Annual Research Journal

Electrical Engineering
Research Experience for Undergraduates

Vol. XI

August 2013

National Science Foundation
Grant No. EEC-1062984

PENNSTATE

Department of Electrical Engineering
University Park, Pennsylvania

EEREU Annual Research Journal

Sven Bilén (editor)

Volume XI
Published in 2013 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/

PREFACE

We are happy to present the ***2013 Annual Research Journal – Electrical Engineering Research Experience for Undergraduates, Vol. XI.*** This volume contains highlights of the EEREU program activities for summer 2013 and twelve technical papers written by EEREU scholars as primary authors.

Summer 2013 marks our eleventh year hosting the National Science Foundation–sponsored Research Experience for Undergraduates (REU) Site Program, in the Department of Electrical Engineering, Penn State. Twelve 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 2013 Annual *EEREU Symposium*, held at University Park, in July 2013.

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 2013
University Park, PA

TABLE OF CONTENTS

Research Articles (* indicates REU student author)

2013 NSF EEREU FACULTY & STAFF MEMBERS

Faculty Mentors:

Prof. Sven Bilén, Co-Director
Prof. Ken Jenkins, Co-Director
Prof. Noel Giebink
Prof. Zhiwen Liu
Prof. John Mathews
Prof. Joan Redwing
Prof. David Salvia
Prof. Victor Pasko
Prof. Jeff Schiano
Prof. Srinivas Tadigadapa
Prof. Julio Urbina

Prof. Phil Boyer, Entrepreneurship Chair
Ms. Amy Freeman, Adviser, Engineering Diversity
Mrs. Lena Getman, Administration
Prof. Andy Lau, Ethics Chair
Prof. Jack Mitchell, Seminar Chair
Prof. David Salvia, Activities and Web Chair
Mr. Jared Price, Assistant Director

NSF EE REU 2013 Summer Program
Research Seminar Series

Room 101 E E East, 9:40 – 10:30 AM

(*Seminars are on Tuesday unless otherwise noted*.)

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

Date	***Topic***	***Speaker***
June 4, 2013	Sensors Using Micro and Nanoscale Structures	Srinivas Tadigadapa
June 11, 2013	Solar Energy & the Terawatt Challenge	Chris Giebink
June 18, 2013	Magnetic Resonance Engineering	Jeff Schiano
June 25, 2013	What's Next? Graduate School, of Course!	David Salvia
July 2, 2013	Systems Design Lab	Sven Bilén
July 9, 2013	Lightning-related Transient Luminous Events in the Middle Atmosphere	Victor Pasko
July 16, 2013	Emerging Techniques in Radar Remote Sensing: Cognitive Radars and Beyond	Julio Urbina
July 25, 2013 (Thursday)	2013 NSF EEREU Symposium	REU Students

NSF EEREU 2013 Summer Program
Ethics & Entrepreneurship Workshops

Room 101 E E East, 10:40 – 11:50 am, Tuesdays

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

Date	***Topic***	***Speaker***
June 4, 2013	Introduction to Ethics: The Responsible Business	Andy Lau
June 11, 2013	Research Ethics	Andy Lau
June 18, 2013	Intellectual Property Issues	Phil Boyer
June 25, 2013	Student Entrepreneur Panel: "What's it take to be an entrepreneur?"	Erik Davidson
July 2, 2013	Sustainability	Andy Lau
July 9, 2013	Ethics Issues in Entrepreneurship	Andy Lau
July 16, 2013	Getting Funding to Pursue Your Idea	Phil Boyer
July 23, 2013	Student Presentations on Tech Transfer Ideas of Research Projects	REU students

2013 NSF EE REU Field Trip Program
Academic/Industrial Sponsors and Tour Hosts

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

Breazeale Nuclear Reactor
Penn State University, University Park, Pennsylvania

Accuweather, Inc.
State College, Pennsylvania

Penn State Center for Sustainability
State College, Pennsylvania

State of the Art, Inc.
State College, Pennsylvania

Energy Efficiency Buildings HUB
Philadelphia, Pennsylvania

NextFab Studio
Philadelphia, Pennsylvania

W. L. Gore and Associates
Elkton, Maryland

Videon Central
State College, PA

2013 NSF EE REU SYMPOSIUM

Final Program

8:30 am to 3:00 pm, Thursday, July 25, 2013
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/ Jared Price)**
8:55 – 9:00 am	**Welcome**	**Ken Jenkins**
9:00 – 10:00 am	**Session I** Session Chairs:	**Pasko/Urbina/ Mathews**
9:00– 9:15	ABSOLUTE POWER CALIBRATION OF THE PENN STATE METEOR RADAR SYSTEM FOR METEOR STUDIES (MENTOR: URBINA)	Zachary Morgan
9:15 – 9:30	DEVELOPMENT OF A SOFTWARE-DEFINED RADAR RECEIVER FOR IONOSPHERIC STUDIES USING INTERFEROMETRIC TECHNIQUES (MENTOR: URBINA)	Benjamin Young
9:30 – 9:45	DEVELOPMENT OF AN EFFICIENT REFERENCE CONDUCTIVITY MODEL FOR GLOBAL ELECTRIC CIRCUIT MODEL CALCULATIONS (MENTOR: PASKO)	Jeremy Pachter
9:45 – 10:00	WE'RE LUCKY TO HAVE THE MOON (MENTOR: MATHEWS)	Niklas Anthony
10:00 – 10:15 am	**Coffee Break**	
10:15 – 11:15 am	**Session II** Session Chairs:	**Bilén/Jenkins/ Schiano**
10:15 – 10:30	COMPARISONS OF ADVANCED ADAPTIVE SIGNAL PROCESSING METHODS FOR REMOVING MATERNAL INTERFERENCE NOISE FROM FETAL ELECTRO-CARDIOGRAMS (MENTORS: JENKINS AND SALVIA)	Ryan M. Collins
10:30 – 10:45	CHARACTERIZATION OF QUADRUPOLE RESONANCE TRANSITIONS IN CYANURIC ACID (MENTOR: SCHIANO)	Amy Hein
10:45 – 11:00	SYSTEM-LEVEL DESIGN OF ACTIVE INFLATABLE FALLING SPHERE FOR USE WITH UPCOMING SUB-ORBITAL FLIGHT OPPORTUNITIES (MENTOR: BILÉN)	Gregory Roberts
11:00 – 11:15	DESIGN OF A MAGNETIC TORQUER SYSTEM FOR CUBESATS (MENTOR: BILÉN)	Cynthia Rojas

2013 NSF EE REU SYMPOSIUM (cont.)

8:30 am to 3:00 pm,

Final Program

Thursday, July 25, 2013
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.)**	(Jared Price)
12:00 – 1:30 pm	**Luncheon (Nittany Lion Inn)**	
1:45 – 2:45 pm	**Session III** Session Chairs:	**Tadigadapa/Liu/ Redwing/Giebink**
1:45 – 2:00	DEVELOPMENT OF A USER INTERFACE BASED PROGRAM TO TRACK IMPEDANCE CHARACTERISTIC OF A MICROMACHINED QUARTZ RESONATOR (MENTOR: TADIGADAPA)	Thenmozhi Elayaperumal
2:00 – 2:15	ETCH RATE AND GEOMETRY FOR ETCHED SI(001) SUBSTRATES FOR SUBSEQUENT GAN FILM GROWTH (MENTOR: REDWING)	Sophia Williams
2:15 – 2:30	AUTOMATED CONTROL OF A G-FRESNEL SPECTROMETER (MENTOR: LIU)	Bill Kim
2:30 – 2:45	SCATTERING ELEMENT DESIGN FOR SOLAR THERMAL CONCENTRATORS (MENTOR: GIEBINK)	Mason Sutorius
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:	Jared Price 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 24. It is strongly recommended that speakers test-run presentation files for audio, video, or special applications on a Windows PC prior to their Thursday presentation. Each presentation is 15 minutes – chair introduction, questions, and answers included.

NSF EE REU PENN STATE Annual Research Journal Vol. XI (2013)

DEVELOPMENT OF AN EFFICIENT REFERENCE CONDUCTIVITY MODEL FOR GLOBAL ELECTRIC CIRCUIT MODEL CALCULATIONS

Jeremy Pachter,* Sotirios Mallios,[+] Andreas Baumgaertner,[§] and Victor Pasko[#]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Columbia University
New York, NY 10027

ABSTRACT

The Global Electric Circuit (GEC) is a circuit that is formed between the Earth's surface, which is a good conductor of electricity, and the ionosphere, a weakly ionized plasma at ~80 km altitude [e.g., Rycroft et al., 2008]. In the absence of any source, the GEC behaves as a leaky spherical capacitor, with the ground being the negatively charged plate and the ionosphere the positively charged plate, which discharges through the weakly conducting atmosphere creating fair-weather current which is about 1 kA integrated over the entire Earth's surface [e.g., Bering et al., 1998].

In the Earth's atmosphere the vertically integrated resistivity is called columnar resistance. This quantity is inversely proportional to the atmospheric conductivity, which increases (usually exponentially) with altitude. Since the fair weather current is linked with the columnar resistance according to Ohm's law, atmospheric conductivity is a very important quantity in the studies of the GEC. Columnar resistance is related to total global resistance, but no global measurements of resistance are possible. Therefore, several extensive conductivity models have been derived from calculations based on measurements of galactic cosmic ray ionization rates, ion recombination rates and ion mobility approximations with the inclusion of aerosol, cloud, and radioactive decay influence [e.g., Tinsley et al., 2006, Zhou et al., 2010, McDonough et al., 2011].

[#] Faculty Mentor
[+] Graduate Mentor
[§] Collaborator, Department of Engineering, University of Colorado, Boulder

However, these models are computationally complicated and provide data that require additional processing before they can be used in GEC models.

Our goal is to create a fast and efficient algorithm that interpolates and extrapolates conductivity data derived from the Whole Atmosphere Community Climate Model (WACCM) [Baumgaertner et al., 2013] at user defined latitudes and longitudes to provide the conductivity distribution along altitude.

INTRODUCTION

The Global Electric Circuit (GEC) was proposed by Wilson [1920] to describe fair weather electricity, and has since been extensively studied. As a product of global thunderstorm activity, the GEC serves as a sensitive indicator of global land surface temperature and moisture levels related to climate change [Bering et al., 1998]. It has been speculated that the circuit itself may cause climate change via electrical effects on cloud microphysics [Tinsley, 2000, 2004; Tinsley and Yu, 2004]. The circuit is formed between the Earth's surface, which is a good conductor of electricity, and the ionosphere, a weakly ionized plasma at ~80 km altitude [e.g., Rycroft et al., 2008]. The ionosphere is held at a constant potential approximately 250 kV above ground, due to charging from thunderstorms, the dynamo effect between the solar wind and the magnetosphere, and the dynamo effect of atmospheric tides [Bering et al., 1998]. A fair weather current density of approximately 1 pA/m^2 discharges the ionosphere, sustaining a global current of 1–2 kA.

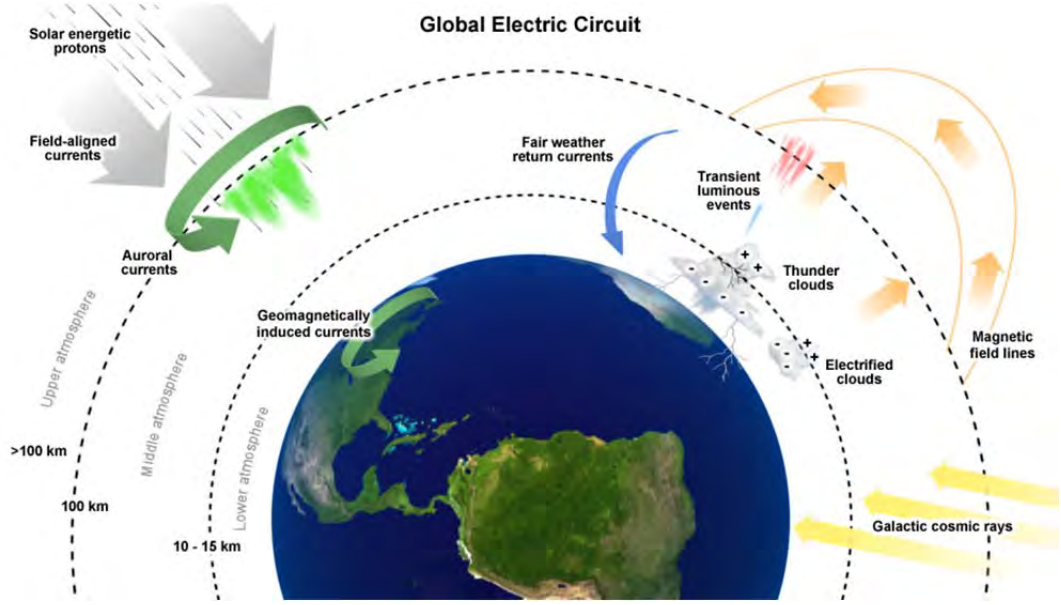

Figure 1: The GEC [http://sisko.colorado.edu/FESD].

Electrically charged constituents are responsible for making the atmospheric conductivity high enough to allow electrical pathways for currents to flow

between the surface of the Earth, clouds, and the ionosphere. Since the fair weather current is linked with the columnar resistance according to Ohm's law, atmospheric conductivity is a very important component in the studies of the GEC. For calculating conductivity, size-dependent aerosol distributions are necessary. The WACCM is a part of the Community Earth System Model (CESM1.1) with contributions from the Community Aerosol and Radiation Model for Atmosphere (CARMA). CARMA simulates all aspects of aerosol life cycle including nucleation, condensational growth, coagulation, and deposition [Baumgaertner et al., 2013]. An improved external dataset allows consideration of the following ion sources: galactic cosmic rays (GCR), solar energetic particles, natural radioactivity from direct α, β, and γ radiation as well as the decay of radioactive gases, and the Radon isotope ^{222}Rn [Baumgaertner et al., 2013].

Figure 2 shows a schematic diagram of the GEC as well as the discrete column resistance after taking into account the major factors that influence the conductivity. The geometry of the global circuit can be treated as plane-parallel because the radius of the Earth is large compared to the thickness of the atmosphere. The discrete column resistances shown in Figure 2 represent the atmospheric continuum at equatorial (E), low (L), high (H), and polar (P) latitudes [Zhou and Tinsley, 2010].

There is a need for a fast and efficient reference conductivity model that interpolates and extrapolates data from the more general Whole Atmosphere Community Climate Model (WACCM) [Baumgaertner et al., 2013]. The present work introduces a new reference model that provides the conductivity distribution along altitude at user defined latitudes and longitudes. The model is tailored specifically to provide the necessities for GEC modeling, and offers the most ergonomic interface for conductivity profiles. Furthermore, the model realistically handles problematic conductivity inputs at the user's discretion, allowing in-depth analysis of resulting GEC behavior.

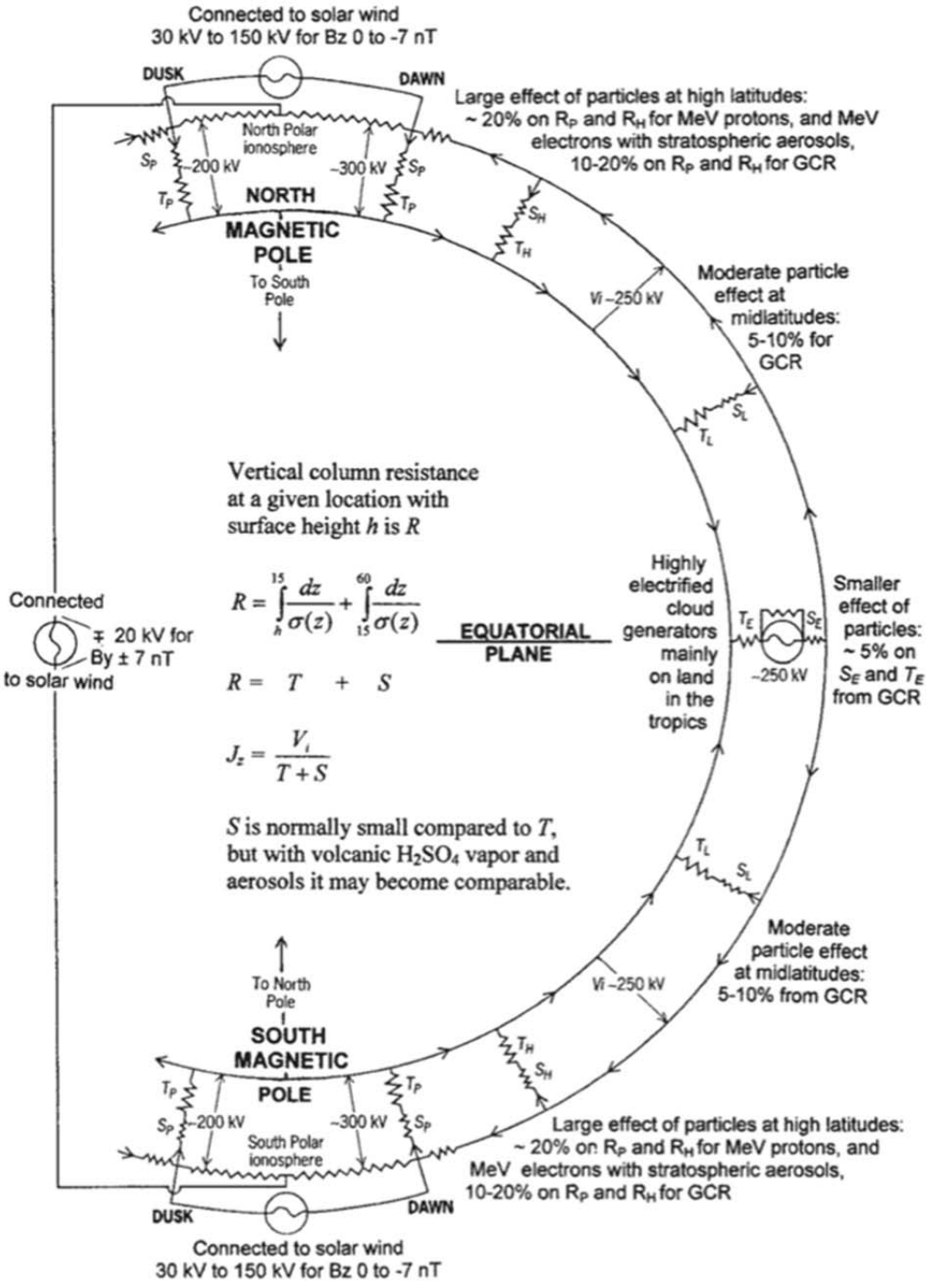

Figure 2: GEC with major factors that influence the conductivity and thus the resistance (aerosols, GCR, and solar energetic particles) [Zhou and Tinsley, 2010].

MODEL DESCRIPTION

The Whole Atmosphere Community Climate Model (WACCM) file contains four dimensions and ten variables. The dimensions are time, altitude, latitude, and longitude; the variables are (1) conductivity with mean galactic cosmic ray flux (GCR), aerosols, radon, and clouds, (2) time, (3) altitude, (4) latitude, (5) longitude, (6) conductivity with mean GCR, aerosols, without clouds or radon, (7) conductivities with maximum GCR, aerosols, radon, and without clouds, (8) conductivities with minimum GCR, aerosols, radon, and without clouds, (9) conductivities with mean GCR, radon, without aerosols or clouds, and (10) conductivities with mean GCR, aerosols, radon, without clouds. The variables are manipulated based on user input to provide the basis of our MATLAB function. Figure 3 shows the help documentation of the function with all the input parameters listed and explained.

```
%   function s = condw(lat,long,month,h,type,condfill,mountain)
%
%   Function provides an interpolated conductivity vector
%   corresponding to user-defined altitudes ranging from 0 to 60 km.
%
%   s           - conductivity array at user-defined altitudes corresponding to
%                 array of altitudes h (S/m);
%
%   lat         - latitude value ranging from -90 (S) to 90 (N) degrees;
%
%   long        - longitude value ranging from -180 (W) to 180 (E) degrees;
%
%   month       - month value ranging from 1 (Jan) to 12 (Dec);
%
%   h           - array of altitudes ranging from 0 to 60 km (km);
%
%   type        - mean GCR, with aerosols, radon, and clouds conductivities     (0)
%               - mean GCR, with aerosols, with radon, no clouds conductivities (1)
%               - mean GCR, with aerosols, no radon, no clouds conductivities   (2)
%               - mean GCR, no aerosols, with radon, no clouds conductivities   (3)
%               - max GCR, with aerosols, radon, no clouds conductivities       (4)
%               - min GCR, with aerosols, radon, no clouds conductivities       (5)
%
%   condfill    - conductivity value to fill in the mountains (S/m) (if 0,
%                 mountains filled with conductivity values at 60 km altitude).
%                 Typical conductivities for mountains are:
%
%                 Pure ice                          - 1e-7 S/m
%                 Soft new snow, density 0.13 g/cm^3 - 1e-9 S/m
%                 Granular snow, density 0.4 g/cm^3  - 1e-7 S/m
%                 Compact wet snow, Colorado         - 1e-6 S/m
%                 Glacial ice, Alberta               - 5e-7 S/m
%                 Dry sand or rock                   - 1e-5 S/m
%                 Frozen earth, Antarctica           - 5e-4 S/m
%                 Partially frozen earth, Alaska     - 2e-3 S/m
%
%                 Conductivity values obtained from [Evans, S., Dielectric properties
%                 of ice and snow - a review, J. Glaciol., 5(42), pp. 773-792, 1965].
%
%   mountain    - remove mountains by performing a series of one-dimensional
%                 linear interpolations at points directly outside of the mountain,
%                 averaging the values, and producing a reasonable conductivity value to
%                 represent that altitude (0) or fill up mountains with conductivity
%                 values corresponding to the condfill parameter (1);
%
%   The function was created by Jeremy Pachter on June 21, 2013 and is
%   based on results obtained by [Baumgaertner et al., 2013].
```

Figure 3: Help documentation of the function.

The user inputs latitude, longitude, month, array of altitudes, cloud options, mountain conductivity values, and mountain handle technique. Latitude values range from −90° (S) to 90° (N), longitude values range from −180° (W) to 180°

(E), month values range from 1 (January) to 12 (December), and the array of altitudes range from 0 to 60 kilometers. This altitude range is chosen because GEC models extend to this region, which maps out the lower boundary of the ionosphere. Above this region, potential is approximately constant. Cloud options (type) correspond to conductivity with mean GCR, aerosols, radon, and clouds (0), mean GCR, aerosols, radon, and no clouds (1), mean GCR, aerosols, no radon or clouds (2), mean GCR no aerosols, with radon, and no clouds (3), max GCR, aerosols, radon, and no clouds (4), or minimum GCR, aerosols, radon, and no clouds (5). Mountain conductivity values (condfill) are used to fill in mountains with user-defined conductivities. If condfill is set to 0, the function fills in mountains with conductivity values corresponding to 60 km altitude. Mountain handle technique (mountain) allows the user to fill up mountains with conductivity values corresponding to the condfill parameter (1), or remove mountains entirely to produce reasonable conductivity values along altitude (0).

The technique that removes the mountains operates by initially creating an interpolated conductivity vector at 12-km altitude, since there is no mountain influence at this altitude. Mountains are distinguished from atmosphere by their large fill-up values, which are introduced during the transformation of coordinate systems from pressure to altitude in the WACCM file, so we set the points inside and directly outside the mountain to high conductivity values. A box consisting of four points is defined around the point of interest, corresponding to the first latitude and longitude points that exist in the original file. Keeping longitude constant, the first latitude points to the left or right of each of the four box points are found. Similarly, keeping latitude constant, the first longitude points above or below each of the four box points are found. This defines the mountain morphology, and we perform a bilinear interpolation to obtain a conductivity value for the point in question if the points are aligned in longitude or latitude. If no points are found after a user specified number of points away from the point of interest, we extrapolate using values at higher altitudes to obtain values inside of the mountain. Figure 4 shows this technique in the form of a grid. As we search for the first points beyond the influence of the mountain, we must generally go further away from the mountain in latitude and in longitude. Figures 5b and 5d show that conductivity varies more over latitude than longitude. Figures 5a and 5c show that conductivity varies more as we decrease in altitude. Therefore, longitude interpolation is preferred because conductivity varies less as a function of longitude than as a function of latitude. The further away from the mountain region we are, the greater the probability of obtaining a conductivity profile that is not smooth. Thus, we limit the scope of our latitude search to nearby points by default. It is important to note that variation in conductivity due to mountains is more prominent as we decrease in altitude, and is dependent on the morphology of the ground. Since the WACCM file lacks conductivity values at altitudes less than 300 meters, we perform a one-dimensional linear extrapolation on log scale between 300-m altitude and sea level (0-km altitude) as a means to complete our conductivity profile.

We emphasize that the mountain removal feature is not intended to create an accurate representation of conductivity, because mountain removal is not physical. We perform the mountain removal to have a reference for conductivity in GEC models; the model's primary concern is smoothness of conductivity profile. More sophisticated regression and prediction techniques must be used to increase accuracy in conductivity profiles. Since we aim to show basic GEC properties, these methods are beyond the scope of this project.

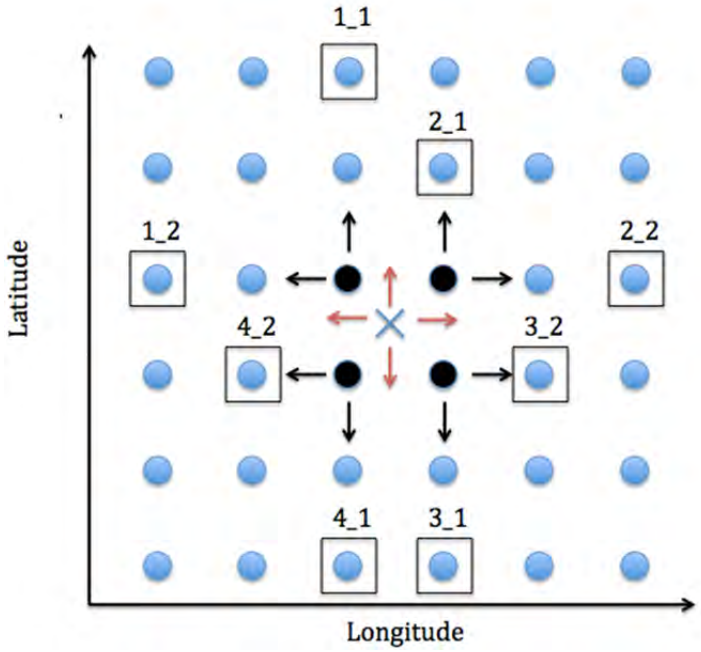

Figure 4: A box is formed around the desired point, formed by the 4 black circles. Searches along latitude and longitude are performed as indictaed by the red and black arrows. The first points outside the mountain are found, as represented by the boxed in points. The numbers above the box represent the longitude and latitude positions respectively.

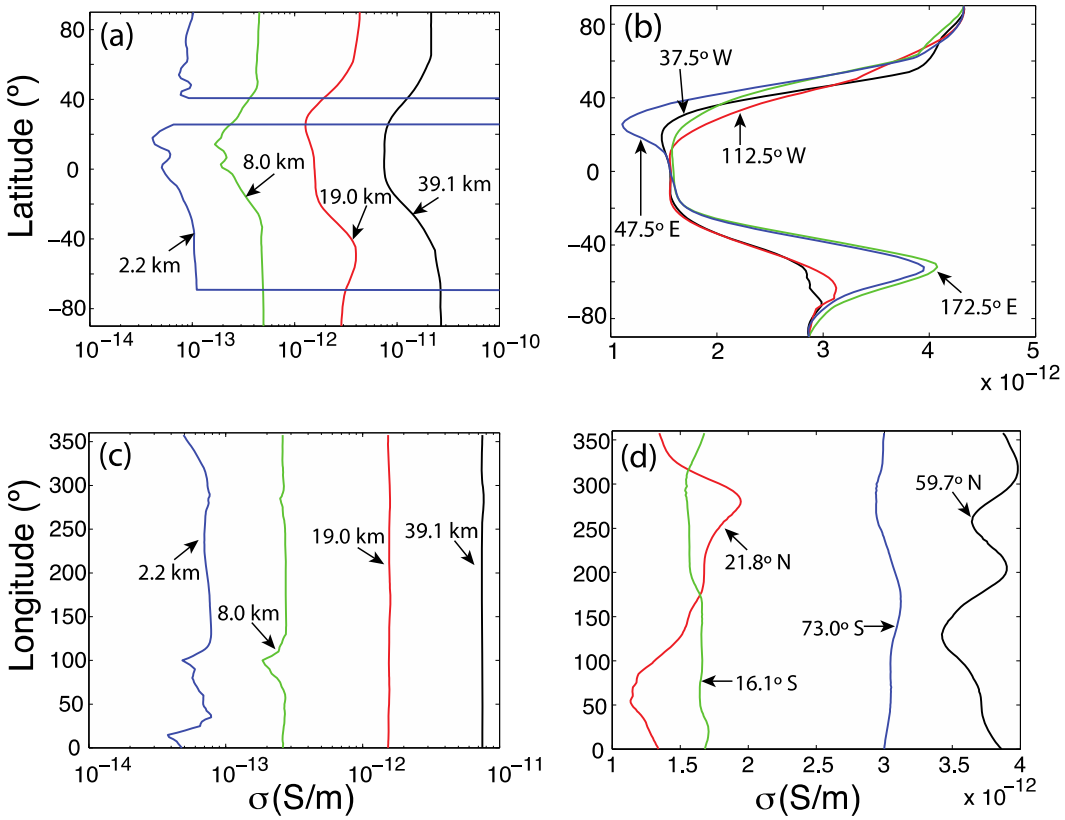

Figure 5: Conductivity variations as a function of (a) latitude at various altitudes and 97.5° E, (b) latitude at various longitudes and 19 km altitude, (c) longitude at various altitude and 2.8421° N, and (d) longitude at various latitudes and 19-km altitude.

The treatment of mountains in the global conductivity profile is important, since the majority of global resistivity is accounted for in the lower-altitude region [e.g., Tinsley et al., 2006]. In Figure 6 we plot conductivity profile as a function of altitude with and without mountains. To remove mountains, we use the technique described above, and to keep mountains, we fill the mountains with conductivities corresponding to 60-km altitude. Figure 6a shows results for the Sierra Mountains in Nevada, at 37.00° N, 119.00° W, and Figure 6b shows the results for the Rocky Mountains in Colorado, at 39.12° N, 106.45° W.

In the absence of pollutants and aerosols, the atmospheric conductivity profile increases exponentially with altitude. The profile follows a linear trend on a log scale, so we perform a three-dimensional linear interpolation of the logarithm of conductivity to provide a conductivity vector corresponding to the input latitude and longitude at each altitude entered by the user. We have tried other kinds of interpolation as well (e.g., cubic, spline), but in principle, the exponential profile of the conductivity makes the linear interpolation on log scale the most robust and

accurate. The linear interpolation obtains the interpolated value by taking a weighted average of the nearest points (weighted by distance).

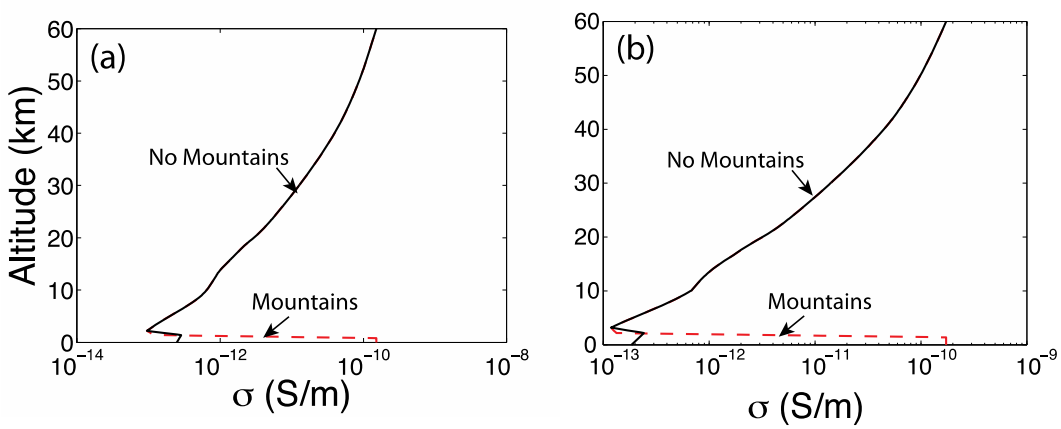

Figure 6: Conductivity as a function of altitude with and without mountains for (a) Sierra Mountains (37.00° N, 119.00° W) and (b) Rocky Mountains (39.12° N, 106.45° W).

RESULTS

We compare the conductivity profile as a function of altitude obtained in the present work with the conductivity profile obtained using the simulations of Tinsley and Zhou [2006] under similar conditions. We remove mountains using our interpolation based on nearby values and include aerosol, radon, and mean GCR contributions in our simulation. Using Tinsley and Zhou [2006] simulations, we include aerosol and radon contributions, and average solar maximum and minimum treatments to match our mean GCR condition. In Figure 7 we compare simulations above land and sea without clouds, and during winter and summer.

In Figure 7c, we notice a knee from ~12 to ~20 km that exists in our simulation and not in the Tinsley and Zhou [2006] model. At 16.3 km, the conductivity value obtained from the Tinsley and Zhou [2006] model is a factor of 2.33 times larger than the conductivity value obtained in our simulation. At 1.6 km altitude, the conductivity value obtained from [Tinsley and Zhou, 2006] is a factor of 3.06 times larger than the conductivity value obtained in our simulation. Other than these discrepancies, the behavioral trends of the two models are similar.

In Figure 8 we compare the conductivity profiles in July and January with cloud influence to the profiles without clouds above the Sahara Desert (20° N, 30° E) and above the South Atlantic Ocean (20° S, 10° W). Above the desert in July at 13 km altitude, we notice that the influence of the clouds lowers the conductivity by ~55.8% when compared to the conductivity without clouds (Figure 8a). Above the ocean there is a difference between conductivities from ~1.2 to ~4.2 km. At 3.1 km, the conductivity without clouds is 1.28 times larger than conductivity with clouds (78.2 % decrease in conductivity due to inclusion of

clouds), as illustrated in Figure 8b. Above this range, the values are in agreement. Above the desert in January, we notice a similar behavior to Figure 8b, and above the ocean during this month there is effectively no difference between conductivity profiles with and without cloud influence.

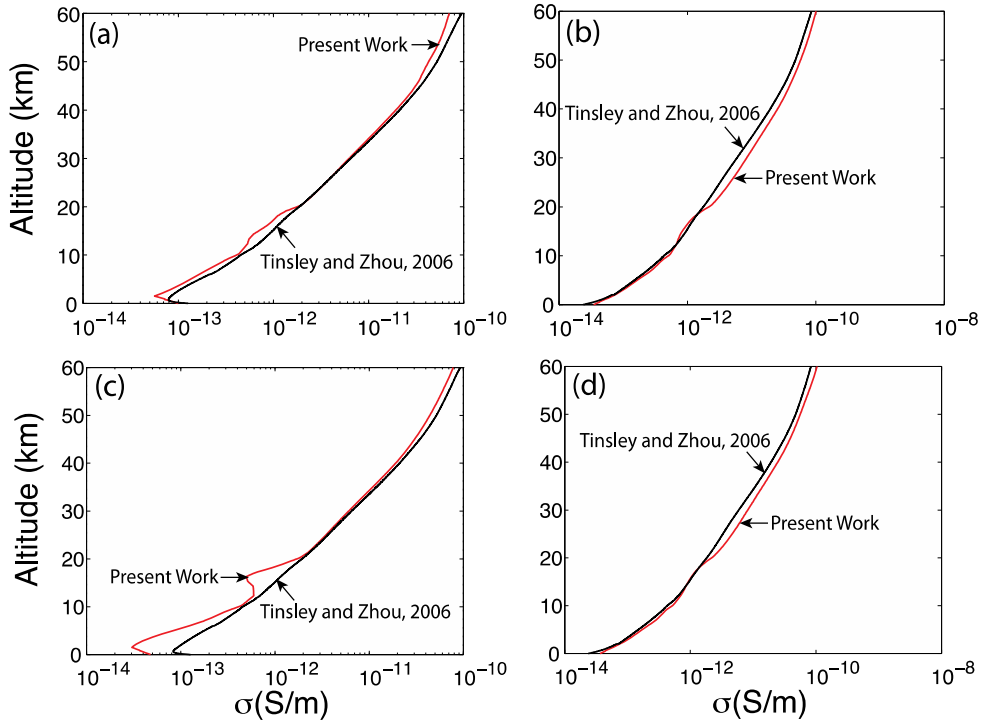

Figure 7: Comparison with conductivity model from Tinsley and Zhou [2006] for January at (a) Sahara Desert, 20° N, 30° E, (b) South Atlantic Ocean, 20° S, 10° W, and June at (c) Sahara Desert, and (d) South Atlantic Ocean.

In Figure 9 we compare the fair-weather total global resistance with and without mountains as a function of month. During each month, the total global resistance is larger when we neglect mountains.

DISCUSSION

In Figures 7a, 7b, and 7d we see that the conductivity profiles obtained by our model are very similar to the profiles obtained by the Tinsley and Zhou [2006] model. The differences that do exist can be attributed to the differences in model calculation techniques. To produce the correct category of topography, the Tinsley and Zhou [2006] model allows the user to choose various types of terrain varying from maritime clean to desert. The model treats aerosol influence differently from our treatment, and defines a layer of ultrafine particles centered on 40 km altitude, with the peak concentration increasing linearly with latitude from ±40° to the geographic poles. In our model, we use latitude and longitude

coordinates to define location, and the corresponding climate is incorporated into the model based on these coordinates.

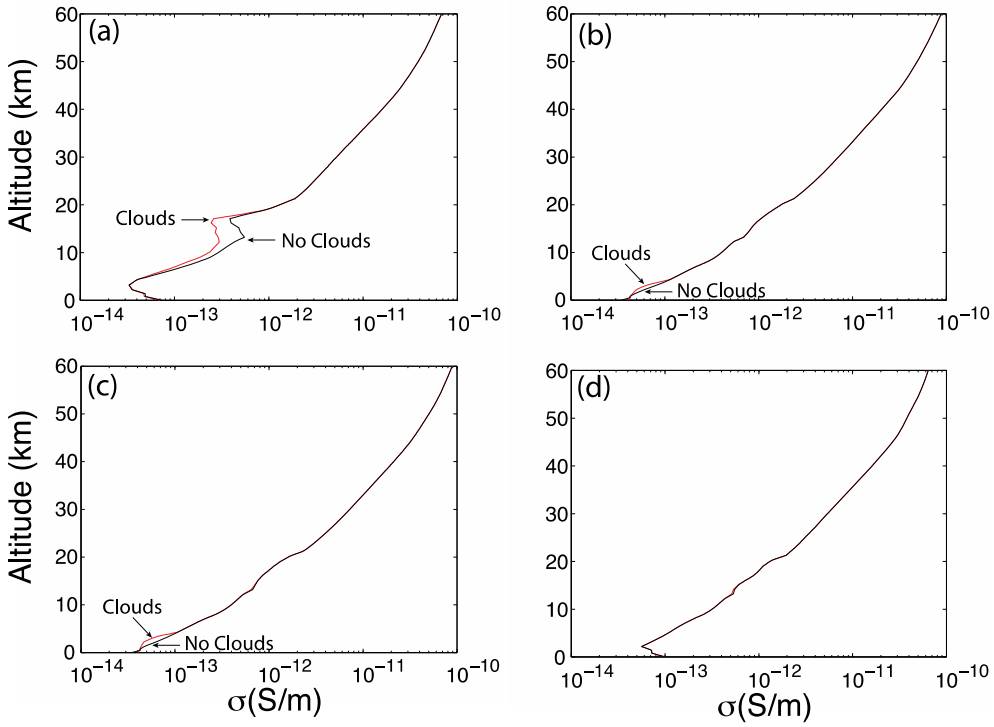

Figure 8: Comparison of conductivity profiles with and without clouds in July at (a) Sahara Desert and (b) South Atlantic Ocean, and January at (c) Sahara Desert and (d) South Atlantic Ocean.

The aerosol data used by Tinsley and Zhou [2006] was based on the GADS (Global Aerosol Data Set) database and assumed size distributions, the cloud information originated from the ISCCP (International Satellite Cloud Climatology Project). Our model enhances this treatment by calculating conductivity using a Earth System Model framework that provides evaluated temperatures, cloud cover, aerosol size distributions, radon distributions, and ionization rates [Baumgaertner et al., 2013].

For aerosol treatment, we do not implement a layer at 40 km; we use an aerosol model that describes the entire life cycle and transport of all important aerosols (CARMA in WACCM). In Figure 7c the knee that exists in our simulation can be attributed to the aerosol model we implement; the Tinsley and Zhou [2006] model includes a relatively coarse aerosol influence below 30 km, which does not create this knee.

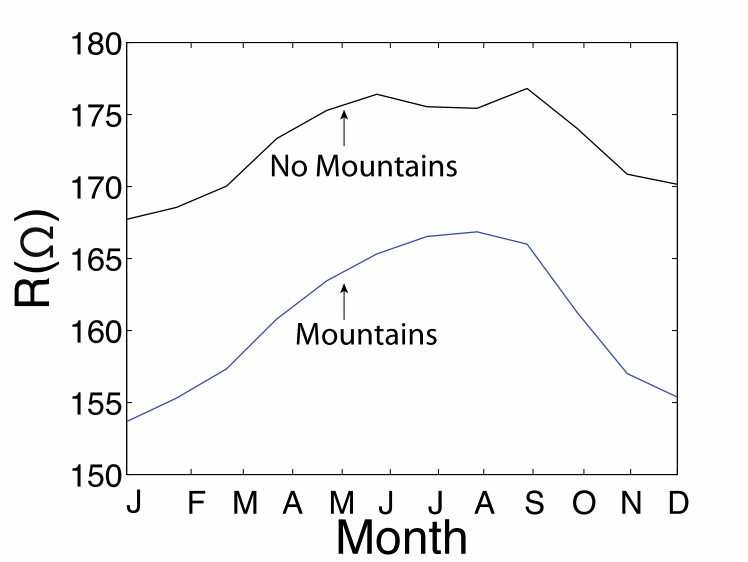

Figure 9: Total global resistance (fair-weather) as function of month with and without mountains.

In Figure 8a, we notice that above the desert in July at 13-km altitude, the influence of the clouds lowers the conductivity by ~55.8% when compared to the conductivity without clouds. This result comes from the fact that the majority of influential clouds exist near this altitude, and cloud effects on conductivity are implemented as a loss of conductivity by a factor of 1/60 in all types of clouds except deep convective clouds [Baumgaertner et al., 2013]. In Figures 8b and 8c, the slight differences in conductivities are due to the nature of clouds as a function of location, altitude, and month. In July over the South Atlantic Ocean, the cloud influence is not overwhelming, since the region is relatively dry; humidity plays a large role in cloud formation. In Figure 8d, the conductivity profiles are effectively the same, since cloud formation is at a minimum in the winter due to low temperatures and low humidity.

In Figure 9, the total global resistance without mountains is larger than the resistance with mountains, since resistance is inversely proportional to conductivity. Column resistance is the reciprocal of conductivity integrated vertically

$$R_c = \int_0^{60000} \frac{dz}{\sigma(z)} \left[\Omega\,\mathrm{m}^2\right] \tag{1}$$

where dz is the model layer thickness, σ is conductivity in S/m as a function of altitude, and we integrate from 0 to 60 km. Since resistance decreases exponentially with increasing altitude, conductivity in the lower atmosphere mainly determines column resistance [Baumgaertner et al., 2013]. From the column resistance, we calculate total resistance as

$$R_{\text{total}} = \left(\iint \frac{d\varphi d\lambda}{R_c(\varphi, \lambda)} \right)^{-1} [\Omega] \tag{2}$$

where $d\phi$ and $d\lambda$ are the model longitude and latitude grid sizes in km. We convert latitude and longitude from degrees to km by using the following formulas [Rapp, 1991, pp. 12-49]:

$$d\lambda = (1111132.954 - 559.822\cos 2\theta + 1.175\cos 4\theta)\Delta\lambda \tag{3}$$

$$d\phi = \left(\frac{\pi 6378137 \cos\theta}{180} \right) \Delta\phi \tag{4}$$

where θ is the latitude in degrees, and $\Delta\lambda$, $\Delta\phi$ are the model latitude and longitude grid sizes in degrees.

CONCLUSIONS

In conclusion, a fast and efficient reference conductivity model that interpolates and extrapolates data from the more general Whole Atmosphere Community Climate Model (WACCM) has been developed. The model is tailored specifically to provide the necessities for GEC modeling, and offers the most ergonomic interface for conductivity profiles. Furthermore, the model realistically handles problematic conductivity inputs at the user's discretion, allowing in-depth analysis of resulting GEC behavior. The model offers the user access to conductivity values along latitude, longitude, and altitude that result from inclusion of radon, aerosols, clouds, and various GCR fluxes. The user can pick and choose which processes to include, providing a powerful tool for investigating the most influential processes in the GEC. The model compares well to existing conductivity models [Tinsley and Zhou, 2006], and uses up to date models and data.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. EEC-1062984 and AGS 1135446.

REFERENCES

Baumgaertner, A. J. G., J. P. Thayer, R. R. Neely III, and G. Lucas (2013), Towards a comprehensive global electric circuit model: conductivity and its variability in WACCM model simulations, J. Geophys. Res., under review.

Bering, E. A., III, A. A. Few, and J. R. Benbrook (1998), The global electric circuit, Phys. Today, 51, 24–30.

Evans, S. (1965), Dielectric properties of ice and snow — a review, J. Glaciol., 5(42), 773-792.

McDonough, S., J. Qin, V. Pasko, (2011), Accurate modeling of ion conductivity in the earth's atmosphere, NSF EE REU Penn State Annual Research Journal, Vol. IX, 13-23.
Rapp, R. H. (1991), Geometric Geodesy Part 1, Department of Geodetic Science and Surveying, The Ohio State University.
Rycroft, M. J., G. R. Harrison, and K. A. Nicoll (2008), An overview of earth's global electric circuit and atmospheric conductivity, Space Sci. Rev., 137(1-4), 83-105.
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, 321–339, AGU, Washington, D. C.
Tinsley, B. A., and L. Zhou (2006) Initial results of a global circuit model with stratospheric and tropospheric aerosols, J. Geophys. Res., 111, D16205.
Zhou, L., and B. A. Tinsley (2010), Global circuit model with clouds, J. Atmos. Sci., 67, 1143-1156.

ABSOLUTE POWER CALIBRATION OF THE PENN STATE RADAR SYSTEM FOR METEOR STUDIES

Zachary Morgan,* Julio V. Urbina,[#] and Freddy Galindo[+]

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

*Undergraduate Student of
Department of Electrical Engineering
The University of Mississippi
University, MS 38677

ABSTRACT

Recently, a technique was developed to derive meteoroid masses from meteor showers by using well-calibrated meteor radar facilities [1]. This technique firstly computes the maximum electron line density from the backscatter signal and then derives the meteoroid mass by theoretical considerations, e.g. chemical properties of the meteor population. The ability to determine meteoroid mass will expand the current understanding of astrophysical properties as well as contributing factors to atmospheric phenomena. We report the absolute power calibration of the Penn State Meteor radar system. This is achieved by hardware and software implementation of several methods of calibration. These techniques include noise generators and transmitted signals acting as ideal echoes, both of which are fed directly to the receiver system. Furthermore, we used a numerical model of cosmic noise sources, which allows the effects of the antenna (e.g. the target in relation to the galactic core) to be considered for calibration at the moment of operation.

INTRODUCTION

Meteoroid studies using radar systems have the capacity to investigate the mesosphere and lower thermosphere. Further, these studies allow us to know the characteristics, e.g., speed and radiant, of the meteor populations that are constantly hitting the upper atmosphere. These radars typically operate in the VHF band and are sensitive to cosmic radio sources. This introduces a range of noise sources (i.e., radio stars and the galactic core) that may be impervious to

[#] Faculty Mentor
[+] Graduate Mentor

common calibration methods [2]. The implementation of power calibration techniques allow the study of electron line densities in the specular point to be more precisely measured. When coupling the added accuracy of these values with new studies that rely on electron line densities (e.g., meteoroid mass determination from ablation models) a higher understanding of a plethora of atmospheric and meteoroid properties is attained.

The absolute calibration of the Penn State Meteor radar system is achieved by the hardware and software implementation of reliable calibration techniques. The calibration methods for the receiving system are noise generators, coherent transmitted signals, which are both fed directly to the system to reveal receiver characteristics, and the application of a Global Sky Model of cosmic noise presented by de Oliveira-Costa [3] to characterize the receiving system. The sky noise calibration process uses a numerical model of cosmic noise observations interpolated to any desired operating frequency and additionally allows the effects of the antenna beam pattern and radar site (e.g., the boresight in relation to the galactic core) to be considered for calibration at the moment of operation.

Calibration with External Noise Generators

Receiving systems often record signal power in digital units based on the precision of the analog-to-digital converter employed. An estimation of the quantity in physical units is made by multiplying the digital value by a calibration factor. Such a factor may be determined by feeding a calibrated noise source directly into the receiving system. Power data collected from the receiving system processed along with the input noise power produces a calibration factor for incoherent detected signals as noise.

Coherent Signal Calibration

A calibration factor may also be produced using coherent signals (e.g., a pulse train) and used to estimate physical units from digital values. This method is quite similar in physical setup in that it feeds the signal right into the receiving system for a direct comparison of what is expected and what is received. Employing a level of pulse repetition for coherent integrations improves the accuracy of the calibration factor.

Sky Noise Temperature Calibration

A final calibration technique involves the use of a cosmic noise data presented by the Global Sky Model to determine what a particular receiving system should read at a given location and time. The data reveals the cosmic noise temperature observed in galactic coordinates and stored in the HEALPix projection for a sphere [4]. From this set of information, new data is interpolated to match the characteristics of the receiving system (e.g., receiving frequency). Any given receiving system may implement sky noise calibration by applying the directional pattern of the receiving antenna.

EXPERIMENT DESCRIPTION

Calibration Factor Determination in Directly Fed Methods

Determining a calibration factor and receiver noise temperature enables the conversion from arbitrary digitizer units to physical units and will reveal characteristics of the receiving system. The calibration factor using a calibrated noise source is a relationship between the input power in known watts and the output power in arbitrary units. The calibrated noise power source that is fed directly into the receiving system can be described by the thermal noise power equation

$$\mathrm{P}_{\mathrm{n.inp}} = k_B B_N T_{n.inp} \tag{1}$$

where $T_{n.inp}$ is the calibrated noise temperature input in kelvin, k_B is the Boltzmann constant approximated at $1.38\times10^{-23}\,JK^{-1}$, and B_N is the receiver noise bandwidth around 1 Mhz [5]. The output power that is received is defined in a similar way

$$\mathrm{P}_{\mathrm{n.inp}} = g_r k_B B_N \left(T_r + T_{n.inp}\right) \tag{2}$$

where g_r is the available gain and T_r is the receiver noise temperature in Kelvin. Output noise power is measured at discrete points chosen at the discretion of the individual calibrating the system and the limits of the noise generator. A noise sample averaged over the duration of the recording process transforms the signal to a single output noise power for each input noise temperature. $\mathrm{P}_{\mathrm{n.out}}$ condenses to a single mean point way

$$P_{n.out} = \frac{1}{n}\sum_{t=0}^{n} D_t \tag{3}$$

where D is the set of data samples, n is the number of samples, and t is the sample index. Latteck et al. simplifies the output power by assigning comprehensive parameters

$$\begin{aligned} a &= g_r k_B B_N T_r \\ b &= g_r k_B B_N \\ P_{n.out} &= a + bT_{n.inp} \end{aligned} \tag{4}$$

where a is an offset mostly described by the receiver noise temperature, available gain, Boltzmann constant, and noise bandwidth and b is a multiplying factor.

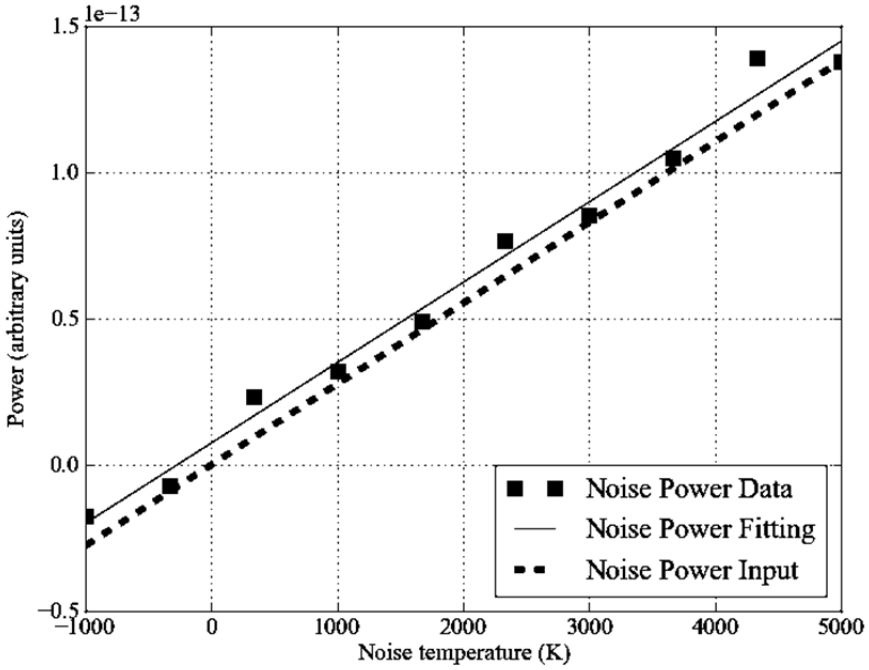

Figure 1: Discrete simulated noise powers with the Levenberg-Marquardt algorithm applied for curve fitting

Calculating *a* and *b* is a simple task if $P_{n.out}$ is assumed to be a linear, but any data collected likely has a level of irregularity resulting in a piecewise linear function (See Figure 1). A curve fitting function from the *SciPy* package of the Python Programming Language is used which employs the Levenberg-Marquardt algorithm for least squares curve fitting. The function returns several optimal values that most accurately predict the output power in terms of the input noise temperature, including the values of *a* and *b*. From the curve fitting the a parameter and the calibration factor, c_n, is found

$$c_n = \frac{P_{n.inp}}{P_{n.out} - a} \tag{5}$$

The *a* parameter divided by the *b* parameter reveals the receiver noise temperature, T_r, during operation. This is way

$$\frac{a}{b} = \frac{g_r k_B B_N T_r}{g_r k_B B_N} = T_r \tag{6}$$

The *b* parameter divided by the Boltzmann constant and receiver noise bandwidth reveals the available receiver gain, g_r, during operation.

$$\frac{b}{k_B B_N} = g_r \tag{7}$$

The calibration factor for a coherent signal is obtained in a similar way as with a calibrated noise source. As before, the signal is fed directly into the receiving system as data is collected. The signals are recorded as they are transmitted and archived by the receiving system. The received peak power of the pulse is the variable of interest for pulsed signals. In a given sample of pulse repetitions the corresponding output signal power is found by integrating each pulse into an average and finding the mean of the peak

$$P_{s.out} = \frac{1}{n \cdot m} \sum_{i=0}^{n} D_i \tag{8}$$

where n is the number of repeated pulse samples, m is the number of samples that define the width of the pulse, i is the index of the sample, and D is the integrated sets of data partitioned based on the inter-pulse period. Comparing each sample in time reveals a coherent signal calibration factor

$$c_s = \frac{P_{s.inp}}{P_{s.out}} \tag{9}$$

where $P_{s.inp}$ is a discrete set of pulse powers produced by the signal generator.

Sky Noise Calibration Techniques

The sky noise observed at a range of frequencies can be a tool for calibrating radar systems. The Global Sky Model presents a reliable map of sky noise temperatures (See Figure 2) stored in the galactic HEALPix representation and is readily available at https://www.cfa.harvard.edu/~adeolive/gsm/.

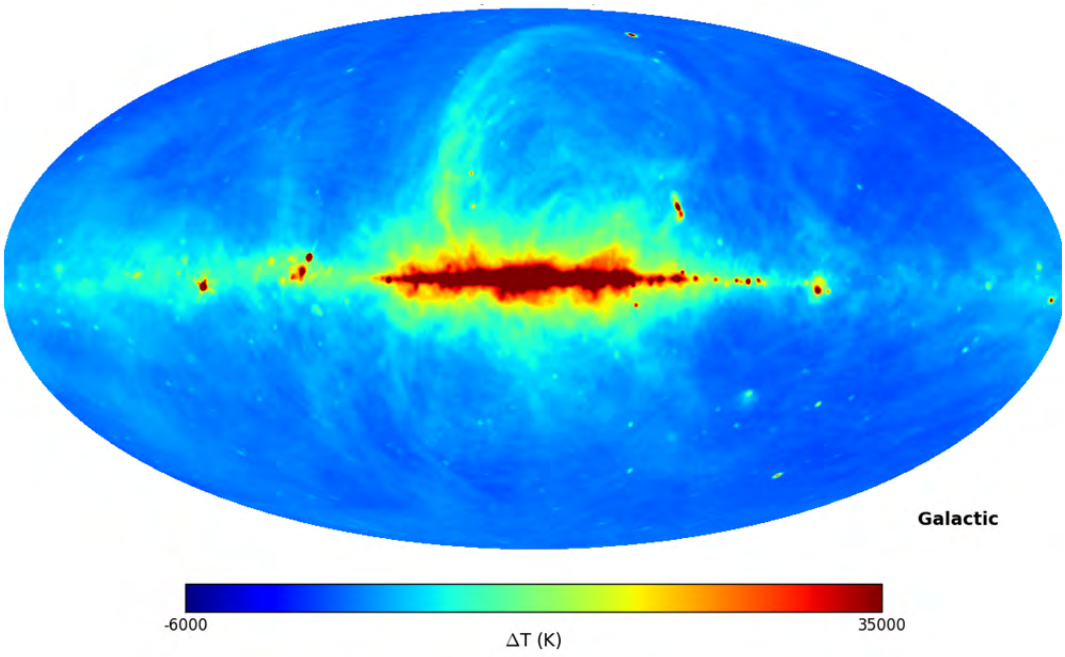

Figure 2: Mollweide projection of cosmic noise temperatures displayed in the galactic coordinate system

The information is applied by determining the noise temperature points of the model that contribute to the overall noise observed by the radar receiving system. Actual data is collected and compared with the expected noise temperatures from the model. For calibration purposes, the resulting sky noise temperature returned from the model may be removed from the overall noise power

$$P_{n.out} = g_r k_B B_N (T_r + T_{sky}) \tag{10}$$

where T_{sky} is the variable that adds to the receiving system's noise power (Stober et al., 2010). Assuming an infinitely narrow antenna beam makes calculating the receiving system's overall sky noise observation simple, but taking into consideration a realistic beam pattern enables a more accurate calibration effort. A coaxial collinear antenna beam pattern generated from a three-dimensional set of data is shown in Figure 3. In addition, Figure 4 shows a cut in the azimuth = 0 direction for the theoretical beam pattern shown in Figure 3.

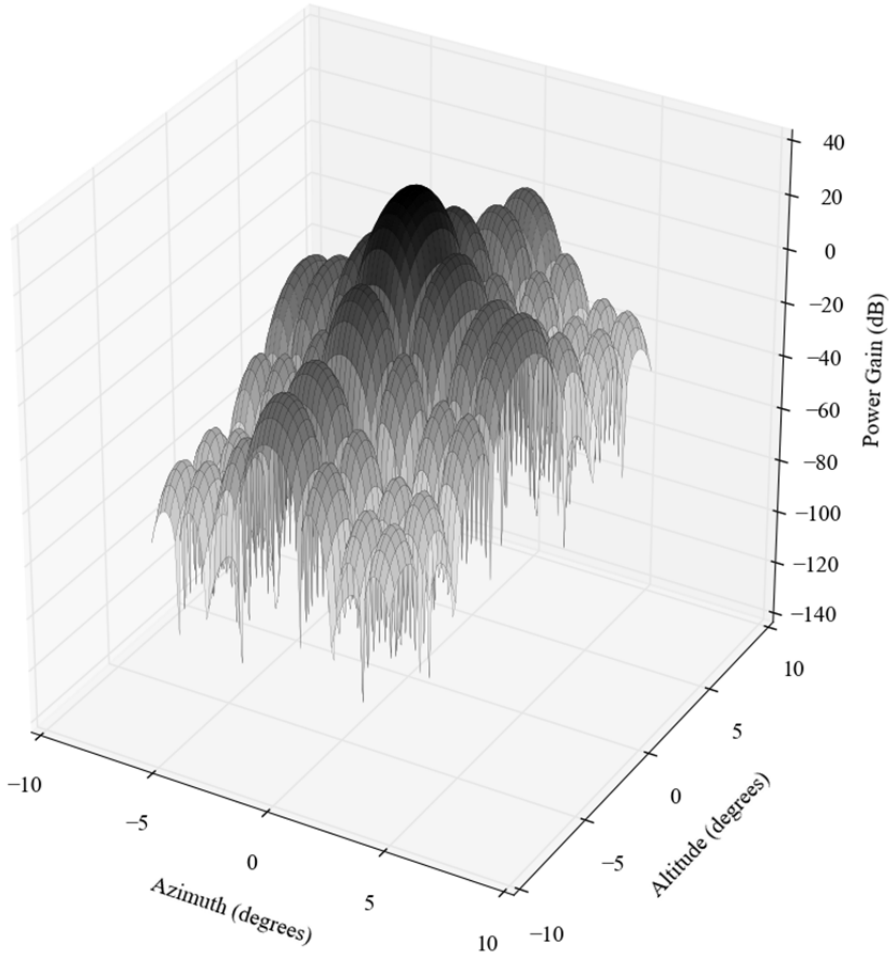

Figure 3: Rectangular radiation plot of a coaxial collinear antenna pattern

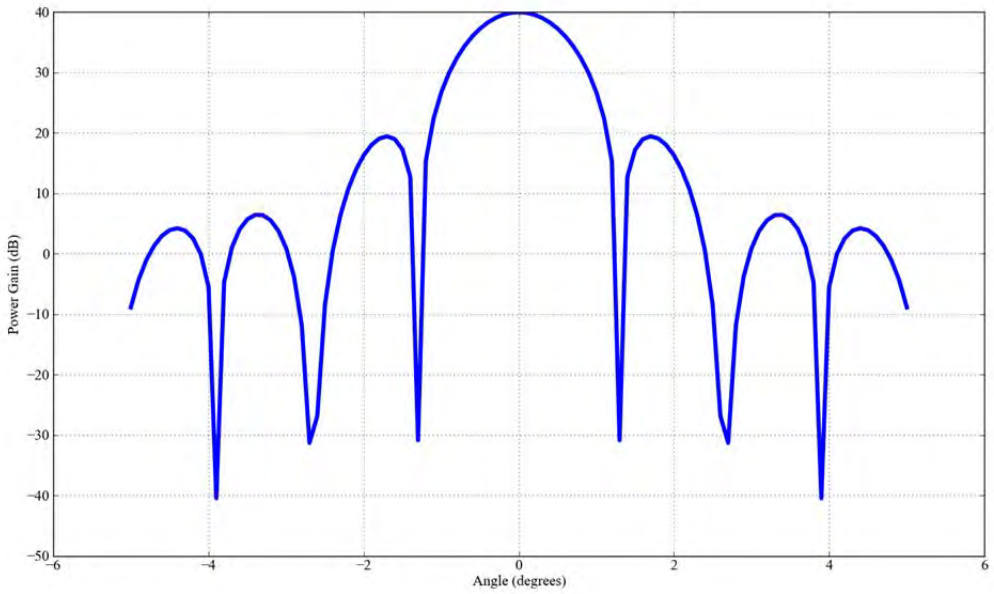

Figure 4: Rectangular radiation plot in a two-dimensional cut of the pattern presented in Figure 3

The radiation pattern is stored in a rectangular fashion where the horizontal axes represent the angles from the boresight and the vertical axis represents the power gain in decibels. The power gain of this particular model is stored as a 101×101 matrix where element [50,50] is the point of maximum gain. The same method for finding a single point of noise from local coordinates, epoch, and direction is applied using the directions indicated by the horizontal axes of the antenna pattern model which is defined by the element indices of the matrix as they are related to the azimuth and altitude difference from the boresight. The observed noise temperature from the model is also stored as a 101×101 matrix making the application of the gain from the antenna model easily accomplished with element by element multiplication of the matrices.

The observed sky noise temperatures with no gain contribution taken into consideration is projected to local coordinates for a single point in time as seen in Figure 5. The noise temperature may vary drastically as the observation window changes.

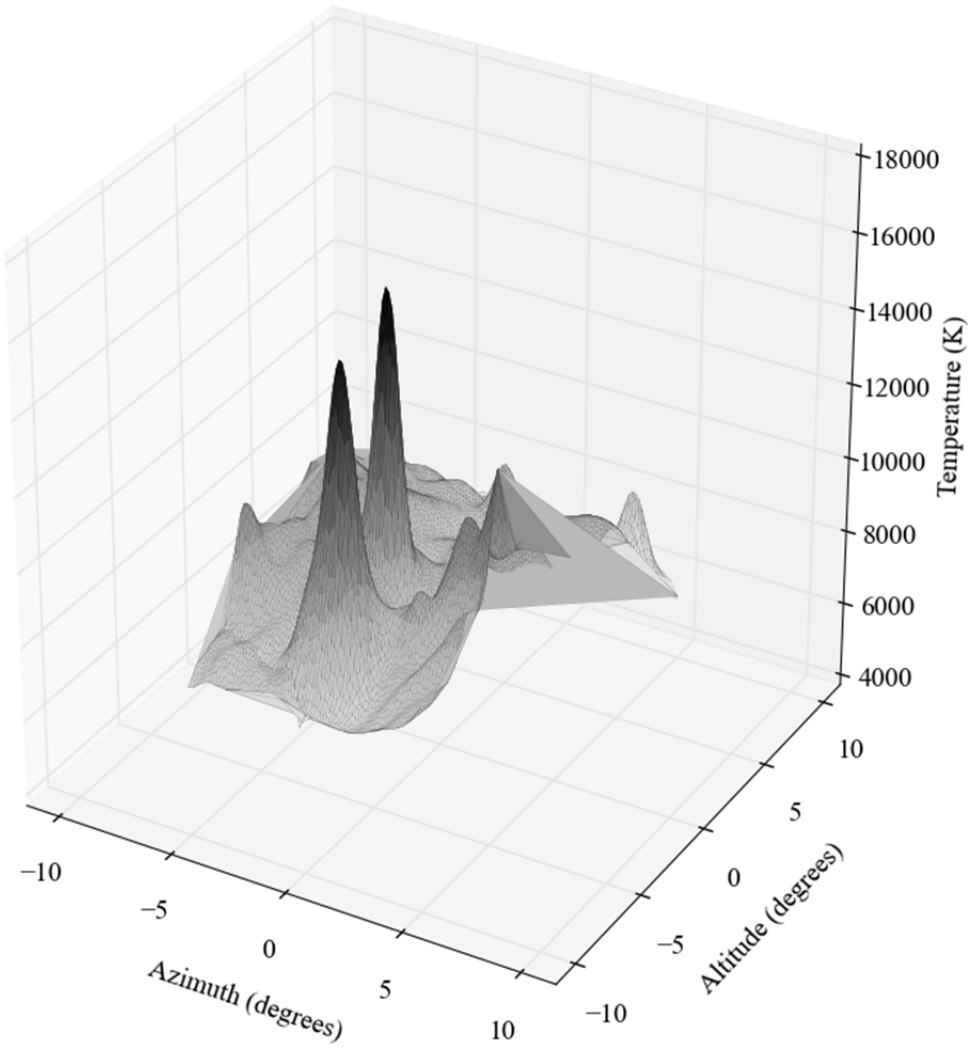

Figure 5: Observed sky noise temperatures for arbitrary beamwidths

The projection offers a set of noise temperatures as they might apply to any antenna beam pattern of more width than a pencil-beam. As the contour plot of Figure 6 shows, calibrating with a single point is not preferred because the noise temperature may fluctuate 10×10^3 K in just a few directional degrees.

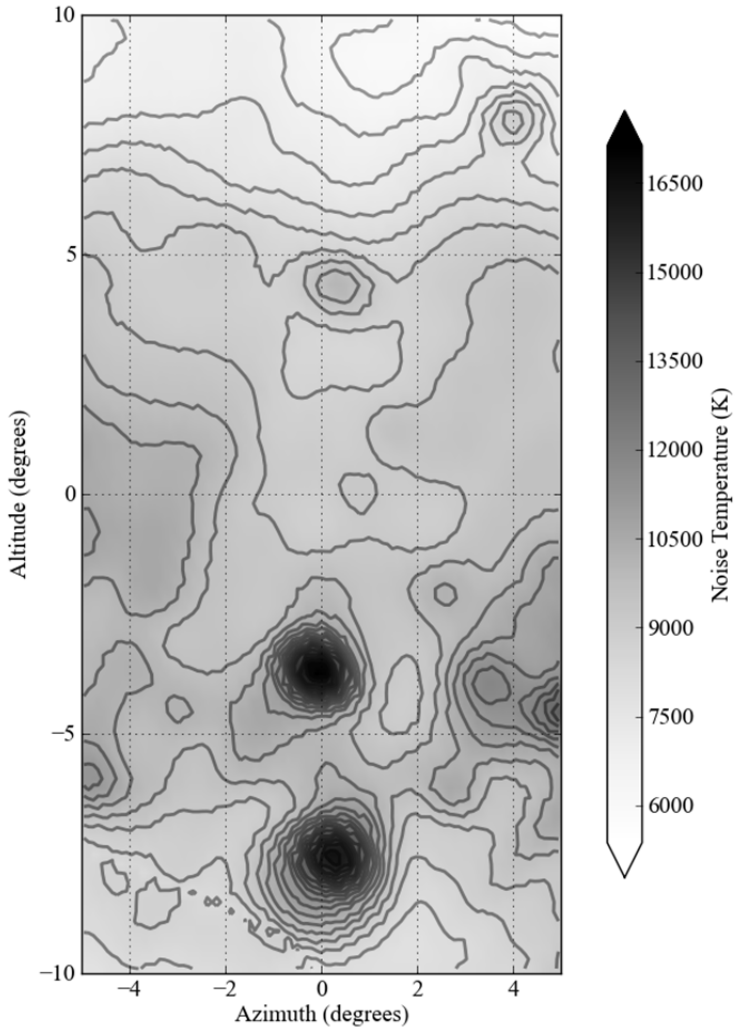

Figure 6: Contour of the observed noise temperatures of Figure 5

The elements of the contributing noise power matrix is defined by

$$P_{ij} = G_{ij} T_{ij} \tag{11}$$

where, i is the row number, j is the column number, P_{ij} is each element of the observed noise matrix, and G_{ij} and T_{ij} are the elements of the antenna gain matrix and corresponding noise temperature matrix, respectively. Once the noise temperatures have specific weights applied by the beam pattern their sum produces the overall noise power observed. This theoretical noise power may be subtracted from the observed noise power of the receiving system to produce the noise power and effectively the noise temperature of the receiver.

$$\frac{P_{n.out}}{g_r k_B B_N} - T_{sky} = T_r \tag{12}$$

Penn State Meteor Radar System Implementation

The three calibration methods previously explained are directly applied to the Penn State Meteor Radar System to characterize system parameters. This system typically operates at 49.8MHz so the calibration techniques implemented are to be mixed with a carrier of that same frequency and demodulated back to the original states. With this process, all characteristics of the receiving system are considered in calibration. The sky noise model is applied by estimating the parameters of the radar site and comparing with actual data obtained from the meteor radar archiving power information as the antenna and all other components are connected and operational.

EXPERIMENT RESULTS

Collecting data from the equipment of the radar site allows the receiving system to be characterized. Both directly fed methods return calibration factors, receiver noise temperatures, and available gain. Applying the Global Sky Model returns an overall system noise temperature.

Pulse trains of varying power were sampled at 1 MHz for two minutes per level. The signal from the function generator was first mixed to 49.8 MHz then directly fed to the receiving system for down conversion and archiving. The data was then processed in the manner described earlier for a curve fitting to be made. Figure 7 shows the calibration cure for the transmitted signal method of calibration.

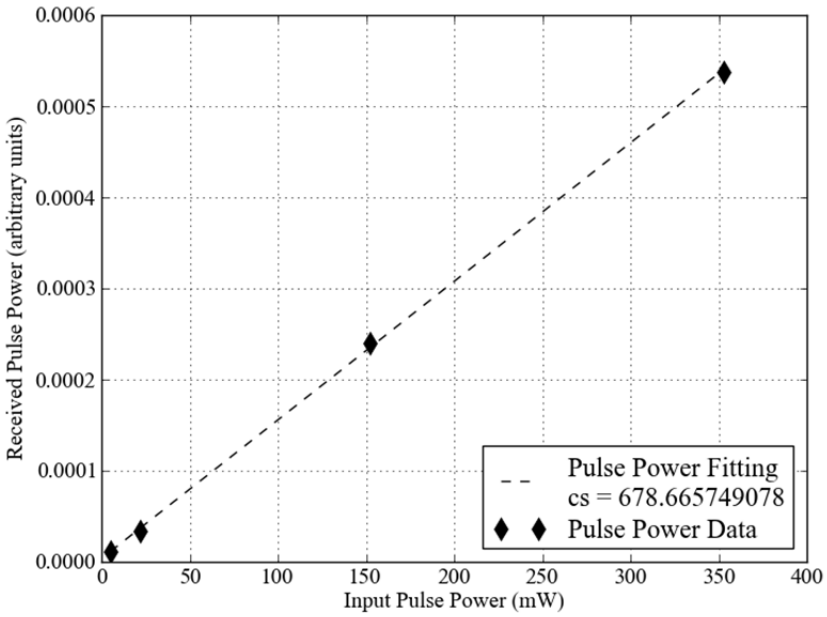

Figure 7: Transmitted signal calibration factor determination by curve fitting

The steps involved in extracting noise temperatures from the Global Sky Model can prove to be a challenge to implement. As a baseline for accurate

results, real 50 MHz Jicamarca Radar Observatory (JRO) sky noise data was replicated with the Global Sky Model

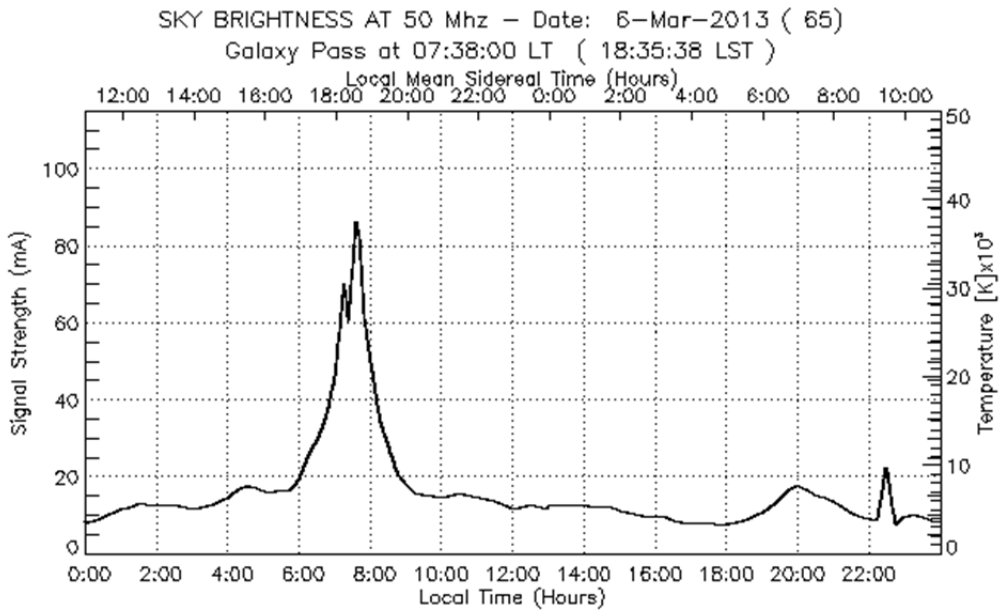

Figure 8: Jicamarca Radio Observatory Sky Noise Data

The JRO radar antenna consists of a large square array which yields a narrow beam width of about 0.8º to 1.1º for operation. Considering this beam width to be nearly a single point in direction yields a sky noise model estimation similar to that of real data as seen in Figure 9.

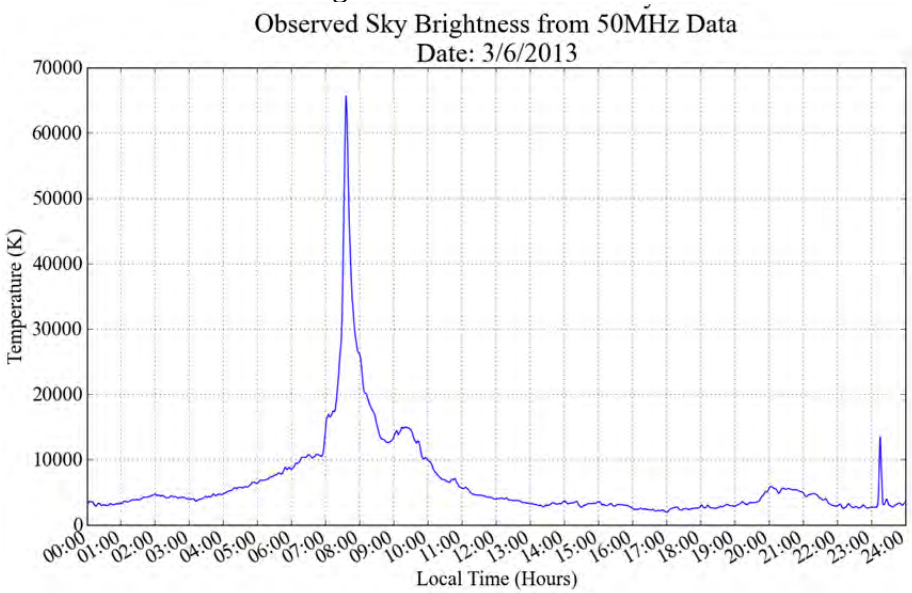

Figure 9: Sky noise data from Jicamarca Radio Observatory parameters generated from the Global Sky Model

The Penn State Meteor Radar System is estimated to have a main lobe of about 20º altitude wise and much flatter in azimuth. The antenna has a gain of 20 dBi at the boresight. Data was collected at the radar site which is directed toward magnetic north with an altitude of about 18º from the ground plane.

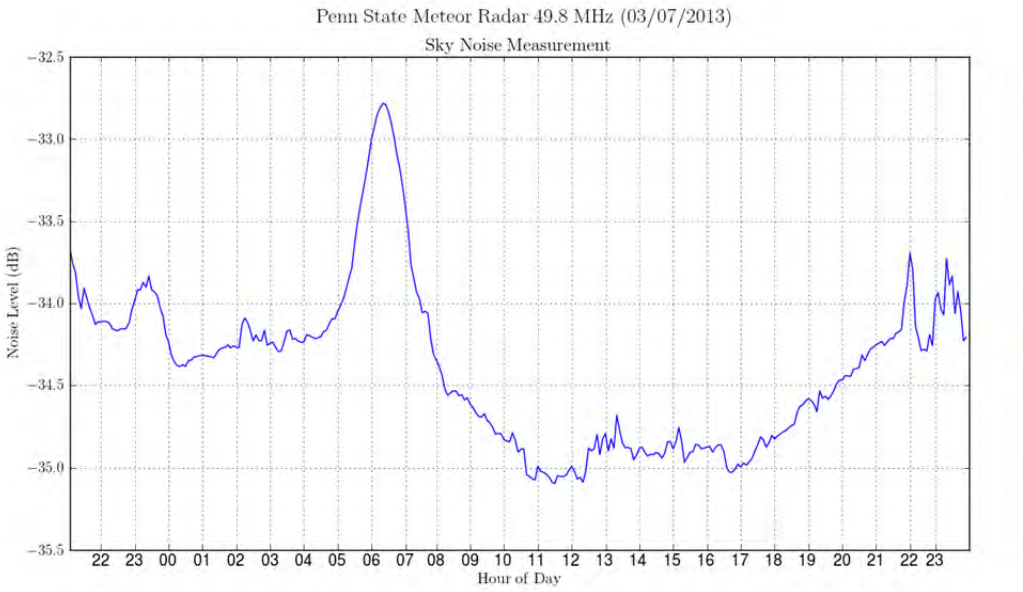

Figure 10: Penn State Meteor Radar System Sky Noise Data

The same parameters are applied to the Global Sky Model to generate a system noise temperature.

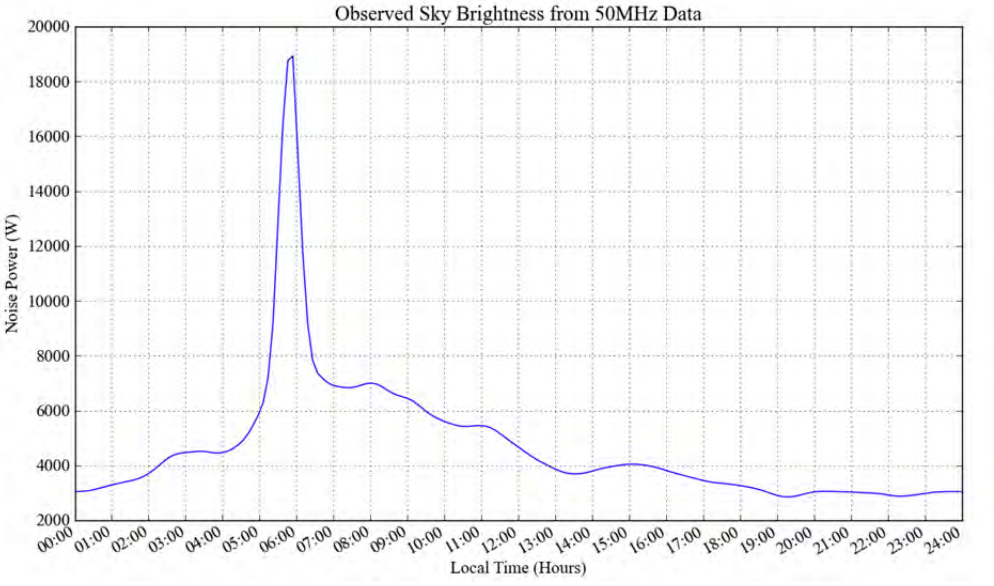

Figure 11: Generated sky noise data for the Penn State Meteor Radar System

Applying Equation (12) at any common time point of Figures 10 and 11 yields a receiver noise temperature.

DISCUSSION

The process of absolute power calibration may seem trivial, but it is close attention to detail that produces favorable results. Take for instance the pulse train of the directly fed transmitted signal calibration. When integrating pulses into an average, there is an easily overlooked cause of error. If integrating is achieved by dividing the sample into segments defined by the pulse repetition frequency, the pulse generating portion of the system must be without fault. In just a few samples, the function generator used in this procedure would allow pulses to drift by slight increments until they were completely out of sync.

Sky noise calibration must also be carefully implemented to have a positive impact. At any given moment sporadic conditions of the atmosphere may affect the noise observed in the receiving system. For this reason, it is recommended that several days of observations are integrated to characterize radar systems.

CONCLUSION

A concise calibration effort is necessary to aid in the increasing sophistication of meteor studies. The outlined procedure demonstrates the absolute calibration of meteor radar systems by three methods: calibrated noise sources, transmitted signals, and the Global Sky Model. Through the directly fed methods of noise sources and transmitted signals, characteristics of receiving systems are obtained. The calibration factors derived from these methods provide a relation of measured power in arbitrary units to physical units such as watts. The software implementation of the Global Sky Model also aids in the characterization of receiving systems by providing a real-time reference for the sky noise temperature component of a power sample. Techniques such as these improve the quality of measurements and sophistication of systems to produce accurate and conclusive results.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would first like to thank Freddy Galindo and Dr. Urbina for all their help on this project as well as the members of the Applied Signal Processing and Instrumentation Research Laboratory. I would also like to thank the friends I have gained during such a rewarding experience and those who had a role in the organization of the REU program.

REFERENCES

[1] Stober, G., W. Singer, and Ch Jacobi. "Cosmic radio noise observations using a mid-latitude meteor radar." Journal of Atmospheric and Solar-Terrestrial Physics 73.9 (2011): 1069-1076.

[2] Stober, G., Ch Jacobi, and W. Singer. "Meteoroid mass determination from underdense trails." Journal of Atmospheric and Solar-Terrestrial Physics 73.9 (2011): 895-900.

[3] Oliveira-Costa, De, et al. "A model of diffuse Galactic radio emission from 10 MHz to 100 GHz." Monthly Notices of the Royal Astronomical Society 388.1 (2008): 247-260.
[4] Gorski, Krzysztof M., et al. "HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere." The Astrophysical Journal 622.2 (2005): 759.
[5] Latteck, R., et al. "Similarities and differences in polar mesosphere summer echoes observed in the Arctic and Antarctica." Annales geophysicae: atmospheres, hydrospheres and space sciences. Vol. 26. No. 9. 2008.3] Huang, W. "On the Selection of Shape Memory Alloys for Actuators". Materials & Design. Vol. 23, Issue 1. 4 December 2001: 11-19

DEVELOPMENT OF A SOFTWARE-DEFINED RADAR RECEIVER FOR IONOSPHERIC STUDIES

Benjamin Young,* Julio Urbina,[#] and Alexander Hackett[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Virginia Commonwealth University
Richmond, Virginia 23220

ABSTRACT

The PSU All-sky Radar Interferometry System (PARIS) at Penn State is an open source project aimed at providing cost efficient and flexible means of studying the atmosphere. ALLSKY is currently undergoing hardware and software upgrades which will enable it to operate using five complex-channels. In this paper, we describe the implementation of this feature to the radar. A full discussion of the system will be given with concentration on the receiver's capabilities and software design. Algorithms and design strategy are discussed as necessary to properly document optimizations and features required by the system.

INTRODUCTION

Concepts and Motivation

PARIS is an ongoing project of the Applied Signal Processing and Instrumentation Lab (ASPIRL) at Penn State beginning in 2007. The base operation of the radar is to collect data parameters from meteors between 90 and 130 km in the ionosphere, such as position and velocity, and subsequently infer neutral wind velocities.

Neutral winds in this region of the atmosphere are difficult to measure. Studies conducted to model their behavior have discovered that they are easily affected by external sources such as radiation from the sun or masses of rising air around areas of formidable and swift altitude differences, which can cause propagating movement called gravity waves [1]. The propagation of neutral winds

[#] Faculty Mentor
[+] Graduate Mentor

causes pressure differences in the atmosphere, which may lead to changes in global weather patterns in addition to acting as a pulling force capable of degrading the orbital path of objects, such as satellites, from low earth orbit (~400 km). Many systems will benefit from studying these neutral winds and how they affect everyday life.

The reason neutral winds are not well understood is due to the lack of ways to probe this specific part of the atmosphere. A weather balloon is typically limited to ~3–4 km and experimental rockets are expensive and very short lived. A satellite cannot study a set point in the sky from orbit, which may leave gaps in the data. Thus the standard means of studying the 90–130 km region of the atmosphere is by use of radar. A technique called interferometry allows low powered radars such as ALLSKY to study neutral winds indirectly by measuring how these winds interact with the ion trail of a meteor. The main parameters of the meteor necessary to study neutral winds are position and velocity. ALLSKY was created as an open source platform to offer research groups the tools necessary to study the atmosphere at an effective price.

Description of ALLSKY

Currently ALLSKY is software defined and consists of a single receiver, or point of reference, which restricts its ability to determine parameters other than size and altitude. To obtain velocity and position relative to the center of the earth, more receivers are necessary for referencing the incoming signal. Since ALLSKY is software defined, most of the necessary implementations needed to increase the number of receivers will be done in a software framework called GnuRadar, created by Ryan Seal [2].

ALLSKY uses various software based signal processing techniques to enhance its operational abilities such as resolution and accuracy, while not leaking into other frequencies. One such technique is called Phase-Coded Pulse Compression, which works by breaking apart the main transmission pulse into a series of shorter pulses each characterized by its own predetermined and optimized phase offset ranging from 0 to 180 degrees [3]. When a long pulse is broken down into consecutive short pulses, the transmitted signal has the range of the long pulse and the resolution of the short pulse [3]. These results have been shown by other radars such as the High-Power Large-Aperture Jicamarca 50-MHz radar (11.95° S, 76.87° W) [4].

Resolution and altitude accuracy alone cannot track a meteor nor can a meteor be tracked with a single antenna since there are no other points of reference from which to compare the signal. To account for this dilemma interferometry is used to track meteors by measuring the phase difference between two or more antennas viewing the same object [5]. An object that interferes with the radar's transmission will reflect a signal that will reach each antenna in the receiver array at a different time and with a different phase corresponding to the angle at which the object is in reference to the center of the radar's field of view. By taking many data points along the meteors trajectory and differentiating the phase difference

with respect to time, it is possible to infer velocity and position. When enough samples are collected the data is sent to a software interface to produce neutral winds.

EXPERIMENT DESCRIPTION

Overview

Since ALLSKY is operating on one receiving antenna, it doesn't have the ability to determine position or velocity. Thus it is necessary to upgrade the radar with four more antennas, which creates a receiver array with the ability to gather indirect data about neutral winds. To safeguard against avoidable errors as well as make the project more manageable, the upgrade design process has been broken down into a finite number of steps. The critical steps required for the radar to properly operate in its new state include selecting a receiver pattern, determining necessary hardware to interface the new receivers with the host computer, and create modulable algorithms, which fit into GnuRadar and appropriately handle the new receivers and their data streams. To test the validity of the upgrade a series of *in-lab* tests were conducted to model the behavior of the hardware interface device.

Receiver Design Parameters

The receiver pattern must be geometrically optimized to detect necessary parameters such as position and velocity, and was determined to take the shape of a diamond, with an antenna at each corner being either 2 or 2.5 wavelengths from a center located antenna [6]. This setup allows the radar to accurately determine the velocity in both the *x*-axis and the *y*-axis; each corner antenna will measure a phase difference with both antennas directly adjacent to itself in addition to the antenna in the center, depicted in Figure 1. This design was spurred not only by the successful geometry of the array design, but also by the success of other radar systems utilizing the same setup such as the SKiYMET from Genesis Software [7] and JASMET at Jicamarca Radio Observatory [8].

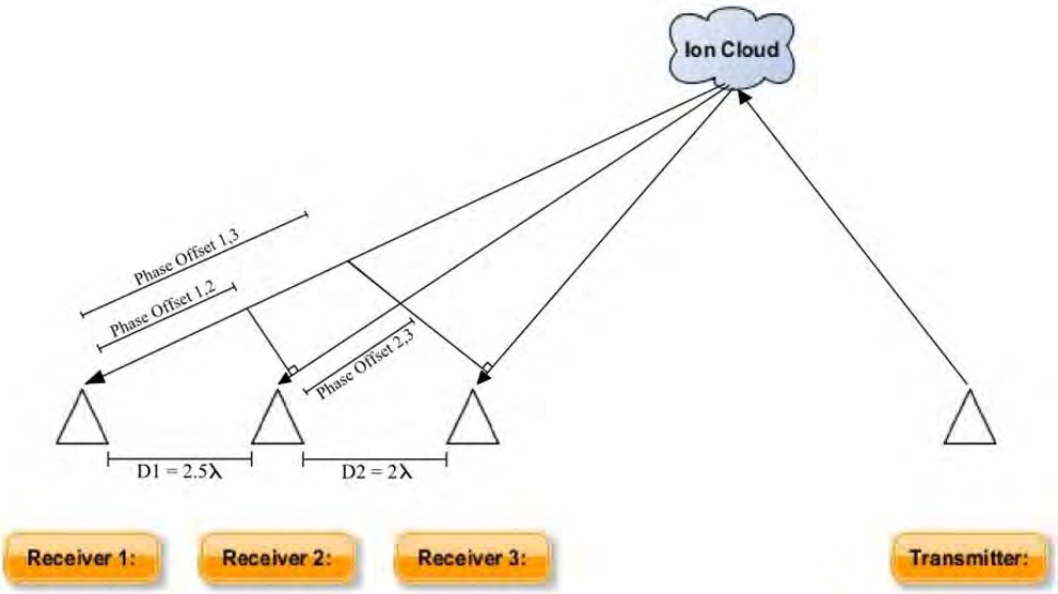

Figure 1: Depiction of a reflected signal being received by all three receivers at different times and with different phases. By careful placement of the antennas and through specific signal processing techniques, it is possible to infer the position of the ion cloud with respect to the center of the field of view of the radar (vertical axis for ALLSKY).

Hardware Interface Parameters

The current means of interfacing the antennas to the host computer utilizes a device called the Universal Software Radio Peripheral 1 (USRP), which digitizes the received analog signal and splits each channel into In-Phase and Quadrature components I and Q respectively, and then passes them to the host computer, shown in Figure 2. As its name implies, the USRP processes signals in software, which allows it to be flexible in the type of signal it's receiving as well as to cut down on bulky and expensive hardware. A received signal is passed through an Analog to Digital Converter (ADC) to generate discrete data-points, then sent through a Digital Down Converter (DDC) to demodulate the signal's frequency from that of the carrier to the local oscillator, and finally through software filters that break it into I and Q components. The USRP is also capable of transmitting by the reverse means of sending the signal through a Digital Up Converter, through a Digital to Analog Converter, and then to an amplifier. In this way the USRP acts as the medium through which signals pass from a computer to a radar.

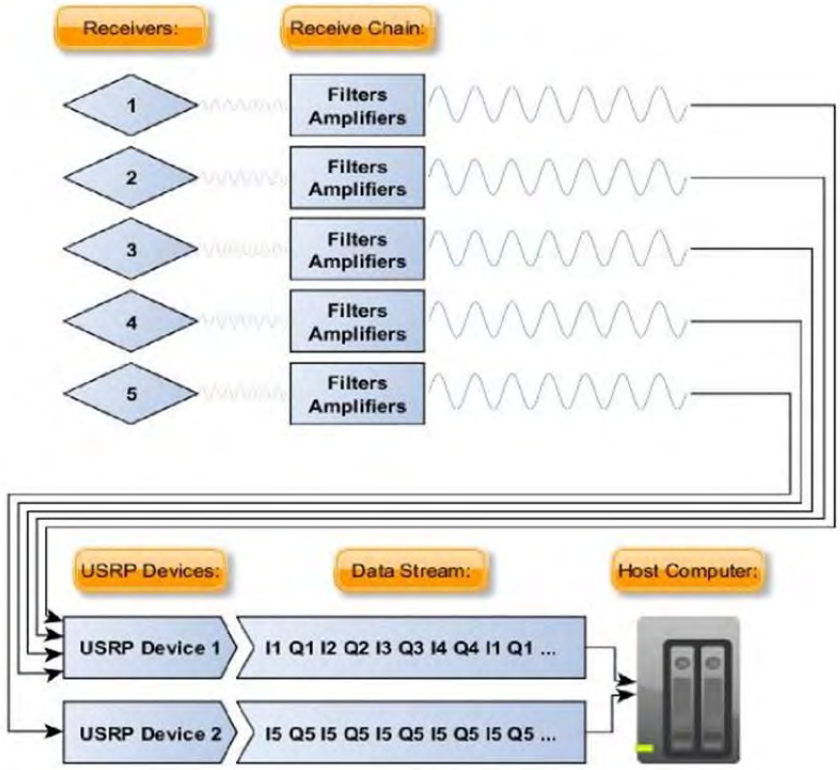

Figure 2: General flow of the data as it passes from the receivers to the host computer. Each receiver transmits the received signal through the receive chain and into the USRP devices, where the signal is processed into discrete data points consisting of In-Phase and Quadrature components and are sent to the host computer.

Each USRP has four channels, where each channel is tied to exactly one antenna. With more than four receiving antenna in ALLSKY's future, one USRP device will not be enough. It is possible to find other hardware with more channels, but the USRP is a cheap and reliable system, making it ideal for the project. The decision is then to use more than one USRP device, and to make the corresponding changes in GnuRadar. Careful measures must be taken when streams of data are pulled from more than one device since they are easily corrupted if not tracked properly. This will be accomplished by adopting a common master-slave device approach. To create a slave device, two of the fifteen Input/Output (I/O) pins on an adapter board called the BasicRX daughterboard are tied together, which the host computer uses to create a specific address for the device based on which pins are tied. The master device has no pins tied; its role is to keep all the slave devices synchronized by feeding them its clock pulse. Utilizing this method will enable two devices, or up to eight channels, to be synchronously connected to the host computer.

Software Design Parameters

As ALLSKY is for research purposes and meant to be expandable, the software will be kept as general as possible to allow for easier future upgrades. The main modification needed in the software is the ability to collect and merge data streams from more than one USRP device since GnuRadar is currently only capable of interfacing with one device.

Each USRP device will be objectified and then stored in a vector which allows for the use of expandable algorithms and clean code. However, using more than one device requires all the receivers to stream data in the exact same way and at the exact same time; otherwise it would not be possible to determine the phase difference between channels. To resolve this problem, each USRP device is set to take data with the same initial parameters such as the receive window and the sampling rate, while the master device feeds its clock pulse into the slaves, while being fed the same clock pulse as the host computer and transmitter. Since the master device is fed the same clock as the host computer, and the slave devices are fed clock pulses by the master, the whole system is synchronized through software making the data streams much more robust and tunable.

With the USRP devices streaming synchronized data, the host computer places sections of data one second in length from each device into a separate vector. This ensures that any slight temporal deviations in a device will be reset rather than accumulate error. The data is formatted as its stored in the vectors; the USRP takes samples from each connected channel and then places the samples sequentially based on its channel number into the vector (see Figure 3) before it takes more samples. When the vector holds one second of data, it is tagged to denote how much information it is storing, which is based on how many channels the device is streaming through and from which device the data originated. At this point, all the vectors are packaged together into one long, organized vector by means of an *interleaving function* shown in Figure 3. This eliminates the need for the data to be stored piece-wise in the computer and prevents the possibility of data getting mixed up, yielding false results that could go unnoticed.

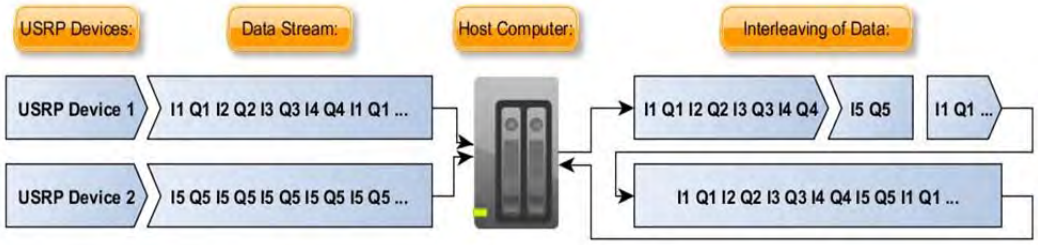

Figure 3: The process of *interleaving* the data. Both data streams enter a function that combines them by order of channel into a new vector, which is then stored in memory.

In-Lab Test Parameters

Once the code has been upgraded and the hardware has been tested, system tests were taken to confirm whether or not the new multi-USRP setup operates properly with its new receiver links. A system overview diagram of the upgraded radar is provided in Figure 4. To filter out as much probability for error as possible, controlled lab tests were conducted so that only necessary parameters were included. This effectively rules out the possibility of gathering false data due to issues related to other parts of the radar, and ensures that any errors with the new receiver array are not mistaken as errors with the radar. Conducting in-lab tests also allows for reliable and predictable results, thus saving time in implementation. These tests benefit from a controlled input signal and behave as expected, whereas a signal received by the radar might not behave predictably, thus increasing the probability of error in the new receiver array.

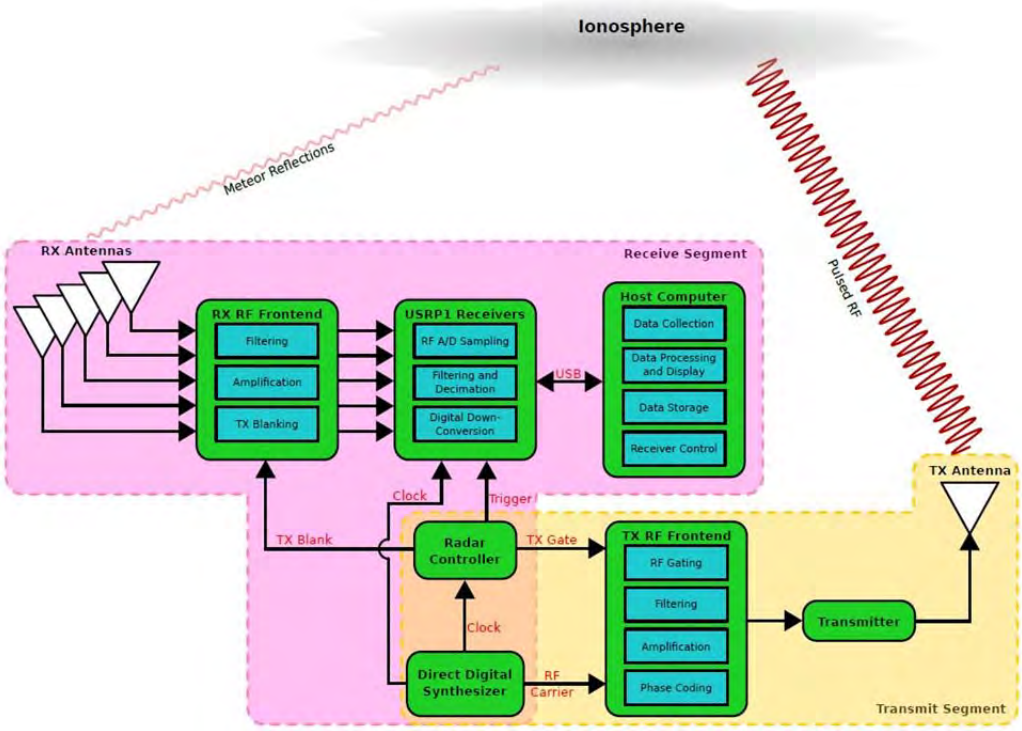

Figure 4: System diagram depicting a system design overview of the upgraded radar [9]

EXPERIMENT RESULTS

The main objective of the experiment was to upgrade the radar from one channel to five channels. To achieve this objective it was necessary to incorporate more than one USRP device, which required GnuRadar to be modified to accept the additional devices as well as handle their data streams. While the experiment to GnuRadar was successful in the fact that the software can now handle more than five channels, unforeseen errors have kept the system as a whole from operating correctly. There were also many low-priority goals, which were meant

to increase the user's operability, that were not incorporated due to time being allocated to debugging these errors.

The problem lies with the data being corrupted by an overflow fault in the system. With data overflow, the computer is too busy handling other tasks to keep up with the data that it is being sent. Each USRP samples data at 64 MB/s per channel, which is equivalent to each channel sending the host computer a new value every 1.56 μs. A typical computer could handle these sampling speeds if it was dedicated to sampling the line, but in GnuRadar the computer has the following tasks: 1) it has to read the data, 2) merge it, and 3) then send it to memory. If the code is not written efficiently enough, then the computer will be less likely to keep up with the sample rate.

To test the performance of the system, a function generator was used to simulate a signal that the receiver would send to the USRP. A simple 50.001-MHz sinusoidal signal was fed into the USRP device with a center frequency of 50 MHz. The signal was sent to each channel individually, while a custom plotting interface built into GnuRadar was used to view the results shown in Figure 5. Since the system was experiencing overflow, the data received from the sinusoid was corrupted, making the bulk of the signal unusable and the plots unreadable.

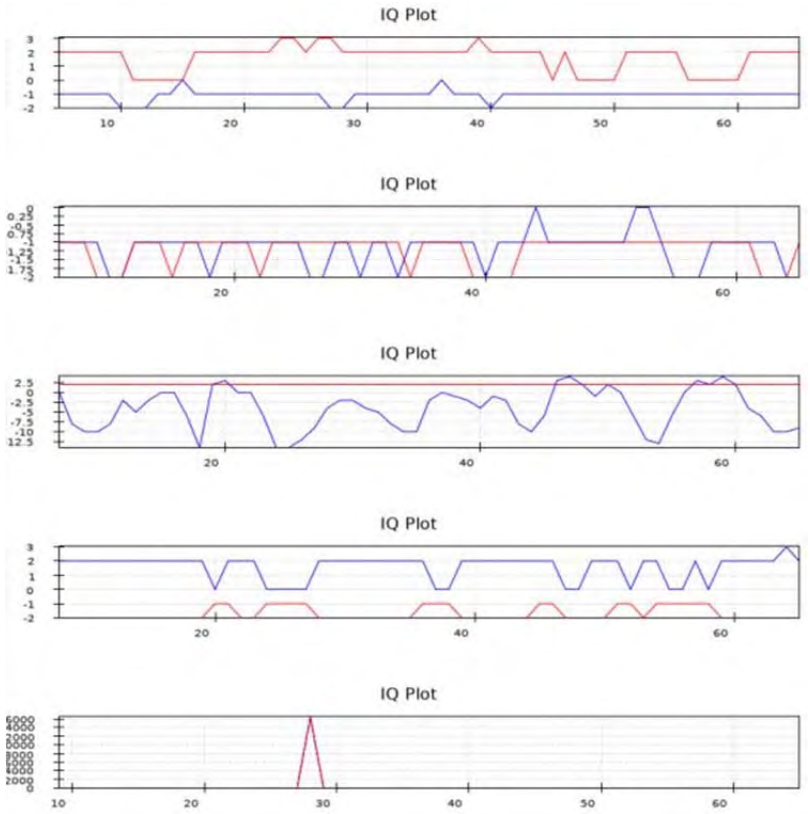

Figure 5: Results of plotting a received sinusoid. The displayed signal on each channel represents a sinusoidal wave, but since the data is corrupt the sinusoid is not visible.

Since the overall system did not operate correctly, it was decided to test individual parts of the system which had been upgraded. In this way it could be shown whether or not the overflow fault was the only error in the system. The *interleave function* was tested to make sure that it was properly merging the data streams. Figure 6 shows a few sample values returned from the *function* indicating that they are in fact properly merged.

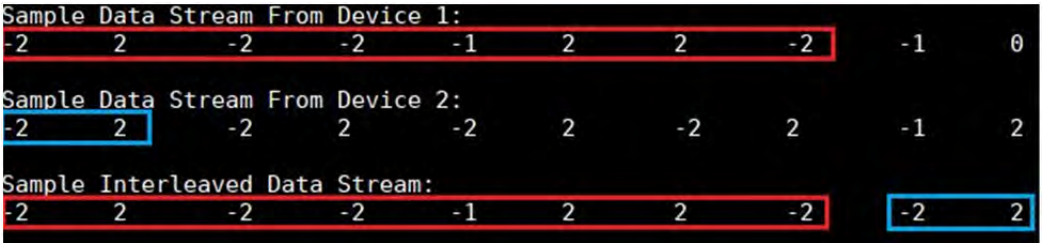

Figure 6: Sample data points demonstrating that the *interleave function* merges data properly.

DISCUSSION

The core of the overflow fault is believed to be the *interleave function* which was kept simple so as not to produce unnecessary errors. The result failed to meet the required time specifications necessary to maintain the sample rate due to this simplicity. Compensation parameters for the data overflow are not well known since several sample rates were tested to determine a relative sample set size by which the machine could properly operate, but every sample rate failed.

To determine whether or not the data overflow was occurring because of the *interleave function*, it was benchmarked by a boost CPU timer function, supplied by Boost C++ Libraries, shown in Figure 7 [10]. Also shown in Figure 7, the benchmark depicts that the *interleave function* takes around 0.8 seconds to complete when attempting to stream 64 MB/s per channel for five channels. This represents ~80% of the time that the computer has to calculate the entire second's worth of data, and the results are proof that the culprit behind the overflow fault was the *interleave function.*

```
//code to benchmark the time it takes to complete this func.
boost::timer::auto_cpu_timer t;
```

```
0.804581s wall, 0.740000s user + 0.050000s system = 0.790000s CPU (98.2%)
```

Figure 7: *Top:* The Boost Timer variable "t" starts a timer when it is created, and stops the timer when it is destroyed at the function's end. *Bottom:* Output generated by the Boost Timer variable "t" at its destruction.

The *interleave function* requires a large quantity of time due to its simplicity; the function utilizes loops to merge the data and replicates objects with a lot of overhead such as vectors shown in Figure 8. Thus the function is wasting time and memory since vectors require both to be created or reallocated. An overview of the function is as such: the data from each device is stored in vectors which are passed by the host computer to the interleave function, where their sizes are measured and a new vector is created. The algorithm used to merge the data is a state machine, which actively tracks a channel's vector index, and knows when to look for that individual channel's data. The state machine then places the data in order in the new vector, which must reallocate space when new data is stored before returning the merged data to the computer. Software loops aid the algorithm, but their benefit of simplicity does not outweigh their slow speed.

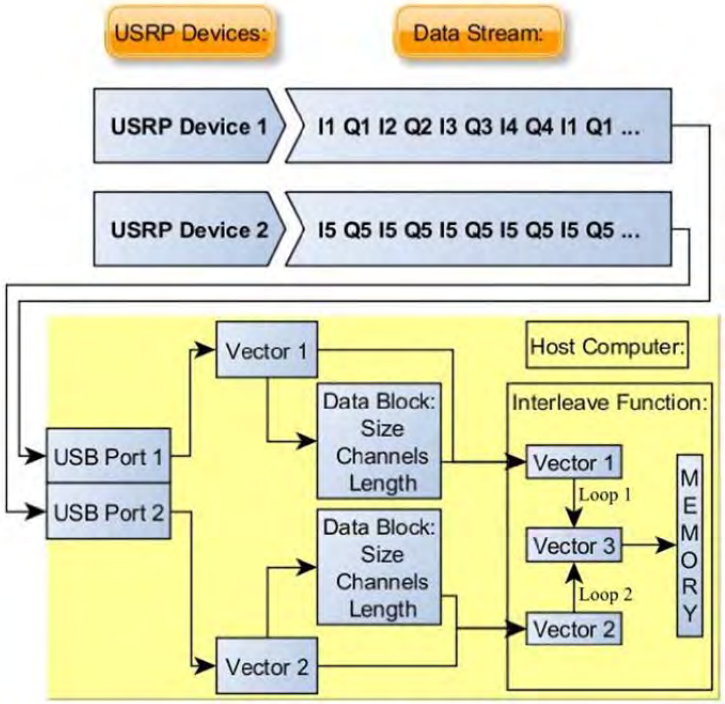

Figure 8: Radar data is sent to the computer which then sends it to the *interleave function* where memory is allocated and loops are run to store the data.

The solution is to optimize the *interleaving function*. One way to accomplish this is to merge the data as it is being received into the host computer as shown in Figure 9, rather than spending time processing it twice. This approach could potentially take a quarter or less time that the current *interleave function* requires by taking advantage of the properties of arrays; it takes the same amount of time for the computer to store a value in the 5^{th} element of the array as it does the 100^{th} or 1000^{th}. An algorithm may be written to take the data as it's being received by the computer, and knowing how many elements need to be offset due to the number of active channels, stores the value in the proper element of the array. A smarter storage technique like this will eliminate the need to merge two sets of data by placing them in the proper order to begin with.

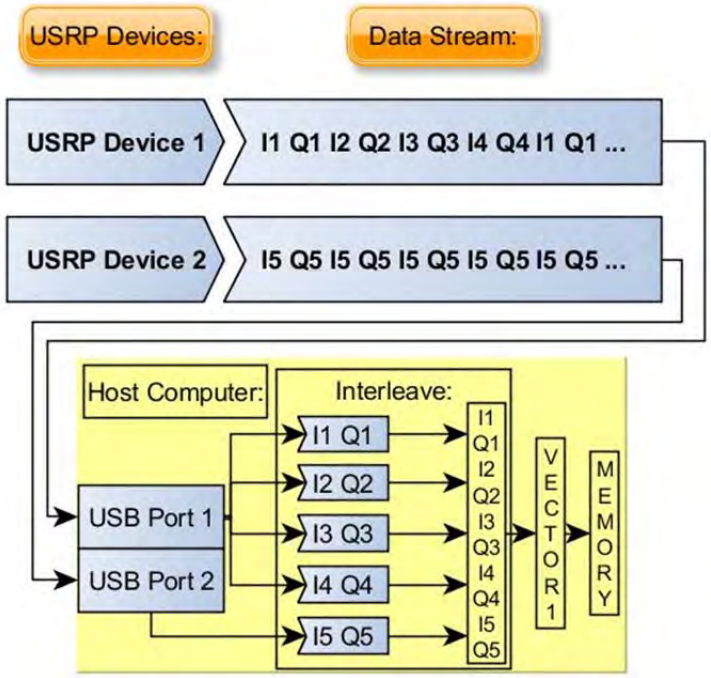

Figure 9: The suggested method to weave the data streams together would occur as the data is entering the host computer. This method is much more efficient than the current implementation.

Even with the source of the overflow known, general precautions should be taken in future upgrades to avoid such errors, as there are numerous sources to trigger them. The flooding of the Southbridge chip on the computer is a primary concern when using more than one USRP device. Tests were not conducted during this experiment to determine whether or not the Southbridge is capable of handling five channels; this test needs to be conducted in future experiments to ensure the radar will not encounter unnecessary data overflow.

When the Southbridge floods, it means that the physical chip cannot keep up with the amount of data being streamed, which may result in lost or misplaced data. To a sensitive operation such as PARIS, the corruption of data is a detrimental effect. The difference between data overflow from flooding of the Southbridge and time consuming code, is the way that the overflow is presented. With time consuming code, the data streams received will vary in length since the computer takes longer to start receiving. With flooding of the Southbridge, data will most likely be jumbled and corrupted since it is unable to keep up, and liable to receive only a portion of data.

With the majority of time spent attempting to solve the data overflow issue, parts of the project not essential to the operability of the upgraded system were cut out. One example of these modifications is the way that the code uses channels on each USRP devices. Ideally the host computer should be able to locate each connected device and probe the ports to determine which channels are

active. Instead, the current setup requires the number of active channels to be recorded in a configuration file by the user. There is also no flexibility in the order that channels are connected; the first four channels must be connected to the first device connected to the computer, and the remaining channels must be plugged into the proper ports on the slave device, else the computer will return an error. Modifications to remedy these nuances will certainly be necessary before this product can be commercialized or released as open source material, since the radar currently requires very specific conditions to operate. However, these modifications are not detrimental to the operation of the radar, but simply ease the user's operability of the radar.

CONCLUSION

Overall results of the experiment were successful, considering that the radar has now been upgraded to utilize five receivers. GnuRadar was successfully modified to accept multiple USRP devices, and has been kept general and easily modifiable. The USRP devices themselves have been shown capable of being synchronized together through a few hardware and software changes.

In order for the new upgrades to be operational, more work is required in the form of optimization on the host computer side by making the *interleave function* more efficient. Post optimization a significant amount of testing will be necessary to ensure the Southbridge chip can handle the amount of data needed for PARIS's operation.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would like to thank Dr. Julio Urbina and Alexander Hackett for all their support and guidance throughout this experiment.

REFERENCES

[1] Chunming Huang, Shaodong Zhang, Qihou Zhou, Fan Yi, and Kaiming Huang. "Atmospheric waves and their interactions in the thermospheric neutral wind as observed by the Arecibo incoherent scatter radar," Journal of Geophysical Research, Vol. 117, 2012 <http://onlinelibrary.wiley.com/doi/10.1029/2012JD018241/pdf>

[2] GitHub Website. 2 March 2008. Ryan Seal. 5 June 2013 <https://github.com/rseal/GnuRadar>

[3] Radartutorial Website, 1997. Christian Wolff. 7 July 2013. <http://www.radartutorial.eu/08.transmitters/tx17.en.html>

[4] J. L. Chau and R. F. Woodman. "Observations of meteor-head echoes using the Jicamarca 50 MHz radar in interferometer mode," Atmosphereic Chemistry and Physics. Atmos. Chem. Phys. 4, 511–521. 28 November 2003 <http://hal.archives-ouvertes.fr/docs/00/29/54/13/PDF/acp-4-511-2004.pdf>

[5] Chunmei Kang and Scott Palo. "Comparison of a Circular and Cross Interferometer Array Configuration for Meteor Radar Applications using a MATLAB Based Planar Array Design Package," University of Colorado. 16 March 2007 <http://sisko.colorado.edu/palo/Meteor_Radar_Info/COSPAR_arrayRevised.pdf>

[6] S. J. Franke, X. Chu, and A. Z. Liu. “Comparison of meteor radar and Na Doppler lidar measurements of winds in the mesopause region above Maui, Hawaii,” University of Illinois. 29 January 2005. <http://cires.colorado.edu/science/groups/chu/pubs/documents/2005JGR_Franke_MauiRadarLidar.pdf>
[7] Genesis Software. SKiMET. 19 July 2013 <http://www.gsoft.com.au/productsandservices/skiymet>
[8] Jicamarca All-sky Specular MEteor Radar (JASMET) Website. <http://jro.igp.gob.pe/english/facilities/JASMET/JASMET.htm>
[9] Alexander Hackett. “On the Development of Modern Ionospheric Sensors Using Software-defined Radio Techniques.” Master's Thesis. The Pennsylvania State University, August 2013.
[10] CPU Timers, Boost C++ Libraries. <http://www.boost.org/doc/libs/1_54_0/libs/timer/doc/cpu_timers.html>

WE'RE LUCKY TO HAVE THE MOON

Niklas Anthony* and John Mathews[#]

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

*Undergraduate Student of
Department of Electrical Engineering
Pennsylvania State University
State College, PA 16801

ABSTRACT

Space is said to be the final frontier. Space, however, is astronomically expensive to explore. The amount of energy needed to move just one kilogram into low-earth orbit is immense. Not only do we need to counteract Earth's gravity, but also its atmosphere. It would be so much cheaper and more efficient if we could get things into space some other way. Luckily, we have a neighbor capable of providing a solution.

The goal of this research project will be to conduct an orbital energetics study of using the moon as a base point to get to various points in the solar system. The only way to launch craft from the moon would be to send fuel there from Earth, which is inefficient, and defeats the purpose of using the moon. An analysis of building a Lunar Launch & Retrieval Rail is conducted as a solution to this issue.

INTRODUCTION

Orbital Equations

An object feels many forces at once. One such force is gravity. Sir Isaac Newton first quantified gravity in his 1687 publication: *Philosophiae Naturalis Principa Mathematica* [1]. His observations led to a very important equation:

$$\vec{F} = -G\frac{Mm}{r^2}\hat{r} \tag{1}$$

He showed that two objects feel an attractive force proportional to their masses, *m* and *M*, and distance between them, *r*, squared, where *G* represents Newton's

[#] Faculty Mentor

gravitation constant. When analyzing how objects work, relative to planets, the objects mass, *m*, is usually vastly smaller than that of the planet, *M*. It is usually ignored for this reason; we can, however, remove its mass parameter by solving for an object's acceleration instead. Substituting the basic equation of force:

$$\vec{F} = m\vec{a} \tag{2}$$

$$\vec{a} = -G\frac{M}{r^2}\,\hat{r} \tag{3}$$

This now states an object's acceleration due to gravity is proportional to its parent planet's mass and inversely proportional to distance from the center of it. This says that, if no other forces are acting on an object, it will always fall towards the center of the planet. In order to keep an object from falling, we give it an orbit. We do this by essentially moving so fast sideways, that as it falls, it misses the planet entirely. The force from this added velocity can be better seen in the equation:

$$\vec{F} = \frac{mv^2}{r}\,\hat{r} \tag{4}$$

Christiaan Huygens wrote this in 1659 [2], showing that an object travelling in a circle feels an outward force proportional to its mass, *m*, tangential velocity, *v*, squared, and inversely proportional to the distance between the object and the focus of its implied ellipse, *r*. Using the same process as before, we can eliminate the mass scalar and analyze its radial acceleration instead:

$$\vec{a} = \frac{v^2}{r}\,\hat{r} \tag{5}$$

We can add Equations 3 and 5 to describe an orbiting object's total force as being its force felt from gravity and its centripetal force as:

$$\vec{a} = (\frac{v^2}{r} - \ G\frac{M}{r^2})\,\hat{r} \tag{6}$$

When in a circular orbit, the force produced by the tangential velocity negates the force of gravity. We know for a fact though that there exist elliptical orbits, and that most massive bodies follow elliptical paths [3]. This occurs when there is a non-constant angular velocity. A body may travel faster or slower than its "circular" velocity at its respective radial distance. If it is travelling faster, it will fly further away on its opposite side of the circle (now ellipse); slower, it will fall. Assuming no outside forces, these orbits are stable, as seen in our solar system. The point at which the body is travelling the fastest is the closest it will be to its

respective parent body, and is called its “periapsis”. The point at which it is travelling the slowest will be its furthest distance, its “apoapsis”.

As orbiting objects traveling in circles and ellipses, we can observe not only radial acceleration, but angular position, velocity, and acceleration. Derived from Noether’s theorem, it is known that in the absence of external torque, a system will have a constant angular momentum [4]. We know that in a circular orbit, angular velocity is constant, and thus, angular acceleration is zero. With now non-constant tangential/angular velocities, we can observe a non-zero angular acceleration, exemplified by the equation:

$$\vec{a} = \frac{-2r'\theta'}{r}\hat{\theta} \tag{7}$$

Where ‘r’’ is radial velocity, ‘θ’’ is it’s angular velocity, and ‘r’ is the distance to the center of the parent body. *Note*: if we change angular velocity, all 3 variables: radial velocity, angular velocity, and radius, are directly affected, effectively changing our orbit. Also note that tangential velocity can be rewritten as its angular velocity multiplied by its distance to the center of the parent body

We can use both Equations 6 and 7 to accurately model a polar representation of orbital dynamics.

Transfer Orbits

A basic necessity when flying in space is the ability to change orbits. Be it transferring from low-Earth to lunar orbits, low-Earth to Martian, lunar to Martian, and so forth, all require getting from one circular orbit to another. There are many ways to do this, but all depend on one thing: increasing or decreasing angular velocity relative to the parent body: Earth, Sun, Moon, etc. Spacecraft do not have to orbit in circles. If there is no need to circularize the orbit, energy can be saved.

There are many factors to include when deciding how to transfer orbits: energy cost, time, and position. A method was developed by Walter Hohmann in 1925 that minimized direct energy input [5]. Known as the Hohmann Transfer Orbit, the idea is to create an elliptical orbit with its apoapsis and periapsis being the starting and finishing circular orbital radii. Using the conservation of angular momentum, one can find the orbital speeds that a body would be travelling at both periapsis and apoapsis. These two values are found using the equation:

$$v^2 = GM\left(\frac{2}{r} - \frac{1}{a}\right) \tag{8}$$

where a is the ellipse’s semi-major axial length. This is found by adding the new and old circular radii and dividing by 2. One finds that these velocities differ from that of their respective circular counterparts. In the case of increasing orbital size, the velocity is higher than the circular speed at periapsis, and less at apoapsis. This means that if one wants to go from a small orbit to a Hohmann transfer

ellipse, an increase in angular velocity is needed. Changing angular velocity requires addition of energy. This is done currently by burning fuel in rockets. For simplicity's sake, we assume that this energy addition is done instantaneously at the correct position in space. In reality, it takes time to increase velocity, which alters calculations, and is not the primary focus of this experiment. Once the angular speed has been increased to the new velocity, thrust is cut, and the body now exhibits an elliptical orbit. Skip forward in time to when the body is now at apoapsis, and another increase in velocity is needed to re-circularize the orbit.

A faster means of transferring orbits is to increase the periapse speed past the transfer orbit; a more complicated and energetic re-circularization burn is needed when it reaches the target distance, however.

Multi-body Problem

We have thus far been able to accurately describe orbits of relatively small bodies around much larger parent bodies. In reality, however, these small bodies respond to gravitational attraction forces from many other objects, planets, stars, moons, etc., at once. And as all of these objects move in their respective orbits, so too does their influence on orbiting bodies.

This complicates orbital calculations, as we must now find a common barycenter, or center of gravity, of all of these objects, plot their respective orbits around it, and only then can analyze how a small body would react in this shifting gravitational field.

David D. Meisel developed a solution space to the 3-body problem involving the Earth and Moon. He followed the works of H. Pollard in creating a rotating Cartesian system [6]. Essentially adding another copy of the orbital equations to the total, he can factor in the Earth and Moon. He experiments with launch velocities, destination orbit, and launch angles (as the Earth is on a tilt) to find the optimal cislunar trajectory [7].

For the purposes of this paper, we assume that orbital transfers are done by using Hohmann transfer orbits to get to desired radial lengths, and that the destination massive body will be there when the body arrives at its orbit, and a respective capture burn is done to get into orbit around it.

Atmospheric Resistance

Earth has a thick atmosphere which shields us from the sun's radiation, keeps us warm, and contains many life-essential molecules. This, however, acts as a gigantic barrier that spacecraft must pass through. Conventional aircraft utilize it as a means of low-altitude flight. In a zero-atmosphere environment, spacecraft are able to orbit around a planet at very low altitudes, as there is nothing in their way to slow them down. The only consideration is the shape of the surface. Crashing into the side of a mountain at orbital speeds will most likely destroy any spacecraft. A formula for drag was developed by Lord Rayleigh that can be applied to any fluid, our atmosphere being an example [8].

$$F = \frac{1}{2}\rho v^2 C A \frac{1}{a} \quad (9)$$

This states that an object feels a resistive force proportional to the mass density of the fluid, ρ, the drag coefficient of the fluid, *C*, its exposed cross-sectional area, *A*, and its velocity, *v*, squared. The last part of this is what concerns rocket scientist the most. The faster one travels in the atmosphere, the more resistive the force becomes. For instance, a low-Earth orbit of around 30 km above the surface requires a tangential velocity of almost 8 kilometers per second. Square that, and one gets massive resistive forces. We observe that mass density decreases as altitude increases, but never quite disappears, even at low-Earth orbits above 300 km. This means periodic addition of energy is needed to maintain normally circular orbits.

Propulsion/Launching

The only way we have been able to get rockets into orbit is by burning chemicals to create thrust. Every rocket has a specific impulse, or its force with respect to the amount of propellant used per unit time. Different propellants and fuels have different properties, leading to different specific impulses. Some rockets are more efficient than others. Tiny home-launch rockets usually have small solid-propellant-based cartridges, whereas orbital launch rockets have gigantic liquid-fuel engines. Regardless of state, burning chemicals wastes a lot of energy as heat instead of thrust. People have looked into different ways of getting things into orbit.

One idea was a space elevator. A space elevator would consist of essentially a large pole that stretched taller than anything anyone has ever seen before. It would extend from the surface to past geosynchronous orbit, so that the tip of it was travelling at proper orbital speed (1 rotation per day). A counterweight attached past geosynchronous orbit would essentially be travelling faster than orbital speed, but forced to maintain its radial distance. This creates tension on the tower. If there is enough tension, the compressive force felt close to Earth's surface would be lessened, and fewer structural supports would be needed to keep the tower upright. Geosynchronous orbit around Earth is at approximately 42,100 km above the surface. At this distance, if a payload was transferred to this height, it would merely have to "let go" of the central pillar, and would already be in orbit, without the need to add energy [9].

Maglev Technology

There are other systems that would be an effective solution to launching without rockets. Another method, which will be the primary focus of this paper, is using electromagnetic power. The concept is similar to that of a maglev train, high-speed roller coasters, and rail guns. The idea is to accelerate a payload or transport cart to orbital velocities using a series of magnets. The advantage of using a maglev system is that it eliminates the negative force of friction between

any wheels and the ground. The rail that the train sits on both levitates the train off of the "track" and propels the train forward. There are two forms of maglev train technology: electromagnetic and electrodynamic. Each uses the precise timing of powering electromagnets to propel a train forward.

Electromagnetic systems use powerful electromagnets and a series of very precise feedback mechanisms on board the train to maintain a certain distance between the train and the rail. This process of levitation is dynamically unstable, meaning if there is a slight change in the distance between the rail and train, resulting forces will be exponentially increased. If it moves closer to the rail, it wants to attract quickly. If it moves away, it wants to repel more. Its primary advantage is that it can be used at low speeds, requiring no wheels or moving parts. Another advantage is that it is proven to work and in use around the world today.

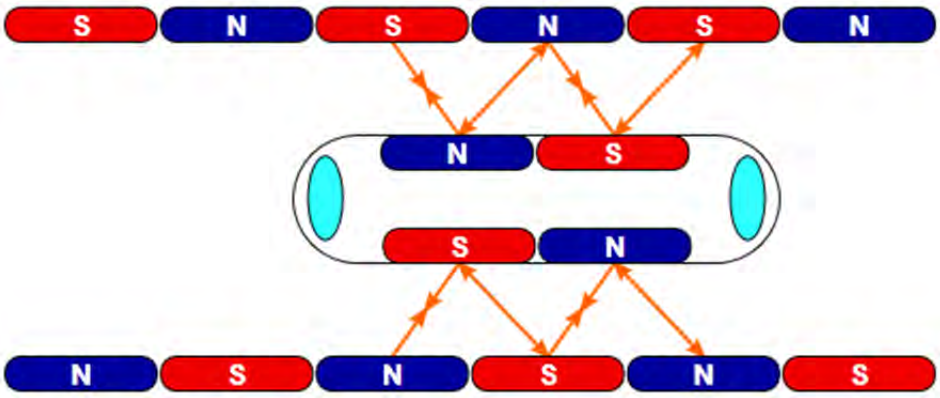

Figure 1: Maglev Propulsion

Electrodynamic systems have magnets both on the train and part of the rail. The attractive/repulsive forces are produced by permanent magnets in the train and induced magnetic fields in the track. The advantage of this technology is that it is dynamically stable; that is, any variation in distance will generate a counter force, pulling the train back to center. The disadvantage of this is that at low speeds, the necessary induced current in the track becomes too high to maintain, and the track becomes incapable of suspending the train. This necessitates the use of wheels or moving tracks at low speeds. At high speeds, this effect doesn't have enough time to build up. Another disadvantage is the high magnetic fields in the train itself. People with pacemakers or magnetic data storage devices would have problems [10].

Both technologies run into trouble when travelling at high speed. The atmospheric drag becomes too strong for the magnets to counteract, and a maximum speed is reached. In a zero-atmosphere environment, however, much higher velocities can be reached.

EXPERIMENT DESCRIPTION

Getting to Low-Earth Orbit

The purpose of this paper is to analyze the energetics of earth–moon transport, and some inter-planetary energetics analysis as well.

To begin, we analyze launching payloads into low-Earth orbit. Here we factor in orbital dynamics and atmospheric resistance. Earth's mass is 5.972×10^{24} kg. Its average radius is 6.371×10^{6} m. If there were no mountains or atmosphere, an orbit of just 1 meter above the surface would require a tangential velocity of 7,873 m/s. An instantaneous acceleration to that speed would require 31.0 MJ/kg. We know this wouldn't work, as the atmosphere would eventually slow the rocket down enough that a crash would result.

Currently, the method of getting payloads into orbit is using rockets. The amount of change in velocity capable of being produced by a rocket is given by the ideal Tsiolkovsky rocket equation:

$$\Delta v = v_e \ln \frac{m_i}{m_f} \tag{10}$$

It shows how much delta-v a rocket can produce given its specific impulse relative to gravity, v_e, and its initial mass, m_i, and final mass, m_f [11]. This does not factor in atmospheric resistance, a huge resistive force source.

The atmosphere is a fluid, but lacks a well-defined upper bound, as it is gravitationally contained. It just trails exponentially off into nothing. For this reason, we approximate the "top" of the atmosphere to be at the Karman Line. At 100 km, it is the boundary where orbiting bodies would have to travel faster than orbital speeds to maintain lift using conventional wings. Sputnik-1 orbited around Earth at around 215 km and the International Space Station orbits at around 340 km. We consider a Low Earth Orbit (LEO) to be a safe 370 km above the Earth's surface. Assuming the initial radius is Earth's surface on the equator, its initial orbital speed to be how fast the earth rotates, to get from the surface of Earth to 300 km, using a Hohmann transfer, would require an initial increase of 7,536.7 m/s, or 28.4 MJ/kg, and an additional 89.4 m/s or 4.0 kJ/kg to circularize the orbit.

Due to the atmosphere, however, a significantly larger amount of energy is needed. Using the general drag equation, and setting mass density of the air to be a decreasing exponential function, relative to r^2, we can predict, generally, the resistive force, and get a more accurate energy quote. However, due to the nature of the program used, which assumes orbital change burns are instantaneous, all attempted orbits became highly eccentric, with initial orbital speeds of 10 billion m/s. For this reason, and applying realistic circumstances, getting out of Earth's atmosphere requires a different kind of transfer orbit. In reality, rocket launches are not nearly parallel to Earth's surface, but directly up. This is to get as far away from Earth's thicker part of the atmosphere as fast as possible. Rockets then

slowly begin to tip over, gaining lateral speed as the atmospheric density decreases.

After some manipulating of the code to run under "constant" acceleration upwards to 1/6 the atmosphere height, then at "constant" acceleration at 45 degrees to the full 100 km, then an instantaneous burn to get to orbital speed, then an instantaneous burn to transfer to 300 km, and finally, the circularization burn. The "constant" acceleration was running the code at relatively small increments, each time adding additional initial velocity. We had to add 1640 m/s for the upwards burn, 5120 m/s during the 45 deg burn, 5237 m/s to get to orbital speed, 59.6 m/s to transfer and 59 m/s to circularize. That's a total of 12115 m/s or 73.4 MJ/kg [12].

From Low-Earth to Beyond

From the 300km orbit, we can safely ignore the atmosphere, focusing simply on orbital dynamics. When travelling to other planets, one must escape Earth's sphere of influence. Using the concept of conservation of energy, in that final velocity would be zero, and final distance is infinity, we find the escape velocity of any payload relative to its parent body to be:

$$v_e = \sqrt{\frac{2GM}{r}} \tag{11}$$

So, to escape Earth from a 300-km orbit, it would require a speed of 10931.3 m/s. From orbital speed, that's an increase of 3201.7 m/s, or 5.13 MJ/kg.

From there, transferring to other planets is done in the same way transferring orbits around Earth is done. Below is a table of delta-*v* requirements to transfer to various points in the solar system.

Table 1: Delta-V for Earth-to-X Transfers

To/From	LEO	SS	GEO
	3201.7	1609.7	1187.0
LEO	×		
SS	3462.8	×	
GEO	3976.9	985.2	×
LUN	3934.5	2076.5	1492.0
VEN	5161.9	<	<
MAR	5562.6	<	<
NEP	15675.9	<	<

The table shows the delta-*v* to both transfer and circularize a Hohmann transfer orbit in m/s. LEO is a Low-Earth Orbit at 300 km. SS is a Semi-Synchronous Orbit with orbital period of half a day at 20200 km. GEO is full Geosynchronous Orbit with orbital period of a full day at 42164 km. The three

values at the top represent the escape velocities at those orbits, thus to get interplanetary delta-v, one must add those values to the values below it. Note that to transfer from LEO to Lunar orbit requires less delta-v than LEO to GEO. This is because the initial burn would be near escape velocity, and requires less delta-v to circularize. Earth to Venus has a negative delta-*v*, implying it will decrease periapsis distance relative to the Sun. Earth-to-Neptune was added to emphasize how far out our solar system reaches [13].

The Moon

The Moon orbits the Earth at around 384,400 km away. It has a mass of 7.348×10^{22} kg and average radius of 1.737×10^{6} m. It is approximately 100 times lighter than Earth, and about 1/3 the radius. It also has virtually no atmosphere. This means that a spacecraft could orbit at 1m above the surface, assuming there is a perfectly flat orbital path.

Getting to Low-Lunar Orbit

NASA states that the highest point on the moon is around 11 km above the surface. Because there is no atmosphere, optimally, one could launch directly sideways, into a Hohmann transfer orbit. Going from the surface to a 15-km orbit, one must add 1683.6 m/s to transfer, and another 3.5 m/s to circularize. This totals to about 1.42 MJ/kg.

From Low-Lunar to Beyond

From a 20-km orbit, the Moon has an escape velocity of 2,362.4 m/s, an increase of 691.94 m/s, or addition of 0.239 MJ/kg. The fact that the moon orbits at around the escape distance of Earth, makes it a little bit easier to move from orbit to orbit. To get to lower planets, one would launch when the moon is closest to the sun, and to get to further planets, one would launch when it is furthest. At almost every point in Lunar orbit, burning to escape velocity will have an apoapsis capture encounter with Earth. If a spacecraft accelerates to escape velocity when it is angularly behind the Moon, relative to Earth, the Moon will catch up to the craft, and either a new orbit will be created, or the Moon will crash into the craft. At Lunar distance from Earth, one needs only to increase its speed relative to Earth by 390 m/s to escape its influence.

EXPERIMENT RESULTS

When we launch from Earth, we pay for being conveniently located near our home planet. The atmosphere is extremely difficult to account for. Two things that were not accounted for in this analysis were wind and launch location. We assumed the atmosphere was a constant mass density, and degraded exponentially to around 125 km. We know that there can be very strong winds at high altitudes, clouds change the mass density, and the “top” of the atmosphere may vary by several dozen kilometers. The Earth also does not “line up” with the solar plane. We know the earth is tilted by about 23 degrees. This implies that we must add

even more delta-*v* to "flatten" our orbit, before we can launch to other planets. We also do not have a launch facility located on the equator, the best area to launch.

When we launch from the Moon, we save over 50 times the energy to get into a low, stable orbit. However, with current technology, we don't have the infrastructure on the moon to be launching from there. Any fuel we would need to launch from the moon would have to come from Earth, or mined from below the Lunar surface. We also note that the moon does not rotate about an axis. It is locked in a direction such that only one side of the moon is visible from Earth. The moon does not have an even mass distribution and is tidally locked to Earth. Theories from when the moon first arrived in Earth orbit suggest a collision. This shifted the center of mass of the moon away from the geometric center. When orbiting, bodies move about their center of mass. This would mean we need a slightly higher orbital distance to prevent lunar surface collision.

DISCUSSION

If we ever have hope of exploring our solar system and local galaxy, we simply cannot launch from Earth. As mentioned, we have no way of launching rockets from the moon, as there is no fuel station there, or means of self-sustaining acquisition of fuel. For this reason, we proposed the Lunar Launch and Retrieval Rail, LLRR. Situated on the far side of the moon, this structure would consist of a powered rail shaped like a transfer orbit between the surface and a low-Lunar orbit, say 20 km. It would accelerate spacecraft from the surface to orbital speed, and release the craft into the orbit. It would also dock and decelerate incoming craft from that orbit to the surface.

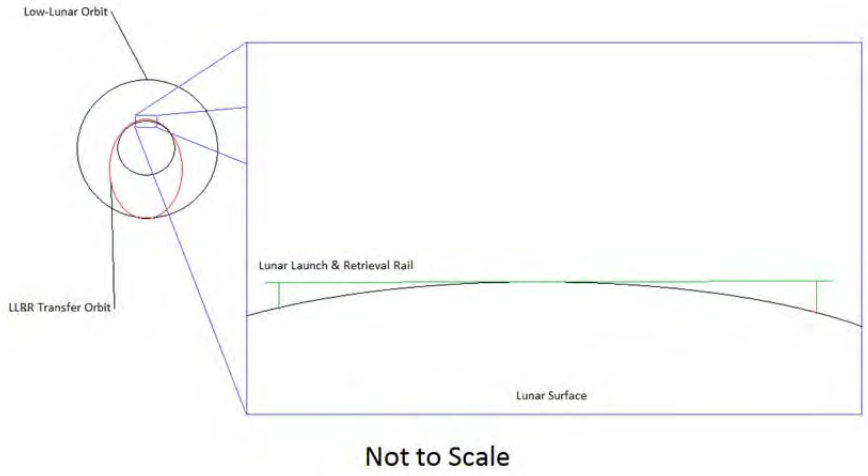

Figure 2: Lunar Launch & Retrieval Rail concept

In order to accommodate human payloads, the acceleration would have to be low, around 40 m/s^2. With this constant, and the values obtained from the transfer orbit code, we can calculate how long the track must be, and how tall it must be at the ends. The orbital speed near the surface of the moon must be around 1650

m/s. This would require an acceleration period of 41.25 seconds. Basic kinematics state the track must then be around 34 km long. Assuming the moon is a perfect circle, it would be only 6.5 m tall. That is assuming a start at the surface. We can shift our starting point to be on the far end of the decelerating rail and still accelerate the 34 km. This would only require both ends to be 1.1 m off of the surface. The track would have to be slightly longer; however, there needs to be room for the empty cart/gondola to accelerate to orbital speed to catch up with the incoming payload, and room for docking time, on both sides.

It would be powered by solar energy, either on the sides of the rail itself, or at a base station nearby. This solar energy would power an electromagnetic rail system, as travelling at both very high speed and low speed is necessary for docking and launching. Some energy can be harnessed by landing payloads and stored for use in launch later. The materials needed to create this will have to be launched from Earth, and can either be assembled by robotic spacecraft, or by humans at a pre-established Moon base, however using robotics would be cheaper.

CONCLUSION

Space is the final frontier. The huge volume of even the inner solar system demands the use of many exploratory spacecraft. The energetics study done in this paper proves that launching from Earth is just not economically feasible. We are lucky enough to have a relatively close Moon, which has low gravity and no atmosphere, a very suitable candidate for a base for exploratory craft to launch and arrive at.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, N.A. would like to thank these people: John Mathews, Sven Bilén, and David Meisel.

REFERENCES

[1] Smith, George, "Newton's *Philosophiae Naturalis Principa Mathematica*", *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/win2008/wentries/newton-principia/>.

[2] Ernst Mach, *The Science of Mechanics* (1919), e.g. p.143, p.172, and p.187

[3] Smith, George, "Newton's *Philosophiae Naturalis Principa Mathematica*", *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/win2008/wentries/newton-principia/>.

[4] Noether, Emmy, "*Invariante Variationsprobleme*". *Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse.* (1918) p. 235-257

[5] Hohmann, Walter, "*The Attainability of Heavenly Bodies*", Washington: NASA Technical Translation (F-44 1960)

[6] Pollard, H, "*Mathematical Introduction to celestial Mechanics*", Prentice-Hall (1966)

[7] Meisel, David D., "DDM_translunar.nb"

[8] Batchelor, G.K., *"An Introduction to Fluid Dynamics"*, Cambridge University Press (1967)

[9] Semon, Theodore, "*Climb*", ISEC Journal of Space Elevator Developments and Technology Volume 1, Number 1 (2012)

[10] Scott, David, "*Maglev: How they're Getting Trains off the Ground"*, Popular Science (Dec. 1973), p. 135
[11] Benson, Tom, "*Ideal Rocket Equation*", NASA Exploration Education, <http://exploration.grc.nasa.gov/education/rocket/rktpow.html>
[12] Anthony, Niklas, "atmosphere.m"
[13] Anthony, Niklas, "transfer.m"

SYSTEM-LEVEL DESIGN OF ACTIVE INFLATABLE FALLING SPHERE FOR USE WITH UPCOMING SUB-ORBITAL FLIGHT OPPORTUNITIES

Gregory Roberts,* Sven G. Bilén,[#] and Robert Capuro[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
University of California, Berkeley
Berkeley, CA 94704

ABSTRACT

Falling sphere systems measure properties of the mesosphere like temperature, density, and winds by monitoring acceleration as they fall. The modern day approach is to outfit an inflatable sphere with sensors and a telemetry system instead of tracking the sphere with radar. This paper presents a first-order design of this device with respect to new methods to realize an antenna for the telemetry system and verification of noise floors for several accelerometers. Antenna simulation results were able to identify promising antenna patterns in terms of gain and beam pattern. A method for fabricating a flexible antenna is also suggested and carried out with good results. Finally, three of five of the tested accelerometers were shown to perform to within or close to within their noise floor specification.

INTRODUCTION

Falling Sphere

Arguably the least well characterized of the Earth's atmospheric layers is the mesosphere due mostly to the difficulty of designing realizable measurement techniques. The mesosphere sits between 60 and 90 kilometers above the Earth's surface, inconveniently located above practical airplane flight and below a satellite's orbit. A falling sphere is one of the more common methods for gaining *in situ* measurements in the mesosphere.

[#] Faculty Mentor
[+] Secondary Mentor

Falling spheres are ejected from a suborbital rocket (or rocketplane) and allowed to naturally fall towards the Earth. From the sphere's kinematic data, information is inferred about atmospheric density, temperature, and winds [1]. In contrast to inflatable spheres, rigid spheres have been designed for rocket delivery to higher altitudes than those achievable by inflatable sphere carriers. Inflatable falling spheres offer larger cross-sectional area at a lighter weight for a given payload size requirement because they are inflated after ejection. A larger cross-sectional area proves useful for the collection of reliable data. The current technology has also moved away from passive falling spheres, which were absent of onboard electronics and downlink systems. These spheres were tracked by radar on the ground and then properties of the mesosphere could be calculated from the sphere's positional data. The concept now is to deploy an active falling sphere, one which contains sensors such as accelerometers, GPS receivers, and thermistors along with a transmitter system to broadcast information back to the ground. These active inflatable spheres have an opportunity for frequent deployments due to NASA's recent sponsoring of XCOR Aerospace for a sub-orbital launch program. Lightweight active inflatable falling sphere systems are once again an interesting and useful technique for the study of the mesosphere.

Positional and kinematic data gathered during the flight can be correlated to properties of the atmosphere for various sphere sizes and weights by overlapping previously made wind tunnel measurements and a fluid mechanics numerical simulation called Direct Simulation Monte Carlo (DSMC). In order for this method to work reliably, the mass of the sphere should not be significantly larger than 160 grams, specifically not more than 290 grams. Outside 290 grams, there is heavy dependence on DSMC, which reduces the integrity of the results. This mass limit places restrictions on mass of the electronics, and also on the radius of the sphere and its inflating gas. The radius is related to the cross sectional area of the sphere, which in turn is positively related to drag force on the sphere. An increased drag force for a given mass creates larger accelerations on the sphere. With larger accelerations, it is easier for the sensing system of the sphere, specifically the accelerometers, to accurately record the sphere's dynamics.

SYSTEM OVERVIEW

Essentially, the electronics in the sphere have to accomplish two tasks. First, several analog signals must be digitized, multiplexed, and delivered to a ground station through a telemetry system [2]. This encompasses almost the entire diagram shown in Figure 1. Accelerometers measure the acceleration vector of the sphere and the magnetometers provide orientation information by detecting Earth's magnetic field with respect to the sphere [1]. The GPS receiver provides position information. Battery levels during the flight also get transmitted back as housekeeping. Second, an inflation system must be triggered when the sphere is launched. This includes the signal to start the inflation process as well as controlling inflation to the right pressure. Overinflating the sphere upon launch is absolutely necessary. Atmospheric pressure increases during descent and, in

order to prolong the useable flight time before collapse, the sphere must be at a high initial pressure.

Many of these pieces have been designed previously and are not performance bottlenecks [1]. However, there are a couple of pieces of the system that would benefit from updated technology and new ideas. First, the acceleration acquisition system benefits from newer low power, lightweight MEMS accelerometers. Reducing power in the electronics translates to a lighter battery or, alternatively, more power to feed the antenna. Assessment of the MEMS accelerometer noise floor is needed, however, to verify they can be used in this application. Second, the antenna used in the previous design was suspended from the bottom of the sphere [1]. This posed limitations to what antennas could be used and affected the assumption that the sphere was perfectly round. Flexible antennas can sit on the surface of the sphere and ideally provide the same performance. These two limiting design factors are explored in this work.

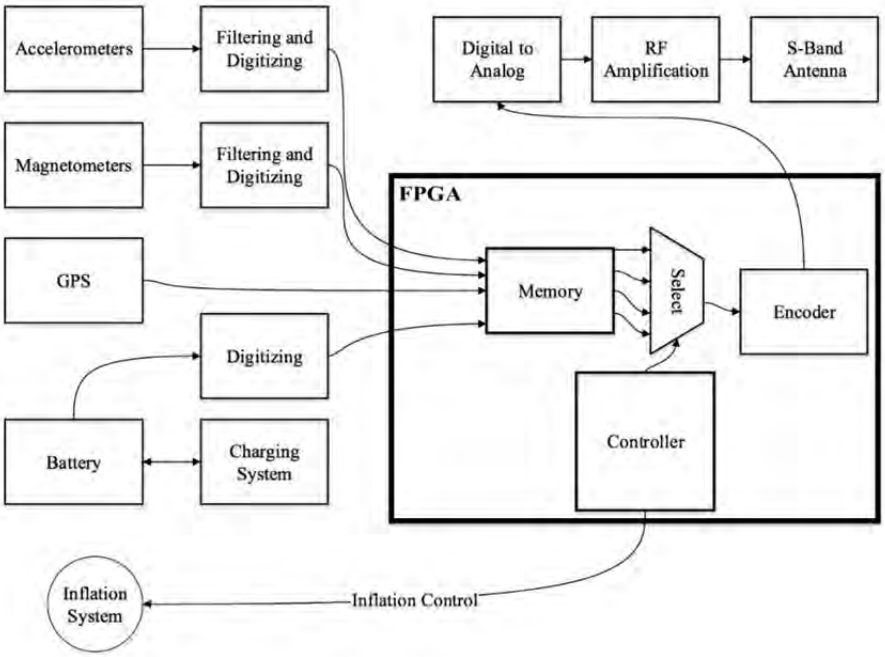

Figure 1: Block diagram of falling sphere instrumentation and telemetry system.

EXPERIMENT DESCRIPTION

Antenna Simulation and Modeling

Several antenna patterns were constructed in CADFEKO, an electromagnetic modeling suite. The software provides an interface to create custom antennas and observe performance characteristics like input impedance, far- and near-field radiation patterns, gain, beam width, and several other useful quantities. In addition, optimization loops can be run to realize desired qualities by tweaking parameters that describe the antenna. In the past, an S-band antenna (2–4 GHz) is used for this application [1]. While a typical transmitting frequency assigned by

NASA would be around 2.27 GHz, antennas were simulated in free space at 2.6 GHz, for convenience.

Antenna patterns were chosen to sit on the outside of the falling sphere. This did not necessarily limit the types of antennas that could be chosen, but instead required that they be modified to wrap around the sphere. Keeping the antenna flush with the spherical surface ensures it will not significantly affect the motion of the sphere as it falls. Several antenna patterns were evaluated in simulation and are presented below; they are a mixture of classic antenna concepts like dipoles and newer concepts like random fractals (see Figures 3, 4, and 5). It is convenient to describe antenna patterns with spherical parametric equations since they exist on the surface of a sphere. For reference, the spherical coordinate system used is shown in Figure 2.

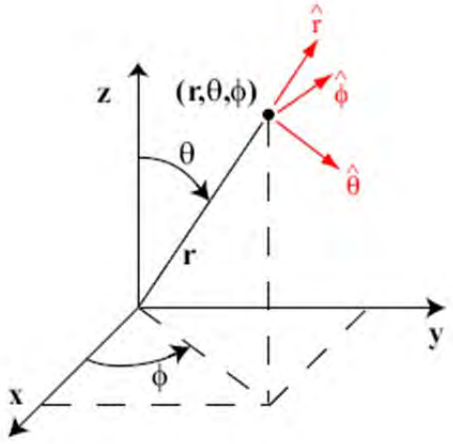

Figure 2: Reference spherical coordinate system used for modeling (image from http://www.livephysics.com/physics-equations/classical-mechanics-eq/coordinate-systems/).

Lambda Wrapped Dipole

A classic antenna concept is the dipole, which was reproduced as a wrapped dipole. Each leg spans one wavelength and curves around the sphere; the dipole is also modified with a width. The basic shape can be described in spherical coordinates by the following parametric equations, while the full version comes from sweeping this line across the surface of the sphere by the width:

$$\frac{-\lambda}{R} \leq t \leq \frac{\lambda}{R}$$
$$r(t) = R$$
$$\theta(t) = t$$
$$\phi(t) = c$$

where R is the radius of the sphere and c is a constant.

Square Patch

The square patch is similarly defined to the wrapped dipole; simply, a dipole is made with identical width and length.

Slotted Ring

The slotted ring is made up of a spherical band with several slots cut out of it. In spherical coordinates, the band is described by the parametric equations:

$$\left(\frac{\pi}{2} - \frac{w}{R}\right) \leq t \leq \left(\frac{\pi}{2} + \frac{w}{R}\right)$$
$$0 \leq s \leq 2\pi$$
$$r(s,t) = R$$
$$\theta(s,t) = t$$
$$\phi(s,t) = s$$

where R is the radius of the sphere and w is the width of the band. The slots follow the same pattern except the parameter s sweeps over a limited range and the width, w, is reduced. Subtracting the slot surfaces from the band surface generates the slotted ring antenna.

Spherical Helices

The spherical helix shape is described by the following parametric equations:

$$0 \leq t \leq N\pi$$
$$r(t) = R$$
$$\theta(t) = v\cos^{-1}\left(\frac{t}{N\pi}\right)$$
$$\phi(t) = t$$

where v is a scaling factor for how fast θ changes and N is twice the number of turns. The above equations describe a single arm, but many spherical helix designs incorporate two or even four arms that all connect at the top of the sphere [2]. A second arm can be added by including another curve with the same equations as above except negating the $\theta(t)$ term. For a third and fourth arm, a new curve can be added for each of the two original arms with a 90 degree shift in the $\phi(t)$ term.

Work has been done in the past with spherical helix antennas [3,4,5]. However, the spherical surface was of a much smaller radius than the falling sphere. Therefore, to get a similar number of terms to what has been done in the past, the total wire length becomes much larger than that of a wavelength. To remedy this, gratings are added to the antenna. A grating is a conductive path from the feed of the antenna to the end or close to the end of the helix. The grating length can be on the order of a wavelength while allowing the helix

structure to contain multiple turns. See Figures 3 and 4, for a summary of what these antennas look like.

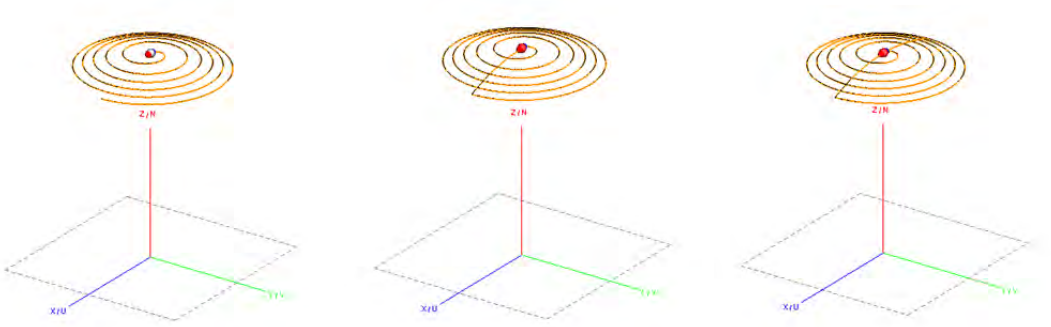

Figure 3: All of these antennas have one arm, R = 0.5 m, N = 11, v = 25. *Left:* no grating; *Center:* single grating; *Right:* double grating.

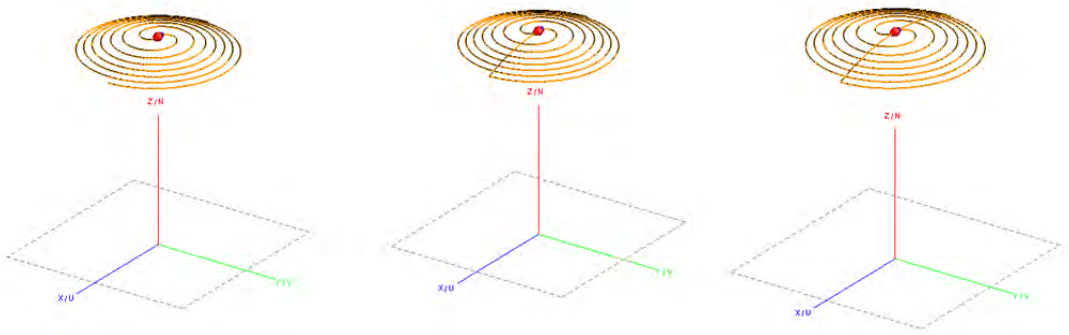

Figure 4: All of these antennas have two arms, R = 0.5 m, N = 7, v = 25. *Left:* no grating; *Center:* single grating; *Right:* double grating.

Brownian Tree Fractal

There are many ways to create fractals and work has been done with fractal antennas as they have interesting properties [6]. The type of fractal created here is a Brownian Tree Fractal. Compared to many other fractal patterns, which often involve iterative geometric processes, this is a randomly generated fractal. Conceptually, the Brownian Tree can be thought of as carrying out the following process:

1. Place a seed in a random location.
2. Place another seed in a random location and let it move randomly.
3. Let the seed move freely until it collides with a stationary seed.
4. Repeat Steps 2 and 3.

This basic concept was modified to randomly generate fractal antennas. The work flow for creating the fractal in a 2-dimensional XY plane looks like:

1. Place a seed in a random location.
2. Place another seed in a random location and let it move randomly.
3. Let the seed move freely until the next move it makes will collide with another seed.
4. Ensure that drawing a line from the last seed placed to this seed does not intersect with any other lines.
5. If it intersects, then go back to Step 2.
6. Repeat Steps 2 through 5.
7. Scale segment lengths with respect to wavelength.
8. Assign z-coordinate so that every point fits on the surface of the sphere.
9. Draw curves between the points in order so that all lines sweep along the surface of the sphere.

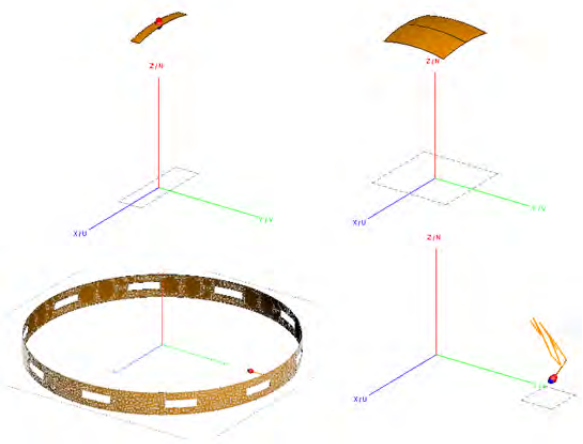

Figure 5: *Top Left:* Lambda Wrapped Dipole; *Top Right:* Square Patch; *Bottom Left:* Slotted Ring; *Bottom Right:* Brownian Tree Fractal

Building a Flexible Antenna

The idea of a flexible antenna that wraps around the outer surface of the sphere is intriguing; however, a fabrication process for the antenna is still necessary. This section presents the technique for thermal aluminum evaporation deposition onto a flexible substrate.

A thermal evaporation chamber cartoon is shown below in Figure 6. At the top of the chamber hangs the substrate being deposited onto and a crystal oscillator, which measures layer thickness by sensing its changing mass. At the base of the chamber is a wire that holds a material, in this case aluminum. After a tight vacuum is created inside with a pressure roughly eight orders of magnitude below that of the atmosphere, a large current is run through the wire. The wire gets hot due to Joule heating and evaporates the metal it is holding. This metal then expands in the chamber and evenly coats the substrate. Patterning the metal layer simply requires using a mask and placing it over the substrate. For time-related restrictions, the mask was cut manually out of plastic.

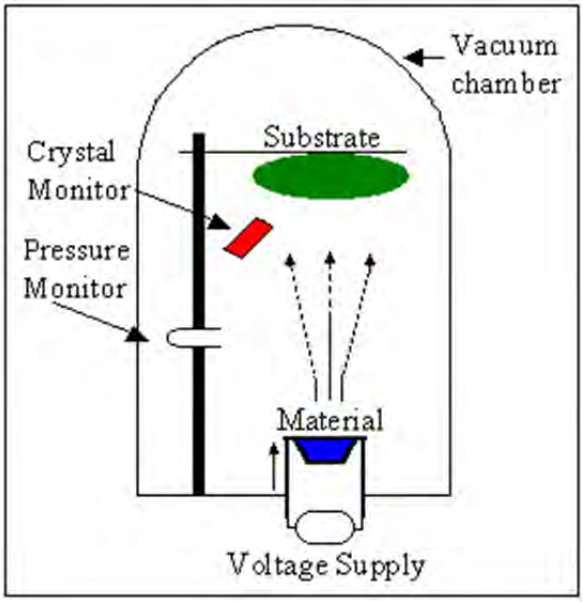

Figure 6: Cartoon of Thermal Evaporation Deposition Chamber (Image Courtesy of http://www.ece.utep.edu/research/webedl/cdte/Fabrication/index.htm)

Accelerometer Noise Floor Testing
Mechanical Test Setup

Figure 7 shows the mechanical test structure for the accelerometer tests. A PVC cube works as a scaffolding with a flat plate suspended inside with bungee cords. Using Zip Ties, a level and accelerometers were easily mounted to the structure. The goal of the mechanical test setup was twofold. It provided a consistent and convenient way of holding the accelerometers for different test static offsets. Additionally, it allowed for quick axis realignment. By simply flipping the PVC structure onto a different side, the *z*-axis could be changed to be put into a 0-g offset as opposed to a 1-g offset, for example.

The entire test setup was placed inside a room in the basement of the Millennium Science Complex at Penn State. These rooms are mechanically isolated from the rest of the building and designed specifically to achieve vibration-free environments. Reduction of vibration noise in the measurements will increase the accuracy of the noise floor measurement.

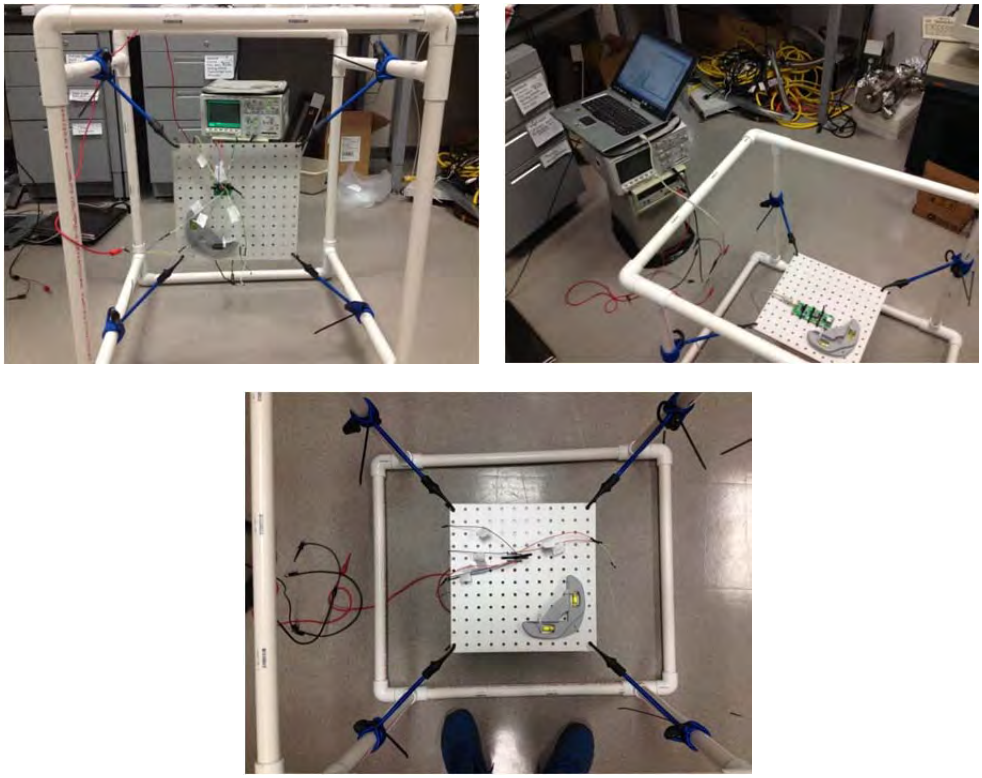

Figure 7: Pictures of Mechanical Test Setup

Test Procedure

Modern accelerometers offer either an analog output or a digital interface; both types were evaluated. All accelerometers were purchased with evaluation boards because producing custom evaluation circuits for each sensor would have been unnecessarily time consuming. Digital evaluation boards communicated with supplied software on the computer, whereas the output of the analog evaluation boards were monitored with an oscilloscope.

After mounting the accelerometer, its time trace was recorded. This time trace was then processed to find noise power to match up with the standard data sheet quote for noise performance. It is a measure of spectral density of the noise or in other words the amount of decibels of power per unit frequency (dB/Hz).

Referring now to Figure 8, integrating the spectral density, $G(f)$, between any two frequencies gives the total amount of noise power expected in that range. Equation 1 shows the integral, from which it is clear the result has units of dB. This is also known as the noise floor. Measurements at or below this level of power are not justifiably significant because there is no way to distinguish signal from noise.

$$\int_{f_0}^{f_1} G(f)df \tag{1}$$

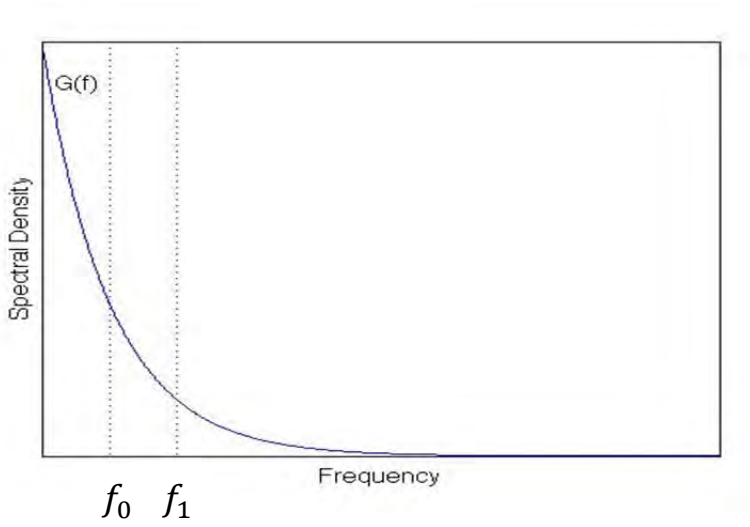

Figure 8: Sample Spectral Density Plot

A voltage output from the accelerometer is quadratically related to power, so the measure can be adjusted to get a plot of spectral density in units of μV/√Hz, which based on the accelerometer can be adjusted to μg/√Hz where g is the unit of acceleration. The figure of merit often cited on data sheets is the average spectral density.

Spectral density is very closely related to a Fourier transform of the time domain data. Data collected typically has a DC offset especially for the analog accelerometers under 0-g conditions. This is because the 0-g point is placed halfway between zero volts and the supply voltage since accelerations are measured plus/minus around zero. A DC offset places a large spike in the Fourier transform at very low and, specifically, zero Hz making the noise floor artificially too high. To correct for this, the mean of the data in the time domain is subtracted out before the spectral density is calculated.

EXPERIMENT RESULTS

Antenna Simulation Results

For a representative subset of simulated antenna patterns, far-field gains are plotted and compared below. Each antenna's gain is plotted along an angle sweep that gives it the best performance.

Spherical Helices

The far-field gain patterns for one- and two-arm spherical helices are compared in Figures 9 and 10, respectively. Comparisons are between helices with like number of arms, but different types of grating.

Square Patch and Wrapped Dipole

These two antennas were very similar as mentioned above. The dipole is a generalization of the square patch, in which the width is not restricted to be the same as the length. Figure 11 is a comparison of the special case of the square patch versus the dipole.

Slotted Ring

The slotted band antenna far field gain pattern is depicted in Figure 12.

Brownian Tree Fractal

Far field gain results from the simulation of the Brownian tree fractal, a randomly generated, fractal antenna pattern are shown in Figure 13.

Top Antennas Comparison

Several antennas performed well, in particular the one arm, single grated spherical helix, the Brownian tree fractal, and the slotted ring. The far field gain patterns of these three were placed on a single plot to get a more direct comparison. This plot is pictured in Figure 14.

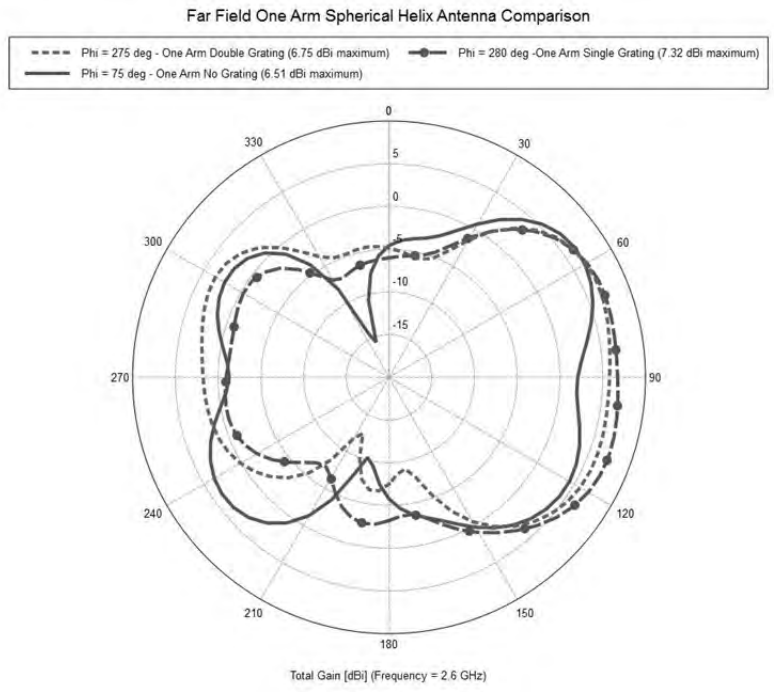

Figure 9: Far Field Comparison for One Arm Spherical Helices

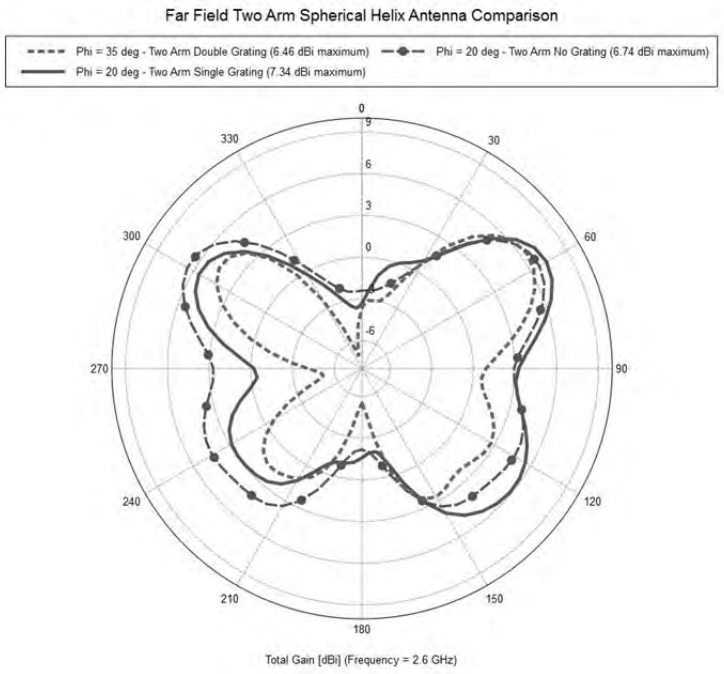

Figure 10: Far Field Comparison for Two Arm Spherical Helices

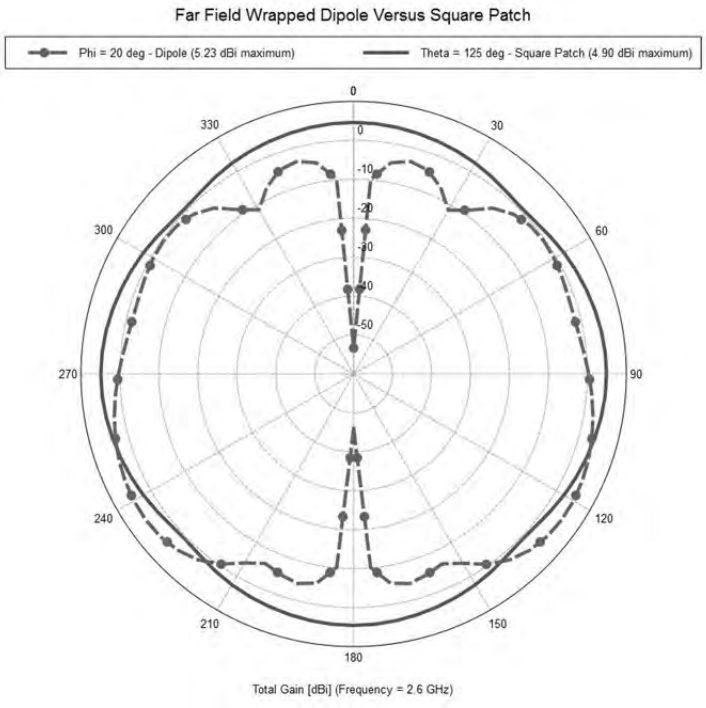

Figure 11: Far Field Comparison for Wrapped Dipole and Square Patch Antennas

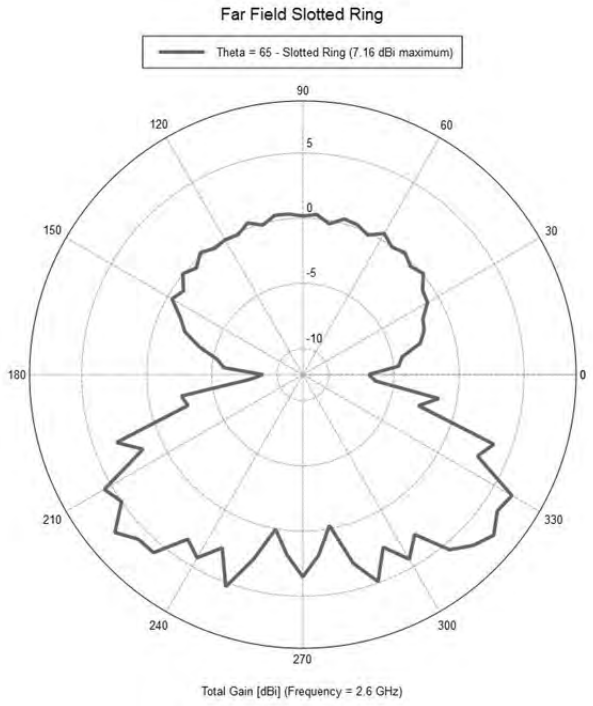

Figure 12: Far Field Comparison for Slotted Ring Antenna

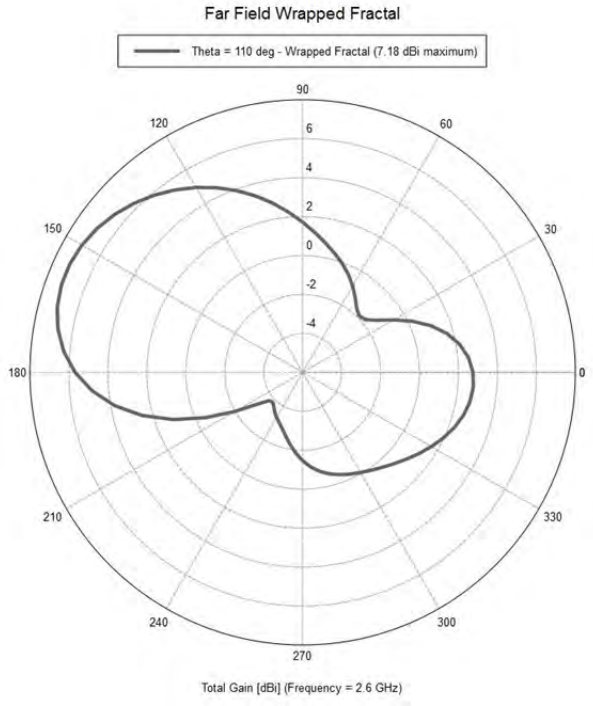

Figure 13: Far Field Plot for Brownian Tree Fractal Antenna

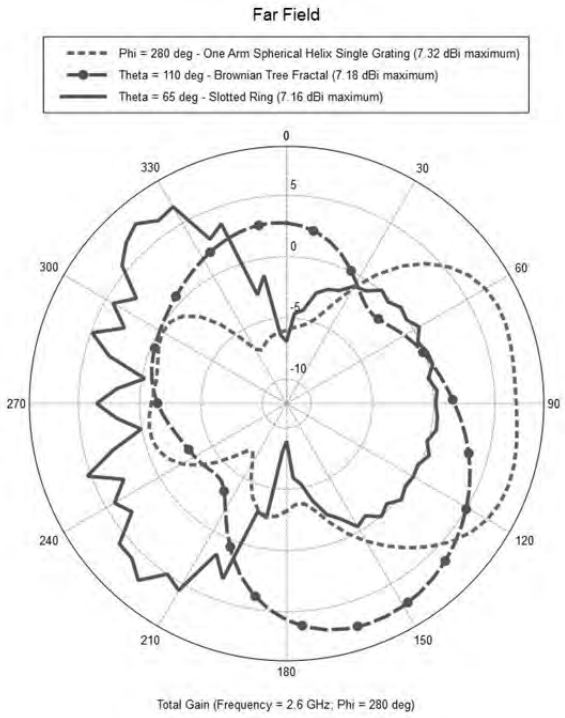

Figure 14: Far Field Plot Comparison for Top Performing Antennas

Metal Deposition Results

Depositing the aluminum in the thermal evaporation chamber was an interesting process. Starting with a piece of flexible plastic, the fractal antenna pattern was scratched onto the surface. It was given a 2-mm width and cut out with a blade. Please note that the resulting mask image pictured below (Figure 15) is from after the deposition, so while it is plastic, it has a metal coating on it. The mask was mounted on top of an identical piece of plastic. Both of these were mounted to a plate that was inserted into the chamber.

Figure 15: *Left:* Fractal Antenna Mask after Deposition Process; *Center:* Resulting Antenna from Aluminum Thermal Evaporation Deposition; *Right:* Demonstration of Flexibility of Antenna.

Overnight, the chamber's pressure was reduced to workable conditions. The metal was then heated and deposited onto the substrates. Figure 15 shows the mask used and the resulting antenna. Table 1 summarizes the characteristics of the aluminum deposition process.

Table 1: Experimental Parameters for Deposition Process

Material	Aluminum
Base Pressure	3×10^{-6} Torr
Current	35 A
Time	7 minutes, 36 seconds
Final Thickness (estimated)	101.5 nm

Accelerometer Noise Floor Testing

Each accelerometer was tested in five different offsets: *x*-axis in 0 g, *y*-axis in 0 g, and *z*-axis in 0 g, 1 g, and −1 g. Along each of the axes, there will not be any static offset acceleration during free fall so all three were tested in 0-g environments. The *z*-axis was characterized more extensively for interest and because it tends to be the noisiest axis for accelerometers.

Analog Output

STMicroelectronics LIS344ALH

Table 2 describes the average noise power for the STMicroelectronics LIS344ALH accelerometer. The plot of spectral density is shown in Figure 16.

Table 2: STMicroelectronics LIS344ALH Noise Floor Results

Axis/Offset	**Noise Floor (µg/√Hz)**
x-axis 0 g	59.5501
y-axis 0 g	45.0739
z-axis 1 g	52.4238
z-axis 0 g	87.5051
z-axis −1 g	47.1743

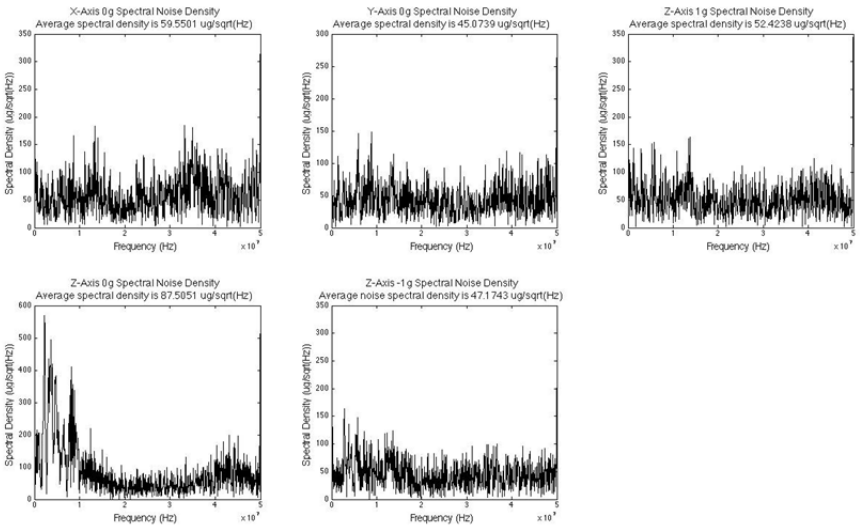

Figure 16: STMicroelectronics LIS344ALH Spectral Density Results

Kionix KXR94-2283

Table 3 describes the average noise power for the Kionix KXR94-2283 accelerometer. The plot of spectral density is shown in Figure 17.

Table 3: Kionix KXR94-2283 Noise Floor Results

Axis/Offset	**Noise Floor (μg/√Hz)**
x-axis 0 g	164.459
y-axis 0 g	186.826
z-axis 1 g	226.832
z-axis 0 g	341.799
z-axis −1 g	150.545

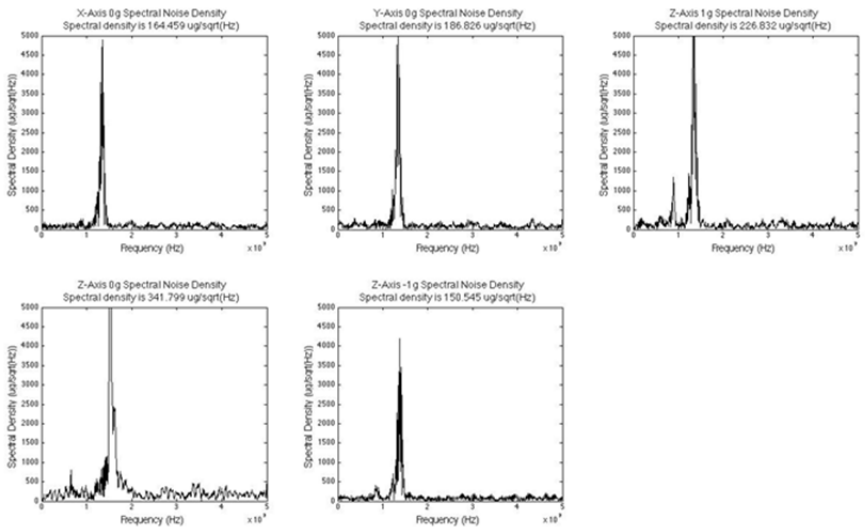

Figure 17: Kionix KXR94-2283 Spectral Density Results

Kionix KXTC9-2050

Table 4 describes the average noise power for the Kionix KXTC9-2050 accelerometer. The plot of spectral density is shown in Figure 18.

Table 4: Kionix KXTC9-2050 Noise Floor Results

Axis/Offset	**Noise Floor (µg/√Hz)**
x-axis 0 g	234.976
y-axis 0 g	227.512
z-axis 1 g	137.189
z-axis 0 g	521.492
z-axis −1 g	144.128

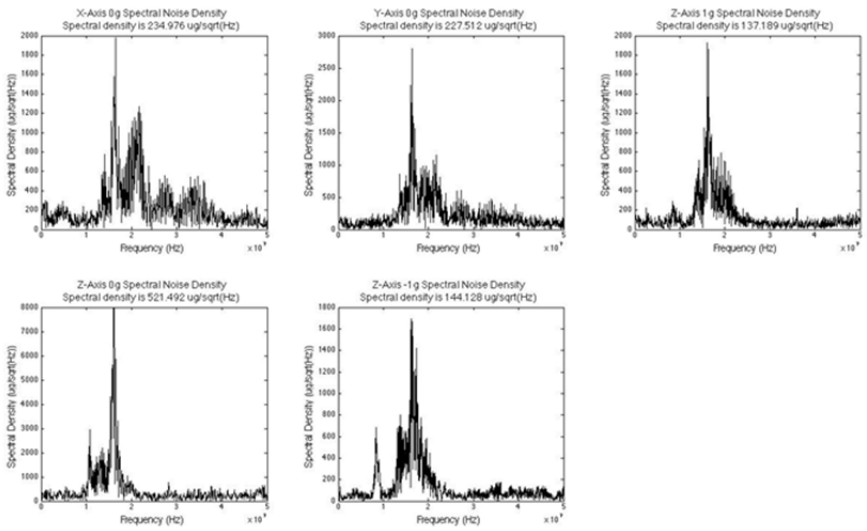

Figure 18: Kionix KXTC9-2050 Spectral Density Results

Digital Output

Analog Devices ADIS16210

Table 5 describes the average noise power for the Analog Devices ADIS16210 accelerometer. The plot of spectral density is shown in Figure 19.

Table 5: Analog Devices ADIS16210 Noise Floor Results

Axis/Offset	**Noise Floor (μg/√Hz)**
x-axis 0 g	18.7369
y-axis 0 g	18.4476
z-axis 1 g	19.2148
z-axis 0 g	18.961
z-axis −1 g	18.9831

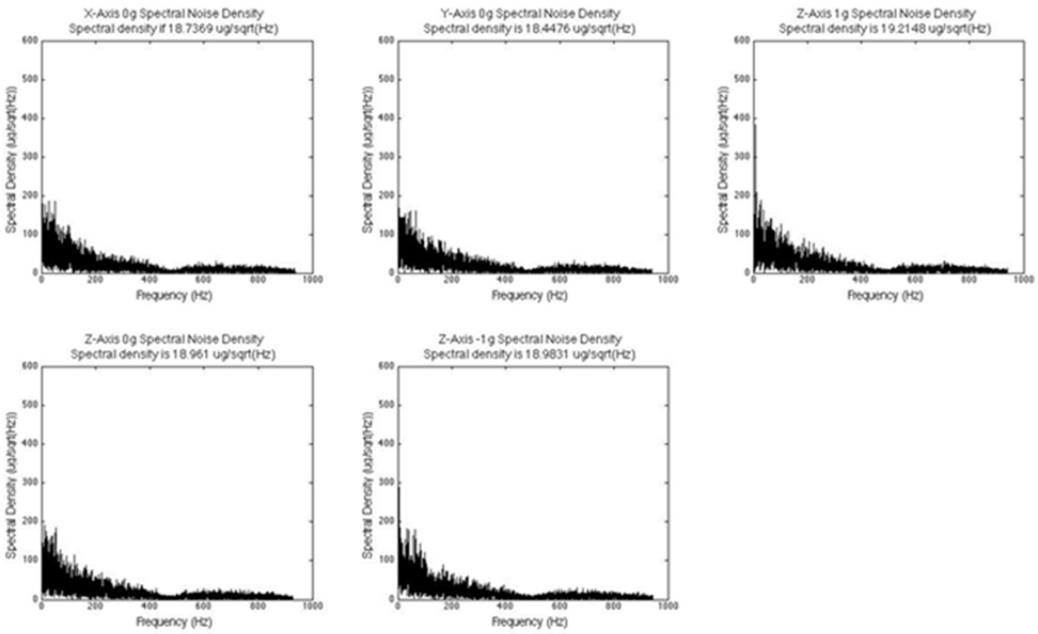

Figure 19: Analog Devices ADIS16210 Spectral Density Results

Analog Devices ADXL362

Table 6 describes the average noise power for the Analog Devices ADXL362 accelerometer. The accelerometer has several programmable settings and three different modes of operation with respect to noise. Presentation of the data aims to show side-by-side the comparison of these different modes. Plots of the spectral densities corresponding to the table can be found in Figure 20.

Table 6: Analog Devices ADXL362 Noise Floor Results

Ultra Low Noise		**Low Noise**		**Normal**	
Axis/Offset	**Noise Floor (µg/√Hz)**	**Axis/Offset**	**Noise Floor (µg/√Hz)**	**Axis/Offset**	**Noise Floor (µg/√Hz)**
x-axis 0 g	43.2406	*x*-axis 0 g	41.0223	*x*-axis 0 g	66.0354
y-axis 0 g	31.0902	*y*-axis 0 g	52.2678	*y*-axis 0 g	72.4382
z-axis 1 g	45.0405	*z*-axis 1 g	51.1062	*z*-axis 1 g	112.881
z-axis 0 g	57.5976	*z*-axis 0 g	61.5399	*z*-axis 0 g	106.224
z-axis −1 g	56.4206	*z*-axis −1 g	59.7572	*z*-axis −1 g	116.679

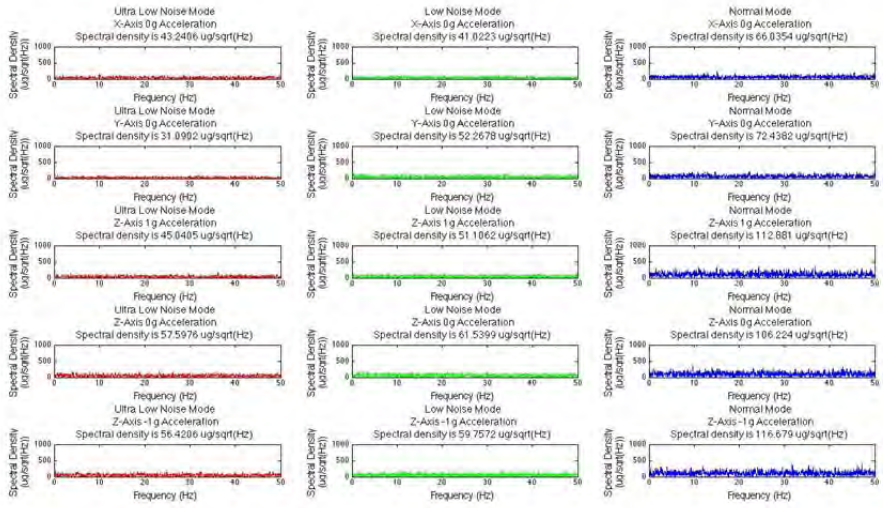

Figure 20: Analog Devices ADXL362 Spectral Density Results

DISCUSSION

Antenna Discussion

Designing the antenna for the system is a multivariable problem. For the purposes of initial evaluation, however, the beam pattern and gain were the driving factors in the analysis. While these facets will not uniquely determine the right choice, they are useful in making decisions on which antennas to discard. The idea of the analysis is to show that antennas on the sphere's surface are viable options for use in the telemetry system. Other features like input impedance for example, will require optimizing one of these antenna shapes that has promising gain and beam pattern.

Spherical Helix Overview

One-arm spherical helices outperformed two arm spherical helices in general; they dominated in beam shape and gain. However, the two-arm single-grated antenna performed on a comparable level to the one-arm helices. Furthermore, the single grating in general provided a gain boost in the dominant direction. Another advantage of adding the gratings was a reduction in input impedance. While this factor was not considered in the analysis because it can be optimized later, it is an added benefit of the grating.

Gain on the one-arm spherical helix with a single grating was similar to antennas from the previous design [1]. The −3 dB beam width of 90 degrees gave this antenna one of the broader beams compared to other simulated antennas.

Square Patch and Dipole

The square patch and dipole underperformed the other antennas. Between the two of them, the square patch has an overall better gain and consistent gain in all directions. If the sphere was not designed to sit in a specific orientation as it fell

and was instead allowed to tumble freely, then an antenna pattern like this would be necessary. However, for the purposes of the design, it can be assumed that the orientation of the sphere will not be changing wildly during the flight.

Slotted Ring

The slotted ring had great directional performance and maximum gain similar to that of the previous design [1]. Its pattern looks bumpy along its main transmitting direction and this affects its −3 dB beam width. However, by visual inspection it can be seen that it produces a fairly wide beam when adding up all of the bumps pointing in the same direction. This performance is a great mix of directional performance and ensures that small movements in the sphere will not cut transmission to the ground station.

Fractal

The fractal antenna created an interesting result. Its performance was one of the best out of the antennas although it was randomly generated. Its gain was comparable to the gain found by antennas in the previous falling sphere design [1]. While it did have a slightly smaller −3 dB beam width (76 degrees), if pointed correctly, it is still a practical solution.

Fractal antennas are not a new concept, but the idea of creating a random antenna and watching it perform well is exciting. Modern day computational abilities enable randomly searching for the best antenna for a given application. Created for this project was a MATLAB script to generate the antennas. Modifications to this script could automatically import the pattern directly into CADFEKO or another modeling program and simulate the results. Leveraging the abilities of fractal patterns for antenna design make this method an intelligent way of randomly searching for an ideal antenna.

Overview

The best antennas, one-arm single-grated spherical helix, Brownian tree fractal, and slotted ring, are compared in Figure 14. All three of them had comparable maximum gain to the ones presented in Soloff's thesis [1]. Also, they each had relatively wide beams. Ideally, the sphere will be set up so that as it falls the orientation of it is controlled; in other words, it will not be tumbling around. Fixing the orientation like this means at all times it is known where the antenna is located. The antenna gain can be very directional if this is the case, but to preserve flexibility in the system, it is important to have an antenna with reasonable power in a large beam width.

Antenna Creation by Metal Evaporation Deposition

The antenna creation successfully showed that the evaporation process works well for custom antenna patterns on a flexible substrate. Its main target would be the smaller scale antennas that need to wrap onto the surface (i.e., a different process would be required for the slotted ring antenna). Due to time limitations,

cutting the mask was not a perfect solution. Below is a blown up look at the resulting antenna, which has a significant imperfection.

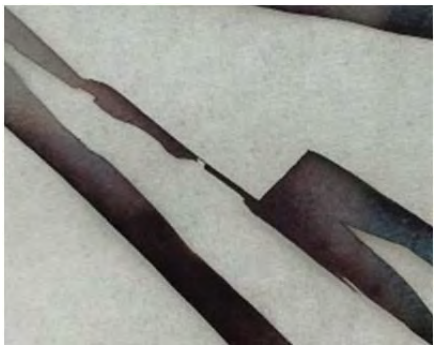

Figure 21: Imperfection in Fractal Antenna

Feature size of the antenna, however, is on a macroscopic level and there is a simple solution to enhancing the mask. 3D printing technology has sufficient resolution. For example, the Makerbot Replicator 2× has 0.011-mm positioning precision in the *xy* plane and this is a consumer-grade printer [7]. The smallest feature on the mask presented here is 2 mm and so 3D printing is a great solution for mask creation. Figure 22 is an example of what the dipole mask looks like when created in Google SketchUp.

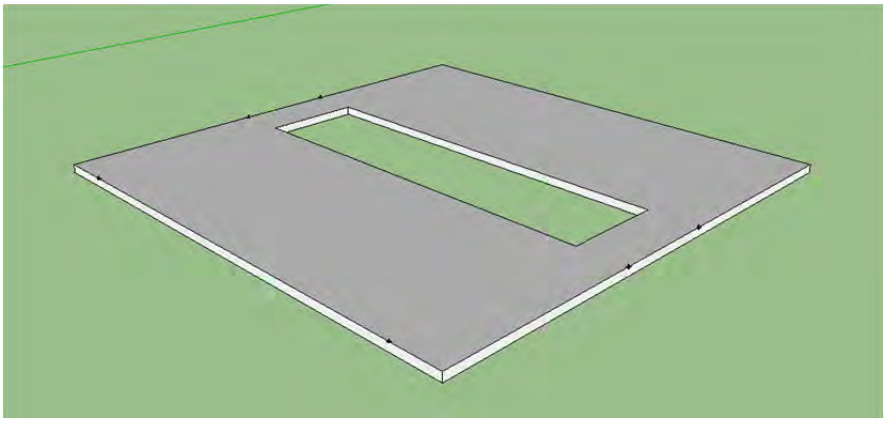

Figure 22: 3D Model of a Mask for a Wrapped Dipole Antenna

Accelerometer Discussion
Sampling Rate

It is important to take note of the span of frequencies in the spectral densities. Noise powers are quoted for conditions that attempt to match up with the data sheet as much as possible. For this reason, the two digital output accelerometers have spectral densities that span different frequency ranges. The importance of the frequency range arises when selecting up to what frequency the accelerometer must be able to measure. This has to do with the science goal of the mission and

other factors not discussed here. It is important to note that to measure larger frequency accelerations with the ADXL362 accelerometer, the noise floor will increase. However, it was found, although not presented here, that the noise floor for higher bandwidth is still well below the lower bandwidth specification.

1/f Noise

The first thing to notice about some of the presented spectral densities is the presence of 1/*f* noise [8]. Strong DC components were eliminated on most spectra by subtracting out the mean from the temporal data. However, the increase in noise with decrease in frequency can be easily seen in the results for the ADIS16210 in Figure 19.

Oscilloscope Noise

The experiment relied on the oscilloscope acquiring an accurate time trace of the accelerometer noise. To gauge the noise introduced by the oscilloscope, a time trace was recorded without anything attached to the probes. Spectral density was computed for that and it was found that this was introducing about 28 $\mu g/\sqrt{Hz}$ of noise into the system. This is a considerable amount of noise considering noise floors for many accelerometers are assessed as low as 45 $\mu g/\sqrt{Hz}$.

Kionix Accelerometers

The Kionix accelerometers did not perform up to their promised specifications. Model KXR94-2283 should have had a noise floor of at least 45 $\mu g/\sqrt{Hz}$ according to the datasheet, but the lowest it performed in test was 150 $\mu g/\sqrt{Hz}$. The KXTC9-2050 performed closer to its noise floor quote of 125 $\mu g/\sqrt{Hz}$, but still only reached 137 $\mu g/\sqrt{Hz}$ at its best.

Both accelerometers exhibit a large spike between 10 and 20 MHz as seen in Figures 17 and 18. This is puzzling because their low pass filters cut off in the kHz range or lower. Therefore, either both accelerometers simply do not provide the noise performance specified by the manufacturer or there is something awry with their evaluation boards. The peak in the spectrum is not likely introduced by the oscilloscope because it does not show up in the LIS344ALH plot in Figure 16.

LIS344ALH

This accelerometer performed well with respect to its advertised noise floor of 50 $\mu g/\sqrt{Hz}$. Along the *z*-axis for a 0-g offset, the noise did increase. During free fall, 0 g is the offset that *z*-axis will be experiencing so this noise increase is important. However, when considering the 28 $\mu g/\sqrt{Hz}$ introduced by the oscilloscope, this number becomes very close to the expected noise floor.

ADIS16210

Expected to have a noise floor of 190 $\mu g/\sqrt{Hz}$, this accelerometer outperformed its specification by an order of magnitude. 1/*f* noise is prevalent in

the spectral densities, but does not significantly increase the average density. Also, between different axes and offsets, the accelerometer performs very consistently in terms of noise.

ADXL362

Below in Table 7 is information from the data sheet for the noise floors. The accelerometer easily outdid its specification. From the results, it can also be seen that the different noise modes do indeed produce significant differences in noise power. This will be useful in tuning the accelerometer performance for the instrumentation system.

Table 7: ADXL362 Data Sheet Noise Floor Specifications

Ultra Low Noise		**Low Noise**		**Normal**	
Axis/Offset	**Noise Floor (µg/√Hz)**	**Axis/Offset**	**Noise Floor (µg/√Hz)**	**Axis/Offset**	**Noise Floor (µg/√Hz)**
x-axis 0 g	250	*x*-axis 0 g	400	*x*-axis 0 g	550
y-axis 0 g	250	*y*-axis 0 g	400	*y*-axis 0 g	550
z-axis 0 g	350	*z*-axis 0 g	550	*z*-axis 0 g	920

CONCLUSION

Antenna Simulation

The antenna simulations established a good foundation for useful antenna configurations that can be attached to the surface of the sphere. Moving forward with the simulation, the one-arm single-grated spherical helix, slotted ring, and Brownian fractal should be optimized for other design points like input impedance. Also, several more random fractals should be created and tested. Automation of testing the fractal antenna after it is randomly generated opens up possibilities for running a large fractal antenna optimization loop.

Experimental verification of these antennas should be performed. This would introduce practical limitations and the associated effects on performance.

Antenna Fabrication

The method for fabricating the flexible antenna was successful. Next steps are to run tests on the antenna and make sure it functions as in simulation when mounted to a sphere. Also, the antenna should be put through stress tests that simulate deflating and inflating the balloon to make sure the antenna is robust enough for deployment. This would involve depositing the antenna onto a substrate, like Mylar, that will more likely be used to construct the sphere as opposed to flexible plastic. Finally, the 3D printed mask needs to be tested to ensure it eliminates macroscopic defects in the fabrication process.

A concern with the process used is the limited area onto which the metal can be deposited. Other methods need to be designed and tested for creating larger antennas like the slotted ring.

Accelerometer Testing

Results of the accelerometer tests are helpful in evaluating the sensors based on noise floors. The top analog performer was the STMicroelectronics LIS344ALH and the top digital performer was the Analog Devices ADIS16210. However, the Analog Devices ADXL362 provides flexibility in how its accelerometer runs with sampling rate and noise levels. Concerns were raised that the digital interface may increase the noise floor over analog output devices. However, this did not appear to be the case. Since the digital interface built in does not pose a problem to the noise floor, this is the recommended type to use in the instrumentation system. Without the need to design custom hardware to interface an analog accelerometer to an FPGA or other digital processing unit, the design is simplified.

Gathering analog accelerometer information needs to be improved. The oscilloscope could record a limited number of samples at a time. Decreasing the sampling rate on the oscilloscope produced much noisier results. This trade-off meant that, for the sake of accuracy, the time trace was recorded for maximum accuracy at a high sampling rate. Spectral densities then spanned beyond the frequencies of interest (acceleration frequencies of interest lie below 1 kHz) and had reduced low frequency resolution. While this should not have posed a problem, the spectral densities for both Kionix accelerometers exposed a large peak between 10 and 20 MHz. Consequently, these accelerometers had average densities much higher than the ones provided by their data sheets. Improving upon the data acquisition system would greatly improve the confidence in the analog output accelerometer noise floor results.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would first like to thank Professor Carlo Pantano for taking an interest in my project and connecting me with Professor Raúl Martin-Palma. Professor Martin-Palma was very helpful in setting up the antenna pattern on a flexible substrate for use in a thermal evaporation chamber. He also carried out the deposition for me so that I could verify the procedure worked. Also, I want to thank Dave Fecko and Josh Stapleton for letting me set up in a vibration free room in the Millennium Science Complex to carry out the accelerometer tests. A special thanks goes out to Professor Sven Bilén, Bob Capuro, Professor Charles Croskey, and Professor Gerald Lehmacher (Clemson) for mentoring me during the project. Additionally, I want to thank the National Science Foundation and program host Pennsylvania State University for providing me with the opportunity to participate in this program. Furthermore, I want to thank my roommates Zach Morgan, Bill Kim,

and Mason Sutorius for their project input and support this summer. Last, but certainly not least, I want to thank my family for their continued support and enthusiasm about my academic development and aspirations.

REFERENCES

[1] Soloff, Jason A. "An Active Instrumented Inflatable Sphere for *In-Situ* Measurement of Temperature and Density in the Mesosphere." M.S. Thesis. The Pennsylvania State University, 2001. Print.

[2] Rashed, Fatemeh M. "Design and Development of a PCM Telemetry System for an Active Inflatable Falling Sphere for Meteorological Measurements." M.S. Thesis. The Pennsylvania State University, 2000. Print.

[3] Ding, Yong, Jing-Hui Qiu, and Wen-Yi Qin. "A new spherical helical antenna." Antennas and Propagation Society International Symposium, 2005 IEEE. Vol. 2. IEEE, 2005.

[4] Best, Steven R. "The radiation properties of electrically small folded spherical helix antennas." Antennas and Propagation, IEEE Transactions on 52.4 (2004): 953-960.

[5] Pfeiffer, Carl, et al. "A printed spherical helix antenna." Antenna Technology and Applied Electromagnetics & the American Electromagnetics Conference (ANTEM-AMEREM), 2010 14th International Symposium on. IEEE, 2010

[6] Cohen, Nathan. "Fractal antenna applications in wireless telecommunications." Electronics Industries Forum of New England, 1997. Professional Program Proceedings. IEEE, 1997.

[7] "MakerBot Replicator™ 2X Desktop 3D Printer." Makerbot® Industries, n.d. Web. 21 July 2013. <http://store.makerbot.com/replicator2x.html>.

[8] Keshner, Marvin S. "1/f noise." Proceedings of the IEEE 70.3 (1982): 212-218.

[9] LIS344ALH Data Sheet, Revision 3, STMicroelectronics, Geneva, Switzerland.

[10] KXR94-2283 Data Sheet, Revision 1, Kionix, Ithaca, NY.

[11] KXTC9 Series Data Sheet, Kionix, Ithaca, NY.

[12] ADXL362 Data Sheet, Analog Devices, Norwood, MA.

[13] ADIS16210 Data Sheet, Analog Devices, Norwood, MA.

DESIGN OF A MAGNETIC TORQUER SYSTEM FOR CUBESATS

Cynthia Rojas,* Sven G. Bilén,[#] and Timothy Brubaker[+]

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, New Mexico 88003

ABSTRACT

CubeSats require attitude control systems to maximize mission capability, yet are highly constrained in terms of mass, power, and size. Hence, simple systems are preferred for attitude control, such as the use of magnetic torquers. The attitude control for the OSIRIS-3U CubeSat will use printed circuit board–based magnetic torque coils, as opposed to torque rods. Implementing the torque coils on a PCB allows their placement behind the solar panels and will save space within the CubeSat and mass. The PCB will have a printed coil of a number of loops that will carry current through it to create a magnetic dipole. The dipole created by the magnetic torquer will interact with the Earth's magnetic field as the CubeSat flies in low Earth orbit. This interaction will cause the magnetic field of the coil to seek to align itself with the Earth's magnetic field. Judicious use of this torquing will allow the CubeSat to orient itself in space. Multiple magnetic torquers will achieve the desired orientation while in orbit. Presented herein is the circuit design for the magnetic torquer.

INTRODUCTION

"CubeSats in low Earth orbit have enabled dozens of universities to develop and place in orbit student-led, student-designed, student-built and student-operated satellites investigating all manner of scientifically exciting phenomena" [1]. CubeSats are used for space exploration, performing *in situ* and remote measurements, and other various missions. The challenge in building these satellites is to meet specific mission requirements on a system that has limited

[#] Faculty Mentor
[+] Graduate Mentor

mass, power, and volume, yet requires attitude control, command and data handling, and other subsystems to ensure functionality.

CubeSat

A CubeSat is a satellite on a small scale that is constrained in mass, power, and size. The satellite is measured in units (or "U"), where 1U is a 10×10×10 cm cube. Figure 1 shows a comparison of typical CubeSat sizes. For every 1U, a max mass of 1.33 kg is allowed. The mass includes the satellite in its entirety; therefore, mass has to be meticulously allocated accordingly to each subsystem. Power is limited for the most part by what can be collected using solar panels and stored in batteries. These constraints have led to many designs that conserve space and weight as well as power consumption. Redesigning the magnetic torquer for the OSIRIS-3U and implementing it on a PCB will minimize volume and mass occupied by the magnetic torque by mounting them behind the solar panels.

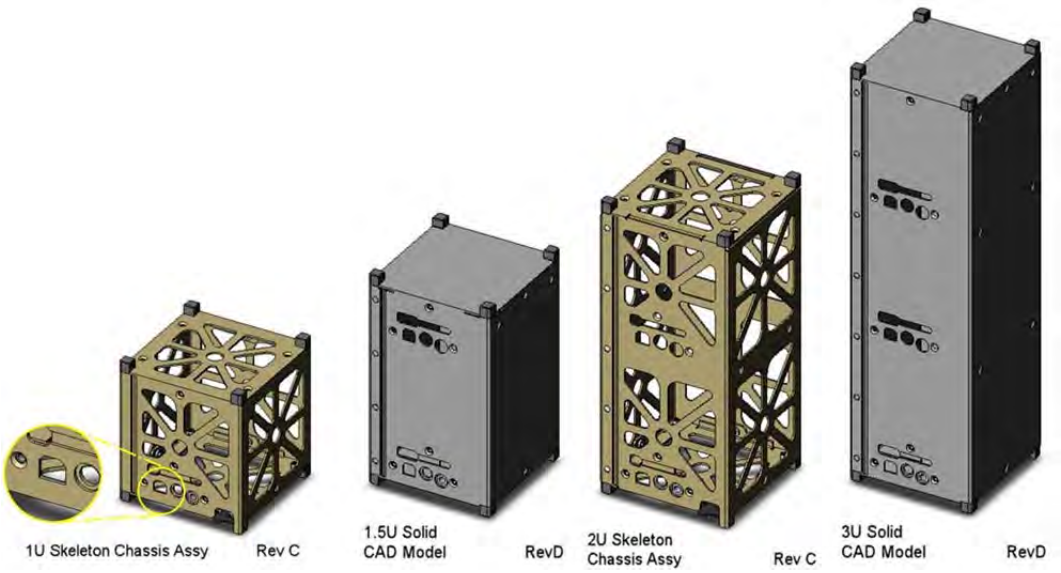

Figure 1: CubeSat structures (from [2])

OSIRIS-3U Mission

There is still much unknown about space weather phenomena, which is important to understand as it affects satellites and communication systems. The OSIRIS-3U mission is to provide *in situ* and remote sensing measurements of the heated ionosphere. The data gathered will be compared with ground-based data and elucidate irregularities that occur in space weather [3]. The dimensions of this CubeSat will be 10×10×30 cm with a mass limit of 4 kg.

Magnetic Torquer

A magnetic torquer functions in the following manner: there is a coil through which a current is carried, which creates a magnetic dipole. As the magnetic

dipole interacts with the Earth's magnetic field, it seeks to align itself with the Earth's magnetic field, thereby creating a torque. When the magnetic torquer is energized at the correct time, this resulting torque orients the satellite in space.

The allotted area for the design is 3800 mil^2 with a keep-out area in the center of 884 mil^2. The coil is traced around the keep-out area and has two layers with 15 turns each. Using a MATLAB program (developed by Kory Knabenshue and Kodie Altvater in the Student Space Programs Lab), we calculated the optimal number of turns and trace width for the magnetic torquer, which were used as a reference for the PCB layout of the coil. As shown in Figure 2, a simple concept drawing of the magnetic torquer shows that it consists of a current source, H-bridge, and a coil. The current source is supplied to the coil to create a magnetic dipole and the H-bridge allows bidirectional flow of current through the coil.

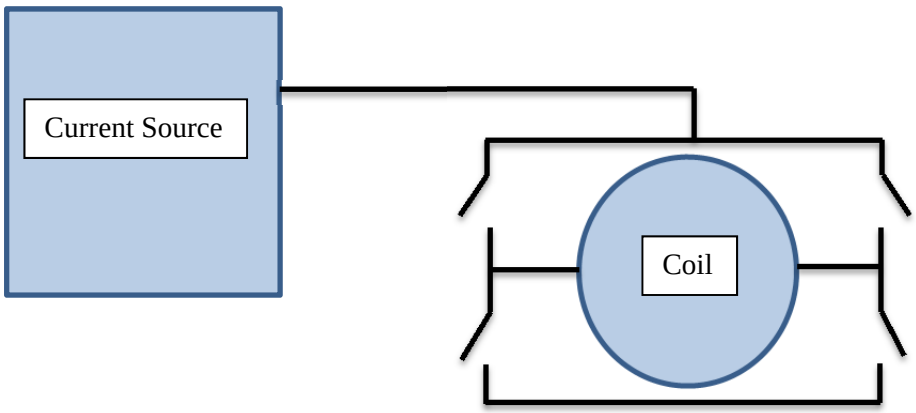

Figure 2: Magnetic Torquer Concept

When choosing components there are a few required specifications that the magnetic torquer will need to make this a practical design. The power source for the torquer will be from a battery with an expected voltage supply of in the range of 6–12.6 V. The conversion frequency range will need to be within 300 kHz–1 MHz to avoid interfering with other components in the CubeSat. The expected regulated coil current is 2–3 A. The following discussion refers to the schematic for the magnetic torquer in Figure 3.

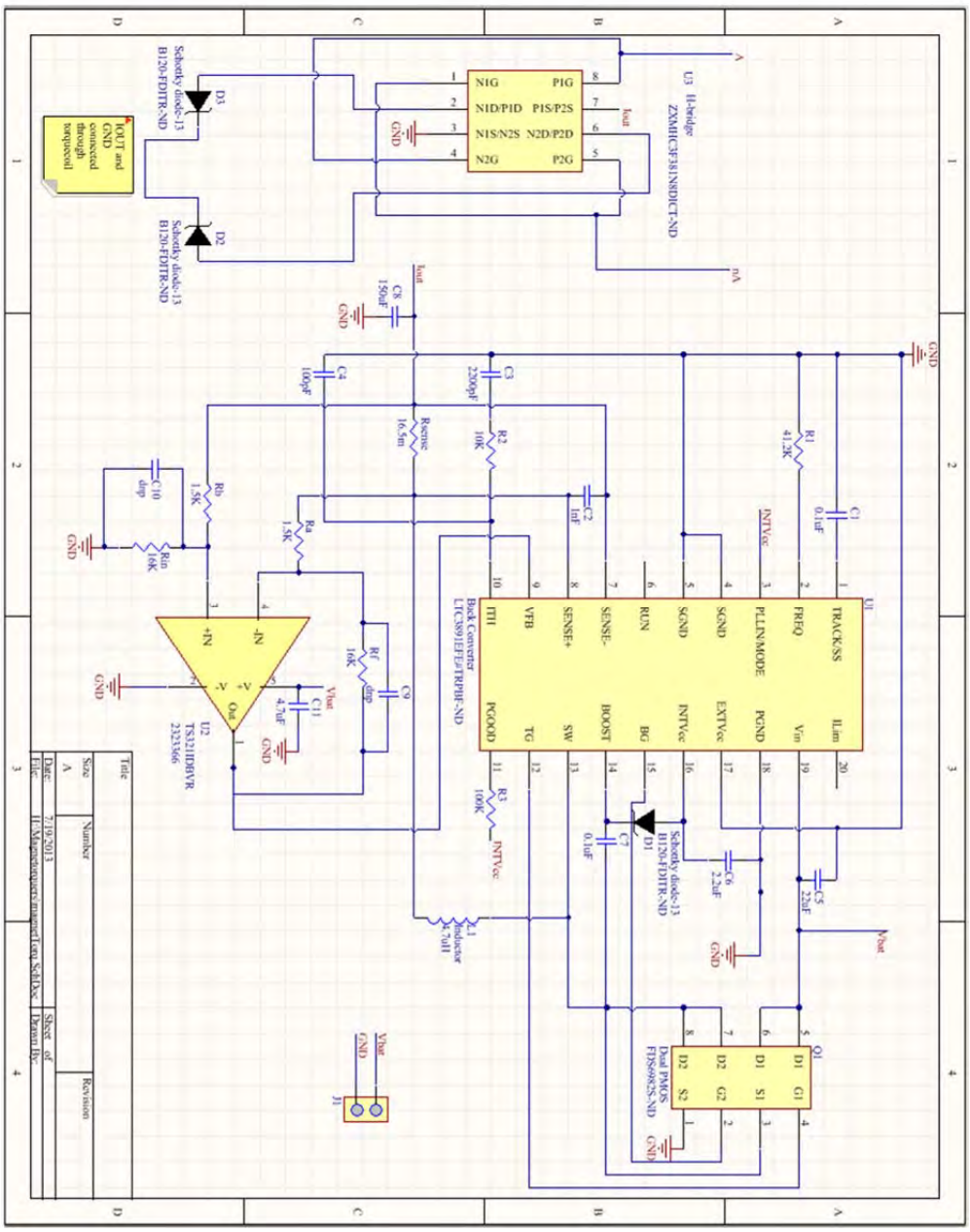

Figure 3: Magnetic Torquer Schematic

MAGNETIC TORQUER DESIGN

The main component of the current source block is the LTC3891 buck converter. It takes a voltage input and outputs a regulated current to the coil. The LTC3891 conversion frequency is selected via the FREQ pin. As the frequency range is 50 kHz–900 kHz, this provides the option of synchronizing to an external clock, whereas tying this pin to ground selects 350 kHz for this application. A

low frequency is preferable to maximize efficiency and minimize losses in the MOSFET switches. There are three modes of operation to select from, which include continuous, burst, and pulse-skipping; these modes are selected by the PLLIN/MODE pin. This pin is connected to INTVcc to select continuous mode, which we desire to be able to supply continuous current to the coil until the desired satellite orientation is reached. To regulate the current a low value resistor is connected between SENSE− and SENSE+. The value of this resistor in part depends on the connection on the Ilim. This pin sets the current comparator sense threshold to one of three options 30 mV, 50 mV, or 75 mV. The value of the low value resistor R_{sense} is determined using [4]

$$R_{\text{sense}} = \frac{V_{\text{sense(max)}}}{I_{\text{max}} + \frac{\Delta I_L}{2}} \ . \tag{1}$$

The values for $V_{\text{sense(max)}}$ have a high and low value for the nominal voltages, which are provided on the data sheet. I_{max} is the maximum output current, in our case it is 3 A. Equation (1) also adds half of the peak-to-peak ripple current. To regulate the current going through R_{sense} such that we get an output current of 3 A, a difference op amp configuration is also connected across R_{sense}. The op amp is set up to have an output of 0.8 V; this is sent to the voltage feedback (Vfb) pin. The buck converter then takes the voltage at the Vfb pin and compares it to a reference voltage of 0.8 V. If there is an offset, it will cause a voltage change in the Ith pin, which will change the inductor current and will regulate it to match load current. This method is essentially regulating the voltage across R_{sense} to maintain the current through it. Several things to keep in mind when laying this out on a PCB is to keep the filter capacitors and the schottky diode close the LTC3891, as these are circuit protection for the sudden changes in voltage caused by the switches being used.

The regulated current is then directed through the coil using an H-bridge, which allows the current to switch direction. The bidirectional flow of current through the coil can also provide redundancy in that, if one of the torquers fails, another torquer can be used in its place by switching the current through the coil changing the polarity of the of the magnetic field of the coil. The H-bridge is comprised of a combination of p-channel and n-channel MOSFETS. The high side MOSFETs are p-channel FETs and the n-channel FETs are on the low side. The p-channel FETs on the high side make it easier to control switching from the gates [4]. To open a PMOS we can just apply a voltage at the gate and closing the switch would just require the gate to be connected to ground. For an NMOS switch as long as the gate-source voltage is equal or less than zero it will remain open but the problem is in closing it, we would require a higher voltage at the gate to ensure it closes [5].Two flyback diodes are used to protect the MOSFETs from voltage spikes caused by switching on the coil [6].

DESIGN EVALUATION

The design was developed using Altium for the schematic capture and the PCB layout. Several component footprints needed to be created and sized to what was listed in the data sheets and add the component characteristics. Once the schematic was captured, it was reviewed by others for any corrections or concerns regarding the design, after which the design went into lay out. The bottom layer is for the solar panels so no changes are made to that layer, the two layers after that are for the coil and the top two layers are for the magnetic torquer circuit.

EVALUATION RESULTS

This circuit was set up originally to have a V_{sense} of 70 mV so the resistors for the op amp were chosen such that there would be a difference of 70 mV between the two inputs and get an output 0.8 V. This had to be changed later because the value of R_{sense} was unavailable. To confirm the op amp configuration, it was breadboarded and two known voltages were applied at each input resistor and 0.8-V output was expected. Initially this was not the case, which was due to an issue with the breadboard pin on the ground rail; once the pin was changed the proper output was confirmed and the op amp was confirmed to be functioning properly. One issue remained, however, which was an unexpected resistance connected between ground and pin 3 of the op amp.

DISCUSSION

This design was originally supposed to be a four-layer PCB; however, after laying out the coil using a PCB design program, it was determined that the space remaining around the coil was very limited. The design was changed to a six-layer PCB, which gave more room for the torquer's circuit and any additional vias that may be needed (Figure 4). The circuit will be placed on the top layer, whereas the two coils will be on the layers labeled MAG1 and MAG2. Figure 5 shows one layer of the coil with 15 turns a second layer of the coil that is connected through a series of vias at the center of the coil.

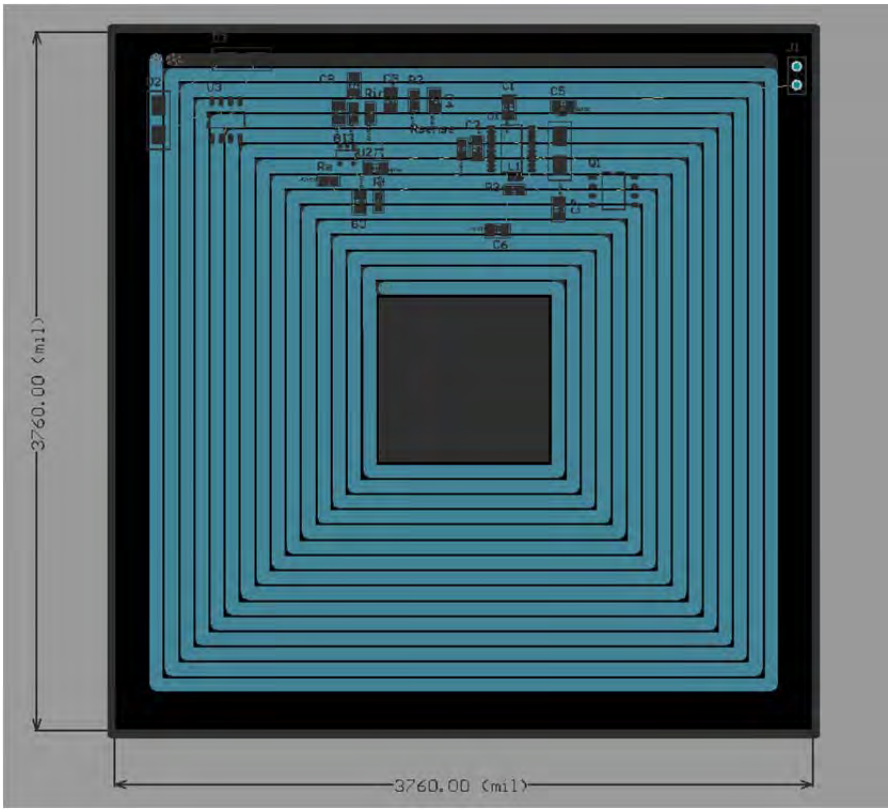

Figure 4: PCB Coil and Circuit Components

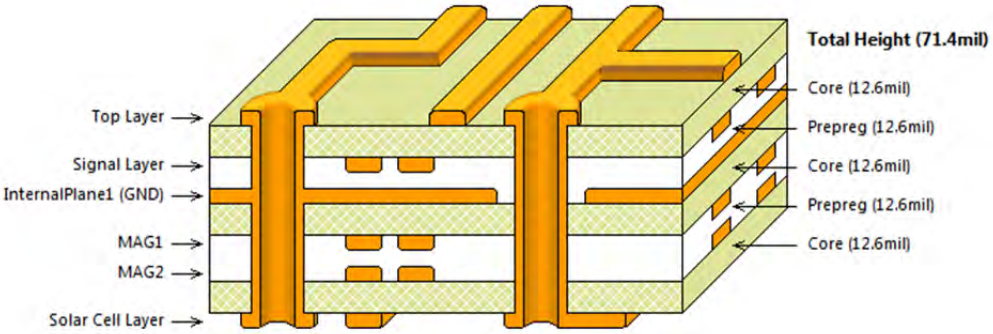

Figure 5: PCB Layer Set Up

CONCLUSION

The magnetic torquer is a better design with a six-layer board than a four-layered board. On a four-layer board space would be limited for any additional vias or adjustments to the circuit. In addition, six layers allow for a better layout for the circuit, specifically the components that need to be placed near the LTC3891 or the op amp for filtering.

As far as the circuit design the op amp configuration will need to be looked at to see what is causing a resistance in parallel with pin 3 of the op amp and ground.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would like to thank Dr. Sven Bilén for the opportunity to work on this project and my experience at Penn State. I would also like to thank Timothy Brubaker for all his help with all of my research, helping me understand the concepts, and getting me in contact with those who also worked on this project to answer any questions I may have had. Thank you to Tyler Boehmer, Ravender Virk, Kory Knabenshue, and Sean McFarlane for all your help and input.

REFERENCES

[1] Staehle, Robert (February 16, 2013). Interplanetary CubeSats: Opening the Solar System to a Broad Community at a Lower Cost. Retrieved from http://www.nasa.gov/directorates/spacetech/niac/staehle_interplanetary_cubesats.html
[2] http://www.triy.org/ENG/CubeSat_Intro.htm
[3] http://www.altium.com/resources/customer_success/Penn_State_EN.pdf
[4] http://cds.linear.com/docs/en/datasheet/3891fa.pdf
[5] http://www.modularcircuits.com/blog/articles/old-h-bridge-secrets/part-1/
[6] http://www.modularcircuits.com/blog/articles/old-h-bridge-secrets/part-1/
[7] http://en.wikipedia.org/wiki/Flyback_diode

COMPARISONS OF ADVANCED ADAPTIVE SIGNAL PROCESSING METHODS FOR REMOVING MATERNAL INTERFERENCE NOISE FROM FETAL ELECTRO-CARDIOGRAMS

Ryan M. Collins,* W. Kenneth Jenkins,[#] and A. David Salvia[#]

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

*Undergraduate Student of
Department of Electrical Engineering
Bucknell University
Lewisburg, PA 17837

ABSTRACT

In the medical field electrocardiograms (ECGs) are a useful tool in observing the overall health of a patient. Likewise, fetal ECGs can be a useful tool in observing the development of a fetus during pregnancy. Currently, techniques to remove maternal interference are not developed enough to allow fetal ECGs to be a practical tool in clinical environments. The goal of this project is to use a variety of adaptive filtering techniques to remove maternal interference in fetal ECGs and to determine which techniques are the most effective under varying circumstances. The techniques being explored are time domain Finite Impulse Response (FIR) adaptive filtering, transform domain FIR adaptive filtering, and multichannel FIR adaptive filtering.

INTRODUCTION

Fetal Electrocardiograms

Fetal ECGs can be measured using two methods. The first method is achieved by attaching an electrode directly to the fetus during labor after the mother's water has broken. The second method is a non-invasive approach in which several electrodes are placed on the abdomen of the pregnant patient. The non-invasive approach has the distinct advantage that it can be utilized through the duration of the pregnancy, not just during labor. Unfortunately, when measuring the fetal signal on the abdomen, the fetal heartbeat has significantly smaller amplitude than the maternal heartbeat. In addition to the maternal heartbeat, the abdominal signal

[#] Faculty Mentor

also contains noise from the maternal heartbeat passing through the thorax to the abdomen.

The channel that the maternal heartbeat passes through, between the thorax and the abdomen, effectively "filters" the maternal signal. This "filter" is unknown and therefore may be modeled as an FIR, Infinite Impulse Response (IIR), or nonlinear system. In addition to the distortion of the maternal heartbeat, other noise is present in the abdominal signals [1,2].

PHYSIONET WEBSITE

The PhysioNet website < http://www.physionet.org/ > offers free web access to large collections of recorded physiologic signals and related open-source software [3]. PhysioNet is one of the four NIH-supported initiatives that was highlighted in a recent White House communication about accelerating the pace of discovery through the use of large realistic biomedical data bases. An important link on this website is PhysioBank, which is a large and growing archive of well-characterized digital recordings of physiologic signals and related data that has been made available for use by the biomedical research community. PhysioBank includes databases of multi-parameter cardiopulmonary, neural, and other biomedical signals from healthy subjects and patients with a variety of conditions with major public health implications, including sudden cardiac death, congestive heart failure, epilepsy, gait disorders, sleep apnea, and aging.

In particular, PhysioBank includes two collections of ECG signals that are relevant to the research conducted in this project. One collection, called "Abdominal and Direct Fetal ECG Database," contains five-minute multichannel fetal ECG recordings with cardiologist-verified annotations of all fetal heartbeats from five women in labor, from the Medical University of Silesia, Poland. The second collection, called "Non-Invasive Fetal Electrocardiogram Database," contains 55 recordings of maternal and maternal+fetal ECGs recorded over a 20-week period from a single pregnant subject. Selected signals from both of these collections are used in the following work. More details are given in the following section under EXPERIMENTAL DESCRIPTION.

Previous ECG fetal research that was conducted under the 2012 REU Program at Penn State [4] performed experiments strictly on the ECG recordings available in the PhysioBank. However, in many of these cases the fetal heartbeat components were not very visible, and it was difficult to determine how well the signal processing methods were working to reduce the maternal component while simultaneously enhancing the visibility of the fetal component. In this current REU project, an approach was used whereby the direct fetal signals that are available in the Abdominal and Direct Fetal ECG Database were added to the abdominal signals in order to produce "synthetic signals" for which the strength of the fetal component could be experimentally controlled. These synthetic signals enable the various signal processing algorithms to be compared in an effort to determine which algorithms are the most effective under various conditions. It must be emphasized here that the experiments that used these

synthetic signals do not guarantee how well the algorithms being investigated produce useful results for clinical practice, but rather the synthetic signals enabled basic research to be conducted to investigate which adaptive methods should be further evaluated in future studies. More details about these synthetically constructed signals are described in detail in the following sections.

EXPERIMENT DESCRIPTION

Modeling of the Fetal Signal

The abdominal signal $d[n]$ is modeled as a superposition of two signals, the fetal signal $f[n]$ and the maternal noise signal $s_0[n]$. The noise signal $s_0[n]$ represents all of the maternal noise, including the maternal heartbeat after it has passed through the unknown system between the maternal heart and the fetus. Thus to achieve a clean fetal signal, $s_0[n]$ must be accurately modeled and then subtracted from the abdominal signal. To accomplish this ECG electrodes are placed on the thorax of the patient to record a clean maternal heartbeat $x[n]$. Then $x[n]$ is used as a reference signal for $s_0[n]$ before passing through the system between the thorax and abdomen.

$$d[n] = f[n] + s_0[n] \tag{1}$$

The non-invasive ECG database consists of 55 recordings from a single patient taken throughout the duration of her pregnancy. Each recording contains two thoracic signals and either three or four abdominal signals. The signals sampled at 1 kHz with a resolution of 16 bits are limited to a bandwidth between 0.01 Hz and 100 Hz.

The direct fetal database contains five records from separate patients during labor. Patients were between 38 and 41 weeks of gestation. Each recording contains a direct signal from the fetal head as well as four signals measured from the maternal abdomen. These signals are also sampled at 1 kHz with a resolution of 16 bits, although are limited to a bandwidth between 1 Hz and 150 Hz.

Kam [5] suggests that an IIR model provides a more accurate model of the system. However, the low frequency nature of ECG signals (on the order of 1 Hz – 2 Hz) suggests that the transfer function for the system is of a low enough order that a low order FIR adaptive filter provides similar results as the IIR model. Therefore for this paper, we chose to focus solely on an FIR implementation of all filters provided in this paper.

The LMS Algorithm

The adaptive filtering algorithm used for all of the adaptive filters is the Least-Mean-Square (LMS) algorithm [6,7,8]. The LMS algorithm is an adaptive algorithm that approximates the steepest descent algorithm. This approximation provides an accurate adaptive algorithm that is simple to implement and not computationally intensive. There are several important factors to be accounted for

when using the LMS algorithm. The first is the algorithm step size parameter μ. Different values of μ will provide different convergence rates for the system, but may also introduce noise. Care must also be taken not to have a value of μ too large or too small; large values of μ will cause the system to become unstable, while small values of μ will prevent the system from converging. The next parameter is the length of the filter (for FIR implementations), and the last parameter is the amount of time delay, if any, in the input signal. As stated above, lower order FIR filters are adequate when using adaptive methods with ECG signals. Therefore for this paper, most tests were performed with filters of order 64.

$$y[n] = \mathbf{w}^T[n]\mathbf{x}[n] \tag{2}$$

$$e[n] = d[n] - y[n] \tag{3}$$

$$\mathbf{w}^T[n+1] = \mathbf{w}^T[n] + 2\mu e[n]\mathbf{x}[n] \tag{4}$$

A variation of the LMS algorithm was also used, called the Power Normalized LMS algorithm. This variation allows large variations in the input signal without causing the system to become unstable. Equation (5) shows the power normalized filter update. The term α is a constant, and for all cases in this paper, both time domain and transform domain, a value of α = 0.001 was used. As seen in Equation (6), $\sigma^2[n]$ is a value based on past values of the input. This memory prevents the filter update from drastically changing between values.

$$\mathbf{w}^T[n+1] = \mathbf{w}^T[n] + \frac{\mu e[n]}{\alpha + \sigma^2[n]}\mathbf{x}[n] \tag{5}$$

$$\sigma^2[n] = \frac{1}{M}\sum_{i=1}^{M} x^2[n-i] \tag{6}$$

Time Domain Adaptive Filtering

The first time domain filter configuration used is an adaptive noise cancelling configuration shown in Figure 1. The noise cancelling configuration utilizes two signals, the noisy abdominal signal and a reference thoracic signal. The thoracic signal provides a clean representation of the maternal heartbeat before passing through the thoracic channel. This signal is fed into the adaptive filter, which then approximates the thoracic channel and subtracts the result from the abdominal signal.

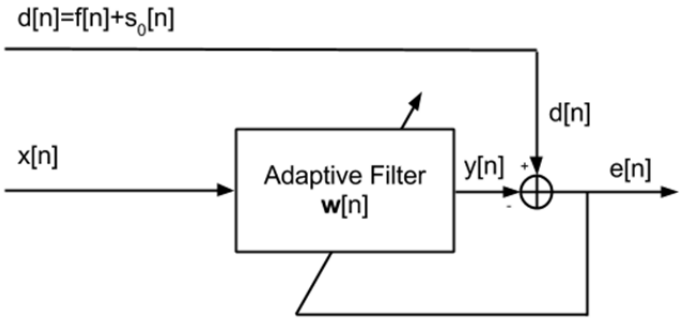

Figure 1: Noise cancelling configuration

The second time domain filter configuration used is an adaptive linear predictive or line enhancement configuration, shown in Figure 2. This system is designed to remove uncorrelated noise from a periodic signal through the use of a time delay. Thus this configuration utilizes only the abdominal signal as an input. The abdominal signal is fed into the adaptive filter, and the output of the filter is subtracted from a time-delayed abdominal signal.

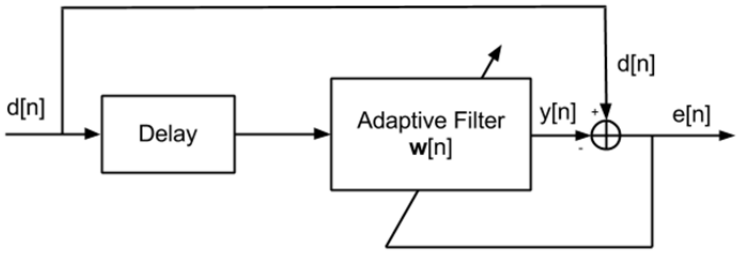

Figure 2: Linear predictive configuration

Transform Domain Adaptive Filtering

The transform domain adaptive filter is used to apply the LMS algorithm to separate frequency channels of the input signal rather than the entire input [7,9,10]. Although several types of mappings can be used to transform the input signal, the one used in this paper is the Discrete Fourier Transform (DFT) that is implemented with the Fast Fourier Transform (FFT) algorithm. The input reference signal is windowed according to the length of the filter and then subsequently mapped to the discrete frequency (DFT) domain. Each frequency channel is subsequently adaptively filtered and summed as illustrated in Figure 3. The summation of the filtered channels provides the output of the system, $y[n]$. This output is then subtracted from the noisy abdominal signal. Once again, the filter was configured in either a linear predictive or noise cancelling configuration.

$$\mathbf{W}[n+1] = \mathbf{W}[n] + 2\mu e[n]\mathbf{X}[n] \tag{7}$$

$$y[n] = \mathbf{X}[n] \bullet \mathbf{W}^H[n] \tag{8}$$

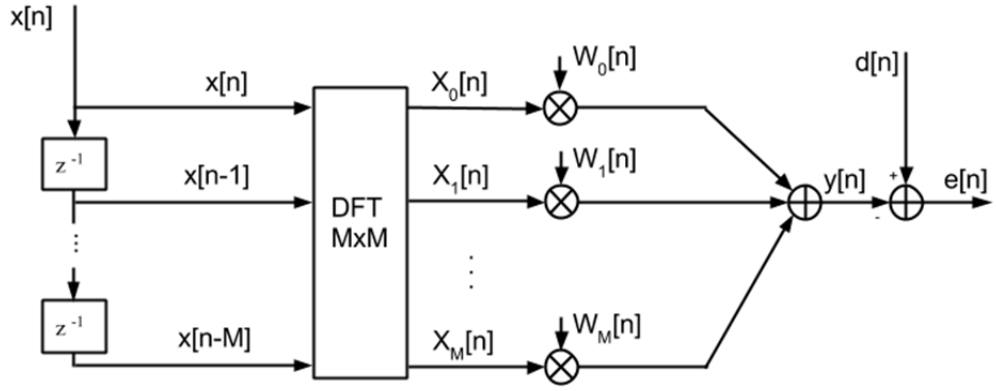

Figure 3: Block diagram for transform domain adaptive filter.

A power normalized variation of the transform domain filter was also applied to the signals. Equation 9 shows that the transform domain power normalization is similar to the time domain power normalization in Equation 5. Rather than a single $\sigma[n]$ value, though, multiple $\sigma_k[n]$ values correspond to the different frequency channels of the transformed signals. $\tilde{\sigma}$[n] is a matrix containing the corresponding $\sigma_k[n]$ values on the matrix diagonal. $\tilde{I}$ is an identity matrix, so that α may be applied to control possible numerical ill-conditioning that may occur each channel of the filter.

$$\mathbf{W}[n+1] = \mathbf{W}[n] + 2\mu e[n]\left(\alpha\tilde{I} + \tilde{\sigma}^2[n]\right)^{-1}\mathbf{X}[n] \tag{9}$$

$$\sigma_k[n] = \mathbf{X}_k[n] \bullet \mathbf{X}_k[n] \tag{10}$$

Multiple Input Transform Domain Filtering

In addition to the transform domain filter, a multiple input system was also built. In the multi-input system, the each signal is filtered in a separate transform domain adaptive filter. The respective errors of each system $e_i[n]$ are then averaged together into $e_{avg}[n]$. This average error is then used as the feedback for each adaptive filter. The averaging was a simple arithmetic mean as seen in Equation (9) where l is the number of inputs.

$$e[n] = \frac{1}{l}\sum_{i}^{l} e_i[n] \tag{11}$$

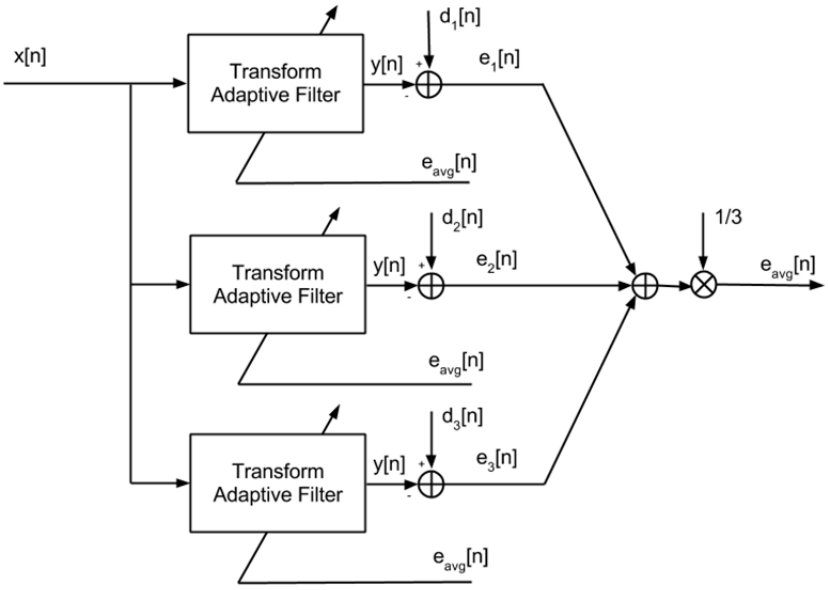

Figure 4: Multi-input adaptive noise cancelling filter.

Error Smoothing

Error smoothing was also applied to the results to reduce any high frequency noise either in the original signal or introduced from by the filter. To accomplish the smoothing, the error $e[n]$ was windowed and all values in the window were averaged into the smoothed error $\varepsilon[n]$.

$$\varepsilon[n] = \frac{1}{L}\sum_{i=0}^{L-1} e[n-i] \tag{12}$$

EXPERIMENTAL RESULTS

The data used to test the filters was generated by adding a scaled fetal signal from the direct fetal database to an abdominal signal found in the indirect database. This allowed for the control of the amplitude of the fetal signal. It was found that the filters performed better when the number of iterations was increased. Figure 5 shows the difference between filtering with N = 2000 and N = 16000, where N is the number of iterations. The separate trials were performed with different values for μ so that both systems remained stable and converged. It can be seen that the trial with more iterations reduced maternal noise more adequately when compared to the N = 2000 trial. All further filters were tested using N = 16000 samples.

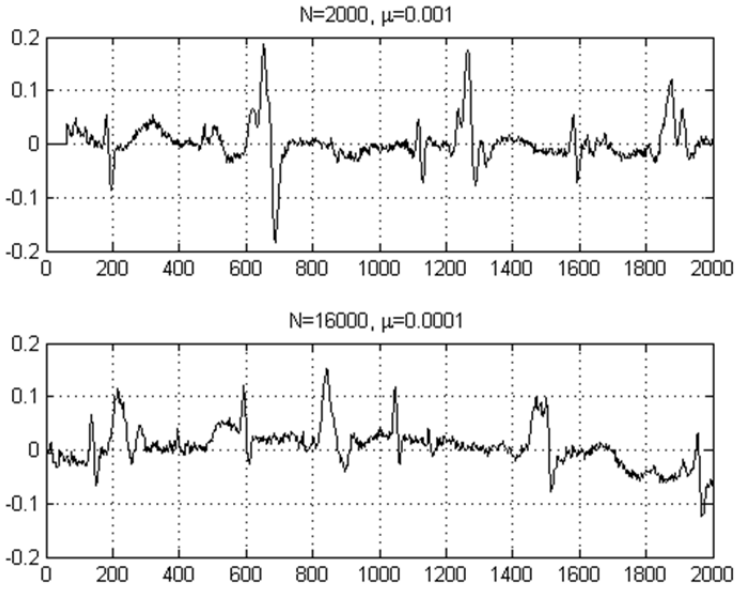

Figure 5: Time domain adaptive noise canceller output. (Top) $e[n]$ for a filter run for 2000 iterations. (Bottom) Last 2000 samples of $e[n]$ for a filter run for 16000 iterations.

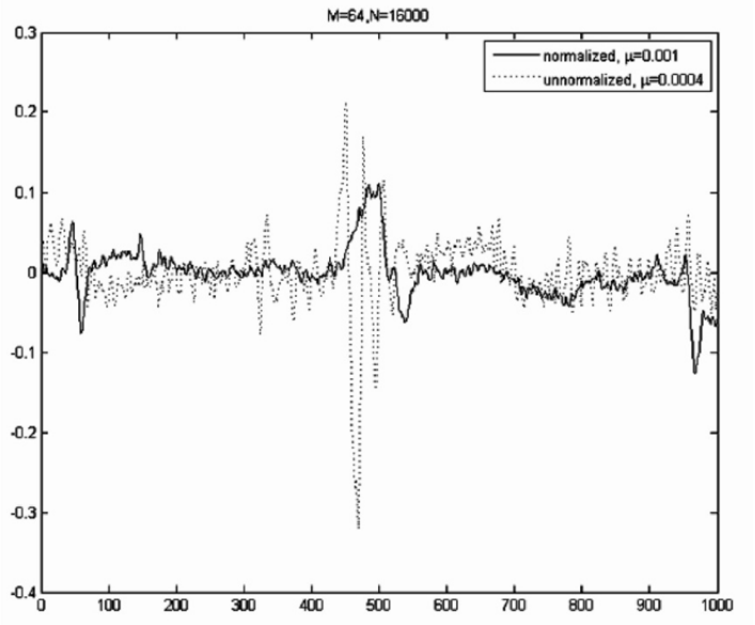

Figure 6: Time domain power normalization comparison. Last 1000 samples.

It can be seen in Figure 6 that power normalization had a large effect on the results of the time domain configurations of the noise canceller. Likewise, power normalization had a large effect on the time domain linear predictive filter.

Without the power normalization, the system would change drastically due to the discrepancy between the large peaks of the maternal heartbeat and the relatively small peaks of the fetal heartbeat. By normalizing each iteration of the filter these wide variations were avoided in the output signal. As a result the maternal signal is much better filtered.

In Figure 6, the unnormalized signal contains much higher noise levels. In addition, the maternal signal is not effectively filtered, and the fetal signal is not preserved. The normalized signal, on the other hand, much more effectively maintains the fetal heartbeat while damping the maternal peaks. The error smoothing successfully removed much of the high frequency noise found in the signals without losing the fundamental signal characteristics. Figure 7 shows a comparison of output signals from the noise cancelling filter with $M = 64$, $N = 16000$, $\mu = 0.001$, and no power normalization. Although the signal is clear in both plots, it can be seen that $e[n]$ contains much more high frequency noise than the smoothed $\varepsilon[n]$.

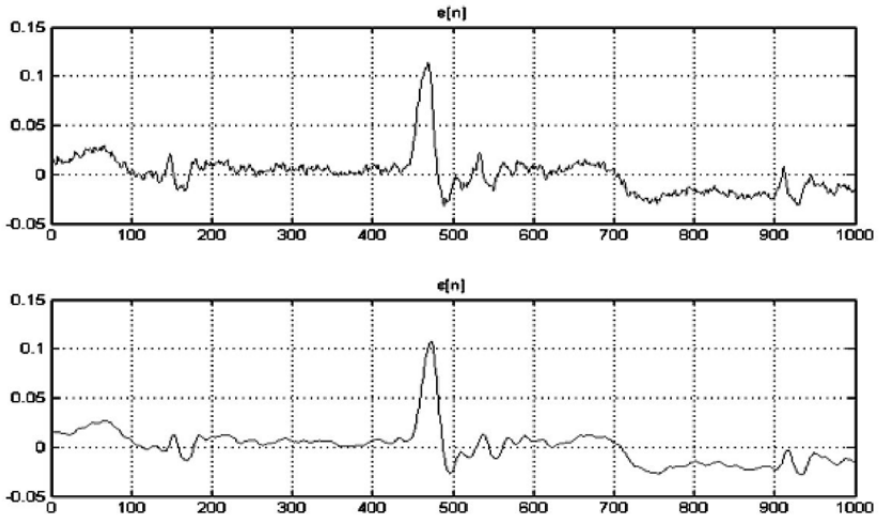

Figure 7: Comparison of error smoothing

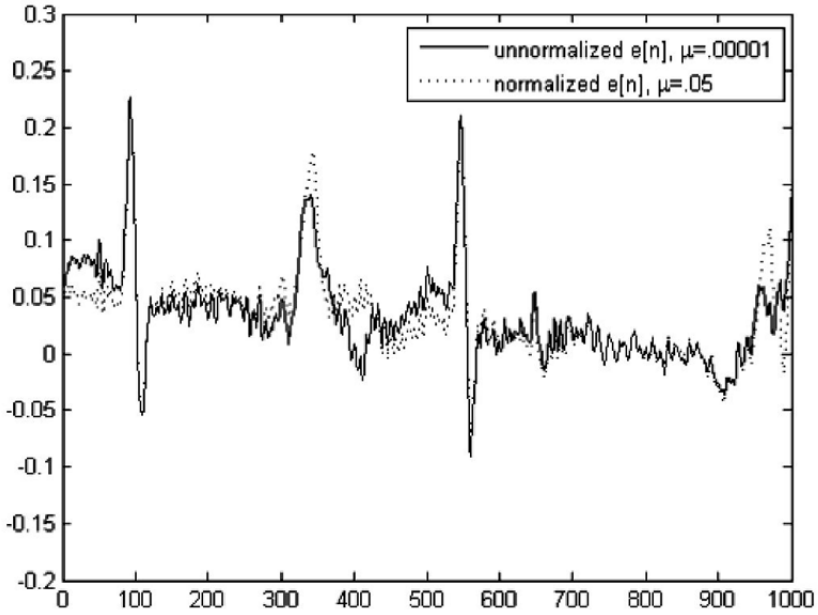

Figure 8: Comparison of power normalized and unnormalized outputs from the transform domain noise canceling filter.

Contrary to the time domain adaptive filters, it was found that the transform domain adaptive filters performed better without power normalization. It can be seen in Figure 8 that although both the normalized and unnormalized filters reduce the maternal component in the output signal, the non-power normalized filter suppresses the maternal peaks more effectively than the normalized filter. This may be due to the low frequency and narrowband nature of the ECG signals. By normalizing all frequency channels in the filter, higher frequency bands that contain only high frequency noise may be normalized as well.

It was also found that the time domain power normalized noise canceller and the unnormalized transform domain noise canceller produced similar results. Figure 9 shows that the last 1500 iterations from both filters are relatively similar. More testing needs to be done to determine whether time domain or transform domain filters are more effective.

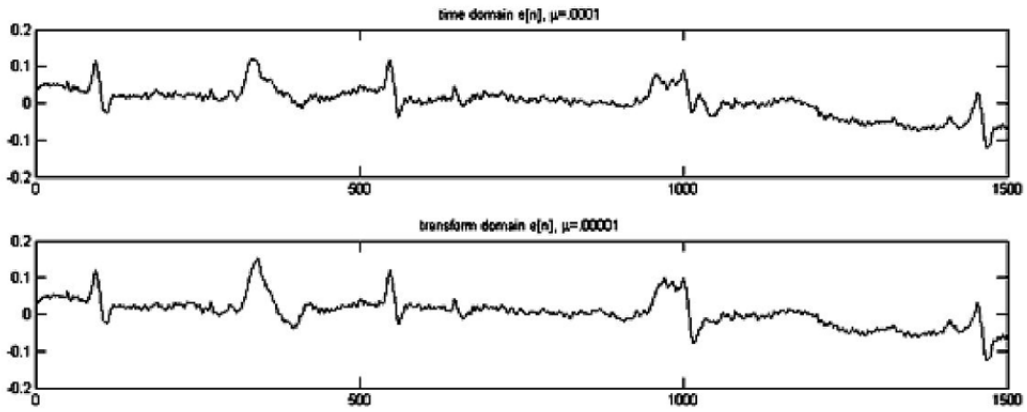

Figure 9: Comparison of time domain $e[n]$ (top) and transform domain $e[n]$ (bottom) for noise cancelling configurations.

The multi-channel transform domain noise canceller appears to perform similarly to the single channel noise canceller. The multi-channel filter, however, has the advantage of multiple outputs that represent the filtered signal. Figure 10 shows a comparison of two outputs from the multi-channel filter and one output from the single-channel filter. It can be seen that the single channel $e[n]$ performs better than the averaged, multi-channel $e_{avg}[n]$. However, the bottom signal $e_3[n]$ reduces the maternal peak better than both $e[n]$ and $e_{avg}[n]$. The multiple inputs and outputs provide a choice for the best output signal from the multi-channel filter.

Linear predictive configurations of the transform domain filter also performed similarly to the noise cancelling configurations. Comparing Figure 11 to the bottom signal in Figure 9, it appears that although both signals are similar the linear predictive configuration more adequately reduces the maternal peaks.

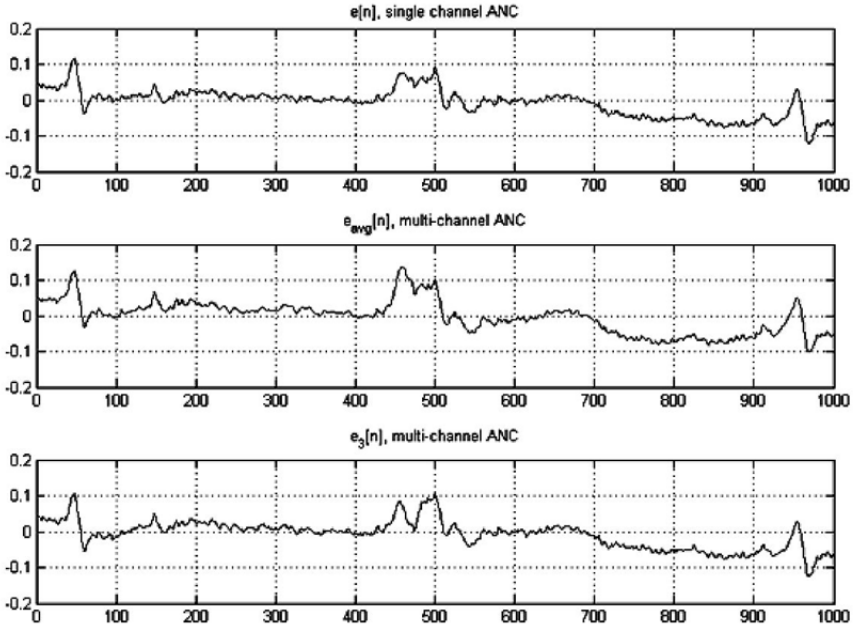

Figure 10: Comparison of single channel transform domain noise canceller and multi-channel transform domain noise canceller.

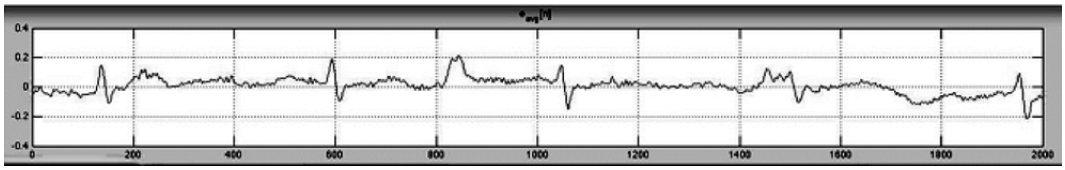

Figure 11: $e_{\text{avg}}[n]$ from multi-channel transform domain linear predictive filter

DISCUSSION

Testing different filter lengths provided results consistent with Kam's assertion that lower order FIR systems were adequate enough to use for the adaptive filtering [5]. Adaptive filters with an order of at least 32 performed similarly to one another. Additionally, filters with a larger number of iterations performed better than those with a smaller number of iterations.

Both the linear predictive and noise cancelling configurations appeared to work well when filtering out the maternal noise. The linear predictive adaptive filter possesses the advantage of not requiring a reference thoracic signal. Despite the lack of a reference, though, the fetal signal consistently remained intact after filtering. Whether one filter performs better than the other is uncertain. This is a test that will need to be done in future experiments. Likewise, both the time domain and transform domain filters performed well. Although it seemed that the transform domain filters had a tendency to outperform the time domain filters, once again further testing is needed to verify that hypothesis. In addition, future

experiments could explore whether multiple filters working in tandem or parallel would yield better results.

It is unsure whether more reference thoracic signals would significantly affect the results of the adaptive noise canceller. While the thoracic signals provided in the indirect fetal ECG database often contained large negative peaks, the abdominal signals only contained large positive peaks. Further research is needed to see whether differently placed reference ECG leads would provide better thoracic references and lead to better performing adaptive noise cancellers.

The results suggest that if a signal can be obtained with a clear fetal component with an amplitude above the noise margin, then the transform domain adaptive filter can effectively remove the maternal noise from the signal. The use of the artificial data ensured that all abdominal signals used contained a strong fetal component to successfully test the adaptive signal processing methods. Although the maternal and fetal signals appear correlated and their spectrums are extremely similar, they contain different fundamental periods. This coupled with their apparent narrowband nature allows them to still be adequately filtered. Further testing is needed to confirm this hypothesis.

CONCLUSION

Adaptive filtering is able to successfully remove maternal noise through the use of linear predictive and noise cancelling configurations. Not all maternal noise is able to be reduced, but large portions are sufficiently suppressed to provide useable fetal ECG data. Current results suggest that transform domain filters perform as well, if not slightly better, than their time domain counterparts. Additionally, results suggest that linear prediction is as effective as noise cancelling without the need of a reference signal.

The results also suggest that the maternal and fetal heartbeats are different enough that they can be successfully filtered without the destruction of the fetal signal. By using the synthetic signals, it was shown that adaptive signal processing methods are adequate in removing the maternal noise in fetal ECGs.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would like to thank my mentors Professor Kenneth Jenkins and Professor David Salvia for all of their help and guidance throughout the duration of the project. I would also like to thank Professor Robert Nickel for his guidance the past few months.

REFERENCES

[1] Jezewski J, Matonia A, Kupka T, Roj D, Czabanski R. >Determination of the fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram. Biomedical Engineering/Biomedizinische Technik 2012 Jul; 57(5):383-394. doi:10.1515/bmt-2011-0130.

[2] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, H. Hearn, J. Zeidler, E. Dong, and R. Goodlin, Adaptive noise cancelling: principles and applications, *Proc. IEEE*, vol. 63, no. 12, p 1692, 1716, 1975.

[3] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

[4] J. Sultanova,* W. K. Jenkins, D. A. Salvia, "Adaptive signal processing methods for removing maternal interference noise from fetal electro-cardiograms," *Annual Research Journal - Electrical Engineering Research Experience for Undergraduates*, Department of Electrical Engineering, Penn State University, University Park, PA (Journal/In House Publication)**,** vol. X, p 83-100, July 2012.

[5] A. Kam and A. Cohen, "Detection of fetal ECG with IIR adaptive filtering and genetic algorithms," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 04, p 2335-2338, 1999 (ISBN: 0-7803-5041-3).

[6] Y. Sharma and A. Shrivastava, "Periodic noise suppression from ECG signals using novel adaptive filtering techniques," *International Journal of Electronics, Computer Science and Engineering*, ISSN-2277-1956, vol. 1, no. 2, p 681-685.

[7] Farhang-Boroujeny, B. Adaptive Filters: Theory and Applications. Sussex: John Wiley & Sons, 1998. Print

[8] A. D. Poularikas and Z. M. Ramadan, *Adaptive Filter Primer with MATLAB*, CRC Press, Taylor and Francis Group: Boca Raton, FL, 2006.

[9] W. K. Jenkins, *Advanced Concepts in Adaptive Signal Processing*, Kluwer Academic: Boston, MA, 1996.

[10] W. K. Jenkins. D. F. Marshall, and C. Radhakrishan, "Transform Domain Adaptive Filtering." *The DSP Handbook*, 2nd ed., Boca Raton: CRS Press, 2010.

CHARACTERIZATION OF QUADRUPOLE RESONANCE TRANSITIONS IN CYANURIC ACID

Amy M. Hein* and Jeff Schiano[#]

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 research characterizes the quadrupole resonance transitions in cyanuric acid using pulsed nitrogen-14 quadrupole resonance (QR) spectroscopy, and is part of a larger investigation aimed at studying the breakdown of energetic materials in the environment. QR spectroscopy affords an exceptionally high sensitivity to chemical structure because the electric field gradient tensor, generated by electronic charges surrounding nitrogen-14 nuclei, determines QR transition frequencies. As a result, chemical shifts are orders of magnitude larger for QR spectroscopy than for nuclear magnetic resonance (NMR) spectroscopy. The broad goal of our research is to study the photocatalytic degradation of a chemical analog for an energetic material to cyanuric acid. Available literature provides the location of QR transitions in cyanuric acid at 77 K, but not the values of the relaxation and decay constants. These values will facilitate optimization of pulse sequences for detecting the end product, cyanuric acid, in degradation experiments. This study has two specific aims. The first aim is to determine the spin-lattice and spin-spin relaxation time constants and the inverse-linewidth parameter of QR transitions in cyanuric acid at 77 K. Because the measured time constant associated with spin-lattice relaxation exceeds five minutes, multi-pulse sequences are necessary to increase the measurement signal-to-noise ratio per unit time. For this reason, the second aim of the study is to demonstrate the use of the spin-locked spin-echo multi-pulse sequence to detect cyanuric acid.

[#] Faculty Mentor

INTRODUCTION

Background

The use of energetic materials and propellants results in soil and groundwater contamination. The environmental fate of 1,3,5-Trinitroperhydro-1,3,5-triazine and 2-Methyl-1,3,5-trinitrobenzene, commonly known as RDX and TNT, are well known [1]. In contrast, less is known regarding the environmental fate of recently introduced nitrogen rich energetic materials, such as the substituted tetrazines and tetrazoles, 2,4-dinitroanisole, nitroguanidine, and 3-nitro-1,2,4-triazol-5-one. The high concentration of nitrogen atoms limits the use of NMR spectroscopy to study breakdown pathways, because the electric quadrupole moment of nitrogen nuclei strongly broadens the NMR linewidth [2]. In contrast, nitrogen-14 QR spectroscopy uses the electric quadrupole interaction to achieve exceptionally high sensitivity to chemical structure. Our broader goal is to determine the feasibility of using nitrogen-14 QR spectroscopy for investigating the breakdown of energetic materials in soil.

For safety and material accessibility, we use chemical analogs for nitrogen rich explosives such as the s-triazine based herbicide atrazine, and melamine, which is used in a host of products including plastics and pharmaceuticals. The adsorption and photocatalytic breakdown pathways of atrazine and melamine both lead to the end product cyanuric acid [3-5]. While the QR transition frequencies in cyanuric acid are known at 77 K, the spin-lattice and spin-spin relaxation times are unknown, as is the decay constant. Knowledge of these parameters is essential for optimizing pulse parameters so that small quantities of cyanuric acid resulting from photo catalysis may be detected. The first aim of this study is to determine the values of the relaxation and decay constants at 77 K for each observable QR transition. This study shows that the spin-lattice relaxation time constant is exceptionally long, thereby increasing the detection time for small quantities of the photodegradation product cyanuric acid. The second aim of this study is to determine whether or not a multi-pulse sequence can be used to improve the signal-to-noise-ratio (SNR) per unit of time.

Quadrupole Interaction

The physical basis for QR spectroscopy is the electrostatic interaction between the electric quadrupole moment tensor of the nucleus and the electric field gradient tensor of the surrounding electronic charges [6,7]. 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, where $0 \leq \eta \leq 1$. By convention, one describes the quadrupole interaction in terms of two parameters, the quadrupole coupling constant, e^2qQ, and 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 (EFG) tensors is quantized, leading to preferred orientations of the nuclei that correspond to specific energy levels. As this study focusses on QR transitions of spin *I = 1* nitrogen-14 nuclei, there are three energy levels

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

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.

Figure 1 shows a sketch of the three energy levels for nuclei with a spin quantum number of one, as is the case for nitrogen-14. Applying an external radio-frequency (RF) magnetic field, whose frequency corresponds to the energy difference between any two energy levels, changes the orientation of the nucleus and the electrostatic interaction energy. In terms of the quadrupole coupling constant and asymmetry parameter, the QR transitions frequencies are

$$\nu_{\pm} = \frac{3}{4}e^2qQ\left(1 \pm \frac{1}{3}\eta\right) \quad (2)$$

$$\nu_d = \frac{1}{2}e^2qQ\eta. \quad (3)$$

When the EFG possesses axial symmetry, η vanishes leaving a single transition frequency.

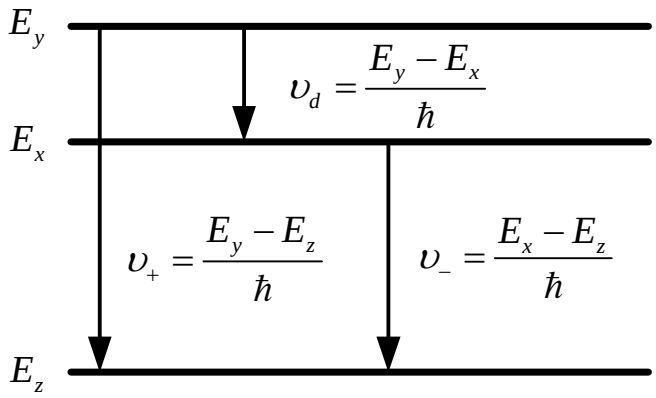

Figure 1: Energy levels and transition frequencies for spin one nuclei

Nitrogen-14 transition frequencies range from about 500 kHz to 5 MHz, and are generally not sharp. Impurities and crystalline strains produce a distribution of resonant frequencies among the nuclei. The resulting lineshape is characterized by a linewidth, typically defined as the full-width half-maximum (FWHM). The Lorentzian function

$$L(f) = \frac{1}{1+[T_2^*(f-\nu_*)]^2} \tag{4}$$

approximates the lineshape, where ν_* is the transition frequency and T_2^* is the inverse-linewidth parameter or decay constant. The FWHM of the Lorentzian line is

$$\text{FWHM} = \frac{1}{\pi T_2^*}. \tag{5}$$

In powdered samples that have been crystallized to remove impurities and relax strains, nitrogen-14 QR linewidths are typically on the order of 1 kHz.

The excitation and detection of a QR response relies on the fact that nuclei with a nonzero electric quadrupole moment tensor also possess a magnetic moment. The magnetic moment, $\vec{\mu} = \gamma \vec{I}$, is proportional to the intrinsic angular momentum, $\vec{I}$, of the nucleus, where the constant of proportionality, γ, defines the gyromagnetic ratio of the nuclei. The interaction of the magnetic moment of the nuclei with an applied RF magnetic field at a transition frequency produces a net torque on the nuclei. Because of the intrinsic angular momentum, the torque causes the nucleus to precess about the axis of the applied RF field at the transition frequency.

Free Induction Decay

A probe coil wound around the sample generates the applied RF field. Gating the applied field generates a series of RF magnetic pulses. During the inter-pulse interval, relaxation processes drive the precessing nuclei to rotate back to the orientation of lowest energy. This motion causes the nuclear magnetic moment to induce a RF voltage across the probe coil terminals at the transition frequency.

An applied RF pulse with transition frequency ν_*, amplitude B_1, and pulse width *tw*, induces an exponentially decaying voltage across the probe coil terminals. The QR receiver demodulates the probe signal to a much lower frequency f_r, resulting in the free-induction decay (FID)

$$s_{\text{FID}}(t) = S_0 \; j_1\left(2\gamma B_1 tw\right) e^{-t/T_2^*} \cos\left(2\pi f_r t\right). \tag{6}$$

The constant parameter S_0 depends on the density of nuclei within the probe coil, the population difference between energy levels, and the magnetic moment of the nuclei.

In powdered samples, the orientation of a given quadrupole nucleus with respect to the axis of the RF excitation field is a random variable with uniform distribution. As a result, not all nuclei are excited and detected with equal weight. The term $j_1\left(2\gamma B_1 tw\right)$, where

$$j_1(\alpha) = \frac{1}{\alpha^2}\left[\sin(\alpha) - \alpha\cos(\alpha)\right] \tag{7}$$

is a spherical Bessel function, accounts for the effect of the random orientation of the nuclei, and depends on the energy of the RF excitation pulse. The parameter $\alpha = 2\gamma B_1 tw$ represents an effective nutation angle of the nuclei in response to an RF pulse. The maximum value of $j_1(\alpha)$ occurs when the effective nutation angle α is 0.66π. Because a $\pi/2$ nutation maximizes the response of the FID in NMR experiments, the nutation angle $\alpha = 0.66\ \pi$ is referred to as a $\pi/2$ pulse in QR experiments.

The exponential term in Equation 6 represents signal loss caused by destructive interference among magnetization signals from nuclei with different precession frequencies. The magnitude-squared Fourier transform of the FID has a Lorentzian lineshape centered at f_r. Figures 2(A) and (B) show the Lorentzian lineshape and FID, respectively. As these signals form a Fourier transform pair, increasing the linewidth decreases the decay constant T_2^*.

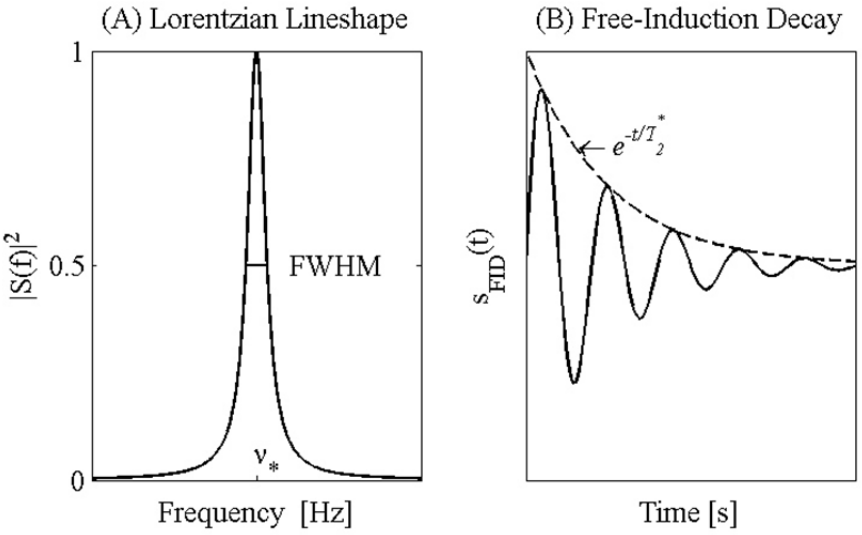

Figure 2: (A) Lorentzian line shape and (B) the FID following a RF pulse

Relaxation Mechanisms

There are two principal relaxation mechanisms in QR, spin-lattice and spin-spin relaxation [8,9]. Spin-lattice relaxation determines the length of time needed for the RF energy absorbed by the nuclei to be dissipated in the surrounding lattice so that the nuclei return to their thermal equilibrium orientation. The lattice is the general name for all other degrees of freedom of the system besides spin, such as translational motion of the molecules. For the three level system shown in

Figure 1, there are two spin-relaxation time constants, termed T_1 short (T_{1s}) and T_1 long (T_{1l}).

The spin-spin relaxation time constant, T_2, describes how energy is exchanged among nuclei through the interaction of their magnetic dipole moments. This interaction produces a perturbation in the transition frequencies that causes the magnetic moments of precessing nuclei to interfere destructively. Unlike T_2^* decay, which is the result of a time-independent disturbance in transition frequencies, the T_2 relaxation is caused by random fluctuations. As random processes are irreversible, the signal loss is not recoverable. It can be shown that $T_{1l} > T_2 > T_2^*$.

Spin Echoes

The signal decay characterized by the inverse linewidth parameter T_2^* is recoverable because it originates from time invariant processes that broaden the QR lineshape. Hahn experimentally discovered that the application of any two RF pulses rephrases the magnetic moments to produce a spin echo response, as shown Figure 3, where *TE* is the time-to-echo [10]. The envelope of the spin-echo response represents two back-to-back exponentials with time constant T_2^*, and its peak amplitude is proportional to e^{-TE/T_2}. Application of a $\pi/2$ pulse followed by a π pulse generates the largest possible echo. The Fourier transform of the spin-echo yields a Lorentzian line with FWHM given by Equation 5.

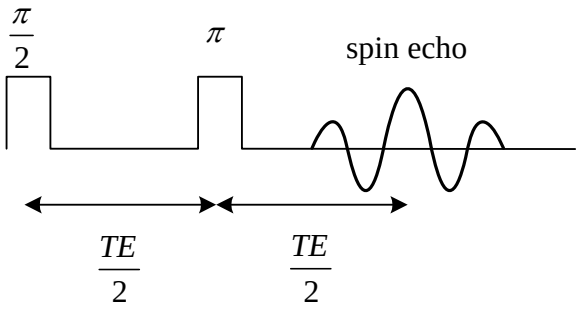

Figure 3: Spin echo pulse sequence

QR Transitions in Cyanuric Acid

Figure 4 shows the molecular structure of cyanuric acid, along with bond lengths (Ångströms) and angles (degrees) determined by neutron diffraction [11]. There are two chemically distinct nitrogen sites resulting from hydrogen bonding within a powered sample. As a result, cyanuric acid admits two ν_+ transitions and two ν_- transitions. In 1964, Widman found QR transitions at 2.7915 MHz and 2.7829 MHz [12]. He did not assign these transitions to specific sites, nor did he classify these transitions as either ν_+ or ν_-. In 1973, Lehn assigned these transitions to a single site, resulting in a value of 3.716 MHz for the electric quadrupole coupling constant e^2qQ and a value of 0.005 for the asymmetry parameter η [2]. A 1977 paper by Subbarao reports that both transitions observed

by Widman are ν_+ lines [13]. Table 1 shows the nitrogen-14 spectra reported by Subbarao for cyanuric acid. A 1983 paper by Stutz agrees with the work of Subbarao up to a single ν_+ line [14]. While Subbarao reports a value of 2.039 MHz, Stutz claims the transition is 2.309 MHz. A 1990 paper by Hiyma on QR transitions in biological systems reports transition locations nearly identical to that of Subbarao [15]. Because of the discrepancies found in the literature, a goal of this study is to verify the location of the QR transitions in cyanuric acid.

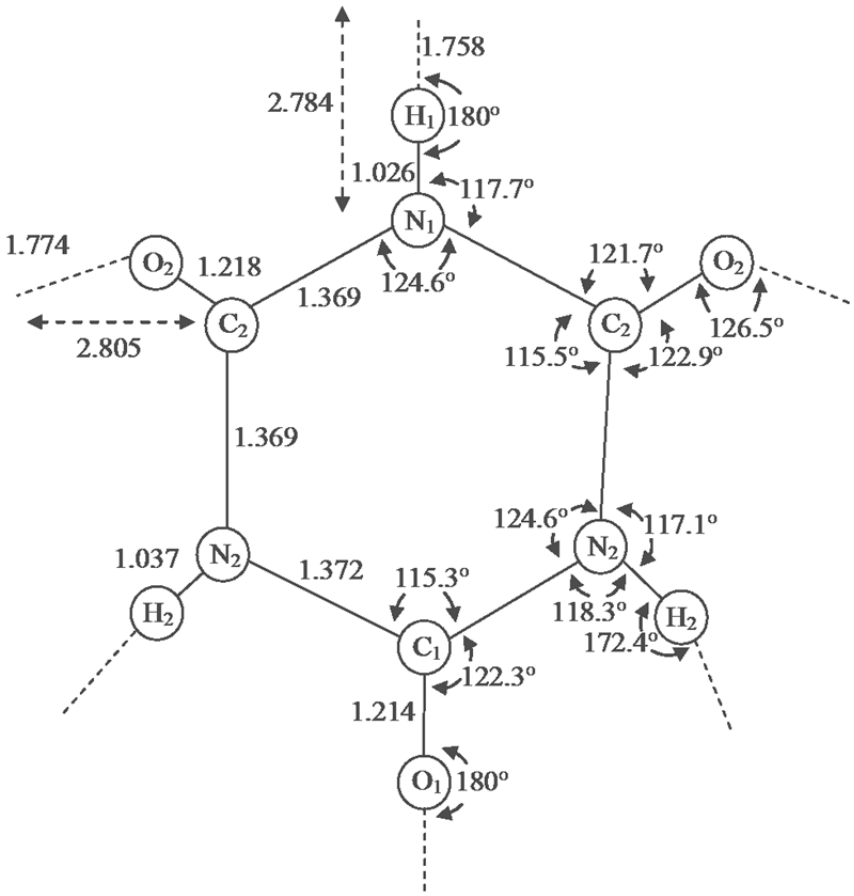

Figure 4: Molecular structure of cyanuric acid

Table 1: Nitrogen-14 spectra of cyanuric acid at 77 K

ν_+ [kHz]	ν_- [kHz]	e^2qQ [kHz]	η
2791	2039	3220	0.467
2783	1980	3175	0.506

EXPERIMENT DESCRIPTION

Materials and Instrumentation

All experiments use a 39.7-g sample of powdered cyanuric acid, CAS Number 108-80-5, contained within a 4.4-inch-long, 1.1-inch-diameter glass vial. The vial

is continually submersed in a reservoir of liquid nitrogen to maintain a sample temperature of 77 K. QR transitions are observed using a pulsed QR spectrometer [16]. The probe coil slips over the sample vial and consists of 38 turns of 18 AWG copper wire uniformly spaced across the vial length. The Quality-factor (Q-factor) of the tuned probe coil is approximately 120. An active Q spoiling system reduces the probe Q-factor to about 10 following RF pulses so that the ringdown time is 50 μs. The amplitude of all RF pulses is approximately 32 G, and the pulse width is varied as needed to obtain desired nutation angles.

Pulse Width Calibration

The duration of a π/2 pulse is determined experimentally by measuring the FID amplitude for a sequence of pulse widths ranging from 5 μs to 50 μs in 5 μs steps. The RF pulse separation is chosen to allow the sample to return to thermal equilibrium between pulses. To maximize the signal-to-noise ratio, both ν_+ transitions are excited by setting the spectrometer frequency to 2.787 MHz. The signal-to-noise ratio is further improved by acquiring and averaging sixteen FIDs for each pulse width. The spectrometer receiver offset frequency is set to 10 kHz so that a 2.787 MHz signal at the probe appears as a 10-kHz signal at the receiver output. The receiver output is filtered by an eight pole Butterworth lowpass filter with a cutoff frequency of 20 kHz to avoid aliasing and reduce the noise bandwidth. The filtered signal is digitized at 500 kHz, and 512 samples are acquired for each FID. The amplitude of the average FID is chosen as the peak magnitude of the Fourier transform of the FID.

The pulse width calibration experiment yields the FID amplitude A_i as a function of pulse width tw_i. From Equation 7, the measured data points (tw_i, A_i) satisfy

$$A_i = \frac{c_2}{\left(c_1 tw_i\right)^2}\left[\sin(c_1 tw_i) - c_1 tw_i \cos(c_1 tw_i)\right] + c_3, \tag{8}$$

where c_i are unknown constants with c_3 representing the DC offset of the receiver. We use a nonlinear least-squares algorithm to estimate the values of the parameters c_i. As $c_1\, tw = 0.66\pi$ maximizes the value of A_i, the duration of the π/2 pulse is $0.66\pi / c_1$.

Estimation of Spin-Lattice Relaxation Times

The spin-lattice relaxation time constants are measured using the two-pulse method, also known as the method of progressive saturation, which is implemented using the pulse sequence in Figure 5 [17]. The delay time, *TD*, is chosen sufficiently large so that the sample is in thermal equilibrium prior to applying the first π/2 pulse. The first pulse saturates the system by generating the largest possible FID signal. If the recovery time, *TR*, is sufficiently greater than the spin-spin relaxation time, the amplitude of the FID following the second pulse,

$$s(TR) = S_0\left[1 - c\,e^{-TR/T_{1l}} - (1-c)\,e^{-TR/T_{1s}}\right], \tag{9}$$

depends only upon *TR* and the spin-lattice relaxation times. The parameter *c* is an unknown constant bounded between zero and one. Ideally, the value of *c* is small so that the amplitude *s*(*TR*) approaches S_0 for small values of *TR*. We measure *s*(*TR*) for several values of *TR*, and then use this data long with a nonlinear least-squares algorithm to estimate the spin-lattice relaxation time constants T_{1s} and T_{1l} along with *c*.

For a given nitrogen cite, the values of the relaxation time constants are the same for both the ν_+ and ν_- transitions, while the value of *c* depends on whether the ν_+ or ν_- transition is excited. By Faraday's law, the FID induces a voltage across the probe coil that is proportional to frequency. For this reason, to maximize the amplitude of the FID, we chose to estimate the spin-lattice relaxation times by exciting both ν_+ transitions with the spectrometer frequency set to 2.787 MHz.

The experiment for estimating the spin-lattice relaxation time uses a *TD* value of 1000 s and *TR* values of 0.1 s, 0.3 s, 1 s, 3 s, 10 s, 100 s, 300 s, and 600 s. The sequence is repeated four times for each recovery time so that the resulting FIDs can be average to improve the measurement SNR. Because of the large value of delay time, the experiment requires about twelve hours to complete.

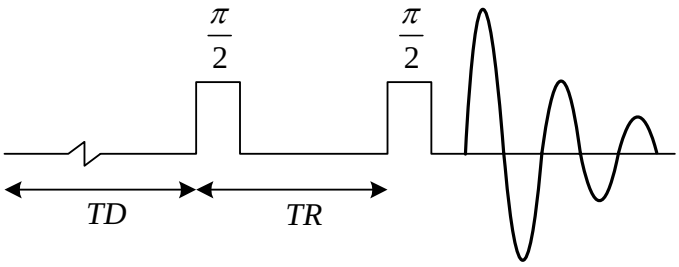

Figure 5: Sequence for estimating spin-lattice relaxation time constants

The spectrometer receiver offset frequency is set to 30 kHz so that a 2.787 MHz signal at the probe appears as 30 kHz signal at the receiver output. The receiver output is filtered by an eight pole Butterworth lowpass filter with a cutoff frequency of 60 kHz to avoid aliasing and reduce the noise bandwidth. The filtered signal is digitized at 1 MHz, and 512 samples are acquired for each FID. For a given recovery time *TR*, the amplitude *s*(*TR*) of the average FID is chosen as the peak magnitude of the Fourier transform of the FID.

Estimation of Spin-Spin Relaxation Times and Decay Constants

The spin-spin relaxation and decay time constant are measured using the Hahn spin echo sequence in Figure 3. If the sample is in thermal equilibrium prior to application of the first π/2 pulse, the amplitude of the spin echo following the π pulse,

$$s(TE) = S_0\left[1 - e^{-TE/T_2}\right], \tag{10}$$

depends only upon *TE* and the spin-spin relaxation time. We measure *s(TE)* for several values of *TE*, and then use this data along with a nonlinear least-squares algorithm to estimate the spin-spin relaxation time constant T_2. Using Equation 5, the decay constant, T_2^*, is estimated from the FWHM of the spin echo as

$$T_2^* = \frac{1}{\pi\,\mathrm{FWHM}}, \tag{11}$$

where the units of FWHM are in Hz. We chose to estimate T_2^* using the spin echo with the largest amplitude, and hence best measurement SNR.

The experiments for estimating the spin-spin relaxation and decay constants time use a *TD* value of 600 s prior to application of the π/2 pulse. The *TE* values are 5 ms, 10 ms, 20 ms, 30 ms, 50 ms, 80 ms, and 100 ms. For each *TE* value, the sequence is repeated ten times and the resulting spin echoes are averaged to improve the measurement SNR.

The spin-spin relaxation and decay constants are measured for both ν_+ transitions by setting the spectrometer frequency to 2.787 MHz. The spectrometer receiver offset frequency is set to 31.250 kHz so that a 2.787 MHz signal at the probe appears as 31.250 kHz signal at the receiver output. The receiver output is filtered by an eight pole Butterworth lowpass filter with a cutoff frequency of 60 kHz to avoid aliasing and reduce the noise bandwidth. The filtered signal is digitized at 1 MHz, and 512 samples are acquired for each FID. For a given *TE* value, the amplitude *s*(*TE*) of the average spin echo is chosen as the peak magnitude of the Fourier transform of the spin echo.

QR Detection using the Spin-Locked Spin-Echo Sequence

In cases where the spin-lattice relaxation time is larger than several seconds, it becomes impractical to improve SNR by signal averaging if each acquired signal, the pulse sequence requires an initial time delay greater than the spin-lattice relaxation time. In order to reduce the effect of noise, QR detection systems employ multi-pulse sequences that facilitate signal averaging. Examples include the spin-lock spin-echo (SLSE) sequence and the strong-off resonant comb (SORC) sequence [18,19]. Multi-pulse sequences generate a periodic train of QR waveforms that are summed to form an average waveform whose SNR is significantly larger in comparison to the SNR for any single waveform. Assuming that the ambient noise is uncorrelated across waveforms, the root-mean-square of the noise in the average waveform decreases as the square-root of the number of averages [20].

This study demonstrates the use of the SLSE sequence in Figure 6 to detect the ν_+ transitions of cyanuric acid. During the interval *TD* the nuclei relax back towards thermal equilibrium. Each RF pulse is chosen as a π/2 pulse. The first two pulses separated by *TE*/2 produce the first spin echo. The subsequent π/2 pulses rephrase the magnetic moments to produce a train of *NE* spin echoes

separated by *TE*. An important aspect of the SLSE sequence is that the peak amplitude of the spin echoes decay toward zero as the pulse sequence progresses. The rate of decay decreases as the pulse separation decreases. Due to this decay, only a limited number of SLSE signals can be averaged together to improve the SNR. Increasing the number of averages above a certain point will result in a diminishing SNR. For this study, *TD* is 1200 s, *NE* is 64, and *TE* is 1 ms.

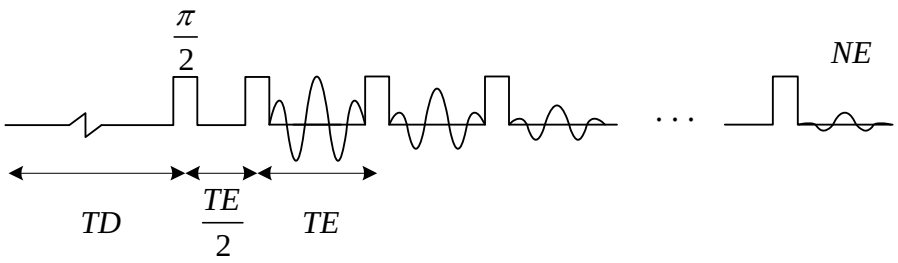

Figure 6: The SLSE sequence

EXPERIMENT RESULTS

Figure 7(A) shows the experimental results from the pulse width calibration experiment. The amplitude of the FID is shown as a function of *tw*. The open circles show the measured data points. These points are used to estimate the unknown c_i parameters in Equation 8. Using the parameter estimates, the solid curve shows the amplitude predicted by Equation 8. The amplitude of the FID is maximized when the pulse width is 17 μs. For the remaining experiments, the π/2 pulse width is set to 17 μs.

Figure 7(B) shows the magnitude spectra for each *TR* value used in the two-pulse sequence in Figure 5. The two spectral ridges coincide with the ν_+ transitions of 2.791 MHz and 2.783 MHz reported by Subbarao [13]. For each spectral ridge, the magnitude *s*(*TR*) of the ridge is recorded as a function of *TR*, and these values are used in conjunction with Equation 9 to estimate the spin-lattice relaxation times.

Figures 7(C) and (D) show the estimation results for the spin-lattice relaxation times for the QR transitions at 2.783 MHz and 2.791 MHz, respectively. The open circles show the measured data points, while the solid curve show the predicted response based on the least-squares estimate of the parameters T_{1s}, T_{1l}, and *c* that appear in Table 2.

By observing the FID following a π/2 pulse, the ν_- QR transitions were found at 2.039 MHz and 1.980 MHz as reported by Subbarao [13]. Transitions were not observed at the 2.309 MHz transition reported by Stutz [14]. We suspect that the value provided by Stutz is a typographical error. The SNR at the ν_- transitions are substantially smaller than that for the ν_+ QR transitions. This study does not provide estimates of the spin-lattice and spin-spin relaxation at the ν_- transitions because of the substantial time required for signal averaging.

Figure 8 shows data for estimating the spin-spin relaxation and decay constants for the two the ν_+ QR transitions. Figure 8(A) shows the magnitude

spectra for each spin echo as a function of *TE*. Consistent with the results in Figure 7(B), the two spectral ridges occur at 2.791 MHz and 2.783 MHz. Figure 8(B) shows the magnitude spectra when *TE* is set to 5 ms. The measured FWHM of the peaks is used to estimate the values of T_2^* for the ν_+ transitions, and these values appear in Table 2.

Along each spectral ridge in Figure 8(A), the magnitude *s*(*TE*) is recorded as a function of *TE*, and these values are used in conjunction with Equation 10 to estimate the spin-spin relaxation times. Figures 8(C) and (D) show the estimation results for the spin-spin relaxation times for the QR transitions at 2.783 MHz and 2.791 MHz, respectively. The open circles show the measured data points, while the solid curve show the predicted response based on the least-squares estimate of the parameters T_2 that appear in Table 2.

Figure 9 shows the results obtained using the SLSE pulse sequence. Figure 9(A) shows the first waveform acquired in the SLSE sequence. The dotted curve shows the waveform when the QR sample is not excited, and represents the thermal noise of the probe and noise introduced by the preamplifier. The solid curve shows a waveform which is the superposition the first QR echo and noise. Figure 9(A) reveals that the measurement noise and spin echo have comparable amplitudes. This point is amplified in Figure 9(b) which shows the magnitude spectra of the two signals in Figure 9(A). The spectrum for a single spin echo is only a few dB above the background noise.

Figure 9 (C) shows the average of the first 64 waveforms obtained in the SLSE signal. As in Figure 9(A), the dotted curve shows the background noise in the absence of a QR response, while the solid curve shows both the QR response and noise. Because each curve represents the average of 64 waveforms, the SNR is significantly improved. Figure 9(D) shows the magnitude spectra of the time domain signals in Figure 9(C). The signal peaks are now more than 30 dB larger than the background noise.

It is important to note that by using the SLSE sequence with *TE* set to 1 ms, 64 signals were acquired over a period of 64 ms. Including the 1200-s time delay preceding the first pulse, the net experiment time is 1200.064 s, or approximately 20 minutes. In comparison, collecting 64 FIDs separated by 1200 s would require 76,880 s, or approximately 21 hours. This comparison shows the time economy afforded by the SLSE sequence.

Table 2: Relaxation and decay constants in cyanuric acid at 77 K

ν_+ [kHz]	T_{1s} [ms]	T_{1l} [s]	c	T_2 [ms]	FWHM [kHz]	T_2^* [μs]
2791	28.7	737	0.288	53.2	3.17	100
2783	1.2	434	0.245	65	3.12	102

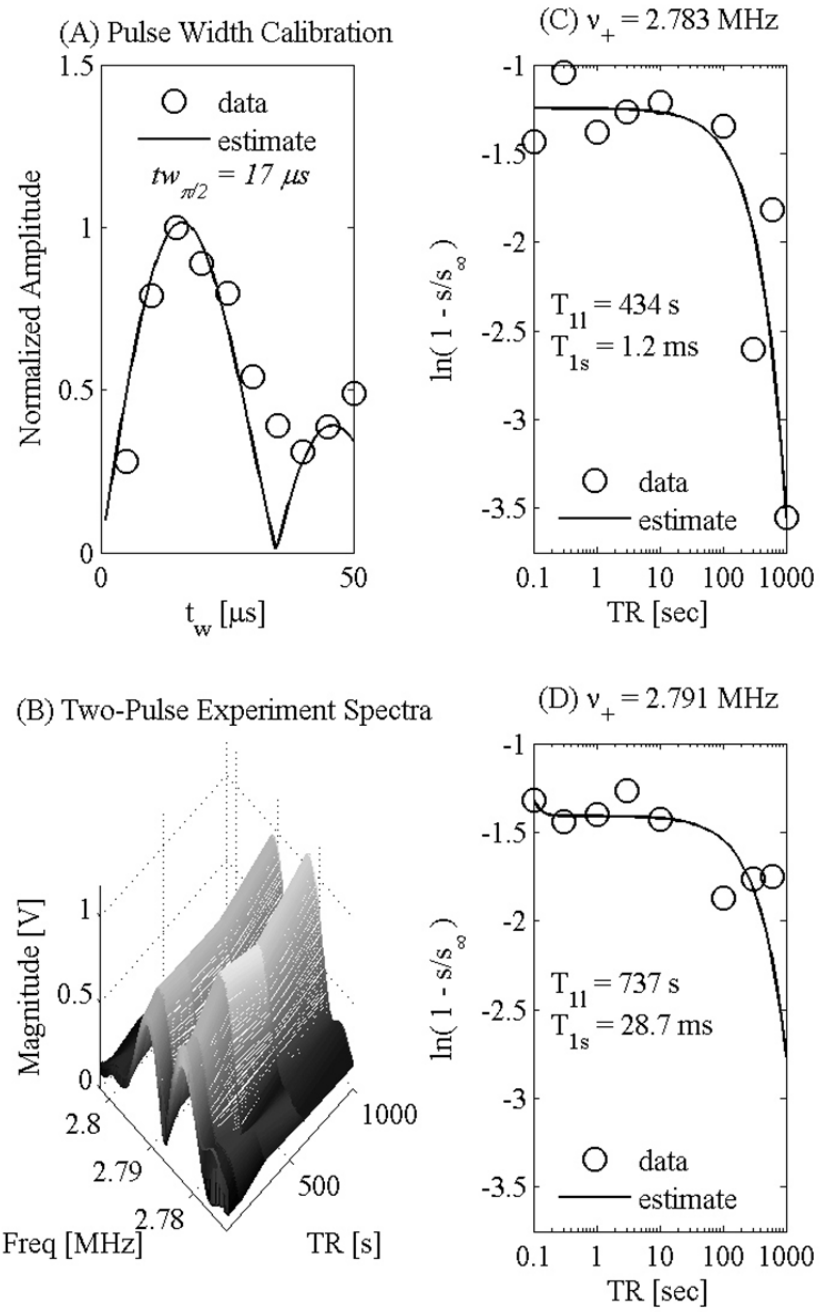

Figure 7: (A) Results from pulse width calibration experiments and (B-D) two-pulse experiments for estimating spin-lattice relaxation times

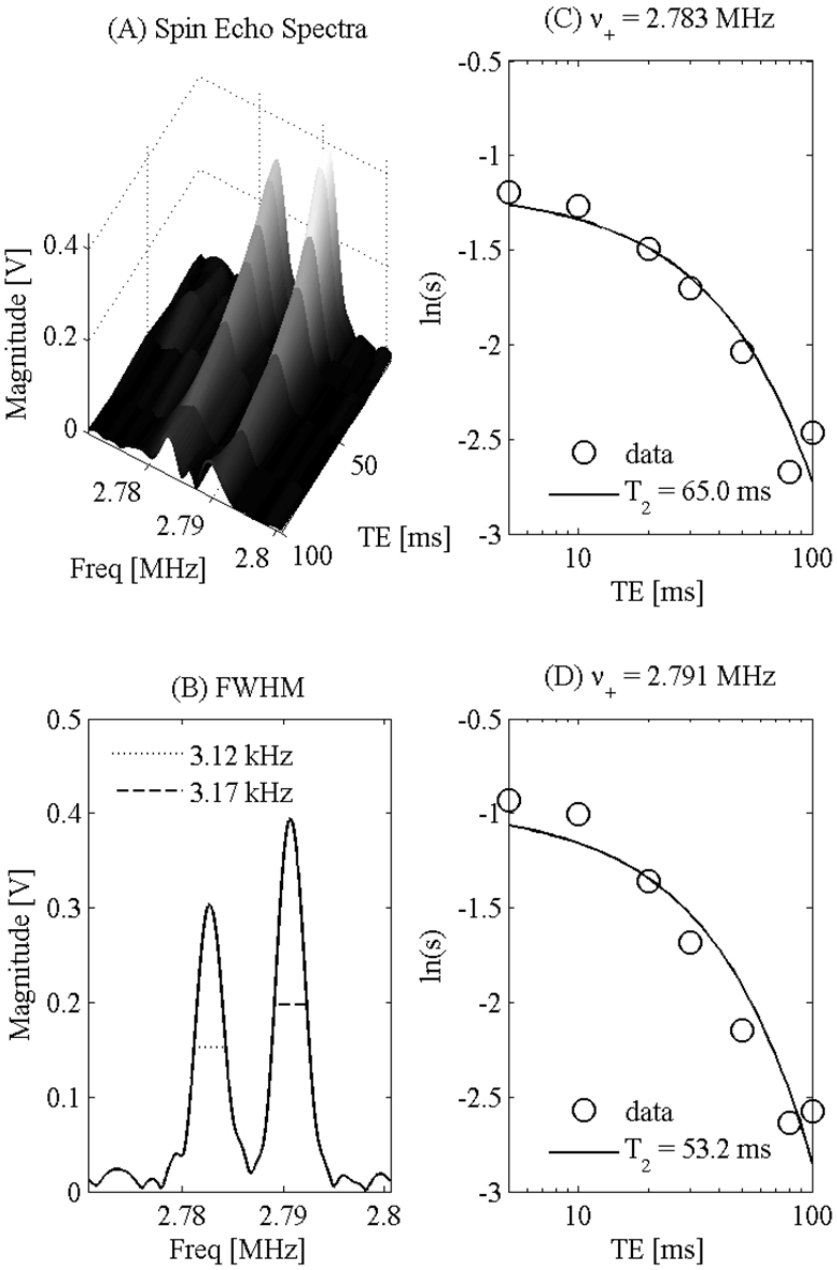

Figure 8: Estimation of decay and spin-pin relaxation times

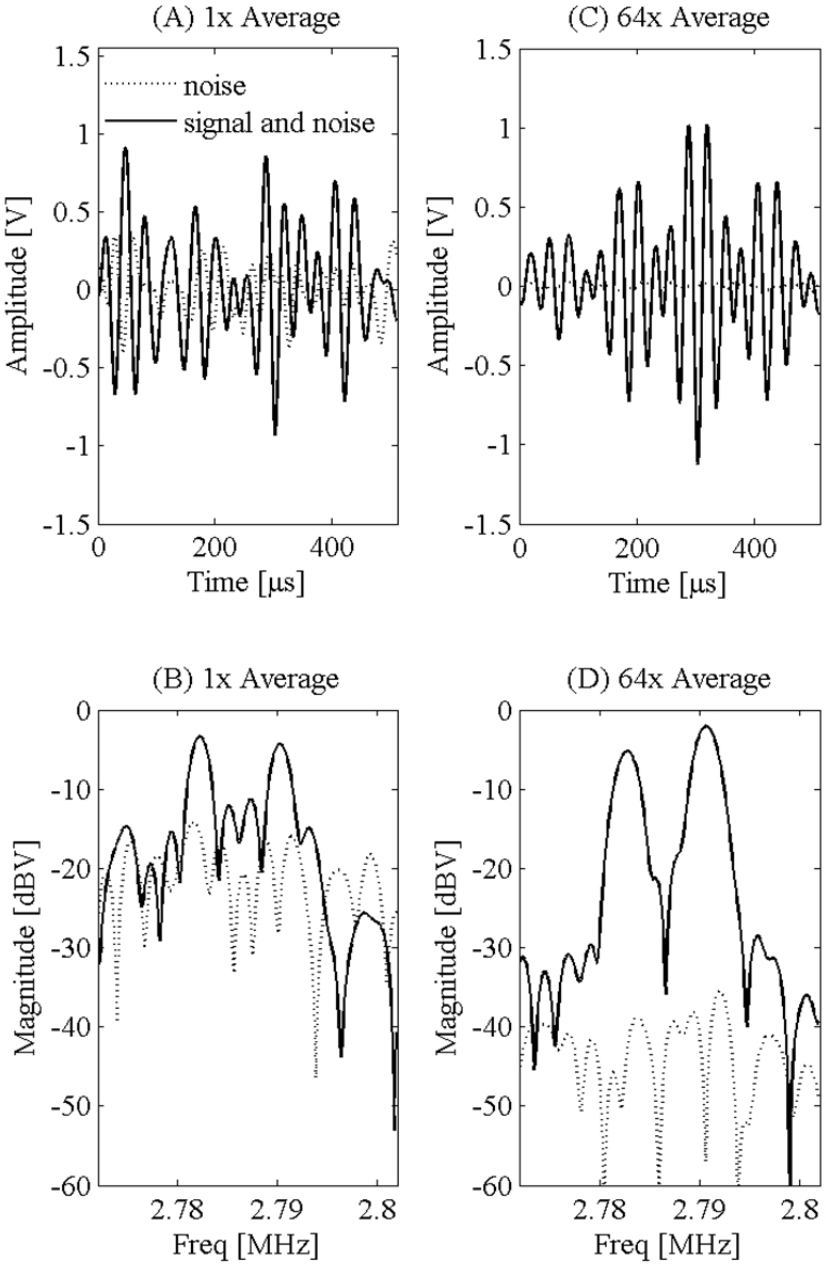

Figure 9: Detection of ν_+ transitions using the SLSE pulse sequence

DISCUSSION

The relaxation and decay times for the ν_+ transitions in cyanuric acid at 77 K are now known. Of significance is the fact that the largest spin-lattice values are in excess of five minutes. This fact motivates the use of multi-pulse experiments to improve the SNR per unit time. The observed linewidths of 3 kHz are about

three times larger than those of other nitrogen-14 transitions and yield decays time of approximately 100 μs. For this range of decays values, the use of active Q-spoiling and spin echoes are essential for observing the QR responses.

The neutron diffraction results of Coppens indicate two chemically distinct nitrogen sites in cyanuric acid, as shown in Figure 4 [11]. Because there are twice as many nitrogen atoms at site 2 than site 1, the ratio of the linewidth areas associated with these sites should be two. Figure 8(B) verifies this relationship and indicates that the ν_+ transitions at 2.783 MHz and 2.791 MHz should be assigned to the nitrogen atoms at sites 1 and 2, respectively.

Acquisition of the data used for estimating the spin-lattice relaxation times required approximately 12 hours. Despite this lengthy experiment that used an averaging factor of four, the measurement SNR was barely sufficient to estimate the spin–lattice relaxation time. It is recommended that the experiment be repeated using additional values of *TR* and an averaging factor of sixteen. Because of the excessively long time required for such an experiment, it also recommended that spin-lattice relaxation times be estimated using the steady-state continuous pulse method introduced by Alexander [17,21,22]. In this method, a continuous sequence of pulses is applied at a transition frequency and the steady-state signal following the pulses is studied as function of pulse separation.

Using the multi-pulse sequence SLSE, it was shown that the SNR per unit time can be significantly increased. A disadvantage of the SLSE sequence is that the peak amplitude of consecutive echoes relaxes towards zero, and as result, increasing the number of acquired echoes beyond a certain point reduces the SNR of the average echo. In contrast, the SORC sequence produces a steady-state response that lasts indefinitely. Unfortunately, depending on the chemical environment of the nitrogen nuclei, not all nitrogen-14 compounds admit a SORC response. It should be determined whether or not cyanuric acid admits a steady-state SORC response.

This study characterized the QR transitions in cyanuric acid at 77 K. A literature search reveals no information regarding the presence of QR transitions at room temperature. If QR transitions exist at room temperature, they are expected to have considerably smaller spin-lattice relaxations, which is advantageous for signal averaging. On the other hand, room temperature transitions are also smaller due to the reduced population difference between energy levels and lower transition frequencies. A search for room temperature QR transitions in cyanuric acid is recommended.

CONCLUSION

This study confirms the location of QR transition frequencies reported by Subbarao, and provides the relaxation and decay constants for the two ν_+ transitions in cyanuric acid. It also reveals the need for developing methods for determining the spin-lattice relaxation times that do not require impractically long experiment times. The results from this study enable the optimization of pulse

sequence parameters in experiments aimed at detecting cyanuric acid in photodegradation experiments.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Pennington, J.C. and Brannon, J.M. “Environmental Fate of Explosives”. Thermochimica Acta 384(1-2)163–172(2002).

[2] Lehn, J.M. and Kintzinger, J.P. “Nitrogen-14 Nuclear Quadrupole Effects” in *Nitrogen NMR*, New York:Plenum Publishing Corporation, 1973, ch. 3, pp. 80-161.

[3] Bozzi, A. “Evolution of Toxicity During Melamine Photocatalysis with TiO2 Suspensions”. Journal of Photochemistry and Photobiology A: Chemistry 162(1)179–185(2004).

[4] Pellzzetti, E. “Photocatalytic Degradation of Atrazine and Other s-Triazine Herbicides”. Environ. Sci. Technol. 24(10)1559-1565(1990).

[5] Lopez-Munoz, M.J and Aguado, J. and Revilla A. “Photocatalytic Removal of s-Triazines: Evaluation of Operational Parameters”. Catalysis Today. 161(1) 153–162(2011).

[6] Dehmelt, H. “Nuclear Quadrupole Resonance”. American Journal of Physics, 22(3)110-120(1954).

[7] Smith, J. “Nuclear Quadrupole Resonance Spectroscopy”. Journal of Chemical Education. 48(1)39-49(1971).

[8] Bloembergen, N., Purcell, E.M., and Pound, R.V. “Nuclear Magnetic Relaxation”. Nature. 160(4066)475-476(1947).

[9] Bloembergen, N., Purcell, E. M. and Pound, R. V. “Relaxation Effects in Nuclear Magnetic Resonance Absorption”. Physical Review. 73(7)670-712(1948).

[10] Hahn, E. “Spin Echoes”. Physical Review. 80(4)580-594(1950).

[11] Coppens, P. and Vos, A. “Electron Density Distribution in Cyanuric Acid. II. Neutron Diffraction Study at Liquid Nitrogen Temperature and Comparison of X-ray and Neutron Diffraction Results”. Acta Crystallographica Section B- Structural Crystallography and Crystal Chemistry. B27(JAN15)146-158(1971).

[12] Widman, R. "Shifts of the Nuclear Quadrupole Resonance Spectra of 14N in Urea, Cyanuric Acid, and Melamine”. Journal of Chemical Physics. 43(8)2922-2923(1965).

[13] Subbarao, S. and Bray, P. “Nitrogen-14 Nuclear Quadruple Resonance Study of Several Hydroxypyrimidines”. Journal of Chemical Physics. 67(3)1085-1090(1977).

[14] Stutz, C. and Early, D. “Temperature Dependence of the Nitrogen-14 Nuclear Quadrupole Resonance Frequencies in Selected Triazine Compounds”. Journal of Molecular Structure.111(DEC)31-35 (1983).

[15] Hiyama, Y. “Multinuclear Quadrupole Resonance Studies of Biological Systems”. Zeitschrift für Naturforschung A Journal of Physical Sciences. 45A(3-4)273-292(1990).

[16] Schiano, J.L. and Ginsberg, M.D. “A Pulsed Spectrometer Designed for Feedback NQR”. Zeitschrift für Naturforschung A Journal of Physical Sciences. 55A(1-2)61–66(2000).

[17] Vega, S. “Theory of T_1 Relaxation Measurements in Pure Nuclear Quadrupole Resonance for Spins I = 1”. Journal of Chemical Physics, 61(3)1093-1100(1974).

[18] Marino, R. “Multiple Spin Echoes in Pure Quadrupole Resonance”. The Journal of Chemical Physics. 67(7)3388-3389(1977).

[19] Kim S.S, Jayakody, J.R.P. and Marino, R. A. “Experimental Investigations of the Strong Off-Resonant Comb (SORC) Pulse Sequence in ^{14}N NQR”. Zeitschrift für Naturforschung A Journal of Physical Sciences. 47A(1-2)415–420, (1992).

[20] Ernst, R.R. “Sensitivity Enhancement in Magnetic Resonance. I. Analysis of the Method of Time Averaging”. The Review of Scientific Instruments. 36(12)1689-1695(1965).

[21] Alexander, S. and Tzalmona, A. “Relaxation by Slow Motional Processes. Effect of Molecular Rotations in Pure Quadrupole Resonance”. Physical Review. 138(3A)A845-A855(1965).
[22] Zussman, A. and Alexander, S. “Quadrupole Resonance Study of Molecular Motion in Solid Hydrazine”. The Journal of Chemical Physics. 49(11)5179-5188(1968).

DEVELOPMENT OF A USER INTERFACE BASED PROGRAM TO TRACK IMPEDANCE CHARACTERISTIC OF A MICROMACHINED QUARTZ RESONATOR

Thenmozhi Elayaperumal,* Srinivas Tadigadapa,[#] and Gokhan Hatipoglu[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
University of West Florida
Pensacola, Florida 32514

ABSTRACT

In this project, a user interface based control programming using LabVIEW is developed to track real time impedance characteristics of a Quartz resonator which is utilized as a sensor to detect frequency and quality factor changes due to mass and viscous loading on the working electrode. A Quartz resonator is an electronic oscillator circuit in which the crystal vibrates at a precise frequency when an alternating electrical signal (input) is applied. This is due to piezoelectric property of the Quartz crystal itself. These resonators can be utilized as sensors for gas detection and protein detection, by tracking the resonance shifts due to any mass or quality factor changes due to viscous loading. In order to detect the indicated impedance changes, precise programming of a Network analyzer (Agilent E5061B: 5 Hz–3 GHz is used in this work), which uses an interactive user interface, is highly desired. Three LabVIEW programs were developed, and the user's main expectation is for LabVIEW to perform real-time tracking of the impedance and phase characteristics of the resonator. In order to satisfy the user's expectations, the network analyzer is transformed to an impedance analyzer by converting s11 parameters to desired impedance parameters such as Z-ϴ (Impedance magnitude and phase) and G-B (conductance and susceptibility).The developed LabVIEW program has sophisticated features, however to highlight the main functionalities of the programs in short, the first program tracks and records the impedance maximum in frequency and records the data. The second program

[#] Faculty Mentor
[+] Graduate Mentor

continuously records the real-time impedance and phase graphs versus the whole spectrum. Third program continuously records the real-time maximum (peak) every iteration until the user terminates the program. Additional features were also added to the program such as IF Bandwidth, offset, noise filtering, and averaging.

INTRODUCTION

Quartz as a Piezoelectric Material

The accuracy of the generated frequency of a sensor is dependent upon the selected reference element. The high accuracy frequency is used to keep track of the clock signals in digital integrated circuits and stabilize the frequency for radio transmitters and receivers. A crystal is a thin "plate of quartz ground to a thickness that causes it to vibrate at a specific frequency when energy is supplied, and it is used as a frequency-control element in radio-frequency oscillators" [1]. There are numerous crystalline substances that have the basic requirements of a reference element, quartz, which has been widely used due to its many desirable characteristics. Quartz reference elements were used since the discovery of piezoelectricity in oscillator circuits. Piezoelectricity is the "electric polarization produced by mechanical strain in crystals belonging to certain classes, the polarization being proportional to the strain and changing sign with it." [2]. The Quartz crystal is made of silicon dioxide (SiO_2) and has a three dimensional geometric pattern formed by the alignment of the identical atoms. Crystals are cut at specific angles to various axes which determine the physical and electrical parameters of that particular resonator. Figure 1 shows the combination of X, Y, and Z rotational cuts in an orthogonal XYZ coordinate system.

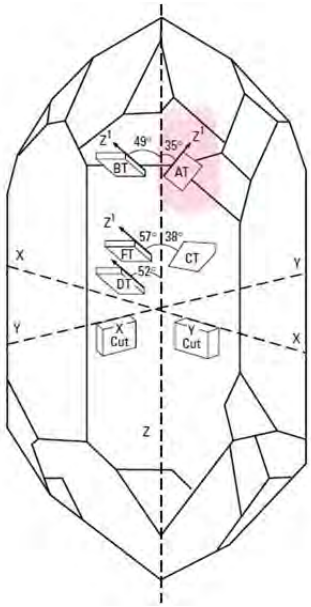

Figure 1: A typical crystal cuts from a doubly terminated quartz crystal

We used the AT-cut crystal that is cut at a 35º angle with respect to the z-axis which is highlighted in Figure 1. Improvements in quartz reference elements are being made by converting the raw quartz crystal into a usable reference element [3]. When the crystal is placed in an amplifier feedback arrangement, the crystal acts as a circuit as shown in Figure 2, which is the electrical equivalence of a quartz plate. It is shown that, as theoretically predicted by Butterworth, "such a resonator can be represented by an inductance, a resistance and a capacity all in series. These are pictured as in parallel with another small condenser and the whole is in series with a third condenser, the additional condensers representing air-gaps. The equations for the current in an oscillatory circuit, to which the resonator is attached, are developed and it is found that almost perfect agreement exists between the forms of current curve obtained theoretically and experimentally" [4].

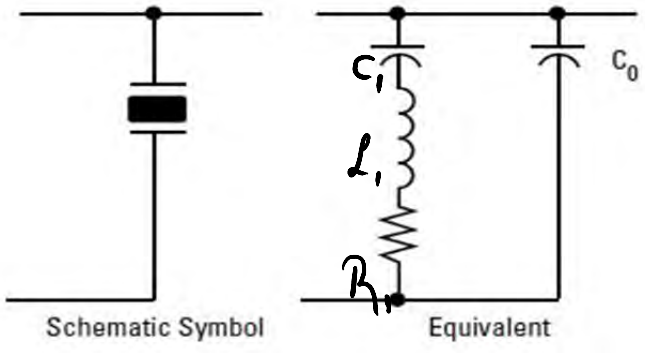

Figure 2: An equivalent circuit of a crystal

The equivalent circuit of a crystal is also known as Butterworth-van Dyke (BvD) circuit. The BvD circuit accounts for the piezoelectric stiffening and includes an "acoustic load on one side of the crystal. In the derivation of the BvD circuits, one assumes small frequency shifts as well as small loads and applies Taylor expansions in the frequency shift (or the load) whenever these variables occur" [5]. When we connect this circuit to an amplifier as in Figure 3, certain amount of energy is fed back to the crystal which causes a vibration. The vibration acts to stabilize the generated frequency at the resonance value. This circuit is also known as an oscillator circuit. An oscillator circuit configuration "accurately [tracks] the resonant frequency and increases the operating range of quartz-crystal-resonator sensor subject to heavy acoustic loading" [6].

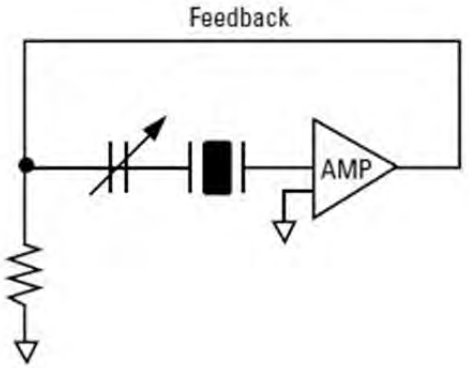

Figure 3: An amplifier feedback circuit using a crystal resonator

Quartz Resonator and Sensor

We use the AT-cut crystal which is the most common cut due to its temperature independent properties and vibration in thickness shear mode. The electric potential applied to the material, in converse to the piezoelectric effect, causes a deformation in the crystal as shown in Figure 4.

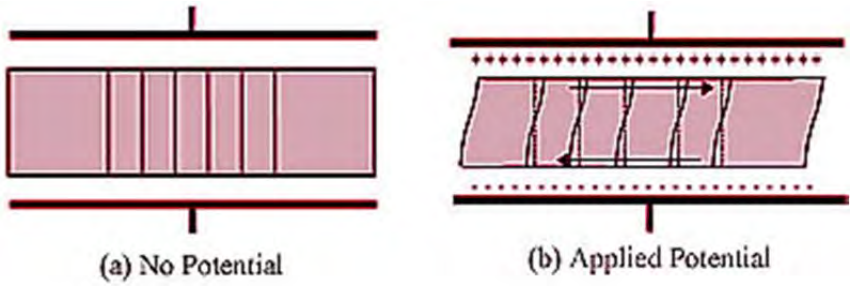

Figure 4: The thickness shear deformation

The crystal oscillates due to the alternating potential, sine wave, applied to the crystal faces. When the acoustic wavelength is half of the thickness of the crystal, a standing wave can be established with the period of the standing wave being twice the inverse of the frequency of the applied potential which is called the resonance frequency, as shown in Equation 1 below.

$$\text{Resonance frequency, } f_0 = \sqrt{\mu_q / \rho_q} \,/ \left(2t_q\right) \tag{1}$$

where μ_q is the ratio of shear stress to strain (known as shear modulus), ρ_q is the density, and t_q is the thickness of the crystal. At this frequency, the magnitude of energy lost during oscillation is at a minimum. The ratio of the stored peak energy to the lost energy per cycle is referred to as the quality factor, Q.

$$Q = \frac{f_c}{\Delta f_{FWHM}} \tag{2}$$

where f_c is the center frequency and Δf_{FWHM} is the full width at half max which is also known as the bandwidth. To track the resonance and quality factor changes, a network analyzer is used.

Impedance and Network Analyzer

Impedance, Z, is a measure of the total opposition to current flow in an alternating current circuit, made up of two components, resistance (ohm) and reactance, and usually represented in complex notation as Z = R + iX, where R is the resistance (ohm) and X is the reactance. To analyze the impedance of the resonator, an Agilent E5061B is used. The frequency ranges from 5Hz to 3 GHz, system impedance 50 Ω and 75 Ω, and S-parameter test set. In fact, the Agilent E5061B is a network analyzer and collects s11 parameter values; however in this work, the network analyzer is first converted into an impedance analyzer which is further explained in the experiment description section below.

EXPERIMENT DESCRIPTION

Experimental Setup

Agilent E5061B network analyzer (Figure 5) is programmed using LabVIEW. The first step performed is the utilization of the network analyzer as an impedance analyzer. The network analyzer is only capable of collecting s-parameters (s11 in this case). Since s-parameters and impedance are related with the mathematical expressions given in Equation 3, the mathematical operations are performed through LabVIEW to achieve impedance analyzing opportunity.

$$Z_{11} = \left[\frac{(1+S_{11})}{(1-S_{11})}\right] Z_o \tag{3}$$

For conversion between z parameters and g parameters, the following relations are used:

$$g_{11} = \frac{1}{z_{11}} \quad g_{12} = \frac{-z_{11}}{z_{11}} \quad g_{21} = \frac{z_{21}}{z_{11}} \quad g_{22} = \frac{\Delta z}{z_{11}} \tag{4}$$

where $\Delta z = z_{11}z_{22} - z_{12}z_{21}$ and Z_o is the internal impedance of the network analyzer which is 50 Ohms.

Figure 5: Agilent 5061B Network Analyzer

The devices that are used to generate signals are so sensitive that slight movements around the resonator or the network analyzer will create noise and disturb to the generated signal. Thus, LabVIEW code was developed that can be used to track the maximum impedance, record the impedance and phase as a function of frequency, and implement the impedance conversion functions such as Z-ϴ to R-X without physically approaching the experiment or the network analyzer while obtaining data.

The LabVIEW program developed gives more options and sophisticated features for users such as selecting the conversion type, trace, etc. The program also requests for a user defined IF bandwidth for noise reduction. In the end, the obtained frequency is smoothed out with the user requested smoothness percentage to analyze the data without noise. Another program was developed to collect data instantly during all iterations, detecting all the changes in real time until the user terminates the data request. However, it is recommended that the user does not use too much smoothing since it might lose the original signal data. The third program was developed with marker posts and the real time peak (maximum) values are plotted on the graph on the front panel and stored in a text file until the user terminates the request. All the programs were successfully developed and completed with no error with all specifications desired by the user. The functionality of each programs are explained in further detail in the following sections.

First LabVIEW program: Real time tracking maximum impedance and phase

In the first program that was developed in LabVIEW, the user selects the appropriate model number of the network analyzer, E5061B or 4294A, on which the data needs to be collected. Then the user selects the conversion type, a trace that needs to be displayed on the window, and path for files to record the gathered data. In addition to selecting desired options for the above mentioned features, the user inputs the center frequency, span, sweep time, offset, IF Bandwidth, and smooth aperture. The data for the selected trace is represented in two graphs on the front panel which look the same as the graph that appears on the network analyzer. The first graph represents the impedance and phase versus frequency

with the user defined span for the selected trace, and the other shows impedance and phase versus time with respective trace 1 and trace 2 with span set to zero. Figure 6 shows the control panel for the user. The advantage of the developed program is that both graphs, the graph with user defined span and zero span, and the one with frequency and time as the functions respectively, are saved and can be compared at any time; whereas, the network analyzer updates the existing graphs every time a trigger command is sent.

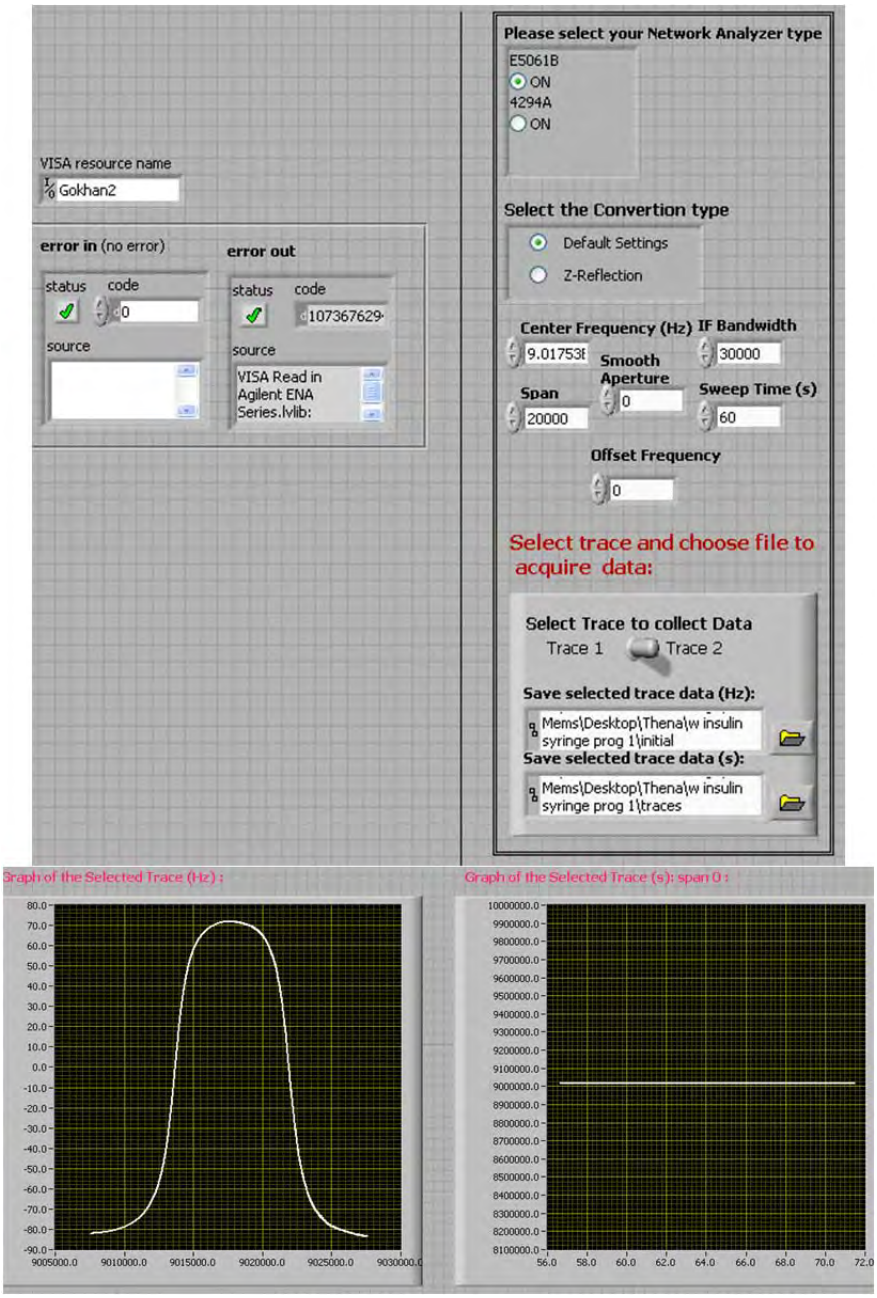

Figure 6: The user control panel in the Front panel of LabVIEW

This initially developed program starts with setting up the display of the network analyzer. The connection to the analyzer is achieved with a global visa resource name calling, where the connection type is a General Purpose Interface Bus (GPIB) connection. To communicate with the network analyzer, we use IEEE-488.2 which is a short-range digital communication bus. The "standard codes, common commands, formats, and protocols for IEEE 488.2 was introduced by the Institute of Electrical and Electronics Engineers (IEEE) in 1987 which was revised in 1992 [7]. Once the communication is established between the computer and the network analyzer, the network analyzer is calibrated via the auto calibration (the user can also manually calibrate with open, close, and a 50Ω resistor). Other interface modifications are made, such as setting the display for two traces allocated as two rows and one column and the assignment of what each trace should represent. The first row displays the impedance and the second row displays phase. Also the frequency, which is the x-axis, is divided into 1601 points. Next, the program accepts all the user input values such as center frequency, span, IF bandwidth, the desired trace and other information that were mentioned earlier. The data for both traces are collected for later analysis, but the major focus is given to the user's desired trace. The block diagram in Figure 7 shows the code written to accomplish the user's request.

Once the network analyzer is triggered, a signal is generated with the user desired frequency and span. The graph is generated on both front panel and on the network analyzer; however, the front panel displays the user desired graph while the data points are appended to the desired text file. Then the marker is set to maximum and the center frequency is set to the marker value. The network analyzer is triggered with span zero sweeping at the user desired sweep time in seconds. The noise in the triggered graph is reduced up to the IF bandwidth requested by the user. The data points are collected and the second graph is displayed on the network analyzer and the front panel of LabVIEW. The signal undergoes further noise reduction with the user input smooth aperture percentage for further investigation of the signal.

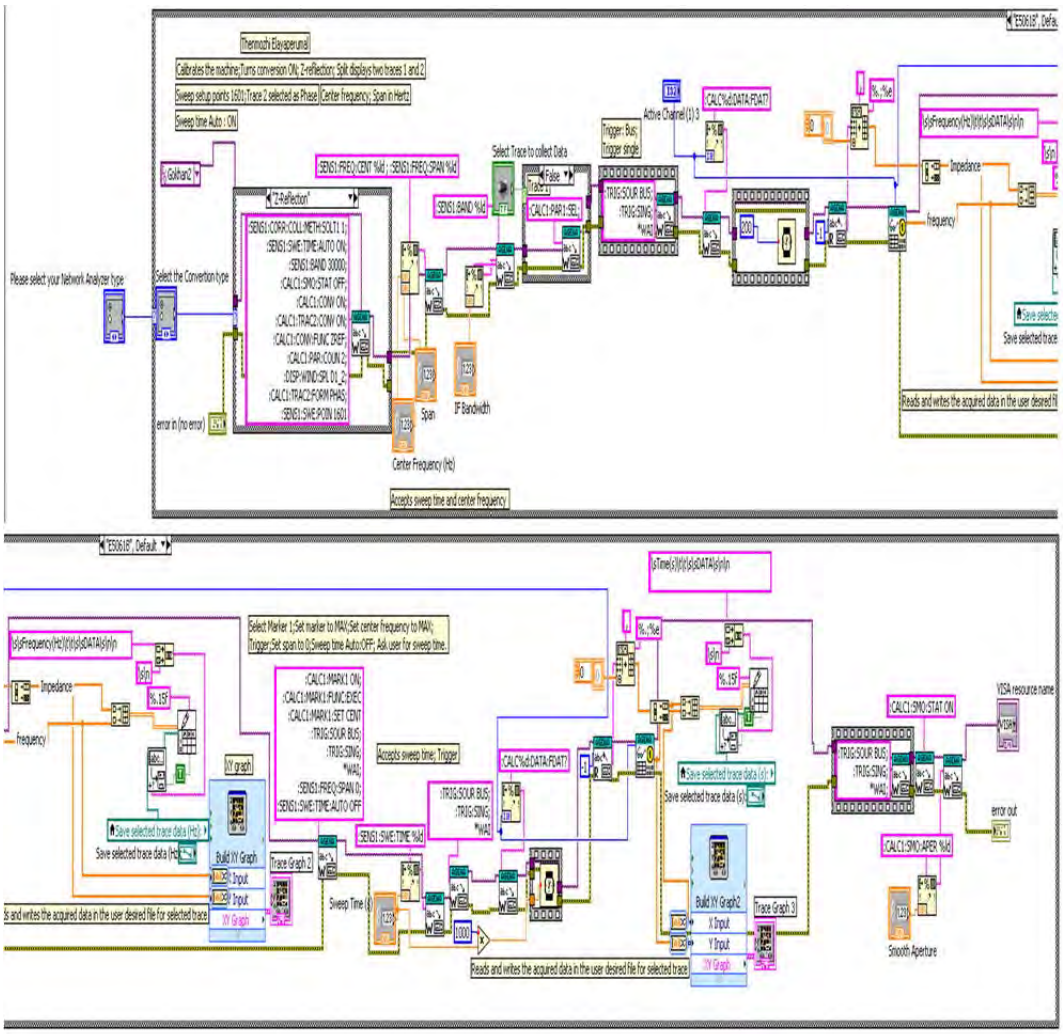

Figure 7: Block diagram of LabVIEW

Second LabVIEW program: Recording full curves within a frequency span

In this program, the user selects a computer path for two separate text files to record the data for trace 1 and trace 2, respectively. When the data is ready to be acquired, the program is executed and data is collected in each file for all iterations. The counter shows the user how many sets of 1601 points data from trace 1 and trace 2 are appended to each text files selected by the user. In addition to that, the maximum of the dedicated scan is displayed in the box for trace 1 and trace 2. The points are plotted and connected in the graph. The front panel of the second program, figure 8, has the user controls where the user can enter the information and read data off the indicators showing the maximum and the graph plotting it accurately plot. The user can stop the iteration by clicking the stop button once the desired amount of data has been acquired.

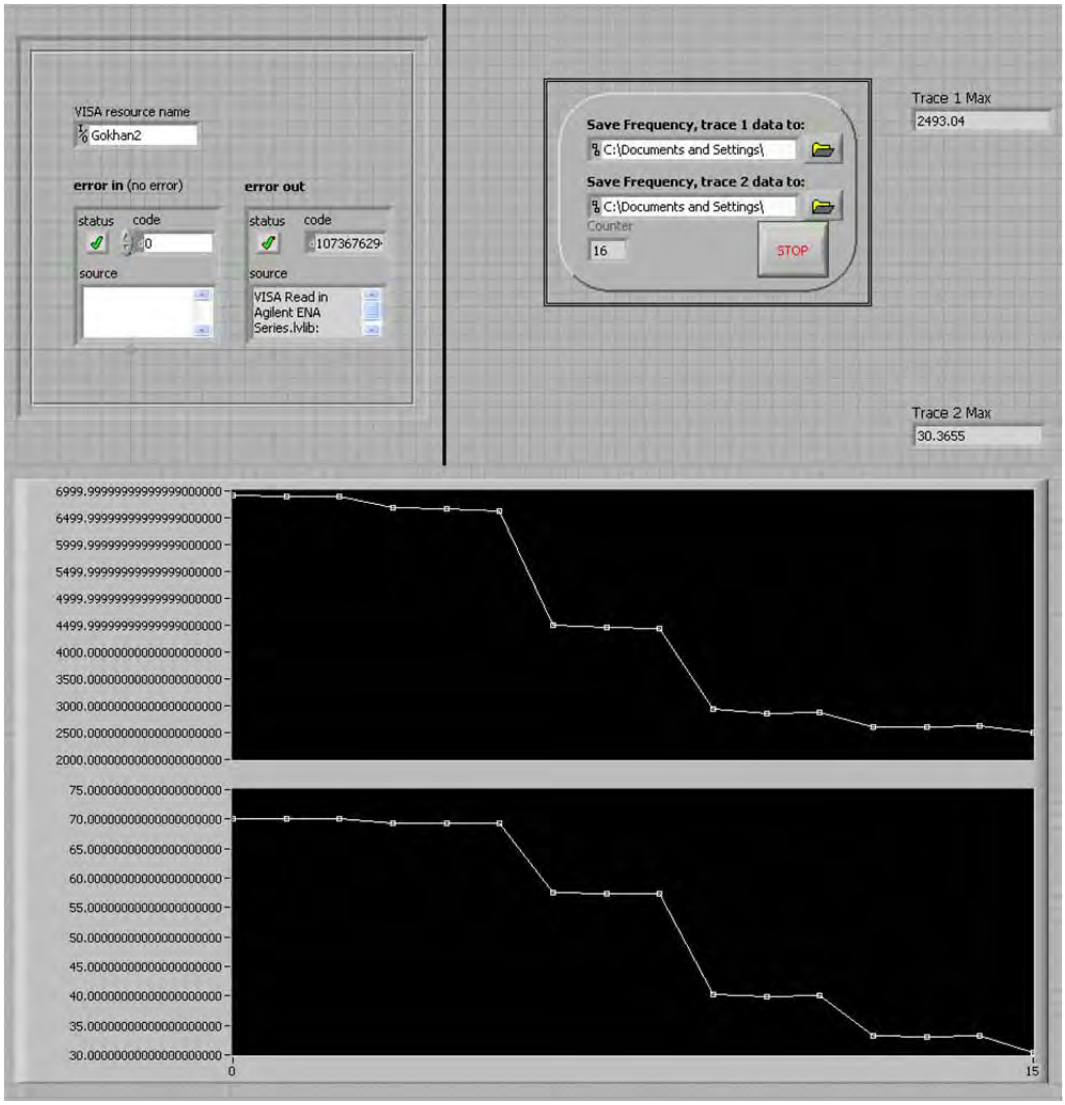

Figure 8: Front panel of LabVIEW with user controls

In this program, there are two traces allocated and the screen is split horizontally for one trace on top and other to the bottom of the screen. The connection to the analyzer is achieved with a global visa resource name calling, where the connection type is again GPIB (General Purpose Interface Bus) connection. Figure 9 shows the block diagram for the program to record full curves within the frequency span.

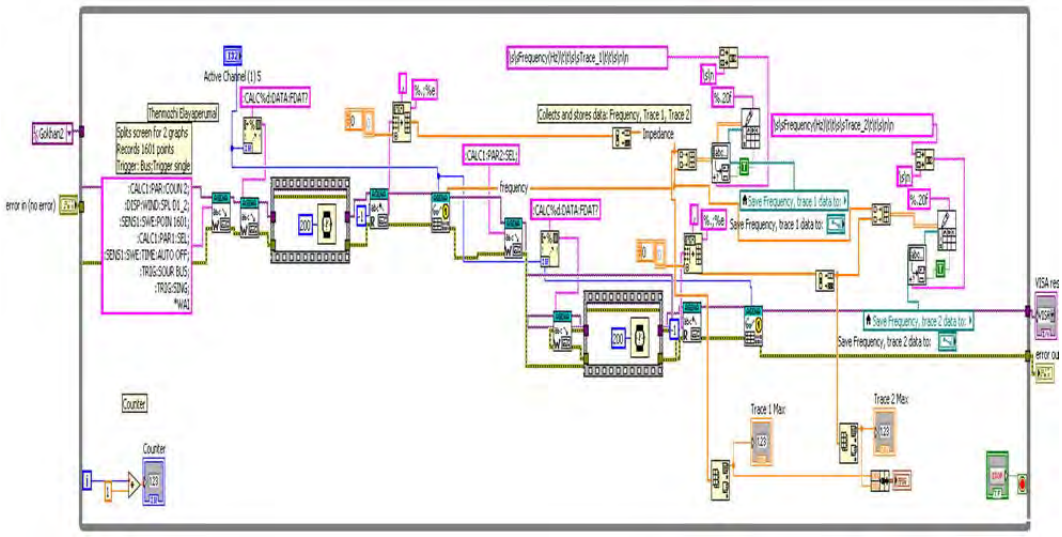

Figure 9: Block diagram of LabVIEW for the second program

When the network analyzer is triggered, 1601 data points are collected for each traces and stored in two separate files as frequency versus impedance for trace 1 and trace 2 respectively. The data is acquired and appended into a text file until the user terminates the program using the stop button. This is achieved with a while loop sequence with a stop button as the terminator. The maximum (peak) for each iteration is plotted on the graph.

Third LabVIEW program: Real-time maximum within the frequency span

In the third program, the user selects a path for one file to record the frequency, trace 1 and trace 2. When the data is ready to be acquired, the program is executed and data is collected for all iterations. The counter shows the user how many sets of 1601 data points of frequency in x-axis is obtained for trace 1 and trace 2. Similar to the previous programs, the collected data is appended to the text file chosen by the user. If the user desires to collect data to different file for every iteration, then the path of the file must be left blank. The program requests the user to select a path for every iteration. The user, however, usually would like to have all the data in one file in most cases. Similar to program two, the maximum for the dedicated scan is displayed in the box for trace 1 and trace 2. The points are plotted and connected in the graph. However, the sole purpose of this program is to record the maximum impedance and phase versus frequency. Figure 10 shows the user control for the third program developed in LabVIEW.

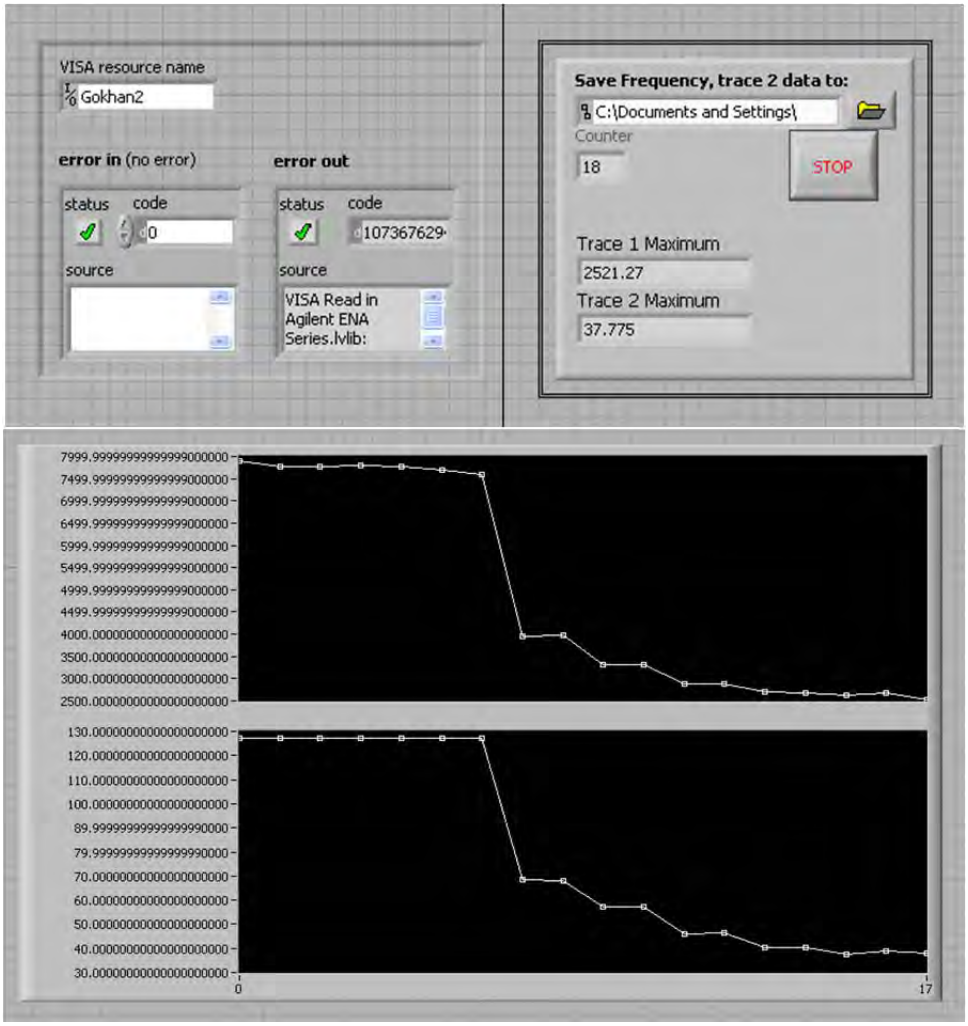

Figure 10: Front panel of LabVIEW with user controls

Similar to program two, there are two traces allocated and the screen is split into two horizontally for one trace on top and other to the bottom of the screen. The connection to the analyzer is achieved with a global visa resource name calling, where the connection type is once again GPIB connection. Figure 11 shows the block diagram for the program to record full curves within the frequency span.

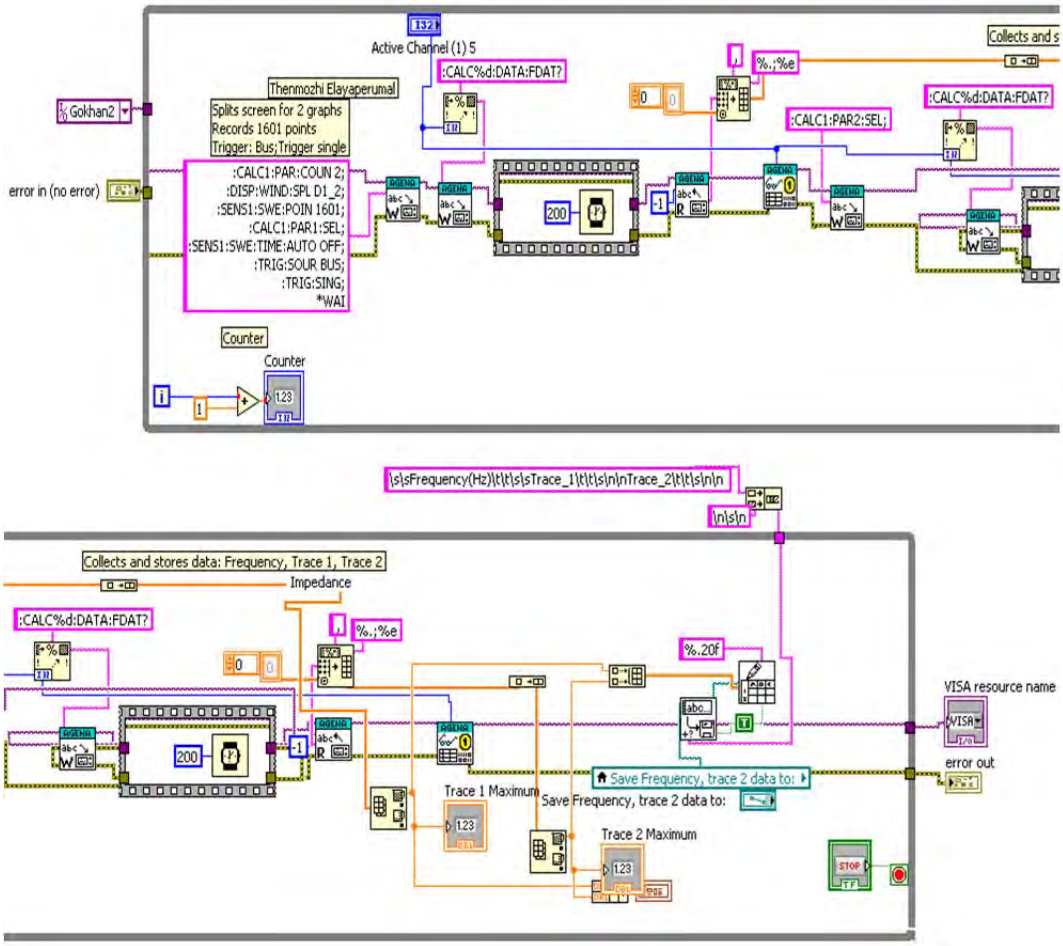

Figure 11: Block diagram of LabVIEW for the third program

When the network analyzer is triggered, 1601 data points are collected for each traces and stored in a single file as frequency versus impedance and phase as trace 1 and trace 2 respectively. The data is acquired and appended into a text file until the user terminates the program using the stop button. The maximum (peak) is plotted on the graph for all iterations.

EXPERIMENT RESULTS

All the three programs work as intended. For demonstration purposes to verify the working of the program, a water droplet test was performed on a 9 MHz Quartz sensor, whose impedance and phase characteristics are depicted in Figure 12 which is right after the program sets up the screen for the experiment under no-loading condition. The loading is performed through a controlled water droplet loading onto the sensor via a 0.3 cc syringe. The volume of each water droplet kept between 2 to 10 µL of water depending on which program is to be verified.

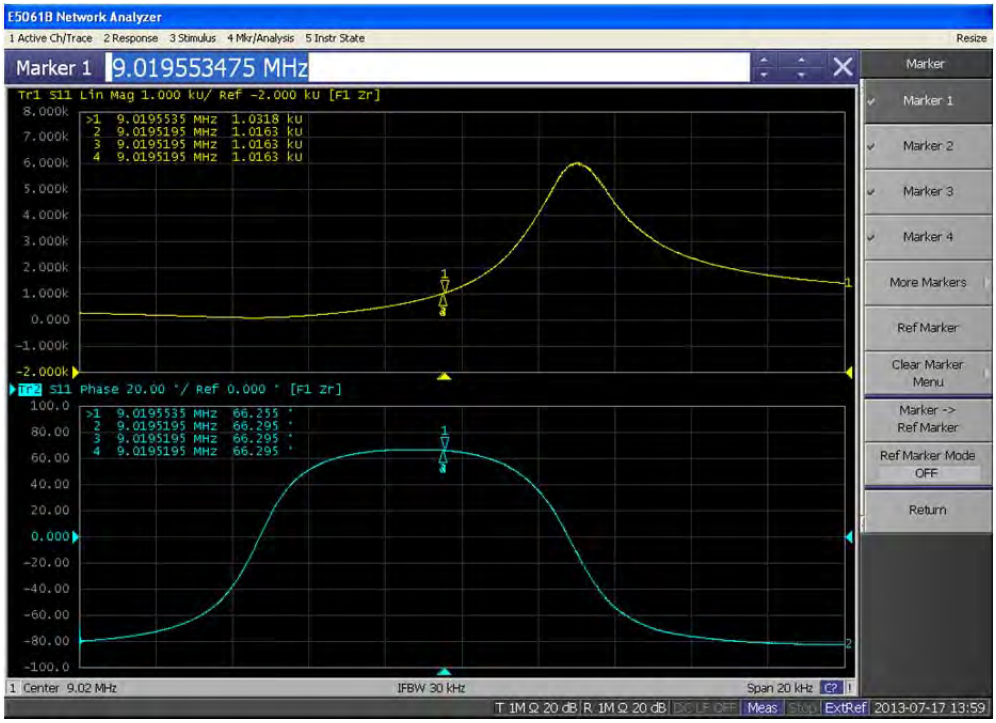

Figure 12: Block diagram of LabVIEW for the third program

Program 1 is a single point sweep tracking, where the data point of the curve maximum is recorded in a specific time that the user determines. For this case, the user requested 60 seconds of a total experiment time and added droplets of 2-3 µL of water onto the resonator. The resonator phase maximum experienced a drop in value (decrease) of approximately 1-2 degree for each droplet. The program successfully performs recording the overall data into a text file. The data is redrawn in excel as shown in figure 13 (a) and the data is then compared with the analyzer output display as shown in figure 13 (b). To demonstrate this program, the user defined settings can be referred from Figure 6. The user has selected trace 2, so as it is observed, the data shown in the network analyzer for trace 2 is the same as the data obtained from the program that was developed in LabVIEW which was plotted on excel.

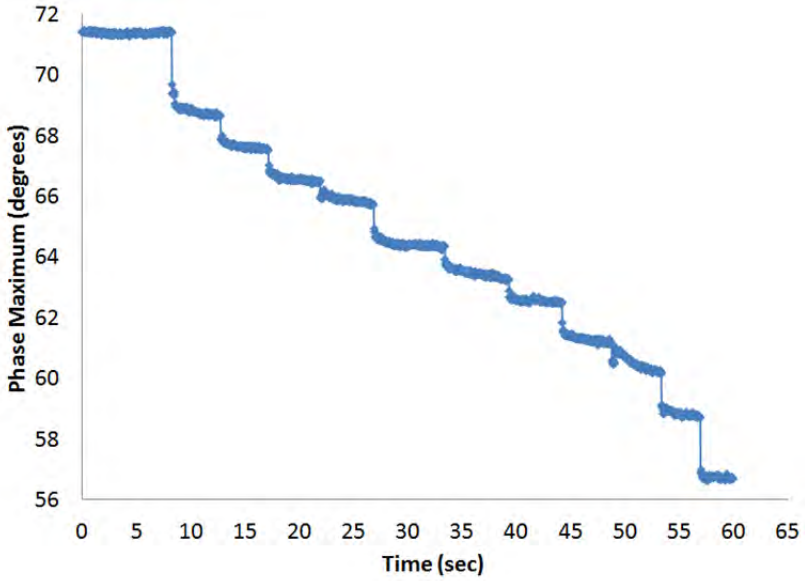

Figure 13 (a): Graph plotted from data recorded by LabVIEW program

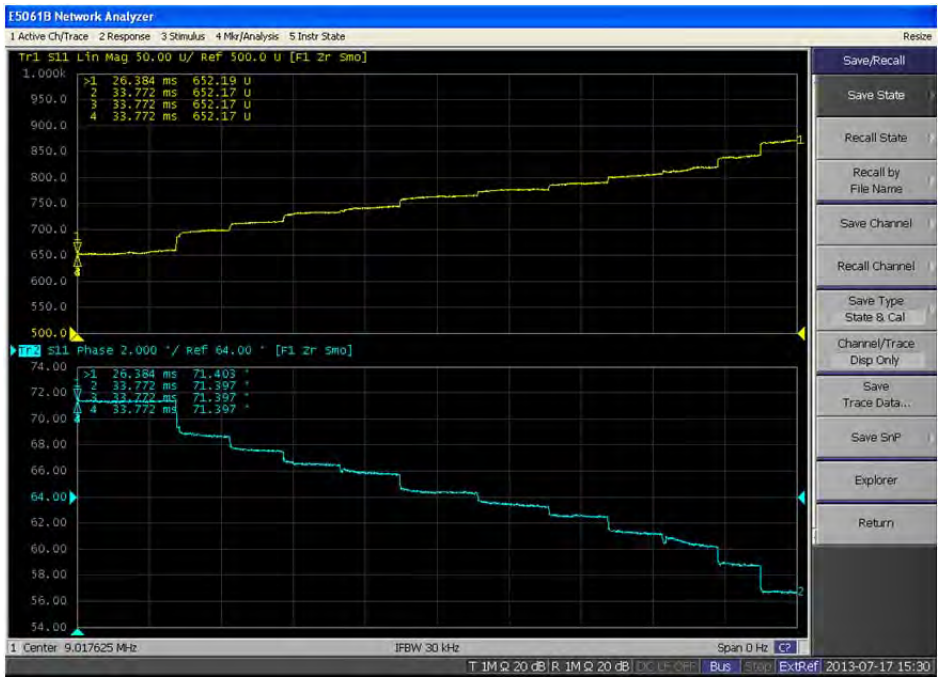

Figure 13 (b): Output from the network analyzer

Program 2, recording the full curves within a frequency span is also verified with water droplet experiments. Water droplets from previous experiment were dried out with a filter paper. In this part of the experiment the water droplet size is approximately 10 µL. Every time, the water is added twice onto the resonator. Since the mass loading is drastic, huge impedance and phase shifts were observed as shown in Figure 14a and Figure 14b. Figure 14 (c) shows the curve on the

network analyzer after the droplets have been added. The top curve represents trace 1 and the bottom curve represents trace 2 and can be compared with the plot obtained in excel from the data collected by the developed program in LabVIEW. In the case where there were no water droplet added, the resonator contacts the air. As water is added, the shift towards left is observed since the resonance shifts toward lesser frequency due to the mass loading.

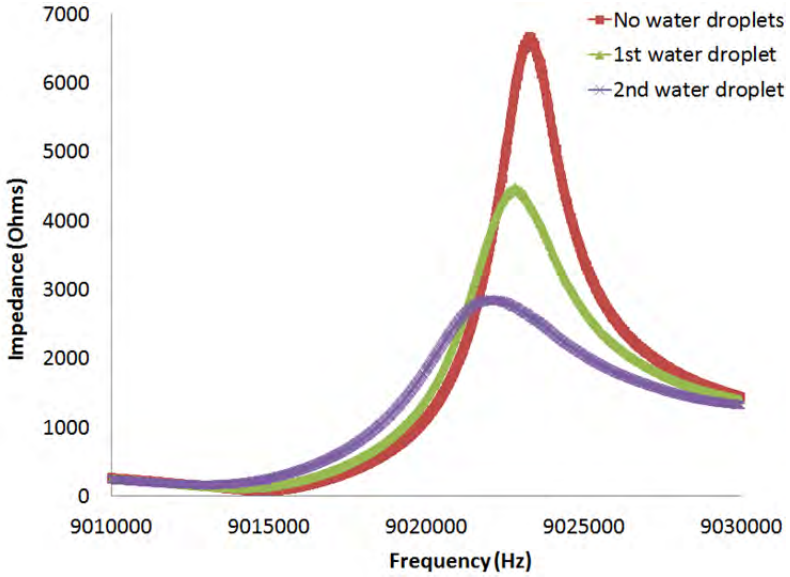

Figure 14 (a): Trace 1 plotted from data recorded by LabVIEW program

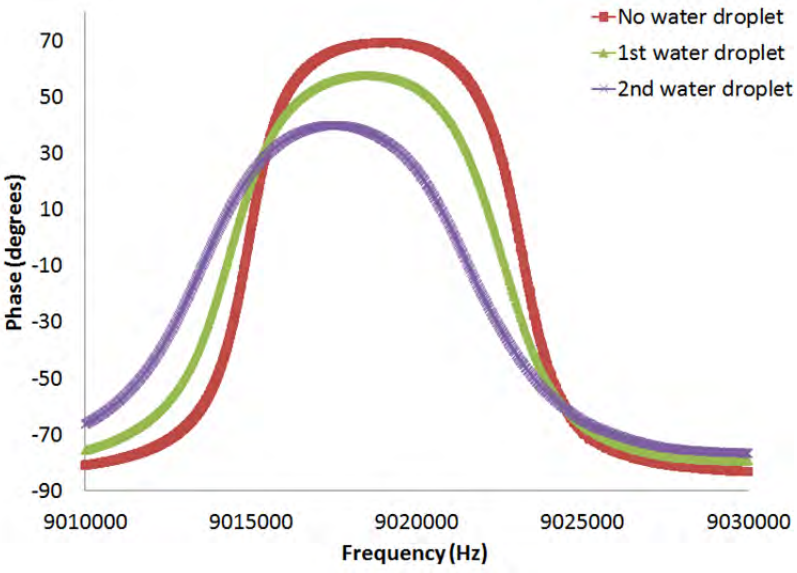

Figure 14 (b): Trace 2 plotted from data recorded by LabVIEW program

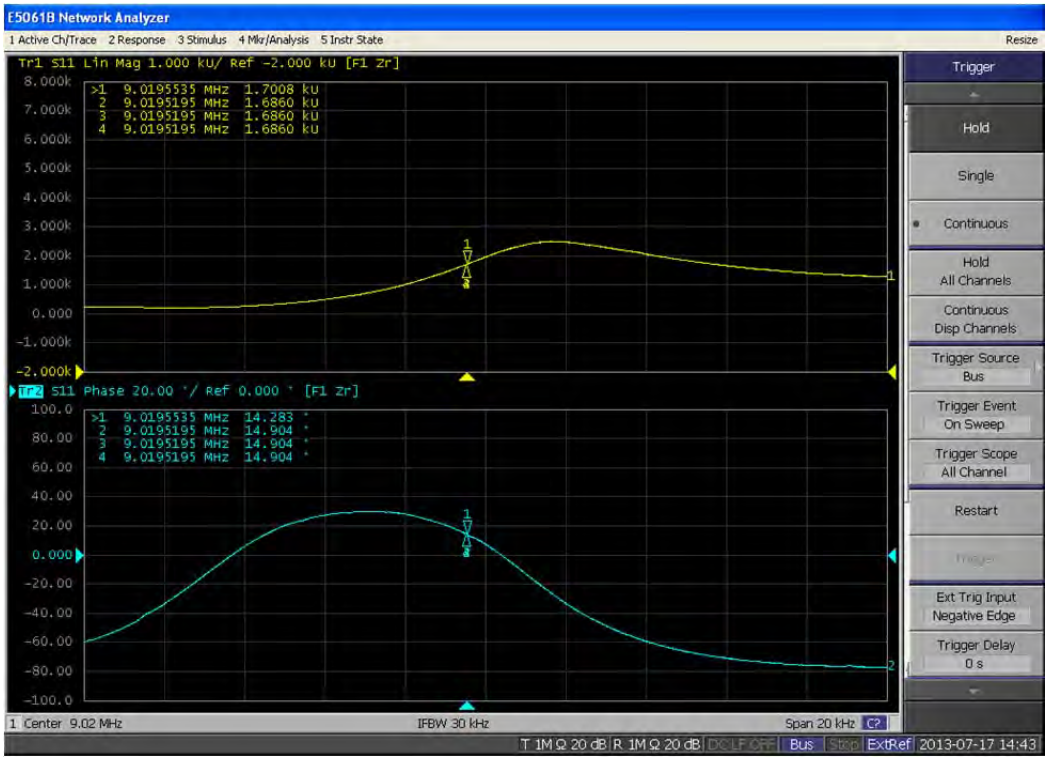

Figure 14 (c): Output from the network analyzer

Program 3, recording the maximum of the full curves within a frequency span is also verified with water droplet experiments. Water droplets from previous experiment were dried out with a filter paper. In this part of the experiment the water droplet size is maintained approximately 10 µL. Every time, the water is added twice onto the resonator. Since the mass loading is drastic, huge impedance and phase shifts were observed as shown in Figure 15 (a) and Figure 15 (b). Figure 15 (c) shows the curve on the network analyzer after all the droplets have been added. In Figure 10, the top curve represents trace 1 and the bottom curve represents trace 2 and can be compared with the plot obtained in excel, Figure 15 (a) and Figure 15 (b), from the data collected by the developed program in LabVIEW. In the case where there were no water droplet added, the resonator contacts the air. As water is added, the shift towards left is observed since the resonance shifts toward lesser frequency due to the mass loading.

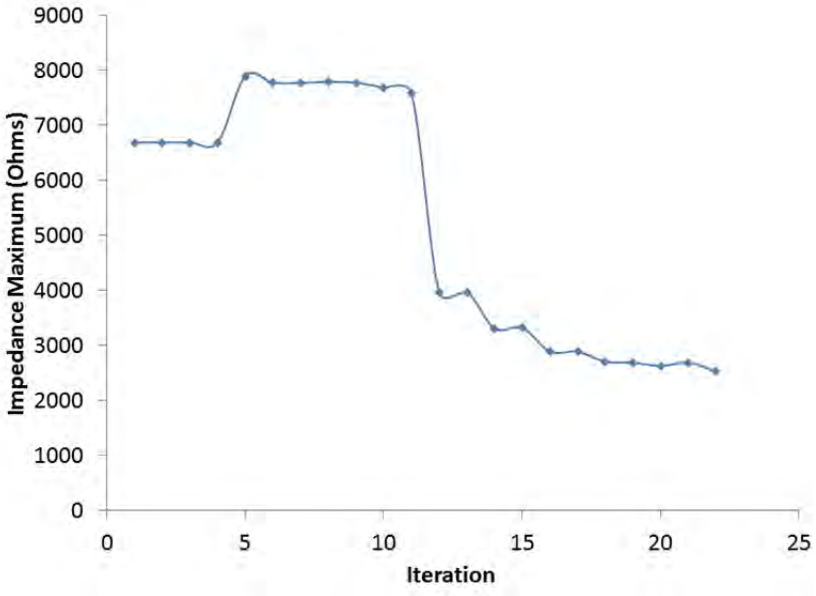

Figure 15 (a): Output from the network analyzer

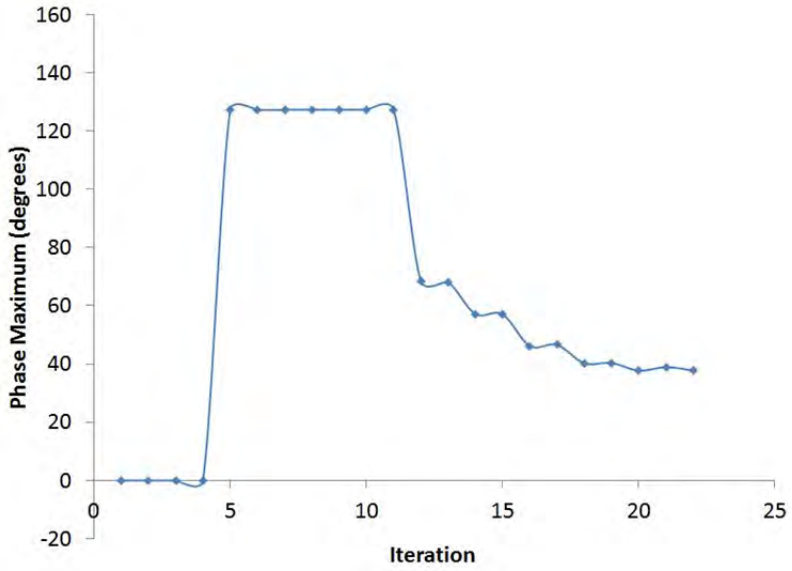

Figure 15 (b): Output from the network analyzer

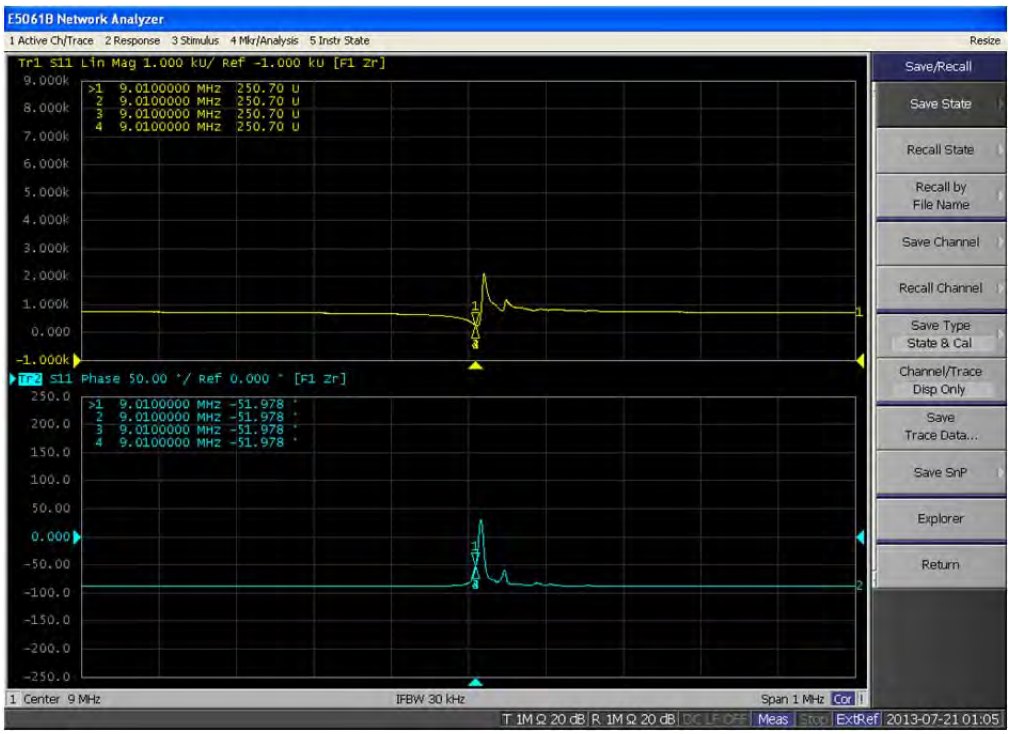

Figure 15 (c): Output from the network analyzer

DISCUSSION

Agilent's programming manual was used as a reference for programming in LabVIEW. In the experiment the size of the water droplets may not be controlled very precisely, thus, the response may not be consistent with the addition of each water droplet. However, the program developed is very precise and accurate in measuring whatever is inputted by the user. This program may also be tested with fluids of various viscosities as well as different sensors. The program will record precise and accurate results.

CONCLUSION

To demonstrate and verify the working of the program, water droplet loading experiment was performed. A 9MHz resonator was loaded with a controlled volume of water droplets while the impedance and phase shifts were recorded in real time. The first program records the data point of the curve maximum and the real-time change in a user determined specific time. The second program records the full curves within a frequency span was recorded and verified. The third program records the real time maximum of the full curves for each iteration within a frequency span. All the programs were developed successfully and its functionality and accuracy were tested and verified.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Rudolf F. Graf, “Modern Dictionary of Electronics, 7th Ed”, 1999: 162

[2] W.G. Cady, “Piezoelectricity, An introduction to the Theory and Applications of Electromechanical Phenomena in Crystals”, (New revised edition in two volumes), Dover Publications, Inc., New York, 1964.

[3] Walter L. Bond, John A. Kusters, “Making Doubly Rotated Quartz Plates” presented at the 31st Frequency Control Symposium, Atlantic City, June 1977.

[4] D. W. Dye, “The piezo-electric quartz resonator and its equivalent electrical circuit”, 1925 Proceedings of the Physical Society of London 38 399.

[5] Matthew A. Cooper, "Piezoelectric Sensors", Springer-Verlag Berlin Heidelberg, pp.100 August 11, 2006.

[6] Ferrari, V., Marioli, D., Taroni, A., "Oscillator circuit configuration for quartz-crystal-resonator sensors subject to heavy acoustic load," Electronics Letters, vol.36, no.7, pp.610,612, 30 Mar 2000.

[7] Institute of Electrical and Electronics Engineers, 1992 New York IEEE Std 488.2-1992.

ETCH RATE AND GEOMETRY FOR ETCHED Si(001) SUBSTRATES FOR SUBSEQUENT GaN FILM GROWTH

Sophia Williams,* Jarod Gagnon,[+] and Joan Redwing[#]

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

*Undergraduate Student of
Department of Engineering
Harvey Mudd College
Claremont, CA 91711

ABSTRACT

GaN on Si is desired as a semiconductor for power electronic applications. Commonly, GaN is grown on Si(111), which has a hexagonal arrangement of surface atoms similar to that of GaN. However, GaN does not grow well on Si(100) wafers because Si(100) does not have a hexagonal crystal structure. This study looks at the etching of Si(100) in order to expose the Si(111) planes for GaN growth. The etching is done using a solution of Potassium Hydroxide (KOH) and Isopropyl Alcohol (IPA). KOH is used because it is an anisotropic etchant that will etch the Si(100) at a much faster rate than Si(111) facets and thus reveal the Si(111) planes. The concentrations of the KOH and IPA, etching time, and etching temperature are varied. Etching rates are calculated by weighing the samples before and after each trial. The resulting pyramidal structure's lateral size and roughness is analyzed using optical microscopy. The etched Si(100) substrates are assessed for their suitability for the growth of GaN films.

INTRODUCTION

Gallium Nitride (GaN) on silicon is revolutionary semiconductor technology that is desirable because of the large band gap, large electron mobility, and high breakdown voltage of GaN [3]. It is cost efficient to grow GaN on silicon because of the comparatively low cost of silicon substrates. Commonly, GaN is grown on Si(111) substrates because the crystallographic orientation is similar to the arrangement of surface atoms on GaN. Since Si(100) is more commonly used in

[#] Faculty Mentor
[+] Graduate Mentor

devices, it would be optimal to grow GaN on these substrates. However, GaN films grown on these substrates tends to have multiple crystalline orientations because the arrangement of the surface atoms of Si(100) does not match the atomic arrangement of hexagonal GaN. Furthermore, the differences in lattice constants and thermal expansion coefficients negatively affect growth and often lead to cracking along the film's surface. This paper presents the optimal etching conditions of Si(100) in which the pyramidal structure of the substrates could be suitable for GaN film growth.

Anisotropic wet etching is a technique used to etch materials in order to expose other (*hkl*) planes [2]. This technique etches the Si(100) and Si(110) planes to expose the Si(111) planes. Usually etching is performed using a mask, however, without a mask the texturization of Si(100) for GaN growth could be more economically feasible. Thus, this study utilizes maskless etching of Si(100) in its experiments.

Etch rate and morphology of the substrate is not only affected by masking. The etching rate and total amount etched are also dependent on the temperature, the concentration of the etchant, and the duration of the etching process. Additionally, Isopropyl Alcohol (IPA), an organic alcohol, can slow etching rate when added to the etching solution [2]. These variables are considered in the investigations conducted in this study.

EXPERIMENTAL PROCEDURE

Every etch was performed using a 300-µm-thick Si(100) substrate with a resistivity of 5–10 Ω. Before etching, an organic cleaning step was performed on the substrate. Then the substrate was dipped in HF in order to remove the SiO_2 (native oxide) off of the sample before it was placed in the KOH solution. The concentrations of KOH and IPA were maintained throughout the etch using a reflux condenser. Unless specified the substrate was etched for 60 minutes. The etch rates of the substrates were calculated by weighing the samples before and after the etch.

Several variables, including IPA and KOH concentration, temperature, and etch time were varied. With a 5%wt solution of KOH in water, the temperature was varied from 80 °C to 95 °C in 5 °C intervals and experiments were conducted with the vol% of IPA at 3, 5, 6, 7, and 9%vol. Furthermore, the substrate was subjected to different etch times of 30, 60, and 90 minutes.

After etching, samples were analyzed using optical microscopy. Each sample's morphology was evaluated to optimize the quantity, size, and uniformity of the pyramidal structures along the surface of the substrate. The pyramid's square bases allowed them to be identified from other structures on the substrate. Pyramids were discerned from pits by raising the focal point of the optical microscope from the square base of the pyramid to the tip.

EXPERIMENT RESULTS

Variability Study

To determine whether outside variables could be influencing the etch rate and morphology of the sample during experimentation three experiments were conducted all under the same conditions. The Si(100) substrate was placed in a 5%wt KOH and 5%vol IPA solution at 90 °C for each trial.

It was found that the etch rate of the samples (Figure 2) had some variability although the substrates were etched under the same conditions for each trial. A standard deviation of 2.5×10^{-4} g/hr was found in the etch rate of the samples. Furthermore, the morphology of the samples (Figure 1) was not consistent throughout all trials. The most homogenous morphology was found in Trial 1 (Figure 1.A) in which many pyramids could be seen. The pyramids on this sample were of a uniform size and were comparatively constant along the surface of the substrate. In the second trial (Figure 1.B), pyramids could still be identified, however, they were fewer and had varying size. The last trial did not reveal any visible pyramids along the surface of the substrate (Figure 1.C).

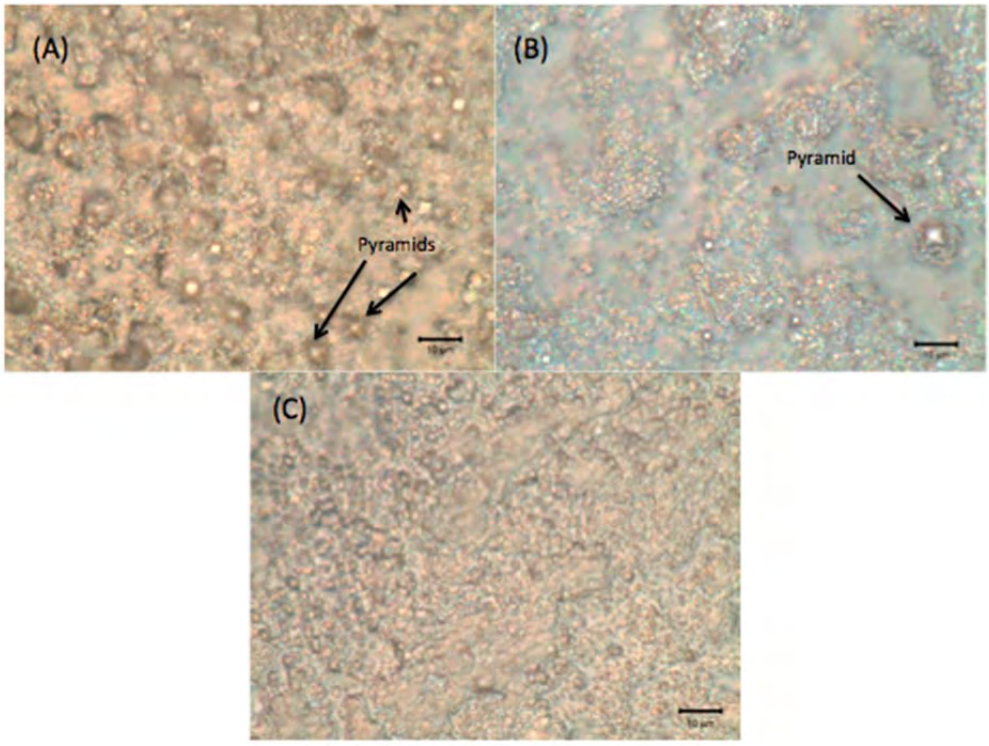

Figure 1: Optical microscopy images of multiple Si(001) wafers with similar etch conditions. Each of the images shown above were taken with a 100× zoom. The images show a section of the sample that is 80/7 µm by 60/7 µm. (A) Trial One, (B) Trial Two, and (C) Trial Three.

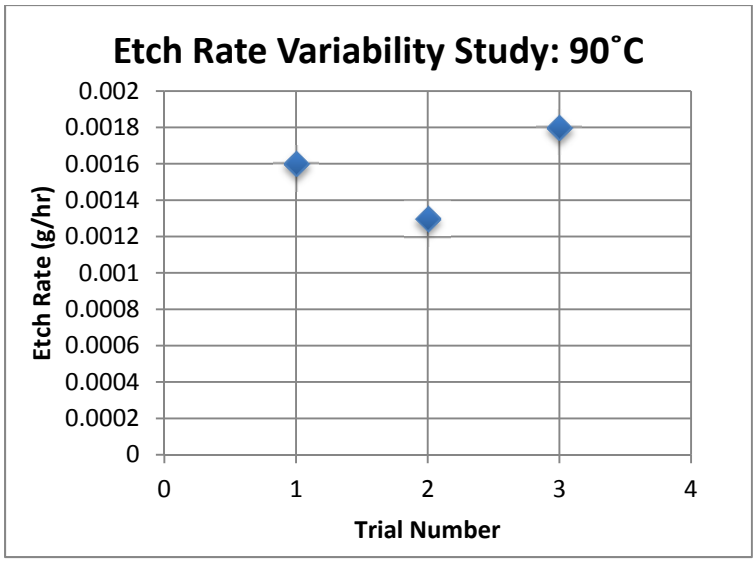

Figure 2: This graph shows the etch rate for all of the variability trials. The graph illustrates that the etch rate is not constant between trials.

Many factors could be contributing to the variability seen in the etch rate. The loss of reaction species, etchant impurities, substrate microstructure, and the amount of substrate already absorbed in the etchant could all be impacting the etch rate [2]. One condition that largely impacts etching conditions is how the mixture is stirred. Without stirring the reaction between the Si(100) and the KOH can change the localized temperature and KOH concentration around the sample. These variability trials were performed without any stirring implemented. Thus, further experiments were conducted to analyze the impact of stirring on the solution (See *Stir Bar Study*).

IPA Study

The IPA study was conducted by varying the volumes of IPA in a 50-mL solution and viewing the resulting etch rates and surface structures. The etches were performed with 5%wt KOH and mixing for 60 minutes. Overall, the substrates with the most uniform pyramid morphologies were found with 3%vol and 5%vol IPA solutions (Figure 3B, 3C). The surface of the substrate after being etched in 3%vol solution showed visible pyramids along the surface. Some of these pyramids were much larger than others, which indicated the pyramids on the substrate had inhomogeneous size. The sample etched in 5%vol IPA showed more uniform pyramids, however, they were much smaller in size compared to the pyramids from the 3%vol solution. In the sample with no IPA, pyramids were scattered across the surface and were largely varied in size. When the IPA %vol was increased to 6%vol and beyond, all pyramidal structures disappeared from the surface of the substrates. Instead, flat planes and pitting was visible along the face.

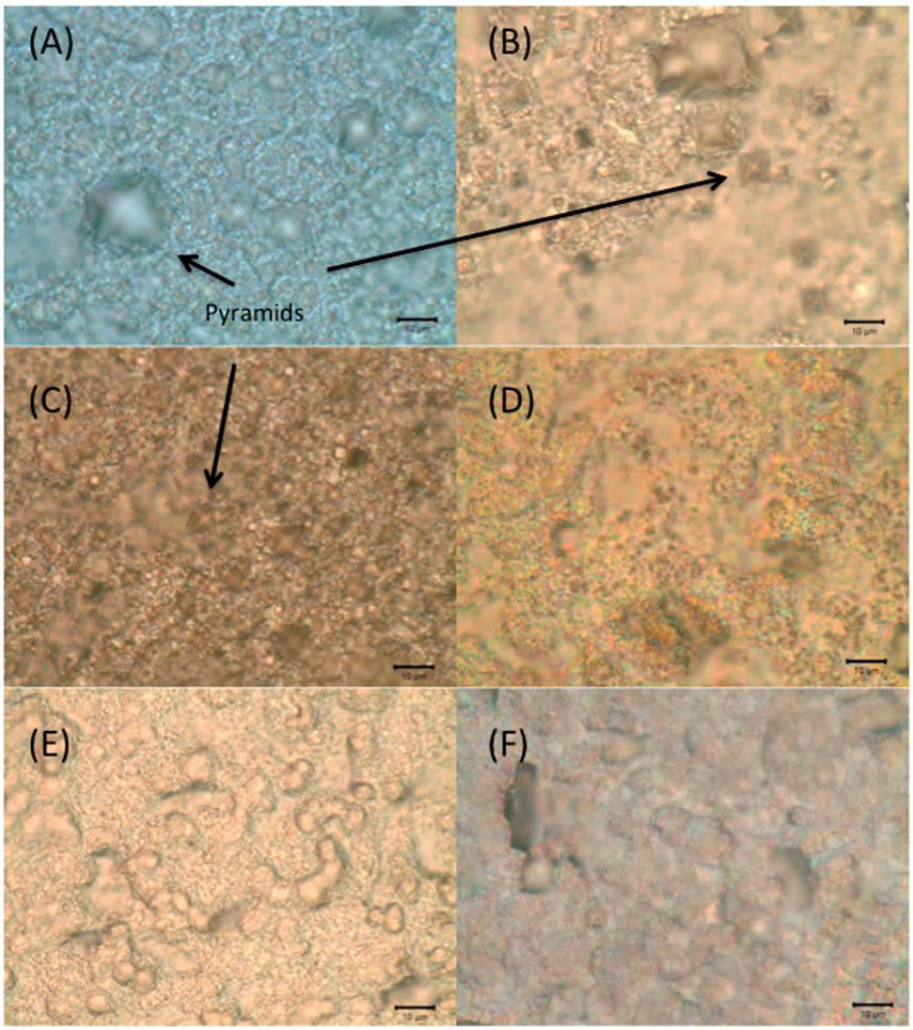

Figure 3: Optical Microscopy of images from IPA Study. These images show the substrates after being etched with no IPA in (A), 3%vol IPA (B), 5%vol IPA (C), 6%vol IPA (D) 7%vol IPA (E), 9%vol IPA (F).

The etch rate of the samples decreases asymptotically as the volume of IPA increases (Figure 4). Thus, the morphologies with pyramids occur with higher etch rates. However, to get achieve uniformity along the surface a lower etch rate is necessary.

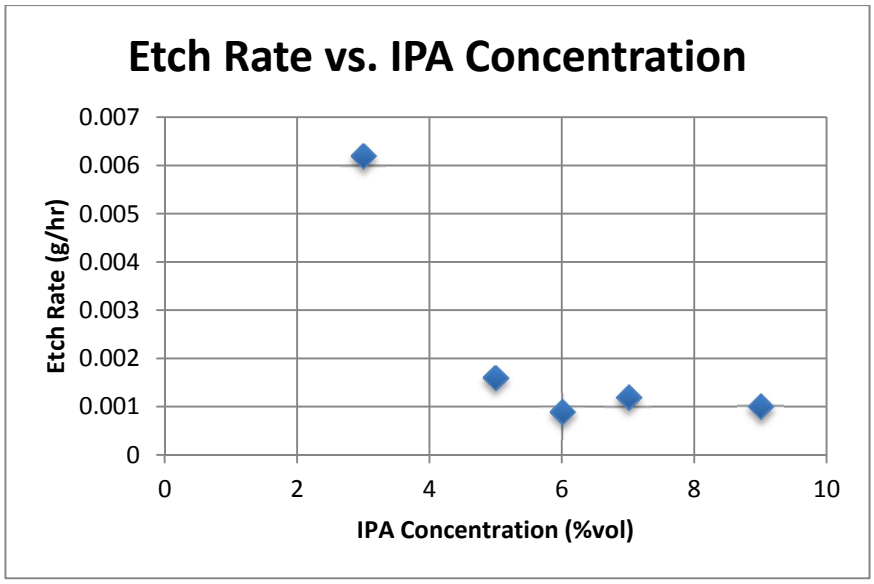

Figure 4: This graph shows the impact of IPA concentration of etch rate. It shows that the etch rate asymptotically decreases as IPA concentration increases. Note that the etch rate with no IPA is unavailable.

Etch Time Study

Etch time is a factor that plays a large role in etch rate [2]. In this study, etch rate was measured and the morphology was analyzed for substrates that were etched for 30, 60, and 90 minutes. These substrates were etched in a solution with 5%wt KOH, 5%vol IPA at 90˚C with no mixing.

After being etched for thirty minutes the sample contained several small pyramids along its surface (Figure 5A). However, the pyramids were not as large or as consistent as the sample etched for 60 minutes (Figure 5B). The substrate etched for 90 minutes had a smooth surface with pits (Figure 5C). Furthermore, the etch rate for the sample etched for 90 minutes was much less than the etch rate of the sample etched for 30 and 60 minutes (Figure 6). These inconsistencies in etch rate could be because the solution was not mixed during etching.

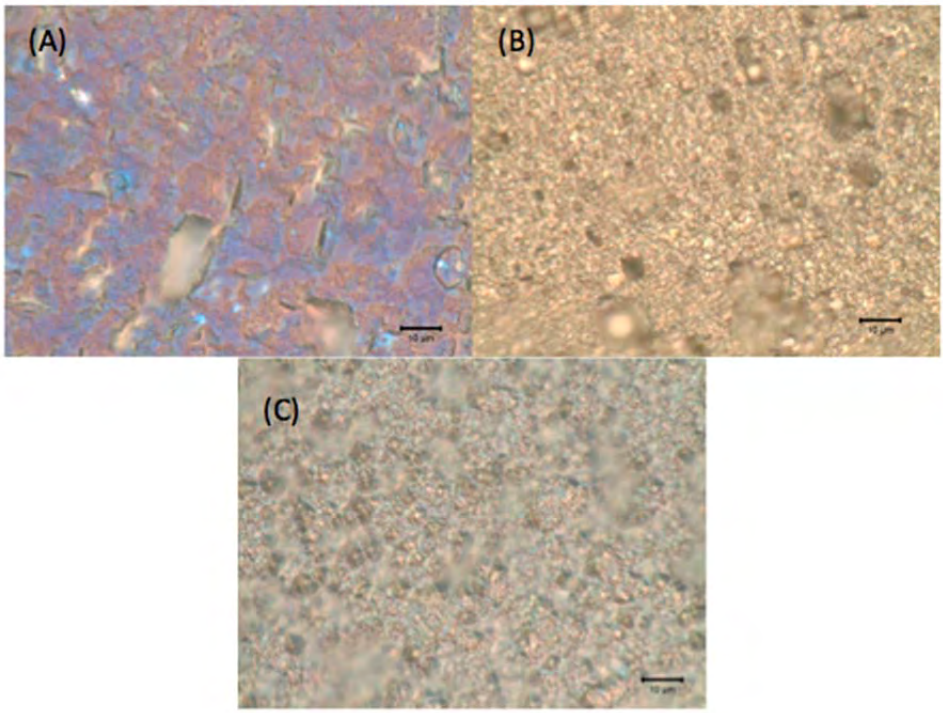

Figure 5: Optical microscopy of samples after different etch times. (A) Image of the substrate after being etched for 90 minutes**.** (B) Image of the substrate after being etched for 60 minutes. (C) Substrate after being etched for 30 minutes.

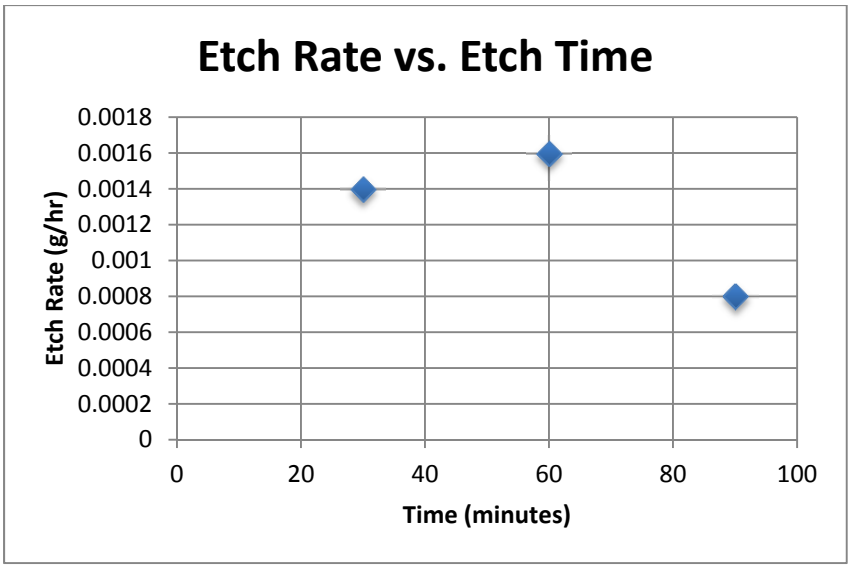

Figure 6: This graph shows the effect on etch rate as one varies the etch time.

Temperature Study

The temperature of the etchant was varied in order to determine temperature's effect on the etch rate of the Si(100). The trials were performed with 5%wt KOH solution with 5%vol IPA and no stir bar. The results showed that 90 °C (Figure 1) had the most pyramidal structures and was the most suitable for GaN growth. The trials performed at 80 °C and 85 °C both had very low etch rates and the surface morphologies looked very smooth (Figure 7.A, 7.B). Some pits and ridges are visible but no pyramids could be identified. The trial at 95 °C shows raised surface structures. However, these structures lack the square base and other native characteristics of pyramids.

The etch rate of the sample increases linearly until 95 °C where a large drop in the etch rate is visible (Figure 8). Munoz et al. found that the etch rate increased with temperature until 85 °C when the temperature ceased to effect the etch rate [4]. The study did not test the etch rate at temperatures higher than 90 °C. The results of this study contradict those results and show a constant increase until 95 °C. The drop in etch rate at 95 °C could be due to a reaction occurring in the KOH solution as it moves closer to its boiling point. This could result in less of the etchant to react with the Si(100) planes on the substrate.

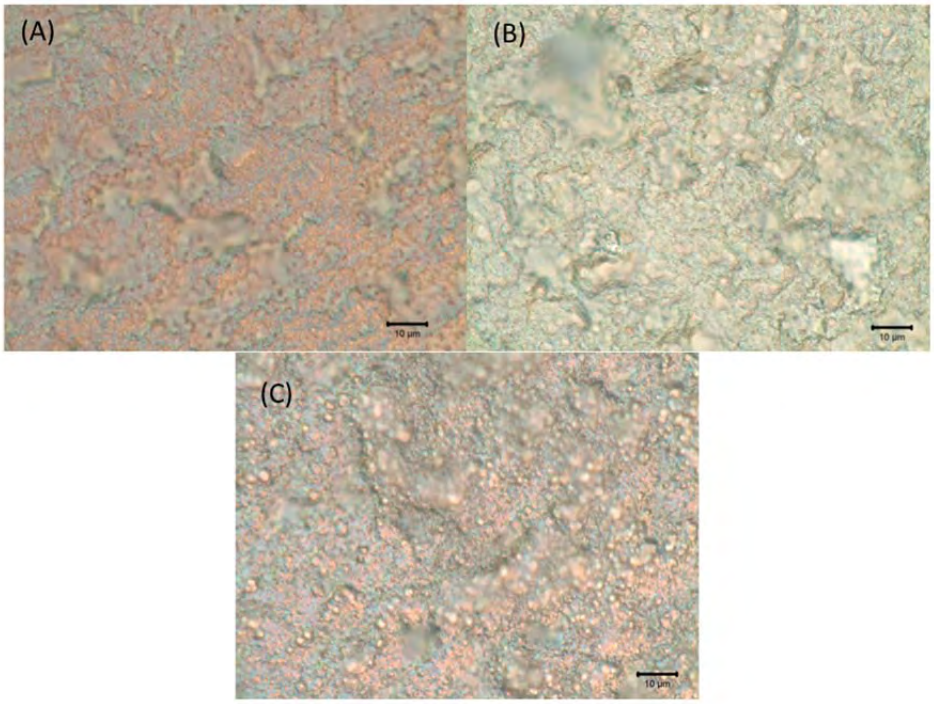

Figure 7: Optical Microscopy of samples after being etched at different temperatures. (A) Sample after being etched at 80 °C. (B) Sample after being etched at 85 °C. (C) Sample after being etched at 95 °C.

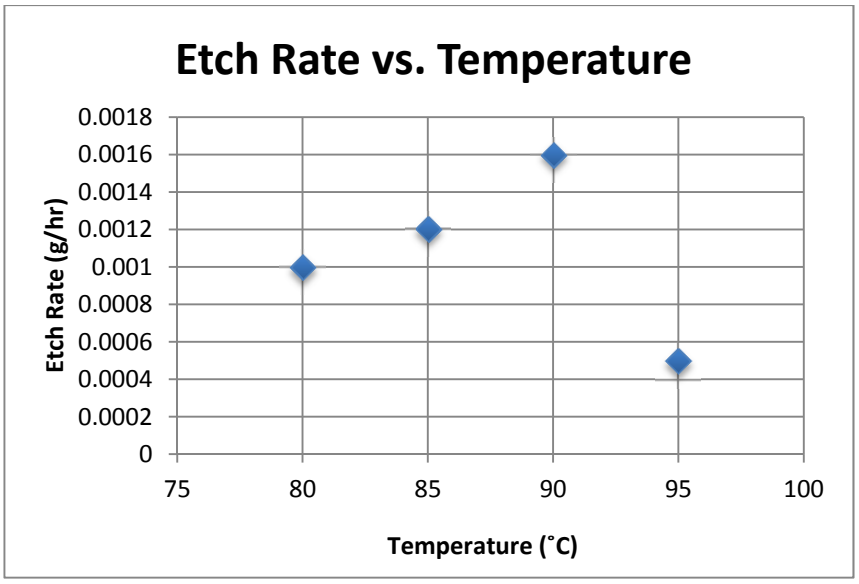

Figure 8: This graph shows the effect of temperature on etch rate at four different temperatures from 80 °C to 95 °C. The results show that there is a constant increase in etch rate until 95 °.

Stir Bar Study

A stir bar study was conducted to determine the effect of mixing on the etched sample. The sample needed to be isolated from the stir bar in order to insure that the substrate was not damaged during the process. Therefore, the magnetic stir bar was placed into a Teflon beaker with small holes along its sides before being placed into the solution. Since it was desired to test the variability of the etching with the stir bar, three trials were completed. Once again, the Si(100) substrate was etched in a 5%wt KOH and 5%vol IPA solution at 90 °C for each trial.

The substrate morphologies with the stir bar at 90 °C and 5%vol IPA are potentially suitable morphologies for GaN growth. The morphologies of the samples (Figure 9) are more consistent from trial to trial with the stir bar. In the first trial (Figure 9A) large uniform pyramids can be seen along the entire surface of the substrate. The third trial had a final structure similar to the first (Figure 9C). The size of the pyramids is smaller but they pyramids are of constant size and cover the sample's surface. The second sample would not be good for GaN growth because the morphology shows fewer and smaller pyramids (Figure 9B). Furthermore, the pyramids are not homogeneously situated across the sample surface.

Although the morphologies of the samples improved with the stir bar, the etch rate variability increased (Figure 10). The etch rate with the stir bar has a standard deviation of 5.1×10^{-4} g/hr compared to the 2.5×10^{-4} g/hr without the stir bar. However, the average etch rate also increased by 63% with the stir bar (Figure 10).

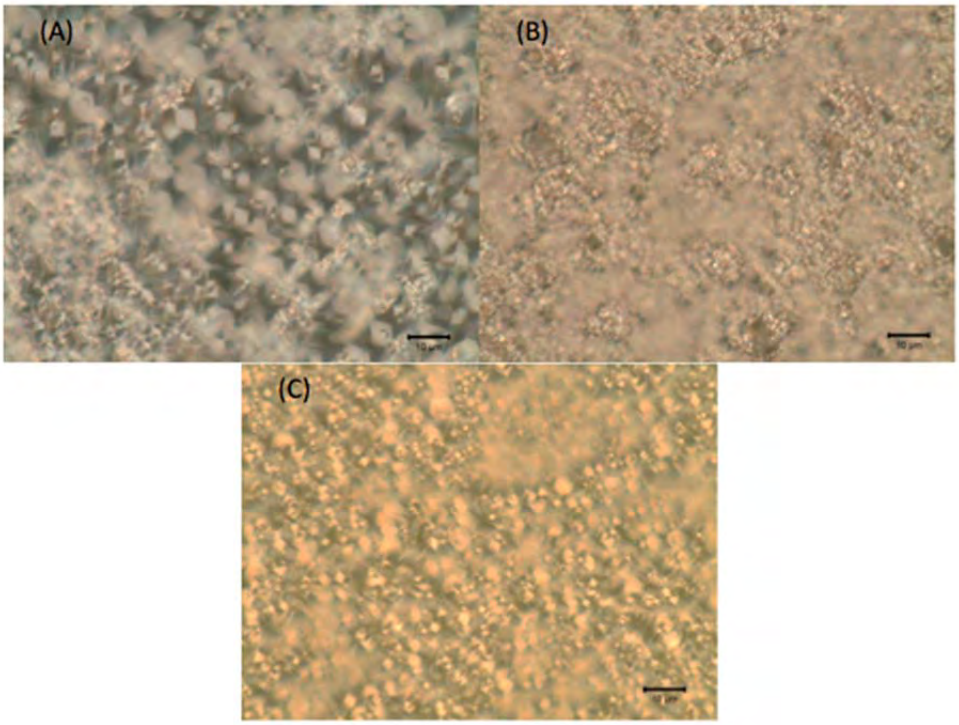

Figure 9: Optical microscopy images of multiple Si(001) wafers with similar etch conditions. (A) Trial one, (B) Trial two, and (C) Trial three.

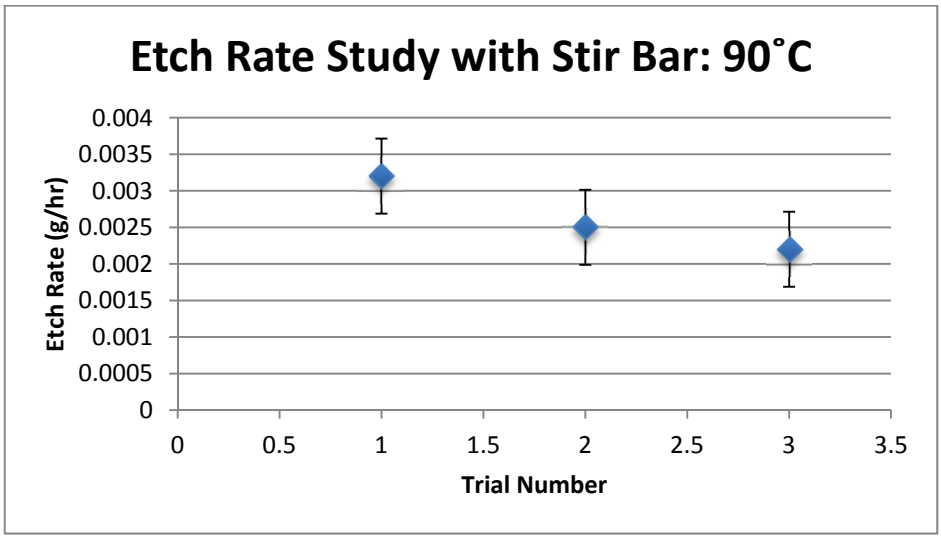

Figure 10: A graph of the etch rate of three samples etched under the same conditions. The etch rate is not constant. However, the etch rate is almost double the results found without the stir bar.

KOH Study

Temperature and IPA both played an important role in effecting the etch rate and morphology of Si(100). However, the impact of the concentration of KOH on the etching of the silicon still needed to be investigated. Thus, two additional etches were performed with 7.7%wt and 12%wt KOH solutions. The solutions were 5 %vol IPA and all tests were implemented at 90 °C with a stir bar for mixing.

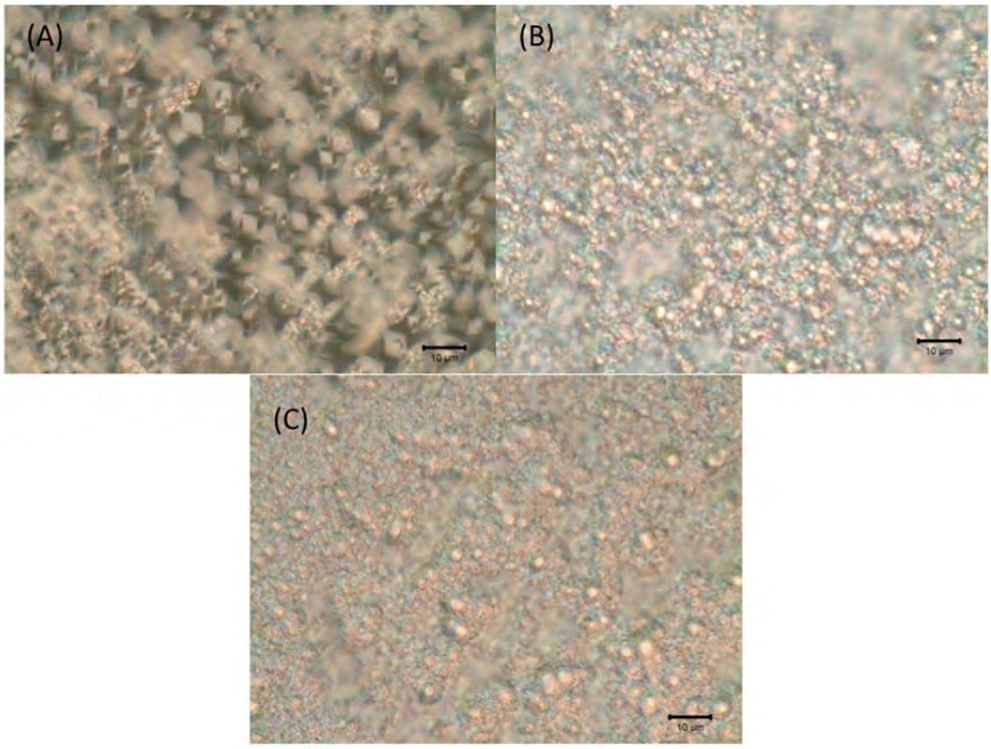

Figure 11: This figure shows a 100× magnification optical microscopy image of the substrate after it was etched in 5%wt KOH solution (A), 7.7%wt KOH solution (B) and 12%wt KOH solution (C).

The 7.7%wt KOH solution and the 12%wt KOH both had small structures along the surface, however, these structures did not have the characteristics of square-based pyramids (Figure 11). It is important to note that the ridges were denser on the surface of the sample etched in 7.7%wt KOH (Figure 11B) which shows that the surface ridges decrease as KOH concentration increases. Thus, higher concentrations of KOH appear to result in smoother substrates. However, the etch rates of the substrates decrease with the increasing concentration of KOH. Thus, even though the surfaces of the substrates appear smoother with higher concentrations of KOH, less silicon has been etched off of the substrates. This trend is contradictory to what previous research has found [3]. The result

could be due to some of some of the other variables that effect etch-rate value like etchant impurities, loss of reaction agent, etc.

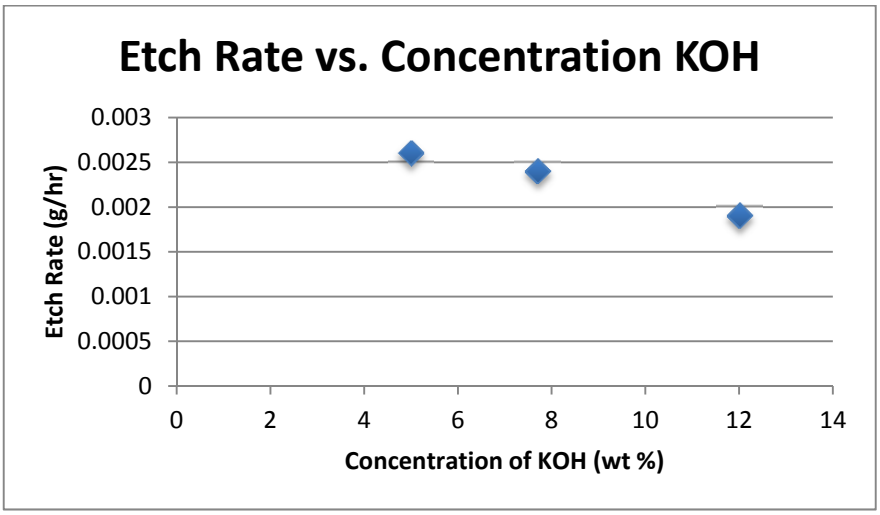

Figure 12: This graph shows the change in etch rate due to a change in the concentration of KOH in the solution. It shows a decrease in etch rate with an increase in KOH concentration.

DISCUSSION

The optimal morphology for GaN growth is a surface with large and uniform pyramids across the entire face of the substrate. It was found that large etch rates and small concentrations of KOH resulted in pyramids on the substrate. However, uniformity only occurred with lower etch rates. Thus, a balance had to be found in order to produce substrates with the appropriate structure. The results of this study reveal ways that variables can be adjusted in order to achieve appropriate morphologies.

When IPA was added, the etch rate decreased but the uniformity of the pyramids increased. However, too much IPA decreased the etch rate such that no pyramids were visible along the surface. Additionally, adding the stir bar greatly improved surface morphology and increased etch rate. Based on these results, optimal pyramidal structures can be found at low concentrations of KOH and IPA. Higher temperatures are preferable for good surface structures. However, it is important to acknowledge that temperatures above 90 °C cause low etch rates. Throughout these experiments, the best morphology was found when using 5%vol IPA, 5%wt KOH, 90 °C, and a stir bar.

CONCLUSION

In this study, the effect of different conditions on the anisotropic etching of Silicon(100) was analyzed and the resulting substrates were assessed for GaN growth. The results showed that the best pyramids and surface homogeneity occurred at low concentrations of KOH and IPA and at high temperatures. It was

also found that mixing the solution during the experiment greatly improved the etch rate and morphology of the samples.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984 and DMR-1006763. Furthermore, S.W. would like to thank the Penn State EE REU for giving me the opportunity to do research at the university.

REFERENCES

[1] Bean, K.E., and K.E. Bean. “Anisotropic Etching of Silicon.” IEEE Transactions on Electron Devices 25.10 (1978): 1185-193. Print.

[2] Gosalvez, M.A., I. Zubel, and E. Viinikka. “Wet Etching of Silicon.” Ed. V. Lindroos. *Handbook of Silicon Based MEMS Materials and Technologies*. Amsterdam: William Andrew/Elsevier, 2010. 375-407.

[3] Li, III-V Compound Semiconductors: Integration with Silicon-Based Microelectronics, ed. T. Li, M. Mastro, and A. Dadgar (CRC Press, 2011), p. 51.

[4] Muñoz, D., P. Carreras, J. Escarré, D. Ibarz, S. Martín De Nicolás, C. Voz, J.M. Asensi, and J. Bertomeu. “Optimization of KOH Etching Process to Obtain Textured Substrates Suitable for Heterojunction Solar Cells Fabricated by HWCVD.” Thin Solid Films 517.12 (2009): 3578-580. Print.

[5] Zubel, Irena, Krzysztof Rola, and Małgorzata Kramkowska. “The Effect of Isopropyl Alcohol Concentration on the Etching Process of Si-Substrates in KOH Solutions.” Sensors and Actuators A: Physical (2011): 436-445.

[6] Zubel, Irena, Malgorzata Kramkowska, and Tomasz Ninierza. “Methods of Silicon Surface Structurization for the Purpose of the Deposition of III-V Epitaxial Layers.” Optica Applicata 4th ser. 29 (2009): 749-59.

AUTOMATED CONTROL OF A G-FRESNEL SPECTROMETER

Bill Kim,* Perry Edwards[+], Ding Ma[+] and Zhiwen Liu#

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Johns Hopkins University
Baltimore, MD 21218

ABSTRACT

This work explores the development of a low-cost, high-performance, miniaturized spectrometer and creation of a mobile spectrometer interface. The spectrometer is constructed using a G-Fresnel diffractive optical element, which represents the combination of the focusing function of a Fresnel lens and the wavelength dispersive capability of a diffraction grating into a single device. The G-Fresnel is capable of maintaining high spectral resolution in a miniaturized package which enables a myriad of applications once unattainable with traditional, bench top spectrometers. To facilitate usability on a mobile platform, a graphical user interface was developed and programmed to communicate with the hardware interface of the G-Fresnel spectrometer. An android application was created to control the functions of the spectrometer and provide a convenient, straightforward interface for the user and cross platform compatibility. The performance of the G-Fresnel spectrometer was evaluated in comparison to a commercial, high performance spectrometer by comparing several filter measurements performed with each device.

INTRODUCTION

Spectrometers

Optical spectroscopy is a widely used technique for providing quantitative analysis of light-material interaction across many scientific and engineering disciplines. For example, spectrometers are used in the environmental sciences to measure water quality; the medical realm for biopsies and diagnoses; and the

[#] Faculty Mentor
[+] Graduate Mentor

manufacturing arena for color measurement. As the demand for spectrometers continues to expand in a plethora of fields, the necessity of a miniature mobile spectrometer, with an easily accessible interface, is growing rapidly. Conventional spectrometers are often bulky and will typically include multiple components. For example, a standard Czerny-Turner style spectrometer requires an entrance slit, focusing mirror, reflection grating, collimating mirror, and then an imaging sensor to collect the data [1]. The task of implementing five discrete optical elements to create a high-resolution spectrometer tends to lead to a bulky and costly piece of equipment. Such an advantageous piece of equipment would ideally be miniaturized and its cost reduced for portability and dissemination.

Development of miniaturized spectrometers

Significant progress has been made in recent years to miniaturize and reduce the cost of spectrometers. Many have focused on miniaturizing the traditional spectrometer design by reducing the size of the components and implementing a more compact design. A common approach has been to design optical elements that can mimic properties of two or more of the discrete optical elements inside a traditional spectrometer. Curved gratings [2] and volume holograms [3] are examples of such devices that can combine functions of multiple elements. However, there exists a limit to the amount of miniaturization that can occur due to geometric constraints. Thus, several efforts to scale down high performance spectrometers have included implementing MEMS-based techniques to create miniaturized spectrometers in silicon, however their resolution performance is limited [4]. Another technique utilizes waveguide grating couplers to achieve miniaturization of a spectrometer [5].

Challenges

The biggest obstacle when downsizing a spectrometer is ensuring the spectral resolution and sensitivity does not suffer on account of the miniaturization. The entrance slit width and groove frequency (the amount of grooves per millimeter on the diffraction grating), are two of the major factors used in determining the spectral resolution of a spectrometer as well as impacting the sensing bandwidth or the spectral range. The entrance slit introduces a tradeoff between the amount of light entering the system versus achievable resolution. A larger entrance slit provides more throughput of light, but worse resolution and vice versa. A higher groove frequency lends itself to a high spectral resolution, but the spectral range must be sacrificed. Clearly, when the goal is to shrink the overall dimensions of a device there will be less surface area on the diffraction grating so the number of grooves, and therefore the spectral resolution, will suffer.

A G-Fresnel Spectrometer

A relatively new, hybrid, diffractive optical element, coined the G-Fresnel, fuses the properties of a grating and a Fresnel lens together to combine the traditional spectrometer element functions of collimation, dispersion, and

collection into a single thin-film optical element [6]. A premier characteristic of the G-Fresnel device is its compact size which can be attributed to its ability to maintain a low f-number as it is reduced in size. Simulations for a transmission type G-Fresnel spectrometer has been analyzed both theoretically and experimentally and show a spectral resolution of around 1 nm can be achieved with a G-Fresnel with a diameter of 4 mm and a focal length of 1 cm [6].

Theory behind a G-Fresnel device

To determine the transmission, or reflection, coefficient of the G-Fresnel, the transmittances of both a Fresnel lens and a linear diffraction grating are multipled together,

$$t(x,y) \propto \eta(\lambda) e^{-j\frac{\pi}{\lambda F}(x^2+y^2)} e^{-j\frac{2\pi}{\Lambda}x}, \quad (1)$$

where λ is the wavelength, F is the focal length, Λ is the grating period, and η is the diffraction efficiency of the device. Thus, the dual functionality of focusing and dispersion is achieved. Using the Fresnel diffraction formula and a paraxial approximation, the field distribution of the G-Fresnel is discussed in detail in [7].

Experimental Description

In this experiment, a reflection configuration G-Fresnel was used (R-G-Fresnel), while a transmission configuration is demonstrated previously [6]. Polydimethylsiloxane (PDMS) soft lithography was used to create the G-Fresnel. A detailed fabrication procedure of the PDMS molded G-Fresnel can be found in [7]. Briefly, a PDMS, pre-polymer was poured onto an existing Fresnel lens. A negative Fresnel mold was obtained after curing. Then the negative mold is aligned with a diffraction grating and more PDMS pre-polymer was inserted between the mold and grating to obtain the G-Fresnel. The G-Fresnel was then treated with a METROLINE M4L Plasma Etcher, for 16 seconds and the grating side was coated with Galinstan, which is known for its high reflectivity. With the reflection configuration the beam passes through the G-Fresnel surface pattern twice, effectively "folding" the optical path length which will reduce the focal length by half. With a shorter focal length the size of the spectrometer setup can then be significantly reduced.

Experimental overview

Using the R-G-Fresnel, a spectrometer prototype was created that was interfaced with mobile wireless technology to increase ease of use of the device. Automated control of the device will be achieved through an android application to provide the convenient interface between users and the spectrometer. This type of automated control on a cell-phone or other mobile devices is extremely practical because field data can be taken without the cost and weight of transporting heavy spectrometers and lab equipment. Thus the concept of a mobile spectrometer is truly realized. The spectrometer will be tested for

performance in regards to resolution and sensitivity compared to a commercial spectrometer with several filters, to demonstrate the performance, mobility, and standalone ability of the device.

The G-Fresnel spectrometer was constructed using three optical components: an entrance slit, the R-G-Fresnel, and a linear imaging sensor. An image of the setup is given in Figure 1. An optical fiber was attached to the 390 μm entrance slit to couple the light into the system. The optical signal will traverse from the entrance slit to the R-G-Fresnel then reflect towards the linear imaging sensor. The linear imaging sensor was placed directly adjacent and nearly perpendicular to the R-G-Fresnel due to the reduced focal length. The linear imaging sensor was placed to the side of the R-G-Fresnel and angled along the locus of foci located on a single line [7] to capture all of the foci. The linear imaging sensor used in the experiment was the Mightex TCN-1304-U CCD camera model. This camera has 3,648 pixels and can be interfaced with through a USB 2.0 connection. Each pixel value represents the intensity of light on that pixel and can range from 0 to 65,536. The minimum and maximum exposure times for the camera are 0.1 ms and 6,500 ms, respectively.

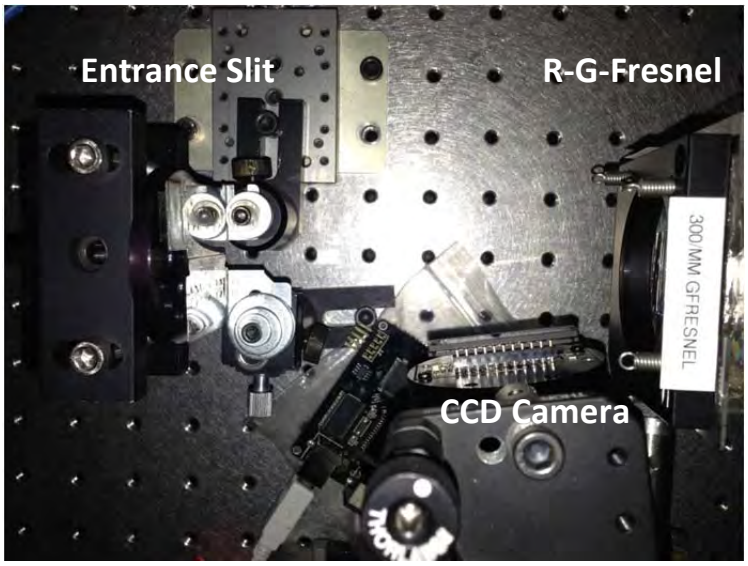

Figure 1: R-G-Fresnel Spectrometer Setup.

Wavelength Calibration

To calibrate the G-Fresnel spectrometer an Argon Ion Laser (Melles Griot 532) with known wavelengths between 457 and 514 nm was directed into the system and the resulting spectrum was obtained [Fig 2]. Two spectrums at different exposure times were needed to uncover all of the argon ion laser lines necessary to create the calibration curve. At the exposure time that all the peaks were visible, 1.2ms, the 488nm line and the 514 nm line were over-saturated due to the fact that they are at least 10 times stronger than the remaining lines [8]. Therefore a lower exposure time, 0.3ms, was also used to discover the unsaturated

pixel values of the peaks. The pixel values of the 8 peaks, specified on the Melles Griot spec sheet [8], were then found and a best fit curve was constructed using the known laser lines of an argon ion laser [Fig 3]. The pixel-wavelength relation was then found to be a cubic function and used to determine the wavelengths of spectrums gained from the G-Fresnel.

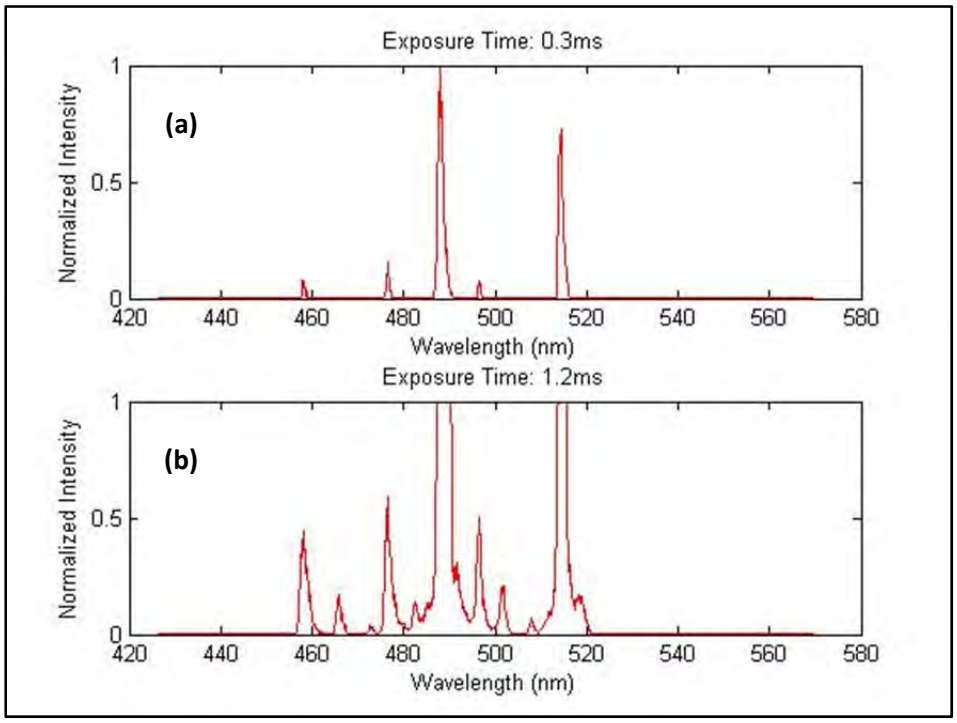

Figure 2: Argon Ion laser lines from R-G-Fresnel Spectrometer at (a) Exposure time 0.3 ms (b) Exposure time 1.2 ms.

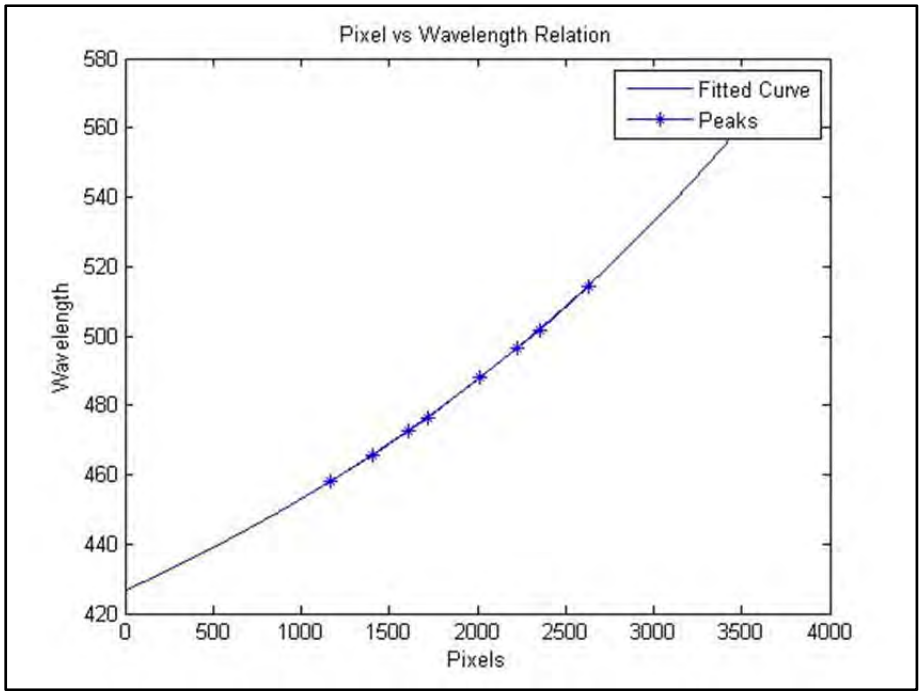

Figure 3: Calibration curve for pixel-wavelength relation.

Performance comparison procedure

After the calibration curve was measured and the pixel-wavelength relation was found several other measurements were taken with the R-G-Fresnel spectrometer. A white light lamp source (ROI 150 Illuminator) was directed onto 6 separate filters: Thor Labs FL532-3 laser line filter, Thor Labs FL488-10 laser line filter, Chroma HQ457/56x bandpass filter, Chroma D455/70X bandpass filter, Chroma E450 shortpass filter, and a 450 nm bandpass filter. Each filter's resulting spectrum was then recorded using the spectrometer system. The white light intensity coming into the spectrometer system without any filter was set to 2.43 mW with a power meter set at 532 nm. The slit width of the setup was 390 μm and the G-Fresnel had a groove frequency of 300 grooves/mm. To analyze the performance of the R-G-Fresnel spectrometer the same filter measurements were taken with the PI/Acton SpectraPro 2500i commercial spectrometer with liquid nitrogen cooling. The slit width was set at 50 μm and a grating groove frequency of 150 grooves/mm and blaze wavelength of 500 nm was chosen. The system was mimicked by aligning a fiber in front of the tunable entrance slit of the commercial spectrometer. Using the same white lamp source the filter measurements were once again conducted, however, the white light intensity recorded coming through the fiber, into the system, was reduced to 0.57 μW due to the commercial spectrometer's superior sensitivity. Each spectrometer measured the spectrum from each filter at three separate exposure times; the exposure time for when the spectrum's intensity first broke the noise floor (value greater than 200), at 100 ms and 150 ms.

Android application overview

The automated control of the R-G-Fresnel spectrometer was achieved through an android applet. A block diagram of the applet is given in figure 4. Because of the specifications of the software the android powered device must have android 3.1 or higher installed in order for the application to run properly. After the user connects the USB device the user can request data from the CCD camera. Once the applet initiates communication and requests data it will continually query data from the camera until the user stops it or the on board camera buffer fills up because an error occurred. The camera will send back the raw pixel data in little endian format, which is then converted to a meaningful integer. While the data is being continuously fetched, the user can choose to graph the data at any time. The graph will not update by itself (i.e. it will not graph in "real time"), however, because the applet continuously pulls data from the camera the applet will store data in "real time". All of the essential spectrometer control options to capture a wide range of measurements is provided by the applet. The exposure time can be set from 0.1 ms to 6,500 ms in steps of 0.1ms. Each axis can be scaled and the applet will graph a maximum of 5 overlay spectrums at once. All of the applet's raw data can be exported to the android powered device's external storage to a file of the user's liking, which can then be accessed by connecting the device to the computer. A screenshot of the application collecting data is given in Figure 5.

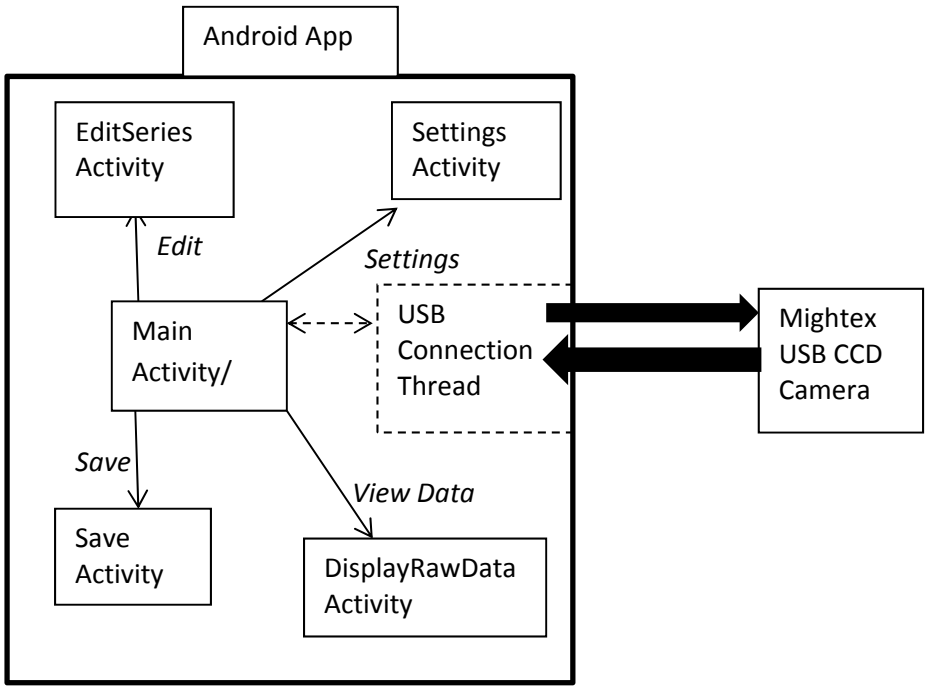

Figure 4: Block diagram of Android application.

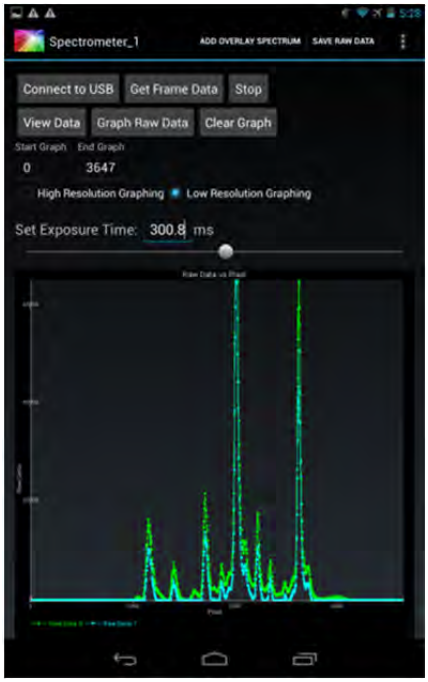

Figure 5: Screenshot of Android application showing the acquired spectra.

EXPERIMENT RESULTS

Several spectra were obtained using the conditions described previously. The R-G-Fresnel spectrometer and commercial spectrometer were each set at an exposure time of 100 ms, and several measurements with a variety of optical filters was performed. In each case, the spectra of the 100 ms exposure time measurement were normalized to enhance the comparison (see Fig. 6). Furthermore the exposure time's at which each spectrometer broke through the noise (intensity greater than 200) are given in Table 1.

Table 1: Exposure times when spectrum breaks through noise for each filter.

	R-G-Fresnel Spectrometer	Commercial Spectrometer
532 Laser Line Filter	59 ms	5 ms
488 Laser Line Filter	30 ms	< 5 ms
455 Band Pass Filter	18.6 ms	< 5 ms
457 Band Pass Filter	21.3 ms	< 5 ms
450 Short Pass Filter	65 ms	5 ms
450 Band Pass Filter	27.1 ms	< 5 ms

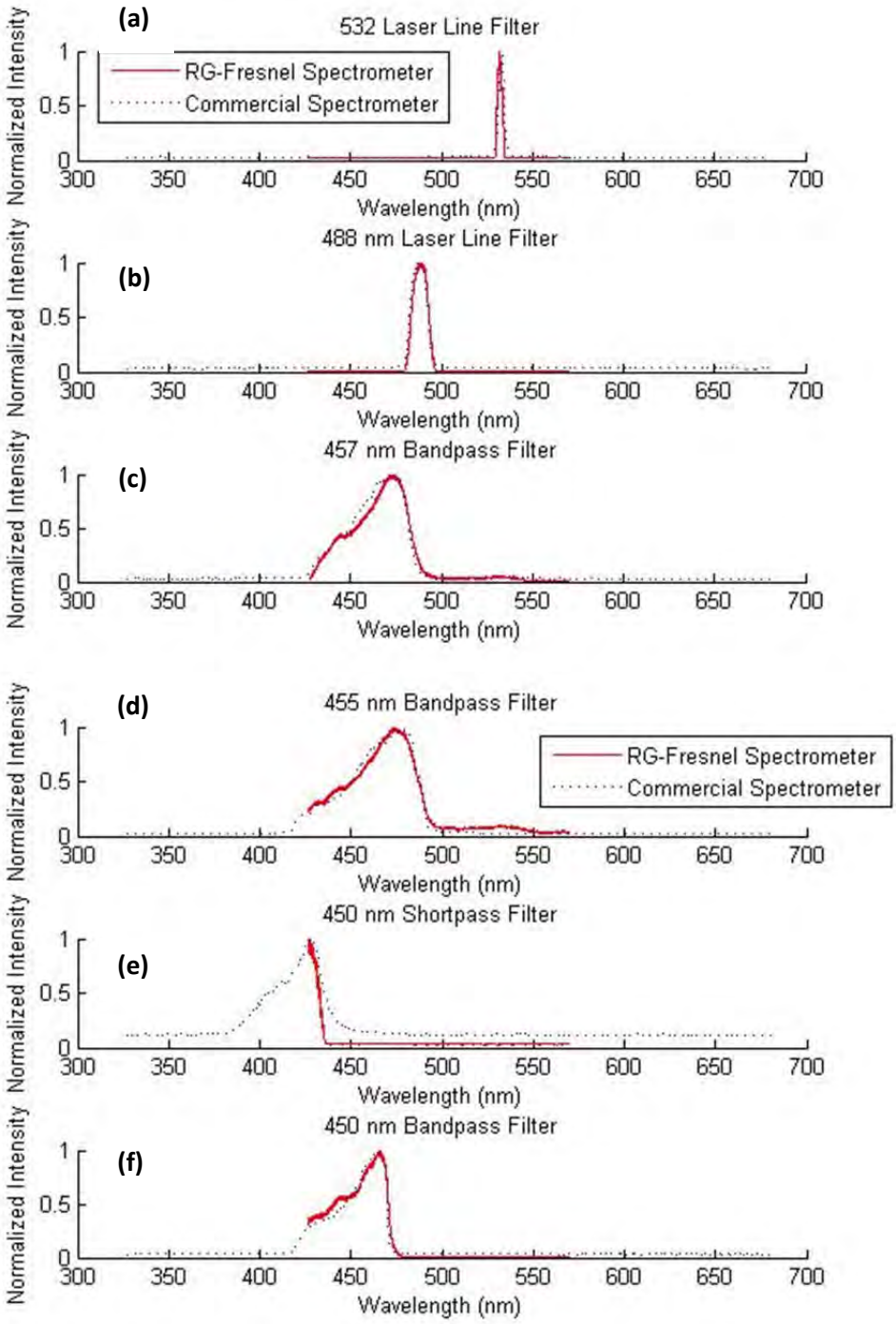

Figure 6: Measured, normalized spectrums of different filters. (a) Thor Labs FL532-3 laser line filter, (b) Thor Labs FL488-10 laser line filter, (c) Chroma HQ457/56x bandpass filter, (d) Chroma D455/70X bandpass filter, (e) Chroma E450 shortpass filter, (d) 450nm bandpass filter.

DISCUSSION

In Fig. 6, one can observe the measured spectrum of the R-G-Fresnel spectrometer correlates well with the commercial spectrometer within the spectral range of the R-G-Fresnel spectrometer. The spectral range of the R-G-Fresnel spectrometer is limited to ~140 nm (430-570 nm). The system resolution of the G-Fresnel spectrometer is determined by measuring the full-width-half-maximum (FWHM) of narrow atomic emission lines or laser spectrum. For this setup, the R-G-Fresnel spectrometer's resolution was determined to be 1.4 nm, which was the measured FWHM of the 476.5 nm Argon Ion laser line. Thus, near 1 nm spectral resolution was achieved with the R-G-Fresnel configuration. Further optimization of the system (alignment, optic quality, etc.) could improve resolution quality. Additionally, an improved design of the G-Fresnel element has been proposed that can have improved resolution (near 1 nm) and much smaller dimensions [6]. With respect to the laser line filters show in Fig. 6a and Fig. 6b the R-G-Fresnel spectrometer performed very well. Specifically, with the 532 laser line filter the R-G-Fresnel spectrometer measured a FWHM of 3.3 nm and the commercial spectrometer measured a FWHM of 3.6nm. For the 488 laser line filter a FWHM was measured to be 10 nm and 10.5 nm for the R-G-Fresnel spectrometer and commercial spectrometer, respectively. In regards to the bandpass filters (see Fig. 6c, 6d, and 6f), the limited range of the R-G-Fresnel is very obvious, extending to 430 nm. The roll off for the 450 nm short pass filter (see Fig. 6e) occurs much more sharply for the R-G-Fresnel, however, it drops off rapidly before it reaches 450 nm.

From the table comparing exposure times when each spectrometer was first able to measure a spectrum that was not noise, it is clear that the commercial spectrometer is much more sensitive than the R-G-Fresnel spectrometer. This is expected, as the commercial spectrometer implements liquid nitrogen cooling to reduce the dark current and improve sensitivity. Additionally, the input power of the white light illuminator was set to four orders of magnitude greater for the R-G-Fresnel spectrometer, highlighting even more the low sensitivity of the system in comparison to the commercial spectrometer. The low sensitivity of the R-G-Fresnel system is partly attributed to the setup of the system, because the grating order used (third order) was not optimized for the operating wavelengths, and only a small portion of the signal at that order was collected. The grating used to create the G-Fresnel was optimized for the first order; however, this order could not be reached due to the dimensional constraints of the CCD linear array. Also, it was noticed the R-G-Fresnel produced a significant amount of stray light. This can be attributed to defects in the mold and coated reflection surface. In this configuration, the overall efficiency of the system was drastically reduced, making it difficult for the CCD camera to pick up low intensity signals.

The results garnered from the comparative measurements signal that there exist significant room for improvement in future device design and fabrication. Many of the working specifications can be improved upon; the spectral resolution, spectral bandwidth, and sensitivity can all be optimized for particular

applications. The G-Fresnel spectrometer can be improved with an optimized Fresnel lens, grating pattern and blaze angle. A better alignment for the CCD detector could also improve sensitivity of the system as well. Considering the small focal length of the device in comparison to the commercial spectrometer, 50 mm versus 500 mm, respectively, and the performance of the proof-of-concept R-G-Fresnel spectrometer versus the commercial spectrometer, the G-Fresnel opens the possibility for creating a high performance, low cost miniature spectrometer.

CONCLUSION

A comparison of a prototype G-Fresnel spectrometer was explored in this work. The operating characteristics of a lab grade, commercial spectrometer were compared with a G-Fresnel spectrometer showing capability for similar spectral performance, in regards to spectral resolution. Although sensitivity and spectral bandwidth of the R-G-Fresnel spectrometer fell much lower compared to the commercial spectrometer, there exists room for improvement. Optimizing the surface pattern of the device as well as other aspects of the spectrometer system (alignment, input-coupling, etc.) could alleviate the G-Fresnel's deficiencies to improve comparability. Considering the agreeable comparisons between an expensive, bulky commercial grade spectrometer and the G-Fresnel spectrometer, in addition to the large ceiling of potential improvement of the G-Fresnel device used in this experiment, the hope of a low-cost, high resolution spectrometer is seemingly realizable. With the booming popularity of portable electronics, a spectrometer integrated with a device such as a tablet or smartphone with an intuitive, easy-to-use control interface could be an invaluable component to society.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would first like to thank Zhiwen Liu, and everyone else in the Ultrafast and Nonlinear Optics Lab. I would also like to thank the Pennsylvania State University and my roommates Mason Sutorius, Greg Roberts, and Zach Morgan for making my summer here at State College an enjoyable experience.

REFERENCES

[1] Arai, Akihiro (JP), Harada, Yoshihisa (JP), Tateno, Ryo (JP) 2001 Czerny-Turner spectroscope SHIMADZU CORP (JP) EP1099942 http://www.freepatentsonline.com/EP1099942A1.html

[2] Palmer, C. and Loewen, E., Diffraction Grating Handbook (Newport Corp., 2005).

[3] Hsieh, C., Momtahan, O., Karbaschi, A. and Adibi, A., "A compact lensless and slitless volume holographic spectrometer for diffuse source spectroscopy," Appl. Phys. B 91, 1-6 (2008).

[4] Wolffenbuttel, R. F., "MEMS-based Optical Mini- and Microspectrometers for the Visible and Infrared Spectral Range." Journal of Micromechanics and Microengineering 15 (2005): 145-52.

[5] Chaganti, K., Salakhutdinov, I., Avrutsky, I. and Auner, G. W., “A simple miniature optical spectrometer with aplanar waveguide grating coupler in combination with a plano-convex lens,” Opt. Express 14, 4064-4072 (2006).
[6] Yang, C., Edwards, P., Shi, K. and Liu, Z., "Proposal and demonstration of a spectrometer using a diffractive optical element with dual dispersion and focusing functionality," Opt. Lett. 36, 2023-2025 (2011).
[7] Yang, C., K. Shi, K., Edwards, P. and Liu, Z., "Demonstration of a PDMS based hybrid grating and Fresnel lens (GFresnel) device," Opt. Express 18, 23529-23534 (2010).
[8] Melles Griot, “Argon-Ion Lasers 532 Series,” 532-GS-A01 Datasheet, 2000, http://www.cvimellesgriot.com/Products/Documents/Downloads/532Ion.pdf.

SCATTERING ELEMENT DESIGN FOR SOLAR THERMAL CONCENTRATORS

Mason Sutorius,* Noel C. Giebink,[#] and Jared Price[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Clarkson University
Potsdam, NY 14616

ABSTRACT

Conventional solar thermal concentrators currently use precise 3-D trackers to align with the sun. These concentrators are large, making them costly and prone to errors due to wind loading. We investigate an alternative approach for solar concentration that mitigates these challenges through 2-D planar 'microtracking', in which light from the focal point of a fixed collection optic is scattered into and trapped within a waveguide by total internal reflection. Here, we explore the efficiency and optimization of the in-coupling process. We find that a geometric compound paraboloidal element derived from nonimaging optics yields a high waveguide in-coupling efficiency, $\eta_{opt} = 90.3 \pm 1.13\%$. These results indicate that the scattering concentrator concept may be a viable alternative to established linear Fresnel and parabolic trough technologies.

INTRODUCTION

As the world uses its unsustainable resources, the need for renewable energy increases. Solar energy has huge potential because it is the only natural energy source that is able to keep up with human consumption [1]. In order for the world to use more solar energy, we must decrease the cost of converting sunlight into either electrical energy via photovoltaics or thermal energy. Figure 1 displays the predicted localized cost of energy generation for 2016 [2].

[#] Faculty Mentor
[+] Graduate Mentor

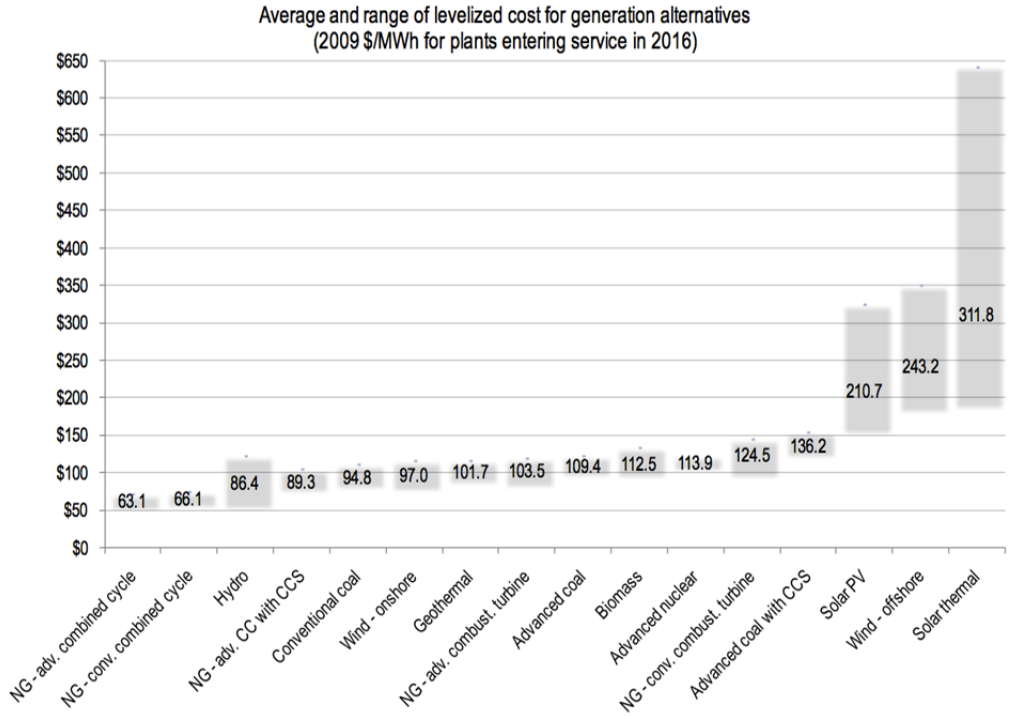

Figure 1: Estimated Levelized Cost of Energy Generation Alternatives (adapted from [2]).

Converting sunlight into heat is of interest because heat can be stored in thermal storage tanks and converted to electricity at night. There are numerous ways to convert solar energy into electricity, but the solar thermal concentrator is to be investigated further. Conventional solar thermal concentrators use a fixed parabolic mirror to concentrate sunlight to a point or sharp line focus. A movable collector is used to track the focal line as the sun moves across the sky [3]. Figure 2 depicts this process below [4].

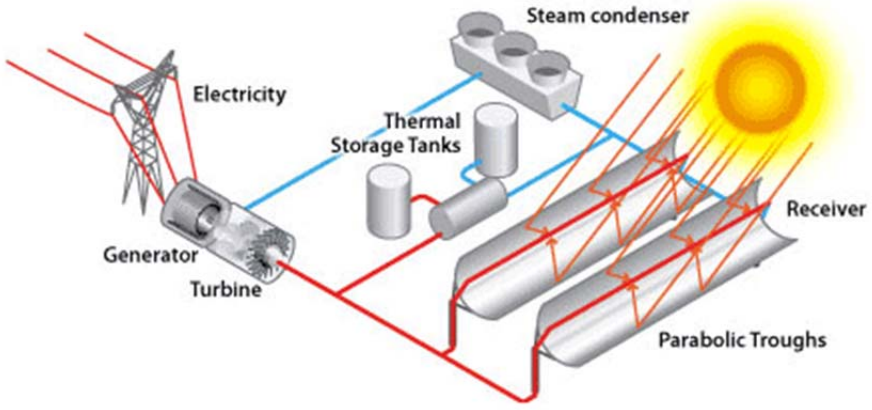

Figure 2: Schematic of a complete concentrating solar power plant (adapted from [4])

A modified approach is also used in which the collector is fixed and a movable mirror is used. Both of these conventional approaches require precise 3-D active tracking of either the focal line or mirror to achieve an optical efficiency between 70 and 80% [5]. Shown in Figure 3 is an example of a parabolic trough solar concentrator used in Solar Energy Generation Systems (SEGS), the largest solar energy generation system in the world [6]. The electrical generation efficiency of these systems is between 29.4 and 37.6% [5].

Figure 3: Parabolic Trough Concentrator, SEGS (from [6])

A novel solar concentrating approach has been designed involving precise 2-D active tracking of a focal point onto a scattering element, which guides light into waveguides. This method has the potential to be inexpensive, small in size, efficient (>70%), and relatively simple to create. Figure 4, as given below, depicts the scattering element process.

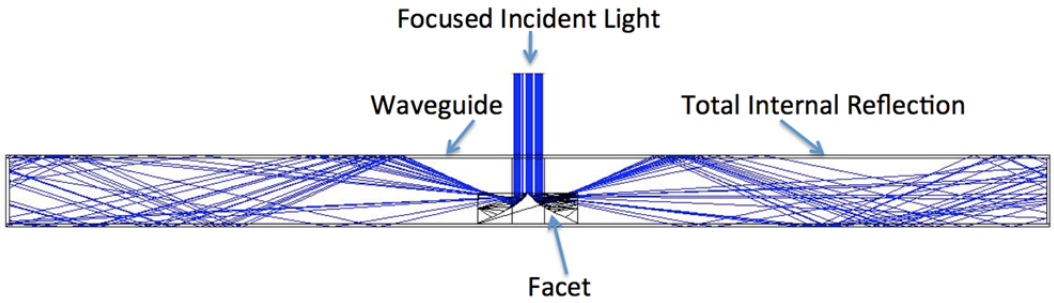

Figure 4: Schematic of light scattering into confined modes of a slab waveguide.

A 2-D active tracking system is required to focus incident concentrated light onto the facet(s) of a scattering element. These facets reflect the light into guided modes of a planar waveguide, propagating toward the edges of the element where

it is absorbed by a blackbody and converted to electricity. The scattering element moves in a planar motion relative to the fixed lens that concentrates the propagated light. As the sun traverses the sky, the focused light begins to distance itself from the reflective facet of the scattering element. The light is recaptured by translating the scattering element relative to the lens that concentrates the light [7]. This approach has numerous advantages, including lower sensitivity to wind loading because of less frontal area and less expensive movement mechanisms due to motion in only one plane.

In this paper we explore the fabrication of scattering elements using a novel approach consisting of a stamping press to imprint specified geometrical properties, and a thermal evaporation system to make the facets reflective. Here we find that scattering elements can be created through this novel approach, with efficiencies as high as $90.3 \pm 1.13\%$. We conclude that this method of fabrication has potential to be a part of a complete scattering concentrator that is a viable alternative to established linear Fresnel and parabolic trough technologies.

EXPERIMENT DESCRIPTION

Fabrication

A stamping press was designed to stamp specified geometric properties of the facets of a scattering element into an acrylic substrate. A load cell and thermocouple were used in the stamping press to make it repeatable by keeping the temperature and applied force relatively constant. Figure 5 depicts the stamping press layout. Two optimized stamps were created using Zemax™ optical ray-tracing software and subsequently implemented using the press. The two shapes, one a parabola and the other a cone, and a resulting substrate are shown below in Figure 6. Acrylic was used as the substrate due to its optical and material properties in addition to its low cost. The entire press sat on a hot plate that was heated to 150 °C, which lies above the glass transition temperature of acrylic, allowing us to indent the plastic. A force of about 100 lbs. was applied to each substrate. The facets were made reflective by depositing ~150 nm of silver through thermal evaporation.

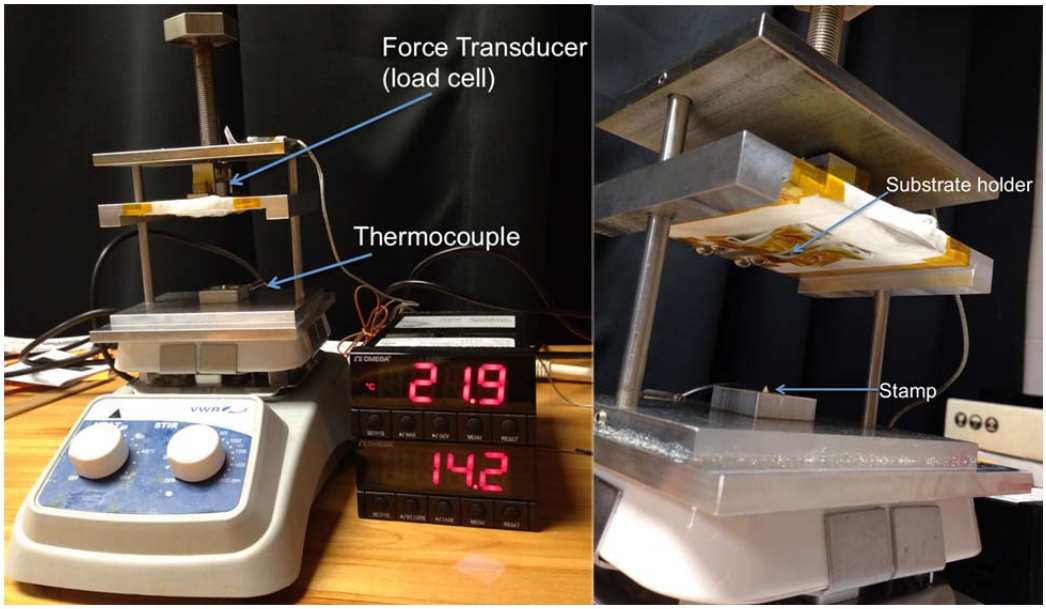

Figure 5: Stamping Press Overview

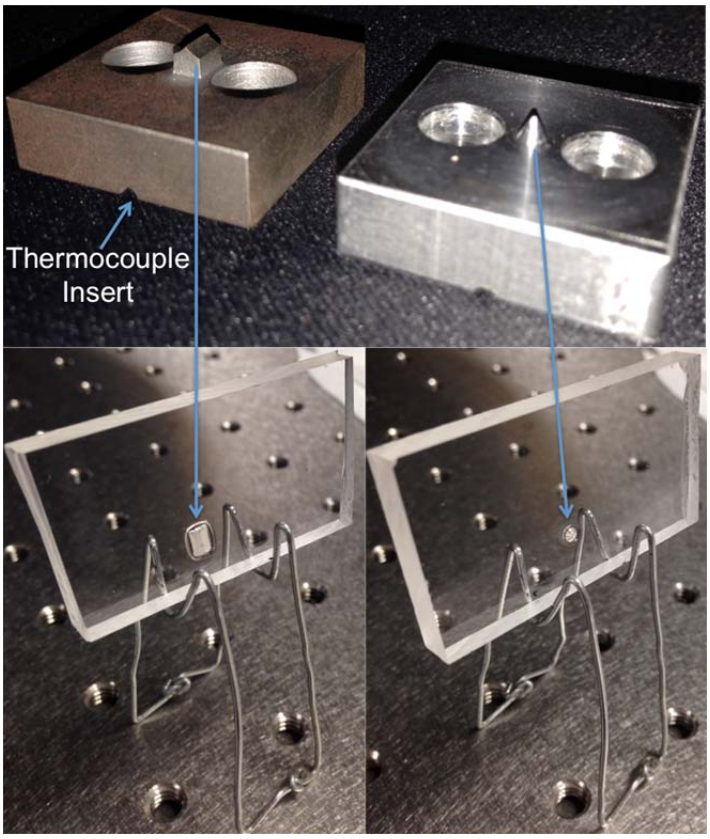

Figure 6: Stamps and Resulting Scattering Elements

Testing

A monochromator was used to vary the wavelength of light from an Energetiq LDLS EQ-99CAL xenon light source. Though the monchromator was able to output light with wavelengths between 300 nm and 1050 nm in increments of 5 nm, substrates were only tested from 400 nm to 1050 nm, since acrylic absorbs light below 400 nm. The light source of varying wavelength was then directed through an optical chopper in order to create an AC signal that could be detected by an Ametek 7230 DSP Lock-in Amplifier. The light hit the scattering element, which was located in an integrating sphere. The integrating sphere homogenized the light being scattered by the element, allowing the use of only one photodiode to detect the intensity of light being scattered through each side of the element. Figure 7 below depicts the experimental setup.

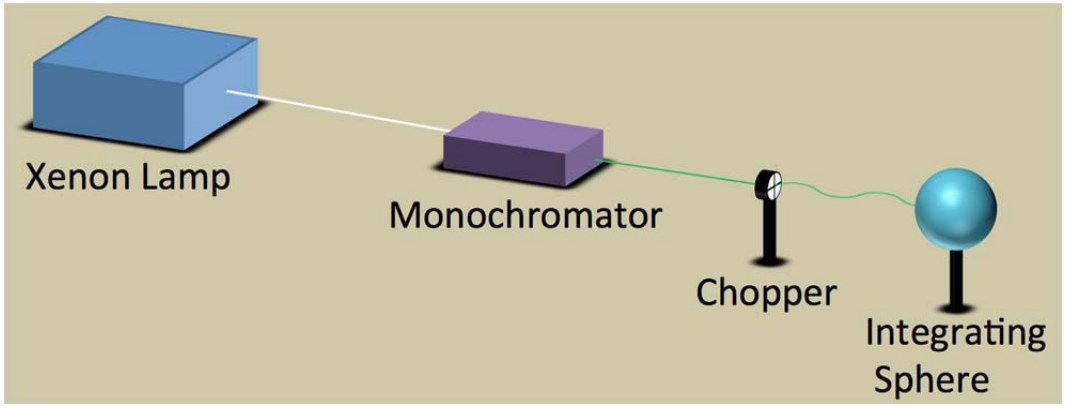

Figure 7: Testing Setup

Each element was tested with and without a blackbody (black electrical tape) surrounding the element in order to calculate the efficiency of the element. The efficiency of the element is given by Eq. 1 below, where *A* represents the amount of current produced without a surrounding blackbody, and *B* represents the current produced with a surrounding blackbody.

$$\eta_{opt} = \frac{A - B}{A} \tag{1}$$

EXPERIMENT RESULTS

The data collected for the experiment are given below in Figures 8, 9, 10, and 11. Scattering element efficiencies across the relevant solar spectrum were calculated using Eq. 1 as given above. Two thicknesses of acrylic were tested, one being 3/16″ (thick) and the other being 1/8″ (thin). Testing with two thicknesses resulted in two different stamping depths into the plastic. Three substrates were tested for both the thick (3/16″) parabolic and conic scattering elements, and two substrates were tested for both the thin (1/8″) parabolic and conic scattering elements. Each substrate was tested three times. The results as given below were

obtained by taking the mean of all substrates and trial per thick/thin and conic/parabolic element.

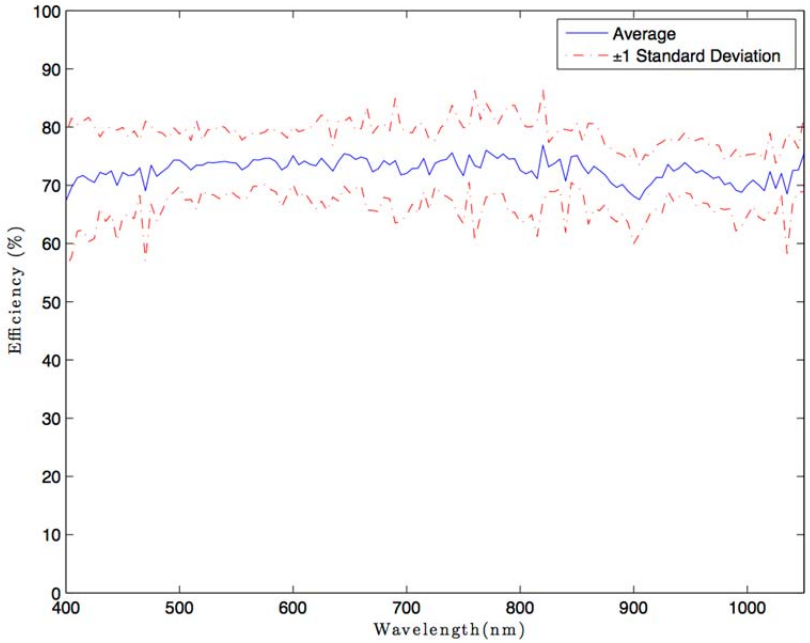

Figure 8: Thick (3/16″) Conic Scattering Element Results

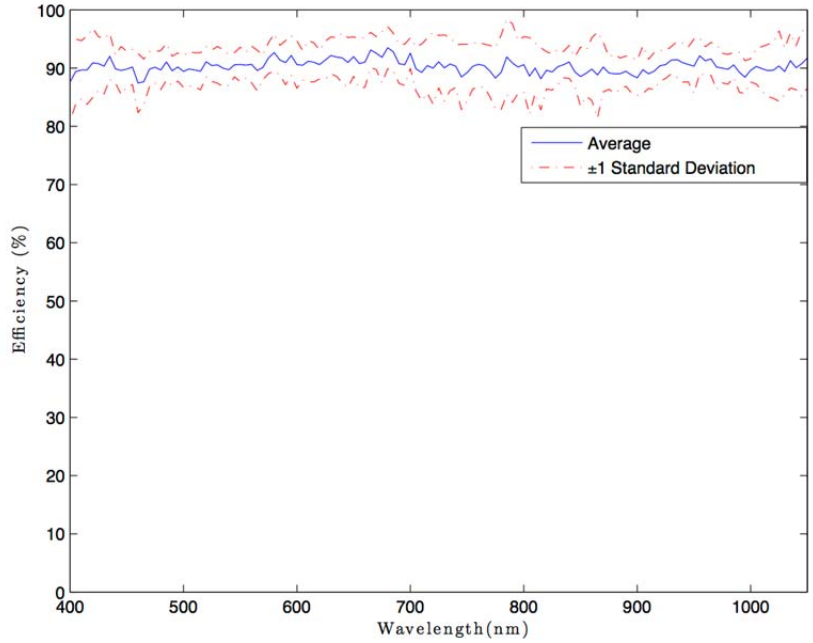

Figure 9: Thick (3/16″) Parabolic Scattering Element Results

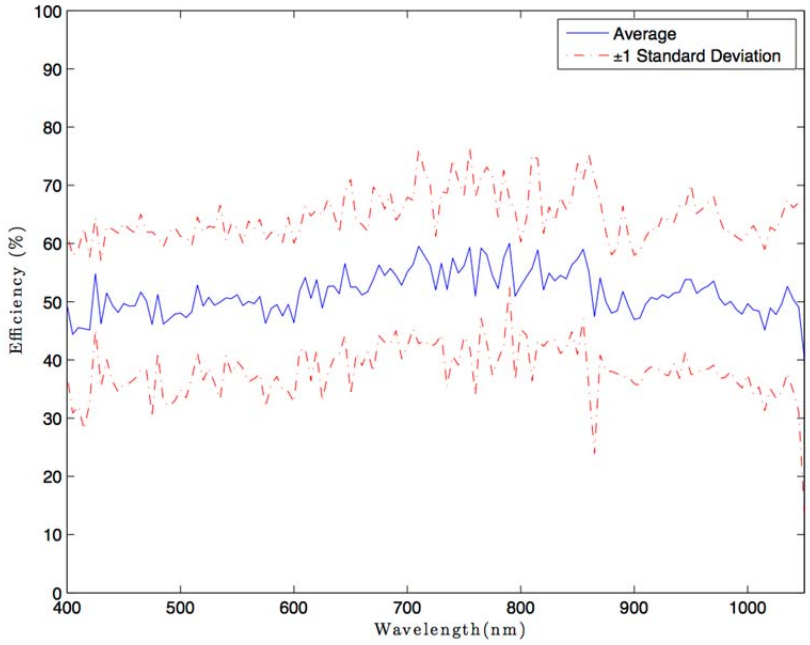

Figure 10: Thin (1/8″) Conic Scattering Element Results

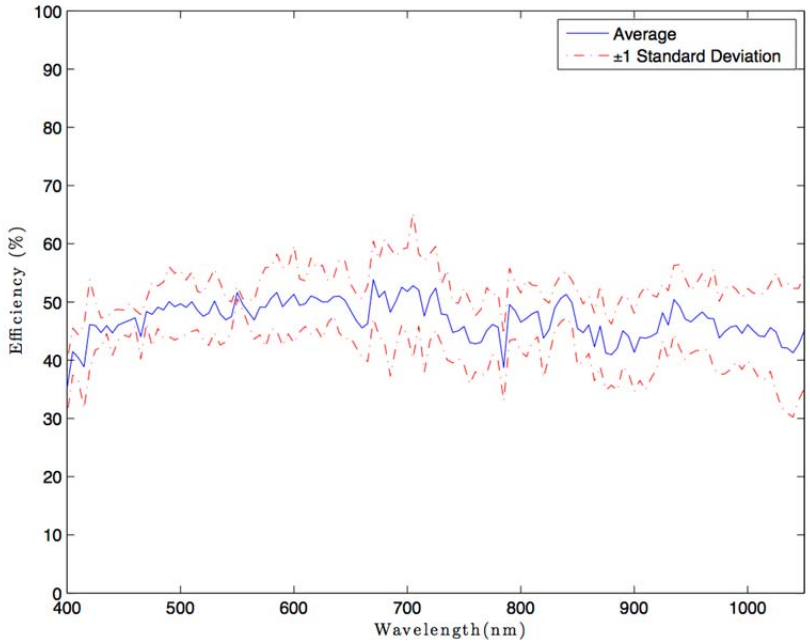

Figure 11: Thin (1/8″) Parabolic Scattering Element Results

DISCUSSION

Based on these results, it is clear that small-scale, lightweight scattering element can achieve >70% efficiency. The efficiency of each type and thickness was averaged across the relevant wavelengths. The efficiencies of the thick and thin parabolic and conic scattering elements are given below in Table 1.

Table 1: Element Efficiencies

Element	Thick Parabola	Thick Cone	Thin Parabola	Thin Cone
Efficiency	90.3±1.13%	72.6±1.92%	47.0±3.30%	51.5±3.66%

One reason that the thin substrates are not as efficient as the thick substrates is because the spot size of the incoming light beam was larger than the facet itself, making it impossible to focus perfectly onto the facets. This phenomenon, depicted in Figure 12 below, also can explain why the efficiency of the conic elements was typically lower than that of the parabolic elements. The noise that appears in our data is caused by errors in the detection of light by the photodiode.

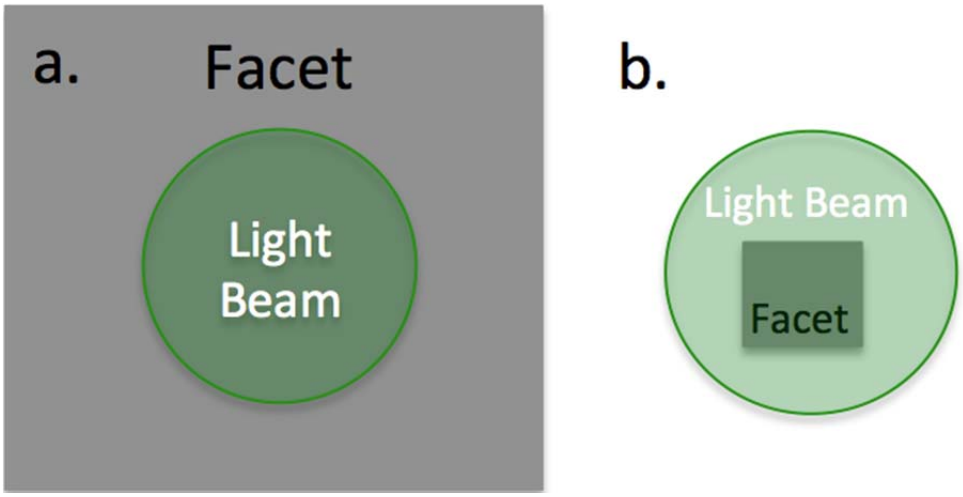

Figure 12: Spot size relative to facet size for (a) thick substrates and (b) thin substrates

CONCLUSION

From the tests performed on each scattering element, it can be concluded that scattering elements made via the process as defined above are an efficient way to couple focused light into a macroscopic waveguide. These results are promising, however, further research is needed at a system level to understand how the efficiency of a complete scattering concentrator compares with established parabolic trough and linear Fresnel technologies.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. EEC-1062984. Furthermore, I would like to thank Jared Price and Noel C. Giebink for their time and effort in helping me with my research project. I would also like to thank Victoria VanMaaren for helping me edit my paper. Thanks to my roommates Bill Kim, Zach Morgan, and Greg Roberts for making my experience here memorable. Lastly, I would like to thank my family for their dedication in furthering my education.

REFERENCES

[1] Lewis, Nathan S. (2007) *Powering the Planet.* Engineering and Science, 70 (2). pp. 12-23.

[2] "Energy Information Administration Electricity Generation Estimates (2011)." IPU-MSU, Apr. 2011. Web. 15 July 2013. <http://ipu.msu.edu/research/pdfs/IPU%20EIA%20Electricity%20Generation%20Estimates%20(2011).pdf>.

[3] John L. Russell, Jr.; Principles Of The Fixed Mirror Solar Concentrator. Proc. SPIE 0085, Solar Energy Utilization II, 139 (February 15, 1977); doi:10.1117/12.954950.

[4] Hamilton, James. "Careers in Solar Power." *U.S. Bureau of Labor Statistics*. U.S. Bureau of Labor Statistics, n.d. Web. 15 July 2013. <http://www.bls.gov/green/solar_power/>.

[5] Price, Hank, et al. "Advances in parabolic trough solar power technology." *Journal of Solar Energy Engineering* 124.2 (2002): 109-125.

[6] *Solaripedia*. Global Swarming LLC, n.d. Web. 15 July 2013. <http://www.solaripedia.com/13/32/362/segs_kramer_junction_solar_collector.html>.

[7] J. Hallas, K. Baker, J. Karp, E. Tremblay, and J. Ford, "Two-axis solar tracking accomplished through small lateral translations," Appl. Opt., 51, 6117-6124 (2012).

AUTHOR INDEX
(* Indicates REU Student Author)