ARMY RESEARCH LABORATORY

Volume I: Select Papers

ARL Summer Student Research Symposium

ARL-TM-2009 August 2009

NOTICES

Disclaimers

Army Research Laboratory
Adelphi, MD 20783-1197

ARL-TM-2009 **August 2009**

Volume I: Select Papers

ARL Summer Student Research Symposium

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE ***(DD-MM-YYYY)*** August 2009	**2. REPORT TYPE** Final	**3. DATES COVERED (From - To)**
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-2009
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 6 August 2009.

Students selected by their directorate for competition participated in the one-day Summer Student Symposium on 6 August 2009. 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** Vallen Emery
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	248	**19b. TELEPHONE NUMBER** ***(Include area code)*** (301) 394-3585

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

Contents

INTENTIONALLY LEFT BLANK.

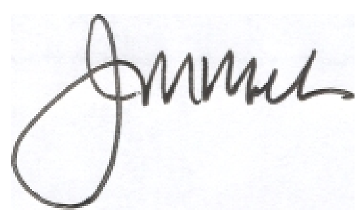
Director's Foreword

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, and 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 research experiences at ARL to students. During the summer of 2009, we supported research experiences at ARL for over 100 undergraduate and graduate students. Each of these students was required to write a paper describing the results of the work they performed while at ARL. While all of the students engaged in our summer research program had remarkable experiences and had excellent results, only a select number could be presented at our symposium. The papers contained herein were those presented at the symposium and they show that the nation truly does have high caliber people pursuing degrees in scientific and engineering disciplines.

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 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 6 August 2009.

Students selected by their directorate for competition participated in the one-day Summer Student Symposium on 6 August 2009. 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.

INTENTIONALLY LEFT BLANK.

Computational & Information Sciences Directorate (CISD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Extended Kalman Filter Implementation and Analysis on Man-Portable Autonomous Robots

by Jason Gregory

Contents

List of Figures

Abstract

Accurate position estimation is a vital factor for precision autonomous navigation and becomes even more important when applied to man-portable robots. Due to physical and atmospheric interference, global positioning system (GPS) data is often unreliable. Similarly, magnetic compass and odometry data can be inaccurate and prone to measurement drift. However, when combined with compass and odometry data, noise present in GPS data can be removed from the raw data by using an extended Kalman filter (EKF). This allows for consistent and accurate position estimations to be made, resulting in more precise navigation. The U.S. Army Research Laboratory's (ARL) Computational and Information Sciences Directorate (CISD) has a semi-autonomous man-portable robotic platform called the PackBot that is capable of GPS waypoint following. As with most small robots, the PackBot has limited processing resources. With autonomy software consuming much of these resources, providing accurate position estimation must be done in a computationally efficient manner. This paper will evaluate the performance of an efficient EKF implementation that provides state estimation on the CISD PackBot's organic processor. Further, the paper will explore alternate solutions to the state estimation problem, namely the unscented Kalman filter.

Acknowledgments

The author wishes to acknowledge the mentorship of Barry O'Brien.

1. Introduction/Background

Recently, man-portable autonomous robots have become ever more present in the battlefield because of their ability to protect warfighters by performing potentially life-threatening tasks. As urban warfare becomes more common, there is a greater demand for smaller robots operating completely autonomously. There already exists much research and development of larger fully autonomous robots such as vehicles; however, there has been less effort involving robots of smaller size and processing power (*1–7*). These smaller autonomous robots face several challenges, mainly due to their size, which include obtaining accurate data, networking and communications, and processing power. In an effort to address the challenge of autonomy, a filtered waypoint follower appears to be a vital component to man-portable robots.

By means of global positioning system (GPS) coordinates, a robot can not only identify its own location, but also determine a destination and navigate to that point. Using several of these destinations, or waypoints, a robot could follow a predetermined path entirely autonomously. However, GPS systems have several deficiencies that can greatly alter the quality of data. For example, GPS systems often lose satellites or have signal blocks, which consequently produce noisy data (*1*). This unreliable data, which occurs quite frequently in urban canyons, can greatly impact other systems when accuracy is a high priority (*1*). By applying a Kalman filter, GPS coordinates that would otherwise be unusable can be accurately estimated and vastly improve the reliability of the data. The Kalman filter implements a type of recursive predictor-corrector algorithm with two major steps (2–6). The first step in the algorithm projects both the current state and error covariance estimates forward to the proceeding time steps. The second part of the algorithm updates state estimates using corresponding feedback, ultimately determining the consistency of data and optimally removing zero-mean white Gaussian noise (*3*). Together, these two steps of the Kalman filter compute a minimum mean-squared error estimate for each state (*3, 8*).

As mentioned, the Kalman filter can recursively estimate the state of a linear system; however, for the case of a nonlinear system (i.e., localization of an accelerating and decelerating robot), the extended Kalman filter (EKF) is used in practice. Similar to the basic Kalman filter, the extended Kalman filter operates in two steps to predict and update measurements. The main difference between the EKF and the basic Kalman filter is the EKF process makes the nonlinear function model simulate linearization around an estimate by computing several Jacobian matrices at each time step (*4, 5*). These Jacobian matrices are calculated using two case-specific nonlinear differentiable equations—the first equation relates the state at time k-1 to the state at k and the second equation relates the current state to the measurement (*2, 5–7*). As a result, the EKF is an ad hoc state, first-order estimation (*1, 3, 4, 6, 8*) that approximates the optimality of

Bayes' rule by linearization (*2, 4, 5*). Using the process noise covariance, measurement noise covariance, state measurements, and calculated Jacobian matrices, the EKF recursively predicts and updates estimates at each time step from nonlinear transformations, as shown in figure 1.

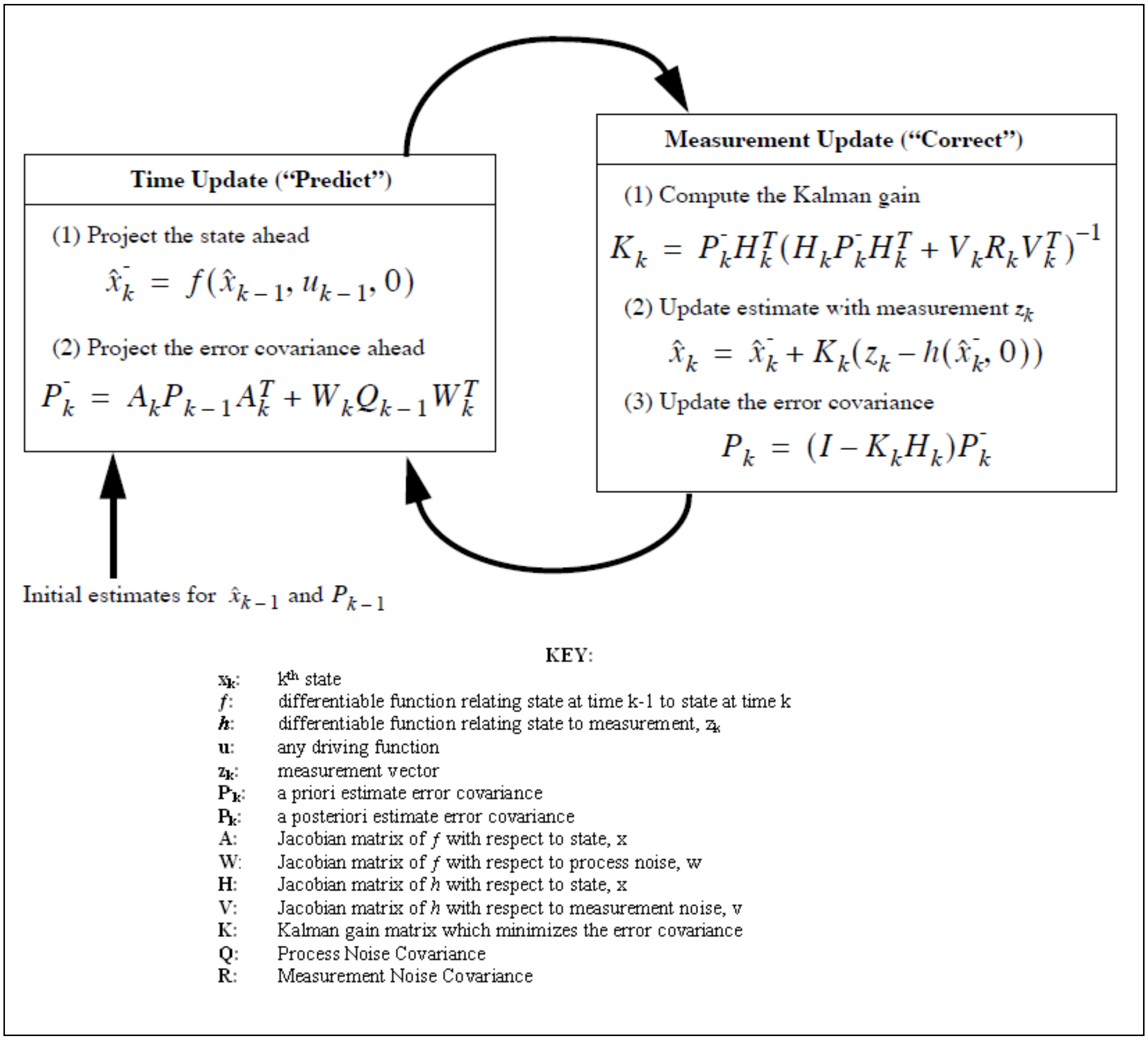


Figure 1. Graphical depiction of the EKF algorithm (*2*).

This paper discusses the efforts of implementing an EKF onto a man-portable robot, the PackBot shown in figure 2, already developed by the Computational and Informational Sciences Directorate (CISD). Several difficulties arise when implementing the Kalman filter and this paper focuses on the current progress in terms of improving the data used in an existing waypoint follower.

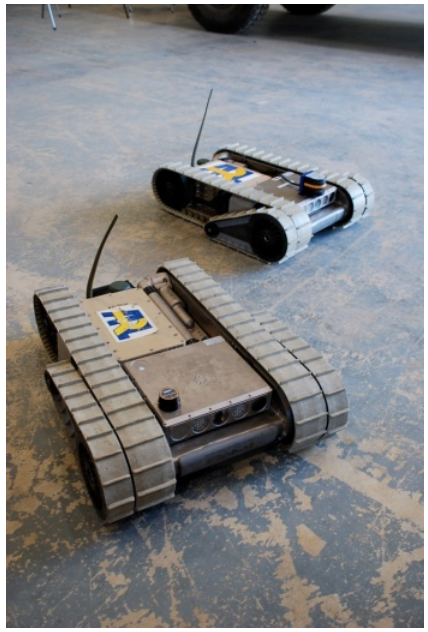


Figure 2. Semi-autonomous man-portable robotic platform called the PackBot.

2. Experiment/Calculations

An experiment was conducted with the primary objective of demonstrating the EKF could successfully remove noise from gathered GPS data in an open area. This experiment attempted to show the estimations produced by the filter provided more reliable and consistent coordinate information than the raw data read from the PackBot. An additional focal point of this experiment was to determine the available resources on the processor with the PackBot simultaneously providing video feed, avoiding obstacles, and filtering GPS data using the EKF. Because the PackBot currently uses the majority of its resources prior to the Kalman filter implementation, continued functionality was a main concern.

This experiment was conducted in the Department of Public Works (DPW) storage area just north of the Shipping and Receiving building. Prior to the experiment, eight waypoint locations were found using a handheld GPS unit. At each location, tape was placed on the ground in the shape of a large "X"' and then labeled with a number to clearly show the exact position. Using a combination of these points, two preliminary courses were developed for the PackBot to navigate, as shown in figure 3. For the test, an ad hoc network was established using a network node and a laptop. This local network allowed communication between the PackBot and the laptop so that the current state of the robot could be monitored. In order to determine the available resources, the laptop tracked CPU time using the “top” command from the Linux terminal. To simulate an external user requiring video and passing waypoint destinations to the PackBot, a second laptop was set up.

Test Course 1: Eight waypoints complete a circular path

Test Course 2: Four waypoints complete a circular path

Note: At the time of the satellite images, there were cars and other objects which were not present for the experiement

Figure 3. Two preliminary test courses from GPS coordinates.

This laptop ran software called CollectControl, which allows users to manually drive the PackBot, control the pan/tilt head, and view the video feed from the camera mounted on the head of the PackBot. This second laptop also ran a test program that allowed for a list of waypoint destinations to be passed to the PackBot. It was quickly determined that the original GPS coordinates found using the handheld GPS unit and the coordinates calculated by the PackBot differed considerably. To address the problem and ensure consistency, the PackBot was used to measure the GPS coordinates of the waypoints. While the physical position of each waypoint did not change, the GPS coordinates determined by the PackBot became the new waypoint locations. At the time of testing, two waypoints were unusable due to construction and other obstructions on the preliminary course paths. Because of these uncontrollable factors, a circular course was not used as planned. Instead, two waypoints from the list of eight were used for initial testing and the remaining four waypoints were plotted for future testing. To start the test, the PackBot was placed at an arbitrary location near the first waypoint and a list containing the selected waypoints was passed to the test program.

3. Results and Discussion

After passing the list of waypoints to the PackBot, the robot unexpectedly turned in the wrong direction and remained stationary. Due to unit conversion issues, the PackBot was unable to find the necessary heading to achieve the first waypoint. This problem was easily solved by correcting the conversion method. The second attempt proved to be more promising; however, the PackBot continued to operate unpredictably. The PackBot turned counter-clockwise toward the waypoint, overshot the heading by approximately 10°, and then remained stationary relatively close to the correct heading. Usually, the PackBot will overshoot the heading and then slowly turn back a few degrees to determine the correct heading for the first waypoint. Due to an unknown cause, the PackBot failed to rotate clockwise and was unable to obtain the correct heading necessary to begin the course. As the PackBot rotated, the robot sent its current state measurements and estimations to the laptop, which were then viewed in the terminal. The filter displayed a valid heading, GPS location, and GPS estimation for the PackBot's current state.

4. Summary and Conclusions

The PackBot has a working EKF, but due to library dependencies and compatibility issues the implemented filter is not fully functional. Prior to the filter implementation, the PackBot could provide 7 frame per second, 320 x 420 pixel video feed while simultaneously avoid obstacles. With the waypoint follower and newly implemented EKF running, the PackBot reported approximately 2% processor cycles to spare, which shows that the onboard processor can handle the new filter and maintain existing functionalities. Even though the processor can handle the new filter, the inevitable change in data structures and additional libraries created a problem with current mobility commands. As with this project, the details in implementing and improving new software are often times not fully considered. When implementing the EKF onto the PackBot, a change in data structures was not carefully examined and resulted in an unexpected malfunction in usual operations. The data structures necessary to implement the EKF were updated versions of the ones used by the obstacle avoidance software. The change of data structures allowed the EKF to estimate state information as planned, but the PackBot was no longer able to maneuver a simple course or avoid obstacles. In conclusion, the implementation of an EKF on the PackBot is an ongoing project and at the time of this publication testing has not been completed. Upon the completion of fully implementing and testing the EKF, further analysis will be presented in future publications.

5. References

1. El-Sheimy, N.; Shin, E.; Niu, X. Kalman Filter Face-off: Extended vs. Unscented Kalman Filters for Integrated GPS and MEMS Inertial. *GNSS* **2006**, 48–53.
2. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*, University of North Carolina at Chapel Hill, 2006.
3. Ahmad, M.; Chang, M., *Comparison of the Simple EKF, IEKF, and UKF for a Sensorless Full-Digital PMSM Drive*, University of Southern California, 2005.
4. Zhang, P.; Gu, J.; Milios, E. E.; Huynh, P., *Navigation With IMU/GPS/Digital Compass with Unscented Kalman Filter*, Dalhousie University, 2005.
5. Wikipedia, *Kalman Filter*, 2009. http://en.wikipedia.org/wiki/Kalman_filter (accessed 2009).
6 LaViola Jr., J. J, *A Comparison of Unscented and Extended Kalman Filtering for Estimating Quaternion Motion*, Brown University Technology Center for Advanced Scientific Computing and Visualization, 2003.
7. Toledo-Moreo, R.; Zamora-Izquierdo, M. A.; Úbeda-Miñarro, B. High-Integrity IMM-EKF-Based Road Vehicle Navigation With Low-Cost GPS/SBAS/INS. *IEEE Transactions on Intelligent Transportation Systems* **September 2007**, *8* (3).
8. Wan, E. A.; Merwe, R. *The Unscented Kalman Filter for Nonlinear Estimation*, 1999.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Stress Driven Surface Instabilities in Ionic Solids With Charged Point Defects

by Steven F. Henke

Contents

List of Figures

List of Tables

Abstract

Oxides exhibit a wide range of functional characteristics that make them suitable for many technological applications, including electronic devices, sensors, solid-state lighting, and catalysis. In many of these applications, the oxides phases form epitaxial systems with the underlying support structure, giving rise to internal stresses that affect mass transport within the system. Stress-Driven Rearrangement Instability (SDRI) theory postulates that diffusion in stressed solids can lead to surface morphological instability, an effect that is currently believed to be important for elevated-temperature deposition or annealing of thin films. Building upon previous works, we present a continuum reformulation of the SDRI theory, with surface diffusion and bulk point defect distributions. In oxides, however, vacancies are charged, and their distribution is influenced by space-charge formation and the associated internal electrostatic fields. Our model considers such electrostatic effects, coupled with elasticity, in modeling point defects within the bulk. The present theory will enable us to understand the coupled electric, elastic, and chemical interactions in ionic thin film systems, as well as the impact of these interactions on the thin film morphology changes during film deposition and annealing.

Acknowledgments

The author wishes to acknowledge the mentorship of Peter W. Chung.

1. Introduction

Lattice mismatched oxide systems are essential components in many electric, electro-mechanical, and electro-optical devices. Such devices have numerous military and civilian applications, including microwave transducers, chemical sensors, catalysis and solid-state lighting. There is current interest in miniaturizing these components to enhance performance, improve properties, and reduce device size and weight. At these small scales, it is critical to have control of the film surface morphology. For example, in capacitors and other applications that require layering, it is necessary to have a flat interface to increase surface contact. However, in self-assembled nanostructures, it may be desirable to have a regularly spaced array of surface bumps.

According to Stress-Driven Rearrangement Instability (SDRI) theory, the inevitable presence of mobile point defects may complicate efforts to control surface morphology. Furthermore, point defects are believed to play an important role in determining the properties and reliability of thin oxide films. Specifically, the distribution of oxygen vacancies within an oxide layer will strongly influence the system's electrical characteristics, since each vacancy is associated with a net electric charge.

In this work, we pose a continuum space-charge model for the distribution for point defects in oxide thin film systems. The equations of elasticity are employed to describe the influence of film-substrate mismatch strain due to thermal expansion, and the difference in film and substrate lattice constants. We then show that the combined effects of the electrostatic and elastic mechanisms determine the updated morphology of the free surface located at the film-vapor interface. We conclude with preliminary results for the stability of barium titanate (BTO) thin films perturbed by "bumps" and "rings", which are representative of surface features that have been previously observed in oxide systems.

2. Theory

2.1 Electrostatics

The fundamental assumption of the electrostatic approximation is that electric currents and magnetic fields, if present, must be time invariant. This is enforced by requiring that the electric and magnetic fields are irrotational, thus decoupling their mutual effects.

$$\textbf{Faraday's Law:} \qquad \epsilon_{ijk}E_{j,i} = -\frac{\partial B_k}{\partial t} = 0 \tag{1}$$

$$\textbf{Ampère's Law:} \qquad \epsilon_{ijk} B_{j,i} = \hat{\mu}_0 \left(J_k + \varepsilon^\circ \frac{\partial E_k}{\partial t} \right) = 0 \tag{2}$$

Employing the electrostatic approximation, the electric field E_i is conservative, and, therefore, can be written as the gradient of a scalar potential (provided the domain is simply connected—that is, has no holes). We interpret the quantity ϕ as the electrostatic potential.

$$E_i = -\phi_{,i} \tag{3}$$

Charges can be categorized based upon whether they are bound, or free to move. Bound charges are described by the electric displacement,

$$D_i = \varepsilon^\circ E_i + P_i \tag{4}$$

which includes the effect of charges bound in space (where ε° is the permittivity of free space), and charges bound in matter P, the material's polarization. Any outward flux of charge from a displacement field must be a free charge; thus, the free charge density $\hat{\rho}$ is given by the divergence of the electric displacement.

$$\textbf{Gauss' Law:} \qquad D_{i,i} = \hat{\rho} \tag{5}$$

The free charge density of an ionic solid containing charged point defects is related to the densities of the defect species

$$\hat{\rho} = e \sum_i z^a N^a \tag{6}$$

where e is the elementary charge, and z is the valency of defect species a. Under the assumptions of Maxwell-Boltzmann statistics (e.g., at high temperature and low defect concentrations), density of a point defect species at thermal equilibrium is given by

$$N^a = N^{\circ a} \exp\left(-\frac{\mathcal{E}^a + e z^a \phi}{kT} \right) \tag{7}$$

where $\mathcal{E}^a$ is the defect formation energy of species a, k is Boltzmann's constant, T is the absolute temperature, and $N^{\circ a}$ is the number concentration of atoms in sub-lattice a in a perfect crystal. Combining these equations yields the Poisson-Boltzmann equation, which determines the space-charge distribution within an ionic crystal containing a dilute concentration of charged point defects.

$$\textbf{Poisson-Boltzmann:} \qquad (-\varepsilon^\circ \phi_{,i} + P_i)_{,i} = e \sum_a z^a N^{\circ a} \exp\left(-\frac{\mathcal{E}^a + e z^a \phi}{kT} \right) \tag{8}$$

We are interested in applying this model to thin film oxide systems, which are used in numerous electric devices. Some members in this class of materials exhibit complex coupling of electrical, thermal, and mechanical properties—examples include pyroelectricity, ferroelectricity, and spontaneous polarization. To simplify analysis, we exclude accounting for these properties in this model and make the assumption of a linear relationship between the electric displacement and electric field.

$$D_i = \varepsilon^\circ E_i + P_i \longrightarrow D_i = \varepsilon_{ij} E_j \tag{9}$$

where the permittivity coefficients are the product of the electric constant and relative permittivity tensor $\varepsilon_{ij} = \varepsilon^\circ \varepsilon^R_{ij}$.

In summary, the electrostatic field within our oxide system arises due to the presence of point defects. We make the simplifying assumption that as the surface rearranges, point defects within the bulk of our system are immediately redistributed to their thermal equilibrium concentrations, which are known using Maxwell-Boltzmann statistics. Our governing equation is the statement of charge conservation, subject to either a specified voltage V or surface charge $\hat{\sigma}$ at all points on the boundary of the computational domain. We write the formal mathematical statement of the problem.

Given: $\hat{\rho} : \Omega \rightarrow \mathbb{R}$, $V : \Gamma_\phi \rightarrow \mathbb{R}$, and $\hat{\sigma} : \Gamma_{\hat{\sigma}} \rightarrow \mathbb{R}$, find $\phi : \bar{\Omega} \rightarrow \mathbb{R}$ such that

$$D_{i,i} = \hat{\rho} \qquad \text{in } \Omega \tag{10}$$

$$\phi = V \qquad \text{on } \Gamma_\phi \tag{11}$$

$$[D_i] n_i = \hat{\sigma} \qquad \text{on } \Gamma_{\hat{\sigma}} \tag{12}$$

2.2 Elasticity

The Cauchy stress tensor σ_{ij} is related to the material strain through the generalized Hooke's law

$$\sigma_{ij} = C_{ijkl}(\epsilon_{kl} - \epsilon^\circ_{kl}) + \sigma^\circ_{ij} \tag{13}$$

where C_{ijkl} are the elastic coefficients, ϵ_{kl} is the material strain, ϵ°_{kl} is the initial strain, and σ°_{ij} is the initial stress. The initial strain ϵ°_{ij} includes the mismatch strain and the eigenstrain

$$\epsilon^\circ_{ij} = \epsilon^{\text{mm}}_{ij} + \epsilon^{\text{th}}_{ij} \tag{14}$$

where the mismatch strain and the thermal strain are given in terms of the substrate lattice constant a^s, film lattice constant a^f, and thermal expansion coefficients α_{ij}.

$$\epsilon^{\text{mm}}_{ij} = (a^f - a^s)/a^s \qquad \epsilon^{\text{th}}_{ij} = -\alpha_{ij} T \tag{15}$$

Since these materials are often brittle, we assume that this system is well-represented by infinitesimal deformation elasticity, which relates film strain to the symmetric part of the displacement gradients.

$$\epsilon_{ij} = u_{(i,j)} = \frac{1}{2}(u_{i,j} + u_{j,i}) \tag{16}$$

Our governing equation is obtained by conservation of linear momentum

$$\sigma_{ij,j} = 0 \tag{17}$$

and is subject to either a prescribed displacement u_i or traction $\hat{t}_i$ at all points on the boundary of the computational domain. We write the formal mathematical statement of the problem.

Given: $f_i : \Omega \rightarrow \mathbb{R}$, $\bar{u}_i : \Gamma_u \rightarrow \mathbb{R}$, and $\hat{t}_i : \Gamma_{\hat{t}} \rightarrow \mathbb{R}$, find $u_i : \bar{\Omega} \rightarrow \mathbb{R}$ such that

$$\sigma_{ij,j} = 0 \qquad \text{in } \Omega \tag{18}$$

$$u_i = \bar{u}_i \qquad \text{on } \Gamma_u \tag{19}$$

$$[\sigma_{ij}]n_j = \hat{t}_i \qquad \text{on } \Gamma_{\hat{t}} \tag{20}$$

2.3 Surface Rearrangement

The surface evolution equation is obtained through mass balance at the film-vapor interface

$$v_i n_i = -\sum_a \Omega^a (q^a_{\alpha,\alpha} + Q^a_i n_i) \tag{21}$$

where v_i is the free surface velocity, n_i is the free surface normal direction (oriented to point from the film into the vapor), Ω is the atomic volume of the defect species, $Q_i n_i$ is the evaporation flux, and $q_{\alpha,\alpha}$ is the divergence of the surface vacancy flux. We note the convention that Greek indices are used for surface coordinates. The surface diffusive flux is given by Fick's law

$$q^a_\alpha = -M^a_{\alpha\beta} \mu^a_{,\beta} \tag{22}$$

where the surface gradient of the chemical potential μ is linearly proportional to the diffusive flux. These proportionality coefficients are given by the mobility tensor $M_{\alpha\beta}$, which is defined in terms of the diffusivity tensor $D_{\alpha\beta}$ and surface concentration $\bar{N}$,

$$M^a_{\alpha\beta} = \frac{\bar{N}^a}{kT} D^a_{\alpha\beta} \tag{23}$$

the surface electro-elasto-chemical potential is postulated to have the form

$$\mu = \mu^\circ + \frac{1}{2}\sigma_{ij}c_{ij} + \frac{1}{2}D_i E_i + \sum_a N^a \mathcal{E}^a + \kappa_{\alpha\beta}\gamma_{\alpha\beta} \tag{24}$$

where we have introduced the surface curvature $\kappa_{\alpha\beta}$ and surface energy $\gamma_{\alpha\beta}$.
In summary, the surface evolution equation governs the motion of the film's free surface, subject to prescribed displacements or velocities along the boundaries of the computational domain.

3. Methods

We solve the fully coupled system using a finite element method, a method in which the computational domain is decomposed into a collection of basic shapes, called elements, and approximate solutions are generated from a chosen set of basis (shape) functions by solving a system of equations. A subclass of more general weighted residual methods, a finite element method is used to solve differential equations in their "weak" form. To formulate a problem in its weak form, the governing differential equation is written as a weighted integral, and the divergence theorem is repeatedly applied to "weaken" the continuity requirements on each field variable. An additional benefit of the weak formulation is that certain boundary conditions are automatically included (called natural), but the remaining boundary conditions must be enforced manually (called essential).

We first write the weak formulation for the electrostatics equations.

Given: $\hat{\rho} : \Omega \rightarrow \mathbb{R}$, $V : \Gamma_\phi \rightarrow \mathbb{R}$, and $\hat{\sigma} : \Gamma_{\hat{\sigma}} \rightarrow \mathbb{R}$, find $\phi \in \mathscr{S}$ such that $\forall\ w \in \mathscr{V}$

$$-\int_{\Gamma_{\hat{\sigma}}} w\hat{\sigma}\ ds - \int_\Omega w_{,i}\varepsilon_{ij}\phi_{,j}\ dv + \int_\Omega w\hat{\rho}\ dv = 0 \tag{25}$$

$$\phi = V \text{ on } \Gamma_\phi \tag{26}$$

Next, we write the weak formulation for the elasticity equations.

Given: $f_i : \Omega \rightarrow \mathbb{R}$, $\bar{u}_i : \Gamma_u \rightarrow \mathbb{R}$, and $\hat{t}_i : \Gamma_{\hat{t}} \rightarrow \mathbb{R}$, find $u_i \in \mathscr{S}$ such that $\forall\ w \in \mathscr{V}$

$$\sum_{i=1}^{n_{sd}} \left(\int_{\Gamma_{\hat{t}}} w_i\hat{t}_i\ ds \right) - \int_\Omega w_{i,j}C_{ijkl}u_{(k,l)}\ dv + \int_\Omega w_{i,j}C_{ijkl}\epsilon^\circ_{kl}\ dv - \int_\Omega w_{i,j}\sigma^\circ_{ij}\ dv + \int_\Omega w_i f_i\ dv = 0 \tag{27}$$

$$u_i = \bar{u}_i \text{ on } \Gamma_u \tag{28}$$

These equations, plus careful approximation of surface geometrical quantities (such as nodal normal vectors n_i and mean curvature κ), determine the quantities necessary to perform a surface update:

$$v_i n_i = \sum_a \Omega^a \left(\left(M^a_{\alpha\beta}\mu_{,\beta} \right)_{,\alpha} + Q^a_i n_i \right) \tag{29}$$

where

$$\mu = \mu^\circ + \frac{1}{2}\sigma_{ij}\epsilon_{ij} + \frac{1}{2}D_iE_i + \sum_a N^a\mathcal{E}^a + \kappa_{\alpha\beta}\gamma_{\alpha\beta} \tag{30}$$

4. Results

Our computational domain (figure 1) is three-dimensional with a film layer occupying the lower portion surrounded by a vapor layer. The vertical sides of the domain are periodic. The substrate is not included in this setting, but the influence of a rigid substrate is included in the initial strain of the film's reference configuration. The Poisson-Boltzmann equation is solved over the entire domain with the film grounded at its base and zero potential far away from the film's surface. The electric field arises due to the difference in charge density between the film and the vapor. The stress equilibrium equation is solved within the film only. The film-substrate interface is fixed at zero displacement. The film's free surface is traction-free. The reference configuration is chosen to have no initial stress, but we again note that lattice mismatch and eigenstrain are included in the initial strain term. For initial analyses, we consider a single defect species, oxygen vacancies ($\Omega = 9.05 \cdot 10^{-4}\,\mathrm{nm}^{-3}$). In both BTO and strontium titanate (STO) (table 1), each oxygen vacancy has an equivalent charge of magnitude two electrons (+2e). We neglect any evaporative fluxes ($Q_i n_i = 0$) and choose a temperature appropriate for an annealing step, $T = 1000\,K$.

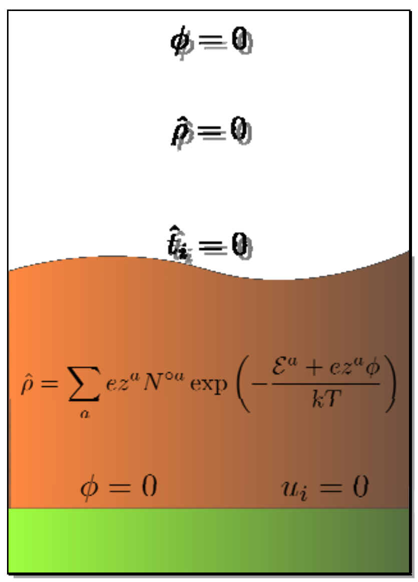


Figure 1. Computational domain—setup and constraints.

Table 1. Descriptions and values for BTO and STO.

Description	**Symbol**	**$BaTiO_3$ (Cubic)**	**$SrTiO_3$ (Cubic)**
Sites/volume	N°	47.02	52.16
Valency	z	$+2$	$+2$
Formation energy	$\mathcal{E}$	$4.49\,\mathrm{eV}$	$2.57\,\mathrm{eV}$
Permittivity	$\varepsilon/\varepsilon^\circ$	200	230
Young's modulus	Y	$114\,\mathrm{GPa}$	$190\,\mathrm{GPa}$
Poisson's ratio	ν	0.31	0.23
Surface energy	γ	$0.20\,\mathrm{eV\,nm^{-2}}$	$0.19\,\mathrm{eV\,nm^{-2}}$

Surface perturbations were chosen to be either bumps (figure 2) or rings (figure 3) with amplitude A and width W in nanometers.

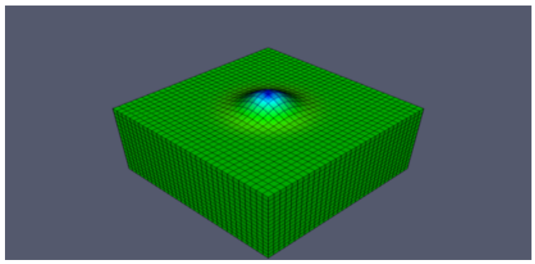


Figure 2. Bump morphology.

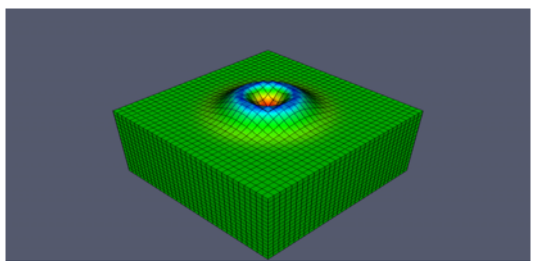


Figure 3. Ring morphology.

We varied the amplitude and width over a range of values that we hypothesized would be both interesting and plausible (figures 4−7). For each value, we computed results for one time step and examined the signed normal velocity at the center of the perturbation. Due to the symmetry of the shapes chosen, we reasoned that initial downward velocity of the bumps (upward velocity of the ring center) would lead to stable configurations due to the reduced curvature. Similarly, initial upward velocity of the bumps (downward velocity of the ring centers) would result in unstable morphologies where curvature values would increase without bound. For each morphology, we vary the amplitude of the initial perturbation $A = 0, \ldots, 10$ and its width $W = 1, \ldots, 100$. We plot initial velocities at the geometric center of the perturbations.

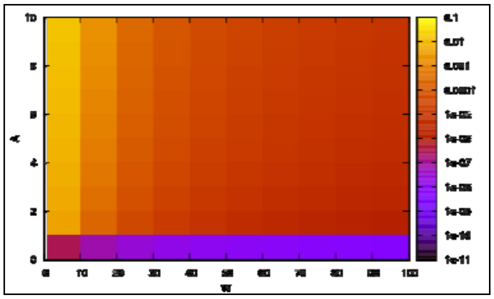


Figure 4. Barium titanate—bump morphology.

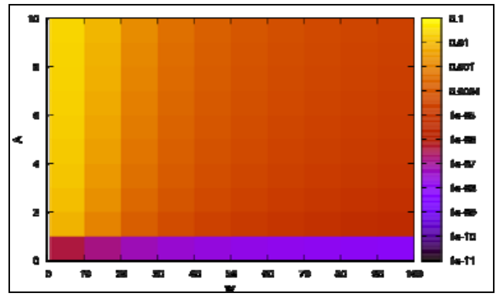


Figure 5. Barium titanate—ring morphology.

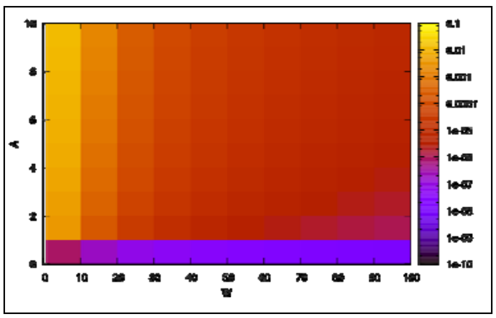


Figure 6. Strontium titanate—bump morphology.

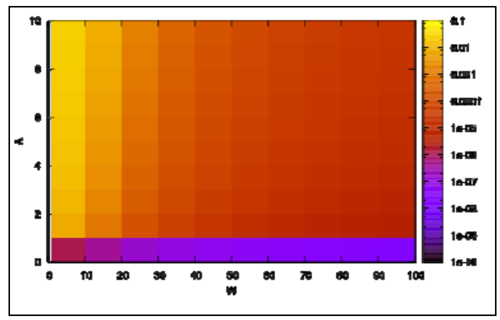


Figure 7. Strontium titanate—ring morphology.

In all cases, the peak velocity was negative for bump morphologies and positive for ring morphologies, so the configurations will be stable. This result is to be expected for the chosen material parameters and geometries, as they are typical of experiments where this instability is not readily observed. Note that normal velocities computed for cases where the perturbation amplitudes are zero, are zero within the chosen tolerance of the solver.

5. Summary and Future Work

A continuum model has been developed for ionic solids containing charged point defects. This model includes thermodynamically consistent coupling of electrostatic, elastic, and surface effects that describe the morphological evolution of the free film-vapor interface. We are using this model to explore the stability of the film surface to various perturbations. Preliminary results were demonstrated using BTO as an example material, and similar data are being computed for STO. Future work will include an accounting for multiple defect species and a more exhaustive investigation of surface perturbations, and will include additional coupling based on our experience with this model.

6. References

1. El-Azab, A.; Liang, Y. *Philosophical Magazine* **2003**, *83*, 3847–3869.
2. Clayton, J. D.; Chung, P. W.; Grinfeld, M. A.; Nothwang, W. D. *International Journal of Engineering Science* **2008**, *46*, 10–30.
3. Clayton, J. D.; Chung, P. W.; Grinfeld, M. A.; Nothwang, W. D. *Mechanics Research Communications* **2008**, *35*, 57–64.
4. Kliewer, K. L.; Koehler, J. S. *Physical Review* **1965**, *140*, 1226–1240.
5. Meyer, M.; Desbrun, M.; Schroder, P.; Barr, A. H. *Visualization and Mathematics III* **2003**, 35–57.
6. Hughes, T.J.R. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Dover, 2000.
7. Belytschko, T.; Liu, W. K.; Moran, B. *Nonlinear Finite Elements for Continua and Structures*, Wiley, 2000.
8. Moulson, A. J.; Herbert, J. M. *Electroceramics: Materials, Properties, Applications*, 2003.
9. Pimpinelli, A.; Villain, J. *Physics of Crystal Growth*, Cambridge, 1998.

Human Research & Engineering Directorate (HRED)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Calibration of an Instrumented Treadmill: Instrumented Pole and Protocol Design

by Michaela McBride

Contents

List of Figures

Abstract

Force plates are used to measure ground reaction forces during gait analysis. Instrumented treadmills are built on or around force plates to collect kinetic data for successive gait cycles. To ensure the accuracy of these instrumented treadmills, frequent and thorough calibration is necessary. One tool used for force plate and instrumented treadmill calibration is an instrumented pole, which is a pole equipped with reflective markers and a load cell to measure force. This paper describes the design of an instrumented pole used for calibration and the protocol that will be used to validate an instrumented treadmill that will be installed in the Human Research Engineering Directorate's (HRED) new Soldier Performance and Equipment Advanced Research (SPEAR) Facility. Data collected during this procedure will be used to determine the validity and reliability of the instrumented treadmill, and for calibration of the treadmill.

Acknowledgments

The author wishes to acknowledge the mentorship of Dr. Michael LaFiandra and Dr. Philip Crowell.

1. Introduction/Background

To optimize Soldier performance, it is necessary to understand how backpacks, footwear and other equipment affect performance. Wearing a backpack and the weight of the load carried during walking and running change the kinetics of the movement (*1*). These changes can influence performance and susceptibility to injury (*2*). Gait analysis, the study of human locomotion, traditionally includes the measurement of the forces, moments, and torques that act on the lower body in addition to kinematic data during walking and running (*3*). Knowledge of Soldier biomechanics is useful for understanding the effects of Soldier equipment on physical performance.

Measuring ground reaction forces using force plates is a key element of gait analysis. The ground reaction force is a three-dimensional vector with one vertical and two horizontal components and is the most common force that acts on the body (*4*). Conventionally, floor-mounted force plates are used to collect three-dimensional kinetic data. Three-dimensional force and moment data are necessary to calculate joint moments, center of pressure, point of force application, and mechanics of the body center of mass (*5*).

While floor-mounted force plates are useful tools, they have several significant shortfalls. Space limitations restrict the number of force plates in gait analysis labs. Because participants only make sufficient contact with the force plates once out of every three steps, collecting gait data during overground walking requires many trials to be collected to ensure useful data for analysis (*6*). Asking participants to target force plates during overground walking trials causes modifications in gait, distorting the data (*3*). Recording numerous trials makes it more difficult to collect trials at a constant velocity and participants may start to fatigue (*7*).

To address these issues, instrumented treadmills were designed. An instrumented treadmill is a treadmill built or mounted on top of force plates or ground dynamometers located under the treadmill feet, allowing kinetic data to be collected during treadmill walking trials (*3*). Instrumented treadmills are able to measure ground reaction forces for many gait cycles in succession. Recording ground reaction forces for successive gait trials enables researchers to calculate accurate and reliable step parameters (*7*).

Instrumented treadmills are available in several different configurations. Treadmills with a single force plate can detect the foot contact and foot off events in the gait cycle using force platform and motion capture data (*8*). It isn't possible, however, for single force plate instrumented treadmills to collect separate ground reaction force profiles for the right and left feet. Instrumented treadmills with multiple force platforms are required to collect data relative to each foot for subjects who are walking. There are three configurations of force platforms available: two force plates placed side-by-side; two platforms arranged in a front-back orientation; and three force plates, one in the front and two placed side-by-side in the back (figure 1). Although instrumented treadmills with multiple force plates allow right and left foot kinetics to be collected separately, if both feet are in contact with the same force plate simultaneously the data is unusable (*3*).

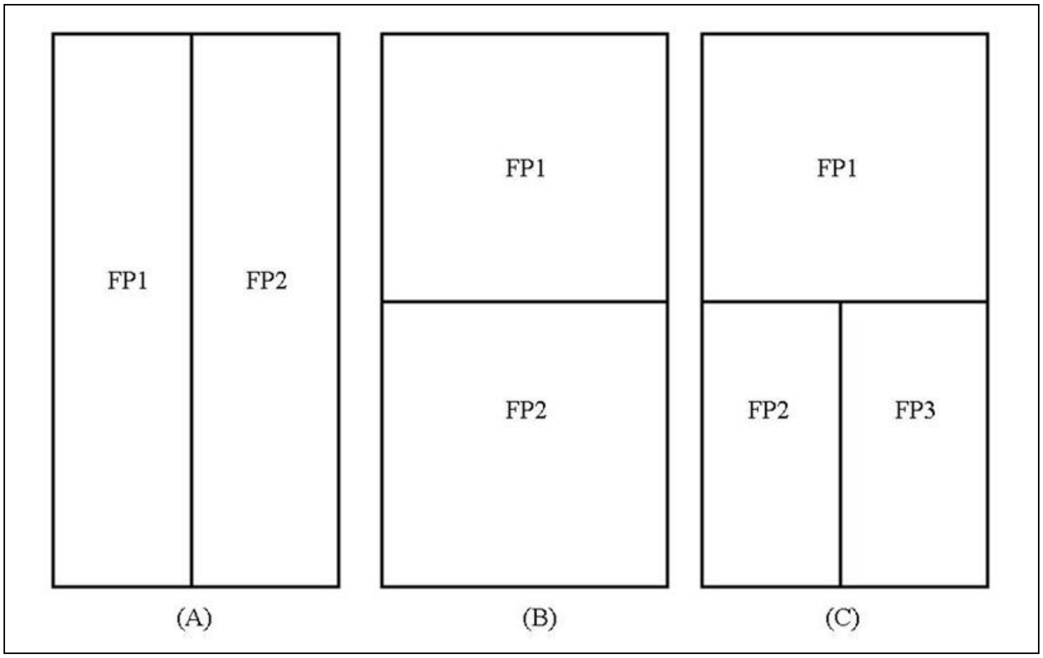


Figure 1. Instrumented treadmill with multiple force places (FPs). Treadmill A has two force places in a side-by-side arrangement (*7*, *9*, *10*). Treadmill B shows the force plates arranged in a front-back setup (*11*). Treadmill C has the three force plate arrangement with one in front and two side-by-side in the back (*3*).

Like floor-mounted force plates, instrumented treadmills collect vertical and horizontal ground reaction forces during walking and running trials. There are many benefits to using instrumented treadmills over traditional, floor-mounted force plates. Instrumented treadmills require less space and allow kinetics to be collected during successive gait trials without the increased cost of adding more floor-mounted force plates. Additionally, since the force plates in an instrumented treadmill are the size of the treadmill itself, foot placement isn't restricted (figure 2). This eliminates the potential for the participants to miss the force plates during the walking trial. Since instrumented treadmills reduce the size of the data collection area, it is easier to collect other data simultaneously (e.g., kinematic data, electromyography [EMG] and oxygen consumption) (*12*). Using an instrumented treadmill allows multiple trials to be collected at constant speed, even across participants, which is difficult with overground walking. Furthermore, treadmill walking allows gait trials to be recorded while the participant is moving at steady state. Acceleration and deceleration affect energy expenditure and gait parameters; consequently, the most representative data is collected during steady-state walking and running (*13*). Instrumented treadmills can be used to quantify the mechanical work done on the participant's center of mass during movement; as well as to study conditions such as gait transitions and response to stability disruptions (*5*).

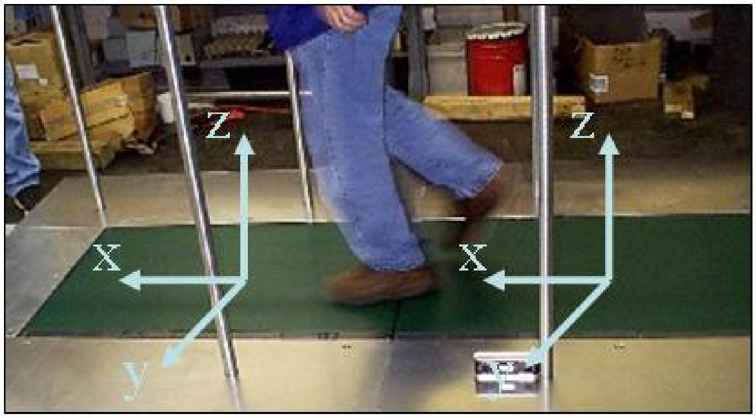


Figure 2. A participant walks on an instrumented treadmill with two force plates.

Despite the many advantages of instrumented treadmills, there are noteworthy issues to take into account. Since instrumented treadmills are built or mounted on the force plates, the accuracy of load measurements may be distorted by the compliance and dynamics of the treadmill structure (*9*). The additional mass of the treadmill can lower the natural frequency of the system, as can the reduced stiffness from a treadmill with a less than rigid frame (*11*). Belt speed variations, high cost, and potential errors of force and center of pressure measurements are additional key disadvantages of instrumented treadmills (*14*). Many of these issues can be rectified by meticulous and accurate calibration procedures.

Frequent calibration is necessary to ensure floor-mounted force plates and instrumented treadmills are configured and working correctly. Gait analysis commonly integrates motion analysis, EMG, and energy expenditure measurement systems with force plates or an instrumented treadmill. Monitoring these systems with regular, thorough calibration is important to identify performance changes and prevent these changes from affecting the data collected (*15*). If data collection systems are not set up or synchronized correctly, errors are introduced into the data, which can then be amplified when joint reaction force is calculated from the non-synchronized ground reaction force and kinematic data (*16*). When collecting data with multiple systems, each system initially relates the data to its own time-space reference system. These measurements are translated into a global time-space reference system, which is obtained through calibration (*17*). Center of pressure location is the most difficult measurement for instrumented treadmills to obtain accurately, so confirming the measurement with another method with known accuracy is a vital part of the calibration process (*14*).

There are many methods in the literature that can be used to calibrate force plates and instrumented treadmills, including using known weights, an instrumented pole, and human participant tests (*9*, *15*, *16*, *17*). Using an instrumented pole is an accurate and efficient method of calibration, allowing many variables (e.g., center of pressure, measurement repeatability, linearity) to be evaluated at one time. The advantage of an instrumented pole is the ability to synchronize the kinematic and kinetic data simultaneously.

An instrumented pole is a rigid rod with reflective markers fixed to it and a load cell that measures applied forces (*9*). Forces are applied to the instrumented treadmill manually using the instrumented pole while the motion capture system, force plate or instrumented treadmill, and instrumented pole are simultaneously collecting data. The pole is used to apply forces at various positions on the force plates and moved through a range of angles at each position. This calibration procedure successfully evaluates system performance and detects potential problems (*15*).

There are various instrumented poles used for calibration presented in the literature (*9*, *15*, *16*). There is a basic design consistent among studies: a rigid pole with a handle, a pointed tip used to apply loads, and multiple cross bars with reflective markers at the ends (figure 3). To allow the measurement of applied forces, a load cell is added near the tip.

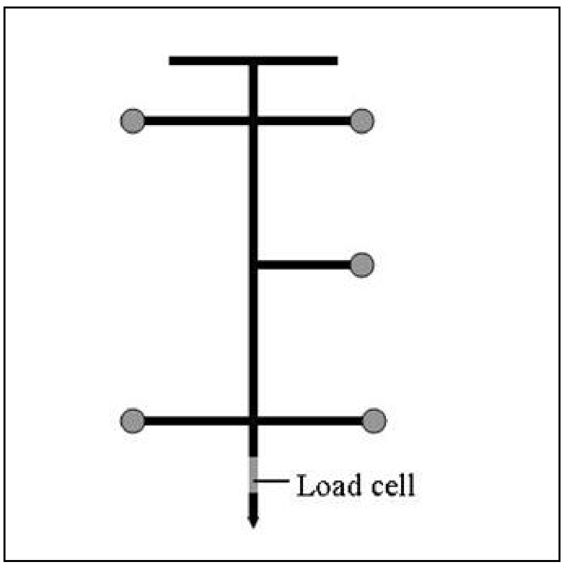


Figure 3. The basic design of an instrumented pole used for force plate and instrumented treadmill calibration.

Motion Lab Systems (Baton Rouge, LA) manufactures the Mechanical Testing Device (MTD-2) (*18*) for calibration of force plates and instrumented treadmills. The MTD-2 is a 1.0-m-long machined rod that has threaded conical points at each end and five 0.16-m rods with tracking markers on one end that can be attached to the 1.0-m rod. A loading plate is used to apply forces with the instrumented pole. The design of the pole allows the researcher to apply force with a negligible moment during calibration. This device has been instrumented with the LC202-3K, a single-axis load cell made by OMEGA Engineering, Inc. (Stamford, CT) (*19*). The miniature universal load cell has a 1361-kg (3000 lb) capacity and comes in a stud mount style. Custom made aluminum parts are required to incorporate the load cell into the MTD-2. When the load cell is added, the total length = 1.079 m and the total weight = 1.4 kg (*9*). The instrumented MTD-2 is used to apply force while the motion capture system and instrumented treadmill are collecting data. A protective plate with two bearing points is used to keep the tip of the pole in place while applying loads using the MTD-2.

The Advanced Medical Technology, Inc. (AMTI) (Newton, MA), walker sensor load cell (MCW-250) was added to a custom-made pole to measure forces in one study (*15*). The AMTI walker sensor measures three-dimensional forces and moments and requires an amplifier. The mass of the pole and force sensor system was 1.1 kg and the total length was 1.4 m. The base of this pole is a rubber-coated rounded metal foot, eliminating the need for a protective plate (*15*).

AMTI has recently developed a load cell created specifically for use with MTD-2. The CalTester three-axis load cell measures forces in three dimensions (x, y, and z). It was designed to fit the MTD-2; consequently, the two tools can be integrated easily by removing the conical tip, securing the load cell on the MTD-2 threading, and replacing the tip. No amplifier is required for this load cell, which provides a direct analog output. The maximum load can be customized up to 22.7 kg (50 lb) vertical force and 9.1 kg (20 lb) horizontal force.

The software used to process calibration data varies across studies. MATLAB (MathWorks, Natick, MA), EVaRT (Motion Analysis Corp., Santa Rosa, CA), and the Vicon Workstation and Body Builder (Vicon Motion Systems, Los Angeles, CA) have been used to process and evaluate the results of calibration tests (*9*, *15*). C-Motion (Germantown, MD) makes the CalTester software package for use with the MTD-2. The software does not currently have the capability to process the data from a load cell; however, C-Motion is currently developing these capabilities.

Calibrating the force plates and instrumented pole ensures that the data used for calibrating the instrumented treadmill is accurate. If there are discrepancies between force plate and instrumented pole force measurements during initial tests, correction factors are calculated and applied to data collected with the instrumented pole (*9*).

The natural frequency and electrical and mechanical noise of the instrumented treadmill measured during calibration is used to design filters used for subsequent data processing (*7*). The maximum belt speed error while the treadmill is unloaded and during participant trials is important to note and take into account during future studies. If the static and dynamic tests using the instrumented pole reveal differences in linearity, point of force contact, center of pressure, force accuracy, or measurement repeatability, appropriate computer programs are written to correct for these errors (*9*).

If drift affects force plate measurements during tests replicating normal use, a computer program is written to correct the drift error (*3*). Differences between overground and treadmill walking ground reaction force and center of pressure profiles are compared to values seen in the literature (*6*). If the data differs significantly more than the published variation between the two conditions, correction factors are calculated.

2. Purpose and Scope of Project

The U.S. Army Research Laboratory (ARL), Human Research Engineering Directorate (HRED) is preparing for the construction of a new research facility, the Solider Performance and Equipment Advanced Research (SPEAR) Facility. The SPEAR Facility will allow researchers to study many aspects of Soldier performance simultaneously, including biomechanical and physiological metrics and cognitive readiness. The biomechanics lab in the SPEAR Facility will include an instrumented treadmill, an EMG system to measure muscle activity, a portable cardio pulmonary exercise testing device to quantify energy expenditure, and portable force plates. This technology will provide the capability to evaluate warfighter performance more completely.

The purpose of this project was threefold: (1) to evaluate existing instrumented poles, (2) to design the instrumented pole that will be used during the calibration process, and (3) to create the protocol that will be used to validate the instrumented treadmill at the SPEAR Facility.

3. Methodology

3.1 Existing Instrumented Pole Evaluation

The instrumented pole consists of two main components: the machined pole and the load cell. The pole must be rigid, equipped with reflective markers for motion tracking, and have a means for the user to apply forces. In order to evaluate the instrumented treadmill at loads representative of normal use, the load cell in the instrumented pole must have a capacity of at least 136 kg (300 lb).

3.2 Equipment

3.2.1 Instrumented Treadmill

A compact tandem force-sensing treadmill (Advanced Medical Technology, Inc., Watertown, MA) will be installed in the SPEAR Facility and validated using the procedure presented in this paper. The two force plates will be set up in a front-back configuration and collect three dimensional (*x*, *y*, and *z*) force and moment data (figure 4). The force capacity will be 8,800 N in the vertical direction and 4,500 N in the horizontal directions. This treadmill has two belts that can be controlled independently or synchronized, each with an area of 0.74 x 0.66 m (1.48 x 0.66 m total). The maximum speed will be 5.55 $m \cdot s^{-1}$ (12.4 mph) and the treadmill will be able to incline up to a 25% grade (*20*). The belt direction will be reversible, allowing uphill and downhill walking and running to be evaluated. Data will be collected at 600 Hz, amplified with MSA-6 amplifiers and stored on a computer.

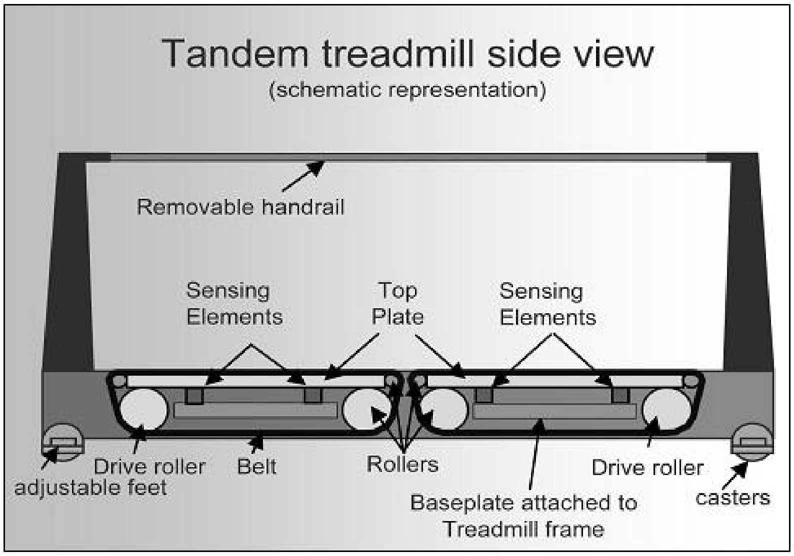


Figure 4. A schematic side view of the AMTI tandem treadmill that will be installed in the SPEAR Facility (*20*).

3.2.2 Force Plate

A force plate will be used to verify the validity of the instrumented pole measurements. A custom 1.22-m square force plate (Advanced Medical Technology, Inc., Newton, MA) that uses strain gauge sensors to measure horizontal and vertical ground reaction forces and moments about the vertical axis will be used in this procedure. The maximum force plate capacity is 8,900 N in the horizontal plane and 17,800 N in the vertical direction. Signals from the force plate will be collected at 600 Hz, amplified and then stored on a personal computer.

3.2.3 Motion Capture System

The Vicon Peak 460 motion capture system will be used to track the positions of reflective markers in three dimensions during testing. The system uses Vicon Workstation software for data collection, management, and analysis and six model M2 cameras with 1.3-megapixel resolution and infrared strobes. For this procedure, motion data will be captured at 60 Hz.

3.2.4 Instrumented Pole

The design of the instrumented pole is based on the evaluation of existing designs and requirements of the calibration protocol (see section 4).

3.2.5 Software

Data will be collected and processed using the Vicon Workstation software. For analysis of the calibration data, software will be written using Vicon Body Builder.

3.3 Protocol Detail

3.3.1 Calibration of the Force Plate with Known Loads

To ensure the accuracy of the measurements, the force plate must first be calibrated using known loads. This will be done by collecting force data while stacking various static weights (25, 50, 75, 100, and 125 kg) on the force plate at 10 locations and measuring the difference between the actual and measured weight (figure 5).

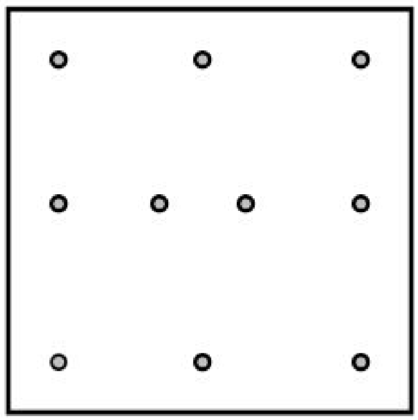


Figure 5. Force place loading locations for static weight calibration and instrumented pole verification.

3.3.2 Instrumented Pole Verification

Before performing tests on the instrumented treadmill force plates, the validity of the instrumented pole force measurements must be confirmed. The following quality checks will be performed using the floor-mounted force plate. First, the motion capture system will be calibrated. While collecting motion capture, force plate, and instrumented pole data concurrently, the instrumented pole will be used to load the force plate with horizontal and vertical force in 10 locations for 5-s trials (see figure 5). At each location, a force, approximately 1,100 N (250 lb), will be applied, beginning with the pole vertical, bringing it to a 45° angle from the floor and moving it in a circle, making a cone shape with the pole (*15*). The force will be applied using the footplate. Collected three-dimensional positions and force data from the instrumented pole will be compared to data from the motion capture and force plate systems and the average error will be determined. This error value will be used to adjust measurements taken by the instrumented pole when calibrating the instrumented treadmill.

3.3.3 Instrumented Treadmill Tests

Natural frequency will be measured by tapping each unloaded force plate with a mallet (approximately 500 N peak force) in the vertical direction while force data is being collected. The natural frequency will be calculated using a fast Fourier transform (FFT) of the force signals using MATLAB (*12, 14*,).

Electrical and mechanical noise will be measured with the belts still and unloaded while the motors are off for 30 s then while the motors are on for 30 s. The power spectral densities of the force signals will be estimated to determine the electrical and mechanical noise. The power spectral density measures how power is distributed over frequency and has units of power^2/frequency. To measure the electrical and mechanical noise with the belts moving, four 30-s trials will be conducted. With unloaded force plates, data will be collected while the belt is moving at 0.45, 1.34, 2.68, 4.02, and 5.36 $m \cdot s^{-1}$ (1.0, 3.0, 6.0, 9.0, and 12.0 mph). For each trial, the power spectral density of the force signals will be estimated using the MATLAB signal processing toolbox (*3*).

Belt speed variation will be measured by fixing several pieces of reflective tape to the edge of each treadmill belt. The motion capture system will collect data for 30-s trials while the treadmill belt is running at 0.45, 1.34, 2.68, 4.02, and 5.36 $m \cdot s^{-1}$. The velocity of the tape pieces will be calculated from deriving the coordinates along the longitudinal axis. The maximum difference between the tape velocity and the belt speed will be calculated for each belt (*12*).

3.3.4 Instrumented Pole Instrumented Treadmill Tests

To measure **linearity** (the force measurement accuracy as weight is incrementally increased), **point of force contact,** and **center of pressure** while the belts are stationary, data will be sampled by the motion capture system, treadmill force plates, and instrumented pole. Force of approximately 1,100 N will be applied at 40 locations for 5-s trials on each instrumented treadmill force plate (figure 6) (*3*). At each location, the pole will begin oriented vertically, then tilted to a 45° angle from the floor and moved in a circle, making a cone shape (*15*). Quantifying the differences between the treadmill force plate and the instrumented pole output will provide information about the accuracy of the instrumented treadmill and allow for error correction. From this data, center of pressure location determined by the instrumented pole and the motion capture system will be calculated and compared to the center of pressure determined from data from the instrumented treadmill.

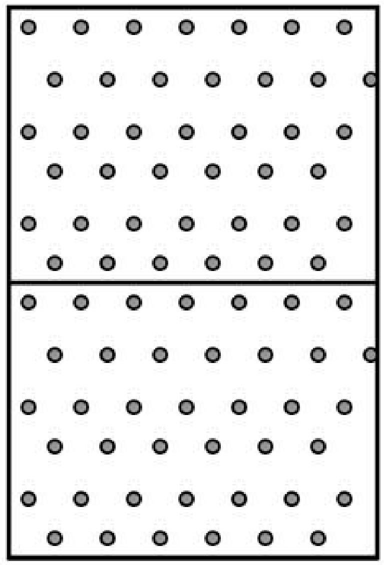


Figure 6. Locations of the instrumented pole force loading during static tests.

To demonstrate the **repeatability of measurements**, a force of approximately 1,100 N will be applied for 5 s, five times at 3-min intervals at three locations on each force plate. Cross talk between vertical and horizontal forces will be calculated by quantifying the horizontal force measured when the pole is applying a vertical force (*3*).

To measure **linearity and center of pressure** while the treadmill belts are moving, data will be collected simultaneously from the motion capture system, treadmill force plates, and instrumented pole. One person will apply a force of at least 1,100 N with varying horizontal and vertical components (*5*). The data output from the instrumented treadmill will be compared to data output from the instrumented pole and motion capture system for linearity, force accuracy, and center of pressure values.

3.3.5 Participant Instrumented Treadmill Tests

Belt speed variation will be measured by fixing several reflective markers to the edge of each treadmill belt. While the motion capture system collects data, a participant will walk or run on the treadmill at 0.45, 1.34, 2.68, and 4.02 $m \cdot s^{-1}$. The velocity of the pieces of tape will be calculated by deriving the coordinates along the longitudinal axis. The maximum difference between the tape velocity and the belt speed will be calculated for each belt (*12*).

Drift is the change in force output during extended use of the treadmill and can be caused by unbalanced sensor heating during trials. Previous studies quantified drift by measuring the forces before, during and after trials of 5 and 30 min at 1.34 $m \cdot s^{-1}$ (2 mph) (*3*, *7*). To thoroughly evaluate the drift, a participant will walk on the instrumented treadmill for 30 min at 1.34 $m \cdot s^{-1}$.

Comparing instrumented treadmill walking to overground walking trials will allow direct comparison of ground reaction forces collected by traditional, floor-mounted force platforms and instrumented treadmill force platforms. The ground reaction force and center of pressure profiles will be compared for 10 strides of overground walking and 10 strides of treadmill walking from the same participant. When evaluating these results, the intrinsic differences between overground and treadmill walking, demonstrated by changes in the joint moments and powers, will be taken into account and resolved with an appropriate computer program (*6*).

4. Results

4.1 Instrumented Pole Selection

A custom-machined pole will be instrumented with five reflective markers and the LC202-3K single-axis load cell (figure 7). The MTD-2 was not selected because the pointed tip makes it difficult to load the pole with forces representative of normal use. The custom pole will have a rubber-coated rounded metal foot at the base. This will protect the treadmill surface and prevent the instrumented pole from slipping during testing. A key requirement of the instrumented pole is the ability to evaluate the accuracy of the instrumented treadmill at loads measured during normal use. The LC202-3K single-axis load cell was selected for its capability to measure these loads. The pole will have a handle and a foot plate located above the load cell that will be used to apply force to the instrumented treadmill. The addition of the foot plate in this design facilitates the application of forces required in the protocol.

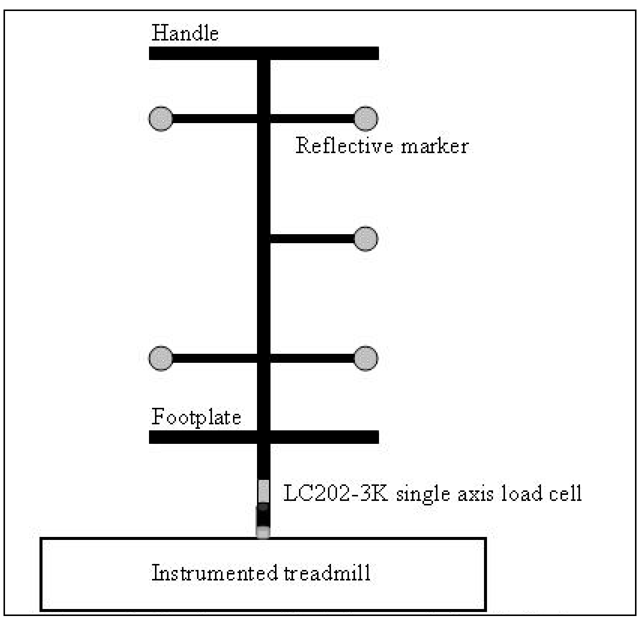


Figure 7. Custom-made instrumented pole that will be used for instrumented treadmill calibration.

4.2 Protocol for Calibration

All motion data will be collected at 60 Hz. Force plate, instrumented pole, and instrumented treadmill data will be sampled at 600 Hz.

4.2.1 Calibration of the Force Plate with Known Loads

The procedure is as follows:

- Record a 30-s trial, beginning with the force plate unloaded. After 5s, place the 25-kg weight on force plate at each location shown in figure 5, leaving it stationary for the rest of the trial.
- Repeat this procedure with 50, 75, 100, and 125 kg weights.

4.2.2 Instrumented Pole Verification

The procedure is as follows:

- Calibrate the motion capture system using standard procedures.
- While recording force plate, instrumented pole, and motion data, use the instrumented pole to load the force plate for 5 s, applying a force of at least 1,100 N with vertical and horizontal components.
- Beginning with the pole vertical, tilt it to a 45° angle with the floor and move it in a circle, making a cone shape with the pole.
- Repeat this procedure ten times, once at each of the locations shown in figure 5.

4.2.3 Instrumented Treadmill Tests

Natural Frequency

The procedure is as follows:

- As force plate data is being collected, strike each unloaded force plate with a mallet (peak force = approximately 500 N) in the vertical direction once every 2 min for 10 min.
- Use MATLAB Fast Fourier Transform (FFT) analysis to calculate natural frequency for all three dimensions.

Electrical and Mechanical Noise

The procedure is as follows:

- Collect force plate data while the belts are still and unloaded and the motors off. Record a 30-s trial.
- Turn the motors on and record another 30-s trial (belts are still and unloaded).

- While the unloaded belts are moving at 0.45 m·s^{-1} collect one 30-s trial.
- Repeat with the belts moving at 1.34, 2.68, 4.02, and 5.36 m·s^{-1}.
- Use MATLAB signal processing toolbox to calculate the power spectral density for each trial.

Belt Speed Variation

The procedure is as follows:

- Fix pieces of reflective tape to the edge of each treadmill belt.
- Recording motion data, collect one 30-s trial while the treadmill belts are running at 0.45 m·s^{-1}.
- Repeat this process while the treadmill belts are running at 1.34, 2.68, 4.02, and 5.36 m·s^{-1}.
- Calculate the marker velocity and maximum difference between belt speed and marker velocity.

4.2.4 Instrumented Pole Instrumented Treadmill Tests

Static Tests

The procedure is as follows:

- Collect motion capture, force plate, and instrumented pole data simultaneously.
- While the treadmill belts are stationary, perform the following test at 40 locations on each force plate, as illustrated in figure 6.
- Using the instrumented pole, apply a force of at least 1,100 N with vertical and horizontal components to one of the instrumented treadmill force plates for 5 s.
- Beginning with the pole vertical, tilt it to a 45° angle with the floor and move it in a circle, making a cone shape with the pole.
- For three locations on each force plate, apply a vertical force of approximately 1,100 N five times for 5 s at 3-min intervals.
- Linearity, point of force contact, center of pressure and repeatability of measurements will be evaluated by comparing instrumented treadmill and instrumented pole results.

Dynamic Tests

The procedure is as follows:

- While the treadmill belts are moving at 1.34 m·s^{-1}, collect motion, force plate, and instrumented pole data.

- One person applies a force greater than 1,100 N with vertical and horizontal components to the treadmill, using the instrumented pole. As the belt moves, the tip of the pole will move towards the end of the treadmill.
- Linearity and center of pressure measurements will be compared for the instrumented treadmill and instrumented pole.

4.2.5 Participant Instrumented Treadmill Tests

Belt Speed Variation

The procedure is as follows:

- Fix pieces of reflective tape to the edge of each treadmill belt.
- Collect one 30-s trial while the treadmill belts are running at 0.45 $m \cdot s^{-1}$ and a participant walks on the treadmill.
- Repeat this process while the treadmill belts are running at 1.34, 2.68, 4.02, and 5.36 $m \cdot s^{-1}$.
- Calculate the tape velocity and maximum difference between belt speed and tape velocity.

Drift

The procedure is as follows:

- Collect force plate data before, during, and after participant trial.
- Participant walks on the treadmill for 30 min at 1.34 $m \cdot s^{-1}$.
- Collect force plate data again 15 min after the completion of the trial.

Comparing Instrumented Treadmill Walking to Overground Walking Trials

The procedure is as follows:

- Collect motion and force plate data while a participant completes 10 strides of overground walking and 10 strides of instrumented treadmill walking.
- Calculate average center of pressure and ground reaction force profiles to compare the two conditions.

5. Discussion

The timelines for the instrumented treadmill ordering and arrival and instrumented pole construction are coordinated so testing will begin early in 2010. The protocol described in this paper will be used for the calibration of the instrumented pole and instrumented treadmill at the SPEAR Facility. Frequent calibration confirms that the data collection systems will properly translate information into the global time-space reference system. This validation and calibration protocol will ensure accurate data are being collected, which will aid in achieving an increased understanding of how Soldier biomechanics and equipment affect physical performance.

6. References

1. LaFiandra, M.; Holt, K. G.; Wagenaar, R. C.; Obusek, J. P. Transverse plane kinetics during treadmill walking with and without a load. *Clinical Biomechanics* **2002***, 17* (2), 116−122.

2. Butler, R. J.; Crowell, H. P.; Davis, I. M. Lower extremity stiffness: implications for performance and injury. *Clinical Biomechanics* **2003***, 18* (6), 511−517.

3. Paolini, G.; Della Croce, U.; Riley, P. O.; Newton, F. K.; Kerrigan, D. C. Testing of a tri-instrumented-treadmill unit for kinetic analysis of locomotion tasks in static and dynamic loading conditions. *Medical Engineering & Physics* **2007***, 29* (3), 404−411.

4. Winter, D. A. *Biomechanics and motor control of human movement,* 3 ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2005.

5. Kram, R.; Griffin, T. M.; Donelan, J. M.; Chang, Y. H. Force treadmill for measuring vertical and horizontal ground reaction forces. *Journal of Applied Physiology* **1998***, 85* (2), 764−769.

6. Lee, S. J.; Hidler, J. Biomechanics of overground vs. treadmill walking in healthy individuals. *Journal of Applied Physiology* **2008***, 104* (3), 747−755. Retrieved Mar, from Article database.

7. Belli, A.; Bui, P.; Berger, A.; Geyssant, A.; Lacour, J. R. A treadmill ergometer for three-dimensional ground reaction forces measurement during walking. *Journal of Biomechanics* **2001***, 34* (1), 105−112.

8. Roerdink, M., Coolen, B., Clairbois, B. E., Lamoth, C. J. C., & Beek, P. J. (2008). Online gait event detection using a large force platform embedded in a treadmill. *Journal of Biomechanics, 41*(12), 2628-2632.

9. Collins, S. H.; Adamczyk, P. G.; Ferris, D. P.; Kuo, A. D. A simple method for calibrating force plates and force treadmills using an instrumented pole. *Gait & Posture* **2009***, 29* (1), 59−64.

10. Verkerke, G. J.; Hof, A. L.; Zijlstra, W.; Ament, W.; Rakhorst, G. Determining the centre of pressure during walking and running using an instrumented treadmill. *Journal of Biomechanics* **2005***, 38* (9), 1881−1885.

11. Dingwell, J. B.; Davis, B. L. A rehabilitation treadmill with software for providing real-time gait analysis and visual feedback. *Journal of Biomechanical Engineering-Transactions of the Asme* **1996***, 118* (2), 253−255.

12. Dierick, F.; Penta, M.; Renaut, D.; Detrembleur, C. A force measuring treadmill in clinical gait analysis. *Gait & Posture* **2004**, *20* (3), 299−303.

13. Macfarlane, P. A.; Looney, M. A. Walkway length determination for steady state walking in young and older adults.(Research Note—Measurement and Evaluation). *Research Quarterly for Exercise and Sport* **2008**.

14. Tesio, L.; Rota, V. Gait analysis on split-belt force treadmills - Validation of an instrument. *American Journal of Physical Medicine & Rehabilitation* **2008**, *87* (7), 515−526.

15. Lewis, A.; Stewart, C.; Postans, N.; Trevelyan, J. Development of an instrumented pole test for use as a gait laboratory quality check. *Gait & Posture* **2007**, *26* (2), 317−322.

16. Holden, J. P.; Selbie, W. S.; Stanhope, S. J. A proposed test to support the clinical movement analysis laboratory accreditation process. *Gait & Posture* **2003**, *17*, 205−213.

17. Rabuffetti, M.; Ferrarin, M.; Benvenuti, F. Spot check of the calibrated force platform location. *Medical & Biological Engineering & Computing* **2001**, *39* (6), 638−643.

18. Motion Lab Systems. *MTD-2 User Guide*. Baton Rouge, LA, 2008 ftp://ftp.emgsrus.com/manuals/mtd2_ug.pdf (accessed July 1, 2009).

19. OMEGA Engineering, Inc. *Miniature Universal Load Cells, Stud Mount Style*, 2009. http://www.omega.com/ppt/pptsc.asp?ref=LC202 (accessed 6 July, 2009).

20. AMTI Force and Motion. AMTI Tandem Force-Sensing Treadmill. *www.amti.biz*. Watertown, MA (accessed 2009).

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Entrainment of Neural Activity in the Cortex Potentially Provides New Information on Individual Differences in Brain Activity: A Proof-of-Concept for Future Neuroscience Research in the Army

by Keith W. Whitaker

Contents

List of Figures

Abstract

This report outlines a new methodology for pre-processing and analyzing electroencephalography (EEG) recordings that uses novel metrics to describe individual differences in brain function. A repetitive background tone entrains cortical neurons while the participant focuses on one of three forced-choice discrimination tasks. Using data from two representative participants, I worked to achieve three aims: (1) identify individual differences in neural processing; (2) perform trial-by-trial analysis of brain activity in response to a repetitive tone for a given individual; and (3) localize cortical areas involved with processing sensory information with patterns of activity that differ within and between participants. I provide proof-of-concept that the first two aims are feasible, but no conclusions can be drawn about the third aim at this time. A novel metric of phase-shifted entrainment can be used to address inter- and intra-individual differences and could be implemented on a trial-by-trial basis. Determination of specific aspects of the EEG recording will require a much greater sample size. Future brain scanning technology may be able to use frequent, repetitive stimuli, such as beeps and tones, to elicit patterns of brain activity to assay mental state without requiring the directed attention of the Soldier.

Acknowledgments

Special thanks to my mentor, Dr. Kaleb McDowell, for his constant help and support. This project would not be possible without Dr. David Hairston's collegial spirit, willingness to share data and patience in teaching me how to obtain and analyze EEG recordings. Thanks to everyone at the U.S. Army Research Laboratory (ARL), Human Research and Engineering Directorate (HRED) (Crewstation) for their helpful comments and criticisms throughout the course of this project.

1. Introduction and Background

Technology is currently being developed to record a Soldier's brain activity in the field (Kerick et al., in prep). It is envisioned that, in real-time or near real-time, commanders will be able to assay individual Soldier's cognitive readiness while the Soldier focuses on soldiering. Fielding equipment capable of measuring a Soldier's cognitive state will require solving a number of technical issues. Once fielded, however, the current metrics of brain activity will not likely be very useful. Current measures have been developed for lab-based studies where the research participant performs one task of interest (i.e., sleeping, Lubenov & Siapas, 2009; controlling a computer, Guger et al., 2009; having a seizure, Franzoni et al., 2009). Future Army technology will require metrics of brain activity that use stimuli that do not distract the attention of the Soldier and provide individualized information in near real-time (McDowell et al., 2009) and do not impinge on the primary goals of the Soldier. Towards this end, the U.S. Army Research Laboratory (ARL), Human Research and Engineering Directorate (HRED) has been pursuing a line of research aimed at identifying observable features of neuronal activity that can be used to assay and define the state of a Soldier's brain.

Although there a number of technologies currently being used to measure brain activity, most require large, immobile scanning equipment. A portable version of acquisition components of the electroencephalogram, or EEG, however, is currently under development (Bloom et al., 2009). Typically, an EEG is a recording of the average electrical activity measured at an array of electrodes in contact with the skull. The technique is widely used in neuroscience research because the equipment required is relatively unobtrusive and non-invasive. There is limited ability to localize scalp-related neural activity to a specific population of neurons, but there is a high degree of temporal resolution (Kandel et al., 2000) that allows for the possibility of real-time data acquisition and processing. Work is being done at a variety of institutions, including the Army (Maxwell, personal correspondence), to be able to understand and interpret specific aspects of EEG activity even in the presence of real-world environmental stimuli and realistic motion (Kerick et al., 2009). In order to take advantage of the real-time recording capabilities of EEG, interference from non-neural sources must be accounted for. For example, muscle contractions, such as eye blinks and limb movements, produce a large amount of electrical activity that is recorded by the electrodes on the scalp and swamps out the neural activity. Although there are experimental methods being developed to actively remove such artifacts (Anderson et al., 2006), trials with identifiable artifacts can be removed from analysis. As the technology develops and matures, there will be a need for new metrics of brain activity that can be used in real-time to compare neural activity of an individual over time as well as compare multiple individuals at a given time.

Towards this end, Hairston and colleagues (2009) are working on a paradigm that elicits cortical and brainstem neuronal responses without requiring the participant's attention. Briefly, a background tone is presented to a participant binaurally, twice a second, while the participant performs a discrimination task. Neurons in the brainstem process this auditory information in a coordinated manner that can be observed in an EEG recording (figure 1). This acoustic brainstem response (ABR) is typically used to test and diagnose hearing and cognitive impairment in infants (Banai et al., 2005), but Hairston and colleagues hypothesize that this response can be modulated by cortical processes associated with directed attention (Hairston et al. 2009). Information about an auditory signal (e.g., a 220 Hz tone) can be observed as neuronal firing following the tone's frequency, at a matching rate of 220 spikes per second, termed the "frequency following response" (FFR). It has been previously shown that the power of the FFR can be enhanced by congruent visual information (Musacchia et al., 2007).

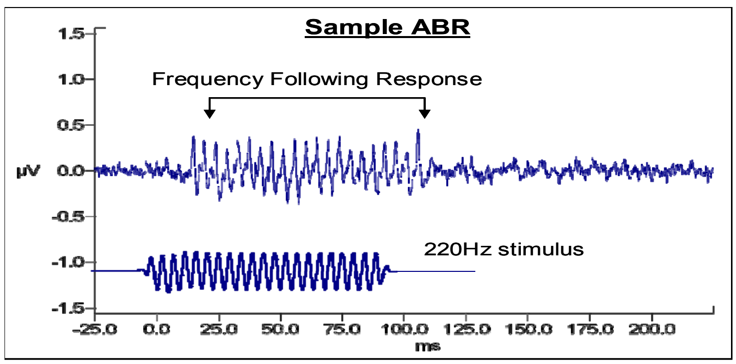


Figure 1. Sample ABR and FFR (from Hairston et al., 2009). Electrical activity of the brain in response to a repetitive background tone is pictured for participant 208. The ABR and FFR components carry the information of the stimulus (220 Hz pure tone).

ABR and FFR, however, are the products of populations of diencephalic neurons with limited roles in cognition. EEG is very sensitive to cortical neurons, which are closest to the recording electrodes. Cortical neurons are known to entrain (synchronize) their activity to the presentation of repetitive sensory cues (Hanslmayr et al., 2007; Model & Zibulevsky, 2006), including an auditory tone (Goldman et al., 2009). Entrained activity may serve different functions in different populations of neurons. In some neural populations, entrained activity may be part of the processes that are actively predicting and compensating for the onset of the next tone while in other groups of neurons, entrained activity could represent the sensation and perception of the previous tone. It is theorized that entrainment is one of the biological mechanisms for focusing attention, perhaps by inhibiting neurons activated by the sensation of each auditory pulse (Banai et al., 2005).

This report describes a methodology that identifies some novel EEG features related to cortical entrainment that may be useful in quantifying individual differences in neural activity. Some individuals, for example, may show entrainment that predicts a stimulus, while other individuals may show a stronger sensation or perception entrainment component. There are three specific aims defined for this project. First, a novel methodology was developed that quantifies individual differences in the latency of neural entrainment compared to the onset of the auditory tone. Second, a cross-correlation analysis for entrainment shows that near-real-time analysis of entrainment is possible on a trial-by-trial basis. Third, I explored features of the complex EEG waveform that may contain additional information on neural activity of cognitive process, such as functional localization to specific areas of the cortex.

2. Experiment and Methods

2.1 Participants and Procedure

Data collection for both subjects was done as part of a separate research project (Hairston et al., 2009). Two participants were selected for analysis in this study based on their comparable age, sex, level of education and experience with EEG recordings. Both were males in their mid-20s with college degrees. The participants had comparable threshold values for the forced-choice discrimination tasks and displayed robust ABRs and FFRs to the background tone.

During EEG recording, the participants were presented with a series of background tones, one every 455 msec. Most of the presentations were a “common tone” (100 ms, 220 Hz sine wave). Randomly, an “oddball” was presented instead (247 Hz). The oddball to common presentation ratio was fixed at 1/32. Participants were instructed to ignore the background tones while performing one of three forced-choice discrimination tasks. Each task was a single-modality task that required the participant to indicate which of two stimuli was longer in duration. The visual discrimination task involved sequentially flashing squares. The second task involved identifying which of two tones in a series had a longer duration. The third task was a control condition; there was no stimulus apart from the background tone. Tasks were presented in two eight minute sessions of 51 trials each yielding 102 trials for each task per participant. More than a thousand common tones were presented during each session. While performing these tasks, participants wore an array of 68 Ag/AgCl electrodes (BioSemi, Amsterdam).

After the data collection, EEG recordings were imported into MATLAB to allow for custom data analyses that is not possible with the commercial software packages used by Hairston et al. (2009) to quantify the ABR and FFR. Custom scripts were written specifically for this project, using functions available by BioSig and EEGLAB (Delorme & Makeig, 2004).

2.2 Tradition Methodologies

Once imported into MATLAB, traditional methodologies of preprocessing EEG data were applied. First, the data was downsampled from 8192 samples per second to 512 samples per second, consistent with analysis of activity of cortical neurons (Supp et al., 2007). After downsampling, two separate bandpass filters were used (figure 2), 0 to 40 Hz and 1.2 to 3.2 Hz. Data from each electrode was visually inspected to ensure the integrity of the data collected. The continuous data file was then epoched; data was broken up into segments around the presentation of each background tone. Each epoch was baseline-corrected and epochs that included non-neural artifacts (< -75mV or >75mV) were rejected. The remaining trials were averaged according to epoch. Although only the results of a single electrode are shown in this report, data from all electrodes is available for future analyses.

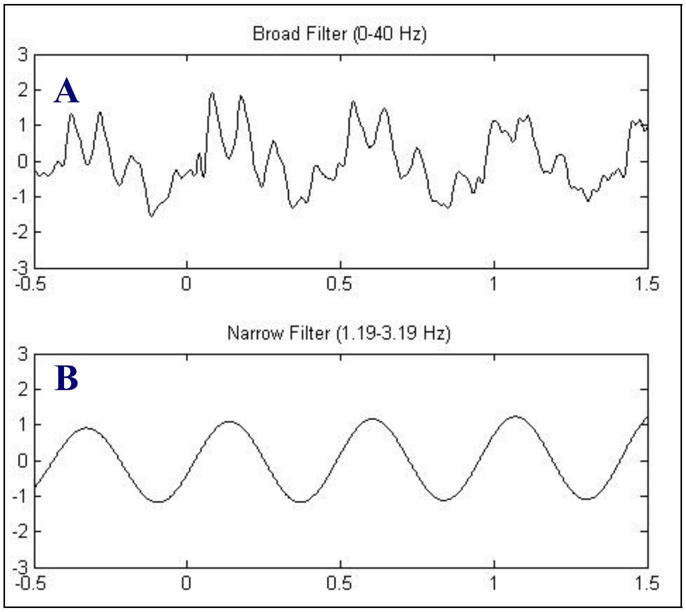


Figure 2. Filtration of EEG recording. Shown are the results of two filters, broad (A, 0−40 Hz) and narrow (B, 1.19−3.19), of the EEG recording from site Fz of participant 208's response to the common tone during the visual task.

2.3 Novel Methodology

I examined epochs of data around each presentation of the common tone in two-second windows that included the presentation of multiple tones (see figure 2b). The two-second window is long enough to display entrainment of neuronal activity to the repetitive tone. Although neural entrainment to a repetitive auditory cute has been observed in other contexts (Will & Berg, 2007; Gao et al., 2009), there are no reports on individual differences in entrainment. The narrow filters of the EEG recordings are set to specifically isolate out cortical entrainment to this tone, if present. The difference between the onset of the auditory tone and the first peak of the EEG recording was defined as the latency of the entrainment. Limited analyses were repeated with the broadly filtered data.

A model of the neural response to the repetitive was built (in radians: $\sin(1.6:0.0269:\pi*8.77+1.6)$) in order to perform a cross-correlation analysis on each trial. This model is a sine wave function where peaks correspond to the onset of the background tone. This methodology was adapted from the phase-coupling analysis of (Hanslmayr et al., 2007).

3. Results and Discussion

3.1 Aim 1: Individual Differences in the Entrainment of Cortical Neural Activity

The average entrainment of neural activity to the background tone provides new information about individual differences. The entrainment of cortical activity to a repetitive tone is a robust, easily visualized phenomena. The latency shift between the tone's onset and peak amplitude of neural activity is a quantifiable difference in brain activity that has not been reported previously.

The latency of neural entrainment recorded at the frontal cortex varies between individuals in a task-specific manner. As shown in figure 3, EEG recordings at site Fz above the frontal cortex of participant 208 (bold line) indicate a strong entrainment to the common tone with a reliable latency. The entrainment latency of the other individual (205, grey line) varies during the visual task.

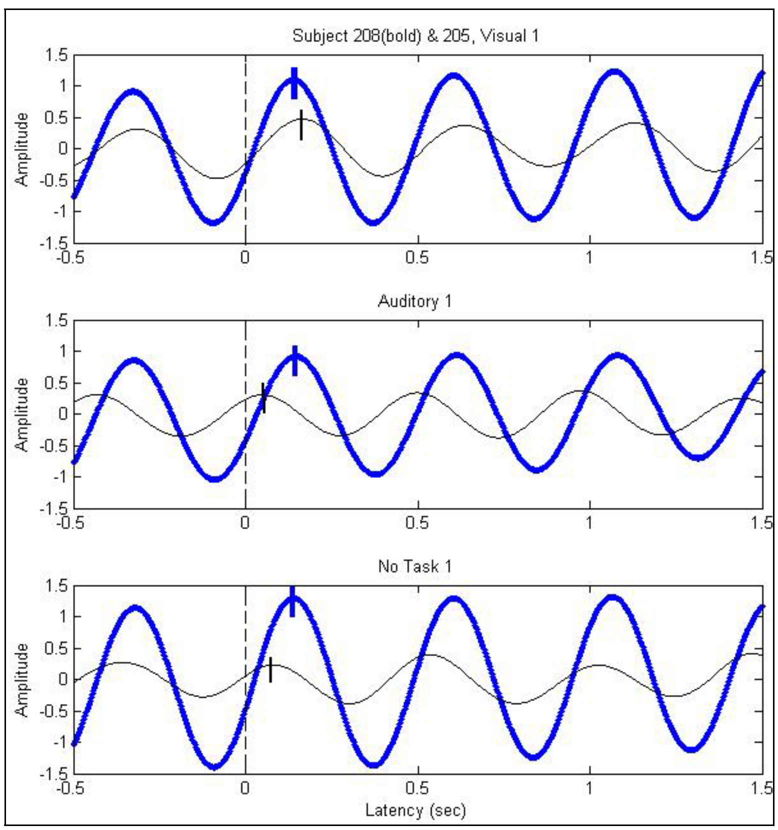


Figure 3. Individual average waveforms recorded at electrode Fz for each task. The narrower bandpass filter shows the general entrainment of electrical neural activity recorded at the front, center of the head, while the background tone is being presented every 0.5 s. The difference in amplitude between participants is the result of non-neural factors (e.g., skull thickness, intracranial space, etc.). The entrainment of each participant appears to phase-shift, which may be a novel metric for individualized trial-by-trial analysis of EEG recordings.

3.2 Aim 2: Feasibility of Near Real-time Snalysis of Cortical Entrainment

In order to quantify differences on single trials, I modeled the auditory stimulus as a sine wave function. Comparing this model to each trial creates a standard measurement to quantify differences in cortical entrainment on each trial. The model is compared against the recording for each trial at electrode Fz. In order to determine the latency the phase-shift of the entrainment, the metric identified in aim 1, separate correlation coefficients are calculated as the model is phase-shifted against each recording in a stepwise manner. The value of the maximum correlation coefficient and corresponding phase-shift for each trial is calculated for each trial. The vast majority of trials were expected to be uncorrelated due to muscle artifacts (eye blinks, limb movements, etc.) and extraneous mental processes (daydreaming, completion of the primary task, etc.) which involve electrical activity unrelated to the neuronal processing of the common tone. Therefore, an arbitrary cut-off for the correlation coefficient was set at 0.7 for this study.

Figure 4 shows the correlation results for participants 205 (left) and 208 (right) for each primary task (from top to bottom: No task, Visual & Auditory). The top histogram is composed of trials with max correlations under 0.7 (blue) and the lower histogram is created from trials with correlations greater than 0.7 (red). The phase-shift that was applied to obtain the max correlation is plotted on the x axis. The bottom figure in each set compares the average waveform of trials with correlation coefficients less than 0.7 (blue dotted) to trials with correlation coefficients greater than 0.7 (red line).

Participant 208 demonstrates a highly stable entrainment, with a large number of trials showing a high correlation when the model is phase-shifted by about 90 frames (about 180 msec). When the participant is doing an auditory task, however, there may be a shift towards a smaller phase-shift. Participant 205 also shows a large number of trials correlating to model, but there is no specific phase-shift that can be applied to fit a large portion of these trials. The average waveforms for each individual also reflect this difference in trial-by-trial stability.

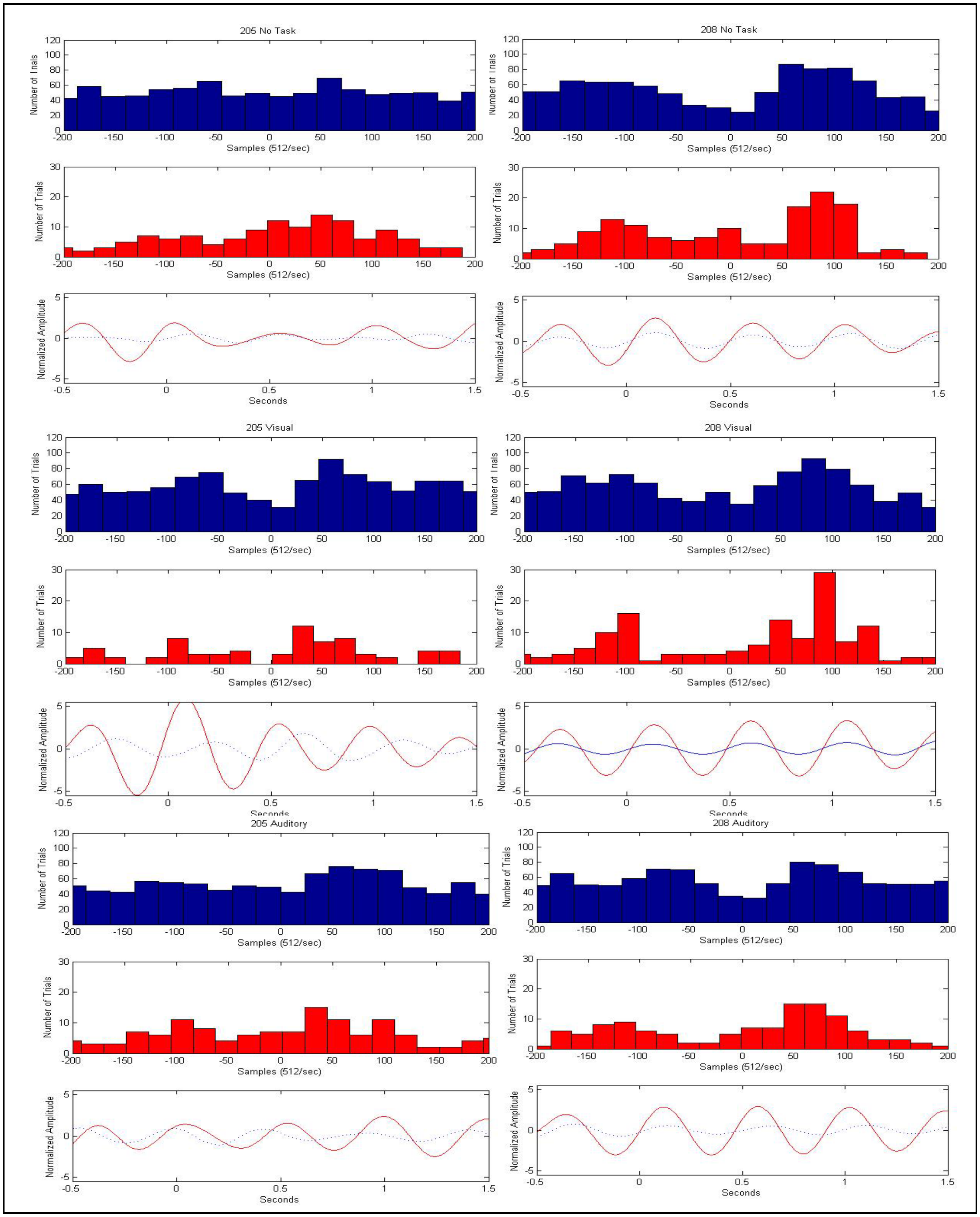


Figure 4. Correlations, latencies and waveforms of phase-shifted entrainment. The results of the trial-by-trial correlations with a model entrainment are presented here. Individual trials with correlations of less than (top, blue) and greater than (middle, red) 0.7 are shown for each of the three primary tasks (Left: participant 205; Right: 208). The bottom figure for each session shows the average waveform for the trials with correlations under 0.7 (blue) and greater than 0.7 (red). Note that inter-task differences in the phase-shift are still only evident in one participant.

3.3 Aim 3: Identification and Localization of Waveform Components

As seen in figure 2, the narrow filters applied to the data remove most of the complexity of the EEG recording. Applying broad filters reveals a number of features that may contain additional information about the neuronal activity of individual participants while they perform a given task. In figure 5, complex waveforms recorded at electrode Fz from each subject shows a number of common features (colored arrows). Although intriguing, two individuals do not provide enough information to develop methodologies or to test hypotheses aimed at understanding the underlying neurobiological and cognitive basis of the specific features of the complex waveforms.

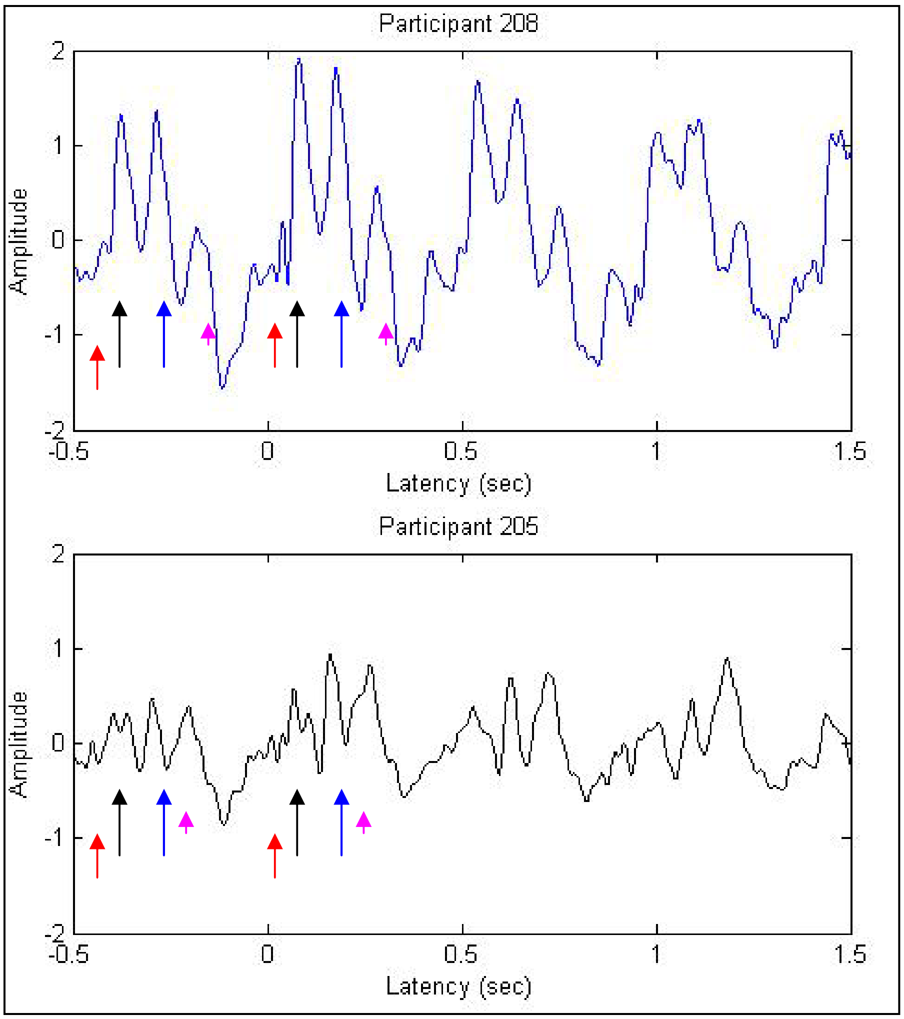


Figure 5. Average waveforms of broadly filtered data reveal details. Relaxing the filters on the raw data before preprocessing reveals the complexity of an averaged EEG waveform, which is a composite of the all neural activity observable by that electrode (here, Fz). For both participants (208: top; 205: bottom), specific features (colored arrows) may be identifiable and phase-locked to the onset of the common tone.

4. Summary and Conclusions

The entrainment of neural activity in the cortex varies between individuals (figure 6). The changes in latency of entrainment may be useful to future Army applications as a metric for individual differences in neural processing of sensory information. Although the biological and cognitive basis of this shift is unknown, the findings reported here will support the development and testing of new hypotheses about how individuals process sensory stimuli. It may now be possible, for instance, to differentiate between individuals predisposed to predicting the next tone from individuals whose neural activity is more heavily influenced by the sensation of the previous tone.

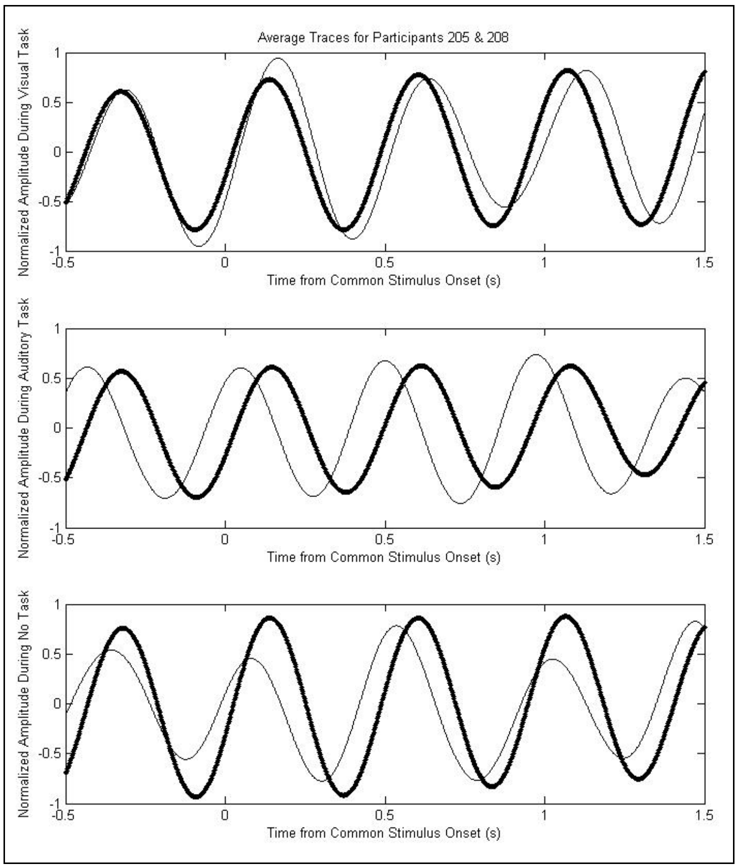


Figure 6. Individual differences in entrainment to auditory tones. Participant 208 (bold) demonstrates entrainment that is task-independent, unlike participant 205 (light). Participant 205's entrainment is phase-shifted when multiple information streams are being processed through a single modality.

In one of the individuals presented here, entrainment was stable across all tasks. In the other, the entrained neurons phase-shifted their activity when the participant was focused on particular tasks. I propose that the latency of the phase-shift could be used as a novel metric for quantifying individual differences in neuronal activity.

Within an individual, cortical entrainment can be quantified on a trial-by-trial basis. Here, I used a sine wave model as a basis for quantifying the amount of phase-shift on each trial. Trials that correlate well with this model can be used to test and generate future hypotheses about the underlying neuronal activation that cannot be addressed with the current metrics. The shift in neuronal entrainment is the result of populations of neurons changing how they process the auditory tone. Although this study was too limited in scope to address the relevance of these changes in neural activation, it will be addressed in future studies. With field-deployable EEG recording equipment currently under development, the entrainment of cortical neurons may provide information that can be used in near real-time to reflect the cognitive state of Soldiers in the field.

The results presented in this study focused on EEG recordings filtered to approximate a sinusoidal waveform. Additional information may be obtained from EEG recording filtered between 0−40Hz. This broad filtration reveals more detailed waveforms. The third aim of this proposal attempted to identify specific components of this waveform, but there is not enough data to draw any conclusions. The peaks noted in figure 5 may be the result of different neuronal populations with different phase-shifted entrainment, for example. This theory could be tested by studying the timing and power of each peak at different electrode sites across the skull. Activity of different populations of neurons can be localized. It is theoretically plausible that the peaks denote different cognitive processes, such as prediction of the next tone or perception of the previous stimulus. Or, these features may be an artifact of these two participants. It would be useful to include additional participants in order to determine the reliability and variation of the features across individuals.

The methodology presented in this report is a novel finding with potential utility for future Army applications. The entrainment of cortical neurons can be easily observed and quantified between individuals. The phase-shift of the entrainment can be quantified on a trial-by-trial basis, which may allow for real-time or near real-time analysis of a Soldier's cognitive state while he or she focuses on soldiering. Studying the more complex waveform of broadly filtered EEG recordings may reveal new sources of information regarding neural activity in the future.

5. Summary of Personal Benefits

Working on this project gave me experience working with human research participants. My research background into the biological basis of behavior has, up until this point, been focused on the use of animal models of behavior. I found that I was able to apply my knowledge about the cellular and molecular basis of behavior and individual differences to develop new and potentially beneficial lines of research for the Army.

I further developed my programming skills by working extensively with MATLAB. I use this software to analyze neurobiological and behavioral data for my doctoral thesis. Further development of this skill set will allow me to be much more efficient with future experiments both here and at The University of Texas at Austin.

6. References

Anderson, C. W.; Knight, J. N.; O'Connor, T.; Kirby, M. J.; Sokolov, A. Geometric Subspace Methods and Time-delay Embedding for EEG Artifact Removal and Classification. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on.* **2002**, *14* (2), 142–146.

Banai, K.; Nicol, T.; Zecker, S. G.; Kraus, N. Brainstem Timing: Implications for Cortical Processing and Literacy. *J Neuro.* **2005**, *25* (43), 9850–9857.

Bloom, F. A. et al. *Committee on Opportunities in Neuroscience for Future Army Applications; National Research Council*; Opportunities in Neuroscience for Future Army Applications. National Academies Press. 2009.

Delorme, A.; Makeig, S. EEGLAB: an Open Source Toolbox for Analysis of Single-trial EEG Dynamics Including Independent Component Analysis. *J Neurosci Methods* **2004**, *134*, 9–21.

Franzoni, E.; Gentile, V.; Pellicciari, A. et al. Prospective Study on Long-term Treatment with Oxcarbazepine in Pediatric Epilepsy. *J Neurol.* **20 July 2009**. [Epub ahead of print]. DOI 10.1007/s00415-009-5157-8, http://www.ncbi.nlm.nih.gov/pubmed/19597919?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum.

Gao, L.; Xiankai Meng, X.; Ye, C. et al. Entrainment of Slow Oscillations of Auditory Thalamic Neurons by Repetitive Sound Stimuli. *J Neuro.* **2009**, *29* (18), 6013–6021.

Goldman, R. I.; Wei, C.; Philiastides, M. G. et al. Single-trial Discrimination for Integrating Simultaneous EEG and fMRI: Dentifying Cortical Areas Contributing to Trial-to-Trial Variability in the Auditory Oddball Task. *NeuroImage.* **2009**, *27*, 136–147.

Guger C.; Daban S.; Sellers E. et al. How Many People are Able to Control a P300-based Brain-computer Interface (BCI)? *Neurosci Lett.* **2009**, *462* (1), 94–98, http://www.ncbi.nlm.nih.gov/pubmed/19545601?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum

Hairston, D. W.; Letowski, T. R.; McDowell, K. Examining Cross-modal Influence of the Auditory Brainstem Response. Poster, 2009.

Hanslmayr, S.; Aslan, A.; Staudigl, T.; Klimesch, W.; Herrmann, C. S.; Bäuml, K-H. Prestimulus Oscillations Predict Visual Perception Performance Between and Within Subjects. *NeuroImage* **2007**, *37*, 1465–1473.

Kandel, E. R.; Schwartz, J. H.; Jessell, T. M. *Principles of Neural Science*; 4th ed. McGraw-Hill, New York, 2000.

Kerick, S. E.; Oie K. S.; McDowell, K. *Assesment of EEG Signal Quality in Motion Environments*; ARL-TN-0355; U.S. Army Research Laboratory: Aberdeen Proving Groumd, MD, 2009.

Lubenov, E. V.; Siapas, A. G. Hippocampal Theta Oscillations are Travelling Waves. *Nature*. **2009**, *459* (7246), 534–539.

McDowell, K.; LaFiandra, M. Army Research Labs. Strategic Technology Area Neuroscience.

Model, D.; Zibulevsky, M. Learning Subject-specific Spatial and Temporal Filters for Single-trial EEG Classification. *NeuroImage.* **2006**, *32*, 1631–1641.

Musacchia, G.; Sams, M.; Skoe, E.; Kraus, N. Musicians Have Enhanced Subcortical Auditory and Audiovisual Processing of Speech and Music. *PNAS* **2007**, *104*, 15894–15898.

Sayers, B. M.; Beagley, H. A.; Henshall W.R. The Mechanism of Auditory Evoked EEG Responses. *Nature*., *247*, 481–483.

Supp, G. G.; Schlögl, A.; Trujillo-Barreto, N.; Müller, M. M.; Gruber, T. Directed Cortical Information Flow During Human Object Recognition: Analyzing Induced EEG Gamma-band Responses in Brain's Source Space. *PLoS One.* **2007**, *2* (1), e684, http://www.ncbi.nlm.nih.gov/pubmed/17668062?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum

Will, U.; Berg, E. Brain Wave Synchronization and Entrainment to Periodic Acoustic Stimuli. *Neurosci Let.* **2007**, *424*, 55–60.

Sensors and Electron Devices Directorate (SEDD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory
SUMMER RESEARCH TECHNICAL REPORT

Transport Phenomena in the Alkaline Anion Exchange Membrane Fuel Cell

by Kyle N. Grew

Contents

List of Figures

List of Tables

Abstract

The U.S. Army depends upon its Soldiers' technological advantage to ensure mission effectiveness. Sustaining this advantage requires the development of safe and reliable power sources for the technologies they depend upon (e.g., sensors, communications, and surveillance). Proton exchange membrane (PEM) based direct methanol fuel cells (DMFCs) have been the primary focus for Soldier portable power; however, these systems require expensive noble metal catalysts, which are subject to degradation and crossover depolarization. An alkaline alternative, known for its favorable kinetics with non-noble metal catalysts, could serve as an enabling technology. This paper takes a computational approach to examine recent demonstrations of alkaline anion exchange membranes in fuel cell (AEMFC) applications. These demonstrations have shown substantial resistive losses and activation losses. A framework for understanding the opportunities, limitations, and technical barriers associated with AEMFCs is being developed. This approach focuses on the transport phenomena to elucidate the underpinnings of the conductive transport, the role of water transport, and their coupled effect on activation processes. The studies are validated with published experimental data and comparisons to the PEM literature.

Acknowledgments

The author wishes to acknowledge the mentorship of Deryn Chu.

1. Background and Introduction

Soldiers in the U.S. Army have growing power requirements to meet mission needs. While these growing demands are supplied by the advanced technologies that provide them a strategic advantage, sufficient fuel or batteries are required to power these devices. In a 2004 review, the Communications-Electronics Research, Development, and Engineering Center's (CERDEC) Patil et al. detail the U.S. Army's research and development efforts that have focused on the development of direct methanol fuel cell (DMFC) as near-term battery replacements in solider portable power and sensor applications (approximately 0−100 W) (*1*). Aqueous methanol's reasonably high specific energy (6.1 kWh/kg), volumetric energy density (4.8×10^3 kWh/m^3), and ease of storage make it suitable for these applications (*2*).

DMFCs typically employ a proton exchange membrane (PEM), because of the relative maturity of the technology (*1–4*). However, PEMs require the use of expensive noble metal derived catalysts to support the oxygen reduction reaction (ORR). Additional catalysts loadings are often used to support the direct and complete six-electron methanol oxidation reaction (MOR) and mitigate depolarization resulting from fuel crossover (*1–4*).

This report examines technical challenges related to the use of an alkaline-anion exchange membrane fuel cell (AEMFC) for near solider portable power applications. The AEMFC is examined for use as a direct alcohol fuel cell (DAFC), where the anion exchange membrane (AEM) is used to provide stability, hydroxyl ion (OH^-) conductivity/transport, and an alkaline environment. The alkaline environment supported by the AEMFCs is favorable for the ORR (compared to the acidic environment in PEMs). By moving to an alkaline environment, the substantial ORR activation losses can be reduced while also providing opportunities for more cost effective and practical non-noble metal catalysts (*2*). Because ionic transport opposes fuel crossover, opportunities for limiting depolarization losses may be available, along with simplified management schemes (*2, 5–6*). These types of opportunities have been limited in previous generations of alkaline fuel cells (AFCs), where aqueous metal-hydroxide electrolytes (e.g., KOH) and electrode architectures presented degradation issues through the precipitation of metal-carbonates in the presence of carbon dioxide (CO_2) (*2*). However, a variety of metal-hydroxide free polymer-based AEMs have recently been reported, which provide sufficient conductivity and stability characteristics (*2*). By removing the aqueous metal-hydroxides from the electrolyte, metal-carbonate precipitation should be negated during operation with CO_2 and other carbon-based species in the fuel and oxidant (*2, 5–6*).

This paper focuses on the two types of AEM: (1) an aminated and alkalized vinyl(benzyl chloride) radiation grafted fluoropolymer (FEP and ETFE), as has been demonstrated by Varcoe, Slade et al. (*2, 5–7*), and (2) a quaternized and alkalized polysulfone (PS), or aromatic polymer-based membrane, formed through a direct chloromethylation process, as has been demonstrated by Hibbs et al. as well as Yanagi and Fukuta using Tokuyama Co. membranes, among others (*8–9*).

In lieu of a complete discussion of these membranes and their synthesis, tables 1 and 2 compare several aspects and figures of merits for these AEMs, along with some details regarding common PEMs. Clarifications of the reported parameters are noted with additional clarifications, definitions, and nomenclature in an Appendix. More complete details as to the measurement methods and membrane architectures are available in the provided references.

Table 1. Comparison of anion (AEM) and cation (PEM) exchange membranes.[a]

Features	AEM	PEM
Backbones	Fluoropolymers: ETFE, FEP Aromatic Polymers: PS, PE	Per-fluorinated polymers: Nafion Fluoropolymers: PTFE, ETFE, FEP, … Aromatic Polymers: PS, PE, Copolymers
Functional groups	Quaternary Ammoniums: $-CH_2N^+(CH_3)_3$	Sulfonic Acids: $-SO_3-$
Conducted ion(s) [b]	OH^- *(HCO_3^-, $CO_3^=$, Cl^-)*	H^+ *(Na^+, K^+, Li^+)*
Advantages	Non-noble metal catalysts Oxygen reduction reaction Cost / Packaging Low fuel permeability	Ionic conductivity Ionomer solutions Chemical stability Mature technology
Issues and Questions	Ionic conductivity H_2O Transport MEA interface Influence of CO_2 Long-term stability	Materials Costs Fuel crossover and depolarization Oxygen reduction activation losses Ostwald ripening and catalyst migration

[a]Figure adapted from [9].
[b]Alternative ionic forms provided in parentheses

Table 2. Properties and performance metrics of anion and proton exchange membranes.

	AEM		PEM		
Property [d]	***Radiation Grafted: ETFE, FEP***	***PS and Aromatic Polymers***	***Nafion***	***Radiation Grafted: ETFE, FEP***	***PS and Aromatic Polymers***
σ (aq.) [b,c] σ (~100% RH): [mS/cm]	15−30 8−15	<35 42	~100 80	190−300	<140
IEC: [mmol-eq./g]	< 2.25	0.7−1.9, 1.7	0.93	2−3	1.0−2.2
λ: [b] [H_2O/funct. group]	21−24 19−21	< 100 wt.%	24 17	22−28	< 140 wt.%
ρ_o: [g/cm^3]	1.5−2.0	X	2.0	1.5	X
ΔH_m: [d] [J/g]	?	5-130	100	?	1−130
References	[2,5−7,10]	[8−9]	[7,10]	[11−12]	[8]

[a] Symbols and description of properties provided in Appendix.
[b] Reported in near ambient conditions (i.e. 25−30 ˚C)
[c] Radiation grafted AEMs reported at IEC ≈ 1.0.
[d] Provided at limits of reported IEC values

In table 1, some of the tradeoffs between the AEM and PEM for fuel cell applications are provided. Specific figures of merit for a number of these properties can be found in table 2. The practical advantages that the AEM may provide for portable power merit further investigation. Specific issues regarding the AEMFC are the low ionic conductivity of the AEM and large activation losses associated with the membrane electrode assembly (MEA) interfaces need to be addressed. The latter is primarily a result of needing improved soluble forms of the AEM for dispersion in the catalyst layer architectures. Not captured in the provided tables, un-optimized demonstrations of AEMFCs have shown considerable losses that are characteristic of resistive losses. Other authors have attributed these losses to mass transport effects and suggest that drying in the AEMFC cathode, where water (H_2O) takes part in the ORR, may be liming present systems (*6–7, 13*). Similar processes, which are very acidic, have been discussed in the PEM literature (*14*). This lack of understanding highlights the need for predictive capabilities to provide insight into the nature of the factors limiting present AEMFCs, as well as opportunities for overcoming these barriers.

In this report, several aspects of the AEM and AEMFC are examined using an analytic and computational approach. The first aspect examined the nature of the physiochemical interactions of the functional group (e.g., benzyltrimethylammonium-hydroxide [BzTMAH]) with the solvent solution. Specifically, the dissociation constant and solvation numbers associated with these groups. This analysis is used to demonstrate a capability for predicting the ionic conductivity of the AEM. Using this analysis, a one-dimensional (1-D) water transport and polarization model is presented. The model is developed using a finite volume method (FVM). While this transport model and the demonstrated results are both preliminary and empirical, considerable insights can be gained from this type of approach. Complete details regarding the morphology, properties, and interactions of present AEMs are not yet available or fixed. At this time, the focus is on the thermodynamic regime prior to the onset of liquid saturation. Ongoing efforts will consider more detailed and rigorous descriptions of the cell, the effects of absorption/desorption, multiphase transport, multidimensional effects, and more complete details of the AEM and AEMFC as they become available or are determined using more sophisticated and rigorous methods.

2. Method of Approach

2.1 Membrane Ionic Transport: Generalized Stefan Maxwell's Membrane Model (GSM)

To provide sufficient details about the approach considered in this study, a brief discussion regarding a previously developed a numeric/analytic framework that is used in this study is necessary. These details are just highlighted, as they have been previously reported (*15*). This model considers a development that employs the general use of the Stefan-Maxwell transport equations for non-ideal solutions (*16*). If the structure plays a prominent role in the transport processes due to its physiochemical interactions with the mobile species (as within polymeric membrane transport processes), the structure can be treated as a discrete species. Using thermodynamic closure (Gibbs-Duhem) along with some detailed development, it can be shown for n-mobile species,

$$-\frac{C_i}{RT}\left(\nabla_T \mu_i - \overrightarrow{F_i}\right) = \sum_{\substack{j=1 \\ i \neq j}}^{n} \frac{C_j \overrightarrow{N_i} - C_i \overrightarrow{N_j}}{C_T D_{ij}^{eff}} + \frac{\overrightarrow{N_i}}{D_{iM}^{eff}} + \frac{B_o C_i}{\nu/\rho \, D_{iM}^{eff}}\left(\nabla P - C_T \vec{F}\right) \tag{1}$$

where n-equations must be written for the n-mobile species. In equation 1, C_i is the species concentration, μ_i species chemical potential, F force vector (electrostatic potential), N_i molar flux, D_{ij} mutual diffusivity, D_{iM} reduced diffusivity due to structure, B_o permeability, ν viscosity, and P pressure. The left-hand side of equation 1 represents terms driving transport, while those on the right represent diffusion/migration and convective processes (e.g., electro-osmotic drag and permeation). Applying equation 1 to the conditions of a closed electrochemical cell for electrochemical impedance spectroscopy (EIS) analysis (i.e., fixed temperature, pressure, and humidity, with equi-molar counter diffusion between the ionic species and water), it can be shown,

$$\sigma = \left(\varepsilon - \varepsilon_o\right)^q \left(\frac{D_{ij}}{1+\delta}\right) \frac{F^2}{RT} \left|z_i\right| C_i \tag{2}$$

where σ is the ionic conductivity. Arriving at equation 2 required the consideration of the ionic flux and Ohm's law in conjunction with equation 1. Key parameters include the mutual diffusion coefficient, D_{ij}, for ionic-H_2O transport rates; an empirical parameter, δ, which accounts for the ratio of the mutual diffusion coefficient (theoretical) and the reduced diffusivity due to the membrane; and pore volume fraction, ε, and its value at the percolation limit, ε_o. Finally a Bruggeman exponent, q, is used to account for the tortuosity factor and the magnitude of the ionic charge number, z_i, can be incorporated for multi-valence ionic species, but is not necessary for this study.

In equation 2, C_i represents the concentration of ionic charge carriers in the pore-regions of the membrane that can participate in conduction/charge transfer process. During initial studies, it was treated using the fixed charge concentration, related to the membrane's ion exchange capacity. Changes in ionic concentration were treated using the swelling and water content in the membrane. These preliminary results are shown in figure 1 for the FEP-AEM, ETFE-AEM and Naifon 115, where the model is compared to experimental data (symbols) from Varcoe (*7*). Specifically, figure 1 compares the ionic conductivity to the water uptake coefficient, which was treated using a Brunauer, Emmet, Teller (BET) multilayer absorption isotherm. While not shown here, the BET parameters and uptake characteristics are quite comparable for the AEMs and Nafion 115 (*15*).

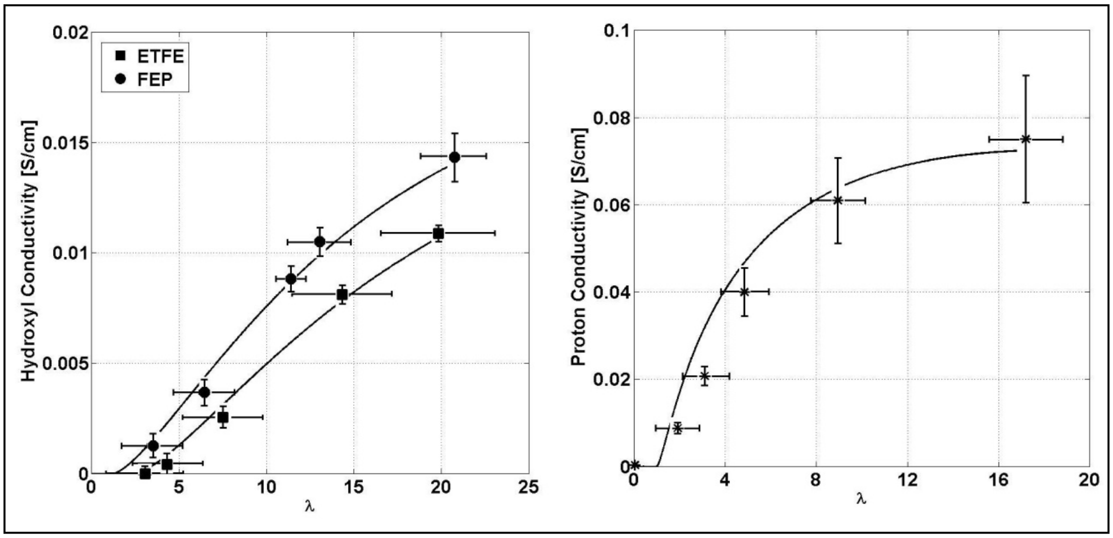


Figure 1. Conductivity prediction for AEM (left) and Nafion (right) using a generalized Stefan Maxwell's equation development. Ionic conductivity is presented versus the water uptake coefficient, λ. Experimental data (symbols) from Varcoe (*7*) and model (lines) presented in (*15*).

However, additional empiricism was necessary to capture the details of the AEM conductivity for the studies provided in figure 1. Further, this empiricism was not required in the Nafion model and followed a dependence on the pore volume fraction of the AEM. This additional empiricism substantially decreased the AEM conductivity model to match the experimental data (nearly an order of magnitude). The pore volume fraction, membrane swelling, and water uptake were also considered and replicated during these initial studies with considerable similarity to Nafion and other PEMs from a macroscopic standpoint. This implies that the additional empiricism required for the AEM is a direct result of its morphology, species, and physiochemical interactions with the water solvent.

2.2 Dissociation and Solvation Models

The additional empiricism that was necessary in these initial studies had a dependence on the pore volume fraction. It is hypothesized that the discrepancy is a result of the physiochemical equilibrium in the AEM. A key approximation made in the initial development was that the functional groups were completely "dissociated." This requires that all OH^- ions be freely available to participate in ionic transport processes. This may not be a bad approximation for the highly acidic Nafion, but it may not be a good approximation for the BzTMAH functional groups used in most present AEMs. Certainly, the concentration of *available* charge carriers scales the resistivity in all sorts of electrochemical systems. Generally speaking, this applies to everything from metallic conductors (e.g., electrons in the conduction bands) to solid state ionics (e.g., vacancy/defects) and traditional electrolytes (e.g., other ions available) (*17*). The AEM conductivity will suffer if the OH^- ions are not dissociated.

A conceptual schematic of how the processes in the AEM may look is presented in figure 2. In this figure, the BzTMAH functional groups tethered to the polymer backbone are observed. As the membrane is hydrated, the polar solvent water coordinates with these hydrophilic groups. With sufficient hydration, the OH^- can ionize, or dissociate on an equilibrium basis, to participate as a charge carrier. These carriers supplement those donated from the electrochemical processes. A solvation shell surrounds any OH- ions in solution, along with the fixed functional groups. These rates and nature of these processes are addressed in this study.

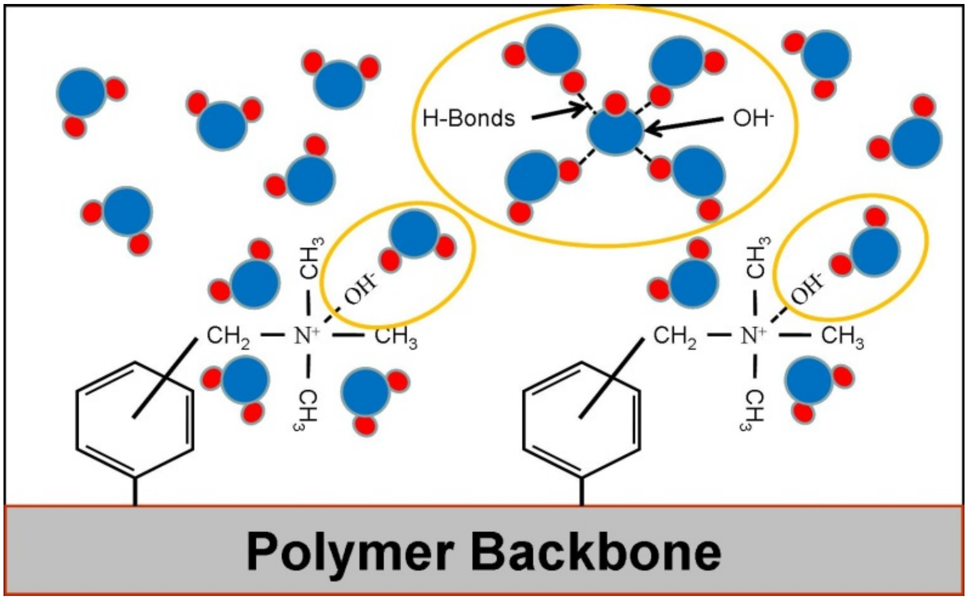


Figure 2. Conceptual representation of the solvation and dissociation in the AEM. The functional groups are solvated with H_2O. Based upon solvent loading, a dissociation constant, K_b, describes the fraction of functional groups dissociated.

To address the physiochemical interactions of the AEM's BzTMAH functional groups with the water solvent, a solvation and dissociation model is developed using analogy to experimental data for an aqueous solution (*18–20*). Similar approaches have been previously demonstrated in the PEM literature, where Nafion is suggested to follow an acidity function comparable to sulfuric acid (*18–22*). PS and aromatic PEMs have also been considered using comparable approaches (*22*). Following the development and validation of the dissociation and solvation model, it is used in the previously discussed conductivity model. Again, experimental AEM and Nafion conductivity data from the electrochemical impedance cell is considered and discussed.

2.3 AEMFC Transport and Polarization Model

With an improved understanding of the AEM, a second part of the transport studies can be considered. The initial framework for a cell-level transport and polarization model is presented. While preliminary, this model is used to provide insight into several discussions regarding AEM water management in the literature (*2, 5–7, 9–10*). It can also provide valuable insight as to more specific experimental characterizations and studies that are needed.

In this study, a 1-D finite volume method (FVM) water transport and polarization model is considered in the AEM. There is very limited experimental data regarding the membrane diffusivities, sorption rates, electro-osmotic drag coefficients, and permeability, so some parameterization is necessary. Properties from the literature and independent analysis are used as appropriate. Because this study is preliminary, multi-phase and multi-dimensional effects are not considered. Transport in the AEM considers diffusion and osmotic drag, as the cathodic ORR consumes H_2O under Faradaic conditions and it is produced in the anode HOR. The electrodes are treated as effective resistances, and polarizations are determined using the H_2O transport study results.

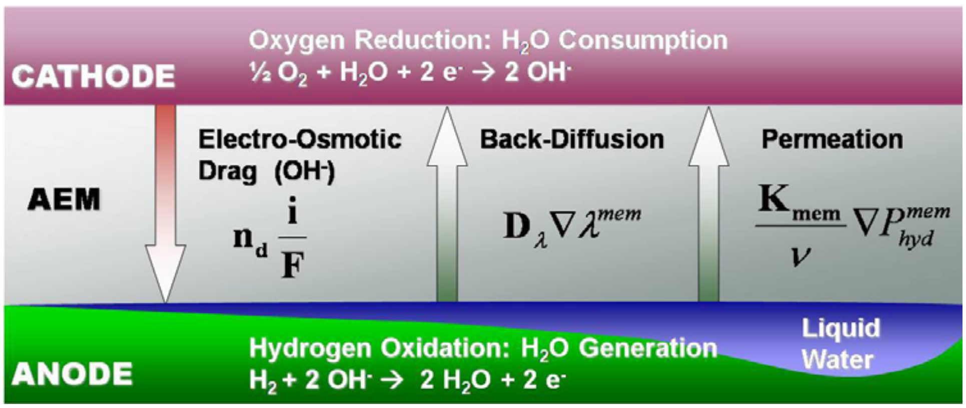


Figure 3. Conceptual schematic of H_2O transport in AEM for an H_2/Air AEMFC. The stoichiometric consumption and production of H_2O associated with the ORR and HOR, respectively, are noted.

3. Calculations and Validation

3.1 Membrane Dissociation and Solvation Model

The acidity functions for a strong acid (Nafion) and base (AEM) are developed using an analogy to solutions with comparable acid and base groups. Experimental data for solutions containing different concentrations of sulfuric acid and BzTMAH is used. The approach is consistent for both, so this discussion will focus on the BzTMAH. An arbitrary acid or base can dissociate (equation 3) and solvate (equation 4),

$$HA \leftrightarrow H^+ + A^- \tag{3}$$

$$H^+ + n \cdot BH \leftrightarrow H^+ \cdot (BH)_n \text{ or } A^- + n \cdot BH \leftrightarrow A^- \cdot (BH)_n \tag{4}$$

where HA is the considered the acid or base, BH a polar solvent, and 'n' a hydration number (solvating) for the H^+ or OH^-. A Hammett function for a strongly basic system is used to determine an equilibrium dissociation constant of an acid or base in water (*23*). For strong bases, this acidity function, H_-, is defined as the negative logarithm of its ionization constant, pK_a, where $pK_b = pK_w - pK_a$. As the ionization ratio of solution approaches unity (e.g., dilute limit), the acidity function approaches the pH of the solution,

$$H_{-} = pK_a - \log([HA]/[A^-]) \tag{5}$$

where the square brackets indicate concentration.

The solution equilibrium is defined using a mass-action kinetics type-expression for the bases and acids considered, i.e., BzTMAH for AEM and sulfuric acid for Nafion. The degree of dissociation, α, represents the fraction of ionized, or dissociated, acid or base groups. It is used to identify the acid and base equilibrium constants, K_a and K_b, respectively. To fit the experimental solution data, a hydration parameter, n, represent the primary solvation shell of the H^+ or OH^- group during dissociation. This provides,

$$\frac{K_b}{\left(C_{H_2O}^{STP}\right)^n} = \frac{\alpha^2}{\left(C^o\right)^{n-1}(\lambda - n\alpha)^n \cdot (1-\alpha)} \quad \text{where} \quad \begin{array}{l} C_{OH^-} = C_{BzNMe_3} = C^o\alpha \\ C_{BzNMe_3OH} = C^o(1-\alpha) \end{array} \tag{6}$$

for the BzTMAH solutions, where C^o is the concentration of the BzTMAH in the dry membrane, and λ is the ratio of H_2O and un-dissociated acid/base concentration and will later represent the H_2O uptake coefficient in the respective membranes. $C_{H_2O}^{STP}$ is the concentration of H_2O under standard conditions (e.g., ~55 M). The solute-solution interactions become prominent as the concentration of the acid or base increases (*17*). These interactions represent changes in the solution activity. The solvation shells can maintain stronger interactions with the water than the water does by itself. This implies that the solution will appear as if the concentration of water has reduced through the formation of intermediate hydrated species. This concept is used to empirically fit the acidity functions to the experimental data for the species of interest. We use the dissociation model (equation. 6), in conjunction with an stoichiometric hydration number for the ion in solution "n", the dissociated base "m", and the un-dissociated base "r" to fit the acid/base scales (*23–24*). The modified BzTMAH system can be shown,

$$pH \approx H_{-} = pK_w + \log\left[C_{OH^-} \middle/ \left(\frac{C_{H_2O}^{STP} - nC_{OH^-} - mC_{BzN^+Me_3} - rC_{BzN^+Me_3OH^-}}{C_{H_2O}^{STP}}\right)^{(n+m+r)}\right] \tag{7}$$

and a similar form can be shown for the acid equivalent. The results of this fit can be found in figure 4. The hydration numbers and the negative log of the acid/base dissociation constants are fit for solutions at different molarities. Both experimental acidity functions (symbols) are shown with increasing acid/base concentration, and the model (lines) show reasonable agreement. The respective pK values in each increase with the acidity functions. Results of this analysis are also provided in table 3, along with some membrane properties that will be used in the proceeding sections.

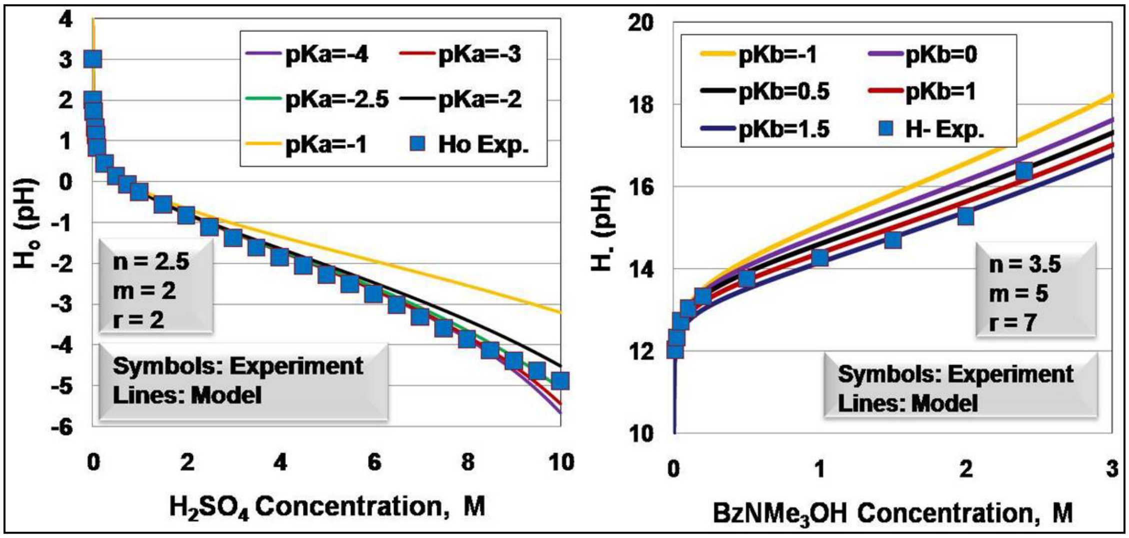


Figure 4. Empirical fit of solvation and dissociation models (lines) to solution data (symbols) with increasing acid/base concentration for the (left) sulfuric acid, and (right) benzyltrimethylammonium-hydroxide (BzTMAH) (*23–24*).

Table 3. Membrane solvation and dissociation properties.

Parameter	*AEM-ETFE*	*AEM-FEP*	*Nafion 115*
pK_b (AEM) or pK_a (Nafion)	1.0 (±1)	1.0 (±1)	-3.0 (±1)
n	3.5 (±1)	3.5 (±1)	2.5 (±1)
m	5.0 (±1)	5.0 (±1)	2.0 (±1)
r	7.0 (±1)	7.0 (±1)	2.0 (±1)
IEC (0% R.H.), mol eq./g [7]	$1.04 \times 10^{-3} \pm 0.07$	$1.03 \times 10^{-3} \pm 0.11$	$0.93 \times 10^{-3} \pm 0.07$
C_M(0% R.H.), mol/dm^3 [7]	1.80 ± 0.16	1.57 ± 0.20	1.84 ± 0.16

3.2 Validation of Dissociation and Solvation Model for AEM

The dissociation and solvation model is now applied to the AEM and Nafion. Equation 6 is used to describe the degree of dissociation associated with the functional groups in the respective membranes. Outside of the hydration number for the dissociated mobile charge carrier—i.e., H^+ and OH^-, respectively—the empirical hydration numbers for both the dissociated and un-dissociated counter-ions (e.g.,, functional groups) are not considered. The fixed charge concentrations in the AEM and PEM membranes (C^o in equation 6) are determined from the ion exchange capacities (IEC) and dry membrane density, ρ_o, (*7*). The concentration of dissociated OH^- and H^+ in the AEM and PEM systems,

$$C_{OH^-} = \alpha \Big/ \left(\lambda_{AAEM} \overline{V_{H_2O}} \right) \qquad C_{H+} = \alpha \Big/ \left(\lambda_{PEM} \overline{V_{H_2O}} \right) \tag{8}$$

where the water uptake coefficient, λ, and partial molar volume of H_2O, $\overline{V_{H_2O}}$, are also needed. The pertinent parameters to the respective membranes, in addition to the insight afforded by the preceding section, are provided in table 3. The confidence values were approximated using a sensitivity analysis.

Validation of the dissociation and solvation models for membrane processes are treated with several unique approaches and discussions. The equilibrium constant and hydration numbers for Nafion compare well with independent measurements and studies using a variety of techniques, ranging from titration and spectroscopic to first principles simulations (*18–22, 25–27*). This is corroborated by recognizing that the dissociation constants identified correspond to those that have been suggested in ion exchange texts (*28–29*). FTIR and Raman spectroscopy studies by Vico et al. suggest large hydration numbers should be anticipated for AEMs, as was recognized (*30*). Similarly, using an approach outlined by Marchael et al*.,* we can compare the solvation properties of the Nafion and AAEM membranes at high relative humidity by comparing the slope of a log-log plot of the membrane conductivity and relative humidity (*31–32*). For the AAEM,

$$-Bz(Me)_3N^+OH^- + p \cdot H_2O \leftrightarrow -Bz(Me)_3N^+(H_2O)_r + OH^-(H_2O)_n \tag{9}$$

$$p = n + r = \propto \frac{1}{2} \cdot \frac{\log(\sigma)}{\log(\%R.H.)} \tag{10}$$

where ‘p’ is the total hydration number. This approach assumes a consistent transport mechanism at high water vapor activities. Using experimental data from Varcoe (*7*), figure 5 provides the total hydration number for Nafion 115, AEM-FEP, and AEM-ETFE as approximately 4→5, 10→11, and 10→11, respectively. This agrees with the hydration numbers presented in figure 4 and table 3. It also agrees with the measurements by Marcael et al. for Nafion and Stoicha et al. on a unique AEM (*31–32*).

Figure 5 also provides normalized OH^- concentration (lines) predictions. These concentrations are predicted using the membrane dissociation model for PS-AEMs, with ion exchange capacities ranging from 0.69 to 1.89 mmol-equiv/g. The respective concentration predictions are normalized and compared to conductivities (symbols) reported by Hibbs et al. (*8*). The conductivities are also normalized for comparison purposes. In this analysis, the dry membrane density, ρ, is parameterized. The model validation is only valid at the discrete data points.

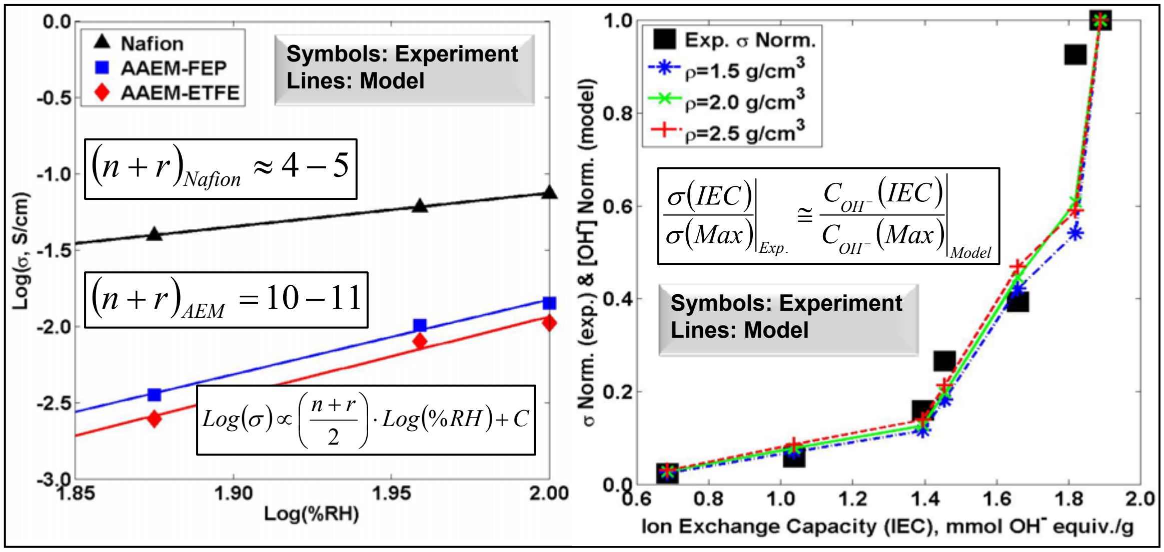


Figure 5. Membrane dissociation model validation. (Left) A hydration numbers of approximately 4–5 for Nafion and 10–11 for the FEP and ETFE AEM is identified. Experimental data (symbols) is from Varcoe (*7*). (Right) The membrane dissociation model was applied to the PS-AEM. The reported conductivities are normalized as a function of ion exchange capacity. This trend is compared to the normalized OH^- concentrations determined by the membrane dissociation model. The membrane density, ρ, is parameterized. Experimental data (symbols) is from Hibbs et al. (*8*).

3.3 Application of Dissociation Model to Conductive Transport in AEM

The dissociation and solvation models are used to understand the effect of the processes on ionic transport in the AEM. The generalized Stefan Maxwell membrane model, equation 2, is used for this process with equation 8 to prescribe ionic charge carrier concentration. In implementing the membrane dissociation and solvation model to conductive transport in the AEM and PEM, a modification has been made. This modification takes the form of an empirical correction to δ, which scales the ratio of the mutual diffusion coefficient of mobile species, and that between the mobile species and membrane structure. This parameter is now described as a function of the water uptake. This modification is shown in equation 11. This results in a scaling effect of the membrane transport as the membrane water content increases, opening pores and more loosely bound regions of water (*18–22, 25, 29, 30*). The correction can be shown as

$$\delta = \frac{\sqrt{2}}{\lambda} \cdot \left(\frac{\overline{V_M}}{\overline{V_{H_2O}}} \right)^{2/3} \tag{11}$$

where all inputs are membrane properties, and λ is the water uptake. It is the only property that depends upon environment conditions. The basis for this correction is empirical, but is supported by the hydration numbers identified in the preceding sections (e.g., activity effect). It is supported by numerous independent NMR methods reported in the literature, which examines how transport rates in PEMs have shown large activation energies at low water contents, indicating strong physiochemical interactions between the solvent solution and membrane. A thermodynamic approach by Datta and coworkers suggests a similar dependence using asymptotic limits (*18–19, 25*). Similar analytic developments support such a form.

3.4 AEMFC H_2O Transport and Polarization Model

The final portion of this study considers a finite volume H_2O transport and polarization model for the AEM in a H_2/Air (O_2) AEMFC. Many of the morphological, transport, and physical properties of the AEMFCs are still being determined. Parameterization and estimates are used in place of these properties at this time. Additionally, a 1-D model is currently used, and multiphase effects are not yet considered. Determining many of these properties will be the subject of ongoing work. A schematic of the model used is provided in figure 6. As shown, it is proposed that the AEM supports electro-osmotic drag, diffusion, and permeation. A single phase model is considered and permeation is neglected.

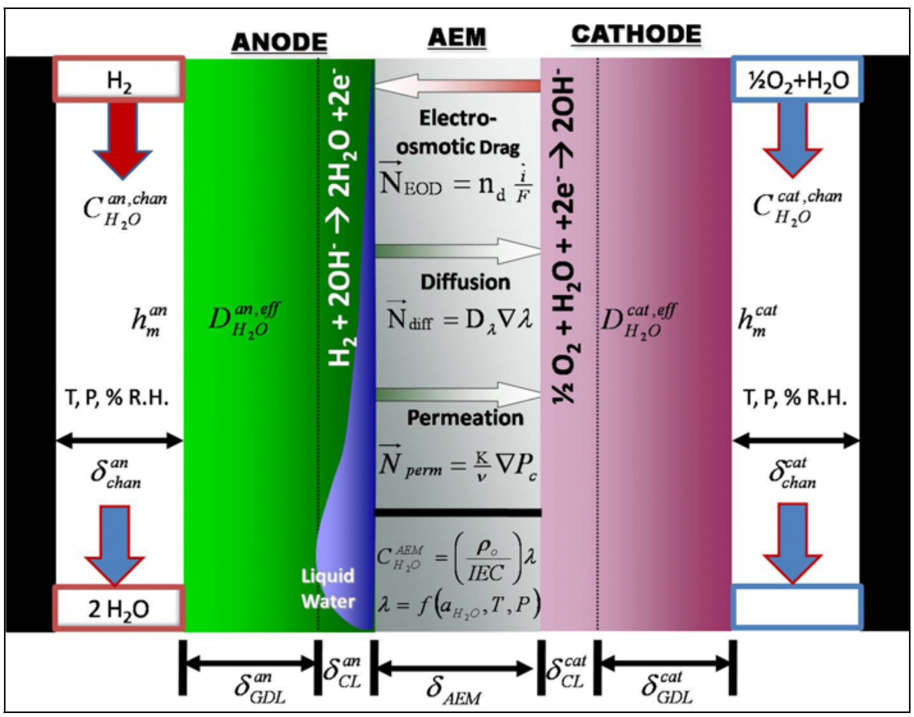


Figure 6. Schematic of the 1-D AEMFC membrane transport and polarization model.

To implement this framework, the fluxes at the AEM/electrode interfaces can be identified through a molar flux balance (flux from cathode to anode defined as positive),

$$\vec{N}_{H_2O}\Big|_{an} = n_d \frac{\dot{i}}{F} - \frac{\rho_o}{IEC} D(\lambda)\cdot\nabla\lambda - \frac{K}{\nu}\nabla P_c + \frac{\dot{i}}{F} \quad \text{and} \quad \vec{N}_{H_2O}\Big|_{cat} = \vec{N}_{H_2O}\Big|_{an} - \frac{\dot{i}}{2F} \tag{12}$$

where n_d is the electro-osmotic drag coefficient, and the membrane water diffusivity, D, is a function of the water uptake coefficient, λ. Preliminary studies suggest that n_d may be considerable in AEMs, but they are not well substantiated (*33*). Other studies suggest that the water diffusion in the membrane may be considerably smaller than in PEMs. This is supported by the solvation numbers and transport rates previously identified in this study. Here, preliminary data from Hibbs et al. (*8*) is used with a dependence upon pore-volume fraction. More detailed estimates will be used in ongoing work, including developments from the conductive membrane transport analysis. Part of the reason this has not yet been expanded is that absorption/desorption in the AEM have yet to be considered. While uptake curves are comparable to Nafion, the resistance of the membrane to sorption is not known. This is analogous to contact resistances, but has been shown as important in recent PEM studies (*34*). These effects can typically be shown as,

$$\vec{N}_{H_2O} = \frac{\rho_o}{IEC} k_{sorp}\left(\lambda^* - \lambda\right) \Rightarrow k_{sorp} = f\left(\lambda, P, T, \pm\vec{N}_{H_2O}\right) \tag{13}$$

where k_{sorp} is mass transfer coefficient for the absorption process. Transient membrane sorption analysis can provide insight with the scaling of dimensionless groups such as the mass Biot and mass Fourier numbers (e.g., ratio of interfacial sorption to diffusive transport). Preliminary efforts are being pursued to determine these properties, but k_{sorp} is set effectively to infinity in this study.

With finite volumes considered for the membrane, concentration boundary conditions at the membrane interface are used to determine the membrane water content, λ. These concentrations are determined using the net water fluxes from equation 12 with treating the electrode transport as transport resistances. Such an approach is valid for steady state diffusive transport with Fick's law. This is shown as

$$R_i = \frac{\Delta C_{H_2O}}{\vec{N}_{H_2O}} \quad \text{and} \quad R_{net}^{Series} = \sum_i R_i = \frac{C_{H_2O}^{i=1} - C_{H_2O}^{i=n}}{\vec{N}_{H_2O}} \tag{14}$$

$$R_{chan}^{an} = \frac{1}{h_m^{an}} \quad R_{GDL}^{an} = \frac{\delta_{GDL}^{an}}{D_{H_2O}^{an,eff}} \quad R_{CL}^{an} = \frac{\delta_{CL}^{an}}{D_{H_2O}^{an,eff}} \tag{15}$$

where comparable resistances for the cathode can also be shown. The convective mass transfer coefficient between the fuel/oxidant channel and the GDL can be determined from a Prandtl and mass Schmidt number. Flow in the channel is laminar (e.g., low Reynolds number); volumetric/mass flow rates have little importance on convection coefficients in this limit. Diffusion coefficients can be taken from standard transport references at the appropriate conditions and are corrected for the effective GDL pore structure with a Bruggeman exponent correction.

The solution procedure for the membrane transport begins with an initialization to the concentrations representative of a linear interpolation of the channel conditions. Transport is allowed to relax and updated fluxes are calculated. An under-relaxation process is employed for stability purposes. Upon completion of under-relaxation, the boundary conditions for the AEM finite volumes are updated. Convergence was set to maximum uptake changes in the membrane less than $1\text{x}10^{-10}$. Grid dependence studies were also completed for all performed studies.

Upon completion of the transport model, the polarizations for the AEM were interpreted. These include the overpotentials due to ohmic losses in the membrane and catalyst layer,

$$\eta_{ohm}^{AEM} = \int_0^{\delta_{AEM}} \frac{1}{\sigma(\lambda)} \cdot \overrightarrow{i_{OH-}} dx \qquad \eta_{ohm}^{CL} = \cdot \overrightarrow{i_{OH-}} \left(\frac{\delta_{CL}^{an}}{\sigma_{AEM}^{an}} + \frac{\delta_{CL}^{cat}}{\sigma_{AEM}^{cat}} \right) \tag{16}$$

where the local conductivities are determined from the dissolution and AEM transport models from the previous sections. The oxygen reduction overpotential was prescribed using Tafel kinetics,

$$\eta_{ORR} = -b \cdot \log\left(\frac{i + i_{xo}}{i_o} \right) \tag{17}$$

where values for the Tafel slope and exchange current were taken from the literature (*5–7,9–10*). Concentration losses and hydrogen oxidation activation losses were neglected at this time because high fuel/oxidant concentrations were used, and the oxidations reactions are considerably more facile than oxygen reduction.

4. Results and Discussion

4.1 Effect of Membrane Dissociation and Solvation Model on Conductivity Predictions

With the updated solvation and dissociation models, the ionic transport model framework was used to describe the conductivity of the AEM and Nafion 115. The results of this model are provided in figure 7, where the degree of dissociation for the respective membranes is plotted as are their respective conductivities. The degree of dissociation represents the fraction of available functional groups that can provide their OH^- to the transport processes. The discrepancy between Nafion and AEM are considerable and demonstrate a fundamental difference in the nature of these membranes. Further, using these updated results, the agreement between the experimental conductivity data (symbols) from (*7*) and the updated model (lines) is recognized.

In figure 7, the Nafion H^+ conductivity and the AAEM OH^- conductivities are plotted versus their respective water uptake coefficients. Note that the y-axis does not maintain the same scale. Nafion 115 has nearly a five times higher conductivity than the AAEMs at a water vapor activity of unity (i.e., maximum shown uptakes). Unlike previous efforts (*15*), the additional empirical dependence on pore volume fraction, ε, was not required. This additional empiricism was noted during previous efforts, but could not fully be explained. Further, it was unique to the AEM. By implementing the dissociation and solvation models to predict the concentration of OH^- charge carriers available to participate in the transport (e.g., donated by the functional groups), this empiricism has been removed, and consistent numeric constructs for the AEM and Nafion are considered. Membrane properties considered for this study are provided in table 3.

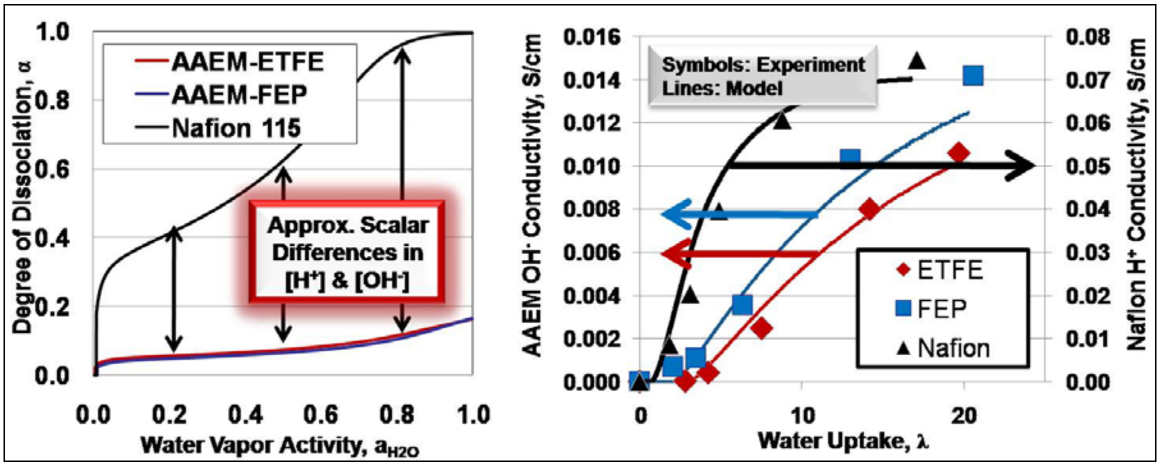


Figure 7. (Left) The degree of dissociation for the Nafion and AEM membranes is presented, representing the fraction of functional groups providing a charge carrier to the solvent. (Right) The updated membrane conductivity model implementing the dissociation model provides good agreement between the model (lines) results and experimental data (symbols) (*7*) without the need of unique empiricism.

The results shown in figure 7 imply that the unique physiochemical interactions between the AEM functional groups and water solvent are important. Higher ion exchange capacities and morphological control can improve these interactions; however, the uniqueness of these systems is demonstrated. Because of the approach taken, additional validation is needed. This is being pursued for different forms of PEMs and several weak acid/base membranes. Titration based methods may also be used on the membranes to determine these factors directly.

4.2 AEMFC Transport and Polarization Model

Results from the AEMFC transport and polarization model are presented in figure 8. This figure considers three membranes of different thickness in an H_2/O_2 AEMFC (*6*). In this figure, the model (lines) is compared to the experimental polarization curves (symbols). The electro-osmotic drag coefficient and water diffusion coefficient are defined as previously discussed and noted in figure 8. The reference value chosen for the diffusion coefficient was $4\text{x}10^{-10}$ m^2/sec, based upon an average from Hibbs et al. (*8*). Additional parameterization of the diffusion coefficient has been considered. Values ranging from functionality reported to that in the Nafion and PEM literature to asymptotic limits (±∞) have been examined. An average is presented here for the sake of simplicity until these dependencies can be better understood. These values are subject to change in future studies with additional understanding. Qualitative validation with the polarization curves provides confidence in the applicability of the approach.

As observed on the right of figure 8, these studies show that the electro-osmotic drag and the Faradaic consumption and production of water at cathode and anode, respectively, are substantial. These processes can flood the anode. Probably, more importantly, the ORR reaction has decreased water content with increasing current density. Such effects are also expected to be cumulative down the flow channel in a planar cell. Figure 8 shows this for the approximately 80 μm AEM-ETFE S80 membrane, as the water uptake decreases substantially on the cathode side with increased current. The effect these processes have on the ORR activation and concentration overpotentials needs to be better understood. The reduced AEM conductivity on the cathode side of the AEM, as well as in the catalyst layer, provides insight in to the losses in the AEMFC. It is also recognized how several authors have suggested they take the form of "mass transport losses," despite their linear nature with increasing current density (*2, 6–7, 13*).

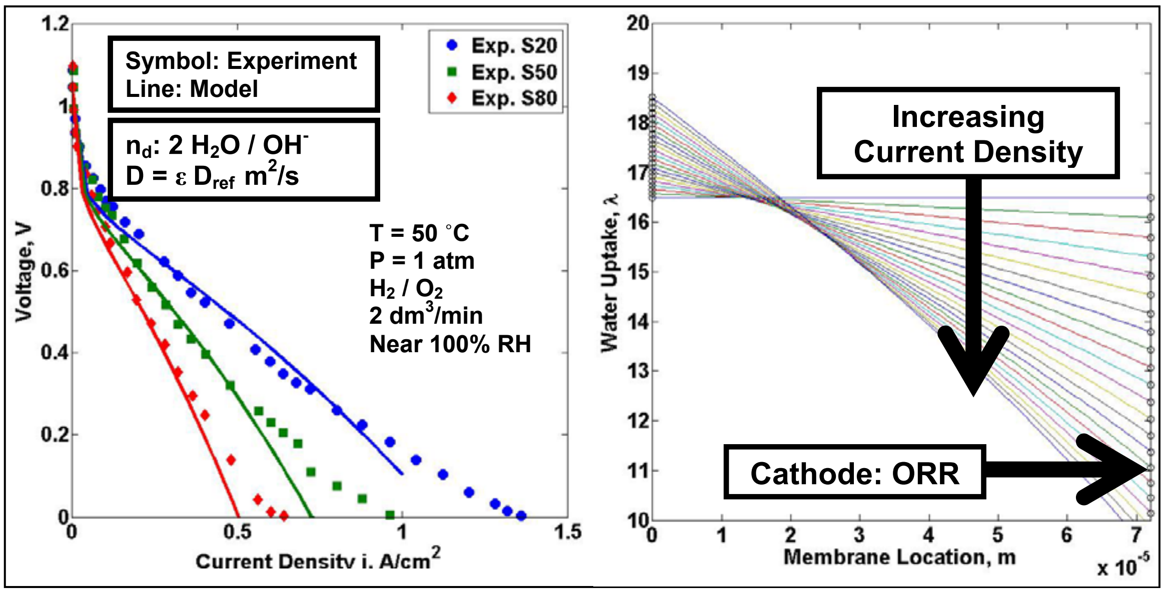


Figure 8. (Left) Preliminary model validation for AEMFC polarization curves presented be Varcoe group (*6*) for different AEM thicknesses. Experimental data (symbols) and AEMFC transport and polarization model (lines). (Right) Sample result for S80 AEM presenting the water uptake coefficient versus membrane location for trends of current density. The drying of the cathode, where the ORR takes place is noted.

Several additional AEMFC results for the S50 membrane are provided in figure 9. In this figure, the overpotentials and net water flux are provided for the above study. The reducing concentration of H_2O in the cathode results in an increase in ohmic losses in the catalysts layer. As recognized in the membrane ionic transport and conductivity studies, the AEM conductivity increases nearly linearly with water uptake. Such a scenario suggests that the incorporation and transport of OH^- in the cathode catalyst layer becomes more difficult with increasing current density; however, it should also improve in the anode due to water crossover effects. The effects of flooding will need to be considered for anodic processes in future studies. The water flux shown in figure 9 resembles that of the osmotic-drag flux due to low membrane diffusivity that was considered. At larger concentration gradients, resulting from large electro-osmotic fluxes and Faradic processes, diffusion effects do begin to show up. Some of these trends are represented with arrows on in this figure.

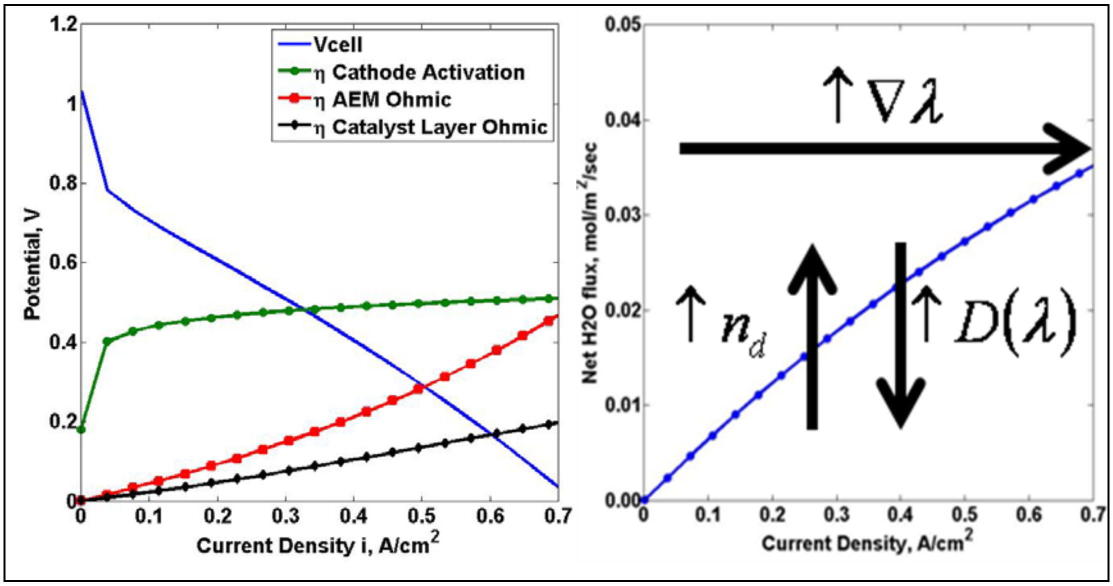


Figure 9. Select results for the S50 AEMFC. (Left) Overpotentials in the H_2/O_2 AEMFC at 50 C, 1 atm and nearly fully humidified, (Right) Net H_2O Flux through AEM. Conditions related to experiments reported in (*6*).

5. Summary and Conclusions

This report presented several developments aimed at enhancing present understanding of AEM and AEMFC technologies. These efforts are being pursued as an attractive and economic opportunity for solider portable power applications in a DAFC.

Initial studies have shown that the AEM are considerably more resistive than PEM. This work demonstrates the belief that these discrepancies are associated with the different equilibrium environments supported in the respective cells. These effects were shown to be further associated with the solvation and hydration numbers of the participating species. This was discussed using analogy to solution data, as has been reported in past literature. While these effects are substantial, it implies that different strategies are required for the control and optimization of the AEM and AEMFC than present in PEMs. Thinner membranes and higher humidification levels can improve these characteristics (*6*). Thinner membranes are problematic in PEM and PEMFCs due to crossover effects; they are more readily possible in the AEM and AEMFC because with decreased transport rates are decreased permeability. Similarly, the opposition of the osmotic fluxes will inhibit crossover losses. Additional manufacturing control strategies, such as cross-linking, block co-polymerization, inclusion of spacer groups, and higher exchange capacities, will assist in these efforts. More detailed guidance will be provided with ongoing efforts—specifically, as more detailed morphological characterization and transport details become available.

With a conductive transport and AEM resistance model in place, an AEMFC transport and polarization model was developed. This model is preliminary in nature, but demonstrates the aid that this type of predictive capability can provide. Several discussions in the AEMFC literature were directly observed. With improved characterization and understanding of present membranes, these models can become more detailed and accurate. These efforts are being pursued. Specifically, the effects down the fuel/oxidant channel and multiphase (e.g., saturation) effects require study. Similarly, consideration of multiphase effects and permeation (due to capillary pressure), and more detailed descriptions of the electro-osmotic drag coefficient, diffusivities, and sorption processes, require exploration. Once these effects can be accurately captured, strategies for optimization and mitigation of losses in the system related to transport can be considered and experimentally validated.

Initially, like the ionic transport conductivity studies, vapor-equilibrated AEMFC studies seem to suggest that thinner membranes are advantageous. Selective catalysts and lower, more selective membranes support thinner membranes under alkaline conditions. If development of higher IECs, cross-linked AEMs, and block-copolymer based AEMs can provide more controlled morphologies, these issues should be improved. At this time, the most pressing issue is the large discrepancies in water content at the cathode, where water is use for the ORR. The effect of that the decreased water concentration on both the conductivity in the catalysts layer and the activation losses needs to be addressed. Similar issues have been presented in the PEMFC literature (*2, 5–7, 9–10, 13–14*). The effect this matter will have on the reduction processes and OH^- activity in the cathode needs additional investigation. These types of insights will be advantageous for identification of the nature of the catalyst layer, gas diffusion layers, and operation strategies as AEMFC development moves forward. This is specifically true for identification of noble-metal free catalysts and appropriate MEA interface binding layers.

Some specific issues regarding the unique nature of the AEM and AEMFC have been identified and discussed in this study. Dramatic increases in performance, manufacturing methods/capabilities, and understanding have already been demonstrated in the associated literature and are likely to continue as understanding of the system improves. There are several issues that need to be addressed; however, none of the technical challenges presented in this study are insurmountable and/or overly daunting. Though, these types of issues do require additional experimental validation, consideration, and thought in ongoing development efforts. Further, a concerted effort to improve the characterization of these systems is needed to aid in these efforts. However, with continued development, the AEMFC-based DAFC is seen as a promising candidate for portable power applications with logistically favorable fuels.

6. References

1. Patil, A. S.; Dubois, T. G.; Sifer, N.; Bostic, E.; Gardner, K.; Quah, M.; Bolton, C. *J. Power Sources* **2004**, *136*, 220–225.
2. Varcoe, J. R.; Slade, R. C. T. *Fuel Cells* **2005**, *5* (2), 187.
3. Costamagna, P.; Srinivasan, S. *J. Power Sources* **2001**, *102*, 242.
4. Borup, R.; Meyers, J.; Pivovar, B.; Kim, Y. S.; Mukundan, R.; Garland, N.; Myers, D.; Wilson, M.; Garzon, F.; Wood, D.; Zelenav, P.; More, K.; Stoh, K.; Zawodzinski, T.; Boncella, J.; McGrath, J. E.; Inaba, K.; Miyatake, K.; Hori, M.; Ota K.; Ogumi, Z.; Miyata, S.; Nishikata, A.; Siroma, Z.; Uchimoto, Y.; Yasuda, K.; Kimijima, K.; Iwashita N. *Chem. Rev.* **2007**, *107*, 3904–3951.
5. Adams, L. A.; Poynton, S. D.; Tamain, C.; Slade, R. C. T.; Varcoe, J. R. *ChemSusChem* **2008**, *1–2*, 79–81.
6. Poynton, S. D.; Kizewski, J. P.; Slade, R. C. T.; Varcoe, J. R. *Solid State Ionics* **2009** (in press), DOI:10.1016/j.ssi.2009.01.019.
7. Varcoe, J. R. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1479–1486.
8. Hibbs, M. R.; Hickner, M. A.; Alam, T. M.; McIntyre, S. K.; Fujimoto, C. H.; Cornelius, C. J. *Chem. Mater.* **2008**, *20*, 2566–2573.
9. Yanagi, H.; Fukuta, K. *ECS Trans*. **2008**, *16* (2), 257–262.
10. Varcoe, J. R.; Slade, R. C. T.; Yee, E. L. H.; Poynton, S. D.; Driscoll, D. J. *J. Power Sources* **2007**, *173*, 194–199.
11. Chuy, C.; Basura, V. I.; Simon, E.; Holdcroft, S.; Horsfall, J.; Lovell, K. V. *J. Electrochem. Soc.* **2003**, *147*, 367.
12. Beatteie, P. D.; Orfino, F. P.; Basura, V. I.; Zuchowska, K.; Ding, J.; Chuy, C.; Schmeisser, J.; Holdcroft, S. *J. Electroan. Chem.* **2001**, *503*, 45–56.
13. Filpi, A.; Boccia, M.; Gasteiger, H. A. *ECS Trans.* **2008**, *16* (2), 1835–1845.
14. Liu, Y.; Murphy, M. W.; Baker, D. R.; Gu, W.; Ji, C.; Jorne, J.; Gasteiger, H. A. *J. Electrochem. Soc.* **2009**, *156* (8), B970–B980.
15. Grew, K. N.; Chiu, W. K. S. *ECS Trans.* **2008**, *13* (23), 61.
16. Krishna, R.; Wesselingh, J. A. *Chem. Eng. Sci.* **1997**, *52* (6), 861–911.

17. Bockris, J. O'M; Reddy, A. K. N. *Modern Electrochemistry 1*; 2nd ed., Plenum Pub., 1998.

18. Thampan, T.; Malhotra, S.; Tang H.; Datta, R. *J. Electrochem. Soc.* **2000**, *147* (9), 3242.

19. Choi, P.; Jalani, N. H.; Datta, R. *J. Electrochem. Soc.* **2005**, *152* (3), E84.

20. Glueckauf, E.;. Kitt, G. P. *Proc. R. Soc. London Ser. A* **1955**, *228*, 322.

21. Seger, B.; Vinodgopal, K.; Kamat, P. V. *Langmuir* **2007**, *23*, 5471.

22. Ma, C.; Zhang, L.; Mukerjee, S.; Ofer, D.; Nair, B. *J. Mem. Sci.* **2003**, *219*, 123.

23. Bowden, K. *Chem. Rev.* **1966**, *66* (2) 119.

24. Paul, M. A.; Long, F. A. *Chem. Rev.* **1957**, *57* (1), 1.

25. Choi, P.; Jalani, N. H.; Datta, R. *J. Electrochem. Soc.* **2005**, *152* (3), E123.

26. Horsfall, A.; Lovell, K. V. *Euro. Polymer Journal* **2002**, *38*, 1671.

27. Chen, J.; Asano, M.; Yoshida M.; Maekawa, Y. *J. Appl. Poly. Sci* **2006**, *101*, 2661.

28. Helfferich, F. G. *Ion Exchange*; McGraw-Hill, 1962.

29. LeVan, M. D.; Carta, G. *Perry's Chemical Engineer's Handbook*; 8th Ed., D. Green W.; Perry R.H., Eds., McGraw-Hill Co., Section 16, 2008.

30. Vico, S.; Palys B.; Buess-Herman C. *Langmuir* **2003**, *19*, 3282.

31. Marechal, M.; Souquet, M. L.; Guindet, J.; Sanchez, J. Y. *Electrochem. Comm.* **2007**, *9*, 1023.

32. Stoicha, D.; Alloin, F.; Marais, S.; Langevin, D.; Chappey C.; Judeinstein, P. *J. Phys. Chem. B* **2008**, *112*, 12338.

33. Yamanaka, U.; Takeguchi, T.; Takahashi, H.;Ueda, W. *J. Electrochem. Soc.* **2009**, *156* (7) B831.

34. Barclay-Satterfield, M.; Benziger, J. B. *J. Phys. Chem. B* **2008**, *112* (12), 3693.

Appendix. Acronyms, Definitions, and Nomenclature

Acronyms:

AEM:	Anion exchange membrane
AEMFC:	Anion exchange membrane fuel cell
AFC:	Alkaline Fuel Cell
BzTMAH:	Benzyltrimethylaommonium-hydroxide
DAFC:	Direct alcohol fuel cell
DMFC:	Direct methanol fuel cell
EIS:	Electrochemical Impedance Spectroscopy
FVM:	Finite volume method
HOR:	Hydrogen oxidation reaction
KOH:	Typical aqueous alkaline electrolyte
MEA:	Membrane electrode assembly
MOR:	Methanol oxidation reaction
ORR:	Oxygen reduction reaction
PEM:	Proton exchange membrane
PEMFC:	Polymer electrolyte membrane fuel cell, proton exchange membrane

Nomenclature:

α:	Degree of dissociation
ε:	Pore volume fraction in membrane or other diffusion media
ε_o:	Membrane pore volume fraction at percolation limit
δ_i:	Thickness of region i, m
δ:	Empirical diffusivity ratio for generalized Stefan-Maxwell equations
λ:	Water uptake coefficient, #H_2O/#Functional group
σ:	Ionic conductivity, S/cm
ρ_o:	Dry membrane density, g/cm^3
η:	Overpotentail, V
ν:	Viscosity, m^2/sec
$D_{i,j}$:	Mutual diffusion coefficient, m^2/sec
$D_{i,M}$:	Reduced diffusion coefficient resulting from structure, m^2/sec
F:	Faraday's constant, 96485 A sec/mol
h_m:	Convective mass transfer coefficient, m/s
ΔH_m	Enthalpy of melting, J/g
i:	Current density, A/m^2
IEC:	Ion exchange capacity, mmol-eq./g
k:	Interfacial sorption mass transfer coefficient, m/s
Ka:	Acid equilibrium dissociation constant
Kb:	Base equilibrium dissociation constant
N:	Molar flux, mol/m^2/sec
n_d:	Electro-osmotic drag coefficient, # H_2O/#OH^-
pKa:	$-\log_{10}$(Ka), pKw=pKa+pKb

P: Pressure, N/m^2
P_c: Capillary pressure, N/m^2
q: Bruggemen tortuosity factor exponent, =1.5
R: Universal gas constant, J/mol/K
T: Temperature, K
z_i: Charge number for ionic species

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Metabolite Analysis of *Clostridium acetobutylicum* Microbial Fuel Cells

by Timothy Mackie

Contents

List of Figures

Abstract

The anaerobe *Clostridium acetobutylicum* has shown promise as an active organism in mediator-less, whole-organism, microbial fuel cells. Such cells typically experience two voltage peaks over a week-long incubation period. This paper will demonstrate that these voltage peaks correspond to oscillating acidogenic and solventogenic phases in the metabolic output of *C. acetobutylicum*. We found that voltage peaks correlated to rapid generation of short chain organic acids, while subsequent periods of declining potential correlated to the reduction of these acids to their corresponding alcohols.

Acknowledgments

The author wishes to acknowledge the mentorship of Dr. James Sumner.

1. Introduction/Background

Microbial fuel cells present an unprecedented opportunity to reclaim energy directly from the fermentation of organic waste. Although their power output is low, they generate this power through a process that must occur for the safe disposal of waste water. Furthermore, these cells produce hydrogen gas as a byproduct of fermentation, which can then power a conventional hydrogen fuel cell. Since Microbial fuel cells can simultaneously process waste and generate power, they can potentially increase the efficiency of Army systems while preventing environmental damage from untreated waste (*1*).

Microbial fuel cells (MFCs) exploit the free electrons generated by anaerobic metabolic pathways to produce a current. All anaerobic respiration uses a molecule other than oxygen as a final electron acceptor, such as lactate or pyruvate (*1*). Some anaerobes do not fully reduce their electron acceptors and produce free electrons at the anode. These electrons then travel through the circuit to the cathode where they reduce O_2 to H_2O (figure 1). The overall reaction can be summarized as follows:

Anode (anaerobic): Substrate(reduced) $\rightarrow$ Substrate(oxidized) + H^+ + e^-

Cathode (aerobic): $O_2 + H^+ + e^- \rightarrow H_2O$

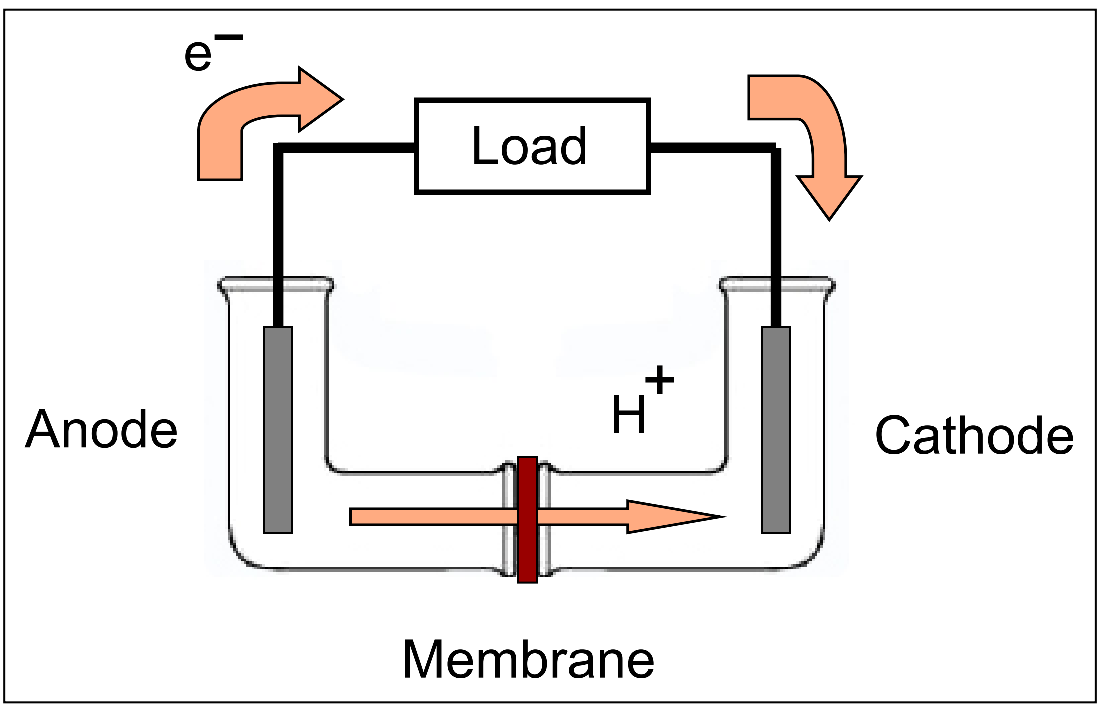


Figure 1. Simplified diagram of a U-tube MFC showing electron flow through the circuit and H^+ flow across the membrane.

Many MFC designs require moderator molecules such as methylene blue. These molecules intercalate in the plasma membranes of the bacteria and provide a channel for free electrons to leave the cell (*2*). Since these mediator molecules are toxic and ecologically persistent, organisms such as *Clostridium acetobutylicum* that can generate voltage without mediators are preferable.

Previous studies conducted on *C. acetobutylicum* MFCs revealed a curious "twin peaks" phenomenon in the current measured across a 10 kOhm resistor (figure 2). Unlike most MFC organisms, which generate a single voltage peak that gradually dies out, *C. acetobutylicum* produced two roughly identical voltage peaks separated by several days of lag. It was hypothesized that these two peaks correspond to oscillating metabolic phases of the organism.

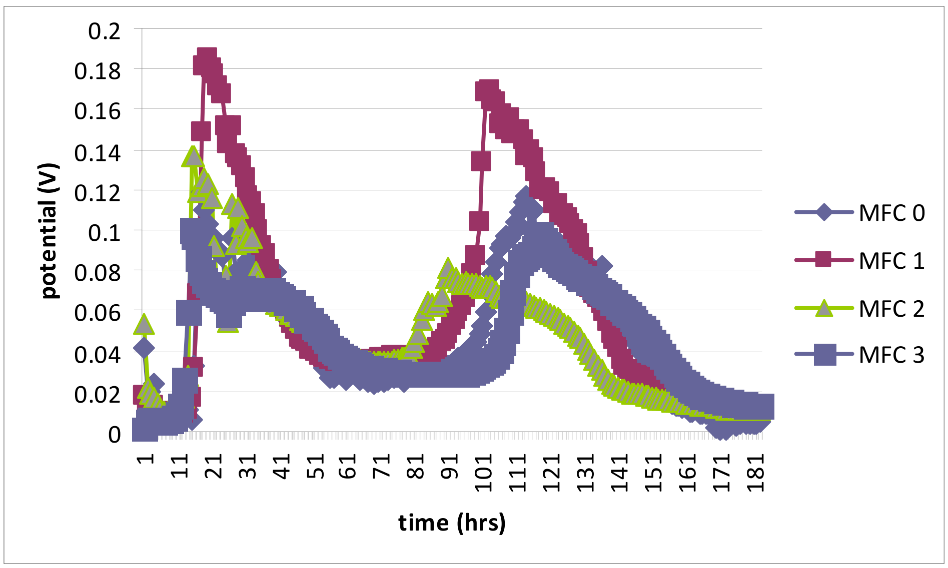


Figure 2. Timescale of MFC output over approximately one week of observation illustrating unique "twin peak" output of *C. acetobutylicum* MFCs.

A growth curve of *C. acetobutylicum* contains two distinct metabolic phases (figure 3). During the acidogenic phase, the organisms ferment glucose to pyruvate through glycolysis and then from pyruvate to small organic acids such as acetate and butyrate, replenishing their stores of NAD^+ and generating extra ATP. The buildup of butyric acid will eventually decouple the proton gradient since undissociated butyric acid is lipophilic enough to ferry protons across the cell membrane. Since *C. acetobutylicum* relies on an extracellular proton gradient to produce ATP, its ability to generate electric current in an MFC depends on how well it is able to deal with the buildup of butyric acid. The organisms then switch to the solventogenic phase using their large stores of NADH and NADPH to reduce acetate and butyrate to ethanol, butanol, and acetone. When the extracellular pH reaches a level that will not decouple the glycolysis reactions, the organisms again begin to ferment glucose in the acidogenic phase (*3*). Cultures of *C. acetobutylicum* will continue to oscillate between these phases until they either run out of food or toxify their environment with an excess of solvents. Presumably, the voltage of the MFC will initially grow exponentially with the cell density of the culture, then gradually fall off as the excess of butyric acid decouples the redox potential. Once the organisms convert enough excess butyrate to butanol, they will experience a second exponential growth phase, accounting for the second voltage peak.

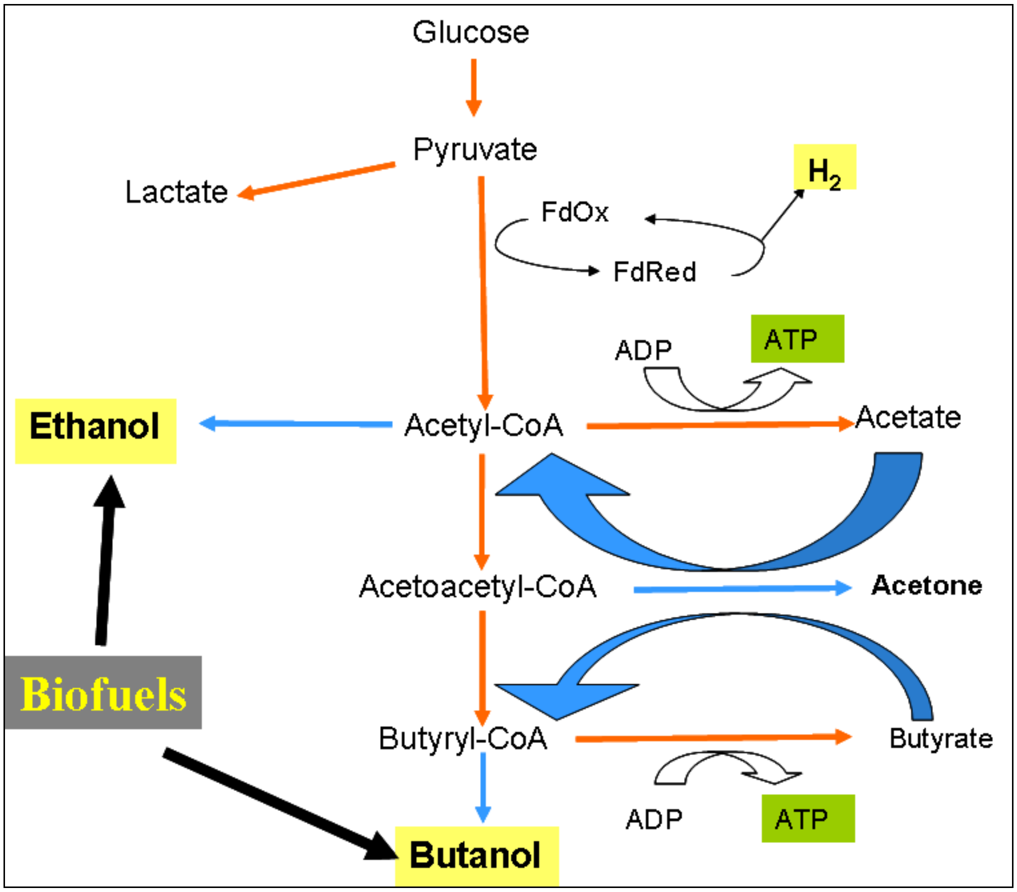


Figure 3. Simplified fermentation pathway for *C. acetobutylicum* with acidogenic reactions shown in orange and solventogenic reactions shown in blue. Adapted from White, D. (*2*).

---

2. Experiment/Calculations

The MFC housings were assembled from glass U-tubes (Ace Glass) separated by a Nafion (fuelcellstore.com) membrane pinched in the tubes' ball-and-socket joint. Both the anode and cathode chambers were filled with 20 mL of Clostridium Growth Media (CGM) based on the recipe developed by Ehrlich et al. (*4*) 500 mL of this media contained 25 g glucose, 2.5 g yeast extract, 1 g asparagine, 1 g $(NH_4)_2SO_4$, 0.5 g NaCl, 0.174 g $MgSO_4$, 5 mg $FeSO_4 \bullet 7H_2O$, 5 mg $MnSO_4 \bullet H_2O$, 0.375 g Na_2HPO_4, 0.410 g $NaH_2PO_4 \bullet 3H_2O$, and 100 µL antifoam C emulsion (all chemicals were purchased from Sigma-Aldrich).

The electrodes themselves were assembled from 1-in graphite cylinders (McMaster-Carr) connected to standard insulated wire by conductive silver epoxy. All connective parts of the electrode were coated with five-minute epoxy, leaving 0.5 in of graphite exposed. The anode was entirely submerged in media while the cathode was submerged just below the surface of the aerobic cathode. A 4-g of lump carbon (Goodfellow Cambridge Limited) was added to the anode chamber to increase the conductive surface area of the electrode. The anode chamber was kept anaerobic by threading the electrode lead through a rubber septum while the cathode chamber was kept aerobic by a loose glass cap.

Cultures of *C. acetobutylicum* (ATCC824) were prepared from sporulated stock by 10 min of heat shock at 80 °C and overnight incubation in CGM under anaerobic conditions at 35 °C. The culture was then diluted with CGM by a ratio of 1:5 before inoculation into the MFCs.

Before inoculation, the electrodes, housings, and media were assembled and autoclaved. The MFCs were placed in a 35 °C incubator, linked to a 10-kOhm resistor, and allowed to discharge overnight. They were inoculated with 100 µL of the diluted *C. acetobutylicum* culture. The anode compartments were vented with 12-gauge syringe needles fed through a 0.2-µm pore sterile filter through a length of plastic Tygon tubing into a flask of nitrogen-sparged water. This prevented buildup of metabolic gasses in the anode chamber while also preventing contamination of the chamber with O_2 or ambient microbes. At several points during the growth curve, 200-µL aliquots were withdrawn from the anode chambers for High Performance Liquid Chromatography (HPLC) analysis. The aliquots were filtered through 2-µm pore centrifuge filters and stored at –20 °C pending analysis. In a set of four MFCs, three aliquots would undergo HPLC analysis while the fourth would be pH tested.

The samples were analyzed by HPLC (Agilent Technologies) with a Multi-Wavelength Detector (MWD) set to 250 nm and a Refractive Index Detector (RID) (figures 4 and 5). The mobile phase consisted of 0.065 M H_2SO_4 in HPLC-grade water pumped at a flow rate of 0.600 mL/min. An Organic Acid Analysis Column (Bio-Rad) heated to 30 °C was used as the stationary phase. The samples were stored at 4 °C until ready for injection at volumes of 20 µL each. Calibration curves for glucose, ethanol, butanol, acetone, butyrate, acetate, lactate, and succinate were calculated using measured quantities of HPLC-grade standards dissolved in HPLC-grade water. Under these conditions, lactate, succinate, and acetone all eluted at roughly the same point, making their quantification impossible.

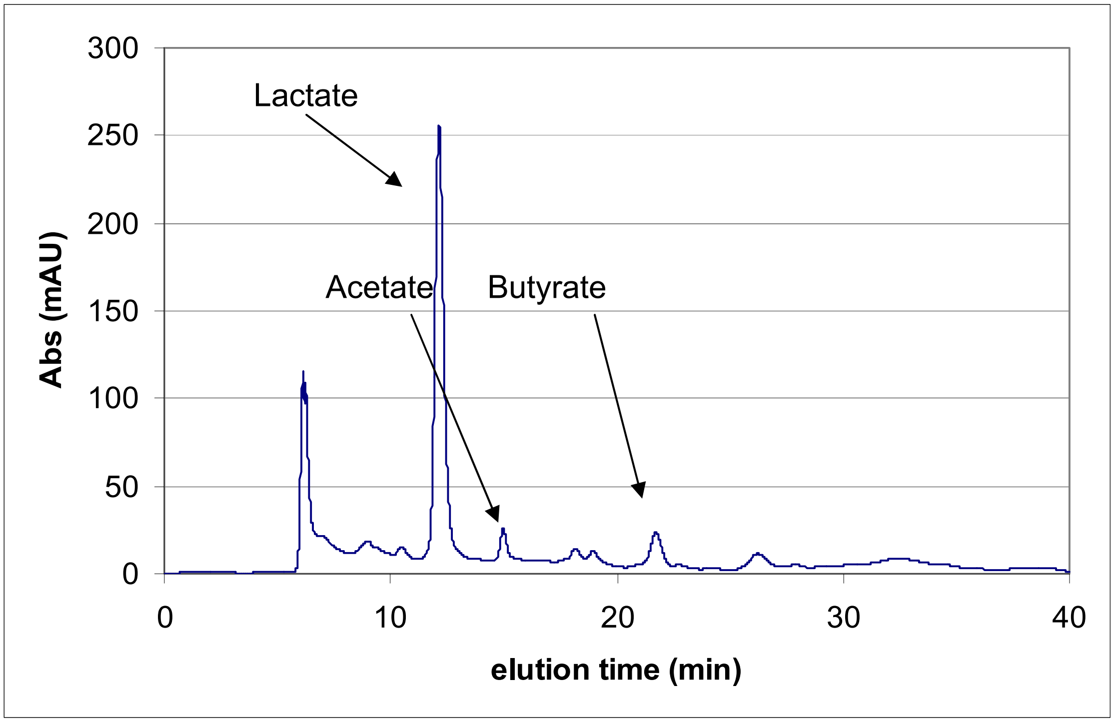


Figure 4. Sample absorbance chromatogram from HPLC showing separation of MFC metabolites.

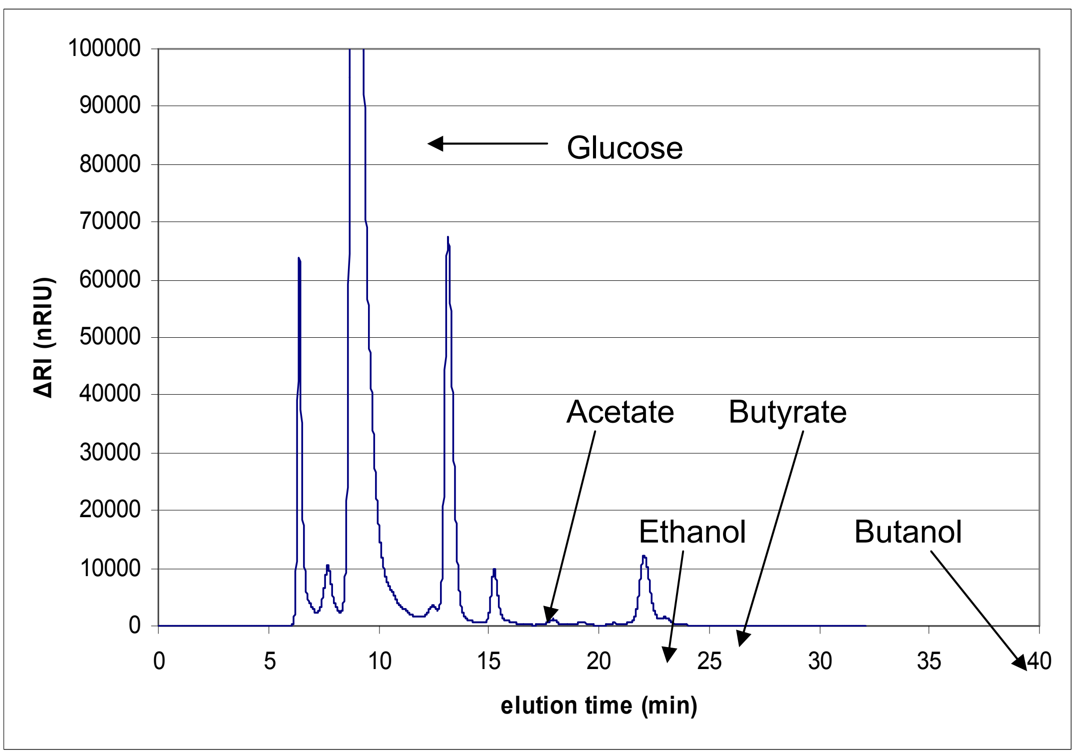


Figure 5. Sample refractive index chromatogram from HPLC showing separation of MFC metabolites.

3. Results and Discussion

HPLC quantification of the metabolic products of *C. acetobutylicum* in the MFCs showed a close correlation between concentration of metabolites and voltage output. During the initial voltage peak at approximately 20 h after inoculation (figure 6), butyrate concentration rose sharply while butanol concentration remained close to zero. After the spike, as voltage output slowly decreased, butyrate concentration leveled off, and then began to decline while butanol concentration began to rise rapidly. Trends were similar for acetate and ethanol, although correlation with expected relative levels was not as ideal as with butyrate/butanol (figure 7). Since acetate is not lipophilic enough to glucose, concentration dropped sharpest during the initial voltage peak, and then leveled off during the solventogenic phase (figure 8). This accounts for slower metabolism due to decoupling of the extracellular proton gradient by excess butyrate. Glucose was more difficult to accurately quantify than the metabolites since it was present in such large concentrations (about 300 mM initially). The total pH of the cells correlated closely with the concentrations of acetate and butyrate, falling sharply during the initial voltage peak, then gradually rising as acid concentrations fall and alcohols increase (figure 9). At present, metabolites from the time of the second voltage peak have not yet been quantified, although parallel trends are expected.

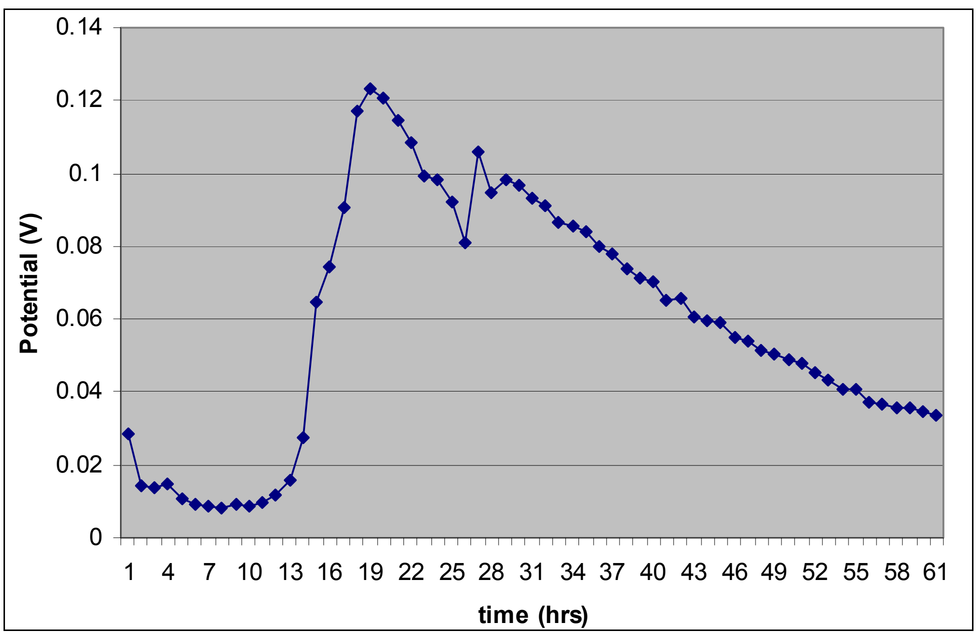


Figure 6. Average MFC potential for the first voltage peak.

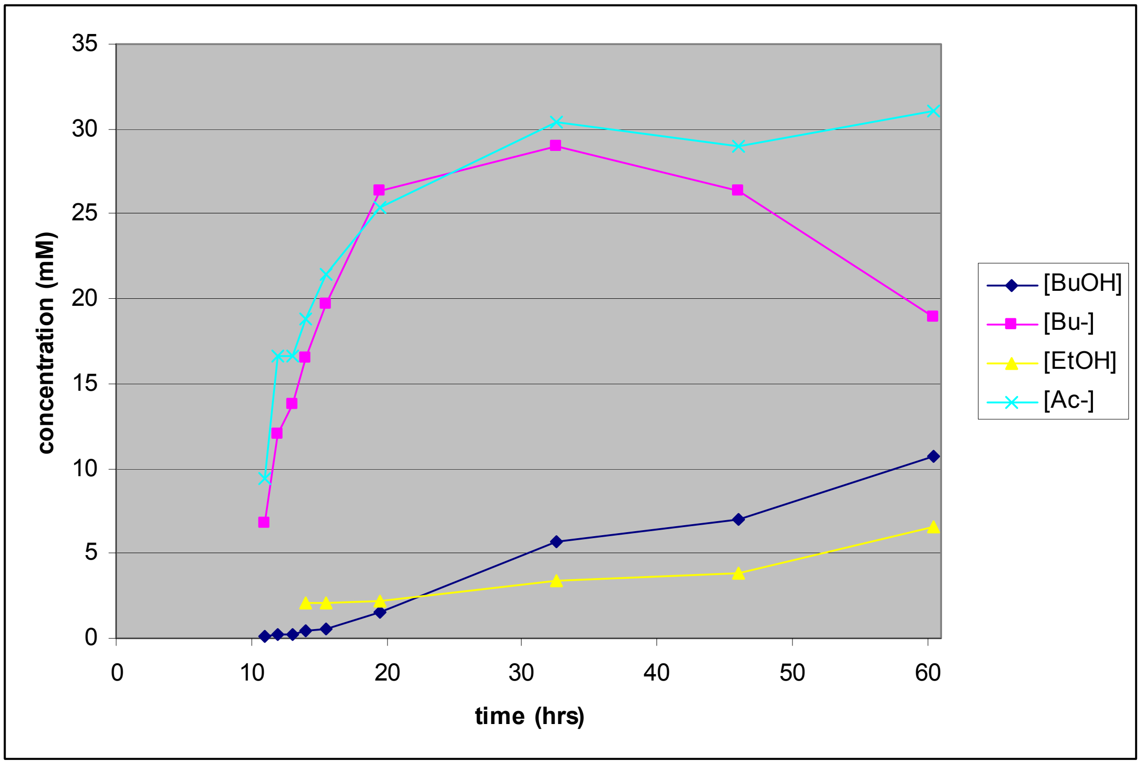


Figure 7. Observed trends in metabolite concentration over the first voltage peak for butanol (blue diamonds), butyrate (magenta squares), ethanol (yellow triangles), and acetate (cyan crosses). The first aliquot was withdrawn 11 h after initial inoculation

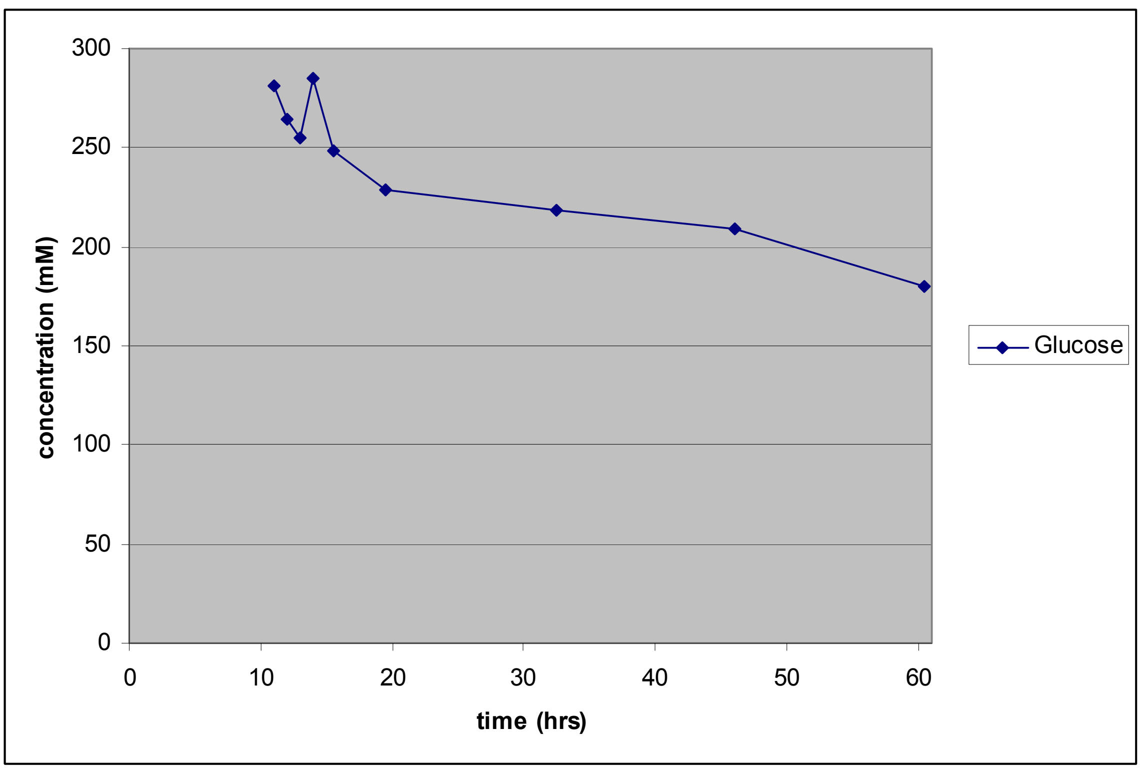


Figure 8. Observed trends in glucose concentration over first voltage peak.

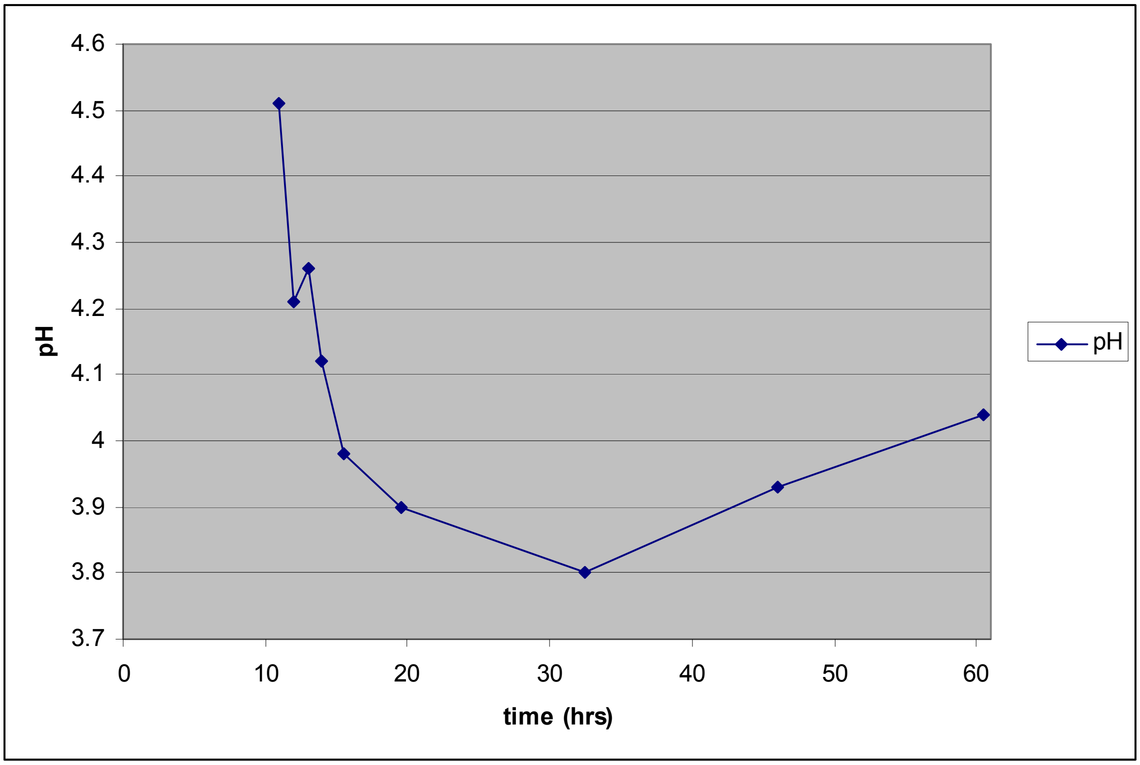


Figure 9. The pH trends over first voltage peak.

4. Summary and Conclusions

Current data suggests a direct correlation between voltage output and metabolic phase, although further research is needed to quantify the metabolite concentrations for the second voltage peak. Furthermore, a method must be developed to overcome the inability of the HPLC to isolate acetone. Such a method will likely involve gas chromatography (GC) analysis.

The correlation between voltage output and solvent formation makes *C. acetobutylicum* an ideal candidate for studying biofuel production and cellular metabolism. In addition to the current generated by the MFC itself, the organism also produces H_2 gas, which may then power a conventional hydrogen fuel cell. The acetone, butanol, and ethanol generated as byproducts of fermentation, although waste from the bacteria's perspective, may be distilled and used as biofuels or industrial solvents. By consolidating the functions of waste management, renewable power generation, and solvent production, *C. acetobutylicum* fuel cells have the potential to reduce the logistics trail and increase the efficiency of Army systems.

5. References

1. Bullen, R. A.; Arnot, T. C.; Lakeman, Walsh, F. C. *Biosensors and Bioelectronics* **2006**, *21*, 2015–2045.
2. White, D. *The Physiology and Biochemistry of Prokaryotes*; Oxford University Press: New York, 2007, pp. 383–403.
3. Sund, C. J.; McMasters, S.; Crittenden, S. R.; Harrell, L. E.; Sumner, J. J. *Applied Microbiology Biotechnology* **2007**.
4. Ehrlich, G. G.; Goerlitz, D. F.; Bourell, J. H.; Eisen, G. V.; Godsy, E. M. *Applied and Environmental Microbiology* **1981**, *42*, 878–885.

Survivability/Lethality Analysis Directorate (SLAD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Development of an Interactive Ray Trace Display for MGED

by Nicholas W. Reed

Contents

List of Figures

Abstract

While the Ballistic Research Laboratory's (BRL) Computer-Aided Design (CAD) Multi-Device Geometry Editor (MGED) offers an interactive display that shows geometry in a wireframe representation, many programs also can provide interactive shaded displays for geometry. In recent years, shaded displays for general geometry have been a frequent request of target modelers and vulnerability analysts who view target models in MGED. Their common problem of being able to quickly identify and understand the construction of a model would be alleviated by a shaded display, which provides a more natural representation of three-dimensional (3-D) models that is much easier to interpret. This report describes the results of my work to create an interactive shaded display for MGED that will work with standard (constructive solid) geometry as well as triangle mesh geometry. The resulting prototype makes use of BRL-CAD's existing ray tracing functionality to render geometry. This is a more accurate and computationally intensive approach compared to the various approximation methods used in other modeling software. For this reason, developing strategies for high performance is a top priority of future work on the prototype.

Acknowledgments

The author wishes to acknowledge the mentorship of Edwin O. Davisson.

1. Introduction/Background

Perhaps the most useful feature seen in many three-dimensional (3-D) modeling applications is the ability for users to interactively edit geometry. Interactive editing typically means that as the user manipulates geometry (e.g., with a mouse and keyboard), the program's visual representation of that geometry is instantly updated to reflect the changes. This gives the user the impression of directly modifying a physical object in real time, which makes the process of modeling faster and more intuitive.

Interactive editing is made possible by the combination of an interactive input system and an interactive display system. However, the processing required to continuously modify geometry based on user input is trivial compared with the processing needed to update the display fast enough to produce a smooth real-time display animation.

A standard solution in 3-D modeling applications has been to use a simplified wireframe representation of geometry, because it is visually useful but computationally fast to create. However, many applications can now take advantage of modern processors and graphics hardware to create a fully shaded representation of geometry fast enough to be used in an interactive display (figure 1).

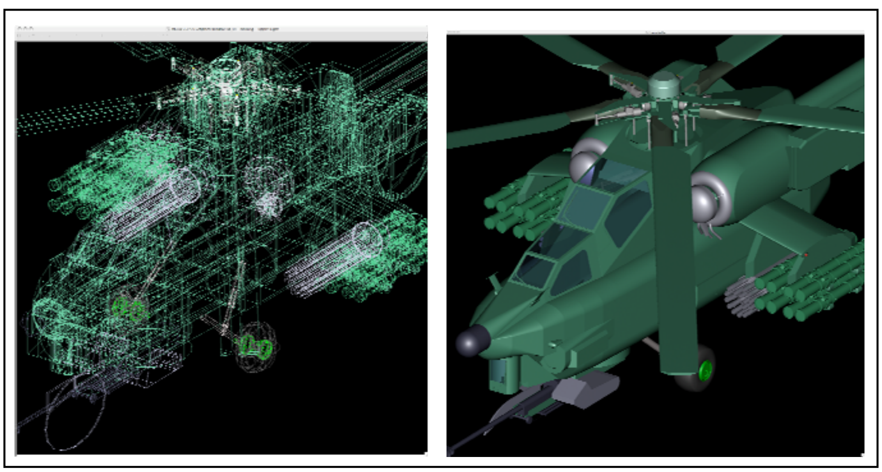


Figure 1. A model of an Mi-28 attack helicopter seen in a wireframe and shaded representation.

High quality shaded renderings are normally created via ray tracing, which is a physically accurate but resource intensive and time-consuming process. However, modeling applications that provide an interactive shaded display usually rely on approximate methods rather than ray tracing to achieve better real-time performance.

The Ballistic Research Laboratory (BRL)-Computer Aided Design (CAD) is a constructive solid modeling and geometric analysis suite. It includes a 3D modeling application called MGED (Multi-Device Geometry Editor) that is primarily used to create models of military targets. MGED provides an interactive display that can represent geometry in wireframe and other forms. MGED's support for an interactive shaded display is currently limited to triangle mesh geometry, but BRL-CAD does have high performance ray tracing capabilities that can be used in the development of an interactive shaded display for general (non-evaluated) geometry.

2. Experiment/Calculations

The goal of my work was to create a prototype interactive shaded display for MGED that utilized BRL-CAD's existing ray trace functionality. In the context of the suite, this translated to writing code for a new display manager. In order to support disparate computing environments, multiple display managers exist which have similar interfaces, but use different graphics specifications (e.g., OpenGL). For the purposes of my work, I chose to base a new ray tracing display manager off of the existing OpenGL display manager. Display managers typically receive lists of points to be displayed as line segments, and significant changes had to be made to incorporate ray tracing into a new prototype.

The most challenging and time-consuming part of work consisted of reviewing the existing code and gaining a working understanding of what it does, and how it does it. During this seemingly never-ending process, I learned how to use some useful tools designed for searching large code bases.

The first step in coding was to write appropriate definitions and hooks into various high level functions and initialization files so that the new display manager would be available for use in MGED. Unfortunately, I encountered some bugs along the way due to some recent changes to the Mac OS X operating system on which I was working. This resulted in some delays and required me to slightly alter my approach in getting the new display manager's framework built.

With the new display manager in place, my focus shifted to finding a place in the display manager code where ray tracing could be done. Once a suitable place was found, I then had to determine what program information was needed to perform ray tracing on visible objects, and where the display manager could get that information. This phase of coding was the most crucial, and once it was completed I was able to begin adding the desired functionality for the prototype display.

Compared with the initial setup work, familiarizing myself with the use of BRL-CAD's ray tracing functions and OpenGL's draw functions was relatively simple. I did some initial testing by firing just a few rays at geometry, and drawing them as points in the display window. With those crucial steps taken, the remainder of my work consisted of improving render quality and performance. First I moved to firing a coarse grid of rays from one direction. This resulted in an interactive point-cloud representation of displayed geometry. However, with rays from only one direction, it was easy to miss surfaces nearly parallel to those rays, resulting in gaps (figure 2). Advancing to three grids of rays fired in the three principles directions (X, Y, Z) provided more even results (figure 3).

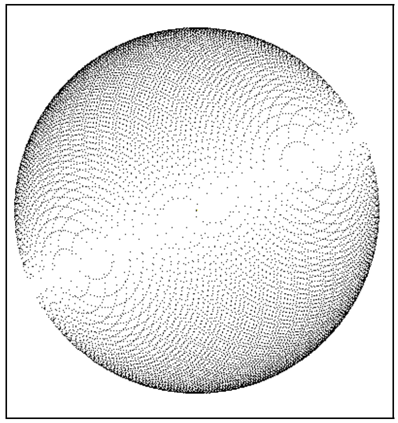


Figure 2. Intersection points between a sphere and a grid of rays in one direction.

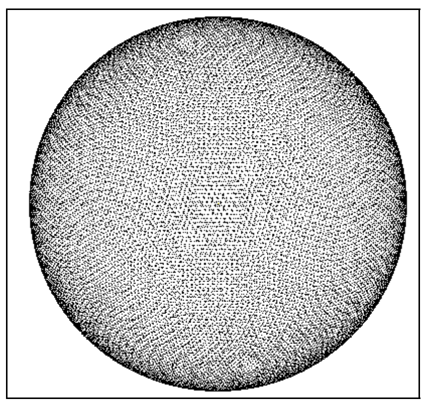


Figure 3. Intersection points between a sphere and three grids of rays in three directions.

3. Results and Discussion

The result of my work, thus far, is a display manager that represents geometry with intersection points from ray tracing. Methods for displaying the geometry are the same as in existing display managers that provide wireframe views. The performance is generally good enough to allow real-time interaction, but the display shows noticeable lag when rendering complex objects composed of a large point set.

The prototype display manager provides a novel view of geometry in its intermediate stage (figure 4), but continued work will bring the rendered image closer to that seen in figure 2.

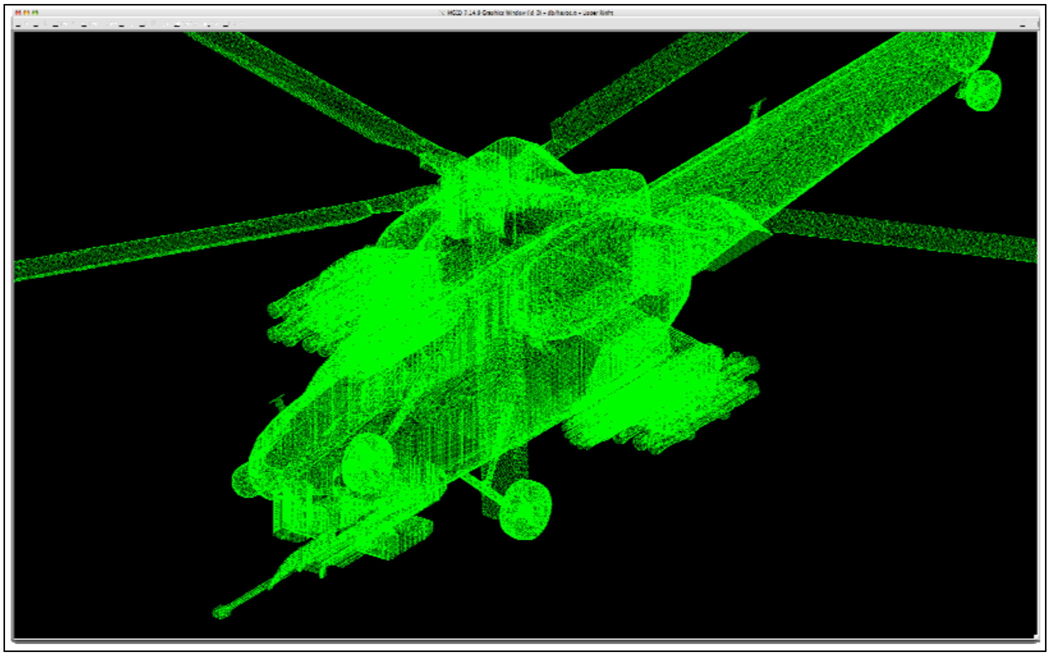


Figure 4. The figure 1 model viewed with the prototype display.

4. Summary and Conclusions

I have written prototype code that displays geometry interactively using ray traced points. In its current form, the prototype display is novel but incomplete. Future work will include increasing point density to avoid gaps, and calculating lighting on a per-point basis to create a shaded view. Optimizations to perform ray tracing progressively should yield a practical interactive shaded display for MGED.

This work required me to learn about the BRL-CAD suite’s display manager and ray tracing code, as well as the OpenGL graphics specification. I have gained valuable programming experience and knowledge of computer graphics that complements my university coursework as a Computer Science major.

5. References

1. Shreiner, D. *OpenGL Programming Guide*. 7th ed., Addison-Wesley Professional, 2009. Print.
2. “Overview.” *brlcad.org*. 16 July 2009.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Scaling Geometry in Operational Requirement-Based Casualty Assessment (ORCA) Using Transformation Matrices

by Eric Weaver

Contents

List of Figures

Abstract

Operational Requirement-based Casualty Assessment (ORCA) is a computer tool used by the U.S. Army Research Laboratory's (ARL) Survivability/Lethality Analysis Directorate (SLAD) to simulate and report the effect of different insults on a Soldier in the field. The model allows the user to insert a number of different parameters into the program such as fragment size, material, velocity, and location on the body that is hit. Given these parameters, ORCA will output injuries sustained by the Soldier, tissues hit, and different incapacitations experienced by the Soldier. The geometry of the modeled Soldier is programmed using a number of three-dimensional coordinates and 4x4 matrices that facilitate the translations and transformations of these points. This report discusses how these matrices can be used to scale the geometry of the three-dimensional anatomical representation when given the desired scaling factors. As the model is scaled, the same fragment will have different effects on different-sized anatomies and result in different injuries.

Acknowledgments

The author wishes to acknowledge the mentorship of Timothy Myers.

1. Introduction/Background

Protecting the warfighter on the battlefield is one of the major and most important challenges the U.S Army faces today. Assessing crew casualties as a result of injury is an extremely important part of improving the quality of protection for personnel. In the Survivability/Lethality Analysis Directorate (SLAD) of the U.S. Army Research Laboratory (ARL), a majority of this assessment and testing is done with the help of computer tools that can simulate the results of an injury to a warfighter. One of the main tools that perform this assessment is the Operational Requirement-based Casualty Assessment (ORCA). ORCA takes a number of different parameters that detail the shape, velocity, and weight of the fragment as well as the path of the fragment and returns injuries to the warfighter. ORCA calculates the AIS for a crew member that will signify the type and severity of the crew member's injuries. Using the underlying detailed injury description and knowledge of the warfighter's mission, an assessment can be made of the ability of the warfighter to complete the mission.

To simulate the crew member, ORCA models the body by dividing the male anatomy into a number of different horizontal cross sections numbered from 1 to 108. This representation of the anatomy was taken from a medical atlas by Eycleshymer and Schoemaker (*1*). This three-dimensional anatomical structure has a width of 535 mm, a depth of 250 mm, and a height of 1750 mm. The use of a coordinate system to describe the geometry allows for a type of grid space to be created to describe each different section of the anatomy. When looking at the anatomy face on, the coordinate system starts at the bottom right and back of the anatomy. The coordinate system is built with the x-axis going from right to left on the back side of the anatomy, the y-axis from back to front along the right-hand side of the anatomy, and the z-axis from bottom to top along the right. Each horizontal cross section is enclosed by a x-y plane, while a y-z plane is used to create columns to further specify sections on the anatomy. This grid creates a series of cubed shaped cells, measuring 5 mm on each dimension, which make up a database of over 124,000 cells. To move the model, the limbs must be translated or articulated in space around a given point. To do this, the tool uses 4x4 matrices to move the sections and rotate or translate the points. The problem discussed in this paper is how to use the same matrices used in the translation to scale the geometry of the computer model to any desired size (*1*).

2. Experiment/Calculations

2.1 Problem Solving/Coding

The first question that must be answered when trying to solve this problem is the question of how to scale figures in three-dimensional spaces. As mentioned previously, many transformations to three-dimensional coordinates are done using 4x4 transformation matrices. The same is true when trying to scale points and make objects larger or smaller (*2*). Figure 1 gives an example of how a three-dimensional point can be scaled using a scaling matrix.

$$S_v p = \begin{bmatrix} v_x & 0 & 0 & 0 \\ 0 & v_y & 0 & 0 \\ 0 & 0 & v_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} = \begin{bmatrix} v_x p_x \\ v_y p_y \\ v_z p_z \\ 1 \end{bmatrix}.$$

Figure 1. Scaling geometry.

In this figure Px, Py, and Pz are the three-dimensional points to be scaled. Using the matrix SvP, with scaling points Vx, Vy, Vz, which can be the same values or three unique values, the two matrices will be multiplied together resulting in a matrix containing the scaled point. This matrix is represented in figure 1 to the right of the equal sign. The end result is having each point being multiplied by the scaling factor corresponding to that point. When they are multiplied, the points will be moved through space into different locations. This will result in the figure being shrunk, stretched or pulled. Figure 2 is an example of the effects a scaling matrix can have on a cube if the right scaling factors are applied.

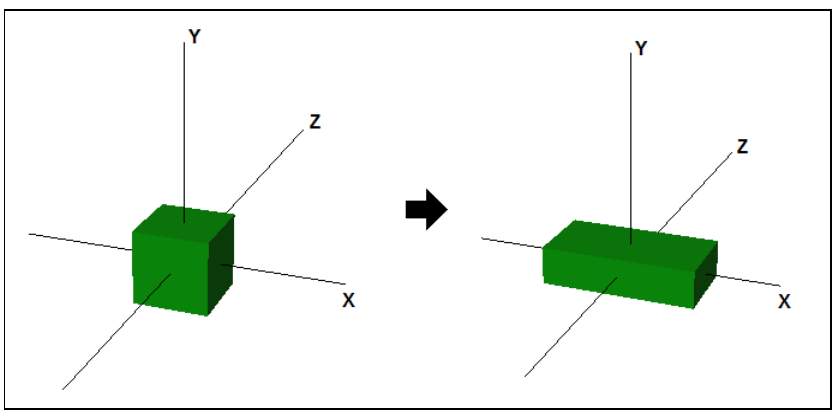


Figure 2. Example of a scaled cube.

While scaling a three-dimensional point is a simple problem of mathematics, figuring out where these points need to be scaled inside of the program code brings in a number of different problems related to computer science. The code for ORCA requires scaling in two different parts of the program. The first will be in the articulation where the form, size, and position of the model are created. The second location is in the anatomy, the portion of the code that returns the tissues and body parts that are damaged or affected by any particular fragment. To understand how the articulation must be changed to accommodate scaling, you must first look at how the articulation structure is defined in the program. In the ORCA source code, articulation is composed of three separate parts. The first is a variable labeled as *bb*. This is a variable of type bb_t, which in the code holds two arrays filled with the min and max points for the bounding boxes of each body region and section within the articulation. A bounding box is an invisible box surrounding a graphical object that determines its size (*3*). The other two parts of articulation are both arrays that hold 4x4 matrices for transforming each body region (variable t) and the inverse of those matrices (variable ti). Figure 3 shows the source code from ORCA that defines the articulation structure.

```
struct _articulation {
  bb_t bb;
  mat_t t[N_BR]; /* transformation matrix for each body region */
  mat_t ti[N_BR]; /* inverse of the same */
};
```

Figure 3. Definition of articulation structure in ORCA.

The next portion of the program that needs to be understood is the method that creates a new instance of articulation. In the source code, this method is called *articulation_new* and takes a pointer to an array of doubles as its parameter. After running this method, it will return a pointer to the location of this instance of articulation in memory. This method is used to calculate the min and max points for the arrays in *bb*. This makes it relatively simple to scale all the points in the newly created articulation since they are all calculated in the same portion of code. The *articulation_new* method starts with three different sets of triples (x, y, z coordinates) called *min_pt*, *max_pt*, and *pivot*. These three points are used as the basis for all of the other points created in the method. These points are calculated by using numbers and arrays that are defined in other parts of the ORCA source code. Since these three points are used later on in the method to calculate the bounding boxes of regions and sections, the points must be scaled immediately after their calculation to ensure that all other parts of the function use the scaled points and not the original points. To scale the original points, a new parameter must be added to the method. To make sure this doesn't interfere with other parts of the program that do not need to use a scaled articulation, we create a new method called *articulation_new_scaled*. This method still takes the pointer to an array of doubles; however, it will now take a series of three numbers, which will be the scaling factors for the x, y, and z coordinates. To use these scaling factors, another 4x4 matrix must first be created that is called *factor*. Once this matrix has been created, it is then initialized by filling it up with all zeros. Next, the matrix is filled up with the scaling factors that were passed into the method so that it is in the same form as figure 1. Before the values of the scaling factors are put into the factor matrix, a short check is done to make sure none of the scaling factors are equal to zero. If a scaling factor is equal zero, it makes all other coordinates on that axis equal to zero. To get around this, the check changes any factor of zero to a factor of one so that all coordinates on that axis will be unchanged. Once the factor matrix has been filled with the proper values, *min_pt*, *max_pt*, and *pivot* can then be multiplied by the newly formed transformation matrix. The multiplication is done with a function defined earlier in the code that multiplies a matrix and a point, when given the two as parameters. The function is called *mat_mul_pt*. This function returns a newly scaled point. After the original point has been passed to the function, it is reassigned as the value returned by *mat_mul_pt*.

As the function continues, it creates bounding boxes and positions for the body regions of the articulation. After the regions are created, the sections of those regions then must be created. This is done much the same way as the body regions, using numbers and information from different parts of the program code. Again as these points are created, they must be scaled using the scaling matrix that was created at the beginning. This time however, these points, which are stored in the variable *pt*, are compared against *min_pt* and *max_pt*, using the VMIN and VMAX functions in the code, respectively. What these functions do are compare each point from each axis from the two points passed in as parameters and select either the smallest of the two or largest of the two depending on the function. The second variable that is passed into the function takes on the value that is returned and is then stored into the articulation.

Once the articulation has been scaled, the program will successfully return the point or points where the articulation is hit based on a specific shot. This point is then used to return from the anatomy file the tissues that are hit. The problem with this is that the anatomy has not been scaled the same way that the articulation has, so the information returned will be inaccurate. To solve this problem, instead of scaling the entire anatomy, all we need to do is adjust the parameters that are passed into the *cellIndex_get* function. This function takes three different parameters. First is a pointer to a *CellIndex*, which is simply a section of the body, a row, and a column. Next is a variable *point*, which is the set of *x*, *y*, *z* coordinates; this is the point where the shot line intersects the articulation. The last parameter is an integer that tells the function what section of the body is hit. The results of this method can be scaled by simply scaling the point that was passed in as a parameter in the opposite direction of the articulation. The trouble with this approach is that the factor by which the articulation was scaled is only relevant inside of the *articulation_new_scaled* function and cannot be seen globally throughout the program. As a result, there must be a way for the *cellIndex_get* to access the scaling factor from articulation. To do this, the factor by which the points are being scaled must be stored inside the definition of the articulation structure. This variable can be called *factor* and will take a triple of *x*, *y*, and *z* coordinates. An example of the new structure definition is seen in figure 4.

```
struct _articulation {
  bb_t bb;
  mat_t t[N_BR]; /* transformation matrix for each body region */
  mat_t ti[N_BR]; /* inverse of the same */
  t3_t factor; /* scaling factor for the articulation */
};
```

Figure 4. New articulation definition.

Now that the scaling factors are stored into the *articulation*, the *articulation* must be passed into the *cellIndex_get* function so that it may retrieve the variable *factor*. Much like the *articulation_new* function, to make sure no parts of the code that do not require scaling are unaffected, a new function must be created called *cellIndex_get_scaled*. In this function, an extra parameter can be added that will take a pointer to a certain articulation. A function must also be added to the articulation that returns the *factor* variable so it can be used in *cellIndex_get_scaled*. The scaling factors can now be accessed; all that's left to do is apply the reciprocal of each scaling factor to the correct coordinate so that it will be scaled in the opposite direction. This can be done with a simple loop that will run through each coordinate of the intersection point as well as each scaling factor. This approach is used as opposed to matrix math, because there is no function in the program that will calculate inverse matrices.

2.2 Testing/Debugging

Whenever writing or changing any sort of code or program, one of the most important parts of the process is testing and debugging the code to make sure that it works properly and functions as intended. One of the easiest ways to check for this is by writing separate pieces of code called test cases. Test cases are defined as documentation-specifying inputs, predicted results, and a set of execution conditions for a test item (*4*). There are a number of different ways to write test cases to make sure that the code is working properly. Methods can be tested to see if they are equal to expected values or even unequal to values, if that is expected. They can also be tested to see if the results are different from the values of another method or the same method with different parameters. Test cases also include things like stress tests, regression tests, and risk-based test (*4*). In the case of the new methods that were created, there were a number of different tests written to make sure that everything was working as expected. The first was to test that the previously existing *articulation_new* function was still working correctly. For this, a test case had already been written so the code was simply modified to fit into the test framework that was provided. This test made sure that the articulation was filled with all the values that were expected and produced the expected results for an intersecting fragment. This test passed, which confirmed that the *articulation_new* method was unchanged. In the next test case, the *articulation_new_scaled* was fed all the same parameters, plus the required scaling factors and tested against the same values of the previous test. As expected this test failed because the articulation had been scaled and produced different values. The last test that was done on the articulation was to make sure that when two instances of articulation were created with different constructors, they would not have any equal values. This test also succeeded, because the two articulations were scaled to different sizes.

Another way to make sure code is working and functioning properly is by using the debugging mode that's available in many development environments. This mode allows one to run the code with break points where one can go step by step into and over functions and operations. This is extremely helpful, because one can view the value of each variable while progressing through the code. If the code is producing errors or the test cases are failing then debug mode can be used to find where exactly the problem is. Figure 5 shows an example of the *articulation_new_scaled* code being run in debug mode in Microsoft's Visual Studio. The break points are set to where the functions start. After being run for a number of steps and iterations, the environment allows one to see the values of different variables as well as their locations in memory.

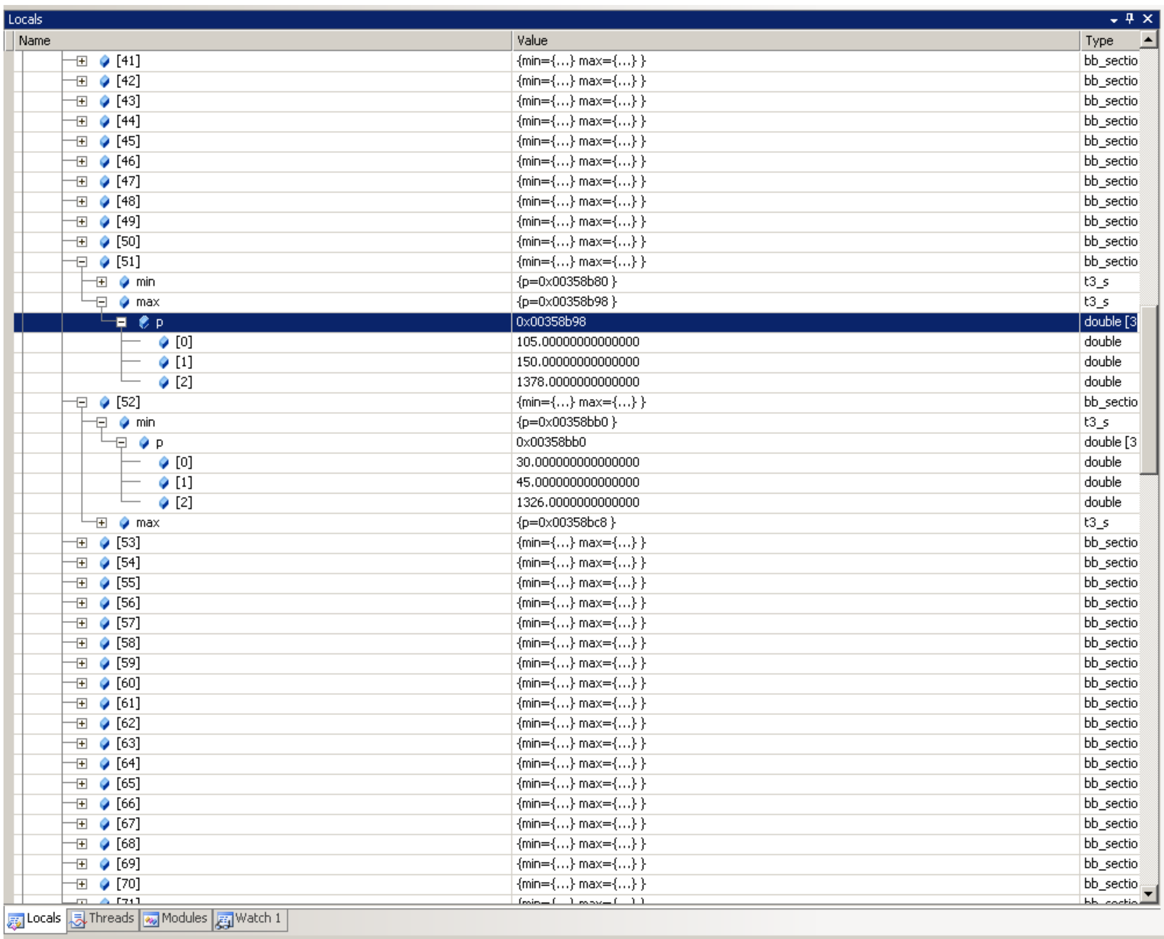


Figure 5. Visual Studio debug mode.

This debugging method was used with ORCA to make sure that all the points were properly scaled in the articulation. As each individual point was examined in the articulation using the pre-existing method, they were compared to the same point from a scaled articulation to make sure that they were in correct portion to each other. So, if a point in the normal articulation had coordinates of 30, 45, and 1326. The same point in the scaled articulation, that was given scale factors of 0.2, 0.2, and 0.2, would have coordinates of 6, 9, and 265.2.

3. Results and Discussion

As a result of the work done in this project, the ORCA model will now support the use of scaled geometry inside of the code. Users can now see the effects of a certain event on a number of different sized geometries if they wish, which could yield some very valuable information about the event. The ORCA model now contains three new functions that allow it to have the new functionality: the *articulation_new_scaled* function, which will scale the articulation; the *cellIndex_get_scaled* function, which will scale the intersection point of the shot line and the articulation; and lastly, a small function, which will return the factor variable from any given articulation. This project also required a slight adjustment to the definition of the *_articulation* data structure in the source code to allow it to store the scaling factors that are being used, if any. The new code that has been written for the program has been done in such a way that it will not affect the normal function of ORCA if the user chooses to use it in its traditional state. The addition of this functionality does not completely make these changes apparent to the user. The work of this project also allows for further changes to the ORCA model to do a number of different things that are outside the scope of this project, such as incorporating the new functionality into the user interface.

4. Summary and Conclusions

This project was done to allow ORCA to support scaled geometry to make it even more versatile than it was before. The new functionality of the program allows ORCA to assess warfighter survivability in a different way than before and expands the information that can be derived from testing. This project also showed that it doesn't always take an overwhelming amount of code to add to the functionality of a program, as long as the code is written correctly and used in the correct functions of a program. Hopefully, the new functions of the program will allow the program to be expanded further in the future so that ORCA can do even more with casualty assessments and geometries of different individuals in order to describe different targets. This project is also a good example of how older code can be added to and expanded upon to make it function quicker and more efficiently, and with more functionality with relative ease. Overall, this project provides a platform for more work on ORCA by providing it with a basic new functionality upon which to build.

5. References

1. Saucier, R.; Kash III, H. *ComputerMan Model Description*. U.S. Army Research Laboratory, August 1994.

2. Molhave, T. Transformations in 3D, October 21, 2000. http://home10.inet.tele.dk/moelhave/tutors/3d/transformations/transformations.html (accessed June 26, 2009).

3. All Business Staff. Business Definition for: Bounding Box, 2006. http://www.allbusiness.com/glossaries/bounding-box/4954964-1.html (accessed June 25, 2009).

4. Kaner, C. *What is a Good Test Case?* Florida Institute of Technology, Department of Computer Science, May 2003.

Vehicle Technology Directorate (VTD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Micro Torque Measurement Using a Flywheel

by Joshua Etchison

Contents

List of Figures

List of Tables

Abstract

Micro-robotic applications exist in many areas including surveillance and miniature simulators. These applications range from micro-aerial vehicles (MAV) to an air-speed simulator. To increase development of micro-robotics, it is imperative that the performance of the relevant micro-rotary motors is known. Having a device that can measure the performance of different micro-rotary motors is essential to increasing prototyping in the field of micro-robotics. One approach to this is a micro-dynamometer using a flywheel. The micro-dynamometer design presented spins a flywheel using a micro-rotary motor to calculate the torque, which can be calculated if the moment of inertia and angular acceleration are found. The micro-dynamometer is able to measure the performance of micro-scale motors quickly and easily to determine the use of a motor for a particular application.

Acknowledgments

The authors wish to acknowledge the mentorship of Justin Shumaker.

1. Introduction/Background

Engineers have been researching the design and use of micro-motors since the 1980s (*1*). Micro-motors are used in a broad assortment of products and industries such as audio and video equipment, security and instrumentation, medical and laboratory equipment, robotics and automation, toys, and in aerospace and defense (*5*). As these motors (figure 1) get smaller, the harder it is to measure their performance (*3*).

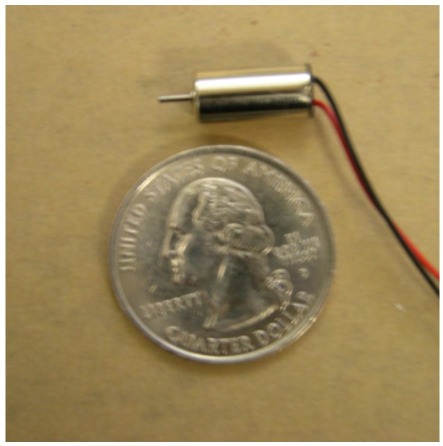


Figure 1. A 6 mm micro-rotary motor.

Since the use of micro-robotics has spread through research and industry, it is essential that the characteristics of the motors implemented are known. This need for micro-torque measurement is recognized throughout the world. Various methods of micro-torque measurement have been researched and used. An established technique of measuring torque is to use a cantilever and measure the force produced as a load is applied to the motor (*3*). Other methods to measure the torque of micro-motors are based on a cable brake principle, differential force between two sensors, and using wind pressure applied to a turbine (*1–3*).

The U.S. Army use for micro-motors range from applications used for surveillance, such as micro-aerial vehicles (MAVs), to simulation, such as the air speed simulator. One example is the concept of the flapping wing MAV (figure 2). The flapping wing MAV flies similar to an insect or bird. If equipped with a micro-camera, it can be used for situational awareness of the Soldier. To accomplish flight in this manor, the motor used must be very small and have enough torque to power the wing movement. This design will further surveillance technology in disguisable or unnoticeable devices.

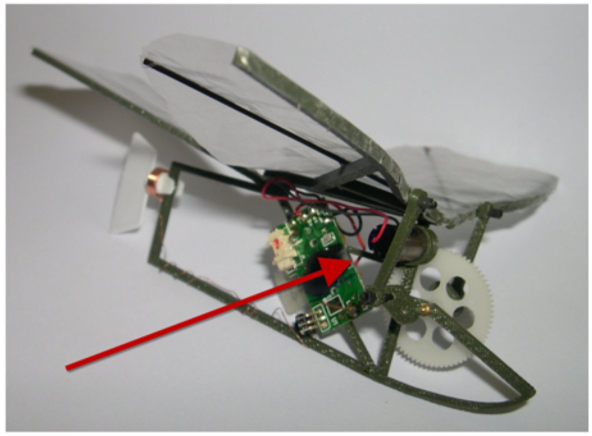


Figure 2. Flapping wing MAV.

The air speed simulator (figure 3) is another application that utilizes a micro-rotary motor. It is a hardware-in-the-loop simulator designed to reproduce airspeed of a MAV in a lab environment. This system will aid in the research to optimize performance of a MAV's autopilot. In order to appreciate these technologies, the performance of the motor must be known so analysis of functionality can be completed.

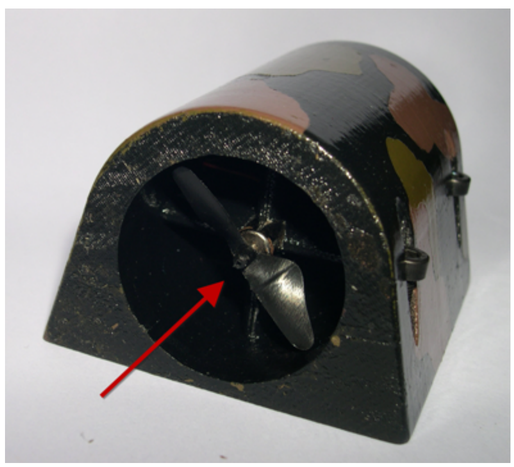


Figure 3. Airspeed simulator.

The flywheel initiated over 100 years ago. Originally, it was used to keep machinery running smoothly from one cycle to another (*4*). Instead of using the flywheel for stored energy, this concept counts the time taken for a flywheel to make each revolution until the flywheel reaches a terminal velocity. If the angular acceleration of the flywheel is found and the moment of inertia for the flywheel is known, then the torque can be calculated.

The ability to measure the performance of various micro-motors will aid in the development of micro-robotics for the U.S. Army. By proving this concept and building a working system, the micro-robotics team and many others within the U.S. Army Research Laboratory (ARL) will be able to characterize multiple micro-motors. This will save valuable time, as opposed to using a motor that is not suitable for an application and having to retreat to the drawing board.

2. Experiment/Calculations

2.1 Summary of Primary Functionality

The system discussed allows a user to power a micro-rotary motor through an external power supply and gather various details regarding the performance of the motor via universal serial bus (USB) to a computer. To begin the process of measuring the performance of a micro-motor, all the connectors need to be plugged into the microcontroller. This includes the light emitting diode (LED), photocell, micro-motor, and the power supply for the motor. Second, the microcontroller needs to be connected to a computer using the HyperTerminal or Minicom. The microcontroller will display "System Ready" on the screen. Then, the user pushes the start button; this will turn the motor on and start to spin the flywheel. Once the microcontroller has acquired all the data points it will turn the motor off and display the performance details. These details can be saved or copied and pasted into a spreadsheet if the user wishes to graph the data. If the user desires to run the motor more than once, all they must do is press the start button again.

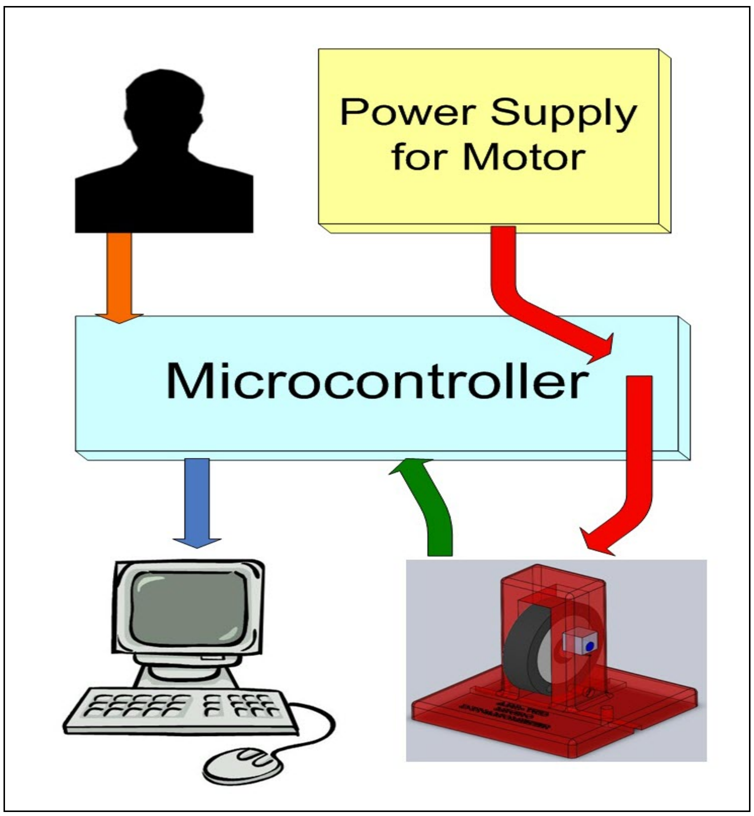


Figure 4. Block model of system.

The micro-dynamometer (figure 5) is made up of various components. It utilizes a geared flywheel, with two holes through it, and two washers, mounted to add weight. A motor cartridge houses the micro-motor undergoing the testing. The motor has a pinion that meshes with the gear on the flywheel. A bright LED and photocell are mounted pointing at each other. As the flywheel spins, the holes through the flywheel will allow the light emitted from the LED to be captured by the photocell. A voltage signal from the photocell will change as a response to the difference in light captured. The change in voltage is recognized as an interrupt to the microcontroller. The microcontroller will then count and store the time it takes the motor to spin the flywheel. The acceleration is calculated allowing the microcontroller to then calculate the torque.

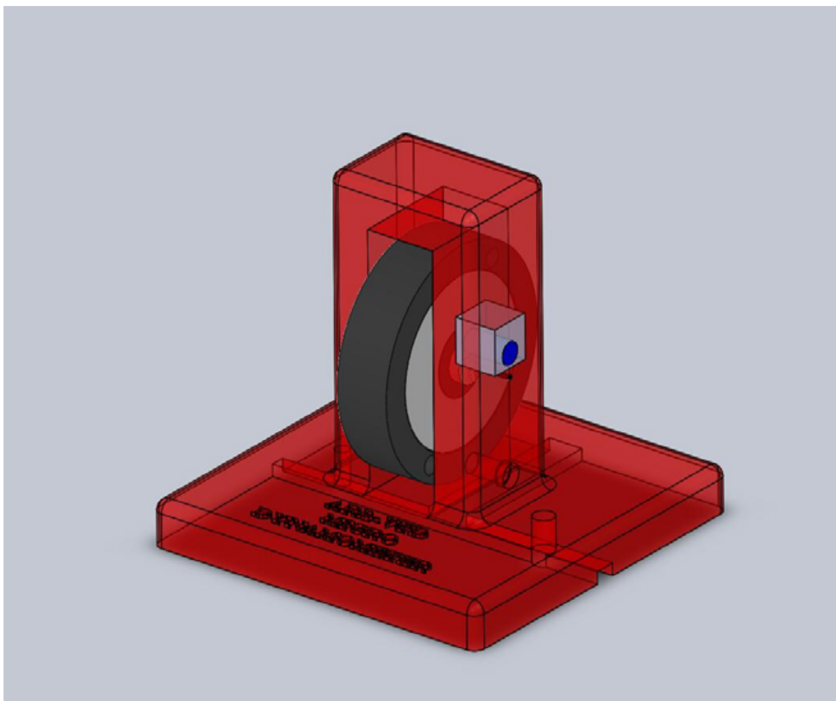


Figure 5. Solidworks model of micro dynamometer.

2.2 Component Descriptions

2.2.1 Micro Dynamometer Housing

The micro-dynamometer's housing (figure 6) was designed on Solidworks and printed on a 3-D printer. The base of the housing is 100 mm x 100 mm x 12 mm. The tower of the housing is 44 mm x 30 mm x 80 mm with a 20 mm x 68 mm gap for the flywheel. There is a square hole on one side of the tower for the motor cartridge. Near the base of the tower is a hole on both sides; one is for the bright LED and the other is for the photocell. There are two holes in the base to allow the wiring of the LED and photocell to be routed underneath the dynamometer.

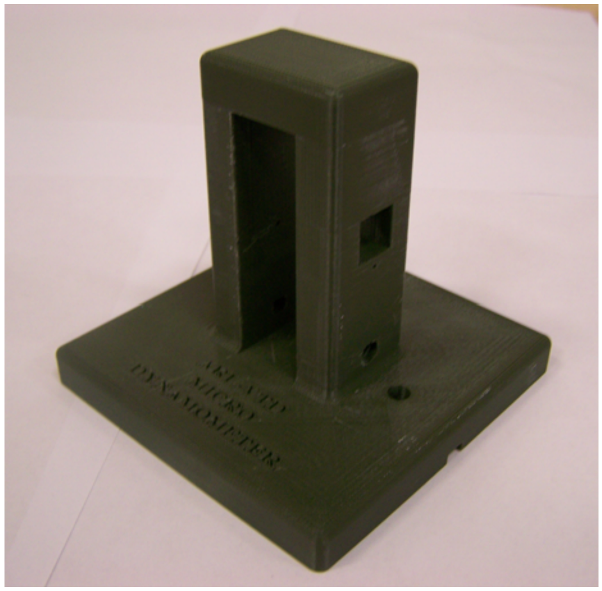


Figure 6. Actual micro dynamometer housing.

2.2.2 Flywheel, Gear, and Pinion

The flywheels were designed in Solidworks and printed on a 3-D printer. There are two 5 mm holes evenly spaced apart, that allow the photocell to detect the light emitted from the LED every half of a revolution. The gear and pinion are from an radio-controlled (RC) car package. There is a 23:6 gear to pinion ratio. The gear is glued onto the flywheel using atmospheric-pressure (AP) adhesive and the pinion slides onto the micro-motor’s shaft. Three flywheels were tested (figure 7). The diameters of the three flywheels are 60 mm; the difference is the width of the flywheel. This change affected the mass of the flywheel. The phase 3 flywheel yielded the best results. This was found through trial and error, based on the amount of time taken for the flywheel to speed up.

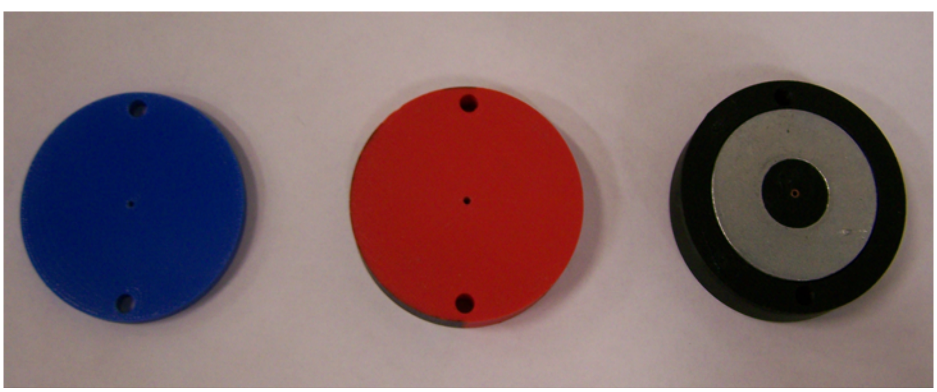


Figure 7. Tested flywheels.

Table 1. Description of flywheels tested.

Phase	1 (Blue)	2 (Red)	3 (Black)
Thickness	5.25 mm	12.5 mm	15.25 mm
Mass	11 gm	23.5 gm	86.8 gm

The heterogeneous flywheel (figure 8) has two 28.4 gram washers recessed in the flywheel, one on each side, to increase the mass of the flywheel without increasing the dimensions excessively.

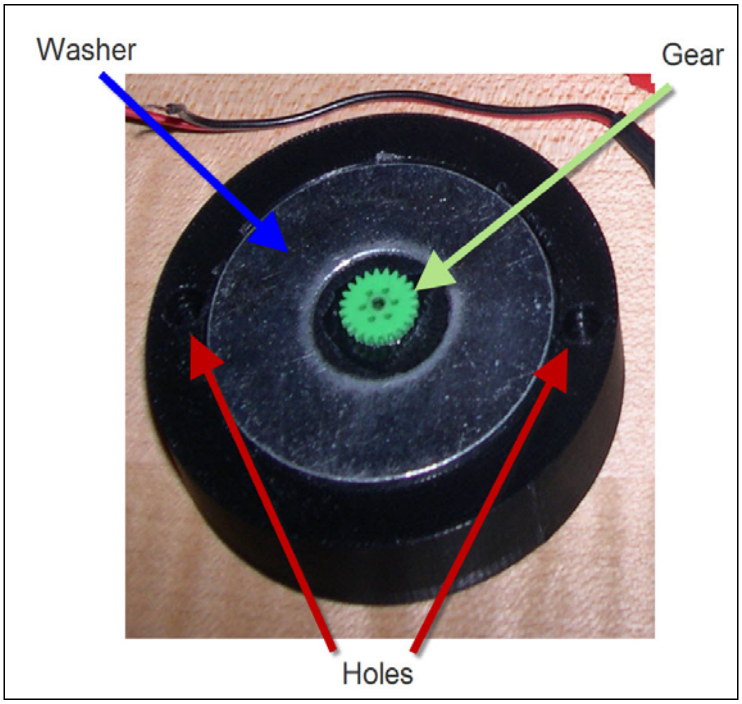


Figure 8. Heterogeneous flywheel.

2.2.3 Motor Cartridge

The motor cartridge (figure 9) design allows almost any micro-motor, within the size of 7 mm to 2 mm, to be characterized. The motor cartridge was also designed in Solidworks. It is a 12 mm cube. The only task needed to prepare a new cartridge is to measure the diameter of the motor desired, and change the diameter of the hole in the Solidworks model. Once this change has been made, the part can be printed on the 3-D printer.

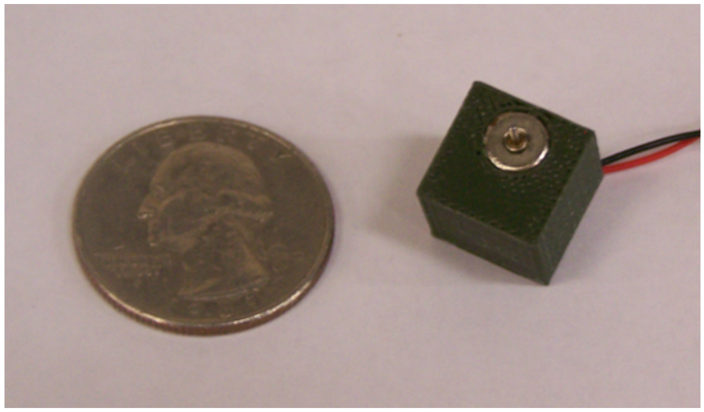


Figure 9. A motor cartridge with 6 mm micro-motor.

2.2.4 Microcontroller

Software for the microcontroller was developed in Linux and is programmed in C language. It uses internal timers to count the rotation time of the flywheel. An interrupt occurs every time the light from the LED is detected by the photocell. Since the microcontroller is triggered twice per revolution of the flywheel, a toggle is enabled in the software to only count every other interrupt.

The print circuit board is 65.5 mm x 52 mm and uses the ARM7 microprocessor. The microcontroller has two USB ports, one for programming and the other for communication. It receives power from the USB on either port. It has a red LED, which turns on when the board is powered, and a blue LED that can be used for status or otherwise programmed. The microcontroller uses one push-button to start the process of performance measurement, and it uses the other push-button to reset the memory for programming. The board has four connectors: External Power (for the motor), Motor, LED, and Photocell. The external power is necessary to allow the user to set the voltage to the motor. The motor is controlled by a field effect transistor (FET). As a byproduct of this design, the use of the FET will also allow control of the motor using pulse width modulation (PWM) for performance measurement at a controlled speed.

The signal from the photocell first goes through a voltage divider, then to an inverting rail to rail op-amp, then inverted again through a 74LS14 chip. This is necessary because the response time of the op-amp, measured at 60 µs using an oscilloscope, was too long and caused the microprocessor to trigger twice on every interrupt. The response time of the 74LS14 was measured to be about 15 µs. By inverting through the 74LS14, the response time was reduce by 45 µs, which corrected the problem of the subsequent triggering.

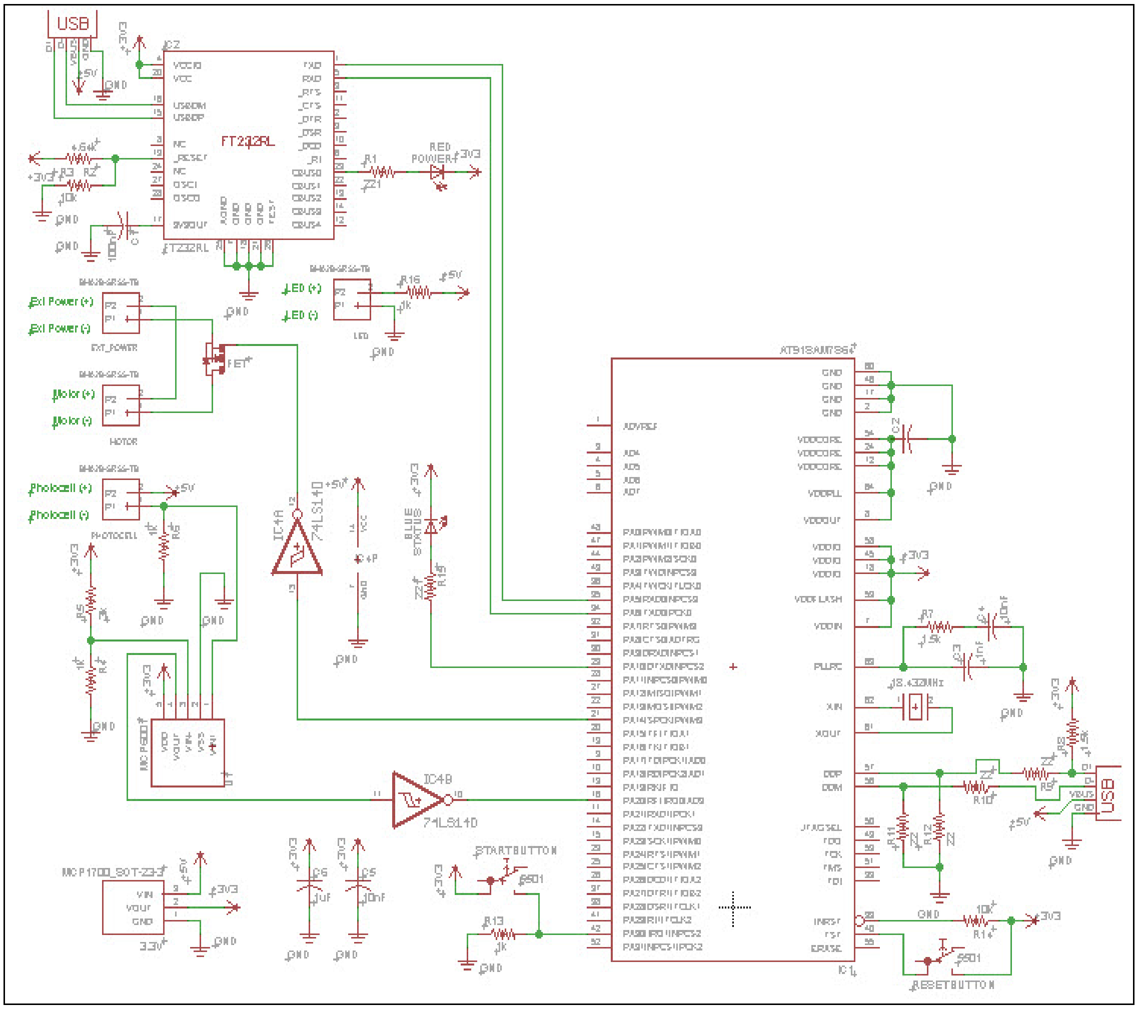


Figure 10. Microcontroller schematic.

2.3 Testing Phase

Testing the concept and design of the micro-dynamometer was done using an AT91SAM7 development board and selected circuitry plugged into a bread board. The motor used for proof of concept was a 6 mm diameter micro-rotary motor. The first dynamometer housing used a clamp to secure the motor. However, this design is not versatile to multiple micro-motors.

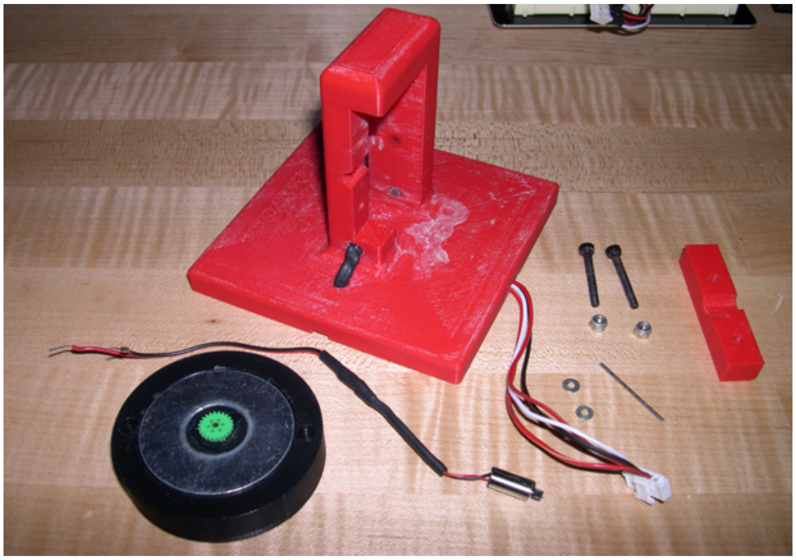


Figure 11. Disassembled dynamometer used for testing.

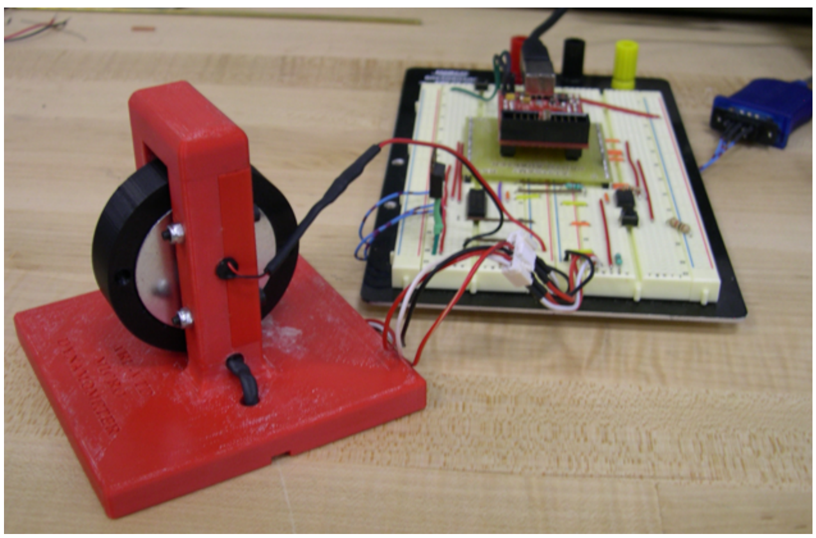


Figure 12. Test dynamometer and circuitry.

2.4 Calculations and Method

The performance measurement of torque, *T*, can be calculated using the moment of inertia of the flywheel, *I*, and the angular acceleration, α, with the formula:

$$T\ (\mu N\ m) = \alpha\ (radians\ /\ ms^2) * I\ (gm\ mm^2)$$

If the flywheel was solid, the moment of inertia could be calculated using the formula:

$$I\ (gm\ mm^2) = \tfrac{1}{2} * m\ (gm) * r^2\ (mm)$$

However, since the flywheel is not a solid cylinder, the mass and materials were entered in Solidworks and the software calculated the moment of inertia for us. The software also gave us the calculated mass, volume, and surface area. We compared the mass calculated by Solidworks to the actual mass measured on a scale accurate to a tenth of a gram. There difference in the mass measured and the mass calculated from Solidworks was about 0.45 g; therefore, we believed the moment of inertia calculated by Solidworks is accurate.

Table 2. Values from Solidworks compared to the measured mass of phase 3 flywheel.

Calculated in Solidworks	$I = 29712.29$ gm mm²
	$m = 86.35$ gm
	$v = 41155.45$ mm³
	$a = 16451.21$ mm²
Measured mass	$m = 86.8$ gm

In order to find the angular acceleration of the flywheel, various calculations need to be made. First, the microcontroller must count the time from one revolution to another—*rotation time (ms)*. As the motor speeds up, the microcontroller stores the rotation time into an array. After the desired amount of data points are acquired, the microcontroller will process the array. Second, the revolutions per second are calculated using the formula below and stored into another array.

$$RPS\ (/ms) = gear\ ratio\ /\ rotation\ time\ (ms) \tag{1}$$

where, *gear ratio* is 23:6 or 3.8333. Third, the change in *RPS*, *ΔRPS*, is calculated using the formula:

$$\Delta RPS\ (/ms) = RPS_{K+1} - RPS_K \tag{2}$$

Then, the angular acceleration is then calculated using the following formula.

$$\alpha\ (radians\ /\ ms^2) = (\Delta RPS\ (/ms)\ /\ rotation\ time\ (ms)) * 2\pi \tag{3}$$

3. Results and Discussion

The microcontroller output the flywheel's rotation time,, the motor's revolutions per millisecond (*RPS*), angular acceleration of the flywheel (α), and the torque (*T*) in columns to Minicom. The data was then saved and opened in a spreadsheet for analysis. The plot of the torque curve (figure 13) was not as expected. In order to locate the cause of the unexpected data, the rotation time, revolutions per second, and the angular acceleration were plotted for analysis.

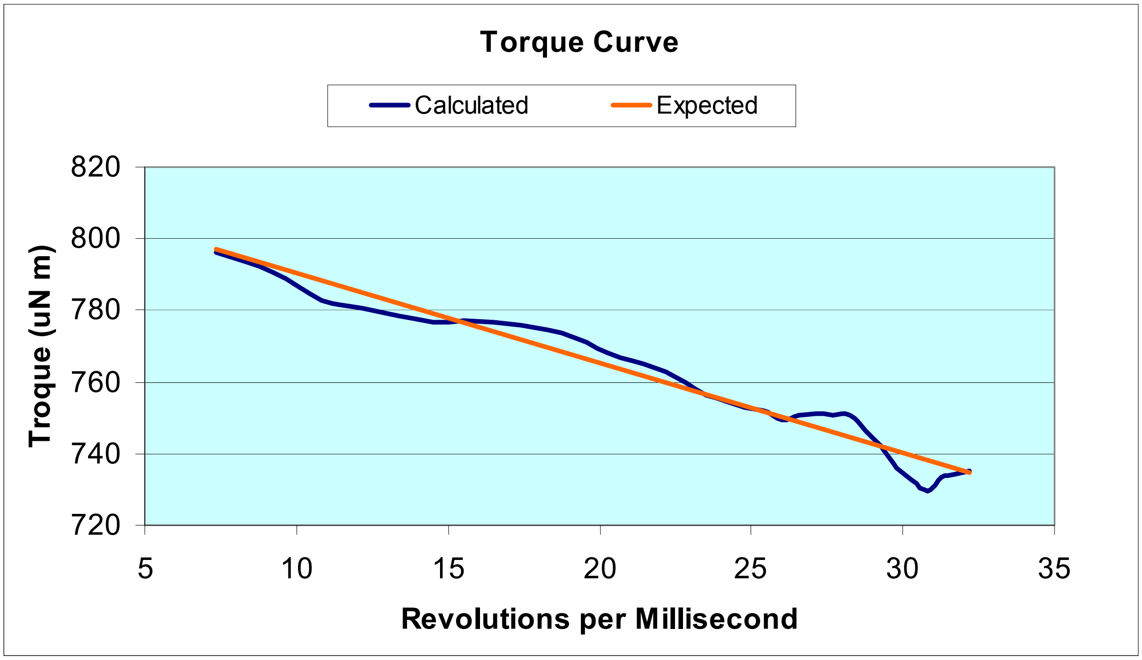


Figure 13. Graph of instantaneous torque curves.

The plot of the rotation time (figure 14) was as anticipated. It displays the time it takes the flywheel to make each revolution, for the first 32 revolutions. The graph illustrates how the flywheel speeds up quickly and begins to reach a terminal velocity.

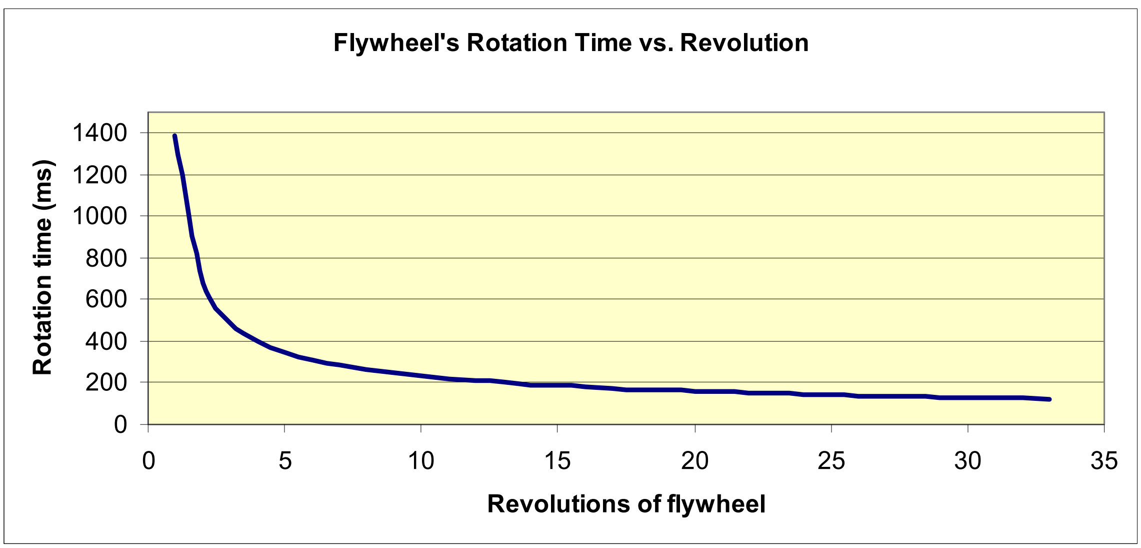


Figure 14. Graph of flywheel's instantaneous rotation time vs. revolutions.

The plot of the revolutions per second (figure 15) was also as expected. The curve is a smooth gradual change. It illustrates the acceleration of the motor.

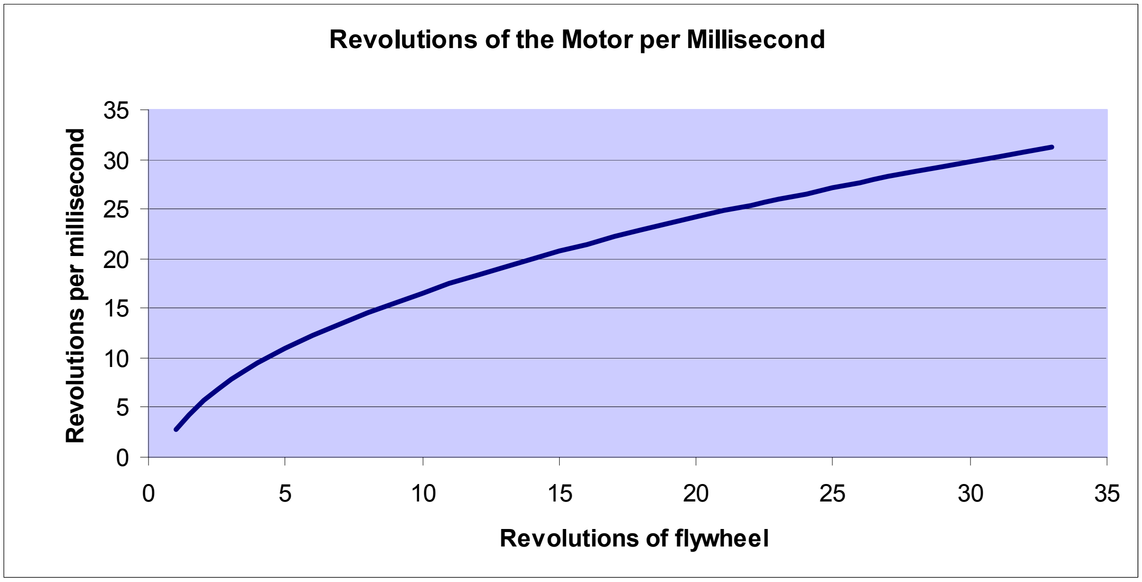


Figure 15. Graph of instantaneous revolutions per millisecond of motor vs. revolutions of flywheel.

The plot of the angular acceleration (figure 16) displayed the characteristics of the unforeseen torque curve.

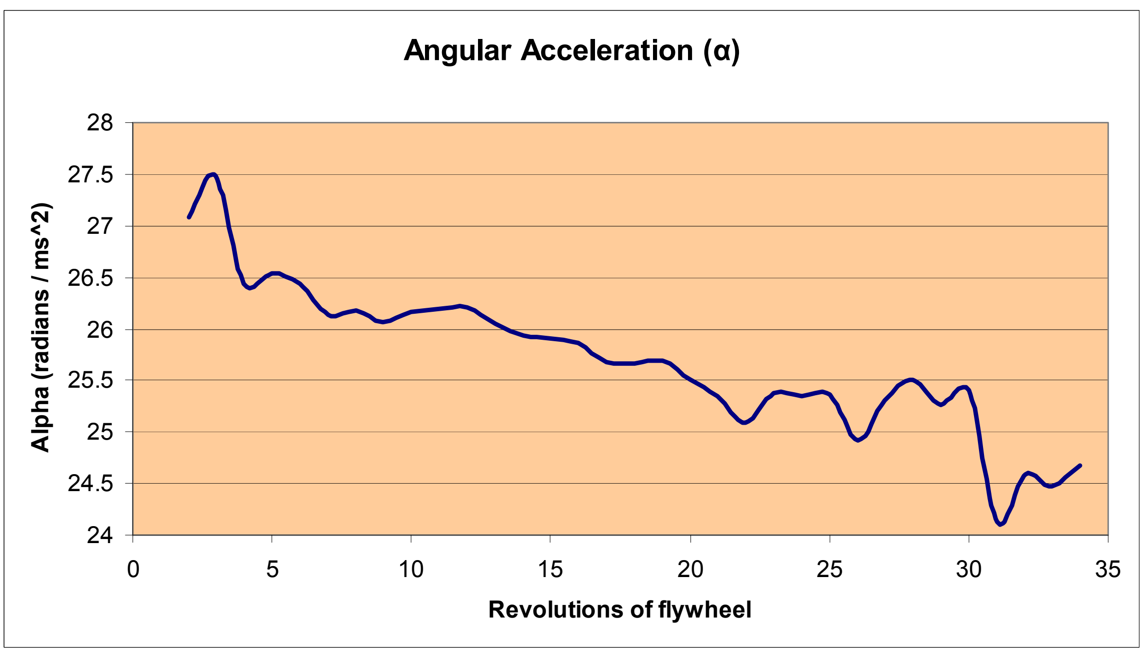


Figure 16. Graph of instantaneous angular acceleration of flywheel.

After analysis of the previous three graphs, the following was observed: the angular acceleration maintained a steady change until the elbow of the curve on the graph of the flywheel rotation time vs. flywheel revolutions (figure 14). Once the change in rotation time between flywheel revolutions begins to level, the angular acceleration becomes magnified with error. Further analysis is currently being conducted on how the change in mass of the flywheel will affect the results. In theory, by increasing the mass of the flywheel, efficient and accurate measurements of motor performance can be acquired. The current phase of this research is to find the appropriate amount of mass needed in order to obtain accurate data.

4. Summary and Conclusions

This study has shown that the method of using a flywheel to measure torque in the micro–Newton-meter range is possible. Although more research is needed on the mass of flywheel and how it affects torque, this dynamometer is capable of measuring the performance of multiple micro-rotary motors. Once completed, it has the ability to measure the performance of multiple micro-motors to determine the use of a motor for a particular application. For example, it can prove that a 4 mm motor is sufficient enough to power a flapping wing MAV, or that a 6 mm motor is necessary to get the flight speed required for flight.

5. References

1. Brenner, Werner; Vujanic, Aleksander; Popovic, Gordana. Micro Torque Measurement Based on the Cable Brake Principle. *Sensors, 2002. Proceedings of IEEE 61.1* **2002**, 936–941.

2. Gass, V.; van der Schoot, B. H.; Jeanneret, S.; de Rooij, N. F. Micro Torque Sensor Based on Differential Force Measurement. *Micro Electro Mechanical Systems, MEMS '94, Proceedings, IEEE Workshop* **1994**, 241–244.

3. Ota, H.; Ohara, T.; Karata, Y.; Nakasima, S.; Takeda, M. Novel Micro Torque Measurement Method for Microdevices. *Journal of Micromechanics and Microengineering Institute of Physics Publishing* **2001**, 595–602.

4. Bitterly, Jack G. Flywheel Technology Past, Present, and 21st Century Projections. *IEEE AES Systems Magazine May 1998*, 13–16.

5. Micro Dyne Motor Testing System Data Sheet. http://www.magtrol.com/datasheets/microdyne.pdf (accessed 2009).

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory
SUMMER RESEARCH TECHNICAL REPORT

Target Geo-location Acquisition from MAV Using Video Imagery: A Preliminary Study

by Ivan Walker

Contents

List of Figures

Abstract

The U.S. Army is constantly working to improve the situational awareness of the Soldier in combat. Unmanned aerial vehicle (UAV) technology has emerged as a major research topic in that effort. This working paper presents a foundation for a system that can be used with fixed-wing micro-aerial vehicles (MAV) to extract target geo-location information quickly and accurately from the video stream. Given accurate attitude, altitude and global positioning satellite (GPS) data, the triangulation of a target's geo-location is well-documented in the literature. The accuracy of such geo-location relies directly on the accuracy of GPS and attitude data and its synchronicity with the video stream. In this research, we propose to synchronize attitude and altitude data with the video stream using a gimbaled platform. Here, the video camera and the Inertial Measuring Unit (IMU) are rocked, generating two streams of data that can be used to find the latency between the IMU and video data. This will allow the system to use the exact attitude data associated with each video frame. GPS is synchronized by flying the camera over targets of known GPS coordinates. The latency of the GPS stream is therefore found by comparing the calculated GPS values to those of the targets.

Acknowledgments

The author wishes to acknowledge the mentorship of Justin Shumaker

.

1. Introduction/Background

Unmanned air systems are prime candidates for tasks involving risk and repetition, or what the military calls the "dull, dirty, and dangerous" (Office of the Secretary of Defense, 2002). For tasks that involve tracking, reconnaissance, and delivery, one objective of unmanned air systems is to accurately determine the location of ground-based objects (*11*).

This paper presents a method for determining the location of objects in world/inertial coordinates using a fixed camera onboard a fixed-wing miniature air vehicle (MAV). We focus on fixed-wing unmanned aerial vehicles (UAV) (as opposed to rotary wing aircraft or blimps) due to the unique benefits available from fixed-wing aircraft, including adaptability to adverse weather, enhanced fuel efficiency, a shorter learning curve for the untrained operator, and extreme durability in harsh environments. Also, minimum airspeed requirements associated with fixed-wing aircraft can provide images from multiple vantage points, allowing for more robust localization.

The acquisition of a target's geo-location from video imagery is well-defined when altitude and attitude of the video camera are known. For larger UAVs, the system works well as these UAVs are rather stable. However, as the airframe gets smaller, the system becomes more vulnerable to weather perturbations. Such perturbations can cause large increases in geo-location errors. This is because the error is highly dependent on the rate of change of attitude angles. The slightest shift in Inertial Measuring Unit (IMU), global positionings satellite (GPS) or video data may cause large errors, especially when the velocity component of the pitch, roll, or yaw is sizable. We, therefore, propose to synchronize the video imagery with IMU data through the use of the gimbaled platform with optical shaft encoders (see figure 9). Our previous work has shown that an IMU can be calibrated using this platform to reduce errors by as much as 30%. Also, it has been shown that this platform can synchronize autopilot systems at the millisecond range. Over and above the calibration and synchronization of the autopilot and video, we propose to synchronize the GPS data stream with the IMU and video data. This will have to be done through actual flight of the system. To further improve geo-location accuracy, we will freeze the selected video frame, allowing for accurate targeting. Upon selection, the system will generate the GPS coordinates for the target. We believe that this system will make it possible to extract accurate target geo-location data from the smallest of miniature aerial vehicles (MAVs).

Smaller MAVs are susceptible to wind perturbations, causing a jittery image and a sizable rate of change on all IMU values. The slightest latency between IMU, GPS, and video data can then cause massive errors in target geo-location. The system we propose will allow for accurate synchronization of these values, greatly reducing errors in geo-location.

The proposed system will use a semi-circular rigid graduated strip (figure 1). This strip will be placed around the platform (figure 2). The platform will move in a triangular fashion, and data from the IMU, Optical Shaft Encoders (OSE), and the corresponding video frames will be collected. The angular deflection of the camera will be extracted from the images and will be plotted against the IMU and encoder data. The IMU and video data will be synchronized using the procedure described in (*2*), namely by finding the time shift that will minimize errors between the two data streams.

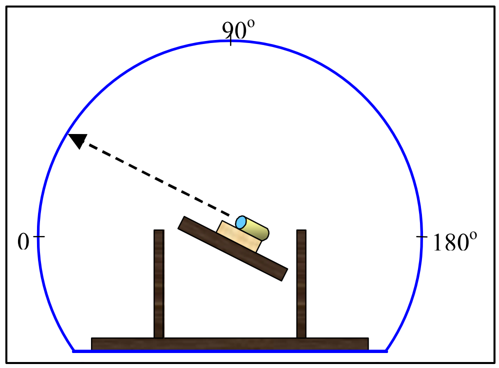


Figure 1. Synchronization setup with graduated strip.

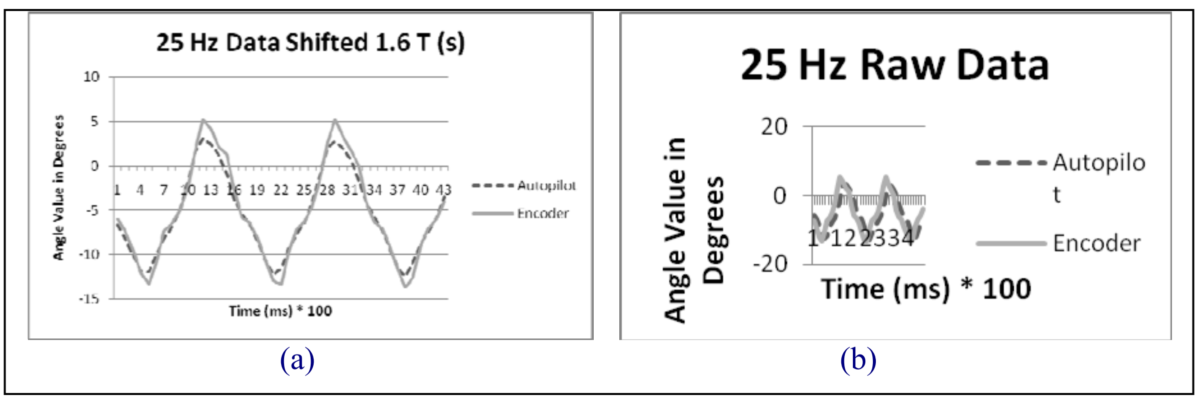


Figure 2. IMU data (a) before and (b) after synchronization.

Once the autopilot and camera are calibrated and synchronized, the system's performance will be verified through real flight. An evaluation of the improvement in accuracy will be conducted (*7*).

1.1 Project Description

1.1.1 IMU Calibration

Previous results have shown that by using a two-degrees-of-freedom gimbaled platform, as shown in figure 2, an IMU can be calibrated (*2*). Effects of this calibration were found to reduce errors in IMU results by over 30% (*1*). IMU errors can be further reduced by using this platform and introducing filters such as Neural Networks and Extended Kalman filters.

1.1.2 IMU and Video Synchronization

IMU and video can be synchronized to the optical shaft encoders of the gimbaled platform. However, GPS will likely be calibrated through actual flight, or through the use of GPS simulation equipment, such as that produced by CAST Navigation. Figure 3 shows IMU and shaft encoder signals before and after synchronization. This method of synchronization has shown that errors can be significantly reduced. Note that IMU is not fully calibrated in the figures 2a and 2b (*2, 3, 4*).

Since the latency between IMU signals *(M)* and the shaft encoder signals *(S)* can be of multiple sampling periods *T*, the first step is to shift IMU data by an integral number of periods until the least amount of error is achieved. The error *(ε)* is defined as the sum of absolute difference between IMU (M_i) and shaft encoder (S_i) data.

$$\varepsilon = \sum_{i=0}^{n} \left| M_i - S_i \right| \tag{1}$$

Synchronization calculations are based on the assumption that the change in pitch and roll information between two consecutive data points is linear. Therefore, if M_i and S_i are the IMU and shaft encoder's i^{th} sample, then $\varepsilon\ (g)$ is the average error in shaft encoder reading versus that of the IMU, where *g* represents the time shift for a fraction of the period *T*. Latency can then be defined as the time *g* for which $\varepsilon\ (g)$ is a minima, where

$$\varepsilon(g) = \{\sum_{i=0}^{n} \left| Mi - ((g+1)S_{i-1} - gS_i) \right| \} / n \tag{2}$$

Using a similar procedure, the angles seen by the video camera will be plotted and synchronized with the shaft encoders. The IMU and video will, therefore, be synchronized.

1.1.3 GPS Synchronization

The simplest and most cost-effective way to synchronize GPS is through real flight. A radio-controlled drone will be used to fly the camera system over targets of known GPS coordinates. Optimal drone GPS values will then be calculated and compared to the measured GPS values. This will allow for the calculation of GPS latency, hence, synchronizing GPS with the IMU and video systems.

1.1.4 Target Triangulation

To obtain the triangulation equations needed to calculate target location, we will assume that:

1. Earth is flat.
2. Exact attitude, altitude, and geo-location values of MAV are known.
3. IMU Video and GPS data are synchronized.

4. Camera viewing angles are small.

We start by assuming that a camera is located in the MAV's center of gravity, pointing downwards. Given that the video images generated by the MAV's camera are represented by $2m$ x $2n$ pixels, with a camera viewing angle of magnitudes $2\omega_x$ and $2\omega_y$, the viewing angle goes from $-\omega$ to $+\omega$, where the pixels run from –m and –n to +m and +n. Each pixel will, therefore, have an angular representation of ω_x/m and ω_y/n. For the pixel $P_{i,j,}$ the x and y vectors can be calculated using equations 3 and 4.

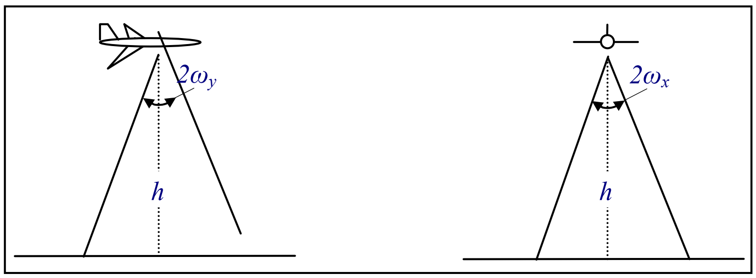


Figure 3. Level flight.

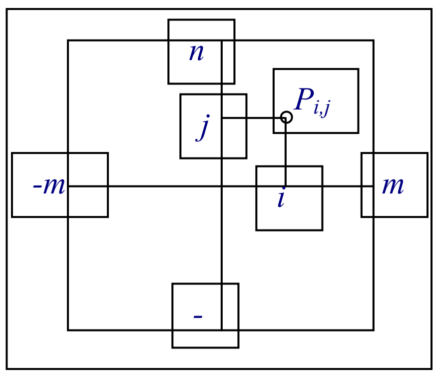


Figure 4. Video frame.

For small camera viewing angles, distortion can be ignored, and it can then be assumed that each pixel represents the same area on the ground. With the MAV at a height h, the dimensions of the viewed area are:

$$x = h * \tan([\omega_x / m] * i) \quad (3)$$

$$y = h * \tan([\omega_y / n] * j) \quad (4)$$

For a frame of m by n pixels, each pixel will represent a space of $(x/m,\ y/n)$ on the ground. Therefore, a pixel $P_{i,j}$ will have the coordinates $(ix/m,\ jy/n)$, as measured from the center.

1.1.5 Effect of Pitch, Roll and Yaw

A non-zero pitch angle θ and roll angle α, are added directly to the pixel's angles, resulting in the new set of x and y vectors calculated as follows:

$$x = h * \tan([\omega_x / m] * i + \alpha) \quad (5)$$

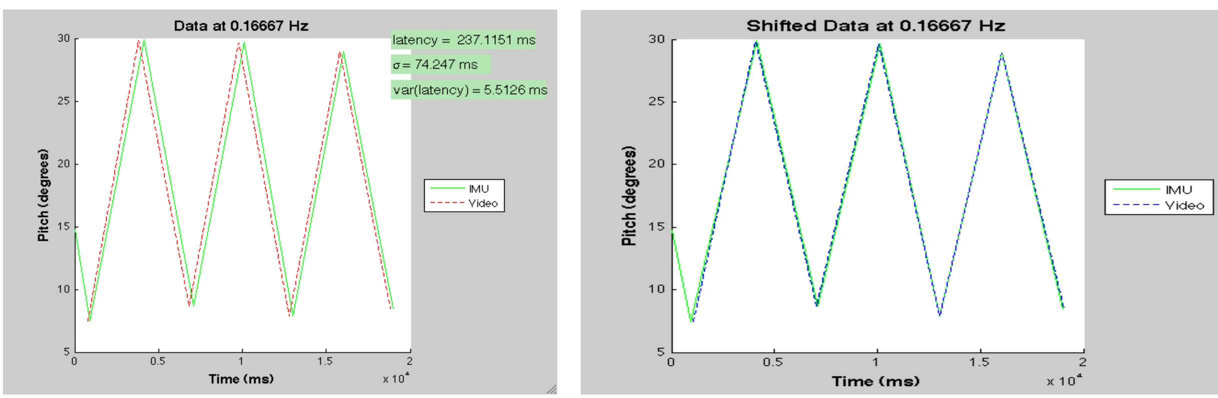
$$y = h * \tan([\omega_y / n] * j + \theta) \quad (6)$$

The yaw angle φ, on the other hand, rotates the vector from the origin to the pixel *P(i, j)* by φ degrees (*6, 11*).

2. Experimental Results

2.1 Video IMU Synchronization

To measure the latency between video and IMU data streams, both the camera and IMU were made to move together, generating an identical waveform. This waveform could then be used to measure latency. Generation of the waveform was achieved by mounting the camera and autopilot onto the gimbaled platform. The platform is capable of computer-controlled motion in two degrees of freedom. The platform was then programmed to move clockwise and counter-clockwise, generating the distinctive triangular waveform. Both the camera and IMU transmit their data via radio to the ground station. The ground station time-stamps the data and stores it for analysis at a later time. This experiment was repeated with the platform rocking at different frequencies, while maintaining the triangular waveform. The data in figures 5 and 6 shows two of the frequencies used. As demonstrated by the figures, latency can be measured more accurately at higher frequencies, namely 1.25 Hz versus 1/6 Hz.



(a) (b)

Figure 5. IMU/video data (a) before and (b) after synchronization.

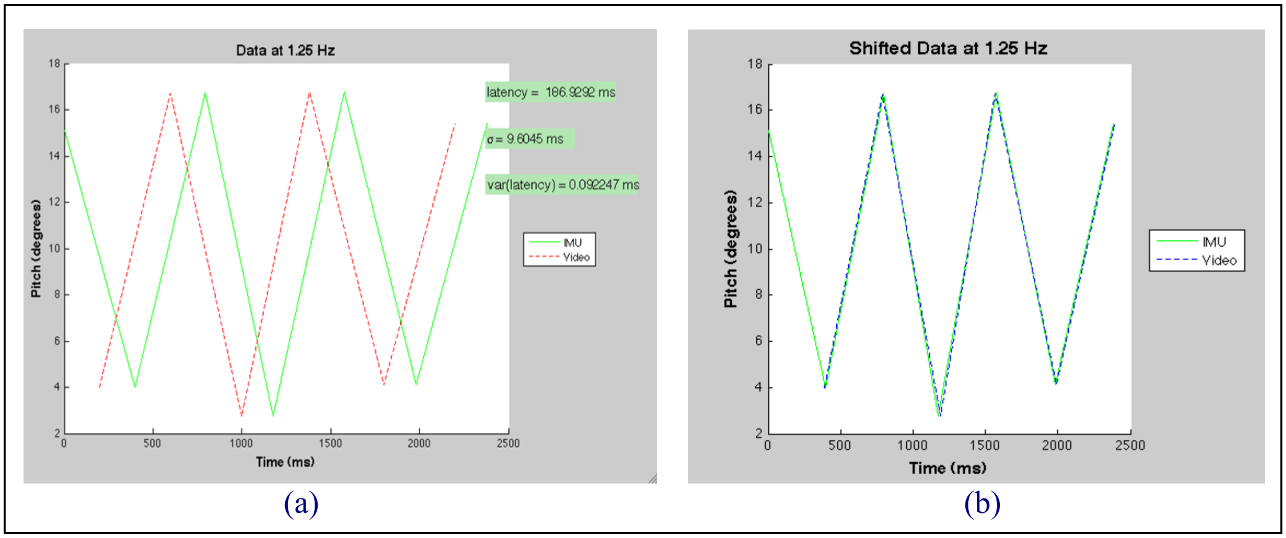


Figure 6. IMU/video data (a) before and (b) after synchronization.

Data from the video stream was handled differently from that of the IMU. The video stream was searched for frames at the peaks and troughs of the triangular wave. The time-stamp for these frames was used to plot the peak/trough point. Because the video stream was arriving at a much higher rate than that at which the platform was moving, finding an actual peak could be done with sufficient accuracy.

The worst-case scenario for peak value error would be when the platform is moving at the highest frequency. In this case, the error is estimated as follows: for a platform frequency of 1.25 Hz, each side of the triangle will cover 0.4 s. With the video streaming at 30 Hz, 0.4 * 30 or 12 points populates each line (triangle side). The maximum error occurs when a peak falls midway between two frames. The maximum error in peak is then 100/(12 * 2), or 4.17%. As the platform frequency decreases, so does the error associated with the peak.

The error calculations for the IMU data stream arriving at 8Hz were more significant (over 15%) than the video stream. Due to the increase in error, an algorithm was developed to analyze IMU data. Since it is known that the platform was moving at a constant angular frequency (the same angular frequency both clockwise and counter-clockwise), it is possible to fit points on the same side of a triangular wave by straight line, and points of intersection of these lines will form the peaks of the waveform. To determine if a point belongs to a particular line group, the slope formed by this point and the one before it is calculated. For a point to remain in the same linear group, the slope sign should remain unchanged, and the value of the slope should be similar to the previously calculated slope (within 20%). Should the slope sign or value change then the point in consideration belongs to the next linear group, and a new linear group is started (see figure 7).

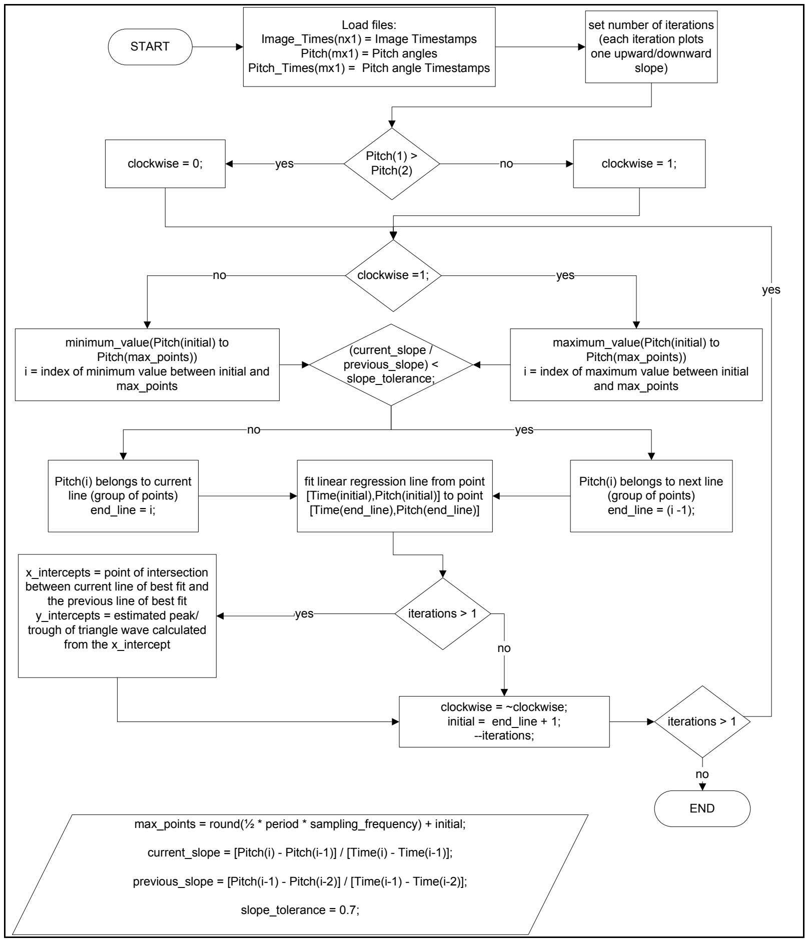


Figure 7. Flowchart of data analysis algorithm.

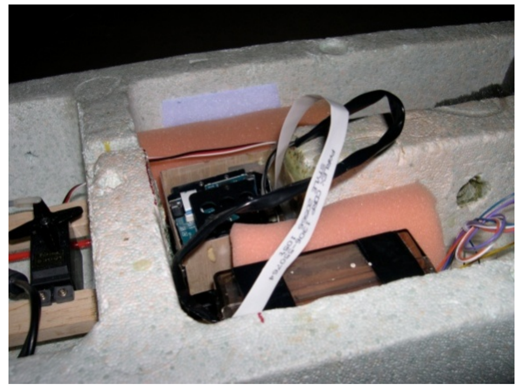
The peaks calculated from the aforementioned algorithm were compared to the peaks indentified in the video stream. A latency of 187 ms was observed. As anticipated, lower platform frequencies resulted in latencies that were, although different, within one standard deviation. The more accurate higher frequency latency will be adopted as the system's latency for use in future calculations.

The measured latency is a value particular to the experimental equipment used. That is, if the radios for video or IMU, the camera, the autopilot code or any other factor were to be modified, this latency will have to be recalculated. The equipment used in this experiment is described in section 3.

2.2 GPS Synchronization

The initial attempt to synchronize the GPS with video involved flying a plane, with the same system used to discover IMU/video latency, over targets of known geo-location. The ideal setting for this experiment was to place the targets along a linear trajectory and fly the UAV over the targets. The UAV should be straight and level over the targets to assure that targets will show up at the center of video frames. The observer will then see the target pass through the center of the video frame. With a UAV flying fast enough, the exact frame for which the target is right below the UAV will be picked out. Obviously, this data will be very sensitive to changes in pitch or roll. Assuming zero pitch and roll, flight right over the targets will allow the generation of two GPS streams, the measured and actual GPS values. Differences between these two streams can easily be translated to latency given the velocity of the UAV.

(a)



(b)

(c)

Figure 8. Various views of the GPS synchronization setup.

Figure 6 shows the experimental set up. The camera was placed to look perpendicularly to the ground on a MIG drone. A 200 mW, 1.2 GHz transmitter was used to stream the video to the ground station. The Autopilot used a MaxStream XT09 radio to transmit both GPS and IMU data to the ground station (see figures 1, 2, 3, and 4).

Upon testing, it was quickly found that it is at best very difficult to keep the drone flying straight, level, and over the targets. This is due to the small size of the craft and wind conditions. Such an experiment will have to be conducted in completely windless conditions if any useful data is to be collected.

Since the idea is to move the camera and the GPS antenna at a constant speed over targets of known geo-location, it was decided to perform this experiment using an automobile. The camera and autopilot will be placed inside the vehicle with the camera looking horizontally out of the passenger window. The ground station and data collection computer will also be placed inside the vehicle. This will assure a clear video signal throughout the experiment. By selecting a road— preferably running north-south or east-west—and by placing targets on the side of the road—preferably using lamp posts as targets—the GPS location of the autopilot can be determined exactly when it is right across from a target. By then driving the vehicle at a known constant speed, two GPS value streams may then be compared; the one transmitted from the autopilot's GPS, and then one calculated from the known target GPS locations. The time needed to synch the two steams is then found. By determining this latency, it will be possible to find the actual GPS data associated with each video frame. When the correct IMU data associated with the same frame can be found, the geo-location of any pixel in that frame can then be calculated.

3. Test Equipment

All latencies measured in this research are only usable for the specific hardware and software used in these experiments. The following subsections list all of the major components of this equipment.

Figure 9. Block camera, radios and autopilot.

3.1 Sony Block Camera (FCB-EX980SP)/Capture Card (IVC-200)

The Sony Block camera was chosen because it operates at approximately 30 fps (28.735 Hz), has relatively good resolution of 320 pixels by 240 pixels, and was readily available. The frame grabber was also readily available, Linux compatible, and PCMCIA (faster communication bus). The camera has a wide viewing angle of 42° (*12*).

3.2 Open Source Autopilot

The autopilot used in this experiment is open source and developed by Dr. Nathan Slegers. The autopilot consists of four peripheral interface controller (PIC) microcontrollers collaborating to form a control system. The diagram shown in figure 7 illustrates the layout of the autopilot. For this experiment the autopilot code was modified to increase the update frequency. By default, the autopilot gets sensor updates at 4 Hz, which is in lock sync with the GPS receiver. Once this tightly coupled link between the GPS and IMU was severed, the source code was modified to get IMU updates at 8 Hz. This means that there are two update strings available for every GPS reading (location). Much of the data is redundant but the critical data of pitch angle is fresh in each string. By default, the autopilot only sends the update with a request command is received. This presented further synchronization issues. The code was further modified to broadcast the autopilot updates with a "!" header to signal the beginning of the update string.

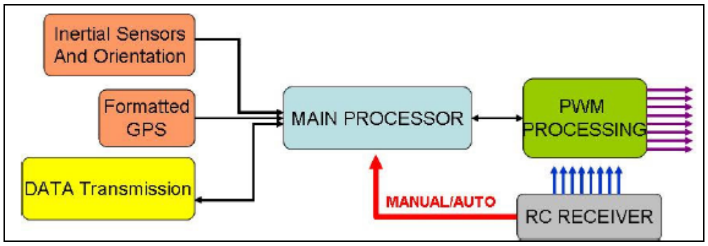


Figure 10. Autopilot diagram.

3.3 Gimbaled Platform

The gimbaled platform was designed to allow an autopilot mounted on it to rotate freely about the *x* and *y* axis. Stepper motors and optical shaft encoders were used to control the platform's rotation. Two motors were used to turn the platform around the x-axis simulating roll, while two larger motors would rotate the platform around the *y*-axis simulating pitch (*2*).

The first generation of the gimbaled platform was utilized. The platform is equipped with four stepper motors. Two NEMA 17 motors rotate the plane in the pitch direction (*x*-axis) and two NEMA 23 motors rotate the pitch plane in the roll direction (*y*-axis).

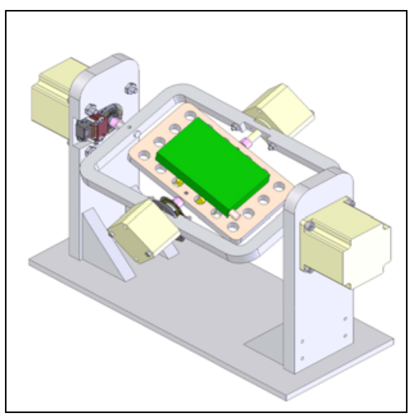


Figure 11. Gimbaled platform.

The control system for the platform was implemented on the ARM7 microcontroller. The control system has two output pins to control the platform. One output pin is the pulse width modulation (PWM) that controls the speed of rotation, and the other output is the direction pin that specifies whether the platform moves clockwise or counter-clockwise. The angle of rotation was determined by changing the output of the direction pin. There was a for loop delay implemented to determine the period of the direction pin output. This method showed inconsistencies when other algorithms were implemented alongside the PWM routine that drives the motor. The routine that controls the period of the direction pin was modified to use the timer subsystem of the ARM7. This allowed for precise control of the platform that was independent of the computation time of the code.

The platform utilizes shaft encoders for precise angular measurement. The shaft encoders were originally interfaced through the AD5 quadrature to a serial encoder interface (SEI) adapter. This system is in the process of being restructured to interface directly to the ARM7 due to the unstable operation of the adapter.

3.4 EffecTV

EffecTV is a real-time video effector. EffecTV was chosen because it is open source and because of its ability to display video effects in real-time. The software comes with over a dozen video effects, such as mosaic, blur, and edge, which could be easily modified. The effect used for this experiment is known as dumb (dumb TV). It simply displays the video frame on the screen. This effect was modified to provide the user with multi-click functionality. The system will continue to display a frame on the screen if the video stream is clicked once. A second click allows the user to locate a target in the frame and saves the location. A third click will resume the live video stream. The multi-click functionality will be utilized in future implementations of the system.

4. Summary and Conclusions

The experimental results show that there is a significant latency in time between GPS, IMU, and video streams of data. Latency is hardware and software dependant. This latency can inject propagating errors into target geo-location. As more localization is performed on the initial source of error, triangulation of objects becomes more accurate. The procedure to find latency has been documented in this paper and should be used any time there is a change in hardware or software.

5. References

1. Civil UAV Team homepage. http://www.nasa.gov/centers/dryden/research/civuav/civ_uav_index.html (accessed 2009).
2. Ali, Kamal S.; Shumaker, Justin. A Gimbaled Platform for MAV Autopilot Calibration. *27th Digital Avionics Systems Conference*, Minneapolis, MN, 2008.
3. Ali, Kamal S.; Carter, Lamarious. Miniature Autopilot Evaluation System. *American Journal of Applied Science, Journal of Computer Science* **2008**, *4* (1), 30–35, ISSN 1549-3636.
4. Carter, L.; Ali, K. Hardware In the Loop Simulator for IMU Calibration, *WMSCI 2008*, Orlando FL, June 2008.
5. Gholkar, A.; Isaacs, A.; Arya, H. Hardware-In-Loop Simulator for Mini Aerial Vehicle. *Sixth Real-Time Linux Workshop, Nanyang Technological University (NTU)*, Singapore. November 3–5, 2004.
6. Davis; Kosicki; Boroson; Kostishack. Micro Air Vehicles for Optical Surveillance. *The Lincoln Laboratory Journal* **1996**, *9* (2).
7. Alvis, W.; Murthy, S.; Valavanis, K.; Moreno, W.; Fields, M.; Katkoori, S. FPGA Based Flexible Autopilot Platform for Unmanned Systems. *2007 Mediterranean Conference on Control and Automation*, Jul. 2007, Athens, Greece.
8. Lyshevski, S. E. Autopilot Design For Highly Maneuverable Multipurpose Underwater Vehicles. *Proceedings of the American Control Conference*, Arlington, VA, June2001.
9. Adiprawita, W.; Ahmad, A. S.; Semibiring, J. Automated Flight Test and System Identification for Rotary Wing Small Aerial Platform using Frequency Responses Analysis. *ICIUS 2007 Conference*, Bali, Indonesia.
10. Jung, S. Design and Development of a Micro Air Vehicle (MAV): Test-Bed for Vision-Based Control. Master Thesis, University of Florida, 2004.
11. Barber, D. B.; Redding, J. D.; McLain, T. W.; Beard, Randal W.; Taylor, Clark N. Vision-based Target Geo-location using a Fixed-wing Miniature Air Vehicle. *J. Intell. Robot Syst.* **2006**, *47*, 361–382
12. Technical Manual, *Sony Color Camera Module*, FCB-EX980S/EX980SP.

INTENTIONALLY LEFT BLANK.

Weapons & Materials Research Directorate (WMRD)

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Multi-Functional Carbon Nanotube Metal Matrix Composites

by Brent J. Carey

Contents

List of Figures

Abstract

The material demands of advanced weapons technologies have pushed the limits of available metals and alloys, emphasizing the need for next-generation lightweight, multi-functional materials. Carbon nanotubes (CNTs), whose tensile strength and thermal and electrical conductivity along their axis exceedingly surpass virtually all known materials, have been shown to provide marked increases in the mechanical strength and hardness of metal matrix composites (MMCs); conversely, their influence on bulk thermal and electrical conductivity has been largely overlooked. Here, we explore the improvement of these three fundamental properties for two novel CNT MMC architectures: (1) a vertically aligned array (forest) of CNTs infiltrated with molten aluminum to achieve an anisotropic, continuous fiber composite, and (2) a CNT-reinforced surface coating applied through cold spraying. The application of a 1 μm sputter-coated layer of aluminum on the CNTs resulted in the encapsulation of all exposed nanotubes, a promising result which strongly suggests an affinity between the two materials and efficient interfacial interaction. This preliminary data also highlights the viability of non-destructively impregnating the CNT forests with aluminum via capillary action, a technique demonstrated previously with a polymer matrix. These composites should significantly improve upon already-existing materials for applications where greater structural integrity, thermal diffusivity and high electrical conductance are desired.

Acknowledgments

The author wishes to acknowledge the mentorship of Jerome Tzeng, Ph.D., and Shashi Karna, Ph.D.

1. Introduction/Background

Since their discovery in the early 1990s (*1*), researchers have strived to exploit the remarkable mechanical, thermal, and electrical properties of carbon nanotubes (CNTs) in a number of diverse applications ranging from nanoscale transistors (*2*) and interconnects (*3*), supercapacitors (*4*), artificial muscle fibers (*5*), and even as axles in molecular vehicles (*6*). While a number of these innovations are still years or perhaps even decades from wide-scale implementation, recent advances in nanotube synthesis (*7*) have afforded the opportunity for the inclusion of CNTs to improve upon current technologies. By the way of composite materials, their contribution to mechanical strength has been seen predominantly with the reinforcement of glassy polymer matrices such as epoxies, which see notable improvement for even modest nanotube concentrations. For these composites, CNTs have been successful as both the sole reinforcing phase (*8*), as well as a supplement to carbon fibers (*9*).

The concept of a metal matrix composite (MMC) is not new, though the processing challenges as compared to their polymer matrix analogs, has somewhat impeded their advancement thus far. Even for MMCs reinforced by macroscopic fibers, it is nontrivial to implement anything other than basic architectures. This problem is compounded for nanomaterials such as CNTs, where homogeneous dispersion (not to mention any form of arrangement or alignment) remains a very challenging problem. Some have reported success with various forms of forced dispersion such as hot extrusion (*10*) and ball mixing (*11*), though the level of homogeneity and the integrity of the nanoparticles after this type of processing are contested. Conversely, nanoscale engineering techniques (*12*) have resulted in truly homogenous dispersion, but the practicality of these methods is contested due to their relative expense and limited scalability. In spite of these challenges, there exist a number of Army applications which would significantly benefit directly from a multifunctional CNT-reinforced MMC if an efficient and scalable process is devised.

In this work, we strive to implement these remarkable molecules as constituents in a aluminum MMC to impart mechanical, thermal, and electrical improvement over the pure metal. A number of studies have shown CNTs to be effective as structural reinforcement in a metal matrix, with one group reporting a 78% increase in hardness with just 5 wt% CNTs dispersed in aluminum (*13*). Amongst the work involving these nanotube MMCs, however, there exists little information about how the electrical and thermal behavior is altered as a result of the presence of the CNTs. To explore the viability of these materials for ordnance applications, we intend to explore two distinct methods to produce: (1) a vertically aligned CNT MMC and (2) a CNT-reinforced MMC surface coating.

2. Experiment/Calculations

2.1 Vertically Aligned Carbon Nanotube Metal Matrix Brush Contact

While sliding electrical contacts are essential in devices such as electrical motors, their utility extends to a litany of other applications where electrical conduction between moving parts is necessary. As far as Army applications, the conductive rails for an electromagnetic gun could benefit greatly from reduced projectile/rail interfacial electrical resistance and wear. In an effort to improve upon the carbon composite and filamentous metal brushes currently used as sliding contacts in high-performance applications, it has recently been shown that forests of vertically aligned carbon nanotubes can serve as compliant, highly effective nanoscale brushes which show almost an order-of-magnitude lower contact resistance as compared to these traditional technologies (*14*). Additionally, the excellent thermal conduction along the axis of the CNTs can serve as a heat sink that draws thermal energy away from the sliding interface, a property exploited previously to show potential in modern electronics (*15*). A downside, however, is their susceptibility to deformation under even modest stress. As such, a reinforcing matrix is necessary for these nanoscale bristles to be functional as robust sliding contacts.

In the engineering of strong composites, an optimal configuration for reinforcement is fibers that span the entirety of the composite. Given their extremely small size and affinity to cling to each other, this has proven to be a particularly difficult task with nanomaterials. It has recently been shown, however, that the intertube spaces between these vertically aligned CNTs can be readily infiltrated by a poly(dimethylsiloxane) monomer and cured in situ (figure 1) (*16*). This continuous reinforcement results in a three-fold improvement as compared to randomly aligned composites with identical CNT loading, a property owed to the unrivaled tensile strength inherent to carbon nanotubes (*17*). Additionally, and perhaps most importantly for our application, the alignment should result in optimal thermal and electrical conductance since these properties are greatest along the axial direction of the CNTs.

The production of these aligned nanotube composites is possible due to the incredible wetting between the CNT surface and the polymer chains (~0° wetting angle) (*18*). This property allows for the spontaneous infiltration and replacement of air space with the polymer matrix. A similar affinity does not exist between the CNT's and aluminum, where the wetting angle is reported to be a discouraging 130–140°, making infiltration unlikely (*19*). However, in this work by Ci et al., they also report that the loosely bound amorphous/pyrolized carbon coating on the CNTs grown by chemical vapor deposition (CVD) will react with aluminum to form aluminum carbide (Al_4C_3), a typically unfavorable byproduct due to its brittleness. In this case however, the 50–70° contact angle between Al and Al_4C_3 should significantly increase the probability of infiltration.

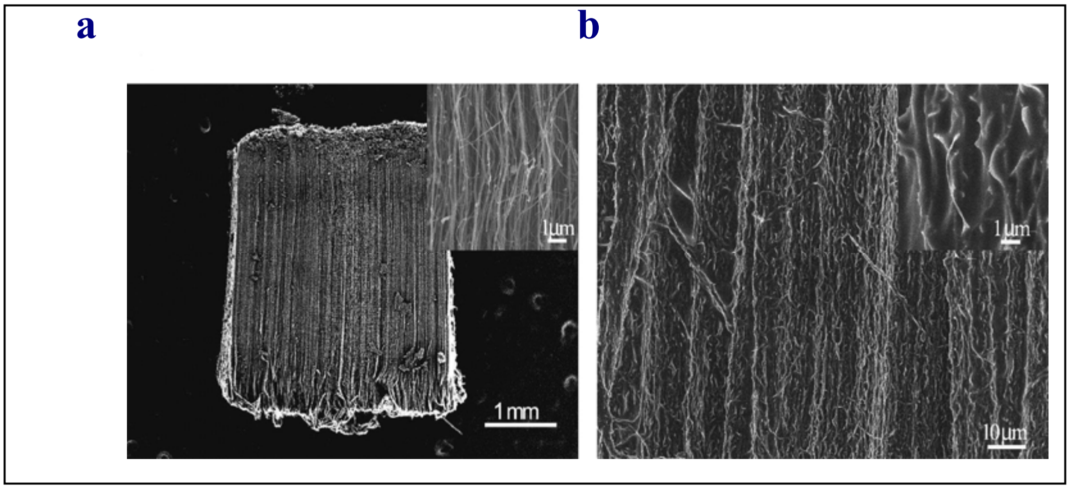


Figure 1. Example of a continuously reinforced nanotube composite[16]. Scanning electron microscope (SEM) images showing an aligned carbon nanotube (CNT) forest both before and after impregnation with a polymer matrix. a, The CNTs are oriented in the same direction, a function of the chemical vapor deposition growth process. The open space between the nanotubes can also be observed. b, After infiltration, it can be seen that the CNTs have remained alignment and that the polymer has seamlessly infiltrated the forest and replaced the air space between the nanotubes.

CVD-grown, vertically aligned MWNTs provided by Rice University (Houston, TX) serve as the infrastructure for this continuously reinforced MMC. To test the viability of this infiltration, the exposed ends of the CNTs were coated with a 1 µm layer of aluminum via sputter coating. The successful wetting initiated by this process (described below in the discussion section) indicates the viability of the above infiltration mechanism, and future plans include heating the sample up past the melting point of aluminum (660 °C) under vacuum (to prevent combustion of the CNTs) in order to observe any morphological changes or infiltration due to capillary action. If successful, efforts will continue to achieve complete infiltration and the realization of a continuously reinforced CNT MMC composite.

2.2 Carbon-Nanotube-Reinforced Surface Coating by Cold Spray Deposition

Cold spray has evolved as a versatile technique to apply coatings or even form freestanding structures at temperatures much lower than traditional spray processes, such as thermal spray coating. By accelerating micron-sized metallic particles to supersonic velocities close to the target surface, plastic deformation occurs during impact, producing up to a 100% dense surface coating. This process has been useful for the application of coatings for corrosion protection, wear resistance, electromagnetic interference shielding, and can even be employed as a method to repair macroscopic damage (*20*).

The application of nanoparticles as constituents in thermal spray processes has been explored by a number of groups over the past few years, and the successful use of CNTs has been reported. Amongst this work, there is only one record of cold spray forming using CNTs, in which micron-sized Al-Si eutectic powders were coated with CNTs through spray drying and then cold sprayed with aluminum powder to produce 500 µm-thick coatings (*21*). Homogenous CNT dispersion is reported, and through nanoindentation, these coatings were shown to display up to a 300% improvement in elastic modulus due to a .5 wt% loading of CNTs. Note, however, that their hardness measurements varied greatly due to the "pockets" of CNTs which were formed where the Al-Si-CNT agglomerates impacted the surface. Also, the addition of CNTs to an aluminum matrix has been recently shown to significantly reduce both the coefficient of friction and the wear rate as compared to the pure metal (*22*).

We intend to mimic (and hope to improve upon) this process (figure 2a) to achieve a highly conducting, resilient surface coating with improved conductivity due to the presence of highly graphitized CNTs. To form the agglomerates for cold spraying, 50–80 nm diameter, 10–20 µm long, graphitized multi-walled CNTs (www.cheaptubes.com, Brattleboro, VT) at concentrations of 1 wt%, 2 wt%, and 5 wt% will be spray-dried with 99% pure, 100 nm diameter aluminum particles (Tekna Plasma Systems, Sherbrooke, Québec, Canada) to produce 20 µm agglomerates (figure 2b). Through this process, the agglomerates become intercalated with CNTs, providing homogeneity during spraying. Using 6061 aluminum as a substrate, the CNT-coated agglomerates are cold-sprayed to deposit a 500 µm surface coating (figure 2c).

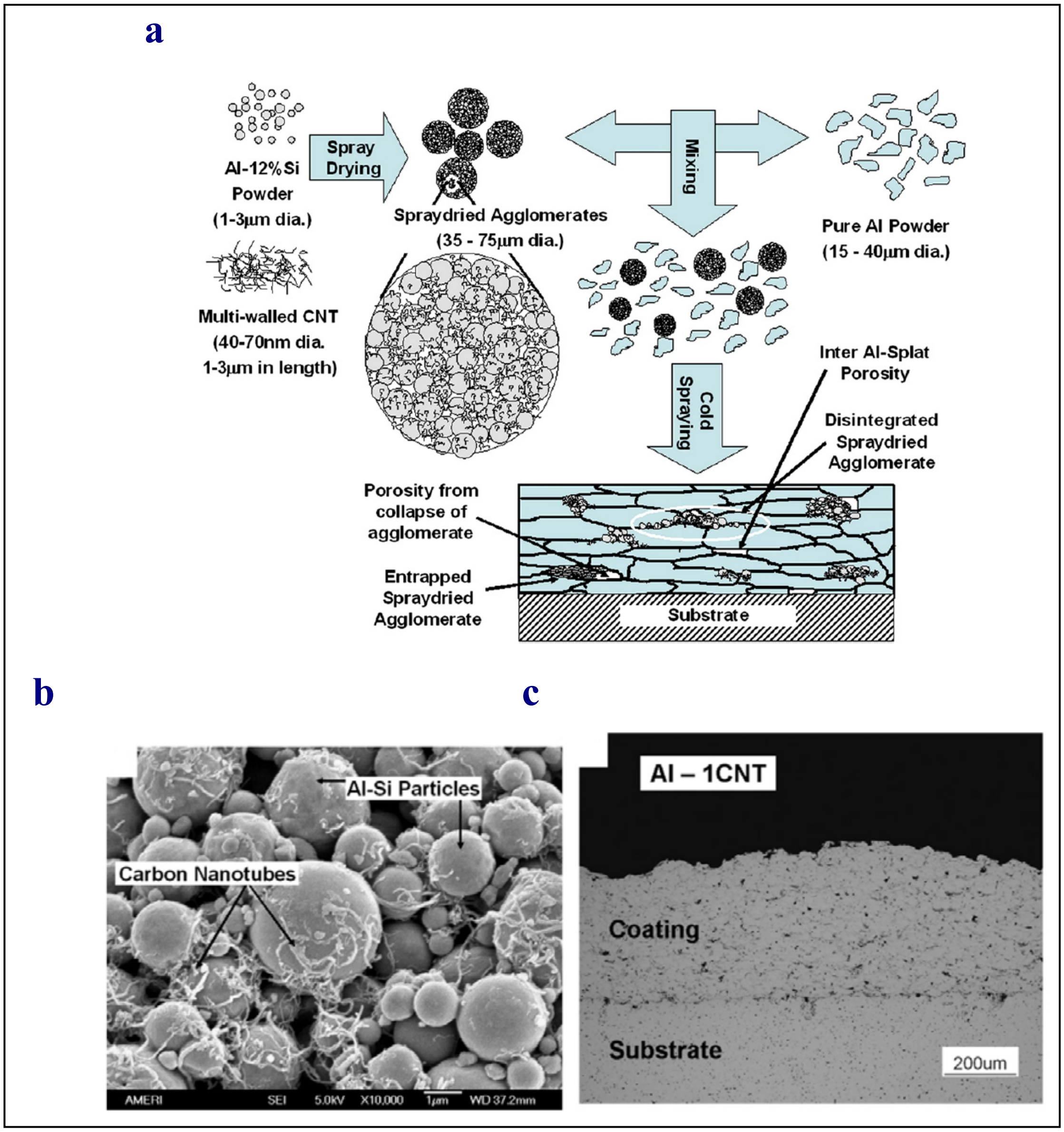


Figure 2. Cold spraying a nanotube-reinforced surface coating[21]. (a) Schematic showing the process to coat aluminum-12% silicon nanoparticles with CNTs for use in cold spraying. (b) Image showing the CNTs on the surface and interspersed between the nanoparticles within the agglomerate (c) As compared to the substrate, the coating has a much greater density of voids which threaten it's integrity. By using a much more pure aluminum and not mixing the agglomerates with Al flakes, the result should be a more homogeneous coating with a lower density of voids.

3. Results and Discussion

3.1 Vertically Aligned Carbon Nanotube Metal Matrix Brush Contact

While there is still much progress to be made, the results from sputter coating aluminum on the CNTs are very promising. It can be seen in the scanning electron microscopy (SEM) images in figures 3 and 4 that the sputtered aluminum on both the tips and the sidewalls of the aligned CNTs resulted in a coating on each individual CNT. This is clearest in figures 4a and 4b, which show a view orthogonal to the deposited surface and into a small void in the coated surface, respectively; here, two assertions can be made: (1) the single CNT protruding from the surface is surrounded by the coating, and (2) the deposition penetrated the surface layer of CNTs somewhat. These observations show that the deposition did not simply occur in the "line-of-sight" from the target to the surface of the CNTs, as is typical of sputter coating. This suggests that the aluminum has an affinity for the CNTs (perhaps via the formation of Al_4C_3), which would make impregnation through infiltration a distinct possibility.

Another interesting observation is the fact that the coating is not smooth, like the surface of the CNTs it surrounds. By comparing figures 3b and 3c this is clearly seen, and there appear to be distinct edges to the raised areas of the coating. While SEM is not adequate to confirm this, this morphology may be due to crystal formation on the surface of the CNTs. After heat treatment the samples will be imaged again. A reduction of the coating thickness indicates further penetration of aluminum into the forest along the CNT surfaces, while coarsening of the "crystallites" would confirm their structure.

3.2 Carbon-Nanotube-Reinforced Surface Coating by Cold Spray Deposition

At the time of writing this report, we are still awaiting the carbon nanotubes for use as the cold spray dispersant. While the work introducing this technique was successful in producing a CNT-reinforced surface coating via cold spraying, there are a few improvements that are planned:

1. **Reducing the density of CNTs coated on the agglomerates.** The collapse of the agglomerates during spraying caused "pockets" of CNTs in the sprayed coating, which caused the surface hardness to vary greatly.
2. **Reduce the agglomerate size.** To improve the homogeneity, we intend to reduce the agglomerate size to reduce the likelihood of void formation.
3. **Spraying with only the CNT-infused agglomerates.** Both in the images and described by the schematic in figure 2a, the dispersion of the CNTs is not truly homogeneous. For the same reasons as above, spraying only the agglomerates should result in more even dispersion.

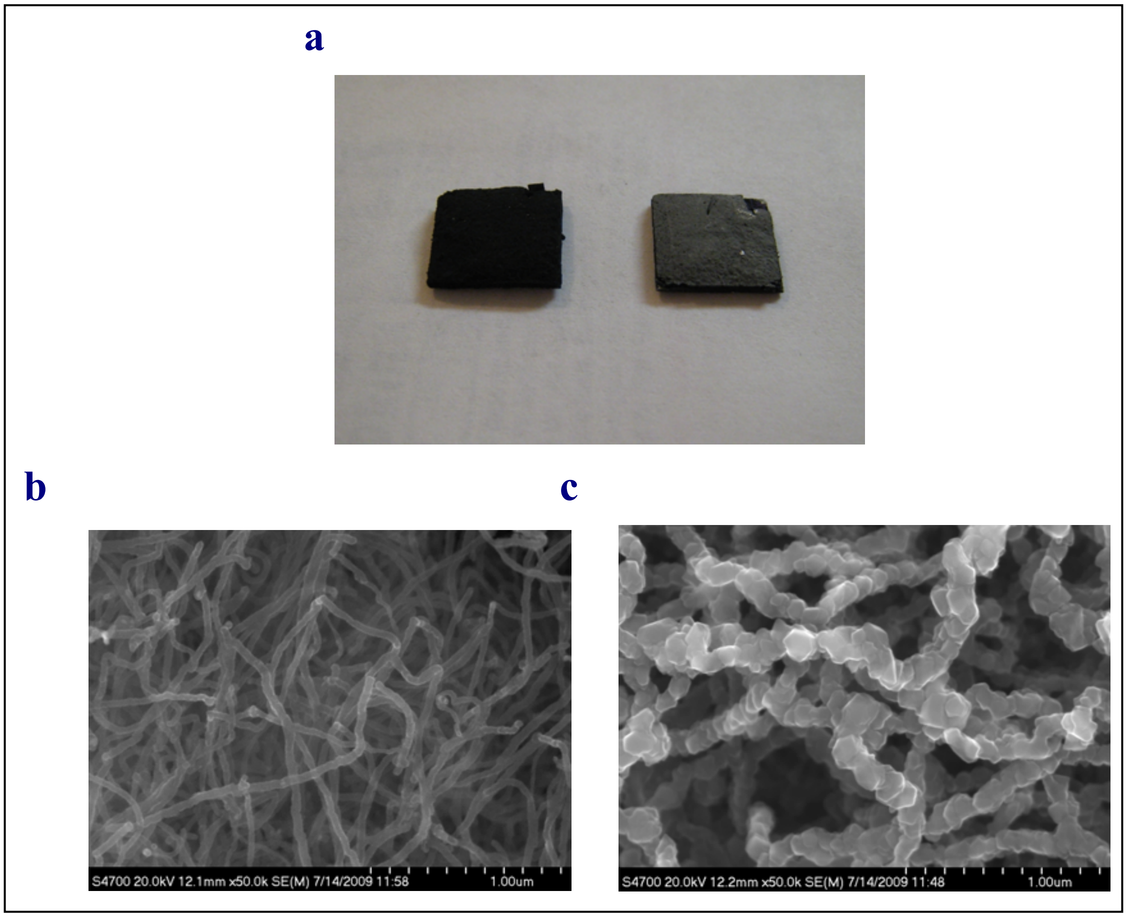


Figure 3. The intimate interaction between carbon nanotubes and aluminum. Images showing a vertically aligned forest before and after sputter coating with 1 µm of aluminum. (a) Photograph showing the optical difference between the pristine and aluminum-coated forest. (b) This SEM image shows that the top of this CNT forest exists as somewhat of a tangled mat, and that the CNTs measure between ~30–80 nm in diameter. (c) It can be seen that after sputter coating, the CNTs are sheathed in aluminum. This strongly suggests good interaction (wetting) and is promising for investigation of impregnation with aluminum to create a continuous CNT metal matrix composite. Also, the morphology of the coating is not smooth like the surface of the CNTs. The crisp sides of the bumps may indicate the formation of metal crystals on the CNTs.

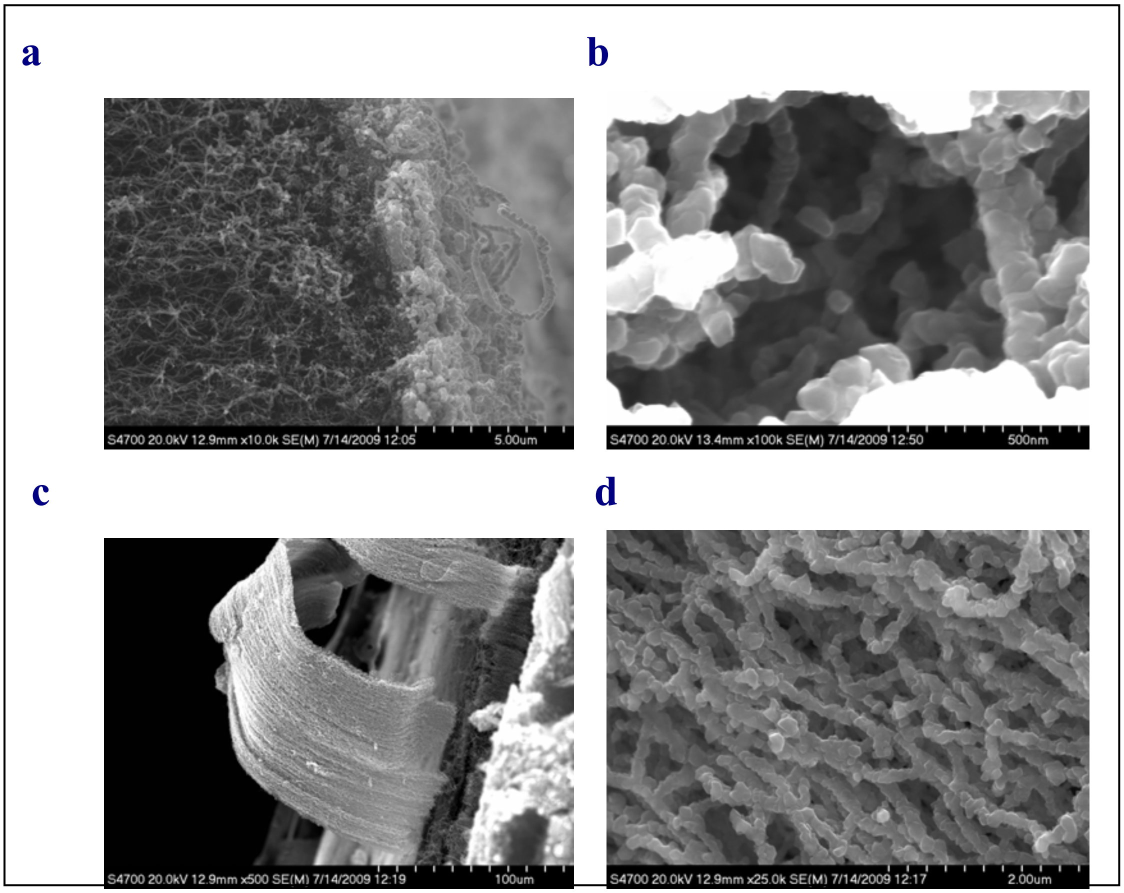


Figure 4. Additional SEM characterization of aluminum-sputtered nanotube forest. (a) Side view of forest after coating. The aluminum partially penetrated the surface and the single CNT protruding from the surface appears to be completely coated, a result not typical of sputter coating. These observations suggest good interaction between the Al and the CNTs. (b) Image down a small void on the sputtered surface. The CNTs underneath are coated, reinforcing the excellent wetting hypothesis. (c) The CNTs in this image actually grew off of the side of the silicon wafer substrate, away from the forest. These CNTs were sputtered on their sidewalls, and (d) also showed complete covering.

4. Summary and Conclusions

With advanced weaponry requiring improved material performance in a lighter, less obtrusive package, engineering at a smaller scale becomes necessary. This proves to be no trivial task, and it becomes necessary to exploit self-assembly and nanoscale dispersion to achieve the desired homogeneity and alignment which provides the desired material properties. For a multi-functional composite, carbon nanotubes are ideal due to the remarkable blend of strength and conductivity which is owed to their sp^2-bonded structure.

In this work, we have proposed two viable methods of producing CNT-reinforced composites which should yield significant improvement in strength and thermal and electrical conductivity over traditional metals and alloys. This work is still in progress, but the results so far are very promising and indicate good interaction between the CNTs and aluminum, the desired matrix for these MMC's. If successful, the recent advances in CNT synthesis techniques will allow for scalability and implementation in both existing and future weapons technologies.

5. References

1. Iijima, S. *Nature* **1991**, *354*, 56–58.
2. Tans, S. J.; Verschueren, A.R.M.; Dekker, C. T. *Nature* **1998**, *393*, 49–52.
3. Kreupl, F. et al. *Microelectronic Engineering* **1994**, *64*, 399–408.
4. Frakowiak, E. et al. *Applied Physics Letters* **2000**, *77*, 2421–2423.
5. Aliev, A.E. et al. *Science* **2009**, *323*, 1575–1578.
6. Shirai. Y. et al. *Nano Letters* **2005**, *11*, 2330–2334.
7. Hata, K. et al. *Science* **2004**, *306*, 1362–1364 (2004).
8. Qian, D. et al. *Applied Physics Letters* **2000**, *76*, 2868–2870.
9. Thostenson, E. T. et al. *Journal of Applied Physics* **2002**, *91*, 6034–6037.
10. Kuzumaki, T. et al. *Journal of Materials Research* **1998**, *13*, 2445–2449.
11. Esawi, A.M.K. et al. *Composites Science and Technology* **2008**, *68*, 486–492.
12. Esawi, A.M.K. et al. *Mat. Sci. and Eng. A* **2009**, *508*, 167–173 (2009).
13. Zhong, R., Cong, H. & Hou, P., *Carbon* **2003**, *41*, 848–851.
14. Toth, G. et al. *Advanced Materials* **2009**, *21*, 2054–2058.
15. Kordas, K. et al. *Applied Physics Letters* **2007**, *90*, 123105.
16. Ci, L. et al. *Nano Letters* **2008**, *8*, 2762–2766.
17. Yu, M.F. et al. *Science* **200**, *287*, 637–640.
18. Barber, A. et al. *Phys. Rev. Lett.* **2004**, *92*, 186103.
19. Ci, L. et al. *Acta Materialia* **2006**, 54, 5367–5375.
20. Alkhimov, A. P. et al. *Doklady Akademii Nauk, SSSR* **1999***, 315*, 1062–1065.
21. Bakshi, S. R. et al. *Surface & Coatings Technology* **2008**, **BI**, 5162–5169.
22. Jiang, J. et al. *Trans. Nonferrous Met. Soc. China* **2007**, *17*, s113–s116 (2007).

INTENTIONALLY LEFT BLANK.

U.S. Army Research Laboratory

SUMMER RESEARCH TECHNICAL REPORT

Investigation of Spin-Yaw Resonance for Maneuvering a Finned Projectile

by Paul R. Jones

Contents

List of Figures

List of Tables

Abstract

The U.S. Army Research Laboratory (ARL) is designing a guided weapon called the Very Affordable Precision Projectile (VAPP). The VAPP projectile uses two canard actuators (movable fins) to maneuver through the air. While canard actuation is well known for guidance, a less conventional method using roll-yaw resonance (also know as spin-yaw resonance) has significant potential. A recent ARL-sponsored 120-mm projectile test unexpectedly demonstrated a course change of 10°. From unpublished ARL data, it is speculated the projectile encountered roll-yaw resonance. This work examines the exploitation of roll-yaw resonance on the VAPP projectile and numerically demonstrates course changes of up to 22°.

Acknowledgments

Mr. Bernard Guidos of the U.S. Army Research Laboratory's Weapons and Materials Research Directorate, Aerodynamics Branch provided a wealth of technical support. This work would not have been possible without him.

1. Introduction

The U.S. Army and Navy both desire projectiles capable of in-flight course correction. Course correction is vital for impacting intended targets. The U.S. Army Research Laboratory (ARL) is currently designing a weapon called the Very Affordable Precision Projectile (VAPP). The VAPP projectile will use two canards to guide the projectile. The canards (movable fins) are capable of achieving 10° course deflections. An alternative method for achieving large course deflection is to resonate the projectile's pitch and yaw frequencies. This phenomenon is known as roll-yaw (or spin-yaw) resonance.

A recent ARL-sponsored 120-mm projectile test unexpectedly demonstrated a course change of 10° over a distance of 300 m. This projectile was launched without canards. From unpublished ARL data, it is speculated the projectile encountered roll-yaw resonance.

This report examines the exploitation of roll-yaw resonance for course correction on the VAPP projectile. Trajectories and projectile characteristics were determined using six-degree-of-freedom (6-DOF) software, Prodas (version 3.4.1), designed by Arrow Tech Associates.

2. Tri-Cyclic Motion

A projectile, subjected to atmospheric forces, will experience cyclic motion about its trajectory. This motion may be decomposed into three vector (arm) components: nutation, precession, and trim. Figure 1 shows an illustration of the components.

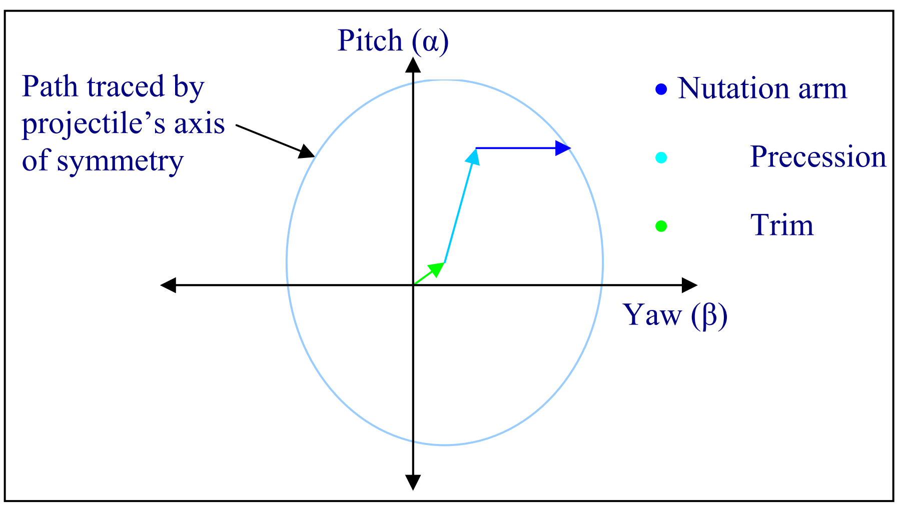


Figure 1. Vector components of tri-cyclic motion.

The trim arm is defined as the total angle of attack that results in a zero net moment on the projectile. The arm represents a projectile's equilibrium orientation during flight. The trim arm spins at the same rate as the projectile's roll rate (equation 1). For a constant roll rate, the nutation (j = 1) and precession (j = 2) frequencies are described by equation 2. During resonance, the nutation arm's spin rate is near equivalent to the roll rate (equation 1) of the projectile as symbolized in equation 3. See page 27 for notation used throughout report. For a detailed explanation of equations used refer to references 1 and 2.

Projectile Roll Rate: $$\phi' = \phi_b' + (\phi_0' - \phi_3')e^{-K_p s} \tag{1}$$

Nutation Frequency: $$\phi_j' = \frac{1}{2}\left[P \pm \sqrt{P^2 - 4M}\right] \quad J = 1, 2 \tag{2}$$

Roll-Resonance: $$\phi' = \phi_j' \tag{3}$$

It is important to include effects caused by the trim arm when resonating a projectile, as opposed to the traditional epicyclic model (nutation and precession arms only), because the trim arm will amplify in magnitude. This influences the projectile's motion.

The magnitude of the trim arm when not in resonance is described by

$$K_{30} = e^{i\Phi_{30}} = \frac{-iA}{(\phi')^2 - P\phi' + M - i(\phi' H - PT)} \tag{4}$$

During resonance, the denominator of equation 4 will decrease in magnitude, since

$$(\phi')^2 - P\phi' \approx 0 \tag{5}$$

This is due to the gyroscopic spin term's (equation 6) dependence upon the roll rate (equation 7):

$$P = \frac{I_Z}{I_Y}\left(\frac{pl}{V}\right), \text{ gyroscopic spin} \tag{6}$$

$$p = \frac{d\phi}{dt}, \text{ roll rate} \tag{7}$$

which will cause the magnitude of the trim arm to increase according to

$$K_{3R} = \left|\frac{-iA}{M - i(\phi' H - PT)}\right| = \left|\frac{A}{\lambda_1 \phi_1' (2 - I_Z / I_Y)}\right| \tag{8}$$

The amplification of the trim arm is obtained by dividing the arm's magnitude at resonance by its magnitude in non-resonance.

$$\frac{K_{3R}}{K_{30}} = \left|\frac{\phi_1'}{2\lambda_1}\right| \tag{9}$$

For a finned projectile, this amplification ranges from zero to 100, increasing with dynamic instability (*3*).

3. Roll-Yaw Resonance

Roll-yaw resonance (roll resonance) occurs when a projectile rolls at a rate near its natural spin frequency (nutation or precession arm frequencies). Only the nutation frequency is considered a criterion for resonance, since the precession frequency will be the same magnitude but opposite in sign. When the two frequencies are nearly the same, the total yaw of the projectile will amplify in magnitude. Roll resonance has traditionally been avoided in projectile design by rolling a projectile well above its resonance frequency. This is done to ensure dynamic stability. The purpose for intentionally resonating the projectile's yaw is to induce large angular deflections, which accompanied by an axial thrust, may invoke large changes in course.

4. VAPP Projectile Design and Trajectory Specifications (Control)

Projectile parameters and initial conditions used as the control trajectory configuration are outlined in tables 1−3. These data were obtained from the 2 June 2009 VAPP specifications. Figure 2 shows a three-dimensional drawing of the projectile.

Table 1. Projectile parameters.

	Before Burnout	**After Burnout**
Weight (kg)	18.40224	17.49812
Axial inertia (kg-m^2)	0.03199	0.03148
Transverse inertia (kg-m^2)	.56067	0.51076
CG from nose (mm)	.4227	0.4115

Table 2. Trim and fin properties.

Trim angle (deg)	0.00
Trim Orientation (deg)	0.00
Fin Cant (deg)	1.50
Fin Count	4

Table 3. Initial trajectory conditions.

Quadrant elevation (deg)	59.000
Gun azimuth (deg)	90.000
Muzzle velocity (m/s)	430.0
Spin at muzzle (Hz)	20.
Twist (cal/rev)	18.00
Exit spin ratio	0.10
Initial X position (m)	0.
Initial Y position (m)	0.
Initial Z position (m)	0.
Initial pitch angle (deg)	0.0
Initial yaw angle (deg)	0.0
Initial pitch rate (rad/s)	0.0
Initial yaw rate (rad/s)	0.0

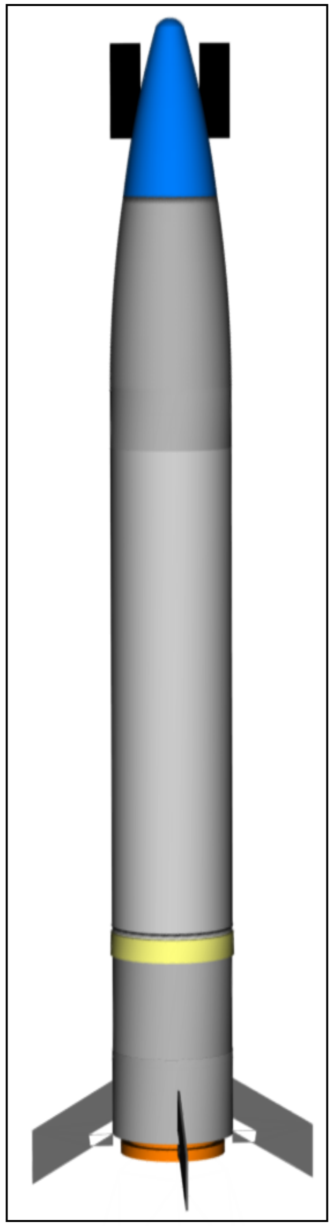


Figure 2. Three-dimensional drawing of VAPP projectile.

The attitude of the control projectile in-flight is described by figures 3–7. Figure 3 is a plot of the projectile's roll rate versus time. The projectile exits the muzzle rolling at 126 rad/s (7200 deg/s) and never rolls less than 50 rad/s (2865 deg/s) during the flight. Figure 4 plots the projectile's nutation frequency versus time. It is clear that after 10 s, the nutation frequency may be assumed constant, 0.045 rad/m (9.05 rad/s, 518.48 deg/s). Figures 5–7 plot the projectile's pitching and yawing motion as a function of time. While these figures appear to show drastic angular changes from 0 to 28 s, the magnitude of the change is less than 0.011 rad (0.63°).

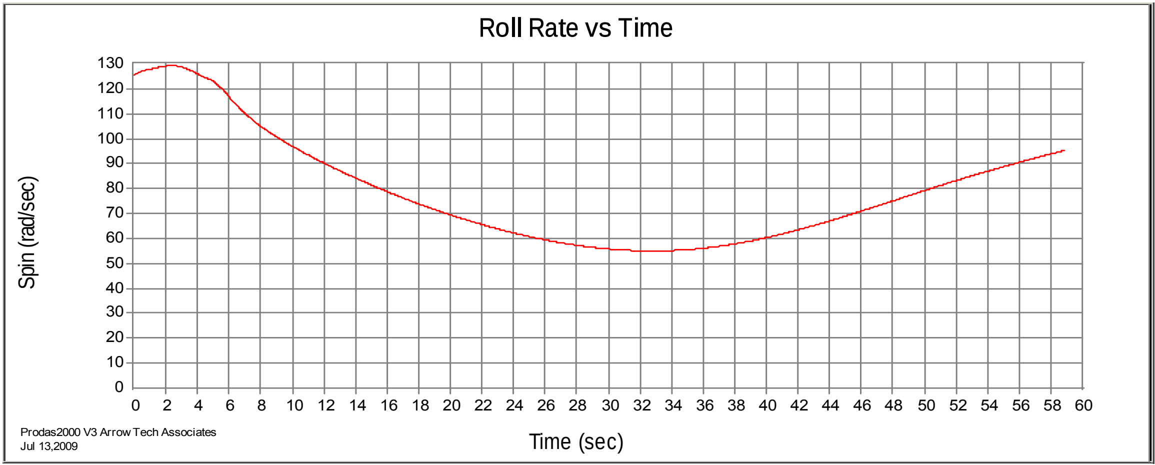


Figure 3. Control projectile, roll rate versus time.

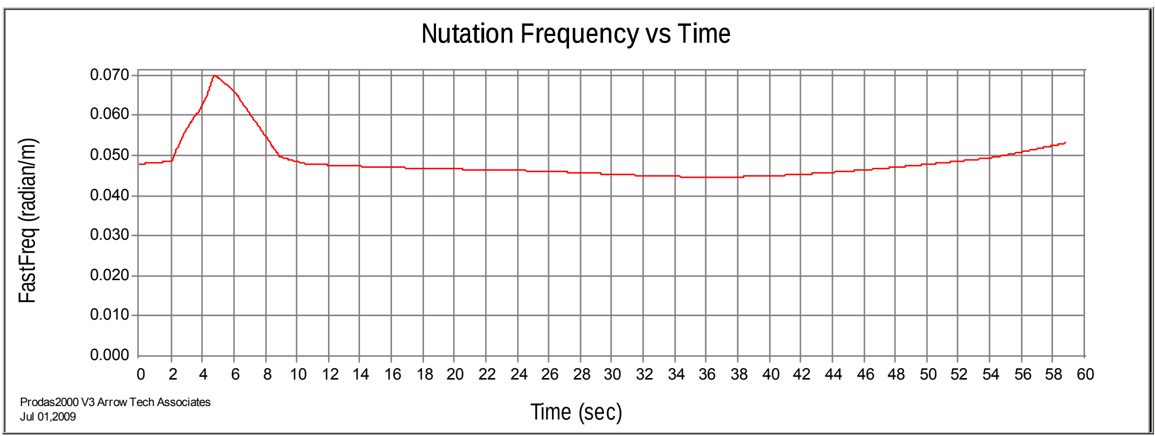


Figure 4. Control projectile, nutation frequency versus time.

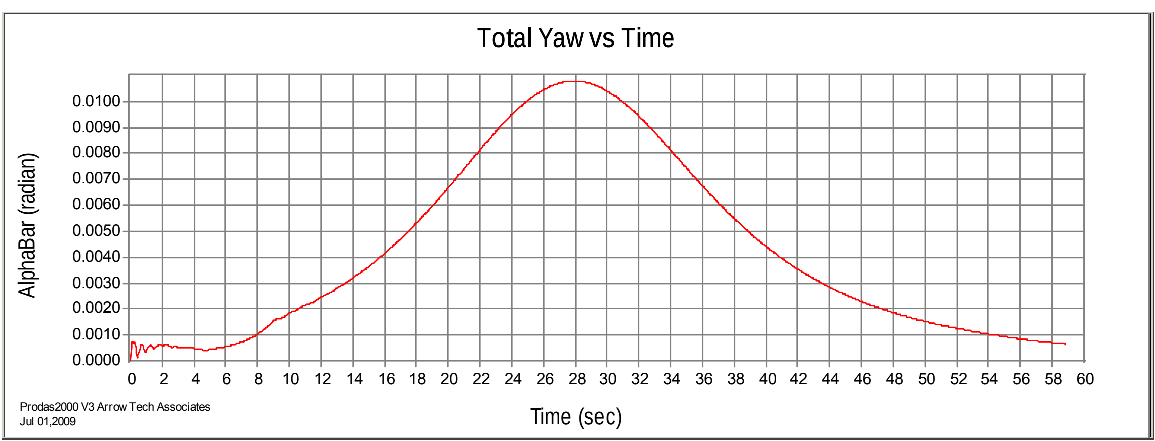


Figure 5. Control projectile, total yaw versus time.

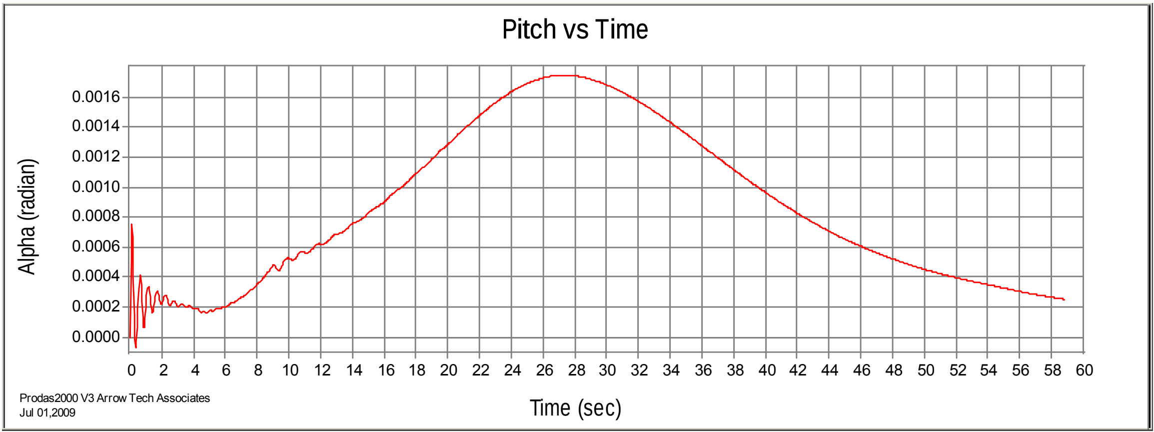


Figure 6. Control projectile, pitch versus time.

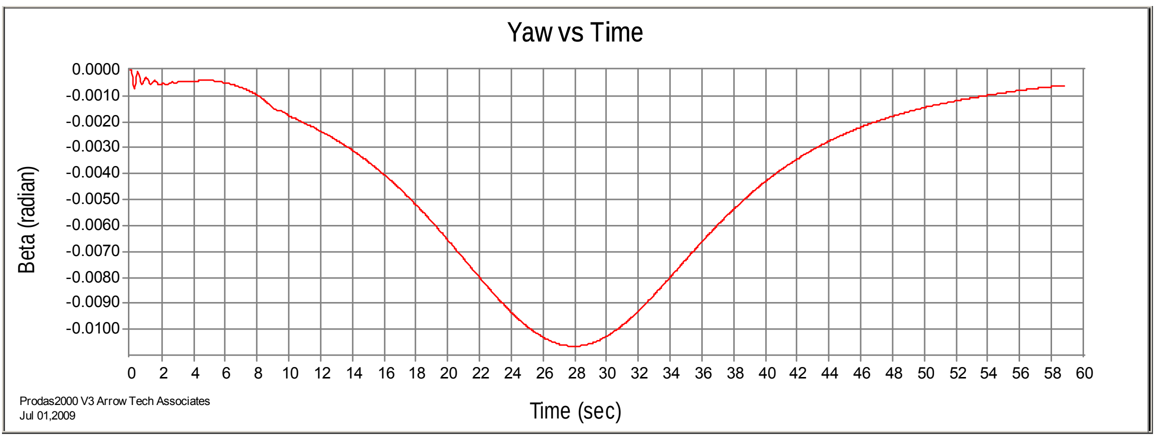


Figure 7. Control projectile, yaw versus time.

5. Methodology

There are five parameters of interest for initiating resonance; however, only the first was explored:

- Fin cant angle
- Axial moment of inertia from nose
- Transverse moment of inertia from nose
- Center of gravity (C.G.) from nose
- Canard actuation

Modifying the fin cant angle directly affects the projectile's roll rate. Figure 8 shows an illustration of fin cant angle. Increasing the cant angle will increase the roll rate. Similarly, decreasing the cant angle will decrease the roll rate. The fin cant was simulated for angles of 0° to 7.9° by increments of 0.1°. Further simulations tested 24 additional cant angles ranging from −2.00° to 0.25° .

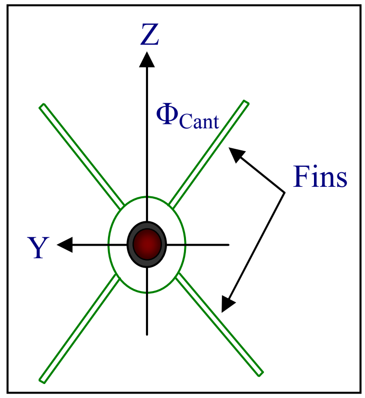


Figure 8. Illustration of fin cant angle from the rear of the projectile.

The axial moment of inertia, transverse moment of inertia, and location of the center of gravity all affect the pitching and yawing motion of the projectile. These parameters were not used, because they are difficult to change on an established weapon.

Canard actuation has a similar effect on projectile roll as modifying the fin cant. While addressing this parameter is beyond the scope of this report, it is important to note that it is a reasonable possibility for encountering resonance during flight.

Trajectories were constructed numerically using commercial 6-DOF software, Prodas (version 3.4.1). The effectiveness of each configuration was assessed using the projectile's pitch, yaw, total yaw, roll rate, and position throughout the entire flight. The control trajectory used for comparison is outlined in the section 4. See figures 3–7 for reference.

6. Results

The fin cant was modified from angles of −2.0° to 7.9°. For brevity, nine variant configurations were plotted simultaneously.

6.1 Comparison of Fin Cant from 0° to 6° (Comparison 1)

The fin cant was simulated for angles of 0° to 6° to outline the scope of possible configurations. Figures 9–12 illustrate the relationship between a projectile's fin cant and its roll rate. Increasing the fin cant increased the projectile's roll rate. It is important to note that the projectiles resonance frequency is around 9 rad/s (518 deg/s). The configurations with the closest roll rates involved fin cants of 0° to 2°.

Separating the configurations into two classes, a low and a high fin cant, angles from 0° to 2° would represent low cant angles, while angles of 2° or more represent high cant angles. High cant angles showed moderate changes in total yaw, pitch, and yaw (see figures 10–12). For a fin cant of 5°, the projectile yawed up to 0.11 radians (6.30°). Although these deflections were significant, the projectile's roll rate did not match the nutation frequency of the control, and therefore, resonance did not occur.

Low cant angles showed roll rates closer to the projectile's nutation frequency. Three additional comparisons were made to determine a configuration that rolls nearest to the resonance frequency.

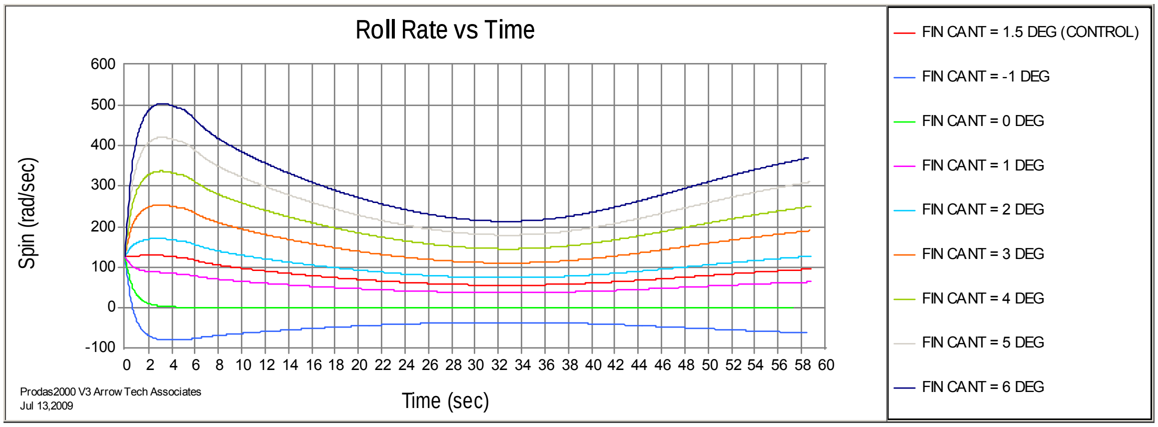


Figure 9. Comparison 1, roll rate versus time.

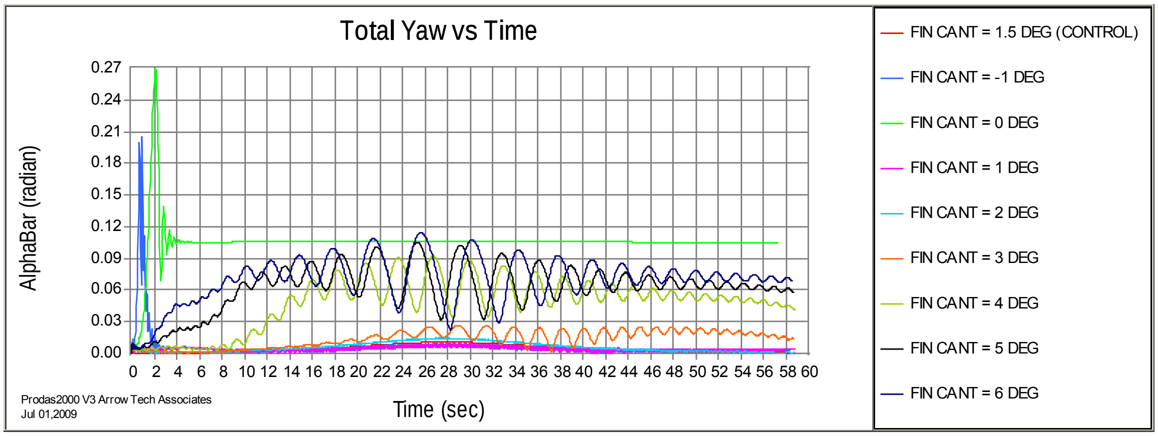


Figure 10. Comparison 1, total yaw versus time.

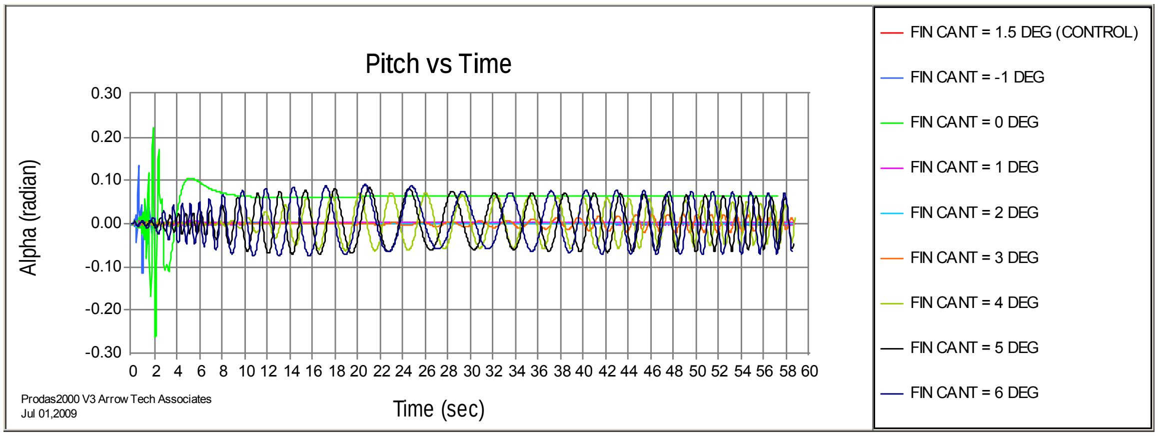


Figure 11. Comparison 1, pitch versus time.

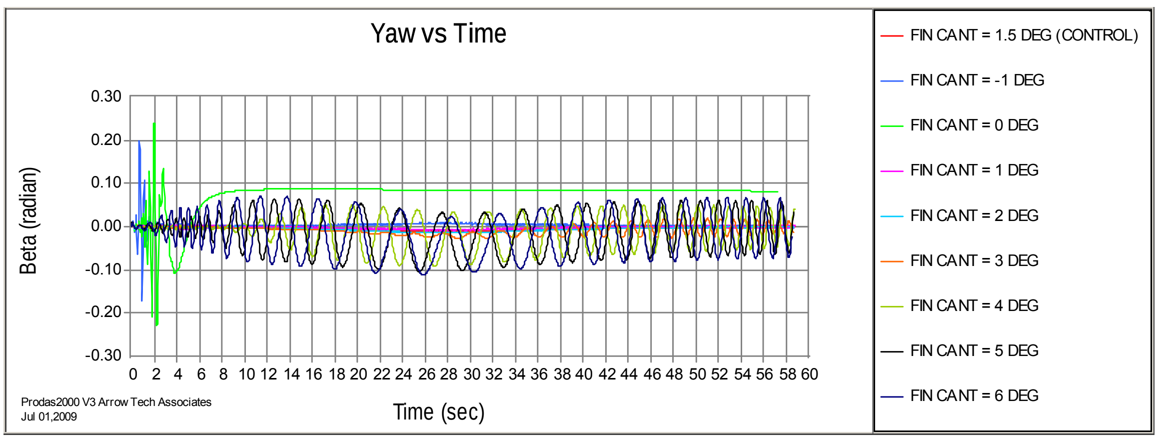


Figure 12. Comparison 1, yaw versus time.

6.2 Comparison of Fin Cant from 0.0° to 1.6° (Comparison 2)

The second comparison simulated fin cant configurations from 0.0° to 1.6°. From figure 13 it is clear that configurations with fin cants of 0.0° and 0.2° came closest to the resonance frequency. This is confirmed by figures 14–16, which show the projectile's total yaw, pitch, and yaw amplifying in magnitude. The configuration with a fin cant of 0° experienced a total yaw of 0.24 rad (14°) around 2 s in flight. The configuration with a fin cant of 0.2° induced a 0.44 rad (25°) total yaw after 5 s of flight.

On average, a 0.0° fin cant was within 103% of the resonance frequency, and a 0.2° fin cant within 43%. This means that the 0.0° fin cant configuration rolled two times slower than the resonance frequency, whereas the 0.2° fin cant configuration rolled 43% faster. These two configurations may effectively serve as bounds for achieving resonance. This will be further justified in section 6.4.

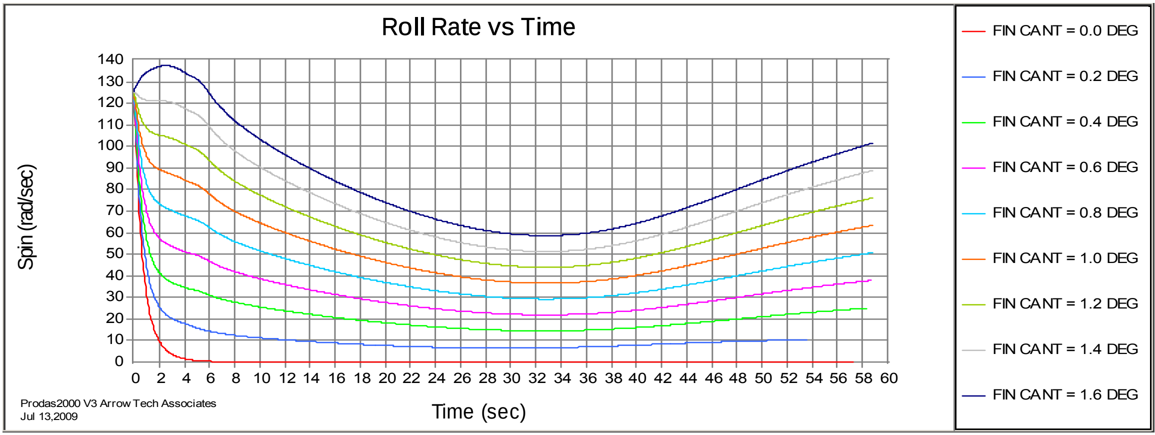


Figure 13. Comparison 2, roll rate versus time.

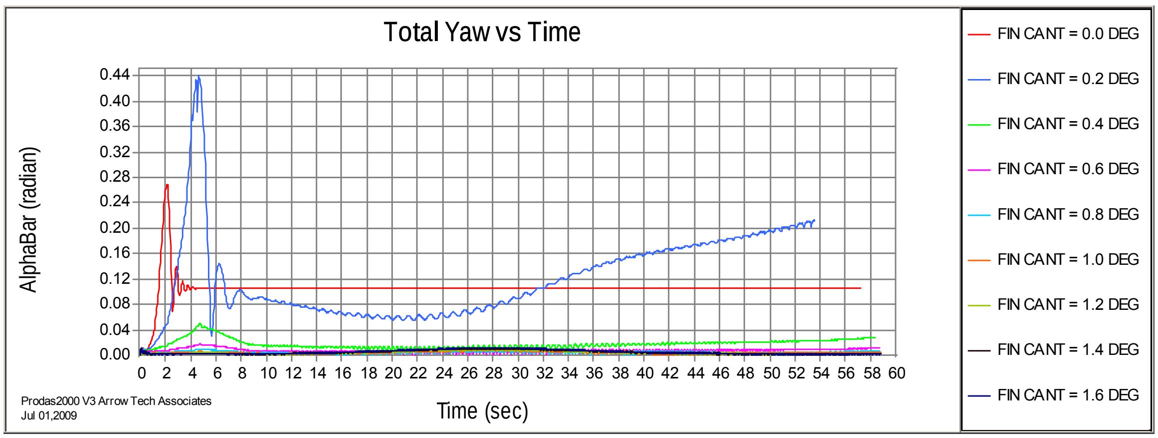


Figure 14. Comparison 2, total yaw versus time.

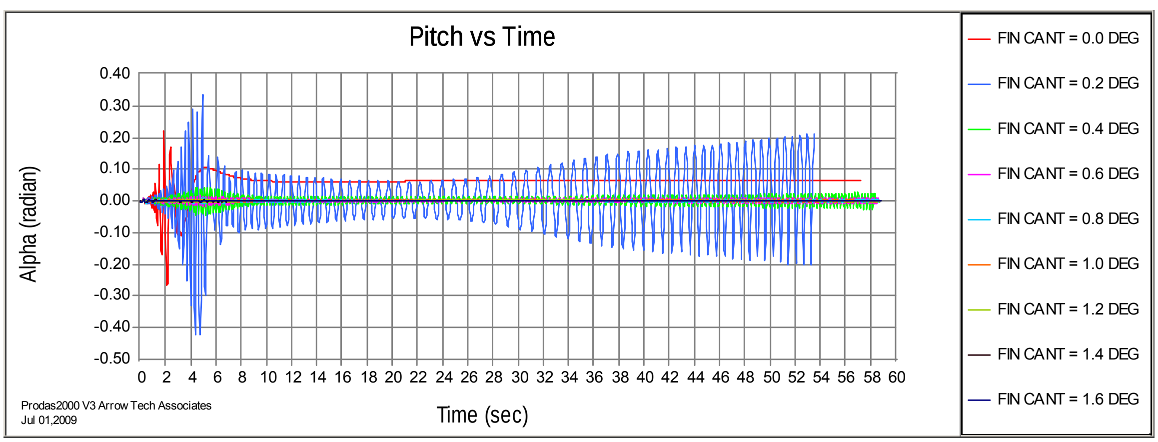


Figure 15. Comparison 2, pitch versus time.

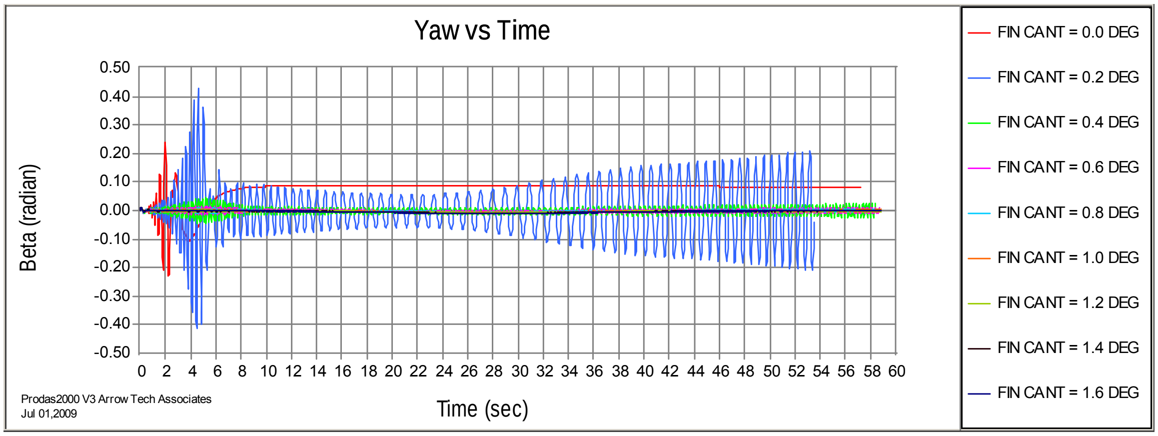


Figure 16. Comparison 2, yaw versus time.

6.3 Comparison of Fin Cant from 0.050° to 0.145° (Comparison 3)

Using the bounds established in the section prior, comparisons 3 and 4 simulated fin cant configurations from 0.050° to 0.145° and from 0.147° to 0.250°, respectively. With the exception of the 0.050° fin cant, all of the configurations approached the resonance frequency as shown in figure 17. This is also illustrated in figure 18 with the increases in total yaw. Most notably, the fin cant of 0.145° demonstrated a rapid increase (spike) in yaw of 0.93 rad (53°) after 39 s in flight. Unfortunately, this configuration completely destabilized and flew sideways (axis of symmetry orthogonal to trajectory) from then on.

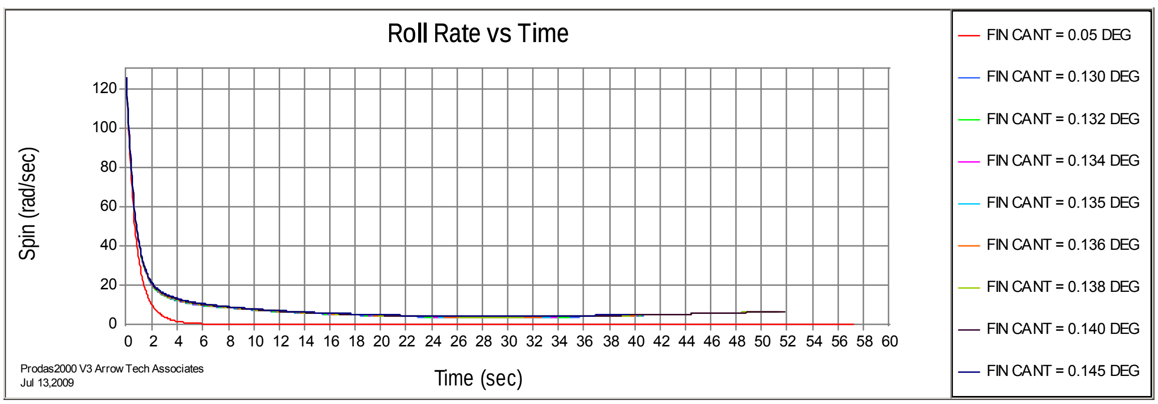


Figure 17. Comparison 3, roll rate versus time.

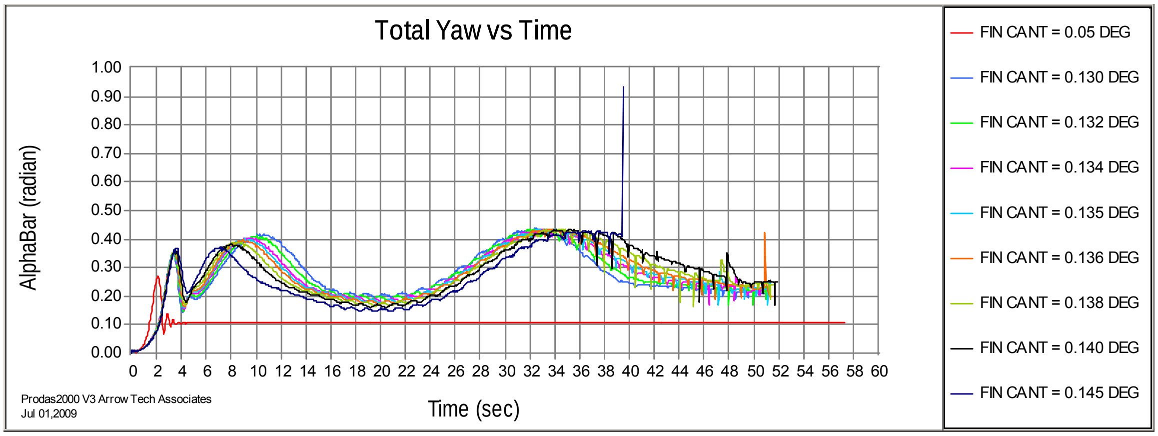


Figure 18. Comparison 3, total yaw versus time.

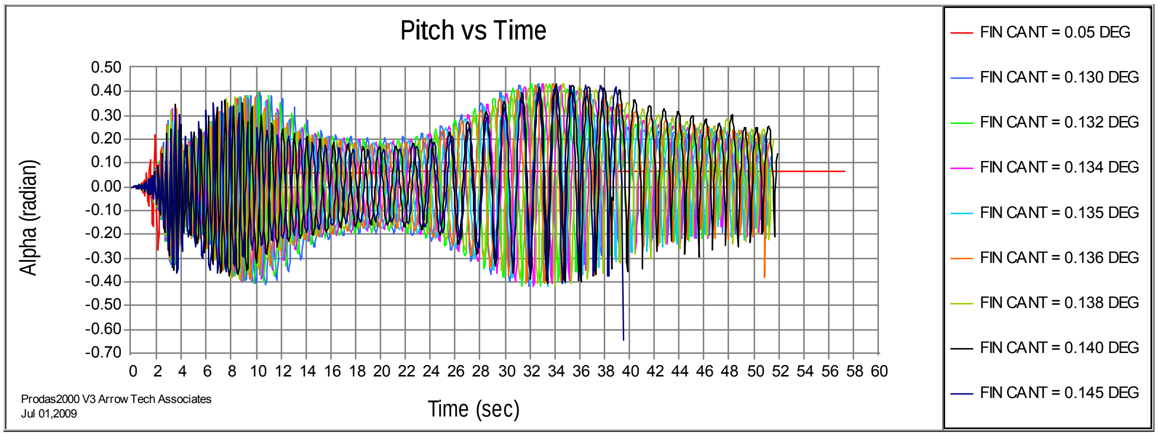


Figure 19. Comparison 3, pitch versus time.

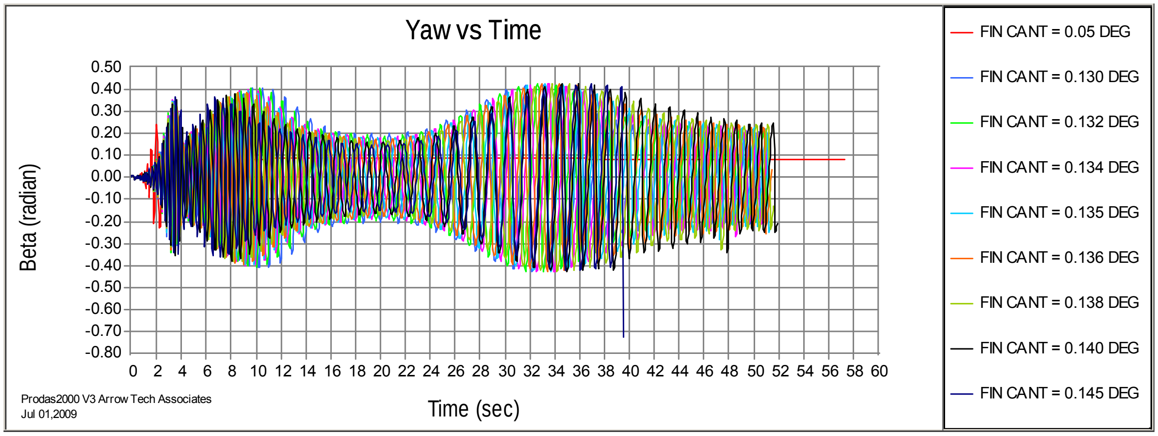


Figure 20. Comparison 3, yaw versus time.

6.4 Comparison of Fin Cant from 0.147° to 0.250° (Comparison 4)

Projectiles with a fin cant below 0.190° rolled close to the resonance frequency as shown in figure 21. For this reason, the total yaw, pitch, and yaw of each projectile were all amplified during flight. This is shown in figures 22–24.

Interestingly, fin cants from 0.147° to 0.180° displayed large yaws between 40 s to the end of flight. Even more interesting is the relationship highlighted between the fin cant and the time at which each spike in yaw occurred. As the fin cant angle increased from 0.147° to 0.180°, so did the time at which each spike occurred. This propagating effect could be useful for initiating large course changes.

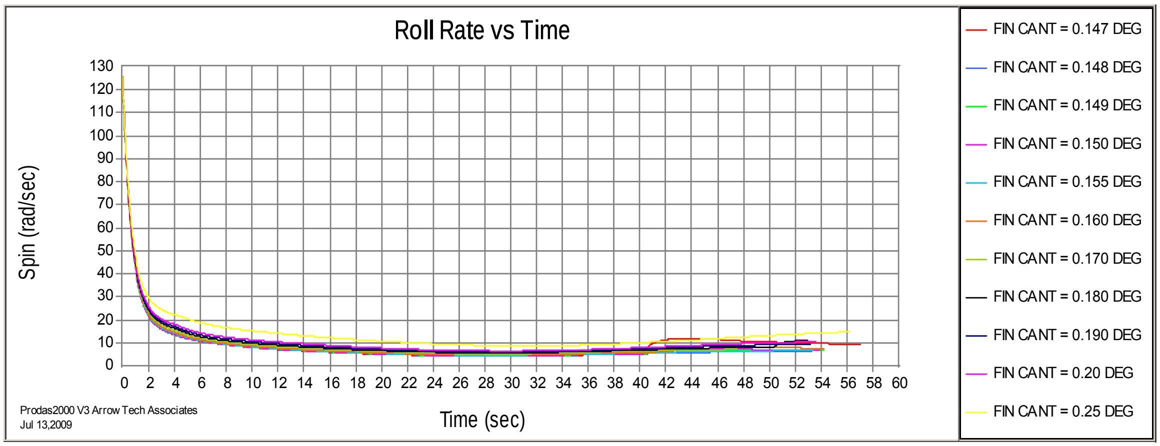


Figure 21. Comparison 4, roll rate versus time.

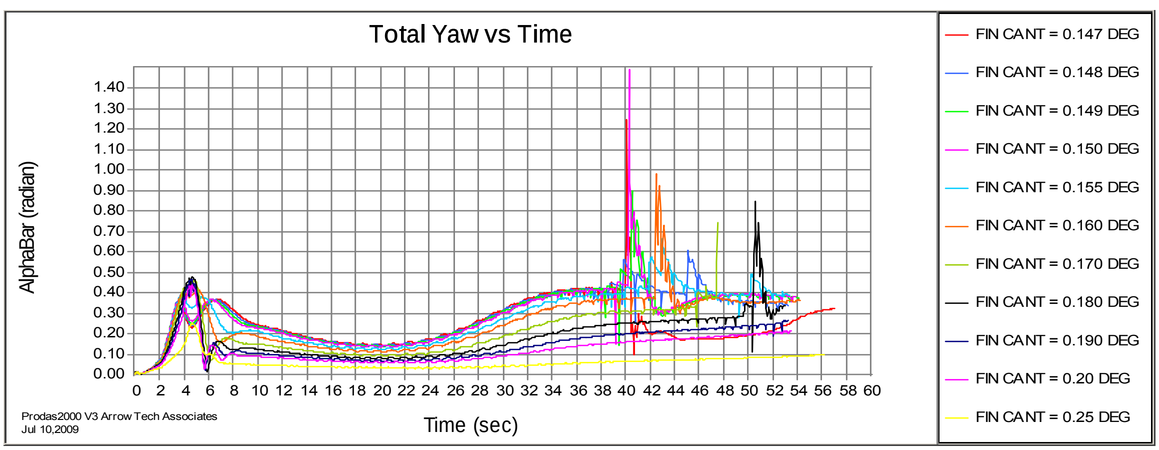


Figure 22. Comparison 4, total yaw versus time.

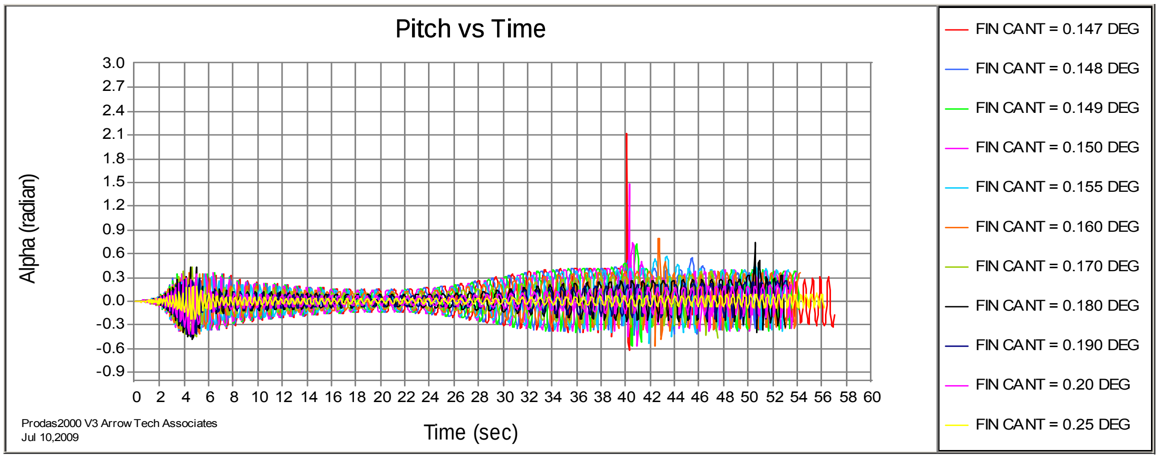


Figure 23. Comparison 4, pitch versus time.

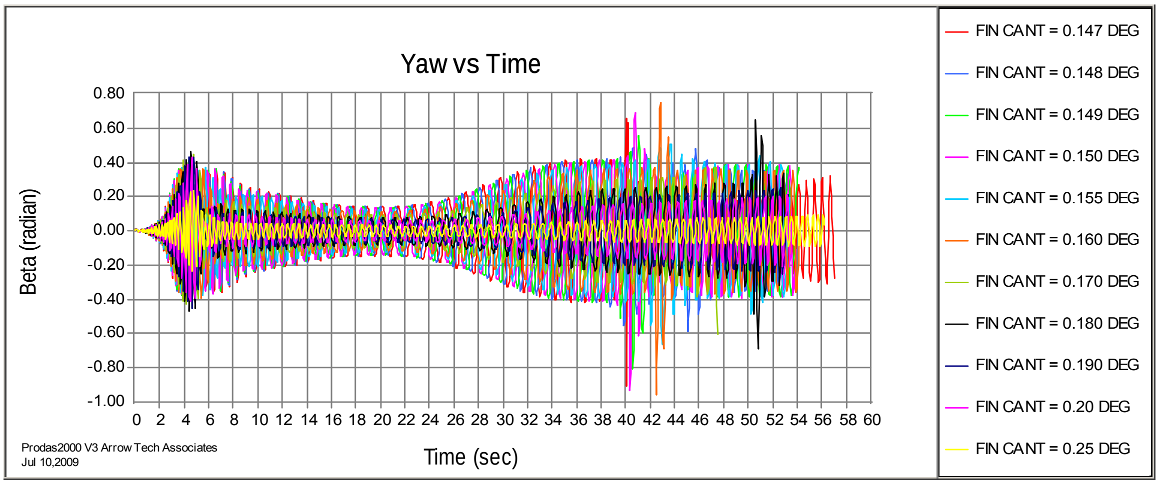


Figure 24. Comparison 4, yaw versus time.

Data from the 6-DOF models used in Comparison 4 were organized into a spreadsheet. Magnitudes of both the minimum and average difference between the projectiles roll and nutation rates were calculated. These values are tabulated in table 4 and plotted in figures 25–26.

Figure 25 indicates a quadratic relationship between the average difference and the fin cant angle. The lowest average difference was 126.3299 deg/s and it occurred with a fin cant of 0.148°.

Figure 26 plots the minimum difference versus the fin cant. Unlike the average difference, there appears to be no trend amongst the minimum difference and the fin cant. The minimum difference was 0.006285 deg/s and occurred with a fin cant of 0.134°.

Table 4. Minimum and average difference between roll and nutation rates.

Fin Cant (deg)	Minimum Difference (deg/s)	Average Difference (deg/s)
0.148	**0.013469**	**126.3299**
0.147	0.059680	126.3452
0.149	0.064782	126.3626
0.150	0.014096	126.4647
0.145	0.051594	126.5984
0.155	0.031684	127.7976
0.140	0.045536	128.2253
0.138	0.092452	129.3420
0.160	0.055052	130.8619
0.136	0.092546	130.9124
0.135	0.006450	131.7920
0.134	0.006285	132.7526
0.132	0.035789	134.9294
0.130	0.084436	137.4502
0.170	0.091012	143.0290
0.180	0.117175	161.6687
0.190	0.291028	184.8541
0.100	10.37837	208.4009
0.200	1.616351	209.3748
0.050	5.094146	339.9853
0.250	129.8683	348.2287
0.000	2.146484	474.1370
-1.000	32.376840	2378.6320
-2.000	94.695750	5201.8160

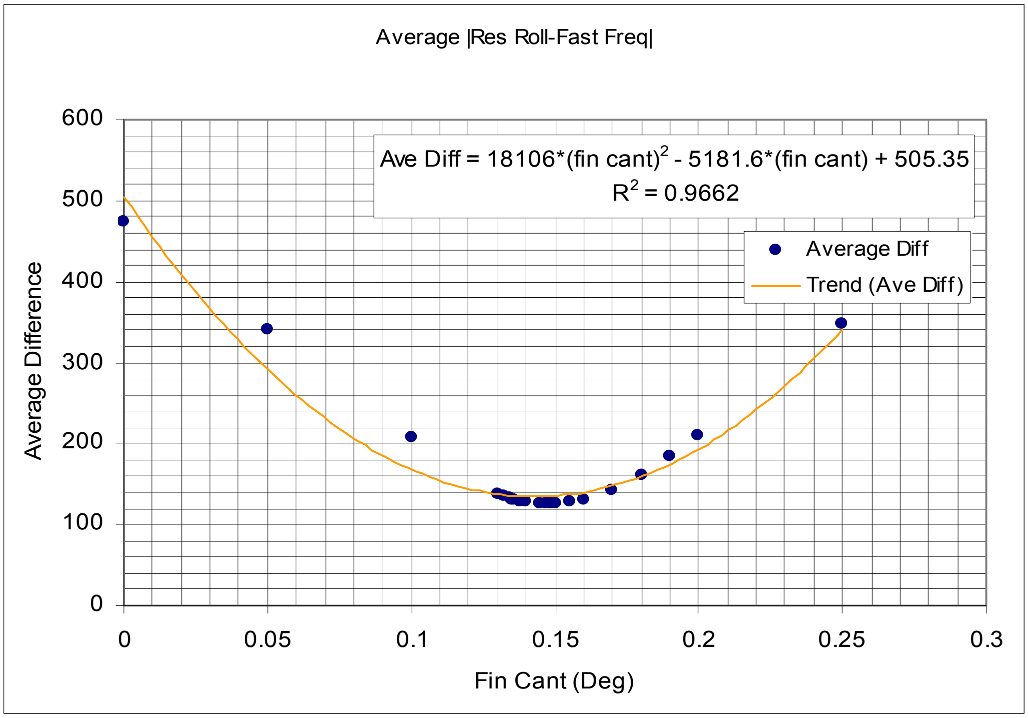


Figure 25. Average difference between roll and nutation rates.

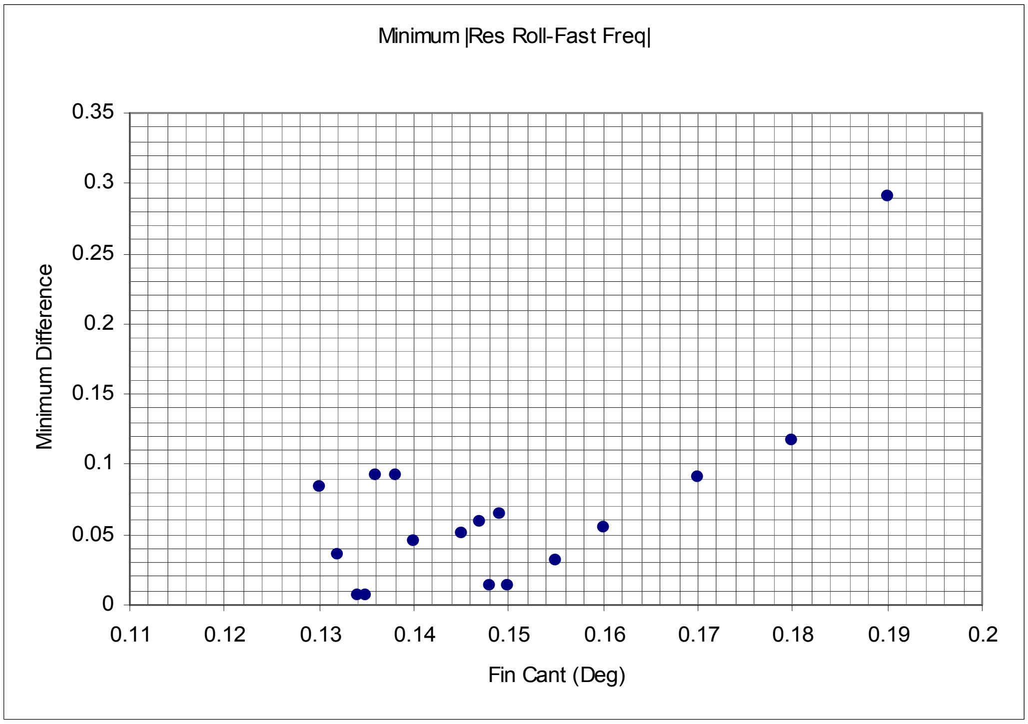


Figure 26. Minimum difference between roll and nutation rates.

6.5 Investigation of 0.148° Fin Cant Configuration

The fin cant of 0.148° was selected for further study. This configuration was selected, because it exhibited the lowest average difference between roll and nutation frequencies. On average, the projectile's roll rate of 9.27 rad/s (531 deg/s) was within 11.8% of the nutation frequency of 8.30 rad/s (475 deg/s) during flight. After 10 s of flight, the projectile rolled at 7.0 rad/s (401 deg/s) and was within 1.1% of its lowered nutation frequency, 6.92 rad/s (397 deg/s). Although this may be unsuitable for fielded projectiles due to instability, for the purpose of this study it was ideal because of the large angles achieved.

The configuration using a fin cant of 0.148° was compared to the control configuration with trim angles of 0° and 6°. Both control configurations were used to show that the trim angle had an insignificant affect between the two trajectories. As a result, the control configurations were virtually indistinguishable.

The projectile maintained a fairly constant roll rate throughout its trajectory. Figure 27 confirms this with the roll rate bounded between 0 and 10 rad/s after 6 s of flight. This roll rate yielded large pitching and yawing motions as shown in figures 28–30. From figure 28, the total yaw is observed to reach 0.61 rad (35°) after 45 s of flight.

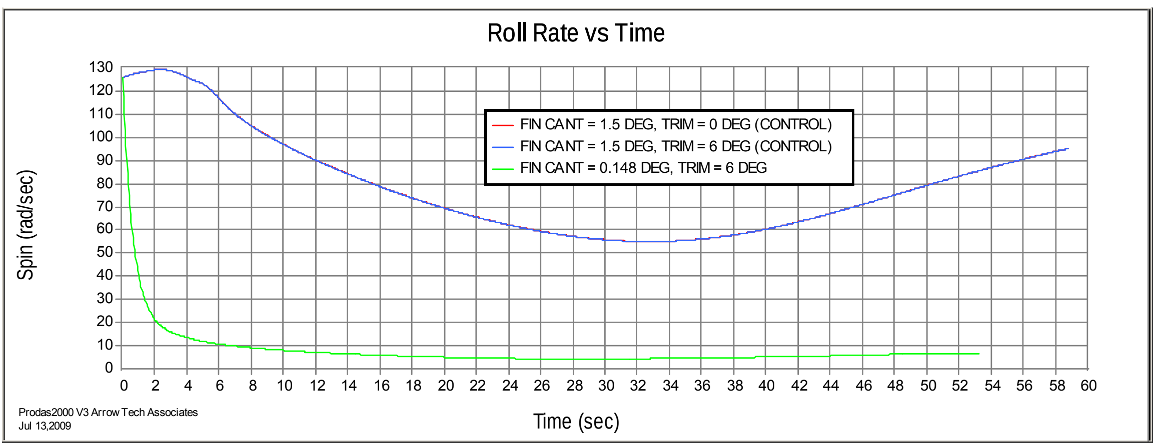


Figure 27. 0.148° fin cant, roll rate versus time.

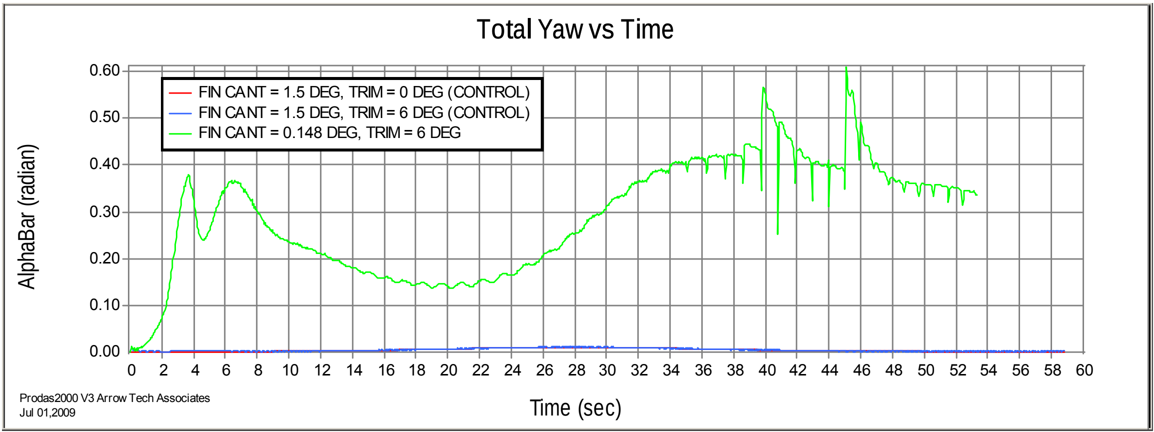


Figure 28. 0.148° fin cant, total yaw versus time.

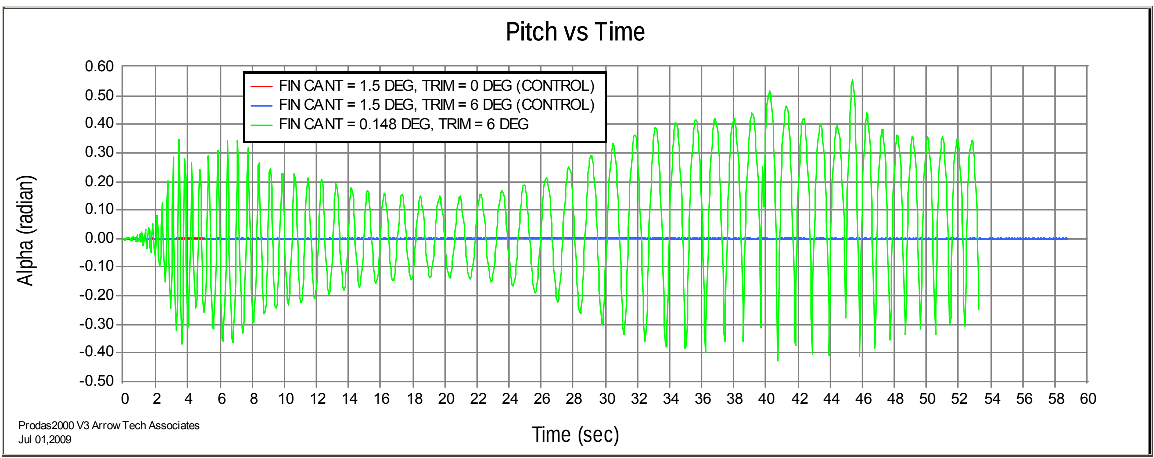


Figure 29. 0.148° fin cant, pitch versus time.

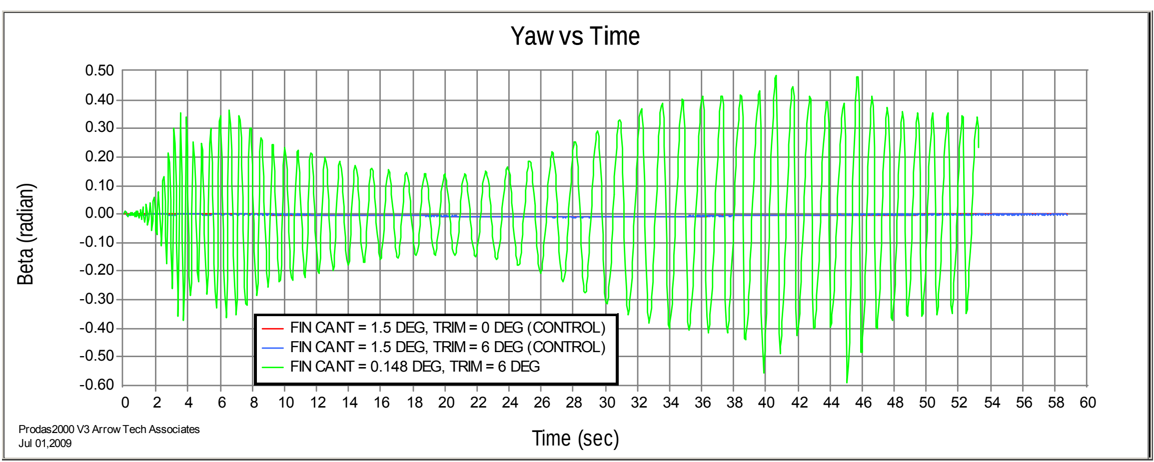


Figure 30. 0.148° fin cant, yaw versus time.

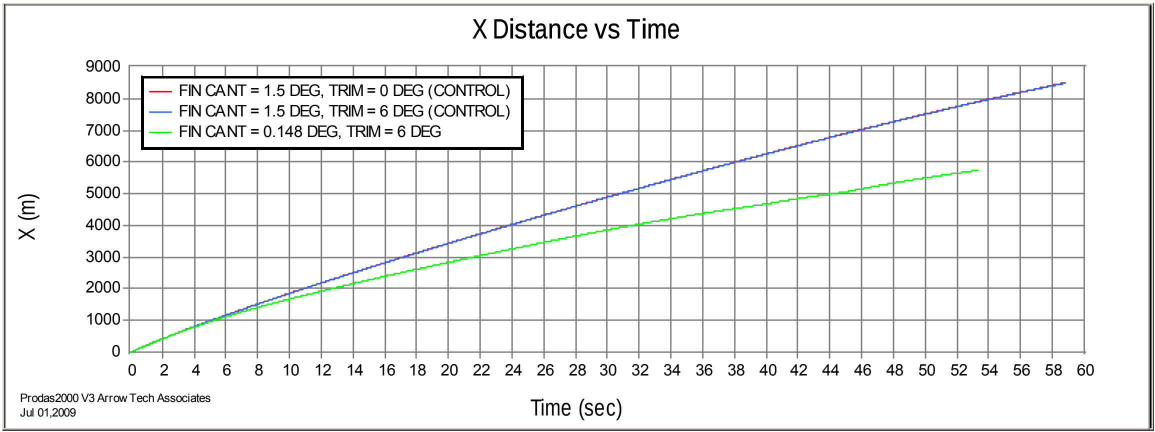


Figure 31. 0.148° fin cant, downrange distance versus time.

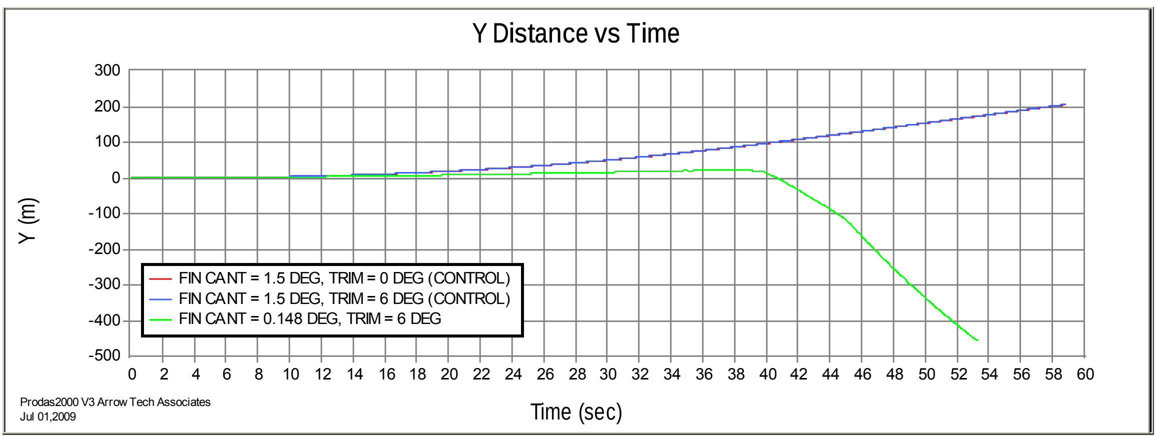


Figure 32. 0.148° fin cant, lateral distance versus time.

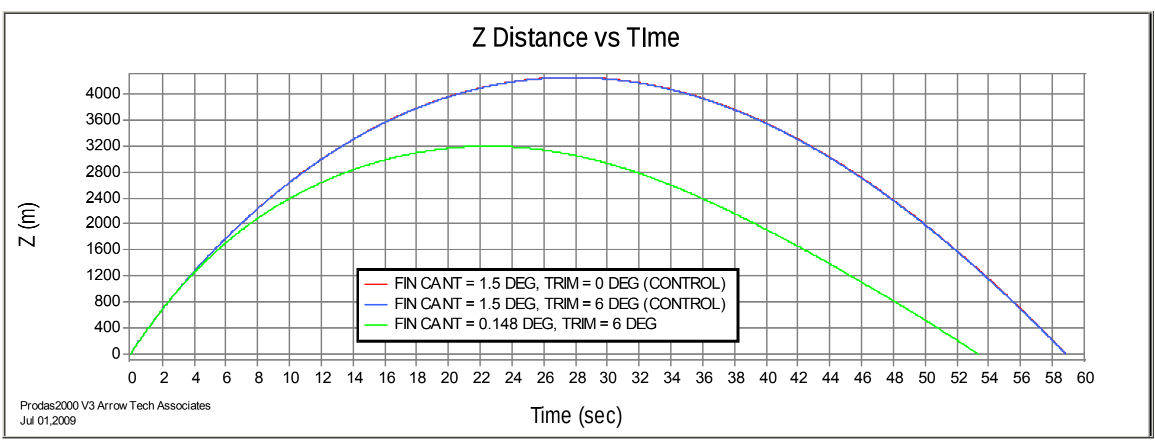


Figure 33. 0.148° fin cant, vertical distance versus time.

Figures 31–33 represent the trajectory of the projectile. Using a right-handed coordinate system located at the gun muzzle's position, the positive x-axis points downrange, the positive y-axis points left, and the positive z-axis points "upward." See figure 34 for illustration of coordinate system.

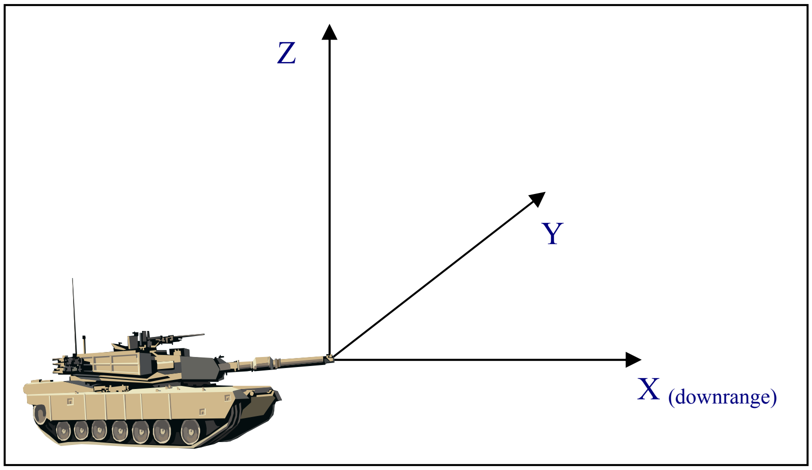


Figure 34. Right-handed coordinate system.

The projectile, which flew in resonance, lost approximately 2500 m in range due to increased drag. This loss in range was accompanied by a 1400-m drop in altitude, and a 600-m change in total side deflection.

6.6 Applied Axial Thrust during Resonance

The VAPP projectile will experience a constant 491.8 N thrust during flight for 3.6 s. If the thrust were applied when the projectile experienced a large yaw, large course changes are possible. To demonstrate this, two projectile configurations were simulated with and without an applied thrust: The control and the 0.148° fin cant configuration. The 491.8 N thrust was applied at 26 s into flight.

Figures 35–37 show trajectories of each projectile, and figures 38–41 show the projectiles attitude. Figure 36 demonstrates the potential of thrusting during resonance. With regards to the 0.148° fin cant configuration, the projectile would have significantly changed course at 40 s into flight. Instead, the thrust applied prior lead the projectile to deviate 22° to the left, since the yawing motion was damped due to the axial thrust. Figures 38–41 illustrate this damping effect.

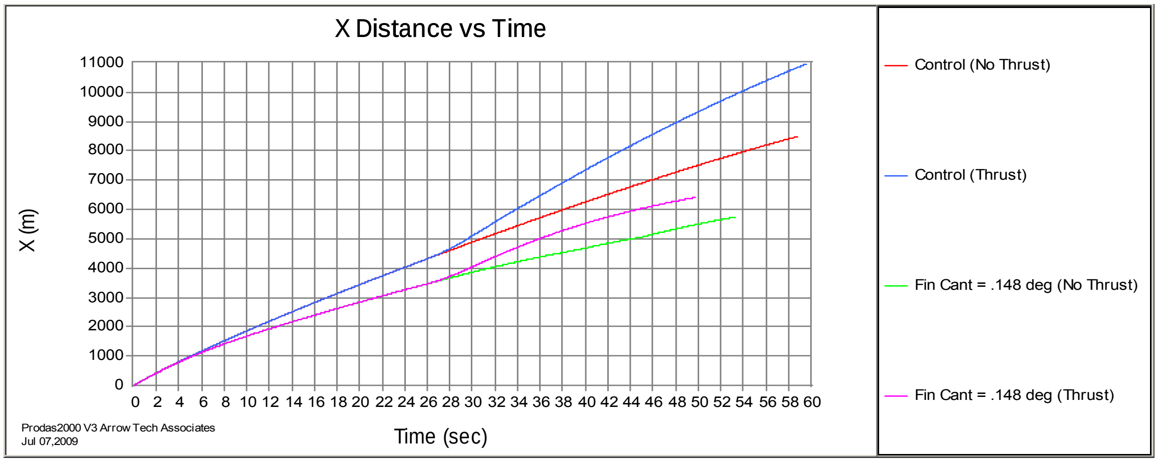


Figure 35. Applied thrust, downrange distance versus time.

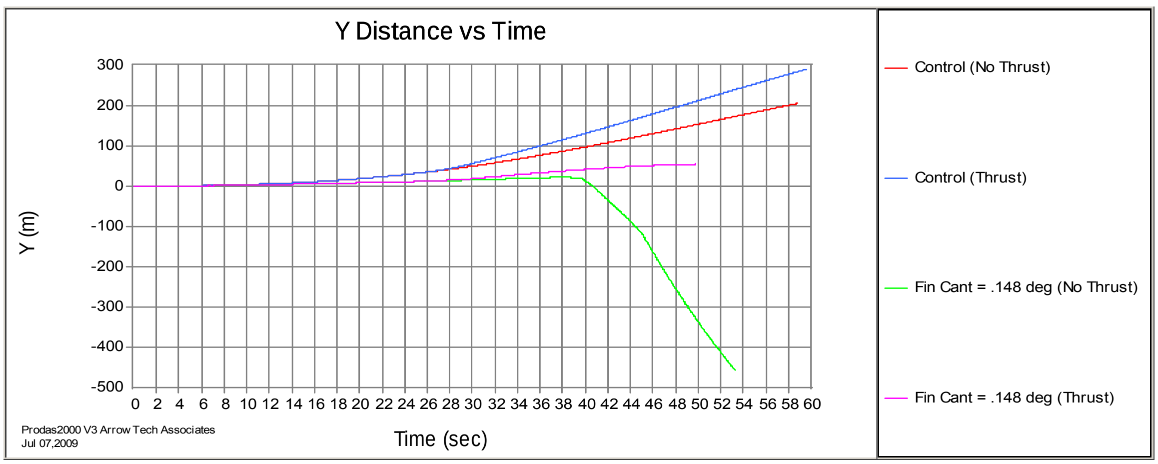


Figure 36. Applied thrust, lateral distance versus time.

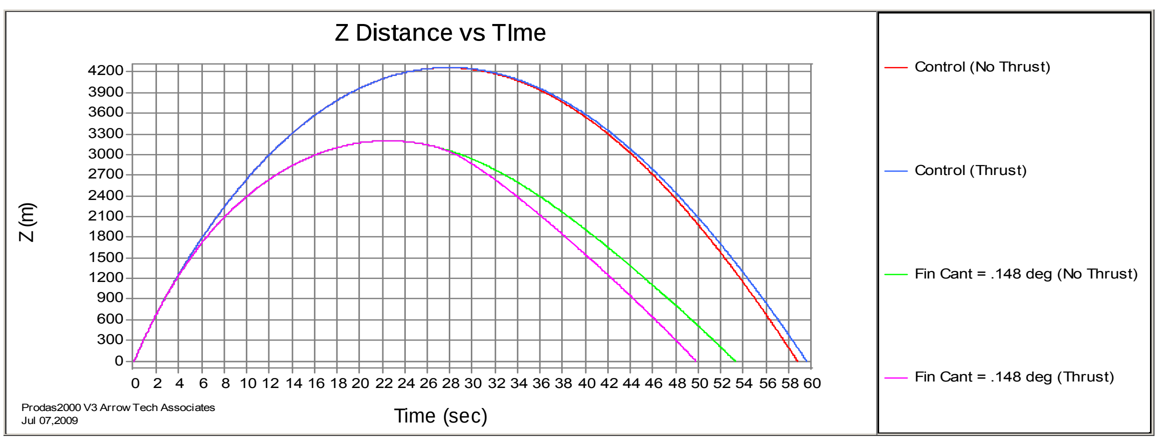


Figure 37. Applied thrust, vertical distance versus time.

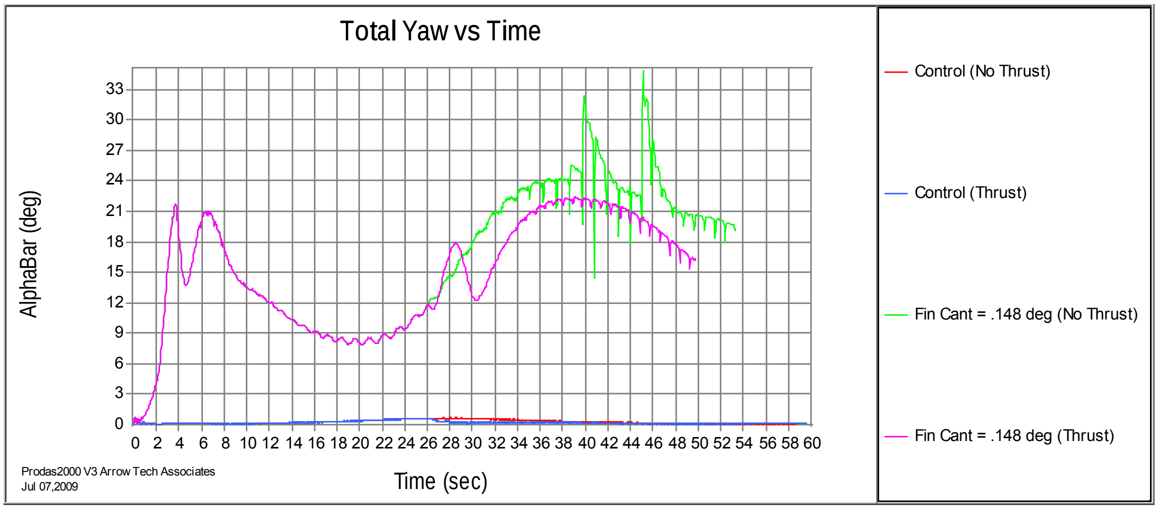


Figure 38. Applied thrust, total yaw versus time.

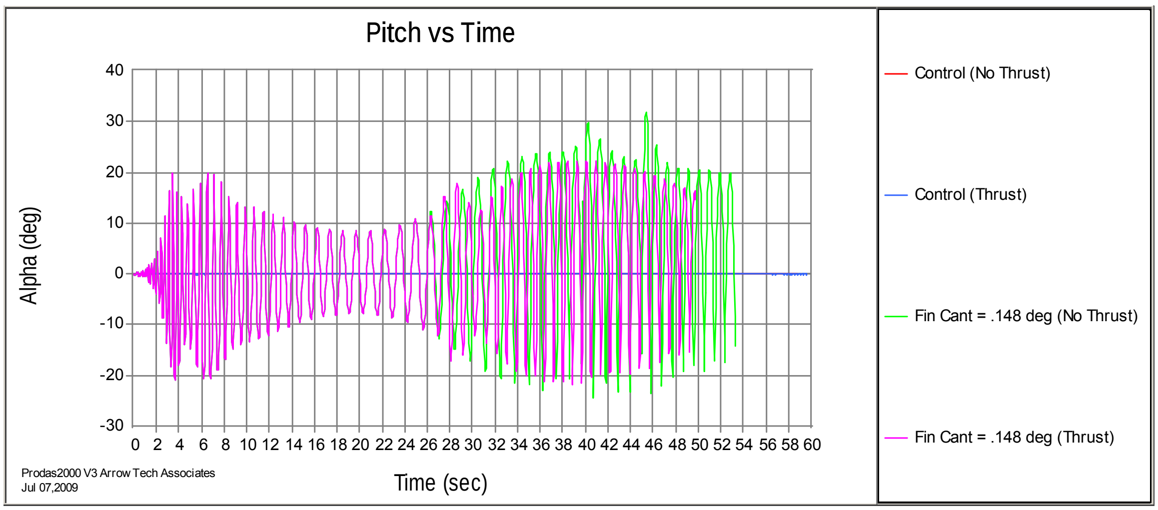


Figure 39. Applied thrust, pitch versus time.

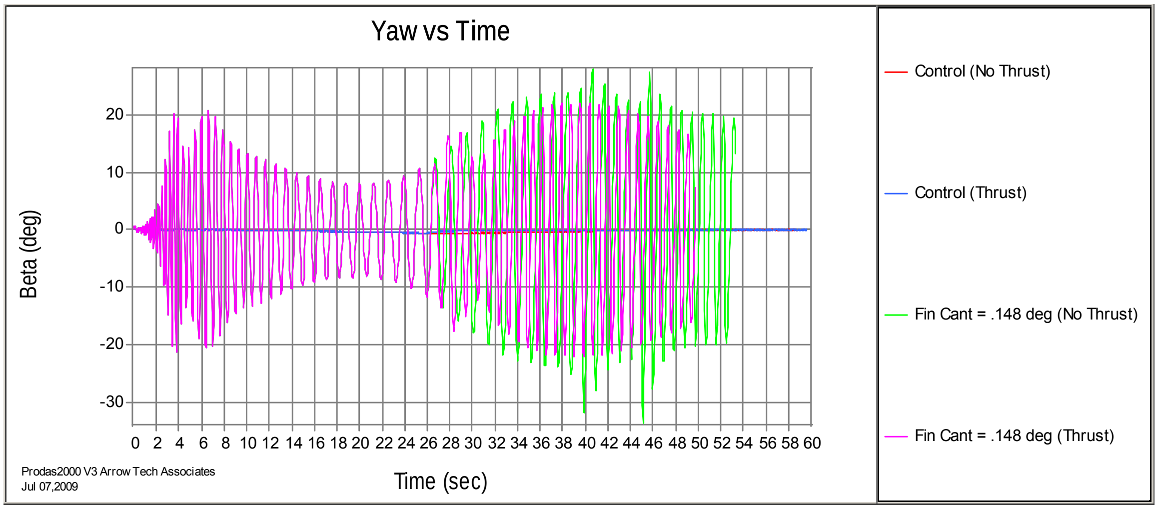


Figure 40. Applied thrust, yaw versus time.

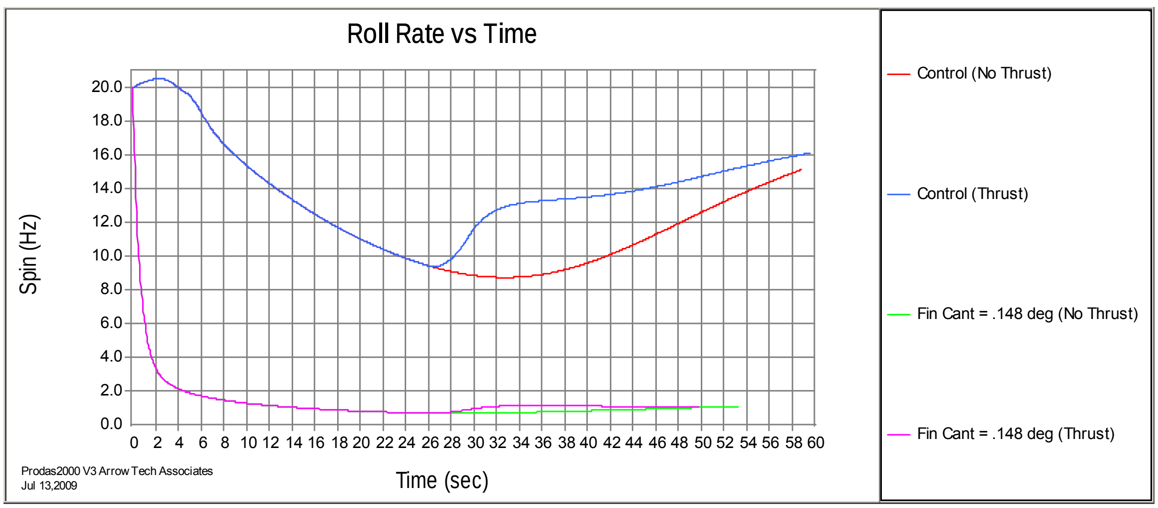


Figure 41. Applied thrust, roll rate versus time.

7. Conclusions

Roll resonance is achieved by rolling a projectile at a rate near its nutation frequency. This was accomplished on the VAPP projectile by modifying the fin cant angle. A fin cant of 0.148° allowed the projectile to roll at 6.92 rad/s (397 deg/s), which was within 1.1% of its nutation frequency after 10 s of flight. This enabled the projectile to experience a 600-m change in total side deflection.

Additionally, a second study was done to demonstrate the effect of an applied axial thrust during resonance. Using the same 0.148° fin cant configuration, a 491.8 N thrust was initiated 26 s into flight for a period of 3.6 s. This single thrust enabled the projectile to deviate from its intended course by 500 m.

Generally, it is good practice to design projectiles with minimum yawing motion for dynamic stability. While I do not suggest flying a projectile in resonance throughout its entire trajectory, for a controlled period of time, this could be useful. Ideally, a projectile would be able to fly steadily, engage in resonance with an accompanied thrust to invoke course change, and then continue on a new stable course. Although it is unrealistic to design actuators for adjusting the projectile's fin cant, it is reasonable to use the VAPP's canard actuators to change the roll rate, placing the projectile in and out of resonance when necessary.

This report was meant to highlight the potential use of roll resonance for course deflection. Further work will include the following:

- Addressing whether equilibrium resonance states exist.
- Addressing whether linear aerodynamic theory can be used to accurately model a projectile during resonance.
- Developing a tri-cyclic model to accurately describe the pitching and yawing motion of a projectile in resonance.
- Using canard actuators to resonate the VAPP projectile.

8. References

1. Murphy, C. H. Response of an Asymmetric Missile to Spin Varying through Resonance. *AIAA Journal* **November 1971**, *9* (11), 2197–2201.

2. Vaughn, H. R. *A Detailed Development of the Tricyclic Theory*; Sandia Laboratory: Albuquerque, NM, February 1968.

3. Nicolaides, J. D. *On the Free Flight Motion of Missiles Having Slight Configurational Asymmetries*; Report 858, AD 26405; Ballistic Research Laboratories: Aberdeen Proving Ground, MD, May 1958.

Appendix A. Nomenclature

The following nomenclature was obtained from reference 1:

A = $-\dfrac{\rho S l^3}{2I_Y}\left(C_{m0} + iC_{n0}\right)$

C_D = drag coefficient

$C_{l\alpha}$ = lift coefficient

C_{M_α} = static moment coefficient

$C_{M_{p\alpha}}$ = Magnus moment coefficient

C_{m0}, C_{n0} = aerodynamic moment coefficients due to asymmetry

$C_{l\delta}$ = roll-producing moment coefficient

C_{l_p} = roll damping moment coefficient

$C_{M_{\dot\alpha}}, C_{M_q}$ = damping moment coefficients

H = $\dfrac{\rho S l}{2m}\left[C_{l\alpha} - C_D - k_t^{-2}\left(C_{M_q} + C_{M_{\dot\alpha}}\right)\right]$

I_Z, I_Y = axial, transverse moments of inertia

k_a = axial radius of gyration $\left(\dfrac{I_Z}{ml^2}\right)^{1/2}$

k_t = transverse radius of gyration $\left(\dfrac{I_Y}{ml^2}\right)^{1/2}$

K_j = amplitude of the j mode (j = 1, 2)

K_3 = amplitude of the response to asymmetric moment

$K_\delta = \dfrac{\rho S l^3}{2I_Z}\delta_f C_{l\delta}$

K_p = $-\dfrac{\rho S l}{2m}\left(k_a^{-2} \cdot C_{lp} + C_D\right)$

l = reference length

m = mass

M_X = axial component of aerodynamic moment

$M_{\tilde{Y}}, M_{\tilde{Z}}$ = transverse components of aerodynamic moment

M = $\dfrac{\rho S l}{2m}\left(k_t^{-2} \cdot C_{M_\alpha}\right)$

p = $\dfrac{d\phi}{dt}$, roll rate

P = $\dfrac{I_Z}{I_Y}\left(\dfrac{pl}{V}\right)$, gyroscopic spin

q = $(s - s_R)\left|\dfrac{\phi_R''}{2}\right|^{1/2}$

$\overline{q}, \overline{r}$ = angular velocity components along $\tilde{Y}, \tilde{Z}$ axis

s = dimensionless distance along flight path $\int_{t_0}^{t} \dfrac{V}{l} dt$

S = reference area

T = $\dfrac{\rho S l}{2m}\left(C_{L_\alpha} + k_a^{-2} \cdot C_{M_{p\alpha}}\right)$

V = magnitude of velocity

$X, \tilde{Y}, \tilde{Z}$ = non-rolling Cartesian coordinate axes

$\overline{\alpha}$ = angle of attack

$\overline{\beta}$ = angle of sideslip

δ_f = control surface deflection

$\theta = 2\overline{K}_p^{-2}\left[e^{-\overline{K}_p q} + \overline{K}_p q - 1\right] = \phi - \overline{\phi}_1 q - \phi_R$

λ_j = damping rate of the j-modal amplitude $\dfrac{K_j'}{K_j}$

$\overline{\mu} = \dfrac{(\overline{q} + i\overline{r})l}{V}$

$\xi = \overline{\beta} + i\overline{\alpha}$, complex angle of attack

ξ_{Δ} = trim response to forcing function

ρ = air density

ϕ = roll rate

ϕ_j = j-modal phase angle, $\phi'_{j0} + \phi'_j s$, j = 1, 2

ϕ_3 = response phase angle $\phi_{30} + \phi$

ϕ'_b = steady state roll rate $= \dfrac{K_\delta}{K_p}$

Superscripts

$\tilde{}$ = components in a non-rolling coordinate system

$(\bar{\ }) = (\_\_)\left|\dfrac{\phi''_R}{2}\right|^{1/2}$

( ) = primes denote derivatives with respect to s

Subscripts

0 = initial value

R = value at resonance

Projectile spin rate

$$\phi' = \phi'_b + (\phi'_0 - \phi'_3)e^{-K_p s}$$

Constant roll rate

Arm magnitude

$K_j = K_{j0}e^{\lambda_j}$ j = 1, 2

Arm damping coefficient

$$\lambda_j = -(H\phi'_j - PT)(2\phi'_j - P)^{-1}$$

Arm spin rate

$\phi'_j = \dfrac{1}{2}\left[P \pm \sqrt{P^2 - 4M}\right]$ J = 1, 2

Trim arm magnitude

$$K_{30} = e^{i\Phi_{30}} = \frac{-iA}{(\phi')^2 - P\phi' + M - i(\phi' H - PT)}$$

Resonance amplitude of trim arm

$$K_{3R} = \left|\frac{-iA}{M - i(\phi' H - PT)}\right| = \left|\frac{A}{\lambda_1 \phi_1' (2 - I_Z / I_Y)}\right|$$

Resonance amplification of trim angle

$$\frac{K_{3R}}{K_{30}} = \left|\frac{\phi_1'}{2\lambda_1}\right|$$

Appendix B. Aerodynamic Coefficients for VAPP Projectile

Table B-1. Aerodynamic coefficients for VAPP projectile.

Mach	Zero Yaw Drag CX0	Normal Force CNa	Forebody Drag CXf	Pitching Moment Cma	Pitch Damping Moment Cmq	Roll Damping Moment Clp	Roll Moment Cld
0.01	0.2296	5.214	0.0999	-1.62	-157	-2.67515	0.038971
0.4	0.2296	5.214	0.0941	-1.62	-179.3	-2.9558	0.038971
0.6	0.2333	5.343	0.0911	-1.651	-190.7	-3.09972	0.039553
0.7	0.2361	5.397	0.0916	-1.657	-197.6	-3.13815	0.039579
0.75	0.2375	5.42	0.0918	-1.659	-202.4	-3.16315	0.039569
0.8	0.2401	5.538	0.0921	-1.788	-207.2	-3.18816	0.039772
0.85	0.244	5.9	0.1101	-2.643	-217.7	-3.23668	0.040037
0.875	0.246	6.082	0.1191	-3.072	-223	-3.26094	0.040273
0.9	0.248	6.261	0.1281	-3.502	-228.3	-3.2852	0.040657
0.925	0.2614	6.375	0.1556	-3.718	-231.9	-3.19856	0.04104
0.95	0.275	6.489	0.1831	-3.932	-235.5	-3.11191	0.041296
0.975	0.3346	6.201	0.2109	-3.098	-239.2	-3.19184	0.04138
1	0.3742	6.064	0.2387	-2.695	-239.6	-3.23219	0.041465
1.025	0.4071	5.973	0.2589	-2.416	-239.1	-3.26265	0.041549
1.05	0.4401	5.883	0.2791	-2.138	-238.7	-3.2931	0.041533
1.1	0.4518	5.809	0.274	-1.52	-234.5	-3.37587	0.041321
1.2	0.4467	5.724	0.2647	-1.514	-227.9	-3.73176	0.041191
1.35	0.445	5.755	0.2552	-1.527	-223.3	-4.28175	0.041213
1.5	0.4444	5.223	0.2447	-1.531	-218.7	-4.83174	0.041213
1.75	0.4108	6.121	0.2297	2.79	-148.1	-1.69693	0.014495
2	0.3888	5.479	0.2223	4.154	-134.2	-1.28613	0.010921
2.25	0.3671	5.262	0.2153	4.608	-129.2	-1.13933	0.009642
2.5	0.3453	5.045	0.2082	5.059	-124.2	-0.99253	0.008363
3	0.3088	4.633	0.1976	5.658	-115.4	-0.78448	0.006551
3.5	0.2902	4.315	0.1978	5.922	-108.3	-0.64897	0.005369
4	0.2716	3.997	0.198	6.184	-101.2	-0.51346	0.004187
4.5	0.2602	3.831	0.1981	6.245	-96.7	-0.45537	0.003681
5	0.2489	3.665	0.1981	6.305	-92.2	-0.39729	0.003176
6	0.2325	3.555	0.1959	6.206	-89.5	-0.37456	0.00298
8	0.2221	3.505	0.2007	6.17	-88.4	-0.36492	0.002905

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 G TECHL LIB T LANDFRIED
BLDG 4600
ABERDEEN PROVING GROUND MD 21005-5066

DIRECTOR
US ARMY RSRCH LAB
ATTN AMSRD ARL RO EV W D BACH
PO BOX 12211
RESEARCH TRIANGLE PARK NC 27709

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