Annual Research Journal

Electrical Engineering Research Experience for Undergraduates

Vol. VII

August 2009

National Science Foundation
Grant No. EEC-0775081

Principal Investigators:
W. K. Jenkins and Sven G. Bilén

PENNSTATE

Department of Electrical Engineering
University Park, Pennsylvania

EEREU Annual Research Journal

Sven Bilén (editor)
John Mitchell (co-editor)

Volume **VII**
Published in 2009 by
The Department of Electrical Engineering
The Pennsylvania State University
University Park, Pennsylvania 16802

NSF EE REU Site Program Contact

316 Electrical Engineering East
The Pennsylvania State University
University Park, PA 16802
Telephone (814) 865-0184
Fax (814) 865-7065
E-Mail: eereu@engr.psu.edu
Web Site: http://www.ee.psu.edu/reu/

ISBN 0-913260-09-6
978-0-913260-09-8

PREFACE

We are happy to present the **2009 *Annual Research Journal – Electrical Engineering Research Experience for Undergraduates, Vol. VII.*** This volume contains highlights of the EEREU program activities for summer 2009 and thirteen technical papers written by EEREU scholars as primary authors.

Summer 2009 marks our seventh year hosting the National Science Foundation–sponsored Research Experience for Undergraduates (REU) Site Program, by the Department of Electrical Engineering, Penn State. Thirteen outstanding young men and women participated in this year's EEREU program at Penn State's University Park Campus. These EEREU scholars, selected from nation-wide applicants, consisted of college first-year, sophomore, or junior students with outstanding academic backgrounds and intense interests in exploring research in electrical engineering and related areas.

During the nine-week summer program, EEREU students carried out research projects under the guidance of his or her faculty mentor(s), in laboratories hosted by the Department of Electrical Engineering and the Materials Research Institute at Penn State. The students also presented their research experience and findings at the 2009 Annual *EEREU Symposium*, held at University Park, in July 2009.

Besides research activities, the EEREU program organized an array of group activities including a ***Weekly Scientific Seminar Series*** that introduced a broad range of research topics to the REU students, a ***field trip program*** that offered EEREU students opportunities to visit prominent local and regional industrial and research sites, and a ***Weekly Workshop on Ethics and Entrepreneurship*** through which students were engaged in debate and analysis of issues in ethics and company start-ups in engineering. For more information about Penn State's EEREU program, please visit our website at: < http://www.ee.psu.edu/reu/ >.

We are confident that readers will find that the series of EEREU journals showcases the achievement of our EEREU students and effective mentorship provided by the faculty and graduate student mentoring teams. We are also hopeful that this publication will not only document original research contributions that are of value for scientific dissemination and publication, but it may also stimulate more college students to consider research careers and to pursue graduate studies in electrical engineering.

W. Kenneth Jenkins and Sven G. Bilén
Co-Directors of the NSF EEREU Site Program
Dept. of Electrical Engineering
The Pennsylvania State University

July 2009
University Park, PA

TABLE OF CONTENTS

Research Articles (* indicates REU student author)

2009 NSF EEREU FACULTY & STAFF MEMBERS

Faculty Mentors:

Prof. Sven Bilén, Co-Director
Prof. Heath Hofmann
Prof. Ken Jenkins, Co-Director
Prof. I. C. Khoo
Prof. Mike Lanagan
Prof. Zhiwen Liu
Prof. John Mathews
Prof. Robert Nickel
Prof. Victor Pasko
Prof. Joan Redwing
Prof. Jeff Schiano
Prof. Srinivas Tadigadapa
Prof. Julio Urbina

Ms. Amy Freeman, Adviser, Engineering Diversity
Prof. Ruyan Guo, Program Advisor (Former Director)
Mrs. MaryAnn Henderson, Administration
Prof. Elizabeth Kisenwether, Entrepreneurship Chair
Prof. Andy Lau, Ethics Chair
Prof. Jack Mitchell, Seminar Chair
Mr. Thomas Tyson, Assistant Director
Prof. David Salvia, Activities and Web Chair

NSF EEREU 2009 Summer Program
Research Seminar Series

Room 101 E E East, 9:40 – 10:30 am, Tuesdays
(*The seminars are open to the public*.)

Department of Electrical Engineering
Pennsylvania State University, University Park, PA 16802

Date	***Topic***	***Speaker***
June 2, 2009	Energy Harvesting	Heath Hofmann
June 9, 2009	The Joy of Quadrupole Resonance	Jeff Schiano
June 16, 2009	Lightning-related Transient Luminous Events in the Middle Atmosphere	Victor Pasko
June 23, 2009	What's Next? Graduate School, of Course!	David Salvia
June 30, 2009	Integrating Nanowire-based Sensors with CMOS Electronics Using Directed Self Assembly	Theresa Mayer
July 7, 2009	Software-defined Radar for Meteor Science	Julio Urbina
July 14, 2009	Sensors Using Micro and Nanoscale Structures	Srinivas Tadigadapa
July 23, 2009 (Thursday)	2009 NSF EEREU Symposium	REU Students

NSF EEREU 2009 Summer Program
Ethics & Entrepreneurship Workshop

Room 101 E E East, 10:50 – 11:50 am, Tuesdays

Department of Electrical Engineering
Pennsylvania State University, University Park, PA 16802

Date	*Topic*	*Speaker*
June 2, 2009	Frameworks of Ethics	Andy Lau
June 9, 2009	Research Ethics	Andy Lau
June 16, 2009	Sustainability and Materials Selection in Product Development	Andy Lau
June 23, 2009	Student Entrepreneur Panel: "What's it take to be an entrepreneur?"	Erik Weir and Erik Davidson
June 30, 2009	Intellectual Property Issues	Liz Kisenwether
July 7, 2009	Ethics Issues in Entrepreneurship	Liz Kisenwether and Andy Lau
July 14, 2009	Student Presentations on Tech Transfer Ideas of Research Projects	REU students

NSF EEREU 2009 Summer Program
Field Trips

All field trips are from 1:30–4:00 pm on Tuesdays except where noted.

Department of Electrical Engineering
Pennsylvania State University, University Park, PA 16802

Date	*Field Trip*	*Contact*
May 28, 2009 Thursday	Shaver's Creek	
June 2, 2009	EE Lab Tour	Tom Tyson
June 9, 2009	Nanofab and MCL	Leeanna McCool
June 16, 2009	Accuweather	Tyler Ishler
June 23, 2009	Videon Central	Todd Erdley
July 2, 2009 Thursday All-day trip	**Pittsburgh Day Trip:** AlphaLab TTC, Inc. MAYA Design	Liz Kisenwether
July 7, 2009	Breazeale Nuclear Reactor	Candace Davison
July 14, 2009	Applied Research Lab	Janet Jonson

2009 NSF EE REU Field Trip Program
Academic/Industrial Sponsors and Tour Hosts

Department of Electrical Engineering
Penn State University, University Park, Pennsylvania

Materials Research Institute
Penn State University, University Park, Pennsylvania

Applied Research Laboratory
Penn State University, University Park, Pennsylvania

Breazeale Nuclear Reactor
Penn State University, University Park, Pennsylvania

Nanofabrication National Facility
Penn State University, University Park, Pennsylvania

Videon Central, Inc.
State College, Pennsylvania

AccuWeather, Inc.
State College, Pennsylvania

AlphaLabs
Pittsburgh, Pennsylvania

The Technology Collaborative
Pittsburgh, Pennsylvania

MAYA Design
Pittsburgh, Pennsylvania

2009 NSF EE REU SYMPOSIUM

Final Program

8:30 am to 3:00 pm, Thursday, July 23, 2009
Room 101 Electrical Engineering East Building
Pennsylvania State University, University Park, PA 16802

Time	Sessions and Topics		Chairs and Speakers
8:30 – 8:55 am	**Symposium Registration (Refreshments Provided)**		**(MaryAnn Henderson/ Tom Tyson)**
8:55 – 9:00 am	**Welcome**		**Ken Jenkins**
9:00 – 10:00 am	**Session I**	Session Chairs:	**Pasko/Urbina/ Mathews**
9:00– 9:15	MINIMUM BREAKDOWN VOLTAGE FOR CORONA DISCHARGE IN CYLINDRICAL AND SPHERICAL GEOMETRIES		Aaron Gibson
9:15 – 9:30	EFFECTS OF HYPERBOLIC ELECTRODE CONFIGURATIONS ON STREAMER PROPAGATION IN AIR		Michael Blackmond
9:30 – 9:45	HIGH TEMPORAL AND SPATIAL RESOLUTION STUDIES OF METEOR HEAD ECHOES AND THE GYRO LINE OBSERVED OVER THE ARECIBO OBSERVATORY		Jim Reuss
9:45 – 10:00	CLASSIFICATION AND RADIO SCIENCE MODELING OF RADAR METEOR ECHOES		Jennifer Cross
10:00 – 10:15 am	**Coffee Break**		
10:15 – 11:30 am	**Session II**	Session Chairs:	**Bilén /Jenkins/ Hofmann/Schiano**
10:15 – 10:30	DEPLOYABLE STRUCTURES BASED ON SHAPE MEMORY ALLOYS		Jennifer Felder
10:30 – 10:45	PEAK POWER CONTROLLER FOR SOLAR ARRAYS ON SPACECRAFT		Arnaldo Negron
10:45 – 11:00	SPEECH PROCESSING USING MATCHED FILTERS BASED ON TRANSFORM DOMAIN PROCESSING		Zehra Raza
11:00 – 11:15	COMPARISON OF NEMA 23 BLDC AND STEPPER MOTORS USED AS GENERATORS		Nicholas Czarnek
11:15 – 11:30	PREDICTION AND MEASUREMENT OF CONVERSION GAIN IN A ROBINSON MARGINAL OSCILLATOR		Devon Miller

2009 NSF EE REU SYMPOSIUM (cont.)

Final Program

8:30 am to 3:00 pm, Thursday, July 23, 2009
Room 101 Electrical Engineering East Building
Pennsylvania State University, University Park, PA 16802

Time	Sessions and Topics	Chairs and Speakers
11:30 – 11:45 am	**Group Photo – All are invited. Please plan to be available.**	(Tom Tyson)
12:00 – 1:30 pm	**Luncheon (Nittany Lion Inn)**	
1:45 – 2:45 pm	**Session III** Session Chairs:	**Tadigadapa/Liu/ Khoo/Lanagan/ Redwing**
1:45 – 2:00	TRANSFER OF SILICON WIRES ONTO GLASS SUBSTRATES FOR USE IN RADIAL SOLAR CELLS	Rachel Christian
2:00 – 2:15	ENERGY STORAGE CHARACTERIZATION OF NONLINEAR DIELECTRICS NEAR BREAKDOWN	Colin Stapleton
2:15 – 2:30	STUDY OF FABRICATION AND ENERGETIC PROPERTIES OF POROUS SILICON	Javier Espinosa Acevedo
2:30 – 2:45	EXPERIMENTAL CHARACTERIZATION OF NONLINEAR OPTICAL MATERIALS	Katie Knox
2:45 – 3:00 pm	**CONGRATULATIONS AND CONCLUDING REMARKS**	**Jenkins/ Bilén**

3:00 pm Adjournment

Picnic at Sunset Park, 5:00 pm – sunset, ALL ARE INVITED

Resources:	Tom Tyson and MaryAnn Henderson
Special Notes:	Laser pointer, desktop PC and LCD PC projector are provided. Speakers please upload presentation files to EEREU group web in Angel by July 22, Wed. It is strongly recommended that speakers test-run presentation files for audio, video, or special applications on a Windows PC prior to their Thursday presentation. Each presentation is 15 minutes – chair introduction, questions, and answers included.

MINIMUM BREAKDOWN VOLTAGES FOR CORONA DISCHARGE IN CYLINDRICAL AND SPHERICAL GEOMETRIES

Aaron S. Gibson*, Jeremy A. Riousset†, and Victor P. Pasko‡

CSSL Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate student of
Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85719

ABSTRACT

The lightning rod was invented in the mid-1700's by Benjamin Franklin, who suggested that the rods should have sharp tips to increase the concentration of electric fields, and also to prevent the rod from rusting.[1, 2] More recently, Moore *et al.* suggested that the strike-reception probabilities of Benjamin Franklin's rods are greatly increased when their tips are made moderately blunt.[2] In this work, we use Townsend's equation for corona discharge, to find a critical radius and minimum breakdown voltage for cylindrical and spherical geometries. We solve numerically the system of equations and present simple analytical formulas for the aforementioned geometies. These formulas complement the classic theory developed in the framework of Townsend theory.[3, 4]

INTRODUCTION

Interest in lightning protection has been renewed in the past decades due to potential hazards to a variety of modern systems, such as buildings, electric power and communications systems, electronic integrated circuit chips, aircraft, and boats.[5] Current lightning protection devices can reduce damage in two different ways: (1)

†Graduate Mentor
‡Faculty Mentor

by hampering the formation of lightning, (2) by intercepting the lightning through a lightning rod.[6, 7] Understanding the complex nature of lightning is made challenging by the difficulty of reproducing atmospheric processes in laboratories. These difficulties are accentuated by the many scales of the lightning phenomenon, which involves microphysical processes developing in a multi-kilometer scale discharge.[7] However, it is now understood that lightning is an atmospheric discharge of the same nature as the spark.[3]

ELECTRON AVALANCHES, CORONA, STREAMERS, AND LEADERS

The building block of any electrical discharge is the electron avalanche.[3] The avalanche begins with a small number of "seed" electrons that appear accidentally, for example due to cosmic rays.[3] Under the influence of an electric field, these electrons gain kinetic energy which is lost in collisions with neutral particles,[8] ionizing these particles and creating new electrons. The so-created electrons experience the same acceleration as the seed electrons, until they collide with more atoms creating new ions and electrons[3, 9]. The process repeats itself forming the "electron avalanche."

A glow corona can be observed in the dark or in bad weather on high-power transmission lines or as St. Elmo's fires on the mast of a ship or airplane.[5] Glow coronae are also responsible in part for the noise of a transistor radio.[5, 9] Physically, a glow corona developing near a grounded object under severe weather conditions is a region of moderately ionized plasma that conducts a very small ionic current, on the order of a few μA, though approaching lightning can cause the current to increase to a few mA.[10] The electric field around the tip of a lightning rod will vary with geometry, but it can hypothetically range from 0.2 kV/cm to 2.7 kV/cm depending on the proximity (or existence thereof) of an approaching lightning.[10]

Glow coronae are formed in strongly nonuniform electric fields[3] by ionization of the neutral particles of air. In order to maintain the corona discharge, the electrons released during the ionization process must replace the electrons lost to attachment to ambient molecules.[3] The rate of ionization is therefore fundamental and is expressed as the ionization coefficient, α, which measures the number of ionization events performed by an electron in a 1 cm path along the electric field.[3] Each electron has an exponential effect towards the avalanche, which can be described by the following equation:[3, 4, 11]

$$\int_{R_1}^{R_2} \alpha \, dr = \ln Q \tag{1}$$

where Q is the number of electrons in the avalanche, R_1 and R_2 are the positions of the first and second electrode respectively, $d = R_2 - R_1$ is the distance between the two electrodes and r is the position between the electrodes. The value, $Q = 10^4$, is used in this paper, since this value is in good agreement with published experimental values for point and wire configurations, and also for positive and negative corona.[11]

In the case of atmospheric discharges, glow corona can greatly influence the initiation of more ionized plasma discharges such as streamers and leaders, discussed hereafter.[10] Streamers are moderately, one can even say weakly ionized plasma channels that are necessary for the extension of leader channels, constituting the lightning branches.[3] In the case of atmospheric discharges, streamers are produced and destroyed at the rate of $\sim 10^9$ Hz[12] in the so-called streamer zone of the atmospheric lightning channel. Streamers can be self-sustained[13] and propagate due to the relatively high electric field at their tip also referred to as the streamer head.[14]

Streamers can lead to perhaps the most notorious atmospheric discharge: lightning. The lightning channel is commonly called a leader.[9] Leaders are highly conductive plasma filaments,[3] with very high ionization currents, on the order of several kA.[5] Bazelyan and Raizer[9] stated that the foremost condition of leader formation in air is an increase in gas temperature, at least to the extent necessary to suppress a decrease in conductivity owing to electron attachment.[9] The temperature of a leader is $\gtrsim$5000 K, which is significantly higher than the 300 K typical of streamer corona.[14, 15] This increase in temperature is due to the merging of the streamer currents starting from the leader tip.[16] An increase in conductivity due to an increase in temperature will result in the narrowing of current flow to a thin channel.[9] The currents of all of the streamers starting from a leader tip are summed up, leading to Joule heating of the region ahead of the tip and therefore to increase in its thermal energy.[14] The increase in thermal energy will increase the conductivity of the region ahead of the tip, thus extending the channel further. The process of leader propagation is not fully understood[17] at present and falls beyond the scope of this present paper. A brief summary comparing the types of discharges discussed previously is given in Table I.

Table I: Atmospheric discharge characteristics at ground level (adapted from [14]).

Parameter	Glow Corona	Streamer	Leader
Temperature	$\sim$330 K	$\sim$330 K[14]	$\gtrsim$5000 K[9, 15]
Electron energy	1–2 eV[4]	5–15 eV[15]	5 keV[18]
Electric field	0.2–2.7 kV/cm[10]	5–7.5 kV/cm[15]	1–5 kV/cm[9]
Electron density	2.6×10^8 cm^{-3}[3]	5×10^{13}–10^{15}cm^{-3}[9, 15]	4×10^{14}cm^{-3}[13]

TOWNSEND BREAKDOWN

Townsend's model deals with dark discharge and glow corona (the primary difference between the two being the luminous "glow"). Townsend discharge occurs in the presence of two parallel plate electrodes with a voltage difference of a few kV in standard conditions in between them. Based on experimental data, Townsend suggested a formula for the ionization coefficient, α:[3, 19]

$$\alpha = Ape^{-\frac{Bp}{E}} \tag{2}$$

where A (1/cm/Torr) and B (V/cm/Torr) are empirically calculated constants that vary with the composition of the gas, p is pressure of the gas in Torr, and E is the electric field in V/cm.[3, 4] Using equations (1) and (2) and Poisson's equation, a critical (minimum) breakdown voltage can be derived and plotted as graphs known as the Paschen curves.[3] Of particular interest in this paper is Stoletov's point, which is the point of maximum current and more importantly minimum voltage for the initiation of corona discharge in a plane-plane configuration.[4]

It is believed that corona discharge may attract approaching downward leaders[21, 22, 26]. Based on repeated observation, it is generally accepted that a lightning leader progresses to about 10 m above the ground until the electric field at a point on the ground surface has increased sufficiently to cause an upward leader to be initiated.[20] The upward-leader evolves from a corona discharge into a high current leader, which can establish a connecting path with the downward lightning leader.[21] Since this observation holds if the object struck is fitted with a lightning rod,[22] an ideal lightning rod should readily produce an upward connecting leader.

In the application of lightning rods, experimental results suggest that lightning rods with blunt tips are more effective than the rods with sharp tips.[1, 2, 23, 24] The effectiveness of the blunt-tipped rod is suggested to be related to the corona onset on the geometry of the rod.[11] To date, a full quantitative description of upward leader development is yet to be developed. A simple first-order approximation of the corona discharge is needed, because the effect it has on the leader initiation and therefore on a lightning protection device's efficiency is not negligible.[25] The conversion of corona into an upward leader is a critical part of the lightning attachment process.[26] Lowke and D'Alessandro[11, 24] formulated theoretical models of glow corona around the lightning rod tip using variations of Peek's formula ($E_c = 30\delta\left(1 + \frac{0.3}{\sqrt{\delta R_1}}\right)$, where E_c is in kV/cm and δ is the relative air density[11]). However, Peek's formula does not allow for the evaluation of the minimum breakdown voltage as a function of the rod's radius. Gary *et al.*[19] suggested the use of $\alpha = k\frac{p}{p_0}\left(\left(\frac{E}{E_0}\right)^2\left(\frac{p_0}{p}\right)^2 - 1\right)$ instead of Townsend's equation (2), where p_0 is a reference pressure ($\approx 10^5$ Pa) and E_0 is the reference electric field ($\approx$ 24-31 kV/cm) at p_0. Gary *et al.*[19] state that Townsend's law is satisfactory for higher values of $\frac{E}{p}$, but dismisses the use of Townsend's equation for lower values.[19] Lowke and D'Alessandro[11] theorized that the Townsend mechanism would suggest that results are dependent on electrode materials. However, the experimental results are remarkably independent of the composition of the electrodes.[11] On the other hand, the use of Townsend's breakdown criterion allows for the estimation of the minimum breakdown voltage.

In this work, the constants A and B are fitted to an exact formulation for α[8]:

$$\alpha = \frac{\nu_i(E) - \nu_{a_2}(E)}{\mu_e(E)E}$$

where ν_i, ν_{a_2}, and μ_e are the ionization frequency, two body attachment frequency, and electron mobility respectively. ν_i, ν_{a_2}, and μ_i are obtained using models formulated by Morrow and Lowke[27]. The coefficients A and B are calculated using an exponential fit to the curve $\frac{E}{N}$ vs. $\frac{\alpha}{N}$, where N is the atmospheric neutral density. In this work, the electric field and minimum breakdown voltage for the initiation of the corona surrounding a lightning rod are calculated numerically and approximated analytically for the cases of cylindrical and spherical geometries.

MODEL FORMULATION

In equation (2), A and B are dependent on the composition of the gas. Raizer[3] uses values of $A = 15$ 1/cm/Torr and $B = 365$ V/cm/Torr; however, better approximations can be obtained based on the numerical results of Morrow and Lowke[§].[27] Using an exponential fit with values of α obtained using these results[27] for ν_i, ν_{a_2}, and μ_e, the optimal values are: $A = 7.0$ 1/cm/Torr and $B = 258$ V/cm/Torr. Figure 1 compares the results.

After the evaluation of A and B, general expression for the electric field is obtained from Poisson's equation in the absence of space charge: $\nabla^2 V = 0$ with the following boundary conditions:

$$E(c) = \delta E_0 \qquad V(R_2) = 0 \tag{3}$$

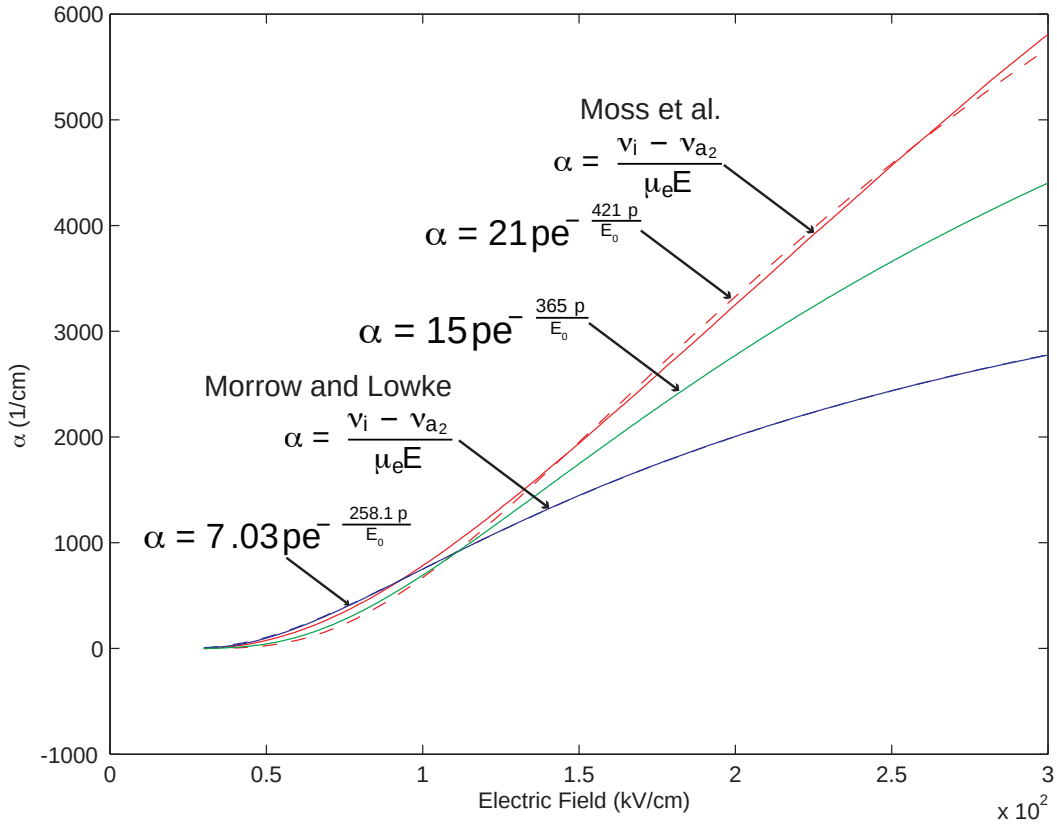
where V is the electric potential ($E = -\vec{\nabla} V$), R_2 is an arbitrarily large distance from the electrode where the voltage is 0 and c is the position of the corona front. The ratio $\delta = \frac{p}{p_0}$ is a scaling factor for different pressures, and E_0 is the classic breakdown electric field. Beyond the corona front ($r > c$), E falls below the classic breakdown electric field, causing ionization to stop, and yielding the first boundary condition. We then substitute E in (1) and solve for c. Finally an expression for the critical voltage is obtained as a function of d for corona discharge in Cartesian geometry and R_1 and R_2 as $V_c = V(R_1)$, for corona discharge around cylindrical or spherical objects.

CARTESIAN COORDINATES

Townsend discharge in the Cartesian case involves two parallel plates, some distance d apart with a gas at a pressure, p. Townsend's equation (2) approximates the experimental data well, though the model fails at higher values of pd (pressure·distance) in part because of the appearance of different ionization phenomena such as the formation of streamers, which require a different model.[9, 19]

The Cartesian solution of Townsend's breakdown is well-known and studied by various authors.[3, 4] Due to symmetry, the electric field is uniform between the two

[§]The air1.m function can also be used, though slightly different values of A and B will be acquired. The air1.m function is explained in detail in [28] and can be found at pasko.ee.psu.edu/air/air1.m. For the purposes of this paper, only formulations of Morrow and Lowke[27] employed.



..

Figure 1: The values of α for different models

plates which simplifies the integral of equation (1). Substituting equation (2) into equation (1), gives:

$$\int_{R_1}^{R_2} \left(Ape^{-\frac{Bp}{E}}\right) dr = \left(Ape^{-\frac{Bp}{E}}\right) \int_{R_1}^{R_2} dr = Ap \underbrace{(R_2 - R_1)}_{d} e^{-\frac{Bp}{E}} = \ln Q \qquad (4)$$

Solving equation (4) for E and noting that $V = Ed$ due to the uniform electric field gives:

$$E = -\frac{Bp}{\ln\left(\frac{\ln Q}{Apd}\right)} \quad ; \qquad V = -\frac{Bpd}{\ln\left(\frac{\ln Q}{Apd}\right)} \qquad (5)$$

This well-known result can be found in [3, 4] under the form: $V = \frac{Bpd}{\ln(pd)+\ln\left(\frac{A}{\ln Q}\right)}$.

Stoletov's minimum values are obtained as[3]:

$$d_{\min} = \frac{e}{Ap} \ln Q \quad ; \qquad E_{\min} = Bp \quad ; \qquad V_{\min} = \frac{eB}{A} \ln Q \qquad (6)$$

CYLINDRICAL COORDINATES

In this section, we calculate the critical breakdown voltage in a cylindrical geometry as a function of the radii R_1 and R_2 of the inner and outer electrodes. The derived equations will be in terms of R_1 and R_2 as opposed to d since it is well established that ionization is a function of the breakdown electric field (E_0) and the curvature radius (R_1) of the electrode.[29] The cylindrical case starts off with Poisson's equation in cylindrical coordinates (r, ϕ, z). Assuming symmetry with respect to ϕ and z, we have $\frac{\partial V}{\partial \phi} = 0 = \frac{\partial V}{\partial z}$. Poisson's equation for an axisymmetric cylindrical geometry with no axial variation gives:[4] $\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \mathrm{V}}{\partial r}\right) = 0$. The solution of this differential equation is:

$$V = k_1 \ln r + k_2 \tag{7}$$

where k_1 and k_2 are integration constants. Using the boundary conditions specified by (3), we have $E(c) = \delta E_0$, for which ionization equals attachment.[11] Additionally, the voltage will be zero at a distance R_2 that is sufficiently larger than R_1. In this work, we used $R_2 = 1$ m. Using these boundary conditions, the equations for voltage and electric field can be found as:

$$E = \frac{\delta E_0 c}{r} \;\; ; \qquad V = \delta E_0 c \ln \frac{R_2}{r} \tag{8}$$

Substituting (8) into Equation (1), we obtain:

$$\int_{R_1}^{c} Ape^{-\frac{Bp}{E}} = \int_{R_1}^{c} Ape^{-\frac{Bpr}{\delta E_0 c}} = \frac{Apc\left(e^{-\frac{Bp}{\delta E_0}} - e^{-\frac{BpR_1}{\delta E_0 c}}\right)}{-\frac{Bp}{\delta E_0}} = \ln Q \tag{9}$$

which we solve for c. An analytical approximation of the solution can be obtained through a Taylor expansion of the exponential term, $e^{-\frac{BpR_1}{\delta E_0 c}} \approx 1 - \frac{BpR_1}{\delta E_0 c}$. This is justified by noting that in reality, $c \gg R_1$.[11] Substituting into Equation (9) and simplifying yields a closed form solution for c, $c = B\left(\ln Q + ApR_1\right) / \left(A\delta E_0 \left(1 - e^{\frac{-Bp}{\delta E_0}}\right)\right)$. Having substituted c into Equation (8) leads to:

$$V(r) = \frac{B \ln\left(\frac{R_2}{r}\right)(\ln Q + ApR_1)}{A(1 - e^{\frac{-Bp}{\delta E_0}})} \quad E(r) = \frac{B(\ln Q + ApR_1)}{rA(1 - e^{\frac{-Bp}{\delta E_0}})} \tag{10}$$

and consequently,

$$V_{\mathrm{c}} = V(R_1) = \frac{B \ln \frac{R_2}{R_1}(\ln Q + ApR_1)}{A(1 - e^{-\frac{Bp}{\delta E_0}})} \; ; \; E_{\mathrm{c}} = E(R_1) = \frac{B(\ln Q + ApR_1)}{AR_1(1 - e^{-\frac{Bp}{\delta E_0}})} \tag{11}$$

The minimum of V_c with respect to R_1 occurs at the root, R_{min} of $\frac{\partial V_c}{\partial R_1} = 0$. Solving for R_1 gives:

$$R_1 = R_{min} = \frac{-\ln Q}{Ap \cdot \text{ProductLog}\left(\frac{e \ln Q}{ApR_2}\right)} \tag{12}$$

where ProductLog is the inverse function of $f(x)$=xe^x. In order to obtain a simpler form of (12), a first order Taylor expansion is used, $\text{ProductLog}(-x) \approx -j\pi + \ln(-x)$[§]. This leads to:

$$R_{min} = -\frac{\ln Q}{Ap\left(j\pi + \ln\left(\frac{-e \ln Q}{ApR_2}\right)\right)}$$

Since $\ln(-e) = j\pi + 1$, the imaginary part of the solution cancels leaving a purely real answer of:

$$R_{min} = \frac{-\ln Q}{Ap\left(1 + \ln\frac{\ln Q}{ApR_2}\right)} \tag{13}$$

The minimum critical electric field and voltage can be found by substituting Equation (13) into (11):

$$\begin{cases} \lambda &= \frac{\ln Q}{ApR_2} \\ \kappa &= \frac{Bp_0}{E_0} \\ V_{min} &= \frac{B}{A}\frac{\ln Q}{1 + 1/\ln\lambda}\frac{\ln\left(-\frac{1}{\lambda}\left(1 + \ln\lambda\right)\right)}{1 - \exp(-\kappa)} \\ E_{min} &= \frac{Bp\ln\lambda}{\exp(\kappa) - 1} \end{cases} \tag{14}$$

SPHERICAL COORDINATES

Deriving the minimum breakdown voltage for corona discharge in spherical coordinates uses the same assumptions as those employed in the cylindrical case. We start with Poisson's equation in spherical coordinates (r, θ, ϕ). The electric field is now independent of θ and ϕ, as opposed to ϕ and z. In spherical coordinates, the general solution for the electric potential becomes:

$$V = -\frac{k_1}{r} + k_2$$

[§]In certain intervals, ProductLog has more than one solution. For sake of continuity, the principal root is generally used, but in this particular application, the correct root was not the principal root (the principal root will give a solution that breaks the assumption that $R_2 \gg R_1$). The Taylor expansion used converges to the desired root as opposed to the principal root. ProductLog is also known as the Lambert-W or Omega function. For more information on the ProductLog function, visit http://documents.wolfram.com/mathematica/functions/ProductLog.

Using the boundary conditions (3), we obtain formulas for potential and electric field.

$$V(r) = c^2 \delta E_0 \left(\frac{1}{r} - \frac{1}{R_2} \right) \;\; ; \qquad E(r) = \frac{c^2 \delta E_0}{r^2} \tag{15}$$

The value of c is obtained from substituting (15) into Equation (1), which yields the following integral:

$$\begin{aligned} \int_{R_1}^{c} Ape^{\frac{-Bpr^2}{\delta E_0 c^2}} \, dr &= \ln Q \\ \frac{Apc\sqrt{\pi}\sqrt{\delta E_0}}{2\sqrt{Bp}} \left(\text{Erf}\left(\sqrt{\frac{Bp}{\delta E_0}} \right) - \text{Erf}\left(\frac{R_1}{c} \sqrt{\frac{Bp}{\delta E_0}} \right) \right) &= \ln Q \end{aligned} \tag{16}$$

where $\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, dt$. As previously, we use a Taylor expansion: $\text{Erf}(x) \approx \frac{2x}{\sqrt{\pi}}$. This approximation is justified by noting that $c \gg R_1$ (and therefore $\frac{R_1}{c} \approx 0$). For typical values ($B \simeq 258$ V/cm/Torr, $E_0 \simeq 31$ kV/cm, $\delta = 1$, and $p = 760$ Torr), $\text{Erf}\left(\sqrt{\frac{Bp}{\delta E_0}} \right) \approx 1$. This leads to:

$$\text{Erf}\left(\sqrt{\frac{Bp}{\delta E_0}} \right) \approx 1 \;\; ; \qquad \text{Erf}\left(\frac{R_1}{c} \sqrt{\frac{Bp}{\delta E_0}} \right) \approx \frac{2}{\sqrt{\pi}} \frac{R_1}{c} \sqrt{\frac{Bp}{\delta E_0}} \tag{17}$$

Substituting Equation (17) into Equation (16) gives, after simplification $c = \frac{2\sqrt{Bp}}{Ap\sqrt{\pi \delta E_0}}$, which with (15) gives:

$$V(r) = \frac{4B \left(\ln Q + ApR_1 \right)^2}{\pi A^2 p} \left(\frac{1}{r} - \frac{1}{R_2} \right) E(r) = \frac{4B \left(\ln Q + ApR_1 \right)^2}{\pi A^2 p r^2} \tag{18}$$

Although imperfect, the above approximation provides a practical analytical solution for the voltage needed to initiate a corona discharge around spherical objects. Since the critical voltage and electric field will occur at the inner electrode, the critical voltage is:

$$\begin{aligned} V_{\text{c}} &= V(R_1) = \frac{4B \left(\ln Q + ApR_1 \right)^2}{\pi A^2 p} \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \\ E_{\text{c}} &= E(R_1) = \frac{4B \left(\ln Q + ApR_1 \right)^2}{\pi A^2 p R_1^2} \end{aligned} \tag{19}$$

The minimum voltage for corona breakdown is obtained by solving $\frac{\partial V_{\text{c}}}{\partial R_1} = 0$. Solving for R_1 results in a cubic equation with three roots:

$$R_1 = -\frac{\ln Q}{Ap} \;\; ; \qquad R_1 = \frac{R_2}{4} \left(1 \pm \sqrt{1 - \frac{8 \ln Q}{ApR_2}} \right) \tag{20}$$

The first root is extraneous (R_1 cannot be negative), and the second root is forbidden by the assumption that $R_2 \gg R_1$. This leaves the third root which is:

$$R_1 = R_{\min} = \frac{R_2}{4}\left(1 - \sqrt{1 - \frac{8\ln Q}{ApR_2}}\right) \tag{21}$$

In reality, we have $R_2 \to \infty$. This limit diverges in cylindrical geometry due to the logarithm, but in spherical coordinates, we get:

$$R_{\min} = \lim_{R_2 \to \infty} \frac{R_2}{4}\left(1 - \sqrt{1 - \frac{8\ln Q}{ApR_2}}\right) = \frac{\ln Q}{Ap} \tag{22}$$

Assuming R_2 is infinite and substituting (22) into (19), the critical electric field and voltage in spherical coordinates are:

$$E_{\min} = \frac{16Bp}{\pi} \;\; ; \qquad V_{\min} = \frac{16B\ln Q}{\pi A} \tag{23}$$

RESULTS

Using the revised values of A and B fitted using Morrow and Lowke's[27] values for ν_i, ν_{a_2}, and μ_e, the obtained analytical solutions are in good agreement with the numerical solution for all three geometries, most notably around the minimum. The values at the minimum for both cylindrical and spherical geometries are shown in Table II.

Table II: Critical Values in Cylindrical and Spherical geometry.

Parameter	Numerical	Analytical	% Error
	Cylindrical		
Critical Radius (cm)	1.4×10^{-4}	1.5×10^{-4}	8%
Critical Electric field (kV/cm)	3.3×10^{4}	3.6×10^{4}	9%
Critical Voltage (kV)	4.9	4.9	0.6%
	Spherical		
Critical Radius (cm)	2.3×10^{-3}	1.7×10^{-3}	26%
Critical Electric field (kV/cm)	2.6×10^{3}	1.0×10^{3}	61%
Critical Voltage (kV)	6.0	1.7	71%

Figure 2(a) plots the breakdown voltage $V_{\rm c}$ (V) for each geometry as a function of the product, pd or pR_1 (cm-Torr) in the same way as the original Paschen curve. Solid lines indicate the analytical solutions and dashed lines indicate the numerical solution using Morrow and Lowke's[27] models. Arrows point to both the numerical

and analytical solutions in Cartesian, cylindrical, and spherical geometries. Also, (car.) indicates the Cartesian solution, (cyl.) indicates the cylindrical solutions, and (sph.) indicates the spherical solutions. An '$\times$' denotes the location of the minimum on each curve.

Figure 2(b) plots the breakdown voltage V_{c} (V) for each geometry as a function of δd or δR_1 (m). As in Figure 2(a), solid lines indicate analytical solutions, dashed lines indicate the numerical models based on Morrow and Lowke's[27] models and (car.), (cyl.), and (sph.) correspond to Cartesian, cylindrical, and spherical geometries, respectively. The minima are also marked with '$\times$'. The axes were scaled to show the location of the minimum in cylindrical geometries, which occurs towards the left side of the graph.

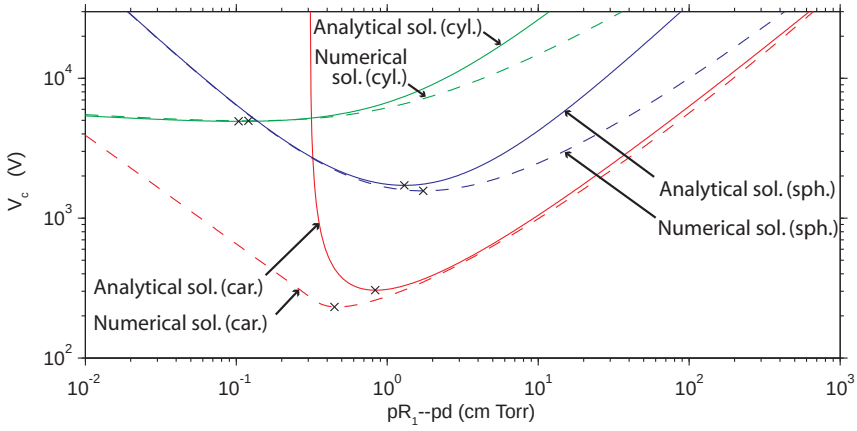
Figure 2(c) plots the breakdown electric field E_{c} (V/m) as a function of δd or δR_1 (m) for each geometry with the same legend as the two prior graphs. Here, '$\times$' denotes the location of $R_{\min}$ and the critical electric field for this value.

Finally, Figure 3 compares breakdown voltage (V) as a function of δR_1 (m) for the models of cylindrical and spherical onset corona proposed by Lowke and D'Alessandro[11] to the analytical and numerical voltages obtained in the previous section, in cylindrical and spherical geometry, respectively. The arrows point to the numerical solution, the analytical solution, and the curve corresponding to the models of Lowke and D'Alessandro.[11] As previously, an '$\times$' is placed at the minimum where applicable.

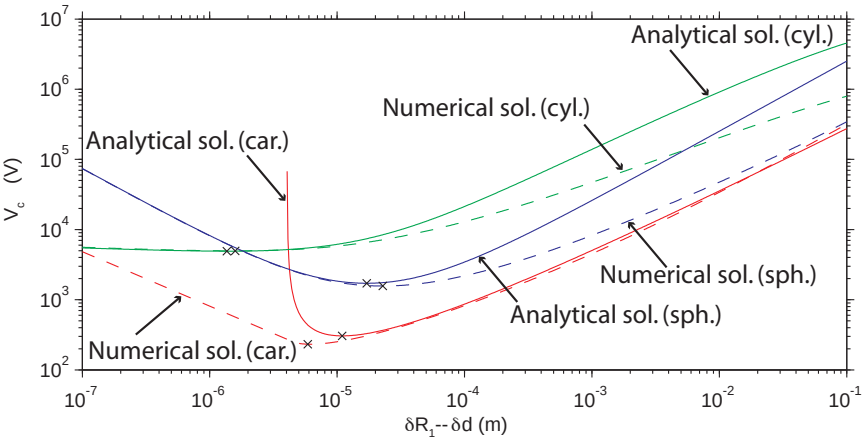
DISCUSSION AND COMPARISON

In this section, we discuss the significance and accuracy of the numerical and analytical solutions to the problem of breakdown voltage for initiation of Townsend discharge in a parallel plate configuration and of corona discharges in cylindrical and spherical geometries. The curves of the Cartesian solution in Figure 2(a) represent the classic solution to the Cartesian problem, and show the well-known form of the Paschen curves [e.g., 3]. For large values of δR_1, the numerical solutions of all geometries converge toward the cartesian solution, since both cylindrical and spherical geometries will locally behave like a plane-plane configuration when the two electrodes are closer together.

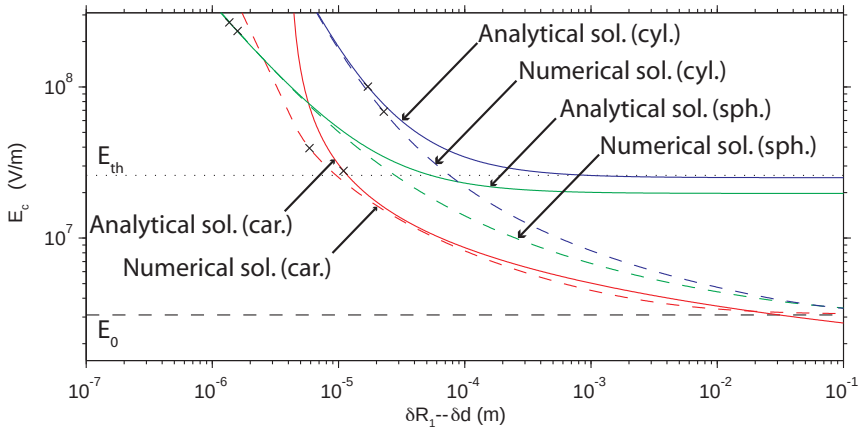
Paschen curves show that the critical electric field is dependent on the product pd, i.e., the corona breakdown voltage follows a similarity law.[3, 29] In the case of the spherical analytical solution, the voltage is a function of the product: pR_1. These results, namely that the breakdown voltage follows a similarity law, suggest that initiation of a corona around a spherical object can be minimized for a given radius by modifying the pressure. In the atmosphere, this could be accomplished by lifting the object to higher altitudes, i.e. to lower pressures. The critical radius, electric field, and breakdown voltage in spherical geometry has a nearly identical form to the values at Stoletov's point in Cartesian geometry, the difference being



(a) Comparison of the breakdown voltages with respect to pd or pR_1. The dashed lines show the numerical solutions, while solid lines show the analytical solutions. The (car.), (cyl.), and (sph.) additions correspond to the Cartesian, cylindrical, and spherical solutions respectively.

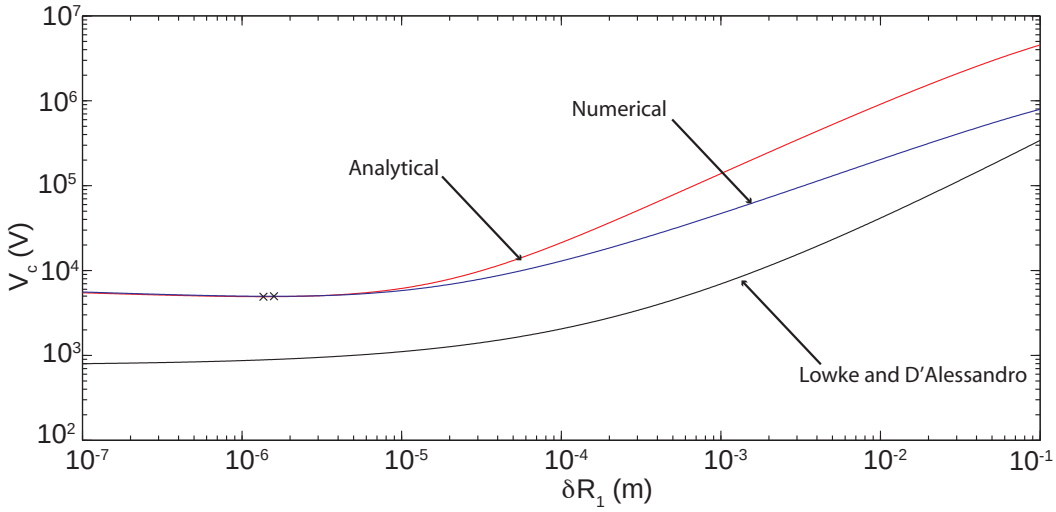


(b) Comparison of the breakdown voltages with respect to δd or δR_1. The legend is the same as in part (a).

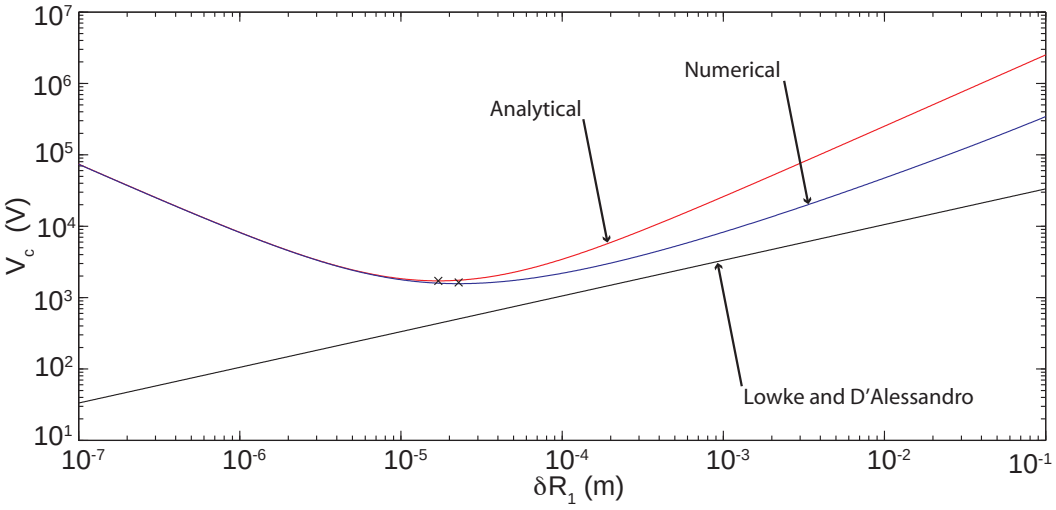


(c) Comparison of the critical electric fields. The legend is the same as in part(a) and (b).

Figure 2: Comparison of numerical and analytical solutions.



(a) Comparison of breakdown voltage curves in the cylindrical case. The top two curves correspond to the analytical and numerical solutions respectively. The bottom curve represents Lowke and D'Alessandro's [11] solution.



(b) Same as plot (a) only for a spherical geometry.

Figure 3: Breakdown curves in various geometries.

a factor of $\frac{16}{\pi}$ as opposed to e, i.e, a factor of $\sim$2 (1.87). It appears that cylindrical geometries do not follow the same similarity law due to the presence of a logarithm in the analytical solution.

While Gary *et al.*[19] states that Townsend's equation is "satisfactory" for higher values of E, the cylindrical and spherical case diverge from the numerical solutions at these values. A possible explanation to this observation lies in the assumption that $\frac{R_1}{c} \approx 0$ in the Taylor expansion of the ProductLog and Erf functions. In reality, it is possible that $c \not\gg R_1$ which may limit the accuracy of the analytical formulas.

Lowke and D'Alessandro[11] use a quadratic model in the cylindrical case and a linear model in the spherical case due to the difficulty in solving these expressions analytically. Figure 3 shows that Lowke and D'Alessandro's[11] models are roughly an order of magnitude different from either the numerical or analytical solution in both geometries. The fit could be improved if numerical parameters in Lowke and D'Alessandro's formulas were adjusted in a similar way to that used in this work for the coefficients of A and B.

In spherical geometry, Lowke and D'Alessandro's[11] model for voltage differs by several orders of magnitude for low values of δR_1, but agreement is reasonable near the minimum $\delta R_1 \approx 2 \times 10^{-5} = 20$ μm. This complements the analytical solution which is more accurate for lower values δR_1, but the linear approximation is overall less accurate than the Townsend equation in spherical geometry. Additionally, Lowke and D'Alessandro's model does not show a minimum point which is suggested by the results of our work.

It should be noted that the minima obtained in this study using Townsend's equation in Cartesian, cylindrical, and spherical geometries appeared to be associated with critical electric fields that exceed the thermal runaway electric fields ($E_{\mathrm{th}} \approx 260$ kV/cm[28]). This suggests that classic estimates of Stoletov's point in Paschen theory as well as the new results presented in this study, should be regarded cautiously as they may be located in a region where ionization and two-body attachment could not be the dominant processes. In addition, theoretical minimum radii are on the order of μm, corresponding to very small or very sharp pointed lightning rods, which does not explain at present the results of the field studies by Moore *et al.*.[1, 2, 23] Consequently, further research is required to determine the lightning rod optimum radius and to give an accurate theoretical explanation for Moore *et al.*'s results.

CONCLUSIONS

Corona onset is fundamental for creation of an upward leader[22] and hence for the development of more effective lightning rods.[11] Golde[21, 22] hypothesized that the lightning rod's efficiency is dependent on how readily upward leaders form from the lightning rod's tip; however, Golde's hypothesis has not yet been confirmed theoretically.[7] Moore *et al.*'s experiments suggest that there is an optimum radius, but none of the published models based on Peek's equation allow for the

evaluation of a minimum.[11, 19] The models presented in this paper introduce analytical formulas as well as numerical solutions, that allow for a first estimation of the critical radius and minimum breakdown voltage for corona discharge around lightning rods in cylindrical and spherical geometries. However, these models predict electric fields that exceed thermal runaway of the electric field for corona onset (≈ 260 kV/cm).[28] To resolve this issue, proposed models may be further improved by fitting the constants A and B not only to different proposed numerical models, but by restricting the fits to more realistic electric fields for corona discharge. Additionally, other factors such as space charge effects need to be considered as well in theoretical models of upward leader development.[26]

ACKNOWLEDGEMENT
This material is based upon work supported by the National Science Foundation under Grant, No. EEC-0755081 and ATM-0652148.

REFERENCES

[1] C. B. Moore, G. D. Aulich and W. Rison, "Measurements of lighting rod responses to nearby strikes," *Geophys. Res. Lett.*, **27** (10) 1487–1490 (2000b).

[2] C. B. Moore, W. Rison, J. Mathis and G. Aulich, "Lightning rod improvement studies," *J. Appl. Meteor.*, **39** (5) 593–609 (2000a).

[3] Y. P. Raizer, *Gas Discharge Physics*, Springer-Verlag, New York, NY, 1991.

[4] R. J. Roth, *Industrial plasma engineering, Vol. 1: Principles*, IOP Publishing Ltd, 1995.

[5] V. A. Rakov and M. A. Uman, *Lightning: Physics and Effects*, Cambridge Univ. Press, Cambridge, U.K.; New York, 2003.

[6] N. L. Aleksandrov, E. M. Bazelyan, F. D'Alessandro and Y. P. Raizer, "Dependence of lightning rod efficacy on its geometric dimensions–a computer simulation," *J. Phys. D: Appl. Phys.*, **38** (1) 1225–1238, doi:10.1088/0022-377/38/8/021 (2005).

[7] N. L. Aleksandrov, E. M. Bazelyan and Y. P. Raizer, "The effect of a corona discharge on a lightning attachment," *Plasma Phys. Rep.*, **31** (1) 75–91 (2005).

[8] N. Y. Liu and V. P. Pasko, "Effects of Photoionization on Similarity Properties of Streamers at Various Pressures in Air," *J. Phys. D: Appl. Phys.*, **39** 327–334, doi:10.1088/0022-3727/39/2/013 (2006).

[9] E. M. Bazelyan and Y. P. Raizer, *Spark Discharge*, Chemical Rubber Company Press, New York, NY, 1998.

[10] N. L. Aleksandrov, E. M. Bazelyan, F. D'Alessandro and Y. P. Raizer, "Numerical simulations of thunderstorm-induced corona processes near lightning rods installed on grounded structures," *J. Electrostat.*, **64** (12) 802–816, doi: 10.1016/j.elstat.2006.02.001 (2006).

[11] J. J. Lowke and F. D'Alessandro, "Onset corona fields and electrical breakdown criteria," *J. Phys. D: Appl. Phys.*, **36** (21) 2673–2682, doi:10.1088/0022-3727/36/21/013 (2003).

[12] N. L. Aleksandrov, A. E. Bazelyan, E. M. Bazelyan and I. V. Kochetov, "Modeling of Long Streamers in Atmospheric-Pressure Air," *Plasma Phys. Rep.*, **21** (60) 57–75 (1995).

[13] N. Aleksandrov, E. Bazelyan and A. Konchakov, "Plasma parameters in the channel of a long leader in air," *Plasma Physics Reports*, **27** (10) 875–885 (2001).

[14] J. A. Riousset, *Fractal modeling of lightning discharges*, Master's thesis, The Pennsylvania State University, University Park, PA (2006).

[15] I. Gallimberti, G. Bacchiega, A. Bondiou-Clergerie and P. Lalande, "Fundamental processes in long air gap discharges," *C. R. Physique*, **3** (10) 1335–1359, doi:10.1016/S1631-0705(02)01414-7 (2002).

[16] N. A. Popov, "Formation and development of a leader channel in air," *Plasma Phys. Rep.*, **29** (8) 695–708, doi:10.1134/1.1601648 (2003).

[17] E. M. Bazelyan and Y. P. Raizer, *Lightning Physics and Lightning Protection*, IoP Publishing Ltd, Bristol, UK and Philadelphia, PA, 2000.

[18] A. V. Gurevich, K. P. Zybin and Y. V. Medvedev, "Runaway breakdown in strong electric field as a source of terrestrial gamma flashes and gamma bursts in lightning leader steps," *Phys. Lett. A.*, **361** 119–125, doi: 10.1016/j.physleta.2006.05.063 (2007).

[19] C. Gary, J. Schmitt and B. Hutzler, "Peeks Law Generalization Application To Various Field Configurations," *IEEE Transactions On Power Apparatus and Systems*, **PA91** (6) 2262–& (1972).

[20] R. H. Golde, "Lightning Conductor," *Journal of the Franklin Institute-Engineering and Applied Mathematics*, **283** (6) 451–& (1967).

[21] R. H. Golde, ed., *Lightning*, vol. 1: Physics of Lightning, Academic Press, London, UK; New York, NY, 1977.

[22] R. H. Golde, ed., *Lightning*, vol. 2: Lightning Protection, Academic Press, London, UK; New York, NY, 1977.

[23] C. B. Moore, G. D. Aulich and W. Rison, "The case for using blunt-tipped lightning rods as strike receptors," *J. Appl. Meteor.*, **42** (7) 984–993 (2003).

[24] F. D'Alessandro, "On the optimum rod geometry for practical lightning protection systems," *J. Electrostat.*, **65** (2) 113–121, doi:10.1016/j.elstat.2006.07.011 (2007).

[25] F. D'Alessandro and J. Gumley, "A "Collection Volume Method" for the placement of air terminals for the protection of structures against lightning," *Journal of Electrostatics*, **50** (4) 279–302 (2001).

[26] F. D'Alessandro and G. Berger, "Laboratory studies of corona emissions from air terminals," *Journal of Physics D-Appied Physics*, **32** (21) 2785–2790 (1999).

[27] R. Morrow and J. J. Lowke, "Streamer propagation in air," *J. Phys. D: Appl. Phys.*, **30** 614–627 (1997).

[28] G. D. Moss, V. P. Pasko, N. Liu and G. Veronis, "Monte Carlo model for analysis of thermal runaway electrons in streamer tips in transient luminous events and streamer zones of lightning leaders," *J. Geophys. Res.*, **111** A02307, doi: 10.1029/2005JA011350 (2006).

[29] N. Spyrou, R. Peyrous, N. Soulem and B. Held, "Why Paschens Law Does Not Apply in Low-Pressure Gas-Discharges With Inhomogeneous Fields," *Journal of Physics D-Appied Physics*, **28** (4) 701–710 (1995).

EFFECTS OF HYPERBOLIC ELECTRODE CONFIGURATIONS ON STREAMER PROPAGATION IN AIR

Michael R. Blackmond*, Sebastien Celestin†, and Victor P. Pasko†

CSSL, Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate student of
Department of Electrical and Computer Engineering
Western Michigan University
Kalamazoo, MI 49007

ABSTRACT

This paper investigates how the shape of a hyperbolic electrode affects the propagation of a streamer at atmospheric pressure in air in a point-to-point configuration. An analysis of numerical methods is performed and their subsequent development for streamer simulations is presented. Some technology transfer opportunities are discussed. An iterative numerical solver is used to compare numerical and analytical solutions of Poisson's equation for a set of representative electrode geometries. A numerical method called the ghost fluid method (GFM), which has been previously developed by Celestin *et al.* [2009], is used to compute the electric potential in order to simulate streamer propagation. It is found that approximations used in models are valid for correct evaluation of time required for connection of two streamers propagating from opposite electrodes in point-to-point configurations, which is critical for the interpretation of the three regimes appearing in the point-to-point nanosecond repetitively pulsed (NRP) discharges.

†Faculty Mentor

INTRODUCTION

The concept of streamer discharges was first introduced in the 1930's by Raether and Loeb to explain spark discharges. Cravath and Loeb also used this concept to explain very fast phenomenon that J. J. Thompson observed at low pressure long tubes in 1893 [Loeb, 1965]. There are many applications of streamer discharges including ozone production, pollution control, plasma-assisted combustion, biological decontamination and air flow control. Streamers often take the form of thin plasma filaments in air at atmospheric pressure and are largely driven by their own *space charge field*. The way a streamer propagates is governed by the high electric field produced by the *streamer head*. There have been many experimental and simulation works done in the past years to better understand this phenomenon. Some earlier studies have used simple electrode configurations such as the plane-to-plane geometry [Kulikovsky, 1997; Dhali and Williams, 1987]. Although these studies are good for certain situations, more complex electrode geometries are much more common in applications. A widely used example of complex geometry is the point-to-plane configuration, where one of the electrodes is a plane and the other is a needle or sphere.

Luque *et al.* [2008] simulated this configuration and left the point electrode out of the computational domain. With only one point in contact with the computational domain, the regions near the point electrode are approximated. Kulikovsky [1997] included the point-shaped electrode in the simulation domain. In this reference, streamer simulations are performed using a rectilinear grid and the finite-volume method. Since the electrode is defined in the computational domain, the shape of the electrode has to be approximated by a 'staircase'. In work presented by Djermoune *et al.* [1995] and Morrow and Lowke [1997], the shape of the electrode is included as accurately as possible by defining a new axial grid point for each radial grid point such that the position defined by the axial and radial coordinates lies on the electrode surface. This approach introduces some difficulties for a study of different electrode configurations because a new grid has to be generated for every new electrode configuration and changing the electrode shape would affect the numerical resolution of transport of charged species.

In present paper, a method is used to take into account the electrode configurations in rectilinear grids. This method, called the ghost fluid method (GFM) Fedkiw *et al.* [1999], has been adapted for the simulation of streamers by Celestin *et al.* [2009]. In Celestin *et al.* [2009], the authors approximated the point-to-point configurations experimentally studied in Pai *et al.* [2008] by hyperbolas with a defined radius of curvature of 324 μm. In Pai *et al.* [2008], the shape of the electrodes (and hence the radius of curvature) is not accurately known. Thus, the goal of the present paper is to study how the shape of electrodes affects the discharge in order to find to what extent an approximation of electrodes is valid.

In the following section, some technology transfer opportunities are discussed. In the Model Formulation Section, the equations for the numerical simulation of streamer discharges are given. Subsequently, Poisson's equation is analytically and numerically solved in hyperbolic coordinates. In the Results Section, the results of the streamer simulations are presented for a few radius of curvature of the hyperbolic electrodes. Finally, the Conclusions Section presents conclusions on the dependance of electrode geometries on streamer propagation.

TECHNOLOGY TRANSFER OPPORTUNITIES

There are numerous applications of streamers. Some of those applications include ozone production, pollution control, biological decontamination, plasma assisted combustion, air flow control, and thin film coatings. Streamers can produce a spark powerful enough to break the bonds between oxygen molecules. The oxygen atoms can then collide with oxygen molecules to produce ozone. As mentioned in Kogelschatz *et al.* [1999], there is an increasing market for ozone production. The main applications are in water treatment and pulp bleaching. Another application of streamers is in the production of AC plasma displays. Also discussed in Kogelschatz *et al.* [1999], a dielectric can be introduced between the electrodes to create a dielectric-barrier discharge (DBD) in order to prevent the transition to arc. This discharge often produces ultra-violet radiation. UV radiation can easily be converted to visible light with the aid of phosphors. By using the 3 primary colors of light (red, blue and green), a screen with many tiny DBD cells can be made to produce a very vibrant display.

Plasma assisted combustion has been getting an increasing amount of attention in recent years. The electrode configurations studied in this paper are of particular interest for this application. In Pilla *et al.* [2006], this application is discussed using NRP applied voltages with pulses of 10 ns. The use of streamers is beneficial for reducing the power used in sustaining the flames that initiate the combustion. While the list is large for current applications of streamers, the future promises even more.

MODEL FORMULATION

When studying the dynamics of a streamers, the most common and effective model is based on the following drift-diffusion equations for electrons and ions coupled with Poisson's equation:

$$\frac{\partial n_e}{\partial t} + \vec{\nabla} \cdot (n_e \vec{v_e}) - \vec{\nabla} \cdot (D_e \vec{\nabla} n_e) = S_e^+ - S_e^-, \tag{1}$$

$$\frac{\partial n_p}{\partial t} = S_p^+ - S_p^-, \tag{2}$$

$$\frac{\partial n_n}{\partial t} = S_n^+ - S_n^-, \tag{3}$$

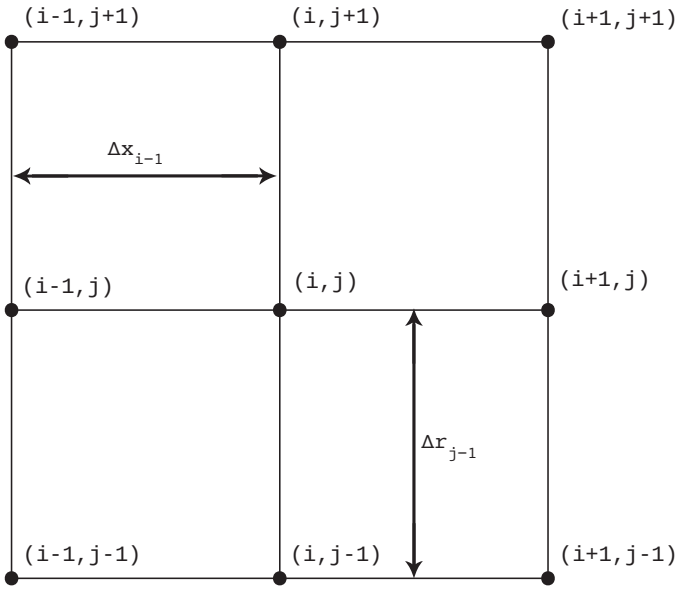


Figure 1: Cartesian grid used in finite-differences method for solution of Poisson's equation.

$$\nabla^2 V = -\frac{q_e}{\epsilon_0}(n_p - n_n - n_e) = -\frac{\rho}{\epsilon_0}, \tag{4}$$

where subscripts 'e', 'p' and 'n' refer to electrons, positive and negative ions, respectively, n_i is the number density of species of *i*, q_e is the absolute value of the elementary charge, ρ is the charge density, *V* is the electric potential, $\vec{v_e} = -\mu_e \vec{E}$ ($\vec{E}$ being the electric field) is the drift velocity of electrons, D_e and μ_e are the diffusion coefficient and the absolute value of the mobility of electrons, respectively, and ϵ_0 is the permittivity of free space. On timescales of interest for the studies presented in this paper, ions are assumed to be motionless. The S^- and S^+ terms stand for the rates of production and loss of charged particles. In this work, the source terms and transport coefficients have been taken from [Morrow and Lowke, 1997].

When simulating streamer propagations, it is critical to have an accurate calculation of the electric field in the domain of the simulation, especially in places where there is a high field gradient. In the following section, we use the finite differences approach as well as the GFM applied to Poisson's equation to get a valid description of the potential and electric fields in streamer simulations.

Poisson's Equation

Writing out Equation (4) in cylindrical coordinates, we get:

$$\frac{\partial}{\partial x}\left(\frac{\partial V}{\partial x}\right) + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) = -\frac{\rho(x,r)}{\epsilon_0}, \tag{5}$$

where *x* and *r* are axial and radial coordinates, respectively. We can use the finite-difference method to discretize Poisson's equation. In reference to Figure 1, we use a rectilinear grid to describe the potential. For an arbitrary point on the grid, the second-order finite-difference discretization of Equation (5) gives the classical five diagonal linear system:

$$V^e_{i,j}V_{i+1,j} + V^w_{i,j}V_{i-1,j} + V^s_{i,j}V_{i,j-1} + V^n_{i,j}V_{i,j+1} + V^c_{i,j}V_{i,j} = -\frac{\rho_{i,j}}{\epsilon_0} \tag{6}$$

where

$$V^e_{i,j} = \frac{1}{\Delta x_i(x_{i+1/2} - x_{i-1/2})}, \tag{7}$$

$$V^w_{i,j} = \frac{1}{\Delta x_{i-1}(x_{i+1/2} - x_{i-1/2})}, \tag{8}$$

$$V^n_{i,j} = \frac{2r_{j+1/2}}{\Delta r_j(r^2_{j+1/2} - r^2_{j-1/2})}, \tag{9}$$

$$V^s_{i,j} = \frac{2r_{j+1/2}}{\Delta r_{j-1}(r^2_{j+1/2} - r^2_{j-1/2})}, \tag{10}$$

$$V^c_{i,j} = -(V^e_{i,j} + V^w_{i,j} + V^n_{i,j} + V^s_{i,j}), \tag{11}$$

with $\Delta x_i = x_{i+1} - x_i$ and $\Delta r_j = r_{j+1} - r_j$. As demonstrated in Celestin *et al.* [2009], this discretization can be modified by the GFM to set the potentials on the interface between the electrodes of complex shape and the computational domain. To solve Poisson's equation, one must set boundary conditions. There are two type of boundary conditions that can be implemented, the homogeneous Neumann boundary condition and the Dirichlet boundary conditions. The Dirichlet boundary condition is in reference to the solution to Laplace's equation where $\rho = 0$ is set. In order to minimize the influence on the streamer propagation, it is necessary to have a much larger computational domain when using the Neumann boundary conditions. Therefore, to solve Laplace's equation, the Dirichlet boundary conditions will be used.

Analytical Solution

To define Dirichlet boundary conditions, the analytical solution to Laplace's equation can be used. For the point-to-plane and point-to-point configurations, it is assumed that the electrodes are infinite hyperboloids and for point-to-point configurations, the two hyperboloids are the same, that is, they have the same radius of curvature and the same distance to the focus, as seen in Figure 2. By considering the following cartesian equation for a hyperbola:

$$\frac{x^2}{a^2} - \frac{r^2}{b^2} = 1, \tag{12}$$

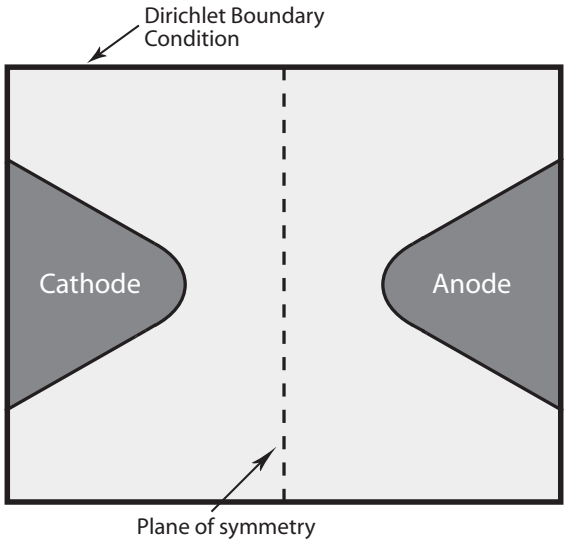


Figure 2: Point-to-point configuration.

the shape of the electrodes can be easily changed by changing *a* and *b*. By changing *a*, the distance from the plane of symmetry is changed. The radius of curvature of the hyperbola is changed mostly by *b*. When modeling these electrodes, it is advantageous to use the hyperbolic transform equations [Eyring *et al.*, 1928]:

$$x = \alpha v\sqrt{1+u^2} \tag{13}$$

and

$$r = \alpha u\sqrt{1-v^2}. \tag{14}$$

Before proceeding further, it is necessary to discuss some of the relations between the constants from Equation (12): *a* and *b*, and from Equations (13) and (14): α. By combining Equations (12)–(14) and with some simple algebra, we arrive at $a^2 = \alpha^2 v^2$, $b^2 = \alpha^2(1-v^2)$, and the squared focal length of the hyperbola, $a^2 + b^2 = \alpha^2$. We can define the electrode surfaces by $v_0^2 = a^2/\alpha^2$. Eyring *et al.* [1928] gives an expression for the potential:

$$V(v) = \frac{A}{2}\log\frac{1+v}{1-v} + B, \tag{15}$$

where *A* and *B* are constants defined by the anode and cathode voltages. They are written as:

$$A = \frac{V_a - V_c}{\log\dfrac{1+v_0}{1-v_0}} \tag{16}$$

and

$$B = \frac{V_a + V_c}{2}. \tag{17}$$

It is important to note that in Equation (15) the voltage is only dependent on the hyperbolic coordinate, *v*. With that in mind, it is useful to have an expression for *v* in terms of x and r. Equations (13) and (14) can be manipulated to get a biquadratic polynomial, which can then be solved leading to:

$$v^2(x,r) = \frac{\alpha^2 + x^2 + r^2 - \sqrt{(\alpha^2 + x^2 + r^2)^2 - 4\alpha^2 x^2}}{2\alpha^2}. \tag{18}$$

To get the electric field, the gradient of Equation (15) in hyperbolic coordinates is evaluated and gives [Eyring *et al.*, 1928]:

$$|\vec{E}(x,r)| = \frac{1}{U(x,r)} \frac{A}{1 - v^2(x,r)}, \tag{19}$$

where

$$U(u,v) = \alpha \sqrt{\frac{u^2 - v^2 + 1}{1 - v^2}}. \tag{20}$$

With all of these equations, we now have the analytical solution to Poisson's equation in hyperbolic point-to-point configuration. These equations are used to define the Dirichlet boundary conditions and to initiate the Poisson solver. We can now proceed to the numerical solution.

Numerical Methods

To numerically solve the transport equations for charges species, an 'upwind' numerical scheme is used [e.g., Pancheshnyi *et al.*, 2001]. It is the simplest scheme for the finite-volume method. We did not carry out any special numerical arrangements near the electrodes when setting up the boundary conditions for the drift-diffusion equations. The boundary conditions are computed in both directions by assuming the interface between the electrode and computational domain is a 'staircase' for fluxes of particles. For simplicity, the gradients of electron density near the electrodes are assumed to be zero. In the work presented in this paper, the D03EBF module of the NAG Fortran library (http://www.nag.co.uk) is used. This is based on the iterative Stone algorithm [Stone, 1968] which produces a very good accuracy and is not time consuming. Celestin *et al.* [2009] performed a comparison of the NAG module and the direct SuperLU and MUMPS solvers. It was found that this solver is very accurate and powerful.

RESULTS

Laplacian Field and Potential

The main focus of the simulations in this work revolve around the point-to-point configuration with different radius of curvature of hyperbolic electrodes. It is first important to check agreement between the analytical and numerical solutions of Poisson's equation at initial moment of time ($t = 0$). This is done for a

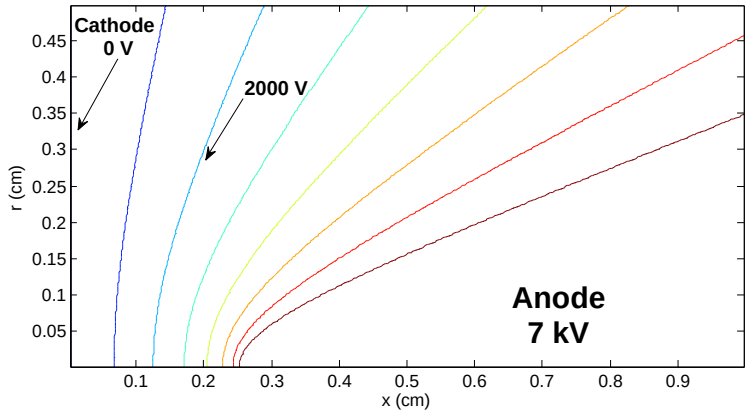


Figure 3: Isocontours of the electric potential in a point-to-plane configuration.

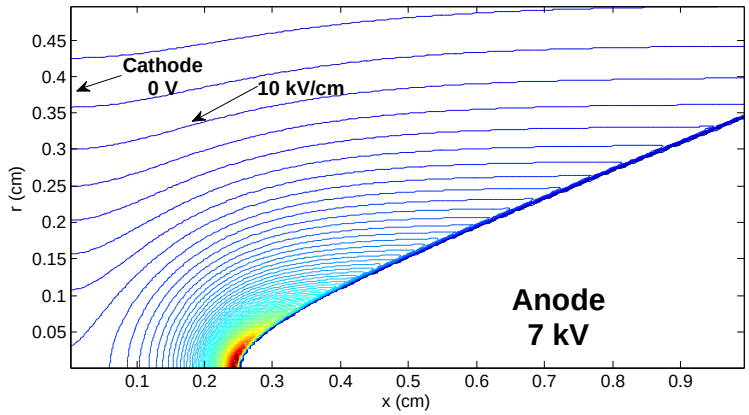


Figure 4: Isocontours of the magnitude of the electric field in a point-to-plane configuration.

point-to-plane geometry. Even though the simulations are done in a point-to-point configurations, these initial tests are merely just to verify the agreement between the analytical and numerical solutions. Moreover, the point-to-plane study presented here has the same hyperbolic geometry as the point-to-point case studied in the following. The accuracy found here is therefore considered as general for any hyperbolic geometry (point-to-plane, point-to-point). In reference to Figures 3 and 4, the hyperbolic anode is set to 7 kV and is located at $x = 0.25$ cm, and the grounded plane cathode is set to 0 V and is located at $x = 0$ cm. For the parameters of the electrode, $a = 0.25$ cm and $b = 0.9$ cm, yielding a radius of curvature at the tip of the electrode of $b^2/a = 296$ μm. For the initial tests, the computational domain dimensions are 1 x 0.5 cm^2 and is discretized using the Cartesian grid in both axial and radial directions. The total number of cells are 400 x 400, giving 25 μm and 12.5 μm cell sizes for the axial and radial directions, respectively. Although these are rather large cells compared to typical cell sizes, it can be seen that this grid is suffi-

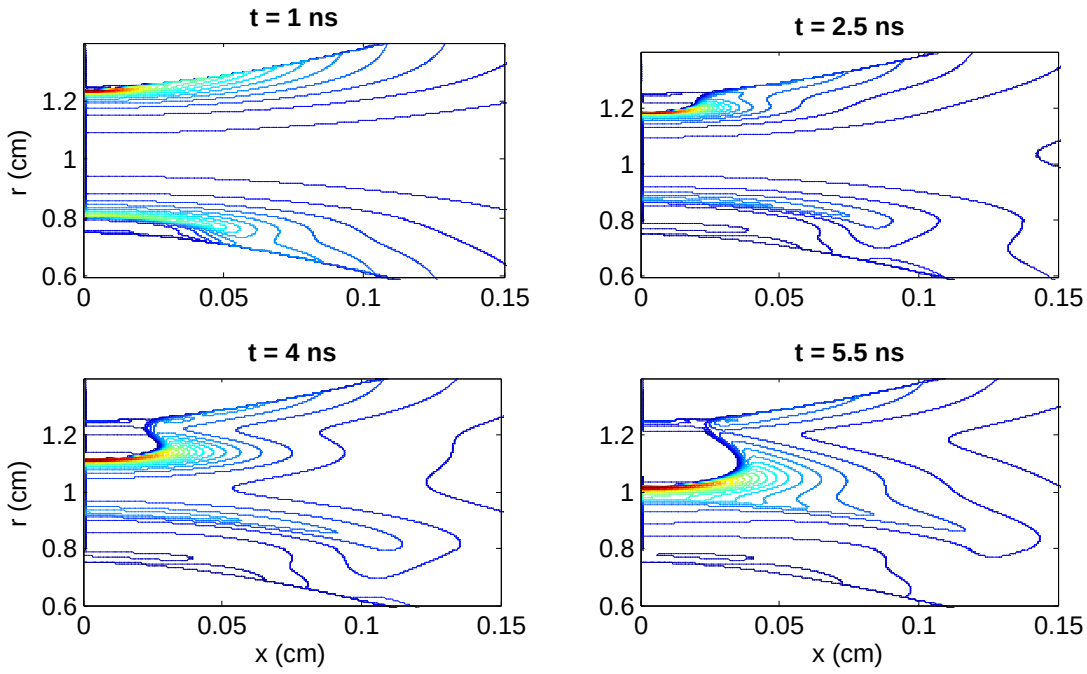


Figure 5: Isocontours of the electric field for a streamer simulation in point-to-point configuration for times at 1, 2.5, 4, and 5.5 ns.

cient to provide good agreement between numerical and analytical results as will be illustrated by subsequent analysis. Figure 3 shows the equipotential contours of the set-up. Equation (15) describes the analytical solution, while the Poisson iterative solver gives the numerical result. There is very good agreement between the two solutions, since they overlap. Figure 4 shows the isocontours of the electric field for the same set-up. Equation (19) describes the analytical solution, while a numerical gradient using the finite-differences method applied to the potentials found by the Poisson iterative solver is used for the numerical result. There is also a good agreement between the two solutions, since they overlap. The biggest discrepancy between the two solutions is, at the most, on the order of 10^{-3} μm between two isocontours. This is very small, but can easily be further reduced by adding more cells, which will be done in the actual streamer simulations. With these agreeing results for the initial data, an analysis of the results given by the streamer simulations for varying radius of curvature for the hyperbolic electrode configurations can be done.

Streamer Simulation Parameters

Before moving on to the results of differing radius of curvatures, it is important to discuss the parameters of the streamer simulations adapted in this work. In the streamer simulations presented here, we used a numerical method known as the ghost fluid method (GFM) introduced by Fedkiw *et al.* [1999] to solve Poisson's equation and adapted to streamer simulations by the works by Celestin *et al.*

[2009]. In order to investigate how streamer simulations depend on electrode configurations, our point of reference will be the configuration shown in Figure 5. The hyperboloid anode (electrode on top) is set to 13 kV and is located at r = 1.25 cm, while the grounded hyperbolic cathode (electrode on bottom) is set to 0 V and is located at r = 0.75 cm. The parameters of the electrode are the same as discussed for the initial conditions (a = 0.25 cm, b = 0.09 cm, radius of curvature at the tip of the electrode is 296 μm). The computational domain is a grid with dimensions 0.5 x 2 cm^2. The initial condition is a continuous pre-ionized neutral background of 10^9 m^{-3}, considered to be due to the repetitive rate of pulses of the power supply [Celestin *et al.*, 2009]. It is important to note that Figure 5 is zoomed in to 0 cm $< x <$ 0.15 cm and 0.6 cm $< r <$ 1.4 cm region. This is merely to get a clearer image of the simulation. The total number of cells is 347 x 1510, which yields cell sizes of about 14.41 μm and 13.25 μm in the axial and radial directions, respectively. It is important to discuss some of the characteristics of a streamer in these simulations. In Figure 5, there is a positive and negative streamer propagating from the anode (top) and cathode (bottom), respectively. The electric field of the head of the positive streamer is much greater than that of the negative streamer. It takes between 5.5 ns and 6.0 ns for the connection between streamer heads. There is no dramatic increase in the electric field right before connection, inferring that the connection is smooth. We next present results for more complex configurations illustrating dependance on electrode shape.

Hyperbolic Electrodes of Varying Radius of Curvature

In this section, the results from the streamer simulations of hyperbolic electrodes with varying shapes in point-to-point configurations are presented. The electrode shape that will be used as reference for comparison is the one with radius of curvature of 296 μm. Although many simulations were done for different electrode curvatures, the two to be focused on in this work are those with radius of curvature of 150 μm and 400 μm (for the rest of this work, to avoid redundancy, each configuration will be referred to as electrode "radius of curvature value in μm", i.e., electrode 296).

When discussing Figure 6, it is important to note that the enhancements of electric field (which look like spikes) on each graph represent the electric field in the streamer head. In reference to Figure 6, it can be seen that, for electrode 296, the electric field in the positive streamer head is increasing with time from the anode side, until a certain time, whereas the electric field in the positive streamer head starts to decrease. Alternatively, the electric field in the negative streamer head is continually decreasing as it migrates from cathode side. It can be seen then, comparatively, that the electrode 400 follows these same trends. The electrode 150 configuration follows the same trend as electrode 400 and 296 from the cathode side, but the interesting part arises from the electric field in the positive streamer head as it propagates from the anode. This smaller radius of curvature produces a

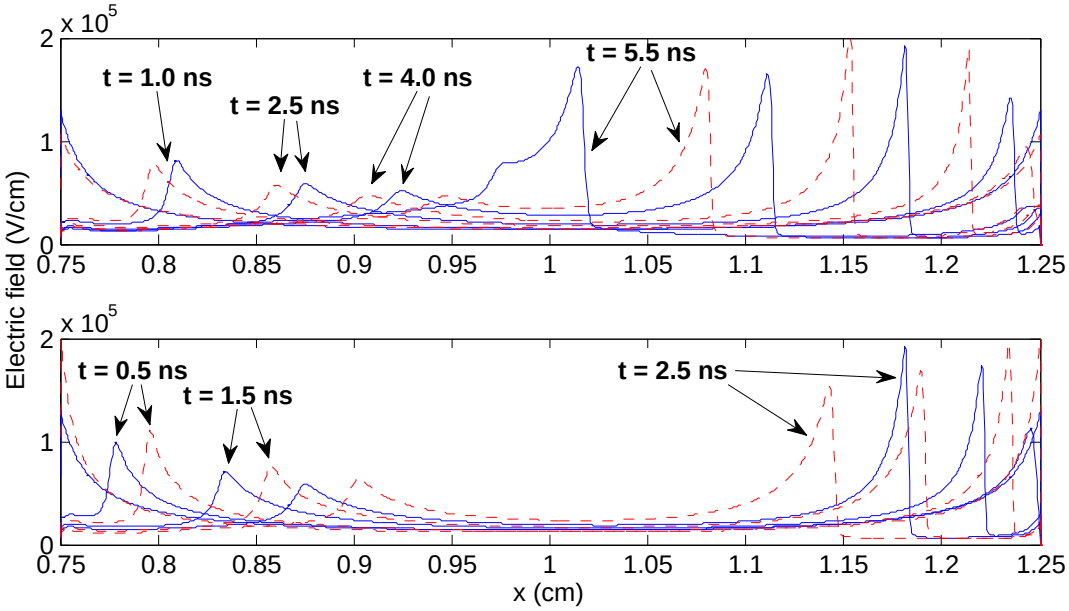


Figure 6: Electric field on the axis of symmetry. Top figure displays times between 1.0 ns and 5.5 ns with a time step of 1.5 ns. Bottom figure displays times between 0.5 ns and 2.5 ns with a time step of 1 ns. The Laplacian field at $t = 0$ is included for each plot. Solid line in both figures represents hyperbolic electrodes with a radius of curvature of 296 μm. The dashed lines for the top and bottom figures represent hyperbolic electrodes with radius of curvatures of 400 μm and 150 μm, respectively.

positive streamer head that has an electric field that decreases as it migrates from the anode. It is important to note that the simulation for electrode 150 stopped at 3 ns. It is unclear whether or not this characteristic of a decreasing electric field in the positive streamer head migrating from the anode continues until the connection.

When comparing electrodes 296 and 400, it can be seen from Figure 6 that the streamer head of electrode 296 moves faster than that of electrode 400. From the anode, it seems that the electric field intensity is increasing faster for electrode 400 than electrode 296 as the positive streamer head migrates away from the electrode. This is interesting because from the cathode, the electric field intensity in the negative streamer head seems to be decreasing at the same rate. These results entail a simple behavior from the cathode and a more complex behavior from the anode. Alternatively, when comparing electrodes 296 and 150, the streamer head of electrode 150 moves faster than that of electrode 296. Electrode 296's electric field intensity in the positive streamer head decreases more slowly than the increase in electrode 150's electric field intensity in the streamer head as they migrate from the anode. From the cathode, the negative streamer head's field intensity tends to decreases at the same rate for both electrodes. It is also interesting to note that at any point in the regions behind the positive and negative streamer heads (streamer body), the electric field for all configurations have similar values.

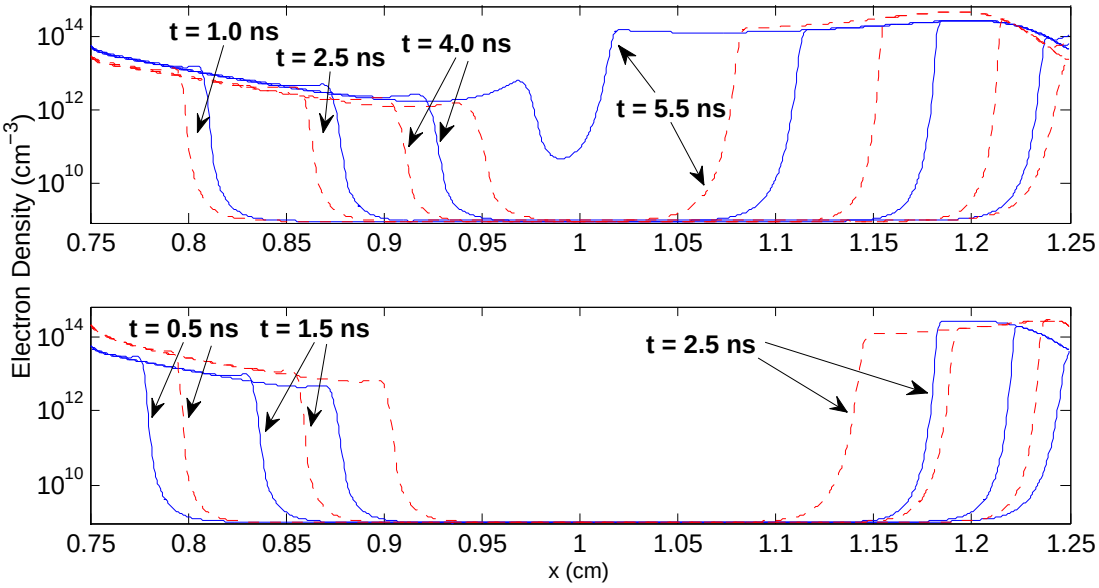


Figure 7: Electron density on the axis of symmetry. Top figure displays times between 1.0 ns and 5.5 ns with a time step of 1.5 ns. Bottom figure displays times between 0.5 ns and 2.5 ns with a time step of 1 ns. Solid line in both figures represents hyperbolic electrodes with a radius of curvature of 296 μm. Dashed lines for top and bottom figures represent hyperbolic electrodes with radius of curvatures of 400 μm and 150 μm, respectively.

Similar results can be observed when comparing electric field and electron density along the axis of symmetry. To avoid redundancies, previously discussed similarities will not be reiterated. To begin, results from the hyperbolic cathode will be discussed, since they follow more logical trends. In reference to the top figure of Figure 7, it can be seen that at any point on the axis before connection, the electron density for electrode 296 is always higher than that of electrode 400. Alternatively, from the bottom figure, electrode 150 always has a higher electron density than that of electrode 296. For all electrode configurations, the electron density in the streamer body tends to decrease as the negative streamer head migrates from the cathode. It is important to note that for electron 296, the electron density has an increase in the negative streamer head, and this increase tends to be larger as the negative streamer head migrates from the cathode. For electrodes 400 and 150, the discussed effect is not noticeable. From the anode side, there are a few effects to notice. For all electrode configurations, there is an increase in electron density as the positive streamer head migrates until a certain point, whereas the electron density starts to decrease. For electrode 150, the increase is sharp and happens quickly. For electrode 400, the increase is more subtle and happens slowly. For electrode 296, the trend is somewhere between what was described for electrode 150 and 400.

It is interesting to note some of the results from simulations that were attained for the same electrode curvatures with the hyperbolic anode set to 7 kV. Results

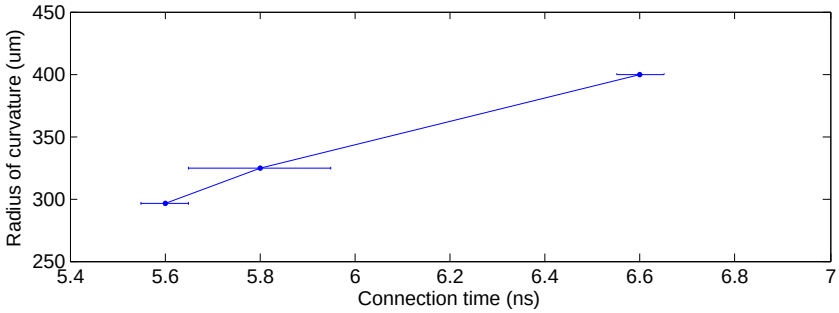


Figure 8: A comparison between the time it takes for a connection between positive and negative streamer heads and the radius of curvature for each hyperbolic electrode configuration studied in the present work.

indicate that for all three electrode curvatures studied, no streamer was produced from the anode. Though there was a streamer head produced from the cathode, it quickly died out before reaching the half-way point. In Figure 8, a plot of the connection time versus radius of curvature is presented. Since the simulation has a 0.5 ns time step, there are error bars associated with the approximated connection times.

CONCLUSIONS

From the results presented in this paper, some conclusions can be made on the dependance of hyperbolic electrode shapes on streamer propagation at atmospheric pressure in air. When considering propagation velocity, it is obvious to see that a smaller radius of curvature produces a higher electric field intensity near the tip of the hyperbolic electrode. From this, we see that the streamer head travels faster from both the anode and cathode with a small radius of curvature and travels slower for a large radius of curvature. While this is true, it is important to take into account the fact that the streamer head produced from the anode of the hyperbolic electrode with radius of curvature of 150 μm had an electric field intensity that initially decreased as the streamer head migrated, while the other two configurations had an initial increase in electric field intensity. It can also be seen that a smaller radius of curvature produces a higher overall electron density. In reference to Figure 8, it is also interesting to note that as the radius of curvature of a hyperbolic electrode in point-to-point configuration increases, so does the connection time between positive and negative streamer heads. To conclude, it was found that the shape of a hyperbolic electrode is important in the modeling of streamer propagation. Additionally, obtained results demonstrate that the shape of the hyperbolic electrode affect the time it takes for a discharge to occur, the diameter of the plasma filament, and the energy dissipated, among other parameters. These findings will prove beneficial for the applications of streamers.

We established that drastic changes in the radius of curvature did not change the connection time much for streamers compared to typical durations of an applied voltage pulse of 10 ns, used for example in experiments reported in Pai *et al.* [2008]. This indicates that the approximations made on the electrode shapes to model experiments (much smaller than the differences in radius of curvature studied here) are valid.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Jeremy Riousset and Mr. Sebastien de Larquier for their help on various software tools. This material is based upon work supported by the National Science Foundation under Grants No. EEC-0755081 and ATM-0734083.

REFERENCES

S. Celestin, Z. Bonaventura, B. Zeghondy, A. Bourdon and P. Segur, "The use of the ghost fluid method for Poisson's equation to simulate streamer propagation in point-to-plane and point-to-point geometries," *Journal of Physics D-Applied Physics*, **42** (6), doi:10.1088/0022-3727/42/6/065203 (2009).

L. Loeb, "Ionizing Waves of Potential Gradient," *Science*, **148** (3676) 1417–& (1965).

A. Kulikovsky, "Positive streamer between parallel plate electrodes in atmospheric pressure air," *Journal of Physics D-Applied Physics*, **30** (3) 441–450 (1997).

S. Dhali and P. Williams, "Two-dimensional studies of streamers in gases," *Journal of Applied Physics*, **62** (12) 4696–4707 (1987).

A. Luque, V. Ratushnaya and U. Ebert, "Positive and negative streamers in ambient air: modelling evolution and velocities," *Journal of Physics D-Applied Physics*, **41** (23), doi:10.1088/0022-3727/41/23/234005 (2008).

D. Djermoune, S. Samson, E. Marode and S. P, "A time resolved two dimensional modelling of the electric behaviour and the chemical yield of the streamer induced discharge," *Proc. 11th Int. Conf on Gas Discharges and Their Applications (Chuo University, Tokyo, Japan)*, pp. 33–4 (1995).

R. Morrow and J. Lowke, "Streamer propagation in air," *Journal of Physics D-Applied Physics*, **30** (4) 614–627 (1997).

R. Fedkiw, T. Aslam, B. Merriman and S. Osher, "A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the ghost fluid method)," *Journal of Computational Physics*, **152** (2) 457–492 (1999).

D. Pai, D. A. Lacoste and C. O. Laux, "Images of nanosecond repetitively pulsed plasmas in preheated air at atmospheric pressure," *IEEE Transactions on Plasma Science*, **36** (4, Part 1) 974–975, doi:10.1109/TPS.2008.924484 (2008).

U. Kogelschatz, B. Eliasson and W. Egli, "From ozone generators to flat television screens: history and future potential of dielectric-barrier discharges," *Pure and Applied Chemistry*, **71** (10) 1819–1828 (1999).

G. Pilla, D. Galley, D. A. Lacoste, F. Lacas, D. Veynante and C. O. Laux, "Stabilization of a turbulent premixed flame using a nanosecond repetitively pulsed plasma," *IEEE Transactions on Plasma Science*, **34** (6, Part 1) 2471–2477, doi: 10.1109/TPS.2006.886081 (2006).

C. Eyring, S. Mackeown and R. Millikan, "Fields currents from points," *Physical Review*, **31** (5) 0900–0909 (1928).

S. Pancheshnyi, S. Starikovskaia and A. Starikovskii, "Role of photoionization processes in propagation of cathode-directed streamer," *JOURNAL OF PHYSICS D-APPLIED PHYSICS*, **34** (1) 105–115 (2001).

H. Stone, "Iterative Solution of Implicit Approximations of Multidimensional Partial Differential Equations," *SIAM Journal on Numerical Analysis*, **5** (3) 530–& (1968).

HIGH TEMPORAL AND SPATIAL RESOLUTION STUDIES OF METEOR HEAD ECHOES AND THE GYRO LINE OBSERVED OVER THE ARECIBO OBSERVATORY

James A. Reuss*, Elijah Hibit†, and Julio Urbina‡

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate student of
Department of Physics
St. Olaf College
Northfield, MN 55057

ABSTRACT

Using the High Power Large Aperture Incoherent Scatter Radar at the Arecibo Observatory we are able to study microgram meteors both as they enter the earth's atmosphere and their gyro line frequencies that are observed when they collide with molecules in the upper atmosphere and liberate electrons. By modeling the signal to noise ratio and frequency shift of the received signal over time and a range of altitudes, meteor head echoes can be examined to determine both the velocity and mass of incoming meteors. By implementing a fast Fourier transform on the same data, plots of the ion and gyro lines can be produced to precisely specify their frequencies at a given time. The data was collected at the Arecibo Observatory with a bandwidth of 4MHz, twice the bandwidth of any previous study. Because of this, an image of higher resolution than any previous work is expected.

This experiment serves to further the method that *Janches and Nicolls* [2007] used in order to get much higher resolution studies of the *E*- and *F*1 regions of the atmosphere. By using the same pulse length and inter-pulse period but more bandwidth, this work uses a sampling rate with twice the frequency from the previous study, allowing both altitude and frequency resolutions in these altitudes that

†Graduate Mentor
‡Faculty Mentor

have never been seen before. Through this, conclusions from earlier work can be refined while having the resolution necessary to observe any other trends hidden in the noise of other experiments. This study observed a new frequency, the plasma line, present in the gyro line plots not seen in earlier research due to bandwidth, and confirmed the findings from earlier reports for both the gyro line and meteor head echoes.

INTRODUCTION

Preliminary research in meteors left many equations for individual phenomenon, but no clear theory linking these or explaining how energy is dissipated into the atmosphere from incoming meteors. Originally, meteors were studied with radar by looking at the head echo. Using the Doppler effect, the velocity of incoming meteors can be calculated, while other methods can estimate mass. Radars became more powerful so smaller meteors could be detected, leading to the current detection of particles of mass 1 μg by using high-power, large-aperture (HPLA) systems. These small meteors impact molecules in the atmosphere, causing a trail of ions to form behind them.

Discovered relatively recently was the gyro line frequency, the frequency at which liberated electrons in this region oscillate due to the Earth's magnetic field. This frequency has been shown to be directly related to the electron density at those altitudes making it a useful study to develop a theory of the atmosphere.

Head echoes can be analyzed from the data collected using similar methods to *Dyrud and Janches* [2008]. Initially, data was only available for large meteors. Even as detection techniques improved, signal processing methods involved dividing by both radar cross section and antenna gain, small numbers that introduced large uncertainties into the results. *Dyrud and Janches* [2008] devised a method that avoids both these problems by using the signal to noise ratio for the calculations, the same method used in this analysis.

This study is concerned with measuring the incoming meteor's velocity and mass. By looking at the Doppler shift present in the data, the meteor's relative velocity can be determined while examining the velocities over a range of altitudes gives a reasonable approximation of the mass. Another approximation used was first implemented by *Dyrud and Janches* [2008] where the angle between the incoming meteor and the horizon is given by,

$$\alpha \approx arctan(\Delta h/B) \tag{1}$$

where α is the angle of the incoming meteor, Δh is the total altitude range of the observed meteor, and B is the radar beam diameter, using 500 m as advised in previous work.[1]

For the second part of this study the gyro resonance line will be plotted. This was predicted to exist in 1961 by *Salpeter* and remained a theory until 1978 when an experiment led by *Behnke and Hagen* [1978] observed this line at Arecibo us-

ing the incoherent scatter radar.[2] In another study, *Bhatt et al.* [2006] gave the approximate frequency offset where this line will appear, given by,

$$\omega^2 = \Omega_e^2 cos^2\alpha(1 + \frac{11k^2v_{th}^2}{4\Omega_e^2} - \frac{\Omega_e^2}{\omega_{pe}^2}) \tag{2}$$

which simplifies to

$$\omega \approx \Omega_e \; cos \; \alpha \tag{3}$$

as the last two terms are usually small, where ω is the gyro wave mode, Ω_e is the electron gyro frequency given by $\Omega_e = eB/m_e$, and α is the angle between the radar beam and the Earth's magnetic field lines. In order to resolve the line, *Behnke and Hagen* [1978] used integrations over two hours. The first experiment to detect this line used a 500μs pulse. Since then, it has been shown in experiments by both *Bhatt et al.* [2006] and *Janches and Nicolls* [2007] that using pulses and integrations of much shorter times could still resolve this line. In this study as well as the study conducted by *Janches and Nicolls* [2007], a pulse length of 45μs was used, more than a factor of 10 shorter than the initial experiments.[3] In addition to this, *Janches and Nicolls* [2007] were able to resolve the line using only seven minute integrations of the data, allowing them to study changes in this region throughout the course of a day, unlike the initial experiment which needed the long integrations. These shorter integrations involved data sampled with higher bandwidths (up to 2MHz), allowing for higher altitude resolution which was as precise as 75 m for the most recent experiment.[3]

Previous studies have concluded that the gyro line is most easily and clearly observed in these data during sunrise and sunset because it is directly related to the electron density in the *E*- and *F*1 regions of the ionosphere.[3] These conclusions confirm that the diurnal changes undergone by the gyro line occur most rapidly and are seen most clearly during sunset and sunrise. This can be accentuated by placing more emphasis on data from those periods during the day and less on observations made during the night or the middle of the day, as the gyro line is not present during the night and is obscured by the low ratio of electron-to-ion energy during the middle of the day.

By increasing the bandwidth, *Janches and Nicolls* [2007] were able to get higher altitude resolution and more precise results of the frequency domain by having a much higher sampling rate than previous studies. This work followed by doubling the sampling rate used by *Janches and Nicolls* [2007], giving unmatched altitude resolution helpful in the analysis of head echoes and more precision when examining the frequencies of the observed gyro line. Data analysis will be similar to the methods used by *Janches and Nicolls* [2007], but should be able to more clearly identify the patterns they describe and look for any that were previously unseen.

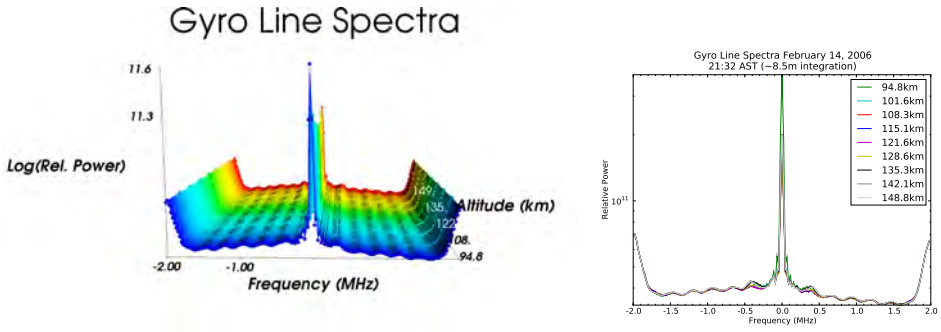


Figure 1: The biggest feature is the ion line, centered in both figures. The gyro line is the first peak on either side of the ion line, offset ∼400kHz.

EXPERIMENTAL DESCRIPTION

Data was gathered using the HPLA incoherent scatter radar at the Arecibo Observatory. A pulse modulated signal of length 45μs was transmitted on a carrier frequency of 430MHz and echoes were listened to during the inter-pulse period (IPP) of 2ms. The only parameter that was changed in this experiment compared to the *Janches and Nicolls* [2007] study with respect to data collection was the increase in bandwidth. Specifically, a 4MHz digital receiver was used as opposed to the 2MHz analog used in the previous experiment.[4] This increase in bandwidth serves to give better altitude resolution and an increased sampling rate as confirmed by *Janches and Nicolls* [2007]. This double in the sampling rate yields a much higher resolution signal. In their paper, *Janches and Nicolls* [2007] indicate that the maximum frequency resolution is the inverse length of the pulse used, giving a resolution of approximately 22kHz. Because the pulse is short, the results are limited to altitudes at or below 170 km, while the much longer 500μs pulse used by *Bhatt et al.* [2006] was able to study phenomina at altitudes up to 345 km.

Gyro Line

In order to get meaningful results about the gyro line, a fast Fourier transform (FFT) was used on the data to see what frequencies were present. The ion line was the central frequency observed an order of magnitude larger than any other features, but this investigation is concerned with the gyro line. Offset approximately 400kHz from this central ion line is the expected gyro line. It appears both above and below the central frequency as the results were symmetric about the ion line. In order to further understand the effect of frequency and magnitude of the gyro line changes on altitude, 3-D mesh plots connecting the varying altitudes were created which allow subtle trends to become more obvious (Figure 1).

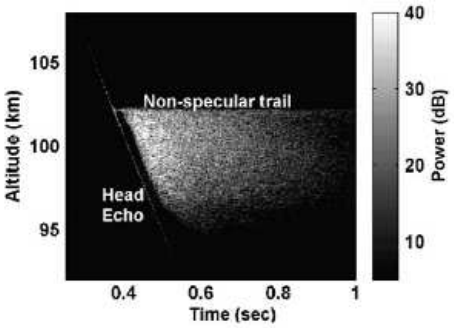


Figure 2: A typical Range-Time Intensity (RTI) plot. The fine diagonal line is the head echo from which meteor velocity can be determined.

To better resolve the gyro line, the resulting signal in each IPP corresponding to altitudes near 300 km is considered. By assuming that the results from this altitude contain only random noise with no signal contributing to the gyro line, the noise can be modeled. For each IPP, the noise is modeled and then subtracted from the remaining data in that IPP. This is performed on each individual IPP, based on the assumption that the noise at 300 km is of the same level as the noise at the altitudes of interest. To verify that these assumptions are correct, the signal will be modeled from 280 to 300 km and manually checked to ensure randomness (i.e. a Gaussian distribution).

Another necessary parameter to consider is the number of points to use in the FFT. If not enough points are used, then the results will be the same as the study sampled at 2MHz. If too many are used, the algorithm would have to interpolate to be able to plot all the points, leading to results that are not a true representation of the data. The simple solution is to use the same number of points in the FFT as gathered, giving no interpolation with the maximum resolution. Unfortunately, the algorithms implemented here, both the fftw routine from fftw.org and CUDA using the CUFFT library, the NVIDIA algorithm, work much more efficiently with a number of points that can be represented as 2^n where n is an integer. To make sure the results have better resolution than the previous work, 256 points are used in the model which provided both clearer results and a reasonable representation of the resolution expected given the sampling rate.

Head Echoes

The second part of this project involved examining the meteor head echoes present in the data. To do this, the signal to noise ratio (SNR) of a sample is modeled with no event present, then compared as a baseline to our data set. Any event that was above this threshold was an incoming meteor. To visualize the results,

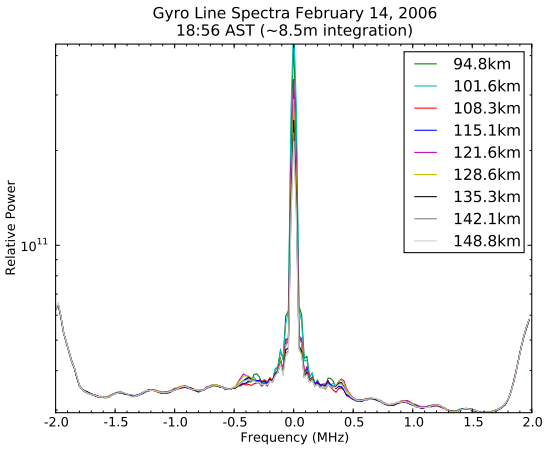


Figure 3: Peaks at the fringes of our bandwidth are present in all the data corresponding primarily to the plasma frequency with a contribution from a harmonic of the electron gyro frequency.

Range-Time Intensity (RTI) plots were used (Figure 2).

Once the data is isolated for each meteor event, velocities are calculated.[1] This can be determined by examining the frequency shift observed with respect to the initial transmitted frequency. Any strong signal that shifted the frequency higher was due to an incoming meteor where greater differences between the transmitted and received frequencies corresponded to higher meteor velocities. An FFT is performed on the received data, looking for strong signals higher than the 430MHz carrier signal. The velocity is then determined based on the precision of the received frequency.

Using the same method as *Dyrud and Janches* [2008], the mass of the meteor can be calculated by looking at how this shift changes over time from the instant it was observed until it no longer gives us an echo above the baseline. More massive meteors will be observed over a much larger range of altitudes as they do not burn up as quickly as smaller meteors. They also take much longer decelerate because they are entering the atmosphere with more momentum, leading to their presence at lower altitudes.

RESULTS

Gyro Line

The initial results involved studying the gyro line phenomenon. Transforms with 256 points were used to quickly get through some data and see the resulting graphs, comparing those to previous work to verify both the data and program used to analyze it. These first plots yielded an unexpected result. Offeset $\sim$1.50MHz

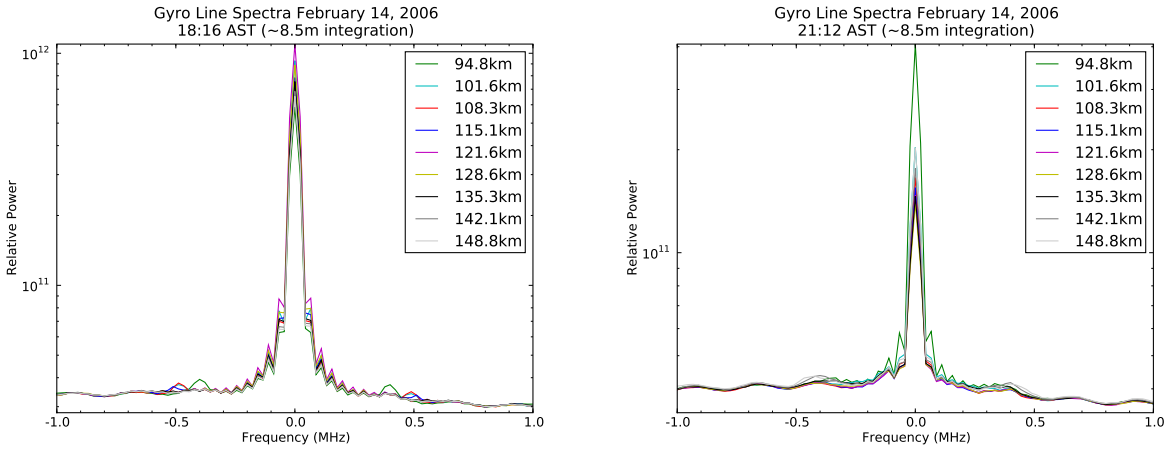


Figure 4: The figure on the left shows the observed gyro line offset from the ion line at various altitudes. The peaks near 400kHz are from the lowest altitudes studied while it is too early to observe the line for higher altitudes. The plot on the right is much later in the day when the line is observed at high altitudes with much lower precision.

from the central ion line, there is a spike observed on either side of the ion line (Figure 3). Upon reviewing the method used, it was concluded that these peaks may be a result of using padding in the arrays of data due to using an FFT with 256 points and not 180 points, corresponding to one point for each point in the data. To resolve this issue, the FFT was recalculated using the slower 180 point method to ensure the resulting graph corresponds to the data. This method was implemented on over fifty data sets collected throughout the day of February 14, 2006. The slower 180 point method confirmed that the faster 256 point method was correct even though padding was present.

A closer look into these peaks reveals that they correspond primarily to the plasma line with a contribution from the second harmonic of the electron gyro line. This phenomenon could not be observed with the 2MHz analog system as its frequency is offset too far from the ion frequency.

The main concern is not this plasma peak, but the gyro resonance line offset by approximately 400kHz from the central ion line. Data collected during the middle of the day confirmed the result of diurnal variability as concluded by *Janches and Nicolls* [2007] where the central ion line overpowers the gyro line and it effectively disappears from the graph. The goal is tracking how this line changes over the course of a day, so the most important data to analyze happens between the transition of the line being completely obscured by the ion line and the disappearance of the line as night comes.

The line could first be resolved at the lowest altitude studied, 94 km, at 17:47 AST. At this altitude, the gyro frequency was initially offset by $\sim$420kHz from the ion line. Tracking this line over time showed that the offset frequency slowly shifts toward zero when the sun finally sets. By 18:41, it is easily seen that the resonance

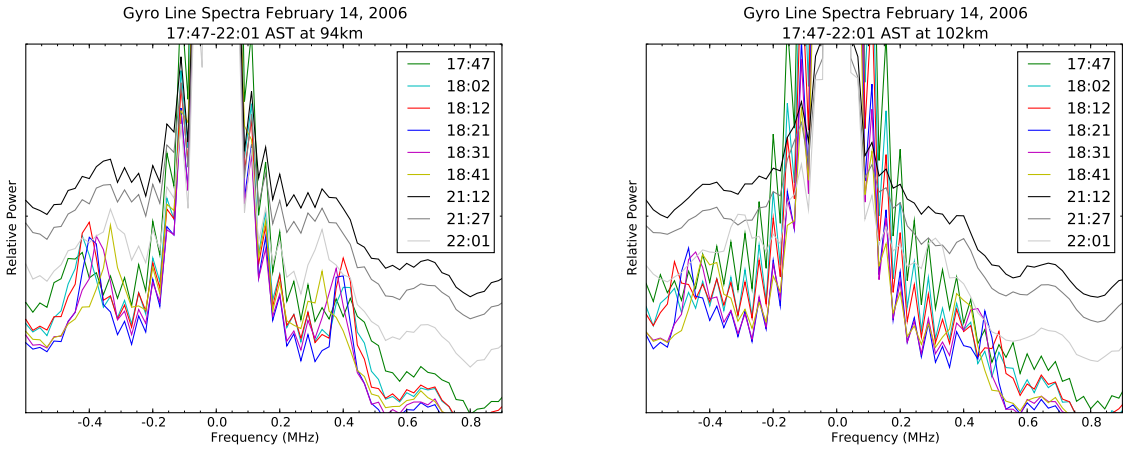


Figure 5: These plots of the two lowest altitudes show the variation at their respective altitude over the course of four hours. The gyro line steadily moves towards zero frequency offset from the central ion line.

line at this lowest altitude has shifted by over 75kHz. This trend continues through all the altitudes studied, though at increasingly later times as it must be later in the day for the gyro line to be detected at higher altitudes. At the highest altitudes this slow frequency shift is still present, though the strength of the signal is greatly diminished.

When comparing the plot of data from 18:16 AST to that from 21:12 AST, it is apparent that the magnitude of this event diminishes as the sun sets, this can be seen in Figure 4. At this later time the higher altitudes are beginning to give a stronger gyro line response, but not nearly as much as the 94 km line did when it displayed its strongest power strength seen over three hours earlier. As the sun finally sets, the frequency offset shifts toward zero for all altitudes, though the time to reach zero appears to be related to the altitude. To clarify this event, plots were made of the lowest two altitudes sampled, 94 and 102 km, showing this trend. These low altitudes were chosen as they displayed the most precise gyro line frequency (Figure 5).

Head Echoes

By using an algorithm that compared each received signal to a sample of background noise, meteors were found. As expected, there was no correlation between the time the data was collected and the number of meteors observed. Of the 50 sets examined, 11 of them contained meteor signals with one set containing more than thirty of these events. A majority of sets contained no echoes from meteors.

A total of 136 meteor events were observed over two hours. They were sequentially analyzed, basing the quality of the event on the clarity of its presence on

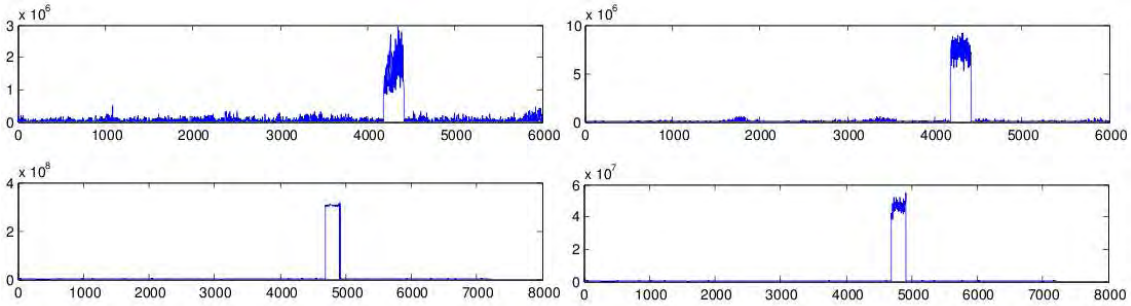


Figure 6: Four plots, clear results of a meteor detection are seen (top). Two different plots detecting the same meteor, the second detection is 12ms after the first (bottom).

a SNR plot. The velocities of these events ranged from 15 to just over 21 km/s, numbers much lower than the results of *Dyrud and Janches* [2008]. Because this research examines much higher altitudes, slower meteors are expected. In previous studies plotting meteor velocities over altitude, speeds of 20-30 km/s were observed, where the majority of meteors were observed with velocities of 22-24 km/s.[1] For this data, meteor detections ranged from 97-219 km, with most of the observations being higher than 150 km, whereas the *Dyrud and Janches* [2008] study only had data ranging from 90-102 km.

This method of signal analysis performs an FFT on the data, and uses the detection time of the signal to determine meteor altitude. This altitude is output when a meteor is detected in the data, giving a numerical value for the height of the meteor when it was detected rather than the traditional RTI plot. A plot is then made showing the SNR over time, but not based on altitude. If more than one meteor is observed in a single IPP, it is received as a very strong signal, but can be resolved as the algorithm will detect two unique events occuring at different altitudes. A typical plot can be seen in Figure 6.

Many times, one meteor appears in many inter-pulse periods as it doesn't burn up prior to being detected by more than one pulse. Through the data, the trajectory of the meteor can be tracked, giving results of the meteor's velocity over time and a range of altitudes, rather than one point per meteor as was seen in previous studies.[1] One meteor was observed in more than ten inter-pulse periods, allowing an in-depth study of its path through the atmosphere with data concerning its velocity and altitude available at each point. For the 20ms period that it was observed, it descended a total distance of 175 km, while being observed to have velocities starting at 14 km/s, accelerating to over 20 km/s, and finally ceasing to give a signal after it decelerated to 10 km/s, the whole time remaining at an altitude around 182 km. Two inter-pulse periods are displayed in Figure 6 (bottom) to show how the received signal by one meteor evolved over 12ms. By being able to compare both the SNR over time and a range of altitudes, and the frequency domain as a result of using an FFT to calculate the velocity, it can be easily determined if the signals are

from one meteor over a range of time or multiple meteors witnessed sequentially.

DISCUSSION

Overall the results confirmed the findings of *Janches and Nicolls* [2007] in how the gyro line behaves throughout the course of a day. More tests must be carried out to further examine the plasma line present, but because this work focuses on the gyro line and meteor head echoes, it is left to future experiments to examine the plasma line.

Background noise was not subtracted from the gyro line results. This is not imperative as intergrations of $\sim$8.5 minutes were used compared with shorter, 7 minute intergations used by *Janches and Nicolls* [2007]. Because of the longer integrations, the SNR can be lower while still yielding clear results. If noise was modeled and subtracted as outlined in the experimental description, the results would be cleaner, but the same overall conclusions would remain. Data from one day was analyzed in this study, by using data from multiple days, either throughout the course of a week or randomly sampled from various months throughout the year, a better understanding of the atmosphere could be gained.

The gyro resonance line first appeared at the lowest altitude at 17:47 AST. The signal from lower altitudes and at earlier times is much more clear and defined than the peaks for higher altitudes, when they finally show up. It could be useful to study even lower altitudes using the same experimental setup to be able to define the frequency of this line more precisely. By using such a short pulse, this study can not effectively study the gyro line at even higher altitudes, and it was apparent that at 150 km the signal was diminished when compared to the lowest altitude, 94 km, results.

The head echoes must be studied further as this analysis used only two hours from one day of data. Automated systems need to be implemented to sort through these events as real-time analysis is necessary to efficiently learn about these meteors. Because meteors are always incoming, they are easily observed and studied using equipment that already is in place. This method yields beneficial results as it was informative to track a single meteor through a range of altitudes and times, gathering information about how the meteor reacts over time, not simply a single velocity and mass value per incoming meteor.

The algorithm used depends upon setting a value to trigger the detection. If the SNR is not above the established baseline noise plus a user-set α value, an event would not be counted. Much work went into finding what the optimum α value is: if it is too high, only very strong signals will be detected, ignoring the many small meteors that do not produce strong signals but are still detectable. If the value is too low, noise is detected, leading to false values for the number of incoming meteors and unnecessary investigation into events that don't exist. It was found that the α value does not change appreciably during the course of a day, but must be

recalibrated when analyzing data from multiple days. There is currently no research into how much this value changes, or if it does so on any predictable basis.

COMMERCIAL APPLICATIONS

This research has two main applications. Primarily, it serves to provide information regarding the diurnal variation in a poorly understood region of the atmosphere. By conducting this study and learning about the phenomena that regularly occur, a basis for both theories and models can be provided, leading to better predictions of weather in space. The second application, using solely the head echo conclusions, is meteor burst communication, a secure and inexpensive way to send information over a thousand miles.

Results concerning both the gyro line frequency and meteor head echoes can be used to help form models of the atmosphere. The specific frequencies observed can be used to test these simulations, helping to learn how this region acts. This is not only beneficial from a knowledge perspective, it can be used to enhance our understanding of how weather patterns happen much closer to the surface of the Earth as well as giving indications of space weather. It also serves as a resource for designers of spacecraft and satellites.

The second application is meteor burst communication. Each meteor creates an ion trail behind it where the trail size is directly related to the mass of the meteor. By using the ion trail, it has been shown that voice signals normally limited to line-of-sight communication can be rebroadcast by the ion trail and sent over 2000 miles. This has been used by amateur ham radio operators to make contact across the United States.[5] Previously, such communication has been limited to short messages as each meteor can only transmit for a fraction of a second. By studying these meteors and designing antenna systems that can adjust to find them, a continuous voice signal can be broadcast.

Meteor burst communication has some advantages over conventional satellite communication, the biggest advantage being cost. Communication over meteors doesn't require building, launching, and maintaining satellites; only an initial investment in developing the technology. The meteors are always incoming, making this a technique that is always available. The second advantage is its security. Satellites rely on complex algorithms for privacy whereas meteor burst communication would require an interceptor to decode each burst they received, thousands per second, in order to decode the message.

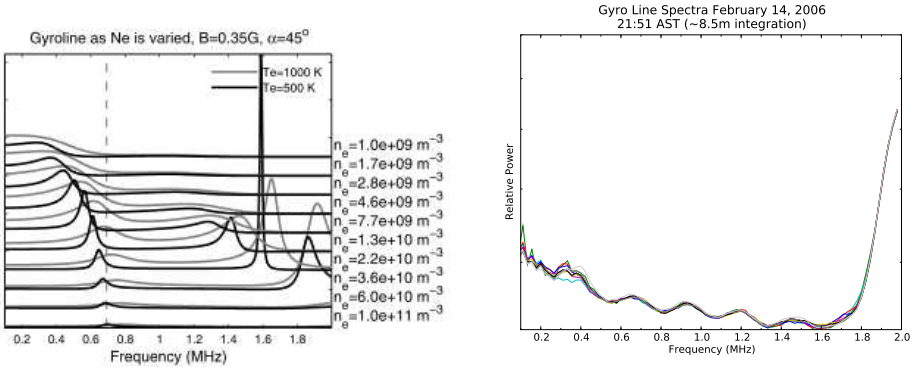


Figure 7: A direct comparison of the theoretical and experimental results showing the frequency for the plasma and gyro lines based on electron density

CONCLUSION

This research has confirmed the findings of *Janches and Nicolls* [2007] concerning the gyro line, showing both that the gyro line is completely concealed by the ion line in the middle of the day, and most clearly defined as it fluctuates, as predicted, during the sunrise and sunset periods. During the middle of the day, the ratio of electron-to-ion energy is too low to resolve the line, but this changes throughout the day, allowing the detection of the resonance line. The data from previous studies is also supported by the results found in this experiment by the primary appearance of the gyro line during the sunset period.

The plasma peaks will have to be examined further. *Janches and Nicolls* [2007] produced a plot of the theoretical frequency for various electron densities, on which the plasma line was present in the same frequency range as the data from this study confirm. To directly compare theoretical versus experimental results, a side-by-side plot is available in Figure 7. The predictions provide lower frequencies when compared to the experimental plot, suggesting the importance of further research. New experiments will be necessary to investigate this event as it only appears at the fringes of our bandwidth, possibly requiring even higher bandwidths to fully resolve and understand this line.

This analysis proved the validity of the algorithms used to detect and visualize meteors while providing both altitude and velocity information about these meteors. It was valuable to track single-meteor events through the life of the meteor after it entered the atmosphere, giving clues as to how energy is disappated as the meteor ablates. A low uncertainty measure of the amount of mass gained by the Earth through events like these does not exist. It is left to future work to conduct long term experiments to study this further.

ACKNOWLEDGEMENT
This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081.

REFERENCES

[1] L. Dyrud and D. Janches, "Modeling the meteor head echo using Arecibo radar observations," *Journal of Atmospheric and Solar-Terrestrial Physics*, **70** 1621–1632 (2008).

[2] B. A. N., "Observations of strong gyro line spectra at Arecibo near dawn," *Geophysicsl Research Lett.*, **33**, doi:1029/2006GL026139 (2006).

[3] D. Janches and M. J. Nicolls, "Diurnal variability of the gyro resonance line observed with the Arecibo incoherent scatter radar at*E*- and *F*1-region altitudes," *Geophysical Research Lett.*, **34**, doi:10.1029/2006GL028510 (2007).

[4] R. Seal, *Multi-channel digital receiver data acquisition system for the Arecibo radar system*, Master's thesis, The University of Arkansas, Little Rock, AR (2008).

[5] C. Green, "Meteor-Scatter Communications," *QST*, pp. 14–17 (1986).

CLASSIFICATION AND RADIO SCIENCE MODELING OF RADAR METEOR ECHOES

Jennifer Cross* and John D. Mathews†

Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802

*Undergraduate student in
Electrical and Computer Engineering
Franklin W. Olin College of Engineering
Needham, MA 02492

ABSTRACT

This paper documents the results of two phases of research: first, the classification of micrometeor events into a collection of specimen events; and second, the creation of a simplified radio science model able to explain a large number of these specimen events. This model is especially equipped to further investigate the existence of fragmentation within approximately 17,000 micrometeor events recorded by the Arecibo Observatory located in Puerto Rico. Our analysis of these events identified many specimens showing the cyclical constructive and destructive interference which we found to be consistent with a micrometeor consisting of a few fragments. This simple interference is indicated by regular interference cycling between multiple pulses (pulse-to-pulse) and within single radar pulses (intrapulse). Our model demonstrates these interference patterns can be completely explained and represented by a model which focuses on radio interference alone and simplifies meteor ablation, plasma physics, the antenna gain pattern and the radar equation into a singular variable. We found that while it is a relatively straight forward process to use the model to emulate these simple interference specimens, the majority of specimen events have more complex interference resulting most likely from a

†Faculty Mentor

larger number of fragments. This compound interference, while theoretically possible to generate using the current model, is too complicated to accurately emulate a specific specimen with our current brute force methods given the large number of interrelated variables.

Further development of this model could include the incorporation of factors such as antenna gain that were previously neglected in order to provide more accurate emulations which allow greater flexibility in the represented meteor science. Other model improvements such as the creation of automated emulation generation algorithms could permit the emulation of specimen with complex interference. This research provided a strong foundation for future modeling attempts while also identifying a set of specimen events which will serve as a tool for highlighting the strengths and weaknesses of current and future models.

INTRODUCTION

Every year, a huge quantity of material from space enters earth's atmosphere. The objects that make up this material flux are known as meteoroids and range in size from a micron or less (known as micrometeoroids) to larger specimens that are multiple meters in diameter. While passing through the earth's atmosphere at high velocities, these objects heat and experience a loss of material through evaporation in a process known as ablation. Meteoroids larger than a millimeter produce visible streaks of light commonly referred to as "shooting stars" that are more specifically known as meteors. The vast majority of these meteoroids disintegrate entirely before reaching the earth's surface leaving unknown quantities of matter and energy in the atmosphere.[1] Those meteoroids that are large or slow enough to be able to pass through the atmosphere to impact the earth's surface are designated meteorites, which are a source of interest for many enthusiasts and scientists who seek to learn more about meteor and planetary science.

Advancing the current state of meteor science requires us to answer many questions about the impacts that this incoming material has on the planet earth. Through the further investigation of meteoroids and meteors, we are able to gain understanding about the amount and composition of material that is deposited in the earth's atmosphere by the ablation of these objects.[1] The effect of the energy generated by the movement of the meteors also has unknown effects. Encounters with these objects may also impact orbiting space devices which do not benefit from the protection of the atmosphere and thus have potential to sustain damage from even the smallest of these meteoroids.

While only large meteors are visible to the naked eye, the introduction of RADAR brought with it the ability to observe far smaller meteors. RADAR which originally served as an acronym for "radio detecting and ranging" is used to describe any of a number of devices which utilize certain principles to collect range data about target objects. Each radar has a transmitter which produces an electromagnetic signal which travels through space to a distant object. The target reflects

some amount of this signal, the echo, and it is retransmitted in all directions. Some of this reflected signal is returned towards the radar's receiver. Once received, the echo can be used to calculate a wide variety of properties of the target, such as it's distance, radial velocity and reflectivity.[2]

As meteoroids undergo ablation through collisional interaction with the atmosphere, they reach extremely high temperatures as particles, usually atoms such as aluminum or iron, are removed from the meteoroid. These ablated particles ionize directly behind the meteoroid forming a "blob" of plasma. While the very small micrometeoroids would be invisible to the radar, this"blob" of plasma serves as a larger target to reflect a stronger echo back towards the radar receiver. The echo that is returned is called the "head echo" of the meteor.[3] These head echoes permit the measurement of the meteor's altitude, velocity, deceleration and very importantly for the focus of this paper, fragmentation. Another type of echo, known as a "anomalous trail echo" (or Range-Spread Trail-Echo) results from the radar's signal reflecting off of the trail of plasma left behind along the path of the meteoroid. These trail and head echoes of micrometeorites are most commonly observed utilizing high power, large aperture (HPLA) radars which transmit very strong signals and have very sensitive receivers.[4]

While effort and focus is being invested in understanding and modeling the plasma physics and meteor science which produce these meteor echoes (as shown in Dyrud and Janches[1]), much of what is observed may also explained and interpreted using radio science and simple radio science models. Such effort towards the characterization and classification of data recorded using HPLA radar helps to better understand the meteor phenomena that it represents. This allows greater accuracy in estimating the properties of meteors and meteoroids that are observed using HPLA radars.

In this paper, we will seek to further classify and explore the effects of fragmentation and the related radio science phenomena. Following this analysis, we will introduce models to further the verification of these classification schemes. This will provide a radio science basis for interpreting different phenomena such as signal interference. In addition, the models permit the estimation of micrometeoroid parameters such as the differential speed between fragments and meteoroids.

OBSERVATIONAL SETUP

The data we analyze in this paper were collected at the Arecibo Observatory, located in Puerto Rico. As a HPLA radar telescope, the Arecibo Observatory (AO) collects regular data on micrometeor events.[3] The data analyzed in this paper collected on June 2, 2008 and June 3, 2008 yielded approximately 10,000 and 7,000 observed micrometeor events, respectively. The data associated with these events were identified utilizing an event detection signal processing method developed by Mathews *et al.*[5] and Briczinski *et al.*[6] The data were then visually inspected as range time intensity (RTI) plots in order to perform case-by-case classification of

Figure 1: Image of Arecibo Observatory, Puerto Rico. Photograph courtesy of the NAIC - Arecibo Observatory, a facility of the NSF

different types of meteor and radio phenomena observed.

The events were simultaneously recorded using VHF and UHF radar frequencies. The UHF radar utilizes a 430 MHz signal with a very narrow (1/6°), sensitive receiver antenna beam possessing side lobes as made visible in Figure 2 and allowing it to pick-up echoes from very small, weak micrometeoroids. For the data associated with this paper, the UHF radar was set to send 20 microsecond pulses with an interpulse period of 1 millisecond. The VHF radar utilizes a 46.8 MHz signal with a wider and less sensitive receiver antenna beam (1.2°) that is co-axial with the 430 MHz antenna. This wider antenna geometry allows it to generally pick-up only stronger echoes but each remains visible in the wider beam for a longer period of time.[4] The VHF radar was set to send 10 microsecond pulses with an interpulse period of 1 millisecond. These simultaneously occurring sets of data, utilizing two radar frequencies, allowed us to have alternative perspectives on each meteor event. This yields additional information about the radio science phenomena and provides greater insight into the observed physical events than the data from either radar taken individually.

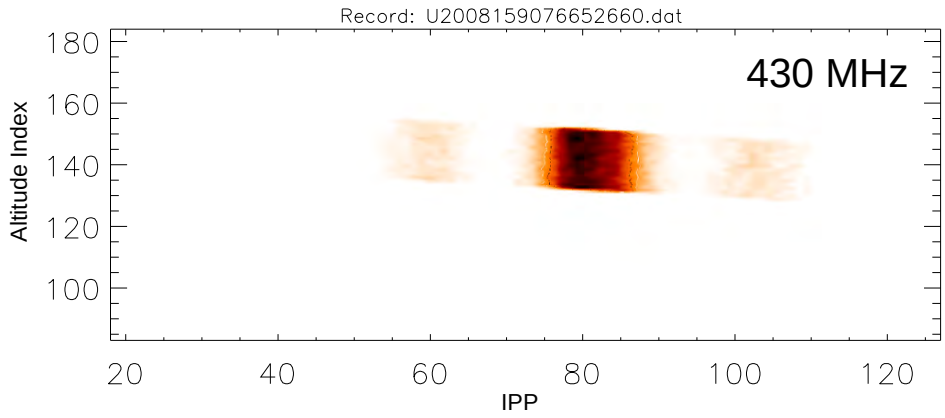


Figure 2: An example of an RTI (range-time-intensity) plot of a meteor head-echo traveling across the Arecibo Observatories UHF antenna gain pattern. In the figure, the echo from first side-lobes are visible surrounding the echo observed with the principal radar beam. The altitude index indicates 150 m changes in altitude in the 80 to 140 km region.

RESULTS AND DISCUSSION

Fragmentation Radio Science

The radar UHF and VHF signals, 430 MHz and 46.8 MHz respectively, are each, as electromagnetic signals, capable of undergoing constructive and destructive interference when two signals of similar frequencies are received simultaneously. Occurring in much the same way that the standard physics "double-slit experiment" demonstration highlights interference created by two visible light sources of similar frequencies, radar scattering targets act as the sources of the echo signals. The light intensity of the double slit at a certain spatial location is then analogous to the echo that is measured by the radar receiver. A single radar scatterer will consistently return an echo of certain power.

The introduction of a second scatterer creates the potential for the existence of signal interference. If the distance between the sources is held constant, the signals received will arrive with a constant phase difference resulting in a certain amount of constructive or destructive interference. However, in order to maintain that constant interference, the radar targets from a fragmented micrometeor must travel at precisely the same velocity towards the receiver. More frequently, as the targets separate or converge, the signal power at the receiver will vary with the cyclical transitions between constructive and destructive interference.

If the targets were simple signal sources, a complete cycle from peak interference to peak interference would occur as the distance between sources increased or

decreased by one wavelength. However since the signals reflecting from the scattering targets are required to travel twice the distance from the radar, as first a pulse then an echo, the effective distance difference between targets is doubled. This means that a complete cycle of interference happens when the distance between the scattering targets increases or decreases by just half a wavelength.

Example Case of Fragment Interference

In Figure 3, the head-echo in the RTI plot is seen clearly in the first 79 pulses at VHF frequency. At 80 pulses, or 80 milliseconds after the initial recording of the event, some meteor science event (which will not be investigated by this paper) occurs. The result appears to be an anomalous trail-echo (based on it's maintained altitude) in addition to the original head-echo which continues descending.

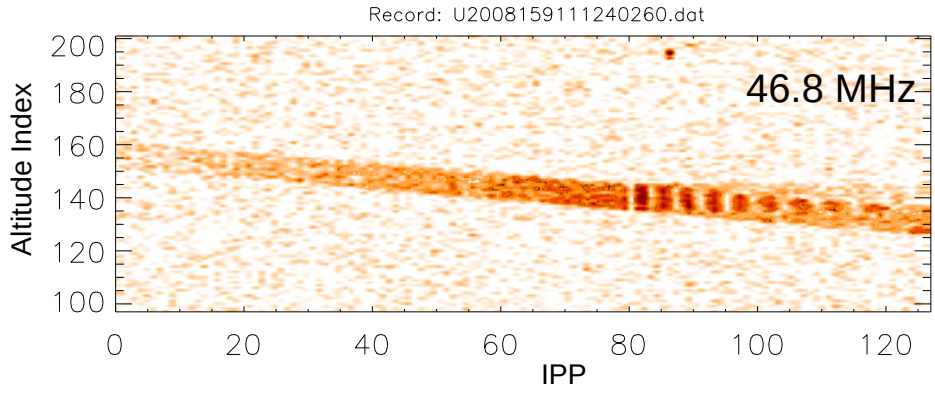


Figure 3: A VHF RTI plot showing initially a single head-echo. At IPP 80, a trail-echo appears which remains at the altitude where it was produced. The head-echo continues descending which results in some pulse-to-pulse interference, seen as the on-off modulation of the full return, between the head-echo and trail-echo.

While these signals are received simultaneously, there is visible pulse-to-pulse fading and strengthening between the two echoes. It is highly likely that this pulse-to-pulse cycling is due to the previously discussed interference, as the radio science model (to be discussed later) to recreate such an interaction is both simple and conceivably possible. As the distance between the stationary trail scatterer and the descending head-echo scatterer increases, the combined signals pass through periods of constructive and destructive interference. Interestingly, the period of these cycles noticeably increases during the observation which is indicative of a reduction in the difference between the scatterers' velocities through either the deceleration of the head-echo scatterer or an (unlikely) acceleration of the trail-echo scatterer. This radio science explanation is further augmented by modeling later in this paper.

Collection of Fundamental Interference Observations

A primary product of the research reported here, was an expansive collection of observed meteor events that were selected from the approximately 17,000 meteor events which were recorded at Arecibo Observatory over the course of two days. The selected meteor events serve as specimens that may be used to further highlight the strengths and flaws in existing and future models of meteor science, plasma physics and radio science in their abilities to explain the different specimens.

Many of the event specimens present very regular cyclical power strengthening and weakening that is consistent with the suggested radio science product of meteoroid fragmentation. In particular, the specimen plots shown in Figure 4 are representive of some of the strongest support for the fragmentation model which is capable of generating such results with the existence of two simply modeled moving point targets. The results of these models will be developed in a later section.

The specimen events were each collected in both the UHF and VHF radar results. While it is rare, likely due to reasons explained by Mathews,[4] a number of specimens exhibit the regular cyclical interference at both frequencies such as those shown in Figure 5. These models further support the radio science models which are able to easily explain the interference cycle period difference when the wavelength differences short wavelength UHF and long wavelength VHF is considered.

Basic Radio Science Model

In order to better explore, experiment and emulate events collected with the radar systems, we generated a model which incorporates as few features as are required to adequately explore the radio science events which are most often seen within the data.

The first assumption that we chose to make for consistency was to assume that all modeled events will be contained in the main antenna beam at both frequencies and move vertically and not horizontally. This allows us to neglect the variations of antenna gain and also reduces the model from the three-dimensional space to a one dimensional model along the major radar axis. For the exploration in this model, we also chose to simplify the model by summarizing the following power equation to a single constant,

$$P_R(meteor) = \frac{P_T L \lambda^2 G^2}{(4\pi)^3 r^4} \sigma_{meteor} \tag{1}$$

where P_R is the power received, P_T is the transmitter power, L is the transmitter system loss coefficient, λ is the signal wavelength, G is the antenna gain, r is the range to the target and σ_{meteor} is the scattering cross-section of the target.[4]

This choice stems from the model's focus on the radio science observations, eliminating the need to immediately consider the effects of antenna gain, the scattering cross-section or the transmitter power which will be held as a constant. Creating this relative power constant also removes the signal power's dependency on the scatterer's distance from the radar. While less than ideal, this approach lim-

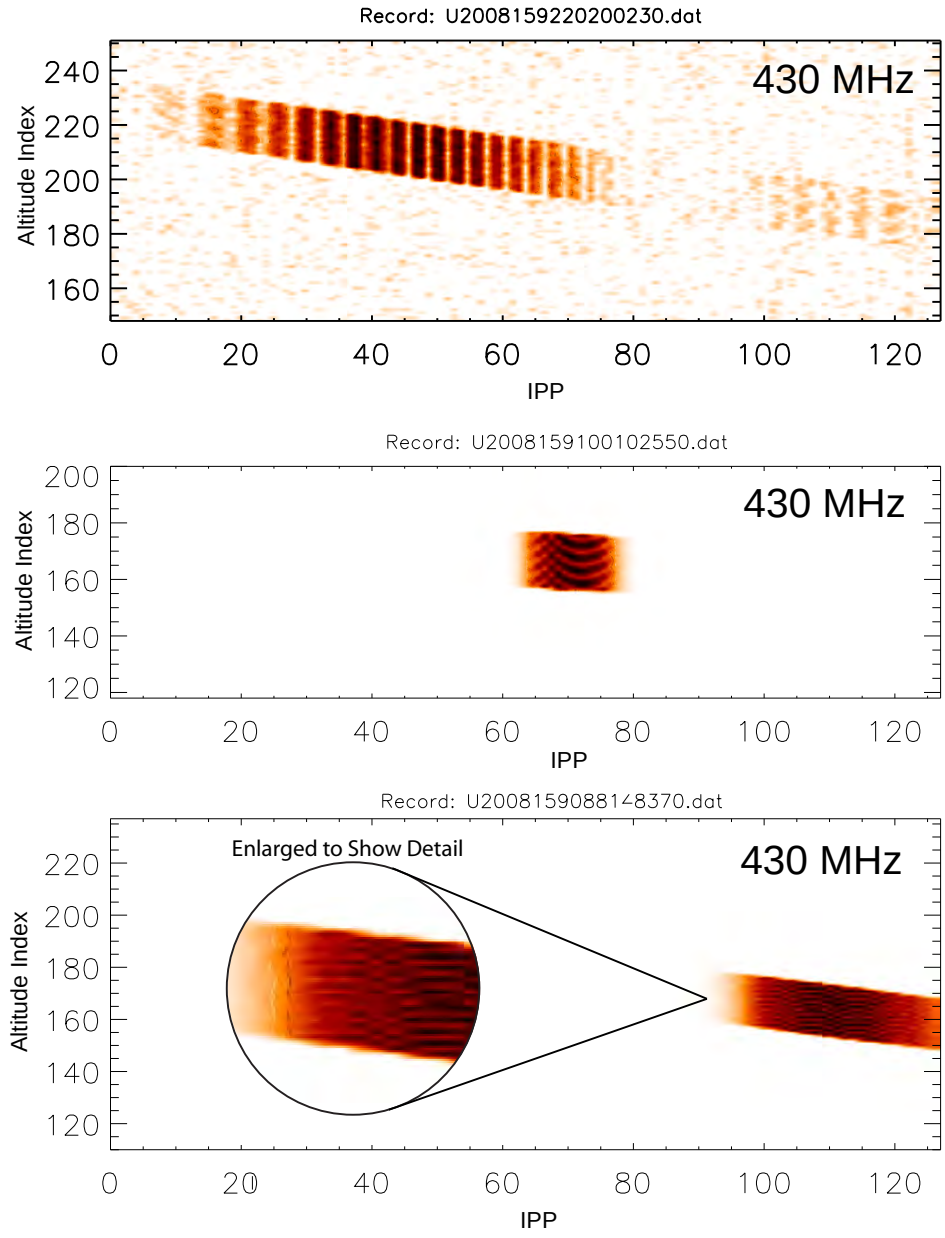


Figure 4: Three UHF RTI plots exhibiting simple interference due to a small number of fragments. Top - The pulse-to-pulse interference cycling results from a velocity difference which is approximately one half wavelength per 4 milliseconds, with much longer period fading likely caused by antenna gain. Middle - Cyclical fading internal to a single pulse indicative of a rapid velocity difference between two fragments which is approximately one half wavelength per 4 microseconds. Bottom - Very fine internal pulse structure indicating a very large velocity difference between fragments, approximately one half wavelength per microsecond.

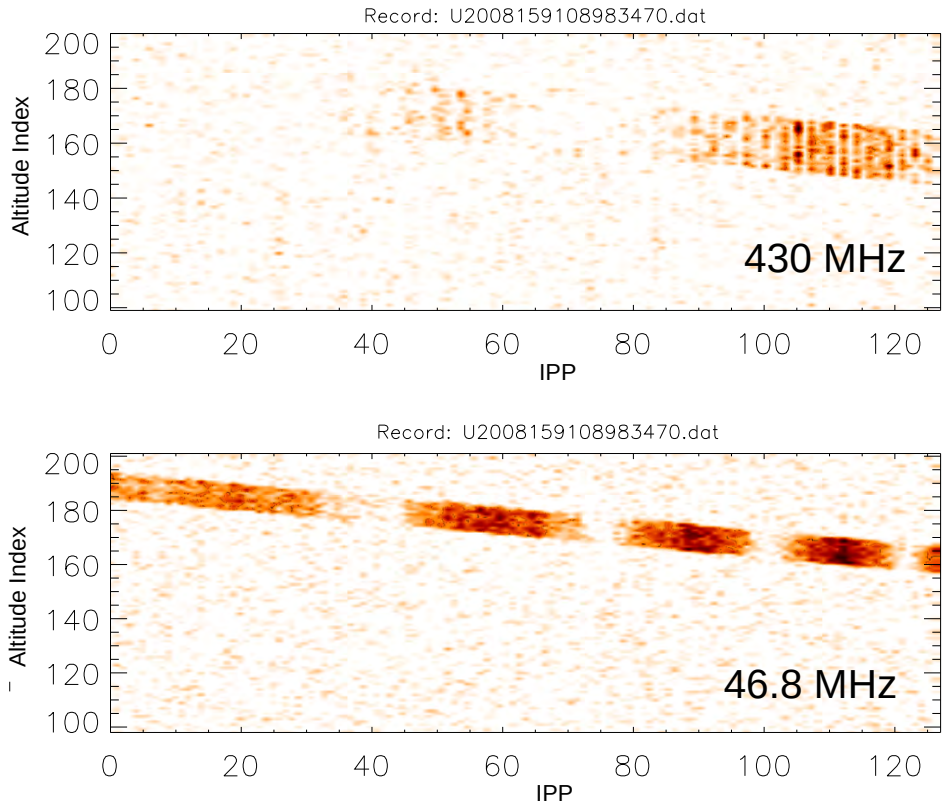


Figure 5: A pair of RTI plots, UHF (top) and VHF (bottom), resulting from the same meteor event. The UHF plot possesses regular quick pulse-to-pulse cycling while the VHF plot has a much slower pulse-to-pulse cycling period reflecting the longer VHF wavelength. Both signals also have some irregular intrapulse structuring either resulting from noise or possibly indicating numerous smaller fragments also contributing to the received signal

its the possible sources of power variations with the model to interference effects alone.

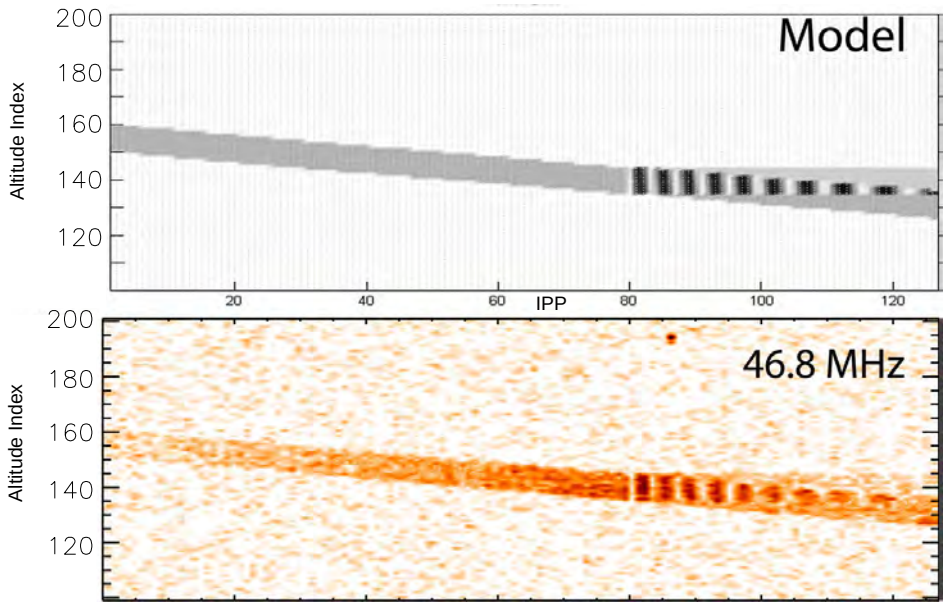


Figure 6: Demonstration of the radio science model showing the model result along side the emulated data from Figure 3

The modeling of the events is broken up into a number of functional blocks within the modeling program. The first feature reflects the movement and modeling of the meteor or plasma blob as radar point targets. Each target (meteoroid head-echo) is represented as a radar point scatterer as was done by Roy *et al.*[7] These target objects contain the physical properties each target, such as a "reflectivity" constant that is representative of variations in the radar backscatter cross-section of the target, velocity of the target along the modeled dimension, acceleration of the target and the target's initial altitude. Additionally, the object may be coded to start and end at arbitrary times in order to compensate for meteor events which would result in the elimination or creation of targets. These objects use the following motion equation based on Roy *et al.*[7] to calculate the the target's range (or altitude):

$$R(t_M) = R_I - v_I * (t_M - t_I) - \frac{a}{2}(t_M - t_I)^2 \tag{2}$$

where $R(t_M)$ is the range of the object at model time t_M, R_I is the initial range of the object when t_M is equal to the starting time of the object t_I, v_I is the initial object velocity of the object (also at time t_I) and a is the acceleration of the object.

The model iterates through individual radar pulses and after each pulse the model discretely increments through time. During each time step, a function representing the ranging functionality of the radar calculates whether or not the emitted

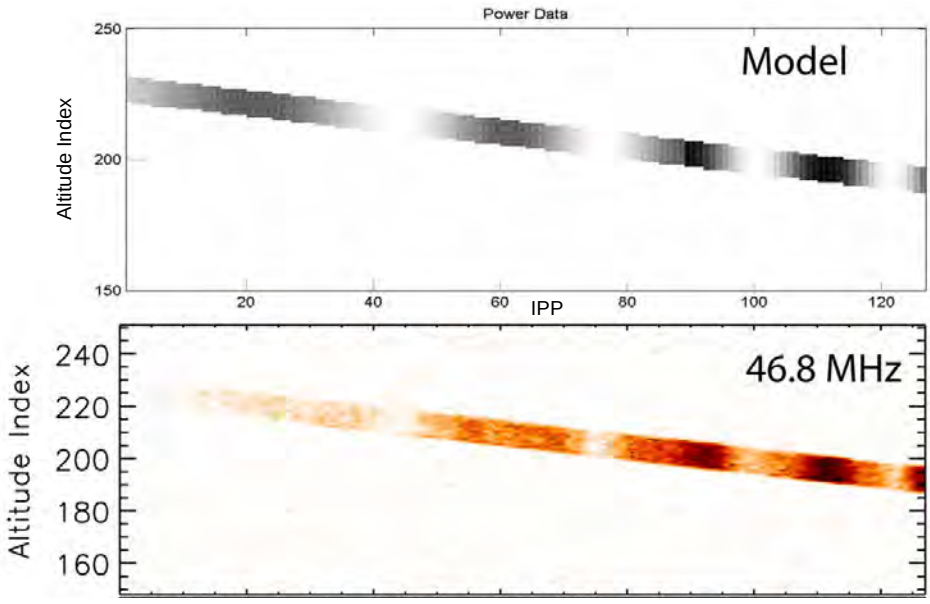


Figure 7: Demonstration of the radio science model, showing the model results along with the emulated data. The emulated data possesses a meteor science phenomena (occurring around pulse 90) which is not accurately explained using the simple radio science model alone.

pulse would have had time to reach each target object and return based on the speed of the signal's propagation and the target's altitude. For echo to have time to return, the model time must fit within the following range:

$$2R_t/C_{radar} < t_M < 2R_t/C_{radar} + t_p \tag{3}$$

where R_t is the range of the target calculated in (2), C_{radar} is the speed of the pulse propagation, t_p is the length of the pulse signal and t_M is the model time.

If there has been adequate time, a function then determines what the baseband signal phase would be given the target's current position (altitude). The complex baseband signal for the target as used in Roy *et al.*[7] is represented as:

$$x_i(t_M) = A_i \exp\left(\frac{i4\pi R(t_M)}{\lambda} t_M\right) \tag{4}$$

where x_i is the baseband signal, A_i is the complex amplitude of the signal, $R(t_M)$ is the range of the target at model time t_M and λ is the radar wavelength.

By accounting for the movement (and resulting decrease in altitude) of the object during the pulse, we are able to automatically account for Doppler shifting of the signal. Adding the signal echo from each object at each discrete delay time permits us to accommodate multiple fragments and allows us to model the phenomena of signal interference from multiple fragments. We then created emulations of

simple event specimens by calculating approximate initial altitudes, velocities, differences in fragment velocities and accelerations based on the recorded data of the specimen. The model parameters were then fine tuned by hand, utilizing trial and error and a developed intuition for correcting the approximations. Two modeled results which were emulated in this way are shown in Figures 6 and 7.

Compounded Interference Observations

By considering combinations of the modeled radio science interference patterns from two meteoroid head-echoes, the majority of events analyzed here are likely emulated using two or more scatterers in the manner of Figures 6 and 7. Figure 8 shows some examples displaying both clear pulse-to-pulse cycling along with well defined intrapulse interference.

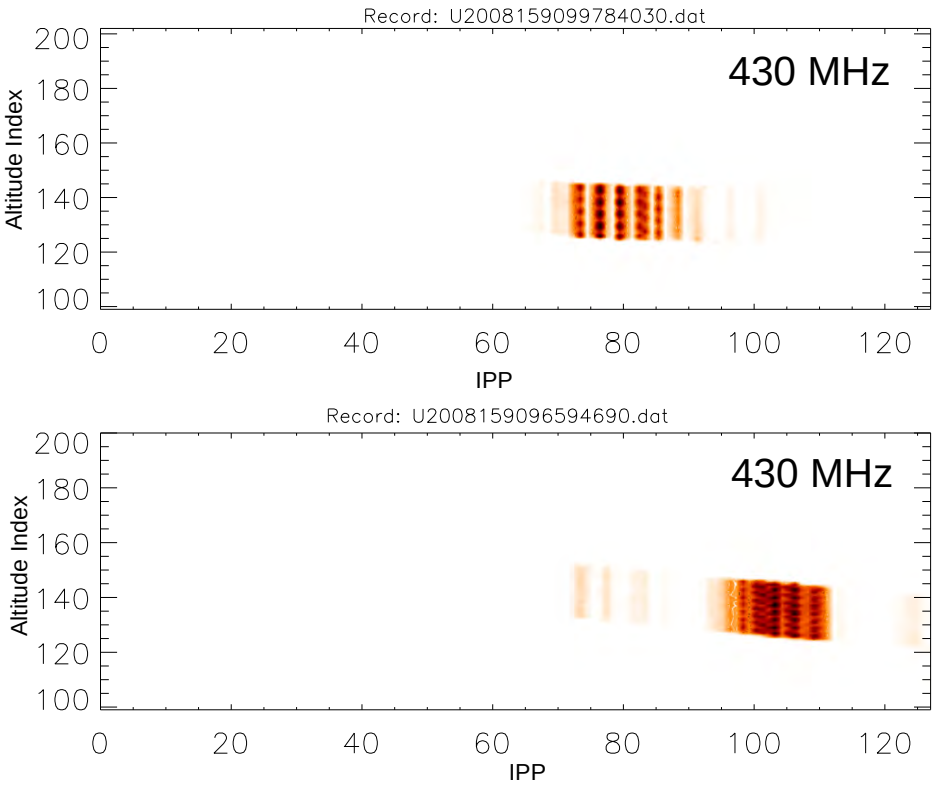


Figure 8: Two UHF RTI plots demonstrating the combination of a small number of fragments, greater than two, which created an interference pattern with clear pulse-to-pulse cycles in addition to intrapulse interference structures.

The intricate relationships between the interference created between each additional scatterer beyond the first interfering pair make the accurate emulation of such events nearly impossible to do by hand and brute force calculations except

for exceptionally simple events. More advanced processing techniques for finding the correct number of scatters and their properties, as shown in Roy *et al.*[7], while requiring greater computational power, would help affirm that these events are actually explained by fragmentation. The advanced processing techniques may even be helpful in determining the number and properties of scatterers of even more complicated events such as the nearly random appearing interference shown in Figure 9.

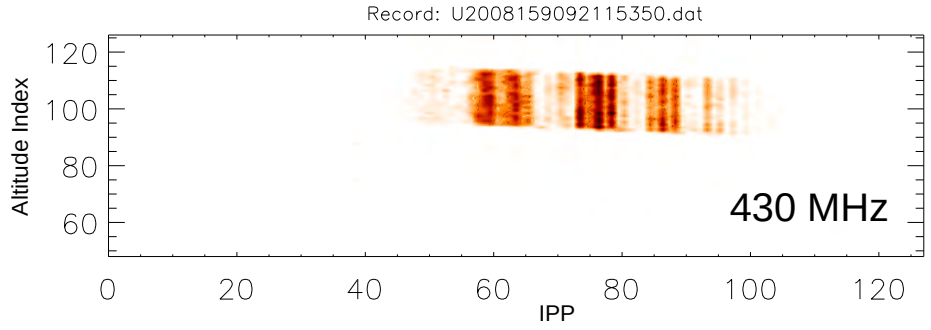


Figure 9: A UHF RTI plot demostrating the combination of a number of fragments of different speeds resulting irregular interference from numerous pulse-to-pulse cycling rates and intrapulse interference.

The event in Figure 10 at first appears to contain no interference patterns, regular or irregular. Closer inspection of that event's RTI plot (and those of many similar events) reveals that the near random noise within single pulses and from pulse-to-pulse is much greater than the background noise, suggesting that these events too are undergoing some amount of fragmentation. It may also be tempting to make the assumption that these radio science explanations are merely coincidentally flexible enough to explain the widely variable types of event specimens that were observed and to speculate that the interference cycling seen is instead the result of some combination of background noise, equipment error or other factors, such as ablation. While meteor science factors definitely play an important role in almost all of the observed events, a number of simultaneous, non-overlapping events suggest that the interference patterns cannot be attributed to radar system inaccuracies alone. Notice that in Figure 11, the lower event possesses noticeable pulse-to-pulse beating while the higher event possesses intrapulse interference along with some reflection of the UHF antenna gain.

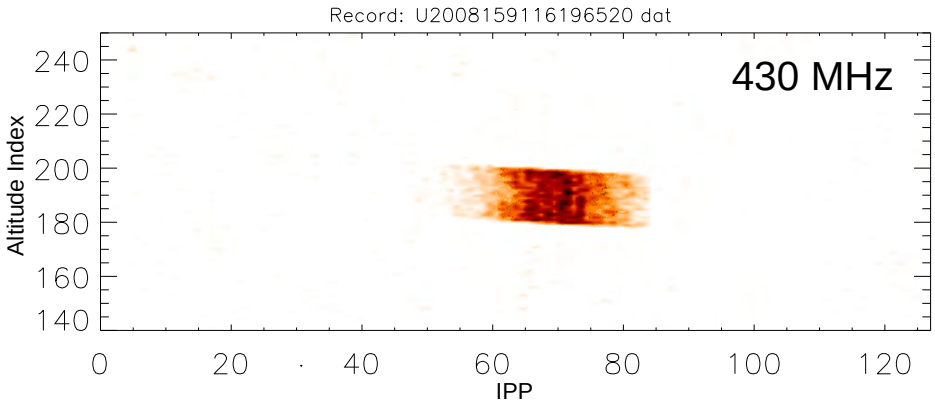


Figure 10: A UHF RTI plot, representative of a majority of meteor events analyzed, at first appears to be the smooth result of a single fragment passing through the main antenna lobe. Close inspection reveals intrapulse and pulse-to-pulse variations that have much greater variations than that seem due to background noise and is further inconsistent with the expected antenna gain. These variations could possibly be the result of interference created by the reflections of numerous small fragments around the larger scatterer that is dominant.

CONCLUSIONS

We have analyzed meteor events from Arecibo Observatory and found strong indications of signal interference and meteor fragmentation presence in nearly 100% of the observed events. We have successfully emulated some of these interference patterns using very simplified models of the involved radio science and helped to further verify the importance of meteor fragmentation.

Our demonstrated model may serve as a starting foundation for further simulations of micrometeors which include these radio signal interference properties. Further work could be incorporated into a 3D model while accounting for the antenna gain geometry. However, for now, the Arecibo Observatory does not possess the interferometric capacities that would contribute very important spatial information to complete these models accurately. Meteor science events, like ablation, as modeled by Dyrud and Janches[1] could also be integrated into a more robust model.

Additionally further investigation is warranted to explain a number of interesting meteor specimens, such as the possible existence of UHF trail-echoes, which have been highlighted by our classification of the Arecibo Observatory data.

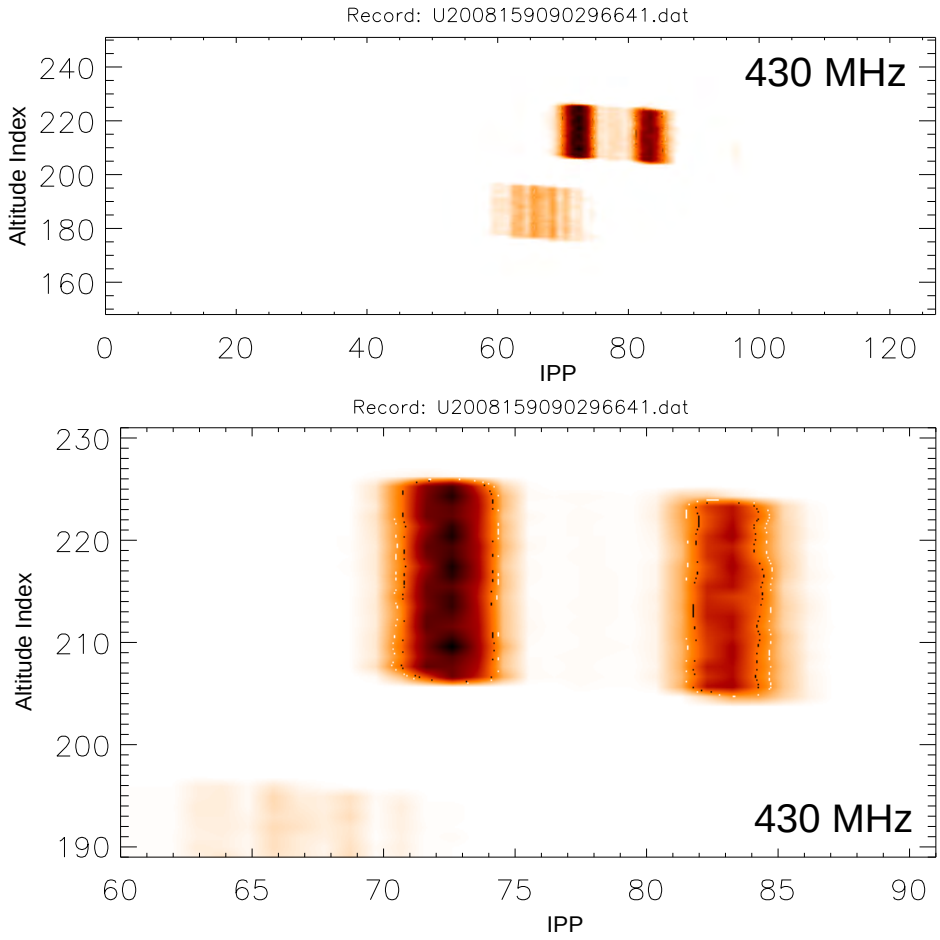


Figure 11: A UHF RTI plot that shows two simultaneous head-echoes, one possessing pulse-to-pulse cycling interference and one with intrapulse interference suggesting that these different interference patterns are not the product of equipment error alone.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with the National Science Foundation.

REFERENCES

[1] L. Dyrud and D. Janches, "Modeling the meteor head echo using Arecibo radar observations," *Journal of Atmospheric and Solar-Terrestrial Physics*, **70** 1621–1632 (2008).

[2] M. I. Skolnik, *Radar Handbook*, 3 ed., McGraw-Hill Professional, New York, 2008.

[3] J. D. Mathews, S. Briczinski, D. D. Meisel and C. Heinselman, "Radio and Meteor Science Outcomes From Comparisons of Meteor Radar Observations at AMISR Poker Flat, Sondrestrom and Arecibo," *Earth Moon Planet*, **102** 365–372 (2008).

[4] J. D. Mathews, "Radio science issues surrounding HF/VHF/UHF radar meteor studies," *Journal of Atmospheric and Solar-Terrestial Physics*, **66** 285–299 (2004).

[5] J. Mathews, J. Doherty, C.-H. Wen, S. Briczinski, D. Janches and D. Meisel., "An update on UHF radar meteor observations and associated signal processing techniques at Arecibo Observatory," *Journal of Atmospheric and Solar-Terrestial Physics*, **65** 1139–1149 (2003).

[6] S. J. Briczinski, J. D. Mathews and D. D. Meisel, "Statistical and fragmentation properties of the micrometeoroid flux observed at Arecibo," *Journal of Atmospheric and Solar-Terrestial Physics*, **114** A04311 (2009).

[7] A. Roy, S. J. Briczinski, J. F. Doherty and J. D. Mathews, "Genetic-Algorithm-Based Parameter Estimation Technique for Fragmenting Radar Meteor Head Echoes," *IEEE Geoscience and Remote Sensing Letters*, **6** (3) 363–367 (2009).

DEPLOYABLE STRUCTURES BASED ON SHAPE MEMORY ALLOYS

Jennifer K. Felder* and Sven G. Bilén[#]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate Student of
Department of Electrical & Computer Engineering
North Carolina State University
Raleigh, NC 27695

ABSTRACT

This paper explores shape memory alloy (SMA) materials, their unique properties and applications, and how they can be controlled using the pulse-width modulation (PWM) technique. Specifically, its focus is on the design of a circuit to generate a PWM signal to control on a small satellite the deployment of a boom that is made of Nitinol, an SMA material. SMA materials are distinctive in that they undergo a solid-state phase change that can be caused by either a current or an increase in loading being applied to them. A Nitinol wire was tested using a power supply and function generator: various duty cycles were tested with input voltages of 3.3 V and 5.5 V and current-draw and deployment time data were recorded for each trial. From this data, an optimal duty cycle and voltage combination was selected to minimize power usage, but still deploy the satellite boom in a reasonable amount of time. Various circuit designs to transmit the PWM signal using the duty cycle and voltage previously selected were then explored, and one using an LPC2468 microcontroller was selected, designed and tested.

INTRODUCTION

Cubesats

This project focuses on the shape memory alloy (SMA) Nitinol wire boom of the OSIRIS cubesat, currently under development in the Student Space Programs Lab at Penn State. Cubesats have a cubic shape, measuring 10 cm on each side, with a mass of no more than 1 kg [1]. This size is defined by the launch system for these satellites, the Poly Picosat Orbital Deployer (P-POD). Because of these size restrictions, deployable structures are imperative on such satellites.

[#] Faculty Mentor

Deployable structures are also important on cubesats because they allow sensors like magnetometers to be deployed far enough away from the satellite such that magnetic interference from the satellite itself is minimal [2]. A novel way to implement deployable structures on satellites is using SMA materials.

Shape Memory Alloys

Two important properties characterize shape memory alloys: the shape memory effect and pseudo-elasticity. Although these properties were discovered in 1938 by Arne Olander, serious SMA research began in the 1960s [4]. CuZnAl, NiTi (Nitinol), and CuAlNi are some of the most commonly used SMA materials. Nitinol was chosen for this project because, when compared with two other popular SMA materials (CuZnAl and CuAlNi), it had the highest output volumetric work, broadest range of transformation temperatures, lowest input power per kg of mass, and highest efficiency (as measured by output work per mass). While it has the highest material cost compared to the others, it has a lower operating cost, as its resistivity is higher (so lower voltage is needed to heat it). Therefore, overall it was determined that Nitinol was the best SMA material to use for this application [3].

The rare properties found in SMA are due to a solid-state phase change that occurs within them, during which molecules rearrange; however, the SMA always takes the form of a solid. Most SMAs need only to have their temperature changed by about 10 °C in order for this phase change to occur. SMAs have two solid states, known as the Martensite and Austenite phases. In the Martensite phase, the lower temperature phase of the two, SMAs are more easily deformed and have a twinned molecular shape. Conversely, in the Austenite phase, the stronger, higher-temperature of the two phases, molecules adopt a cubic shape. The Martensite phase, when not deformed, appears macroscopically identical to the Austenite phase [4].

For the shape memory effect to occur, an SMA material is first cooled below a characteristic temperature, such that it is in its Martensite phase. It can then be deformed into virtually any shape. The material then recovers a preset shape when it is heated above another characteristic temperature, after which it is in its Austenite phase. This effect occurs because the heat transferred through the wire causes the molecules within the wire to rearrange [4]. This shape memory effect of SMA materials can take one of two forms, based on how they are trained: one-way or two-way. When a one-way memory effect occurs, an SMA is deformed at low temperature, and when heated returns to its preset shape. Conversely, with the two-way memory effect, it has one shape to which it reverts when it is cooled and a different shape to which it returns when it is heated [5].

The other unique characteristic of SMAs is their pseudo-elasticity. For this to be observed, the SMA must first be completely in its Austenite phase. Pseudo-elasticity is similar to the shape memory effect, with the exception that it is caused by an increase in loading, and not a temperature change. When the loading applied to the material is sufficient, the SMA changes from its Austenite

phase to its Martensite phase, as the Martensite form of the material absorbs the loading. When the loading is decreased, the SMA reverts to is Austenite phase, and returns to its original shape [4].

SMA materials have been implemented in many applications: deploying the wings of an airplane, replacing hydraulic systems, creating bone plates and robotic limbs, simulating muscle movement realistically, and being employed in everyday appliances like coffeepots, thermostats, and cell phone antennas [4]. Early concerns with SMA materials being used in construction resulted from the fact that these materials change their geometric structure once deployed, which is difficult to account for with precision in designs. Because of such problems, SMA materials are not used in construction as much as they are used to increase the functionality of other materials. An example of SMA materials used in conjunction with other materials occurred on the container of the Nimbus satellite: SMA wires were used to open the flaps of the satellite by contracting when their temperature increased and closing them when they cooled. Another possible application of SMAs is using them to drive long trusses in open space, because they have such a high damping ability. They also are replacing pyrotechnical materials in deployment devices, because the device can be activated multiple times with negligible damage to the SMA and they are much safer than pyrotechnical materials, allowing workers to work safely in close proximity to their projects. Shock-load that occurs when such devices are released is also not a concern when using SMA materials. Future goals include designing SMA devices powered by the Sun's heat and using them as construction, eliminating the need for other materials in devices [6].

There are several benefits of using SMA materials, as opposed to traditional materials: they can be used in biological tissue without harming it, they have a wide range of applications, and they have many useful mechanical properties, among them being corrosion resistance and great strength [4]. They are also used specifically in deployable structures because they have a high efficiency, are safe to use, have a high specific working capacity, are easy to control, and do not require shock suppressors. They are also especially useful in zero-gravity environments, making them appealing in aerospace applications [6]. However, they are still relatively expensive, and some fatigue more quickly than other materials, so they are not being as widely used as they could be [4].

Pulse Width Modulation

The circuit that deploys the SMA boom employs a pulse width modulation (PWM) technique. This technique is used to digitally control an analog signal, turning a DC supply fully on or completely off at any time. By altering the duty cycle, or the percentage of the period the DC supply is on, PWM encodes various analog signal levels. This significantly lowers both power consumption and cost. PWM has other benefits as well, when compared to analog control. As the signal is always digital, there is no need for any digital-to-analog conversion and noise interference is minimal. A PWM signal is also easy to implement and occupies

relatively little circuit area [7]. Furthermore, using PWM with SMA materials heats the materials more evenly and prevents hot spots, which avoids overheating and damage to the material. Also, as the amount the SMA contracts when heated (or expands when cooled) is proportional to the duty cycle, PWM allows more accurate and easier control than other techniques [8].

This project uses PWM to control the SMA boom deployment for the OSIRIS satellite. Specifically, it focuses on the circuit that controls the deployment of this boom, which was designed by choosing a duty cycle/voltage combination that minimizes both power usage and deployment time. In contrast to applying a constant voltage to deploy the boom, using PWM decreases power usage; however, it does not necessarily decrease deployment time.

EXPERIMENT DESCRIPTION

Preliminary Testing

Nitinol deployment testing was conducted in two parts. First, preliminary PWM testing of the Nitinol wire was performed. To do this, a Nitinol wire 0.772 m in length was formed into a loop and clamped between two insulating blocks, as shown in Figure 1. The wire was then connected, through a transistor (IRF 540), to both a power supply (Agilent E3631A Triple Output DC Power Supply) and a function generator (Agilent 33220A 20 MHz Function/Arbitrary Waveform Generator). However, the function generator used in the preliminary testing initially replaced the microchip (LPC2468) shown in the schematic in Figure 2. To simulate an orbital environment, in which gravity would not have an effect, the wire was deployed horizontally across a low-friction surface, as in Figure 3. While the wire did not have to overcome the force of gravity, knowing the maximum mass it could support and still deploy successfully is important, as even in outer space the boom must still overcome inertia. To find this maximum mass, washers were successively added to the wire, until it could no longer deploy. It was found that the wire could have a maximum of 13 g attached to it and still deploy successfully.

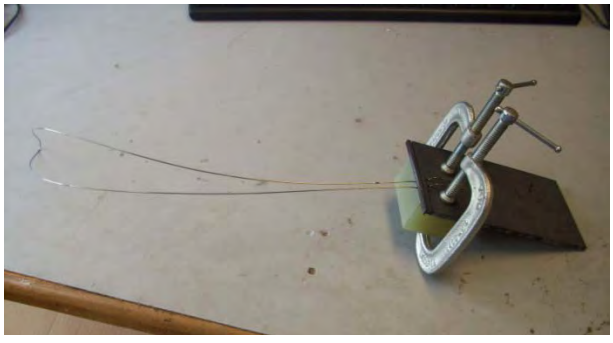


Figure 1: Nitinol Wire Setup

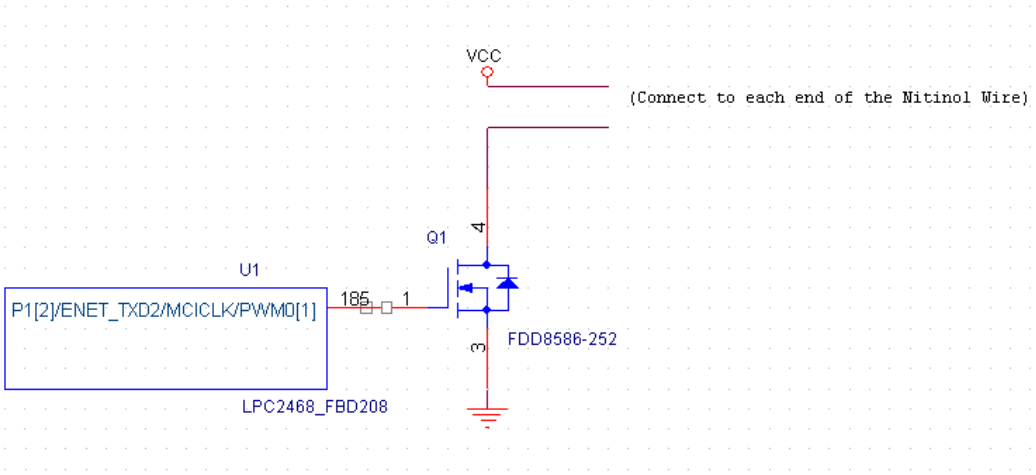


Figure 2: PWM Circuit Schematic

To obtain the initial data, the voltage limit of the power supply was set to 3.3 V and the current limit to 3 A. Because the voltage was constant, the current draw depended only on the duty cycle. The 'pulse' function on the function generator was then selected and the duty cycle set to 5%. It was also determined that the period is arbitrary, and does not affect the power usage or deployment time for any given duty cycle. In this case, a period of 50 ms was used. The function generator, power supply, and a timing device were then turned on simultaneously, and the deployment time—the time it took the wire to return from a curled position (Figure 3) to a straight position (Figure 4)—was recorded. The current draw was also recorded from the DC power supply. This was repeated twice more, for a total of three trials.

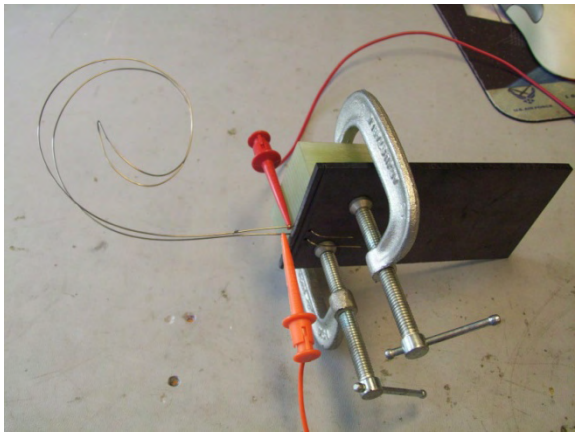


Figure 3: Wire Initial (Curled) Position

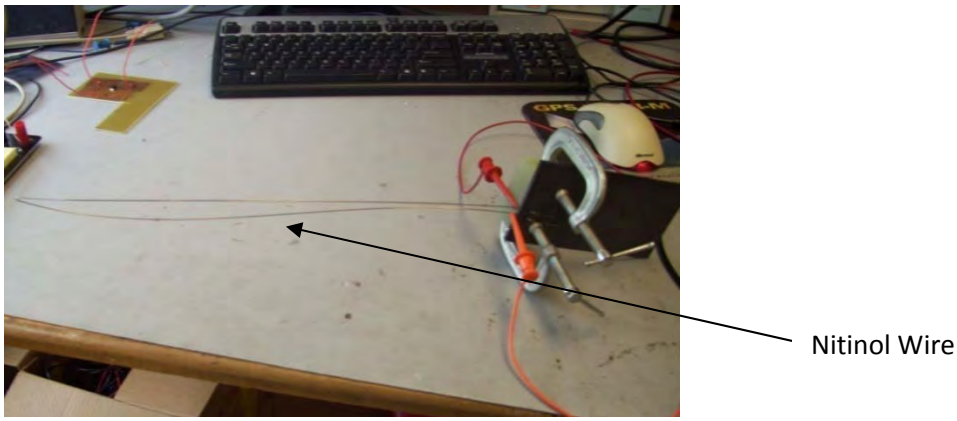


Figure 4: Wire Final (Straight) Position

The above procedure was then repeated, increasing the duty cycle by 5% every three trials, until a duty cycle of 95% was reached. As a control, data were also collected without using the function generator, thereby obtaining data for a 100% duty cycle. Then, the same data were collected using the above procedure, except limiting the voltage emitted by the power supply to 5.5 V. Only these two voltages were tested, as they are the two available on the satellite on which the circuit and boom would be implemented. The 3.3-V data collected began at a 45% duty cycle, because that was the minimum duty cycle that was able to deploy the wire. The 5.5-V data began at a 25% duty cycle, because that was the lowest duty cycle that deployed the wire in a reasonable amount of time.

Finally, from this data, power and power usage (in watthours) were calculated using the formulas

$$P = I^2R \tag{1}$$

$$\text{Power Usage} = Pt \tag{2}$$

where P is power, I is current, R is resistance, and t is deployment time in hours. This determined the optimal voltage and duty cycle combination that minimizes both power usage and deployment time.

The PWM Circuit

The second part of the experiment involved implementing a circuit to control the PWM signal that is used to deploy the boom of the satellite, instead of connecting the Nitinol wire to a function generator. Several methods were explored. The first was circuits in which the values of resistors or capacitors in the circuit are changed to determine various duty cycles. One such circuit involved connecting a 555 timer (LM555) to a power supply, two resistors, two capacitors, and a diode. In this method, the duty cycle is calculated from the values of two resistors, R_1 and R_2 [9]. For duty cycles, D, greater than 50%

$$D = (R_1 + R_2)/(R_1 + 2R_2), \quad (3)$$

whereas, for duty cycles less than 50%

$$D = R_1/(R_1 + R_2). \quad (4)$$

The 74AC14 Hex Inverter with Schmitt-Trigger Inputs uses a similar technique to change the duty cycle. Once the 74AC14 is connected to a power supply, two capacitors, two diodes, and a variable resistor, the output charges through two variable resistor paths; altering the potentiometer changes the duty cycle [10]. Similarly, the CD4093 can be used for PWM once connected with a 220-kΩ potentiometer, a capacitor, two diodes and a 2.2-kΩ resistor; the duty cycle is selected by changing the value of the capacitor [11].

Another method of controlling PWM implements circuits that use programmable chips. One such chip is a PIC (12F675 Servo controller). To do so, it should be connected to a capacitor, three resistors, and a serial interface. The duty cycle is altered directly in the program, which is transferred to the PIC using the serial interface [12].

Another means to control a PWM signal is through circuits whose duty cycle is dependent upon the input voltage, such as the HTC 852 chip connected to five 10-kΩ resistors. The duty cycle is altered by varying the input voltage, and can be calculated using the fact that the frequency of the output signal is fixed at 31250 Hz [13]. The TL1454A also controls PWM, outputting different duty cycles that are determined by the input voltage, as calculated by

$$V_{DC} = V_{O(MAX)} - D(V_{O(MAX)} - V_{O(MIN)}), \quad (5)$$

where V_{DC} is the input voltage needed to limit the duty cycle and $V_{O(MAX)}$ and $V_{O(MIN)}$ are voltage values specified on the data sheet). To be used in this way, it must first be connected to two resistors and a capacitor [14].

The final PWM control circuit explored was a multivibrator circuit, which uses the LM339. When connected to a capacitor and nine resistors of various specified values, the PWM duty cycle is determined by changing the value of a specific resistor (R_4 on the datasheet schematic) [15].

Because the OSIRIS satellite already contains a microcontroller (LPC 2468), a circuit that uses the microcontroller to generate the PWM signal was chosen. The microcontroller was then programmed to control the power source using PWM such that the duty cycle could be altered simply by typing the desired duty cycle directly in the source code. To test the program and microcontroller, the Nitinol wire was connected through a transistor (FDD8586 N-channel MOSFET) to the power supply and microcontroller, as in Figure 2. Various duty cycles were then tested and verified with an oscilloscope to confirm that the circuit functioned correctly.

EXPERIMENT RESULTS

The data collected in the first part of the experiment, testing the Nitinol wire using a power supply and function generator, are displayed is Table 1. The voltage data are constant for the two voltages tested, as that is limited by the voltage limit of the power supply on the cubesat and is predetermined. The current draw data were read directly from the power supply. The duration (deployment time) was obtained using a stopwatch. Using that data collected directly, power and power usage were calculated as in Equations 1 and 2. The specific method of calculating power (Equation 1) was chosen because both the resistance of the wire and voltage applied to the wire are constant, while the current passing through the wire varies with duty cycle. Therefore, power usage depends on the duty cycle.

Table 1: Nitinol Wire PWM Data

Voltage (V)	Current Draw (A)				Duty Cycle	Duration (s)				Power (W)	Power Usage (Wh)
	Trial 1	Trial 2	Trial 3	Average		Trial 1	Trial 2	Trial 3	Average		
3.3	0.851	0.945	0.896	0.897	45.00%	63	66	64	64	3.38	0.0271
3.3	1.041	1.024	1.03	1.032	50.00%	45	42	45	44	4.47	0.0272
3.3	1.103	1.084	1.121	1.103	55.00%	35	38	41	38	5.11	0.0296
3.3	1.24	1.284	1.241	1.255	60.00%	37	32	34	35	6.62	0.0381
3.3	1.337	1.315	1.326	1.326	65.00%	26	27	28	27	7.38	0.0363
3.3	1.428	1.429	1.437	1.431	70.00%	24	25	23	24	8.60	0.0399
3.3	1.526	1.556	1.547	1.543	75.00%	20	20	20	20	10.0	0.0414
3.3	1.692	1.694	1.69	1.692	80.00%	18	16	17	17	12.0	0.0451
3.3	1.813	1.794	1.762	1.790	85.00%	15	15	16	15	13.5	0.0479
3.3	1.914	1.901	1.926	1.914	90.00%	14	13	14	14	15.4	0.0530
3.3	2.054	2.065	2.037	2.052	95.00%	12	12	12	12	17.7	0.0570
3.3	2.099	2.104	2.098	2.100	100.00%	13	13	13	13	18.5	0.0677
5.5	0.846	0.945	0.875	0.889	25.00%	25	23	25	24	3.32	0.00561
5.5	1.017	1.051	1.033	1.034	30.00%	16	16	17	16	4.49	0.00607
5.5	1.253	1.181	1.208	1.214	35.00%	13	15	13	14	6.19	0.00834
5.5	1.352	1.369	1.356	1.359	40.00%	11	12	11	11	7.76	0.00988
5.5	1.625	1.614	1.604	1.614	45.00%	8	8	8	8	10.9	0.0114
5.5	1.812	1.863	1.792	1.822	50.00%	8	8	8	8	13.9	0.0149
5.5	1.987	1.97	2.007	1.988	55.00%	7	7	7	7	16.6	0.0176
5.5	1.845	1.907	1.932	1.895	60.00%	7	7	7	7	15.1	0.0167
5.5	2.068	2.052	2.085	2.068	65.00%	6	6	6	6	18.0	0.0203
5.5	2.23	2.245	2.223	2.233	70.00%	6	6	6	6	20.9	0.0234
5.5	2.653	2.371	2.377	2.467	75.00%	5	6	5	6	25.6	0.0300
5.5	2.514	2.51	2.518	2.514	80.00%	5	5	5	5	26.5	0.0303
5.5	2.622	2.631	2.657	2.637	85.00%	5	5	4	5	29.2	0.0314
5.5	2.786	2.797	2.799	2.794	90.00%	4	5	4	4	32.8	0.0354
5.5	2.935	2.937	2.935	2.936	95.00%	3	4	4	4	36.2	0.0345
5.5	3.497	3.497	3.497	3.497	100.00%	5	4	5	4	51.4	0.0608

Data regarding the properties of the Nitinol wire itself were also collected. Cross-sectional area was found using the diameter of the wire, which was measured with a caliper. The resistance of the wire was also measured directly

using a multimeter. The length of the wire was measured with a tape measure. Finally, the resistivity of the wire was calculated using

$$\rho = RA/L, \tag{6}$$

where R is the resistance of the wire, A is the cross-sectional area of the wire, and L is the length of the wire.

Table 2: Nitinol Wire Data

Resistivity (Ω*m):	7E-07		Nitinol Length (m):	0.772
Cross-sectional area (m^2):	1.3E-07		Resistance (Ω):	4.2

From this data, the trend was observed that, for each voltage tested, both power and power usage vary directly with duty cycle. Therefore, the minimum duty cycle that effectively deploys the wire in a reasonable amount of time should be chosen. Also, in terms of choosing between the two input voltages tested (3.3 V and 5.5 V), it was observed that in each comparable duty cycle, 5.5 V used less power; this is because using 5.5 V to deploy the boom drastically decreases deployment time.

DISCUSSION

The hypothesis originally presented in this work was that pulse width modulation decreases both power usage and deployment time, as compared to deploying the boom by applying a constant voltage to it. The data collected in the preliminary setup, before the actual PWM circuit was designed and tested, supports the power-usage part of this hypothesis. In the 3.3-V data, the 100% duty cycle trial used 0.01 Wh more power than the PWM trial that used the most power. Similarly, in the 5.5-V data, the 100% duty cycle trial used approximately twice as much power as the PWM trial that used the most power. The tradeoff for this decrease power is that the 100% duty cycle trial deploys the wire in less time than most of the PWM trials. However, for this application time is not a constraint, while power is.

A few further experiments could be conducted strengthen the claim that the method of deployment chosen is actually the most efficient one in terms of power usage. First, Nitinol wires of various diameters that were still able to support the sensor that will ultimately be attached to the end of the wire could be tested to see if a wire of larger or smaller diameter, which would therefore have lower and higher resistances, respectively, would use less power and maintain an acceptable deployment time. Wires composed of different SMA materials could also be tested using the same procedure to see how their power usages compared with that of Nitinol. These tests could verify that the material and size wire selected are actually the most efficient.

CONCLUSION

From the tests performed on the Nitinol wire, it can be concluded that pulse width modulation is effective in reducing power usage, at the tradeoff that deployment time may not be decreased or may remain unchanged. From the data collected, a duty cycle of 25% with an input voltage of 5.5 V was selected to implement on the PWM circuit controlling the boom on the actual satellite. This combination was selected because it used the least power out of all the voltage/duty cycle combinations tested.

When designing the PWM circuit to control the satellite boom, while several options of producing a PWM signal were explored, ultimately an LPC2468 microcontroller was selected, in conjunction with an N-channel MOSFET, because there was already a microcontroller on the satellite that could be used. It terms of limited space, size, and mass on the satellite and in terms of cost, this made the most practical sense. Since duty cycle and voltage combinations were previously tested, the microcontroller was programmed to implement the 25% duty cycle with 5.5-V combination that was already decided upon in earlier testing. The circuit was tested ensure that it functioned correctly, but it was assumed from prior data that that combination would be optimal.

This research has many applications and commercialization opportunities. One area of application in which SMA structures have particular interest is cubesats. Because these satellites use a common design standard and launching system, costs are minimized, making outer space more accessible. However, as a result of this launching system, they also have relatively small size restrictions, making deployable structures imperative [2]. SMA materials are particularly useful in these structures, as they are simple to deploy and use very little power.

Outside of satellite applications, Nitinol and other SMA materials are being employed in other areas as well, including medicine, military planes, robotics, and household appliances. In the medical industry, SMA materials are being implemented in tweezers and catheter guides that are able to remove objects from the body using minimally sized incisions and eyeglasses frames that are more easily repaired with little heat when bent. They are also being incorporated in orthodontic wires that do not have to be tightened or adjusted as much as other wires and that work more quickly because of the constant pressure they apply as they return to their original shape [16]. Such materials are also used to construct bone plates: they allow the bones to heal more easily and more quickly, as they apply constant pressure to the break [4]. In military planes, SMA materials are used in F-14 fighter planes to link hydraulic lines tightly and with greater ease. SMA materials are also being used in devices that prevent scalding in shower heads and water faucets (the devices turns off when the SMA reaches a certain temperature) and fire sprinklers that respond more quickly. Finally, they are used in robotics for their smooth motion, to simulate muscle movement more realistically; this may ultimately lead to improvement in prosthetics as well [16].

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081. Furthermore, I would first like to thank Dr. Sven Bilén for giving me the opportunity to do this research, and for his help and support. Thanks also to David Zhang for programming the microcontroller used in the PWM circuit and to Brian Schratz for helping me with the program OrCAD, which was used to create my schematic. Finally, I would also like to thank Matthew Becker and Paul Pribula for helping me get accustomed to the lab and helping me numerous times throughout my research.

REFERENCES

[1] Cubesat Community Website. 2 October 2008. California Polytechnic State University. 5 June 2009. <http://cubesat.atl.calpoly.edu/media/CDS_rev11.pdf>.

[2] Thomsen, Michael. "FEATURE: Cubesats-How Small Can Satellites Get?". SatMagazine. January 2009. 26 June 2009. <http://www.satmagazine.com/cgi-bin/display_article.cgi?number=1556330976>.

[3] Huang, W. "On the Selection of Shape Memory Alloys for Actuators". Materials & Design. Vol. 23, Issue 1. 4 December 2001: 11-19. 11 June 2009. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TX5-44KKMNS-2&_user=209810&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000014439&_version=1&_urlVersion=0&_userid=209810&md5=728ee7230f11f2aade2f49728b81fa91>.

[4] Shape Memory Alloys. 17 August 2001. SMA/MEMS Research Group. 1 June 2009. <http://www.cs.ualberta.ca/~database/MEMS/sma_mems/sma.html>.

[5] Shape Memory Alloy. 2009. AbsoluteAstronomy.com. 1 June 2009. <http://www.absoluteastronomy.com/topics/Shape_memory_alloy>.

[6] Razov, A. I., and A. G. Cherniavsky. "Applications of shape memory alloys in space engineering: past and future". Space Mechanisms and Tribology, Proceedings of the 8th European Symposium. Vol. 438. 1999: 141-145. 8 June 2009. <http://articles.adsabs.harvard.edu//full/1999ESASP.438..141R/0000143.000.html>.

[7] Barr, Michael. Introduction to Pulse Width Modulation (PWM). 2 July 2003. O'Reilly. 1 June 2009. <http://www.oreillynet.com/pub/a/network/synd/2003/07/02/pwm.html>.

[8] Activating Nitinol w/PWM Page 7. 2007. Images Scientific Instruments. 8 June 2009. <http://www.imagesco.com/articles/nitinol/07.html>.

[9] LM55 Timer. July 2006. National Semiconductor. 26 June 2009. <http://www.national.com/ds/LM/LM555.pdf>.

[10] 74AC14 – 74ACT14 Hex Inverter with Schmitt Trigger Input. December 1999. Fairchild Semiconductor. 26 June 2009. <http://www.datasheetcatalog.org/datasheet/fairchild/74AC14.pdf>.

[11] CD4093BC Quad 2-Input NAND Schmitt Trigger. April 2002. Fairchild Semiconductor. 26 June 2009. <http://www.fairchildsemi.com/ds/CD/CD4093BC.pdf>.

[12] PIC12F629/675 Datasheet. 2007. Microchip Technology Inc. 26 June 2009. <http://ww1.microchip.com/downloads/en/DeviceDoc/41190E.pdf>.

[13] <http://www.hightechips.com/Download/htc852.pdf>.

[14] TL494 Pulse-Width Modulation Control Circuits. May 2002. Texas Instruments. 26 June 2009. <http://www.datasheetcatalog.org/datasheet/texasinstruments/tl494.pdf>.

[15] LM139/LM239/LM339/LM2901/LM3302 Low Power Low OffsetVoltage Quad Comparators. March 2004. National Semiconductor. 26 June 2009. <http://www.national.com/ds/LM/LM339.pdf>.

[16] Lin, Richard. Shape Memory Alloys and Their Applications. 22 February 2008. Stanford University. 1 June 2009. <http://www.stanford.edu/~richlin1/sma/sma.html>.

PEAK POWER CONTROLLER FOR SOLAR ARRAYS ON SMALL SPACECRAFT

Arnaldo Negron,[*] Matthew G. Anderson,† and Sven G. Bilén[#]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

[*]Undergraduate student of
Department of Electrical Engineering
377 Ponce de Leon Ave.
Hato Rey, PR 00918

ABSTRACT

As the power supplied by solar arrays depends upon the varying temperature, load conditions, and satellite attitude variations, it is necessary to control its operating point to draw the maximum power from the array. It is our goal to implement a control system that can extract the maximum output power (i.e., the "peak power") from the solar array under these various conditions.

INTRODUCTION

Most satellites derive their power from solar arrays, either directly or via batteries that are charged by the arrays. These batteries must be recharged when the satellite is in sunlight and provide enough power for the satellite while it is in eclipse. It is desirable to operate the solar array at its peak power point so that the maximum power can be delivered under varying temperature, load conditions, and satellite attitude variations [1]. Our goal is to implement a control system that can extract the maximum output power (i.e., "peak power") from the solar array under these various conditions. To accomplish this, a microcontroller implements a peak power tracker (PPT) algorithm. The advantages of using a microcontroller are its flexibility and ability to be shared with the command and control function of the satellite. The PPT is based on the principle that the current available from a solar panel depends on operating voltage, and that by varying the operating voltage you can maximize the power extracted from the array.

---

†Student Mentor
[#] Faculty Mentor

Solar Panel Fundamentals

What is intriguing about solar power is that all the electricity is generated from the material of which the solar panels are made and the energy from the Sun. The solar panels are fabricated primarily out of semiconductor material, with silicon being the most used. The benefit of using semiconductor material is being able to control its conductivity, whereas insulators and conductors generally cannot be altered. The electrons of the semiconductor material can be located in one of two different bands: the conduction band or the valence band. The valence band is initially full with all the electrons that the material contains. When the energy from sunlight, known as photons, strikes the electrons in the semiconductor, some of these electrons will acquire enough energy to leave the valence band and enter the conduction band. When this occurs, the electrons in the conduction band begin to move creating electricity. As soon as the electron leaves the valence band, a positively charged hole will remain in the location the electron departed. When this occurs, the valence band is no longer full and can also play a role in the current flow. This process basically describes how photovoltaic (PV) systems function. However, PV systems further enhance the rate at which the electrons are sent into the conduction band through the process of doping [2].

Characteristics of Solar Cells

An *I–V* characteristic of a solar cell, when neglecting the internal shunt resistance, traditionally is given by the following equation [1]:

$$I_o = I_q - I_{\text{sat}}\left\{\exp\left[\frac{q}{AkT}(V_o + I_o R_s)\right] - 1\right\}$$

where I_q is the light generated current, I_{sat} is the reverse saturation current, q is the electronic charge, A is a dimensionless factor, k is the Boltzmann constant, T is the temperature in Kelvin, R_s is the series resistance of the cell [3]. Figure 1 shows a typical *I–V* characteristic curve of a solar cell.

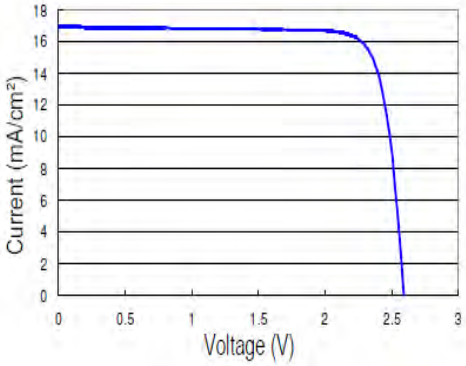


Figure 1. Typical *I–V* curve for a solar cell

Peak Power Tracker (PPT)

Figure 2 shows the circuit schematic of a PPT. A voltage sensor is connected in parallel to the array sending the data acquired to the LPC2378 microcontroller at ADC [0] from pin 13 of the OP467. The current sensor is connected in series through the shunt resistor and sends the data to the LPC2378 microcontroller at ADC [1] from pin 11 of the OP467. An inverting amplifier is connected to the gate of the P-type MOSFET and its input is controlled by the channel 1 output of the PWM located in the microcontroller at pin 107.

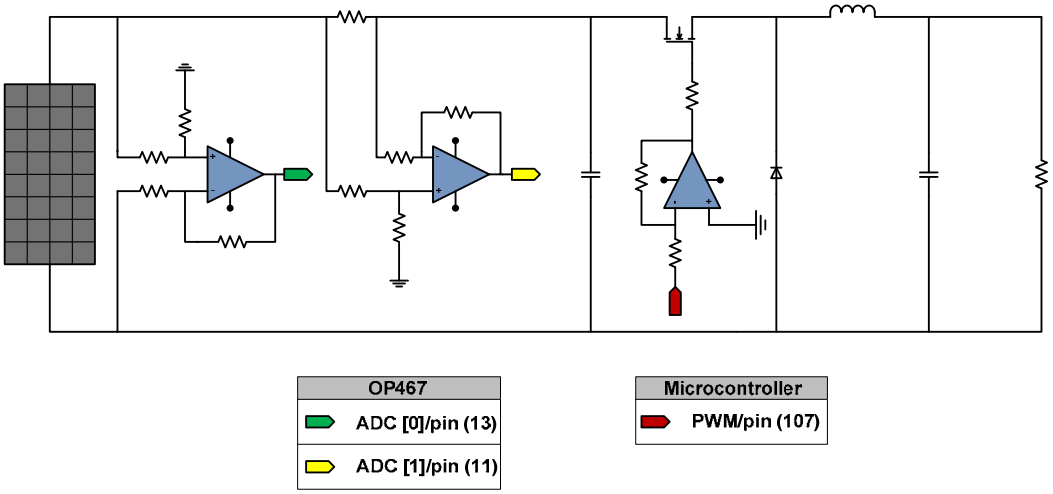


Figure 2. Peak power tracker schematic

EXPERIMENTAL DESCRITPION

Materials Used

The research project was performed in the Student Space Programs Laboratory at The Pennsylvania State University. The materials and equipment used in the research project were the following:

- (3) Solar Cell
  - ❖ Tecstar TRW Part No. 817715-3, size 31 cm², 140 μm(5.5 mil) thickness
- (1) 500-watt Halogen Lamp
  - ❖ Utilitech Part No. 0156853
- (1) Agilent Oscilloscope
  - ❖ Model No. 54622D
- (1) Fluke Multimeter
  - ❖ Model 87 III True RMS Multimeter
- (2) DC Voltage Supply
  - ❖ Hewlett Packard Model No. E3612A
  - ❖ Agilent Model No. E3615A
- (1) Breadboard/Wires/Connectors

Our solar array consists of three solar cells connected in parallel and located perpendicular to the 500-watt halogen lamp (which is used to simulate the Sun). The basic design for the test setup is shown below (see Figure 3). Among our hardware solution is a voltage and current sensor connected to the output of the array in order to provide the voltage and current readings from the solar array to the microcontroller. In our system we need an output voltage that is smaller than the input voltage; hence, the buck converter was used for this task. We also decided to implement the incremental conductance method for peak power tracking as this algorithm is more efficient than the perturbation and observation method, which are alternative methods. Because panel terminal voltage is changed according to its value relative to the maximum power point voltage, this method is independent of solar panel characteristics.

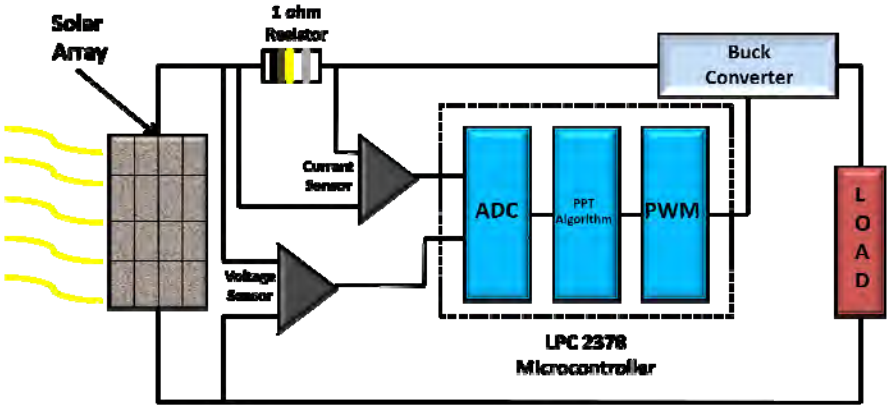


Figure 3. Hardware design

Hardware Design

To design our voltage sensor, current sensor, and buck converter, we first had to determine the values for the open circuit voltage and short circuit current of the solar array. The open-circuit voltage, V_{oc}, is the maximum voltage available from a solar cell, and this occurs at zero current. The open-circuit voltage corresponds to the amount of forward bias on the solar cell due to the bias of the solar cell junction with the light-generated current. The open-circuit voltage is shown on the *I–V* curve below in Figure 4. On the other hand, the short-circuit current is the current through the solar cell when the voltage across the solar cell is zero (i.e., when the solar cell is short circuited). Usually written as I_{sc}, the short-circuit current is shown on the *I–V* curve below in Figure. 5. The open-circuit voltage was 3.3 V and short-circuit current was 70 mA.

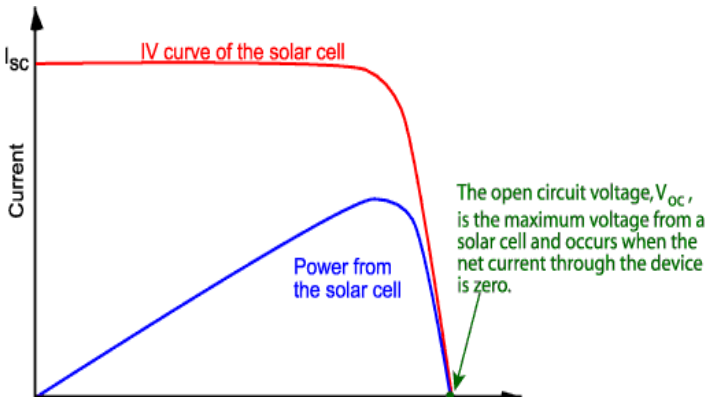


Figure 4. *I–V* curve of a solar cell showing the open-circuit voltage

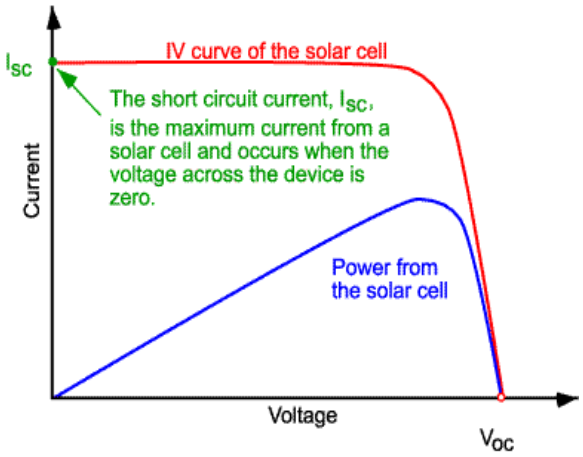


Figure 5. *I–V* curve of a solar cell showing the short-circuit current

Voltage Sensor

The purpose of the voltage sensor is to measure the voltage across the array for input into the algorithm as implemented, which varies the duty cycle as needed. For our voltage sensor we established our output voltage range to be from 0 to 3 V. Our gain was calculated with the following equation which is described as the open-loop gain of the op-amp:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{3\ \text{V}}{3.3\ \text{V}} = 0.909\ \text{V/V}$$

To measure the voltage supplied from the solar array, we designed our voltage sensor based on a differential op-amp circuit and used an OP461 quad, high speed, precision amplifier. This amplifier offers the performance of a high speed op-amp combined with the advantages of a precision op-amp in a single package.

$$V_{\mathrm{OUT}} = \frac{R_2}{R_1} \times V_{\mathrm{ARRAY}} = \frac{42.3\ \mathrm{k}}{46.14\ \mathrm{k}} \times V_{\mathrm{ARRAY}} = 0.917 \times V_{\mathrm{ARRAY}}$$

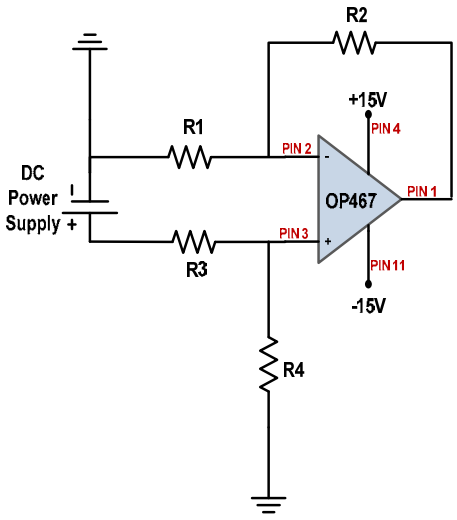


Figure 6. Voltage sensing circuit

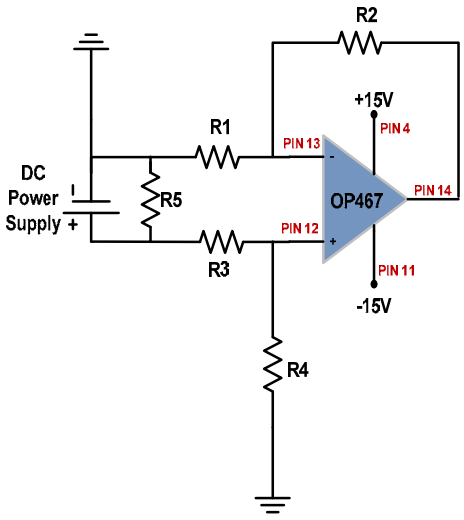


Figure 7. Current sensing circuit

Current Sensor

The purpose of the current sensor is to measure the current flowing through the shunt resistor for input to the algorithm. The sensor was designed by implementing a differential amplifier with a 1-ohm shunt resister located across the differential voltages (as shown above in Figure 7).

Buck Converter

The DC–DC buck converter uses energy storage components such as inductors and capacitors to control the energy flow from the solar array to the load by continuously opening and closing a switch (in this case it is a MOSFET P-type). This converter is suitable for use when the array voltage is high and the battery voltage is low. The power flow is controlled by adjusting the on/off duty cycle of the switching from the output of the PWM1 at pin 107 from the microcontroller. As a result, the converter (shown below in Figure 8) was implemented in our design.

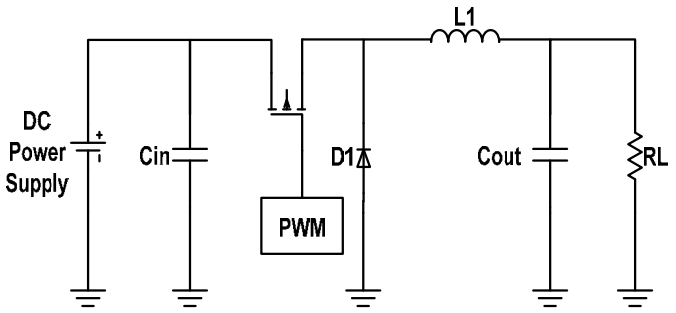


Figure 8. Buck converter schematic

Microcontroller

For the microcontroller we used the LPC2378 because of its use in other projects and familiarity in SSPL. The NXP (founded by Philips) LPC2378 is an ARM7TDMI-S based high-performance 32-bit RISC microcontroller with thumb extensions, 512 kB on-chip flash ROM with In-System Programming (ISP) and In-Application Programming (IAP), 58 kB RAM, CPU clock up to 72 MHz, on-chip crystal oscillator, on-chip 4 MHz RC oscillator, on-chip PLL [4]. This microcontroller includes PWM capabilities as a simple method for digital control logic to create an analog equivalence. Since the application of a PPT calls for a specific "on" and "off" pulse width to control the gate at the P-type MOSFET, the PWM provides this capability for the research project. In addition, with double edge controlled PWM output, specific match registers control the rising and falling edge of the output. This allows both positive going PWM pulses (when the rising edge occurs prior to the falling edge), and negative going PWM pulses (when the falling edge occurs prior to the rising edge).

Software Algorithm

For our software algorithm the incremental conductance method was used. The IncCond algorithm is based on the slope of the power. When the operating point in the *V–P* plane (as shown in Figure 9) is on the right of the maximum peak power (MPP), then the voltage is reduced, whereas when the operating point is on the left of the MPP, then the voltage is increased.

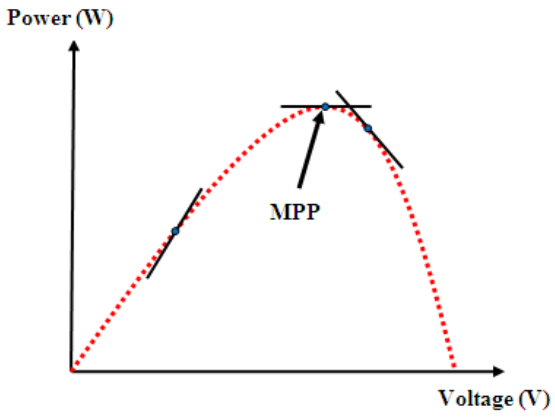


Figure 9. The *V–P* curve

Figure 10 provides a flowchart of the IncCond method. First, the algorithm obtains an initial measurement from the voltage and current sensors. Next, we initially increase the duty cycle and apply a delay to wait for the slow voltage and current sensor transient.

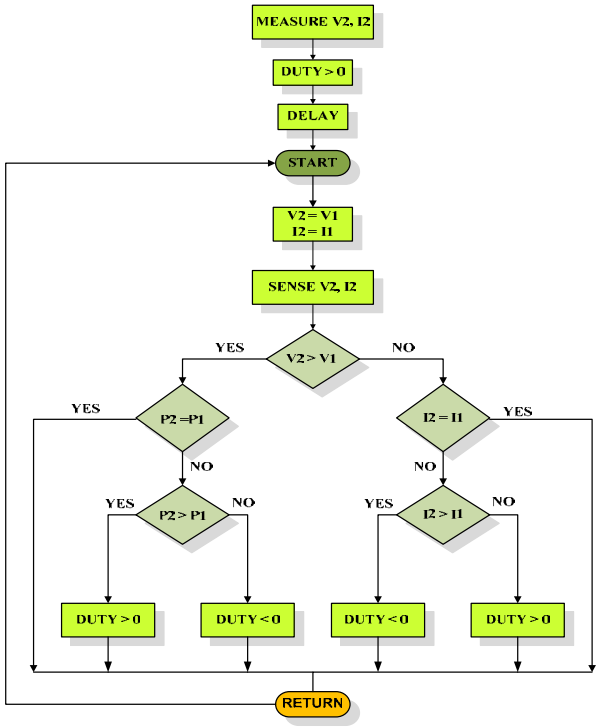


Figure 10. Flowchart of the incremental conductance method

TEST RESULTS

Indoor Testing

The indoor testing included the use of a halogen lamp that illuminated the solar panel with 500 watts of light. The set-up as shown in Figure 11 below was used for our circuit testing.

Figure 11. Photograph of the indoor solar panel testing set-up

Voltage Sensor

To test our voltage sensor design we connected the voltage sensor in parallel to the array. We then placed the 500-W halogen lamp at various distances, ranging from 36 inches to 8 inches. By measuring with both the Agilent Oscilloscope at the input and the Fluke multimeter at the output we obtained the following values as shown in Table I. According to the circuit design from Figure 6 it is set up to have a gain of 0.909. The measurements from Table I gave us an average gain of 0.916 which is a reasonable value compared to the calculated gain of 0.909.

Table I. Output/Input Voltage Measurements for Voltage Sensor

Distance (inches)	Vout (V)	Vin (V)	Gain (V/V)
36	2.258	2.472	0.913430421
34	2.257	2.48	0.910080645
32	2.272	2.492	0.911717496
30	2.294	2.503	0.9165002
28	2.302	2.519	0.913854704
26	2.313	2.534	0.912786109
24	2.333	2.548	0.915620094
22	2.335	2.549	0.916045508
20	2.358	2.58	0.913953488
18	2.365	2.584	0.915247678

16	2.376	2.603	0.912792931
14	2.394	2.616	0.915137615
12	2.401	2.623	0.915364087
10	2.423	2.635	0.919544592
8	2.5052	2.652	0.944645551

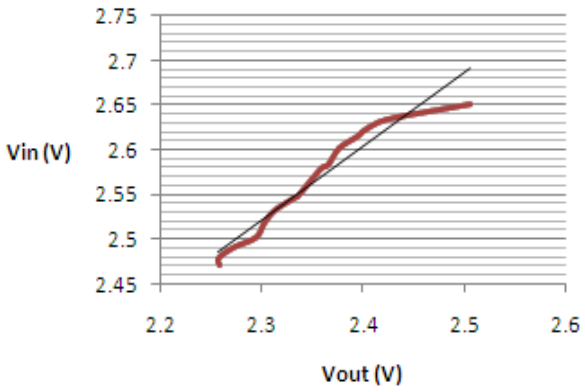


Figure 12. Output voltage vs. input voltage for voltage sensor

Current Sensor

To test our current sensor design, we connected the shunt resistor of the current sensor in series with the array. We then placed the 500 W halogen lamp at various distances as mentioned previously in the voltage sensor testing. Next, we measured with both the Fluke multimeter connected in series with the current sensor at the array output and the oscilloscope at the output of the current sensor to obtain the values shown in Table 2. According to the circuit design from Figure 7 it is set up to have a gain of 42.857. The measurements from Table 2 gave us an average gain of 43.88 which is also a reasonable value compared to the calculated gain of 42.857.

Table 2. Output/Input Voltage Measurements for Current Sensor

Distance (inches)	Vout (V)	Vin (V)	Gain (V/V)
36	0.6	0.01132	53.00353357
34	0.6	0.01244	48.23151125
32	0.6	0.01337	44.87658938
30	0.6	0.01486	40.37685061
28	0.9	0.01667	53.98920216
26	0.9	0.01907	47.19454641
24	0.9	0.02135	42.15456674
22	1.1	0.02384	46.1409396
20	1.1	0.02726	40.35216434

18	1.4	0.03256	42.997543
16	1.7	0.03862	44.01864319
14	1.9	0.0465	40.86021505
12	2.3	0.0605	38.01652893
10	3.6	0.0923	39.00325027
8	6.1	0.1648	37.01456311

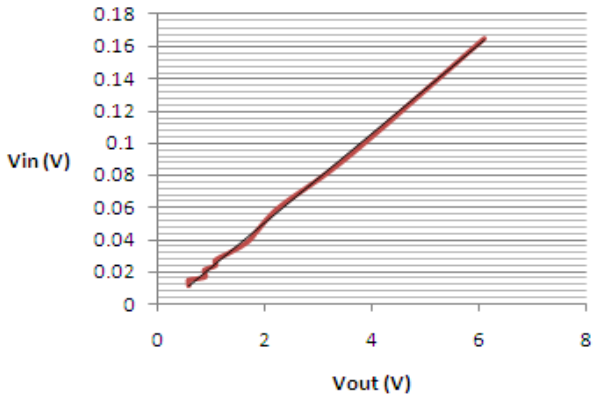


Figure 14. Output voltage vs. input voltage for current sensor

Buck Converter

The buck converter functionality was tested by applying a 52.3-ohm load and sweeping the duty ratio of the buck converter from 13.3 to 100%. This resulted in the graph of the buck converters full range of operation with a specific load. The results of the test are shown below in Figure 15. The output of the buck converter proved to be linearly dependent on the duty cycle of the power MOSFET as expected.

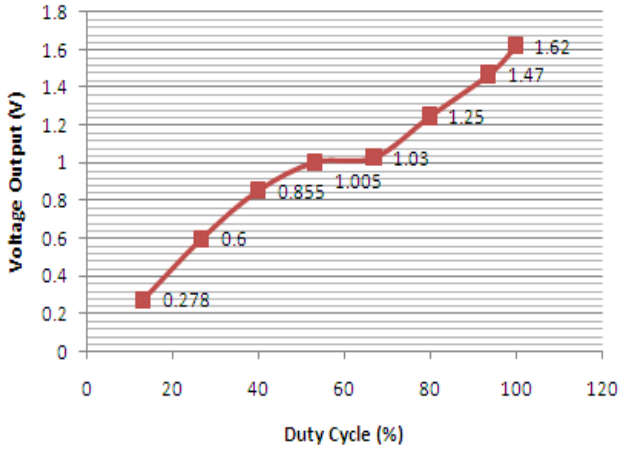


Figure 15. Buck converter voltage output as a function of duty cycle

CONCLUSION

The objective of this research is to develop a solar array peak power tracker that provides a load with maximum power. The design of a peak power tracking system proved to be a serious design challenge. There are many factors involved when designing a circuit that relies on both digital and analog aspects of circuitry. The experience is a valuable lesson in the problems that may occur when designing, assembling, and testing parts. Even though work is still under way and further analysis and testing are being planned, the process has provided an invaluable experience for the future.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081. I would like to thank Professor Sven Bilén and senior student at Penn State, Matthew Anderson, for their more than helpful contributions over the course of the research program. I would also like to thank them for the use of the Student Space Programs Laboratory and all of its equipment.

REFERENCES

[1] Aldous, Scott. “How Solar Cells Work.” How Stuff Works. Nov. 2, 2002. http://fitness.howstuffworks.com/solar-cell.htm?printable=1.

[2] Zweibel, Ken. Harnessing Solar Power: The Photovoltaics Challenge. New York: Plenum Press, 1990.

[3] ChihChiang Hua and Chihming Shen,”Comparative Study of Peak Power Tracking Techniques for Solar Storage System.” *IEEE* 1998.

[4] http://www.keil.com/dd/chip/4153.htm.

[5] Chad Alberts, Justin Braun and Anna Fowler,”Design of a Solar Array Peak Power Tracking System.” NASA1 Engineering, Inc., May 1, 2006.

[6] http://www.datasheetcatalog.com/

SPEECH PROCESSING USING MATCHED FILTERS BASED ON TRANSFORM DOMAIN PROCESSING

ZEHRA RAZA[*], Robert Nickel[#] and W. Kenneth Jenkins[#]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

[*]Undergraduate student of
School of Engineering and Applied Science
Department of Electrical Engineering and Computer Engineering
The George Washington University
Washington, DC 20052

ABSTRACT

Speech with additive noise found in many communication channels and systems ranging from military computer based systems to speech being transmitted over standard telephone circuits will be denoised and treated using the optimal matched filter. The primary goal of this work is to use the matched filter to implement high speed correlation/convolution using efficient transform based methods ranging from the Discrete Fourier Transform (DFT) to the Fermat Number Transform (FNT), which is among the most efficient methods with reduced computational complexity implemented in a finite field that possesses the same cyclic convolution property as the DFT. The process of correlation will take a small speech segment extracted from a segmented speech signal and will find the best enhanced speech segment within a large database. The best segment will replace the corrupted speech segment to allow clean speech communication.

INTRODUCTION

In the physical world we are surrounded by signals that are transmitted or recorded in a reverberant environment. Signal corruption can range from echoes in a recording studio, audio recording, and even conference telephony. However specifically in this research, we are focusing on speech signals with additive noise that can be found in many communication channels and even speech enhancement

[#] Faculty Mentor
[#] Faculty Mentor

has led to the research of creating optimal and adaptive microphones for speech in automobiles and aircrafts[1]. The usage of mobile telephones in cars has also increased the demand for hands free operation in car installations [1]. But in that environment speech is still affected by poor sound quality, additive noise, and acoustic feedback. Speech applications in communication systems are more significant because if distortion in speech occurs such as in the aircraft where the pilot is trying to communicate with air traffic controllers, serious consequences may occur. As a result, correct and clean speech has to be produced to assure accuracy and reliability in these types of situations.

This research will utilize a database of speech files with additive noise. Matched filters will be used to implement correlation using convolution that is achieved by DFT/FNT transform domain processing. Matched filtering has been chosen because it produces a maximum output when the segmented speech signal is time reversed and convolved with input signal, and then maximum amplitude detection is performed on the resulting convolution signal. It will be implemented using correlation. Correlation of two signals produces a measure of similarity between two signals. The approach used in this work is to correlate one signal with a large database of speech data to find the best segment that has similar frequency content and is in phase with one another. The basic goal is to determine the location of a digitized “signature” waveform residing within a large time slice”. The matched filter is optimal because the peak is well above the noise and amplitude of each point and the output signal is a measure of how well the filter matches corresponding sections of input signal. Matched filtering with correlation processing will be tested with two transforms:1) Discrete Fourier Transform (DFT) and a 2) Fermat Number Transform.

Some of the modern approaches to speech enhancement are being researched today such as using waveform filtering based methods to denoise speech [2]. But the speech processing done in this research has taken it in another direction through matched filtering. Many researchers are trying to find the most efficient method to eliminate speech with additive noise found in jet pilot communication systems. Whereas eliminating noise in speech is the secondary goal of this research. The primary goal has been defined to use matched filtering as an efficient method to implement a process that takes a small speech segment from the many speech segments within the large speech database. Once that best speech segment is found, it can be replaced with the original segment. In a way, we are enhancing speech, while at the same time additive noise is being eliminated. However matching a speech segment within a large database can take a lengthy period of time if an efficient method is not used. Efficiency, decreasing computational complexity, and striving towards low powered devices in speech led to use of discrete Fourier transforms to implement cyclic convolution through block processing to realize the matched filter. The most significance of this research has been to find an efficient way to handle the convolution of long data sequences. If the complete input speech signal is convolved with a test segment, it can become computationally expensive when computing a large length DFT. The

delay between the first input sample and full result becomes very long. As a result, it has proved better to divide the data sequence into subsequences and implement shorter DFTs on each sub-sequence. Basically, block processing will be used where the input signal is segmented into smaller blocks and each section is processed using a DFT. Now in order to combine results from these individual subsequences and to be certain that a correct result is obtained, the overlap save (or overlap add) method is used where the input signal x(n) is partitioned into sections but each will overlap with the previous one by (M-1) samples saving the last (N-M+1) output samples. This whole stage of processing the input signal through segmentation, circular convolution by a DFT, and block processing x (n) will implement the desired correlation.

Lastly, this research will experiment with ways to implement the process of cyclic convolution. In the beginning, the discrete Fourier transform (DFT) is being used which works on "sequences in the complex number field" [3]. The DFT requires multiplications that increase the computation time. When implementing the matched filter, the convolution property is demonstrated to "efficiently compute cyclic convolution of two long discrete sequences" [3]. In order for the cyclic convolution to be performed efficiently, it is possible to substitute the complex number field with a finite field where only a finite ring of integers are used. This research will introduce the Fermat Number Transform (FNT) where the computation of length N requires only "Nlog(2N) additions, bit shifts, and subtraction, but no multiplication" [3]. The cyclic convolution property in the implementation of the matched filter will be provided with a finite ring of integers. So as a result, cyclic convolution of the speech input signals will be implemented faster with less computational complexity and potentially less power consumption.

MATCHED FILTERS

In this study, match filtering is used to correlate a known signal with an unknown signal to detect the presence of the known signal. A matched filter is an optimum linear filter and an LTI system that maximizes signal to noise ratio in presence of stochastic or additive noise and minimizes bit error rate. The received signal is correlated with a filter that is parallel with the signal and that filter is orthogonal to noise. In order to design such a filter, we want to maximize signal power and minimize noise. In a matched filter, the signal-to-noise (S/N) ratio is maximized when the impulse response of that filter is exactly time reversed and is a time delayed copy of transmitted signal. The output of the filter is sampled at time T and the matched filter's waveform shape is a time reversed version of the transmitted waveform shape. If transmitted pulse shape is h(t) for $0 \leq t \leq T$, then the ideal matched filter response is $h_m(t) = h(T - t)$ for $0 \leq t \leq T$.

The matched filter algorithm is characterized by the following steps:

1.) Reverse the filter in the time domain
2.) Normalize the energy of each filter

3.) Convolve each filter with input signal
4.) Take maximum amplitude of the resulting convolution signals

The idea behind the matched filter is correlation. The amplitude of each point in the output signal is a measure of how well filter kernel matches corresponding section of input signal. As another way to put it, a matched filter is actually the same as correlating a signal with a copy of itself. Correlation is a weighted moving average:

$$r[n] = \sum_{k=0}^{L-1} x[k] * y[n+k] \tag{1}$$

Correlation is actually a measure of the similarity between two signals as a result of the time shift between them. If there are two similar signals and they are unshifted, then their product is positive. However if the shift increases, some parts of the product becomes negative and as a result, the correlation process allows us to see where signals are similar and not shifted. The value that is largest of the correlation process tells when two signals were similar and in phase with one another. Correlating a signal with another can be used to identify a signal by comparing it with a database of reference signals.

A similar process is used in this study to implement a matched filter where a copy of a reference signal is extracted from a segmented speech file and is correlated with a database of other speech signals. The correlation will be high if the extracted reference signal is similar to the unknown signal within the speech database. The largest correlation value will be the most likely match. The difference between convolving two signals and correlating them is convolution involves multiplying the frequency spectra of two signals whereas in correlation, multiplication of the complex conjugate of the frequency spectrum of one signal by the other is done.

In summary correlation is expressed by

$$r[n] = \sum_{k=0}^{L-1} x[k] * y[n+k] \ , \tag{2}$$

whereas convolution is also a weighted moving average with one signal flipped end-to-end:

$$r[n] = \sum_{k=0}^{L-1} x[k] * y[n-k] . \tag{3}$$

EXPERIMENTAL DESCRIPTION

This research on speech processing using Match Filtering based on the Fermat Number Transform was conducted through many range of simulations using the algorithmic level language: MATLAB. The research was done in a modular and top down design format. In the beginning, a simple scenario for matched filtering was created by setting the block length, doing FFT Processing

in blocks, and implementing matched filters with block processing. Matched filter was initially being implemented by following the matched filter algorithm where filters are reversed in the time domain, each filter is convolved with the input signal and the maximum amplitude is taken of the resulting convolution signals. MATLAB cconv function was used to zero pad the two signals before multiplying the FFTs. Since research is more of a trial and error procedure, efficient ways of implementing matched filters was experimented such as taking input data, slicing it into segments, and then extracting one segment to conduct the correlation match with the database. Implementation of different algorithms was done by programming in MATLAB.

Matched Filter Algorithm

1. The first step was to read a speech file called arctic_a0001.wav into matlab by using the wavread function and necessary information was found of the file such as duration of the sound file, sampling rate, and bit resolution.

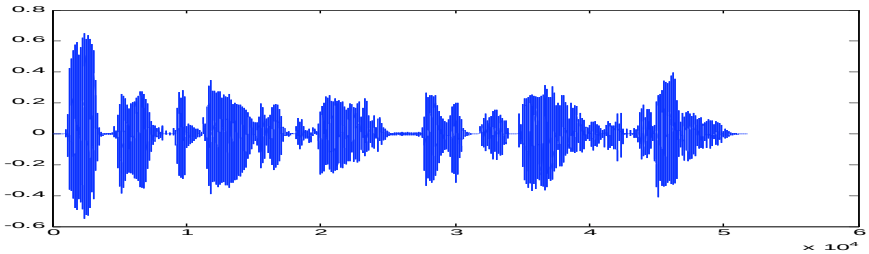


Figure 1: Speech Reference File 1- Arctic_a0001.wav

Information of the sound file "arctic_a0001.wav":
Duration = 3.23506 seconds
Sampling rate = 16000 samples/second
Bit resolution = 16 bits/sample

---

2. Next in order to avoid computing a large DFT and increasing computation time, the infinite length input was segmented into small sections using overlap and save method. A function called **zsegment** that takes the speech file, M-length of segment, and L which is length of segment minus overlap as its inputs. This function created a matrix S which was M by i. The (i) parameter represented number of segments created using this segmentation function. Zero padding was done by computing a parameter z which represented number of zeros to add. After testing this function, the input speech file arctic_a0001.wav was segmented properly.

   Below is the Segmentation Algorithm given in the MATLAB code:

```
function S=zsegment(wave,M,L)
%This function segments the input
%signal: wave.
%Example Test: segment(1:10,4,5)
%Wave is the input: So in this
%example, I am using 1:10 as my
%input.
%In this Example: M=4 and L=5.
%Since M represents my rows, I have 4
%rows in the output.
%You can see that according to the
%code below, zero padding has been
%done at the end of each row.
%So there are 3 segments (i).
% Answer is given below
% ans =

  %   1     6     0
  %   2     7     0
  %   3     8     0
  %   4     9     0

N=length(wave); %Length-L1
i=ceil((N-M)./L+1); % # of segments
L2=L*(i-1)+M; %Length-L2
Z=L2-N; % # of zeros to add-(Zero
%Padding)
wave=wave(:).'; % Make it a row
%vector
wave=[wave zeros(1,Z)]; %Attach zeros
S=zeros(M,i); % Initialize matrix

for n=1:i;
    % Assign the columns of the
    % matrix one at a time.
   S(:,n)=wave((n-1)*L+1:(n1)*L+M).';
end
```

3. It was time to start implementing the matched filter. Since the matched filter implements correlation through convolution. A function called correlation was created that had 2 inputs called xlong and xpart. Xlong represented the segmented reference speech file 1 and xpart represented the impulse response that was extracted from xlong to correlate both signals and find a measure of similarity between the 2 signals. This function correlated one signal with a large database of speech data to find the best one that has similar frequency content and is in phase with one another. The function returned the parameters xmatch, rr_max, and idx. Xmatch was the best matching segment in xlong, rr_max was the maximum correlation value, and idx was the index where the maximum correlation value was located.

   Below is the Correlation Algorithm given in the MATLAB code:

```
function[xmatch,rr_max,idx]=correlation(xlong,xpart)

%This function implements correlation which is done of
%two signals and it will be a measure of similarity
%between two signals.

% convert to row vectors
  xlong=xlong(:).'; xpart=xpart(:).';

% find the length of the sequences
  Nlong=length(xlong); Npart=length(xpart);

% computation of the correlation
  rr=conv(xlong,fliplr(xpart));

% normalization of the correlation
  q=ones(1,length(xpart)); qq=conv(xlong.*xlong,q);
  rr=rr./(sqrt(qq).*sqrt(sum(xpart.*xpart)));

% extracting the relevant subset of correlation values
  rr=rr(Npart:end-Npart+1);

% finding the index of the maximum correlation value
  [rr_max,idx]=max(abs(rr));

% extract the best matching segment in xlong
  xmatch=sign(rr(idx))*xlong(idx:idx+Npart-1);
```

4. Next step was to implement the computation of correlation using the discrete Fourier transform utilizing the fft and ifft commands in MATLAB. Previously, the (conv) command was used to compute correlation that implemented linear convolution. At this stage, circular convolution was being done using the DFT. Again the same procedure applied where the computation of the correlation was being done on the long signal (xlong) and the impulse response (xpart). The new function created was called **correlationdft.**

   The MATLAB code that was changed from the previous correlation function was the part where the computation of correlation was being done.

```
% computation of correlation

rr=conv(xlong,fliplr(xpart));
```

→

```
Nz= Nlong+Npart-1;
z=[fliplr(xpart) zeros(1,Nz-
Npart)];

% computation of DFT correlation
rr=ifft(fft(xlong,Nz).*fft(z));
```

5. Next algorithm that was implemented was the Overlap Add function that added the output segments together to form a continuous output signal. This was done by doing linear cross-fading between segments to reduce distortion and produce better sounding speech. An important point to

mention is that this Overlap Add function is completely different from the Overlap-Add block processing algorithm that is described later for efficient transform domain processing of long input data sequences.

Below is the MATLAB code for the Overlap Add function:

```
function wavenew=zovrlpadd(S,L)
%The purpose of this function is to add the
%output processed speech segments using linear
%cross-fading.
M=size(S,1); i=size(S,2); K=M-L;

Wavenew = S(:,1).';

for k=2:i;

     s=S(:,k).';
    %Beginning part of first output wave
    %segment
    x1=wavenew(1:(k-1)*L);
    %Fading in the first output wave segment
    %linearly
    x2=wavenew((k1)*L+1:end).*linspace(1,0,K);
    %Fading out the second output wave segment
    %linearly
    x3=s(1:K).*linspace(0,1,K);
    %End part of the second output wave segment
    x4=s(K+1:end);
    %Adding the segmented output wave together
    wavenew=[ x1 x2+x3 x4 ];

end
```

6. The last algorithm written in MATLAB was the zdenoise algorithm. This algorithm combined all the written MATLAB functions together such as the zsegment, correlation, and overlap-add. The main purpose of this algorithm was to denoise the input signal by first segmenting the input speech signal, extracting one segment from the segmented speech signal representing the impulse response, and correlating the single segment with 10 other speech databases to find the best clean segment. Once the best segment is found within the database, then the overlap add function will replace the corrupted segmented with the denoised speech segment and add all the output speech segments together. Length M of each segment was set to 320 samples with 50% overlap; so L was 160 samples long. At the end, two waveforms were generated; one was the input signal and the second one was the output correlated signal.

   Below is a flow chart that explains the Matched Filter process:

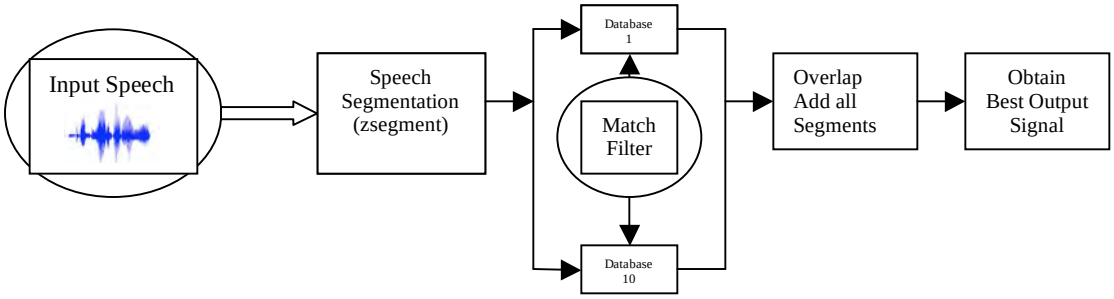


Figure 2: Matched Filter Flow Chart

DERIVATION OF THE MATCHED FILTER

We will determine the best optimum receiver filter H(f) in frequency domain. If p(t) is the received pulse then

$$\rho^2 = \rho^2(t_m)/\sigma_n^2 \ , \tag{4}$$

where ρ^2 is the signal-to-noise ratio (SNR) at time $t = t_m$. Then

$$F\{p(t)\} = P(f).$$

In the following discussion the Pulse Spectral Density of the channel noise will be denoted by $S_n(f)$.

Optimum Receiver Filter in Frequency Domain

$$p_o(t) = F^{-1}[P(f)H(f)] = \int_{-\infty}^{\infty} P(f)H(f)\, e^{-2\pi f t}\, df$$

Therefore at $t = t_m$

$$p_o(t_m) = \int_{-\infty}^{\infty} P(f)H(f)e^{-2\pi f t} df \tag{5}$$

Once the noise has been filtered out, the mean is zero, while its variance is

$$\sigma_m^2 = \int_{-\infty}^{\infty} S_n(f)|H(f)|^2\, df \tag{6}$$

So as a result, the SNR at time $t = t_m$ is

$$\rho^2 = \frac{p_o^2(t_m)}{\sigma_n^2} = \frac{\int_{-\infty}^{\infty} P(f)H(f)\,e^{-2\pi f t}\,df}{\int_{-\infty}^{\infty} S_n\,(f)|H(f)|^2\,df} \tag{7}$$

Now by applying the Cauchy Schwarz Inequality the optimum filter H(f) is determined to be:

$$X(f) = H(f)\sqrt{S_n(f)} \tag{8a}$$

$$Y(f) = \frac{P(f)e^{-j2\pi f\,t_m}}{\sqrt{S_n(f)}} \tag{8b}$$

So, by applying the Cauchy Schwarz Inequality to the numerator of equation (7),

$$\rho^2 = \frac{\left|\int_{-\infty}^{\infty} X(f)H(f)df\right|^2}{\int_{-\infty}^{\infty}|X(f)|^2\,df} \leq \frac{\int_{-\infty}^{\infty}|X(f)|^2 * \int_{-\infty}^{\infty}|Y(f)|^2 df}{\int_{-\infty}^{\infty}|X(f)|^2 df}$$

$$= \int_{-\infty}^{\infty}|Y(f)|^2 df = \int_{-\infty}^{\infty}\left|\frac{P(f)e^{-j2\pi f\,t_m}}{\sqrt{S_n(f)}}\right|^2 = \int_{-\infty}^{\infty}\frac{|P(f)|^2}{S_n(f)}\,\mathrm{df}\,, \tag{9}$$

ρ^2 is maximized if and only if

$$X(f) = k[Y(f)]^* \text{ and } H(f)\sqrt{S_n(f)} = \mathrm{k}\left[\frac{P(f)e^{-j2\pi f\,t_m}}{\sqrt{S_n(f)}}\right]^*.$$

So the SNR is maximized if and only if:

$$H(f) = \frac{k\,P(-f)e^{-j2\pi f\,t_m}}{S_n\,(f)} \tag{10}$$

It is concluded that matched filter is optimal because it maximizes the SNR and minimizes bit error rate.

MATCHED FILTER USING FOURIER BLOCK PROCESSING

In this study, the matched filter has been implemented through block processing experimenting with Overlap and Add and Overlap and save methods. It has been known that most input signals have long lengths and when it comes to processing the signal, limited computer memory is available. When implementing the matched filter using the discrete Fourier transform (DFT) on a large number of points, the DFT becomes too large to compute and as a result, a long delay is created. To avoid processing delay, the input signal is segmented into blocks of length L. Then each block can be convolved with a finite impulse response and the DFT can be computed. This procedure is developed as follows.

Assume the input signal is x(n) with N samples. The input signal is segmented into i number of blocks of length L.

N - Number of samples in the “long” input signal
M - Length of impulse response
L - Length of each individual block
i - number of blocks or segments

The equation for the number of segments will be computed:

$$(i-1)*L + M \geq N$$
$$(i-1)*L \geq N-M$$
$$(i-1) \geq (N-M) / L$$
$$i \geq (N-M) / L +1$$
$$\mathbf{i = ceil ((N-M)/L+1)} \quad (11)$$

Equation (11) indicates how many segments have been created by segmenting input signal x(n). After x(n) convolves with h(n), the output filtered block will have a length of L+M-1.

The first well known block processing algorithm is the *Overlap-Save* algorithm where L is selected to be some suitable value with L > M. The signal x[n] is divided into blocks of length L and they overlap by M-1 data points. Therefore, the i^{th} block is

$$x_i[n] = x[n + i(L - M + 1)], n = 0, \dots, L - 1.$$

The filter impulse response h[n] is augmented with L-M zeros to produce:

$$h_{pad}[n] = \begin{bmatrix} h([n], & n = 0,..., N-1 \\ 0, & n = N,..., N_{DFT}-1 \end{bmatrix} \quad (12)$$

Then, the DFT is utilized to get The DFT is then used to obtain $Y_{pad}[n] = DFT\{h_{pad}[n]\} \cdot DFT\{s_k[n]\}$, and $y_{pad}[n] = IDFT\{Y_{pad}[n]\}$. From the $y_{pad}[n]$ array the values that correctly correspond to the linear convolution are saved; values that are erroneous due to wrap-around error caused by the circular convolution of the DFT are discarded. The k^{th} block of the filtered output is obtained by

$$y_k[n] = \begin{bmatrix} y_{pad}([n], & n = 0,..., N-1 \\ 0, & n = N,..., N_{DFT}-1 \end{bmatrix} \quad (13)$$

So for this algorithm, every time a block is processed, M-N+1 points are saved and M-1 points are discarded. Each block moves ahead by M-N+1 data points and overlaps the preceding block by N-1 points.

The second well known block processing algorithm is the *Overlap-Add* algorithm in which it is similar to the *Overlap-Save* except the i^{th} input block is designated to be:

$$s_k[n] = \begin{bmatrix} s[n], & n = 0,..., L-1 \\ 0, & n = L,..., N_{DFT}-1 \end{bmatrix} \quad (14)$$

where L= M-N+1. Again, the filter function $h_{pad}[n]$ is increased with zeros to produce $h_{pad}[n]$, and then the DFT is computed. Now in each output block $y_{pad}[n]$, the first N-1 points are incorrect and the last N-1 points are incorrect. The center M-2(N-1) points correspond to linear convolution. But however in order to obtain linear convolution, last N-1 points from block I are overlapped with the first N-1 points from block i+1 and then are added. So after doing addition, the correct number of points for each block is M-N+1.This study has focused on lowering the computational complexity and look toward low powered systems. As a result, these block filtering methods helped contribute to minimizing computation time and allowing high data rates.

COMPARISON OF DIFFERENT TRANSFORM DOMAIN PROCESSES

This study used the Discrete Fourier Transform (DFT) to lay the platform to arrive to transforms that have low computational complexity and high efficiency in terms of low power. DFT is obtained by sampling the continuous frequency domain at N points that are $w_k = (2\pi k/N),\ k = 0, 1, \ldots, N-1$. Discrete Fourier Transform is the resulting Fourier representation for finite duration sequences in the interval $0 \leq n \leq N-1$ that is discrete in both time and frequency [4]. As a result, DFT transform pair is:

$$X[k] = \sum_{n=0}^{L-1} x[n] e^{-j\left(\frac{2\pi}{N}\right)kn} \; for\; k = 0,1,\ldots,N-1 \quad (15a)$$

$$x[k] = \frac{1}{L}\sum_{k=0}^{L-1} X[k]\, e^{-j\left(\frac{2\pi}{N}\right)kn} \; for\; n = 0,1,\ldots,N-1 \quad (15b)$$

The DFT treats signal x[n] as though it were a single period of a periodic sequence [5]. An essential application of the DFT is the execution of fast convolution which has been focused on in this study. The structure of the DFT demonstrates the cyclic convolution property. Since DFTs are periodic and have to be nonzero for $n \geq N$, the circular convolution operation allowed convolved signal to be zero outside the range n= {0, 1… N-1}. But multiplication of two DFTs results in circular convolution of discrete time sequences [5]. As stated that "circular convolution is simply a linear convolution of the periodic extensions of the finite sequences being convolved" [5] where one period is of a finite length N.

For example, when implementing a filter with a finite impulse response h[n], the output to input x[n] is:

$$r[n] = \sum_{k=0}^{L-1} x[k] * y[n-k] \tag{16}$$

Basically, the steps in conducting the cyclic convolution of two signals are:

1. Using the DFT, convert h[n] and x[n] into H[k] and X[k].
2. Multiply the both DFTs point-wise to get Y[k] = H[k] X[k].
3. Then use inverse DFT to get :

$$y[n] = DFT^{-1}\{Y[k]\}.$$

The output of the cyclic convolution computed by the DFT will realize linear convolution if and only if block length of the DFT is ≥ L+M-1 and h[n] and x[n] are zero padded to form blocks of length M [5]. This research has focused on utilizing transforms similar to the DFT to implement cyclic convolution in an efficient manner. The DFT can be implemented efficiently with fast Fourier transform (FFT) algorithm. "FFT is a computationally efficient algorithm that decreases complexity of the computing DFT from Order $\{N^2\}$ to Order $\{N\log_2 N\}$.

THE FERMAT NUMBER TRANSFORM-FINITE NUMBER FIELDS

In order to process signal with a digital computer, efficient computational algorithms have to be implemented. In this study, it has been found that processing infinitely long sequences is a difficult task, as a result, the main goal has been to find best signal processing algorithm in the case of finite length sequences. In this study, implementing a matched filter using a method that is efficient, has low computational complexity, and can implement a DFT algorithm with less complex multiplications has been the objective of this study. Since correlation is implemented through convolution, minimization of the number of the multiplications for circular convolution has been extremely necessary to decrease computation time and to quickly process smaller length DFTs.

The bridge that helps in traversing between certain transforms is the cyclic convolution property. This property is crucial in computing cyclic convolution of two long discrete sequences. In the beginning, researchers have explored in using the DFT with the fast Fourier Transform to compute the DFT in the complex number field. This method involved $(N/2)\log_2(N/2)$ complex multiplications when computing the DFT of a length N signal. This method completed the job but it significantly increased computation time [7]. So as a result, the idea that "the complex number field of the continuous domain can be substituted with a finite field or a finite ring of integers" [7] was accepted. In this ring, the cyclic convolution property can be used very efficiently by introducing a transform similar to the DFT which a number theoretic transform is called the Fermat Number Transform. Fermat Number Transform (FNTs) are defined in a Galois

field (finite field) on the basis of Fermat primes of the form $F_t = 2^b + 1$, where $b=2^t$ and t is a positive integer. For a discrete time sequence x[n], the FNT is given by the following pair of equations [8]:

$$\langle X[j] \rangle \, mod \, F_t = \langle \sum_{n=0}^{N-1} x[n] \, \alpha^{nj} \rangle \, mod \, F_t \tag{17a}$$

$$\langle x[n] \rangle \, mod \, F_t = N^{-1} \langle \sum_{j=0}^{N-1} X[j] \, \alpha^{-nj} \rangle \, mod \, F_t \tag{17b}$$

The indexes j and n range from 0,…, N-1, α is an integer of order N,

$$\alpha^N = 1 \, mod \, F_t$$
$$F_t = 2^{2^t} + 1$$

and F_t is as the t^{th} Fermat Number. It was stated in [7] that this transform resembles an "FFT structure of length N and it requires on the order of $(N)log_2(N)$ additions, bit shifts, and subtraction, except no multiplication. If the block length N is selected to be a power of 2, then the FNT can be implemented with an FFT based method and multiplications are reduced [3]. FNT based computation again implements cyclic convolution to obtain an output that resembles linear convolution. It is better for α to have a simple binary representation, so that multiplications are easy to implement. Since this transform is defined in the algebraic system of integers modulo, F_t then implementing circular convolution will give circular convolution modulo F_t [3]. If the circular convolution Y_n

$$Y_n = \sum_{k=0}^{N-1} x_k h_{(n-k)}$$

is calculated using the FNT, the resulting output would be

$$Y_n = \left\{ \sum_{k=0}^{N-1} x_k h_{(n-k)} \right\} mod \, F_t .$$

RESULTS

After writing all the MATLAB algorithmic level functions to do segmentation, correlation, and combining output segments, the final simulation was conducted. The input speech signal arctic_a0001.wav was read in MATLAB, segmented into blocks of M samples long-chosen to be 320 and L-samples of overlap chosen to be 160. The input speech file and one of its extracted segment was correlated with a database that consisted of 9 other speech files to de-noise

the input signal and find the best speech segment to enhance the intelligibility of the input signal and allow clean speech communication.

The first final simulation was done between the input signal and four other speech files in the speech database. The simulation took about 5 minutes to complete due to the process of correlating long speech files. The output waveform is presented in Figure 3.

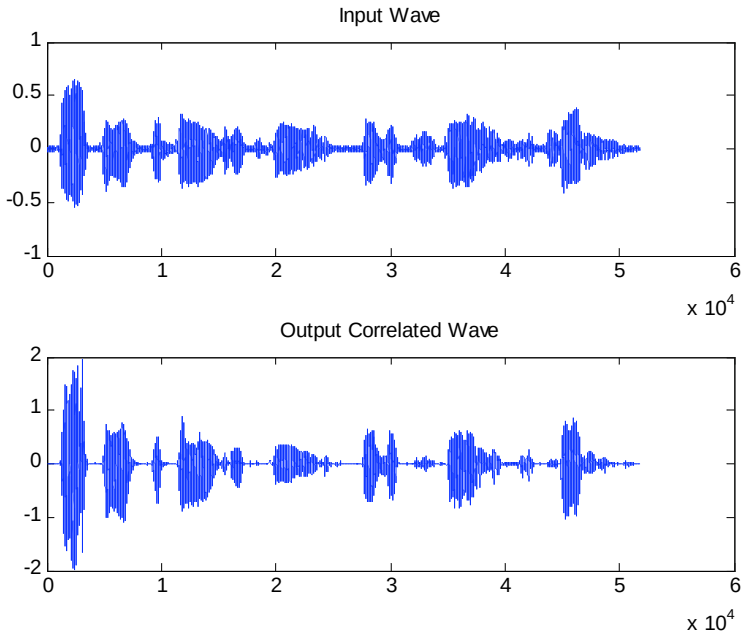


Figure 3: Final Simulation 1-Correlation with 4 Speech Databases

The second final simulation was done between the input signal and 9 other speech files in the speech database. The simulation again took 5 minutes to complete due to the process of correlating long speech files. The output waveform is presented in Figure 4.

DISCUSSION

After conducting final MATLAB simulations with the input signal and the speech database consisting of 9 other speech files, 2 waveforms were obtained showing the input wave and the output correlated wave. In Figure 4, the input has significant amounts of additive noise being shown in the waveform. However after the denoising algorithm has been used to process the input signal, the output correlated wave shows reduced amounts of noise. The matched filter implemented by process of correlation through high speed DFT convolution did an excellent job in finding the best speech segment to allow clean speech communication because after listening to the output correlated wave, the speech was 85% clean and intelligible compared to the input signal. Some acoustic noise was still present in the output wave, although there is always a certain level of ambient

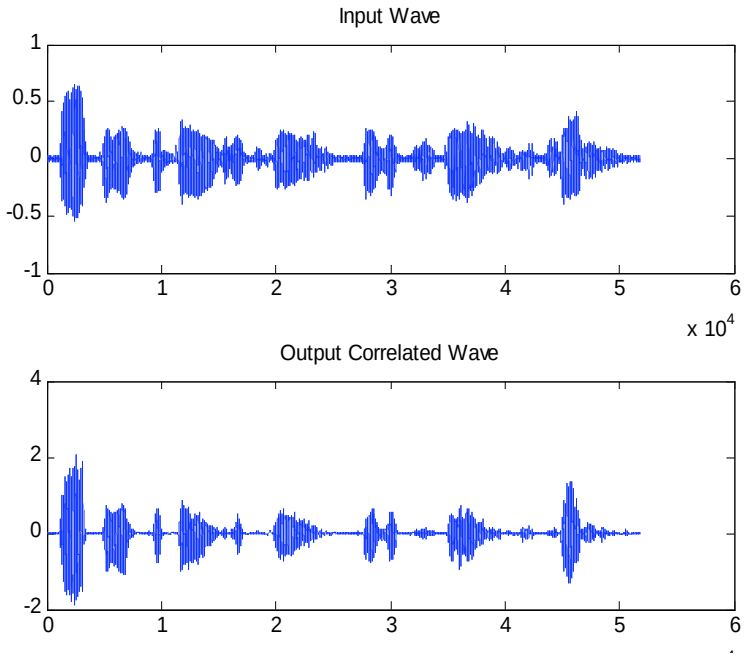


Figure 4: Final Simulation 2-Correlation with 9 Speech Databases

background noise present in any speech. After listening to the output signal 2 to 3 times, the spoken words and phrases were clearly understood. The discrete Fourier transform was an excellent choice for a transform domain process because speech processing was done in a timely manner and it didn't take a lengthy duration for the computation of the correlation to be done. As this research proceeds forward in the future, the Fermat Number Transform as discussed in the paper can be implemented in MATLAB to decrease the computation time and matched filter algorithm will contain less computational complexity. Low powered systems as always needed by the military can be effectively implemented.

CONCLUSION

The goal of this study was to use the matched filter to implement high speed correlation/convolution using efficient transform based methods ranging from the Discrete Fourier Transform (DFT) to the Fermat Number Transform (FNT). Successful implementation of the matched filter through process of correlation allowed speech processing to be done effectively. The input speech signal was successfully denoised and the best speech segment was found within the large database of speech files and it replaced the corrupted speech segment which allowed clean speech communication. The additive noise was eliminated from the output speech signal and the speech was intelligible; every spoken word was clearly understood.

ACNOWLEDGEMENTS

One of the authors (ZR) would like to thank Dr. Robert Nickel and Dr. W. Kenneth Jenkins for their extensive knowledge and help on this project. ZR would also like to thank Penn State for the opportunity to be a part of the EE REU program. This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081.

REFERENCES

[1] Lars Hakansson, Sven Johansson, Mattias Dahl, Per Sjosten and Ingvar Claesson., "Noise Cancelling Headsets for Speech Communication., pp. 305-306. Edited by: Gillian M. Davis., CRC Press

[2] Xiaqiang Xiao, Peng Lee, and Robert M. Nickel, "Inventory Based Speech Denoising with Hidden Markov Models," Proceedings of the 16th European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland, August 25-29, 2008.

[3] Ramesh C. Agarwal and Charles S. Burrus, "Fast Convolution Using Fermat Number Transforms with Applications to Digital Filtering", IEEE Trans. *Acoustics, Speech, and Signal Processing.,*Vol. Assp-22., pp. 87-88, April 1974.

[4] Alan V. Oppenheim, Ronald W. Schafer, The Discrete Fourier Transform, pp. 99, 112-113, Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[5] W. K. Jenkins, "Fourier Methods for Signal Analysis and Processing," Chapter 1 of *The Fundamentals of Circuits and Filters Handbook* (vol. I of the *Circuits and Filter Handbook)*, 3nd edition, Wai-kai Chen, ed., CRC Press, 2009.

[6] B.P. Lathi, Zhi Ding, Optimum Receiver Filter-Matched Filter, pp. 607-611, Modern Digital Analog Communication Systems, Fourth Edition, Oxford University Press, New York, 2009.

[7] James H. McClellan, Charles M. Rader, Number Theory in Digital Signal Processing, pp. 84, Prentice Hall, Inc. Englewood Cliffs, NJ, 1979.

[8] C. Radhakrishnan and W. K. Jenkins, "Fault Tolerant Adaptive Filters Based on Number Theoretic Transforms," *Proceedings of the Forty-Second Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, pp. 261-265, October 26-29, 2008.

[9] James H. McClellan, "Hardware Realization of a Fermat Number Transform", IEEE Transactions on Acoustics, Speech, and Signal Processing, VOL. ASSP-24, NO. 3, June 1976.

COMPARISON OF NEMA 23 BLDC AND STEPPER MOTORS USED AS GENERATORS

Nicholas Czarnek,[*] Heath Hofmann,[#] and Junyu Yao[†]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

[*]Undergraduate student of
Department of Electrical Engineering
University of Pittsburgh
Pittsburgh, PA 15904

ABSTRACT

The purpose of this paper was to compare the effectiveness of two types of machines used as generators for energy harvesting applications. A NEMA 23 brushless direct current (BLDC) motor and three NEMA 23 stepper motors were bracketed to a large DC motor that served to turn the shafts of the smaller motors, and the output power was measured for each of the machines. The power density of the machines was then calculated normalized to their mass, and used to conclusively determine the effectiveness of the machines as generators to be used in various energy harvesting applications.

INTRODUCTION

Electronic devices are becoming more and more common in today's world, from cell phones to output displays for various equipment. Powering these devices requires batteries or some other electrical power source, like a wall outlet. Either way, energy must be extracted, converted into usable electrical form, and used to power these devices. Concern over the Earth's limited energy resources has also grown along with the growth of the electronic device industry. This concern is the leading drive behind research into energy harvesting, or converting ambient energy that would otherwise be wasted into a usable form. Many solutions have already been proposed and implemented, such as wind mills and hydroelectric dams. On a smaller scale, generators have been incorporated into

[#] Faculty Mentor
[†] Graduate Student

bicycles [1] or elliptical machines [2] to capture the energy that humans use during exercise. Saez et al. [3] studied human power generation and discovered that one minute of bicycle pedaling could charge a car battery with enough energy to power a PC for five minutes. This led to the thought that perhaps someday, gyms with modified exercise equipment could be used to recharge batteries of the users' electronic devices, such as iPods or phones.

Following this concern about Earth's limited resources, this study focuses on the use of motors as generators to collect ambient energy, specifically focusing on the comparison between the effectiveness of a Brushless DC and a hybrid stepper machine both run at low shaft velocities to simulate the conditions in which the energy would be harvested.

BACKGROUND

Brushless DC (BLDC) machines are synchronous electric machines that are commutated electronically instead of with brushes as used by brushed DC machines. The winding through which electric current is run in the machine is attached to the machine housing, or stator, while the magnets are on the rotor of the machine. Since the windings of the machine are attached to the stator, heat elimination is efficient as the hot stator is cooled by the air surrounding it [4]. A cutaway image of a BLDC motor is shown in Figure 1 [5].

Figure 1: BLDC motor cutaway

For the following discussion about hybrid stepper motors, refer to Figures 2 and 3 [6, 7]. Whereas in the BLDC motor, the rotor is in the presence of a magnetic field created by two or more permanent magnets, in a hybrid stepper motor, the rotor contains a single permanent magnet. All of the motors used in this experiment had two stacks, one at each end of the magnet. The stacks are attached to the north and south ends of the rotor's permanent magnet. Also, the stacks and stator of the motor have stamped teeth, which cause the torque to be created. All of the motors tested had two phases. The stator of each is made up of eight windings in a circle, with every other winding corresponding to the same phase. Figure 2 shows a basic diagram of a three pole, two phase hybrid stepper motor with four windings.

Stepper motors are called such because they usually move in distinct steps with easily measurable angles by exciting phase windings in sequence. In Figure 2, part a, the vertical winding is excited, meaning that a current is passed through it. By the right hand corkscrew rule, this causes the top portion of the winding to

be a north pole and the bottom portion to be a south pole. This causes the southern end of the rotor to align with the top, and the northern end of the rotor (shown with dotted lines) to align with the bottom. This can be seen as the starting position for the stepping motor. Now, in order to make a step, the vertical winding's current must stop, and the horizontal winding must be excited, as shown in part b. This excitation, again by the right hand corkscrew rule, causes the rotor to realign so that the north and south ends of the rotor are now matched up to the induced magnetized south and north, respectively, portions of the stator. Notice that the rotor moved by 30º in order to achieve this alignment; this is one step. To make another step, a current must be run through the vertical winding, shown in part c, in the opposite direction to the current originally run through part a. This causes the previous north end of the winding to become the south end, and the previous south end to become the north end. Again, the rotor realigns, making another 30º step. This process continues to make a series of precise steps. To recap the currents that cause the magnetic fields in the winding, first the vertical winding is given a current in the positive direction. Next, the horizontal winding is given a current in the positive direction. A negative current is then sent through the vertical winding, followed by a negative current sent through the horizontal winding. This causes four steps to be taken, after which the process can repeat itself.

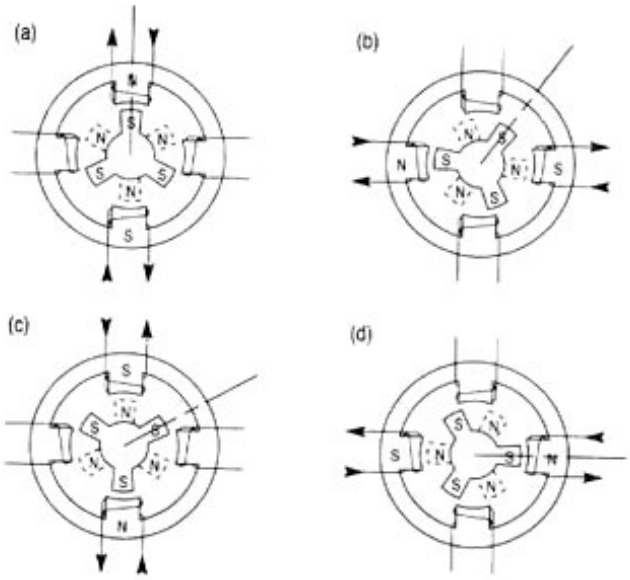


Figure 2: Frontal view of hybrid stepper motor

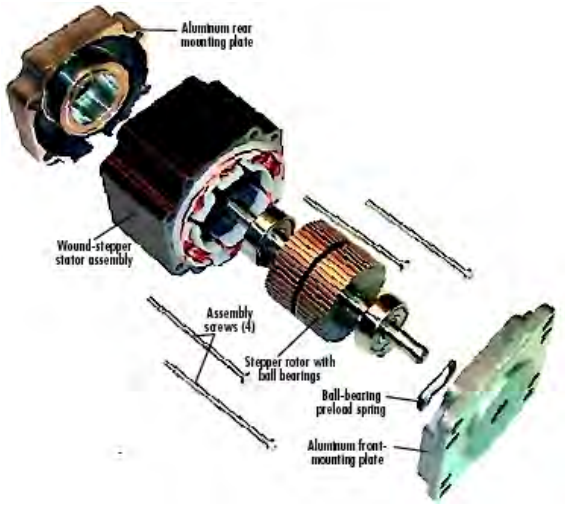


Figure 3: Hybrid stepper motor expanded

Figure 4 shows a chart of all of the relevant information for each of the motors tested in this experiment.

	Motor			
	Stepper			BLDC
	5704M-02	5709M-02S	5718M-02S	BLDC23F-23
Size	NEMA 23	NEMA 23	NEMA 23	NEMA 23
Mass (kg)	0.68	0.68	0.68	0.64
# of Pole Pairs	200	100	50	8
Step Angle	0.45°	0.9°	1.8°	
Line to Line Resistance (Ohm)	3.6	2.1	2.4	1.5
Line to Line Inductance (mH)	4	7.8	8.5	0.69

Figure 4: Motor information

EXPERIMENTAL SETUP

Four NEMA 23 size motors were tested in this experiment: a BLDC motor with eight pole pairs and three stepper motors with 50, 100, and 200 pole pairs. Since the motors were being used as generators, the shafts had to be turned in some controllable way. In order to turn the shafts, a large 25 HP Sabina EE DC motor was used. To provide the necessary DC input voltage to the motor, the 120 V AC wall voltage had to be rectified using a simple 8A/400V full wave bridge rectifier circuit. A stabilizing polarized 11 mF capacitor was connected in parallel to the bridge rectifier through a current limiting resistor to create a low ripple DC voltage. This DC input voltage caused the current in the field winding of the motor, which created a magnetic field for the armature winding of the shaft. The

armature input winding was connected to a DC power supply, which in turn receives its power from a 208V source. Now that the driving motor was connected, a bracket was made to couple the shaft of the large DC motor to the shafts of the smaller NEMA 23 motors for testing. This completed the setup for the large DC motor.

When the stepper and BLDC machines were used as generators, they created an AC output voltage that had to be rectified to DC in order to measure the output power generated. For the stepper motors, according to Geiger [8], a capacitor was needed to overcome the inductive impedance from the coils of the stator windings according to

$$C = \frac{1}{\left(\frac{P}{2}\omega\right)^2 L}$$

where $\frac{P}{2}$ = number of pole pairs
ω = angular velocity = 2π (frequency)
L = impedance.

For the tested frequencies of 5-25 Hz in 5 Hz increments, this resulted in the following values for the AC capacitors as shown in Table 1.

Table 1: Stepper Capacitors to Cancel Inductive Impedance

	# of Pole Pairs	Inductance (mH)	5 Hz Capacitance (µF)	10 Hz Capacitance (µF)	15 Hz Capacitance (µF)	20 Hz Capacitance (µF)	25 Hz Capacitance (µF)
5718M-02S	50	8.5	48	12	5.3	3	1.9
5709M-02S	100	7.8	13	3.2	1.4	0.81	0.52
5704M-02S	200	4	6.3	1.6	0.7	0.4	0.25

The BLDC machine did not require a capacitor for the AC output because of the differences in construction that resulted in extremely low inductance. However, both types of motors required a diode rectifier: two two-phase rectifiers for the steppers and one three phase rectifier for the BLDC machine. Low-forward-voltage Schottky diodes were used for these rectifiers. Both types of machines also required capacitors to stabilize the rectified pulsating DC voltage. Because the steppers had so many more poles, the output voltage generated had a higher frequency than the output of the BLDC motor. Therefore, smaller capacitors were required to filter the DC pulses in the steppers than in the BLDC. A 230 µF film capacitor was used to filter the output from the stepper motors, and a large 11 mF polarized electrolytic capacitor was used to stabilize the BLDC output. Thus, the schematics for the BLDC and stepper rectifiers were generated as shown in Figures 5 and 6, respectively, with C-AC as the capacitor to cancel out the inductive impedance in the steppers and C-DC as the capacitor to stabilize the

pulsating DC output. Note that in Figure 5, the J1 connector represents the BLDC motor.

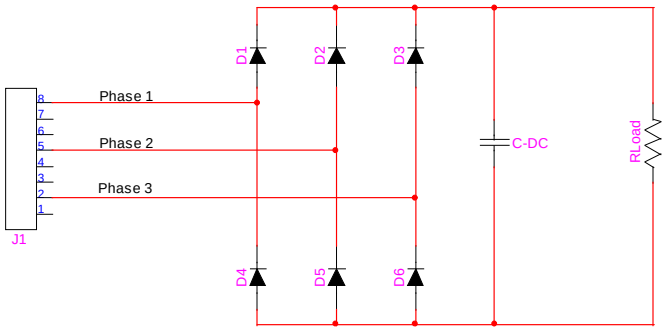


Figure 5: BLDC rectifier circuit

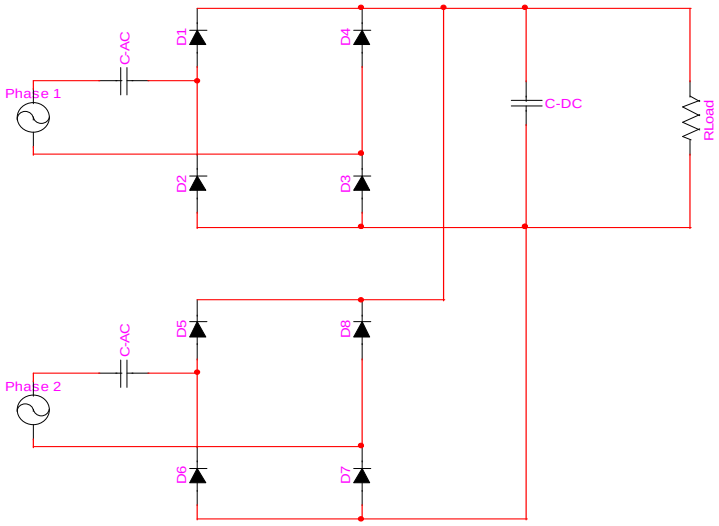


Figure 6: Stepper rectifier circuit

In order to produce the necessary capacitances for the C-AC on the Stepper Rectifier, a solder breadboard was used with ceramic capacitors. 11 switches were used to control 11 parallel capacitances of 0.05, 0.1, 0.2, 0.4, 0.4, 1, 2, 4, 8, 16, and 32 µF to allow any combination of capacitors up to 64.15 µF with an accuracy of 0.05 µF. A picture of the experimental setup is shown below in Figure 7.

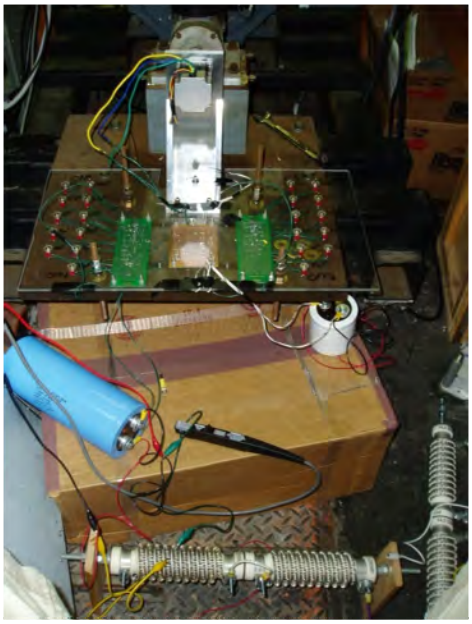


Figure 7: Experimental setup

After soldering the rectifying circuits in place, the testing was started. The large DC motor was run from 5-25 Hz shaft frequency and used to spin the smaller NEMA 23 motors. Theoretically, the maximum power from the motors should have been generated when RLoad was set equal to the resistance of the motor windings. Then, the output power of the motors at each frequency would be measured from the resistance of RLoad from Figures 5 and 6 and the voltage across the resistors. Therefore,

$$I = \frac{V}{R}$$
$$P = IV$$
$$P = \frac{V^2}{R}$$

According to this method, the resistances for the three stepper motors, 5718-02S, 5709M-02S, and 5704M-02, and the BLDC motor, BLDC23F-23, should have been, respectively, 2.4, 2.1, 3.6, and 1.5 Ω. However, because of the presence of the diode rectifier and the capacitor bank, the actual resistance of the load to optimize power output varied. Because of this, a variable resistor was manually adjusted while monitoring power output to find the maximum power generated. The output voltage across the resistor was measured using a four channel oscilloscope. The output current through the resistor was connected through a current probe to a second channel of the oscilloscope. Finally, a third channel was used to monitor the frequencies at which the stepper motors were run by

connecting the leads across the AC output voltage of the motors. The frequency at which the motors rotated was calculated, in Hz, by the following equation:

$$frequency = \frac{AC\ output\ frequency}{\#\ of\ pole\ pairs}$$

Using the math functions of the oscilloscope, the output power was calculated by multiplying the output voltage and current together to produce another waveform. This was monitored in real time as the resistance was varied to yield the maximum output power.

RESULTS

All results were preliminary as of 18 July 2008. After putting together the experimental setup, testing was completed. While the motors were running, the current and voltage running through the load resistor were monitored on an oscilloscope. This yielded a real time display of the power generated using the math functions of the scope. After adjusting the resistance of the variable resistor to maximize the power, the power was recorded and graphed using Microsoft Excel. Figure 8 shows a graph of the power harvested for the frequency range of 5-25 Hz normalized to the masses of 0.68 kg for the three stepper motors and 0.64 kg for the BLDC motor.

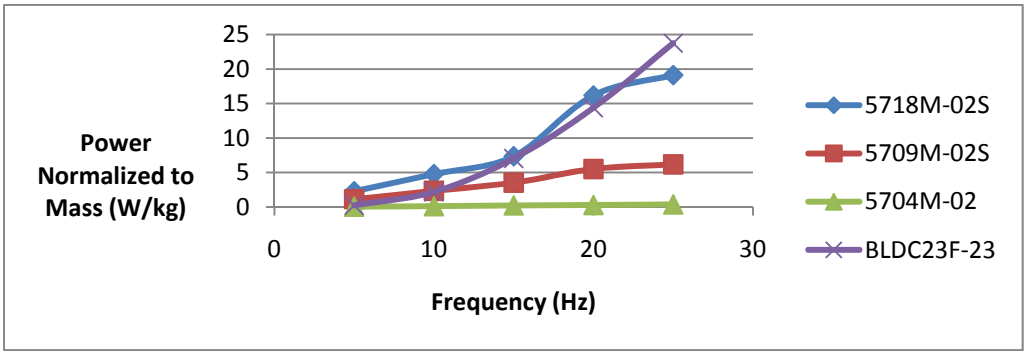


Figure 8: Power harvested for different motors

DISCUSSION

The results of the tests were indicative of the usefulness of the motors. The 5704M-02 and 5709M-02S stepper motors generated relatively low power relative to the 5709M-18S and BLDC23F-23 motors. This can be explained from analysis of the number of pole pairs in each stepper motor. As the limit of the number of pole pairs in the stepper motor increases to infinity, the rotor becomes closer and closer to a smooth cylinder. A smooth cylinder would experience no torque in a motor and therefore generate no power when operated in reverse as a generator. This is consistent with the data collected because as the number of pole pairs increases, the power harvested decreases. Therefore, the 5704M-02S

and 5709M-02S motors would not be the best motors to be used in energy harvesting applications. These motors would be more suitable for use as motors in positioning systems, for example, in printers or fax machines, where very precise movement is necessary.

The 5718M-02S and BLDC23F-23 motors performed more favorably for energy harvesting. However, the applications of the two motors should be completely different from each other. The range of power generated by the BLDC motor was 150 mW to 15.2 W from 5 to 25 Hz, while the range of power generated by the stepper motor was 1.55 to 13 W. Therefore, for low frequency applications, the stepper motor would be the best option to harvest the most power by a significant margin, over ten times as much as the BLDC motor at similar frequencies. However, for higher frequency applications, the BLDC gives a slight improvement for harvesting power.

CONCLUSION

This experiment was performed in order to find the best type of motor with the options of a stepper and BLDC motor for use as a generator in energy harvesting applications. The results indicated that not only was the type important, but also the characteristics of the individual motors was important. For low frequency, ~5 Hz, applications of the motors, stepper motors with large step angles proved to be the most effective collectors of energy. For higher frequencies, >20 Hz, the BLDC motor was the best candidate for collecting energy. However, since most applications in which ambient energy would be harvested deal with low frequencies, the stepper motor, in this case, 5718M-02S, is the best motor for energy harvesting. This motor could see applications in gyms in cardiovascular workout equipment such as rowing machines or stationary bicycles. It could also be coupled to the axis of doors to power small electronic devices in stores or businesses.

ACKNOWLEDGEMENTS

Nicholas Czarnek (NC) would like to thank Dr. Heath Hofmann, his faculty advisor, for his help in suggesting this project, explaining the physics behind it, helping to build the setup required for testing, and revising this paper. Also, NC thanks Junyu Yao, his graduate student advisor, for his help in acquiring data during testing. Finally, NC would like to thank his father, Dr. Robert Czarnek, for his constant support and help in overcoming many obstacles with the electronics of this experiment. This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081.

REFERENCES

1. Blechman, Adam, George Braker, Brad Chodnicki, Esther DuBow, Kelly Pernia, Timothy Sy, Matthew Thompson, Jordan Tucker. "A study of the benefits of retrofitting cardiovascular exercise equipment of a gym with human energy

harvesting technology." Accessed 25 Jun 2009. <http://hdl.handle.net/1903/9083>. Gemstone Team Research Collections. University of Maryland. May 2009

2. Henderson, Tessa. "Power harvested from gym workouts." *Energy Harvesting Journal*. USA. 26 Mar 2009. Accessed 25 Jun 2009. <http://www.energyharvestingjournal.com/articles/power-harvested-from-gym-workouts-00001310.asp?sessionid=1>.
3. Saez, Maria Loreto Mateu. "Energy Harvesting from Passive Human Power." January 2004. PhD Thesis. Technical University of Catalonia. Accessed June 2009.
4. Sloan, Irvine. "Brushless DC Motors in Medical Applications." *MachineDesign.com* 2 June 2009. Accessed 19 Jul 2009. <http://machinedesign.com/article/brushless-dc-motors-in-medical-applications-0602>.
5. *Bodine Electric Handbook*. 5th Edition. Chicago, IL: Bodine Electric Company, Chicago, IL, 2009.
6. Lefebvre, Mike. "Dynamics of Hybrid Stepping Motors." *MachineDesign.com*. 17 Feb 2005. 25 Jun 2009 <http://machinedesign.com/article/dynamics-of-hybrid-stepper-motors-0217>.
7. Fitzgerald, A.E., Charles Kingsley Jr., Stephn D. Umans. *Electric Machinery*. 5th. USA: McGraw-Hill, Inc., USA, 1990.
8. Geiger, Richard, Heath Hofmann. "Energy Harvesting Using AC Machines with High Effective Pole Count." Published 15-19 June 2008. Power Electronics Specialists Conference, 2008. Olive Roads, Greece. IEEE. pages 2229-2234. Accessed 5 June 2009

COMPARSION OF PREDICTED, SIMULATED, AND EXPERIMENTALLY MEASURED CONVERSION GAIN IN A ROBINSON MARGINAL OSCILLATOR

Devon Miller,[*] Jeffrey L. Schiano,[#] and Tom Tyson[+]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

[*]Undergraduate student of
Department of Physics and Engineering
West Virginia Wesleyan College
Buckhannon, WV 26201

ABSTRACT

A Robinson marginal oscillator (RMO) is an instrument for measuring small changes in the losses within a harmonic oscillator. The key figure of merit for a RMO is the conversion gain, which is the sensitivity of the oscillation amplitude to changes in the losses. The characteristics of the memoryless nonlinear element that limits the amplitude of the oscillation determine the conversion gain. Viswanathan[1] provides an analytical method for estimating the conversion gain given an input/output description of the nonlinear element. This study compares the estimated conversion gain against that observed in numerical simulations and experimental measurements, and provides a framework for testing alternative analytical methods for predicting the conversion gain.

INTRODUCTION

Marginal Oscillator

Articles by Roberts[2] in 1947 and Rollins[3] in 1948 credit Pound with developing the marginal oscillator for detecting the absorption of radio frequency (RF) energy by the nuclei of a solid material.[4] By measuring the change in absorption as a function of frequency, Pound was able to investigate the molecular structure of the solid. The use of marginal oscillators has since extended to ion cyclotron resonance spectroscopy,[5] monitoring mechanical properties of thin-films,[6] the characterization of defects in silicon,[7] and the measurement of penetration depth in superconductors.[8]

[#] Faculty Mentor
[+] Graduate Mentor

Figure 1 represents the marginal oscillator as a parallel resistor-inductor-capacitor (RLC) circuit driven by a dependent current source. The resistance *R* represents the electrical losses within the circuit. The voltage *v* across the parallel RLC circuit determines the output current *i* of the dependent current source, and this feedback path enables the circuit to oscillate at a frequency determined by the values of *L* and *C*. The oscillator is said to be *marginal* because the gain of the feedback path is made just large enough to sustain oscillation. By appropriately choosing the function $i=G_i(v)$, small changes in the losses *R* are revealed by proportionately larger changes in the amplitude of oscillation.

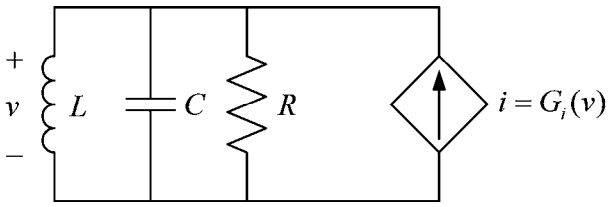


Figure 1: Marginal oscillator

The dependent current source is also known as a limiter, because it reduces the gain between the voltage input and current output as the amplitude of the input voltage increases. Figure 2 shows a limiter function $G_i(v)$ that is piecewise linear with two symmetrical limiting segments with slope g_B and a symmetrical amplifying segment with slope g_A. When the absolute value of the input voltage exceeds a threshold value v_T, the slope of the limiter characteristic decreases from g_A to g_B. Because the limiter does not satisfy the homogeneity property, that is, the ratio of the limiter output to the limiter input is not constant, the limiter is a nonlinear system.

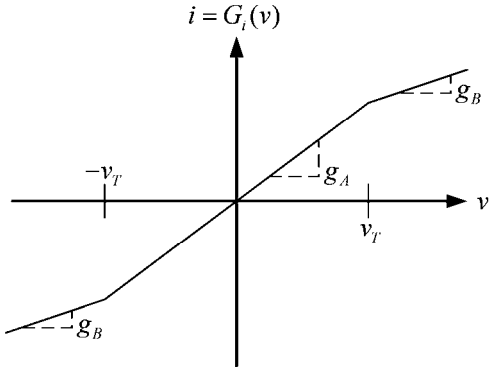


Figure 2: Piecewise linear limiter characteristic

Early limiter designs rely on a single active device to amplify and limit the input signal. By using separate devices for amplification and limiting, Robinson demonstrated a significant improvement in the signal-to-noise (SNR) ratio of

oscillator amplitude measurements.[9] A marginal oscillator that uses this approach is known as a Robinson marginal oscillator (RMO).

Using a method proposed by Rychak[10] to estimate conversion gain requires characterization of the parallel RLC circuit in terms of its natural frequency

$$\omega_n = \frac{1}{\sqrt{LC}}, \tag{1}$$

and its quality-factor (Q-factor)

$$Q = \frac{R}{\omega_n L}. \tag{2}$$

The Q-factor is defined as the ratio of the average energy stored in a system per cycle to the energy converted to heat by losses within the system. For the parallel RLC circuit, it can be shown that the Q-factor is the ratio of the natural frequency to the 3-dB bandwidth of the transfer function relating the input current *i* to the output voltage *v*. In general, a high Q-factor is desirable because the measurement SNR of oscillator amplitude is proportional to the square-root of the Q-factor.[11]

Conversion Gain

The key figure merit of a RMO is the ratio of the percent change in oscillator amplitude *A* to the percent change in electrical losses *R*. This ratio is known as the sensitivity of *A* with respect to *R*, and is denoted as S_R^A. To avoid confusing the sensitivity S_R^A with detection sensitivity, which is the smallest change in amplitude that can be observed in the presence of thermal noise, Viswanathan[1] relabeled S_R^A as the conversion gain

$$G_c = S_R^A = \frac{\%\text{ change in A}}{\%\text{ change in R}} = \frac{\Delta A / A}{\Delta R / R}. \tag{3}$$

For a parallel RLC circuit driven by an independent current source at the natural frequency, it can be shown that $S_R^A = 1$. A RMO uses a nonlinear limiter characteristic *G(v)* to achieve conversion gains significantly larger then unity.

A large conversion gain is desirable as it increases the measurement SNR oscillator amplitude. For example, in quadrupole resonance (QR) spectroscopy the value of *R* is typically 1 Ω, while Δ*R* is on the order of 1 μΩ . For nominal oscillation amplitude *A* of 1 V, the corresponding change in amplitude Δ*A* is 1 μV. Because the amplitude change is so small, it is observed using lock-in detection that improves the measurement SNR at the expense of increasing detection time. As the measurement SNR scales directly with conversion gain,

increasing the conversion gain allows the experimenter to trade between improved measurement SNR and shorter detection times.

Deriving an expression for the conversion gain is complicated by the fact that the limiter is a nonlinear system. Using an approximation to the solution of a nonlinear second-order system developed by Kryloff and Bogoliuboff,[12] Viswanathan derived two expressions that can be solved for the amplitude of oscillation and the conversion gain

$$\sin 2\theta + 2\theta = \frac{\pi}{\left(1 - \frac{g_B}{g_A}\right)}\left[\frac{1}{g_A R} - \frac{g_B}{g_A}\right], \ \theta \in \left[0, \frac{\pi}{2}\right] \tag{4}$$

$$G_C = S_R^A = \frac{\pi}{2\left(1 - \frac{g_B}{g_A}\right) g_A R \sin 2\theta}, \tag{5}$$

where

$$\sin\theta = v_T / A\,. \tag{6}$$

Given the numeric value of g_A, g_B, and v_T, Equation (4) is solved for θ, and the result is used in Equations (5) and (6) to obtain the conversion gain G_c and the amplitude A of oscillation, respectively.

Specific Aims

Viswanathan's method for estimating the conversion gain relies on approximating the solution of a nonlinear second-order differential equation. The central question that this study addresses is whether or not the estimated conversion gain is accurate. To this end, for a specific experimental RMO test bed, the conversion gain predicted by Viswanathan's approach is compared against the conversion gain observed in numerical simulations and experimental measurements. This work also provides a framework for testing new methods of analytically predicting the conversion gain, and a means for verifying that the desired limiter parameters g_A, g_B, and v_T are correctly set in an experimental system.

METHODS

RMO Test Bed

In order to facilitate characterization of the RLC circuit and limiter system, these subsystems are implemented on separate circuit boards shown in Figure 3. The leftmost board contains a voltage follower for observing the voltage across the RLC circuit and an envelope detector for monitoring the oscillator amplitude. The inductor is located in a separate shielded enclosure and connects to the SMA

jack located at the top of the printed circuit board. The rightmost circuit board implements the limiter system and the circuits on the two boards are connected by semi-flexible coaxial lines. The oscilloscope probes in Figure 3 are for monitoring the voltage at the limiter input and output. The limiter parameters g_A, g_B, and v_T are estimated from the observed waveforms.

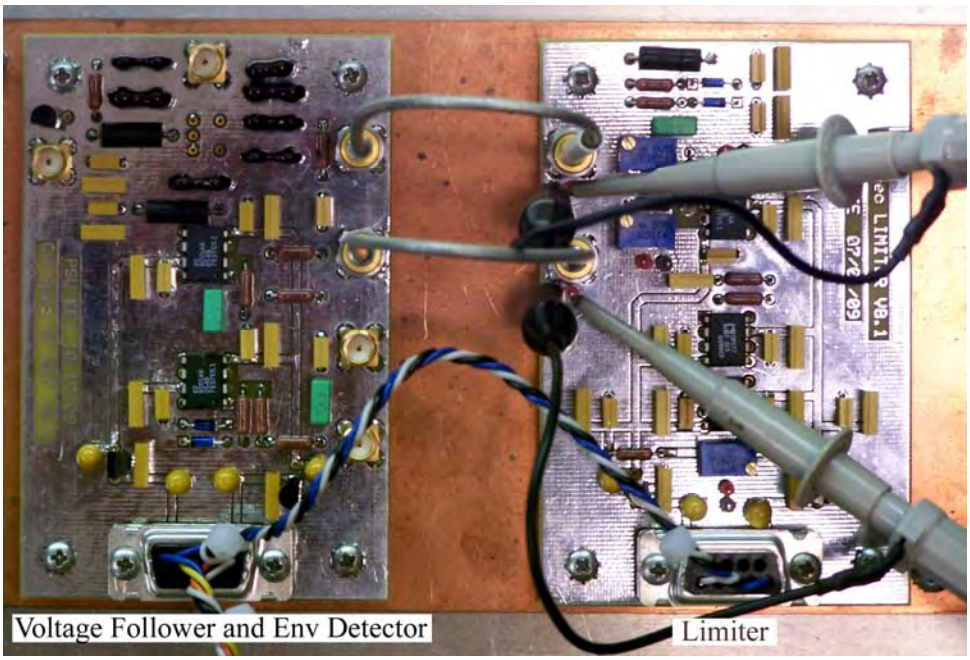


Figure 3: RMO test bed

Figures 4 and 5 show the schematic representation of the voltage follower/envelope detector and limiter circuits, respectively. Murphy provides a detailed discussion of the circuit operation.[13] The capacitor and inductor values are chosen so that the nominal frequency of oscillation is 3.608 MHz. For this study, the key detail is that separate potentiometers located in the limiter circuit allow the experimenter to set the parameters g_A, g_B, and v_T. Figure 4 also shows a calibration circuit for measuring conversion gain.[10] When the MOSFET is turned on, the electrical losses within the RLC are increased.

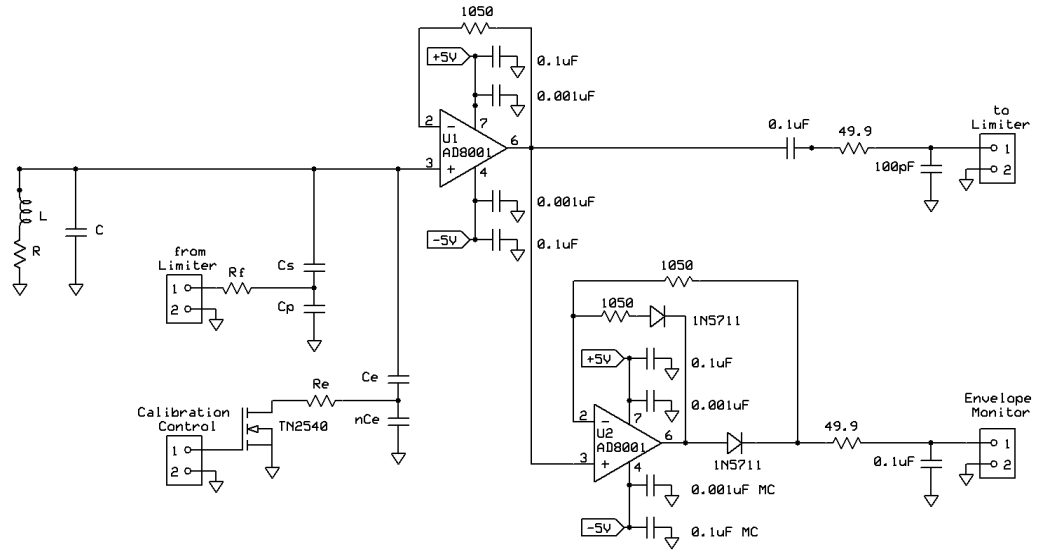


Figure 4: Voltage follower and envelope detector

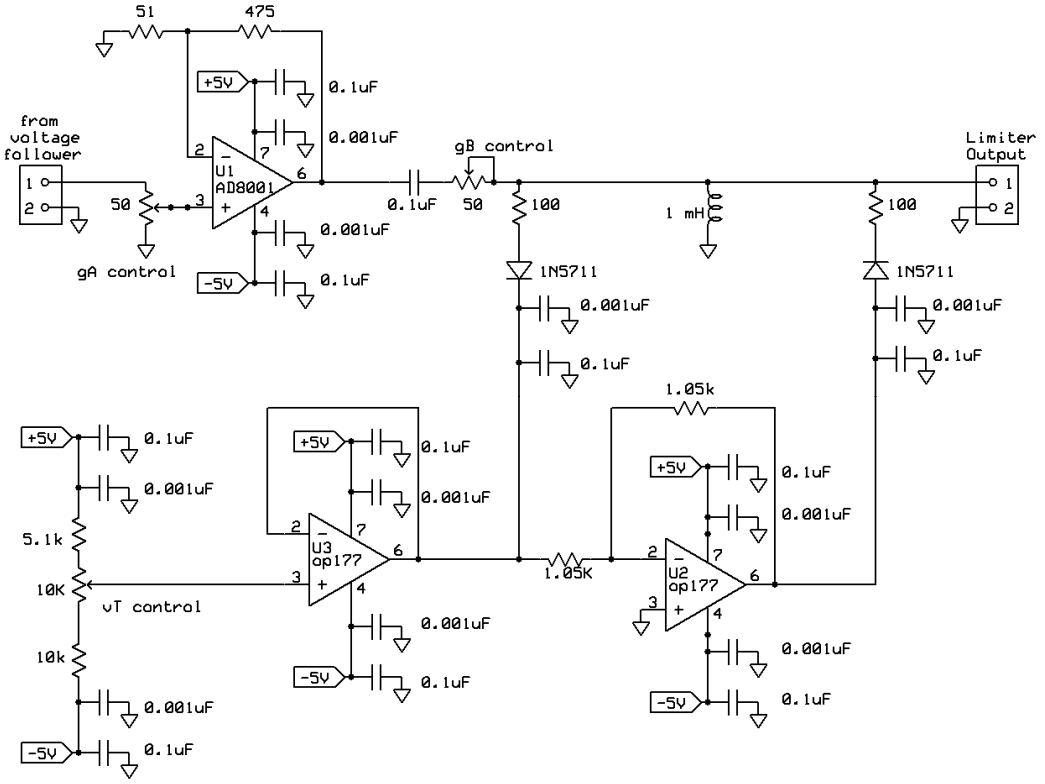


Figure 5: Limiter circuit

In order to apply Viswanathan's method for estimating the conversion gain, the RMO system must be represented by the parallel RLC circuit and dependent current source in Figure 1. Figure 6 shows a simplified circuit representation of

the test bed RMO. The limiter circuit is represented by the dependent voltage source at the far right of the diagram. Except for the inductor, the remaining components are located on the voltage follower/envelope detector circuit board. Depending on the state of the MOSFET switch in the calibration circuit, the overall dynamic model of the RMO test bed is either sixth-order when the switch is turned off or fifth-order when the switch is turned on.

An approximate second order model for the RMO test bed is obtained in three steps. The first step is to reduce the limiter circuit in the rightmost dashed box to its Norton equivalent circuit in the lower diagram. This transformation is valid as long as the operating frequency is greater than $1/R_f(C_s + C_p)$. The second step is to replace the RLC circuit in the upper diagram with the RLC circuit in the lower diagram. The two RLC circuits behave identically near the natural frequency as long as the Q-factor is much larger than unity. The third step is to represent the calibration circuit by the parallel RC circuit in the lower diagram.[10] This transformation holds for values of n greater than ten. Regardless of whether or not the MOSFET is turned on, the calibration circuit places a small capacitance C_e in parallel with C. When the MOSFET is turned on, an effective resistance n^2R_e is placed in parallel with the RLC circuit, thereby increasing electrical losses. The parallel combination of the resistances and capacitances in the lower circuit result in a second-order model equivalent to that shown in Figure 1.

Limiter Parameter Measurements

As a means of verifying the operation of the limiter and obtaining a measurement of the limiter parameters g_A, g_B, and v_T, a LabVIEW program was written that acquires the waveforms present at the limiter input and output using an oscilloscope, measures the DC bias voltage applied to the diodes in the limiter circuit, and records the frequency of oscillation using a frequency counter. Figure 7 shows the front panel of the LabVIEW program. The graphs show the limiter input and output over several cycles. The measured frequency and DC bias are shown on the right side of the panel along with a least-squares estimate of the limiter parameters. Also shown are the oscillation amplitude derived from the limiter input waveform and an estimation of θ defined in Equation (6).

The original goal was to estimate the limiter parameters while the RMO is operating. However, because the fraction of each cycle that exceeds the threshold value v_T is small, there is insufficient number of data points for estimating the slope g_B of the limiting segments in Figure 2. For example, the leftmost plot in Figure 8 shows a plot of the limiter input versus the limiter output while the RMO is operating. The region corresponding to the amplifying segment is bounded by vertical lines. There are significantly less points appearing on the limiting segments. In order to obtain a good estimate of the slope g_B, the limiter is driven by an external function generator whose output amplitude is chosen larger than the threshold voltage v_T. For this case, the rightmost plot in Figure 8 shows that there are many data points along the limiting segments. For this reason, data for

estimating the limiter parameters is obtained by driving the limiter with an external function generator.

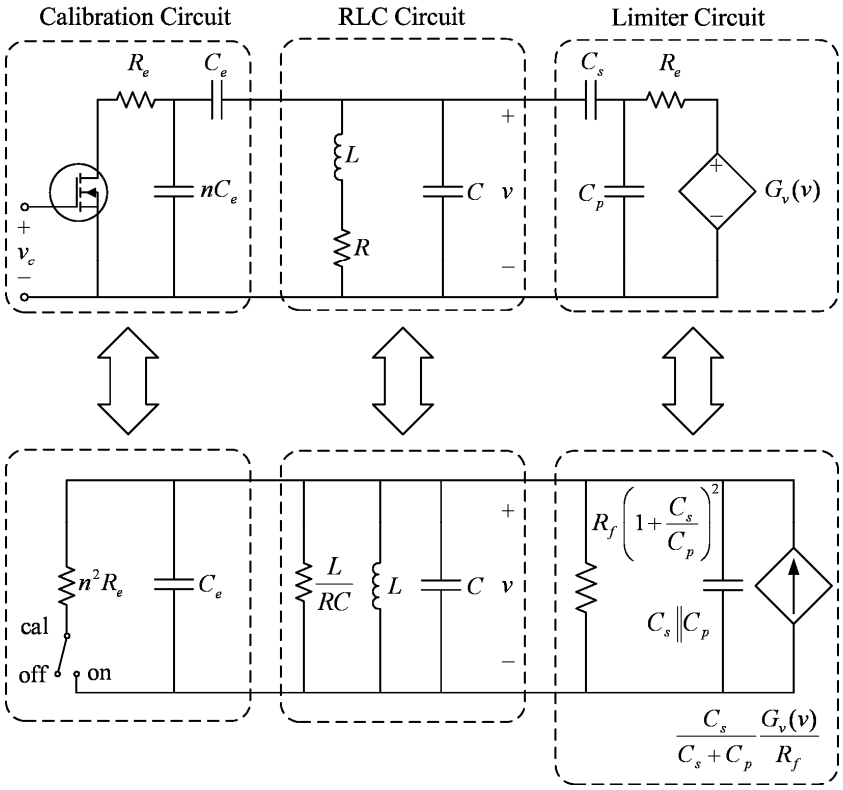


Figure 6: Representation of the RMO test bed as a second order system

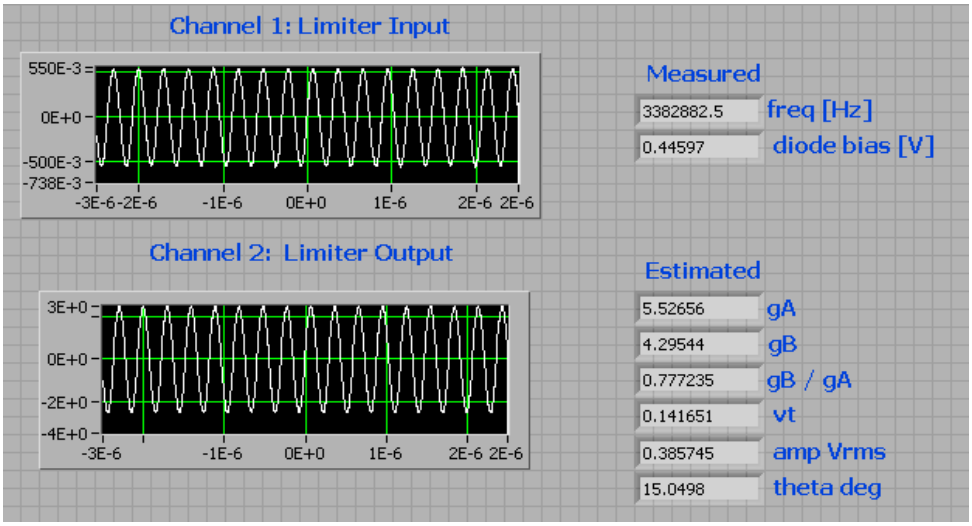


Figure 7: LabVIEW front panel

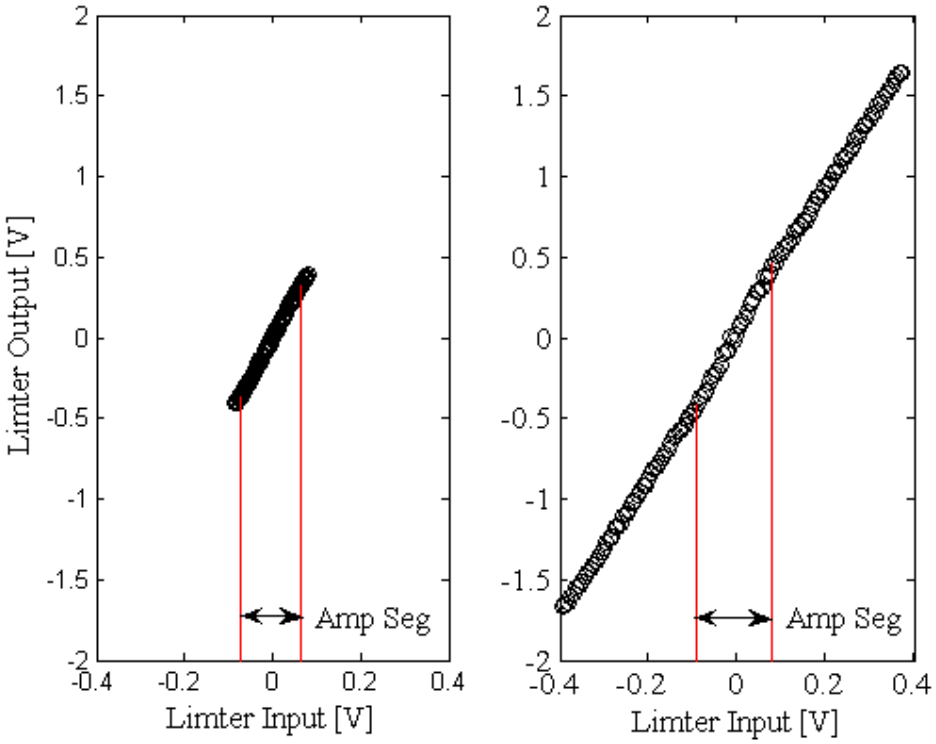


Figure 8: Limiter input/output characteristics obtained while the RMO is operating (left), and when the limiter is driven by an external function generator (right)

Measuring Conversion Gain

While a change in oscillation amplitude is easily observed using an oscilloscope, measurement of the change of resistance caused by the calibration circuit is difficult. For this reason, it is useful to express the conversion gain as

$$G_C = S_R^A = S_Q^A S_R^Q \ . \tag{7}$$

For a parallel RLC circuit, S_R^Q is unity at the natural frequency, while the sensitivity S_Q^A is estimated as

$$S_Q^A = \frac{\Delta A / A}{\Delta Q / Q} = \frac{\dfrac{A_{\text{cal off}} - A_{\text{cal on}}}{A_{\text{cal off}}}}{\dfrac{Q_{\text{cal off}} - Q_{\text{cal on}}}{Q_{\text{cal off}}}} \ . \tag{8}$$

The change in oscillator amplitude resulting from turning the calibration circuit on and off is observed using an oscilloscope, while the change in Q-factor is

observed using a vector network analyzer (VNA). The latter measurement is made by first disconnecting the limiter circuit. The output of the VNA is loosely coupled to the inductor using a small loop, and the output of the voltage follower is connected to the input of the VNA. Measurement of the magnitude of the S_{21} scattering parameter reveals the frequency response of the RLC circuit, and hence its Q-factor. In order to measure the small difference in Q-factor when the calibration circuit is turned on and off, the span of the VNA was set to the smallest value that would permit observation of both 3 dB points and the number of measurements points was set to its maximum value. It should be noted that the latter measurement is only made once for each state of the calibration circuit as the limiter parameters do not affect the value of $\Delta Q/Q$.

In the experiments, the slope g_A of the amplifying segment of the limiter characteristic and the threshold voltage v_T are held fixed. The range of g_B is chosen so that expected conversion gain varies from about 10 to 50 in nine steps. For each set of limiter parameters, the oscillator amplitude is measured for the calibration circuit turned on and off. At the end of the experiment, the Q-factor was recorded for the two states of the calibration circuit.

Simulating Conversion Gain

Numerical simulation of the circuit in Figure 1 provides a means of determining the oscillation amplitude A for a given value of R. To mimic the RMO test bed, the value of L is chosen as the value of inductance used in the experiments. The value of C is chosen so that the natural frequency of the circuit in Figure 1 matches the frequency of oscillation observed in the experiments. Two values of R are used in the simulations, $R_{\text{cal off}}$ and $R_{\text{cal on}}$, and these values are chosen using equation (2) so that the corresponding Q-factors matched the experimentally measured values of $Q_{\text{cal off}}$ and $Q_{\text{cal on}}$. The limiter parameters g_A, g_B, and v_T are set to the values estimated for the RMO test bed.

Numerical simulations are performed in MATLAB using the solver ODE113. The initial capacitor voltage is set to 90% of the threshold value and the simulation duration is chosen long enough so that the oscillator amplitude reaches a steady-state value. For each set of limiter parameters, the RMO is simulated for both values of R and the conversion gain is calculated as

$$Gc = S_R^A = \frac{\Delta A / A}{\Delta R / R} = \frac{\dfrac{A_{\text{cal off}} - A_{\text{cal on}}}{A_{\text{cal off}}}}{\dfrac{R_{\text{cal off}} - R_{\text{cal on}}}{R_{\text{cal off}}}}, \tag{8}$$

where $A_{\text{cal off}}$ and $A_{\text{cal on}}$ are the steady-state oscillation amplitudes observed for the resistances $R_{\text{cal off}}$ and $R_{\text{cal on}}$, respectively.

RESULTS

For the RMO test bed, the parameter g_A is fixed and its estimated value is 5.11 V/V. The threshold voltage v_T is also fixed and its estimated value is 76.6 mV. The nominal frequency of oscillation is approximately 3.308MHz. The Q-factor of the RLC circuit is 125.21 when the calibration circuit is turned off, and 124.15 when the calibration circuit is turned on. The resulting change in Q-factor between the calibration states is $\Delta Q/Q$ = 0.00847 or 0.847%. The measured value of the inductance L is approximately 22 μH.

Key results from the experimental measurements, numerical simulations, and predications based on Viswanathan's method are summarized in Table I.The first column shows the nine measured values of g_B/g_A used in the RMO test bed. The second and third columns show the corresponding amplitude of the oscillation measured for the calibration circuit turned on and off, respectively. Using the measured value of the operating frequency, inductance L, Q-factors $Q_{\text{cal off}}$ and $Q_{\text{cal on}}$, and estimated limiter parameters g_A, g_B, and v_T, numerical simulations were performed to determine the oscillation amplitude for both states of the calibration circuit. The fourth and fifth columns of Table I show the oscillation amplitude obtained from numerical simulations for the calibrator circuit turned on and off, respectively. Using the measured value of the limiter parameters g_A, g_B, and v_T, and the values of $R_{\text{cal off}}$ and $R_{\text{cal on}}$ estimated from the measured Q-factors $Q_{\text{cal off}}$ and $Q_{\text{cal on}}$, Equations (4) and (5) were used to obtain a theoretical estimation of the conversion gain for the two states of the calibration circuit. These estimates appear in the sixth and seventh column of Table I, while the average of the two conversions gains appear in the eighth column. The ninth column shows the conversion gain obtained using numerical simulations and Equation (8), while the tenth column shows the conversion gain obtained from experimental data using Equation (8).

Table I. Summary of results

	Amplitude [mVrms]				Conversion Gain				
Exp	Exp		Sim		Theory			Sim	Exp.
g_B/g_A	Cal. On	Cal. Off	Cal. On	Cal Off	Cal. On	Cal. Off	Avg.		
0.928	99.2	166	113	165	33.0	42.4	37.7	37.6	47.5
0.902	81.3	103	85.1	105	17.1	18.4	17.8	22.7	24.9
0.891	64.1	75.2	80.3	93.7	13.6	14.0	13.8	16.9	17.4
0.879	59.5	67.8	80.0	86.1	11.2	11.2	11.2	8.47	14.5
0.863	57.8	64.7	77.8	83.3	10.2	10.1	10.1	7.80	12.6
0.842	56.1	61.9	75.3	78.9	8.95	8.81	8.88	5.36	11.1
0.818	52.8	57.9	74.6	79.2	8.68	8.52	8.60	6.93	10.4
0.805	50.6	55.1	74.1	78.9	8.59	8.43	8.51	7.22	9.65
0.799	50.8	54.8	73.4	78.2	8.03	7.84	7.93	7.23	8.62

Figure 9 provides a plot of the conversion as a function of the ratio g_B/g_A. Each square symbol represents the average of the conversion gains predicted by Viswanathan's method for the calibration circuit turned on and off. The open circles represent the conversion gain obtained from numerical simulations, while the open diamonds show the conversion gain estimated from the measured oscillation amplitude in the RMO test bed.

DISCUSSION

The results in Figure 9 answer the central question posed by this study. Viswanathan's method for predicting conversion gain yields results that are reasonably close to the conversion gains predicted by numerical simulations and measured in the laboratory. As important, the results demonstrate that is possible to obtain conversion gains significantly larger than unity.

The results in Figure 9 indicate that as the ratio g_B/g_A approaches unity, the RMO test bed yields a conversion gain that is larger than what is predicted by either theory or numerical simulations. This discrepancy may result from the fact that both the theoretical predictions and numerical simulations assume the piecewise linear characteristic shown in Figure 2. In comparison, the limiter characteristic of the RMO test bed is a smooth polynomial whose order is greater than one because of the properties of the diodes used in the limiter circuit. Another potential explanation is that Viswanathan's method is applicable to second-order systems, while the order of the RMO test bed is several degrees higher.

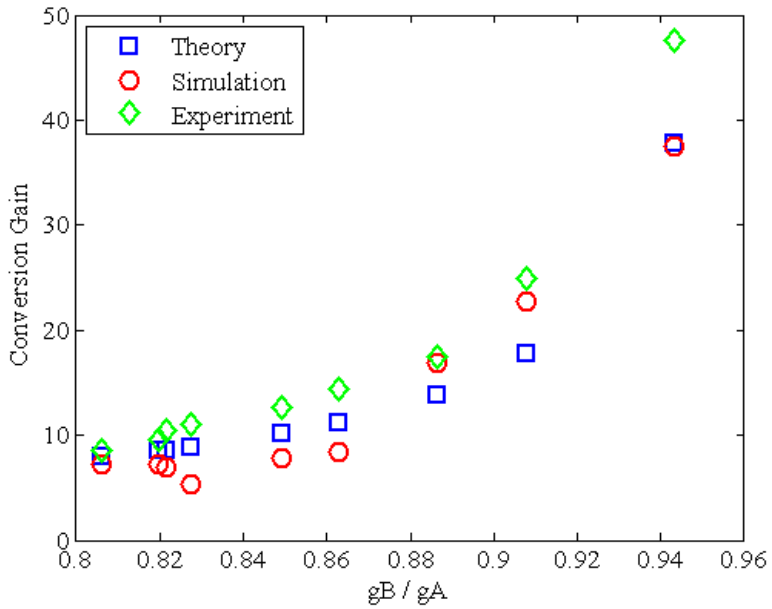


Figure 9: Conversion gain obtained from theoretical predictions, numerical simulations, and experimental measurements

Ongoing work is directed along three fronts. The first direction is to develop an analytical method for predicting conversion gain that does not limit the order of the system to two. The second direction is to find a method for determining the limiter characteristic that achieves specified performance criteria such as conversion gain and amplitude of oscillation. The third direction is to design a RMO test bed that can implement an arbitrary limiter characteristic.

CONCLUSION

This study verifies that Viswanathan's method provides an accurate estimate of conversion gain in a RMO. Experimental results demonstrate that it is possible to achieve conversion gains greater than ten. Further work is needed to understand why the conversion gain measured in the RMO test beds exceeds that predicted by theory and numerical simulations when the conversion gain becomes greater than ten.

ACNOWLEDGEMENTS

I would like to thank both Dr. Jeff Schiano and Tom Tyson for their tremendous help throughout the project. Also would like to commend both on their late night availability. I would also like to thank Penn State University for giving me the opportunity to take part in the EEREU this summer. This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081.

REFERENCES

[1] T.L. Viswanathan, T.R. Viswanathan, K.V Sane, "Study of Marginal Oscillator Behavior," *IEEE Transactions on Instrumentation and Measurement*, **IM24** (1) 55-61 (1975).

[2] A. Roberts, "Two New Methods for Detecting Nuclear Radio Frequency Absorption," *The Review of Scientific Instruments*, **18** (11) 845-848 (1947).

[3] B.V. Rollin, "Nuclear Paramagnetism," *Reports on Progress in Physics*, **12** (1) 22-33 (1949).

[4] R.V. Pound and W.D. Night, "A Radio Frequency Spectrograph and Simple Magnetic-Field Meter," *The Review of Scientific Instruments*, **21** (3) 219-225 (1950).

[5] R.T. McIver, "A Solid-State Marginal Oscillator for Pulsed Ion Cyclotron Resonance Spectroscopy," *The Review of Scientific Instruments*, **44** (8) 1071-1074 (1973).

[6] G.L. Miller, M. Soni, R.L. Fenstermacher, "A Technique for Investigating the Properties of Surfaces, Thin Films, and Interfaces by Means of a Mechanical Marginal Oscillator," *Journal of Applied Physics*, **53**(2) 979-983 (1982).

[7] D. Alexiev, M.I. Reinhard, L. Mo, A. Rosenfeld, "A Transient Conductance Technique for Characterization of Deep-Level Defects in Highly Irradiated Detector-Grade Silicon," *Nuclear Instruments and Methods in Physics Research Section A*, **434**(1) 103-113 (1999).

[8] A. Gauzzi, J. Le Cochec, G. Lamura, B.J. Jönsson, V.A. Gasparov, F.R. Ladan, B. Plaçais, P.A. Probst, P. Pavuna, J. Bok, "Very High Resolution Measurement of the Penetration Depth of Superconductors by a Novel Single-Coil Inductance Technique," *The Review of Scientific Instruments*, **71**(5) 2147-2153 (2000).

[9] F.N.H. Robinson, "Nuclear Resonance Absorption Circuit," *Journal of Scientific Instruments*, **36**(12) 481-487 (1959).

[10] R. Rychak. "Method and Instrumentation for Determining the Sensitivity of a Robinson Marginal Oscillator," Honor's Thesis, The Pennsylvania State University, 2009.

[11] C.D. Motchenbacher and J.A. Connelly, *Low-Noise Electronic System Design*, John Wiley and Sons, New York, 1993.

[12] N. Kryloff and N. Bogoliuboff, *Introduction to Non-linear Mechanics*, Kraus Reprint Company, New York, 1970.

[13] M. Murphy and J.L. Schiano, "Evaluation of Circuit-Blocks for a Robinson Marginal Oscillator," *NSF EE REU PENN STATE Annual Research Journal*, **6**, Edited by J. Mitchell and R. Guo, 117-131, (2008).

TRANSFER OF SILICON WIRES ONTO GLASS SUBSTRATES FOR USE IN RADIAL SOLAR CELLS

Rachel Christian*, Chito Kendrick†, and Joan Redwing‡

Department of Materials Science
The Pennsylvania State University, University Park, PA 16802

*Undergraduate student of
Department of Electrical and Computer Engineering
Binghamton University
Binghamton, NY 13902

ABSTRACT

The high aspect ratio semiconductor heterojunction (HARSH) solar cell project uses the vapor liquid solid growth process to grow arrays of silicon wires. These wires are used to fabricate radial silicon solar cells which have the potential to be more efficient than traditional solar cells. Research is being done into methods that would make implementing this technology cheaper. One such method would be for the silicon wires to be transferred from the silicon growth substrate onto a silicon coated glass substrate; allowing the silicon substrate to be reused as a base for growing more silicon wires. The glass substrate is cheaper to produce than the required silicon wafers, while still allowing support for the wire arrays for the rest of the solar cell fabrication. Two different methods for transfer have been attempted. The first method entailed encasing the wires in a layer of polymer, removing them from the growth substrate, and then annealing them to the new substrate. The second was a direct growth from the substrate onto the alternative substrate. The first method has so far been unsuccessful; however, the second method has shown some promising preliminary results.

---

†Graduate Mentor
‡Faculty Mentor

INTRODUCTION

Silicon solar cells make up the majority of the solar cell market. However, their production cost has ensured that they are not a dominatnt source of energy and is why they have yet to replace fossil fuel based energy production. The main cost issue is the actual silicon wafer, which is about 65% of the total cost for a silicon solar cell.[1] An alternative structural design to a planar solar cell is one that has a radial geometry. Planar silicon solar cells require extremely pure silicon, to ensure long minority carrier diffusion lengths in order to perform at reasonable efficiencies, while they need to be thick enough to ensure maximum photon absorption. Whereas a radial geometry silicon solar cells can cope with minority carriers with short diffusion lengths, due to the short radial direction, while maintaining maximum absorption of the incident photons; thereby decoupling the two mechanisms that are required for a solar cell to function (Fig. 1).[2] Device physics modeling indicates that solar cells made with impure and thus less expensive silicon wire arrays have the potential to match the efficiencies of planar solar cells made with high purity silicon.[2]

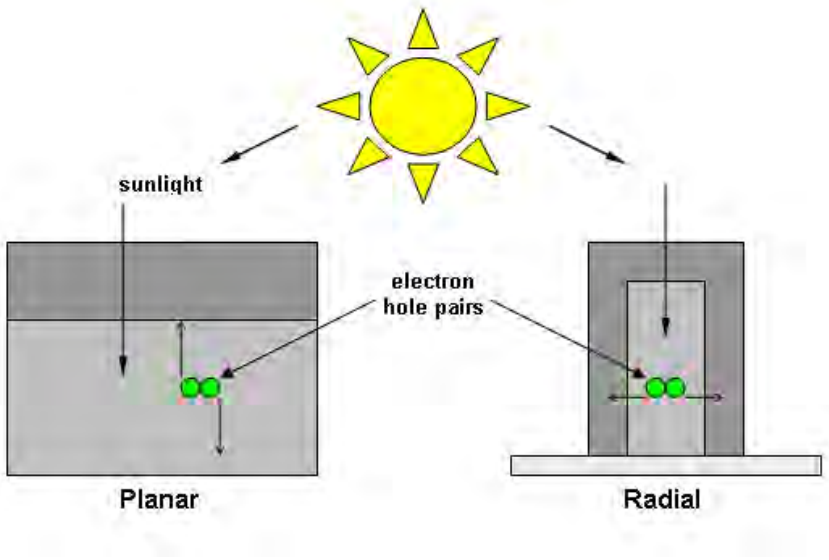


Figure 1: Comparison of planar and radial solar cell geometry

There are several fabrication techniques to produce silicon wires. An approach that is being explored is the bottom-up Vapor-Liquid-Solid Mechanism (VLS).[3] The VLS growth process uses a metal catalyst, typically gold, to drive the wire growth (see Fig. 2). Gold is first deposited onto a silicon wafer and heated, thereby forming an Au-Si liquid phase alloy. At the growth temperature, silicon tetrachloride ($SiCl_4$) is introduced and decomposes before being incorporated into the alloy.

Once the liquid becomes supersaturated with silicon from the vapor, a silicon wire starts to grow.[3] The VLS growth process allows for direct control over the position and size of the wire, through placement of the gold seed and diameter of the Au-Si droplet respectively.

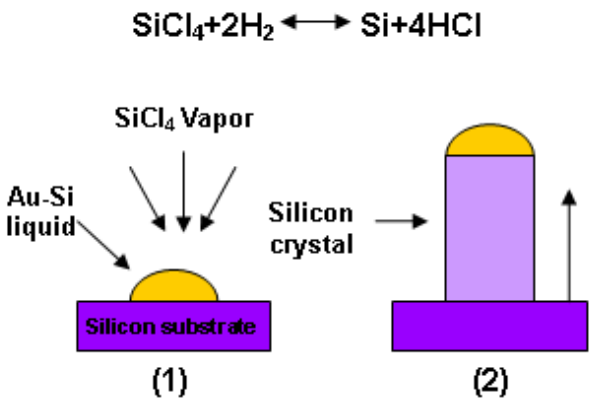


Figure 2: VLS growth mechanism for growing silicon wire arrays

The purpose of this project is to investigate ways of transferring silicon wire arrays onto alternative substrates, ideally glass, in order to further reduce fabrication costs of radial silicon solar cells. Two approaches have been considered. The first relies on a method put forth by researchers at the California Institute of Technology who are working toward developing a flexible solar cell using silicon wire arrays embedded within the polymer polydimethylsiloxane (PDMS).[4] Starting with an array of silicon wires, grown by VLS, they coat the wires with PDMS and then mechanically remove them from their growth substrates.[4] This method will be used to maintain the structure of the silicon wire arrays while they are removed from the silicon growth substrate and then annealed onto silicon coated glass substrates. The second approach involves the direct growth onto an alternative substrate. It has been shown that silicon nanobridges will grow between parallel etched pillars.[5] Based on these results, intersubstrate growth was attempted. Both approaches would allow the silicon substrate to be reused as a base for growing more silicon wires; while a cheaper material is used to support the wire arrays for the rest of the solar cell fabrication.

EXPERIMENTAL DESCRIPTION

Materials Used

The silicon wire arrays used in this study were grown using $SiCl_4$ and a H_2 carrier gas at a growth temperature of 1050 °C. The substrates used held five distinct square arrays (Fig. 3) each with different wire diameters and wire spacing combinations.

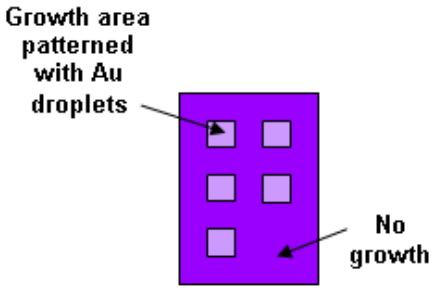


Figure 3: Top view of growth substrate showing five unique areas of wire growth

The polymer used was polydimethylsiloxane (PDMS) and it was prepared using a 10:1 weight ratio of base to curing agent. When the polymer proved to be too viscous it was diluted with dichloromethane. The final recipe for preparing the polymer was 1 gram of base, 0.1 gram curing agent, and 4 grams of dichloromethane. Alternative substrates used included silicon, gold coated silicon, and silicon coated quartz.

Methodologies

For the PDMS approach, two methods were used to coat the wire arrays. The first was drop casting. A pipette was used to drop the desired amount of the polymer over the sample, which was left to cure overnight. The additional time also helped the polymer planarize over the array. Often the polymer was still tacky and an additional heat cure for 10 min at 150 °C was used to further solidify the PDMS. The second method was spin casting. The sample was thoroughly coated with PDMS and then spun on a spinner at 1000 RPM for 2 min. The number of layers applied to a sample indicates the number of times this was repeated. Spin cast samples were then cured immediately. Thermal heating was done using a hot plate at either 100 °C for 45 min, 125 °C for 20 min, or 150 °C for 10 min.

Once the wires were covered in a layer of PDMS that left the gold tips exposed, they were removed from their growth substrates by mechanically scraping the arrays off with a razor blade. For the annealing process, the flexible silicon wire arrays were then sandwiched between two pieces of the alternate substrate and weighted down on the hot plate. They were heated for 1 hour at 370 °C. A tube furnace was also used when higher temperatures were necessary.

For the direct growth approach the growth substrate was positioned in the boat, as usual, and small pieces of sapphire were used to support the alternate substrate above the growth substrate, as shown in Fig. 4. Alternate substrates included silicon and silicon coated quartz wafers. The wires were allowed to grow for one hour. Other than the longer growth time, the process was not altered. The alternate substrates were removed from the growth substrates by prying the two wafers apart using tweezers; when this was not enough to seperate the wafers, a razor blade was

used to begin the seperation.

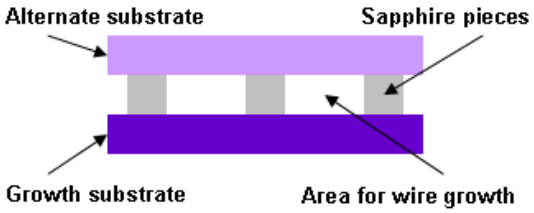


Figure 4: Side view of direct growth setup

RESULTS

Removal and Transfer of Arrays Using PDMS

The first method of applying the PDMS to the silicon wire array was drop casting. Fig. 5 shows an undiluted mix of PDMS drop cast over a silicon wire array. As can be seen from the figure, this simple method resulted in complete coverage of the wires by PDMS. In an attempt to create a thin enough layer to expose the gold tips of the wires, the PDMS was diluted with dichloromethane. However this did not improve the results of drop casting.

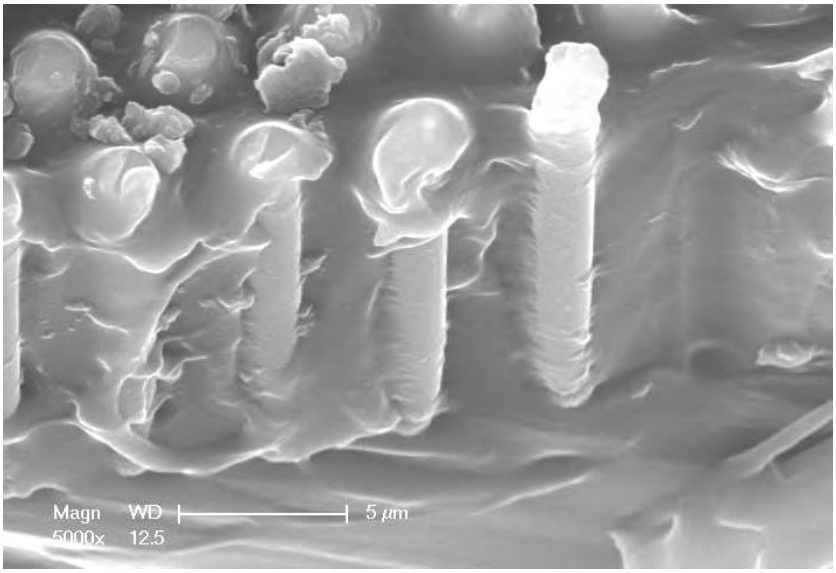


Figure 5: Drop cast PDMS covered silicon wire array

The second method of applying the PDMS was spin casting. This method resulted in thinner, more even coatings of PDMS. The thickness of the PDMS could

be controlled by adjusting the layers of PDMS applied to the substrate as shown in Fig. 6.

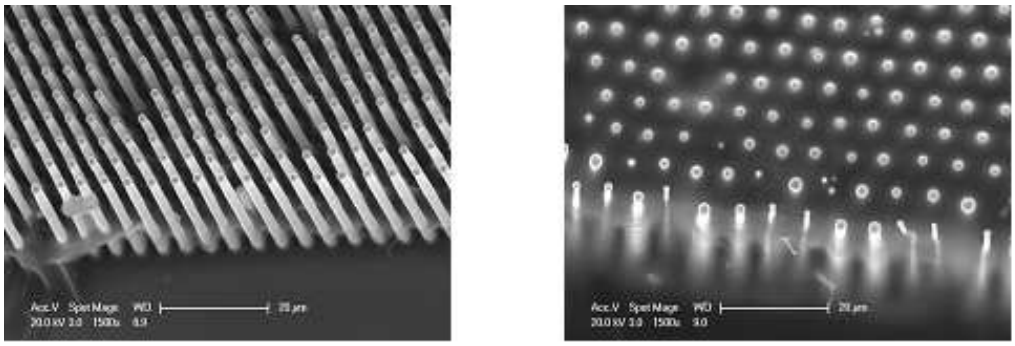


Figure 6: (a) Spin cast 1 layer PDMS (b) Spin cast 2 layers PDMS

Once the wires were covered in a layer of PDMS that left the gold tips exposed they were removed from their growth substrates mechanically. The resulting squares of silicon wires were flexible but maintained their structure. Several attempts were made at annealing the wire arrays to alternate substrates. Our results did not show a strong bond between the new substrate and the gold tip.

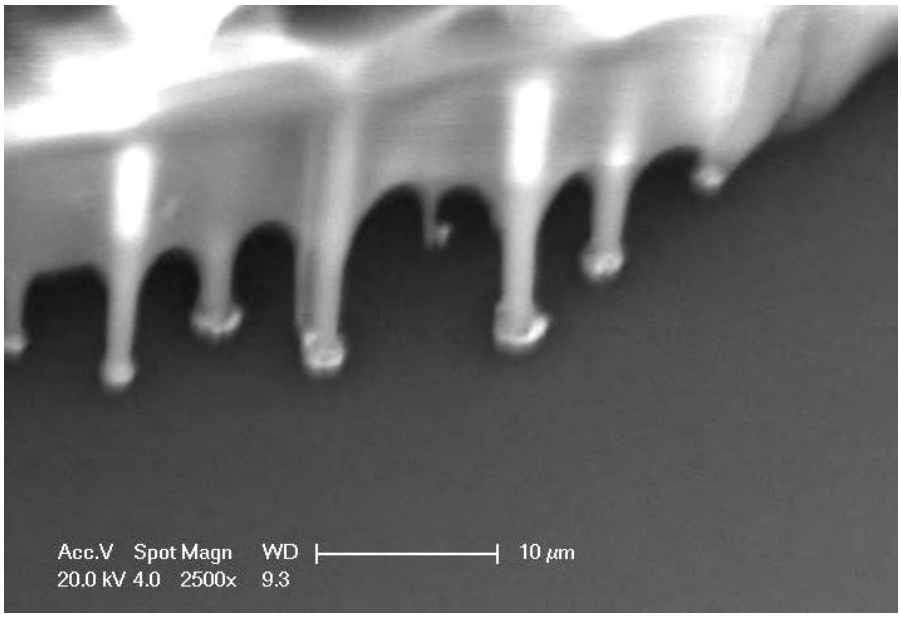


Figure 7: PDMS coated wire array annealed using hot plate to silicon

Fig. 7 shows the gold tips after being annealed to the new silicon wafer at 370 °C for one hour. From the scanning electron miscroscopy (SEM) micrographs it is hard

to determine if the gold has melted so it is doubtful that the temperature was high enough for the gold and silicon to alloy. In order to achieve higher temperatures the experiment was repeated this time using a tube furnace at 700 °C and 600 °C. At these temperatures, the PDMS decomposed and the wire arrays broke apart. No wires were successfully annealed to the silicon substrate.

Direct Growth

The first alternate substrate used for the direct growth method was a silicon substrate of the same size as the growth substrate. The growth substrate was completely covered by the alternate substrate. The growth of the wires was significantly stunted in the middle of the substrates (Fig 8). The stunted growth meant that only the outer perimeter of the growth areas was transferred onto the alternate substrates (Fig. 9). The stunted growth is possibly due to the reduced flow of $SiCl_4$ into the middle of the pattern. Next a silicon coated quartz substrate of the same size as the growth substrate was used. Even less of the perimeter was transferred, and the wires were not upright or orderly (Fig. 10).

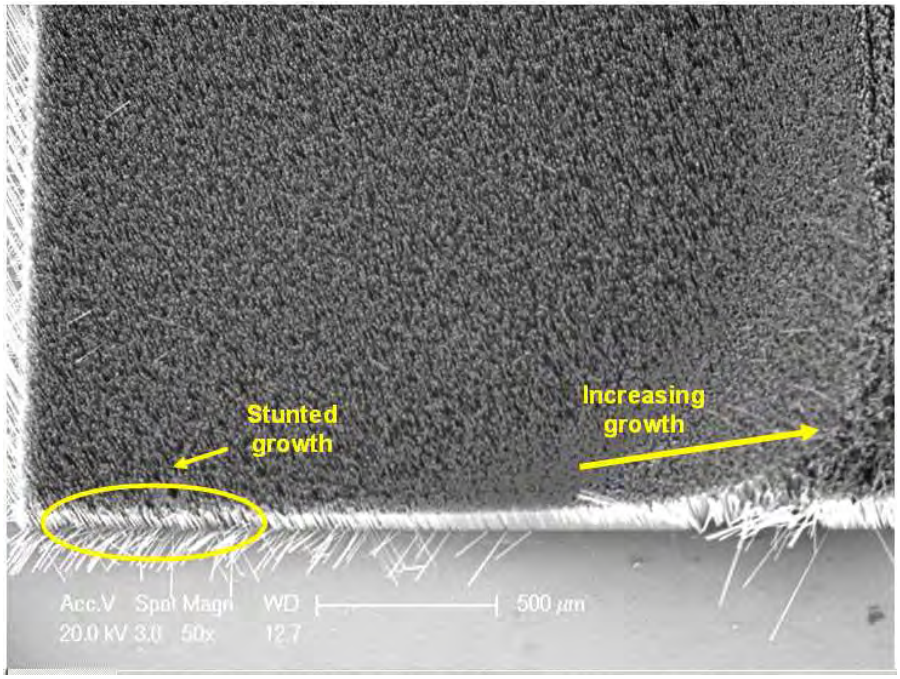


Figure 8: One array of SI wires on growth substrate showing the varying growth

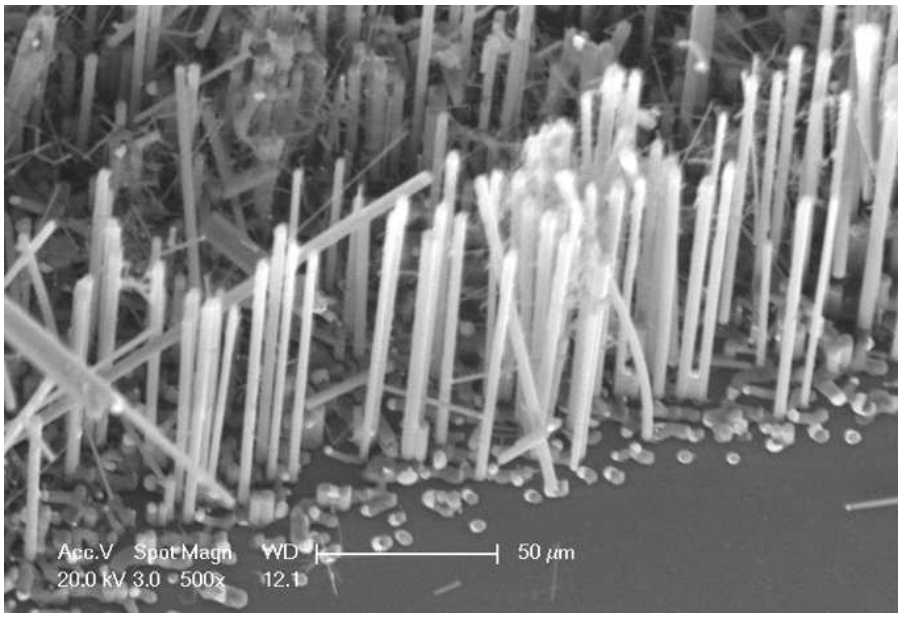


Figure 9: Silicon wires grown onto new silicon substrate

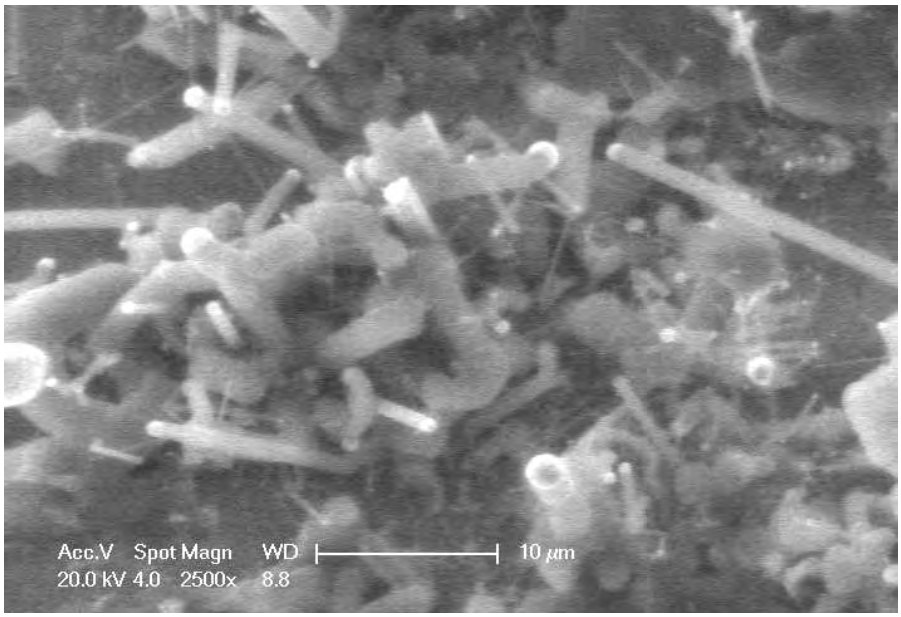


Figure 10: First attempt at silicon wire growth onto Si coated quartz

To remedy the poor growth in the center of the substrate, small pieces of silicon were placed over each growth area individually. This growth resulted in more uniform transfer of the wire arrays, including one complete square (Fig.11a, though the centers of some were still barren. Fig. 11b shows the bond between the silicon substrate and the gold tips. Another attempt was made at growth onto the silicon coated quartz (Fig. 12). This time one large growth area was used as opposed to the five area growth substrates previously used and described in the experimental section. This substrate was covered with one large piece of the silicon coated quartz. The perimeter of the growth was transferred onto the quartz. The quantity of wires transferred was improved from the first growth onto silicon coated quartz and more of the wires were upright (Fig. 12).

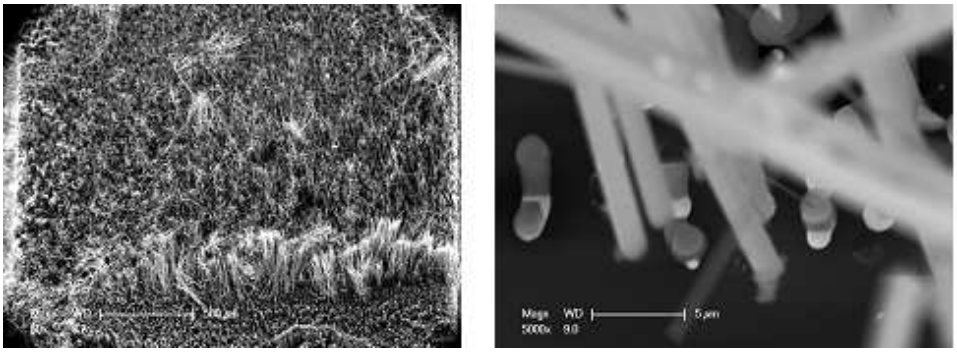


Figure 11: (a) Improved growth onto Si (b) Close up on melted Au tips

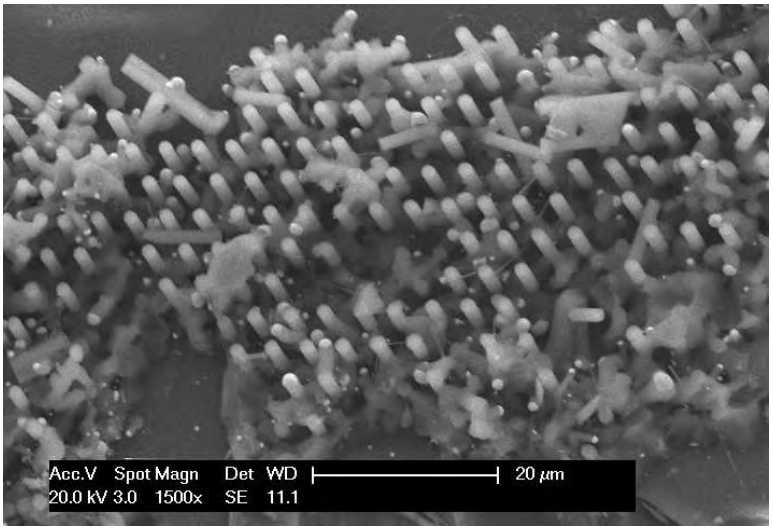


Figure 12: Improved silicon wire growth onto Si coated quartz

DISCUSSION

Of the two methods studied it appears that the direct growth method is the most promising technique for wire transfer. Though spin casting produced a PDMS coated wire array with exposed tips that was easily removed from the growth substrate; the PDMS could not withstand the high temperatures needed to anneal the gold tips to the new substrate. This could possibly be overcome if a polymer that could withstand these temperatures was used.

The direct growth method has proven it is capable of significant growth onto another silicon substrate and limited growth onto silicon coated quartz. The next step is to improve the results on silicon coated quartz. The latest results on quartz showed a transfer of wires around the perimeter of the growth area. These are comparable to the results on silicon when one large piece of the alternate substrate was used. It was not until smaller pieces were used that the amount of growth transferred improved on silicon. The results up to this point indicate that the gold tips will anneal with the silicon coated quartz but some parameters may need to be adjusted for example size and placement of alternate substrate, growth time, or gas flow.

CONCLUSION

Transferring silicon wire arrays onto alternative substrates is a promising technique for making radial silicon solar cells a more cost effective technology. The successful transfer of wires onto a silicon wafer, and the partial transfer of wires onto a silicon coated quartz wafer indicate that with further study the complete transfer of a silicon wire array onto a cheaper alternate substrate like silicon coated glass is feasible. This cost reduction would widen the gap between the cost of planar silicon solar cells and radial silicon solar cells thereby increasing the appeal of the radial geometry silicon solar cell.

ACKNOWLEDGEMENTS

I would like to thank Dr. Chito Kendrick for his guidance throughout this project and Prof. Joan Redwing for the opportunity to be a part of her research group this summer. This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081 and Department of Energy Award No. DE-FG36-08GO18010.

REFERENCES

[1] A. Luque and S. Hegedus, *Handbook of Photovoltaic Science and Engineering*, John Wiley and Sons., Hoboken, NJ, 2003.

[2] B. Kayes, H. Atwater and N. Lewis, “Comparison of the device physics principles

of planar and radial p-n junction nanorod solar cells," *J. Appl. Phys*, **97** 114302–11 (2005).

[3] A. P. Levitt, *Whisker Technology*, Wiley Interscience., Hoboken, NJ, 1970.

[4] K. E. Plass, M. A. Filler, J. M. Spurgeon, B. M. Kayes, S. Maldonado, B. S. Brunschwig, H. A. Atwater and N. S. Lewis, "Flexible Polymer-Embedded Si Wire Arrays," *Adv. Mater.*, **21** 325–328 (2009).

[5] M. S. Islam, S. Sharma, T. I. Kamins and R. S. Williams, "Ultrahigh-density silicon nanobridges formed between two vertical silicon surfaces," *Nanotechnology*, **15** L5–L8 (2004).

ENERGY STORAGE CHARACTERIZATION OF NON-LINEAR DIELECTRICS NEAR BREAKDOWN

Colin Stapleton[*] and Michael Lanagan[#]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

[*]Undergraduate student of
Department of Electrical Engineering
Milwaukee School of Engineering
Milwaukee, WI 53202

ABSTRACT

Near breakdown, dielectrics exhibit non-linear polarization characteristics, which strongly influence the energy storage and loss values in capacitors. This paper discusses a method of modeling this non-linearity, for use in future research. The model developed was used successfully for a particular ceramic-polymer dielectric, provided the parallel resistivity of the dielectric was known.

INTRODUCTION

Capacitors are used in many applications to store energy, and to release energy rapidly. The energy stored by capacitors can be used as an alternative to batteries in such applications as implanted heart defibrillators and electric cars. Capacitors can also be discharged more rapidly than batteries, and are used in systems requiring large pulses of power, such as pulsed lasers. As the Ragone Chart shows, capacitors have a lower energy density than batteries, but are able to deliver greater power.

[#] Faculty Mentor

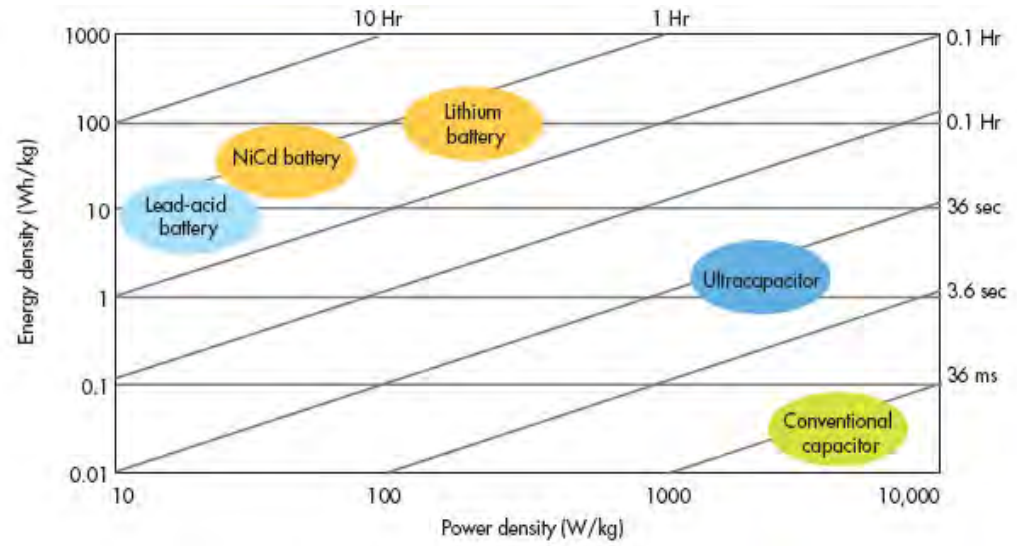


Figure 1: Ragone chart[1]

Comparison of capacitor energy densities can be performed by measuring the polarization of a dielectric, for a given applied electric field.

$$P = \frac{Q}{A} \tag{1}$$

and

$$E = \frac{V}{d} \tag{2}$$

where P is the dielectric polarization, Q is the electrode charge, A is the area of the electrode, E is the electric field, V is the applied voltage, and d is the space between the two electrodes. Given the polarization and electric field, the energy density can then be determined:

$$\frac{Energy}{vol} = \int EdP \tag{3}$$

where $\frac{Energy}{vol}$ is the energy density of the dielectric.

The energy stored in a capacitor depends on the physical dimensions of the capacitor, and its dielectric constant (relative permittivity).

$$C = \frac{KA\varepsilon_0}{d} \tag{4}$$

where C is the capacitance, K is the relative permittivity, A is the surface area of the capacitor, d is the distance between the plates of the capacitor, and ε_0 is the absolute permittivity.

The area and distance are constrained by physical limitations, and ε_0 is a constant, thus the relative permittivity is important in maximizing the energy density of the capacitor. The relative permittivity is determined by the material properties of the dielectric. A dielectric that has a higher polarizability will have a greater relative permittivity.

For practical applications, the method of dielectric failure must also be considered. Two common dielectric materials, polymers and ceramics, have different methods of breakdown. Polymer dielectrics exhibit graceful failure, in which the capacitance slowly lowers over the life of the capacitor. Ceramic dielectrics short between the electrodes when breakdown occurs. Because ceramic dielectrics have a greater relative permittivity, it is desirable to combine the high relative permittivity of ceramics with the graceful failure of polymers. One method of combination is to distribute nano-particles of ceramic into a polymer matrix. A computer model to predict the performance of these ceramic-polymer dielectrics will be useful for future research.

Stored Energy of Non-Linear Dielectrics

Predicting the energy density of the ceramic-polymer compositions can be difficult, because of their non-linear resistivity. For a linear dielectric, the energy density can be expressed as a function of the electric field and permittivity:

$$\frac{Energy}{Vol} = \frac{1}{2}\varepsilon_0 K E^2 \tag{5}$$

where $\frac{E}{Vol}$ is the energy density, E is the electric field, ε_0 is the absolute permittivity and K is the relative permittivity. Where V is voltage and Q is charge.

Each dielectric composition will break down at a particular voltage. When voltages near breakdown are applied to the dielectric, the resistivity of the material will increase non-linearly. Equation (5) cannot be used in this case. However, by plotting the charge as a function of voltage, the stored energy can be determined:

$$Energy = \int V dQ \tag{6}$$

Equivalent Circuit Model

A capacitor can be modeled as an equivalent circuit, such as in Figure 2.

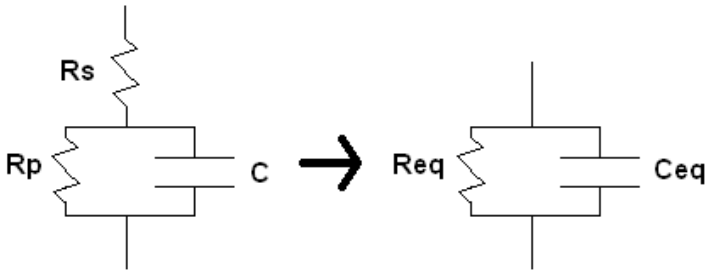


Figure 2: Capacitor equivalent circuit

This equivalent circuit is useful for modeling purposes, since the charge developed across the circuit is the integral of the current through the circuit. When the circuit is simplified into an equivalent parallel resistance and capacitance, the charge can then be calculated by integrating the sum of the currents through the capacitor and resistor:

$$Y_T = \frac{\frac{1}{R_s}(\frac{1}{R_p} + j\omega C)}{\frac{1}{R_s} + \frac{1}{R_p} + j\omega C}$$

$$= \frac{R_p + R_s + R_s(\omega C R_p)^2}{(R_p + R_s)^2 + (\omega C R_p R_s)^2} + j\omega\left(\frac{C{R_p}^2}{\left(R_p + R_s\right)^2 + (\omega C R_p R_s)^2}\right)$$

$$= \frac{1}{Req} + j\omega Ceq$$

where Y_T is the total admittance of the circuit, Ceq is the equivalent capacitance, and Req is the equivalent resistance. The current through a capacitor (I_C) is a function of the capacitor's voltage and capacitance, therefore:

$$I_C = C\frac{dV}{dt} \tag{7}$$

$$Q = \int (Ceq\frac{dV}{dt} + \frac{V}{Req})dt \tag{8}$$

where Q is the charge, V is the voltage and t is time.

The equivalent circuit can be used to plot charge versus voltage, producing a hysteresis loop. A sample with infinite parallel impedance, and zero series impedance, will resemble Figure 3.

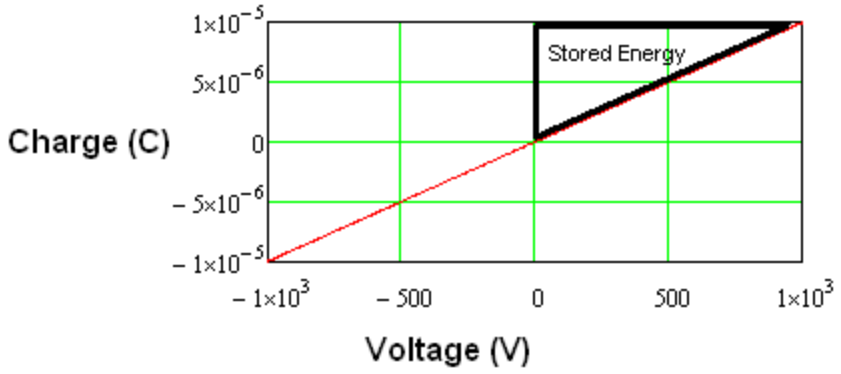


Figure 3: Purely capacitive hysteresis loop

As an input voltage is applied, the capacitor will charge and discharge, with no loss of energy. The energy stored will be the region indicated in Figure 3, as determined by Equation (6). A sample with zero capacitance will resemble Figure 4. As an input voltage is applied, no energy is stored by the resistor.

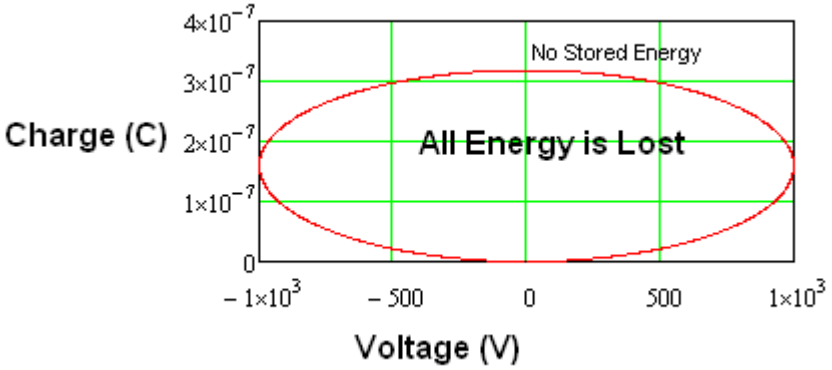


Figure 4: Purely resistive hysteresis loop

EXPERIMENTAL DESCRIPTION

Polarization Measurement

To accurately model the non-linear characteristics of a dielectric near breakdown, the equivalent circuit model must be compared with empirical data. The polarization of a dielectric sample was measured by the equipment in Figure 5.

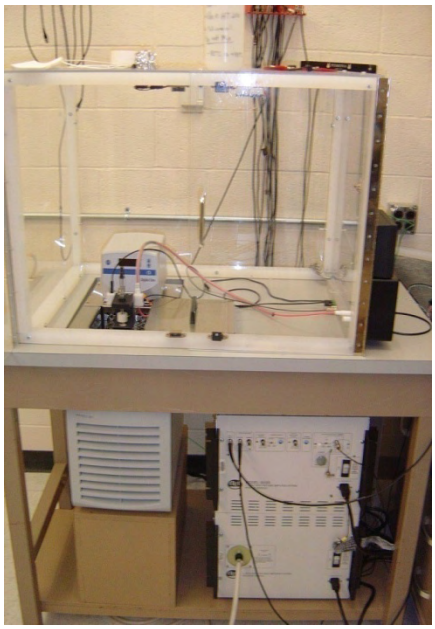


Figure 5: Polarization measurement equipment

The sample was placed between two electrodes, and an AC voltage was applied, at a specified frequency and peak. The polarization equipment measured the charge on the sample, and plotted the charge as a function of voltage. Figure 6 shows the hysteresis loop of the ceramic-polymer dielectric near breakdown. The non-linear characteristics of the loop are clearly visible.

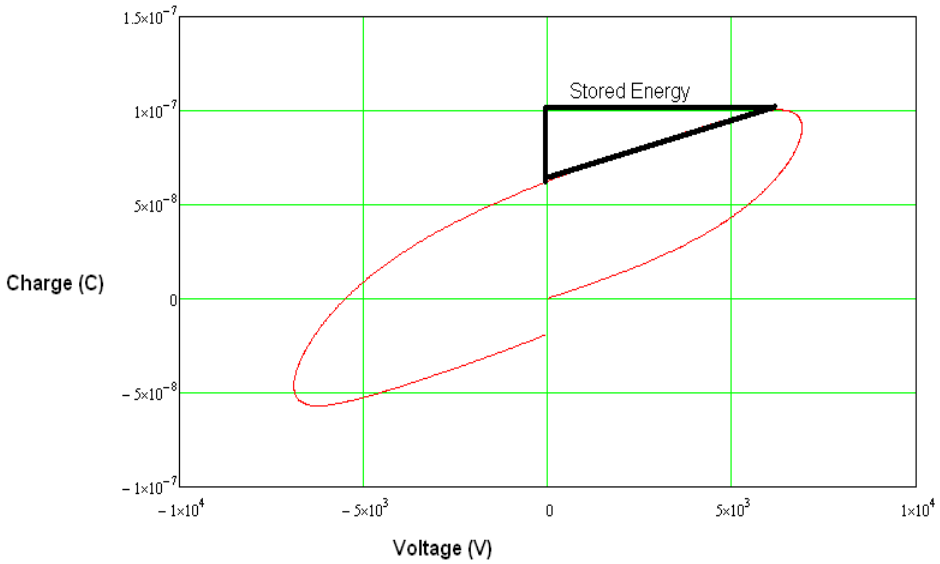


Figure 6: Ceramic-polymer hysteresis loop

Varistor Model

A varistor exhibits very high resistance up to a specified voltage, at which its resistance decreases at a greater than linear rate. The current through a varistor above the specified clip voltage is proportional to the voltage, raised to a constant exponent.

$$I_V = XV^{\gamma} \tag{9}$$

where I_V is the varistor current, X and γ are constants, and V is the voltage.

The equivalent circuit model of Figure 2 was modified by the above varistor equation in MathCad, and used to develop a hysteresis loop model of varistor charge. The equivalent circuit was then measured in the polarization equipment, and the two loops were compared.

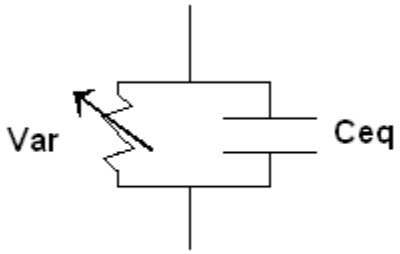


Figure 7: Equivalent circuit with varistor

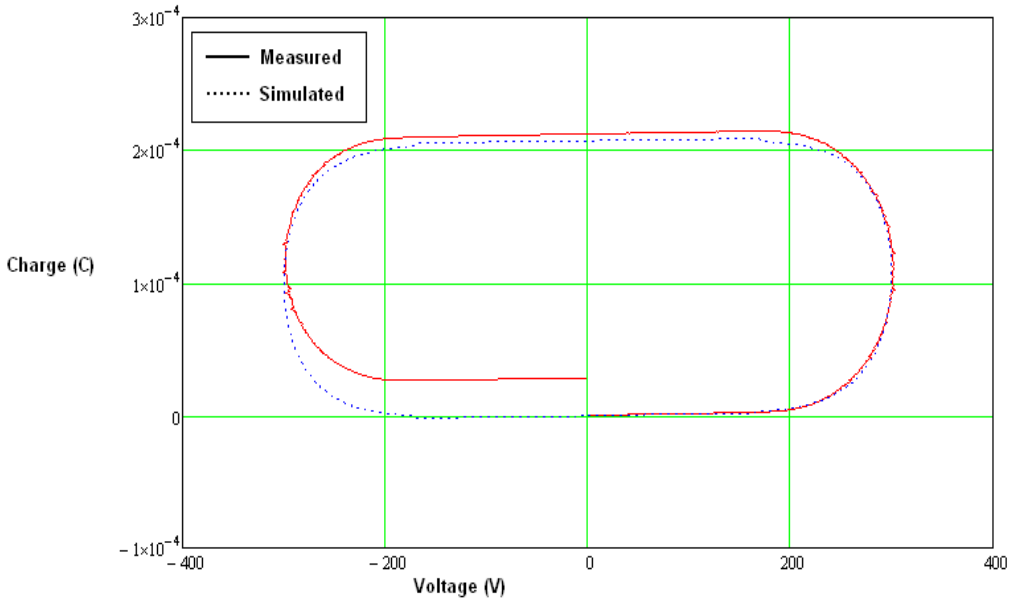


Figure 8: Varistor hysteresis loop

The varistor used was a Harris Corporation 130L20B, with a rated clip voltage of 175V. The MathCad model provided a close approximation of the varistor charge at this particular range of voltages. The varistor was not, however, a useful model of the ceramic-polymer dielectric. As Figure 8 shows, the change in parallel resistance of the varistor is more dramatic than that of the ceramic-polymer dielectric.

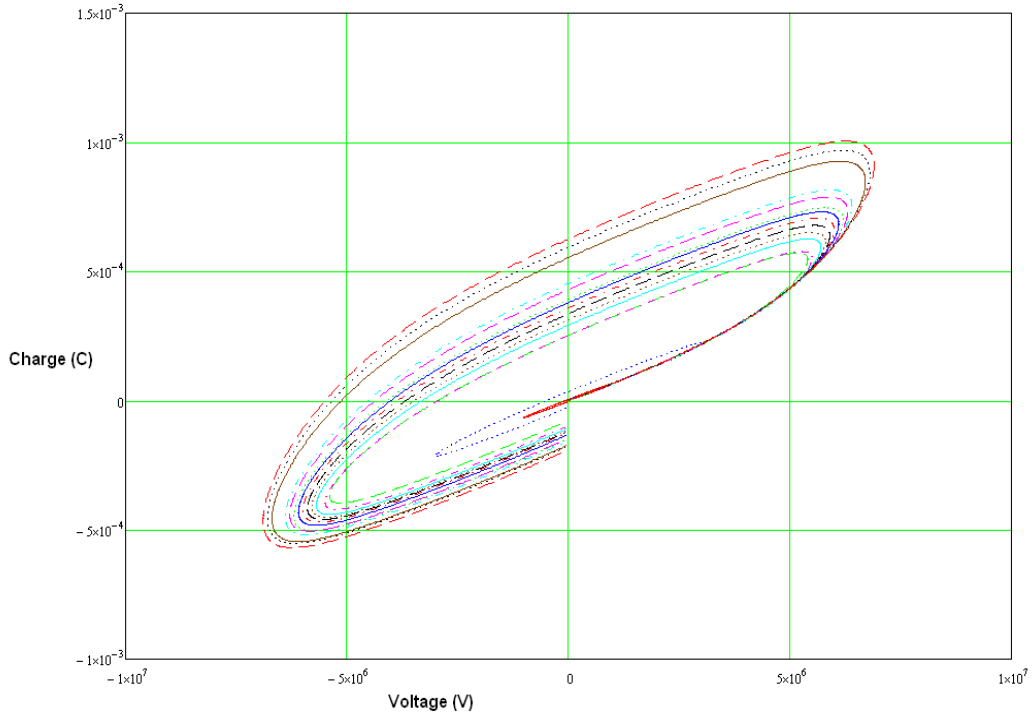


Figure 9: Ceramic-polymer polarization loops

As the voltage applied to the ceramic-polymer dielectric increases, the parallel resistance changes at a non-linear rate. To approximate this change, the magnitude of a sine function, from 0^o to 90^o was used. The resistances at minimum voltage and breakdown voltage were determined, and this difference was multiplied by the magnitude of a sine wave to give a close approximation of the change in parallel resistance. The non-linear characteristic was modeled by multiplying the equivalent circuit charge by an exponential function.

In addition to modeling the polarization of the ceramic-polymer dielectrics, it was desirable to measure the stored energy and hysteresis loss of the samples. This was done through a numerical approximation of area, using the charge and voltage data produced by the polarization equipment.

RESULTS

The ceramic-polymer dielectric model provided a close approximation of the polarization data.

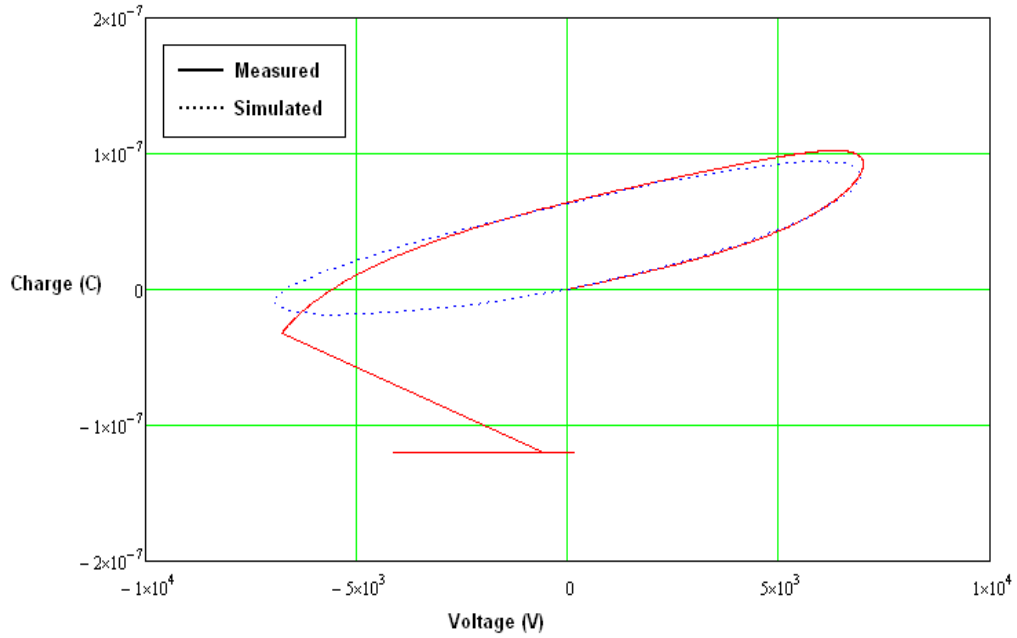


Figure 10: Ceramic-polymer polarization data and model

Figure 10 shows the polarization loop of the sample at breakdown. The sample breakdown is visible in the third quadrant. The model requires input of the sample's capacitance and breakdown voltage. The model also does not account for variations in the resistivity of different samples. Multiple sets of collected data showed variations in the parallel resistances of different samples, for the same dielectric material.

CONCLUSION

Simplifying the dielectric as an equivalent electrical circuit was a successful method of simulating the dielectric's polarization loop. The non-linear polarization loop of the ceramic-polymer composite can be modeled, provided its change in resistivity is known. Varistors also exhibit non-linear resistivity, but are not a useful simulation of the ceramic-polymer composite dielectric.

As a dielectric, the energy loss of the ceramic-polymer composite may be too large for practical use. Figure 11 shows the smaller hysteresis loss of a glass dielectric for comparison. Future work may reduce the hysteresis loss of the composite, or it may show that glass dielectrics are a more useful material for applications.

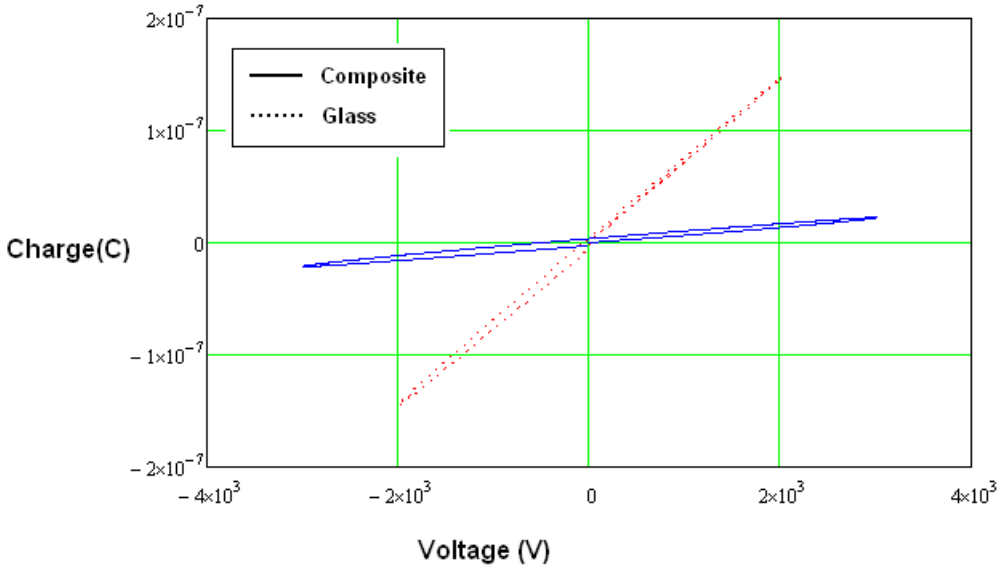


Figure 11: Comparison of ceramic-polymer composite to glass

ACNOWLEDGEMENTS

One of the authors (CS) would like to thank Steve Perini, Jeff Long, Ben Koch and Badri Rangarajan for their knowledge and support of this project. This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081.

REFERENCES

1. http://electronicdesign.com/Articles/Print.cfm?ArticleID=17465

STUDY OF FABRICATION AND ENERGETIC PROPERTIES OF POROUS SILICON

Javier Ivan Espinosa Acevedo,[*] Srinivas Tadigadapa,[#] and Venkata Sharat Chandra Parimi[+]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate Student
Department of Electrical Engineering
Inter American University of Puerto Rico
Bayamon, PR 00957-6257

ABSTRACT

Porous Silicon (PS) is a variant of the element silicon (Si) that has nanometer scale pores in its microstructure, rendering a large surface to volume ratio in the order of $500m^2/cm^3$. PS is formed by an electrochemical dissolution of crystalline Si wafers in Hydrofluoric acid (HF). In this study we analyze the etching process and its effect on the properties of porous silicon during its formation. We observed how the doping level, etching time, current density and HF concentration levels affect the porous structure. The porous silicon samples were then filled with an oxidizer and ignited and the combustion reaction was studied. The long term goal of this project is to understand the combustion process of silicon. Understanding the fundamental aspects of the chemical reaction will permit control of the reaction which can be used for several purposes such as using silicon as an energy source and in the fabrication of Si integrated self destructive chips.

INTRODUCTION

Silicon (Si) is one of the most commonly used elements. It is a basic material that makes up about 25–28% of the Earth's crust. It is the 14^{th} element in the periodic table and is chemically less reactive than carbon. Silicon commonly occurs as an oxide and as a silicate. Due to its semiconductive properties silicon has revolutionized technology and is now found in almost all electronic

[#] Faculty Mentor
[+] Graduate Mentor

applications. In addition it is an important ingredient in steel and it is used in lasers, glasses, transistors, solar cells and many solid state electronic devices.

Silicon was discovered by Lavoisier in 1787 and first isolated by Berzelius in 1823. In 1907 Potter studied the interaction of silica with carbon, which prepared the way to the process of obtaining silicon for commercial purposes. Silicon is prepared through a process which involves the heating of silica and carbon with an electric furnace, utilizing carbon electrodes. The resulting product is in the form of amorphous silicon. The silicon crystals used for electronic devices are obtained using Czochralski method. These thin slices of semiconductor material are then used in the fabrication of microchips.

Porous silicon (PSi) is a form of silicon which has unidirectional nanometer scale pores on its microstructure rendering a large surface to volume ratio. Silicon seems to be the only material that spontaneously gives rise to well define pores with nm dimensions. Porous silicon was discovered accidentally in 1956 by Arthur Uhlir Jr. and Ingeborg at the Bell laboratories. Porous silicon can be broadly classified into three regimes as follows:

1. Microporous: pore diameters <10 nm.
2. Mesoporous: pore diameters between 10–50 nm.
3. Macroporous: with geometries >50 nm.

Though porous silicon was discovered in the 1950s, not much attention was focused on porous silicon till 1990 when unexpected optical properties of microporous silicon were discovered [1]. This discovery of visible light emission from porous silicon generated significant interest in hopes of developing an electrically excited LED or laser. Another important discovery regarding porous silicon is its reactivity, which was discovered accidentally. Silicon-based explosive behavior was described for the first time in 1992, when it was reported that potassium nitrate or nitric acid could create explosive-like reactions when reacting with porous silicon. Researchers were investigating the chemiluminescence of anodized silicon, when the addition of concentrated nitric acid to freshly etched porous silicon reacted violently and produced a “flash of light with an audible pop [2]. The next report on explosive porous silicon appeared in 2001 [3]. Nano-explosions were accomplished at very low cryogenic temperatures, in the range 4.2 K to 90 K. When porous silicon in a vacuum was cooled to these low temperatures, and oxygen was allowed into the vacuum system, the porous silicon exploded. Placing porous silicon into liquid oxygen also results in an explosive reaction. These discoveries have initiated research into porous silicon as an energetic material.

Porous silicon is formed by means electrochemical etching of silicon in a HF medium. The formation of the pores occurs on the initiation of the rising part of the curve $0 < V < V_{ep}$, where V_{ep} is the electropolishing voltage (Figure 1).

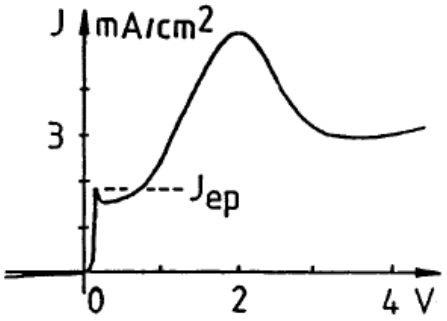


Figure 1 Typical *J–V* curve of p^+-Si in dilute aqueous HF solution. Porous silicon is obtainable for $J > J_{ep}$ (from [4])

Several factors are involved in the formation of porous silicon. Table I below shows a few of the most known factors in porous silicon formation.

Table I Proposed Factors in Porous Silicon Formation (from [4]).

Pore initiation		**Pore propagation**	
Dopant induced dissolution	Dubin (1992)	Image force effects	Beale (1985)
Interface instability	Kang et al (1993) Valence (1995) Wehrspohn (1996)	Hole diffusion	Smith (1988)
Hydrogen induced defects	Allongue (1995)	Crystallography	Chuang (1988)
Diffusion effect	John (1995)	Charge transfer	Gaspard (1989)
Vacancy supersaturation	Corbett (1995)	Quantum Confinement	Canham (1990) Lehmann (1991)

The initial wafer is weighed before and after the anodizing process. It is also weighed after the separation of the entire PSi layer in an aqueous solution. The porosity is given by the following equation.

$$P = \frac{(m_1 - m_2)}{(m_1 - m_3)}$$

where:
m_1 : Weight of the Silicon wafer
m_2 : Weight of the Porous silicon
m_3 : Weight after the separation of PSi

Using the previous masses it is also possible to obtain the thickness of the layer.

$$W = \frac{(m_1 - m_2)}{S \times d}$$

where:
d: the density of bulk silicon
S: the wafer area exposed to HF during anodization.

The measurements of the thickness and porosity can also be obtained with Non-destructive optical techniques such as ellipsometry. However these methods are model dependent and further studies are needed to understand the ellipsometric data.

Porous silicon has several interesting applications in widely different areas ranging from explosives to biological applications. A few of the applications or porous silicon are:

1. Small porous silicon samples filled with oxidizers can be used as to ignite booster charges of automobile air bags.
2. Due to its large surface to volume ratio, porous silicon can store large amounts of hydrogen. Thus, porous silicon can be used as a hydrogen storage media. Porous silicon can release up to 4% by weight of hydrogen which satisfies the requirements on materials accumulating hydrogen in fuel cells.
3. Porous silicon can be integrated into electronic circuits to protect sensitive technology by forming self destructible chips.
4. The energy yield that can be obtained per unit weight from optimized porous silicon can be much higher than the most existing energetic materials of today such as TNT and several other high power explosives.

To summarize, porous silicon has several potential commercial applications because of its unique properties. The high surface to volume ratio of porous silicon resulting from its nanoscale structure gives rise to several interesting properties such as photoluminesce (PL), high reactivity and a very high specific surface area. A short list of some of the applications of porous silicon is:

1. Initiator for booster charges of air bags.
2. Gas sensors
3. Humidity sensors
4. Waveguides for light emitting devices due to the small optical loses in oxidized porous silicon.
5. Capacitors
6. Bioactive implants
7. Light emitting devices
8. Improving solar cells efficiencies
9. Environmental monitoring

10. Medical diagnostics
11. As fuel for MEMS devices.

In the current work, the effect of the etching conditions on the pore structure and thus the oxidation reactions of porous silicon were studied.

EXPERIMENTAL DESCRIPTION

There are several methods to fabricate porous silicon [5]. The three most common methods for anodizing porous silicon are lateral anodization cell, the conventional single-tank cell and the double-tank cell. In this work the method used to etch porous silicon is the conventional single-tank cell method (Figure 2). This method was selected because of the ease of producing porous silicon in a specific area and allows obtaining a good control of porosity and thickness and facilitates back side illumination of the wafers if required.

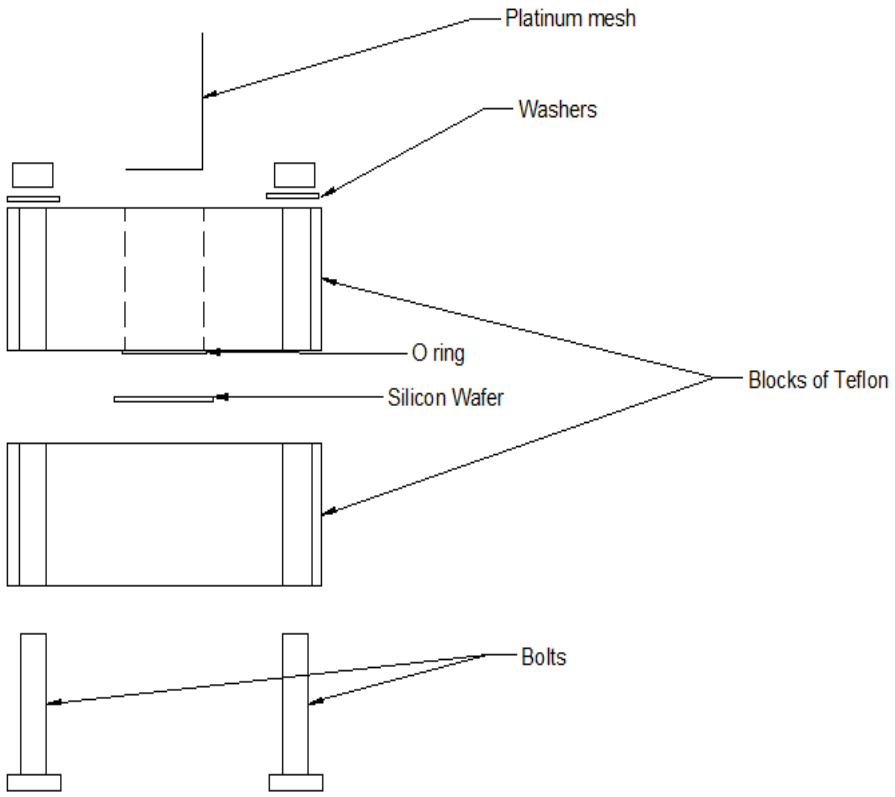


Figure 2 Schematic of the experimental set up for etching porous silicon

The anodization cell was formed using two Teflon blocks, and the volume of the electrolyte that can be held in the cell is about 3 ml. Teflon was used since it is resistant to chemical attack by hydrofluoric acid, and the volume of the electrolyte was minimized so that the least amount of hydrofluoric acid would be handled. An O-ring made of Teflon was used to prevent any leakage of the electrolyte. The

area of the wafer enclosed by the O-ring is the etched area, which is a circle with a diameter of 10/16″. Uniform back side electrical contact was obtained using an aluminum foil while etching highly doped wafers. For wafers with lower doping levels aluminum was deposited on the backside to make ohmic contact. P-type silicon wafers doped with boron and <100> orientation were etched.

In the electrochemical etching process, the silicon wafer acts as the anode, and the cathode is made of a material inert to hydrofluoric acid, which is a platinum mesh in this case. A pulsed DC power supply (Dynatronix DuP10-1-3) with 1 mA current resolution and 0.1 ms pulse time resolution was used as the power source.

The electrolyte consists of HF and ethanol/acetic acid. In the anodic etching process, hydrogen bubbles are generated at the silicon surface. Also, the pore diameters are in the order of nanometers, which makes it difficult for both removing the hydrogen bubbles generated within the pores and the electrolyte to reach the end of the pores. To overcome these problems, acetic acid/ethanol is added to the hydrofluoric acid to lower the surface tension and ease the removal of the hydrogen bubbles. A minimum volume concentration of 15% of ethanol is required to accomplish this whereas 5% by volume acetic acid will suffice. Much higher HF concentrations can be obtained using acetic acid or other commercial surfactants.

The electrolyte used had 3:1 or 2:1 or 1:1 by volume of 49% HF and ethanol or 5:1 or 11:1 by volume 49% HF and acetic acid. The pore diameters reduce upon increasing the HF concentration in the electrolyte, and the HF/acetic acid electrolytes were used to etch pores of smaller diameters.

The pore size is also affected by the current density, with higher current densities resulting in larger pores. The samples were etched with current densities between 20 mA/cm^2 and 45 mA/cm^2. Also, the etching process was started at a much higher current density, of 125 mA/cm^2 and reduced gradually to the required level. This creates “funnel shaped” pores, with wide mouths (Figure 3). This further facilitates the removal of hydrogen bubbles from the pores. The total time for which the samples were etched was between 30–90 minutes. A duty cycle of 50% with pulse width of 9.9 ms was used for the etching process. During the etching process, the acid in the pores decomposes, and has to be rejuvenated to continue the etching process. Introducing breaks in the etching process allows replenishing of the acid within the pores and also helps in removing the hydrogen bubbles. The pulse width was chosen to allow sufficient time to replenish the acid inside the pores [6]. After the completion of the electrochemical etching, the silicon wafers were washed with ethanol to remove any traces of hydrofluoric acid and then dried under a stream of nitrogen.

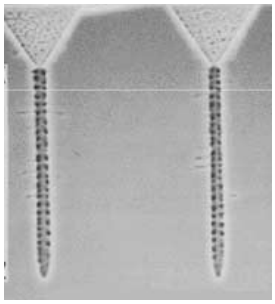


Figure 3 Current densities effects on porous silicon (from [7])

Table II HF and Current density effects on pore size

HF Concentration	Pore Size	Current Density
↑	↓	↓
↓	↑	↑

The high reactivity of porous silicon arises from its high surface to volume ratio. Having smaller pores results in higher surface to volume ratio, and the current densities and HF concentrations were changed through the work to attain smaller pore diameters (Table II). Also, the combustion of the porous silicon wafers is affected by the pore depth with the samples with greater pore depths resulting in faster reactions. It was observed that under some conditions, the porous layer breaks away from the silicon wafer being etched rendering the wafers useless (Table III). So, different etching conditions were tried out to check which conditions would yield the desired (i.e., smallest pore diameter and largest pore depth) without breaking off from the surface of the wafer.

Impregnation with Oxidizers

The porous silicon wafers etched must be filled with oxidizers before they can be ignited. The oxidizers tried were potassium nitrate, sulfur and potassium permanganate. One of the problems encountered is the small size of the pores and the hydrophobic nature of the silicon surface. These factors prevent liquids with high surface tension such as water from diffusing into the pores. So oxidizers soluble in alcohols were mainly used. Potassium nitrate is a very good oxidizer, but is insoluble in alcohols. Thus for potassium nitrate, an aqueous solution was used. The porous silicon sample was left in a saturated solution of potassium nitrate for 30 minutes and dried out. This was repeated twice. However, it was found that filling the pores with aqueous nitrate was unreliable.

The other oxidizers used were sulfur and magnesium perchlorate. With sulfur, the samples were filled with the oxidizer using molten sulfur or sulfur dissolved in carbon disulfide. To fill the pores with molten sulfur, the porous silicon with sulfur was placed in an oven at a temperature at 160 °C for 30 minutes. With carbon disulfide, the porous silicon sample was left in a saturated solution of

carbon disulfide and sulfur for 30 minutes and dried. This process was repeated twice. To fill the pores with magnesium perchlorate, a solution of magnesium perchlorate in ethanol was used. It was observed that the porous silicon samples had to be left in the solution much longer and two or more fillings had to be used for better results.

Ignition

The porous silicon samples filled with the oxidizer can be ignited using an electric spark or heat or friction. Since the goal was to observe the flame propagation, a CO_2 laser with a wavelength of 10,600 nm (IR) was used to thermally ignite the samples. The flame propagation was captured using a high speed camera. The sampling rates used were between 20,000 fps to 80,000 fps. The flame propagation rate and the plume created upon the ignition of the sample were observed.

Table III: Results of Electrochemical Etching

Resistivity of Si wafer (mΩ cm)	Current (mA/cm^2)	Ratio	Electrolyte Composition	Anodization Time (min)	Result
21.9	60	3:1	HF: ETHANOL	60	Broken
21.9	60	2:1	HF: ETHANOL	60	Broken
21.9	60	1:1	HF: ETHANOL	60	Broken
21.9	60	2:1	HF: ETHANOL	30	Broken
21.9	20	1:1	HF: ETHANOL	60	Good
21.9	45	2:1	HF: ETHANOL	60	Broken
19.2	40	11:1	HF: ETHANOL	90	Broken
19.2	20	11:1	HF: ETHANOL	60	Broken
19.2	20	11:1	HF: ETHANOL	60	Broken
19.2	20	5:1	HF: ETHANOL	90	Good
19.2	20	5:1	HF: ETHANOL	60	Good
19.2	45	1:1	HF: ETHANOL	30	Good
19.2	45	1:1	HF: ETHANOL	40	Good
19.2	45	1:1	HF: ETHANOL	45	Good
19.2	45	2:1	HF: ETHANOL	30	Good
19.2	45	2:1	HF: ETHANOL	40	Partially Broken
17.4	40	2:1	HF: ETHANOL	30	Good
17.4	40	2:1	HF: ETHANOL	40	Good
17.4	40	2:1	HF: ETHANOL	50	Good
17.4	40	1:1	HF: ETHANOL	30	Good
17.4	40	1:1	HF: ETHANOL	40	Broke
17.4	40	1:1	HF: ETHANOL	30	Good
17.2	45	1:1	HF: ETHANOL	30	Good
17.2	45	1:1	HF: ETHANOL	40	Good
17.2	45	2:1	HF: ETHANOL	30	Good
17.2	45	2:1	HF: ETHANOL	40	Partially Broken
14.6	45	1:1	HF: ETHANOL	30	Broke
14.6	45	1:1	HF: ETHANOL	40	Good
14.6	45	3:1	HF: ETHANOL	30	Good
14.6	40	3:1	HF: ETHANOL	40	Good
14.6	45	2:1	HF: ETHANOL	55	Broke
14.6	45	1:1	HF: ETHANOL	55	Good
1.8Ω	45	3:1	HF: ETHANOL	30	Good
1.8 Ω	45	3:1	HF: ETHANOL	40	Good

RESULTS AND DISCUSSION

The porous silicon wafers etched were studied using a SEM. Also, the porous silicon was impregnated with oxidizers and ignited and the combustion of the wafers was observed. From the SEM images, it was found that the pore diameters of the etched wafers are in the order of a few nm (50–100 nm) (Figures 4 and 5).

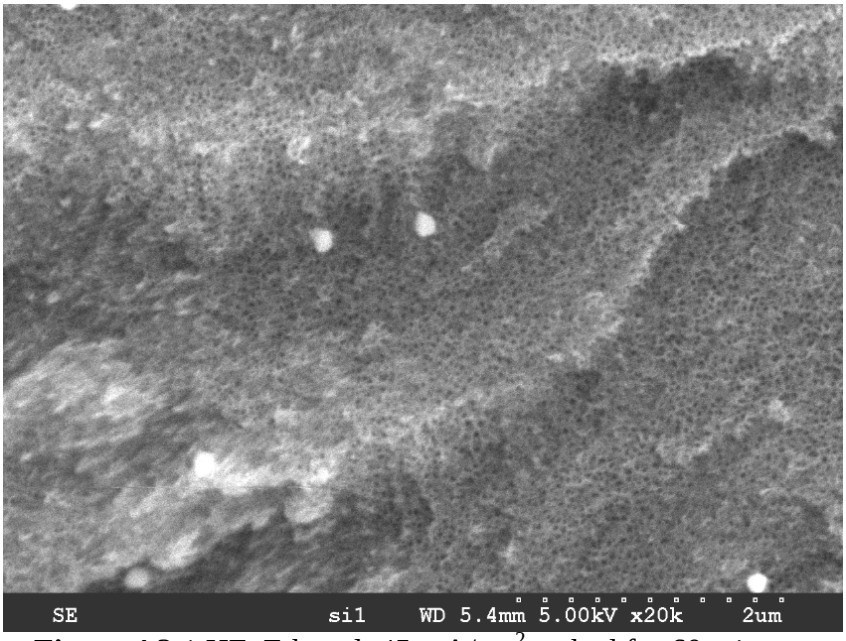


Figure 4 3:1 HF: Ethanol, 45 mA/cm^2 etched for 60 minutes

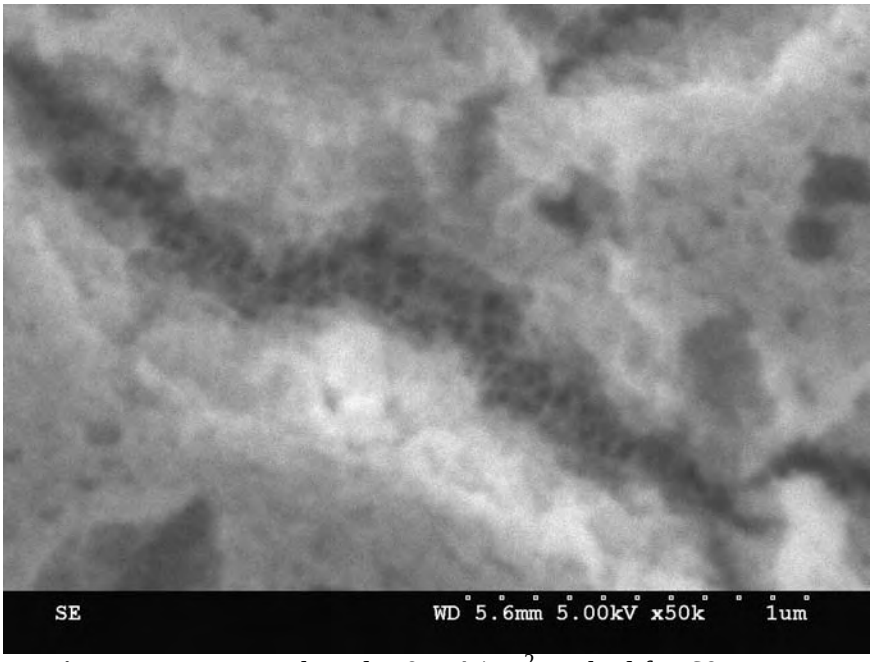


Figure 5 2:1 HF: Ethanol, 40 mA/cm^2, etched for 60 minutes

For the samples etched at lower current densities and higher HF concentrations (using acetic acid), it is estimated that the pore diameters are < 10 nm. This was found out while trying to obtain SEM images of the pores as the SEM was not able to resolve the fine structure on the sample. To resolve such small pore sizes, FE SEM must be used.

The SEM images along the edge of the wafers were used to determine the depth of the pores etched (Figure 6). For the samples used, it was found that the etching rate was nearly constant and was found to be approximately 1.9 μm/min for a 2:1 HF: Ethanol mixture at a current density of 45 mA/cm^2.

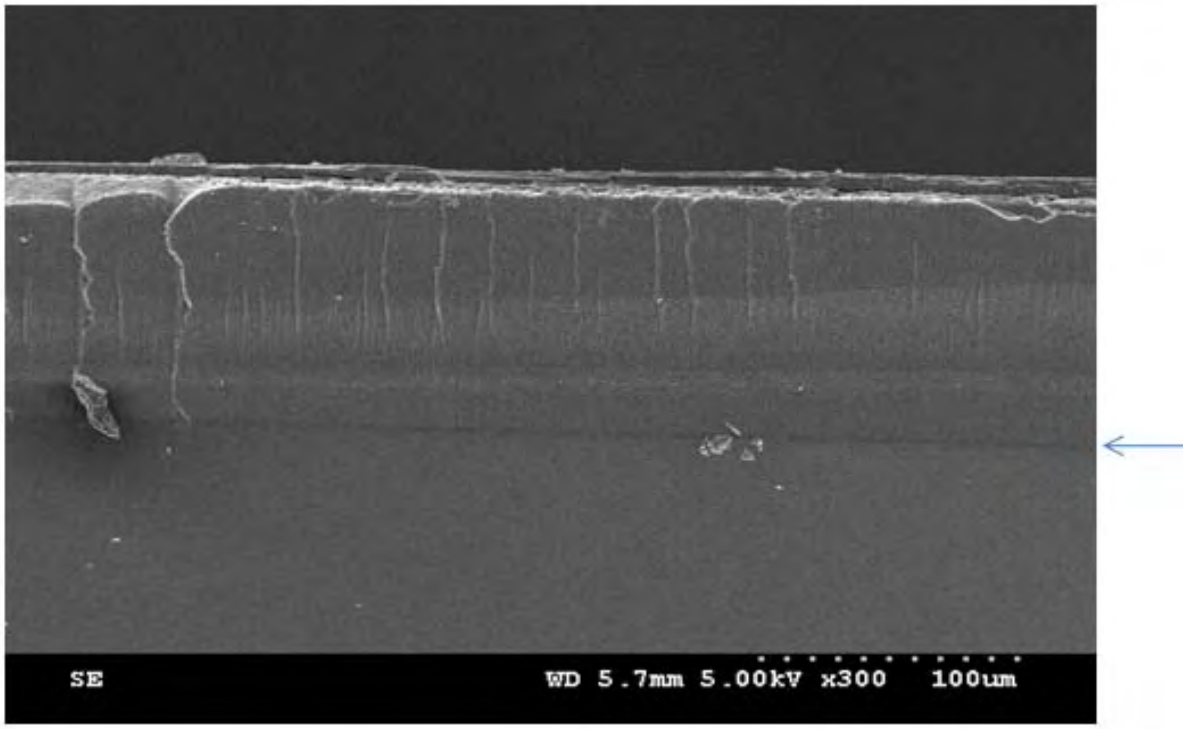
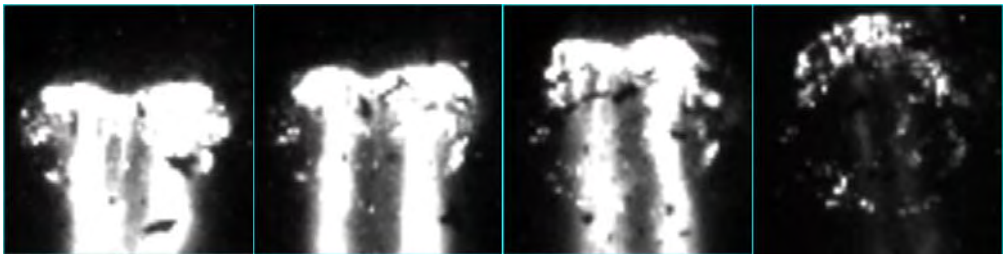


Figure 6 Side view of a porous silicon sample. The part above the arrow is porous silicon.

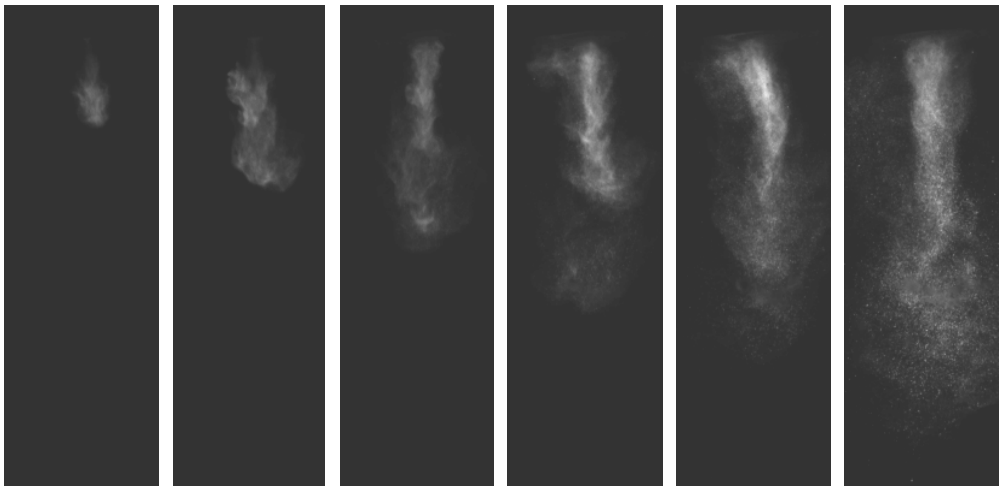
The flame propagation speed was measured on samples impregnated with sulfur and magnesium perchlorate. It was observed that filling with sulfur using a carbon disulfide solution gives better results than using molten sulfur. The samples filled with dissolved sulfur ignite on friction, heat and impact whereas the samples filled with molten sulfur ignite only upon heating. The flame speed for the samples filled with sulfur and magnesium perchlorate was found to be about 1.5 m/s. The following images capture the flame propagation on a circular sample ignited at an edge. In the first set of images the burning wafer is observed normal to the surface (Figure 7). The second image is of the plume observed during the combustion of the wafer. This image is obtained by looking at the wafer from a side (Figure 8).

t = 0 μs t = 3922 μs t = 5322 μs t = 6482 μs



t = 7392 μs t = 8082 μs t = 9262 μs t = 10542 μs

Figure 7 Front View of the burning wafer



t = 238 μs t = 577 μs t = 1076 μs t = 1863 μs t = 2861 μs t = 6431 μs

Figure 8 Side views of the burning wafer

SUMMARY

Several porous silicon samples were etched, varying the etching conditions for each sample. The SEM images of these samples indicate that the samples etched at higher current densities have larger pores in the order of 50 μm and the ones etched with lower current densities and higher HF concentrations have much smaller pores (<10 nm) as expected. The etching rate was estimated for a particular case, and was found to be fairly constant. The samples were impregnated with various oxidizers and the flame propagation was studied. Further work is needed to characterize the samples with lower pore diameters

with FE SEM. Also, the etching rates need to be determined for other etching conditions and the porosities of the samples is to be estimated.

ACKNOWLEDGEMENTS

I would like thank my mentor Dr. Srinivas Tadigadapa and grad student mentor Sharat Parimi for sharing their knowledge and permitting me to work with them. Without their help this research paper would not been possible. I also wish to thank the National Science Foundation (Grant No. EEC-0755081) for funding this research and Penn State University for giving me the opportunity to participate in their EE REU program.

REFERENCES

[1] L.T. Canham, *J. Appl. Phys. Lett*. **57** (1990) 1046

[2] P. McCord, S.-L. Yau, and A.J. Bard, "Chemiluminescence of anodized and etched silicon: Evidence for a luminescent siloxene-like layer on porous silicon," Science, vol. **257**, pp. 68–69, 1992.

[3] D. Kovalev, V.Y. Timoshenko, N. Künzner, E. Gross and F. Koch, "Strong explosive interaction of hydrogenated porous silicon with oxygen at cryogenic temperatures," *Physical Review Letters*, vol. **87**, pp. 68301/1-68301/4, 2001.

[4] Canham, Leigh, Institution of Engineering and Technology, 1997, p. 9.

[5] A. Halimaoui, NOYES publication, 1997 p. 12.

[6] J Escorcia-Garcia, O Sarracino Martínez, J M Gracia-Jiménez and V Agarwal, "Porous silicon photonic devices using pulsed anodic etching of lightly doped silicon," 2009 *J. Phys. D: Appl*. 42 145101 (7 pp)

[7] H. Föll, M. Christophersen, J. Carstensen, G. Hasse, Formation and application of porous silicon, *Materials Science and Engineering* R 39 (2002) 93–141

EXPERIMENTAL CHARACTERIZATION OF NONLINEAR OPTICAL MATERIALS

Katie Knox,[*] Chuan Yang,[†] and Zhiwen Liu[‡]

Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

[*]Undergraduate student of
Department of Engineering Physics
John Carroll University
University Heights, OH 44118

ABSTRACT

An optical system based on a half-wave plate—polarizer attenuator was designed to determine the transmitted light through materials that behave nonlinearly in the presence of high-energy picosecond laser pulses. The sample's nonlinear absorbance of a frequency-doubled Nd:YAG laser was measured as the energy of the incident light was regulated by the rotation of a half-wave plate. The optical system was used to measure the absorption of two linear samples as well as a nonlinear liquid crystal. Experimental results revealed that the optical material, which was composed of gold nanospheres in the nonlinear liquid L34, behaved as expected in the presence of lower input energies, but the attenuated beam did not reach a high enough energy to observe the full limiting effects of the nonlinear material.

INTRODUCTION

Nonlinear Optics

The research conducted in the following experiment lies in the field of nonlinear optics, which concentrates on the behavior and interaction of light with nonlinear materials. The optical phenomena studied within the field are considered nonlinear from the understanding that "they occur when the response of a material system to an applied optical field depends in a nonlinear manner upon the strength of the optical field."[1] The field of nonlinear optics first emerged in the 1960s following P.A. Franken, A.E. Hill, C.W. Peters, and G. Weinreich's first observation of second harmonic generation at the University of

[†] Graduate Mentor
[‡] Faculty Mentor

Michigan. This discovery had to wait until the first laser was successfully developed, since natural light is not intense enough to modify the optical properties of a nonlinear material system.

Nonlinear and Reverse Saturable Absorption

Nonlinear optical materials are unique in that they behave very differently from linear mediums, particularly in the presence of intense electromagnetic waves produced by high-powered lasers. When light passes through nonlinear materials, the output intensity of the transmitted light is not proportional to the input intensity that actually strikes the medium. For this reason, nonlinear materials seem to be the key to optical limiting. Figure 1 shows the theoretical behavior of ideal nonlinear materials when they interact with highly intense light. The intensity of the light transmitted through the medium initially increases until it reaches a maximum output value where no matter how strong the intensity of the incident beam, the energy that is transmitted by the sample remains constant. In other words, at low powers, the transmission increases along with the incident power, but at high input powers, the optical property is limited by the sample at a saturated output intensity.[2]

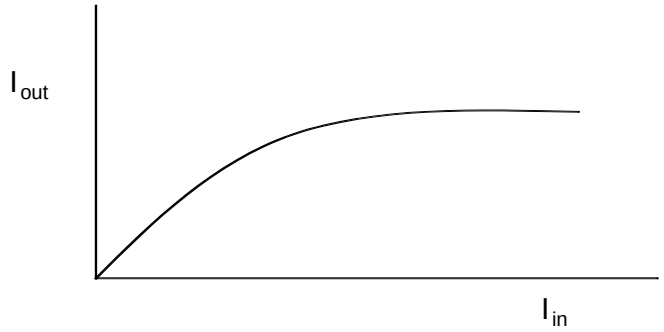


Figure 1: Theoretical model of the behavior of an ideal nonlinear material.

Reverse saturable absorption (RSA) is a photophysical mechanism that occurs on the molecular level of chromophores in optical materials. When visible light hits a chromophore, it can absorb energy by exciting electrons from their ground states into an excited state. When a chromophore exhibits qualities of RSA, an excited state has a larger absorption cross-section, σ_e, than the ground-state cross-section, σ_g. The higher the intensity of the laser, the more molecular states are involved in the absorption. Therefore, if an optical pulse passes through these types of chromosphores, the transmittance decreases as the intensity, energy, and fluence (energy per unit area) increases.[3]

The RSA response of nematic liquid crystals and nonlinear liquids can be enhanced by a variety of dopants to alter the optical properties of the crystal. Gold nanoparticles are an example of one such doping agent that has been found to provide stronger reverse saturable absorption response of carbon allotropic

materials such as fullerene and C_{60} placed inside liquid solvents. The theory is that by exciting the gold nanoparticles, energy will be transferred to the surrounding nonlinear liquid, exciting its own molecular state, and improving its RSA response.[4]

Benefits of Optical Limiting

While a better understanding of nonlinear materials and their different characteristics is a definite benefit for the field of nonlinear optics, there are also a number of avenues in industry that can be impacted by further experimentation on nonlinear mediums. One of the possibilities for such research to be commercialized and transferred to technology lies in the need for protective eyewear, specifically designed for pilots, that can block out all visible and infrared light emitted from lasers of all intensities and pulsing frequencies. Pilots flying aircraft have recently become easy targets for individuals on the ground to aim various types of lasers into the cockpits of commercial airplanes during the most crucial phases of flight, take-off and landing. Since the Federal Aviation Administration (FAA) began accounts of such events in 2004, there have been well over 900 incidents reported nationally.[5] While there are currently goggles on the market that researchers can use to safely work with lasers in the laboratory, each pair only works for a specified range of wavelengths, intensities, and pulses. A better understanding of nonlinear materials seems to be the key to optical limiting and sensory protection, not only for human eyes, but for other sensitive systems as well.

Other possible opportunities for research applications of optical limiters include the transition from optic-electronic-optic communications to optical fiber communications, photonic switching, optical computing and optical signal processing.[6] Each of these areas would exploit the nonlinear properties of different mediums, including optical fibers, to improve the efficiencies and functionalities of a wide range of applications.

THEORY

Polarization of Light

The polarization of electromagnetic waves is an important concept to be considered in understanding the design of the experimental set-up. The polarization of waves is a property that describes the orientation of their oscillations by defining the direction of the wave's electric field. Most natural sources of electromagnetic radiation, such as the sun, are unpolarized. An unpolarized electromagnetic wave is one in which the wave's electric field oscillates randomly in all possible planes or orientations in relation to the wave's direction of propagation. In contrast to unpolarized light sources, high energy lasers emit polarized electromagnetic waves whose electric field vector and directional propagation vector create a specific plane of polarization.[7] However, the plane of polarization, or plane of oscillations for the electric field vector, can

be manipulated by certain optical elements, including wave plates and linear polarizers.

Wave Plates and Linear Polarizers

Wave plates are optical elements that can be used experimentally to rotate the polarization plane of an incident polarized wave. Wave plates, also known as retarders or polarization rotators, change the polarization of electromagnetic waves such as those emitted from a laser by creating a phase difference between the extraordinary (*e*-) and ordinary (*o*-) orthogonal components of the incident beam as it passes through the plate. The resulting phase difference between the *e*- and *o*- waves depends on the type of wave plate used. For example, the quarter-wave plate introduces a phase shift of $\pi/2$ radians, whereas the half-wave plate introduces a phase shift of π radians or 180°. As shown in Figure 2, when the wave plate is rotated along its slow axis (axis parallel to the incident beam), the polarization plane will make an angle θ with the fast axis (axis perpendicular to the incident beam). Once the polarized light has passed through the plate, the wave will have been rotated by an angle 2θ.[8]

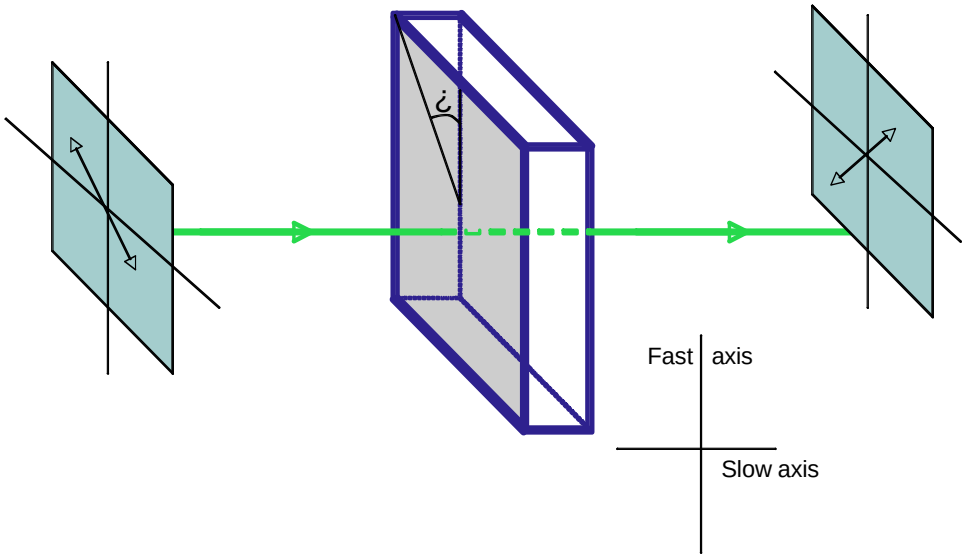


Figure 2: Half-wave plate. The initial plane of polarization is rotated so that wave which first oscillated in the first and third quadrants now oscillates in the second and fourth quadrants.[8]

The linear polarizer is another type of optical element that has different effects on the polarization states of electromagnetic waves. One of the basic functions of linear polarizers is to act like a filter. It separates the two components or polarization states of incident light, discarding one and passing on the other. The polarizing beamsplitter, shown in Figure 3, creates two orthogonal, linearly

polarized beams. In this experiment, the beamsplitter allows the horizontal components of the polarized wave to be transmitted, whereas the vertical components of the incident light are reflected. The reflected light is also known as the s-polarized beam and the transmitted light is also referred to as the p-polarized beam.

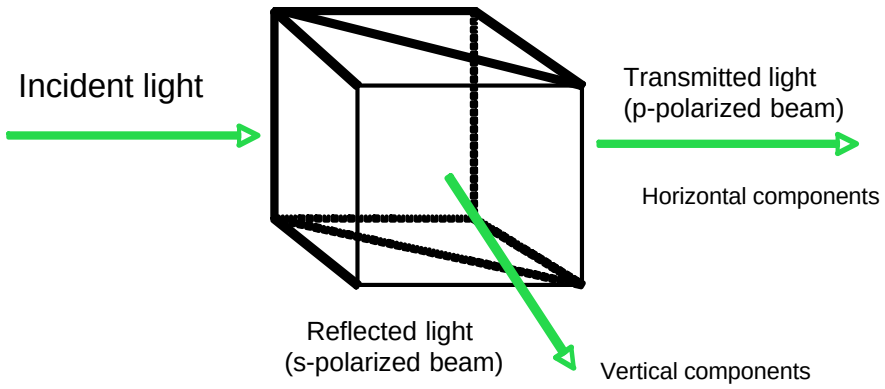


Figure 3: A typical polarizing beamsplitter that selectively transmits one component of light while reflecting the other.

Malus's Law

Malus's Law was theorized in 1809 by Étienne Malus after studying the effects of placing two linear polarizers in series. In his experiment, natural light was directed at a linear polarizer tilted at an angle θ with respect to the vertical, and was then passed through a second polarizer, or analyzer.[8] According to this law, the intensity of the light transmitted through the first polarizer is I_0, but after it passes through the analyzer, the intensity of the light is[7]

$$I_{\text{transmitted}} = I_o \cos^2\theta \text{ (incident light polarized).} \quad (1)$$

With respect to the following experiment conducted to characterize nonlinear optical materials, a half-wave plate will be used in place of a linear polarizer since the incident electromagnetic waves passing through it are already polarized from a pulsed laser. Despite this modification, Malus's Law is still applicable since the half-wave plate will rotate the light's plane of polarization. A polarizing beamsplitter will also be equivalent to the analyzer, such that the intensity of the light transmitted by the beamsplitter will be defined by Equation 1. Since the total irradiance of the light is conserved and $I_{\text{reflected}}$ is the light reflected by the beamsplitter, the sum of $I_{\text{transmitted}}$ and $I_{\text{reflected}}$ should equal I_o. Therefore, the intensity of the light reflected by the beamsplitter is

$$I_{\text{reflected}} = I_o \sin^2\theta. \quad (2)$$

EXPERIMENT

Optical Material Samples

The first two optical samples that were tested included toluene, an aromatic hydrocarbon, also known as methylbenzene and phenylmethane. One sample contained only the clear liquid toluene solution, but another sample also contained gold nanospheres, resulting in a dark purple colored solution. These two linear samples were used to observe how the gold nanospheres alter the absorption properties of the hydrocarbon.

The third sample tested was made out of L34, a nematic liquid crystal fabricated by researchers at The Pennsylvania State University, and gold nanoparticles, used as the dopant. This nonlinear sample was tested to observe the sample's limiting effects of RSA in the presence of high-energy picosecond pulses.

Optical Setup

A specific optical system was created and employed for the purpose of determining the transmitted energy of light through each optical sample. Figure 4 shows a schematic diagram of the experimental system. The light that passed through the system was split from a high-energy solid state Nd:YAG laser. The frequency of the picosecond pulsing laser was doubled to produce a wavelength of 532 nm. A high-energy beamsplitter was used to separate a portion of the electromagnetic wave required for the experiment, while directing the excess energy into a dumper. The remaining fraction of the laser that was transmitted through the high-energy beamsplitter was subsequently directed into the system by a high intensity mirror. After first being narrowed by an initial Iris aperture, the beam then passed through a half-wave plate, a polarizing beamsplitter, a second aperture, and then two focusing lenses. The purpose of the two lenses was to converge the beam at a focal point where the optical samples would be placed during the experiment.

Two photo detectors also shown in Figure 4 were included in the system and connected to a ratio meter so that the energy of the light reflected from the beamsplitter, as well as the energy of the transmitted light, could be measured and averaged over a series of 10 pulses. An external pulse trigger was also needed to signal the cycle of data collection to begin, so the ratio meter was triggered from the output of the laser itself.

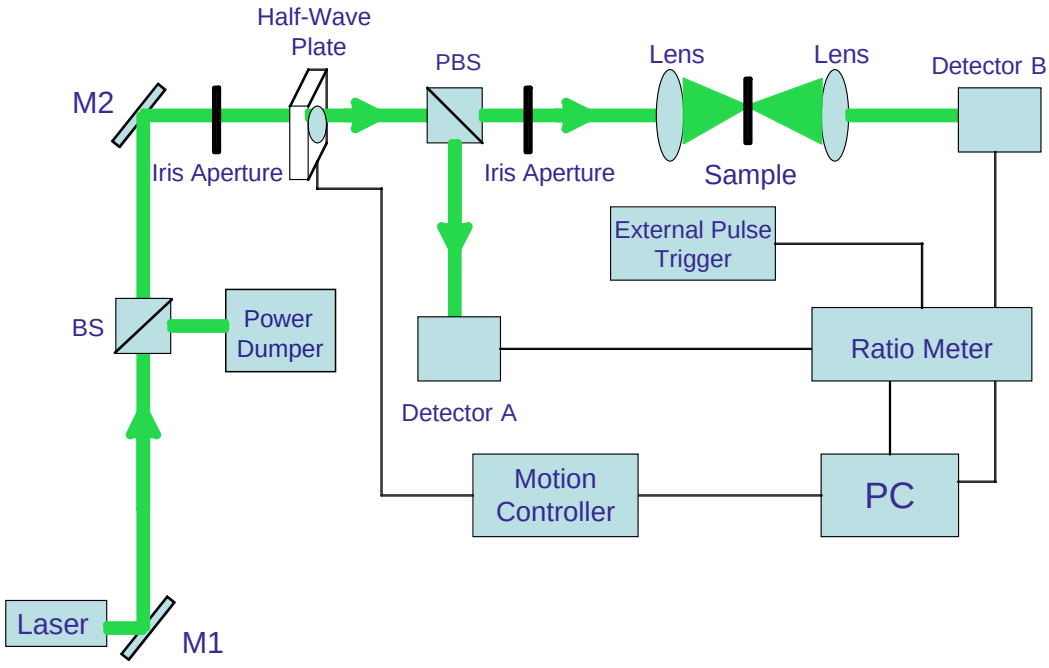


Figure 4: Schematic diagram of the initial experimental setup.

The key feature of this optical set-up was the external attenuator created by the combination of the half-wave plate and polarizing beamsplitter. As Y. Zhao et. al. describe in their paper, the attenuation of polarized light, or reduced intensity of electromagnetic radiation due to the absorption or scattering of photons, can be achieved by using a half-wave plate together with a linear polarizer.[9] The half-wave plate was mounted inside a rotational stage that was connected to a motion controller. The motion controller was also connected to a computer through a General Purpose Interface Bus (GPIB) card. Then through the GPIB interface, the rotation of the half-wave plate was controlled by a command code designed in MATLAB.

The code was designed to turn the rotational stage and half-wave plate *n* degrees at a time, for a total of 180 degrees of rotation, which represented two full periods of attenuation. As the half-wave plate was rotated to different angles along its slow axis, the polarization state of the wave changed, impacting the vertical and horizontal components of the polarized wave. In other words, when the components of the light that passed through the polarizing beam splitter varied, the intensity of the transmitted and reflected waves of the linear polarizer also varied.

For every *n* degrees of rotation, the ratio meter collected readings from both photo detectors to determine the intensity of the horizontal and vertical components of the wave at each angle. Since the ratio meter was also connected to a computer code generated in MATLAB through an IEEE interface, the data was stored in matrices that could be analyzed after each experimental run.

By doing preliminary runs of the experiment without the optical sample in place, detector B collected measurements representative of the intensity of the energy that would be incident upon the medium once it was inserted into the system. As a result, each energy measurement from detector B would correspond to a relative energy measured by detector A. Therefore, once the sample was put in place, the energy of the light transmitted by the sample could be matched with the incident energy by comparing data collected by detector A, enabling the observation of each sample's absorption patterns.

RESULTS AND DISCUSSION

The first experimental runs were executed without the optical samples in order to determine if the attenuation system would operate according to theoretical predictions. As shown by the graph displayed in Figure 5, measurements taken from the two photo detectors yielded sine and cosine curves that were theoretically predicted by Malus's Law. The energy measurements collected by detector B are relatively lower than those measured by detector A because the second aperture narrowed the diameter of the beam transmitted by the polarizing beamsplitter, causing the additional reduction in the beam's energy.

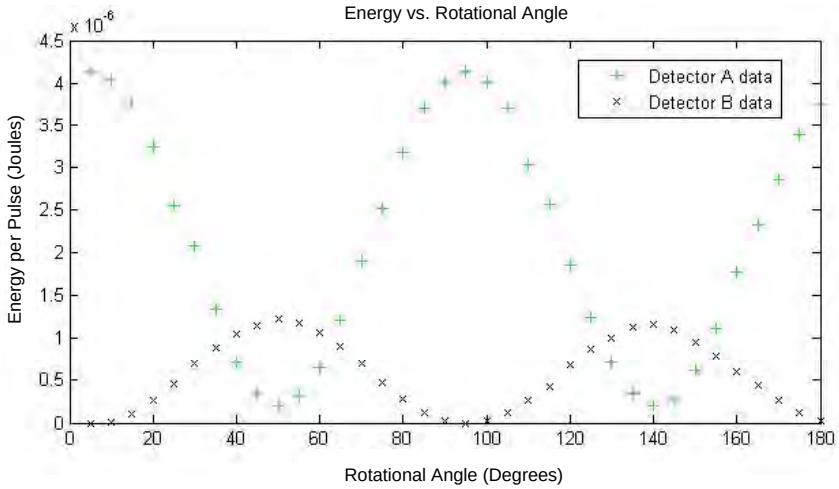


Figure 5: One example of a plot generated by a preliminary run to calibrate the system before the sample was put in place. The half-wave plate was rotated every 5° for a total of 180°. The graph corresponds to the theoretical predictions described by Malus's Law.

After a number of preliminary runs were completed, yielding data similar to that displayed in Figure 5, a test sample was put in place at the focal point between the two lenses in order to determine if the energy emitted by the laser would be strong enough to pass through the medium and be measured by detector B. The sample was tested, only to reveal that after rotating the half-wave plate through a full 180°, the light was completely absorbed by the sample and never

reached the second photo detector. Therefore, it was determined that modifications needed to be made to increase the intensity of the laser that entered the system and the half-wave plate was relocated. It was moved from its initial position to the opening between the first high intensity mirror and beamsplitter, whose locations are shown in Figure 4.

The test sample was again used for a trial at this increased energy, however, when the modified system was experimented, the laser beam was too strong and surpassed the damage threshold of the glass sample cell holder. There was determined to be a narrow range of 12° every $\pi/4$ radians within which the glass would be strong enough to withstand the irradiance of the laser. Since this range was too small for the purpose of the experiment, an additional modification needed to be made.

The intensity of the electric field at the focal point, the location of the sample, was determined to be too energetic for the glass to handle so the second lens was removed from the system. The sample was then placed at a location in front of the focal point so that the slightly larger radius of the beam at this position would not burn through the glass. Figure 6 represents a portion of the final experimental set-up used to determine the transmitted intensities through the nonlinear sample as well as the two linear samples used for comparison.

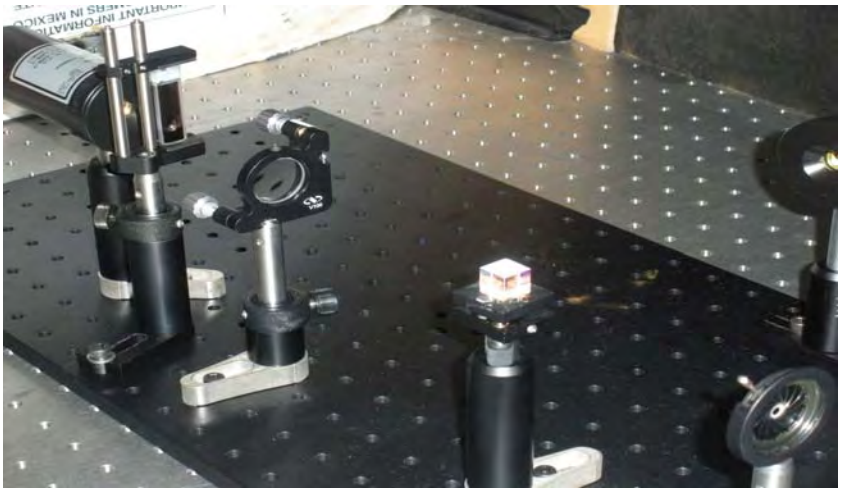


Figure 6: A photograph illustrating a portion of final experimental system. The laser enters the aperture (bottom right) and then hits the polarizing beamsplitter which then either reflects the beam into detector A or passes it on to the lens, followed by the sample and detector B.

Using the modified optical system, the absorption of energy by the two samples containing toluene was observed. The experimental run conducted without the sample was used to calibrate the system and measure the input intensity that would strike the samples. Figure 7 shows one set of the three

experimental runs plotted against the half-wave plate's angle of rotation. For each run, the wave plate was rotated every 10° for a total of 180°.

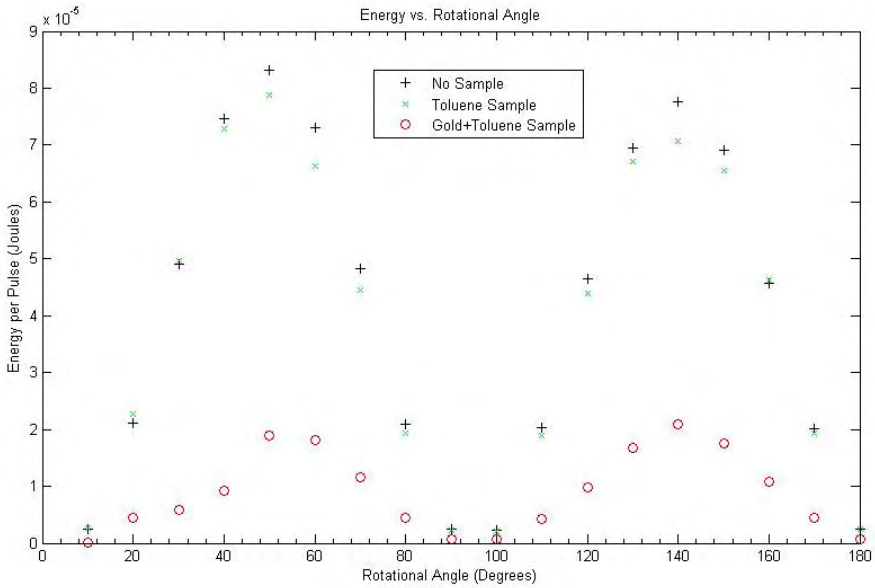


Figure 7: Energy measurements recorded by detector B in relation to the half-wave plate's rotated angle. The data collected without a sample present also represents the intensity of the laser that was incident upon the linear and nonlinear samples.

As expected, the sample containing the gold nanoparticles significantly increased and enhanced the material's absorption of the incident energy emitted from the picosecond pulsing laser. In order to show the magnitude of the difference in the samples' absorptions, the transmitted energy that passed through each sample was graphed against the input energy. A graph of this data is shown in Figure 8. While the output energy increased proportionally to the input energy, the sample that contained both gold and toluene was determined to have a much stronger absorption of higher incident energies than the sample containing only toluene.

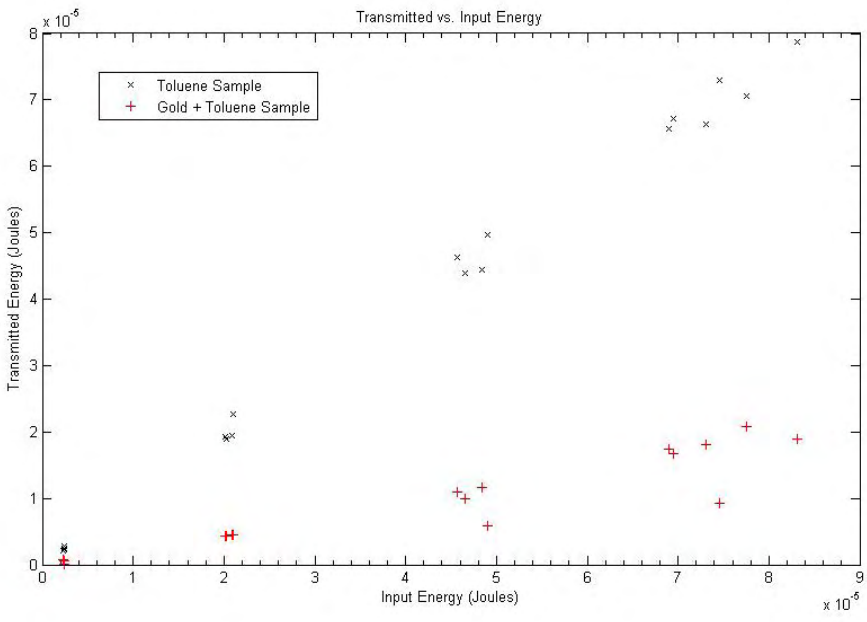


Figure 8: The transmitted energy of two linear samples containing toluene as a function of the incident energy.

The nonlinear sample containing L34 and gold nanoparticles was then tested by the optical system. As shown in Figure 9, the transmitted energy through the nonlinear sample was also graphed as a function of the input energy. According to the data, the energy incident upon the sample was not high enough to reach the level of saturated absorption, where the transmitted energy becomes constant. Due to the damage threshold of the glass containing the liquid, the optical system was unable to deliver input energies higher than around 1.085×10^{-4} Joules to the nonlinear L34-gold sample. At the highest input energies as plotted by *Data2* in Figure 9, it appears that the sample was close to reaching its maximum transmitted energy value, but the sample's molecular properties were not excited to a high enough energy state to observe its full optical limiting power.

The transmission of energy through the nonlinear sample was also graphed to show how the transmittance changed as the input energy increased. As evident by the data in Figure 10, the transmission appeared to be constant throughout the entire range of energy levels, and did not decrease with an increase in the beam's intensity. Therefore, the intensity of the beam was not yet high enough to induce the RSA of the nonlinear material.

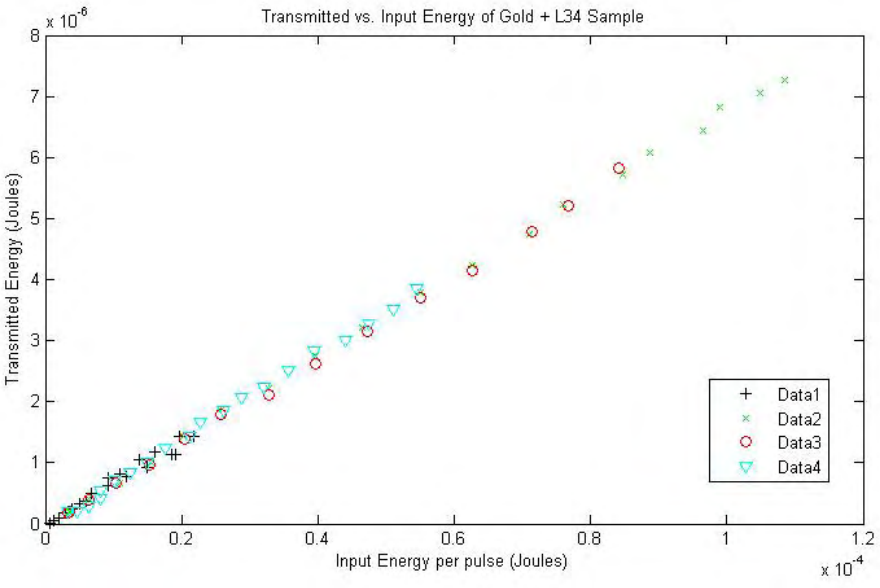


Figure 9: Data from four experimental runs revealing the low-energy range where the sample still behaves linearly.

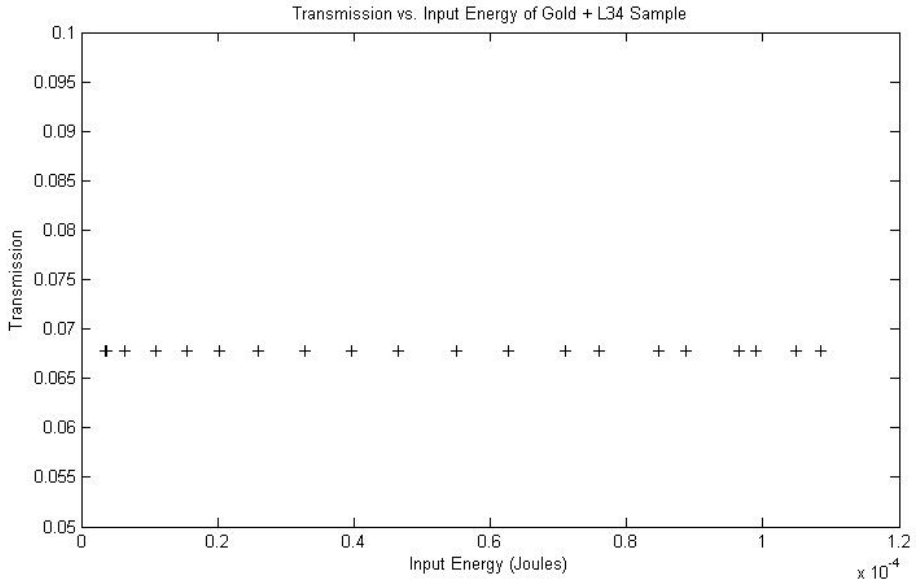


Figure 10: The measured transmission of the nonlinear sample.

CONCLUSION

The influence of gold nanoparticles on the toluene's optical absorption was successfully observed by the experiment. The presence of the gold nanoparticles significantly enhanced the material's absorption of 532 nm picosecond pulses.

The optical system was also successful in measuring the linear absorption of the L34-gold sample at low-energy pulses, and by improving the system to achieve an input energy high enough to induce the sample's RSA characteristics, it will also be possible to more fully observe the optical limiting qualities of the L34-gold nonlinear sample.

ACKNOWLEDGMENTS

KK would like to thank the Department of Electrical Engineering at The Pennsylvania State University for supporting her participation in the 2009 REU program. Special thanks to Kebin Shi, Haifeng Li, and members of I. C. Khoo's research team, Andres Diaz, Junbin Huang, and Mike Stinger, for helpful discussions. Also, thank you to J. Huang for providing the samples used for the experiment. This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081.

REFERENCES

[1] R. W. Boyd, *Nonlinear Optics*, 2nd ed. Academic Press, San Diego, 2003.

[2] M. I. Miah, "Stimulated photoluminescence and optical limiting in CdI_2," *Optical Materials* **20** (4) 279-282, (2002).

[3] H. S. Nalwa and S. Miyata, *Nonlinear Optics of Organic Molecules and Polymers*, CRC Press, (1996).

[4] V. Amendola, G. Mattel, C. Cusan, M. Prato, and M. Meneghetti, "Fullerene non-linear excited state absorption induced by gold nanoparticles light harvesting," *Synthetic Metals* **155** (2) 283-286, (2005).

[5] *'Pilot' Laser Study Launched*, WWW Document, (http://www.photonics.com/Content/ReadArticle.aspx?ArticleID=30198).

[6] H. A. Abdeldayem, D. O. Frazier, B. G. Penn, D.D. Smith, and C. E. Banks, "Non-linear optothermal properties of metal-free phthalocyanine," *Thin Solid Films* **350** (1-2) 245-248, (1999).

[7] R. D. Knight, *Physics for Scientists and Engineers: A Strategic Approach*, Addison Wesley, San Francisco, 2004.

[8] E. Hecht, *Optics*, 4th ed. Addison Wesley, San Francisco, 2002.

[9] Y. Zhao, Y. Jia, J. Yang, and X. Zhu, "Influence of a Half-Wave Plate—Polarizer Attenuator on Broadband Femtosecond Laser Pulses," *Optical Engineering* **46** (4) 1-6, (2007).