

# *SunFEST*

([www.seas.upenn.edu/~sunfest](http://www.seas.upenn.edu/~sunfest))



**2006**

*SUMMER UNDERGRADUATE FELLOWSHIPS  
IN  
SENSOR TECHNOLOGIES*



*TECHNICAL REPORT  
TR-CST28NOV06  
Center for Sensor Technologies  
University of Pennsylvania  
Philadelphia, PA 19104*

# SUNFEST 2006

SUMMER UNDERGRADUATE FELLOWSHIP IN SENSOR TECHNOLOGIES  
Sponsored by the National Science Foundation (EEC-0244055)

## PREFACE

*This report is the result of 11 undergraduate students' research efforts during the summer of 2006. From May 30<sup>th</sup> through August 06, 2006 11 students from Penn and other colleges participated in the SUNFEST program, which is organized by the Center for Sensor Technologies of the School of Engineering and Applied Science at the University of Pennsylvania. This unique "Summer Experience for Undergraduates in Sensor Technologies" program was initiated in 1986 and has grown considerably in size. It is now recognized as one of the most successful summer programs for undergraduates in the country. I would like to express my sincere gratitude to the National Science Foundation for their continued support for this REU Site, as well as Microsoft Corporation for sponsoring two of our students.*

*The purpose of the SUNFEST program is to provide bright, motivated undergraduate students with the opportunity to become involved in active research projects under the supervision of a faculty member and his graduate student(s). The general area of research concentrates on sensor technologies and includes projects such as materials and technology for sensors, microstructures, smart imagers, and neural networks for sensory processing and robotics. By providing the students with hands-on experience and integrating them with a larger research group where they can work together with other students, the program intends to guide them in their career choices. By exposing the students to the world of research, we hope they will be more inclined to go on for advanced degrees in science and engineering.*

*The students participated in a variety of hands-on workshops in order to give them the tools to do first-rate research or enhance their communication skills. These included "Giving Effective Presentations", "Ethics in Science and Engineering", "Use of Electronic Databases" and "Writing Technical Reports". Students also had plenty of opportunity for social interactions among themselves or with faculty and graduate student advisors.*

*This booklet contains reports from this year's projects, the quality of which testifies to the high level of research and commitment by these students and their supervisors. I would like to express my sincere thanks to the students for their enthusiastic participation; the help of the faculty members, graduate students and support staff is very much appreciated. I would also like to thank Ms. Delores Magobet, Shelley Brown, Sherri Butler, Sid Deliwala, Scott Slavin, and the rest of the ESE staff for their invaluable help in making this program run smoothly.*

Jan Van der Spiegel, Director  
Center for Sensor Technologies

**FINAL REPORT  
2006 SUMMER UNDERGRADUATE FELLOWSHIP  
IN SENSOR TECHNOLOGIES  
Sponsored by the National Science Foundation**

<http://www.ese.upenn.edu/~sunfest/pastProjects/Projects06.html>

**TABLE OF CONTENTS**

<b>LIST OF STUDENT PARTICIPANTS</b>	<b>iv-x</b>
<b>THE ROLE OF LEG DIFFERENTIATION IN HEXAPEDAL RUNNING</b>	<b>1 – 26</b>
<i>Sam Burden, (Electrical Engineering), University of Washington</i>	
<i>Advisors: Dr. Koditschek, Dr. Clark, and Joel Weingarten</i>	
<b>IMPLEMENTATION OF SINTERED LTCC FOR THE FABRICATION OF A 3D CYLINDRICAL MICRO-COMBUSTOR</b>	<b>27 – 42</b>
<i>Jose M. Castillo Colon, (Physics and Electronics), University of Puerto Rico at Humacao</i>	
<i>Advisor: Dr. Jorge Santiago Aviles</i>	
<b>FORMING VESICLES FROM CARBON NANOTUBES</b>	<b>43 – 75</b>
<i>Alexsandra Fridshtand, (Bioengineering), Lehigh University</i>	
<i>Advisor: Dr. Jennifer Lukes</i>	
<b>QUANTITATION OF OXYGEN CONCENTRATION</b>	<b>76 – 91</b>
<i>Shakera Guess, (Chemistry, Spanish) Lincoln University</i>	
<i>Advisor: Dr. Britton Chance</i>	
<b>ELECTROCHEMICAL METHODS FOR THE DETERMINATION OF REDOX PROPERTIES IN THIOL-TERMINATED PORPHYRIN SELF-ASSEMBLED MONOLAYERS</b>	<b>92 – 107</b>
<i>Journee B. Isip (Physics) Columbia University</i>	
<i>Advisor: A.T. Charlie Johnson</i>	
<b>ELECTROMECHANICAL MODELING OF A NEW CLASS OF CONTOUR-MODE AlN MEMS RF RESONATORS</b>	<b>108 – 125</b>
<i>Raúl Pérez Martínez, (Department of Physics and Electronics) University of Puerto Rico, , Humacao</i>	
<i>Advisor: Dr. Gianluca Piazza</i>	
<b>FABRICATION OF PVA MICROPOLARIZE ARRAYS FOR A CMOS IMAGE SENSON</b>	<b>126 - 141</b>
<i>Nathan Lazarus (Electrical Engineering), University of Pennsylvania</i>	
<i>Advisors: Dr. Jan Van der Spiegel, Dr. Viktor Gruev</i>	

<b>PEDIATRIC DYNAMOMETER</b>	<b>142 - 188</b>
<i>Armand O'Donnell, (Electrical Engineering), University of Pennsylvania</i>	
<i>Advisor: Dr. Jay Zemel</i>	
<b>APPLICATION OF FERROFLUID AS A VALVE/PUMP FOR POLYCARBONATE MICROFLUIDIC DEVICES</b>	<b>189 - 205</b>
<i>Helen Schwerdt (Biomedical Engineering), Johns Hopkins University</i>	
<i>Advisors: Professor Haim H. Bau, Jason Thompson</i>	
<b>SPATIAL RESOLUTION OF CHONDROCYTE RESPONSE TO MECHANICAL SIGNALS FOR CARTILAGE TISSUE ENGINEERING</b>	<b>206 - 223</b>
<i>Xiaoning Yuan (Biomedical Engineering) – Duke University</i>	
<i>Advisors: Dr. Robert L. Mauck, Alice H. Huang</i>	

**LIST OF STUDENT PARTICIPANTS  
IN THE SUNFEST PROGRAM SINCE 1986**

**Summer, 2006**

Sam Burden	University of Washington
Jose M. Castillo Colon	University of Puerto Rico at Humacao
Alexsandra Fridshtand	Lehigh University
Shakera Guess	Lincoln University
Journee Isip	Columbia University
Nathan Lazarus	University of Pennsylvania
Armand O'Donnell	University of Pennsylvania
William Peebles	Lincoln University
Raúl Pérez Martínez	University of Puerto Rico at Humacao
Helen Schwerdt	Johns Hopkins University
Xiaoning Yuan	Duke University

**Summer, 2005**

Robert Callan	University of Pennsylvania
David Cohen	University of Pennsylvania
Louie Huang	University of Pennsylvania
Roman Geykhman	University of Pennsylvania
An Nguyen	University of Pennsylvania
Olga Paley	University of California at Berkeley
Miguel Perez Tolentino	University of Puerto Rico at Humacao
Ebenge Usip	University of Southern California
Adam Wang	University of Texas, Austin
Kejia Wu	University of Pennsylvania

**Summer, 2004**

Benjamin Bau	Massachusetts Institute of Technology
Alexander H. Chang	University of Pennsylvania
Seth Charlip-Blumlein	University of Pennsylvania
Ling Dong	University of Rochester
David Jamison	Johns Hopkins University
Dominique Low	University of Pennsylvania
Emmanuel U. Onyegam	University of Texas, Dallas
J. Miguel Ortigosa	Florida Atlantic University
William Rivera	University of Puerto Rico, Mayaguez
Matthew Saucedo	Texas A&M University, Kingsville
Olivia Tsai	Carnegie Mellon University

### **Summer, 2003**

Emily Blem	Swarthmore College
Brian Corwin	University of Pennsylvania
Vinayak Deshpande	University of Virginia
Nicole DiLello	Princeton University
Jennifer Geinzer	University of Pittsburgh
Jonathan Goulet	University of Pennsylvania
Mpitulo Kala-Lufulwabo	University of Pittsburgh
Emery Ku	Swarthmore College
Greg Kuperman	University of Pennsylvania
Linda Lamptey	University of Pennsylvania
Prasheel Lillaney	University of Pennsylvania
Enrique Rojas	University of Pennsylvania

### **Summer, 2002**

Christopher Bremer	Colorado School of Mines
Aslan Ettehadien	Morgan State University
April Harper	Hampton University
Catherine Lachance	University of Pennsylvania
Adrian Lau	University of Pennsylvania
Cynthia Moreno	University of Miami
Yao Hua Ooi	University of Pennsylvania
Amber Sallerson	University of Maryland/Baltimore County
Jiong Shen	University of California-Berkeley
Kamela Watson	Cornell University
John Zelena	Wilkes University

### **Summer, 2001**

Gregory Barlow	North Carolina State University
Yale Chang	University of Pennsylvania
Luo Chen	University of Rochester
Karla Conn	University of Kentucky
Charisma Edwards	Clark Atlanta University
EunSik Kim	University of Pennsylvania
Mary Kutteruf	Bryn Mawr College
Vito Sabella	University of Pennsylvania
William Sacks	Williams College
Santiago Serrano	Drexel University
Kiran Thadani	University of Pennsylvania
Dorci Lee Torres	University of Puerto Rico (Humacao)

### Summer, 2000

Lauren Berryman	University of Pennsylvania
Salme DeAnna Burns	University of Pennsylvania
Frederick Diaz	University of Pennsylvania (AMPS)
Hector Dimas	University of Pennsylvania (AMPS)
Xiomara Feliciano	University of Turabo (Puerto Rico)
Jason Gillman	University of Pennsylvania
Tamara Knutsen	Harvard University
Heather Marandola	Swarthmore College
Charlotte Martinez	University of Pennsylvania
Julie Neiling	University of Evansville, Indiana
Shiva Portanova	University of Pennsylvania

### Summer, 1999

David Auerbach	Swarthmore
Darnel Degand	University of Pennsylvania
Hector E. Dimas	University of Pennsylvania
Ian Gelfand	University of Pennsylvania
Jason Gillman	University of Pennsylvania
Jolymar Gonzalez	University of Puerto Rico – Mayaguez
Kapil Kedia	University of Pennsylvania
Patrick Lu	Princeton University
Catherine Reynoso	Hampton University
Philip Schwartz	University of Pennsylvania

### Summer, 1998

Tarem Ozair Ahmed	Middlebury University
Jeffrey Berman	University of Pennsylvania
Alexis Diaz	Turabo University, Puerto Rico
Clara E. Dimas	University of Pennsylvania
David Friedman	University of Pennsylvania
Christin Lundgren	Bucknell University
Heather Anne Lynch	Villanova University
Sancho Pinto	University of Pennsylvania
Andrew Utada	Emory University
Edain (Eddie) Velazquez	University of Pennsylvania

### Summer, 1997

Francis Chew	University of Pennsylvania
Gavin Haentjens	University of Pennsylvania
Ali Hussain	University of Pennsylvania
Timothy Moulton	University of Pennsylvania
Joseph Murray	Oklahoma University
O'Neil Palmer	University of Pennsylvania
Kelum Pinnaduwage	University of Pennsylvania
John Rieffel	Swarthmore College
Juan Carlos Saez	University of Puerto Rico, Cayey

### Summer, 1996

Rachel Branson	Lincoln University
Corinne Bright	Swarthmore College
Alison Davis	Harvard University
Rachel Green	Lincoln University
George Koch	University of Pennsylvania
Sandro Molina	University of Puerto Rico-Cayey
Brian Tyrrell	University of Pennsylvania
Joshua Vatsky	University of Pennsylvania
Eric Ward	Lincoln University

### Summer, 1995

Maya Lynne Avent	Lincoln University
Tyson S. Clark	Utah State University
Ryan Peter Di Sabella	University of Pittsburgh
Osvaldo L. Figueroa	University of Puerto Rico-Humacao
Colleen P. Halfpenny	Georgetown University
Brandeis Marquette	Johns Hopkins
Andreas Olofsson	University of Pennsylvania
Benjamin A. Santos	University of Puerto Rico-Mayaguez
Kwame Ulmer	Lincoln University

### Summer, 1994

Alyssa Apsel	Swarthmore College
Everton Gibson	Temple University
Jennifer Healy-McKinney	Widener University
Peter Jacobs	Swarthmore College
Sang Yoon Lee	University of Pennsylvania
Paul Longo	University of Pennsylvania
Laura Sivitz	Bryn Mawr College
Zachary Walton	Harvard University

### Summer, 1993

Adam Cole	Swarthmore College
James Collins	University of Pennsylvania
Brandon Collings	Hamilton University
Alex Garcia	University of Puerto Rico
Todd Kerner	Haverford College
Naomi Takahashi	University of Pennsylvania
Christopher Rothey	University of Pennsylvania
Michael Thompson	University of Pennsylvania
Kara Ko	University of Pennsylvania
David Williams	Cornell University
Vassil Shtonov	University of Pennsylvania

### Summer, 1992

James Collins	University of Pennsylvania
Tabbetha Dobbins	Lincoln University
Robert G. Hathaway	University of Pennsylvania
Jason Kinner	University of Pennsylvania
Brenelly Lozada	University of Puerto Rico
P. Mark Montana	University of Pennsylvania
Dominic Napolitano	University of Pennsylvania
Marie Rocelie Santiago	Cayey University College

### Summer, 1991

Gwendolyn Baretto	Swarthmore College
Jaimie Castro	University of Puerto Rico
James Collins	University of Pennsylvania
Philip Chen	University of Pennsylvania
Sanath Fernando	University of Pennsylvania
Zaven Kalayjian	University of Pennsylvania
Patrick Montana	University of Pennsylvania
Mahesh Prakriya	Temple University
Sean Slepner	University of Pennsylvania
Min Xiao	University of Pennsylvania

### Summer, 1990

Angel Diaz	University of Puerto Rico
David Feenan	University of Pennsylvania
Jacques Ip Yam	University of Pennsylvania
Zaven Kalayjian	University of Pennsylvania
Jill Kawalec	University of Pennsylvania
Karl Kennedy	Geneva
Jinsoo Kim	University of Pennsylvania
Colleen McCloskey	Temple University
Faisal Mian	University of Pennsylvania
Elizabeth Penadés	University of Pennsylvania

### Summer, 1989

Peter Kinget	Katholieke University of Leuven
Chris Gerdes	University of Pennsylvania
Zuhair Khan	University of Pennsylvania
Reuven Meth	Temple
Steven Powell	University of Pennsylvania
Aldo Salzberg	University of Puerto Rico
Ari M. Solow	University of Maryland
Arel Weisberg	University of Pennsylvania
Jane Xin	University of Pennsylvania

### Summer, 1988

Lixin Cao	University of Pennsylvania
Adnan Choudhury	University of Pennsylvania
D. Alicea-Rosario	University of Puerto Rico
Chris Donham	University of Pennsylvania
Angela Lee	University of Pennsylvania
Donald Smith	Geneva
Tracey Wolfsdorf	Northwestern University
Chai Wah Wu	Lehigh University
Lisa Jones	University of Pennsylvania

### Summer, 1987

Salman Ahsan	University of Pennsylvania
Joseph Dao	University of Pennsylvania
Frank DiMeo	University of Pennsylvania
Brian Fletcher	University of Pennsylvania
Marc Loinaz	University of Pennsylvania
Rudy Rivera	University of Puerto Rico
Wolfram Urbanek	University of Pennsylvania
Philip Avelino	University of Pennsylvania
Lisa Jones	University of Pennsylvania

**Summer, 1986**

Lisa Yost  
Greg Kreider  
Mark Helsel

University of Pennsylvania  
University of Pennsylvania  
University of Pennsylvania

# The Role of Leg Differentiation in Hexapedal Running

Sam Burden

Dr. Koditschek, Dr. Clark, and Joel Weingarten  
SUNFEST REU 2006

## ABSTRACT

Designing a robot that can autonomously traverse a variety of terrain types is difficult. For this reason, one may refer to nature for inspiration and produce robots that mimic biological organisms. EduBot is one such device, a hexapedal robot that resembles a cockroach. Building a robot that mimics a living creature offers numerous advantages, including design efficiency and locomotive stability. In addition, mathematical models that describe animal walking and running can be applied to the device, so an entire existing body of analysis can be used to characterize the robot's movement; this saves time and increases intuitive understanding. One such model, the spring-loaded inverted pendulum (or SLIP) model for animal running, has been successfully applied to EduBot's forward motion. However, like many animals, EduBot deviates from this model in that it exhibits a rolling motion about its forward axis. The goal of the current research is to describe this motion with a simple model and use the model to motivate modifications to the robot that will eliminate the roll. To characterize the roll, we measure the ground reaction forces from each of the robot's legs with a force plate and use a video-based motion capture system to record the robot's trajectory through space. We show that the maximum roll appears to decrease as the middle legs are made stiffer, which matches the prediction from the model.

# Contents

<b>1 INTRODUCTION</b>	<b>3</b>
1.1 OVERVIEW . . . . .	4
<b>2 BACKGROUND</b>	<b>4</b>
2.1 SPRING-LOADED INVERTED PENDULUM MODEL . . . . .	4
<b>3 CONSERVATIVE MODEL FOR ROLL</b>	<b>5</b>
3.1 PRELIMINARY JUSTIFICATION . . . . .	5
3.2 STANCE EQUILIBRIUM . . . . .	7
3.3 HYBRID DYNAMICAL EQUILIBRIUM . . . . .	9
3.3.1 ANALYTICAL METHODS . . . . .	9
3.3.2 NUMERICAL METHODS . . . . .	12
<b>4 DAMPED MODEL FOR ROLL</b>	<b>14</b>
4.1 ELIMINATING ROLL . . . . .	14
<b>5 RESULTS</b>	<b>16</b>
5.1 SIMULATION FORCES AND TRAJECTORIES . . . . .	18
5.2 ROBOT FORCES AND TRAJECTORIES . . . . .	18
<b>6 DISCUSSION</b>	<b>20</b>
<b>7 RECOMMENDATIONS</b>	<b>20</b>
<b>8 ACKNOWLEDGMENTS</b>	<b>21</b>
<b>A SIMPLE POINCARÉ MAP PROBLEM</b>	<b>23</b>
<b>B FORCE PLATE</b>	<b>25</b>
<b>C NUMERICAL METHODS</b>	<b>25</b>
<b>D TRAJECTORY MEASUREMENTS</b>	<b>26</b>

# 1 INTRODUCTION

Calling upon nature to inform the design of a robotic device is a common technique in robotics; there exist numerous insect-like robots [1] [2], biologically-inspired implements like whiskers are common extensions for robotic sensing [3], and investigations into biological locomotion have potential applications to robotic movement [4].

The benefits of biomimesis in legged locomotion are numerous. First, a biological organism's gait is often efficient and passively stable. As a consequence, their artificial counterparts frequently exhibit these same qualities [5, 1]. Second, mathematical models created to describe a mimicked biological organism can be readily applied to the mimicking robotic device [6] [2]. Aside from simplifying analysis by permitting the use of established mathematical tools, this crossover provides a useful intuitive context for studying the robot.

One such biomimetic device is RHex, a cockroach-like hexapedal robot [1]; it is fast, highly mobile, and relatively stable [7]. A model from biological investigations which has found fruitful application to RHex is the spring-loaded inverted pendulum (SLIP) model for running and hopping [8]. In the SLIP model, the organism in question is abstracted as having only a single spring-loaded leg. This leg contacts the ground at one end and is attached to the body's mass at the other. The SLIP model has been successfully applied to a wide variety of animals [9], as well as RHex [6]. Applying the SLIP model to RHex's motion is useful for two reasons. First, it provides an underlying intuition for why the robot moves so efficiently and so stably. Second, it provides a simple enough *template* for the robot's motion that actual analysis can be performed on the stability [5, 10], providing a mathematical explanation for the observed robustness of the robot's gait. We work here with a smaller version of RHex called EduBot whose design and movement is similar to its predecessor. We are thus motivated to seek similarly simple and descriptive models for other aspects of EduBot's motion with the dual purpose of understanding the basic mechanical properties of the robot intuitively and of providing a rigorous justification for its design.

In particular, EduBot exhibits a seemingly periodic and stable *rolling motion* about the forward axis. It is unclear what, exactly, is causing this lateral roll. More importantly, it is unclear how this roll affects the overall forward motion. It is possible the rolling contributes to stability or efficiency, but it is also perfectly possible that the rolling motion adversely affects both. We present here a simple model for the roll and use the model to inform a modification to the robot's hardware intended to eliminate the roll.

Various techniques have been developed for modeling and analyzing equilibrium gaits. We take as a simple and canonical example the passive walker [11] [12] and attempt to perform a similar sort of analysis using techniques from the field [13] [14]. Since we find that our model's dynamics cannot be integrated directly, we are forced to pose the problem in less direct terms in an attempt to prove gait properties [5, 10] [15].

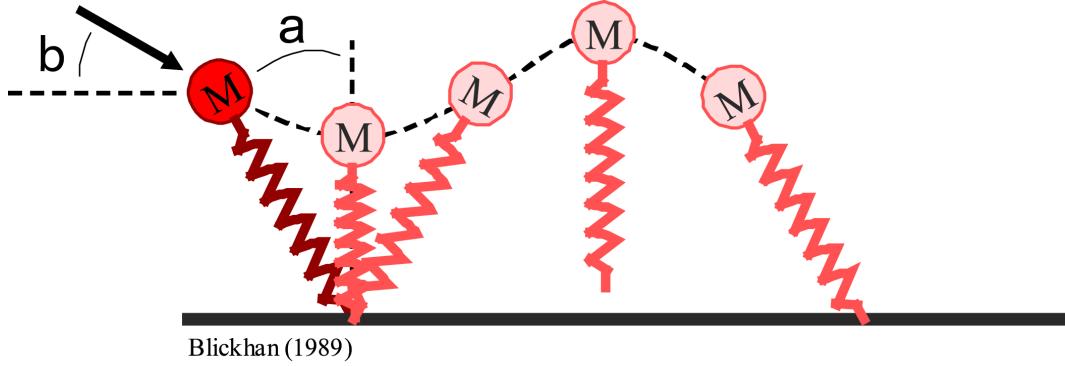


Figure 1: Example trajectory for the Spring-Loaded Inverted Pendulum Model for animal running.

## 1.1 OVERVIEW

We begin with a discussion of the SLIP model for animal running, which is meant to introduce the reader to the type of modeling problems that have been studied before. We then present some data recorded with EduBot and note some of the salient characteristics of the robot’s gait. Then we develop a conservative model for EduBot’s roll and discuss some analytical and numerical results. Finally, we describe a damped model for the robot’s rolling motion and present numerical results that suggest a certain physical modification to EduBot will eliminate the roll and present the results of the physical modification.

## 2 BACKGROUND

### 2.1 SPRING-LOADED INVERTED PENDULUM MODEL

We were motivated in the design and analysis of our rolling model by the success of the simple Spring-Loaded Inverted Pendulum (or SLIP) hybrid dynamical model for animal running. SLIP has been applied to a variety of animals [9], and even artificial creatures [2] [6], and analysis of the model has yielded insight into the fundamental nature of RHex and EduBot’s stability [5, 10].

The SLIP model consists of a single point mass attached to one end of a spring; see Figure 1. This mass  $M$  is projected toward the ground with velocity  $\vec{v} = (v \cos \beta, v \sin \beta)$  while the spring is fixed at an angle  $\alpha$  from the  $y$ -axis. When the spring strikes the ground, the point of contact is fixed. The spring then compresses, pivots, and uncompresses. When the spring reaches its nominal length  $l_0$ , it leaves the ground and the mass-spring system experiences a flight phase.

During flight, the system’s dynamics are transparent. Namely,

$$x(t) = x_0 + \dot{x}_0 t, \quad y(t) = y_0 + \dot{y}_0 t - \frac{1}{2} g t^2, \quad (1)$$

where  $(x_0, \dot{x}_0, y_0, \dot{y}_0)$  represents the system's state at liftoff. During stance, the system can be described via its Lagrangian  $L = T - V$  in terms of the mass's position  $(x, y)$ ; specifically,

$$T = \frac{1}{2}M(\dot{x}^2 + \dot{y}^2), \quad V = gMy + \frac{1}{2}(l_0 - \sqrt{x^2 + y^2})^2, \quad (2)$$

which leads to the equations of motion for the mass during stance,

$$\ddot{x} = x \frac{k}{m} \left( \frac{l_0}{\sqrt{x^2 + y^2}} - 1 \right), \quad \ddot{y} = y \frac{k}{m} \left( \frac{l_0}{\sqrt{x^2 + y^2}} - 1 \right) - g. \quad (3)$$

These equations cannot be solved analytically as they stand. Koditschek [10] has shown that if gravity is neglected during stance (corresponding physically to the use of relatively stiff springs), then a closed-form solution exists, up to the evaluation of a (difficult) integral.

When working with the conservative SLIP model, one typically only considers equilibrium trajectories, which are those with a constant forward velocity during consecutive flight phases.

### 3 CONSERVATIVE MODEL FOR ROLL

Noting the success the SLIP model has had in describing the robot's sagittal-plane motion and the success researchers have had in producing analytical results about the stability of the model, we attempt to apply similar methods of analysis to the robot's roll about its forward axis.

It seems possible, if not plausible, that this rolling motion is due primarily to an imbalance in weight distribution during running. Specifically, EduBot utilizes an *alternating tripod gait* when running: the two front legs on one side of the body are in contact with the ground while only the middle leg touches down on the other. Since the legs are compliant, this imbalance likely compresses one more than the other, resulting in a lateral rolling motion.

To model this behavior, we imagine two massless linear springs attached to either side of a rod of mass  $M$ , moment of inertia  $J$ , and width  $2w$  (see Figure 2); for the majority of our experiments, we assume that mass is distributed evenly throughout EduBot. Assuming the robot's legs are very similar due to a fairly precise construction process (supported by preliminary data), the springs in the model have identical nominal lengths,  $l_0$ . But as there are two legs in contact with the ground on one side of the robot while only one on the other, the spring constants  $k_1$  and  $k_2$  are in general unequal.

#### 3.1 PRELIMINARY JUSTIFICATION

As EduBot runs, it takes full advantage of its rigid body's six degrees of freedom: the robot rolls, twists, pitches, slides, jumps, and runs. To simplify the analysis, we proceed by assuming that the jumping and rolling motions are decoupled from the rest and analyze them exclusively. This is a strict assumption, since the full-body dynamics certainly have an effect on the rolling motion.

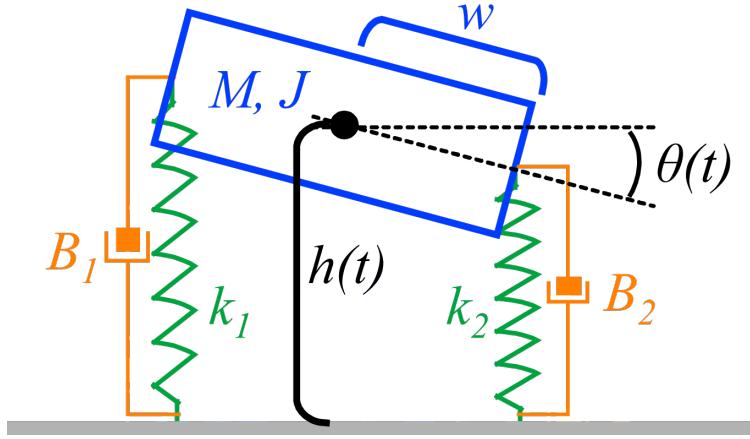


Figure 2: Simple spring-mass model for EduBot’s rolling motion. In general,  $k_1 \neq k_2$ . We set  $B_1 = B_2 = 0$  when we wish to study a conservative model.

When running at equilibrium, EduBot follows roughly periodic rolling and bounding trajectories. Figure 3 shows such a trajectory (see Appendix D) next to one obtained from the conservative model discussed here. There are a few distinctive qualitative behaviors to note about EduBot’s trajectory: first, the center of mass oscillates up and down with a fixed frequency; second, the robot’s roll has a distinct, periodic form that the model roughly captures.

The previous discussion does not completely specify the model, as there are some constraints we must impose on the dynamics. First, we fix the center of mass of the rod to only move in the vertical direction (as if fixed on a rail), and allow the ‘feet’ of the springs to slide frictionlessly across the ground. This removes one degree of freedom from the system, and is a reasonable simplification, since the robot doesn’t move much laterally on average. Second, it doesn’t make any sense to allow the rod to flip around—if it rotates beyond  $\pm\frac{\pi}{2}$ , the model is nonphysical. We get around this by simply not allowing simulations or analysis to proceed into that region of state space. Third, we allow the springy legs to come off the ground, and thus for the rod to attain a *flight phase*. When one or both legs are off the ground, the dynamics should alter accordingly. This puts the model into the analytically challenging realm of *hybrid dynamics* [13], but is necessary to capture the dynamics we are interested in.

Now we have an analytical choice to make about the set of coordinates for our two degree-of-freedom system. Two obvious choices are to record the two spring lengths ( $x_1, x_2$ ) or the height of the center of mass of the rod and the rod’s rotation ( $h, \theta$ ); we found the equations in  $(h, \theta)$  space to be more amenable.

Note that  $\mathbf{X} = \{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}\}$  is a linear vector space, whereas  $\Theta = \{(h, \theta) \in \mathbb{R} \times e^{i\mathbb{R}}\}$  isn’t. The transformation  $\mathbf{X} \leftrightarrow \Theta$  is therefore nonlinear. In fact, a simple computation yields (with  $\theta$  increasing counterclockwise)

$$h = \frac{x_1 + x_2}{2}, \quad \sin \theta = \frac{x_2 - x_1}{2w}, \quad (4)$$

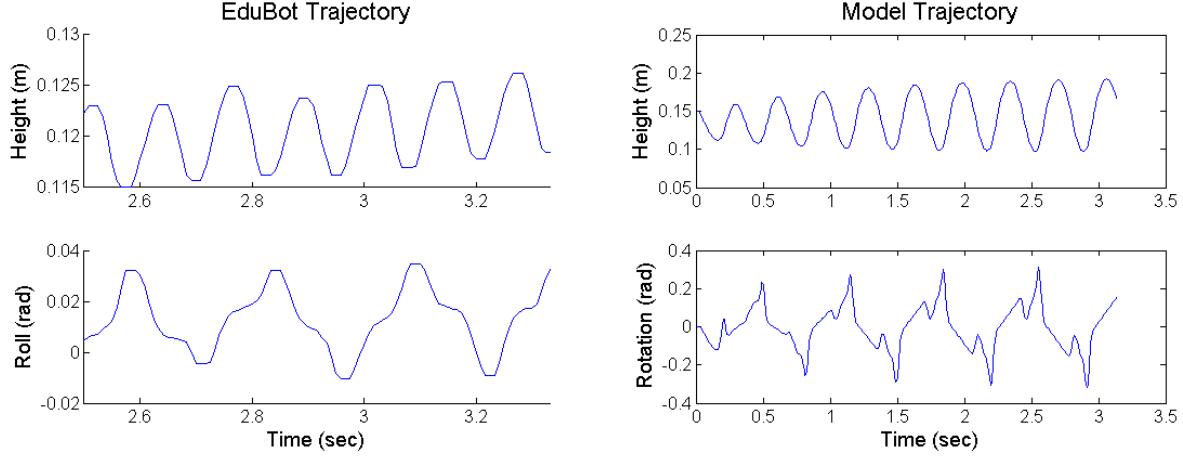


Figure 3: Rolling and height trajectories from EduBot data and from simulation; experimental data was filtered (see Appendix D). There is significant deviation in both coordinates, but the model qualitatively captures some of the distinctive behavior exhibited by the robot.

a decidedly nonlinear transformation in  $\theta$ .

Now that we have settled on a coordinate system and have determined all the model's parameters, we can write down and analyze the equations of motion for the rod. We choose to find these equations using the *Lagrangian*  $L = T - V$ , where  $T$  is the kinetic energy of the system and  $V$  the potential in the generalized coordinates. (Use of the Lagrangian is not necessary; the same result follows using Newton's laws directly.) Our equations of motion then become

$$\frac{\partial L}{\partial h} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{h}} = 0, \quad \frac{\partial L}{\partial \theta} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{\theta}} = 0. \quad (5)$$

Analysis of the model during stance yields

$$T = \frac{1}{2}M\dot{h}^2 + \frac{1}{2}J\dot{\theta}^2, \quad V = Mgh + \frac{1}{2}k_1(l_0 - (h - w \sin \theta))^2 + \frac{1}{2}k_2(l_0 - (h + w \sin \theta))^2. \quad (6)$$

Solving the equations of motion for  $\ddot{h}$  and  $\ddot{\theta}$ , we find

$$\begin{aligned} \ddot{h} &= \frac{1}{M}(-gM + w \sin \theta(k_1 - k_2) - (h - l_0)(k_1 + k_2)), \\ \ddot{\theta} &= \frac{1}{J}(w \cos \theta((h - l_0)(k_1 - k_2) - w \sin \theta(k_1 + k_2))). \end{aligned} \quad (7)$$

### 3.2 STANCE EQUILIBRIUM

We now turn to the task of finding stationary points of the system defined by eq. 7 when the robot never leaves the ground; specifically, we're interested in fixing  $\dot{\theta} \equiv 0$ . Intuitively, it seems there can be only a few, and it seems unlikely that there would be any dynamic solutions (i.e. solutions with  $\dot{h}(t) \neq 0$  for some  $t$ ). We therefore seek to back up this intuition with precise mathematical reasoning.

**STATIC SOLUTION** The static solution that ensures  $\dot{\theta} \equiv 0$  can be found by applying static analysis to the system. When stationary, the springs will have to support the entire weight of the block, so  $k_1\Delta x_1 + k_2\Delta x_2 = Mg$ . An additional equation relating the  $\Delta x$ 's comes from observing that the sum of torques acting on the block must be zero at equilibrium, and therefore  $k_1\Delta x_1 \cos\theta - k_2\Delta x_2 \cos\theta = 0$ . Combining these two equations (and excluding the case<sup>1</sup>  $\theta = \pm\frac{\pi}{2}$ ), we see that  $\Delta x_i = \frac{gM}{2k_i}$ ,  $i = 1, 2$ . Using eq. 4, the resulting equilibrium condition is

$$h_0 = l_0 - \frac{gM}{4} \left( \frac{1}{k_1} + \frac{1}{k_2} \right), \quad \dot{h}_0 = 0, \quad \sin\theta_0 = \frac{gM}{4w} \left( \frac{1}{k_1} - \frac{1}{k_2} \right), \quad \dot{\theta}_0 = 0. \quad (8)$$

Substituting these initial values into eq. 7 yields  $\ddot{h} = \ddot{\theta} = 0$ , so the system is at static equilibrium for all time. Note, conveniently, that if  $k_1 \neq k_2$ , these equations imply

$$h_0 = l_0 - \frac{gM}{2}, \quad \dot{h}_0 = 0, \quad \theta_0 = 0, \quad \dot{\theta}_0 = 0, \quad (9)$$

exactly the static equilibrium condition expected for a simple spring-mass system.

**DYNAMIC SOLUTION** At this point, the question arises: are there any dynamic solutions that ensure  $\dot{\theta} \equiv 0$ ? Using physical intuition, it would seem that if one found a point where the sum of torques was zero, raising or lowering the mass would cause a torque imbalance to appear, thus giving the body a nonzero angular velocity. This argument is nearly rigorous, and can probably be made quite precise, but we prefer instead to turn to the equations of motion to find a mathematical justification for this intuitive explanation.

From eq. 7, we can solve for  $h$  in terms of  $\ddot{h}$  and  $\sin\theta$ :

$$h = \frac{1}{k_1 + k_2} \left[ -gM - \ddot{h}M + k_1(H + w \sin\theta) + k_2(H - w \sin\theta) \right]. \quad (10)$$

Also from eq. 7, assuming  $\ddot{\theta} \equiv 0$  (required for  $\dot{\theta} \equiv 0$ ) yields

$$\sin\theta = \frac{(k_1 - k_2)(h - l_0)}{w(k_1 + k_2)}. \quad (11)$$

Now, substituting eq. 10 into eq. 11,

$$\sin\theta = \frac{M(g + \ddot{h})}{4w} \left( \frac{1}{k_1} - \frac{1}{k_2} \right). \quad (12)$$

Note that if  $\ddot{h} \equiv 0$ , this is exactly the case we found for static equilibrium. If  $\ddot{h}(t_0) = \eta \neq 0$ , then  $\ddot{h} \equiv \eta$ , because otherwise  $\theta$  changes in time, which means  $\dot{\theta} \neq 0$  at some time. But it would require an infinite amount of energy to ensure  $\ddot{h} \equiv \eta \neq 0$ , so this case is nonphysical. Therefore *there does not exist* a system trajectory that sets  $\dot{\theta} \equiv 0, \dot{h}(t) \neq 0$ .

---

<sup>1</sup>This is physically justified, since we don't intend to analyze the case where the robot rolls onto its side.

### 3.3 HYBRID DYNAMICAL EQUILIBRIUM

Now that we have shown no dynamic stance equilibria can exist, we turn our attention to the generally more difficult realm of *hybrid dynamical systems*, where the governing equations of motion are allowed to change as functions of state. We use the formalism given by Guckenheimer [13], which treats a hybrid system as a collection of 4-tuples  $(F_\alpha, V_\alpha, \mathcal{H}_\alpha^\beta, \mathcal{T}_\alpha^\beta)$  indexed by  $\alpha, \beta \in \mathbf{D}$ , where  $\mathbf{D}$  is a set of dynamical regimes. In this framework,  $F_\alpha : \mathbb{R}^5 \rightarrow \mathbb{R}^4$  governs how the system progresses through state space in a certain regime  $\alpha$  and can be written implicitly as the solution to a system of differential equations,  $V_\alpha \subset \mathbb{R}^4$  represents the region of state space reachable in  $\alpha$ ,  $\mathcal{H}_\alpha^\beta : \mathbb{R}^4 \rightarrow \mathbb{R}$  are *threshold functions* that go to zero when the system transitions from regime  $\alpha$  to  $\beta$ , and  $\mathcal{T}_\alpha^\beta : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  are *transition functions* that describe how to transition between  $V_\alpha$  and  $V_\beta$ . In the model presented here, there are four regimes: flight, left leg stance, right leg stance, and full stance, denoted by  $\mathcal{F}$ ,  $\mathcal{L}$ ,  $\mathcal{R}$ , and  $\mathcal{S}$ , respectively.

#### 3.3.1 ANALYTICAL METHODS

**AN APPROXIMATION** To approach the problem of finding equilibrium gaits analytically, we first make an approximation. Since the robot tilts less than  $15^\circ$  in the regime we are interested in, we make a small-angle approximation,  $\sin \theta \mapsto \theta$ . This has the effect of reducing our system to a fourth-order system of linear differential equations, which vastly simplifies the analysis.

**EQUATIONS OF MOTION** Our first step is to describe the equations of motion that govern each regime, the  $F_\alpha$ 's. During flight, we can solve explicitly for  $z(t) = (h(t), \dot{h}(t), \theta(t), \dot{\theta}(t))$ . Indeed,

$$F_{\mathcal{F}}(t, (h, \dot{h}, \theta, \dot{\theta})) = \begin{bmatrix} h_k + \dot{h}_k t - \frac{1}{2}gt^2 \\ 0 \\ \theta_k + \dot{\theta}_k t \\ 0 \end{bmatrix}, \quad (13)$$

where  $(h_k, \dot{h}_k, \theta_k, \dot{\theta}_k)$  is the state of the system at the  $k^{th}$  liftoff event.

During stance, we can put the equations in the form  $\dot{z}(t) = Az + b$ , where  $A$  is  $(4 \times 4)$  and  $b$  is  $(4 \times 1)$ . During full stance,

$$\begin{bmatrix} \dot{h}(t) \\ \ddot{h}(t) \\ \dot{\theta}(t) \\ \ddot{\theta}(t) \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{M}(k_1 + k_2) & 0 & -\frac{w}{M}(k_1 - k_2) & 0 \\ 0 & 0 & 0 & 1 \\ \frac{w}{J}(k_1 - k_2) & 0 & -\frac{w^2}{J}(k_1 + k_2) & 0 \end{bmatrix} \begin{bmatrix} h(t) \\ \dot{h}(t) \\ \theta(t) \\ \dot{\theta}(t) \end{bmatrix}_{\mathcal{S}} + \begin{bmatrix} 0 \\ H(k_1 + k_2) - gM \\ 0 \\ wH(k_1 - k_2) \end{bmatrix}. \quad (14)$$

To compute the partial stance equations of motion, simply omit one of the spring constants in eq. 14: remove  $k_1$  to get the equations for  $\mathcal{R}$ ; remove  $k_2$  to get the equations for  $\mathcal{L}$ . To be explicit,

$$\begin{bmatrix} \dot{h}(t) \\ \ddot{h}(t) \\ \dot{\theta}(t) \\ \ddot{\theta}(t) \end{bmatrix}_{\mathcal{L}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{M}k_1 & 0 & -\frac{w}{M}k_1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{w}{J}k_1 & 0 & -\frac{w^2}{J}k_1 & 0 \end{bmatrix} \begin{bmatrix} h(t) \\ \dot{h}(t) \\ \theta(t) \\ \dot{\theta}(t) \end{bmatrix}_{\mathcal{L}} + \begin{bmatrix} 0 \\ Hk_1 - gM \\ 0 \\ wHk_1 \end{bmatrix}, \quad (15)$$

$$\begin{bmatrix} \dot{h}(t) \\ \ddot{h}(t) \\ \dot{\theta}(t) \\ \ddot{\theta}(t) \end{bmatrix}_{\mathcal{R}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{M}k_2 & 0 & \frac{w}{M}k_2 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{w}{J}k_2 & 0 & -\frac{w^2}{J}k_2 & 0 \end{bmatrix} \begin{bmatrix} h(t) \\ \dot{h}(t) \\ \theta(t) \\ \dot{\theta}(t) \end{bmatrix}_{\mathcal{R}} + \begin{bmatrix} 0 \\ Hk_2 - gM \\ 0 \\ -wHk_2 \end{bmatrix}. \quad (16)$$

**THRESHOLD FUNCTIONS** A threshold function is denoted  $\mathcal{H}_\alpha^\beta(z)$ , where  $\alpha$  is the starting dynamical regime and  $\beta$  the ending one; a transition occurs when  $\mathcal{H}_\alpha^\beta(z) = 0$ . Each  $\mathcal{H}$  is accompanied by a corresponding *transition function*  $\mathcal{T}_\alpha^\beta(z)$  that specifies how to map points in the space associated with  $\alpha$  to the one associated with  $\beta$ , although in all cases presented here,  $\mathcal{T}_\alpha^\beta(z)$  is the identity function.

It is somewhat easier to think about the threshold equations using a different set of coordinates. Specifically, we will rely on the substitutions

$$x_1 = l_0 - (h - w\theta), \quad x_2 = l_0 - (h + w\theta), \quad (17)$$

$$\dot{x}_1 = -\dot{h} + w\dot{\theta}, \quad \dot{x}_2 = -\dot{h} - w\dot{\theta}. \quad (18)$$

When in flight,

$$\begin{aligned} \mathcal{H}_{\mathcal{F}}^S(z) &= (l_0 - h)^2 + (\theta)^2, -\dot{h} - w|\dot{\theta}| \leq 0, \\ \mathcal{H}_{\mathcal{F}}^L(z) &= (x_1)^2, \quad \dot{x}_1 \leq 0, \quad \theta > 0, \\ \mathcal{H}_{\mathcal{F}}^R(z) &= (x_2)^2, \quad \dot{x}_2 \leq 0, \quad \theta < 0. \end{aligned} \quad (19)$$

When in stance,

$$\begin{aligned} \mathcal{H}_{\mathcal{S}}^F(z) &= (l_0 - h)^2 + (\theta)^2, -\dot{h} - w|\dot{\theta}| > 0, \\ \mathcal{H}_{\mathcal{S}}^L(z) &= (x_1)^2, \quad \dot{x}_1 > 0, \quad \theta < 0, \\ \mathcal{H}_{\mathcal{S}}^R(z) &= (x_2)^2, \quad \dot{x}_2 > 0, \quad \theta > 0. \end{aligned} \quad (20)$$

When in left stance,

$$\begin{aligned} \mathcal{H}_{\mathcal{L}}^R(z) &= (l_0 - h)^2 + (\theta)^2, \quad \dot{x}_1 > 0, \quad \dot{x}_2 < 0, \\ \mathcal{H}_{\mathcal{L}}^F(z) &= (x_1)^2, \quad \dot{x}_1 > 0, \quad \dot{x}_2 > 0, \\ \mathcal{H}_{\mathcal{L}}^S(z) &= (x_2)^2, \quad \dot{x}_1 < 0, \quad \dot{x}_2 < 0. \end{aligned} \quad (21)$$

When in right stance,

$$\begin{aligned} \mathcal{H}_{\mathcal{R}}^L(z) &= (l_0 - h)^2 + (\theta)^2, \quad \dot{x}_1 < 0, \quad \dot{x}_2 > 0, \\ \mathcal{H}_{\mathcal{R}}^F(z) &= (x_1)^2, \quad \dot{x}_1 > 0, \quad \dot{x}_2 > 0, \\ \mathcal{H}_{\mathcal{R}}^S(z) &= (x_2)^2, \quad \dot{x}_1 < 0, \quad \dot{x}_2 < 0. \end{aligned} \quad (22)$$

**POINCARÉ MAP** The definition of an equilibrium gait in the hybrid dynamical case departs significantly from the definition given in Section 3.2. Here we require that since the spring constants are exchanged during flight (to model EduBot’s *alternating tripod gait*<sup>2</sup>), each touchdown state  $(h_k, \dot{h}_k, \theta_k, \dot{\theta}_k)$  must be the mirror image of the previous touchdown state  $(h_{k-1}, \dot{h}_{k-1}, \theta_{k-1}, \dot{\theta}_{k-1})$ ; i.e.

$$(h_k, \dot{h}_k, \theta_k, \dot{\theta}_k) = (h_{k-1}, \dot{h}_{k-1}, -\theta_{k-1}, -\dot{\theta}_{k-1}). \quad (23)$$

Note that  $h_k$  and  $\theta_k$ , the height and rotation at the  $k^{\text{th}}$  touchdown, are related:

$$l_0 = h - w \sin |\theta|, \quad (24)$$

Thus, we can define a *Poincaré section*  $\Sigma$  (see Appendix A),

$$\Sigma = \{(h, \dot{h}, \theta, \dot{\theta}) \in \mathbb{R}^4 \mid l_0 = h - w \sin |\theta|\} = \{(h_k, \dot{h}_k, \theta_k, \dot{\theta}_k)\}, \quad (25)$$

Note that the restriction  $l_0 = h - w \sin |\theta|$  implies that  $\Sigma$  is a three-dimensional hypersurface in the four-dimensional state space.

Technically speaking,  $\Sigma$  is only a Poincaré section if it is *transverse* to the system trajectories that we are interested in. Formally, this means if the system touches down at time  $t_k$ ,  $k > 0$ ,  $\phi_{t_k}(x_0) \cdot n_\Sigma(x_k) \neq 0$ , where  $x_k = (h(t_k), \dot{h}(t_k), \theta(t_k), \dot{\theta}(t_k)) \in \Sigma$  is the state of the system at touchdown,  $n_\Sigma(x_k)$  is the normal vector to  $\Sigma$  at  $x_k$ , and  $\phi_t(x_0)$  is a system trajectory that passes through  $\Sigma$ . Intuitively, this means the trajectory cannot travel along  $\Sigma$ , i.e., trajectories intersect  $\Sigma$  at discrete points.

In the physical system this condition is always met. To see this, look at a trajectory  $\phi_t(x_0)$  that intersects  $\Sigma$ . Suppose at time  $t_k$  the body just contacts the ground, yet  $\dot{h}(t_k) = \dot{\theta}(t_k) = 0$  (otherwise at time  $t_k + \epsilon$ ,  $0 < \epsilon \ll 1$ , the system has passed through  $\Sigma$ ). Now, considering a point on  $\Sigma$ ,

$$l_0 = h(t_k) - w \sin |\theta(t_k)|. \quad (26)$$

Substituting this into the equations of motion for the system (eq. 7) we find that if  $\theta = 0$ ,  $\ddot{h} \neq 0$  and the trajectory passes through  $\Sigma$ . If  $\theta \neq 0$ , then  $M\ddot{h} = -gM \pm 2k_i w \sin |\theta|$ , where  $i \in \{1, 2\}$ . Now  $\ddot{h} = 0 \iff \sin |\theta| = \frac{gM}{2k_i w}$ . For the range of parameters we use in our model, this angle lies in the right half-plane, so it is not *a priori* an impossible choice. The choice for  $\sin |\theta|$  implies  $J\ddot{\theta} = -w \cos |\theta| (\pm gM)$ , which, given that the robot doesn’t land on its side, is nonzero. Thus all system trajectories we are interested in are transverse to  $\Sigma$ , so  $\Sigma$  is a Poincaré section.

With the formal definition of what we mean by an equilibria in the hybrid dynamical system, we would like to proceed by finding the equilibria analytically and comment on their stability. Unfortunately, this is such a difficult task that we have made little progress. To use the Poincaré section, we would need to find an explicit formula for the corresponding Poincaré map, which

---

<sup>2</sup>The front and rear legs on one side of EduBot are synchronized with the middle leg on the other. In a gait with a flight phase, this means only three legs are in contact with the ground during stance.

we have not yet been able to do. To find the equilibria by some other insight may be possible, but we have yet to discover the method. Thus, we resort to numerical analysis to investigate the system's equilibria.

### 3.3.2 NUMERICAL METHODS

**VECTOR FIELD** We start the system simulation (see Appendix C) from a mesh of initial conditions  $x_0^{i,j} \in \Sigma$  and record the next point  $x_1^{i,j}$  in the trajectory that intersects  $\Sigma$ . We then make a vector field plot of this data, connecting the initial condition with the final condition<sup>3</sup> by a directed line segment (head toward the final condition).  $\dot{\theta}_k$  is plotted along the  $x$ -axis and  $\theta_k$  along the  $y$ -axis, and since  $\dot{h}_k$  is determined up to a sign difference by  $\dot{\theta}_k$  (by conservation of energy), it determines the color of the vector. Since  $h_k$  is determined by  $\theta_k$  by the threshold equation, it is not contained in the plot.

Unfortunately, due to the unstable nature of the system,  $x_1$  generally lands very far from  $x_0$ , so these plots initially appear as a mess of intersecting lines. To make the plots useful, they show vectors at  $1/10^{th}$  or  $1/20^{th}$  their original magnitude; without carefully noting this fact, the data is distorted. For instance, the scaled-down plots give the impression that the flow field has stable equilibria, which is almost certainly false.

The point of generating these plots is to gain an intuitive understanding of how the system progresses from one set of initial conditions to another. One important goal is to identify equilibrium gaits and determine (roughly in this case) their stability. To identify a fixed point, we iteratively produce a vector field plot over a given range of parameters, choose the smallest vector in the plot, and shrink the field range around that point. After ten iterations, we typically achieve estimates on the fixed point good to about 1%.

**RESULTS** Figure 4 contains a plot of the conservative vector field with the period-one fixed point highlighted and enhanced. By narrowing our vision around the approximate fixed point and remembering that the vectors in the plots are scaled by a factor of  $1/20$ , we begin to see behavior that suggests an unstable equilibrium. Namely, the fixed point rests on a very narrow well, and any deviation from the point is amplified by successive strides.

Even if it turns out that the fixed point is *locally* stable with a small basin of attraction, we find this physically irrelevant; EduBot clearly exhibits a high degree of stability, so we have no interest in discussing infinitesimally-attracting equilibrium gaits. For this reason, we turn toward the investigation of a damped version of the model discussed here to describe EduBot's rolling motion.

---

<sup>3</sup>We actually plot the final condition *after applying the map*  $(\theta, \dot{\theta}) \mapsto -(\theta, \dot{\theta})$ ; this takes into account the fact that the leg spring constants are exchanged during flight, and means shorter vectors correspond to more symmetric gaits.

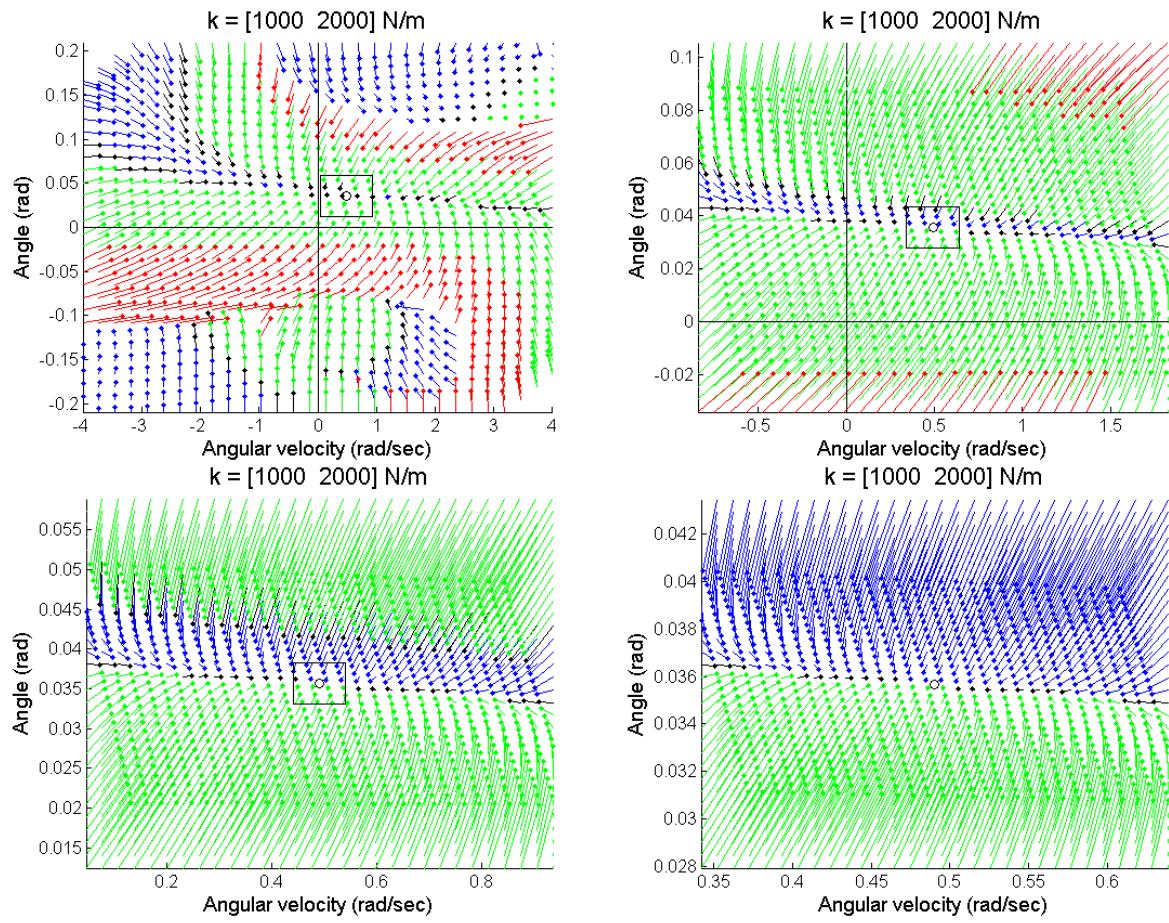


Figure 4: Conservative vector field; unstable period-one fixed point highlighted. Vectors scaled by  $1/20^{\text{th}}$  for clarity.

## 4 DAMPED MODEL FOR ROLL

Given substantial numerical evidence that the period-one equilibrium gaits in the conservative model for EduBot’s roll are (at the very least) not significant attractors, we proceed to add damping to the model with the hope that the attractive basins surrounding fixed points will enlarge considerably, making stable equilibrium gaits.

Note that if we were to solely add viscous damping to the model in Section 3, our simulations would become degenerate very quickly. Without a corresponding method to add energy to the system, there would be precisely one (very stable!) globally attracting equilibrium gait with infinite (or zero) stride period: the static gait described by

$$(h, \dot{h}, \theta, \dot{\theta}) \equiv \left( l_0 - \frac{gM}{4} \left( \frac{1}{k_1} + \frac{1}{k_2} \right), 0, \sin^{-1} \left( \frac{gM}{4w} \left( \frac{1}{k_1} - \frac{1}{k_2} \right) \right), 0 \right). \quad (27)$$

Thus, we simultaneously add an energy-addition scheme to the conservative model.

**ENERGY ADDITION** We add energy to the system in a similar manner to Koditschek and Buehler [16]: when a spring reaches maximum compression in the model, we temporarily increase its spring constant, thus increasing the energy stored in the spring. When a spring leaves the ground, its spring constant is reset. Maximum compression is detected when a leg’s velocity goes to zero,

$$-\dot{h} \pm w\dot{\theta} = 0, \quad (28)$$

where the  $\pm$  is determined by which leg is being considered. At that instant, the spring constant  $k_i$  is multiplied by an increase factor  $\gamma > 1$ . When the leg leaves the ground, the spring constant is reset to its initial value.

This energy addition scheme was chosen because it is easy to implement and can facilitate analysis better than many other methods. However, it is a poor model for the physical system and could be a source of qualitative disagreement between experimental and simulation results; we do not investigate the affect of changing the energy addition scheme here.

Figure 5 contains a plot of the damped vector field with one of the period-one fixed points highlighted and enhanced. Note that vector scaling in this figure is half what it was for the conservative system plots. There is actually a second fixed point, which can be seen in the lower half-plane of the upper-left plot; it corresponds to a gait which lands first on the stiff leg. The period-one equilibrium gait we found for the conservative system lands first on the soft leg, so we consider this the *dominant* equilibrium gait, and choose to highlight it in Figure 5. The second fixed point will be referred to as the *suppressed* equilibrium gait.

### 4.1 ELIMINATING ROLL

**SPRING CONSTANTS** When we constructed the current model, we assumed that the difference in spring constants drives the rolling motion observed in EduBot, and that letting

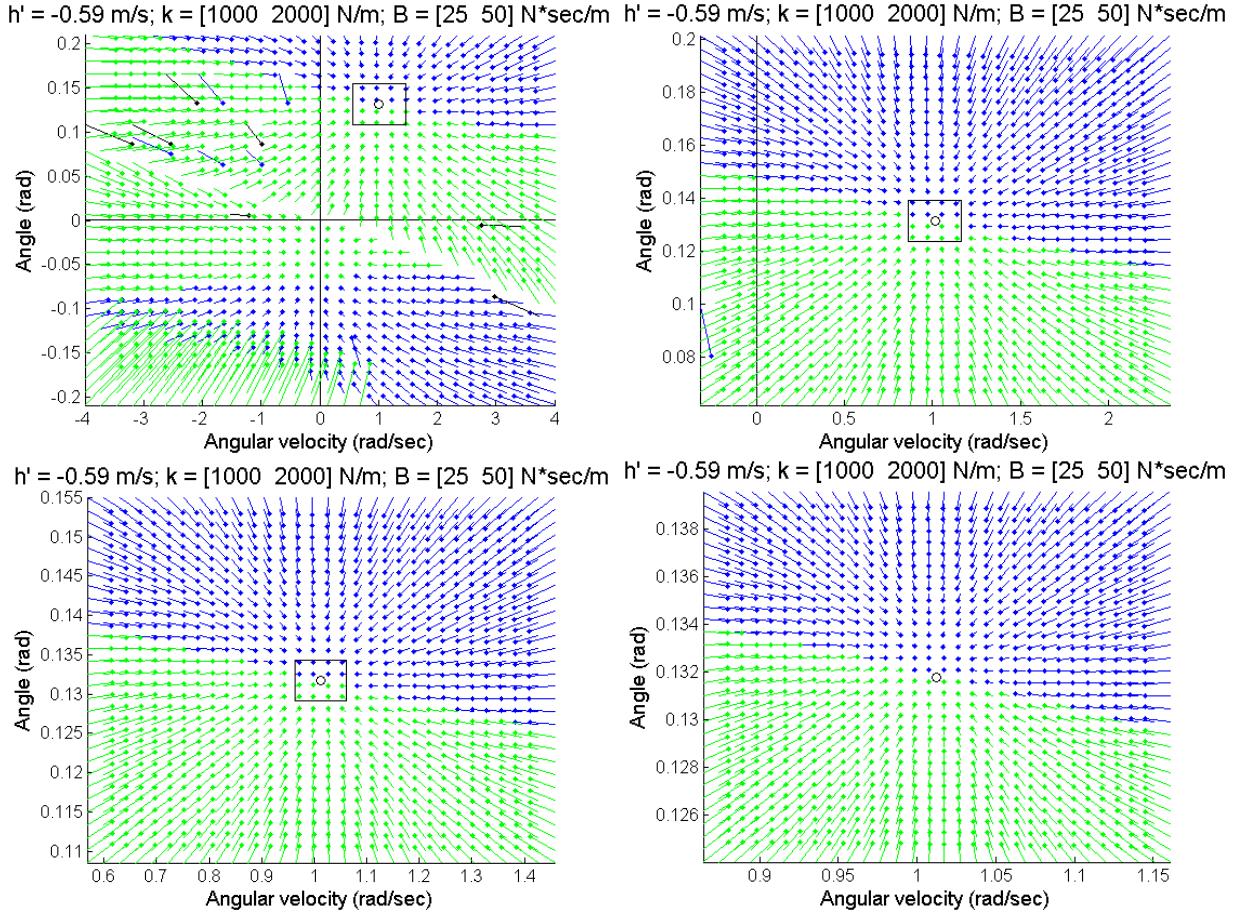


Figure 5: Damped vector field; dominant stable period-one fixed point highlighted. Vectors scaled by  $1/10^{th}$  for clarity. Though  $\theta$  determines  $h$  at touchdown as in the conservative model, energy dissipation implies  $\dot{h}$  isn't determined by  $\dot{\theta}$ . Thus, these plots show a particular  $\dot{h}$ -slice;  $\dot{h} = -0.59$  m/sec.

$k_1 \rightarrow k_2$  would likely eliminate the roll<sup>4</sup>. However, we were not sure how the equilibrium gaits would move toward the origin and what would happen to the gaits' stability as  $k_1 \rightarrow k_2$ . To study the progression of the equilibrium gaits toward the origin, we first compute fixed points of the Poincaré map  $\mathcal{P}$  associated with  $\Sigma$  to desired error tolerances, then numerically compute the Jacobian of  $\mathcal{P}$ ,  $\frac{\partial \mathcal{P}}{\partial(h, \dot{h}, \theta, \dot{\theta})}$ , to determine how the stability of the gaits changes.

**EQUILIBRIUM GAITS** Based on physical intuition and qualitative numerical evidence, we proceed by assuming that equilibrium gaits are stable in the damped model. This means if the system is allowed to flow from initial conditions close to a fixed point, the Poincaré map will tend asymptotically to the point. Thus, to find a fixed point to a given tolerance, we run a forward simulation and average the values of the Poincaré map; the standard deviation in these values provides our estimate of the error.

**NUMERICAL STABILITY** Since we have yet to provide an analytical result regarding the stability of fixed points of the damped model, we turn to numerical methods to estimate stability. Specifically, we compute the Jacobian of the  $4 \times 4$  Poincaré Map  $\mathcal{P}$ ,

$$\frac{\partial \mathcal{P}}{\partial(h, \dot{h}, \theta, \dot{\theta})} = \begin{bmatrix} \partial_h \mathcal{P}_h & \partial_{\dot{h}} \mathcal{P}_h & \partial_\theta \mathcal{P}_h & \partial_{\dot{\theta}} \mathcal{P}_h \\ \partial_{\dot{h}} \mathcal{P}_h & \partial_{\dot{h}} \mathcal{P}_{\dot{h}} & \partial_\theta \mathcal{P}_{\dot{h}} & \partial_{\dot{\theta}} \mathcal{P}_{\dot{h}} \\ \partial_h \mathcal{P}_\theta & \partial_{\dot{h}} \mathcal{P}_\theta & \partial_\theta \mathcal{P}_\theta & \partial_{\dot{\theta}} \mathcal{P}_\theta \\ \partial_{\dot{h}} \mathcal{P}_\theta & \partial_{\dot{h}} \mathcal{P}_{\dot{\theta}} & \partial_\theta \mathcal{P}_{\dot{\theta}} & \partial_{\dot{\theta}} \mathcal{P}_{\dot{\theta}} \end{bmatrix} \quad (29)$$

**PERTURBED TRAJECTORIES** As  $k_1 \rightarrow k_2$ , we observe that the system becomes more susceptible to perturbations in that it may begin to switch between equilibrium gaits. For instance, Figure 6 shows two system trajectories, one with  $k_1 = 1000$  N/m,  $k_2 = 2000$  N/m and the other with  $k_1$  set to 1500 N/m. At the times indicated by the vertical bars, the system was perturbed. Perturbations take the form of instantaneous forces: the first perturbation set  $\dot{h}^+ = \dot{h}^- + 0.2$  m/sec,  $\dot{\theta}^+ = \dot{\theta}^- - 2$  rad/sec; the second set  $\dot{h}^+ = \dot{h}^- - 0.4$  m/sec,  $\dot{\theta}^+ = \dot{\theta}^- - 4$  rad/sec. In the first case, the perturbations are rejected, and the system returns to its equilibrium gait. As  $k_1 \rightarrow k_2$ , however, the perturbations switch the system back and forth between the dominant equilibrium gait and the suppressed.

## 5 RESULTS

Figure 7 shows the model's prediction for how the state of the body at touchdown changes as  $k_1 \rightarrow k_2$ : the angle and angular velocity approach the origin. Ideally, we would generate an identical plot experimentally, however, the state at touchdown is difficult to measure directly with EduBot. Thus we present additional results from the model which are more amenable to experimentation, and compare this data with results from the physical system.

---

<sup>4</sup>This is intuitively clear since the body in the model will almost certainly roll when, for instance,  $k_1 = k_2/2$ , but will not roll at all when  $k_1 = k_2$ , so long as  $\theta_0 = \dot{\theta}_0 = 0$  (since in this case the model is equivalent to one with a single spring with spring constant  $k = 2k_1 = 2k_2$ ).

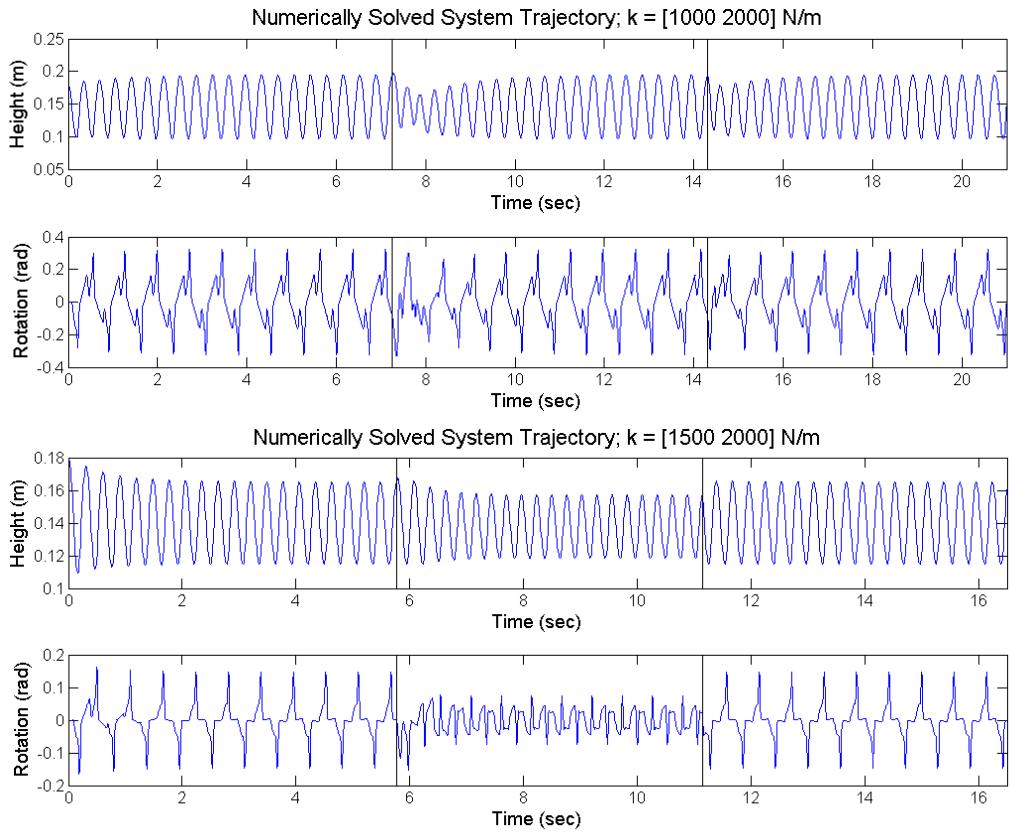


Figure 6: Numerically computed trajectories perturbed at the 20<sup>th</sup> and 40<sup>th</sup> touchdown. When  $k_1 = 1000$  N/m, the system asymptotically approaches the same equilibrium gait after each perturbation. When  $k_1 = 1500$  N/m, the system switches between equilibrium gaits when perturbed.

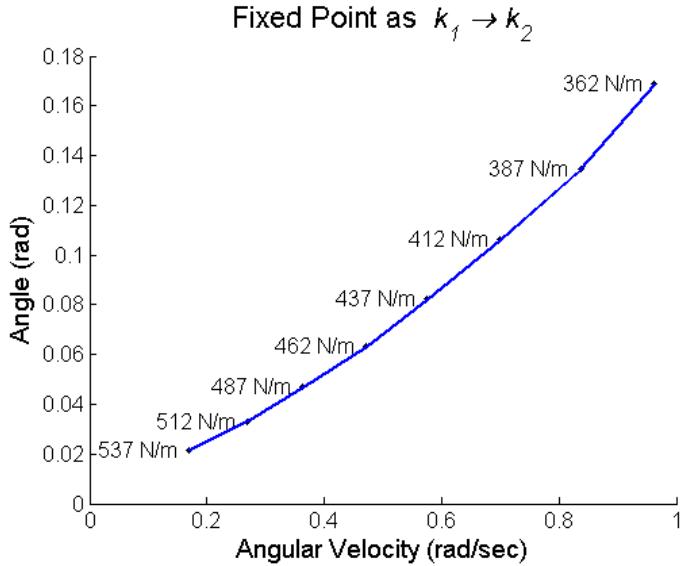


Figure 7: Body state at touchdown as  $k_1 \rightarrow k_2$ .

$k_1$ (N/m)	Average maximum roll (degrees)	Number of samples
300	$4.1 \pm 0.4$	42
450	$2.5 \pm 0.6$	42
600	$1.9 \pm 0.5$	68

Table 1: As  $k_1 \rightarrow k_2$ , EduBot's roll decreases. Though the decrease is not enormous, it is statistically significant.

## 5.1 SIMULATION FORCES AND TRAJECTORIES

Figure 8 clearly demonstrates that the average maximum roll at equilibrium decreases as  $k_1 \rightarrow k_2$ .

Figure 9 demonstrates that individual leg ground reaction forces equalize as  $k_1 \rightarrow k_2$ . In the model, if  $k_1 = k_2$ , then the force patterns would be precisely equal between the legs.

## 5.2 ROBOT FORCES AND TRAJECTORIES

Table 1 contains the result of averaging the single stride maximum roll EduBot exhibits over a series of experiments. Though the trend is not overwhelming, it does appear that roll decreases as  $k_1 \rightarrow k_2$ .

Figure 10 demonstrates that the forces exerted on the ground by each leg appear to equalize as  $k_1 \rightarrow k_2$ , however, this data is preliminary.

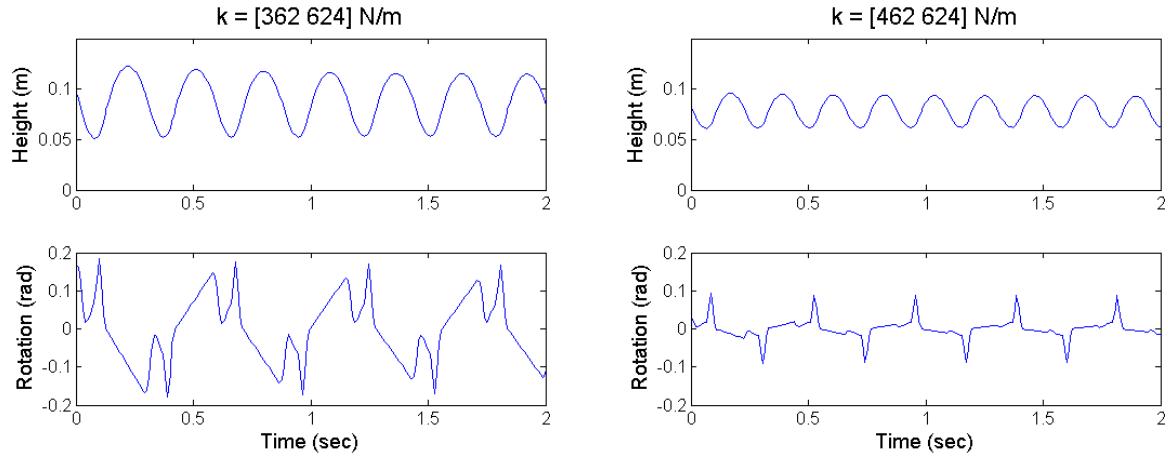


Figure 8: Simulated roll trajectory as  $k_1 \rightarrow k_2$ .

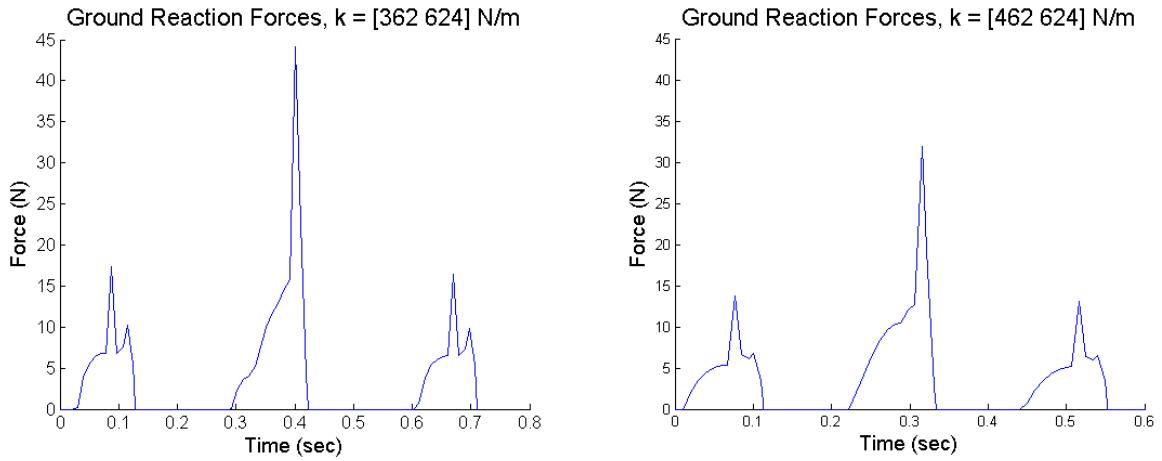


Figure 9: Simulated individual leg forces. We present the data as we would record it experimentally (see Figure 10) for ease of comparison; the first and third ground strikes are identical here.

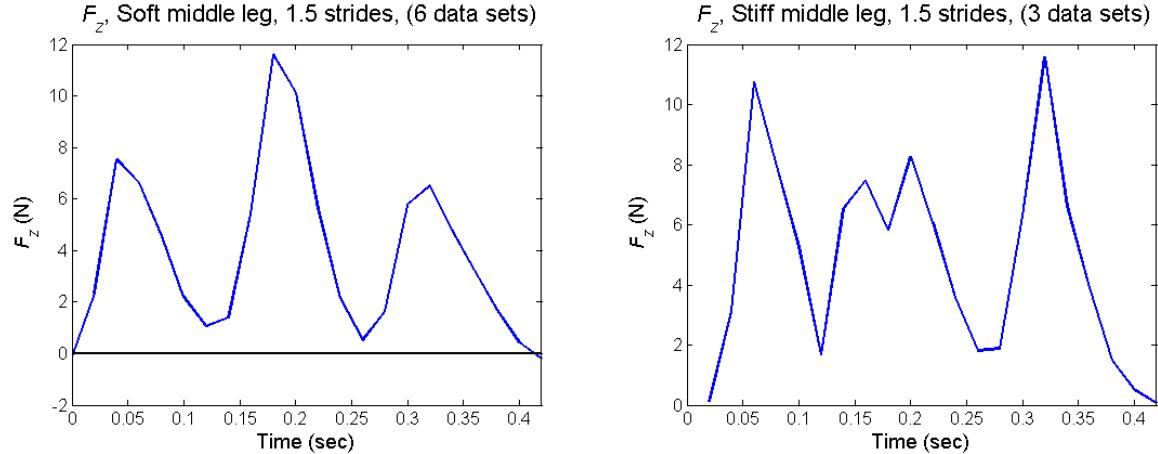


Figure 10: Experimental individual leg forces. From left to right, the three main peaks correspond to the front, middle, and back legs striking the force plate.

## 6 DISCUSSION

The major contributions of this work are the new analytical model for EduBot’s roll and the correspondence between numerical and experimental results that suggest one way to eliminate the robot’s roll. This is significant for several reasons. First, despite the success of simple models in describing the sagittal-plane and lateral-plane dynamics of legged locomotion, little work has been put into the analysis of rolling dynamics. Second, since we were able to use this model to predict the effect of a parameter variation in the physical system, we have provided the initial evidence that this model captures something fundamental about the rolling dynamics exhibited by EduBot.

The significance of producing an experimentally valid simple mathematical description of a certain aspect of legged locomotion is subtle. The immediate utility of such a model is that it provides an intuitive framework to think about legged creature’s rolling; we are justified in thinking of rolling motion independent from other aspects of a robot’s gait, and we understand, qualitatively, the effect of physical parameter variations on roll. In farther-reaching terms, this work could motivate the design of controllers to manipulate rolling in legged robots by providing a plant model, or could inform the design of legged robots themselves by giving insight into the predominant physical processes that lead to rolling.

## 7 RECOMMENDATIONS

We succeeded in posing the analytical problem of identifying equilibrium gaits in the conservative model using a formal mathematical framework, but we did not make any progress toward actually identifying such fixed points. Future work should therefore be concentrated on producing analytic results in the conservative model. A similar analytical statement can

be constructed for the damped model, and it would be an even more laudable goal to produce analytical results in that case, though we consider this to be a more difficult problem.

We elected to investigate the effect of one parameter variation on roll; this restriction was motivated primarily by the desire to produce an experimentally-verifiable experiment from the model in the time available. There are many more avenues of inquiry that could be taken up in the model, however, like studying the effect of using different energy-addition schemes in the damped model and designing controllers to stabilize the equilibrium gaits of the conservative model. These investigations would likely yield useful insight into how to design and control legged robots.

## 8 ACKNOWLEDGMENTS

Without the profound support I've received from Dr. Klavins over the past two years, I would not have been able to attend this summer program. Without the significant time and thought invested by Dr. Koditschek, Dr. Clark, and Joel Weingarten, I would not have been nearly so productive or successful this summer as I have been. And without the assistance of Goran Lynch, Dr. Komsuoglu, and the other members of Dr. Koditschek's lab, I would have often been stymied and directionless, not to mention robotless. I would also like to thank the GRASP lab for the use of their motion capture system. Finally, my sincere gratitude goes to Dr. Van der Speigel, the SUNFEST program, and the NSF REU system for organizing and sponsoring this project—I have learned a ridiculous amount this summer.

## References

- [1] U. Saranli, M. Buehler, and D. Koditschek, Rhex: A simple and highly mobile hexapod robot, *The International Journal of Robotics Research*, 20(7) (2001), 616–631.
- [2] S. Bailey, J. Cham, M. Cutkosky, and R. Full, Comparing the locomotion dynamics of a cockroach and a shape deposition manufactured biomimetic hexapod, International Symposium on Experimental Robotics, December 2000.
- [3] M. Fend, Whisker-based texture discrimination on a mobile robot, European conference on Advances in Artificial Life, September 2005.
- [4] M. Srinivasen, S. Zhang, J. Chahl, G. Stange, and M. Garratt, An overview of insect-inspired guidance for application in ground and airborne platforms, *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.*, (2004).
- [5] D. Koditschek, R. Altendorfer, and P. Holmes, Stability analysis of legged locomotion by symmetry-factored return maps, *International Journal of Robotics Research*, 23(10-11) (2004), 979–999.

- [6] H. Komsuoglu, R. Altendorfer, U. Saranli, D. Koditschek, H. Brown, M. Buehler, M. Moore, D. McMordie, and R. Full, Evidence for spring loaded inverted pendulum running in a hexapod robot, International Symposium on Experimental Robotics, December 2000.
- [7] J. Weingarten, G. Lopes, M. Buehler, R. Groff, and D. Koditschek, Automated gait adaptation for legged robots, vol. 3, International Conference on Robotics and Automation, 2004, pp. 2153–2158.
- [8] R. Blickhan, The spring-mass model for running and hopping, *Journal of Biomechanics*, 22(11-12) (1989), 1217–1227.
- [9] R. Blickhan and R. Full, Similarity in multilegged locomotion: Bouncing like a monopode, *Journal of Computational Physiology*, A(173) (1993), 509–517.
- [10] D. Koditschek, R. Altendorfer, and P. Holmes, Stability analysis of a clock-driven rigid-body SLIP model for RHex, *International Journal of Robotics Research*, 23(10-11) (2004), 1001–1012.
- [11] T. McGeer, Passive dynamic walking, *International Journal of Robotics Research*, 9(2) (1990), 62–82.
- [12] M. Garcia, A. Chatterjee, A. Ruina, and M. Coleman, The simplest walking model: Stability, complexity, and scaling, *Journal of Biomechanical Engineering*, 120(2) (1998), 281–288.
- [13] J. Guckenheimer and S. Johnson, Planar hybrid systems, *Hybrid Systems II*, Springer-Verlag, 1995, pp. 202–225.
- [14] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, American Mathematical Sciences, no. 42, Springer-Verlag, New York, 2002.
- [15] J. Schmitt and P. Holmes, Mechanical models for insect locomotion i. theory, *Biological Cybernetics*, 83 (2000), 501–515.
- [16] M. Buehler and D. Koditschek, Analysis of a simplified hopping robot, *International Journal of Robotics Research*, 10(6) (1991), 587–605.
- [17] Advanced Mechanical Technology Inc., He6x6 force plate, (2006), <http://www.amtiweb.com/index.htm>.
- [18] Y. Kwon, Force-plate issues, (2006), <http://kwon3d.com/theory/grf.html>.
- [19] The MathWorks Inc., Matlab, (©1994-2006), <http://www.mathworks.com/>.

## A SIMPLE POINCARÉ MAP PROBLEM

When confronted with the task of discovering the long-term behavior of a dynamical system, one approach is to look at the sequence of points where the system intersects a *Poincaré section* [14]. Essentially, a Poincaré section  $\Sigma$  is a surface of codimension 1 embedded in state space. By studying the action of the system on  $\Sigma$ , we can glean information about the system's long-term behavior.

We present here a solution to a simple linear dynamical system problem which utilizes a Poincaré section. This exposition is intended to motivate the analysis we later set up to apply to the conservative model for EduBot's rolling.

### DEFINE

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \dot{x} = (\sigma I + \omega J)x, \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

$$q = \begin{bmatrix} \rho \\ \theta \end{bmatrix} = \begin{bmatrix} \sqrt{x_1^2 + x_2^2} \\ \tan^{-1}(x_1, x_2) \end{bmatrix}, \quad \Theta = \{(\rho, \theta) \in \mathbb{R} \times \mathbb{S}^1 \mid \theta = 0\}.$$

(We adopt the convention that  $\theta = \tan^{-1}(y, x)$  implies  $x = \cos \theta$ ,  $y = \sin \theta$ .)

Define a *Poincaré map*  $\mathcal{P} : \Theta \rightarrow \Theta$  in the following way: if  $x(t, x_0)$  solves the differential equation<sup>5</sup> and  $\Theta(x(\tau_k)) = 0$  with  $\rho(x(\tau_k)) = \rho_{\tau_k}$ , then  $\mathcal{P}(\rho_{\tau_k}) = \rho_{\tau_{k+1}}$ , where  $\tau_{k+1} = \min_{t > \tau_k} \{\theta(x(t)) = 0\}$ .

So, intuitively, if  $x_1$  lies along the  $y$ -axis and  $x_2$  along the  $x$ -axis,  $\Theta$  is the set of points where a solution to the differential equation crosses the positive  $x$ -axis, and  $\mathcal{P}$  maps such points onto the next such point in (temporal) order of occurrence.

**FIND** An explicit formula for  $\rho_{\tau_{k+1}} = \mathcal{P}(\rho_{\tau_k})$ .

**GENERAL SOLUTION** The general solution to the differential equation is given by

$$x(t, x_0) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} e^{\sigma t}(\bar{x}_1 \cos(\omega t) + \bar{x}_2 \sin(\omega t)) \\ e^{\sigma t}(-\bar{x}_1 \sin(\omega t) + \bar{x}_2 \cos(\omega t)) \end{bmatrix}, \quad x_0 = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}. \quad (30)$$

**STRUCTURE OF  $\Theta$**  Given  $\theta = \tan^{-1}(x_1, x_2)$ , it is clear that  $\theta = 0 \iff x_1 = 0, x_2 > 0$ . But

$$x_1(t) = 0 \iff \bar{x}_2 \sin(\omega t) = -\bar{x}_1 \cos(\omega t) \iff \tan(\omega t) = -\frac{\bar{x}_1}{\bar{x}_2}, \quad (31)$$

$$x_2(t) > 0 \iff -\bar{x}_1 \sin(\omega t) + \bar{x}_2 \cos(\omega t) > 0. \quad (32)$$

---

<sup>5</sup>In this case, we find it convenient to drop the  $x_0$  argument and simply call the solution  $x(t)$ .

The condition in eq. 31 gives two possible choices for  $\omega t$  which lie in opposing quadrants (i.e. I and III or II and IV); the condition in eq. 32, when taken with the result from eq. 31, selects one of these two choices (the choice where  $\sin(\omega t) < 0$ ).

Note that since  $\sin(\omega t)$  and  $\cos(\omega t)$  are  $\frac{2\pi}{\omega}$ -periodic, a solution  $\tau_0$  determined from eqs. 31 & 32 gives rise to an entire sequence of solutions of the form  $\{\tau_k = \tau_0 + \frac{2\pi}{\omega}k \mid k \in \mathbb{N}\}$ . Note also that consecutive terms in this sequence provide the proper temporal ordering on events in  $\Theta$ , by which we mean  $\tau_{k+1} = \min_{t > \tau_k} \{\theta(x(t)) = 0\}$ .

**FINDING  $\mathcal{P}(\rho_{\tau_k})$**  We are nearly finished. We have the sequence of times corresponding to points in  $\Theta$ , so all that remains is to evaluate  $\rho$  at the  $(x_1, x_2)$  coordinates corresponding to those times.

First, note that an event in  $\Theta$  requires  $x_1 = 0$ , so  $\rho_{\tau_k} = x_2(\tau_k)$ . Next, note that by construction  $\tan(\omega\tau_k) = -\frac{\bar{x}_1}{\bar{x}_2}$ ,  $\sin(\omega\tau_k) < 0$ , so

$$\sin(\omega\tau_k) = \frac{-\bar{x}_1}{\sqrt{\bar{x}_1^2 + \bar{x}_2^2}}, \quad \cos(\omega\tau_k) = \frac{\bar{x}_2}{\sqrt{\bar{x}_1^2 + \bar{x}_2^2}}. \quad (33)$$

So by combining eq. 33 with eq. 30, we have

$$\rho_{\tau_k} = x_2(\tau_k) = e^{\sigma\tau_k} \left( \frac{\bar{x}_1^2 + \bar{x}_2^2}{\sqrt{\bar{x}_1^2 + \bar{x}_2^2}} \right). \quad (34)$$

Now, recognizing that the next event occurs at time  $\tau_{k+1} = \tau_k + \frac{2\pi}{\omega}$ ,

$$\rho_{\tau_{k+1}} = \mathcal{P}(\rho_{\tau_k}) = e^{\sigma(\tau_k + \frac{2\pi}{\omega})} \left( \frac{\bar{x}_1^2 + \bar{x}_2^2}{\sqrt{\bar{x}_1^2 + \bar{x}_2^2}} \right) = e^{\frac{2\pi\sigma}{\omega}} \rho_{\tau_k}, \quad (35)$$

and, more generally,

$$\rho_{\tau_{k+n}} = \mathcal{P} \circ \cdots \circ \mathcal{P}(\rho_{\tau_k}) = \mathcal{P}^n(\rho_{\tau_k}) = e^{\frac{2\pi\sigma}{\omega}n} \rho_{\tau_k}. \quad (36)$$

**SYSTEM BEHAVIOR** Assuming at least one of the initial conditions is nonzero, choices of  $\sigma$  and  $\omega$  determine the qualitative behavior of the system.

If  $\sigma = 0$ , the system moves in a circle. If  $\sigma < 0$ , the system spirals toward the origin. If  $\sigma > 0$ , the system spirals toward infinity. Note that these qualitative statements are reflected in eq. 36, and given precise analytical meaning—*this is the primary utility of such a solution*.

If  $\omega = 0$ , the system moves along a ray: toward infinity, toward the origin, or remaining stationary, depending on the choice of  $\sigma$ . In this case, the conclusion that the system's dynamics are  $\frac{2\pi}{\omega}$ -periodic is invalid, and either  $\Theta \equiv \emptyset$  or  $\Theta \subset \mathbb{R}$ , depending on the choice of initial conditions.

## B FORCE PLATE

We measure the *ground reaction forces* for EduBot using a commercially-available force plate [17]. The force plate uses Hall Effect sensors to measure the net force  $\mathbf{F} = (F_x, F_y, F_z)^T$  and moment  $\mathbf{M} = (M_x, M_y, M_z)^T$  applied to the *center of pressure* of the surface.

The plate either outputs analog signals which can be combined to compute  $\mathbf{F}$  and  $\mathbf{M}$  or a digital signal which can be used with proprietary software to save the net force and moment to an ASCII text file at 12 bit resolution and up to 200 Hz.

Given  $\mathbf{F}$  and  $\mathbf{M}$ , we wish to find the net  $z$ -axis torque  $T_z$  applied to the plate. From [18], we know that

$$\mathbf{M} = (x - a, y - b, \tau - c) \times \mathbf{F} + (0, 0, T_z), \quad (37)$$

where  $(a, b, c)$  is the actual origin for the force plate (given by the manufacturer's *calibration matrix*),  $\tau$  is the thickness of whatever padding is added to the surface of the force plate, and  $T_z$  is the net torque applied to the plate about the  $z$ -axis.

Writing this equation in matrix form, we have

$$\mathbf{M} = \begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \begin{bmatrix} 0 & -(\tau - c) & y - b \\ \tau - c & 0 & -(x - a) \\ -(y - b) & (x - a) & 0 \end{bmatrix} \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ T_z \end{bmatrix}, \quad (38)$$

so we arrive at

$$x = -\frac{M_y + (c - \tau)F_x}{F_z} + a, \quad (39)$$

$$y = \frac{M_x - (c - \tau)F_y}{F_z} + b, \quad (40)$$

$$z = \tau, \quad (41)$$

$$T_z = M_z - (x - a)F_y + (y - b)F_x. \quad (42)$$

So now given  $\mathbf{F}$  and  $\mathbf{M}$  from the force plate, we can extract the *center of pressure* and the net force and torque exerted by the robot.

## C NUMERICAL METHODS

We performed numerical simulations in MATLAB [19] using the `ode45` numerical solver with `RelTol` =  $1 \times 10^{-8}$ , `AbsTol` =  $1 \times 10^{-6}$ . We computed equilibria in the conservative model to within 1% in all coordinates unless otherwise specified. Vector field plots contain 30 subdivisions in each coordinate; trajectories that fail to produce a touchdown event are omitted. We used `ode45`'s built-in *event function* feature to implement the hybrid dynamics of the model.

## D TRAJECTORY MEASUREMENTS

Robot trajectories were recorded using a motion-capture system. The system treats EduBot like a rigid body in  $\mathbb{R}^3$ , recording the robot’s position and three-axis rotation at roughly 120 Hz with sub-millimeter accuracy. Occasionally the motion capture system loses track of the robot for one to five samples; in this case, we interpolate linearly across the lost samples. We then apply a 5-sample median filter to smooth jump discontinuities in the data that appear to be introduced by noise.

University of Pennsylvania  
SUNFEST  
REU Program  
Summer 2006

**Implementation of sintered LTCC for the fabrication  
Of a 3D cylindrical  
Micro combustor**

Jose M Castillo Colon (Physics and Electronics)-University of Puerto Rico at Humacao  
Advisor: Jorge Santiago Aviles

**ABSTRACT**

The micro-combustor is a compact, sub-millimeter device that burns hydrocarbon fuels homogeneously as a source of power. It efficiently converts heat generated by combustion into electric power. We want to design a cylindrical structure, using the FemLab simulation program it demonstrates that doing this geometry can reduce impedance fluid, and several gas inputs can be placed under and several outputs can be placed around the mixing chamber so it can be more efficient.

The materials to be used for the construction of this device are fundamentally Low Temperature Co-Fired Ceramic (LTCC) and Graphite. The fabrication of this device will rely essentially on a thermal process (sintering of the LTCC tapes). The instruments that will be used for the fabrication include: a furnace for sintering the ceramics, a heated press for the ceramics lamination, a thermal laser and a numerically controlled milling machine for the patterning and machining of the tape. It is hoped that the combustor fabrication will be completed as designed. The parameters that characterize its combustion and power are expected to be consistent with its application as an electrical generator by means of the thermoelectric effect.

The main objective of this project is to complete the fabrication of a small combustor that contains fundamentally four inputs, four output, mixer and burning chamber. In two of its inputs a combustible gas (Propane) is injected, and oxygen from the air as an oxidizer flows through the other two inputs. The gases are mixed in the mixer area then goes to the burning chamber, a flame is initiated in the burning area to burn the fuel / oxidizer mixture.

## **CONTENTS**

<b>Abstract.....</b>	<b>1</b>
<b>Introduction .....</b>	<b>3-5</b>
<b>Background.....</b>	<b>5-6</b>
<b>Theory.....</b>	<b>6-7</b>
<b>AutoCad architecture Diagrams.....</b>	<b>8-9</b>
<b>Laser process detail.....</b>	<b>10-11</b>
<b>Lamination Process.....</b>	<b>11-13</b>
<b>Furnace process.....</b>	<b>13-14</b>
<b>Conclusion.....</b>	<b>15</b>
<b>Recommendation and acknowledgments.....</b>	<b>16</b>
<b>References.....</b>	<b>17</b>

## 1. INTRODUCTION

When the air is mixed with a combustible and ignited to form a flame producing high temperatures, the process is known as combustion. The combustion is a chemical reaction in which a fuel (element or component) is combined with an oxidizer (generally oxygen in form of gaseous O<sub>2</sub>), giving off heat and producing an oxide. Frequently used types of element for the combustion are the carbon and hydrogen. The combustion process happens as often in living beings as in devices used as sources of energy.

When this process happens inside a device, is known as a combustor. Combustors are commonly seen in mechanical motors such as in cars, airplanes, boats, etc. As one knows these are made to move and climb, as they are designed for the displacement of great weights that require enormous amounts of energy, which implies the consumption of great amounts of fuel. Nevertheless components exist that do not require large amounts of energy. These in their majority are electronics systems, which are designed to consume energy supplied by means of batteries and electricity.

To make a combustor at a small scale, one that will work for devices requiring lower energy levels, it has been proposed that one must construct a combustor of proportionally smaller dimension. This is known as a micro-combustor. The micro-combustor is a compact, millimeter length device that burns hydrocarbon fuels homogeneously as a source of power.

The main objective of this project is to complete the fabrication of a cylindrical micro-combustor out of LTCC tapes, which contains fundamentally four inputs, four outputs a combustion and a burning area. In the two inputs a combustible gas (probably propane) is injected, and air flows through the other two inputs. The gases are mixed in the mixer area. A flame is initiated in the burning area to burn the fuel / oxidizer mixture by means of a capacitor discharge or a piezoelectric element.

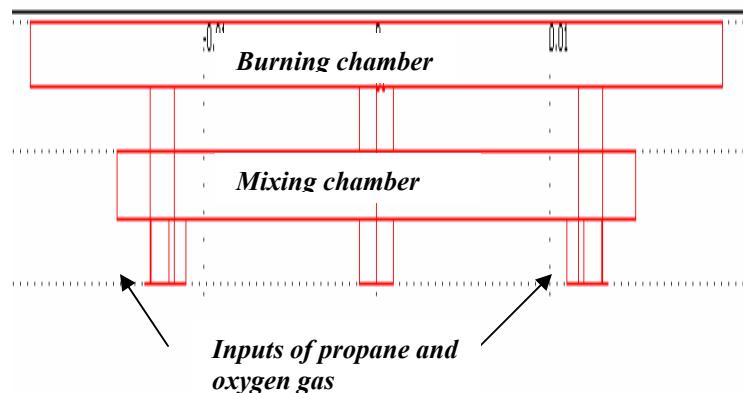


Fig1. Three-dimensional micro-combustor (on one side)

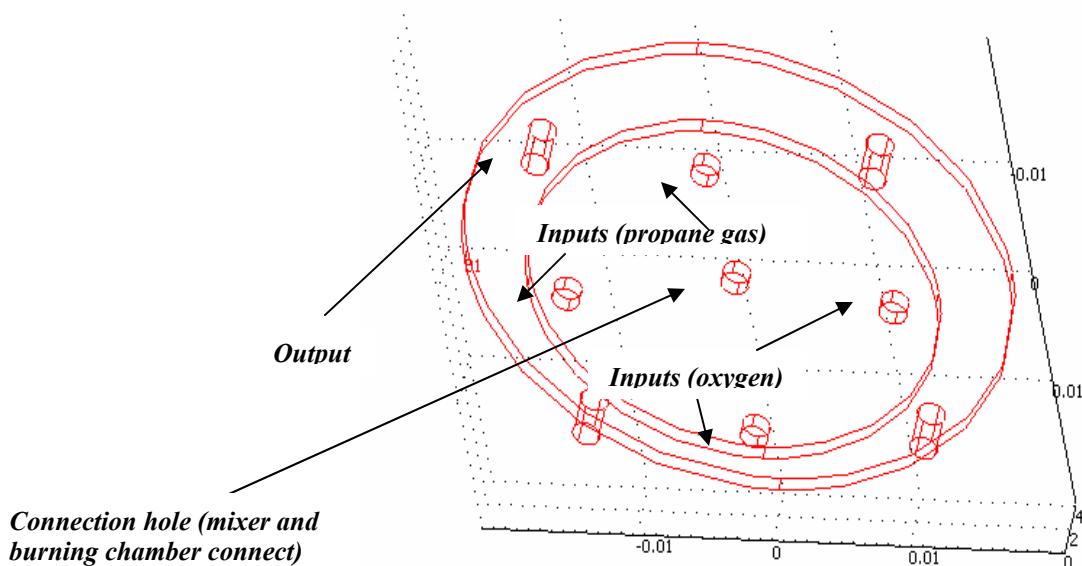


Fig2. Three-dimensional micro-combustor (on the bottom)

The materials to be used for the construction of this device are Low Temperature Co-Fired Ceramic (LTCC) tape. The LTCC represent an important alternative to be used as substrates for machining in the meso and micro scale. They provide several advantages including: electronic circuits can be integrated because of their hybrid nature, tapes of different compositions can be formulated to obtain desired layer properties (e.g. magnetic permeability), possibility of fabrication of hybrid structures consisting of ceramics, silicon, metals and/or some other suitable materials, layer count can be high, possibility of self-packaging, fabrication techniques are relatively simple, inexpensive and environmentally benign.

The fabrication of this device will rely essentially on a thermal process (sintering of the LTCC tapes). The instruments that will be used for the fabrication / characterization include: a furnace for sintering the ceramics, a heated press for lamination of the ceramics, and a thermal laser.



Fig3. Example of a LTCC Tape

## 2. BACKGROUND

The typical commercial battery has become the staple solution for portable power in today's society of mobile electronic devices. Unfortunately, current commercial batteries possess low energy density, short life spans, and are harmful to the environment upon disposal. The goal of the micro-combustor project is to combat the aforementioned disadvantages of the commercial battery by providing a competitive, portable energy source.

A promising alternative to electrochemical batteries involves the combustion of liquid hydrocarbon based fuels. Since liquid hydrocarbon based fuels employ energy densities two orders of magnitude greater than commercial batteries, these fuels are able to provide an ideal source for mobile power generation. However, conventional methods of electrical generation require the combustion of hydrocarbons to drive a mechanical generator. The inclusion of the mechanical generator increases both the size and weight of such a system and reduces its efficiency due to intermediate energy conversions.

Unlike a conventional generator, the micro-combustor will harness the heat produced from combustion and directly convert it to electricity through a thermoelectric element. The absences of mechanical parts in a micro-combustor allows for a much smaller size and quiet operation while generating power. These features allow a micro-combustor to provide portable energy similar to a battery while allowing it to utilize greater energy density through hydrocarbon-based fuels.

Figure four shows potential energy of various sources in terms of energy density. Compared with liquid fuel such as gasoline, these alternative possibilities posses far smaller energy densities.

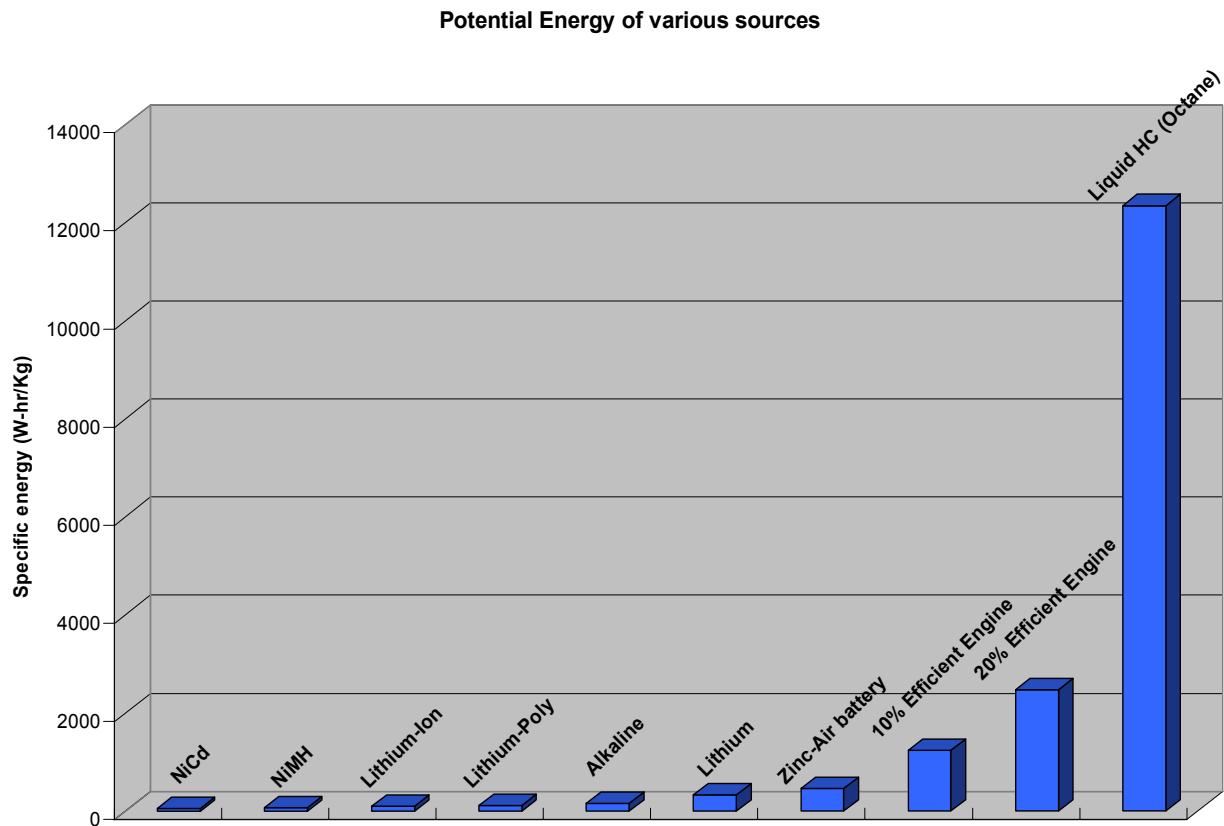


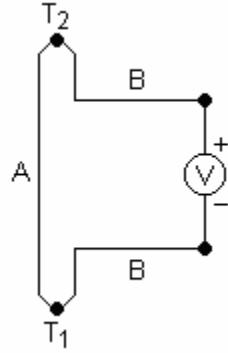
Fig4. Energy density of batteries in various forms

### 3. Theory

The theory behind the micro-combustor is largely summarized by the seebeck effect. The seebeck effect is a phenomenon, which results in a voltage difference being produced across the terminals of an open circuit with each junction held at a different temperature. The produced voltage difference is directly proportional to the difference of the hot and cold junction temperatures and is not dependent on the distribution of temperature along the metal between the junctions. As result of the seebeck effect, the temperature gradient across the hot plate and the cold plate produces electric power.

A voltage, the thermoelectric EMF, is created in the presence of a temperature difference between two different metals or semiconductors. This usually causes a continuous current to flow in the conductors. The voltage created is on the order of several micro volts per degree of difference.

In the circuit:



(Which can be in several different configurations and be governed by the same equations), the voltage developed can be derived from:

$$V = \int_{T_1}^{T_2} (S_B(T) - S_A(T)) dT$$

$S_A$  and  $S_B$  are the Seebeck coefficients (also called *thermoelectric power* or *thermopower*) of the metals A and B, and  $T_1$  and  $T_2$  are the temperatures of the two junctions. The Seebeck coefficients are non-linear, and depend on the conductors' absolute temperature, material, and molecular structure. If the Seebeck coefficients are effectively constant for the measured temperature range, the above formula can be approximated as:

$$V = (S_B - S_A) \cdot (T_2 - T_1)$$

Thus, a thermocouple works by measuring the difference in potential caused by the dissimilar wires. It can be used to measure a temperature difference directly, or to measure an absolute temperature, by setting one end to a known temperature. Several thermocouples in series are called a thermopile. This is also the principle at work behind thermal diodes and thermoelectric generators (such as radioisotope thermoelectric generators or RTGs) which are used for creating power from heat differentials. The Seebeck effect is due to two effects: *charge carrier diffusion* and *phonon drag*.

## Materials and tools

The tools and materials utilized in the fabrication of these devices were:

1. Green tape type 951AT from DuPont, (approximately 100  $\mu\text{m}$  thick).
2. Isotemp Programmable Forced-Draft Furnace (Fisher Scientific)
3. Heated press (Carver Model C)
4. X-660 Laser Platform (Universal Laser Systems, Scottsdale, AR, U.S.A., 60 W CO<sub>2</sub> laser, wave length of 10.6  $\mu\text{m}$ )

## COMBUSTOR ARCHITECTURE

Below is a detailed description of the combustor morphology and configuration.

### 3.1 AUTOCAD2000i BASED ARCHITECTURE DIAGRAMS.

Below is a detailed scheme of the combustor architecture. Each layer has the measure of every space in the LTCC layer.

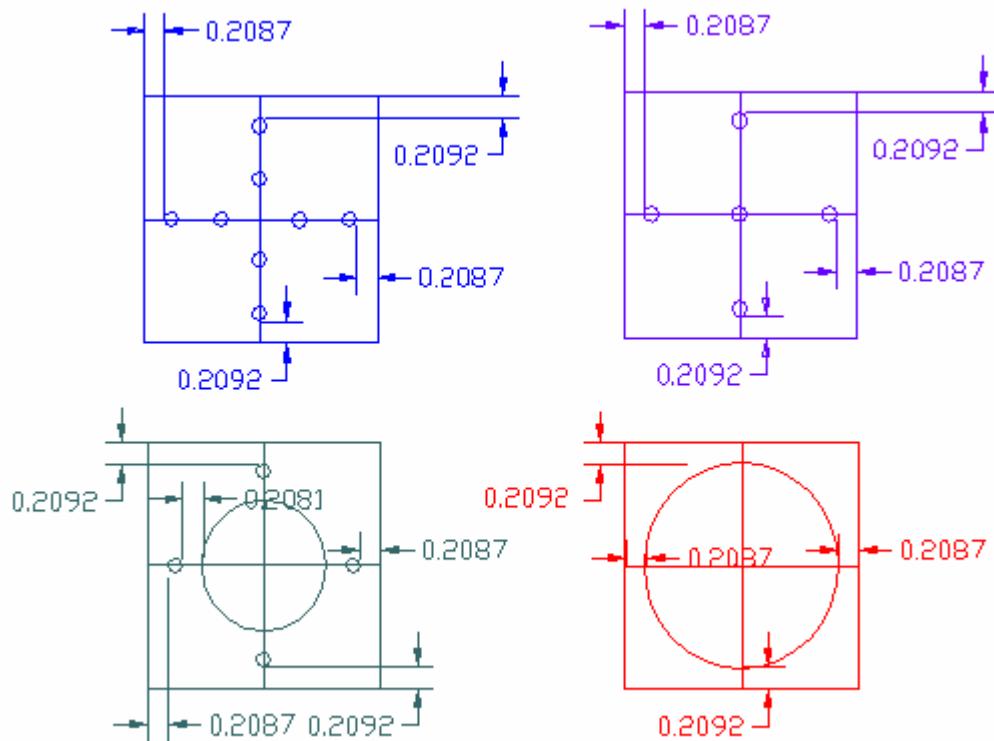


Fig5. Layers and measure of the layers in AutoCAD (Each layers is 2.4" x 2.4")

Layer Color	Circle diameter	Quantity of layer created
Blue layer	.0729" every circle	8
Violet layer	.0729" every circle	8
Green layer	.0729" the small circle and the bigger circle have .6391 "	23
Red layer	1.000"	23

Each layer has a different type of purpose and its build it in that way to have the cylindrical geometry we want to built. The idea of doing this geometry is because having several gas inputs and several outputs can reduce fluidic impedance so the efficiency of the micro combustor increases.

The blue layer it's called the bottom layer, in this layer it's were all the inputs of gas and oxygen goes. The violet layer it's called the inter layer because it goes between the blue layer and the green layer, so its connect the outputs to the green layer and the hole in the middle it's were all the mixture of the oxygen and the propane gas going to be passing trough the mixer chamber and finally goes to the burning chamber. The green layer it's called the mixer layer because here is where all the gases are going to be mixing thanks to the design of the chamber. And the red layer it's called the burning layer because here is where the device produces the flame thanks to the combustion process.

The following illustration show how the structure looks from the top, using AutoCAD layers.

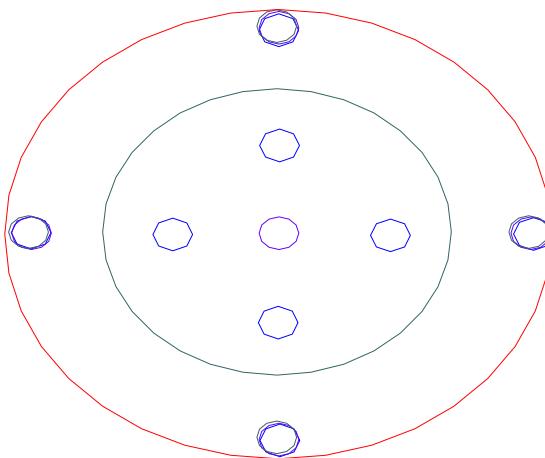


Fig6. Top part of the micro combustor using the auto CAD software

#### 4. LASER PROCESS

When all the layers are made it in the AutoCAD software the next step to do, is take the file to the computer, the laser can read this file and make a process similar to a printing paper when some one like to print a page on a computer.

Before attempting the fabrication of the combustor, we calculated the number of LTCC tape sheets needed for the laminated structure. The DuPont 951 LTCC is about 4 mils thick (around 100 µm) and for one of the possible combustor inserts thickness (2mm), at least 50 layers were needed. ***"Miguel Perez Tolentino, Study on the Implementation of sintered LTCC and Graphite as a sacrificial material for the fabrication of Microcombustors"***



Fig7. X-660 Laser Platform

The X-660 laser platform makes the layer similar of the one that have been defined in AutoCAD. Spending some times putting LTCC layers in the machine and programming the laser to use a speed and power of 5% give us a layer that look like the one on the picture:

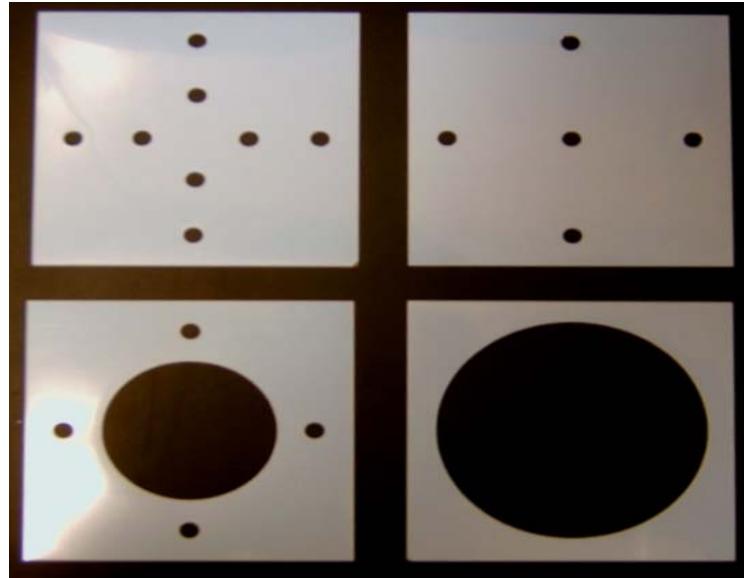


Fig8. Layers created using the laser printing process

#### 4.1 Lamination Process

Having the layers finished on the laser the next process is to laminate using the LTCC tape that we get from the laser. Using a high quantity of layers can be more difficult to make this process because there are four types of different geometry and second it have to be patient when putting this quantity of layer, sometimes they broke.

Lamination is a process utilized to bond all the LTCC tape sheets as to construct 3-D structure upon heating under a stress. When laminating LTCC tapes, it is important to keep all the sheets consistently with the same side up, that is, the LTCC tapes are fabricated over a Mylar sheet. To facilitate the release of the LTCC from the Mylar, a lubricant is utilized. It is important that the “shiny” side (side facing the Mylar) is always up or down. ***“Miguel Perez Tolentino, Study on the Implementation of sintered LTCC and Graphite as a sacrificial material for the fabrication of Microcombustors”***

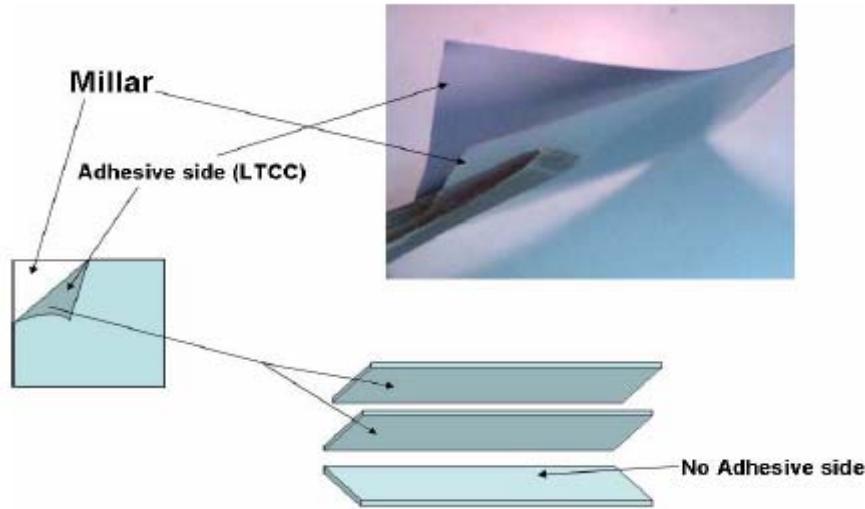


Fig.9 Consistent orientation of LTCC sheets before lamination.

The LTCC sheets at both ends of the laminate (top and bottom) are protected from the hot platens in the hydraulic press by Mylar sheets. During lamination, the stress and time are controlled for best results. In our case we laminated at 1000 psi for 20 minutes at a platen temperature of 80 C.

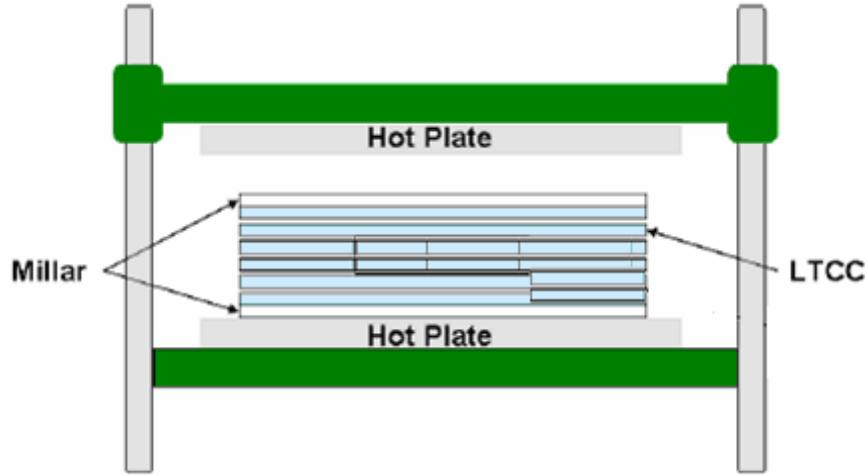


Fig.10 Scheme of the Lamination Process for the combustor using heated press (Carver Model C)

When we remove the structure we measure the thickness that was around .5mm using at least 50 layers and it looks this way when a picture from the top part was taken.



Fig.11 Top view of the micro combustor when the lamination process was done.

This is the top part of the micro combustor where the transducer is going to be placed. The flame is going to be heating the transducer and we the seebeck effect, we can transfer heat to electrical power.

## 4.2 FURNACE PROCESS

The last process to do is putting the structure in a furnace. The heating schedule as programmed in the resistance furnace:

- From room temperature to 300° C at a rate of 10° C/min.
- Kept at 300 C for 30 minutes
- Ramp from 300 C to 850° C at a rate of 10° C/min
- Kept at 850° C for 1 hour and 45 minutes.
- Turn furnace off and let it cool to room temperature.



Fig11. Isotemp Programmable Forced-Draft Furnace (Fisher Scientific)

Because we ran out of time we don't finish this process, but the configuration of the furnace is the same that in the past were done by other students. But this process can improve the tight of the structure, in that way when the experimentation goes on the micro combustor can hold the pressure inside the structure.

## **CONCLUSIONS**

First we obtained that FemLab simulation show a better mixture of oxygen and propane gas in the 3D cylindrical micro combustor. In that way we improve the micro combustor compared for the one that have been done in the past. The process of design and build of this cylindrical micro combustor it is not too easy because the multiple layers and the four design in AutoCad it takes some times when doing the lamination process.

We learn too that the laser process can do the layers in terms of minutes. Using a 5% of speed and power we can obtain the four layer using LTCC tape. The dimension of the micro combustor was 2.5"x2.5" it can be improve to doing much smaller, but simulation and analysis have to be done to achieve this idea.

We don't make any experimentation with this device because we don't have much time left to do it, but as I said before simulation show a nice mixture of gases because the inputs and outputs the device have it. And it works more efficiency because it have different chamber to do the combustion process.

## **RECOMMENDATIONS**

A good recommendations is to first do a lot of simulation using FemLab because in that way can see the effect of the mixture of the gases in the structure. Another one is to improve the dimension of the structure so it can be done smaller that the one we design.

Because we don't have much time left another good recommendation is to put graphite inside the structure so in that way when some one put the devices in the heat plates and in the furnace the LTCC can't break easily.

Last recommendation is to put some type of glue around the device so in that way it can prevent any gas leakage and can make the layers more stronger in that way.

## **ACKNOLEDGMENTS**

I wish express the thanks, first for the Professors Jorge Santiago and Rogerio Furlan for having recommended me to apply to the Sunfest Program. I also thank to Dr. Jan Van Spiegel, the University of Pennsylvania, and University of Puerto Rico at Humacao, Sunfest Program and NSF.

## REFERENCES

- [1] Lesley Millar, “Novel Microcombustor for Highly Efficient Generation of Electric Power from Fuel”, *Office of Technology Management*, University of Illinois at Urbana.
- [2] Shuji Tanaka, Takashi Yamada, Shinya Sugimoto, Jing-Feng Li, and Masayoshi Esashi “Silicon nitride ceramic-based two-dimensional microcombustor”, *Journal of Micromechanics and Microengineering*, 13 (2003) 502–508.
- [3] Lars Sitzki, Kevin Borer, Ewald Schuster and Paul D. Ronney, “Combustion in Microscale Heat-Recirculating Burners”, *The Third Asia-Pacific Conference on Combustion*, Seoul, Korea, June 24-27, 2001
- [4] D. G. Norton, K. W. Voit, T. Brüggemann, and D. G. Vlachos, “Portable Power Generation Via Integrated Catalytic Microcombustion-Thermoelectric Devices”, Department of Chemical Engineering, University of Delaware.
- [5] Miguel Perez Tolentino, Rogerio Furlan, Idalia Ramos and Jorge J. Santiago Aviles “Study of Laser Milling Of Sintered LTCC and Quartz Substrates for Microfluidic Applications”, *The National Conference On Undergraduate Research (NCUR) 2005*, Washington and Lee University Virginia Military Institute Lexington, Virginia, April 21– 23, 2005.
- [6] M.R. Gongora-Rubio, P. Espinoza-Vallejos, L. Sola-Laguna, J.J. Santiago-Aviles, “Overview of low temperature co-fired ceramics tape technology for meso-system technology (MsST)”, *Sensors and Actuators A* 89, 2001, pp. 222-241.
- [7] E.W. Simoes, I. Ramos, L. Garcia, R. Furlan, J.J. Santiago-Aviles, and M.T. Pereira, “Numerical modeling of a process for definition of features in low temperature co-fired ceramics”, in *Electrochemical Society Proceedings*, Volume 2002-8, Porto Alegre (Brazil), SBMicro 2002, September 9-14, 2002, pp. 244-252.
- [8] Miguel Perez Tolentino, Jorge Santiago Aviles, “*Study on the Implementation of sintered LTCC and Graphite as a sacrificial material for the fabrication of Microcombustors*” SUNFEST Report TR-CST01DEC05, Department of Electrical and Systems Eng. University of Pennsylvania, 2005, pp. 136-155.

*University of Pennsylvania*

SUNFEST

NSF REU Program  
Summer 2006

## **FORMING VESICLES FROM CARBON NANOTUBES**

NSF Summer Undergraduate Fellowship in Sensor Technologies  
Alexsandra Fridshtand (Bioengineering) – Lehigh University  
Advisor: Dr. Jennifer Lukes

### **ABSTRACT**

Carbon nanotubes (CNTs) are essentially layers of graphite seamlessly wrapped into cylinders. They have shown great potential in a wide variety of applications. One prospect in bioengineering is to utilize CNTs for drug delivery by forming micelles or vesicles from aligned tubes. These structures would essentially be spherical nanocontainers with short carbon nanotube arrays forming the wall of the chamber. The nanotubes that constitute these arrays can be made amphiphilic by functionalizing only one end of the tube with a hydrophilic molecule, since CNTs are naturally hydrophobic. Amphiphilic molecules naturally configure so that their nonpolar ends are away from aqueous solution (in the core of the structure) while the polar ends are next to the aqueous solution (on the outside surface). Nanotubes that have been functionalized with a polar molecule on one end could be influenced to aggregate into such arrangements by altering the dimensions of the tubes, since short and straight amphiphilic nanotubes have a good chance of forming micelles spontaneously. The nanotubes would mimic the behavior of phospholipids in water, self-assembling into micelles in aqueous solutions. If successfully developed, these micelles could be used for transport of poorly soluble drugs, such as many anticancer agents. Vesicles, which have a bilayer membrane of nanotubes, could also be formulated, and could be used to transport water-soluble drugs since both the inner and outer walls would be hydrophilic.

## Table of Contents

1. Introduction.....	3
2. Background.....	4
2.1 Micelles and Vesicles.....	4
2.2 Carbon Nanotubes.....	7
3. One-End Functionalization.....	9
3.1 Chemical Process.....	11
3.2 Confirmation of Functionalization.....	13
4. Dispersion.....	14
5. Phase Determination.....	16
5.1 Concentration.....	16
5.2 Amphiphilic Molecule Dimensions.....	18
6. Experimental Variables.....	20
6.1 Nanotube Concentration.....	20
6.2 Types and Amount of Surfactant.....	21
7. Visualization.....	21
7.1 Electron Microscopy.....	21
7.2 Fluorescence.....	22
8. Conclusions and Recommendations.....	23
9. Acknowledgements.....	23
10. References.....	24

Appendix A: Data Sheet for  $\gamma$ -Secretase Inhibitor IX

Appendix B: Price Quotations

Appendix C: Data Sheets for Fluorophores

## **1. INTRODUCTION**

In recent years, new drug delivery systems have been developed that are more efficient, safer, and more beneficial than the traditional forms of drug administration. Such new approaches have allowed novel therapies to be established that can be used to treat conditions that have been previously incurable. One of the most prominent drug delivery methods is the entrapment of drugs in small spherical containers called vesicles which can then be injected into the bloodstream of a patient. This mechanism can be used to deliver medicine directly to the site in need of treatment, and has been specifically tailored to target cancerous tumors.

Vesicles can be made of substances such as proteins, carbohydrates, and synthetic polymers, but vesicles made from lipids, called liposomes, are by far the most widely studied. These structures are generally very tiny, their dimensions being on the nanoscale. They exhibit remarkable properties such as biocompatibility, degradability, and nontoxicity that allow them to travel with ease throughout a patient's body. However, existing vesicles are not without their problems, such as the poor stability of liposomes *in vivo* [17]. In this paper, carbon nanotubes will be introduced as a possible alternative in creating vesicles to transport drugs, instead of lipids and polymers.

The goal of this project was to investigate how an artificial vesicle could be developed from carbon nanotubes in order to be used to encapsulate and transport drugs within a patient's body. Nanotubes would need to align parallel to each other (just as lipids do) to make up the membrane. In order for this to be achieved, the carbon nanotubes would need to mimic the behavior of phospholipids, which make up naturally occurring vesicles found in cells. Phospholipids are amphiphilic molecules, that is, they have a hydrophobic (water-fearing) region and a hydrophilic (water-loving) region. Because the two ends of phospholipids have different polarities, they spontaneously assemble into different phases when placed in aqueous solutions so that they hydrophilic ends are near the water and the hydrophobic ends are away from the water. To form such phases from carbon nanotubes, they too must be made into amphiphilic molecules so that self-assembly is made possible. This can be achieved by functionalizing one end of the nanotubes with a hydrophilic molecule, since carbon nanotubes are intrinsically hydrophobic.

Once carbon nanotubes have been functionalized on one end with a polar molecule, experiments can be run to determine what types of conditions are most preferable for the nanotubes to assemble into aggregates such as bilayers, micelles, and vesicles. The dimensions of the nanotubes as well as their concentration in solution can be altered in order to observe how these variables affect what phase the nanotubes will form in aqueous solutions. Also, different surfactants can be added to the solution to better disperse the nanotubes and establish whether they aid the formation of aggregates from carbon nanotubes.

## 2. BACKGROUND

### 2.1 Micelles and Vesicles

Naturally occurring micelles and vesicles are made up of phospholipids, which are amphiphilic molecules consisting of a water-loving (hydrophilic) region and a water-fearing (hydrophobic) region. These amphiphilic molecules spontaneously aggregate into structures such as micelles, inverse micelles, bilayers, and vesicles when placed in aqueous solutions such as water in order to keep the hydrophobic regions hidden from water and the hydrophilic regions near water. These types of formations are commonly found in animal cells: a bilayer forms the cell membrane as well as organelle membranes, and vesicles are used to transport substances in and out of the cell, typically to either excrete waste or to import nutrients. Because of the tiny size of typical vesicles inside a cell (about 20-50nm), and their inherent function to hold and transfer various materials, they are the perfect candidates for drug delivery systems [0].

Employing micelles and vesicles made of various materials as drug carriers is currently an area of intense research. These types of structures have many attractive properties such as good biocompatibility, small size, and high stability both in vitro and in vivo [1]. They can also be made to have other, more specific, characteristics including slow degradability or reaction to certain stimuli (temperature, pH, etc.). The carriers should also be long circulating, so that they can “slowly accumulate in pathological sites with affected and leaky vasculature (such as tumors, inflammations and infarcted areas) via the enhanced permeability and retention effect (EPR) and enhance drug delivery in these areas” [3]. In this way micelles and vesicles can be used to transport drugs, such as anti-cancer agents, to a target site such as a tumor.

Since micelles have an interior that is hydrophobic, they can be used to carry many poorly soluble pharmaceuticals. If the drug to be transported is hydrophilic, liposomes (artificial vesicles) can be created. These structures have a bilayer membrane, so that hydrophilic tails make up both the outside and inside surfaces while the hydrophobic regions are trapped in the middle of the membrane. Images of a micelle and a liposome are shown below:

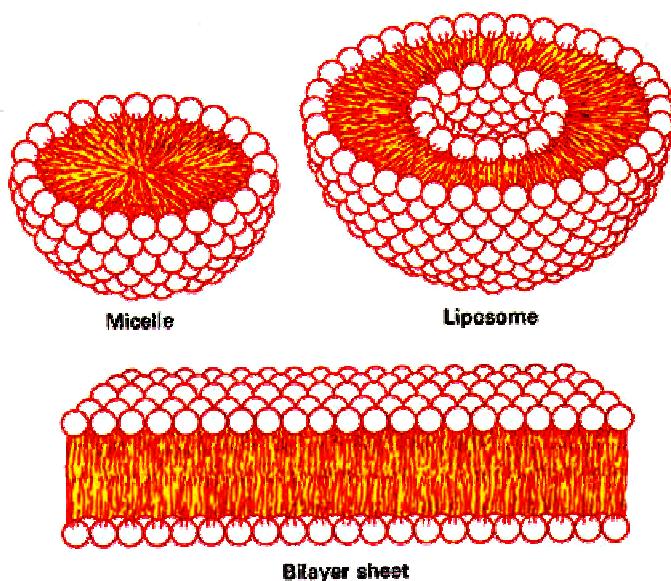


Figure 1 – Illustration of the different phases of amphiphilic molecules

Amphiphilic molecules can also spontaneously form straight bilayer sheets, also shown above. Which

of these three configurations are actually formed when amphiphilic molecules are placed in water depends on several conditions, the main ones being amphiphilic molecule concentration and the actual dimensions of the molecules.

Although the molecules used to form micelles are amphiphilic, they will still disperse in water in very low concentrations. The onset of formation of micelles is called the *critical micelle concentration* (CMC); when the concentration is increased above this value, the number of micelles formed increases. When the concentration reaches very high values, an inverted micelle phase can be formed, in which the hydrophobic regions are actually on the outside of the sphere and the little water present in solution is actually trapped inside the core of the micelle [0].

The shape and dimensions of the amphiphiles have a lot do with what type of phase will be formed in an aqueous solution. When the cross-section of the hydrophilic region of the amphiphile is large compared to that of the hydrophobic area, the hydrophobic regions will bundle in the core of the micelle since it is easy for them to fit all together, and a regular micelle will be formed. When the two regions are close to the same size, it is likely that a bilayer will be formed because it is not easy for either portion to cramp together in the center. If the hydrophilic part of the amphiphile is too small, an inverted micelle may form (as mentioned above) since the hydrophilic parts can fit together in the core while the larger hydrophobic regions will be on the outside of the micelle [30].

Both the concentration and shape of the amphiphiles in an aqueous solution tie in with the energy considerations that really determine which phase is most stable in a particular environment. Forming micelles reduces entropy, an unfavorable operation, since order is being made from a disorganized solution. However, it is favorable when there is minimal energy present, which occurs when a micelle is formed since the hydrophobic regions are away from water while the hydrophilic parts are near the water. Therefore, the respective amounts of entropy and energy must be balanced for micelles, or other aggregates, to form in a given situation [0].

By experimenting with different concentrations of amphiphilic molecules and by varying their dimensions, certain conditions can be established that would be preferable for the formation of vesicles (for the transport of hydrophilic drugs) or micelles (for hydrophobic drugs) in aqueous solutions. One anti-cancer drug that can be potentially used is  $\gamma$ -Secretase Inhibitor IX, which is hygroscopic, meaning it attracts moisture [Appendix A]. In this case, a vesicle would be needed to hold and transport this drug since it would behave like a hydrophilic substance. However, many other anti-cancer agents are intrinsically hydrophobic, so micelles that could be formed to carry such substances would also be of great benefit [1].

This project's goal was to learn more about the conditions that would allow carbon nanotubes to form such aggregates as micelles and vesicles. The notion of utilizing micelles and/or vesicles to transport various drugs has been investigated rather thoroughly, but most of these structures have been formulated from polymers, not carbon nanotubes. In these cases, polymer monomers have been made into amphiphilic molecules so that they will aggregate into polymeric micelles in aqueous solutions, with their hydrophobic regions in the core of the micelle and their hydrophilic tails toward the water [3]. These polymeric monomers are made to mimic the behavior of phospholipids, which is what will be attempted here with carbon nanotubes. Since carbon nanotubes are intrinsically hydrophobic, hydrophilic molecules will need to be attached at one end of each nanotube to make them into amphiphilic molecules. If this can be accomplished, the nanotubes will function in a self-assembling system when placed in aqueous solution by forming spontaneously under the right conditions [3, 4].

Carbon nanotubes are more suitable materials for this purpose than are polymers or other molecules for a few reasons. The intrinsic properties of nanotubes provide one advantage, and the mechanisms of drug release from the vesicle/micelle once it arrives at its target site provide another. The small size of carbon nanotubes allows them to form vesicles/micelles that could travel anywhere throughout the body without much obstruction. Since carbon nanotubes "are of the size where cells do not recognize them as harmful intruders", they are able to enter cells with ease [19]. The dimensions of most cells are in the micrometer range and the space within them is very congested, so ideal drug delivery systems must be in the nanometer range in order to properly interact with or enter a cell [20]. Also, nanotubes are very stiff and exceptionally strong, yet have remarkable flexibility and can resist fracture [19]. The rigidity of their structure would make aggregates formed from them very stable, with the ability to withstand harsh conditions present within the body [18, 33].

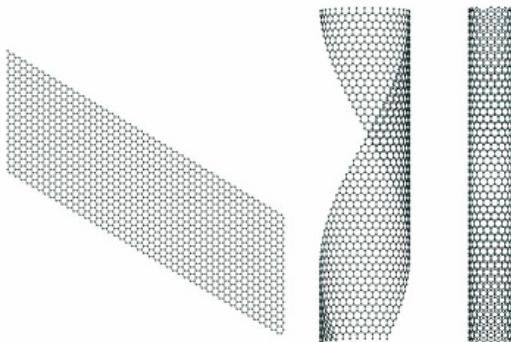
However, some may argue that these properties can be accomplished with certain types of polymers, or combinations of polymers. So the deciding factor is really how these micelles/vesicles will rupture at the target site and release their contents. The structures that have currently been made from polymers either degrade over time, which means micelles/vesicles that have not reached the target site will deliver their contents wherever they are in the patient's body, or are ruptured by a change in pH. This can be harmful for the patient's body since the pH balance of the blood is altered as a result [4]. Vesicles or micelles that are made from carbon nanotubes, however, can be made to rupture when a specific stimulus is applied, such as increased temperature [30]. This will avoid the problem of vesicles decomposing during delivery and releasing drugs that could be toxic. The heat can be applied directly to the tumor region using an infrared light, which is not harmful to the patient's body since skin and other biological systems are transparent to it [10]. This technique is also beneficial since it does not require the pH balance of the patient's body to be modified. Because carbon nanotubes intrinsically absorb infrared light, they can be optically stimulated to carry out different functions [10]. In this case, the carbon nanotubes would be heated up and caused to disassemble by the infrared light source, but no harm would be done to the healthy regions of the patient's body.

It is clear that if vesicles can be developed from carbon nanotubes, there will be many opportunities to make advances in drug delivery systems. Transporting pharmaceuticals in vesicles made of nanotubes will provide more precision in controlling the structure and organization of drug-delivery mechanisms [20]. But forming these vesicles may be a difficult

task, since the properties of carbon nanotubes are still being thoroughly investigated. In order to understand how to approach the venture of formulating vesicles or micelles, the structure and characteristics of carbon nanotubes must first be recognized.

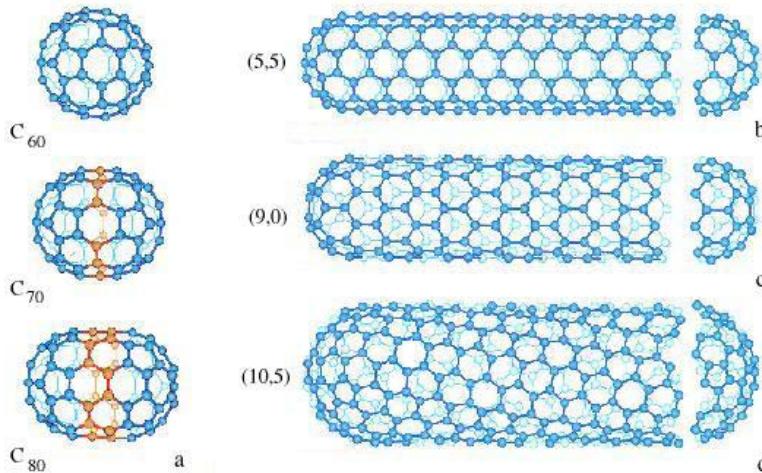
## 2.2 Carbon Nanotubes

A carbon nanotube is essentially a sheet of hexagonally arranged carbon atoms (a sheet of graphite, or graphene sheet) rolled seamlessly into a cylinder, as pictured below:



**Figure 2 – Illustration of a graphene sheet being rolled into a carbon nanotube**

A graphene sheet can be rolled into a cylinder in several ways, leading to different orientations of the lattice with respect to an arbitrary tube axis. These varying configurations are referred to as chiralities, and there are three types that nanotubes are generally grouped in: armchair, zigzag, and chiral nanotubes [33]. Carbon nanotubes (CNTs) can also be pictured as a hollow sphere of carbon atoms, or fullerene, that has been stretched along one axis to produce a cylinder that is capped with a hemisphere on each end. The three different types of nanotubes, along with their respective fullerenes, are shown below:



The varying  
as a functi

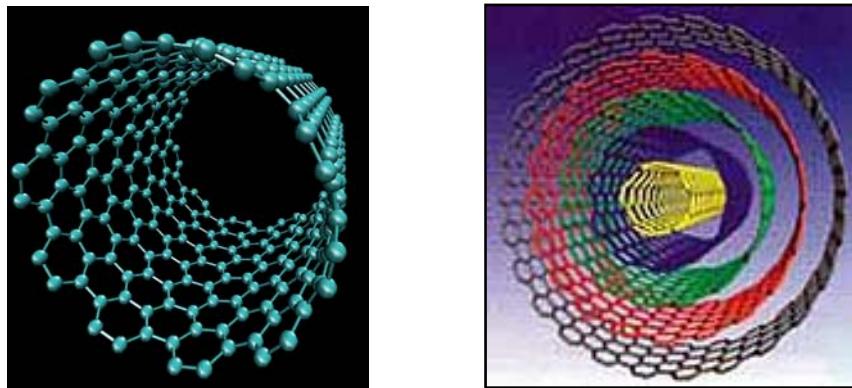
on properties change  
anotubes are  
metallic, whereas zigzag and chiral nanotubes can be either metallic or semiconducting. The configuration of a nanotube can be defined in terms of an integer pair  $(n, m)$ , which characterizes the tube's direction and diameter by the following equations:

$$d = [a \sqrt{(m^2 + mn + n^2)}] / \Pi$$

$$\theta = \arctan [-(\sqrt{3}n) / (2m + n)]$$

where  $d$  is the diameter,  $a$  is the lattice constant in the graphene sheet and  $\theta$  is the chiral angle of the carbon nanotube. The relation between  $n$  and  $m$  describes the three categories of nanotubes: 1) armchair  $\rightarrow n = m$  and chiral angle is 30 degrees, 2) zigzag  $\rightarrow n = 0$  or  $m = 0$  and chiral angle is 0 degrees, and 3) chiral  $\rightarrow$  other values of  $n$  and  $m$  and chiral angles between 0 and 30 degrees [33, 34]. Example integer pairs for each of the nanotube types are shown in Figure 3.

Besides the varying chiralities and diameters of nanotubes, there are also two major types of CNTs: single-walled nanotubes (SWNTs) and multi-walled nanotubes (MWNTs). Single-walled nanotubes consist of only one graphene cylinder, while multi-walled nanotubes have many concentric graphene cylinders nestled within one another and can have diameters much larger than single-walled tubes [23]. SWNTs typically have a diameter of about 1.4 nm, which is very close to the diameter of an ideal (10, 10) carbon nanotube. The spacing between the concentric layers in a MWNT is about 0.34 nm, which is close to the interlayer distance in graphite [18, 19, 33]. Examples of single-walled and multi-walled nanotubes are pictured below to allow for contrast:



**Figure 4 – Cross-sectional images**

Both single-walled and multi-walled nanotubes are insoluble in most solvents due to the great hydrophobicity of carbon [6]. Also, strong van der Waals forces between adjacent nanotubes compel them to bunch together in clusters or ropes [2, 22, 36]. Hence, both types of CNTs will bundle together in solution, especially if the solution is polar or aqueous [7]. In order to allow the manipulation of nanotubes, many steps have to be taken to disperse the nanotubes since they will clump together otherwise. Dispersion in various solvents is an important step to enhance the performance of CNTs, and usually involves techniques such as ultrasonication and centrifuging [2]. Other mechanisms have been described that use polymers to “wrap carbon nanotubes and render them soluble in water or organic solvents”, but this technique is not used in this project [5].

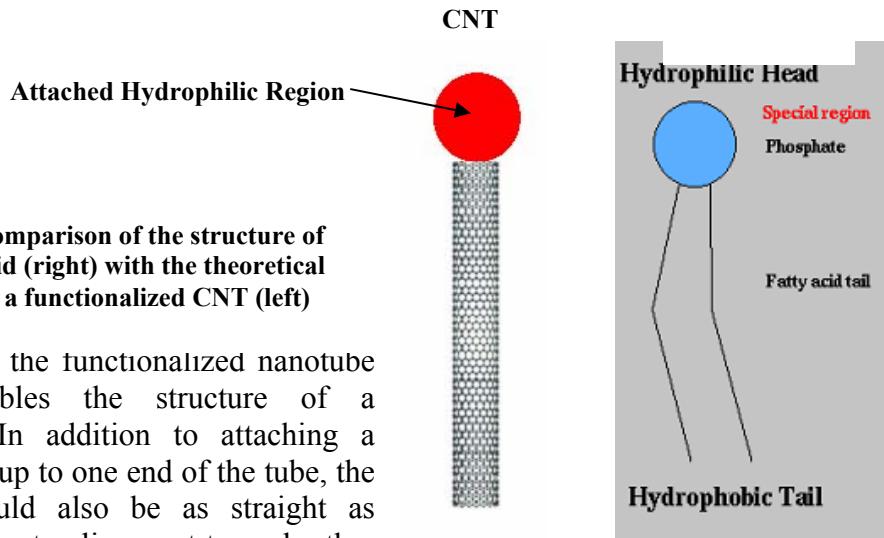
Solubilization of carbon nanotubes in different solutions can be achieved by functionalization with certain molecular groups or by adding surfactants to the solution [9, 31,

32]. Functional groups such as amines, alkylaryl amines, or carboxylic acids allow CNTs to be dissolved in various solvents [5, 21]. Numerous surfactants can be used to produce “stable aqueous dispersions” of carbon nanotubes by “coating the nanotubes and increasing the surface interaction between the nanotubes and the water” [32]. Sonication can also be used in combination with functionalization and surfactants in order to maximize the dispersion of CNTs in solution. Improved solubility of carbon nanotubes increases their ability to be processed and allows their unique properties to be coupled with other materials to produce innovative applications in a variety of fields [21]. Functionalization of nanotubes and coating with surfactants can also be used to achieve biocompatibility, which is especially important for applications such as drug delivery systems that will take place in vivo [8].

### 3. ONE-END FUNCTIONALIZATION

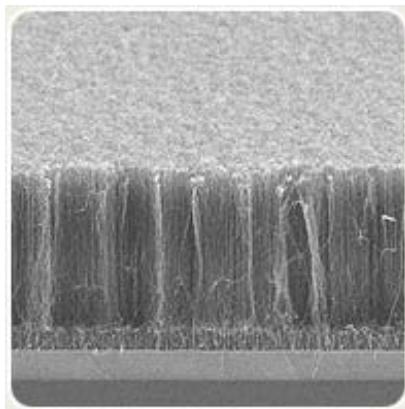
In order for carbon nanotubes to have the ability to self-assemble in aqueous conditions, a hydrophilic region must be introduced to one end of the nanotube so they become amphiphilic molecules. Since spontaneous self-aggregation is best known in phospholipids, carbon nanotubes will be used to mimic the unique behavior of these molecules.

Phospholipids are made up of two fatty acid tails, which are hydrophobic, and phosphate head, which is hydrophilic. This structure makes the molecule amphiphilic, and they naturally assemble to keep their tails away from water and their heads near water [0, 4]. To give carbon nanotubes the same feature, a hydrophilic molecule (or numerous molecules) needs to be attached to one end to mimic the phospholipids head, and the nanotube itself will act as the hydrophobic tail. The structure of a phospholipid is compared to a theoretical model of a functionalized nanotube below:



Clearly, the functionalized nanotube closely resembles the structure of a phospholipid. In addition to attaching a hydrophilic group to one end of the tube, the nanotubes should also be as straight as in order for them to align next to each other forming a micelle, vesicle, or bilayer. Because nanotubes ordinarily “have very small diameters and large length, they have very high aspect ratios” and are therefore “susceptible to structural instability” [19]. This property (along with the strong van der Waals attractions) causes CNTs to bend and twist around each other, resembling thin hairs that have been entangled into networks [22]. Hence, if straight nanotubes are desired, the CNT dimensions need to be modified to lower

the aspect ratio by increasing the diameter and shortening the length [16]. Specifying the length and diameter of carbon nanotubes is still a methodology that is being perfected, so it is difficult to control the exact proportions of the nanotubes. Nanotubes can be cut into smaller portions by ultrasound, but the broken pieces are not uniform in length and would not be well-suited to form aggregates such as micelles and vesicles [18]. In order to obtain carbon nanotubes of uniform dimensions, they are grown in parallel, aligned arrays on a substrate. This method allows more precision in controlling the length and diameter of the nanotubes. Specifically, the size of the catalyst particle at the tip of the nanotube determines the diameter of the tube, and the growth process can simply be stopped when the desired length is reached [29]. An image of a carbon nanotube array is shown below:



**Figure 6 – An aligned array of carbon nanotubes grown on a silicon substrate**

Since nanotube dimensions are on such a tiny scale, it becomes very difficult grow arrays of tubes that have a length smaller than about 1 micron (1000nm), purely because it becomes challenging to work with precision. The diameters, however, can be substantially smaller, ranging anywhere from 5-100nm, depending on the size of the catalyst particle used to grow each tube [29]. In order to acquire straight and uniform nanotubes, the diameter to length ratio should be relatively small so that the tubes are short and wide. Since the smallest length that can be obtained from the process is 1000nm, the diameter was chosen to be about 100nm, so that a ratio of 1:10 would be produced. Making the nanotubes short and thick would ensure that they would be straight and also would make them better able to align next to each other in different patterns [16]. Also, let it be clarified that the CNTs used will be multi-walled nanotubes since the diameter necessary is quite large and could not be achieved with single-walled tubes [23]. This is also an advantage because a tube with more walls has superior resistance to bending than one with just one wall [12].

The next step was to determine what molecule, or molecules, to attach to the ends of the nanotubes. The conditions that had to be met were that the molecule be hydrophilic and that it would bind relatively easily to the carbon structure of the CNTs. Numerous hydrophilic molecules were found that have been frequently attached to carbon nanotubes for various purposes. These included the hydroxyl group (-OH), the carboxyl group (-COOH), poly(*m*-aminobenzene sulfonic acid) (PABS), and polyethylene glycol (PEG) [3, 5, 8]. Since both PABS and PEG molecules are polymers, it was decided that they should not be used because their effects on how the nanotubes would aggregate together were unknown. The hydroxyl group was also disregarded due to the fact that this molecule reacts so readily with water and might separate from the carbon nanotubes when they are placed in aqueous solution. Therefore, it was decided

that the carboxyl group would be used to functionalize the ends of the nanotubes because of its hydrophilic properties, easy attachment to CNTs, and its frequent and customary applications.

Nanotubes that have been functionalized on just one end with a hydrophilic molecule, such as carboxylic acid (COOH) are not readily available. Because the chemistry behind one-end attachment is rather complex (the process is described in the next section), the decision was made to purchase nanotubes that have already been functionalized and put into solution. Many commercial companies were contacted; the majority of them did not know how to approach the task of attaching a molecule on just one end of the nanotube. Certain companies did offer functionalization, but this was usually just the addition of molecules all over the nanotubes – sidewalls as well as ends – accomplished simply by immersing the nanotubes in different substances, such as acids [11, 21]. Finally, a vendor, NanoLab, Inc., was found that was willing to perform a technique that would leave only one set of ends of the CNTs functionalized with carboxyl groups. The price quotations for the nanotubes can be found in Appendix B, and two 15x15mm<sup>2</sup> arrays (to be dispersed in distilled water) were ordered based on the third quotation (page 31) to take advantage of the discount offered when numerous arrays were ordered.

### 3.1 Chemical Process

The process to be performed by NanoLab, Inc. to functionalize just one end of the nanotubes with carboxyl groups is a unique technique that has not been done prior to this order. Because of the novel nature of the procedure, the lead time offered was very long (8-10 weeks). Unfortunately, this means that the nanotubes will arrive after the time allotted for this research project has passed.

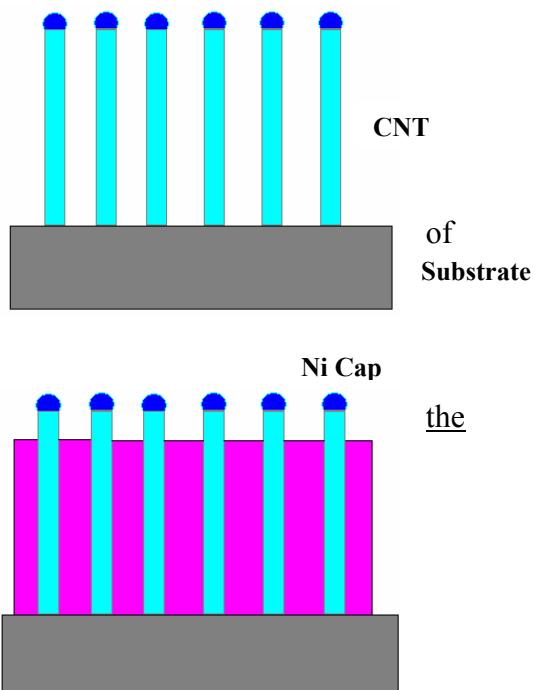
The rest of this paper will discuss what is to be done to continue this project once the functionalized nanotubes have arrived. The recommended methods and materials to be employed when this project continues were derived from other research endeavors that were either related to working with carbon nanotubes or with amphiphilic molecules that can act as models of nanotube behavior in aqueous solutions.

In order to attach carboxyl groups (COOH) to just one end of the nanotube NanoLab, Inc. will take the following measures:

1. Grow the nanotubes aligned on a substrate

This is achieved by placing nickel dots on a chromium coated silicon wafer, which will act as the substrate. The nickel catalyst will get lifted up and stay at the tip of the growing nanotube. The dot size determines the diameter the nanotube, which will be between 50-150nm. The growth time determines the length, and at the end there will be a cap of nickel covered by a closed-end nanotube.

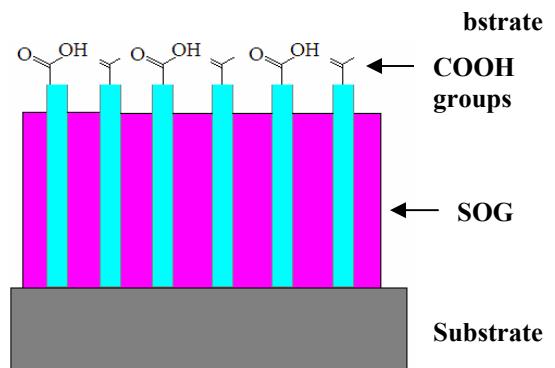
2. Fill in a matrix between the nanotubes, leaving only tips exposed



To avoid functionalizing the sidewalls, the array will be filled with spin-on-glass (SOG) to SOG the entire length of the tubes. The surface will then be polished with hydrofluoric acid (HF) to expose just the tips of the nanotubes.

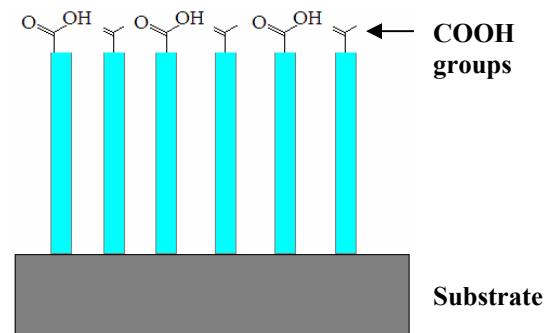
### 3. Functionalize the tips

To attach carboxyl groups to the tips, the array will be exposed to nitric acid ( $\text{HNO}_3$ ) or sulfuric acid ( $\text{H}_2\text{SO}_4$ ), which would open up the nanotube ends, remove the nickel, and leave -COOH groups at the tips.



### 4. Dissolve the matrix, leaving a free-standing nanotube array

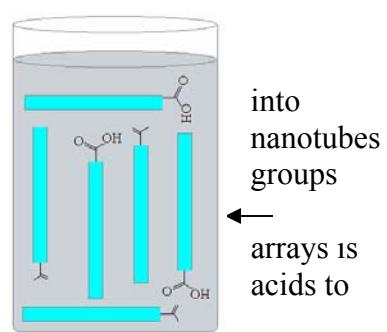
The sealed array would be placed in hydrofluoric acid (HF) to etch away the spin-on-glass matrix and leave the nanotubes functionalized and standing aligned on the substrate.



### 5. Strip the nanotubes from the substrate

The nanotubes need to be cleaved from the substrate, and put distilled water. The end result of this detailed process will be in surfactant-free solution (distilled water) that have carboxyl attached on one end.

A very similar approach to functionalizing nanotube discussed in one source, but with the goal of attaching nucleic the nanotubes instead of carboxyl acid groups [14]. In this examination, multi-walled carbon nanotubes were also grown in an array using metal catalyst particles at the tips, and spin-on-glass (SOG) was deposited within the gaps between the aligned nanotubes. The researchers found that the “SOG film provided structural support to the carbon nanotubes, enabling them to retain their vertical configuration” during “oxidative treatment for the opening of closed CNT ends” and also made the nanotubes more compatible for coupling chemistry in aqueous solutions [14]. This demonstrates that the spin-on-glass will not only keep the sidewalls from getting functionalized in this current experiment, but will also ensure the mechanical stability of the nanotube array, which can often collapse during various treatments [14].



### 3.2 Confirmation of Functionalization

Once the carbon nanotube order has arrived, the presence of carboxyl groups ( $\text{COOH}$ ) on the ends of the CNTs must be confirmed. This can be accomplished by using Fourier Transform InfraRed (FTIR) spectroscopy, which is used to identify organic materials by measuring their absorption of various infrared light wavelengths [28]. Different materials can be identified by comparing the spectrum of an unknown substance to that of a known, and preferably similar, substance. In this experiment, it would be preferable to compare the spectrum of the functionalized carbon nanotubes to that of plain carbon nanotubes to see if the carboxyl group peak can be observed. It would also be advantageous for the two types of nanotubes to have been prepared by the same method, so that peaks that occur as a result of amorphous carbon or impurities will appear on both spectra. Absorption bands in the frequency range of 4000-1500 $\text{cm}^{-1}$  are typically due to functional groups.

More specifically, the carboxyl group ( $\text{COOH}$ ) has three different types of bonds,  $\text{C-O}$ ,  $\text{C=O}$ , and  $\text{O-H}$ , and therefore has three different peaks that will appear. The  $\text{C-O}$  peak will typically appear in the 1260-1000  $\text{cm}^{-1}$  region, the  $\text{C=O}$  in the 1760-1670  $\text{cm}^{-1}$  region, and the  $\text{O-H}$  in the 3000-2500  $\text{cm}^{-1}$  region. Besides frequency, peaks can also be classified by their intensity and shape. For example, the  $\text{C-O}$  and  $\text{C=O}$  bonds have strong peaks, and the  $\text{O-H}$  bond has a broad peak. Below is an example of an FTIR spectrum of 2-bromobutanoic acid, a carboxylic acid:

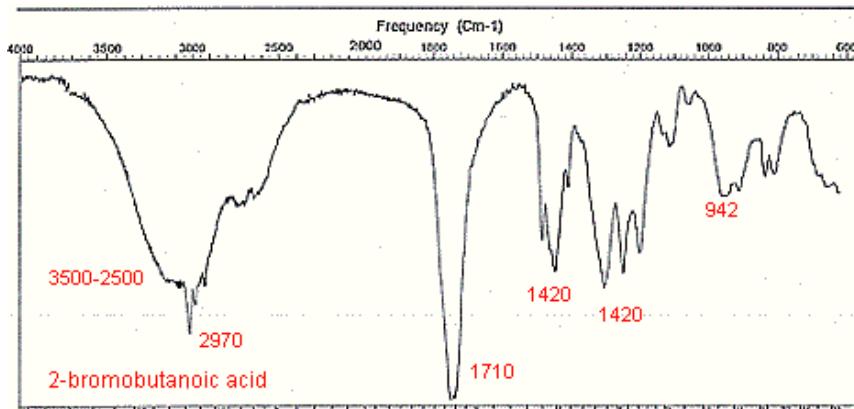


Figure 7 - FTIR spectrum of 2-bromobutanoic acid showing the peaks for  $\text{COOH}$  bonds

Although FTIR spectroscopy can be used to detect types of bonds, it is very difficult to quantify results with this method. For example, it would be very useful in this experiment if the number of carboxyl groups on each end of the nanotubes could be determined. But this would be nearly impossible to do with FTIR because the only way concentration can be ascertained is if it is compared to an established curve of known concentration of the same material.

In order to get quantitative results, a method called thermogravimetric analysis (TGA) could be used to assess the mass-fractions of different materials. This type of testing determines changes in the weight of a sample as a function of either time or temperature. The TGA technique would allow the evaluation of the number of carboxyl groups that are attached at the terminus of each nanotube, which would provide information as to what the dimensions of the

hydrophilic region are like in comparison to the hydrophobic region of the tubes. One source mentions using acid-base titrations to determine the concentration of carboxyl groups attached to carbon nanotubes, which can also be investigated when doing this experiment [11]. This information would assist in determining what type of phase the nanotubes would form most readily in aqueous solution, and will be discussed further in Section 5.

#### 4. DISPERSION

As previously mentioned, carbon nanotubes naturally clump together into ropes or bundles due to their hydrophobicity and strong van der Waals interactions [24]. Dissolution of CNTs is almost always necessary to conduct proper chemical reactions and for utilization of the tubes [23]. In order to disperse individual carbon nanotubes in solution, the “thermodynamic drive towards bundling must be overcome” [24]. This is commonly done by either ultrasonication to physically disperse the CNTs or by coating the tubes with surfactants to increase their solubility in solution. Surfactants provide an additional repulsive force to separate the nanotubes and prevent aggregation while sonication acts as an “external mechanical energy” source which helps to overcome the attractive van der Waals forces [13].

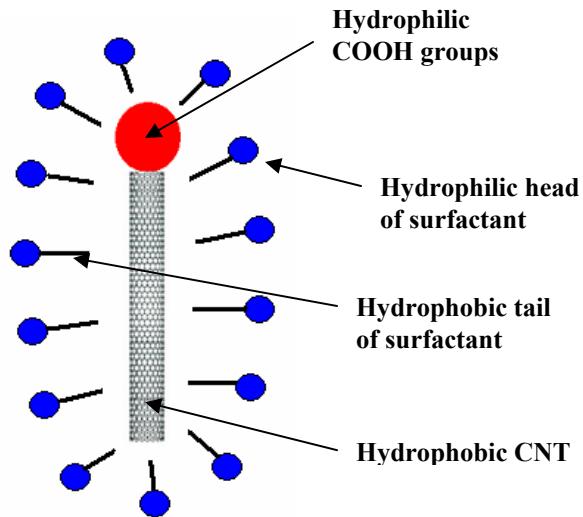
It is suggested here that when the functionalized carbon nanotubes arrive, a small portion should be taken out and sonicated while the rest of the functionalized nanotube solution be stored in a safe spot. The sonicated portion can then also be divided into different amounts, one of which will be used for FTIR in order to confirm that the carboxyl groups are indeed attached. The other samples should be sonicated for different amounts of time and observed, to see what the optimal sonication time is for maximum dispersion of tubes. As a side note, if there are impurities in the solution when the functional nanotubes arrive, they can be removed by centrifugation. This process is commonly used to remove graphitic particles or other constituents larger than 500nm from solution [13].

Since the goal of this experiment is to form vesicles or micelles from carbon nanotubes, it would be preferable to achieve good dispersion with just the sonicating technique and to avoid using surfactants. The reason for this is that surfactants themselves are amphiphilic molecules that will strive to self-assemble into micelles or other such aggregates when placed in aqueous solutions [25]. This is precisely why they work well to separate nanotubes and make them more soluble – they surround each nanotube with the hydrophobic regions near the tube and the hydrophilic regions away from the tube to form elongated micelles that surround the entire length of the CNT [31].

While this is a useful outcome when solubilizing non-functional carbon nanotubes, this experiment makes the matter a bit more complex since the nanotubes themselves have been made into amphiphilic molecules. Since it is desired that the nanotubes aggregate together into certain structures, it cannot be known what type of effect the surfactant molecules will have on their self-assembly since they will coat the entire perimeter of the nanotubes [25]. Although the surfactant molecules will help to separate individual nanotubes, they might also prevent the tubes from aligning parallel to each other to form bilayers and vesicles. Also, since there will be a hydrophilic region on the functionalized nanotubes, it is not clear how the surfactants will behave in those areas, as they might induce unwanted chemical reactions [13]. Another concern

is that the nanotubes will be made entirely hydrophilic when the surfactant molecules surround them, since all the hydrophilic regions from the surfactant will be facing the outside. This is illustrated below:

**Figure 8 – Demonstration of how amphiphilic surfactant molecules surround a functional nanotube. Since all of the hydrophilic surfactant heads are on the outside, this could potentially make the entire nanotube hydrophilic**



If the functional nanotubes are made entirely hydrophilic, they will no longer have the amphiphilic properties needed for self-assembly into aggregates such as micelles, vesicles, or bilayers. Hence, the attachment of the carboxyl groups to single ends of the nanotubes will have just been a wasted effort since they will still not display amphiphilic properties. There is no way to know exactly how the surfactant molecules will affect the nanotubes' ability to self-aggregate except to run experiments and observe the results. For these reasons, it is proposed that ultrasonication of the nanotubes be performed numerous times first to establish whether that process is substantial for dispersing the nanotubes. If it is found that the nanotubes cannot be separated well by sonication alone, then different surfactants can be tested to see how they influence the behavior of functional carbon nanotubes. If surfactants are indeed utilized, there might also be a need for the added complication of removing the surfactant after processing so they do not interfere with the nanotube self-assembly [15]. In summary, surfactants should not be used unless it has been determined that they do not agitate the amphiphilic characteristics of the functional CNTs, which are vital in order for the nanotubes to spontaneously assemble into formations when placed in aqueous solution.

## 5. PHASE DETERMINATION

Probably the vaguest aspect of this research project is how to establish what phase (micelle, vesicle, bilayer, or other) will be formulated when the amphiphilic nanotubes are placed in aqueous solution. There has been no prior research conducted on making carbon nanotubes amphiphilic or how these nanotubes will assemble in water. Therefore, alternate amphiphilic molecules must be researched and it must be assumed that carbon nanotubes will act similarly when in the same environment. By far the most well-known amphiphilic molecules are phospholipids, which make up the cell membrane as well as many other cell organelles. In order to understand the conditions required for functional nanotubes to self-assemble, phospholipids will be used as models of their behavior. It should be understood that the way in which the

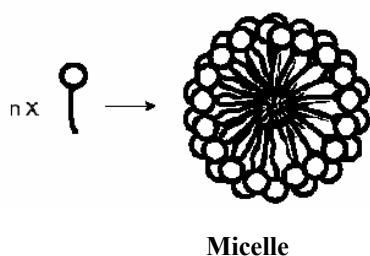
amphiphilic nanotubes function might vary from how phospholipids function since phospholipids are much smaller (~5-10nm in length) than the nanotubes used here (~100nm in length), and they have some differing characteristics due to their chemical composition [0]. Nevertheless, assuming that the nanotubes will act similarly to the way phospholipids act in specific situations is the only way to establish certain parameters for experimentation. The following information about amphiphilic molecules was derived from sources discussing phospholipid self-assembly and behavior.

There are two main variables that determine what phase amphiphilic molecules will fall into when they exist in aqueous solution: the concentration of the amphiphile and the shape of the amphiphilic molecules. The concentration ascertains whether or not the molecules will aggregate at all, while the dimensions of the molecules determine which phase will most likely be formed – micelle, inverted micelle, or bilayer. One may notice that a vesicle is not included in the phase options; this is because a vesicle is formed by bending an already existing bilayer, meaning that a bilayer would have to be formed in solution first. It may be possible to create recipes for specific aggregates by altering the concentration and shape of the amphiphilic nanotubes during experimentation and analyzing the way in which they assemble.

## 5.1 Concentration

The concentration of amphiphiles in a given aqueous solution is an important factor in concluding whether or not a phase will form. The effect of varying concentration is mainly analyzed by discussing the onset of formation of micelles in solution. The reason for this is that micelle formation is affected primarily by concentration while the formation of a bilayer is affected by the shape of the molecule. To clarify, a bilayer will only form when the appropriate concentration for micelle formation is reached, but the shape of the amphiphilic molecules prefers a bilayer. So when the concentration to form micelles is discussed, one can think of it as the concentration necessary to form either the micelle or bilayer phase [0].

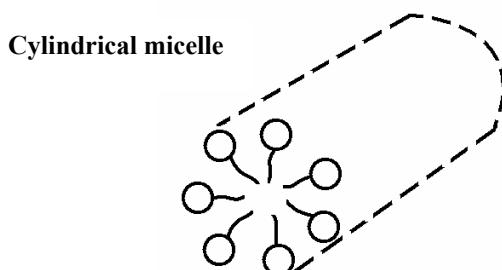
When the concentration of amphiphilic molecules in solution is below the *critical micelle concentration* (CMC), the molecules form a single layer on the liquid surface and disperse throughout the solution. At the critical micelle concentration, the molecules will organize into spherical micelles, with the hydrophobic tails on the inside and the hydrophilic heads on the outside, as shown here [0].



**Figure 9 – A spherical micelle formed from  $n$  amphiphilic molecules**

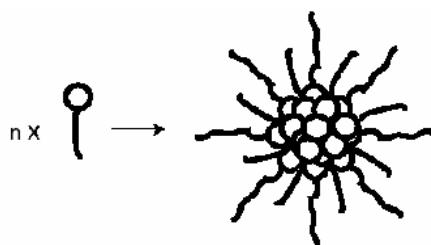
The CMC depends on the chemical composition of the amphiphilic molecules, mainly on the ratio of the head area and the tail length. As the concentration of amphiphilic molecules is increased beyond the CMC, more spherical micelles will be formed and the amphiphiles might also start to form elongated cylindrical micelles, also called hexagonal micelles, shown below

[30]. If the concentration is slightly further increased, there is an interesting effect: the number of micelles will increase but the sizes of the micelles stay pretty constant and are only weakly affected [0].



**Figure 10 – A cylindrical micelle formed from amphiphilic molecules**

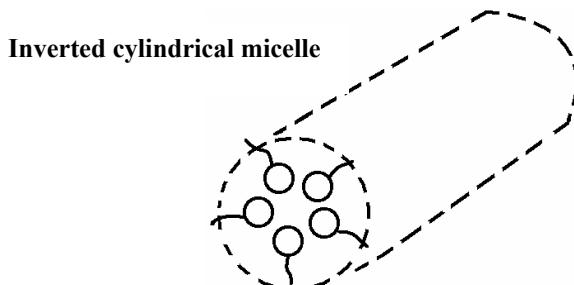
Increasing of the concentration still further will lead to increasingly more micelles formed until such a high concentration of amphiphiles is reached that inverted micelles will begin to be formed. At this point the amount of amphiphilic molecules is so high that the aqueous solution is actually trapped in the interior of the micelle, and the micelle is formed with the hydrophobic regions on the outside and the hydrophilic heads on the inside [0, 30]. A spherical inverted micelle is illustrated below.



**Figure 11 – An inverted spherical micelle formed from  $n$  amphiphilic molecules**

**Inverted micelle**

Finally, when the amphiphile concentration is kept extremely high, inverted cylindrical micelles may also form in addition the spherical inverted ones. Again, the aqueous solution will actually be trapped on the inside of the structure, next to the hydrophilic heads, since there will be far more amphiphiles than water molecules in solution [30].



**Figure 12 – An inverted cylindrical micelle formed from amphiphilic molecules**

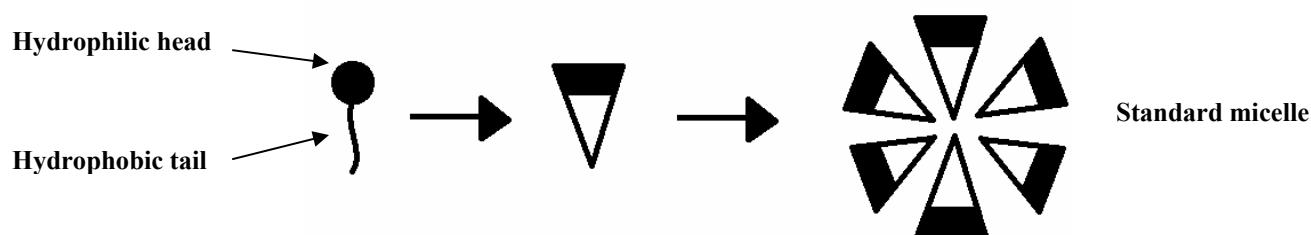
The amphiphilic molecules exemplified here have only one hydrocarbon tail, whereas phospholipids have two hydrophobic tails. They were illustrated in this way because the functional nanotubes that this project is concerned with will be more similar to a lipid with just one hydrophobic tail, since the CNT itself will function as the single hydrophobic region. The difference between single- and double-chain molecules is that single-chain amphiphilic molecules have a higher critical micelle concentration than double-chain molecules. The reason

for this is that the double-chain molecules have twice as many carbon atoms for the same length. Since they have a larger hydrophobic region than single-chain molecules, they have a higher tendency to form aggregates at low concentrations, and therefore have lower CMC values. By the same principle, CMC values decrease as the lengths of the hydrophobic regions are increased in either single- or double-chain amphiphiles [0]. In this case, since the functional carbon nanotubes will act like single-chain lipids, the critical micelle concentration is expected to be higher than that of double-chain lipids or other amphiphilic molecules.

## 5.2 Amphiphilic Molecule Dimensions

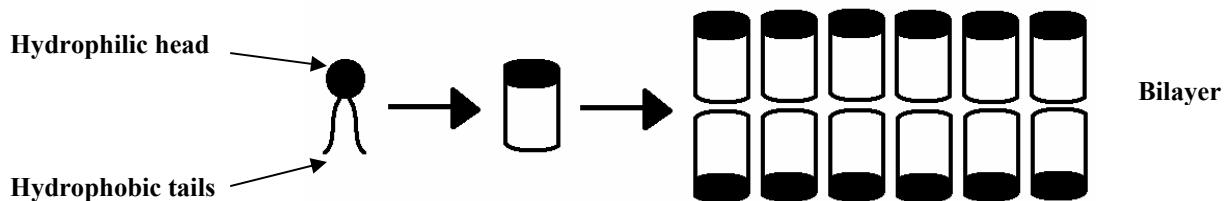
How amphiphilic molecules aggregate depends on the way in which they are able to align with each other, which depends on their spatial geometry [30]. The ratio of the hydrophilic head cross-sectional area to the cross-sectional area of the hydrophobic tail plays a major role in determining whether the amphiphilic molecules will form micelles, bilayers, or inverted micelles [0]. The size of the respective regions can vary due to hydrophilic head group size, the number of hydrocarbon chains present, and sometimes attractive or repulsive forces between adjacent head groups can also have an effect. Another important factor that influences what phase is most favorable in a given circumstance is temperature. When the temperature of a solution is increased, the motion and mobility of the hydrocarbon chains increases, which results in the hydrophobic region becoming larger than when the temperature was at a lower value. Hence, if a solution contained amphiphilic molecules that were forming micelles because the hydrophilic region was bigger than the hydrophobic, and then the temperature was increased, the phase would most likely change to the bilayer formation since the area of the hydrophobic region has increased [30].

When the head cross-sectional area is large compared to that of the hydrophobic region, the hydrophobic tails get cramped together in the center and a micelle is formed. This usually occurs if the hydrophobic chains are short or if there is just one hydrocarbon tail. In these cases the molecule resembles a wedge shape with the hydrophilic region as the wide portion and the hydrophobic region at the tip [0, 30]. It is easily visualized that the smaller hydrophobic regions will gather in the center of the micelle while the hydrophilic regions stay on the outside, as picture below.



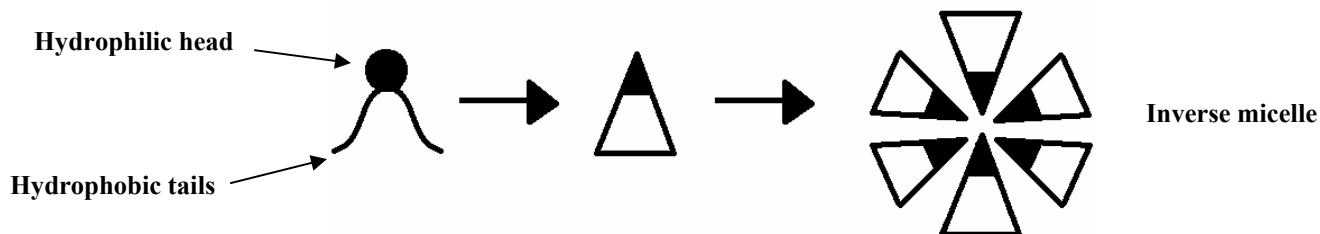
**Figure 13 – A micelle formed from amphiphilic molecules with large hydrophilic regions (black) compared to the hydrophobic portions (white)**

When the cross-sectional areas of the hydrophilic and hydrophobic regions are about equal, the amphiphilic molecules line up in parallel to each other to form a bilayer structure. These molecules have hydrocarbon tails that take up about the same width as their hydrophilic heads, and so resemble cylinders. The hydrophobic tails are again shielded from the water by being on the inside of the bilayer, while the hydrophilic heads are on the perimeter of the bilayer, next to the water. An illustration of this is shown:



**Figure 14 – A bilayer formed from amphiphilic molecules with hydrophilic (black) and hydrophobic (white) regions of about the same size**

If the cross-sectional area of the nonpolar hydrocarbon chains is larger than that of the polar (hydrophilic) head, the amphiphilic molecules will again look like wedges, but now with the hydrophobic region as the wide part and the hydrophilic region as the tip. According to the structures of the molecules, the wedges will want to align so that the hydrophilic tips are in the center of the micelle and the wide hydrophobic parts are on the outside, thus forming an inverse micelle [0]. However, if these molecules are located in aqueous solutions, this formation will be against energy considerations since the nonpolar regions would be placed near water, a very detrimental state. Therefore, if an inverse micelle is to be formed, a very high concentration of amphiphilic molecules is needed in addition to the reverse wedge shape. This way the water molecules can be trapped within the nonpolar interior of the micelle, while the hydrophobic regions stay on the outside next to the hydrophobic portions of the neighboring inverse micelles. Hence, a high amphiphilic concentration is needed so that all of the aqueous solution present can be confined within the structures.



**Figure 15 – An inverse micelle formed from amphiphilic molecules with small hydrophilic regions (black) compared to the hydrophobic regions (white)**

Since it would be quite difficult to try to alter the dimensions of the functional nanotubes once they arrive, only the concentration of the nanotubes could be experimented with in this project. However, determining the amount of carboxyl groups on the ends of the nanotubes will be of great benefit since that will help to determine the ratio between the hydrophilic and hydrophobic cross-sectional areas. Although the ratio could not be modified, it would still be of great use to know which phase would be favored by the geometry of the molecules alone. During experimentation, the concentration of nanotubes in solution should be altered and the samples observed in an attempt to find an optimal concentration to form each of the phases.

## 6. EXPERIMENTAL VARIABLES

### 6.1 Nanotube Concentration

As mentioned, experiments should be run at various nanotube concentrations in order to establish whether there is a critical micelle concentration (CMC) for the functionalized nanotubes. Optimal concentrations for the formation of bilayers or inverted micelles may also be found, in which case specific recipes can be developed for each type of phase.

One difficulty that arises is how to measure the concentration of nanotubes when they are in solution. Their size is so small that traditional methods of computing the quantity present in solution do not suffice. In one source, optical absorption spectroscopy is discussed as a technique to determine nanotube concentration. This approach involves recording the absorbance of light as a function of CNT concentration, which allows the amount of nanotubes to be measured even when dispersed in aqueous solution [35].

There have also been studies done to assess how nanotubes behave in relation to each other when placed in water, which would be a useful point to understand when trying to pick appropriate conditions for CNTs to self-assemble. One source discusses the forces that are present between two nanotubes in water, and suggests there are maximum and minimum attractive forces depending on the specific distance present between the carbon nanotubes [7].

### 6.2 Types and Amount of Surfactant

If using various amounts of nanotubes in combination with ultrasonication still does not show any useful results, different surfactants can be used to see if they facilitate the self-aggregation of the nanotubes. Although it is not clear whether the surfactant molecules will aid in assembly of the tubes or actually interfere with it, this is another variable that can be tested to see if there are any beneficial outcomes. Several surfactants that have been used successfully with carbon nanotubes are Triton X-100 [25], poly(methyl 2-methylpropenoate) (PMMA) [6], Gum Arabic (GA) [22], dimethylformamide (DMF) [28], and sodium dodecyl sulfate (SDS) [31]. Each of these, and other surfactants, can be examined to see what type of effects they have on the self-assembly properties of the functional CNTs.

## **7. VISUALIZATION**

One of the biggest challenges that arises when working with nanotubes is how to visualize them. Because of the tiny size of carbon nanotubes, traditional methods using microscopes do not have enough resolution to examine the CNTs closely. For this reason, two alternative techniques for visualizing the functional carbon nanotubes are discussed here.

### **7.1 Electron Microscopy**

The most common approach that many researchers utilize to visualize carbon nanotubes is electron microscopy, of which there are several forms. This category includes scanning electron microscopy (SEM), transmission electron microscopy (TEM), freeze-fracture electron microscopy, and many variations of each of them. However, preparation for each of these techniques is intensive, and the sample to be investigated must be fixed according to specific standards for each type of equipment. For example, SEM and TEM both require that samples must be dehydrated and fixed by either chemical or physical methods so that they can withstand a high vacuum environment. Since this research project is based on observing the behavior of carbon nanotubes while in aqueous solution, both of these techniques will not be successful since it is impossible to investigate CNTs in solution using either of these methods [27].

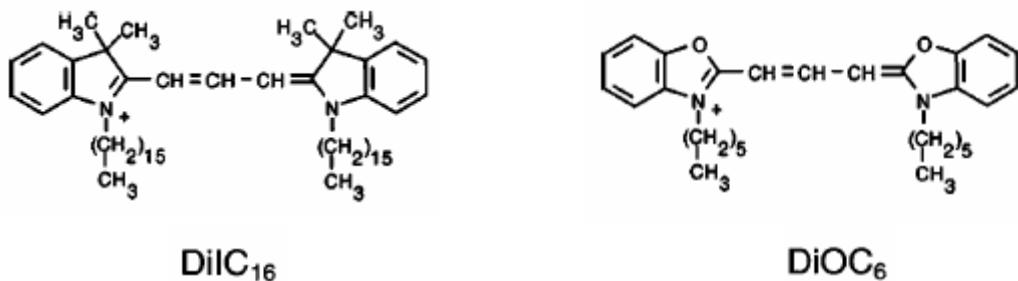
The only possible electron microscopy technique that could be applied to this project is the freeze-fracture method, in which a sample (which could be a solution) is frozen by liquid nitrogen and then cracked to expose the interior region of the sample. This approach would work in this case because the nanotubes and their aqueous surroundings would be frozen together in one block, and the nanotube positions could be examined when the sample is fractured. The advantage of this process is that the resolution that can be achieved by electron microscopy is very high, to about 0.2 nanometers, so the nanotubes could be observed very closely. However, the disadvantage is that the sample must be frozen, so it is not possible to detect how the nanotubes move around or interact with each other in real time. In order to examine these dynamic properties of nanotubes, another technique, fluorescence microscopy, is introduced.

### **7.2 Fluorescence**

Staining with fluorescent molecules, or fluorophores, allows carbon nanotubes to be distinctly visualized in solvents using simple light microscopes [27]. Although optical microscopes typically only have a resolution of about 200 nanometers, fluorescence of the nanotubes greatly eases their observation and manipulation, which enhances the visualization significantly [26]. This process allows nanotubes to be observed in real time under a simple microscope, so the motion and assembly of nanotubes can be assessed, and even recorded with video enhancement. Also, the conditions in which the nanotubes are present can be altered during observation so the effects can be immediately monitored. For example, the sample on the slide can be diluted with a drop of water while it is on the microscope and the reaction of the nanotubes can be immediately analyzed.

Two specific fluorophores have been proven to successfully stain carbon nanotubes in a study that tested six common dyes. The two that labeled the CNTs best were 3,3'-

Dihexyloxacarbocyanine iodide, DiOC<sub>6</sub>(3), and 1,1'-Dihexadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate, DiIC<sub>16</sub>(3). A suggested explanation of the success of these two fluorophores in comparison to others was that they had long nonpolar, hydrocarbon chains that could interact with the hydrophobic nanotubes. The structures of both molecules are shown below [26].



**Figure 16 – An illustration of the chemical structures of two common fluorophores that successfully label CNTs**

Although fluorescence microscopy does provide the opportunity to study the dynamics of carbon nanotubes by giving them fluorescent tags, there are several problems that may arise with this technique. The resolution with an optical microscope is always much worse than with electron microscopy methods. This may not be a dilemma at first because the nanotubes that are being worked with are relatively large in size (~100nm in diameter and 1.5 microns in length), but if this type of research continues on to work with smaller nanotubes, visualization of them may not be possible with fluorescence microscopy. Another issue that may arise is how these molecules will affect the ability of the nanotubes to assemble into phase formations. Just like with the surfactant molecules, the response of the functional nanotubes to the fluorophore molecules cannot be predicted. The only way to find out how the nanotubes will be affected is to visualize two identical samples using both fluorescence microscopy and freeze-fracture electron microscopy to assess whether the fluorophore molecules have any impact on the self-aggregation of the functional CNTs.

## **8. CONCLUSIONS and RECOMMENDATIONS**

In summary, this paper has discussed the topics that need to be addressed when attempting to form various phases, such as micelles, vesicles, and bilayers, from nanotubes that have been made amphiphilic. Once they arrive, the functionalized carbon nanotubes will be analyzed under diverse conditions, for instance, with different amounts of surfactant present in the water and different nanotube concentrations. Throughout experimentation with different variables, it would be worthwhile to determine whether there are optimal conditions for these nanotubes to form certain aggregates. Ideally, a sort of recipe could be developed for each type of phase so that they could be easily produced when desired. Things such as the type and amount of surfactant that is best to use would also be beneficial information to obtain. It would also be valuable to establish a critical micelle concentration for nanotubes that have been functionalized at one end with carboxyl groups, since nothing of the sort has yet been investigated. There may also be other concentration values that lead to bilayers or inverse micelles that would be useful to determine. Correlations between nanotube concentration and/or amount of surfactant in solution could be developed to show which aggregates form in which situations. This type of assay will illustrate which, if any, circumstances allow the nanotubes to form micelles, vesicles, bilayers, or other such aggregates most readily.

Another very important issue that should be investigated is how to make a carbon nanotube bilayer form into a vesicle. The only mechanisms currently known to form vesicles or liposomes are to alter the conditions of pre-existing bilayers. For example, if excess water is added to a solution that contains a lipid bilayer formation, the end areas are able to detach and round off to form a sealed spherical liposome [30]. After a bilayer has been successfully formed from CNTs, this will be the key step in forming a carrier for pharmaceuticals. The sort of information that could be acquired from this research project will surely lead to huge advances in the biomedical and drug delivery fields.

## **9. ACKNOWLEDGEMENTS**

I would like to thank my advisor, Dr. Jennifer Lukes of the University of Pennsylvania, for all of her advice and encouragement. I would also like to thank Dr. Jan Van der Spiegel of the University of Pennsylvania for coordinating and supervising the SUNFEST program. Finally, I would also like to thank the National Science Foundation for their ongoing support of the SUNFEST program, which presents undergraduate students with the opportunity to do valuable research.

## 10. REFERENCES

0. D. Boal, *Mechanics of the Cell*, Cambridge University Press, 2002, p. 137-150.
1. V.P.Torchilin, A.N.Lukyanov, Z.Gao, B.Papahadjopoulos-Sternberg, *Immunomicelles: Targeted pharmaceutical carriers for poorly soluble drugs*, Proc. Natl. Acad. Sci. USA, 100, 6039-6044, 2003.
2. Jeffrey L. Bahr, Edward T. Mickelson , Michael J. Bronikowski , Richard E. Smalley \* and James M. Tour \*, *Dissolution of small diameter single-wall carbon nanotubes in organic solvents?*, Department of Chemistry and Center for Nanoscale Science and Technology, Rice University, Houston, Texas, USA. The Royal Society of Chemistry 2001.
3. V.P.Torchilin, *Targeted polymeric micelles for delivery of poorly soluble drugs*, Cell. Molec. Life Sci., 61, 2549-2559, 2004.
4. V.P.Torchilin, *Recent advances with liposomes as pharmaceutical carriers*, Nature Rev. Drug Discov., 4, 145-160, 2005.
5. Robert C. Haddon, Hui Hu, and Bin Zhao, *Synthesis and Properties of a Water Soluble Single-Walled Carbon Nanotube-Poly(m-aminobenzene sulfonic acid) Graft Copolymer*, Adv. Funct. Mater. 2004, 14, No. 1 January.
6. Jing Liu, Tao Liu, and Satish Kumar, *Effect of solvent solubility parameter on SWNT dispersion in PMMA*, Science Direct, Polymer 46, 3419-3424, 2005.
7. T. Werder, J.H. Walther, and P. Koumoutsakos, *Hydrodynamics of Carbon Nanotubes – Contact Angle and Hydrophobic Hydration*, Institute of Computational Science, Zurich.
8. Hongjie Dai, Moonsub Shim, Nadine Wong Shi Kam, Robert J. Chen, and Yiming Li, *Functionalization of Carbon Nanotubes for Biocompatibility and Biomolecular Recognition*, Nano Letters, Vol. 0, No. 0, A-D, 2002.
9. S. Joseph, R.J. Mashl, E. Jacobsson, and N.R. Aluru, *Ion Channel Based Biosensors: Ionic Transport in Carbon Nanotubes*, Nanotech 2003 Vol. 1, 158-161.
10. H. Dai, N. Wong Shi Kam, M. O'Connell\*, and J. A. Wisdom, *Carbon nanotubes as multifunctional biological transporters and near-infrared agents for selective cancer cell destruction*, PNAS, 102, 11600-11605, (2005).
11. Bin Zhao, Hui Hu, Aiping Yu, Daniel Perea, and Robert C. Haddon, *Synthesis and Characterization of Water Soluble Single-Walled Carbon Nanotube Graft Copolymers*, J Am Chem Soc. 2005 Jun 8;127(22):8197-203.
12. David E. Luzzi and Brian W. Smith, *Formation Mechanism of Fullerene Peapods and Coaxial Tubes: A Path to Large Scale Synthesis*, Chemical Physics Letters, 2000 April 21: 169-174.

13. J. M. Bonard, T. Stora, J. P. Salvetat, F. Maier, T. Stoeckli, C. Duschl, L. Forro, W. A. de Heer and A. Chatelain, *Purification and size-selection of carbon nanotubes*, Advanced Materials 9, 827 (1997).
14. C.V. Nguyen, L. Delzeit, A.M. Cassell, J. Li, J. Han, and M. Meyyappan, *Preparation of Nucleic Acid Functionalized Carbon Nanotube Arrays*, NanoLetters, Vol. 2, pp. 1079-1081 (2002).
15. Carter Kittrell; Haiqing Peng; Richard E. Smalley; John E. Fischer; W. Eduard Billups; Robert H. Hauge; Wade Adams; Sivarajan Ramesh; Wei Zhou; Lars M. Ericson; Virginia A. Davis; Rajesh K. Saini; Matteo Pasquali; Yuhuang Wang; Juraj Vavro; Csaba Githy; Hua Fan; Joseph Sulpizio; Richard Booker; A. Nicholas G. Parra-Vasquez; Myung Jong Kim; Gerry Lavin; Howard Schmidt; Wen-Fang Hwang, *Macroscopic, Neat, Single-Walled Carbon Nanotube Fibers*, Science, Volume 305, Issue 5689, Pages 1447 - 1450, 2004-09-03
16. Vladislav A. Ryzhkov, *Bulk Production and Applications of Naturally Short Carbon Multi Wall Nanotubes*, Rosseter Holdings Ltd, Cyprus, [slava@e-nanoscience.com](mailto:slava@e-nanoscience.com).
17. Robert Langer, *New Methods of Drug Delivery*, Science, New Series, Vol. 249, No. 4976. (Sep. 28, 1990), 1527-1533.
18. P. M. Ajayan, *Nanotubes From Carbon*, Rensselaer Polytechnic Institute, Chem. Rev. 1999, 99, 1787-1799.
19. Niraj Sinha and John T.-W. Yeow, *Carbon Nanotubes for Biomedical Applications*, IEEE Transactions on Nanobioscience, Vol. 4, No. 2, June 2005, 180-195.
20. Dan Luo, *Nanotechnology and DNA Delivery*, MRS Bulletin, Vol. 30, September 2005, 654-658.
21. Andreas Hirsch, *Functionalization of Single-Walled Carbon Nanotubes*, MiniReview, Angew. Chem. Int. Ed. 2002, Vol. 41, No. 11, 1853-1859.
22. Rajdip Bandyopadhyaya, Einat Nativ-Roth, Oren Regev, and Rachel Yerushalmi-Rozen, *Stabilization of Individual Carbon Nanotubes in Aqueous Solutions*, Nano Letters 2002, Vol. 2, No.1, 25-28.
23. Dimitrios Tasis, Nikos Tagmatarchis, Vasilios Georgakilas, and Maurizio Prato, *Soluble Carbon Nanotubes*, Concepts, Chemistry: A European Journal, 2003, Vol. 9, 4000-4008.
24. James M. Tour and Christopher A. Dyke, *Overcoming the Insolubility of Carbon Nanotubes Through High Degrees of Sidewall Functionalization*, Concepts, Chemistry: A European Journal, 2004, Vol. 10, 812-817.

25. Howard Wang, Wei Zhou, Derek L. Ho, Karen I. Winey, John E. Fischer, Charles J. Glinka, and Erik K. Hobbie, *Dispersing Single-Walled Carbon Nanotubes with Surfactants: A Small Angle Neutron Scattering Study*, Nano Letters 2004, Vol. 4, No. 9, 1789-1793.
26. R. Prakash, S. Washburn, R. Superfine, R. E. Cheney, and M. R. Falvo, *Visualization of individual carbon nanotubes with fluorescence microscopy using conventional fluorophores*, Applied Physics Letter, August 2003, Vol. 83, No. 6, 1219-1221.
27. K. Otobe, H. Nakao, H. Hayashi, F. Nihey, M. Yudasaka, and S. Iijima, *Fluorescence Visualization of Carbon Nanotubes by Modification with Silicon-Based Polymer*, Nano Letters 2002, Vol. 2, No. 10, 1157-1160.
28. Sandeep Kumar, Rajesh Kumar, Ranvinder Singh, Rakesh Kumar, Awdhesh Kumar Shukla, V.K. Jindal, and Lalit M. Bharadwaj, *Binding of Carbon Nanotubes Dispersed by Optical Tweezer on Silicon Surface*, Azojono – Journal of Nanotechnology Online, June 2006, Vol. 2.
29. Z. F. Ren, Z. P. Huang, J. W. Xu, J. H. Wang, P. Bush, M. P. Siegal, and P. N. Provencio, *Synthesis of Large Arrays of Well-Aligned Carbon Nanotubes on Glass*, Science 1998, Vol. 282, 1105-1107.
30. E. Mathiowitz, *Encyclopedia of Controlled Drug Delivery*, Volumes 1-2, 1999 John Wiley & Sons, p. 461-492.
31. K. L. Anderson, E. Manias, R.A. Vaia, and B.L. Farmer, *Dispersion of Single Walled Carbon Nanotubes by Sodium Dodecyl Sulfonate Surfactants in Aqueous Soultion: Molecular Dynamics Simulation*, Materials and Manufacturing Directorate, Dept. of Materials Science and Engineering, The Pennsylvania State University.
32. Overview - “Anatomy of a Patenting Area: Solubilization of Carbon Nanotubes”, Based on the NanoSPRINT Encyclopedia of Carbon Nanotubes, Carbon Nanotubes Monthly, No. 4, Dec. 2005.
33. M. S. Dresselhaus, G. Dresselhaus, and Ph. Avouris, *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*, Springer, New York, Volume 80, 2000, p. 1-29.
34. C.N.R. Rao and A. Govindaraj, *Nanotubes and Nanowires*, RSC Publishing, Cambridge, The Royal Society of Chemistry 2005, p. 1-66. \
35. S.H. Jeong, K.K. Kim, K.H. An, S.Y. Jeon, S.H. Lee, and Y.H. Lee, Determination of Carbon Nanotubes Concentration in Solvent by Using Optical Absorption Spectroscopy, *NT06: Seventh International Conference on the Science and Application of Nanotubes*, Nagano, Japan, June 2006.

36. Y.Y. Huang, S.V. Ahir, and E.M. Terentjev, *Dispersion Rheology of Carbon Nanotubes in a Polymer Matrix*, Physical Review B 73, No. 125422, 2006.

## APPENDIX A:



Data Sheet 565784 Rev. 23-November-04 JSW

### **$\gamma$ -Secretase Inhibitor IX in Solution Cat. No. 565784**

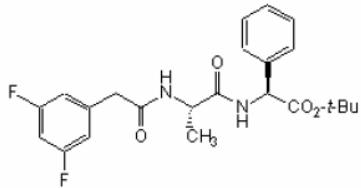
**Description:** A cell-permeable dipeptide that inhibits  $\gamma$ -secretase activity and suppresses A $\beta$  production (A $\beta$  total IC $_{50}$  = 115 nM; A $\beta$ <sub>42</sub> IC $_{50}$  = 200 nM). Reported to be functionally active in both HEK293 cells overexpressing human APP751 and neuronal cultures, however, it does not affect the secretion of amyloid- $\beta$  precursor protein (APP). Effective in acutely lowering Ab in APPV717F-transgenic mice. Also effective in inhibiting the production of an ~ 6 kDa C-terminal cleavage product (CTFg) in HEK293 cells stably transfected with the Swedish mutant,  $\beta$ APP<sub>695</sub>. Reported to reduce extracellular A $\beta$  plaques and intracellular A $\beta$  accumulation in 3xTgAD transgenic mice.

**Form:** Liquid. Supplied as a 25 mM (5 mg/462  $\mu$ l) solution of  $\gamma$ -Secretase Inhibitor IX (Cat. No. 565770) in DMSO. Packaged under inert gas.

**Molecular Weight:** 432.5

**Molecular Formula:** C<sub>23</sub>H<sub>26</sub>F<sub>2</sub>N<sub>2</sub>O<sub>4</sub>

**Structure:**



**Purity:**  $\geq 95\%$  by HPLC

**Storage:** REFRIGERATOR (+4°C). Protect from light. Hygroscopic. Following initial use, aliquot and refrigerate (+4\*deg;C). This product is stable for 2 years as supplied.

**Toxicity:** MSDS available upon request.

**References:**  
Oddo, S., et al. 2004. *Neuron* **43**, 321.  
Dovey, H.F., et al. 2001. *J. Neurochem.* **76**, 173.  
Vandermereren, M., et al. 2001. *Neurosci. Lett.* **315**, 145.

**USA and Canada**  
Tel (800) 628-8470  
[technical@calbiochem.com](mailto:technical@calbiochem.com)

**Germany**  
Tel 0800 100 3496  
[techservice@merckbiosciences.de](mailto:techservice@merckbiosciences.de)

**United Kingdom and Ireland**  
UK Freefone 0800 622935  
Ireland Toll Free 1800 409445  
customer service  
[@merckbiosciences.co.uk](mailto:@merckbiosciences.co.uk)

**All Other Countries**  
Contact Your Local Distributor  
[www.calbiochem.com](http://www.calbiochem.com)  
[technical@calbiochem.com](mailto:technical@calbiochem.com)

A Brand of EMD Biosciences, Inc., an Affiliate of Merck KGaA, Darmstadt, Germany  
[www.calbiochem.com](http://www.calbiochem.com)

FOR RESEARCH USE ONLY. NOT FOR HUMAN OR DIAGNOSTIC USE.

**APPENDIX B:**

# **QUOTATION**

**To:** Alexsandra, alf8@Lehigh.EDU.  
**From:** Zhongping Huang  
**Date:** 12/4/2006  
**Re:** NanoLab Carbon nanotube arrays

Item	Quantity	Description	Price Each	Total Amount
Functionalized low site density of array	1	8x8 mm <sup>2</sup> low site density of carbon nanotube array on 15x15 mm <sup>2</sup> , 300 nm chromium-coated Si (400-500 microns thick). Target nanotube diameter ~100 nm, and target length 2 microns. Nanotube site density 5x10 <sup>6</sup> CNTs/cm <sup>2</sup> .  Seal the low site density of nanotube array with spin-on-glass, sealing height 1 micron.  Treat nanotube tips with HF to remove the spin-on-glass from surface, and then with HNO <sub>3</sub> and H <sub>2</sub> SO <sub>4</sub> acid to functionalize nanotube tips.  After tips functionalized, put the sealed array into HF acid to remove 0.5 micron deep spin-on-glass and remain 0.5 micron thick to support nanotubes in freestanding.  Making the functionalized carbon nanotubes into DI water.	\$2000.00  \$200.00	

Standard lead time is 8-10 weeks after a purchase order is received. The actual lead time could be short than it, depending on concurrent order load. Each array will be accompanied by SEM images verifying length, diameter, etc. Terms: P.O may be faxed to 617 581 6749, Terms Net 30 or credit card. **Shipping and taxes are not included.**

Zhongping Huang, NanoLab, Inc.

# QUOTATION

**To:** Alexsandra, alf8@Lehigh.EDU.  
**From:** Zhongping Huang  
**Date:** 12/4/2006  
**Re:** NanoLab Carbon nanotube arrays

Item	Quantity	Description	Price Each	Total Amount
Function -alized low site density of array	1	15x15 mm <sup>2</sup> low site density of carbon nanotube array on 25x25 mm <sup>2</sup> , 300 nm chromium-coated Si (400-500 microns thick). Target nanotube diameter ~100 nm, and target length 2 microns. Nanotube site density 5x10 <sup>6</sup> CNTs/cm <sup>2</sup> .  Seal the low site density of nanotube array with spin-on-glass, sealing height 1 micron.  Treat nanotube tips with HF to remove the spin-on-glass from surface, and then with HNO <sub>3</sub> and H <sub>2</sub> SO <sub>4</sub> acid to functionalize nanotube tips.  After tips functionalized, put the sealed array into HF acid to remove spin-on-glass.  Making the functionalized carbon nanotubes into DI water.	\$2600.00	\$2600.00
			TOTAL	\$2600.00

Standard lead time is 8-10 weeks after a purchase order is received. The actual lead time could be short than it, depending on concurrent order load. Each array will be accompanied by SEM images verifying length, diameter, etc. Terms: P.O may be faxed to 617 581 6749, Terms Net 30 or credit card. **Shipping and taxes are not included.**

Zhongping Huang, NanoLab, Inc.

# QUOTATION

**To:** Alexsandra, alf8@Lehigh.EDU.  
**From:** Zhongping Huang  
**Date:** 12/4/2006  
**Re:** NanoLab Carbon nanotube arrays

Item	Quantity	Description	Price Each	Total Amount
Function -alized low site density of array	2	15x15 mm <sup>2</sup> low site density of carbon nanotube array on 25x25 mm <sup>2</sup> , 300 nm chromium-coated Si (400-500 microns thick). Target nanotube diameter ~100 nm, and target length 2 microns. Nanotube site density 5x10 <sup>6</sup> CNTs/cm <sup>2</sup> .  Seal the low site density of nanotube array with spin-on-glass, sealing height 1 micron.  Treat nanotube tips with HF to remove the spin-on-glass from surface, and then with HNO <sub>3</sub> and H <sub>2</sub> SO <sub>4</sub> acid to functionalize nanotube tips.  After tips functionalized, put the sealed array into HF acid to remove spin-on-glass.  Making the functionalized carbon nanotubes into DI water.	\$2340.00	\$4680.00
			TOTAL	\$4680.00

Standard lead time is 8-10 weeks after a purchase order is received. The actual lead time could be short than it, depending on concurrent order load. Each array will be accompanied by SEM images verifying length, diameter, etc. Terms: P.O may be faxed to 617 581 6749, Terms Net 30 or credit card. **Shipping and taxes are not included.**

Zhongping Huang, NanoLab, Inc.

## APPENDIX C:

<b>Product Name</b>	<b>DiIC16(3)</b> [1,1'-Dihexadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate]
<b>Size</b>	100 mg
<b>Catalog #</b>	84905
<b>US\$</b>	\$95
<b>Molecular Weight</b>	877.8
<b>Spectral Properties</b>	Abs/Em = 549/565 nm
<b>Solvent System</b>	Ethanol
<b>Description</b>	Lipophilic neuronal tracer
<b>References</b>	Cheng Z, et al. (2004). Differential control over postganglionic neurons in rat cardiac ganglia by NA and DmnX neurons: anatomical evidence. <i>Am J Physiol Regul Integr Comp Physiol</i> 286, R625-33; Koo YE, et al. (2004). Real-time measurements of dissolved oxygen inside live cells by organically modified silicate fluorescent nanosensors. <i>Anal Chem</i> 76, 2498-505; Wu CC, et al. (2004). High-throughput morphometric analysis of individual neurons. <i>Cereb Cortex</i> 14, 543-54; Suksaweang S, et al. (2004). Morphogenesis of chicken liver: identification of localized growth zones and the role of beta-catenin/Wnt in size regulation. <i>Dev Biol</i> 266, 109-22; Iriyama A, et al. (2004). Effects of indocyanine green on retinal ganglion cells. <i>Invest Ophthalmol Vis Sci</i> 45, 943-7; Vest RS, et al. (2004). Divalent cations increase lipid order in erythrocytes and susceptibility to secretory phospholipase A2. <i>Biophys J</i> 86, 2251-60; Zhang X and Kirsch LE (2004). Correlation of the thermal stability of phospholipid-based emulsions and the microviscosity measurements using fluorescence polarization. <i>Pharm Dev Technol</i> 9, 219-27; Hefnawy MA and Abou-Zeid AM (2003). Differential adaptation of membranes of two osmotolerant fungi, <i>Aspergillus chevalieri</i> and <i>Penicillium expansum</i> to high sucrose concentrations. <i>Acta Microbiol Pol</i> 52, 53-64; Jordan MB, et al. (2003). Liposomal clodronate as a novel agent for treating autoimmune hemolytic anemia in a mouse model. <i>Blood</i> 101, 594-601; Bianchi L, et al. (2003). A potassium channel-MiRP complex controls neurosensory function in <i>Caenorhabditis elegans</i> . <i>J Biol Chem</i> 278, 12415-24; deAzevedo LC, et al. (2003). Cortical radial glial cells in human fetuses: depth-correlated transformation into astrocytes. <i>J Neurobiol</i> 55, 288-98; Ramani K and Balasubramanian SV (2003). Fluorescence properties of Laurdan in cochlear phases. <i>Biochim Biophys Acta</i> 1618, 67-78.
<b>Storage</b>	4°C

<b>Product Name</b>	<b>DiOC6(3)</b> [3,3'-Dihexyloxacarbocyanine iodide]
<b>Size</b>	100 mg
<b>Catalog #</b>	84715
<b>US\$</b>	\$49
<b>Molecular Weight</b>	572.5
<b>Spectral Properties</b>	Abs/Em = 484/501 nm
<b>Solvent System</b>	DMSO
<b>Description</b>	Most widely used for measuring membrane potential
<b>References</b>	Cheng Z, et al. (2004). Differential control over postganglionic neurons in rat cardiac ganglia by NA and DmnX neurons: anatomical evidence. <i>Am J Physiol Regul Integr Comp Physiol</i> 286, R625-33; Koo YE, et al. (2004). Real-time measurements of dissolved oxygen inside live cells by organically modified silicate fluorescent nanosensors. <i>Anal Chem</i> 76, 2498-505; Wu CC, et al. (2004). High-throughput morphometric analysis of individual neurons. <i>Cereb Cortex</i> 14, 543-54; Suksaweang S, et al. (2004). Morphogenesis of chicken liver: identification of localized growth zones and the role of beta-catenin/Wnt in size regulation. <i>Dev Biol</i> 266, 109-22; Iriyama A, et al. (2004). Effects of indocyanine green on retinal ganglion cells. <i>Invest Ophthalmol Vis Sci</i> 45, 943-7; Vest RS, et al. (2004). Divalent cations increase lipid order in erythrocytes and susceptibility to secretory phospholipase A2. <i>Biophys J</i> 86, 2251-60; Zhang X and Kirsch LE (2004). Correlation of the thermal stability of phospholipid-based emulsions and the microviscosity measurements using fluorescence polarization. <i>Pharm Dev Technol</i> 9, 219-27; Hefnawy MA and Abou-Zeid AM (2003). Differential adaptation of membranes of two osmotolerant fungi, <i>Aspergillus chevalieri</i> and <i>Penicillium expansum</i> to high sucrose concentrations. <i>Acta Microbiol Pol</i> 52, 53-64; Jordan MB, et al. (2003). Liposomal clodronate as a novel agent for treating autoimmune hemolytic anemia in a mouse model. <i>Blood</i> 101, 594-601; Bianchi L, et al. (2003). A potassium channel-MiRP complex controls neurosensory function in <i>Caenorhabditis elegans</i> . <i>J Biol Chem</i> 278, 12415-24; deAzevedo LC, et al. (2003). Cortical radial glial cells in human fetuses: depth-correlated transformation into astrocytes. <i>J Neurobiol</i> 55, 288-98; Ramani K and Balasubramanian SV (2003). Fluorescence properties of Laurdan in cochleate phases. <i>Biochim Biophys Acta</i> 1618, 67-78.
<b>Storage</b>	4°C

# Oxygen Absorption in Hemoglobin and Brain Activity Measurement

## Quantitation of Oxygen Concentration

NSF Summer Undergraduate Fellowship in Sensor Technologies  
Shakera Guess (Chemistry, Spanish) – Lincoln University  
Advisor: Dr. Britton Chance

### **ABSTRACT**

Many studies today are executed with the goal of inventing a non-invasive method to measure blood oxygen levels in the brain. The current methods include taking blood samples during surgery or using a pulse oximeter. Yet, neither of these methods yields direct or immediate data about brain oxygen levels. The goal of this project was to measure the concentration and saturation of hemoglobin with oxygen and associate this with how Near Infrared Light Spectroscopy (NIRS) can be used in a non-invasive way to measure brain activity. The protocol of this experiment was to use the Clarke oxygen electrode to measure absorption levels of oxygen in yeast. After executing this experiment many times, the same method was to be transferred to hemoglobin, an animal's brain, and finally the human brain. The Cognosensor (brain imager) invented by Dr. Britton Chance, which uses NIRS to correlate changes in oxygen levels to measure brain activity, was used during the hemoglobin experiment. During all of these experiments, levels of oxygen concentration were measured and plotted respectively. The results of these experiments were used to correlate the methods used on the aforementioned brain imager. This information could yield to developing a low-cost, non-invasive method to measuring oxygen concentration and saturation of hemoglobin in the brain.

## Table of Contents

1.	Introduction .....	3
2.	Background .....	3
2.1	Current Methods.....	6
2.2	Further Background.....	7
3.	Project Goals .....	6
4.	Experimental Set-Up.....	8
4.1	Oxygen Absorption in Yeast.....	8
4.2	Measurement of Oxygen Level Concentration Changes in Hemoglobin.....	10
4.3	Measurement of Oxygen Level Concentration Changes in Hemoglobin Using the Clarke Oxygen Electrode and the Cognosensor (Brain Imager).....	12
5.	Discussion and Conclusion.....	13
6.	Recommendations.....	14
7.	Acknowledgements.....	14
8.	References.....	15

## **1. Introduction**

The human brain is one of the few components in the body of which we know relatively very little. This is largely due to the fact that the human brain is very sensitive, and we cannot invasively measure changes in its activity. Therefore, the ability in the present day to accurately predict a particular cognitive function given a pattern in the change of blood volume levels is limited. The current methods for measuring cognitive function are very innovative. Yet, many of these techniques require the use of large or invasive machinery in addition to immobilization of the subject. Also, the equipment costs of many of these techniques further puts their use outside the realm of most environments. [1]

There are methods available today for measuring cognitive function such as functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), and positron emission tomography (PET). Yet, Near Infrared Spectroscopy (NIRS) is non-invasive, and relatively cheap when compared to the aforementioned methods.

Near Infrared spectroscopy (NIRS) is a safe, non-invasive way for measuring hemodynamic response using an absorption band of deoxyhemoglobin (Hb) and oxyhemoglobin ( $\text{HbO}_2$ ) between wavelengths of 700-900 nm. In the 1980s, a single unit, stand-alone NIRS was made available, but the application of Near-Infrared spectroscopy was focused more on chemical analysis. It became a much more powerful tool for scientific research in more fields such as physics, physiology, and medicine in the early nineties, with the introduction of light-fiber optics and monochromator-detector developments.

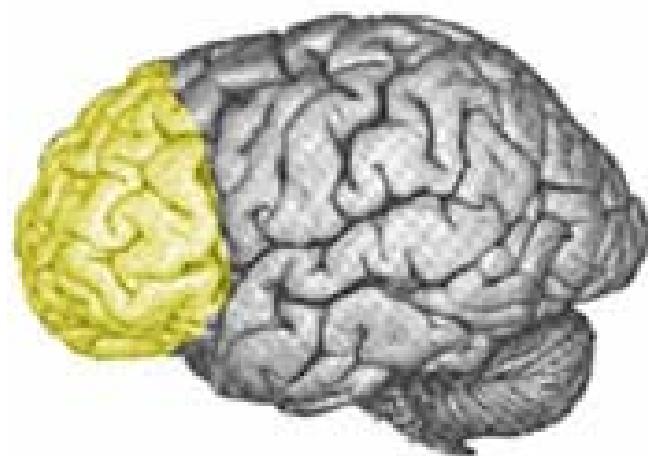
Dr. Britton Chance's laboratory at the University of Pennsylvania School of Medicine has created a brain imager which employs Near Infrared Spectroscopy to measure changes in oxygen levels. Furthermore, Dr. Chance and his team have designed a remote, handheld sensor device which is simple, cheap, and is used to detect changes in the blood volume levels of the brain. [1] This study focuses on NIRS, how it is used to quantitate oxygen concentration levels in hemoglobin, and how hemoglobin serves as a biosensor of brain functional activity.

## **2. Background**

Near Infrared Spectroscopy is a technique used for safely measuring the changes in blood volume during cognitive function for *in vivo* experiments. It is an innovative approach to cognitive function analysis and has recently been picking up in popularity because it is non-invasive, easy to implement, and relatively cheap. It is used in a variety of procedures including tissue density monitoring, cancer detection, oximetry, and brain function detection. For the past ten years, it has been used for *in vivo* measurement for breast cancer detection, tissue vascular oxygenation, and functional brain imaging. [1]

Though discovered in the nineteenth century, Near Infrared Spectroscopy was introduced to the industry in the 1950s. At first, it was used only as an accompaniment to other optical devices already familiar to the industry such as ultraviolet (UV) and visible (Vis) spectrometers. In the 80s a stand-alone NIRS system was made available, but its application was focused on chemical

analysis. At present, Near Infrared Spectroscopy is used as a powerful tool for medical research and in a number of scientific fields including physics, physiology, and as a medical tool for monitoring patients. NIRS is also used in optical tomography (OT) for functional mapping of the human cortex, which is illustrated in (**Figure 1**).



**Figure 1:** The human brain with the pre-frontal cortex highlighted in yellow. [2]

## 2.1 Current Methods

Since 1973, which was the year when the first metabolic image of the human brain was produced, technology has been developed to produce better and more ways to produce images. One such method, positron emission tomography or PET scan (**Figure 3**) acquires physiological images based on the detection of radiation from the emission of positrons (tiny particles emitted from a radioactive substance administered to the patient) injected into the skin.



**Figure 3:** A positron emission tomography (PET) scanner. [3]

Though most commonly used for cancer detection and to examine the effects of cancer therapy by characterizing biochemical changes in the cancer, PET scans can also be used for brain

imaging. PET scans of the brain are used to evaluate patients who have memory disorders of an undetermined cause, suspected or proven brain tumors, or seizure disorders that are not responsive to medical therapy and are therefore candidates for surgery. The PET scanner bears a resemblance to a large doughnut and includes multiple rings of detectors that record emission of energy from radioactive substances in the body. The subject is to remain still (immobilization) during the scanning procedure which is about 45 minutes in length. [3]

Electroencephalography (**Figure 4**) measures the electrical activity of the brain by recording from electrodes most commonly on the scalp but in special cases subdurally or in the cerebral cortex. [4]



**Figure 4:** A patient undergoing EEG.

The results are the electroencephalogram (EEG), which represent an electrical signal from a large number of neurons. It is most useful for classifying epilepsy and syncope, examining sleep disorders, diagnosing comas, and assessing brain death. [4]

Functional Magnetic Resonance Imaging (fMRI) is a technique used to detect subtle increases in blood flow associated with activation of parts of the brain. (**Figure 5**)



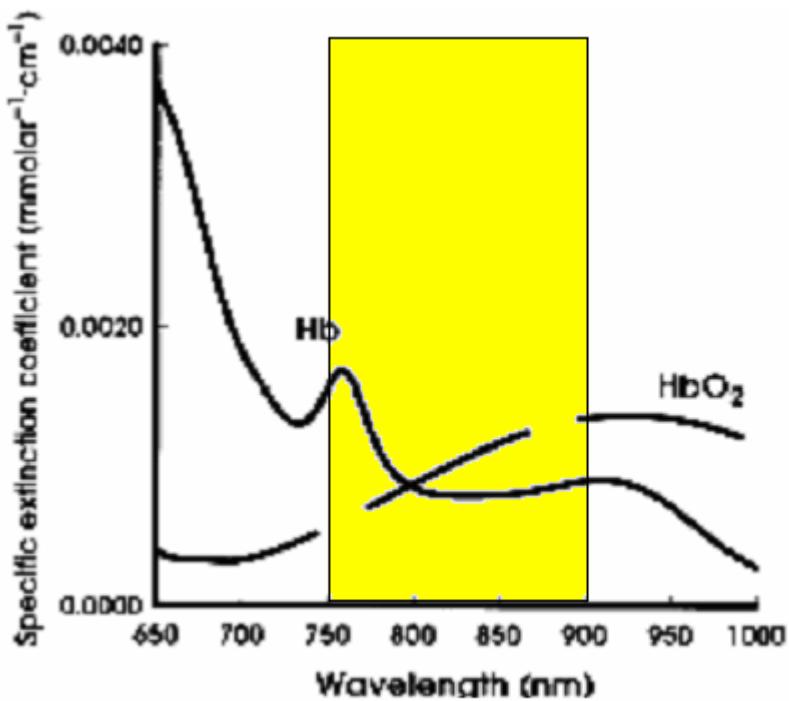
**Figure 5:** Berkeley's 4T fMRI scanner. [5]

It is useful for what is frequently called brain “mapping”. [6] It takes advantage of the fact that activation of an area of the brain causes an increase in blood flow to that area that is greater than that needed to keep up with oxygen demands of tissues. This results in a net increase in intravascular oxyhemoglobin (Hb) and a decrease in deoxyhemoglobin ( $\text{HbO}_2$ ). It uses radio waves and a strong magnetic field rather than x-rays to provide clear and detailed pictures of internal organs and tissues. [6] The procedure causes no pain, but similar to PET imaging, the subject must be immobilized. This instrument is also very expensive. The fMRI machine itself could cost between \$1.5 and \$4 million.

## 2.2 Further Background

It has been suggested that certain cognitive functions induce changes in the blood volume levels of the prefrontal cortex of the brain, and that observing these changes in blood volume levels can be used to predict these same cognitive thought processes. Because of the sensitivity and inability to measure these changes upon stimulation of cognitive function, non-invasive measures are necessary. Biological tissue is transparent to light in the near infrared range between 700-900 nanometers. Therefore, this wavelength range acts as an optical window for the non-invasive assessment of brain tissue. Several NIRS studies in recent years have demonstrated that changes in brain activity can be assessed non-invasively in adult human subjects. [7]

Visually opaque media are omnipresent in nature. While some materials are called opaque because they strongly absorb visible light, others are called opaque because photons traveling within them are largely scattered rather than absorbed. Light is transported through these materials in a process similar to heat diffusion. These so called “diffusing photons” are used in the medical and biophysics communities to view the functions and structure of the body. (**Figure 2**) This is made possible by a spectral window that exists within tissues in the 700-900 nm region, in which photon transport is dominated by scattering rather than absorption. [8]



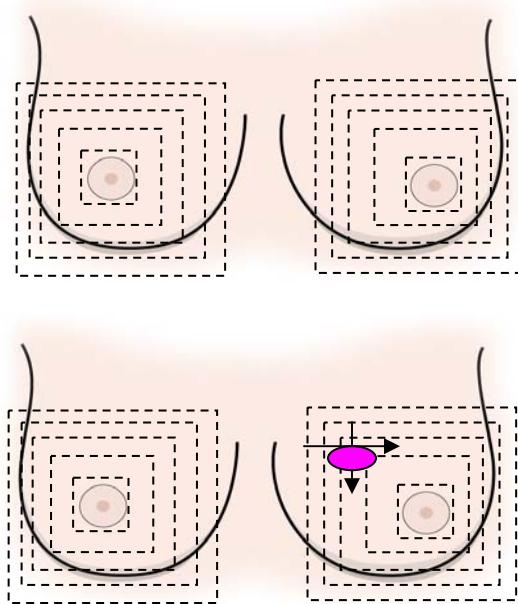
**Figure 2:** The absorption spectrum of oxyhemoglobin ( $\text{HbO}_2$ ) and

deoxyhemoglobin (Hb) in the near-infrared region 65—1000 nanometers. The highlighted yellow region represents the “spectral window” of 750-900 nm that exists within tissues.

### 3. Future Uses

Many ground-breaking studies are being executed to explore the use of NIRS in lie detection. NIRS measures changes in Blood oxygenation due to neural activity in the cortex of the brain. Because neural pathways in the brain mediate both the cognitive components of conscious deception, as well as emotional responses NIRS can be used to differentiate intentional deception as well as guilt.

A future use of near infrared spectroscopy as mentioned earlier may be early breast cancer detection. It has been shown that the mammogram does not work well on women under the age of 40. This is partly because women under 40 have more radiographically dense tissue. In addition, many women with a high risk of getting breast cancer may get it before the age of 40. The handheld sensor device under development by Dr. Chance and his team uses NIR penetrating ~3 cm into breast tissue. Moved around the breast in a circular motion, it detects any lumps in the breast. This would benefit many women including those in underserved populations who are less inclined to obtain a mammogram.



**Figure 6:** Two example images of results from breast cancer detection exam. The top image is of a normal breast, and the bottom demonstrates where the tumor may be located.

## **4. Project Goals**

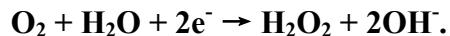
The goal of this project was to use different methods to quantitate oxygen concentration levels in the human brain and correlate these findings near infrared spectroscopy and how it measures these concentration changes in oxyhemoglobin and deoxyhemoglobin. The Cognosensor (brain imager) invented by Dr. Britton Chance, Ph. D, Sc. D, MD was to be used with human blood to represent how Near Infrared Spectroscopy measures oxygen concentration changes in the human brain.

## **5. Experimental Set-Up**

### **5.1 Oxygen Absorption in Yeast**

#### **Introduction**

The Instech Clark Oxygen Electrode is a machine that is used to connect with the non-metallic part of a circuit using the equation



This machine could be used to measure the absorption of oxygen in yeast. Yeast was a good starting material because it was cheap, easy to use, and responded to the electrode as the human brain would if the oxygen supply was to be cut off.

#### **Materials**

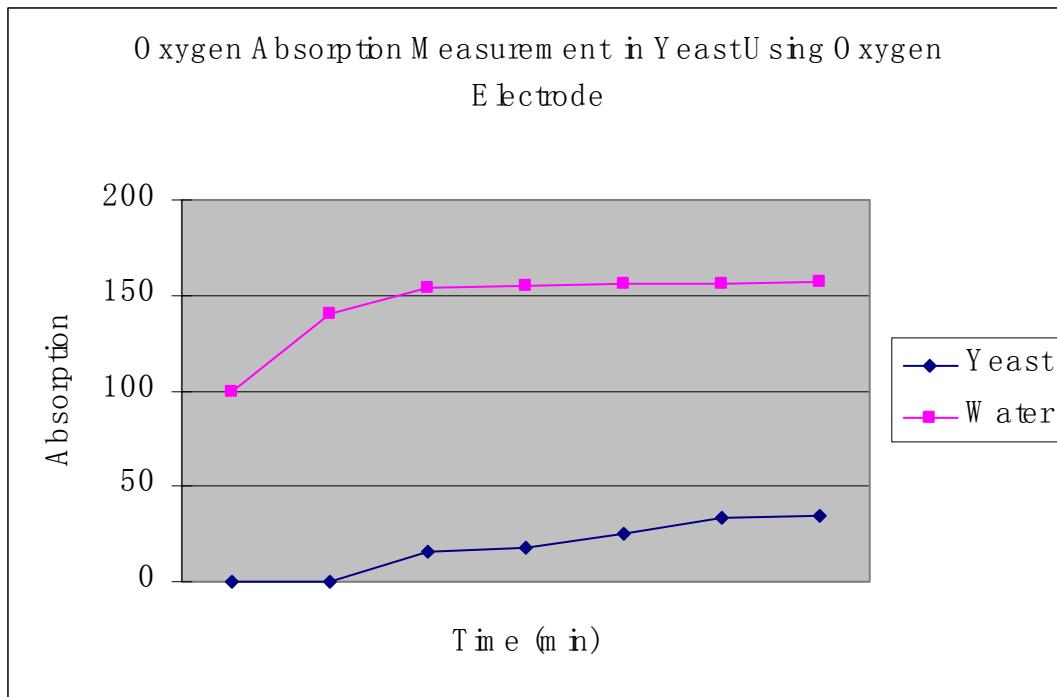
- 1 gram PBS Saline Buffer pH 7.4
- Instech Clarke Oxygen Electrode
- Magnetic Stir Bar
- Magnetic Stirrer
- Salt
- Water
- 7 grams yeast
- 400 ml beaker

## Methods

Saline Buffer, salt, and yeast were added to a 400 ml beaker. Water was added to the 100 ml mark. The magnetic stirrer was used to mix the solution. A separate beaker of water was placed next to the beaker. The electrode was placed in the water and set to 100% absorbency reading. Next, the electrode was placed inside of the yeast solution and the time was taken for the reading to reach 0% absorbency. The electrode was again placed in the water and the time was taken for the electrode to reach a stable reading. This procedure was repeated for a duration of seven minutes.

## Data

The data was recorded and graphed. (**Figure 6**)



**Figure 7:** Oxygen Absorption in Yeast Experiment 1. Absorption plotted against time.

## Conclusions

As water was introduced into the least solution, the percent absorption of oxygen increased.

## Discussion

This experiment was conducted many times throughout the ten-week span. The basic purpose of the experiment was to introduce me to the Instech Oxygen Electrode.



**Figure 8:** Clarke Electrode.

## 5.2 Measurement of Oxygen Concentration Level Changes in Hemoglobin

### Introduction

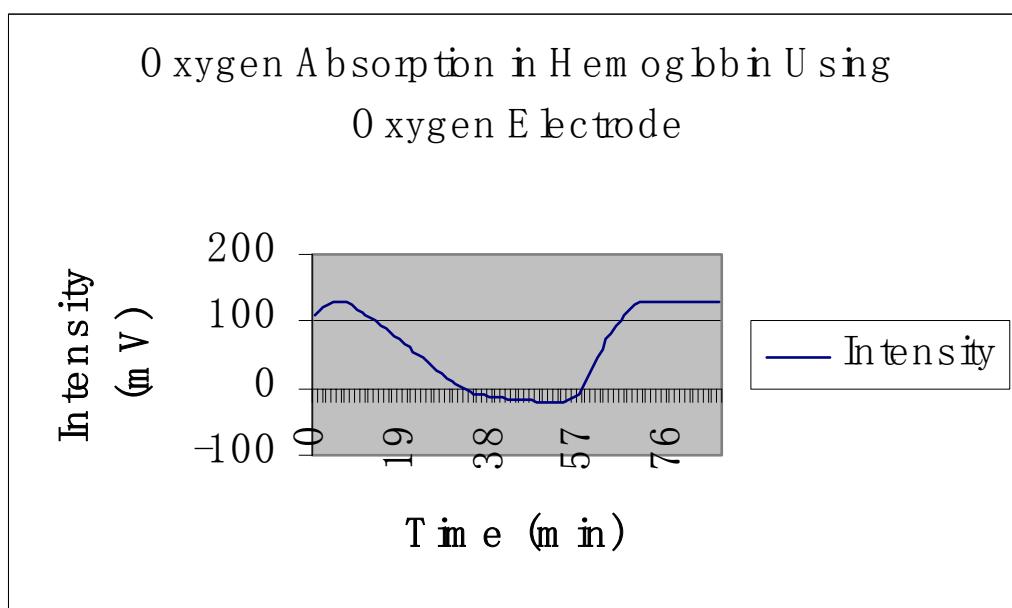
### Materials

- Instech Clarke Oxygen Electrode
- Magnetic Stir Bar
- Magnetic Stirrer
- Water
- Hemoglobin
- PBS Sigma Saline Buffer pH 7.4
- Salt
- Liposyn Intralipid
- Oxygen tank
- Thermometer
- Syringe
- Yeast
- 2000 ml beaker

### Methods

The apparatus was set up. Next, 200 ml water was added to the 2000 ml beaker. One package Saline buffer and one gram salt were added. A syringe was used to add 4 ml hemoglobin. The magnetic stirrer was used to mix the solution, a thermometer was used to ensure the solution remained at optimal temperature between 24-25° C. A separate container was used to mix 1.5 g yeast with a sample of the hemoglobin solution, which was then added to the beaker. A syringe was used to add 2 ml Liposyn Intralipid. The Electrode was connected to the solution. The Oxygen tank was connected to the solution and turned on. This introduces oxygen to the solution, and forms oxygenated (oxy) hemoglobin ( $\text{HbO}_2$ ). The beaker was sealed. An initial reading was taken and subsequent readings were taken every minute. After the a stable reading was taken, the Oxygen tank was turned off and a reading was taken every minute until the Electrode read stable absorption level.

## Data



**Figure 9: Results graph Oxygen Absorption in Hemoglobin Using Oxygen Electrode**

### 5.3 Measurement of Oxygen Concentration Level Changes Using Both the Clark Oxygen Electrode and the Cognosensor (Brain Imager)

#### Introduction

Next, I graduated to using the same technique with the Electrode and adding the Cognosensor to study how Near Infrared Spectroscopy detects changes in blood oxygen levels in the human brain.

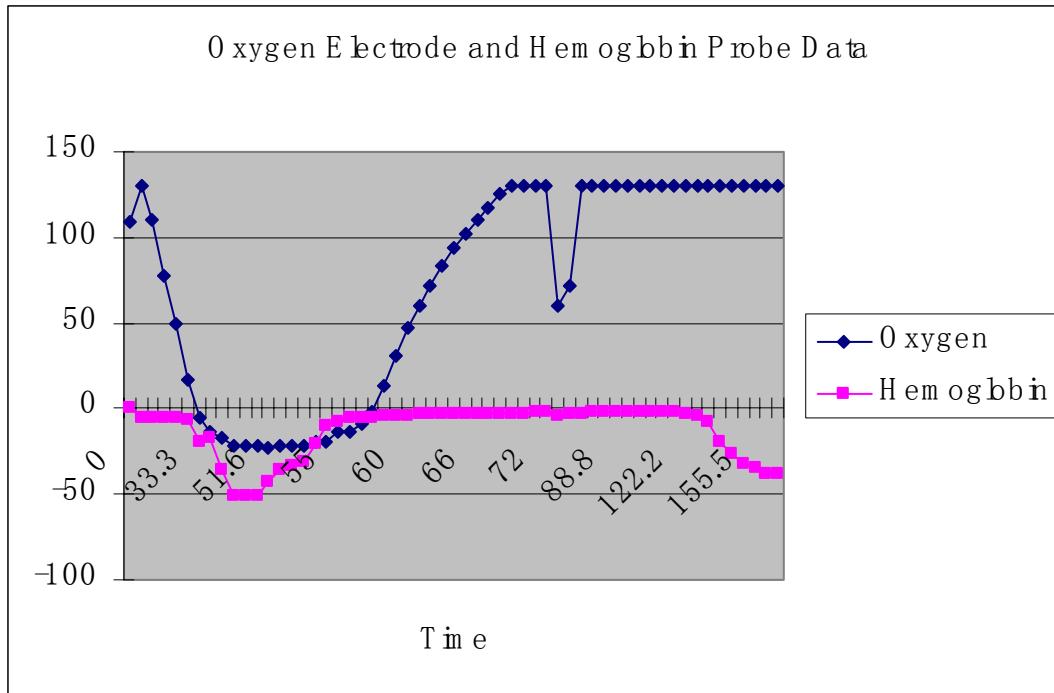
## **Materials**

- Salt
- Water
- Magnetic Stir Bar
- Magnetic Stirrer
- Thermometer
- Liposyn Intralipid
- Hemoglobin
- Duct tape
- Instech Clarke Oxygen Electrode
- Yeast
- PBS Sigma Saline Buffer
- Hemoglobin
- Oxygen Fluorometer
- 2000 ml beaker

## **Methods**

One thousand milliliters water was added to the beaker. Next, 17 cc Liposyn Intralipid, salt, and 1 pack saline buffer were added. A syringe was used to add 20 ml hemoglobin. The magnetic stirrer was used to mix the solution. The thermometer was used to ensure the solution stayed at optimal temperature of 25-30° C. The electrode was placed inside of and the O<sub>2</sub> tank was connected to the solution. The beaker was sealed. Oxygen was turned on, making oxygenated hemoglobin (HbO<sub>2</sub>). Once a stable reading was taken, the tank was turned on and a reading taken every minute until stable.

## **Data**



**Figure 10:** Graphed data from experiment Oxygen Electrode and Hemoglobin Probe Data.

### Conclusion

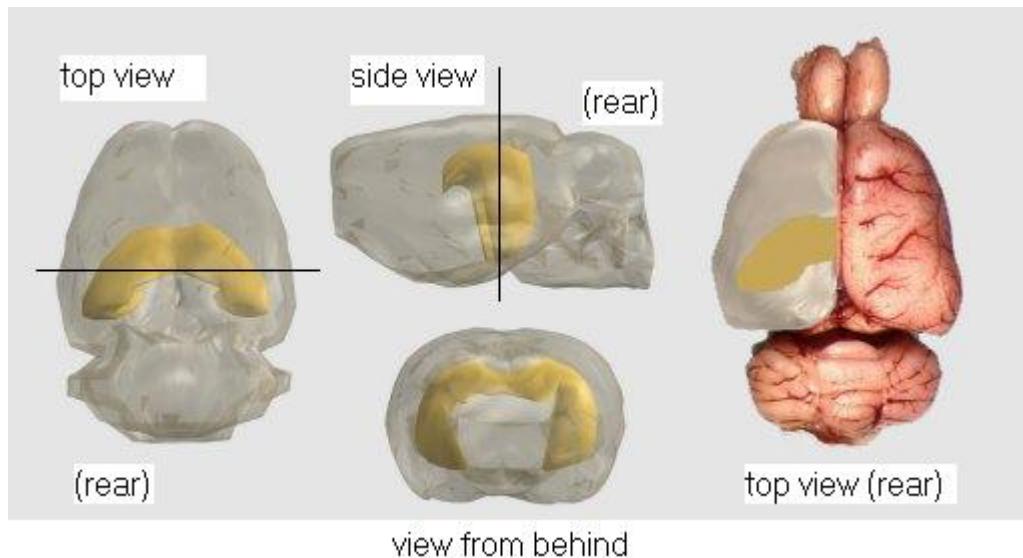
The electrode and brain imager should show increases and decreases in oxygen level concentration level changes at the same time. The results on this graph may be due to systematic error of the electrode.

## 6. Discussion and Conclusions

Throughout the 10 week research period, there were several dilemmas with equipment, predominately the Clarke Oxygen Electrode. At one point, it became necessary to order another machine. This hindered my research greatly.

## 7. Recommendations

The next step would be to employ the same technique used on the experiment and to conduct an experiment using the Cognosensor and a rat brain.



**Figure 8:** View of the entire rat brain

## 8. Acknowledgements

I have learned several lessons throughout this summer, including that a chief principal of research is that it is not always as organized as you want it to be.

I would like to thank the National Science Foundation for providing and funding this opportunity. This was my first research experience and I can truly say that I am not only a better student, but a stronger person because of it. Thank you, all of my fellow SUNFEST students. Because of you this experience has not only been an educational one, but pleasurable as well. I would like to thank Dr. Jan Van der Spiegel for his kindness, genuine concern for the SUNFEST students and their research, and his dedication to the SUNFEST program. I would like to thank all of Dr. Van der Spiegel's administrative assistants. I would like to thank Ms. Delores Magobet for helpfulness with a few "technicalities" that occurred this summer.

Finally, I would like to thank Dr. Britton Chance and his laboratory, especially Juan Du for her supervision and assistance. Dr. Chance has been a blessing this summer.

He was always available for help; he was patient and honestly cared about my SUNFEST experience.

## 8. References

- [1] D. Kaplan, B. Tseng, M. Vergheze, Using Near Infrared Spectroscopy to Detect Brain Function, May 1, 2006
- [2] Vanderbilt University School of Engineering, NSF Grant EIA-0325641, A Biologically Inspired Adaptive Working Memory System for Efficient Robot Control and Learning, [eecs.vanderbilt.edu/CIS/CRL/wm.shtml](http://eecs.vanderbilt.edu/CIS/CRL/wm.shtml)
- [3] Johnson City Medical Center, Positron Emission Tomography (PET) Scan, <http://www.msha.com/body06.cfm?id=14&action=detail&ref=89>
- [4] E. Niedermeyer, F. Lopes de Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott, Williams, and Wilkins, Philadelphia, 2005, p. 139
- [6] S. Ogawa, *et al*, Intrinsic Signal Changes Accompanying Sensory Stimulation: Functional Brain Mapping with Magnetic Resonance Imaging, Proceedings of the National Academy of Sciences of the United States of America, 1992
- [7] G. Boas, Noninvasive Imaging of the Brain, *Optics and Photonics News*, 15 (2004) 52-55
- [8] A. Yodh, B. Chance, Spectroscopy and Imaging with Diffusing Light, *Physics Today*, 48 (1995) 34-40

Electrochemical Methods for the Determination of Redox Properties in Thiol-Terminated  
Porphyrin Self-Assembled Monolayers

Submitted to the Summer Undergraduate Fellowship in Sensor Technologies

Journee B. Isip (Physics) – Columbia University

Advisor: A.T. Charlie Johnson

**ABSTRACT**

A simple potentiostat was adapted from an existing experimental design, with the purpose of emulating both the functionality and the accuracy of commercial potentiostats. The precision and accuracy of this prototype was compared against that of a commercial potentiostat through a series of cyclic voltammograms of sulfuric acid, ferrocene, and porphyrin, at varying scan rates. A reference voltage was ramped, yielding a current which peaked at certain critical redox points, depending on the analyte and the scan rate. A relationship between scan rate and redox peak height was investigated, and an attempt to detect thiol-terminated gold surface-adsorbed porphyrin molecules was made.

## **TABLE OF CONTENTS**

- 1. INTRODUCTION**
- 2. CYCLIC VOLTAMMETRY**
  - 2.1 BASIC CHARACTERISTICS OF CYCLIC VOLTAMMOGRAMS**
  - 2.2 REVERSIBLE REACTIONS**
  - 2.3 REACTION KINETICS AND IRREVERSIBLE REACTIONS**
- 3. POTENTIOSTAT**
- 4. ELECTROCHEMISTRY SETUP**
  - 4.1 ELECTRODES**
  - 4.2 ARGON CHAMBER**
  - 4.3 ELECTROLYTES**
- 5. SULFURIC ACID BLANK**
- 6. FERROCENE**
- 7. PORPHYRIN SELF-ASSEMBLED MONOLAYER**
- 8. DISCUSSION**
- 9. FURTHER RECOMMENDATIONS**
- 10. CONCLUSIONS**
- 11. ACKNOWLEDGEMENTS**
- 12. REFERENCES**

## **1. INTRODUCTION**

Molecular electronics, or the use of single molecules in electrical components, is a growing field due to its small size scale and its promise for very high speeds<sup>3</sup>. Although no one has created a reliable molecular electronic component, many advances have been made. For example, nanometer-sized junction gaps between two gold conduction leads have been fabricated in a reproducible and relatively convenient manner<sup>2</sup>. Single molecules can then be bonded on either end to the two gold leads to allow for electron transport, in the hopes of eventually producing a single-molecular transistor. However, molecules in the break junction bonded to the gold leads exhibit different electrochemical properties than free molecules in solution<sup>4</sup>.

In investigating these properties, self-assembled monolayers (SAMs) of these molecules have surface-adsorbed to gold and cyclic voltammetry measurements have been conducted<sup>4</sup>. However, different conductance measurements on SAMs do not yield consistent data.

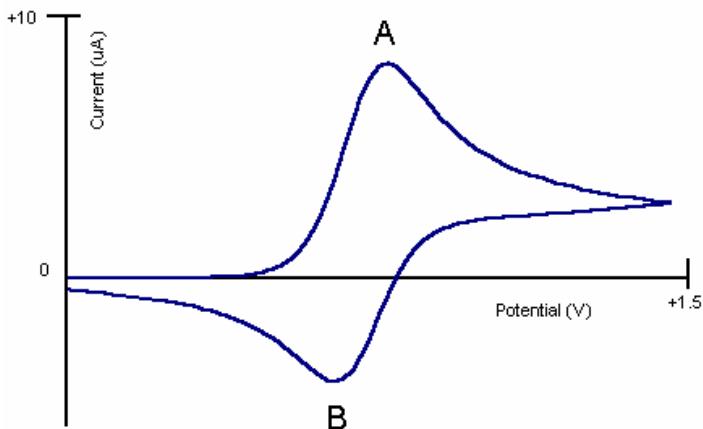
This paper examines the oxidation-reduction properties of thiolated porphyrin molecules bonded to gold working electrodes using a standard three-probe electrochemistry system. Emphasis is placed on the versatility and convenience of the device used to regulate current/voltage output. A simple potentiostat was constructed for this purpose<sup>1</sup>. By varying the scanning voltage sweep rates, a characteristic relationship between the sweep rate and peak heights was determined in order to test the functionality of the potentiostat. In addition, results were compared to those of a commercial potentiostat and found to be very similar.

## **2. CYCLIC VOLTAMMETRY**

### **2.1 BASIC CHARACTERISTICS OF CYCLIC VOLTAMMOGRAMS**

Cyclic voltammetry is a very powerful tool in determining properties of oxidation-reduction reactions in electrolyte solution. The analyte, or species to be studied, undergoes both reduction and oxidation. First, the voltage drop between the working and reference electrodes is set at a point where no reaction can occur (typically 0 V). This voltage is then ramped linearly with time. As the potential rises, reduction of the analyte begins and the observed current varies exponentially with voltage.

This method exploits the fact that the reaction will only occur at the working electrode/solution interface. As the analyte reacts in this interface, the concentration of its oxidized state is decreased which results in a drop of the observed current. This yields a distinct current peak (point A, fig. 1). As the voltage increases further, the reaction is limited by diffusion effects, and the current continues to drop until the voltage ramp is reversed. A similar process will occur for the reoxidation of the analyte (point B, fig 1). This method produces a curve of current as a function of voltage, the peaks of which indicate a maximum rate of charge flow, and therefore maximum reaction speeds.



**Fig. 1.** Idealized cyclic voltammogram depicting the reduction peak (A) and oxidation peak (B).

## 2.2 REVERSIBLE REACTIONS

Reaction peaks in a cyclic voltammogram provide telling distinctions between reversible and irreversible reactions. In a reversible reaction, the respective concentrations of the oxidized and reduced states of the analyte at the working electrode are constantly maintained at the equilibrium defined by the Nernst equation:

$$V = V^0 - \frac{RT}{nF} \ln \frac{[\text{red}]}{[\text{ox}]} \quad (1)$$

At Nernstian equilibrium, eqn. 1 dictates the concentrations of the reduced and oxidized states given the total applied voltage between the working and reference electrodes ( $V$ ) and preset constants and parameters ( $R$ ,  $F$ ,  $V^0$ ,  $n$ ,  $T$ ). In a reversible process, the magnitudes of the peak heights are the same. Additionally, the rate at which the voltage is ramped (V/s) should not affect the positions of the peaks, which will always be separated by the following relation, where  $n$  denotes the number of electrons per molecule transferred in a redox reaction:

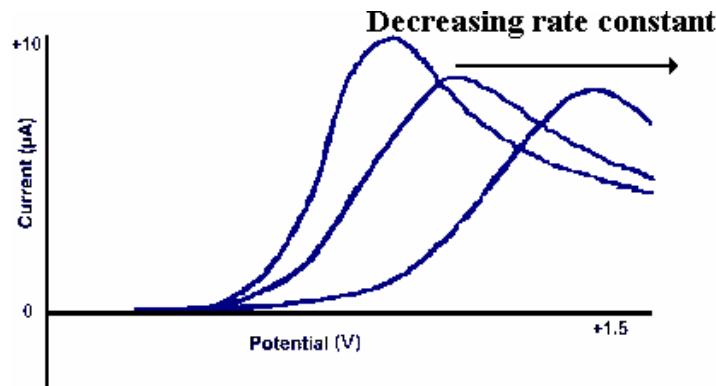
$$\Delta V = \frac{58}{n} \text{ mV}, \quad (2)$$

The peak heights, however, vary with the square root of the scan rate. This can be qualitatively explained if diffusion relations are considered. An increase in the voltage scan rate will yield a steeper diffusion gradient between the electrode interface and the bulk solution. Therefore, the rate of diffusion of the analyte to the electrode interface will be quicker than in slower scans, thus producing a more pronounced current peak.

## 2.3 REACTION KINETICS AND IRREVERSIBLE REACTIONS

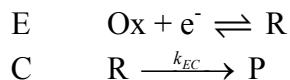
By contrast, irreversible processes appear quite differently on cyclic voltammograms. In such cases the electron transfer may be governed by a reaction constant,  $k_{\text{red}}$ , which determines the favorability of a reaction to proceed in either direction. As  $k_{\text{red}}$  decreases, the reaction takes

longer to proceed at any given scan rate, resulting in a delayed peak relative to the reversible process (fig. 2).



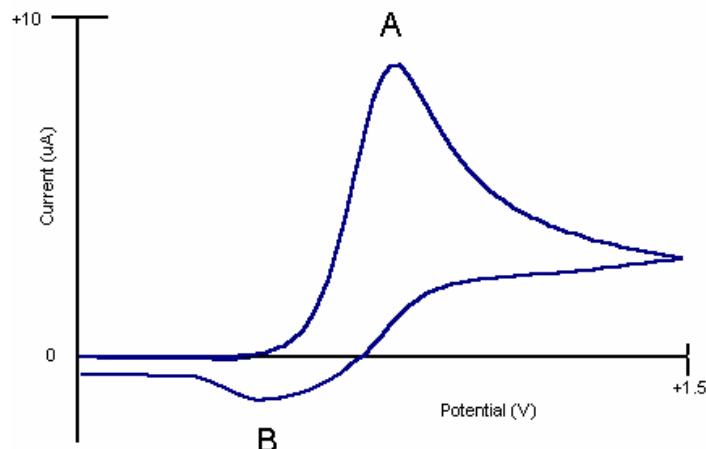
**Fig. 2. Idealized irreversible reaction diagram.** As the rate constant,  $k_{\text{red}}$  decreases, a higher voltage must be maintained to achieve the reduction peak.

In addition, an irreversible process may be defined by reactions which do not take place at the electrode (fig. 3). These are known as chemical, or C reactions.



**Fig. 3. Example of an EC mechanism.** The E (electrode) reaction is fully reversible, but the C (chemical) reaction is not, which disrupts peak positions and heights on cyclic voltammograms.

Whereas a reversible process operates under the E (electrode) reaction, an irreversible process is governed by the E as well as one or more C reactions. For the mechanism in fig. 3, the reduced state of the analyte can react irreversibly to form a compound P. The propensity of the C reaction to occur is dictated by the reaction constant  $k_{EC}$ . This is depicted as a decrease in the reoxidation peak (B) relative to the reduction peak (A) in an irreversible cyclic voltammogram (fig. 4).



**Fig. 4. Idealized irreversible cyclic voltammogram.** As the analyte is reversibly reduced (A), it irreversibly reacts as per the C reaction, subsequently decreasing the reoxidation peak (B).

### 3. POTENTIOSTAT

The potentiostat (fig. 5) is composed of three probes: the working, counter, and reference electrodes. A constant voltage  $V_{\text{ref}}$  is maintained between the working and reference electrodes. The working electrode is kept at ground, while the reference electrode is kept stably at  $V_{\text{ref}}$ . The counter electrode injects the amount of current necessary to sustain the constant  $V_{\text{ref}}$  voltage. In addition, a current-voltage converter allows  $V_{\text{out}}$  to be sent to a computer to be read by LabView.

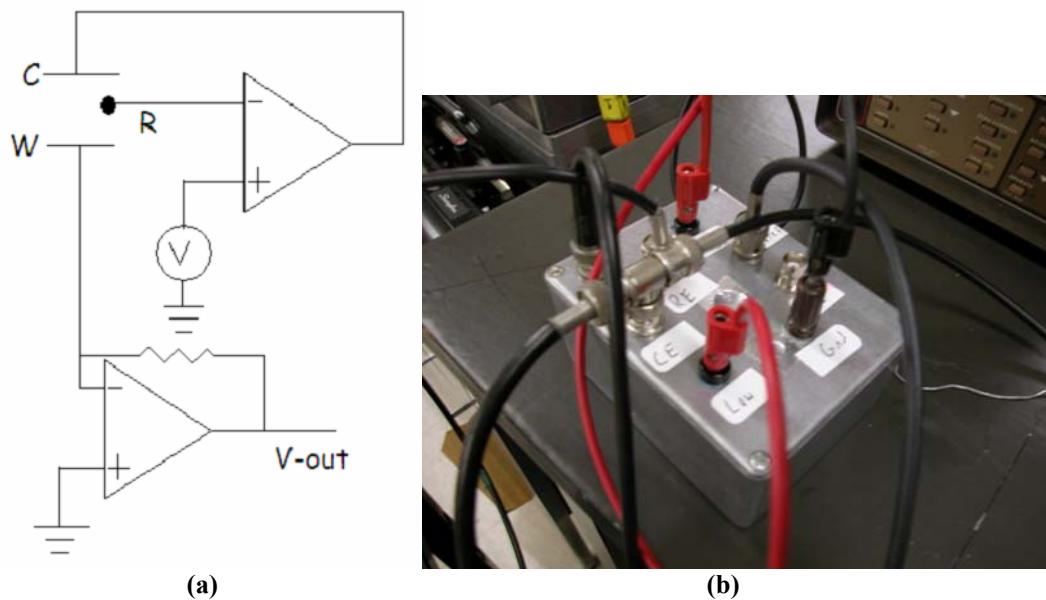


Fig. 5. (a) Schematic of potentiostat used for cyclic voltammetry. (b) Actual prototype potentiostat device, connecting to electrochemical cell and breakout box.

Some advantages of this potentiostat compared to commercial potentiostats (fig. 6) are that it is both inexpensive and versatile. The simple op-amp circuit in fig. 5 (a) can be constructed very easily and at almost no cost. Also, because the circuit is so simple, the consequences of any modifications made to it can be fully known, making it very easy to integrate with other studies. Though the commercial potentiostat in fig. 6 can be used for a variety of other measurements, the potentiostat constructed above has been implemented in the cyclic voltammetry for the foregoing reasons.

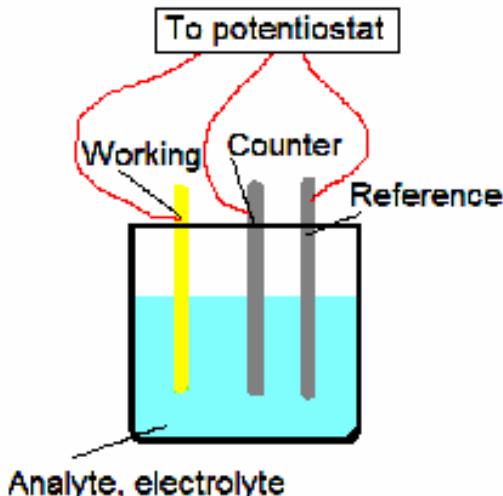


Fig. 6. PAR potentiostat/galvanostat model 363.

### 4. ELECTROCHEMISTRY SETUP

## 4.1 ELECTRODES

The electrochemical cell consists of three electrodes (fig. 7). We have elected to use a gold working electrode because the principal aim of this project is to provide insight into the electron transfer of molecules self-adsorbed to gold. Since it is a very good conductor, gold is also a common working electrode in solution-interface reactions as well, so the same electrode can be conveniently used to run multiple cyclic voltammograms.



**Fig. 7. Schematic of electrochemical cell, comprised of gold working electrode, silver counter electrode, silver reference electrode, conducting electrolyte solution, and analyte.**

For no electrochemical cell is it feasible to measure the potential drop across a single electrode interface. To resolve this difficulty, a reference electrode is used; the potential drop is observed across both the working and reference electrode interfaces and the recorded potential drop is observed as the difference between the voltages across both surfaces. The reference electrode must be stable in order to maintain a precise voltage drop  $V_{\text{ref}}$  between the working and reference. In this study we have used silver wire as the reference electrode.

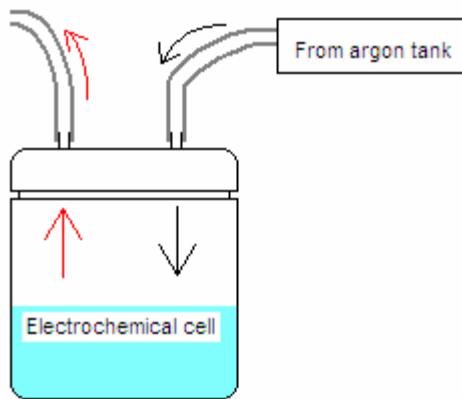
In many setups, only the reference and the working electrodes would be needed to perform electrochemical measurements. These cases are *potentiostatic*, with no applied voltage being delivered to the system. However, when an external voltage is applied for cyclic voltammetry a counter electrode is necessary. Silver was used as the counter electrode in this experiment. Typically in electrochemical experiments it is necessary to isolate the counter electrode from the working electrode, as the redox reaction sustained at the counter electrode can interfere with analyte measurements. However, since the time scale for conduction of a cyclic voltammogram is sufficiently short, any reaction at the counter electrode interface is rendered negligible.

## 4.2 ARGON CHAMBER

Oxygen can undergo a variety of reactions which can be measured in cyclic voltammograms. To minimize the effect of oxygen, it was flushed out of the cell with an argon gas line. For CVs

involving aqueous electrolytes, argon was continually pushed into and out of the cell throughout the duration of the measurements. Due to the volatility of organic solvents, however, the CVs involving ferrocene required closed systems, thus precluding the possibility of a continuous gas flow.

In this setup, air flow could be readily stopped by constriction of the gas lines leading into and out of the chamber (fig. 8). The outlet was also connected to tubing in order to minimize the backflow of oxygen into the chamber.



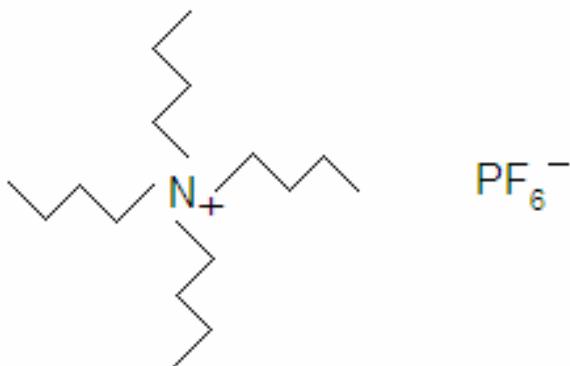
**Fig. 8. Schematic of electrochemical cell with argon line. Black arrows indicate flow of argon from tank into cell. Red arrows show flushed gas exiting cell.**

#### 4.3 ELECTROLYTES

Two different conducting electrolyte solutions were used in this study. In particular, sulfuric acid ( $\text{H}_2\text{SO}_4$ ) and Tetrabutylammonium hexafluorophosphate ( $\text{TBAPF}_6$ ) were used to test the measurement accuracy of the potentiostat.

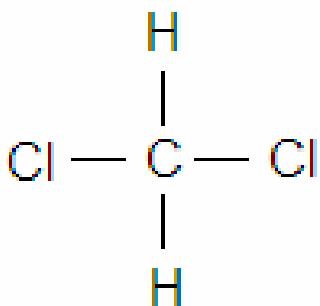
$\text{H}_2\text{SO}_4$  is a strong acid, dissociating to form  $\text{H}^+$  and  $\text{HSO}_4^-$ , making it a very suitable conductor in solution. However, because of its corrosive properties, care was taken not to make a high concentration. As in previous studies<sup>5</sup>, we have used a 50 mM ( $\text{pH} \sim 1.3$ ) as the electrolyte for the blank sulfuric acid test as well as the SAM test.

$\text{TBAPF}_6$  is a common electrolyte solution used for the detection of organic compounds (fig. 9). A much higher concentration of  $\text{TBAPF}_6$  was used relative to  $\text{H}_2\text{SO}_4$  because (1) it lacks the corrosive properties of sulfuric acid, and (2) a more conducting electrolyte solution will allow for easier detection of current in the CV diagram. Since the cyclic voltammogram involving this solvent serves only to compare the commercial potentiostat with a simpler prototype, it is beneficial to achieve an appreciable current output.



**Fig. 9.** Tetrabutylammonium hexafluorophosphate in dissociate form.

This molecule is very stable, so cyclic voltammograms can be made repeatedly with the same electrolyte solution. TBAPF<sub>6</sub> appears as a white crystalline substance and dissolves readily in methylene chloride, an organic solvent (fig. 10). In its dissolved form, TBAPF<sub>6</sub> forms an ionic solution. Methylene chloride (CH<sub>2</sub>Cl<sub>2</sub>) is a colorless liquid which is slightly denser than water (~1.3 g/mL). Because CH<sub>2</sub>Cl<sub>2</sub> is both highly volatile and potentially toxic, care was taken to ensure minimum exposure to the atmosphere during the experiment.



**Fig. 10.** Diagram of methylene chloride (CH<sub>2</sub>Cl<sub>2</sub>) solvent, also known as dichloromethane.

## 5. SULFURIC ACID BLANK

Various cyclic voltammograms were performed using a 0.1 M H<sub>2</sub>SO<sub>4</sub> electrolyte, in both the prototype and the commercial potentiostat. The argon chamber and electrochemical cell was implemented as stated in section 4. However, a minor adjustment was made to the potentiostat setup. While the prototype potentiostat (fig. 5) connected to the counter and reference electrodes, a picoammeter provided the proper connection to the working electrode. Therefore, the current-voltage converter portion (fig. 5a, bottom) from the prototype was not used. In later studies this will be incorporated with the rest of the studies.

The plot below (fig. 11) shows the cyclic voltammogram measurements taken for a sulfuric acid blank at four different speeds with both potentiostats. Fig. 11 plots the height of the bottom peak vs. scan rate for both potentiostats.

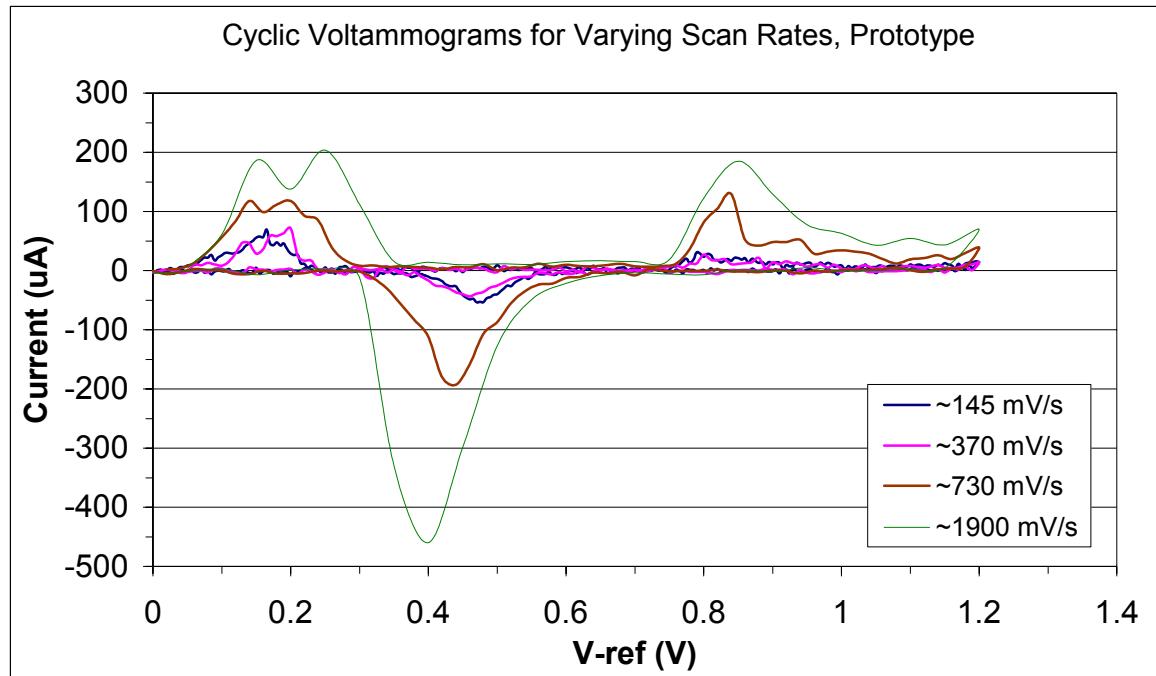


Fig.

11. Cyclic voltammograms for prototype potentiostat at different approximate scan rates: ~145 mV/s (blue), ~370 mV/s (pink), ~730 mV/s (brown), ~1900 mV/s (green).

The general pattern of this cyclic voltammogram can be characterized by the increasing amplitudes of the peaks with respect to the scan rate in mV/s. Fig. 12 shows the peak heights for the cyclic voltammograms in the commercial potentiostat by Princeton Applied Research.

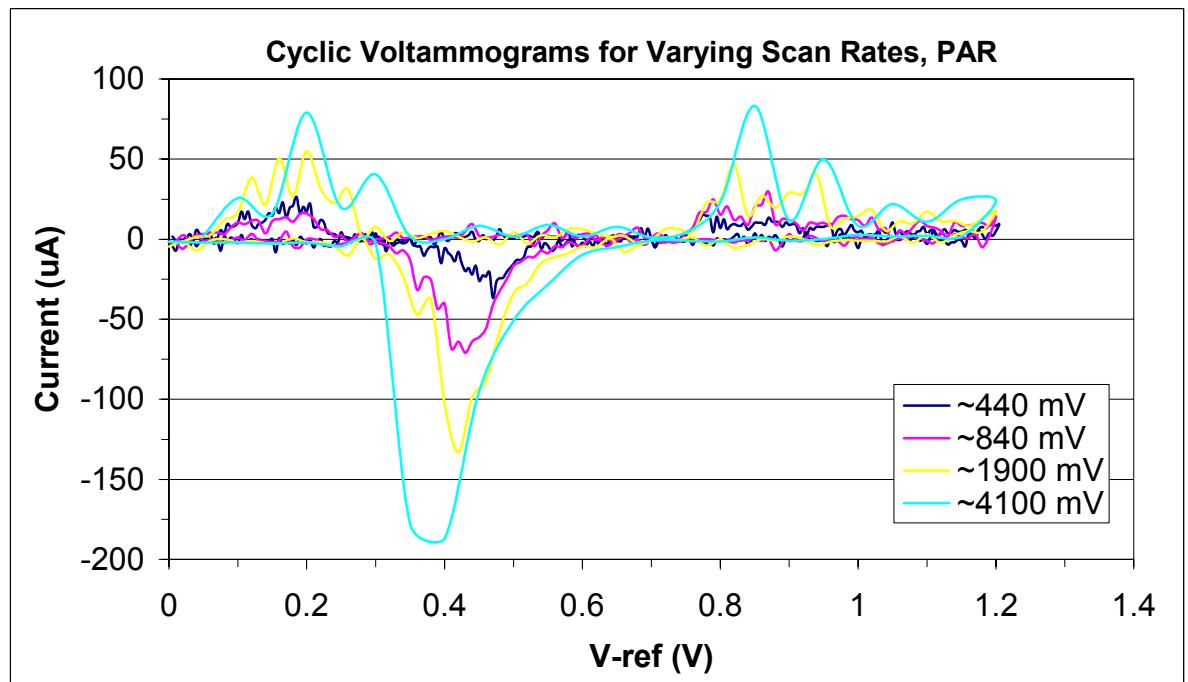


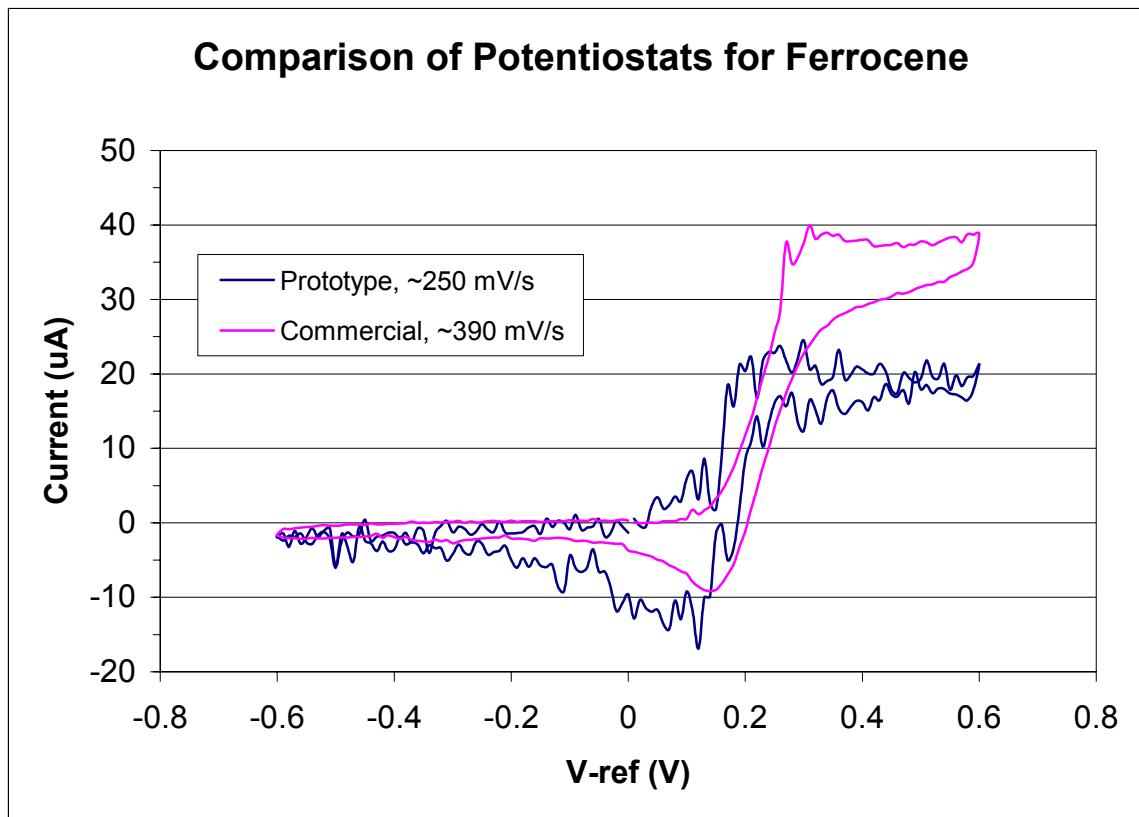
Fig.

12. Cyclic voltammograms for commercial prototype potentiostat at different approximate scan rates: ~440 mV/s (blue), ~840 mV/s (pink), ~1900 mV/s (yellow), ~4100 mV/s (cyan).

The above plots depict that the cyclic voltammogram peaks occur at the same critical voltages (~0.2 V, ~0.4 V, ~0.9 V) in both the prototype and the commercial potentiostats. However, at roughly corresponding voltage rates, the commercial potentiostat CV plot peaks are less pronounced than the prototype potentiostat peaks, by a scale factor. Additionally, more noise is present in the voltammograms performed with the commercial potentiostat, though this can be attributed to the lower scale of the current output. Though there is a lack of direct correspondence in scan rates, there is a direct correspondence in voltage step sizes (e.g. data points were taken in 50 mV steps in both the ~145 mV prototype scan and the ~440 mV commercial scan). The disparity in scan rates is due to the speed at which the data was taken with both potentiostats.

## 6. FERROCENE

A test using a ferrocene electrolyte was conducted using a method similar to the test used in the sulfuric acid blank test. Like with the sulfuric acid blank, the picoammeter provided the current which the potentiostat would have produced otherwise. Below, both plots using the prototype and the commercial potentiostat indicate the current to reference potential, at various scanning rates.



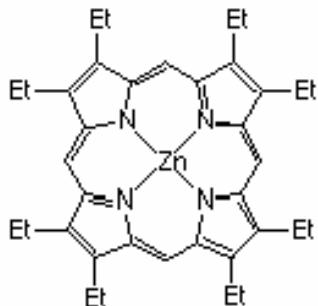
**Fig. 13.** Cyclic voltammogram comparison of two potentiostats: prototype (blue) at ~250 mV/s vs. commercial (pink) at ~390 mV/s.

The above plot shows the cyclic voltammograms of ferrocene peaks. Again, the scan rates are substantially different (~250 mV/s for the prototype and ~390 mV/s for the commercial potentiostat). However, their voltage *steps* are identical; both measurements were taken at a rate of 10 mV/step. The different resultant scan rates can be attributed to delays in the instruments and software that were used. Possible further studies would try to examine the two peaks at identical scan rates. However, valuable information can still be extracted from this plot. In particular, it is encouraging to note that the peaks achieved by the prototype are in approximately the same location as those achieved by the commercial potentiostat (0.18 V, 0.25 V). This indicates that the voltages produced by the prototype are correct; this test can be helpful for calibration purposes.

## 7. PORPHYRIN

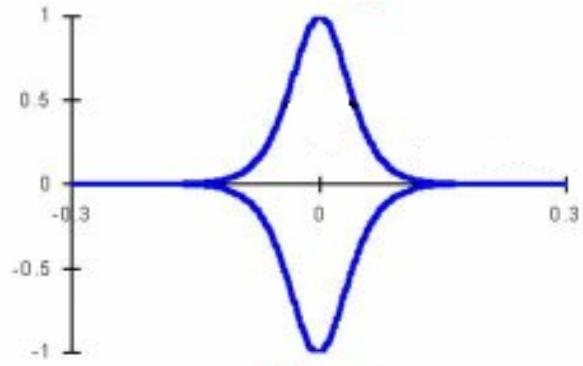
A cyclic voltammogram was conducted for the detection of peaks of porphyrin self-assembled monolayers using a TBAPF<sub>6</sub> electrolyte in a methylene chloride solvent. The gold working electrode was immersed in a solution of porphyrin overnight and a CV was taken in an otherwise analyte-free solution.

The porphyrin molecule (fig. 14) is thiol-terminated at one end. This serves two purposes: (1) thiol groups adsorb relatively well to gold surfaces compared to other organic groups, such as nitrate, and (2) since thiol groups spontaneously bond to each other to create chains of molecules, thiolating porphyrin in a single end will ensure that all of the porphyrin adsorbed to the gold will be a single molecule, and not a chain.



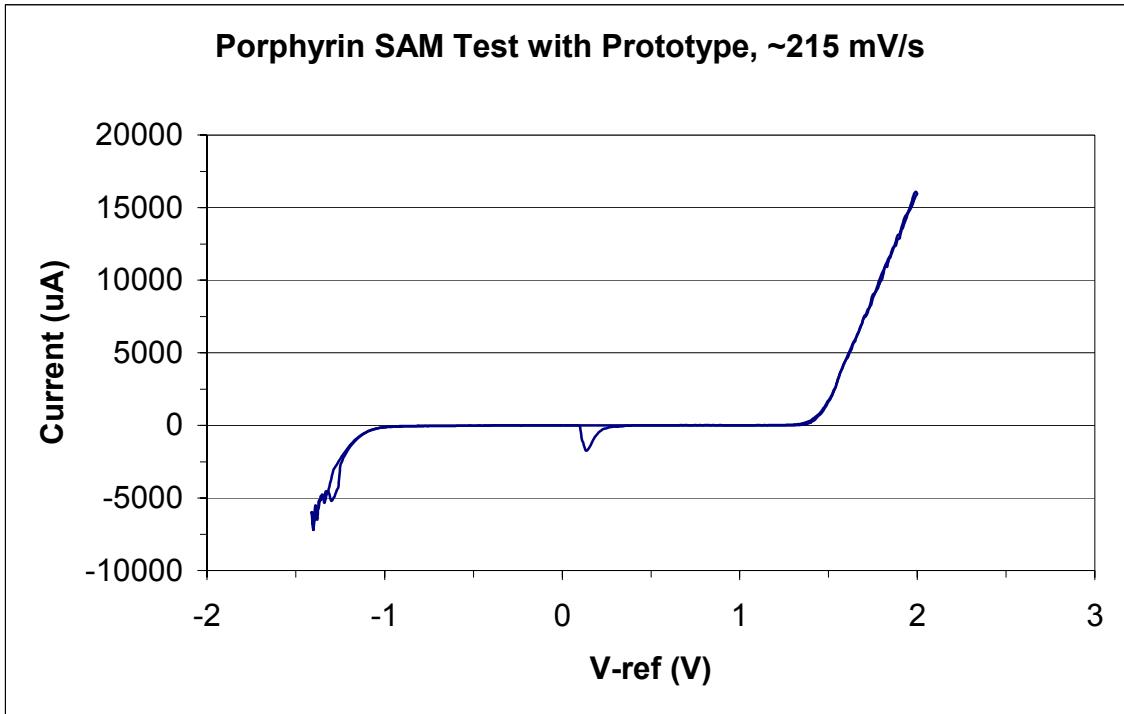
**Fig. 14. Porphyrin molecule with a zinc core.**

Ideally, a surface-adsorbed monolayer would have perfectly symmetrical peaks corresponding to the instantaneous oxidation and reduction of the molecule. Since diffusion is not a factor in the redox reactions, the peaks should readily occur at precisely the same voltage (fig. 15).



**Fig. 15.** Idealized cyclic voltammogram for surface-adsorbed monolayer.

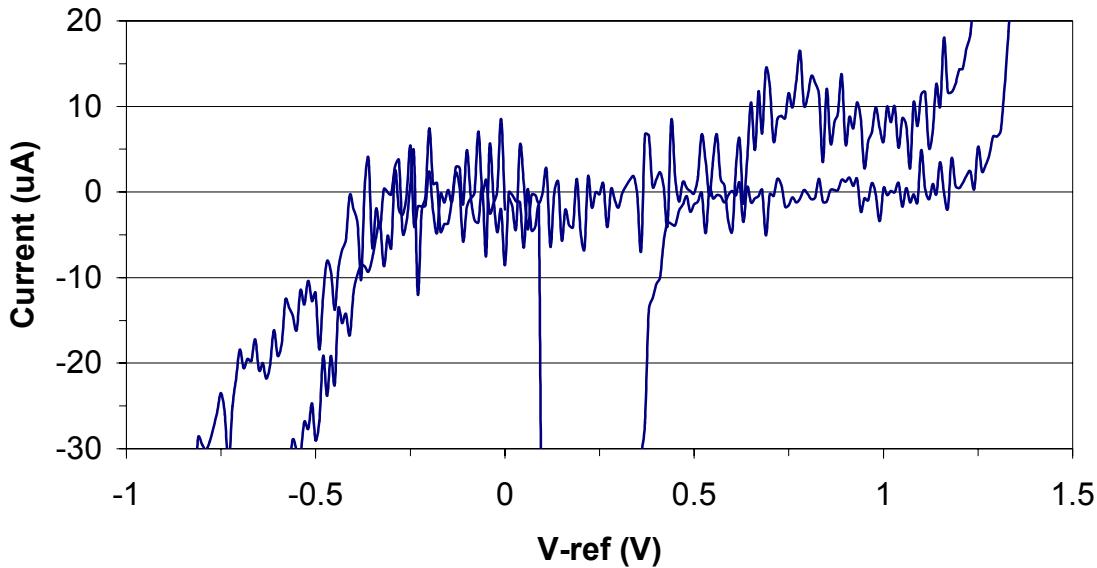
However, similar results were not obtained using the prototype potentiostat (figs. 16, 17). The plots below display the same cyclic voltammogram in two different scales, in consideration of the large current obtained at higher voltages.



**Fig. 16.** Porphyrin SAM detection test, scaled to show entire plot.

Though a peak was observed at ~0.3 V, the magnitude of its height (~2500 uA) is so large given the scan rate that this cannot be attributed to the very small concentration of thiolated porphyrin molecules adsorbed to the gold surface. Additionally, there is no symmetric peak of the same magnitude, but at the same location. The plot below (fig. 17) rescales the CV in order to examine a possible peak of a different order of magnitude.

### **Porphyrin SAM Test with Prototype, ~215 mV/s, Rescaled**



**Fig. 17. Porphyrin SAM detection test, scaled to neglect large current at higher voltages.**

## **8. DISCUSSION**

The sulfuric acid blank test allowed for the observation of increased peak heights with respect to scanning rates. In this way we were able to show that the prototype and the commercial potentiostats had similar functionalities. Both plots show three distinct peaks, the heights of which vary with the scan rates. No reverse peaks were observed from the cyclic voltammograms, reaffirming that the sulfuric acid blank test exhibited peaks due to an irreversible reaction.

The ferrocene cyclic voltammograms also yielded similarities between the two potentiostats. The peaks occurred in the same locations on the two superimposed plots, but the magnitudes of their heights were different. Additionally, a significant level of noise was observed in the prototype potentiostat, possibly attributed to its simple design.

The thiol-terminated gold surface-bonded porphyrin cyclic voltammogram could not confirm the adsorption of the analyte onto the gold working electrode. This could be due to the scan rate: in other tests, self-assembled monolayers were detected with scan rates which were orders of magnitude higher than our initial attempt. Also, a relatively low concentration of porphyrin was used, and the gold working electrode was only immersed in the solution for 24 hours. All such factors can contribute to the lack of detection of the molecule by the prototype potentiostat. A cyclic voltammogram using the commercial potentiostat was not taken.

## **9. FURTHER RECOMMENDATIONS**

Since it is possible that the scan rate and the working electrode contributed to our inability to confirm a self-assembled monolayer of porphyrin, it is recommended that the investigation be continued with a much faster scan rate and a longer immersion time for the electrode. To further increase the chance of detection, the working electrode should have a larger surface area to allow for more porphyrin adsorption.

Additional experimentation warrants the use of new, identical electrodes after each substance is tested, in order to prevent contamination and error from a worn electrode.

The argon chamber flushed out the oxygen from a system with a non-volatile solvent. For an organic solvent such as methylene chloride, the argon flushing caused the solvent to evaporate, possibly skewing the ferrocene data. However, since it is important to have a de-oxygenated environment in the experiment, perhaps the design of the argon chamber should be reconsidered.

The prototype potentiostat can be used to examine the effects of different resistances and different operational amplifiers, since both are detachable in the design of the circuit. Running tests over a range of op-amps and resistors will allow for optimal versatility and flexibility in taking measurements.

With more precise instrumentation, it is important to confirm the relationship between scan rate and peak redox peak heights. Literature states that peak height varies with the square root of scan rate; this is a relatively simple relationship to measure. Running cyclic voltammograms with the prototype potentiostat offers an easy method of checking the precision of the device.

## 10. CONCLUSIONS

A prototype potentiostat (voltage clamp) was constructed using a simple design to incorporate versatility, flexibility, and price convenience. The prototype was compared against a commercial potentiostat for experiments in both precision and accuracy. The cyclic voltammograms confirmed that the prototype exhibited peak heights which varied with the scan rate. Additionally, a test using ferrocene, which is often used in calibrating instruments, determined that the prototype yielded redox reactions at approximately the same voltages as the commercial potentiostat.

However, a self-assembled monolayer of thiol-terminated porphyrin could not be detected using the prototype. A number of causes could be attributed to this, including insufficient immersion time and CV scan rate. Further study may involve both examination and minimization of these aspects of the initial experimental design.

## 11. ACKNOWLEDGEMENTS

I thank Dr. Jan Van der Spiegel and the SUNFEST program, including the NSF REU Grant EEC-0244055, for providing ample financial and administrative support. In addition, I would

like to acknowledge Douglas Strachan, Danvers Johnston, Samuel Khamis, Yaping Dan, Sujit S. Datta, Tae-Hong Park, and A.T. Charlie Johnson for their invaluable discussion and assistance throughout the duration of this project.

## 12. REFERENCES

1. Horowitz, Paul and Hill, Winfield. The Art of Electronics. Cambridge, UK: Cambridge University Press, 1989.
2. Appl. Phys. Lett. **86**, 043109 (2005).
3. Ann. N.Y. Acad. Sci. 1006: 1-20 (2003).
4. Nuckolls, Colin, et. al. "Molecular Electronic Measurements in Conducting Solutions." Molecular Imaging Corporation, 2005.
5. Appl. Phys. Lett. **84**, 12 (2004).
6. Fisher, A.C. Electrode Dynamics. Oxford, UK: Oxford University Press, 1996.
7. Compton, Richard G. and Sanders, Giles H. W. Electrode Potentials. Oxford, UK: Oxford University Press, 1996.

SUNFEST Technical Report, Center for Sensor Technologies, Dept of Electrical and Systems Eng, Univ. of Pennsylvania, Philadelphia, PA 2006

*University of Pennsylvania*

SUNFEST  
NSF REU Program  
Summer 2006

## **Electromechanical Modeling of a New Class of Contour-Mode AlN MEMS RF Resonators**

NSF Summer Undergraduate Fellowship  
Raúl Pérez Martínez, University of Puerto Rico, Department of Physics and Electronics,  
Humacao  
Advisor: Dr. Gianluca Piazza

### **ABSTRACT**

A new class of contour-mode aluminum nitride micro electro mechanical radio frequency resonators (AlN MEMS RF resonators) is modeled. In the attempt to provide design guidelines for the resonator layout, fully understand the resonator electrical response and improve its performance, a thorough electrical characterization of the contour-mode RF resonators is performed using a high frequency probe station. The equivalent electrical parameters of the resonator are extracted from the obtained plots of admittance versus frequency. Specific attention is placed in modeling any external parasitic capacitances and resistive losses due to the layout configuration, the silicon substrate or the resonator design itself. An extended Butterworth Van Dyke (BVD) is proposed to thoroughly account for the electromechanical response of the resonator in the frequency range of interest.

### **TABLE OF CONTENTS**

- Introduction
- Background
- Rectangular Plate Resonator
- Ring-shaped Resonator
- Experimental Procedure
- Results
- Conclusion
- Recommendations
- Acknowledgment
- References

## 1. INTRODUCTION

The development of RF filters and frequency based devices with reduced power consumption, size and price is receiving lots of attention from the electronics industry. This research shows how this new class of contour-mode MEMS RF resonators works and fits in the vision for the development of these electrical components. The resonators that were electrically characterized in this research use AlN, a piezoelectric material, for the transduction mechanism. When an external electric field is applied across the thickness of this device, the structure undergoes large amplitude vibrations at a fundamental frequency, known as the resonant frequency, and set by the in-plane geometry and material properties of the structure. For the electrical characterization of these devices the scattering parameters are obtained using a Network Analyzer and the equivalent electrical parameters and performance characteristics are extracted, such as the quality factor (Q factor), motional resistance and admittance.

The Q factor of a system describes the sharpness of the systems response to an input signal; it is the ratio of the center (fundamental) frequency to the 3dB bandwidth. For  $Q < 1$ , the system is so strongly damped that it never completes a single oscillation, and in the limit of  $Q = 0$ , it basically decays exponentially toward equilibrium. The resonant behavior of a system depends strongly on  $Q$ ; at the fundamental frequency a system with a high  $Q$  resonates with larger amplitude than one with a low Q factor, and its response falls off more rapidly as the frequency changes. Mathematically for a driven system it is given by,

$$Q = \frac{f_o}{\Delta f} \quad (1)$$

where  $f_o$  is the fundamental frequency and  $\Delta f$  is the 3dB frequency bandwidth.

The motional resistance can be seen as a representing the equivalent impedance of the resonating device at its natural resonant frequency.

Scattering parameters (s-parameters) will be acquired from the electrical measurements and then converted to admittance. The scattering parameters are used for complex circuits to describe the scattering and reflection of a traveling wave signal when a network is inserted into a transmission line. These parameters are usually used but not limited to high frequency applications, and are displayed in terms of frequency. The individual parameters for a 2-port network (contains one input and one output), dimensionless and complex numbers normally expressed as magnitude and phase, in most cases are:

- S<sub>11</sub>: input reflection coefficient
- S<sub>21</sub>: forward transmission coefficient
- S<sub>12</sub>: reverse transmission coefficient
- S<sub>22</sub>: output reflection coefficient

For the one-port network only S<sub>11</sub> exists. Since the resonators and filters are symmetric devices the S<sub>21</sub> and S<sub>12</sub> are equal in most cases. The scattering parameters are measured by sending a signal (in a certain frequency range) into the device and detecting what returns as a reflection of

the signal. For the two-port devices its output is terminated with a  $50 \Omega$  resistor during measurement.

Admittance ( $\mathbf{Y}$ ) is defined as the inverse of the impedance. It is an expression of the ease with which alternating current flows through a complex system. The admittance is composed of a real part (the conductance  $\mathbf{G}$ ), and an imaginary part (the susceptance  $\mathbf{B}$ ), shown by the equation

$$\mathbf{Y} = \mathbf{G} + \mathbf{B} \cdot j \quad (2)$$

## 2. BACKGROUND

Several electrostatically-transduced micromechanical resonators have been demonstrated in the Very High Frequency (VHF, 30 MHz – 300 MHz) and the Ultra High Frequency (UHF, 300 MHz – 3 GHz) range; even though they have a sheer high Q (order of 10,000), they exhibit large values of motional resistance (100s k $\Omega$ ), which complicates the coupling of these electrostatic devices to  $50 \Omega$  RF systems.<sup>[2]</sup> In order to reduce the equivalent motional resistance research trying to act on geometrical parameters, such as the effective actuation area, has been done but not shown significant improvements. The use of piezoelectric transduction has tremendous advantage over electrostatically-driven contour-mode resonator and solves the problem of large motional resistance.<sup>[2]</sup>

When compared to surface-based electrostatic forces, piezoelectric body forces offer higher electromechanical coupling coefficients making possible low values of motional resistance. The RF MEMS resonators used in this research are called contour-mode resonators since the fundamental frequency is defined by the in-plane dimensions of the structure (figure 1). This type of design allows for batch fabrication of arrays of piezoelectric resonators with different frequencies on a single chip unlike with other designs (film bulk acoustic wave resonators and shear mode quartz resonators) in which the frequency is set by the film thickness instead of its in-plane dimensions.<sup>[3]</sup> The contour-mode resonator design, developed in the shape of disks, plates, or rings, have the unique features of providing multiple frequencies with the ability to interface directly with  $50 \Omega$  systems.

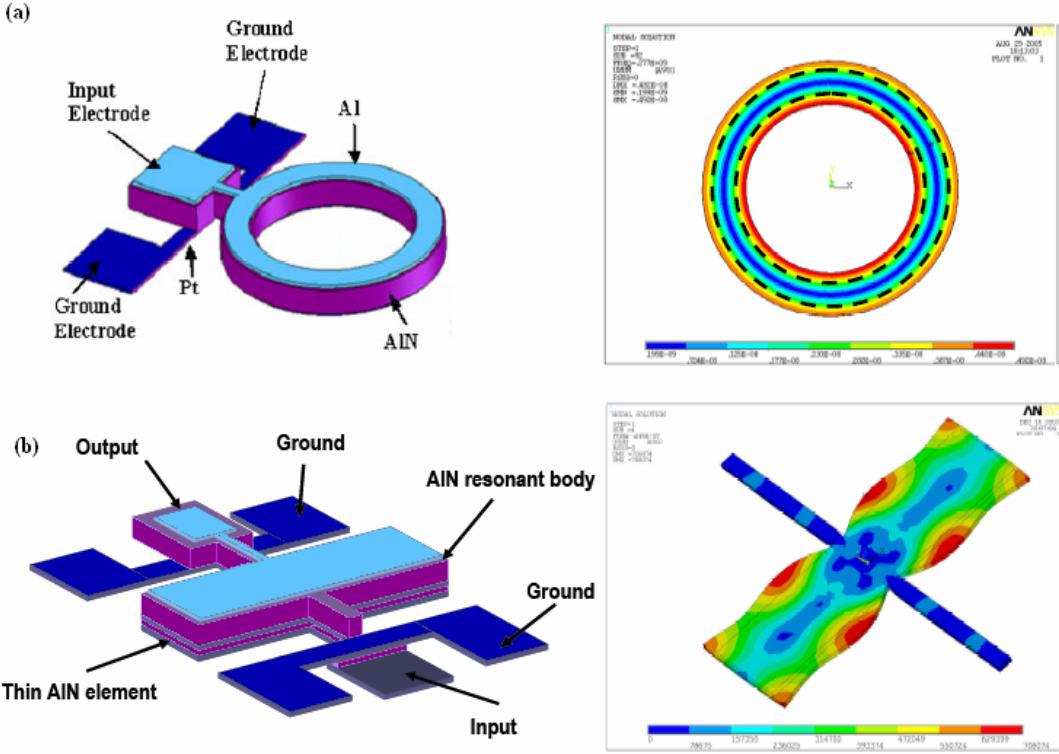


Figure 1: Resonating elements and their vibrating modes shown for the one-port ring resonator (a) and for the two-port rectangular resonator (b).

The piezoelectric resonator consists of a piezoelectric material that mechanically vibrates when an alternating electric field at the proper frequency is applied to it, and in which the highest amplitude of vibrations coincides with the lowest impedance of the system. The mechanical vibrations near the resonant frequency depend on the inertia, elastic compliance and stiffness of the resonant system. These devices can be looked at electrically as an equivalent Butterworth-Van Dyke (BVD) model. The BVD model consists of an inductance, capacitance and resistance, corresponding respectively to inertia, compliance and damping of the mechanical system; connected in series with a parallel capacitance, which is due to the dielectric properties of the piezoelectric crystal, capacitive coupling between the electrodes (figure 2).<sup>[2]</sup>

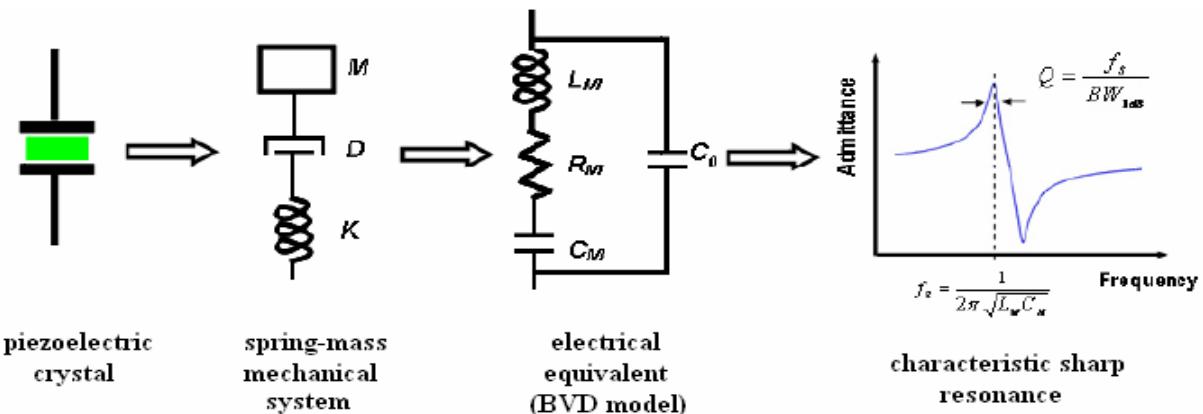


Figure 2: Analogy between the mechanical and electrical domain of a piezoelectric crystal. The resonating piezoelectric material can be seen as a lumped spring, mass, damper mechanical system; this in the electrical domain can be seen as an equivalent Butterworth-Van Dyke model which is characterized by a sharp resonance.

Two different configurations of the ring-shaped and the rectangular shaped resonators were used: the one-port and the two-port configurations. The one-port configuration contains only one input, while the two-port configuration contains one input and one output (figure 3).

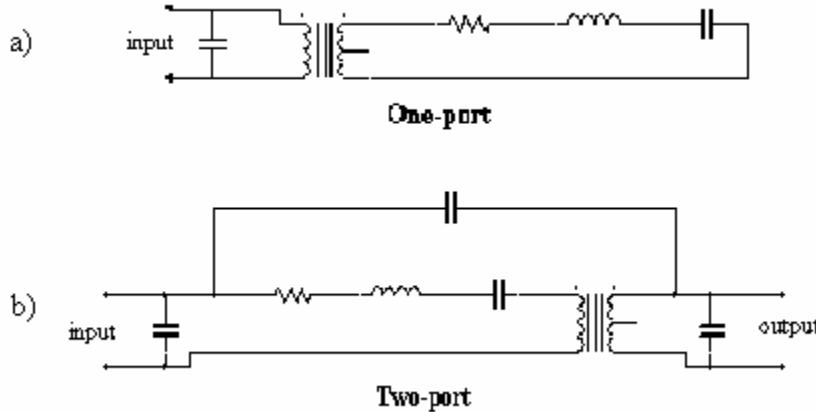


Figure 3: a) Electrical equivalent for a one-port configuration (notice the single input); b) Electrical equivalent for a two-port configuration (notice the input and output).

Both configurations have their advantages and disadvantages. Some advantages for the one-port configuration are: it avoids unwanted modes, minimizes motional resistance and maximum energy is coupled into the system; the two-port configuration eliminates all kinds of feedthrough and it maintains a high electromechanical coupling. A disadvantage for the one-port

configuration is that it encounters parasitic feedthrough at high frequencies, while for the two-port configuration spurious modes are encountered and its fabrication process is more complicated than that of the one-port configuration.

### 3.1 RESONATING DEVICES

To study the behavior of the device the equivalent impedance was calculated. The individual impedance for the capacitance, resistance and inductance are as follow:

$$Z_C = \frac{1}{C \cdot w \cdot j} \quad Z_R = R \quad Z_L = L \cdot w \cdot j \quad (3)$$

where  $w$  is the angular frequency,  $j$  is the complex or imaginary component, and  $C$ ,  $R$ , and  $L$  correspond to the capacitance, resistance and inductance, respectively. Since these components are in series, the equivalent impedance is simply their sum, but there is also a parallel capacitance ( $C_o$ ) to the **RCL** components. The impedance for the two port circuit is:

$$Z_{eq} = \frac{1}{C_o \cdot w \cdot j} \left\| R + \frac{1}{C \cdot w \cdot j} + L \cdot w \cdot j \right. \\ Z_{eq} = \left( C_o \cdot w \cdot j + \frac{1}{R + \frac{1}{C \cdot w \cdot j} + L \cdot w \cdot j} \right)^{-1} \quad (4)$$

$$Z_{eq} = \frac{R + j \cdot (w \cdot L - 1/w \cdot C)}{1 - w^2 \cdot L \cdot C_o + C_o/w + j \cdot w \cdot R \cdot C_o}$$

Multiplying the numerator and denominator by the complex conjugate of the previous equations gives [1]:

$$Z_{eq} = \frac{R + j \cdot (w \cdot L - 1/w \cdot C - w^3 \cdot L^2 \cdot C_o + 2 \cdot w \cdot L \cdot C_o/C - C_o/w \cdot C^2 - w \cdot R^2 \cdot C_o)}{C_o^2/C^2 + 1 - 2 \cdot w^2 \cdot L \cdot C_o + 2 \cdot C_o/C + w^2 \cdot R^2 \cdot C_o^2 - 2 \cdot w \cdot L \cdot C_o^2/C + w^4 \cdot L^2 \cdot C_o^2} \quad (5)$$

For the BVD circuit to be resonant the imaginary part should be considered equal to zero, doing so the equation changes to:

$$w^2 \cdot L \cdot C^2 - C - w^4 \cdot L^2 \cdot C^2 \cdot C_o + 2 \cdot w^2 \cdot L \cdot C \cdot C_o - C_o - w^2 \cdot R^2 \cdot C^2 \cdot C_o = 0 \quad (6)$$

The equation can be further simplified by eliminating the last term which contains  $R$  since for most practical resonators it is negligible compared to other terms in calculating the resonant frequencies. By doing so and using the quadratic formula the resonant frequency and the anti-resonant of parallel frequency are obtained:

$$w_R^2 = \frac{1}{L \cdot C} \rightarrow f_R = \frac{1}{2 \cdot \pi} \cdot \sqrt{\frac{1}{L \cdot C}} \quad (7)$$

$$w_A^2 = \frac{1}{L \cdot C} + \frac{1}{L \cdot C_o} \rightarrow f_A = \frac{1}{2 \cdot \pi} \cdot \sqrt{\frac{1}{L \cdot C} + \frac{1}{L \cdot C_o}} \quad (8)$$

Subtracting the two previous equations the following relationship is obtained

$$w_A^2 - w_R^2 = \frac{1}{L \cdot C_o} \quad (9)$$

$$2 \cdot w \cdot \Delta w = \frac{1}{L \cdot C_o} \quad (10)$$

Since the resonant and anti-resonant frequencies are nearly the same, equation (10) is assumed true. Therefore the following parameters can be derived:

$$L = (8 \cdot \pi^2 \cdot f \cdot C_o \cdot \Delta f)^{-1} \quad (11)$$

$$C = \frac{2 \cdot \Delta f \cdot C_o}{f} \quad (12)$$

$$Q = \frac{w \cdot L}{R} = (4 \cdot \pi \cdot R \cdot C_o \cdot \Delta f)^{-1} \quad (13)$$

where  $\Delta f = f_A - f_R$ .

The network analyzer measures the scattering parameters in a frequency range. These data is converted to admittance and different methods were used for the one-port and two-port configuration. For the one-port configuration the following equation was used,

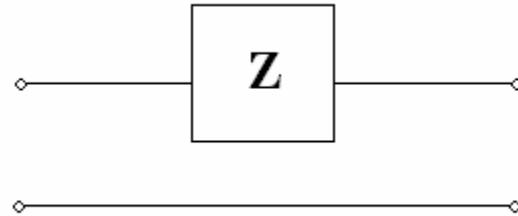
$$Y = \frac{1 - S_{11}}{Z_0 (1 + S_{11})} \quad (14)$$

in which  $Z_0$  is the  $50\Omega$  resistance connected to the port of the device.

For the two-port device ABCD parameters are used. These are more convenient for cascading networks than the scattering parameters or impedance parameters. Mathematically they are defined as,

$$\begin{pmatrix} V_1 \\ I_1 \end{pmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{pmatrix} V_2 \\ I_2 \end{pmatrix} \quad (15)$$

in which the subscript indicates the input (1) and output (2). Schematically it can be seen as shown below,<sup>[4]</sup>



where  $A = 1$ ,  $B = Z$ ,  $C = 0$ , and  $D = 1$ .

The  $S_{12}$  parameter is obtained and using the values shown above and the following equation,

$$S_{12} = \frac{2 \cdot (A \cdot B - B \cdot C)}{A + B/Z_0 + C \cdot Z_0 + D} \quad (16)$$

we can simplify and obtain,

$$S_{12} = \frac{2}{2 + B/Z_0}, B = Z \rightarrow Y = \frac{1}{\left(2/Z_{12} - 2\right) \cdot Z_0} \quad (17)$$

To better understand the impact of the electrical parameters on the resonators response, a simulation program was written using SigmaPlot. For this program the values of each parameter were changed a certain percentage and the variations on the resonance curve were recorded.

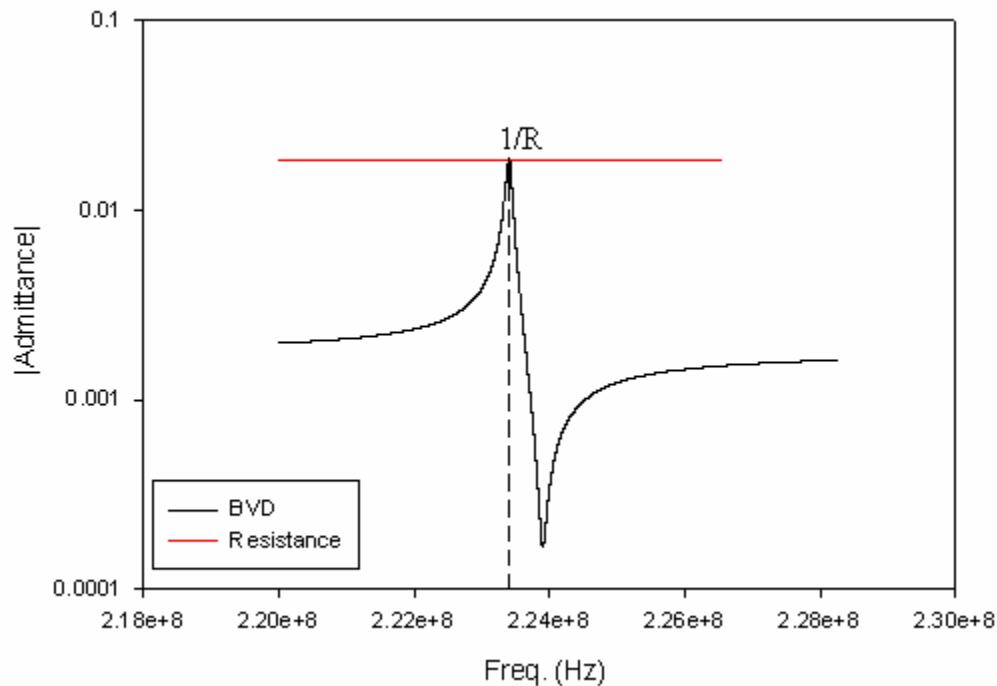


Figure 4: Admittance plot of the resonator system and the single resistance that is part of such system. Note the point where the two plots intersect.

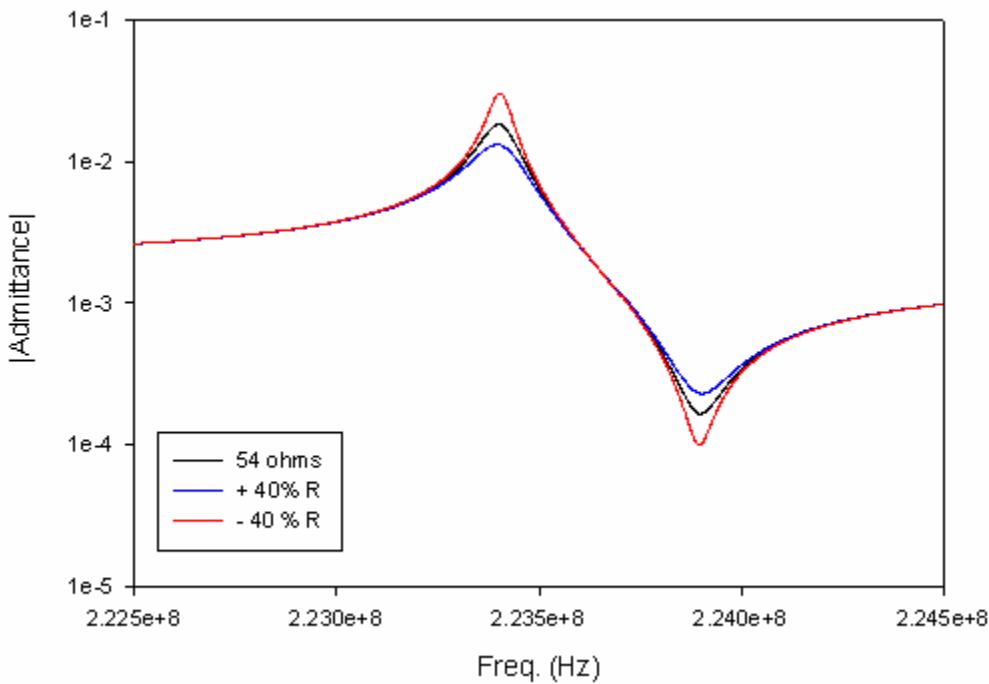


Figure 5: Variations on the resonator response due to changes in the value of the resistance.

Figure 4 shows the admittance plot of a resistance with a value identical to the one of the resonator and the plot of the resonator system; it shows how the plots coincide at the resonant

peak showing that the resonator behaves as a resistor at resonance. The highest amplitude of the peak of the resonator system is in the point of the lowest impedance; this behavior is seen clearly in figure 5 with the variations on the curve based on different values of the resistance of the BVD system in which the highest peak is when a lower resistance value is used in the resonator.

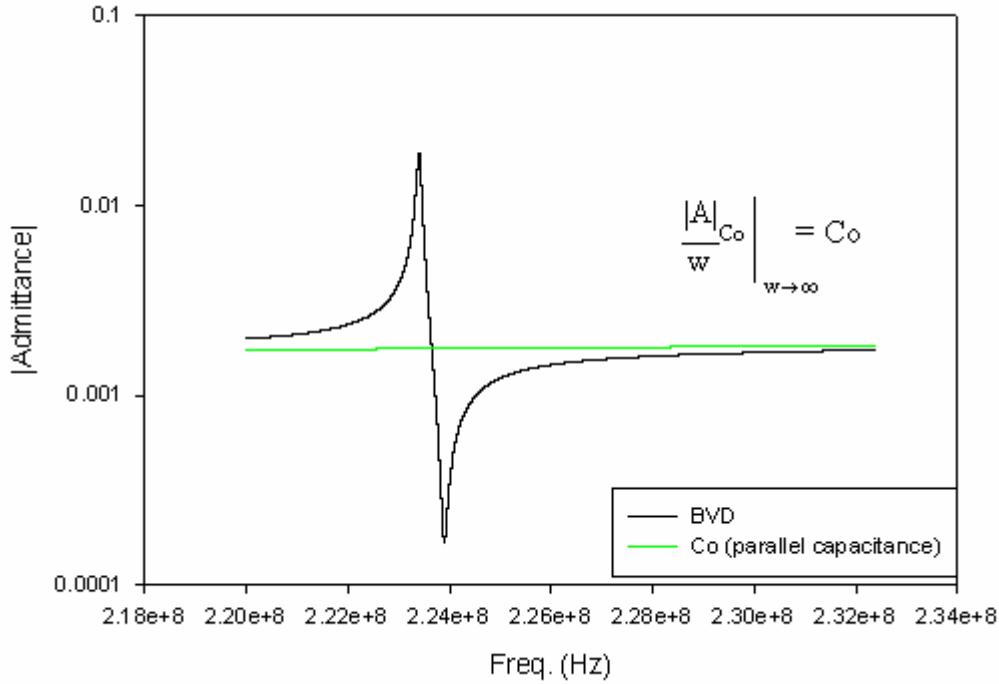


Figure 6: Admittance plot of the resonator system and the parallel capacitance that is part of such system. Note that the curves tend to coincide as the frequency increases.

Figure 7 shows how  $C_0$  defines the anti-resonance peak when changing the inductance and capacitance while keeping the resonance frequency constant. This capacitance has no effect over the resonance frequency, but it affects directly the anti-resonance peak of the curve, as predicted by equations (7) and (8), if  $C_0$  is too big the resonator signal is then masked by the capacitor.

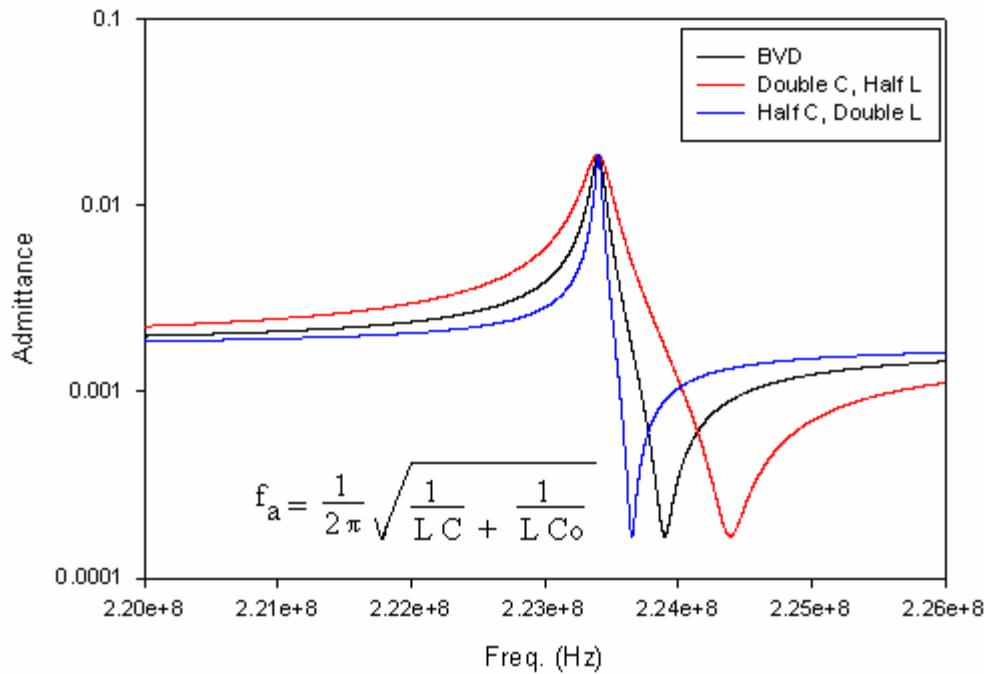


Figure 7: Admittance versus frequency keeping the resonant frequency and the parallel capacitance constant, and changing the values of the capacitance and inductance.

### 3.2 RECTANGULAR PLATE RESONATORS

A vertical electric field applied across the film thickness induces in plane dilation of the structure and excites the resonator either in length-extensional or width-extensional mode shapes, depending whether the structure vibrates primarily across its length or width and the excitation frequency.

Several modes of vibrations can be found in the rectangular plate and its detection depends on both the strength of the electromechanical coupling associated with the particular mode and the quality factor of the mechanical structure; just those mode shapes that undergo a net area change over the whole electrode surface generate a net motional current and can be detected electrically.

The designer can change the width for length-extensional mode and length for width-extensional mode to independently set the motional resistance.

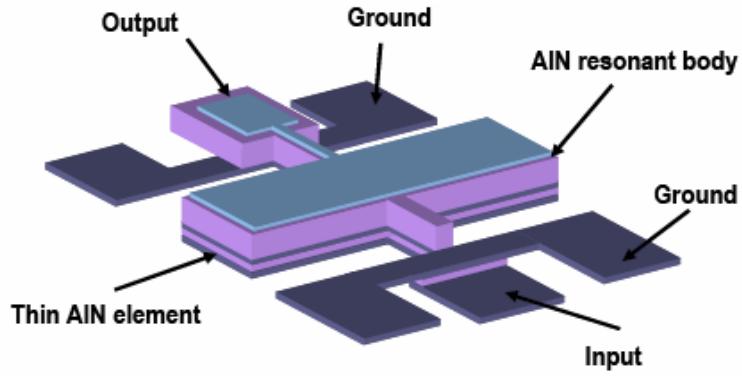


Figure 8: Schematic diagram of rectangular plate two-port resonator.

### 3.3 RING-SHAPED RESONATORS

The one-port rings used have a  $20\mu m$  wide and  $90\mu m$  inner radius circular ring (figure 9). The width of the ring fundamentally sets the center frequency. The ring is suspended over the substrate and held by the tether which is connected to the electrodes. One-port devices in which the whole top surface is electroded, guarantees that maximum energy is coupled into the desired mode. It also allows for minimal motional resistance obtained and minimizing undesired modes.<sup>[3]</sup>

An electric field applied across the thickness of the film tends to dilate and contract the structure and makes it vibrate in a breathing mode across its width.

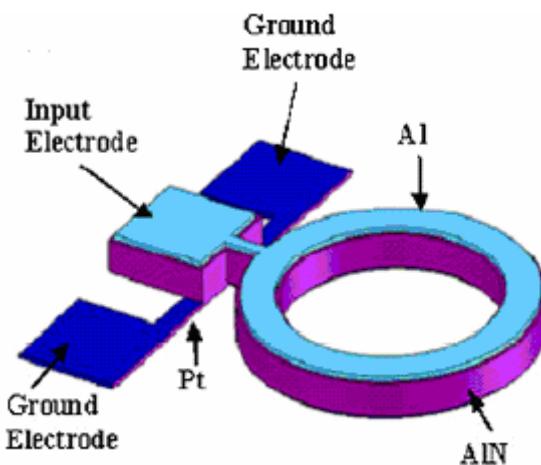


Figure 9: Schematic diagram of ring-shaped one-port resonator.

### 3.4 EXPERIMENTAL PROCEDURES

The resonator devices that were experimentally tested and analyzed in this report had been previously manufactured and just need to be cleaned and released. The devices were placed in a ion plasma etcher with a gas flow of 16 scm of  $O_2$  and applying a power of 350 watts for 5 minutes; this step was performed to remove any organic material on the surface of the device including a photo resist layer. The same plasma etcher was also employed to release the vibrating element from the silicon substrate (See figure 10). A flow of 35 mm (8.2 cc) of  $SF_6$  and a power of 150 watts were used. A Desert Cryogenics high frequency probe station was used in order to probe the resonating devices and an Agilent Network Analyzer PNA-L series (300 kHz to 20 GHz) to obtain the  $S_{11}$  parameters and at the same time convert them to admittance (figure 11). The one-port  $S_{11}$  parameter calibration (SOL) was performed using short (S) open (O) and load (L) on a ceramic substrate, while for the two-port calibration (SOLT) was performed using short (S) open (O) load (L) and thru (T) on the same ceramic substrate.

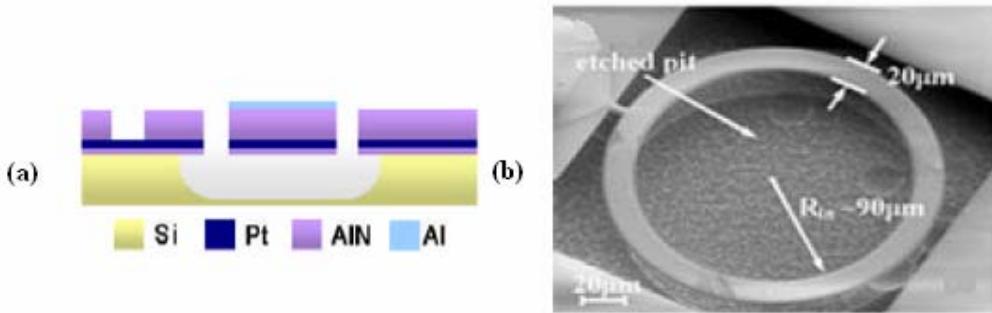


Figure 10: (a) Cross-sectional schematic view of the release of the vibrating elemet; (b) SEM image of ring suspended over the substrate by the tether.

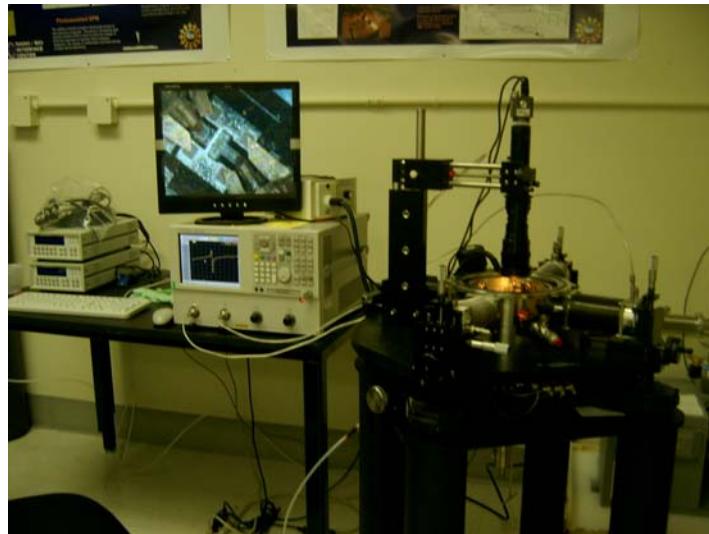


Figure 11: Work station: probe station located at the right the left and next to it the network analyzer taking measurements.

### 3.5 RESULTS

Scattering parameters were measured using the network analyzer and were then converted to admittance using equations (14) for the one-port devices and (17) for the two-port devices. From the admittance plot the electrical parameters were calculated using,

$$R = \frac{Y_A}{w_A^2 \cdot C_0^2} \quad (18)$$

$$\frac{1}{Y_R} = \frac{R}{1 + (w_R \cdot R \cdot C_0)^2} \quad (19)$$

$$w_A^2 - w_R^2 = \frac{1}{L \cdot C_0} \quad (20)$$

$$w_R = \frac{1}{\sqrt{L \cdot C}} \quad (21)$$

Substituting equation (18) in (19) we obtain  $C_0$  and then obtain the value of  $R$  from equation (18); then using the remaining equations we obtain the rest of the unknown parameters,  $L$ , and  $C$ . The quality factor is calculated using equation (13).

After extracting the electrical parameters an admittance plot was derived and compared to the experimental results. Figure 12 shows the plots for the one-port rings with their respective electrical parameters and quality factor. Notice the slight shift of the curve in the frequency, this is due to issues with the calculation of the inductance and the capacitance of the system which define the resonant and anti-resonant peak.

In figure 13 the admittance plots measured with the network analyzer and calculated of the extrapolated electrical parameters for the two-port rectangular devices. In this case using the same technique as for the one-port devices, the curve is not shifted but the curves do not coincide at frequencies lower or higher than the resonant and anti-resonant peaks respectively; which is due to a miscalculation of  $C_0$  since this parameter defines the curve at both extremes.

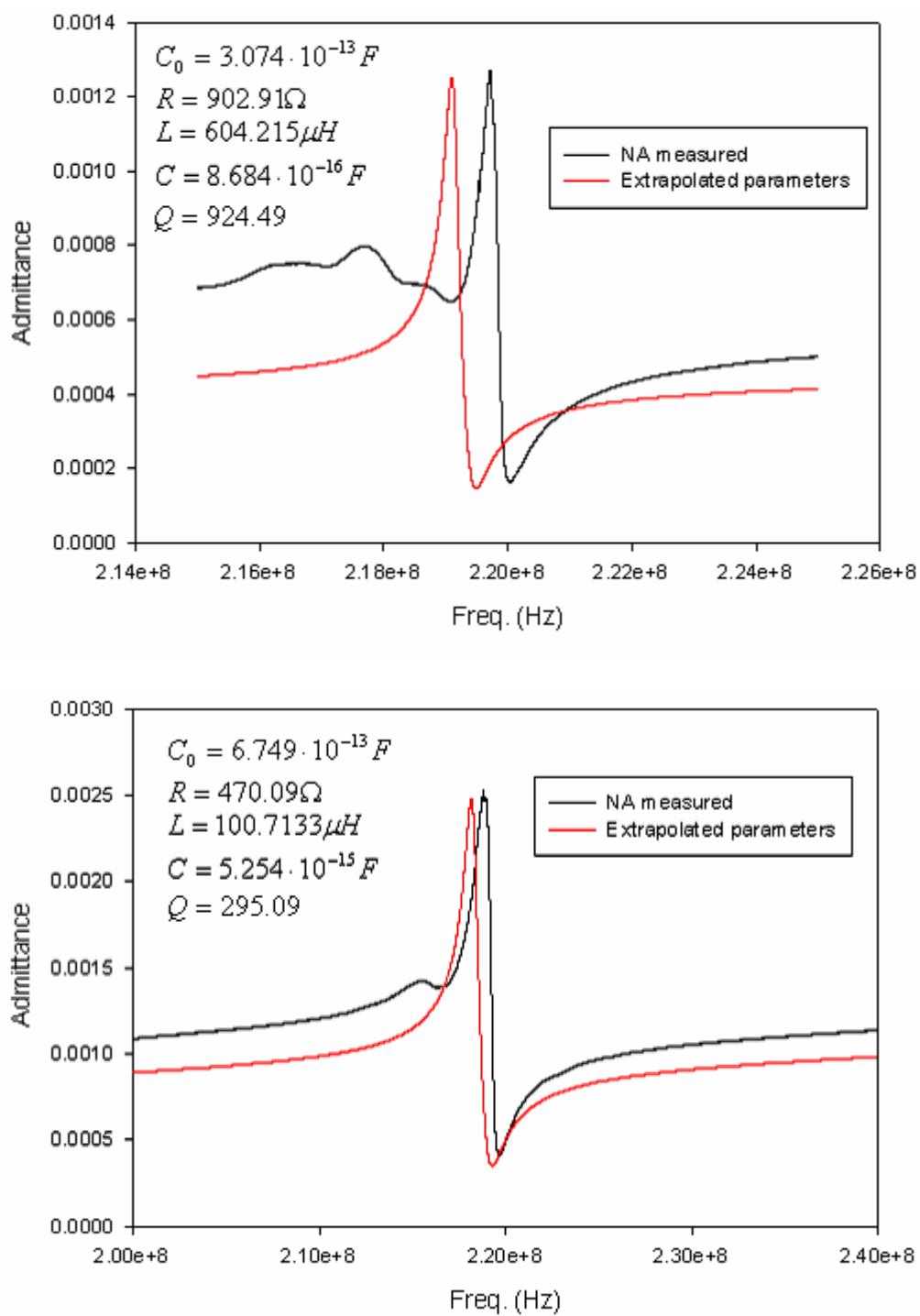


Figure 12: Admittance plot of the two one-port ring resonator systems compared to the plot of their respective extrapolated parameters.

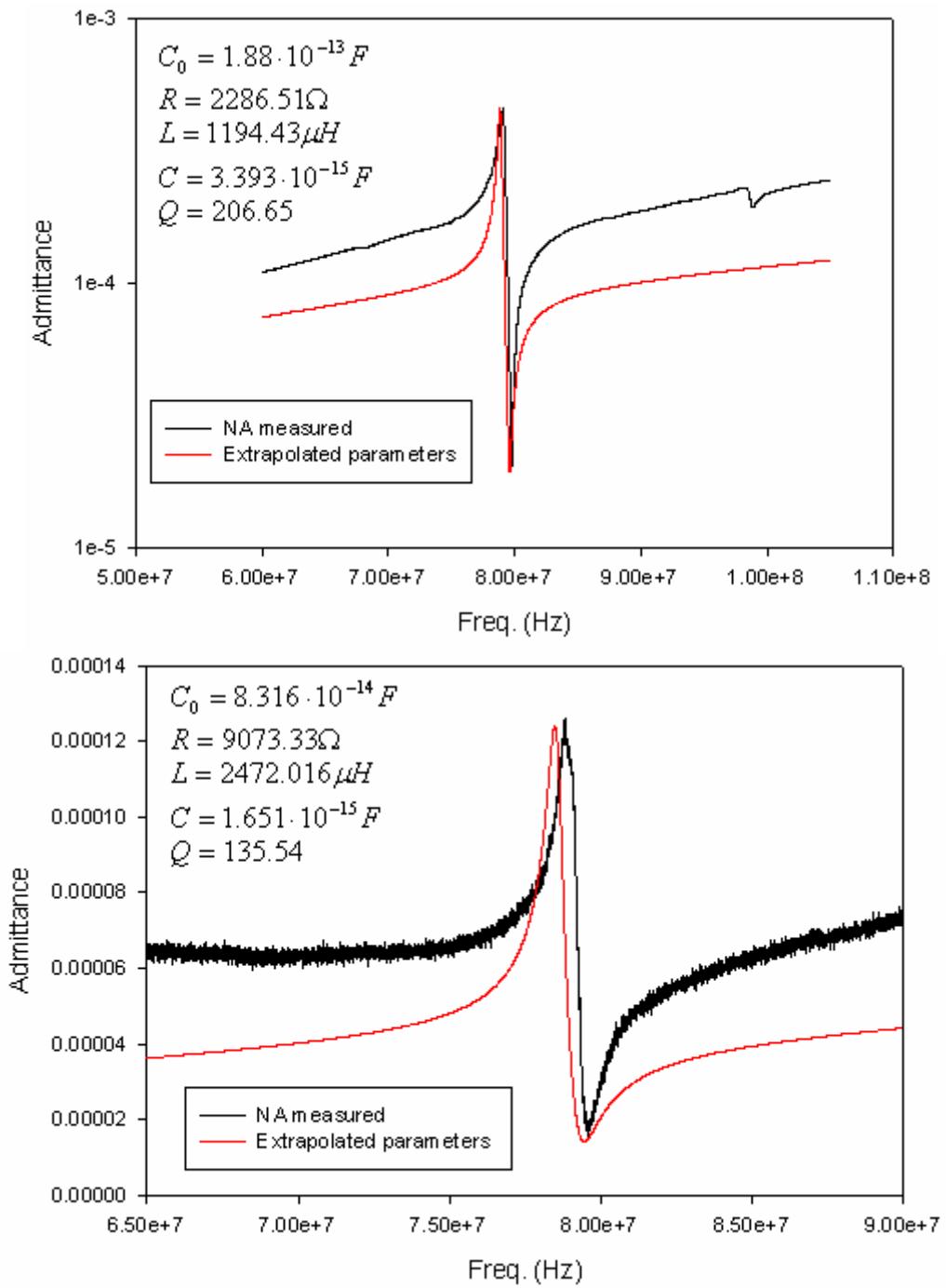


Figure 13: Admittance plot of the two two-port rectangular resonator systems compared to the plot of their respective extrapolated parameters.

## **4. CONCLUSION**

The extrapolation technique used is a good first approximation for the admittance curve of the resonator systems. Improvement could be achieved by using more critical points in the admittance plot to extrapolate the electrical parameters and by adding elements to the electrical model of the resonating system.

## **5. RECOMMENDATIONS**

Devices measured were not fabricated recently, and contact with the atmosphere probably caused oxidation on the electrodes; to avoid this measurements on new devices is recommended. Contact issues were also encountered with the probes of the probe station, dealing with this would allow for better measurements. The probe station also has some features that were not used. For example, pressure can be lowered to level around 5  $\mu$ Torr, and temperature can be varied from 400 K – 4.5 K. Further characterization can be done using these features.

## **6. ACKOWLEDGEMENTS**

I wish express my deepest gratitude first to my advisor, Dr. Gianluca Piazza for giving me the opportunity to work with him and helping me become a better scientist. I also want to acknowledge the advice, help and guidance offered to me by my fellow lab partners: Carlos Perez, Nipun Sinha and Dwo Chu I also want to thank Dr. Van Der Spiegel, the University of Pennsylvania, and University of Puerto Rico, Humacao, the SUNFEST Program and NSF.

## **7. REFERENCES**

- [1] V.E. Bottom, Introduction to Quartz Crystal Unit Design, *Van Nostrand Reinhold Electrical/Computer Science and Engineering Series*, 1982, pp. 82-133
- [2] G. Piazza, *Piezoelectric Aluminum Nitride Vibrating RF MEMS for Radio Front-End Technology*, 2005, Doctorate Dissertation
- [3] G. Piazza, P.J. Stephanou, A.P. Pisano, *Piezoelectric Aluminum Nitride Vibrating Contour-Mode MEMS Resonators*, 2006
- [4] D.M. Pozar, *Microwave Engineering* Second Edition, John Wiley & Sons Inc., 1998, pp.182-221
- [5] J.D. Larson III, P.D. Bradley, S. Wartenberg, R.C. Ruby, *Modified Butterworth-Van Dyke Circuit for FBAR Resonators and Automated Measurement System*, 2000 IEEE ULTRASONICS SYMPOSIUM 2000 pp. 863-868

[6] J.D. Larson III, R.C. Ruby, P.D. Bradley, J. Wen, S. Kok, A. Chien, *Power Handling and Temperature Coefficient Studies in FBAR Duplexers for the 1900 MHz PCS Band*, 2000 IEEE ULTRASONICS SYMPOSIUM 2000 pp. 869-874

[7] W.G. Cady, *Piezoelectricity: An Introduction to the Theory and Applications of Electromechanical Phenomena in Crystals* Vol. 1, Dover Publications Inc. New York, 1964, pp. 284-307

[8] J.D. Larson III, Y. Oshmyansky, *Measurement of Effective  $k_t^2$ , Q, Rp, Rs vs. Temperature for Mo/AlN FBAR Resonators*, 2002 IEEE ULTRASONICS SYMPOSIUM 2002 pp. 939-943

# Fabrication of PVA Micropolarizer Arrays for a CMOS Image Sensor

NSF Summer Undergraduate Fellowship in Sensor Technologies

**Nathan Lazarus (Electrical Engineering), University of Pennsylvania**

Advisors: Dr. Jan Van der Spiegel  
Dr. Viktor Gruev

## ABSTRACT

Most image sensors ignore the polarity of light signals, primarily because the human eye is not sensitive to polarization. However, it is possible to gather valuable information about geometry and composition based on the polarity of light reflecting off of an object. A polarization sensor has been designed combining a CMOS image sensor with micropolarizers fabricated out of polarizing polyvinyl alcohol (PVA) sheets. This project was focused on fabricating these micropolarizers using photolithography and etching. In order to obtain complete characterization for each pixel, it is necessary to obtain the intensity of light polarized at 0 degrees, at 45 degrees, and in total. As a result, the micropolarizer array must contain two layers of 10-micron PVA structures oriented 45 degrees to each other. A single layer of micropolarizers was created in the PVA film using plasma etching, but exhibited significant etching underneath the structures created in the photoresist by lithography. A second etching technique, reactive ion etching (RIE), was evaluated and determined to cause significantly less underetching. Procedures for gluing layers of micropolarizers together and for removing the remaining photoresist were also developed. After a mask is fabricated with the proper alignment markers to allow layers of micropolarizers to be aligned relative to each other and to the CMOS chip, a complete polarization image sensor will be assembled using these techniques.

## Table of Contents

<b>1. INTRODUCTION</b>	<b>3</b>
<b>2. OVERVIEW OF POLARIZATION IMAGE SENSORS</b>	<b>3</b>
<b>2.1 Polarization</b>	<b>3</b>
<b>2.2 Polarization Applications</b>	<b>4</b>
<b>2.3 Polarization Image Sensors</b>	<b>4</b>
<b>2.4 Micropolarizer Fabrication</b>	<b>5</b>
<b>2.5 Polarization Sensor Design</b>	<b>5</b>
<b>3. SINGLE LAYER MICROPOLARIZER FABRICATION</b>	<b>6</b>
<b>3.1 Sample Preparation</b>	<b>6</b>
<b>3.2 Lithography</b>	<b>7</b>
<b>3.3 Etching</b>	<b>8</b>
<b>3.4 Optical Properties</b>	<b>9</b>
<b>3.5 Plasma Etching vs. RIE</b>	<b>10</b>
<b>4. DOUBLE LAYER MICROPOLARIZER ARRAYS</b>	<b>13</b>
<b>4.1 Methods for Gluing Multiple Layers</b>	<b>13</b>
<b>4.2 Aligning Multiple Layers</b>	<b>13</b>
<b>4.3 SU-8 Removal</b>	<b>14</b>
<b>5. DISCUSSION AND CONCLUSIONS</b>	<b>14</b>
<b>6. RECOMMENDATIONS</b>	<b>15</b>
<b>7. ACKNOWLEDGMENTS</b>	<b>15</b>
<b>8. REFERENCES</b>	<b>15</b>

## 1. INTRODUCTION

Polarization imaging is an emerging field with important applications in computer vision. Traditional image sensors are not built to be sensitive to polarization, focusing on other aspects such as color or total light intensity. However, the polarization of light reflecting off an object does carry useful information about the shape and composition of that object. This could allow computers to track objects, or robots to determine their own motion based on the movement of the scenery around them. The current generation of polarization sensors is dependent on bulky mechanically moving polarizing filters, limiting their usefulness. Viktor Gruev and Jan Van der Spiegel have designed a CMOS sensor that will be able to calculate the polarization parameters of natural light without mechanically moving filters, overcoming this limitation.

This project is focused on creating the micropolarizer array necessary to complete a CMOS image sensor. It is a continuation of a 2005 SUNFEST project carried out by Kejia Wu of the University of Pennsylvania which developed techniques to create a single layer of micropolarizers [1]. This project continues that work, and centers on preparing to fabricate the micropolarizers directly on the image sensor.

## 2. OVERVIEW OF POLARIZATION IMAGE SENSORS

### 2.1 Polarization<sup>1</sup>

The orientation of the electric fields of light is known as the polarization. There are numerous ways to represent this phenomenon mathematically, but one common form is the Stokes Vector. The Stokes Vector is composed of four elements:  $S_0$  is the total intensity,  $S_1$  is the degree of polarization at two perpendicular axes,  $S_2$  is the degree of polarization at two perpendicular axes  $45^\circ$  to the right of the previous set, and  $S_3$  is the degree of circular polarization. Circular polarization is a phenomenon uncommon in most light sources of interest, and is not sought in our application. These elements can be expressed as shown below:

$$\begin{aligned}S_0 &= I_{\text{TOTAL}} \\S_1 &= 2*I_0 - I_{\text{TOTAL}} \\S_2 &= 2*I_{45} - I_{\text{TOTAL}} \\S_3 &= I_{\text{TOTAL}} - 2*I_{45, \pi/2}\end{aligned}$$

$I_{\text{TOTAL}}$  is the total intensity of light,  $I_0$  is the intensity of light polarized at 0 degrees,  $I_{45}$  is the intensity of light polarized at 45 degrees, and  $I_{45, \pi/2}$  is the intensity of light polarized at 45 degrees after a phase shift of  $\pi/2$ .

---

<sup>1</sup> This overview of polarization was compiled with the help of [2] and [3]; for a more detailed presentation, refer to those sources.

## 2.2 Polarization Applications

The polarization of light varies significantly based on the composition of and angle of reflection off of an object. This has relevance in a number of different applications. Polarization filters are commonly used in sunglasses and photography to minimize glare. Past research has also been done to demonstrate a number of other applications of light polarization. Wolff and Boult [4] demonstrated that metals and dielectrics can be distinguished by the polarization of light reflected off of them; this could have valuable applications in computer vision and robotics. Figure 1 below shows one application of polarization imaging.

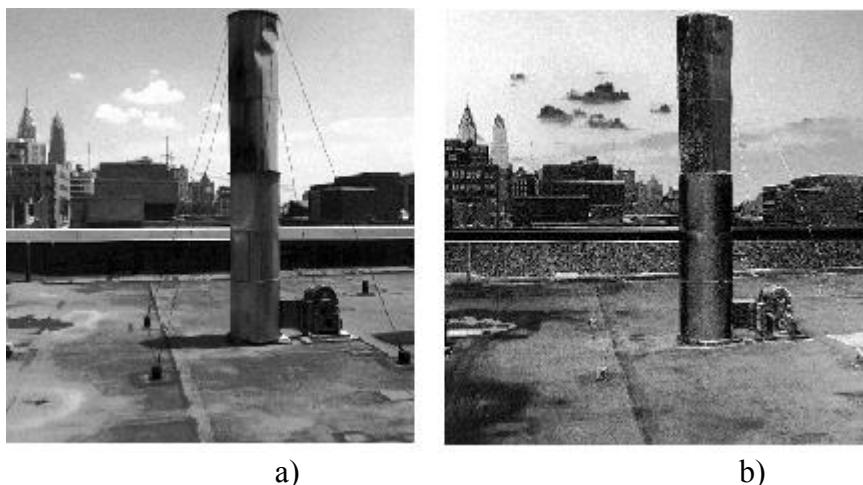


Figure 1: The Philadelphia skyline a) as an intensity image, b) with the degree of polarization shown.

The picture on the left is a conventional grayscale picture of the Philadelphia skyline, using the intensity of light at each pixel. The picture on the right shows the same skyline using the degree of polarization. The pictures are very similar, but the polarization picture shows much better detail in areas of low light, such as the wall and in the shadow of the smokestack. Another important observation is the difference between the sky, which is uniformly polarized, and the clouds which exhibit random polarization.

## 2.3 Polarization Image Sensors

In the 1980s and early 1990s, most work on polarization imaging used polarization image sensors based on mechanically rotating filters in front of a standard camera (e.g. in [4]). These sensors were able to retrieve the polarization components from the incoming light, but tended to introduce optical distortions and required the polarization image frame rate to be significantly slower than that of the camera used. In 1995, Wolff and Andreou [5] introduced two alternative implementations of polarization imagers. The first technique used liquid crystals controlled by an electric field to replace the mechanically rotating polarization filter; this avoided some of the optical distortion effects of the rotating filter. Their second technique was to use a beam splitter to divide orthogonal polarization components of the incoming light and direct that light onto two

cameras. This allowed the sensing of polarization components, but required complicated optics and two distinct image sensor arrays. Our research is focused on a third technique proposed in [5]: this involves coating the image sensor directly with polarizing filters known as micropolarizers to retrieve the components of light polarized at  $0^\circ$  and  $45^\circ$ . Andreou and Kalayjian [6] and Momeni and Titus [7] implemented a polarization sensor using micropolarizers, but their sensor had much more limited resolution.

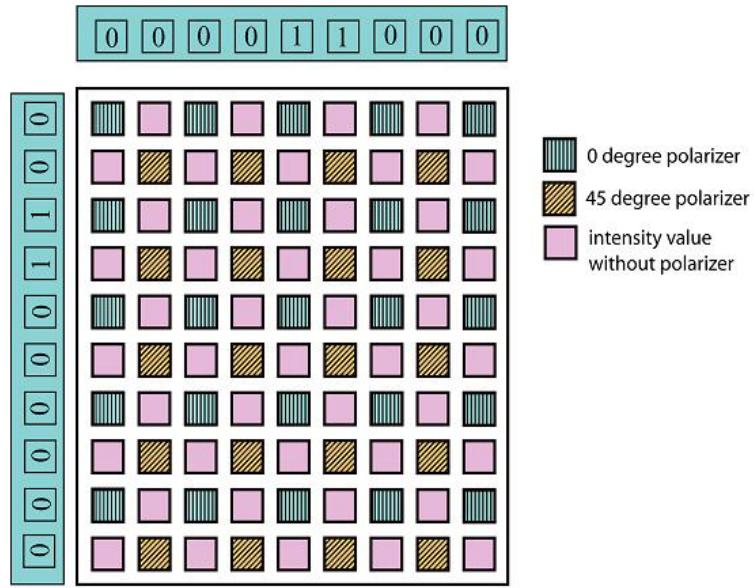
## 2.4 Micropolarizer Fabrication

Micropolarizer arrays can be fabricated with a variety of different techniques. One technique that has been used successfully in the past is to create micropolarizers using various structures in patterned metal. One research team successfully created micropolarizers using chromium strips fabricated on silica [8]. Another group has worked on the possibility that micropolarizers could be fabricated using columns of anodized alumina [9]. Micropolarizer effects could also be achieved with several other techniques. Research has been done developing micropolarizers by etching away segments of birefringent  $\text{CaCo}_3$  crystal [10]. Another possibility that has been developed has been created using liquid crystal micropolarizers created on a thin gold film [11].

One of the most promising areas of micropolarizer research has been creating micropolarizer arrays using thin polymer-based polarized films. Junpeng Guo and David Brady [12] were able to create 5-micron pixel sizes using 1-micron thick film of a dichroic dye solution called Polacoat. Our goal is to create an array of filters using the more common PVA polarizer film, easily found in commercial polarizer filters. Guo and Brady [13] successfully have made dual-axes 25-micron structures in PVA; we would like to fabricate a micropolarizer array with structures as small as 10 microns.

## 2.5 Polarization Sensor Design

Our polarization image sensor is intended to obtain the first three Stokes parameters for every pixel. This requires the total light intensity and the intensities filtered with  $0^\circ$  and  $45^\circ$  polarization filters. Figure 2 below shows a simplified layout of the micropolarization pattern on the image sensor.



**Figure 2: Layout of polarization image sensor.**

As shown, each group of four adjacent photodiodes contains photosensitive elements below polarizing filters oriented at  $0^\circ$  and  $45^\circ$ , as well as two photosensitive elements without filters to obtain the total intensity at that pixel. By turning on two adjacent rows and two adjacent columns (denoted by 1's in the figure), the necessary components can be obtained to find the Stokes parameters at that pixel. In order to achieve this, two layers of micropolarizers oriented at 45 degrees to each other are required.

### 3. SINGLE LAYER MICROPOLARIZER FABRICATION

#### 3.1 Sample Preparation

The first step in creating a micropolarizer array is to isolate the thin film of PVA in a commercial polarizing filter. Most commercial polarizing filters containing PVA use an arrangement similar to that shown in Figure 3 below.

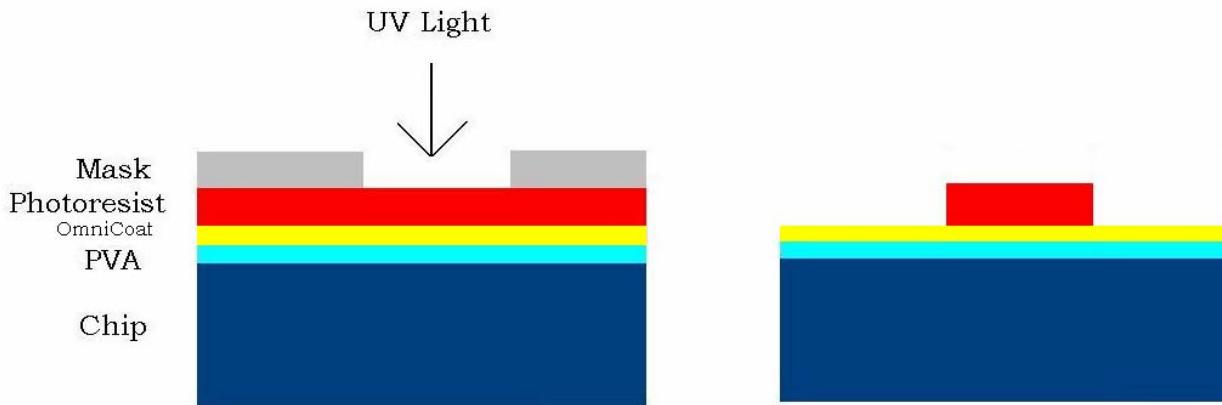


**Figure 3: Commercial polarizing filter a) before and b) after the top CAB layer is removed.**

A thin layer of PVA, generally about 20 microns thick, is sandwiched between two thicker (300 micron) layers of transparent backing material, in this case cellulose acetate butyrate (CAB). Since the micropolarizers need to be formed in the PVA layer, one of the two layers of CAB needs to be removed. The second CAB layer is left in place to provide backing; this is accomplished by protecting the back of the sample with chemical resistant tape. The top layer of CAB is then weakened by soaking the sample in acetone. The acetone attacks the CAB but does not react with the PVA film. Most of the CAB is removed by wiping it off the sample by hand. The final layers close to the PVA are dislodged using de-ionized water to avoid scratching the PVA.

### **3.2 Lithography**

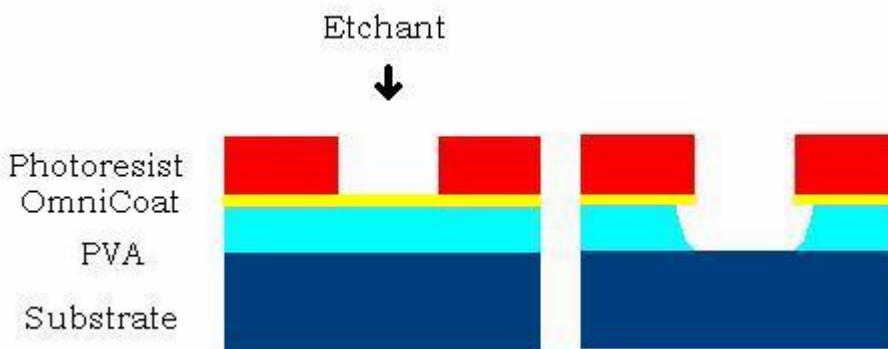
In order to create a layer of micropolarizers, the pattern must be printed on top of the PVA film using photolithography. First the PVA sample is cleaned with acetone to eliminate dust particles on the surface. A thin layer of a chemical called OmniCoat is then added to the surface of the PVA film. The OmniCoat is applied to the surface using a pipette, then accelerated to 500 revolutions per minute (RPM) over 5 seconds, then accelerated to 2000 RPM in 5 seconds and spun at 2000 RPM for 20 seconds. The OmniCoat serves two purposes. It improves the adhesion of the final photoresist structures to the PVA film by providing a better surface. OmniCoat also allows the photoresist structures to be removed more easily once the PVA film itself has been patterned. Next the sample is heated to 110° C for a period of 10 minutes to remove water from the PVA and to harden the OmniCoat layer. A thin layer of a substance known as a negative photoresist is deposited on top of the PVA; a negative photoresist is a substance that hardens when exposed to UV light, but can be easily removed in unexposed areas. A uniform layer of photoresist is made using spin coating. Photoresist is first added to the sample using a pipette. The sample is then spun at 500 revolutions per minute (RPM) for 10 seconds, then accelerated to 3000 RPM over 5 seconds, and spun at 3000 RPM for 45 seconds. The photoresist used was SU-8 2015, and based on the spin coating procedures, the SU-8 layer should be between 15 and 20 microns thick. Next the areas where the polarizers should be need to be exposed to UV light. This is done by covering with a mask the areas where the SU-8 should be removed, and then exposing the sample. A Karl Suss MA4 mask aligner was used for this purpose. The sample is exposed to 365 nm UV light at 8 mW/cm<sup>2</sup> for 25 seconds. The sample is then developed, removing the unexposed SU-8 and leaving the desired structures. Figure 4 below shows this process.



**Figure 4: Photolithography step to define the pattern of the PVA to be etched.**

### 3.3 Etching

Following lithography, there are SU-8 structures on top of the polarizing film; the next step is to create the micropolarizers themselves. This is done by removing the PVA in the areas not covered by photoresist in a process known as etching. Figure 5 below shows the etching process.



**Figure 5: Etching process illustrating the concept of underetching**

As shown, a substance known as an etchant is used to eat through the PVA, without attacking the photoresist layer. Depending on the etching procedure, a variety of etchant can be used. One common form of etching called wet etching uses a chemical such as an acid to remove the desired areas. This is not possible in this case, because wet etching removes the layer equally in both the horizontal and vertical directions. Since the PVA layer is 20 microns thick, and the desired structures are 10 micron, the structures would be destroyed in this process. An alternative technique called plasma etching uses high energy plasma to react with and remove the desired areas. This technique allows for sharper side walls, although there will be some degree of etching underneath the SU 8 structures, as shown in Figure 5. The OmniCoat is first developed

(removed in the areas not covered by SU-8) by putting the sample in oxygen plasma for 90 seconds using a Technics PlasmaEtch II machine. The etching of the PVA is done with a mixture of oxygen and  $\text{CF}_4$  in a three-to-one ratio using the same instrument. Another etching method being considered is reactive ion etching (RIE). RIE combines the oxygen plasma, which chemically reacts to etch away the PVA, with an equal concentration of high energy argon particles. Since argon is inert, it physically removes the PVA by hitting it at high speeds. This can be done by using a combination of oxygen and argon in the plasma etching machine.

### 3.4 Optical Properties

Initially, work was done with mask with a 15 micron circular pattern to develop the lithography and etching procedure and evaluate the resulting micropolarizers. Figure 6 below shows three images from a single micropolarizer as the incoming light shifts from being orthogonal to the polarizer to being parallel to the polarity of the polarizer.

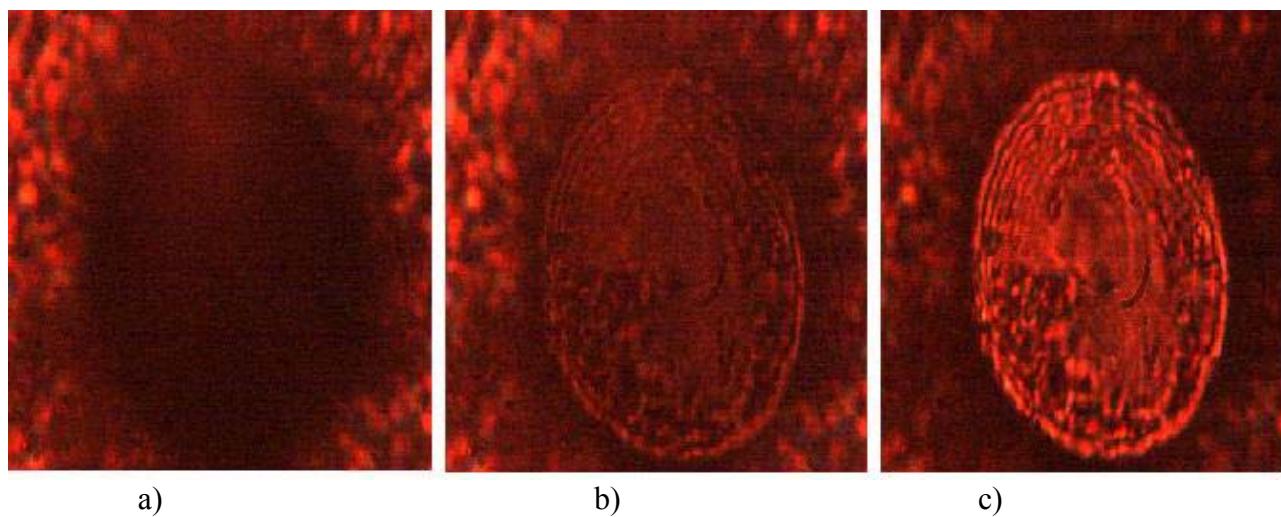


Figure 6: A single micropolarizer a) when incoming light is polarized at 90 degrees to the polarity of the polarizer, b) with incoming light at 45 degrees, c) with incoming light at 0 degrees

As shown in 6(a), when the light is orthogonal to the micropolarizer, very little light passes and the micropolarizer appears dark. At 45 degrees to the polarity of the micropolarizer, seen in 6(b), some light is able to pass, and the polarizer appears only slightly less bright than the background. In 6(c), the polarizer is at the same orientation as the incoming light, and the micropolarizer appears transparent. Figure 7 below shows the amount of light passing for each degree of polarization.

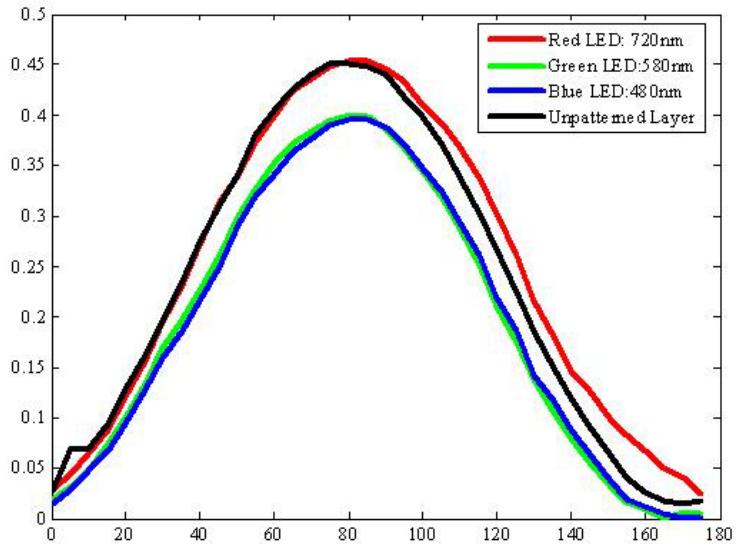
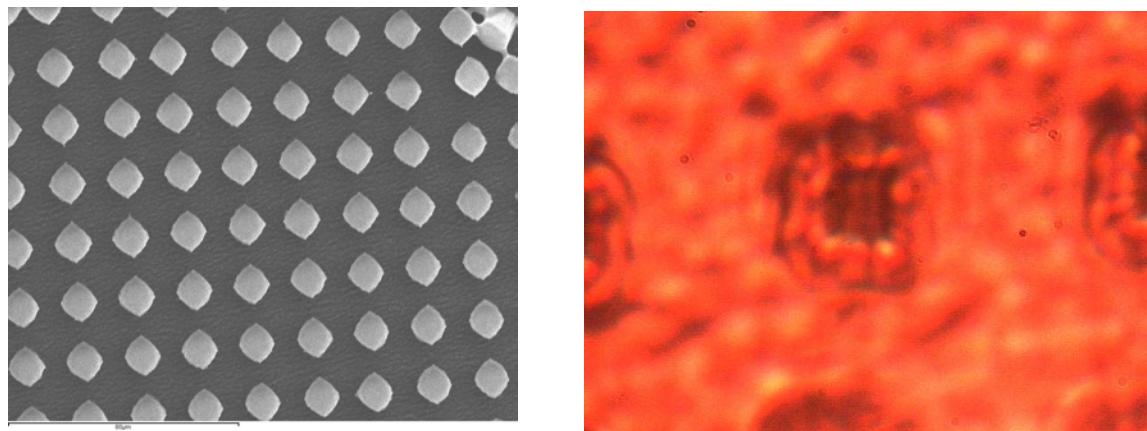


Figure 7: Graph of transmission percentage vs. polarization angle of light source

As indicated on Figure 7, the micropolarizer passes about 40% of light when aligned with the polarization angle of the light source, and less than 0.1% when orthogonal to that angle for both green and blue light. Longer wavelengths, such as red, pass at about 45% when aligned, and 3% when orthogonal. Hence, the extension ratio, defined as the ratio of the maximum to the minimum light intensity, is 100 for red light and 1000 for blue light. The black dotted line on the graph shows the characteristics of the unpatterned PVA film. We can conclude that the micropolarizer structures retain the polarization properties of the original unpatterned polarization filter.

### 3.5 Plasma Etching vs. RIE

A second layer of micropolarizers was created using a mask that contains the pattern that will be needed for the final sensor. The micropolarizers were etched using the plasma etching technique. Figure 8 below shows this micropolarizer array.



a) b)

Figure 8: Micropolarizers fabricated using plasma etching. a) A section of the micropolarizer array taken using a scanning electron microscope (SEM). b) A single micropolarizer at higher magnification using an optical microscope.

Figure 8a shows the structures created on the PVA imaged under a SEM. Every other space in each row and column is covered by a micropolarizer, as desired. Figure 8b shows a single micropolarizer at higher magnification imaged under an optical microscope. The larger transparent square is the SU-8 structure on top of the micropolarizer; the smaller dark square is the region that is actually polarized. A large percentage of the original pattern on the mask is lost due to underetching, approximately 3 microns on each side of the 12 micron structure. One possible method to avoid this problem is to compensate by creating a mask with larger features. The alternative is to switch to an alternative etching method such as reactive ion etching. Figure 9 below shows similar micropolarizers created using RIE.

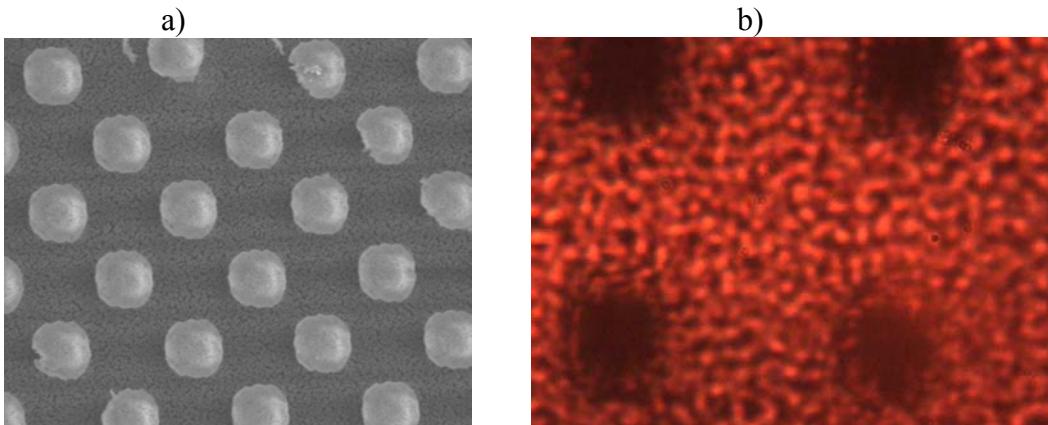
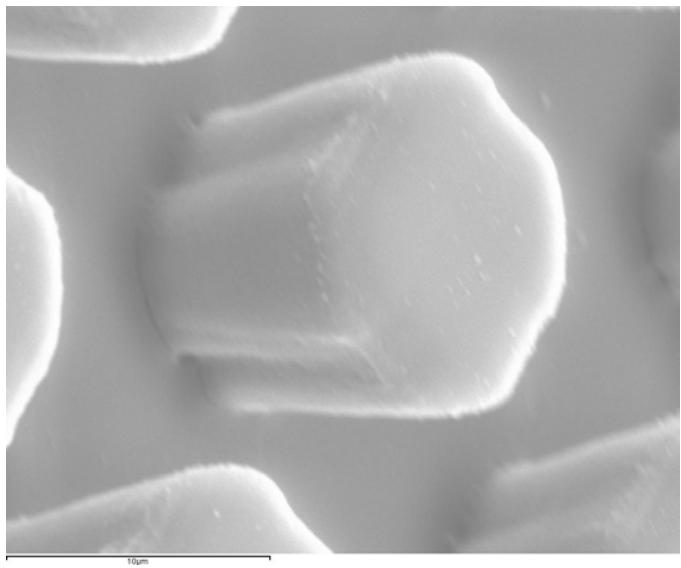
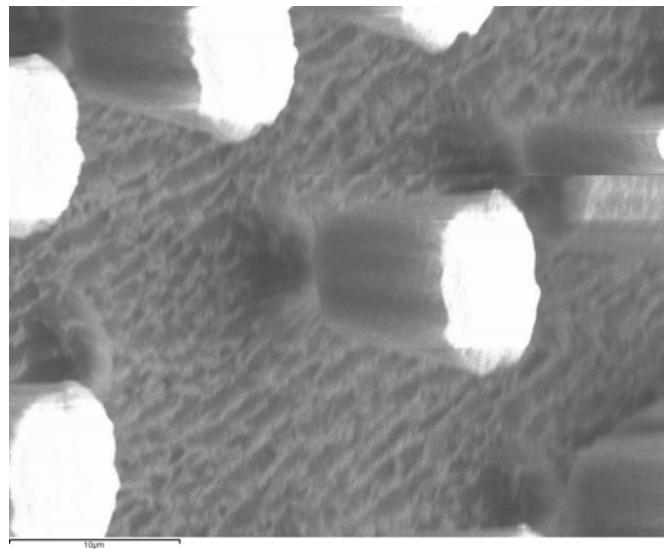


Figure 9: Micropolarizers fabricated using plasma etching. a) A section of the micropolarizer array taken using an SEM. b) The array at higher magnification taken through an optical microscope.

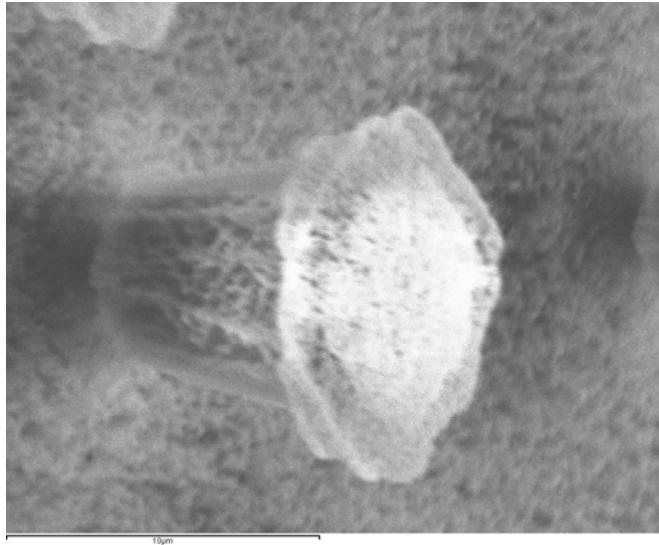
Figure 9a shows the structures created on the PVA. Figure 9b shows a smaller group of the micropolarizers at higher magnification; the dark polarized area is much larger relative to the size of the overall structure. To better evaluate the level of underetching in each case, the structures were examined using a scanning electron microscope, as shown in Figure 10.



a)



b)



c)

Figure 10: A single structure a) before etching, b) with plasma etching, and c) with reactive ion etching. The scale bar in each case is 10 microns.

Figure 10a shows a single structure after lithography; the height of the structure was measured to be approximately 14 microns, with a diameter of approximately 12 microns. 10b is a single micropolarizer etched using the plasma etching technique. The large structure on top is the SU-8, which is resistant to the plasma and remains mostly intact. The actual micropolarizer underneath is much smaller, as shown by the small root-like structure underneath the SU-8. 10c is a micropolarizer created using RIE; the SU-8 structure is nearly destroyed, since the argon attacks it as well as the PVA. However, the actual micropolarizer is nearly as large as the original SU-8 structure; the underetching is much smaller, approximately one micron on each side. The grainy appearance of the structure in 10c is due to the RIE; the argon tends to damage the surface of the structures. This may degrade the optical properties; this is still being evaluated.

## **4. DOUBLE LAYER MICROPOLARIZER ARRAYS**

### **4.1 Methods for Gluing Multiple Layers**

In order to create the required micropolarizer array, a method to create micropolarizers directly on the image sensor is required. Previously the micropolarizers were created using the CAB layer as backing material. Since the CAB layer is 300 microns thick, this layer would decrease the light transmission rate as well as introduce undesirable optical distortions. As a result, we would like the PVA to be glued directly onto the image sensor, allowing the removal of CAB using acetone.

Initially this was attempted using SU-8, our photoresist, as an adhesive. SU-8 is an epoxy based chemical, similar to many glues, and will stick to PVA. It will also adhere to the SiO<sub>2</sub> that forms the top surface of the CMOS image sensor, based on experimentation using glass. However, gluing with SU-8 requires that the solvents within the SU-8 evaporate, allowing the photoresist to harden. In this case, the SU-8 is sandwiched between the chip and the PVA, limiting the evaporation and leading to a poor bond between the layers.

To avoid this, the process was repeated using an optical glue from Dymax called OP-30. OP-30 is a UV curing adhesive that contains no solvents; it hardens completely upon exposure for a short period of time to UV light. This allowed the sample to be glued by administering the glue, placing it under a glass plate, and exposing it using the mask aligner. The mask aligner also allowed the application of pressure to press out air bubbles in the glue. Upon attempting to remove the CAB layer, however, the acetone tended to attack the OP-30, destroying the bond. A second UV curing glue, Loctite 349, was then tested and determined to have similar difficulties with acetone. Since the acetone is only able to attack the small boundary areas of the glue, it may be possible to overcome this problem by protecting these boundaries with chemical resistant tape or by using large samples and cutting out the areas lost due to acetone. An acetone resistant two part epoxy, Loctite E30CL, was tested as well, but was found to have difficulties evaporating solvents similar to the problems with the SU-8.

### **4.2 Aligning Multiple Layers**

For the micropolarizer array to be useful, the micropolarizer on each layer need to be placed in the proper locations relative to each other and the image sensor. This can be done using the optics on the mask aligner, but some points of reference on each layer are required. This is traditionally done using elements known as alignment markers. Figure 11 below shows a basic alignment marker system.

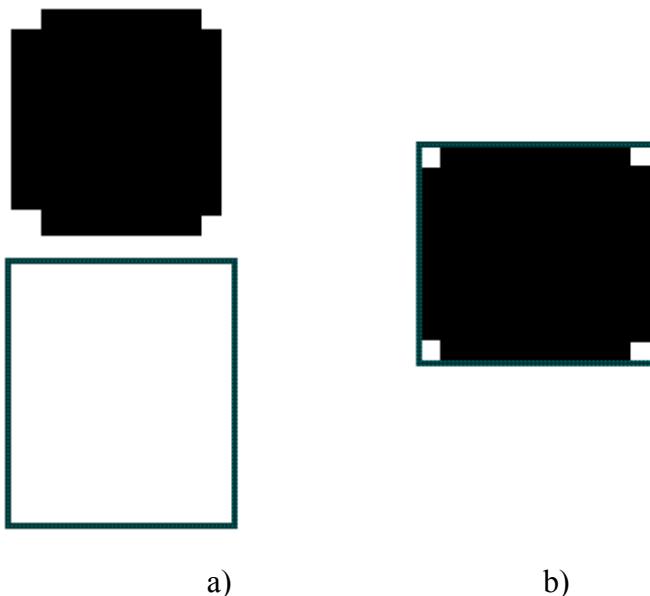


Figure 11: Traditional alignment markers. a) The shapes of two simple alignment markers. b) The appearance of the markers when brought into the proper alignment

As shown, the traditional alignment marker system is a cross and a square. When placing a layer, the person aligning places the cross within the box created by the square. The cross and square system work particularly well because it allows the person aligning to not only place a point on the two layers relative to each other, but also enables the person to achieve the proper angle of rotation to achieve the proper alignment. Our image sensor does not contain any specific alignment markers, but the large square input and output pads can be used for that purpose. The alignment marks on the final mask will have to be placed to allow this use.

#### 4.3 SU-8 Removal

Although SU-8 is mostly transparent, the 15-20 micron thick photoresist structures tend to attenuate light to some degree. As a result, the photoresist should be removed after etching the PVA to create the micropolarizers. SU-8 generally bonds strongly after photolithography but the layer of OmniCoat does allow its removal. The samples can be dipped in an 85° C bath of a chemical called Remover PG for 30 minutes to remove the photoresist. This procedure has been used to remove SU-8 structures successfully from a glass slide. Since Remover PG attacks the CAB substrate of the PVA structures, this has not been completed with an actual micropolarizer array. Another possibility is to remove the SU-8 during etching. In reactive ion etching, the argon attacks the SU-8, as shown in Figure 10c. By changing the spin acceleration to change the thickness of photoresist, this can be used to eliminate the SU-8 as well. The CF<sub>4</sub> and oxygen mixture does not attack the SU-8, as shown in Figure 10b, so this would not be possible for plasma etching.

#### 5. DISCUSSION AND CONCLUSIONS

Polarization image sensors are an exciting new development with numerous possible applications. Thin film micropolarizer arrays will be necessary to make polarization sensors

practical. An effective technique for fabricating a single layer of 10-micron micropolarizers in thin polyvinyl alcohol film has been presented. The micropolarizers were demonstrated to have a transmission rate similar to that of the original polarizing film; they can, therefore, be used effectively in a polarization image sensor. Two different etching techniques, plasma etching and reactive ion etching, are evaluated. Plasma etching was found to cause a large degree of etching underneath the photoresist structures, degrading the quality of the micropolarizers formed. Reactive ion etching, however, was found to cause significantly less underetching, maintaining most of the 10 micron structures formed in the photoresist. Reactive ion etching was also found to remove most of the photoresist from the polarizer, unlike the plasma etching technique. As a result, the reactive ion etching will be used in the final polarization sensor. This paper also presents preliminary work towards fabrication of a multiple layer micropolarizer array.

## 6. RECOMMENDATIONS

Microfabrication is a delicate procedure that is extremely sensitive to environmental conditions. Many of the difficulties encountered can be mitigated or even eliminated with better-controlled temperature and lower humidity. As a result, my primary recommendation is to repeat some of the more promising efforts outside of the warm summer months; this has great potential to help improve our results. Future work should be devoted towards expanding on the efforts toward creating a multi-layer micropolarizer array. A new mask should be designed and fabricated using the proper alignment markers so that it can be used for multiple layers.

## 7. ACKNOWLEDGMENTS

I would like to thank my advisors, Dr. Jan Van der Spiegel and Dr. Viktor Gruev, for entrusting me with this project and aiding my efforts. I also would like to thank the other students of the Analog VLSI lab and Scott Slavin of the Microfabrication Lab for their assistance. I would like to thank the National Science Foundation for their REU grant, and the School of Engineering for their additional funding through the ROPE fellowship. Finally, I would like to thank the other ESE staff members who make SUNFEST possible.

## 8. REFERENCES

1. K. Wu, Fabrication of Micro-Polarizer Array with Polymer Thin Film, Technical Report, TR- CST01DEC05, pp. 201-219, 2005
2. W. McMaster, Polarization and the Stokes Parameters, *American Journal of Physics* 22 (1954), 351-362
3. V. Gruev, K. Wu, J. Van der Spiegel, and N. Engheta, Real time extraction of polarimetric information at the focal plane, *Proc. Of SPIE Polarization:Measurement, Analysis, and Remote Sensing VII, Orlando, FL, USA, May 2006*
4. L. Wolff and T. Boult, Constraining Object Features Using a Polarization Reflectance Model, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13 (1991), 635-657

5. L. Wolff and A. Andreou, Polarization Camera Sensors, *Image and Vision Computing* 13(1995), 497-510
6. A. G. Andreou and Z. K. Kalayjian, Polarization Imaging: Principles and Integrated Polarimeters, *IEEE Sensors Journal*, 2 (2002), 566-576
7. M. Momeni and A. H. Titus, An analog VLSI chip emulating polarization vision of octopus retina, *Neural Networks, IEEE Transactions on* 17 (2006)., 222-232
8. T. Clausnitzer, H. Fuchs, E. Kley, A. Tuennemann, U. Zeitner, Polarizing metal stripe gratings for a micro optical polarimeter, *Proc. Of SPIE Lithographic and Micromachining Techniques for Optical Component Fabrication II, San Diego, CA, USA, August 2003* 8-15
9. M. Saito and M. Miyagi, Micropolarizer using anodized alumina with implanted metallic columns: theoretical analysis, *Applied Optics*, 28 (1989) 3529-3533
10. Y. Chang, H. Tien, C. Sung, C. Lee, C. Wang, J. Chang, Micropolarizer fabricated from CaCO<sub>3</sub> by anisotropic wet etching, *Applied Optics* 42 (2003), 4423-4426
11. C. K. Harnett and H. G. Craighead, Liquid-crystal micropolarizer array for polarization difference imaging, *Applied Optics*, 41 (2002) 1291-1296
12. J. Guo and D. Brady, Fabrication of thin-film micropolarizer arrays for visible imaging polarimetry, *Applied Optics* 39 (2000), 1486-1492
13. J. Guo and D. Brady, Fabrication of high-resolution micropolarizer arrays, *Opt. Eng.* 36(1997), 2268-2271

# PEDIATRIC DYNAMOMETER

NSF Summer Undergraduate Fellowship in Sensor Technologies  
Armand O'Donnell (Electrical Engineering) – University of Pennsylvania  
Advisor: Dr. Jay Zemel

## ABSTRACT

Weight-bearing activity has been demonstrated to benefit childhood bone development<sup>[1]</sup>. To further investigate this connection, the Growth and Development Laboratory at the Children's Hospital of Philadelphia wishes to study the effects of strenuous activity among school-age children. Acquiring quantitative information about the magnitude and duration of force on children's feet has posed a challenge, as surveys serve only to collect qualitative information about exercise. While the equipment necessary to perform human kinetic analyses exists, devices used for directly monitoring the activities of a child can be large, awkward, and obtrusive. A small, inconspicuous, self-contained mobile device to collect and store data about the physical activity of the individual wearing it would facilitate medical research in pediatric bone health studies.

The data acquisition device proposed by Dr. Jay Zemel of the Moore School of Electrical engineering is a small, battery-powered, in-shoe physical activity dynamometer (Foot-PAD). It uses a strip of piezoelectric plastic as a force sensor along with circuitry capable of taking measurements and storing data over a period of three to fourteen days. After this period, information can be uploaded from the dynamometer to a computer, at which point the data can be analyzed or stored in an archive for later analysis.

Earlier projects established the feasibility of using a piezoelectric polyvinylidene fluoride (PVDF) sensor for the dynamometer and presented a small but non-operational design. As a result of this summer's work, the pediatric dynamometer has been programmed to sample the piezoelectric sensor, make simple calculations based on these readings, store relevant data on a small flash memory chip, and extract this information, uploading it to a personal computer after the acquisition has finished. The hardware design has also been optimized to take up as little space as possible and consume minimal power.

## Table of Contents

<b>1. INTRODUCTION.....</b>	<b>144</b>
1.1. Background.....	144
1.2 Preceding Work .....	144
1.3 Goals of This Project .....	145
<b>2. Building the Pediatric Dynamometer.....</b>	<b>145</b>
2.1 Device Overview .....	145
2.2 PIC16F88 Flash Microcontroller .....	146
2.3 PVDF Sensor .....	146
2.4 Charge Amplifier .....	148
2.5 AUSART and Voltage Level Shifter (Serial Module).....	152
2.6 Synchronous Serial Communications with EEPROM/Flash .....	157
2.7 Integrating AUSART/SSP Components in Software.....	158
<b>3. Final Results .....</b>	<b>161</b>
3.1 Hardware Design Results .....	161
3.2 Software Design Results.....	165
<b>4. Conclusions .....</b>	<b>166</b>
<b>5. Acknowledgments .....</b>	<b>167</b>
<b>6. References.....</b>	<b>168</b>
<b>Appendix A. Recommendations for Future Work .....</b>	<b>169</b>
A.1 Digital Signal Regression .....	169
A.2 PC Software Component .....	175
<b>Appendix B. Source Code listing .....</b>	<b>176</b>

## 1. INTRODUCTION

### 1.1. ***Background***

The correlation of a child's exercise habits to bone density<sup>[1]</sup> has been a topic of recent interest. The Children's Hospital of Philadelphia (CHOP) is attempting to quantitatively document the kinetic activities of a child over the course of days or weeks, specifically the average and peak force on a child's foot. Measurements of force shall be taken continuously over a period of time. Surveys and questionnaires are common sources of data used medical research, but they are susceptible to sampling and reporting biases and are inherently inaccurate, particularly if data is collected from children or adolescents<sup>[2]</sup>. Due to these issues, CHOP cannot use surveys for this childhood bone health study; more reliable means are necessary.

If instrumentation is used to record childhood activities, it must be sensitive and designed for the specific application. According to an Iowa State University study<sup>[3]</sup>:

To accurately assess children's activity patterns, an instrument must be sensitive enough to detect, code, or record sporadic and intermittent activity. Care also must be used to select criterion measures that reflect appropriate physical activity guidelines for children.

The study also concluded that in studies involving children, there is an "accuracy/practicality" tradeoff such that more accurate devices are accordingly expensive and obtrusive. A device is needed that can accurately measure the force on a child's feet over the course of a day or longer without affecting the child's daily routine.

Portable devices have been patented for use in more general kinetic studies, mainly for calculating energy expenditure over the course of a day. These patents describe detailed methods for sensing force on the feet of an injured patient<sup>[4]</sup>, the use of a piezoelectric sensor to analyze human locomotion<sup>[5]</sup> and calculation of the metabolic expenditure of energy by placing a force-sensitive resistor, or strain gage, inside of a shoe<sup>[6]</sup>.

The Electrical Engineering department of the University of Pennsylvania is collaborating with CHOP to find a solution. Embedding a portable dynamometer inside the shoe has been proposed and is under development (see Preceding Work). Dr. B Zemel of CHOP suggests that it should not be noticeable or else attention will be drawn to the device, and the participant, a child, may not want to wear it. It is also necessary to store the data in a central location, preferably a personal computer, after the sampling has taken place. The device should be relatively inexpensive to fabricate so that multiple instances of the device can be used by different participants simultaneously. Once adequate data has been received by the hospital, it can be interpreted by specialists in the field of childhood bone development.

### 1.2 ***Preceding Work***

The pediatric dynamometer under development at Penn is fundamentally different from what has been done in that it is a completely self-contained unit. The embedded microprocessor combined with the sensor and memory allows data to be processed and recorded autonomously and locally, not relying on an external communications link to a storage device. The nature and sheer quantity of information this device is capable of recording may enable medical researchers to address other research issues beyond monitoring children's activity and calculating their energy expenditure.

Work on this project was started in the summer of 2004 by Sunfest fellow Olivia Tsai<sup>[7]</sup>. Since then, it has progressed as a senior-year design project during the 2004-05<sup>[8]</sup> and 2005-06<sup>[9]</sup> academic years. The dynamometer has evolved from a conceptual plan to a functioning wireless data acquisition device worn attached to the leg to a two-square-inch small circuit that fits inside the heel of a young child's shoe with room to spare.

### **1.3 Goals of This Project**

As of the start of the Sunfest project in May 2006, a hardware design for the pediatric dynamometer had been established and most of the details for its design and operation were close to implementation. The following conditions needed to be met before this device could be utilized by volunteer children:

1. The microcontroller used by this device requires a program that will oversee sensor sampling, signal processing (if necessary), documentation of results to an onboard memory chip, and serial communications with a personal computer for data retrieval and archival.
2. The existing hardware layout of the pediatric dynamometer was a generic design that would have worked, but could be streamlined to include fewer components. Furthermore, the hardware layout must be consistent with the software algorithms used in (1).
3. The pediatric dynamometer must be tested under real-world conditions inside a shoe, which should include exposing the sensor and circuits to moisture, mechanical stress, and a range of temperatures.
4. A simple software front-end was needed to easily interface the device to the communications port of a PC in a format readily accessible to the researchers' existing software.

Programming the device and finalizing its layout were the highest priorities for the 2006 Sunfest project. With a solid software foundation and a reliable hardware backbone, the pediatric dynamometer could be reconfigured and reprogrammed to record the desired data in flash memory. The optimal sampling rate could then be decided upon. An important goal was that the microcontroller's software could accommodate modified algorithms for extracting useful information from the sensor's output.

## **2. Building the Pediatric Dynamometer**

### **2.1 Device Overview**

In order to meet the pediatric dynamometer's size, storage capacity, and power consumption constraints, a number of design decisions have been made over the course of this project's development. The heart of the pediatric dynamometer is the PIC16F88 flash microcontroller. This device oversees the interaction of the other components, makes calculations, and performs some simple data interpretation. The polyvinylidene diflouride (PVDF) thin film sensor's output is proportional to the force on the sole of a shoe. The sensor's output is conditioned by a charge amplifier circuit connected to the analog-to-digital converter inside of the PIC. After processing, the resulting information is recorded on a 16 megabit serial EEPROM that was chosen for its extremely small size and relatively high storage capacity.

Each of the device modules and associated software are described in detail with explanations as to how each stage interacts with the others. An overview of the operation of the

completed device is then presented. A number of issues arose during the project that may require alternative solutions as discussed in section 5, Recommendations for Future Work.

## **2.2 PIC16F88 Flash Microcontroller**

Microcontrollers, the backbones of “embedded systems,” have become ubiquitous over the past few decades. As processors they provide appreciable computing power and can execute millions of instructions per second. As fabrication technology has improved, the microcontroller’s voltage and power requirements have decreased so that they can run for extended periods of time off of 3.3V button-cell battery power. The PIC and all of the other devices in the pediatric dynamometer are powered by a 3.3V button cell battery. Most importantly, microcontrollers offer a wealth of features, including general digital inputs and outputs, analog-to-digital converters, serial interfaces, and timers.

The PIC series of microcontrollers offers a line of units which best meet the needs of the pediatric dynamometer. The PIC16F88 was chosen because of its low power consumption, small size, and availability of features: it provides an asynchronous serial port for interfacing with a PC, and a synchronous serial port used to read from and write to a small serial EEPROM (electrically-erasable read-only memory) chip. There is a multiplexed analog-to-digital converter on this PIC, essential for sampling information from the sensor. In addition, the PIC16F88 has an internal clock that removes the need for a timing crystal, saving space on our circuit board.

PIC microcontrollers typically contain EEPROM (electrically erasable programmable read-only memory) or flash memory in which the program instructions are stored. The PIC is compatible with in-circuit serial programming, which allows users to program and debug the PIC in a prototype dynamometer. In-circuit programming simplifies the prototyping phase by eliminating the external programming of the PIC, which requires removing the PIC from the circuit.

Finally, and of significant importance, is the PIC’s low cost. Since the childhood activity study would involve dozens or even hundreds of volunteers, it is desirable to create the device from inexpensive parts. If purchased directly from Microchip™, one PIC16F88 costs between two and three dollars, depending on the quantity ordered.

One of the surprises encountered when working with the PIC over the course of the design process was the fact that some of the microcontroller’s features frequently caused unexpected results. The watchdog timer, which automatically resets the microcontroller after a specified number of instruction cycles, is enabled by default and must be explicitly disabled if its operation is not desired. The 16F88 employs a low-voltage brown-out detect that automatically switches the device off when the supply voltage drops below 4V, which naturally presented a problem when working with the 3.3V lithium battery voltage.

## **2.3 PVDF Sensor**

Measuring the vertical force of a foot inside a shoe with a PVDF sensor had been established by previous groups<sup>[8, 9]</sup>. Numerous specific experiments were carried out on these sensors under a variety of circumstances.

The principle behind these sensors is piezoelectricity. PVDF is a piezoelectric thin-film polymer. When the sensor is bent, charges accumulate on the surfaces, inducing current in the

external circuit. Figure 1 shows the basic operation of the PVDF sensor in response to a force in the horizontal direction.

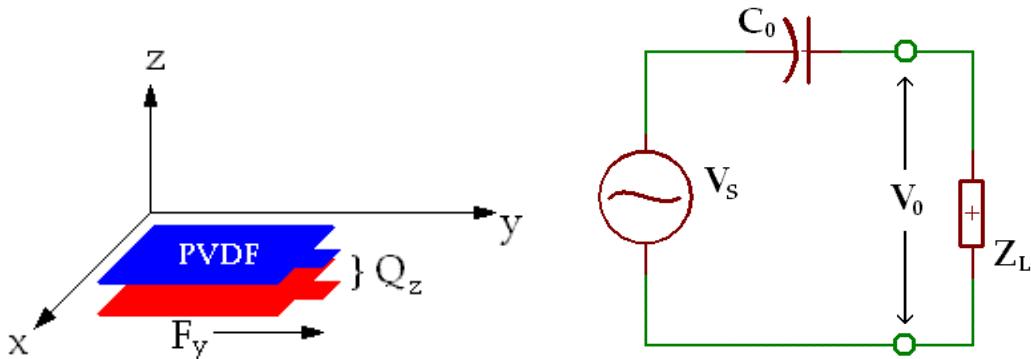


Figure 1a (left): PVDF Sensor Force-Charge Relation  
1b (right) PVDF Equivalent Circuit<sup>[11]</sup>

The linear relationship governing this charge displacement is:  $Q_z = d_{zy}F_y$  where  $Q_z$  is the charge displaced across the planar surface,  $d_{zy}$  is the longitudinal piezo strain coefficient, and  $F_y$  is the force along the y-axis. The equivalent circuit<sup>[11]</sup> shows a voltage source  $V_s$  in series with the sensor's capacitance  $C_0$ . The external circuit is the load,  $Z_L$ . For further reference, see [10].

In order to be used with other circuitry, the PVDF is coated with a thin layer of conducting material, typically silver, and is then laminated with a protective plastic coating. A pair of terminals is attached for connection to the external circuit. Charge induced in the external circuit during an instant of time is fundamentally expressed as the definition of current:  $dQ/dt = I$ . The behavior of the PVDF sensor in a circuit analysis context is as follows: any change in the deflection of the device results in an immediate current flow through the load  $Z_L$ . At any period of time when the sensor is stationary the net current through  $Z_L$  is zero.

Earlier projects done on the pediatric dynamometer investigated the PVDF sensor's response to mechanical stress. It was shown that the most satisfactory mechanical arrangement for the PVDF sensor inside a shoe was to mount it on a cantilever of spring steel<sup>[8a]</sup>. A series of calibration experiments were made by placing standard weights on top of the sensor one after the other in set increments and viewing the output of the PVDF sensor in each case. For deflection within the bounds of normal human locomotion, it was found that the amount of charge displaced was proportional to the vertical force on the sensor<sup>[8b]</sup>.

The pediatric dynamometer's primary goal is to record *force*. The piezoelectric sensors' response to a change in applied force produces a temporary current so long as the force is changing, and not over the entire duration of the force. The force is obtained from the charge displaced as a function of time, not just at a single instant. Mathematically, the force is proportional to the time integration of the signal.

Integration can be done in two ways: using the microcontroller's ADC to sample the current and store digital signals in an accumulator, and analog integration with a charge amplifier. Software in the PIC provides successive sampling of the current flow in and out of the sensor. The accumulator value is then a digital representation of the total charge displaced by the sensor. This method produced promising results, as the sample data presented both the original PVDF output alongside the integration of the data. Software integration has two drawbacks: analog device offset and sampling rate. Signal offset is added to the sensor output causing the

accumulator to introduce an error each sampling cycle. Over time, the offset accumulation introduces a drift, potentially rendering the data useless. The sampling rate must also be sufficiently high or else data may be lost if the force changes suddenly.

Analog integration of the PVDF sensor signal is an alternate and widely used approach. One form of analog integrator, called a charge amplifier, was investigated this summer and was chosen for use in the pediatric dynamometer. The next section provides an in-depth explanation of the charge amplifier.

The PVDF sensor also responds to a change in temperature, the pyroelectric effect. In the dynamometer application, the low-frequency signal introduced by the pyroelectric effect is unwanted. When a test subject first puts on a shoe, the inside of the shoe (particularly the sensor inside it) will warm up from room temperature to near body temperature. This introduces a slow drift in the input signal that can be reduced by adding a charge-leakage resistor to the charge amplifier as discussed in the next section.

## 2.4 Charge Amplifier

The signal current from the PVDF sensor can be integrated continuously with a charge amplifier (CA) to provide a voltage proportional to the charge displaced in the PVDF sensor. The CA initially examined is composed of three stages, each controlled by an operational amplifier. The Electronics Workbench schematic for the CA is shown in Figure 2.

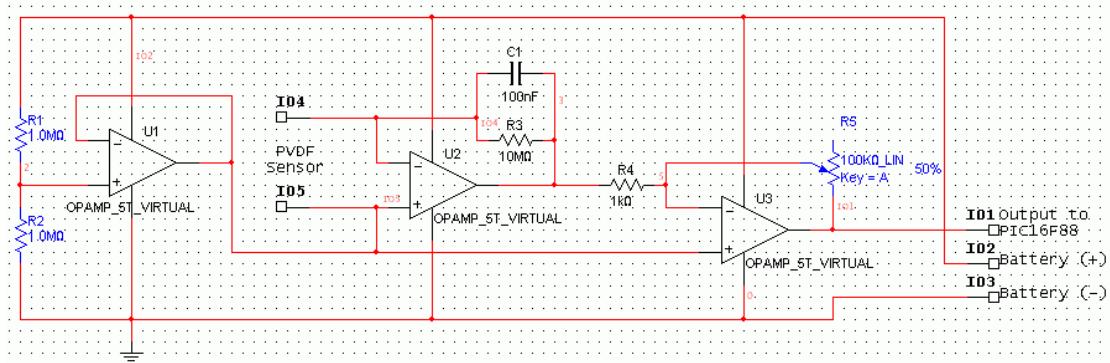


Figure 2: Charge Amplifier Prototype—Electronics Workbench™

The first stage provides a voltage reference of half the source voltage for the rest of the circuit. The PVDF sensor moves current in both directions, resulting in positive and negative voltages across its terminals. Since the entire system is powered from a single source, a button-cell battery, the op-amp's input and output range is from 0V to the source voltage. An offset is needed for negative voltages to be represented. A logical choice is the half-source reference so that an equal amplitude signal can be represented above and below the reference without saturating the amplifier (reaching the max or min voltage of the amplifier).

The second stage is the charge amplifier, which integrates the PVDF output signal connected between IO4 and IO5. Referring to the voltage at IO5 as  $V_{ref}$ , the voltage across C1 as  $V_C$ , and the output voltage of the amplifier U2 as  $V_{CA}$ , and applying the negative feedback condition of the op-amp ( $V_{IO4} = V_{IO5}$ ) results in the following relationship:  $V_{CA} = V_{ref} + V_C$ . The capacitor voltage  $V_C$  depends on the charge on it:  $V_C = Q_C/C$ . Differentiating with respect to time yields:

$$\frac{dV_C}{dt} = \frac{1}{C} \frac{dQ_C}{dt} = \frac{1}{C} i_C(t)$$

Integrating with respect to time yields:

$$\int \frac{dV_C}{dt} dt = \int \frac{1}{C} i_C(t) dt \Rightarrow V_C = \frac{1}{C} \int_0^{t_0} i_C(t) dt$$

where the CA is powered at time 0 (such that  $V_C = 0$ ), and  $t_0$  is the time of measurement. The current induced by the PVDF sensor is balanced by the op-amp, which outputs a voltage that cancels out the capacitor voltage. **The output voltage of the charge amplifier with respect to the reference voltage is proportional to the charge on the feedback capacitor.** A resistor is connected in parallel with the feedback capacitor to slowly discharge it. This attenuates the integration of very low-frequency signals, but preserves the integration of desired signals within the 1-5 Hz range.

The third stage is a simple amplifier with variable gain. The third stage was not used in the final design but helped during prototyping because it allowed the gain to be adjusted as needed.

A photograph of the prototype built to test the CA's operation is shown in Figure 3. Note that C1 and R3 can be replaced easily, as the prototype contains a 4-pin socket so we can experiment with different R-C pairs. The potentiometer, shown on the right of Figure 2 and the bottom of Figure 3, is used to adjust the gain of the output stage.

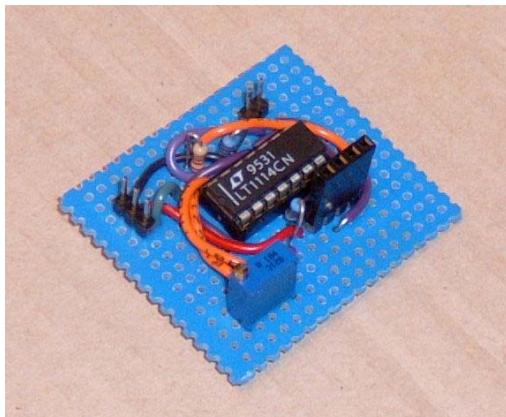


Figure 3: Charge Amplifier Prototype

The charge amplifier was tested with a PVDF sensor placed inside of the heel of a shoe. The CA does not integrate the input signal from the PVDF element appreciably for low values of R (below  $10M\Omega$ ) or C (below  $10nF$ ). When R or C are low, the charge is removed from the capacitor quickly, and the integration is attenuated. Controlled, even steps were taken to test the effectiveness of the CA. The CA cannot integrate without a feedback capacitor. Even with a feedback capacitor, charge leaks off very quickly when a  $1M\Omega$  resistor is used, canceling out the integration.

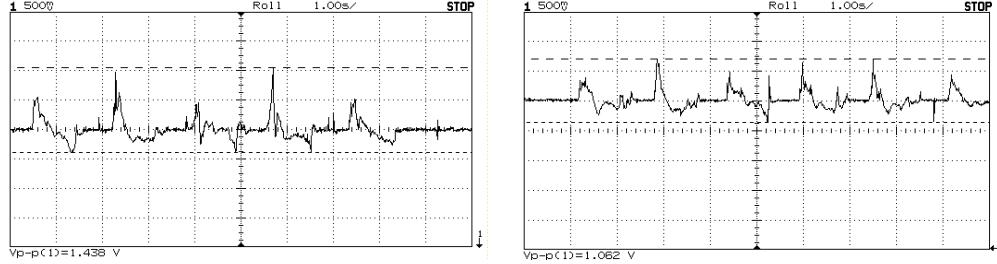


Figure 4 (left): CA using a  $1\text{M}\Omega$  resistor and no capacitor

Figure 5 (right): CA using a  $1\text{M}\Omega$  resistor and  $100\text{pF}$  capacitor

A larger capacitor is able to retain charge for a longer amount of time. However, the amplitude of the integration is lower with large capacitors because more charge needs to be displaced to attain the same voltage increase ( $Q=CV$ ).

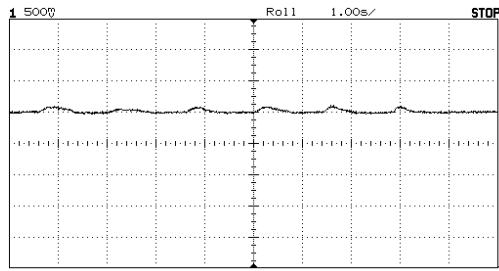


Figure 6: CA using  $1\text{M}\Omega$  resistor and  $1\mu\text{F}$  capacitor

Using a  $10\text{M}\Omega$  resistor reduces the charge leak from the capacitor, but not completely. Taking short steps with the PVDF introduces and removes charge quickly enough so that the  $10\text{M}\Omega$  doesn't draw substantial charge from the capacitor. For long steps, however, there is noticeable charge leakage, and on the up-step, the counter-charge results in an offset that takes a while to restore.

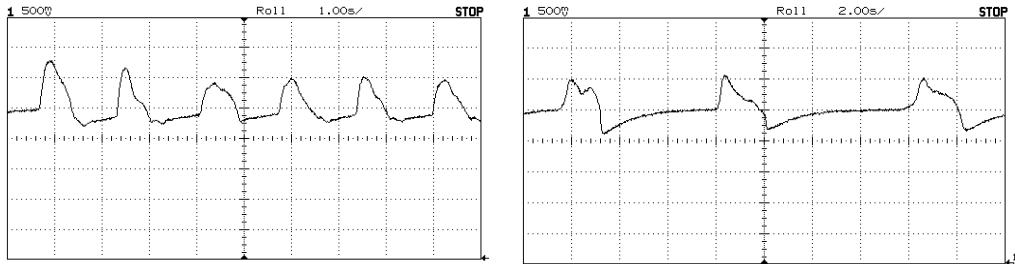


Figure 7 (left): CA using  $10\text{M}\Omega$  resistor and  $100\text{nF}$  capacitor—short steps

Figure 8 (right): Same CA feedback network—longer steps (note 2s/div)

Omitting the resistor altogether yields surprisingly clean integration.

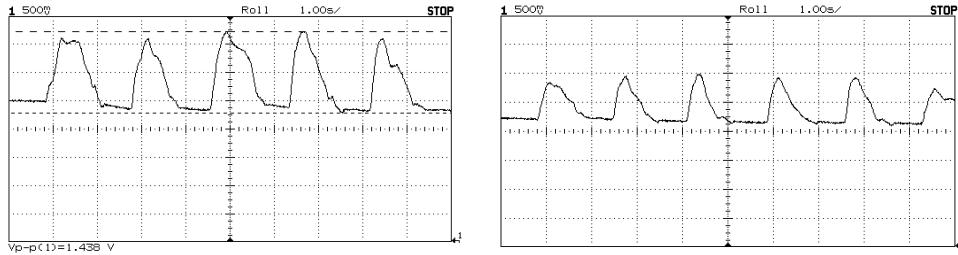


Figure 9 (left): CA using a 47nF capacitor, and without a resistor.  
 Figure 10 (right): CA using a 100nF capacitor, and without a resistor.

Also, care must be taken not to choose too small a capacitance. If this happens, the op-amp may drive the CA into saturation, since for a small capacitor, little charge exchange is needed for a large voltage swing.

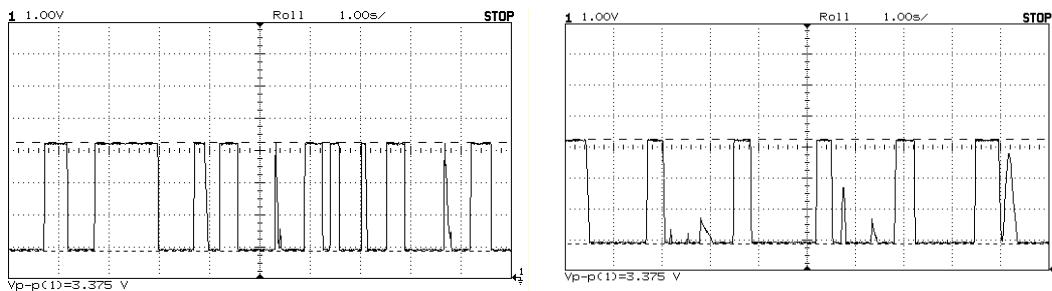


Figure 11 (left): CA using a 100pF capacitor.  
 Figure 12 (right): CA using a 1nF capacitor.

Another problem lies with omitting the shunt resistor across the capacitor. Although no drift due to mechanical operation of the PVDF sensor was detected, pyroelectric effects caused the CA to drift considerably from the half-source voltage, and at times even drove the CA into saturation. This effect is particularly noticeable when the user first puts on his/her shoe or takes it off, thereby introducing thermal variation as the sensor changes from room temperature to body temperature or vice versa.

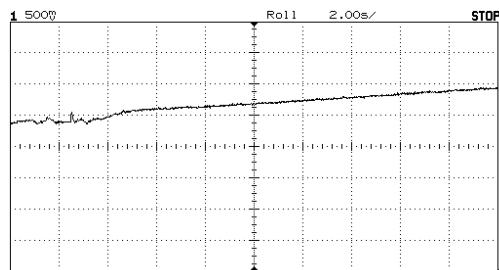


Figure 13: Removing sensor from shoe:  
 CA output, using no resistor and 100nF capacitor.

This effect is less problematic when a shunt resistor is connected across the capacitor. Since the pyroelectric effect is a low-frequency problem, a large RC time constant can be used to remove very low frequencies on the order of tenths of Hz. Using a  $10M\Omega$  resistor eliminates the CA's response to the pyroelectric effect.

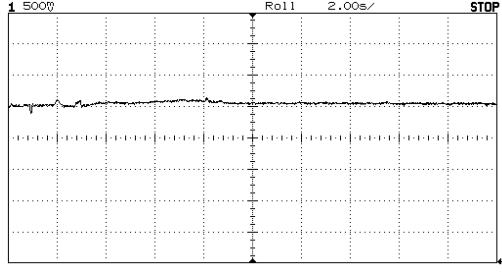


Figure 14: With a  $10M\Omega$  resistor in the feedback network, pyroelectric drift is attenuated.

Despite these low-frequency concerns, the charge amplifier was used for the final design because it simplified PIC measurements. The PIC sampled the output of the charge amplifier to obtain the force. Using a suitable high-impedance resistor, as much as  $50M\Omega$  - $200M\Omega$ , is an effective compromise. Such a high resistance only attenuates extremely low frequencies. It is also possible to use a smaller RC feedback circuit on the CA. Though this causes distortion on the signal, extracting the integration can be done in software by multiplying each previous sample by:

$$e^{-T_{\text{sample}}/RC}$$

and subtracting the difference from the next sample. This particular solution has been proven in MATLAB, and will be discussed more in-depth in section 5.

Of particular interest for childhood bone density studies are measurements of the peak force, the average force over the duration of a step, and the amount of time each step takes. In the case of a low-frequency offset, as long as the CA has not been driven into saturation, the peak-to-peak values are all that is needed to extract useful information about the step. This means data resembling Figures 7 and 8 may be useful after all.

## 2.5 AUSART and Voltage Level Shifter (Serial Module)

The PIC microcontroller can communicate with a personal computer using compatible signals, requiring only a voltage-level shifter circuit in between the PIC and computer. The AUSART, or Addressable Universal Synchronous/Asynchronous Receiver and Transmitter, is set up in software and configures two ports on the PIC, one as a receiver (Rx) and another as a transmitter (Tx). As long as the computer and PIC are configured to receive and transmit data at the same number of bits per second, the “baud rate,” no common clock or handshaking is necessary for normal two-way communications when the module is operating in asynchronous mode.

The standard for this asynchronous protocol is to hold the line at a high level until a byte is transmitted. This is often referred to as “idle high”. Each byte that is sent in either direction begins with a start bit, logic “0”, to notify the target device that a byte is coming. Then, at the pre-determined frequency, the bits to be transmitted appear on the line, least-significant bit first. A number of test programs were written to utilize the serial port along with other features of the device. The following example in Figure 15 demonstrates a pair of bytes, the second generated by the PIC’s analog-to-digital converter (ADC), being sent out of the serial port to the computer. A variable resistor connected to provide a range of voltages was connected to one of the PIC’s analog inputs. The ADC, which converts a voltage signal to a 10-bit binary number, was set to run continuously. After the analog-to-digital conversion completed, the most significant eight

bits were sent over the serial line. Having this example program readily available facilitated later use of both the serial port and ADC.

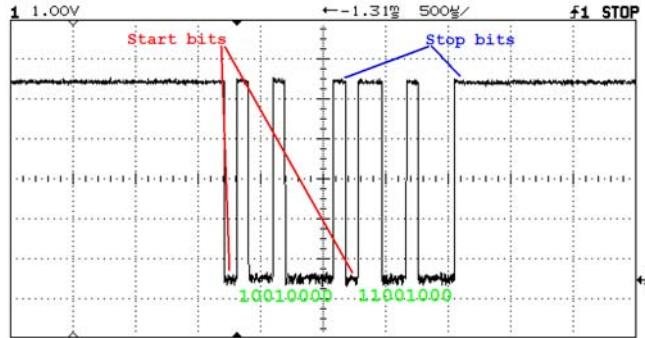


Figure 15: Sending signals 0x09, 0x13 over the serial line

Beginning with the LSB, the signal can be reversed to obtain: 00001001 00010011. In hexadecimal, these correspond to 0x09, 0x13. The analog-to-digital converter, which will be discussed in subsequent sections, is producing the value:  $(1 * 2^4) + (3 * 2^0) = 19$  in decimal. The first bit, 0x13, is simply a byte that was used to signal the receiving program that a series of bytes has ended. For the sake of this example, only one byte was sent, the eight most significant bits of the analog-to-digital converter.

An additional piece of hardware that is required for these communications is a level shifter, which changes the signal levels from the rs232 (serial communications port) to a level the PIC uses. Since the PIC is powered by a small 3.3V lithium button-cell battery, it only has a 3.3V source to work with. As a result, the PIC represents a “0” digitally by setting a pin to 0V, its low voltage, and a “1” by with a signal around 3.3V, its highest available voltage. The rs232 communications port, however, represents a ‘1’ with -12V and a “0” with +12V. These voltages are outside the ranges which the PIC can provide from its power source alone.

A small serial module was built that performed these voltage level conversions. The MAX231 integrated circuit performs a standard 0-5V to  $\pm 12V$  conversion with minimal external hardware. It contains a built-in inverting charge pump, so applying a +12V supply voltage to this chip allows a -12V signal to be sent from the MAX231, as well. Our PIC circuit, however, uses a 0-3.3V configuration. Therefore, a separate conversion must be made with an up-shifter, which converts signals transmitted by the PIC from 3.3V to 5V, and a down-shifter, which performs the opposite conversion for signals to be received by the PIC.

The up-shifter consists of a pair of NPN bipolar transistors each arranged in an inverting configuration. A voltage divider at the base of the first transistor halves the input voltage. As these transistors generally have a base-emitter junction voltage of about .7V at room temperature, the first transistor begins inverting when the input signal reaches around 1.4V, about half of the PIC’s 0-3.3V range. Since these are digital signals, the temperature dependence of the bipolar transistor’s gain or the saturation current of the base-emitter junction are not important. Whether the inverter switches at 1V or 2V or anywhere in-between is immaterial; it is only essential that the device shift somewhere reasonably in between 0V and 3.3V. The second transistor simply inverts the inverted signal presented by the first, preserving the binary value of the original signal. This two-inverter configuration reproduces and amplifies these digital signals very reliably over a range of frequencies.

### Upshifter Voltage Characteristics

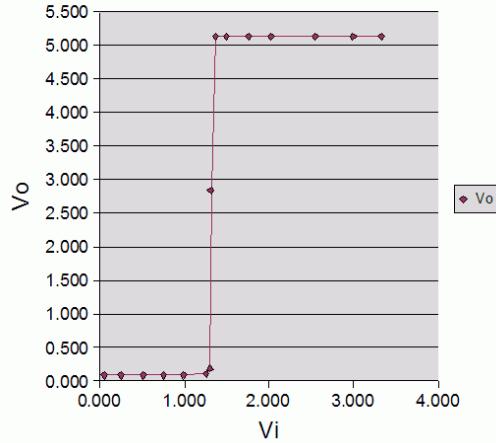


Figure 16: DC Transfer Function of Digital Level Up-shifter

Testing has shown that the PIC configuration used for the pediatric dynamometer communicates most efficiently at 38.4kBaud, or 38,400 bits per second. Start/stop bits must also be transmitted, so each byte send and received is essentially 10 bits. Using a function generator, 38.4kBaud transmission with a 3V peak is simulated by a square wave signal. Figure 17 shows the input signal on channel 2 and the output of the level shifter on channel 1.

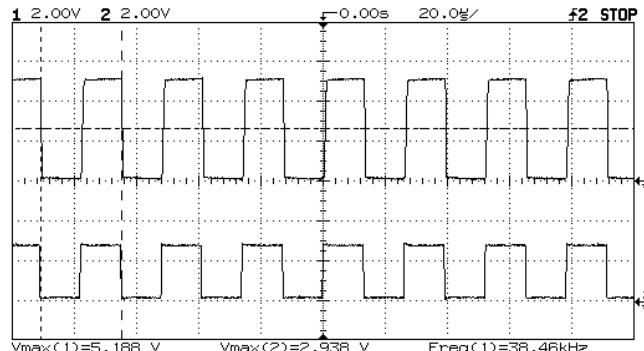


Figure 17: Digital Level Up-shifter operating at 38.4kHz

The down-shifter is even simpler, consisting of a zener diode with a 3V drop and series resistor. This fundamental design exhibits minimal distortion below 500kHz, which is considerably higher than the transmission rates the dynamometer employs.

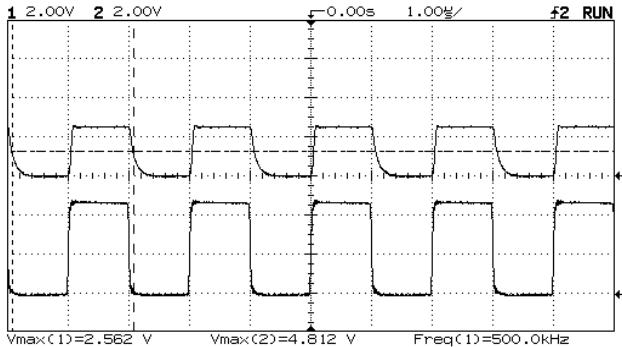


Figure 18: Digital Level Down-Shifter operating at 500kHz

Creating a permanent design for this device was facilitated by the use of Eagle™ Layout Editor software. A schematic of the level shifter's final design is documented in Figure 19.

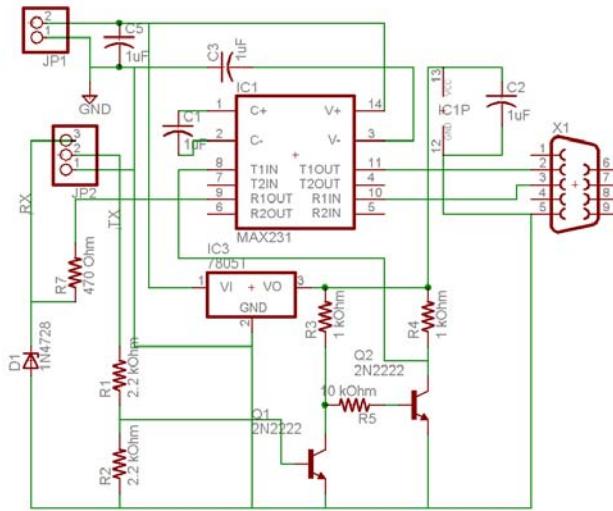


Figure 19: Schematic of Digital Level Shifter (Serial Module)

The hardware was then arranged using the same software suite. Figure 20 shows the final component layout for the serial module. It was milled on a T-Tech numerically controlled micro-mill and cut from standard copper-surfaced PCB.

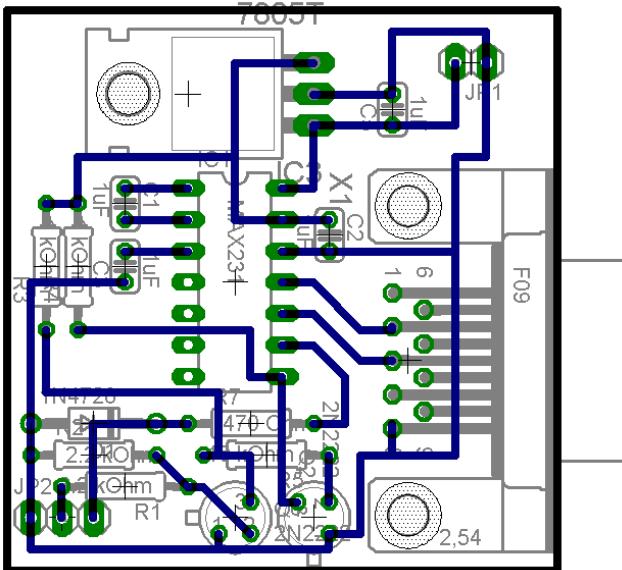


Figure 20: Layout of Digital Level Shifter (Serial Module)

This board was populated with the necessary components and its operation verified by testing the communication link between the PIC and a PC in the lab. A photo of the completed Serial Module is shown in Figure 21. The circuit board measures less than 1.8" x 1.8".

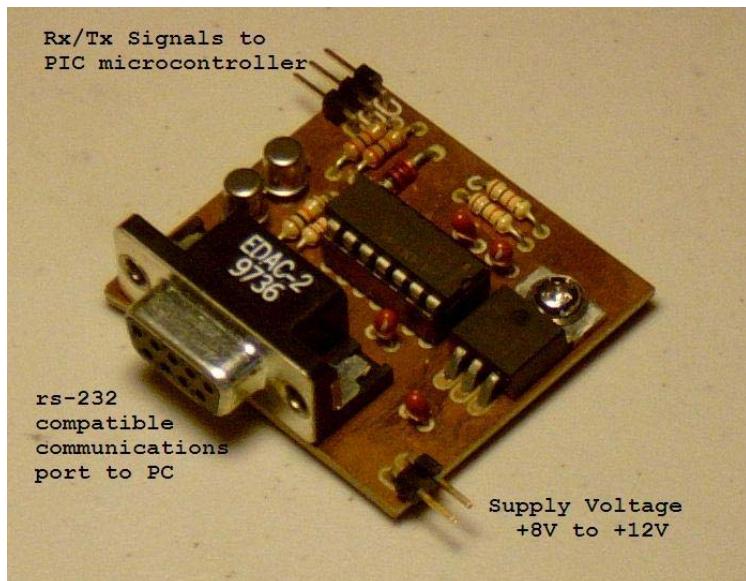


Figure 21: Completed Digital Level Shifter (Serial Module)

When the pediatric dynamometer is ready to be used for clinical research, it will be necessary to produce a number of these Serial Modules, one for each location where information will be downloaded from the dynamometer. This digital level shifter can operate with a supply ranging from 8 to 12 volts, so a standard 9V battery would be a likely candidate due to its ubiquity and low cost. Another alternative is to use a 9V or 12V DC wall adapter to provide a more reliable and maintenance-free source of power.

## 2.6 Synchronous Serial Communications with EEPROM/Flash

Microcontrollers frequently communicate with peripheral devices via a protocol called Serial Peripheral Interface, or SPI. SPI is similar to the serial protocol used for rs232 communications described above, except that SPI is a synchronous serial protocol (SSP). It requires a special clock signal (SCK) provided by the PIC in addition to serial data in (SDI) and serial data out (SDO). There is one additional signal called chip select (CS) which enables the particular device with which the PIC intends to communicate. This way, multiple devices can be connected to the SPI bus, and the PIC can specify which one it is communicating to by setting the appropriate chip select signals.

As an example, the 8-bit character 'W' represented by "01010111" in binary is written to an SPI device. Figure 22 shows the serial clock SCK on channel 1 of the oscilloscope and SDO on channel 2. A common standard dictates that the data is latched to the receiving device on the rising edge of the clock, when SCK switches from low to high. An important distinction must be made in the order of bits transmitted in synchronous and asynchronous modes: SSP bytes are transferred with the most significant bit (MSB) first, while AUSART communications begin with the LSB as discussed in section 2.4.

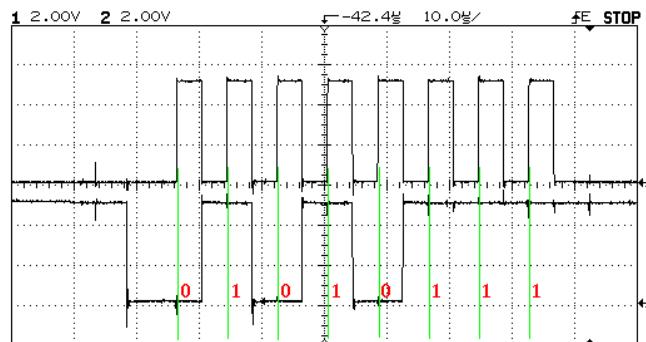


Figure 22: Transmission of 'W' via SSP

Each device connected to a microcontroller has its own dedicated chip select signal. By convention, CS is normally high when the microcontroller is not communicating with the device. Once the microcontroller sets CS low, instructions and data can be exchanged with the peripheral. Figure 23 shows the chip select signal (CS, channel 2) going low while the clock signal (SCK, channel 1) oscillates 8 times, indicating transmission a byte to a peripheral device.

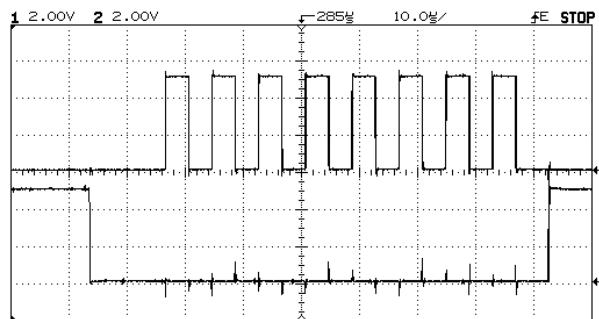


Figure 23: CS Low as SCK oscillates 8 times

The SPI protocol allowed the PIC to read and write a peripheral memory chip serially. Two devices were used for testing the SSP module in software: first, an Atmel 25256 EEPROM was used because of its higher supply voltage specification and larger package, which made it easier to connect to our prototype. The ATMEL has a 256kbit capacity. The Spansion S25FL016A serial flash chip has a 16Mbit capacity and is much smaller in size than the Atmel model, making it an attractive solution for the data storage needs of the pediatric dynamometer. Because of its small size and use of surface mount technology, though, the Spansion flash was used only in later prototypes.

The Atmel and Spansion memories first receive instructions from the microcontroller. These instructions are eight bit codes that initiate a certain operation, such as read, write, erase, and write-enable/disable. The instruction may be followed by data for appropriate instructions. Read and write instructions require a 24-bit address indicating where the data will be read from or written to. Erase and write-enable instructions do not require a specific address.

Another subtle difference between the Atmel 25256 and Spansion S25FL016A exists in their fabrication technology: the Atmel memory is EEPROM, electrically-erasable programmable read-only memory, while the Spansion device is flash memory. The Atmel device can be re-written to freely, because the internal circuitry automatically erases a block of memory when a write instruction is received. The Spansion device can only write 0's to memory blocks, not 1's. In order to re-write 1's over existing 0's, sectors of memory must first be cleared. This required writing a software routine that cleared the Spansion flash memory upon receiving instruction.

## **2.7 Integrating AUSART/SSP Components in Software**

It is necessary for the dynamometer to be able to transfer data from the serial flash memory to the serial port of a computer. General-purpose functions were written to communicate with a device connected to the SSP bus and a computer connected to the AUSART. Additional functions enable and disable the SSP and AUSART interfaces as needed. The PIC16F88 cannot use the SSP and AUSART features simultaneously, as pin 8 of the microcontroller can act exclusively as an input Rx from the AUSART module or an output SDO to the SSP module.

Fortunately, disabling one protocol frees up the resources to enable the other. Continuously switching between SSP and AUSART is not as intuitive as it may seem, and requires the PIC to carefully manage the states of the pins and features of the protocols so that no conflicts occur. The following four functions oversee this operation: `enableAUS()`, `disableAUS()`, `enableSSP()`, and `disableSSP()`.

```
// Set up reception/transmission from serial port
void enableAUS(){
    // Set up PORTB
    PORTB.2 = 1;      // Remove artifact of SSP
    TRISB.5 = 0;      // PortB5 output for Tx
    TRISB.2 = 1;      // PortB2 input for Rx
    // Set up AUSART
    TXEN = 1;         // Enable Transmission
    CREN = 1;         // Enable Continuous Receive
    SPEN = 1;         // Enable Port
    RCIF = 0;         // Clear Serial Receive Interrupt Flag
    RCIE = 1;         // Enable Serial Receive Interrupt
}
```

Table 1: `enableAUS()` method

Before the AUSART can be used, it is necessary to configure the pins on the PIC. Since asynchronous serial communications are idle high, it is important that the PIC begins AUSART mode with the Tx line on `PORTE.2` high; otherwise, the PIC will interpret a low signal as a “start bit” as explained in 2.5. Furthermore, it is necessary to set `PORTE.5` as an output Tx to the serial module and `PORTE.2` as an input Rx from the serial module.

The last five lines enable specific features of the AUSART. Transmission and reception must be enabled before use; theoretically the PIC can use the asynchronous serial interface as only a receiver or transmitter, but the dynamometer uses two-way communication with the PC. When in communications mode, the dynamometer uses an interrupt to receive data from the PC: whatever the PIC is doing is interrupted when a byte is received on the serial line. The two modes of the pediatric dynamometer will be discussed in greater detail in the software results section, 3.2.

```
        SPEN = 0;      // Disable Port  
    }
```

Table 2: `disableAUS()` method

Disabling the AUSART interface requires freeing up the resources needed by the SSP interface. Before this can begin, it is salutary to ensure that the serial port has finished transmitting its last byte of data. This is done by repeatedly polling the bit `TRMT` until it is set (`while (!TRMT);`). If the serial port is turned off before transmission has completed, important data may be lost.

Even after disabling the AUSART, the serial module is still connected to the PIC. The output line (Tx, `PORTE.5`) to the computer must be held high, or else the computer will interpret the low state as a start bit and receive spurious data (usually `0x00`, a zero byte). This is done by setting the pin high (`PORTE.5 = 1;`) before disabling the AUSART.

In a manner similar to enabling the port, each of the port's features must be disabled or else unwanted operation may occur. Particularly, disabling serial receive interrupt (`RCIE = 0;`) will prevent the AUSART from receiving data on `PORTE.2`. Since `PORTE.2` is also used as SDO, data sent from the PIC to the serial flash memory will be received by the PIC's AUSART unless serial receive is disabled. If serial receive is not disabled, the next time AUSART is enabled, the SDO data will have been read into the AUSART receive flag, causing unpredictable results.

```

// Synchronous serial port for off-chip EEPROM
void enableSSP(){
    // Set up PORTB
    TRISB.1 = 1;      // PortB1 input for SDI
    TRISB.2 = 0;      // PortB2 output for SDO
    TRISB.4 = 0;      // PortB4 output for SCK
    // Set up SSP
    SSPEN = 1;        // Enable SSP
//    SSPIE = 1;        // Enable SSP Interrupt

```

Table 3: enableSSP() method

Enabling the SSP interface requires fewer instructions than enabling the AUSART. Again, the appropriate `PORTE` pins of the PIC need to be set to input/output for proper SPI operation. One line (`SSPEN = 1;`) enables the SSP interface. Another optional line (`SSPIE = 1;`) can be used to enable interrupts whenever an SPI byte transmission is complete. The pediatric dynamometer uses polling with the SPI interface because it is less of a priority than sampling the ADC and reading from the serial port, which must be done in real-time.

Table 4: disableSSP() method

Finally, before disabling SSP interface, it is necessary to ensure that the clock SCK remains low while SSP is not in use. Disabling SSP while `PORTE.4` is set high will result in a rising edge, which initiates data reception on the peripheral device. After clearing bit `PORTE.4`, the SPI can then be disabled by clearing (`SSPEN = 0;`).

Armed with these four functions, the PIC can interface with the serial port of the computer and the serial flash memory by switching modes as needed. Table 5 shows an example of how the pediatric dynamometer sends the contents of the flash memory to the computer one byte at a time.

```

sendByteSSP(READ);
sendByteSSP(0x00);
sendByteSSP(0x00);
sendByteSSP(0x00);
// Read each byte and write to terminal
// until we receive a blank byte
do {
    readBuf = receiveByteSSP();
    disableSSP();
    enableAUS();
    serialout(readBuf);
    disableAUS();
    enableSSP();
} while(readBuf != 0xFF);
disableSSP();
end_InstSSP

```

Table 5: Reading from the serial flash memory

SSP must first be enabled. The PIC then sets chip select low in the instruction `begin_InstSSP`. Once CS is low, the serial flash memory begins receiving instructions. The next four lines send four bytes of data to the memory: first, a read instruction is sent, followed by the twenty-two bit address 0x000000, the very beginning of the flash memory. After these four bytes are sent, the memory sends bytes as long as the clock SCK keeps oscillating and CS remains low. The line (`readBuf = receiveByteSSP();`) stores a byte sent by the memory into the buffer `readBuf`. Then, SSP is disabled and AUSART is enabled. The contents of the buffer is sent out to the serial port in the statement (`serialOut(readBuf);`). Once transmission of this byte is complete, the AUSART is disabled, SSP is enabled, and the cycle can continue. Before it does, however, the PIC checks the buffer `readBuf` to see if it is 0xFF, which in binary is 11111111. All 1's is characteristic of flash memory that has either not been written to or has been cleared. The PIC interprets this condition as the end of data stored in memory and stops reading. SSP is disabled and the CS signal is set high to indicate the end of the read instruction.

The read/write cycle was demonstrated using the HyperTerminal® program. A sequence similar to the one in Table 5 was used to write a character to the flash memory every time a character was typed to the terminal. Pressing Ctrl + R initiates the memory dump shown in Table 5.

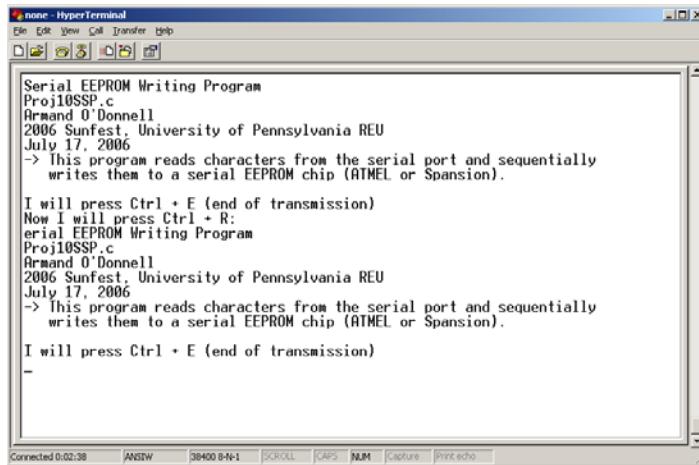


Figure 24: Screen capture of serial flash memory I/O

### 3. Final Results

#### **3.1 Hardware Design Results**

This summer's work resulted in a fully functional pediatric dynamometer. The final design consumed less space than any previous design. Figure 25 shows the schematic of the dynamometer in Eagle Layout Editor®.

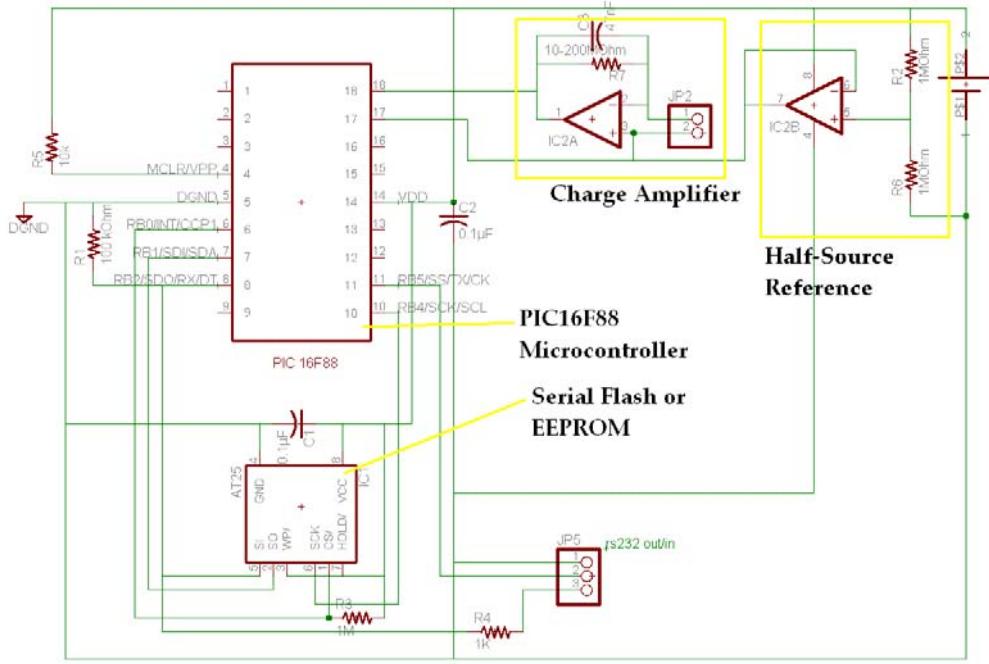


Figure 25: Schematic of Pediatric Dynamometer

A 3.3V lithium cell provides power to this circuit, shown on the upper right. The charge amplifier, shown in the top right corner of Figure 25, does not have a gain stage like the one discussed in Section 2.4. Using a 47nF capacitor in the feedback loop of the charge amplifier provided sufficient voltage swing for mechanical stress on the PVDF sensor without saturating the amplifier. A 200MΩ resistor is used in the feedback loop to minimize drift while attenuating very low frequency signals. Note that both the half-source reference and the output of the charge amplifier are connected to the PIC16F88. These are analog inputs. The PIC's ADC samples both the half-source reference and the output of the charge amplifier. A comparison is done so that if the reference voltage is not exactly  $3.3V/2 = 1.65V$ , the PIC can still calculate the CA's output relative to the reference. How this is done in software will be discussed in more detail in section 3.2.

There are two inputs to this system; the PVDF sensor is connected to the two pins on the right side of the charge amplifier. The bottom of Figure 25 shows a three-pin connector marked "rs232 out/in" to which the serial module is connected. The 1kΩ resistor (R4) between pin 3 of this connector and pin 8 (Rx) of the PIC serves to prevent the signal from the serial module from interfering with the SDO signal from the PIC in SPI mode. When the AUSART is enabled, pin 8 of the PIC becomes a high-impedance input and the 1kΩ resistor does not attenuate the signal from the serial module. When the PIC is in SPI mode, pin 8 becomes a low-impedance output providing SDO data to the serial flash memory. The serial module still provides an idle-high signal, but the 1kΩ resistor prevents the signals from loading each other. Figure 26 shows the Rx/SDO pin 8 of the PIC on channel 1. Note how both AUSART and SSP signals are at full strength. Channel 2 shows the signal at pin 3 of the "rs232 out/in" connector. In this case, only the AUSART signal is preserved and the SPI signal is attenuated.

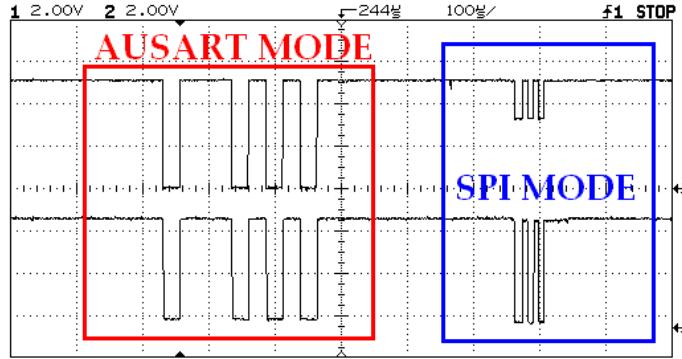
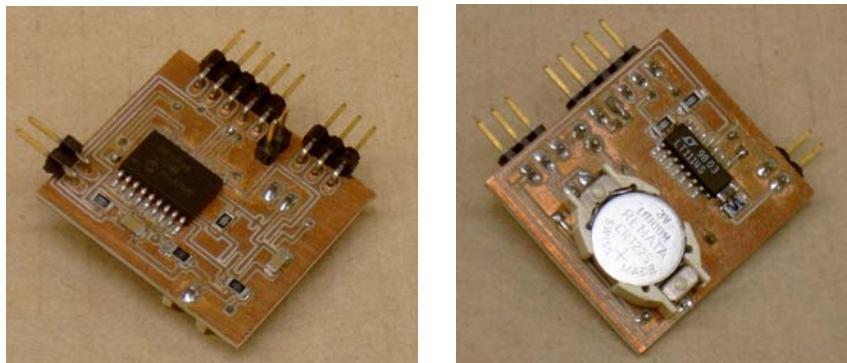


Figure 26: Comparison of SDO, Rx signals

The  $100\text{k}\Omega$  resistor between ground and pin 8 allows the PIC to detect a serial connection. When SSP is disabled and PORTB.2 is an input, the  $100\text{k}\Omega$  resistor to ground pulls this input low. When the serial module is connected, the idle-high signal from the serial module pulls high, directing the PIC to stop sampling the ADC and enter communications mode. A discussion of the two modes is presented in section 3.2.

To test the pediatric dynamometer's design, a prototype was milled on the T-Tech 5000 CNC milling machine and populated with surface-mount technology (SMT) components.



Figures 27, 28: Top and bottom of pediatric dynamometer prototype

The prototype shown in Figures 27 and 28 has some extra pins for programming and debugging the PIC and for an external power source. These photos were taken while testing the AUSART interface, when the serial flash memory had not yet been soldered on. The prototype measures  $1.215'' \times 1.335''$  ( $3.086\text{cm} \times 3.391$ ). A photo comparing the 2004-2005 and 2005-2006 Senior Design projects with this summer's prototype is shown in Figure 29. Space was saved by avoiding use of a timing crystal, since the PIC contains its own internal RC oscillator that is satisfactory for the dynamometer.

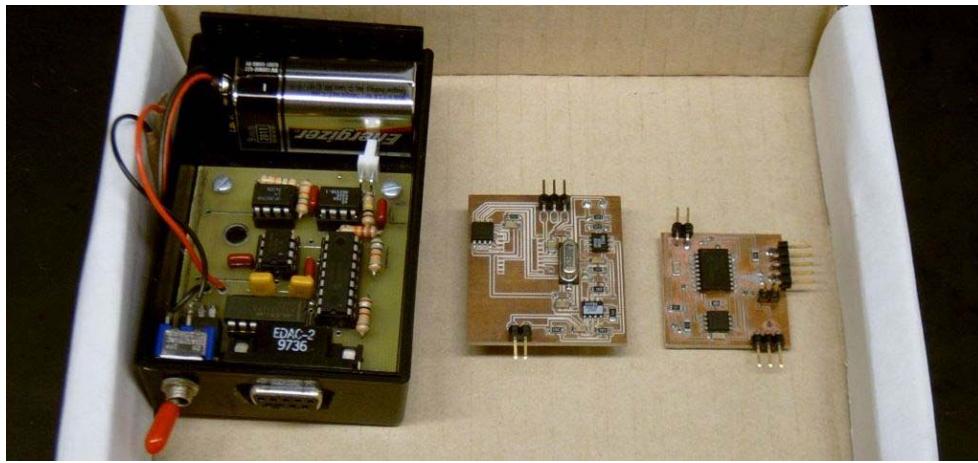


Figure 29: Pediatric Dynamometers:  
 (left) Senior Design 2004-2005  
 (center) Senior Design 2005-2006  
 (right) Sunfest 2006—Prototype

The digital components of the prototype function as expected. The PIC is able to read and write the Spansion serial flash memory and communicate with a personal computer via the serial module. The analog components are more sensitive to slight imperfections in fabrication and sometimes did not perform as desired in the prototype. The half-source reference circuit provided a 1.63V reference, which is satisfactory. The charge amplifier worked but some connection in the circuit introduced a parasitic resistance much lower than the desired  $200\text{M}\Omega$ . As a result, the charge amplifier suffered from decay signals similar to what is shown in Figure 8. Future options for preventing this from occurring include placing a small guard ring around the charge amplifier feedback loop and using “rub out” to remove all of the copper from the area surrounding the charge amplifier, minimizing the risk of parasitic impedances.

Removing the test and power pins from the prototype reduces the size of the board. Since only two operational amplifiers are needed, the LT1112 is more space-efficient than the LT1114 quad op-amp used by the prototype. The resulting pediatric dynamometer design could be used in the bone health studies outlined in the background. Figures 30 and 31 show the layout of the final design of the pediatric dynamometer. The board measures 1.135" x 1.150".

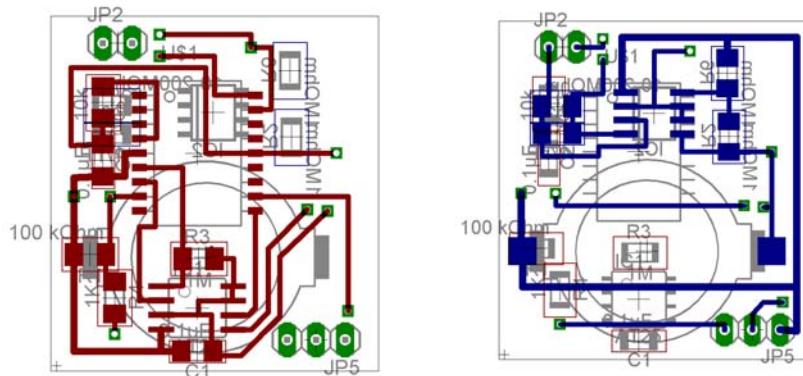


Figure 30 (left): Top layer of pediatric dynamometer  
 Figure 31 (right): Bottom layer. Final 2006 SUNFEST design.

The battery used in this design cannot provide sufficient power to the pediatric dynamometer for more than two days. Using a larger lithium battery like the one used by the 2005-2006 Senior Design team should allow this year's design to run for a week or more. Further work can be done to improve this year's design by incorporating a larger battery. Care must be taken to keep the design small in size.

### 3.2 Software Design Results

The pediatric dynamometer operates in two modes: sampling mode and communications mode. In sampling mode, the PIC consumes less power by operating at a lower frequency, 250kHz. It samples the ADC connected to the charge amplifier 200 times per second and the ADC connected to the half-source reference every 256 cycles (roughly once every 1.28s). Every sample, the PIC compares the current measurement with previous measurements and updates the appropriate data as necessary.

The PIC's goal is to determine the beginning of a step, record important data for that step, and determine the end of the step, writing important data to the flash memory. This includes: a timestamp of when the step started, the duration of time the step lasted, the average force on the PVDF sensor, and the max force over the period of the step. Table 6 is a flow chart that outlines this operation.

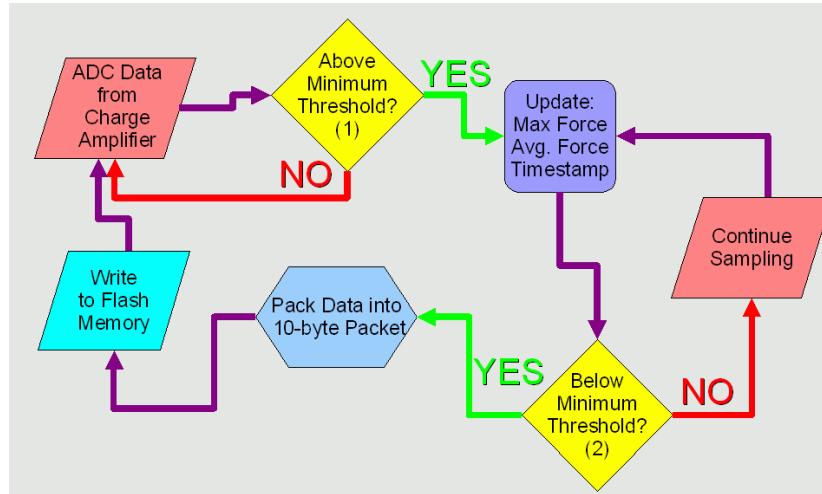


Table 6: Sampling mode algorithm

The PIC determines the beginning of a step by comparing the ADC of the CA with the ADC of the half-source reference. If the CA's value is above the reference by a set threshold, the PIC interprets it as the beginning of a step and records the timestamp. Processing of the max and accumulated force continues every sample. Once the CA's measurement is below a different threshold, the PIC detects the end of a step, and calculates the length of the step by subtracting the timestamp from the current time. The threshold for detecting the beginning of a step is higher than the threshold for the end of a step. Hysteresis prevents the PIC from detecting a small variation above the threshold as a complete step.

After every step, the PIC writes a 10-byte packet of information to the flash memory. The first three bytes encompass the timestamp, represented in 5ms increments from the time the PIC begins sampling. This counter overflows once every 23.3 hours, almost once a day, but upon

inspection of the data it should be evident where one day ends and another begins. The next two bytes contains the duration of the step in 5ms increments. The max step length is therefore 327.7 seconds. This is followed by a three-byte number that represents the accumulated sum of each ADC measurement made during the step. This number can be divided by the two-byte step length to obtain the average force. The PIC does not support division in hardware, and software division takes hundreds of cycles. Therefore, it is preferred to carry out the average force calculations on the computer once the data has been uploaded. Finally, the last two bytes contain the maximum ADC measurement over the course of the step, corresponding to the maximum force.

Once a serial connection is detected, the PIC immediately stops sampling the ADC, and switches to 8MHz operation. At this speed, the PIC can communicate with a PC's serial port at 34.8kbaud with perfect data fidelity. In communications mode the PIC can receive the following instructions: dump the raw contents of the flash memory (Ctrl + R), begin a hex dump of the flash memory (Ctrl + X), clear the flash memory (Ctrl + E), and begin sampling (Ctrl + S). Using HyperTerminal<sup>©</sup>, each of these operations was tested.

If the user enters Ctrl + S, the PIC slows down to 250kHz, and waits for the user to detach the serial module. Once this happens, sampling proceeds until the cable is connected again. If the user wishes to erase the flash memory and enters Ctrl + E, the PIC first prompts the user for confirmation, and only proceeds if the user presses 'y'.

By the end of the summer, the software interface between the PIC and the computer worked perfectly. The PIC provides a raw data or hex dump of the contents of the flash memory upon request. A calibration mode was added to test the operation of the algorithm in Table 6. Pressing Ctrl + C at the terminal begins real-time sampling of the ADC. Measurements of the half-source reference and charge amplifier are continually updated on the screen along with the accumulated force for the current step and max force. An 'S' appears at the end of the data whenever the PIC detects the beginning of a step, and disappears at the end of the step. Connecting a small PVDF sensor to the board in Figures 27 and 28 and gently deflecting it allowed testing of both the hardware and the algorithm when in calibration mode.

One aspect of the software that needs more work is writing to the flash memory in sampling mode. The flash memory can be written to using other programs that run in 8MHz mode, but for some unknown reason, the flash memory does not receive data when the PIC is in sampling mode. Inspection of the CS, SCK, and SDO signals with an oscilloscope verify that the PIC does operate correctly by writing to the flash memory at the end of a step; however, the flash memory is empty or contains old data whenever the PIC tries to read it after exiting sampling mode. Future work will require isolating this software problem and fixing it so that data acquired in sampling mode can be recorded on the flash memory.

#### 4. Conclusions

As a result of this summer's work, a functioning pediatric dynamometer was developed. The digital end of the hardware prototype functioned as desired. The basis for communicating among the PIC, a personal computer, and flash memory has been established. The current configuration has enough storage capacity to record 200,000 steps, enough for weeks or even months of measurements. Furthermore, the current digital subsystem can be used for other types of studies, since the hardware for recording data, storing it, and retrieving it does not depend on the type of data being recorded.

The analog circuitry has been reconsidered this summer. Adding the analog charge amplifier allowed us to remove the signal conditioning circuits used by the two previous senior design teams—saving the space of an 8-pin instrumentation amplifier. The CA also served as an alternative to software integration. It made prototyping easier, since connecting an oscilloscope probe to the output of the CA displayed a signal proportional to the force on a scope; software integration required the integral to be sent over the serial port into a computer and displayed in software, a more time consuming proposition. The CA introduced its own set of problems, not the least of which was trying to attenuate low-frequency signals like the pyroelectric effect while keeping the RC time constant high enough not to leak too much charge off of the capacitor. The CA built on our final prototype operated with unexpectedly high parasitic charge loss, which was not a problem in the large CA built on a separate board discussed in Section 2.4. Ideally, a controlled experiment should compare the results of software integration and the charge amplifier.

Another issue to take into consideration is that when the charge amplifier contains a resistor in the feedback loop, charge will be drawn off of all signals, including higher frequency signals. This effect is less immediately evident with higher frequency signals. As time goes on, the charge amplifier will settle to a state where the average signal is the half-source reference. This means that after minutes of walking, steps will no longer introduce a signal beginning at the half-source reference; the signal will settle to a level below the half-source reference such that if the signal is plotted in time, the area above the half-source reference will equal the area below it. This isn't necessarily a problem, except for the fact that the current algorithm measures all signals with respect to the half-source reference. Further work should include sampling the minimum force and taking the average with respect to *it*.

Much was learned over the course of this summer, as the project required reading numerous datasheets for specifications and instructions, milling circuit boards for prototypes, and programming the PIC microcontroller in both C and assembly. The PIC and Spansion flash memory often acted unexpectedly, but consulting the datasheet frequently clarified the issue. Using the T-Tech 5000 CNC milling machine became easier with practice; fortunately, since even slight changes in the design required building new prototypes, there were plenty of opportunities. Circuit layout in Eagle was another skill that was developed over time, as the software takes a bit of getting used to.

The pediatric dynamometer is ready to enter the final stages of design and testing. Once the last few software bugs are ironed out, it will be ready for testing in the field. Controlled experiments can be done to acquire calibration data as needed. A variety of batteries can be tested with the dynamometer, allowing designers to make an informed decision on the best overall cell in terms of cost, size, and capacity. Once these hurdles are overcome, the pediatric dynamometer will be ready for mass production and use in pediatric bone health studies.

## 5. Acknowledgments

This pediatric dynamometer would not have been possible if it were not for the perseverance of Dr. Jay Zemel in continuing the project. Many thanks for your advice, not only about electronics and design, but about engineering and education in general. Thank you Dr. Van der Spiegel, for organizing Sunfest and our summer accommodations. Sid Deliwala, for being available to dispense the necessary parts—on such short notice—and for your help with lab equipment and

PIC references. Finally, Dr. Haim Bau and Dr. Howard Hu deserve a special thank you for allowing us to use our noisy T-Tech 5000 in their lab.

## 6. References

- [1] Mackelvie KJ, McKay HA, Khan KM, Crocker PR. *A school-based exercise intervention augments bone mineral accrual in early pubertal girls*. J Pediatr. 2001 Oct; pp. 139(4):501-8.
- [2] Elgar F., Roberts C., Tudor-Smith C., Moore L.; *Validity of self-reported height and weight and predictors of bias in adolescents*. J Adolesc Health. 37 (5): pp 371-5. Nov 2005.
- [3] Welk G., Corbin C., Dale D.; *Measurement issues in the assessment of physical activity in children*. Res Q Exer Sport, 71 (2nd Suppl): S 59-73. Jun 2000.
- [4] United States Patent #5,269,081: Gray, *Force Monitoring Shoe*; 1993.
- [5] United States Patent #4,814,661: Ratzlaff et al., *Systems for measurement and analysis of forces exerted during human locomotion*; 1989.
- [6] United States Patent #5,925,001: Hoyt et al., *Foot Contact Sensor System*; 1999.
- [7] Tsai O., *Dynamometer – The New Activity Monitor*, Report no. TR-ST01NOV04, Center for Sensor Technologies, Univ. of Pennsylvania, pp. 199-221, 2004.
- [8] Unpublished Senior Design Report: Lamptey L., Sin J., Wong M.; *Pediatric Step Monitor for Bone Health Studies*. Univ. of Pennsylvania, (a) 7-8, (b) 39-43. 2005.
- [9] Unpublished Senior Design Report: Kam M., Tran D.; *Flexible Pediatric Dynamometer for Bone Health Studies*. Univ. of Pennsylvania, pp. 6-7, 2006.
- [10] Bauer S., BauerGogonea S., Dansachmüller M., Dennler G., Graz I. Kaltenbrunner M., Keplinger C., Reiss H., Sariciftci N.S., Singh T.B., Schwödiauer R.; *Piezoelectric Polymers*, Mater. Res. Soc. Symp. Proc. Vol. 889, pp. 1-4. 2006.
- [11] Measurement Specialties, Norristown, PA. *Piezo Film Sensors: Technical Manual*. April 1999.

## Appendix A. Recommendations for Future Work

### A.1 Digital Signal Regression

After using a “charge amplifier” to determine the amount of deflection done on the PVDF sensor, it became clear that two conflicting goals arose from the RC feedback loop. A small time constant ensured that the system would not drift away from the 1.65V reference, yet a substantial amount of charge bled off of the capacitor very quickly, canceling out the integration. On the other hand, a large time constant led to fewer losses during integration, but whenever the PVDF sensor returned to its original position, the signal took a very long time to return to the 1.65V reference.

This situation can be modeled in Matlab, fabricating a few simple mathematical examples that are theoretically accurate. Figure A1 shows a pair of unit impulses, one with amplitude +1 and the other with amplitude -1, 50 samples apart. In this case, the PIC will take 200 samples/second, so 50 samples corresponds to a quarter-second. The second graph shows the signal after integration.

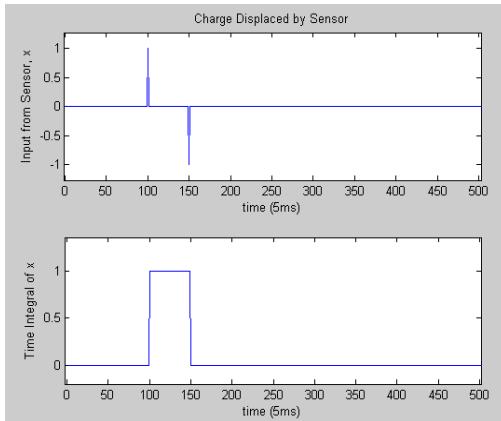


Figure A1: Integral of a “unit impulse”

Taking into account the exponential decay of the charge stored in the capacitor, the exponential function  $e^{-t/RC}$  is loaded into an array, where the time constant is .796. Matlab's convolution function simulates the integration with this exponential decay taken into account. Integrating with respect to time in the continuous domain is essentially the same as taking the convolution of the “unit step” function in discrete domain with respect to samples.

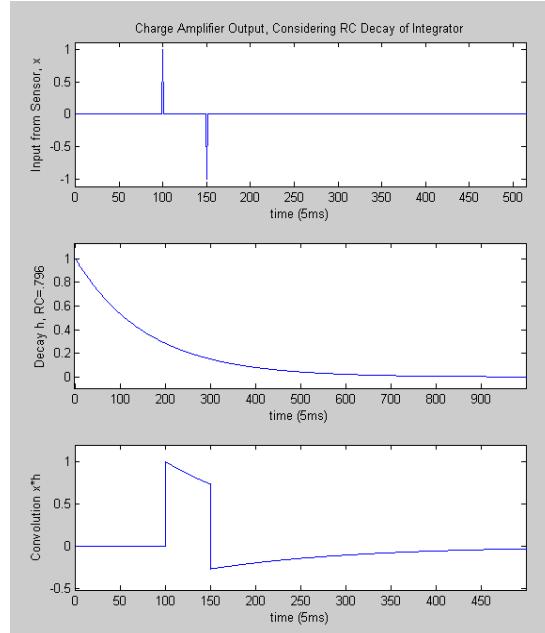


Figure A2: The RC time constant degrades charge

The RC parasitic effect degrades the integration from what it should be, as shown in Figure A1. In particular, right after the first impulse, charge is lost from the capacitor and after the second pulse; it takes very long for the signal to return to zero.

Two unit impulses are hardly a general case, so a more continuous series of inputs is considered in Figure A3. The next example shows how the system responds to a pair of triangular current pulses close together.

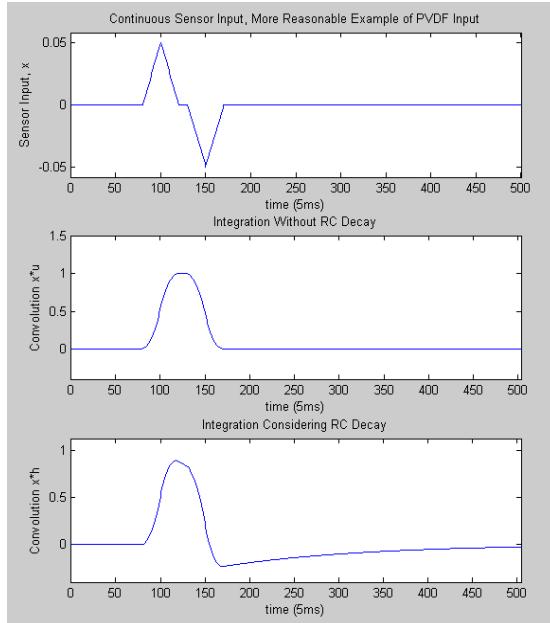


Figure A3: Triangular impulse

The analog signal is not that distorted in this case. However, the rising and falling pulses occur very closely together. This might be characteristic of a quick step, but there is no guarantee that all steps will be like this. Longer steps will have the rising and falling impulses further apart, and as a result the feedback loop will have more time to lose charge.

In this case, the degradation of the data is more severe. It will be impossible to directly apply this sampled data to the pediatric dynamometer.

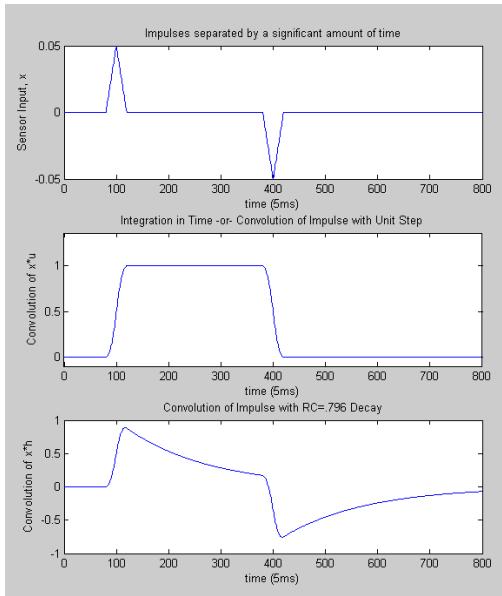


Figure A4: Separated impulses

The effects of convolution can be reversed using a function called deconvolution, in the Matlab function *deconv*. It is possible to reconstruct the original function from the convoluted result if the response of the system is known:

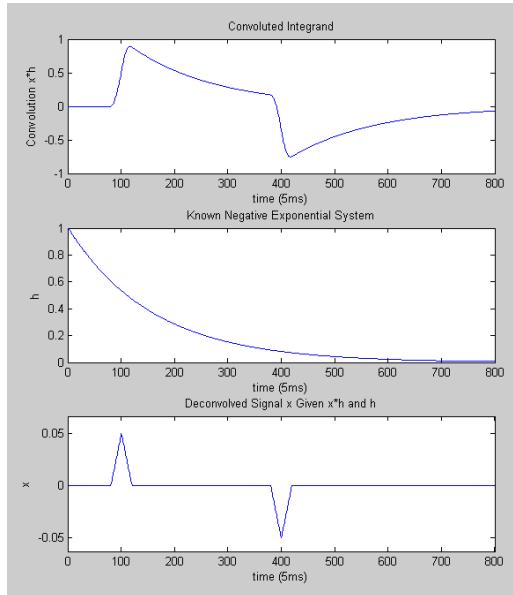


Figure A5: Reconstruction of original data

Unfortunately, there are two problems with this solution: deconvolution requires a fair amount of processing power, and would require the PIC to do a lot of computation, if real-time deconvolution were even possible. The other problem, more importantly, is that the deconvolution of the analog integration is completely useless to us, as it contains nothing more than the data from the sensor; we might as well just read that in directly.

Fortunately, a small algorithm written for Matlab takes the distorted integrand and performs a regression on it, expecting the decay due to the RC time constant.

```
% Armand O'Donnell
% June 2, 2006
% SUNFEST, Pediatric Dynamometer
% DeconvCA, an algorithm to perform data regression on a signal
% whose integral has undergone exponential decay.

% RC Time constant, dependent on amplifier feedback loop
RC = .796; % for Rf = 700kOhm, Cf = 1uF

reconInt(1) = 0;

for iCnt = 2:1000;
expNext = y(iCnt-1)*exp(-.005/RC);
if y(iCnt) < (expNext - .001) | y(iCnt) > (expNext + .001)
reconInt(iCnt) = reconInt(iCnt - 1) + (y(iCnt)-expNext);
else
reconInt(iCnt) = reconInt(iCnt - 1);
end
end

clear expNext;
```

Table A1: DeconvCA.m

A few advantages of this code are as follows: first of all, although this example calculates the next expected input from the value of  $e^{-0.005/RC}$ , one could easily arrange the code so that the program never needs to calculate the exponent. For each iteration, a constant  $e^{-0.005/RC}$  will be multiplied to the value calculated in the former iteration. If there is a significant discrepancy, it must be due to a signal from the PVDF sensor; so the integral is adjusted accordingly. One can easily change the RC time constant by altering the first line of code. Although the program uses intervals of .005 seconds to correlate with the 200 samples/second, the rate could also be changed with minimal code alteration.

This program was run on the distorted integral on the bottom of Figure A4 and is compared to the theoretical (fully digital) integration. As far as I am concerned, the regression produces a result that is strikingly similar to what we are looking for:

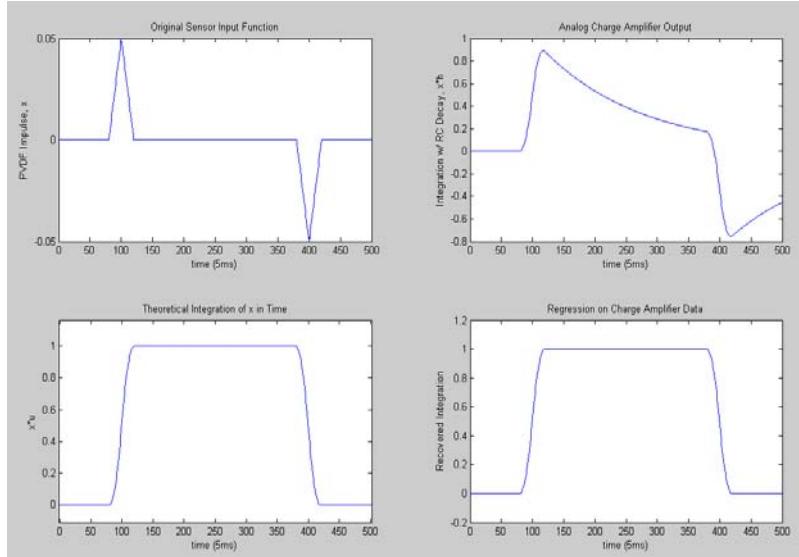


Figure A6: Regression from distorted integral to reconstructed integral.

The next step is to figure out the effect of an incorrect RC time constant. Since each iteration multiplies the former by the exponent, the algorithm is somewhat sensitive to errors in the exponent. This produces an offset so that the error “corrects” itself by compensating either above or below center in the steady state.

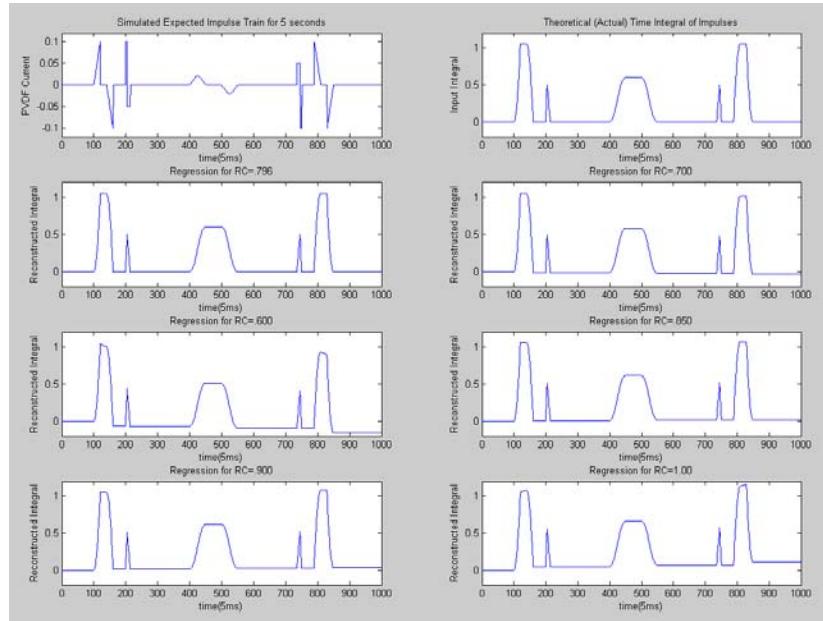


Figure A7: System response with incorrect RC time constants

Figure A7 shows that the signal only produces significant offset for RC time constants more than 10% outside of the true value. Otherwise, the small offset produced is negligible. Components within 5% of the rated value are readily available, in which case the RC constant can be measured experimentally and assumed to be the same for all devices made. Furthermore, it should be possible to write a module in the software that calibrates the time constant to match that of the analog charge amplifier device.

## **A.2 PC Software Component**

The PIC's communication with the PC has been proven with a common serial terminal program. This is fine in the lab, but medical researchers will desire more user-friendly software. Furthermore, the terminal program is incapable of storing data in a file for processing or later use. An appropriate software interface must be written to accomplish these tasks.

Perhaps a Java graphical program would be a practical solution. Java is preferred because it is compatible across platforms. A graphical user interface can feature function buttons such as “read flash”, “erase flash”, and “begin sampling”. Pressing one of these buttons would send a signal to the PIC and initiate the corresponding function.

The Java program would also be able to capture the raw data dump from the PIC, process the data into decimal values, and store the measurements in a standard format. For example, the comma-separated values (\*.csv) format can be used by a number of spreadsheet programs, and a little formatting would allow scientific software such as Matlab to do statistical analysis on the data presented.

## Appendix B. Source Code listing

```
/*
Armand O'Donnell
July 25, 2006
Sunfest-University of Pennsylvania
Proj12ADCStore.c

This software can be used on a fully functional pediatric
dynamometer. The PIC samples the ADC 200 times per second,
and can figure out when a step begins and ends. When a
step begins, a timestamp is set. A summation of the force
begins, so that the average force can be calculated. The
max force is also documented here.

At the end of a step, the statistics for that particular
step are saved to the EEPROM.

When a serial module is connected to the unit, the clock
frequency is increased to 8MHz and the device enters
serial communications mode. In this setting, the PIC can
communicate with a personal computer. The following
control codes bring about the corresponding functions:

Ctrl + R: Raw dump of EEPROM data. This is the fastest
method of reading the pediatric dynamometer, but the
data is unreadable on a terminal; instead, it should be
read by a special program used for interfacing with the
dynamometer. Ideally, it could create a list of
measurements stored in a comma-separated values file (.csv).
This can be read into an Excel spreadsheet or Matlab
program for proper data analysis.

Ctrl + X: Hex dump of EEPROM data. The read information
will be displayed on a terminal screen, one line of data
corresponding to a step in the following format:
BBBBBB LLLL AAAAAA MMMM
B - Beginning timestamp in 5ms units
L - Length of step in 5ms increments
A - accumulated force over duration of step
M - max force measurement during step.

Ctrl + S: Sample mode. PIC will switch to 250kHz mode, wait
for serial module to be removed, then it will start
sampling ADC continuously, in data acquisition mode

Ctrl + E: Erases entire EEPROM. PIC will poll EEPROM's
status register once a second until EEPROM signals that
erasure is complete. This key combination also clears a
number of variables, like the master timer, so that the
data acquisition can begin from scratch.

Uses a PIC16F88, a Spansion S25FL016A EEPROM or Atmel
25256 Serial EEPROM, and a properly assembled charge
amplifier.

*/
// Interrupt handler library
#include "int16CXX.h"

// Spansion S25FL016A Serial Memory Programming Codes
#define WREN 0x06 // Write enable
#define WRDI 0x04 // Write disable
#define READ 0x03 // Read data bytes
#define SE 0xD8 // Sector erase
#define BE 0xC7 // Bulk (chip) erase
#define RDSR 0x05 // Read Status Register
#define WRSR 0x01 // Write to Status Register
#define PP 0x02 // Page program
```

```

// Chip select pin, drive low to use memory
#pragma bit CS @ PORTB.0

// Chip select low to initiate instruction to EEPROM
#define begin_InstSSP CS = 0; \
                           nop(); \
                           nop(); \
                           nop();

// Notifies EEPROM of end of instruction
#define end_InstSSP           nop(); \
                           nop(); \
                           nop(); \
                           CS = 1;

bit sample_mode_flag;      // 0 for serial comms mode, 1 for sampling mode
bit write_ready;           // Signals that a step is ready to be written to EEPROM
bit insideStepFlag;        // Currently sampling a step or not?
long ADCData;              // 16-bit variable containing 10-bit result of last ADC
uns24 masterTimer;         // Will oversee timestamp implementation, incremented every 5ms
uns24 stepStartTime;       // Timestamp contents at beginning of current step
long calData;              // Digital representation of half-source reference
long relSample;            // ADC measurement of Charge Amplifier relative to calData
long forceMax;             // Max relative ADC measurement for current step
uns24 forceAccum;          // Summation of CA measurements, used to calculate average on PC

// Temporary variables for writing to EEPROM
uns16 wr_stepLength;       // Number of 5ms cycles for previous step
uns24 wr_stepStartTime;
uns24 wr_forceAccum;
long wr_forceMax;

/*
   ADC units (about 3mv) of high/low threshold for signal to be considered a step
   Uses hysteresis: if ADC signal is STEPHIGHTH above calData, begin measuring inside
   of a step. Continue until signal is STEPLOWTH above calData.
*/
#define STEPLOWTH    10
#define STEPHIGHTH   20

// Pre-define serial I/O functions
char serialIn;
void serialOut(uns8 txByte);
void enableAUS();
void disableAUS();
void enableSSP();
void disableSSP();
void sendByteSSP(char wrData);
char receiveByteSSP();

// Hexadecimal formatting functions for hex dump
char hiHex(char binData);
char lowHex(char binData);

#pragma origin 4
interrupt serverX(){
    int_save_registers;

    // Serial Port Receive Interrupt
    if(RCIF){
        // Receive input byte
        serialIn=RCREG;
        RCIF = 0;           // Clear USART Receive Flag
    }

    // Analog-to-Digital Conversion Completion Interrupt
    if(ADIF && ADIE){

        ADIF = 0;           // Clear ADC flag
    }
}

```

```

// Read from ADC
ADCData.low8 = ADRESL;
ADCData.high8 = ADRESH;

// ADCON0.3 -> AN1 selected, sample is from Charge amplifier
if(ADCON0.3){
    // Calculate ADC relative to calibration
    relSample = ADCData - calData;

    // If not yet inside of a step and threshold is reached
    if(!insideStepFlag && (relSample > STEPHIGHTH)){
        // Initialize accumulators for new step
        insideStepFlag = 1;
        stepStartTime = masterTimer;
        forceAccum = 0;
        forceMax = 0;
    }

    // If currently inside of a step:
    if(insideStepFlag){
        // update accumulator with current force
        forceAccum += relSample;
        // test upper bound
        if(relSample > forceMax)
            forceMax = relSample;
        // end of step?
        if(relSample < STEPLOWTH){
            // Bail if still writing to EEPROM
            if(!write_ready){
                insideStepFlag = 0;
                // Store useful info into temp variables
                wr_stepStartTime = stepStartTime;
                wr_stepLength = masterTimer - stepStartTime;
                wr_forceAccum = forceAccum;
                wr_forceMax = forceMax;
                // Signal EEPROM write
                write_ready = 1;
            }
        }
    }

    // Every 256 cycles, sample half-source reference
    if(masterTimer.low8 == 0){
        ADCON0.3 = 0; // Switch to AN0 for ADC
        GO = 1; // Begin ADC for AN0
    }
}

// !ADCON0.3 -> AN0 selected, sample is half-source voltage
else{
    // Update calibration data
    calData = ADCData;
    ADCON0.3 = 1; // Next sample from AN1, Charge Amplifier
}

// Timer1 overflow interrupt
if(TMR1IF && TMR1IE){

    // Next int in 5ms (312 cycles for 250 kHz)
    TMR1L = 0xC8;
    TMR1H = 0xFE;

    // Increment timestamp
    masterTimer++;

    // Begin an A/D conversion
    GO = 1;

    TMR1IF = 0;
}

```

```

        int_restore_registers;
    }

// Send to rs232 chip (MAX231)
void serialOut(uns8 txByte){
    while(!TXIF);           // Check to see if buffer is Empty
    TXREG = txByte;          // Transmit data from ADC to AUSART
}

// Set up reception/transmission from serial port
void enableAUS(){
    // Set up PORTB
    PORTB.2 = 1;             // Remove artifact of SSP
    TRISB.5 = 0;              // PortB5 output for Tx
    TRISB.2 = 1;              // PortB2 input for Rx
    // Set up AUSART
    TXEN = 1;                // Enable Transmission
    CREN = 1;                // Enable Continuous Receive
    SPEN = 1;                // Enable Port
    RCIF = 0;                // Clear Serial Receive Interrupt Flag
    RCIE = 1;                // Enable Serial Receive Interrupt
}

// Stop receiving/transmitting from serial port
void disableAUS(){
    while(!TRMT);            // Ensures that all transmission is complete
    PORTB.5 = 1;              // Trick rs232 into thinking there's no signal
    RCIE = 0;                // Disable Serial Receive Interrupt
    TXEN = 0;                // Disable Transmission
    CREN = 0;                // Disable Continuous Receive
    SPEN = 0;                // Disable Port
}

// Synchronous serial port for off-chip EEPROM
void enableSSP(){
    // Set up PORTB
    TRISB.1 = 1;              // PortB1 input for SDI
    TRISB.2 = 0;              // PortB2 output for SDO
    TRISB.4 = 0;              // PortB4 output for SCK
    // Set up SSP
    SSPEN = 1;                // Enable SSP
//    SSPIE = 1;                // Enable SSP Interrupt
}

void disableSSP(){
    PORTB.4 = 0;                // Clock at a low level
    SSPEN = 0;                // Disable SSP
//    SSPIE = 0;                // Disable SSP Interrupt
}

void sendByteSSP(char wrData){
    SSPBUF = wrData;
    while(!BF);
}

char receiveByteSSP(){
    SSPBUF = 0x00; // Dummy value
    while(!BF);
    return SSPBUF;
}

// High 4 bits to hex character
char hiHex(char binData){

    binData >= 4;
    binData &= 0x0F;

    if(binData < 10)
        return ('0' + binData);
}

```

```

        else
            return ('A' + binData - 10);
    }

// Low 4 bits to hex character
char lowHex(char binData){

    binData &= 0x0F;

    if(binData < 10)
        return ('0' + binData);
    else
        return ('A' + binData - 10);
}

void main(){

    // Set up internal Oscillator
    OPTION_REG = 0x87;      // Disable PORTB Pull-ups, TRM0 256 Prescalar
    // OSCCON = 0x78;          // 8MHz Internal Clock, internal RC for system clock
    OSCCON = 0x28;           // 250kHz, internal RC for system clock

    TRISB.0 = 0;             // Chip select of serial EEPROM
    TRISB.2 = 1;              // Ground RB2 through 100kOhm resistor

    // General AUSART Settings
    BRGH = 1;                // High baud rate setting
    CREN = 1;                // Enable Port, enable continuous receive
    // SPBRG = 51;             // 9.615 Kbaud for 8MHz Clock, BRGH = 1
    // SPBRG = 25;             // 19.231 Kbaud for 8MHz Clock, BRGH = 1
    // SPBRG = 16;             // 29.412 Kbaud for 8MHz Clock, BRGH = 1
    SPBRG = 12;               // 38.462 Kbaud for 8MHz Clock, BRGH = 1
    // SPBRG = 8;              // 55.556 Kbaud for 8MHz Clock, BRGH = 1

    // Set up Timer1
    T1CON = 0x0D;             // No Prescaler, enable Timer1, internal clock

    // SSP Settings
    CKE = 1;                  // Clocking mode compatible with the serial EEPROM
    SSPCON = 0x00;             // Data latched on rising edge,
                                // SPI Master mode, clock = OSC/4
    TRISB.0 = 0;               // Output for Chip select

    // Set up A/D
    TRISA |= 0x03;             // RA0, RA1 are inputs
    ANSEL = 0x03;              // Enable AN0, AN1
    ADCON0 = 0x00;             // Use system clock/2, Turn A/D on, sel AN0
    ADCON1 = 0x80;              // Right-justified output
    ADON = 1;                  // Prepare first A/D conversion

    // Enable Peripheral Interrupts (like serial int), Global Interrupt Enable
    INTCON = 0xC0;

    // Instantiate local variables
    uns24 addrEEPROM;          // Pointer to next byte in memory
    char readBuf = 0;            // Read buffer
    char deadStream = 0;          // Counter to detect end of EEPROM

    // Initialize global variables
    write_ready = 0;             // Do not write yet
    insideStepFlag = 0;           // Not yet inside of a step
    serialIn = '\0';             // No serial data to interpret
    sample_mode_flag = 1;

    // First int in 250ms
    TMR1L = 0xF7;
    TMR1H = 0xC2;

    TMR1IE = 1;                  // Enable Timer1 overflow interrupt
    ADIE = 1;                    // Enable ADC Conversion Completion Interrupt
}

```

```

PEIE = 1;                      // Enable peripheral interrupts
GIE = 1;                        // Global interrupt enable

// Drive chip select high
end_InstSSP

// Main program loop
while(1){

    // Low-frequency sample mode
    if(sample_mode_flag){

        // Check to see if serial module is connected
        if(PORTB.2){

            sample_mode_flag = 0;
            // Disable Timer1 and Interrupts
            TMR1IE = 0;
            TMR1IF = 0;
            T1OSCEN = 0;
            // Disable ADC
            ADIE = 0;
            GO = 0;
            ADON = 0;
            // Clear sample mode flags
            insideStepFlag = 0;
            write_ready = 0;

            // Wait for one second (250kHz clock)
            TMR1L = 0xDC;
            TMR1H = 0x0B;
            TMR1IF = 0;
            while(!TMR1IF);

            // 8MHz Internal Clock, internal RC for system clock
            OSCCON = 0x78;
            // Wait for stable frequency
            while(!OSCCON.2);

            enableAUS();
        }

        // Write step data to EEPROM
        if(write_ready){

            enableSSP();

            // Write enable instruction--obligatory
            begin_InstSSP
            sendByteSSP(WREN);
            end_InstSSP

            // 64us Delay in between instructions
            nop();
            nop();
            nop();
            nop();
            nop();

            // Page program instruction
            begin_InstSSP
            sendByteSSP(PP);
            // Address bytes (2 or 3)
            // Disable next line for ATMEL, only 16-bit addresses!
            sendByteSSP(addrEEPROM.high8);
            sendByteSSP(addrEEPROM.mid8);
            sendByteSSP(addrEEPROM.low8);
            // Write timestamp to EEPROM first (3 bytes)
            sendByteSSP(wr_stepStartTime.high8);
            sendByteSSP(wr_stepStartTime.mid8);
            sendByteSSP(wr_stepStartTime.low8);
            // Write step length to EEPROM (2 bytes)
        }
    }
}

```

```

        sendByteSSP(wr_stepLength.high8);
        sendByteSSP(wr_stepLength.low8);
        // Write force summation to EEPROM (3 bytes)
        sendByteSSP(wr_forceAccum.high8);
        sendByteSSP(wr_forceAccum.mid8);
        sendByteSSP(wr_forceAccum.low8);
        // Write max force to EEPROM last (2 bytes)
        sendByteSSP(wr_forceMax.high8);
        sendByteSSP(wr_forceMax.low8);
        end_InstSSP

        disableSSP();

        addrEEPROM += 10;           // Move forward 10 bytes in EEPROM

        PORTB.2 = 0; // Low so that serial signal can be recognized
        TRISB.2 = 1; // Return RB2 to input for sampling
                      // for serial module connection

        write_ready = 0;           // Clear write flag
    }

}

// Serial Communication mode
else {

    // No serial input -> do nothing
    if(serialIn == '\0')
        continue;

    // Ctrl + C: Calibration mode
    if(serialIn == 3){

        serialIn = '\0';

        ADON = 1;
        ADIE = 1;

        while(!serialIn){
            // ADC TEST PROCEDURE
            GO = 1;
            while(GO);

            masterTimer++;
            write_ready = 0;

            // Continuously displays the following quantities:
            // Current sample relative to calibration
            serialOut(hiHex(relSample.high8));
            serialOut(lowHex(relSample.high8));
            serialOut(hiHex(relSample.low8));
            serialOut(lowHex(relSample.low8));

            serialOut(' ');

            // Calibration data (should be near 0x0200)
            serialOut(hiHex(calData.high8));
            serialOut(lowHex(calData.high8));
            serialOut(hiHex(calData.low8));
            serialOut(lowHex(calData.low8));

            serialOut(' ');

            // Maximum force for current/last step
            serialOut(hiHex(forceMax.high8));
            serialOut(lowHex(forceMax.high8));
            serialOut(hiHex(forceMax.low8));
            serialOut(lowHex(forceMax.low8));

            serialOut(' ');
        }
    }
}

```

```

        // Accumulator value for last step
        serialOut(hiHex(forceAccum.high8));
        serialOut(lowHex(forceAccum.high8));
        serialOut(hiHex(forceAccum.low8));
        serialOut(lowHex(forceAccum.low8));

        serialOut(' ');

        //The letter S will appear if inside a step
        if(insideStepFlag)
            serialOut('S');

        serialOut(0x0D);
        serialOut(0x0A);

        // END ADC TEST PROCEDURE

    }
    serialIn = '\0';
}

// Ctrl + S: sample mode
if(serialIn == 19){

    serialOut(0x0D);
    serialOut(0x0A);
    serialOut('R');
    serialOut('e');
    serialOut('m');
    serialOut('o');
    serialOut('v');
    serialOut('e');
    serialOut(' ');
    serialOut('C');
    serialOut('a');
    serialOut('b');
    serialOut('l');
    serialOut('e');
    serialOut('.');
    serialOut(0x0D);
    serialOut(0x0A);

    disableAUS();

    // Turn RB2 into an input
    PORTB.2 = 0;
    TRISB.2 = 1;

    // 250kHz, internal RC for system clock
    OSCCON = 0x28;

    // Enable Timer1
    T1OSCEN = 1;

    // Wait for user to remove Serial Module
    // for one second (250kHz clock)
    while(PORTB.2);
    TMR1L = 0xDC;
    TMR1H = 0x0B;
    TMR1IF = 0;
    while(!TMR1IF);

    // Next int in 5ms (312 cycles for 250 kHz)
    TMR1L = 0xC8;
    TMR1H = 0xFE;

    // Enable Timer1 Overflow interrupt
    TMR1IF = 0;
    TMR1IE = 1;
}

```

```

// Enable ADC
ADON = 1; // ADC on
ADIE = 1; // Enable ADC Conversion Completion interrupt

sample_mode_flag = 1;

serialIn = '\0';
}

// Ctrl + E? Erase entire EEPROM (Spansion only)
else if(serialIn == 5){

    // Ask for confirmation
    serialOut('S');
    serialOut('u');
    serialOut('r');
    serialOut('e');
    serialOut('?');

    while( serialIn != 'y' &&
           serialIn != 'Y' &&
           serialIn != 'n' &&
           serialIn != 'N' ) ;

    if(   serialIn == 'n' ||
          serialIn == 'N' ) {
        serialOut(' ');
        serialOut('C');
        serialOut('a');
        serialOut('n');
        serialOut('c');
        serialOut('e');
        serialOut('l');
        serialOut('e');
        serialOut('d');
        serialOut('.');
        serialOut(0x0D);
        serialOut(0x0A);
        serialIn = '\0';
        continue; }

    GIE = 0;           // Disable all interrupts

    disableAUS();
    enableSSP();

    // Write enable instruction--obligatory
    begin_InstSSP
    sendByteSSP(WREN);
    end_InstSSP

    // 2us Delay in between instructions
    nop();
    nop();
    nop();
    nop();

    // Write status register, to write-enable all sectors
    begin_InstSSP
    sendByteSSP(WRSR);
    sendByteSSP(0x00);      // Free all sectors of flash
    end_InstSSP

    // 2us Delay in between instructions
    nop();
    nop();
    nop();
    nop();

    // Read status register
    begin_InstSSP

```

```

        sendByteSSP(RDSR);
        // Wait for status register to be written to
        while(receiveByteSSP() & 0x01);
        end_InstSSP

        // 2us Delay in between instructions
        nop();
        nop();
        nop();
        nop();

        // Write enable instruction--obligatory
        begin_InstSSP
        sendByteSSP(WREN);
        end_InstSSP

        // 2us Delay in between instructions
        nop();
        nop();
        nop();
        nop();

        // Bulk Erase instruction
        begin_InstSSP
        sendByteSSP(BE);
        sendByteSSP(0x00); // 24-bit fill value (not important)
        sendByteSSP(0x00);
        sendByteSSP(0x00);
        end_InstSSP

        T1OSCEN = 1; // Enable Timer1
        readBuf = 0x01;
        while((readBuf & 0x01)){
            // Wait for 33ms (8MHz clock)
            TMROIF = 0;
            TMR1L = 0x00;
            TMR1H = 0x00;
            while(!TMR1IF);

            // Read status register of EEPROM
            begin_InstSSP
            sendByteSSP(RDSR);
            readBuf = receiveByteSSP();
            end_InstSSP

            // Display periods while waiting.....
            disableSSP();
            enableAUS();
            serialOut('.');
            disableAUS();
            enableSSP();
        }

        T1OSCEN = 0; // Disable Timer1 and Interrupts

        // Confirm memory deletion to user
        disableSSP();
        enableAUS();
        serialOut('[');
        serialOut('0');
        serialOut('K');
        serialOut(']');
        serialOut(0x0D);
        serialOut(0x0A);

        serialIn = '\0';
        addrEEPROM = 0; // Reset EEPROM pointer
        masterTimer = 0; // Reset timestamp

        GIE = 1;
    }
}

```

```

// Ctrl + R? Dump raw contents of EEPROM
else if(serialIn == 18){
    GIE = 0;                      // Disable all interrupts
    addrEEPROM = 0;
    deadStream = 0;

    disableAUS();
    enableSSP();
    // Begin reading at address 0
    begin_InstSSP
    sendByteSSP(READ);
    sendByteSSP(0x00);
    sendByteSSP(0x00);
    sendByteSSP(0x00);
    // Read each byte and write to terminal until
    // we receive 10 blank bytes (a blank step)
    do {
        readBuf = receiveByteSSP();
        disableSSP();
        enableAUS();
        serialOut(readBuf);
        disableAUS();
        enableSSP();
        addrEEPROM++;
        if(readBuf == 0xFF)
            deadStream++;
        else
            deadStream = 0;
    } while(deadStream < 10);
    disableSSP();
    end_InstSSP
    enableAUS();

    // Go back to first blank byte
    addrEEPROM -= deadStream;
    // Clear pseudo-flag, wait for next keypress
    serialIn = '\0';
    GIE = 1;                      // Enable all interrupts
}

// Ctrl + X? Formatted Hex Dump (useful for Terminals)
else if(serialIn == 24){
    GIE = 0;                      // Disable all interrupts
    addrEEPROM = 0;
    deadStream = 0;

    disableAUS();
    enableSSP();
    // Begin reading at address 0
    begin_InstSSP
    sendByteSSP(READ);
    sendByteSSP(0x00);
    sendByteSSP(0x00);
    sendByteSSP(0x00);
    // Read each byte and write to terminal until
    // we receive 10 blank bytes (a blank step)
    do {

        deadStream = 0;

        // Read timestamp from EEPROM (3 bytes)
        wr_stepStartTime.high8 = receiveByteSSP();
        wr_stepStartTime.mid8 = receiveByteSSP();
        wr_stepStartTime.low8 = receiveByteSSP();
        // Read step length from EEPROM (2 bytes)
        wr_stepLength.high8 = receiveByteSSP();
        wr_stepLength.low8 = receiveByteSSP();
        // Read force summation from EEPROM (3 bytes)
        wr_forceAccum.high8 = receiveByteSSP();
        wr_forceAccum.mid8 = receiveByteSSP();
}

```

```

        wr_forceAccum.low8 = receiveByteSSP();
        // Read max force from EEPROM last (2 bytes)
        wr_forceMax.high8 = receiveByteSSP();
        wr_forceMax.low8 = receiveByteSSP();

        disableSSP();
        enableAUS();

        // Write timestamp to terminal (6 hex chars)
        serialOut(hiHex(wr_stepStartTime.high8));
        serialOut(lowHex(wr_stepStartTime.high8));
        serialOut(hiHex(wr_stepStartTime.mid8));
        serialOut(lowHex(wr_stepStartTime.mid8));
        serialOut(hiHex(wr_stepStartTime.low8));
        serialOut(lowHex(wr_stepStartTime.low8));
        serialOut(' '); // Insert a space
        // Write step length to terminal (4 hex chars)
        serialOut(hiHex(wr_stepLength.high8));
        serialOut(lowHex(wr_stepLength.high8));
        serialOut(hiHex(wr_stepLength.low8));
        serialOut(lowHex(wr_stepLength.low8));
        serialOut(' '); // Insert a space
        // Write force summation to terminal (6 hex chars)
        serialOut(hiHex(wr_forceAccum.high8));
        serialOut(lowHex(wr_forceAccum.high8));
        serialOut(hiHex(wr_forceAccum.mid8));
        serialOut(lowHex(wr_forceAccum.mid8));
        serialOut(hiHex(wr_forceAccum.low8));
        serialOut(lowHex(wr_forceAccum.low8));
        serialOut(' '); // Insert a space
        // Write max force to terminal (4 hex chars)
        serialOut(hiHex(wr_forceMax.high8));
        serialOut(lowHex(wr_forceMax.high8));
        serialOut(hiHex(wr_forceMax.low8));
        serialOut(lowHex(wr_forceMax.low8));
        // Insert Newline at end of each step
        serialOut(0x0D);
        serialOut(0x0A);

        disableAUS();
        enableSSP();
        addrEEPROM+=10;

        // Check to see if we have reached the end of data
        if(wr_stepStartTime == -1)
            deadStream++;
        if(wr_stepLength == -1)
            deadStream++;
        if(wr_forceAccum == -1)
            deadStream++;
        if(wr_forceMax == -1)
            deadStream++;

    } while(deadStream < 4);

    disableSSP();
    end_InstSSP
    enableAUS();

    // Go back to first blank byte
    addrEEPROM -= 10;
    // Clear pseudo-flag, wait for next keypress
    serialIn = '\0';
    GIE = 1;           // Enable all interrupts
}

// Echo character to terminal
else {
    serialOut(serialIn);
    // Enter? Proper newline
    if(serialIn==0x0D)

```

```
        serialOut(0x0A);
        // Backspace? Make last char disappear
    else if(serialIn==0x08){
        serialOut(' ');
        serialOut(0x08);
    }
    serialIn = '\0';
}

}
}
```

# **APPLICATION OF FERROFLUID AS A VALVE/PUMP FOR POLYCARBONATE MICROFLUIDIC DEVICES**

NSF Summer Undergraduate Fellowship in Sensor Technologies 2006

Helen Schwerdt (Biomedical Engineering) - Johns Hopkins University

Advisors: Professor Haim H. Bau, Jason Thompson

## **ABSTRACT**

Integrated pumps and valves make up an essential part of most microfluidic devices and produce increased control in the device. Some existing micro-pumps and valves utilized in microfluidics involve electrical, pneumatic, and thermal actuation. However, electrical micro-pumps and valves generally entail a complicated and expensive fabrication process, and are limited to samples free of any ions or charges. We attempt to create a phase-changing pump/valve out of a mixture of ferrofluid and paraffin wax, called “ferro-wax”, that is operated by magnetic force. We utilize oil based ferrofluid solution, which is made up of magnetic iron nanoparticles and surfactant. Ferrofluid may be used as immiscible slugs that are able to separate and pump water based samples in a polycarbonate (PC) channel modified to exhibit hydrophilic properties, without leaving noticeable film. We treat different polycarbonate samples using oxygen and argon plasma, spin-on-glass (SOG), and polyvinylpyrrolidone (PVP) to find the optimal procedure to produce low contact angles with water, while maintaining high surface energy and resistance to heat. Although most of our experiments involved using ferrofluid alone, melting the ferro-wax allows it to become a fluidic pump that can be easily manipulated with an external magnet. This same procedure allows the ferro-wax to operate as a valve, by directing the fluid across a branching side channel. We work on finding the most effective ratio of ferrofluid to wax, what properties the substrate require for this pump/valve to function properly, and how to create such a microfluidic device that integrates these ferrofluid based pumps/valves.

## TABLE OF CONTENTS

<b>1. INTRODUCTION</b>	<b>3-4</b>
<b>2. BACKGROUND AND PRELIMINARY PROCEDURE</b>	<b>4-7</b>
<b>3. SURFACE PROPERTIES OF CHANNEL SUBSTRATE</b>	<b>7-14</b>
<b>3.1 Observations</b>	<b>7-8</b>
<b>3.2 Surface Treatment</b>	<b>8</b>
<b>3.3 Oxygen Plasma</b>	<b>9-10</b>
<b>3.4 Argon Plasma</b>	<b>10</b>
<b>3.5 PVP</b>	<b>11-13 3.6</b>
<b>Spin-on Glass</b>	<b>13-14</b>
<b>4. FUTURE APPLICATIONS OF FERROFLUID VALVE AND PUMP</b>	<b>14-16</b>
<b>4.1 Ferro-wax Valve</b>	<b>14</b>
<b>4.2 Ferrofluid Pump</b>	<b>15-16</b>
<b>5. RESULTS AND DISCUSSION</b>	<b>16</b>
<b>6. RECOMMENDATIONS</b>	<b>16-17</b>
<b>7. ACKNOWLEDGEMENTS</b>	<b>17</b>
<b>8. REFERENCES</b>	<b>17</b>

## 1. INTRODUCTION

Microfluidics incorporates the study and application of precise control and manipulation of micro to nano liter volumes of fluid. Common materials that form the basis of microfluidic chips include silicon, polymers (PDMS, PMMA, PC), and glass. Handling liquids at a micro level differs greatly from a macro scale in that factors such as surface tension, energy, and fluidic resistance start to dictate the behavior of the fluid. Additionally, due to a low Reynold's number,

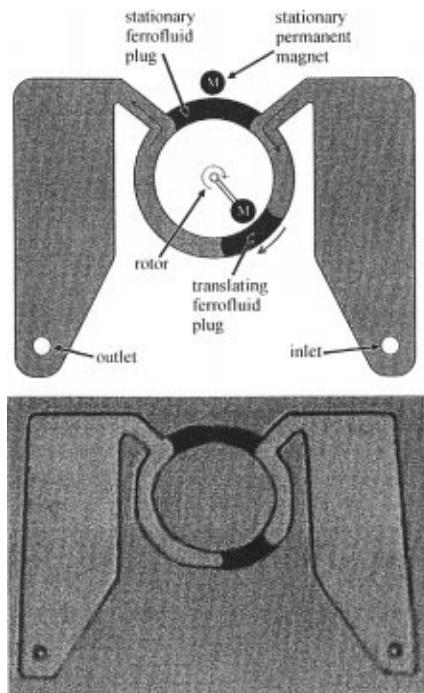
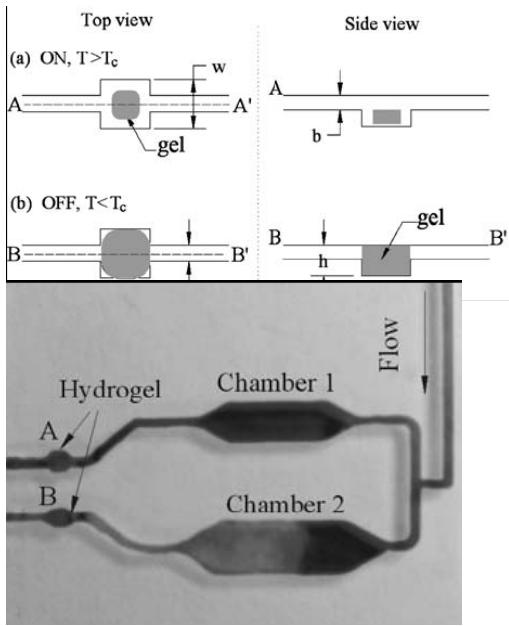
$$Re = \frac{v_s L}{\nu},$$

where  $v_s$  denotes fluid velocity,  $L$  is the characteristic length of the channel, and  $\nu$  is kinematic fluid viscosity defined as

$$\nu = \frac{\mu}{\rho},$$

where  $\mu$  is the dynamic viscosity, and  $\rho$  is the fluid density, fluid flow remains laminar. Microfluidic chip applications include detection within a solution, biological assay operations (such as protein or DNA analysis), and Lab-on-a-Chip (LOC) systems.

Most microfluidic chips integrate pumps and valves in order to produce increased control of the device. Common microfluidic control elements utilize electromechanical or electrokinetic effects such as peristaltic micro-pumps, which are operated by piezoelectric disks [1], or an electromagnetic actuator for a microvalve [2]. However, such devices generally entail a complicated fabrication process and are uneconomical, and therefore a pump/valve that can be driven simply by a magnet is more favorable. Electrically actuated pump/valves limit the device's functionality by disallowing electrically sensitive samples from being tested; such as solutions containing a noticeable charge or pH. Another example of a micro-pump/valve is a hydrogel valve that closes as liquid approaches it because the liquid causes the hydrogel to swell and shut the chamber. The valve opens again when heated [3]. An example of a liquid phase change valve is one that closes as a liquid is cooled below its freezing temperature and opens when heated [4]. In our experiments we chose polycarbonate (PC) because it is an inexpensive, transparent, biocompatible material that is easily machined into chips. Additionally, surface treatments of PC, an essential part of our design, necessitate less complicated procedures than those of other polymers.



**Figure 1.** Hydrogel valves in microfluidic channels (left) [3]. **Figure 2.** Ferrofluid valve controlled by rotating magnet and stationary magnet (right) [5].

## 2. BACKGROUND AND PRELIMINARY PROCEDURE

Part of the ferro-wax valve/pump we employ consists of an oil based ferrofluid, a magnetic fluid consisting of nanoparticles of 5% magnetic solid mixed with a 10% surfactant and 85% carrier, manufactured by FerroTec Corp (Lot#F021606A). The nanoparticles of ferrofluid behave like molecules in a paramagnetic gas, where in the absence of a magnetic field, the particles are randomly oriented [6]. When a magnetic field is applied, the tendency for dipole moments to align with field lines is somewhat overcome by thermal agitation. Magnetization of ferrofluid can be described by the following equations

$$\frac{M}{\Phi M_d} = \coth \alpha - \frac{1}{\alpha} \equiv L(\alpha),$$

$$\alpha = \frac{\pi}{6} \frac{\mu_0 M_d H d^3}{kT},$$

where  $\Phi$  is the volume fraction of magnetic solids ( $\pi n d^3 / 6$ ),  $M$  is the magnetization of

ferrofluid ( $\text{A m}^{-1}$ ),  $M_d$  is the domain magnetization of the particles ( $\text{A m}^{-1}$ ),  $L(\alpha)$  is the Langevin function,  $d$  is the particle diameter ( $\sim 10 * 10^{-9} \text{ m}$ ),  $H$  is the applied magnetic field ( $\text{A m}^{-1}$ ),  $k$  is Boltzmann's constant ( $1.38 * 10^{-23} \text{ N m K}^{-1}$ ),  $T$  is the absolute temperature (degrees Kelvin), and  $\mu_0$  is the permeability of free space ( $4\pi * 10^{-7} \text{ H m}^{-1}$ ). The ferrofluid exhibits saturation magnetization at  $L(\alpha) = 1$ . There are two mechanisms by which ferrofluid magnetization relaxes after modification of the applied field. Relaxation can take place by particle rotation in the liquid, which is represented by a Brownian rotational diffusion time  $\tau_B$ . *Extrinsic superparamagnetism*, where the Brownian mechanism dominates, occurs when

$$\frac{\tau_N}{\tau_B} \gg 1$$

It may also occur by rotation of the magnetic moment within the particle, which is known as the Néel mechanism, with a time constant,  $\tau_N$ . *Intrinsic superparamagnetism*, where the Néel mechanism dominates, occurs when

$$\frac{\tau_N}{\tau_B} \ll 1$$

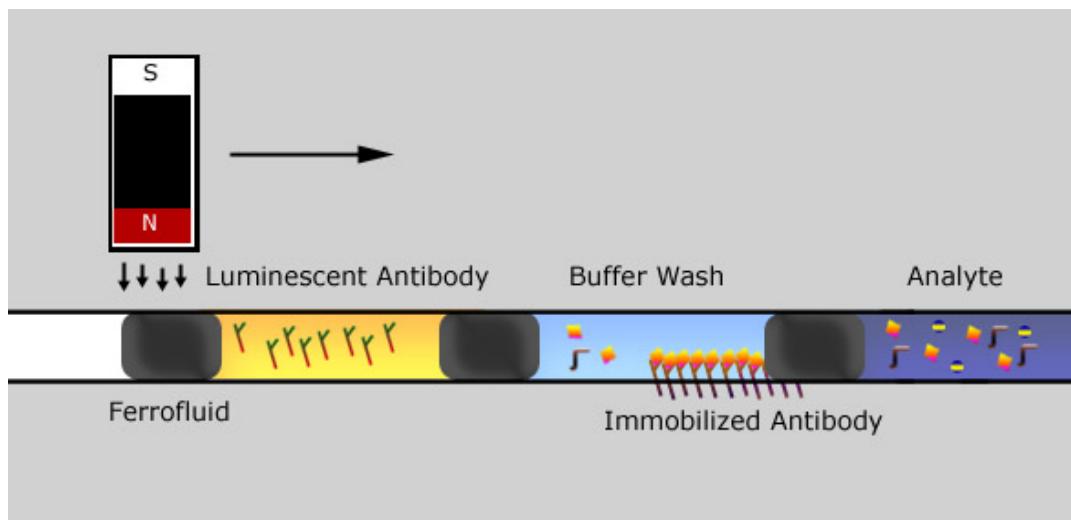
Microfluidic devices have employed the use of ferrofluid in the past for various applications. Some noteworthy examples include using ferrofluid to mix different solutions as well as to guide proteins or other molecules. Oil based ferrofluid is an attractive substance, as it is cheap, easily manipulated, immiscible with aqueous solutions, and conforms to the shape of the channel geometry.

The other portion of our valve/pump is made up of ordinary paraffin wax (Acros Organics, Lot#A017441501), such as may be found in birthday candles. This wax helps resolve any leakage difficulties ferrofluid has by itself. At room temperature the wax will remain solid and operate as an exemplary seal in most channels, and when heated above 50°C its melted form combined with ferrofluid can be easily repositioned along a channel with an external magnet.



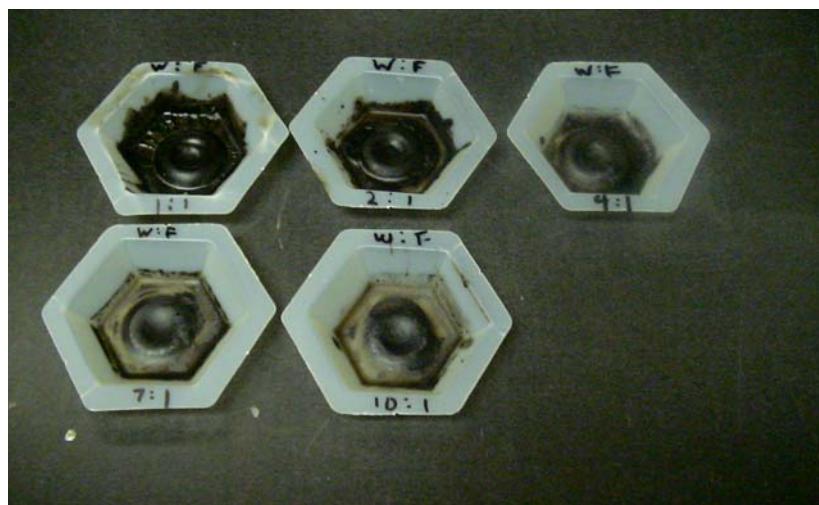
Ferrofluid by itself may also function as a pump and ideally we would like to make a train of fluids separated by immiscible slugs, propagated by one magnet. However, with this mechanism leakage issues need to be considered.

**Figure 3.** Ferrofluid in vial, with square magnet on side.



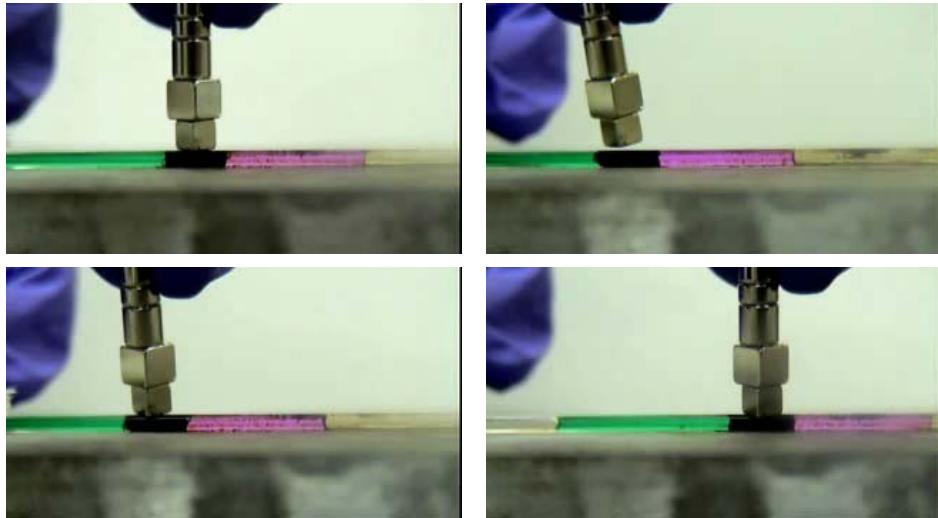
**Figure 4.** Fluid train prototype design.

To create ferro-wax we heated paraffin wax pellets on a Thermolyne hotplate at 50°C to 80°C. When the pellets had melted, we added ferrofluid and stirred the solution. We then let the ferro-wax cool and examined the stiffness of the solid for different ratios of ferrofluid and wax.

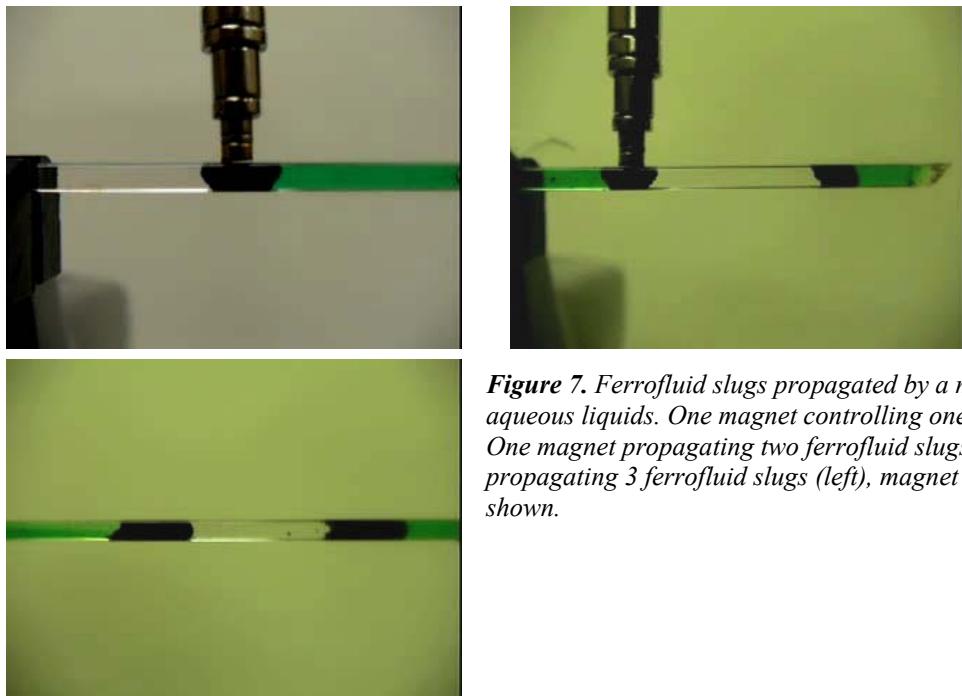


**Figure 5.** Ferrowax (wax : ferrofluid) solidified to shape of an external ring magnet at different mix ratios.

We tested ferrofluid and ferro-wax separately in glass capillaries to observe their pumping capability at the macro scale. The glass capillaries had square cross sections of approximately 2.2 mm on a side. Water was placed in between ferrofluid slugs and dyed liquids were used to note any leakage. Samples were inserted into the tube using a syringe and a magnet was placed directly above the surface of the channel. Advancing the slugs at slow velocities showed clean pumping action with no observable film or leakage. However, at faster speeds, we detected traces of film and some leakage.



**Figure 6.** Ferro-wax was inserted between two dyed aqueous solutions in a square shaped glass capillary and controlled by a magnet as shown above. The device was placed on top of a heating plate at 85°C.



**Figure 7.** Ferrofluid slugs propagated by a magnet, separating dyed aqueous liquids. One magnet controlling one ferrofluid slug (top left). One magnet propagating two ferrofluid slugs (top right). One magnet propagating 3 ferrofluid slugs (left), magnet and last ferrofluid slug not shown.

### 3. SURFACE PROPERTIES OF CHANNEL SUBSTRATE

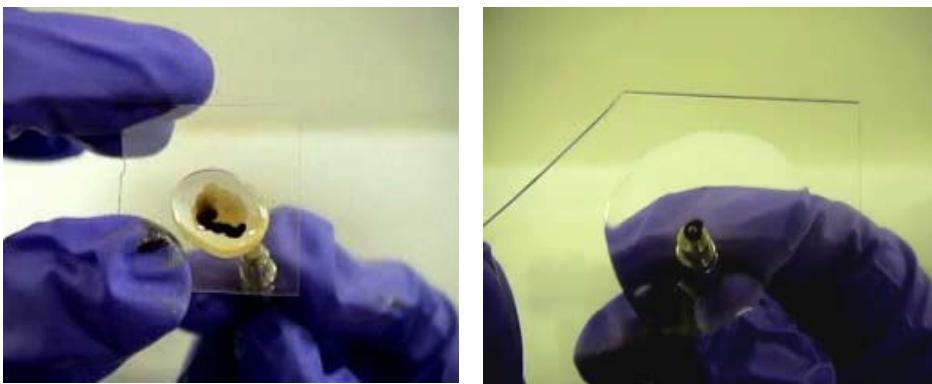
#### 3.1 Observations

We used contact angle estimates of water droplets as a measurement of hydrophilicity/hydrophobicity on the surfaces of glass, polycarbonate without any

treatment, and polycarbonate with various surface modifications. Polycarbonate and glass retain contact angles of about  $80^\circ$ , and  $35^\circ$  respectively [7]. Higher contact angles correspond to more hydrophobic materials whereas hydrophilic surfaces express low contact angles. Ferrofluid slugs in hydrophilic channels, such as glass, left no observable film during slow movement. However, in hydrophobic substrates, such as polycarbonate, apparent film deposition occurred.



**Figure 8.** Contact angle differences can be observed on PC (left) vs glass surfaces (right).



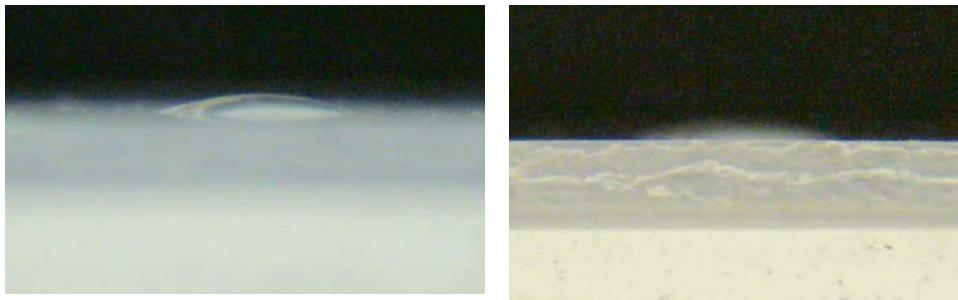
**Figure 9.** Because of differences in wetting properties of PC vs glass, ferrofluid floats above the surface on glass, creating a thin film of water underneath, and preventing a film from depositing (right). However, this is not the case for PC (left).

### 3.2 Surface Treatment

To make the channels fabricated in polycarbonate hydrophilic, such that ferrofluid slugs matched or surpassed the pumping action in glass tubes, we tested treatments by oxygen and argon plasma, PVP (polyvinylpyrrolidone), and SOG (spin-on-glass). Some issues arising with plasma treatment include complications with bonding. Bonding is a process in which a PC chip containing milled channels and another flat piece of PC are pressed together and heated to generate an enclosed microfluidic chip. Presently, our procedure for PC bonding involves the use of a bonding press in which chips must be heated to  $145^\circ\text{C}$ . However, plasma treatments are reversed at temperatures higher than approximately  $50^\circ\text{C}$  [8]. Therefore, some research and experiments were conducted to bond PC chips at low temperatures.

### 3.3 Oxygen Plasma

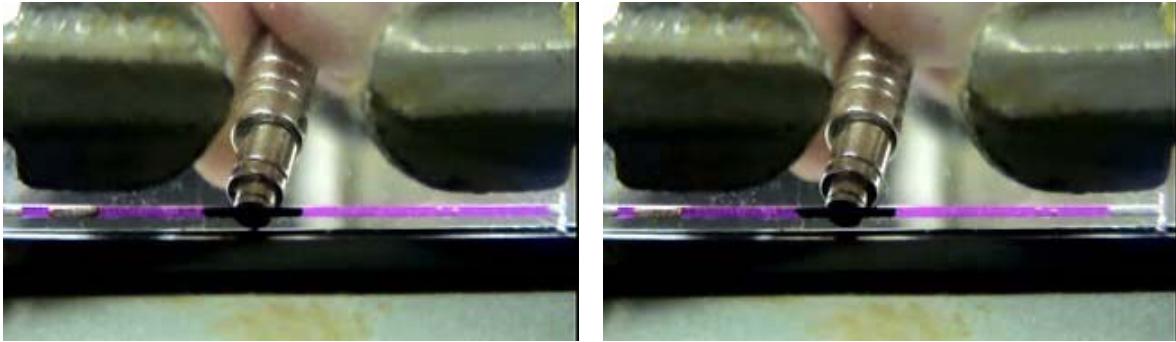
Oxygen plasma activation can yield surface contact angle measurements of 10° or less on PC [8]. However, difficulties arise from its limited stability towards washing, heating, and time extent of maintaining these low contact angles. Basic plasma treatment settings include power (watts), gas flow rate (sccm), and duration of treatment. At the most intense settings, corresponding to the highest W/F (power : flow rate) ratio, modified PC could be stored up to six months in aluminum foil before contact angles approached 20° [8].



**Figure 10.** Water droplets on oxygen plasma treated PC surface (50 W, 10 minutes, 3 sccm).

We tested our PC chips at various settings and our results agreed with earlier studies. Higher W/F ratios resulted in the lowest contact angles and greatest resistance to washing with ethanol. Our initial best two treatment settings were found to be 270 W, 3 sccm, for 3 minutes (W/F = 90), and 406 W, 3 sccm, for 3 minutes (W/F = 133). However, the latter treatment slightly melted the chip. In later PC channel treatments, to test our ferrofluid pump, we would use power in the range of 300-350 W for 3-5 minutes at the same flow rate. We based our conclusions on how hydrophilic a particular treatment was by visually estimating contact angles with the aid of a protractor. However, using too much power for extended periods warped the shape of the PC because higher power introduces more heat that can begin to melt the chip.

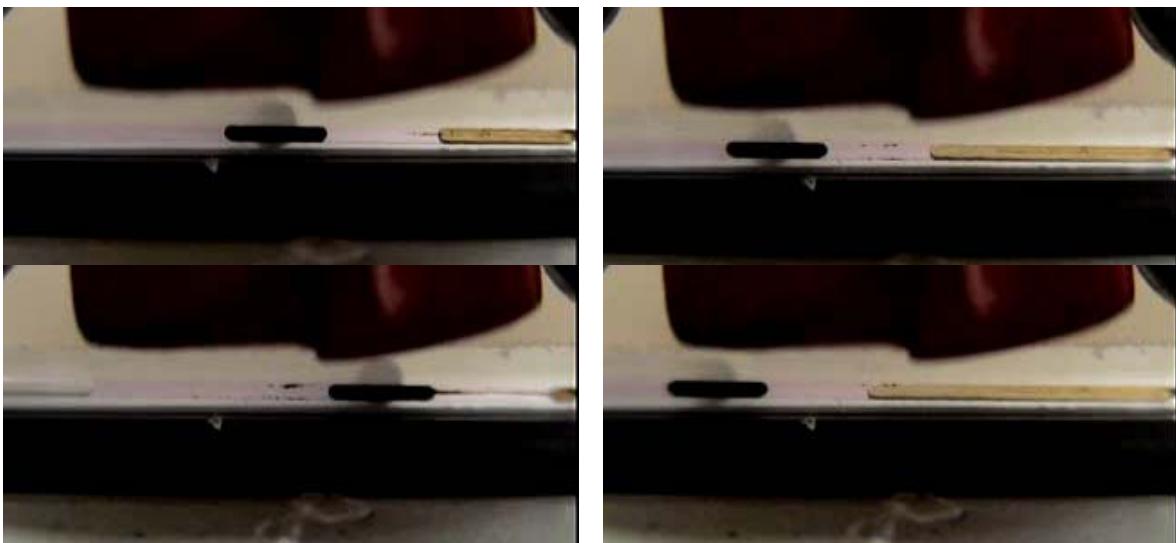
Finally we applied this treatment to a PC chip with a channel about 800x800 µm and fixed it to another piece of treated PC using a vice (to avoid thermally bonding the chip and maintain previous plasma treatments) to observe ferrofluid and dyed water in the modified channel. Ferrofluid surrounded by aqueous solutions left no noticeable film or leakage when manipulated carefully and slowly by an external magnet. Successful pumping action was also observed.



**Figure 11.** Oxygen plasma treated PC channels, bonded temporarily by vice. Ferrofluid guided by external magnet, surrounded by aqueous dyed liquid.

### 3.4 Argon Plasma

Argon plasma activates the surface of PC with physics similar to oxygen, and was performed in the same machine with comparable settings. However, Ar plasma does not activate the surface as well as oxygen and therefore film deposition was an issue. Water contact angles on argon plasma treated PC have been measured to be around 47° [9].

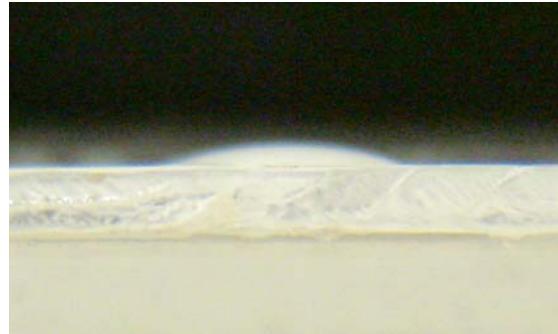


**Figure 12.** Argon plasma treated PC channels, temporarily bonded by a vice. Ferrofluid deposited some film, especially when moved at faster speeds.

Argon and oxygen plasma together have been used to successfully bond silicon to PC and is known to increase adhesion properties on the surface of PC [10]. Argon by itself has been shown to improve adhesion of PC to SiO<sub>2</sub> [11]. However, both of these methods were ineffective in PC-to-PC bonding in our experiments.

### 3.5 PVP

PVP (polyvinylpyrrolidone) is a polymer material used in many biomedical applications because of its biocompatibility within the human body [9]. It is used in toothpaste and can even be eaten in some cases. A powder form of PVP (Fluka K90, EC No. 2018004) was combined with water to form a homogeneous solution (1%, w/v). This solution can be used to further enhance the hydrophilic properties of a surface. Without previously activating PC, the PVP solution will not adhere to the surface because of its hydrophobicity. Therefore, in order to obtain a wetting surface, argon plasma (50 W, 5-10 minutes) was used to activate the surface, and afterwards, the chips were dipped in PVP solution a few times. Spinning the chips guaranteed a thin layer through the channel and on the surface. To dry and let the PVP solidify on the surface, the chips were heated at 50°C for approximately one hour.

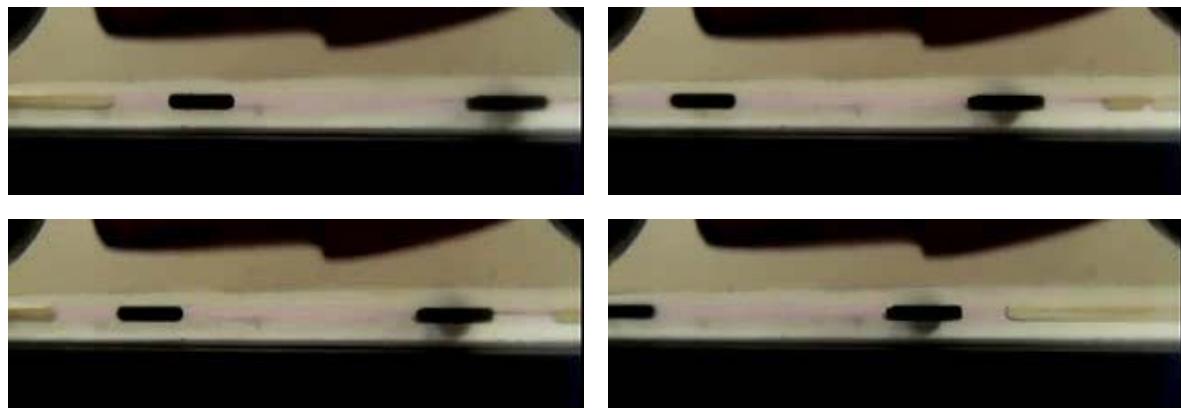


*Figure 13. Water droplets on PVP coated PC surface.*

We observed the cleanest pumping action of ferrofluid between aqueous solutions with this treatment. Furthermore, pumping multiple slugs of ferrofluid was possible with PVP coated PC channels. With previous treatments, our attempts to create a pumping mechanism with multiple immiscible slugs and one ferrofluid slug controlled by a magnet failed due to leakage across slugs not influenced by the magnetic field.

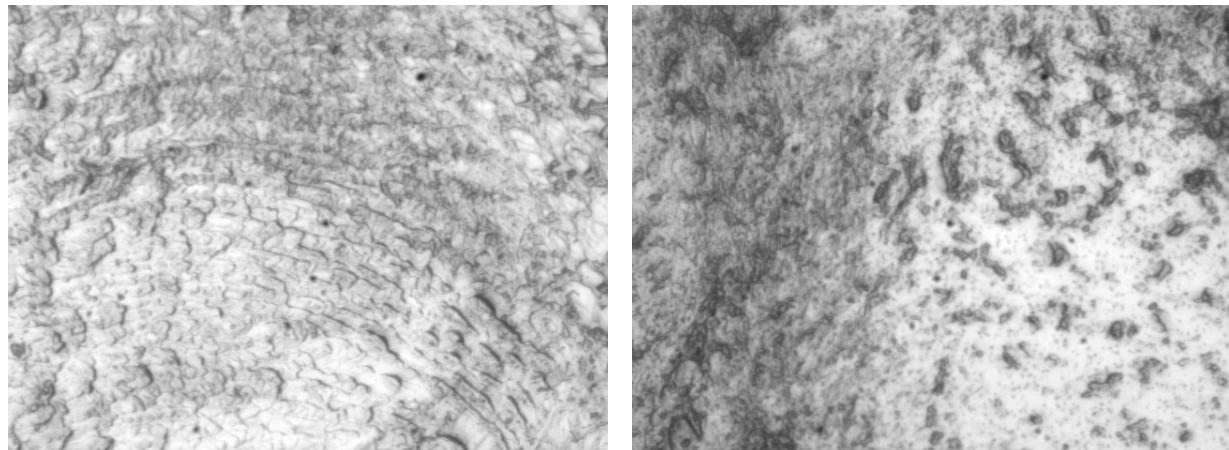


*Figure 14. Pumping action of one ferrofluid slug in argon treated, and PVP coated channel.*

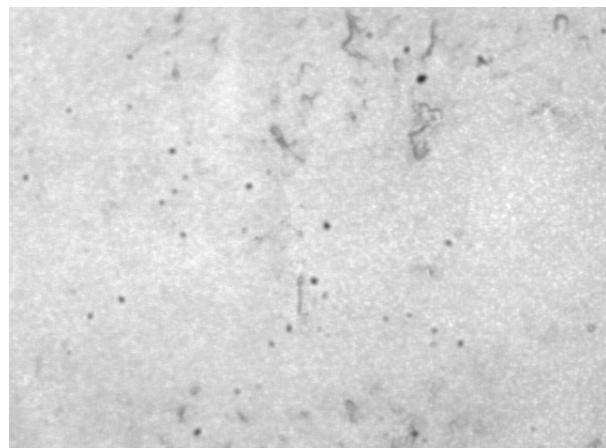


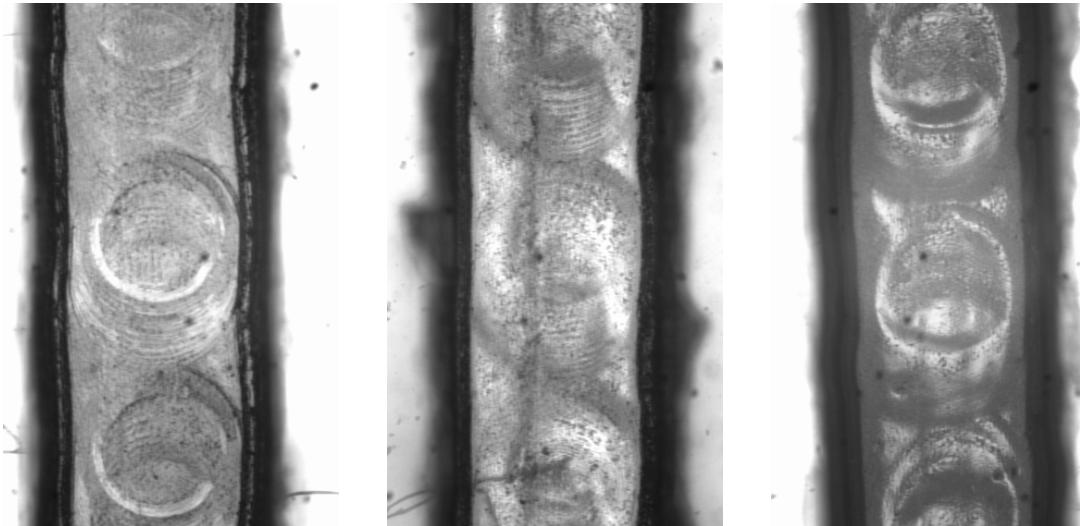
**Figure 15.** One magnet pumping 2 ferrofluid slugs and 3 water slugs in same channel as above.

A method of temporarily activating the surface of PC was by application of acetone prior to dipping in the PVP solution. Although acetone eats through PC, depending on the amount applied, PVP creates a coating that seems to cover pores and dents smoothly. Additionally with the acetone application, hydrophilic modification of a microfluidic channel in PC can be achieved following thermal bonding of the PC chips.



**Figure 16.** Drilled channel in regular PC, no modifications (top). Acetone applied in same channel for 5 seconds and air dried (top right). Acetone applied again and 5 seconds later, PVP solution inserted, then heated at 50°C for 15 minutes (right). 20x magnification.





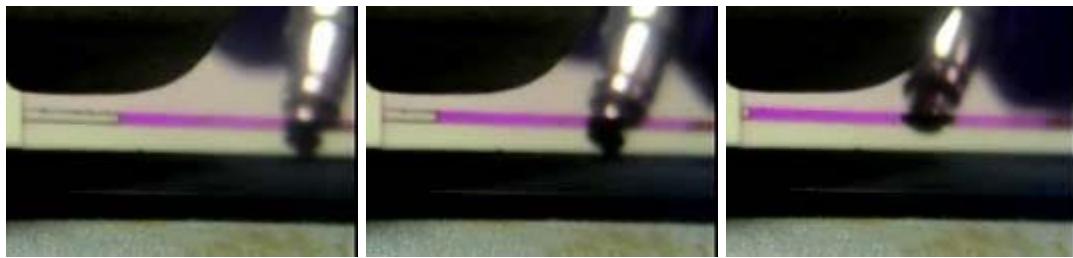
**Figure 17.** Same channels as above in order: unmodified, acetone applied, PVP treatment (from left to right). 2x magnification

A potential low temperature bonding approach for PC was found by directly bonding PVP dipped, argon plasma treated PC chips, and pressing them together with 300 lbf at 50°C. Although the seal in between the PC chips was very weak, the bond successfully enclosed the channel. Additional modifications could be made to better fasten the PC microfluidic chip following PVP sealing, such as screwing or taping the two pieces together.

Another possible low temperature bonding technique was to wet both surfaces with acetone and immediately press them together, before acetone could evaporate. This could be done at room temperature. However, complications arose in attempting to apply enough acetone on both sides of PC while simultaneously shielding the modified channels.

### 3.6 Spin-on Glass Treatment

Spin-on glass (SOG) is mainly used to coat silicon wafers with hard films of pure SiO<sub>2</sub> and was generously donated to us by Filmtronics (silicate, 700Å). Generally SOG is cured at temperatures of 750°C to 1000°C, and thin coats of 1000Å-1500Å are applied in order to avoid cracking. However, polycarbonate surfaces were heated up 140°C to prevent warping the PC. SOG was pipetted on to the PC surfaces and spun in order to create an even, thin layer (1600 rpm, 30 seconds). The PC was heated gradually to 140°C (about 30-45 minutes to heat) and maintained at this temperature for about 90 minutes.



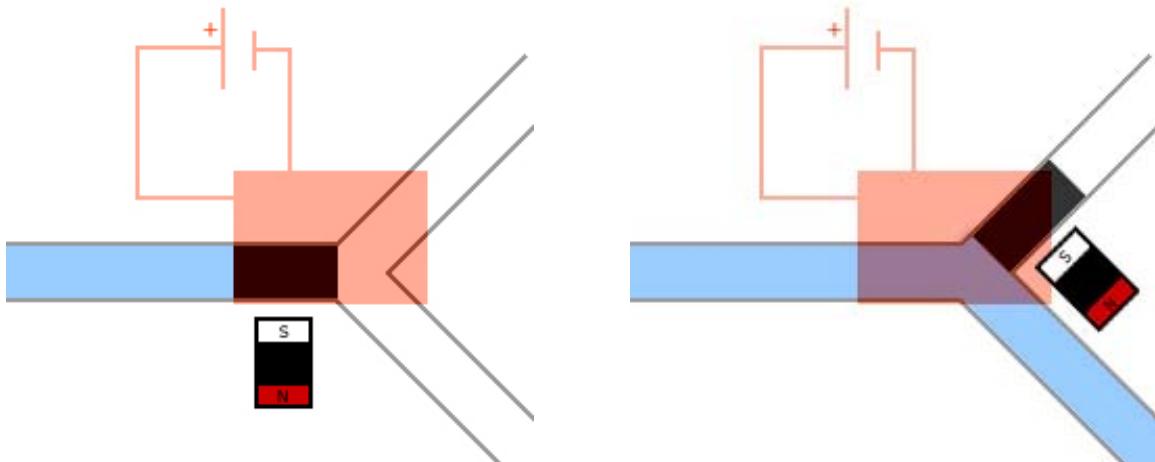
**Figure 18.** Ferrofluid slug guided by magnet in between two aqueous dyed liquid in SOG coated PC. Channel held together temporarily by vice.

Channels treated with SOG displayed similar results as the macroscale glass capillaries. Although the ferrofluid slug left no noticeable film, very slow movements were necessary.

#### 4. FUTURE APPLICATIONS OF FERROFLUID VALVE AND PUMP

##### 4.1 Ferro-wax Valve

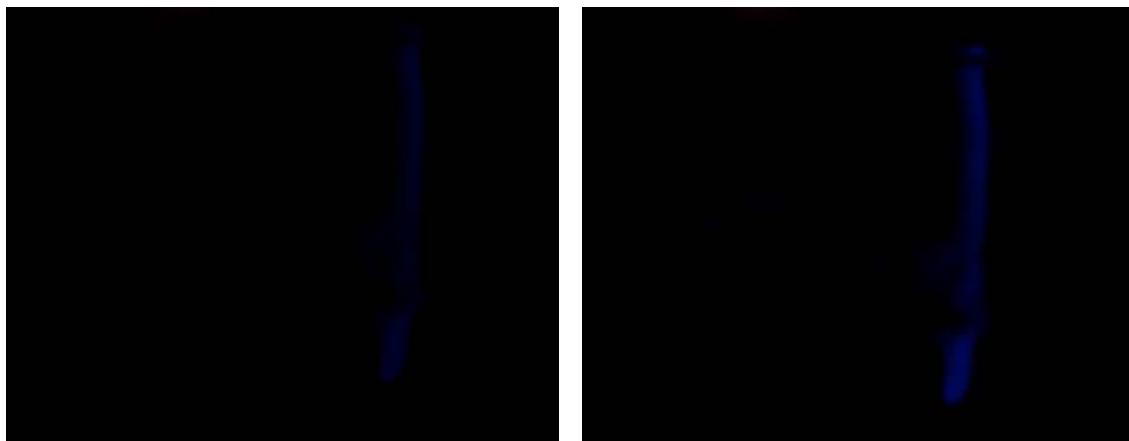
Although most of our experiments dealt with ferrofluid as a pumping device, ferro-wax can be applied to function as a prospective valve in microfluidic chips. As shown below, in this mechanism, a heating pad would be required to heat above 80°C, the region in which the ferro-wax valve will be situated and transfer itself by a magnet. However, with this technique, heat sensitive solutions could not be located in the vicinity of the heating pad.

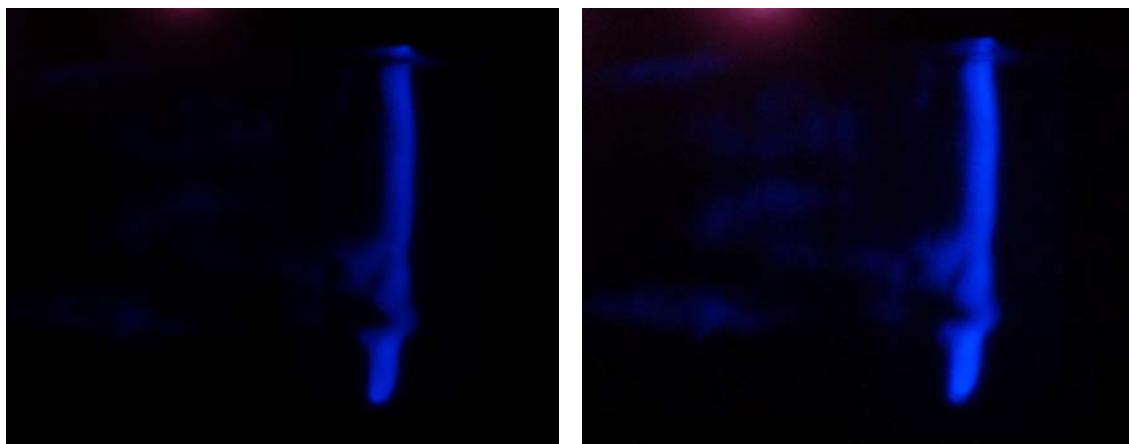


**Figure 19.** Ferro-wax valve prototype design, closed valve (left), open valve (right).

## 4.2 Fluid Train

As modeled in figure 4, ferrofluid can be used to direct fluids separated by immiscible slugs (possibly ferrofluid). In this example, a fluid train operates as a protein detection device utilizing chemiluminescence technology. This cost efficient technology is advantageous for microfluidic devices because it has a high level of sensitivity, does not generate hazardous waste, and the light produced by the chemical reaction lasts for a substantial duration. Such a device could be used to detect a particular protein which may suggest the presence of cancer. First, an aqueous buffer solution slug flows through the section containing immobilized antibodies or a membrane of antibodies directed against the target protein. The second slug contains the sample, possibly a body fluid, being investigated. As the sample flows across the antibody section, corresponding target protein molecules bind to these antibodies if they are present in this sample. To wash out any remaining molecules in the sample not attached to the immobilized antibodies, another buffer flows through this area. Next, tagged antibodies directed against the target protein bind to the available target protein sites, followed by another buffer. Finally, a slug of chemiluminescence detection reagents at the end of the train activates the tagged antibody sites to emit light. Using a camera suitable to detect very low intensity light signals, the light emitted by the chemiluminescent reaction can be captured. If no light is emitted, the target protein was not present in this sample, and the test would yield a negative result. Initially this chemiluminescence procedure was tested on a small piece of membrane outside of the microfluidic device. To test whether ferrofluid had a negative effect on the immobilized biological system, a small amount was applied on the membrane and rinsed off after a few minutes. Even after applying ferrofluid, light was detected from the membrane (this membrane was prepared and known to contain target proteins and their corresponding tagged antibodies). Pictures were taken of the membrane at different exposure times following application of the detection reagents, as seen in figure 20.





**Figure 20.** Membrane with antibodies and chemiluminescence emitting light at 38.2 s exposure (top left), 65.0 s (top right), 182.8 s (bottom left), 423.2 s (bottom right).

## 5. RESULTS AND DISCUSSION

Ferrofluid was shown to effectively pump aqueous solutions in a hydrophilized PC channel. Although ferrofluid could also pump in an unmodified PC channel, it coated the walls of the channel, which may be undesirable in a microfluidic device because of contamination of biological entities or inability to see through the surface of the device. Even though this may be true, initial experiments demonstrated that ferrofluid does not adversely affect the chemiluminescence procedure. The most appealing results were found using argon plasma or acetone activation, followed by a coating of PVP solution. Only with this treatment did ferrofluid have the ability to pump successive aqueous solutions and ferrofluid slugs without noticeable leakage. Additionally this treatment could be performed following thermal bonding of the PC chips by flowing acetone in the channel directly followed by PVP solution. For these reasons, fabricating a device that incorporates a fluid train, with the capability of utilizing chemiluminescence for protein detection, shows promise.

## 6. RECOMMENDATIONS

Although PVP treatment seemed to yield the best ferrofluid pump in polycarbonate channels, the duration of this modification should be tested. Additionally, further experimentation of bonding PC chips using PVP or plasma should be conducted. PVP seemed to seal the chips under certain conditions, but the bond was so weak that supplementary screws or adhesives to fix the two pieces together would be necessary. However, PVP application was only tested following argon plasma or acetone activation. Oxygen plasma also hydrophilizes PC and should therefore be investigated for PVP bonding. Our plasma bonding trials were limited to testing a few different settings, while higher power (around 1000 W) that was used in previous research of plasma bonding [10] was not tested. After finding the best method of hydrophilizing and bonding PC channels, exact leakage amounts across ferrofluid plugs should be measured and the maximum leakage permitted for applications of these pumps should be studied.

## 7. ACKNOWLEDGEMENTS

I would like to thank the SUNFEST coordinator, Professor Jan van der Spiegel, and my advisor, Professor Haim H. Bau, for giving me the opportunity to conduct research in a microfluidics laboratory at the University of Pennsylvania. I would also like to thank my mentor, Jason Thompson, for all of his guidance in this research and in the lab. In addition, I would like to thank Joseph Grogan for his help in the microfabrication lab. Furthermore, I would like to thank Erica Falls for constructing numerous chemiluminescence membranes for us to test. Finally, I would like to thank everyone else in Dr. Bau's lab for their support.

## 8. REFERENCES

1. J. G. Smits, *J. Sensors Actuators. A* **21** (1990) 203–206.
2. D. Bosch, B. Heimhofer, G. Muck, H. Seidel, U. Thumser and W. Welser, *J. Sensors Actuators. A37-A38* (1992) 684-692.
3. J. Wang, Z. Chen, M. Mauk, et al., *Biomedical Microdevices. 7:4* (2005) 313-322.
4. Z. Chen, J. Wang, S. Qian, H. Bau, *Lab Chip. 5* (2005) 1277-1285.
5. A. Hatch, E Kamholz, G. Holman, P. Yager, K. Bohringer, *J. Microelectromechanical Systems. 10:2* (2001) 215-221.
6. R. Rosensweig, *Ann. Rev. Fluid Mech. 19* (1987) 437-463.
7. H. Shadpour, H. Musyimi, J. Chen, S. A. Soper, *J. Chromatogr. A1111* (2006) 238-251.
8. A. Larsson, H. Derand, *J. Colloid and Interface Science. 246* (2002) 214-221.
9. L. Kessler, G. Legeay, et al., *J. Biomater. Sci. Polymer Edn. 10* (2003) 1135-1153.
10. S. Kitova, M. Minchev, G. Danev, *J. Optoelectronics and Advanced Materials. 7* (2005) 2607-2612.
11. S. Vallon, B. Drevillon, F. Poncin-Epaillard, J. Klemberg-Sapieha, L. Martinu, *J. Vac. Sci. Technol. A14:6* (1996) 3194-3201.

# **SPATIAL RESOLUTION OF CHONDROCYTE RESPONSE TO MECHANICAL SIGNALS FOR CARTILAGE TISSUE ENGINEERING**

NSF Summer Undergraduate Fellowship in Sensor Technologies

Xiaoning Yuan (Biomedical Engineering) – Duke University

Advisors: Dr. Robert L. Mauck, Alice H. Huang

## **ABSTRACT**

Cartilage is a supportive structure able to transmit and distribute loads within the body, as a result of its composition of chondrocytes and extracellular matrix (ECM). Damage to cartilage associated with osteoarthritis and injury is difficult to repair because of the tissue's avascular nature. Growth and maintenance of cartilage tissue *in vivo* result from the synergistic effects of biochemical molecules and mechanical stimulation. Specifically, the chondrocyte functions as a biological sensor that detects and transduces extracellular mechanical signals (mechanotransduction). However, many of the fundamental mechanisms of chondrocyte mechanotransduction remain unclear. The mitogen-activated protein kinases (MAPKs) are a family of molecules involved in responding to extracellular stimuli and regulating subsequent intracellular activities, a series of events known as the MAPK signaling cascade. In this study, a bioreactor capable of dynamic compression was used to stimulate ECM gene transcription in chondrocyte-laden agarose disks (2.25 mm × 5 mm Ø,  $30 \times 10^6$  cells/mL) cultured in chondrogenic medium with or without TGFβ-3 for one- and four-hour loading periods. After loading, disks were divided into inner core (2 mm Ø) and outer annulus regions and analyzed for ECM gene expression (aggrecan, type II collagen) by real-time PCR, and for active MAPK signaling molecules (ERK 1/2, SAPK/JNK, p38 MAPK) by Western blotting. In the future, this characterization of specific MAPK signaling cascades involved in chondrocyte mechanotransduction will be valuable in efforts to repair damaged cartilage tissue through functional tissue engineering solutions and in further understanding the role of the chondrocyte as a sensor.

## Table of Contents

1.	Introduction	3
2.	Background	4
3.	Experimental Methods	5
3.1	Chondrocyte and MSC Isolation	5
3.2	Chondrocyte Encapsulation in Agarose	6
3.3	Dynamic Loading	6
3.4	Real-time PCR	7
3.5	Western Blotting	8
4.	Results	9
4.1	Aggrecan and Type II Collagen Gene Expression Vary Regionally in Two-Day Cultured Chondrocyte-Laden Disks Subjected to One- and Four-Hour Dynamic Loading in CDM+ Media	9
4.2	Aggrecan and Type II Collagen Gene Expression is Sustained for Two-Day Cultured Chondrocyte-Laden Disks After One- and Four-Hour Dynamic Loading in CDM+ Media	10
4.3	Aggrecan Gene Expression is Elevated in the Outer Region for Three-Week Cultured Chondrocyte-Laden Disks Subjected to Four-Hour Dynamic Loading in CDM+ Media Relative to Two-Day Cultured Disks	10
4.4	Protein Extraction Levels and Total ERK1/2 Detection Vary with MSC Pellet Density	10
4.5	Phosphorylated ERK1/2 Protein Levels in MSC Pellets Vary Over Three-Day Culture in Chondrogenic Media	11
4.6	Protein Extraction from Intact Chondrocyte-Laden Agarose Disks and SDS-PAGE of Extracted Proteins Suggest Functionality of Revised Western Blotting Methods	12
4.7	Total p38 MAPK Protein Was Detected in Chondrocytes Extracted from the Outer Region of Agarose Disks	13
5.	Discussion and Conclusions	13
6.	Recommendations	15
7.	Acknowledgments	16
8.	References	16

## **1. INTRODUCTION**

Cartilage is a unique structure within the body that transmits and distributes physiological loads. In order to properly maintain these functions, cartilage tissue is dynamic and constantly remodels itself in response to mechanical changes in its environment. This load-bearing ability of cartilage results from a combination of specialized cells known as chondrocytes and a structural foundation of extracellular matrix (ECM) consisting of collagen, proteoglycans, and fluid. Specifically, one type, articular cartilage, is commonly found lining bone in joints [1]. Articular cartilage ECM contains predominantly type II collagen and proteoglycans composed of glycosaminoglycans with an aggrecan core protein [2].

The avascularity of cartilage causes the tissue to rely on diffusion for transport of nutrients and other cell materials and also limits the extent of native tissue repair possible [3]. This limited healing potential is especially a concern in the treatment of osteoarthritis, a common form of arthritis which is characterized by degradation of cartilage tissue. An estimated 21 million Americans suffer from stiffness, pain, and even loss of function associated with osteoarthritic joints [4].

As a result, functional tissue engineering solutions have since been sought to produce biological replacements for damaged native cartilage tissue. The field of tissue engineering combines the principles of multiple disciplines, including biochemistry and materials science, with the goal of restoring the lost or compromised functions of organs by producing new, healthy tissue. Tissue-engineered constructs are created through the combination of cells with biodegradable scaffolds, which are later implanted into the body [5]. While chondrocytes are the natural cell source for use in cartilage tissue engineering, recently mesenchymal stem cells (MSCs), a related cell type, have also seen application. MSCs are multipotential cells capable of differentiating along several lineages, such as chondrogenesis, adipogenesis, and osteogenesis. Although MSCs can be conditioned toward differentiation into chondrocytes, their functional properties have not been shown to be as robust as chondrocytes in studies involving their use in tissue-engineered constructs [6].

In cartilage tissue engineering, one key feature that constructs must exhibit in order to be functionally useful are mechanical properties similar to native cartilage. To achieve this, tissue-engineered constructs have been subjected to treatments such as culture in medium containing biochemical factors that promote chondrogenesis (differentiation into cartilage) and mechanical stimulation reminiscent of physiological conditions by loading regimes. In particular, this loading can be controlled through the design and use of bioreactors, devices capable of supporting biologically active environments [5].

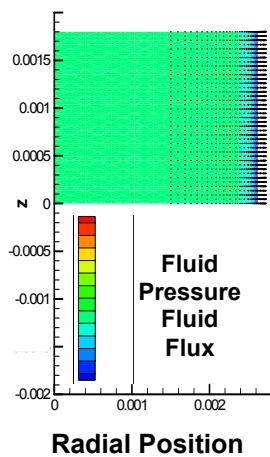
Despite the work of many in cartilage tissue engineering, results thus far demonstrate one fundamental problem: the functional properties of tissue-engineered constructs remain far less than that of native cartilage, even with enhancement by chondrogenic factors and mechanical stimulation. The search therefore continues for methods by which the functionality of tissue-engineered constructs can be increased to more closely resemble the original tissue.

## 2. BACKGROUND

Normal cartilage tissue growth and development *in vivo* is governed by both dynamic mechanical stimulation and the presence of biochemical growth factors. Results from previous studies suggest that cartilage is a tissue strongly influenced by mechanical changes in its environment and requires physiological loading to maintain its proper functions and homeostasis [2]. Mechanical loading of cartilage explants by methods such as static and cyclic compression and shear with variations in duration and frequency *in vitro* has been shown to influence ECM gene expression and remodeling of cartilage [7]. In general, these studies have shown that dynamic loading increases biosynthesis of ECM components, proteoglycans and collagen [8].

Loading studies with cylindrical constructs have also shown that regional variation in biological response exists between the inner core and outer annulus regions of constructs. Distinct patterns of gene expression and ECM biosynthesis in the inner and outer regions result from the different mechanical forces experienced by each region. Specifically, the inner core region experiences high fluid pressures, while the outer annulus experiences high fluid fluxes [9]. Thus, the regional response of chondrocytes to dynamic loading is of interest for study.

Hydrogel-based systems have been widely used as models for the study of chondrocyte response to biochemical and mechanical stimulation, because of their characteristic high water content reminiscent of native cartilage and capacity for even dispersion of cells within the scaffold during preparation. Agarose, a polysaccharide derived from red algae, is one type of hydrogel that is commonly used for chondrocyte encapsulation. The ability of chondrocyte-laden agarose hydrogel constructs to be consistently and uniformly manufactured makes them simple and popular model systems for use in studies [5]. Finite element modeling of agarose constructs as a biphasic (fluid and solid) medium yields the same regional variation in mechanical forces discussed previously (Fig. 1) [10].



**Figure 1** Finite element modeling of a hydrogel construct shows the mechanical forces experienced by a construct vary radially from its center. High fluid pressures predominate at and directly around the center of the construct, and high fluid fluxes are present at its outer edges. In the axial direction, fluid pressures within the inner core region and fluid fluxes are uniform.

While the pathways involved in mechanical loading are very complex, the biochemical mechanisms beginning with the application of a mechanical stimulus and leading to the response of individual cells remain largely unclear. In the cell, the mechanism by which extracellular, mechanical stimuli are interpreted and converted into biochemical signals is known as mechanotransduction. One well-studied family of biochemical signaling molecules, the mitogen-activated protein kinases (MAPKs), may be a critical component in chondrocyte mechanotransduction. Although their exact role is yet to be determined, these molecules are known to be involved in many cellular activities such as mitosis and gene expression, by sensing stimuli and initiating a cascade of events which lead to a biological response [11, 12].

Protein kinases such as MAPKs are capable of regulating biological events by phosphorylating other molecules. Currently, the MAPK family has been separated into five groups: extracellular signal-regulated kinases (ERKs) 1 and 2 (ERK 1/2), p38 MAPK and its isoforms ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ), c-Jun amino-terminal kinases (JNKs) 1, 2, and 3, and the less-studied ERKs 3, 4 and ERK5. In general, the MAPK signaling cascade consists of three conserved kinases, MAPK, MAPK kinase (MAPKK), and MAPKK kinase (MAPKKK), which act sequentially in response to stimuli such as mitogens, cytokines, and stress (Fig. 2) [11].

**Stimulus → MAPKKK → MAPKK → MAPK → Biological Response**

**Figure 2.** The general sequence of events in the MAPK signaling cascade.

Previous studies involving static compression of cartilage explants have suggested that several members of the MAPK family, ERK1/2, p38 MAPK, and SAPK/JNK, have time-dependent roles in chondrocyte mechanotransduction [13]. ERK1/2 was found to be rapidly activated at 10 min, followed by a rapid decay in activity, with a final elevation in activated ERK1/2 levels for at least 24 hr. Active p38 MAPK levels were observed to be transiently elevated at 10 min, while activated SAPK/JNK levels were greatest at 1 hr. Alternatively, cartilage explants exhibited elevated phosphorylated ERK1/2 levels when subjected to 0.5 and 16 hr of dynamic loading relative to static compression [14].

In this study, chondrocytes encapsulated in agarose hydrogel constructs were cultured in media containing TGF $\beta$ -3, a chondrogenic factor, and subjected to short-term dynamic loading regimes and evaluated for regional changes in gene expression and MAPK activation. The purposes of this study are to evaluate regional differences in ECM gene expression and to identify members of the MAPK family that are activated in response to dynamic loading and culture containing TGF $\beta$ -3, as well as to determine a time course for these gene expression and MAPK signaling events.

### **3. EXPERIMENTAL METHODS**

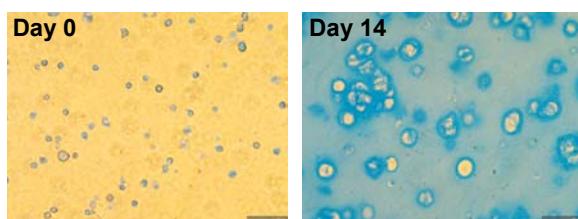
#### **3.1 CHONDROCYTE AND MSC ISOLATION**

Articular cartilage was harvested from the carpalmetacarpal joint of freshly slaughtered three to six-month old calves (Fresh Farms Beef, Rutland, VT, USA) and stored in serum-containing medium (SCM; high glucose Dulbecco's Modification of Eagle's Medium [hgDMEM], 10% fetal bovine serum, 1x penicillin/streptomycin/fungizone [PSF]). Chondrocytes were isolated from the cartilage pieces after digestion in a solution of hgDMEM containing 2.5 mg/mL pronase (Calbiochem, San Diego, CA) for one hour, followed by a 0.5 mg/mL collagenase type II (Sigma Chemicals, St. Louis, MO) solution for four hours at 37°C with stirring. The resulting cartilage digest was poured through a 70 µm sterile filter to remove any undigested parts, and cells were collected by centrifugation at 300×g for 5 min. The resulting cell pellet was resuspended in fresh hgDMEM, counted, and diluted to a concentration of  $60 \times 10^6$  cells/mL.

Bovine MSCs were harvested from the tibiae of freshly slaughtered three to six-month old calves (Fresh Farms Beef). Tibiae marrow pieces were mixed in an anti-coagulation solution (hgDMEM, 1x PSF, 300 U/mL heparin) and centrifuged at 300×g for 5 min after agitation. Resulting pellets were plated onto 30 cm plates and cultured in fresh SCM twice weekly. When cells reached confluence, sub-culturing was performed at a 1:3 expansion ratio in 60 cm plates. To harvest MSCs for studies, cells were washed with 1x PBS and incubated for 5 min at 37°C in 10 mL trypsin-EDTA. After cells detached from the plates, the reaction was terminated by adding 20 mL SCM, and the resulting solution was collected into a 50 mL tube and centrifuged for 5 min at 300×g. The cell pellet was resuspended in fresh hgDMEM, counted, and resuspended to a concentration of  $10^6$  cells/mL.

#### **3.2 CHONDROCYTE ENCAPSULATION IN AGAROSE**

The chondrocyte solution containing  $60 \times 10^6$  cells/mL was mixed with an equal volume of 4% type VII agarose in PBS and casted into a 2.25 mm thick gel at a final concentration of 2% agarose at  $30 \times 10^6$  cells/mL. The cell-seeded gel was cultured overnight in chemically defined medium (CDM-; 1x PSF, 0.1 mM dexamethasone, 50 mg/mL ascorbate 2-phosphate, 40 mg/mL L-proline, 100 mg/mL sodium pyruvate, 1x ITS+). Disks were punched from the gel the next day using a 5 mm Ø biopsy punch. For J1–3, the disks were dynamically loaded after two days of culture in CDM-. For J4, disks were pre-cultured for three weeks in CDM- prior to loading. After three weeks of culture in CDM-, chondrocytes within the agarose scaffold have begun to synthesize matrix and form cell-cell interactions (Fig. 3).

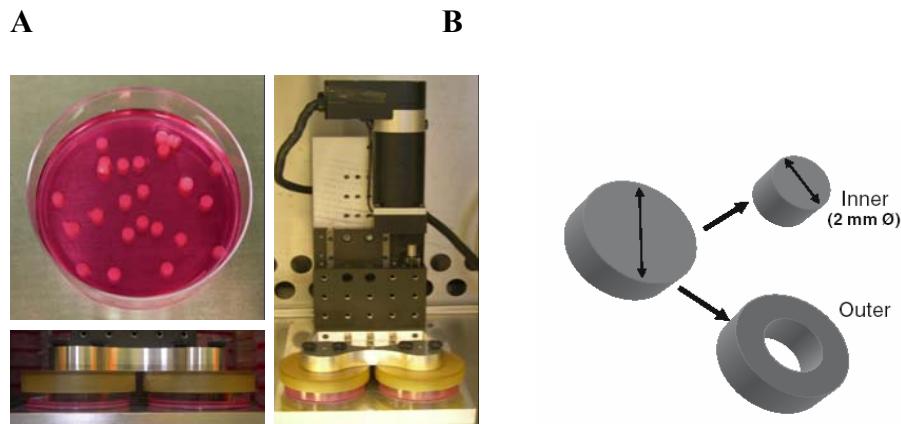


**Figure 3** Alcian Blue staining of proteoglycans shows increased proteoglycan content in chondrocyte-laden agarose disks at day 14 of culture in CDM- relative to day 0.

### 3.3 DYNAMIC LOADING

In preparation for loading, 4% Type VII agarose gels were cast in 60 cm plates, and 6 mm Ø holes were punched into the plates using a biopsy punch to serve as placeholders for the disks. Samples were loaded onto these plates directly before loading with 7 mL of chemically defined medium containing TGF- $\beta$ 3 (CDM+) at a final concentration of 10 ng/mL. TGF- $\beta$ 3 is a growth factor previously shown to enhance chondrogenesis (R&D Systems, Minneapolis, MN) [6]. Using a compressive loading bioreactor, dynamic (2% tare + 10% cyclic load at 1 Hz) and tare loading (2% tare) were applied to disks for one- or four-hour periods (Fig. 4A). For studies J1–3, both one- and four-hour loading durations were performed. For J4, disks were subjected only to the four-hour loading regime.

After loading, disks were harvested into a 2 mm Ø inner core and remaining outer annulus regions using a 2 mm biopsy punch for RNA and Western blotting (Fig. 4B). Subsequently, one sample in these studies represents 3–4 disks (inner or outer region) collected into one for RNA, and 3–5 disks for Western blotting. For RNA, samples for each loading condition were stored into one tube in 1 mL TRIZOL reagent (Invitrogen), while samples for Western blotting were flash-frozen in liquid nitrogen. Both RNA and Western blotting samples were stored at -80°C until further processing.



**Figure 4** A compressive loading bioreactor was used to apply specific dynamic loading regimes on chondrocyte-laden agarose disks ( $2.25\text{ mm} \times 5\text{ mm } \varnothing$ ) placed into Petri dishes (A). After loading, each disk was harvested into inner core (2 mm Ø) and outer annulus regions for further analysis (B).

### **3.4 REAL-TIME PCR**

Aggrecan and type II collagen gene expression were evaluated by real-time PCR. RNA was isolated from the TRIZOL-stored samples by a double chloroform extraction method. Samples were initially grinded by a motorized pestle and extracted by an initial 200 µL of chloroform. After centrifugation of vortexed samples at 4°C for 15 min at 12,000 rpm, the aqueous layer containing RNA was transferred to a new tube along with an additional 500 µL of TRIZOL reagent. Another 200 µL of chloroform was added to the sample, and the remainder of the RNA extraction was performed according to the manufacturer's instructions. The final purified RNA product was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Rockland, DE).

Complementary or cDNA was synthesized with random hexamers for 2000 ng of product using the First Strand cDNA Synthesis Kit (Invitrogen), according to manufacturer's protocol. 7 µL of RNA sample was added if the concentration was insufficient to produce 2000 ng of cDNA. The resulting cDNA product was diluted to 50 or 100 µL in DEPC water prior to use in real-time PCR.

Real-time PCR was performing using an Applied Biosystems 7300 Real-time PCR machine. The real-time PCR reaction consisted of cDNA sample, SYBR Green, and forward and reverse primers specific for bovine aggrecan, type II collagen, and GAPDH. After the reactions were completed, the initial quantities of aggrecan and type II collagen gene transcripts were determined using the respective standard curves and normalized by corresponding GAPDH values to account for variations in starting cDNA concentrations. Data for dynamically loaded samples were further normalized by data for tare loaded samples of the same conditions.

### **3.5 WESTERN BLOTTING**

To extract protein, frozen samples were lysed in 100 µL RIPA buffer (1x PBS, 1% nonidet P-40, 0.5% sodium deoxycholate) and sonicated twice on ice for five seconds each. In addition, chondrocyte-laden agarose disks were ground using a motorized pestle prior to sonication. Lysed samples were then centrifuged at 12,000 rpm for 10 min at 4°C, and the supernatant was collected. An additional 25 µL RIPA buffer was added to the collected supernatants, and the concentrations of protein within these samples were measured using the BCA Protein Assay Kit (Pierce).

Neat samples were diluted to a 20 µL volume in RIPA buffer, such that the concentration of all samples matched the sample with the lowest concentration of protein. Samples were then mixed with 5 µL of 5x ImmunoPure Lane Marker Non-Reducing Sample Buffer (Pierce) containing 100 mM β-mercaptoethanol, and heated at 99°C for six minutes to denature the protein.

Denatured protein samples were resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on a 12% resolving and 5% stacking gel of 0.75 mm thickness for ~1 hr at 130 V. After SDS-PAGE, proteins from the gel were electrophoretically transferred to a polyvinylidene fluoride membrane for 1 hr at 100 V and 30 min at 60 V. After transfer, the gel was stained using Gel Code Blue (Pierce) to visualize protein banding patterns.

The transferred membrane was incubated at 25°C for 1 hr in 10 mL blocking buffer (5% w/v nonfat dry milk in 1x PBS/0.1% Tween-20) with gentle agitation and washed three times for 5 min each with 1x PBS/0.1% Tween-20 (PBS-T) to remove residual blocking buffer prior to incubation over night at 4°C in a primary antibody solution (10 µL primary antibody in 10 mL 5% bovine serum albumin/PBS-T) with gentle agitation. The membrane was washed three times for 5 min each with PBS-T the next day to remove any unattached primary antibody, before incubation for 1 hr at 25°C in a secondary antibody solution (5 µL secondary antibody in 10 mL blocking buffer) with gentle agitation. Finally, the membrane was washed three times for 5 min each with PBS-T. To develop the blot, the membrane was mixed with a 1:1 solution of SuperSignal West Pico Chemiluminescent Substrate (Pierce) and exposed to film using a Kodak RP-X-Omat automatic X-ray film processor.

For MSC1, harvested MSCs were collected at  $10^6$ ,  $5 \times 10^5$ ,  $2.5 \times 10^5$ ,  $10^5$ , and  $5 \times 10^4$  cells/pellet and pelleted at 10 min at 300×g. These pellets were immediately flash-frozen in liquid nitrogen and stored at -80°C until they were probed for BiP/GRP78 (BD Biosciences), a constitutively expressed chaperone protein, by Western blotting.

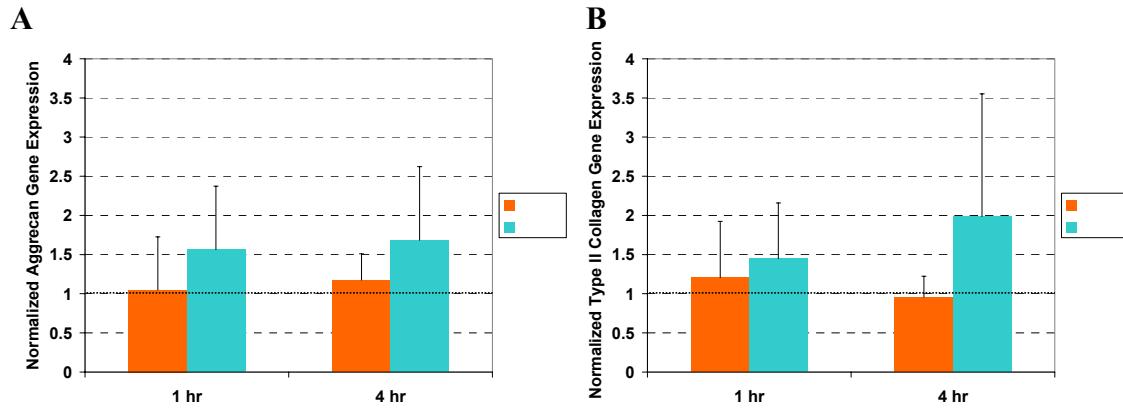
For MSC2, cells were pelleted at a density of  $2.5 \times 10^5$  cells/pellet for 10 min at 300×g. These pellets were cultured in either CDM- or CDM+ to induce chondrogenesis for 24 hr, with 12-14 pellets collected per sample at 0, 1, 3, 4, 8, and 24 hr for both media conditions [15]. Collected pellets were flash-frozen in liquid nitrogen and stored at -80°C. For these samples, 300 µL RIPA buffer was used to lyse the cells, since the total number of cells per sample was much higher than in MSC2. Finally, the lysed samples were probed for total ERK1/2 (Cell Signaling Technologies) by Western blotting.

For J1, protein extracted from the collected chondrocyte-laden agarose disks was probed for total p38 MAPK (Cell Signaling Technologies) by Western blotting.

#### 4. RESULTS

##### **4.1 AGGREGAN AND TYPE II COLLAGEN GENE EXPRESSION VARY REGIONALLY IN TWO-DAY CULTURED CHONDROCYTE-LADEN DISKS SUBJECTED TO ONE- AND FOUR-HOUR DYNAMIC LOADING IN CDM+ MEDIA**

Real-time PCR was used to quantify aggrecan and type II collagen gene expression after one- and four-hour tare and dynamic loading regimes in CDM+ media (Fig. 5). All data for dynamically loaded samples were normalized by data for tare loaded samples with the same loading time and media conditions. A normalized value of 1 therefore corresponds to no difference in gene expression between the tare and dynamically loaded samples. For both genes, gene expression levels for inner region samples are ~1, while the outer region samples exhibited higher levels of gene expression (~1.5 – 2). Thus, spatial resolution between the inner and outer regions of chondrocyte-laden agarose disks is substantially different for investigation, because the two regions exhibit different levels of gene expression.



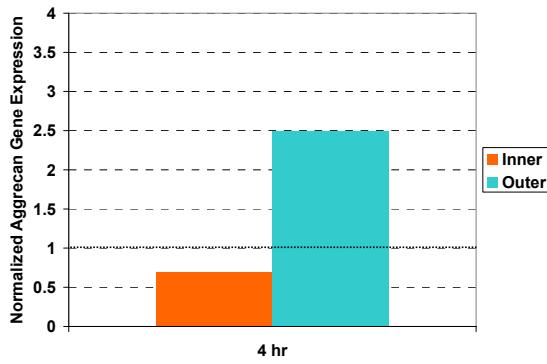
**Figure 5** Aggrecan (*A*) and type II collagen (*B*) gene expression vary between the inner and outer regions of two-day cultured samples subjected to dynamic loading for one or four hours in CDM+ media (n = 3).

#### 4.2 AGGREGCAN AND TYPE II COLLAGEN GENE EXPRESSION IS SUSTAINED FOR TWO-DAY CULTURED CHONDROCYTE-LADEN DISKS AFTER ONE- AND FOUR-HOUR DYNAMIC LOADING IN CDM+ MEDIA

For both aggrecan and type II collagen, gene expression levels for inner region samples subjected to one- or four-hour dynamic loading times do not vary considerably (Fig. 5). Likewise, outer region samples do not exhibit appreciable differences in aggrecan and type II collagen gene expression levels between one- or four-hour loading times. This similarity suggests that the level of gene expression activity is sustained from one to four hours of loading, and more importantly, that activity within the samples does not substantially change between one and four hours of loading.

#### 4.3 AGGREGCAN GENE EXPRESSION IS ELEVATED IN THE OUTER REGION FOR THREE-WEEK CULTURED CHONDROCYTE-LADEN DISKS SUBJECTED TO FOUR-HOUR DYNAMIC LOADING IN CDM+ MEDIA RELATIVE TO TWO-DAY CULTURED DISKS

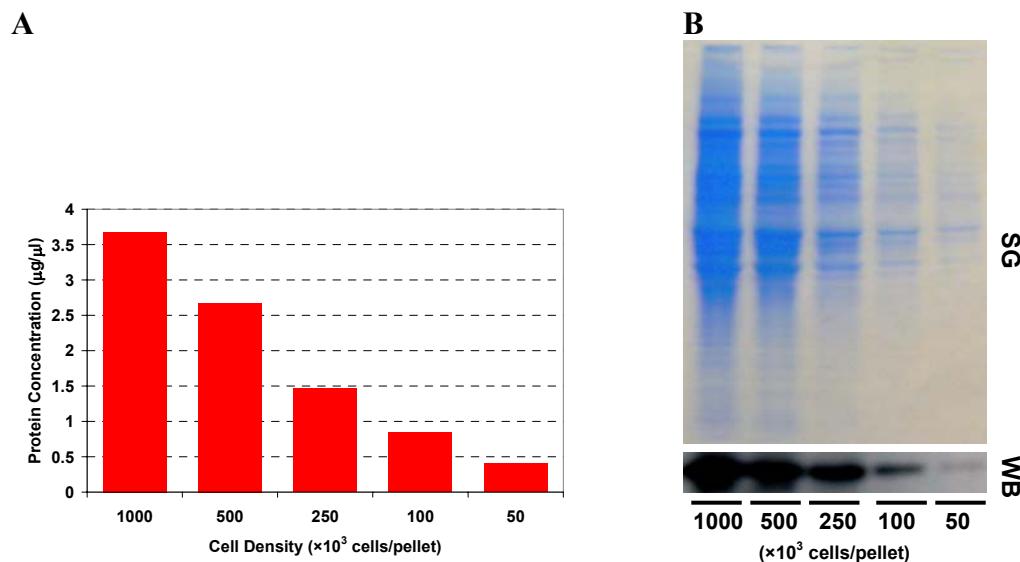
The differences observed in two-day cultured samples between aggrecan gene expression levels of the inner and outer regions were maintained in three-week cultured samples (Fig. 6). Samples after three weeks of culture exhibit elevated levels of aggrecan gene expression in the outer region (2.494) relative to samples after two days of culture (1.683). Type II collagen gene expression (*not shown*) was also assessed for three-week culture samples, but inner and outer region levels were not observed to be substantially different.



**Figure 6** Aggrecan gene expression varies between the inner and outer regions of three-week cultured samples subjected to dynamic loading for one or four hours in CDM+ media.

#### 4.4 PROTEIN EXTRACTION LEVELS AND TOTAL ERK1/2 DETECTION VARY WITH MSC PELLET DENSITY

In MSC1, the concentrations of protein extracted from MSC pellets of various cell densities were quantified (Fig. 7A). A linear positive correlation was found between pellet cell density and the concentration of protein extracted from the pellet, with the greatest concentration of protein extracted ( $3.679 \mu\text{g}/\mu\text{L}$ ) from the sample of the highest cell density ( $1000 \times 10^3$  cells/pellet). Gel electrophoresis and staining of the samples further supported this relationship, as the intensity of staining was greatest for the sample of the highest cell density and lowest for the sample of the lowest cell density ( $50 \times 10^3$  cells/pellet) (Fig. 7B – SG). Lastly, BiP/GRP78 protein levels as shown by band intensity were also found to be greater in samples of higher cell density (Fig. 7B – WB). Thus, protein levels can be quantified through a BCA assay and correlated with staining intensity on polyacrylamide gels and band intensity on Western blots.

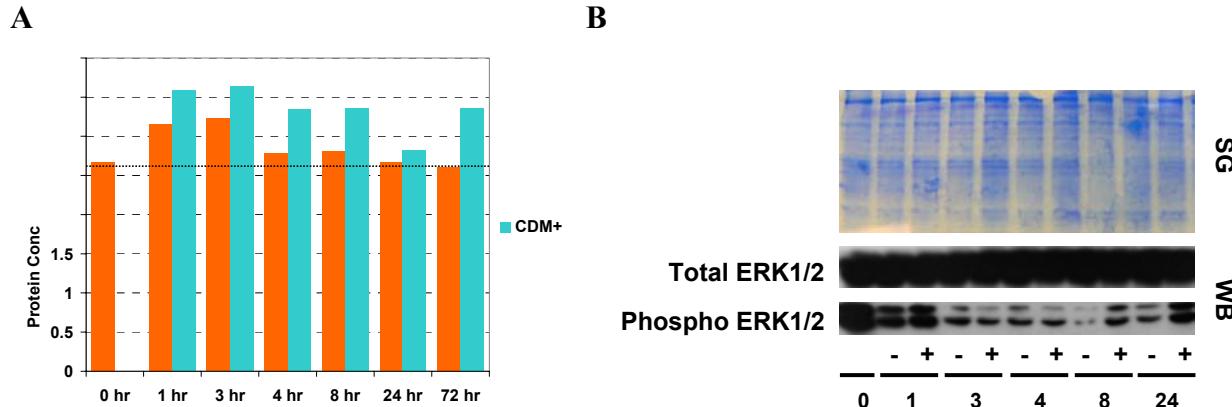


**Figure 7** The concentration of protein extracted from MSC pellets exhibits a linear and positively correlated relationship with cell density (A). Equal volumes of neat samples

loaded onto a polyacrylamide gel show distinct variation in staining intensity, with the greatest staining associated with the sample of highest cell density (*B*). Bands on the Western blot corresponding to BiP/GRP78 detection of the samples show the same relationship between chemiluminescence intensity and sample cell density.

#### **4.5 PHOSPHORYLATED ERK1/2 PROTEIN LEVELS IN MSC PELLETS VARY OVER THREE-DAY CULTURE IN CHONDROGENIC MEDIA**

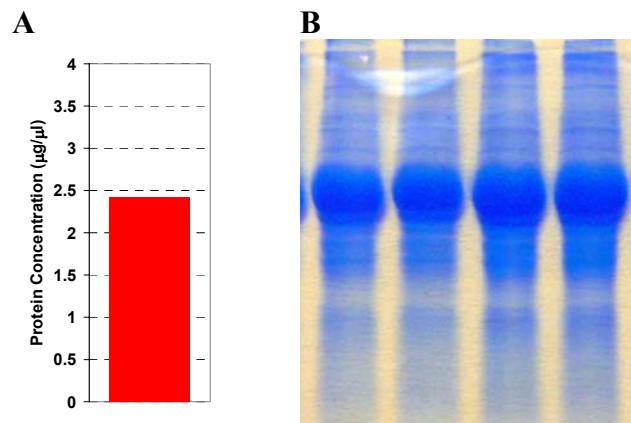
The concentration of protein extracted from the MSC pellets collected at various time points during three days of culture in CDM- or CDM+ were all found to be at least 2.5 µg/µL (Fig. 8A). All samples were diluted to match the sample with the lowest concentration (2.606 µg/µL) for Western blotting, and gel staining verified that these samples have similar staining intensity and were loaded with approximately the same amount of protein in each well (Fig. 8B – SG). Total ERK1/2 levels were observed to be high for all samples, although phosphorylated ERK1/2 levels vary with time and media conditions over the 24 hours of culture (Fig. 8B – WB). Specifically, phosphorylated ERK1/2 levels are highest at early and late times (1 hr; 8, 24 hrs) and lower between these times. The presence of TGFβ-3 in CDM+ media is known to stimulate chondrogenesis in MSCs, and the appearance of stronger band intensities for samples cultured in CDM+ suggests that TGFβ-3 is also related to ERK1/2 activation in MSCs. From these results, cell lysis methods to extract protein from MSC pellet tissue were verified, and the functionality of total and phosphorylated ERK1/2 antibodies was confirmed for subsequent use with Western blotting of chondrocyte-laden agarose disks.



**Figure 8** The concentrations of protein extracted from MSC pellets exceed 2.5 μg/μL for all samples, with the black dotted line designating the sample of lowest concentration (A). All other samples were diluted to match this sample for Western blotting. Gel staining after SDS-PAGE confirms equal loading of protein sample in each well (B). Total ERK1/2 protein levels are constant for all samples, while phosphorylated ERK1/2 protein levels vary on the Western blot.

#### 4.6 PROTEIN EXTRACTION FROM INTACT CHONDROCYTE-LADEN AGAROSE DISKS AND SDS-PAGE OF EXTRACTED PROTEINS SUGGEST FUNCTIONALITY OF REVISED WESTERN BLOTTING METHODS

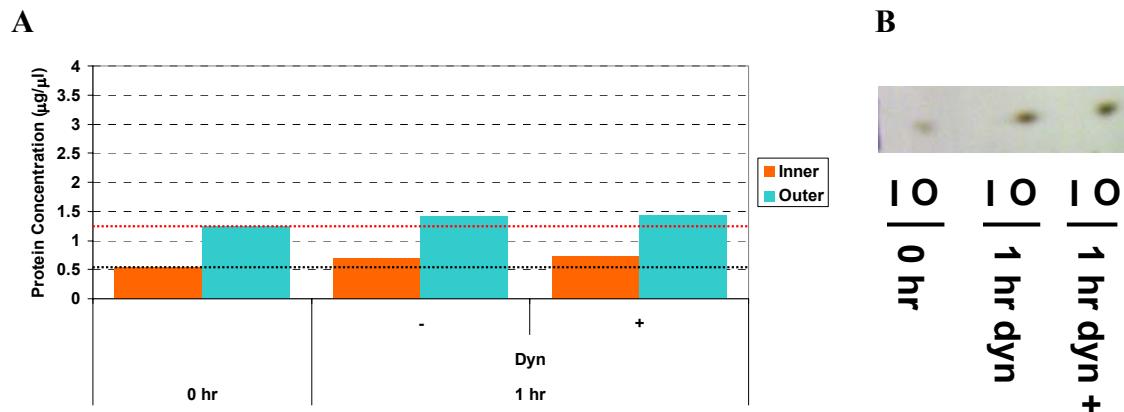
Intact two-day cultured chondrocyte-laden agarose disk seeded at  $30 \times 10^3$  cells/mL were tested for protein extraction efficacy and verified by SDS-PAGE separation of proteins. The concentration of protein extracted from one sample of three intact disks was 2.422 μg/μL (Fig. 9A). Gel staining after SDS-PAGE verified the presence of protein extracted from the disks (Fig. 9B), although the banding pattern of the proteins was unique from that of MSC pellets (Fig. 8B – SG). The modified cell lysis protocol was therefore determined to be successful for extraction from agarose disks.



**Figure 9** The concentration of protein extracted from three intact chondrocyte-laden agarose disks is 2.422 μg/μL (A). Gel staining after SDS-PAGE shows a distinct banding pattern for protein extracted from chondrocytes in agarose disks from that of MSCs in pellets (B).

#### 4.7 TOTAL P38 MAPK PROTEIN WAS DETECTED IN CHONDROCYTES EXTRACTED FROM THE OUTER REGION OF AGAROSE DISKS

For Western blotting, all inner region samples were diluted to match the sample of lowest concentration ( $0.533 \mu\text{g}/\mu\text{L}$ ), and likewise for outer region samples ( $1.245 \mu\text{g}/\mu\text{L}$ ) (Fig. 10A). However, the diluted inner region samples contained insufficient amounts of protein for p38 MAPK antibody detection, and only bands for the outer region samples appeared after developing (Fig. 10B). Thus, MAPK proteins can be detected in chondrocyte-laden agarose disks, although samples of higher protein concentrations are necessary ( $> 0.533 \mu\text{g}/\mu\text{L}$ ).



**Figure 10** The concentration of protein extracted from the inner ( $\sim 0.6 \mu\text{g}/\mu\text{L}$ ) and outer regions ( $\sim 1.4 \mu\text{g}/\mu\text{L}$ ) of chondrocyte-laden agarose disks are much less than that of intact disks (A). The black and red dotted lines designate the inner and outer region samples of lowest concentration, respectively. Total p38 MAPK protein levels appear only for the outer region samples and are elevated in samples with dynamic loading relative to time zero (B).

#### 5. DISCUSSION AND CONCLUSIONS

One of the aims of this project was to evaluate regional differences in ECM gene expression in response to TGF $\beta$ -3 induction and dynamic loading. Real-time PCR results from three replicate studies show that two-day cultured chondrocyte-laden agarose disks exhibit varying levels of aggrecan and type II collagen gene expression between the inner and outer regions. For the inner region, gene expression levels for both aggrecan and type II collagen are  $\sim 1$ , suggesting that short-term dynamic loading (one and four hours) at 10% strain and 1Hz is insufficient for initiating changes in gene expression. In contrast, gene expression levels for both aggrecan and type II collagen are elevated for the outer region. Since the two regions differ in the type of mechanical forces experienced, the regional variation in gene expression is not unexpected. However, the differences in gene expression levels present between the inner and outer regions do validate one of the first aims of these studies.

Although gene expression levels differ between the two regions, they do not vary greatly for duration of loading time. Specifically, aggrecan gene expression levels are not vastly different after one hour of loading versus four hours ( $\sim 1.5$ ; Fig. 5A). While type II collagen levels show more variance between one and four hours of loading, this difference is not appreciable (Fig. 5B). This sustained response between one and four hours of loading suggests that the gene

expression activity exhibits greatest change within the first hour of loading and is steady during the latter three hours. These results give us a time frame to focus upon in future studies.

The standard deviation between values from replicate studies J1–3 is noticeably large, which are likely due to low yields of RNA extracted from the two regions of samples. The inner region was especially difficult to extract large quantities of RNA from because of its small volume. Even with three disks collected into one sample for processing, resulting RNA concentrations were often low, which may account for much of the variation in starting gene transcript quantities.

In J4, a preliminary study of three-week cultured chondrocyte-laden agarose disks, results from one replicate show aggrecan gene expression levels are elevated in the outer region (~2.5) in comparison to results for two-day cultured disks (~1.7; Fig. 6). The elevation in aggrecan gene expression may result from the presence of ECM synthesized by cells with additional time in CDM- culture prior to loading (Fig. 3). In contrast, type II collagen gene expression levels were not observed to be substantially different between the inner and outer region. This study must be replicated to evaluate any robust trends in gene expression and differences between two days and three weeks of pre-culture in CDM- prior to loading.

Western blotting techniques were tested on MSC pellets initially, in order to develop a working protocol for eventual use with inner and outer region samples of chondrocyte-laden agarose disks. The concentration of protein extracted (~2.6-3.7  $\mu\text{g}/\mu\text{L}$ ) from the two MSC pellet studies, MSC1 and MSC2, and used for blotting (~2.6  $\mu\text{g}/\mu\text{L}$ ) were found to be adequate for detection of BiP/GRP78 chaperone protein and total ERK1/2 proteins (Figs. 7, 8). From MSC1, a linear, positively correlation was observed between pellet cell density and the amount of protein extracted from the pellet (Fig. 7A), a relationship which extended to the band intensity of stained gels and probed blots (Fig. 7B). From MSC2, the functionality of the MAPK antibodies was verified to advance towards use on protein extracted from chondrocyte-laden agarose disks (Fig. 8B).

Next, Western blotting techniques were tested on intact chondrocyte-laden agarose disks. Difficulty arose in the initial steps of cell lysis and protein extraction for the agarose disks. MSC pellets are primarily composed of cells, and sonicating the samples twice on ice was sufficient to lyse cells for the release of proteins. In contrast, chondrocytes encapsulated in agarose must be extracted from the gel before they can be lysed and proteins are released. The original cell lysis protocol was thus modified by using a motorized pestle to grind the agarose before sonication, similar to the method described previously for RNA extraction. Despite this additional step to extract protein, the concentration of protein extracted from the disks (2.422  $\mu\text{g}/\mu\text{L}$ ) was lower than that of MSC pellets, although the volume of lysis buffer was less for the disks (100  $\mu\text{L}$ ) than for the pellets (300  $\mu\text{L}$ ). Nonetheless, SDS-PAGE and gel staining showed the steps of protein extraction, denaturation, and separation by electrophoresis were successful, prompting the final transition to performing Western blots with inner and outer region samples from dynamic loading studies, J1–4 (Fig. 9). Notably, the proteins extracted from the chondrocyte-laden disks did exhibit a distinct banding pattern than that of the MSC pellets, but this observation is not unexpected given the differences between the two cell types (Fig. 9B).

Inner and outer samples harvested from J1 were used for testing of Western blotting. The concentration of protein for both inner ( $\sim 0.6 \mu\text{g}/\mu\text{L}$ ) and outer ( $\sim 1.4 \mu\text{g}/\mu\text{L}$ ) samples were both much less than that of the MSC pellets used previously for Western blot testing (Fig. 10A). The concentration of protein extracted from one sample containing three agarose disks ( $2.25 \times \varnothing 5.0 \text{ mm}$ ) seeded at  $30 \times 10^6$  cells/mL and lysed in 100  $\mu\text{L}$  was only 2.422  $\mu\text{g}/\mu\text{L}$  (Fig. 9A). Thus, the concentration of protein extracted from one sample containing 3-4 inner cores ( $2.25 \times \varnothing 2.0 \text{ mm}$ ) or outer annuli seeded at  $30 \times 10^6$  cells/mL and lysed in 100  $\mu\text{L}$  was not expected to be much greater.

For the first trial blot, all samples were diluted to match the sample of the lowest concentration, an inner region sample (0.533  $\mu\text{g}/\mu\text{L}$ ) (Fig. 10A). However, this concentration proved to be insufficient for antibody detection of phosphorylated ERK1/2 and phosphorylated p38 MAPK, resulting in blank sheets of film after developing (*not shown*). Thus, before performing more blots, sample concentrations were increased to improve protein detection. Since the concentration of outer region samples is generally at least two times greater than that of the inner region, all outer region samples were diluted to match the outer region sample of the lowest concentration (1.245  $\mu\text{g}/\mu\text{L}$ ), and similarly for the inner region samples (0.533  $\mu\text{g}/\mu\text{L}$ ) (Fig. 10A). Probing for total p38 MAPK would test if the amount of protein loaded onto the gel for the inner or region samples was sufficient for detection, since total MAPK levels are greater than phosphorylated MAPK levels in samples (Fig. 10B).

Overall, protein was successfully extracted from chondrocyte-laden agarose disks and probed for MAPKs, although difficulty was encountered in insufficient concentrations of protein loaded on the gel, similar to the trouble encountered during RNA extraction for real-time PCR.

## 6. RECOMMENDATIONS

For J1–3, these studies should be repeated with additional samples harvested for real-time PCR and Western blotting analysis. Real-time PCR results and trends between different loading durations would likely be more robust with higher yields of extracted RNA. Likewise, the blotting process would be much more efficient with additional samples to increase the concentration of protein loaded onto gels for Western blotting. With sufficient protein extracted from samples, the active and total MAPK profiles can be completed for the dynamically loaded chondrocyte-laden agarose disks in these studies.

The time frame established in J1–3 should also be explored in future studies in order to evaluate the changes that occur in aggrecan and type II collagen gene expression within the first hour of loading. Correspondingly, phosphorylated MAPK activity is also of interest relative to these changes in gene expression.

In J4, the differences in gene expression activity between two days and three weeks of pre-culture prior to loading were considered, but additional replicate studies must be conducted to assess the accuracy of the observed elevation in aggrecan gene expression levels with additional time in culture and ECM biosynthesis. Similarly, the protocol for these studies can also be modified to test other durations of pre-culture time and their effects on gene expression and MAPK activity.

For Western blotting, the initial studies performed with MSC pellets should be repeated with chondrocytes to distinguish differences in banding patterns among MSCs, chondrocytes, and agarose-encapsulated chondrocytes. In addition, this test will also check for the presence of any contamination from residual agarose.

Finally, these studies in elucidating chondrocyte mechanotransduction can be translated to MSCs in the future, since these cells are an important alternative source in cartilage tissue engineering. With the knowledge of the biochemical mechanisms underlying chondrocyte and MSC response to mechanical signals, functional cartilage tissue engineering can be vastly improved for application to osteoarthritis and other diseases involving cartilage damage and degradation.

## 7. ACKNOWLEDGMENTS

The author would like to acknowledge Dr. Robert L. Mauck and Alice H. Huang of the McKay Orthopaedic Research Laboratory for their support and mentorship throughout the summer, as well as Dr. Paul Billings of the Center for Fibroplasia Ossificans Progressiva for his instruction in and help with Western blotting. The research discussed in this paper was accomplished through funding from the National Science Foundation and the Summer Undergraduate Fellowship in Sensor Technologies at the University of Pennsylvania.

## 8. REFERENCES

1. V.C. Mow, X.E. Guo, Mechano-Electrochemical Properties of Articular Cartilage: Their Inhomogeneities and Anisotropies, *Annu. Rev. Biomed. Eng.*, 4 (2002) 175-209.
2. F. Guilak, C.T. Hung, Physical Regulation of Cartilage Metabolism, In V.C. Mow, R. Huiskes, *Basic Orthopaedic Biomechanics and Mechano-Biology*, Lippincott Williams & Wilkins, Philadelphia, PA, 3rd ed., 2004, pp. 259-300.
3. E.B. Hunziker, Articular cartilage repair: basic science and clinical progress. A review of the current status and prospects, *Osteoarthr. Cartil.*, 10 (2001) 432-463.
4. Arthritis Foundation, Disease Center: Osteoarthritis - Overview, Retrieved August 2, 2006, from [http://www.arthritis.org/conditions/diseasecenter/oa/oa\\_overview.asp](http://www.arthritis.org/conditions/diseasecenter/oa/oa_overview.asp).
5. C.K. Kuo, W-J. Li, R.L. Mauck, R.S. Tuan, Cartilage tissue engineering: its potential and uses, *Curr. Open. Rheumatol.*, 18 (2006) 64-73.
6. R.L. Mauck, X. Yuan, R.S. Tuan, Chondrogenic differentiation and functional maturation of bovine mesenchymal stem cells in long-term agarose culture, *Osteoarthr. Cartil.*, 14 (2006) 179-189.
7. A.J. Grodzinsky, M.E. Levenston, M. Jin, E.H. Frank, Cartilage tissue remodeling in response to mechanical forces, *Annu. Rev. Biomed. Eng.*, 2 (2000) 691-713.
8. J.H.C. Wang, B.P. Thampatty, An introductory review of cell mechanobiology, *Biomech. Model. Mechanobiol.*, 5 (2006) 1-16.
9. F. Guilak, V.C. Mow, The mechanical environment of the chondrocyte: a biphasic finite element model of cell-matrix interactions in articular cartilage, *J. Biomech.*, 33 (2000) 1663-1673.
10. R.L. Mauck, B.A. Byers, X. Yuan, R.S. Tuan, Regulation of Cartilaginous ECM Gene Transcription by Chondrocytes and MSCs in 3D Culture in Response to Dynamic Loading, *Biomech. Model. Mechanobiol.*, (2006) in press.

11. P.P. Roux, J. Blenis, ERK and p38 MAPK-Activated Protein Kinases: a Family of Protein Kinases with Diverse Biological Functions, *Microbiol. Mol. Biol. Rev.*, 68 (2004) 320-344.
12. R. Seger, E.G. Krebs, The MAPK signaling cascade, *FASEB J.*, 9 (1995) 726-735.
13. P.J. Fanning, G. Emkey, R.J. Smith, A.J. Grodzinsky, N. Szasz, S.B. Trippel, Mechanical regulation of mitogen-activated protein kinase signaling in articular cartilage, *J. Biol. Chem.*, 278 (2003) 50940-50948.
14. K.W. Li, A.S. Wang, R.L. Sah, Microenvironment Regulation of Extracellular Signal-Regulated Kinase Activity in Chondrocytes, *Arthritis Rheum.*, 48 (2003) 689-699.
15. R. Tuli, S. Tuli, S. Nandi, X. Huang, P.A. Manner, W.J. Hozack, K.G. Danielson, D.J. Hall, R.S. Tuan, Transforming Growth Factor- $\beta$ -mediated Chondrogenesis of Human Mesenchymal Progenitor Cells Involves N-cadherin and Mitogen-activated Protein Kinase and Wnt Signaling Cross-talk, *J. Biol. Chem.*, 278 (2003) 41227-41236.