

Annual Research Journal

Electrical Engineering
Research Experience for Undergraduates

Vol. IX

August 2011



National Science Foundation
Grant No. EEC-1062984

PENNSTATE



Department of Electrical Engineering
University Park, Pennsylvania

EEREU Annual Research Journal

Sven Bilén (editor)

Volume IX

Published in 2011 by

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

NSF EE REU Site Program Contact

121 Electrical Engineering East
The Pennsylvania State University

University Park, PA 16802

Telephone (814) 863-6315

Fax (814) 865-7065

E-Mail: eereu@engr.psu.edu

Web Site: <http://www.ee.psu.edu/reu/>



Summer REU
Penn State
ELECTRICAL ENGINEERING

PREFACE

We are happy to present the ***2011 Annual Research Journal – Electrical Engineering Research Experience for Undergraduates, Vol. IX.*** This volume contains highlights of the EEREU program activities for summer 2011 and thirteen technical papers written by EEREU scholars as primary authors.

Summer 2011 marks our ninth 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 2011 Annual *EEREU Symposium*, held at University Park, in July 2011.

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 EE REU Site Program
Dept. of Electrical Engineering
The Pennsylvania State University

*July 2011
University Park, PA*

TABLE OF CONTENTS

2011 NSF EEREU Faculty and Staff Members	iii
2011 NSF EEREU Summer Program Research Seminar Program	iv
2011 NSF EEREU Summer Program Ethics and Entrepreneurship Workshops	v
2011 NSF EEREU Field Trips and Academic/Industrial Sponsors and Tour Hosts	vi
2011 NSF EEREU Annual Symposium Program	vii

Research Articles (* indicates REU student author)

TITLE OF ARTICLE	Author	Page
POTENTIAL EFFECTS OF THUNDERCLOUD INDUCED RADIATION BURSTS ON AIRCRAFT PASSENGERS AND CREW <i>Christopher A. Curwen, * Sebastien Celestin, Wei Xu, and Victor P. Pasko</i>	1
ACCURATE MODELING OF ION CONDUCTIVITY IN THE EARTH'S ATMOSPHERE <i>Sean S. McDonough, * Jianqi Qin, and Victor P. Pasko</i>	13
STUDY OF DOPPLER VELOCITY ESTIMATION TECHNIQUES ON METEOR-HEAD RADAR REFLECTIONS <i>Christopher Hong, * Julio Urbina, and Freddy Galindo</i>	25
STUDY OF PROCESSING TECHNIQUES FOR REMOVAL OF EQUATORIAL ELECTROJET ECHOES TO ENHANCE METEOR DETECTION AT JICAMARCA <i>Dimitri Ressetar, * Julio Urbina, and Freddy Galindo</i>	41
CONSIDERATIONS FOR AN INTRA-SOLAR SYSTEM LASER SATELLITE DATA NETWORK <i>Christopher W. Jones* and John D. Mathews</i>	57
SYSTEM DEVELOPMENT AND INTEGRATION OF COMMUNICATION AND POWER LINK BETWEEN OLITE 2 SATELLITE AND HIGH ALTITUDE STUDENT PLATFORM <i>Christopher A. Galvan, * Sven G. Bilén, Robert Capuro, and Allen Kummer</i>	67

DEVELOPMENT OF A CHARGING METHOD FOR THE GEOPEBBLES USING WIRELESS POWER TRANSMISSION	87
<i>Luis E. Olique Ortiz, * Sven G. Bilén, Robert Capuro, Michael Conway, and Tyler Boehmer</i>		
ROBUSTNESS AND FAULT TOLERANT CAPABILITIES OF TRANSFORM DOMAIN FIR FILTERS WORKING ON REAL SIGNALS	97
<i>Danielle Sova, * W. Kenneth Jenkins, David Salvia</i>		
VERIFICATION OF A MARGINAL OSCILLATOR FOR CONTINUOUS-WAVE QUADRUPOLE RESONANCE SPECTROSCOPY	117
<i>Jacob Wilson, * Jeffery Schiano, and Thomas Tyson</i>		
OPTICAL AND ELECTRICAL CHARACTERIZATIONS OF FREE STANDING MICROELECTROMECHANICAL STRUCTURES	129
<i>Bahareh Adami Ardestani, * Srinivas Tadigadapa, and Hwall Min</i>		
GROWTH AND OPTICAL ANALYSIS OF BRANCHING SILICON WIRE ARRAYS	141
<i>Christina DiMarino, * Chito Kendrick, and Joan Redwing</i>		
DEVELOPMENT OF A NEAR-FIELD SCANNING OPTICAL MICROSCOPE SYSTEM	157
<i>Ruth Nan, * Chuan Yang, Hans Hallen, and Zhiwen Liu</i>		
MRI MICROCOILS FOR IMAGING INDIVIDUAL CELLS	169
<i>Matthew Feldman, * Michael Lanagan, and Steven Perini</i>		
<i>Author Index</i>	181

2011 NSF EEREU FACULTY & STAFF MEMBERS

Faculty Mentors:

Prof. Sven Bilén, Co-Director
Prof. Ken Jenkins, Co-Director
 Prof. I. C. Khoo
Prof. Michael Lanagan
 Prof. Zhiwen Liu
 Prof. John Mathews
 Prof. Joan Redwing
 Prof. David Salvia
 Prof. Victor Pasko
 Prof. Jeff Schiano
Prof. Srinivas Tadigadapa
 Prof. Julio Urbina

Ms. Amy Freeman, Adviser, Engineering Diversity
Prof. Ruyan Guo, Program Advisor (Former Director)
 Mrs. Lena Getman, Administration
 Prof. Phil Boyer, 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 2011 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 7, 2011	Systems Design Lab	Sven Bilén
June 14, 2011	Integrating Nanowire-based Sensors with CMOS Electronics Using Directed Self Assembly	Theresa Mayer
June 21, 2011	Magnetic Resonance Imaging	Jeff Schiano
June 28, 2011	The Wild Historical Tug-a-War Between DSP Theory and IC Technology	Ken Jenkins
July 5, 2011	What's Next? Graduate School, of Course!	David Salvia
July 12, 2011	Lightning-related Transient Luminous Events in the Middle Atmosphere	Victor Pasko
July 19, 2011	Emerging Techniques in Radar Remote Sensing: Cognitive Radars and Beyond	Julio Urbina
July 28, 2011 <i>(Thursday)</i>	2010 NSF EEREU Symposium	REU Students

NSF EEREU 2011 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

<i>Date</i>	<i>Topic</i>	<i>Speaker</i>
June 7, 2011	Frameworks of Ethics	Andy Lau
June 14, 2011	Research Ethics	Andy Lau
June 21, 2011	Intellectual Property Issues	Phil Boyer
June 28, 2011	Sustainability and Materials Selection in Product Development	Andy Lau
July 5, 2011	Student Entrepreneur Panel: “What’s it take to be an entrepreneur?”	Erik Davidson
July 12, 2011	Ethics Issues in Entrepreneurship	Andy Lau
July 19, 2011	Getting Funding to Pursue Your Idea	Phil Boyer
July 26, 2011	Student Presentations on Tech Transfer Ideas of Research Projects	REU students

2011 NSF EE REU Field Trip Program
Academic/Industrial Sponsors and Tour Hosts

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

Applied Research Laboratory
Penn State University, University Park, Pennsylvania

Breazeale Nuclear Reactor
Penn State University, University Park, Pennsylvania

RTD Embedded Technologies Inc.
State College, Pennsylvania

Accuweather
State College, Pennsylvania

Videon Central, Inc.
State College, Pennsylvania

AlphaLabs
Pittsburgh, Pennsylvania

MAYA Design
Pittsburgh, Pennsylvania

2011 NSF EE REU SYMPOSIUM

Final Program

8:30 am to 3:00 pm, Thursday, July 28, 2011
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)	(Lena Getman/ Tom Tyson)
8:55 – 9:00 am	Welcome	Ken Jenkins
9:00 – 10:00 am	Session I	Session Chairs: Pasko/Urbina
9:00– 9:15	POTENTIAL EFFECTS OF THUNDERCLOUD INDUCED RADIATION BURSTS ON AIRCRAFT PASSENGERS AND CREW (MENTOR: PASKO)	Christopher Curven
9:15 – 9:30	ACCURATE MODELING OF ION CONDUCTIVITY IN THE EARTH'S ATMOSPHERE (MENTOR: PASKO)	Sean McDonough
9:30 – 9:45	STUDY OF DOPPLER VELOCITY ESTIMATION TECHNIQUES ON METEOR-HEAD RADAR REFLECTIONS (MENTOR: URBINA)	Christopher Hong
9:45 – 10:00	STUDY OF PROCESSING TECHNIQUES FOR REMOVAL OF EQUATORIAL ELECTROJET ECHOES TO ENHANCE METEOR DETECTION AT JICAMARCA (MENTOR: URBINA)	Dimitri Ressetar
10:00 – 10:15 am	Coffee Break	
10:15 – 11:30 am	Session II	Session Chairs: Mathews/Bilén Jenkins/Schiano
10:15 – 10:30	CONSIDERATIONS FOR AN INTRA-SOLAR SYSTEM LASER SATELLITE DATA NETWORK (MENTOR: MATHEWS)	Christopher Jones
10:30 – 10:45	SYSTEM DEVELOPMENT AND INTEGRATION OF COMMUNICATION AND POWER LINK BETWEEN OLITE 2 SATELLITE AND HIGH ALTITUDE STUDENT PLATFORM (MENTOR: BILEN)	Christopher Galvan
10:45 – 11:00	DEVELOPMENT OF A CHARGING METHOD FOR THE GEOFEBBLES USING WIRELESS POWER TRANSMISSION (MENTOR: BILEN)	Luis Olique
11:00 – 11:15	ROBUSTNESS AND FAULT TOLERANT CAPABILITIES OF TRANSFORM DOMAIN FIR FILTERS WORKING ON REAL SIGNALS (MENTORS: JENKINS/SALVIA)	Danielle Sova
11:15 – 11:30	VERIFICATION OF A MARGINAL OSCILLATOR FOR CONTINUOUS-WAVE QUADRUPOLE RESONANCE SPECTROSCOPY (MENTOR: SCHIANO)	Jacob Wilson

2011 NSF EE REU SYMPOSIUM (cont.)

8:30 am to 3:00 pm,

Final Program

Thursday, July 28, 2011

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
1:45 – 2:00	OPTICAL AND ELECTRICAL CHARACTERIZATIONS OF FREE STANDING MICROELECTROMECHANICAL STRUCTURES (MENTOR: TADIGADAPA)	Bahareh Ardestani
2:00 – 2:15	GROWTH AND OPTICAL ANALYSIS OF BRANCHING SILICON WIRE ARRAYS (MENTOR: REDWING)	Christina DiMarino
2:15 – 2:30	DEVELOPMENT OF A NEAR-FIELD SCANNING OPTICAL MICROSCOPE SYSTEM (MENTORS: LIU)	Ruth Nan
2:30 – 2:45	MRI MICROCOILS FOR IMAGING INDIVIDUAL CELLS (MENTOR: LANAGAN)	Matthew Feldman
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 Lena Getman	
Special Notes:	Laser pointer, desktop PC and LCD PC projector are provided. Speakers please upload presentation files to EEREU group web in Angel by Wednesday, July 27. It is strongly recommended that speakers test-run presentation files for audio, video, or special applications <u>on a Windows PC</u> prior to their Thursday presentation. Each presentation is 15 minutes – chair introduction, questions, and answers included.	

POTENTIAL EFFECTS OF THUNDERCLOUD INDUCED RADIATION BURSTS ON AIRCRAFT PASSENGERS AND CREW

Christopher A. Curwen,* Sebastien Celestin,[#] Wei Xu,⁺ and Victor P. Pasko[#]

Communications and Space Sciences Laboratory
Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

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

ABSTRACT

In the past 20 years, large research efforts have been directed at studying radiation emissions originating from within and around thunderclouds. The mechanisms behind such events are still up for debate; however, characterization of the radiation produced by these events is sufficient to develop computer models of these events. Although it is widely accepted that aircraft crew and passengers are subjected to radiation mostly due to cosmic rays, the effects of radiation originating from thunderstorms have been studied in only one article [Dwyer *et al.*, 2010]. In the present paper, we investigate the potential effects of exposure to thundercloud radiative events on aircraft crew and passengers. A Monte Carlo simulation of photon transport through the atmosphere is developed and implemented to model radiation given off by thunderclouds and lightning events. It is found that the maximum potential effective dose from energetic photons is ~0.014 mSv and represents no danger to air travelers.

INTRODUCTION

Thundercloud Radiative Events

Theories describing energetic radiation produced in association with thunderclouds have been investigated since C.T.R. Wilson [1925] first suggested such a phenomenon in 1925. However, it was not until 1996 that Eack *et al.* [1996] successfully detected such radiation for the first time using scintillation

⁺ Graduate Mentor

[#] Faculty Mentor

detectors flown through thunderclouds. Since this discovery, many radiation measurements related to thunderstorm activity have been recorded. In fact, ground-based measurements have shown a correlation between the stepping process in the lightning discharge propagation, and the production of X-ray bursts [e.g., *Moore et al.*, 2001; *Dwyer et al.*, 2005]. In addition to those measurements, terrestrial gamma-ray flashes (TGFs) have been serendipitously discovered by satellites initially launched to observe celestial gamma-ray sources, and these events have been correlated to thunderstorm activity [*Fishman et al.*, 1994].

It is commonly agreed upon that the radiation observed from thunderclouds is the result of bremsstrahlung produced by high-energy electrons, called runaway electrons. Runaway electrons are electrons with high energy and therefore low probability of collision with gas molecules, propagating in an applied electric field so that the energy they acquire from the field is higher than the energy losses due to collisions. These electrons are therefore capable of efficiently gaining energy from an electric field in air. This relationship is demonstrated in Figure 1. One can distinguish between thermal runaway processes, for which a very high electric field E exceeds the friction force at low-energy (~ 100 eV) and brings electrons to regimes where they continuously accelerate ($E > E_c \sim 240$ kV/cm in air at ground pressure), and relativistic runaway processes [*Gurevich et al.*, 1992] for which initial high-energy electrons are already present (e.g., cosmic ray secondary electrons) and can initiate relativistic runaway electron avalanches (RREAs) in electric fields higher than $E_t \sim 2$ kV/cm in air at round pressure.

Since energetic radiation is produced in thunderclouds, *Dwyer et al.* [2010] studied its potential danger for aircraft passengers and crew members. This study by *Dwyer et al.* [2010] focused primarily on the radiation of electrons and used a model for thundercloud radiative events in which runaway electrons were generated in a circular area with a varying diameter that measured up to 2 km, and assumed an avalanche multiplication factor. *Dwyer et al.* [2010] concluded that these emissions may be dangerous depending on the geometry of the runaway electron source.

Recent studies have suggested that TGFs and X-ray bursts might be both generated from lightning discharges without invoking further amplification in RREA [*Celestin and Pasko*, 2011]. In this paper, we study the potential effects of energetic photons emitted by thundercloud radiative events on aircraft crew and passengers by considering that photons are produced at the tip of a lightning leader, which we approximate as a source point. Specifically, we model the production of energetic radiation (X-ray bursts and TGFs) from lightning discharges based on results from [*Celestin and Pasko*, 2011].

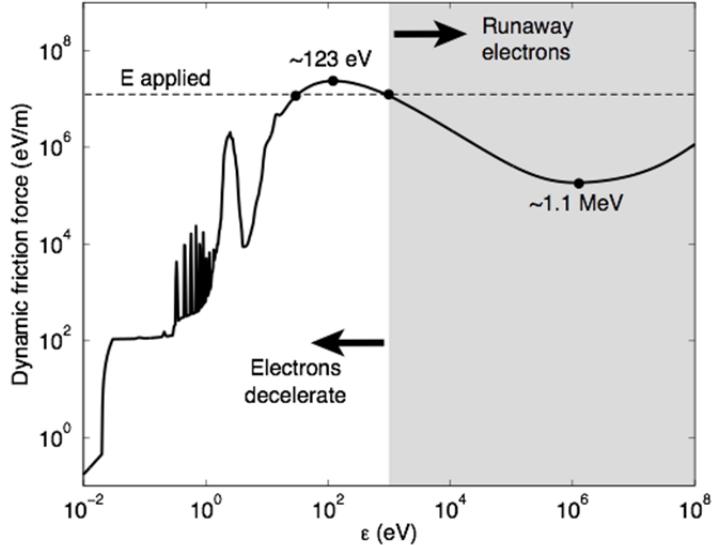


Figure 1: Friction force on energetic electrons moving through air at ground pressure.

Radiation Dose Calculations

It is difficult to specify a definitive effect of human health due to radiation exposure, however, an SI measurement of such effects has been established by the International Bureau of Weights and Measurements. This measurement is named ‘equivalent dose’ and is measured in Sieverts. Equivalent dose is a weighted summation based on the tissue exposed and the type of radiation. These relationships are described by the following equations [Pelliccioni, 2000]:

$$H_T = \sum_R \omega_R D_{T,R} \quad (1)$$

where H_T is the equivalent dose received by tissue T , $D_{T,R}$ is the average absorbed dose from radiation R in tissue T , and ω_R is a quality factor determined for different types of radiation at different energies. The equivalent dose for an organism is then a weighted sum of the equivalent doses in all tissues and organs in the organism [Pelliccioni, 2000]:

$$E^* = \sum_T \omega_T H_T \quad (2)$$

where E^* is the effective dose for an organism, ω_T is the weighting factor for tissue T , and H_T is the equivalent dose received by tissue T .

Calculating the effective dose received by the human body using Equations (1) and (2) is a complicated process. In this paper, we use a conversion factor

rigorously calculated by *Pelliccioni* [2000] using Equations (1) and (2). The effective dose received by the human body is then given by:

$$E^* = h_E \Phi_{re} \quad (3)$$

where E^* is the effective dose, Φ_{re} is the flux of radiation, and h_E is the conversion factor.

Radiation Risks

As described in *Barish* [2004], risks involved with radiation exposure are typically categorized into ‘stochastic’ and ‘nonstochastic’ risks. Stochastic risks are those characterized by a direct, linear relationship between exposure and probability of occurrence. There is no threshold for stochastic effects to occur, and the probability model does not distinguish between severity of effects. For fetuses (those most sensitive to radiation), an exposure to 1 mSv causes an estimated 1/10,000 increased chance that the unborn child will experience stochastic effects as a child [*Barish*, 2004]. For adults, an exposure to 10 mSv causes an estimated 5/10,000 increased chance that he/she will experience stochastic effects [*U.S. Environmental Protection Agency*, 2011]. Common stochastic effects include cancer and hereditary effects.

Nonstochastic risks are characterized by severity of effect as a function of dose. Nonstochastic risks include nausea, hair loss, and vomiting, as well as miscarriage, mental retardation, and congenital malformations for unborn fetuses [*Barish*, 2004; *Martin*, 2003]. While there is no guaranteed safe limit for nonstochastic risks, the threshold for nonstochastic effects on fetuses is generally accepted to be ~20 mSv [*Barish*, 2004], and the threshold for nonstochastic effects on adults is commonly accepted as ~0.5 Sv [*U.S. Environmental Protection Agency*, 2011].

MODEL DESCRIPTION

Photon Collisions

To simulate the radiation emitted by a thundercloud event, we simulate the transport of energetic photons through the atmosphere and calculate the fluence of these photons through different surfaces. A Monte Carlo code is used to simulate the transport of these energetic photons. The modeling of such transport is based on the three main types of collisions experienced by energetic photons occurring with corresponding probabilities: photoelectric absorption, Compton scattering, and electron-positron pair production. Given the number, energies, beam geometry, and initial coordinates of a source of photons, their paths and energies were tracked through a three dimensional space and the resultant energy distributions and radiation fluence could be calculated for any given location.

The relative probability of a collision is given by the so-called collisional cross-section. The cross-section of a given collision type varies with the energy of the photon. The relevant cross-sections are plotted in Figure 2. To determine the relative probability of a given collision, we simply divide the collision's cross-section by the sum, of all three cross-sections.

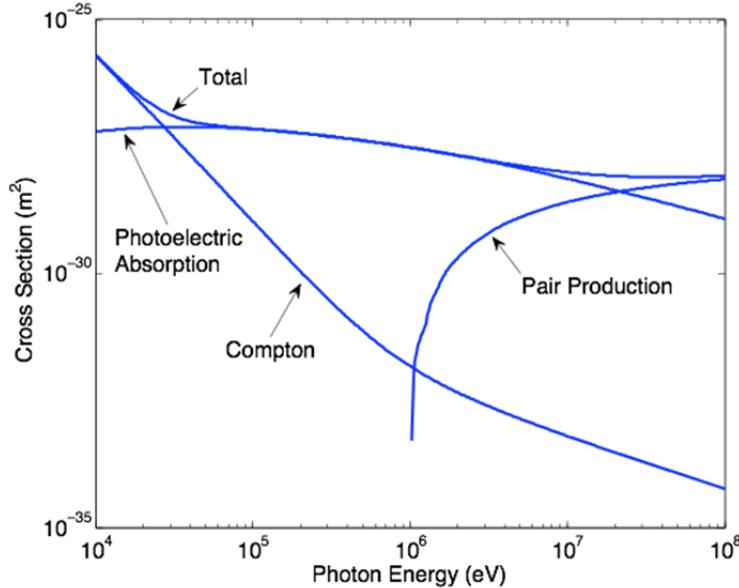


Figure 2: Cross sections for energetic photon propagating through air (80% N₂ + 20% O₂)

As demonstrated by Figure 2, photoelectric absorption is the dominant collision type for photon energies less than ~ 30 keV. This is the process by which a photon is fully absorbed by a molecule, transferring its energy to the molecule, which frequently results in the ejection of a K-shell electron from the molecule. To determine the probability that a photon is absorbed, the corresponding atomic cross section is calculated analytically using the following relationships [Lehtinen, 2000]:

$$\sigma_P = \sigma_0 F \left(\frac{3}{2} \right) \left(\frac{Z_a^5}{134^4} \right) \left(\frac{mc^2}{\epsilon_{ph}} \right)^5 (\vartheta)^{3/2} (\vartheta) \\ \vartheta = \frac{4}{3} + \frac{\gamma(\gamma-2)}{\gamma+1} \left(1 - \frac{1}{2\gamma\sqrt{\gamma^2-1}} \log \left(\frac{\gamma+\sqrt{\gamma^2-1}}{\gamma-\sqrt{\gamma^2-1}} \right) \right) \quad (4)$$

where ϵ_{ph} is the energy of the photon in joules, m is the mass of an electron (9.1094×10^{-31} kg), c is the speed of light (2.9979×10^8 m/s), Z_a is the atomic number of the element with which the photon interacts, $\sigma_0 = (8/3)\pi r_0^2$ ($r_0 =$

2.818×10^{-15} m is the so-called classic electron radius), and $\gamma = 1 + \varepsilon_{ph}/(mc^2)$ is the photoelectron relativistic factor. The absorption edge factor is

$$F = 2\pi \left(\sqrt{\frac{\varepsilon_{ion}}{\varepsilon_{ph}}} \right) \frac{e^{-4\xi \cot^{-1} \xi}}{1 - e^{-2\pi\xi}}; \quad \xi = \sqrt{\frac{\varepsilon_{ion}}{\varepsilon_{ph} - \varepsilon_{ion}}}$$

where

$$\varepsilon_{ion} = \frac{Z_a^5 mc^2}{2 \times 137^2} \quad (5)$$

is the K-shell electron ionization energy.

At photon energies between ~ 30 keV and ~ 20 MeV, Compton scattering is the dominant collision type. Compton scattering is characterized by a change in direction and magnitude of a photon's momentum after colliding with an electron orbiting a molecule. This results in a decrease in energy of the photon and often the ejection of the electron from the molecule. In our model, the post collision energy of the photon is described by the following equation [Lehtinen, 2000]:

$$\gamma' = \frac{\gamma}{1 + sX + (2\gamma - s)X^3} \quad (6)$$

where $s = \gamma/(1 + 0.5625\gamma)$, $\gamma = \varepsilon_{ph}/mc^2$ is the photon's energy normalized to the electron rest energy, and X is a uniformly distributed random variable between [0,1]. Equation (6) is an approximated relationship given by the Klein-Nishima formula [Lehtinen, 2000].

The deflection angle with respect to the axis along the initial momentum is given by the energy-momentum conservation law:

$$\theta = \arccos \left(1 + \frac{1}{\gamma} + \frac{1}{\gamma'} \right) \quad (7)$$

The azimuthal angle of deflection is uniformly distributed using a random angle between 0 and 2π . The cross sections for Compton scattering are taken from the XCOM database of the National Institute of Standards and Technology (<http://www.nist.gov/pml/data/xcom/index.cfm>).

Above photon energies of ~ 20 MeV, the dominant collision type is pair production, which is characterized by the transformation of the photon into an electron-positron pair. This process occurs in extremely large electric fields near the nucleus of an atom. Because it is an energy to mass conversion, for pair production to occur, the photon energy must exceed the threshold energy of $2m_e c^2$

= 1.022 MeV. The resulting electron is neglected in our simulations, however, the positron will annihilate with another electron, producing two new 511 keV photons in opposite directions that must be considered in the simulation. The cross section data for pair production was also taken from the XCOM database of National Institute of Standards and Technology.

3-Dimensional Modeling

Successful modeling of photon transport through the atmosphere requires continuous tracking of the photons position and momentum through a three dimensional space. To achieve this, the following parameters are tracked and used to calculate trajectories between collisions as well as momentum changes:

- θ , the angle between the photon's momentum and the z -axis (orthogonal to the Earth's surface).
- ψ , the azimuth angle of the photon's momentum around the z -axis.
- χ , the angle between the photon's initial momentum and final momentum after a collision.
- φ , the azimuth angle of the photon's momentum after a collision around its previous momentum before the collision.
- The energy of the photon.
- The spatial coordinates of the photon.

The path length between collisions is calculated using the equation [Østgaard, 2008]:

$$s = \left(-\frac{1}{a \cos(\theta)} \right) \ln \left(1 + \frac{\ln(1 - rand) a \cos(\theta)}{\sigma_T \rho_z} \right) \quad (8)$$

where $\rho_z = \rho_0 e^{-az}$ is the atmospheric density at the current altitude ($\rho_0 = 2.688 \times 10^{25}$ molecules/m³ is the ground level atmospheric density; z is the current altitude, $a = 1/7000$ m⁻¹), σ_T is the total cross section, and $rand$ is a uniformly distributed random number between [0,1]. The unit vector, \hat{i} of the photon's momentum is calculated as follows:

$$\begin{aligned} \hat{i} = & [\cos(\chi) \sin(\theta) \cos(\psi) + \sin(\chi) \cos(\phi) \cos(\theta) \cos(\psi) - \sin(\psi) \sin(\chi) \sin(\phi)] \hat{x} + \\ & [\cos(\chi) \sin(\theta) \sin(\psi) + \sin(\chi) \cos(\phi) \cos(\theta) \sin(\psi) + \sin(\chi) \sin(\chi) \cos(\psi)] \hat{y} + \\ & [\cos(\chi) \cos(\theta) - \sin(\chi) \cos(\phi) \sin(\theta)] \hat{z} \end{aligned} \quad (9)$$

Radiation Detection

In order to detect the radiation content of the simulated photons, a series of ring surfaces was placed in the simulation domain with centers along the axis of

the source of photons, orthogonal to the z-axis. The flux of photons through these surfaces was used to generate a two dimensional cross section through the x-z plane depicting radiation fluence with a resolution of 1 km. A visualization of this process is shown in Figure 3.

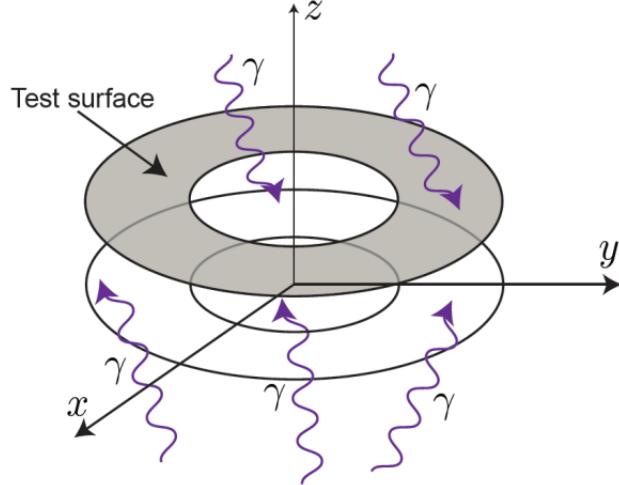


Figure 3: Disk surface oriented orthogonally to the z-axis with center along axis of source photons used to detect radiation flux [Wei, 2011].

RESULTS AND DISCUSSION

Radiation Distributions

In considering the radiation exposure due to thundercloud radiative events, a standard TGF was simulated. The photon energy distribution of a TGF was calculated using theories of TGF production described in [Celestin and Pasko, 2011] for a 100 MeV unbranched lightning leader and is shown in Figure 4. Lower TGF altitudes are estimated to be around 12 km [Wei, 2011], so Figure 5 shows the resulting radiation and effective dose distributions using a TGF source at 12 km.

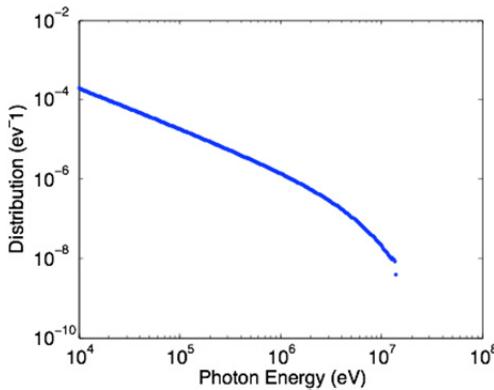


Figure 4: TGF photon energy distribution at the source altitude

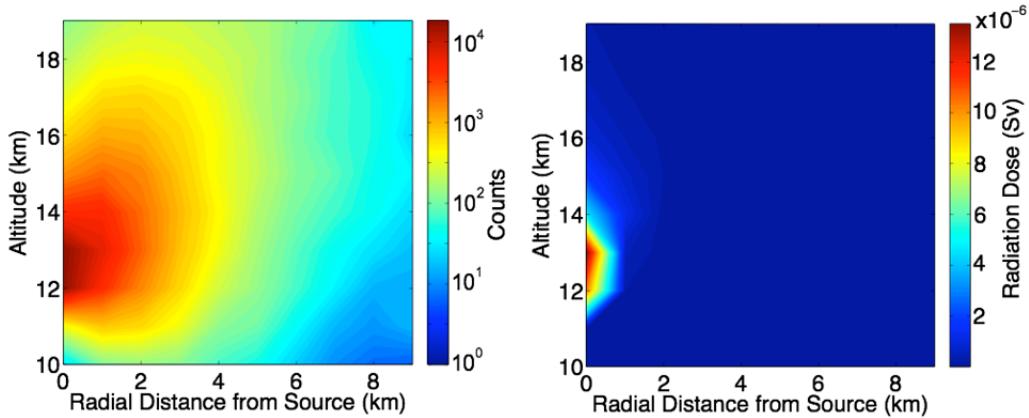


Figure 5: Left: distribution of photon hits on test surfaces. Right: radiation dose (Sv) corresponding to the left panel.

Commercial airline flight altitudes are typically around 10 km, but can reach 12 km at times. It can be seen from Figure 5 that at these altitudes, the worst-case radiation exposure is $\sim 14 \mu\text{Sv}$, or .014 mSv, if the plane is located within 1 km of the TGF and flying at 12 km. These results are well below the potentially harmful doses described earlier in the *Radiation Risks* section. Furthermore, in [Barish, 2004], it is estimated that the exposure of aircraft passengers and crew to cosmic radiation during a transcontinental flight can typically range from 0.060–0.150 mSv [Federal Aviation Administration, 2004]. Thus, not only is the potential TGF radiation exposure to aircraft travelers and crew less than a fraction of a percent of the recommended safe doses for all humans from embryo to adult, it is also a negligible exposure relative to the typical cosmic radiation exposure during a commercial airline flight.

In [Celestin and Pasko, 2011], X-ray bursts and TGFs correspond to the same event, but TGFs are more energetic and consist of more photons. Thus, it is safe to extend the conclusion that if TGFs are of no threat to aircraft crew and passengers, neither are X-ray bursts from a lightning discharge.

CONCLUSIONS

In this paper, we have used a Monte Carlo model to simulate the transport of energetic photons through the atmosphere in order to estimate the radiation distribution given off by a radiative thundercloud event. It was discovered that the maximum radiation dose potentially received by aircraft crew and passengers is $\sim 0.014 \text{ mSv}$. This quantity has been determined to be insignificant to humans of all ages and developmental stages and is trivial in comparison to the average exposure to cosmic radiation during continental and transcontinental flight of $\sim 0.060\text{--}0.150 \text{ mSv}$ respectively.

ACKNOLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grants No. EEC-1062984 and AGS 0734083.

REFERENCES

- Barish, R.J. (2004), In-flight radiation exposure during pregnancy, *The American College of Obstetricians and Gynecologists*, 103(6), 1326-1330.
- Brent, R.L. (1989), The effect of embryonic and fetal exposure to x-ray, microwaves, and ultrasound: sounseling the pregnant and nonpregnant patient about these risks, *Seminars in Oncology*, 16(5), 347-68.
- Carron, N. J. (2007), *An Introduction to the Passage of Energetic Particles through Matter*, Taylor & Francis Group, Boca Raton, FL.
- Dwyer, J. R. (2008), Source mechanisms of terrestrial gamma-ray flashes, *J. Geophys. Res.*, 113, D10103, doi:10.1029/2007JD009248.
- Dwyer, J. R., D. M. Smith, M. A. Uman, Z. Saleh, B. Grefenstette, B. Hazelton, and H. K. Rassoul (2010), Estimation of the fluence of high-energy electron bursts produced by thunderclouds and the resulting radiation doses received in aircraft, *J. Geophys. Res.*, 115, D09206, doi:10.1029/2009JD012039.
- Dwyer, J. R., et al. (2004), Measurements of X-ray emission from rocket-triggered lightning, *Geophys. Res. Lett.*, 31, L05118, doi:10.1029 / 2004GL018770.
- Dwyer, J.R., et al. (2005), X-ray bursts associeated with leader steps in cloud-to-ground lightning, *Geophys. Res. Lett.*, 32, L01803, doi: 10.1029 / 2004GLO21782.
- Eack, K. B., W. H. Beasley, W. D. Rust, T. C. Marshall, and M. Stolzenburg (1996), Initial results from simultaneous observations of x-rays and electric fields in a thunderstorm, *J. Geophys. Res.*, 101, 29637-29640.
- U.S. Environmental Protection Agency. Radiation Protection Web Site. Health Effects. Available at: http://www.epa.gov/radiation/understand/health_effects.html. Retrieved July 21, 2011.
- Federal Aviation Administration. Radiobiological Team Web Site. What commercial aircraft crewmembers should know about their occupational exposure to ionizing radiation. Available at: <http://www.cami.jccbi.gov/AAM-600/Radiation/trainingquestions.htm>. Retrieved March 25, 2004.
- Fishman, G. J., et al. (1994), Discovery of intense gamma-ray flashes of atmospheric origin, *Science*, 264(5163), 1313.
- (2004), Guidelines for diagnostic imaging during pregnancy, *Obstetrics and Gynecology*, 104(3), 647-651.
- Gurevich, A. V. (1961), On the theory of runaway electrons, *Sov. Phys. JETP, Engl. Transl.*, 12, 904.
- Gurevich, A. V., G. M. Milikh, and R. Roussel-Dupré (1992), Runaway electron mechanism of air breakdown and preconditioning during a Thunderstorm, *Phys. Lett. A*, 161(1-2), 119-125.
- Lehtinen, N. G. (2000), Relativistic runaway electrons above thunderstorms, PhD dissertation, Stanford University.
- Martin, J.E. (2000), *Physics for Radiation Protection*, John Wiley & Sons, Inc., New York, New York.
- Moss, G. D., V. P. Pasko, N. Liu, and G. Veronis (2006), 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.
- Østgaard, N., T. Gjesteland, J. Stadsnes, P. H. Connell, and B. Carlson (2008), Production altitude and time delays of the terrestrial gamma flashes: Revisiting the Burst and Transient Source Experiment spectra, *J. Geophys. Res.*, 113, A02307, doi:10.1029/2007JA012618.

Pelliccioni, M. (2000), Overview of fluence-to-effective dose and fluence to ambient dose equivalent conversion coefficients for high energy radiation calculated using FLUKA code, *Radiat. Protect. Dosimetry*, 88, 279-297.

Wei, X., S. Celestin, and V.P. Pasko, 2011, Monte Carlo Simulation of Terrestrial Gamma-ray Flashes, CEDAR Workshop, 26 June to 01 July, Santa Fe, NM.

Wilson, C.T.R., The acceleration of beta-particles in strong electric fields such as those of thunderclouds, *Proc. Cambridge Phil. Soc, London*, 22, 534-538, 1925.

ACCURATE MODELING OF ION CONDUCTIVITY IN THE EARTH'S ATMOSPHERE

Sean S. McDonough,* Jianqi Qin,⁺ and Victor P. Pasko[#]

Communications and Space Science Laboratory
Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate Student of
Department of Electrical & Computer Engineering
University of Rochester
Rochester, NY 14627

ABSTRACT

This paper focuses on the numerical modeling of the ion conductivity profile in the Earth's atmosphere up to 90 km by expanding upon the model established by *Tinsley and Zhou* [2006]. The related knowledge is important, in particular, for understanding lightning-driven transient luminous events called sprites and halos. It has been demonstrated that ambient conductivity at sprite initiation altitudes (~65–85 km) is critical for the establishment of a streamer initiation region in the sprite-halo events induced by positive cloud-to-ground lightning discharges [*Qin et al.*, 2011]. In this work we improve the model of *Tinsley and Zhou* [2006] by including new sources of high altitude ion pair production, more accurately modeling the ion-ion recombination coefficient, along with including new aerosol concentrations that expand the ion-aerosol attachment rate to the mesosphere. The model accounts for geographical, latitudinal, and seasonal variations in the Earth's atmosphere. The results of this model are consistent with previous theoretical and experimental [*Hale*, 1984] evaluations of the Earth's atmospheric ion conductivity.

INTRODUCTION

The global atmospheric electric circuit (see Figure 1), which was initially proposed by *Wilson* [1920] in order to describe fair weather electricity, has been extensively studied since the early 20th century. In Wilson's model, the Earth's surface and ionosphere can be thought of as the highly conductive plates of a

[#] Faculty Mentor

⁺ Graduate Mentor

capacitor separated by the poorly conducting atmosphere. Thunderstorms produce upward current, uniformly charging the system, and the highly resistive atmosphere provides the return current path to ground for the displaced current [Tinsley and Zhou, 2006]. This global circuit is important since it is fundamentally tied to other atmospheric phenomena. It has been found that this circuit is a sensitive indicator of global moisture and land temperature levels [Bering et al., 1998], and possibly is a source that causes climate change [Tinsley, 2000, 2004; Tinsley and Yu, 2004]. One active research area in this field has been devoted to accurately modeling of the conductivity profile in the atmosphere from the ground to the D-region of the ionosphere.

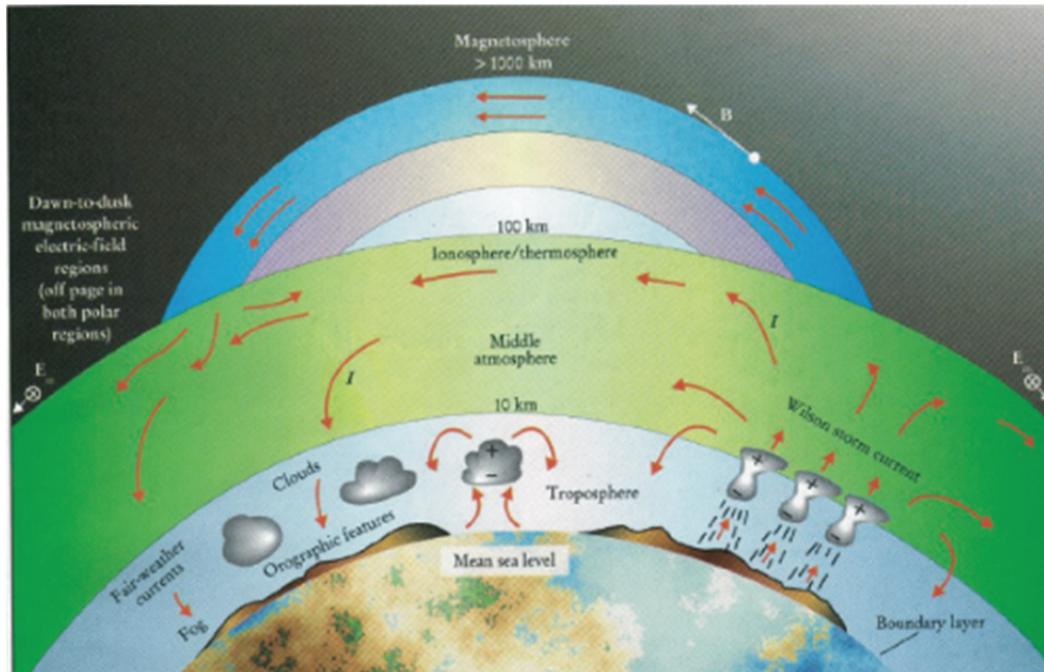


Figure 1: Flow of the electric current in the global circuit. Unlabeled arrows represent current flow.

Sprites are transient luminous events (TLEs) that are usually produced by positive cloud-to-ground lightning discharges. These optical phenomena exhibit a diffuse glow in the lower ionosphere, referred to as a sprite halo, and at lower altitudes develop into fine structure sprite streamers [Qin et al., 2011]. Wilson [1925] theoretically predicted the possibility of large-scale gas discharge events above thunderstorms, but it was not until 1989 that these events were documented by researchers at the University of Minnesota serendipitously during a test of a low-light television camera [Franz et al., 1990]. Recent studies highlight the importance of the conductivity variations in the upper atmosphere, since the atmospheric conductivity at sprite initiation altitudes ($\sim 65\text{--}85$ km) is critical for the establishment of a streamer initiation region in the sprite-halo events induced

by positive cloud-to-ground lightning discharges [Qin *et al.*, 2011]. Therefore, accurate modeling of the ion conductivity variations up to the sprite initiation altitudes is an important topic to be investigated. Although the global electric circuit model established by Tinsley and Zhou [2006] can calculate ion conductivity up to 60 km, it cannot be applied to the study of sprite-halo events due to its limited scope.

The purpose of this work is to improve the current treatment of the ion conductivity profile by expanding the upper boundary of the model from 60 km in altitude to 90 km. The improved model is expected to provide useful information to better understand the initiation mechanisms of sprites. We note that the previous work in [Tinsley and Zhou, 2006] has only modeled the ion-ion recombination rate up to 40 km and the ion-aerosol attachment up to 60 km for several species of aerosols. In order to improve this model, the research will consist of accurately modeling the ionization rate in the upper atmosphere due to Lyman-alpha radiation and other sources, and modeling the ion loss due to ion-ion recombination up to 90 km and modeling aerosol concentrations and ion-aerosol attachment up to 90 km.

MODEL DESCRIPTION

The conductivity profile of the atmosphere is modeled in high resolution for one dimension using equidistant grid points. Expanding upon the work of Tinsley and Zhou [2006], the conductivity profile is calculated by first modeling the ion pair production rate (q), which at altitudes below 60 km is dominated by ionization due to galactic cosmic ray (GCR) flux and at higher altitudes is dominated by Lyman-alpha radiation ionizing nitric oxide [Reid, 1986]. The ion-ion recombination coefficient (α), which is dominated by ternary recombination below 30 km and binary recombination above 30 km, is then calculated along with the ion-aerosol attachment rate (β), which varies widely between geographical locations and latitudes. The ion concentration (n) is obtained by solving the equation

$$dn / dt = q - \alpha n^2 - \beta n \quad (1)$$

for the steady state condition where $dn/dt = 0$, with t being time. The conductivity (σ) can then be calculated by

$$\sigma = ne(\mu_+ + \mu_-) \quad (2)$$

where e is the elementary charge and μ is the mobility of the positive and negative ions [Tinsley and Zhou, 2006].

Ion-Pair Production

The ion-pair production at low altitudes (under 5 km) is dominated by ionization due to the natural radioactivity released by the ground. The principle sources of radiation are direct alpha, beta and gamma radiation from the earth

along with the radioactive gasses ^{222}Rn , ^{220}Rn and their subsequent daughter products. We utilize the near ground ion pair production rate as a function of altitude as given by *Tinsley and Zhou* [2006].

The main source of ion-pair production in the upper troposphere, stratosphere and lower mesosphere is the GCR flux. While ionization due to bremsstrahlung X-rays is present at stratospheric altitudes [*Vampola and Gorney*, 1983], the ion pair production is several orders of magnitude smaller than that due to GCR flux, and its effect is negligible. Ion pair production at these attitudes is related to geomagnetic latitude and proportional to air density [*Neher*, 1967]. In this model we utilize the ionization rate at standard temperature and pressure given in *Tinsley and Zhou* [2006] and we later scale this data for the variation of air density using data from the U.S. Standard Atmosphere [1976].

Above 65 km the ionization of nitric oxide by Lyman-alpha radiation becomes the dominant source of ionization. Although the intensity of Lyman-alpha radiation is well known, the ion pair production due to this phenomena cannot be accurately measured due to the variability in nitric oxide concentration in the upper atmosphere [*Ratnasiri*, 1977]. We therefore model the ionization due to Lyman-alpha radiation using the daytime mid latitude ion-pair production rate of *Reid* [1986, Figure 13.1] with no respect to latitudinal variation. The change in ionization rates for daytime and nighttime condition occurs rapidly at twilight therefore we only model the two states. For nighttime ion pair production due to scattered Lyman-alpha radiation [*Reid*, 1986] we simply raise the height of the Lyman-alpha production layer by 8 km, roughly consistent with [*Vampola and Gorney*, 1983].

Two additional sources of ionization in the upper atmosphere are the ionization of singlet oxygen ($\text{O}_2 \ ^1\Delta_g$) by ultra violet radiation and ionization due to solar X-rays. The $\text{O}_2 \ ^1\Delta_g$ is a product of O_3 photo-disassociation and is present in the daytime; it is thought to have higher concentrations at lower latitudes [*Ratnasiri*, 1977]. We model the concentration of ions produced by $\text{O}_2 \ ^1\Delta_g$ ionization and X-rays using the data of *Reid* [1986, Figure 13.1] for mid-latitude conditions. We model these sources of ionization for daytime conditions since these forms of solar radiation are not present at night.

Ion–Ion Recombination

The ion–ion recombination coefficient (α) has little temporal or spatial variation, except in altitude [*Tinsley and Zhou*, 2006]. The altitude distribution of α is not well known due to the uncertainty of the ion species involved as well as the temperature uncertainty [*Smith and Church*, 1976]. The value of the ion–ion recombination coefficient has been calculated experimentally in [*Rosen and Hoffman*, 1981], theoretically by *Bates* [1982] and from the laboratory measurements in [*Smith and Church*, 1976; *Smith and Adams*, 1982].

In the present model, the ion–ion recombination coefficient below 20 km is calculated by interpolating the data provided in [*Bates*, 1982], in which the

treatment of ternary recombination, which is the prevailing reattachment phenomena at low altitudes, is highly accurate. For the altitude range from 20 km to 60 km, we use a modified equation similar to that in [Smith and Adams, 1982] to obtain the ion-ion recombination coefficient

$$\alpha_T = 1.63 \times 10^{-5} \exp((z - 5) / 7.38) + 6.38 \times 10^{-8} \text{ cm}^{-3} \text{ s}^{-1} \quad (3)$$

and for altitudes between 60 km and 90 km we use the constant values of α for given altitude profiles given in [Smith and Church, 1976].

Ion-Aerosol Attachment

The ion-aerosol attachment term (β) can be expressed as $\sum_i \sum_r \beta_r (dN_i / dr) \Delta r$, which as described by Tinsley and Zhou [2006] is highly variable in time and space, especially over land and downwind of continents over the oceans. This is due to the variability of the concentrations, size distributions, and species of the aerosol particle concentrations, N_i . The coefficient of attachment (β_r), representing attachment of aerosols of radius r to ions of species i , varies with the radii of the particles. Thus $\sum_i \sum_r \beta_r (dN_i / dr) \Delta r$ is a summation over size (r) as well as species (i).

The coefficient of attachment (β_r), which describes the rate of aerosol-ion neutralization, is taken from Tinsley and Zhou [2006]. The expression for β_r , as a function of aerosol particle radius (r), is given as:

$$\beta_r(r) = 4.36 \times 10^{-5} r - 9.2 \times 10^{-8} \text{ cm}^3 \text{ s}^{-1} \quad (4)$$

for a radius r greater than 0.01 μm where (r) is in micrometers. For aerosol radii between 0.004 μm and 0.01 μm we use the expression

$$\log_{10} \beta_r(r) = 1.243 \log_{10}(r) - 3.978 \quad (5)$$

where r is in micrometers.

This work includes the detailed treatment of both tropospheric and stratospheric aerosols as found in Tinsley and Zhou [2006], with modifications. We base this model on the software package Optical Properties of Aerosols and Clouds (OPAC) and the accompanying paper by Hess *et al.* [1998], in contrast to the latitude-longitude bin method used by Tinsley and Zhou [2006] for tropospheric aerosols. We use the size and species distribution given by Hess *et al.* [1998] for tropospheric altitudes. The size distribution of tropospheric aerosols is applied to each species i :

$$\frac{dN_i}{dr} = \frac{N_i}{r\sqrt{2\pi} \log 10 \log_{10} \sigma_i} \exp \left[-\frac{1}{2} \left(\frac{\log_{10} r - \log_{10} r_{\text{mod}N,i}}{\log_{10} \sigma_i} \right)^2 \right] \quad (6)$$

where N_i is the total particle number density of an aerosol species, σ_i is the width of the distribution and $r_{\text{mod}N,i}$ is the mode radius.

The thirteen aerosol types given in the Optical Properties of Aerosols and Clouds (OPAC) include urban, continental, desert, oceanic, and polar conditions. As described in *Tinsley and Zhou* [2006] these conditions are described in terms of mixtures in different proportion of ten primary aerosol components, such as soot, minerals, and sulfate droplets, each with their own size distributions. These aerosol types along with a background aerosol layer for nonvolcanic periods can be used to simulate 13 types of geographical conditions as modeled in OPAC and by *Tinsley and Zhou* [2006]. The vertical distribution of aerosols in the troposphere is specified as a series of layers, starting at ground level as given in *Hess et al.* [1998], with different boundaries and scale heights depending on the geographical conditions. The OPAC tropospheric data provide more detailed and representative geographic, altitude, and attachment rate variations than those used previously [*Tinsley and Zhou*, 2006].

Aerosols above the troposphere are much less diverse in composition, with some layers being composed of only one aerosol species. For stratospheric aerosol layers we use the bimodal Gaussian altitude distribution given in *Tinsley and Zhou* [2006] along with the accompanying size distributions. For the layer of high latitude aerosols at 40 km we use the altitude and size distribution as given in *Tinsley and Zhou* [2006].

Aerosol concentrations above the stratopause are not well known and are subject to much controversy. This model incorporates two new sources of aerosols for mesospheric altitudes, namely a high altitude summer polar sulfate aerosol layer and a substantial layer of meteoric smoke extending down to the stratosphere. To model the polar sulfate aerosol layer we use the altitude distribution given in *Mills et al.* [2005, Figure 6] for volcanic conditions and the size distribution at 83 km for 1992 conditions [Figure 7]. This layer is only present in the local summer and is similar to spatial distribution of the 40 km layer, and we arbitrarily set the aerosol concentration to zero at $\pm 40^\circ$ latitude and twice the concentration at the poles. We neglect to model volcanic variations of this layer, as there is only a small effect on conductivity. For the layer of meteoric smoke we utilize the altitude and size distributions given by *Megner et al.* [2006, Figure 2]. This layer of meteoric smoke is highly variable with respect to meteoric input, coagulation, and vertical winds, therefore we do not attempt to model temporal or spatial variations in this layer and instead treat it as uniform.

Ion Mobility

The ion mobility (μ) can be expressed as a function of temperature (T) and pressure (p) given by the equation

$$\mu = \mu_0 \left(\frac{T}{273.15} \right) \left(\frac{1013}{p} \right) \quad (7)$$

where μ_0 is the reduced ion mobility, T is in kelvin and p is in millibars. Although the experimental values of μ for positive and negative ions vary by tens of

percents [Gringel *et al.*, 1986], in this model we simply use a value of $3.3 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ for the sum the positive and negative reduced ion motilities and assume that it is constant with altitude. Moreover, we use the U.S. Standard Atmospheric Model to calculate the temperature and pressure as a function of altitude.

RESULTS

In order to present the full variation of the model we have chosen to vary a wide range of parameters depending on the type of profile. We vary the ionization rate and conductivity profiles by temporal parameters and we vary the ion concentration and attachment rate profiles by latitudinal parameters.

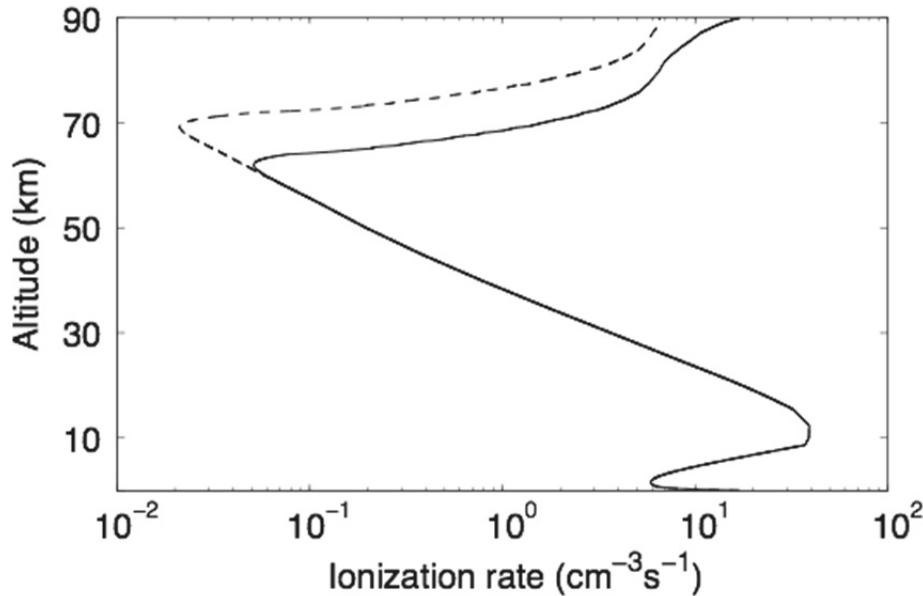


Figure 2: Vertical profile of the ionization rate at 45° latitude for December over land. The solid line represents daytime conditions while the dashed line represents nighttime conditions.

The temporal variation in the ion-pair production rate for mid latitudes can be seen in Figure 2 where daytime and nighttime conditions are represented by solid and dashed lines, respectively. There is little difference in the ion pair production rates for December and July. The knee where ionization due to Lyman-alpha radiation overtakes ionization due to GCR flux is approximately located at 60 km during the daytime and 70 km at night. A refinement to the current model would be the inclusion of spatial variations of Lyman-alpha radiation along with spatial and temporal variations of NO concentrations.

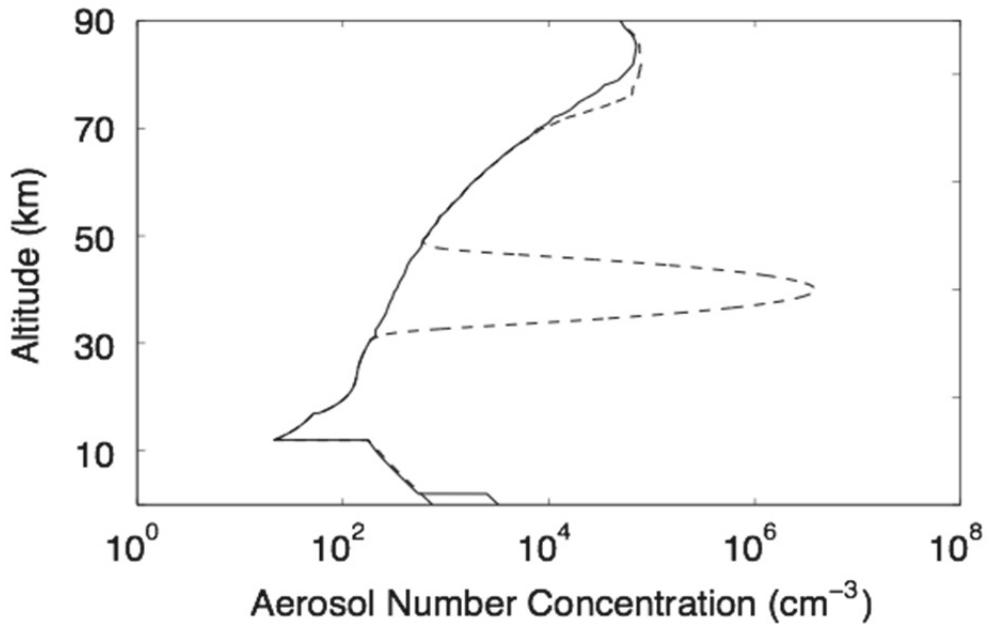


Figure 3: Vertical profiles of the aerosol particle concentration for equatorial (solid line) and polar (dashed line) conditions in July.

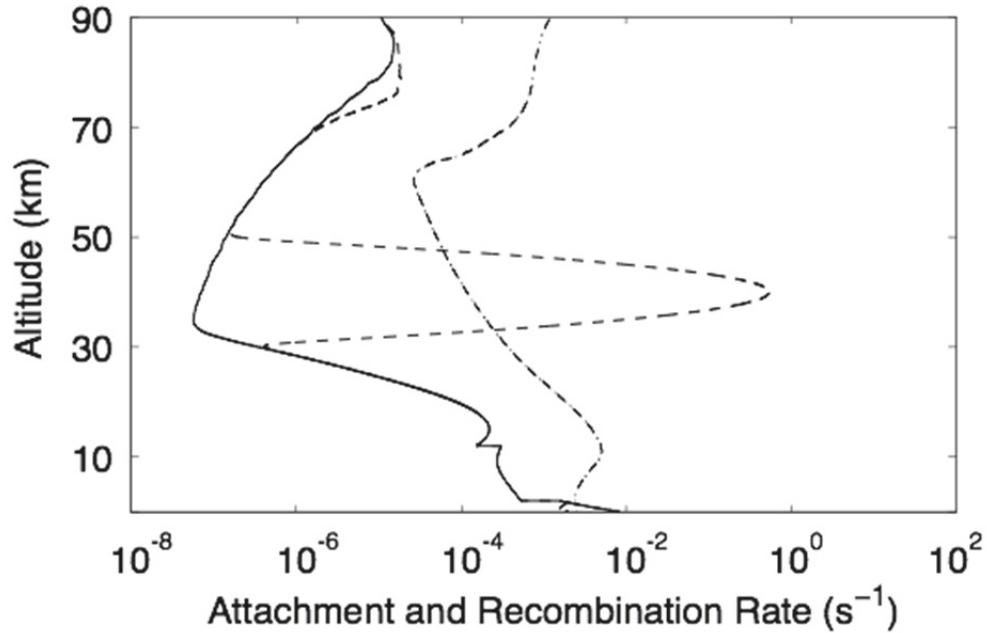


Figure 4: Vertical profiles of the ion loss mechanisms, the ion aerosol attachment rate coefficient β for equatorial conditions (solid line), polar condition (dashed line) and the rate of ion-ion attachment αn (dot-dash line) for equatorial conditions. The data is for the daytime of July over ocean.

Figure 3 shows the aerosol concentration profile for summertime equatorial and polar conditions. There is little variation in the corresponding curve for winter conditions for most altitudes. It is important to note the increase in aerosol concentration at high altitudes in the region between 70 and 85 km for polar conditions, this is the layer of ultrafine sulfate particles. It should also be noted that the aerosol components above 35 km have a significantly lower mass concentration than that of the tropospheric and stratospheric aerosols due to the significantly smaller particle radius.

The ion-aerosol attachment rate for summertime equatorial and polar conditions is represented in Figure 4 as solid and dashed lines, respectively. The profile of the ion-aerosol attachment rate for high altitudes is significantly smaller than the aerosol concentration profile for high altitudes as seen in Figure 3. This discrepancy is due to the small radius of the particles, and thus a smaller effective attachment rate. Also plotted in Figure 4 is the ion-ion recombination rate αn as a dash-dot line, it is evident therefore that the ion-ion loss process is dominant at all but low altitudes and lower mesospheric altitudes for polar conditions.

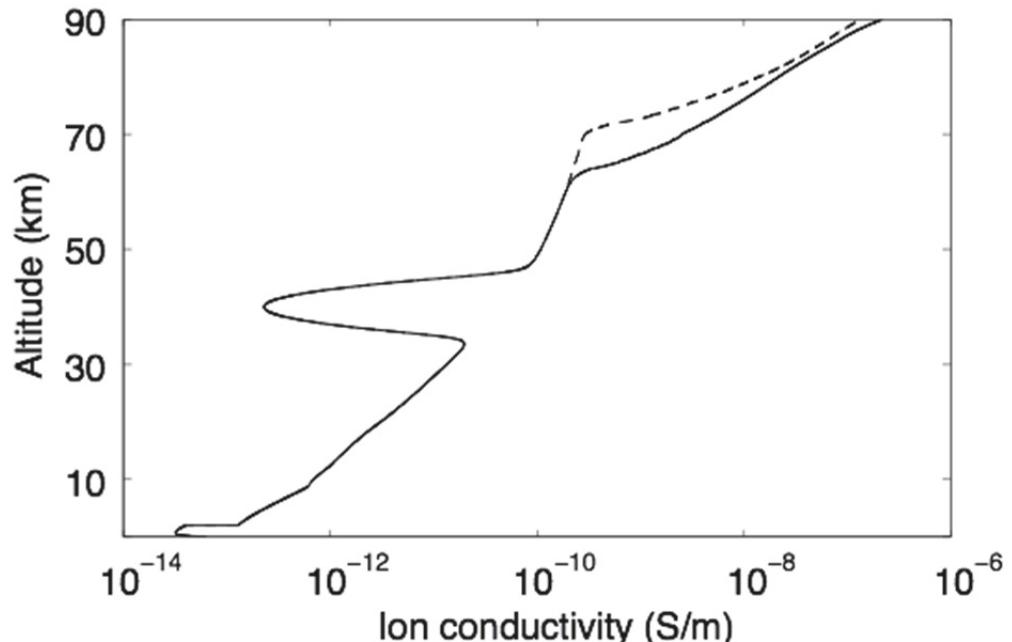


Figure 5: Vertical profile of ion conductivity for a July day (solid line) and night (dashed line) for a latitude of 45° over a continental location.

The conductivity profile modeling for summer mid latitude continental conditions is presented in Figure 5. It can be seen that there is a knee present at approximately 60 km during daytime conditions and at 70 km during nighttime conditions. This sharp knee is due to the ionization rate of Lyman-alpha radiation overtaking that due to GCR flux during the day. The only source of high altitude ionization at night is that due to scattered Lyman-alpha radiation and this

reduction in ion pair production leads to a knee of 70 km. Although the ion pair production rate is smaller at upper mesospheric altitudes than at ground the high ion mobility of the upper mesosphere leads to higher conductivity at those heights.

DISCUSSION

The conductivity profile presented in Figure 5 agrees well with the experimental results shown in *Hale* [1984] and the theoretical model given by *Reid* [1984]. For daytime conditions the conductivity profile above 60 km and below 30 km agrees to within a factor of 3 with *Hale* [1984, Figure 4, Curve 3] and *Reid* [1984, Figure 13.9]. The disagreement between the two model and the experimental results for mid-altitudes is due mainly to the dense volcanic aerosol layer present in [*Tinsley and Zhou*, 2006]. For nighttime conditions the model agrees to within a factor of 2 of *Hale* [Figure 4, Curve 1] above 70 km. However, for 40 km to 70 km in altitude the conductivity profile deviates by approximately an order of magnitude. The deviation of the experimental data may be due to a dense layer of aerosols present at those altitudes [*Hale*, 1984] for nighttime conditions.

Any deviation from the experimental conductivity profile at altitudes above 60 km is most likely due to the high variability of NO concentrations and therefore ion-pair production due to ionization by Lyman-alpha radiation. At high altitudes the ion pair production rate has the largest effect on the total ion conductivity profile above 60 km. It is also important to note the high variability in experimentally measured values of conductivity, since there is sufficient difficulty in obtaining and interpreting measurements of bulk parameters in the Earth's atmosphere [*Reid*, 1986].

CONCLUSIONS

In this work we have developed a model that is capable of simulating the ion conductivity in the Earth's atmosphere from the ground and up to 90 km altitude. The model has expanded upon the model in [*Tinsley and Zhou*, 2006] up to a height of 90 km and includes a more accurate ion-ion recombination profile along with additional sources of ionization due to Lyman-alpha radiation, UV radiation, and solar X rays in the upper mesosphere. Moreover, aerosol production has been expanded into the upper mesosphere with the inclusion of meteoric smoke and high altitude summer polar aerosols. The model compares well with other experimental and theoretical measurements of the ion conductivity, notably the theoretical model of *Reid* [1986] and the experimental results of *Hale* [1984].

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984.

REFERENCES

- Bates, D. R. (1982), Recombination of small ions in the troposphere and lower stratosphere, *Planet. Space Sci.*, 30(2), 1275–1282.
- Bering, E. A., III, A. A. Few, and J. R. Benbrook (1998), The global electric circuit, *Phys. Today*, 51, 24–30.
- Franz, R. C., R. J. Nemzek, and J. R. Winckler (1990), Television image of a large upward electrical discharge above a thunderstorm system, *Science*, 249, 48–51.
- Gringel, W., J. M. Rosen, and D. J. Hoffman (1986), Electrical structure from 0 to 30 kilometers, in *The Earth's Electrical Environment*, pp. 166–182, Natl. Acad. Press, Washington, D. C.
- Hale, L.C. (1984), Middle atmospheric electrical structure, dynamics and coupling. *Adv. Space Res.*, 4, pp. 175–186.
- Hess, M., P. Koepke, and I. Schult (1998), Optical properties of aerosols and clouds: The software package OPAC, *Bull. Am. Meteorol. Soc.*, 79(5), 831–844.
- Megner, L., M. Rapp, and J. Gumbel (2006), Distribution of meteoric smoke - sensitivity to microphysical properties and atmospheric conditions, *Atmos. Chem. Phys.*, 6, 4415–4426.
- Mills, M. J., O. B. Toon, and G. E. Thomas (2005), Mesospheric sulfate aerosol layer, *J. Geophys. Res.*, 110, D24208, doi:10.1029/2005JD006242.
- Neher, H. V. (1967), Cosmic ray particles that changed from 1954 to 1958 to 1965, *J. Geophys. Res.*, 72, 1527–1539.
- NOAA (1976), U.S. Standard Atmosphere 1976, U.S. Govt. Print. Off., Washington, D. C.
- Qin, J., S. Celestin, and V. P. Pasko (2011), On the inception of streamers from sprite halo events produced by lightning discharges with positive and negative polarity, *J. Geophys. Res.*, 116, A06305, doi:10.1029/2010JA016366.
- Ratnasiri, P.A.J. (1977), D-region processes at equatorial latitudes, *J. Atmos. Terr. Phys.*, Volume 39, Issues 9-10, Pages 999-1009
- Reid, G.C. (1986), Electrical structure of the middle-atmosphere, in Study in Geophysics: the earth's electrical environment. In: Krider, E.P., Roble, R.G. (Eds.), National Academy Press, Washington, DC, pp. 183–194.
- Rosen, J. M., and D. J. Hoffman (1981), Balloon borne measurements of electrical conductivity, mobility, and the recombination coefficient, *J. Geophys. Res.*, 86(C8), 7406–7410.
- Smith, D., and M. J. Church (1976), Ion-ion recombination rates in the earth's atmosphere, *Planet. Space Sci.*, 25, 433–439.
- Smith, D., and N. G. Adams (1982), Ionic recombination in the stratosphere, *Geophys. Res. Lett.*, 99, 1085–1087.
- Tinsley, B. A. (2000), Influence of the solar wind on the global electric circuit, and inferred effects on cloud microphysics, temperature, and dynamics of the troposphere, *Space Sci. Rev.*, 94, 231–258.
- Tinsley, B. A. (2004), Scavenging of condensation nuclei in clouds: Dependence of sign of electroscavenging effect on droplet and CCN sizes, paper presented at International Conference on Clouds and Precipitation, World Meteorol. Organ., Bologna, Italy, 18–23 July.
- Tinsley, B. A., and F. Yu (2004), Atmospheric ionization and clouds as links between solar activity and climate, in *Solar Variability and its Effects on Climate*, *Geophys. Monogr. Ser.*, vol. 141, edited by J. Pap and P. Fox, pp. 321–339, AGU, Washington, D. C.
- Tinsley, B. A. and Zhou, L (2006) Initial results of a global circuit model with stratospheric and tropospheric aerosols, *J. Geophys. Res.*, 111, D16205.
- Vampola, A. L., and D.J. Gorney (1983). Electron energy deposition in the middle atmosphere, *J. Geophys. Res.* 88, 6267.
- Wilson, C. T. R. (1920), Investigations on lightning discharges and on the electric field and thunderstorms, *Phil. Trans. R. Soc. Lond. A* 221, 73–115.
- Wilson, C. T. R. (1925), The acceleration of Beta-particles in strong electric fields such as those of thunderclouds, *Proc. Camb. Philos. Soc.*, 22, 534.

STUDY OF DOPPLER VELOCITY ESTIMATION TECHNIQUES ON METEOR-HEAD RADAR REFLECTIONS

Christopher Hong,* Julio Urbina,# and Freddy Galindo⁺

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
The Cooper Union
New York, NY 10003

ABSTRACT

This paper examines current techniques used to determine the Doppler velocity of meteor head echoes. Our study is important because different derived meteor parameters, such as mass and orbit, depend on the Doppler velocity for their computation. Therefore, accurate Doppler velocity estimation is critical to infer accurate derived parameters. In our study, we analyzed and compared four techniques for computing the Doppler velocity from meteor head echoes: 1) spectral analysis, 2) pulse-to-pulse computation 3) decoding processes, and 4) linear fitting of the received echoes.

First, we worked with simulated data in order to evaluate and improve the different techniques mentioned above. We concluded that pulse-to-pulse computation provides the best results, but it requires the usage of an a priori range-time velocity estimation to remove aliasing effects. We also analyzed the impact of noise on the other techniques, which led to inaccurate results with a large standard deviation.

Second, we worked with experimental data. This data was acquired on April 26th, 2005 from the Jicamarca Radio Observatory. The setup at Jicamarca uses two polarizations of the main radar to observe a meteor-head echo simultaneously with two types of long pulses: non-coded and coded. The experimental data has different parameters from our modeled data and our results show that pulse-to-pulse computation does not provide accurate Doppler velocity when applying the a priori range-time velocity estimation. For non-coded long pulses, linear fitting of the received echoes can replace range-time velocity while for coded long

Faculty Mentor

⁺ Graduate Mentor

pulses, decoding techniques are needed. We showed that using proper radar parameters, such as low inter pulse period and high sampling rate, these problems can be minimized.

I. INTRODUCTION

Every day, billions of meteoroids enter and impact the Earth's atmosphere. These meteoroids can affect space weather, damage satellites [1] and affect space travel [2]. As the meteoroids enter the atmosphere, they ablate and form plasma in the altitude range of approximately 70 to 140 km. A meteor head echo is the radar return from this plasma. In order to observe this phenomenon, a high-power large-aperture (HPLA) radar is necessary. From the received data, several important meteor parameters, like position, velocity, and direction [3–7], can be determined. For example, it was shown that the micrometeoroid population is dispersed around the Apex's direction [3] and a new technique to determine the meteoroid mass was reported [4]. Although we do not emphasize the importance of computing basic meteor parameters, like radial velocity and signal to noise ratio, an accurate estimation of these basic parameters allows us to infer precise information regarding the meteoroid. For example, the accurate computation of the meteor's velocity is important for computing the meteor's mass and determining information about the parent meteoroid.

The Jicamarca Radio Observatory (JRO), located in Lima, Peru (Latitude 11.95° South, Longitude 76.87° West), is one of the main facilities in the world for Equatorial ionospheric studies. Jicamarca is equipped with a HPLA 49.92-MHz incoherent scatter radar that is composed of 18,432 half-wave dipoles, divided up into 64 modules, which can be manually adjusted. The radar is able to operate with a beam width of 1° and it has a frequency bandwidth of approximately 1 MHz allowing it to accurately obtain data from the ionosphere [8, 9].

Jicamarca is able to probe the lower, middle, and upper atmosphere. Some of the data that Jicamarca can obtain are state parameters of the ionosphere, accurate data for computing drift velocity and electric field, data regarding the plasma drifts of meteors, and Equatorial Electrojet (EEJ) echoes. EEJ echoes are radar returns from the current carried by free electrons in the ionosphere [10]. Figure 1 shows a typical range time intensity (RTI) graph of data collected around 100 km of altitude from Jicamarca. The *x*-axis represents time in hours and the *y*-axis represents altitude in kilometers. In the altitude range of 85 to 115 km, EEJ echoes and meteor reflection echoes are present. Figure 2 shows an RTI graph in the range of 85 to 115 km that shows two different meteor echoes observed at Jicamarca. In addition, this figure shows the EEJ.

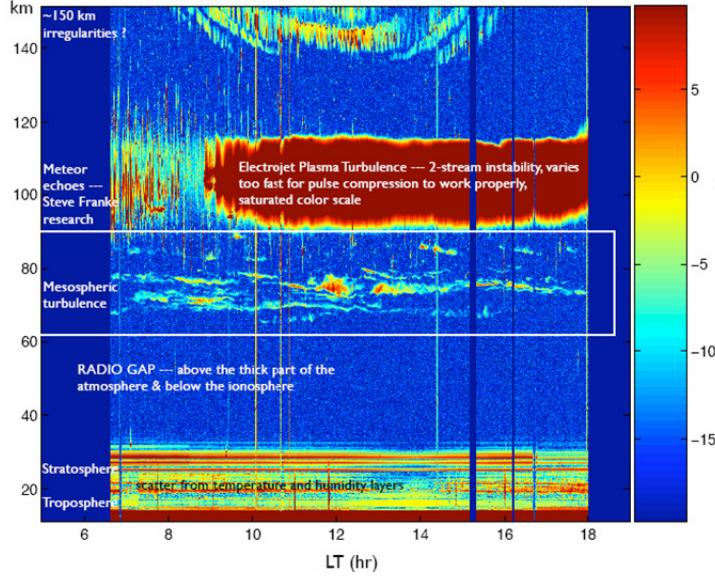


Figure 1: Meteor reflection echoes and the electrojet plasma co-exist in the altitude range of 85 to 115 km. The electrojet echoes interfere with the meteor echoes resulting in contaminated meteor-head radar returns [Courtesy of Professor E. Kudeki].

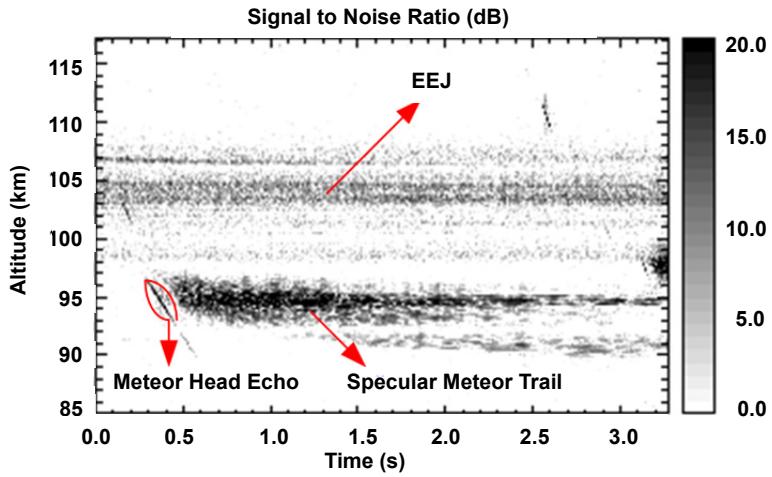


Figure 2: A range time intensity graph in the range 85 to 115 km showing meteor reflection echoes and the Equatorial electrojet (EEJ) echoes. The graph shows meteor head echo and a specular meteor trail.

The three types of meteor reflection echoes are: head echoes [11, 12], specular echoes [13], and non-specular echoes [14]. A meteor-head echo is an instantaneous radar return from the plasma that immediately surrounds the disintegrating meteoroid. This plasma is moving at approximately the same velocity as the meteor [15]. Specular echoes are detected when the meteor is

moving in a direction perpendicular to the radar wave vector. Lastly, non-specular echoes are detected when the radar wave vector is oriented perpendicular to Earth's magnetic field.

To obtain meteor head echo data, the radar must send out a pulse modulated carrier signal of 49.92 MHz and then detect the carrier signal when it returns. The waiting time for the pulse to travel from and return to the radar is known as the inter pulse period (IPP). The experimental data that this project uses is composed of two types of carrier pulses: non-coded long and coded long. Non-coded long pulses are carrier pulses that allow for a good amount of sampling, but have a poor range resolution. Due to this poor range resolution, two or more meteors that are nearby may be detected as one "merged" meteor. To rectify this problem, coded long pulses are long pulses that are compressed through coding. They have good range resolution and allow for more data sampling.

This paper examines current techniques used to determine the Doppler velocity on non-coded and coded long pulses. We worked with and compared four techniques for computing the Doppler velocity: 1) spectral analysis, 2) pulse-to-pulse computation 3) decoding processes, 4) linear fitting of the received echoes, and 4) pulse-to-pulse computation.

This work is organized as follows: In Section II, each Doppler velocity estimation technique is discussed. Then in Section III, each technique is compared first with modeled data which represent non-coded and Barker-13 coded long pulses. In Section IV, the techniques are then compared with experimental data from April 26th, 2005, obtained with the HPLA radar at Jicamarca Radio Observatory. Lastly in Section V, the best technique for determining the radial Doppler velocity of an incoming meteor is determined from the results and a comparison between non-coded and coded long pulses is made. Table 1 shows the parameters of the modeled data and the experimental data.

Table 1: Parameters of Modeled and Experimental Meteor Head Echoes

	Modeled Data	Experimental Data from Jicamarca
inter pulse period (IPP)	60 km	400 km
sampling time	0.15 km or 1 μ s	0.6 km or 4 μ s
pulse length	13-bits, 13 μ s	28-bits, 112 μ s
carrier frequency sizes	13 bits	28 bits

II. DOPPLER VELOCITY ESTIMATION TECHNIQUES

The four techniques that compute a meteor's Doppler velocity actually determine the meteor's Doppler frequency. We transform the Doppler frequency, f_D , to Doppler velocity, v_D , using the following equation:

$$v_D = \frac{\lambda f_D}{2}, \quad (1)$$

where λ represents the wavelength of the signal [16]. Jicamarca's radar has a wavelength of 6 meters so the relation to convert Doppler frequency to Doppler velocity for Jicamarca data is

$$v_D = 3f_D. \quad (2)$$

Frequency Spectrum Analysis

The fast Fourier transform (FFT) is performed directly on the complex voltage data detected from the radar's antenna to obtain its corresponding frequency spectrum. Figure 3 is an example of an application of the FFT on a sample simulated pulse. The corresponding frequency of the maximum amplitude is the meteor's Doppler frequency.

It should be noted that the FFT is a discrete Fourier transform. Therefore, the resulting frequency spectrum may not have a clearly defined maximum amplitude. In order to make the resulting spectrum more useful, zero padding is a technique applied to the original signal prior to the FFT [17]. The result is a more refined and “continuous” plot. Figure 4 shows the FFT of the same pulse as Figure 3, but with zero padding added. In this case, the Doppler frequency obtained is more accurate.

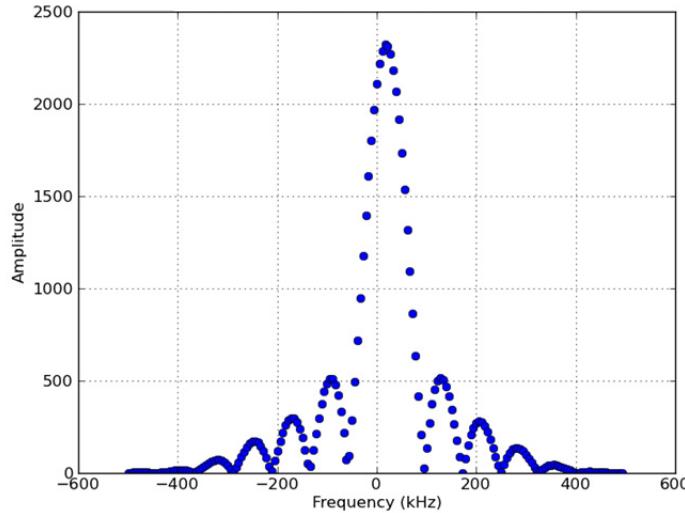


Figure 3: The fast Fourier transform of a sample simulated pulse of 180 samples and frequency resolution 5,555 kHz.

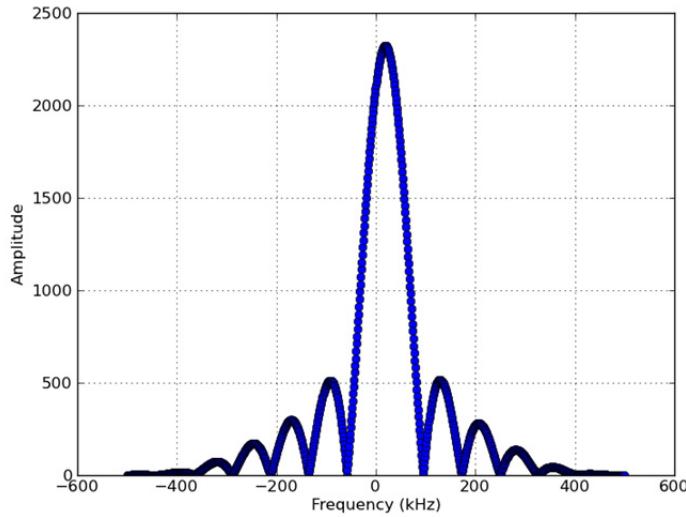


Figure 4: The fast Fourier transform of the same simulated pulse as Figure 3 with zero padding added totaling 1024 samples.

Least square fitting is also used in frequency spectrum analysis. This technique fits a curve to the discrete frequencies to get a more refined and continuous spectrum allowing for more accurate results. Note that like zero padding, fitting does not add new information, but rather, it refines the information already present. Figure 5 shows an example of fitting on Figure 3.

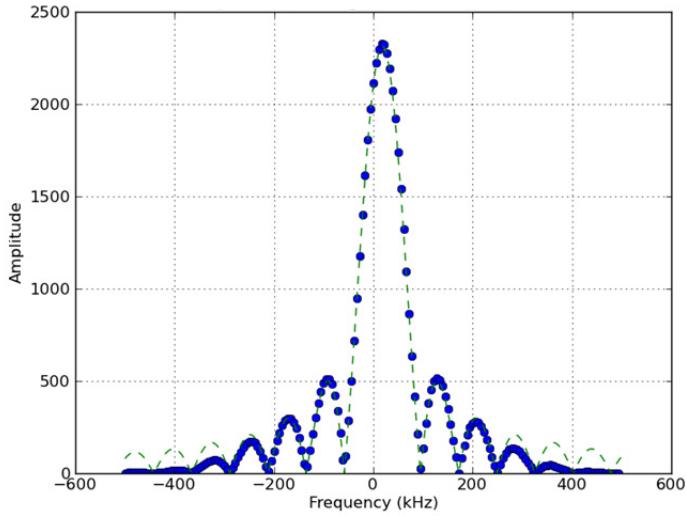


Figure 5: Least square fitting on the fast Fourier transform of the same simulated pulse as Figure 3.

Pulse-to-Pulse Computation

Pulse-to-pulse computation is performed on non-coded and coded long pulses. The Doppler frequency is obtained through finding the autocorrelation between two back to back profiles. The two profiles are given by S_1 and S_2 :

$$S_1 = \exp[-j(-\omega t + \theta)] \quad (3a)$$

$$S_2 = \exp[-j(-\omega(t + IPP) + \theta)], \quad (3b)$$

where $\omega = 2\pi f_D$ is the angular frequency of the meteor in radians per second, IPP is the inter pulse period in seconds, and θ is the phase of the original pulse in radians. The autocorrelation function is then calculated as

$$ACF = S_1 S_2^* = \exp[2\pi f_D(IPP)]. \quad (4)$$

Thus, solving for the Doppler frequency, f_D , we obtain

$$f_D = \frac{\phi(ACF) + 2\pi N}{2\pi(IPP)}, \quad (5)$$

where $\phi(ACF)$ denotes the angle of the autocorrelation function in radians and IPP is in units of seconds. It should be noted that the phase of the autocorrelation function is aliased in the computation. To anti-alias the phase, $2\pi N$, where N is an integer, is added to the phase. The approximated value of Doppler velocity obtained through a range-time velocity estimation is used to compute the proper value of N in order to retrieve an accurate value of the Doppler frequency, f_D [18, 19].

Pulse Decoding

Pulse decoding is performed on coded long pulses. These pulses are generally decoded by cross-correlating the original transmitted coded pulse with the head echo signal returned by the meteor to remove the phase. Then, the result is analyzed for the meteor's Doppler velocity [19, 20]. There are three ways to obtain the Doppler frequency from a coded long pulse through decoding.

The first way is to take windows of samples from the received data and decode those samples by multiplying the windows with the code. The fast Fourier transform is then performed on this long pulse to determine the Doppler frequency of the meteor.

The second pulse decoding technique is the convolution of the data and the code. The data obtained by the radar, S_r , is in the form

$$S_r = B(t) \exp[j\omega t], \quad (6)$$

where $\omega = 2\pi f$ and $B(t)$ is the code. Then, to obtain the non-coded signal of the meteor, S_m , we have

$$S_m = (S_r \exp[-j\omega_D t]) \otimes B(t) \text{ where } \omega_D = 2\pi f_D, \quad (7)$$

$$S_m = \mathfrak{I}^{-1} \left\{ \mathfrak{I} \{ S_r \exp[-j2\pi f_D t] \} \times \mathfrak{I}^* \{ B(t) \} \right\}, \quad (8)$$

where \mathfrak{I} represents the fast Fourier transform, \otimes represents convolution, and f_D is the Doppler frequency of the meteor. The Doppler frequency can be found directly by analyzing the frequency spectrum prior to the final inverse Fourier transform of Equation (8).

The third decoding technique divides out the Fourier transform of the code:

$$S_m = \mathfrak{I}^{-1} \left\{ \frac{S_r}{\mathfrak{I} \{ B(t) \}} \right\}. \quad (9)$$

Similar to the second decoding technique, the Doppler frequency is found by analyzing the frequency spectrum prior to the final inverse Fourier transform of Equation (9).

Linear Fitting of the Received Echoes

Linear fitting of the received echoes is performed on non-coded long pulses. The natural logarithm of the complex part of Equation (3a) is in the form of a linear function:

$$y = \omega t + \theta, \quad (10)$$

where $\omega = 2\pi f_D$ and θ represents the phase of the received echo data. A linear least square fitting is applied to the natural logarithm of the carrier pulse and the resulting slope can be solved for the Doppler frequency. Figure 6 shows a graph of this technique on a simulated pulse in one profile (the data collected in one IPP).

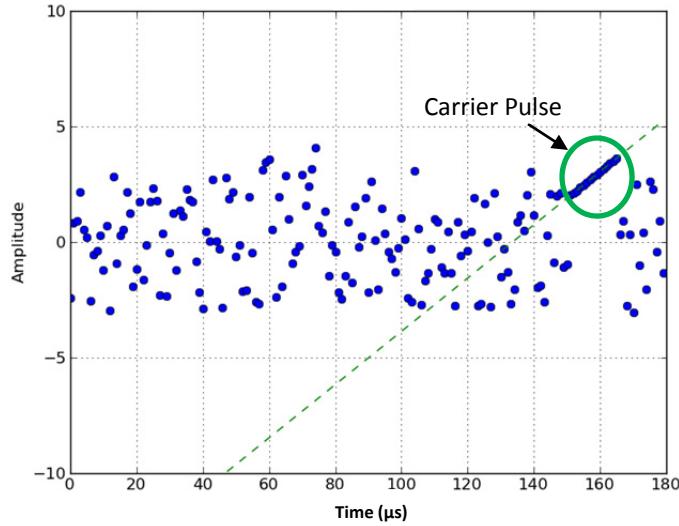


Figure 6: Linear fitting of the received echoes of one profile. Note that only a small set of points clearly define a line. The rest of the points are due to noise.

III. ANALYSIS OF TECHNIQUES

Each technique was tested and analyzed with modeled data representing data received by Jicamarca. The result, computation time, standard deviation, and root mean square were obtained from each technique for analysis. The modeled data was first analyzed with no added noise and then analyzed with noise. Finally, the optimal technique was determined for the analysis of the experimental data from Jicamarca. Figure 7 shows the two examples of the different types of long pulses that were analyzed—non-coded and Barker-13 coded—with noise. The modeled Doppler velocity is 55 km/s.

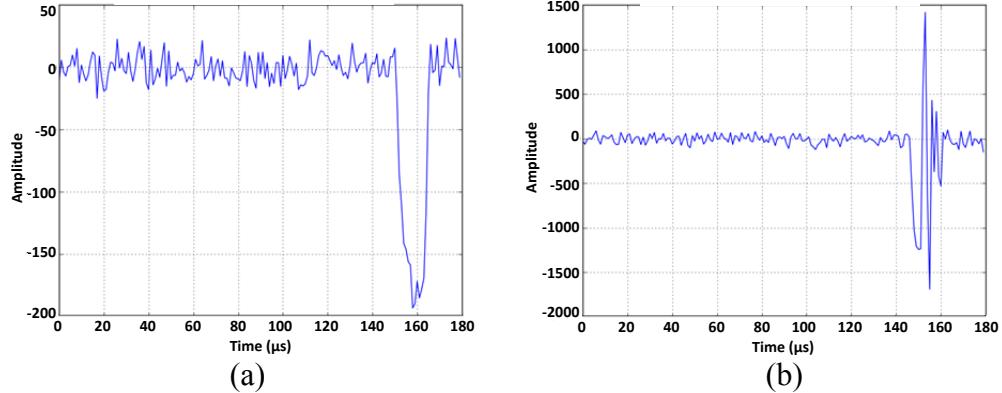


Figure 7: These figures were produced with a head echo moving at 55 km/s. (a) A profile of a long non-coded pulse with noise. (b) A profile of a long coded pulse with noise.

Spectrum Analysis

With a small number of samples, frequency spectrum analysis provides consistent, but inaccurate results. Frequency spectrum analysis without noise obtained a Doppler velocity of about 50 km/s. When zero padding was used on a long non-coded signal without noise, more refined results were obtained at the cost of computation time. The frequency spectrum analysis with zero padding obtained a Doppler velocity of 55.2 km/s, which is much closer to the modeled velocity.

When noise was introduced to the signal, the spectrum analysis led to scattered, inaccurate results. This is because the non-coded long signal is 13 bits wide, making the frequency spectrum, which is almost like a sinc function, wider. When the noise is introduced to the discrete frequency spectrum, an adjacent point near the peak may become the new peak and result in a false Doppler frequency. Figure 8(a) shows the frequency spectrum of Figure 7(a). Figure 8(b) shows the results of frequency spectrum analysis with zero padding. Notice that the results fluctuate between about 50 and 67 km/s. The noise that was introduced to the discrete pulses affected the peaks randomly and resulted in a large standard deviation and a larger root mean square.

Least square fitting for frequency spectrum analysis provides results near the modeled velocity and a relatively low standard deviation. One main drawback with this technique is the necessity to assume certain values for the fitting of the sinc function. With the presence of noise, these values may not always converge to the optimal function. Additionally, least square fitting took the longest amount of time to compute compared to the FFT and zero padding. Figure 9(a) shows a fitting to Figure 8(a). Figure 9(b) shows the results of fitting to the frequency spectrum with noise. Note that the results are scattered, but they do center about the simulated velocity of 55 km/s.

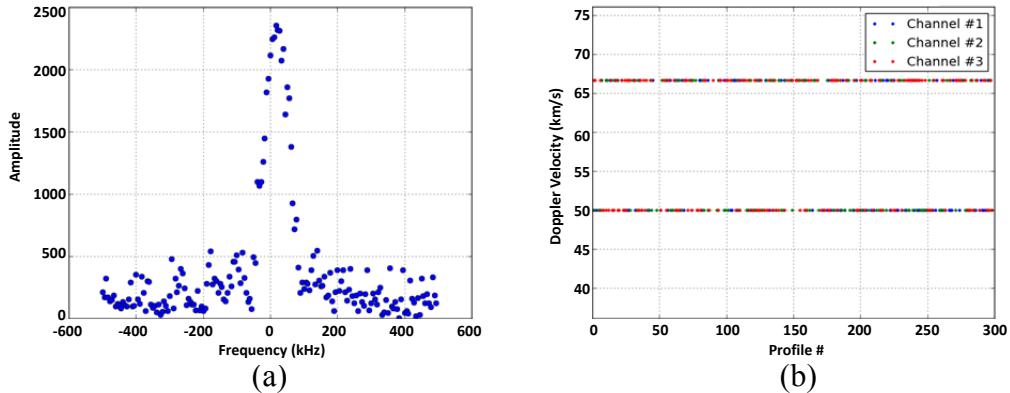


Figure 8: (a) The frequency spectrum of Figure 8(a). (b) The results of frequency spectrum analysis.

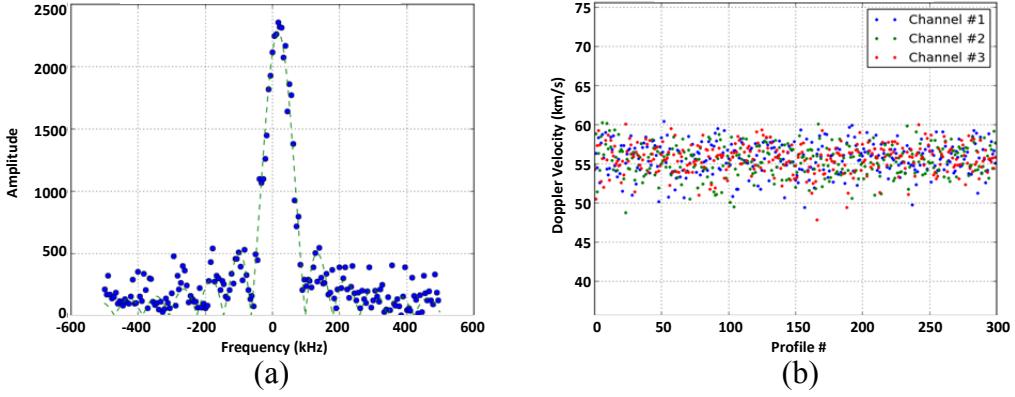


Figure 9: (a) A typical fitting to the frequency spectrum. (b) The results to the frequency spectrum with fitting.

Pulse-to-Pulse Analysis

Pulse-to-pulse computation provides good results for both non-coded and coded long pulses. The phase shift between each profile is easily determined and is anti-aliased through the application of the range-time velocity estimation. The resulting Doppler velocities are all within 0.1 km/s of the theoretical value, and additionally, the standard deviation is low even with a noisy signal. This technique takes a shorter amount of time to compute than fitting on the frequency spectrum and pulse decoding. Figure 10 shows the results of pulse-to-pulse computation for both types of pulses with noise.

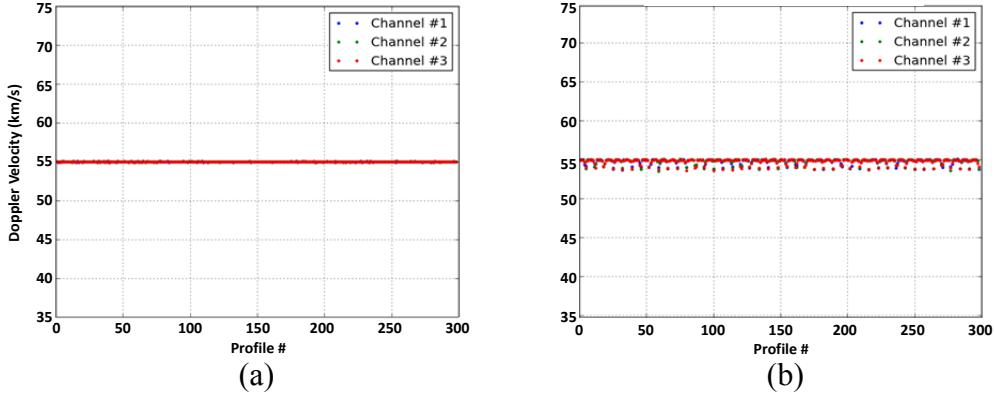


Figure 10: (a) The results of pulse-to-pulse computation for non-coded long pulses with noise. (b) The results of pulse-to-pulse computation for coded long pulses with noise.

Pulse Decoding Analysis

The first decoding technique is not accurate and has a large standard deviation as shown in Figure 11(a). The second and third decoding techniques provide worse results as shown in Figure 11(b). They both provide the same outputs despite different approaches adapting to the way the modeled data was produced.

The large standard deviations of pulse decoding techniques are due to the necessity of performing the FFT on the noisy data. Additionally, the time and memory required for computation are excessive. Thus, the pulse decoding techniques are not the best techniques for Doppler velocity estimation.

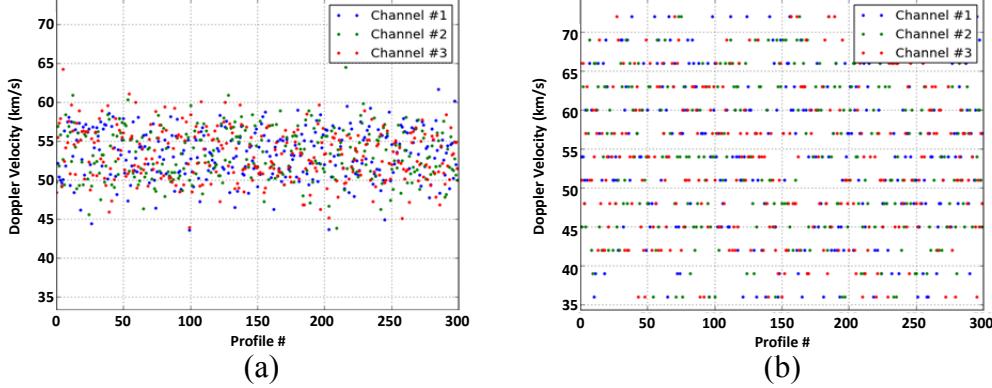


Figure 11: (a) First pulse decoding technique with noise. (b) Second and third decoding technique with noise.

Linear Fitting of the Received Echoes Analysis

The linear least square fitting on the natural logarithm of the complex signal provides very good results. Figure 12(a) shows the natural logarithm of the amplitude of the complex portion of Figure 7(a). Note that there is a short section in time where there is a relatively straight line. The slope of that line represents the Doppler frequency and is relatively constant for all the profiles of the meteor head echo as shown in Figure 12(b). In general, the results are in the vicinity of the theoretical Doppler velocity of 55 km/s. As a result, this technique is very useful for calculating an accurate estimate of the Doppler velocity for long non-coded pulses in a short amount of time.

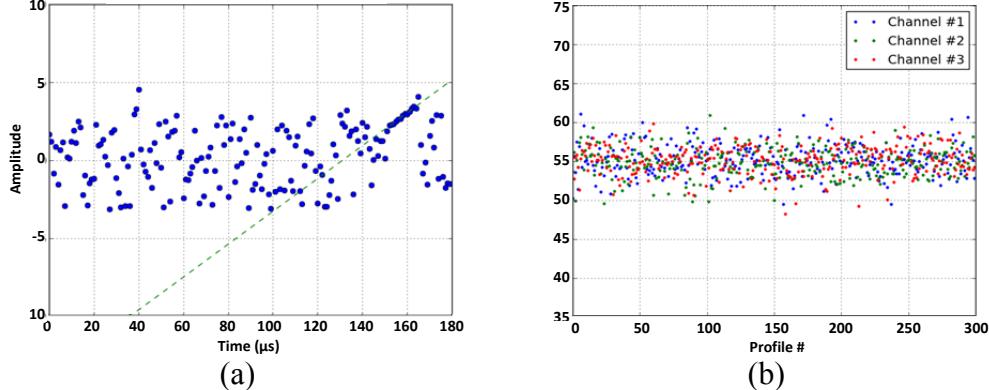


Figure 12: (a) The natural logarithm of the pulse with noise. (b) The results of linear fitting to the natural logarithm of the pulse with noise.

Selection of Optimal Doppler Estimation Techniques

Table 2 shows the data collected from analyzing these techniques. This table shows the average computed velocity, computation time, standard deviation (STD), and root mean square (RMS).

Table 2: Results of Modeled Data Analysis

Technique	Samples	Average Velocity (km/s)	Time (s)	STD (km/s)	RMS
Just FFT	180	56.11	0.08	22.38	1.85
FFT + Fitting	180	55.55	18.84	1.98	0.092
FFT + Zero Padding	256	55.69	0.10	21.84	1.92
	1024	55.77	0.18	22.08	1.93
	4096	55.64	0.58	22.07	1.92
	16384	55.64	3.43	22.07	1.93
	256	55.34	22.26	1.98	0.088
FFT + Zero Padding + Fitting	1024	55.33	55.55	1.91	0.078
	4096	55.32	197.50	1.91	0.077
Fitting of the Received Echoes	-	55.07	2.88	1.90	0.081
Pulse-to-Pulse Analysis	-	55.02	1.84	0.025	0.0007
First Decoding	-	53.31	51.6	2.05	0.084
Second Decoding	-	54.85	225.88	18.38	1.53
Third Decoding	-	54.85	203.28	18.38	1.53

The best technique for Doppler velocity estimation for both non-coded and coded long pulses is the pulse-to-pulse computation. This technique has the lowest standard deviation and root mean square, meaning it provides the most precise results. Although the average time of computation is moderate, the accuracy outweighs the time requirement. One drawback of this technique, however, is the necessity of an a priori range-time velocity estimation for anti-aliasing the phase. If an insufficient amount of profiles are provided, it would be impossible to anti-alias the phase and obtain the proper Doppler velocity. Compared to pulse-to-pulse computation, the impact of noise on the other techniques is quite large.

IV. EXPERIMENTAL DATA

There were several differences between the experimental data from Jicamarca collected on April 26th, 2005 and the modeled data. First of all, the parameters are different (see Table 1). The pulse length is longer, meaning the range resolution of the data is lower and the data may consist of “merged” meteor head echoes. Additionally, the experimental data has several meteor head echoes per 2.6 seconds unlike the modeled data, in which there was only one per set of data.

Thus, a section or window of the entire data must be partitioned for analysis of the Doppler velocity.

Since it was concluded from the modeled data that pulse-to-pulse computation is the most accurate technique, it was used on the experimental data. At first, the results obtained from pulse-to-pulse computation from the non-coded and coded data were the same, but once the window size or position was changed, the result changed. Thus, there is a problem somewhere with the pulse-to-pulse computation. One problem is regarding the size of the IPP of the experimental data. From Equation (5), the larger the IPP, the harder it is to anti-alias the result. In the experimental data, the IPP is 400 km and thus, the Doppler velocity estimate must be within $2\pi/\text{IPP}$, which is 1.125 km/s. If there was a smaller IPP, pulse-to-pulse computation would be more forgiving on the velocity estimation. In this case, a very accurate estimation is necessary to obtain the best Doppler velocity. Unfortunately, a very precise estimation of Doppler velocity is not available through range-time velocity estimation, which is required. The range resolution of the data is poor, since the pulse length is 16.8 km or 112 μ s, and thus, this estimation can be less accurate. In addition, the sampling rate is 4 μ s and quite noisy, resulting in inaccuracy. One way to rectify this problem for non-coded long pulses is to replace range-time velocity estimation with linear fitting of the received echoes. For coded pulses, decoding can be used. This will provide a relatively accurate Doppler velocity estimate despite the radar parameters.

V. CONCLUSION

From the modeled data, we concluded that pulse-to-pulse computation is the most accurate technique for Doppler velocity estimation of meteor-head echoes. The other techniques, especially those that use the FFT, were greatly affected by noise and provided inaccurate results. Pulse-to-pulse computation did not provide fruitful results when used on experimental data and weaknesses of this technique were exposed. One such weakness is this technique's dependence on its range-time velocity estimation. The acceptable error of this estimation is dependent on the IPP of the received echo. In the experimental data from Jicamarca on April 26th, 2005, the large IPP makes this technique dependent on a very accurate a priori Doppler velocity estimate. Unfortunately, due to the large pulse length, noise, and slow sampling rate, range-time velocity estimation does not provide an accurate a priori estimation for anti-aliasing. Linear fitting of the received echoes can replace range-time velocity estimation for non-coded long pulses while decoding can be used for coded pulses to rectify this problem. To minimize this problem, data should be collected from a radar with a small IPP and a fast sampling rate, making pulse-to-pulse computation the best Doppler velocity estimation technique. Further testing is necessary to see how reliable these modifications to pulse-to-pulse computation are. Overall, this project paves the way for an improvement of pulse-to-pulse computation and has the potential to make pulse-to-pulse computation accurate regardless of the radar parameters.

ACKNOWLEDGMENTS

We would like to thank Dr. Kenneth Jenkins and Dr. Sven Bilén for organizing the 2011 Penn State Electrical Engineering Research Experience for Undergraduates program. In addition, we would like to thank those involved in the program, especially Tom Tyson and Lena Getman. The Jicamarca Radio Observatory is part of Instituto Geofísico del Perú and mainly supported by the United States National Science Foundation through a cooperative agreement with Cornell University. This material is based upon work supported by the National Science Foundation under Grants No. EEC-1062984, ATM-0457156, and ATM-0638624.

REFERENCES

- [1] J. A. M. McDonnell and D. J. Gardner, "Meteoroid Morphology and Densities: Decoding Satellite Impact Data," *Icarus*, 133, 25-35, Jan 1998.
- [2] C. G. Gros, "Meteoroids as a possible hazard to space vehicle operations: An annotated bibliography," May 1963.
- [3] J. L. Chau and R. F. Woodman, "Observations of meteorhead echoes using the Jicamarca 50MHz radar in interferometer mode," *Atmos. Chem. Phys.*, 5, pp. 511-521, 2004. Available: www.atmos-chem-phys.org/acp/4/511/
- [4] S. Close, M. Oppenheim, D. Durand, and L. Dyrud, "A new method for determining meteoroid mass from head echo data," *Journal of Geophysical Research*, vol. 110, 2005.
- [5] D. Janches, J. D. Mathews, D. D. Meisel, and Q.-H. Zhou, "Micrometeor Observations Using the Arecibo 430 MHz Radar," *Icarus*, vol. 145, pp. 53-63, 2000.
- [6] D. Janches, A. Pellinen-Wannberg, G. Wannberg, A. Westman, I. Haggstrom, and D. Meisel, "Tristatic observations of meteors using the 930 MHz European Incoherent Scatter radar system," *Journal of Geophysical Research*, vol. 107, no. A11 1389, 2002.
- [7] Elizabeth Bass, Meers Oppenheim, Jorge Chau, and Alice Olmsted, "Improving the Accuracy of Meteoroid Mass Estimates from Head Echo Deceleration," *Earth Moon Planet*, 102, pp. 379-382, 2008.
- [8] Jicamarca Radio Observatory, "Jicamarca Radio Observatory." [Online]. Available: <http://jro.igp.gob.pe/english/index.htm>
- [9] W. Swartz, D. Farley, and R. Woodman, "The Jicamarca Radio Observatory." [Online]. Available: <http://jicamarca.ece.cornell.edu/overview.html>
- [10] D. T. Farley, "The equatorial *E* region and its plasma instabilities: A tutorial," *Ann. Geophys.* 18 Feb. 2009.
- [11] C.H. Wen, J.F. Doherty, J.D. Mathews, and D. Janches, "Meteor detection and non-periodic bursty interference removal for Arecibo data," *Journal of Atmospheric and Solar-Terrestrial Physics*, 67, pp. 275-281, 2005.
- [12] L. Dyrud et al., "Plasma and Electromagnetic Simulations of Meteor Head Echo Radar Reflections," *Earth Moon Planet*, 102, pp. 383-394, 2008.
- [13] W. G. Elford, "Novel Applications of MST radars in meteor studies," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol 63, pp. 143-153, 2001.
- [14] M. M. Oppenheim, G. Sugar, E. Bass, and Y. S. Dimant. "Day to night variation in meteor trial measurements: Evidence for a new theory plasma trail evolution," *Geophysical Research Letters*, vol. 35, L03 102, Feb 2008.
- [15] G. Wannberg, A. Pellinen-Wannberg, and A. Westman, "An ambiguity function based method for analysis of Doppler decompressed radar signals applied to EISCAT measurements of oblique UHF-VHF meteor echoes," *Radio Science*, vol. 31, no. 3, pp. 497-518, May-June, 1996.

- [16] Bassem R. Mahafza, *Radar Systems Analysis and Design Using MATLAB*. Boca Raton: Chapman & Hall, 2000.
- [17] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 3rd Ed. Upper Saddle River, New Jersey: Simon & Schuster, 1996.
- [18] J. D. Mathews, D. D. Meisel, K. P. Hunter, V. S. Getman, and Q. Zhou, “Very High Resolution Studies of Micrometeors Using the Arecibo 430 MHz Radar,” *Icarus*, 126, 157-169, 1997.
- [19] F. R. Galindo. “Caracterizacion de meteoros, utilizando el radar VHF del Radio Observatorio de Jicamarca durante la presencia de ecos coherentes de Electrochorro Equatorial,” Master Thesis, Universidad Nacional de Ingenieria, Lima, Peru, Jan 2007.
- [20] C.-H. Wen, J. F. Doherty, and J. D. Mathews, “Time-Frequency Radar Processing for Meteor Detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, Mar. 2004.

STUDY OF PROCESSING TECHNIQUES FOR REMOVAL OF EQUATORIAL ELECTROJET ECHOES TO ENHANCE METEOR DETECTION AT JICAMARCA

Dimitri Ressetar,* Julio Urbina,[#] and Freddy Galindo⁺

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

*Undergraduate Student of
Electrical Engineering Program
School of Science, Engineering, and Technology
The Pennsylvania State University – The Capital College
Middletown, PA 17057

ABSTRACT

Meteor echoes typically appear on high-power large-aperture (HPLA) radars at Equatorial latitudes between 70 and 140 km altitude. Coherent and strong echoes caused by the Equatorial Electrojet (EEJ) overlap the same heights and inhibit the study of meteor-head echoes, especially at the Jicamarca Radio Observatory. This project studies different schemes to remove EEJ echoes. Implementation of these EEJ removal schemes will allow for the future design of a robust algorithm for meteor detection, which can lead to the successful extraction of more meteors from that data than is presently possible by collecting usable data 24 hours per day. In order to accomplish this task, different methods are studied, such as: signal statistics traditionally used in meteor detection, coherence and Doppler velocity, pulse-to-pulse subtraction, and high and low threshold detectors. These techniques are based in discriminating the main and critical differences between head and EEJ echoes.

Traditional meteor signal statistical methods are found to be unable to discriminate meteor-heads from EEJ echoes and noise. However, the pulse-to-pulse subtraction technique is shown to be extremely beneficial because of its simplicity in removing much of the EEJ. Also, schemes combining individual methods helped diminish enough of EEJ echoes to enable successful meteor-head detection. Using predefined constant thresholds, meteor-head echoes can be accurately detected. Future work will be needed in automating the threshold and

[#] Faculty Mentor

⁺ Graduate Mentor

detection process to enable current technologies and signal processing to operate when EEJ echoes are present.

I. INTRODUCTION

The Jicamarca Radio Observatory (JRO), located near Lima, Peru (latitude 11.95° South, longitude 76.87° West), was constructed in 1960–61 by the United States National Bureau of Standards under the direction of Dr. Kenneth Bowles. It was originally designed to study the earth's ionosphere from the ground as the first large radar implementing the incoherent scatter technique proposed by Professor William E. Gordon of Cornell University in 1958 [1].

Atmospheric research continues today at Jicamarca, including a study of micro-meteors [2, 3]. Every day billions of meteoroids, most less than 1 mm in diameter, enter Earth's atmosphere. These meteors are too small to be seen visually, or with optical instruments, or even directly with radar. Instead, they are detected as they travel through the upper atmosphere and ablate, forming electrically conductive plasma. This occurs at around 70 to 140 km in altitude, with average meteor velocities around 60 km/s [3, 4]. Because of its conductivity, the plasma reflects radio waves and is detectable on Jicamarca's high-power large-aperture (HPLA) radar as a meteor-head echo [5]. These meteor-head echoes are radar data that represent the meteor's movement and can provide much information about the meteoroids and how they interact with the earth: velocity, deceleration, orbital parameters, etc. The mass and population can be derived from the aggregate data [4].

The main problem with meteoroid detection at Jicamarca is the presence of atmospheric interference at around 100 km altitude. Much of this interference is caused by the presence of the Equatorial Electrojet (EEJ) [3], a narrow moving stream of electrically charged particles about 600 km wide centered near the equator which flows across the earth in an east to west direction. The Electrojet ranges in altitude from approximately 90 to 150 km, with most of the interference appearing in the range of 90 to 110 km [6, 7]. Unfortunately this range overlaps with a portion of the ideal heights for meteor detection, 70 to 140 km [3].

Current meteor studies at Jicamarca generally collect data at times when the EEJ is absent or weak compared to head echoes [3]. Two processing schemes are used. The first is automatic but conservative; only one meteor is detected for a given time period. It does not work when the data are contaminated with EEJ interference [3]. The other scheme uses manual detection. It is robust but time consuming because a skilled operator must select each individual meteor by hand [4].

Since no automatic scheme can presently operate during periods when EEJ echoes are present, there are several hours each day during which information about meteors cannot be known due to atmospheric interference. Thus, this project explores methods to reduce EEJ interference so current techniques and signal processing can be modified to extract meteors even during times when the EEJ is present.

II. EXPERIMENT DESCRIPTION

First, signal statistics used for meteor detection are analyzed for information that may decipher the difference between meteor and EEJ echoes. These include received power, coherence, correlation, as well as auto-correlation and Doppler velocity. Then, pulse-to-pulse subtraction of the radar data is performed to reduce slow moving and periodic interference. Finally, various schemes to remove EEJ are presented, which combine signal statistic methods with mathematical and logical functions.

Four different sets of data from Jicamarca were used in this experiment. Sets 0 and 1 contain both meteor-head and EEJ echoes, set 2 contains no noticeable meteor-heads during a time when the EEJ was present, and set 3 contains only meteor-heads when the EEJ was not present. Unless otherwise noted, all figures showing meteor data are from data set 0.

A. Signal Statistics

Raw signals from the antenna are recorded as complex voltages in digital form. These voltages can be quickly arranged into image profiles, which correspond to time, and ranges. This formatting allows for a quick, organized calculation of signal statistics, which can easily be seen visually with altitude (ranges) on the *y*-axis and time (profiles) on the *x*-axis.

1) Power

Voltages collected from the antenna can be changed into a range-time-intensity (RTI) plot, as seen in Figure 1, from calculating the power in logarithmic form with units of decibels by taking the mean of the complex voltages squared, that is:

$$P = 10 \log_{10} |\bar{V}^2| \quad (1)$$

In the figure, dark areas indicate reflected power of 60 dB. The white background is roughly 40 dB or less. Unfortunately the EEJ echoes present in bands between 95 and 100 km altitude have strong returns at up to 60 dB, while the circled meteor-head echo between 105 and 110 km on the right side is barely visible above the background at only around 52 dB.

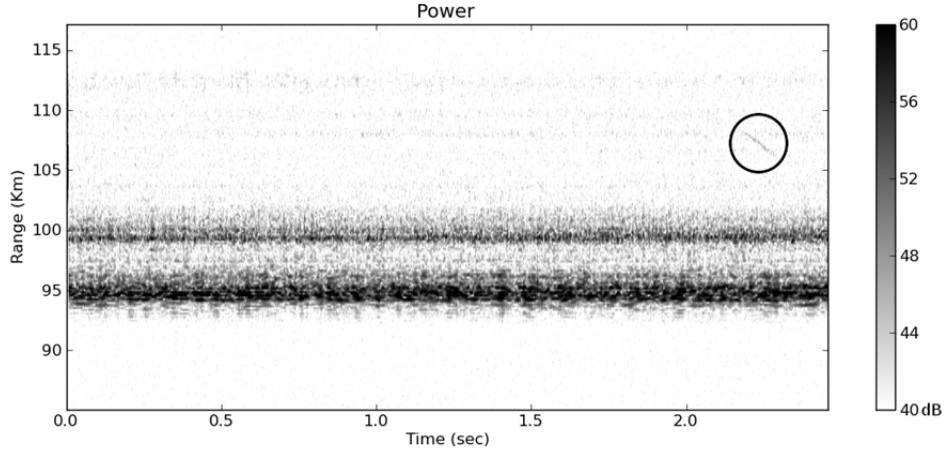


Figure 1: Typical range-time-intensity plot of received power from 80 to 120 km (in decibels). Note that the EEJ echoes contaminate the data. It is worst in the bands between 95 and 100 km altitude, where it reaches 60 dB. The only visible meteor-head echo is circled, and has power intensity around 52 dB.

2) Correlations

Correlations compare signals received from different sources for similarity. Two are used here: coherence, derived from the cross-correlation function (CCF), and the auto-correlation function (ACF). Both methods can be used to help separate meteor signals from noise.

a) Coherence

Coherence compares the data received from different antennas, sometimes called channels, through several mathematical steps. First, CCF is calculated by multiplying the complex voltages received by one channel with the complex conjugate of the voltages received from another channel, that is:

$$\overrightarrow{CCF}_{AB} = \vec{V}_A \cdot \vec{V}_B^* \quad (2)$$

Then, coherence is calculated by normalizing the cross-correlation through dividing its absolute value by the square root of the power function:

$$Coherence = \frac{|\overrightarrow{CCF}_{AB}|}{\sqrt{P}} \quad (3)$$

From Equation 3, plots can be generated which compare the data received from any two radar antennas. Figure 2 compares channels A and B of Jicamarca's antenna. The values range between zero and one, with a 1, black in the figure, indicating complete coherence, that is, the data are the same. Note that in the figure the circled meteor-head echo and bands of EEJ echoes appear dark, indicating high coherence for both. Coherence alone cannot discriminate between

the EEJ and meteors; however, it does indicate the large amount of incoherent noise as the grey background region present in the signal.

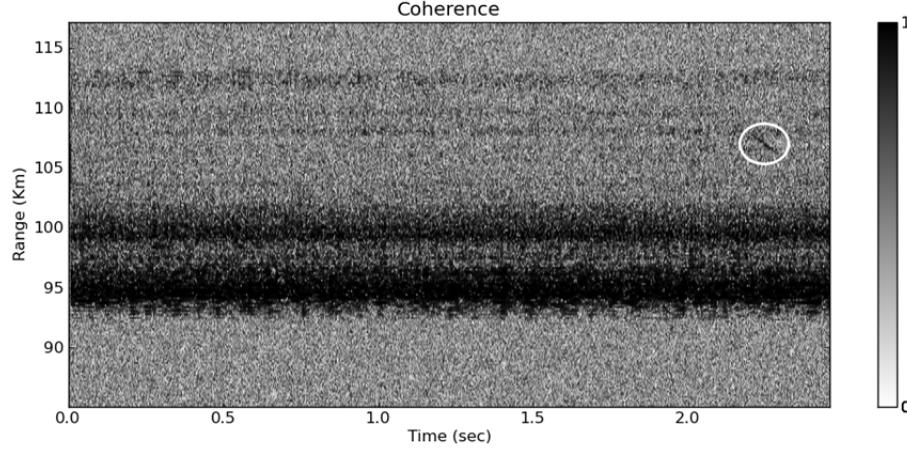


Figure 2: Coherence map between channels A and B. Dark areas indicate high coherence, or similarity, of the data received by each antenna section. Note that the horizontal bands of EEJ and the circled meteor-head echo are dark, indicating that both have high coherence.

b) Auto-Correlation Function (ACF)

The auto-correlation function is the second correlation method which can also provide the Doppler velocities of the meteor-head echoes. It is similar to cross-correlation, but instead of comparing data received from different channels, data from the same channel are compared offset by time. This is accomplished by multiplying the voltage of one profile by the complex conjugate of next profile's voltage, and repeating until all profiles have been processed:

$$ACF_A = V_n \cdot V_{n+1}^* \text{ for } (0 \leq n < \text{Number of Profiles} - 1) \quad (4)$$

Next, the arc-tangent is calculated using both imaginary and real values to obtain the correct phase angle. An example is shown in Figure 3 using the ACF of Channel A. Note that in this case the meteor-head echo has a highly positive phase angle shown in black, while the large bands of EEJ noise have a low phase angle near zero, indicated in grey. In situations such as this, the ACF phase angle can be used to discriminate between the EEJ and meteors. Strong noise is still evident in this ACF phase plot which interferes with meteor detection.

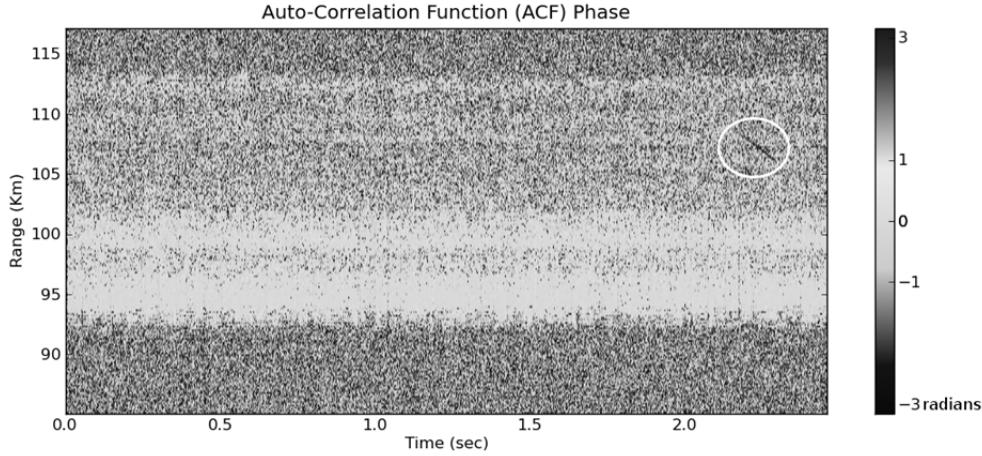


Figure 3: Map of auto-correlation function phase angle (in radians) using Channel A. Angles deviating the most from origin are dark, including the circled meteor-head echo. The horizontal bands of EEJ are light grey, indicating low phase angle.

The two correlation methods provide multiple ways to see the received radar data, but neither have the ability to reliably separate meteor head echoes from noise. This is primarily due to the fact that the EEJ echoes tend to be coherent like meteor-head echoes.

B. Pulse-to-Pulse Subtraction

Pulse-to-pulse subtraction is the single best performing signal modification technique for EEJ removal because of its ability to remove interference in one step. Each profile is subtracted from the adjacent profile to remove “slow moving” signals which do not change between adjacent vertical slices of the plot. This is accomplished in a similar manner to auto-correlation, except using subtraction instead of multiplication: The voltages of each profile are subtracted from the voltage of the previous profile, repeated for all of the data. Fortunately, most of the EEJ moves relatively slowly and linearly while meteors move quickly and diagonally. An example is Figure 4, which now very clearly shows two if not three meteor-head echoes, all circled. Compared to Figure 1, this makes meteors much more obvious but traces of strong noise still appear in the lower regions. Therefore, while greatly beneficial, this method cannot be used alone to remove enough of the EEJ.

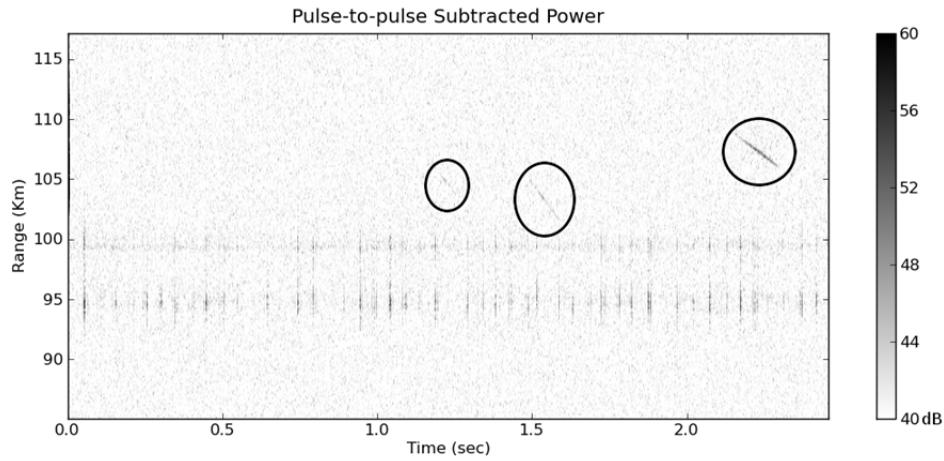


Figure 4: RTI plot of pulse-to-pulse subtracted power (in decibels). Dark areas indicate up to 60 dB are reflected back to the radar. Note that compared to Figure 1, the EEJ echoes have been substantially reduced and a total of three circled meteors are now visible.

C. Schemes for Removal of EEJ

Since no signal statistical method or pulse-to-pulse subtraction can wholly remove EEJ echoes, three different processes were developed in this project by combining individual methods. While each scheme is unique, all of them use the pulse-to-pulse subtraction technique because of its great ability to remove noise.

Before the EEJ is removed, however, it must be determined if its echo is present in the data. This is accomplished by averaging horizontal slices, the ranges, of the plot in two ways. First, the mean of all the horizontal slices is calculated, called the “overall mean”. Second, the mean of each horizontal slice that is in the EEJ region is calculated. If 90% or more means in the EEJ band exceed the overall mean, the EEJ is considered to be present and the data will be processed through the schemes developed to reduce the interference.

1) EEJ Removal Scheme using Pulse-to-Pulse Subtraction and Coherence

The first EEJ removal scheme uses the pulse-to-pulse subtraction technique followed by power and coherence analysis. A flowchart outlining the process is Figure 5. The raw voltage data are immediately processed using the pulse-to-pulse subtraction technique to produce subtracted voltage data. The result is converted to power and coherence data concurrently. Separate high and low thresholds are applied on each data stream. The threshold values are manually defined constant values chosen to capture as many meteor-head echoes as possible while also removing as much noise as possible. A binary array with potential meteor-heads recorded as True values is outputted from each function. The two resulting arrays are combined together using a bitwise AND function that works like an AND gate.

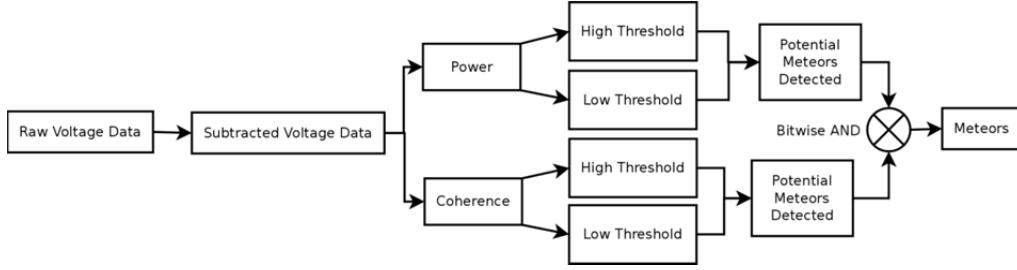


Figure 5: Flowchart demonstrating the EEJ reduction scheme using manually defined thresholds on both the pulse-to-pulse subtracted power data and the coherence function.

Since this scheme uses two separate thresholds each based on different data, the diversity increases the likelihood that most of the noise should be eliminated.

2) EEJ Removal Scheme using ACF and Subtracted Power

Figure 6 documents the flow of the second scheme. Like the first, two separate paths are used and the results combined at the end. However, in this case only half of the procedure use subtracted voltages; the ACF phase is calculated directly using the raw voltages.

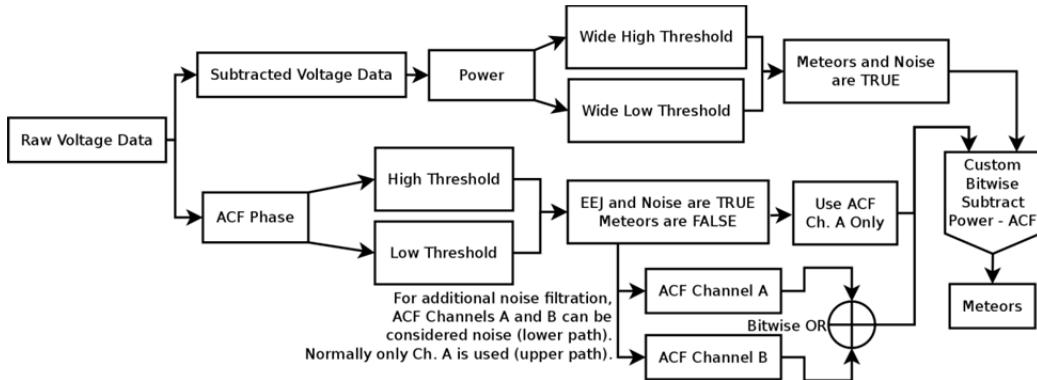


Figure 6: Flowchart demonstrating the EEJ reduction scheme using manually defined thresholds on the ACF phase angle and the pulse-to-pulse subtracted power data. Note that normally only Channel A of the ACF is used but for additional noise reduction the combination of Channels A and B can be used. Also note the custom bitwise subtract which combines the outputs correctly.

Unlike the first scheme, the manually defined constant thresholds set for the power allow more data to pass through. This lets all meteor-heads as well as some of the EEJ and noise to pass through. Since this method is coupled with others, the extra noise will ideally be removed by the end of the process.

Once the ACF is calculated from the raw voltages and passed through manually defined low and high thresholds, the output has the suspected meteor-

heads recorded as *False* values and both the EEJ and noise recorded as *True* values. The values are assigned opposite compared to other methods but for this scheme it simplifies the processing. If additional noise filtration is desired or necessary, a combination of ACF Channels A and B can be used in tandem. Theoretically, even channel C could also be used for further reduction. The bitwise OR function is used on the result as an OR gate to pass all of the True values. The remaining False values are considered candidates for meteor-head echoes.

A custom “bitwise subtract” function was defined to properly combine the binary arrays from power and ACF. Only values that are True in the power array, either meteor-heads or noise, can be True in the result; the ACF array can only be used to remove additional noise or interference. If a value is True in the ACF array because it is detected as EEJ or noise, the value in the result will be False. Otherwise, the result has the same values as the Power array.

This scheme is more complicated than the first since it draws on two different kinds of data that are independent. The pulse-to-pulse subtracted power and the ACF are not derived from a common intermediate source. Because of this separation, this scheme has the potential to more accurately identify meteors by using the ACF to correctly determine EEJ echoes and noise.

3) EEJ Removal Scheme using Subtracted Power Averaging

A final scheme for EEJ removal depends on the use of averages from the data, in a similar way to how the EEJ was detected. The raw voltage data are immediately processed using the pulse-to-pulse subtraction technique. The result is converted to the subtracted power (Figure 4). This time the vertical slices of the plot, the profiles, are averaged. First, the mean of each vertical slice is calculated only within the EEJ bands. Second, the mean of all vertical slices within the EEJ bands is calculated, so-called the “overall mean”. Finally, each vertical mean is individually compared to the overall mean. Any mean that exceeds the overall mean most likely contains the thin dark lines of the EEJ echo that were not removed by the pulse-to-pulse subtraction technique. The corresponding vertical slice is therefore blanked by setting its values within the EEJ region to zero in the array.

Since this scheme begins with the pulse-to-pulse subtracted power data, much of the EEJ echo and incoherent noise has already been removed. The averaging detects the remaining EEJ echo and removes the thin vertical lines of power intensity greater than approximately 50 dB, which should help in the accurate detection of meteor-head echoes since ideally no long lines of interference will pass through this scheme. Also, since operations are only performed within the EEJ echo region of 95 through 100 km altitude, meteor-head echoes contained in other regions of the plot cannot be inadvertently removed. Meteor-heads in the EEJ echo region may be accidentally removed, but their shape may still be seen. Therefore, it is anticipated that this scheme will provide the best overall EEJ reduction.

III. EXPERIMENT RESULTS

A. Signal Statistics

Simple signal statistics methods used in meteor detection have already been demonstrated to be insufficient for fully identifying meteors under times when the EEJ echo is present. Figures 1–3 consistently show high levels of noise in addition to EEJ echoes clouding the meteor-head echoes. This was expected, and necessitated this exploration of additional detection methods.

B. Pulse-to-Pulse Subtraction

The pulse-to-pulse subtraction technique provides the greatest reduction of EEJ echoes and slow noise of any one step method. It also has the ability to “uncover” meteor-heads that have been hidden by dense noise (compare Figure 4 with Figure 1). However, as previously mentioned, the remaining noise prevents reliable meteor-head detection. This technique has become the foundation of the three schemes developed by this project for EEJ removal because of its simple and effective EEJ echo reduction.

C. Schemes for Removal of EEJ

Before the EEJ can be removed, its echo must be identified in the data. The EEJ detector routine accurately detected presence of the EEJ echoes in data sets 0, 1, and 2. Only data set 3 did not contain EEJ; it was correctly identified as a signal without EEJ.

1) EEJ Removal Scheme using Pulse-to-Pulse Subtraction and Coherence

Figure 7 shows the result of the first EEJ removal scheme. For data set 0, two meteors out of the three visually identified in Figure 4 are still present. While this method worked well for data set 0, it did not work at all for data set 1 which contained both one meteor-head echo and EEJ. Figure 8 shows the missing meteor outlined. It is not discernable since it appears the same as the remaining background noise.

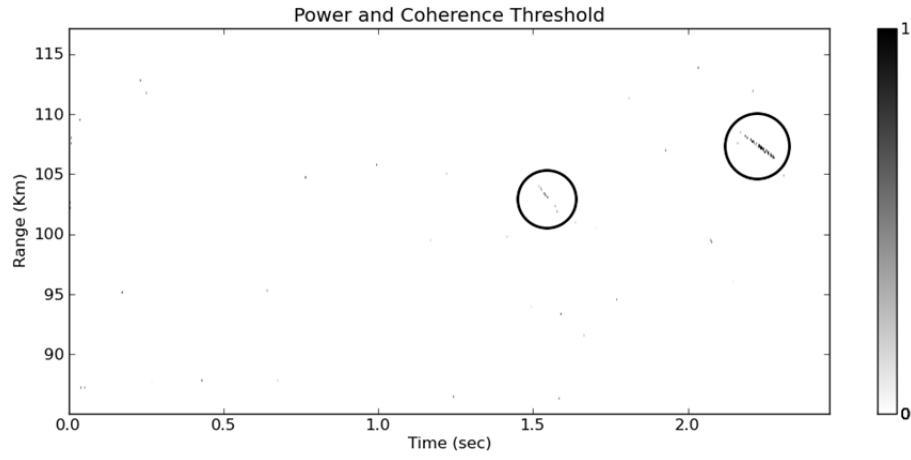


Figure 7: Result of the first EEJ removal scheme on data set 0 using a manually defined constant threshold on both pulse-to-pulse subtracted power and coherence. Note that the plot is binary and contains two out of the three visually identified meteors-heads from Figure 4. Very little noise is present as random small black dots in the background.

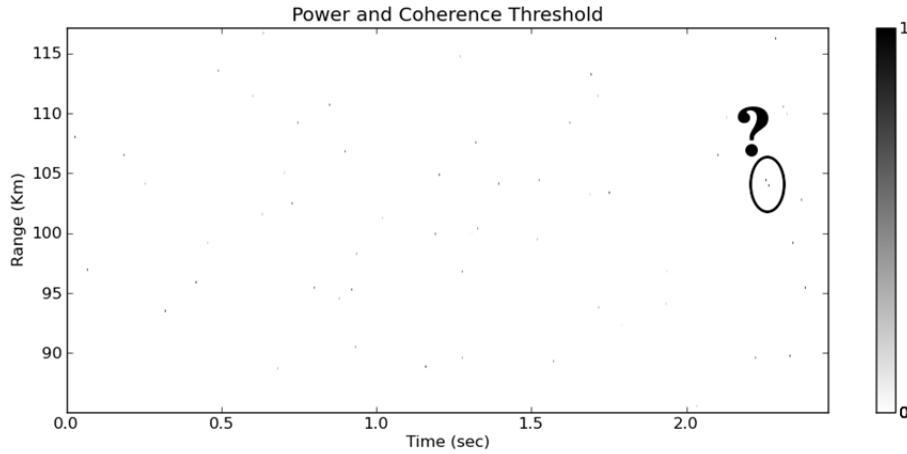


Figure 8: Result of the first EEJ removal scheme applied to data set 1. The one meteor-head present in the signal was not detected. Its location is circled and labeled.

2) EEJ Removal Scheme using ACF and Subtracted Power

The second EEJ removal scheme had results similar to the first, as shown in Figure 9.

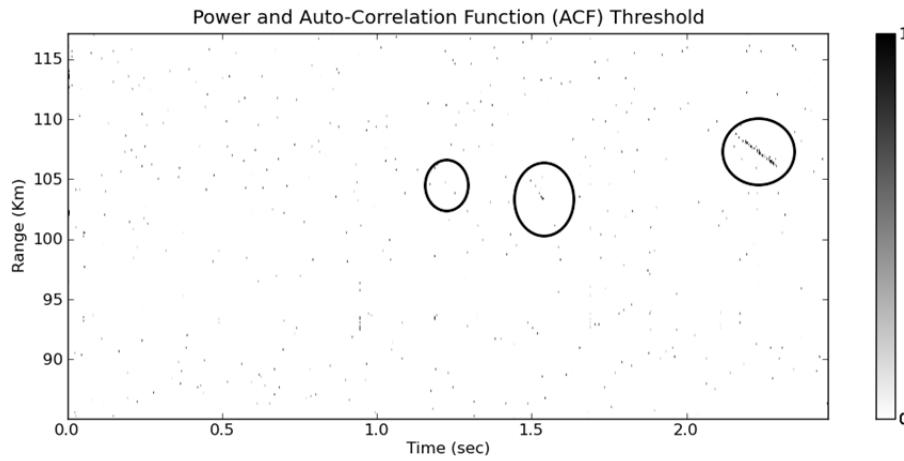


Figure 9: Result of the second EEJ removal scheme using a manually defined constant threshold on both the ACF and pulse-to-pulse subtracted power. Note that the plot is binary and noticeably contains two meteor-heads, inside the right and center circles. The remaining meteor-head circled on the left is not discernable in the relatively high noise still present in the background.

This scheme identified many noise points as potential meteors, as evidenced by the plethora of black dots in the background. Two out of the three meteor-head echoes previously visually identified are still noticeable as trails. The third meteor-head on the left became lost in the background noise because it had low relative power intensity.

3) EEJ Removal using Subtracted Power Averaging

Comparing with Figure 4, removing the remaining EEJ echo remnants without disturbing the rest of the data in the region visually shows less clutter and makes the three meteor-head echoes more prominent, as shown in Figure 10. Figure 11 is the same plot but additionally processed through the pulse-to-pulse subtracted power and coherence scheme. Like Figures 7 and 9, only two out of three meteor-heads were identified against a background of light noise.

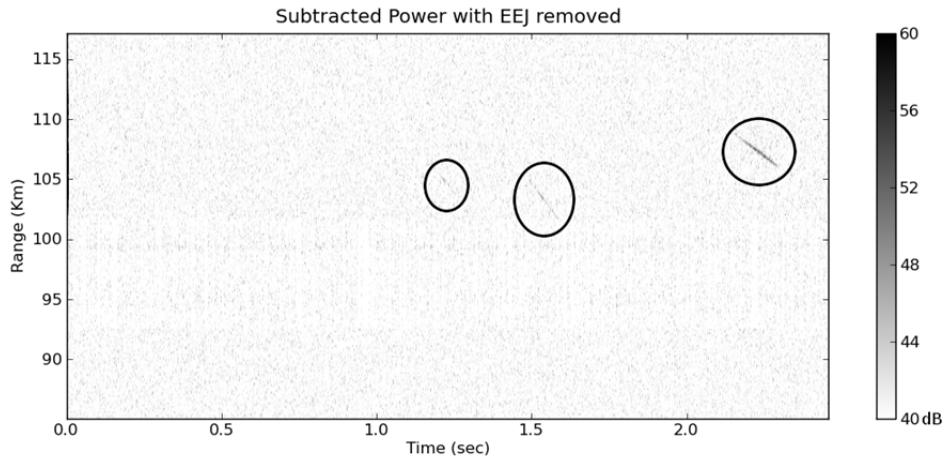


Figure 10: Remaining EEJ echoes from Figure 4 successfully removed, leaving the three meteor-head echoes identifiably present.

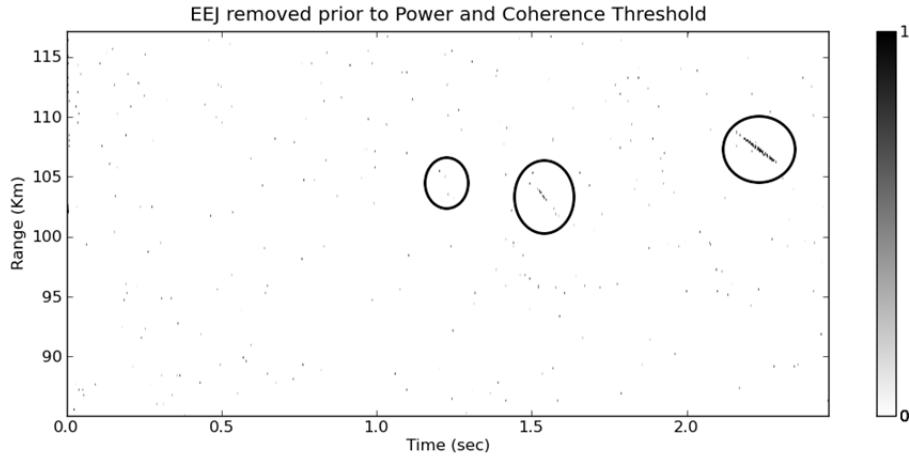


Figure 11: Result from the subtracted power averaging EEJ removal scheme further processed through the pulse-to-pulse and coherence scheme. The two right circled meteor-heads are identifiable; the left has been lost in the noise. Black dots in the background indicate noise identified as meteor-head candidates.

The pulse-to-pulse subtraction and coherence scheme on a data set 1 incorrectly removed all traces of a meteor-head echo in Figure 8. When the same data were first processed through the subtracted power averaging scheme, the one meteor-head became a slightly noticeable trail. Figure 12 illustrates this, though it also is clouded with many black dots incorrectly identified as meteor candidates.

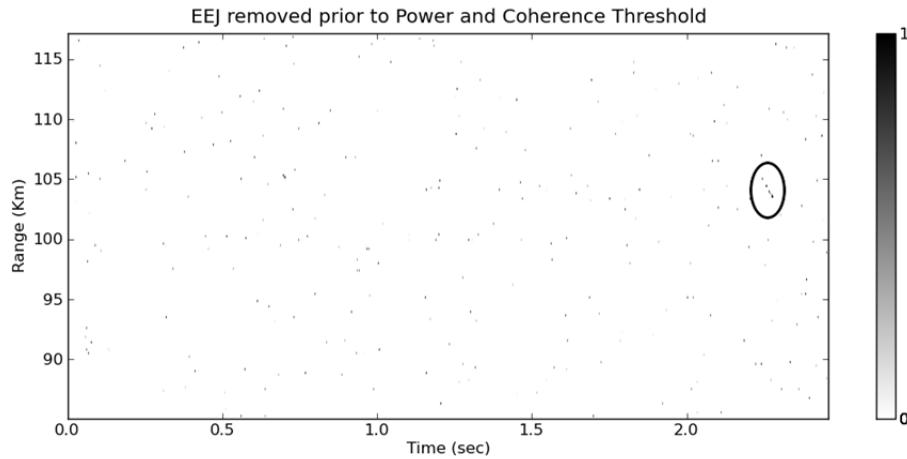


Figure 12: Data set 1 processed first with the subtracted power averaging scheme followed by the pulse-to-pulse subtracted power and coherence threshold scheme. One meteor-head, circled, is revealed. The meteor-head echo was not detected by the threshold scheme alone.

IV. DISCUSSION

Due to the high power and high coherence of the EEJ echoes, it quickly became apparent that no traditional meteor signal statistics method would be able to discern the meteor-head echoes. Even though visually it is possible to see a meteor trail in the data, the computer cannot be easily programmed to detect this pattern when surrounded with intense noise.

The pulse-to-pulse subtraction technique was used in each EEJ removal scheme and should be considered in future works because in all cases it has removed most of the EEJ and slow moving noise without completely obscuring meteor-head echoes. In this project, it has enabled two additional meteor-heads to become visible on data set 0 (Figure 4).

Each of the three newly developed schemes for EEJ removal show promising preliminary results as tested in this project. While all the test data are from Jicamarca, providing the best testing environment, there are natural variations in the amount and intensity of EEJ echoes, noise, and meteor-heads. Variations are expected since the radar monitors the ever-changing atmosphere, but the schemes presently are unable to adapt to the changes. Further work will be necessary to facilitate this adaption before use in meteor detection experiments at Jicamarca while EEJ echoes are present.

The first two schemes are dependent on manually defined constant thresholds, which were picked during the programming process to best detect the most meteor-head echoes while filtering out as much of the remaining noise as possible. Constant thresholds do work as demonstrated above, but between data set, time of day, or season, preset values will quickly categorize the data incorrectly. In the course of this experiment, this happened between data sets 0

and 1. Figure 8 shows this failure. Therefore, in order to complete these processing schemes, a method for automatically determining the best threshold value for each data set will need to be developed before these schemes can be used in production.

The last scheme relies mostly on averaging which produces more consistent results between data sets, allowing it to out-perform any other method. Figure 10 shows that the EEJ is removed almost completely. The averaging automatically adapts to the makeup of each data set. In some cases, this allows meteor-head echoes to be detected when other methods failed to do so, as in Figure 12 which illustrates a meteor-head detected from data set 1 that was ignored by the first two schemes.

Power averaging does not detect meteor-head echoes, even though its EEJ reduction is superior to other schemes. Instead, it relies on the other schemes and their manually defined thresholds to identify meteors. Therefore, in any case, the manual threshold is the main feature of this project that will need additional work, even though many of the aspects of each scheme yielded beneficial results.

V. CONCLUSION

Traditional meteor signal statistical methods discussed earlier have been shown to be inadequate for identifying meteor-head echoes in the presence of EEJ echoes. They are useful, however, because they are component parts of the three schemes for EEJ removal that have been developed by this project. While the pulse-to-pulse subtraction technique remains the strongest tool for EEJ reduction across data sets, it is only when used with other operations that its full utility is realized.

Each of the three developed schemes demonstrated fruitful results in their initial stages. It is indeed possible to separate meteor-head echoes from EEJ echoes and background noise. Future work to tweak and innovate on these schemes, by way of an automatic threshold detection method, will allow for many beneficial meteor studies. Eventually it may be possible to integrate these detection methods into current signal processing at Jicamarca.

ACKNOWLEDGMENTS

The Jicamarca Radio Observatory is operated by the Instituto Geofísico del Perú (IGP) with financial support from the United States National Science Foundation (NSF) through a cooperative agreement with Cornell University. Thank you to Julio Urbina and Freddy Galindo, mentors for this project, for sharing their expertise in this field. Gratitude is expressed to W. Kenneth Jenkins and Sven Bilén, directors of the 2011 Penn State Electrical Engineering Research Experience for Undergraduates program, and to Tom Tyson and Lena Getman for their assistance to the program. Complaints are directed to the students who constantly abused the printers in the Hammond computer lab and to the staff who did not repair or refill the printers once during the course of this project. This

project was sponsored under NSF grants ATM-0457156, ATM-0638624, and EEC-1062984.

REFERENCES

- [1] R. T. Frost, W. P. Richardson, and C. M. Benedict, *Scatter Radar: Space Research From the Ground*. Washington, DC: National Bureau of Standards, United States Department of Commerce, 1963.
- [2] W. Swartz, D. Farley, and R. Woodman, *The Jicamarca Radio Observatory (JRO)*, November 1997. [Online]. Available: <<http://jicamarca.ece.cornell.edu/overview.html>>.
- [3] J. L. Chau and R. F. Woodman, "Observations of Meteor-Head Echoes Using the Jicamarca 50 MHz Radar in Interferometer Mode," *Atmospheric Chemistry and Physics*, vol. 4, no. 2, pp. 511-521, March 2004.
- [4] J. L. Chau, R. F. Woodman, and F. Galindo, "Sporadic Meteor Sources as Observed by the Jicamarca High-Power Large-Aperture VHF Radar," *Icarus*, vol. 188, no. 1, pp. 162-174, May 2007.
- [5] L. Dyrud et al., "Plasma and Electromagnetic Simulations of Meteor Head Echo Radar Reflections," *Earth, Moon, and Planets*, vol. 102, no. 1-4, pp. 383-394, June 2008.
- [6] C. A. Onwumechili, *The Equatorial Electrojet*. Amsterdam: Overseas Publishers Association, 1997.
- [7] D. T. Farley, "The Equatorial E Region and its Plasma Instabilities: A Tutorial," *Annales Geophysicae*, vol. 27, no. 4, pp. 1509-1520, April 2009.

CONSIDERATIONS FOR AN INTRA-SOLAR SYSTEM LASER SATELLITE DATA NETWORK

Christopher W. Jones* and John Mathews[#]

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

*Undergraduate Student of the
School of Electrical Engineering and Computer Science at
Washington State University
Pullman, Washington 99163

ABSTRACT

Current technology is sufficient to implement a Solar-System wide satellite communications network. Possible system specifications are explored here at distances of up to 250 AU or about 37 billion km. At these distances light travel time is nearly 1.5 days. By shortening the distance between satellite transponders we can maintain a robust network with gigabit-per-second transfer rates when using near-IR lasers and modest sized optics.

INTRODUCTION

Current Technological Capability

The present technology that allows us to use Global Positioning Satellites (GPS) gives us, through GPS, the ability to measure points in space and time with precision. Because relativistic corrections are needed to reach the current level of precision, we have been able to experimentally verify certain aspects of special and general relativity as predicted by Einstein. In addition to its value in math and physics, the GPS system and its ability to maintain time and frequency allows for reliable, global high speed data transmission, i.e., the Internet. In the future, a similar system will be able to support space exploration, test large scale physical theories, and effectively place communications beacons outside the Sun's theorized bowshock, around 250 AU.

A recent paper [1] explores a hypothetical two-way laser communications link between Earth and Mars. In this paper we will extend this idea to an Exploration data NETwork (eNET). At 250 AU the light, and data, travel time is almost 1.5 days. A narrow-beam laser signal would be very difficult to find/detect from earth

[#] Faculty Mentor

at these distances. One way to increase robustness of communications at these distances is to find suitable locations to place transponders. By using stable orbital positions near planets known as Lagrangian points we can shorten the distances to the Interstellar Medium (ISM). We will present a table with distances, over a 200-year period, between a set of planets to show what the path options look like as we approach Earth from the ISM. The paths will be evaluated on their data transfer ability and the characteristics that affect it. In our

EXPERIMENT DESCRIPTION

Beamwidth Divergence versus “Look-ahead”

In the above-mentioned paper [1] by Degnan there was a need for a “look-ahead” pointing mechanism. An example of look-ahead pointing is the view from Mars. From Mars’ surface the image of the Earth appears as it would one light transit time ago and in order to effectively reach the receiver we must look-ahead two light transit times from the image of the Earth. However, at larger distances the beam width divergence will have a pattern cross section that is greater than two light transit times, thereby eliminating the need for look-ahead pointing. Differential pointing is achieved in Degnan’s paper through the use of Risley prisms. The basic function here is the implementation of fine beam steering by rotating two or three prisms in the beam path [2]. This allows the target to remain centrally located in the beam path for a longer time interval than coarse pointing allows. In the Experimental Results section below we will evaluate the appropriateness of this look-ahead pointing versus coarse pointing and beam width divergence. This information will be used to suggest different designs for transponders with different average path lengths.

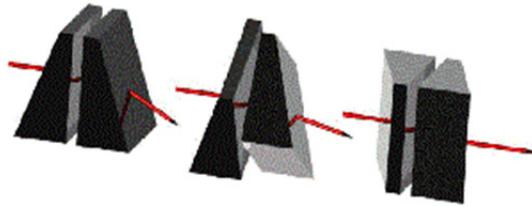


Figure 1: Risley Prisms. By rotating the prisms differential beam-steering is possible. The image on the left shows maximum deflection.

Pulsar Positioning

In a 2009 paper [3], pulsars are explored as a means of galactic positioning. Using Minkowski’s space-time coordinates, (x,y,z,t) , provides the four pieces of information necessary to say where and when an event occurred. The 2009 paper enlists four signals, $(\tau_1, \tau_2, \tau_3, \tau_4)$, from four different pulsars to provide the necessary information to define an event. The four pulsars, 0751+1807 (3.5 ms), 2322+2057 (4.8 ms), 0711–6830 (5.5 ms) and 1518+0205B (7.9 ms), have a nearly even angular distribution which resembles the vertices of a tetrahedron. The position of our Solar System would place us inside the tetrahedron. There is up to a 4-ns uncertainty associated with any of the signals from its travel through the ISM. The uncertainty allows for a spatial resolution of about 1m. They note that, while this is not highly precise, it is extremely stable and has an enormous

domain of validity. Finally, an origin of the system is chosen and all satellites leaving Earth must continually record the four signals in order to bring the chosen coordinates along. This information along with a catalog of other transponders will be used in the acquisition of target receivers.

The drawback to the pulsar positioning scheme is the necessity of large radio telescopes to receive the signals.

Future research may include finding a system that acquires these signals and makes them readily available.

Time/Frequency Standard

In the absence of the pulsar positioning system, onboard atomic clocks can provide the necessary temporal precision at sub-nanosecond levels. This would lead to accuracies in the decimeter range [4]. That paper recommends using a rubidium frequency standard and a ‘timing distribution module’. It appears that the true limits of accuracy and data transfer are mostly a product of our ability to maintain synchronicity through clock distribution. Whether time/frequency is kept by pulsars or atomic clocks, it is this fundamental characteristic that will measure the performance of the network.

Future research may include research on the most economical and robust way to transfer time throughout the system.

System Catalog and Ephemerides

It is necessary for there to be an onboard catalog of objects of interest in the solar system and their related ephemerides, or when and whereabouts. It must be extendable and continuously updated as the permanence of any of the satellites orbits is not guaranteed. It would be an object of future research to investigate the best ‘base’ coordinate system, i.e., (x,y,z,t) or $(\tau_1,\tau_2,\tau_3,\tau_4)$. The elliptic equations involving used to generate an ephemeris have been solved [5] to include solutions in both Cartesian and spherical coordinate systems.

Future research may include the solution in a pulsar coordinate system.

Stable Orbits and Lagrange Points

When considering where the nodes or transponder should be placed in the solar system, we looked for places of long term stability. This is seen in nature with the Trojan and Greek asteroids in stable co-orbit with Jupiter. They occupy the L4 and L5 points. In a simplified two-body system The L4 and L5 points occur at the vertices of equilateral triangles with the two bodies at the other vertices. In an elliptical orbit they are approximately 60 degrees ahead and behind the smaller orbiting body. The L4 point precedes the body and the L5 follows it.

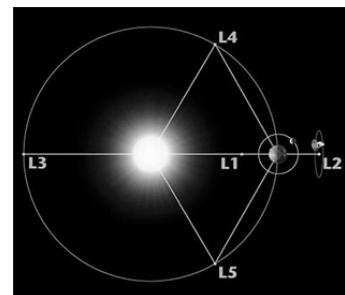


Figure 2: Lagrange points in a two-body system. They indicate gravitationally stable locations.

Antenna Size and Its Dependant Equations

The main feature of this system will be the large mirror used for transmitting and receiving the laser signals. The following equations relate the system characteristics to the theoretical rate of data transfer.

In this paper we use the following relationship between wavelength λ and diameter d to calculate a 3-dB beamwidth $\Delta\theta$ in radians.

$$\Delta\theta = \frac{4\lambda}{\pi d} \quad (1)$$

Calculating the number of photons is done per pulse of time length τ , where τ is equivalent to the inverse of the bandwidth. The energy per pulse E_T is divided by the energy per photon, giving the number of photons per pulse.

$$N_{\text{photons}} = \frac{E_T}{hc/\lambda} \quad (2)$$

In order to calculate the number of photons received by a dish of diameter d at distance R from the transmitter the number of photons per pulse is multiplied by the ratio of the receiving dish area to the area of the beam. This in turn is affected by the efficiencies of the transmitter and receiver, ε_T and ε_R , respectively.

$$N_{\text{received}} = \frac{E_T}{hc/\lambda} \cdot \frac{\pi(d/2)^2}{\pi(R\cdot\Delta\theta/2)^2} \cdot \varepsilon_T \cdot \varepsilon_R \quad (3)$$

This can be simplified to:

$$N_{\text{received}} = \frac{E_T}{hc/\lambda} \cdot \left(\frac{d}{R\cdot\Delta\theta}\right)^2 \cdot \varepsilon_T \cdot \varepsilon_R \quad (4)$$

Assuming that the given bandwidth is Nyquist sampled we can calculate the noise energy E_{noise} over a pulse of time length τ .

$$E_{\text{noise}} = k_B T_{\text{sys}} \quad (5)$$

where k_B is Boltzmann's constant and T_{sys} is the system temperature. Using the noise to generate a dark count average would require E_{noise} to be divided by the energy per photon.

$$N_{\text{dark}} = \frac{k_B T_{\text{sys}}}{hc/\lambda} \quad (6)$$

With the calculations for N_{received} and N_{dark} we can determine the signal-to-noise ratio (SNR). This also allows us to use the Shannon equation for channel capacity C in bits per second [6].

$$C = BW \cdot \log_2 \left(1 + \frac{N_{\text{received}}}{N_{\text{dark}}} \right) \quad (7)$$

BW is the bandwidth over which the signal is transmitted and is measured in Hz.

It is through these equations that the theoretical performance of the system will be evaluated.

EXPERIMENTAL RESULTS

Planetary Distances

Several graphs of distances will be presented illustrating the variability of path distance from planetary orbits that will be frequently used. It is noteworthy that at certain times the shortest distance from Neptune to Earth is the direct route.

The distances are calculated from NASA JPL ephemeris [7] data over a roughly 200 year period from January 1900 until September 2099. The distance to the Solar System's bowshock is estimated to be about 230 AU or 34 billion km. The maximum path length considered is 250 AU or 37 billion km. An AU is an Astronomical Unit which is approximately 149 million km, the distance from the Sun to the Earth.

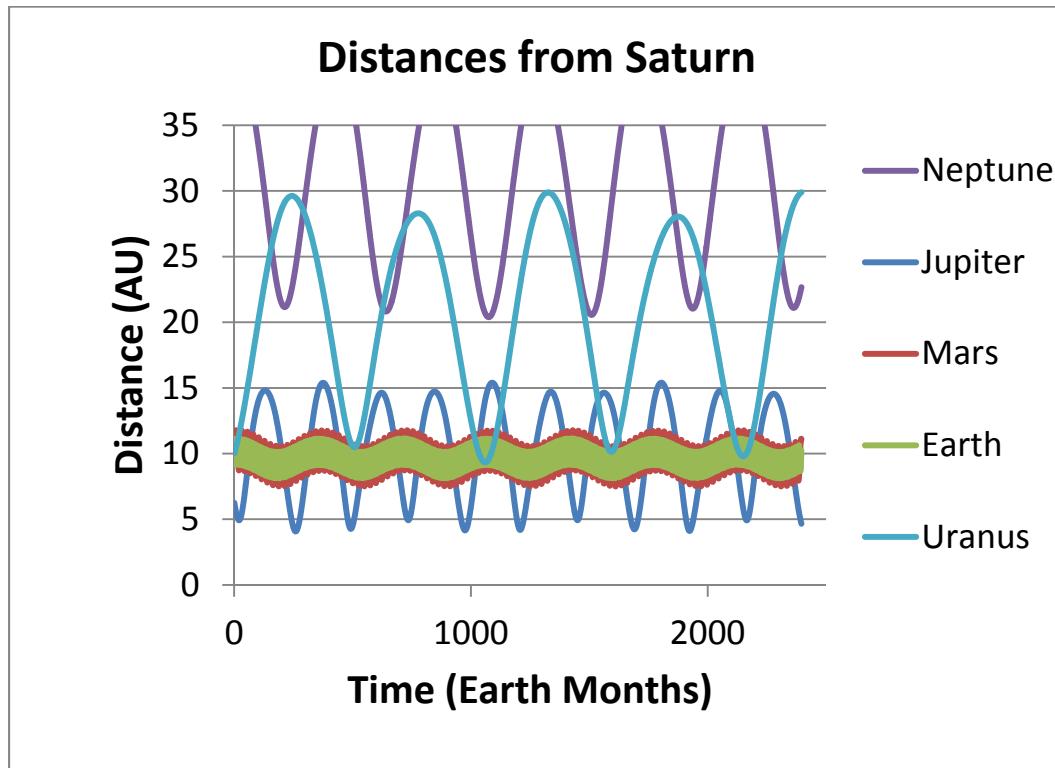


Figure 3: Distances from Saturn

Data on the minimums for each month in the above charts is summarized in the following table. Distances are to the next closest target and given in km. To

clarify, the ‘Maximum’ column has the highest minimum value recorded over the 200 year period.

Table 1: Minimum Distances to Next Target. These figures were calculated from all the minimum values used in Figure 3.

Planet	Minimum (AU)	Mean (AU)	Maximum (AU)
Neptune	10.7	23.6	31.5
Uranus	9.33	16.0	21.3
Saturn	4.08	8.03	11.1
Jupiter	3.56	4.80	6.52
Mars	0.38	1.72	2.70

Beamwidth Divergence and Look-Ahead Pointing

Using Equation (1) we calculate an upper and lower bound for antenna size. All subsequent calculations will be evaluated with a minimum dish size of 24 cm and a maximum of 10 m.

Table 2: Beamwidths. Divergence is a measure of how much ‘spread’ a beam has.

Aperture Diameter	0.24 m	10.0 m
BW	5.64 μ rad	0.135 μ rad

Knowing the targets mean speed in orbit will help to determine the amount of ‘Look-Ahead’ needed, if any. This table contains the mean speeds of the planets in question. The information will indicate whether or not differential pointing is necessary.

Table 3: Mean Planetary Speeds (km/s). The angular velocity of the planet with respect to their orbital centers.

Planet	Neptune	Uranus	Saturn	Jupiter	Mars	Earth
Mean Speed	5.432	6.795	9638	13.050	24.077	29.783

We will use the Table 3 to determine the length of time that the receiver will be inside the 3-dB beamwidth after an initial acquisition. Table 4 gives the minimum and maximum radii of beamwidths that correlate to the distances in Table 2. Table 4 also shows the minimum time in seconds that a particular target will occupy the beam radius.

Table 4: Minimum Beam Times. Once the transmitter is pointed at the location of the receiver the appropriate amount of light travel time ahead, these times indicate how long the aim will be good for.

Neptune

Aperture Diameter	0.24 m	10.0 m
Min r (km)	8,918	213.5
Uranus (s)	1312	31.4
Earth (s)	299	7.17

Uranus

Aperture Diameter	0.24 m	10.0 m
Min r (km)	7,791	186.5
Neptune (s)	1434	34.3
Saturn (s)	808	19.4
Jupiter (s)	597	14.3
Mars (s)	324	7.75
Earth (s)	262	6.26

Saturn

Aperture Diameter	0.24 m	10.0 m
Min r (km)	3,403	81.46
Jupiter (s)	261 – 712	6.24
Mars (s)	141 – 386	3.38
Earth (s)	114 – 312	2.73

Jupiter

Aperture Diameter	0.24 m	10.0 m
Min r (km)	2,969	71.06
Saturn (s)	308	7.36
Mars (s)	123	2.95
Earth (s)	99.6	2.40

Mars

Aperture Diameter	0.24 m	10.0 m
Min r (km)	316.5	7.576
Earth (s)	10.6	0.25

Channel Capacity

Using Equation (7) we can now look at the data transfer rates over the given path distances.

The approximate working distances of 40 million km to 37 trillion km, or up to 250 AU, leave transfer rates from about 10 kbits per second to about 2.8 Gbits per second. The lower line is for the 24-cm dish using 1-mJ pulse energy. The upper line is for the 10-m dish using 4-J pulses. Notably, the Voyager 2 probe is at a distance of about 96 AU and has transfer rates of 160 bits/s using RF

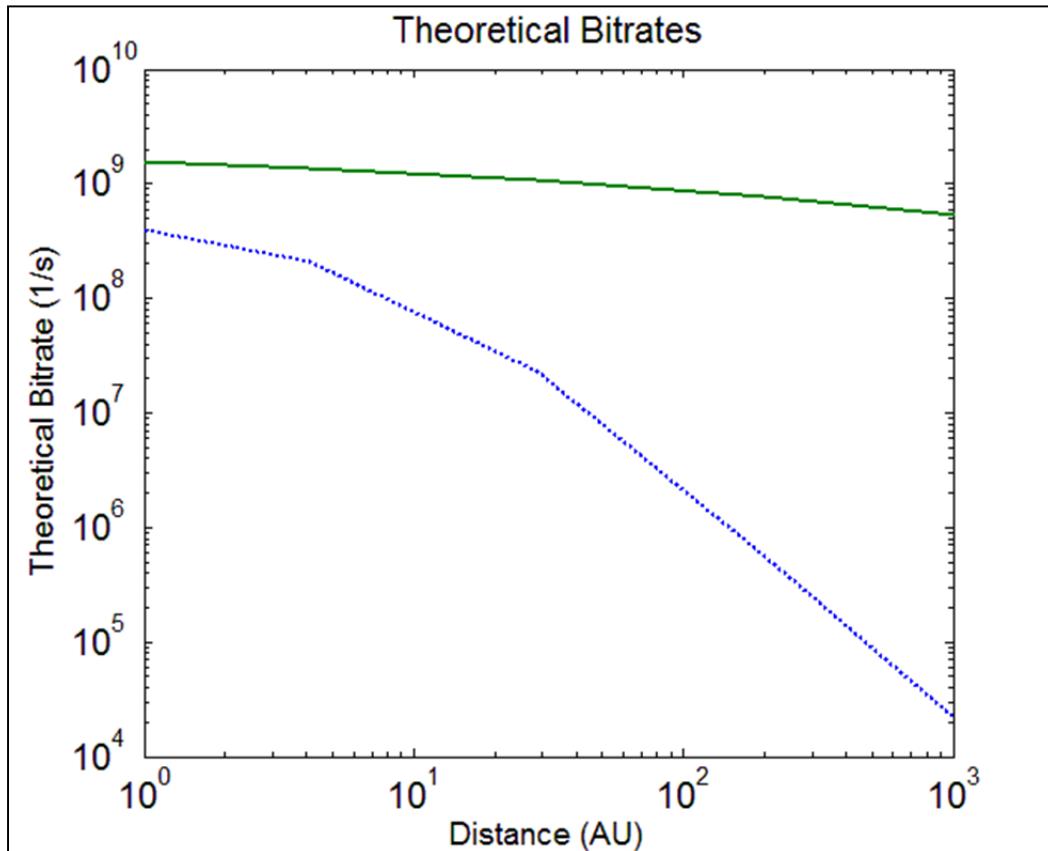


Figure 4: Theoretical Bitrate – The solid upper line is the 4-J pulse with the 10-m dish. The lower line is the 1-mJ pulse with the 24-cm dish. The right boundary is deep in the ISM.

DISCUSSION

With the beam transit times from the ‘Beamwidth Divergence’ section of the Experimental Results it seems more appropriate to select a transponder with differential pointing. The 0.25-second transit time would still likely be able to transmit about 300 Mb before a re-pointing was necessary. And certainly there is an optimization available that addresses the range of transit times with a certain dish size and pulse energy.

Getting the satellites to their position may likely involve the numerous asteroids that populate the orbits of Earth and Mars. The satellites that would

reside further from the Sun would have smaller orbital velocities and this fact may lead to a different approach in their transit.

With the global ability to maintain more precise time and to transmit that time standard around the Earth we have greatly expanded the human species technological purview. Everything from cell phone service, GPS, air traffic control, and even the internet itself rely on that ability. The creation of a more robust GPS network led to direct applications of both General and Specific Relativity. What sort of scientific, technical, and sociological paradigms will be encountered with a system that operates throughout the Solar System?

There are also the questions that arise if we put a satellite outside of our Solar System: will anyone notice it? What happens when they do? It is almost like putting a roadside sign outside of our Heliosphere to indicate to the Universe that, yes indeed, there is life inside this little bubble. The ability to get data to and from the outer reaches of our Sun holds many possibilities.

CONCLUSION

The theoretical calculations certainly indicate that the system is feasible with current technology. Even with conservative figures the data rates can remain near the gigabit per second range. Future exploration would be greatly enhanced and our ability to collect information and develop infrastructure in the Solar System is vital to any movement that involves leaving the Earth.

Future research in these areas is offered near the end of the respective sections above.

ACKNOWLEDGMENTS

C. Jones would like to thank Dr. John Mathews for this tremendous, positive experience and to thank Thomas Tyson and the others who facilitated the NSF Summer EE REU program this year at Penn State. Your support and leadership was greatly appreciated. This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984.

REFERENCES

- [1] Degnan, et al. (1997). "Design and Test of a Breadboard Interplanetary Laser Transponder." 11th International Workshop on Laser Ranging, (pp. 716-728). Deggendorf, Germany.
- [2] Ostaszewski, et al. (2006). "Risley Prism Beam Pointer." Free-Space Laser Communications VI. Proceedings of SPIE, Vol 6304, 630406.
- [3] Coll, Tarantola (2009). "Using Pulsars to Define Space-Time Coordinates." arXiv:0905.4121v1.
- [4] Degnan, J.J. (1996). "Compact laser transponders for interplanetary ranging and time transfer." Proc. 10th International Workshop on Laser Ranging, (pp. 24-31). Shanghai, China.
- [5] P. Bretagnon, G. Francou (1988). "Planetary theories in rectangular and spherical variables.VSOP87 solutions." Astronomy and Astrophysics 202 (pp. 309-315).
- [6] Shannon, C. (1949). "Communication in the Presence of Noise." Proceedings of the IRE, Vol 37, No. 1, (pp.10-21).
- [7] <http://ssd.jpl.nasa.gov/horizons.cgi#top>.

SYSTEM DEVELOPMENT AND INTEGRATION OF COMMUNICATION AND POWER LINK BETWEEN OLITE 2 SATELLITE AND HIGH ALTITUDE STUDENT PLATFORM

Christopher A. Galvan,* Sven G. Bilén,[#] Robert Capuro,^x and Allen Kummer⁺

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
New Mexico State University
Las Cruces, NM 88001

ABSTRACT

Students at Penn State are developing the OSIRIS Lite 2 high altitude balloon payload to demonstrate a CubeSat spacecraft bus. The spacecraft bus is targeted for use on SSPL's proposed OSIRIS CubeSat mission. This paper reviews the design process employed to build a communication and power link between the OLite 2 satellite and the High Altitude Student Platform. A systems engineering approach was employed to develop and integrate the system.

INTRODUCTION

This project focuses on the system design and integration of a power and communication connection between OSIRIS Lite 2 satellite (OLite 2) and the High Altitude Student Platform (HASP). The data collected during this high altitude environmental test will enable further development of the Orbital System for Investigating the Response of the Ionosphere to Stimulation and Space weather (OSIRIS). OLite 2 will demonstrate the main functions of the OSIRIS system such as communications, power generation, attitude determination, and command and data handling. The success of the OLite 2 mission, therefore, is imperative for progress in the design of OSIRIS. Equally, a successful communication and power link between OLite 2 and HASP is essential for success of OLite 2.

[#] Faculty Mentor

^x Second Mentor

⁺ Graduate Mentor

OSIRIS

Ionospheric irregularities are variable conditions in the Earth's space environment that are unpredictable and are important because they can affect technologies on which society relies. Such effects include inaccuracies in GPS location, damage to power grids, and damage to satellites already on orbit. The OSIRIS satellite mission aims to better understand these phenomena and research the broader area of space weather. The primary objectives of the OSIRIS mission are to take measurements of the stimulated ionosphere produced by ground-based heaters, to correlate these measurements with ground-based measurements, and investigate spatial and temporal characteristics of the heated ionosphere by measuring plasma properties through several gradually separating probing satellites [1].

OSIRIS Lite 2

OLite 2 is a precursor mission to the OSIRIS mission. The main focus of OLite 2 is to test system and sub-system components in a near-space environment and to demonstrate new technologies critical to the success of the future OSIRIS satellite mission. OLite 2 is composed of three subsystems, which include the satellite Support Box (SB), the CubeSat Simulator (CSS), and a recently added technology demo for the Penn State Lunar Lion (a camera phone). A CAD rendering of the OLite 2 design (without camera phone), developed in the Student Space Programs Laboratory (SSPL) at Penn State, is given in Figure 1.

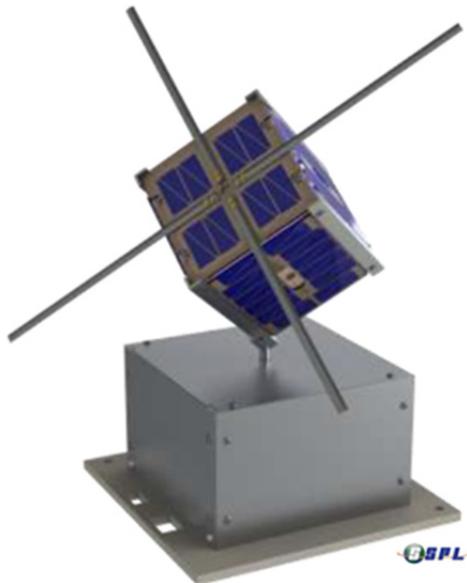


Figure 1: SSPL conceptual design of OLite2 (CubeSat Simulator above and Support Box below)

A CubeSat is a small, cube-shaped satellite with mass less than 1.3 kg that follows the Cal Poly standard for picosatellites [2]. In its “1U” configuration it has dimensions of 10 cm each in length, width, and height. At the top of Figure 1

is the CSS with solar panels and its antenna, while at the bottom is the SB. The SB was designed as part of this project.

High Altitude Student Platform

For the current mission, OLite 2 will utilize a cost effective space test platform known as the High Altitude Student Platform (HASP). HASP uses a zero pressure, small volume polyethylene film balloon to reach an altitude of over 36 km [3]. It allows students to conduct extensive environmental and flight-testing on CubeSats, prototypes, and small experimental payloads. Furthermore, it allows for an inexpensive alternative to testing prototype picosatellites in low earth orbit. HASP offers flight test opportunities each academic year. The 2011 HASP mission will carry OLite 2, along with 11 other experiments, on a high altitude flight for approximately 15 to 20 hours. HASP also provides payloads with standard mechanical, power, and communication interfaces.



Figure 2: Picture of High Altitude Student Platform launch [4]

The communication and power link between the CSS and the HASP system is critical to both OLite 2 and OSIRIS. Due to the incompatibility of the components (caused by the relative higher voltage supplied by HASP) and the need for a versatile system in terms of power and data flow, a decision was made to create an interface converter between OLite 2 and HASP.

System Integration

Known as the CubeSat Support Box, this system link provides several utilities. Employing a microprocessor on a command and data handling board (C&DH board) along with several other components, the SB needs to meet specific requirements to fit the needs of the mission. A systems engineering approach was taken to capture the requirements for both HASP and the CSS. Requirements were developed for the SB, taking into account the HASP and CSS

interfaces, and associated high level functional and physical diagrams were subsequently produced. A system was then designed that met those requirements strongly considering commercial-off-the-shelf (COTS) and heritage components. The system was prepared for integration with the CSS by designing and executing thorough system tests. Finally, the SB was successfully integrated with the CSS this system was taken through environmental testing, including simulated HASP interfacing.

SYSTEM DESIGN

The research project presented here is hardware based and involves several features in the design and implementation of a system. System components from OLite 2 will be tested in a near space environment for future use on OSIRIS. Many sensors were placed on components of the OLite 2 system to monitor their performance in an environment with low pressure and extreme temperature variations. This data will provide useful information on necessary changes or improvements to OLite 2 hardware before use on OSIRIS. For this reason, it is essential that OLite 2 maintains enough power to run the system, the corresponding sensors, and communication with HASP for C&DH purposes. Though the solar cells on OLite 2 were developed to provide the system with enough power, they have not been tested in a near space environment nor is it certain that the battery will have enough charge to power the system at night on the balloon flight (which is considerably longer than the approximately 45 minutes that would occur in orbit). To maintain the necessary power needed, we will take advantage of the power supplied by the HASP platform. Furthermore, HASP provides data downlink for system monitoring purposes. This data communication from OLite 2 to HASP is critical for the success of our mission. Data transfer to and from HASP therefore must be facilitated.

Design Process

1. Develop the requirements for the SB taking into account HASP and CSS interfaces
2. Develop the associated high level functional and physical diagrams
3. Design a system that meets these requirements
4. Prepare the system for integration with the CSS by designing and executing thorough system tests
5. Oversee the successful integration of the SB with the CSS and see the integrated system through environmental testing

Design Process Level 1

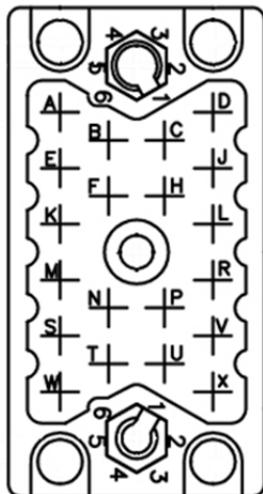
Requirements

1. The SB shall receive power from the HASP power system
2. The SB shall supplement power to the battery onboard OLite 2 as needed
3. The SB shall establish data communication with the HASP
4. The SB shall retrieve data commands uplinked through HASP

5. The SB shall execute a command from HASP if command is for itself
6. The SB shall send a HASP command to the CSS if command is for the CSS
7. The SB shall receive data from the CSS
8. The SB shall send data received from the CSS to HASP for downlink
9. The SB shall provide continuous power to an onboard cell phone

HASP Power

HASP power must be accessed through an EDAC 516 connector. A diagram of the power connection is shown in Figure 3. Wires from the connector are color coded according to their function as listed in Figure 4, which is a table taken from the HASP interface manual [5]. The connector consists of 20 pins with analog downlink channels and discrete HASP commands for shutting off power. Continuous power will be drawn from HASP, so it was not necessary to worry about the discrete HASP commands, nor were the analog downlink channels needed. The HASP power interface will be used merely for constant power to the SB, so only +30 VDC lines and power ground lines were needed



Function	EDAC Pins	Wire Color
+30 VDC	A,B,C,D	White with red stripe
Power Ground	W,T,U,X	White with black stripe
Analog 1	K	Blue
Analog 2	M	Red
Signal Return	L, R	Black
Discrete 1	F	Brown
Discrete 2	N	Green
Discrete 3	H	Red with white stripe
Discrete 4	P	Black with white stripe

Figure 4: Table of HASP power pin layout [5]

Figure 3: EDAC 516-020 receptacle pin layout from HASP platform [5]

HASP Data

The HASP data interface uses a RS-232 serial DB9 connection (Figure 5): signal ground is pin 5, transmitted data is pin 3, and received data is pin 2. Data coming from HASP consists of eight data bits, no parity, one stop bit, and no flow control [5]. Due to the size and weight of OLite 2, it is considered a small payload and therefore will have a serial port speed of 1200 baud. Also noted in the HASP

interface manual, it is necessary to wire the DB9 pigtail as a null modem to OLite 2 because it is considered Data Terminal Equipment.

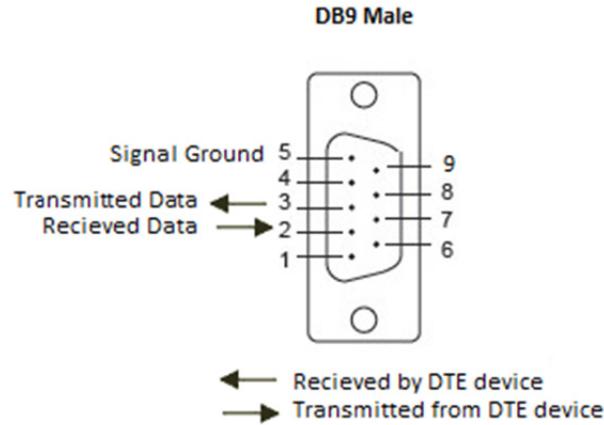


Figure 5: RS-232 serial connector from HASP platform [5]

CubeSat Simulator

Primary power for the CSS is provided by a 4.2-V lithium battery that is charged by solar cells. The solar cells act as a constant current source and are attached to an external interface connector on the bottom of the C&DH board in the CSS. Fortunately for the mission, the same connector also contains pins for receiving and transferring data. This simplifies the design to some extent in that the data communication lines and power output by the SB to the CSS can be sent to the same connector in the CSS. A caution taken into consideration was that the lithium battery on board the CSS must not exceed 4.2 volts.

Physical Parameters

In the beginning of the design process, a large portion of the physical structure of the SB was already available to the SSPL team working on OLite 2. This was a major physical parameter for the SB. Figure 6 shows the dimensions of the SB. The SB system components were required to fit within this $3.75 \times 5.856 \times 5.856$ inch ($9.53 \times 14.87 \times 14.87$ cm) volume. A second level of requirements includes the following:

- The SB shall interface with HASP power through EDAC 516 Connector
- The SB shall communicate with HASP using an RS-232 Serial connection
- A mounting support for system shall provide support for a 10-g vertical and 5-g horizontal shock
- Heaters shall be placed near components that may need heat in a cold temperature environment
- A cell phone shall be mounted to top of SB

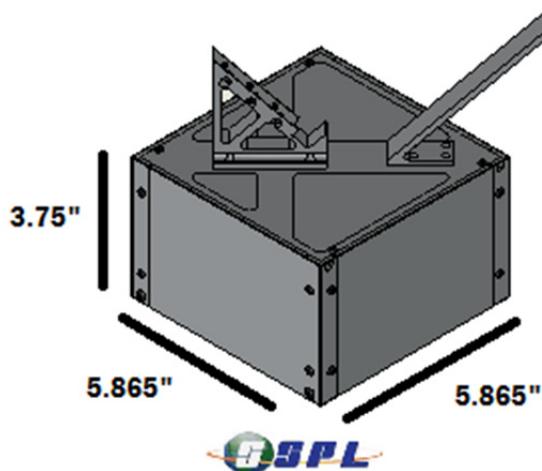


Figure 6: CAD rendering of OLite 2 Support Box

Power and Data Parameters

Although HASP provides power at a nominal 30 V, this voltage can vary from 32 V to 33 V at the beginning of the flight to a voltage of 29 V at the end of the flight. Therefore, the SB system is required to accommodate a voltage range from 29 to 33 V. Also, in the HASP interface manual it is made clear that power from HASP is fused for a maximum current draw of 0.5 A. Furthermore, at startup any small payload that exceeds 1.5 A longer than one second will open the fuse. This time delay gives the SB system enough time for in-rush current. A data parameter considered during the design process was that the HASP serial port speed is 1200 baud, so data being transferred through the SB shall be sent at a correlating speed and data format.

Data Communication

A Linux operating system is used to control data communication. After startup, the software selects special programs to run functions that take measurements, create data packets, accept commands, and transfer data. For testing, a software routine called `minicom` is used to monitor what is going on in the system. It allows the OLite 2 team to monitor errors when they occur. Overall, the OS, software, and programs that run the system are robust in that they provide an easy method for communicating with OLite 2, sending commands, and reprogramming the C&DH board.

Design Process Level 2

After defining requirements, parameters, and studying the CSS and HASP interfaces, we then developed a functional diagram for the SB. Figure 7 shows the function of the SB and its role in the overall system.

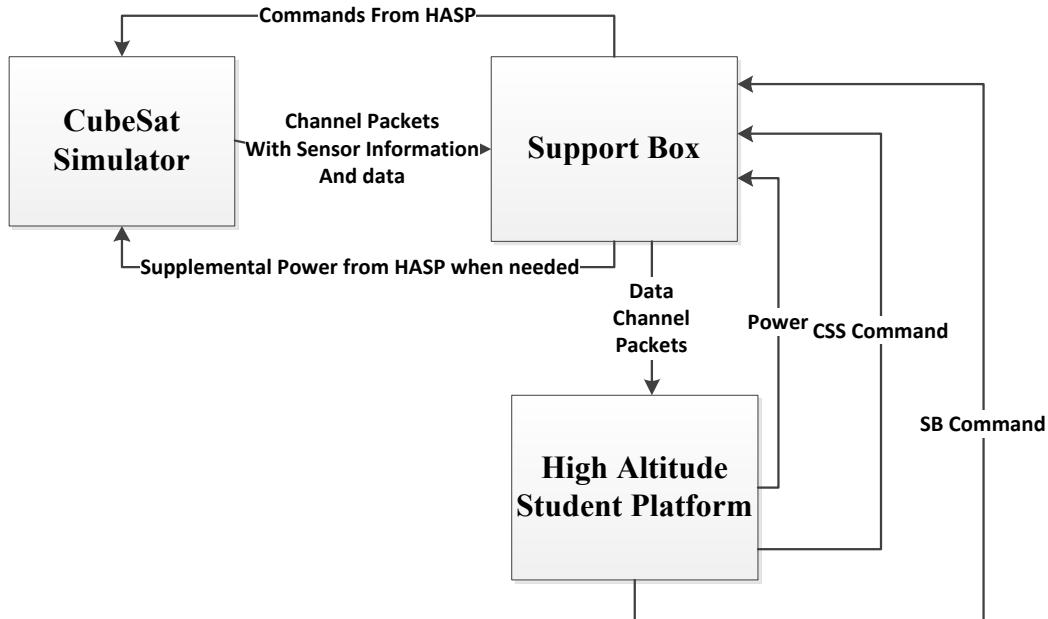


Figure 7: Support Box high level functional diagram

Although at this stage a system diagram would seem appropriate, system components have not been defined so a SB system diagram will be illustrated in level three of the design process.

Design Process Level 3

The thought process and reasoning for choosing each of the following components is provided below.

Battery Charger

The first task to consider was how to charge the battery inside the CSS. Because constant current is required by the battery during charging, it was necessary to design a constant current battery charger. The lithium battery used on OLite 2 could not exceed 4.2 volts. A Linear Technology 1512 constant-current/constant-voltage battery charger was selected. Its accuracy is 1% or better in constant-voltage mode and its max 1.5-A switch current allows for charging currents up to 1 A for a single lithium-ion cell [6]. An LT 1512 evaluation board developed by SSPL was adapted for use in the SB. Several other benefits of using this battery charger include providing the user the capability to change the current delivered to the battery and to change the voltage established after the battery is charged. Once the battery is charged at a constant current up to a predetermined voltage, the battery charger switches to a constant voltage source, which we set not to exceed 4.2 V. However, one drawback of this battery charger is its maximum voltage rating is 30 V, yet HASP could supply up to 33 V. To address this issue a voltage regulator was added.

Voltage Regulators

A National Semiconductor LM5576 voltage regulator was selected as it can accept a wide range of voltages (from 6–75 V). Several voltage regulators are required: one 5-V regulator for the battery charger, one 3.3-V regulator for the C&DH board, and another 5-V regulator for another input to the C&DH board. The benefit to using several voltage regulators is that each has a shutdown pin, which provides the ability to shut down the battery charger on command by sending a low output from the C&DH board into the corresponding 5-V regulator, which provides a second safeguard against exceeding the 4.2-V limit on the battery. These voltage regulators give the SB a flexible power system as they cover the range of voltages supplied by HASP.

Command and Data Handling Board

For the data processing in the SB, a decision was made to use a copy of CSS C&DH board. This allowed the use of the same software routines for data processing in the CSS. Furthermore, as shown in Figure 8, the bottom of the C&DH board contains an external interface connector that provides an easy connection to incoming commands, outgoing received data, and supplemental power and ground to the battery charger. On the opposite side of the C&DH board is an attached adaptor board that connects to its input and output pins, which is where the 3.3-V line and the 5-V power lines were connected.

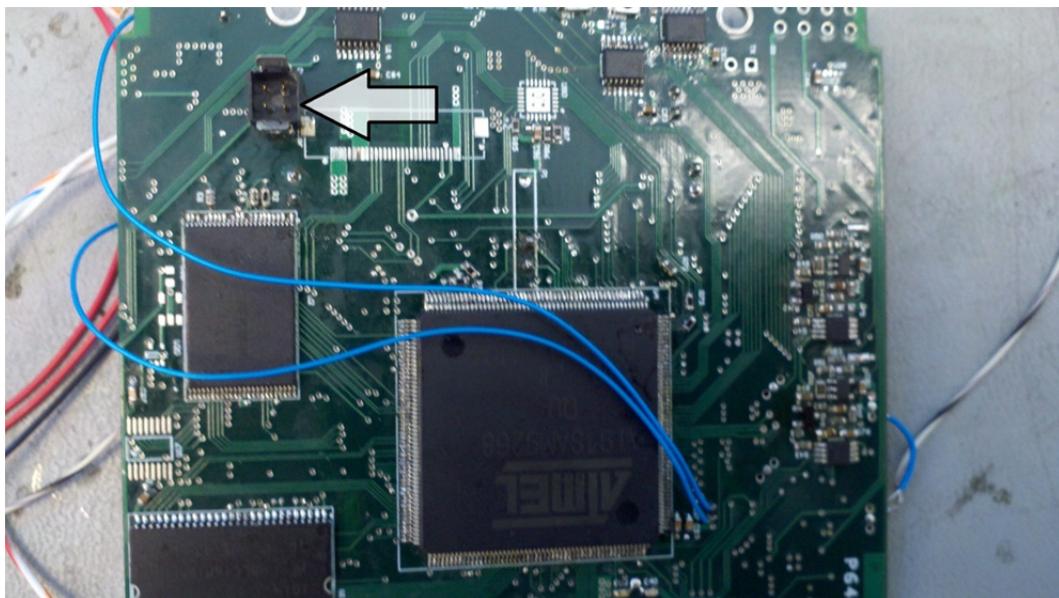


Figure 8: C&DH board for Support Box with external interface connector marked by arrow

Power Distribution Block

A power distribution block was added to simplify wiring inside the SB. As it was readily available and simple to mount inside the SB, an eight-gang power

terminal block was chosen. The three gangs on the far left of this terminal block were used for power and three on the right were used for ground. The two in between were left alone to provide some spacing between power and ground. This mount gave the benefit of organizing the wires, which later helped in making wire connections from component to component.

RS-232-to-UART Converter

For data parsing from HASP, it was necessary to first convert the information from a serial to a parallel form. A simple RS-232-to-UART conversion board was selected. It consists of receiving, transmitting, ground, and 3.3-V pins.

Resistor Heaters

Due to the temperature variations the system will undergo, operating temperatures of each component must be considered. Though all components in the SB have an operating range of temperatures that can sustain a near-space environment, several resistor heaters were added as a precaution.

DB9 Connectors

The physical wiring from the SB to the CSS was required to be sturdy during connection and disconnection, yet simple. The solution was a wire bundle with male and female DB9 connectors. This provides a method to easily change wires if needed and a sturdy support due to the screws in the male to female connectors. This also provided flexibility to easily isolate components.

System Diagram

With the components defined, a physical system diagram (see Figure 9) was created for a better understanding of the SB. This helped in maintaining a clear idea of each component's responsibility in the system as a whole.

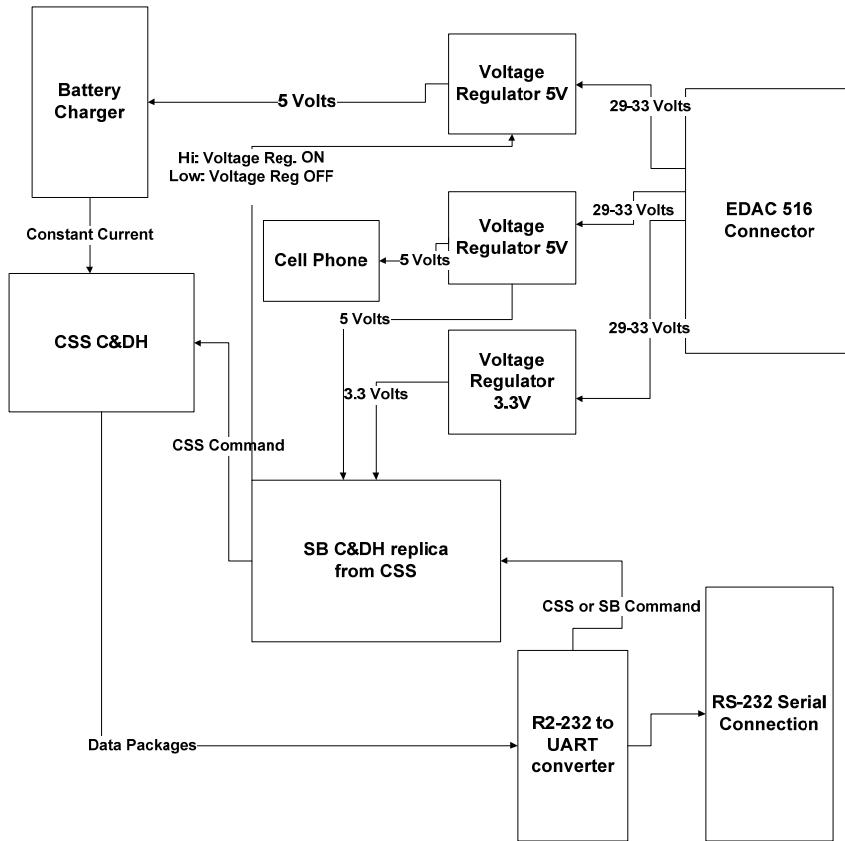


Figure 9: Support Box system diagram

SYSTEM DEVELOPMENT

An explanation of how the SB was physically put together and any process involved in modifying any of the components is provided below.

Adjusting Voltage on Voltage Regulator

Two 5-V regulator outputs and one 3.3-V regulator output are required. Because 5 V was the manufacturer setting, only one of the voltage regulators had to be configured. From the design equations given in the LM5576 Evaluation Board data sheet [7], resistors 5 and 6 were replaced with $3.01\text{ k}\Omega$ and $1.780\text{ k}\Omega$, respectively.

Adjusting Current and Voltage Output of Battery Charger

Figure 10 shows the schematic of the battery charger circuit board. From the LT 1512 data-sheet [6], it is given that the maximum charging current at 5 V into the lithium battery is a 0.6 A. From the design equations specified by the LT 1512 data-sheet [6], it is calculated that resistor 3 must be replaced with approximately $0.16\text{ }\Omega$. Only $0.2\text{ }\Omega$ was available so the constant output current was calculated as 0.5 A. This value was close enough to the desired current input to the battery while maintaining a safe current value for max rating.

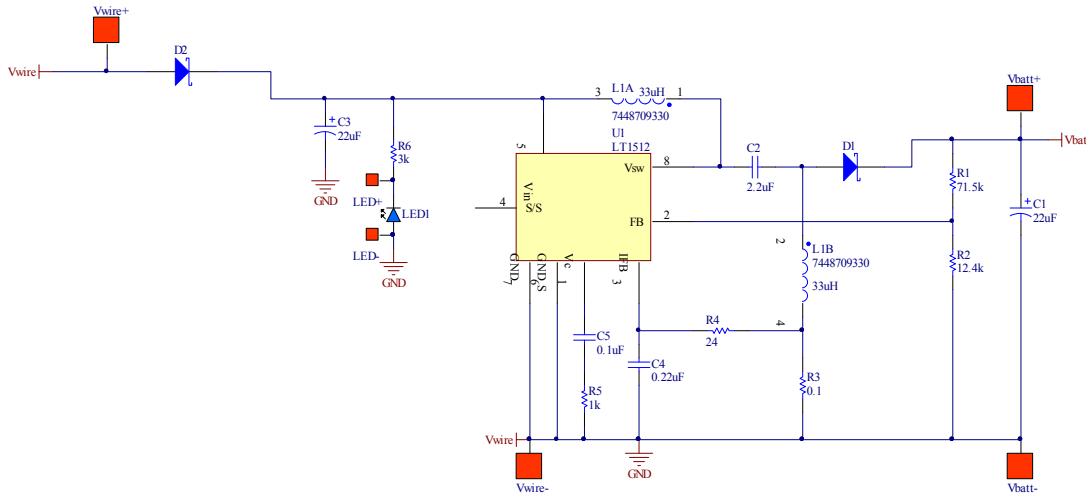


Figure 10: LT1512 battery charger schematic

Next, the constant voltage output was also altered to 4.2 V. Again, using the design equations from LT 1512 data-sheet [6], resistors 1 and 2 were changed to be $61.9\text{ k}\Omega$ and $27\text{ k}\Omega$, respectively, giving an constant output of 4.18 V. This was sufficiently close to the desired voltage without causing a safety issue of going over the maximum value of 4.2 V.

Circuitry Configuration

The circuitry inside the SB was the next priority. In developing the wiring scheme, connection diagrams were created: one of the SB and the other of the overall system. Figures 11 and 12 show both the overall link between HASP and the CSS and of the circuitry inside the SB.

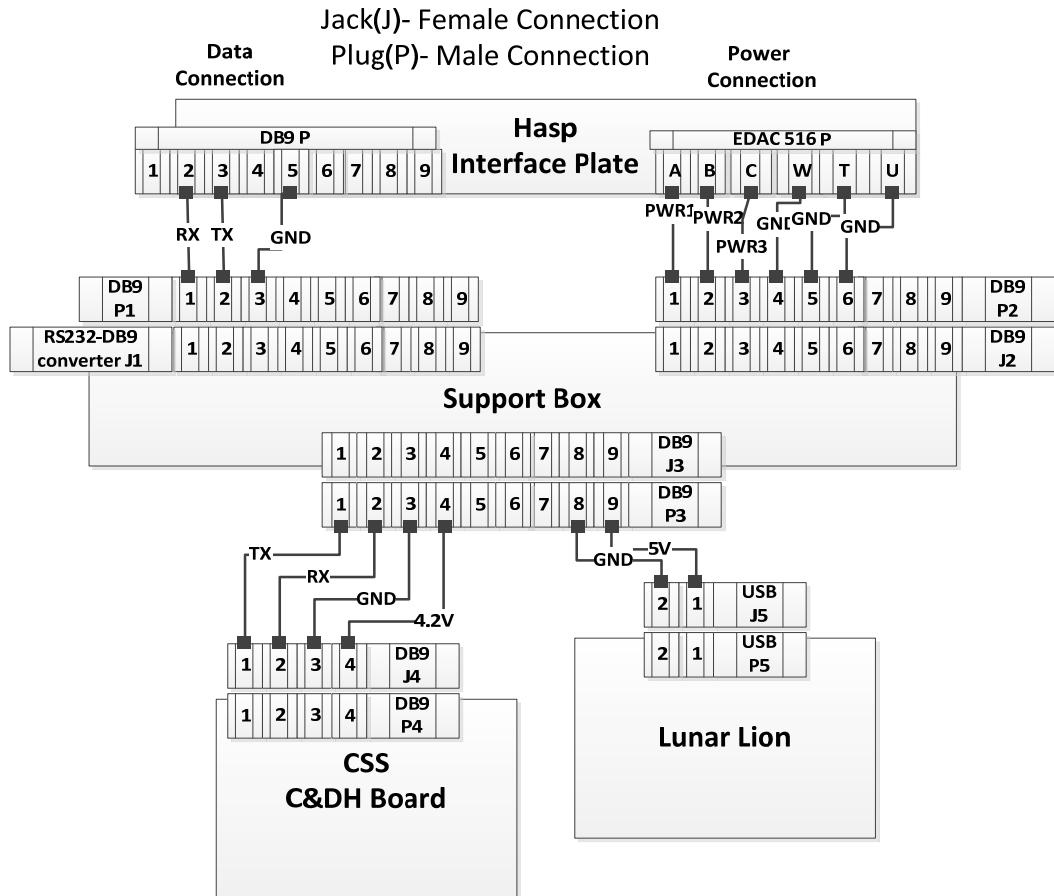


Figure 11: HASP-SB-CSS system connection diagram

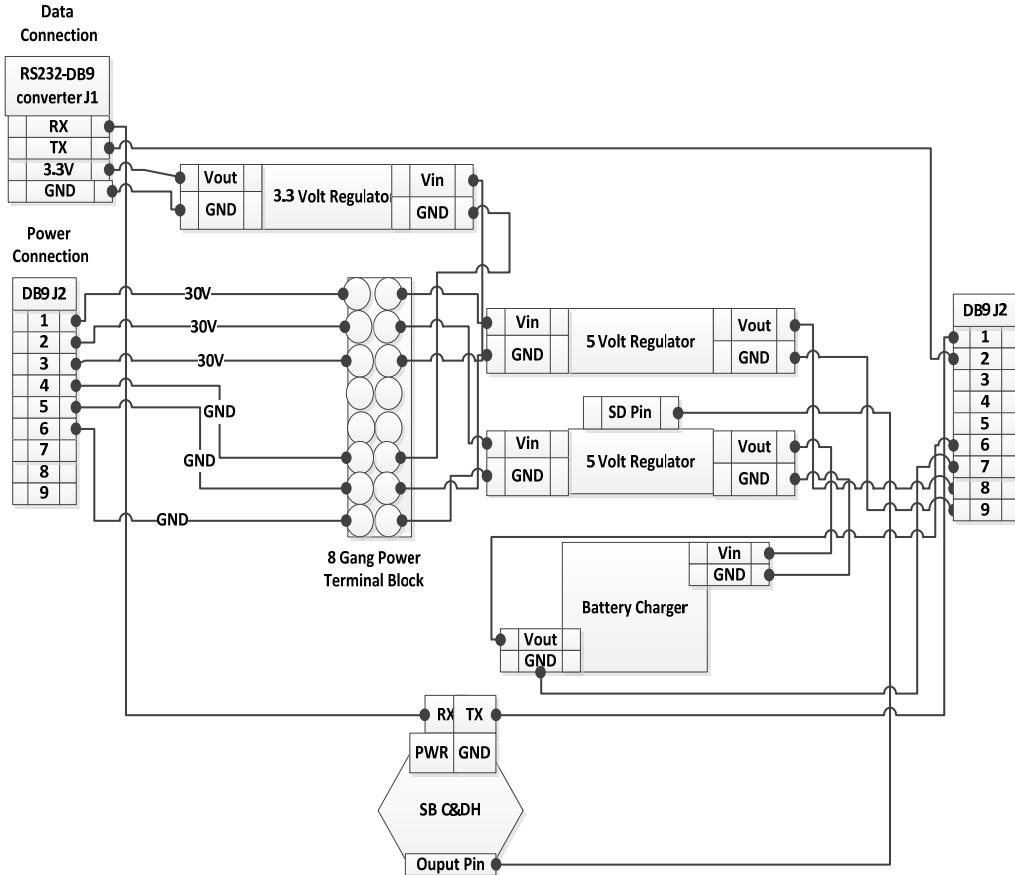


Figure 12: Support Box connection diagram

Physical Structure Development placement

Next came was the task of situating components inside the SB. The dimensions of each component were measured, and a physical configuration was selected. The voltage regulators were stacked on top of each other. In placing the other components, three precautions were taken into consideration. First, the voltage regulators could not be too close to each other because of the possibility of two voltage pins coming into contact. Second, the C&DH board must be as far as possible from the power distribution block because of the sensitivity of its circuitry. Third, components must be secured well in the SB.

With the help of several aerospace and mechanical engineering students on the OLite 2 team, a 3D representation of the SB was created. As you can see in Figure 13, all three precautions were addressed in the design. All components were secured well by the bottom aluminum plate inside of the SB. The C&DH board was elevated relatively far from the power distribution block. The voltage regulators were separated by spacers, which along with several bolts provided a secure support for all components in the SB.

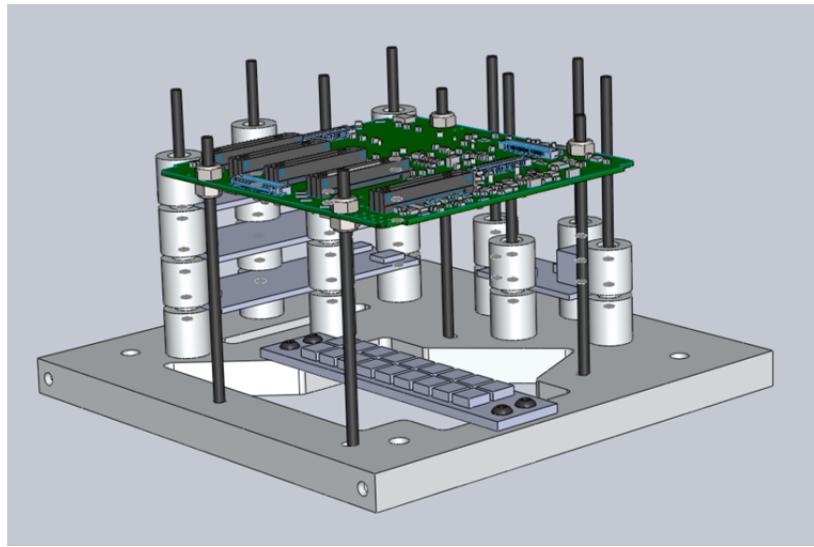


Figure 13: 3D representation of configuration inside Support Box

The manufacturing process of the entire SB could then be completed. The structure of the SB was designed through SolidWorks and then manufactured out of aluminum at the Penn State Learning Factory. An exploded view of the structure of the SB is shown here in Figure 14. After production, holes were punched through the side panels to fit all DB9 connections.

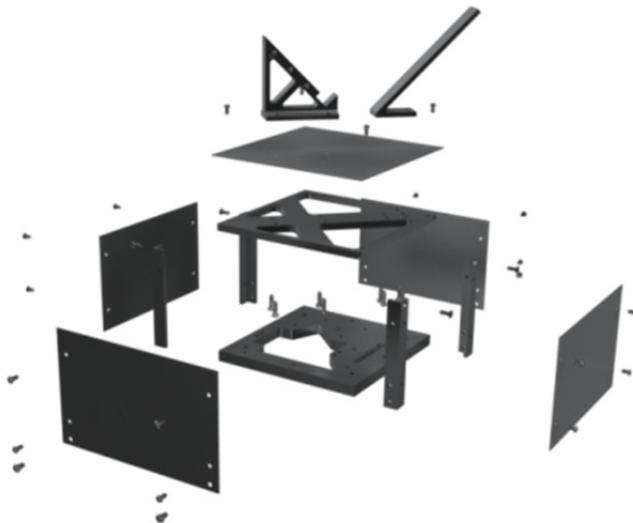


Figure 14: CAD Rendering of Support Box Physical Structure

Design Process Level 4

System Testing

Before the SB could be put together, system testing played a crucial part in assuring a safe integration with the CSS and HASP. Each component had to be

tested individually first, each part that was connected to another to be tested in conjunction, until finally the entire system could be tested together. Two aspects of the SB had to be tested: power and data.

Power

First, it was necessary to check each individual voltage regulator as they make up a big part of the power system. Two of the voltage regulators were still at manufacturer setting, so they required a 5-V output verification. Both were tested with and without a load at a wide range of input voltages up to 33 V. In every case a constant output of 5 V was measured with a multimeter. The same process was done with the 3.3-V regulator and again the correct output was measured.

The battery charger was tested next. A 5-V input was given to the battery charger to simulate a 5-V regulator. A simple power resistor was then connected to the battery charger and a constant current of approximately 0.5 A was measured. The battery charger was then connected to the battery and the voltage was monitored again with a multimeter. A constant current was delivered to the battery until it reached 4.18 V, at which point the current dropped to about 3% capacity. At that point the constant voltage source took effect and a voltage of around 4.2 V was measured.

The voltage regulator was then connected to the input of the battery replacing the power supply. The 5-V regulator gave a correct output. The battery charger provided a constant 0.5 A to the battery. Once again, when the voltage of the lithium battery reached 4.18 V the charger converted to a constant voltage source as around 4.2 V.

The same setup was then tested with the introduction of the shutdown pin on the voltage regulator. A wire was connected from a power supply to the shutdown pin. Output of the power supply was set to about 0.5 V, which acted as a low input value, signaling the voltage regulator to shut off. The voltage of the power supply was gradually increased until it reached a critical value of 0.7 V, at which point the voltage regulator went into a standby mode. The voltage was again increased to a second critical value of 1.225 V rendering the voltage regulator fully operational. This test was performed several times, and the outputs of each component were monitored to make sure that turning on and off the voltage regulator through the shutdown pin did not negatively affect any of the other components.

Finally, the last 5-V regulator was connected to the camera phone. The phone was turned on and a visual verification was established that the voltage regulator was indeed charging the phone. The 5-V regulator was monitored and the test continued until the phone's battery was about half charged.

With the components each tested individually, they could then be tested together with the C&DH board. The system was connected according to the connection diagrams. For the first test, 25 V was supplied to the voltage regulators, with the current limited to 0.25 A. Though the maximum draw allowed from HASP for OLite 2 is 0.5 A, the system should not exceed about 0.16

A as that was the measured current draw from both the CSS and the components in the SB. If the current was to limit at 0.25 A the system could be shut off the quickly and hardware problems could be debugged. After examination of the voltages and currents it was established that everything was working properly.

Data Communication

With the circuitry configured, data communication could then be tested. The programs were flashed to the C&DH board and the Linux system was booted. Using a computer to simulate HASP commands, software called `minicom` was used to monitor data flow between the CSS, the SB, and HASP simulator. The first test was to verify if a heartbeat was received from the CSS. The rest of the data communication tests carried out included resetting the CSS, resetting the SB, and turning off the power to the SB. After several sessions of debugging, everything tested worked properly.

ENVIRONMENTAL TESTING

Design Process Level 5

The final step was to test the integrated system in an environment similar to that which will be experienced on HASP. This was done by placing the CSS connected to a computer, simulating HASP, and the SB inside of a thermal-vacuum chamber (Figure 15). The integrated system was tested at temperatures varying from -10°C to $+40^{\circ}\text{C}$ and pressures from 722 Torr to 1 Torr. Two environmental tests were carried out. In the first test, the system was fully operational for 6.5 hours. After that time channel data from the CSS ceased. The SB had reset to a non-support box mode and the SB would not accept a command to return to SB mode. After further hardware and software testing, various SB reset problems were fixed. During the second environmental test the system ran through an abridged version the timeline onboard HASP. Reaching a desired 10 hours, the SB and CSS operated continuously throughout the test. Through command, the SB supplemented power to the CSS for approximately two to three hours of the test. Furthermore, data packets were transferred from sensors in the CSS through the SB to the simulated HASP platform. A third environmental test will be done in the near future for further system and sensor performance testing purposes.

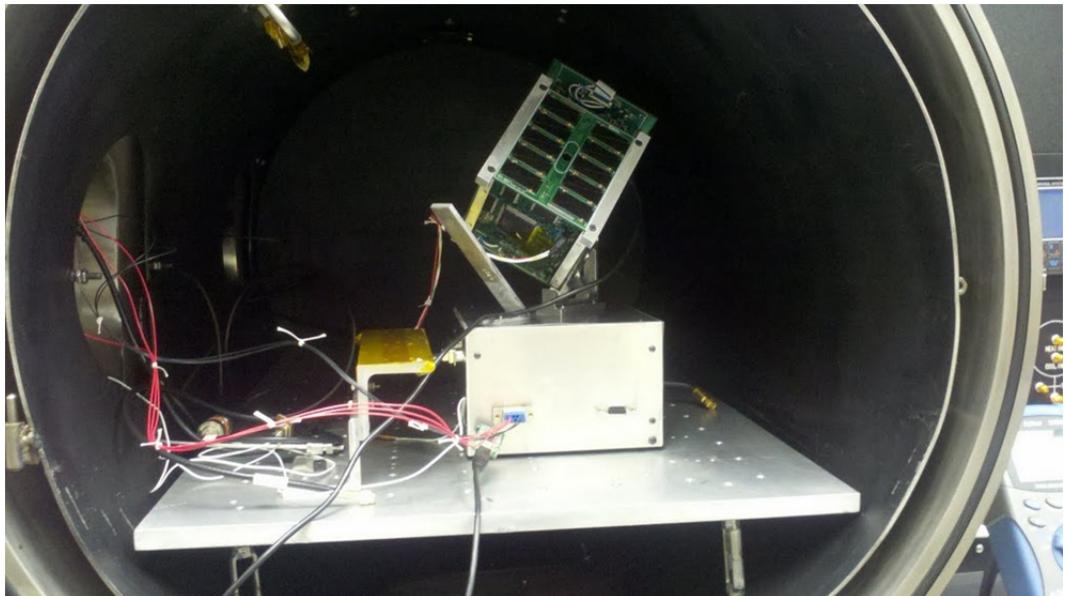


Figure 15: OLite 2 in vacuum chamber for environmental testing

Failure Prevention

Because the SB plays such a vital role in the OLite 2 mission, it is very important to provide a safety net in case something happens to the SB. In the event that any of the components fail before flight, back-up components have been produced. This includes populating a new battery charger circuit board, ordering and configuring new voltage regulators, and making sure that there is a spare of each part in the SB. Because of the simple design of the SB, it is easy to switch out any component making this a proper safeguard.

DISCUSSION

The SB established a power and data link between HASP and the CSS. The structure of the SB and the circuitry within it was such that the SB was able to meet the requirements established at the outset of this project. Through high level functional and system diagrams, connection diagrams, and 3D models of the system, a design of the SB was developed. Through system and environmental testing, the safety of the CSS and the C&DH board in the SB was ensured.

CONCLUSION

Supporting data received from system and environmental testing of the integrated system shows that this project was successful. All nine of the requirements regarding establishing the link for both power and data from HASP to the CSS were accomplished. By employing a systems engineering approach, a system design was taken through design and test phases. By understanding HASP and CSS interfaces, an interface was created that fits the mission requirements for OLite 2.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-0755081. C. Galvan would like to thank his mentor Prof. Sven Bilén, his secondary mentor Bob Capuro, and his graduate mentor Allen Kummer for their extremely helpful contributions over the course of the research project. C. Galvan would also like to thank all of the students at the Student Space Programs Laboratory working on OLite 2 who also had helpful contributions over the course of the research project.

REFERENCES

- [1] "OSIRIS Proposal." The Pennsylvania State University, 2009.
- [2] CubeSat Community. California Polytechnic State University. Internet: http://cubesat.atl.calpoly.edu/images/developers/cds_rev12.pdf. October 2 2008 [12 June 2011].
- [3] LSU Space Sciences Group. "High Altitude Student Platform" Internet: <http://laspace.lsu.edu/hasp/> [June 4,2011]
- [4] Guzik, T., and J. P. Wefel. "The HASP 2007 Flight and Call for Payloads 2008." Presented at Southeast Regional Space Grant, Shreveport, Louisiana, 2007.
- [5] "HASP – Student Payload Interface Manual." Internet:http://laspace.lsu.edu/hasp/documents/public/HASP_Interface_Manual_v21709.pdf Feb 19, 2009[June 4, 2011].
- [6] Linear Technology. "LT1512 SEPIC Constant-Current/Constant-Voltage Battery Charger."
- [7] National Semiconductor. "LM7576 SIMPLE SWITCHER® 75V, 3A Step-Down Switching Regulator." June 14, 2010.

DEVELOPMENT OF A CHARGING METHOD FOR GEOFEBBLES USING WIRELESS POWER TRANSMISSION

Luis E. Olique Ortiz,* Sven G. Bilén,[#] Robert Capuro,[†] Michael Conway,⁺ and Tyler Boehmer⁺

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

* Undergraduate Student of
Department of Electrical & Computer Engineering
University of Puerto Rico at Mayagüez
Mayagüez, PR 00681

ABSTRACT

Wireless power transfer uses magnetic resonance to transfer energy via a non-radiating field. Wireless energy transfer is not a new concept: Nikola Tesla in 1891 made a coupled tuned oscillator and Tesla coil to transfer energy wirelessly. Today, this technology finds diverse applications in devices such as cell phones, laptops, and also TVs. A simple wireless power battery charging sub-system is proposed in this paper for the GeoPebble in order to minimize wired interfaces between battery charger and power supply. Power transfer was demonstrated; however, efficiency still must be maximized. Several circuit topologies are presented in this paper with experimental results to show which has the best efficiency for this application.

INTRODUCTION

Ice sheet flow is a complex geophysical process that is the subject of active theoretical, computational modeling, and field data gathering/analysis work. Recent successes in satellite remote sensing measurements have revolutionized the study of the surface of ice sheets. What is still poorly known is the internal structure of ice sheets and properties at the base of the ice. One of the most important tools for determining these properties is seismic data gathering and analysis. These seismic experiments have been conducted using widely-spaced grids of geophysical-profiling electronic systems. Deployment, operation, and

[#] Faculty Mentor

[†] Second Mentor

⁺ Graduate Mentor

retrieval of these systems in harsh environments are problematic. Legacy systems are deployed in arrays consisting of 100 or devices interconnected via complex cabling and connectors. A device known as GeoPebble is being developed to enable geoscientists to more easily gather data on ice sheet and glacier morphology. The simplification comes through eliminating inter-device cabling and minimizing wired interfaces between the GeoPebble and its support system. One GeoPebble subsystem that is desired to have a wireless interface is the battery charging subsystem. To meet this requirement we designed and implemented a wireless energy transfer subsystem based on magnetic resonance.

The basic principle employs two coils, separated by an air gap, that resonate at the same frequency to provide high efficiency energy transfer (Figure 1). The flow of an alternating current through the primary circuit creates a magnetic field that links the primary and secondary coils. This induces a voltage on the secondary coil that provides wireless energy to the GeoPebble battery charging circuit.

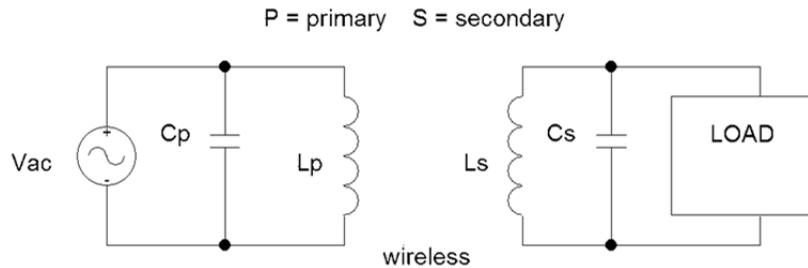


Figure 1: Simple diagram of wireless power transmission

A diagram of a simple wireless power transmission system is shown in Figure 1, in which L_p and L_s are the primary coil and the secondary coil inductances, respectively. Capacitors C_p and C_s are selected to obtain the desired resonant frequency for the coupling.

Magnetic Resonance Coupling

Physical objects and circuits are capable of oscillating or resonating with high amplitude at a particular frequency known as its natural frequency. A vintage example of exciting a natural resonance using sound waves is a singer that shatters a crystal glass using his/her voice. When the singer hits the note, or frequency, at which the glass has its natural resonant frequency, the oscillation amplitude in the glass builds up and the glass will shatter. Similarly, it is possible to achieve resonance through an electromagnetic field.

The electromagnetic field is an energy-transfer medium where we can use resonance in an electronic circuit, with sufficient amplitude to attain energy transfer. In order to provide magnetic resonance coupling an LC circuit is used.

Basic Magnetic Resonance Theory

As mentioned above, a wireless power transfer circuit utilizes a primary and a secondary coil. An alternating current through the primary coil produces a magnetic field according to Ampere's Law, which states that each current has a magnetic field associated with it. Using Lenz's law the induced voltage can be found as shown in the Equation 1.

$$V = -N \frac{d\varphi}{dt}, \quad (1)$$

where V is the induced voltage in the secondary coil, N is the number of the turns in the secondary coil, and φ is the magnetic flux that links the secondary coil with the primary coil. Then, energy is used by the load after being converted from an alternating current to a direct current.

Resonance Coupling Using a Transformer Model

Wireless power transmission can be analyzed using a transformer model. The first consideration is mutual inductance, which determines the voltage that will be induced by the current passing through the other circuit. Mutual inductance, M , is given by

$$M = k \sqrt{L_p L_s}, \quad (2)$$

where k is the coupling coefficient. Using the circuit in Figure 1 and knowing the impedance of the secondary circuit, one can determine the reflected impedance. Ignoring the self-resistance of the coils, the reflected impedance is given by [1]

$$Z_r = \frac{\omega^2 M^2}{Z_s}, \quad (3)$$

where ω is the angular frequency of the system and Z_s is the impedance of the secondary circuit using the transformer equation. The value for Z_s depends on the circuit topology (series or parallel) and is given by [1]

$$Z_s = j\omega L_s + \frac{1}{j\omega C_s} + R_L \quad (\text{series}) \text{ and} \quad (4)$$

$$Z_s = j\omega L_s + \frac{1}{j\omega C_s + \frac{1}{R_L}} \quad (\text{parallel}), \quad (5)$$

where R_L represents the resistive load of the system, which in the case of the GeoPebble application is the battery charger circuit.

Circuit Efficiency

Power loss in wireless power transfer systems is strongly dependent on topology and the quality factor of the circuit. There are four possible circuit topologies for tuning the primary and secondary coils, series-series, series-parallel, parallel-series, and parallel-parallel. The definition of these topologies is

determined by the capacitor and the coil connections as shown in Figure 2. The circuit efficiency of each circuit topology was determined by conducting characterization tests of each circuit topology.

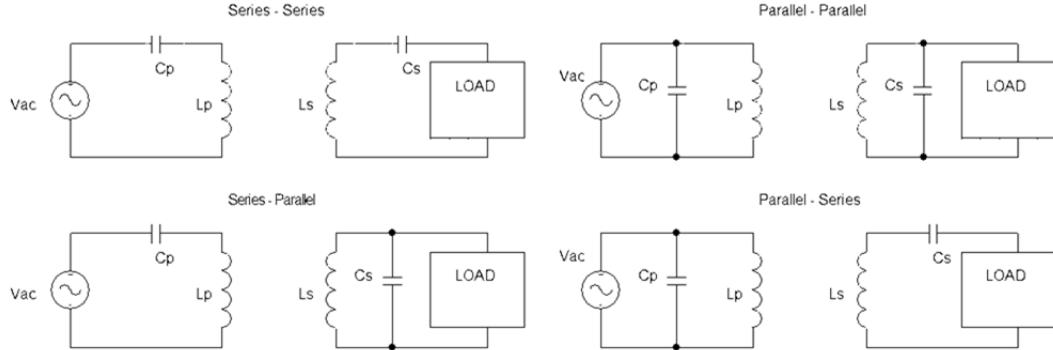


Figure 2: Different circuit topologies for resonance tuning

The quality factor, Q , depends on coil material, size, shape, and operating frequency. Although one might suspect that high Q factors would be better for power transfer between the coils, this is not always true in the case of the wireless power transmission. With a high Q factor we are limited to operating very close to one frequency, since any small variation will be critical for the amplitude. However, we can achieve high amplitude constructing the coils with good quality factor, without risk of losing a very large amount of power. The general definition of the quality factor is based in the operating frequency, the inductance of the coils and the DC resistance of the coil. From this, the quality factor is:

$$Q = \frac{\omega L}{R}, \quad (6)$$

where ω is the angular frequency, L is the inductance of the coil, and R is the DC resistance of the coil.

PROPOSED STRUCTURE FOR INDUCTIVE CHARGING

Shown in Figure 3 is the structure proposed for wireless power transfer to the GeoPebble. The power supply is an AC source that energizes a primary coil located outside of the GeoPebble. Then, the energy will couple wirelessly through an air gap and the case of the GeoPebble to the secondary circuit. The secondary circuit will provide energy to the charge circuit that will condition the power for charging the battery.

Due to all the systems in the GeoPebble case, we have limited space to set up the inductive charging circuit. We are using a Pelican 1200 case for this. There is no space limitation for the

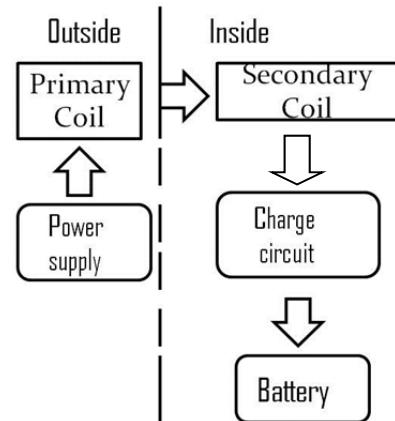


Figure 3: Proposed structure

primary coil since it is not inside the GeoPebble. The secondary coil compartment in the GeoPebble was measured to provide a range of possible sizes for the secondary coil. The diameter cannot be more than 120 mm and the height cannot exceed 34 mm.

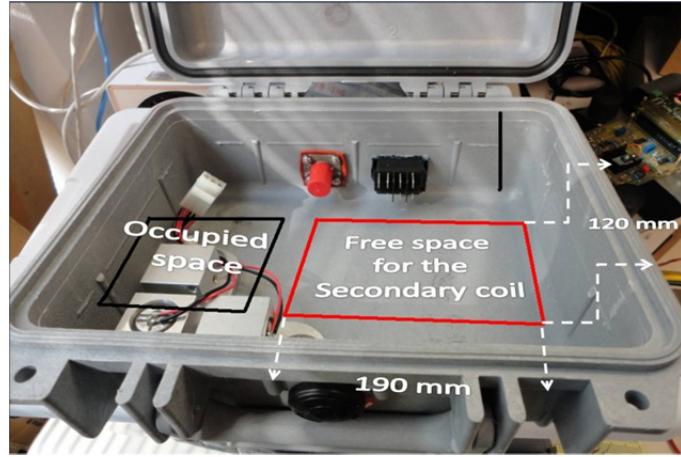


Figure 4: Open top view and free space available in a GeoPebble

The wire used to construct the two coils has a 2.1-mm diameter and each coil is 118-mm in diameter. The circular shape was chosen because the equations that describe this type of coil are well known. The inductance calculation for the coils proposed for this research is given by

$$L = N^2 r \mu_0 \left[\ln \frac{8r}{a} - 1.75 \right], \quad (7)$$

where N is number of coils turn, μ_0 is the permeability of a vacuum, r is the radius of the coil, and a is the wire radius. The circuit topology proposed consists of two circuits with capacitance in parallel to the coil.

To generate the alternating current in the primary circuit we used an AC power supply outputting a square wave with 50% duty cycle and frequency of 125 kHz. The circuit construction consisted of selecting a correct capacitance and inductance such that the circuit had a natural frequency of 125 kHz. With the inductance measured, the capacitance was calculated via

$$C = \frac{1}{L(2\pi f)^2}. \quad (8)$$

In order to find the circuit model yielding the maximum transfer efficiency, we used Equation 8 to determine the tuning capacitor and ran two sets of experiments to find the best circuit model. The first test determined how much voltage the coupled circuit provides to the load as a function of distance between coils. The second test calculated power in the primary circuit by measuring its current and the voltage. Then, the voltage (and hence current) at the load were

measured to calculate power received. These measurements were made with 1-cm gap distance between coils. This information was used to calculate efficiency, which is the ratio of power in the secondary circuit to that in the primary circuit, i.e.,

$$\text{Efficiency} = \frac{P_s}{P_p} \times 100\% . \quad (10)$$

Charging Method

The induced AC voltage in the secondary must be converted to a DC voltage to enable charging of the GeoPebble battery. To perform this function a full bridge rectifier was proposed. This circuit is built with the use of four diodes. This takes the received power from the secondary coil and turns it into a regulated current or voltage used to charge the GeoPebble battery.

EXPERIMENTAL RESULTS

The constructed coils, shown in Figure 5, have an inductance of $5.6 \mu\text{H}$. Using Equation 8, the value of the capacitance to resonate at 125 kHz is $0.3\mu\text{F}$. As described above, we wanted to find which circuit model had the highest efficiency. In order to find this, a resistance of 200Ω was used to represent the load. The first test shows how these circuits are affected by the distance (Figure 6). The second test determined the efficiency of circuits.

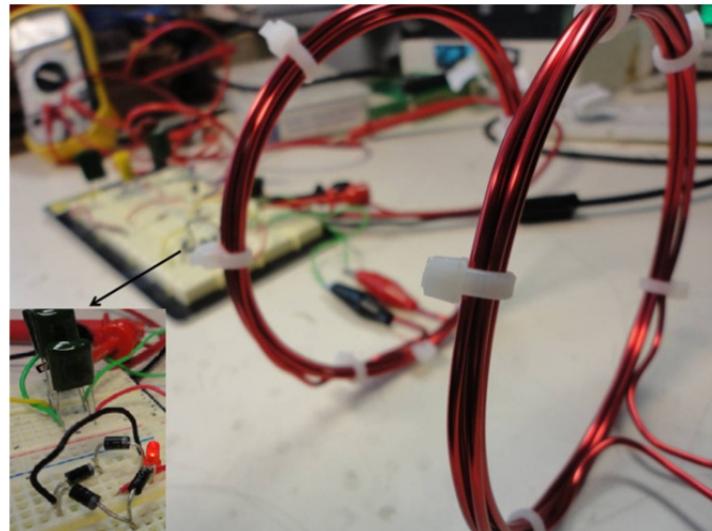


Figure 5: Constructed coils

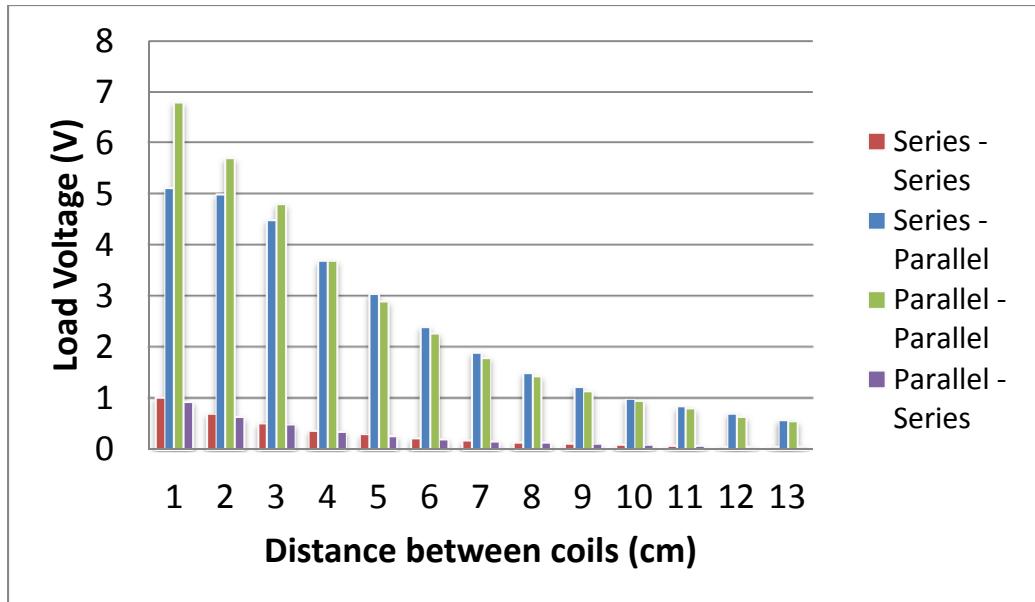


Figure 6: Test of distance for the circuit models

DISCUSSION

As shown in Figure 6, the parallel-parallel circuit has maximum voltage at the load when the distance between the coils is 1 cm. However, as the distance between the coils increases the series-parallel circuit produces higher voltage at the load. At 4 cm, both circuit models have the same voltage at the load. Because of this, the efficiency of these two circuits was measured to determine which is better.

The results of this test are interesting. Although parallel-parallel circuit has a higher voltage at the load than series-parallel circuit, the efficiency test demonstrates that series-parallel has a better efficiency. The series-parallel circuit had an efficiency of 41% versus 20% for the parallel-parallel circuit. The ceramic capacitors used in this study were dissipating heat during the both tests indicating power loss in these components. For future work capacitors that can operate at a power level of 10 watts are needed for this application and to eventually use a smaller case as shown in Figure 7.



Figure 7: A smaller GeoPebble case

CONCLUSION

Based on the characterization tests, a series-parallel connection is better for this application. That is to say the primary circuit should be series connection between the capacitor and the coil and the secondary circuit should be a parallel connection between the capacitor and the coil.

Although battery charging was not attempted in these two tests, wireless power transfer was successfully demonstrated. Also, these tests demonstrate that the material of which the case is made does not affect the magnetic field of the wireless power transfer. Something interesting noticed during this research is a square wave with 50% of duty cycle is better than a sine wave for this technology.

With further research, it may be possible to improve the efficiency and be able to transfer 10 watts that are needed for the GeoPebble battery charging. With more research this kind of technology may be applied to more applications. One interesting application is in medical devices. Because it is a non-radiating technology, it can go through the skin without affecting it. An example of such an application is a pacemaker charged by using wireless power transmission.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, L. Oblique Ortiz would like to thank Prof. Sven Bilén, Prof. Ken Jenkins, Prof. Jeff Schiano, Mr. Robert Capuro, Mr. Mike Conway, Mr. Tyler Boehmer, Mr. Thomas Tyson, and Mr. Allen Kummer for their help and guidance during this research experience. I would like to give special thanks to Lena Getman who made this great opportunity possible to me.

REFERENCES

- [1] Conway, Michael, and Tyler Boehmer, "Inductively-Powered Battery System for Harris Corporation," The Pennsylvania State University. (2011), pp. 8–14.

- [2] Fotopoulos, Kyriaki and Brian W. Flynn, "Wireless Power Transfer in Loosely Coupled Links: Coil Misalignment Model," *IEEE Transactions on Magnetics*, Vol. 47, No. 2, February 2011. <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05639082>>.
- [3] Zhu, Chumbo, Kai Liu, Chunlai Yu, Rui Ma, Hexian Cheng, "Simulation and Experimental Analysis on Wireless Energy Transfer Based on Magnetic Resonances," *IEEE Vehicle Power and Propulsion Conference*, September 3–5, 2008, Harbin, China. <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4677400>>.
- [4] Cheon, Sanghoon, Yong-Hae Kim, Seung-Youl Kang, Myung Lae Lee, Jong-Moo Lee, and Taehyong Zyung, "Circuit-Model-Based Analysis of a Wireless Energy Transfer System via Coupled Magnetic Resonances," *IEEE Transactions on Industrial Electronics*, Vol. 58, No. 7, July 2011. <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5560805>>

ROBUSTNESS AND FAULT TOLERANT CAPABILITIES OF TRANSFORM DOMAIN FIR FILTERS WORKING ON REAL SIGNALS

Danielle Sova,* W. Kenneth Jenkins,[#] A. David Salvia[#]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
George Mason University
Fairfax, VA 22030

ABSTRACT

Transform domain Fault Tolerant Adaptive Filters (FTAFs) rely on inherent learning capabilities of the adaptive process to compensate for transient (soft) or permanent (hard) errors in the hardware implementation. These transient and permanent errors are of increasing concern in the downscaling sizes of nanotechnology. In this project, the robustness and fault tolerance capabilities have been investigated in transform domain FTAFs that improve the quality of corrupted speech signals. Two different transform domain adaptive FIR filters working on noise corrupted speech signals have been used in this study, one based on the zero padded FFT and the other based on a Modified Discrete Fourier Transform (MDFT).

INTRODUCTION

There is often a need to remove undesirable components of a signal without knowing what exactly should be filtered out in terms of frequencies. Adaptive digital filters are commonly used in this situation. The tap weights (filter coefficients) of adaptive digital filters are adjusted to minimize the error between the desired or predicted signal and the actual filter output. Although the adaptive filters with no fault tolerance that minimize the complex error suffice in having the capacity to eliminate unwanted portions of both real and complex signals, they cannot inherently overcome the challenges of hard and soft errors that are becoming increasingly more abundant with the downsizing of integrated circuits by way of nanotechnology. Due to either soft or hard errors in hardware, some errors would be fatal to the function of the filter, possibly corrupting the signal

[#] Faculty Mentor

entirely. Thus, adaptive fault tolerance that couples redundancy with the learning capabilities of an adaptive filter to overcome hardware errors becomes more desirable [3]. In situations where only real signals (such as speech) are being processed, the number of computations in the FTAFs can be significantly reduced, and different forms of adaptive filtering can be introduced.

When filtering real signals, the imaginary part of the signal is equal to zero. Instead of using tap weights to drive the imaginary error to zero, these tap weights are used to minimize the real error. With more tap weights than necessary to properly filter the signal, the conjugate pairs of tap weights can adjust to zero the updating values that have been corrupted by tap weight errors. The DFT filters take advantage of this aspect; however, these filters do not provide full fault tolerance (each extra tap weight with the capability to perfectly overcome any single error of the filter, like with zero padding). This problem stems from the fact that if both members of a conjugate pair have errors, they may prevent the filter from converging properly [8]. Also, when there are real valued tap weights that have no conjugate pair, any error is detrimental. The MDFT rotates the frequency domain sampling points by a small angle in the z -plane so that there are no conjugate pairs and no values with only a real or imaginary part.

Recently published research results show that the MDFT has the capability to provide more robust adaptive fault tolerance than the conventional DFT/FFT adaptive filter when operating on real-valued signals [9]. However, this previous research has not tested the MDFT algorithm with speech signals. In this project, the enhanced robustness of adaptive fault tolerance provided by the MDFT as compared to the conventional DFT/FFT adaptive filter will be demonstrated with real-valued speech signals. Should the MDFT prove to be a viable source of fault tolerance, the reduction in computational complexity (relative to zero padding) could lead to power savings that would be greatly beneficial to the voice-communications industry.

Explanation of Hard/Soft Errors

Single-event effects (SEE) are classified into two categories, hard and soft errors. A soft error is not permanent and can be fixed upon resetting the device. A single event upset (SEU) is a soft error in which a charged particle passes through circuit elements, ionizing the medium and leaving a trail of electron–hole pairs [2]. Interconnect noise is a second type of soft error caused by the interactions of electric and magnetic fields of nearby circuit components. For example, two nearby parallel wires are two conductors separated by an inductor (air), forming a capacitor. Single bit upsets are one result of a soft error [1].

A hard error, on the other hand, is permanent and can only be fixed by replacing the hardware. A single event burnout (SEB) is an event in which circuit elements are pumped an amount of current above device specifications and the result is device failure. One potential result of this error is frozen bits [6].

ADAPTIVE FILTERING

In this project three types of adaptive filters, all based on adaptive linear prediction, are used to remove various typed of noise from real-valued speech signals—time domain adaptive filters, DFT/FFT adaptive filters, and MDFT/MFFT adaptive filters.

Linear Prediction Using a Time Domain Adaptive Filter

Consider the input signal $x(n)$ to be a single tone (sinusoidal signal) corrupted with a small amount of white noise. If a filter with linear prediction were to be applied (as it is in the block diagram below), the filter would be able to provide an estimation of the tone ($y(n)$) because it is highly correlated, and thus ‘visible’ in the previous samples viewed by the filter.

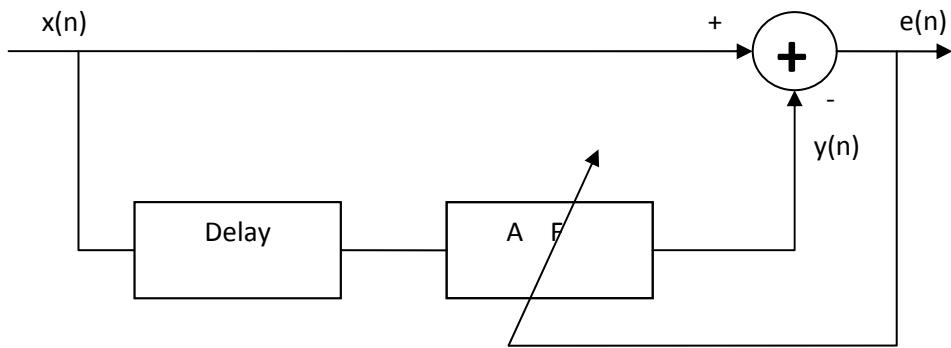


Figure 1: Time domain adaptive filter with linear prediction.

The algorithm of the adaptive filter (AF), in this case a least mean square (LMS) algorithm, minimizes the error ($e(n)$) given by the difference of $x(n)$ and $y(n)$ by an approximation to the steepest descent method, as shown below. Defining the error as

$$e(n) = x(n) - y(n), \quad (1)$$

the output of the adaptive filter is the input convolved with adaptive filter coefficients [7]

$$y(n) = h(n) * x(n), \quad (2)$$

where the coefficients of the adaptive filter ($h(n)$) are updated using the following equation, where delta is the step size of the LMS algorithm:

$$h(n+1) = h(n) + \Delta e(n)x(n) . \quad (3)$$

Linear Prediction Using a Transform Domain Adaptive Filter

This linear prediction concept can also be used in the frequency domain, as shown in Figure 2 [5]. In Figure 2 the $*$ indicates an LMS adaptive filter, D is representative of a delay between the current iteration and the beginning of the

window iterations and $\mathbf{x}[n]$ is a row vector of N samples of the original input signal that are passed through an N -point FFT to create frequency-domain samples $\mathbf{X}[k]$. $\mathbf{X}[k]$ then is filtered by an LMS adaptive filter such that the following equation (the inverse fast Fourier transform (IFFT)) is essentially performed. The transform domain LMS algorithm is explained below.

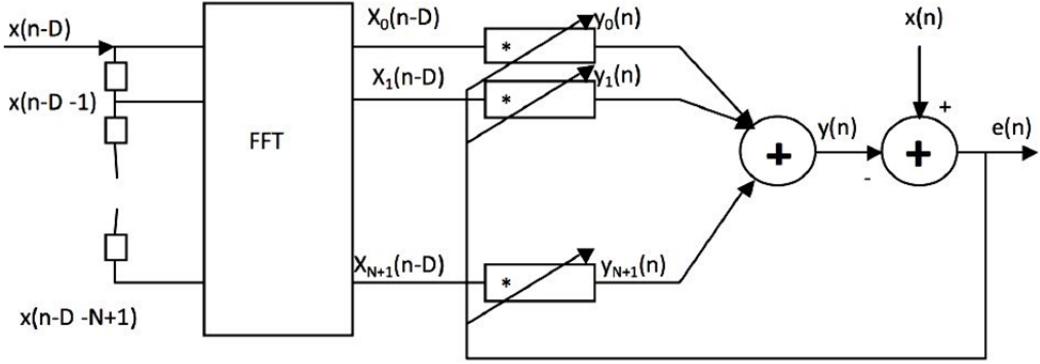


Figure 2: Transform domain adaptive filter with linear prediction.

The equation for the IFFT is as follows [4]:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j(2\pi/N)kn} \quad n = 0, 1, 2, \dots, N-1. \quad (4)$$

Note that each element of $\mathbf{x}[n]$ is the average of rotated values of the elements of $\mathbf{X}[k]$. Given that $\mathbf{x}[n]$ and $\mathbf{X}[k]$ have the same correspondence as $\mathbf{y}[n]$ and $\mathbf{Y}[k]$, the x input values can be substituted for y output values

$$\mathbf{y}[n] \leftrightarrow \mathbf{Y}[k] \quad (5)$$

$\mathbf{Y}[k]$ can then be substituted by the product of the transform domain tap weight (\mathbf{W}) and input (\mathbf{X}) vectors. $\mathbf{W}[k]$ is row vector:

$$\mathbf{Y}[k] = \mathbf{W}[k] \mathbf{X}[k]. \quad (6)$$

These two sets of substitutions lead to the formula

$$y[n] = \frac{1}{N} \sum_{k=0}^{N-1} W[k] X[k] e^{j(2\pi/N)kn} \quad n = 0, 1, 2, \dots, N-1 \quad (7)$$

In this particular situation, instead of reconstructing all N elements of the output in the time domain, only the first element is desired, thus $n = 0$. To compute a single frequency component of only the first time domain element, only the first elements of the W and X elements are used (indicated by a subscript of 0). The coefficient and summation remain because all of the frequencies must still be incorporated. The exponential in Equation 7 disappears because $n = 0$, and anything to the power of zero equals 1. Thus the following equation is derived:

$$y[0] = \frac{1}{N} \sum_{k=0}^{N-1} W_0[k] X_0[k]. \quad (8)$$

In the transform domain code, the first elements of the tap weight coefficients are combined into a column vector (W_0) and then multiplied (inner product) with $X(k)$

$$y[0] = \frac{1}{N} \mathbf{W}_0[k] \mathbf{X}_0[k] \quad (9)$$

The coefficients of the adaptive filter ($W(n) = W_0(k)$) are updated with the following equation

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \Delta e(n) \mathbf{X}^*(k), \quad (10)$$

where $X^*(k)$ is the conjugate vector of $X(k)$ and Δ is the step size of the function. With a smaller step size, a filter can converge with greater accuracy. However, a larger delta allows the filter to converge faster, provided it does not go instable.

Power Normalization in the Frequency Domain

Power normalization in proper conjunction with delta optimizes the convergence rate [5]. The frequency domain channels that have more power tend to converge faster than weaker channels [5]. The power normalization makes the filter converge faster by making all of the frequency channels approximately the same strength, so that the deviation between strongest and weakest channels is reduced, enabling the entire filter to converge faster. The power normalization factor is applied uniquely to each individual element of $\mathbf{X}[k]$ to form a power normalized input, $\mathbf{X}_p[k]$, which equals

$$X_p(k) = \frac{X(k)}{\alpha + \mathbf{X}_c(k)^* \mathbf{X}_c(k)}, \quad k = 0, 1, \dots, N-1 \quad (11)$$

The power normalized element k of X_p is the element k of original signal divided by the power of the channel signal. In Equation 11, $X_c(k)$ is a vector of N previous values of the k th frequency channel and alpha is a small factor relative to the magnitude of the channel and prevents the denominator from possibly equaling 0.

FAULT TOLERANCE

Soft and hard errors occur in the computer hardware and software and the purpose of fault tolerance is to overcome such errors. In the case of errors in computation can be especially problematic. Two transform domain methods discussed in this paper that address the serious problem of tap weight errors are zero padding and the modified discrete Fourier transform.

Zero Padding

Zero padding, or redundancy, has the same function as the TDAF algorithm with slight exceptions [8]. For each window of x , Z zeros are concatenated to the

end of window. Also the number of tap weights increases by Z . Although adding zeros does not change the signal, it permits more frequency channels to be created. With more frequency channels, each frequency channel contains fewer frequencies, and represents the frequency response with greater accuracy.

The zero padded input signal is represented as $x_z(n)$ (a row vector) and is the result of Z zeros being added to the N -sample portion of the input signal.

$$x_z(n) = [x(n) \ 0 \ \dots \ 0] \quad (12)$$

The DFT of $x_z(n)$ is computed to be $X_z(n)$ and the output $y(n)$ is the inner product of $X_z(n)$ and column vector of tap weights $W_z(n)$, expressed as:

$$y(n) = X_z(n)W_z(n) \quad (13)$$

Fault Tolerance of the Transform Domain Adaptive FIR Filter (DFT/FFT) Operating on Real-Valued Signals

Transform Domain Adaptive filters work on signals in the frequency domain, where complex arithmetic is used. The extra computations of the complex math can be avoided when working with real-valued signals. These extra computations, however, provide fault tolerance for real signals if the coefficient tap weights rely on the minimization of the real value of the error $e(n)$ [8]. Suppose that $x(n)$ is a real-valued input signal that produces a complex output $y(n) = y_r(n) + jy_i(n)$. The minimum error using the LMS algorithm is then:

$$|e(n)|^2 = [x(n) - y_r(n)]^2 + y_i(n)^2 \quad (14)$$

$$\min_w \{ |e(n)|^2 \} = \min_w \{ [x(n) - y_r(n)]^2 \} + \min_w \{ y_i(n)^2 \} \quad (15)$$

Because $x(n)$ is a real valued signal, the tap weights will minimize the imaginary value to equal zero.

$$\min_w \{ y_i(n)^2 \} = 0 \quad (16)$$

There are N real adjustable parameters in w (the tap weight coefficients), $N/2$ real and $N/2$ imaginary parameters that uniquely identify N real coefficients in the time domain. If only the real part of the error is calculated, then the imaginary value of the output is no longer constrained by any parameters. The over parameterization is created by having more than N parameters in the frequency domain used to define the N real valued tap weights in the time domain.

In the below equation for the tap weights (w), the matching subscripts indicate conjugate pairs and the subscripts 0 and $N/2$ are the real-valued tap weights that have no conjugate pair.

$$w = [w_0 \ w_1 \ w_2 \ \dots \ w_{N/2} \ \dots \ w_2 \ w_1] \quad (17)$$

The fault tolerance inherent in the FFT/DFT adaptive filter is dependent on the

fact that entire conjugate pairs (or the real valued tap weights which have no conjugate pair) do not have errors [3].

Fault Tolerance of the Transform Domain Adaptive FIR Filter Using the MDFT and Real-Valued Signals

Above, it was described how an adaptive DFT/FFT filter could provide significant fault tolerance if there were no errors in both members of the conjugate pair or the real values. The MDFT/MFFT exploits the fault tolerance of the use of minimizing only the real error, but also circumvents the problems of the conjugate pairing and real values by rotating the values by a value greater than 0, but less than π . Because of this rotation, no values lie exclusively on the real or imaginary axis, nor are there conjugate pairs, solving the problem of the DFT filter [9].

This shift occurs in the time domain. The vector $x(n)$ is the window portion of the input and $x_m(n)$ is the modified window:

$$x_m(n) = x(n)e^{-j\frac{\pi n}{2N}} \quad n = 0, 1, \dots, N-1 \quad (18)$$

EXPERIMENT DESCRIPTION

The experiment is separated into three parts: the time domain adaptive filter, the transform domain (DFT/FFT) adaptive filter, and the Modified Transform Domain adaptive filter. Each test will involve filtering a noise corrupted version of a given speech signal ('answers3.wav') which has the following time and frequency domain representations:

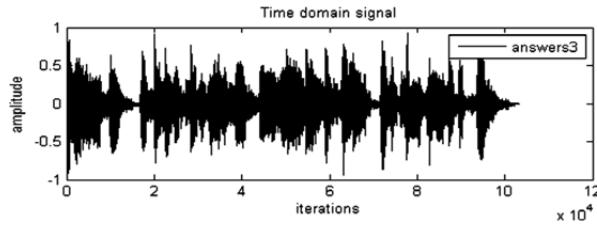


Figure 3: answers3.wav time domain

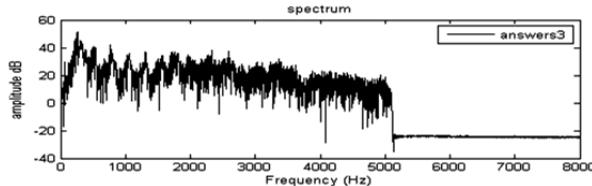


Figure 4: answers3.wav frequency spectrum

The original speech signal, 'answers3.wav' has a sampling frequency of 16,000 samples (16 bits per sample) per second, which does not violate the Nyquist sampling theorem. All of the corrupted speech signals have been created

in the program CoolEdit by mixing the answers3.wav signal with various noises and tones. All of the filters have been coded in MATLAB. (The MATLAB code is available from the authors upon request.)

Part 1: Time Domain AF

Part one of the project demonstrates the ability of the time domain adaptive filter to remove varying tones from a speech signal with linear prediction. No power normalization will be used in this portion of the project.

Part 2: Transform Domain AF (DFT/FFT)

Part two of the experiment tests the Transform Domain Adaptive Filters. Case a of this section repeats the test from Part 1, except that it will be in the transform domain. The complex error is used to update the tap weight coefficients, and delta will be selected so that it removes the tone as well as the time domain filter.

The corrupted signal ‘answers3-stepping5tones.wav’ will be used for the remainder of the tests of the DFT as well as the MDFT.

Case b tests a power normalized filter with a zero pad of 32 zeros that uses the complex error to update the tap weight coefficients by forcing errors into the filter and then comparing it against the same filter without any errors.

Case c of part two will employ a filter whose coefficients update using only the real portion of the error and have no zero padding.

Part 3: Modified Transform Domain AF (MDFT/MFFT)

Part 3 of the experiment employs the MDFT/MFFT with no zero padding. This part of the project demonstrates the ability of the MDFT adaptive filters in the same tests that the DFT/FFT adaptive filters are tested.

For quantitative analysis of the frequency domain adaptive filtering, the $L_{1/2}$ value of the error in every frequency component of the time domain between the desired output and the output of the tested filter is used. By breaking down the signals into the components of the time domain, the differences in the speech frequencies (degradations) are compared as well as the ability to remove the corruption.

$$L_{1/2}(n) = \left(\frac{1}{N} \sum_{k=0}^{N-1} |d_k(n) - x_k(n)|^{1/2} \right)^2 \quad (19)$$

The graph of the $L_{1/2}$ values of the signal characterizes the overall error of the signal and is used as a basis of quantitative comparison. When tones and short band noise are added to the original speech signal, the greatest error generated between the input of the signal and original speech signal is at the frequency of the added tone. To place greater value on the error in the degradation of the speech, the fractional value of L is chosen.

The major peaks in the subplots of Figure 5 are characteristic of the tone changing four times. The step sizes (delta) are chosen such that the same TD filter removes a set of five time dependent tones at notably different rates. With a relatively small delta (top graph), the filter removes the tone at a much slower rate which is seen as a constant downward slope from the major peaks. A relatively large delta causes speech degradation which is seen as jagged lines caused by the difference of other channels (such as the channels that contain the frequencies of speech) in portions after the major peaks.

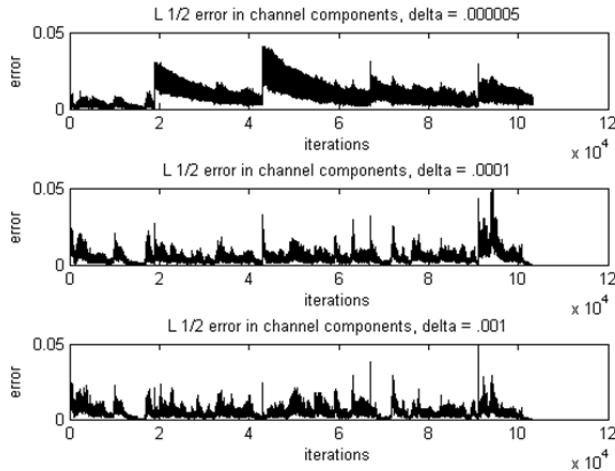


Figure 5: $L_{1/2}$ error – slow convergence or speech degradation

EXPERIMENT RESULTS

Part 1: Time Domain AF

In the first part of the experiment, the original speech signal was corrupted with a constant 4000 Hz tone (Figure 6).

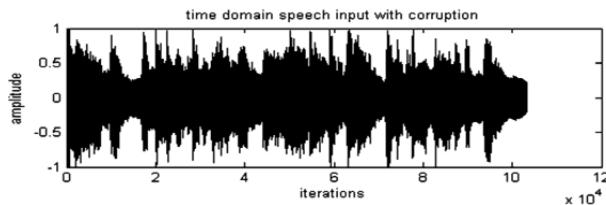


Figure 6: Speech signal corrupted with a single tone of 4000 Hz

Using the time domain adaptive filter, the non-power normalized filter was set to have 32 tap weights, a delay of 5 samples and a step size (delta) of 0.0005. The time domain output (Figure 7) has visibly removed much of the sinusoidal tone and better represents the original speech signal (Figure 3).

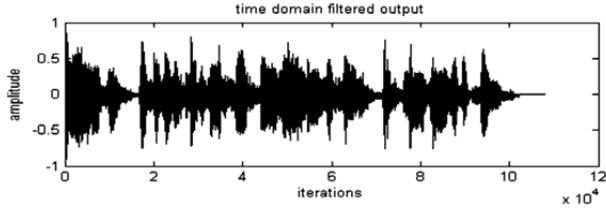


Figure 7: 4000-Hz tone removed from corrupted speech signal

Figure 8 depicts the frequency spectra of the adaptive filter output, input, and original speech signals after the 10,000th and 100,000th iterations. The peak at 4000 Hz is due to the added tone and is visible in the inputs, but not visible in the outputs. The filter has converged closer to its optimal value after 100,000 iterations, resulting in better removal of the tone.

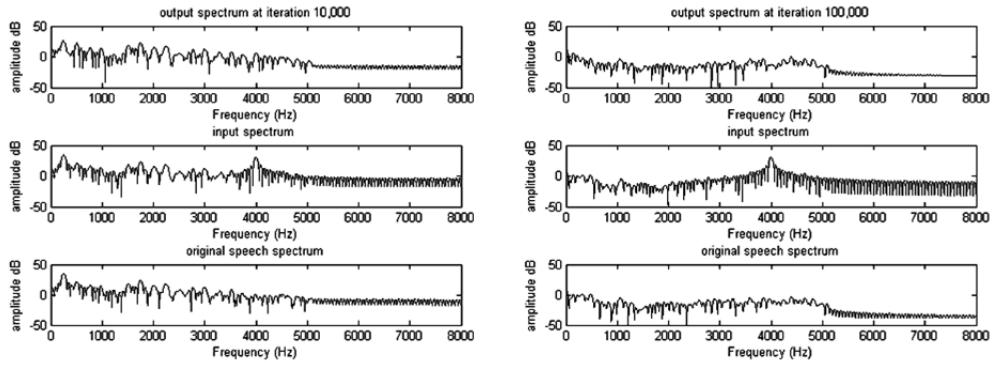


Figure 8: Frequency Spectra after 10,000th and 100,000 iterations

Part 2: Transform Domain AF (DFT/FFT)

Case a of Part 2 runs the same filter scenario as the part one except that the filter is a transform domain filter, not a time domain filter. The same original speech signal is corrupted with the same constant 4000-Hz tone. The non-power normalized filter itself has 32 tap weights, a delay of 5 samples, and a step size (delta) of 0.00001. Figure 9 shows the output of the transform domain adaptive filter in the time domain. The output is comparable to Figure 7, the output of the time domain adaptive filter.

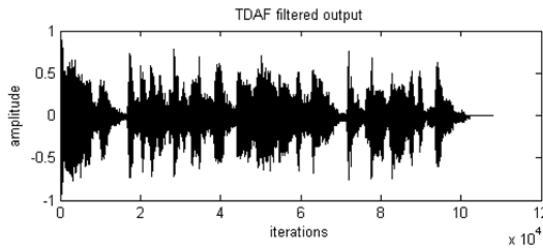


Figure 9: Transform domain output

Figure 10 depicts the frequency spectra of the output, input, and original speech spectra (from top to bottom respectively). Comparable to Figure 8, Figure 10 shows the same attenuation of the 4000-Hz peak from the input to the output. As before, these figures show that the filter's performance after 100,000 iterations is better than after 10,000 iterations.

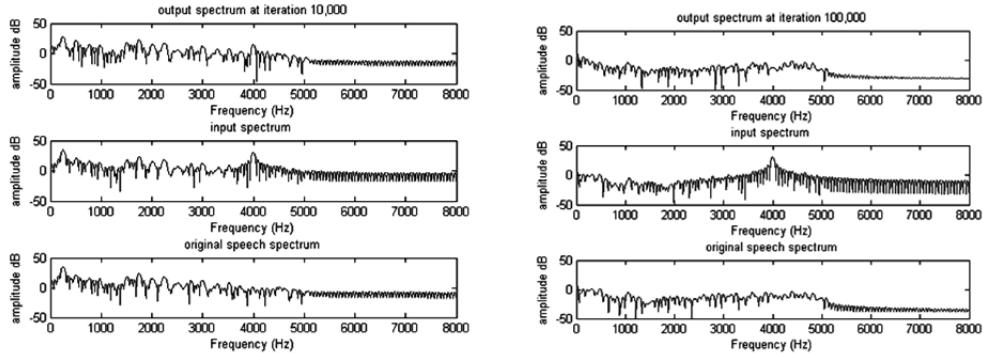


Figure 10: Transform domain output

For case b of Part 2, the zero padding algorithm is tested in the transform domain with power normalized frequency channels and complex error is used to update the tap weight filter coefficients. In this example, 32 tap weights are used with 32 zeros padded to each window and the step size (delta) is 0.0001. The corrupted signal ‘answers3-stepping5tones.wav’ has multiple tones each lasting 1.5 seconds. The tones, in order of time, are 1000 Hz, 2222 Hz, 3333 Hz, 4444 Hz, and 5555 Hz. In the first step, no errors are forced into the filter; however, in the second step, Error 1 is forced upon the filter. Error 1 affects one of the real valued tap weights, as well as 9 other conjugate pairs. Starting with iteration 42,000, the specified tap weight coefficients are no longer updated.

In Figure 11, the $L_{1/2}$ of the filter with Error 1 (top) and the filter without Error 1 (bottom) are shown. The four major peaks visible in both graphs are evidence of the shift in tone corruption of the original signal. The original calculated error at iteration 42,000 (in the second tone) is not visible because the filter has already converged at this point, and the tap weight filter changes are much less significant at this point. The error is first seen in the second major peak (located between iterations 40,000 and 60,000). Certain errors, such as Error 1, in the filter, visible in the $L_{1/2}$ graphs, cause slower convergences to the minimal amount of error.

Each of the frequency spectra in Figure 12 illustrates one of the five tones being removed without Error 1 in the filter; at iteration 10,000, the corrupting tone is 1000 Hz, at iteration 30,000, the corrupting tone is 2222 Hz, at iteration 50,000, the corrupting tone is 3333 Hz, and at iteration 100,000, the corrupting tone is 5555 Hz. Note that in the frequency spectrum of the 10,000 iteration, the linear predictive filter identifies the 1000 Hz tone, as well as a tone of surrounding tones between 0 and 2000 Hz. However, at the 100,000th iteration,

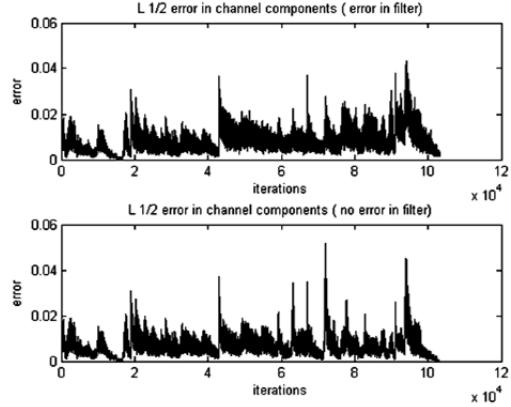


Figure 11: $L_{1/2}$ error – Error 1, zero padding, complex error

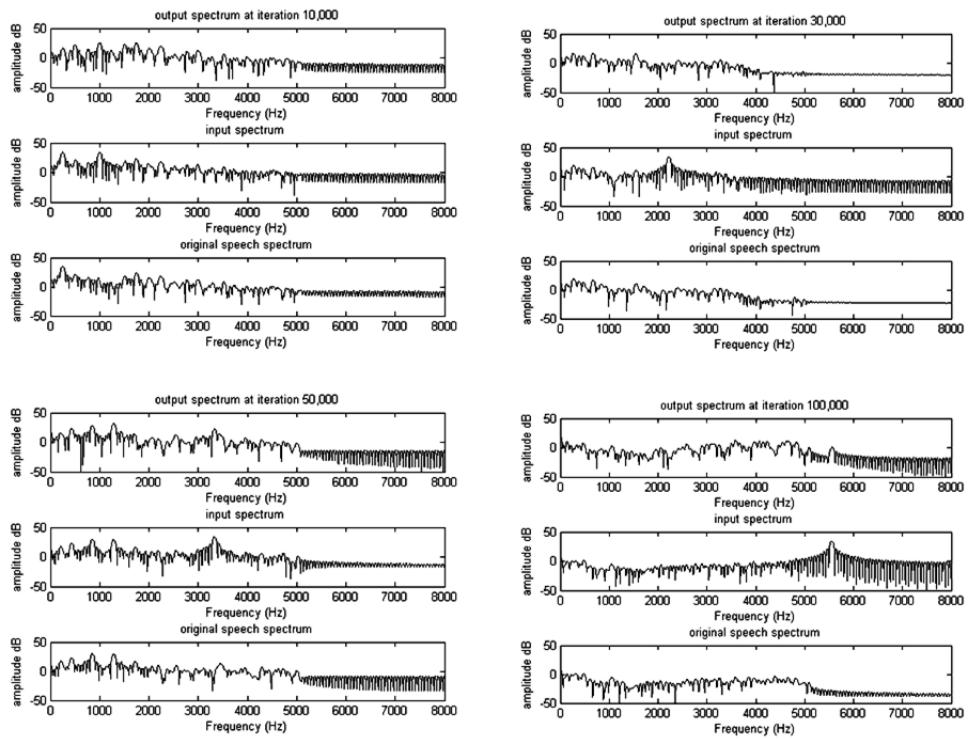


Figure 12: Spectrum frequencies – no error, complex error, zero padding

the linear predictive filter identifies and removes only one tone, the 5555 Hz added tone.

Because the Error 1 occurs at the 42,000 iteration, the frequency spectra of the first two iterations of figure 13 are not affected by the error in the filter. Before Error 1 has occurred (iteration 10,000) the frequency spectrum shows that the tone added has not been removed. After the error has occurred, in iterations 50,000 and 100,000 the filter is unable to remove the tone. The last frequency

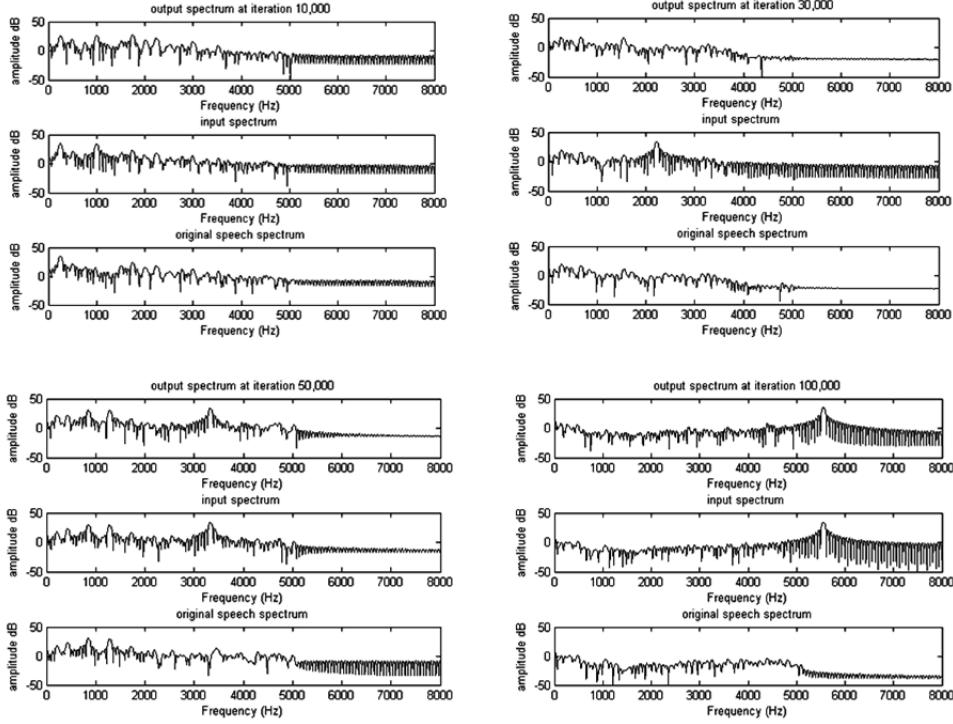


Figure 13: Spectrum frequencies – Error 1, complex error, zero padding

spectrum of the filter shows that the tone was not removed and added noise at 4500 Hz.

For case c of Part 2, the TD filters have no zero padding, power normalized frequency channels and the real-valued error is used to update the tap weight filter coefficients. In this example, 32 tap weights are used and the step size (delta) is 0.0001. The corrupted signal ‘answers3-stepping5tones.wav’ is the same as the previous experiment.

In the first step, Error 1 is forced upon the filter and in the second, Error 2 is forced upon the filter. Error 2 is designed to implement the maximum number of theoretical errors prescribed by [9]. Error 2 affects all conjugate pairs, but not the real-valued tap weights. The real portions of the first members and the imaginary portions of the second members of the conjugate pairs are not updated from their values at iteration 42,000.

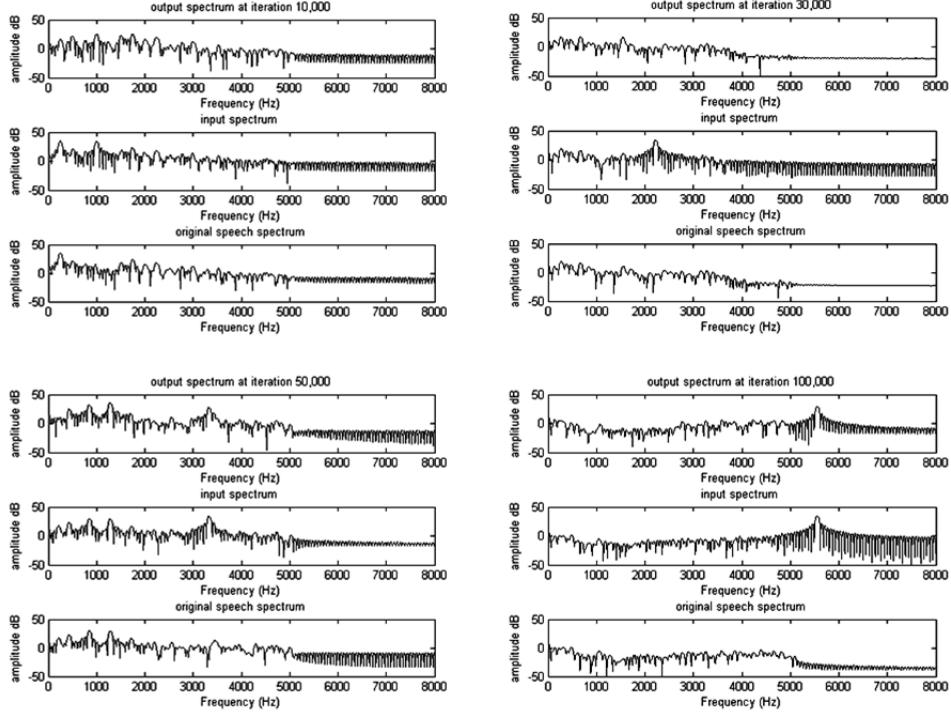


Figure 14: Spectrum frequencies – Error 1, real error, no zero padding.

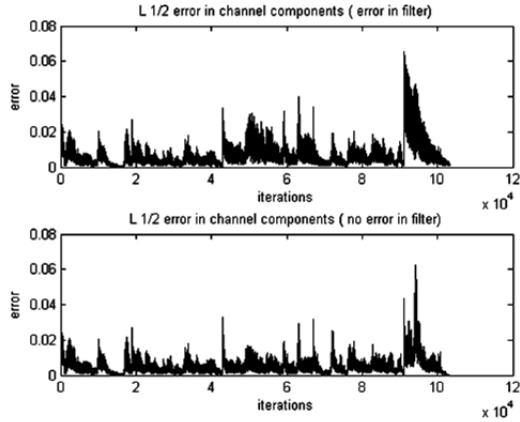


Figure 15: $L_{1/2}$ error – Error 1, real error, no zero padding.

Note the difference between the frequency spectrums of the 30,000th iterations in Figures 13 and 14. Both figures demonstrate a filter working on the same error. However, Figure 14 shows that the TDAF filter with no zero padding, but tap weight coefficients updated by the real of the error removing the 2222 Hz tone added. Figure 12, in comparison to Figures 13 and 14, shows that the zero padded filter based on complex error does remove the tone, however, at a slower rate.

Note the differences in the $L_{1/2}$ values for the filters with errors in Figures 11 and 15. The real error based filter with no zero padding (Figure 15) converges more rapidly and to a lower error floor than the zero padded filter with complex error (figure 11).

In the next part, Error 1 is replaced by Error 2 (Figure 16). In iteration 30,000, before the error, the filter removes the tone of 2222 Hz completely. After the error, in iteration 50,000 and beyond, the filter is able to identify, but only slightly attenuate the added tones.

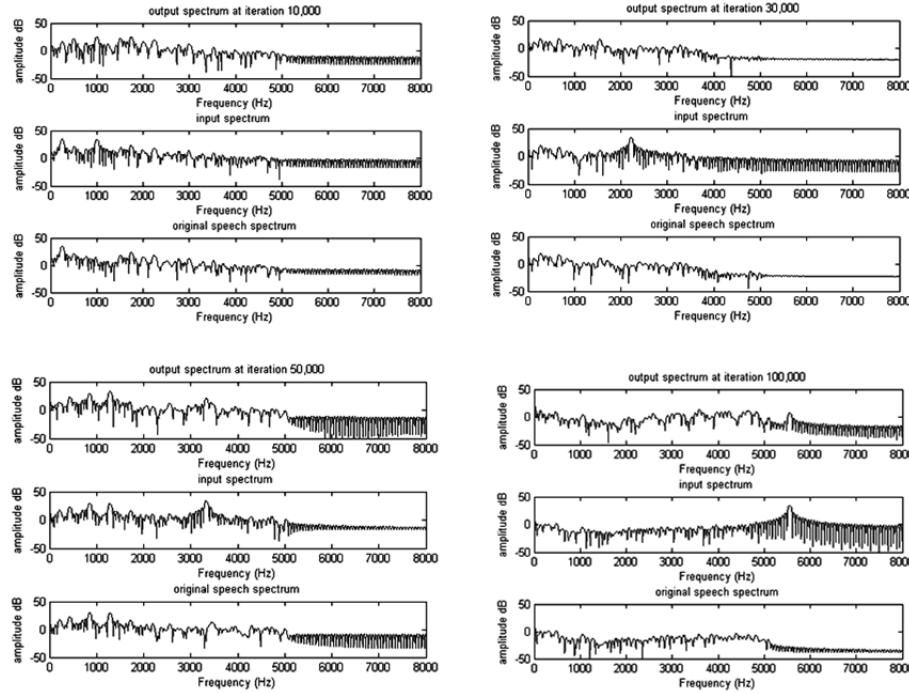


Figure 16: Spectrum frequencies – Error 2, real error, no zero pad

The $L_{1/2}$ values of the TD adaptive filters with real error and no zero pad (Figure 17) are nearly identical, meaning that even with the error, the filter was able to converge properly at the same rate as if there were no error.

For Part 3, MDFT filters have power normalized frequency channels, no zero padding, and real-valued error is used to update the tap weight filter coefficients. In this example, 32 tap weights are used and the step size (delta) is 0.0001. In the first step, Error 1 is forced upon the filter. In the second step, Error 2 is forced upon the filter. In both cases, the number of tap weights is set to 32, the delay to 5, and delta to 0.0001 (Figure 18). The filter is able to identify the correct tone in

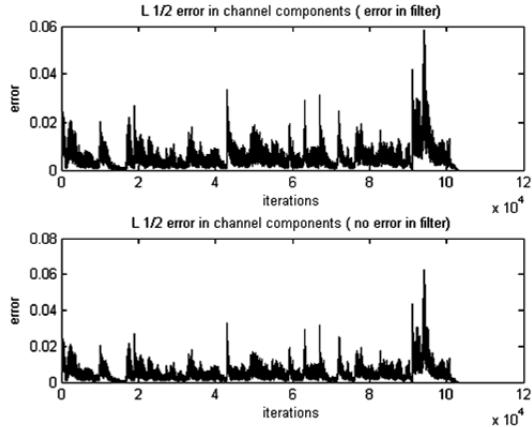


Figure 17: $L_{1/2}$ error – Error 2, real error, no zero padding
Part 3: Modified Transform Domain AF (MDFT/MFFT)

iteration 10,000 as well as increase tones in the 1500 to 2000 Hz range. After the error, however, the filter is unable to properly converge in that it does not completely remove the tones added, but does identify them.

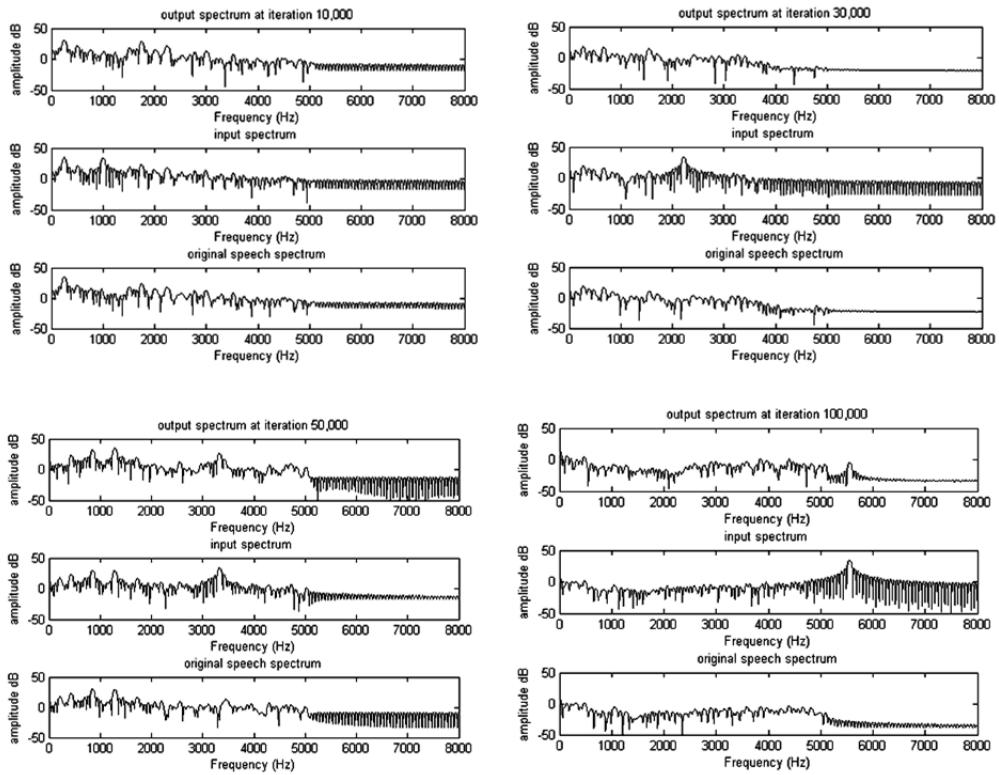


Figure 18: Spectrum frequencies – Error 1, MDFT

The $L_{1/2}$ error for the MDFT filter with Error 1 (Figure 19), has approximately the same convergence rate and noise floor as the TDAF with real error and no zero padding with the same error (Figure 15). The first major peak of Figure 19 is higher than that of Figure 15, the initial amplitude of the added signals is higher in the MDFT.

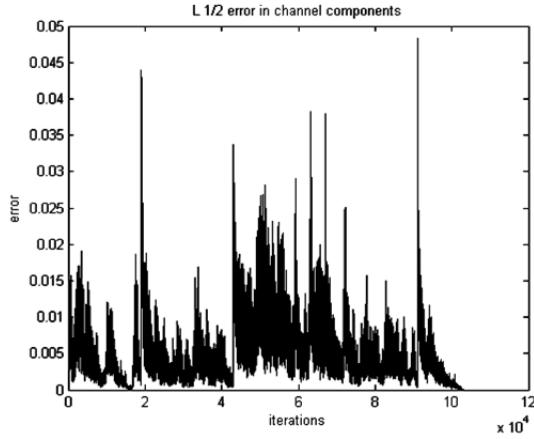


Figure 19: $L_{1/2}$ error - Error 1, MDFT

The MDFT filter spectrum outputs (Figure 20) show that before the error occurs at iteration 42,000, the first two filters are able to identify and remove the tones completely. Compared to Figure 16 at these iterations, the MDFT appears to identify and remove the tones faster than the real error with no zero pad filter. After the error occurs, the filter is able to identify, but not completely remove the added tone, which are the same results as the Figure 18.

Compared to the same error occurring in a DFT/FFT adaptive filter (Figure 17), Figure 21 performs with approximately the same amount of error. As with Error 1, the first major peak of the MDFT filter is higher than that of DFT with real error, the initial amplitude of the added signals is higher in the MDFT.

DISCUSSION

Speech signals have a majority of their signal energy within the 0 to 3400 Hz range, with the range of 0 to 2000 Hz having the greatest magnitude. Each filter was able to remove tones from the signal, but tones added in this 2000 Hz range were harder for the linear predictive adaptive filters to identify. With more iterations, the filters better removed the added tones.

Due to the nature of linear prediction, it works best when filtering a tone from white noise (two extremes complete correlation and no correlation). Speech is in the middle of this spectrum, as it has only some correlation. The linear prediction methods “predict” highly correlated parts of signals. Testing showed that they are most successful at removing unwanted noise from speech when the noise from

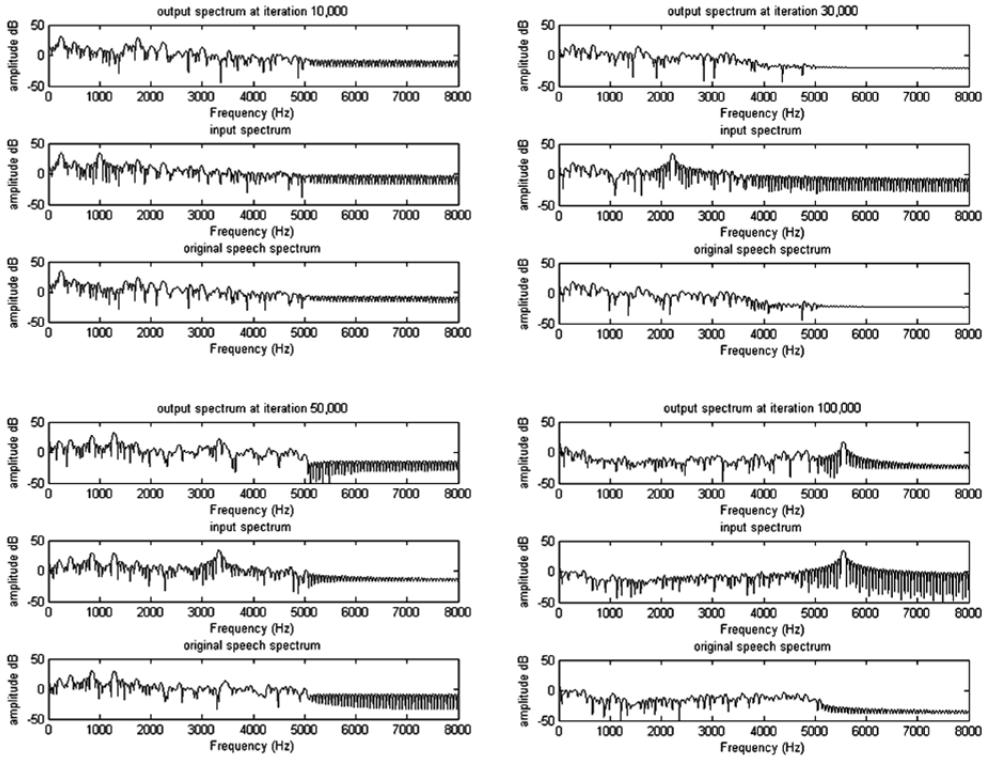


Figure 20: Spectrum frequencies – Error 2, MDFT

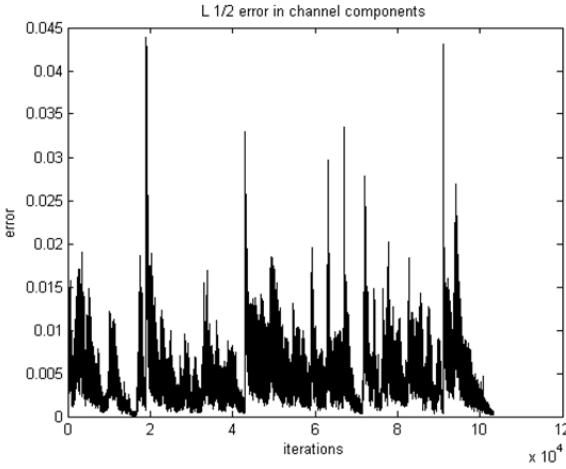


Figure 21: $L_{1/2}$ error - Error 2, MDFT

speech is highly correlated, such as a single tone, especially if the tone is more pronounced in the speech (higher amplitude and at a frequency outside of the speech range). However, 10 Hz wide (3990 to 4000) noise is difficult for the linear prediction to track in speech, which is why tones are used in the experiment.

The problem arises in testing the fault tolerance when only a very narrow band of noise (a tone) is being filtered because only two taps are necessary to remove a single tone (only two taps are required to filter a tone). However, fewer taps working on the filter degrades the speech signal because the notch filter is wider.

Comparing Figures 16 and 21 in iterations 10,000 and 100,000, the difference between the MDFT and the DFT with real error and no zero padding is made clear. The MDFT is able to identify and remove the tones better in certain situations. The similarities in Figures 20 and 24 show the DFT and MDFT filters having approximately the same performance levels.

CONCLUSION

At the most rudimentary level of adaptive filtering with no fault tolerance or power normalization, the time and transform domain filters worked equally well in removing the corruption quickly and efficiently without degrading the signal. The transform domain has the enhanced capability of power normalizing the channels in the frequency domain which allows for faster convergence without added speech degradation.

Adding zero padding to a TD filter based on complex error makes the filter converge at a slower rate, with greater degradation in the speech, relative to TD adaptive filters with no zero padding and tap weights that are updated with real error. The MDFT has advantages over the DFT with real error and no zero padding in certain cases of errors in the filter.

In future experiments, to fully test the adaptive filters with speech, a filter that employs a method other than linear prediction, such as system identification, would be better. This alternate experiment would enable the speech to be corrupted by a wider variety of noises that reflect the realistic corruptions of speech providing a more accurate perspective as to the function of the adaptive filters working on signals.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, D. Sova would like to thank her mentors, David Salvia and Kenneth Jenkins for their hard work and dedication to the project.

REFERENCES

- [1] Ekekwe, N., "Power dissipation and interconnect noise challenges in nanometer CMOS technologies," *IEEE Potentials*, vol. 29, no. 3, pp. 26-31, May-June 2010, doi: 10.1109/MPOT.2010.935825.
- [2] Holbert, Keith E., "Single Event Effects," Arizona State University. Retrieved from <http://holbert.faculty.asu.edu/eee560/see.html>.
- [3] Jenkins, W. Kenneth, Andrew W. Hull, Jeffrey C. Strait, Bernard A. Schnaufer, and Xiaohui Li, *Advanced Concepts in Adaptive Signal Processing*, Kluwer Academic Publishers, 1996.

- [4] Jenkins, W. K., "Fourier Methods for Signal Analysis and Processing," Chapter 4, Vol. 1, *The Fundamentals of Circuits and Filters Handbook*, 3rd edition, Wai-kai Chen, ed., CRC Press, 2009.
- [5] Jenkins, W. K., D. F. Marshall, and C. Radhakrishnan, "Transform Domain Adaptive Filters," in *The DSP Handbook*, Vijay Madisetti and D. Williams, eds., 2nd edition, CRC Press, 2010.
- [6] Matipuram, Ritesh, Edwin C., Wee: (September 30, 2004) "Soft Errors' on System Reliability," *Electronics Design, Strategy, News*, Retrieved from http://www.edn.com/article/473992-Soft_errors_impact_on_system_reliability.php.
- [7] McClellan, James H., W. Ronald W. Schafer, and Mark Yoder, *DSP FIRST A Multimedia Approach*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [8] Radhakrishnan, C. and W. K. Jenkins, "Special fault tolerant properties of FFT-based transform domain adaptive filters," *Proceedings of the International Symposium on Circuits and Systems*, pp. 1140-1143, May 18-21, 2008.
- [9] Radhakrishnan, C. and W. K. Jenkins, "Modified discrete Fourier Transforms for fast convolution and adaptive filtering," *Proceedings of the International Symposium on Circuits and Systems*, pp. 1611-1614, May 30 - June 2, 2010.
- [10] Shivakumar, P., M. Kistler, S. W. Keckler, D. Burger, and L. Alvisi, "Modeling the effect of technology trends on the soft error rate of combinational logic," *Dependable Systems and Networks, 2002, DSN 2002*, pp. 389-398, 2002.

VERIFICATION OF A MARGINAL OSCILLATOR FOR CONTINUOUS-WAVE QUADRUPOLE RESONANCE SPECTROSCOPY

Jacob Wilson,* Jeffrey Schiano,[#] and Thomas Tyson[†]

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

*Undergraduate Student of
Department of Physics & Engineering
West Virginia Wesleyan College
Buckhannon, WV 26201

ABSTRACT

This study verifies the operation of a continuous wave (CW) spectrometer by detecting and characterizing the quadrupole resonance (QR) transition in hexamethylenetetramine (HMT). In comparison to pulsed spectroscopy methods, CW spectroscopy provides two advantages. First, the amplitude of the applied magnetic field is smaller by orders of magnitude. As a result, the cost of a CW spectrometer is less than that of a pulsed spectrometer and the magnetic fields present no health hazards in applications such as explosives detection at aviation security checkpoints. Second, for materials where the ratio of the spin-lattice and spin-spin relation time is large, such as the explosive PETN, the signal-to-noise ratio per unit time is higher for CW spectroscopy. The CW spectrometer in this study is a marginal oscillator (MO). The key figure of merit for a MO is the conversion gain, which is the ratio of the percent change in oscillation amplitude to the percent change in electrical loss due to the absorption of energy by quadrupole nuclei. Previous work by Tyson shows that the conversion gain of the MO used in this study exceeds ten. Using this MO as a CW spectrometer, the width and location of the QR transition in HMT are shown to be consistent with data obtained from a pulsed spectrometer.

INTRODUCTION

Quadrupole Resonance Spectroscopy

The physical basis for QR spectroscopy is the electrostatic interaction between the electric quadrupole moment tensor of the nucleus and the electric field

[#] Faculty Mentor

[†] Graduate Mentor

gradient tensor of the surrounding electronic charges.^[1, 2] The quadrupole moment tensor describes the deviation of the nuclear charge distribution from spherical symmetry. The quadrupole moment tensor is described by a single parameter, the electric quadrupole moment, eQ , that is, a measure of the deviation of the nuclear charge density from spherical symmetry. The electric field gradient tensor is a description of the electric field gradient at the nucleus caused by the surrounding electronic charge. The electric field gradient tensor is described by two parameters, the largest gradient, eq , and an asymmetry parameter, η .

A quadrupolar nucleus possesses a magnetic moment whose value is proportional to the intrinsic angular momentum, or spin, of the nucleus. Because the spin is specified by the quantum number I , the electrostatic interaction energy between the quadrupole moment and the electric field gradient tensors is quantized, leading to preferred orientations of the nuclei that correspond to specific energy levels. This study is concerned with the QR transition associated with nitrogen-14 nuclei where $I = 1$ and the energy levels are

$$E_z = \frac{-e^2 q Q}{2}, E_x = \frac{-e^2 q Q}{4}(1-\eta), E_y = \frac{-e^2 q Q}{4}(1+\eta). \quad (1)$$

The lowest energy level, E_z , occurs when the principal axes of the electric quadrupole moment and electric field gradient tensors coincide. Equivalently, the magnetic moment of the nucleus is collinear with the direction of the largest electric field gradient.

In nitrogen-14 compounds the transition frequencies range from about 500 kHz to 5 MHz. The transition frequencies of a sample are, generally, not sharp. Due to the presence of impurities and crystalline strains, each nucleus possesses a slightly different resonant frequency, resulting in a distribution of transition frequencies known as a lineshape and characterized by a linewidth. In powdered samples that have been crystallized to remove impurities and relax strains, linewidths are typically on the order of 1 kHz.

Pulsed QR Spectroscopy

Pulsed spectroscopy provides a means for observing QR transitions.^[3] As in nuclear magnetic resonance (NMR), the sample is placed within a radio-frequency (RF) coil. The orientations of the quadrupolar nuclei are perturbed by applying a rotating magnetic field whose frequency corresponds to the energy difference between any two preferred orientations. After the external magnetic field is removed, the nuclei precess about the principal axis of the electric field gradient tensor. The precession of the nuclear magnetic moment induces a voltage across the RF coil terminals, revealing the presence of the quadrupolar nuclei. Because the lineshape is not sharp, the nuclei precess at different rates, and owing to destructive interference, the induced voltage relaxes to zero. For this reason, the QR response to a single RF pulse is known as a free-induction decay (FID).

Figure 1 shows the Lorentzian lineshape defined by its central transition frequency ν_*

$$L(f) = \frac{1}{1 + [T_2^*(f - \nu_*)]^2}, \quad (2)$$

and full-width half-maximum (FWHM)

$$FWHM = \frac{2}{T_2^*}. \quad (3)$$

In response to a RF pulse applied at the transition frequency ν_* , a QR transition characterized by a Lorentzian distribution induces an exponentially decaying voltage

$$s(t) = S_0 e^{-t/T_2^*} \cos(2\pi f_r t) \quad (4)$$

across the probe coil terminals. The QR receiver demodulates the probe signal to a much lower frequency f_r , resulting in the received signal $s(t)$ in Figure 1. The time domain signal in Figure 2 defines the FID, and its magnitude-squared Fourier transform is the Lorentzian lineshape shifted in frequency so that it is centered at f_r .

After the application of an RF pulse, the population of the three energy levels relaxes to their thermal equilibrium value through interactions of the nuclei and the surrounding crystal lattice. This relaxation is characterized by the spin-lattice time constant T_1 . The time necessary for the nuclear magnetic moment orthogonal to the precession axis to return to zero after excitation is T_2 , or the spin-spin relaxation time constant. This relaxation time is a result of the magnetic interactions among different nuclei. The time constant T_1 is always larger than T_2 .

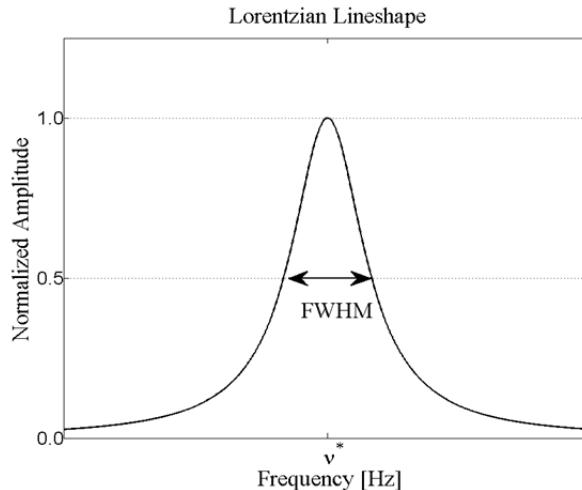


Figure 1: The Lorentzian lineshape is characterized by its center frequency ν_* and full-width half-maximum (FWHM).

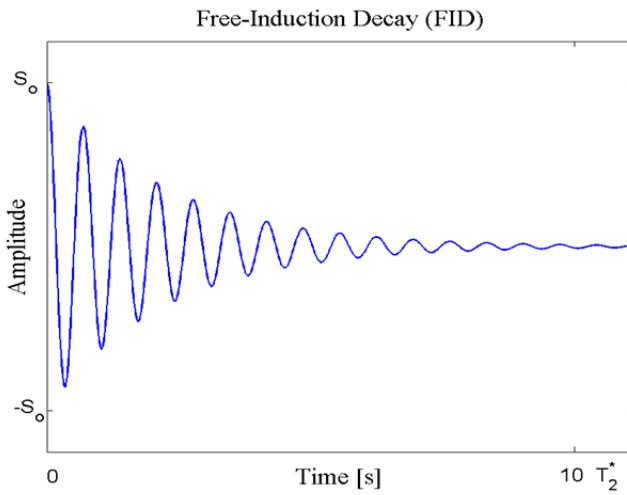


Figure 2: The free-induction decay (FID).

Continuous Wave QR Spectroscopy

The first measurements of QR and NMR transitions were observed using continuous-wave (CW) spectroscopy. In this approach, a sample is placed within a tuned RF circuit that is driven by a RF current whose frequency is slowly varied across the region where the QR transition exists. The amplitude of the sinusoidal voltage across the tuned circuit is recorded as function of excitation frequency. When the excitation frequency is within the linewidth of the transition, the nuclei absorb energy. The maximum adsorption of energy occurs when the frequency of oscillation matches the transition frequency and the magnetic field H_1 seen by the nuclei satisfies the constraint^[4]

$$\gamma^2 H_1^2 T_1 T_2^* = 1, \quad (5)$$

where γ is the gyromagnetic ratio which is 1933 rad/sec/Gauss for nitrogen-14. The adsorption of energy reduces the amplitude of oscillation, which reveals the presence of the quadrupolar nuclei. One can draw a comparison between pulsed and CW techniques using linear system concepts. In pulsed methods, the RF pulse produces an impulse response that is observed using the search coil, while the CW method generates a Bode plot that represents the frequency response function of the system.

It has been shown that sensitivity of pulsed and CW techniques are equivalent in many situations.^[5] In general, pulsed spectroscopy is superior as its signal-to-ratio per unit measurement time is significantly larger than that for the CW approach. The reason for this difference is that a narrow RF pulse has a broad spectrum which excites all the nuclei within the lineshape. In contrast, the CW approach uses a monochromatic source that excites a limited number of nuclei at a time.

There are several reasons for revisiting CW methods. First, unlike pulsed-methods, the peak power for CW detection systems is significantly smaller. As a result, the instrumentation is less expensive and the RF levels are sufficiently small that they pose no risk when scanning people for explosives at aviation security checkpoints. Another advantage is that for materials where the ratio of the spin-lattice and the spin-spin relaxation times is much greater than unity, the sensitivity of CW methods is superior to that of pulsed methods.^[5] This fact is of importance in explosives detection as the ratio is about 600 for the explosive pentaerythritol tetranitrate (PETN).

Marginal Oscillator CW Spectrometer

Marginal oscillators provide a way to detect small changes in the losses within a harmonic oscillator. Roberts^[6] in 1947 and Rollin^[7] in 1948 credited Pound for developing the first marginal oscillator to observe nuclear magnetic resonance within solids. The use of the marginal oscillators now extends to many other areas including the characterization of defects in silicon,^[8] determining the mechanical properties of thin-film super-conductors,^[9] measuring penetration depth in superconductors,^[10] and ion cyclotron resonance spectroscopy.^[11]

Figure 3 shows a conceptual representation of a marginal oscillator, which consists of a tuned circuit connected to a nonlinear negative resistance. The loss mechanisms associated with the tuned RLC circuit are represented by the resistance R . The nonlinear negative resistance, implemented as a dependent current source, $G(v)$, is placed in the circuit in order to obtain self-sustaining oscillations. The dependent current source is a memoryless, nonlinear element controlled by the voltage, v , across the tuned RLC circuit. To sustain a steady-state oscillation, the dependent current source must provide enough power to exactly offset the losses represented by R . This implies that the dependent current source must appear as a negative resistance with an equivalent value $-R$, so that the parallel resonant circuit sees an open circuit.

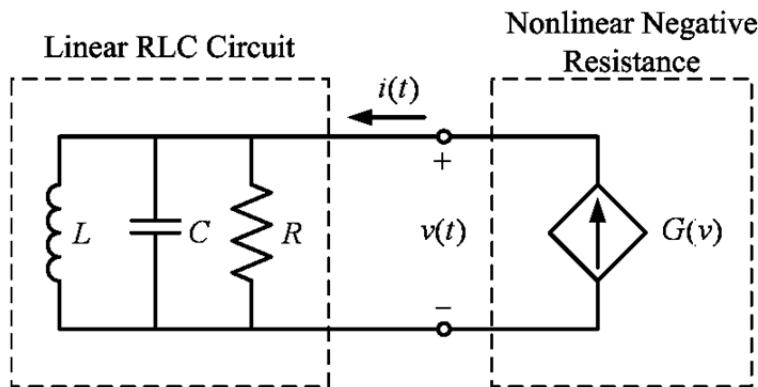


Figure 3: Schematic representation of a marginal oscillator

In 1975, Viswanathan presented the first analysis that predicts how small changes in the losses effect the oscillation amplitude.^[12] Viswanathan defined the sensitivity of the oscillator amplitude with respect to the changes in the losses R of the tuned circuit as the conversion gain

$$G_c = \frac{\% \text{ change in } A}{\% \text{ change in } R} = \frac{\Delta A / A \times 100}{\Delta R / R \times 100} = \frac{\Delta A}{\Delta R} \frac{R}{A}. \quad (6)$$

In the limit as ΔR approaches zero,

$$G_c = \frac{\partial A}{\partial R} \frac{R}{A}. \quad (7)$$

Viswanathan determined that by appropriately designing the shape of $G(v)$, one is able to achieve conversion gains that are significantly larger than unity.

For nitrogen-14 samples, the maximum change $\Delta R/R$ in losses is typically on the order of 10^{-6} , while the amplitude, A , is on the order of 1 V in order to satisfy Equation (5). Even for conversion gains greater than ten, in order to observe the change ΔA in the presence of thermal noise the measurement bandwidth must be restricted to a fraction of a Hertz using a lock-in amplifier.

For lock-in detection, the frequency of the CW spectrometer is modulated by electrically varying the capacitance in Figure 3. Let $f(t)$ denote the instantaneous frequency of the MO, f_0 the nominal frequency of the MO, f_{lock} the frequency of lock-in detection, and f_d the frequency dither in the MO oscillation frequency so that

$$f(t) = f_0 + f_d \sin(2\pi f_{\text{lock}} t). \quad (8)$$

When the value of $f(t)$ lies within the linewidth of a QR transition, the variation in oscillation frequency will result in a variation of adsorption and hence oscillation amplitude. By demodulating the envelope amplitude at the lock-in frequency f_{lock} and lowpass filtering the result, the measurement bandwidth is reduced to a fraction of hertz. It can be shown that the output of lock-in detection approximates the first derivative of the lineshape as long as the dither frequency f_d is a small fraction of the transition FWHM.^[4,13] Figure 4 shows the derivative of the Lorenztian lineshape. Note that the derivative vanishes at the transition frequency. It can also be shown that the spacing between the peaks is equal to $\text{FWHM}/\sqrt{3}$.^[13]

Objectives

The specific aim of this study is to detect a QR transition using the test bed MO investigated by Tyson.^[14] Using a pulsed spectrometer, the transition frequency and linewidth are observed for a nitrogen-14 compound that has a known QR transition. The sample is then placed in the probe coil of the MO, and the frequency of the CW spectrometer is swept over the known location of the transition. If the CW spectrometer is working as expected, the output of the lock-in amplifier as function of spectrometer frequency should resemble the derivative

of the Lorentzian line shape as shown in Figure 4. In addition, the location of the transition and the separation of the peaks should match the values predicted by data obtained from the pulse spectrometer.

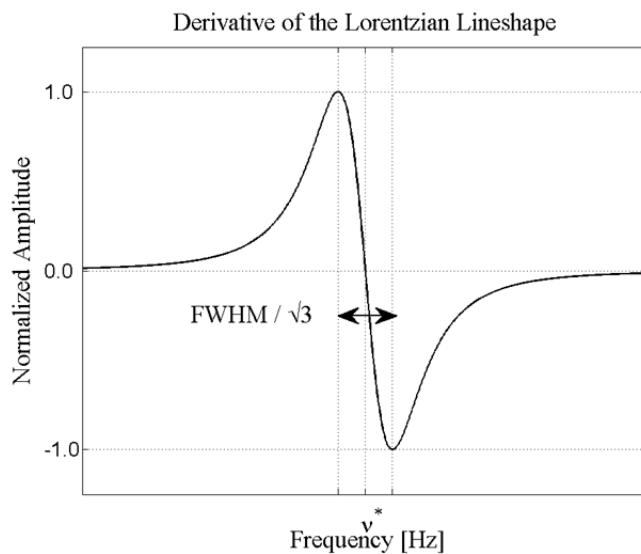


Figure 4: First derivative of the Lorentzian lineshape.

EXPERIMENTAL METHODS AND MEASUREMENTS

Material Selection

The sample is 9 g of hexamethylenetetramine (HMT) contained in a 19 mm by 65 mm glass vial. The QR spectrum of HMT is well studied as it is used in the synthesis of many compounds including plastics and pharmaceuticals. Because the asymmetry parameter η is zero, HMT has a single transition frequency that is about 3.308 MHz at room temperature.^[15]

Pulsed Spectrometer Measurements

A detailed description of the pulsed spectrometer used in this study appears in a paper by Schiano.^[16] The amplitude of the pulsed magnetic field is set to 10 G. From experiment, the pulse width that maximizes the amplitude of the QR response is approximately 17 μ s. Figure 5(A) shows a FID following the 17- μ s pulse, while Figure 5(B) shows the squared-magnitude of the Fourier transform of the FID. In order to improve the measurement signal-to-noise ratio, the data in Figure 5 represents the average of 36 experiments, each separated by several T_1 time constants. The receiver frequency of the spectrometer is set about 10 kHz above the QR transition. The FWHM is indicated by the solid line draw across the lineshape in Figure 5(B). The transition frequency is 3.809 MHz, the FWHM is 725 Hz, and the corresponding value of T_2^* is 440 μ s. As the transition frequency is sensitive to temperature, small variations on the order of 1 kHz are expected due to fluctuations in room temperature. Using a saturation recovery pulse sequence, the estimated spin-lattice relaxation T_1 is 18 ms.^[15]

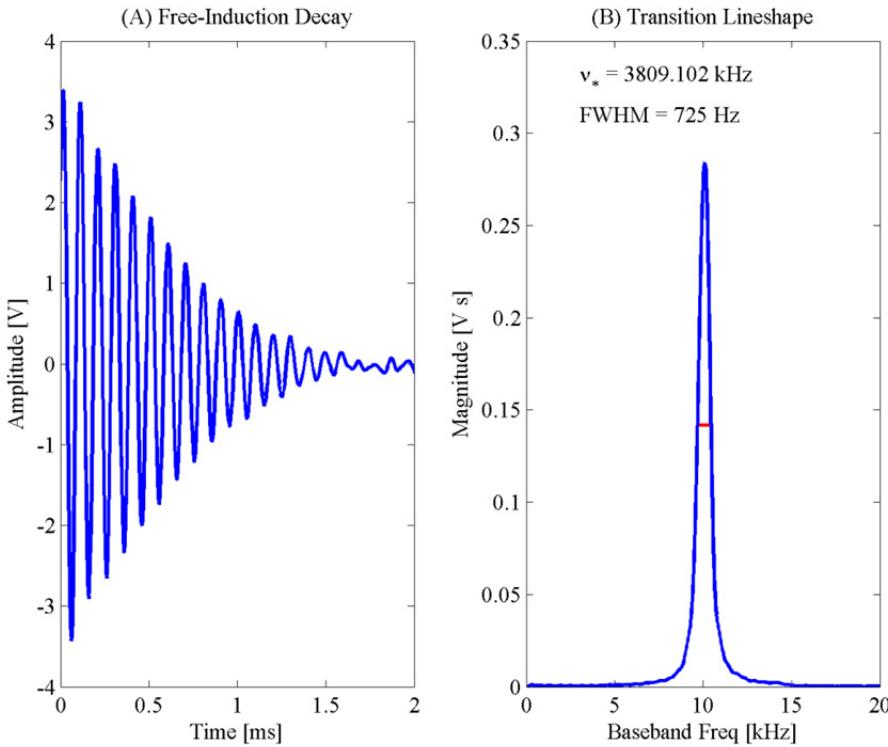


Figure 5: (A) Free-induction and (B) lineshape of HMT observed at room temperature using a pulsed spectrometer.

The measurements obtained using the pulse spectrometer provides a means of predicting the shape of the derivative of the Lorentzian lineshape that is generated by the CW spectrometer. In specific, the peaks in Figure 4 are expected to be separated by $\text{FWHM}/\sqrt{3} = 419 \text{ Hz}$, and the peaks should be centered on a transition frequency of 3.809 MHz. In order to achieve the maximum difference between the amplitude of the peaks, the amplitude of the magnetic field of the CW spectrometer should satisfy Equation (5). Using the measured values of T_1 and T_2^* , the optimal amplitude of the field is 185 mG, which is about 50 times smaller than the RF amplitude used in the pulsed experiments.

CW Spectrometer Measurements

Tyson describes the MO that serves as the CW spectrometer in this study.^[14] The MO is adjusted to have a conversion gain of approximately 10. The amplitude of oscillation is adjustable from about 1 mG up to 90 mG. The same 9-g HMT sample and probe coil used in the pulsed experiments is also used in the CW experiments. The frequency of the CW spectrometer is swept from 3.305 MHz to 3.311 MHz in 50-Hz steps. For a conversion gain of 10, the transient response of the envelope amplitude to changes in losses is about 20 ms, and so the

period of the lock-in reference is 50 ms, or equivalently, f_{lock} is 20 Hz. To obtain the derivative of the Lorentzian lineshape at the lock-in amplifier output, the dither frequency, f_d , is set to 50 Hz, so that it is more than a factor of ten smaller than the expected FWHM of 725 Hz. The time constant of the lock-in amplifier is set to 20 s to achieve a noise measurement bandwidth of 0.050 Hz. The CW spectrometer dwells 40 s at each sweep frequency to allow the lock-in amplifier output to reach steady-state. The amplitude of the CW spectrometer is set to 32 mG. Because the applied field is about one-sixth the optimal value of 185 mG predicted by Equation (5), the measurement signal-to-ratio will be reduced by about 15 dB. The time required to complete a single sweep using these parameters is 80 minutes.

The CW spectrometer is controlled using LabVIEW, and Figure 6 shows the graphical user interface. This interface allows the user to set the sweep range, the increment between sweep frequencies, the lock-in frequency, and the dither frequency. The conversion gain and amplitude of the CW oscillator are set by potentiometers on the MO. The lock-in parameters are set manually. The LabVIEW interface also provides the ability for the user to directly measure the conversion by varying the losses in the tuned circuit by a known amount.

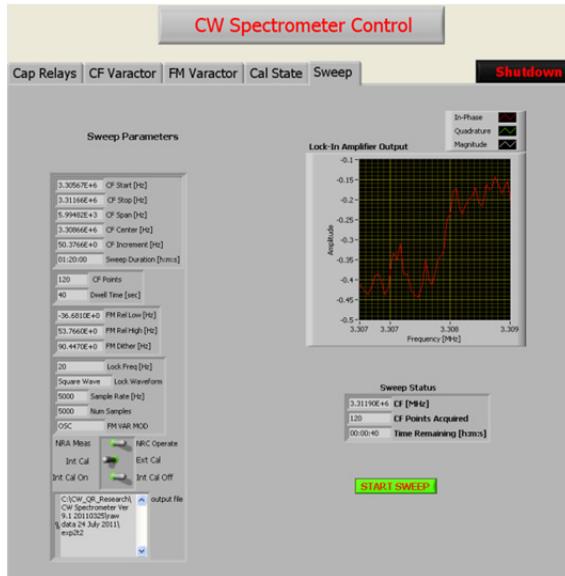


Figure 6: Graphical user interface for controlling the CW spectrometer.

Figure 7 shows the output of the lock-in amplifier. As expected, the shape of the curve matches the derivative of the Lorentzian lineshape shown in Figure 4. The characteristics of the data are consistent with the results predicted by the pulsed spectroscopy measurements. The distance between the peaks is 600 Hz, while the peaks are centered on 3.3079 MHz. The expected values from pulsed spectroscopy are 419 Hz and 3.3089 MHz, respectively. The differences are well within measurement error and most likely result from fluctuations in room

temperature. The data in Figure 7 verifies that the MO described by Tyson can detect a QR transition.

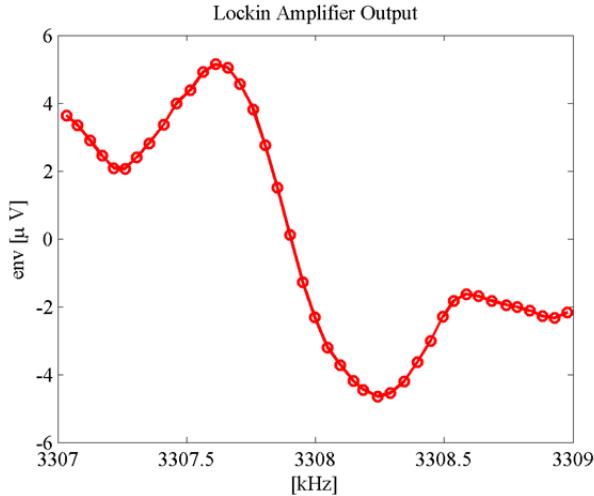


Figure 7: QR transition of HMT revealed by the CW spectrometer.

SUMMARY AND CONCLUSIONS

The goal of this study was to determine whether or not the MO described by Tyson could detect a QR transition. As a first step, pulsed spectroscopy was used to measure the transition frequency, FWHM, and spin-lattice relaxation time of the 9-g HMT sample used in the CW spectroscopy experiments. This information provides a benchmark for determining if the output of the CW spectrometer is correct. A comparison of the CW spectrum in Figure 7 with the parameters obtained from pulse spectroscopy, confirm that the MO design can detect QR transitions.

The results of this study also present a direct comparison of pulsed and CW QR spectroscopy. The 10-G magnetic field used in the pulsed spectrometer is about three hundred times larger than the 32-mG field used in the CW spectrometer. Because of the significantly lower field amplitude, instrumentation for the CW spectrometer is about 50 times less expensive than that for the pulsed experiment. The additional cost for the pulse spectrometer reflects the need for a high power RF amplifier and probe tuning network to accommodate the larger field amplitudes. The time required to perform the pulsed measurements leading to the data in Figure 6 is about one-half minute, while the CW spectrometer requires about 80 minutes to sweep through the QR transition. The ratio of the measurement time between CW and pulsed spectroscopy is about 300, or roughly the same as the ratio of RF field amplitudes. This study reveals that the choice between CW and pulsed spectroscopy presents a tradeoff between RF magnetic field amplitude, and hence spectrometer cost, and data acquisition time.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. J. Wilson would like to thank Dr. Jeffrey Schiano and Thomas Tyson for all their help during this process. Also, The Pennsylvania State University for giving me the opportunity to participate in the REU.

REFERENCES

- [1] H. G. Dehmelt, "Nuclear Quadrupole Resonance," *American Journal of Physics*, **22**(3)110-120(1954).
- [2] J. A. S. Smith. "Nuclear Resonance Quadrupole Spectroscopy," *Journal of Chemical Education*, **48**(1)39-49(1971).
- [3] J. A. S. Smith. "Nuclear Resonance Quadrupole Spectroscopy. Part 3 – Pulse Methods; Chemical Applications," *Journal of Chemical Education*, **48**(3)A147-A170(1971).
- [4] A. Abragam. Principles of Nuclear Magnetism. Oxford Science Publication, Clarendon Press, Oxford, 1961. (Reprinted in 1994).
- [5] R. R. Ernst and W.A. Anderson, "Application of Fourier Transform Spectroscopy to Magnetic Resonance," *The Review of Scientific Instruments*, **37**(1)93-102(1966).
- [6] A. Roberts, "Two New Methods for Detecting Nuclear Radio Frequency Absorption," *The Review of Scientific Instruments*, **18**(11):845-848 (1947).
- [7] B.V. Rollin, "Nuclear Paramagnetism," *Reports on Progress in Physics*, **12**(1)22-33(1949).
- [8] 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).
- [9] 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).
- [10] A. Gauzzi, J. Le Coche, 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).
- [11] R.T. McIver, "A Solid-State Marginal Oscillator for Pulsed Ion Cyclotron Resonance Spectroscopy," *The Review of Scientific Instruments*, **44**(8)1071-1074(1973).
- [12] 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).
- [13] J.A.W. Johnson, "Modulation Broadening of Second-Derivative Gaussian and Lorentzian Lines," *Journal of Magnetic Resonance*, **24**(1)21-25 (1976).
- [14] T. J. Tyson, "Marginal Oscillator Conversion Gain: Prediction, Simulation, and Experimental Measurements," M.S. thesis, Dept. Elect. Eng., Penn State Univ., State College, PA, 2011.
- [15] S. Alexander, A. Tzalmona, "Relaxation by Slow Motional Processes. Effect of Molecular Rotations in Pure Quadrupole Resonance," *Physical Review*, **138**(3A)A845-A855(1965).
- [16] J.L. Schiano and M. D. Ginsberg. "A Pulsed Spectrometer Designed for Feedback NQR." *Zeitschrift fur Naturforschung*, **55a**(1)61-66(2000).

OPTICAL AND ELECTRICAL CHARACTERIZATIONS OF FREE STANDING MICROELECTROMECHANICAL STRUCTURES

Bahareh Adami Ardestani,* Srinivas Tadigadapa,[#] and Hwall Min⁺

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
American River College
Sacramento, CA 95841

ABSTRACT

Ever since the advent of the atomic force microscope, free standing structures like cantilevers fabricated, using silicon micromachining techniques have been extensively investigated for chemical and biochemical sensing applications, due to their high sensitivity to surface loading. Typically these devices are used with optical readout since the use of position sensitive detectors makes it possible to measure nanometer level deflections. The aim of this paper is to compare the deflection sensitivity of microelectromechanical free standing structures using electrical and optical methods. Controlled deflection in the cantilever structure will be induced using electrostatic actuation force while the deflection will be quantified (i) electrically using ac capacitance measurement and (ii) optically using a position sensitive detector. The range of the resonance frequency of these cantilevers is between 50 kHz to 200 kHz. The project will involve understanding the electrically induced flexural motion of the cantilever using the two methods and critically evaluating the obtained sensitivity (signal to noise ratio) using the two methods.

INTRODUCTION

The history of biosensors started in 1962 with the development of enzyme electrodes by scientist Leland C. Clark, but biosensors using microelectromechanical systems (MEMS) [1, 2] technology have come into existence only in the last two decades. Microcantilevers and free standing plates are the most simplified MEMS based devices. Diverse applications of free

[#] Faculty Mentor

⁺ Graduate Mentor

standing structures in the field of sensors have been explored by many researchers. Several groups have also shown the possibility of using microcantilevers for the diagnosis of prostate cancer [3], myocardial infarction [4] and glucose monitoring [5]. Molecular diagnostic devices are getting smaller with the advancement of miniaturization technologies. As a result of spectacular movements in this technology, there is increasing interest in the field of biosensor and chemical sensors research on miniaturized platforms. Functionality of MEMS sensors is based on mechanical movements and deformations of their micromachined components, such as single-clamped suspended beams (cantilevers), double-clamped suspended beams (“bridges”), or suspended diaphragms. Cantilever structures are some of the simplest MEMS that can also be considered as basic building blocks for a variety of more complex MEMS devices. [6]. Its flexibility and versatility make it a popular component for a variety of applications in microsystems devices. Recent studies show that cantilever sensors have the ability of measuring the mass down to 10^{-21} gram [7, 8], which allows the possibility of single molecule detection. The general idea behind these types of sensors is that physical, chemical, or biological stimuli can affect mechanical characteristics of the micromechanical free standing structure in such a way that the resulting change can be measured using electronic, optical, or other means [6]. It is very important to implement the best method to determine this resulting mechanical change for enhancing chemical and biological sensors sensitivity. The higher the sensitivity of the cantilever, the better it will function. The comparative analysis of electrical and optical methods in this paper gives the fundamental understanding of advantages and disadvantages of both methods in determining the sensitivity of cantilevers and free standing plates.

METHOD AND MATERIALS

Materials and Fabrication

The starting materials used in this process were 4" p-type single-side polished <100> Si wafers (Young's modulus $E = 163$ GPa, Poisson ratio $\nu=0.23$, plane strain Young's modulus $E' = 172$ GPa) and 4" Pyrex 7740 borosilicate glass wafers. The fabrication process of the cantilever structure is shown in Figure 1. Olivier Pierron, member of the MEMS and Nanoscale Devices Group fabricated the device under test. A close up photograph of a free standing plate is shown in Figure 2.

Silicon (Si) is the dominant material for integrated circuit (IC) and microsystems devices fabrication. Silicon has an atomic number of fourteen 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 applications. In addition it is used in lasers, glasses, transistors, solar cells and many solid state electronic devices. Table I shows the material properties of silicon.

For the fabrication first, the top 4–5 μm of Si wafers were boron doped at a concentration higher than 7×10^{19} atoms· cm^{-3} . Then, the Si wafers were patterned

twice in order to define the gap ($\sim 1 \mu\text{m}$) between the structure and the bottom electrode on the wafer glass, and to define the structure themselves. Etching was done using RIE. After that, a metal (Cr/Au) layer was deposited and patterned on the glass wafers using a lift-off process. This metal layer acts as electrodes for electrostatic actuation of the structures. Then, both Si and glass wafers were diced in 24 mm units prior to bonding. Finally, the bonded structures were diced into individual chips (3.5 mm by 12 mm). Each chip was then released using an EDP anisotropic etch-stop process. The final dimensions of resulting plate within the chip is shown in Table II.

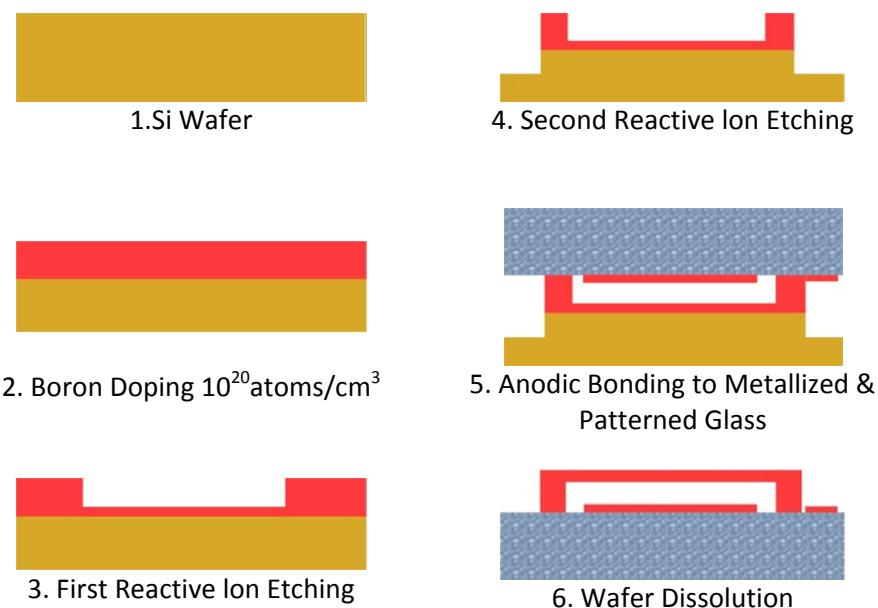


Figure 1: Fabrication process of Silicon Free Standing Plate

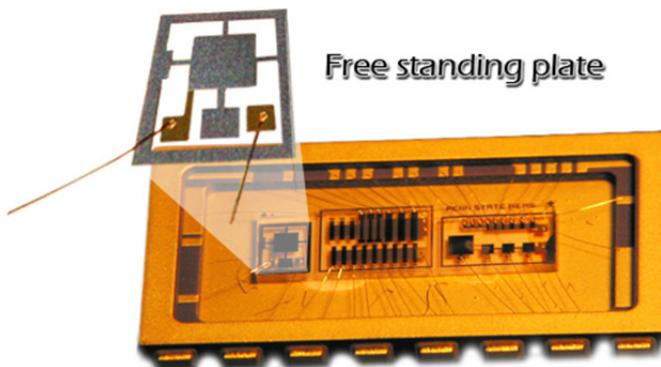


Figure 2: Photograph of packaged device and zoomed photograph of Free Standing Plate

Table I: Material and Physical Properties of Silicon

Properties of Silicon	
Name, symbol, number	silicon, Si, 14
Element category	metalloid
Group, period, block	14, 3, p
Standard atomic weight	28.0855(3)
Electron configuration	[Ne] 3s ² 3p ²
Phase	solid
Relative Permittivity	11.7
Liquid density at m.p.	2.57 g·cm ⁻³
Thermal Conductivity (solid)	1.412 W/cm-K
Thermal Conductivity (liquid)	4.3 W/cm-K
Specific Heat	0.70 J/g-K
Melting point	1687 K, 1414 °C, 2577 °F
Boiling Point	3538 K, 3265 °C, 5909 °F
Critical Temperature	5159 K
Vapor pressure	1e-7 Torr at 1050 °C 1e-5 Torr at 1250 °C
Molar heat capacity	20.00 J/mol-K

Table II: Geometric Parameters of Fabricated Free Standing Plate Shown in Figure 2

	Width	Length	Gap
Central square plate	1 mm	1 mm	1 μm
Beams connected the central plate	100 μm	300 μm	1 μm

Electrical Characterization

The capacitive method [9] is based on the principle that when the cantilever deflection takes place due to the adsorption of the analyte, the capacitance of a plane capacitor is changed. A free standing structure like cantilever or plate is a device that can act as a physical, chemical or biological sensor by detecting changes in cantilever bending which shows the sensor's sensitivity. The sensitivity is a quantity that reflects the sensor capability in transducing the input signal into an electrical output signal. The structure of the assessed cantilevers and freestanding suspended plates in this technique can be likened to parallel plate capacitors. A capacitor is constructed from two conducting plates separated by an insulating material called a dielectric. This insulating material can be paper, plastic, film, mica, air or vacuum. In our case the freestanding suspended plate is connected to its substrate through a gold electrode that is fabricated underneath the plate from one end, but the other side is free and the space between the substrate and the gap is filled with air.

To measure capacitance and deflection, a small oscillating voltage with a DC bias voltage is applied across the plate causing the freestanding suspended plate to deflect. This deflection causes a change in capacitance that can be measured. Special consideration should be made to the applied voltage to avoid ‘pull-in,’ where the movable cantilever collapses on the lower plate when the bias voltage goes beyond certain limit. The electrostatic force which is generated between the top cantilever and bottom gold surface make the cantilever bend and reduces the separation dx . It is important to note that the gravitational force of microcantilevers can be neglected in this experiment due to their small sizes and mass.

At small voltages the electrostatic and elastic forces are in equilibrium, however as voltage increases the electrostatic force increases much more dramatically. Eventually, the electrostatic force will cause the “pull-in” event, in which the plates will snap together.

Capacitance between two parallel plates can be measured using:

$$C = \varepsilon \frac{A}{x} \quad (1)$$

where ε is the permittivity of air, A is the area of the electrode, and x is the separation between the two electrodes. The partial derivative of Eq. (1) with respect to voltage is obtained using the chain rule

$$\frac{\partial C}{\partial V} = \frac{\partial C}{\partial x} \frac{\partial x}{\partial V} = -\frac{\varepsilon A}{x^2} \frac{\partial x}{\partial V} \quad (2)$$

which relates the change in capacitance to a change in voltage. The deflection is a function of voltage and can be determined from the force equilibrium of the elastic and electrostatic forces:

$$F_s = F_{el} \quad (3)$$

$$\rightarrow kx = \frac{1}{2} \frac{\varepsilon A}{x^2} V^2 \quad (4)$$

The deflection of the free standing plate can be solved using Eq. (4); however, it is beyond the scope of this work. The sensitivity of the capacitance with respect to voltage is the figure of merit of the free standing plate for use in sensing applications.

The deflection of the beam and the “pull-in” voltage can be measured by the increase in capacitance between the air gap. For performing capacitance measurement an Agilent 4294A Precision Impedance Analyzer with frequency range 40 Hz to 110 MHz is used. For each experiment, two of the chip’s pins are connected to the device through impedance measurement ports. One pin is connected to either free standing plate or cantilevers and the other one is attached to the common. The range of resonant frequency of assessed plates and cantilevers is from 50 kHz to 200 kHz, so for the capacitance measurement the

frequency of the device is fixed on 20 kHz, a frequency far below resonant frequency to avoid sharply increase in mechanical vibration amplitude. The oscillator output signal (OSC) is set on 200 mV and the Bias monitor on volt. For each cantilevers or plates two series of experiments are implemented. One for noise measurement which will obtain data without applying any voltage for actuation of the free standing components, and the other one is done for capacitance measurement. The latter is performed by applying voltage and constantly increase it by small amount until a sudden rise is seen in capacitance graph on the device monitor (see Figure 4).

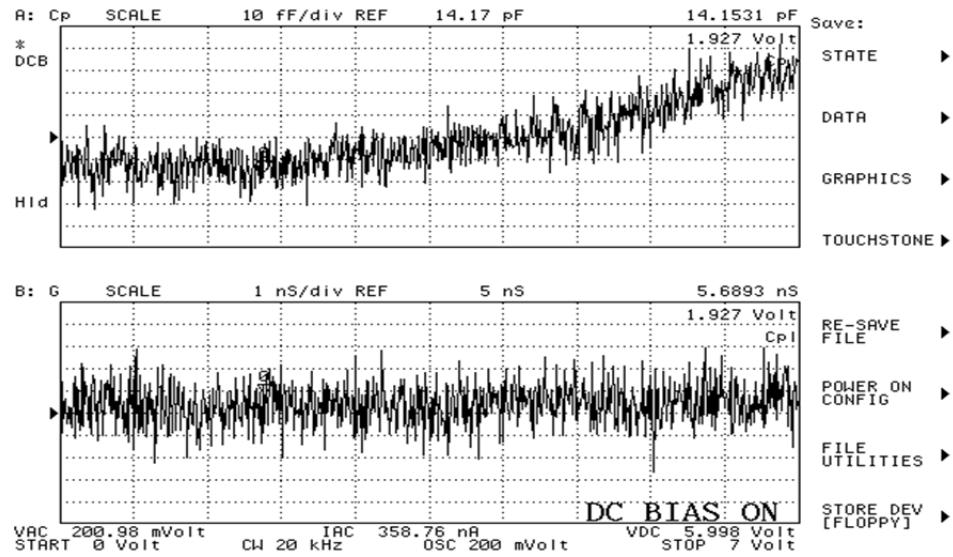


Figure 4: Output of Agilent 4294A Precision Impedance Analyzer showing onset of pull-in voltage

Optical Characterization

The optical beam deflection technique introduced by Meyer and Amer [10] is a very reliable and simple detection method. (See Figure 5) Because of this, the optical beam deflection method is widely used in laboratory and commercially available AFMs. For implementing the optical method, a laser beam is directed and aimed at the top of suspended plate structure (See Figure 6). The laser beam, shot by a laser source, passes through an attenuating filter to reduce reflection in the photo sensor. After passing through the first mirror in the setup and the focus, the beam is directed to the freestanding suspended plate inside the chip which is set in the chamber. Next, the reflected beam is captured and forwarded by another mirror to the photo sensor. The reflected laser beam is directed to a position sensitive detector (PSD) that is based on the quadrant photodiode of four cells. A, B, C, and D. Each of the cells is coupled to the input of a separate transimpedance amplifier the output voltages of which, V_A , V_B , V_C , and V_D , are proportional to the

illumination of the respective quadrant. The normalized differential output, $V_{\text{out}} = [(V_A + V_C) - (V_B + V_D)/(V_A + V_B + V_C + V_D)]$, depends linearly on the vertical displacement of the weighted center of the light spot projected by the cantilever [6]. A position sensitive detector is connected to the computer and is controlled by specific software interfaces created on LabVIEW by Sangmin Jeon from the Nanoscale Science & Device Group. This user interface can measure frequency spectrum, deflection and room temperature and record data for analysis.

The experimental procedure of the optical method consists of two steps as well. First, noise measurement which can be done by running the setup and shingling the laser beam onto the tip of free standing cantilevers or plates, without applying any voltage to them. And second step is measuring the deflection of the nanoscale components by applying voltage to them. The applied voltage starts from 0 and goes up to 5 volts, a voltage below pull-in voltage. A triangular wave with 5 Vpp and offset of 2.5 V fits to this experiment perfectly.

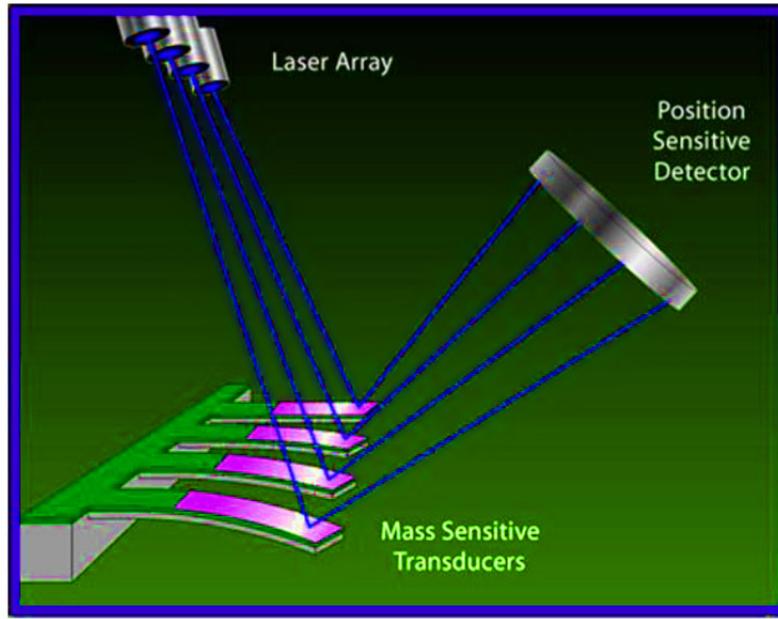


Figure 5: Schematic of the optical characterization method

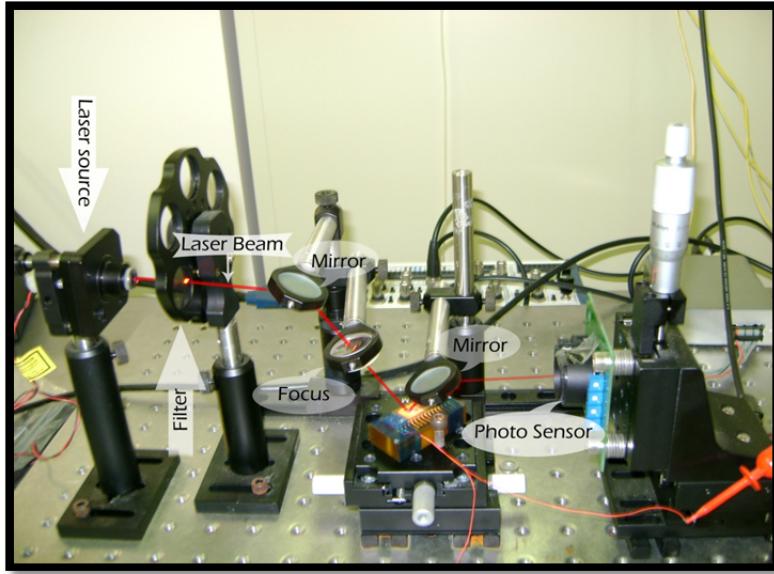


Figure 6: Experimental setup of the optical characterization method

RESULTS AND DISCUSSION

In this project, about twenty silicon cantilever beams and bridges and one free standing plate sitting on a silicon wafer were assessed. The following graphs (see Figures 7 and 8) represent the electrical and optical measurement results for the free standing plate shown in Figure 2. Table 3 shows the comparison of the Electrical and Optical Characterization Methods. It is important to note that the behavior of that free standing plate is fundamentally similar to those of cantilevers.

These experiments both electrical and optical are similar to using the plate or cantilever as a mass sensor. When a small tiny substance is placed on top of cantilever, it causes it to bend, exactly parallel to what happens to it by applying voltage. (See Figure 5) This voltage is just a control parameter. Putting a very small mass on top of cantilever is a very complicated experiment. Instead something equivalent to that is implemented to the measure the sensitivity of the plate. The application of the voltage allows a nice independent reversible control over the system.

By implementing the experiments over and over it becomes evident that both methods have some advantages and some disadvantages. Optical method which most extensively are used for measurements of freestanding suspended plate deflections has excellent readout efficiency in the case of free standing structures including cantilevers and bridges with reflecting areas of at least a $10 \times 10 \mu\text{m}^2$. An optical readout technique is somehow inefficient when applied to nanostructures. The shortcomings of some optical techniques, in particular the optical deflection method, are related to loss of intensity and directionality of optical beams scattered by nanosize cantilevers. One of the main advantages of

capacitance readout is that it can be used in integrated MEMS devices that are fully compliant with standard CMOS technology. Capacitance readout, on the other hand, suffers from interference with variations in the dielectric constant of the medium. While differential schemes may eliminate this interference, electrically conductive media, such as electrolytes, make capacitance readout more challenging [6].

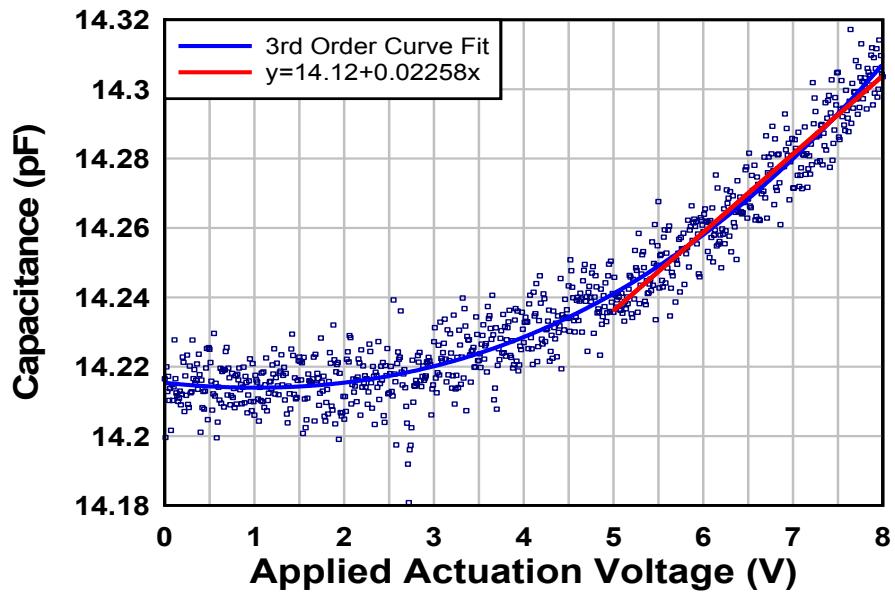


Figure 7: Capacitance versus voltage bias

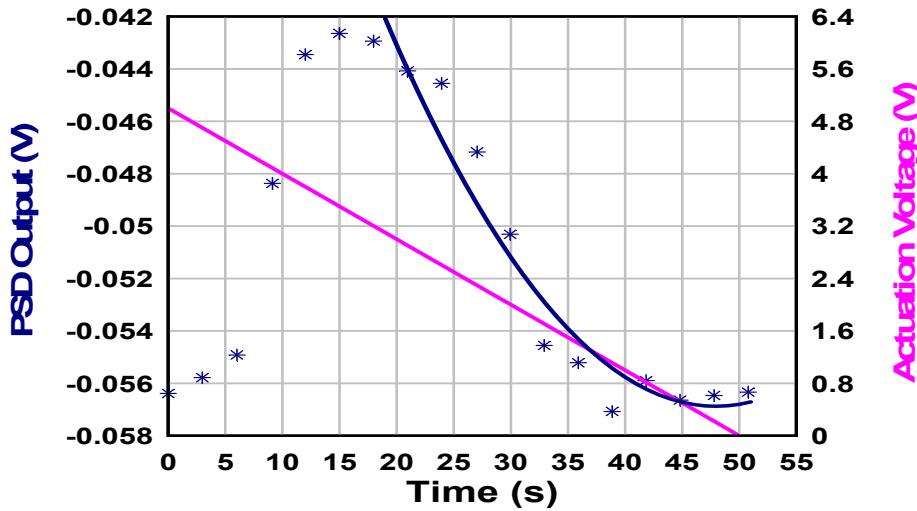


Figure 8: PSD voltage output and actuation voltage output of optical characterization method

Table III: Comparison of the Electrical and Optical Characterization Methods

Electrical Noise	Electrical Sensitivity	Electrical Resolution
0.006795 pF	0.0225825540284 pF/V	0.3 V
Optical Noise	Optical Sensitivity	Optical Resolution
33.22115 μ V	-0.864 mV/V	0.038 V

CONCLUSION

Many experiments were performed using electrical and optical methods. Based on the results that are obtained in these experiments and by comparing the signal to noise ratio of both methods, the optical detection technique is ~ 13 times more sensitive. This result can be fully used for enhancing biological and chemical sensors sensitivity.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant, No. EEC-0755081 and ATM-0734083. Furthermore, B. Adami Ardestani would like to thank Dr. Srinivas Tadigadapa for giving her the opportunity to do this research, and for his help and support. Thanks also to Hwall Min for mentoring her and Matt Chang for helping her with the optical setup. Finally, she would like to specially acknowledge Kiron Mateti and Gokhan Hatipoglu for helping her get accustomed to the lab and assisting her numerous times throughout her research.

REFERENCES

- [1] Grayson, A.C.R., Shawgo, R.S., Johnson, A.M., Flynn, N.T., Li, Y., Cima, M.J. & Langer, R. A BioMEMS Review: "MEMS technology for physiologically integrated devices," *Proc. IEEE*, **92**(1), 6-21(2004).
- [2] Polla, D.L., Erdman, E., Robbins, W.P., Markus, D.T., Diaz, J.D., Rinz, R., Nam, Y. & Brickner, H.T. "Microdevices in Medicine," *Ann. Rev. Biomed. Eng.*, **2**, 551-76 (2000).
- [3] Wu, G.H., Datar, R.H., Hansen, K.M., Thundat, T., Cote, R.J. & Majumdar, A. "Bioassay of prostate-specific antigen (PSA) using microcantilever," *Nat. Biotechnol.*, **19**, 856-60 (2001).
- [4] Arntz, Y., Seelig, J.D., Lang, H.P., Zhang, J., Hunziker, P., Ramseyer, J.P., Meyer, E., Hegner, M. & Gerber, C. "Label-free protein assay based on a nanomechanical cantilever array," *Nanotechnology*, **14**, 86-90 (2003).
- [5] Subramanian, A., Oden, P.I., Kennel, S.J., Jacobson, K.B., Warmack, R.J., Thundat, T., Doktycz, M.J. "Glucose biosensing using an enzyme-coated microcantilever," *Appl. Phys. Lett.*, **81**, 385-87 (2002).
- [6] N. V. Lavrik, M. J. Sepaniak, P. G. Datskos, "Cantilever transducers as a platform for chemical and biological sensors," *Journal of Applied Physics*, **75**(7), 2229-54 (2004).
- [7] B. Ilic, H. G. Craighead, S. Pavlov, W. Senaratne, C. Ober, and P. Neuzil, "Attogram detection using nanoelectromechanical oscillators," *Journal of Applied Physics*, **95**(10), 3694-3703 (2004).

- [8] K. L. Ekinci, Y. T. Yang, and M. L. Roukes, "MultiMate limits to inertial mass sensing based upon nanoelectromechanical systems," *Journal of Applied Physics*, **95**(5), 2682-2689 (2004).
- [9] Blanc, N., Brugger, J., Rooij, N.F.D. & Durig, U. "Scanning Force Microscopy in the Dynamic Mode Using Microfabricated Capacitive Sensors," *J Vac. Sci. Technol.*, B, **14**(2), 901-05(1996).
- [10] G. Meyer and N. M. Amer, "Novel optical approach to atomic force microscopy," *Applied Physics Letters* , **53**(12), 2400-2402 (1988).

GROWTH AND OPTICAL ANALYSIS OF BRANCHING SILICON WIRE ARRAYS

Christina DiMarino,* Chito Kendrick,[†] and Joan Redwing[‡]

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

*Undergraduate student of
School of Engineering
James Madison University
Harrisonburg, VA 22801

ABSTRACT

This study analyzed the optical properties of branched, single-crystalline silicon microwire arrays to determine their potential usage in photovoltaic cells. These arrays are preferable to silicon wafer based photovoltaic cells because they use less material and have lower purification requirements, thereby making them more cost-effective. Furthermore, the radial geometry of the microwires allows for sufficient light harvesting and carrier extraction, two key factors in solar cell efficiency. The arrays were grown using the chemical vapor deposition (CVD) method and vapor-liquid-solid (VLS) growth mechanism. After the growth of the primary wires, the samples go through a post-growth process and are then returned to the growth chamber where the wires are branched. Using a spectrophotometer, it was determined that the branching reduced the reflectivity of the microwires for wavelengths in the visible light region. It was also concluded that increasing the lengths of the primary wires reduces the overall reflectivity of the arrays. Further investigation into the solar cell efficiency of these branched arrays is needed to determine their application in photovoltaic devices.

INTRODUCTION

Silicon Wafer versus Silicon Microwires

Silicon photovoltaic cells fabricated from monocrystalline silicon wafers currently make up a substantial amount of the solar cell market [1]. However,

[†] Postdoc Mentor

[‡] Faculty Mentor

high manufacturing costs remain a significant issue. Replacing the wafer with single-crystalline silicon microwire arrays offers a potential solution because the wire arrays use less material than the present wafer based photovoltaic cells, and, consequently, are predicted to be more economical. Silicon wire arrays can be grown using chemical vapor deposition (CVD) and the vapor-liquid-solid (VLS) growth mechanism; however this can lead to the contamination of the silicon due to the use of a metal catalyst to control the wire growth. Metal continuation leads to short minority carrier diffusion lengths, which for microwires are $1\text{-}10 \mu\text{m}$ long, compared to photovoltaic grade silicon wafers that have minority carrier diffusion lengths up to hundreds of microns. However, Kayes et al. reported that these silicon wire photovoltaics actually allow for the photon absorption and minority carrier collection process to be decoupled, as shown in Figure 1 [2]. Therefore, a sufficiently long silicon wire could completely absorb the incident photons [3] and, if the wire radius is equal to the length of the minority carrier diffusion length, carrier collection should still occur before the carriers recombine. This is significant because light harvesting and electron collection are two essential factors for effective solar cells [4,5]. It is unexpected that the microwire solar cells will surpass the efficiencies of present silicon wafer based solar cells; however, they should be able to outperform present low-cost photovoltaics [6]. Additionally the wire arrays have the opportunity to be transferred to low-cost substrates such as glass, making them even more cost-effective [6].

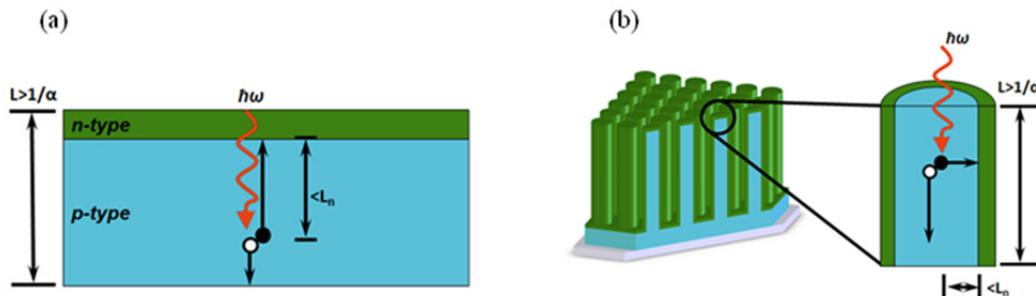


Figure 1: Schematics illustrating pn junctions for (a) planar solar cell, and (b) wire array solar cell. The electron hole pair generated by a photon with energy $\hbar\omega$ is shown, where L_n represents the carrier diffusion lengths [2].

Branching

It is believed that growing nanowire branches along the sides of the primary microwires could improve the efficiency of low-cost solar cells. For dye-sensitized solar cells (DSSCs), the increased surface area from the branching will be beneficial. A high surface area is essential in DSSCs because it allows for a greater volume of dye to be in contact with the electrodes (silicon branched wires). Therefore, branching ensures sufficient dye loading for the absorption of photons and provides more conduction pathways for faster electron transport [7,8]. A recent study by Ko et al. determined that branched ZnO nanowire DSSCs have higher efficiencies than ordinary ZnO nanowire DSSCs. The branched

nanowires had an efficiency of 2.51%, while the unbranched nanowires had an efficiency of only 0.71% [8].

The application of branched microwire arrays in solid-state solar cells will also be investigated. Whereas in DSSCs the dye is responsible for photon absorption, in a solid-state device the semiconductor absorbs the photons and surface texturing is often required to reduce the reflection of photons from the surface of the wafer. Consequently, this study analyzed the reflectivity of the branched microwires to determine if branching could reduce the surface reflectance. This paper will discuss the optical results of branched wire arrays that were grown using the chemical vapor deposition (CVD) method in conjunction with the vapor-liquid-solid (VLS) growth mechanism with gold (Au) as the metal catalyst. If the branched wire arrays prove to have a lower reflectance, their use as a cheaper alternative to crystalline silicon wafers for both solid-state and DSSCs photovoltaic cells will be further assessed.

Superhydrophobic Analysis

These branched wire arrays will also be tested to determine if they can be made superhydrophobic by coating with the appropriate organic molecules. The superhydrophobic property will be established by using a goniometer to measure the contact angles of water droplets on the surfaces of the samples. If the interior contact angle is greater than 150°, then the sample is considered superhydrophobic [9]. If the arrays are in fact superhydrophobic, then they may find applications such as coatings for airplane wings. Flying through clouds, freezing rain, fronts, and low-pressure areas often results in the buildup of ice on the aircraft. Ice on the leading edge of the wings increases drag and reduces lift, which could cause the aircraft to stall [10]. Working with Dr. Mallouk and Liz Seibel in the Chemistry department, the pragmatism of this application will be evaluated. Superhydrophobic surfaces are also good for self-cleaning, which would be beneficial in solar cells.

EXPERIMENTAL METHODS

Primary Growth Process

In the CVD and VLS methodologies, silicon wafers containing gold patterns are placed in a growth chamber where the substrate is heated to 1050 °C, causing liquid droplets of Au–Si to form (Figure 2a). Vaporized silicon tetrachloride (SiCl_4) is then allowed to flow into the chamber. The liquid Au–Si alloy acts as a preferred nucleation site for the Si atoms, and as the alloy supersaturates with silicon it precipitates a silicon wire (Figure 2b) [11]. The patterned wafer used in this study was made up of a square array with a pore diameter of 3 μm , 10- μm pitch, and an Au layer thickness of 200 nm.

Post-Growth and Branching Processes

After the primary wire growth, the wires go through a post-growth process. First, a 10:1 buffered oxide etchant (BOE) was used to remove the patterning and

native oxide from the substrate and wire array surfaces. The remaining gold from the first growth was then removed from the wires using Transene TFA gold etchant. Next, the samples were sputtered with a layer of gold, at an argon pressure of 1 mTorr, which acts as the catalyst for the branching growth. The gold thicknesses used were 20 nm, 50 nm, and 100 nm, with the 20 nm Au layer being employed the most. Etching prior to sputtering ensures that the gold layer has uniform coverage (Figure 2c). A process was also developed to prevent larger branches from growing on the tops of the wires. In this process, a photoresist layer that only leaves the tips of the wires exposed is first placed on the samples (Figure 2d). The wires are then placed in the gold etch, which only removes the gold that is exposed at the tips of the wires. The photoresist is then removed with acetone, followed by a piranha etch (Figure 2e). Due to time constraints, this process for removing the sputtered Au layer from the tops of the wires was not performed for every sample. After this post-growth process, another growth is performed to produce the branched nanowires (Figure 2f).

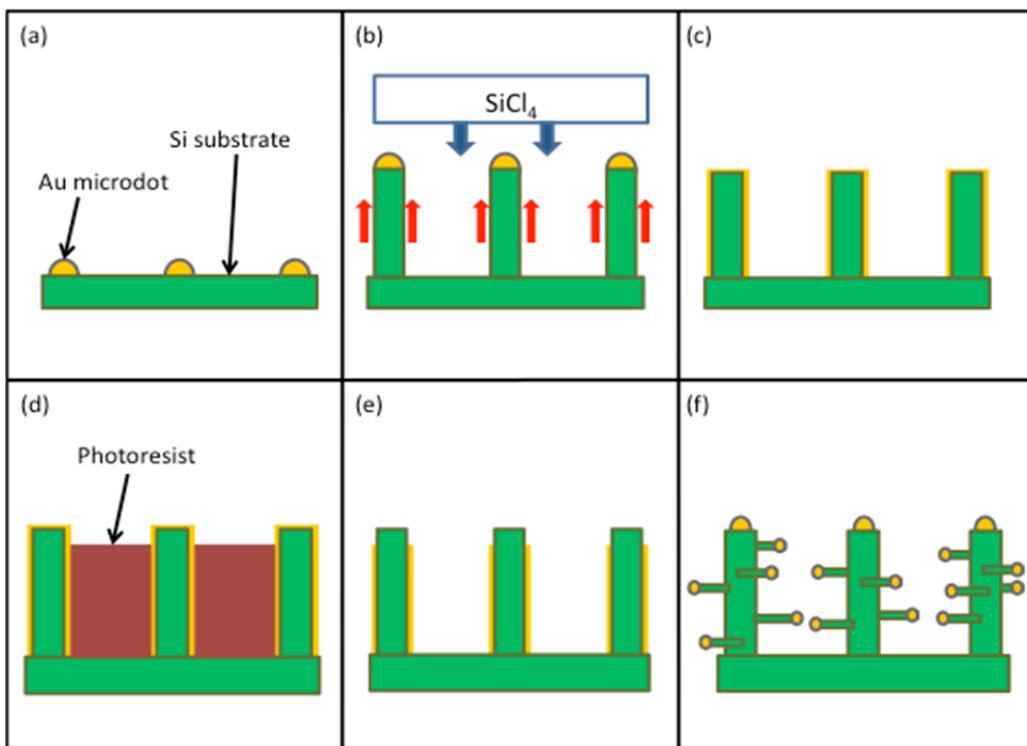


Figure 2: Wire Growth Process, (a) patterned substrate heated to 1050 °C, (b) wires growing from CVD and VLS methods, (c) sample sputtered with Au, (d) photoresist layer applied, (e) gold tips and photoresist removed, (f) branches grown along the primary wires.

Variables

Several variables were analyzed to determine their impacts on the optical properties of the samples. These variables were: primary wire length, primary

wire diameter, branch length, and branch diameter. In this study, the primary wires had lengths in between 20 μm and 200 μm and diameters of approximately 2 μm . The branch lengths and diameters had ranges from 1 μm to 3 μm and 0.1 μm to 2 μm , respectively. When examining one of these variables, all other factors were held constant. For instance, to determine the affect of the primary wire length on the optical properties of the arrays, the optical reflection values of wires with different primary wire lengths, but similar branch diameters and lengths, were compared.

Nanowire Growth Process

For comparing the affects of the primary wire diameter on the optical properties of the arrays, nanowires were also branched. For these growths, the samples were not removed from the growth chamber after the synthesis of the primary nanowires. Instead, after the growth of the primary wires, the flow of SiCl_4 to the reactor was stopped and the furnace cooled. This allowed for the gold to diffuse down the sides of the wires. After the furnace had time to cool, it was once again heated to 1050 °C. Upon reaching the growth temperature and allowing the sample to stabilize for 10 minutes, the flow of SiCl_4 through the chamber was resumed, restarting the growth process.

Optical Testing

The optical testing was done using a Perkin-Elmer Lambda 950 UV-Vis-NIR Spectrophotometer, which measures the percentage of incident light that is reflected from the sample. This data was acquired using the Labsphere DRA-CA-301 [12] and outputted to a plot of reflection versus wavelength. The wavelength range for these optical measurements was 300 to 1300 nm, which spans from the ultraviolet to the infrared regions. In particular, the reflectivity in the 400 to 800 nm range was analyzed because absorption of wavelengths in the visible light region is typical for solar cells and the majority of the solar spectrum is concentrated within this range of wavelengths. The transmittance was assumed to be negligible because the silicon substrates were 0.5 mm thick. Therefore all of the light that was not reflected by the samples was presumed to be absorbed. The graphs outputted by the spectrophotometer were used to evaluate the reflectivity of various samples. Several branched wire samples were compared to a set of unbranched primary microwires as well as a planar silicon sample, which served as controls. For the branched samples, the lengths and diameters of the primary wires and branches were compared to see which combination would yield the lowest reflectivity. The same wafer pattern was used for all of the microwire samples, thus the densities of the arrays and diameters of the primary wires were constant throughout this study. The nanowires, however, were grown on an unpatterned substrate, which produced denser arrays. The gold was not etched away from the samples after growth, which may have impacted the optical measurements.

Superhydrophobic Testing

Prior to performing the superhydrophobic testing, the samples needed to be prepared. First, the oxide and gold were removed using BOE and a gold etch. The samples were then left for 2 hours to allow the native oxide to re-grow. After the oxide re-growth, the samples were coated with a self-assembled monolayer (SAM). The SAM used was (3,3,3-Trifluoropropyl)methyldimethoxysilane. The superhydrophobic testing was then performed. This testing involved placing a droplet of water on the samples with a syringe and measuring the advancing and receding contact angles. A camera captured images of the droplets, which were used to measure the contact angles.

RESULTS

Silicon Microwires

First, using the Perkin-Elmer Lambda 950 UV-Vis-NIR Spectrophotometer, the light reflection values of ordinary microwire arrays and planar silicon were compared (Figure 3a). From these measurements, it was determined that the wires had lower reflectance values than the planar silicon. The affect of the primary wire length on the reflectance of the samples was then investigated.

The primary wire lengths of the microwire arrays tested in this study ranged from approximately 20 μm to 200 μm . It was observed that, overall, the reflectivity decreased as the primary wire length increased. This is demonstrated in Figure 3(a), which compares the optical reflection values of arrays with wire lengths of 36.2 μm , 52.6 μm , and 95.8 μm . Of these three array samples, the 95.8- μm -long wires had the lowest reflectance. The cross-sections of the plotted microwire samples are shown in Figures 3(b), (c), and (d). These images, taken using a scanning electron microscope (SEM), illustrate the increase in the primary wire lengths of the samples. In the spectrophotometer measurements, extensive noise was present in the infrared region because the gain was set to a high value in order to reduce the spike that occurred at 860 nm, which is the wavelength at which the spectrophotometer switches detectors.

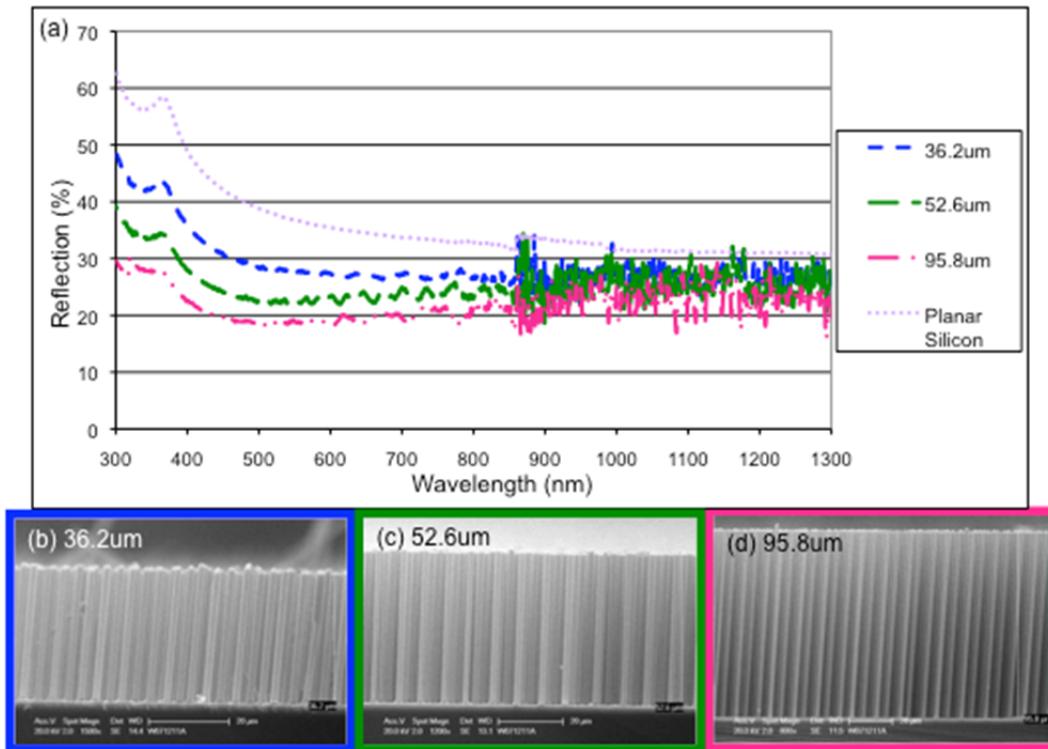


Figure 3: (a) Plot of unbranched microwires of various lengths and planar silicon. The cross-sections of the 36.2 μm , 52.6 μm , and 95.8 μm microwire samples are shown in (b), (c), and (d), respectively.

Branched versus Unbranched Microwires

Next, the reflectivity of branched microwires was compared to unbranched samples. For this analysis, the primary wire lengths of the samples were held constant. Ideally, the optical testing would have been conducted on the microwires prior to and after branching to ensure that the branching was the only variable. While this was possible for the 50-nm-Au and 100-nm-Au samples, discussed later, time restraints made this method impractical. Instead, microwires with similar characteristics, such as primary wire length and Au thickness, were compared. Figure 4(a) shows a plot of microwires with two primary wire length ranges: approximately 30–35 μm and 45–50 μm . Each range includes both an unbranched and branched sample. Both branched samples were sputtered with 20 nm of Au prior to branching, so they have similar branch diameters. Shown in Figure 4(a) is the reflection spectrum from the 20-nm coatings. For the 45–50 μm samples, the branching caused the optical reflection to decrease significantly across the wavelength range of 300 nm to 500 nm. However, for wavelengths greater than 550 nm, the reflectance of the 44.7 μm branched arrays, unlike other branched samples, increased considerably. This could be due to the long branches extending from the tops of the primary wires, which the other samples lacked. For the 30–35- μm samples, the optical reflection, overall, decreased as a result of the

branching. SEM images for each of the samples plotted in Figure 4(a) are shown in Figures 4(b), (c), (d), and (e). As shown by Figures 4(c) and (e), the branches grew hexagonally along the sides of the wires. This was expected because the single crystalline silicon pillars have a hexagonal crystal structure.

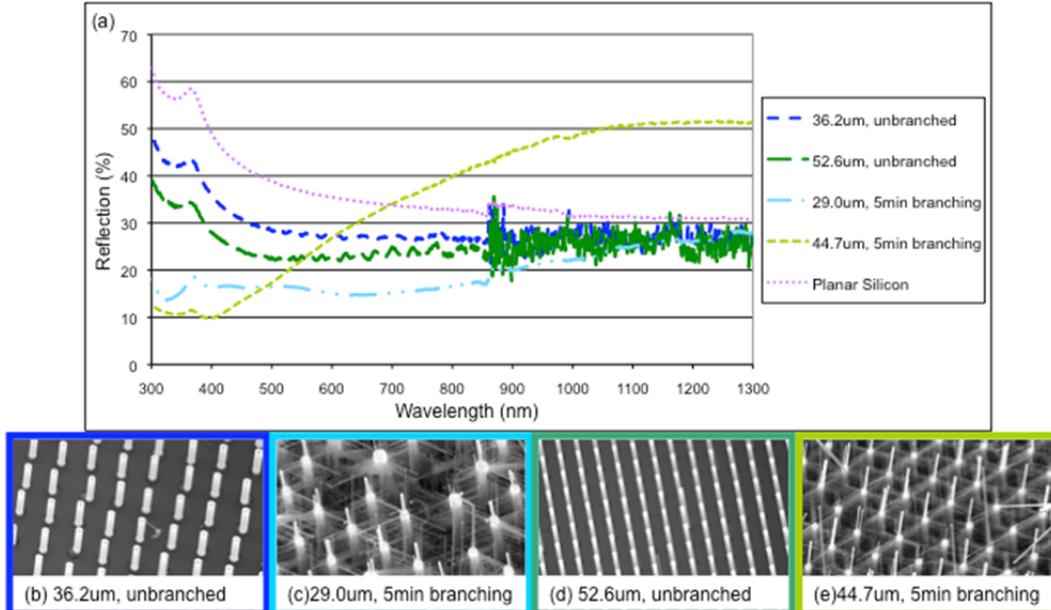


Figure 4: With the primary wire lengths held relatively constant, the affects of branching (with 20 nm of Au) on reflectivity are shown in (a). Planar silicon is also plotted as a comparison. 15° tilted SEM images of the unbranched and branched microwires in the 30 μm to 35 μm primary wire length range are shown in (b) and (c). 15° tilted SEM images of the unbranched and branched microwires in the 45 μm to 55 μm primary wire length range are shown in (d) and (e).

Nanowires

Optical measurements were also performed on branched nanowires to determine the impact of the primary wire diameter on the reflectivity of the arrays. After several attempts, branched nanowires were finally achieved. The initial problem was that the length of time needed to allow the gold to diffuse down the wire was assumed to be considerably shorter than actually required. Allowing the primary nanowires to slowly cool in the furnace for one hour granted the gold enough time to diffuse down the sides of the wires. After the one hour, the furnace was heated back up to 1050 °C and the flow of SiCl₄ was run through the growth chamber. This branching method was attempted for microwires; however, the gold catalyst was too stable and consequently did not diffuse down the sides of the wires. As shown in Figure 5, the branched nanowires have a lower reflectance in the 300 nm to 400 nm range compared to the 20 nm Au branched microwires. However, for wavelengths greater than 400 nm the reflectance increases drastically to ~60%. The long, thin branches

extending from the tops of the 44.7 μm branched samples may resemble the branched nanowires, which would explain the similar reflection of infrared wavelengths. These results indicate that branched nanowires have similar reflectance to the branched microwires, but only over a small wavelength range.

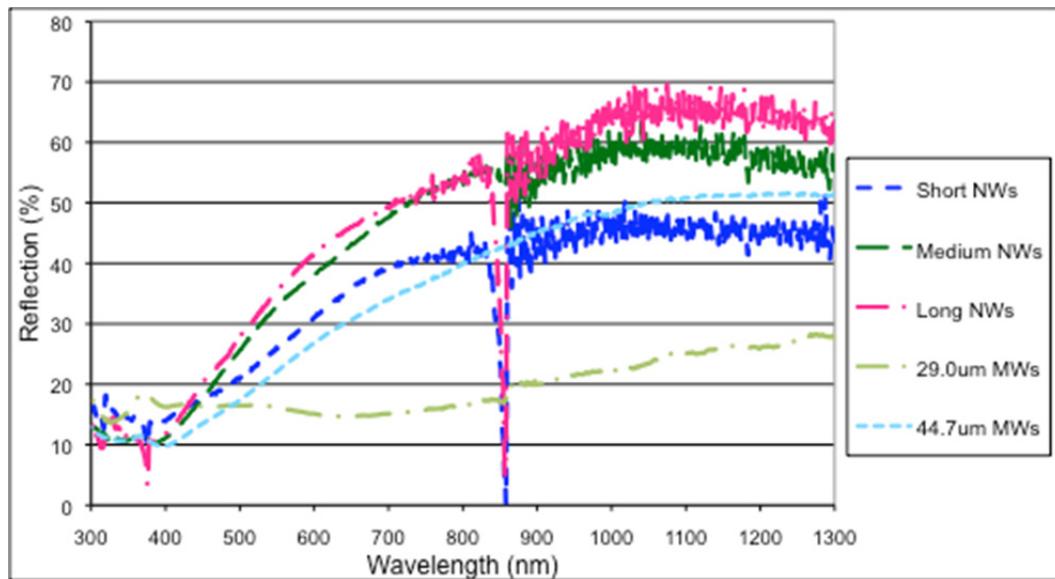


Figure 5: Plot of the optical reflectance versus wavelength for short, medium, and long branched nanowires (NWs) and 29.0- μm - and 44.7- μm -long branched microwires (MWs).

Branch Diameter

Three different branch diameters were analyzed. This was achieved by varying the Au thickness that was sputtered onto the primary wires. Thicknesses of 20 nm, 50 nm, and 100 nm were used. For the branching growth with the 50 nm and 100 nm Au thickness coatings, it became evident that the branching growth time needed to be increased by a factor of 5 or more in order to obtain branch lengths comparable to those of the 20 nm Au coated wires, which were typically branched for a period of 1 minute.

The branch diameters for the 20 nm, 50 nm, and 100 nm Au thicknesses were approximately 0.1 μm , 0.3 μm , and 1.7 μm , respectively. Figure 6(a) shows a plot of unbranched and 50 nm Au branched wires of various primary wire lengths. Unlike the results presented in Figure 4(a), these optical measurements are from the same set of samples, meaning the reflectivity of these arrays was tested before and after branching. As shown in Figure 6(a), although branching lowered the reflectivity of the 73.6 μm and 112.0 μm wires, it actually increased the reflectivity of the 200 μm wires. Also, the reflectivity between the pre- and post-branching wires decreased as the primary wire length increased. These results suggest that as the primary wire length increases branching becomes less beneficial, and past a critical primary wire length branching actually results in a

more reflective surface. This is because, as shown in Figures 3(a), 4(a), and 6(a), long unbranched primary wires have lower reflectance. The results shown in Figure 6(a) suggest that the critical primary wire length is in between 112.0 μm and 200 μm .

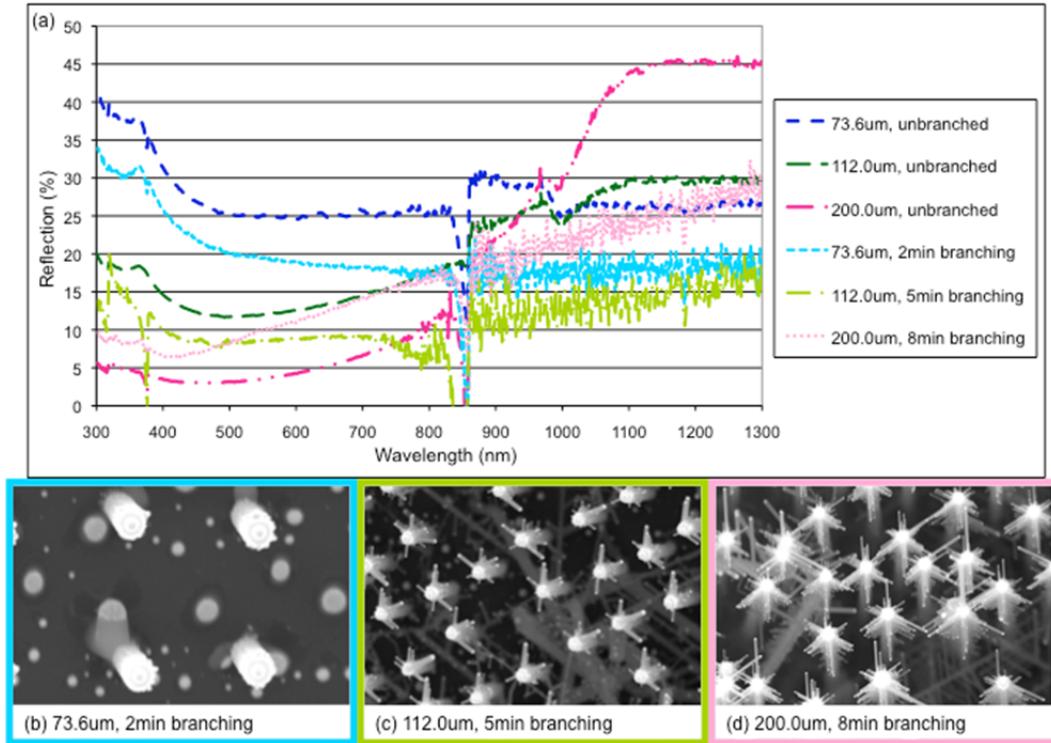


Figure 6: The optical reflection of the unbranched and 50 nm Au branched samples are plotted in (a). The 15° tilted SEM images of these branched microwires can be seen in (b), (c), and (d).

For the 100 nm Au branched microwires, the diameters were approximately 1.7 μm . For this branch diameter, the branching resulted in a lower reflectance compared to the unbranched samples, as shown in Figure 7(a). The lower reflectivity of the branched samples could be attributed to the primary wire lengths of 19.3 μm , 32.2 μm , and 66.0 μm being less than the critical value. Unlike the previous samples discussed, the branch lengths of these samples do not increase with increasing primary wire length. The shortest branches, less than 0.5 μm long, were grown on the 32.3- μm primary wires, which produced the most reflective branched sample. The longest branches, roughly 3.0 μm in length, were grown on the 19.3 μm primary wires, which had the second lowest reflective branched sample. The branched 68.0 μm wires had the lowest reflectivity and branches that were roughly 1.5 μm long. These results suggest that the optimal branch length for high absorption is in between 1.5 μm and 3.0 μm . Furthermore, as shown by Figure 7(a), regardless of branch length, the increase in primary wire

length leads to a decrease in reflectance.

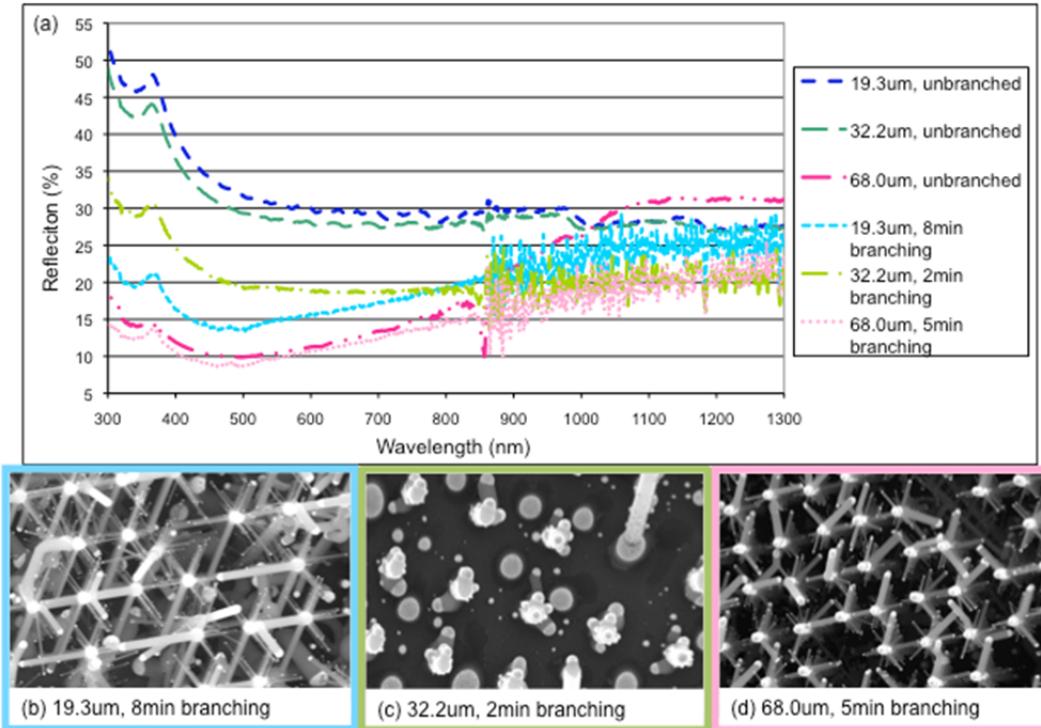


Figure 7: The percent reflection values of unbranched and 100-nm-Au-branched microwires are plotted in (a). Aerial images of the branched microwires are shown in (b), (c), and (d).

Unfortunately, it is difficult to conclude whether or not the branch diameter had a significant affect on the reflectivity of the wires. This is because there were no 20 nm, 50 nm, and 100 nm Au branched samples with similar primary wire and branch lengths. This was largely due to the fact that the branching rate unexpectedly decreased with increased Au thickness. The 20 nm Au branch time was roughly 1 minute, while the 50 nm and 100 nm Au thicknesses required a branch time of at least 5 minutes to achieve similar branch lengths. It was not anticipated that the growth time would need to be increased this significantly, and therefore samples of different branch thicknesses but similar primary wire and branch lengths were not obtained. Granted more time, the necessary samples would have been grown and their reflectance spectra would have been compared.

Branch Length

Microwires with similar primary wire lengths and branch diameters, but different branch lengths were also compared in order to determine the impact of branch length on the optical properties of the samples. Figure 8(a) shows a plot of microwires with an approximate primary wire length of 30 μm and Au thickness

of 20 nm. The branch lengths for the 31.2- μm , 27.4- μm , and 29.0- μm wires are roughly 1.2 μm , 2.0 μm , and 3.0 μm , respectively. As shown by the plot, for wavelengths of 450 nm and greater, the shorter branches have a lower reflectance than the longer branches. However, due to longer growth times, the 27.4 μm and 29.0 μm samples have long branches extending from the tips of the wires, which could be the cause of the higher reflection of infrared wavelengths. Further testing of samples with the gold tips removed prior to branching is needed to obtain definitive conclusions regarding the impact of branch length on the reflectivity of the samples.

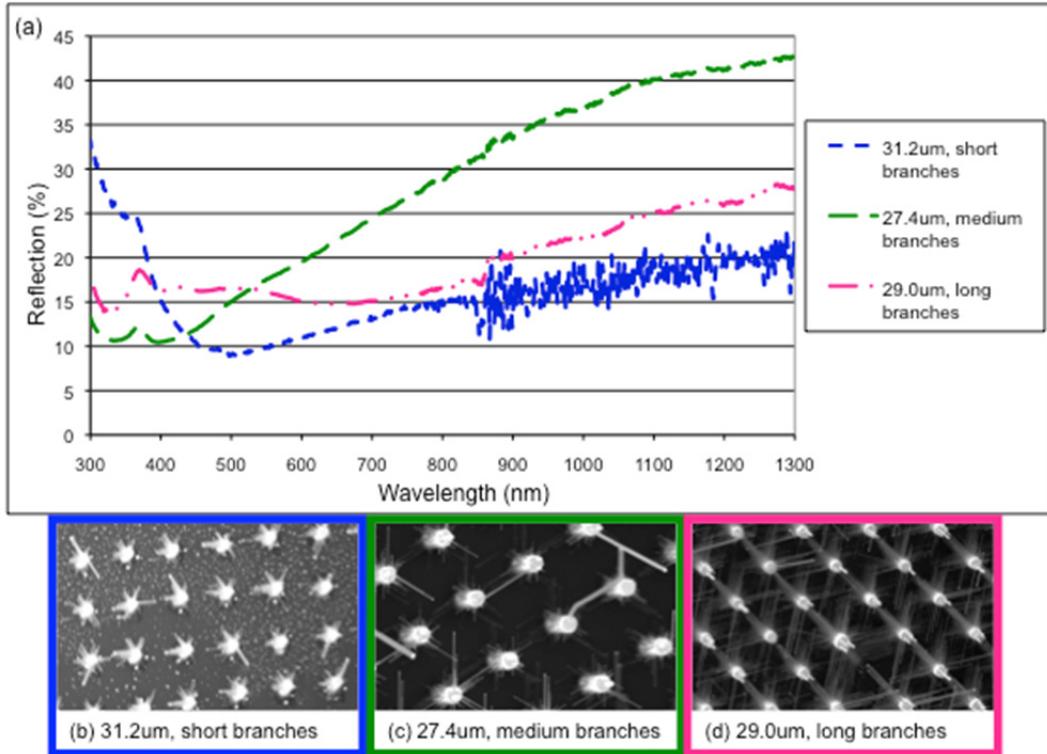


Figure 8: Reflectance measurements from the samples with primary wire lengths of approximately 30 μm and varying branch lengths are plotted in (a). SEM pictures of the samples with short, medium, and long branches are shown in (b), (c), and (d), respectively.

Superhydrophobic Results

The contact angle measurements from the wettability testing revealed that, with the SAM coating, some of the branched arrays are quite hydrophobic and their contact angles approach superhydrophobic. The average advancing contact angle for the branched samples was 132°. The average advancing contact angle for the unbranched arrays was 113°, indicating that the branching reduced the wetting of the wires. All of the arrays had greater contact angles than the planar silicon, which had an advancing contact angle of only 43°. Images from the

goniometer with advancing and receding contact angle measurements for the planar silicon, unbranched microwire, and branched microwire samples are shown in Figures 9(a), (b), and (c), respectively.

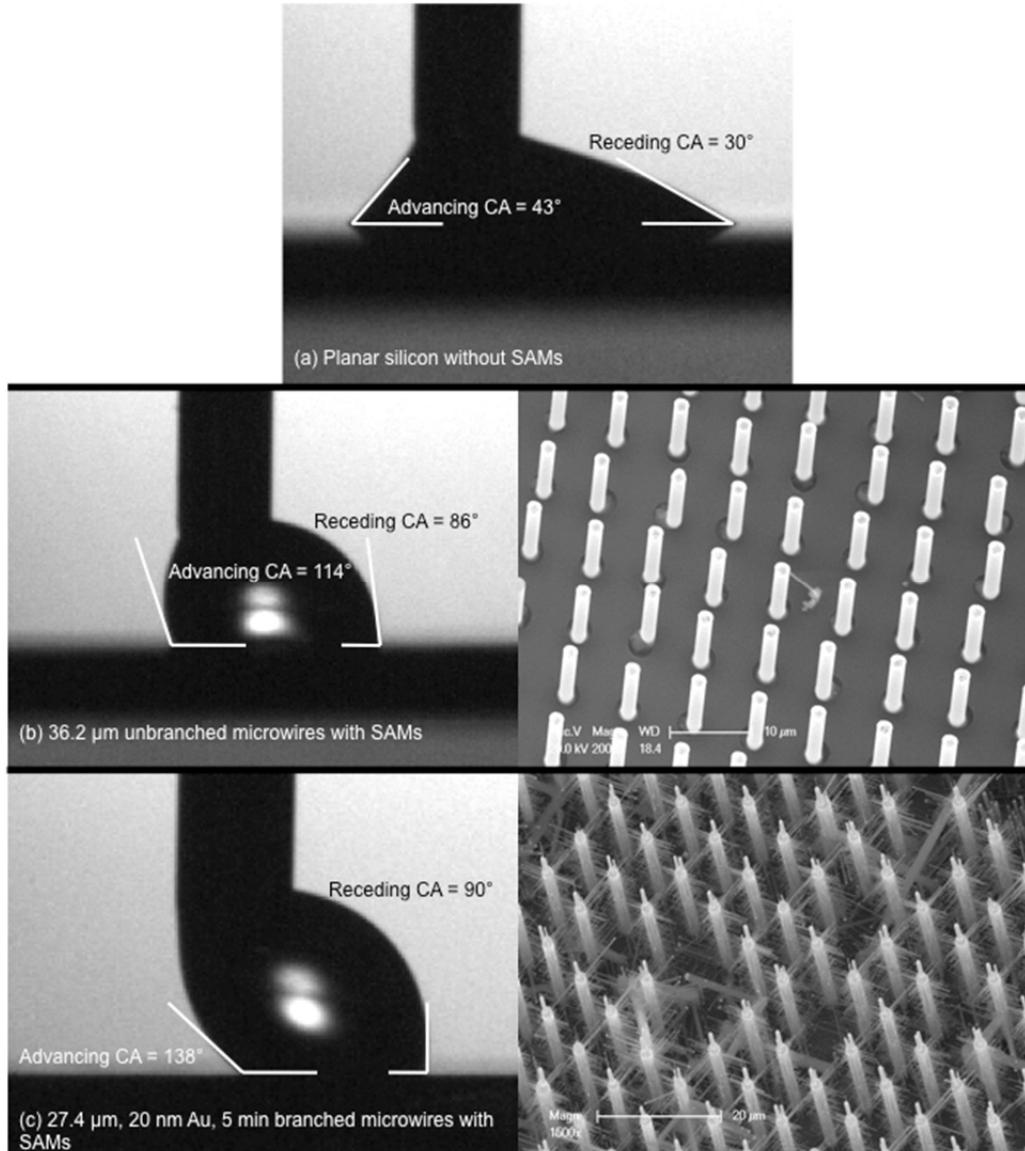


Figure 9: Contact angle measurements for (a) planar silicon, (b) unbranched microwires, and (c) branched microwires with 20 nm of Au are shown. The white lines show the advancing contact angles on the left and the receding contact angles on the right

DISCUSSION

The hypothesis presented in this work was that branching would reduce the reflectivity of microwire arrays, thereby increasing their absorption of incident

light. Using a spectrophotometer, the optical testing showed that this is true for microwires with primary wire lengths below roughly 200 μm . Longer than this length, the branching becomes less beneficial and can even result in a higher reflectivity than the unbranched samples. Also, as shown in Figure 5, the branched nanowires and 44.7 μm branched sample reflected more than 50% of infrared wavelengths. This similar reflectivity among these samples could be attributed to the long branches extending from the tops of the wires in the 44.7 μm branched sample. Nonetheless, all of the microwire samples, both unbranched and branched, had significantly lower reflectance overall than planar silicon. This indicates that these microwires are more efficient absorbers than a bare silicon wafer.

Additionally, the growth process of the branched microarrays was successfully controlled. Adjusting the Au thickness that is sputtered onto the primary wires during the post-growth process altered the branch diameters, while changing the growth time varied the branch lengths. Also, the cross-section SEM pictures verified that branches were able to grow all the way down to the base of the wires, indicating that the sputtering successfully coated each wire entirely with Au. A growth process for branching nanowires was also established. After the primary growth, the nanowires needed to cool for an hour prior to branching in order to give the gold enough time to diffuse down the sides. This process was unsuccessful for the microwires because of their larger size leading to a more stable gold droplet.

The affects of branch diameter and length on the reflectivity of the samples were inconclusive. Samples with similar characteristics, other than the variable being analyzed, were not obtained due to inconsistent growths. Further testing is needed to determine optimal branch diameters and lengths. In the future, prior to performing the optical testing, the gold should be removed from the samples to ensure that it does not alter the results.

The contact angle measurements indicated that some of the branched arrays coated with the (3,3,3-Trifluoropropyl) methyldimethoxysilane SAM were close to superhydrophobic. The branched samples also tended to reduce the wetting of the arrays. More testing is needed to determine the impact that the SAMs had on the wetting of the samples. Furthermore, other SAMs should be tested to optimize the superhydrophobic properties of the arrays.

CONCLUSION

The use of branched silicon microwire arrays in dye-sensitized and solid-state solar cells is a promising alternative to crystalline silicon wafers. The branched microwires proved to be better absorbers than planar silicon, and decreased the reflectivity of most unbranched arrays. The increased surface area of the branched arrays is desirable for dye-sensitized solar cells, and their low reflectivity makes the wires good candidates for solid-state devices. Additional research on the solar cell efficiency of branched microwire arrays is needed to determine their pragmatism in the photovoltaic market.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. C. DiMarino would like to thank Chito Kendrick for his guidance and support throughout the course of this project, and Joan Redwing for providing me with this research opportunity. She would also like to thank Hoating Shen for his help using the spectrophotometer, and Dr. Tom Mallouk and Liz Seibel for assisting with the superhydrophobic testing.

REFERENCES

- [1] Wolden, Colin A., Juanita Kurtin, Jason B. Baxter, Ingrid Repins, Sean E. Shaheen, John T. Torvik, Angus A. Rockett, Vasilis M. Fthenakis, and Eray S. Aydil. "Photovoltaic Manufacturing: Present Status, Future, Prospects, and Research Needs." *Journal of Vacuum Science and Technology A* 29.3 (2011). American Vacuum Society. Web.
- [2] Kayes, Brendan M. "Radial pn Junction, Wire Array Solar Cells." Thesis. California Institute of Technology, 2009. Print.
- [3] Lopez, Rene, Joan Redwing, and Kathleen Melde. "Bio-Inspired Electro-Optic Structure for Silicon Photovoltaics." (2011). Print.
- [4] Picraux, S.T., J. Yoo, I.H. Campbell, S.A. Dayeh, and D.E. Perea. "Semiconductor Nanowires for Solar Cells." *Semiconductor Nanostructures for Optoelectronic Devices-Processing, Characterization and Applications*. Los Alamos: Center for Integrated Nanotechnologies. Print.
- [5] Kayes, Brendan M., Harry A. Atwater, and Nathan S. Lewis. "Comparison of the Device Physics Principles of Planar and Radial P-n Junction Nanorod Solar Cells." *Journal of Applied Physics* 97.11 (2005). Print.
- [6] Garnett, Erik C., Mark L. Brongersma, Yi Cui, and Michael D. McGehee. "Nanowire Solar Cells." *Annual Review of Materials Research* 41.1 (2010): 269-95. Print.
- [7] Wu, Chen-Te, Wen-Pin Liao, and Jih-Jen Wu. "Three-dimensional ZnO Nanodendrite/nanoparticle Composite Solar Cells." *Journal of Materials Chemistry* 21 (2011): 2871-876. Print.
- [8] Ko, Seung H., Daeho Lee, Hyun W. Kang, Koo H. Nam, Joon Y. Yeo, Suk J. Hong, Costas P. Grigoropoulos, and Hyung J. Sung. "Nanoforest of Hydrothermally Grown Hierarchical ZnO Nanowires for a High Efficiency Dye-Sensitized Solar Cell." *Nano Letters* (2011): 666-71. Print.
- [9] Wang, Shutao, and Lei Jiang. "Definition of Superhydrophobic States." *Advanced Materials* 19.21 (2007): 3423-424. Print.
- [10] "Aircraft Icing." AOPA (2008). FAA, Flight Safety Branch. Web. 8 July 2011. <<http://www.aopa.org/asf/publications/sa11.pdf>>.
- [11] Wagner, R. S., and W. C. Ellis. "Vapor-Liquid-Solid Mechanism of Single Crystal Growth." *Applied Physics Letters* 4.5 (1964): 89-90. Print.
- [12] "MCL - UV-Vis Instrumentation." *Materials Research Institute*. Web. 8 July 2011. <<http://www.mri.psu.edu/facilities/MCL/techniques/UV-Vis/UV-VisInstrument.asp>>.

DEVELOPMENT OF A NEAR-FIELD SCANNING OPTICAL MICROSCOPE SYSTEM

Ruth Nan,* Chuan Yang,⁺ Hans Hallen, and Zhiwen Liu[#]

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

*Undergraduate student of
Department of Biomedical Engineering
Washington University in St. Louis
St. Louis, MO 63130

ABSTRACT

Traditional optical microscopy has a diffraction limit determined by wavelength and numerical aperture. The smallest resolution achieved using traditional optical methods is around 200 nm. Near-field scanning optical microscopy (NSOM/SNOM) was developed in the late 1980's as a means to break the diffraction limit, while collecting data on the topography of a sample. NSOM utilizes near-field interactions and raster scans point-by-point in order to resolve at a 50–100 nm range. This discovery has opened the way for many nano-applications in the fields of materials science, biology, and optics.

The goal of this project is to set up the near-field scanning optical microscope, with all of its components, obtained from our collaborator at NC State. The set up of a near-field scanning optical microscope involves electronics, optics, and software. This paper will explore the set-up of an NSOM microscope and detail the methods in operating it. The result of the project was a set of working connections for an NSOM as well as a frequency scan generation.

INTRODUCTION

Recently, a push towards nanoscience and nanotechnology has prompted research in a smaller scale. The drive towards the ability to measure, to manufacture, and to obtain structures on the nanometer scale has called for a microscopy method that is able to resolve at that level. Several nano-optical methods have been developed to image at the nanometer scale, including both far-field and near-field attempts to break the diffraction limit. Far-field methods used

⁺ Graduate Mentor

[#] Faculty Mentor

to attain higher resolution include stimulated emission depletion (STED or STORM) and photo-activated localization microscopy (PALM) or stochastic optical reconstruction microscopy (STORM). Another method employed by scientists to break the diffraction limit is scanning electron microscopy (SEM). Near-field methods of imaging include apertureless near-field scanning optical microscopy (ANSOM) and near-field optical microscopy (NSOM).

Traditional Microscopy Methods: Optics

The field of optics has long used Maxwell's equations for light to describe the electromagnetic behavior of light. Since their discovery, Maxwell's equations have had strong implications on the study of light as an electromagnetic wave. Although evanescent waves have been speculated about for decades, only recently have they been used to super-resolve images because of their near field properties. Evanescent waves exponentially decay because of the plane wave solution to Maxwell's equations. The uniform plane wave solution to Maxwell's equations give the following equations:

$$\mathbf{E}(\mathbf{r}) = E_0 e^{-ik\mathbf{r}} \quad (1)$$

$$\mathbf{B}(\mathbf{r}) = B_0 e^{-ik\mathbf{r}} \quad (2)$$

Using far-field detection and illumination, optical microscopy has been the traditional method for imaging.

Limitations of Traditional Optical Microscopy

The point-spread function is also known as the Airy function, and it defines the resolution limit through its width. The point-spread function measures the resolving power of an optical system by defining the spread of a point source, as shown in Figure 1.

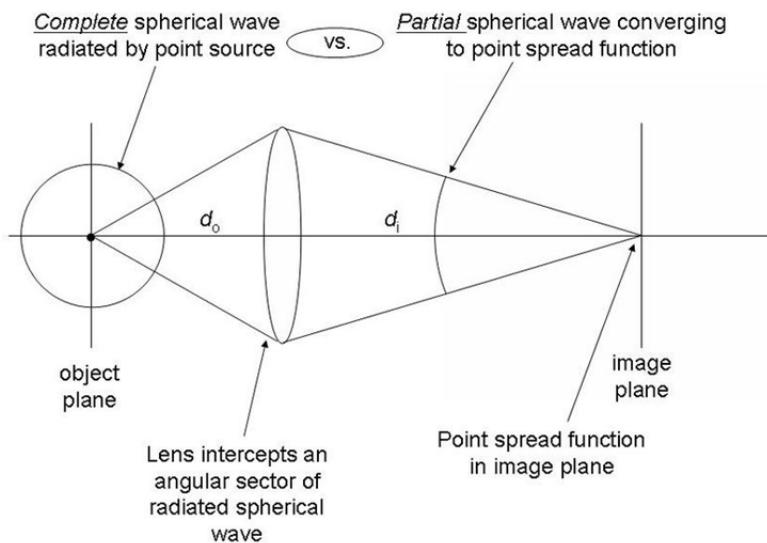


Figure 1: The configuration used for the point-spread function calculation.

The width of the point-spread function Δx is denoted as the Airy disk radius and gives the resolution limit of the paraxial point-spread function. Two point-spread functions must have their maxima be separated by a minimum distance of the characteristic width of an individual point-spread function in order to be distinguished. In the early 1870's, Ernst Abbe defined this resolution limit for the far-field microscope, where NA is the numerical aperture defined by the index of refraction:

$$\Delta x = 0.6098 \frac{\lambda}{NA} \quad (3)$$

$$NA = n \sin \theta \quad (4)$$

Because the Abbe limit is defined by the wavelength of the emitter and the index of refraction of the lens, it has been manipulated by means of using the compound oil-immersion microscope, which has an index of refraction of 1.4, and using low wavelengths of light (around 400 nm). However, the lowest resolution attainable by manipulation of parameters is still a couple hundred nanometers.

With the invention of near-field scanning optical microscopy, or NSOM, the Rayleigh and Abbe diffraction limit no longer existed, and optical resolution of less than 50 nm is achievable. With NSOM, evanescent fields near the tip can be detected by the probe, giving a distance-dependent increase in the number of photons.

Near-field Scanning Optical Microscopy

Near-field scanning optical microscopy (NSOM) was first conceived by E. H. Synge in 1928. He proposed the idea of using a pinhole aperture to image a surface at a sub-wavelength resolution. By raster scanning at a distance less than the aperture diameter, the resolution would be the size of the diameter. However, the technology at that time had not yet advanced to the point where it was possible to generate the technology needed for NSOM, Synge's idea was unable to be realized and soon forgotten [2]. In 1956, J. A. O'Keefe developed similar theories on the concept of near-field microscopy, without knowing about Synge's earlier papers [3]. O'Keefe writes, "The realization of this proposal is rather remote, because of the difficulty providing for relative motion between the pinhole and the object, when the object must be brought so close to the pinhole."

Finally, in 1972, Ash and Nichols performed the first experimental realization in the microwave region using a 1.5-mm aperture with a wavelength of 3 cm. Ash and Nichols demonstrated a resolution of $\lambda/60$ [4]. The invention and development of scanning probe microscopy in the early 1980's paved the way for the realization of Synge's idea at optical frequencies. In 1984, the first papers on NSOM appear, showing that near-field microscopy is practical possibility and spurring the growth of a new scientific field.

Modes of NSOM

Several configurations are possible for near-field optical microscopy. The configurations are simply the different combinations of near-field and far-field illumination and detection, as illustrated in Figure 2. In all cases, the light can be detected using different equipment—a photodiode, a photomultiplier tube (PMT), or a charge-coupled device (CCD). The four modes of operation for the NSOM system are transmission, reflection, collection, and illumination/collection. In transmission mode, the probe illuminates the sample, and the light that reaches the other side of the sample is collected and detected. In reflection mode, the probe also illuminates the sample, and the reflected light from the surface is collected and detected. In collection mode, the sample is externally illuminated from either above or below, and the light from the surface is collected into the probe. The last mode uses the probe to both illuminate the sample and collect the reflected signal from the sample. The signals from the detector are used to create an image of the surface.

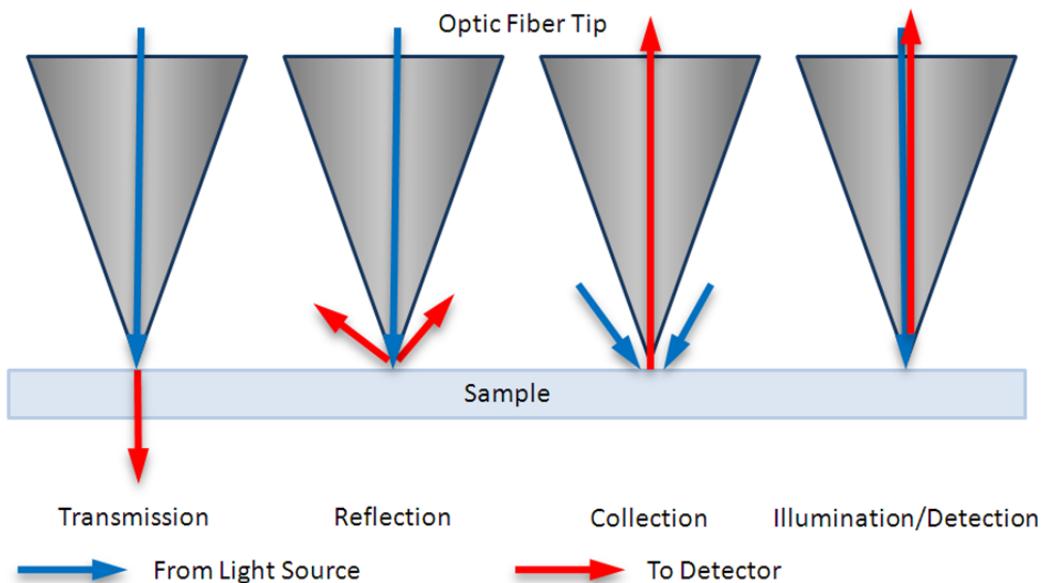


Figure 2: Schematic illustration of NSOM modes of operation.

EXPERIMENTAL DESCRIPTION

The near-field scanning optical microscope is composed of several components: the parts, the connections, and DataTaker, a data acquisition tool created in National Instrument's LabVIEW. The parts consist of the tip holder (tuning fork and tip), microscope, piezoelectric tube, equipment (oscilloscope, lock-in amplifier, quiet box, stepper motor box), connector box, and the power supply. To set up the NSOM and make it functional, the connections must be checked, a tip holder made, and resonance and the Q factor determined.

Checking Connections

Table 1 shows a list of the connections to and from the components: connector box, “quiet box,” lock-in amplifier, and the patch out box. The other Z mon output is to an oscilloscope, so you can observe what is happening along the z-axis. The output of the preamplifier board goes to the two inputs (Signal In) of the lock-in amplifier via an A-B adaptor, and the output of the lock-in (Sine Out) feeds back into the board.

Table 1: Modified table of connection inputs and outputs [6].

FROM	TO
<u>PCI-MIO Connector Box</u>	<u>Stepper Card</u>
<i>Digital I/O</i>	
0	Step
1	Direction
<u>PCI-MIO Connector Box</u>	“Quiet Box”
<i>Digital I/O</i>	
2	Sample/Hold Switch
<i>DAC (Analog Output)</i>	
0	X20 X IN
1	X20 Y IN
<i>Differential Inputs (ADC)</i>	
0	Feedback In
1	Reference In
2	Z Mon Out
e	Error Monitor
“Quiet Box”	Patch Out Box
+X	
-X	
+Y	
-Y	
HV Z Out	Input to patch out box and from there via Cooner wire to the XYZ tube piezo
<u>Lock-in Amplifier</u>	<u>PCI-MIO Connector Box</u>
<i>Back</i>	<i>Differential Input</i>
Fast X (X)	0
DAC 0 (Aux Out 1)	Reference In

The computer has a Peripheral Component Interconnect (PCI) or a personal computer bus card, the National Instruments (NI) PCI-6251, which is connected to the Connector Box, the NI BNC-2110. A one-to-one transfer of the cables was applied from the PCI-MIO-16XE-50 connector box to the BNC-2110 to make the NSOM work on different computers.

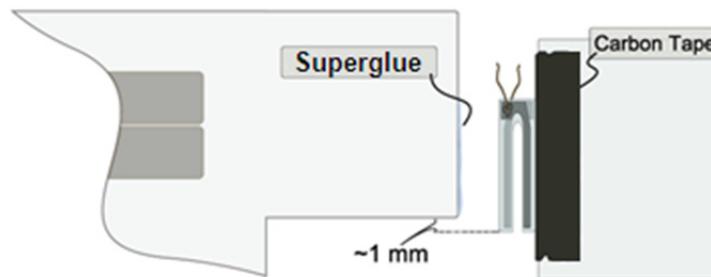
Making Tip Holder

The NSOM operates in a feedback mechanism to keep the tip at a constant distance from the surface of the sample. The tip holder is composed of a glass

slide, a 32.768-kHz tuning fork, and an etched or pulled optical fiber. The tuning fork is first placed on a glass slide with double-sided carbon tape. Using a three-dimensional stage for the tip holder glass slide with metal rods and a clamp with a carbon-taped glass slide for the tuning fork, the tip holder is carefully placed against the tuning fork. A small amount of UV epoxy is put on the edge of the holder beforehand using an optical fiber or needle, and after the tuning fork is on the holder, it is cured under a black light for 30 minutes. After hardening, the holder and tuning fork are gently removed from the glass slide with the carbon tape. The carbon tape is for easy transfer of the tuning fork to the holder and prevents any residue from sticking on the tuning fork.

The optical fiber is aligned with the other side of the tuning fork, and the entire length of the tip is flush with the tuning fork. The tapered region of the fiber (about 0.5 to 1 mm of the tip) should extend past the end of the tuning fork. A bit of UV-cure optical epoxy glue is applied on the other side of the tuning fork in the same manner as before. The optical fiber is mounted entirely against the edge of the tuning fork, and the glue will wick all the way up the joint. It is then cured under a UV lamp for another 30 minutes. Figure 3 shows a schematic of how to glue the tuning fork and fiber tip to the tip holder.

Gluing the Tuning Fork onto Tuning Fork Holder



Gluing the Optical Fiber Tip onto Tip Holder



Figure 3: Gluing the tuning fork and fiber tip to make a tip and tuning fork holder, reproduced based on reference [7].

Resonance and Q

To determine the resonance of the tip holder, the bridge is first balanced. The variable capacitor and resistor are adjusted to obtain the lowest background voltage signal possible, less than 0.1 mV. The leads of the tuning fork must be at least a couple of inches apart, to prevent vibrations from affecting the capacitance between the leads. In the Resonance Scan vi in DataTaker, a scan through the frequencies from 30 to 40 kHz determines the resonance peak. Quality is determined using the formula:

$$Q = \frac{f_0}{\Delta f} = \frac{\omega_0}{\gamma\sqrt{3}} \quad (5)$$

where $f_0 = \omega_0/2\pi$ is the resonance frequency and γ is the damping constant. Δf is the full-width at half-maximum (FWHM) of the resonance, and these parameters determine the Q -factor of the fork. The quality factor, or the Q -factor is used to determine how underdamped a system is—a higher Q indicates a lower energy loss relative to the stored energy of the oscillator, or a lower damping constant.

DataTaker LabVIEW Settings

The data acquisition tool, DataTaker, is the software used to interface with the NSOM. For a Mac, the two folders in the installed driver folder are named ‘myDAQmxRoutines’ and ‘myFullDAQmxRoutines,’ and for a PC, the folders are called ‘myDAQmxBaseRoutines’ and ‘myDAQmxRoutines,’ respectively. On DataTaker, there are several options to choose. The buttons we will be primarily concerned with are ‘Instr. Set-up,’ ‘Watch8,’ ‘Resonance Scan,’ and ‘Scan,’ shown in Figure 4.



Figure 4: DataTaker panel to communicate with the NSOM via the connector box and PCI Data Acquisition (DAQ) card [6].

Resonance Scan is used to scan through an internal range of frequencies from 30 kHz to 40 kHz and to monitor the voltage output while scanning. From the resonance peak, the quality, operating frequency, and feedback values can be determined. After setting the operating frequency, the frequency at which the amplitude is 85–95% of the value at resonance, a z-approach curve is desired. The ‘Step-Test Approach’ sub-vi is used for this purpose.

EXPERIMENTAL RESULTS

Figure 5 shows the amplitude versus frequency plot of the bare tuning fork. The quality (Q) factor for this tuning fork is 3374.06, and the resonance is at 32.768 kHz. Table 2 compares the resonant frequencies and quality of the bare tuning fork and the tuning fork attached to the tip holder and optical fiber.

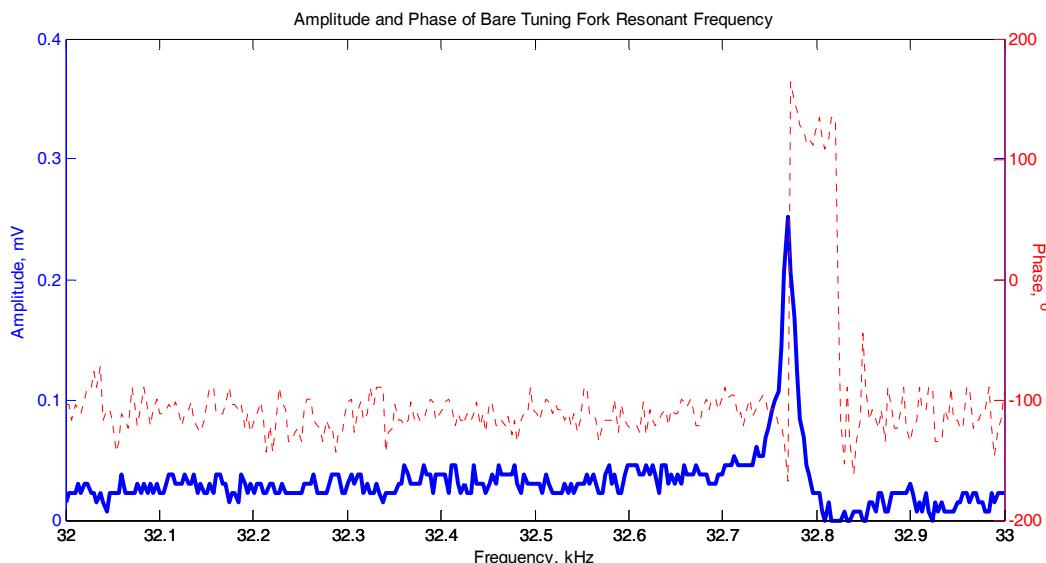


Figure 5: Bare tuning fork amplitude and phase at resonant frequency.

A total of eight fiber optic tips, or probes, are available for making a tip holder. Three of these are shown below in Figure 6. The first tip was used in the tip holder, which consisted of a glass slide, two metal rods, a tuning fork, and the optical fiber. A monochromatic CCD was used to observe the tip, and the captured image is shown below in Figure 7. This optical fiber was the tip used on the holder for data collection. The taper of the tip was created via meniscus etching or tube etching.

Table 2: Comparison of characteristics of bare and attached tuning fork.

Fork	Resonance Frequency	Quality
Bare	32.768 kHz	3374.06
Attached	31.473 kHz	395.64

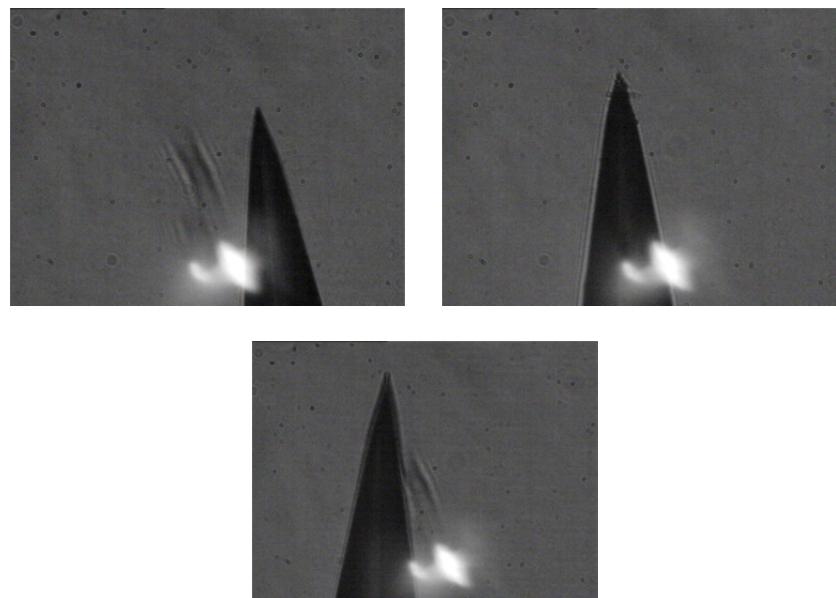


Figure 6: Three different fiber optic tips as observed by an optical microscope.



Figure 7: The optical fiber tip, coated with aluminum.

The tip holder resonance was determined to be 31.473 kHz by scanning through a range of frequencies and locating the highest output voltage peak. Figure 7 zooms in on the resonance scan from 34.44 kHz to 34.50 kHz, generating a plot of amplitude versus frequency.

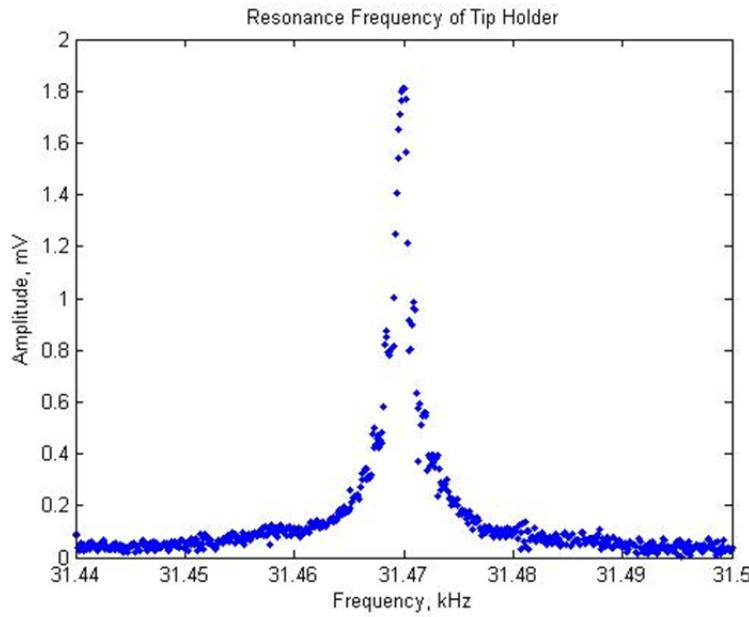


Figure 7: Amplitude vs. frequency plot of the tip holder, showing the resonance.

By scanning through a large range of frequencies, 30–40 kHz, the significance of a peak at 31.473 kHz can be determined. Here, the quality of the peak, calculated using Equation 5 and listed in Table 2, is about 400, which is in the desirable range for Q 's of 100–500 for feedback.

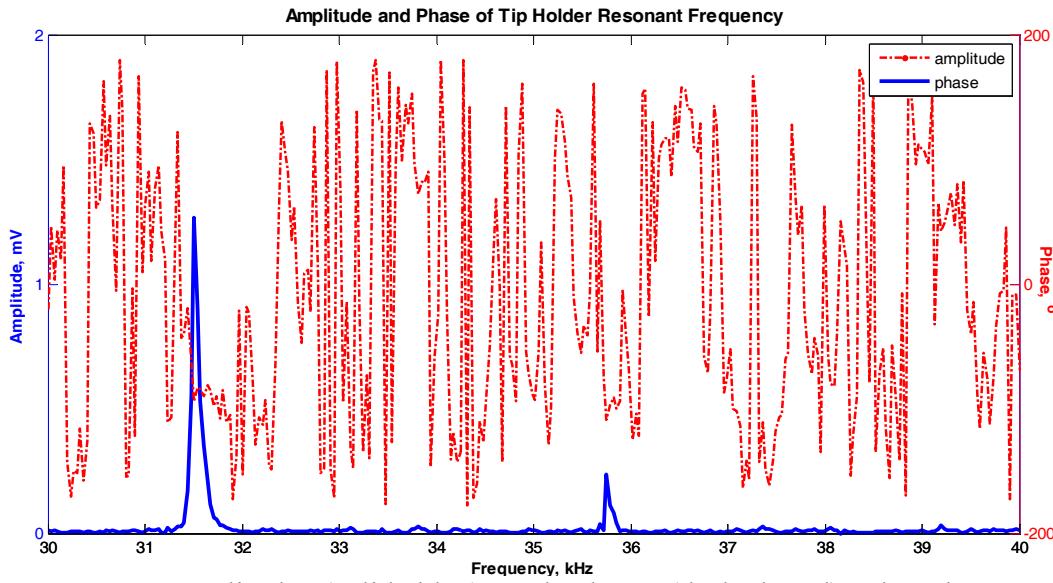


Figure 8: Amplitude (solid blue) and phase (dashed red) plotted versus frequency.

DISCUSSION

The bare tuning fork and the tuning fork attached to the glass slide and optic fiber show varying resonance frequency and quality values. The resonance frequency shifts from its natural resonant frequency of 32.768 kHz when the bare tuning fork is attached to the glass slide and tip. The shift in frequency is expected because the fiber and glass slide add a mass onto both arms of the tuning fork, causing it to resonate at a different frequency. The quality of a bare tuning fork is almost 10 times higher than an attached tuning fork, which is expected. A bare tuning fork can oscillate for a long time without much dampening, but when there is mass added to both arms, the quality drops because the full-width at half-maximum is much larger, which can be seen in Equation 5.

The scan through the frequencies can be compared between the bare tuning fork and the attached tuning fork in Figures 7 and 8. Observing the scale, the peak of the bare tuning fork is much narrower than the peak of the attached tuning fork. The peak resembles a Dirac delta function when it spikes at its respective resonant frequency. Because the peak is so narrow, it is difficult to detect the resonant frequency of a bare tuning fork since it is easy to miss. By sampling at higher rate and using closer limits, the resonant peak of a bare tuning fork is detectable.

The noisy baseline of the bare tuning fork can be caused by several reasons. If there are any drifts in the air circulation of the room, it may be picked up by the tip and recorded onto the resonance scan. Also, an unbalanced resistor capacitor bridge circuit on the pre-amplifier board will cause an increase in background voltage with an increase in frequency.

CONCLUSIONS

The NSOM system operates using a feedback loop with the lock-in amplifier, the stepper motor, and a tuning fork. An optical fiber collects light point-by-point and images a sample. By following the contours of the sample, the NSOM is able to produce both a topographic and an optical image in the nanometer scale.

The connections for the near-field scanning optical microscope set-up were accomplished, and a resonance frequency scan was finished. The bare tuning fork exhibited a much higher quality-factor than the attached tuning fork, and the resonance frequency shifted when mass was attached to either arm. Although the approach curve has not yet been generated, when a voltage is applied to the stepping motor, the motor turns proportional to the amount of voltage applied to the stepper motor box. The next steps are to set the feedback settings to obtain an approach curve and conduct a scan.

ACKNOWLEDGMENTS

One of the authors (RN) would like to thank Dr. Zhiwen Liu for the opportunity to work in his lab this summer, Chuan Yang for his mentorship and guidance, and the members of the Ultrafast and Nonlinear Optics Lab at Penn State for their help and support. RN would also like to thank Dr. Hans Hallen from North Carolina State University for the many Skype meetings and emails

required to set up the near-field scanning optical microscope. Lastly, RN would like to acknowledge Penn State University and the National Science Foundation for providing the funding for her summer research project under Grant No. EEC-1062984.

REFERENCES

- [1] L. Novotny and B. Hecht, *Principles of Nano-optics*. New York: Cambridge University Press, 2006.
- [2] Synge, E. H. "A suggested method for extending the microscopic resolution into the ultramicroscopic region". *Phil. Mag.* **6**: 356 (1928).
- [3] J. A. O'Keefe, "Resolving power of visible light," *J. Opt. Soc. Am.* **46**, 359 (1956).
- [4] Ash, E.A., and G. Nicholls. "Super-resolution aperture scanning microscope," *Nature* **237**: 510—513 (1972).
- [5] B. E. A. Saleh and M. C. Teich. *Fundamentals of Photonics*. New York: John Wiley & Sons, Inc., 1991.
- [6] Hallen, H. and S. Huerth. 1999. *Users Manual for NCSU NSOM*. Unpublished manuscript, North Carolina State University, Raleigh, NC.
- [8] "NSOM (SNOM) Overview." *HTSKorea*. 2003. Accessed July 11, 2011. <http://www.htskorea.com/technote7/data/board/kkkinfo5/file_in_body/1/gen03.pdf>.

MRI MICROCOILS FOR IMAGING INDIVIDUAL CELLS

Matthew Feldman,* Michael Lanagan,[#] and Steven Perini⁺

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

*Undergraduate Student of

Department of Electrical & Computer Engineering
University of Florida
Gainesville, FL 32612

ABSTRACT

With high-magnetic-field MRI scanners, it becomes necessary to finely tune the impedance and resonating frequency of a receive-coil electronically. B_1 -producing coils for large magnets, especially those that target 600 MHz or greater, are extremely sensitive to external influences and coupling with the copper shielding, metal inside the machine, and the sample being imaged. Signal-to-noise ratio and resolution can be sharply increased if the resonant frequency of the coil can be tuned simultaneously while all of these factors have been introduced, rather than by trial-and-error by taking apart the set up each time to make an adjustment.

Varactors are voltage tunable capacitors that are a potential option for allowing the user to tune the coil electronically from a distance. This study has shown that they are capable of tuning a circuit to a desired frequency, as viewed from the S_{11} parameter on a Vector Network Analyzer. However, the Q-factor of hyperabrupt-junction varactors is so low that the S_{21} does not show any discernable peak. Therefore, this method of tuning is not the most desirable for high-frequency applications, as loss is directly related to frequency.

A Helmholtz coil proved to be a valuable design for creating a uniform magnetic field. Though the dimensions of the setup in this study were not ideal, they seemed to indicate that a properly designed coil will exhibit rewarding behavior. The magnetic field passing through the midplane seemed to be very uniform, except for where there was a deliberately-placed gap in the coil for design purposes. The behavior of the magnetic field along the axis passing through the centers of the two loops was inconclusive with the results from this

[#] Primary Faculty Mentor

⁺ Secondary Faculty Mentor

study, but further experimentation can be done with existing prototypes to assess this.

It would be advantageous to assess whether it is the complication of the circuit with a varactor setup or the varactors themselves that are interfering with the emission of signal. Numerous steps can be taken in the future, including using alternative electronic tuning methods or using different components to test specific parts of the circuit, in order to improve the SNR and resolution. In the future, these coils can be used to advance biological knowledge of tissue structures and real-time understanding of flow, diffusion, and perfusion of substances.

INTRODUCTION

Magnetic Resonance

The goal of this project is to use the nuclear magnetic resonance (NMR) signal to obtain images of individual biological cells. In NMR imaging, magnetic fields induce atoms with non-zero nuclear spin to give off radio frequency (RF) signal. Specifically, a B_0 magnetic field aligns the net nuclear spin of hydrogen atoms within substances such as adipose and water. A B_1 field is then used to cause these spins to precess at a “Larmor” frequency. This frequency is determined by both the gyromagnetic ratio of the atom and the value of B_0 . Three orthogonal gradient coils are used to select a slice and cause a position-based alteration of frequency and phase that can be used to discern voxels in an image. The B_1 process involves both transmission in the form of an RF electromagnetic wave in the shape of a sinc function and the reception of the RF signal emitted by the nuclear spins using Faraday’s Law. All of these various coils, except for B_1 receive, are fully contained within the 14-T MRI scanner that is to be used in this project. However, the receive coil must be custom built.

In recent years, some companies have already built coils that can be used to image small objects, but the signal-to-noise ratio and resolution of these images are less than desirable to biologists and researchers.¹ The signal-to-noise ratio can be improved by using a stronger B_0 field, improving the sensitivity of the receive coil, and lengthening the imaging sequence. The density of k -space, field of view, and slice thickness determines the spatial resolution. The coil built in this project is designed to improve these parameters.

Helmholtz Coils

The coil design used in this resonator is a Helmholtz coil. This coil design consists of two loops of equal radius separated by some distance, as shown in Figure 1. When the distance between them is equal to their radius, the magnetic field is strong and uniform. It is a volume coil, which makes it better than a surface coil, considering its field strength and uniformity. This is ideal for preventing artifact. The magnetic field, as a function of the distance from the midplane, r , is given by:

$$B(r) = I \times 10^{-7} \int_0^{2\pi} \frac{2a(a - r\cos(\theta))d\theta}{(a^2 + b^2 + r^2 - 2ar\cos(\theta))^{\frac{3}{2}}}$$

where I is the current through the coils, a is the radius of the coils, and b is the distance between them. By changing the values of parameters in this equation, it is possible to get various field strength designs. If the distance between the coils is too small, there will be strong magnetic field when r is close to a and weak magnetic field when r is close to zero. If the distance between the coils is too large, the magnetic field will peak in when r is close to zero, rather than creating a plateau across all values of r . Water phantoms can be used to assess the uniformity of the magnetic field in an MRI scanner.

Tunability

Since the inductance of the coil is fixed, a capacitor in parallel must be adjusted to tune the resonant frequency of the circuit, while a capacitor in series must be adjusted to tune the impedance of the circuit. Mechanically tunable capacitors have a high quality (Q) factor and are the most straightforward way of achieving tunability, but are impractical because the coil circuit is inserted deep in the imaging coil and inaccessible during the imaging sequence. Therefore, a digital means of tuning must be used. This project attempts to use varactor diodes to tune these two capacitances. These components vary capacitance with potential difference introduced across them. Therefore, cables can be fed through the MRI tube and out of the room, where DC power supplies can be used to supply the potential differences across both varactors. The junction capacitance, C_j , of a varactor as a function of reverse bias voltage, V_r , can be determined by the equation:

$$C_j(V_r) = \frac{C_{j0}}{\left(1 + \frac{V_r}{\Phi}\right)^\gamma}$$

where γ is a constant (equal to 0.5 for hyperabrupt junction and greater than 0.5 for abrupt junction varactors) and Φ is contact potential (about 0.75 for Si). The type of junction refers to the way the varactor was processed. Silicon and Gallium Arsenide are the only two materials a varactor can be made of.

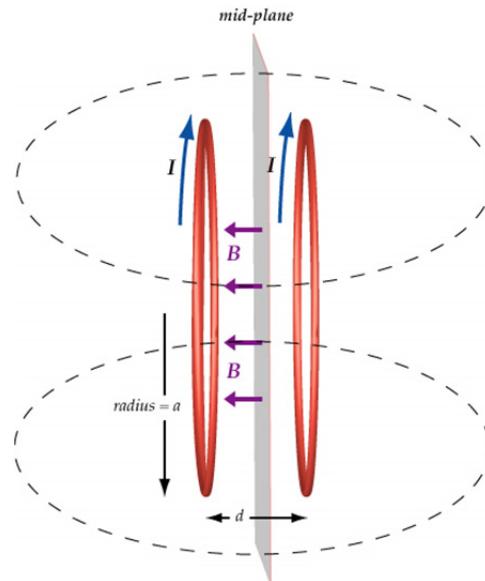


Figure 1: The variables that determine the uniformity of a magnetic field caused by a Helmholtz pair are the radius of the coils and the distance between them

EXPERIMENT DESCRIPTION

Equipment

A vector network analyzer (VNA) made it possible to see where the coil was resonating and how sharp the peak was. This piece of equipment is capable of testing a device at a range of frequencies in the microwave range and returning various data of interest. Specifically, the VNA was used to determine the S_{11} and S_{21} parameters of the chip. These parameters are ratios and measured in decibels. S_{11} represents the ratio of the signal being introduced into the chip through the $50\text{-}\Omega$ coaxial cable to the signal being returned back through the coax. A loop antenna was used as a load to obtain S_{21} , which represents the ratio from signal introduced through port 1 (the coil) to signal received through this antenna (port 2). The VNA also has a Smith Chart feature, which can be used to see whether the value of impedance of the circuit and whether it is inductive, capacitive, or resistive. This is helpful in determining if a peak in the frequency-versus-dB chart is actually a resonance frequency or a coincidental reflection of waves.

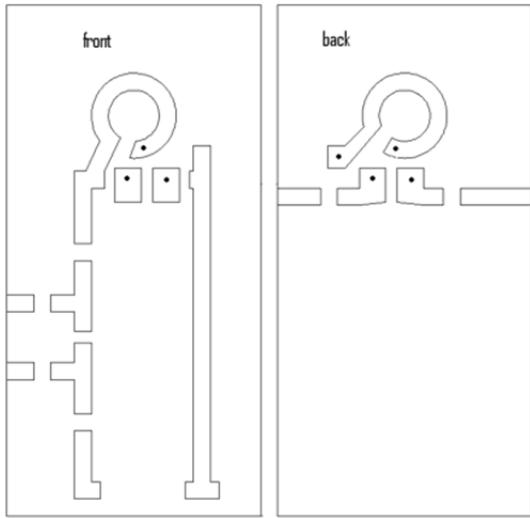


Figure 2: Actual layout for chip, with gaps for components and dots for through-connections

Proof of Concept

The feasibility of varactors was tested in many different steps. Due to cost restraints, magnetic varactors were used in the early models. The varactor requires two capacitors, two resistors, and biasing lines in order to operate, as shown in Figures 2 and 3. A DC reverse bias across the varactor is what changes its capacitance. Two large resistors must be placed in the DC circuit so that there is little current running across the varactor, and the actual signal going through the coil's circuit is not affected. Two large capacitors must isolate the DC circuit so that the potential difference is applied across only the varactor of interest and not the rest of the circuit. Using large isolating capacitors ensures that relatively small varactor capacitances dominate the equivalent capacitance. Two possible pitfalls with varactor tuning are that varactors are relatively lossy, and

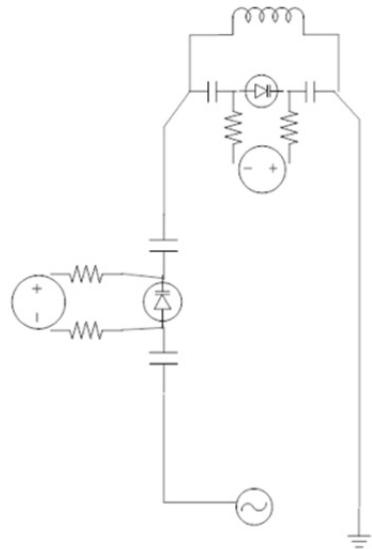


Figure 3: Schematic circuit diagram of 600-MHz resonator.

integrating them into a small chip may overcomplicate the circuit and cause incidental capacitance and inductance in the traces, as can be seen in Figures 2 and 3. Also, the varactors have a range of just a few picofarads. This means that the values of the isolating capacitors must be chosen carefully to allow the equivalent capacitance to be the necessary amount.

A CO₂ laser was used to cut out an AutoCAD-drawn copper trace design onto a 30 mm × 15 mm piece of Roger RO3000. This material is a ceramic and woven glass substrate with very low dielectric loss, and was used because high permittivity ceramic dielectrics can potentially ameliorate some of the challenges of high frequency coil design and encourage strong magnetic fields in a compact design.² The excess copper was chemically etched off, leaving a ceramic a sturdy platform for soldering pieces on and off. With this board, the experimental components were attached to get a sense of the frequency range with certain values of capacitance. An SMA connector was used to connect to the waveguide port.

After this initial test, nonmagnetic components were used. The nonmagnetic varactors are hyper-abrupt junction, which means they have a lower Q . This final product can be seen in Figure 4. Note that the varactors were attached using silver paint, rather than solder, due to a coating that quickly wore off the varactors, preventing them from bonding to solder. The final step of the process is to drill a hole in the Helmholtz coil and grow cells. Once this is done, the coil is to be mounted on an MRI probe and inserted into the scanning machine. Once the coil is in the machine, the presence of new materials and a large magnetic field will make the tuning process more complicated than it is in a lab. The circuit must be out of tune while the MRI machine undergoes its B_1 transmission process, but tuned quickly after in order to receive the signal. Various switches may be used to achieve this rapid tuning.

CST Microwave Studio was used to design a model of the coil and test the magnetic field uniformity and viability of the created circuit. MathCAD was used to mathematically calculate the theoretical distribution of the magnetic field in a coil, given our dimensions.

EXPERIMENT RESULTS

Theoretical Models

According to the CST Microwave Studio model of the circuit, it was possible

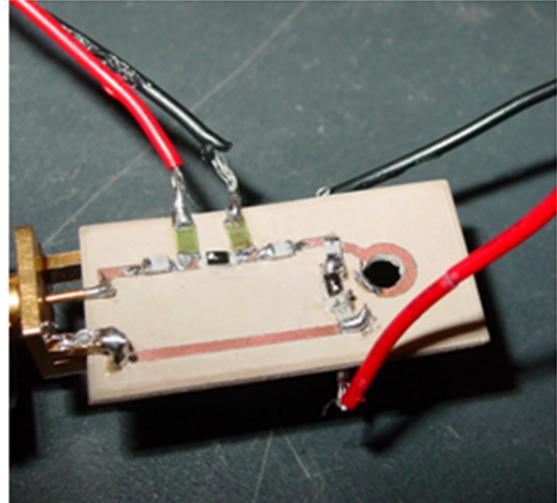


Figure 4: Photograph of finished product.

for our circuit to resonate at 600 MHz, given its geometry. The magnetic field within the Helmholtz coil clearly indicated that the field strength was somewhat hourglass shaped. There were a few limitations with CST in predicting what the circuit would look like. Firstly, there are no preset capacitors or varactors in the program. Therefore, having a specific gap in the copper trace set capacitance values. In the physical prototype, the ceramic capacitors and varactors actually have their own loss and inductance associated with their geometry. CST also did not allow for resistors or biasing lines to be added to the circuit. Had this been possible, it may have been possible to see if there is any coupling between the additional traces added to allow biasing lines.

After the chip was fabricated, the exact dimensions were input into a MathCAD file to calculate what the magnetic field should look like. The graph, Figure 5, showed that there is a noticeable dip in the midway between the coils. The ideal geometry would yield a nearly perfectly uniform magnetic field. When the imperfect coil is used in an MRI scanner, this deformity should be manifested as the middle of the image being darker than the ends of the image.

Physical Results

After the coil was constructed with varactor setups and was tuned to 600 MHz, data from the VNA were recorded with a LabVIEW data acquisition program. According to S_{11} , the varactors were a viable method of tuning the circuit. The impedance was able to be matched at 50Ω when the resonant frequency of the system was 600 MHz.

However, looking deeper into the setup and measuring S_{21} , the

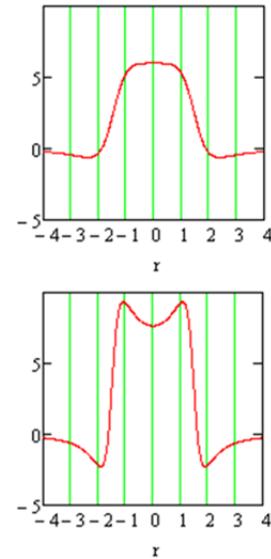


Figure 5: Magnetic field caused by Helmholtz coil with 1.5-mm radius. The top image represents magnetic field as a function of distance from the midplane if the separation between the two coils is ideal (1.5 mm). The bottom image represents the actual separation of the Helmholtz coil used in this study, 0.75 mm.

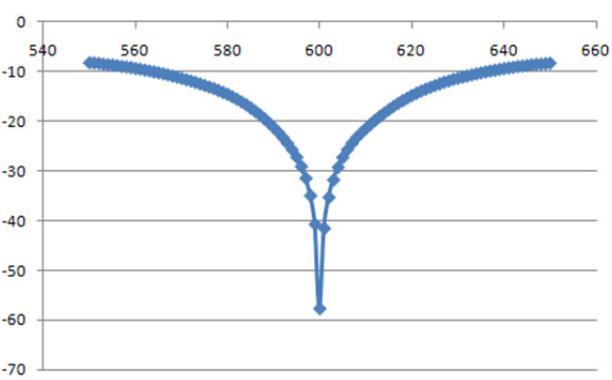


Figure 6: Plot of S_{11} versus frequency, when varactors are used to tune the circuit to 600 MHz.

varactors proved to be far too lossy for MRI applications. When a loop antenna was brought close to the coil, there was no discernable peak in the S_{21} logarithmic magnitude (Figure 7). This means that although the circuit was resonating at 600 MHz and very little signal was reflected back into port 1, very little signal was actually emitted as RF through the coil. It is believed that the low Q value of the varactors is the reason for this loss in energy. They may be absorbing most of what would normally be emitted.

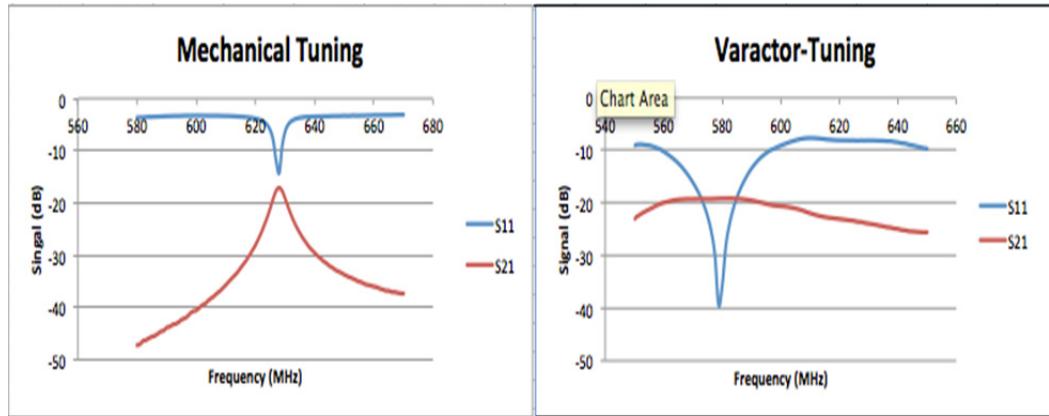


Figure 7: S_{11} and S_{21} comparison between the varactor-tuning and mechanical-tuning. The bandwidth of S_{21} for the mechanical setup has a much smaller bandwidth.

Unfortunately, when the chip was mounted on a 14-T MRI probe connected to the VNA with a BNC cable, it was impossible to recreate the S_{11} peak. Jarring the DC bias wires had a major impact on the S_{11} chart. It is unclear why this happened, as there should not be much coupling happening with the introduction of DC to the chip. A possible solution would be to use solid wire rather than bundled wire to connect to the power supply.

The S_{21} graphs shown in Figure 7 have implications from the perspective of MRI scans. When these particular coils are placed in the MRI scanners, the S_{11} parameter will be used to tune them, but S_{21} indicates of how well they will receive signal. Since these are receive-only coils, their function is to resonate under an induced oscillating *emf*, as per Faraday's Law, caused by T_2 precession of magnetic spin in the target substance. This whole process is the reverse of the S_{21} parameter, which makes it a good indication of how well the coil will operate. If the coil can emit RF energy at a certain frequency, it is also capable of picking up signal at that same frequency. When a varactor-tuned circuit is tested in an MRI scanner, the wide bandwidth will manifest itself as a low signal-to-noise ratio and resolution. However, it implies that the circuit does not necessarily need to be tuned to 600 MHz, as the coil will pick up signal from a range of frequencies about its peak.

B_1 Field Uniformity

A prototype of the Helmholtz coil design that used mechanically tunable capacitors was put into a 14-T MRI scanner to view the uniformity of the magnetic field. Images of various planes of a water-filled capillary were taken, using a 3D spin-echo sequence. Figure 8 describes the exact details of the sequence used to obtain the images below.

The most notable feature in Figure 9 is the meniscus. With the parameters of the spin-echo sequence, the meniscus came out very well and clear; however, it poses a problem in assessing the uniformity of the B_1 magnetic field of the Helmholtz coil. Figure 10 shows the signal intensity as a function of distance along the red line in Figure 9. Approaching the center of the coil, signal intensity increases, as is predicted by the simulations. Note that the scale is much larger than that of the MathCAD simulation. However, because of the meniscus, there is no water on the other side of the coil to assess whether or not the magnetic field created the “hourglass” shape that was predicted. There is a steep drop in intensity once the water stops. Therefore, the results of this particular scan are inconclusive in this respect and another scan should be done with a more carefully designed capillary.

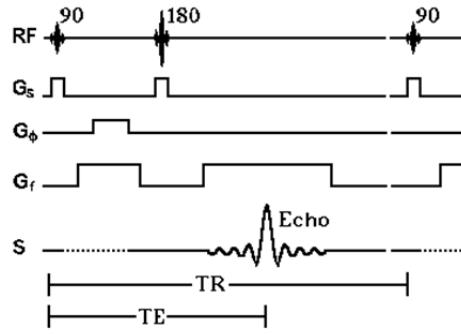


Figure 8: The field-of-view was $5 \times 3 \times 3$ mm and the data matrix was $166 \times 100 \times 100$; therefore, the voxel size was $0.0301 \times 0.03 \times 0.03$ mm. In the spin echo sequence, the repetition time, T_r was 1 s and the echo time, T_e , was 15 ms. The 90 degree pulse lasted 300 μ s at 15.5 dB and the 180 degree pulse lasted 600 μ s at 15.5 dB. Total scan time was approximately 11 hours. (Adapted from “The Basics of MRI” by Joseph Hornak³)

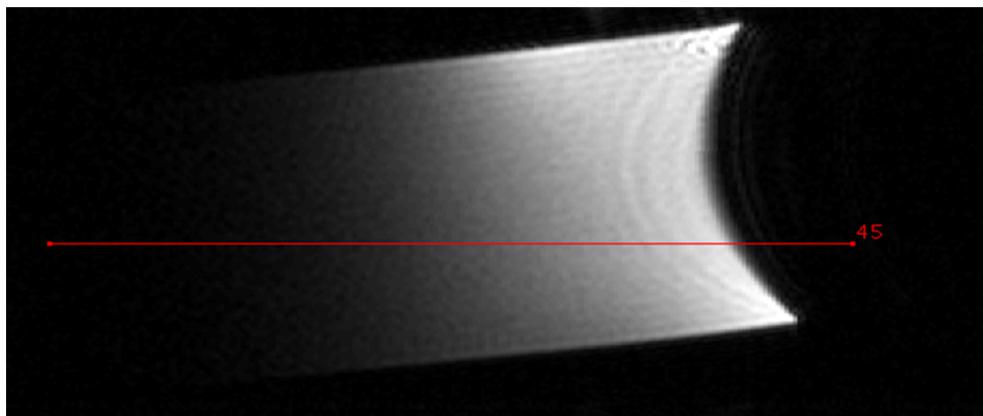


Figure 9: A meniscus is visible and clearly resolved, but it is a problem in assessing the uniformity of the B_1 magnetic field.

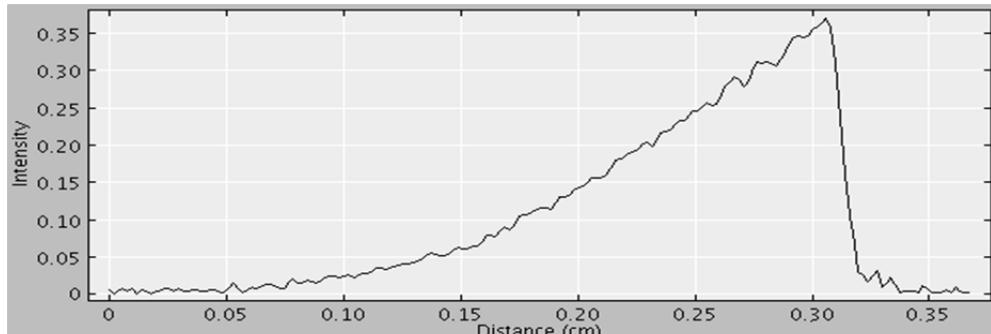


Figure 10: A graph of normalized intensity versus position begins similar to that predicted by models, but the meniscus causes a steep drop in intensity, which makes it impossible to see whether there is actually an “hour-glass” shape in the intensity of the magnetic field. Further scanning must be done.

A peculiarity visible in a different plane, shown in Figure 11, is a dark spot appearing towards the bottom right of the image. This is most likely due to the design flaw that the two Helmholtz loops are not complete circles. Using copper trace and two sides of a substrate, it is impossible to have two complete loops, as the trace must go through the board and continue on the other side so assure the circuit does not short. Besides the dark spot, the field passing through the midplane of the Helmholtz coil appears to be very uniform.

Some of the images (not shown) contained artifacts. It is unclear at this point in time what was causing those artifacts, whether it was a design flaw of the coil, material on the outside of the capillary, or the sudden jump in intensity at the point of the meniscus. However, further testing can be done on various subjects to better assess the validity and usefulness of the current design.

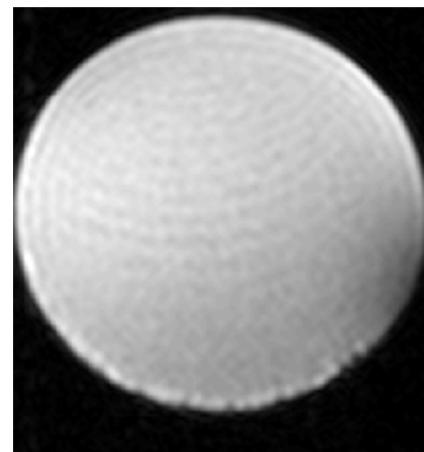


Figure 11: A dark spot appears in the bottom-right section of this image. This is probably due to the design of a copper-trace Helmholtz coil on a substrate. The coil is not a full 360 degrees, which causes a weakness in the B_1 field at the location of the gap.

DISCUSSION

As predicted, the Q factor of the varactors proved to be a limitation of their feasibility in tuning the resonance circuit. Though they seemed to be useful from the S_{11} sweep, they were not very useful according to the S_{21} sweep. The Q factor is inversely proportional with the frequency of the resonator, which means they are even less useful for MRI magnets stronger than 14 T.

However, the results of this study lead to numerous possible future projects. One such experiment would be to have the entire varactor setup, minus the

varactors. Replacing them with mechanical capacitors would and measuring S_{21} would prove if the extra traces and components are actually affecting the signal. The potential bias across the varactor may be interfering with the signal, which would mean that larger resistors must be used. If this is not the case, varactors may be unusable because the extra traces and components required add too much coupling to the circuit. Likewise, some of the signal may be lost through the entire loop created by the circuit on the substrate. Due to the lack of real-estate, the whole trace may be acting as its own inductor and lose some of the signal. This phenomenon was experienced to a small extent in the S_{21} parameter, though it was clear that the Helmholtz coil emitted most of the signal.

Another branch from the results of this study would be to experiment with different forms of digital tuning methods. A corollary to the Aeroflex Inc. hyperabrupt varactors used in this study would be high- Q abrupt-junction nonmagnetic varactors sold by companies, such as Cobham plc. Another possible component is a single-layer capacitor. Furthermore, it may actually be the silver paint used to bond the varactors to the trace that caused the issues at such a high frequency. This substance should be individually assessed to understand its properties at high frequency.

Future work may also include changing the Helmholtz coil design. With thicker materials, a successful Helmholtz coil with a 1.5-mm radius may be achievable. However, it may also be possible to wind thin wire to create a volume coil. The advantages of a wire-coil is that it would be very interchangeable, thus making interchangeable coils for various frequencies and it would be a round wire, as opposed to a flat trace where a magnetic field may not be created as uniformly. Due to the x - y grid that the laser etcher moves on, it had difficulty cutting a smooth arc on with such a small radius, leaving the edge of the trace jagged. The disadvantage of interchangeable coils is that they would not be physically as sturdy as a copper-trace coil fixed to a ceramic board. It is unclear whether this form of coil will reduce or augment the observation of artifact.

CONCLUSION

From the results gathered, it appears that images can be produced with varactors, but the signal-to-noise ratio of varactor-tuned circuits is not high enough to produce satisfying results for outside researchers looking to scan cells and other small objects. The varactor-tuned circuit can undergo more refinement to possibly produce better results, but the low Q of these components at high frequencies is a major pitfall. They do not have a very sharp S_{21} peak, which is indicative of the need for a change in either design or components. The Helmholtz coil showed the uniformity distribution predicted by the equations in MathCAD. This shows how the Helmholtz design is very sensitive to geometry and should be reconstructed more carefully.

Overall, using 10-pF isolating capacitors, 2-to-9-pF varactors, 1-M Ω resistors, and a 1.5-mm radius Helmholtz coil, a circuit can be tuned and impedance can be matched so that the circuit will resonate at 600 MHz when connected to a 50- Ω

coaxial cable. This is a useful result in the sense that it has pushed the learning curve of working with high frequency resonators, expanded a database for troubleshooting, and pointed future work in directions that will be more successful and closer to commercial transfer. Some potential applications of this research in the technological environment include the ability to study flow, diffusion, and perfusion of chemicals in real time, improved signal-to-noise ratio to understand structures of microscopic tissues, and alternative methods for finely tuning antennas.

ACKNOWLEDGMENTS

M. Feldman would like to thank Jeff Long and Amanda Baker for their knowledge and generosity with equipment that supported the project. Also, he would like to thank Haolun Zhang for his theoretical model and help in CST Microwave Studio. Lastly, he would like to thank Thomas Neuberger for his contribution in using the 14-T machine to scan a phantom. This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984.

REFERENCES

- [1] Weber, H., N. Baxan, D. Paul, J. Maclare, D. Schmidig, M. Mohammadzadeh, J. Hennig, and D. Von Elverfeldt. "Microcoil-based MRI: Feasibility Study and Cell Culture Applications Using a Conventional Animal System." *Magnetic Resonance Materials in Physics, Biology and Medicine* 24.3 (2011): 137-45.
- [2] Haines, Kristina Noel. *Applications of High Dielectric Materials in High Field Magnetic Resonance*. Dissertation. The Pennsylvania State University, 2010.
- [3] Hornak, Joseph P. *The Basics of MRI*. Henrietta, NY: Interactive Learning Software, 1996. Web. <<http://www.cis.rit.edu/htbooks/mri/>>.

AUTHOR INDEX
(* Indicates REU Student Author)

<i>Adami Ardestani, Bahareh</i> *	129
<i>Bilén, Sven G.</i>	67, 87
<i>Boehmer, Tyler</i>	87
<i>Capuro, Robert</i>	67, 87
<i>Celestin, Sébastien</i>	1
<i>Conway, Michael</i>	87
<i>Curwen, Christopher A.*</i>	1
<i>DiMarino, Christina*</i>	141
<i>Feldman, Matthew*</i>	169
<i>Galindo, Freddy</i>	25, 41
<i>Galvan, Christopher A.*</i>	67
<i>Hallen, Hans</i>	157
<i>Hong, Christopher*</i>	25
<i>Jenkins, W. Kenneth</i>	97
<i>Jones, Christopher W.*</i>	57
<i>Kendrick, Chito</i>	141
<i>Kummer, Allen</i>	67
<i>Lanagan, Michael</i>	169
<i>Liu, Zhiwen</i>	157
<i>Mathews, John D.</i>	57
<i>McDonough, Sean S. *</i>	13
<i>Min, Hwall</i>	129
<i>Nan, Ruth*</i>	157
<i>Olique Ortiz, Luis E. *</i>	87
<i>Pasko, Victor P.</i>	1, 13
<i>Perini, Steve</i>	169
<i>Qin, Jianqi</i>	13
<i>Redwing, Joan</i>	157
<i>Ressetar, Dimitri*</i>	41
<i>Salvia, David</i>	97
<i>Schiano, Jeffrey L.</i>	117
<i>Sova, Danielle*</i>	97
<i>Tadigadapa, Srinivas</i>	129
<i>Tyson, Tom</i>	117
<i>Urbina, Julio V.</i>	25, 41
<i>Wilson, Jacob*</i>	117
<i>Xu, Wei</i>	1
<i>Yang, Chuan</i>	157

