Adaptive routing usually uses more than simply the destination node number. To alleviate network congestion an adaptive scheme may use additional information like resource conflicts, presence of other flits and alternative routes, thus adapting to current network conditions. So adaptive schemes dynamically choose paths based on a wide variety of information. But this is accomplished at the cost of increased hardware complexity as well as more potential for deadlock. Adaptive routing has faced sharp criticism [15] while others suggest a careful and judicious consideration of various factors [16]. Fixed routing, also known as dimension–order routing, does not consider network conditions and follows a fixed path depending solely on the destination. This may lead to higher latency but the implementation is very simple. In this project we consider fixed routing only.

Routing can be minimal or non–minimal. Minimal routing, as the name suggests, considers only shortest paths (with minimum number of hops). Non–minimal routing, on the other hand, allows routing around certain links (called

misrouting) mainly to increase performance by avoiding busy links and achieve fault tolerance by avoiding faulty links. But this flexibility comes at the cost of increased hardware complexity and possibility of livelock. For this project we consider minimal routing only.

Virtual Channels

Our network should perform well for both long and short messages representing data blocks and control information respectively. Message blocking has a high probability in our low dimension direct network with wormhole routing. To mitigate the adverse effects of message blocking virtual channels (VCs) become essential. We propose to have a fixed number of virtual channels *dedicated* to every physical channel as opposed to having a centralized pool of virtual channels per node. Some of the design issues regarding the virtual channels are outlined below.

- *Number of VCs*: In previous research the number of VCs is determined to avoid deadlocks in adaptive routing methods and to achieve fault–tolerance. But for our real–time system, the need for VCs is primarily to satisfy the stringent requirement of guaranteed on–time delivery of urgent messages. In other words, VCs are needed to minimize priority inversions as well as deadlock. The number of virtual channels is thus an important design parameter and we will study it in detail with our simulator.
- *Width of a VC*: The width of a VC will be an integral multiple of the flit width. The maximum number of flits that a VC can accommodate is an important parameter and will also be experimented extensively with our simulator. The efficacy of wormhole routing lies in deeper pipelining of message transmission which may be facilitated by wider VCs to some extent, but a very wide VC will essentially reduce the benefits of wormhole routing to a store–and–forward scheme. Also, every flit transmission is linked with an explicit acknowledge signal from the receiver which means that a wide VC may not be useful. The router should not only perform well in the presence of multiple messages waiting for a physical channel, but should also handle a single message without introducing unnecessary delays between consecutive flits. The hardware trade–off of a wider VC will be considered in conjunction with the number of VCs per node since both have the same consequences in terms of increasing physical channel utilization and decreasing message latency at the cost of increased memory resources. It is also important to study the interplay between the memory and bookkeeping logic requirements as one increases the number and width of VCs. Once a basic bookkeeping overhead (of at least two flits) is included, increasing the width of a VC is then a relatively localized growth issue because it will not substantially affect the other parts of the router design. But the same is not true for the number of virtual channels as it affects the crossbar size, amount of VC sate information, arbitration logic, control logic, etc.
- *Flit width*: This is also a parameter of the simulator but will not be varied as widely as the number of VCs or the width of a VC. From the latency point of view it pays to keep the flit width small. But we must realize that a router with all the features of wormhole routing with VCs and message priorities will incur an unavoidable baseline complexity which cannot be reduced by reducing the flit width. Thus the flit width should be somewhere between four and sixteen bytes.

Message Priorities

Perhaps the most important consequence of the unique requirements of real–time systems on the router design is the association of priority levels with the messages. Messages must be treated differently with deliberate *unfairness* to ensure the completion of urgent tasks. Thus to support hard deadline systems the baseline router will use message priorities for all arbitration/allocation decisions wherever applicable.

- *Granularity of priority levels*: Most often, the priority mechanisms implementing RMS offer a limited number of priority levels. For instance, the number of priority classes is restricted to eight in the 802.5 Token–Ring standard. If the number of different periods in the system is bigger than the number of levels, the tasks must be partitioned in priority classes and therefore, no matter how different their periods may be, tasks belonging to the same class are not distinguishable. Ideally, there should be as many priority levels as the scheduling algorithm requires. But that is not feasible. When we have a smaller number of priority levels available compared to the number needed by the scheduling algorithm, the schedulability of a resource decreases. In such a case, employing a constant–ratio priority grid for priority assignments reduces this loss. For RMS, the percentage loss in worst–case schedulability due to imperfect priority representation is given by [17]:

$$\text{Loss} = 1 - [\ln(2/r) + 1 - 1/r]/\ln 2 \quad \text{when } r < 2$$

$$\text{Loss} = 1 - 1/(r \ln 2) \quad \text{when } r \geq 2$$

where r is the grid ratio and *ln* is the natural logarithm. Typically given that the shortest and longest periods in a system are 1msec and 100sec (such that $r=(10^5)^{n/c}$, where c is the number of priority levels) then the variation of schedulability loss with the number of priority bits is given by Fig. 1. From this

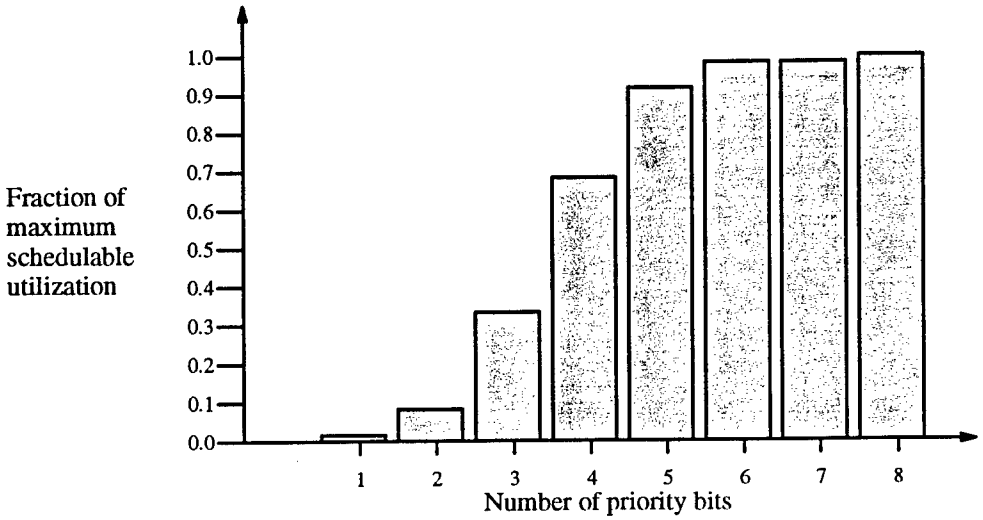


Fig. 1. Schedulability loss versus the number of priority bits

figure one can see that the schedulability loss is tolerably small for six or more priority bits, but it deteriorates appreciably as we decrease the number of priority bits below six. This is an interesting theoretical result derived for a single resource and will be verified in the multiple resources environment by our simulator by making the number of priority levels a parameter. In this regard we will also study the variation of average latency versus priority levels.

- *VC priorities*: The priority of a VC is the priority of the message associated with that VC. The VC retains its priority until the tail flit of the message leaves it. We reiterate that a VC may be empty but still have some priority because of it being reserved by a message whose tail flit has not yet visited that VC. The priority of a VC that has not been visited by any message or relinquished by a message is undefined. This does not cause any ambiguity because every VC has a field that denotes whether it is free or not.
- *VC allocation policy*: The VCs are allocated to physical channels solely based on their priorities. According to the allocation policy, a router output port is allocated based on the VC priorities of **all** messages that are destined for that port. This does not imply that we maintain a centralized pool of virtual channels. Instead, each dedicated group of VCs are queried *simultaneously* and then the global maximum is selected from these local maximums. Ties between VCs with same priorities are broken arbitrarily.

4. The Simulator

A concise, time–driven, flit–based, priority–driven, wormhole–routed, network simulator has been designed in C with an user–friendly GUI to experiment with the numerous design issues. The simulator is based on the simple packet–switched simulator developed by M. Pertel at Cal Tech [18] which provides us with the starting point.

Data Structures

The three basic data structures are described in detail.

1. *Flit* – This is the fundamental unit of communication. The structure, reproduced from the simulator code, is as follows:

```
struct flit
{
int  id;        /*unique message identifier,for debugging purposes only*/
int  flittype; /*0=header, 1=body, 2=tail*/
```

```
int  dest;     /*destination address in decimal, header only*/
int  tsent;    /*time of injection*/
int  tin;      /*time a flit arrives in a virtual channel of a node*/
int  nhops;    /*number of hops a tail takes to reach destination*/
int  priority; /*message priority, in header only*/
int  len;      /*message length in flits, for statistical purposes only*/
flit *next;    /*for linked list purpose*/
};
```

2. *Fifoq* – A First In First Out (FIFO) buffer that denotes a virtual channel. The structure, reproduced from the simulator code, is as follows:

```
struct fifoq
{
flit *head;    /*pointer to head of FIFO which is enqueued*/
flit *tail;    /*pointer to tail of FIFO where flit is dequeued*/
int  len;      /*width of a VC*/
int  id;       /*-1 if VC is free, else equal to message id*/
int  priority; /*-1 if VC is free, else equal to message priority*/
int  nextnode; /*which node to go to, for non-headers*/
int  iport;    /*which port in nextnode to go to, for non-headers*/
int  oport;    /*which port in this node to take, for non-headers*/
int  vc;       /*vc number at nextnode, for non-headers*/
};
```

3. *Nodestate* – This data structure implements a whole node which consists of the five ports in addition to some timing information. The structure, reproduced from the simulator code, is as follows:

```
struct nodestate
{
fifoq *ipqueue[IPL]; /*collection of injection VCs as a priority queue*/
fifoq *pqueue[NIN][Q];/*NIN-1 priority queues of VCs, one for each port*/
int  len[NIN];       /*total number of VCs so far including free ones*/
int  len_valid[NIN]; /*total number of VCs so far excluding free ones*/
int  tnh[NIN][Q];    /*time-next-head for each VC, to avoid input overlap*/
int  itnh[IPL];      /*same as tnh but for injection port only*/
int  tfree[NOUT];    /*when output port is free, to avoid output overlap*/
int  tls;            /*time-last-send, to prevent injection overlap*/
};
```

Theory of Operation

The time–driven approach is chosen over the commonly used discrete–event driven technique because of its simplicity, speed of execution, and ease of understanding. Fig. 2 shows the block diagram of the basic router architecture as implemented by the simulator. The five ports marked 0, 1, 2, 3, and 4 in the diagram correspond to the host, east (or *X–predecessor*), west (or *X–successor*), south (or *Y–predecessor*), and north (or *Y–successor*) respectively. Each physical link is unidirectional and is assumed to be 8–bits wide. Only the datapaths are shown in Fig. 2; the control signals are omitted from the figure to avoid cluttering.

There is a collection of virtual channels organized in the form of prioritized queues (PQs) corresponding to each port. Each virtual channel is organized as a FIFO queue made up of an integral number of flits. The host port is different from the rest (the NEWS) of the ports. The host PQ essentially represents the injection channels (the consumption channels are not explicitly created in the simulator) and so its depth *IPL* is much larger compared to the depth *Q* of the other PQs. The total number of FIFO buffers is $IPL + 4*Q$ per node. The width of an injection channel *FIFOL_IQ* must be at least equal to the maximum message length *MSGL*, whereas the width of all the other virtual channels is *FIFOL* which is generally smaller than *FIFOL_IQ* and is an important parameter of the simulator.

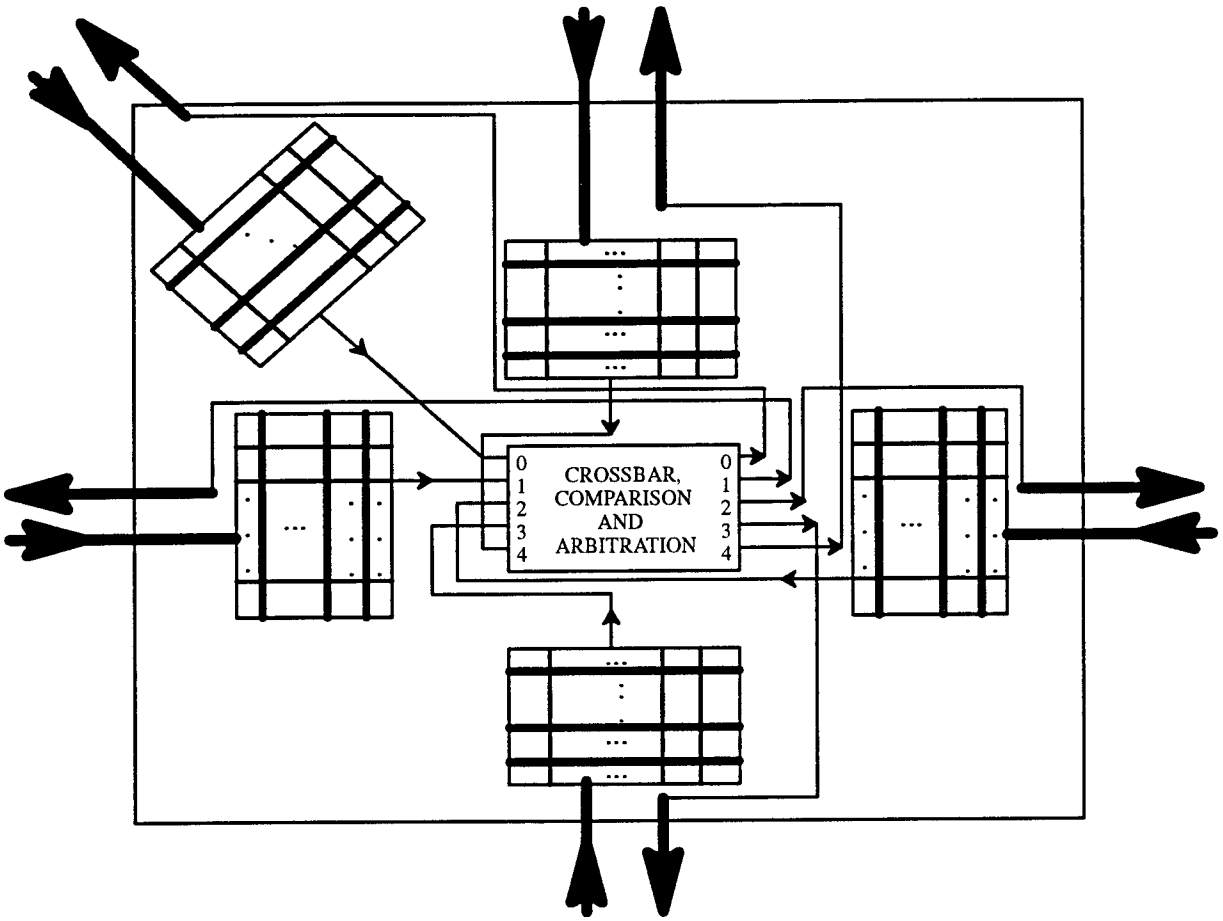


Fig. 2. Block diagram of router configuration in simulator.

For each cycle of simulated time, every node in the network is simulated. Messages of arbitrary length (maximum message length is *MSGL*) in terms of flits are created and the simulator deals with individual flits. As the header flit moves through the network, it establishes the route by reserving a virtual channel in every node in its path. The non–header flits of the message simply follows the established path by occupying the appropriate VC in the nodes. A VC is relinquished by a message only when the tail flit leaves that channel. As a flit traverses the network, it may be delayed for one or more of the following reasons:

- The flit is not at the head of the FIFO queue.
- No VC is free.
- Appropriate VC has been reserved, but it is full.
- Output link is busy.
- Loses out to a flit with a higher priority.
- Loses out to a flit with a lower priority to avoid deadlock but leads to priority inversion.

During each cycle of simulated time, each node injects a message with probability (4*A)/(R*L*AVL) where A = applied load, R = network radix, L = flit length in bytes, and AVL = average message length in flits. This is derived as follows. Let N = total number of nodes in the network, which implies that the bisection width of the network is N/R. The destination addresses are generated randomly with a uniform distribution. At any time the number of flits that may move is N*L*AVL*p where p = probability of injection at every node. Now consider the network in two halves, each with (N*L*AVL*p)/2 flits; half of which moves from one half of the network to the other half. In other words (N*L*AVL*p)/4 flits move from one half to the other. But the bisection width is N/R. Thus the load per link is A = [(N*L*AVL*p)/4]/(N/R) which yields p to be (4*A)/(R*L*AVL).

Overlap Prevention

One must carefully avoid different overlaps that are possible in a time–driven simulator. The three types of overlaps and their prevention mechanisms are:

- *Injection Overlap* – This can happen if in a given node, a flit is injected before the last byte of the previously injected flit is accounted for. This is prevented by computing, for that node, $\texttt{tls}_{new}$ = max(`curtime`, $\texttt{tls}_{old} + L$) where `curtime` is the current simulated time. Then the `tin` field of the flit is set equal to $\texttt{tls}_{new}$.
- *Input Overlap* – This happens when a flit is removed from the head of a VC at time T and then the next flit reaches the head before T+*L*. To prevent this we have the `tnh` field for every VC. A flit is not allowed to move at time `curtime` if `curtime` < `tnh`.
- *Output Overlap* – This occurs when a flit is forwarded on an output link at time T and another at time less than T+*L*. To avoid this we have the `tfree` field for every output link, and if a flit is forwarded at time T, then $T \geq \texttt{tfree}_{old}$ and $\texttt{tfree}_{new} = T+L$.

Deadlock Prevention

Deadlock prevention is critical for any router design. Without prioritizing messages, if the flits are required to traverse the dimensions in a fixed order then the *consumption assumption* is sufficient deadlock freedom. In other words, in the conventional networks if a node always consumes a flit that arrives for it then no flit can be blocked at its destination. The proof of this result is inductive and trivial. But the situation is different in our case in the presence of prioritized queues. In essence, to prevent deadlock we force consumption at the cost of priority inversion. Instead of forwarding the flit with the highest priority among *all* flits destined for a particular link, we send out the flit with the highest priority among those flits destined for that port *and* acceptable to receiving node (i.e., receiving node has space for that flit).

Convergence

Our simulator contains a simple test for the convergence of the average latency measurement, and it stops when convergence is detected. The throughput is monitored to verify that it converges to the applied load. In all our simulation results the throughput matches the applied load before the latency converges. Thus the throughput results are omitted from the convergence test but are recorded for verification purposes. Also, the difference between the number of messages sent and the number received is monitored and we find that it is a small fraction of the total number of messages. As a precautionary measure, once the latency converges we stop injection of messages and recording of statistics but let the simulator run till all the leftover messages reach their destinations. This rules out the possibility of deadlock in those particular cases. Our simple check for convergence provides sequential snapshots of performance so that the user can monitor convergence, as well as a more flexible criterion for termination than merely specifying a simulation time interval or number of messages to be simulated. The user specifies an initial simulation *INTERVAL* and a convergence tolerance *TOL*. The simulator runs until the end of a time interval, and then the interval is doubled. If the change in combined average latency measurement is less than *TOL* when the interval is doubled, then the simulator stops.

GUI

To make the simulator user–friendly we have included a GUI based on Motif. Fig. 3 shows a sample display. All parameters can be set from the main window using scale panels that let the user see the values of all the parameters at the same time. For some other options, like convergence criteria and output selection, submenus are created with radio buttons which ensures that only one option is selected at a time from any submenu. For best view the simulator should be run in the foreground. The GUI is easily expandable to accommodate future modifications.

In this section we present some of the results obtained from experimentation with our simulator. In most cases we plot the average latency per priority level. As shown in Fig. 4, the actual latency values do not monotonically decrease with the increase in priority levels. Instead, there are some *spikes* in the plot which are mainly attributed to pseudo random nature of the destination addresses, priority level, and message length in the simulator, and to priority inversion to some extent. For comparison purposes, as shown in Fig. 4, we concentrate on the fifth order best fit curve of the latency values which is sufficient to convey the relevant information. We show in Fig. 5 that the plot become smooth when the load factor is reasonably high because a large number of messages are simulated before the simulator converges under those circumstances and *initialization transients* are not predominant.

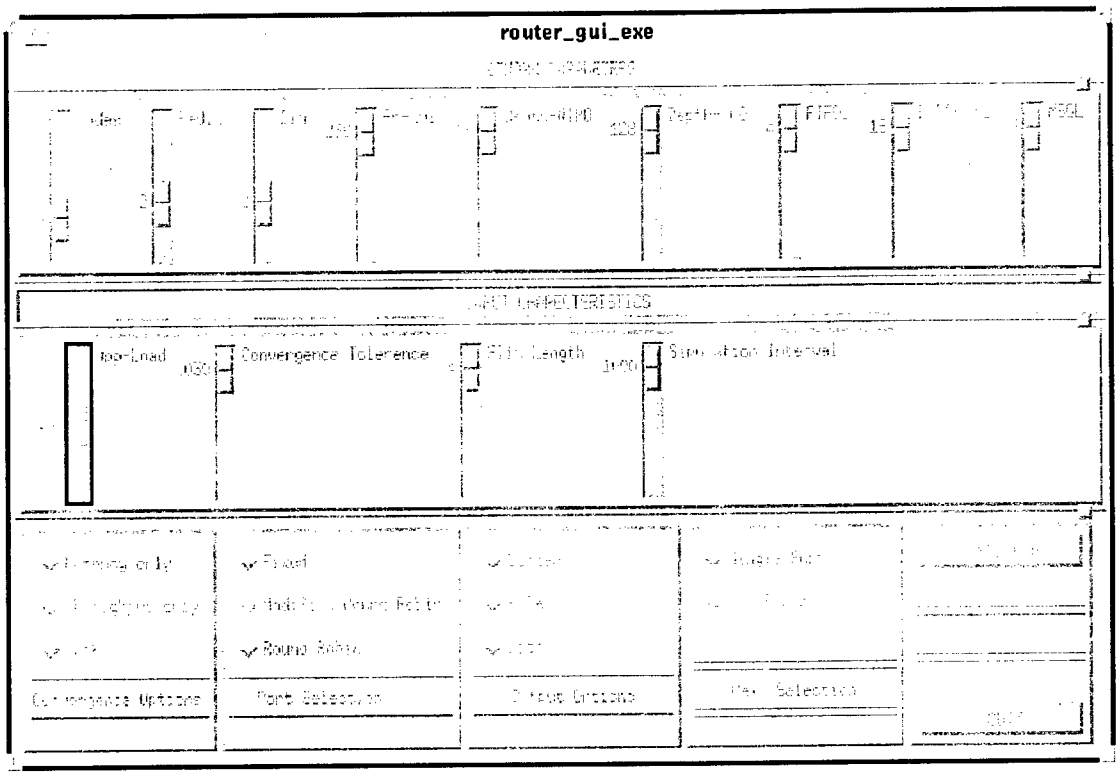


Fig. 3. A sample window display of the GUI of the simulation package.

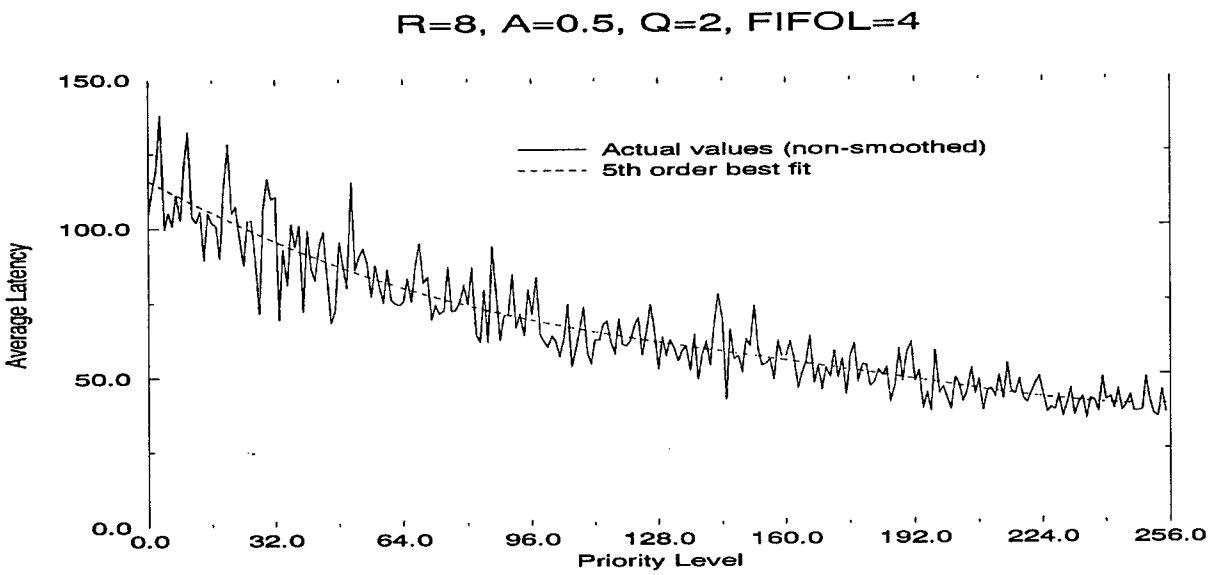


Fig. 4. The actual and best fitted plots for a typical case.

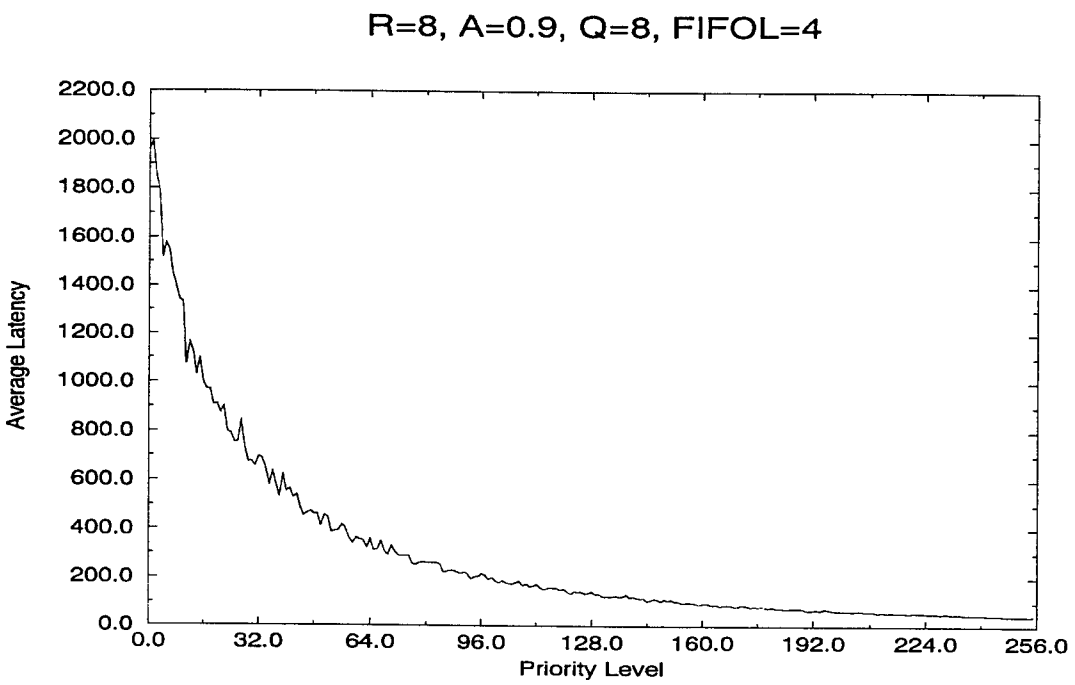


Fig. 5. Actual latency values under heavy load.

The effect of changing the number of VCs is shown in Fig. 6. We notice that the change is negligible beyond four virtual channels. More VCs favor messages with higher priorities. There is a crossover point around priority level 100 below which the latency is less with two VCs than four. To analyze the overall effect we plot, in Fig. 7, the combined average latency (for all messages) versus the number of VCs which shows that there is a sharp decline in latency form two to four but beyond the decrease in latency is rather small.

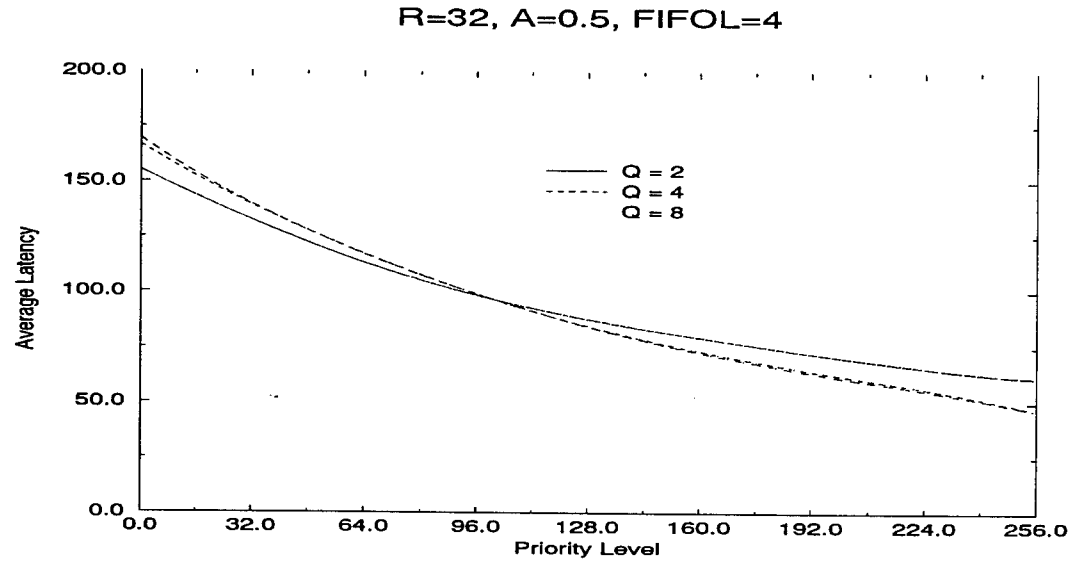


Fig. 6. Effect of number of virtual channels on average latency for an 1024 node mesh.

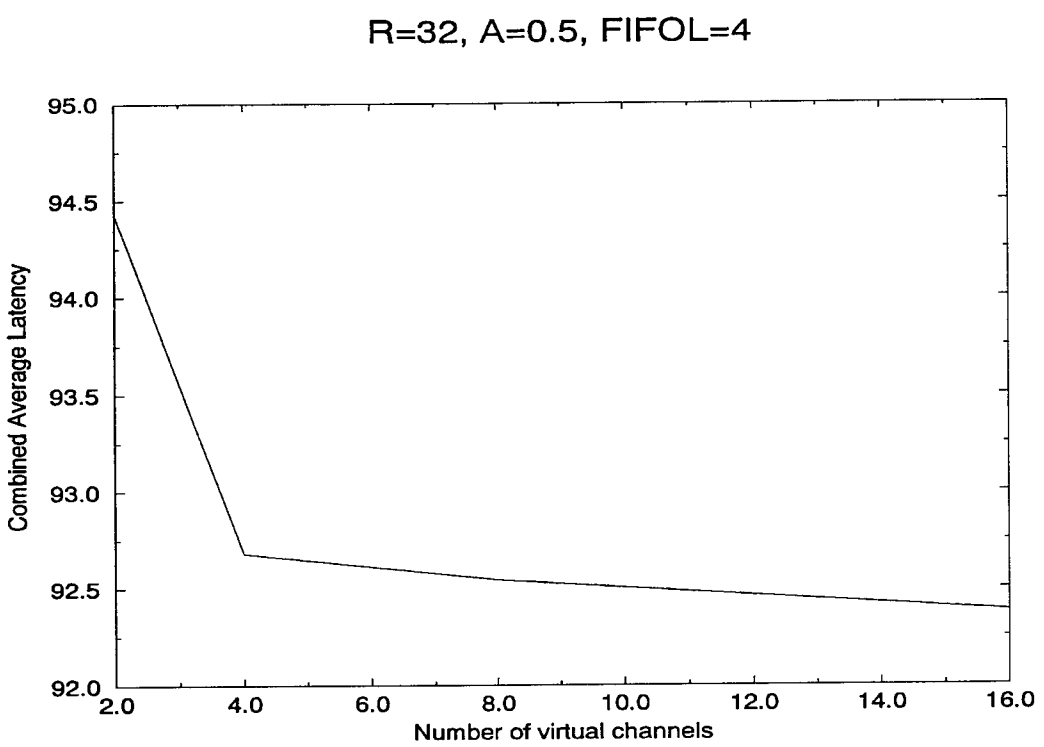


Fig. 7. Effect of number of virtual channels on combined average latency for an 1024 node mesh.

The next parameter varied is the width of VCs. For an 64–node mesh with eight VCs and an applied load factor of 0.5 we plot the variation of average latency with the 256 priority levels for different VC widths in Fig. 8. We find that for critical messages with priority levels beyond 160 there is no measurable difference. But for low priority messages the average latency is greatly improved by increasing the VC width from 2 to 3 flits and beyond. To analyze the overall effect we plot, in Fig. 9, the combined average latency versus the VC widths which shows that there is a steep decrease from 2 to 3 flits, a moderate decrease from 3 to 7 flits, and the decrease saturates.

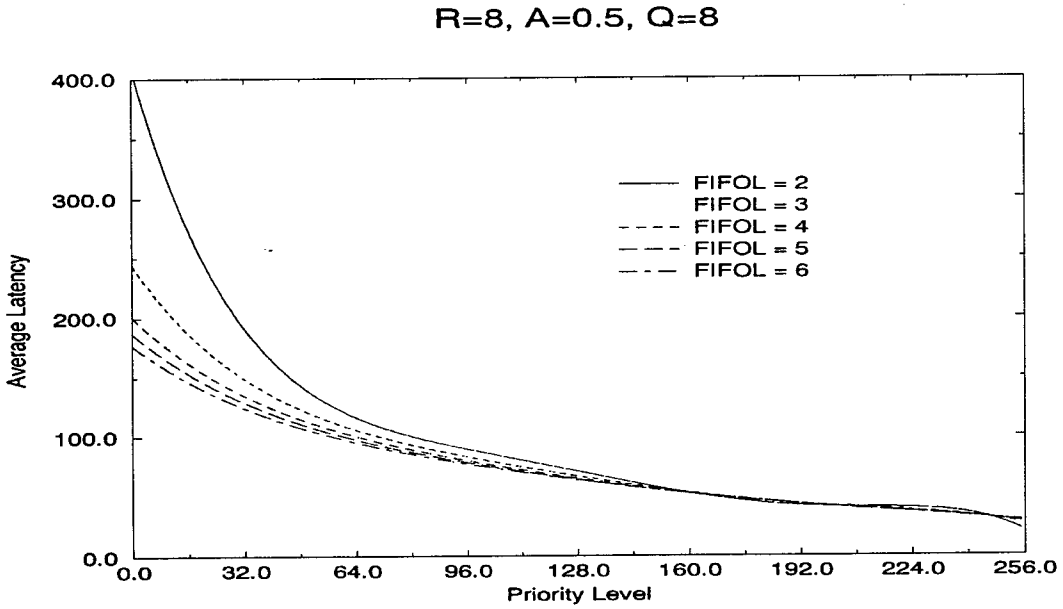


Fig. 8. Variation of average latency with priority levels for different virtual channel widths.

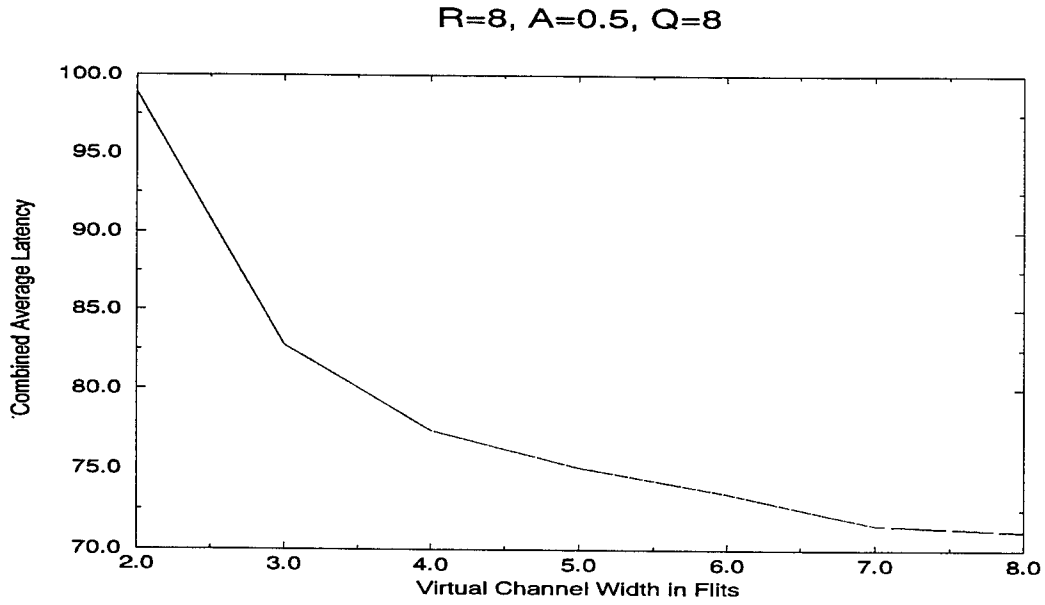


Fig. 9. Variation of combined average latency variation with virtual channel width

So far the applied load in the network was kept constant. The applied load factor A dictates the rate of message injection and when the network stabilizes the throughput or the link utilization matches A. We vary A from 0.2 to 0.9 in an 64–node mesh with 8 VCs where each VC is 4–flit wide. The results are shown in Fig. 10. For any given load factor the average latency decreases with the increase in priority levels. But the decrease is very sharp when the network is heavily loaded. In fact, as shown in Fig. 11, the combined average latency increases very rapidly as the load is increased beyond 0.7, below which the increase is modest. We do not show A=0.1 because the under such light load the network converges very fast with very little meaningful statistical data. Similarly, the case A=1.0 is omitted because the simulator failed to converge in this case even after three days!

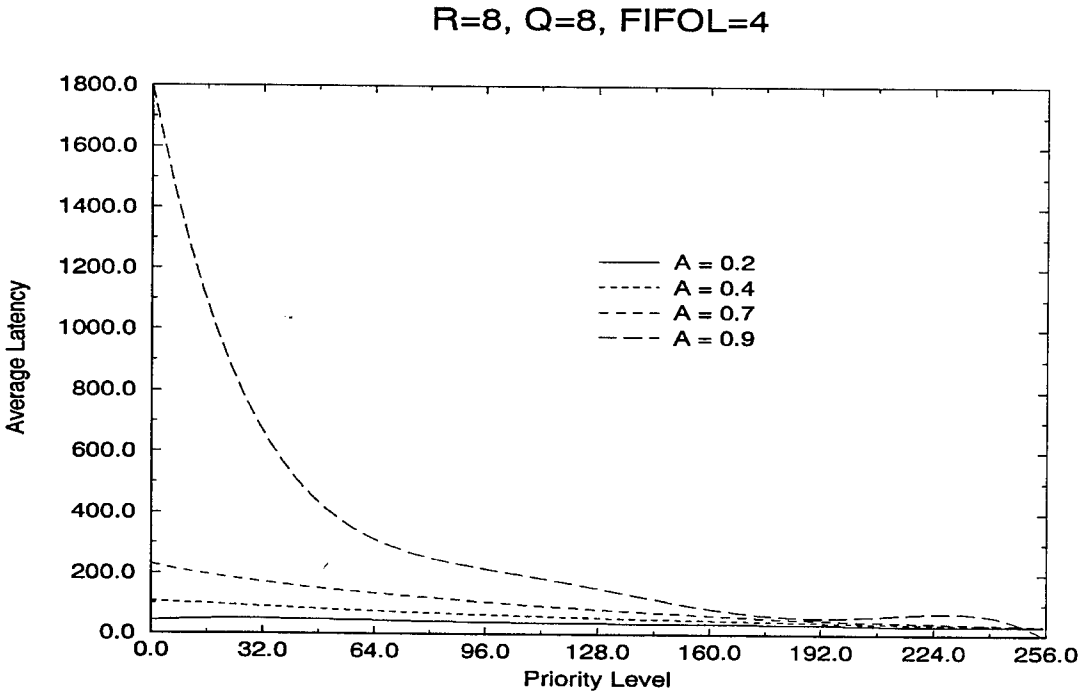


Fig. 10. Variation of average latency for different priority levels with load factor.

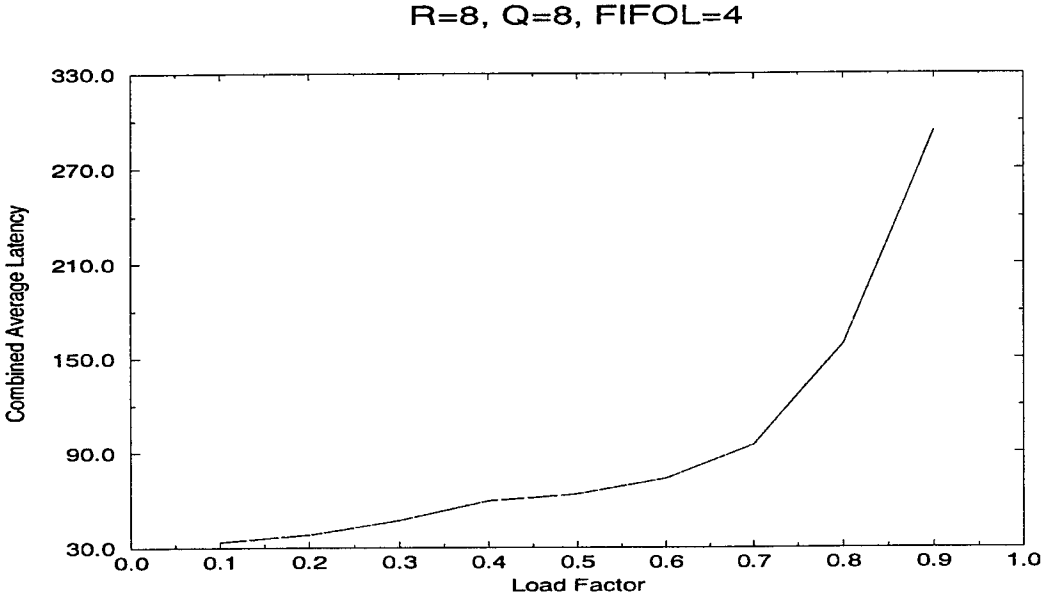


Fig. 11. Variation of combined average latency with load factor.

As mentioned in section 3, the target network size varies from tens to hundreds of nodes. So we analyzed the network performance for different sizes and the results are shown in Figs. 12 and 13, where the applied load factor, the number of VCs, and VC width are kept constant at 0.5, 4, and 4 respectively. Recall that the size of the network is RxR. In larger networks the average latency is more due to the fact that messages travel longer distances on average. Fig. 12 shows that even in larger networks, the messages with high priorities have lower latencies than messages with lower priorities. The case of R=4 is not monotonic because for such a small network the load factor of 0.5 is too small to stabilize the network. This anomaly can also be seen from Fig. 13, where the combined average latency curve has two slopes – one from radix 4 to 8, and the other from radix 8 and beyond.

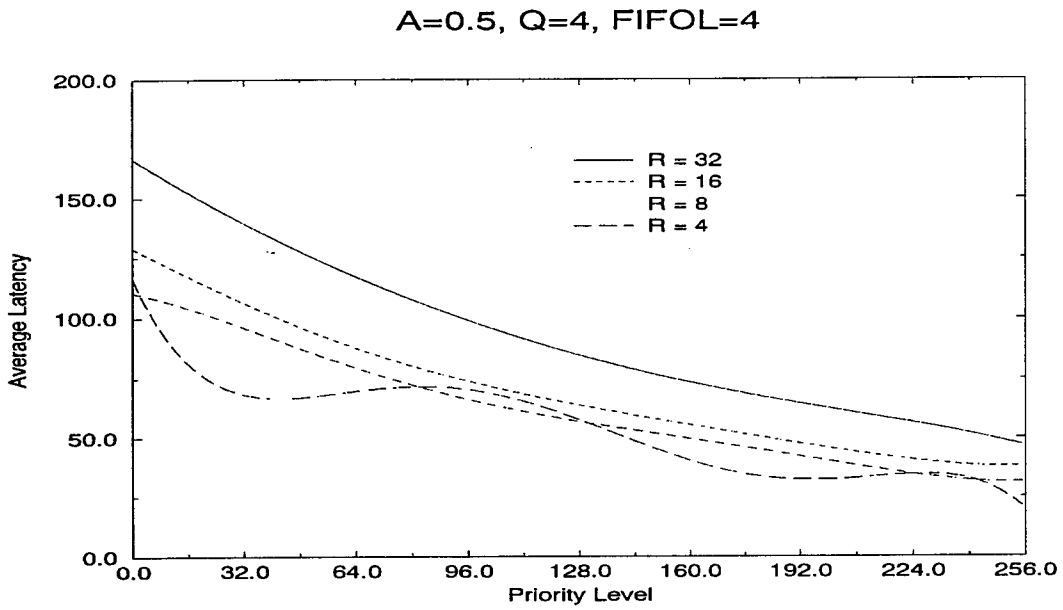


Fig. 12. Variation of average latency with network size.

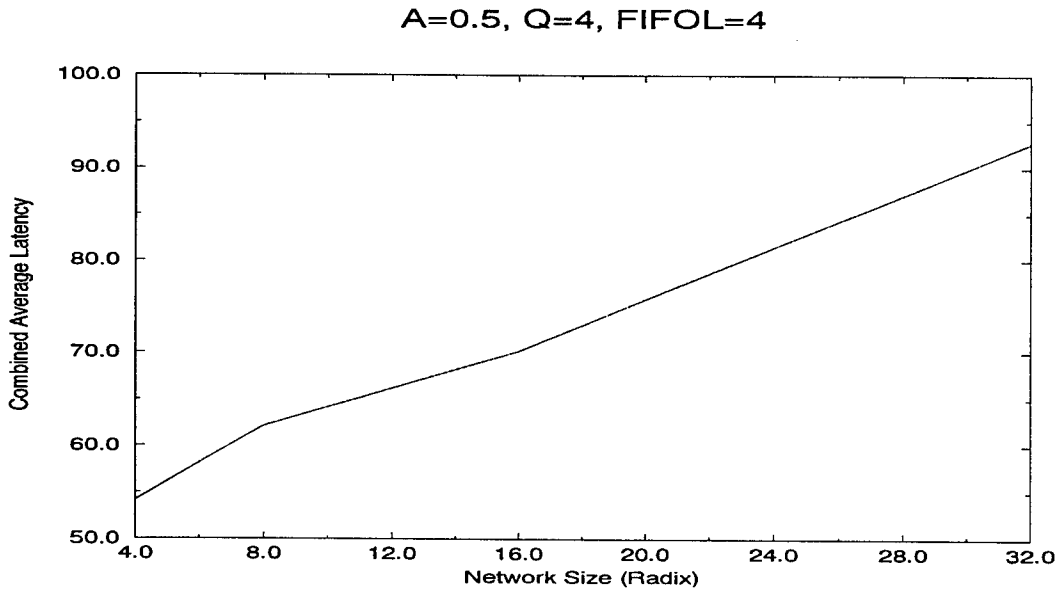


Fig. 13. Variation of combined average latency with network size.

Currently we do not have sufficient resources in the simulator to measure the effect of a change in the granularity of priority levels. The reason is that the priority is generated randomly with a uniform distribution independent of the message period and length. Under these circumstances changing the number of priority levels does not yield a lot of information. We experimented with 32, 64, 128 and 256 priority levels. We noticed that the behavior of the average latency is very similar for different priority levels except that the curve is smoother with a lower number of priority levels. No plots are shown for this case as the information obtained is not substantial and due to space shortage.

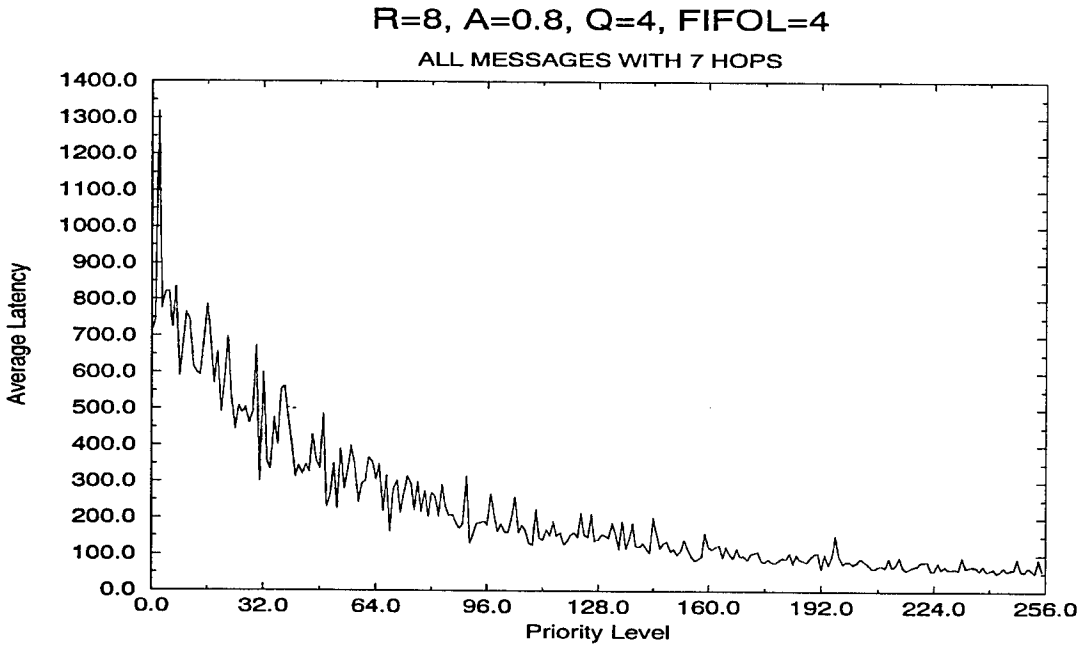


Fig. 14. Variation in average latency for messages that traversed 7 hops only.

So far we have not distinguished between messages within the same priority level or class. To get more insight into messages within a priority level we first restrict our attention to messages that traversed a fixed number of hops. Fig. 14 shows such a case for a 64–node mesh where we look at messages that traversed 7 hops from source to destination. We find that the latency not only decreases with increasing priority levels but the spikes in the plot are very few

indicating a low degree of priority inversion. We then examine messages that traversed 7 hops but keep the entire network load to 5–flit long messages only. For this case the results are shown in Fig. 15.

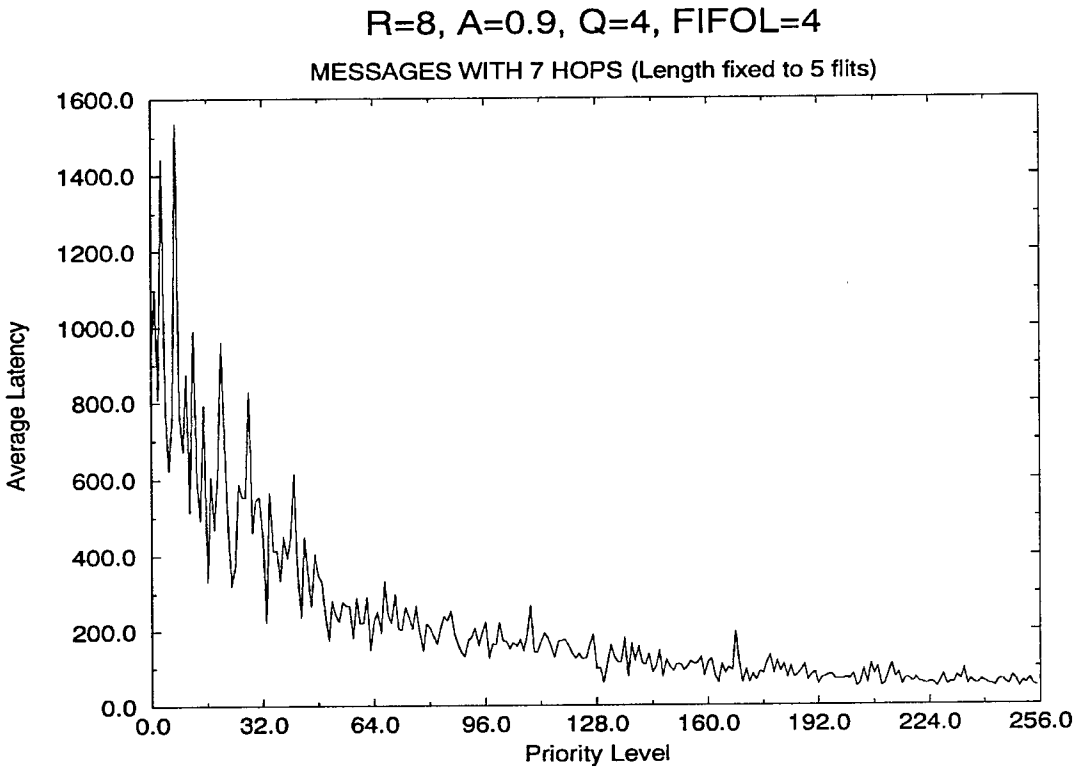


Fig. 15. Variation in average latency for messages with 7 hops with load restricted to only 5–flit long messages.

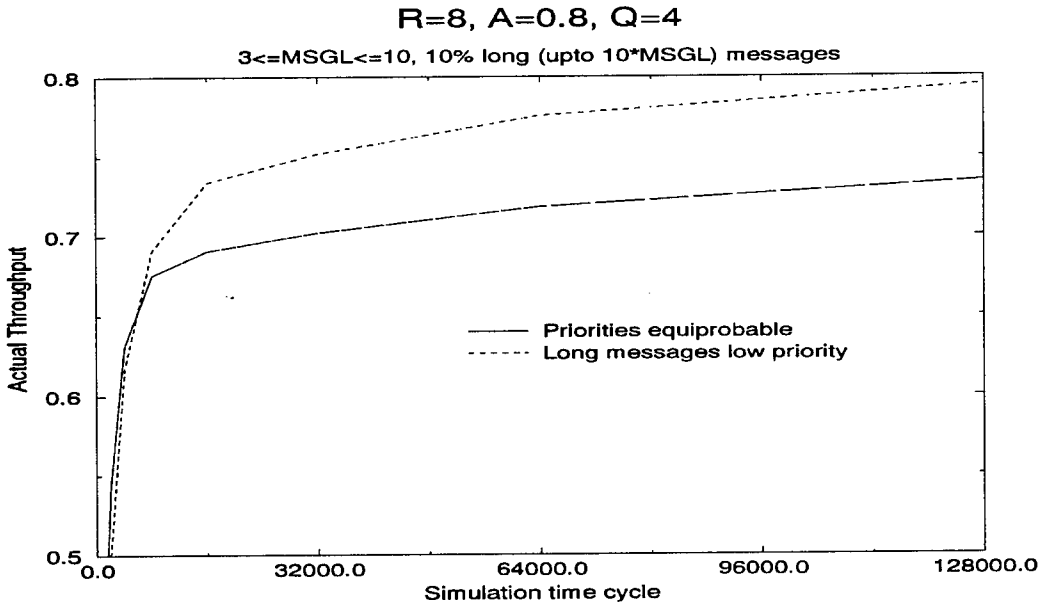


Fig. 16. Measured throughput variation for a particular non–uniform distribution.

We now introduce various forms of non–uniformities in the message distributions and analyze the effects. First we make 10% of all messages longer than usual (ten times longer to be precise) but let the priorities still remain equi-

probable. Then we make the priorities of the unusually long messages very low. The results are depicted in Figs. 16 and 17. We find, from Fig. 16, that the throughput approaches the applied load if the priorities are kept low for the long messages. Similarly, Fig. 17 shows that the combined average latency increases rapidly if the priorities of the long messages are not lowered deliberately.

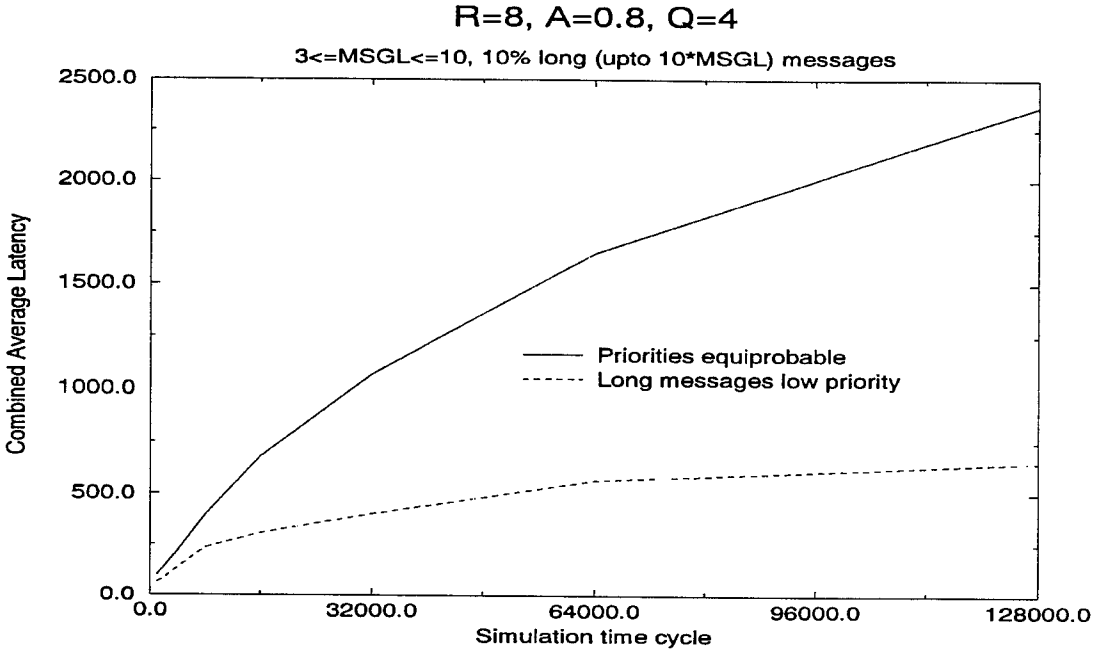


Fig. 17. Combined average latency variation for the same non–uniform distribution (as in Fig. 16).

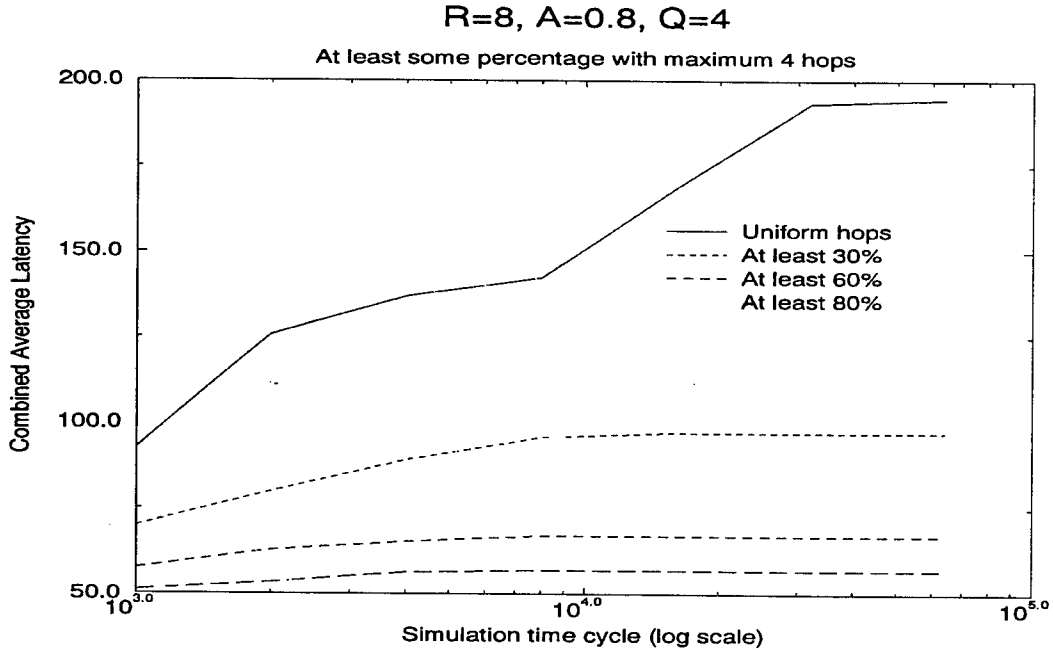


Fig. 18. Variation in combined average latency with simulation time for fixed number of hops (log–linear).

Another form of non–uniform distribution is to bias local communication over global. To test this, we modify our simulator such that destination addresses are generated within 4 hops for at least a certain percentage of the messages in an 8x8 mesh. The results are shown in Fig. 18. We see that when the destination addresses are chosen with

the random uniform distribution not only is the latency value large but it also takes much longer to converge. On the other hand, as we increase the number of messages with at most 4 hops the latency decreases and converges faster.

For an application to run efficiently on a parallel computer, it should be partitioned such that urgent messages do not travel large distances. In other words, the message priorities should be inversely proportional to the number of hops. Such a scenario is simulated on an 8x8 mesh with 256 priority levels. Note that the number of hops range

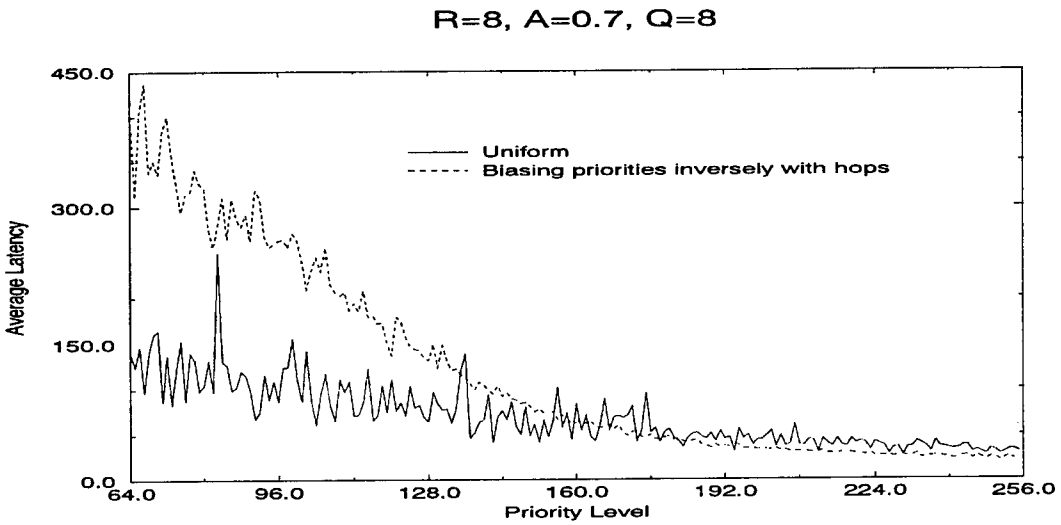


Fig. 19. Comparison of uniform priorities and biasing priorities inversely with number of hops.

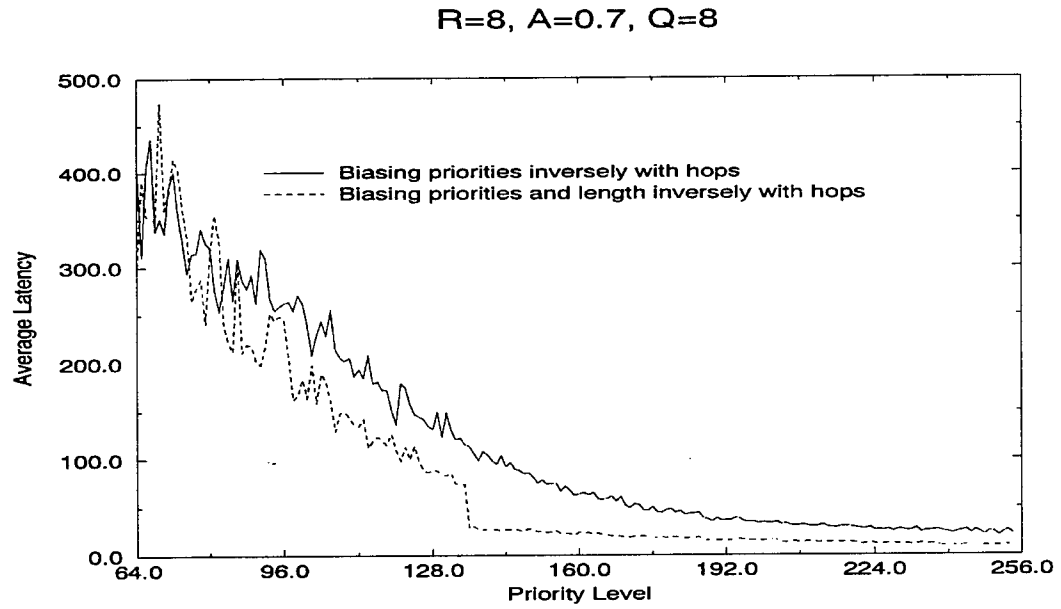


Fig. 20. Comparison of biasing priorities and message lengths inversely with number of hops.

from 0 to 14. Thus we partition the priority range into fifteen equal subranges, where subrange 0 is [0...16], 1 is [17...33], so on. Priorities are assigned according to the number of hops, e.g., the number of hops 14 corresponds to subrange 0, 13 corresponds to subrange 1, so on. The results are shown in Fig. 19. Due to biasing no messages were received for very low priorities and so the graph starts from priority level 64. Note that for high priority messages the non–uniform distribution not only reduces the average latency but also the graph is smoother indicating very little priority inversion. Lower priority messages, on the other hand, suffer as a result of this biasing. The effect is more pronounced if we make the message lengths smaller for messages with smaller hops as shown in Fig. 20.

6. Conclusions

We formulated a network function by carefully considering various characteristics. These network properties cater to the unique demands of real–time systems. The topology is chosen to be a 2D mesh with no wraparound edges. The information is handled in the form of messages or packets. The fundamental unit of communication is a flit. We have three types of flits – header, body and tail. We do not distinguish between logical and physical flits. Wormhole switching is our choice for the switching technique. Wormhole switching is ideally suited for real–time systems. It has lower communication latency and smaller buffer requirements than most other techniques. The routing method is implicit, dimension order or fixed, and minimal. A very important characteristic of the network is a fixed number of virtual channels dedicated to every physical channel. The number of virtual channels, the width of a virtual channel, and the width of a flit are important network parameters. As a consequence of real–time systems all messages are associated with priorities and every virtual channel is associated with the priority of the message it stores. The virtual channels are allocated to the physical channel based on priorities and state information. The arbitration logic queries the dedicated pool of virtual channels simultaneously and then finds the global maximum among the local maximums.

A concise, time–driven, flit–based, priority–driven, wormhole–routed, network simulator has been designed in C with an user–friendly Graphical User Interface. The simulator enables us to analyze the performance cost of the various parameters of the network function. Special care is taken to prevent various possible kinds of overlaps (in time) and deadlock. The simulator contains a simple test for the convergence of the average latency, and it stops when convergence is detected. The throughput is always monitored to verify that it converges to the applied load.

The simulation results are plotted and analyzed in great detail. The simulation sheds light on the network performance based on various parameters like the number of virtual channels, the width of a virtual channel, applied load factor, size of the network, among others. Initially the destination address, message length and message priority are generated randomly with an uniform distribution. Various non–uniformities are put for realistic workload.

References

[1] A. van Tilborg and G. Koob, *Foundations of Real–Time Computing: Scheduling and Resource Management*, Kluwer Publishing, Norwell, MA, 1991.

[2] C. Warren, "Rate monotonic scheduling," *IEEE Micro*, vol. 11, no. 3, pp. 34–38, 102, 1991.

[3] "Futurebus+ Recommended Practice," *IEEE Standard 896.3* (IEEE, 345 East 47th St., NY, NY 10017), 1993.

[4] C.L. Liu and J.W. Layland, "Scheduling algorithms for multiprogramming in a hard real time environment," *J. ACM*, vol. 20, no. 1, pp. 46–61, 1973.

[5] J.P. Lehoczky, L. Sha and Y. Ding , "The rate monotonic scheduling algorithm: Exact characterization and average case behavior," *Proceedings Real–Time Systems Symposium*, IEEE CS, Los Alamitos, CA 1989.

[6] L. Sha, R. Rajkumar and S.S. Sathaye, "Generalized rate–monotonic scheduling theory: A framework for developing real–time systems," *Proceedings of the IEEE*, vol. 82, no. 1, January 1994.

[7] L. Sha, R. Rajkumar and J.P. Lehoczky, "Priority inheritance protocols: An approach to real–time synchronization," *IEEE Trans. Computers*, pp. 1175–1185, September 1990.

[8] L. Sha and J.B. Goodenough, "Real–time scheduling theory and Ada," *IEEE Computer*, pp. 53–62, April 1990.

[9] D.A. Reed and R.M. Fujimoto, "The performance of multicomputer interconnection networks," *IEEE Computer*, vol. 20, no. 6, pp. 63–73, June 1987.

[10] W.J. Dally, "Wire–efficient VLSI multiprocessor communication networks," *Proceedings Stanford Conference on Advanced Research in VLSI*, MIT Press, Cambridge, MA, pp. 391–415, March 1987.

[11] W.J. Dally, "Performance analysis of k–ary n–cube interconnection networks,' *IEEE Trans. Computers*, vol. 39, no. 6, pp. 775–785, June 1990.

[12] D. Bertsekas and R. Gallagher, *Data Networks*, Prentice–Hall Inc, 1987.

[13] W.J. Dally and C.L. Seitz, "The torus routing chip," *J. of Distributed Computing*, vol. 1, no. 3, pp. 187–196, 1986.

[14] W.J. Dally and C.L. Seitz, "Deadlock–free message routing in multiprocessor interconnection networks," *IEEE Transactions on Computers*, vol. C–36, no. 5, pp. 547–553, May 1987.

[15] M.J. Pertel, "A critique of adaptive routing," Caltech Computer Science Technical Report, Caltech–CS–TR–92–06, June 1992.

[16] A.A. Chien, "A cost and speed model for k–ary n–cube wormhole routers," March 1993.

[17] L. Sha, R. Rajkumar and J.P. Lehoczky, "Real–time computing with IEEE Futurebus+," *IEEE Micro*, vol 11, no. 3, pp. 30–33, 95–99, June 1991.

[18] M.J. Pertel, "A simple simulator for multicomputer routing networks," Caltech Computer Science Technical Report, *Caltech–CS–TR–92–04*, March 1992.

TURBINE BLADE FILM COOLING WITH FREE STREAM TURBULENCE

John J. Schauer
Professor
Mechanical and Aerospace Engineering Department

University of Dayton
300 College Park
Dayton OH 45469-0210

Final Report for:
Summer Faculty Research Program
Aero Propulsion and Power Directorate
US Air Force Wright Laboratory

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

and

Aero Propulsion and Power Directorate
US Air Force Wright Laboratory

August 1994

TURBINE BLADE FILM JET COOLING WITH FREE STREAM TURBULENCE

J. J. Schauer
Professor
Mechanical and Aerospace Engineering
University of Dayton

ABSTRACT

Experimental measurements and closed form analytical predictions are presented for the spread and mixing of a film cooling jet injected into a high free stream turbulence main stream. Blowing ratios near one were studied at free stream turbulence levels from 0.9% to 17%. With 17 percent turbulence the spreading rate of the coolant fluid is doubled and the peak temperature difference is reduced to one fourth of the low free stream values by 15 jet diameters downstream. The experimental protion of this work was done primarily by Captain Jeffrey P. Bons of the Aero Propulsion and Power Directorate, Wright Laboratory, WPAFB.

TURBINE BLADE FILM JET COOLING WITH FREE STREAM TURBULENCE

J. J. Schauer

INTRODUCTION

Film cooling applications in modern gas turbine engines are greatly complicated by a harsh aero-thermodynamic environment. To adequately design cooling systems for these applications, it is important to understand the effects of high levels of free stream turbulence, pressure gradients, anisotropy, etc... The inspiration for this experimental and analytical study of the effect of main flow Free Stream Turbulence (FST) on the spreading and mixing of film cooling jets came from the work on film cooling effectiveness done by Bons, et. al. (1994). In this work, the author found that high levels of FST had a dramatic effect on the development of film cooling jets. MacArthur (1986), and more recently Maciejewski and Moffat (1989) and Bogard et. al. (1992) have also reported dramatic effects of high levels of FST on wall heat transfer. To adequately model the development of a film cooling jet injected into a high FST flow, this study utilized the work of Schetz (1980) for jet eddy diffusivity forms and the combined work of Thomas and Hancock (1976) with Hunt and Graham (1977) for the wall effects on the FST.

Experimental data were collected with flow parameters in the range of typical turbine applications. The Reynold's number based on film cooling hole diameter was 20,000 and the blowing ratios were near one (and modeled as identically one in the analytical work). Since the single jet cooling flow spread rapidly, particularly with high FST, the distance down the zero pressure gradient flat plate where the cooling flow produced measurable effects was limited. The data presented corresponds to x/D from 0 to 16 and Reynold's numbers based on distance down the plate of up to 400,000. The density ratio between the cooling flow and the main stream was approximately 0.93.

The analytical modeling utilized the strong assumption that the velocity was uniform everywhere. This clearly is valid only for blowing ratios near one. At blowing ratios fifty percent away from one, however, the velocities were within five percent of the no blowing case by six hole diameters downstream. Integral solutions of the radial constant property energy equation and this equation multiplied by temperature (called the energy concentration equation in this work) provided two equations to predict maximum temperature difference decay and jet spreading rate. As a further refinement of the model, the nearby wall strongly attenuates the fluctuating vertical component of velocity and enhances the fluctuating transverse component of velocity. Incorporating this flow feature into the model permitted it to match experimentally observed asymmetric spreading of the jet.

NOMENCLATURE

D	film cooling hole diameter (1.905cm)
M	blowing ratio ($\rho_{fc}U_{fc}/\rho_{fs}U_{fs}$)
T	static temperature (degrees K)
T*	local temperature difference from the free stream value (deg K)
T_m	difference between maximum temperature on a constant x/D plane and the associated free stream value (deg K)
Tu	turbulence intensity (u'/u) (%)
c	gas specific heat at constant pressure (J/kg degK)
q"	local heat flux (W/m^2)
r	radial distance from center of axisymmetric coolant jet (m)
u	mean local streamwise velocity (m/s)
u'	fluctuating streamwise velocity component (m/s)
v	mean local vertical velocity (m/s)
v'	fluctuating vertical velocity component (m/s)
w	mean local transverse velocity (m/s)
w'	fluctuating transverse velocity component (m/s)
x	streamwise distance from downstream lip of injection hole (m)
y	vertical distance from injection surface (m)
z	spanwise distance from centerline of injection hole (m)
Δ	edge of the film cooling jet temperature profile
Δ_2	midpoint of the film cooling jet temperature profile
ρ	fluid density (kg/m^3)

subscripts:

fc	in the film cooling fluid
fs	in the free stream fluid

EXPERIMENTAL FACILITY

The wind tunnel used for the experiments has a 0.38m (width) by 0.18m (height) cross-section. An elliptical-edge bleed with vacuum suction is located 12.07cm upstream of the downstream lip of the film cooling injection hole (designated as x/D=0). A 1.59mm diameter steel rod located 2.54cm from the elliptical-edge bleed is used to trip the new boundary layer and insure a spanwise uniform turbulent boundary layer profile at the injection point. Without employing turbulence generation devices, the tunnel's free stream turbulence level is 0.9% (±0.05) and velocity uniformity is within ±2.5%. Free stream turbulence generation is accomplished by two methods for the present experimental data. A secondary flow is injected from two opposing rows of holes located on the top and bottom of the wind tunnel 1.02m upstream of the boundary layer

bleed at a velocity ratio (jet to free stream velocity) of 14 to produce a turbulence level of 17% (±0.85) at the film cooling injection station. To provide a turbulence level of 6.5% (±0.3), a standard square grid is installed 0.94m upstream of the coolant injection point. The film cooling is injected through a 1.9cm diameter, 35 degree inclined hole centered in the test section. The injection pipe length from the coolant access plenum to the exit is 3.5 hole diameters.

The data presented in this report were taken using a 4mm diameter tungsten hot wire and a 0.15mm diameter boundary layer split film. The hot wire (or split film) and a flow temperature thermocouple (0.33mm bead diameter) are mounted on a vertical traverse. A velocity map of the injection plane (x-z plane at y = 0) was used to determine the injection jet mean velocity . This average velocity, u, and the local free stream velocity, u_{fs}, are used to calculate the film cooling blowing ratio. The test surface downstream of the film cooling injection point is insulated with 10cm thick urethane foam and is essentially adiabatic. Uncertainty in the velocity measurement, estimated at ±2.0%, is attributed to the calibration fit accuracy and the horizontal displacement between the probe and the flow thermocouple. The uncertainty in the temperature ratio data is ±0.7%.

ANALYTICAL DEVELOPMENT

The constant property radial energy equation in the form

$$u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial r}=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{q^{"}}{\rho c}\right) \qquad (1)$$

is the basis of this modeling. The energy equation is simplified by assuming that the velocity field is relatively uniform and can be approximated by $u = u_{fs}$. The continuity equation then implies that v = 0 throughout the flow field. With these simplifications, the energy equation is multiplied by rdr and integrated from r = 0 to Δ, the edge of the jet. Because the rq" product is zero at both limits, the right hand side of (1) is zero. Defining $T^* = T-T_{fs}$, a dimensionless temperature profile $f(\eta) = T^*/T_m$ (where $\eta = r/\Delta$) is employed which is assumed to be independent of x. T_m here is defined as the maximum temperature in the flow at a given x/D location and is considered to be on the jet centerline (since u is constant the centerline is the center of the temperature distribution). Substituting these non-dimensional quantities into (1) and using Leibnitz's rule for moving the derivative outside the integral, we obtain

$$\frac{d}{dx}\left(\Delta^2 T_m^*\right)=0 \qquad (2)$$

This equation simply states that the total energy in the flow is constant. That is, if the maximum temperature decays then the spread of the jet , Δ, increases to cause the total energy deviation from free stream conditions to be constant. To obtain a second equation in Δ and T_m, we multiply (1) by T, simplify, multiply by rdr, and again integrate from 0 to Δ. The right hand side is integrated by parts to give an energy concentration integral equation of the form

$$\frac{d}{dz}\left(\Delta^2 T_m^*\right)\int_0^1 \eta f^2 d\eta = -\frac{2\Delta T_m^*}{u_{fs}}\int_0^1 \eta \frac{q''}{\rho c} f' d\eta \qquad (3)$$

The heat transfer, q", is modeled with an eddy diffusivity approach. Following Schetz' analysis, we find eddy diffusivity proportional to u' Δ dT*/dr. This turbulent heat transfer comes from two sources. The mixing associated with the film cooling jet is a dominate factor when the FST is small. This mixing is modeled as

$$\frac{q''}{\rho c} = C_{fc} u_{fc}' \Delta \frac{dT^*}{dr} \qquad (4)$$

where u'_{fc} is the maximum streamwise velocity fluctuation in the jet flow. The decay of this turbulence was measured and fit with $u_{fc}' = u_{fc}'(0)\exp(-.11x/D)$. The initial turblence in the jet was proportional to the jet velocity and taken as $u_{fc}'(0)/u_{fs} = 0.17M$. This was the only effect of blowing ratio on the analysis.

The second source of mixing in the eddy diffusivity model is associated with the FST. This second term dominates at high FST levels and is modeled as

$$\frac{q''}{\rho c} = C_{fs} u' \Delta \frac{dT^*}{dr} \qquad (5)$$

The decay of the FST was measured and fit with $u'_{fs} = u'_{fs}(0)(1-.015x/D)$. The mixing heat transfer from equations (4) and (5) were added and substituted into equation (3). Substituting from (2) to eliminate T_m, the differential equations can be solved for Δ giving

$$\Delta = \Delta(0) + \frac{C_{fc} I_1}{0.11 I_2} Tu_{fc}\left(1 - e^{-0.11\frac{x}{D}}\right) + \frac{C_{fs} I_1}{I_2} Tu_{fs}\left[\frac{x}{D} - 0.0075\left(\frac{x}{D}\right)^2\right] \qquad (6)$$

where

$$I_1 = \int_0^{\Delta} \eta f^2 d\eta \quad ; \quad I_2 = \int_0^{\Delta} \eta f'^2 d\eta \quad ; \quad Tu_{fs} = u'/u_{fs} \qquad (7)$$

The modeling was done using $f(\eta) = (1-\eta^2)^2$. For this profile the half velocity point is $\Delta_2 = 0.54\Delta$.

The jet spread depends on the level of mixing specified by the sum of equations (4) and (5). To predict the spread, Δ, using the model, we applied equation (6) to first the vertical spreading on the top of the jet and then to the transverse spreading at the sides of the jet. To predict the vertical spread of the jet, the Tu_{fs} value used in equation (6) was v'/v'_{fs} Experimental data of v'/v'_{fs} decay approaching the wall with no jet injection matched the data of Thomas and Hancock (1976) and the analysis of Hunt and Graham (1977). This v'/v'_{fs} data was fit with $(v'/v'_{fs})^2$ = 1-exp(-3y/L) where L is the integral turbulence scale in the free

stream.. An average y location of the jet centerline was used in equation (6). The value of $v'/v'_{fs} = 0.6$ was used in equation (6) for jet generated FST of 17 percent. The value of $v'/v'_{fs} = 0.8$ was used for grid generated turbulence of 6.5 percent. The different v'/v'_{fs} value used for the grid turbulence is representative of the smaller turbulent length scale of this turbulence compared to the jet generated turbulence.

The references' work show that u' is nearly constant with y and that the turbulent kinetic energy associated with a reduction in v' is compensated with an increase in w'. For the current work, kinetic energy was assumed to be conserved and the free stream turbulence was assumed homogeneous. For the transverse spread, a Tu_{fs} value corresponding to the above criteria for w'/w'_{fs} was used in (6) to find the transverse spread rate. The product of Δ_y and Δ_z are used in equation (2) in place of Δ^2 to solve for T_m, the temperature decay.

RESULTS

A typical velocity profile taken across the film cooling jet exit plane is shown in Fig 1. Rather than the uniform velocity profile that might be assumed, we see that the velocity distribution is shaped like a volcano with a high velocity rim surrounding a lower velocity "crater". The velocity in the "crater" center is 0.6 of the maximum velocity on the rim for the data shown. This data compliments the numerical simulation of Leylek and Zerkle (1993). They compute a high velocity point at the upstream edge of the jet. They also predict a strong counter rotating vortex structure which sweeps this high velocity region around the rim in both directions from the upstream edge. The volcano shaped velocity profiles measured here seem to be consistent with this numerical work.

As the film cooling jet flow goes from the exit plane to the downstream edge of the cooling hole (defined as $x/|D = 0$), the flow is mixed to such an extent that the volcano shape is mostly washed out. The velocity differences quickly mix out, particularly with high FST.

The vertical temperature profiles at various x/D locations are shown in Fig.2. The temperature distribution, initially uniform in the cooling jet, develops bell shaped distributions in both the vertical and transverse directions by the time the flow reaches the downstream edge of the cooling jet hole. The temperature profiles at the downstream lip ($x/D = 0$) are nearly independent of blowing ratio or FST level for the range of blowing ratio from 0.96 to 1.87. (Data not shown in this paper show that at a blowing ratio of 0.6 the initial velocity and temperature profiles are flattened in the vertical direction considerably.) The temperature decay rate is considerably less than the velocity decay rate. This difference is associated with the strong pressure gradient near the downstream lip which effects the velocity field directly. The dominating effects of high FST levels on the decay of the temperature distribution are clearly evident in figure 2.

Fig. 3 shows the spread of the coolant flow in the transverse z direction. Of note is the marked increase in transverse (z) jet spread with increased FST levels, accompanied by the streamwise decay noted previously.

The lobe in the Tu=0.9% profile at x/D of 1.6 is associated with vortex interaction and the volcano shape of Fig. 1 at the exit plane.

Data at three levels of FST and two blowing ratios showing the decay of the maximum temperature deviation from the free stream conditions are shown in Fig. 4. The analysis splits the 0.9 percent FST level data with the spread associated with different blowing ratios being slightly under predicted. The analysis predicts about ten to twenty percent lower values of temperature ratio when compared to the 6.5 percent data. The 17 percent FST data follow the analytic modeling well, with the analysis being about twenty percent high on temperature ratio at the final data station of x/D=15.8.

Fig. 5 and 6 show the spread of the cooling fluid energy. The spread is necessarily larger for the conditions where the decay is greater since the total energy carried into the flow is constant with an insulated plate below the flow. That is, the energy deviation from the free stream datum is constant and as the peak of the concentration decays the energy must be convected further to the sides. The midpoint of the z temperature profile for the jet injected into the Tu=17% free stream increases by a factor of nearly 3 by x/D=15.8, while the low FST data (and model) show only a modest increase. The high FST flows develop from a round, axisymetric distribution in a plane at constant x to an elliptical shape. This trend is predicted by the analysis since the attenuation of v' and the amplification of w' which cause this trend are incorporated into the model. The discrepancy between the data and the analytical results at x/D = 0 & 1.6 in Fig 6 is associated with large deviations of the measured profiles from the assumed profile shape $f(\eta) = \left(1-\eta^2\right)^2$. This approximation was much more accurate by x/D = 6.5. The data follows the analytical trends well except for a significant spread with different blowing ratios of the z direction data at low FST levels.

DISCUSSION AND CONCLUSIONS

Several significant features of the film cooling jet flow have been illustrated by the experimental data and the analytical modeling.

With an abrupt entry to the film cooling hole, the exit flow does not exhibit a traditional parabolic mean velocity profile. Instead, the profile is shaped like the top of a volcano with the velocity at the "crater" bottom at about 60 percent of the rim velocity. This finding is supported by the numerical work of Leylek and Zerkle (1993). This initial velocity profile contributes to the "lobed" shape of the coolant fluid at small x/D values. The "lobed" phenomena is generally attributed to vortex structure interactions. This data seems to clearly indicate that both phenomena contribute.

The film cooling jet is quickly turned to a direction nearly parallel to the plate and the main flow. The angle that the jet flow centerline makes with the plate at the downstream edge of the hole is dependent on the blowing ratio.

The initial turbulence in the film cooling jet hole dominates the spreading rate for blowing ratios near one and low FST levels. With a significant difference in velocity between the jet and the main flow the normal

jet entrainment mixing, neglected in this analysis, would be of about the same significance as the injection turbulence at low FST levels. Evidence of this was in some incomplete data taken at M = 0.6.

At Tu=17% the FST generated mixing dominates the decay and spreading of the coolant flow. The amplification of the w' component of the fluctuating velocity at the expense of the v' component is the key to understanding the enhanced spreading of the cooling jet flow in the transverse, z, direction. The amplification is most effective when the length scale of the turbulence is large compared to the distance from the jet centerline to the plate. Accordingly, this effect would be particularly significant with large scale, high intensity FST and low blowing ratios. This is in contrast to the 0.9% FST data where, at least for blowing ratios from 0.96 to 1.87, the jet flow remains essentially "round" except for plate effects.

Analytically the use of the energy equation multiplied by T was effective in permitting a simple integral analysis to predict the main features of the coolant flow mixing. Since this "energy concentration" equation requires modeling of the mixing heat transfer, it gives us the opportunity to introduce both the internal jet fluctuation contribution and the FST contribution to mixing. While the energy equation gives the simple conclusion that the energy in the mixed flow remains constant, the energy concentration equation permits modeling of the rate at which the energy becomes less concentrated.

While a "standard" value for the eddy diffusivity coefficient in equations (4) and (5) might be 0.086 (Schetz, 1980), the values of C_{fc}=0.04 for the internal mixing and C_{fs}=0.12 worked well for this flow. The smaller length scale corresponding to the turbulence in the cooling hole and the decay of this length scale in the mixing region correspond to the reduced value of the eddy diffusivity coefficient for the internally introduced fluctuations. The large length scales associated with the free stream turbulence and the constant reinforcement of these scales and fluctuations by the free stream correspond to the higher value of the eddy diffusivity coefficient for the free stream introduced fluctuations. Corresponding to the reduced length scale of the 6.5 percent FST level flows relative to the 17 percent, a slightly smaller value of C_{fs} for the 6.5 percent cases would have improved the agreement between the analysis and the data. Since the improvement was small and for simplicity, the single value of C_{fs} was retained for all the flows.

In summary, the dramatic effects of high FST levels on the spread of the film coolant is measured and predicted by a closed form model. This is intended as a first step toward the prediction of the change in film cooling effectiveness with FST as measured by Bons et. al. (1994).

REFERENCES

Bons, J.P., MacArthur, C.D. & Rivir, R.B., 1994. *The effects of High Freestream Turbulence on Film Cooling Effectiveness.* ASME 94-GT-51.

Hunt, J.C.R. & Graham, J.M.R., 1977. *Free Stream Turbulence near Plane Boundaryies.* J. Fluid Mech. Vol 84 pp. 209-235.

Leylek, J.H. and Zerkle, R.D., 1993. *Discrete-Jet Film Cooling: A Comparison of Computational Results with Experiments.* ASME 93-GT-207.

MacArthur, C.D., 1986. *Fluid Dynamics and Heat Transfer of the Circular Tangential Wall Jet.* Ph.D. Dis., Univ. of Dayton.

Maciejewski, P.K. & Moffat, R.J., 1989. *Effects of very High Turbulence on Heat Transfer.* Turbulent Shear Flow 7,2,20-3.

Schetz, J.A., 1980. *Injection and Mixing in Turbulent Flows.* Progress in Astronautics and Aeronautics, Vol. 68.

Thomas, N.H. & Hancock, P.E., 1976. *Grid Turbulence Near a Moving Wall.* J. Fluid Mech. Vol 82 pp.481-496.

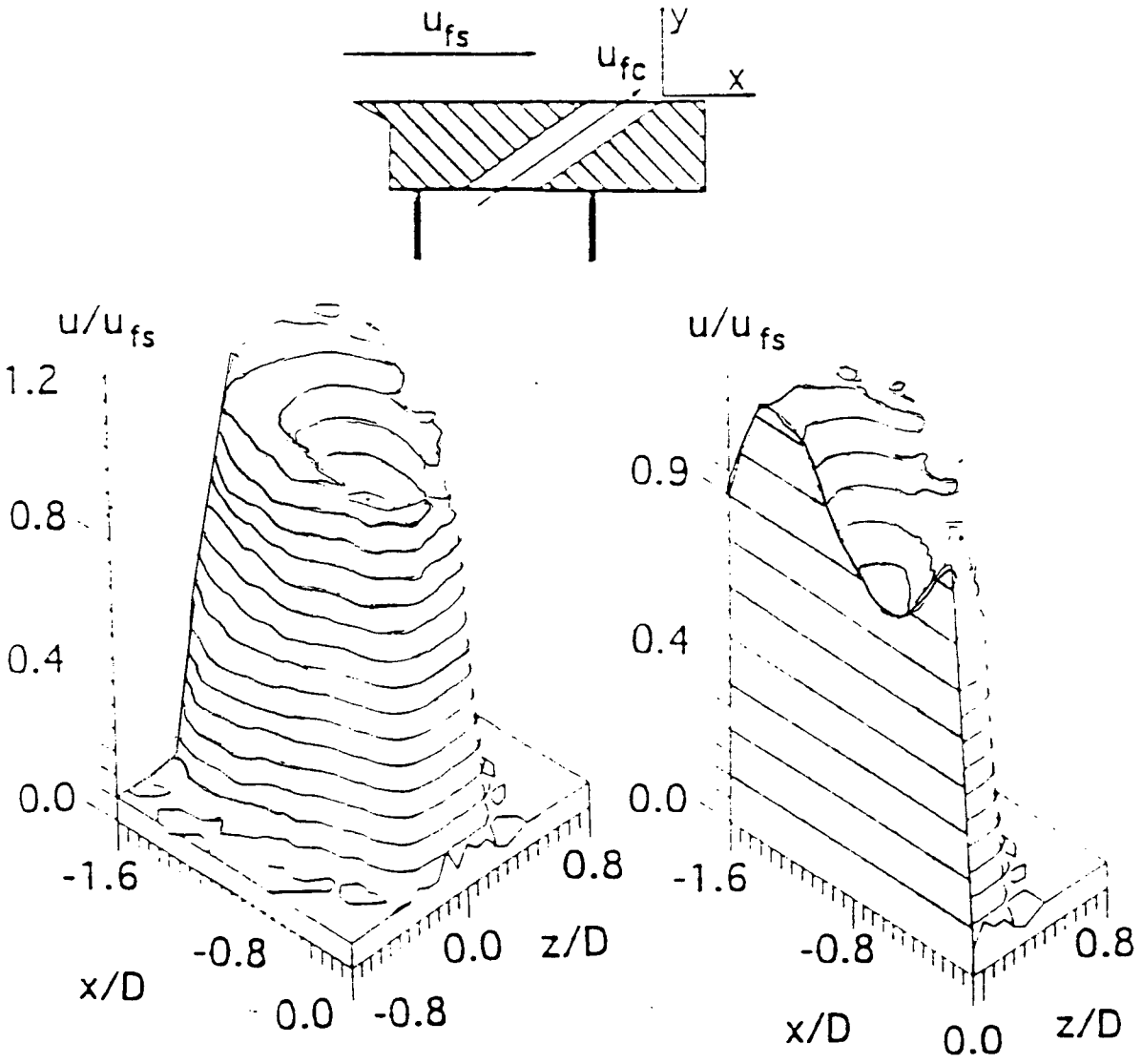


Figure 1: Mean velocity profile at the film cooling injection hole exit plane (cross-section also). Tu = 0.9% and M = 1.07.

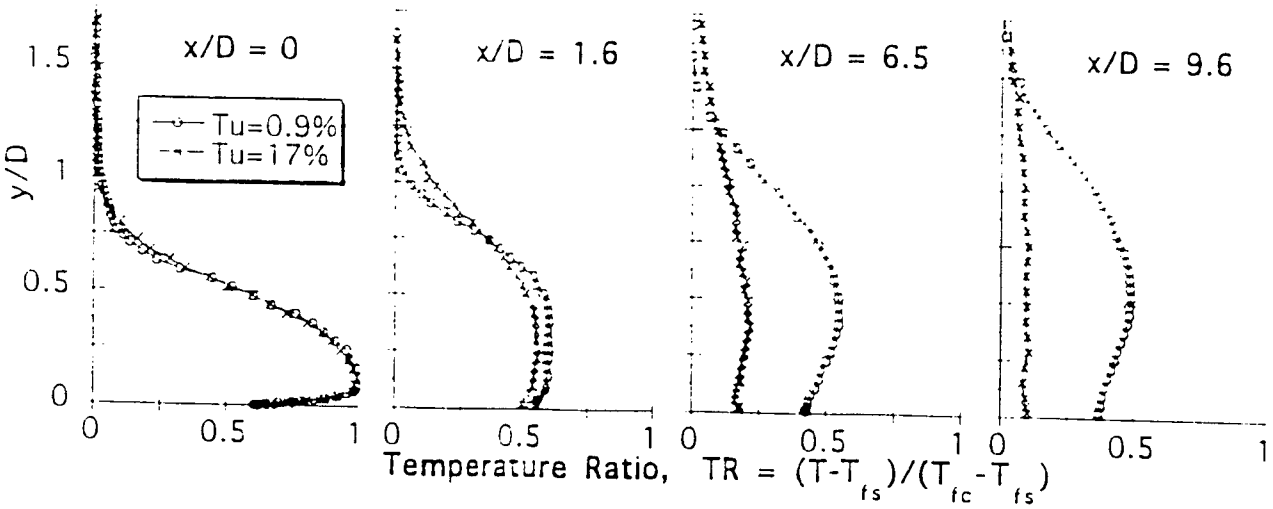


Figure 2. Vertical (y) Jet Tempurature Ratio Profiles at 4 x/D Locations Comparison of Data a 2 FST Levels and Identical Blowing Ratio, M = 1.07

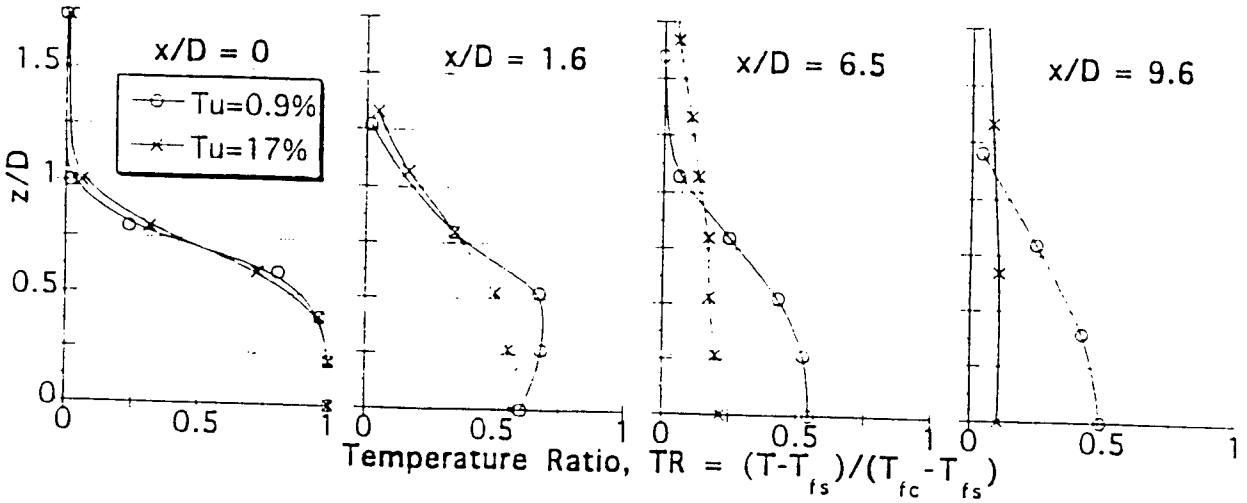


Figure 3. Transverse (z) Jet Tempurature Ratio Profiles at 4 x/D Locations Comparison of Data a 2 FST Levels and Identical Blowing Ratio, M = 1.07

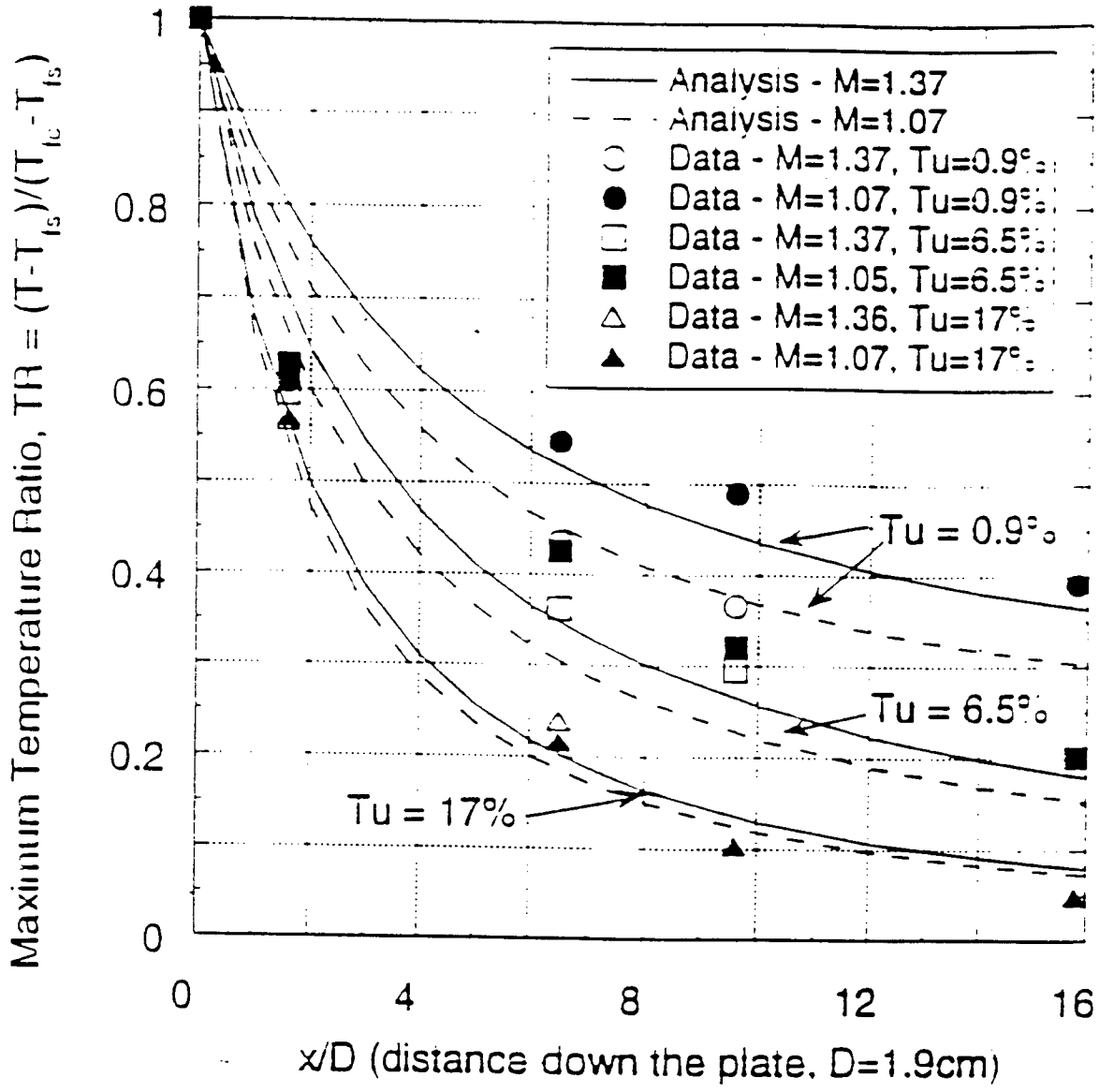


Figure 4. Maximum Temperature Ratio Decay with x/D. Analysis Compared to Experimental Data for 3 FST Levels and 2 Blowing Ratios

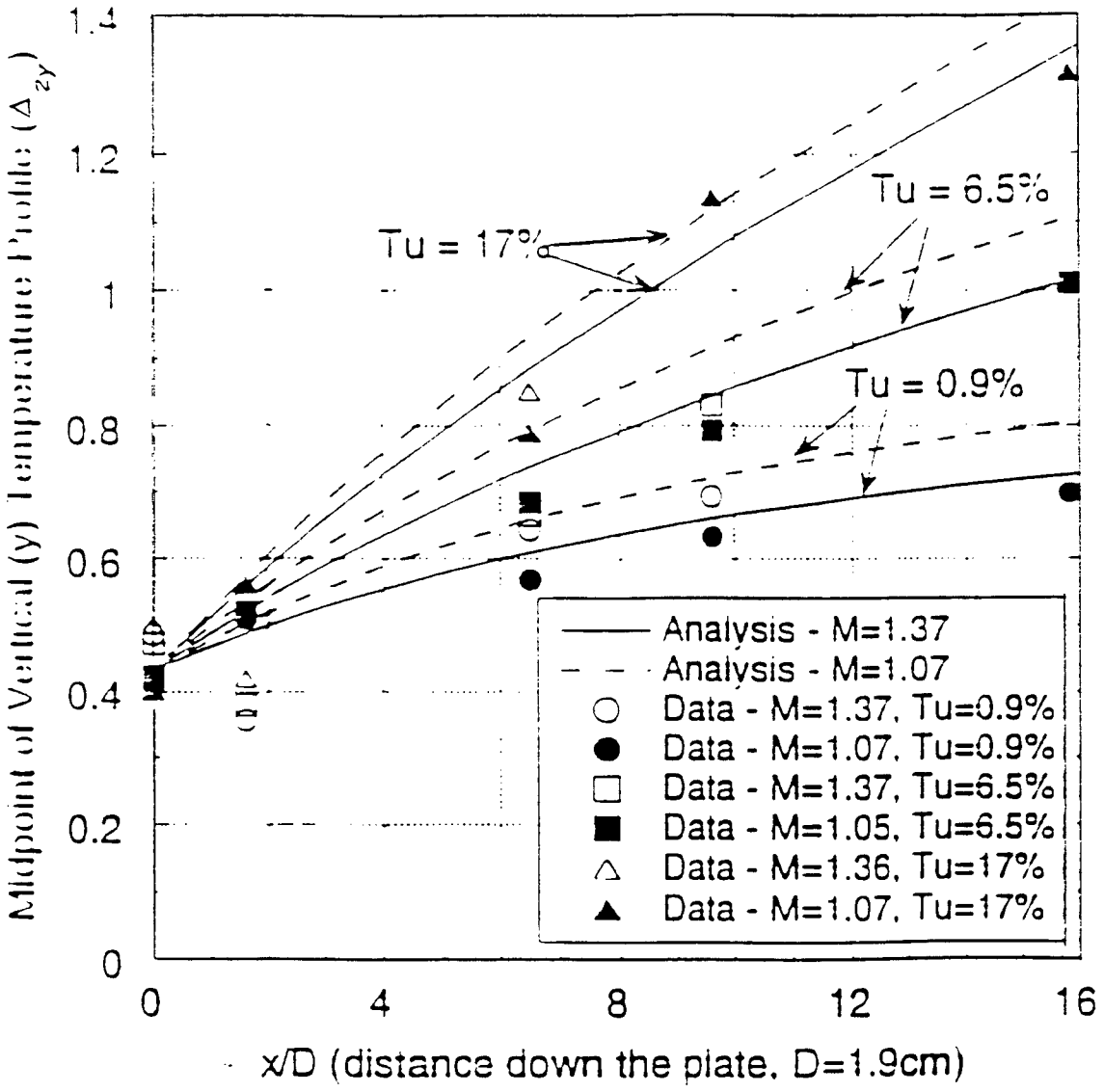


Figure 5. Vertical (y) Jet Spread with x/D. Analysis Compared to Experimental Data for 3 FST Levels and 2 Blowing Ratios

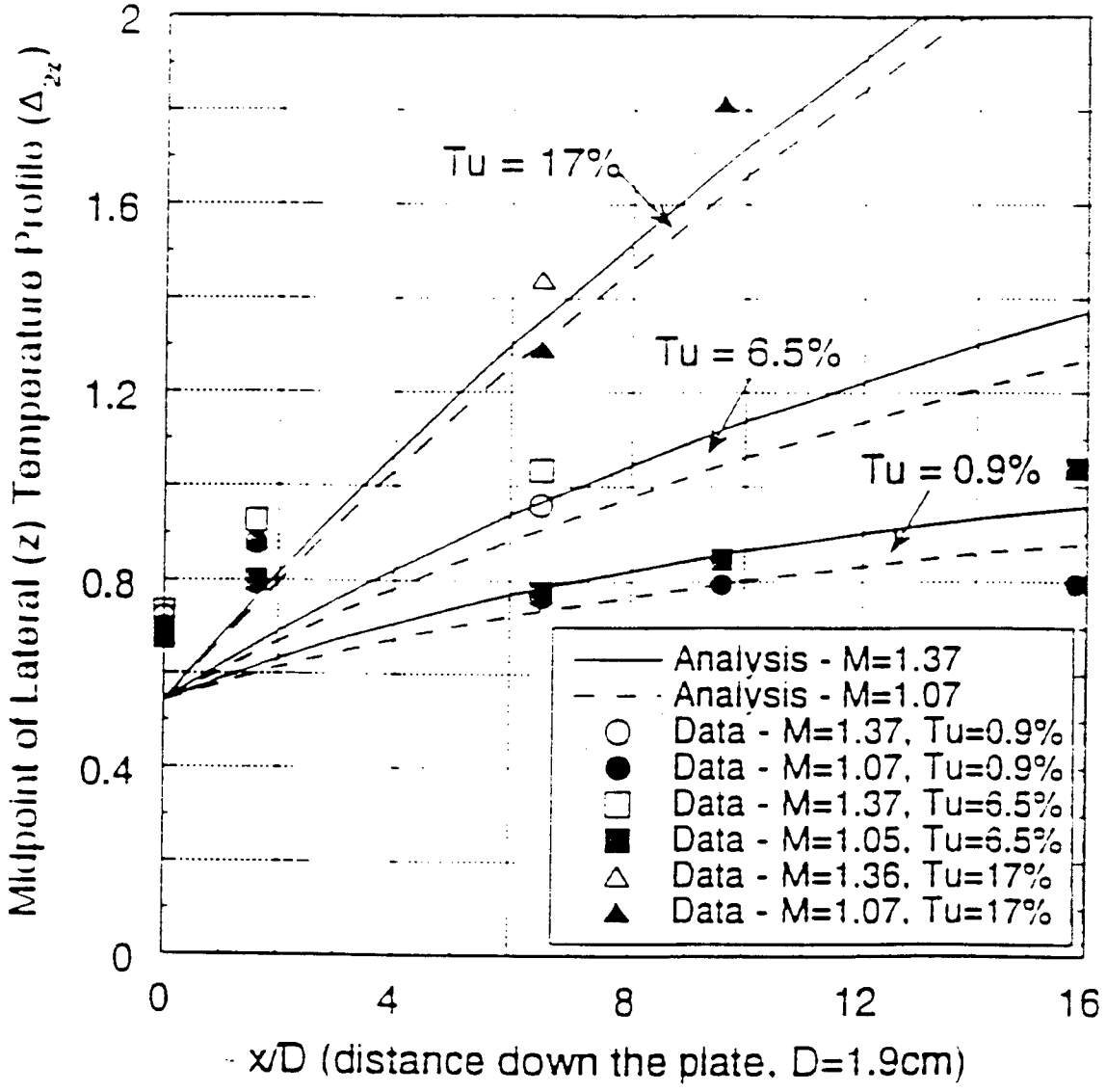


Figure 6. Transverse (z) Jet Spread with x/D. Analysis Compared to Experimental Data for 3 FST Levels and 2 Blowing Ratios

Neural Network Identification and Control in Metal Forging

Carla A. Schwartz
University of Florida
Department EE
405 CSE, Building 42
Gainesville, FL 32611-2024
USA

Final Report for:
Summer Faculty Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC
and
Wright Laboratory
AFOSR Contract F49620-93-C-0063

September, 1994

Neural Network Identification and Control in Metal Forging

Carla A. Schwartz
University of Florida
Dept of Electrical Engineering
405 CSE, Building 42
Gainesville, FL 32611-2024
email: carla@neptune.ee.ufl.edu

Abstract

Forged metal is used as a strengthening technique to form parts for a wide variety of applications. Currently, many aspects of the forging process are imprecise and/or ad hoc. Thus, the metal forging area is ripe for the application of process control techniques to improve the forging process for efficient and effective manufacturing.

There have been some attempts to apply control techniques to forging in the past. Most of these involve using open loop control. This report describes the application of neural network identification and feedback control techniques to some metal forging processes.

2 Introduction

Metal forging, the process of applying stress to metal in order to harden and form it, has been in use since ancient times. Traditionally, forging has been an imprecise process, relying on trial and error, rather than analysis, to make decisions such as initial billet[1] shape, forging temperature, and selection of ram[2] velocity profiles. This is still largely true, despite the high performance of modern equipment and the current understanding of fundamental material behavior. Typical industrial forge presses cannot provide accurate velocity or force profiles. Although precise computer-controlled[3] presses are available, they are not in widespread use. This is mainly due to the lack of systematic procedures for selection of forging process design variables.

Thus, even though sophisticated materials models are available, there is still much room for improvement in the metal forming industry, in order to achieve an efficient process for manufacturing. A typical objective in a forging is to achieve a particular microstructure,[4] which may not be uniform throughout the finished piece, and that the process not have any failures such as cracks, or failure of the billet to completely fill the die.[5] Process errors may be due to any of several aspects of the forging process: improper ram velocity profile designation,[6] poor choice of initial billet shape, or lack of proper lubrication.

Problems present in the metal forging industry seem ripe for the application of control ideas in order to utilize information provided by sophisticated mathematical models to identify parameters for use in optimization of forging process. In this work, we proposed to use feedback control to improve the metal forging process.

The work of Malas, et.al. [11] and Grandhi et.al. [7], represents one of the first attempts to include some form of control in the forging process. Other work in this area has been carried out in [3, 2]. Real-time model-based closed loop control has been proposed for hot isostatic pressing by Meyer and Wadley [14], using detailed models of the powder consolidation process, and assuming direct measurement of certain microstructural variables. These types of models are unavailable for the forging problem.

Berg et. al., [3], used a commercial finite element (FEM) code, *ANTARES*, to generate strain rate profiles for use in an open loop optimization of ram velocity profiles. That

[1]The billet is the workpiece to be formed.

[2]The ram is attached to and drives one of the dies as it stresses the billet to form it. Dies are the pieces used to stress the billet into its form.

[3]open loop control

[4]measured by grain size and percent of material recrystalized

[5]improper shape

[6]possibly due to insufficient knowledge of the process parameters

study optimizes the choice of ram velocity profiles in order to track a desired strain rate profile. Mathematical models relating strain, strain rate, and temperature to microstructure of the material form the basis of the open loop optimization procedure. This technique was demonstrated on a simulation of an upsetting of steel, with and without consideration for friction.

The FEM code, *ANTARES*, uses a flow stress model of the billet material, along with the complete thermo-mechanical characterization of the forging environment, to predict, among other quantities. strain, strain rate, and temperature profiles as a function of stroke, (position of the ram-driven die/billet interface), throughout the workpiece. *ANTARES* may also calculate the ram force as a function of stroke.

The inputs *ANTARES* requires for its computation include die velocity profiles, (as a function of stroke) initial temperature, heat transfer coefficient, and a lubrication factor. The heat transfer coefficient, (HTC), is difficult to measure, hard to predict accurately, and may change during the forging. Changes in HTC can only be accounted for in *ANTARES*, if the time variation of the heat transfer is known.

The uncertainty of model parameters and material dependent microstructural models naturally leads to the idea of implementing adaptive feedback control in a forging process.

Although *ANTARES* models almost any aspect of the forging process, it is based on highly complex nonlinear partial differential equations models. *ANTARES* takes hours to run one simulation of a forging process with a fixed set of parameter values. It is impossible to invert *ANTARES*, and impossible to use it as a model for parameter identification or optimization in real-time control.

Due to these limitations, in order to use control techniques to operate on a forging in real time, it is necessary to have fast computational models relating input and output variables which are identified as important for precision control of forging. With this in mind, in this report it is proposed to use neural network models to model input/output behavior of the forging process, in order to use these models for real-time parameter identification and control.

This report is organized as follows. The next section contains a small introduction to the use of neural networks in control. Following this is a description of how neural networks will be used for identification and control of a forging process. Finally, conclusions will follow.

3 A Brief Introduction to Neural Networks in Control

Over the past few years, there has been a flurry of activity in the area of applying neural networks to identification and control of nonlinear systems, ([4, 5, 9, 13, 15, 18, 8, 6], for example).

A major premise in much of the literature pertaining to the application of neural networks to the adaptive control of nonlinear systems is that the nonlinear systems models are feedback linearizable with stable zero dynamics, ([4, 15]). The reason for this assumption is that a simple, causal, stable controller may be solved for in this case. This assumption is unnecessary provided that the neural network model has the property that its targeted outputs are controllable from the control variables in question. Feedback linearizability of a nonlinear model is sufficient, but not necessary for the model in question to be controllable. For other nonlinear control systems models, the concept that the output variables are controllable from the input, or, equivalently, that target tracking can be achieved for some permissible input, is strongly related to the invertibility of the network. By invertibility of the network, we mean the ability of the network to trace backwards through the network, starting from a arbitrary output, to reach an input which generated that output through the network. Although the method of iterative inversion, [10], has been applied in several applications, [8, 6], for neural network models, it is unnecessary to address the issue of inverting causal models in order to implement control, since inverse models may be trained directly from the observed *input* and *output* data.

The standard back propagation algorithm for neural network models, [17], is a gradient descent algorithm for determining the weights and biases for a neural network during the training. This method uses target outputs, and successively works backwards through the layers of the network, to generate the gradients used for updating the weights at each layer.

There are several important decisions a modeler makes in choosing a configuration for a neural network model. The number of hidden layers, the nonlinear activation function acting at each layer, whether or not the model is recurrent, the learning algorithm, the amount of training data, the learning rate, and the error tolerance must all be chosen. These choices will influence the speed and accuracy of the model, as well as its ability to generalize.

Teixeira et. al., [18], derive a method, in terms of the size of the training data and the number of inputs and outputs, for calculating the minimum number of hidden elements in a neural network, with either one or two hidden layers, in order for the neural network to represent an input/output relation with no error.

There are some rules of thumb for chosing some of these other parameters, but sometimes it is necessary to try several configurations before a good neural network model is derived.

The next section discusses the use of neural networks in a feedback configuration for real-time adaptive control of a forging process.

4 Feedback Control in Forging Using Neural Networks

As was stated in the introduction, metal forging is a complex process for which there are no fast computational models. The models currently available are quite complex, and are based on highly nonlinear partial differential equations models depending on some unknown parameters, as well as possibly inaccurate microstructural models. Here a neural network based adaptive control scheme for real-time control of a forging process will be described. The process is currently being tested.

Ideally, the forging process itself would serve as a plant in the proposed adaptive control scheme, for which real-time measurement of data such as ram force, temperature, and billet shape would be used as feedback for estimation of the neural network based controller parameters. It is difficult to make many experiments with actual forgings. Additionally, at this point in time, it is too early to expect that sensors for several appropriate signals can be implemented in experiments, but there is a future possibility for this hope to be realized, [16].

As a result of these two factors, the FEM forging simulation program, *ANTARES*, will be used, at least initially, to generate plant data (over several operating conditions) which will be used as learning data by the neural network models. Additionally the only measured signal proposed to be used for training the networks in these first experiments is the ram force, since this signal is currently feasible for measurement in a forging. The initial experiment described in this work will be the extrusion of a cylindrical piece of steel.

Two neural network models will be used for learning the input/output behavior of *ANTARES*. The first model, the controller, will learn from training data generated by open loop optimal velocity[7] profiles which are generated by *ANTARES* for different sets of (constant and varying) heat transfer coefficients,[8] and output training data generated by the average weighted strain rate over the billet.[9] The learned network will act as a model for *ANTARES* (with optimization) which relates average strain rate, and HTC, (as a function

[7] as a function of stroke, which is the height of the billet, and acts as a marker of time

[8] also a function of stroke

[9] The weighting is based on strain information and the microstructural models.

of stroke), as inputs, to velocity, (as a function of stroke), as output. The second model, the HTC estimation model, is similar to the controller neural network, only it is trained on a wider range of velocity data and uses ram force instead of strain rate as the other input training variable.

Figure 1 depicts the two neural networks which will be trained on *ANTARES* data. Each network will be a three-input,[10] one-output,network with two hidden layers of four nodes each. The nonlinear activation functions at the hidden layers will be hyperbolic tangent sigmoidal functions, and the output layer will be linear.

Figure 1(a) represents the training of the controller network. The controller network is trained to reproduce, as closely as possible, the optimal ram velocities, (as a function of stroke), generated by the optimization algorithm. The desired weighted average strain rate is an input to both the optimization algorithm and the controller network. It is calculated based on microstructural material models, and may be stroke dependent. This allows enough flexibility to specify nonuniform grain sizes in the finished product. The training set will include about 240 input/output data triples, (HTC and strain rate, (as functions of stroke), as inputs and velocity, (as a function of stroke), as the output).

Figure 1(b) represents the training of the HTC estimation network. The estimation network is trained to reproduce, as closely as possible, the relationship between ram velocity, ram force, and HTC, (all as functions of stroke), as predicted by *ANTARES*. Note that the HTC and ram velocity are inputs to *ANTARES*, and ram force is an output, while the estimation network takes ram force and ram velocity as inputs and generates the HTC. Causality is not an issue in training the neural network, which is simply building a static mapping between values. The training set will include about 720 input/output data sets.

The result of each of these neural network trainings will be a set of weights (and biases) which characterize input/output models of a forging, (*ANTARES*).

Figure 2 depicts the use of the controller and estimation networks in an adaptive feedback control scheme. Initial experiments will be simulations, and the plant will be ANTARES, but future work will replace ANTARES with an actual forge press. Note that the trained networks, with weights held fixed, are used exactly as they were trained, except the controller takes the estimated HTC value instead of the actual value. The locations of the various inputs and outputs have been slightly rearranged in the figure.

As this work develops, other parameters and measured variables may be included in the adaptive control scheme.

[10]The third input will correspond to stroke.

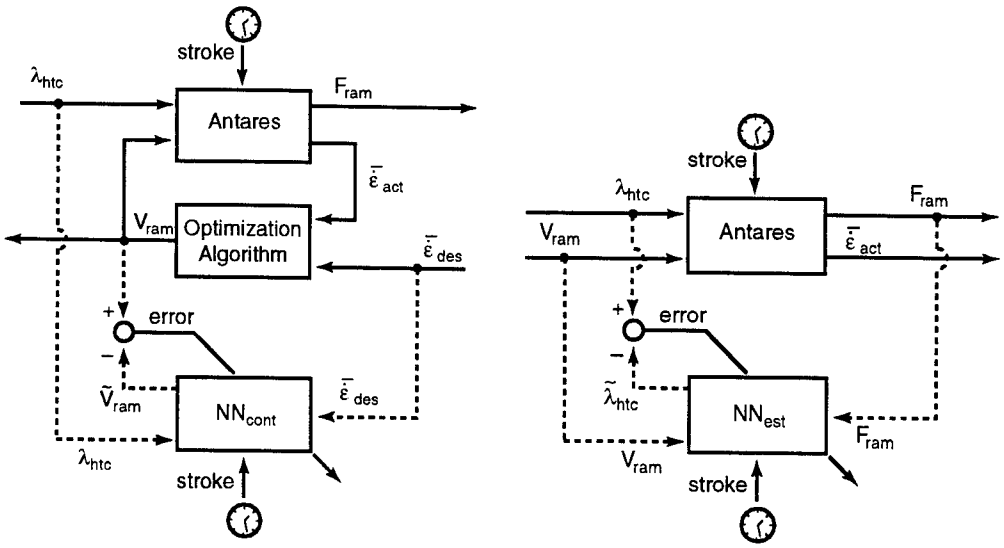


Figure 1: Training of the forward neural network from ANTARES: (a) For the controller (b) For the heat transfer coefficient estimator

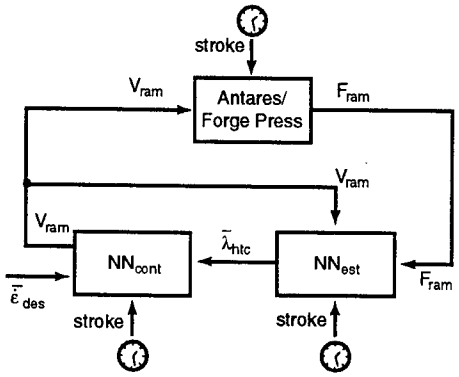


Figure 2: Adaptive Feedback Control for Metal Forging

5 Conclusion

A neural network based adaptive control scheme for real-time control of a metal forging process has been presented. The approach relies on neural network models relating stroke, velocity, and heat transfer coefficient to load and strain rate measurements, and the use of back propagation within these models for control optimization. The proposed scheme is currently under test using data generated by *ANTARES*, and will be tested in forging experiments. The scope of the proposed method will be expanded as appropriate sensing devises are made available. This work represents a new approach to the use of feedback control in metal forging. A novelty of the method presented here is the use of neural networks to train inverse models for direct identification and control.

References

[1] H. Demuth and M. Beale. *Antares User's Manual Version 3.0.* United Energy Systems, Inc., Dayton OH, 1993.

[2] J. M. Berg, R. J. Adams, J. C. Malas, and S. S. Banda. Design of Ram Velocity Profiles for Isothermal Forging via Nonlinear Optimization. *Proceedings, American Control Conference* , pp. 323-327, 1994.

[3] J. M. Berg, R. J. Adams, J. C. Malas, and S. S. Banda. Nonlinear Optimization-Based Design of Ram Velocity Profiles for Isothermal Forging. *IEEE Trans. Control Applications*, submitted, 1994.

[4] F-C. Chen and H.K. Khalil. Adaptive Control of Nonlinear Systems Using Neural Networks– A Dead-Zone Approach. *Proceedings of the American Control Conference*, pages 667-672, 1991.

[5] S.R. Chu and R. Shoureshi. A Neural Networks Approach for Identification of Continuous-Time Nonlinear Dynamic Systems. *Proceedings of the American Control Conference*, pages 667-672, 1991.

[6] D.T. Davis,Z. Chen, L. Tsang, J-N Hwang, and A.T.C. Chang. Retrieval of Snow Parameters by Iterative Inversion of a Neural Network. *IEEE Trans. Geoscience and Remote Sensing*, vol. 31, no. 4, pages 843-852, 1993.

[7] Grandhi, R.V., A. Kumar, A. Chaudhary, and J.C. Malas State-Space Representation and Optimal Control of Non-Linear Material Deformation Using the Finite Element Method. *Int. J. for Num. Meth. in Engr.*, Vol. 36, 1993, pp. 1967-1986.

[8] D.A. Hoskins, J-N Hwang, and J. Vagners. Iterative Inversion of Neural Networks and Its Application to Adaptive Control. *IEEE Trans. Neural Networks*, vol. 3, no. 2, pages 292-301, 1992.

[9] K.J. Hunt, D. Sbarbaro, R. Zbikowski, and P.J. Gawthrop. Neural Networks for Control Systems—A Survey. *Automatica*, vol. 28, no. 6, pages 1083-1112, 1992.

[10] A. Linden and J. Kindermann. Inversion of Multilayer Nets. *Proceedings, International Joint Conference on Neural Networks*, vol 2, pp 425-430, 1989.

[11] J.C. Malas and V. Seetharaman. Using Material Behavior Models to Develop Process Control Strategies. *Journal of Materials*, vol 44, pp 8-13, 1992.

[12] H. Demuth and M. Beale. *Neural Network Toolbox User's Guide.* MathWorks, Inc., 1992-93.

[13] D. McGrane R. Smith, and M. Mears. A Study of Neural Networks for Flight Control. *Proceedings, American Control Conference*, Baltimore, MD, 1994.

[14] Meyer D., and H.N.G. Wadley, Model-Based Feedback Control of Deformation Processing with Microstructure Goals. *Metallurgical Transactions B*, Volume 24B, April 1993, pp. 289-300.

[15] K.S. Narendra and K. Parthasarathy. Identification and Control of Dynamical Systems Using Neural Networks. *IEEE Trans. Neural Nets*, vol 1, no 1, pp 4-26, 1990.

[16] NMAB Committee on On-Line Control of Metal Processing, *On-Line Control of Metal Processing* Cambridge, MA, MIT Press, 1986.

[17] D. Rummelhart, J. McClelland and the PDP Research Group, *National Materials Advisory Board, Report no. NMAB-444* February, 1989.

[18] E. Teixeira, K. Loparo, and F.A.C. Gomide. Design of Multi-Layer Neural Networks for Accurate Identification of Nonlinear Mappings. *Proceedings of the American Control Conference*, pages 14-15, 1991.

A PRELIMINARY STUDY OF MULTIPLE JET MIXING AND ATOMIZATION IN REACTING AND NONREACTING FLOWS

James P. Seaba
Assistant Professor
Department of Mechanical & Aerospace Engineering

University of Missouri-Columbia
E2403 Engineering Building East
Columbia, MO 65203

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

September 1994

MULTIPLE JET MIXING AND ATOMIZATION IN REACTING AND NONREACTING FLOWS

James P. Seaba
Assistant Professor
Department of Mechanical & Aerospace Engineering
University of Missouri-Columbia

Abstract

The use of smaller multiple jets instead of a large single jet to promote mixing and atomization for gas turbine combustor applications is investigated. The mixing characteristics using different orifice patterns in a circular configuration are characterized using three different techniques; Reactive Mie Scattering (RMS), preformed Al_2O_3 seeds, and acetone based Laser Induced Fluorescence (LIF). The spray was characterized by light sheet visualizations where the spray was introduced at the center of the circular orifice pattern issuing dry air. The air jets and spray behave as an air blast atomizer, seven different orifice configurations were assessed. Two different liquids were used for the cold spray visualizations, water and stoddard solvent. It was observed that the stoddard solvent best resembled the atomization characteristics of JP8 plus 100 as compared to water. The air blast atomizers were also tested in a single dome combustor, where the multiple jets were introduced in the liner wall 45 mm downstream of the center dome. A single 1/2" orifice jet was positioned opposite the multiple jet forming an opposed jet configuration 45 mm downstream of the dome. Only the top multiple jet and the dome contained fuel. The inlet air temperature was 500 °F for the side jets and dome. The side jet injection using fuel and air created a more stable combustor, the lean blowout was decreased by 50%. This preliminary investigation of side fuel injection shows promise for better combustion efficiencies and possible pollutant reduction.

MULTIPLE JET MIXING AND ATOMIZATION IN REACTING AND NONREACTING FLOWS

James P. Seaba

Introduction

The purpose of this experimental investigation is to access the atomization and mixing characteristics of multiple jets as compared to a single large jet of equivalent area. It is desired for multiple jets to enhance the mixing and atomization of fuel in the near field region of a gas turbine combustor. This would provide better stability, higher combustion efficiency, and possibly lower pollutant emissions as compared to combustors with large single holes in the liner. It has been observed that the spray characteristics of a nozzle centered in a single large hole collapses the spray into an undesirable liquid jet stream.

Multiple jet configurations lend themselves to better mixing compared to a single jet. An experimental study of four round jets placed in a square configuration were compared to a single jet (Moustafa and Rathakrishnan, 1993). The characteristics of the multiple jet configuration were; the potential core of the individual jets decreases compared to a single jet, axis switching is characteristic of this configuration meaning better mixing and noise reduction, and the mean velocity profile and decay are unaffected by stagnation pressure. Multiple two-dimensional slot jets have also be studied (Krothapalli, 1979, Grandmaison and Zettler, 1989, and Ng and Grandmaison, 1989). Again, the potential core was shorten, and the jets spread out much more quickly than a single slot jet. Also, a low pressure region between the slot jets was observed when the jets were placed within one slot width of each other.

Previous experimental work has investigated multiple jets with applications to thrust augmentation and noise reduction in the aerospace industry. Presently, no previous work has been located which investigates the combination of atomization and mixing using multiple jets. This preliminary investigation is aimed at determining the important characteristics and parameters associated with multiple jet mixing and atomization relative to jet engine combustor applications.

Experimental Procedures

Multiple Jet Visualization Setups

The seven multiple jet nozzles used in this investigation are shown in Figure 1. The small hole in the center of each of the configurations is for a spray nozzle, the nozzle assembly consisting of the spray nozzle and the multiple air jet disk is shown is Figure 2. The seven multiple air jet configurations and the 0.5 inch diameter hole have approximately the same overall exit area of 0.196 $in.^2$ The seven multiple jet nozzles are labeled relative to the number of holes and the angle of the holes, i.e., S12A0 represents a 12 hole circular pattern, the holes drilled straight through the disk (0 angle). A more complicated geometry is the S16A15S15, 16 holes drilled at an outward angle of 15 degrees and swirled 15 degrees radially from vertical. The outer holes issue air to atomize the spray and mix air with the vaporized liquid spray. The individual jet diameters of the multiple air jet nozzles are listed in Table 1.

Three different Delevan "peanut" spray nozzles were investigated using spray angles of 60, 70, and 90 degrees with corresponding flow numbers (Fn) of 2.0, 1.0 and 1.1, respectively. Flow visualization techniques were used to qualitatively determine and compare the mixing of the multiple jets and a single half inch hole. The visualizations techniques consisted of Reactive Mie Scattering (RMS), preformed Al_2O_3, and Laser Induced Fluorescence (LIF). The liquid spray was also characterized by visualizations using a laser sheet (532 nm) and 35 mm photographs. Two fluids were studied; water and stoddard fluid. The stoddard fluid has similar fluid properties and atomization characteristics as compared to JP8+100 fuel.

Table 1 Multiple Air Jet Conditions

Nozzle Name	Qair scfm	Jet dia in	Jets No.	Air area in2	Avg. velocity m/s
S12A0	21.8	0.144	12	0.195	82
S16A0	20.5	0.125	16	0.196	76
S28A0	19.9	0.093	28	0.190	77
S12A25-3S10	21.5	0.144	12	0.195	80
S16A30	19.9	0.125	16	0.196	74
S16A15S15	21.5	0.125	16	0.196	80
0.5 hole	21.8	0.5	1	0.196	82
S.Annular	22.4	NA	NA	0.195	84

The pressure drop across the multiple air jet disk was measured using a water filled manometer. The pressure tap was placed approximately 12.5 mm upstream of the multiple jet disk on the tube housing as shown in Figure 2. This is the identical pressure tap placement as used for the combusting experiments discussed herein. The air flow rate was regulated using a Fisher Porter rotameter (Tube # FP3/4 21-G-10/55) using a 60 psig line pressure. The maximum flow rate for the rotameter is 22.8 SCFM with a minimum flow rate of 2 SCFM and an accuracy of ±0.25 SCFM. The manometer has an accuracy of ±0.25 in. H_2O with a full scale of 25 in. H_2O. The spray had little effect on the pressure drop measurements using S12A0 at 20 SCFM. Therefore, the pressure drop was measured using air only, no spray from the center nozzle was present during the measurements. The experimental schematic is shown in Figure 4. The pressure drop was recorded for the seven different multiple jet configurations and a 0.5 in. hole over the range of the flow meter, 2.0 to 22.8 SCFM. The exit air conditions of the multiple jet configurations are summarized in Table 1.

RMS and Al_2O_3 Visualization facility

The Reactive Mie Scattering (RMS) and preformed seeds of micron sized Al_2O_3 particles provided scattering media in the flow field of multiple jets. The flow visualizations were photographed at the flow conditions listed in Table 1. The RMS and Al_2O_3 test setups are shown in Figure 4 . The 532 nm laser sheet from a YAG laser cut through the center vertical plane of the multiple jet configuration. Photographs were taken with a 35 mm camera using a macro-lens. The camera settings were F3.3 and1/8 sec shutter speed using 400 ASA film. The laser was pulsed at 10 Hz, with no timing required between the camera and laser.

Acetone based LIF

The air jets issuing from multiple jet configurations S16A0 and S28A0 were doped with acetone vapor. The flow field was viewed from two different axis, both the vertical and horizontal planes were analyzed. Horizontal planes at 0.5 mm, 5 mm, 10 mm, 15 mm, 25 mm, and 50 mm were taken at two flow rates, 3 and 20 SCFM. The vertical plane was taken on the center axis. Due to laser power limitations a 25 mm vertical light sheet at the test section was used. The vertical plane was recorded at three different downstream locations to provide a total vertical sheet length of 75 mm on the center plane. Acetone is excited in a range of 250 to 290 nm with a peak absorption at 280 nm. The output of the fluorescence is at 310 nm. The second harmonic output (532 nm) of the YAG was doubled using a BBO crystal to provide the fourth harmonic (266 nm) output. The 266 nm wavelength

was formed into a light sheet which was focus at the test section to excite the acetone as shown in Figure 5. A photometrics CCD array system provided the data acquisition of the acetone fluorescence as shown in Figure 6.

Cold Spray Apparatus

The spray pattern was qualitatively analyzed by using a pulsed vertical light sheet from a YAG laser at 532 nm. The light sheet was positioned on the centerline from the nozzle exit to 15.24 cm downstream. The nozzle is pointed downward for all spray conditions, unlike the gas mixing visualizations discussed previously. A shallow pan was positioned approximately one meter downstream to collect the fluid during test. The pan had negligible effects on the spray pattern. Two different fluids were accessed; water and stoddard solvent. Water is a fluid commonly used in atomization studies, the stoddard solvent has fluid properties similar to that of jet fuel. Photographs were taken with a 35 mm camera using a macro-lens. The camera settings were F3.3 and1/8 sec shutter speed using 400 ASA film. The laser was pulsed at 10 Hz, with no timing required between the camera and laser.

Combustor Apparatus

A single dome combustor was used to test the multiple air jet spray in the side injection mode as shown in Figure 7. The top liner injection used both air and fuel with a multiple jet disk, the bottom liner injection used only air with a 0.5 inch diameter hole. Both side injection nozzles maintained the same air pressure drop. The inlet air to the main and side injectors was maintained at 500° F. The fuel used in the main and top side injectors was JP8+100.

Results and Discussion

The Results and Discussion section are divided into 4 areas; pressure drop across the nozzles, mixing of the air jets, multiple jet atomization of fluids and combustion studies of the multiple jets in a single dome combustor. The majority of the results are qualitative, however, it provides valuable insight for future studies.

Pressure Drop Across Nozzles

The pressure drop versus air flow rate of the multiple jet configurations and the 0.5 inch diameter hole is shown in Figure 8. At 22.8 SCFM, the nozzle pressure drops are between 12.5 and 18 inches of water, corresponding to a 3 to 5 percent pressure drop across the combustor liner. The pressure drop across the nozzles increases as the hole diameters decrease, and swirl and jet angles are introduced. The 0.5 inch hole maintained

nearly an identical pressure curve as compared to S12A0. The S16A30 nozzle had the largest pressure drop for the multiple jet configurations tested.

Multiple Air Jet Mixing

$TiCl_4$ visualizations for the seven different multiple jet configurations and the 0.5 inch hole are shown in Figure 9 (8 photographs). The photograph sequence correlates to the nozzle configuration pattern in Figure 1 with the exception of the 0.5 inch jet. For example, the upper left photograph represents the S.annular nozzle. The 0.5 inch jet is shown in the lower right corner. The individual air jet velocities are between 74 and 84 m/s for the nozzle configurations tested, each nozzle flow condition is summarized in Table 1. The upper four photographs (S.annular, S12A0, S16A0, and S28A0) reveal similar near field flow characteristics. The jets merge together, and the outer jet shear layers (overall jet diameter) move radially inward until the jets merge downstream. Once the jets merge, the overall jet diameter starts to spread out, similar to a single jet. The jets merge at approximately 20 mm downstream of the jet exits. The core region of the multiple jet configurations is defined as the middle of the multiple jets prior to the jets merging. The core region for S.annular, S12A0, S16A0, and S28A0 exhibit a strong recirculation zone as seen by the even intensity of Mie scattering TiO_2 particles. LIF visualization of the S16A0 nozzle flow field reveal a relative even intensity of acetone fluorescence even near the exit (0.5 mm above the multiple jet exits) as shown in Figure 10. A white line was drawn on the LIF images which define the jet boundary. The horizontal planes at 0.5, 10, 20, and 30 mm downstream of the jet exit indicates the overall jet diameter decreases between the jet exit and 20 mm downstream. After 20 mm the jet starts to spread out, becoming more dilute from the entrained ambient air. This trend agrees with the $TiCl_4$ visualizations previously discussed. It is also conjectured that the near field recirculation zone is responsible for the atomization characteristics associated with these nozzles.

The other three nozzle configurations, S16A30, S16A15S15, and S12A25-3S10 do not reveal decreasing overall jet diameters in the near field. The multiple jet spreads out starting from the exit and shows better mixing as compared to the other four nozzles. The degree of mixing is qualitatively shown by the lack of intensity gradients in the downstream flow field. The 0.5 inch diameter round jet reveals the characteristic potential core region where no TiO2 particles have formed due to the lack of external mixing. The preformed Al_2O_3 particle visualizations did not add any additional information to the RMS visualizations.

Spray Characteristics

The eight spray photographs in Figures 10 and 11 represent the seven multiple air jet nozzles in the same pattern as shown in Figure 1. The lower right corner is a photograph of the Delevan "peanut" nozzle with no air, liquid only. The air flow conditions are defined in Table 1 where the jet velocities are between 74 and 84 m/s, identical to the $TiCl_4$ conditions. A 90 degree peanut nozzle at 30 psig, Fn = 1.1 is used to illustrate the spray characteristics. Figure 10 reveals the spray characteristics of stoddard solvent for the air jet configurations tested. The top four photographs corresponding to S.annular, S12A0, S16A0, and S28A0 nozzles, expand the liquid sheet outward to 180° at the exit and form a well focused atomized stream. The atomization characteristics are attributed to the strong recirculation zone formed at the exit of the air nozzle as discussed in the previous section. The three other nozzles (S16A30, S16A15S15, and S12A25-3S10) form sprays that expand out from the center spray nozzle. The 90° Delevan peanut nozzle was also placed in the middle of a 0.7 inch diameter hole, providing an annular air region with the area equivalent to a 0.5 inch diameter hole. The high velocity air (80 m/s) collapsed the spray into a non atomized liquid stream. It is desired to have a high momentum spray for liner injection applications in order to overcome the momentum induced by the cross flow produced by the main dome injector. The four spray patterns (S.annular, S12A0, S16A0, and S28A0 nozzles) appear to be the best candidates for this application.

A comparison between stoddard fluid and water atomization is illustrated in Figures 10 and 11. The Stoddard solvent (Figure 10) does not atomize as well as water, but better represents the atomization characteristics of real fuel. The use of water to model JP8+100 characteristics is overly optimistic and should be avoided.

The effects of blockage at a single hole was tested using the S12A0 air nozzle with the 60 degree Delevan spray nozzle and using water as the fluid. One hole was plugged and the flow rate was reduced to create the same velocity (82 m/s) through each hole as per the unplugged case. The blockage of one hole has a significant effect on the atomization process as shown by Figure 12. The plugged nozzle has little effect on the atomization process, while the unplugged nozzle forces the liquid sheet out and forms a much better atomized spray. It is conjectured that the plugged hole allows external air to entrain into the central core, eliminating the intense recirculation zone.

The other spray nozzle configurations, Delevan peanut nozzles with 60 and 70 degrees were also characterized. They showed the same trends as discussed for the 90 degree nozzle. Also, a 0.375 inch diameter jet

of air at 3 m/s was directed near the nozzle exit normal to the spray using S12A0. No significant effects on the spray pattern were observed.

Combustor Study

A preliminary study on the use of the multiple air jet and spray nozzle as a side jet liner in a single dome combustor was initiated. The S.annular, S12A0, S16A0, and S28A0 nozzles showed promising results relative to lean blow out and combustor stability. The lean blowout was decreased by 50% using the S12A0 nozzle with JP8+100 fuel. The side liner fuel and air injection showed a blue flame starting near the multiple jet exit. The most intense flame zone occurred near the middle of the opposed side injectors. For S16A0 and S28A0 nozzles, the intense central flame structure would move toward the top air and fuel injector. This is due to the imbalance of the opposed jet where the bottom 0.5 inch hole jet delivers more air at the same pressure drop as shown by Figure 7. The higher momentum air jet pushes the flame region slightly above the midpoint between the opposed injector region. The opposed jet configuration combined with the main dome cross flow combustion generated vortices between the side and main injection points as shown in Figure 8. The recirculation zones presented a stable burning environment, a light blue, low luminosity flame zone.

The S12A25-3S10 did not perform well due to poor atomization characteristics, and the S16A30 and S16A15S15 were not tested. The combustor tests also demonstrated the importance of using a gasket between the peanut nozzle housing and the multiple jet disk as shown in Figure 2. The atomization characteristics become poor if an air leak exists at the center hole, between the spray nozzle exit and the multiple air jet disk. This phenomena was also duplicated in the lab, the gasket is presently the best solution for this problem. It is unclear why the nozzle is so sensitive to such a small leak, more work is required to address this issue. The combustor tests were documented using two video cameras and VCRs.

Conclusion

The use of multiple air jets to improve mixing and fuel atomization for combustor applications has been investigated. The nozzles S.annular, S12A0, S16A0, and S28A0 have improved the lean blowout characteristics of a single dome laboratory combustor. These nozzles behave as an air blast atomizer and spray the fuel in a well

atomized, high momentum jet. The most critical and sensitive region in the atomization process for these nozzles is the jet core, the volume in the middle of the merging air jets. This area is sensitive to air leaks by the center liquid spray nozzle, and obstructions to the any of the individual air jets. The use of different peanut nozzles does not appear to effect the spray pattern significantly. The atomization characteristics of the nozzles in the laboratory combustor and cold flow tests using stoddard fluid were consistent. This may provide a simple test for new spray nozzle designs.

An effort to characterize the mixing of the multiple jets with its surroundings was attempted. The flow visualizations could not distinguish the individual jet structures or the core recirculation zone, however, they did provide information relative to overall jet merging and spreading. The acetone based LIF visualization has the best potential to access the flow characteristics. This is due to the added information of the horizontal plane, where the seeded visualization techniques are restricted to the vertical plane.

References

Grandmaison, E.W., and Zettler, N.L. (1989). "Turbulent Mixing in Coflowing Plane Jets," The Canadian Journal of Chemical Engineering, Vol. 67, pp. 889-897.

Krothapalli, A.D. (1979). "An Experimental Study of Multiple Jet Mixing," Ph.D. Thesis, Stanford University, Stanford, CA.

Moustafa, G.H., and Rathakrishnan E. (1993). "Studies on the Flowfield of Multijet with Square Configuration," *AIAA Journal*, Vol. 31, No. 7, pp. 1189-1190.

Ng, S., and Grandmaison, E.W. (1989). "Mixing Indices in Coflowing Plane Jets: High Momentum Conditions," The Canadian Journal of Chemical Engineering, Vol. 67, pp. 898-905.

Figure 1. Multiple jet configurations

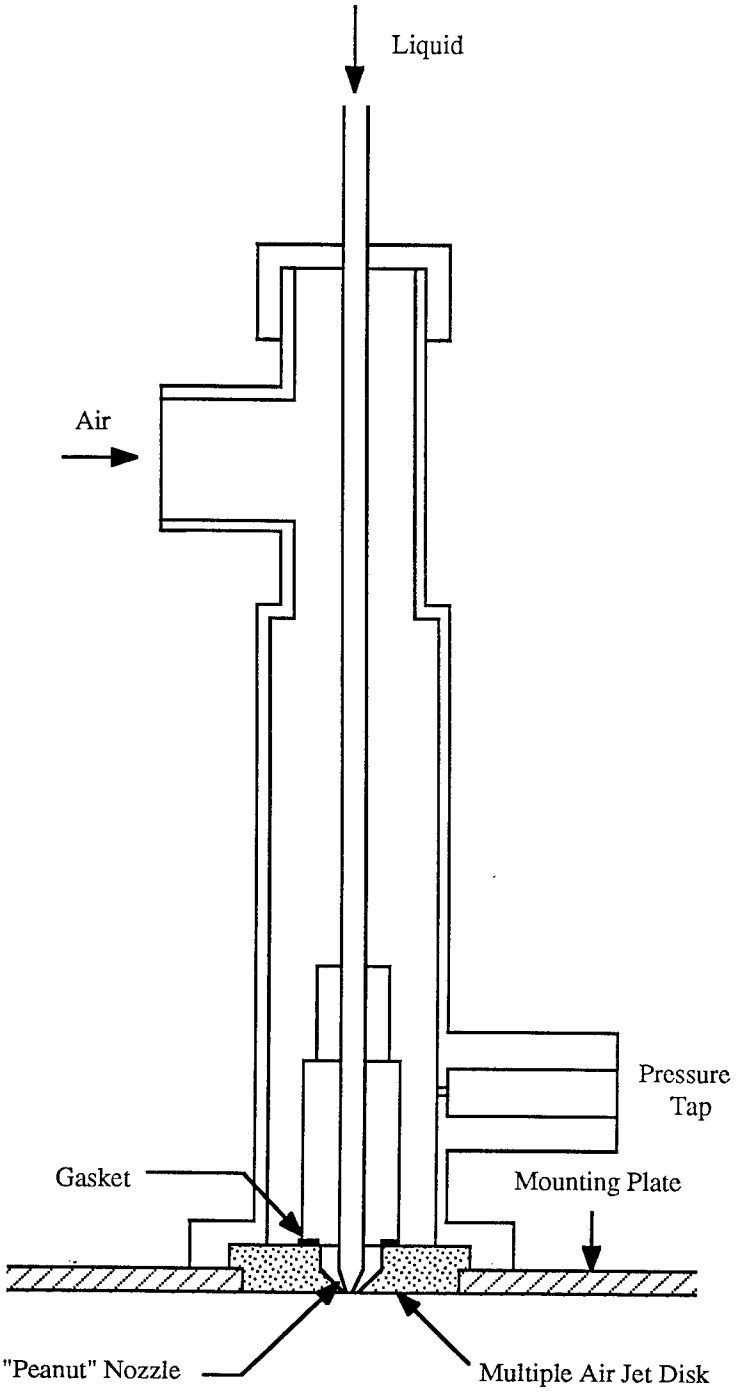


Figure 2 Nozzle assembly

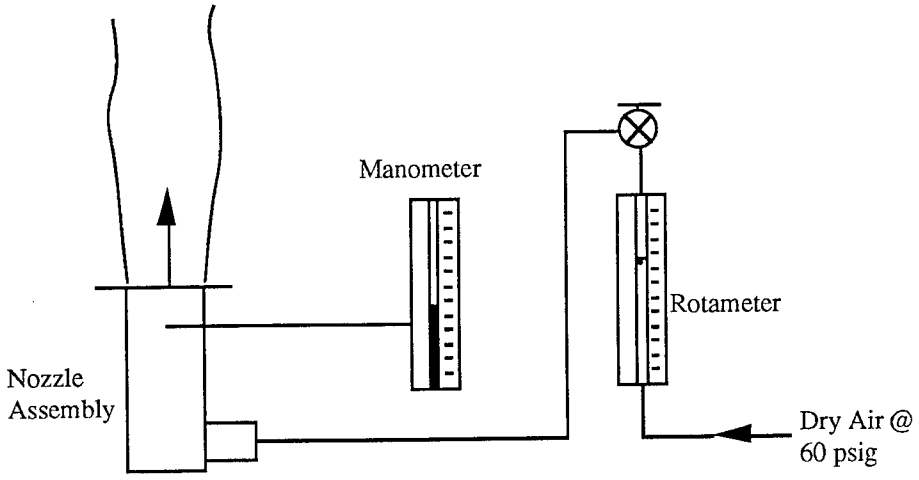


Figure 3. Pressure measurement schematic

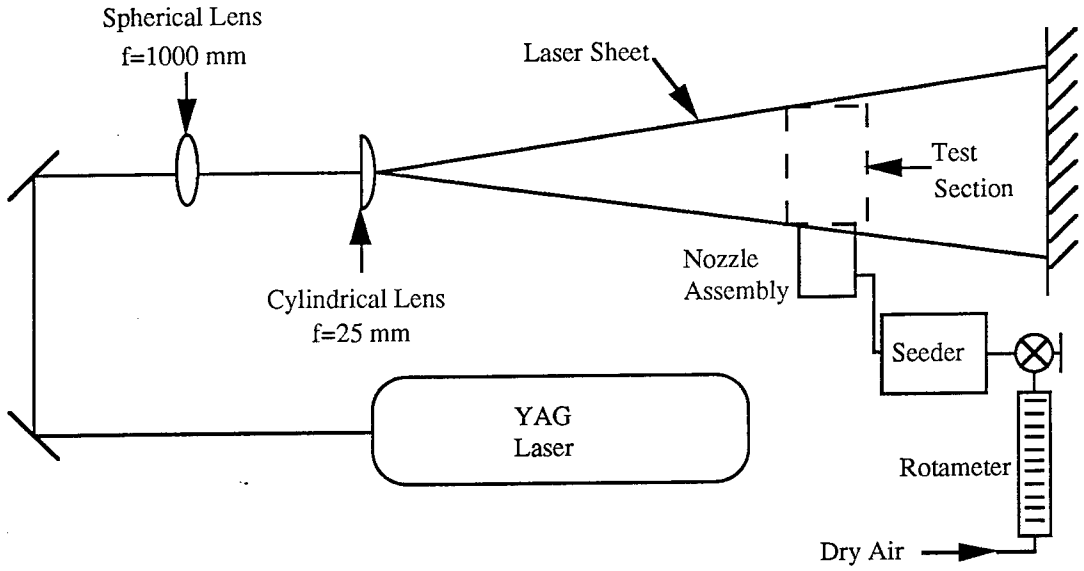


Figure 4. RMS and Al_2O_3 test apparatus

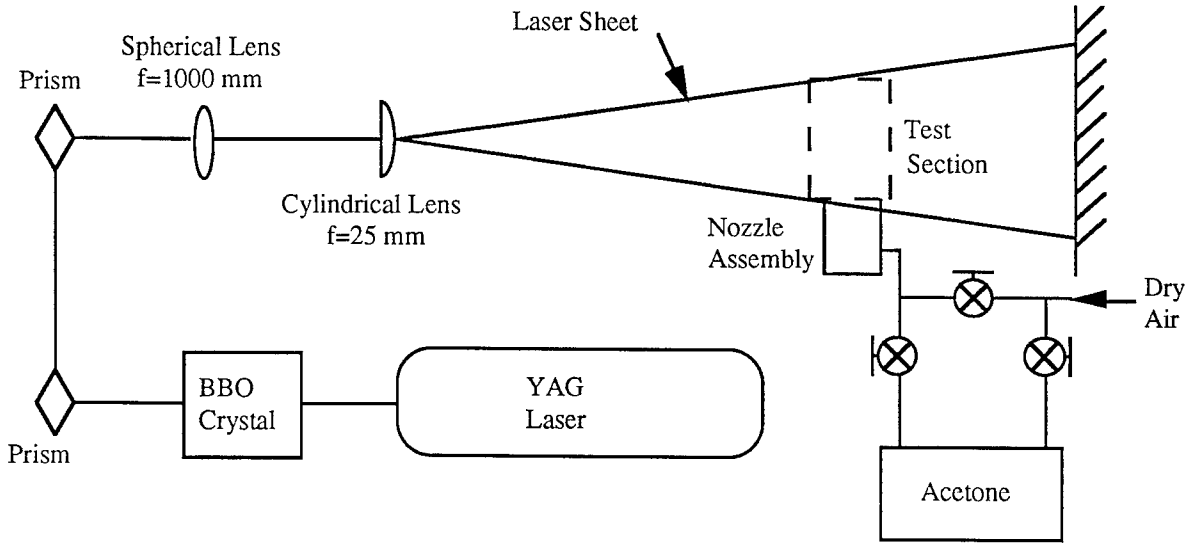


Figure 5. Acetone based LIF

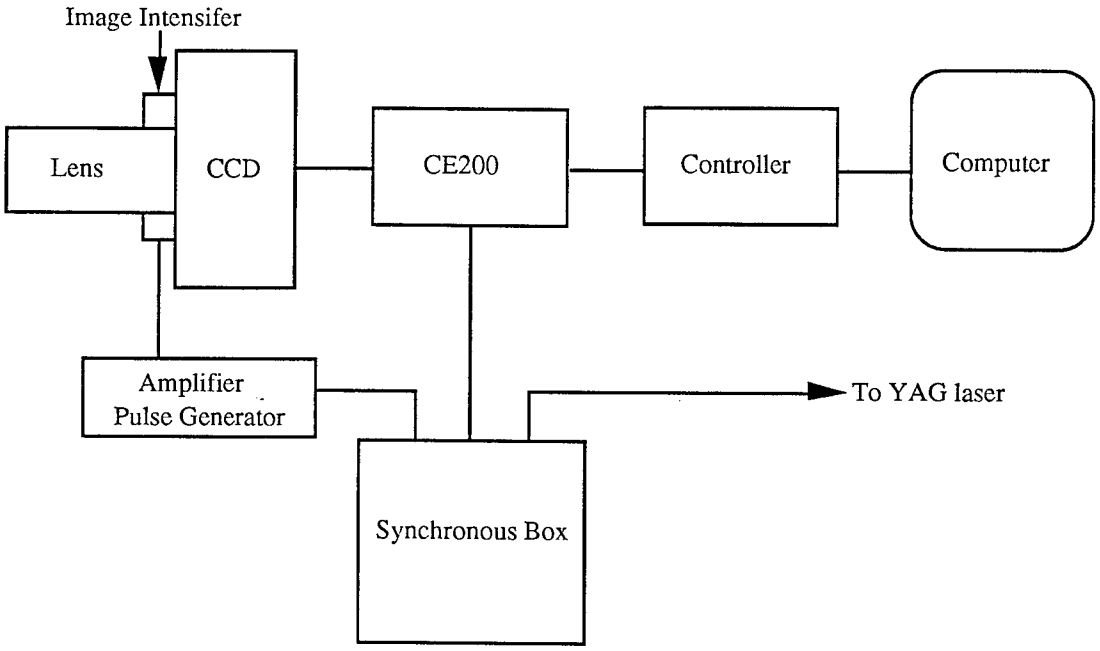


Figure 6. CCD image system

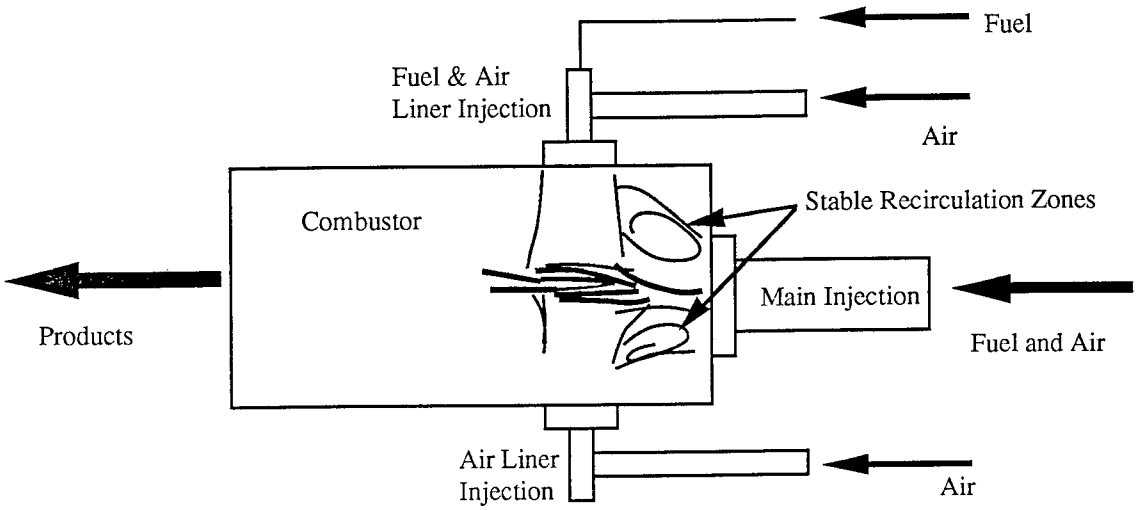


Figure 7. Combustor Schematic

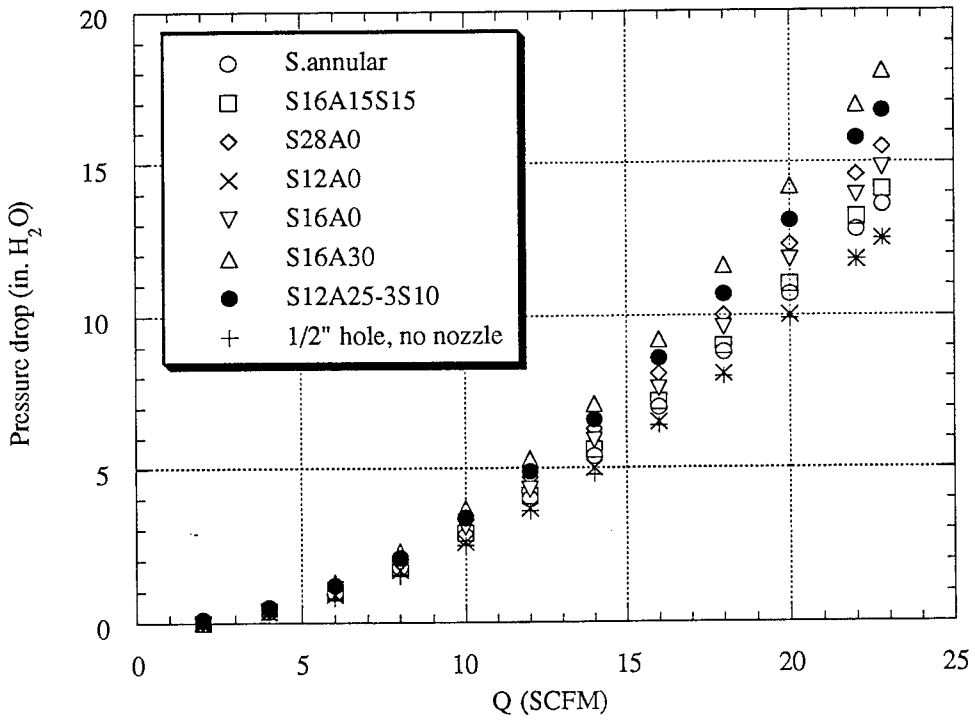


Figure 8. Pressure drop across nozzles

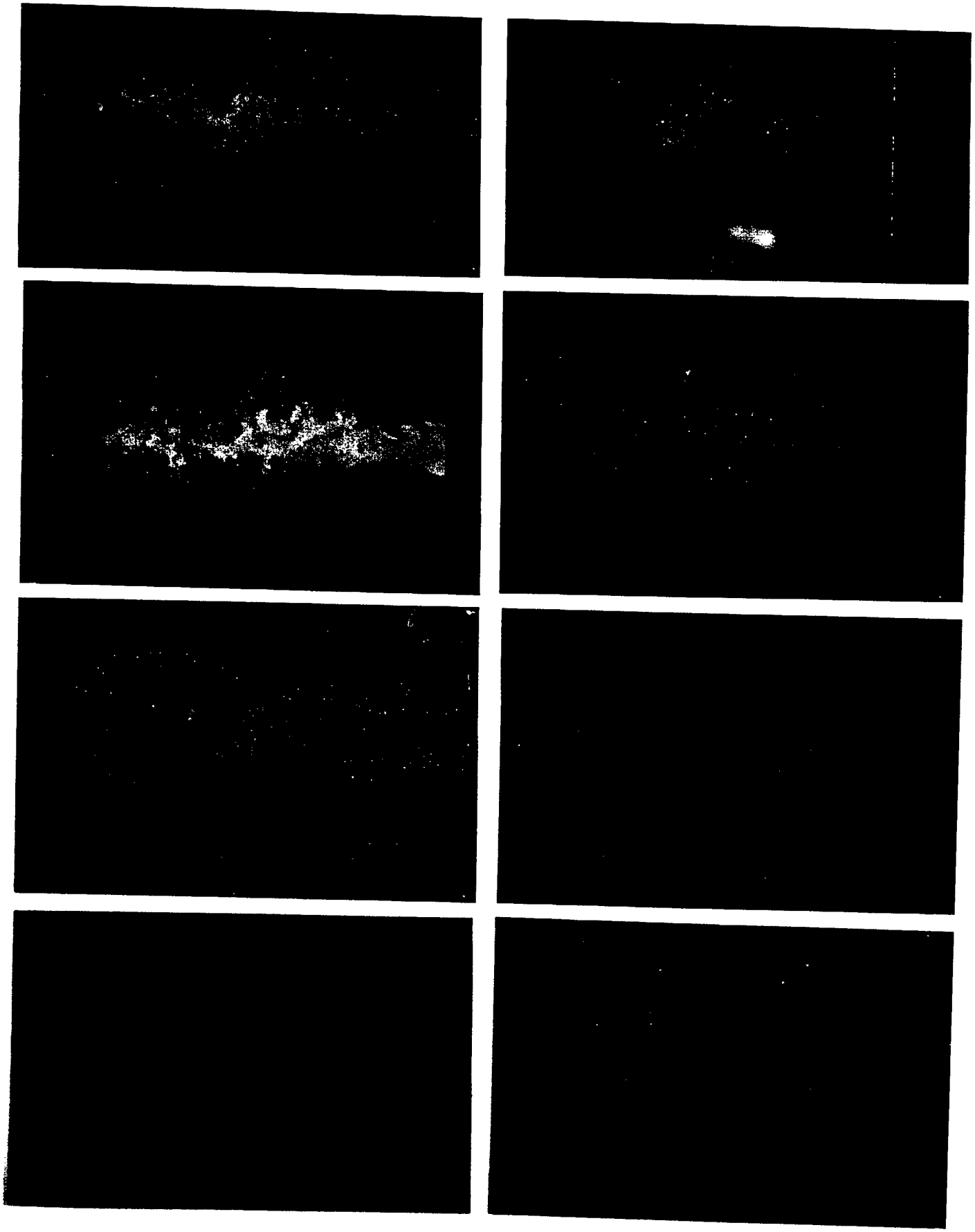
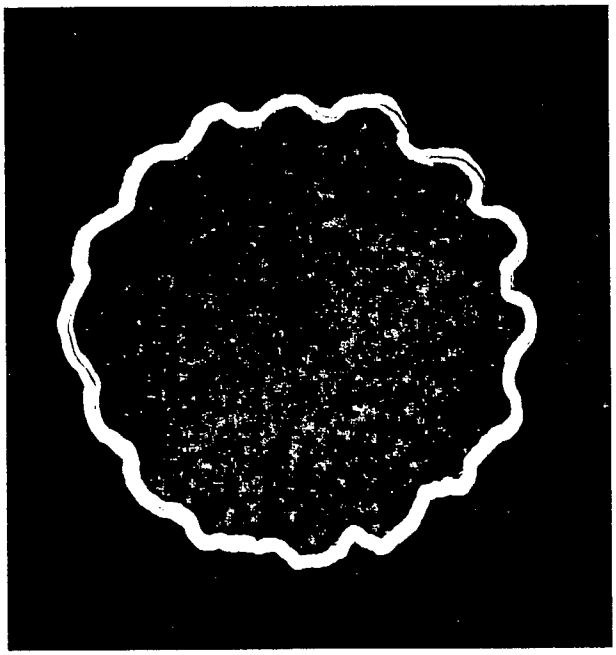
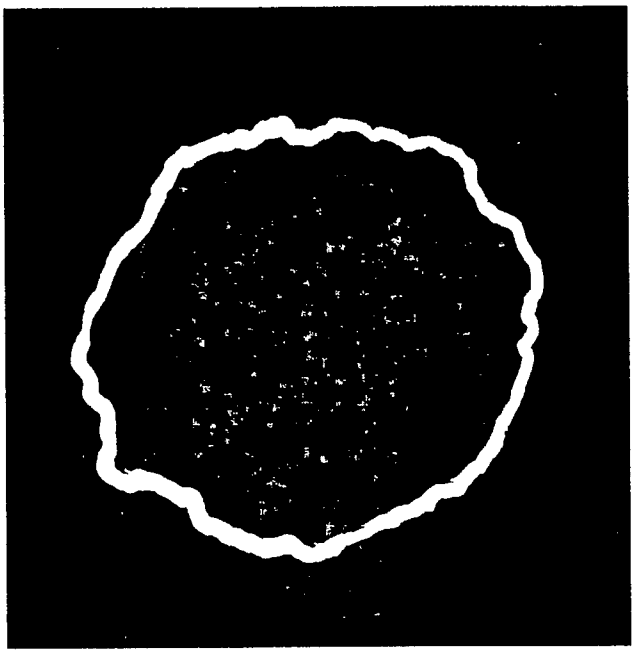


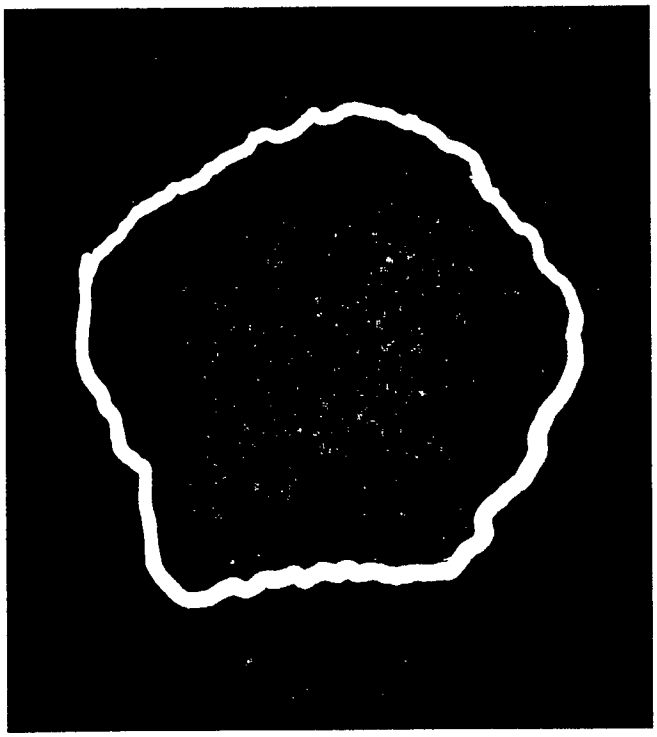
Figure 9. TiCl4 visualizations



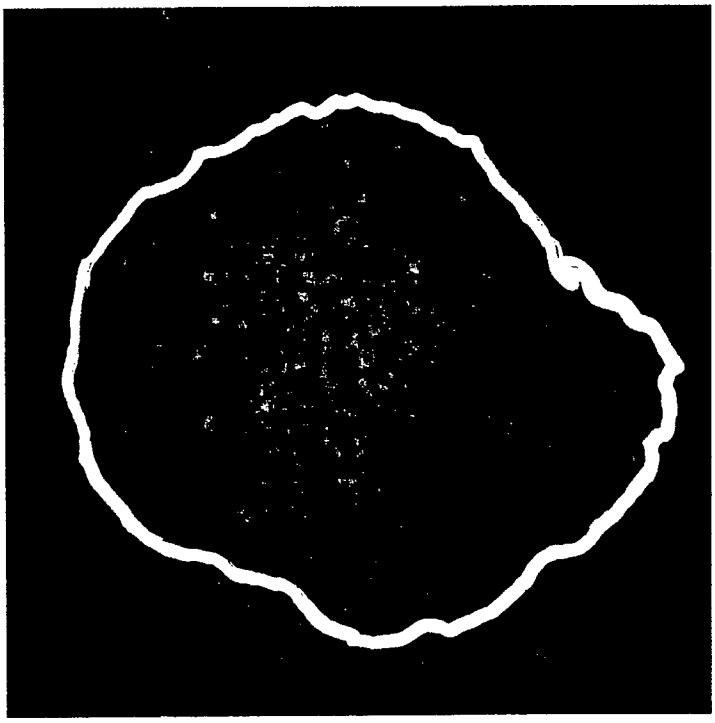
a. S16A0 at z=0.5 mm Horizontal plane



b. S16A0 at z=10 mm Horizontal plane



c. S16A0 at z=20 mm Horizontal plane



d. S16A0 at z=30 mm Horizontal plane

Figure 10. LIF visualizations

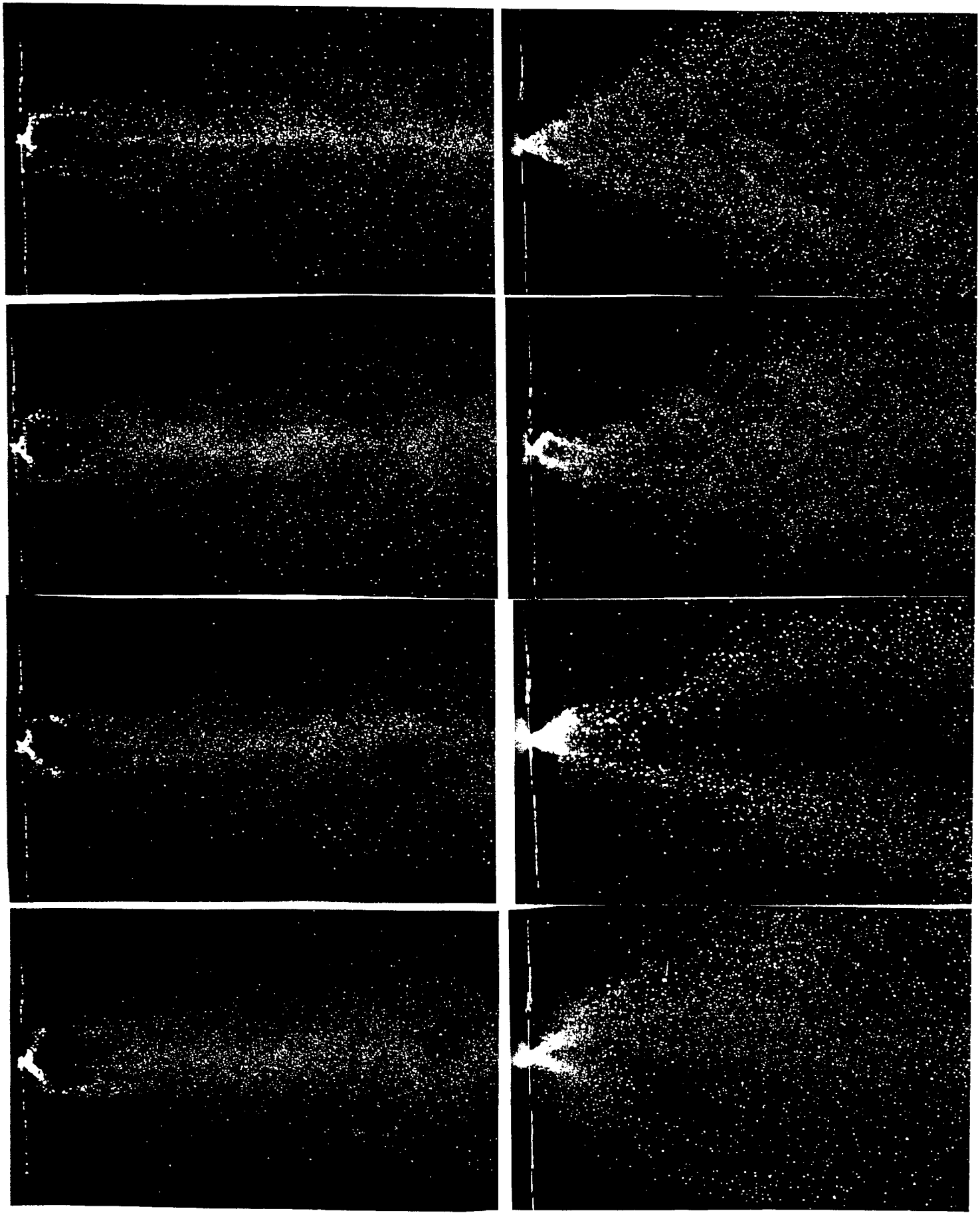


Figure 11. Stoddard fluid, 90 degree spray

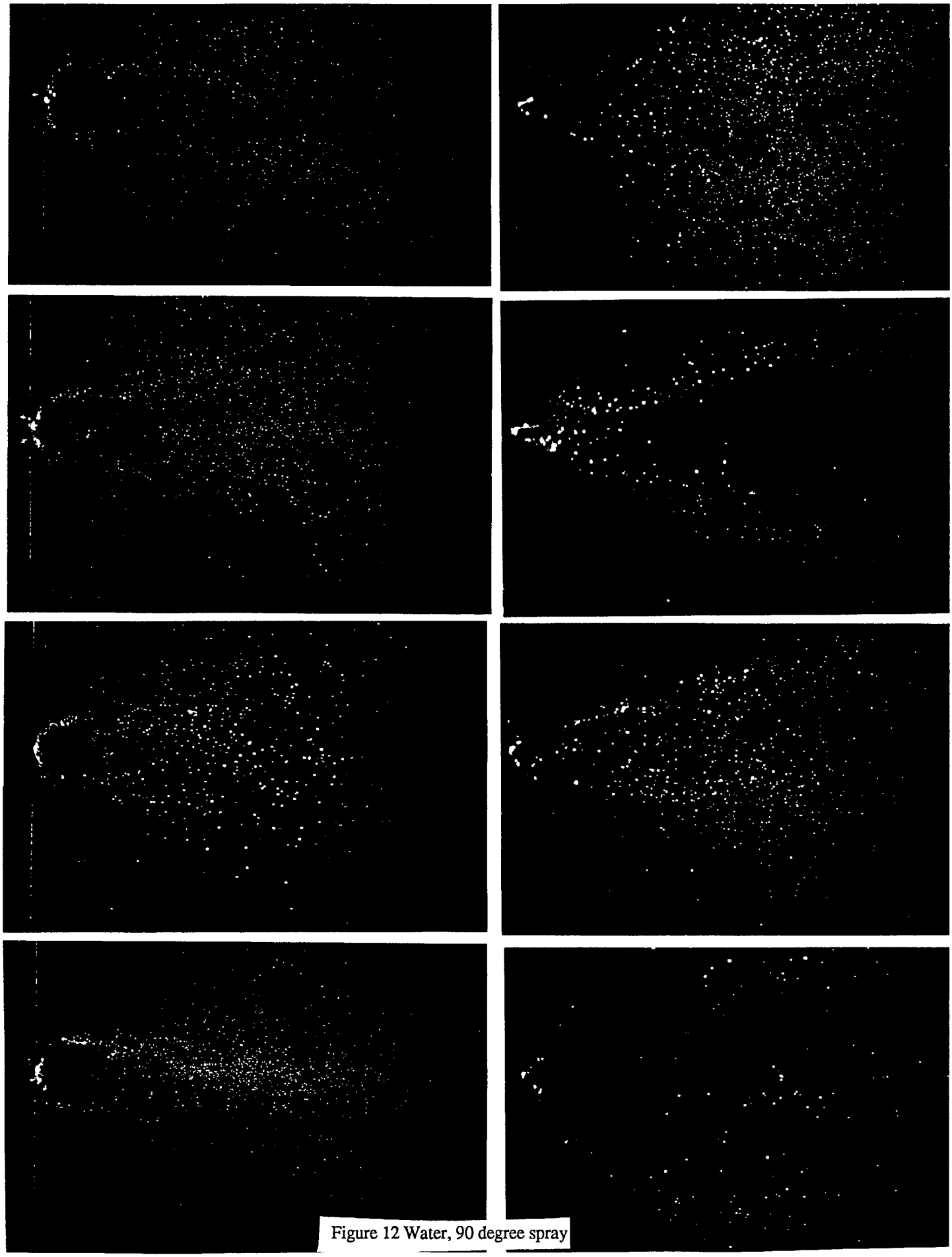
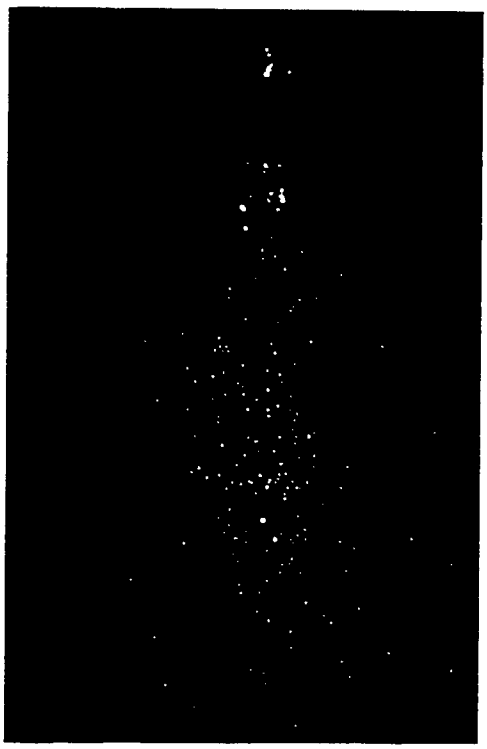
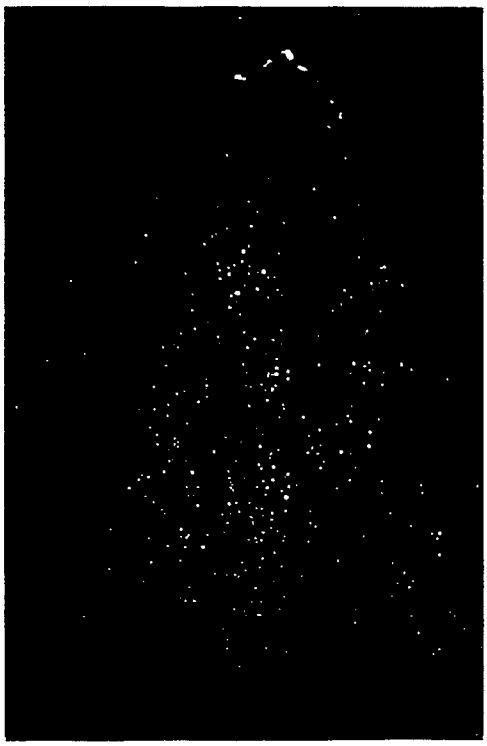


Figure 12 Water, 90 degree spray



One jet blocked



All jets unblocked (normal mode)

Figure 13 S12A0 using water spray, partially blocked and unblocked

HRR Radar Based Target Identification

S. Simanapalli
Assistant Professor
Electrical Engineering Department

University of North Carolina, Charlotte, NC 28223

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

HRR Radar Based Target Identification

S. Simanapalli
Assistant Professor
Electrical Engineering Department
University of North Carolina, Charlotte, NC 28223
P: (704) 547-4828; F: (704) 547-2352; email: siman@uncc.edu

Abstract

A high range resolution radar sensor is under investigation for an automatic target identification system. In this report, a model for the return signal is formulated and applied to stretch-radar processing. For point scatterers this leads to a sum of sinusoids plus noise model and so spectral estimation techniques may be employed to extract low dimension feature vectors that can be subsequently applied to a feature classifier. An approach to the optimal design of classifier is outlined and a procedure for template generation is proposed. Experimental results on characterizing the radar signatures using spectral estimation techniques is presented. Linear and quadratic classifiers using range profiles as the feature vectors are also implemented.

HRR Radar Based Target Identification

S. Simanapalli

Introduction

The problem of target identification by active radar has been of interest for sometime now [1], [2]. The overall process of identification involves the stages of target detection, feature extraction and target classification.

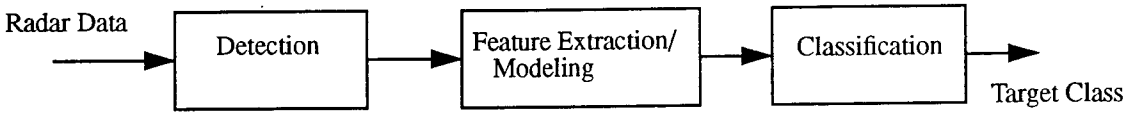


Fig. 1. Target Identification Process

The basic techniques are either two dimensional (2D) based, e.g. Synthetic Aperture Radar (SAR) and Inverse SAR (ISAR) imagery, or one dimensional (1D) based. 2D approaches require longer processing time and visual inspection during target classification. 1D approaches include the natural mode response method (also termed singularity expansion method) [3] and the high range resolution (HRR) method. In the former, short pulses are transmitted and the late-time portion of the received waveforms is processed for natural modes of a target. Although this method is aspect independent, it may be of limited practicality in the field [4]. In HRR, the waveform has sufficient bandwidth to achieve range resolution of orders smaller than the target length. Due to scattering centers of the target which dominate the scattering in a high frequency sense, HRR returns contain target specific information.

Next, a mathematical model for the return signal is formulated. This is followed by experimental results on modeling HRR radar signatures. Subsequently, classifier design and template generation is discussed, and experimental results on linear and quadratic classifiers are presented.

Mathematical Model for Return Signal

From a systems theory perspective, the received/return signal is given by

$$r(t) = h_R(t) * h_B(t) * h(t) * h_F(t) * h_T(t) * s(t) \tag{1}$$

where $s(t)$ is the transmitted signal, * denotes convolution, and $h_{(.)}(t)$ are the various impulse response terms that are explained below. $h_T(t)$ is due to the transmitter, and assuming an ideal transmitter over the operating bandwidth

$$h_T(t) = \alpha_T \delta(t) \tag{2}$$

where α_T is a scaling constant and $\delta(t)$ is the delta function. $h_F(t)$ is the channel impulse response on forward propagation from transmitter to the target. For a direct line of sight path and time invariant channel, it corresponds to a delay and so

$$h_F(t) = \frac{\alpha_F}{R}\delta\left(t-\frac{R}{c}\right) \tag{3}$$

where R is the range from the transmitter to the front tip of target and c is the velocity of propagation. For a non-LOS propagation, multipath and clutter should be taken into consideration. $h(t)$ is the target impulse response and it describes the scattering from the target. Different models have been proposed like the singularity-expansion method, general cone scattering, high-frequency method, linear array scattering, and scattering-center forms [1]. Our interest here is in the last form which is a high-frequency/short-time model and is orientation/aspect dependent. Invoking the scattering center representation [2] wherein the target is modeled as consisting of scatterers

$$h(t) = \sum_i c_i(t) \tag{4}$$

where $c_i(t)$ is the ith scattering center response to an impulse. Although an all inclusive study on scattering behavior from an electromagnetic theory perspective is yet to be reported in literature [5], point-scatterer approximations have been found to be reasonable. For point scatterer

$$c_i(t) = \alpha_i\delta(t-2R_i'/c) \tag{5}$$

where R_i' is its range from the front tip and α_i is the strength/reflectivity. $h_B(t)$ in (1) is due to the backward propagation and has similar expression as (3) provided the transmitting and receiving antennas are collated which results in monostatic path [6]. Finally, $h_R(t)$ is the receiver impulse response and is expressed as in (2). It now follows that after demodulating with a signal $d(t)$, the demodulated signal is given by

$$x(t) = r(t)d(t) = \left[\frac{\alpha}{R^2}h\left(t-2\frac{R}{c}\right)*s(t)\right]d(t)+n(t) \tag{6}$$

where $\alpha=\alpha_T\alpha_F\alpha_B\alpha_R$ and the additive term $n(t)$ includes noise due to receiver and modeling errors [20]. Expression (6) is generic and involves the transmitted signal, target scattering and demodulating signal. The design of a suitable transmitted signal $s(t)$ to enhance SNR is of interest in itself and the details may be found in relevant texts [7], [8]. For HRR, pulse compression based on Linear Frequency Modulation is a common method. We next discuss expression (6) for a specific case of LFM referred to as stretch radar.

Stretch Radar: Here the transmitted signal is a Linear Frequency Modulated pulse (chirp) and the radar applies a LFM ramp for demodulation [9]. Using analytic signal form [10], [11]

$$s(t) = e^{j(\omega_0 + \mu t)t}, \quad 0 \le t \le T \quad ; \qquad d(t) = e^{-j(\omega_r + \mu(t-\tau_r))(t-\tau_r)} \tag{7}$$

where T is the pulse width, μ is the frequency slope, μT is the chirp bandwidth, and the reference frequency $\omega_r > \omega_0$ ($<\omega_0$) when $\mu > 0$ (<0). For point scatterers, using (4), (5) and (7), in (6) yields

$$x(t) = \sum_i \beta_i s(t-\tau_i)\, d(t) + n(t) = \sum_i \gamma_i e^{-j(\omega_r - \omega_0 + 2\mu(\tau_i - \tau_r))t} + n(t) \tag{8}$$

where $\beta_i = \alpha\alpha_i / R^2$, $\tau_i = 2(R + R_i')/c$, $\gamma_i = \beta_i e^{j\left((\omega_r\tau_r - \omega_0\tau_i) - \mu\left(\tau_r^2 - \tau_i^2\right)\right)}$. Since $\tau_r < \tau_i$ as the demodulator is on before the return signal arrives, the bracketed radian frequency in the exponent term above is >0 (<0) when $\mu > 0$ (<0). Rewriting the equation,

$$x(t) = \sum_i \gamma_i e^{-j(\omega_d + \omega_i)t} + n(t) \tag{9}$$

where the frequency offset $\omega_d = (\omega_r - \omega_0) - 2\mu\tau_r$ and, the frequency $\omega_i = 2\mu\tau_i = 4\mu R_i / c$ is a scaling of the range $R_i = R + R_i'$ of the ith scatterer from the receiver. From (8) and (9), it is seen that the demodulation converts time of arrival (needed to measure range) into a frequency translation. Therefore scatterers at different ranges generate return pulses of different frequencies.

The demodulated signal (9) is a sum of complex sinusoids in noise model. Any of the classical or high resolution spectral estimation techniques, including FFT may be used to determine the frequencies and hence ranges, and their amplitudes. These characteristics then form the feature vector for target classification.

Experimental HRR Radar System

A HRR radar sensor is under investigation for automatic target identification. Details of the system and the processing can be found in [12] (although the measured signature data is unclassified, much information about the data remains classified). The scenario of a flying target/aircraft and a ground radar was a direct line of sight propagation. Fig. 2 shows a possible configuration for the system [6]. The radar used to collect data transmits 64 LFM pulses per dwell and upon receiving each pulse, applies a LFM ramp to demodulate the signal. The data is acquired at a frequency lower than the radar operates (refer to (8)), so coherent sampling (in-phase and quadrature) is used to maintain proper phase relationships. This sampling occurs 820 times per pulse. Each pulse is phase-shifted/motion-compensated to remove the effect of doppler frequency shift caused by the target's radial velocity. Predetermined amplitude and phase weighting coefficients are applied to the 820 x 64 complex array of each dwell to correct transmission/reception LFM non-linearities. These weights reduce the range sidelobes level for each target scatterer. The 64 pulses are suitably summed to improve the SNR and the resultant signature is considered for analysis if the SNR is above 35 dB.

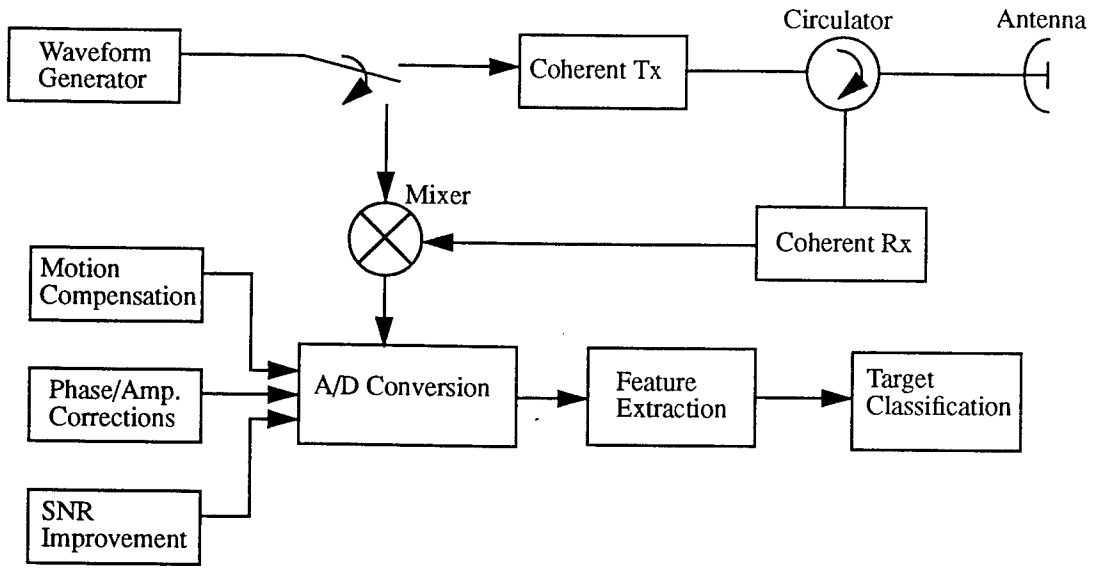


Fig. 2. Simplistic HRR Radar System and Processing.

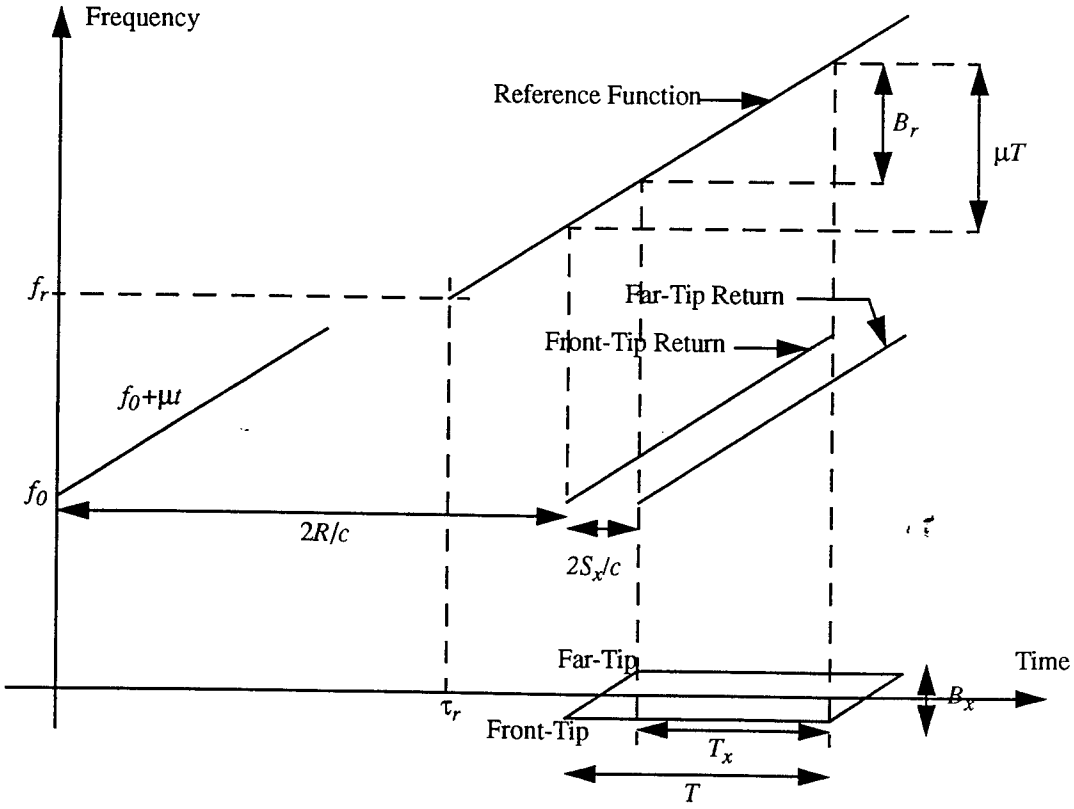


Fig. 3. Frequency vs. Time Characteristics for Stretch Radar Decoding of Chirp

The frequency vs. time characteristics for a single radar pulse is shown in fig. 3. The generation and transmission of the LFM waveform begins at time zero with chirp rate μ. The return signal beginning with the front-tip return and ending with the far-tip is shown occurring with appropriate delay associated with round-trip propagation. Mixing with the reference frequency during demodulation produces a constant frequency signal for each return from a point scatterer. The total set of signals have a bandwidth B_x related to the total range swath width S_x by the scale factor $2\mu/c$ (refer to the definition of ω_i in (9)). Note that the signals from all ranges do not completely overlap in time and so for efficient processing only the central overlapped region T_x is A/D converted. The effective RF bandwidth B_r which determines the range resolution ρ_r is seen to be less than the full transmitted bandwidth μT. ρ_r is given by [8]

$$\rho_r = c/2B_r \doteq c/2\mu T \tag{10}$$

Experimental Results on Modeling

In this section the application of spectral estimation techniques for characterizing HRR radar signatures is presented. Field data of 200 signatures each for three different targets/aircrafts that we denote as A1, A2, A3 was obtained using stretch radar technique. The azimuth and elevation angular sectors covered by these signatures are as follows.

A1:	ϕ=(-15.4° to 0.3°)	θ=(9.7° to 12°)
A2:	ϕ=(-25° to -9°) and (6° to 18.5°)	θ=(10.4° to 15.3°)
A3:	ϕ=(-11° to 2°)	θ=(12° to 14°)

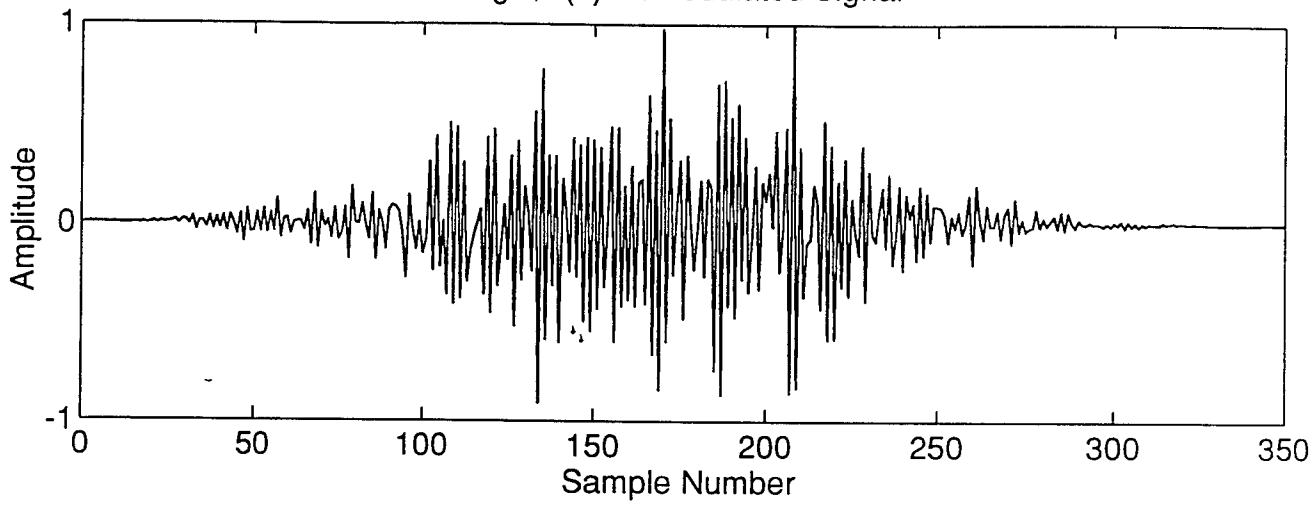
We shall illustrate our results here for an example signature without a loss in generality. Fig. 4a indicates the in-phase component of the demodulated signal $x(t)$ of 350 contiguous samples (the rest of samples from 820 were of negligible values) for A1 at azimuth and elevation angles of $\phi = -15.4°, \theta = 12°$. In order to obtain the ranges and amplitudes of its scatterers, the following techniques [13], [14] were implemented.

(i) *FFT*: Fig. 4b is the magnitude plot of the 820-point fft of the signal $x(t)$, commonly known as the range profile. Referring to (9), the ranges can be obtained from frequencies by applying an offset ω_d and a scale factor inversely proportional to LFM sweep slope μ.

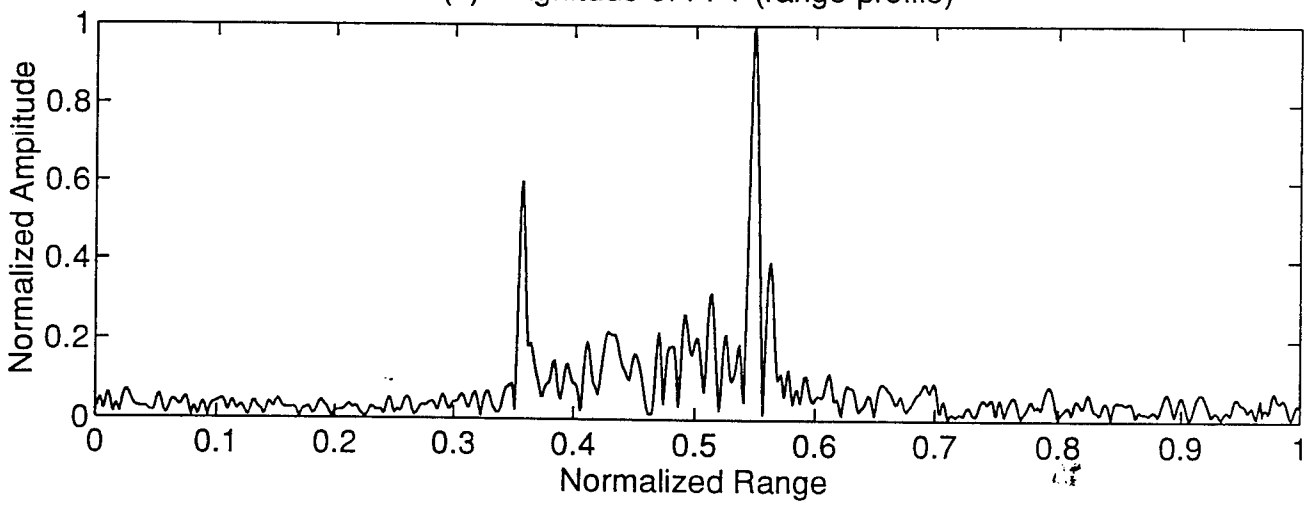
(ii) *Linear Prediction*: A sum of sinusoids plus noise is an ARMA process which is equivalent to an infinite order AR process. Therefore linear prediction methods are applicable. A forward linear prediction method with biased correlation estimates (unbiased estimates did result in lower prediction-error power) is implemented. The prediction order may be determined by Akaike Information Criterion, Final Prediction Error, or Minimum Description Length criteria wherein the p that minimizes the expression

$$AIC(p) = N\ln\sigma_p^2 + 2p; \qquad FPE(p) = \sigma_p^2\left(\frac{N+p+1}{N-p-1}\right); \qquad MDL(p) = N\ln\sigma_p^2 + p\ln N \tag{11}$$

Fig. 4. (a) Demodulated Signal



(b) Magnitude of FFT (range profile)



is chosen as the model order. Here σ_p^2 is the prediction-error power and N is the data length. However, it is known that these criteria tend to underestimate and so an order of $N/3$ has been suggested in the literature [15]. Figs. 5 and 6 depict the criteria functions and the AR amplitude spectra for orders 30 (dotted) and 100 (solid) along with fft (dashed) for comparison purposes using N=350 samples. It is seen that order 30 results in spectral shape similar to fft, but that order 100 almost replicates the fft spectrum with same resolution. The prediction-error power is found to be 19% of the total power for order 30 and 15% for order 100. It will be shown later that the range bins with extremely small amplitudes like the first and the last few samples in fft have negligible effect in target classification as they effectively correspond to noise. In terms of feature vector, prediction of order 100 reduces the dimensionality by eight over fft. Fig. 7 is a plot of the zeros of the prediction-error polynomial. For order 30 there are two distinct zeros close to the unit circle corresponding to the two main peaks in the fft spectrum. For order 100 most of the zeros are close to the unit circle and there appear two zeros in close proximity which result in achieving resolution at the largest peak. It may be noted that for amplitude estimates, the Prony technique is more reliable than AR peaks.

(iii) *Burg Algorithm*: of forward-backward linear prediction using data directly instead of correlations resulted in identical prediction order criteria curves and spectral plots as discussed in (ii) above.

(iv) *Minimum Variance*: The performance of this technique is comparable to fft when the correlation matrix dimension is 100 as can be seen from fig. 8. There appears no advantage in using this technique over fft as the resolution remains same. It may be noted that the peaks in this estimator are approximately linearly proportional to the sinusoid amplitudes.

(v) *MUSIC*: This method is based on an eigenanalysis of the correlation matrix and hence is computationally more involved than linear prediction. However, unlike the latter, it is ideally suited to fit a sum of sinusoids plus noise model. The number of sinusoids is determined by selecting the p corresponding to minimum values of the functions [16]

$$AIC(p) = l_p + p(2P-p); \quad MDL(p) = l_p + \frac{p}{2}(2P-p)\ln(N); \quad l_p = N(P-p)\ln\left(\frac{\frac{1}{P-p}\left(\sum_{i=p+1}^{P}\lambda_i\right)}{\left(\prod_{i=p+1}^{P}\lambda_i\right)^{\frac{1}{(P-p)}}}\right) \tag{12}$$

where λ_i, i=1,....,P are the eigenvalues of the correlation matrix formed using N P-vector samples. Fig. 9 is a plot of the two functions when P=60. This value of P was chosen to minimize numerical problems due to extremely small eigenvalues that influence the geometric mean in the denominator term in l_p. Fig. 10 is the MUSIC spectra for correlation matrices of orders 30 and 100 when using eigenvectors corresponding to the 10 least eigenvalues for noise space calculations (by so doing, no explicit signal space dimensionality is set). Clearly, the significant frequencies in fft are also captured by MUSIC, and the noise range bins are seen to be the first and the last few. It should be

Fig. 5. Prediction Order Criteria Functions

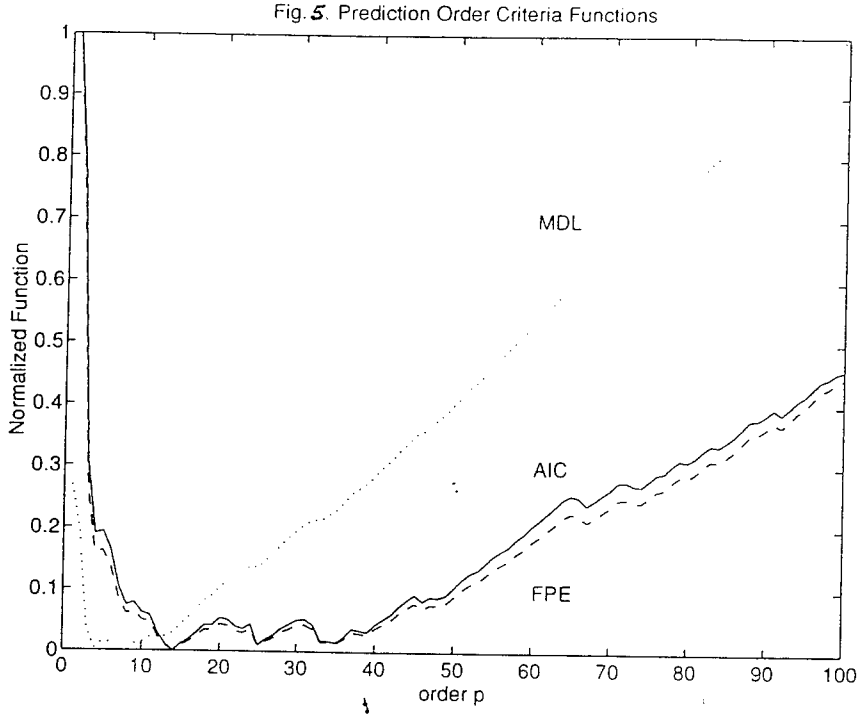


Fig. 6. Spectral Estimate from Linear Prediction

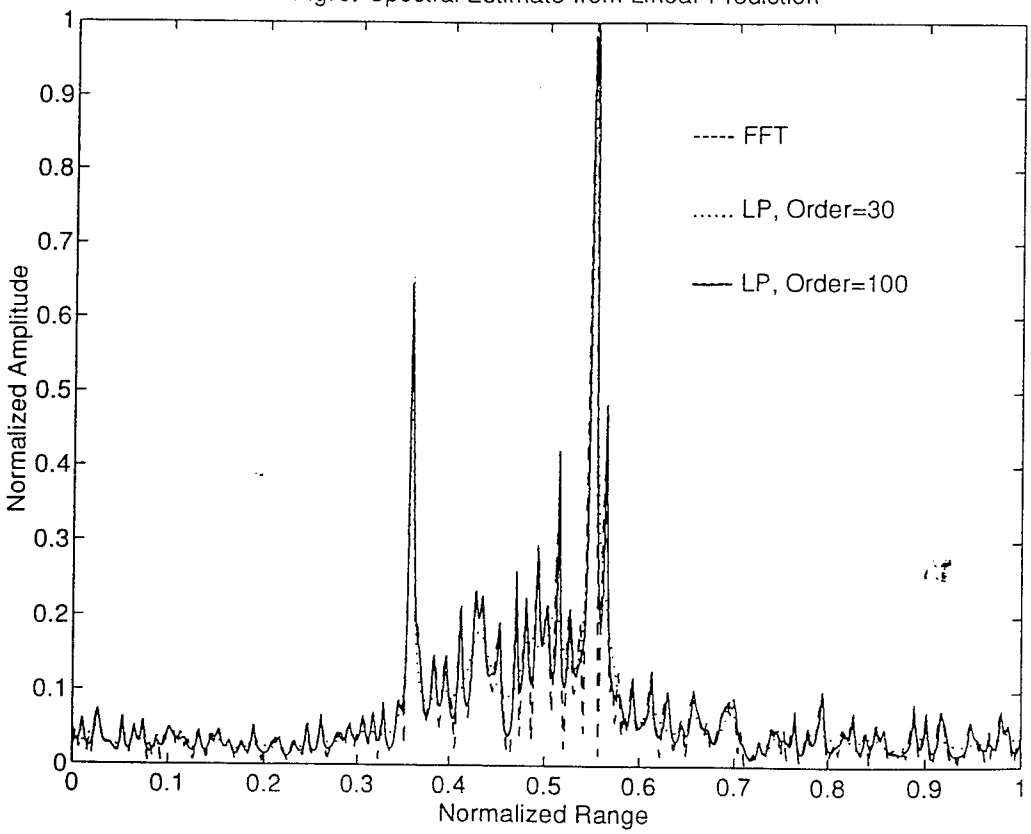


Fig. 7. Location of Zeros of Prediction Polynomial

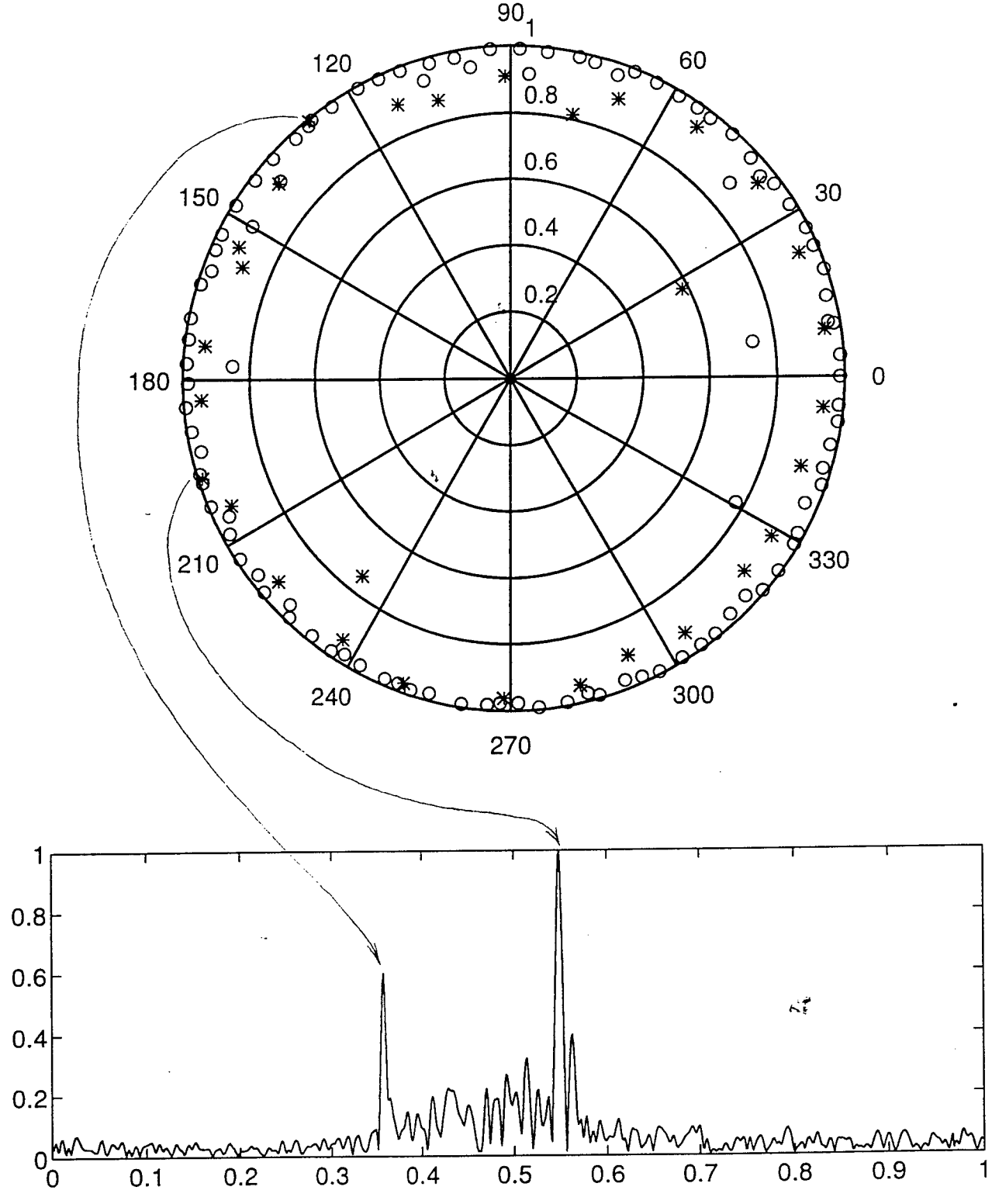


Fig. 8. Spectral Estimate from Minimum Variance Technique

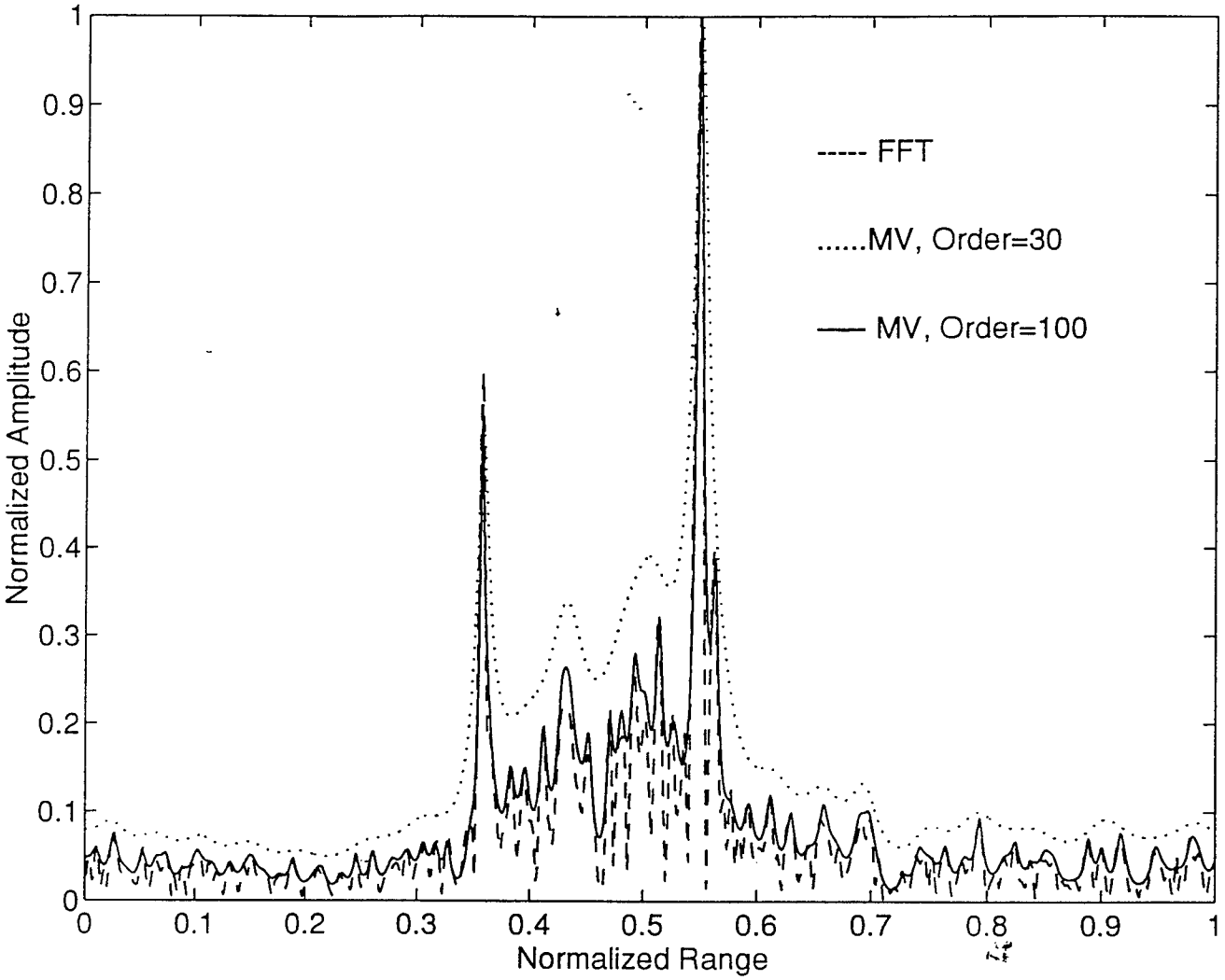


Fig. 9. Order Selection Criteria Functions

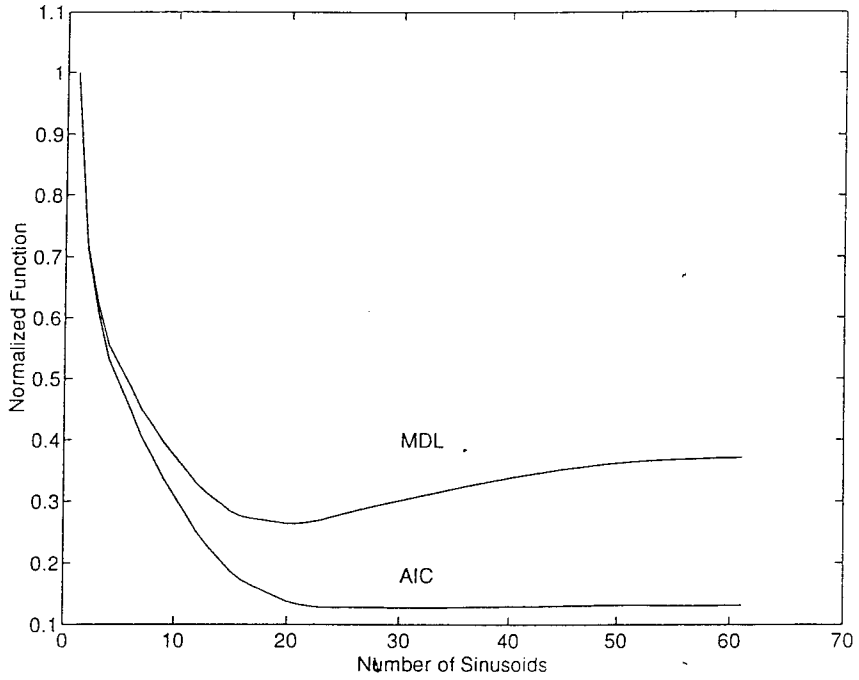
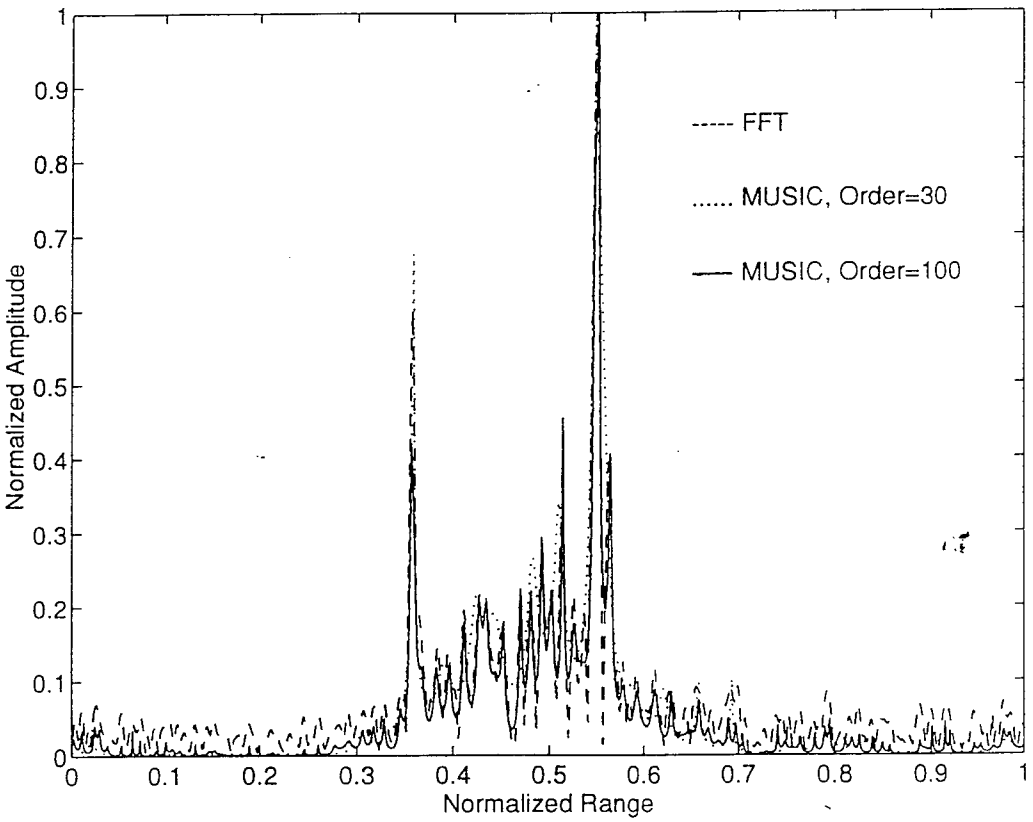


Fig. 10. MUSIC Spectral Estimate



noted that the MUSIC spectrum only gives the frequency locations; calculation of the corresponding amplitudes is a separate step.

(vi) *Min-Norm*: In this method the minimum norm vector in noise subspace is used to obtain either the spectrum or the zeros. The spectrum obtained was similar to that of MUSIC with spurious peaks in the noise range bins.

In summary, the performance of various spectral estimation techniques, in particular linear prediction, is comparable to fft. Although fft is computationally less expensive than the other techniques, parametric modeling may be used to extract suitable low dimension feature vectors for target classification. One such successful application has been reported in an air traffic control environment [17].

Design of Classifier

Given a set of *features* $\boldsymbol{f}$ (vectors are denoted by bold case) extracted from the measured radar data, the task of a classifier is to identify a target type from a library by *matching* the $\boldsymbol{f}$ to the stored *templates*.

Feature Vector: Various choices for feature vectors are the complex-valued fft [18], the magnitude of fft or range profile [19], [20], the squared magnitude of fft or periodogram, peak locations and their amplitudes in the spectrum obtained by fft or any other technique [21], location of zeros of prediction error polynomial close to the unit circle, prediction polynomial, sequence of reflection coefficients in prediction [17], etc. It is not known how the choices compare in terms of performance and much research is yet to be done. For our purposes we chose the range profile.

Template Generation: Since HRR signatures are aspect dependent, templates are required for increments in aspects. Too small an increment will increase storage for data base. It is proposed that the aspect increment be such that the range motion of the scattering centers not exceed a resolution cell ρ_r. Correspondingly, if the length of target is L, then the angular increments in azimuth and elevation should satisfy [6]

$$\Delta\phi < \rho_r / (L/2) \quad ; \quad \Delta\theta < \rho_r / (L/2) \tag{13}$$

where the origin of the coordinate system is assumed at the target midpoint. Suppose that M training signatures $\boldsymbol{g}_1,\ldots,\boldsymbol{g}_M$ over a sector less than $(\Delta\phi, \Delta\theta)$ are available. (Our signatures refer to the range profiles.) Three concerns are noteworthy.

(i) In order to minimize the effect of considerable variability in magnitude due to nonconstant gain adjustment in the different signatures, each signature is normalized by the square-root of its power to have a variance of one.

(ii) It is assumed that these signatures do not contain flashes (large gains due to reflectivity from target antenna). The presence of flashes should be treated separately by forming alternate templates.

(iii) The different signatures may not be aligned range binwise to begin with.

The averaging of the signatures to form a template $\boldsymbol{g}$ proceeds as follows.

1) Initialize template with the first signature, i.e $\boldsymbol{g}=\boldsymbol{g}_1$.

2) For $m=2,\ldots,M$

a) find alignment/shift l which minimizes the mean square error between g and $g_{m,l}$ [22]. However mmse is equivalent to maximizing the correlation since

$$\min_{l} (g_{m,l}-g)^T (g_{m,l}-g) = \min_{l} \left(g_{m,l}^T g_{m,l} + g^T g - 2 g_{m,l}^T g \right) \Rightarrow \max_{l} g_{m,l}^T g \tag{14}$$

Here superscript T denotes transpose operation. Compared to mmse, correlations are computationally simpler and may also be obtained by fft. If the correlation exceeds a set value (say, 0.85), then the aligned signature is averaged with g, i.e. $g=(g+g_{m,l})/2$. If not, g_m is discarded.

3) g is now the template mean. The covariance matrix is obtained as

$$\Sigma = \frac{1}{M-1} \sum_{m=1}^{M} (g_{m,l}-g)\,(g_{m,l}-g)^T \tag{15}$$

Matching: The generic approach is a probabilistic analysis based on hypothesis testing. The two decision criteria of most interest are Bayes and Neyman-Pearson. The former provides optimum decision rule when statistics of the problem are known. Without delving into the details, if we assume that all target classes/hypothesis are equally likely to occur and that all misclassifications are equally costly, the classification rule assigns feature vector f to class c_j if [23], [24]

$$p(f/c_j) > p(f/c_i), \qquad \text{all } i \neq j \tag{16}$$

where $p(.)$ is a probability density function. Statistically, the Fourier transform $X(k)$ of a sampled signal $x(n)$, $n=1,...,N$ are asymptotically $(N \to \infty)$ independent over k and follow a complex normal distribution [25] $\mathcal{N}(.,.)$ as

$$X(k) \sim \mathcal{N}\left(0, \sigma_k^2\right), k \neq 0; \qquad X(0) \sim \mathcal{N}\left(m, \sigma_0^2\right), m = E[x(n)] \tag{17}$$

where E denotes expectation operator and σ_k^2 is the variance at fft bin k. (It is assumed that the process is wide-sense stationary.) The magnitude of fft will then follow a Rayleigh distribution [26]. The vector formulation of these distributions and the discriminant function resulting from Bayes rule (16) is a topic for future research. It may be noted that most theoretical developments of classifiers assume multivariate normal distribution and to use this theory, the Rayleigh variable should be transformed to a normal by a suitable power transformation [23].

In this report we explore linear and quadratic classifiers. Linear classifiers are the simplest as far as implementation is concerned. Although in the Bayes sense, they are optimum only for certain distributions under restricted conditions e.g. normal distributions with equal covariance matrices [23], [27], their simplicity and robustness more than compensate for the loss in performance in many cases. The linear discriminant function is

$$\text{If} \quad V^T f > v_0 \quad \text{then} \quad c_j \tag{18}$$

The optimum V and v_0 may be obtained by an iterative procedure as described in [23]. For simplicity, we use the mean template for the j*th* class g_j for V. (18) now refers to the correlation coefficient between g_j and f (recall that the range profiles are normalized to unit variance), and f must be aligned for maximum correlation. The value of v_0 is obtained during the training phase of the template generation for different classes. A more general form can be taken as

$$\text{If} \quad g_j^T \Sigma^{-1} f > v_0 \quad \text{then} \quad c_j \tag{19}$$

However, in practice one has to be cautious about rank deficiency of Σ (sufficient training samples may not be available to evaluate Σ) and the noise range bins (for which $\sigma_k^2 \cong 0$). It is customary to only consider the diagonal elements of Σ in analysis.

For normal distributions, the Bayes discriminant function is quadratic. Even if non-normal, the quadratic classifier is popular as it works well for many applications. However, it tends to have significantly larger bias than linear classifier particularly when the number of design samples is small. The quadratic discriminant function is expressed as

$$\text{If} \quad f^T W f + V^T f > v_0 \quad \text{then} \quad c_j \tag{20}$$

For simplicity, a form used is

$$\text{If} \quad (f - g_j)^T \Sigma^{-1} (f - g_j) > v_0 \quad \text{then} \quad c_j \tag{21}$$

Since f is normalized to unit variance and

$$lhs\,(21) = f^T \Sigma^{-1} f + g_j^T \Sigma^{-1} g_j - 2 g_j^T \Sigma^{-1} f = f^T \Sigma^{-1} f + constant - 2\,lhs\,(19) \tag{22}$$

(19) and (21) are equivalent in terms of performance when Σ is a scaled identity matrix.

Experimental Results on Classification

With the available data, first a template for A1 covering an azimuth of 5°, which corresponds to the first 70 signatures of the 200, was generated by the averaging procedure described earlier. Fig. 11 indicates the mean and the variance (diagonal elements of the covariance matrix) of the template. Although the range profile is of length 820, neglecting the first and the last 100 or 200 samples still resulted in identical templates over the mid 400 range bins which contain the significant peaks. Hence the first and last few samples correspond to noise range bins.

Fig. 11a. Template Mean

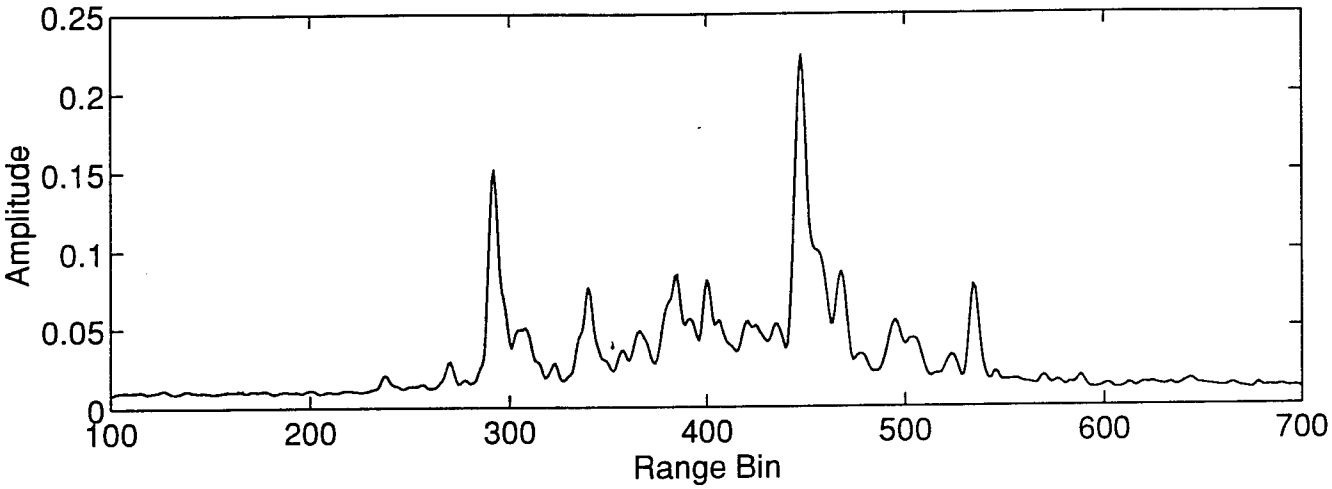


Fig. 11b. Template Variance

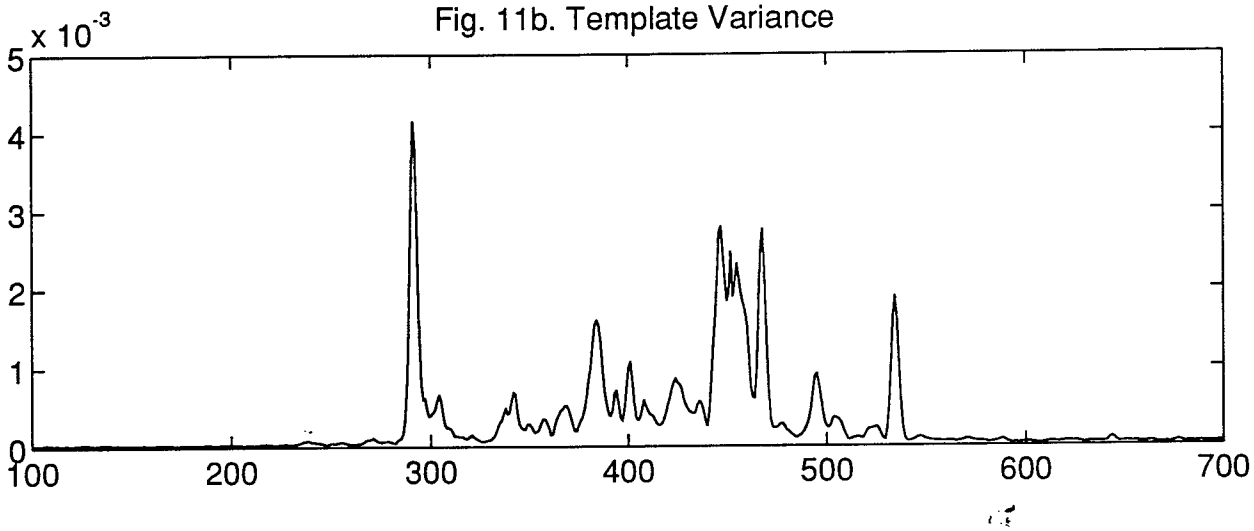


Fig. 12 is a plot of the correlation coefficient as in (19) with $\Sigma=I$, the identity matrix, for 200 signatures each of A1, A2, A3. Since the template was generated for A1, the high correlation with A1 especially for the first 70 signatures is not surprising. The correlations with A2 and A3 are relatively lower. Fig. 13 is a plot of the mse as in (21).

Final Remarks

From the experimental study conducted the following remarks are in order.

1) The resolution provided by the various spectral estimation techniques is at best that of fft. This may be due to the transmitted chirp signal having enough bandwidth to resolve the different scatterers.

2) The performance of various spectral techniques is comparable. This may be attributed to the high enough SNR in the received waveforms.

3) The choice of a technique and a feature vector will ultimately depend on the performance of the classifier. Complete studies are yet to be conducted.

4) Minimizing the mse to obtain translation of training signatures in template generation is equivalent to maximizing the correlation. The performance of the linear and quadratic classifiers may be equivalent under certain conditions.

A framework for Bayes classifier when using the range profile as the feature vector has also been outlined.

Acknowledgments

Working at the Wright Laboratories was a pleasant experience. I thank Dr. (Capt.) Tom Burns and Mr. Rick Mitchell for several thoughtful discussions. Thanks are also due to Mr. Michael Bryant, Mr. Rob Dewall of Veda, and Mr. Bill Pearson for their helpful input, Mr. Vince Velten for the hospitality, and the computing staff in MBV lab and Mr. Randy Riegsecker for their support. Finally, I thank Dr. Jon Sjogren of AFOSR for his recommending me to the summer faculty program.

Fig. 12. Linear Classification

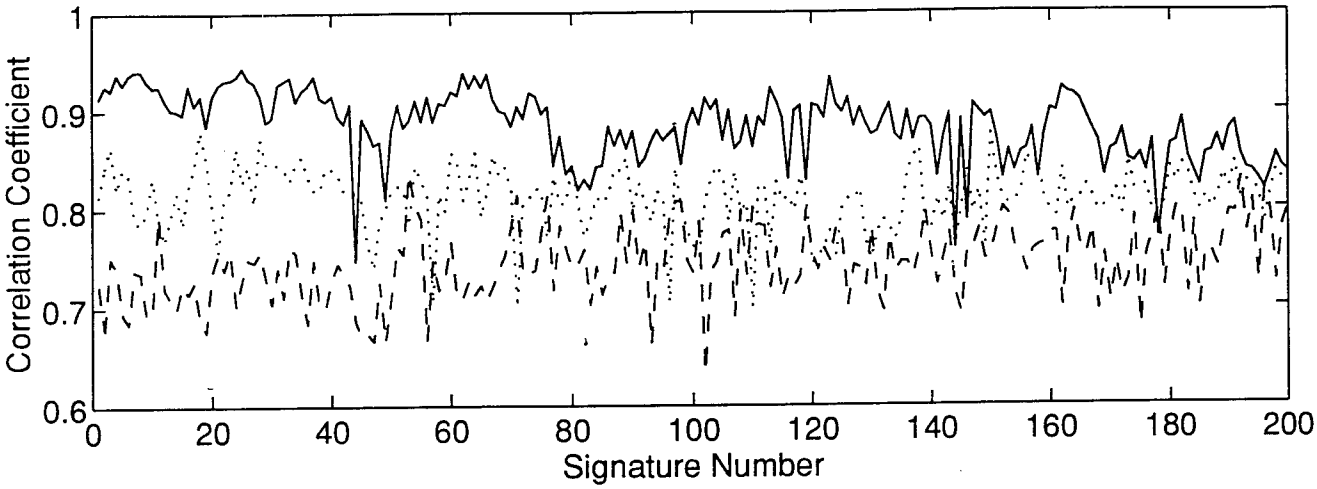
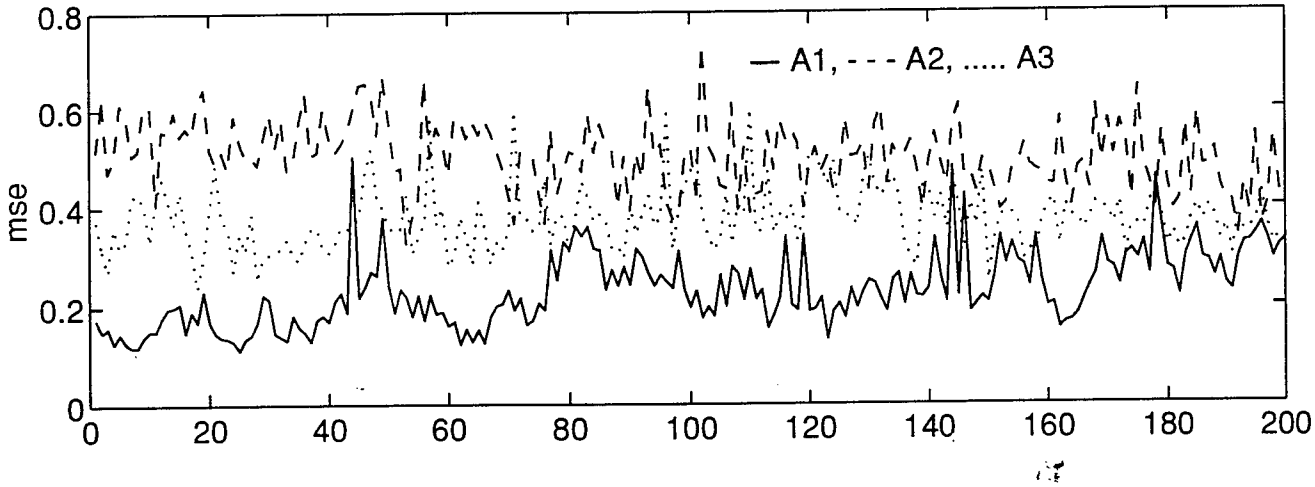


Fig. 13. Quadratic Classification



References

[1] C. E. Baum, "Signature-based target identification and pattern recognition," *IEEE Antennas Propagat. Magazine*, June 1994.

[2] C. R. Smith and P. M. Goggans, "Radar Target Identification," *IEEE Antennas Propagat. Magazine*, April 1993.

[3] D. L. Moffatt *et al.*, "Transient response characteristics in identification and imaging," *IEEE Trans. Antennas Propagat.*, Feb. 1981.

[4] D. G. Dudley, "Progress in identification of electromagnetic systems," *IEEE Antennas Propagat. Soc. Newsletter*, Aug.1988.

[5] J. J. Sacchini, *Development of 2D parametric radar signal modeling and estimation techniques with applications to target identification*, Ph. D. Thesis, Ohio State University, 1992.

[6] D. A. Ausherman, *et al.*, "Developments in radar imaging," *IEEE Trans. Aerospace Electronic Sys.*, July 1984.

[7] M. I. Skolnik, *Introduction to Radar Systems*, McGraw-Hill, 1980.

[8] B. Edde, *Radar Principles, Technology, Applications*, Prentice-Hall, 1993.

[9] G. W. Stimson, *Introduction to Airborne Radar*, Hughes Aircraft Co., 1983.

[10] J. L. Eaves and E. K. Reedy, *Principles of Modern Radar*, ed., Van Nostrand Reinhold, 1987.

[11] C. E. Cook, "Pulse compression - key to more efficient radar transmission," *Proc. of the IRE*, March 1960.

[12] *ARTI Data Processing*, General Dynamics, Pomona Division, June 1991, Contract No. F33615-86-C-1048.

[13] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*, Prentice-Hall, 1987.

[14] S. Haykin, *Adaptive Filter Theory*, 2nd ed., Prentice-Hall, 1991.

[15] S. W. Lang and J. H. McClellan, "Frequency estimation with maximum entropy spectral estimators," *IEEE Trans. Acoust. Speech Signal Process.*, Aug. 1980.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech Signal Process.*, April 1985.

[17] S. Haykin *et al.*, "Classification of radar clutter in an air traffic control environment," *Proc. of IEEE*, June 1991.

[18] H. J. Li and S. H. Yang, "Using range profiles as feature vectors to identify aerospace objects," *IEEE Trans. Antennas Propagat.*, March 1993.

[19] F. Beckner *et al.*, *Automatic Radar Target Identification*, Tech. Report WRDC-TR-90-1003, General Dymanics (Pomona) and U.S. Air Force Wright Labs, March 1990.

[20] E. W. Libby, *Application of sequence comparison methods to multisensor data fusion and target recognition*, Ph. D. Thesis, Air Force Institute of Technology, 1993.

[21] R. Carriere and R. L. Moses, "High resolution radar target modeling using a modified prony estimator," *IEEE Trans. Antennas Propagat.*, Jan. 1992.

[22] K. M. Hansen, *Statistical classification of UHRR signatures: A single-look algorithm*, Sandia National Labs, May 1994.

[23] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, 1990.

[24] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley, 1973.

[25] D. R. Brillinger, *Time Series: Data Analysis and Theory*, expanded ed., Holden-Day, 1981.

[26] A. D. Whalen, *Detection of Signals in Noise*, Academic Press, 1971.

[27] T. W. Anderson and R. R. Bahadur, "Classification into two multivariate normal distributions with different covariance matrices," *Ann. Math. Stat.*, 1962.

DOCUMENTATION OF BOUNDARY LAYER CHARACTERISTICS FOR LOW CHORD-REYNOLDS-NUMBER FLOW ON THE SUCTION SURFACE OF A LOW-PRESSURE TURBINE AIRFOIL

Terry W. Simon
Professor
and
Ralph Volino
Graduate Student
Department of Mechanical Engineering

University of Minnesota
111 Church St. S. E.
Minneapolis, MN 55455

Final Report for:
Summer Faculty Research Program
Graduate Student Research Program
Wright Laboratory
Wright-Patterson AFB

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

and

Wright Laboratory

September 1994

DOCUMENTATION OF BOUNDARY LAYER CHARACTERISTICS FOR LOW CHORD-REYNOLDS-NUMBER FLOW ON THE SUCTION SURFACE OF A LOW-PRESSURE TURBINE AIRFOIL

Terry W. Simon, Professor
Ralph Volino, Graduate Student
Department of Mechanical Engineering
University of Minnesota

Abstract

It is recognized that the low-pressure turbine has, because of its low chord Reynolds number, regions of strong acceleration and diffusion effects. Consequently, there are extended regions of transition from laminar to turbulent flow and there is a strong likelihood of having regions of flow separation. To investigate this low-Reynolds number flow, a program was initiated where a representative low-pressure turbine airfoil configuration is installed in a wind tunnel facility and run at chord Reynolds numbers of 40,000 and 80,000. High background turbulence and disturbances from passing wakes are imposed upon the flow to simulate the turbine environment. The boundary layer state, laminar-like or turbulent, separated or attached, is characterized for representative operating conditions. Instrumentation includes hot-wire anemometry and surface-mounted thin film sensors. When without wake passing disturbances, surface static pressure taps are used to document surface static pressure coefficient, Cp, distributions. With background turbulence present but without wakes, cases are run for TI levels of 1.0 and 20 % for chord Reynolds numbers of 40,000 and 80,000. These cases are repeated with wake generation at representative blade velocities. For documenting the approach flow, turbulence spectra and turbulence intensities are taken for the various TI levels and Re_C values with and without wake generation. From this, the integral length scales are computed. Measurements within the cascade include, on the suction surface, the transition location, the separation location, and the point of reattachment. Instruments for locating these regions are surface-mounted thin-film gages and a hot-wire sensor positioned very near the wall at various streamwise locations. To document the statistical quantities when operating with wake generation, rms fluctuation levels of an ensemble of records are taken behind the wakes and between the airfoils. They are encoded off the translation device and plotted versus t/τ (the dimensionless time within the wake passing period). These data are used to characterize the unsteadiness.

List of Symbols

Symbol	Definition (units)
Lower-case symbols:	
f	Wake passing frequency (Hz).
s	Streamwise distance.
t	Time (sec).
u'	Fluctuation component of streamwise velocity (m/sec).
w'	Fluctuation component of cross-span velocity (m/sec).
Upper-case symbols:	
C	Chord length of the airfoil (m).
C_p	Static pressure coefficients, $(P - P_{static,1})/(P_{tot} - P_{static,1})$.
K	Acceleration parameter, $\nu/U_\infty^2 \, dU_\infty/ds$.
P	Pressure (kPa).
Re_C	Chord Reynolds number, $V_2 C/\nu$.
S	Dimensionless wake passing frequency, Strouhal number, $2\pi fC/V_1$.
TI	Turbulence intensity, $\sqrt{\overline{(u')^2}}/U$, $\sqrt{\dfrac{\overline{(u')^2}+2\overline{(v')^2}}{3}}\Big/U$, or $\sqrt{\dfrac{\overline{(u')^2}+\overline{(v')^2}+\overline{(w')^2}}{3}}\Big/U$.
U	Mean streamwise velocity (m/sec).
V	Flow velocity (m/sec).
Lower-case Greek symbols:	
τ	The wake-passing period (sec).
ν	Kinetic viscosity (m^2/sec).
Subscripts	
∞	at the edge of the boundary layer
static	static, or thermodynamic, value
tot	total, of stagnation, value
1	in the plenum upstream of the airfoil row
2	in the plenum downstream of the airfoil row

DOCUMENTATION OF BOUNDARY LAYER CHARACTERISTICS FOR LOW CHORD-REYNOLDS-NUMBER FLOW ON THE SUCTION SURFACE OF A LOW-PRESSURE TURBINE AIRFOIL

Terry W. Simon
Ralph Volino

Introduction

Compressor and turbine design models have been moderately successful in predicting losses and heat transfer rates in the high-pressure components of the gas turbine where chord Reynolds numbers are large and separation and transition regions are small. When applied to low chord Reynolds number flows, these models fail to match experimental data and often fail to converge. Being able to predict transition at low Reynolds number operation is important for the design of some components of the gas turbine engine. For instance, in the low pressure turbine where designs are aft-loaded, 90% of the blade suction surface can be covered with transitional boundary layer flow (Mayle, 1991). When the chord Reynolds number on an airfoil is decreased, the flow acceleration effect on the boundary layer rises in magnitude and the boundary layer is more likely to separate. An approximate onset of separation Re_c value is 400,000, as given on Fig. 1 which was taken from Sharma, Ni, and Tanrikut (1994). Elevated free-stream turbulence, would reduce the separation Reynolds number. The effects of disturbances due to wakes generated by upstream airfoil rows on this separation Reynolds are more uncertain. Wakes tend to increase the turbulence level (not always, but in general) and they change the momentary angle of attack to off-design angles, favoring increased separation on the suction surface, if the flow were behaving in a quasi static manner. Under high-frequency oscillation of the flow due to passing wakes, the effects are not certain. Needed are transient measurements of the boundary layer flow over the downstream portions of the suction surface, ensemble-averaged on the wake passage event. Such measurements provide information about the state of the boundary layer and the receptivity of the boundary layer to external disturbances, including those from the passing wakes. Possible wake generators may be cylinders, representing wake turbulence with large-scale turbulence, thin plates oriented parallel to the flow so that the wake is decaying boundary layer turbulence which consists of mostly small-scale turbulence, or actual airfoils. Comparisons of measurements in cascades and rotating rigs indicate that losses and heat loads are higher in the unsteady flow than in steady flow (Sharma et al., 1994, Hodson, 1983, Blair et al., 1988, Doorley et al., and Sharma et al. 1990). Sharma et al., 1994 proposed scaling on a relative time scale to capture this effect. The appropriate ratio is the wake passage period divided by the transit period for fluid

flow through the airfoil row. Cases of similar values of this ratio display similar augmentation due to unsteadiness. Such scaling is applied in the present study. Operation at low Reynolds numbers would create separation zones allowing a study of the effect of the wakes on the incipience to the separation process, on the free-shear layer transition length, and on the separation bubble length. Under low-Reynolds-number conditions, upstream wakes can result in smaller separation zones and lower losses relative to steady slow. Presently, no model has been developed to capture this effect (Sharma et al., 1994). Thus, wake effects on the separated flow must be included for a low-Reynolds number study to be accurate and complete. Hot-wire anemometry measurements can be made in the free-shear layers over the separation bubbles without affecting the flow. Measurements within the separated flow zone are difficult, but possible, and needed. With these measurements, one could assess whether the change in incidence angle during the wake passing event is of importance or whether the main wake-related effect is the turbulence washing over the boundary layer and free-shear-layer flows. Thus, careful documentation of the turbulence and wake disturbance effects is needed in support of design model development.

Objectives

The objectives of this program are to determine the transition and separation zone locations for a particular low-pressure airfoil under representative conditions, including an assessment of the effects of free-stream turbulence and of wake passings on the locations of these zones. A secondary objective is to document the pre-separation boundary layer receptivity to external disturbances and the instability of the free-shear layer over the separation zone. In doing so, the cases of Table 1 are investigated, the transition and separation zones for each are located, and measurements are taken in the pre-separation boundary layer and in the separation zone. For cases with wakes, this documentation is given for various times within the wake passing period.

Method

First, the wake generator and the turbulence generation equipment are taken out of the flow and the entry flow is documented for Cases 1 and 3. Documentation is by measuring the transverse uniformity of mean velocity and turbulence intensity (based upon the streamwise component of velocity only) then characterizing the turbulence with single-velocity-component power spectral density distributions and the isotropy of the flow with measurements of the streamwise and cross-span components of turbulence intensity on a streamline which passes though the center of the cascade. Once the turbulence is characterized, the static pressure coefficient distribution on the surface of the blade is documented and the locations of critical points related

to transition and separation are found for the two cases. Next, the turbulence generator is activated and the above measurements are repeated (Cases 2 and 4). When this is complete, the wake generator is activated and Cases 5 through 8 are documented. These cases require a somewhat different measurement program. The single-component power spectral density distribution is taken downstream of the wake generator and within the center of the cascade channel. Single-component turbulence intensity measurements are time resolved according to the fraction of the wake-passing period; values of RMS fluctuation values from an ensemble of 50 records are taken. Time resolution for these readings is 1/100 of a passing period. Next, the unsteady locations of critical points regarding transition and separation are located for various times within a wake-passing period, again with the 1/100 period resolution and with a spatial resolution of about 2.5 mm (0.10 inch). This test program is outlined in Table 2. The separate effects of Reynolds number, turbulence intensity, and wake passage disturbances are assessed by comparing these 8 data sets.

Experimental Apparatus

Cascade Facility

The experiments are conducted in the AFIT cascade facility. This wind tunnel is driven with a Buffalo Forge Model No. BL-365 centrifugal blower operating in the suction mode. Flow velocity control is with an Allen-Bradley Model No. 1336S motor controller, with inlet dampers on the fan, and with a bypass vent on the ductwork between the test section and the fan. A schematic of the wind tunnel layout is given at Fig. 2. The cascade consists of four geometrically identical blades of chord length 11.4 cm (4.5 inches).

Turbulence Generator

Upstream of the cascade is a turbulence generation device which consists of a passive square grid of 13 mm by 13 mm (0.5 inch by 0.5 inch) square bars arranged with a 25.4 mm (1.0 inch) center-to-center spacing. This resides 1.5 m (57 inches) upstream of the cascade row. A distance of 0.6 m (24 inches) downstream of the passive grid is an active jet grid with 3 tubes with 6 blowing holes each distributed and oriented as shown in Fig. 3. The passive grid/jet grid arrangement is a replica of one described by Sahm and Moffat, 1992.

Wake Generator

At a distance of 7.6 cm upstream of the leading edge of the cascade row is a series of 6 cylinders which are traversed across the tunnel cross-section in the transverse direction. These cylinders simulate the wakes that are generated by the airfoil row which resides just upstream of the airfoil row of interest in an actual turbine. The cylinders which simulate the wakes in

the facility are 9.5 mm (0.375 inch) in diameter and are separated by the same transverse spacing as that of the airfoils, 91.7 mm (3.611 inches). These cylinders are driven by a device which is capable of translating them at selected velocities from 0.5 m/sec to as high as 5.5 m/sec. The total translation distance is 43 cm (17 inches) and it is estimated that about 6.5 cm (2.5 inches) each is required for acceleration and deceleration phases of the translation. Early tests with the translation driver, alone, indicate that over the central 30 cm (12 inches) of travel, the velocity is uniform to within 6% of the nominal velocity at the low end of the velocity range and within 1.5% at the high end of the range. The translation is powered by a spring, to aid with the initial acceleration, and with a 1 hp, DC electric motor to continue the acceleration and sustain the velocity by driving a push bar with a 78 mm (3 inch) diameter friction wheel. This drive motor is model number DC-1 of the Reliance Electric Co. Maximum rotational speed is 1725 RPM. The drive device is shown on Fig. 4. The translation slide is sketched in Fig. 5. The wake generating tubes are inserted in holes in the runner and the drive bar is driven by the drive device. Attached to the drive mechanism is a photo-diode sensor which indicates when the translation runner has moved a distance of 10 cm (4 inches). The signal from this sensor is used to activate the data acquisition trigger. The sensor's voltage rises from 0 to ~ 2 v as a fin on the drive bar of the slide mechanism interrupts a light beam. This signal is input to a Hewlett-Packard 8012B Pulse Generator which converts this to a stronger signal which is adjustable but is usually set to give a 0 to 3 v rise when the fin interrupts the light beam. This signal is available to the data acquisition system. The duration of the measurement portion of the translation is from one to five wake passings, which ranged from 16 ms to 0.2 sec for the cases tested.

Hot Wire Anemometer

Instantaneous local velocities are measured using a TSI Model 100 (IFA-100) Intelligent Flow Analyzer, constant-temperature anemometer with a TSI Model No. 1210-T1.5 Tungsten single-wire, hot-wire probe or a TSI Model No. 1243-20 platinum cross-sensor, hot-film probe. The cross-sensor probe is used to document the isotropy of turbulence of the flow ahead of the cascade. It has two 51 μm (2 mil) diameter sensors, mounted perpendicular to one another, parallel to the airfoil axes, and 45° to the cascade endwall and to the flow. The single-wire is used for streamwise velocity measurements and for indicating the state of the boundary layer on the suction surface of the airfoil by traversing it to very near the surface and measuring the fluctuation intensity and interrogating the waveform to detect signs of flow reversal. The probe can be traversed to near the surface with an optical rotator device with a precision of one second of rotation which results in a precision of translation of 4 μm (16 μinch). The single-wire probe has one 3.8 μm (150 μinch) diameter wire, mounted perpendicular to the flow and to

the cascade endwall. Each hot-wire output signal, an analog voltage, is amplified and filtered using the TSI Model No. 158 signal conditioner. Each is read with the data acquisition unit.

Surface-Mounted Thin-Film Gages

The surface-mounted, thin-film gages are small, hot-film elements vapor-deposited on a thin sheet of polymeric substrate and sandwiched between another such sheet. The sensor resistance is approximately 2 ohms. The resistance changes with temperature similarly to that of the wire of the hot-wire anemometer sensor discussed above. The surface-mounted sensor is driven with the same anemometer bridge as used for the hot-wire sensor, the IFA-100. The sensor streamwise separation distance is 2.5 mm (0.1 inch). The sheet employed in the present study contains 30 sensors aligned along the streamwise distance. This sensor is wrapped around an airfoil so that it covers, approximately, the downstream 70% of the suction surface and the downstream 15% of the pressure surface.

Surface Static Pressure

The surface static pressures are read using 22 static pressure ports installed on a blade surface. They are connected to steel tubing which extends through the endwall of the facility connecting to tygon tubing outside which, in turn, lead to a 36-port Scanivalve selector. The Scanivalve is programmed to bring the ports, one at a time, into communication with the diaphragm of a Validyne, 860 Pa- (3.5 inch)- maximum-pressure-difference, variable-reluctance, pressure transducer driven by a Validyne model CD-15 Carrier Demodulator. The pressure transducer provides an analog voltage signal which is read with the data acquisition unit. Ten seconds is allowed between each reading for the pressure to equalize with the diaphragm.

Temperatures

Flow temperature is measured with an iron-constantan (Type J) thermocouple. The digitizer is used to acquire the thermocouple signal and convert it to temperature. Additionally, a combination of a single Type J thermocouple and an Omega Digicator displays a readout of the flow temperatures for visual reference by the operator.

Data Acquisition

All the data acquisition is with a Hewlett-Packard High Speed Data Acquisition System, Model No. HP3852A. This unit has a 13-bit (sign plus 12 digits) A/D converter driven by a High-Speed FET Multiplexer, allowing multiple channels of high speed acquisition with nearly simultaneous sampling. For X-wire measurements, two channels are sampled; for all other measurements, a

single channel is sampled. Sampling rates of as high as 100,000 samples/sec can be achieved and record lengths of up to 8192 readings can be acquired (for X-wire measurements 50,000 samples/sec can be acquired and 4096 points per channel can be stored). The data acquisition unit also has a 16 bit integrating digital voltmeter which is used for acquisition of pressure readings from the pressure transducer and thermocouple readings of the air temperature in the wind tunnel. Temperature readings are stored along with all hot-wire readings for correction of the hot-wire voltages to the calibration temperature. A relay actuator on the HP3852A controlled the Scanivalve drive motor. Data are transferred to a Unisys model PW^2 386-based PC computer via a National Instruments IEEE-488 interface board. Data acquisition is controlled with the PC using software written in QuickBASIC. Data processing is done on the PC using software written in QuickBASIC or on a Sun workstation in Minnesota as explained below.

Data for spectral measurements are acquired with the single-wire probe located 5.4 cm (2 inches) upstream of the blades in the center of the channel. Data for the analysis are acquired in three sections. The first section is acquired with a 100 kHz sampling rate and low-pass filtered at 10 kHz. The second section is acquired with a 10 kHz sampling rate and low-pass filtered at 1 kHz. The third section is acquired at 1 kHz with low-pass filtering at 100 Hz. For each section, 10 traces of 8192 points are digitized. Raw voltage data from the IFA-100 as well as calibration constants and the ambient temperature are stored to disc. These data are then transferred to a Sun workstation at the University of Minnesota for processing. A program written in C is used to perform a Fast Fourier Transform (FFT) and compute the power spectral density of each half-trace of data (i.e. 20 records of 4096 points each are processed with the FFT for each of the three sampling rates). The 20 spectra at each sampling rate are then averaged to create a composite spectrum. The three sections of the spectrum are then pieced together for presentation. Acquiring the spectra in sections allows better resolution of both high and low frequencies, maximizing the quality of the spectrum, given the limited amount of data that can be acquired and stored in the digitizer for each record.

The uniformity of the flow in the wind tunnel is checked with both the single-wire probe and the X-wire probe. With the single-wire, the probe is moved to various positions in the flow, were 4000 data points are digitized at a 100 Hz sampling rate. The instantaneous data are converted to velocities and averaged to give U and u'. These two values are then stored. The procedure for the X-wire is similar. A total of 250 data points are acquired in 60 seconds. The values of U, V, u' and v' are calculated and stored. All hot-wire data are low-pass filtered at 5 kHz, with the exception of the spectral data described above.

For unsteady measurements with the traversing wakes in use, raw voltage traces from the IFA-100 are digitized and stored for later ensemble averaging. Traces could be acquired from the single-wire probe or from one of the hot-film sensors mounted on an airfoil. The sampling rate and record length of each trace are adjusted so that 100 samples are acquired between each wake passing event. Fifty traces, corresponding to 50 traverses of the slider mechanism are acquired at each position of interest for later conversion to velocities and for ensemble averaging.

Static pressure measurements are taken at 22 positions around the airfoil and from a pitot tube located 5.4 cm (2 inches) upstream of the airfoil. Total pressure is also measured with the pitot tube. From these measurements, the pressure coefficient distribution around the airfoil is calculated and stored.

Status of the Program

The entire test facility has been modified to accept the new airfoils, the new wake generator, and the new turbulence generator. Four new airfoils have been constructed, two without instrumentation, one with the surface-mounted thin-film sensors, and a fourth with the static pressure taps. The thin-film sensor sheet of 30 hot-film elements has been attached to bus wires. The slider mechanism and the device which drives it has been designed, constructed, and qualified. A photo-diode device to create a trigger signal for the data acquisition has been constructed and checked out. All data acquisition methodology has been developed (including modification of the software) and checked. The tunnel modification is essentially complete. The facility now stands essentially ready for the beginning of data acquisition. Unfortunately, a few small but necessary machining operations remain that have prevented the actual data-taking portion of the program from beginning during the Summer period allotted for the program. Thus, arrangements have been made for continuation.

ACKNOWLEDGMENTS

We wish to express our gratitude to the Aero Thermal research team of the Wright Propulsion Laboratory for their aid and support; Richard Rivir, Task Manager, and, within the Laboratory, Greg Cala, David Pestian, Shichuan Ou, and Chris Murawski. We especially appreciate the sound technical advice and encouragement of our host Dick Rivir. We also thank Professor Paul King of the Air Force Institute of Technology for the use of the AFIT facilities and for his technical help. The fabrication of the new airfoil blades and modification work on the wind tunnel were done by Dave Driscoll at the AFIT shop. We thank him for his steady and careful efforts, especially for the cast airfoils which are truly works of art. We also appreciate the skill and patience of James Higgins, Shop Foreman, and Tim Klopfenstein at the Dayton Research Institute Shop at the University of Dayton. Finally we thank AFOSR for their financial support and RDL for their fine management of the program.

References

Blair, M. F., Dring, R. P., and Joslyn, H. D., 1988, "The Effects of Turbulence and Stator-Rotor Interactions on Turbine Heat Transfer, Part I, Design Operating Conditions," ASME Paper #88-GT-125.

Doorley, D. J., Oldfield, M. L. G., and Scrivener, C. T. J., "Wake Passing in a Turbine Rotor Cascade," AGARD CP-390, Paper No. 7, Bergen, Norway.

Hodson, H. P., 1983, "The Development of Unsteady Boundary Layers in the rotor of an Axial-Flow Turbine," AGARD Proceedings No. 351, *Viscous Effects in Turbomachines.*

Mayle, R.E., 1991, "The Role of Laminar-Turbulent Transition in Gas Turbine Engines, ASME Paper No. 91-GT-261.

Sahm, M. K., and Moffat, R. J., "Turbulent Boundary Layers with High Turbulence: Experimental Heat Transfer and Structure on Flat and Convex Walls," Report No. HMT-45, Thermosciences Division, Mechanical Engineering Department, Stanford University, 1992

Sharma, O. P., Pickett, G. F., and Ni, R. H., 1990, "Assessment of Unsteady Flows in Turbines," ASME Paper #90-GT-150.

Sharma, O. P., Ni, R. H. and Tanrikut, S., "Unsteady Flows in turbines - Impact on Design Procedure," from *Turbomachinery Design Using CFD*, AGARD Lecture Series No. 195, given at the Ohio Aerospace Institute, May 1994.

Table 1

Cases of the study

Case number	Re_c	TI (%)	Case number	Re_c	TI (%)
without wake generation			with wake generation		
1	40,000	1	5	40,000	1
2	40,000	20	6	40,000	20
3	80,000	1	7	80,000	1
4	80,000	20	8	80,000	20

Table 2 Outline of the Test Program

I. No Wake Generator
   A. For Re_c = 40,000 and 80,000
      1. For no turbulence generator and an active turbulence generator
         a. Transverse distributions
            - U
            - $\sqrt{\overline{(u')^2}}/U$
         b. For the centerspan and mid-passage streamline
            - 1-D PSD
            - $\sqrt{\overline{(u')^2}}/U$, $\sqrt{\overline{(w')^2}}/U$
         c. For the airfoil surfaces
            - C_p
            - Locations of laminar & turbulent, attached & separated flows

II. With the Wake Generator
   A. For Re_c = 40,000 and 80,000
      1. For no turbulence generator and active turbulence generator
         a. For the centerspan and mid -passage streamline
            - 1-D PSD
            - Time-resolved:
               > $\sqrt{\overline{(u')^2}}/U$
               > Locations of laminar & turbulent, attached & separated flows

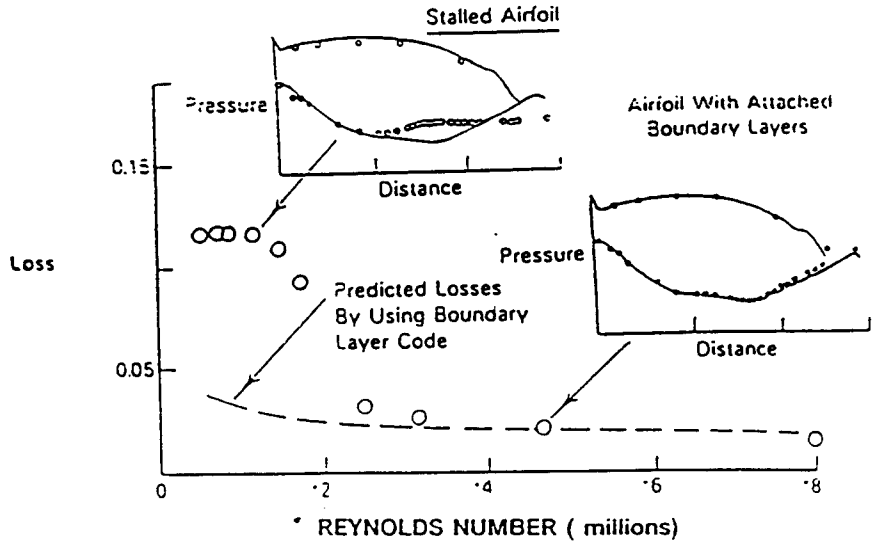


Fig. 1 High loss levels measured for airfoils at low Reynolds number. Boundary layer flow separation evident on airfoil suction surfaces. Taken from Sharma, Ni, and Tanrikut, 1994.

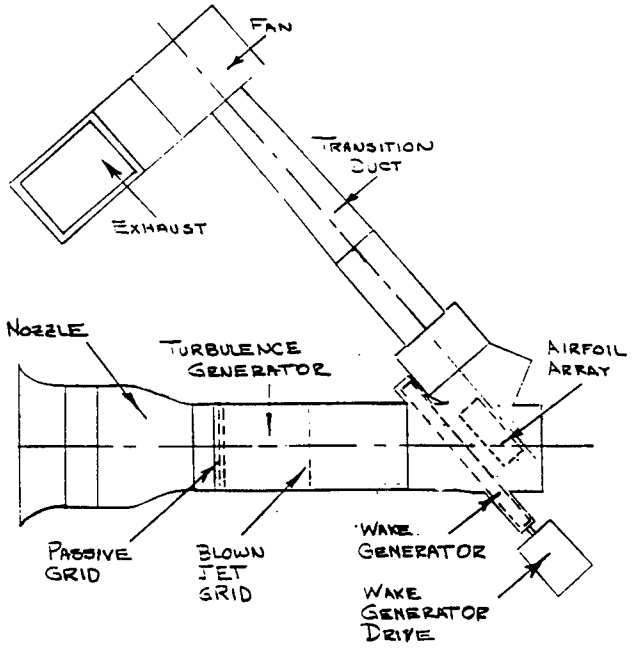
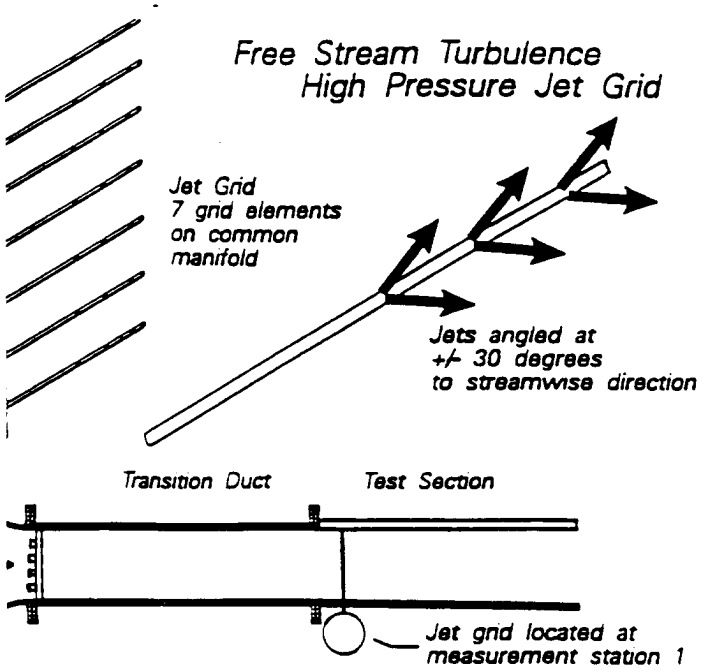


Fig. 2 Layout of the wind tunnel facility.



Schematic of the jet grid. From Sahm and Moffat, 1992.

A photograph of the wake generator driving mechanism.

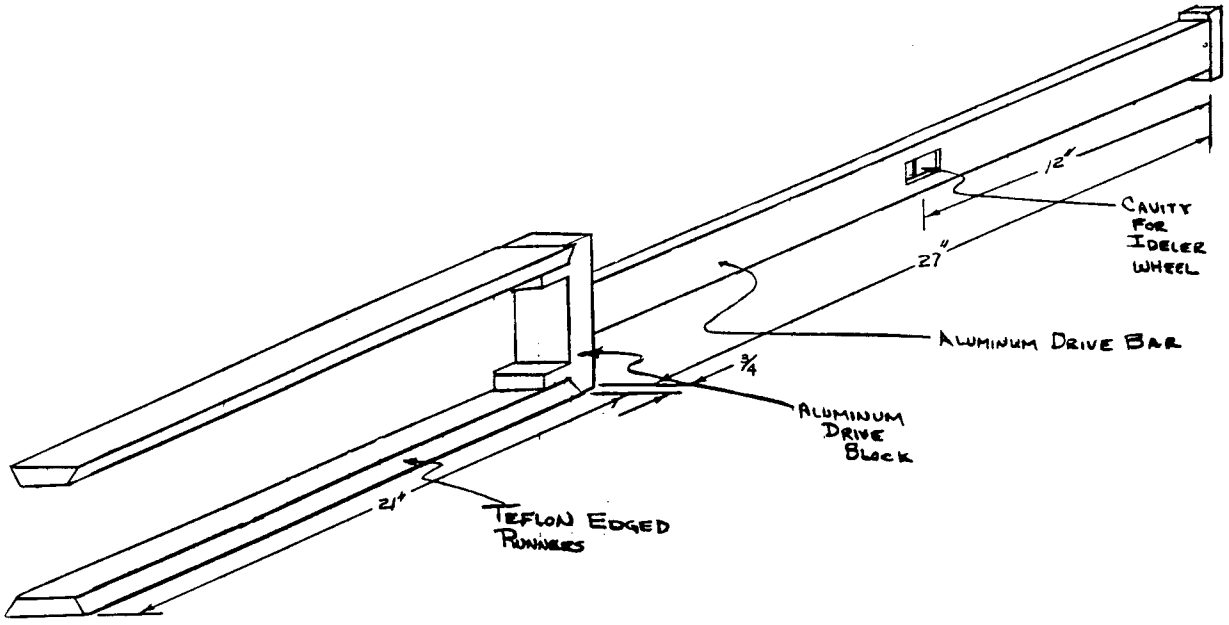


Fig. 5 A sketch of the traverse slider and the push bar

MECHANISM FOR INDIUM SEGREGATION IN $In_xGa_{1-x}As$ STRAINED QUANTUM WELLS DURING MBE GROWTH

Marek Skowronski
Associate Professor
Department of Materials Science and Engineering

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

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

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

September 1994

MECHANISM FOR INDIUM SEGREGATION IN $In_xGa_{1-x}As$ STRAINED QUANTUM WELLS DURING MBE GROWTH

Marek Skowronski
Associate Professor
Department of Materials Science and Engineering
Carnegie Mellon University

ABSTRACT

Indium surface segregation was investigated during deposition of $In_{0.22}Ga_{0.78}As$ strained quantum wells on GaAs by Molecular Beam Epitaxy. The indium concentration profiles have been studied by Desorption Mass Spectrometry and amount of surface indium was assessed by Reflection High Energy Electron Diffraction. In particular, the In profiles in as-grown and in situ annealed strained layers have been studied. The analysis of results led us to propose a new mechanism for surface segregation, namely that of solid state diffusion enhanced by high concentrations of cation vacancies in the subsurface region with thickness of approximately 8 monolayers. Consequences of the proposed model for deposition of square InGaAs quantum wells will be discussed.

MECHANISM FOR INDIUM SEGREGATION IN $In_xGa_{1-x}As$ STRAINED QUANTUM WELLS DURING MBE GROWTH

Marek Skowronski

INTRODUCTION

In recent years there has been a marked increase of interest in surface segregation phenomena during deposition of semiconductor alloys. It is related to our advancing ability to characterize these materials on the ever diminishing scale and increasing importance of control of interfaces between dissimilar materials. It is essential in either for example Pseudomorphic High Electron Mobility Transistors, Quantum Well Lasers, or type-II superlattice infrared detectors to control to potential well profiles on the monolayer scale. Deviations from the desired potential shape and/or nonuniformities of the well thickness are known to result in lower carrier mobilities, significant shifts of the emission wavelength, or detectivity. The presence of surface segregation prevents us from producing ultra-sharp interfaces the device designs call for. It is, therefore, important to understand the nature of the underlying processes and predict deposition conditions and procedures which will allow for deposition of desired structures.

In this report, we focus on the effects and mechanisms of indium surface segregation in $In_{0.22}Ga_{0.78}As$ strained layers deposited on GaAs substrates as investigated by Desorption Mass Spectrometry (DMS) and Reflection High Electron Energy Diffraction (RHEED). Results of both techniques are presented and their consequences for deposition of ultra-abrupt interfaces is discussed.

BACKGROUND

Although the surface segregation in metals have been well known and investigated for several decades,[1] its importance for the Molecular Beam Epitaxial deposition of semiconductor alloys have been recognized only recently[2-8]. The driving force for this phenomenon is the same in both material systems, namely, the difference in surface energy of different species for example indium and gallium on the surface of $In_xGa_{1-x}As$ layer. This energy, in principle, has two components due to difference in bond energy and strain energy associated with difference in ionic radii. Because of the above one could expect the surface segregation in semiconductors to be as prevalent as in metals. However, both systems are very much different in terms of kinetics of segregation processes. In metals the primary segregation mechanism is bulk solid state diffusion. The samples are prepared in the bulk form of an homogenous alloy, cut to appropriate shape with polished surface, and cleaned in Ultra High Vacuum by combination of thermal annealing and sputtering. Clean samples are then annealed for a time long enough to reach thermal equilibrium between surface segregation and thermal evaporation. Surface compositions are assessed by in situ Auger Spectroscopy or X-ray Photoelectron Spectroscopy. The typical values of surface concentration enhancement of a diluted component is anywhere between several percent and close to a 100 percent coverage depending on the annealing temperature and difference in surface energies. In most of investigated cases it was possible to approach a state of local thermal equilibrium between bulk and surface concentrations.

In semiconductor alloys the bulk diffusion coefficients at typical processing and deposition temperatures are much lower than in metals. For example the indium bulk diffusion coefficient obtained by investigation of long terms anneals of InGaAs quantum wells is only $2x10^{-18}$ cm^2/s at typical growth temperatures[9]. Such low D values could not result in any appreciable surface segregation during the deposition process. This conclusion led

researchers to propose a different segregation mechanism based on the assumption of a two different states for surface indium atoms. One was the covalently bonded chemisorbed state and the second weakly interacting "floating indium layer".[10] The "floating" indium layer is expected to be populated in the initial stags on $In_xGa_{1-x}As$ layer growth at the expense of lower indium concentration in the layer. This would account for experimentally observed indium deficiency in the first to be deposited InGaAs. Once it reaches the steady state population, it should move with the growth front.

There are number of potential problems with this model. First, there is no independent experimental verification of existence of physisorbed indium. Such proof could be obtained by thermal desorption experiment which should show two much different activation energies for surface indium. This, in turn, should result in two distinct peaks of desorbed indium at temperatures low enough not to allow for gallium desorption. No such peaks were observed. Also, the growth interrupt during $In_xGa_{1-x}As$ deposition should result in slow incorporation of physisorbed indium into the layer and produce a local maximum in the indium content. No such increase has been reported so far.

The above discussion indicates that there is a need for a new mechanism of surface segregation. Such model and the experimental evidence supporting it is presented below.

RESULTS AND DISCUSSION

A major research tool used to investigate the indium segregation phenomena in InGaAs layer was Desorption Mass Spectrometry (DMS) which was used to assess indium distribution profiles in as-deposited and annealed layers. It was, therefore, important to determine the depth resolution of the DMS technique which was tested the structure shown in Fig. 1.

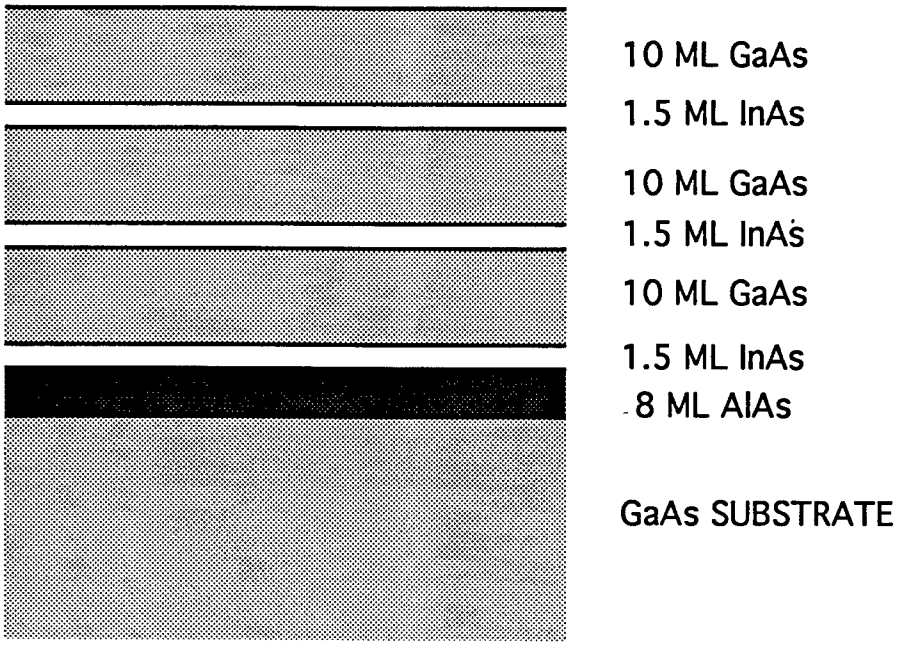


Fig. 1 InAs / GaAs structure used to assess the resolution in Desorption Mass Spectrometry

The structure consisted of initial 8 monolayers of AlAs followed by 1.5 ML of InAs and 10 ML of GaAs repeted three times. The AlAs layer served as the etch stop. The indium and gallium desorption transients obtained using constant current in the heater are shown in Fig. 2.

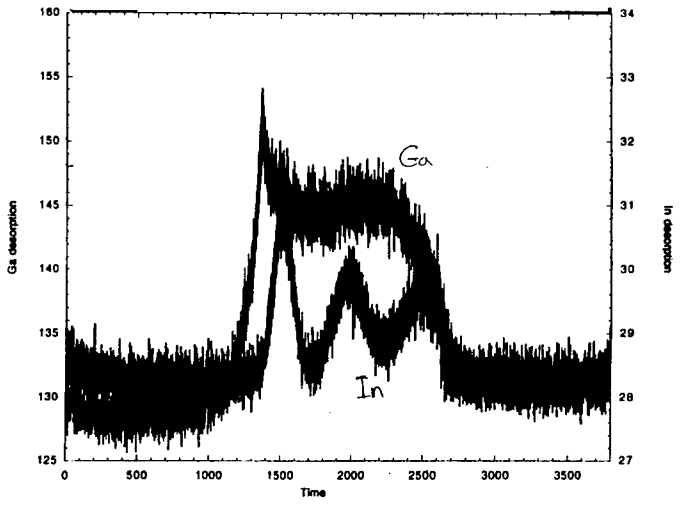


Fig. 2 Indium and gallium desorption transients from structure shown in Fig. 1.

It is apparent that indium desorption exhibits three well defined peaks corresponding to three thin InAs layers. The width of peaks increases with the distance from the surface which was most likely due to non uniform temperature distribution in the wafer and consequently uneven thermal etching rates. However, the fact that all layers are resolved assures that resolution is better than 2 nm in layers approximately 15 nm thick.

The typical desorption transients from 12.5 nm thick $In_{0.22}Ga_{0.78}As$ well deposited on AlAs are shown in Fig. 3.

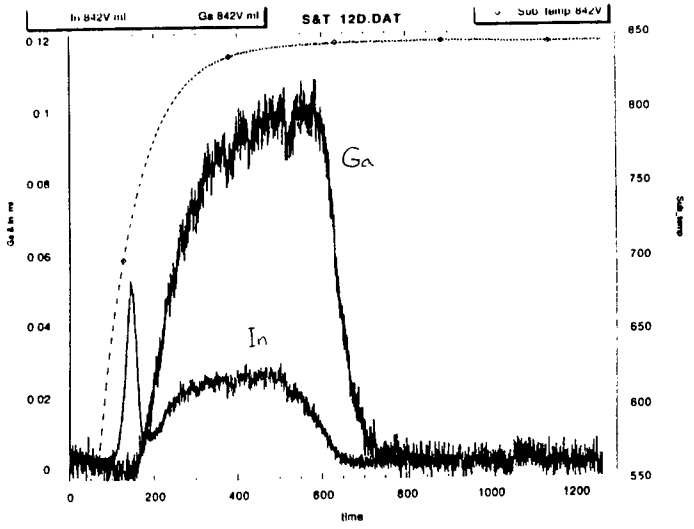


Fig. 3 Indium and gallium desorption transients from $In_{0.22}Ga_{0.78}As$ quantum well

The first to desorb species is indium. This is to be expected as In-As bonds are weaker than that of gallium. The indium desorption stops after approximately 1.5 monolayers (integrated area under the first peak). This amount is lower than the total indium content in the layer and roughly corresponds to amount of surface indium as detected by RHEED experiments. At this point, the surface of the layer is gallium terminated and this prevents further desorption of indium. With increase of temperature both indium and gallium start to desorb at the same time. This process is limited by gallium desorption, as soon as gallium atom leaves the

surface exposing underlying indium it follows instantly. As a result one would expect In signal to parallel that of gallium. It is largely true with two small deviations.

(i) At the end of desorption run indium signal decreases faster than gallium signal. There are two contributions to this deviation. An obvious one is that if the remaining thickness is less two monolayers all indium atoms are in effect exposed and should desorb immediately. At this point only gallium would remain. However, this effect can account only for a deviation of about 0.5 ML of indium. The second and more important effect is the indium which segregated to the surface. The starting profile of indium concentration is shown in Fig. 4.

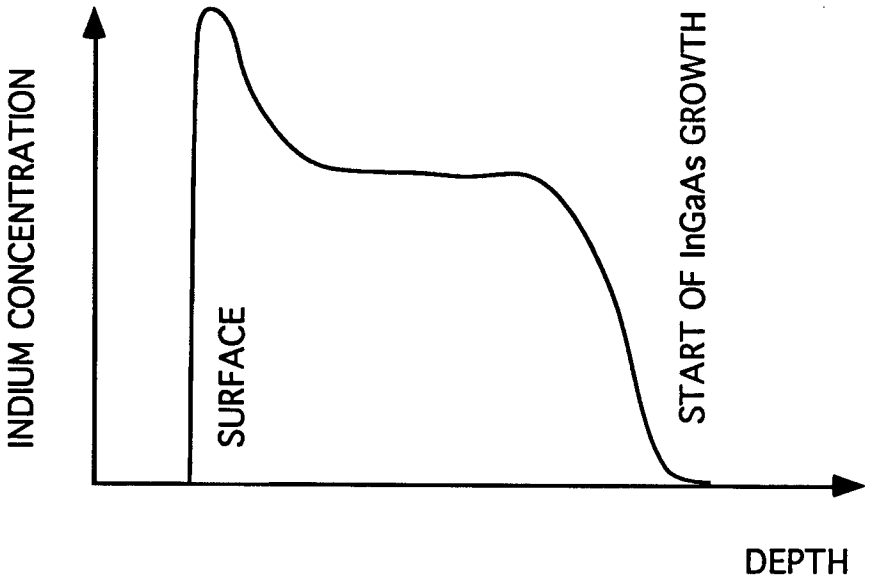


Fig. 4 Indium concentration profile at the end of deposition.

The part of the well close to the AlAs interface is depleted of indium while the surface concentration is enhanced. The magnitude of depletion closely corresponds to the area under the surface peak.

It should be pointed out that the temperatures of indium and gallium desorption are relatively close to each other implying that the binding energies of both species are similar. The reported activation energies are 4.0 and 4.9 eV, respectively. This proximity indicates that both atoms are covalently bonded to the surface and surface indium is not physisorbed.

(ii) The second deviation appears on the rising slope of gallium desorption. Namely, the indium signal increases at slightly faster rate than that of gallium. For certain heating rates, this results in a shoulder or even a peak on indium desorption. This increase implies that there is an additional mechanism which supplies indium to the surface faster. It is proposed here that this mechanism is subsurface diffusion. In particular, we postulate that the diffusion coefficient is a function of distance from the surface and is enhanced by several orders of magnitude in the top 10 monolayers. If this assumed model is correct then indium desorption rate should be equal to the sum of gallium desorption rate (weighted by the indium concentration 0.22/0.78) and the diffusion rate. As the desorption experiment continues, subsurface diffusion is expected to deplete indium in subsurface layer and as a consequence diffusion flux should decrease. At steady state condition the diffusion flux should decrease to zero. This mechanism, therefore, should produce a temporary increase in indium desorption rate before attaining steady state value as observed in Fig. 4.

Direct verification of the enhancement of subsurface diffusion was obtained in the in situ annealing experiment. The structure consisted of 12.5 nm of $In_{0.22}Ga_{0.78}As$ deposited on AlAs thermal etch stop layer. The thermal desorption cycle consisted of surface indium desorption performed at 720 °V followed by in situ anneal time at the same temperature for different periods of time. At the end of the anneal the temperature was ramped over 20 minutes to gallium desorption temperature of 850 °V The indium desorption transients

obtained for the anneal time of 5 and 40 minutes are shown in Fig. 5. The indium content in subsurface region in layer annealed for 40 minutes is significantly lower than in the reference layer. The difference is increasing with time of anneal and is expected to saturate for long anneal times at the value corresponding to indium content in layer with enhanced subsurface diffusion. The long term in situ annealing experiment is planned for October 28, 1994.

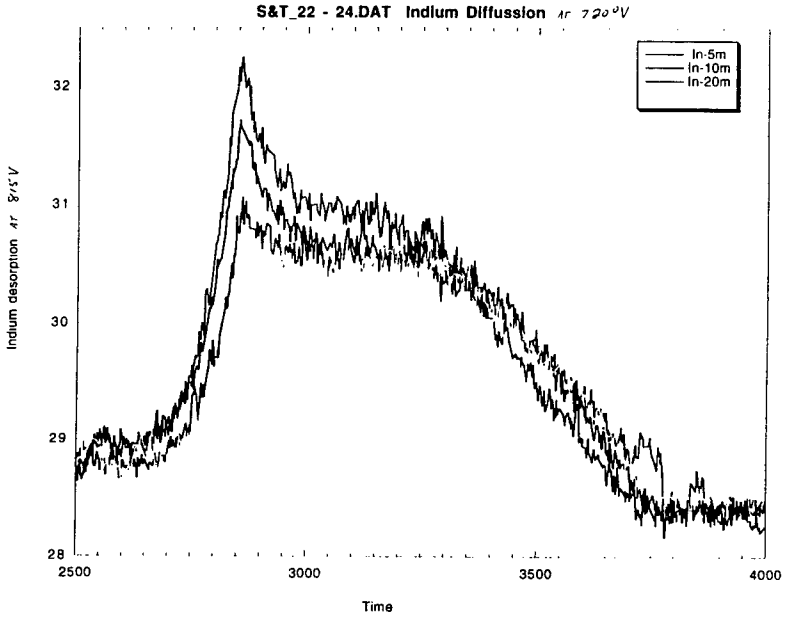


Fig. 5 Indium desorption profiles from as-deposited and annealed InGaAs layers.

The concept of enhanced subsurface diffusion, although not widely accepted or quantified, was used to explain other experimental observations. Philips et al.[11] used a depth dependent diffusion to explain ordering in semiconductor alloys. The driving force for ordering is the same as in the surface diffusion: the difference in size between constituents of an alloy. In ordering, of III-V mixed cation alloys, the cation site underneath the group V dimers (it is present in the second cation layer and fourth layer overall) is experiencing compressive stress and therefore is preferentially occupied by a smaller atom. In order to account for large order parameters observed in these alloys one has to allow for almost all atoms in the second monolayer exchanging sites. The stresses are reversed when an additional monolayer is deposited. The only way to reconcile this reversal with ordering is to assume that diffusion in

the third monolayer down from the surface is much slower than in the second one and the ordered state is "frozen in". The above model is well documented and explains multitude of experimental observations. It is also fully consistent with our observation of subsurface diffusion.

The dominant mechanism of autodiffusion in semiconductor compounds is a vacancy mechanism. The diffusion rate is, in fact, limited by the concentration of vacancies. The enhanced subsurface diffusion coefficient should, therefore, be reflected in vacancy concentration as well. An observation to this effect was made in silicon wafers diffused with gold[12,13]. Gold is know to be a fast diffuser which moves interstitially but can occupy substitutional sites. The concentration of Au_{Si} is expected to reflect the local silicon vacancy concentration. The experimental profiles show an increased concentration of Au_{Si} at both front and back surface with low concentrations in between, again in accordance with our interpretation.

CONTINUATION OF THE PRESENT RESEARCH

The current results on effects of in situ annealing on indium surface depletion (Fig. 5) should be extended to longer anneal times. As stated above it is expected that because of the strong dependence of subsurface diffusion enhancement on the distance from the surface the depletion width will saturate. The total indium evaporated during annealing will give the depth at which diffusion coefficient drops to the bulk value.

Also, although the results of DMS and RHEED experiments are a strong argument for the indium surface segregation the existence of subsurface diffusion at the growth temperatures complicates the studies of indium profiles. There is a clear need for the characterization of indium profiles at room temperature when all atoms are "frozen" in place. The resolution of

the technique should be on the order of 1-2 monolayers and sensitivity at the level of 2-3 indium atomic percent. Several transmission electron microscopy based techniques approach both resolution and sensitivity requirements. The most promising two are Z-contrast mapping developed by S. Pennycook of Oak Ridge National Laboratory and by imaging energy filter microscopy. The latter instrument (probably the first in the world to be coupled with atomic resolution microscope) is being installed at Carnegie Mellon University. Drs. Skowronski and De Graef are planning to perform quantitative measurements of indium profiles in PHEMT structures.

ACKNOWLEDGMENTS

The work summarized above is a result of close collaboration between myself K.N. Evans, R. Kaspi, J. Ehret, and E. Taylor with significant input provided by other members of the WL / ELRA branch.

REFRERENCES:

1. P. Wynblatt and R.C. Ku, in Interfacial Segregations, (ASM, Metals Park, OH, 1979) 115 (1979).
2. J.M. Moison, F. Houzay, F. Barthe, J.M. Gerard, B. Jusserand, J. Massies and F.S. Turco-Sandroff, J. Crystal Growth **111**, 141-150 (1991).
3. O. Brandt, L. Tapfer, K. Ploog, R. Bierwolf and M. Hohenstein, Appl. Phys. Lett. **61**, 2814-2816 (1992).
4. W. Braun and K.H. Ploog, J. Appl. Phys. **75**, 1993-2001 (1994).
5. K. Muraki, S. Fukatsu, Y. Shiraki and R. Ito, J. Crystal Growth (1993).
6. J.M. Gerard, C. d'Anterroches and J.Y. Marzin, J. Crystal Growth **127**, 536-540 (1993).

7. Y.C. Kao, F.G. Celii and H.Y. Liu, J. Vac. Sci. Technol. B **11**, 1023-1026 (1993).

8. J.F. Zheng, J.D. Walker, M.B. Salmeron and E.R. Weber, Phys. Rev. Lett. **72**, 2414-2417 (1994).

9. M.C. Joncour, M.N. Charasse and J. Burgeat, J. Appl. Phys. **58**, 3373-3376 (1985).

10. O. Brandt, K. Ploog, L. Tapfer, M. Hohenstein, R. Bierwolf and F. Phillipp, Phys. Rev. B **45**, 8443-8453 (1992).

11. B.A. Philips, A.G. Norman, T.Y. Seong, S. Mahajan, G.R. Booker, M. Skowronski, J.P. Harbison and V.G. Keramidas, J. Crystal Growth **140**, 249-263 (1994).

12. R.K. Graupner, J.A.V. Vechten and P. Harwood, J. Vac. Sci. Technol. B **10**, 1852-1855 (1992).

13. J.V. Vechten, U. Schmid and Q.S. Zhang, J. Electron. Mater. **20**, 431-439 (1991).

QFT CONTROL OF AN ADVANCED TACTICAL FIGHTER AEROELASTIC MODEL

Joseph C. Slater
Assistant Professor
Department of Mechanical and Materials Engineering

Wright State University
3640 Colonel Glenn Highway
Dayton, OH 45435
jslater@valhalla.cs.wright.edu

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

September 1994

QFT CONTROL OF AN ADVANCED TACTICAL FIGHTER AEROELASTIC MODEL

Joseph C. Slater
Assistant Professor
Department of Mechanical and Materials Engineering

Abstract

This report discusses Quantitative Feedback Control (QFT) of an aeroelastic model typical of advanced tactical fighter aircraft. The aeroelastic model is high order, unstable, and nonminimum-phase. Control is attempted on the original plant, the partially dynamically inverted plant, and the dynamically inverted plant. Application of the Nyquist stability criterion using Nichols plots is discussed and suggestions for future efforts are given.

QFT CONTROL OF AN ADVANCED TACTICAL FIGHTER AEROELASTIC MODEL

Joseph C. Slater

Introduction

Currently, flight control used in fighter aircraft is based on classical frequency domain methods (i.e. PID with notch filters, band pass filters, lead/lag control) and is scheduled throughout the flight conditions of the aircraft. The dynamics of the aircraft are assumed to be linear (or close enough to linear that linear control techniques can be applied) within each scheduled region of the flight envelope. If enough flight conditions are used in the scheduling, the assumption of linearity remains reasonably valid, and care is taken to assure that the controlled aircraft dynamics change gradually through the flight envelope, then control scheduling is transparent to the pilot and flying qualities are preserved throughout the flight envelope under nominal flight conditions. However, if any of these assumptions are violated or the aircraft is damaged in combat, the controller/aircraft system can become unflyable, resulting in severe performance degradation and/or possible total system failure[1,4]. For this reason, robust control of fighter aircraft has been the focus of much attention in recent years.

At this time, three robust control methodologies are being developed and tested by control groups for possible use in flight control: H_∞, Quantitative Feedback Theory (QFT), and variable structure control. Both H_∞ and QFT are frequency domain (and thus linear) control methods while variable structure control is a nonlinear control technique. For a comparison between H_∞ and QFT the reader is referred to Bailey, Helton, and Merino[5]. Slotine provides an introduction to variable structure (also known as sliding mode) control[9].

This report concentrates on attempts to apply QFT control to flight control of aeroelastic models (provided by Max Given of Northrop) typical of a modern advanced fighter configuration. The model includes both rigid body and flexible modes. Actuator and sensor models were also provided, but were neglected due to their high bandwidth relative to the desired flight controller bandwidth.

Problem Description

The goal of the problem is to design a robust closed loop control that stabilizes the plants provided while providing sufficient tracking bandwidth and "reasonable" closed loop disturbance rejection. Sufficient tracking bandwidth was considered to be above 15 rad/sec while disturbance rejection criterion were not defined explicitly. Four plant cases (provided by Max Given of Northrop) were provided for both lateral and longitudinal motions. The cases are as follows:

Case	Mach no.	Alt. (ft)
1	.4	sea level
2	.9	20,000
3	1.2	20,000
4	1.6	40,000

For the lateral case, the inputs in order are torques (in-lb) to the tail (rudder), inboard flap, outboard flap, and leading edge flap. Outputs are: tail displacement (deg), inboard flap displacement (deg), outboard flap displacement (deg), leading edge flap displacement (deg), side slip (deg), side velocity (ft/sec), roll rate (deg/sec), yaw rate (deg/sec), y acceleration at the aerodynamic center (g's), y acceleration at the center of gravity (g's), roll angle (deg), and yaw angle (deg). For the longitudinal case, the single input is torque on the tail (elevator). Outputs are: tail displacement (deg), angle of attack (deg), true velocity (ft/s), pitch rate at station 1 (deg/sec), pitch rate at station 2 (deg/sec), vertical acceleration at the aerodynamics center (g's), vertical acceleration at the pilot station (g's), vertical acceleration at the center of gravity (g's), pitch (deg), forward velocity (ft/s), and vertical velocity (ft/s). The sign convention used is that of Blakelock[6].

It is common in flight control for inputs to be control surface deflections as opposed to torques on the surfaces. Here the inputs are torques as a result of the model being aeroelastic. In a real fly-by-wire aircraft, the command signal is passed as a voltage to an actuator (which are usually hydraulic, but may be piezoceramic (smart actuators), electric...). The actuator, in turn, applies a torque to the control surface. The surface deflection is a result of the applied control torque and aerodynamic loading and flexibility of the surface. Its aerodynamic effectiveness is a results of the surface deflection combined with elastic deflection of the surrounding substructure. Higher dynamic

loading and lower flexural stiffness cause reduced control surface deflection, contributing to loss in control surface effectiveness. The inputs to the state space model are the torques acting on the control surfaces. Transfer function models of the actuators can be used to model the dynamics responses of the actuators (giving torque/control voltage). However, the actuator models were not available in time for implementation in the control design, and the bandwidth of the actuators was later observed to be well above that of the desired controller bandwidth.

The outputs of the state space model can thus be chosen to be the aircraft rigid body motions, stresses or strains in the aircraft, as well as control surface deflections. For the purpose of flight control, however, only certain outputs (such as rigid body motions) are physically feasible outputs to design flight control around. However, smart structure technology is heading in the direction of being able to provide precise information with respect to aircraft structural motion (modes of vibration), allowing stress and strain constraints to be directly incorporated into the control law.

The plants provided by Northrop are "...typical of an advanced fighter configuration." (Max Givens). The longitudinal cases are 64 state models, while the lateral cases are 70 state models. Pole-zero maps between tail torque and pitch rate for each plant are shown below. As can be observed, the plants are nonminimum phase and have unstable open-loop poles. Many of the poles are marginally stable, making it likely that instability will result from closed loop control if the control is not sufficiently robust.

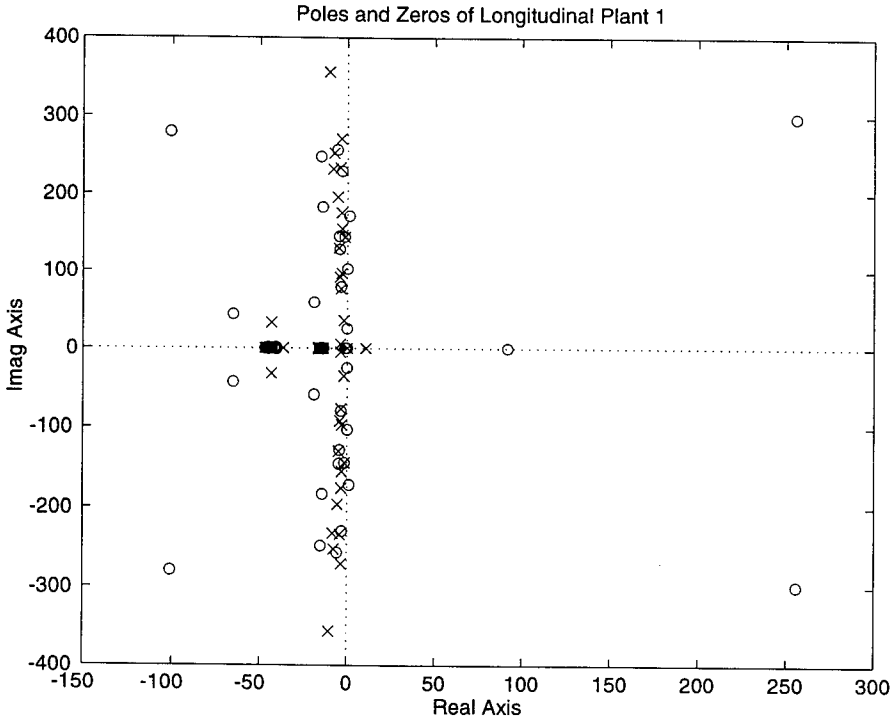


Figure 1: Pole-zero plot of plant 1 between elevator torque and pitch rate.

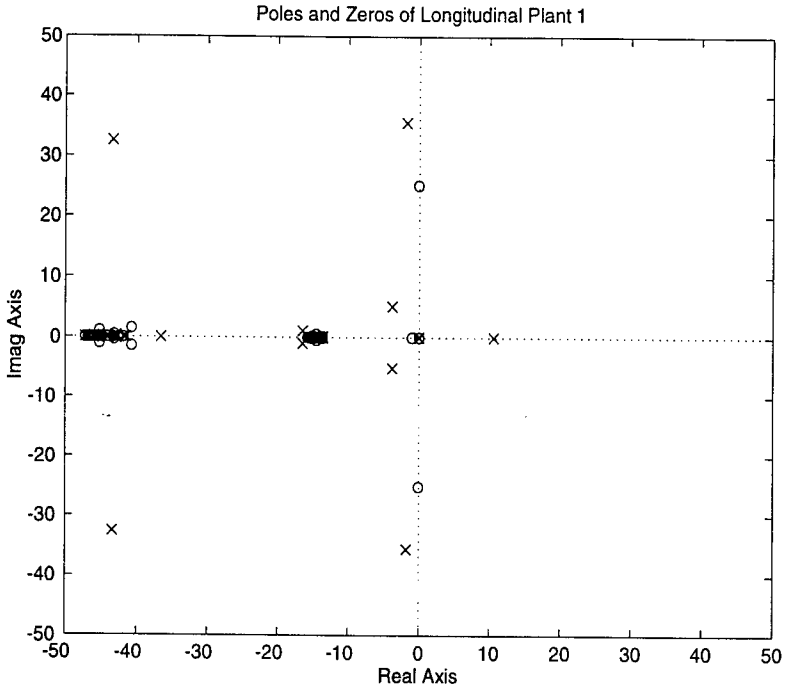


Figure 2: Pole-zero plot of plant 1 between elevator torque and pitch rate (low frequencies).

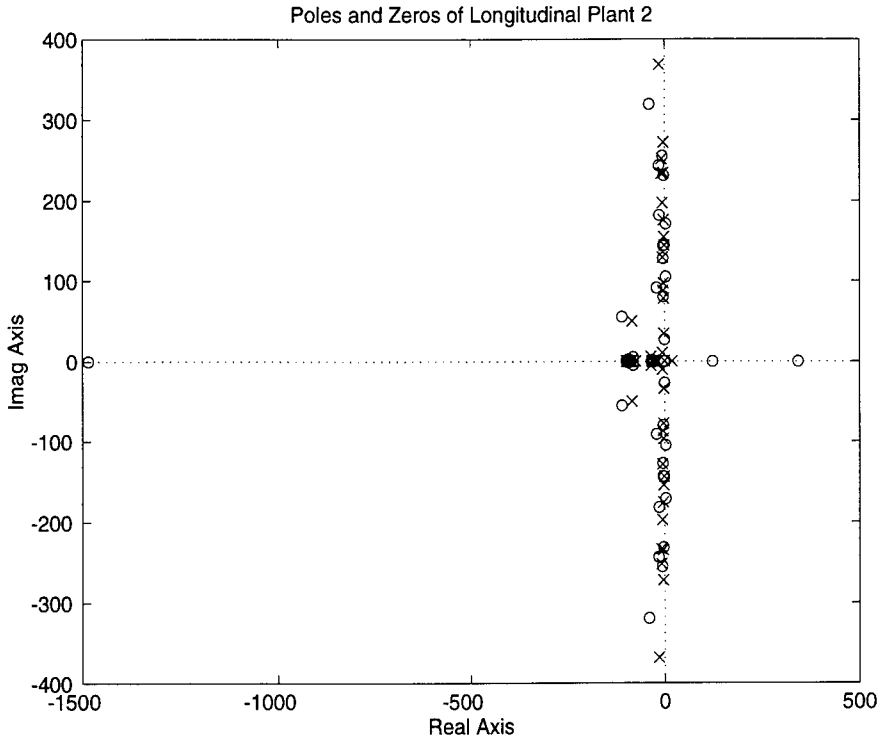


Figure 3: Pole-zero plot of plant 2 between elevator torque and pitch rate.

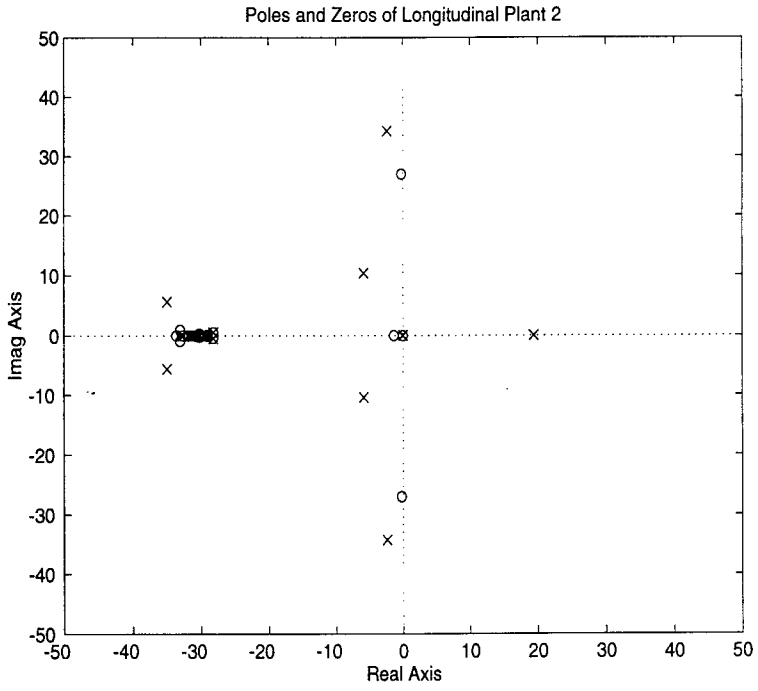


Figure 4: Pole-zero plot of plant 2 between elevator torque and pitch rate (low frequencies).

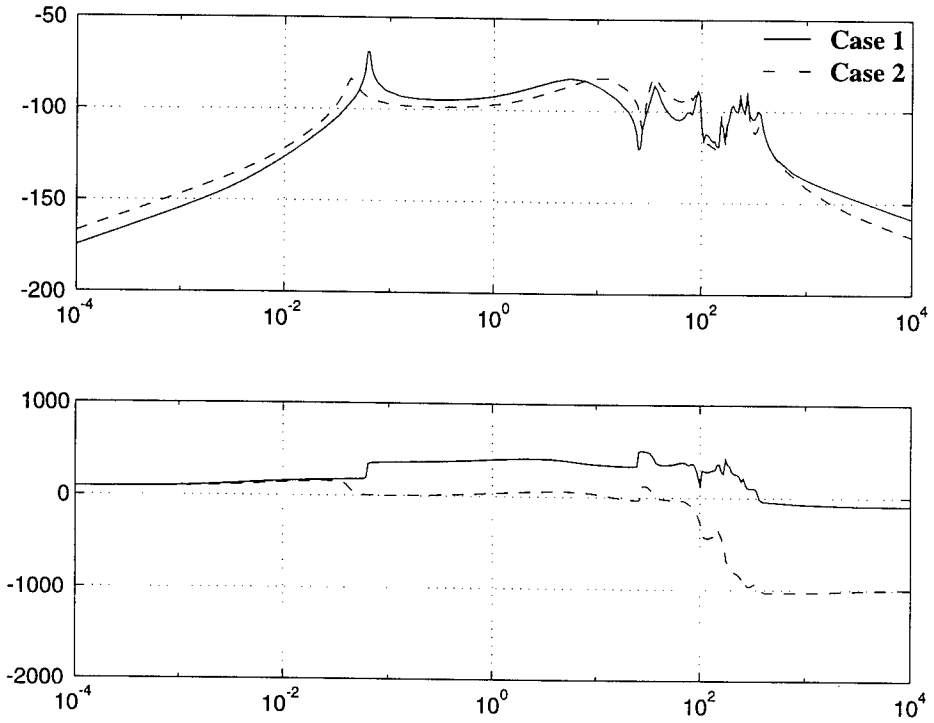


Figure 5: Bode plot of plants between elevator torque and pitch rate.

Approach

At high frequencies, structural modes must be considered since surface deflections can begin to show excessive response (e.g. resonance) or limit cycles (e.g. flutter). Conventional aircraft and control design attempts to place structural modes at high frequencies and design the flight control with a low enough bandwidth such that the flight control system cannot significantly excite structural modes. However, this is not always possible. The F-18 has been shown to have significant vibrational problems under certain operating conditions, requiring the redesign of the flight control system to alleviate undesirable vibrations. Notch filters are often used at frequencies of structural mode shapes in order to filter out energy in the control system at that frequency. These "piecemeal" control methods run the risk of causing instability and poor performance in the face of small variations in the aircraft (launching of munitions, emptying of storage tanks, damage...) and flight conditions (altitude, mach number...). The goals of this control law design are to partially account for variation in flight conditions by using robust control (thus minimizing the need for control scheduling and guaranteeing stability under a set of flight conditions) and the inclusion of flexible modes.

Increased performance demands and increased flexibility will bring structural effects into the desired bandwidth of the flight control, adding to the likelihood of instability if structural modes are not accounted for.

QFT control appears to be the better linear robust control approach to take for this type of problem since H_∞ would produce a very high order controller. However, a stabilizing QFT controller was not found using the QFT CAD package[11]. The procedure used in this design was identical to the design procedure outlined by Houpis. A description of how to design QFT control can be found in Houpis[10] and D'Azzo and Houpis[12]. A number of contributing factors were involved in preventing a successful controller design. First, the large order of the system caused the CAD process to be slow simply due to the large amount of computational effort involved. Second, the CAD package was not designed to deal with high order plants. Although no limit in plant size is intrinsic to the code, modifications to the code could greatly ease its use on larger order problems. Third, the number of unstable open loop poles made the design process difficult. It is not intuitively obvious in QFT how to shape the Nichols plot to stabilize an unstable system. Although Cohen, Chait, Yaniv, and Borghesani[7] demonstrate stability analysis using Nichols plot, they do not describe how to design for stability. There is also no example of an open loop unstable system stabilized by closed loop control. Referring to their paper, theorem 3 states that a closed loop system is stable if "the one-sheeted Nichols plot of $L(s)$ does not intersect the point $\mathbf{q} := (-180°, 0\text{dB})$ and the net sum of its crossings of the ray $R_0 := \{ (\phi, r') : \phi = -180°, \ r' > 0\text{dB} \}$ is equal to n" (n being the number of unstable poles)[7]. This definition, however, is not sufficient in that it does not take into account the possibility of poles on the imaginary axis, and according to Kuo[8], is insufficient in the case of non-minimum phase systems.

Results

Since a satisfactory design was not found, efforts were taken to provide a future foundation for designing QFT control for typical plants. The first part was to determine how to shape the Nichols plot of an open-loop unstable system such that the close loop system will be stable. The remainder of the results are suggestions for finding a nominal open loop control law, and possible improvements to the CAD package that would aid in designing/redesigning compensators for large order systems.

Since the Nyquist and Nichols chart of any given function can each be found from the other plot, it seems intuitive that although the Nyquist stability criterion is based on the use of the Nyquist chart, one should be able to apply it using the Nichols chart. A brief explanation of how this can be done is given in the following. A more sound derivation has been left for future work.

The Nyquist stability criterion can be stated as: "the closed-loop system is stable if

$$\Phi = -(0.5P_{\omega} + P_{-1})\,180° \tag{1}$$

is true."[8] Here P_{ω} is the number of poles of $L(s)$ on the imaginary axis and P_{-1} is the number of unstable poles of $L(s)$. "The angle Φ is the angle traversed by the Nyquist plot (couterclockwise) of $L(s)$ with respect to the $(-1,j0)$ point..." as $j\omega$ goes from ∞ to 0. Taking $j\omega$ from 0 to ∞ gives the negative expression. Example 10-2 of Kuo is a non-minimum phase system with a pole on the imaginary axis. This is very similar to the type of plant we are dealing with, so a brief overview will be given here, and then the results will be correlated to the Nichols plot.

Consider a system with the open loop transfer function $L(s) = \frac{K(s-1)}{s(s+1)}$. Here P_{ω}=1 and P_{-1}=0. Hence, the closed-loop stability requirement is

$$\Phi = -(0.5P_{\omega} + P_{-1})\,180° = -90° \tag{2}$$

Observing Figure 6 one can see that the phasor drawn from the point $(-1,j0)$ to the Nyquist plot of $L(s)$ traverses positive 90° as ω goes from ∞ to 0. Thus, the closed loop system is unstable. Viewing Figures 7 and 8, one can see that of the three plots shown, the only value of K that causes the system to be closed loop stable is -.5. The actual range of values of K for which the system is stable are $-1<K<0$.

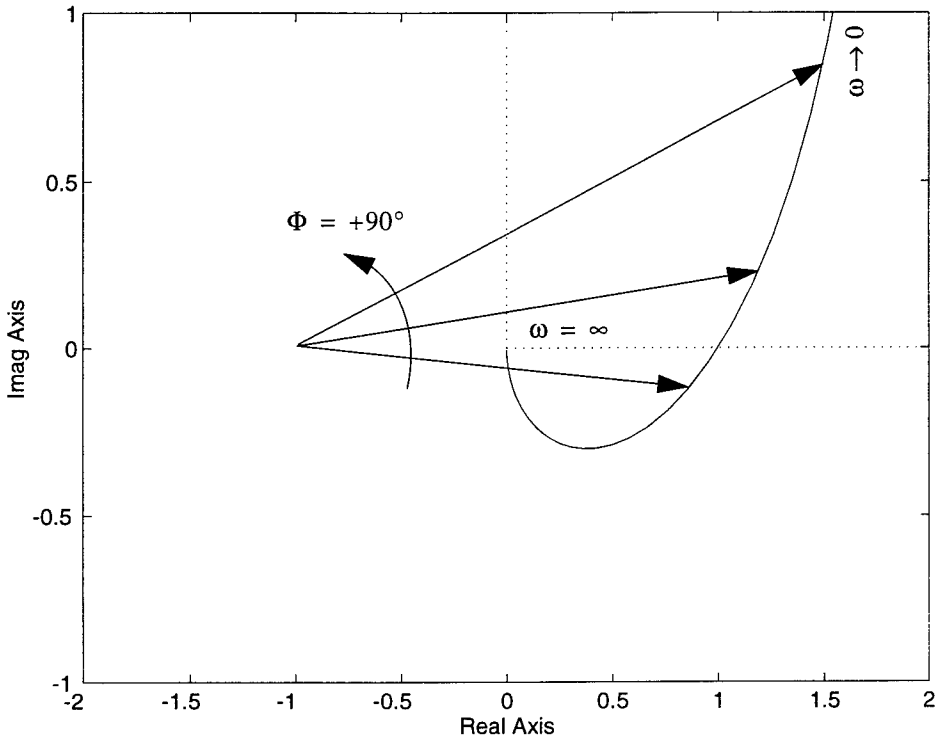


Figure 6: Nyquist plot of $L(s) = (K(s-1))/(s(s+1))$ with K=1 (unstable).

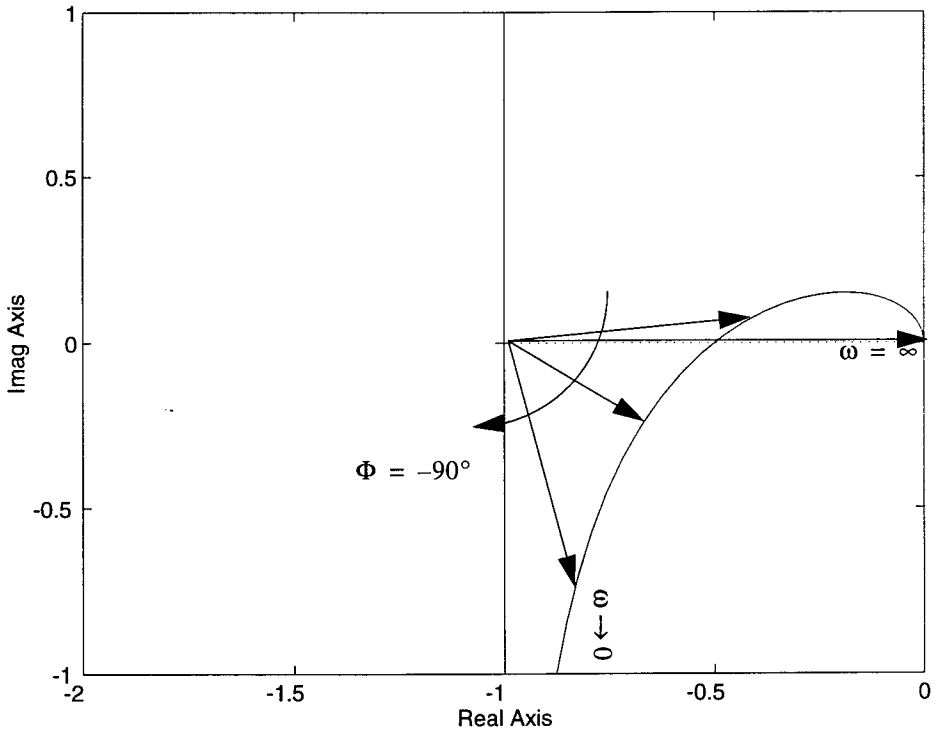


Figure 7: Nyquist plot of $L(s) = (K(s-1))/(s(s+1))$ with K=-.5 (stable).

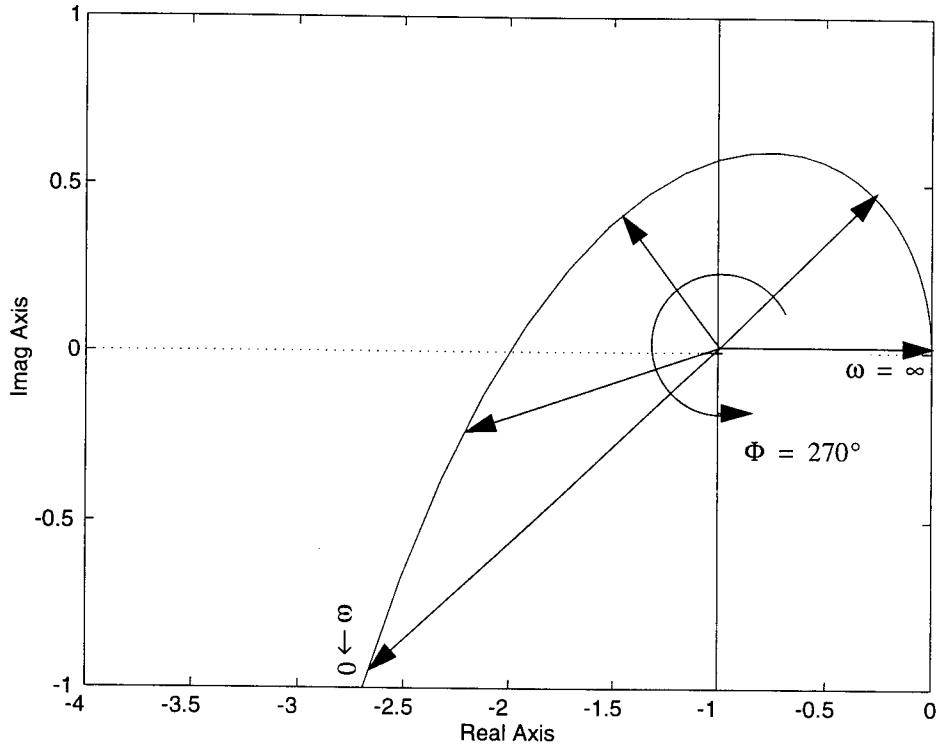


Figure 8: Nyquist plot of $L(s) = (K(s-1))/(s(s+1))$ with K=-2 (unstable).

The next three figures correspond to the previous three in that they are the same system with corresponding gains, but on a Nichols plot. The angle Φ can be found from the plots using the following guidelines. Here we define positive direction along the open loop plot to go from $j\omega$=0 to $j\omega$=∞. These guidelines simply correlate "similar" parts of the Nichols and Nyquist charts to one another by noting that once the magnitude of the open loop transfer function is less than 0 dB, the angle traversed about the point (-1,j0) on the Nyquist chart is zero, with the final angle at high frequencies going to 0° (or a multiple of 360°) assuming the magnitude decreases to 0 in the limit as $j\omega$ goes to 0.

First, find the open loop phase angle at $j\omega$=0. Call this Φ_1. Second, find the phase angle (Φ_2) at an open loop amplitude of 0 dB. From this point, go to the nearest open loop phase of 0° or a multiple of 360° (positive or negative). Call this angle Φ_3. Make sure to use a "multiple sheet" Nichols chart or keep track of how many times you pass through the plot from right to left. The angle Φ can be found by the equation

$$\Phi = \Phi_1 - \Phi_3 \tag{3}$$

For example, observing Figure 9 for K=1, the initial open loop phase is $\Phi_1 = 90°$. This is found by "unwrapping" the Nichols chart so that the plot does not jump from one side of the graph to the other. When the open loop magnitude is 0 dB, the open loop phase is $\Phi_2 \approx 0°$. The nearest multiple 360° or 0° is 0° itself. Thus $\Phi_3 = 0°$ and the total angle traversed is $\Phi_1 - \Phi_3 = 90°$. From equation (2), this system is unstable.

Observing Figure 10 for K=-.5, the initial open loop phase is $\Phi_1 = -90°$. When the open loop magnitude is 0 dB, the open loop phase is $\Phi_2 \approx -135°$. The nearest multiple 360° or 0° is again 0° itself. Thus $\Phi_3 = 0°$ and the total angle traversed is $\Phi_1 - \Phi_3 = -90°$. From equation (2), this system is stable.

Observing Figure 11 for K=-2, the initial open loop phase is $\Phi_1 = -90°$. When the open loop magnitude is 0 dB, the open loop phase is $\Phi_2 \approx -225°$. The nearest multiple 360° or 0° is −360°. Thus $\Phi_3 = -360°$ and the total angle traversed is $\Phi_1 - \Phi_3 = -90° + 360° = 270°$. From equation (2), this system is unstable.

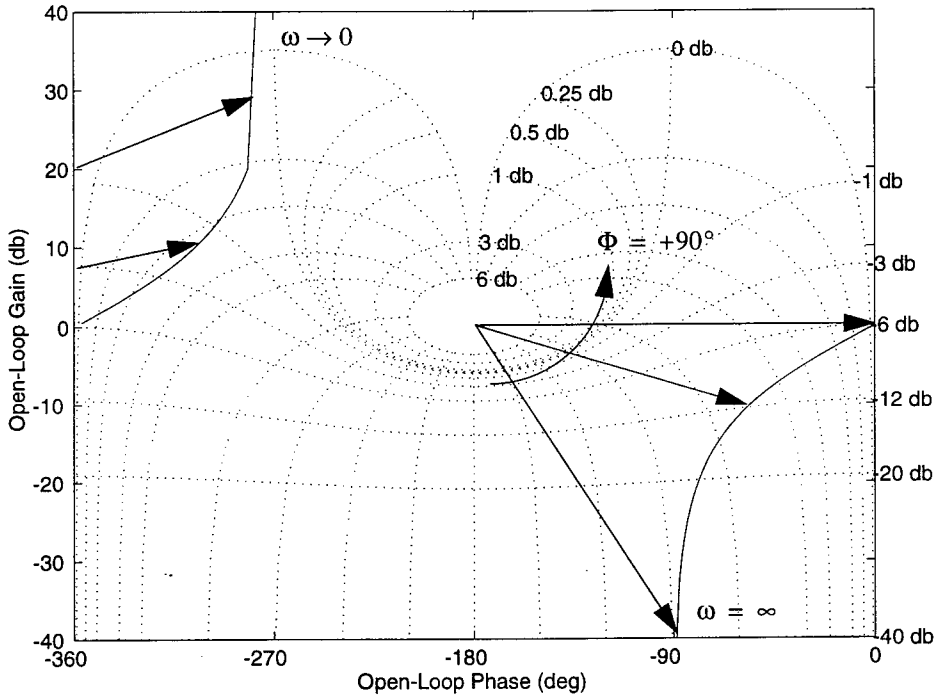


Figure 9: Nichols plot of L (s) = (K (s − 1)) / (s (s + 1)) with K=1 (unstable).

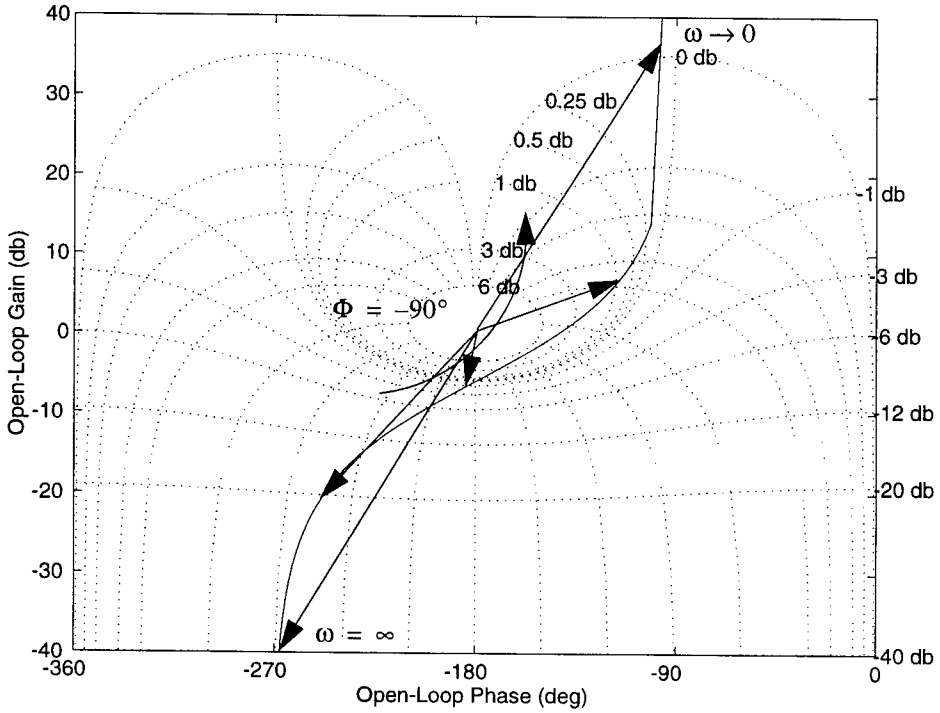


Figure 10: Nichols plot of L (s) = (K (s – 1)) / (s (s + 1)) with *K*=-.5 (stable).

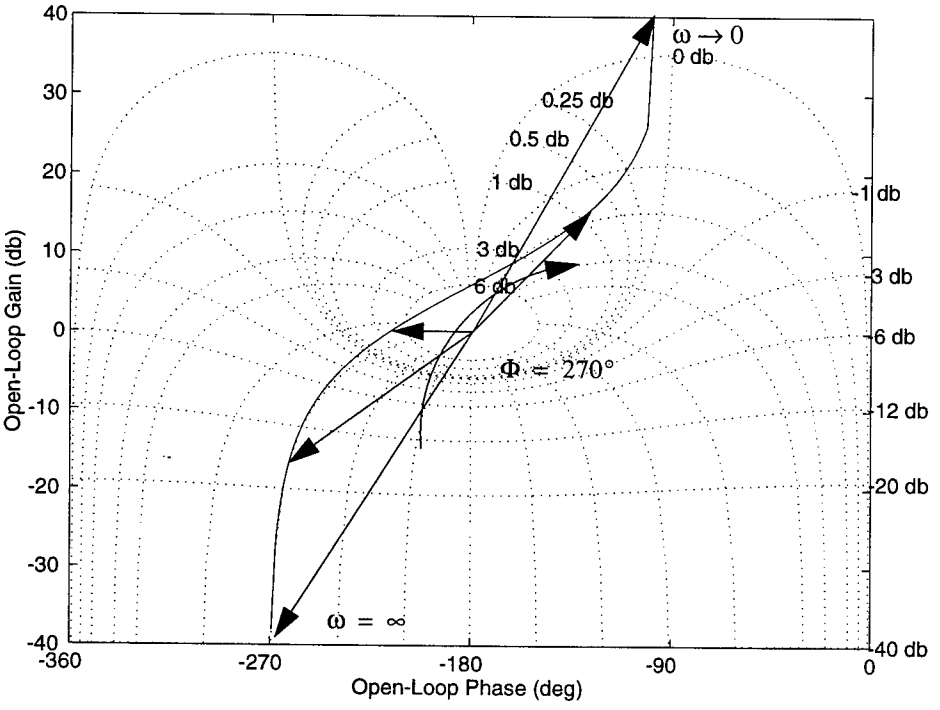


Figure 11: Nichols plot of L (s) = (K (s – 1)) / (s (s + 1)) with *K*=-2 (unstable).

Consider a second example with one unstable pole. The objective here is to determine how to modify the open loop transfer function so that the closed loop transfer function will be stable. The open loop transfer function is given by $L(s) = k(s+1)/(s^2+9s-10)$. Equation (1) states $\Phi = -(0.5P_\omega + P_{-1})\,180°$ must be true for the system to be stable. The open loop poles are -10 and 1. Since their are no poles on the $j\omega$ axis and there is one unstable pole, $P_\omega = 0$ and $P_{-1} = 1$. Thus, for a stable closed loop system, we require that $\Phi = -180°$. The Nichols plot of $L(s)$ is shown in Figure 12 for a gain of k=10. The initial phase at ω=0 is $\Phi_1 = -180°$. Of course this is only true if the gain is raised slightly such that the initial magnitude is greater than 0 dB. The nearest multiple of 360° when the magnitude of $L(s)$ is 0 dB is 0°. Thus, if the gain is greater than 10, $\Phi_1 - \Phi_3 = -180° + 0° = 180°$ as required for stability. The root locus of Figure 13 shows this to be the case.

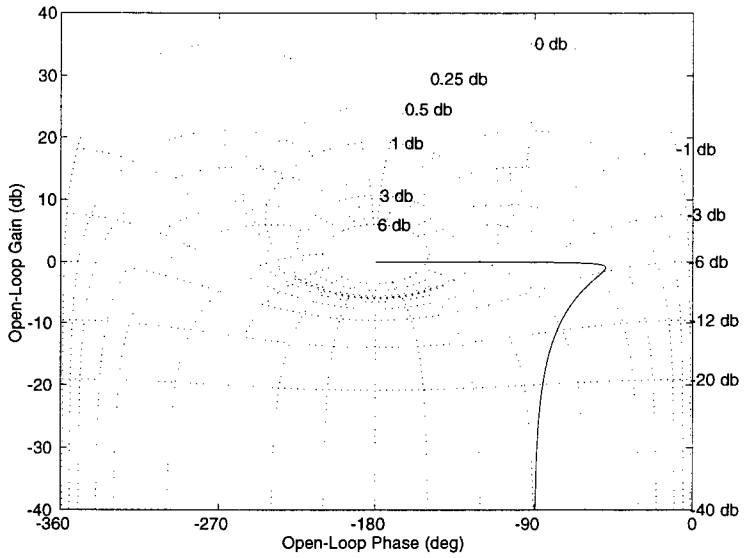


Figure 12: Example 2, stabilizing an unstable system ($L(s) = k(s+1)/(s^2+9s-10)$).

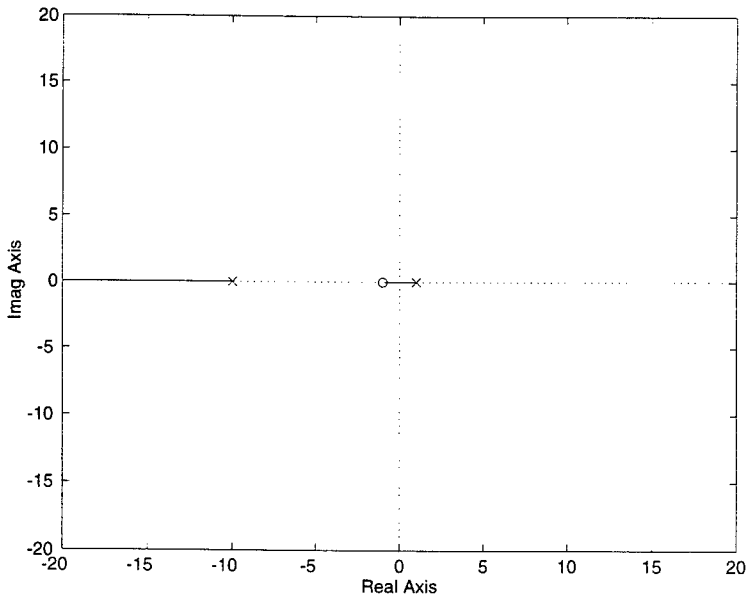


Figure 13: Root Locus of $L(s) = k(s+1)/(s^2+9s-10)$.

For aeroelastic plant 1, three unstable poles exist: 10.62, and $0.00042 \pm 0.0628i$. Whether or not the complex pair is really unstable (or just a numerical result) is of significant importance because plant 2 has only 1 unstable pole. From Figure (2), one can see that they are close enough to zero that their practical effect is two poles at the origin. Figure (4) shows that plant two has two very similar poles as well. Observing the Bode diagram (Figure 5), however, shows that the phase effects of these two pairs of poles are distinctly different. From a practical standpoint, we are only interested in stabilizing the higher frequency pole of each plant. How to stabilize specific unstable poles using QFT is still unknown and is an important area of future study. Considering for the moment that this problem can be handled (we now have only 1 unstable pole for each plant), equation (1) says that for stability $\Phi = \Phi_1 - \Phi_3 = -180°$. Since the initial phase is at a multiple of $360°$, and we need the total phase change to be $-180°$, two zeros will need to be added at the origin and the phase will have to be wrapped to the right around the set of M-curves directly below the start (ω=0) of the Nichols plot. How to do this is a matter of learning (better yet "discovering") how to preform loop shaping on the Nichols chart[13,14].

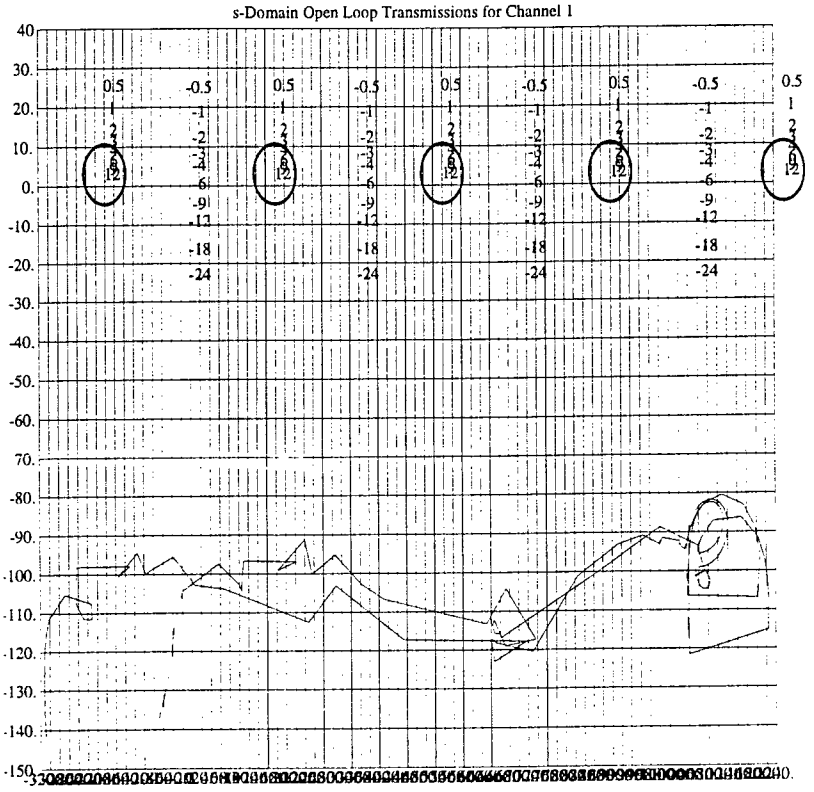


Figure 14: Nichols Plot of the Open Loop Aeroelastic Aircraft Model

Future Efforts

The high order, and thus high degree of complexity, of the plants made design of even a stabilizing compensator extremely difficult. The number of right-half plane zeros makes loop shaping difficult due to the drastic phase changes (see Figure 5). Because of this, it is suggested that a nominal stabilizing design be found for the nominal plant using LQG design and then optimized on the Nichols chart for robustness using QFT. It has also been suggested by some that H_∞ be used as a first cut at the control design and then QFT be used to further optimize the compensator by reducing the order of the compensator and reducing unnecessary robustness intrinsic to H_∞ [5].

As stated earlier, the Sating QFT /CAD package[11] is not designed specifically with high order systems and high order compensators in mind. A few minor flexibility improvements would greatly enhance the package's ease of use for larger systems. Redrawing the Nichols plot after each modification of the compensator and automatically upon entering the compensator design module is a slow and computationally expensive procedure when the phase changes

drastically from low to high frequency (Figure 14) and the order of the compensator and plant are very high. The ability to set an option disabling this feature would be a great time saver in these cases. Also, one of the easiest ways to start design of a compensator is to begin with the Bode diagram, either to invert the plant, or get a good first design ignoring robustness. Having a menu choice that would enable the automatic generation of the open loop Bode plots would save quite a few key strokes, creating the same ease of use that the automatic Nichols plot generation does for loop shaping on the Nichols plot.

The addition of explicit lead/lag compensators would aid in the design process by allowing the designer to modify parameters such as the center frequency and the spread of the compensator. The ability to tie together a pole and zero in this manner can help when dealing with higher order compensators.

The range command, for the range of the open loop plant/compensator data to be plotted, is now restricted to the range of the templates. With a high order plant, the Nichols plot of the open-loop data can wrap on top of itself many times, making it look like a wad of yarn. It would be helpful to be able to see these smaller ranges.

When adding poles and zeros, the entire compensator is redisplayed. It would be helpful to be able to turn this feature off so that the menu remains in sight and you do not have to wait for the entire compensator to scroll by on slower machines or over the network.With a 30th order compensator, the screen is completely taken up by the compensator, and the menu is gone from the screen.

Conclusions

A QFT control law design was attempted for a high order aeroelastic model typical of modern fighter configurations. Compensator design was attempted on the original plant, the partially inverted plant, and the identically inverted plant (without inverting poles and zeros in the right half plane). Difficulty in dealing with such an unwieldy plant model and lack of a method for determining how to stabilize an open-loop unstable plant prevented a successful controller design. A method was presented that will allow the future design of stabilizing control for unstable non-minimum phase plants. Suggestions for future efforts with high order plants are given, and recommendations are given for improvement of Sating's QFT/CAD package to improve design flexibility.

References

[1] R. Y. Chiang, M. G. Safonov, K. Haiges, K. Madden, and J. Tekawy, "A fixed H_∞ controller for a supermaneuverable fighter performing the Herbst maneuver," *Automatica*, vol. 29, pp. 111-127, Jan. 1993.

[2] M. G. Safonov and R. Y. Chiang, "CACSD using the state-space L^∞ theory–a design example," *IEEE Transactions on Automatic Control*, vol. AC-33, pp. 477-479, May 1988.

[3] M. G. Safonov, A. J. Laub, and G. Hartmann, "Feedback properties of multivariable systems: the role and use of the return difference matrix," *IEEE Transactions on Automatic Control*, vol. AC-26, pp. 47-65, Feb. 1981.

[4] B. T. Clough, "Longitudinal handling qualities approximations and bandwidth minimization for robust automatic flight control system design using quantitative feedback theory," AFWAL-TR-86-3107, Appendix D, 1987.

[5] F. N. Bailey, J. W. Helton, O. Merino, "Alternative approaches in frequency domain design of single loop feedback systems with plant uncertainty," *Proceedings of the American Controls Conference*, 1994.

[6] J. H. Blakelock, *Automatic Control of Aircraft and Missiles*, 2nd ed., John Wiley & Sons, New York, 1991.

[7] N. Cohen, Y. Chait, O. Yaniv, and C. Borghesani, "Stability analysis using Nichols charts," *Quantitative Feedback Theory Symposium Proceeding*, August, 1992

[8] B. C. Kuo, *Automatic Control Systems*, 6th ed., Prentice-Hall, New York, 1991.

[9] J. J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice Hall, New York, 1991.

[10] C. H. Houpis, "Quantitative feedback theory (QFT), technique for designing multivariable control systems," AFWAL-TR-86-3107, January, 1987.

[11] R. Sating, "MIMO/QFT CAD Program user's manual," Air Force Institute of Technology, ed. by C. H. Houpis and M. Pachter, 1993.

[12] J. J. D'Azzo and C. H. Houpis, *Linear Control System Analysis and Design: Conventional and Modern*, 3rd ed., McGraw-Hill, New York, 1988.

[13] D. F. Thompson and O. D. I. Nwokah, "Analytic loop shaping methods in quantitative feedback theory," *Journal of Dynamics Systems, Measurement, and Control*, vol. 116, pp. 169-177, June 1994.

[14] F. N. Bailey and J. C. Cockburn, "Loop gain phase shaping design for SISO robust controllers," *Proceedings of the American Controls Conference*, Boston, 1991.

PERFORMANCE ANALYSIS OF QUADRATIC CLASSIFIERS FOR SYNTHETIC APERTURE RADAR TARGET RECOGNITION

John A. Tague
Associate Professor
Department of Electrical and Computer Engineering

Ohio University
Athens, OH 45701

Final Report for:
Summer Faculty Research Program
Wright Laboratories

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

and

Wright Laboratories

August 1994

PERFORMANCE ANALYSIS OF QUADRATIC CLASSIFIERS FOR SYNTHETIC APERTURE RADAR TARGET RECOGNITION

John A. Tague
Associate Professor
Department of Electrical and Computer Engineering
Ohio University

Abstract

We study the performance of quadratic pattern classification algorithms as applied to a prototype problem in automatic target recognition. We generated synthetic aperture radar images of four vehicles using the Xpatch electromagnetic scattering code and a SAR image formation algorithm. Each class contained eleven images of the target appearing at slightly different azimuths with respect to a fixed reference direction. We used the images to train a quadratic classifier algorithm, and we evaluated its performance when a test image was distorted by random pixel fluctuations. Several variations of the quadratic classifier were tested. We compare their performance against one another. We found that an adaptive gain factor, used to scale the test images before classification, improved the classifier's performance significantly.

PERFORMANCE ANALYSIS OF QUADRATIC CLASSIFIERS FOR SYNTHETIC APERTURE RADAR TARGET RECOGNITION

John A. Tague

I. Introduction

In this paper we study the performance of a quadratic pattern classification algorithm as applied to a problem in automatic target recognition. Before we describe the specific issue addressed in this report, let us define the target recognition problem in general terms and break down its solution into its generic steps.

The objective of automatic target recognition is to examine a volume of space and identify vehicles, aircraft, and other military assets from electromagnetic sensor data. More often than not, an image of the test region is formed before recognition is performed. Optical systems, infrared systems, and radar are used to probe the test region. When radar from a moving platform is used to probe the test region, its motion can be exploited to improve image resolution by way of aperture synthesis techniques; they are called synthetic aperture radar (SAR) systems [1]. In this report, we will focus on processing SAR images by algorithms designed to recognize objects within the field of view.

The automatic target recognition process can be decomposed into four generic steps. The first step is called *data preprocessing.* The signals measured by the sensors are amplified, filtered, and calibrated. In addition, if imaging is performed, it is implemented in this stage.

The second step is the *detection* step. An algorithm is applied to the image to locate bright spots called regions of interest which may contain a target. The regions of interest, also called image chips, are located and extracted from the scene.

The third step is the *discrimination* step. The goal of discrimination is to seperate targets from non-targets. This is a delicate step at radar wavelengths; returns from the ground, from trees, and from other vegetation are often similar to those from targets. The false returns, usually called clutter, are difficult distinguish from target-like returns. After discrimination, those chips classified as target-like are sent to the next stage.

The final step is the *identification* step. Each target-like chip is processed by a pattern recognition algorithm which extracts characteristic features from the image and compares them to templates stored in a feature catalog. The output of the identifier is a list of those targets which best match the image chip. Sophisticated algorithms can glean more information from the image chip; for example, they can provide a list of most likely targets annotated by correct choice probabilities.

In this paper we focus on the identification problem. We will assume that the image chips have made

it through the preprocessing, detection, and discrimination stages. The scenario under consideration is as follows. A moving platform such as an aircraft illuminates a fixed test region on the ground. We can assume that the aircraft is circling the test region at a fixed altitude. The radar antenna measurements are stored and processed by a spotlight SAR image formation code. The detection and discrimination algorithms are applied to the image, and a region of interest is identified. The chip is input into the classifier stage of the ATR system.

The goal of the classifier is to process the image chip and decide among several trial target classes. Usually, the target's orientation within the chip is unknown. In addition, sensor noise, target occlusion, and target articulation compound our ability to classify the chip. These problems must be taken into account in the classifier design.

We will apply quadratic classifiers to the SAR image recognition problem [2] [3]. This algorithm can be derived from first principles of statistical decision theory, and it has already been applied to other kinds of pattern recognition problems. When it is brought to bear on SAR image problems, certain *ad hoc* modifications to the basic algorithm can be used to enhance its performance [4]. In this report, we will compare several data preprocessing schemes and algorithm modifications against each other and against competing types of pattern recognition techniques.

This report is organized as follows. In Section II, we will describe the signal models and the basic quadratic classifier algorithm. In Section III, we describe various modifications to the basic algorithm and describe how the classifier is "taught" to recognize patterns. In Section IV, we describe the data set which was used to test the classifier. The experimental design is described in Section V. The results are presented and critiqued in Section VI, and our conclusions appear in Section VII.

II. The Quadratic Classifier Algorithm

The quadratic classifier is a solution to the *statistical* pattern recognition problem [3]. This approach to classification accounts for the statistical properties of pattern classes in the design process. What does this mean within the context of SAR target recognition? To answer this question we point out that not all of the target characteristics are known in advance. For example, in practical problems, the target's position and orientation within the chip are unknown. Small changes in target orientation can produce large variations in a target's SAR image. Furthermore, we must account for perturbations in the image produced by sensor and electronic noise. Taken together, these effects introduce uncertainty into the test images which must be considered by the algorithm designer.

How do we pose the classification problem statistically? Let us represent the image chip by the column

vector $\mathbf{x}$. This vector is constructed by stacking the image pixels column by column into $\mathbf{x}$. Suppose there are M targets and let ω_i denote the ith target class. The conditional probability that $\mathbf{x}$ belongs to ω_i is $p(\mathbf{x}|\omega_i)$. Now let us describe the Bayesian approach to the problem. We are given the prior probabilities $p(\omega_1)$, $p(\omega_2)$, and so on; $p(\omega_i)$ = the probability that the true target class is class i. We then define a loss matrix whose elements $\{L_{ij}\}$ define the cost of deciding that the image is target j when its true identity is target i. The performance metric is the conditional average risk function $r_j(\mathbf{x})$ which measures the loss incurred by the classifier when it decides that $\mathbf{x}$ is target j:

$$r_j(\mathbf{x}) = \sum_{i=1}^{M} L_{ij} p(\omega_i|\mathbf{x}). \tag{1}$$

The classifier must choose one of M alternatives. If it computes $r_1(\mathbf{x}), r_2(\mathbf{x}), \ \ldots \ , r_M(\mathbf{x})$, and then assigns the image to that class with the smallest conditional risk, the total average loss will be minimized. It can be proven that decision rule reduces to computing the decision functions

$$d_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)p(\omega_i) \tag{2}$$

for $1 \leq i \leq M$. The image or pattern $\mathbf{x}$ is classified as target i if $d_i(\mathbf{x}) > d_j(\mathbf{x})$ for all $j \neq i$. This is called the *Bayes* classifier; the quadratic classifier is a special case of this class of algorithms.

What is a quadratic classifier and under what conditions is it an optimum classifier? Suppose that $\mathbf{x}$ is Gaussian distributed with mean $\mathbf{m}_i$ and covariance matrix $\mathbf{C}_i$ when it belongs to the ith target class. Then

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_i|^{1/2}} \exp(-(\mathbf{x}-\mathbf{m}_i)'\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)), \tag{3}$$

where $|\cdot|$ means matrix determinant and the prime symbol denotes transpose. Then $d_i(\mathbf{x})$ is the product of equation (3) times the prior probability $p(\omega_i)$. But this is complicated to implement; an equivalent decision function which is easier to implement would be very useful. It is well known that any monotonicaly increasing function of $d_i(\mathbf{x})$ can be used instead of $d_i(\mathbf{x})$ itself; therefore, we can use the natural logarithm of $p(\mathbf{x}|\omega_i)$ without penalty. Taking the natural logarithm of equation (3) and substituting it into equation (2) yields

$$d_i(\mathbf{x}) = \log p(\omega_i) - \frac{n}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{C}_i| - \frac{1}{2}(\mathbf{x}-\mathbf{m}_i)'\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i) \tag{4}$$

for $1 \leq i \leq M$. Equation (4) can be simplified if we assume that the prior probabilities $p(\omega_i)$ are equal. Then the $\log p(\omega_i)$ terms are all the same and can be eliminated. In addition, $n/2 \log 2\pi$ can be eliminated because it is not a function of i; $d_i(\mathbf{x})$ then becomes

$$d_i(\mathbf{x}) = -\frac{1}{2}\log|\mathbf{C}_i| - \frac{1}{2}(\mathbf{x}-\mathbf{m}_i)'\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i). \tag{5}$$

The decision rule is to pick the largest (most positive) $d_i(\mathbf{x})$. If we assume that $\log|\mathbf{C}_i| \approx$ the same for all i, then maximizing $d_i(\mathbf{x})$ is equivalent to minimizing the quadratic form $(\mathbf{x}-\mathbf{m}_i)'\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)$ over the i target classes. This result, the quadratic classifier, is implemented as follows. First, compute the quadratic forms

$$d_i(\mathbf{x}) = (\mathbf{x}-\mathbf{m}_i)'\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i) \tag{6}$$

for $1 \leq i \leq M$. Pick the smallest result; this minimizes the Bayes risk. The image is identified with the smallest quadratic form.

III. Applying the Quadratic Classifier to SAR Imagery

A. Image Phase Information Is Discarded

In this section we describe how the quadratic classifier is used in the context of SAR image recognition. The images under consideration are "spotlight mode" SAR images formed by processing range profiles from a platform which circles the target scene. The output of the image synthesis algorithm is a matrix of complex numbers which are called "image domain" data. But should we use the complex data to classify targets? If not, what should we use?

Most researchers within the SAR target recognition community argue that the phase information in the complex SAR image should be discarded prior to classification. Phases are very sensitive to changes in range between the imaging sensor and the target's phase center. For example, when the imaging radar's center frequency is on the order of tens of gigahertz, very small range perturbations produce large changes in phase. The classifier would not be robust with respect to range errors if image phases were to be used. Therefore, only the magnitude information is used for SAR image classification.

When the image phases are discarded then the pixels are no longer Gaussian distributed. This in turn means that the quadratic classifier is no longer optimum. Nevertheless, it is still a useful algorithm because it computes weighted distances between the image chip $\mathbf{x}$ and the M image means $\mathbf{m}_i$. But if we could perform additional preprocessing on the pixel magnitudes such that the result were approximately Gaussian, then the algorithm would be nearly optimum *and* robust with respect to range errors. To this end, it has been proposed that "quarter power" data be used to classify SAR images [4]. The quarter power image is computed by taking the square root of the magnitude data pixel by pixel. One can argue that the quarter power data is approximately Gaussian even though it is always non-negative. In this report, we present test results using magnitude data and quarter power data.

B. The Classifier Is Trained With Sets of Images

Next, we discuss how the classifer is trained to identify image chips. So far, we have modeled the image as a set of Gaussian distributed pixels whose mean and covariance depend on target class. Each class is distinguished by its own unique mean vector and covariance matrix, and we assume the mean and covariance are known. But in SAR imagery recognition problems we are not given mean and covariance information. Instead we are given sets of images for each target class. For example, each image set may consist of a target viewed from a slightly different elevation or range. We are left with the problem of finding an appropriate mean and covariance matrix given the set of "training" images; computing the mean and covariance matrices which are infused into the classifier is called training the classifier [3]. These results are called target templates and are incorporated into the classifier. Our approach is intuitive and simple to implement: within each class, we perform magnitude or quarter power preprocessing, and then we compute the sample mean and sample covariance of the training data on a pixel-by-pixel basis.

C. An Adaptive Calibration Factor Accounts for Sensor Uncertainties

The motivation behind this modification is the fact that the radar sensor output gain fluctuates from image to image. This mean that a target's apparent brightness can vary due to random gain fluctuations in the sensor electronics. From a target recognition point of view, this means that a given image could be misclassified simply because its average brightness is much larger or smaller than the mean template's average brightness. If we knew the gain we could scale the image chips; however, this information is typically not available. An alternative approach, investigated in this report, is to estimate the unknown gain from each test image, and then scale the test image by the estimated gain. We state the formula without proof below; the derivation appears in reference 4.

D. The Modified Quadratic Classifier

Let us denote the training data by a set of vectors $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \ldots, \mathbf{x}_L^{(i)}$, where L is the number of training images per class and the superscript i denotes the target class. The training data is obtained by collecting complex sar images, stacking their pixels into column vectors, and then taking either the magnitude or the square root of the magnitude on an element-by-element basis. The mean template for target class i is the sample mean of the training vectors:

$$\widehat{\mathbf{m}}_i = \frac{1}{L}\sum_{l=1}^{L} \mathbf{x}_l^{(i)}. \tag{7}$$

The covariance or "variance" template is sample covariance matrix of each training data set:

$$\widehat{\mathbf{C}}_i = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{x}_l^{(i)} - \widehat{\mathbf{m}}_i)\,(\mathbf{x}_l^{(i)} - \widehat{\mathbf{m}}_i)'. \tag{8}$$

The classifier is implemented as follows. First, the complex SAR image chip is preprocessed, yielding magnitude or quarter power pixels. The pixels are stacked into a test image vector x. Next, when there are M target classes, we compute the M discriminant scores using the function

$$d_i(\mathbf{x}) = (\mathbf{x} - \widehat{\mathbf{m}}_i)'\widehat{\mathbf{C}}_i^{-1}(\mathbf{x} - \widehat{\mathbf{m}}_i). \tag{9}$$

The target is classified according to which discriminant score is the smallest.

If the image chips have large numbers of pixels, the sample covariance matrices can become huge. For example, suppose we want to process 64×64 SAR images. Then x is a 4096×1 vector and $\widehat{\mathbf{C}}_i$ is a 4096×4096 matrix! It is difficult and time consuming to invert these large matrices. However, if we assume the pixels are statistically independent the covariance matrices become diagonal, which greatly simplifies the classifier's implementation.

In order to estimate the unknown sensor gain α, we compute the ratio

$$\widehat{\alpha} = \mathbf{x}'\widehat{\mathbf{C}}_i^{-1}\widehat{\mathbf{m}}_i / \mathbf{x}'\widehat{\mathbf{C}}_i^{-1}\mathbf{x}, \tag{10}$$

and then scale the test image by $\widehat{\alpha}$:

$$d_i(\mathbf{x}) = (\widehat{\alpha}\mathbf{x} - \widehat{\mathbf{m}}_i)'\widehat{\mathbf{C}}_i^{-1}(\widehat{\alpha}\mathbf{x} - \widehat{\mathbf{m}}_i). \tag{11}$$

Once again, the target is classified according to which discriminant score is the smallest.

IV. The SAR Image Data

In this section we shall describe the SAR images used to test the quadratic classifier as well as other pattern recognition algorithms. The images were obtained by processing simulated range profiles of four vehicles: a fire truck, a school bus, and M-1 tank, and a T-72 tank. The Xpatch ray tracing code was used to generate the range profiles.

The SAR radar was operated in spotlight mode across a three degree circular track. Eleven SAR images per target were synthesized. The images were a collection of looks at each vehicle as its orientation was adjusted in half-degree increments. The orientations varied from 23.5 degrees to 28.5 degrees. The radar elevation was 10 degrees with respect to the ground plane, and its center frequency was 10 GHz. The

signal bandwidth and synthetic aperture were adjusted so that the images were resolved to 1 foot by 1 foot resolution. The images had 4096 complex pixels in a 64×64 format.

V. Experimental Procedure

We designed an experiment to evaluate the classifier's performance for a wide range of operating conditions. We implemented the process described above using the Matlab programming language. The codes are available to any interested reader. The mean and variance templates were generated using the "leave-one-out" technique described by Fukunaga [3]. This means that only ten out of the eleven images in each class were used to train the classifier. The image which was not used to train the classifier was used to test the classifier. A total of eleven such experiments were performed. During the first experiment, the 23.5 degree image was left out of the training step. During the second experiment, the 24 degree image was left out of the training step and so on.

In order to evaluate the classifier's performance, a way to perturb the test image pixels away from their nominal values must be found. There are many ways to do this; for example, we could introduce occlusion into the orginal test image. We could add clutter to the scene, or we could add or remove specular highlights from the test images. We choose to perturb the images by adding complex Gaussian noise to the test images on a pixel-by-pixel basis. Additive Gaussian noise simplifies the analysis of pattern recognition algorithms. Other kinds of test image perturbations are more realistic; however, since a central goal of this project is algorithm analysis, we decided to begin our studies with a simple additive noise model.

To stress the algorithm we therefore added zero mean, complex, Gaussian noise to the test images pixel by pixel. We added the noise before computing the magnitude or quarter power data, and we adjusted its variance to add greater or lesser amounts of noise to the test images. We expect the classifier to work better as the noise power decreases because the images are perturbed less. Performance of statistical signal processing algorithms are often quantified in terms of a signal-to-noise ratio (SNR). To compute the signal-to-noise ratio, we computed the average energy of all 44 SAR images and defined the result "average image power." The total average image power was 0.0760. Taken individually, the fire truck average image power was 0.06233, the T-72 tank average image power was 0.009527, the M-1 tank image power was 0.002719, and the school bus average image power was 0.001437.

When additive noise is used to perturb the images, its effect must be incorporated into the variance templates. We estimated the noise variance for the magnitude and quarter power tests and added the estimate to each element of the four variance templates.

We used the probability of correct classification as the performance metric. At any signal-to-noise ratio,

it will always lie between zero and one; the closer to one the better. In order to estimate the classification probability, we conducted ten Monte Carlo trials per signal-to-noise ratio and per left out image. The data were compiled from the eleven independent experiments and averaged. In the next section, we will describe the results.

VI. Experimental Results

A. Case 1: Magnitude Data and Fixed Alpha

In the first experiment we processed image magnitudes. The adaptive gain factor α was set to one. We tested the classifier at signal-to-noise ratios between and including -20 to $+20$ dB. The results are shown in Figure 1. We find that between -20 and -13 dB, the correct classification probability is 0.25. The classifier does not work at these very low signal-to-noise ratios because of the large amount of noise added to the four test images. Note that a classification probability of one-fourth is the same as that obtained by classifying a test image by flipping a (four-headed) coin.

As a signal-to-noise ratio of -13 dB, the classification probability begins to improve. Notice the plateau between -6 and $+2$ dB. In this range, the classification probability is one-half. Why? The classifier scores the fire truck and the T-72 correctly; these are the two most powerful image sets. The same amount of noise is added to each of the four test images because the signal power is an average across all 44 images. Since the fire truck and T-72 tank are the most powerful, the signal-to-noise ratio of these two images is large. The pixels are perturbed less; therefore, the classifier is able to score them correctly. Continuing on, we see a second plateau between $+6$ dB and $+10$ dB. The classification probability is 0.75; the classifier now scores the three most powerful images correctly. All four images are scored correctly when the signal-to-noise ratio exceeds $+13$ dB.

B. Case 2: Magnitude Data and Adaptive Alpha

In this test, we calculated the adaptive gain factor α using Equation 10 and implemented the classifier given by Equation 11. The results are shown in Figure 2. We find that the classification probability curve does not display the same kinds of pronounced plateaus that were seen in Figure 1. Instead, we find that the classifier does not work for signal-to-noise ratios less than -3 dB. Then, as the signal-to-noise ratio increases, the classification probability increases fairly smoothly up to one for $+6$ dB and higher. The adaptive gain operates as a scale factor which tends to even out the classifier input image powers. The most powerful images are scaled down and the weakest images are scaled up. Therefore, the classification curve does not

manifest prounced plateaus. The net effect of the adaptive gain is that the minimum signal-to-noise ratio required for unity classification probability is pushed back from +15 dB to +6 dB. This is a significant difference, and we conjecture that in general, adaptive scaling is worth the computational overhead required to implement it.

C. Case 3: Quarter Power Data and Fixed Alpha

In the third set of experiments, we used quarter power images to train and test the classifier. The quarter power images are formed by taking the square root of the magnitudes of the complex SAR images pixel-by-pixel. The results are shown in Figure 3. We found that the performance curve exhibits the typical plateaus we found in the first experiment. However, when quarter power images are used, the minimum signal-to-noise ratio required to classify two images correctly was 0 dB, which is 5 dB more than the magnitude data classifier required. The quarter power image mean templates are closer to one another is a square sum sense; this degrades the quadratic classifier's performance. However, the performance of the magnitude and quarter power classifiers is roughly equivalent for signal-to-noise ratios greater than 0 dB.

D. Case 4: Quarter Power Data and Adaptive Alpha

In the final set of experiments, we used quarter power images to train and test the classifier. The gain factor α was computed with Equation 10. The results are shown in Figure 4. No performance plateaus are seen. When the signal-to-noise ratio exceed 0 dB, the classification probability ramps up to one at +6 dB. Comparing this result to the previous experiment, we find that the adaptive gain factor improves the classifier's performance a great deal. Comparing this result to Figure 3, we find that the signal-to-noise ratio required for unity correct classification probability decreases from +11 dB to +6 dB. Now let us compare the quarter power and magnitude image tests in which the adaptive gain is used. Comparing Figure 4 to Figure 2, we find that only significant difference is the signal-to-noise ratio where the curve begins to ramp. The signal-to-noise ratio required for certain classification is the same (+6 dB) and is less than the non-adaptive gain classifiers.

VII. Conclusions

We tested four types of quadratic classifiers on SAR image data. We found that scaling the magnitude and quarter power images by an adaptive gain factor improved the classifier's performance significantly. We explained why it improves performance in qualitative terms; quantifying its effect on quadratic classifier

performance will be a topic for additional research. In addition, we found that when the adaptive gain factor was used, the performance of the quarter power and magnitude data classifiers was about the same.

How does the performance of the quadratic classifier compare to other kinds of classifiers? Our colleagues have experimented with correlation-based classifiers such as the MACE and SDF classifiers [5] [6]. So far, when additive noise is used to perturb the images, the performance of the correlation classifiers is better. We conjecture the reason for this is the correlation processors are generalized matched filters and, as such, provide coherent gain against the noise that the quadratic classifiers do not provide. In fact, the quadratic classifier design criteria are different than correlator-based classifiers; therefore, there is no reason to believe their performance should be identical.

This work provides the basis for many interesting open problems. There is a need to quantify and compare the performance of ATR algorithms in the face of target articulation and occlusion. We need to compare the quadratic classifier against correlation-based classifiers within these contexts. The pressing need is a better way of comparing ATR algorithms against each other using a combination of analytical tools and data-specific tests.

References

1. The University of Michigan Extension, **Synthetic Aperture Radar Technology and Applications**, short course notes, December 1993.

2. J.T. Tou and R.C. Gonzalez, **Pattern Recognition Principles**, Addison-Wesley, 1974.

3. K. Fukunaga, **Introduction to Statistical Pattern Recognition**, Second Edition, Academic Press, 1990.

4. J. Salazar, "The Sandia National Laboratories Modified Quadratic Classifier," unpublished notes.

5. L.M. Novak *et. al.*, "Radar Target Identification Using Spatial Matched Filters," **Pattern Recognition**, Volume 27, Number 4, pp. 607-617, 1994.

6. B.V.K. Kumar, "Tutorial survey of composite filter designs for optical correlators," **Applied Optics**, Volume 31, Number 23, pp. 4773-4801, 1992.

Figure 1

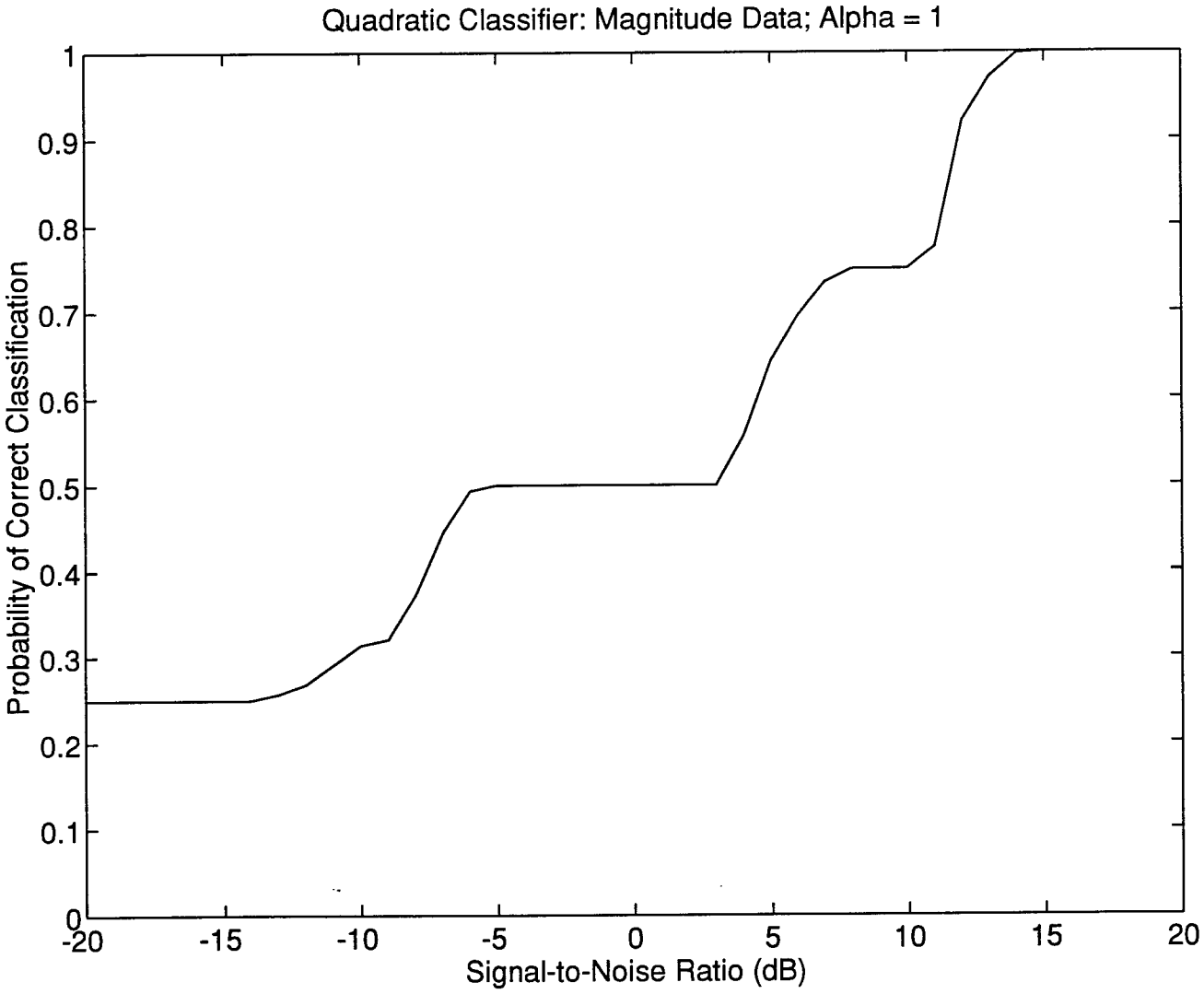

Quadratic Classifier: Magnitude Data; Alpha = 1
Probability of Correct Classification
Signal-to-Noise Ratio (dB)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-20
-15
-10
-5
0
5
10
15
20

Figure 2

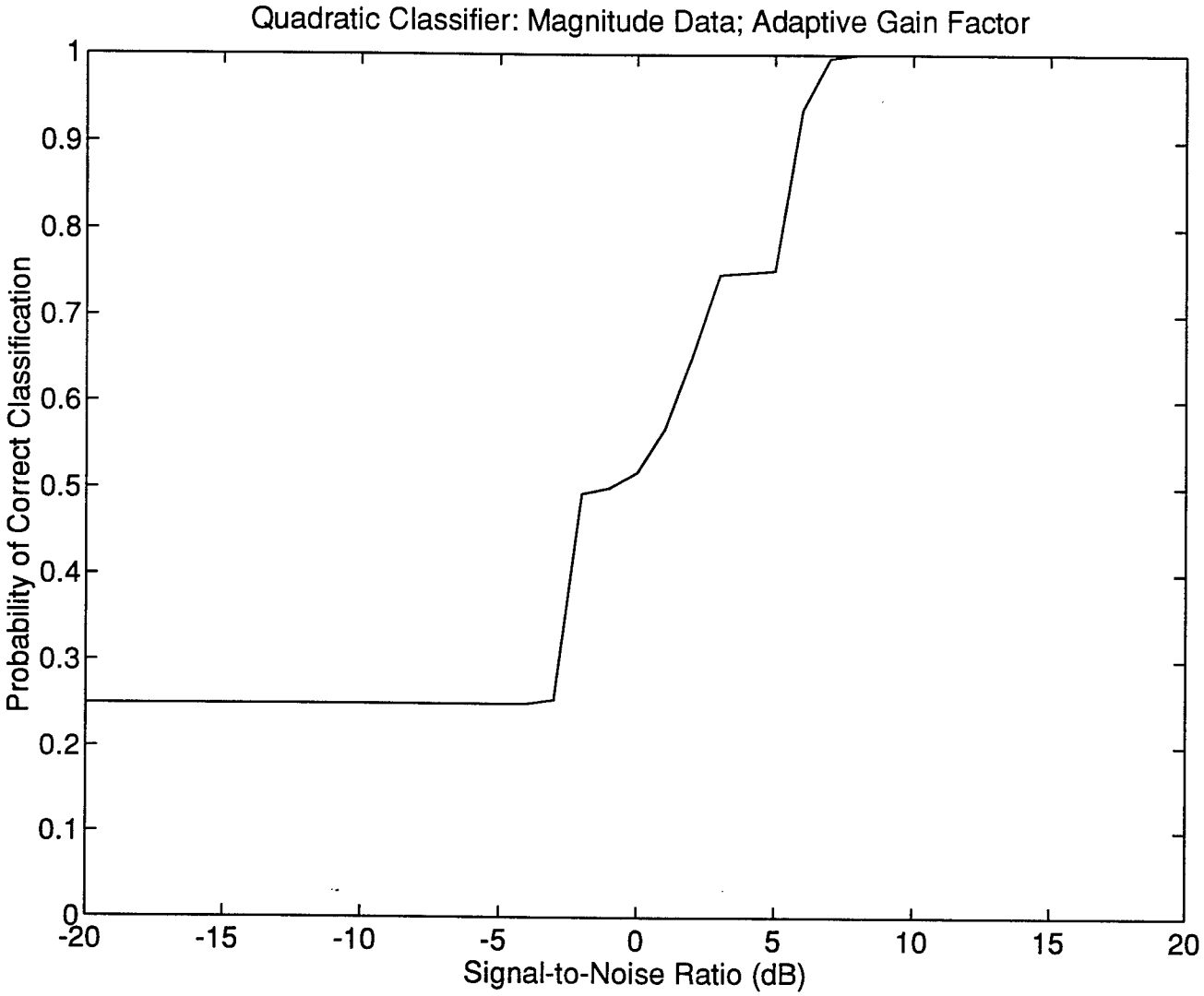

Quadratic Classifier: Magnitude Data; Adaptive Gain Factor
Probability of Correct Classification
Signal-to-Noise Ratio (dB)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-20
-15
-10
-5
0
5
10
15
20

Figure 3

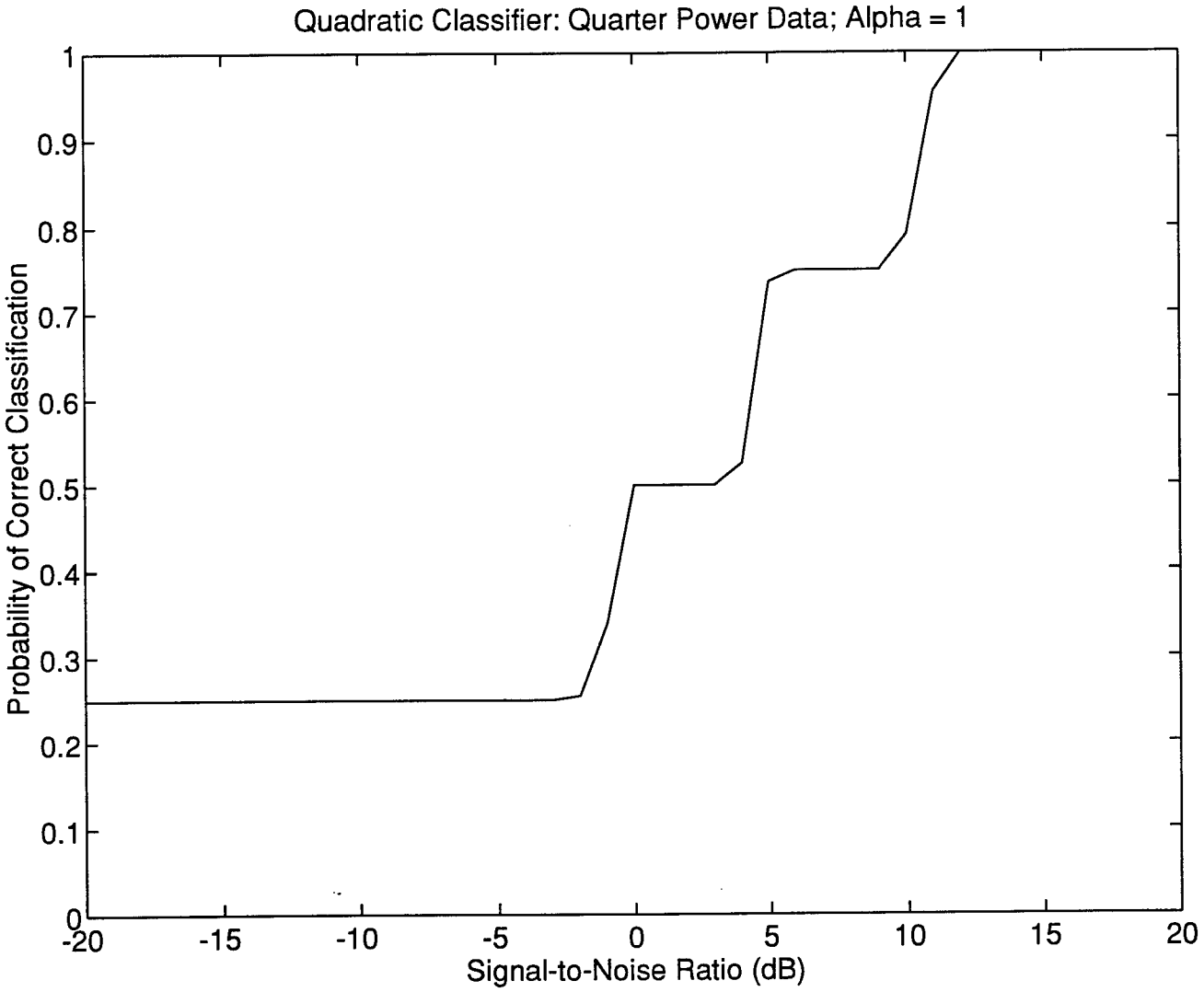
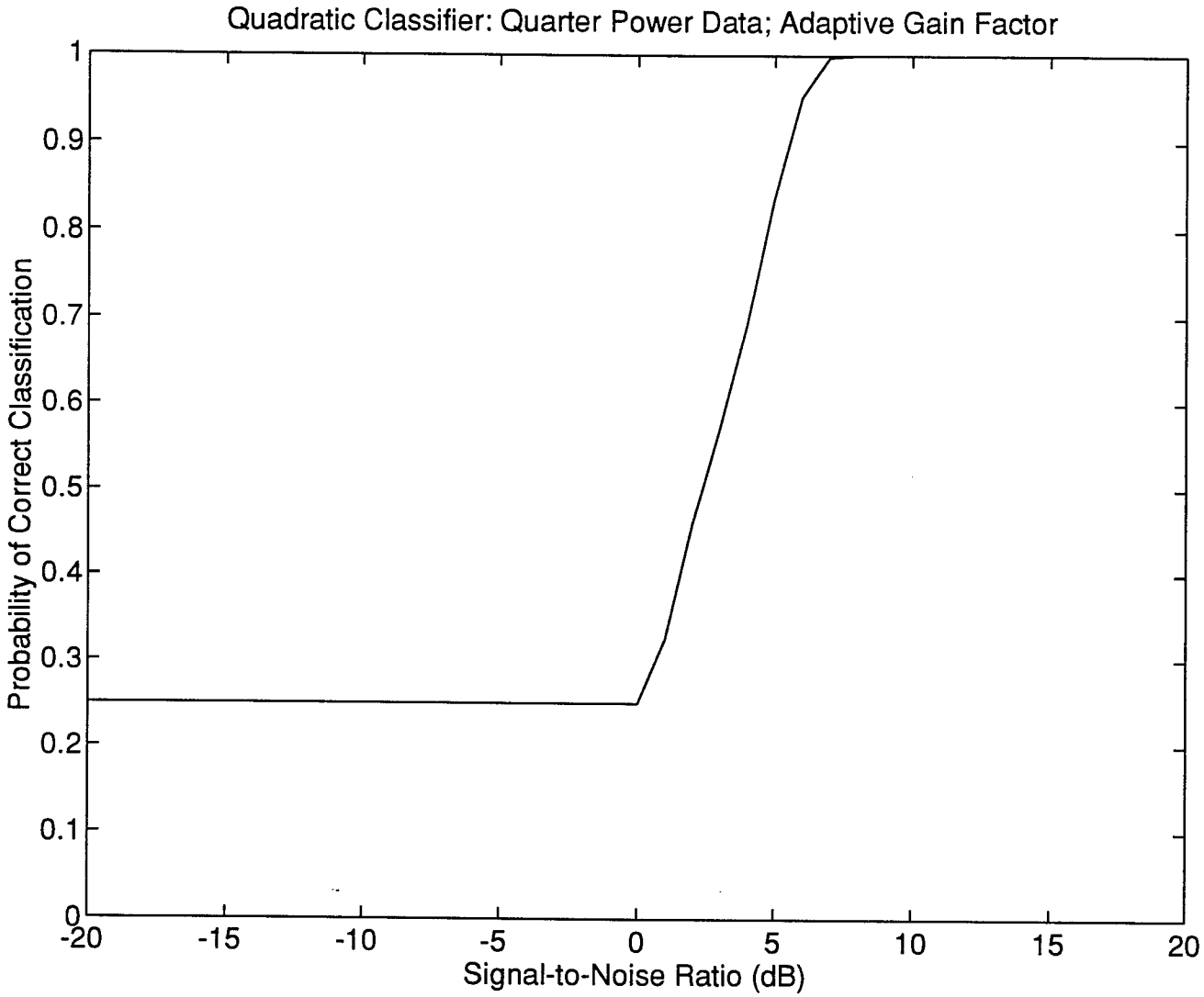

Quadratic Classifier: Quarter Power Data; Alpha = 1
Probability of Correct Classification
Signal-to-Noise Ratio (dB)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-20
-15
-10
-5
0
5
10
15
20

Figure 4


Quadratic Classifier: Quarter Power Data; Adaptive Gain Factor
Probability of Correct Classification
Signal-to-Noise Ratio (dB)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-20
-15
-10
-5
0
5
10
15
20

ELECTROLUMINESCENCE STUDIES OF THE RIGID ROD POLYMER

Poly (p-phenylenebenzobisthiazole)(PBZT)

Barney E. Taylor

Visiting Assistant Professor

Department of Physics

Miami University - Hamilton

1601 Peck Blvd.

Hamilton, OH 45011

Final Report for:

Summer Faculty Research Program

Wright Laboratory

WL/MLBP

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, DC

September 1994

ELECTROLUMINESCENCE STUDIES OF THE RIGID ROD POLYMER

Poly (p-phenylenebenzobisthiazole) (PBZT)

Barney E. Taylor
Visiting Assistant Professor
Department of Physics
Miami University - Hamilton

Abstract

Poly (p-phenylenebenzobisthiazole)(PBZT) is a rigid rod polymer with excellent thermal and mechanical properties. Electroluminescent devices with PBZT as the active layer were fabricated upon an indium-tin oxide (ITO) coated glass slides. Magnesium metal was evaporated as the injection electrode, and the magnesium was overcoated with an evaporated silver layer to reduce oxidation of the magnesium. The devices were characterized using current-voltage and spectral techniques. The current-voltage measurements indicated that the devices acted as a diode with a significant resistive component; while the spectral studies showed an emission from 850 nm (the long wavelength limit of the apparatus) up through about 525 (nm) at device voltages ranging from 2.35 V to 4.0 V for various devices studied. The light output of the devices decreased with usage until the device ultimately failed. The devices also became more resistor like with use until failure occurred.

ELECTROLUMINESCENCE STUDIES OF THE RIGID ROD POLYMER

Poly (p-phenylenebenzobisthiazole) (PBZT)

Barney E. Taylor

Introduction

Electroluminescent polymers offer great hope for the display technology of the future by making thin, flexible flat panel color displays to become a possibility if suitable materials and processing techniques can be found. This document reports on the research performed on the electro-luminescent characterization of the rigid rod polymer PBZT (Poly (p-phenylenebenzo-bisthiazole)) by means of electrical and spectral techniques. The structure of PBZT is shown in Figure 1.

Previous work [1], has shown that diodes fabricated upon PBZT films emit light under forward bias. The more successful diodes were obtained from the thinnest PBZT polymer films. Unfortunately, the PBZT diodes were reasonably short lived and quite erratic in light output. The major thrust of this study was to successfully characterize the spectral response of the PBZT devices. In order to do this, operating and experimental conditions had to be found that would extend the device lifetime and uniformity of light emission. One step was to extend the lifetime of the devices by limiting the electrical power applied to the device. Earlier investigations by the Physics Group had shown device lifetime could be extended by using a current pulse rather than a constant current through the device. It was speculated that heating, analogous to "thermal runaway" in conventional semiconductors, was in part responsible for the short lifetimes. Pulses with a low duty cycle (short "on" time) allowed the device to cool appreciably between pulses and keeping the overall temperature of the device down compared to a continuous current. Thus, the experimental system needed required modification in order to measure the total light output of the device as a function of time. That data could be used as a means of normalizing the spectral data to be equivalent to that of a device of constant brightness. Care was exercised to minimize the exposure of the devices to air or water vapor -- sources of oxygen that lead to severe contact degradation due to oxidation of the Mg contact.

EXPERIMENTAL

Sample Fabrication

The experiments were performed on PBZT films cast upon indium-tin oxide (ITO) coated glass by the doctor blade technique of Agrawal[1]. After repeated washing in distilled water and drying in a vacuum oven at 60 °C, magnesium injection electrodes were evaporated upon the PBZT. Magnesium was chosen for its desirable work function. Approximately 500 nm of Mg was evaporated on the PBZT film as 1 mm diameter pads spaced about 3 mm on center. To reduce the tendency of Mg to very rapidly oxidize, the Mg pads were overcoated with 500 nm of silver during fabrication. The devices were stored in a vacuum dessiccator and exposed to atmospheric moisture and air as little as possible.

Experimental Apparatus

All experiments were performed with the devices mounted in a Specac ™ cryostat with quartz optical windows. A "pogo-pin" (spring loaded electrode pin) was mounted in a Teflon insert to make contact with the evaporated silver pad of the device under study. Electrical contact to the ITO was made by scraping away the PBZT film at an edge of the glass and attaching a fine wire with indium-mercury amalgam and transparent tape. The glass plate was held to the "pogo-pins" by rubber bands over the cold head of the Specac. The usual vent tubes in the Specac were replaced with electrical feedthroughs - one for the "pogo-pin" and another for the wire to the ITO. All studies were performed at room temperature and under a forepump vacuum of 2 to 20 milliTorr. Absorption measurements were performed on the windows of the Specac cryostat, and the ITO coated glass prior to characterization of PBZT devices. No noteworthy absorption was observed [2].

Light from the device was focused on the entrance slit of a Spex 500M monochromator with a 1200 groove/mm grating blazed at 500 nm as shown in Figure 2. A photomultiplier mounted at the exit slit was used to detect the light at the wavelength of interest. A small fraction of the light was split out using a microscope slide as a beamsplitter and directed upon a silicon photodiode or another photomultiplier to monitor the overall light output of the device. A block diagram of the electronics is shown in Figure 3. Light

from each detector (spectral component and overall) was measured using synchronous detection via a lock-in amplifier. A computer system was used to perform the data acquisition.

In order to use synchronous detection, the data must be modulated in a periodic manner. The best device lifetimes had been obtained by pulsing the device with a 4 kHz, 20% duty cycle pulse. Usually a lock-in amplifier performs best with a 50% duty cycle pulse. A study was performed using a GaAs light emitting diode as a prototype sample. The operating conditions (frequency and duty cycle) of the LED were set to mimic the PBZT devices. Spectra were obtained by chopping the 4 kHz 20%, emitted light stream at about 13 Hz with a mechanical chopper (50% duty factor) and by using the pulse generator as a reference. Although not an optimum signal for the lock-in amplifiers, the pulse reference provided a signal approximately 2.5 times greater than that obtained with the mechanical chopper at the lower frequency. Thus, it was decided to use the pulse period of the device as the reference for synchronous detection for all runs.

The previous studies had used a current amplifier to provide the operational bias for the sample. Only the intervention of the operator could limit the power available to a device that suddenly showed a drop in resistance -- a frequently observed phenomena. It was decided to design and construct a voltage controlled current source that would maintain a constant current through the device in response to a given input voltage. A suitable design was developed that allows a controlled current pulse to applied to the device. The level of the pulse is determined by the level of a constantly applied reference voltage, and the period and duty cycle of the pulse is determined by a digital pulse generator. The computer was integrated into the experiment so that it controlled the level of the reference voltage and monitored for over voltage conditions while taking the data. As a result of this modification, the device current, device voltage, overall light output and light output at the wavelength range being passed by the monochromator were measured for all experiments -- once the system was fully implemented.

Two types of experiments were performed: electrical characterization of voltage and current (I-V) and a spectral measurement over a range of wavelengths under preset operating conditions. Three experimental control

computer programs were developed: Voltage vs. Current; Current vs. Voltage and Spectral.

A more complete discussion of the modified experimental system is in the Appendix.

Data and Discussion

Electrical Characterization

The electrical characterization of most of the PBZT diodes showed a "leaky" diode behavior that indicated that a relatively low shunt resistance was in parallel with the diode. Figure 4 shows a typical current versus voltage plot for a PBZT device. The data is adequately described by the function

$$\mathbf{I(V) = \frac{V}{R} + I_o \cdot \left(e^{A \cdot V} - 1 \right)} \qquad \text{Eqn. 1.}$$

where V is the device voltage and I is the device current. R is the magnitude of the shunt resistance. R and A were determined by non linear least squares fitting of the equation to the data. In order to check the validity of the equation, a known resistor was placed in parallel with a GaAs LED and data was taken as if the sample were a PBZT device (except that the maximum allowed voltage was kept lower to prevent damage to the GaAs device). Application of the above equation yielded a correct value for the shunt resistor and a consistent exponential factor. For a nonradiative diode the exponential factor is usually expressed as $V_t = kT/q_e = 1/A$. Evaluation at room temperature gives a value of about 40 for A. The GaAs LED had a value of around 4 for A and the typical PBZT device had a value of about 0.2~0.4 for A. The exact nature of the variation is not known, but could be related to the larger "turn on" voltage of the PBZT devices compared to GaAs (2.5 V versus 1.6 V, respectively). Occasionally, a better fit could be obtained by the inclusion of an additive term to the above equation. Presumably, the additive term helped to remove a "zero offset" in the instrumentation.

One PBZT device studied presented the highest intrinsic shunt resistance of any device studied -- by at least an order of magnitude. It was not necessary to include the shunt resistance term in fitting the data. The value of A for this sample is about 0.4 -- higher than obtained by the combined fit

but still within reason. The current-voltage data is shown in Figure 5. The sample of Figure 5 was exceptional in other ways. It was the most stable device studied with only a gradual decrease with time. It was unusually long lived (several hours) for PBZT devices studied in this laboratory. The voltage measured across the device was much lower when detectable light was emitted than the typical PBZT device (2.4 V versus 3.5 V or more). The possible reasons for this particular device to be so unusual will be discussed later.

As mentioned in the experimental section, the light emitted from the devices (if any) was monitored during the I-V runs. Figure 6 presents the intensity of light at a wavelength of 600 nm(determined by the monochromator). The curve shows a response similar to the diodes. Some success was obtained fitting these equations with a form similar to the exponential term in the equation above:

$$I_{photo} = I_{photo_0} \cdot (e^{A \cdot V} - 1) + C \qquad \text{Eqn 2.}$$

I_{photo} is the intensity of the emitted light. The names of the constants have been changed in order to avoid confusion (except for A). Success of the fitting process is seen by the overlaid best fit line. At this point the significance of the numerical values of the exponential factor, A, is unknown.

Some other variations of Current versus Voltage measurements were performed. The voltage controlled current source (VCCS)described in the experimental section functions by monitoring the voltage across a resistor of known value within the circuit. The control condition is that the voltage across the resistor is equal to the current through the device times the resistance. By physically swapping the device's and resistor's connections, it was possible to use the VCCS as a voltage controlled voltage source. Since there was no protection for the device if its resistance dropped (the normal aging mode) this method was not used very often. In a few cases, it allowed excellent I-V data to be obtained by controlling the device voltage and measuring the current. Figure 5 is one of the better sets of data obtained in this manner.

The VCCS current sources used to bias the devices is by nature unipolar (i.e. it cannot change polarity). However, the sample could be placed under

reverse bias conditions by interchanging the electrical leads to the device. No interesting or unexpected results were obtained in the reverse bias mode.

Spectral Characterization

The spectral characterization of the PBZT devices proved to be very challenging. The earlier work [1] had shown that the intensity of the light output was sufficiently unstable to perform a scan using a monochromator based spectrometer system. An optical multichannel analyzer would be ideal since the entire spectrum would be acquired simultaneously, but none were available during the time of this study. A method to obtain a representative spectra was devised in which a small part of the total light from the device was reflected into an independent full spectrum detector. The intensity recorded by the independent detector could be used to normalize the spectral data point by point to provide a spectrum representative of a device of constant brightness. Intrinsic in this methodology is the assumption that the spectral distribution of the device is independent of the light output.

The technique was simulated using a GaAs LED as a source and varying its intensity significantly and randomly during the spectral run. The noise level was very low due to the high efficiency of the GaAs LED; and normalization yielded a spectrum whose shape was identical to that obtained at constant intensity. With this validation of the experimental technique, PBZT devices were studied. Several devices were studied during the term of this investigation. Only three were sufficiently bright and long lived to obtain a good baseline normalization. Since only about 6% of the light from the device was split off and being sampled, the amount of baseline signal above the noise level of the electronics was usually marginal. Figure 7 is a family of spectral runs obtained from a PBZT device under various experimental conditions. This particular device was the most stable with respect to time of all studied. Even so, the inset shows the degradation in the overall brightness during one of the scans. Since the baseline was so linear, normalization was performed by both a point by point normalization and

performing a linear fit to the baseline intensity. No significant spectral difference was noted between the two.

The shape of the spectra has not been adjusted for the response of the monochromator and photomultiplier used to obtain the spectral data. Previous photoconductivity studies using the monochromator indicate that the grating is most efficient in the region of greatest intensity. The response of the photomultiplier falls off very rapidly at 850 nm and is the limiting factor in the measured experimental spectral shape. A photomultiplier with a different type of photocathode was not available to insure that the intensity of the device is indeed falling below 850 nm (about 1.45 eV). The earlier work of Ferguson, et. al.,[1]shows electroluminescence at energies slightly lower than the photoluminescence data. Since both sets of data were obtained using the same monochromator as used for this investigation, there might indeed be more interesting physics at still longer wavelengths. The absorption coefficient of the PBZT films is in its asymptotic region toward zero. Thus it is likely that the intensity of the electroluminescence is falling off. It is desirable to know the exact nature of the decay in intensity with increasing wavelength (decreasing photon energy).

Near the end of this investigation, an optical multichannel analyzer in another laboratory was used to characterize a yellow GaAs LED. The same diode was characterized in this laboratory as though it was a prototypical PBZT sample. The resulting spectra are shown in Figure 8. It can be seen that the OMA yielded a spectra of greater width than the monochromator/ photomultiplier combination. This again confirms that the response of the monochromator falls off at shorter wavelengths, and that the photomultiplier is less sensitive at longer wavelengths. The reasonable agreement between the shapes of the two spectra in Figure 8 suggests that the high energy features of the obtained PBZT spectrum Figure 7 are indeed a valid representation of the electroluminescent nature of PBZT.

SUMMARY AND SUGGESTIONS FOR FURTHER STUDY

PBZT is a very interesting and challenging material for electroluminescence studies. Its desirable physical and thermal properties could yield much more rugged and durable light emitting devices than presently

available. The short lifetime of the devices is a cause of concern. It is believed that this is a materials / fabrication problem. The films are cast by a manual technique and then allowed to dry. Before electrodes were deposited the general appearance of the films were noted. There were regions of different color indicating regions of different thickness across the sample. This material is not soluble in water or other common solvents, hence, spin coating uniform layers at this time is not a viable prospect. The lack of a good uniform film is likely the source in the variability in the devices fabricated. Some devices were totally resistive with no diode nature of light emission. These devices likely had pin holes that allowed conductive regions to form during deposition or very thin regions that broke down under the electric field and heating effects. Other pads were initially diode like in nature and exhibited some electroluminescence, but were very erratic in both resistance and light output. This behavior could be caused by nonuniform regions that were sufficiently small in diameter that when the film broke down, a sizable component flowed through the good regions nearby. Occasionally one or two spots were observed on a glowing pad that were extremely small and very bright in comparison with the rest of the device. The spots would quickly die away.

The electrical and thermal stress in the typical PBZT device of this investigation is nontrivial. The operating voltage would typically be 4 volts, across an extremely thin film, yielding electric fields in the range of 10^5 V/cm. Currents of 200 mA were not uncommon in some of the more resistive devices, leading to a current density of 20 A/cm^2. The power is very high -- typically 0.8 watts in a region that is extremely small. The power density involved might indicate why longer lifetimes have been observed using pulses to the device. With a typical duty cycle of 20% at a 4 kHz pulse rate, the device is only active for about 50 μs and has about 200 μs to dissipate the heat before being active again. The device lifetime has been reported to be significantly shortened by using pulse frequencies of less than 2 kHz [2]. It is likely that the active region of the sample becomes so hot during the active time that it cannot cool sufficiently and overheats and degrades during successive pulses.

In comparison to the "typical" PBZT device, the very good device shown in Figures 5 and 7 had a voltage of about 60% of typical and a current that was about 10% of normal. Hence, it was dissipating only about 6% of the electrical power of the "typical" device. It is felt that the lowered thermal and electrical stress on this particular sample gave rise to its excellent (for PBZT) lifetime and optical stability.

Contact degradation is a very real problem with the Mg electrodes. Overcoating with Ag reduces the oxidation of the contacts, but degradation still occurs even when the samples are stored in a vacuum dessicator under forepump vacuum. The contacts are initially very silvery, smooth, and fairly durable. The "pogo pin" does not abrade away the contact material of a freshly prepared contact. The same is not true for an aged contact. The visual appearance is powdery, and the contacts are extremely fragile. It is possible that residual acids/solvents in the film are diffusing to the surface to react with the electrodes, although surface science tests have not yet been performed to test that possibility. A second, but no more convincing, scenario is that oxygen is diffusing along the surface and oxidizing the Mg underneath the Ag overcoated layer. Due to the long hoses and small apertures involved with attaching the forepump to the dessicator or Specac, it is likely that the vacuum is not a particularly high vacuum. Unfortunately, the Specac was not designed for high vacuum use, and constant pumping is necessary to maintain a vacuum of a few torr in the cryostat.

An optical multichannel analyzer (OMA) currently being procured will allow refinement of the spectral shape beyond the experimental limitations of this investigation. The techniques of sample mounting, pulsing and gathering the current-voltage data while obtaining the present spectra will be of great use when the OMA arrives and is in place. Since the OMA acquires the entire spectrum simultaneously, there is no need to know the instantaneous brightness of the device for normalization. However, as long as the light output is erratic and the lifetime short, having that knowledge is useful in trying to learn the true nature of electroluminescence in PBZT.

Appendix

The Voltage Controlled Current Source (VCCS) mentioned in the experimental section will be described. A block diagram of the circuit is in Figure A1. The output current level is determined by the analog reference voltage applied to the input of the analog switch. The analog switch is alternated between its second input (at ground potential) and the reference analog input by the pulse coming from the digital pulse generator. When the pulse is high, the control voltage to the VCCS stage is the analog reference voltage, and when the pulse is low, the output of the VCCS is set to zero. In the present experimental conditions, the analog control voltage is provided by the Digital to Analog output of the PAR 5210 lock-in amplifier. The reference oscillator of the 5210 is set to the desired operating frequency and the TTL output is connected to the external trigger input of the HP digital pulse generator. Thus, a pulse is generated for every reference cycle of the lock-in. Two differential amplifiers are included in the circuitry. One measures the actual voltage across the resistor that provides the negative feedback for the VCCS section. The second measures the voltage across the device under study. None of the operational amplifiers have been provided with zero offset compensation, however, the offsets are typically quite small. The largest was constant offset of 13 mV in the voltage measured across the VCCS resistor.

The circuitry is powered by an 18 V bipolar power supply obtained from four nine volt batteries. The lifetime of the batteries is relatively short if the circuit is not switched off when not in use. It would be desirable to redesign the circuitry using higher voltage operational amplifiers to allow greater voltages to be measured by the differential amplifiers. With the present design, device voltages of 4 V can be accommodated when the current is such that the voltage across the VCCS resistor is about 4 V while using a 10 V external power supply.

REFERENCES

1. J. B. Ferguson, R. J. Spry, A. K. Agrawal, J. W. Lee, S. J. Bai, S. M, Hegde, C. S. Wang, Bull. Am. Phys. Soc. 39, No. 1, 99, (1994)
2. J. B. Ferguson (Private Communication)

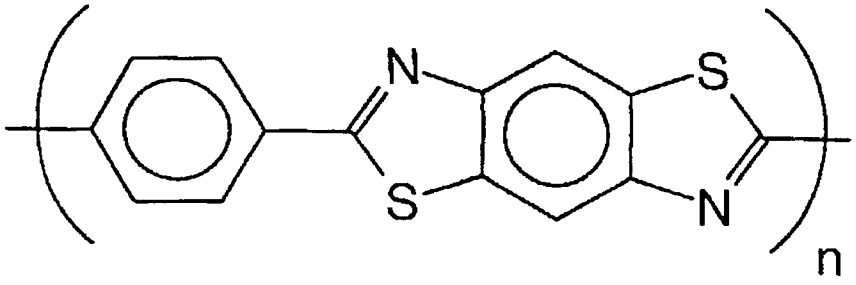


Figure 1. Structure of PBZT.

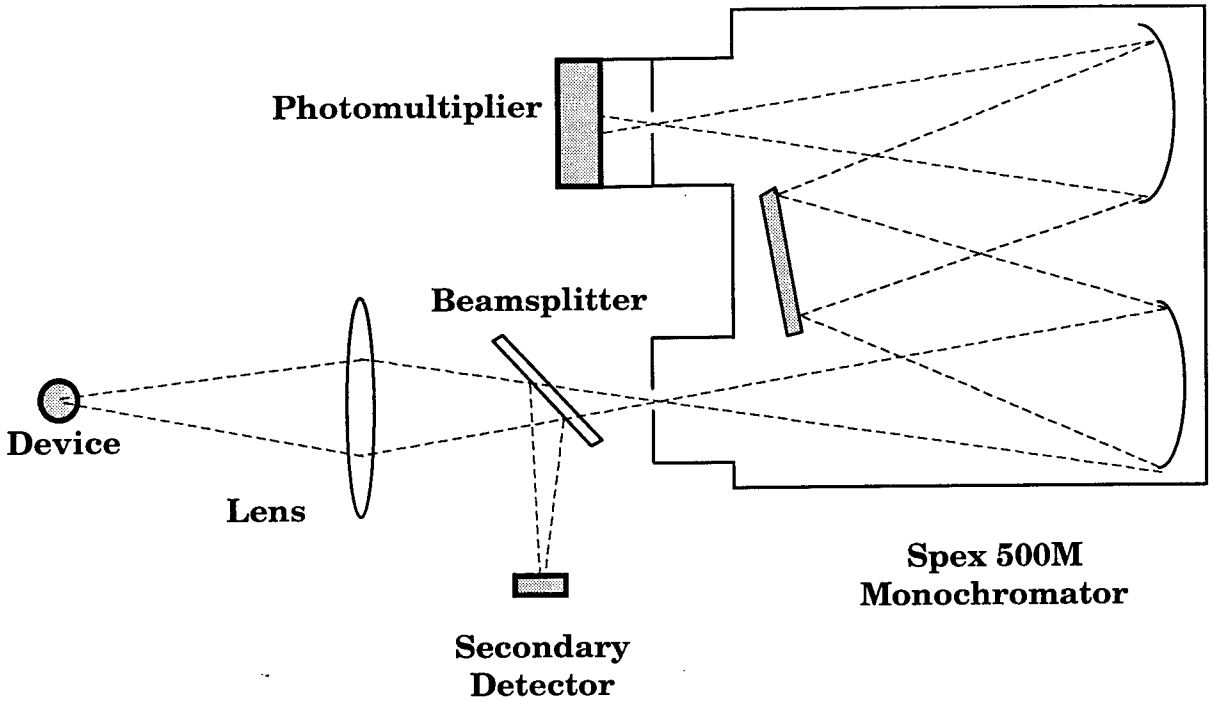


Figure 2. Optical block diagram of the EL system.

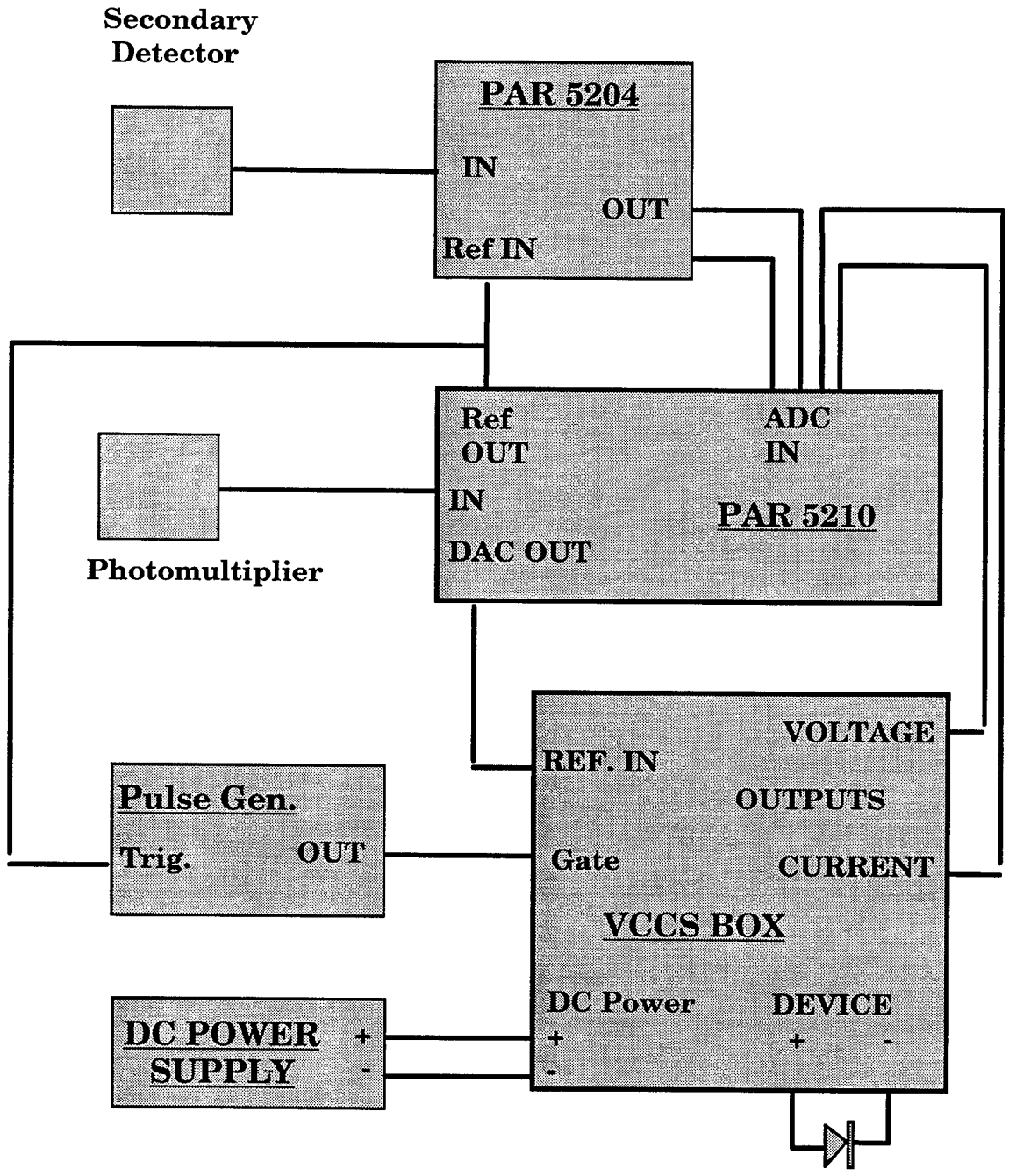


Figure 3. Electronics block diagram after inclusion of the VCCS circuitry.

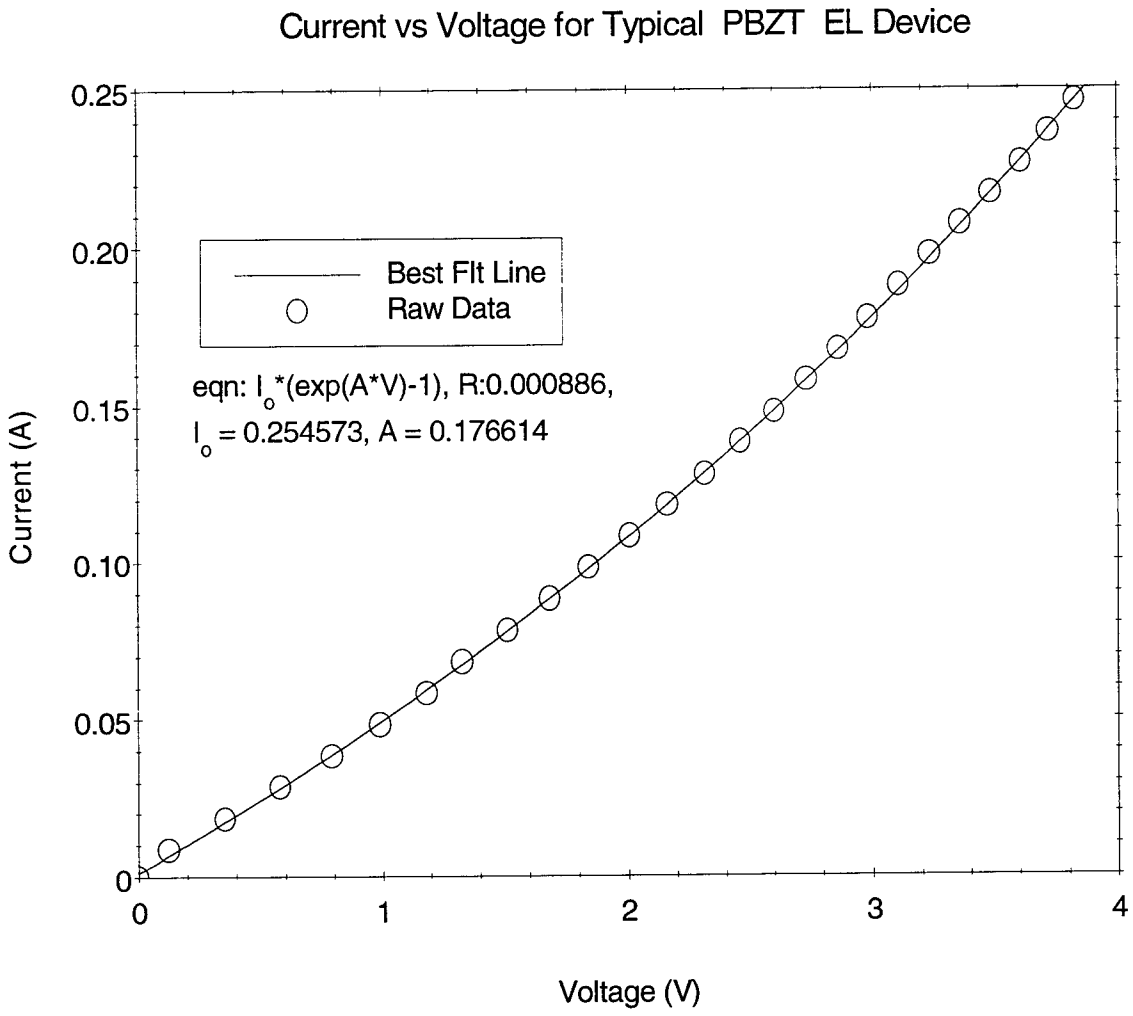


Figure 4. Typical Current vs. Voltage for PBZT devices. Note 'on' power is approximately 1 watt.

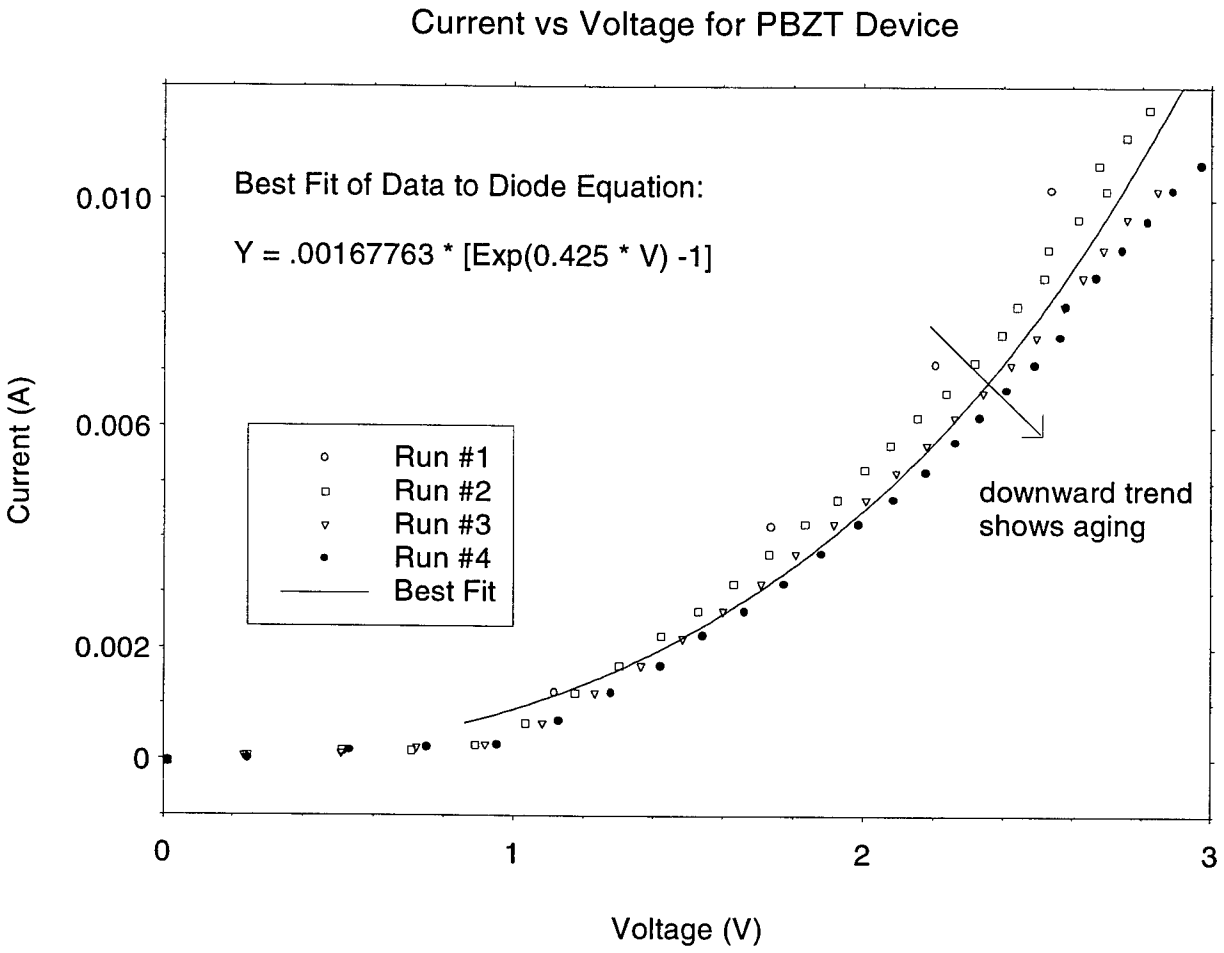


Figure 5. Current vs. Voltage data for the 'Best' PBZT EL sample. Note that the peak power is about 35 milliwatts. The device also shows aging. After repeated measurements, the current-voltage behavior turned into the more typical behavior of the previous figure.

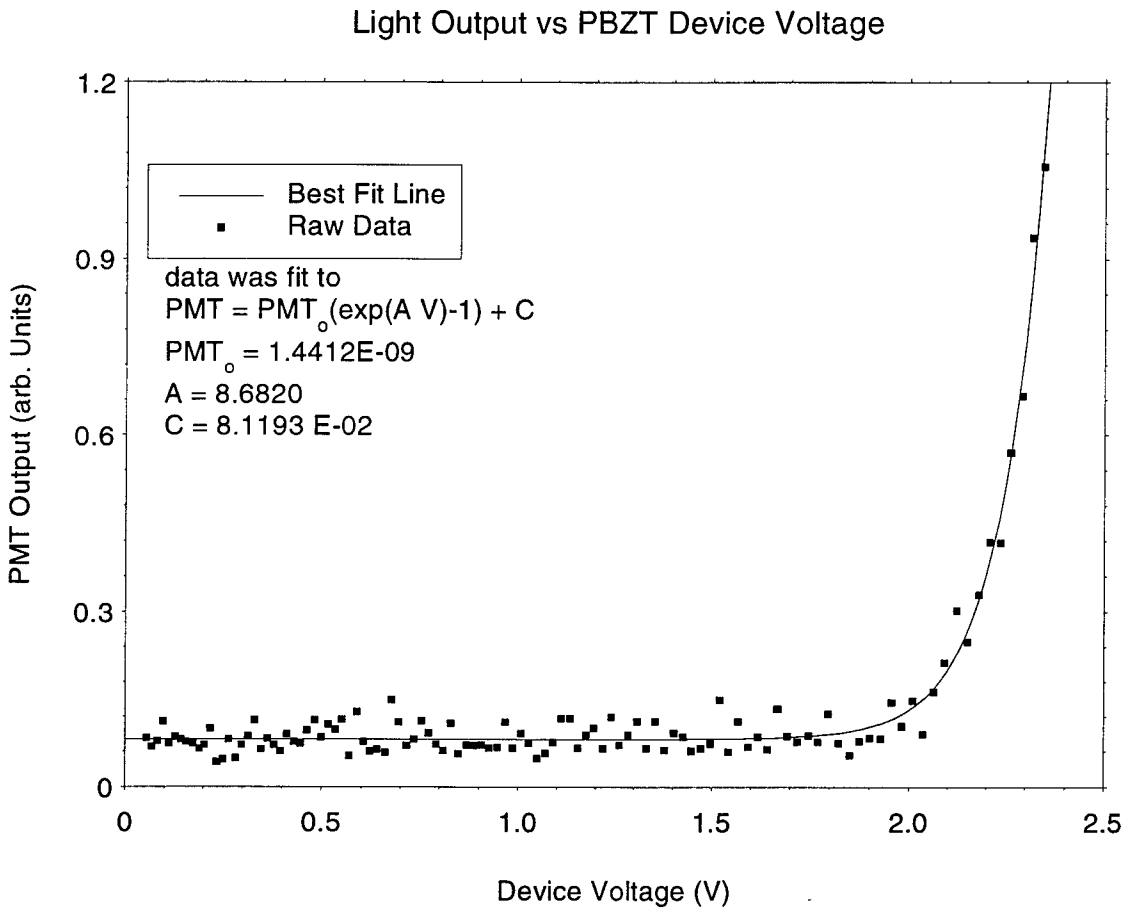


Figure 6. The light output of the 'Best' PBZT EL device as a function of applied voltage. The light output was measured at 600 nm through the Spex monochromator using a photomultiplier as the detector. The lock-in amplifier was operated in its magnitude mode, so that the baseline represents the average noise in the system The light output fits the same type of equation that the electrical behavior follows.

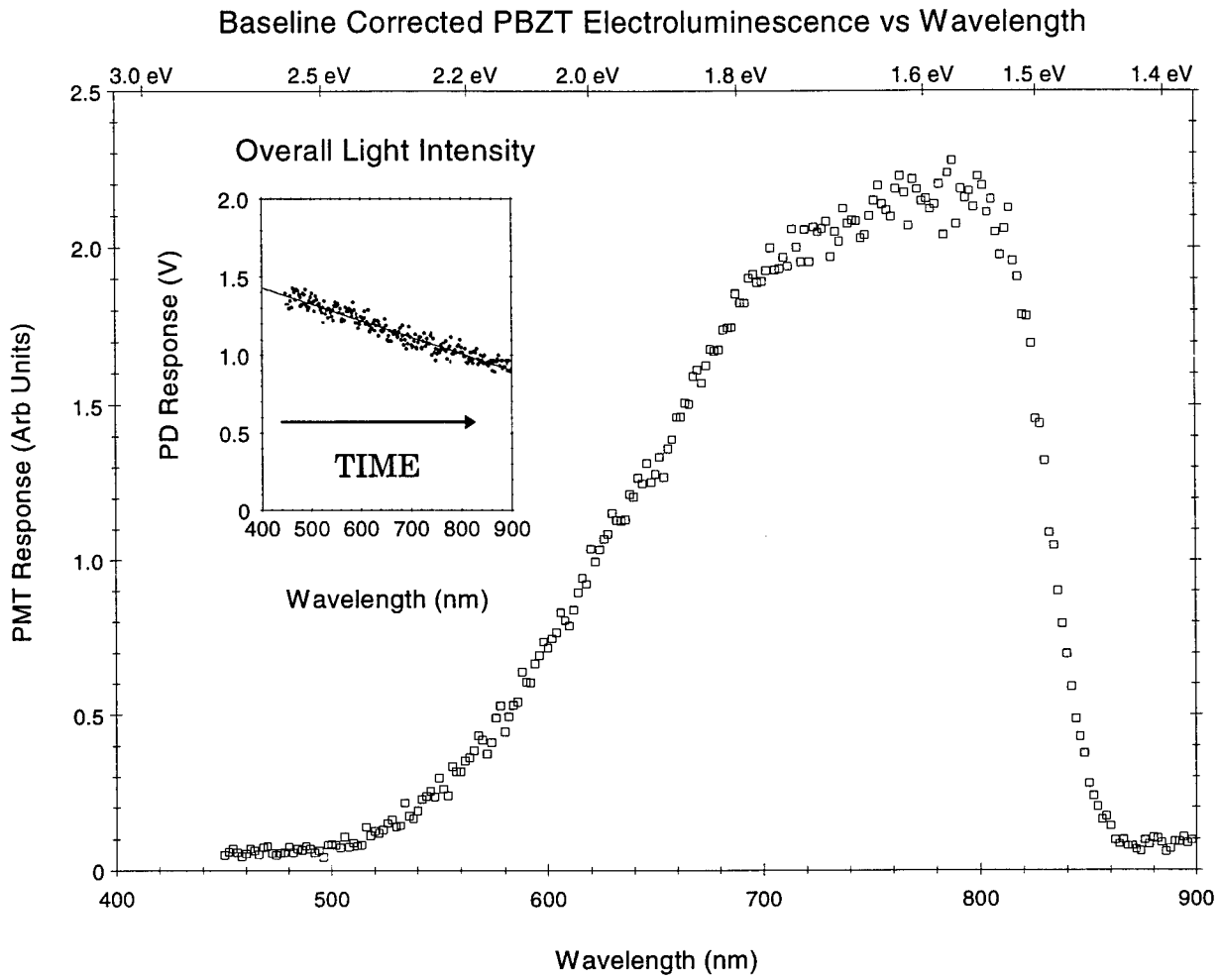


Figure 7. Electroluminescent spectrum of PBZT. The spectra has been corrected for the change in device brightness during the scan. The inset shows that the total light intensity decreased by about 40% during the run. The device was operated at 2.5 V. Other scans of this device under different bias conditions resulted in essentially the same shape after baseline correction. The falloff at 850 nm is due to the loss of sensitivity of the photomultiplier's cathode.

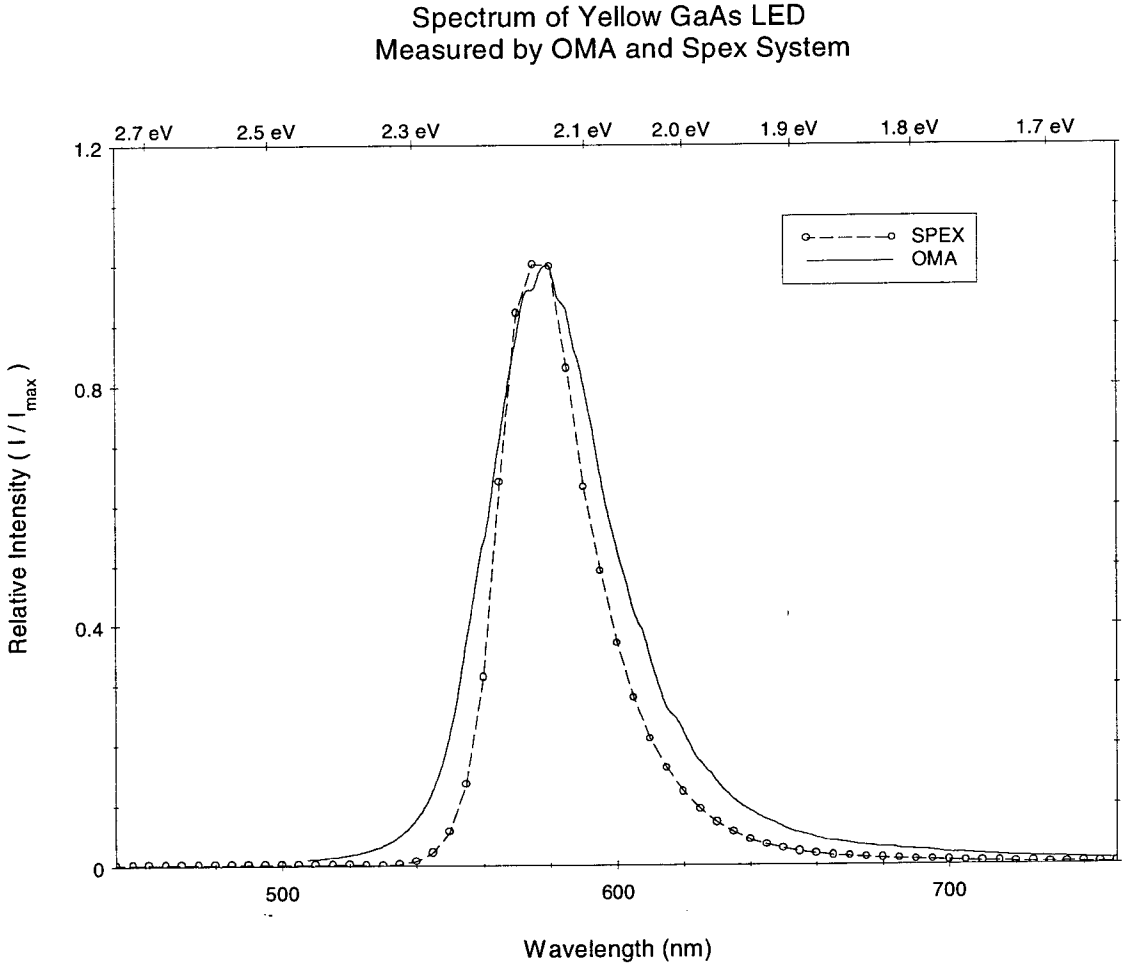


Figure 8. Spectrum of a yellow GaAs Led as measured with a commercial optical multichannel analyzer (OMA) and the experimental system used for this study. Each spectrum has been normalized to its maximum value. Both instruments show the same shape, but the OMA indicates a greater width to the peak.

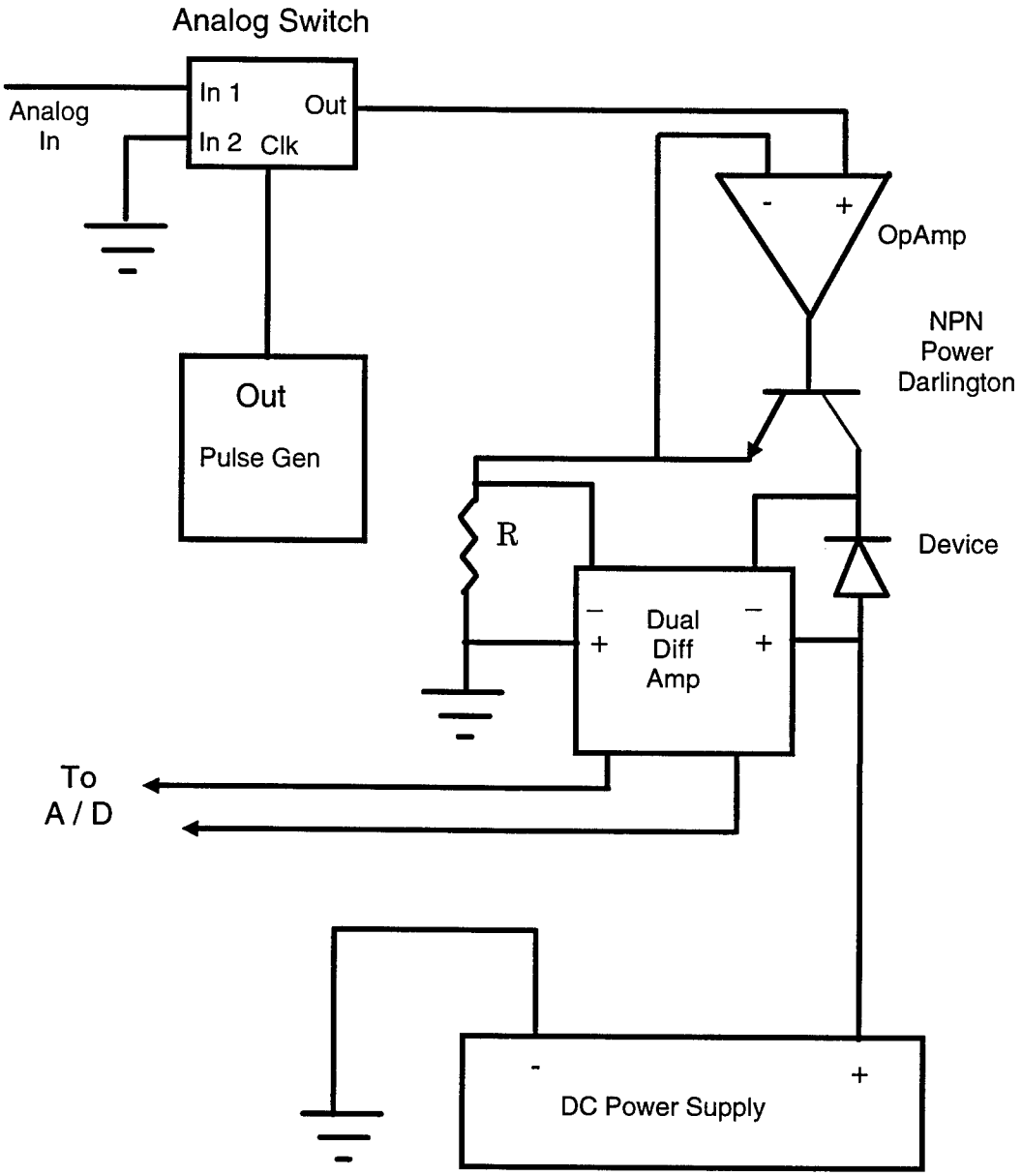


Figure A1. Block diagram of the VCCS circuitry used to bias the PBZT devices.

VHDL-93 PARSER IN PROLOG

Krishnaprasad Thirunarayan
Assistant Professor
Department of Computer Science and Engineering

Wright State University
3640, Col. Glenn Highway
Dayton, OH 45435.

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

September 1994

VHDL-93 PARSER IN PROLOG

Krishnaprasad Thirunarayan
Assistant Professor
Department of Computer Science and Engineering
Wright State University

Abstract

This report describes the upgrading of a VHDL-87 parser written in Prolog (by Peter Reintjes) to conform to the IEEE Standard 1076-1993 of VHDL. The original parser, which is based on a simplified grammar of VHDL-87 (given by Lipsett, Schaefer and Ussery), was first modified to conform to the IEEE Standard 1076-1987. Subsequently, it was extended by incorporating additional features supported by VHDL-93. The new parser has been tested on the VHDL-87 test suite and a few VHDL-93 programs. Our VHDL-93 parser takes about 30% longer to parse the VHDL-87 test suite.

VHDL-93 PARSER IN PROLOG

Krishnaprasad Thirunarayan

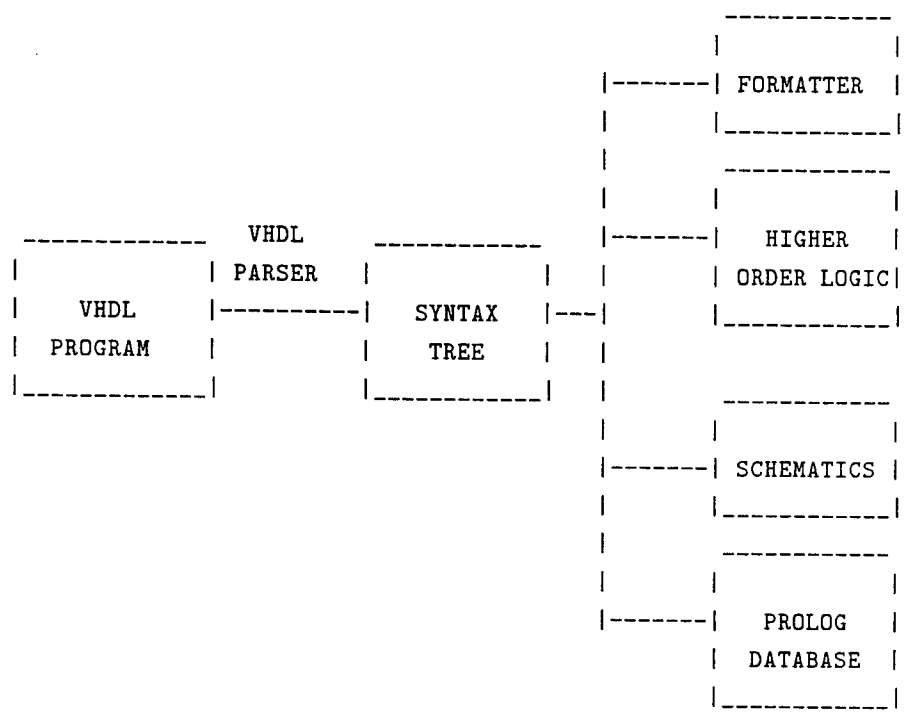
Introduction

VHDL is an acronym for VHSIC (Very High Speed Integrated Circuits) Hardware Description Language [2]. It is a "formal" language for specifying, documenting, and communicating hardware designs and models at many levels of abstraction ranging from the behavioral level to the gate level. The description can also be used to test the correctness of the circuit via simulation, or to synthesize a gate level description.

As the hardware designs get larger, the corresponding VHDL descriptions become unwieldy. To access and organize information about the hardware components of the designs (such as gates, adders etc) and the software components of the design descriptions (such as subprograms, processes etc), we need an "intelligent browser". This can be created by assimilating the VHDL description in a suitable database, and then, querying the database to obtain relevant information about the hardware and the software components.

The goal of the VHDL parser written in Prolog by Peter Reintjes [9] was to transform a VHDL program into a Prolog *term* representing its abstract syntax. Once this term is asserted into the Prolog database, simple Prolog programs can be written to access information stored in it. Furthermore, the abstract syntax tree can in fact serve as the basis for building a formatter/pretty printer; with appropriate graphics support, a schematics generator can also be written. We can even translate suitable subsets of VHDL into languages such as HOL [4] or Prolog [1] to verify the functional correctness, or into Prolog to check structural correctness of the layout [3]. These potential uses of the parser are shown in the block diagram on the next page.

This report describes the upgrading of a VHDL-87 parser [9] to conform to the IEEE Standard 1076-1993 of VHDL [6]. The original parser, which is based on a simplified grammar of VHDL-87 [8], was first modified to conform to the IEEE Standard 1076-1987 [5]. Subsequently, it was extended to incorporate the novel features of VHDL-93 as described below.



Discussion of the Problem

We now enumerate the errors in the original parser and the extensions made to it to conform to IEEE Std 1076-1993 of VHDL. We assume that the reader is familiar with [9] (or at least [8]) and has access to [6].

Errors in Tokenizer

The VHDL-87 tokenizer [9] does not conform to IEEE Std 1076-1987 [5] as follows.

- In VHDL-87, the keywords and the ordinary identifiers are not case-sensitive, while strings are. The original tokenizer did not fully implement this aspect.
- The based literals (numbers in base 2 to 16) were not recognized.

- The comments at the end of a program file caused error.
- Strings and extended identifiers were permitted to contain "non-tab-format-effectors".
- The use of ticks in VHDL-87 in different contexts, such as heralding an attribute or delimiting a character literal, poses problems to the tokenizer. For instance, constructions such as `character'('a')`, `a'b'c'd`, etc. are tokenized incorrectly. (However, the intepretation of `range` as an attribute or as a keyword has been dealt with correctly.)

Extensions to Tokenizer

The IEEE Std 1076-1993 [6] enriches the lexical structure of the language in the following ways.

- The alphabet has been extended from seven-bit ASCII to ISO eight-bit coded character set (ISO 8859-1987).
- Extended identifiers have been introduced to allow identifiers with non-alphanumeric characters. These identifiers are case-sensitive (in contrast with ordinary identifiers and keywords).

Errors in Parser

The VHDL-87 parser [9] does not conform to IEEE Std 1076-1987 [5] as follows. (However, it must be mentioned here that the original parser explicitly assumed that the input VHDL-program is syntactically correct.)

- The original VHDL-87 parser incorrectly permitted all possible declarative items for all construct types (that can contain declarations). For instance, it permitted subprograms in configuartion statement, attribute specifications in package body etc.
- The definition of `interface_element` imposes constraints on the mode, the bus-type, and the initialization expression based on the object-class. The original parser permitted all combinations.
- The original parser permits the names of entity, configuration, component, unit, function etc. to be only an identifier or a selected name. This is not

entirely correct because it does not accomodate function calls such as `Clock'Delayed'Stable(T)` etc.

- It incorrectly allows records with no fields, but omits zero-arity function calls and incomplete type declarations.

- The original parser does not enforce restrictions such as non-associativity of `nand`, `nor`, and `**`, or ensure proper use of sign in expressions like `( 5 + - 5 )` etc. Furthermore, it does not permit valid strings such as `( - abs 4 * 2 + 8 )` etc. which contain a legal sequence of unary operators.

Extensions to Parser

The IEEE Std 1076-1993 [6] extends the syntax of the language in the following ways. (The first three changes contribute to the *orthogonality* of the language.)

- The *end* statements are now more uniform in that they may optionally specify the construct type and the corresponding identifier/label. For example, `end entity NAME`, `end package body NAME`, `end if LABEL`, `end function NAME`, etc. are now permited.

- The use of keyword `is` is more consistent in VHDL-93. For instance, `block NAME is`, `component NAME is`, and `process NAME is` etc. are also permitted.

- The use of labels is more consistent in VHDL-93. For instance, the *if* statement and the *case* statement can have a label now.

- VHDL-93 introduced the *report* statement that is actually a special case of the assertion statement in wide use.

- VHLD-93 introduced the keywords `pure` and `impure` in subprogram declaration, and the keyword `shared` in variable declaration.

- VHLD-93 introduced the keywords `inertial` and `reject` in delay mechanism (in signal assignment statement), the keyword `unaffected` in waveform, and the keyword `postponed` in concurrent assertion statement, concurrent procedure call statement, concurrent signal assignment, process statement etc.

- The syntax of file declarations has been changed. In fact, the VHDL-87 file declarations are no longer compatible with the VHDL-93 standard. Furthermore, the file names can now be passed as parameters to subprograms.
- The *generate* statement now has an optional declaration.
- The clause for conditional-waveform has been modified to accomodate `waveform when condition`-form (that is, an else-part is not required).
- VHDL-93 supports *group*, *group template* and *signature* constructs.
- VHDL-93 generalizes component_name to *instantiated_unit* in component instantiation statement.
- VHDL-93 introduces the *shift operators* and the *shift expressions*.

Methodology and Results

The tokenizer and the parser were modified to reflect the corrections to the original parser, and to incorporate changes to meet the VHDL-93 standard. Here, we describe only a few important changes to the parser, in detail.

Changes to Tokenizer

The tokenizer code has been substantially revised to overcome the problems stated above. Here, we discuss only the tick problem, in detail.

The handling of ticks in VHDL is relatively complex. This is because, ticks can occur in several different contexts that can "interact" in complicated ways, as explained below. A single tick can be used as a *separator* in a qualified expression (such as `typename'(value)`) and in attribute names (such as `prefix'attribute`). A pair of ticks can be used as *delimiters* in character literals. Given these facts, the lexical analysis of constructions such as `character'('a')`, *A'B'C*, `T'range of range'a'to'b'`, and `"or"'A'B` (borrowed from internet postings of Joerg Lohse and Jacques Rouillard and [7]) becomes rather involved. However, the following disambiguation rule (adapted from Gary Beihl's posting on the internet) can be used to guide the tokenizer:

> A tick begins a character literal *if and only if* the previous token is *not* a non-keyword identifier or attribute name or a close paren or a

string literal and there is a matching tick to terminate the character literal exactly two characters ahead in the input stream. In other cases, the tick can either flag a potential attribute designator or is a token by itself.

To implement this rule, we need to know the previous token type and have the ability to peek one character ahead (in addition to the one lookahead we already have in the tokenizer). The VHDL-93 parser compiled for SUNs handles the quirky examples involving ticks satisfactorily, while the VHDL-93 parser compiled for Ultrix does not, because the underlying Quintus Prolog implementation for Ultrix does not support the built-in `peek_char/1`. In summary, with the additional "peeking" capability, it is possible to make the tokenizer (lexical analyzer) an independent module of the parser.

Lexical Overhead

We now discuss certain elements of the lexical structure of VHDL-93 that seem to complicate the implementation without really contributing much to programmer convenience [7].

Identifiers and Abstract literals

In the syntax of identifiers and abstract literals (numbers), successive underscores are not permitted. The implementation can be simplified by permitting sequence of underscores without sacrificing compatibility with VHDL-87. Similarly, not allowing an underscore to start the fractional part or the exponent part of a number seems unnecessary. If the fractional part is zero, a zero need not be required to follow the period. Banning an empty extended identifier does not seem justified.

Replacement Characters

In the syntax of based literals, the two matching double quotes delimiters can be replaced by two matching percent characters. The implementation can be simplified by allowing double quotes to be matched with percent. However, this simplification will not work for strings.

In the syntax of strings, the VHDL-93 does not treat the double quotes delimiter and the percent delimiter similarly. In particular, the "reasonable" string %...#...% is illegal.

The Tick Problem

The use of tick as a separator in attribute names, as a token in qualified expressions, and as a delimiter in character literals, permits us to create VHDL-93 constructs that are "difficult" to tokenize. Replacing the tick with another character in attribute names and qualified expressions could have simplified matters. However, the backward compatibility will pose a problem then.

Changes to Parser

The original parser code was corrected and then revised to reflect the "cosmetic" changes and new additions made to upgrade VHDL-87 to VHDL-93, as explained before. In particular, parsing of expressions has been completely redone to enforce restrictions such as non-associativity of **nand**, **nor**, and ******, and to ensure proper use of sign in expressions like `(- 5 + 5 )`. A complete description of the design and implementation of the VHDL-93 parser is documented in [10].

Overall, the performance of the VHDL-93 parser has degraded somewhat on the available VHDL-87 test suite, compared to the original parser. When compiled, the VHDL-93 parser takes about 30% longer. When interpreted, the latter number can be as high as 50%.

Conclusions

The entire VHDL-93 parser is implemented in a little over 2000 lines (about 1000 clauses) of Prolog, thus making it between one fifth and one tenth the size of other VHDL implementations [9]. It has been tested on the VHDL-87 test suite and a few VHLD-93 programs. The implementation exploits the "first-argument indexing" capability of Prologs and is reasonably efficient. The error detection and reporting in the tokenizer is satisfactory, but the error reporting in the parser needs work. The parser described in [10] can become the basis for a wide variety of tools whose input is a VHDL program. For instance, a VHDL pretty printer, and a VHDL program browser can be built on top of this parser. The parser can also be effectively used to translate VHDL programs into other languages (like HOL and Prolog) for specific applications (like hardware verification or circuit extraction) [1, 4, 3].

References

[1] Barrow, H., "VERIFY: A Program for Proving Correctness of Digital Hardware Designs," Artificial Intelligence, Vol. 24, No. 1-3, 1984.

[2] Bhasker, J., *A VHDL Primer*, Prentice Hall, Inc., 1992.

[3] Dukes, M., Brown, F., and DeGroat, J., *Verification of Layout Descriptions using GES*, 29th Design Automation Conference, pp. 63-72, 1992.

[4] Gordon, M., "HOL : A Proof Generating System for Higher-Order Logic," *VLSI Specification, Verification, and Synthesis*, Boston: Kluwer Academic Publishers, pp. 73-128, 1988.

[5] *IEEE Standard VHDL Language Reference Manual, Std 1076-1987*, IEEE, NY, 1988.

[6] *IEEE Standard VHDL Language Reference Manual, Std 1076-1993*, IEEE, NY, 1993.

[7] Levia, O., Maginot, S., and Rouillard, J., "Lessons in Language Design: Cost/Benefit Analysis of VHDL Features," 31st Design Automation Conference, pp. 447-453, 1994.

[8] Lipsett, R., Schaefer, C., and Ussery, C., *VHDL: Hardware description and design*, Boston: Kluwer Academic Publishers, 1989.

[9] Reintjes, P., "VHDL Parser in Prolog", Technical Report, Microelectronics Center of North Carolina, Reasearch Triangle Park, 1990.

[10] Thirunarayan, K., Reintjes, P., and Ewing, R., "VHDL-93 Parser in Prolog", Technical Report, Wright Laboratory, Wright Patterson Air Force Base, 1994.

DEVELOPING QUALITATIVE PROCESS CONTROL DISCOVERY SYSTEMS

Robert B. Trelease, Ph.D.
Adjunct Assistant Professor
Department of Anatomy and Cell Biology

UCLA School of Medicine
10833 Le Conte Avenue
Los Angeles, CA 90024-1763

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

September 1994

DEVELOPING QUALITATIVE PROCESS CONTROL DISCOVERY SYSTEMS

Robert B. Trelease, Ph.D.
Adjunct Assistant Professor
Department of Anatomy and Cell Biology
UCLA School of Medicine

Abstract

The theory and application of artificial intelligence and qualitative process modeling methodologies were studied in the contexts of process control for automated composite materials curing and hyperbaric oxygenation. Knowledge bases, quantitative/qualitative data encoders and decoders, serial port data acquisition routines, expectation/exception handling, envisionment evaluation, and rule mutation heuristics were developed to support discovery capabilities. The principal objective was to develop a discovery system capable of controlling materials processes, recognizing sensor errors and cure exothermy, and generating new process rules based on evaluation of real experimental data against the existing knowledge base heuristics of a fundamental qualitative process model. A secondary objective was to begin the construction of a system to model and to conduct discovery for processes involved in hyperbaric wound healing. The operation of fundamental qualitative process models was demonstrated, and system heuristics were successfully created for identifying unexpected states, generating new rules, and controlling on-line process parameters.

DEVELOPING QUALITATIVE PROCESS CONTROL DISCOVERY SYSTEMS

Robert B. Trelease, Ph.D.

Introduction

Contemporary models and analyses of complex physical systems conventionally rely on quantitative methods (e.g., multiple transfer functions, differential equations, and parametric statistics) to characterize functional relationships between different natural and artificial processes. Certain classes of model problems however, deal with unquantifiable relationships and/or incompletely understood processes. To model these problems, qualitative reasoning methods have been developed using cognitive theory and symbolic computer programming techniques employed by artificial intelligence (AI) research modeling of cognitive processes in expert problem-solving. (Weld and De Kleer, 1992).

Knowledge-based qualitative reasoning approaches such as qualitative process (QP) theory (Forbus, 1984) recognize that human experts typically comprehend processes and scientific concepts in qualitative terms (e.g., current increases as voltage rises given a constant electrical resistance) despite the fact that such relationships might be most deterministically described in formal quantitative terms (e.g., I=E/R). As used by scientific experts and other problem-solvers, qualitative reasoning can thus be thought of as underlying or preceding the formulation of quantitative descriptions of systems and dynamics (Weld and De Kleer, 1992).

Computer-based qualitative reasoning methods have been successfully used to solve physics problems that are intractable when approached with quantitative techniques (De Kleer and Weld, 1992). Qualitative simulation and modeling have been used for testing system theories, for controlling processes, for diagnosing problems, and for developing "discovery systems" that evolve new process theories based on model behaviors (Weld and De Kleer, 1992). In the realm of process control, QP systems have been used to model and regulate the production of polymer-based composite materials cured in the complex pressure/temperature/time environment of an autoclave (LeClair and Abrams, 1989; LeClair et al., 1989).

The purpose of the research project described here was to study and to assist in the development of knowledge-based discovery systems for process control based on qualitative process modeling. **The principal focus was the application of QP system discovery methods to support automated knowledge base development for autoclave curing of composite materials.** *A secondary focus, constituting applied technology transfer, was to begin the creation of a QP modeling system to support discovery in hyperbaric oxygenation, a medical treatment method (in regular use by the Air Force Medical Corps) that is comparable in aspects of its application to autoclave curing of composite materials.*

Work on composite curing discovery system issues was performed in collaboration with Frances L. Abrams (Wright Laboratory, Wright-Patterson AFB, Dayton, OH) who created the autoclave knowledge base and prototype discovery system and generously provided hours of instruction and discussion on process control and discovery, the basis of her dissertation research. Work on hyperbaric oxygenation was performed in collaboration with Colonel Richard Henderson USAF MC, Chief of Medical Operations of the Hyperbaric Medicine Unit, 645th Medical Group, Wright-Patterson AFB, Dayton Ohio. Dr. Henderson provided key insights into critical issues and problems in hyperbaric medicine and processes of wound healing. Assistance with qualitative process and discovery theories and system internals was obtained from Jack Park, developer of the symbolic programming tools used for this work.

Methodology and Discussion of Problem Domain Issues

All work for this project was performed on Apple Macintosh II computer systems with 10 or more megabytes of random access memory and 105 megabyte or larger hard disks. All software systems development was performed in a Macintosh-based symbolic programming environment constituting an extensible, knowledge-based, qualitative process modeling system. System development was rooted in accepted practices and theories in AI, cognitive sciences, epistemology, and scientific discovery. In Section I, below, we discuss some of the basic tenets and methodology of AI discovery systems. In Section II, we outline fundamental considerations for building discovery systems using The Scholar's Companion. In Section III, we examine specific problem domain issues for autoclave curing of composite materials and for hyberbaric oxygenation treatments for wound healing.

I. Strategies for Discovery in Knowledge-Based Systems

Kocabas (1991) has distinguished four loosely defined, overlapping types of knowledge necessary for research and discovery: Commonsense, technical, theoretical, and methodological. All of these types are currently computationally representable with symbolic programming methods.

In expert systems and related knowledge-based systems, knowledge is typically represented in the form of heuristics or production rules composed in an object-oriented symbolic programming language (typically a LISP dialect). The collection of rules for a given system is referred to as a knowledge base (KB), and it encompasses theories, hypotheses, and functional descriptions for entities and processes for the problem/model domain. Inferences about real-world problems can be made by an expert system's inference engine, which evaluates data (i.e., from external processes) via forward- or backward-chaining application/evaluation of production rules.

Early work with expert systems focused on strategies for supporting diagnosis (identification of faults and problems) via the interaction of simulation and inference (Davis, 1983). One such method was termed discrepancy detection and relied on handling discrepancies between simulation inferences and real device data. Given the context of diagnosis, discrepancy detection thus qualified as a strategy for discovery.

In the current qualitative process modeling environment, a comparable method for achieving discovery involves relating the expectations explicit in simulation heuristics to exceptions (discrepancies) encountered in real systems data. Specifically, a process simulation (envisionment) including behavioral expectations is compared with experimentally derived data, and different expectation failure process rules (handlers) are used to generate new envisionment heuristics and/or to output instructions (via decoders) for control of the real process being modeled. This procedure seeks to emulate symbolically the high-level logical processes involved in human scientific discovery. A good scientist will create a hypothesis system (model), perform experiments to test it, evaluate results in terms of the specific hypothesis and current beliefs, revise hypotheses, and design additional evaluative experiments. In this frame of reference, experimentation is thus a part of most discoveries. Recognizing this, a few machine discovery systems have utilized experimentation as integral parts of their structures (Shrager and Langley, 1990).

QP reasoning methodology, in fact, emerged from expert systems research directed at emulating related problem-solving. While expert systems demonstrated domain-specific and global limitations related to the need for extremely large KBs to represent adequately real experts' knowledge, useful qualitative modeling tasks and problem-solving could be performed with relatively compact KBs. In the context of discovery, QP models might be effectively employed in systems for evaluating experimental data and evolving new model heuristics (hypotheses).

In order to function effectively in process discovery, QP systems must possess some special characteristics related to handling functional actor and state changes. Discrete states must be recognized as changes occur. Model must be general enough to allow for variations in real world data, but expectation failures must recognized and handled. Discovery systems can work to generate new theories by induction, deduction, or abduction.

Practical discovery systems remain somewhat like the "Holy Grail" of some AI researchers, and relatively few working systems exist (Kocabas, 1991).

II. Discovery System Issues and The Scholar's Companion

Discovery system development for this project was undertaken using The Scholar's Companion (TSC; ThinkAlong Software, Brownsville, CA) a symbolic programming environment composed of nested, extensible multiple language interpreters/compilers, run-time utilities, an envisionment builder, and basic tools supporting expectation definitions, failure handling, data file and database loading, device/process control, envisionment study, and guided process rule mutation.

A. Knowledge Bases, Tasks, and Envisionments

Comparable to a more conventional AI expert system, TSC allows the definition of actors (e.g., different types of cells), taxonomic and functional relationships, physical and biochemical processes, and predicated interactions and behaviors in hierarchical sets of symbolic production rules. TSC also compiles the description of initial conditions and functional designs for simulated experimental trials to be run using defined KB actors (cells) and process behaviors. As distinguished from an expert system that uses a logical inference engine to evaluate acquired data in the context of logical relationships defined by the KB, TSC employs an "envisionment builder" to record and to visualize the evolution of process behaviors and cell interactions resulting from initial experimental conditions given the defined actors, substances, processes, and functional relationships. An envisionment is thus the depiction of all possible behaviors that can be produced given the initial conditions of a simulation/experiment and the specific actors, states, relations, predicates, and process rules defined in the related KB.

TSC performs its functions under the direction of TASK statements. FILLIN tasks, for example, set up the system behaviors necessary to perform a simulation based on actors, states, relations, predicates, and process rules defined in the KB. Other tasks can be created to set up system behaviors to study specific envisionments. STUDY tasks can result in the identification of the sequence of process rule firings necessary for achieving a desirable process outcome. Other study tasks can be created to compare real world data with "expectations" defined in envisionment (process simulation) episodes, a crucial behavior in setting up TSC to produce discovery about processes. Rule mutation behaviors can be employed following study tasks to produce new process rules that can be used in building new envisionments. Finally, control tasks can be set up to handle process control based on process simulation envisionments.

B. Encoders and Decoders

Encoders and decoders are TSC code tools that (respectively) translate real-world data for use by process rules and translate process rule output into a form that can be used for interactions with processes external to the TSC environment, such as process control. For example, in order for temperature data to be available for TSC process rule heuristics, specific encoders must be written translating real numerical data into variables compatible with the qualitative environment.

C. Expectations and Expectation Failure Handlers

Expectations are outcomes (actors and/or states and/or relations) that are reasonably expected to occur as process rule conditions are satisfied: That is, they are conditions anticipated in encoded real-world process data consistent with the terms of the envisionment model. Expectations are coded in process rules by including the left-hand side (consequent) slot THEN. EXPECT as in :

THEN.EXPECT ((increasing (*temp) true))

This specific example creates the expectation that *temp should be increasing as a consequence of the firing of the process rule in which the slot is included. Such rules transform a simple qualitative domain model into an anticipatory model capable of supporting discovery via behavioral rules that respond to expectation failures (failure of encoded real-world data to fulfill expectations coded in the process model).

TSC is currently equipped to handle two different kinds of expectation failures:

```
events that do not occur when expected

events that occur when unexpected.
```

Expectation failures of the first type occur when real processes fail to yield consequences (actors, states, or relations) as defined in the process rules of the model envisionment. Expectation failures of the second type occur when real processes yield consequences undefined or different from those encoded in process rules of the model envisionment.

The two previously designated classes of expectation failures presume that process rules and encoders are properly written AND that incoming data is not erroneous. De facto expectation failures may occur if erroneous data is input or if encoded data terms (actors, states, and relations) do not match those of process rules containing expectation slots. This latter condition may result in encoded data failing to match any envisionment process rules. During testing and debugging of prototype discovery systems it is thus necessary for the developer to compare encoded data episodes with envisionment process rules to insure that a given system is capable of properly handling expectation failures likely to be encountered during control or study of a modeled process or during the study of a database in terms of a defined data model.

In order to support discovery, expectation failures must build tasks for TSC to execute. The effect of these tasks will depend on the domain-specific behavior rules created in the KB to handle the specific expectation failures anticipated.

D. Explanation, Rule-building Strategies, and Building New Envisionments

Explanation seeks to account for system behaviors in terms of defined process heuristics. When new system behaviors are encountered, explanation must involve building of new processes rules to account for those behaviors. This involves rule-building heuristics that mutate or build on existing process rules, in addition to system tasks that study experimental data and generate new envisionments based on new rules.

1. Process Rule Mutation: Directed Evolution

TSC is intended to support new rule building by methodology referred to as directed evolution.

a. By Simple Heuristics

In a simple heuristic approach, defined behavioral rules guide selected mutations (changes in antecedent and/or consequent terms) applied to the original set of rules. Different types of behavioral rules are applied to the processes of creating new process rules and of performing mutations to existing rules.

In the following example, we show how a full Eurisko-like set of behaviors is programmed into a TSC KB. To do this, we look at the notion of a *cycle* in which the following behaviors are involved:

```
Find something to do (top priority on agenda)
Do it
Study what you just did
```

This central cycle is the internal "loop" of TSC. To make it work, behavior rules, process rules, and sense rules must support the cycle.

b. By Genetic Algorithm

In a genetic algorithm (GA) approach, a collection of effective process rules are mutated by random combinations splicing terms (slots) from different rules. The GA serves as a guided optimization system in that, over a number of "generations," it selects elements and combines them into predictive rules. Rules are successively rated for their predictive accuracy during the learning exercise and successful outcomes are reinforced. This process attempts to mimic environmental factors which favor particular genes in natural selection processes occurring in populations of biological organisms. This feature is supported but not yet implemented in TSC.

III. Problem Domain Issues

A. Autoclave Curing of Composite Materials

The principal focus of discovery system development during this project was control of polymer composite material curing within an autoclave. This process has several discovery-relevant aspects involving details of the composite part, the part container and tool, sensors, and the autoclave and its control system.

Composite parts are constructed of layers of fibrous cloth-like material (parallel carbon, aramid, or glass fibers) impregnated with monomer resin. Layers of this material, called prepreg, are laminated to produce the desired part shape/size. The alternating configuration of the fibers in prepreg layers give the finished part its characteristic strength and stiffness properties when the resin is cured (polymerized).

The laminated prepreg part is placed on a metal mold (tool) and covered with a flexible bag that is sealed over the part. A teflon sheet and matting is placed between bag and part to prevent sticking. A vacuum line attached to the

bag removes the air and volatile materials which may be in or surrounding the part. This process helps to reduce voids that might form between layers in the part. The part/tool/bag assembly is referred to as a lay-up.

For controlled curing, the lay-up is placed inside of the autoclave, a large pressure vessel that can be heated by circulating gas. Pressurizing the autoclave with heated gas initially lowers resin viscosity, presses fiber layers together into a continuous laminate, and provides the thermal activation energy for resin polymerization. A properly cured part is composed of fibers embedded in a solid piece of polymer without cracks or voids. If the resin is sufficiently cured, the strength and stiffness of the composite part is determined by the number and orientation of fiber layers and by the ratio of fiber to polymer (matrix).

A number of undesirable features can arise during an inadequately controlled cure cycle. As previously noted, voids within the part may be created by the formation of gas bubbles. Large temperature gradients can cause cracks or residual stresses. Finally, since heat is generated by polymerization, there can be a runaway exothermic reaction that can burn the part.

Variables controlling the curing process include temperature and pressure of the autoclave and the magnitude of pressure or vacuum applied to the lay-up bag. Heat applied by increasing autoclave temperature can increase the degree of cure. However, a sustained temperature increase can slow the cure because of viscosity changes in the resin. Increasing pressure increases the efficiency of heat transfer within the pressure chamber, and too much or too little pressure at the wrong time can cause the formation of bubbles and voids. There are very complex, non-linear relations between autoclave temperature, pressure, bag pressure, heat transport, resin composition and viscosity, and the desired strength, temperature and chemical resistance and other characteristics of a cured part. These relationships are far too complex and poorly understood to discuss in this report. The complexity of these relationships is, in fact, the very reason why artificial intelligence (expert system) approaches have been employed in the quest to improve composite parts curing and autoclave control.

During a cure cycle, only temperature and pressure measurements can be obtained from the part and the autoclave. Pressure measurements made with diaphragm transducers are only presently reliable for the autoclave and not for the laminate/part or for the interior of the bag. Thermocouple temperature transducers will work for the part and the autoclave, and a spatial array can be used to infer rate and direction of heat transfer.

Autoclave heating, pressurization, lay-up back pressurization/vacuum, and gas supplies are controllable via a computer serial port interface. For development of simulation and control software, an autoclave simulator is available. The autoclave simulator runs on a separate personal computer and communicates with the discovery system (TSC) host computer via a standard RS-422 serial port interface.

To support optimization of composite parts curing, autoclave control, and the automatic creation of new process heuristics, an iterative process of KB construction, envisionment building, process data acquisition (real or simulated), envisionment-data study, and new rule design was designated. As illustrated in Figure 1, below, this process was intended to allow both on-line (during cure) and off-line (after cure) discovery (expectation failure handling) as well as building of new KB heuristics and control of composite curing based on KB heuristics of successful envisionments.

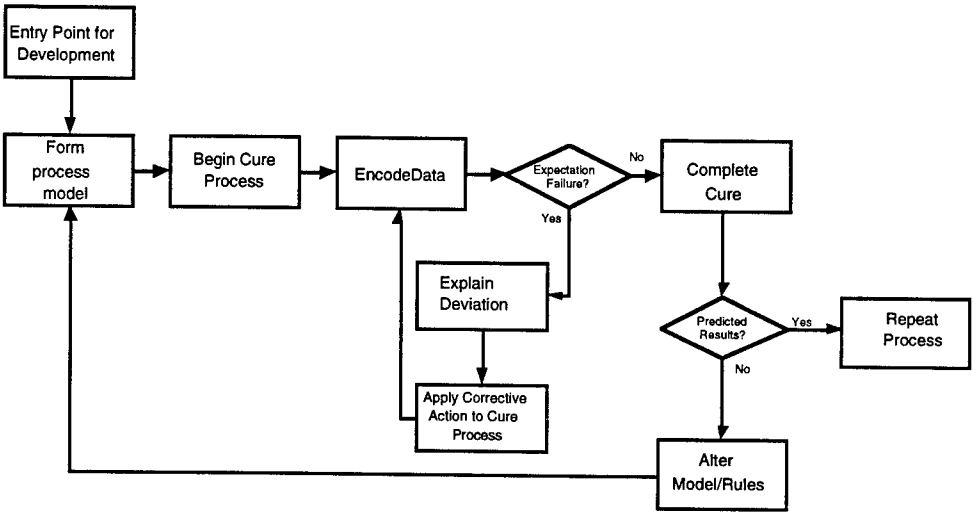


Figure 1: Flow Diagram of Desired System Behaviors for Automated Knowledge Base Development for Composite Curing

B. Hyperbaric Oxygenation

Hyperbaric oxygenation (HBO) is a medical treatment that in certain aspects of its application resembles the use of an autoclave in the curing of composite materials. Individuals with specific types of diseases or pathological conditions are subjected to elevated atmospheric pressures (typically in excess of 2 atmospheres absolute [ATA]) in a treatment chamber for repeated periods under alternating conditions of pure oxygen (O2) and normal air breathing. HBO is an accepted and efficacious treatment for decompression sickness (DCS), healing of problem wounds and compromised skin grafts/flaps, mixed soft tissue infections and gangrene, burns, soft tissue and bone radiation necrosis, and carbon monoxide and smoke poisoning (Hyperbaric Oxygen Committee, 1992).

Specific HBO treatment protocols for individual conditions have been empirically derived, and many of the underlying processes involved (whether cellular, physiologic, or systemic) are poorly understood or unknown. **Given this incomplete understanding, it is virtually prohibitive to develop improvements in HBO protocols or to tailor treatment regimens to individual patients and conditions.**
Because of the variety of problems treated with HBO, it was decided to focus on one treatment paradigm for discovery system development. Because of the frequency with which such treatments were conducted at WPAFB Medical Center (daily) and the large number of patients treated, it was decided to focus on wound healing treatments and understanding the interaction of basic cellular repair processes with the wound healing protocol. Figure 2 is a flow diagram showing the details of the HBO wound healing protocol.

In order to support discovery in the HBO-wound-healing environment, we began development of a TSC KB containing details of cell biology for human tissues, knowledge of HBO-relevant concepts and protocols and wound-healing process rules. The cell biology KB component (HISTO.T) was built upon a previously constructed biology KB developed for modeling immune system processes and behaviors (Trelease, 1994). All concepts and process rules were derived from current theories and objective information published in the current general biological (Purves et al., 1992), systematic (Fortuner, 1993) histological (Weiss, 1983), specialized immunological (Abbas et al., 1991), HB0 (Hyperbaric Oxygen Committee, 1990) and wound-healing literatures (Davis and Hunt, 1988; Kloth et al., 1990).

Compiled after the cell biology component was a set of actors, predicates, states, and process rules for the HBO treatment environment. One particular set of process rules was created for implementing HBO treatment cycles (pressurization to 2.4 A.T.A. and 100% O2 alternating with air breathing.

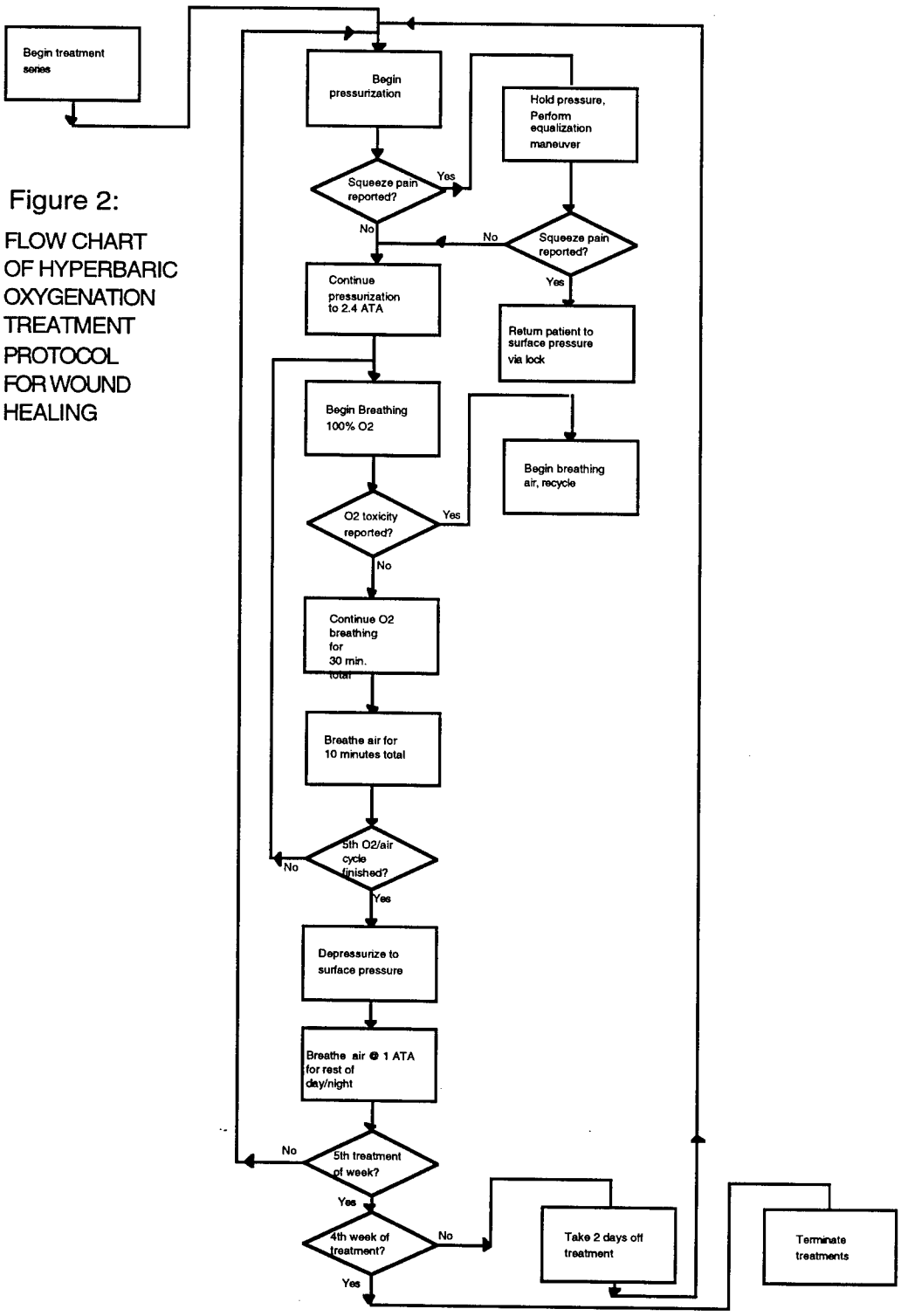


Figure 2:

FLOW CHART OF HYPERBARIC OXYGENATION TREATMENT PROTOCOL FOR WOUND HEALING

IV. Instrumentation and Sensors

Composite Curing

For the real application, the discovery/control system must deal with multiple temperature sensor (thermocouple), pressure transducer, and process cycle data communicated via an RS-232 serial port controlling the autoclave. For the purposes of KB, decoder, and process rule development, temperature, pressure and cycle data were provided by a Quick BASIC autoclave simulation running on an MS-DOS personal computer connected to the TSC Macintosh system via a null modem cable.

Hyperbaric Oxygenation

Numerous clinical monitoring devices can be used in HBO therapy. A central assumption in HBO theory relates accelerated wound healing to increased tissue oxygen tension in the affected region. Practical tissue clinical oxygen and carbon dioxide measurements are typically obtained by oximetry/carboximetry from electrodes attached to the wound and to unaffected tissue regions. We therefore developed programs for acquiring transcutaneous O2 and CO2 measurements from an array of Radiometer TCM3 clinical gas monitors. Given that the most immediate measure of patient O2 toxicity would be electroencephalographic (EEG; brain wave) activity changes (Kinneyt et al., 1980; Torbati et al., 1981; Dolmierski et al., 1990), provisions were made to interface and to acquire data from clinical EEG recorders.

Results

Composite Curing

At the beginning of the project period, there was no capability for TSC to handle serial port communications (i.e., with the autoclave simulator). There were no operational encoders or decoders, and work had just begun on process rules for the composite curing KB.

During the early weeks of the project, the serial port interface encoders were completed, and it became possible for TSC to communicate with the autoclave simulator computer. Subsequently, work progressed on writing symbolic process rules for the TSC KB and encoders and decoders for autoclave data. Initial difficulties were experienced with encoder and decoder function, but these problems were resolved as a matter of program code that was erroneously not being INCLUDEd (loaded) into TSC before operation.

Initial trials of envisionment building demonstrated that the system ran very slowly: It took all night to complete the cure simulation run. After examining rules involving temperature relations between specific sensors, it was decided to rewrite these rules, changing them from TSC RELATES slots to STATES slots. This yielded a great increase in envisionment building speed.

During the course of the project, additional functions were added to TSC to allow the user to specify desired outcomes and to create a task for determining the rules fired in proceeding from initial conditions to each desired outcome.

Initial trials with the composite curing model KB and the autoclave simulator yielded expectation failures that could not be handled by the existing STUDY task structure. First, expectation failures occurred because there was inadequate correspondence between process rules' actors/states/relations, and those of a given autoclave simulator (data) episode. These exceptions were handled by rewriting encoders and process rules to achieve a greater concordance between modeled and "real" (simulator) antecedents and consequents.

Hyperbaric Oxygenation Models

Initial experiments with process heuristics yielded a qualitative model that produced an acceptable HBO treatment cycle (chamber pressurization followed by alternating 100% O2 and air breathing periods). Figure 3, below shows the graphical envisionment of a typical wound healing protocol of 5 O2/air breathing cycles at depth per daily treatment.

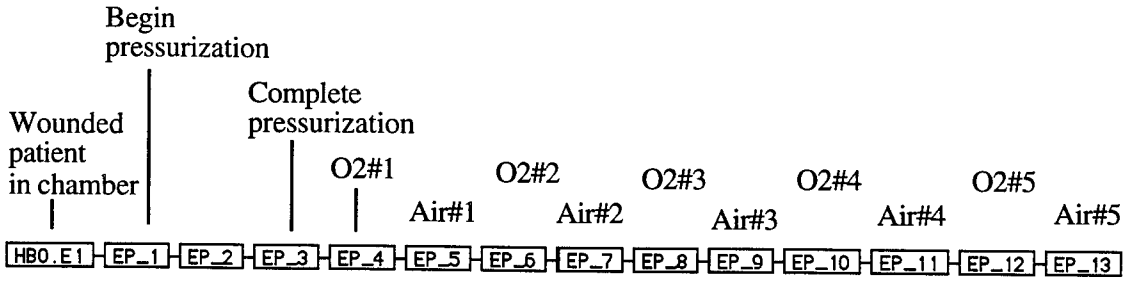


Figure 3: TSC frame-browser envisionment of a complete HBO chamber treatment session (5 O2/air breathing cycles).

Process rules were created for a phasic model of wound healing (Kloth et al., 1990) and enclosing the afflicted patient in the cycling hyperbaric chamber. At the end of the initial research, the overall model was demonstrating phasic outgrowth of new blood vessels with reestablishment of perfusion from the periphery to the center of the wound.

Conclusions and Discussion

During the course of the study period, practical knowledge bases were developed that constituted the cores of QP models against which experimental data could be evaluated. These data could be loaded from spreadsheet files, a computer-based process simulator, or from conventional data acquisition systems and instrumentation. System study heuristics and tasks were developed to allow favorable and unfavorable process outcomes to be recognized and to be identified, and process expectations to be designated. Real process data episodes could be compared to model envisionment episodes, and lists of matches (or the lack thereof) with process heuristics could be constructed.

Initial success was obtained in producing simple heuristic mutations of extant process rules by adding or subtracting antecedent and/or consequent terms (slots) based on differences between related envisionment and experimental episodes.

Full composite curing discovery system and control functions were not completed during the period of study, and additional work remains to be done to enhance the behavioral and task heuristics for producing new rules and thus to allow basic knowledge bases to be expand automatically as a result of evaluations of experiments. Furthermore, discovery system function once completed must be tested in experiments involving real composite parts curing in a functioning autoclave. Finally, to satisfy the principal investigator's research objectives, the system must be able to identify and to respond to mis-identification of temperature sensors as well as to the occurrence of an exothermic reaction during the curing process.

Much more work remains to be done in the HBO environment. A suitable, unified high-level model must be completed to allow patient wound-healing to be assessed on the basis of data acquired during treatments. Interfacing and data acquisition engineering must be completed to support encoding of patient information for use by the TSC system. Tissue O2 and CO2 measurements are currently being obtained and stored in spreadsheet format for system use. New methods were proposed for automated laser wound-scanning to assess the healing process, and work is currently under way to set up and to implement scanning of HBO patients. Given that neurological toxicity is the most likely problem to be encountered during treatment, provisions were made to set up and acquire patient electroencephalographic (brain wave) data during chamber sessions. Plans were made to study these data for frequency and synchronization changes known to accompany oxygen toxicity (Kinney et al., 1978; Dolmierski et al., 1990). Finally, to "close the loop" and link changes in the HBO process parameters with new discovery system heuristics, plans were made to use the TSC system in animal experiments in which gas/cycle parameters could be altered during the course of wound healing treatments.

TSC demonstrated itself to be a very powerful system for creating qualitative process models in the physical and biological materials domains. Given the extensibility of TSC and the modular approach adopted in KB construction, discovery should be supportable in other problem domain areas within materials sciences and biology once all discovery-related procedures (e.g., expectation failure-handling and rule mutation) have been fully implemented and testes. In particular, given the global approach taken to creating biological process heuristics, it should be possible to create systems handling problems in other areas of biosynthesis and molecular biology, pathology, physiology, neuroscience, and internal medicine. Since this work has emphasized integrating discovery with process control, it should be possible to build complete intelligent systems that conduct experiments, apply model-based reasoning to experimental data obtained, and automatically build new heuristics (refine models) for the processes being observed and controlled.

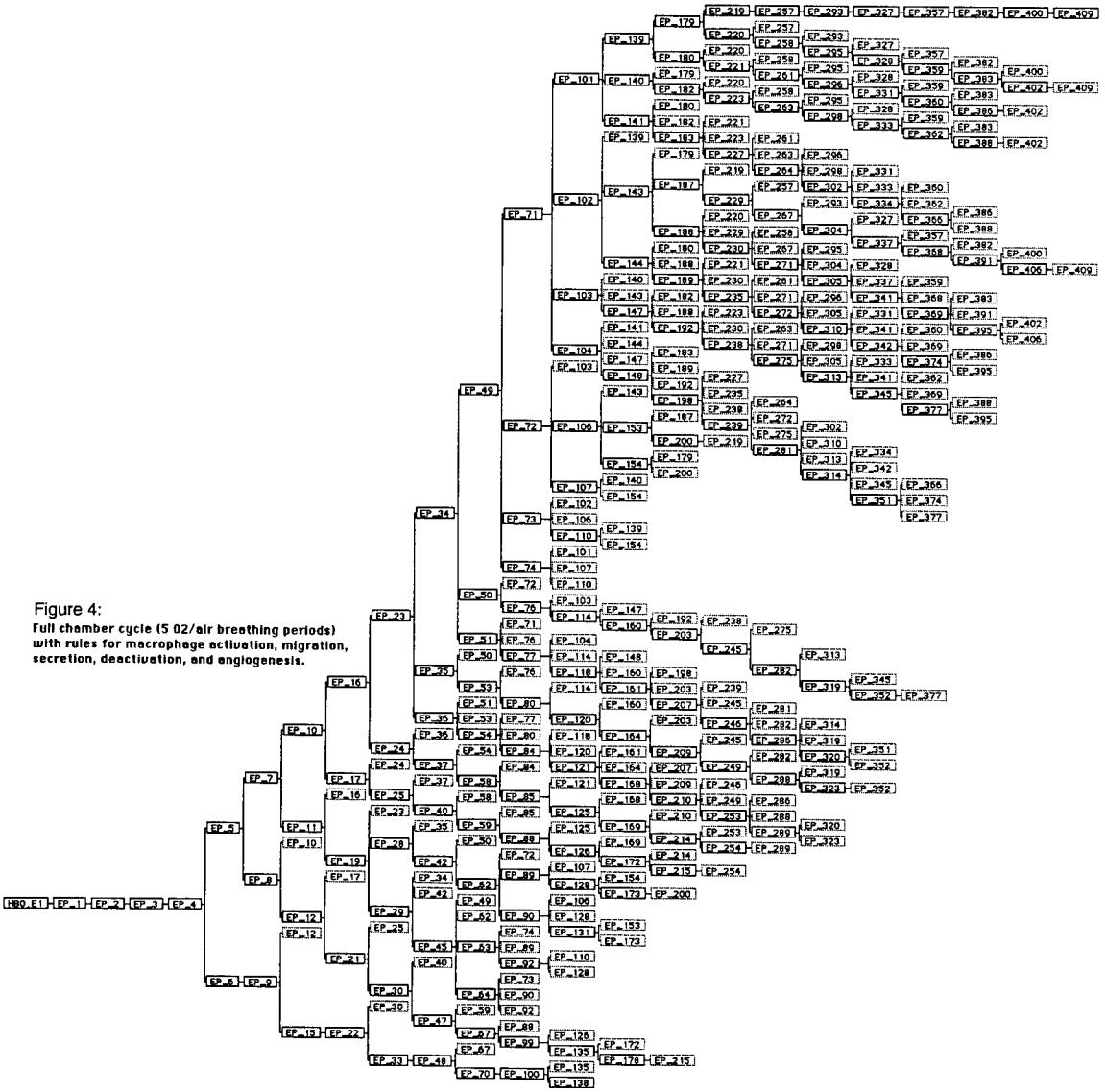


Figure 4:
Full chamber cycle (5 02/air breathing periods) with rules for macrophage activation, migration, secretion, deactivation, and angiogenesis.

References and Bibliography

Abbas, A.K., Lichtman, A.H., and Pober, J.S. Cellular and Molecular Immunology. W.B. Saunders Company: Philadelphia, 1991.

Bylander, T., Smith, J.W. and Svirbely, J.J. Qualitative representation of behavior in the medical domain. Computers and Biomedical Research 21:367-380, 1988.

Cohn, A. Rosenbaum, S., Factor, M., and Miller, P.L. DYNASCENE: An approach to computer-based intelligent cardiovascular monitoring using sequential clinical scenes. Methods of Information in Medicine 29:122, 1990.

Davis, J.C. and Hunt, T.K. Problem Wounds. The Role of Oxygen. Elsevier: New York, 1988.

Davis, R. Diagnosis via causal reasoning: Paths of Interaction and the locality principle. Proceedings AAAI-83 88-94, 1983.

De Boer, R.J. and Perelson, A.S. Size and connectivity as emergent properties of a developing immune network. Journal of Theoretical Biology 149:381-424, 1991

De Boer, R.J. Segel, L.A., and Perelson, A.S. Pattern formation in one and two dimensional shape-space models of the immune system. Journal of Theoretical Biology 155:295-333, 1992

De Kleer, J. and Forbus, K. D. Building Problem Solvers. MIT Press: Cambridge, MA, 1993.

De Kleer, J. and Weld, D.S. Qualitative Physics: A Personal View. In:Weld, D.S. and De Kleer, J. (Editors). Readings in Qualitative Reasoning About Physical Systems. Morgan Kaufmann: San Mateo, 1992. Pp 1-8.

Della Corte, F., Barelli, A., Arcangeli, A., Anile, C., Roselli, R., and Proietti, R. Computerized systemic and neurophysiological monitoring in patients with severe head injuries. Journal of Clinical Computing 16:144-155, 1988.

Dolmierski R., Maslowski J., Matousek M., and Petersen I. EEG changes measured by spectrum analysis under hyperbaric conditions. Aviation, Space, and Environmental Medicine 61(1):33-7, 1990.

Downing, K. The qualitative criticism of circulatory models via bipartite teleological analysis. Artificial Intelligence in Medicine. 2:149, 1990.

Forbus, K. D. Qualitative process theory. Artificial Intelligence 24:85-168, 1984.

Fortuner, Renaud (Editor). Advances in Computer Methods for Systematic Biology. The Johns Hopkins University Press: Baltimore, 1993.

Gaglio, S., Giacomini, M., Nicolini, C. and Ruggiero, C. A qualitative approach to cell growth modeling and simulation for cancer chemotherapy. IEEE Transactions on Biomedical Engineering 38(4):386-389, 1991.

Horvits, E.J. Automated reasoning for biology and medicine. In: Fortuner, Renaud (Editor). Advances in Computer Methods for Systematic Biology. The Johns Hopkins University Press: Baltimore, 1993. Pp. 3-27.

Hyperbaric Oxygen Committee. Hyperbaric Oxygen Therapy: A Committee Report. Undersea and Hyperbaric Medical Society: Bethesda, MD, 1992

Karp, P.D. Hypothesis formation as Design. In: Shrager, J. and Langley, P. (Editors). Computational models of scientific discovery and theory formation. Morgan Kaufman: San Mateo, CA, 1990. Pp. 276-317.

Karp, P.D. Frame representation and relational databases: Alternative information-management technologies for systematic biology. In: Fortuner, Renaud (Editor). Advances in Computer Methods for Systematic Biology. The Johns Hopkins University Press: Baltimore, 1993. Pp. 275-85.

Kinney J.A., Hammond R., Gelfand R., and Clark J. Visual evoked cortical potentials in men during compression and saturation in He-O2 equivalent to 400, 800, 1200 and 1600 feet of sea water. Electroencephalography and Clinical Neurophysiology 44(2):157-71, 1978.

Kloth, L.C., McCulloch, J.M., and Feedar, J.A. (Editors). Wound Healing: Alternatives in Management. F.A. Davis Company: Philadelphia, 1990

Kocabas, S. Computational models of scientific discovery. The Knowledge Engineering Review 6(4):259-305, 1991.

Kuipers, B. Qualitative simulation. Artificial Intelligence 29:289-388, 1986.

Kuipers, D. Causal reasoning in medicine: Analysis of a protocol. Cognitive Sciences 8:363, 1984

Kulkarni, D. and Simon, H.A. Experimentation in machine discovery. In: Shrager, J. and Langley, P. (Editors). Computational models of scientific discovery and theory formation. Morgan Kaufman: San Mateo, CA, 1990. Pp. 255-273.

LeClair, S.R., Abrams, F.L., and Matejka, R.F. Qualitative processs automation: Self-directed manufacture of composite materials. Artifical Intelligence for Engineering Design, Analysis and Manufacturing. 3(2):125-136, 1989.

LeClair, S.R., Abrams, F.L. Qualitative process automation. International Journal of Computer Integrated Manufacturing 2(4):205-211,1989.

Moeller, D.P.F. Advanced simulation in biomedicine. Springer-Verlag: New York, 1994.

Nordhausen, B. and Langley, P. An integrated approach to empirical discovery. In: Shrager, J. and Langley, P. (Editors). Computational models of scientific discovery and theory formation. Morgan Kaufman: San Mateo, CA, 1990. Pp. 98-128.

O'Rorke, P., Morris, S., and Schulenburg, D. Theory formation by abduction: A case study based on the chemical revolution. In: Shrager, J. and Langley, P. (Editors). Computational models of scientific discovery and theory formation. Morgan Kaufman: San Mateo, CA, 1990. Pp.195-224.

Purves, W.K., Orians, G.H., Heller, H.C. Life, the Science of Biology, Third Edition. Sinauer Associates: New York, N.Y., 1992.

Ruggiero, C., Giacomini, M., and Gaglio, S. A qualitative model of the dynamics of blood glucose and its hormonal control. Computer Methods and Programs in Biomedical Research 40(2):117-130, 1993.

Shrager, J. and Langley, P. Computational approaches to scientific discovery. In: Shrager, J. and Langley, P. (Editors). Computational models of scientific discovery and theory formation. Morgan Kaufman: San Mateo, CA, 1990. Pp.1-25.

Shrager, J. and Langley, P. (Editors). Computational models of scientific discovery and theory formation. Morgan Kaufman: San Mateo, CA, 1990.

Torbati, D., Simon, A.J. and Ranade, A. Frequency analysis of EEG in rats during the preconvulsive period of O2 poisoning. Aviation, Space, and Environmental Medicine 52(10):598-603, 1981.

Trelease, R.B. Development of a feature-based knowledge base for biologically-based materials and processes. Contributed Research and Development 145:1-10, 1994.

Uckun, S. Model-based reasoning in biomedicine. Critical Reviews in Biomedical Engineering 19(4):261-292, 1992.

Weiss, Leon. Histology. Cell and Tissue Biology. Fifth Edition. Elsevier Biomedical: New York, 1983.

Weld, D.S. and De Kleer, J. (Editors). Readings in Qualitative Reasoning About Physical Systems. Morgan Kaufmann: San Mateo, 1992.

A STUDY OF MASSIVELY PARALLEL COMPUTING ON EPIC HYDROCODE

C.T. Tsai
Associate Professor
Department of Mechanical Engineering

Florida Atlantic University
P.O. Box 3091
Boca Raton, FL 33431-0991

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

A STUDY OF MASSIVELY PARALLEL COMPUTING ON EPIC HYDROCODE

C.T. Tsai
Associate Professor

Department of Mechanical Engineering
Florida Atlantic University

Abstract

The massively parallel computing on EPIC hydrocode was studied. In order to apply the latest massively parallel programming language (MPPL) to the EpIC hydrocode for increasing the computing speed, several key algorithms are examined for the suitability of parallelism. Those algorithms are coded and tested in both MasPar MP-1 massively parallel computer and Cray T3D Massively parallel emulator. The results show that massively parallel computer can drastically increase the computing speed over single processor computer.

A STUDY OF MASSIVELY PARALLEL COMPUTING ON EPIC HYDROCODE

C.T. Tsai

Introduction

The rapid research and development of weapon system can respond quickly to the combat needs of operational commanders. The computer simulation of the warhead penetration can save huge amount of time and money on warhead design by allowing prototypes to be constructed, tested and accepted before investing in tooling. It allows the design to be refined on a computer instead of during expensive, time consuming test. The simulation of complex warhead penetration behavior requires a long CUP time even they are run in the Cray supercomputer. It is a bottle neck in the rapid research and development of warhead system.

The warhead penetration were simulated using EPIC hydrocode in Armament Directorate of Air Force Wright Laboratory. In order to reduce the CPU time of warhead penetration simulation, the suitability of massively parallel computing on EPIC hydrocode have to be investigated. Several key subroutines are coded and tested in Cray T3D and MasPar MP-1 massively parallel computer.

Background on Massively Parallel Computing

Until recently, advances in single CPU computer architecture and component technology have accounted for a significant portion of the increase in computational speed. However, the majority of the performance improvements in those computers came through increases in clock speed. As shown in Fig. 1, all single processor systems - whether they be supercomputers, minicomputers,

or workstations - are confronting the same fundamental law of physics as they work to shorten their clock cycles. Even modest increases in performance will become increasingly expensive in the coming years, and hence all of the interest in - and movement to - parallelism.

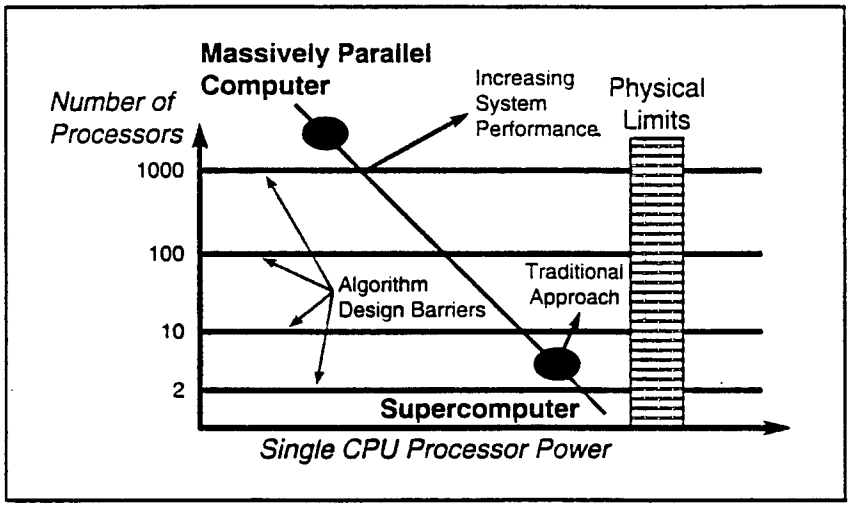


Fig. 1. Growth Path for Computer Systems

The fundamental idea behind parallel computing is to team CPUs together to solve single or multiple problems. If a single processor can solve a problem in 60 minutes, then two processors should, theoretically, be able to solve the problem in 30 minutes provided that the problem can be evenly divided between the CPUs. Fig. 2 depicts the theoretical improvements possible by moving multiple CPU architectures. However, some obstacles - primarily communications between CPUs - may reduce the theoretical improvements shown in Fig. 2.

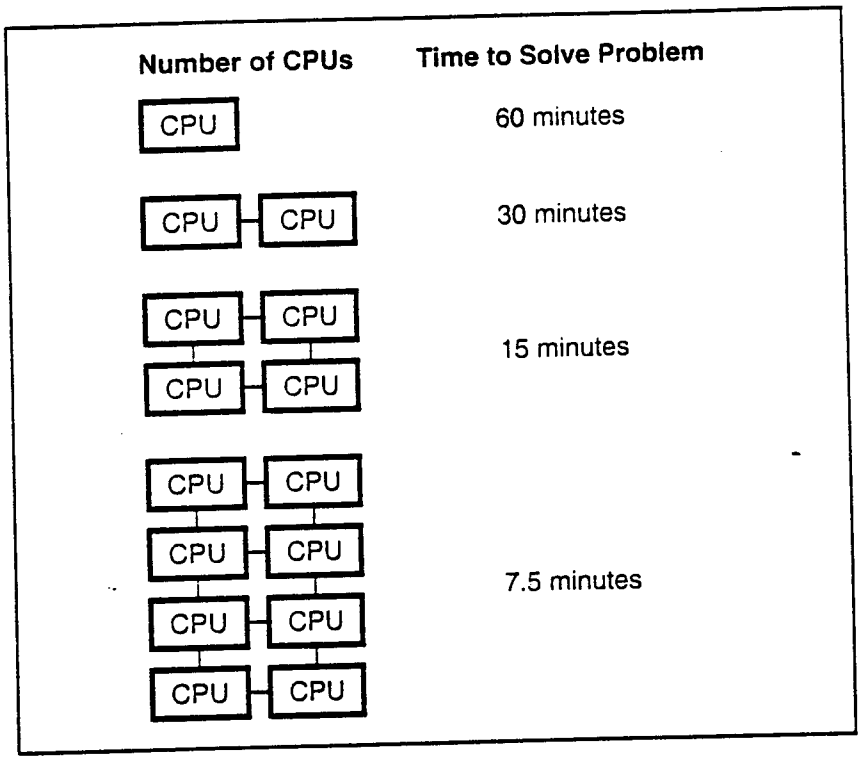


Fig. 2. Theoretical Improvements by Moving to Multiple Processors

To make maximum use of massively parallel computer, the data-sharing and work-sharing programming methods should be wisely used. For data-sharing method, all processors (PEs) assigned to a job should be used whenever possible, through the use of the most appropriate data distribution method. For work-sharing method, the statements of the program have to be distributed among the computer's PEs with the goal of executing them in parallel. Based on those concepts, several subroutines from EPIC hydrocode will be coded and tested in next section.

Results and Discussions

A subroutine called SOLID which computes the deviator and shear stresses for solid materials is converted to a massively parallel program. We first use the MasPAr MP-1 massively parallel computer having 4096 PEs to solve this problem. The CPU time and speedup vs. number of elements for sequential on a single PE and Parallel on multiple PEs are shown in Fig. 3. The CPU time is almost kept constant for parallel computing when the number of elements increase. It expects the constant CPU time will last until the number of elements reach the number of PEs (i.e., 4096 in this machine). The speedup chart shows that a linear relation between the speedup and number of elements. Based on this chart, the maximum speedup can reach about 400 as the number of element increases to 4096. The results shows that tremendous speedup can be achieved for this subroutine.

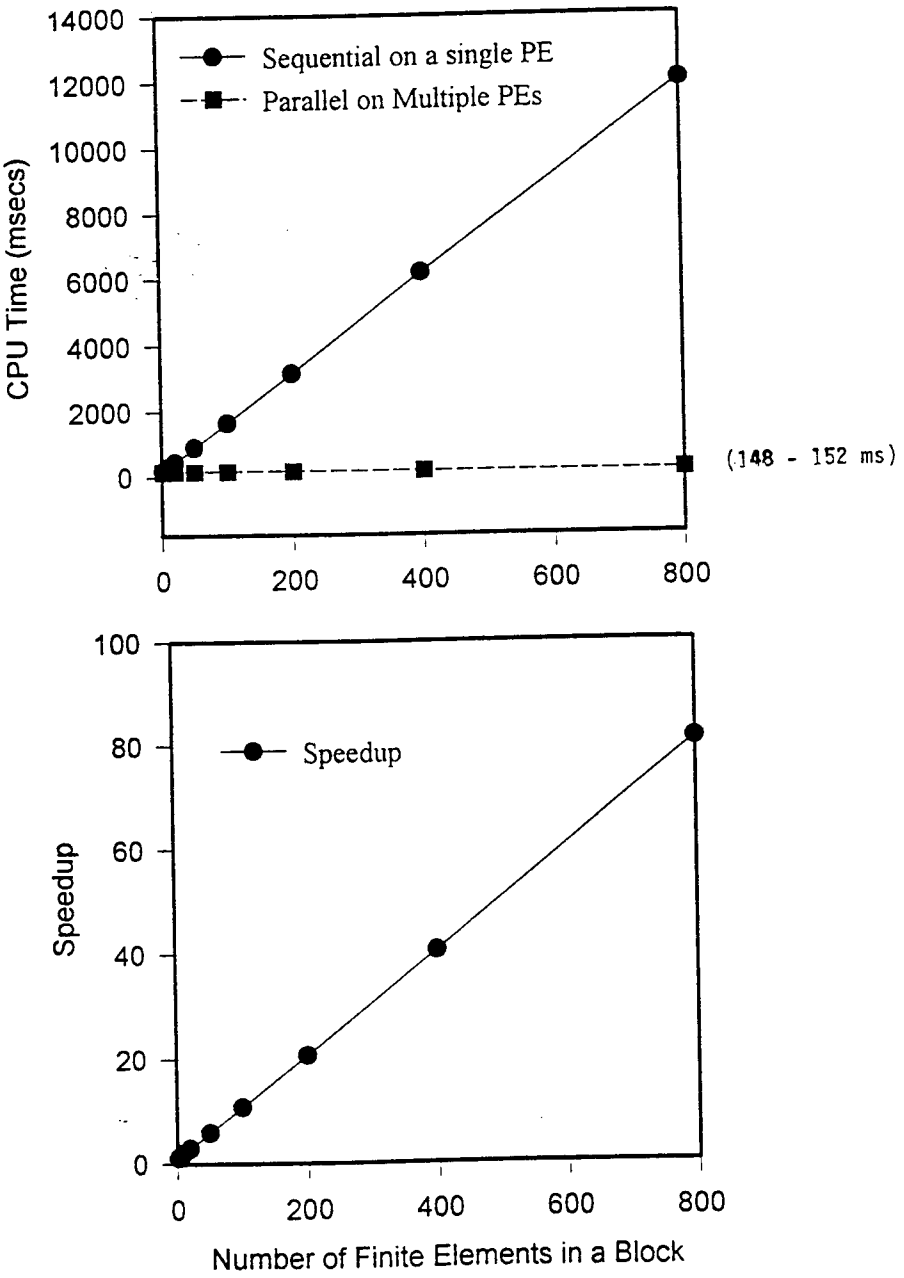


Fig. 3. Sequential and Parallel Versions of SOLID Subroutine Running on MasPar MP-1 Massively Parallel Computer (4096 PEs)

The same subroutine is also coded in Cray T3D massively parallel computer. Since the Cray T3D will not be installed until late this year or early next year, we will use Cray T3D emulator in Cray YMP supercomputer to test this program. The emulator can not tell the real CPU time to run the program in Cray T3D, but it can tell the operation references for each or all variables distributed in each PEs. The results for operation references distributed in one PE and eight PEs are shown in Figs. 4 and 5. Fig. 4 shows that a total of 20215 operation references are required in a single PE process. Fig. 5 shown that the operations are distributed in 8 PEs. Because the operations can not be 100% parallelized, the operation references is not uniformly distributed in each PEs. The maximum operation references of 2982 is distributed in PE0. Even the distribution is not uniform, it is very close. It indicates that this subroutine can be highly parallelized. When 32 PEs is used, the maximum operation references of 1143 is distributed in PE0. The estimated speedup is about 6.8 for 8 PEs and 17.7 for 32 PEs.

Subroutine VOLUME, which computes volumes, volumetric strains and strain rates, is also selected to be converted to a massively parallel program in Cray T3D. The results are shown in Figs. 6 and 7. Fig. 6 shows that 5528 operation references are required for single PE. Fig. 7 shows that the operation references are distributed over 8 PEs. The maximum operation references of 712 is distributed in PE0. The estimated speedup is 7.8 and 15 for 8 PEs and 16 PEs over single PE. The results show that subroutine VOLUME is more suitable for parallelism than SOLID.

For a typical three dimensional penetration example as shown in Fig. 8, subroutine VOLUME takes about 39% of total CPU time and SOLID takes 8.6%. Two subroutines already take about 48% of the total CPU time for this example. This is the reason why we selected them for this initial study. It means that the entire code can have a speedup of about 1.8 if VOLUME and SOLID subroutines are converted to have a speedup of 10. Even these two subroutine have the most impact on the performance of the entire code, the rest of the subroutines have to be converted to massively parallel code also in order to achieve a higher performance. We found that subroutine

ELOOP, which includes subroutines VOLUME and SOLID, takes about 76% of the total CPU time. Subroutine ELOOP is suitable for parallelism. Other subroutines, which solve the slide interface problem, may not be highly parallelized. However, certain speedup can still be achieved through careful massively parallel programming.

Conclusions and Future Work

Two key subroutines have been converted to massive parallel versions. The speedup over the single PE computer is very impressive. For the future work on converting EPIC hydrocode entirely to massively parallel version, we suggest to convert the subroutines ELOOP and NLOOP first. If those two subroutines can speedup the entire EPIC by 3 times or more, it is worth to invest the effort to convert the entire code.

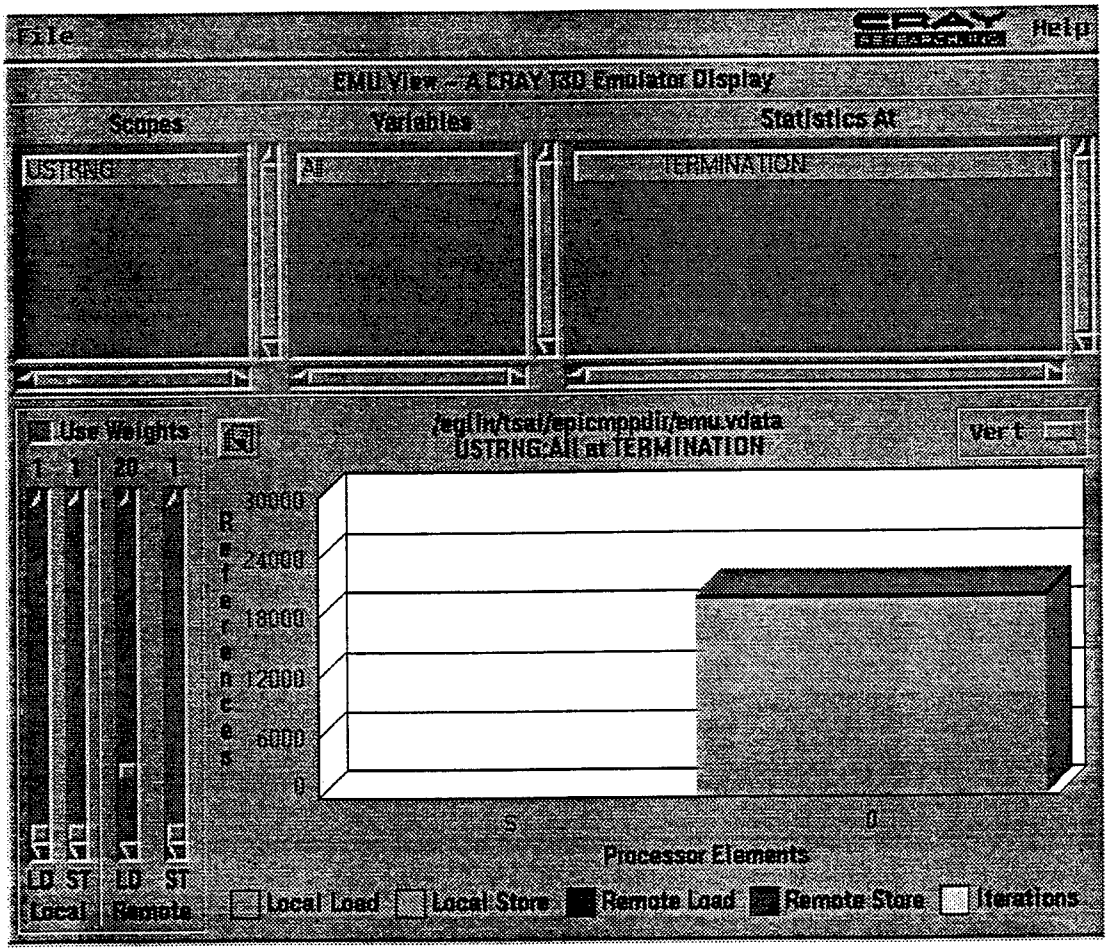


Fig. 4. Operation References of Single PE for Subroutine SOLID

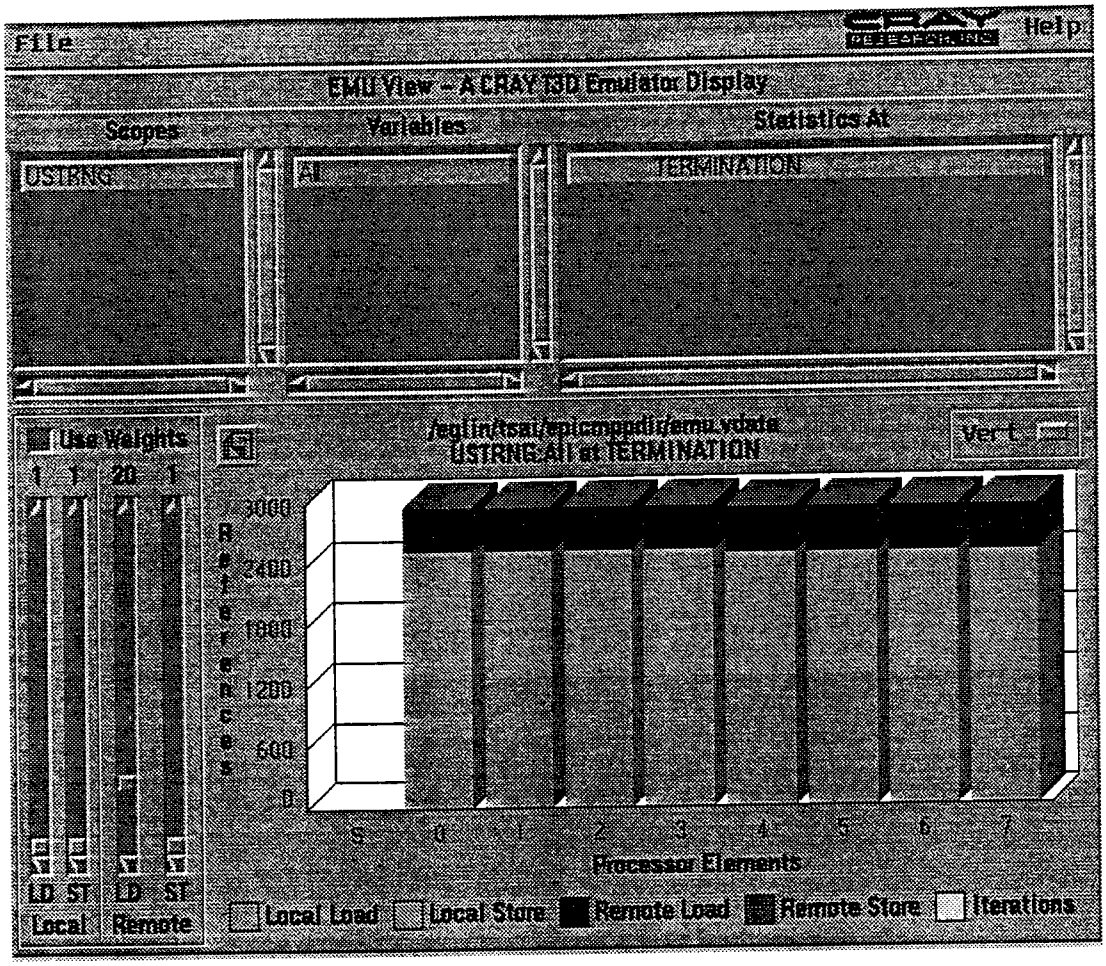


Fig. 5. Operation References of 8 PEs for Subroutine SOLID

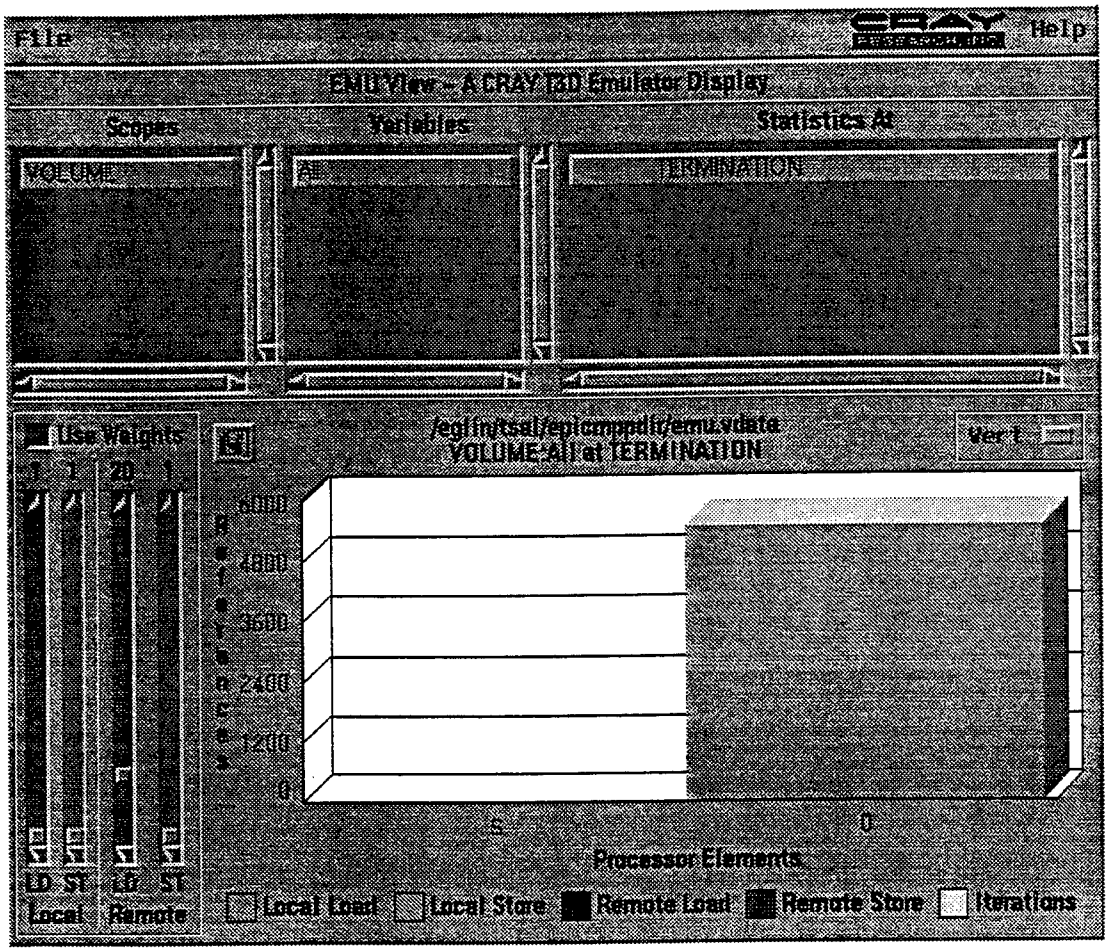


Fig. 6. Operation References of single PE for Subroutine VOLUME

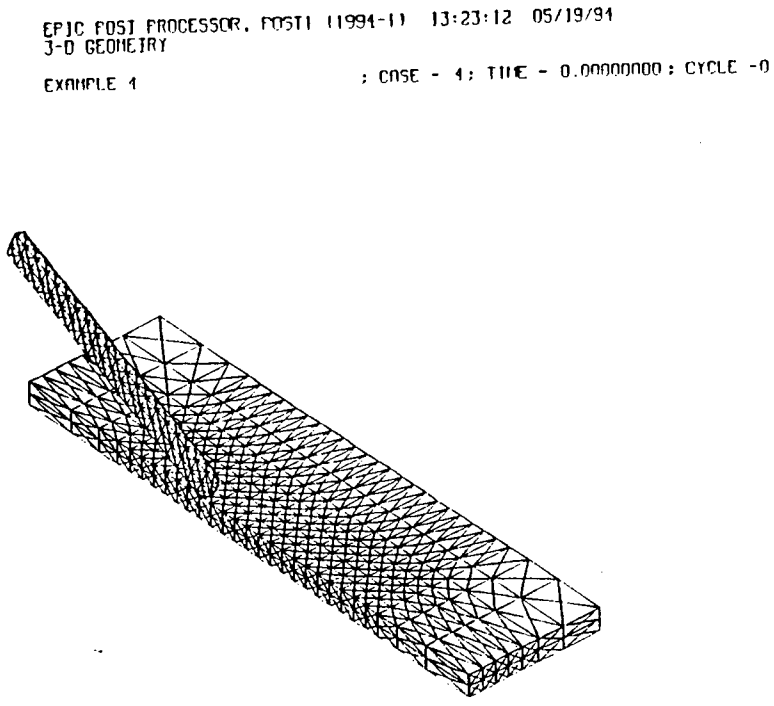


Fig. 8. A Typical Three Dimensional Penetration Problem

STRESS ANALYSIS OF THE V-NOTCH (IOSIPESCU) SHEAR TEST FOR COMPOSITE MATERIALS

James M. Whitney
Professor
Department of Civil & Environmental Engineering and Engineering Mechanics

University of Dayton
300 College Park
Dayton, OH 45469-0240

Final Report for:
Summer Faculty Program
Wright Laboratory

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

and

Wright Laboratory

August 1994

STRESS ANALYSIS OF THE V-NOTCH (IOSIPESCU) SHEAR TEST FOR COMPOSITE MATERIALS

James M. Whitney
Professor
Department of Civil & Environmental Engineering and Engineering Mechanics
University of Dayton

Abstract

Use of higher order beam theories is investigated in conjunction with stress analysis of the V-Notch Shear Test for composite materials. Higher order displacements are utilized which lead to a third order polynomial in the inplane normal stress, σ_x, relative to the thickness coordinate. Through-the-thickness shear stress, τ_{xy}, and normal stress, σ_y, are obtained by integrating the equations of classical theory of elasticity in conjunction with σ_x. Numerical results indicate that transverse normal strain, ε_y, can have a significant effect on the shear stress distribution at the notch cross-section. In addition, a singularity analysis at the notch tip is performed. No singularity was obtained for isotropic materials. However, singularities were obtained in conjunction with 0° unidirectional composites and 0°/90° laminates utilizing state-of-the-art graphite/epoxy material properties.

STRESS ANALYSIS OF THE V-NOTCH (IOSIPESCU) SHEAR TEST FOR COMPOSITE MATERIALS

James M. Whitney

Introduction

The V-Notched beam (Iosipescu) shear test has drawn considerable interest by the composite materials community. Both interlaminar and inplane shear properties are to be determined by this method.

The genesis of this test method can be found in a paper published by Iosipescu [1]. In this paper he proposed the test specimen shown in Figure 1 for determining the shear properties of metals. Iosipescu argued that the presence of a 90° v-notch in the presence of a cross-section with zero moment would produce a state of pure shear which is uniformly distributed across the width, w, between the notches. Because the sides of the notches are parallel to the normal stress direction, Iosipescu also concluded that no stress concentration, at least for isotropic materials, existed at the bottom of the notch.

In 1978 Slepetz, Zagaeski, and Novello [2] modified the loading scheme introduced by Iosipescu and applied the specimen to composite materials. The modified loading scheme was referred to as the asymmetrical four-point bending (AFPB) test. Another specimen configuration involving a circular disk with cutouts which resulted in a small test region was introduced by Arcan, Hashin, and Voloshin [3]. Although this specimen was designed primarily for the purpose of introducing a general case of biaxial loading, pure shear was included in a manner similar to Iosipescu's original configuration (i.e. a zero bending moment was used to produce pure shear across the width between two notches).

Considerable work has been done by Adams and Walrath [4-7] in developing the Iosipescu test for composite materials. Results of a finite element analysis are contained in Reference 6. A number of additional finite element solutions can be found in the literature [8-14]. Three major conclusions can be drawn from these various studies:

1. The state-of-stress between the two notches appears to be pure shear.

2. The shear stress distribution between the notches is uniform for isotropic materials.

3. A shear stress concentration occurs at the notch tip for orthotropic materials. This stress concentration can be alleviated by using a small radius at the notch tip.

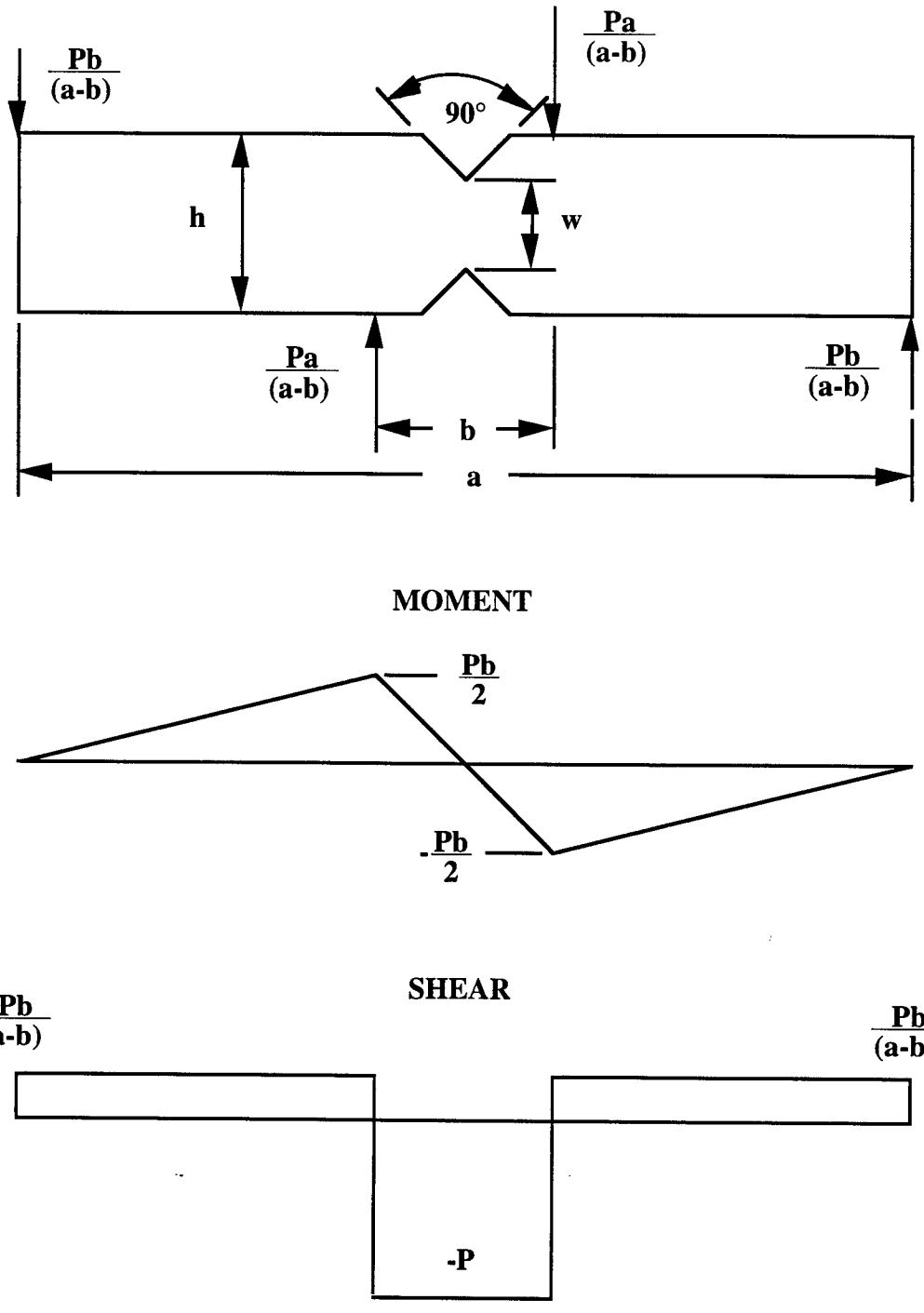


Figure 1. Specimen Geometry for Iosipescu shear test.

Despite the considerable number of research studies accomplished to date in conjunction with the Iosipescu test method, there are still a number of fundamental issues that need to be addressed, including the three conclusions mentioned above. Part of the problem stems from the fact that all of the published analytical results currently available appear to be from finite elements.

Recently a strength-of-materials analysis has been performed in conjunction with the Iospescu shear specimen [15]. In this analysis the inplane normal stress, σ_x, is assumed to be linear through-the-thickness in the usual manner. The through-the-thickness shear stress, $\tau_{xy,}$, and the transverse normal stress, σ_y, are determined by integrating the equations of equilibrium from classical theory of elasticity. The following conclusions were drawn from the strength-of-materials model:

1. The shear stress distribution is parabolic between the notches, not uniform as concluded by Iosipescu. This behavior is driven by the fact that the inplane normal stress σ_x vanishes on that cross-section inducing zero stresses at the notch tips.

2. Zero tractions along the notch boundaries induce a significant transverse normal stress component, σ_y, in the notch region of the specimen.

3. A sharp discontinuity in σ_y exists at the cross-section between the notches. This behavior is a direct consequence of a sharp discontinuity in the slope of the shear stress distribution (i.e. $\tau_{xy,x}$) along the length of the notch section.

4. All of the stress distributions are independent of material properties, i.e. the assumption that σ_x is linear in y in conjunction with the equations of equilibrium and the stress-free boundary conditions along the notch surface lead to stress distributions which are independent of material properties.

Obviously the strength-of-materials model yields considerably different results than those obtained from the finite element method. In regard to these discrepancies, we note that the geometry of the Iosipescu specimen presents a particular challenge to the finite element method. In particular, details of the strength-of-materials analysis show that stress distributions in the notch section of the specimen are very much influenced by the stress-free boundary conditions along the notch surface. Conventional finite element formulations, which are usually based on displacements, cannot directly satisfy traction-free boundary

conditions, as these are natural conditions to be satisfied in the limit. For results presented in the open literature authors do not address the extent to which their solution actually satisfies the traction-free conditions along the notch surface. In fact, stress-free boundary conditions are among the last of the elasticity relations to converge as a finite element mesh is refined. This is further complicated by the fact that the nature of the finite element method makes it difficult to precisely determine boundary stresses unless a very fine mesh is utilized in the boundary region. In addition, the sharp discontinuity in σ_y at the cross-section between the notches, as determined by the strength-of-materials analysis, suggests a very fine mesh is required in this region.

Difficulties with the strength-of-materials model hinge on two key assumptions:

1. Sufficient distance exists between the interior loads (dimension b in Fig. 1) to invoke St. Venant's principle (i.e. the loading can be modeled as shown in Fig. 2).

2. The stress component σ_x is linear with respect to y and, as a result, directly proportional to the bending moment.

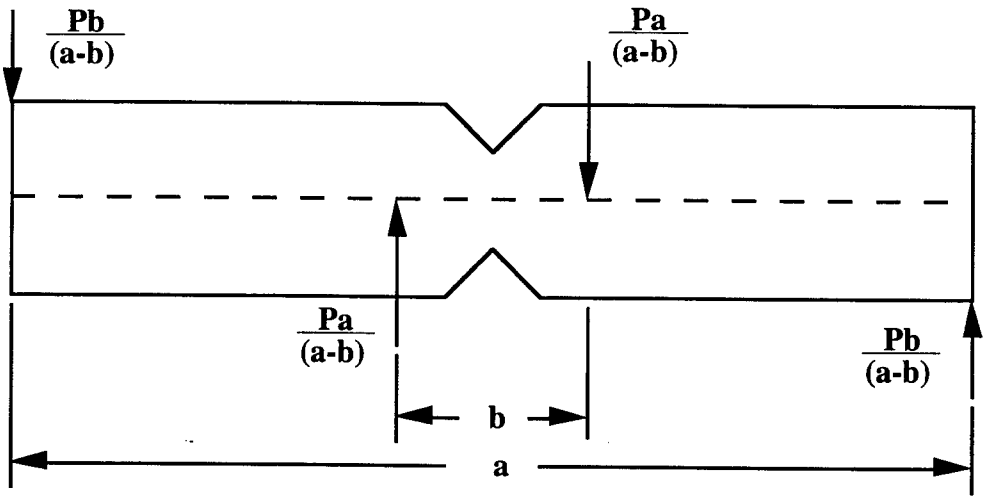


Figure 2. Idealized loading with St. Venant effect dissipated.

The first condition is a function of material properties. In particular, the dimension b required to satisfy the St. Venant effect is likely to be much larger for highly orthotropic materials than for isotropic materials. In

fact, the stress concentration near the notch tip as determined by the finite element method in conjunction with highly orthotropic materials may be due more to the St. Venant effect than to notch geometry. The second assumption takes on even more importance as it determines the shape of the distribution through-the-depth of the beam for both τ_{xy} and σ_y via integration of the equilibrium equations from classical theory of elasticity. As seen in the model, a linear distribution of σ_x leads to a second order polynomial in y for τ_{xy} and a third order polynomial in y for σ_y.

In the present work higher order beam theories are investigated in conjunction with stress analysis of the V-Notch shear test for composite materials. Higher order displacements are utilized which lead to a third order polynomial in the inplane normal stress, σ_x, relative to the thickness coordinate. Through-the-thickness shear stress, τ_{xy}, and normal stress, σ_y, are obtained by integrating the equations of classical theory of elasticity in conjunction with σ_x. In addition, a singularity analysis at the notch tip is performed.

Higher Order Beam Analysis

In performing a higher order beam analysis we must assume that there is sufficient distance between the interior loads so that St. Venant's principle applies. In such a case the loading is idealized as shown in Figure 2. Under this condition we can consider a simplified analysis using the free-body diagrams shown in Figure 3.

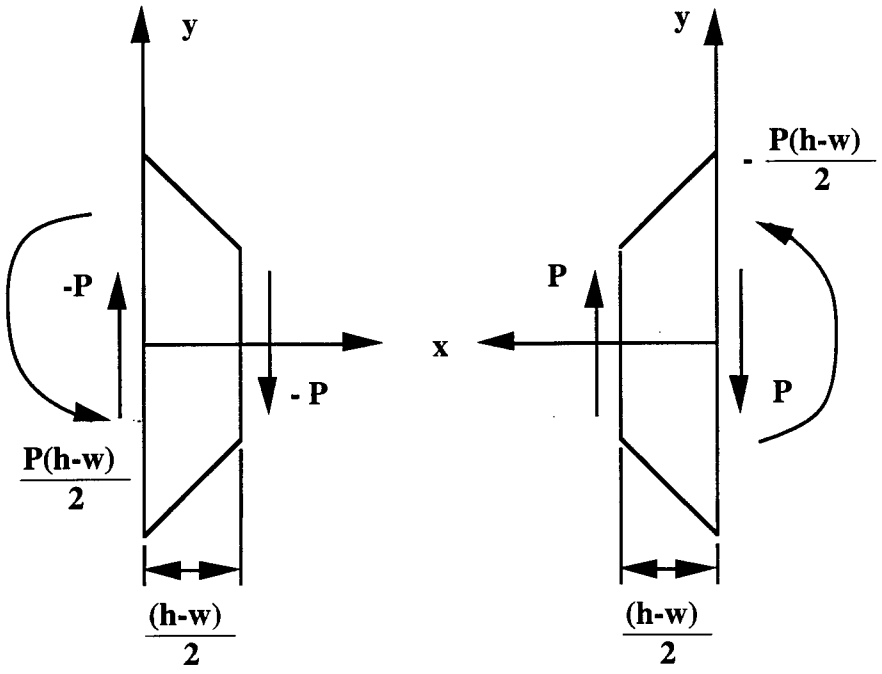


Figure 3. Free-body diagram for analytical model.

Using the free-body diagram on the left-hand side of Figure 3, we denote the beam depth by H(x). Thus,

$$H(x) = (h-2x) \tag{1}$$

In developing a higher order beam theory applicable to the Iosipescu shear test, we retain the usual assumptions associated with bending loads. In particular, the inplane normal stress σ_x is assumed to be anti-symmetric with respect to the through-the-thickness coordinate. For the present work we choose a displacement based theory which produces a third order polynomial for σ_x relative to the thickness coordinate, y. If a displacement based formulation is utilized, two approaches can be implemented in obtaining the desired distribution of σ_x. In the first case the effect of the thickness normal strain, ε_y, is neglected, which leads to the displacement field:

$$u = y\psi(x) + \frac{y^3}{3}\phi(x),\ v = v^o(x) \tag{2}$$

These displacements lead to the following normal stress, moment, and shear force resultants

$$\sigma_x = yE_1\left(\psi_{,x} + \frac{y^2}{3}\phi_{,x}\right),\ \tau_{xy} = G_{12}\left(\psi + v^0_{,x} + y^2\phi\right)$$

$$M = t\int_{-H/2}^{H/2}\sigma_x y dy = E_1\frac{tH^3}{12}\left(\psi_{,x} + \frac{H^2}{20}\phi_{,x}\right) \tag{3}$$

$$Q = t\int_{-H/2}^{H/2}\tau_{xy}dy = G_{12}tH\left(\psi + v^o_{,x} + \frac{H^2}{12}\phi\right)$$

where a comma denotes partial differentiation, t denotes the beam thickness in the z direction, E_1 is the beam modulus in the x-direction, and G_{12} denotes the shear modulus relative to the x-y plane.

If we include the effect of the transverse normal strain, ε_y, the displacement in eq. (2) takes the form

$$u = y\psi_x(x) + \frac{y^3}{3}\phi_x(x),\ v = v^0(x) + \frac{y^2}{2}\psi_y(x) + \frac{y^4}{4}\phi_y(x) \tag{4}$$

which leads to the following normal stress, moment, and shear force resultant

$$\sigma_x = y\left(Q_{11}\psi_{x,x} + Q_{12}\psi_y\right) + y^3\left(\frac{Q_{11}}{3}\phi_{x,x} + Q_{12}\phi_y\right)$$

$$M = Q_{11}\frac{tH^3}{12}\left[\psi_{x,x} + \frac{Q_{12}}{Q_{11}}\psi_y + \frac{H^2}{12}\left(\phi_{x,x} + \frac{Q_{12}}{Q_{11}}\phi_y\right)\right] \tag{5}$$

$$Q = G_{12}ht\left[\psi_x + v_{,x}^0 + \frac{H^2}{12}\left(\phi_x + \frac{\psi_{y,x}}{2}\right) + \frac{H^4}{320}\phi_{y,x}\right]$$

where Q_{ij} are the reduced stiffnesses for plane stress.

Equation (4) contains all of the polynomial terms in the y coordinate which contribute to both a third order inplane stress, σ_x, and a fourth order transverse shear stress, τ_{xy}. Integration of the first of eqs. (5) in conjunction with the first equilibrium equation from classical theory of elasticity also produces a fourth order polynomial in y for the transverse shear stress, τ_{xy}. However, the expression for τ_{xy} obtained from assumed kinematic relations does not satisfy appropriate boundary conditions. This can only be accomplished by integrating the first equilibrium equation from classical theory of elasticity with the result

$$\tau_{xy} = -\int_{-H/2}^{y}\sigma_{x,x}d\zeta + \tau_{xy}(x,-H/2) \tag{6}$$

The second term in this relationship represents the value of τ_{xy} at the top of the cross-section, which in general does not vanish. Stresses relative to the x-y coordinate system along the free boundary are shown in Figure 4. These results along with a similar consideration of the lower notch boundary at y = - H(x)/2 lead to the conditions

$$\sigma_y(x,-H/2) = \sigma_x(x,-H/2),\ \tau_{xy}(x,-H/2) = \sigma_x(x,-H/2)$$

$$\sigma_y(x,H/2) = \sigma_x(x,H/2),\ \tau_{xy}(x,H/2) = -\sigma_x(x,H/2) \tag{7}$$

A major effect produced by the 90° notch is the introduction of a significant σ_y stress component. In conventional beam problems this stress component is not present except in the vicinity of a vertical surface traction. The distribution of σ_y can now be determined by integrating the second equation of equilibrium from classical theory of elasticity in conjunction with the resulting expression for τ_{xy} as determined from

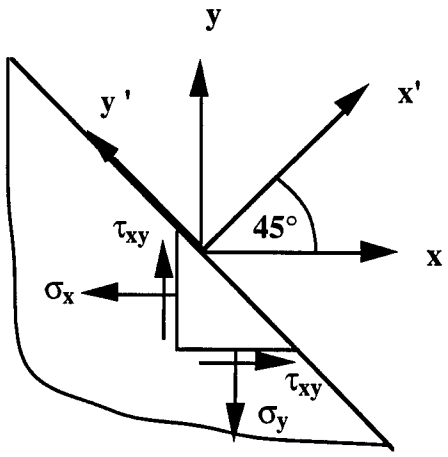


Figure 4. Stresses in x-y coordinate along free-boundary.

eq. (6), i.e.

$$\sigma_y = -\int_{-H/2}^{y} \tau_{xy,x} d\zeta + \sigma_y(x,-H/2) \tag{9}$$

It should be noted that eq. (9) will yield a third order polynomial in y when used in conjunction with axial strain effects only, while a fifth order polynomial will result when transverse strain effects are added. If σ_y is determined from kinematic assumptions, the resulting function will be of the same order in y as the corresponding function for σ_x.

For either kinematic relations, eq. (2) or eq. (4), two variables can be eliminated from the analysis by using M and Q as determined from the shear and moment diagrams, i.e.

$$M = \frac{P}{2}(h - w - 2x), \quad Q = -P \tag{10}$$

The remaining equilibrium equations in terms of displacement variables are determined from the principle of minimum potential energy. For the case of axial strain only, this results in a single non-homogeneous differential equation in the variable ϕ. When transverse strain is also included, three non-homogeneous, coupled, differential equations in the variables ϕ_x, ψ_y, and ϕ_y are obtained. In each case the differential equations can be solved in a straight forward manner. For the sake of conserving space the governing

equations and corresponding solutions are not shown. The following boundary conditions are utilized in the problem:

1. For the case of axial strain only

$$\text{At } x=0,\ \frac{(h-w)}{2}:\ \phi_{,x}=0 \tag{11}$$

2. For the case of both axial and transverse strain

$$\text{At } x=0:\ \phi_x=\phi_{x,x}=\phi_y=0$$

$$\text{At } x=\frac{(h-w)}{2}:\ \phi_{x,x}=\psi_y=\phi_y=0 \tag{12}$$

Conditions at the notch cross-section lead to vanishing of σ_x for both cases under consideration and continuity of v for the second case. Symmetry (or anti-symmetry) as dictated by Figs. 2 and 3 assure continuity of u and τ_{xy} at the notch cross-section. The boundary conditions at the beginning of the notch region (x = 0 in Fig. 3) leads to linear distributions with respect to y for σ_x in both cases and for u in the second case.

Numerical Results

We now consider a unidirectional composite with the fibers oriented along the axis of the beam [0°] composite. The following material properties are assumed:

$$\frac{E_1}{E_2}=15.4,\ \frac{G_{12}}{E_2}=0.615,\ \nu_{12}=0.3 \tag{13}$$

where E_2 is the modulus in the y-direction. Here ν_{12} denotes the major Poisson's ratio as determined from a uniaxial tensile test in the x-direction while measuring contraction in the y-direction. The properties in eq. (13) are typical of state-of-the-art graphite/epoxy unidirectional materials. When transverse strain effects are included in the kinematic relations, eq. (4), the reduced stiffnesses can be determined from the relationships

$$Q_{11}=\frac{E_1}{\left(1-\nu_{12}^2E_2/E_1\right)},\ Q_{12}=\frac{\nu_{12}E_2}{\left(1-\nu_{12}^2E_2/E_1\right)} \tag{14}$$

A comparison of shear stress distributions at the notch-tip for the two kinematic assumptions, eqs. (2) and (4), and classical beam theory are shown in Fig. 5. The notch depth is given by w/h = 0.5 and shear stress values are normalized by the average shear stress at the notch cross-section, i.e.

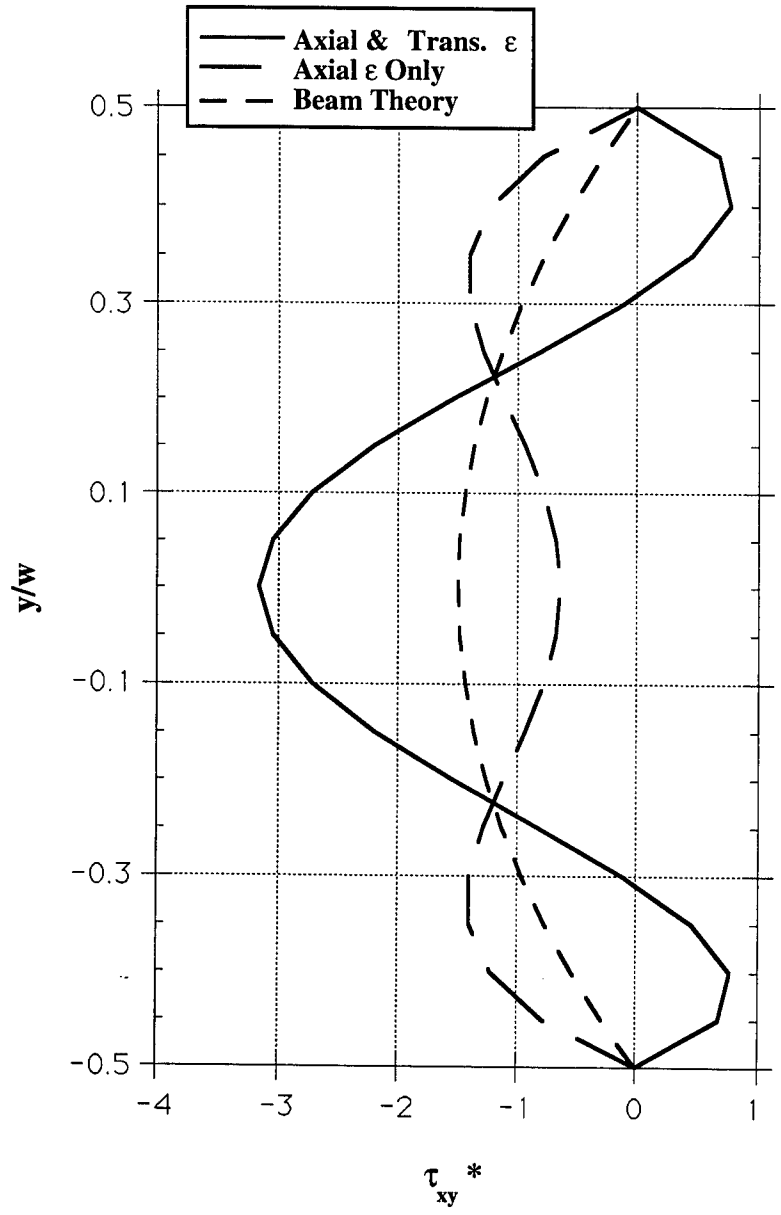


Figure 5. Shear stress distribution at notch cross-section, [0°] composite, w/h = 0.5.

$$\tau_{xy}^{*} = \frac{wt\tau_{xy}}{P} \tag{15}$$

The shear stress distribution is determined from equilibrium consideration, i.e. eq. (6). As previously noted the beam solution yields a classical parabolic shear stress distribution which is independent of material properties.

Results in Fig. 5 indicate a large effect of transverse normal strain as well as a significant departure from elementary beam theory. In fact these results indicate some instability in the shear stress distribution and leaves serious doubt that any of the three cases represents an accurate solution.

As a second example we consider the same unidirectional composite with the fiber oriented perpendicular to the beam axis, [90°] composite. In this case the material properties relative to the x-y coordinates are as follows:

$$\frac{E_1}{E_2} = 0.065, \quad \frac{G_{12}}{E_2} = 0.04, \quad \nu_{12} = 0.0195 \tag{16}$$

A comparison of models for the shear stress distribution at the notch cross-section is shown in Fig. 6 for w/h = 0.5. As in the case of the [0°] composite considerable departure from elementary beam theory is displayed by the higher order theories. However, in this case similar results are obtained with and without transverse strain included. Thus, the solution appears to be more stable than for the case of the [0°] unidirectional composite.

Singularity Analysis

The unstable nature of the solution displayed in Fig. 5 for the [0°] composite suggests the possibility of a singularity at the notch tip. In the singularity analysis we consider the right-side of the notch section, as shown in Fig. 3, and focus on the region near the notch tip. For convenience we choose the coordinate system shown in Fig. 7 with the y axis running along the length of the beam and the x axis oriented in the thickness direction. For a 90° notch the angle α is obviously 45°.

For the V-Notch specimen, the beam will be under plane stress conditions. A stress function is chosen in the form

$$\Phi = \frac{A_1(Z_1^{\delta+2} + \overline{Z}_1^{\delta+2}) + A_2 i(Z_1^{\delta+2} - \overline{Z}_1^{\delta+2}) + A_3(Z_2^{\delta+2} + \overline{Z}_2^{\delta+2}) + A_4 i(Z_2^{\delta+2} - \overline{Z}_2^{\delta+2})}{(\delta+1)(\delta+2)} \tag{17}$$

where Z_i are functions of the complex variables $x + i\mu y$ and $\overline{Z}_i$ are complex conjugates $x - i\mu y$. In

general μ is complex, however, for most engineering materials it is real and of the form

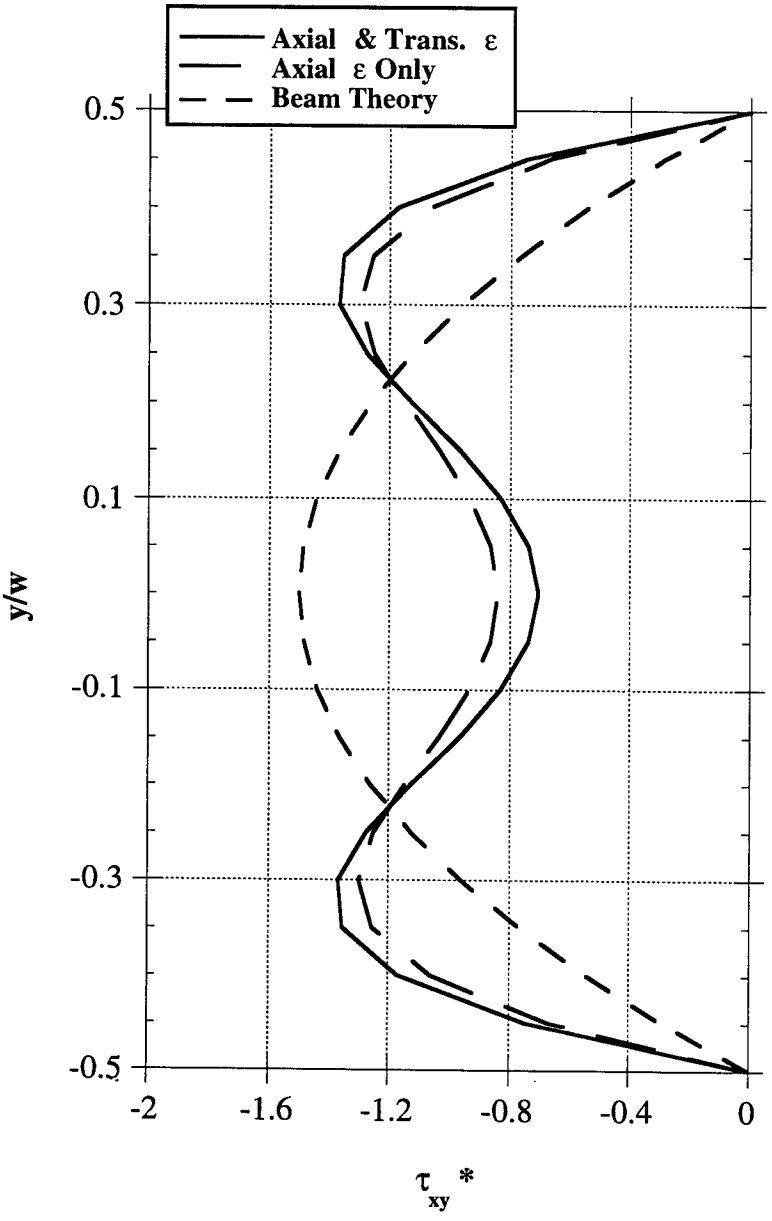


Figure 6. Shear stress distribution at notch cross-section, [90°] composite, w/h = 0.5.

$$\mu = \sqrt{\frac{1}{2}\left[\left(\frac{E_1}{G_{12}} - 2\nu_{12}\right) \pm \sqrt{\left(\frac{E_1}{G_{12}} - 2\nu_{12}\right)^2 - 4\frac{E_1}{E_2}}\right]} \tag{18}$$

Equation (18) is applicable to any composite laminate in which the effective inplane properties are orthotropic. In most engineering applications this would imply symmetric, balanced angle-ply laminates. Balanced in this context requires that every $+\theta$ (fiber orientation relative to the coordinate axis along the length of the beam) layer must be balanced by a $-\theta$ layer with $\theta \neq 0°$ or 90°.

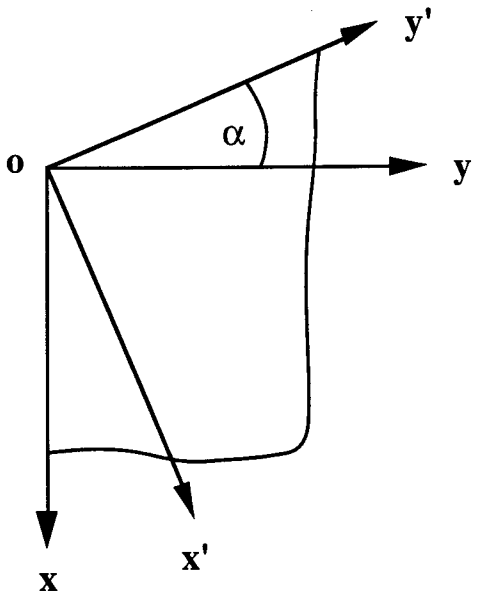


Figure 7. Nomenclature for singularity analysis ($\alpha = 45°$ for a 90° notch).

The stress function Φ yields the following stresses:

$$\sigma_x = -A_1\mu_1^2(Z_1^\delta + \overline{Z}_1^\delta) - A_2\mu_1^2 i(Z_1^\delta - \overline{Z}_1^\delta) - A_3\mu_2^2(Z_2^\delta + \overline{Z}_2^\delta) - A_4\mu_2^2 i(Z_2^\delta - \overline{Z}_2^\delta)$$

$$\sigma_y = A_1(Z_1^\delta + \overline{Z}_1^\delta) + A_2 i(Z_1^\delta - \overline{Z}_1^\delta) + A_3(Z_2^\delta + \overline{Z}_2^\delta) + A_4 i(Z_2^\delta - \overline{Z}_2^\delta) \tag{19}$$

$$\tau_{xy} = -A_1\mu_1 i(Z_{1i}^\delta - \overline{Z}_1^\delta) + A_2\mu_1(Z_1^\delta + \overline{Z}_1^\delta) - A_3\mu_2 i(Z_2^\delta - \overline{Z}_2^\delta) + A_4\mu_2(Z_2^\delta + \overline{Z}_2^\delta)$$

The displacement in the x-direction is given by

$$u = -\frac{\left(\mu_1^2 - \nu_{12}\right)}{E_1(\delta+1)}\left[A_1\left(Z_1^{\delta+1} + \overline{Z}_1^{\delta+1}\right) + A_2 i\left(Z_1^{\delta+1} - \overline{Z}_1^{\delta+1}\right)\right]$$

$$-\frac{\left(\mu_2^2 - \nu_{12}\right)}{E_1(\delta+1)}\left[A_3\left(Z_2^{\delta+1} + \overline{Z}_2^{\delta+1}\right) + A_4 i\left(Z_2^{\delta+1} - \overline{Z}_2^{\delta+1}\right)\right] \tag{20}$$

The coefficients A_i are determined from the following boundary conditions:

$$\text{At } y = 0,\ u = \sigma_y = 0\ ;\ \text{At } x = -y,\ y > 0,\ \sigma_x' = \tau_{xy}' = 0 \tag{21}$$

Using stress transformation relations, the second of eqs. (21) become

$$\text{At } x = -y,\ y > 0,\ \sigma_x - \sigma_y = 0;\ \sigma_y + \tau_{xy} = 0 \tag{22}$$

Combining eqs. (19) - (22), we obtain 4 homogeneous, algebraic equations in the variable δ. In order to avoid the trivial solution, δ is chosen such that the determinant of the coefficient matrix vanishes. This results in the following transcendental equation:

$$\left(1+\mu_1^2\right)^{\delta/2} \sin\ \delta\theta_1\left(\sin\ \delta\theta_2 - \mu_2 \cos\ \delta\theta_2\right) - \left(1+\mu_2^2\right)^{\delta/2} \sin\ \delta\theta_2\left(\sin\ \delta\theta_1 - \mu_1 \cos\ \delta\theta_1\right) = 0 \tag{23}$$

where

$$\theta_1 = \pi - \arctan\ \mu_1,\ \ \theta_2 = \pi - \arctan\ \mu_2$$

There are an infinite number of values of δ, either real or complex, which will satisfy this relationship. For the purpose of identifying singularities at the notch tip, we are interested in $-1 < \delta < 0$. For $\delta \leq -1$, unbounded displacements will occur.

For the case of isotropic materials, the form of the solution changes. The governing equations in terms of a stress function have been presented by Williams [16]. Singularities are determined in analogous fashion to the procedure presented here for orthotropic materials.

The first five eigenvalues to eq. (23) are shown in Table 1 for the [0°] and [90°] unidirectional composites previously discussed. A bidirectional, $[0°/90°]_S$, laminate is also shown in Table along with an isotropic

material with a Poisson's ratio $\nu = 0.3$. Bidirectional laminate properties are as follows:

$$\frac{E_1}{E_2} = 1, \quad \frac{G_{12}}{E_2} = 0.0748, \quad \nu_{12} = .0366 \tag{24}$$

Eigenvalues in Table 1 are determined from *Mathematica* which employs the Newton-Rapson method. Real roots were cross-checked using an iteration scheme. It should be noted that 0 is a root to eq. (23).

TABLE 1. First Five Non-Zero Eigenvalues to Eq. (23)

[0°]	[90]	$[0°/90°]_S$	Isotropic
-0.2466	0.1435	-0.0586	0.2094 ± 0.5832 i
1.1161 ± 0.1665 i	1.6060 ± 0.1981 i	1.2233	1.5662 ± 0.8007 i
2.3549 ± 0.0495 i	2.8685	1.5361	2.9156 ± 0.9418 i
3.2878	3.4194	2.2731	4.2603 ± 1.0470 i
3.9133	4.0689	3.3208 ± 0.1104 i	5.6020 ± 1.1310 i

Results in Table 1 reveal a singularity in conjunction with both the [0°] and $[0°/90°]_S$ composites. We note, however, that the singularity associated with the bidirectional laminate is very weak.

Summary and Conclusions

A higher order beam analysis of the V-Notch shear test for composite materials has been presented in which the axial normal stress is represented by a third order polynomial in the thickness coordinate. Two kinematic models have been investigated. One considers only axial strain, while the second model considers the effect of both axial and transverse normal strains. Transverse shear and normal stresses are determined by integrating the equations of equilibrium from classical theory of elasticity in conjunction with the axial normal stress as determined from the higher order beam models.

Numerical results for a [0°] graphite/epoxy composite show a large difference in shear stress distribution at the notch cross-section between the two kinematic models. In addition both of these models show a large discrepancy compared to elementary beam theory. Analysis of the stress distribution in the vicinity of the notch tip using classical theory of elasticity reveals a singularity. Thus, the higher order beam models are incapable of representing the correct shear stress distribution because of the absence of representation of the singularity.

Numerical results for a [90°] graphite/epoxy composite reveal good correlation between the two kinematic models for the distribution of the shear stress at the notch cross-section. As in the case of the [0°] composite a large discrepancy with elementary beam theory was observed. A singularity analysis revealed no singularity at the notch tip for the [90°] graphite/epoxy composite.

A singularity analysis was also performed in conjunction with a $[0°/90°]_S$ graphite/epoxy laminate and an isotropic material. No singularity was found for the isotropic material, while a weak singularity was obtained in conjunction with the bidirectional laminate.

Future work should focus on cases where singularities occur at the notch tip. In such cases the correct singulariity should be built into the higher order beam model. For the cases where no singularities are observed, a higher beam theory in which the axial stress is represented by a fifth order polynomial should be investigated and compared to the higher order models presented here.

One important point should be noted when considering finite element results. Physics of the V-Notch specimen dictate that in an elastic analysis either the shear stress at the notch tip must vanish or it is singular. Neither of these conditions appear to be met in solutions presented in the literature.

References

1. N. Iosipescu, "New Accurate Procedure for Single Shear Testing of Metals, "*Journal of Materials*, Vol. 2, No. 3, September, 1967, pp. 537-566.

2. J. M. Slepetz, T. F. Zagaeski, and R. F. Novello, "In-Plane Shear Test for Composite Materials," Report No. AMMRC TR 78-30, Army Materials and Mechanics Research Center, Watertown, MA , July, 1978.

3. M. Arcan, Z. Hashin, and A. Voloshin, "A Method to Produce Uniform Plane-stress States with Application to Fiber-reinforced Materials," *Experimental Mechanics*, Vol. 18, No. 4, December, 1978, pp. 141-146.

4. D. E. Walrath and D. F. Adams, "The Iosipescu Shear Test as Applied to Composite Materials," *Experimental Mechanics*, Vol. 23, No. 1, March, 1983, pp. 220-224.

5. D. F. Adams and D. E. Walrath, "Iosipescu Shear Properties of SMC Composite Materials," *Composite Materials: Testing and Design (Sixth Conference),* ASTM STP 787, I. M. Daniel, Editor, American Society for Testing and Materials, Philadelphia 1982, pp. 19-33.

6. D. F. Adams and D. E. Walrath, "Further Development of the Iosipescu Shear Test Method," *Experimental Mechanics*, Vol. 27, No. 2, June, 1987, pp. 113-119.

7. D. F. Adams and D. E. Walrath, "Current Status of the Iosipescu Shear Test Method," *Journal of Composite Materials*, Vol. 21, 1987, pp. 494-507.

8. J. L. Sullivan, B. G. Kao, and H. Van Oene, "Shear Properties and A Stress Analysis Obtained from Vinyl-ester Iosipescu Specimen," *Experimental Mechanics*, Vol. 24, No. 3, September, 1984, pp. 223-232.

9. M. Kumosa and D. Hull, "FEM Analysis of Mixed Mode Fracture in the Iosipescu Shear Test", *Proceedings of the Sixth International Conference on Composite Materials/Second European Conference on Composite Materials*, Vol. 3, Edited by F.L. Matthews, N. C. R. Buskell, J. M. Hodgkinson, and J. Morton, Elsevier Applied Science, London, 1987, pp. 3.243-3.253.

10. J. A. Barnes, M. Kumosa, and D. Hull, "Theoretical and Experimental Evaluation of the Iosipescu Shear Test," *Composite Science and Technology*, Vol. 28, No. 4, 1987, pp. 251-268.

11. M. J. Pindera, G. Choksi, J. S. Hidde, and C. T. Herakovich, "A Method for Accurate Shear Characterization of Unidirectional Composites," *Journal of Composite Materials*, Vol. 21, No. 12, December, 1987, pp. 1164-1184.

12. M. G. Abdallah and H. E. Gascoigne, "The Influence of Test Fixture Design on the Iosipescu Shear Test for Fiber Composite Materials," *Test Methods for Design Allowables for Fibrous Composites, 2nd Volume,* ASTM STP 1003, C. C. Chamis, Editor, American Society for Testing and Materials, Philadelphia, 1989, pp. 231-260.

13. H. T. Budiman, H. Ho, M. Y. Tsai, and J. Morton, "Effect of Specimen Length and Notch Geometries on the Performance of Notched Shear Specimens, "*Proceedings of the 33rd AIAA SDM Conference, Part 5 - Design Engineering/Design Optimization, Materials, Work-in-Progress*, AIAA, Washington, D.C., 1993, pp. 2905-2914.

14. H. Ho, M. Y. Tsai, J. Morton, and G. L. Farley, "Numerical Analysis of the Iosipescu Specimen for Composite Materials," *Composite Science and Technology*, Vol. 46, No. 2, 1993, pp. 115-128.

15. J. M. Whitney, "Some Observations Concerning the Iosipescu Shear Test for Composite Materials," *Proceedings of the 39th International SAMPE Symposium and Exhibition,* Vol. 9, Book 2, Society for the Advancement of Material and Process Engineering, Covina, CA, 1994, pp. 2923-2934.

16. M. L. Williams, "Stress Singularity Resulting From Various Boundary Conditions in Angular Corners of Plates in Extension," *Journal of Applied Mechanics*, Vol. 19, Sept. 1952, pp. 526-528.