ARMY RESEARCH LABORATORY

Volume I: Select Papers

ARL Summer Student Research Symposium

ARL-TM-2011 **August 2011**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1197

ARL-TM-2011 **August 2011**

Volume I: Select Papers

ARL Summer Student Research Symposium

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
August 2011	Final	

4. TITLE AND SUBTITLE
Volume I: Select Papers

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
ARL Summer Student Research Symposium

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TM-2011

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The ARL Summer Student Research Symposium is an ARL Director's Award Program for all the students participating in various summer scholarship and contract activities across ARL. The goal of the program is to recognize and publicize exceptional achievements made by the students and their mentors in the support of Army science.

All college undergraduate and graduate students receiving research appointments and conducting summer studies at ARL are automatically enrolled in the symposium program. As an integral part of their summer study, all students are required to write a paper on their work which summarizes their major activity and its end product.

The program is conducted on two separate competitive levels: undergraduate and graduate. The format of the paper in both levels is the same. However, the evaluation will take into consideration the difference in the academic level of the students.

All students submitted their research paper for directorate review. Directorate judging panels selected one or two papers from each competition category for the laboratory-wide competition at the Summer Student Symposium on 11 August 2011.

Students selected by their directorate for competition participated in the one-day Summer Student Symposium on 11 August 2011. At the symposium the students presented their papers to the ARL Director and an ARL Fellows panel.

This volume of the Summer Student Symposium Proceedings contains the papers presented at the symposium.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Vallen Emery
Unclassified	Unclassified	Unclassified	UU	230	19b. TELEPHONE NUMBER (*Include area code*) (301) 394-3585

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

Director's Foreword **v**

Introduction **1**

Computational & Information Sciences Directorate (CISD) **3**

Examining a Terrorist Network Using Contingency Table Analysis....5

Optimizing Strategies for an Observation-nudging-based Four-Dimensional Data Assimilation Forecast Approach with WRF-ARW....21

Human Research & Engineering Directorate (HRED) **39**

Motion Simulation in the Environment for Auditory Research....41

Weighted Phase Lag Index (WPLI) as a Method for Identifying Task-Related Functional Networks in Electroencephalography (EEG) Recordings during a Shooting Task....53

Sensors and Electron Devices Directorate (SEDD) **69**

Growth and Transfer of Graphene for Device Fabrication....71

Integrated Thin-film Piezoelectric Traveling Wave Ultrasonic Motor....85

Survivability/Lethality Analysis Directorate (SLAD) **103**

Investigation of Integrating Three-Dimensional (3-D) Geometry into the Visual Anatomical Injury Descriptor (Visual AID) Using WebGL....105

Target Modeling for Ground Mobile Branch (GMB)....119

Vehicle Technology Directorate (VTD) **131**

Digital Image Correlation of Flapping Wings for Micro-Technologies....133

Development of a Lumped Element Circuit Model for Approximation of Dielectric Barrier Discharges....151

Weapons & Materials Research Directorate (WMRD) 167

Microstructural Investigation and Evaluation of Mechanical Properties in Friction Stir Welded Joints 169

Tunable, Highly Ordered TiO_2 Nanotube Arrays on Indium Tin Oxide Coated PET for Flexible Bio-sensitized Solar Cells 189

Tactical Means to Stow Super-Caliber Tailfins of a Developmental Flight-Controlled Mortar 207

Director's Foreword

The U.S. Army Research Laboratory (ARL) mission is to "Provide innovative science, technology, and analyses to enable full spectrum operations." As the Army's corporate laboratory, we provide the technological underpinnings critical to providing capabilities required by our current and future Soldiers.

Our nation is projected to experience a shortage of scientists and engineers. ARL recognizes the criticality of intellectual capital in generating capabilities for the Army. As the Army's corporate laboratory, addressing the projected shortfall is a key responsibility for us. We have, therefore, identified the nation's next generation of scientists and engineers as a key community of interest and have generated a robust educational outreach program to strengthen and support them. We have achieved many successes with this community. We believe that the breadth and depth of our outreach programs will have a significant positive effect on the participants, facilitating their journey toward becoming this Nation's next generation of scientists and engineers.

A fundamental component of our outreach program is to provide students research experiences at ARL. During the summer of 2011, we supported research experiences at ARL for over 100 undergraduate and graduate students. Each of these students writes a paper describing the results of the work they performed while at ARL. All of the papers were of high quality, but only a few could be presented at our student symposium. The abstracts for all papers prepared this summer are contained in this volume of the proceedings and they indicate that there were many excellent research projects with outstanding results. It is unfortunate that there was not enough time for us to have all of the papers presented. We would have enjoyed hearing them all.

We are very pleased to have hosted this outstanding group of students for the summer. It is our hope that they will continue their pursuit of technical degrees and will someday assist us in providing critical technologies for our Soldiers.

INTENTIONALLY LEFT BLANK.

Introduction

The ARL Summer Student Research Symposium is an ARL Director's Award Program for all the students participating in various summer scholarship and contract activities across ARL. The goal of the program is to recognize and publicize exceptional achievements made by the students and their mentors in the support of Army science.

All college undergraduate and graduate students receiving research appointments and conducting summer studies at ARL are automatically enrolled in the symposium program. As an integral part of their summer study, all students are expected to write a paper on their work which summarizes their major activity and its end product.

The program is conducted on two separate competitive levels: undergraduate and graduate. The format of the paper in both levels is the same. However, the evaluation will take into consideration the difference in the academic level of the students.

All students submitted their research paper for directorate review. Directorate judging panels selected one or two papers from each competition category for the laboratory-wide competition at the Summer Student Symposium on 11 August 2011.

Students selected by their directorate for competition participated in the one-day Summer Student Symposium on 11 August 2011. At the symposium, the students presented their papers to an audience of ARL scientists and engineers, including the ARL Director and an ARL Fellows panel.

This volume of the Summer Student Symposium Proceedings contains the papers presented at the symposium.

INTENTIONALLY LEFT BLANK.

Computational & Information Sciences Directorate (CISD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Examining a Terrorist Network Using Contingency Table Analysis

ALLISON MOORE
MENTOR: DR. ELIZABETH K. BOWMAN
COMPUTATIONAL AND INFORMATION SCIENCES DIRECTORATE
ABERDEEN PROVING GROUND, MARYLAND

Contents

List of Figures 7

List of Tables 7

Abstract 8

Acknowledgments 9

Student Bio 10

1. Introduction/Background 11

2. Experiment/Calculations 11

3. Results and Discussion 13

4. Summary and Conclusions 17

5. References 18

List of Figures

Figure 1. Initial logistic model. 15
Figure 2. Second logistic model 16
Figure 3. Proportional odds ratios. 16

List of Tables

Table 1. Original Ali Baba variable table. 12
Table 2. After categorization Ali Baba variable table. 12
Table 3. Truth vs. location. 13
Table 4. Truth vs. event. 14
Table 5. Location vs. event. 14

Abstract

The information revolution of the 21st century has changed the nature of war to focus on the area of network-centric warfare. While the number and strength of dark networks continue to increase, the U.S. Department of Defense seeks to identify, predict, and counteract possible terrorist threats. The intelligence community is undertaking the seemingly impossible task of using the information that is currently available through social networks and military reports. This report will examine the Ali Baba data set that was created in 2003 for the National Security Agency (NSA) by Mark Jaworoski and Steve Pavlak. The Ali Baba data set contains fictitious Word documents that have recorded the actions of a suspected terrorist network. This report will demonstrate the use of statistics in examining terrorist organizations. Specifically, it will determine if contingency table analysis using the R programming language can be used to analyze a terrorist network.

Acknowledgments

I wish to acknowledge the mentorship of Elizabeth Bowman.

Student Bio

I will be a senior this fall at Virginia Polytechnic Institute and State University in Blacksburg, Virginia. I am a double-major in Mathematics and Statistics, with a minor in Actuarial Science. This is my second year as a summer student at the U.S. Army Research Laboratory (ARL). After graduation, I plan on either attending graduate school to concentrate in applied statistics or becoming a mathematical statistician for the government. Prior to this summer, I had no experience working with the R programming language.

1. Introduction/Background

The information revolution of the 21st century has changed the nature of war (*2*). The areas of cyber-warfare and network centric warfare (*2, 4*) are an expanding focus of research within the United States military and the Department of Defense (DoD) (*7*). While the number and strength of networks continue to increase, the intelligence community is developing programs and techniques to identify, predict, and counteract possible terrorist threats (*3, 4*). The intelligence community is undertaking the seemingly impossible task of using the information that is currently available through social networks and military reports. Faced with an endless supply of data, this is a very difficult problem.

The purpose of this paper is to demonstrate the use of statistics in examining terrorist organizations. Specifically, we wish to determine if Contingency Table Analysis (CTA) using the R programming language can be used to analyze a dark network. While the statistical scope of this paper is limited, it is my desire to see an increase in the use of statistics in solving problems within the intelligence community. The data set that will be used for all subsequent analysis in this paper is entitled "Ali Baba." In 2003, Mark Jaworowski and Steve Pavlak originally created the data set for the National Security Agency (NSA) as a way to test visualization software since there were no reliable data sets available for evaluation. They sought to develop a data set that an analyst would be able to piece together and determine the intentions of a terrorist cell.

The Ali Baba data set contains fictitious word documents that have recorded the actions of a suspected terrorist network from April to September 2003. The documents include local communication intercepts and various intelligence reports (police, human intelligence (HUMINT), Foreign Broadcast Information System (FBIS), and detainee). There are a core group of messages (n=75) that contains the main activities of the cell. The remaining messages contain possibly interesting information from peripheral members. The scenario of the data set follows a suspected terrorist cell of radical Islamic fundamentalists that are centralized in England. The FBIS reports in southern Egypt from spring of 2003 describe an outbreak of cholera among large groups of schoolchildren. The radical Islamic world was blamed for the attack, and this data set reveals their desire to revenge the claim. It is the organization's plan to contaminate London's water supply by blowing up a water treatment facility (*9*).

2. Experiment/Calculations

The Ali Baba data set contains 609 entities with nine original variables, as shown in table 1. In order to perform statistical analysis, it is necessary to categorize the data. Table 2 holds the

categorized version of the data set that contains a total of five variables. This specific process of categorization is labor-intensive, as each line has to be assigned by hand due to variability in the data.

Table 1. Original Ali Baba variable table.

Variable Name	Variable Type	Description
Document	Character	Source of message and identification number
Date	Character	Date of original message
Associate1	Character	Author of message
Associate2	Character	Receiver of message or mentioned in message
Associate3	Character	Receiver of message or mentioned in message
Location	Character	Where the message originated
Event/Activity	Character	Summary of message contents
Organization	Character	Organization involvement of the author
Targeted Entity	Character	Target of interest

Table 2. After categorization Ali Baba variable table.

Variable Name	Variable Type	Description
Truth	Numeric	Villain level of Associate 1, 2, 3
Location	Character	Region of the message's origin
Event	Character	Categorization of event/activity
binlocal	Numeric	If Location = "SE" or "London" then binlocal = 1, else = 0
binTruth	Numeric	If Truth = 1 then binTruth = 1, else = 0

The biggest change between table 1 and table 2 is the switch of the "Associate" variables from character to the ordinal numeric "Truth" variable. This is done to denote the members through what we will call Core Villain Levels, which range from 1 to 5. A Ground Truth document created by Jaworoski and Pavlak (*8*) is used to rank the villains Black, Dark Blue, Medium Blue, Light Blue, and Unknown, or "1", "2", "3", "4", and "5," respectively. Two issues arise when categorizing the "Associate" variables. First, if an individual is ranked as a "5," we will say that they are not involved with the terrorist organization, and is, therefore, not a suspected threat. Second, if an individual was a possible member of two core villain levels, they were given the higher ranking. The levels provided by the Ground Truth document could also be determined by a social network graph like those created in Analyst Notebook.

The variable "Location" is categorized to represent the change from individual locations to regions (*11*). Additional entities for "Location" include "Unknown", "Outside" for locations outside of Britain, and "Britain" for unspecific messages. Several obvious errors were found while going through the "Location" variable of the original data set. These errors include London being spelled "Lodnon," or "Brigton" instead of Brighton. These mistakes are corrected to ensure more accurate results. The "Event/Activity" variable is also changed to "Event," which describes the change to categorized groups based on certain keywords found in the data set. For example, the "Brit" grouping is based on messages that mentioned Tony Blair, British

government or politics, and anti-British or anti-UK sentiment. This particular grouping is the most controversial, as it is determined particularly by me.

The final two variables in table 2 are binary variables called "binlocal" and "binTruth". In regression, binary variables also known as indicator variables take on the values "0" or "1" to indicate failure or success of a categorical effect (*1*). The binary results for "Location" are recorded as "1" when the villain sends a message from "SE" or "London," and recorded as "0" otherwise. Similarly, the "binTruth" is set as a "1" when the villain is a Black Villain, and recorded as "0" otherwise.

3. Results and Discussion

Now that all of the data is categorized, we are able to fully use CTA within the R programming language. CTA can be used to investigate the relationship between two or more variables (*1*).

First we can examine the Villain Truth Level vs. Location table, which is table 3. The primary step is to determine if the two variables are independent or not. We conduct a Pearson's Chi Square test (*1*) in R, and determine that $X^2 = 95.1836$, degrees of freedom = 30, and p-value = 1.053×10^{-8}. The degrees of freedom are determined by (# of rows – 1)*(# of columns – 1). Since the p-value is less than .05, we can reject the null hypothesis and conclude that Truth and Location are dependent.

$$X^2 = \sum \frac{(n_{ij} + \mu_{ij})^2}{\mu_{ij}} \quad (1)$$

Table 3. Truth vs. location.

	Brit.	**E**	**EM**	**Lon.**	**NE**	**NW**	**Out**	**SE**	**SW**	**Unk.**	**WM**	**Totals**
Black	*2*	*4*	*1*	*75*	*2*	*0*	*11*	*34*	*5*	*5*	*1*	140
Dark	*10*	*8*	*0*	*50*	*1*	*0*	*15*	*34*	*15*	*20*	*1*	154
Medium	*4*	*4*	*0*	*56*	*4*	*0*	*3*	*30*	*0*	*9*	*5*	115
Light	*1*	*2*	*0*	*40*	*5*	*3*	*6*	*48*	*10*	*2*	*0*	117
Totals	17	18	1	221	12	3	35	146	30	36	7	526

We can also calculate odds and odds ratios from table 4, using formula (*2*).

$$\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} \quad (2)$$

Looking at villains overall, we determine that they are 5.3277 times more likely to have sent their message from the Southeast or London than any other location. We will, thus, hypothesize that the terrorist cell is focused in the Southeastern region of Britain. This fact will be useful later when we are interested in creating a model. For all further analysis, we will classify "higher level villains" as Black or Dark Blue Villains, and "lower level" as Medium or Light

Blue. Again, looking at table 3, we can calculate that lower level villains are 4.07 times more likely to send messages from Northern Britain than higher level villains.

Second, we can examine the Villain Truth Level vs. Location table. The Pearson's Chi Square in table 4 is 155.7799 with 27 degrees of freedom, and a calculated p-value of less than $2.2x10^{-16}$. Therefore, we can reject the null hypothesis and conclude that the variables are dependent. Various odds and odds ratios can once again be calculated from the table. For example, upper level villains are 8.482 times more likely to be found discussing non-suspicious items, while Light Blue villains are over five times more likely to discuss casing than their counterparts.

Table 4. Truth vs. event.

	Alq	**Assoc**	**Bomb**	**Brit**	**Case**	**Islam**	**911**	**Not**	**Rec**	**Sus**	**Totals**
Black	*4*	*36*	*16*	*4*	*1*	*11*	*4*	*33*	*20*	*11*	140
Dark	*11*	*33*	*9*	*4*	*3*	*45*	*9*	*21*	*16*	*3*	154
Medium	*6*	*35*	*10*	*7*	*2*	*27*	*4*	*7*	*11*	*6*	115
Light	*1*	*26*	*11*	*4*	*29*	*25*	*2*	*0*	*6*	*13*	117
Totals	22	130	46	19	35	108	19	61	53	33	526

Lastly, we can look at the Location vs. Event table (table 5). The chi-square calculation from this table is 315.5827, with 90 degrees of freedom and a p-value of less than $2.2x10^{-16}$. Therefore the "Location" and "Event" variables are dependent on each other. Interesting odds ratios from the table are that villains in the Southeast or London are 3.7796 and 2.1237 times more likely to discuss suspicious things and Al Qaeda, respectively, than villains in other locations. Using the data in the three previous tables, various graphs, figures, and odds ratios can be created. Due to the required brevity of this paper, these additional items have not been included. Through the information gathered in previous tests, we are able to begin the process of developing a model to fit the data.

Table 5. Location vs. event.

	Alq	**Assoc**	**Bomb**	**Brit**	**Case**	**Islam**	**911**	**Not**	**Rec**	**Sus**	**Totals**
Brit.	*1*	*1*	*11*	*0*	*0*	*0*	*0*	*1*	*1*	*1*	16
E	*6*	*7*	*3*	*6*	*1*	*10*	*5*	*8*	*1*	*3*	50
EM	*0*	*0*	*1*	*1*	*0*	*0*	*2*	*3*	*0*	*0*	7
Lon.	*13*	*35*	*16*	*14*	*3*	*43*	*7*	*29*	*24*	*21*	205
NE	*0*	*6*	*0*	*4*	*0*	*1*	*6*	*1*	*1*	*1*	20
NW	*0*	*0*	*2*	*0*	*0*	*1*	*1*	*0*	*1*	*0*	5
Out	*7*	*7*	*3*	*0*	*0*	*4*	*7*	*3*	*3*	*1*	35
SE	*18*	*25*	*6*	*8*	*23*	*22*	*34*	*20*	*3*	*22*	181
SW	*0*	*1*	*3*	*4*	*4*	*8*	*1*	*8*	*3*	*5*	37
Unk.	*1*	*4*	*1*	*4*	*4*	*10*	*21*	*1*	*0*	*1*	47
WM	*0*	*1*	*4*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	6
Totals	46	87	50	41	35	99	84	74	38	55	609

Using R and the "Design" package, we wish to construct a model (*8, 12*) from the Ali Baba data set. The data has a two-level variable called "binlocal" that is used as our response or dependent variable. The predictors or independent variables from table 2 are "binTruth" and "Event." Before the model is created, we must check the various tables to see if any cells are small or empty. There are several methods in R that can be used to create easy-to-read tables from an inputted data set, as seen earlier. The two-way tables for binTruth vs. binlocal, binlocal vs. Event, and binTruth vs. Event are satisfactory to continue the analysis. We then wish to determine if the variables in the three tables are independent or dependent. This is done with a chisq.test line of code in R and compared at the 5% significance level. The three variable pairings have a p-value of less than .05, so we can conclude that the respective variables are dependent.

Now that the basic assumptions have been satisfied, we are able to look at our first logistic regression model. Figure 1 contains the partial output from R for the initial model that uses all three of the variables discussed previously.

Looking at the table of coefficients in the output (figure 1), we see that the intercept and several of the factors of "Event" are not significant in the model. A second model (figure 2) is created that removes the "Event" variable to see if we can improve our results.

```
Logistic Regression Model

lrm(formula = binlocal ~ binTruth + Event, data = fullset, na.action =
na.pass)

Frequencies of Responses
  0   1
159 367

       Obs  Max Deriv Model L.R.       d.f.          P
       526      1e-05      56.96         10          0

            Coef    S.E.   Wald Z P
Intercept   -0.3321 0.4359 -0.76  0.4462
binTruth     0.8224 0.2556  3.22  0.0013
Event=assoc  1.3448 0.4818  2.79  0.0052
Event=bomb  -0.4098 0.5342 -0.77  0.4431
Event=brit   0.7158 0.6475  1.11  0.2689
Event=case   2.6822 0.7447  3.60  0.0003
Event=isl    1.1722 0.4841  2.42  0.0155
Event=nin    0.2715 0.6361  0.43  0.6695
Event=not    0.4826 0.5178  0.93  0.3513
Event=rec    1.4077 0.5530  2.55  0.0109
Event=sus    1.6087 0.6289  2.56  0.0105
```

Figure 1. Initial logistic model.

```
Logistic Regression Model

lrm(formula = binlocal ~ binTruth, data = fullset, na.action = na.pass)

Frequencies of Responses
  0   1
159 367

       Obs  Max Deriv Model L.R.       d.f.          P
       526      2e-14       6.15          1     0.0131

          Coef   S.E.   Wald Z P
Intercept 0.7009 0.1081 6.48   0.0000
binTruth  0.5564 0.2305 2.41   0.0158
```

Figure 2. Second logistic model.

The outputs in figures 1 and 2 show logistic regression models that are based on the Ali Baba data set. The frequency and observation sections allow us to check that our data was read into R correctly and that we can continue analysis. If these numbers did not match, we would need to go back to our initial R code to check for hidden errors that would not return error messages on the output screen. From figure 1, we can see that the likelihood ratio chi-square for the full model is 56.96, with 10 degrees of freedom, and a p-value of approximately 0. The model without "Event" has a chi-square value of 6.15, degrees of freedom 1, and p-value of .0131. This tells us that both of our models as a whole are statistically significant as compared to a model with no predictors.

The rest of the results show the table of coefficients, their standard errors, the Wald z-test, and the p-values. In figure 2, we can see that both the intercept and the "binTruth" variables are statistically significant at the 5% level. It is important to note that the output is given in ordered logits since we are using logistic regression. To interpret "binTruth," we can say that for a one-unit increase (going from 0 to 1), we expect a .56 increase in the expected value of "binlocal" on the log odds scale. The "Design" package also allows us to output the coefficients as proportional odds ratios as seen below.

```
Effects              Response : binlocal

 Factor       Low High Diff. Effect S.E. Lower 0.95 Upper
0.95
 binTruth     0   1    1      0.56   0.23 0.10       1.01
  Odds Ratio 0   1    1      1.74     NA 1.11       2.74
```

Figure 3. Proportional odds ratios.

Since "binTruth" is an indicator variable, the low and high will always be 0 and 1. For "binTruth," going from 0 to 1 increases the odds of the message location being in the Southeast

or London by 1.74 times, given that the other variables are held constant. It is important to note that the proportional odds assumption does hold for the coefficients in this regression.

4. Summary and Conclusions

The process of categorization was the first necessary step to examine the Ali Baba data set using CTA. Once that was completed, we were able to determine if the variables were related with any type of dependence. After this analysis, the location of the messages was centralized and various statistical odds were provided, as well. Further CTA statistics were generated but were not included due to the brevity of this paper. Finally, initial models were created using R packages, but additional research will be conducted this summer on the topic.

With the massive amounts of information available today, there is an increasing need to understand and interpret data provided by social networks and military reports. The modeling and predictive power of statistics in examining dark networks is currently underused. The use of CTA on this data set was time-consuming because of the layout of the data. However, a data set with additional entries, variables, or categories would better have exploited the abilities of CTA. With respect to R, there are various other statistical packages on the market that can provide equivalent results. However, it was the purpose of this paper to focus specifically on the R programming language. The R programming language is preferred by some because of the open-source nature of the language and for the availability of help resources. Overall, the R programming language has proved effective in analyzing the Ali Baba data set in the realm of CTA.

5. References

1. Agresti, Alan. *An Introduction to Categorical Data Analysis.* 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc., 2007.
2. Alberts, David S.; Garstka, John J., Stein, and Frederick P. *Network Centric Warfare.* Center for Advanced Concepts and Technology, 1999. Print.
3. Allanach, Jeffrey; Tu, Haiying; Singh, Satnam; Willett, Peter; Pattipati, Krishna. Detecting, Tracking, and Counteracting Terrorist Networks via Hidden Markov Models. University of Connecticut: Dept. of Electrical and Computer Engineering, n.d. Web. 22 June 2011. www.engr.uconn.edu/~sas03013/docs/Aerospace_HMM.pdf.
4. Arquilla, John; Ronfeldt, David. The Advent of Netwar. *Networks and Netwars: The Future of Terror, Crime, and Militancy*. Rand, 19 Mar. 2002. Web. 22 June 2011. http://faculty.cbpp.uaa.alaska.edu/afgjp/padm610/networks%20and%20netwar.pdf.
5. Boik, John. Lab 4: Correlation, Regression, Contingency Tables. *R Computing*. 30 Nov. 2007. Stanford University. Web. 13 June 2011. www.stat.stanford.edu/~jcboik/stat-141-2007/R_labs/07Lab4.pdf.
6. "Entering Data."*R Class Notes*. UCLA: Academic Technology Services, Statistical Consulting Group. Web. 9 June 2011. www.ats.ucla.edu/stat/r/notes/entering.htm.
7. Flynn, Michael T.; Pottinger, Matt; Batchelor, Paul D. Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan." *Center for a New American Security*. 04 Jan. 2010. Web. 12 July 2011. http://www.cnas.org/node/3927.
8. Harrell, Jr. Frank E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression and Survival Analysis.* Springer, New York, 2001.
9. Jaworowski, Mark; Pavlak, Steve. Ali Baba Scenario 1 Description. 2003. Print. 1 June 2011.
10. Kabacoff, Robert I. Frequencies and Crosstabs. *Quick-R: Accessing the Power of R. Web. 10 June 2011.* www.statmethods.net/stats/frequencies.html.
11. Map of England's Regions. Outdoor Sport & Leisure. n.d. Web. 21 June 2011. www.outdoor-sport-leisure.net/activities-UK.htm.
12. Ordinal Logistic Regression. *R Data Analysis Examples*. UCLA: Academic Technology Services, Statistical Consulting Group. Web. 11 July 2011. www.ats.ucla.edu/stat/r/dae/ologit.htm.

13. Quinn, Kevin. Log Linear Example.” 11 Dec. 2002. University of Washington. Web. 13 June 2011. www.stat.washington.edu/quinn/classes/536/S/loglinexample.html.

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Optimizing Strategies for an Observation-nudging-based Four-Dimensional Data Assimilation Forecast Approach with WRF-ARW

ANDRE PATTANTYUS
MENTOR: ROBERT DUMAIS, JR.
BATTLEFIELD ENVIRONMENT DIVISION
WHITE SANDS MISSILE RANGE, NEW MEXICO

Contents

List of Figures 23

List of Tables 23

Abstract 24

Acknowledgments 25

Student Bio 26

1. Introduction 27

2. Experimental Design 29

3. Results and Discussion 30

4. Summary and Conclusion 36

5. References 38

List of Figures

Figure 1. Domain configurations for (a) domain 1 and (b) domain 2. Assimilated observations are represented as circles (surface stations) and triangles (radiosondes) and their radii of influence are approximated by the open circles with radii equal to 120 km and 60 km in (a) and 40 km and 20 km in (b). Verification stations are also labeled with 4-letter station ids. Location of domain 2 in (a) represented by box. 29

Figure 2. Estimated one day accumulated precipitation from radar ending 12 UTC 1 December 2007. 31

Figure 3. Experimental accumulated precipitation for the entire simulation over domain 1. Experiment labels top right. 32

Figure 4. Observed and experimental accumulated precipitation in domain 1 for (a) KNYL, (b) KBLH, and (c) SWLC1. 32

Figure 5. (a) Temperatures and (b) dewpoints at KNYL valid 07 UTC 30 Nov 2007 - 06 UTC 1 Dec 2007. 33

Figure 6. Model reflectivity and upper-level winds from experiment (L) HLRAD and (R) SMRAD valid 02 UTC 1 December 2007 (top) and 03 UTC 1 December 2007 (bottom). 34

Figure 7. Experimental accumulated precipitation for entire simulation period over domain 2. Experiment labels top right. 35

Figure 8. Observed and experimental accumulated precipitation in domain 2 for (a) KNYL, (b) KBLH, and (c) SWLC1. 35

List of Tables

Table 1. Model physics options used in all experiments. 30

Table 2. List of model experiments with descriptions. 30

Abstract

The Weather Research and Forecasting model is an open-source numerical weather prediction model with numerous features, such as the four-dimensional data assimilation system, an option to nudge forecasts to observations in order to improve forecast quality and performance. This process can act as a dynamical initialization prior to the forecast period to reduce errors that result from interpolation across the grid. For data void regions, assimilating observations from a temporary network could vastly improve forecasts in the region. However, users must specify how much influence each observation will get in the model and over how large an area each observation will influence. These weighting factors will have a great effect on the forecast quality. This research attempts to determine the optimal setting of the radius of influence for upper air and surface observations within the assimilation system. A case study over Yuma, AZ, is examined in which a low-pressure system produced dynamic and orographic precipitation across the region. Observations from the U.S. Air Force in Yuma will be assimilated into Weather Research and Forecasting for the 06 coordinated universal time forecast for an entire 24-h forecast cycle. Results suggest upper air observations have a greater impact on forecast accuracy given a large radius of influence.

Acknowledgments

I would like to acknowledge the mentorship provided by Robert Dumais, Jr., for providing leadership and focus on research goals. Technical support was provided by Richard Padilla and Robert Flanigan. Thanks to Yuma Proving Ground (YPG) for providing the data for this project. Thanks also to the rest of the U.S. Army Research Laboratory (ARL) staff at White Sands Missile Range (WSMR) for a pleasant working environment and for making me feel welcome for the summer internship.

Student Bio

I received my B.S. in Meteorology from Plymouth State University, NH, in 2007. I received my M.S. in Meteorology from Florida Institute of Technology in 2010. My master's thesis titled "Numerical Investigations of Mountain Flows in the Las Vegas and Owens Valleys" involved high resolution mesoscale modeling using the WRF-ARW. In June 2011, I was the lead author of "Improving High-Resolution Model Forecasts of Downslope winds in the Las Vegas Valley," which was published in the *Journal of Applied Meteorology and Climatology*. This paper was the result of my master's thesis research. I also presented this research at the Florida Academy of Sciences Annual Meeting in March 2010, the 5th International Symposium on Computational Wind Engineering in May 2010, the International symposium for the Advancement of Boundary Layer Remote Sensing in June 2010, and the AMS Annual Meeting in January 2011. I am currently in the Environmental Science Ph.D. program at Florida Tech and continue to research the use of the WRF-ARW at high-spatial resolution over complex terrain and the use of four-dimensional data assimilation (FDDA). Other interests include the use of the WRF-ARW and FDDA for short-term cycling forecasts for wind farms. I plan to pursue a career in applications meteorology and mesoscale modeling.

1. Introduction

U.S. Army test ranges, garrison installations, and tactical battle regions are often located in regions of complex terrain that depend on accurate, high-resolution weather analyses and forecasts to carry out routine tests, daily activities, and mission execution operations. An efficient yet sophisticated high-resolution modeling system is required to both ingest local and regional sources of weather data, and to produce detailed local forecasts/nowcasts, which capture the unique time-dependent flow and sensible weather conditions related to complex terrain environments under varying synoptic and mesoscale forcing states. Unfortunately, neither local observations nor the courser resolution global and mesoscale model outputs from national centers are able to resolve these local factors in their data analyses and forecasts (*1*). This is especially the case in many areas where military personnel are stationed or deployed. In such areas, sufficient or reliable weather data is often limited or may lack altogether, despite the presence of complex local terrain and land use features. This poses a serious challenge for high-resolution mesoscale numerical weather prediction (NWP) models that rely on denser networks of weather observation sites and platforms to provide both surface and upper air data for forecast initialization. If observation data is lacking over most of a mesoscale forecast region, the high-resolution model tends to interpolate the sparse observations across the entire region (and rely more on interpolated background model states provided by a previously executed courser external model run). This will certainly have a negative impact on the high-resolution forecasts if the observations and background model states do not resolve the critical and dominant mesoscale and synoptic details. The Advanced Research version of the Weather Research and Forecasting (WRF-ARW) model has the capability to ingest observations and relax (or nudge) the model towards these observations in order to produce a more accurate short-range forecast.

A form of continuous and dynamic atmospheric data assimilation called four-dimensional data assimilation (FDDA), or observation/station nudging, has demonstrated the potential to improve mesoscale model initializations and short-range forecasts based upon the insertion of local/regional asynoptically observed direct weather quantities (temperature, winds, moisture, pressure) into model integration through fictitious relaxation "forcing" terms. Each grid point in observation-nudging FDDA is relaxed or "nudged" based upon weighted averages of each innovation (i.e., differences between the observation and external background model value at a given location). These weights are based upon a grid-dependent horizontal radius of observation influence, observation time window, vertical influence, and parameter-specific nudging/relaxation time-scale magnitude values determined through user-specified namelist values. The time window allows one to ramp observational nudging up and down. A new FDDA user-specified parameter in the WRF-ARW v3.2.1 namelist is "obs_sfcfact", which allows the user to specify a smaller radius of influence for surface observations while maintaining a larger radius of influence for upper air observations. Additional namelist flexibility in v3.2.1 is

allowing a user to specify the vertical influence depth within the model planetary boundary layer of surface observations, based also upon the particular boundary layer stability regime.

Incorporation of the observation-nudging FDDA system into the WRF-ARW framework has been followed by numerous studies highlighting the improvement in forecasts. Studies by Penn State (*2*) found improvements at all model levels and grid resolutions (down to 4 km) by using FDDA over CONUS and Alaska for all variables being ingested. Attempts to improve precipitation forecasts in Israel via FDDA, and with grid resolutions down to 1 km, verified favorably against observations (*3*). Researchers noted forecast skill degraded with decreasing resolution due to inadequate representation of orographic forcing. A study over White Sands Missile Range in 2005 (*4*) found that a model surface cold bias reduced explicit precipitation forecasts. However, the model was able to capture the dominant orographically-generated precipitation. Initial results from Dugway Proving Ground, UT, suggest that the model FDDA was able to better capture a precipitation event than its predecessor MM5, even though the FDDA had not been fully updated to match that of the MM5 FDDA yet (*1*). Although improvements to traditional model forecasts have been shown through the use of FDDA, a number of questions remain on the effect of the user-specified weighting factors and the use of isotropic weighting. In observation nudging, experience-based ad-hoc weighting functions act as the "Kalman gain" used in more sophisticated analysis/data assimilation systems such as 3DVAR, 4DVAR, and EnKF. The Kalman gain in these more sophisticated systems relies upon optimally-estimated values of background and observation errors. Computationally, FDDA observation nudging is a more efficient approach, and its continuous versus intermittent assimilation method is advantageous for high spatial/temporal scales, and asynoptic data rates and observing frequencies.

Given the high cost of upper air versus surface observations and the difficulty in deploying systems in the field, it is necessary to quantify the impact of both types of observations on the model's FDDA, which may, in turn, be synoptic/mesoscale-forcing and location-dependent. In particular, what size to make the radius of influence for observations? How much improvement is there from surface versus upper air observations? In this way, strategies for optimizing FDDA can be developed, and the number of observations needed can be determined based on forecast needs. It is hypothesized for this study of a dynamically forced southwestern U.S. precipitation event that upper air data will have a larger impact on forecast accuracy.

A series of model experiments were carried out over Yuma Proving Grounds (YPG), AZ, from 0600 UTC 30 November to 0600 UTC 1 December 2007. Observations were ingested into WRF-ARW FDDA from the YPG base surface mesonet. Special radiosondes were released across YPG at close to a 90-min frequency throughout much of this day in support of special ongoing range exercises, and these were incorporated through the FDDA. Verification of the model results was done using available NWS and Mesowest surface stations outside of the YPG mesonet but within the model domain both in close and far proximity to YPG FDDA stations.

2. Experimental Design

A one-way, double-nest domain configuration of the WRF-ARW model, with horizontal grid spacing of 3 km and 1 km, respectively, was run for all experiments using 90 terrain-following vertical levels. Figure 1 displays the horizontal domain configuration. The center of the model domain is located within YPG. Observations included in the FDDA were taken from the YPG surface mesonet and radiosondes, and are shown on figure 1. There are five surface stations and three sounding locations included. As is shown, the surface stations were tightly clustered in one area, while the sounding locations were more evenly spaced. Two sets of sounding and surface stations are collocated. The large open circles represent our user-defined radii of influence for the observations that were tested in the experiments. The largest circles represent a radii = 120 (40) km, and the smaller circles are radii = 60 (20) km in domain 1 (2). The five station identifiers represent stations used in model verification. The model physics are held constant for all experiments and are summarized in table 1. Initial and lateral boundary conditions come from the NAM-218 forecasts produced by NCEP (~12 km).

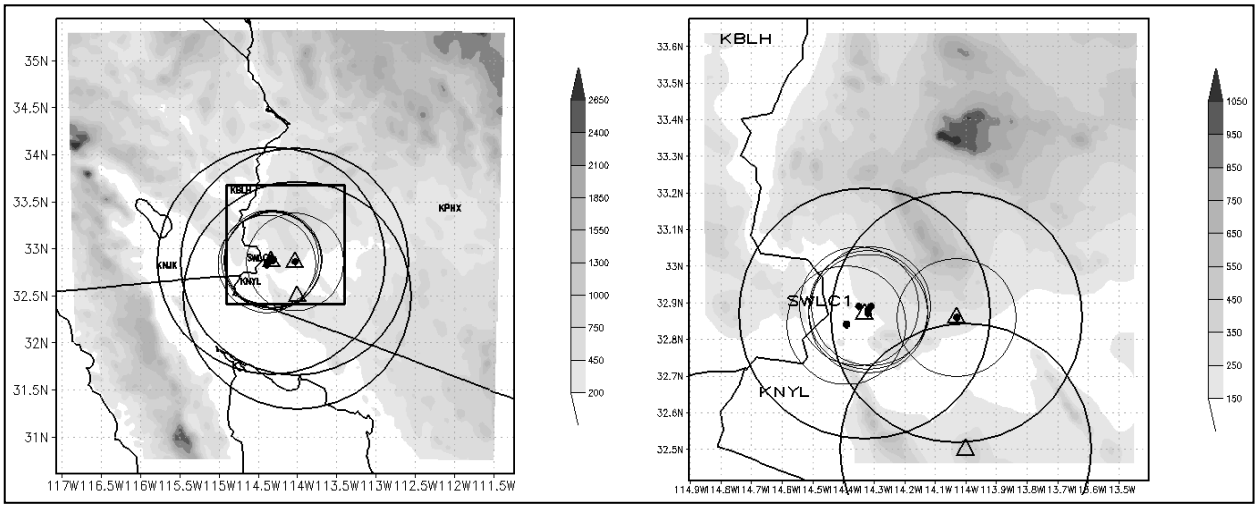

Figure 1. Domain configurations for (a) domain 1 and (b) domain 2. Assimilated observations are represented as circles (surface stations) and triangles (radiosondes) and their radii of influence are approximated by the open circles with radii equal to 120 km and 60 km in (a) and 40 km and 20 km in (b). Verification stations are also labeled with 4-letter station ids. Location of domain 2 in (a) represented by box.

Additionally, an adaptive time-stepping method was used during integration of all experiments, except where specified. This method allows for shorter running time and less computational expense. Certain microphysics options were found to cause a positive bias in precipitation forecasts (5). However, these results are only from course resolution grids (>10 km), where convection is parameterized. Results have not been published on grids smaller than 5 km, in which cumulus physics are solved explicitly. The dynamic time-step used hereafter begins with a time-step of 18 s (6) in domain 1 (2). The maximum (minimum) allowed time-step is 3× (.5×)

the starting value. Runs with fixed, static time-steps were also executed to determine if and how dynamic time-stepping affects the explicit precipitation process.

Table 1. Model physics options used in all experiments.

Physics	Scheme
Land surface	Noah
surface Layer	QNSE
PBL	QNSE
SW radiation	Dudhia
LW radiation	RRTM
Cloud microphysics	Thompson

All experiments were initialized at 0600 UTC 30 November 2007 and integrated for 24 hours. Observations were assimilated over the entire period of integration, except for the experiment labeled COLDS. The list of experiments and brief descriptions are found in table 2.

Table 2. List of model experiments with descriptions.

Experiment	Description
LGRAD	radii of influence for all obs domain 1 = 120 km, domain 2 = 40 km
HLRAD	radii of influence for UA obs remains same but surface obs 60 km, 20 km
SMRAD	radii of influence for all obs domain 1 = 60 km, domain 2 = 20 km
HSRAD	radii of influence for UA obs remains same but surface obs 30 km, 10 km
DFRAD	radii of influence for all obs set to 18 km, 6 km
STATIC	same as HLRAD but with fixed time-step of 9 sec for domain 1, 3 sec for domain 2
COLDS	same as STATIC but with no fdda

Analysis of model results will focus on the ability of the model to capture the rain event intensity and timing over the entire domain. Model rainfall totals compared to the verification stations totals will also be addressed. Variables that were assimilated will also be analyzed at the verification stations to determine how and if the observations had an impact. Observations will be compared to model results from the simple nearest grid point, as neither the number of stations nor the time period to validate was long enough to produce statistical significance or warrant the use of more robust methods.

3. Results and Discussion

The rain event that occurred across the domain from 30 November to 1 December 2007 was the result of a surface low pressure system moving up the coast of California and the passage of a strong upper-level trough. The coastal range of California and Baja, Mexico, provided additional lift and moisture advection up from the Sea of Cortez, which fed the system as it progressed

through the region. A distinct convective band progressed through YPG from the west at 0200−0300 UTC 1 December and produced much of the precipitation in that local area. Overall, the system produced significant orographic, dynamic, and convective precipitation, as seen in the radar estimated precipitation in figure 2.

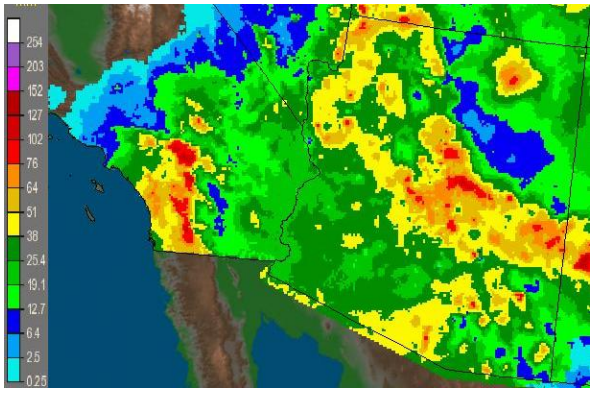

Figure 2. Estimated one day accumulated precipitation from radar ending 12 UTC 1 December 2007.

Objective analysis of the accumulated precipitation from domain 1 (figure 3) showed the same general features: a band of heavy precipitation over the coastal range with an area of low precipitation on the leeward side, a band of heavier precipitation near the California/Arizona border, and orographically forced precipitation to the NE in the Mogollon Rim of Arizona. Larger radii of influence showed more expansive regions of higher precipitation near southern Nevada and to the SE of the domain, while producing smaller regions of low precipitation in the lee of the coastal range, compared to runs with smaller radii of influence. The smaller radii of influence produced more orographic precipitation in the northwest of the domain and to the northeast in Arizona. However, all nest 1 simulations appear to have significantly under-predicted the magnitude of the precipitation amounts in the central and eastern Mogollon Rim country of Arizona. There areas were resolved explicitly on this 3-km outer nest. There were no major differences between precipitation from adaptive and static time-stepping runs. DFRAD showed less convective precipitation in the Yuma area and along the California/Arizona border, while producing slightly heavier orographic precipitation, especially to the NE. COLDS also showed a similar trend with noticeably less precipitation in the area of YPG.

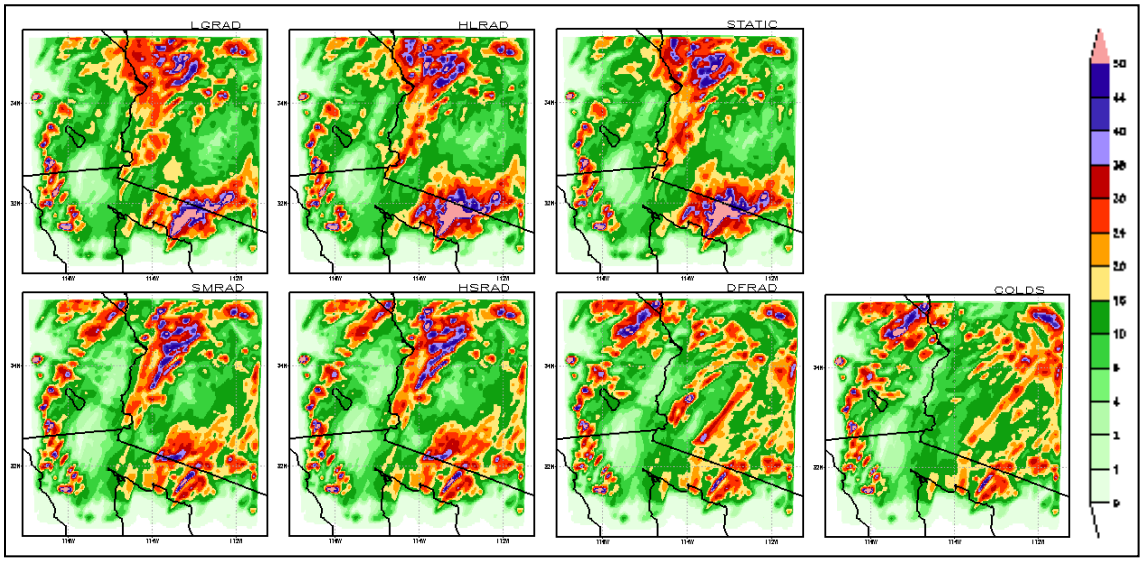

Figure 3. Experimental accumulated precipitation for the entire simulation over domain 1. Experiment labels top right.

The validation stations in close proximity to the FDDA sites showed differing trends in total precipitation, depending on how and if the passing convective bands were resolved. In figure 4, the staggering of heavy precipitation by approximately an hour was evident from observations. Model results showed similar trends in performance at KNYL and KBLH, with improvement from LGRAD to HLRAD and STATIC, while performance decreased in the remaining runs. At SWLC1, the smaller radii runs matched the precipitation totals better than the larger radii runs, which overestimated precipitation.

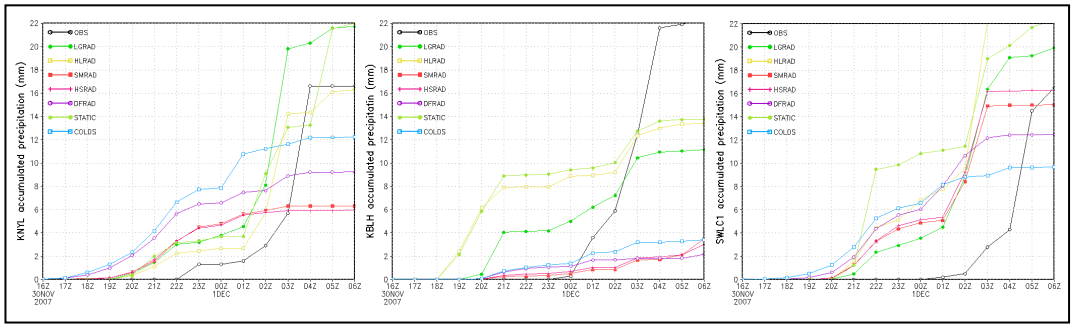

Figure 4. Observed and experimental accumulated precipitation in domain 1 for (a) KNYL, (b) KBLH, and (c) SWLC1.

It was evident that the model initiated precipitation too early across the entire domain for all experiments, often 3−4 hr prematurely in the YPG area. This was possibly due to slightly premature low-level moisture advection into the region from the Sea of Cortez (Note: this is not clear from the surface dewpoint comparisons at KNYL other than COLDS). The model showed a cold bias at all verification stations and can be seen in a sampling of experimental values at KNYL compared to observations (figure 5). Station dewpoints tracked observations during the later part of the runs after an early moist bias was eliminated. These two factors, the cold bias

and moist bias, increased the amount of latent instability at the surface and promoted convection to occur earlier. From observations, convection appeared to be limited by availability in moisture. Also, low-level convergence likely aided in this process to provide additional lift and produced a longer rain event than was observed.

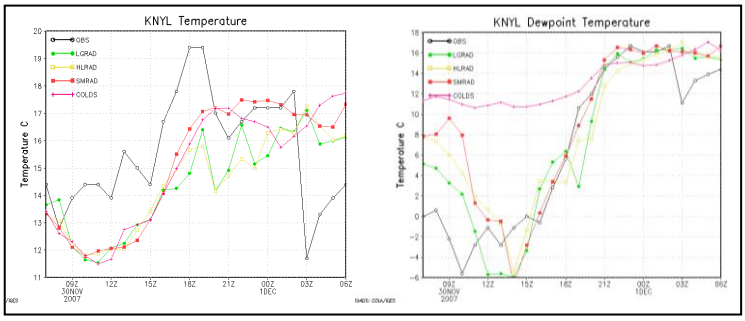

Figure 5. (a) Temperatures and (b) dewpoints at KNYL valid 07 UTC 30 Nov 2007 - 06 UTC 1 Dec 2007.

As shown in the observed accumulated precipitation graphs in figure 4, much of the YPG-area rainfall fell within a 2-h window near the end of the period. This was the result of a convective rain band that formed and moved in from the W/SW out of SE California and Northern Mexico. The model developed this band with larger radii of influence for upper air observations but to a lesser degree in runs with smaller upper air radii (figure 6), resulting in the gap in accumulated precipitation seen in the experiments (figure 4). As widespread a precipitation shield as observed over the coastal range was not seen in any experiment, but this is almost certainly due to the mountains' location so close to the model outer nest lateral boundary. The radii size seemed to influence convection downstream, as well as upstream, as there was more reflectivity evident in the SMRAD runs to the NE. There was a substantial difference in precipitation produced downstream at Phoenix, AZ, (KPHX) between using larger or smaller radii of influence for upper air observations (c.f. figure 3).

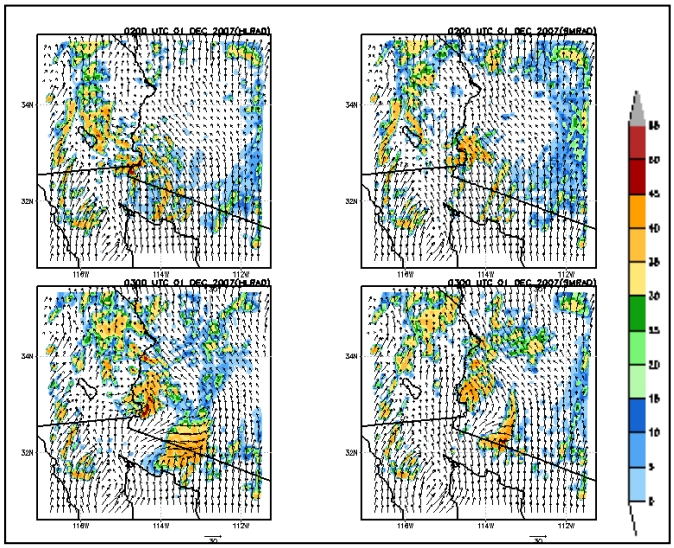

Figure 6. Model reflectivity and upper-level winds from experiment
(L) HLRAD and (R) SMRAD valid 02 UTC 1 December 2007
(top) and 03 UTC 1 December 2007 (bottom).

The effects of the varying radii of influence were also seen in the temperatures and dewpoints at both KBLH and SWLC1, as the influence of the nudging diminished throughout the experiments (not shown). LGRAD and HLRAD performed better on average than SMRAD, HSRAD, and DFRAD. Wind speed and direction showed more spread over the experiments due to the nature of the complex flows surrounding the convective cells, which were represented differently in each run, and caused varied responses to the winds at each location.

The results for domain 2 followed those of domain 1, with LGRAD and HLRAD outperforming most other experiments (figure 7). This was expected since these experiments developed the outer nest convection the best, which fed into domain 2. The experiments with larger radii for the upper air observations produced more precipitation upstream of the observations (south) than experiments with smaller radii. The maximum region of precipitation for the small radii runs was found over the high point in the domain and verifies against observations. This feature was less defined in LGRAD and HLRAD but was clear in STATIC. DFRAD produced more (but less organized) precipitation upstream from observations than SMRAD and HSRAD but still well below observations. COLDS, with no observations, had similar results to DFRAD but with less downstream precipitation. However, there was a more noticeable difference in the performance between the adaptive and static runs in domain 2.

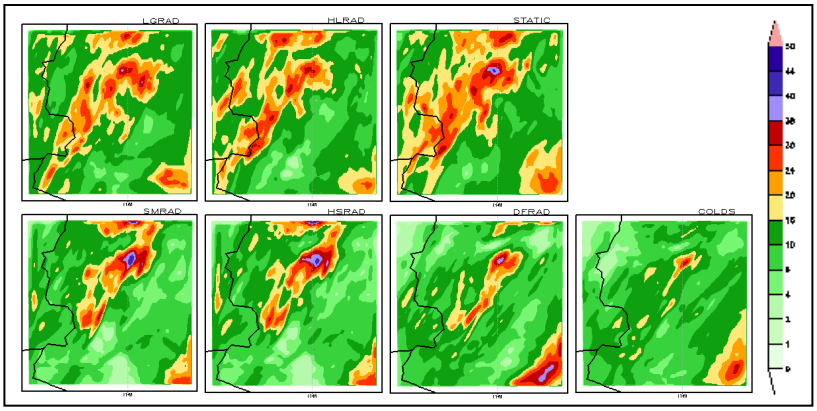

Figure 7. Experimental accumulated precipitation for entire simulation period over domain 2. Experiment labels top right.

The performance at the verification stations was similar in domain 2 as in domain 1, with KBLH and KNYL performing better with large radii experiments and SWLC1 with small radii experiments (figure 8). The results at KBLH were relatively unchanged from those in domain 1, which may be due to its close proximity to the boundaries of domain 2. SWLC1 was closer to the observations than KNYL and so experienced more nudging than KNYL (c.f. Fig 1). Regardless, experiments with larger radii of influence produced greater amounts of precipitation, but this may be more of a result from the boundary forcing from domain 1 rather than from nudging in domain 2. There was also an improvement at KNYL and SWLC1 in temperatures and dewpoints in experiments with large radii in domain 2 (vs from the domain 1 values not shown). Values did not change in experiments with small radii at KNYL since it was outside of the radii of influence, but values did increase at SWLC1 going from too dry to too moist. At SWLC1, experiments where the radii encompassed the station showed better performance with temperatures and dewpoints than those that did not.

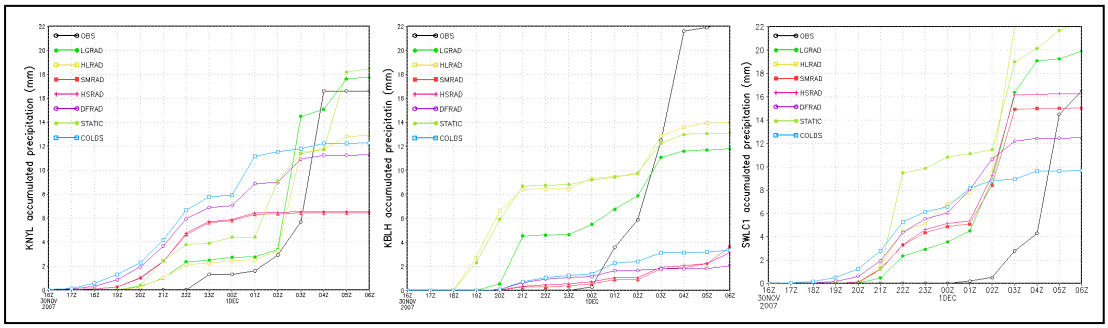

Figure 8. Observed and experimental accumulated precipitation in domain 2 for (a) KNYL, (b) KBLH, and (c) SWLC1.

4. Summary and Conclusion

Results showed that the use of large radii of influence appeared necessary for both surface and upper air observations in the WRF-ARW FDDA. In this case, upper air observations appeared to provide the greatest improvement in forecasted precipitation, while surface observations showed the most improvement prior to the onset of precipitation. Adaptive time-stepping did not display a bias in precipitation while cutting down on runtime by approximately 50%, which is critical to the Army's interest in running frequent FDDA cycling using such nesting configurations.

Significant differences in rainfall were produced over YPG, in both domains, between experiments using large radii and those with small radii. The large radii experiments produced more precipitation that matched observed radar estimated precipitation patterns over the region and local station precipitation totals. These runs also produced more coherent reflectivity values, representing the convective band that produced heavy precipitation in the area, which was not produced in experiments with small radii (especially for upper air observations). Surface variables also showed improvement in these experiments, which appeared to be more of a result of upper air observations rather than surface observations. More research is needed on surface variables to confirm these findings. Downstream effects from upper air observations could be seen in domain 1 precipitation fields, likely as a result of the current isotropic weighting approach used in the FDDA nudging. This may be solved by using a slightly smaller radius of influence, or in the future new approaches to FDDA that can apply anisotropic weightings.

Based on the precipitation totals produced from the experiments, upper air observations appeared to have a greater impact on performance than surface observations for this highly dynamic case. Only surface dewpoints showed a dramatic improvement from the inclusion of surface observations in FDDA. Wind speeds and directions showed some improvement but were difficult to verify against suspect station data. However, in conditions of weak synoptic/dynamic forcing and significant terrain and/or land use heterogeneity, these findings could be completely different. Much more research must be done.

Adaptive time-stepping produced impressive savings in computational cost and runtime, while producing results similar from a static time-step. The omission of the STATIC series in figure 5 was due to its striking resemblance to its adaptive counterpart HLRAD. The improvement in runtime was approximately 50%. This is slightly higher than previously reported savings because the extra computational costs from the relaxation terms in FDDA are significantly reduced at larger time-steps with a minimal impact on performance. On the 1-km domain, there were larger differences in the precipitation field between the static and adaptive experiments than on the 3-km domain.

Further research is needed to continue to improve WRF-ARW FDDA. Potential experiments should investigate the use of FDDA on the coarse domain only; the inclusion of upstream sounding data to improve upstream precipitation and convection; the performance of different microphysics schemes; the assimilation of one variable at a time (i.e., temperature, moisture, or winds); and, most importantly, the reduction of the model daytime cold bias at the surface.

5. References

1. Liu, Y.; Bourgeois, A.; Warner, T.; Swerdlin, S.; Hacker, J. 6th Annual WRF Users' Workshop, Boulder, CO, 2005.
2. Deng, A.; Stauffer, D.; Gaudet, B.; Dudhia, J.; Hacker, J.; Bruyere, C.; Wu, W.; Vandenberghe, F.; Liu, Y.; Bourgeois, A. 10th Annual WRF Users' Workshop, Boulder, CO, 2009.
3. Rostkier-Edelstein, D.; Liu, Y.; Roux, G.; Givati, A.; Pietrokowski, A.; Ge, M.; Hahmann, A.; Pinto, J.; Warner, T.; Swerdlin, S., 9th European Conference on Application of Meteorology, Toulouse, FR, 2009.
4. Yu, W.; Liu, Y.; Warner, T.; Bullock, R.;, Brown, B.; Ge, M., 7th Annual WRF Users' Workshop, Boulder, CO, 2006.
5. Hutchinson, T. A. 19th Conference on Numerical Weather Prediction, Omaha, NE, 2009.

Human Research & Engineering Directorate (HRED)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Motion Simulation in the Environment for Auditory Research

BRAXTON BOREN
MENTOR: MARK ERICSON
HUMAN RESEARCH AND ENGINEERING DIRECTORATE
ABERDEEN PROVING GROUND, MARYLAND

Contents

List of Figures 43

Abstract 44

Acknowledgments 45

Student Bio 46

1. Introduction 47

2. Streaming Audio in MATLAB 47

3. Panning Algorithms 48

4. Signal Processing 49

5. Discussion 50

6. References 51

List of Figures

Figure 1. Plot of a symmetrical “dogbone” flight pattern.....48

Abstract

Virtual sound source motion has been implemented in the U.S. Army Research Laboratory's Environment for Auditory Research, which contains a 57-channel spherical loudspeaker array located in a semi-anechoic chamber. Using the low-latency PortAudio application programming interface from the Psychophysics Toolbox Version 3, we are able to dynamically update 57 channels of streaming audio in real time using MATLAB for signal processing. Both Distance-Based Amplitude Panning (DBAP) and Vector Base Amplitude Panning (VBAP) have been implemented in MATLAB for controlling source motion. Sources are defined on a given path, such as a circle, ellipse, or the "dog bone" pattern often used in aviation. Although DBAP works convincingly for virtual sources located on the sphere defined by the loudspeaker array, VBAP is needed to position sources outside the array. Source motion paths are defined parametrically with respect to time, and the playback buffer updates the panned position every 11.5 ms. Based on the source's instantaneous distance, diffuse-field or free-field amplitude attenuation is added in MATLAB, as is air absorption filtering. This virtual sound source method will be used for a variety of audio simulations and auditory experiments.

Acknowledgments

Many thanks to my advisor, Mark Ericson, for giving me the tools and information needed to quickly get the ropes of an incredible audio reproduction facility at the Environment for Auditory Research. Mark's hospitality and guidance were crucial to the quick completion of this research. Thanks most of all to my wife Laura, who graciously put up with me being gone to undertake this work.

Student Bio

I studied Music Technology at Northwestern University, graduating as valedictorian of the School of Music in 2008. In 2009, I was awarded the Gates Cambridge Scholarship to fund a postgraduate degree at the University of Cambridge. I read for a Master of Philosophy (MPhil) in Physics at the Cavendish Laboratory, working in archeological acoustics. I am currently a second-year doctoral student at New York University, where I am working as part of the Music and Audio Research Laboratory's 3-D Audio group. My future plans include, but are not limited to, learning and helping others to learn.

1. Introduction

The Environment for Auditory Research is a state-of-the-art facility designed for a wide range of basic and applied auditory research (*1*). Its multiple loudspeaker configurations can create immersive audio simulations designed to test human auditory perception, detection thresholds, and angular and distance localization. One of its three main reproduction environments, the Sphere Room, contains a 57-channel spherical loudspeaker array, allowing a dense reproduction field along both azimuth and elevation dimensions.

The Sphere Room allows for easy reproduction of static soundscape simulations through a simple multichannel audio editor, such as Adobe Audition. However, the system must reproduce both static and moving sources to convincingly simulate real-world environments. Rendering a moving source via individual pans between 57 channels becomes quite tedious in a software environment designed for static sources. In Chowning's pioneering work on source motion, a signal-processing system was designed to precompute synthesized signals and change amplitudes on a quadraphonic reproduction system (*2*). This method has been adapted in the Sphere Room to precompute channel gains over time for a given source path and panning algorithm with recorded rather than synthesized source signals. The array of channel gains is then used to move the source by updating each loudspeaker's gain with real-time streaming audio. While this system was developed for experiments in the Sphere Room, it can also simulate source motion in less dense audio reproduction facilities.

2. Streaming Audio in MATLAB

MATLAB was used to implement the source motion system, which allowed low-level control of panning and signal processing. MATLAB interacts with the Sphere Room's RME Hammerfall DSP audio interface via the PortAudio application programming interface available through MATLAB's Psychophysics Toolbox Version 3 (PTB-3) extensions (*3*). PortAudio allows high-fidelity, low-latency audio control using static or streaming buffers. The streaming algorithm works by initiating an audio buffer and then continually appending small amounts of audio data onto the end of the buffer while playback is active.

The creators of PTB-3 recommend that each append to the buffer be no longer than half the latency of the audio system. Although lower latency allows the system to handle more iterations of the streaming loop in a given amount of time, the latency must allow time to process all real-time signal operations or the buffer will underflow. A series of tests using white noise and sinusoidal signals showed that the Hammerfall's internal buffer size should be set to 1024 samples, which corresponds to a system latency of 23 ms. This allows the streaming algorithm to

update every 11.5 ms without underflowing. This update speed is sufficient to provide perceptually smooth panning paths across hundreds of individual positions. Multiple sources can be processed simultaneously, but the buffer size must be increased proportionally to handle real-time operations for each source.

3. Panning Algorithms

The Spatial Audio MATLAB Toolbox (*4*) was used for panning the virtual sources within the Sphere Room. This toolbox contains functions for both Distance-Based Amplitude Panning (DBAP, also called Vector Distance Panning in the toolbox) and Vector Base Amplitude Panning (VBAP). For both panning algorithms, the source's motion path was defined parametrically with respect to time. Test paths included a circle, ellipse, and the "dogbone" flight pattern often used in aviation—two long parallel lines with a larger curve at the end where the pilot turns around to come back along the other side, as shown in figure 1.

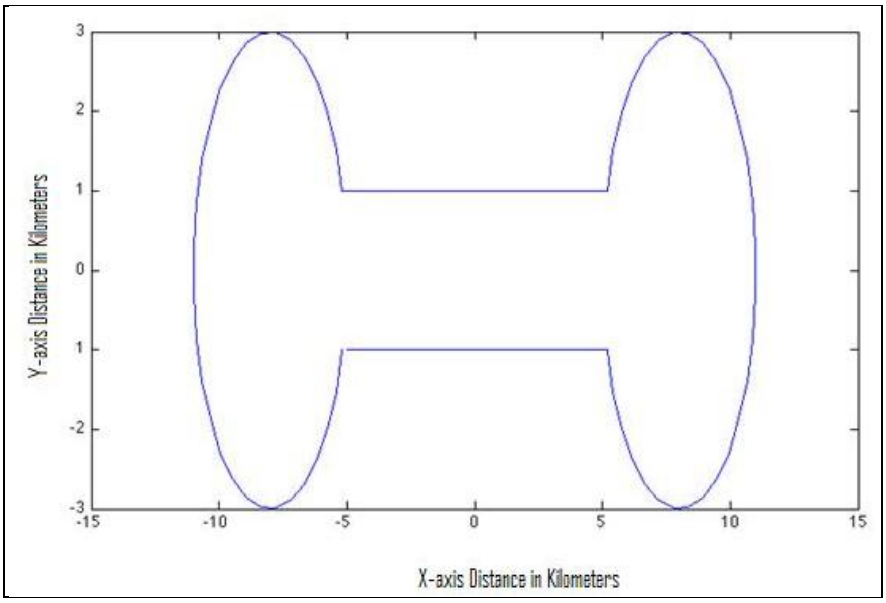

Figure 1. Plot of a symmetrical "dogbone" flight pattern.

DBAP pans a sound by simultaneously manipulating the gains of all of a system's loudspeakers based on only the source's distance to that loudspeaker and disregarding listener position (*5*). DBAP is relatively new but has performed well in subjective tests against more standard spatialization techniques, such as VBAP and Ambisonics (*6*). Because DBAP works for any number of loudspeakers at a single moment, it was the first panning algorithm implemented. DBAP panning gives a convincing impression of virtual sources along the sphere represented by the loudspeaker array. Source motion implemented with DBAP is smooth with a widened source impression because of the additional loudspeakers used to produce each virtual source. However,

the drawback of DBAP is that it cannot accurately represent positions outside the loudspeaker array, since the difference between the loudspeakers' individual distances to a faraway source becomes negligible. While the creators of DBAP suggest other workarounds for such situations, such as projecting a distant source onto the loudspeaker array and then applying attenuation effects, it was decided to use only DBAP for virtual sources within the bounds of the loudspeaker array itself.

VBAP (*7*) is a robustly developed panning algorithm based on the following principle: given a set of three linearly independent vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ arranged in a triangle, any fourth vector $\vec{d}$, the desired source location, within that triangle may be created by a linear combination of the initial three vectors:

$$l_1\vec{a} + l_2\vec{b} + l_3\vec{c} = \vec{d}\ . \tag{1}$$

If the first three vectors represent the positions of three loudspeakers arranged in a triangle, then a virtual source may be placed within that triangle by scaling and adding the first three vectors together. The coefficients of each vector thus become gain factors for each of the three loudspeakers in the triangle. While the Spatial Audio MATLAB Toolbox contained a function for calculating the VBAP gains for any given triangle of loudspeaker vectors, unconstrained source motion would require the system to automatically find which triangle the source vector intersected at any given point in time.

Sunday's algorithm (*8*) was implemented to calculate the vector-triangle intersections. The 57 loudspeakers in the Sphere Room define 110 distinct triangles, using the smallest angles possible in each case. These triangles' position vectors each define a unique plane, and the intersection algorithm first determines whether the source vector intersects that triangle's plane. If it intersects, the algorithm finds the plane's parametric coordinates **s** and **t** that determine the point of intersection. If the sum of **s** and **t** is between 0 and 1, then the vector intersects the given triangle. Finding these parameters requires only five distinct dot-product calculations, and the normal vectors of each plane may be precomputed and stored to optimize computation for static loudspeaker arrays. Because of this optimization, this algorithm is extremely efficient for VBAP systems whose loudspeaker triangles usually do not change position. The system then returns the current loudspeaker triangle based on the source's position and uses the Spatial Audio MATLAB Toolbox's VBAP function to calculate the appropriate gains within the loudspeaker triangle. These gains are then applied to the current point in the output signal and appended to the corresponding channels of the streaming 57-channel array via PortAudio.

4. Signal Processing

The method just described was initially intended for the reproduction of high-quality spatial recordings of aviation vehicles that were made available along with detailed x-y-z position data

over time. When these data are used, reproduction may be perfectly adapted to a given recording, calculating loudspeaker gains in response to the recorded source's interpolated movement over time. This method maintains physical accuracy because distance-based attenuation, low-pass filtering due to air absorption, and Doppler effects are all contained within the original recording.

To make the system more flexible, however, we have begun to add additional signal processing capabilities to allow movement of an arbitrary signal. The current system has options of including free-field (inverse square) or diffuse-field (inverse alone) amplitude attenuation. This attenuation is added in real time to the streaming audio, although the attenuation coefficients are precomputed for each panning point once the source's motion path is defined. For air absorption filtering, following the suggestion of Huopaniemi (*9*), the yulewalk function from the MATLAB Signal Processing Toolbox (*10*) was implemented to reverse-engineer a one-pole filter based on known frequency-based attenuation data. Using the ANSI standard (*11*), we altered the system to find attenuation by octave bands as a function of distance and humidity, interpolating between known values if necessary. The filter coefficients are precomputed for the source path, but the filter can be applied to each 11.5 ms "chunk" of audio in real time without causing an underflow.

5. Discussion

The method outlined in the previous section incorporates streaming audio to allow low-latency virtual source motion in a large spherical loudspeaker array. DBAP and VBAP are both implemented, although VBAP is only needed for large-scale simulations. This method can be used with high-quality recordings to reproduce the exact path taken by an object recorded in motion. Amplitude attenuation and air absorption filtering have been added to allow flexible control of arbitrary source signals.

The drawback of this approach is that it makes sense for only signals whose audio content does not change based on their movement. While this might be the case for a person talking while walking at a low speed or even a white noise burst, which is fairly abstract in the first place, it does not really make sense for recordings of vehicles, whose engines' outputs vary greatly with velocity, not to mention the difficulty of modeling sound radiation patterns and source orientations. It is also difficult to obtain high-quality recordings of vehicles moving at a constant velocity, since the microphone is usually contained within the vehicle rather than outside it. An outside recording of a vehicle's engine idling may be used with this method to obtain variable results, but such a simulation will not sound convincing to an experienced listener. The inclusion of a flexible Doppler shift based on the source's relative velocity will be a future goal, although as mentioned previously, the sources that move fast enough to create a significant Doppler shift would usually be used in functions tailored specifically to the motion path of the recorded source. For slower-moving objects, the Doppler shift would have a negligible effect on the perception of the moving sound source.

6. References

1. Henry, P.; Amrein, B.; Ericson, M. The Environment for Auditory Research. *Acoustics Today* **2009**, *5* (3), 9–16.
2. Chowning, J. The Simulation of Moving Sources. *J. Audio Eng. Soc.* **1971**, *19*, 2–6.
3. Kleiner, M.; Brainard, D.; Pelli, D. What's New in Psychtoolbox-3? *Perception* [Online] **2007,** *36*, ECVP Abstract Supplement. http://www.perceptionweb.com/abstract.cgi?id=v070821 (accessed July 27, 2011).
4. Kostadinov, D.; Reiss, J. Spatial Audio MATLAB Toolbox, Centre for Digital Music, Queen Mary University of London, 2009. http://www.isophonics.net/content/spatial-audio-matlab-toolbox (accessed July 27, 2011).
5. Lossius, T.; Baltazar, P.; de la Hogue, T. DBAP - Distance-Based Amplitude Panning. *Proceedings of the 2009 International Computer Music Conference*, 2009.
6. Kostadinov, D.; Reiss, J.; Mladenov, V. Evaluation of Distance Based Amplitude Panning for Spatial Audio. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010.
7. Pulkki, V. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *J. Audio Eng. Soc*. **1997**, *45*, 456–466.
8. Sunday, D. Intersections of Rays, Segments, Planes and Triangles in 3D, 2003. http://www.softsurfer.com/Archive/algorithm_0105 (accessed July 27, 2011).
9. Huopaniemi, J.; Savioja, L.; Karjalainen, M. Modeling of Reflections and Air Absorption in Acoustical Spaces - A Digital Filter Design Approach. *IEEE 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, 1997.
10. MATLAB, Signal Processing Toolbox, MathWorks Inc. http://www.mathworks.com/products/signal/ (accessed July 27, 2011).
11. American National Standard (ANS) 21.26-1995, Method for Calculation of the Absorption of Sound by the Atmosphere, 1995.

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Weighted Phase Lag Index (WPLI) as a Method for Identifying Task-Related Functional Networks in Electroencephalography (EEG) Recordings during a Shooting Task

SANDHYA RAWAL
MENTOR: JEAN M. VETTEL, PH.D.
TRANSLATIONAL NEUROSCIENCE BRANCH
HUMAN RESEARCH AND ENGINEERING DIRECTORATE
ABERDEEN PROVING GROUND, MARYLAND

Contents

List of Figures 55

Abstract 56

Acknowledgments 57

Student Bio 58

1. Introduction/Background 59

2. Experiment/Calculations 60

2.1 Participants 60

2.2 EEG Acquisition 60

2.3 Task and Event Extraction 60

2.4 EEG Pre-processing 61

2.5 Analysis 61

3. Results and Discussion 61

4. Summary and Conclusions 66

5. References 67

List of Figures

Figure 1. (a) The map of electrodes depicted highlights the electrodes with significance in black. (b) The map of electrodes depicted highlights the nonzero phase lag statistical interdependencies identified by WPLI with black lines between the interdependent electrodes. The location of each electrode is the same as in figure 1(a).....63

Figure 2. (a) The WPLI analysis depicted was conducted in the theta band, 4–8 Hz, on the Fz and CP3 electrode. Thirty significant points were found from 290–0 ms before the trigger pull. (b)The WPLI analysis depicted was conducted in the theta band, 4–8 Hz, on the FC3 and Pz electrode. Thirty significant points were found from 2990–2700 ms before the trigger pull. (c) The WPLI analysis depicted was conducted in the delta band, 1–4 Hz, on the F8 and P02 electrode. Thirty-two significant points were found from 2,910–2,600 ms before the trigger pull. (d) The WPLI analysis depicted was conducted in the theta band, 4–8 Hz, on the C3 and Pz electrode. Sixty-four significant points were found from 2,990–2,360 ms before the trigger pull.....65

Abstract

Weighted Phase Lag Index (WPLI) is a methodology used to identify nonzero phase lag statistical interdependencies between electroencephalography (EEG) time series from pairs of electrodes. Identifying nonzero phase lags may be useful to identify neural interaction among regions based on the known delay for brain region-to-region communication. This project applies WPLI analysis to previously collected EEG data from a Soldier performing a shooting task. WPLI was examined on EEG data epoched around trigger pull events, which includes both brain activity and movement artifacts from weapon recoil. This functional connectivity measure was compared to a traditional time-frequency analysis at individual electrode sites. In the individual studied, WPLI identified a left lateralization in the network communication as well as neural activity from the occipital electrodes. Both of these findings were not evident in the channel-based analysis. This project suggests that WPLI may be able to detect task-related functional networks, despite artifactual contamination that is typical in real-world environments.

Acknowledgments

I wish to acknowledge the mentorship of Dr. Jean Vettel, Dr. Kaleb McDowell, and Dr. Scott Kerick. I also wish to acknowledge the collaboration of Samantha Luo.

Student Bio

I just completed my junior year at the Massachusetts Institute of Technology in Cambridge, MA. I am currently studying Brain and Cognitive Sciences and Management Sciences with a minor in Economics. I have previously worked at the Columbia University in the Program for Imaging and Cognitive Sciences as a summer intern. There, I ran Volume Based Morphometry, an application of Statistical Parametric Mapping that was new to the Hirsch lab. I specifically used the program to contribute measurements of cortical volume to the project that studied the cognitive basis of social anxiety disorder. This effort resulted in second authorship on a presentation entitled "Regional differences in brain volume in panic and social anxiety disorder" that was presented at the conference of American College of Neuropsychopharmacology, 2009. I am a Science, Mathematics & Research for Transformation (SMART) scholar and look forward to joining the U.S. Army Research Laboratory (ARL), Human Research and Engineering Directorate (HRED) as a fulltime employee in the Translational Neuroscience Branch in the summer of 2012.

1. Introduction/Background

Electroencephalography (EEG) records electrical activity from the brain by placing electrodes along the scalp, and the study of this activity through EEG is a prevalent methodology used in neuroscience. EEG has several advantages over other prevalent methodologies, in that it can be used outside of the laboratory and has a high temporal resolution. EEG data have been analyzed at individual electrode sites to identify patterns of brain activity when an event occurs. However, the brain relies on communication between brain regions, so developing analysis methods to investigate brain activity between regions may provide additional information over analyzing electrical activity at single locations. Specifically, identifying the communication between brain regions that occurs during tasks may provide information regarding the cognitive processes involved in the tasks. The brain regions and their connectivity involved in this communication during a task are referred to as the task-related functional network.

One approach to identify task-related functional networks is to investigate the synchronization of neural activity between different brain regions. A common approach calculates statistical interdependencies between electrical activity recorded from pairs of electrodes. Each electrode can capture electrical data from different local regions of the brain, so the interdependencies between two electrodes may suggest neural communication. Although neural communication is fast, occurring in milliseconds, EEG has a high temporal resolution, making it a suitable method for identifying communication among brain regions. However, EEG often has a low signal-to-noise ratio based on movement and muscle artifacts. Another phenomenon that contributes to the low signal-to-noise ratio is often referred to as the volume conduction problem, when two nearby electrodes tend to pick up activity from the same source, giving rise to false correlations between the time series.

To address these problems, Stam and colleagues (*1*) introduced a measure called Phase Lag Index (PLI) to address volume conduction issues. PLI is based upon the idea that nonzero phase lag between two time series is not caused by volume conduction from a single source. Therefore, it may capture the communication between brain regions, as these communications have a delay, or a nonzero phase lag due to axonal transmission properties. The delay exists because brain regions must communicate through a physical biological medium, neurons, and thus cannot communicate instantly. PLI is less affected by common source problems than the traditional measures of functional connectivity (*1*).

However, according to Vinck and colleagues (*2*), PLI is limited by the discontinuity of phase lag estimates between zero and infinitesimally small non-zero phase lags. They claimed that both the sensitivity to noise, including volume conduction, and the capacity to detect changes in phase synchronization is hindered by the discontinuity of PLI. To improve upon PLI, the Weighted Phase Lag Index (WPLI) was introduced by Vinck and colleagues (*2*). In WPLI, the contribution

of the observed phase leads and lags is weighted by the magnitude of the imaginary component of the cross-spectrum (*3*). Vinck and colleagues (*2*) demonstrated two advantages of WPLI over PLI, reduced sensitivity to noise sources and increased power to detect changes in phase synchronization.

The advantage offered by WPLI of eliminating noise in EEG data is particularly useful for Army purposes because it allows for future studies and product development based on noisy EEG data that would be encountered in real-life situations outside of the laboratory. Artifacts introduced in real-world settings seriously exacerbate the low signal-to-noise ratio, but the ability to identify cognitive activity within the noisy data may allow for discovery of cognitive processes in real-world settings. Identifying nonzero phase lag statistical interdependencies via WPLI may allow for the identification of this cognitive activity. WPLI, therefore, presents itself as a methodology to provide functional connectivity data in studies performed in real-world conditions.

Accordingly, this project applies WPLI analysis to previously collected EEG data from a Marine performing a shooting task. WPLI was computed on epochs of data surrounding the trigger pull event, which consists of both brain activity and movement artifacts from the weapon recoil. This project evaluates the use of WPLI as a method to provide functional connectivity data that identifies task-related cognitive activity not identified using traditional time-frequency analyses.

2. Experiment/Calculations

2.1 Participants

In the original study (*3*), 18 participants volunteered, and 17 of the participants were U.S. Marines and 1 was a U.S. Army Ranger. All had completed basic training and shooting qualifications. The data collected from 1 participant of the 18 was selected for analysis.

2.2 EEG Acquisition

Data were previously collected by Kerick and colleagues (*3*). Data were acquired at 500 Hz using a Neuroscan system with 34 active electrodes, including 2 horizontal and 2 vertical eye channels and 2 mastoid reference channels. These analyses use the 28 non-reference electrodes.

2.3 Task and Event Extraction

The EEG data analyzed in this paper were acquired from the participants during a simulated shooting task. The original study by Kerick and colleagues (*3*) involved eight shooting conditions. The conditions included single and dual task conditions in which the Soldier was required to perform only a shooting task or a shooting task while performing an arithmetic task; with short (2–4 s) and long target (4–6 s) exposure, and all enemy targets or mixed enemy and friendly targets (*3*) . However, this paper investigates only the condition that had a single task load, all enemy targets, and a long target exposure.

The shooting scenarios were a simulation of an outdoor shooting range at Aberdeen Proving Ground, MD, and the rifle provided to the participant was demilitarized to emit a laser beam for fire. The rifle also simulated the recoil that follows an actual shot in order to imitate the type of artifact obtained in a real-world shooting scenario. A digital terminal board sent event markers, including onset and offset times of targets, trigger responses, and target hits and misses, from the simulation system to the EEG recording system. Since all targets were enemy targets, no decision to shoot was necessary. The participant had to search for the enemy target, orient the weapon towards the target, aim, and pull the trigger.

Thirty trials of the event type in which participants aimed and shot successfully were extracted by epoching the data around the event from –3 to 2 s, where 0 s corresponds to the event marker from the digital terminal board indicating that a shot was fired at the enemy target.

2.4 EEG Pre-processing

All pre-processing, including bandpass filtering and some data rejection, was performed and described by Kerick and colleagues (*3*). EEGLab scripts were used to automatically remove bad channels. Independent components (ICs) were obtained from independent component analysis (ICA) decomposition. Those with eye movements or muscle artifacts were removed. The channel data without the eye movement and muscle artifact ICs were then analyzed.

2.5 Analysis

WPLI and time-frequency data in six frequency bands, delta through gamma, were calculated. The derived WPLI and time-frequency was extracted in the epochs surrounding the trigger pull (–3 s to 2 s). For each electrode (pair) and frequency band, significance was determined by comparing each time point (–3 to 0 s) across all 30 shots to a baseline determined by averaging across the whole data epoch. The significance level was obtained by dividing 0.05 by 501 ($p < 9.98003992 \times 10^{-5}$) to account for the chance occurrence of a significant test (Type I error). To further correct for the Type I error, at least five consecutive time points had to be significant in order to be considered a significant point grouping for WPLI and at least 25 consecutive time points were required for significance in the time-frequency analysis. The requirement for consecutive time points for significance is based on the estimated time course of communications between brain regions.

3. Results and Discussion

WPLI and time-frequency measures were calculated from EEG data acquired from a Soldier during performance of a shooting task. This project attempts to identify task-related functional networks to further understand cognitive processes that occur before the trigger pull; however, the movement artifact caused by the simulated recoil makes the signal-to-noise ratio low. A

technique that is impervious to artifacts of this type may allow for identification of cognitive processes occurring concurrently with the movement artifact. In this project, we investigate WPLI to determine if task-related functional networks can be identified after removing the zero phase lag statistical interdependencies between pairs of channels. A traditional time-frequency analysis at individual channel locations is used as a point of comparison.

The channels that were identified as significant in time-frequency data are indeed different from the ones identified in WPLI. Twenty-seven groups of points were significant prior to the trigger pull in the 28 electrodes and 6 frequency bands studied using time-frequency analysis. Fifty-eight groups of points were significant in the 378 electrode pairings in the 6 frequency bands studied using WPLI analysis. In figure 1, the results are plotted on a schematic view of the electrodes on the head, where the front of the brain is located at the top of the graph (Fz). In (a), all electrodes are present and labeled on the figure in red. The Fz electrode is marked by a larger red circle. The significant channels that were identified in time-frequency data are highlighted in black. Figure 1(b) illustrates the significant channel pairs identified by WPLI with black lines between the independent electrodes. The location of each electrode in (b) is the same as in (a).

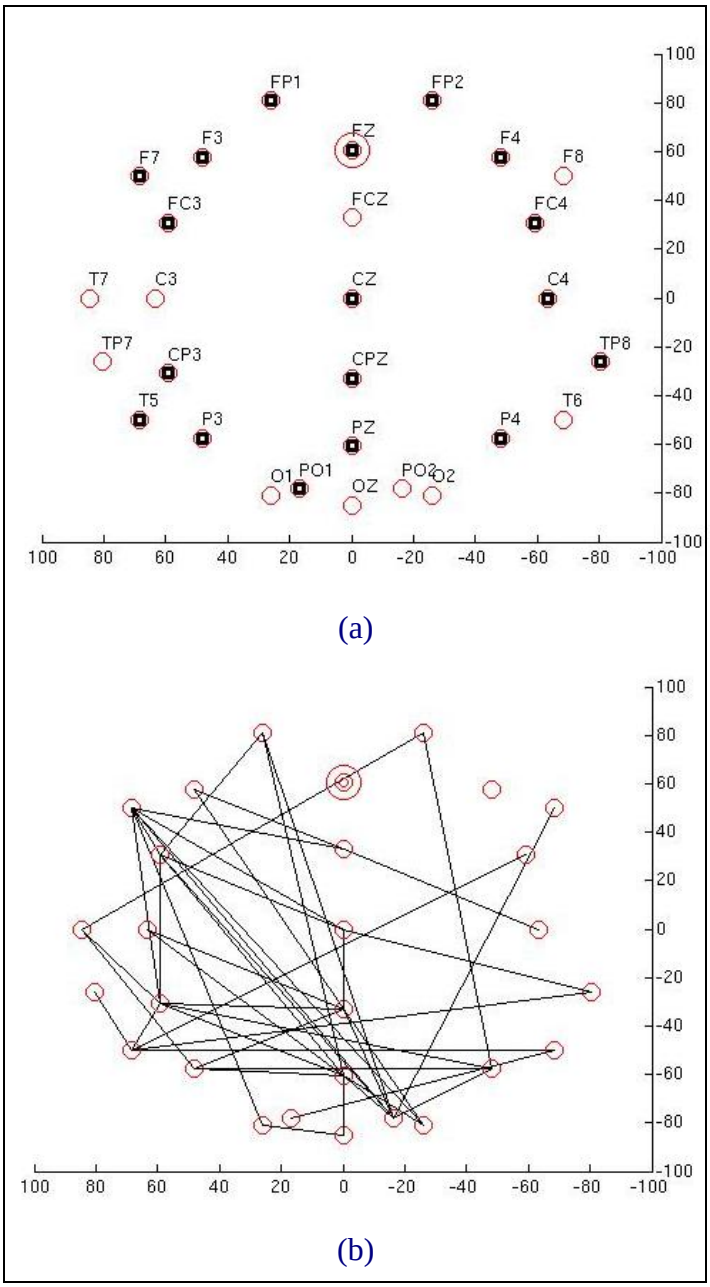

Figure 1. (a) The map of electrodes depicted highlights the electrodes with significance in black. (b) The map of electrodes depicted highlights the nonzero phase lag statistical interdependencies identified by WPLI with black lines between the interdependent electrodes. The location of each electrode is the same as in figure 1(a).

The differences between figure 1(a) and (b) highlight the additional information provided by the WPLI measure. The significance of individual channel locations indicated by the electrodes highlighted in black in figure 1(a) appears to be distributed symmetrically. However, the distribution in the WPLI analysis in figure 1(b) is more heavily localized to the left hemisphere. Interestingly, the WPLI analysis reveals a left lateralization of the paired network communication among brain regions, providing an insight not captured in the channel analysis.

The involvement of the frontal motor cortex and the parietal cortex suggests that the cognitive activity identified may be motor-spatial planning, as the frontal cortex is involved in planning and motion and the parietal cortex integrates sensory information from different modalities, particularly for spatial sense. The increased interdependencies in the left hemisphere is consistent with previous literature, as left temporal alpha power is consistently observed in marksmen few seconds before the trigger pull (*1*).

In addition to the lateralization, the WPLI analysis also revealed activity in the occipital electrodes not found in the channel-based analysis. All occipital electrodes show interaction with the F7 electrode, suggesting communication between the visual cortex and the frontal cortex. This communication may reflect that the shooting task requires planning in response to a visual target before the trigger pull. However, the possibility that communication between the occipital electrodes and the F7 electrode reflect the eye movements occurring near the F7 electrode must be addressed in further research, as this paper does not address the potential issue of brain activity correlated to muscle activity.

Figure 1 highlights the promise of WPLI to reveal network communication among brain regions. The plots show all pairs of electrodes that are statistically significant point groupings, but we were also interested in what pairs reveal an extended window of communication during the 3 s before the trigger pull. As a preliminary exploration, we plotted the four longest groups of significant periods, which are illustrated in figure 2. The blue line represents the WPLI waveform and the black lines enclosing the blue line represent the confidence interval.

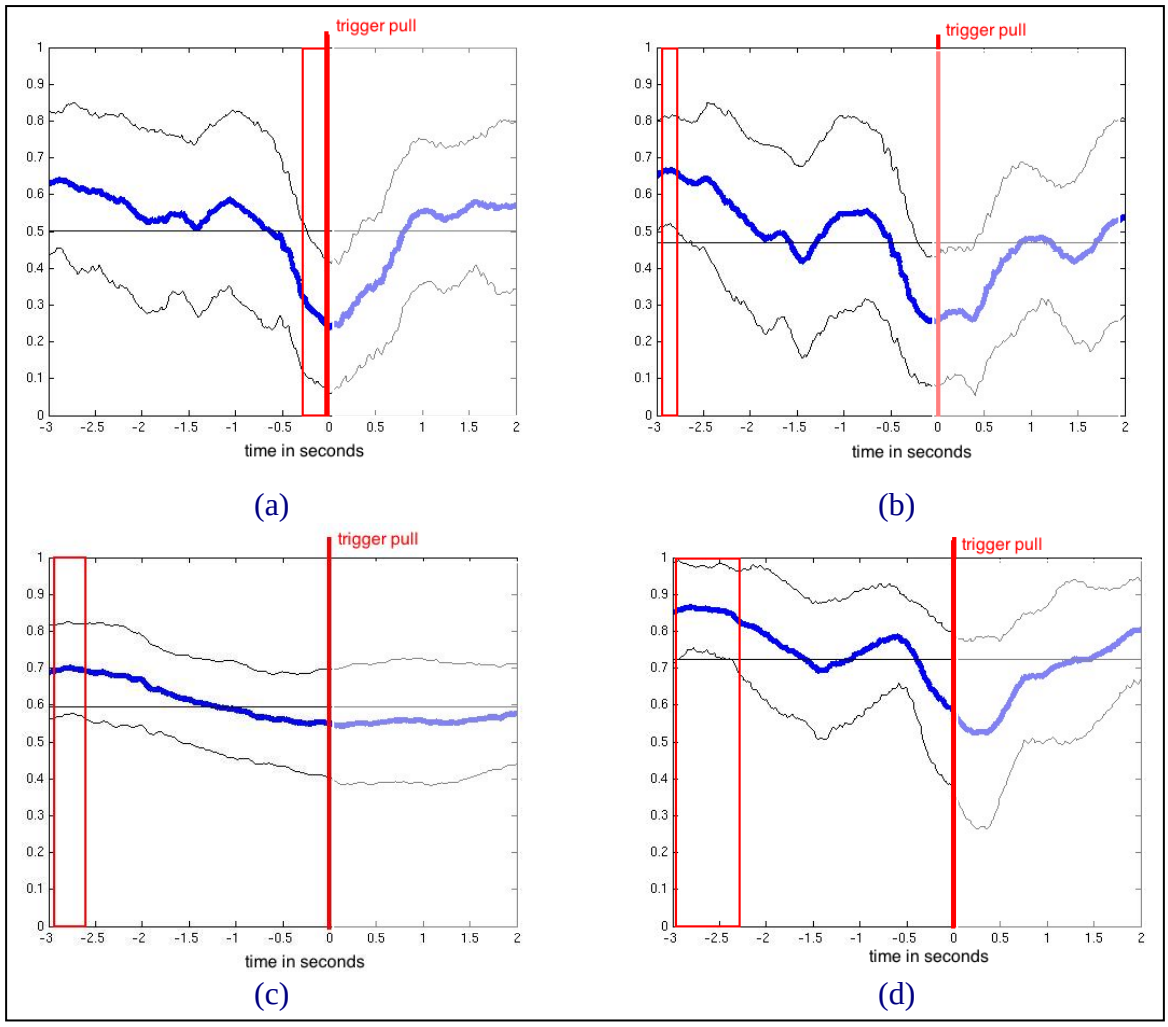

Figure 2. (a) The WPLI analysis depicted was conducted in the theta band, 4–8 Hz, on the Fz and CP3 electrode. Thirty significant points were found from 290–0 ms before the trigger pull. (b)The WPLI analysis depicted was conducted in the theta band, 4–8 Hz, on the FC3 and Pz electrode. Thirty significant points were found from 2990–2700 ms before the trigger pull. (c) The WPLI analysis depicted was conducted in the delta band, 1–4 Hz, on the F8 and P02 electrode. Thirty-two significant points were found from 2,910–2,600 ms before the trigger pull. (d) The WPLI analysis depicted was conducted in the theta band, 4–8 Hz, on the C3 and Pz electrode. Sixty-four significant points were found from 2,990–2,360 ms before the trigger pull.

All four pairs included a frontal and parietal electrode (Fz and CP3, FC3 and Pz, F8 and P02, and C3 and Pz). All waveforms follow a similar downward trend from 3 s before the trigger pull to the trigger pull. However, the first pair (2a) contains a significant point grouping immediately before the trigger pull, while the other three pairs (2b–d) contain significant groupings around 3 s before the trigger pull. Another similarity is the frequency bands in which the significant point groupings occur, as three of the four longest groups occur in the theta (4–8 Hz) band. Figure 2(c) is the only significant point grouping that occurs in the delta (1–4 Hz) band. Further analyses are needed to clarify the significance of these pairs, although these particular pairs reflect the general pattern of communication between parietal and frontal regions shown in figure 1.

There are several ways to improve on these preliminary results. Primarily, all 18 participants in the study must be analyzed. The patterns observed in the single participant strongly support the extension of these analyses to all participants. Secondly, the statistical approach may have been suboptimal. As was discussed regarding figure 2, all four waveforms followed a generally similar pattern, yet the statistical approach revealed significance at different times relative to the trigger pull. Alternative approaches conducted in all 18 participants may reveal a clearer result. Finally, this paper does not address the potential issue of brain activity that is correlated to muscle activity, which may be the case for the activity of the F7 electrode due to the electrode's proximity to the eye. However, further research is necessary.

4. Summary and Conclusions

WPLI analysis suggests that the measure may be able to capture task-related functional networks during cognitive tasks, even in the midst of large-scale movement artifacts. Further research looking across more individuals and at more tasks is needed to provide more definitive results. However, the results from this one participant support that conducting pair-wise electrode analysis in addition to individual electrode analysis may provide more information than individual electrode analysis alone. While EEG time-frequency reveals a symmetric distribution of electrodes with significance, WPLI analysis finds statistical interdependencies more heavily lateralized to the left hemisphere. In addition, WPLI reveals communication with occipital electrodes, whereas these individual channels are not marked significant in the time-frequency analysis. The preliminary findings reported here suggest that the F7 electrode may be a hub of communication, with several significant pairings with other electrodes. Therefore, the methodology used in this project shows that WPLI may provide functional connectivity data not accessible in time-frequency data. WPLI provides an interesting method for investigating how the brain performs cognitive tasks even with large artifacts in the signal, allowing for further investigation and understanding of Army tasks performed in complex environments.

5. References

1. Stam, C. J.; Nolte, G.; Daffertschofer, A. Phase Lag Index: Assessment of Functional Connectivity from Multi Channel EEG and MEG with Diminished Bias from Common Sources, Hum. *Brain Mapp.* **2007**, *28* (11), 1178–1193.
2. Vinck, M.; Oostenveld, R.; van Wingerden, M.; Battalglia, F.; Pennartz C.M.A., 2011: An Improved Index of Phase-Synchronization for Electrophysiological Data in the Presence of Volume-Conduction, Noise, and Sample-Size Bias, *NeuroImage* **2011**, *55* (4), 1548–156.
3. Kerick, S. E.; Hatfield, B. D.; Allender, L. E. Event-Related Cortical Dynamics of Soldiers During Shooting as a Function of Varied Task Demand, *Aviation, Space, and Environmental Medicine* **2007**, *78* (5), B153–B164.

INTENTIONALLY LEFT BLANK.

Sensors and Electron Devices Directorate (SEDD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Growth and Transfer of Graphene for Device Fabrication

KEVIN HAURI
MENTOR: EUGENE ZAKAR
ELECTRONICS & RF DIVISION
ADELPHI, MD

Contents

List of Figures **73**

List of Tables **73**

Abstract **74**

Acknowledgments **75**

Student Bio **76**

1. Introduction/Background **77**

2. Experiment **77**

2.1 Graphene Growth 77

2.2 Graphene Transfer 78

3. Results and Discussion **79**

3.1 Graphene Growth 79

3.2 Graphene Transfer 81

4. Summary and Conclusions **81**

5. References **83**

List of Figures

Figure 1. The graphene growth process involves the following: I. Ramp up phase to 1000 °C with Ar and H_2 gases; II. Anneal phase with Ar and H_2 gases for grain growth; III. Growth phase, where CH_4 is introduced into the system; and IV. Ramp Down phase at 5 °C/min with continued exposure to Ar, H_2, and CH_4. 78

Figure 2. The graphene transfer process off a nickel-based substrate involves undercut etching of the SiO_2 layer, then of the nickel layer to float the graphene off the surface of the substrate. The new substrate of choice is then used to manually scoop the graphene film onto its surface. 79

Figure 3. Graphene samples were grown at 975 °C with 800 sccm Ar and 200 sccm H_2 gas flows and 5 °C/min ramp down. The methane content was decreased, and smoother graphene with fewer multilayer patches was grown as methane decreased. 80

Figure 4. Raman spectroscopy data peaks correspond to spots tested. Higher intensity peaks occurred at points where there were darker patches of graphene. (Data provided by Center for Advanced Material Characterization at the University of Oregon.).....80

Figure 5. Graphene is pictured before and after transfer from the nickel-based substrate of grown to an oxidized silicon substrate. SiO_2 is an ideal substrate because of its insulator properties and because graphene is visible on its surface.....81

List of Tables

Table 1. Interpretation of Raman spectroscopy data approximates the number of graphene layers on a nickel sample (*6*). 80

Abstract

The unique electronic properties of graphene enable the development of field effect transistors with potential for speeds much higher than the existing technology. Diluted methane chemical vapor deposition growth of graphene on a nickel substrate has resulted in potentially useful multilayer graphene layers. These graphene layers were grown on nickel substrates deposited onto silicon (Si)/silicon dioxide (SiO_2) wafers by evaporation and sputter methods. The graphene must be transferred from the nickel substrate onto a template more suitable for device fabrication. However, the current graphene transfer process is not fully mature and presents a number of challenges. In the transfer process, a polymer coating is applied to the graphene, and the original materials underneath are removed. A coating of photoresist is spun and hardened on top of the graphene, and then the substrate is removed with a buffered oxide etch and a nickel etch. One challenge is removal of the photoresist from the surface of the graphene without film wrinkles, cracks, and contamination. Alternative solutions that minimize damage to the graphene during the transfer process are currently being explored.

Acknowledgments

Special thanks go to Eugene Zakar, Barbara Nichols, and Greg Meissner for guiding my research.

Student Bio

I am a currently in my first year as a Student Temporary Employment Program (STEP) student at the U.S. Army Research Laboratory (ARL) in Adelphi, MD. I will be entering my second year as an undergraduate in Materials Science and Engineering at the University of Maryland this fall. My research at ARL has involved processing work within the cleanroom (sputtering, etching, and characterization) and optimization of the graphene growth and transfer processes. I plan on continuing my work at ARL in the future to fully optimize these processes and develop skills to characterize graphene using different methods.

1. Introduction/Background

Graphene, an atomically thick layer of carbon, has recently emerged as a material with the potential to greatly improve semiconductor devices. The zero bandgap width of graphene enables high electron and hole concentrations, which gives a high field effect mobility of about 10,000 cm^2/Vs (*1*). While graphene is most commonly referred to as a single layer, bilayer and multilayer graphene up to about 10 layers in thickness exhibit graphene-like properties; 10 or more layers tend to exhibit bulk graphite characteristics (*2*). Bilayer graphene is unique because it can form a bandgap between the two atomically thin layers (*3*). Large sheets of single layer graphene have already been produced through exfoliation methods (*4*), but the exfoliation process is limited. Exfoliation involves cleaving graphene from bulk graphite a single atomic layer at a time, which limits production to single-layer graphene. Additionally, the film produced is not uniform because it conforms to the roughness of the graphite sample being cleaved.

However, the growth of graphene using atmospheric pressure chemical vapor deposition (APCVD) is promising. APCVD graphene growth involves flowing diluted methane gas over a nickel sample in an isolated environment at about 1000 °C. The nickel samples are either sputtered or evaporated onto silicon (Si)/silicon dioxide (SiO_2) wafers. Nickel is an ideal substrate for graphene growth because it is approximately 0.9% with carbon (*5*), which causes the nickel to form a solution with the carbon it is exposed to from the diluted methane gas. When this solution is cooled at a controlled rate, the carbon segregates itself from the nickel substrate and forms a layer of graphene on the surface of the nickel. The challenges presented in APCVD graphene growth are optimizing the process to limit the number of graphene layers grown and minimizing defects and pitting caused by hydrogen gas exposed to the sample. The goal is to work toward a process that gives consistent smooth areas of bilayer graphene.

Before this graphene can be used for device fabrication, it is necessary to transfer the graphene onto a substrate with insulator properties. The current transfer process is still preliminary and presents a number of challenges. Since the process is a manual wet etch of the nickel substrate the graphene was grown on, the graphene can easily become contaminated, wrinkled, or torn. The process is currently being optimized to produce the best possible quality graphene.

2. Experiment

2.1 Graphene Growth

The graphene growth process (figure 1) requires the nickel substrate to be in an isolated environment at atmospheric pressure. Once isolated, the system is heated to about 1000 °C

while exposed to argon (Ar) and hydrogen (H_2) gas. The argon acts as a carrier gas, while the hydrogen prevents oxidation of the nickel at high temperature. Once the system is ramped to temperature, the sample is annealed while exposed to argon and hydrogen. After the anneal period, methane gas (CH_4) is introduced into the system and temperature is held for about 10 min while the carbon forms a solution with the nickel substrate. After this growth period, the system is ramped down to room temperature at a rate of 5 °C/min, which allows the carbon to segregate from the nickel sample and form graphene on top of the nickel.

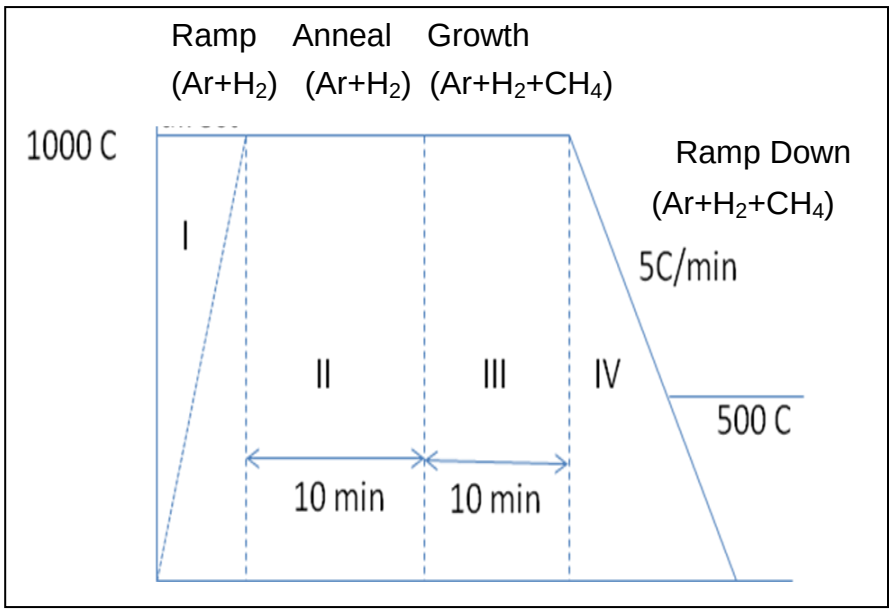

Figure 1. The graphene growth process involves the following: I. Ramp up phase to 1000 °C with Ar and H_2 gases; II. Anneal phase with Ar and H_2 gases for grain growth; III. Growth phase, where CH_4 is introduced into the system; and IV. Ramp Down phase at 5 °C/min with continued exposure to Ar, H_2, and CH_4.

In order to control the number of graphene layers grown on the sample, the methane content was adjusted in search of an optimum value. Graphene was grown at 975 °C with the gas content held at 800 sccm Ar and 200 sccm H_2 and a 5 °C/min ramp down. Four different runs were performed with the previous conditions held constant, but with varying methane content ranging from 5 to 2 sccm. In order to find conditions for graphene growth that minimized defects from hydrogen, graphene was grown in a series of experiments that used ranges of hydrogen from 300 to 700 sccm. Growth was also attempted at temperatures ranging from 975 to 900 to 800 °C.

2.2 Graphene Transfer

Once graphene suitable for device fabrication is produced, the graphene is transferred onto a substrate suitable for device fabrication by a wet etch process (figure 2). First, a layer of poly(methyl methacrylate) (PMMA) is manually spun on the surface of the nickel-graphene sample to protect the graphene layer. After the PMMA is spun on the surface, the sample is floated in a BOE for 1 h and rinsed. The sample is then floated in a nickel etchant until the

nickel layer is visibly removed. After a float rinse in water, a new oxidized Si substrate is used to scoop the layer of graphene and PMMA out of the water. This substrate was used because graphene is visible on the surface of SiO_2. After baking the graphene onto the Si/SiO_2 substrate, an acetone vaporization method was used to removed the PMMA from the graphene's surface. By heating acetone to 110 °C and submerging the sample in the acetone four separate times, a clean graphene sample on SiO_2 is obtained.

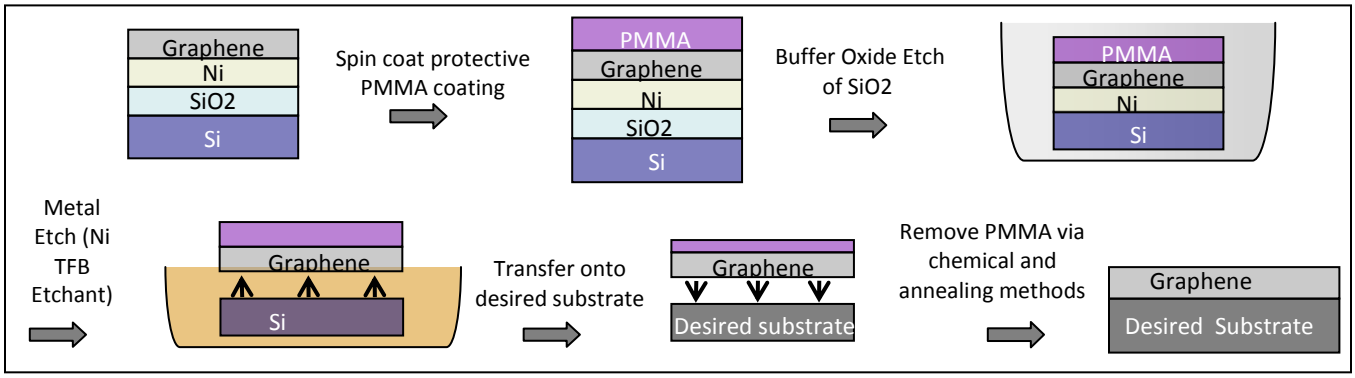

Figure 2. The graphene transfer process off a nickel-based substrate involves undercut etching of the SiO_2 layer, then of the nickel layer to float the graphene off the surface of the substrate. The new substrate of choice is then used to manually scoop the graphene film onto its surface.

3. Results and Discussion

3.1 Graphene Growth

When graphene was grown at different ranges of methane content from 0.5% to 0.2%, it was found that decreased methane content caused less graphene to grow (figure 3). GPH100, the sample grown with 0.3% methane content, was tested using Raman spectroscopy to characterize the approximate number of layers in the sample (figure 4). It was found that darker patches on the sample surface were found to be thick multilayer graphene, lighter patches were approximated to be about five layers, and transparent spots were few layers or potentially bilayer (table 1). Since as methane content was decreased, graphene samples displayed few dark and light patches and more consistently smooth transparent patches, the interpretation of the Raman spectroscopy data shows as methane content was decreased, multilayer graphene patches also decreased and larger areas of smooth few-layer graphene increased.

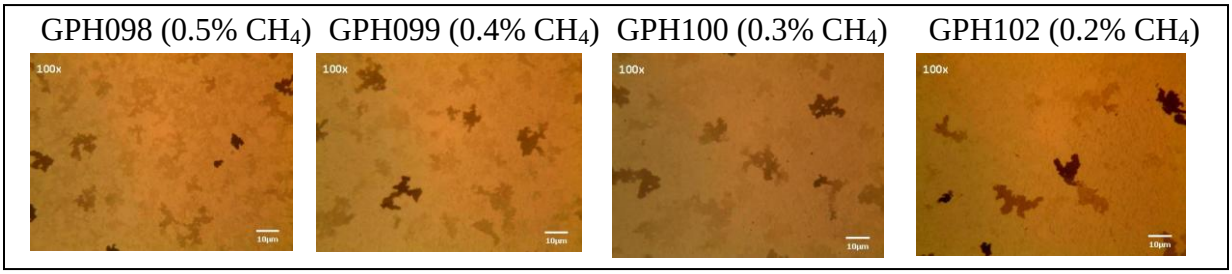

Figure 3. Graphene samples were grown at 975 °C with 800 sccm Ar and 200 sccm H_2 gas flows and 5 °C/min ramp down. The methane content was decreased, and smoother graphene with fewer multilayer patches was grown as methane decreased.

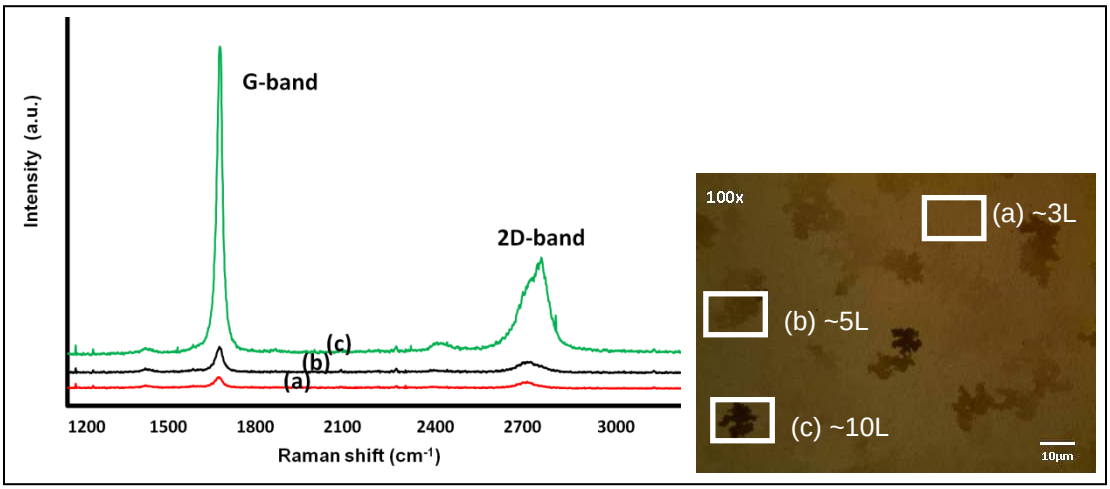

Figure 4. Raman spectroscopy data peaks correspond to spots tested. Higher intensity peaks occurred at points where there were darker patches of graphene. (Data provided by Center for Advanced Material Characterization at the University of Oregon.)

Table 1. Interpretation of Raman spectroscopy data approximates the number of graphene layers on a nickel sample (*6*).

Raman Spectroscopy Data Interpretation

Ratio (G/2D)	Graphene Layers	2-D Peak
<0.5	Single Layer	Sharp
~1	Bi-Layer	Broad
~3	3 Layers	Broad
~5	Multi-Layer	Broad

When testing growth with differing ranges of hydrogen, it was found that as the hydrogen content increased at temperatures of ~1000 °C, the defects increased in number and became larger. However, it was found that at decreased temperature more hydrogen could be used without causing defects to the nickel surface. At 900 °C, 400 sccm of hydrogen was effectively used to aid in graphene growth, while this same flow of hydrogen caused defects at 975 °C. By lowering the temperature at which graphene growth takes place, hydrogen defects were successfully eliminated.

3.2 Graphene Transfer

By following the previous graphene transfer process, graphene was successfully transferred onto a Si/SiO_2 substrate (figure 5). However, the process is still preliminary and presented a number of difficulties. Since the PMMA is spun on manually, it actually attaches to the sides of the sample and prevents the buffered oxide etchant from removing the SiO_2 layer. This problem was combated by manually cleaving the PMMA from the sides of the sample with a razor, and then using acetone to clean the sides of the sample. This helped the BOE etch more quickly, but also caused tears in the sides of the graphene.

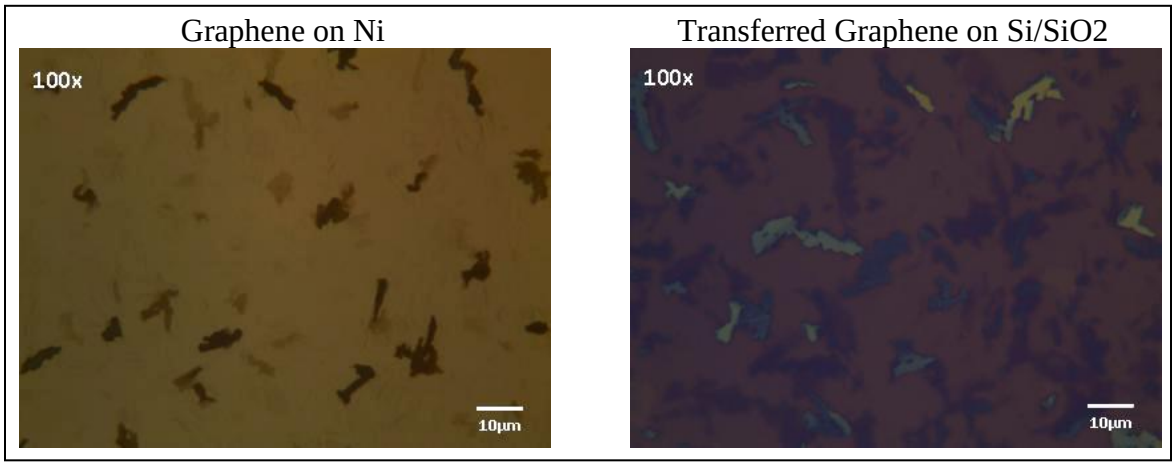

Figure 5. Graphene is pictured before and after transfer from the nickel-based substrate of grown to an oxidized silicon substrate. SiO_2 is an ideal substrate because of its insulator properties and because graphene is visible on its surface.

The BOE was able to etch part of the SiO_2 layer, but it was not particularly effective because the nickel etchant took multiple days to completely release the graphene from the surface of the nickel. Even when the nickel was completely etched, the graphene was still partially attached to the surface of the substrate it was grown on and had to be removed with a razor. Although the graphene was successfully transferred onto the Si/SiO_2 substrate, the process was not completely optimized to prevent damage to the graphene.

4. Summary and Conclusions

By varying the methane and hydrogen content in a series of graphene growth experiments, optimum conditions for temperature and methane and hydrogen content were found. By reducing the methane to the lowest possible content at which growth still occurs and lowering the temperature to 900 °C to eliminate hydrogen pitting, larger areas of smooth few-layer graphene was grown with fewer defects. Graphene samples grown at 900 °C with 0.3% methane content exhibited surfaces suitable for nanoscale devices. More work must be done to continue

to optimize different variables in the growth process and further reduce the number of layers grown, with bilayer graphene being the goal.

In order to optimize the transfer process, it is important to find a successful etchant for the SiO_2 layer of the nickel-graphene substrate. Since the PMMA on the sides of the sample, even when manually cleaved, seemed to hinder the BOE from etching, future plans for optimizing the graphene transfer process involved the use of a different etchant for the SiO_2 layer. Once the SiO_2 layer can be successfully etched, the nickel etchant will be able to fully remove the nickel layer so the graphene can lift off the surface and enable transfer that minimizes tears and defects in the graphene. Future efforts will include the use of a 49% HF solution and HF vaporization to attempt to better etch the SiO_2 layer.

Graphene's advantages in electrical mobility and its unique electronic properties show promise for eventual industrial-scale device production. In order to reach the ability to mass produce graphene devices, it is important to create a process that is easily executable and produces quality results. Through the optimization of temperature and gas content in the growth of graphene and the successful transfer of graphene onto a Si/SiO_2 substrate, the process of APCVD graphene production has evolved to give graphene with fewer layers and better quality. By finding graphene layers can be reduced through controlling methane content, this process can be applied with optimized hydrogen flow and temperature to form nickel-grown graphene with the fewest layers possible. Through successful transfer of this graphene, device fabrication becomes practical with few-layer graphene grown on a nickel substrate.

5. References

1. Novoselov, K. S. et al. Electric Field Effect in Atomically Thin Carbon Films. *Science* **2004**, *306*, 1095–9203.
2. Geim, A. K.; Novoselov, K. S. The Rise of Graphene. *Nature Materials* **March 2007**, *6*.
3. Zhang, Yuanbo et al. Direct Observation of a Widely Tunable Bandgap in Bilayer Graphene. *Nature* **11 June 2009**, *459*, 820–823.
4. Hernandez, Y. et al. High Yield Production of Graphene by Liquid Phase Exfoliation of Graphite. *Nature Nanotechnology* **2008**, *3*, 563–568.
5. Li, Xuesong. Evolution of Graphene Growth on Ni and Cu by Carbon Isotope Labeling. *Nano Letters* **2009**, *9* (12), 4268–4272.
6. Malard, L. M.; Pimenta, M. A.; Dresselhaus, G.; Dresselhaus, M. S. Raman Spectroscopy in Graphene. *Physics Reports* **2009**, *473*, 51–87.

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Integrated Thin-film Piezoelectric Traveling Wave Ultrasonic Motor

RYAN RUDY
MENTOR: RONALD G. POLCAWICH
RF & ELECTRONICS BRANCH
SENSORS AND ELECTRON DEVICES DIRECTORATE
ADELPHI, MARYLAND

Contents

List of Figures 87

List of Tables 87

Abstract 88

Acknowledgments 89

Student Bio 90

1. Introduction 91

2. Design and Modeling 91

3. Results and Discussion 94

4. Conclusions 100

5. References 101

List of Figures

Figure 1. Exploded view of a piezoMEMS-based USM showing the anticipated position of the stator element in the USM. 92

Figure 2. Micrograph of a fabricated 1-mm stator showing six PZT elements intended to test standing waves. The stator is anchored in the center and released from the substrate. 92

Figure 3. B_{04} mode shown from ANSYS FEA model. Analysis showed a resonance of 1.4 MHz for a 1-mm all-silicon stator disk at this resonant mode. 93

Figure 4. B_{13} mode shown from the ANSYS FEA model. Analysis predicts a resonance frequency of 3.3 MHz for a 1-mm all-silicon stator disk. 94

Figure 5. Sequence of LDV data for a 3mm diameter disk illustrating the B_{13} clockwise traveling wave for the mode shape, shown in figure 6, created by using two input 252.2-kHz sine waves separated in phase by 90°. The black square is included to help track the progression of the traveling wave. 95

Figure 6. LDV test data from an on a 2-mm disk actuated with thin- film PZT actuated at afrequency of 718.8 kHz. This shape matches the corresponding B_{13} mode. 95

Figure 7. LDV test data from standing wave in a 2-mm disk actuated with thin-film PZT actuated at a sine wave frequency of 313.9 KHz. This shape matches the predicted B_{04} mode. 96

Figure 8. Frequency response of a 3-mm-diameter disk up to 1 MHz, illustrating numerous resonance peaks, including the highlighted B_{13} mode. 96

Figure 9. Maximum displacement for different electrodes, frequencies, and excitation voltages shows the amplitude is significantly affected by both the voltage and excitation frequency. 97

Figure 10. Maximum deflection as a function of applied voltage shows a significant cubic contribution to the response. 98

Figure 11. Frequency response of the stator to a square wave input at the resonance frequency displays no significant high frequency components. 99

Figure 12. B_{13} mode shape in a 3-mm-diameter disk from square wave actuation at the resonance frequency shows that a square waveform does not significantly affect the mode shape. 99

List of Tables

Table 1. Example of modeled motor parameters. 93

Table 2. Calculated Q-factor for various conditions. 98

Abstract

An integrated approach to the fabrication of thin-film piezoelectric traveling wave ultrasonic motors (USMs) at the millimeter-scale is being developed for low power, high torque motors for small-scale robotics, biomedical, and sensing applications. This paper describes the realization of ultrasonic motor stators ranging in diameter from 1–3 mm using wafer-scale micro-electromechanical system (MEMS) fabrication techniques with lead zirconate titanate (PZT) thin films. Using laser Doppler vibrometry (LDV), controlled traveling waves were demonstrated in the bulk silicon elastic medium of the stator and the standing wave behavior was characterized for control purposes. Furthermore, the resonant modes of the fabricated stators were modeled using finite element models, and the experimental results agree well with this analysis.

Acknowledgments

I wish to acknowledge the mentorship of Gabe L. Smith and Ronald G. Polcawich, and the assistance in device fabrication by Luz Sanchez, Brian Power, and Joel Martin.

Student Bio

I am currently attending the University of Maryland – College Park and am sponsored by the Science, Mathematics & Research for Transformation (SMART) fellowship program. I have just finished my first year in a Ph.D. program. I received both my bachelor's (2009) and master's degrees (2010) from the University of Michigan. My area of study is mechanical engineering with a focus on micro-electromechanical systems. I have previous research experience in ground mobile millimeter-scale robotics. After completion of my degree, I will work at the U.S. Army Research Laboratory.

1. Introduction

While a number of commercial traveling wave ultrasonic motors (USMs) have been demonstrated at the centimeter and larger scale, effective USMs at the millimeter (mm)-scale and below have eluded researchers (*1*). Rotary USMs have many advantages over electrostatic or electromagnetic motors in that they have excellent high torque, low speed performance; zero power off state with high holding torque; high efficiency; and small size (*2*). Rotary USMs can be created using either standing waves or traveling waves. Standing wave USMs use the vertical vibration of an inclined beam to push the rotor when the beam bends. Traveling wave USMs, first reported by Sashida in 1983 (*3*), use the horizontal motion of the surface of the stator to move the rotor.

Traveling wave USMs are desirable over standing wave USMs due to bidirectional actuation capabilities. Previous efforts in this area have involved the assembly of bulk-fabricated piezoelectric components, with insufficient precision to maintain and control the desired traveling waves. Flynn et al. (*5*) demonstrated an assembled USM using bulk ceramic lead zirconate titanate (PZT) as the actuator with an 8-mm diameter (*4*) and thin-film USMs as small as 5 mm. However, the thin-film devices were unable to provide bidirectional motion control, presumably due to the device configuration that prevented a well-defined traveling wave within the device. Recently, a surface micromachined ultrasonic motor, driven by a bonded bulk piezoelectric, has been reported by Kaajakari et al. (*6*). This device demonstrated clockwise and counterclockwise rotation while also allowing for easy integration of surface micromachined components; however, the traditional ultrasonic motor benefits of no gearing and high holding torque cannot be realized with this actuation scheme.

In the present work, we are exploiting high quality PZT thin-film materials to enable wafer-level batch fabrication with photolithographic patterning to improve manufacturing precision and reduce cost. Using wafer-scale micro-electromechanical system (MEMS) fabrication techniques with thin-film PZT on silicon-on-insulator (SOI) processing, USM stators ranging in diameter from 1–3 mm have been successfully realized.

2. Design and Modeling

The motor design leverages an integrated fabrication process using a combination of surface and bulk micromachining. Figure 1 shows an exploded view of the conceptual design, displaying the rotor, stator, and other motor components. Figure 2 shows the details of a fabricated stator.

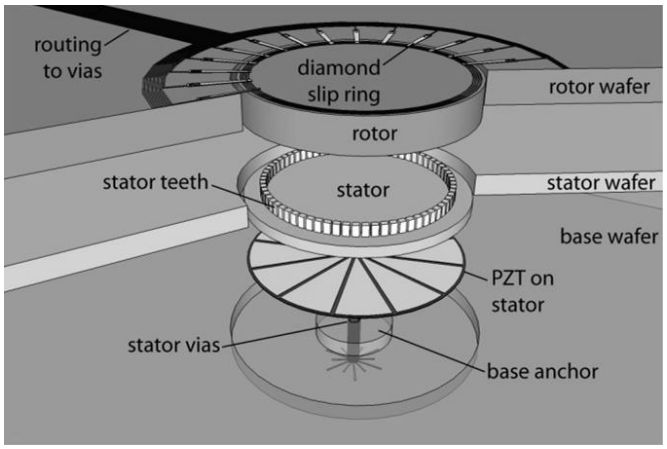

Figure 1. Exploded view of a piezoMEMS-based USM showing the anticipated position of the stator element in the USM.

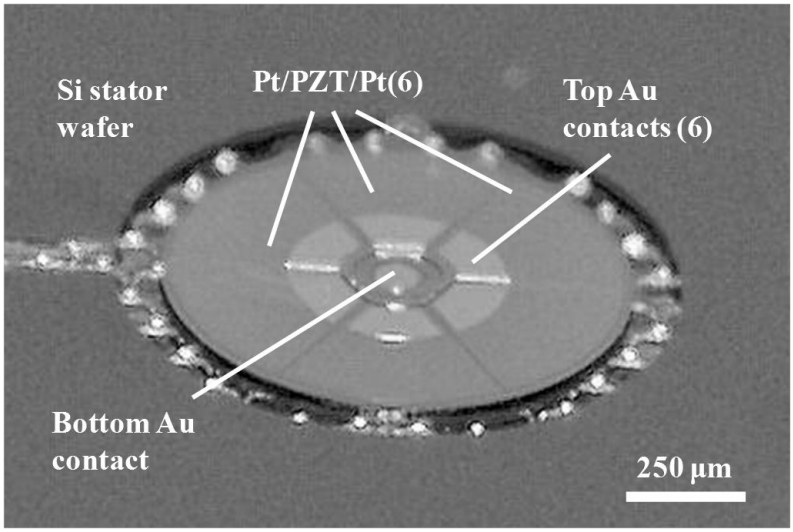

Figure 2. Micrograph of a fabricated 1-mm stator showing six PZT elements intended to test standing waves. The stator is anchored in the center and released from the substrate.

Because the demonstrated fabrication process takes advantage of photolithographic patterning and alignment for all of the components, it does not suffer from the constraints inherent to macro-scale assembly that have traditionally limited the minimum USM dimensions. The fabrication process flow for the stators uses chemical solution deposited PZT, metal sputtering and evaporation, reactive ion etching, etching via vapor hydrofluoric acid (HF), metal liftoff, ion milling, and atomic layer deposition of aluminum oxide (Al_2O_3) (*9*).

In order to predict the response of the USM stators, both equivalent circuit modeling and ANSYS Finite Element Analysis (FEA) modal analysis were performed. Since traveling wave motors operate through a set of complex electromechanical phenomena, an optimal design of the proposed µUSM will require an effective model to guide the design process. While various

vibration modes are under consideration for future USM motors, the B_{13} mode, consisting of 1 nodal circle and 3 nodal diameters, is being explored first. For this case, the stator is patterned with 12 conductive regions grouped into four sets of three alternating electrodes. To estimate the performance of our initial motor design, an equivalent circuit model was constructed assuming a conservative mechanical quality factor of 100. Performance estimates for this early design are shown in table 1.

Table 1. Example of modeled motor parameters.

Parameter	Value
Voltage input	±1.8 V
Stator diameter	1 mm
Stator height (without teeth)	40 µm
Stator tooth height	30 µm
PZT thickness	0.25 µm
Eigenfrequency (B13 mode)	3.4 MHz
Max output speed	8.1 Hz
Maximum output (stall) torque	0.14 N mm
Input power at 1 Hz	3.7 mW

ANSYS FEA modeling was used to determine modal frequencies of the standing waves. Of primary interest were the B_{03}, B_{04}, B_{13} and B_{14} modes consisting of 0 and 1 circumferential nodes and 3 and 4 nodal diameters, respectively. Figures 3 and 4 show finite element simulations of two of these mode shapes.

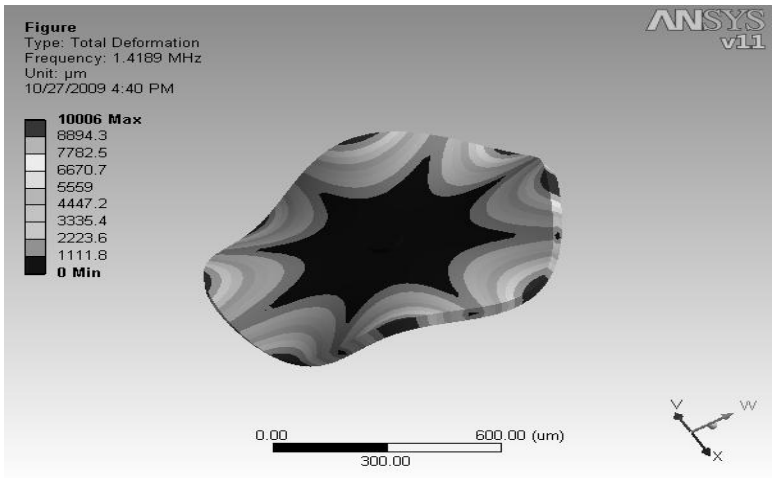

Figure 3. B_{04} mode shown from ANSYS FEA model. Analysis showed a resonance of 1.4 MHz for a 1-mm all-silicon stator disk at this resonant mode.

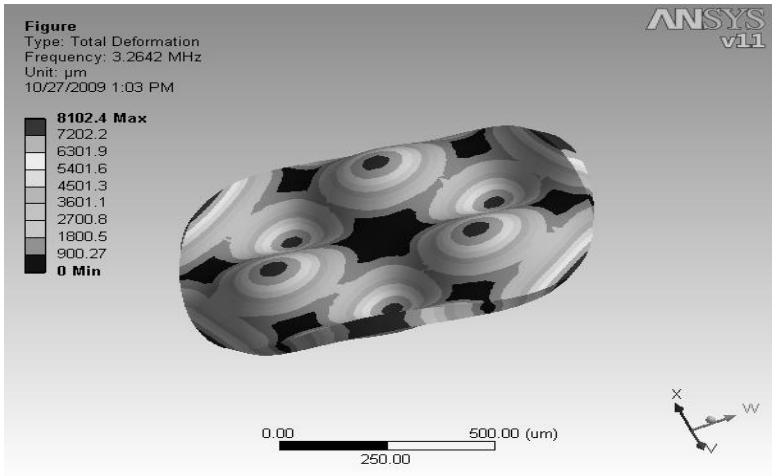

Figure 4. B_{13} mode shown from the ANSYS FEA model. Analysis predicts a resonance frequency of 3.3 MHz for a 1-mm all-silicon stator disk.

These modes are desirable as they ensure rotor stability by maintaining at least three contact points of the stator to rotor, maximize vertical displacement compared to higher order modes, and minimize the required number of actuation regions.

3. Results and Discussion

A controlled traveling wave was induced in the fabricated stators using two out-of-phase input waveforms. This controlled traveling wave, which can be driven in either a clockwise or counterclockwise direction, allows the motor to operate in both directions as opposed to the standing wave rotary USMs reported in reference 10. Figure 5 illustrates the traveling wave using phase-stepped laser Doppler vibrometry (LDV) data. This representation illustrates the deformation of the stator over a single cycle in time, equivalent to 4 μs, resulting in the wave progressing approximately 120° around the disk.

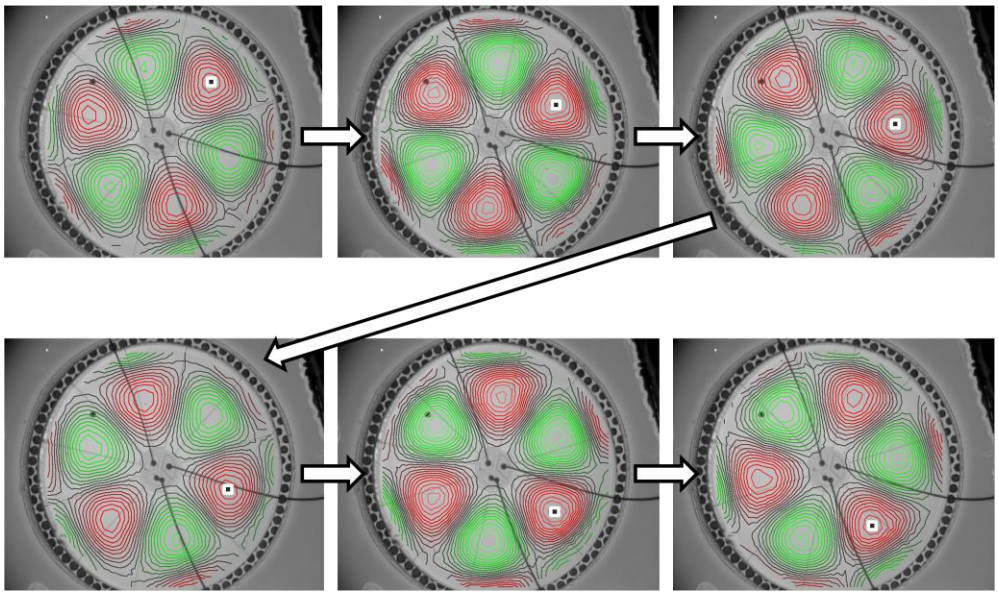

Figure 5. Sequence of LDV data for a 3mm diameter disk illustrating the B_{13} clockwise traveling wave for the mode shape, shown in figure 6, created by using two input 252.2-kHz sine waves separated in phase by 90°. The black square is included to help track the progression of the traveling wave.

The traveling waves were generated by exciting standing waves in phase and apart in space according to the identity,

$$\cos(n\theta)\cos(\omega t) + \sin(n\theta)\sin(\omega t) = \cos(n\theta - \omega t). \quad (1)$$

To characterize the standing waves, individual electrodes were used to excite the resonance mode shapes within the stators. LDV was then used to determine the mode shapes present. The resonance mode shapes for the B_{04} and B_{13} modes, shown in figures 6 and 7, match the mode shapes determined by FEA in figures 3 and 4.

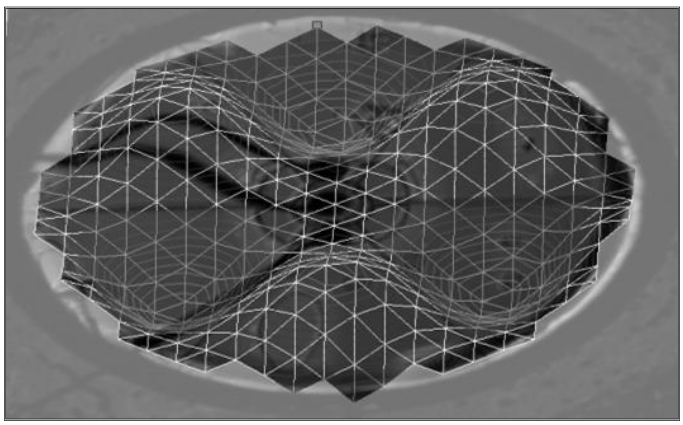

Figure 6. LDV test data from an on a 2-mm disk actuated with thin-film PZT actuated at a frequency of 718.8 kHz. This shape matches the corresponding B_{13} mode.

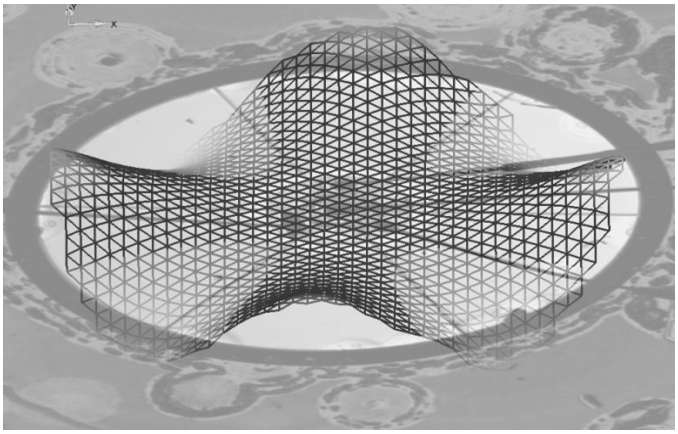

Figure 7. LDV test data from standing wave in a 2-mm disk actuated with thin-film PZT actuated at a sine wave frequency of 313.9 KHz. This shape matches the predicted B_{04} mode.

Experimental determination of the resonance frequencies was determined by inputting a 0–5 V white noise signal, using a waveform generator, and measuring the response using LDV. A Fourier transform was performed on this data to determine the frequencies of greatest amplitude. These frequencies were then isolated and examined to determine the mode shape present at each resonance frequency. Figure 8 shows the frequency response of a 3-mm-diameter disk up to 1 MHz, suggesting a number of resonances.

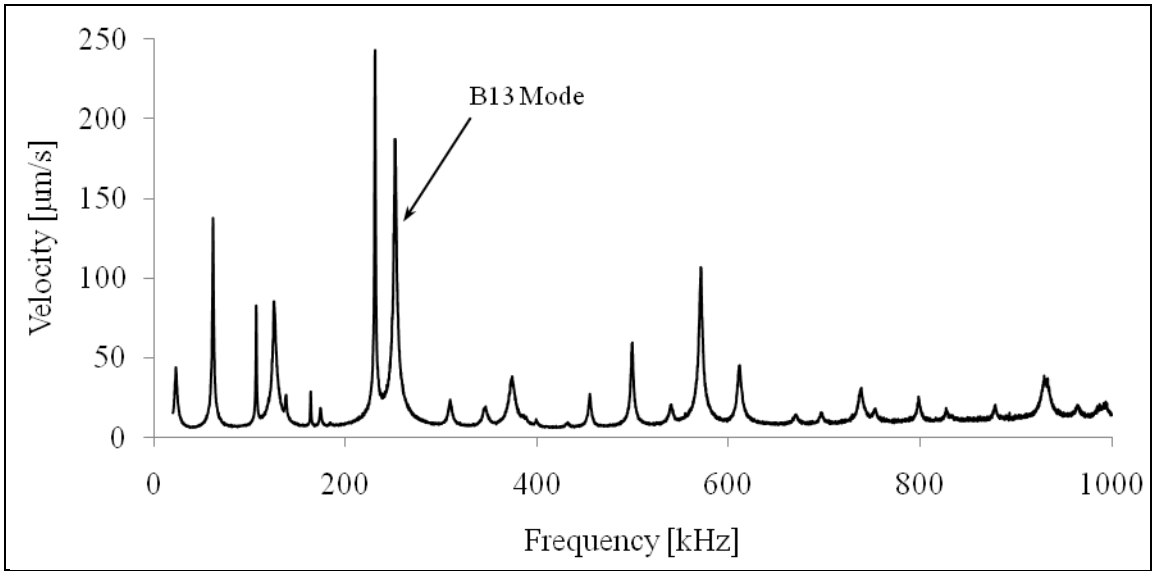

Figure 8. Frequency response of a 3-mm-diameter disk up to 1 MHz, illustrating numerous resonance peaks, including the highlighted B_{13} mode.

The desired B_{13} mode was further characterized for the effects of frequency and voltage on the deflection of the surface. To accomplish this, sinusoids of single frequencies were input to one electrode on the stator and the displacement was determined by LDV. The results of this characterization are shown in figure 9. From this plot it is clear that frequency and voltage both significantly affect the response.

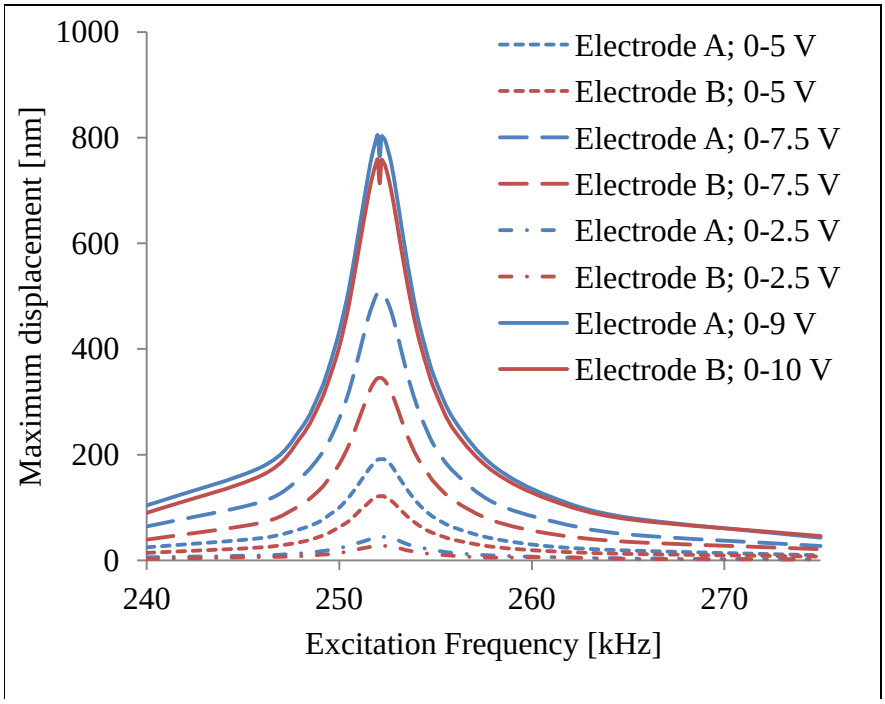

Figure 9. Maximum displacement for different electrodes, frequencies, and excitation voltages shows the amplitude is significantly affected by both the voltage and excitation frequency.

Beyond the simple qualitative conclusion that voltage and frequency affect amplitude, one can note specifics about the system. First, it is an easy extrapolation to approximate the Q-factor:

$$Q = \frac{f_p}{\Delta f}, \tag{2}$$

where f_p is the peak frequency and Δf is the bandwidth, defined as the difference between the upper and lower frequency, where the response reaches half of the maximum amplitude. Performing these calculations for each of the electrodes and voltages gives the Q-factors as recorded in table 2.

Table 2. Calculated Q-factor for various conditions.

Condition	Q-factor
Electrode A; 0–1 V	50.6
Electrode A; 0–2.5 V	53.9
Electrode A; 0–5 V	54.9
Electrode A; 0–7.5 V	54.6
Electrode A; 0–9 V	54.1
Electrode B; 0–1 V	55.9
Electrode B; 0–2.5 V	57.2
Electrode B; 0–5 V	56.8
Electrode B; 0–7.5 V	54.6
Electrode B; 0–10 V	54.9

From this, we can see that for the 3-mm-diameter stator all conditions display a Q-factor near 50. For large Q-factors, Q can be approximated:

$$Q = \frac{1}{2\zeta}. \tag{3}$$

Under this approximation, the evidence suggests that damping in the stator system is not related to displacement magnitude. From this, it can be posited that squeeze film damping, which is highly related to vibration amplitude, is not active in the analyzed stator system. Another point that can be made from the data is that the relationship between voltage and maximum displacement is cubic in nature. The points tested and the equations for the two electrodes tested are shown on figure 10. The nonlinearity is quite strong, accounting for over 60% of the calculated displacement at just a 5-V actuation. In evaluating the equations in figure 10, it is assumed that the displacement relationships should be point symmetric about this origin, thus only allowing odd functions.

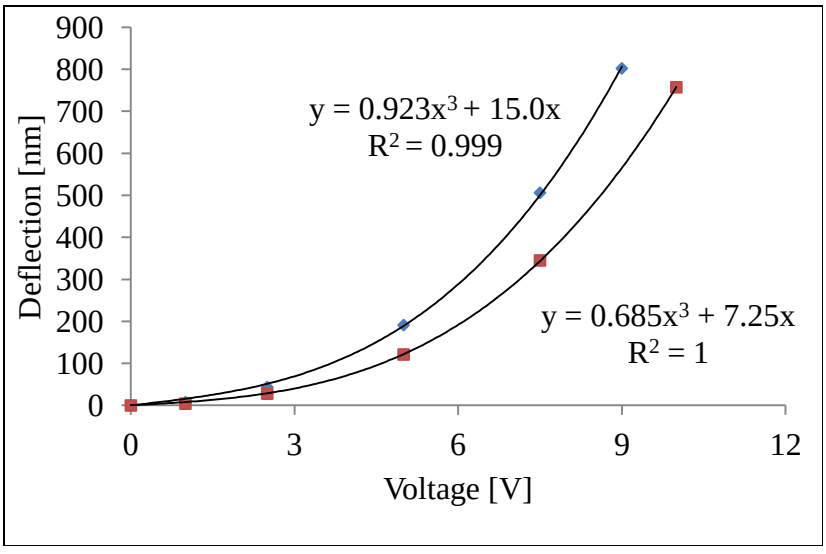

Figure 10. Maximum deflection as a function of applied voltage shows a significant cubic contribution to the response.

This sort of nonlinearity, specifically, a softening cubic nonlinearity, has been observed by others, such as Sattel (*11*). The nonlinear catastrophic jump phenomenon described by Sattel (*11*) was not observed, however. This could be due to the relatively small amplitudes of displacement; however, definitively proving this would require further investigation.

Beyond characterization of stator deflection based on sinusoidal inputs, the stators were also actuated using square waves. No significant high frequency components were observed in the frequency response, as shown in figure 11.

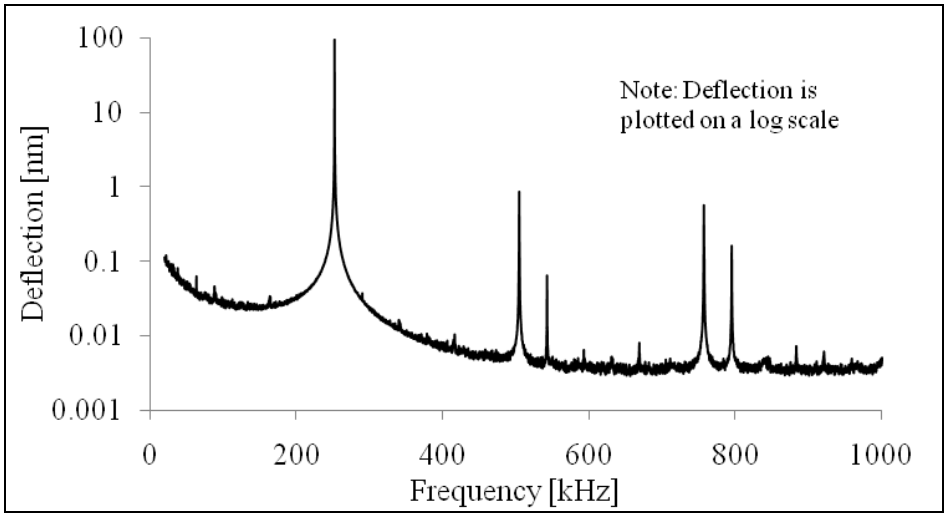

Figure 11. Frequency response of the stator to a square wave input at the resonance frequency displays no significant high frequency components.

From square wave excitation, both traveling and standing waves were induced in the stator. A standing wave from square wave excitation is shown in figure 12, illustrating the B_{13} mode shape unchanged by substituting a square wave for the sinusoidal input waveform.

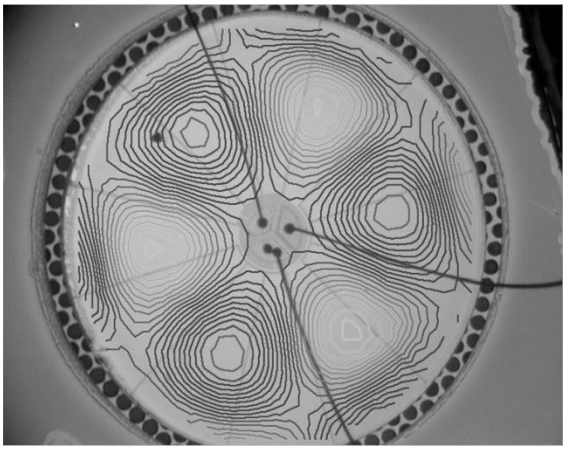

Figure 12. B_{13} mode shape in a 3-mm-diameter disk from square wave actuation at the resonance frequency shows that a square waveform does not significantly affect the mode shape.

4. Conclusions

Both standing wave and traveling wave actuation of the bulk silicon elastic medium of a microfabricated USM stator have been successful achieved using thin-film PZT. The resonant modes of the fabricated stators have been modeled with FEA, with experimental results agreeing well with these simulations, and an analytic model to predict performance characteristics of the motors has been developed. Demonstration of square wave operation without significant high frequency components allows for the use of on-off switching rather than sinusoidal input signals. This makes integration into a portable, untethered system significantly easier. With the excellent predicted performance metrics for the microfabricated USMs and the experimental results from stator testing, the devices are expected to open the door to new opportunities for integrating miniature rotary actuation into a wide range of commercial and military applications.

5. References

1. Uchino, K.; Giniewicz, J. R. *Micromechatronics* **2003**, 417–430.
2. Uchino, K. Piezoelectric Ultrasonic Motors: Overview. *Smart Material Structures* **1998**, *7*, 273–285.
3. Sashida, T. Approach of the Development for the Ultrasonic Motor. *Mechanical Automation Japan* **1983**, *15* (12), 31–35.
4. Flynn, A. M. Performance of Ultrasonic Mini-motors Using Design of Experiments. *Smart Material Structures* **1998**, *7*, 286–294.
5. Flynn, A. M. et al. Piezoelectric Micromotors for Microrobots. *Journal of Microelectromechanical Systems* **March 1992**, *1* (1), 44–51.
6. Kaajakari, V.; Lal, A. Micromachined Ultrasonic Motor Based on Parametric Polycrystalline Silicon Plate Excitation. *Sensors and Actuators A: Physical* **June 2007**, *137* (1), 120–128.
7. Bronson, J. et al. Thin Film PZT Flapping Wing Actuators for Insect-Inspired Robotics. *22nd IEEE International Conference on MEMS*, Sorrento, Italy, Jan 2009.
8. Pulskamp, J. S. et al. Two Degree of Freedom PZT MEMS Actuated Flapping Wings with Integrated Force Sensing. *Hilton Head Workshop*, June 2010.
9. Smith, G. L. et al. Atomic Layer Deposited Alumina for Use as an Etch Barrier Against Xenon Diflouride Etching. *Solid-State Sensors, Actuators, and Microsystems Workshop*, Hilton Head, SC, June 2010.
10. Sashida, T. Trial Construction and Operation of an Ultrasonic Vibration Driven Motor. *Oyo Butsuri* **1982**, *51*, 713–720.
11. Sattel, T. Dynamics of Ultrasonic Motors, Doctoral Thesis, Technische Universität Darmstadt, Germany, 2002.

INTENTIONALLY LEFT BLANK.

Survivability/Lethality Analysis Directorate (SLAD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Investigation of Integrating Three-Dimensional (3-D) Geometry into the Visual Anatomical Injury Descriptor (Visual AID) Using WebGL

AUTUMN KULAGA
MENTOR: PATRICK GILLICH
WARFIGHTER SURVIVABILITY BRANCH, BALLISTICS & NBC DIVISION,
ABERDEEN PROVING GROUND, MARYLAND

Contents

List of Figures **107**

List of Tables **107**

Abstract **108**

Acknowledgments **109**

Student Bio **110**

1. Introduction/Background **111**

2. Support for Implementing 3-D **112**

2.1 Why 3-D? 112

2.2 Why WebGL? 114

2.3 Costs of Implementation 114

3. WebGL Prototype and Discussion **115**

4. Summary and Conclusions **117**

5. References **118**

List of Figures

Figure 1. Visual AID user interface. 112
Figure 2. Inappropriate overlap caused by use of 2-D images. 113
Figure 3. Visual AID interface (left) next to test of Zygote geometry in WebGL environment. 116

List of Tables

Table 1. AIS severity levels. 111

Abstract

The Visual Anatomical Injury Descriptor (Visual AID) is a graphical computer tool developed to illustrate injury and severity on an anatomical figure, allowing for communication of trauma. Currently, the tool uses two-dimensional images rendered from three-dimensional (3-D) geometry to convey injury. While integrating 3-D geometry was a part of the initial development plan, this step was delayed while waiting for technologies to mature. This paper discusses the Web-based 3-D environment prototype being developed to understand the feasibility of integrating WebGL into Visual AID. Using WebGL will permit us to display the current anatomical geometry used by Visual AID in a 3-D Web-based environment. Employing a 3-D environment allows us to more accurately display the geometry, thus providing a better navigation system and the potential to add annotations fixed to 3-D points. Further developments must be made to better understand integration of highly detailed geometry as well as changing geometry attributes at runtime, such as Abbreviated Injury Scale (AIS) code, severity, and color association. WebGL is a novel cross-platform technology that must be further explored as standardization occurs, thus enabling future advancements. The features that have already been explored will enhance Visual AID's injury illustrating capabilities, facilitating the communication of injury data.

Acknowledgments

Thanks to Ronald Weaver, Timothy Myers, and Zachary Hamman for their technical insight and programming expertise on this project, and also to Patrick Gillich for excellent mentorship.

Student Bio

I am a graduate student with a bachelor of fine arts in Communication Arts from Virginia Commonwealth University in Richmond, VA. This undergraduate degree has provided me with the ability to visually communicate information, allowing for a more concise and precise understanding of the information at hand. In order to fulfill the medical illustration track, I completed a concentration in science, focusing on human anatomy. In the fall of 2011, I will be attending the University of Illinois Chicago to obtain a master of science in Biomedical Visualization. This program is one of only five accredited by the Association of Medical Illustrators. For the past five summers, I have been a student employee with the Survivability/Lethality Analysis Directorate in the Warfighter Survivability Branch helping to improve visualization of anatomical geometry.

1. Introduction/Background

Visual Anatomical Injury Descriptor (Visual AID) is a graphical tool developed to illustrate injury and severity on an anatomical representation. Illustrations produced by Visual AID visually communicate trauma. Injuries are based on information from casualty medical records or generated from modeling and simulation. Visual AID illustrations allow for communication of injuries to audiences outside the medical community. The first generation of Visual AID uses two-dimensional (2-D) images of prerendered three-dimensional (3-D) geometry (*1*). The development plan of Visual AID included future integration of 3-D rendering within the application. However, a 3-D implementation was delayed while waiting for 3-D Web technology to mature.

Visual AID enables efficient communication of injury data. Injury data is input by manual selection of tissues or using Abbreviated Injury Scale (AIS) (*2*) codes.

AIS is an anatomical scoring system that ranks the severity of injury, providing a consistent technique for describing trauma. AIS is a seven-digit code for unique injury coding. The first six digits detail the location, nature, and type of injury. The seventh, or post dot, digit encodes the severity level. Injury severity is classified according to a six-point ordinal scale with the addition of a seventh value denoting an unknown injury level (table 1). Within the U.S. Army Research Laboratory (ARL), this scale is associated with color values. The associated colors allow Visual AID to display the severity of injured tissues.

Table 1. AIS severity levels.

AIS	Injury Level	Type of Injury	Example of Injury: HEAD
1	Minor	Superficial	Minor laceration of scalp
2	Moderate	Reversible Injuries; medical attention required	Major laceration of scalp, blood loss < 20%
3	Serious	Reversible Injuries; hospitalization required	Fracture of skull, penetration < 2 cm
4	Severe	Non-reversible injuries; not fully recoverable without care	Depressed skull fracture, penetration > 2 cm
5	Critical	Non-reversible injuries; not fully recoverable even with care	Depressed skull fracture, laceration of spinal artery
6	Maximal	Nearly Unsurvivable	Massive brain stem crush
9	Unknown		Died of head injury without further substantiation of injuries or no autopsy confirmation of specific injuries.

Injury images are generated using the Zygote Human Anatomy 3-D model (*3*). Use of a reference anatomy independent of personal identification, such as Zygote, allows Visual AID to anonymously display actual patient trauma. The Zygote geometry was broken into components in Autodesk Maya 2009, a 3-D modeling program. Each component was modeled to describe a specific location of injury. Individual tissue components were rendered into 2-D images exported from Maya. Each tissue was rendered in all of the seven colors associated with AIS to

allow for display of severity. All of the tissue images are stored in a directory organized by location on the body and tissue type. A black and white background image was made by rendering geometry from Maya and editing in Adobe Photoshop CS3. When a list of injury codes is input into Visual AID, a stack of prerendered images are displayed on the screen over the prerendered background image (figure 1).

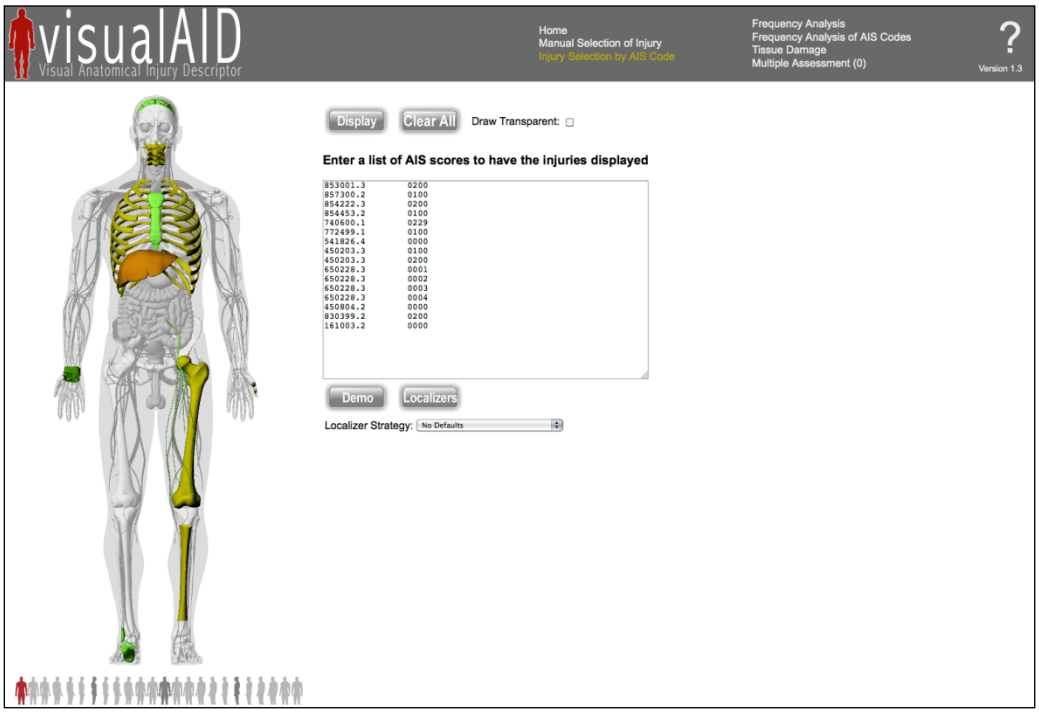

Figure 1. Visual AID user interface.

Recent maturation of 3-D Web technology makes the planned transition from prerendered 2-D images to rendering 3-D geometry within Visual AID a possibility. Web-based Graphics Library (WebGL) is an application programming interface (API) that allows for viewing of 3-D geometry within a Web browser without the need of any plug-ins. WebGL Specification Version 1.0 was released in February 2011 (*4*). This report documents the current effort to understand the advantages, capabilities, and possible costs of integrating 3-D models into Visual AID using WebGL.

2. Support for Implementing 3-D

2.1 Why 3-D?

The development of Visual AID has allowed for easy production of detailed injury illustrations. Implementation of 3-D geometry allows for more accurate injury illustrations, a better navigation system, and the ability to fix placement of annotations.

When 3-D geometry is prerendered from Maya to 2-D images, the geometry is flattened. When tissues located within close proximity of each other are both displayed, they can produce inappropriate overlap. Visual AID places one image on top of the other based on a predetermined layering order. The current layering order is determined by the distance of the geometry from camera Maya uses to render the image. In figure 2, three images—the lung, the heart, and the coronary arteries, respectively—overlay the background image. The lungs and coronary arteries surround the heart. Portions of the lungs should be hidden behind the heart while other portions of the lungs should overlap the heart. Also, the coronary arteries wrap around the heart. No unambiguous layering is available to portray anatomically correct visualizations. Analysts with less knowledge of anatomy examining the data may confuse and misconstrue the illustrations. In figure 2, arteries displayed could be misinterpreted as existing in front of the heart, or that the lung tissue affected is only behind the heart.

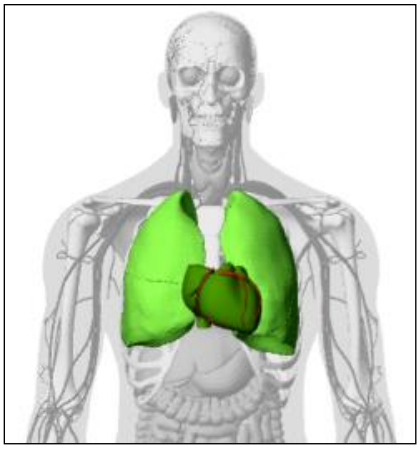

Figure 2. Inappropriate overlap caused by use of 2-D images.

Use of 3-D geometry resolves current issues of overlapping geometry depicted with 2-D images. The Zygote geometry used is defined in 3-D space. If the 3-D geometries representing lungs, heart, and coronary arteries are combined directly in Visual AID, the composite image will display proper interleaving of the tissues. The desire to replace images with 3-D geometry allows Visual AID to provide users with a better understanding of the affected anatomy.

Another advantage to using 3-D rendering is the ability to add a more flexible navigation system. Initial implementation only allowed visualization of injury from the front. The current navigation of Visual AID allows for a view of the anatomy from every 15° along the x-axis. Images are prerendered for each tissue from each supported view angle.

Navigation in 3-D allows for observation from any angle. In addition to rotation along the x-axis, rotation along the y-axis and a zoom option will be added. Adjustment of camera angle can be input to capture any perspective of the injury being illustrated. Once the screen is loaded, the application interchanges injury illustrations to display different angles with little lag.

Furthermore, use of 3-D provides the potential to add annotations that fixed to a specific 3-D point. The current version of Visual AID allows the user to draw a dot in the illustration. The dot is connected to an annotation. If the user decides to view the injury illustration from a different angle, the dot stays in the previous location. The inability to move this annotation along with the geometry is a limitation of Visual AID. This limitation could be fixed in the 2-D version. However, fixing this issue within 2-D implementation would require a rudimentary 3-D context to be developed.

Two functions of the annotation tool are to add notes to specific locations on the illustration and visualizing shot lines. Notes allow for specific comments related to injured tissues. Shot lines are representations of a path taken by a threat, i.e., bullet or fragment. Shot lines are described with the annotation tool by labeling entrance and exit wounds. With the potential to add drawing capabilities in a 3-D context, more mobile annotations can be created.

2.2 Why WebGL?

Previously established methods of Web-based 3-D model visualization use plug-ins associated with specific browsers and platforms to add a 3-D context to HTML4. HTML5 adds a canvas tag, which allows for JavaScript embedded in HTML pages to dynamically draw graphics. WebGL, through JavaScript, provides a means to render 3-D geometry in the HTML5 canvas tag. Web browsers with HTML5 and WebGL allow the rendering of 3-D geometry without the use of any plug-ins. The Khronos Group, a nonprofit industry consortium, developed WebGL. WebGL's cross-platform API is derived from Open Graphics Language on Embedded Systems (OpenGL ES) 2.0. OpenGL is a widely adopted 2-D and 3-D graphics API used to communicate with the graphics processing unit (GPU). The use of an already established cross-platform graphics library allows Visual AID to be more adaptable to future developments.

2.3 Costs of Implementation

As a new development, WebGL is a maturing technology that has yet to be fully developed and supported. WebGL's lack of full development creates a need for a specific browser, a need to import partitioned geometry, and a lack of documentation.

To use WebGL, a new version of a Web browser is necessary. The Khronos WebGL working group has developed specifications for WebGL version 1.0; however, not all Web browsers support this specification. Mozilla Firefox 4.0 and Google Chrome 10 are the current browsers that support WebGL implementation. Internet Explorer has no plans to develop WebGL capabilities, but a third-party plug-in exists. Also, it is problematic to import very detailed geometry. Detailed geometry equates to more points, vertices, in 3-D space. Vertex indices are used in WebGL to define the points associated with respective polygons. The WebGL working group has placed limitations on the amount of vertex indices associated with one piece of geometry in an effort to create a cross-platform capacity, specifically to support embodied systems such as smart phones. This polygon limitation affects Visual AID by limiting the detail

of tissue geometry or forcing the segmentation of tissue geometry into portions and referencing multiple portions to display a single tissue. The final issue to be discussed is the lack of WebGL documentation needed to allow for efficient learning and implementation. While OpenGL ES 2.0 is the basis for WebGL, and references are available for OpenGL ES, no WebGL-specific books are available. WebGL's current capabilities are tested by referencing online tutorials, wiki pages, and the WebGL Work Group's Specification page (*5, 6*).

As WebGL further matures, the aforementioned developmental deficiencies should be resolved. The WebGL specification is being adopted across browsers on existing and emerging platforms. Also, thorough documentation will follow the expanded support of this novel technology.

3. WebGL Prototype and Discussion

To understand the feasibility of integrating WebGL, insight to the WebGL process has been acquired and implementation tests have been performed. In order to provide benefits of 3-D in a Web browser, the abilities that have been explored include understanding the methods of input, assignment of different color shaders to tissues, means to dynamic scene lighting, and control over navigation. To better understand the prospect of integrating WebGL into Visual AID, methods of loading and displaying multiple highly detailed geometries as well as representing separate AIS severities on different tissues must be further explored.

To efficiently run the next version of Visual AID, equipped with WebGL, researchers provided a better understanding of the process WebGL uses to display graphics. JavaScript variables hold buffer and shader information. Buffers hold the data that describe geometry, such as vertex points, vertex indices, and color. A vertex shader and a fragment shader hold different types of shader information, which later define the buffers. Shader programs are built to link variables and shader instructions. These shader instructions are later sent to the GPU along with instructions on how to draw the scene. Knowledge of this process facilitates better development and enhancement of Visual AID.

By following tutorials provided online (*5*), the ability to load geometry and create a dynamic 3-D environment has been prototyped. The development of this project has incorporated heart geometry from Visual AID (figure 3). After understanding the information WebGL needs defined in buffers, such as vertex points, vertex indices, and normal vectors, the information was exported from Maya into text files. The data from the text files were loaded into buffer variables. Once the data-defining geometry is loaded into buffer variables, color can also be defined. Once the GPU has a color defined, the vertex shader facilitates applying the color to the vertex. The fragment shader draws the pixel on the screen and fills in the space between vertices with the correlating color. Once normals are loaded and applied to the vertices, lighting can be integrated into the scene. The normal vectors allow the vertex shader to calculate the direction light reflects off of the geometry, allowing for the appearance of depth of the 3-D shape.

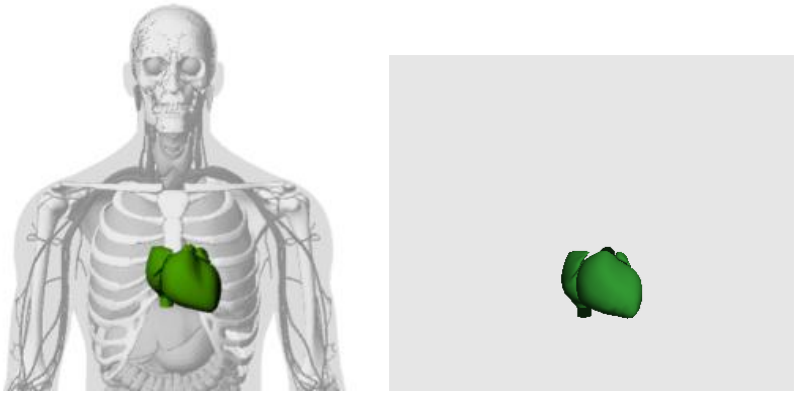

Figure 3. Visual AID interface (left) next to test of Zygote geometry in WebGL environment.

After the tissue geometry was integrated into the scene with color and dynamic lighting, the navigation system was enhanced. A 3-D matrix represents the space where all of the geometry is drawn. In order to adjust the view of the geometry, the matrix is adjusted. To allow for viewing of the heart from multiple angles, the left and right cursor keys adjust the x-axis, the up and down cursor keys adjust the y-axis, and the page-up and page-down keys adjust the zoom view. The geometry data is preloaded into the scene and the view is manipulated, constantly adjusting the position of this data and redrawing the screen.

Development explorations involve fully understanding how to load detailed geometry and how to associate geometry with AIS code to turn on associated color, and methods of adding draw capabilities to add annotations. Currently, it is unknown if latency is an issue when loading all of Visual AID's detailed anatomical geometry. How the application will handle large amounts of data is currently being examined. Geometry with large amounts of indices will have to be reduced or broken into portions and displayed in unison.

Another feature to be evaluated is associating geometry, with correct AIS codes and displaying geometry on the screen. AIS codes can easily be mapped to array variables defining tissue data. Localizers are a numeric code in AIS that further specifies location of an injury. Given the common use of localizers when analysts code AIS, we need to determine if association of only portions of tissue data can be mapped to AIS codes without duplicating entire data files. The benefits or costs of not duplicating this data must be further understood.

Finally, the best method to load geometry and control visibility at run time needs to be better understood. There is the potential to have differences in application performance with preloaded geometry vs. geometry loaded while running the application. The GPU runs the shader programs once when the page is initially loaded; this is the point when the geometry is loaded onto the scene. The current options to explore include loading the geometry at once and adjusting visibility, or reloading the scene to adjust displayed tissue data and AIS severity. The most efficient way to perform such operations will fully be understood once large amounts of data can be properly loaded into the scene.

Current external efforts on the Web, such as Google Labs (*7*), support the prospect of integrating WebGL into Visual AID. In this prototype, benefits of 3-D have been realized by understanding the methods of input, assigning different color shaders to tissues, and adding dynamic scene lighting and navigation. To provide an injury illustration tool with better capabilities and control over the environment, methods of loading and displaying multiple highly detailed geometries, as well as representing separate AIS severities on different tissues, must be better understood and implemented.

4. Summary and Conclusions

As a computer graphics tool, Visual AID allows for effective communication of injury through descriptive illustrations. The current method to display injury or trauma data is by displaying 2-D images of 3-D geometry. Implementation of a 3-D environment will facilitate Visual AID's ability to provide more accurate injury illustrations, navigate fully around the geometry, and more accurately place annotations. These implementations will allow for more effective representations of trauma. WebGL is a viable means to implement the 3-D environment given its cross-platform/cross-browser abilities and lack of need for Web browser plug-ins. The full support of WebGL by Web browsers is still under development; however, multiple sources (*4*–*6*) are available to support integration of this technology into Visual AID.

Understanding the prospect of adding 3-D into the Web environment was done by integrating WebGL to control the data input, assigning different color shaders, and adjusting dynamic lighting, as well as improving navigation. As further investigation and creation of a WebGL prototype continues, the methods of loading multiple highly detailed geometry and assignment of AIS code with severity to tissue data will be examined. With the enhancement of Visual AID through integration of a 3-D environment, more accurate and descriptive injury illustrations can be made. The use of an established cross-platform graphics library, such as WebGL, will facilitate many of Visual AID's future advancements.

5. References

1. Kulaga, A. R.; Gillich, P. J. *Visual Anatomical Injury Descriptor (AID)*; Survivability/Lethality Analysis Directorate, U.S. Army Research Laboratory, July 16, 2010.
2. International Injury Scaling Committee. *Abbreviated Injury Scale (2005 Revision)*; Association for the Advancement of Automotive Medicine: Arlington Heights, IL, 2008.
3. Zygote Media Group Web site. Zygote Human Anatomy 3D Model, 2010. http://www.zygote.com/ (accessed July 26, 2011).
4. Khronos Group Web site. Khronos to Create New Open Standard for Advanced Device and Sensor Input, April 12, 2011. http://www.khronos.org/news/press/releases/khronos-webgl-initiative-hardware-accelerated-3d-graphics-internet (accessed 26 July 2011).
5. Learning WebGL Web site. http://learningwebgl.com (accessed 26 July 2011).
6. Khronos Group Web site. WebGL Specification, 10 February 2011. https://www.khronos.org/registry/webgl/specs/1.0/ (accessed 26 July 2011).
7. Google Body Web site, 2011. http://bodybrowser.googlelabs.com/ (accessed July 26, 2011).

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Target Modeling for Ground Mobile Branch (GMB)

MATTHEW SCHULZ
MENTOR: SCOTT HORNUNG
GROUND MOBILE BRANCH
SURVIVABILITY/LETHALITY ANALYSIS DIRECTORATE
BUILDING 247
ABERDEEN PROVING GROUND, MARYLAND

Contents

List of Figures **121**

Abstract **122**

Student Bio **123**

1. Introduction **124**

2. Description of Effort **124**

2.1 Data Collection Through Metrology 124

2.2 Modeling 126

2.3 Test Shot Data 127

3. Results and Discussion **128**

4. Summary and Conclusions **129**

5. References **130**

List of Figures

Figure 1. iGPS unit used by TMT to capture reference points. Points are collected through the tip of the arm shown at the top. 125

Figure 2. Scanning arm used to collect point data and capture curves. 125

Figure 3. Basic geometry that has been isolated for use in constructing a representative model 126

Figure 4. A completed model of a global positioning system (GPS) unit and its casing. The model is derived from conditioned basic geometry. 127

Figure 5. Snapshot of a BAD test prior to munition detonation. The penetrator is not displayed. 128

Figure 6. Witness plate after being hit with spall generated from the testing process. 128

Abstract

This report documents my efforts and the insights I have gained while serving as a target modeler for the Ground Mobile Branch (GMB) of the Survivability Lethality Analysis Directorate (SLAD). In support of the Branch's role of providing modeling and simulation support through the ballistic analysis of current and future vehicular systems, my primary task was to gain knowledge and insight into the analysis process through understanding the role SLAD plays in the development of combat systems, specifically, how SLAD's analysis results leverage vehicle evaluation and assessment in order to lead to modifications, which will ultimately increase the survivability of a particular system. To achieve this, I participated in vehicle measurements using metrology equipment (using three-dimensional [3-D] BRL computer aided design (CAD) target geometries and the MUVES vulnerability model), supported target geometry development through the production of component parts, and observed various test-shots intended to produce data for use in MUVES analysis.

Student Bio

I am currently an undergraduate student at the University of Maryland, Baltimore County (UMBC) majoring as a mechanical engineer in my junior year of study. I served as a target modeler for the Ground Mobile Branch (GMB) in the previous summer of 2010 as a Student Temporary Employment Program (STEP) hire and was invited to return to further my developing computer-aid design (CAD) skills and deepen my understanding of the interconnectivity of the Branches in their ultimate mission. My future goal is to transfer to Weapons Materials Research Directorate (WMRD) in order to aid in the study and design of current and future armor systems and their component materials.

1. Introduction

The Ground Mobile Branch (GMB) contributes to the U.S. Army Research Laboratory's (ARL) mission by modeling and analyzing ground-based vehicular systems. This is typically achieved through the use of high-fidelity computer-aided design (CAD) models and use of the MUVES-S2 (referred to as MUVES) analysis software. The results generated from the analyses of ground-vehicle systems are used to make vehicle assessment recommendations for modifications aimed at achieving vulnerability reductions.

Due to the nature of our mission, a large percentage of our efforts are involved in producing the data and models necessary to support the MUVES software. MUVES is a suite of programs designed to aid in the prediction of vehicle vulnerability by incorporating vehicle models, penetration algorithms, and a wide variety of other data relating to performance of both the threat and target (*1*). These inputs are required before MUVES is capable of running a quality analysis and developing them is a primary task for any analysis.

An ongoing project GMB is currently involved in is the development of a target vulnerability model of the Special Operations Command (SOCOM) variant of the Mine Resistant Ambush Protected (MRAP) All Terrain Vehicle (MATV). I have participated directly in this project through my work in applying modern data collection techniques using metrology equipment and my contributions relating to creating vehicle target geometries. In addition to this, I have observed the testing process used to develop behind armor debris (BAD) models for the MUVES software.

2. Description of Effort

2.1 Data Collection Through Metrology

The development of a target description begins with the identification of its intended use. Radar signature analyses require highly accurate surface representations compared to a standard vulnerability analysis, which requires accurate CAD geometries of all components and internal systems. These requirements shape the data collection process: accurate surfaces can be derived from extremely dense point cloud data, whereas relative positions and basic geometries can be obtained through three-dimensional (3-D) scanning technologies and hand measurements. Traditionally, the requirements of GMB require the constraints of the latter.

For the SOCOM MATV modeling project, our collectors used several coordinate measuring machines (CMMs) to ease and shorten the collection process. One CMM (figure 1) uses two rotating laser emitters to identify the location of the CMM and maintain a constant global

coordinate system as data are collected. This system is known as an Indoor Global Positioning System (iGPS) and is used for the capture of reference points over the vehicle. The second CMM used by our collectors is known as a scanning arm (figure 2). Data are collected through an interchangeable tip at the head of the arm; a small sharp tip is used for collecting point data and a small ball-head tip is used for capturing curves, lines, and edges.

Figure 1. iGPS unit used by TMT to capture reference points. Points are collected through the tip of the arm shown at the top.

Figure 2. Scanning arm used to collect point data and capture curves.

Proper data collection techniques outline a general procedure for collecting basic geometries and reference points. The process begins with the staging of the vehicle; this is a blanket term that includes aligning articulation points and setting the vehicle into a typical combat configuration, as well as any other preparation that must be done on either the vehicle or the workspace environment. After staging the vehicle, reference points are identified on the vehicle. Reference points allow collectors to overcome the range limitations of their metrology equipment by providing common points between multiple data sets (when moving the positioning equipment,

small discrepancies in the global coordinate system may arise due to the sensitive nature of these devices), which serve as references for fitting all data sets to one common coordinate system (*2*). These points may be either predetermined standards or simply symmetry points that will aid in aligning the future target model's data points. Once the reference points have been identified, the actual collection of data can begin. Data are collected in the form of isolated points from surfaces and captured edges. Complex shapes are broken down such that only the edges and points necessary to rebuild them within a CAD package are taken.

2.2 Modeling

GMB places a large emphasis on the development of high-fidelity CAD to ensure the geometries used for analysis in MUVES do not introduce unnecessary error through the improper representation of targets. A wide variety of CAD packages (including but not limited to Pro-E/Creo, Spaceclaim, and Rhino) are used to develop the necessary geometry depending on the situation and personal preference of the modeler in question. All models must eventually be converted into a BRL-CAD™ format as it is the only CAD package currently supported by MUVES, making it critical to the operations of GMB.

The final product of the data collection process provides a foundation for a developing CAD model. Parts and component level structures are built through the isolation (figure 3) and conditioning of captured data; a process that entails normalizing and projecting captured curves and points to planar regions. To accomplish this, modelers use analysis tools within the CAD packages to determine critical values for the data in question (circle radii, line segment lengths, etc.) and rebuild the curves according to both ideal and measured values. This accounts for error in the collection stage and eases the final step of aligning every component within a master file. Conditioned data and geometric relations are then used to create the solids and surfaces necessary to build up the component parts.

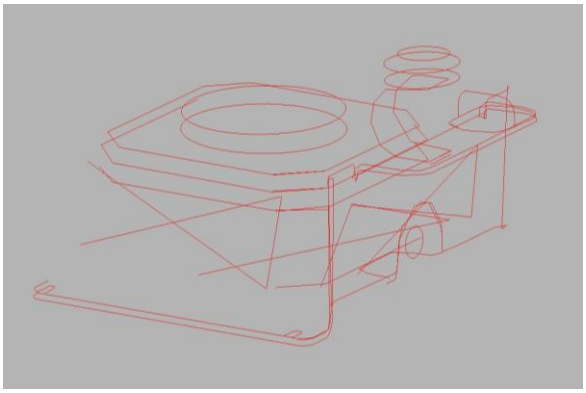

Figure 3. Basic geometry that has been isolated for use in constructing a representative model.

The SOCOM MATV modeling process saw the simultaneous development of two distinct models within both Rhino, a CAD package based on non-uniform rational B-splines (NURBS),

and BRL-CAD™ format. This dual-package development allowed for rapid development of components within Rhino and the subsequent importing of these parts into BRL-CAD™ to be used as references for rebuilding them natively. Rebuilding parts in BRL-CAD™ that have been imported from Rhino is preferred to the actual conversion of components as it aids in minimizing the amount of error introduced when the conversion process approximates surfaces. My contributions to the modeling project are the addition of completed components for the Rhino model (figure 4), including various antenna and their mounts, stretcher brackets, and assorted detail work among other components.

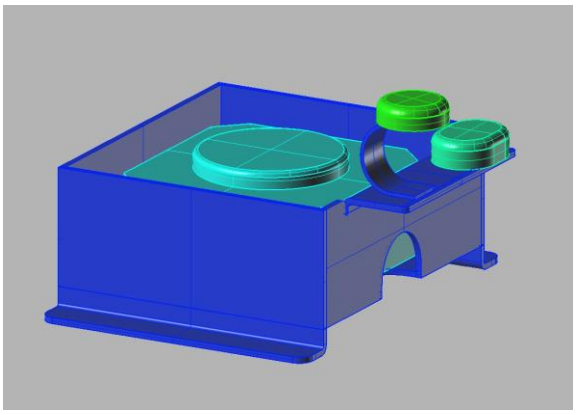

Figure 4. A completed model of a global positioning system (GPS) unit and its casing. The model is derived from conditioned basic geometry.

2.3 Test Shot Data

Just as the accuracy of target geometry can be a limiting factor in a MUVES analysis, the characteristics and material properties of both the threat and target need to be accurate in order to ensure the software operates as accurately as possible. Other complications that affect vehicle survivability, such as behind armor debris (BAD) and other secondary effects, must be considered through the development of suitable simulation inputs.

Coupon test data provide the input that defines both the threat characteristics and the performance and material properties of the armor being fired upon. Threats are reduced through this experimentation to a series of tables and values for properties including impact velocity, fragment mass, and relative penetrative power. The strength of vehicle armor is typically approximated by an experimentally determined equivalence to a given thickness of rolled homogeneous armor (RHA). Tests that calculate these values and approximations are typically conducted by the Weapons Materials Research Directorate (WMRD). These values are input into MUVES as part of a series of penetration algorithms.

BAD, also known as spall, can be a significant concern in determining vehicle survivability. In order to model more effectively and better predict its impact on crew and internal components,

testing is routinely performed by the Survivability Lethality Analysis Directorate (SLAD) to investigate and develop new characterizations of spall proliferation. Current test methods involve firing a penetrator at a coupon (figure 5) to develop spall fragments that impact and penetrate a series of witness packs (figure 6). The path and depth of the channels the spall fragments generate allow for approximations of fragment mass/velocity and fragment trajectory (*3*). These data allow the development of an approximated "spall cone" model for use within MUVES analysis.

Figure 5. Snapshot of a BAD test prior to munition detonation. The penetrator is not displayed.

Figure 6. Witness plate after being hit with spall generated from the testing process.

3. Results and Discussion

Every aspect of my modeling experience and the analysis process is more than a self-contained unit—each part acts as a link in a chain, passing on information and products that are used by the

next step. Metrology provides the foundation for the development of target geometries through the accurate and efficient acquisition of basic shapes and data. Testing data produce properties and relations that are used to model threat and armor performance on an analytical level. Both products are then integrated into the analysis directly through the use of the MUVES software. The results and revelations these analyses yield represent the next step of the chain and are the final product that SLAD produces.

Because every aspect of the process has some influence on the final result, it becomes clear that every input at every stage is important. Work in any one link cannot falter or produce sub-par products as it leads to inefficiencies and inaccuracies in all following links. This concept of interconnectivity helped clarify the concept that my work and its quality directly correlate to my utility. For every measurement I err in accuracy or any target description I incorrectly develop, I have produced a flawed product that will, in turn, lead to a less accurate assessment of vehicle survivability. The idea that my errors could result in excess effort to rectify my mistakes within later links of the chain provides enough initiative to ensure I produce the highest caliber of work that I am capable of.

The ultimate link in our design chain is the Soldier—to ensure they can operate effectively and with success, we must be certain that every link in our chain is secure.

4. Summary and Conclusions

This paper documents my experiences and understanding of the analysis process while serving as a target modeler for GMB. In my course of duty, I have performed work in areas including the data gathering processes and the development of target geometries. Additionally, I have attended coupon testing and gained an understanding of the subsequent analysis of the results.

Through training and practical application of skills, I have furthered my proficiency in multiple CAD packages; developed a sense for the techniques and methodology used within modern metrology; and gained an understanding of the procedures and standards used for armor characterization.

In gaining an understanding of both the direct impact of my work and the consequences it will have on those who rely on it, my own appreciation of the interconnectivity of the work both GMB and SLAD as a whole perform within the structure of ARL has deepened. Each Branch supports the whole in our ultimate mission—the betterment of the Soldier.

5. References

1. Orn, J. U.S. Army Research Laboratory, Survivability/Lethality Analysis Directorate, Aberdeen Proving Ground, MD. Private communication, MUVES analysis, June 2011.
2. Reynolds, F. U.S. Army Research Laboratory, Survivability/Lethality Analysis Directorate, Aberdeen Proving Ground, MD. Private communication, Data Collection/Metrology, June 2011.
3. Abell, J. U.S. Army Research Laboratory, Survivability/Lethality Analysis Directorate, Aberdeen Proving Ground, MD. Private communication, Behind Armor Debris Testing, June 2011.

Vehicle Technology Directorate (VTD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Digital Image Correlation of Flapping Wings for Micro-Technologies

LESLIE HALL
MENTOR: DR. RAJNEESH SINGH
VEHICLE TECHNOLOGY DIRECTORATE
ABERDEEN PROVING GROUND, MD

Contents

List of Figures 135

List of Tables 135

Abstract 136

Acknowledgments 137

Student Bio 138

1. Introduction and Background 139

2. Experiments and Calculations 141

2.1 Load Cell Design and Manufacturing 141

2.2 Design Space and Design of Experiments 143

2.3 Wing Design and Manufacturing 145

2.4 Experimental Setup and Process 146

3. Results and Discussion 147

4. Summary and Conclusions 147

5. References 148

List of Figures

Figure 1. Wing lever mechanism. 140
Figure 2. Wheatstone bridge circuit. 142
Figure 3. Strain gage locations. 142
Figure 4. Final load cell design with actuator and wings. 143
Figure 5. Wing shape inspirations: locust (left), balsa (center), and drosophila (right). 143
Figure 6. DOE wing models. 144
Figure 7. Additional wing models. 145
Figure 8. Speckle reference. 145
Figure 9. DIC wing mount. 146

List of Tables

Table 1. Wing shape DOE. 144

Abstract

The unpredictable, and therefore inherently dangerous, situations warfighters experience have created a demand for advanced environmental knowledge of impending threats. Small, stealthy, and versatile micro-technologies have great potential in this area and can provide autonomous reconnaissance and nonlethal protection while either crawling or flying. One particular challenge with flying systems is designing a lightweight vehicle while maintaining critical function performance. This project focuses on using digital image correlation software and high-speed cameras to analyze lightweight flapping wings for micro-system use. To understand the parametric design space, a design of experiments is created, varying wing span, chord, shape, and spar count. Then the wings are manufactured using a three-dimensional printer and thin plastic sheeting. A custom-made load cell measures the thrust and lift of the speckled wings, which are mounted on and powered by a bimorph actuator. The stereo setup assesses the wing's speckle pattern and measures the change in light intensity, which correlates to wing structure movement or deformation. This enables the calculation of important wing parameter trends, including strain, position, velocity, and acceleration. Computational modeling verifies these results while identifying design features that enhance flight performance and support a lightweight and efficient micro-system design.

Acknowledgments

I would like to extend my gratitude to the following people, for without their assistance, the efficiently, accuracy, and success of this project would not have been possible:

- Dr. Dimitri Mavris, Aerospace Systems Design Laboratory director at the Georgia Institute of Technology, and Dr. Elias Rigas, Vehicle Applied Research Division Chief at the U.S. Army Research Laboratory (ARL), for providing me this opportunity and for their support.
- Dr. Rajneesh Singh and John Gerdes of ARL, for their mentorship and guidance.
- Asha Hall, Paul Moy, Christopher Kroninger, and Howard Carpenter of ARL, for their expertise and the use of their laboratory space and materials.
- Pierre Valdez, Dr. Mulugeta Haile, and Mark Foster of ARL, for all of their technical assistance.

Student Bio

I am a second-year graduate student at the Georgia Institute of Technology in the Aerospace Systems Design Laboratory. I received my bachelor's degree in aerospace engineering from The University of Michigan, Ann Arbor. My interests at Georgia Tech include systems engineering analysis of the U.S. Army Research Laboratory's micro-autonomous technologies as well as designing processes and methods for optimizing the engineering collaborative design process for the Office of Naval Research. I have also been involved in researching heat shields and several other entry, descent, and landing projects at the National Aeronautics and Space Administration Jet Propulsion Laboratory. During my free time, I am involved in community service regarding healthy diets for the elderly and disabled, and have served as a tutor for grade school students. I enjoy playing the tuba in the Georgia Tech Pep Band, playing classical piano, reading science fiction books, and jogging.

1. Introduction and Background

Micro air vehicles (MAVs), as defined by the Defense Advanced Research Projects Agency (*1*), are maneuverable aerial robotic mechanisms that weigh <100 g and are smaller than 150 mm. MAVs are of great interest to the military because in life-threatening situations, they have the potential to scan, gather, and relay critical information back to the unit commanders. MAV missions could include interior reconnaissance, convoy assistance, nonlethal protection, biochemical sensing, and deployment of light payloads.

MAV research is difficult due to vehicle requirements; the vehicle must be small, lightweight, and expendable, while maintaining the endurance and stealth required for these missions (*2*). Many aerodynamic issues arise when operating at low Reynolds numbers as well as difficulties computationally modeling the thin wing skin. Due to these issues, no MAVs exist that operate with the precision and efficiency of insects; investigating MAVs is an incredibly intriguing field to many researchers.

Whereas most motion capture systems are costly and permanent, digital image correlation (DIC) is an alternative that can be temporary, faster, and less expensive. DIC is a displacement measuring technique that uses a large number of images that are taken during the loading of a specimen. The concept behind DIC is simple and only holds two assumptions. First, the motion of the object in the images directly corresponds to the actual test specimen, and second, the object has adequate variations in light intensity. The software is capable of acquiring, interpreting, and analyzing the variations in light intensity reflected from a surface pattern, referred to as speckle. Any change in light intensity at a given point on the object correlates to structure movement and deformation. If the light intensity from the speckle remains constant, then any movement is purely translational; this can be resolved into the position, velocity, and acceleration of the specimen over time. However, a variation of light intensity correlates to material deformation at this point, which can be resolved into strain.

DIC analysis can be done with a single camera; however, adding another camera enables three-dimensional (3-D) analysis of the system, which is desirable for flapping wings because of their rotational movements. The DIC system used in this project is the Vic-3D, which will resolve the wing motion, allowing the strain, position, velocity, and acceleration to be calculated. The Vic-3D setup is simple and allows the cameras to be anywhere from 15° to 45° apart to obtain stereo viewing. The software (*3*) is extremely useful because it is simple to set up and is capable of interpreting the data from both cameras and compiling in order to attain a complex 3-D stereo image of the test subject.

The flapping wings will be mounted on a layered Pb(Zr0.55Ti0.45)O3 lead zirconate titanate (PZT) bimorph actuator (*4*, *5*). Electrical currents sent through the actuator's lead-based

piezoelectric material results in a few millimeter displacement of the material tip at frequencies up to 200 Hz. The tip of the actuator material is connected by a simple lever to the wing base, as shown in figure 1, which dissipates practically no power. The frequency of the current is tuned to vary the wing-flapping frequency and is generally adjusted to match the wing resonant frequency for maximum tip displacement (around 5 to 30 Hz). The small displacement in the actuator material is amplified at the wing tip; at resonant frequencies, the actuator can provide wing-flapping angles up to 70°.

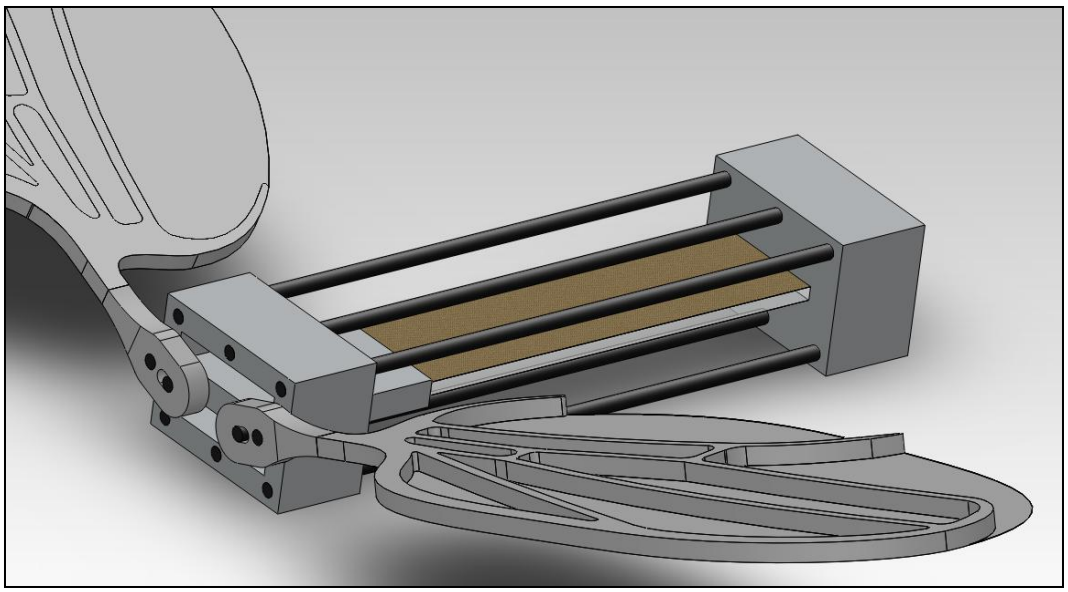

Figure 1. Wing lever mechanism.

The wing actuator base is clamped to a load cell custom designed to measure the experienced thrust and lift. Piezo-resistive semiconductor strain gages are used because of their small size and high sensitivity. The load cell design is optimized to be sufficiently sensitive in resolving very small forces while also maintaining adequate stiffness to avoid natural frequencies near those of the flapping wings. A detailed overview of the load cell design process can be found in section 2 of this report.

To determine the wing configurations to be tested, a design of experiments (DOE) is created varying the wingspan, chord, shape, and number of spars. Twelve sets of wings are manufactured using a 3-D printer for the structure and a thin plastic sheet for the skin. The use of a strobe light to capture the wing movement with fewer frames per second was considered; however, for simplicity, a Photon high-speed video camera is used instead. The load cell strain gages are linked to a voltage meter to record the resulting thrust and lift generated from the flapping wings. After the load cell and DIC data has been processed, trends between design parameters and wing performance are analyzed and compared to similar research results. Results are also validated by detailed wing runs and computational analysis results; details of these aspects can be found in section 2 of this report.

2. Experiments and Calculations

The process of this project consisted of several defining steps. First, load cell designs and methods are researched to understand characteristics required to resolve small forces; this information is used to design and manufacture the load cell. Concurrently, design characteristics of past studies and future areas of interest are compiled to create a flapping wing parametric design space. Next, using the design space, we designed and manufactured a set of test wings. Last, the flapping wings are set up, and high-speed cameras are used to gather information for the DIC software. The load cell is then used to gather forces data. The coupling and comparison of the final sets of data are discussed in section 3 of this report.

2.1 Load Cell Design and Manufacturing

Load cells generally consist of a beam with strain gages attached as specific points. The strain gage conductor is stretched or compressed when the material is deformed, changing its resistance. This resistance can be measured by the resulting change in voltage, which correlates to the strain in the material. The equation for engineering strain is shown in equation 1, where ε is the engineering strain, L_0 is the initial length, and L_1 is the final length.

$$\varepsilon = \frac{(L_1 - L_0)}{L_0}. \tag{1}$$

Piezo-resistive semiconductor strain gages are ideal for the flapping wing application because they are small, measuring only 0.008 in wide, and have high sensitivity, more than 50 times greater than that of foil gages. However, due to the high changes in gage resistance, care must be taken when using semiconductor gages with the Wheatstone bridge, shown in figure 2. The Wheatstone bridge is characterized by equation 2. The strain measurement from the Wheatstone bridge circuit is shown by equation 3 and is dependent on the gage factor (GF).

$$V_{out} = V_{in}\left[\frac{R_4}{R_4 + R_2} - \frac{R_3}{R_1 + R_3}\right]. \tag{2}$$

$$\varepsilon = \frac{\frac{\Delta R_2}{R_2}}{GF}. \tag{3}$$

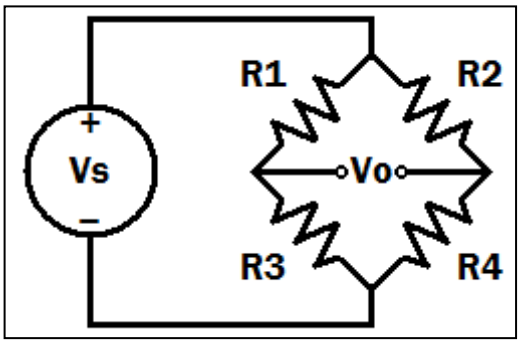

Figure 2. Wheatstone bridge circuit.

Micron Instruments SS-060-033-1000P-S1 practice strain gages were applied to 16 locations on the load cell, as shown in figure 3. The load cell is designed to be sensitive in the corresponding lift and thrust directions, where no lateral force is expected because of symmetrical wings. The final design features create a balance between system sensitivity and stiffness. The system must have adequate sensitivity to resolve small forces but enough stiffness to avoid low natural frequencies near those of the flapping wings (5 to 30 Hz). The final load cell design, shown in figure 4, is computer numerical control manufactured from 6065 aluminum, can resolve forces down to 0.1 g, and has a first-order natural frequency around 90 Hz.

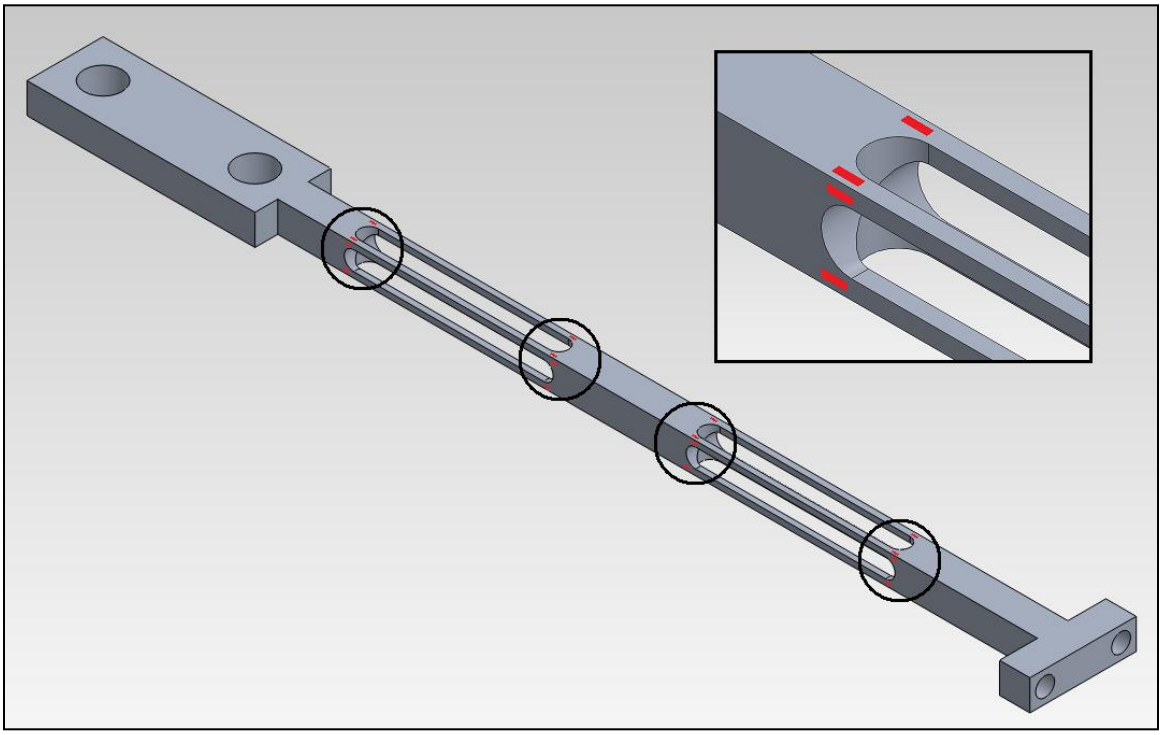

Figure 3. Strain gage locations.

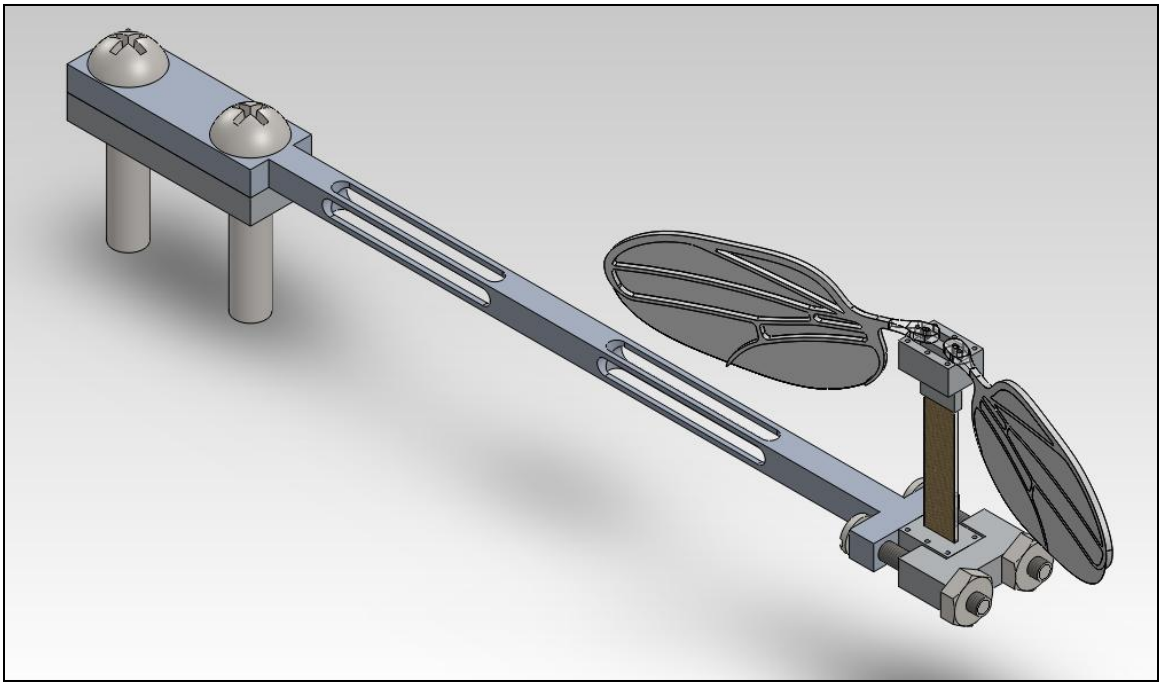

Figure 4. Final load cell design with actuator and wings.

2.2 Design Space and Design of Experiments

The parametric design space is a compilation of previously researched wing designs as well as design characteristics that may be of future interest. Three main wing shapes have been researched in the past: wing shape 1 is based off the wing of a locust, wing shape 2 is based off Kumar's balsa wood design (*6*), and wing shape 3 is based off the drosophila wing. These three wing types are used as a starting point in the wing design for this project and are shown in figure 5. The wingspan varies between 3.5 and 5.5 cm, the wing chord varies between 1 and 3 cm, and the spars vary from zero to two. When parameters from previous studies are used, it is important to keep the Reynolds number accurate depending on the test size and fluid viscosity. Reynolds number is dependent on chord length *l*, velocity *U*, and kinematic viscosity ν, shown in equation 4.

$$Re = \frac{lU}{v}. \tag{4}$$

Figure 5. Wing shape inspirations: locust (left), balsa (center), and drosophila (right).

The DOE is determined using Taguchi's L_9 orthogonal array, allowing the four parameters to each have three levels, as shown in table 1. This array results in 9 test runs, shown in figure 6, rather than 81 tests if done using the full factorial. Three additional tests were added to the test plan: two to test rectangular shape and confirm material deformations known from Kirchhoff-Love flat plate bending theory and one design that includes a large-scale version of the drosophila wing vein structure, shown in figure 7. With this experimental data, response surface methodology (RSM) will be used to explore the relationship of the design parameters to the experimental outputs. JMP statistical software will be used for the RSM analysis.

Table 1. Wing shape DOE.

Run	Shape	Chord (cm)	Half Span (cm)	Spars
1	Locust	1	4.5	0
2	Locust	2	5.5	1
3	Locust	3	6.5	2
4	Balsa	1	5.5	2
5	Balsa	2	6.5	0
6	Balsa	3	4.5	1
7	Drosophila	1	6.5	1
8	Drosophila	2	4.5	2
9	Drosophila	3	5.5	0

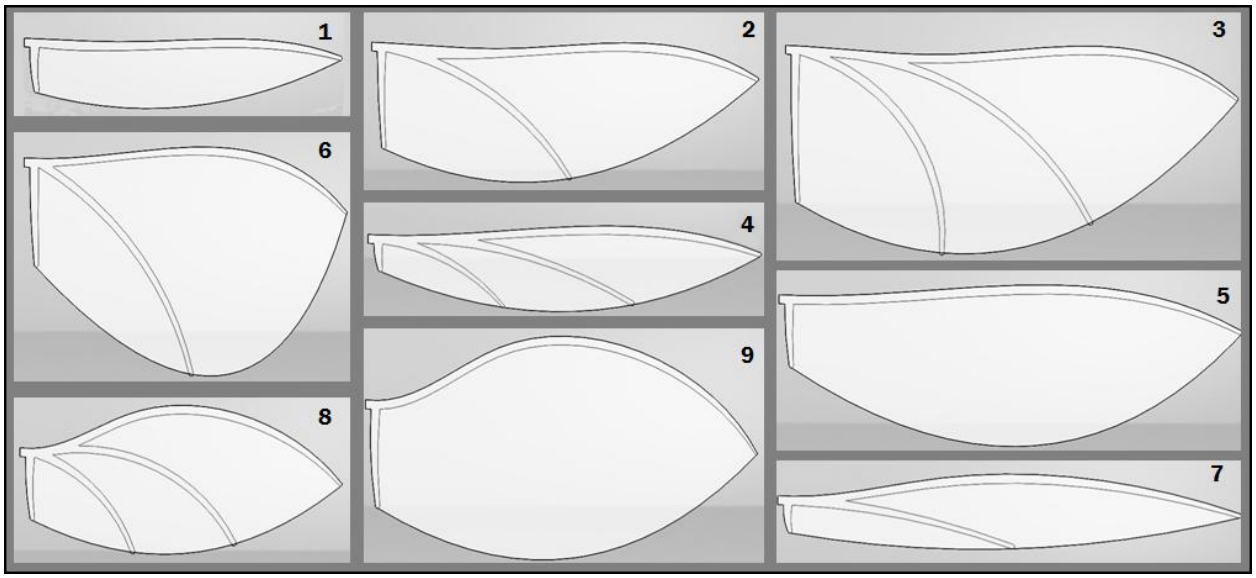

Figure 6. DOE wing models.

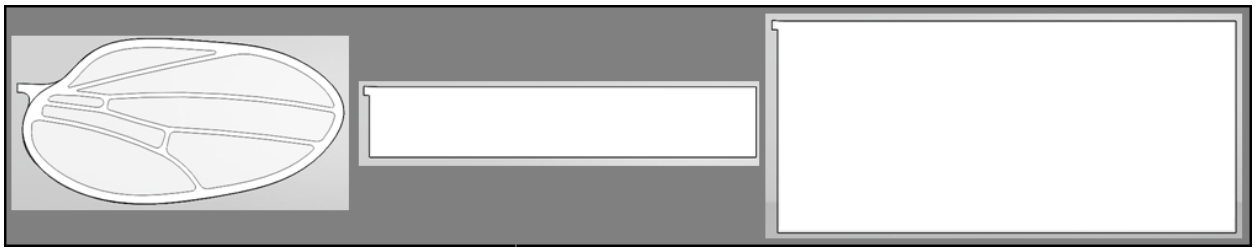

Figure 7. Additional wing models.

2.3 Wing Design and Manufacturing

Using the design specifications indicated from the DOE, we modeled each wing in SolidWorks. The wing structure comprised leading edge, the inner spar, and up to two outer spars, depending on the configuration. The wing structure was made of Vero White using an Object Eden260V 3D printer where two copies of each wing were created. A soft and flexible low-density polyethylene (LDPE) 0.03-mm-thick plastic was used for the wing skin, which was tautly stretched and taped to a solid surface. The LDPE is adhered to the wing structures using Loctite adhesive and allowed to set before the wing shape is cut using a razor blade.

There are many methods available to achieve a distinguishable speckle, including patterned paper adhesives and water-soluble temporary "tattoos." Spray paint is used to create the speckle for simplicity purposes and to maintain minimal skin thickness. Unfortunately, the effect of adding the paint to the LDPE is unknown since the wing must be painted to be tested. To achieve an even background, the wing is first coated with a thin layer of flat white spray paint, and then a fine mist of flat black spray paint is layered over the white to achieve the desired speckle pattern. A good speckle has dots similar in size and evenly distributed, and has about a high contrast, 50-50 ratio between black and white areas (*7*), as shown in figure 8.

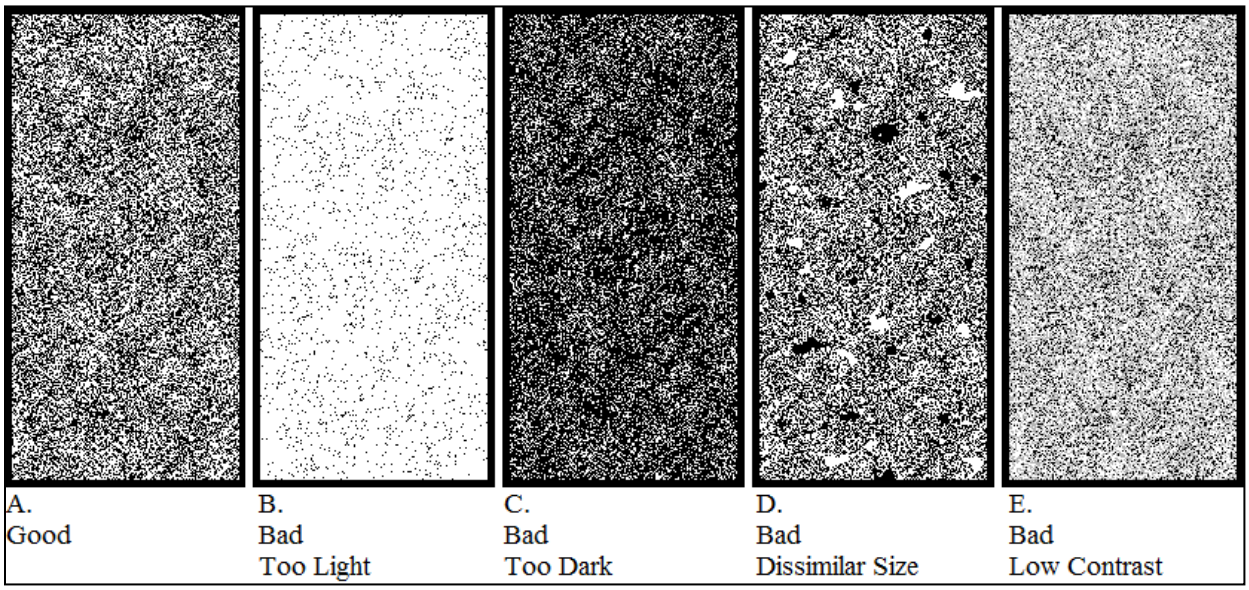

Figure 8. Speckle reference.

2.4 Experimental Setup and Process

Due to limited laboratory space, the DIC tests and load cell tests are run separately. For the DIC tests, the actuator is mounted on a clear, stable stand (figure 9) designed by Miles C. Pekala (*8*). The vibration of the actuator is tuned to match the wing's resonant frequency, causing the largest wing displacement. This frequency is noted for each wing design to ensure the same frequency is matched in the load cell portion of the test. The high-speed Photon cameras are set up at 20° angles to each other to view the specimen in stereo at 2000 fps. The cameras are calibrated, and after the wing motion is steady, about 2 s of data is collected, resulting in about 20 entire wing flap motions. This process is done for the 12 sets of wings. Once the frames are acquired, they are processed through the Vic-3D software, which will give strain, position, velocity, and acceleration data for each wing design.

The load cell portion of the experiment is set up using an optics board to mount the load cell. The actuator fits directly into the load cell mount, and again the vibration of the actuator's PZT material is adjusted to match the wing's resonant frequency. All of the load cell's strain gages are connected to the voltage meter to record the resulting thrust and lift generated from the flapping wings. Each set of wings is tested for about 10 s. and the voltage readings are recorded and saved.

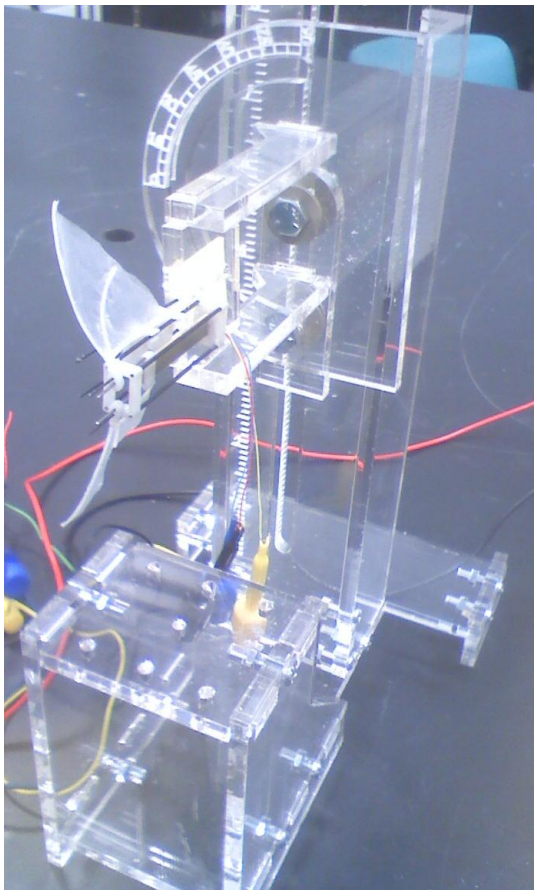

Figure 9. DIC wing mount.

3. Results and Discussion

Trends between the wing design parameters and performance will be examined and compared to previous research (*6*, *9*–*12*). Similar wing designs will also be tested using the test setup described here and to verify test performance results. In-depth computational analysis of the same wing designs will be able to validate project results, and providing an accurate model will be used to optimize and provide a final wing design.

4. Summary and Conclusions

The goals of this project are to gain a better understanding of small flapping wing motion. This is done by designing and documenting the process of analyzing small flapping wings with the DIC system and by building a load cell that can resolve minute stresses. This data will be used to validate a computational model that will assist in further design optimizations.

Specifically, there is interest in understanding the amount of strain, stress, and load experienced by the wing as well as what portion of the wing is under the most strain, stress, and load. Once the trends in the wing deformations are known, how the wing shape affects strain, stress, and loads will be determined. How the wing shape affects the wing position, velocity, and acceleration can be resolved, which will lead to the determination of the best wing shape.

The accuracy of this project data will be verified by comparing the results to computational models and simulations as well as historical data and other similar research projects. Upon completion of the testing and data analysis, we will recommend a final wing design and document the DIC testing process to aid in further testing.

5. References

1. Svanberg, C. E. *Biomimetic Micro Air Vehicle Testing Development and Small Scale Flapping-Wing Analysis*; Air Force Institute of Technology, Graduate School of Engineering and Management: Wright-Patterson Air Force Base, OH, March 2008.

2. Huber, A. F. *Death by a Thousand Cuts: Micro-Air Vehicles (MAV) in the Service of Air Force Missions*; Air War College Air University: Maxwell Air Force Base, AL, April 2001.

3. Vic-3D 2007 Testing Guide, Correlated Solutions. http://classes.engr.oregonstate.edu/mime/winter2009/me453-001/VIC-3D%20Testing%20Guide.pdf (accessed July 28, 2011).

4. Hall, A. *Overview of Piezoelectric Actuator Displacement Measurements Utilizing a MTI-2100 Fotonic Sensor*; ARL-TN-429; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, April 2011.

5. Riddick, J. C.; Hall, A. Functionally-Modified Bimorph PZT Actuator for cm-Scale Flapping Wing; *Proc. SPIE 7978* **2011**, 79780A.

6. Kumar, R.; Silin, E.; Shkarayev, S. *Experimental Study of Aerodynamic Performance of Locust and Model Wings*. 49th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, 2011.

7. Lecompte, D.; Smits, A.; Bossuyt, S.; Sol, H.; Vantomme, J.; Van Hemelrijck, D.; Habraken, A. M. Quality Assessment of Speckle Patterns for Digital Image Correlation. *Optics and Lasers in Engineering* **2006**, *44*, 1132–1145.

8. Hall, A; Pekala, M. C. Micro-Electro-Mechanical Flapping Wing Technology for Micro Air Vehicles, *IMAV 2010 Proceedings*, Braunschweig, Germany, July 2010.

9. Aono, H.; Chumakurthi, S. K.; Wu, P.; Sallstrom, E.; Stanford, B. K.; Cesnik, E. S.; Ifju, P.; Ukeiley, L.; Shyy, W. *Computational and Experimental Study of Flexible Flapping Wing Aerodynamics*. 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, 4–7 January 2010, Orlando, FL; AIAA 2010-554.

10. Bryson, D. E.; Weisshaar, T. A. *Aeroelastic Optimization of a Two-Dimensional Flapping Mechanism*. 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 12–15 April 2010, Orlando, FL; AIAA 2010-2961.

11. Norris, A. G.; Palazotto, A. N.; Cobb, R. G. *Structural Dynamic Characterization of an Insect Wing*. 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 12–15 April 2010, Orlando, FL; AIAA 2010-2790.

12. Wu, P.; Ifju, P. Flapping Wing Structural Deformation and Thrust Correlation Study With Flexible Membrane Wings. *AIAA JOURNAL* **September 2010**, *48* (9).

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Development of a Lumped Element Circuit Model for Approximation of Dielectric Barrier Discharges

THOMAS C. UNDERWOOD
MENTOR: BRYAN GLAZ
VEHICLE TECHNOLOGY DIRECTORATE
ABERDEEN PROVING GROUND, MD

Contents

List of Figures 153

List of Tables 153

Abstract 154

Acknowledgments 155

Student Bio 156

1. Introduction/Background 157

2. Theoretical Development 157

3. Results and Discussion 163

4. Summary and Conclusions 164

5. References 165

List of Figures

Figure 1. Electric circuit model of a dielectric aerodynamic plasma actuator.....158
Figure 2. Region of homogeneous potential, i.e., "hot spot."....158
Figure 3. Sketch of the air capacitive element.....159
Figure 4. Sketch of the dielectric capacitive element. 159
Figure 5. Input voltage, V_{app} vs. time.....161

List of Tables

Table 1. Experimental parameters. 163

Abstract

This work presents a circuit model for calculating the total energy dissipated into neutral species for pulsed direct current (DC) dielectric barrier discharge (DBD) plasmas. Based on experimental observations, it is assumed that nanosecond pulsed DBDs, which have been proposed for aerodynamic flow control, can be approximated by the two independent regions of a homogeneous electric field. An equivalent circuit model is developed for the homogeneous region near the exposed electrode, i.e., the “hot spot,” based on a combination of a resistor, capacitors, and a zener diode. Instead of fitting the resistance to an experimental data set, a formula is established for approximating the resistance by modeling a plasma as a conductor with DC voltage applied to it. Various assumptions are then applied to the governing Boltzmann kinetic energy equation to approximate electrical conductivity values for weakly ionized plasmas. The model is compared with experimental data sets of the total power dissipated by a plasma to validate its accuracy.

Acknowledgments

I wish to thank my mentor, Dr. Bryan Glaz, for the guidance and support he provided throughout the period of research. I would also like to acknowledge the support provided by Dr. Surya Dinavahi and Tomas Houba regarding the direction of research as well as the many enlightening discussions shared with them.

Student Bio

I am currently an undergraduate student attending the University of Florida (UF), majoring in both nuclear engineering and physics. At UF, I am a member of the Applied Physics Research Group headed by Prof. Subrata Roy, where I conduct research on plasma physics. In May of 2011, I completed his second year of studies at UF and was lucky enough to have an opportunity to work with plasma actuators on a project of interest for the U.S. Army. Upon completion of my undergraduate education, I hope to pursue the subject of applied plasma physics in graduate school in order to obtain a doctoral degree.

1. Introduction/Background

The need for improved control over aerodynamic flow separation has increased interest in the potential use of plasma actuators. The inherent advantages of plasma actuator flow control devices include minimal size and weight, lack of moving parts, cost efficiency, and low drag penalties. However, the actuators that affect the flow via directed momentum transfer are not effective at Mach numbers associated with most subsonic aircraft applications. Recently, Roupassov et al. (*1*) demonstrated that pulsed plasma actuators, in which energy imparted to the flow appears to effectively control flow separation, seem to be suitable at Mach numbers ($M \approx 0.3$) beyond the capabilities of the current plasma-induced momentum-based approaches.

Given the fundamental differences between the novel pulsed discharge approach and the more conventional momentum-based approaches, there is a need to develop an effective and efficient model for the energy delivered to the flow by the plasma. Once calculated, that value can be input to a computational fluid dynamics solver as an energy source term resulting in a coupled fluid/plasma dynamics model. Multiphysics models of this type are required in order to study detailed flow characteristics. However, detailed numerical calculations that are primarily used for isolated plasma simulations are not suitable for a variety of coupled fluid/plasma dynamic studies because of their excessive computational expense. To address this issue, efficient circuit element models have been introduced to approximate the complex dynamic processes within plasmas. However, models such as those by Orlov et al. (*2*) rely on empirical constants that are tuned to experiments which are not applicable to nanosecond pulsed discharges. To date, an approximate model of nanosecond pulsed plasma actuators has not been developed. This paper deals primarily with establishing a flexible model that could be implemented as an approximation for the energy dissipated within a plasma for any pulsed direct current (DC) dielectric barrier discharge (DBD) configuration. Among the other fundamental goals in this paper is to establish a deeper understanding of the background processes that occur within plasma and incorporate that knowledge into the model.

2. Theoretical Development

One of the primary assumptions in creating this model is that a small volume near the exposed electrode can be idealized as consisting of a homogeneous energy density. The “hot spot,” observed by Roupassov et al. (*1*), plays an important role in the power dissipated into the air and thus the resulting fluid dynamics. In this study, only the spatially homogeneous hot-spot region was considered. Therefore, the model presented in this paper consists of a single network containing a resistor, capacitors, and a diode.

As shown in figure 1, circuit elements that were used to model the plasma include an air capacitor C_a, a dielectric capacitor C_d, a resistor R_f, and a zener diode D_f. The air capacitor represents the capacitance between the dielectric surface and the exposed electrodes. The dielectric capacitor represents the capacitance between the dielectric surface and insulated electrodes and is proportional to the thickness of the dielectric layer. Thus the dielectric layer in the form of both its thickness and the value of its dielectric constant plays an important role in determining the effectiveness of the plasma actuator. Finally, the zener diode, introduced by Orlov et al. (*2*), is used in the model to enforce an energy threshold value below which plasma will not form.

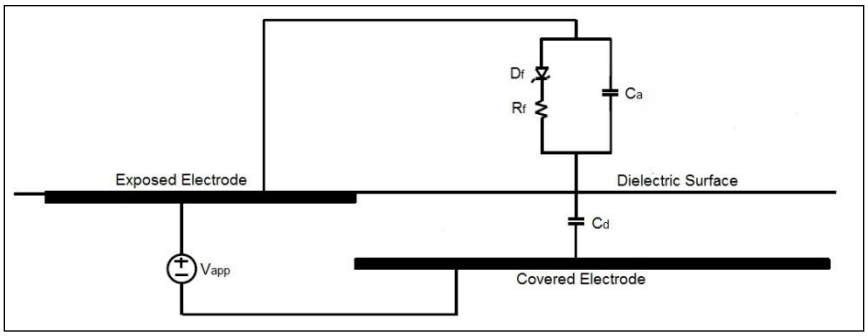

Figure 1. Electric circuit model of a dielectric aerodynamic plasma actuator.

Since a uniform charge distribution along the top of the dielectric is assumed, the typical asymmetric plasma actuator geometry featured in figure 1 can be simplified to a homogeneous symmetric region. This assumption results in a one-dimensional (1-D) model, and any variation along the horizontal direction depicted in figure 2 is neglected.

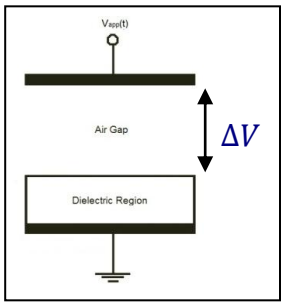

Figure 2. Region of homogeneous potential, i.e., "hot spot."

As displayed in figure 1, the lumped element circuit is a function of the two capacitance values, C_a and C_d. In this model, the air is treated as both a conductor to generate a physical relationship for the resistance R_f and a parallel plate capacitor to generate C_a. An advantage of modeling the plasma as a conductor in addition to a parallel plate capacitor is that it generates a physical relationship for the resistance, R_f—a value that is traditionally empirically determined. The air gap capacitor can be modeled as (*3*)

$$C_a = \frac{\epsilon_0 \epsilon_a A_a}{h_a}, \tag{1}$$

where A_a is the cross-sectional area of the air and h_a is the approximate height of the plasma region of interest. As displayed in figure 3, A_a is the product of the spanwise length of the actuator z_a, and l_a is the chordwise distance from the exposed electrode to the end of the dielectric region.

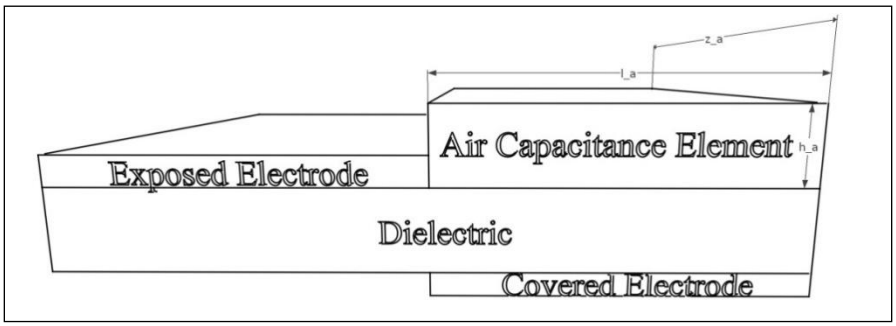

Figure 3. Sketch of the air capacitive element.

The capacitive element corresponding to the dielectric can be modeled as (*3*)

$$C_d = \frac{\epsilon_0 \epsilon_d A_d}{h_d}, \tag{2}$$

where A_d is the cross-sectional area of the dielectric capacitive element and h_d is the height of the dielectric barrier layer. As displayed in figure 4, A_d is the product of the spanwise length of the actuator z_a and d_d is the width of the dielectric region. Treating the plasma as a conductor, the resistance for DC voltage is proportional to the electrical conductivity, σ_p, as well as A_a, and h_a and can be given as (*4*)

$$R_f = \frac{h_a}{\sigma_p A_a}. \tag{3}$$

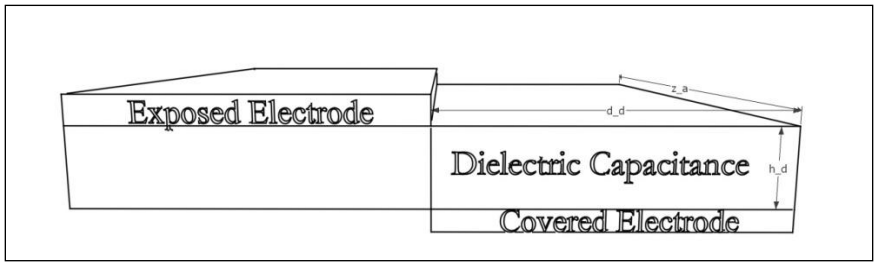

Figure 4. Sketch of the dielectric capacitive element.

Starting from Kirchhoff's laws (*3*), the governing differential equation for the voltage drop experienced by the air gap ΔV is given by

$$\frac{d\Delta V(t)}{dt} = -\frac{dV_{app}(t)}{dt}\left(\frac{C_a}{C_a + C_d} - 1\right) - \kappa \frac{\Delta V(t)}{R_f(t)(C_a + C_d)}, \tag{4}$$

and

$$\kappa = \begin{cases} 1 \; if \; |\vec{E}| > \overrightarrow{|E|}_{crit} \\ 0 \; if \; |\vec{E}| < \overrightarrow{|E|}_{crit} \end{cases}, \tag{5}$$

where V_{app} is the applied voltage and κ is the contribution from the zener diode. If the electric field magnitude, given as

$$|\vec{E}| = \frac{|\Delta V|}{h_a}, \tag{6}$$

is greater than some threshold value, then κ takes on a value of one, otherwise it is zero to signify that plasma has not formed. For nanosecond high-voltage pulses, the applied electric field will virtually always be over the critical breakdown field (*1*), thus κ can be set equal to one for the case considered in this study.

The energy imparted to neutral species from the plasma can be calculated from knowledge of the voltage drop over the air gap by

$$E(t) = \int_0^t \frac{\Delta V}{R_f(t)} dt. \tag{7}$$

To effectively calculate the resistance governed by equation 3, an expression must first be developed for the electrical conductivity of the plasma. This value is one that traditionally requires a numerical approach. To simplify the problem to a point where an analytic formulation can be used, numerous simplifying assumptions were used and are described in the following paragraphs.

For any plasma, the resulting electric current is composed of two primary terms: the current from electrons and that from ions. As the drift velocity, $\boldsymbol{w_e}$, which represents the velocity induced by an electric field, is significantly higher in a nonequilibrium plasma for electrons compared to ions, the current density can be approximated as only the portion from electrons as long as the number densities, N_e and N_i, are approximately the same (*4*). When a form of the generalized Ohm's law is used, the current density vector $\boldsymbol{J}$ and plasma conductivity, respectively, can be written as

$$\boldsymbol{J} \approx -eN_e\boldsymbol{w_e} = \sigma_p\boldsymbol{E}, \tag{8}$$

and

$$\sigma_p = e(N_e\mu_e + N_i\mu_i), \tag{9}$$

where μ_i and μ_e represent the ion and electron mobilities, respectively. Much like equation 8, the electrical conductivity relation can be simplified using the concept of quasineutrality, which is defined as having approximate equal number densities of both ions and electrons. Thus, as μ_e is typically approximately three orders of magnitude larger than μ_i for any plasma that is not in thermal equilibrium, if N_e is at least the same order of magnitude as N_i, it is a good assumption to approximate the electrical conductivity as only coming from electrons (*4*). Quasineutrality itself

is a typical assumption that is valid as long as the plasma being modeled is far away from the cathode to avoid the boundary layer in plasma physics called the sheath.

Since a pulsed DC voltage is assumed, the activation of the external electric field will follow the voltage waveform as a step function. Thus two expressions will be required for the σ_p, where the first is valid for the period when a constant external electric field is applied over the peak of the voltage pulse, as shown in figure 5, from 13–40 ns, and the second when the external voltage is zero. For the portions of the voltage waveform that the voltage is zero, the power is also zero according to Ohm's law, and thus the conductivity during this time is of no importance.

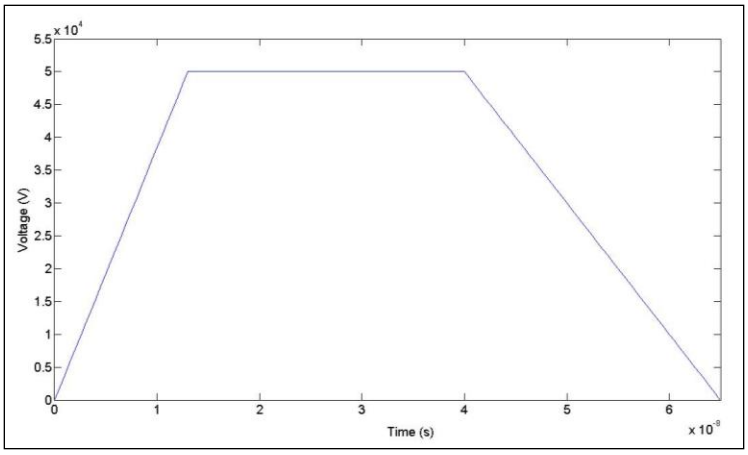

Figure 5. Input voltage, V_{app} vs. time.

Given an assumption of a constant electric field during the peak, an analytic formulation can be generated for the electrical conductivity with knowledge of $\boldsymbol{w_e}$ (*5*) by starting with the Boltzmann equation for electrons written as

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}}{m} \cdot \frac{\partial f}{\partial \boldsymbol{r}} = \frac{df}{dt}, \tag{10}$$

where *df*/*dt* is typically called the collision integral and refers to the variation in the number of particles within some phase space control volume. In order to approximate this, a relaxation time, τ, is introduced, which is defined as the time taken for the system to be reduced to an equilibrium distribution function. This approximation can be given as $\tau \approx (N_a \sigma_{ea} \nu)^{-1}$, where σ_{ea} is the collision cross section, ν is the average collision velocity, and N_a is the number density of atoms (*5*). According to Smirnov (*5*), an analytic formulation is obtained by assuming that the force term can be given as the Lorentz force,

$$\boldsymbol{F} = -e\boldsymbol{E}\exp(-i\omega t) - \frac{e}{c}(v \times \boldsymbol{B}), \tag{11}$$

where ω is the frequency the electric field, $\boldsymbol{B}$ is the magnetic field and assumed to be directed along the z-axis (vertical), and $\boldsymbol{E}$ is assumed to belong only in the x(horizontal)-z plane. Taking

the tau approximation of electron kinetic energy equation and integrating over the electron velocity, we can then write the equation of motion for an electron as

$$m_e \frac{d\boldsymbol{w_e}}{dt} + m_e \frac{\boldsymbol{w_e}}{\tau} = -e \exp(-i\omega t) - \frac{e}{c}(\nu \times \boldsymbol{B}). \quad (12)$$

Substituting the cyclotron frequency $\omega_H = eB/m_e c$, $a_x = -eE_x/m_e$ and $a_z = -eE_z/m_e$, the following scalar equations representing the different components of equation 12 can be produced:

$$\frac{dw_x}{dt} + \frac{w_x}{\tau} = a_x \exp(-i\omega t) - \omega_H w_y, \quad (13)$$

$$\frac{dw_y}{dt} + \frac{w_y}{\tau} = \omega_H \omega_x, \quad (14)$$

and

$$\frac{dw_z}{dt} + \frac{w_z}{\tau} = a_z \exp(-i\omega t). \quad (15)$$

Equations 13–15 can be solved to obtain three component expressions for the electron drift velocity. If a constant electric field and a negligible magnetic field are assumed, ω and ω_H can be set equal to zero and the resulting vector equation can be obtained:

$$\boldsymbol{w_e} = -\frac{e\boldsymbol{E}\tau}{m_e}. \quad (16)$$

When equation 16 is solved in conjunction with equation 8 for σ_p, the following expression independent of the applied electric field can be obtained:

$$\sigma_p = \frac{N_e(t)e^2}{m_e N_a \sigma_{ea} v}. \quad (17)$$

In equation 17, σ_{ea} is a function of the electron temperature, T_e, and can be obtained for various molecules found in air (*6*). As shown in equation 17, the inputs required for this approximation include N_e , the number density of electrons, N_a, and T_e. Among these values, N_a is assumed to be constant in time as the number density of atoms is significantly higher than that of free electrons. T_e is solved for quantity from the energy transport equation and changes primarily with reduced electric field strength, E/N_a. Results for the electron temperature are generated from BOLSIG+, a program aimed at solving for the transport properties of electrons.

The final required input, the electron number density N_e, is traditionally solved from the drift diffusion approximation of the conservation of momentum equation. Starting from the species continuity equation for electrons and neglecting diffusion, an equation for the time rate of change in electron density can be obtained:

$$\frac{dn(t)}{dt} = \alpha \left| n\mu_e \vec{E} \right| - \beta n^2, \quad (18)$$

where α determines the ionization frequency and β establishes the recombination rate of electrons. As equation 18 is function of the electric field experienced by a plasma, a complete

system of equations will require equations 4, 6, 17, and 18 solved in a coupled manner. Equation 18 is also a function of empirical coefficients α and β which are both dependent on the pressure, p, of the system and can be given by

$$\alpha = Ap \exp\left(\frac{-Bp}{\vec{E}}\right), \tag{19}$$

and

$$\beta = C\left(\frac{300}{T_e}\right)^{\frac{1}{2}}, \tag{20}$$

where A, B, and C are tabulated values for air at atmospheric pressure. According to Raizer (*7*), appropriate values based on experimental measurements are: 38.1 (in·Torr)$^{-1}$ (15 [cm·Torr]$^{-1}$) for A, 927.1 V/(in·Torr) (365 [V/(cm·Torr)]) for B and 1.2 $\times 10^{-8}$ in^3/s (2×10^{-7}) cm^3/s for C.

3. Results and Discussion

In order to validate the accuracy of the model described in this paper, comparisons with data presented in Roupassov et al. (*1*) are provided. The experimental parameters that were mentioned and used in the circuit model are given in table 1.

Table 1. Experimental parameters.

h_a	0.016 in (0.4 mm)
h_d	0.01 in (0.3 mm)
ϵ_d	2.7
ϵ_a	1
A_n	0.05 in^2 (30 mm^2)
A_d	0.05 in^2 (30 mm^2)
V	50 kV
T	50 ns

An approximation of the 50-kV applied voltage pulse employed by Roupassov et al. (*1*) was considered for this study and is shown in figure 5.

For an E/N_a value of 4921 Td, BOLSIG+ calculates that the mean electron energy should be ~1.18 $\times$ 10^{-20} Btu (78 eV). Using this result, we calculated a peak number density of ~6.5 $\times$ 10^{17}1/ft^3 (2.3 $\times$ 10^{19}1/m^3) using equation 18 coupled with equation 4 where N_0, the number density before the pulse, was set equal to 2.83 $\times$ 10^{13}1/ft^3 (10^{15} 1/m^3). The height of the plasma, h_a, was determined to be a constant 0.02 in (0.4 mm) over time based on experimental observations by Roupassov et al. (*1*). Thus using the calculated values for $N_e(t)$ and T_e as well as the experimentally observed h_a, we obtained a value of 2.3 $\times$ 10^{-6} Btu (2.6 mJ) by using

equation 7. Gas heating results from Roupassov et al. (*1*) correspond to a total energy of 4.2 mJ that was imparted to air per 65 ns. Therefore, the approximate model presented in this study provides an answer with an absolute error of ~38%. The accuracy of higher-fidelity numerical models based on discretization of the drift diffusion equations has not been established.

4. Summary and Conclusions

A new lumped element circuit model was presented that is valid for any pulsed DC DBD plasma. In addition, an approximate expression was formulated to calculate the resistance value for the air gap as a function of the conductivity of the plasma. Results of the model were verified against a pulsed DC experiment conducted by Roupassov et al. (*1*), and an order of magnitude agreement was obtained for the energy imparted into the plasma in a homogeneous "hot spot" region. Future work will focus on extending the circuit model to account for the plasma region outside of the "hot spot" as well as determining the geometric dimensions of the plasma regions as functions of the applied voltage.

5. References

1. Roupassov, D. V.; Nikipelov, A. A.; Nudnova, M. M.; Starikovskii, A. Y. Flow Separation Control by Plasma Actuator with Nanosecond Pulse Periodic Discharge. *Proceedings of the 17th International Conference on Gas Discharges and Their Applications,* 7–12 September 2008; pp 609–612.
2. Orlov, D.; Corke, T.; Patel, M. Electric Circuit Model for Aerodynamic Plasma Actuator. AIAA Paper 2006-1206, 2006.
3. Griffiths, D. J. *Introduction to Electrodynamics*; Prentice Hall: Upper Saddle River, NJ, 1999.
4. Mitchner, M.; Kruger, C. H. *Partially Ionized Gases*; John Wiley Sons, Inc.: Indianapolis, IN, 1973.
5. Smirnov, B. M. *Physics of Ionized Gases*; John Wiley: New York, 2001.
6. Phelps, A. Compilation of Electron Cross Sections Used by A. V. Phelps. http://jila.colorado.edu/~avp/collision_data/electronneutral/ELECTRON.TXT (accessed 27 July 2011).
7. Raizer, Yu P. *Gas Discharge Physics*; Springer: Berlin, Germany, 1991.

INTENTIONALLY LEFT BLANK.

Weapons & Materials Research Directorate (WMRD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Microstructural Investigation and Evaluation of Mechanical Properties in Friction Stir Welded Joints

BRIAN JUSTUSSON
MENTORS: DR. CONSTANTINE FOUNTZOULAS AND DR. CHIAN-FONG YEN
U.S. ARMY RESEARCH LABORATORY, WMRD, RDRL-WMM-B,
ABERDEEN PROVING GROUND, MD 21005-5069

Contents

List of Figures **171**

List of Tables **171**

Abstract **172**

Acknowledgments **173**

Student Bio **174**

1. Introduction/Background **175**

2. Experiment/Calculations **176**

2.1 Material 176

2.2 Grinding/Polishing Method 177

2.3 Tensile Testing Method 177

2.4 Microscopy/EBSD Method 177

3. Results and Discussion **178**

3.1 Tensile Results 178

3.2 Microscopy: Grain Structures 179

3.3 Electron Backscatter Diffraction: Grain Analysis of Upper Weld Nugget 181

3.4 Observations on Increased Ductility 182

4. Summary and Conclusions **185**

4.1 Summary 185

4.2 Future Work 185

5. References **186**

List of Figures

Figure 1. FSW tool. 175
Figure 2. Tensile results in X-direction. 178
Figure 3. Tensile results in the Y-direction. 178
Figure 4. Tensile results in Z-direction. 179
Figure 5. Post-etching image of weld. 180
Figure 6. Post-etching image highlighting specific regions. 180
Figure 7. Transition between weld nugget and TMAZ. 180
Figure 8. Transition inside of weld nugget and TMAZ. 180
Figure 9. Fine equaixed grains observed in the weld nugget. 181
Figure 10. Transition between weld nugget and TMAZ. 181
Figure 11. TMAZ zone showing elongated grains. 181
Figure 12. The as-imaged SEM. 182
Figure 13. Identification of grain boundaries. 182
Figure 14. EBSD results. 182
Figure 15. Results of grain analysis. 182
Figure 16. Strain distribution evolution in 2X from center weld. 183
Figure 17. Strain distribution evolution in 2Z from center weld. 183
Figure 18. Strain distribution evolution in 11X. 184

List of Tables

Table 1. FSW parameters. 176
Table 2. Polishing procedure. 177

Abstract

With involvement of the United States Military in various places in the world, increased improvements in personnel and vehicle protection are needed. Traditionally, armoring techniques have performed well against particular threats; however, a new threat has emerged from improvised explosive devices (IEDs), which has exposed the vulnerability of the undercarriage of the vehicle to explosions. Various solutions have been proposed to combat this problem, such as implementation of a unibody construction of the chassis. While traditional joining techniques have shown promise, the variability in mechanical properties of the weldment can be costly and needs to be addressed. Friction Stir Welding (FSW) is a solid-state welding technique, which involves local softening of the material subjected to severe plastic deformation, coupled with a localized heat flux, creates a unique microstructure in and around the weld that is dependent on the FSW parameters. The current work involves the characterization of the microstructures that develop during FSW as it relates to the locally spatial mechanical properties in the weld zone. The microstructure will be characterized by using scanning electron microscopy (SEM) and electron backscatter diffraction (EBSD) techniques, which, when coupled with mechanical testing data, can be used to shed light on the spatial stress-strain behavior of the welded sample.

Acknowledgments

I am grateful for the support of all the employees of the U.S. Army Research Laboratory (ARL), but in particular of Dr. Constantine Fountzoulas and Dr. Chian-Fong Yen for their continued mentorship and encouragement to become a better research scientist. I am also grateful for the training and general help provided by Dr. Jian Yu, Dr. Brian Shuster, and Dr. Tomoko Sano. Additional help was provided by David Gray, Theresa Dillon, Ashiq Quabili, and Seth Ghiorse. A special thanks goes to Jessica Medintz, who was a valued partner on this work, and Dr. Mark Pankow of ARL and The University of Michigan for helping train my mind to be the research scientist I am today.

Student Bio

I am a third-year Ph.D. pre-candidate at the University of Michigan in Mechanical Engineering, mentored by Prof. Anthony Waas. I currently hold a master's degree in Mechanical Engineering from The University of Southern California and a bachelor's degree from The University of Michigan in Mechanical Engineering. My current research involves the rate-dependent mechanical properties of composites materials, but also in some metals. After completion of my Ph.D. work, my objectives are to continue expanding my current research interests as part of a recognized institution and work in a research laboratory in either the public or private sector, but also, at some point in the future, become a collegiate lecturer.

1. Introduction/Background

Aluminum is becoming an increasingly desirable structural metal for replacing steel due to its lower density, 2.67g/cc and 7.87 g/cc, respectively, with comparable specific bending stiffness. To date, armored vehicles rely heavily on steel armor, which increases the vehicle payload. To replace steel with light metals, there is a need for new lightweight personal and vehicular armoring techniques. The creation of a unibody aluminum chassis for armored vehicles can greatly reduce the weight of the vehicle; however, a joining process is needed.

Friction stir welding (FSW) is a solid-state joining technique developed in 1991 (*1*) and currently used extensively in aluminum alloys. This process involves the use of a welding tool, which consists of concentric cylinders composed of a thinner cylinder, known as a pin, and a wider cylinder known as a shoulder (figure 1). The pin is pushed into the material after a critical temperature is reached and the flow stress is easily overcome due to local softening as a result of the heat flux and the tool can be pulled through the material. As the tool is pulled through the material, distinct and unique microstructures develop.

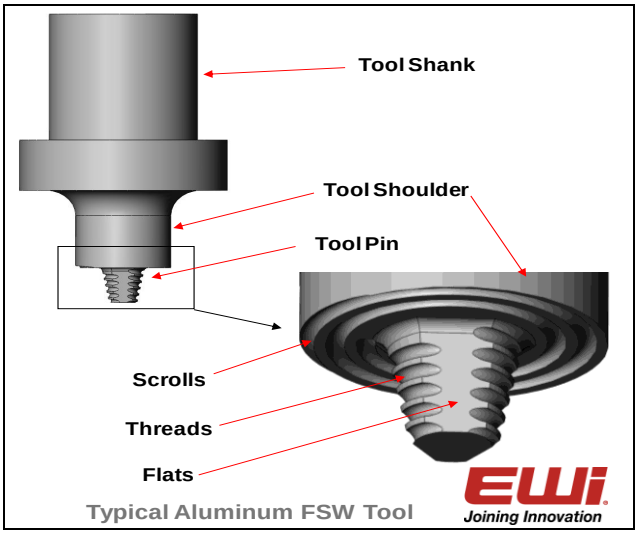

Figure 1. FSW tool.

These microstructures are generally referred to as the Heat Affected Zone (HAZ), which is a result of the heat dissipating through the material, which promotes grain growth, or the Thermo-Mechanically Affected Zone (TMAZ), which is a result of competing mechanisms due to the local heat flux promoting grain growth and localized severe plastic deformation, which results in decreased grain size. A final unique microstructure, termed the weld-nugget, is a fine, equiaxed grain structure located near the center of the weld. These zones have been shown to display unique mechanical properties (*2*), where the yield strength and ultimate tensile strength can be reduced while the ductility remains largely constant. Studies have shown that the microstructures

are a result of dynamic recrystallization and are very dependent on the stir parameters. Hirata et al. (*3*) showed that grain size decreases with weld velocity and increases with rotational speed. As of today, weldability and survivability of welds of advanced aluminum alloys subjected to blast loading have not been established. There is a need to accurately simulate FSW processes to create predictive models for blast loading.

Finite element models have been created to capture the FSW process. Sequential work by Fratini and Buffa (*4*–*6*) has been able to accurately simulate the various zones in welds of a quarter inch, while capturing the effects of recrystallization. Additional work by Grujicic et al. (*7*–*8*) was able to simulate FSW of AA-5083 and AA-2139 and clearly identify the development of microstructures. The developed model has been validated through experimental hardness tests (*9*). These models predict the results well, but are very computationally expensive.

The goal of this work is to understand the property distribution in and around the weldment for implementation into the finite models to reduce computational costs and develop a representative response of the weld zone. This will allow for development of a cohesive element to be used in all future simulations. The current work reports the following:

1. The spatial stress-strain response and mechanical properties models
2. The spatial distribution of the grains, grain size, and grain structure
3. Identification of microstructures that lead to variations in spatial mechanical properties

2. Experiment/Calculations

2.1 Material

Two plates of Aluminum 2139 were welded together as a butt joint using the parameters shown in table 1, which have been shown to be optimal parameters in past welds. The single plate was welded at EWI in Dayton, OH. The FSW tool was a two-piece tool with four flats and a left-handed thread, similar to the one shown in figure 1. Global tensile and bending tests were performed at EWI for the weld and were determined to be within specification for weld quality.

Table 1. FSW parameters.

Parameter	Specification
Shoulder diameter	1.625 in
Pin length	0.972 in
Plunge depth	0.65–0.005 in
Spindle speed	150–250 RPM
Travel speed	2 IPM
Total length	18 in

2.2 Grinding/Polishing Method

Polishing was completed using a Stuer RotoPol-31 fitted with Stuer Rotoforce 4 automatic polisher head using the steps shown in table 2. The grinding steps involved sandpaper with the given grit size. The polishing steps for 6 and 3 microns were performed on trident polishing cloths with water-based diamond suspension solutions. The final step for 0.05 micron was performed using colloidal silica. Between steps, the specimens were examined for scratches and cleaned using ethanol. A sodium hydroxide (NaOH) etchant was used for an application time of 1.5 min.

Table 2. Polishing procedure.

Step Type (Grit)	Time (min)
Grinding (120 paper)	Until flat
Grinding (320 paper)	1:30
Grinding (600 paper)	6:00
Grinding (1200 paper)	8:30
Polishing (6 um)	10:00
Polishing (3 um)	15:00
Polishing (0.05 um)	2:00

2.3 Tensile Testing Method

Tensile testing was performed on a screw-driven Instron 1125. Dog bone samples from the plate discussed in section 2.1 were cut using electric discharge machining (EDM) at various places in the weldment. Specimens were cut along the weld (Y), perpendicular to the weld (X), and through thickness (Z). The test was displacement controlled at a rate of 0.004 in/min (0.1 mm/min) to create a quasi-static loading condition. Photos were taken every 15 s using a Nikon D300. The resulting force/displacement data were analyzed and correlated to the photographs using ARAMIS digital image correlation software.

2.4 Microscopy/EBSD Method

Polished samples were examined for grain structures as a function of distance from the weld. Samples were imaged using a Nikon EPIShot 300. Electron backscatter diffraction (EBSD) was performed using a Nova NanoSEM 600. Images were taken with a step size of 1 micrometer and an image space of 250x250 micrometers was used. Images were analyzed using software ,which allows for pole analysis to show how the grains are oriented and textured.

3. Results and Discussion

3.1 Tensile Results

The results of the mechanical testing showed an interesting behavior, as seen in figures 2, 3, and 4. Figure 2 shows the tensile results in X or transverse direction, whereas figure 3 shows the Y or longitudinal direction and figure 4 shows the through thickness results. Of particular interest, groups 2 and 3 show the highest failure strain, but the lowest ultimate tensile strength. This is directly a result of where the samples were taken from. Sample 2 is taken inside of the weldment, whereas sample 3 was taken from an area along the weldment but contained both areas of the weldment and outside.

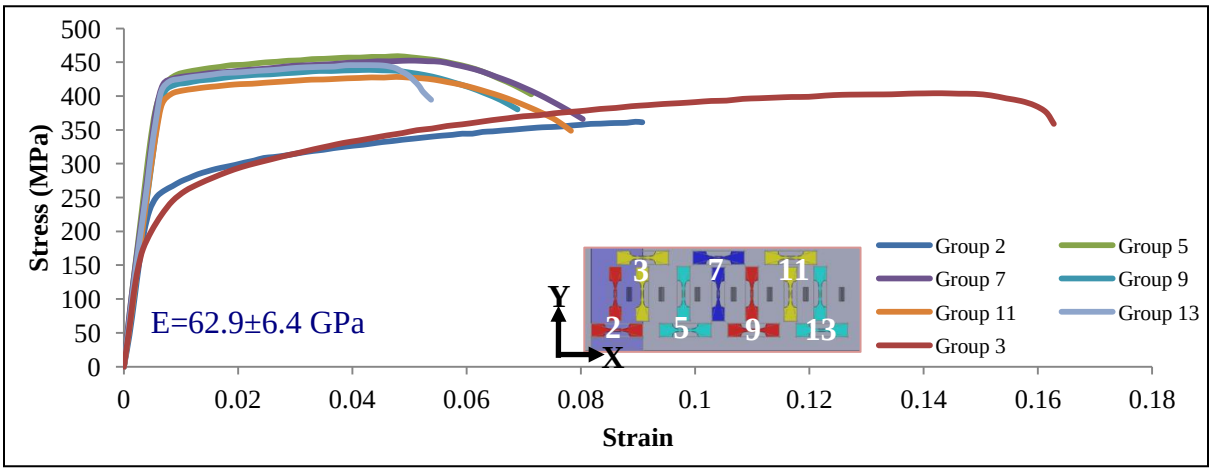

Figure 2. Tensile results in X-direction.

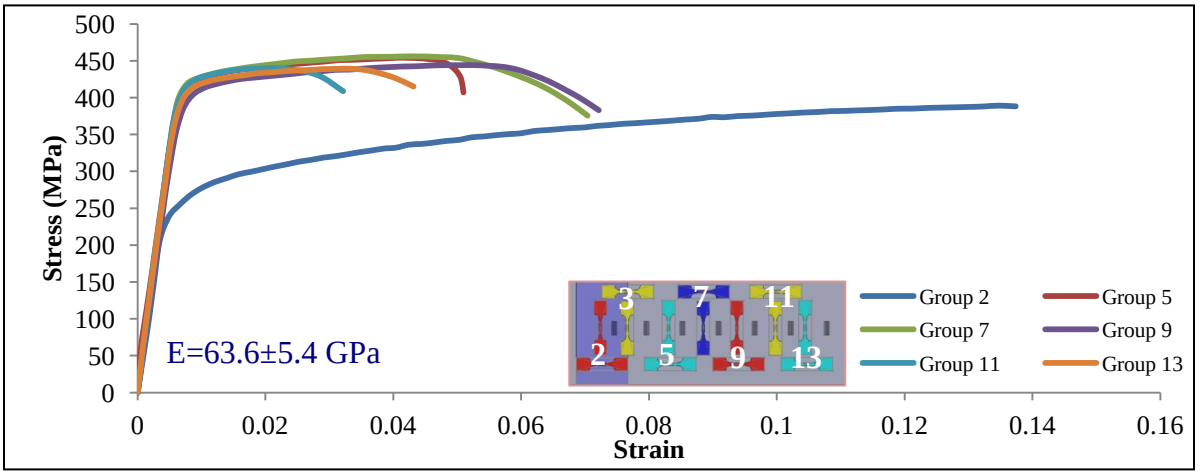

Figure 3. Tensile results in the Y-direction.

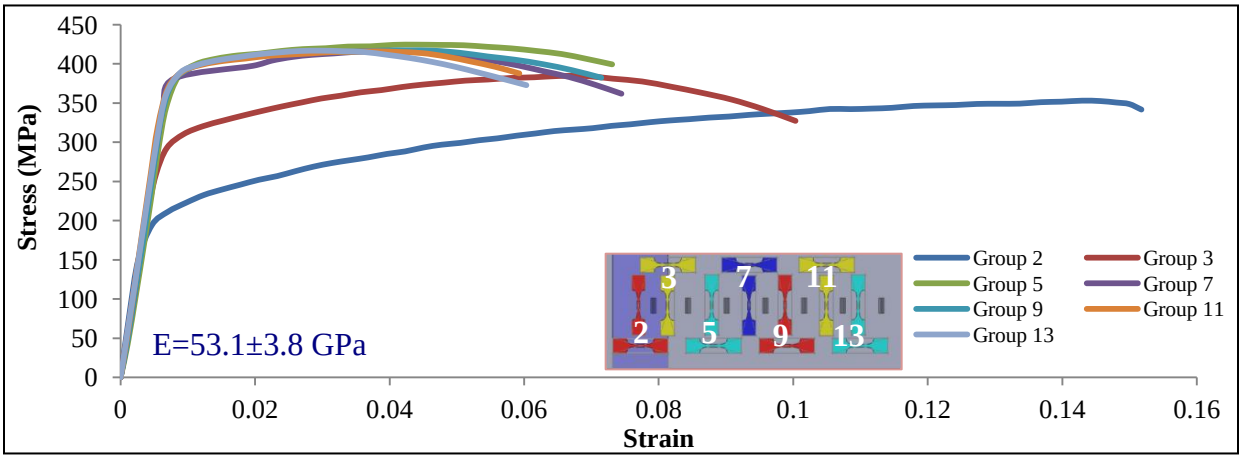

Figure 4. Tensile results in Z-direction.

Conversely, groups 5 and higher showed remarkable similarities in behavior in failure strain as well as ultimate tensile strength and elongation to failure in all directions. This would indicate that there are only spatial considerations for mechanical properties in the direct area of the weld.

A general trend was noticed that the modulus in the through thickness (Z) direction was the lowest in all the samples. This is most likely a result of the texturing in the grains that develops during cold working (*10*) of the plates prior to welding and the texturing that is created by the severe plastic deformation that is created through the welding process (*11*).

3.2 Microscopy: Grain Structures

From the tensile results, the entirety of the center weld was polished and etched, and is shown macroscopically in figure 5. From these images, the various microstructural zones are shown in figure 6, which is a result of different etching rates due to changes in grain sizes. In figure 6, sections A and B show the microstructures that develop in the weld nugget; however, A and B show different grain sizes. This is a result of the heat flux gradient that is generated during welding since previous studies have shown that as much as 86% of the heat flux is generated from the shoulder of the tool (*12*). For this reason, finer grains are observed in section A due to less grain growth occurring as a result of less heat. Section C is the TMAZ, D is the HAZ, and E is the unaffected material. The transition between structures is sometimes very distinct, as between A/B and A/C, but is harder to distinguish, as in the case of C/D.

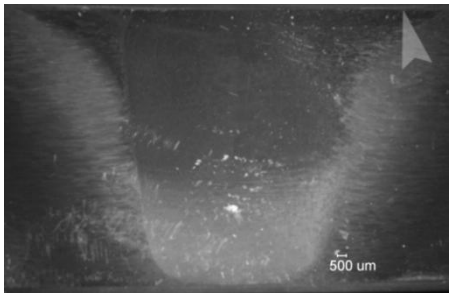

Figure 5. Post-etching image of weld.

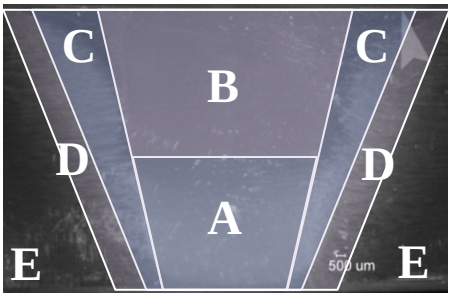

Figure 6. Post-etching image highlighting specific regions.

Examinations of the interfaces were done to show the distinct characteristics in the transition between zones and are shown in figures 7 and 8. Figure 7 shows a stark contrast between the very fine grains shown in section A of the weld nugget and the coarse grains of the TMAZ. This creates a very undesirable situation in which there is a large mismatch in mechanical properties. This mismatch during mechanical loading will cause an elevated local stress and likely lead to failure. The differences in A to B of the weld nugget are subtle, as shown in figure 8, but will still create an increased local stress at the interface with the TMAZ.

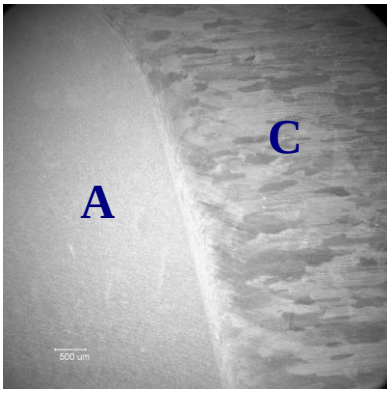

Figure 7. Transition between weld nugget and TMAZ.

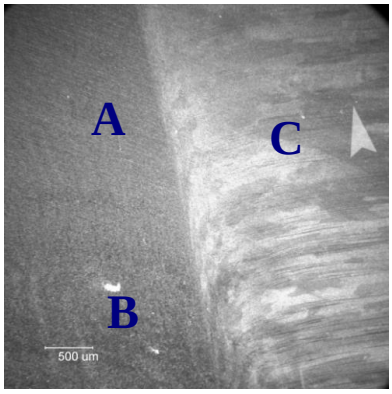

Figure 8. Transition inside of weld nugget and TMAZ.

Finally, the interface was once again examined to determine the relative difference in grain structures between the interfaces, as shown in figure 10. From figure 9, the fine equiaxed grains have an average grain size around 10–20 microns, while partially elongated grains on the major axis of the TMAZ, shown in figure 11, have a significantly larger grain size of over 100 microns. This once again shows a large and sudden change in the grain structure, which would lead to an elevated localized stress.

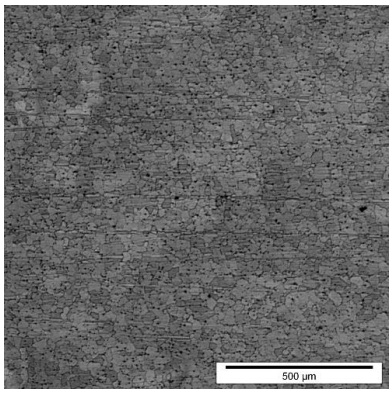

Figure 9. Fine equaixed grains observed in the weld nugget.

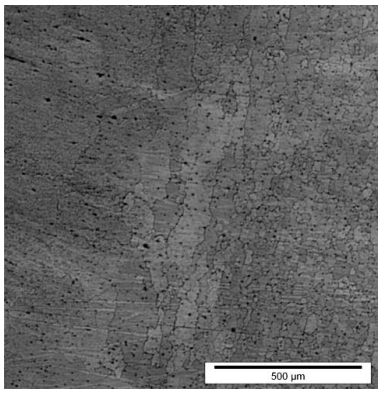

Figure 10. Transition between weld nugget and TMAZ.

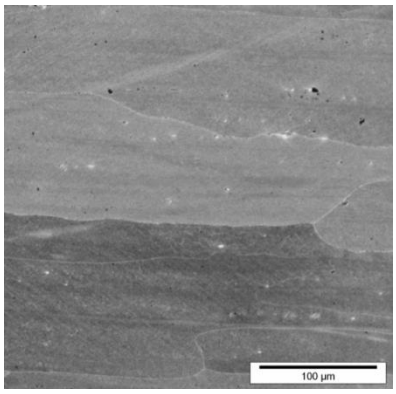

Figure 11. TMAZ zone showing elongated grains.

The existence of higher stresses caused by dissimilar mechanical properties can be seen in the results of the tensile test for the samples from region 2 and 3, shown in figures 2 through 4. From the current study, tensile specimens in the X-direction from region 2 would traverse the entire length of the weld, which would result in a very fine equiaxed grain structure through the gauge section, as that shown in figure 9. The Hall-Petch relationship indicates that yield stress would be higher than that of large grained materials shown in regions 5–13 (*14*). This, however, is not observed and is in fact calculated as a lower value. This would indicate that the yield strength, as characterized by Hall-Petch, must have been reached in order to cause plastic flow in the gauge section. The mismatch in material parameters causes elevated local stresses at the interface, which allows failure to occur.

3.3 Electron Backscatter Diffraction: Grain Analysis of Upper Weld Nugget

EBSD was performed on a sample taken from the *x-y* plane in upper part of the weld nugget. The as-imaged scanning electron microscopy (SEM) photo is shown in figure 12. Scratching of the sample was observed; however, it did not affect the ability to detect the grains, as shown in figure 13, which accurately mapped the grain boundaries. Figure 14 shows the results of the EBSD analysis. This shows that the grains are slightly elongated in the upper part of the nugget in the direction of the tool path. This is most likely a result of a dynamic recrystallization process taking place as the tool is passing through. This would result in texturing of grains in the Y-direction, which would result in directional hardening. These results are consistent with the findings of Medintz (*13*), the elongation to failure and ultimate tensile strength are higher for the X-direction, while modulus stays generally intact.

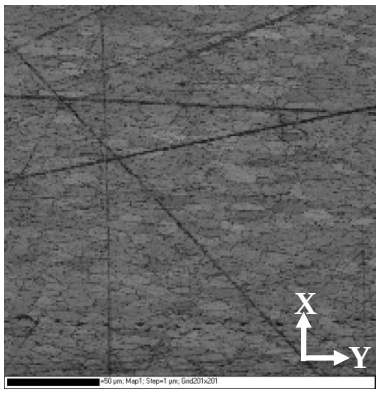

Figure 12. The as-imaged SEM.

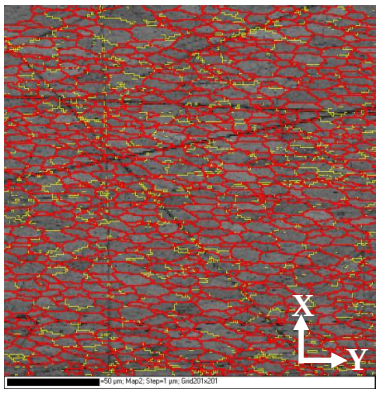

Figure 13. Identification of grain boundaries.

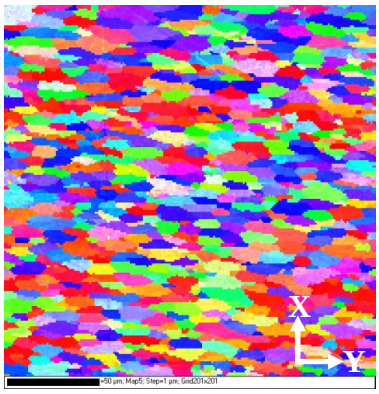

Figure 14. EBSD results.

The results of the pole analysis shown in figure 15 showed that the grains were generally well distributed in the crystallographic directions; however, there was slight preferential alignment toward the 101 and 111 directions. This may account for the anisotropic behavior of the tensile testing. The Z-direction shows a slight preferential orientation towards the 101 direction. This accounts for the lower modulus observed in the Z-direction as reported by Medintz (*13*). This is due to the fact that the single crystal of aluminum shows a highest modulus along the 111 direction (*14*). Having fewer grains aligned in this direction would lead to an overall lower relative modulus. This accounts for some of the issues related to the anisotropy found in testing.

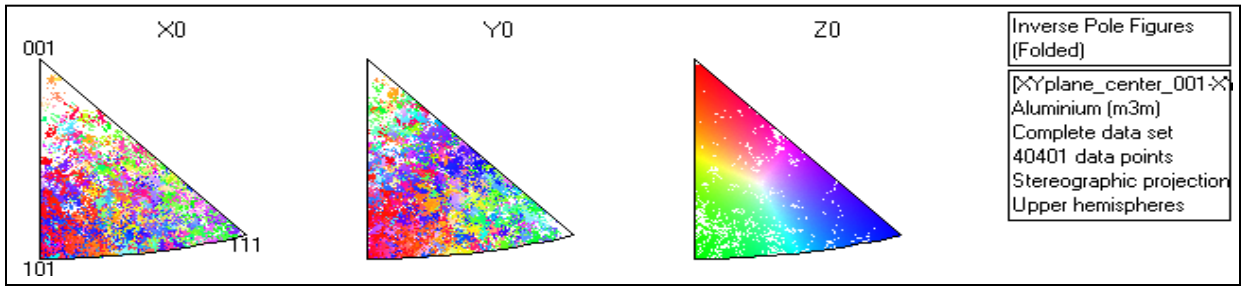

Figure 15. Results of grain analysis.

3.4 Observations on Increased Ductility

As seen from figures 2 through 4, the area in and around the weldment, shown in regions 2 and 3, exhibits the largest failure strain and highest ductility. When taking the strain fields from ARAMIS and plotting them during the elastic region, this phenomenon can be explained by the aforementioned work, in particular, the difference in microstructure. After plasticity occurs, geometric instabilities develop, which does not accurately characterize the strain distribution. These results are shown in figures 16 through 18. When comparing the results of the tensile test to the strain profiles along the tensile direction, it can be seen that this is a result of the microstructural distribution along the tensile axis.

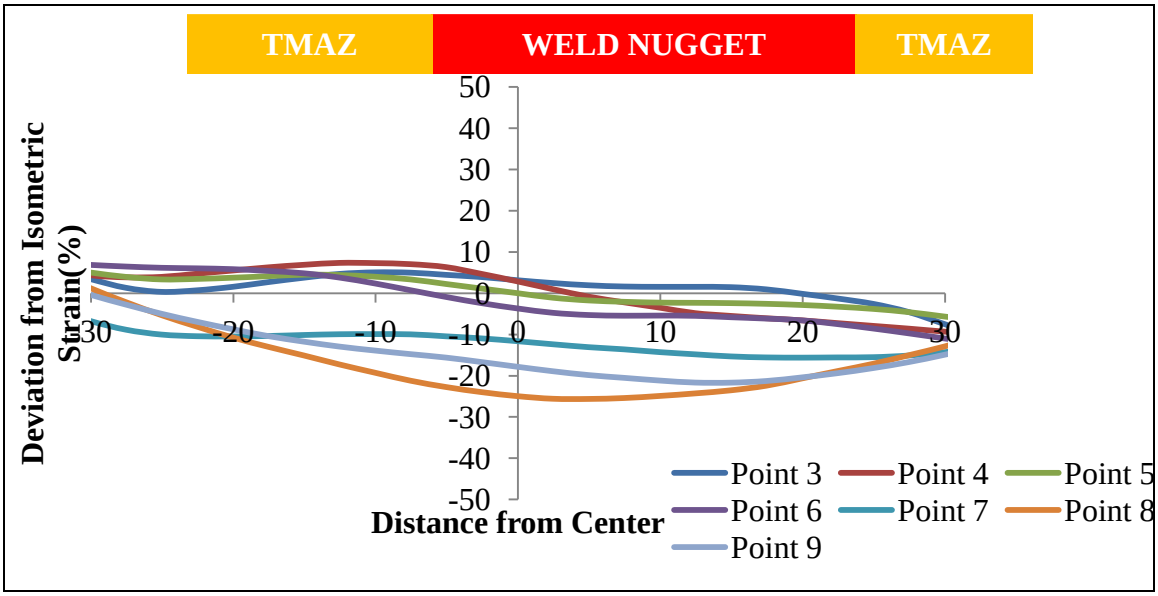

Figure 16. Strain distribution evolution in 2X from center weld.

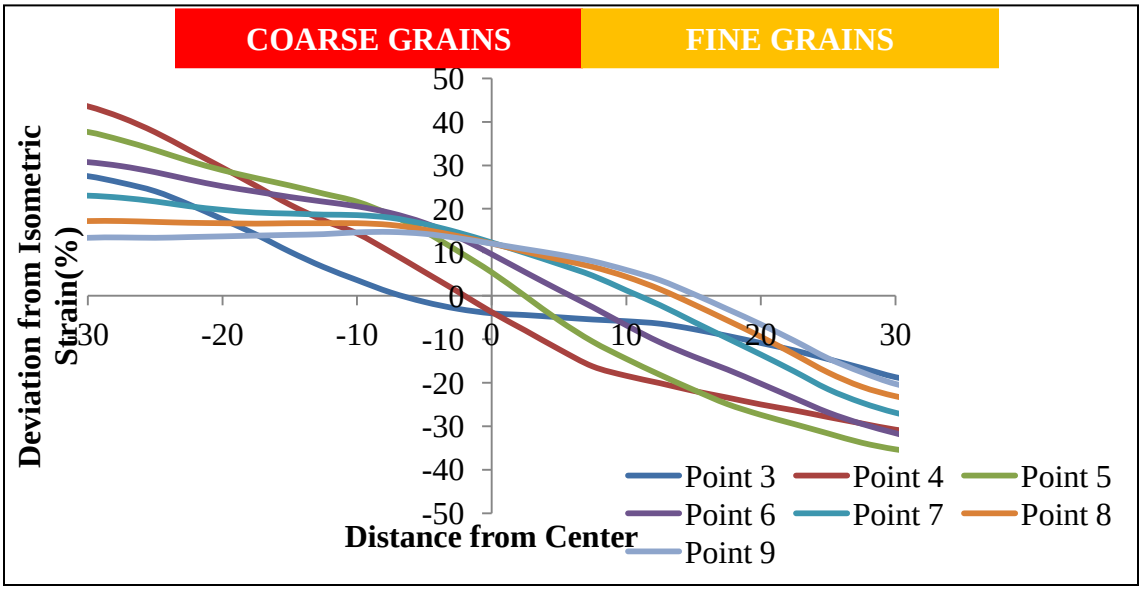

Figure 17. Strain distribution evolution in 2Z from center weld.

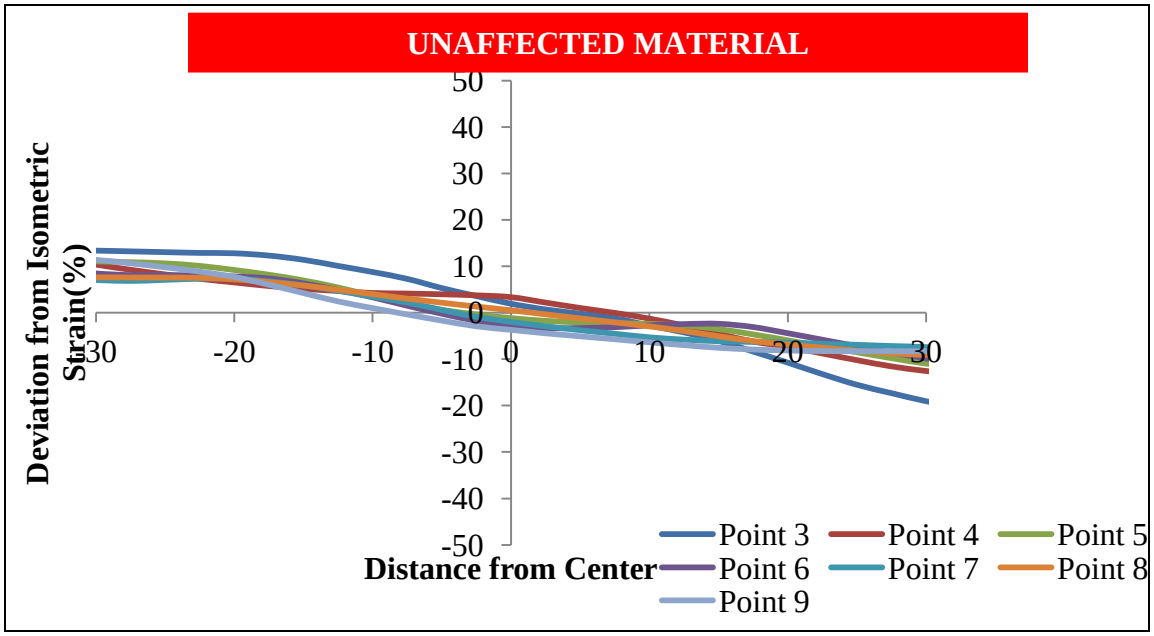

Figure 18. Strain distribution evolution in 11X.

The results from figure 16 indicate that as the material is strained, it does not strain isometrically, but rather is dependent on the microstructure in the given specimen. From figure 16, the center of the specimen strains at a rate 25% lower than the specimen average. This is related to the fine grains located in the center of the weld nugget.

The coarse grains of the TMAZ are straining at a higher gradient because of the grain size effect. These results demonstrate the development of local yielding in the more favorable zones. As indicated, the microstructurally related strain distribution evolution develops as the preferential grain structure begins to yield, in this case, the TMAZ. This result is in agreement with the Hall-Petch relationship (*14*).

Likewise, this same behavior is seen in the Z-direction of the weld and shows a distinct zone of both fine and coarse grains. This grain distribution effect on non-isometric strain is captured in figure 17, which shows the evolution of the strain distribution in the 2Z sample. As the strain distribution evolves, a preference toward straining is observed by the top nugget, which strains 12% higher than the average, whereas the bottom nugget strains as much as 15% below average. This again is an effect of the grain size. The finer grains are stronger and strain less than the coarse grains. This once again is in agreement with the Hall-Petch relationship.

Lastly, the strain distribution of the unaffected material far from the weld of section 11 is shown in figure 18. These results show that while there is slight preferential straining toward the top end of the sample, this pattern generally stays intact and there is no microstructural related strain evolution that develops during elastic loading. These results confirm the existence of a microstructural-related strain distribution evolution between zones in the weldment.

4. Summary and Conclusions

4.1 Summary

Tensile specimens were cut from an FSW butt joint of AA-2139 and tested. The mechanical properties show little spatial variation; however, in the weldment, there are significantly different mechanical properties. There is a higher failure strain and lower ultimate tensile strength in the weld zone. This may be attributed to the specific microstructures that form during the FSW process characterized by the TMAZ, HAZ, and weld nugget. Optical microscopy showed an abrupt change in the microstructure, which results in elevated local stresses due to the mismatch in mechanical properties. Additionally, a difference in local strain gradients was observed due to the microstructural differences, which allows for increased ductility. Analysis of the grain structure in these zones shows that the grains have a slight preferential alignment in the weld direction, which may account for the higher strains to failure. Anisotropic behavior can be explained by the preferential alignment of the grains towards the 101 and 111 directions in the X- and Y-directions while a preference toward the 101 is observed for the Z-direction.

4.2 Future Work

Since a model has been developed to accurately simulate the welding and relative microstructures, a novel technique for simulating the blast is needed. One way to do this would be to develop a homogenized response characterized by the spatial mechanical response of the weld. In order to do this, a better understanding of the mechanical properties derived in this experiment would be necessary. Further testing to understand the mechanical response in the weld zone will allow for successful implementation of the data into a homogenized center weld zone that will not require modeling of the TMAZ, HAZ, and weld nugget. When this is completed, a finite element model and simulation will be used to allow for a better design methodology to implement a FSW join on a vehicle.

5. References

1. Thomas, W. M.; Nicholas, E. D.; Needham, J. C.; Murch, M. G. Friction Stir Butt Welding. Patent Application Number 5,460,317 United States of America, 1995.

2. Mahoney, M. W.; Rhodes, C. G.; Flintoff, J. G.; Spurling, R. A.; Bingel, W. H. Properties of Friction-Stir-Welded 7075 T651 Aluminum. s.l.: *Metalurgical and Materials Transactions A* **July 1998**, *29A*.

3. Hirata, T.; Oguri, T.; Hagino, H.; Tanaka, T. Influence of Friction Stir Welding Parameters on Grain Size and Formability in 5083 Aluminum. 1-2, s.l.: *Material Science & Engineering A* **2007**, *456*, 334–349.

4. Buffaa, G.; Fratinia, L.; Shivpurib, R. CDRX Modelling in Friction Stir Welding of AA7075-T6 Aluminum Alloy: Analytical Approaches. 1-3, s.l.: *Journal of Materials Processing Technology*, *191*, 356–359.

5. Fratini, L.; Buffa, G. CDRX Modelling in Friction Stir welding of Aluminium Alloys. 10, s.l.: *International Journal of Machine Tools and Manufacture*, *45*, 1188–1194 .

6. Fratini, L.; Buffa, G.; Palmeria, D. *Computers & Structures.* 17–18, *87*, 1166–1174.

7. Grujicic, M.; Arakere, G.; Yalavarthy, H. V.; He, T.; Yen, C.-F.; Cheeseman, B. A. Modeling of AA5083 Material-Microstructure Evolution During Butt Friction-Stir Welding. s.l.: *Journal of Materials Engineering and Performance* **July 2010**, *19*, 672–684.

8. Grujicic, M.; Arakere, G.; Yen, C.-F.; Cheeseman, B. A. Computational Investigation of Hardness Evolution during Friction-Stir Welding of AA5083 and AA2139 Aluminum Alloys. s.l.: *Journal of Materials: Design and Applications* **January 2010**.

9. Grujicic, M.; Arakere, G.; Pandurangan, B.; Hariharan, A.; Yen, C-F.; Cheeseman, B. A.; Fountzoulas, C. Computational Analysis and Experimental Validation of the Ti-6Al-4V Friction Stir Welding Behavior. s.l.: *Journal of Engineering Manufacture January 2010*.

10. Yuana, H.; Wanga, Q. F.; Zhanga, J. W.; Liua, W. C.; Gaob, Y. K. Effect of Grain Shape on the Texture Evolution During Cold Rolling of Al–Mg Alloys. s.l.: *Journal of Alloys and Compounds* **2011**, *508*, 922–928.

11. Suhuddin, U.F.H.R.; Mironov, S.; Sato, Y. S.; Kokawa, H. Grain Structure and Texture Evolution During Friction Stir Welding of Thin 6016 Aluminum Alloy Sheets. s.l.: *Materials Science and Engineering A* **2010**, *527*, 1962–1969.

12. Schmidt, H.; Hattel, J.; Wert, J. An Analytical Model for the Heat Generation in Friction Stir Welding. s.l.: *Modelling and Simulation in Material Science and Engineering* **2004**, *12*, 143–157.

13. Medintz, J. Private Communication. US, ARL: s.n., July 2011.

14. Callister, William D.; Rethwisch, David G. *Materials Science and Engineering: An Introduction;* s.l.: John Wiley & Sons, 2009.

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Tunable, Highly Ordered TiO_2 Nanotube Arrays on Indium Tin Oxide Coated PET for Flexible Bio-sensitized Solar Cells

JOSHUA J. MARTIN, UNIVERSITY OF DELAWARE
MENTORS: DR. SHASHI KARNA AND DR. MARK GRIEP
U.S. ARMY RESEARCH LABORATORY, WMRD, RDRL-WMM-A
ABERDEEN PROVING GROUND, MARYLAND

Contents

List of Figures **191**

Abstract **192**

Acknowledgments **193**

Student Bio **194**

1. Background Information **195**

1.1 DSSC Overview 195

1.2 Critical Factors for Efficiency 197

1.2.1 Growth of TiO_2 Nanotube Arrays 198

1.2.2 Length, Diameter, and Pore Size 198

1.2.3 Crystalline Structure 199

2. Experiment/Calculations **200**

3. Results and Discussion **201**

4. Summary and Conclusions **203**

5. References **204**

List of Figures

Figure 1. Principle of a DSSC.....196

Figure 2. (a) Random electron walk in nanoparticles. (b) 1-D electron transport in nanotubes. 197

Figure 3. Proposed Concept design. 197

Figure 4. Anodization set up. 198

Figure 5. (a) Without methanol wetting, (b) with methanol wetting, (c) side SEM view of hexagonal, highly-oriented nanotube array, and (d) top view. 201

Figure 6. (a) Anodization with a high NH_4F Concentration and (b) anodization with a normal NH_4F concentration. 202

Figure 7. (a) Nanograss formations and (b) honey-comb structure. 203

Abstract

Highly ordered, free-standing titanium dioxide (TiO_2) nanotube (TNT) arrays have been of intense interest in the alternative energies field in recent years due to their barrier-free electron conduction pathway versus TiO_2 nanoparticles in dye sensitized solar cell (DSSC) designs. TNT arrays prepared by electrochemical anodization of titanium (Ti) foils and combined with a transparent, indium tin dioxide coated polyethylene terephthalate (PET) film are attractive candidates for efficient, flexible DSSCs. Flexible solar cells offer great benefits because of the potential for low-cost, roll-to-roll production and the increase in applications due to superior robustness. This approach uses a two-step anodization procedure coupled with implementation of a rapid inert gas dehydration and ultrasonic agitation detachment method. By controlling the reaction conditions during anodization (voltage, duration, and concentration), TNT arrays with specific morphology, lengths, and diameters can be tailored to satisfy a particular application such as incorporating specialized protein dyes, in particular, bacteriorhodopsin (bR). The free-standing arrays, comprised of hexagonally closed-packed and regularly ordered TNT membranes, have been synthesized and detached from the original Ti substrate. Once the TiO_2 sol-gel is created, the free-standing arrays will be attached to the flexible PET film for improved photovoltaic properties and overall performance.

Acknowledgments

I wish to acknowledge the mentoring support of Dr. Shashi P. Karna and Dr. Mark Griep, as well as all fellow Oak Ridge Institute for Science and Education (ORISE) and Science and Engineering Apprentice Program (SEAP) interns working with me this summer.

Student Bio

I am currently a rising senior at the University of Delaware studying mechanical engineering with a biomedical engineering minor. I am a member of the Society for the Advancement of Material and Process Engineering (SAMPE) and Engineers Without Borders Club on campus, as well as the President of the Rock Climbing Club. This is my first research period at the U.S. Army Research Lab. After graduating from UD, I plan on pursuing my interest in renewable energy. After a few years, I intend on applying everything I have learned to a term with the Peace Corps.

1. Background Information

Soldiers and vehicles are particularly constrained by a dependence on bulky batteries to meet their power needs. Future force Soldiers will be equipped with a wide array of technology, essentially all of which requiring electrical power for operation. Integrating solar harvesting capabilities with future equipment would relieve the burden of batteries and increase overall performance of Soldiers, as well as provide relief toward the need for renewable energy sources. While modern silicon (Si) cells are costly to produce and adversely affect mobility, dye-sensitized solar cells are providing very advantageous properties. These systems are prospectively capable of achieving higher efficiencies, more attractive costs, and lower environmental impact, and are more easily manufactured than current Si solutions. Dye-sensitized titanium oxide (TiO_2) platforms do not require expensive semi-conductor substrates, nor do they require highly complex processing steps, making roll-to-roll production possible (*1*). Most importantly, they can be geared to be highly robust and flexible. These attributes qualify dye-sensitized solar cells (DSSCs) as a chief candidate for future solar energy harvesting systems, and will facilitate a successful entry to the Si-dominated market.

The objective of the research presented in this paper is to develop and optimize TiO_2 nanotubes for a flexible, dye-sensitized solar cell. Particular focus will go into producing highly ordered nanotube arrays with a tunable pore diameter to compensate for a bacteriorhodopsin (bR) dye, thus increasing the organic nature of the cell. Removal of the nanotube arrays from the Ti substrate using methanol wetting and N_2 blowing will make it possible to make a front-side illuminated solar cell, while replacing the conducting substrate with indium tin dioxide-coated PET increases the flexibility and conductivity.

1.1 DSSC Overview

The classic DSSC is composed of a layer of nanocrystalline TiO_2 particles on a conducting substrate, a platinum counter electrode, an electrolyte, and an adsorbed ruthenium (Ru)-dye as a sensitizer (*2*). In contrast to conventional Si systems, where the semiconductor assumes both the light absorption and charge carrier transport, the two functions are separated here (*3*). The operating principle is demonstrated in figure 1. On the surface of the TiO_2 is an adsorbed dye, attached by specific functional groups, which serves as a light absorber and photon-to-electron converter. A key requirement for the dye is that the lowest unoccupied molecular orbital

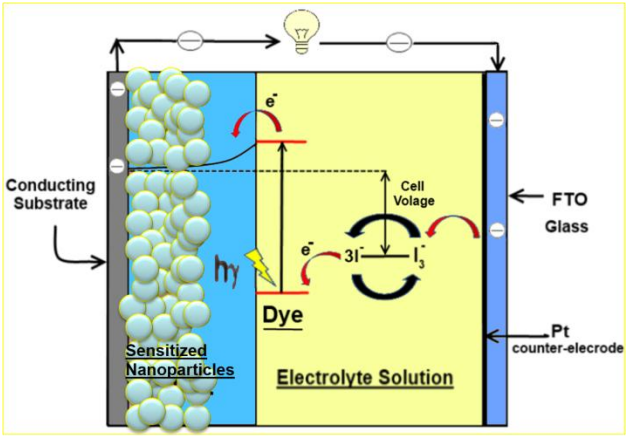

Figure 1. Principle of a DSSC.

(LUMO) of the dye molecule is energetically positioned slightly higher than the conduction band of TiO_2. Under the illumination of sunlight, highest occupied molecular orbital (HOMO)–LUMO transitions in the dye occur. Photoexcitation of the dye leads to rapid injection of electrons into the TiO_2. These electrons are then collected to power an external load. For typically used Ru dyes, the electron injection from the dye sensitizer to the conduction band of TiO_2 occurs through a metal-to-ligand charge transfer (MLCT) pathway (*4*). Electron donation from the redox couple in the electrolyte, usually an ionic liquid containing an I_3^-/I system, regenerates the oxidized dye and returns it to its original state. The tri-iodide ions formed in the dye regeneration process diffuse through the liquid phase to the cathode, where they are reduced back to iodide ions to complete the cycle (*5*). The efficiency of collection of the photoinjected electrons, which is a critical factor in device performance, is determined by competition between electron transport to the anode and electron transfer to the I_3^- ions electrolyte. Research using such constructions has achieved light-to-electricity conversion efficiencies of 11% (*6*).

While this demonstrates much prospect, the electron transport time in TiO_2 nanoparticles is relatively slow when compared to the rate of de-excitation of the dye and the regeneration time constant of the dye (*7*). Electron flow is slowed down due to defects, surface states, grain boundaries, and other sites of electron trapping within the nanoparticles. These factors enhance recombination and adversely affect electron transfer (*8*). In other words, the majority of excited electrons recombine with the dye before they can be collected to power a load. To minimize this effect, recent research has explored the use of more organized TiO_2 structures such as nanotubes.

TiO_2 nanotubes have a highly ordered structure with vertical pore geometry, which appears to be very suitable for the fabrication of solid-state junction cells (*9*). This is due mainly to the 1-D conductive path of nanotubes vs. the 3-D unsystematic walk network and grain boundary effects of randomly associated nanoparticles, displayed in figure 2. It has been reported that compared to conventional TiO_2 nanoparticle films of the same thickness, nanotube arrays give enhanced light scattering and improved collection efficiencies (*10*). Therefore, this study will be using the TiO_2 nanotube platform displayed in figure 3.

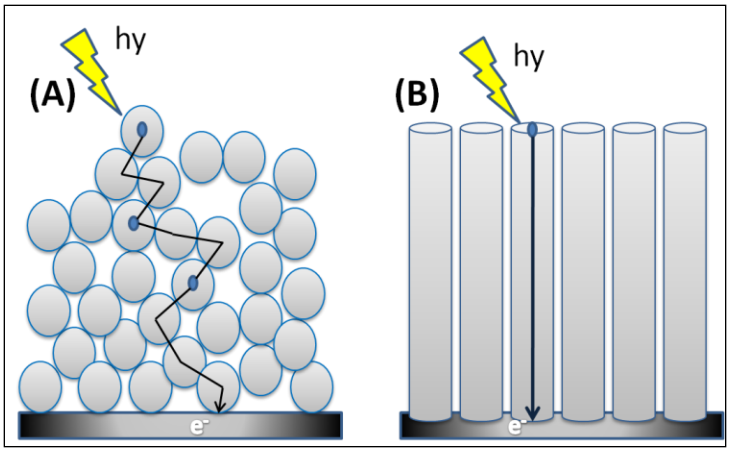

Figure 2. (a) Random electron walk in nanoparticles. (b) 1-D electron transport in nanotubes.

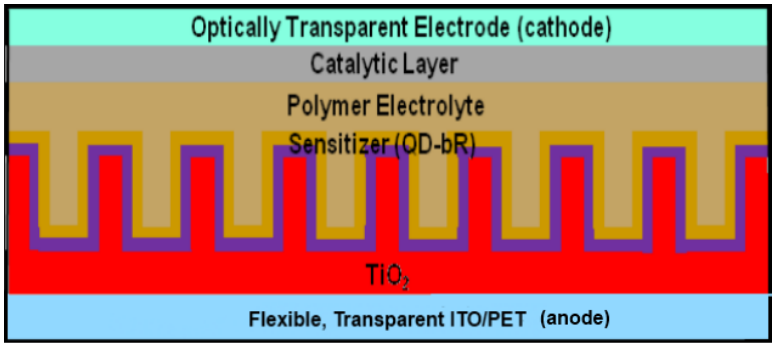

Figure 3. Proposed Concept design.

1.2 Critical Factors for Efficiency

It has been established that the conversion efficiency of DSSCs is mainly governed by molar absorption coefficient and HOMO-LUMO levels of the dye, the effective surface area of the electrodes available for dye anchoring, the transport kinetics of electrons to the substrate, the efficiency of dye regeneration via a redox couple, and the losses arising from recombination and back reactions of injected charge carriers (*11*). Many papers exist with the goal of optimizing key components of DSSCs, such as the TiO_2 structures, the dye, and the electrolyte used. However, it is beyond the scope of this paper to go into a detailed review of each topic. Critical optimization factors pertaining to the TiO_2 arrays will be outlined, with a particular focus in maximizing pore diameter to accommodate for the bR dye.

The performance of TiO_2 nanotube arrays in a DSSC as an efficient semi-conductor and electron recipient is directly related to several factors: morphology, crystallinity, and geometry (*2*). Many studies have been conducted in order to control the dimensions of TNT arrays via anodic oxidation (anodization). It has been shown that uniform TiO_2 nanotube arrays with various lengths, diameters, and wall thicknesses can be fabricated in fluoride containing solutions (HF

and NH_4F) by tailoring the electrochemical conditions (anodization voltage, fluoride concentration, anodization time, and temperature) (*12–14*).

1.2.1 Growth of TiO_2 Nanotube Arrays

The simple electrochemical process of anodizing Ti foils, displayed in figure 4, to produce TNTs can be accomplished using a two- (or three-) electrode system, with Ti-foil as the working electrode (+), a counter electrode such as platinum (Pt) (–), and a voltage source. Nanotube formation in fluoride-ion-bearing electrolytes occurs due to three simultaneously occurring processes: the field assisted oxidation of Ti metal to form TiO_2, the field-assisted dissolution of Ti metal ions in the electrolyte, and the chemical dissolution of Ti and TiO_2 due to etching by fluoride ions (which is substantially enhanced by fluoride ions) (*16*). A thin dioxide film develops at the Ti/electrolyte interface due to the following chemical reaction (*10*):

$$Ti + 2H_2O \rightarrow TiO_2 + 4H^+ - 4e^-. \quad (1)$$

As this oxidation occurs, one can note that the current measured will reduce as time goes on. This is due to the increase in resistance provided by the oxide. As the oxygen ions (O^{2-}) are transported from the solution to the Ti, titanium ions (Ti^{4+}) transported from the titanium to the electrolyte interface are dissolved into the solution. A simultaneous reaction occurs at the TiO_2/electrolyte interface, forming soluble fluoride complexes. These create pits at the surface according to the following reaction (*10*):

$$TiO_2 + 6F^- \rightarrow TiF_6^{2-}. \quad (2)$$

These complexes dissolve and form pits on the surface of the newly formed TiO_2. As anodization time increases, all three processes continuously increase the depth of the pores (pits) and, therefore, produce highly ordered nanotubes (*19, 20*).

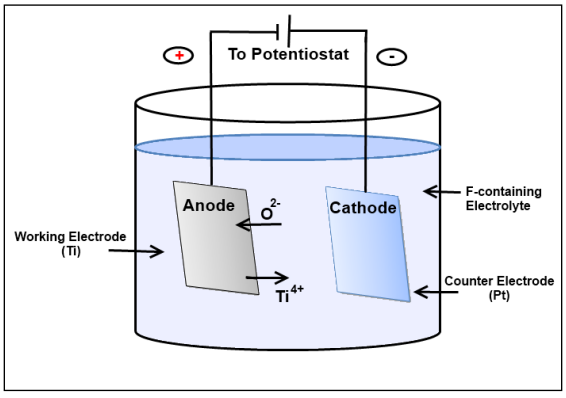

Figure 4. Anodization set up.

1.2.2 Length, Diameter, and Pore Size

Certain geometrical parameters of the nanotube arrays have a large effect on performance. It has also been found that by controlling the anodization setting, these parameters can be fairly

tunable. Length, diameter, and pore size (inner diameter) are largely influenced by anodization time, voltage, and NH_4F concentration (*15*). Increasing the overall length of the array allows for more surface area, resulting in more attached dye. The easiest way to modify the length of the array is to increase the anodization time. Due to the higher number of dye molecules that can attach to the increased length of the tubes, increasing thickness leads to an increase in the short circuit photocurrent. It is important to note that although the short circuit current is increased with tube length, so is the chance of recombination and electron trapping. It is reported that when the TNT length is significantly more than 20 μm, the diffusion length of the electron becomes shorter than the length of the TNT (*17*). Hence, the photovoltaic performance shows an initial increase and following decrease with the increasing TiO_2 nanotube length. The same study was able to achieve an outstanding conversion efficiency (Π) of 8.07% using a 20.8 μm nanotube array with an open-circuit voltage (V_{OC}) of 0.814 V, a short-circuit current (J_{SC}) of 15.46 mA cm^{-2}, and a fill factor (FF) of 64.1%. This agrees with other studies that report the maximum internal conversion efficiency is achieved for tubes of a length of approximately 20 μm (*2*).

In order to further modify the dye "absorbing" capabilities of the nanotubes, diameter and pore size are very important. These factors are critical depending on certain situations. In order to take full advantage of nanotubes having both inner and outer walls, the dye being used must be able to fit inside of the inner diameter. According to Schmuki's studies (*9*), the diameter of the tubes can be adjusted by the anodization voltage. In addition, modifying the concentration of the electrolyte used during anodization further modifies the pore diameter and wall thickness. This occurs due to the increase in the dissolution rate, which increases due to a higher amount of F^- ions within the electrolyte solution, leading to an increase of the inner diameter—thus reducing wall thickness and increasing the pore size (*10*).

1.2.3 Crystalline Structure

The crystal structure of TiO_2 tube walls is a significant factor that affects the electronic properties of the array. When anodically formed, the TiO_2 arrays are initially in an amorphous state. Due to a large amount of defects, impurities, and other recombination sites, the amorphous TiO_2 has hardly any conversion efficiency (*11*). In order to harness the maximum potential of the TNT array, the nanotubes must be annealed to a crystallite form.

When annealed at temperatures between 300 and 500 °C for about 3 h, the anatase crystalline form of TiO_2 can be obtained. At higher temperatures—550 °C and above—rutile-based TiO_2 will begin to develop. In addition to annealing temperature, the ramping speed has an effect on solar cell performance (*2*). High ramping speeds (>30 °C/min) can adversely affect performance. Furthermore, annealing for longer periods of time increases crystallinity. Little attention has been paid to the rutile form of TiO_2, although it is used as a base in most paints due to its light-scattering ability. The anatase form of TiO_2 is perceived as more active than rutile because of its surface chemistry and potentially higher conduction-band energy (*18*). Rutile films consist of homogenously distributed rod-shaped particles. These particles show no preferred orientation

and, therefore, have a much smaller packing density compared to anatase-based particles. The smaller packing density results in a smaller particle connectivity. In other words, electrons move slower through rutile-based TiO_2 than in anatase-based. In addition, anatase arrays have a larger surface area per unit volume (about 25%); this allows for more dye to be absorbed. As a result, anatase structures can achieve a 30% higher J_{SC} than rutile. Maintaining a pure anatase TiO_2 nanostructure is essential for achieving a high efficiency DSSC using nanotubes.

2. Experiment/Calculations

Highly ordered free-standing TiO_2 nanotube arrays were prepared using a two-step anodization process. Ti foils (99.7%, 0.25 µm) were purchased from Sigma-Aldrich and electrochemically anodized in a standard two-electrode cell using a Platinum counter-electrode and an Agilent E3649A DC power supply (figure A-1). The electrodes were kept at a fixed distance of 1.5 cm. The foils were first mechanically polished and cut into 1cm × 2 cm samples. Then, the substrates were separately sonicated in acetone, isopropanol, and ethanol, each for 5 min, before being rinsed with deionized (DI) water. The anodization process was performed at 60 V for 2–6 hr using an electrolyte consisting of 0.25 wt % NH_4F and 0.75% H_2O in ethylene glycol. The anodized samples were rinsed with DI water and then soaked in methanol for 30 s to initiate the detachment procedure. Free-standing TNT membranes were separated from the Ti substrate by drying the methanol-wetted samples with a stream of N_2 gas. This caused the freshly formed TNT array to delaminate from the substrate. To effectively remove the TNT membranes, methanol wetting and N_2 blowing may need to be repeated several times. The Ti substrates from the first step were then anodized a second time using a fresh electrolyte solution of the same composition. The separation procedure was repeated in order to achieve highly ordered free-standing TiO_2 nanotube arrays with clear, open-top ends.

To crystallize the amorphous TNT arrays into the anatase phase, a two-step annealing process was used. First the TNT arrays were subjected to 200 °C for 1 hr using a ramp rate of 1 °C s^{-1}. The arrays were then further heated at 475 °C for 3 h using a heating rate of 30 °C s^{-1}. After annealment, the arrays are to be bonded to a PET film using a TiO_2 sol containing titanium butoxide and polyethylene glycol. The TNT/PET was then annealed at a low temperature of 200 °C.

Following the annealing process, the TNT/PET films are to be sensitized using a bacteriorhodopsin dye (bR) for 16 hr. To evaluate their performance, the dye-sensitized TNT/PET films will be closed together using an electrolyte and indium tin dioxide-coated PET.

3. Results and Discussion

Figure 5 shows the effects of methanol wetting and N_2 blowing on freshly anodized TNT arrays. Due to the low surface-tension of methanol, soaking the TNT arrays allows the methanol to penetrate the oxidized layer and aid in the delaminating of the nanotube layer. This resulted in a nanotube array nearly twice the size of one collected without methanol wetting.

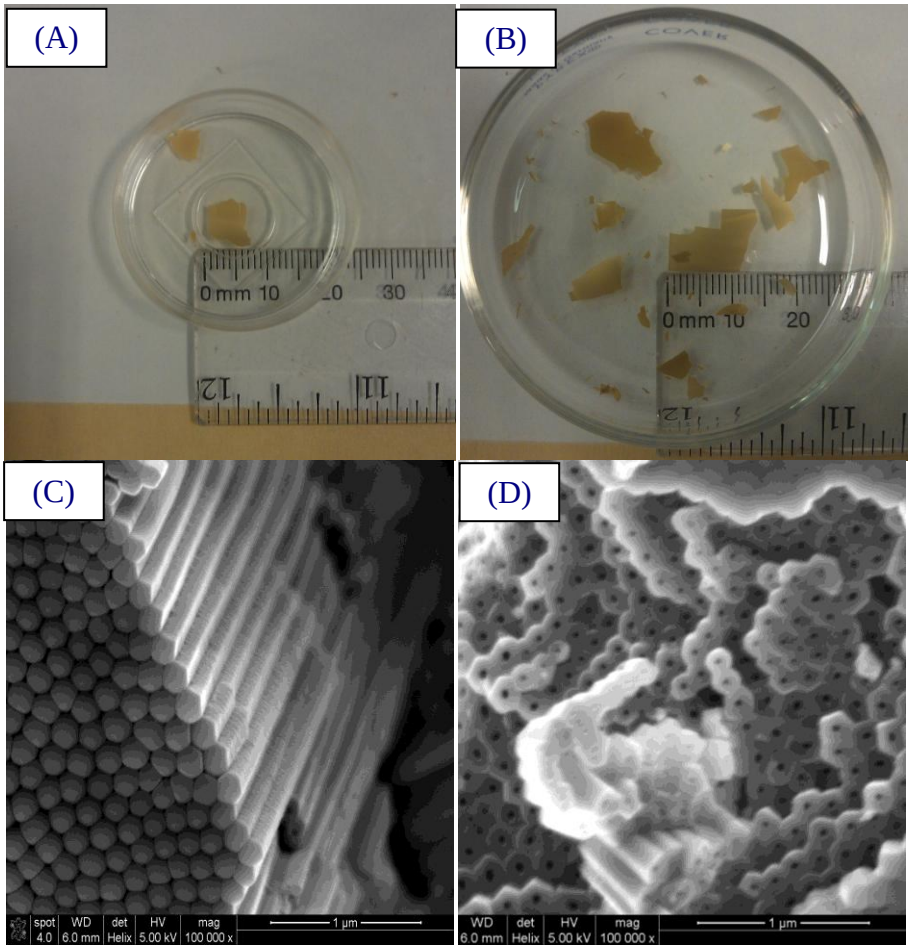

Figure 5. (a) Without methanol wetting, (b) with methanol wetting, (c) side SEM view of hexagonal, highly-oriented nanotube array, and (d) top view.

SEM images in figure 5 (c and d) show highly packed, hexagonally oriented nanotube arrays. This result was highly unexpected, because the arrays were formed using a single-step anodization procedure. These highly organized arrays were anodized at 60 V for 3 h, and had a length of 19 µm, tube thickness of 64 nm, and a pore diameter of 32 nm. The high wall thickness compared to pore size is most likely due to a low concentration of F^- ions in the electrolyte solution, which slowed down the dissolution rate, thus increasing the inner diameter. Since this was made immediately after mixing a batch of electrolyte solution (0.25 wt % NH_4F and 0.75%

H_2O in ethylene glycol), it is possible that the NH_4F was not equally distributed. The opposite affect can be seen in figure 6, in which the uneven distribution of NH_4F right after mixing a batch of electrolyte solution resulted in extraordinarily high concentration of F^- and the complete dissolution of the titanium electrode.

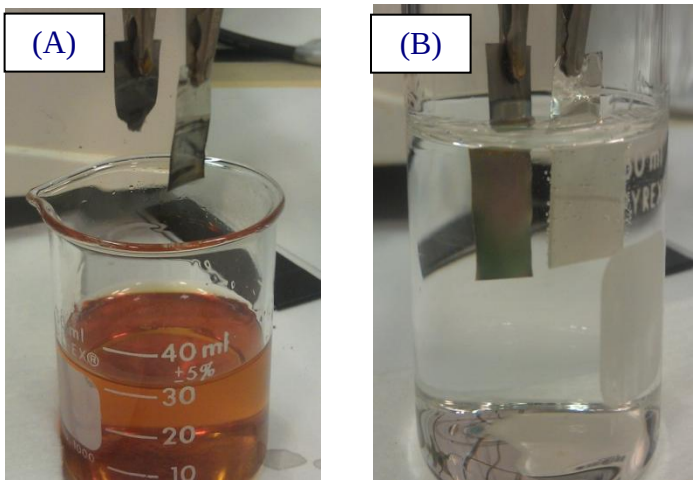

Figure 6. (a) Anodization with a high NH_4F Concentration and (b) anodization with a normal NH_4F concentration.

Although SEM images of samples anodized using different solution concentrations are not shown here, the effects of a longer anodization time at the same voltage and concentration are demonstrated in figures 5 and 7. The SEM images report samples that were anodized for 6 h at 60 V, using the same solution as the samples shown in figure 5. However, it is clear that the results are drastically different. This shows the effects of electrode preparation, as well as the sensitivity of the nanotubes. The SEM images report grass-like structures, also known as nano-wires, of approximately 24 nm thickness sitting on top of a highly porous honeycomb-like nanostructure with wall thickness of only 36 nm. The “nano-grass” is most likely the result of not mechanically polishing and sonicating the electrode prior to anodization. This allowed the chemical attack on the tube ends and the formation of nanograss. The top surface was left non-uniform compared to that of figure 5, resulting in an uneven dissolution of the Ti. The longer anodization time presumably contributed the rapid breakdown of the nanotubes.

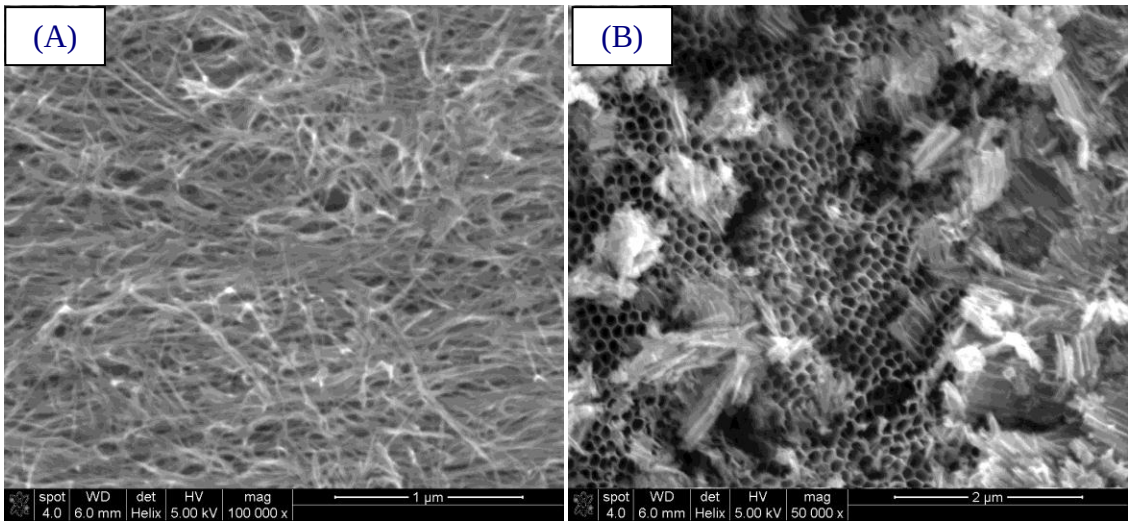

Figure 7. (a) Nanograss formations and (b) honey-comb structure.

4. Summary and Conclusions

This study presents a method of obtaining free-standing, highly organized TiO_2 arrays for use in flexible dye-sensitized solar cells. Free-standing, highly oriented, hexagonally packed nanotubes were made. Future efforts will be directed towards combining the nanotubes with an indium tin dioxide PET using a sol-gel consisting of Ti nanoparticles. In order to improve the reported results, a higher concentration of NH_4F should be used in the electrolyte solution. This will increase the inner diameter reported in figure 5, allowing for the use of a bR dye. An anodization time of less than 6 h should be used in order to avoid the dissolution at the nanotube tops into nanograss. Although the nanograss and honeycomb structures found in this study are not idea for use in a DSSC, they may have other valuable uses besides photovoltaics, such as bacteria filters and other biomedical applications.

5. References

1. Kang, M. G.; Park, N.-G.; Ryu, K. S.; Chang, S. H. 2001, p. 7.
2. Grätzel, M. *Inorg. Chem.* **2005**, *44*, 6841.
3. Jennings, J. R.; Ghicov, A. *J. Am. Chem. Soc.* **2008**, *130*, 13364–13372.
4. Chiba, Y.; Islam, A.; Watanabe, Y.; Komiya, R.; Koide, N.; Han, L. *Jpn. J. Appl. Phys.* **2006**, *45*, L638.
5. Koops, S. E.; O'Regan, B. C.; Barnes, P.R.F.; Durrant, J. R. *J. Am. Chem. Soc*. **2009**, *131*, 4808.
6. Nelson, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 15374, LP.
7. Shankar, K.; Mor, G. K.; Prakasam, H. E.; Varghese, O.K.; Grimes, C. A. *Lamg,ior* **2007**, *23*, 12444–12449.
8. Zhu, K.; Neale, N. R.; Miedaner, A.; Frank, A. J. *Nano Lett.* **2007**, *7*, 69–74
9. Ghicov, Andrei; Albu, Sergiu. $Ti0_2$ Nanotubes in Dye-Sensitized Solar Cells: Critical Factors for the Conversion Efficiency. *Chem. Asian J.* **2009**, *4*, 520–525.
10. Elsanousi, A.; Zhang, J. Self-Organized $Ti0_2$ Nanotubes with controlled dimensions by anodic Oxidation. *J Mater Sci* **2008**, *43*, 7219–7729.
11. Cai, Q.; Paulose, M.; Varghese, O. K. *J Mater Res* **2005**, *20*, 230, doi: 10.1557/JMR.2005.0020.
12. Macak, J. M.; Taveroa, L. V.; Tsuchiya, H.; Sirotna, K.; Macak, J.; Schumuki, P. *J Electroceram, 16*, 29.
13. Mor, G. K.; Carvalho, M. A.; Varghese, O. K.; Pisho, M. V.; Grimes, C.A. *J Mater Sci* **2004**, *19*, 628.
14. Prakasam, H. E.; Shankar, K.; Paulose, M.; Varghese, O. K.; Grimes, C. A. A New Benchmark for $Ti0_2$ Nanotube Array Growth by Anodization. *J. Phys. Chem* **2007**, *111*, 7235–7241.
15. Lei, B.; Liao, J.; Zhang, R.; Wang, J.; Su, C.; Kuang, D. Ordered Crystalline $Ti0_2$ Nanotube Arrays on Transparent FTO Glass for Efficient Dye-Sensitized Solar Cells. *J. Phys. Chem.* C **2010**, *114*, 15528–33.
16. Park, N. G.; Frank, A. J. Comparison of Dye-Sensitized Rutile and Anatase-Based $Ti0_2$ Solar Cells. *J. Phys. Chem. B* **2000**, *104*, 8989–8994.

17. Yang, D. J.; Kim, H. G.; Cho, S. J.; Choi, W. Y. *Mater Lett* **2008**, *62*, 775, doi:10.1016/j.matlet.2007.06.058.

18. Macak, J. M.; Tsuchiya, H.; Ghicov, A.; Yashuda, K.; Schumuki, P. *Curr Opin Solid State Mater Sci* **2007**, *11*, 3, doi:10.1016/j.cossms.2007.08.004.

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Tactical Means to Stow Super-Caliber Tailfins of a Developmental Flight-Controlled Mortar

CHRISTIAN A VALLEDOR
MENTOR: JOHN CONDON
GUIDANCE TECHNOLOGIES BRANCH, LETHALITY DIVISION
WEAPONS AND MATERIALS RESEARCH DIRECTORATE
ABERDEEN PROVING GROUNDS, MARYLAND

Contents

List of Figures **209**

Abstract **210**

Acknowledgments **211**

Student Bio **212**

1. Introduction/Background **213**

2. Design Process **213**

2.1 Literature Review 214

2.2 Modeling 215

2.3 Prototyping 217

3. Evaluation and Discussion **218**

4. Summary and Conclusions **219**

5. References **220**

List of Figures

Figure 1. Patent No. 7851734 (left) and Patent No. 6314886 (right). 214
Figure 2. Conceptual drawings, active (left) and passive (right). 215
Figure 3. FRD revisions. 216
Figure 4. Stress plot of device under a 10-lbF load, applied at snap position. 216
Figure 5. Final revisions. 217
Figure 6. SolidWorks rendering and prototypes assembled FRD on existing fin set. 217

Abstract

The active control systems being used in new “smart” munitions often require greater aerodynamic stability to properly operate. In the case of the Flight Controlled Mortar (FCM), this stability is provided by a set of super-caliber, spring-loaded, deployable fins that must be stowed prior to launch. For the testing and evaluation process, the FCM tail fins were secured using a single strand of thin Kevlar string. In an attempt to produce a more rugged tactical solution, research and development into alternative Fin Retention Devices (FRDs) was initiated at the start of the summer. After a thorough review of existing FRDs, I determined that a custom solution was needed to fulfill the needs of FCM. Using an iterative design process, concepts were modeled using SolidWorks, and physical prototypes were produced. Based on the analysis of these prototypes, further changes have been identified for ongoing revision. I believe that at the end of this project, a tactical FRD will be produced for testing and evaluation, and may one day be used for the FCM or similar systems employing folding fins.

Acknowledgments

First, I wish to acknowledge my mentor Mr. John Condon for his constant support and guidance. Furthermore, the members of the Guidance Technologies Branch at large are thanked for accepting me as a valued addition to the team and supporting me in all ways possible.

Student Bio

I am a member of the Massachusetts Institute of Technology's (MIT) class of 2012, majoring in Aeronautics and Astronautics. In my time at MIT, I have participated in many hands-on research and engineering projects through the MIT Rocket Team as well as through coursework. In the 2010–2011 school year, I lead the Rocket Team in the National Aeronautics and Space Administration's (NASA) University Student Launch Initiative, placing second overall with the team's novel rocket-launched unmanned aircraft system. In the spring of 2011, as part of a senior capstone course, I and a partner developed and planned an experiment to test a face shield addition to the Advanced Combat Helmet in an effort to mitigate blast-induced mild traumatic brain injuries for Soldiers. The experiment is slated to be conducted in fall 2011. After completing my undergraduate degree, I plan to explore opportunities to gain secondary degrees while gaining hands-on research experience.

1. Introduction/Background

In an effort to provide the warfighter with precision indirect fire capabilities the U.S. Army Research Laboratory (ARL) is working in conjunction with the Naval Surface Warfare Center (NSWC), Dahlgren, VA, on an 81-mm Flight Controlled Mortar (FCMortar or FCM) program. The FCM project focuses on the development of a guidance and control system that is compatible with the standard 81-mm mortar system already in use. However, due to the active control surfaces used in this system, the standard M24 fin set cannot adequately provide the necessary aerodynamic stability. Therefore, a fin set employing spring-loaded super-caliber folding fins has already been developed for use with the FCM.

In the folded state, the new fin set allows the FCM to function with the standard M252 mortar tube. However, because the springs on the fin hinges are pre-tensioned for rapid deployment upon firing and muzzle exit, a fin retention mechanism must be employed to allow for proper travel down the length of the mortar tube upon "hand-drop." Furthermore, due to stability requirements, the fins must be deployed immediately upon mortar tube exit, within about 30 ms. For the purposes of testing and evaluation, this task was accomplished with the use of a single length of 0.014-in-diameter Kevlar (para-aramid) string located in notches on the trailing edge of the fins and secured with a single hand-tied square knot. Upon ignition of the mortar's propelling charges, the para-aramid string burns through and the FCM moves up and through the mortar tube, allowing for fin deployment upon muzzle exit.

Although this retention mechanism has worked for the test flight events in the testing and evaluation phase, there are two main concerns about its tactical use. First, the use of a single string lends itself to getting caught on items during storage and handling by the user. This increases the likelihood of the string being cut or broken, leaving the mortar in an unusable state until replaced. Secondly, properly installing the sting and tying the knot requires a great amount of time and effort. The increased difficulty in installing the string has tactical disadvantages as it would be difficult to accomplish in the field should the string need to be replaced. As such, it was tasked over the summer to research and develop possible tactical solutions to this fin retention issue. A tactical solution would be easier to install and be less prone to snagging or cutting damage, while still allowing for proper FCM operation.

2. Design Process

To complete the design of a fin retention solution in a systematic method, the design process was broken into three distinct phases: open literature review, modeling, and prototyping. Using this method it was possible to quickly move through the design process while minimizing time spent

on fruitless paths. Furthermore, by using iterative design in the modeling and prototyping phases allows for the fin retention device to be further refined until all design requirements are met or exceeded.

2.1 Literature Review

By researching published documentation on deployable fin retention systems, as well as comparing existing products on the market, it was possible to learn how other systems tackle the challenge of fin retention. To begin this process, a market survey was completed, examining similar projectiles using deployable fins. It quickly became apparent, however, that the challenges of using spring-loaded fins on a mortar-type projectile are fairly unique in the field. This is primarily due to the fact that mortars are muzzle loaded and must travel through the launch tube under the influence of gravity alone before being propelled out of the barrel. Most systems employing super-caliber fins are breach loaded, and therefore, only travel through the barrel when being propelled by their primary charges. In many cases, these products also do not require fin retention systems because the launch tube can constrain the fins throughout the launch process. Of all the systems investigated, the only other mortar system identified was the 120-mm XM395 Precision Guided Mortar Munition. Some versions of this mortar system used a folding fin design very similar to that used in the FCMortar; however, no information was openly available on the fin retention system (*1*).

Finding no suitable solution in the market survey, a patent search was conducted in an attempt to identify other solutions applicable to the FCMortar project. Using the U.S. Patent and Trademark Office's searchable patent database and Google Patent Search, approximately 2,000 patents describing deployable fin retention methods were examined. Again due to the unique nature of the FCM's tail boom, no patents were identified that would be directly applicable. However two promising patents were identified providing the groundwork for future solutions. The drawings from these two patents can be seen in figure 1.

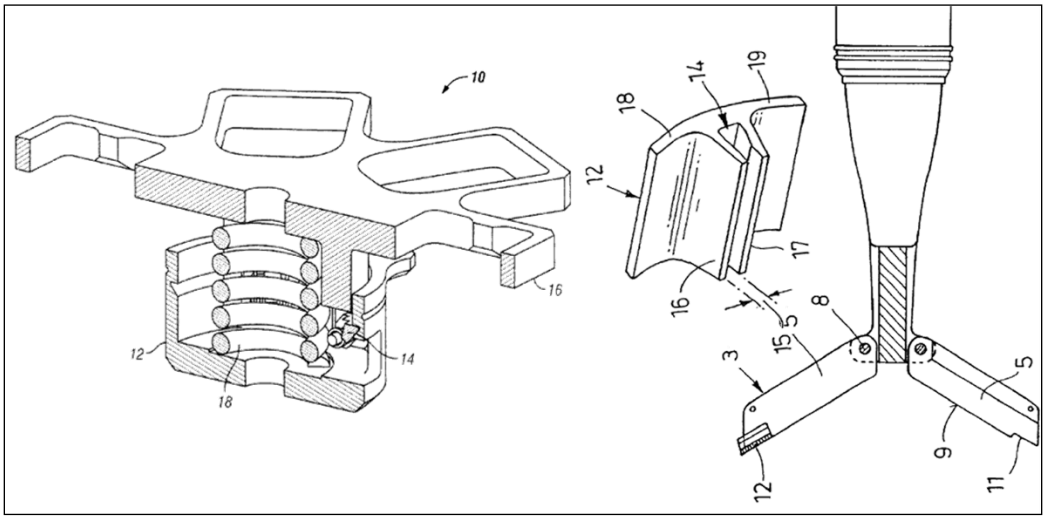

Figure 1. Patent No. 7851734 (left) and Patent No. 6314886 (right).

U.S. patent No. 7,851,734, seen on the left, describes a latch-based system that is actuated by the setback acceleration seen at launch. The mechanism shown in this drawing is located within the central core of the intended round, with the upper ring structure securing the fins until the initial setback acceleration of launch dissipates (*2*). Because of its central location, this patent would not work on the FCM. However, based on this patent, a spring-loaded ratcheting ring concept was developed, which would allow for external mounting around the tail boom. A conceptual diagram of this idea can be seen on the left of figure 2. U.S. patent No. 6,314,886 on the other hand described a completely passive device, labeled as number 12 in the drawing, which attaches to notches in the fins. The fins in this patent, however, are not spring loaded. Instead the patent describes a method of allowing the fins to ride along the gun barrel, and then deploy upon exiting the muzzle due to aerodynamic forces (*3*). Drawing on the topic of this patent, a ring structure was conceptualized, which would attach to the fins like in this patent, securing them while in barrel. A graphic of this ring structure can be seen on the right side of figure 2.

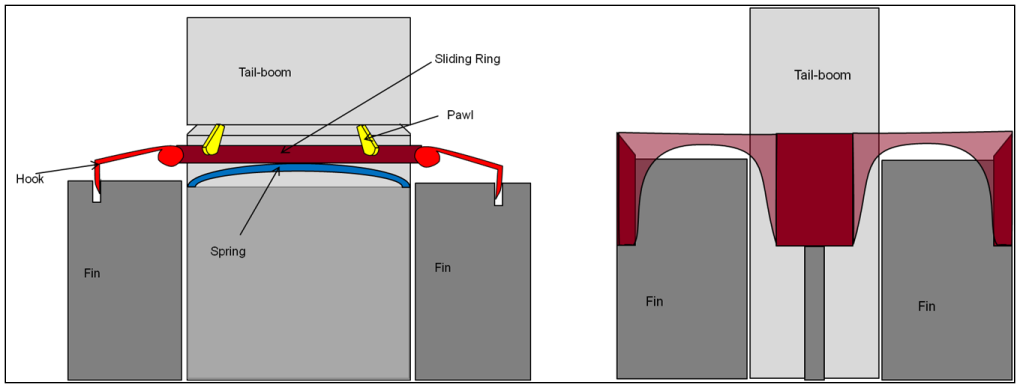

Figure 2. Conceptual drawings, active (left) and passive (right).

2.2 Modeling

After completing the literature review, concepts for possible Fin Retention Devices (FRDs) were explored and roughly grouped into two categories: active or passive. These categories were selected based upon the way the FRD would release the fins. After a brainstorming period, the best examples of passive and active FRDs were translated into conceptual drawings and briefed to some of the engineers and technicians working on the FCM project for additional input. These conceptual drawings can be seen in figure 2. From this meeting, it was decided that it was best to pursue a passive FRD as it would reduce the complexity of the tail section and simplify integration with existing aspects of the FCM.

With the determination to develop a passive FRD, the original concept, shown on the right of figure 2, was developed further and modeled in SolidWorks 2010 Professional. Using SolidWorks allowed for the rapid transition from concept to a virtual three-dimensional (3-D) component. SolidWorks also offered a chance to test interactions between the FRD and the fin set in a virtual state. This fact allowed for multiple revisions of the FRD in a very limited time span. Through this revision process, the original concept of a single section ring retainer

transformed into a set of four simple snap-fit devices that attach to each fin and link together in the span between fins. A selection of revisions created in this modeling process can be seen in figure 3, with the original concept on the far left and the most recent revision on the far right.

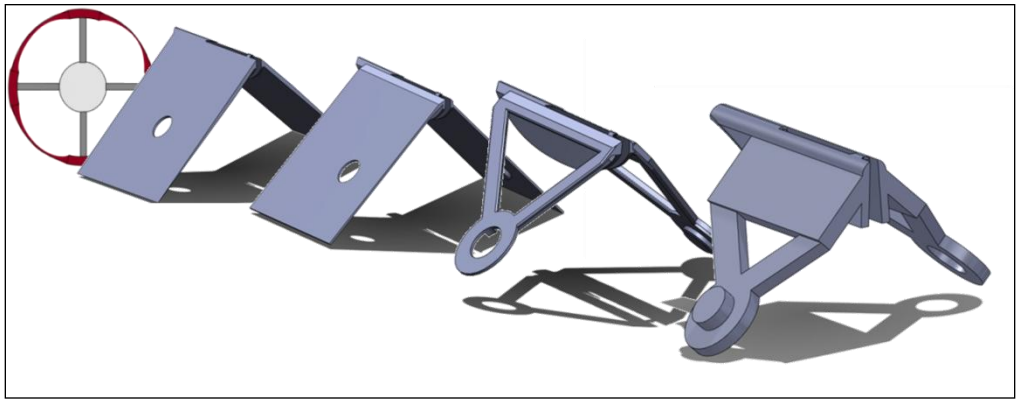

Figure 3. FRD revisions.

An additional benefit of using SolidWorks is that it allows for basic structural analysis using its built-in simulation tool. By employing these tools, a preliminary determination on the strength of each model was determined, allowing for each successive revision to become stronger. One way this was done was by analyzing the deflection and resultant stress in the arms that span the space between the fins. A graphical representation of this analysis can be seen in figure 4. On this plot brighter colors represent higher stresses, and the sections on the inner corners show plastic deformation.

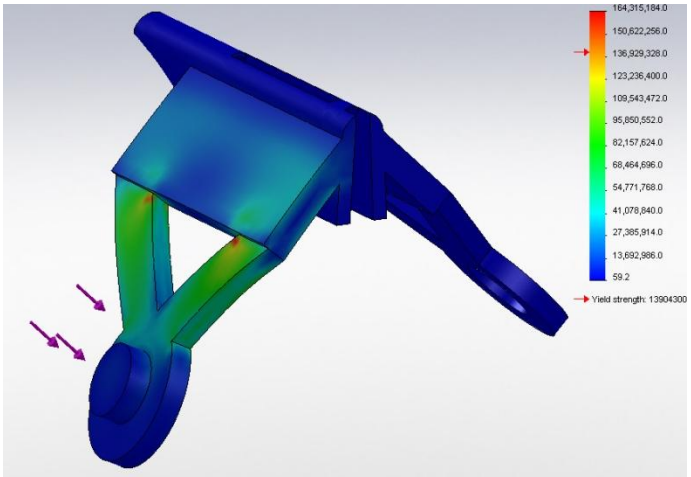

Figure 4. Stress plot of device under a 10-lbF load, applied at snap position.

The plastic deformation observed in the simulation was a sign that the arm section could fail in some cases. Therefore, one final revision was to add a leg extension to one of the arms to help support the structure. The addition of the support leg would allow for a reduction of the stress in

the overall structure. This allowed for only elastic deformation to result in the new model when further simulations were conducted. These two versions can be seen in figure 5.

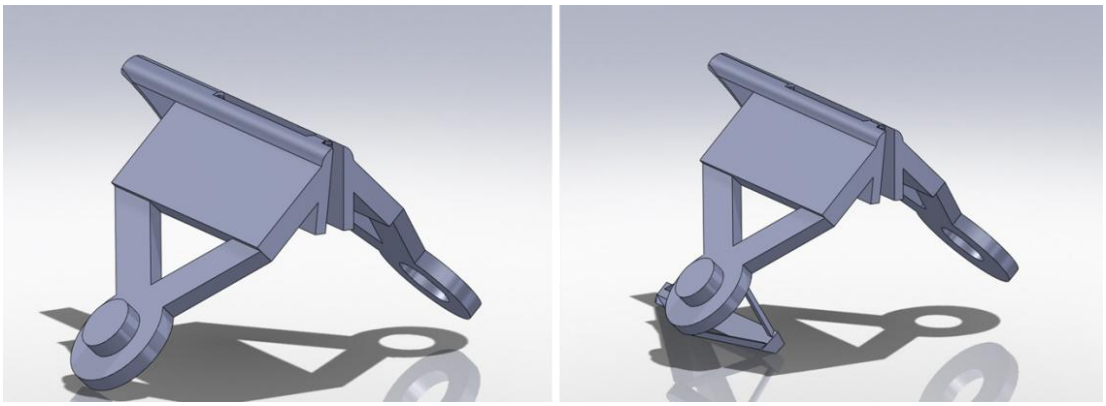

Figure 5. Final revisions.

Note: The leg is seen on the right figure, opposite the circular tab feature.

2.3 Prototyping

Although SolidWorks allows for the virtual testing of integration and loading of the parts, it is not enough to work based on computer models alone. Therefore, it was decided to send the final two revisions of the FRD to the rapid prototype shop so that physical models could be used for evaluation purposes. To keep the prototypes as accurate to the computer model as possible, they were constructed out of a photosensitive polymer using a stereolithography 3-D printing process. This process resulted in highly accurate prototypes with physical properties close to what would be used in a production setting. With a full set of these prototypes it was then possible to physically inspect the strength and performance of the FRD design. A SolidWorks rendering and the assembled physical prototype can be seen in figure 6.

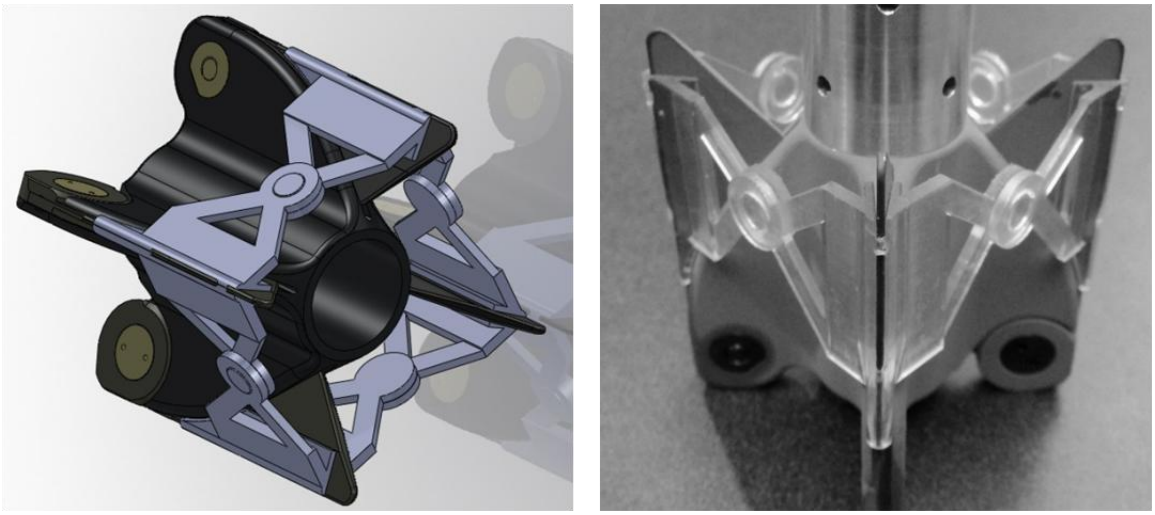

Figure 6. SolidWorks rendering and prototypes assembled FRD on existing fin set.

3. Evaluation and Discussion

With the arrival of the physical prototypes, it was possible to test the interaction between the existing fin set and the FRDs. From this simple test, it became immediately apparent that there were many small issues with the design. Most importantly, the notch in the fins that was previously used for securing the string during testing and evaluation was not large enough for the FRD design. Because of this fact, there was an obvious clearance issue between the FRD and the mortar tube. This clearance issue is one of the most critical issues, but could be solved in a variety of ways. One such method would be to increase the size of the notch; another would be to reduce the amount of material on this portion of the FRD. One other key issue discovered was that the leg addition did little to support the FRD because the supports would buckle under "finger-loading." A solution to this would be to replace the thin supports with solid plate structures, which would be less likely to buckle under similar loading conditions.

The evaluation of the FRD also showed a number of positive aspects not seen in the computer model. First, the snap mechanism proved to work well even when a clearance fit was selected for the prototypes. This aided in the simplicity of assembling the full FRD and suggests that the tolerances for a production run could be relaxed, lowering costs and easing manufacturing. An additional unexpected aspect of the design is that it has built-in redundancy, needing only one of its two connections to secure the fin in the stowed position, suggesting two key things. First, this means the chances of the FRD failing (prior to mortar firing) are reduced in the current design, because both sides of the device must fail before a fin would be released. Secondly, it means the system could be simplified by removing one side of the span attachments. A simplification of the design would remove the redundancy but would also simplify the assembly procedure and reduce production costs.

With these observations in mind, it is now possible to return to the design and prototyping phase of development to further refine the design of the FRD. In the next iteration of design changes, a new version will be made based on team feedback to address the issues previously mentioned. This iteration will also allow for a greater degree of structural analysis with a more powerful version of the SolidWorks simulation tool that was recently made available to this project. With the combination of better analysis and a new design, the next iteration should fulfill a greater number of the requirements identified at the start of this project while further simplifying the part design.

4. Summary and Conclusions

The task of designing an FRD for the FCM system is an interesting challenge due to the specific requirements of the system. However, the development of a simple and reliable system may have an extended effect on a larger group of projects using deployable fin sets. With this goal in mind, a systematic approach was applied to the design process and a working prototype was developed. With the evaluation of this prototype, many design changes were planned and are pending the next design iteration. With further work on this project, a novel FRD will be completed to the point where it can be used for later testing and evaluation of the FCM project.

5. References

1. M395 Precision Guided Mortar Munition (PGMM). *Defense Update*. 2006. http://defense-update.com/products/x/xm395.htm (accessed 2011).
2. Hash, G. D.; Bojanowski, A. J.; Hightower, R. M.; Lishin, G. U.S. Patent No. 7,851,734. Washington, DC: U.S. Patent and Trademark Office, 2010.
3. Kuhnle, J.; Heitmann, T.; Niemeyer, T.; Arendt, N. U.S. Patent No. 6,314,886. Washington, DC: U.S. Patent and Trademark Office, 2001.

ADMNSTR
DEFNS TECHL INFO CTR
ATTN DTIC OCP (ELECTRONIC COPY)
8725 JOHN J KINGMAN RD STE 0944
FT BELVOIR VA 22060-6218

US ARMY RSRCH LAB
ATTN RDRL CIM P TECHL PUB
ATTN RDRL CIM L TECHL LIB
ATTN IMNE ALC HRR MAIL & RECORDS MGMT

INTENTIONALLY LEFT BLANK.