Annual Research Journal

Electrical Engineering
Research Experience for Undergraduates

Vol. VIII

August 2010

PENNSTATE

Department of Electrical Engineering
University Park, Pennsylvania

EEREU Annual Research Journal

Sven Bilén (editor)

NSF EE REU Site Program Contact

PREFACE

We are happy to present the ***2010 Annual Research Journal – Electrical Engineering Research Experience for Undergraduates, Vol. VIII.*** This volume contains highlights of the EEREU program activities for summer 2010 and thirteen technical papers written by EEREU scholars as primary authors.

Summer 2010 marks our eighth year hosting the National Science Foundation–sponsored Research Experience for Undergraduates (REU) Site Program, by the Department of Electrical Engineering, Penn State. Thirteen outstanding young men and women participated in this year's EEREU program at Penn State's University Park Campus. These EEREU scholars, selected from nation-wide applicants, consisted of college first-year, sophomore, or junior students with outstanding academic backgrounds and intense interests in exploring research in electrical engineering and related areas.

During the nine-week summer program, EEREU students carried out research projects under the guidance of his or her faculty mentor(s), in laboratories hosted by the Department of Electrical Engineering and the Materials Research Institute at Penn State. The students also presented their research experience and findings at the 2010 Annual *EEREU Symposium,* held at University Park, in July 2010.

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 2010
University Park, PA

TABLE OF CONTENTS

Research Articles (* indicates REU student author)

TITLE OF ARTICLE	Author	Page
EQUIVALENCE PRINCIPLE APPLIED TO MODELING OF THUNDERSTORM CHARGE CONFIGURATION *Heather. N. Graffius,* Victor P. Pasko, and Jeremy A. Riousset*	………...	1
COMPUTATION OF RELATIVISTIC RUNAWAY ELECTRON AVALANCHE RATES USING FULL ENERGY RANGE MONTE CARLO MODELS *James I. McDonald,* Sebastien Celestin, and Victor P. Pasko*	………...	13
STUDY OF THE POSSIBLE EFFECT OF THE SIGNAL PROCESSING IN THE METEOR-HEAD DATABASE FROM JICAMARCA RADIO OBSERVATORY *Reinaldo Mulero,* Julio Urbina, and Freddy Galindo*	………...	23
VHF/UHF RADAR METEORS AT ARECIBO: THE USUAL SUSPECTS AND SOME BOLIDES *Jonathan Ore,* John D. Mathews, and Akshay Malhotra*	………...	37
THE CHARACTERIZATION OF RF ANTENNAS FOR BEAMED POWER *Charles Perkins,* Aseem Singh, and Sven Bilén*	………...	51
INTEGRATING PRESSURE MEASUREMENTS INTO WIRELESS SENSOR SYSTEMS *Julie Georgiev* and Sven Bilén*	………...	57

2010 NSF EEREU FACULTY & STAFF MEMBERS

Faculty Mentors:

Prof. Sven Bilén, Co-Director
Prof. Ken Jenkins, Co-Director
Prof. I. C. Khoo
Prof. Michael Lanagan
Prof. Zhiwen Liu
Prof. John Mathews
Prof. Vishal Monga
Prof. Robert Nickel
Prof. Victor Pasko
Prof. Jeff Schiano
Prof. Srinivas Tadigadapa
Prof. Julio Urbina

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

NSF EEREU 2010 Summer Program
Research Seminar Series

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

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

Date	*Topic*	*Speaker*
June 8, 2010	Robust Signal Hashing – Theory and Applications	Vishal Monga
June 15, 2010	What's Next? Graduate School of Course!	David Salvia
June 22, 2010	The Joy of Quadrupole Resonance	Jeff Schiano
June 29, 2010	Lightning-related Transient Luminous Events in the Middle Atmosphere	Victor Pasko
July 6, 2010	Integrating Nanowire-based Sensors with CMOS Electronics Using Directed Self Assembly	Theresa Mayer
July 13, 2010	Software-defined Radar for Meteor Target Detection	Julio Urbina
July 20, 2010	Sensors Using Micro and Nanoscale Structures	Srinivas Tadigadapa
July 29, 2010 (Thursday)	2010 NSF EEREU Symposium	REU Students

NSF EEREU 2010 Summer Program
Ethics & Entrepreneurship Workshop

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

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

Date	Topic	Speaker
June 8, 2010	Frameworks of Ethics	Andy Lau
June 15, 2010	Research Ethics	Andy Lau
June 22, 2010	Sustainability and Materials Selection in Product Development	Andy Lau
June 29, 2010	Student Entrepreneur Panel: "What's it take to be an entrepreneur?"	Erik Weir Erik Davidson Phil Boyer Andy Lau
July 6, 2010	Intellectual Property Issues	Phil Boyer
July 13, 2010	Ethics Issues in Entrepreneurship	Phil Boyer
July 27, 2010	Student Presentations on Tech Transfer Ideas of Research Projects	Phil Boyer

2010 NSF EE REU Field Trip Program
Academic/Industrial Sponsors and Tour Hosts

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

Materials Research Institute
Penn State University, University Park, Pennsylvania

Applied Research Laboratory
Penn State University, University Park, Pennsylvania

Breazeale Nuclear Reactor
Penn State University, University Park, Pennsylvania

Nanofabrication National Facility
Penn State University, University Park, Pennsylvania

Videon Central, Inc.
State College, Pennsylvania

AlphaLabs
Pittsburgh, Pennsylvania

MAYA Design
Pittsburgh, Pennsylvania

2010 NSF EE REU SYMPOSIUM

Final Program

Time	Sessions and Topics		Chairs and Speakers
8:30 – 8:55 am	**Symposium Registration (Refreshments Provided)**		(MaryAnn Henderson/ Tom Tyson)
8:55 – 9:00 am	**Welcome**		**Ken Jenkins**
9:00 – 10:00 am	**Session I**	Session Chairs:	**Pasko/Urbina/ Mathews**
9:00– 9:15	EQUIVALENCE PRINCIPLE APPLIED TO MODELING OF THUNDERSTORM CHARGE CONFIGURATION		Heather Graffius
9:15 – 9:30	COMPUTATION OF RELATIVISTIC RUNAWAY ELECTRON AVALANCHE RATES USING FULL ENERGY RANGE MONTE CARLO MODELS		James McDonald
9:30 – 9:45	STUDY OF THE POSSIBLE EFFECT OF THE SIGNAL PROCESSING IN THE METEOR-HEAD DATABASE FROM JICAMARCA RADIO OBSERVATORY		Reinaldo Mulero
9:45 – 10:00	VHF/UHF RADAR METEORS AT ARECIBO: THE USUAL SUSPECTS AND SOME BOLIDES		Jonathan Ore
10:00 – 10:15 am	**Coffee Break**		
10:15 – 11:30 am	**Session II**	Session Chairs:	**Bilén /Jenkins/ Monga/Schiano**
10:15 – 10:30	THE CHARACTERIZATION OF RF ANTENNAS FOR BEAMED POWER		Charles Perkins
10:30 – 10:45	INTEGRATING PRESSURE MEASUREMENTS INTO WIRELESS SENSOR SYSTEMS		Julie Georgiev
10:45 – 11:00	AN INVENTORY SIZE FEASIBILITY STUDY FOR FAST ALGORITHM ANALYSIS/RE-SYNTHESIS BASED SPEECH ENHANCEMENT		Carnell Hunter
11:00 – 11:15	ROBUST VIDEO HASHING USING THE THREE DIMENSIONAL DISCRETE COSINE TRANSFORM		Miloud Id Elhaj
11:15 – 11:30	DETERMINING CONVERSION GAIN IN A MARGINAL OSCILLATOR USING NUMERICAL SOLUTIONS		Michael Scruggs

2010 NSF EE REU SYMPOSIUM (cont.)

Final Program

8:30 am to 3:00 pm, Thursday, July 29, 2010
Room 101 Electrical Engineering East Building
The Pennsylvania State University, University Park, PA 16802

Time	Sessions and Topics	Chairs and Speakers
11:30 – 11:45 am	**(Group Photo – All are invited. Please plan to be available.)**	(Tom Tyson)
12:00 – 1:30 pm	**Luncheon (Nittany Lion Inn)**	
1:45 – 2:45 pm	**Session III** Session Chairs:	**Tadigadapa/Liu/ Khoo/Lanagan**
1:45 – 2:00	MEASURING THE THERMOELECTRIC POWER OF NANOMATERIALS ON A MICROMACHINED WORKBENCH	Matthew Chang
2:00 – 2:15	COMPARATIVE ANALYSIS OF SILICON CANTILEVER AND MICROMACHINED QUARTZ GRAVIMETRIC SENSORS	Emmanuel Carpena
2:15 – 2:30	NONLINEAR OPTICAL LIMITING IN FIBER ARRAY WITH L34	Seonwoo Lee
2:30 – 2:45	ELECTRICAL CHARACTERIZATION OF CONDUCTING MATERIALS AT MICROWAVE FREQUENCIES	Cameron Patterson
2:45 – 3:00 pm	**CONGRATULATIONS AND CONCLUDING REMARKS**	**Jenkins/Bilén**
	3:00 pm Adjournment	
	Picnic at Sunset Park, 5:00 pm – sunset, ALL ARE INVITED	
Resources:	Tom Tyson and MaryAnn Henderson	
Special Notes:	Laser pointer, desktop PC and LCD PC projector are provided. Speakers please upload presentation files to EEREU group web in Angel by Wednesday, July 28. 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.	

EQUIVALENCE PRINCIPLE APPLIED TO MODELING OF THUNDERSTORM CHARGE CONFIGURATION

Heather. N. Graffius,* Jeremy A. Riousset,[+] and Victor P. Pasko[#]

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

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

ABSTRACT

The equivalence principle has been used to model the changes in the electrical structure of a thunderstorm caused by lightning strikes.[1] However a rigorous justification of the applicability of this principle to model the electric field above the cloud remains to be established. This work uses an observation-based model of the electrical structure of thunderstorms to explore the effects of lightning on this structure and how related electric fields can be modeled using the equivalence principle. This study looks into applicability of the equivalence principle for the modeling of the electric field above the cloud. The electric field changes following a variation of charge before and after a lightning strike are calculated numerically and the results obtained with and without using the equivalence principle are compared. The results establish a proof of the validity of the equivalence principle for modeling the electric field change above the thundercloud following the occurrence of lightning.

INTRODUCTION

The first studies aimed to determine the electrical structure of thunderstorms started in the early 1900's.[2–6] The structure was initially believed to be a vertical dipole with a positively charged region above a negatively charged region. Around 1940[7,8] a tripole structure was adopted, with a lower positive charge region below the dipole. The two upper charge regions are considered to be the main charges and are specified to be approximately equal in magnitude. They form a positive dipole, i.e., the positive charge is above the negative, which gives

---

[#] Faculty Mentor
[+] Graduate Mentor

an upward-directed dipole moment. The electric field intensity due to the three vertically stacked charges can be found by replacing the conducting ground with three image charges using the principle of superposition. The total electric field then becomes the vectorial sum of all six contributions. Because of the symmetry of the problem the magnitudes of the contributions from the actual charge and its images are equal at $z = 0$ km.[9] In 1996, it was suggested that the tripole and dipole charge structure models were too simplistic.[10,11] An improved model of the charge structure of the thunderstorm was in presented in 1998 by *Stolzenburg et al* as shown in Figure 1.[11] Yet, the tripole model appears to correctly represent the electrical structure in the convective region (see Figure 1).

Storms, similar to the one illustrate in Figure 1, commonly produce lightning, and so-called transient luminous events. Lightning is an electrical discharge between and within the three main charge regions of the cloud (lower positive, central negative, and upper positive charge regions) accompanied by charge transfer between the charge regions or to the ground. Transient luminous events (TLEs) are different types of electrical discharges produced by large thunderstorms in the altitude range 15–90 km, i.e, between the altitude of cloud tops and the lower ionosphere. The most common types of TLEs are blue jets, gigantic jets, sprites, and elves as illustrated in Figure 2 below.[12] Blue jets are relatively slow-moving fountains of blue light which emanate from the top of thunderclouds up to an altitude of 40 km at speeds ~100 km/s and are characterized by a blue conical shape.[12]
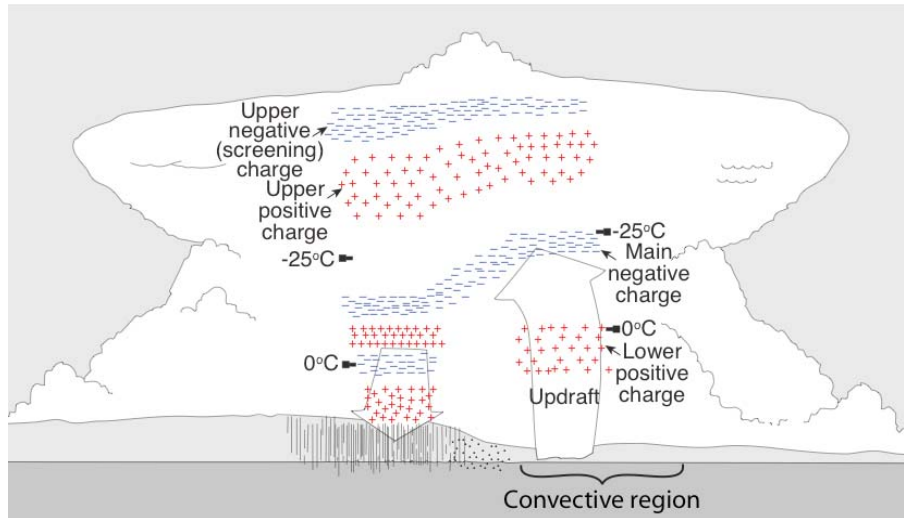

Figure 1: Schematic of charge structure of a thunderstorm.[11] A tripolar structure is visible in the convective region (right hand side) with a central negative, upper positive, and lower positive charge regions. The tripole is topped by an upper negative screening charge.

Gigantic jets are visually similar to blue jets and establish a direct path of electrical contact between thundercloud tops and the lower ionosphere.[12] Sprites

are luminous glows occurring above thunderstorms at altitudes typically ranging from ~50 to 90 km. Sprites develop at the base of the ionosphere and move rapidly downward at speeds up to 10 000 km/s. In video, they exhibit a red color at their top, which gradually changes to blue at lower altitudes. The lateral extent of "unit" sprites is typically 5–10 km, and they last for several milliseconds.[1] Elves are rapid (<1 ms) optical emissions at 80–95 km altitudes with lateral extents up to 300 km.[1] Our study particularly applies to the investigation of the electric field that initiates the sprites.

The goal of this paper is to show how the equivalence principle can be used to represent the changes in the electrical structure of the thunderstorm induced by a lightning strike and the resulting variations of the electric field above the cloud up to the lower ionospheric region.[1] The principle of equivalence states that electric field change following the charge removal in a complex system of electrical charge regions can be modeled in the far field by the addition of an equivalent charge but with opposite polarity at the location of the original removal of charge. In this paper, we use the basic tripole charge structure of the cloud described above to model thunderstorm in order to demonstrate the applicability of this equivalence model in the framework of lightning and TLE studies.
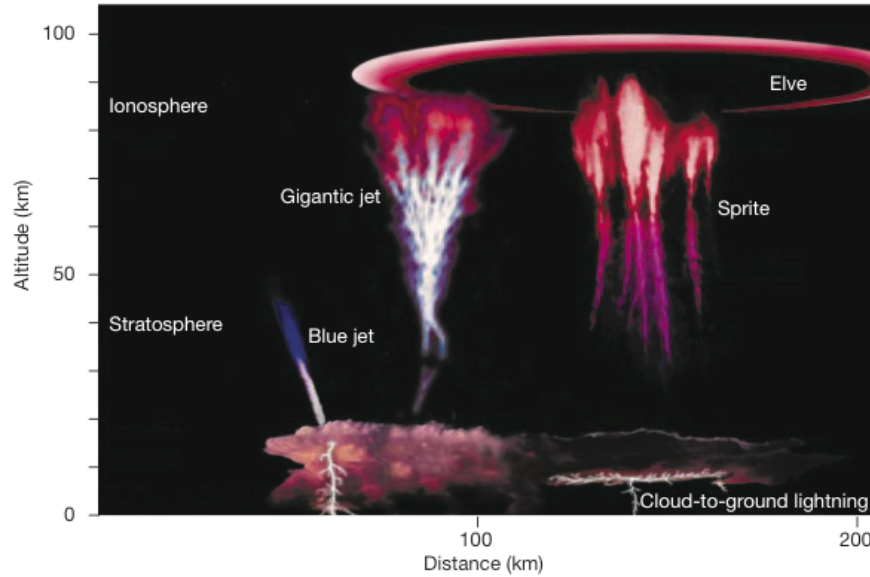

Figure 2: The different types of transient luminous events (TLEs) in relation to the altitude.[12]

MODEL FORMULATION

To test the equivalence principle in the framework thunderstorm studies, a classic tripolar charge structure[13] is used inside a dielectric cylinder of radius r_c and vertical extent z_c (Table I) that represent the limits of the cloud. Each charge layer in the structure is assumed to be cylindrically symmetric and disk shaped.[14] The dimensions, locations, and net charge contents of the charge regions for this

test case are shown in Table II and were developed based on observations of a storm over Langmuir Laboratory, New Mexico on 31 July 1999.[14,15]

The value of the net charge content of each charge region is assumed prior a simulation run and reported in Table II. The source charges are uniformly distributed within each charge disk and produce time- and space-varying induced free charges ρ_f. The source charge densities ρ_s can be calculated by the assumed charge Q_i of the region i (positive, negative, lower positive charge layer) under consideration using the equation $\rho_{s,i} = \dfrac{Q_i}{\pi r_i^2 d_i}$ where the radius r_i and depth d_i of the region are given in Table I. Together ρ_s and ρ_f create an electric potential ϕ both inside and outside the storm. The basic set of equations relating ρ_f, ρ_s and ϕ are:

$$\nabla^2 \phi = -\frac{\rho_s + \rho_f}{\varepsilon_o} \tag{1}$$

$$\frac{\partial \rho_f}{\partial t} - \nabla \sigma \cdot \nabla \phi = -\sigma \frac{\rho_s + \rho_f}{\varepsilon_o} \tag{2}$$

where σ is the atmospheric conductivity. The total charge density can be represented by $\rho_t = \rho_s + \rho_f$. The above equations express Gauss's law and conservation of charge, in which the conduction current $\boldsymbol{J}$ is assumed to be Ohmic and replaced by $\boldsymbol{J} = \sigma \boldsymbol{E} = -\sigma \nabla \phi$ in (2).[1,16]

The conductivity σ at any location $(r,\ z)$ in the simulation domain is expressed by

$$\sigma(r,z) = \sigma_o e^{\frac{z+z_{gnd}}{h}} \left(1 - \frac{1 - \tanh\left(\dfrac{r - r_c}{\alpha}\right)}{2} \times \frac{1 - \tanh\left(\dfrac{z - z_c}{\alpha}\right)}{2} \right) \tag{3}$$

where the parameter α determines the thickness of the conductivity transition region between the cloud interior and the surrounding clear air (Table I). The conductivity inside the thunderstorm is reduced because of the larger value of the ion attachment coefficient in thunderclouds.[17]

Table I. Cloud Boundaries

	Symbol	Parameter [km]
Radius	r_c	5
Height	z_c	9
Boundary thickness	α	0.75

The conductivity outside the cloud increases exponentially with altitude z (term $\sigma_o e^{\frac{z+z_{gnd}}{h}}$ in (3)) with an altitude scaling factor $h = 6$ km and a conductivity at sea level σ_o defined as 5×10^{-14} S/m. [1,16] Inside the cloud the conductivity is reduced to zero by the factor in brackets following the exponential term

$$\left(1 - \frac{1 - \tanh\left(\frac{r - r_c}{\alpha}\right)}{2} \times \frac{1 - \tanh\left(\frac{z - z_c}{\alpha}\right)}{2} \right)$$

in (3) with a smooth transition at the boundary of width $\sim 2\alpha$. [16] For the simulation presented in this paper α is set equal to 0.75 km.

In order to accurately reflect the effect of a negative cloud-to-ground discharge on the thunderstorm, the cloud-to-ground discharge is assumed to transfer –20 C to the ground from the central negative charge region as shown in Figure 3. The typical value of negative cloud-to-ground charge transfer ranges between 20–80 C.[9]

The time scale of the charge dissipation $\tau_\sigma(z)$ is given by $\tau_\sigma = \varepsilon_o / \sigma(z)$ where ε_o is the free space permittivity, $\sigma(z)$ is the conductivity of the medium as a function of the altitude z, and σ_o is given by (3). It is used to find the amount of time that the system needs to reach steady state, and consequently the time the model needs to run. At the top of the cloud $\tau_\sigma(12 \text{ km}) \sim 14.5$ s and therefore the steady state is achieved after $\sim 3\, \tau_\sigma \approx 43.5$ s.

Table II. Charge Values, and Dimensions for Charge Regions Used in the Model

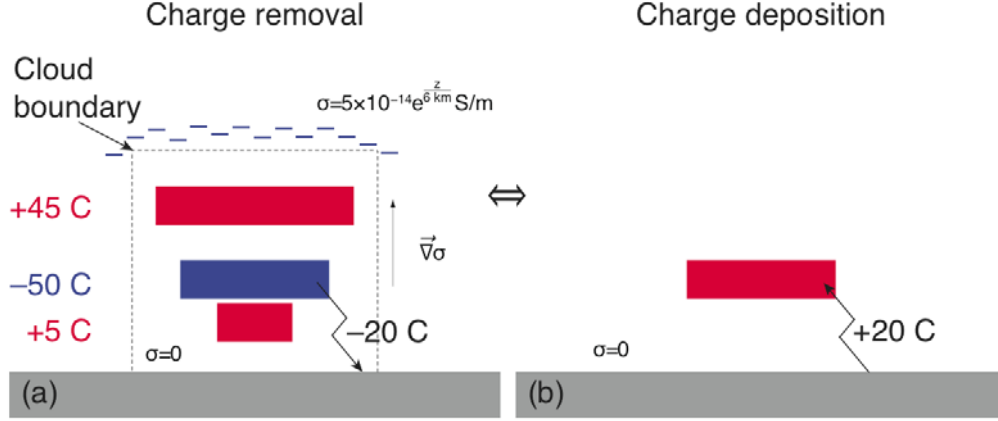| Charge Layer	Altitude [km]*	Depth [km]	Radius [km]	Charge [C]
Positive	6.75	1.5	4.0	45
Negative	3.75	1.5	3.0	-50
Lower Positive	2.00	1.5	1.5	-5

*above ground level (ground level set at 3 km)

Figure 3: Schematic of the full and equivalent models. Panel (a) shows the thundercloud structure and the –20 C discharge from the central negative region. Panel (b) shows the equivalent model of the thunderstorm in which +20 C charge is deposited at the location from which –20 C charge was removed.

For this model $\tau_{min} = \tau_{\sigma_{max}} = 11.1$ μs. To find a suitable time step to satisfy the Courant–Friedrick–Lewy Condition[16] for stability of numerical scheme δt was defined as $\delta t < \frac{1}{2}\tau_{min} = \frac{1}{2}(11.1 \mu s)$. For the run presented in this simulation, we choose $\delta t = 0.5$ μs.

The principle of equivalence, as defined for the purposes of studies presented in this paper, states that the removal of a –20 C charge from the midlevel negative charge region can be modeled by an addition of an equivalent +20 C charge deposited at the same location. This can be represented by a single positive charge region with dimensions of the central negative region from and at the same location (see Table II and the Figure 3a), with net charge content equal to +20 C, in a dielectric atmosphere as shown in Figure 3b. Unlike for the case of the modeling of a tripolar cloud, the absence of a conductive atmosphere leads to a static solution in that latter case.

RESULTS AND DISCUSSION

In this section we present and discuss the results from the simulation of the variations in electric field following a negative cloud-to-ground discharge removing –20 C from the negative charge region of a classic, normally electrified, tripolar thundercloud. Hereafter, we refer to this case as "full model case." We also present the results from an alternative model based on the equivalence principle, and we refer to this second case as "equivalent model case."

From the simulation results we can compare the total charge density of the full model case over the entire simulation domain right before and at the moment when the cloud-to-ground discharge occurs. Figure 4 shows the cloud right before

the cloud-to-ground discharge at $t = 37.5$ s. Figure 4a shows the total charge density ρ_t just before the initiation of the cloud-to-ground discharge, while Figure 4b shows the magnitude of the electric field at the same instant of time. The Figure 4c shows the profiles of the source charge density ρ_s, induced free charge density ρ_f, and total charge density ρ_t at $r = 0$ km. Figure 5 shows the same results as Figure 4, except at the time when the discharge occurs. From these figures it is clear that the magnitude of the electric field in the vicinity of thundercloud decreases as the discharge occurs.
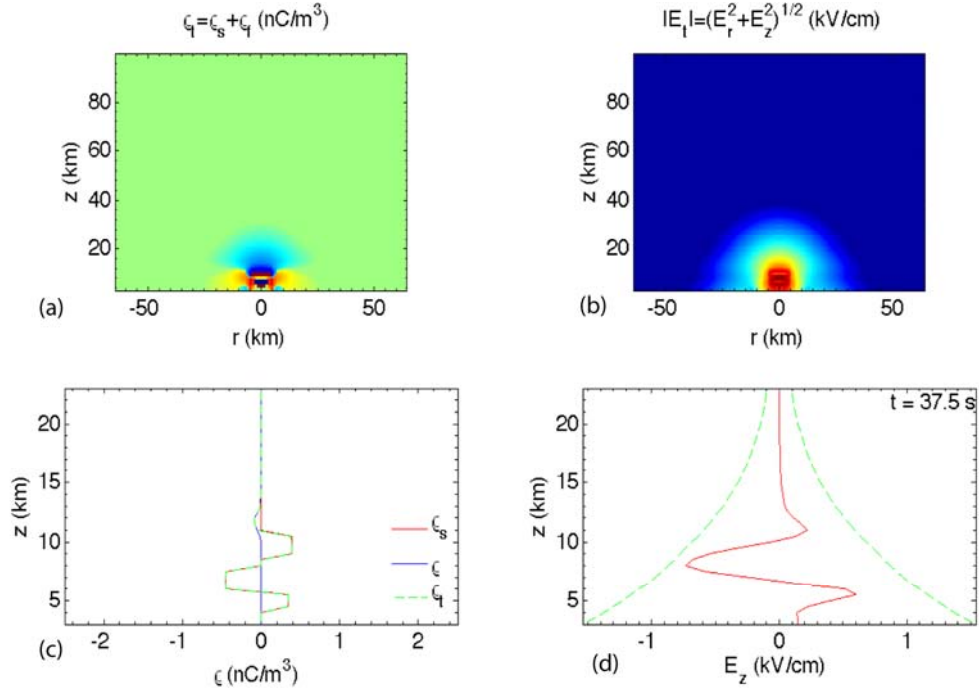

Figure 4: Simulation of full model case at $t \sim 37.5$ s. (a) Total charge density and (b) electric field magnitude, (c) source charge density ρ_s, induced free charge density ρ_f, and the total charge density ρ_t, and (d) electric field with respect to altitude z, at $r = 0$ km.
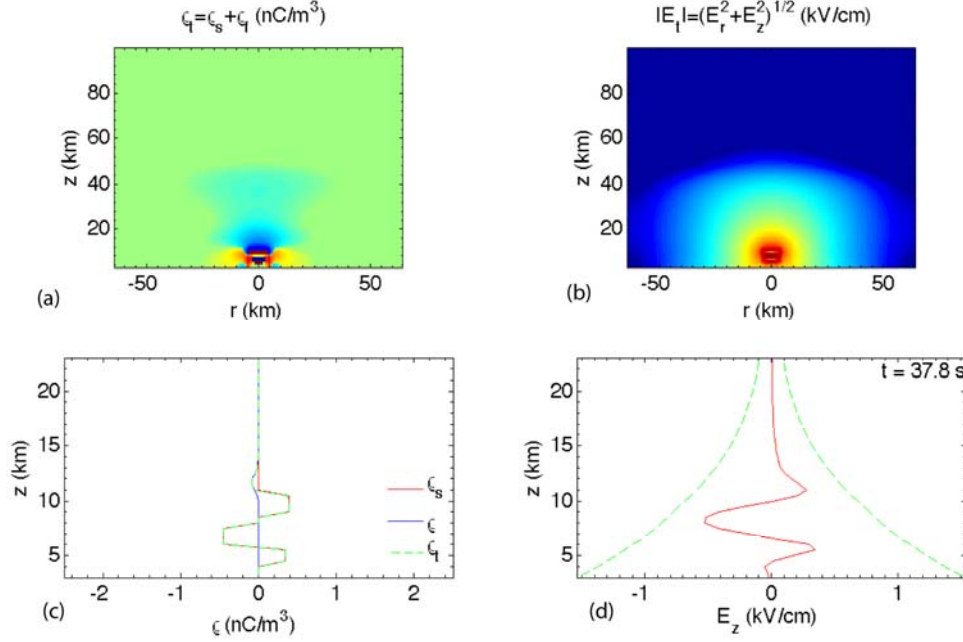
Figure 5. Same as Figure 4, except at $t = 37.8$s, as the -20 C cloud-to-ground discharge occurs.

Figure 6 shows the electric field with respect to the altitude of the full model case and the equivalent model on the same graph on the axis of the simulation domain at $r = 0$ km. From this figure it can be seen that at higher altitudes (i.e., above ~25 km) the electric field of the full model case and the equivalent model case are in excellent agreement. The figure also shows the lightning initiation threshold, which is the electric field, required to initiate the lightning.[14] At lower altitudes, the electric fields of the two models become significantly different. Figure 6 displays three local minima corresponding to the field inversions at the locations of the lower positive charge, central negative, and upper positive regions of the altitudes 5 km, ~7 km, and ~10 km, respectively. In contrast the equivalent model only presents one field inversion at the location of the unique positive charge ($z \sim 7$km).

The equivalence principle is sometimes misinterpreted when applied to the charge in thunderclouds by implying that the excess of positive charge remaining in the cloud after the discharge is the source of the electrostatic field in the lower ionosphere and associated with the occurrence of some TLEs (in particular sprites).[18]
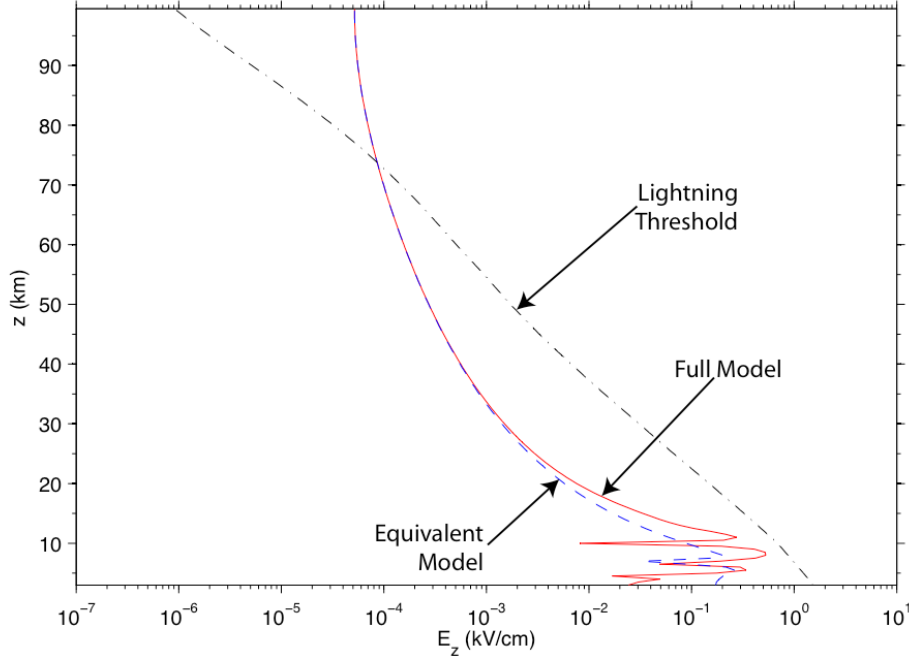
Figure 6: Electric field amplitude E_z as a function of the altitude z at $r = 0$ km. The solid line represents the full model case, the dotted line is the equivalent model and the dotted-dashed line shows the lightning initiation threshold assuming that it scales proportionally to air density with altitude.

The equivalence principle actually states that the removal of charge subsequent to the occurrence of a cloud-to-ground discharge is equivalent to the addition of a new charge of opposite polarity. *Pasko et al.*[1] applied the equivalence principle to simple, dipolar, normally electrified thundercloud. More specifically *Pasko et al.*[1] further showed that as the thundercloud charges slowly build up before a lightning discharge, high-altitude regions are shielded from the quasi-electrostatic fields of the thundercloud charges by space charge induced in the conducting atmosphere at lower altitudes. The appearance of this shielding charge is a consequence of the finite vertical conductivity gradient of the atmosphere above the thundercloud. When one of the thundercloud charges is quickly removed by a lightning discharge, the remaining charges of opposite sign in and above the thundercloud produce a large quasi-electrostatic field that appears at all altitudes above the thundercloud, and endures for a time equal to approximately the local relations time ($\tau_\sigma = \varepsilon_o / \sigma(z)$) at each altitude. This large electric field can be thought of as being the difference between the electrostatic field dictated (via Coulomb's law) by the dipole configuration of thundercloud charges and the polarization charge which are in effect before the discharge, and that required by the combination of the single thundercloud charge remaining after the discharge and polarization charge.[1] In cases of more complex charge distributions in the thundercloud which sometimes involve up to six charge layers

in the vertical direction,[19] each of the charge centers can be viewed as generating its own polarization charge in and above thundercloud, and the resultant configuration of the electric field and charge density can be obtained by using principle of superposition. This consideration is helpful in visualization of the fact that the electric field appearing at mesospheric altitudes after the charge removal by cloud-to-ground lightning discharge is defined mostly by the absolute value and altitude of the removed charge and is essentially independent of the complexity of the charge configuration in the cloud. The charge removal can also be viewed as the "placement" of an identical charge of opposite sign. The initial field above the cloud is simply the free space field due to the "newly placed" charge and its image in the ground which is assumed to be perfectly conducting.[1] These conclusions are further supported and demonstrated by the simulation results presented in the present paper.

CONCLUSIONS

In this paper, we demonstrate how the equivalence principle can be used to represent the changes in the electrical profile of the thunderstorm induced by a lightning strike and the resulting variations of the electric field above the cloud up to the lower ionospheric region.[1] The model simulations for the full model case and the equivalent model case are shown to have very similar electric fields, especially when comparing the field farther away from the cloud. The full model case simulates a thundercloud that has a three charge region electric structure (lower positive, central negative, and upper positive charge layers) enclosed in the conductive atmosphere. The cloud has parameters of a 5 km radius, 9 km height, 0.75 km thickness of the cloud boundary, and is placed above the perfectly electrically conducting ground located at altitude of 3 km to simulate the condition of a New Mexico thunderstorm. The full model simulation showed a cloud-to-ground discharge removing –20 C from the negative charge region. The equivalent model simulated the equivalent system of the full model case in a non-conductive atmosphere. The agreement above 25 km between the full model, including a tripolar cloud in the conducting atmosphere, and the equivalent model as shown in Figure 6 demonstrates the applicability of the equivalence principle to model the electric field induced by a cloud-to-ground discharge in a realistic thundercloud model above altitudes 20–30 km.

ACKNOWLEDGMENTS

REFERENCES

1   Pasko, V.P., U. S. Inan, T. F. Bell, and Y. N. Taranenko, Sprites produced by   quasi-electrostatic heating and ionization in the lower ionosphere, *J. Geophys. Res., 102,* 4529–4561 doi:10.1029/96JA03528 (1997).

2   Wilson, C. T. R., On some determinations of the sign and magnitude of electric discharges in lightning flashes, *Proc. R. Soc. London, Ser. A, 92,* 555–574 (1916).

3   Wilson, C. T. R., Investigations of lightning discharges and on the electric field of thunderstorms, *Phil. Trans. R. Soc. London, Ser. A, 221,* 73–115 (1920).

4   Wilson, C. T. R., The electric field of a thundercloud and some of its effects, *Proc. R. Soc. London, 37, 32D–37D* (1925).

5   Simpson, G. C., On the electricity of rain and snow, *Proc, R. Soc. London, Ser. A., 83,* 394–404 (1909).

6   Simpson, G. C., The mechanism of a thunderstorm, *Proc. R. Soc. London, Ser. A., 114,* 376–401 (1927).

7   Simpson, G. C., and F. J. Scrase, The distribution of electricity in thunderclouds, *Proc. R. Soc. London, Ser. A., 161,* 309–353 (1937).

8   Simpson, G. C., and G. D. Robinson, The distribution of electricity in thunderclouds, II, *Proc. R. Soc. London, Ser. A. 177,* 281–328 (1941).

9   Rakov, V.A., and M. A. Uman *Lightning: Physics and Effects,* Cambridge Univ. Press, Cambridge, U. K. (2003).

10   Rust, W. D., and T. C. Marshall, On abandoning the thunderstorm tripole-charge paradigm, *J. Geophys. Res., 101,* 23, 499–23, 504 (1996).

11   Stolzenburg, M., W. D. Rust, and T. C. Marshall, Electrical structure in thunderstorm convective regions – 3. Synthesis, *J. Geophys. Res.,* 103(D12), 14, 097–14,108, doi:10.1029/97JD03545 (1998).

12   Pasko, V. P., Blue jets and gigantic jets: Transient luminous events between thunderstorm tops and the lower ionosphere, *Plasma Phys. Control. Fusion, 50*(12), 124050, doi:10.1088/0741-3335/50/12/124050 (2008).

13   Williams, E. R., The tripolar structure of thunderstorms, *J. Geophys. Res., 94,* 13,151–13,167 (1989).

14   Riousset, J. A., V. P. Pasko, P. R. Krehbiel, R. J. Thomas, and W. Rison, Three-dimensional fractal modeling of intracloud lightning discharge in a New Mexico thunderstorm and comparison with lightning mapping observation, *J. Geophys. Res., 112,* D15203, doi:10.1029/2006JD007621 (2007).

15   Krehbiel, P., W. Rison, R. Thomas, T. Marshall, M. Stolzenburg, W. Winn, and S. Hunyady, Thunderstom charge studies using a simple cylindrical charge model, electric field measurements, and lightning mapping observations, *Eos Trans. AGU, 85*(47), Fall Meet. Suppl., Abstract AE23A–0843 (2004).

16   Riousset, J. A., V. P. Pasko, P. R. Krehbiel, W. Rison, and M. A. Stanley, Modeling of thundercloud screening charges: Implications for blue and gigantic jets, *J. Geophys. Res., 115,* A00E10, doi:10.1029/2009/JA014286 (2010).

17   Rycroft, M. J., A. Odzimek, N. F. Arnold, M. Fullekrug, A. Kulak, and T. Neubert, New model simulations of the global atmospheric electric circuit driven by thunderstorms and electrified show clouds: The roles of lightning and sprites, *J. Atmos, Sol. Terr. Phys., 69*(17–18), 2485–2509, doi:10.1016/j.jastp.2007.09.004 (2007).

18   Luque, A., U. Ebert, Emergence of sprite streamers from screening-ionization waves in the lower ionosphere, *Nature Geoscience,* 757-760, doi:10.1038/NGE0662 (2009).

19   Marshall, T. C., and W. D. Rust, Two types of vertical electrical structures in stratiform precipitation regions of mesoscale convective systems, *Bull. Am. Meteorol. Soc., 74,* 2159 (1993).

COMPUTATION OF RELATIVISTIC RUNAWAY ELECTRON AVALANCHE RATES USING FULL ENERGY RANGE MONTE CARLO MODELS

James I. McDonald,* Sebastien Celestin,[#] and Victor P. Pasko[#]

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

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

ABSTRACT

This paper focuses on the development of Monte Carlo models to simulate Relativistic Runaway Electron Avalanches (RREAs), which are believed to be the causes of hard x-rays and terrestrial gamma-ray flashes (TGFs) observed in the Earth's atmosphere in association with lightning activity. The accurate modeling of the physical processes at play in RREAs requires keeping track of electrons with energies from sub-eV to MeVs, which represents a major numerical challenge. In the present work, the growth rates of RREAs are documented for homogeneous applied electric fields from approximately 4 kV/cm to E_k, where E_k is the conventional breakdown electric field threshold, allowing for direct comparison with rates documented in the existing literature and application to TGF generation processes.

INTRODUCTION

Runaway Electrons

This project focuses on the simulation of electron scattering and acceleration in air in an applied electric field as a means of simulating runaway electrons through Monte Carlo Models. Specifically, this paper refers to electrons above thunderstorms, which accelerate to higher and higher velocities as a result of the electric fields that arise during lightning. Low energy electrons reach a limit of velocity, when the accelerating force of the electric field on the electrons is equal

---

[#] Faculty Mentor

to the drag force of the collisions due to the molecules in air. This drag force is called the *dynamic friction force,* which is the energy lost per unit distance traveled in air. However, if an electron's energy reaches past approximately 123 eV, regardless of altitude, the dynamic friction force begins to drop as the electron energy increases and may drop to below the electric force felt by the electron. This will cause this electron, called a runaway electron, to only continue to accelerate. The dynamic friction force also has a local minimum for electrons with energies around 1.1 MeV (see Figure 1). The dynamic friction force rises afterwards due to relativistic effects and irradiative processes. As long as the applied electric field on the electrons is higher than this minimum, runaway electrons have a chance of developing if they are energetic enough. Once a runaway electron develops, its energy will continue to increase as its velocity will asymptotically approach the speed of light.

    Figure 1 represents the dynamic friction force experienced by one electron as a function of its energy. The electric field E_{run} is shown to intersect with the minimum point on the dynamic friction force $F_D(\varepsilon)$ at 1.1 MeV such that $E_{run} \cong 1.82 \times 10^5$ V/m. The E_{run} field is the minimum applied electric field required for runaway electrons to exist.[1] This value for E_{run} is calculated from our own work and will be defined as $E_{run} = 2.15 \times 10^5$ in order to compare with the existing literature[2,3]. This net force can be best understood by considering the net force on the electron such that $F_e = q_e E - F_D(\varepsilon)$, where q_e is the net charge of an electron, and F_D is the dynamic friction force as a function of electron energy ε (eV). This shows that when $E > E_{run}$ and $F_e > 0$ each individual electron will accelerate in the direction of the net force, since $m_e\, dv/dt = F_e$, where m_e is the mass of the electron and v is electron velocity. We emphasize that due to their negative chare, electrons accelerate in the direction opposite to the direction the applied electric field.
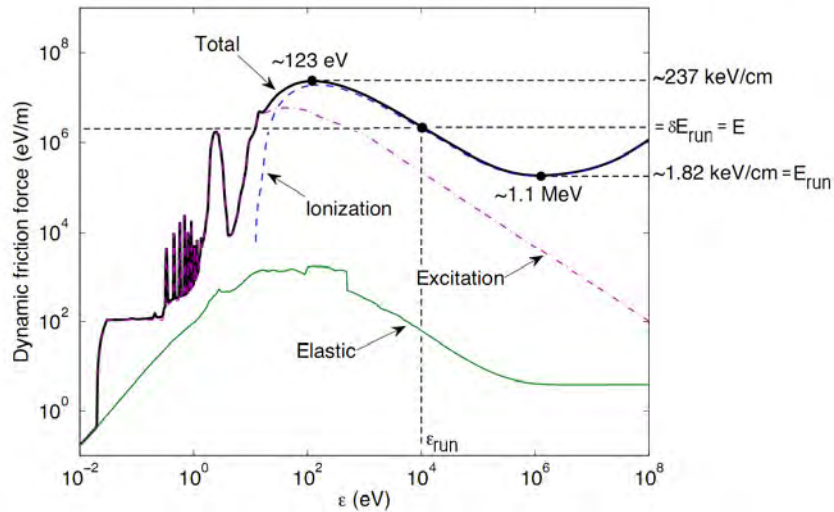

Figure 1: The Dynamic Friction Force in air with 80% N_2 and 20% O_2 and with total number density $N = 2.688 \times 10^{25}$ m^{-3}.

Relativistic Runaway Electron Avalanches (RREA)

RREAs involve the process in which runaway electrons collide with gas molecules in the atmosphere and ionize them, removing secondary electrons from atmospheric molecules' electron shells. In this ionization process, a fraction of the energy of the incident electron is donated to the secondary electron. Only few secondary electrons have enough energy to runaway, but it appears that enough secondary electrons are created that surpass the kinetic energy threshold to maintain the avalanche and allow it to grow in size exponentially.[4] The number of runaway electrons increases exponentially with a time constant depending on the applied electric field and the gas density.[4] Incredibly, this process requires just one runaway electron moving at relativistic speeds to start the entire chain reaction.

Terrestrial Gamma-ray Flashes

Before the evidence from the Burst and Transient Source Experiments (BATSE) on board the Compton Gamma Ray Observatory (CGRO),[5] the belief was that hard X-rays and gamma rays only came from the cosmos. However, it has been shown with this new evidence from CGRO that hard X-rays and gamma rays can originate from Earth's own atmosphere in correlation with lightning activity. RREAs that produce numerous high-energy, relativistic speed electrons colliding with molecules in the atmosphere are believed be a feasible source to these X-rays and gamma ray bursts produced through the brehmsstrahlung process,[6] in which electrons collide with atmospheric molecules at high speeds, emitting high energy radiation. It has been suggested that the thermal runaway electron process operating in leaders may contribute to the production of TGFs.[1]

In the present work we use a full energy range Monte Carlo model to simulate RREAs under application of various magnitude electric fields, and document their rate of multiplication.

MODEL DESCRIPTION

In this work, we have used two Monte Carlo codes. The first one, developed in MATLAB, simulates the propagation of low energy electrons in Argon. This code allows us to get preliminary results on the dynamics of electrons subjected to both elastic and inelastic scattering in a gas and under an applied electric field. The second Monte Carlo code used in this study is a research code developed in the Atmospheric Electrodynamics Group in the department of electrical engineering at Penn State University.[8] It simulates the electron scattering and acceleration in air from sub eV energies to over MeV energies.

Cross Sections and Collision Frequencies

In air, electrons have certain frequencies of collisions with atoms and molecules (i.e., N_2, O_2) and there are several types of collisions that can occur. While more detailed collisions were included in the complete code, those implemented for the Argon model are elastic, electronic excitation, and ionization

collisions. The frequencies of related processes depend on the density of the gas, the velocity of the electrons, and the cross section for a type of collision and gas, which corresponds to a probability of an electron colliding with a particle at any moment in time. In our codes, this cross sectional data must be loaded in from an external source containing the recorded cross sections for gases in air. With this, frequencies for each type of collision can be calculated, and then with these data, we can compute the probability of a given collision during one time step in the simulation.

Collisions

Determining whether a particle will collide or not, what type of collision it will be, and how much energy will be lost is a stochastic process. To simulate these processes, we use a null collision technique.[1] In this technique, the probability P_{coll} of a collision with a molecule is represented by equation:

$$P_{coll} = 1 - e^{-(\frac{\Delta t}{\tau_c})},$$
(1)

where τ_c is the smallest mean time between collisions of an electron and Δt is the time step of the simulation. In order to have a good accuracy in the description of collisions, we choose the time step $\Delta t = \delta \tau_c$ where $\delta \ll 1$. The τ_c is related to maximum collision frequency ν_{max} as:

$$\tau_c = \frac{1}{\nu_{max}}$$
(2)

$$\Delta t = \frac{\delta}{\nu_{max}} = \delta \tau_c$$
(3)

$$P_{coll} = 1 - e^{-(\frac{\delta \tau_c}{\tau_c})} = 1 - e^{-\delta} \cong \delta = .1$$
(4)

We can then define all the other probabilities from their total frequencies of collisions $\nu_t(\varepsilon)$ as relative to ν_{max}. From the comparison between a random number generated between 0 and 1 P_{real}:

$$P_{real} = \frac{\nu_t(\varepsilon)}{\nu_{max}}$$
(5)

One can then determine if the collision is going to be considered real or null. Lastly, if the collision is real, determining the type of collision an electron will have is done by normalizing the frequencies of each type of collision with the sum of all collisions as:

$$P_j = \frac{v_j}{\sum_{j=1}^{n} v_j} \tag{6}$$

where *j* represents a collision type and n is total number of collisions. Then, checking if a collision will occur for a specific electron and type of collision can be done with the following 3 steps after generating three random numbers A, B, and C within the interval from 0 to 1.

1. Is the following statement true? If TRUE, skip step 2. If FALSE, continue to step 2.

$$A < \delta \tag{7}$$

2. Is the following statement true? If TRUE, skip step 3. If FALSE, continue to step 3.

$$B < \frac{v_t(\varepsilon)}{v_{max}} \tag{8}$$

3. For what lowest k is the following statement true? The collision is of the kth type.

$$C < \frac{\sum_{j=1}^{k} v_j}{\sum_{j=1}^{n} v_j} \tag{9}$$

If the program makes it through step 3, that particle will lose a fraction of its energy, depending on the collision type, and also scatter at an angle determined from the differential scattering cross section [1]. Its new velocity components will then be calculated. If the collision is an ionization collision, a new electron will be added to the simulation. Regardless of whether a collision happens or not, the velocity of the electron will be updated due to acceleration by the electric field. This process is performed for every particle within the simulation.

RESULTS
Low Energy Monte-Carlo Model in Argon
    The focus of this numerical experiment is to represent the dynamics of electrons subjected to an applied electric field in a gas as accurately as possible. To verify the results of the Monte Carlo models created, we graphed the Electron Energy Distribution Function as shown below in Figure 2. The simulated time was 1 ns. The results are compared to the well known Boltzmann equation solver, *Bolsig+*[9] for the same conditions of applied electric field and gas density.
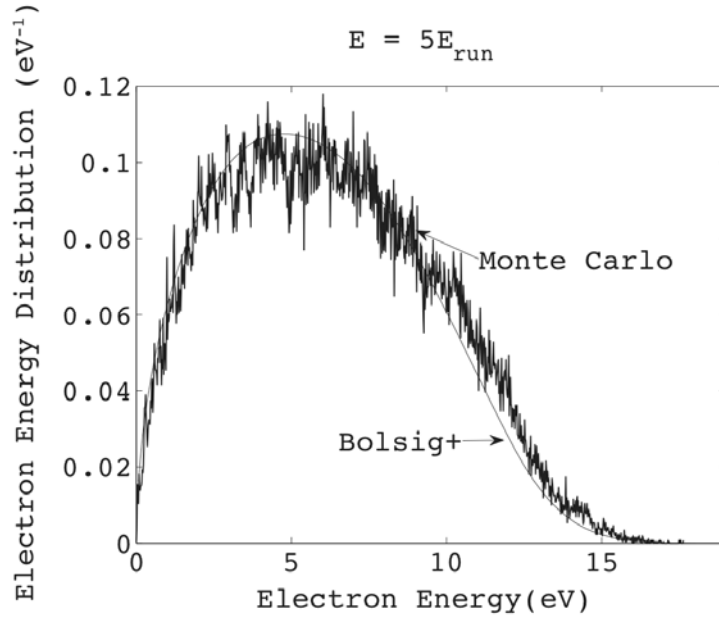
Figure 2: Electron Energy Distribution function in Argon gas for applied electric field with E = $5E_{run}$ = 1.078×10^6 V/m, and gas density N = 2.688×10^{25} m^{-3}, and the number of particles 14,000.

The *Bolsig+* Solver and the Monte Carlo model agree nearly perfectly. This verifies our Monte Carlo models with the established methods.

Full Energy Range Monte-Carlo Model In Air

In this section, we present the results obtained on RREA rates in air for different applied electric fields. The comparison of the electric field applied with the dynamic friction force lead to the determination of the critical runaway energy ε_{run}. We document the values of ε_{run} for the different electric fields studied in this work in Table I. These values have been computed as described in [2].

Table I: Values of the critical runaway energy ε_{run} with respect to the applied electric field E. This parameter δ is defined as $\delta = E / 2.15 \times 10^5$ (V/m).

δ	2	5	8	10	12	15
E (V/m)	4.3127E5	1.0782E6	1.725E6	2.1566E6	2.5876E6	3.2345E6
E/N (Td)	16.0443	40.1108	64.1773	80.2216	96.2659	120.3324
ε_{run} (eV)	4.4944E5	9.5462E4	5.0442E4	3.7688E4	2.9786E4	2.2371E4

Using the full energy range Monte-Carlo code in air, we define N_0 = 8000 electrons at $t = 0$ with an energy of 1 MeV and velocity in the anti-parallel direction of the electric field. The electric field applied E is characterized by

parameter δ = E (V/m)/2.15×10^5 = 2, 5, 8, 10, 12, and 15 for direct comparison with the existing literature [2]. From the values of ε_{run} shown in Table 1, we can track all electrons with energies higher than ε_{run}. The related number of electrons $N_{run}(t)$ is then of exponential form $N_{run}(t) = N_0\exp(\gamma t)$ and we can compute the related time constant γ for a given applied electric field. These rates are presented in Figure 3.
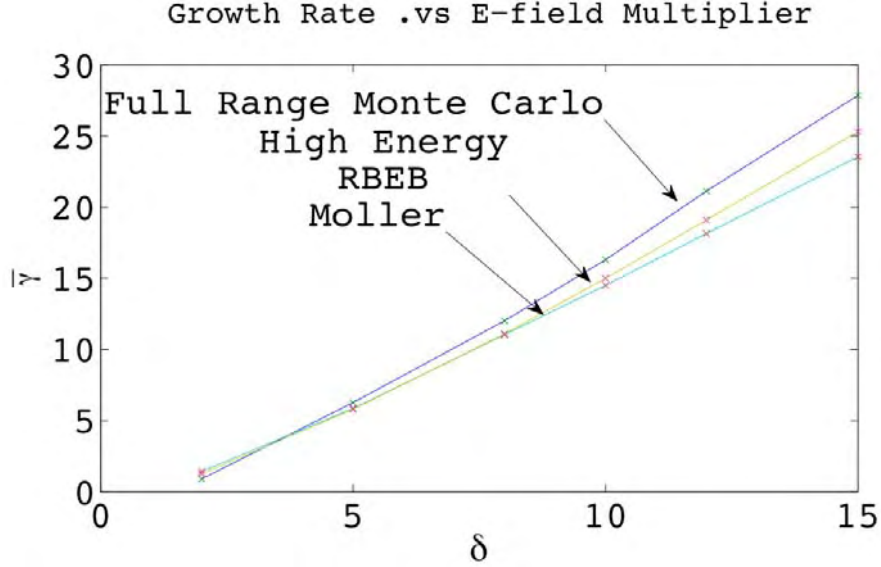

Figure 3: RREA rates presented by $\bar{\gamma} = \gamma \times 171.53$ vs. δ in which δ is a parameter defined by $\delta = E(V/m)/2.15 \times 10^5$ for δ = 2, 5, 8, 10, 12, and 15.[2,3]

DISCUSSION

Low Energy Argon Model

The Monte Carlo code that created the distribution function in Figure 3 was executed for a full 1 ns, which is more than enough time to converge to a steady state. Figure 2 shows that this model agrees very well with the *Bolsig+* solver, assuring the validity of the Monte Carlo code. Our code is expected to have variation due to the stochastic process we are simulating. This explains some small deviations with respect to *Bolsig+* solutions observed in Figure 2. We emphasize that the Argon model simulates low energy electrons only. The significant relativistic effects only become important when the relativistic factor

$\gamma_r = \left[1 - \left(\dfrac{v_e}{c} \right)^2 \right]^{-(1/2)}$, where v_e is electron velocity and c is velocity of light,

significantly deviates from 1, i.e. $\gamma_r \neq 1$. Electron energies in this model reached at most 20 eV, producing a relativistic factor very close to 1.

Full Energy Range Monte Carlo Model in Air

The results obtained in Figure 2 are in good agreement with the existing results in the literature. It is essential to note that the models producing the results shown in Figure 2 are very different. The RBEB and Moller models presented in [3] only simulate high-energy electrons (> 2 keV) and low energy ionizing collisions are taken into account through the continuous approach of a dynamic friction force made for this purpose. In the full range Monte Carlo model used here, all the energies accessible to electrons and related possible collisions are explicitly taken into account. Another difference comes from the fact that the elastic scattering is simulated for every elastic collision in the full energy range Monte Carlo code, although it is taken into account using a multiple scattering model in the high-energy code reported in [3].

It is interesting to note that the simulations reached a limit of the runaway electron count after some duration. In the simulations, the number of electrons is obviously limited, and a remapping technique is used [1,3] so that it is possible to take into account a large number of low energy electrons created. For the sake of comparison, the number of remapped electrons is not changed from one run to another. However, for higher electric fields we have to take into account more runaway electrons, and since the same proportion is remapped, the remapping method introduces a numerical loss term that would finally lead to a steady state of the number of runaway electrons and the growth of electrons stops. Nevertheless, $N_{run}(t)$ grows exponentially during a sufficient time allowing to extract the growth rate.

CONCLUSION

In this work we have used both a low energy Monte Carlo model in Argon as well as a full energy range Monte Carlo model in air. The low energy model in Argon compares very well with the well known *Bolsig+* solver and demonstrates the accuracy of the Monte Carlo codes. The full energy range model was used for a range of applied electric fields characterized by $\delta = E(V/m)/2.15 \times 10^5 = 2$, 5, 8, 10, 12, and 15. The present results are more accurate than those obtained previously with the RBEB and Moller models due to our models' explicit calculations of all types of collisions and full energy range simulations. We observed a limit of the runaway electron count that occurred after some time in the simulations, however, it was still possible to extract a sufficient amount of time steps from the simulations to measure avalanche rates accurately.

ACKNOWLEDGMENTS

REFERENCES

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

[2] Lehtinen, N. G., T. F. Bell, and U. S. Inan (1999), Monte Carlo simulation of runaway MeV electron breakdown with application to red sprites and terrestrial gamma ray-flashes, *J. Geophys. Res.*, *104*, 24699.

[3] Celestin, S. and V. Pasko (2010), Soft collisions in relativistic runaway electron avalanches, *J. Phys. D: Appl. Phys.*, *43*, 315206.

[4] Gurevich, A.V., G. M. Milikh, R. Rousseldupre (1992), Runaway electron mechanism of air breakdown and preconditioning during a thunderstorm, *Phys. Letters A*, *165*, 463.

[5] Fishman, G. J., P.N. Bhat, R. Mallozzi, J. M. Horack, T. Koshut, C. Kouveliotou, G. N. Pendleton, C. A. Meegan, R. B. Wilson, W. S. Paciesas, S. J. Goodman, H. J. Christian (1994), Discovery of intense gamma-ray flashes of atmospheric origin, *Science, 264*, 1313.

[6] Dwyer, J. R., and D. M. Smith (2005), A comparison between Monte Carlo simulations of runaway breakdown and terrestrial gamma-ray flash observations, Geophys. Res. Lett., 292, 64.

[7] Carlson, B. E., Lehtinen, N. G. and Inan (2009), Terrestrial gamma ray flash production by lightning current pulses, *J. Geophys. Res., 114*, A00E08.

[8] Celestin S. and V. P. Pasko (2010), Monte-Carlo Models for Studies of Electron Runaway Phenomena in Air, CEDAR Workshop W#2 ECCR245, University of Colorado, Boulder, CO, USA.

[9] Hagelaar, G. J. M., and L. C. Pitchford (2005), Solving the Boltzmann equation to obtain electron transport coefficients and rate coefficients for fluid models, *Plasma Sources Sci. Technol.*, *14*, 722.

STUDY OF THE POSSIBLE EFFECT OF THE SIGNAL PROCESSING IN THE METEOR-HEAD DATABASE FROM JICAMARCA RADIO OBSERVATORY

Reinaldo Mulero,* Julio Urbina,[#] and Freddy Galindo[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
University of Puerto Rico
Mayaguez, 00680

ABSTRACT

Meteor head echoes studies using the High Power–Large Aperture (HPLA) 50 MHz radar at Jicamarca Radio Observatory (JRO) have been performed from an astronomical perspective over the last few years. For example: *Chau and Woodman, 2004* reported that the radiant distribution of all detected meteor is concentrated in relative small angles centered around the Earth's Apex; and *Chau and Galindo, 2008* showed the first meteor shower observation using a HPLA radar. Although, HPLA radars has shown to be great facilities for meteor studies in comparison to other instruments, there is still a lot of work to be conducted in order to understand their biases and characteristics.

In this work, we present the study of two particular signatures observed in signal to noise ratio (SNR) the meteor-head echoes detected at Jicamarca. The first signature called Signature Type 1 seems to be a result of meteoroid fragmentation but and in-depth study of this type of signature will be left for future work. The second signature called Signature Type 2 shows a relationship between the changes in range (or the value of the radial velocity) and the periodic pattern observed in the SNR. In the most trivial case, the SNR shows ripples that are strongly correlated with the changes in range. This peculiar Range-SNR relationship seems to indicate a possible systematic problem present in the "meteor-head echo-digital receiver" system. We present a simple model that aims to understand the Signature Type 2 effect in meteor echoed and validate the assumption with experimental data.

[#] Faculty Mentor
[+] Graduate Mentor

INTRODUCTION

The return of the Leonid meteor shower in the 90's triggered an increase and interest in meteor studies. This led, on November of 1998, to the first meteor-head studies using the HPLA-VHF radar Jicamarca and since then different meteor campaigns have been conducted in order to study different aspects related to the meteors (e.g., *Chau et al,* 2007; *Oppenheim et al.,* 2009).

When a meteoroid enters Earth's atmosphere, its collision with the air molecules leads to the ionization of ablated material, resulting in the formation of clouds of electrons. JRO is able to detect the radar signal scattered back from the cloud of electron around a meteoroid moving at or near its speed. This radar signals are known as meteor-head echo (*Sparks et al*., 2009).

A process that may observed in meteor-head echoes is called fragmentation, taking place due to thermally induced stresses or the separation of a molten metal droplet from the lower density chondritic compounds of a heated meteoroid (*Kero et al*., 2008). For example: *Kero et al.* (2008) reported two examples of pulsating meteor events observed with the tristatic 930 MHz EISCAT; *Mathews et al.* (2010) concluded that a majority of the studied events exhibit fragmentation.

We report in this paper, two particular signatures observed in signal to noise ratio (SNR) of the meteor-head echoes detected at Jicamarca. Although, our report focuses in the data from May 2007, we would like to indicate that we have looked over different meteor-head echo campaigns (e.g., Dec 2008) and in all these datasets the signatures are presented. The first signature called Signature Type 1 seem to be of the same origin as it was reported in *Kero et al.* (2008) and *Mathews et al.* (2010). The *second signature called Signature Type 2* shows a relationship between the changes in range and the periodic pattern observed in the SNR. In the most clear case, the SNR shows ripples which are strongly correlated with the changes in range. Signature Type 2 seems to be a natural response (or issue) of the system "head echo-digital receiver". In addition, the *head echo-digital receiver* model developed to understand the effect of the like-point movement of the head echo and the digital receiver reproduces most of the main characteristics of the Signature Type 2 (within the limitation of our model). The last allows us to support the idea about a possible effect of the signal processing during the acquisition of the data.

The preliminary results showed that for the meteor-head echo campaign of May 2007 (05 and 06 May 2007), a 0.4% of events shows the Signature Type 1, 22.1% the Signature Type 2 and 77.5% presents no signature. The total population observed for this campaign was 38,007 head echoes. The larger percentage of head echoes that do not show neither signature Type 1 nor 2 is easily understood observing the SNR statistics and noticing that this population show lower SNR. Moreover, because radial velocity (or change in range) defines the shape of the Signature Type 2 and at high radial velocities this shape is hard to distinguish, therefore we do not expect to see the Signature Type 1 and 2 in most

of the events since most of them are detected around the sunrise (i.e., radial velocities around 50 km/s).

EXPERIMENT DESCRIPTION
Meteor-head experiments at Jicamarca

The data was collected using Jicamarca VHF radar located about half an hour drive inland (east) from Lima, Peru and 10 km from the Central Highway (11°5705S 76°5227.5W / 11.95139°S 76.874306°W / -11.95139; -76.874306, 520 meters ASL). The VHF radar that operates at 50 MHz (i.e., 6m of wavelength) and consists of 18432 cross-polarized half-wavelength dipoles occupying an area of ~300 m × ~300 m. Currently, JRO has three transmitters, capable of delivering 1.5 MW peak power each one.

The meteor-head observations have been made using the whole Jicamarca antenna (~300 m × ~300 m) for transmission and three quarters for reception (~150 m × ~150 m). The use of three non-collinear receivers is to locate the meteors inside the transmitting beam (technique called interferometry. The delivered power was ~2 MW. Table I summaries the experimental parameters used in our observations. Notice from Table I that we have used digital receivers. More details about the experimental setup can be found in *Chau and Woodman* (2004).

Table I. Table of experimental parameters

Parameters	Mode 1
Interpulse period (km)	60
Code	Barker-13
Baud width (km)	0.150
Altitude coverage (km)	80 - 120
Sampling (μs)	1
Nyquist Doppler (km/s)	±3.750
Peak transmitter power	2MW
Duty cycle (%)	3.25
Number of receivers	3
Receiver type	Digital

Meteor Parameters

At JRO, the small volume were the meteoroid ionize consists of a narrow truncated cone, defined by the antenna pattern and a properly selected sampled altitude range. At Jicamarca, such truncated cone has ~2° full-width (including the first sidelobe) and a height of 40–50 km around 100–105 km altitude. When the meteor head echo is received, the following parameters are obtained directly: mean signal-to-noise ratio (SNR), duration of meteor event (Δt), radial velocity (v_r), initial range (R_0), change in range (ΔR) and angular coverage ($\Delta \alpha$). Figure 1 illustrates the meteor trajectory for the time that the head echo is detected. Using the Figure 1 and the direct parameters, the extra parameters can be derived:

From the obtained parameter it can be derived:

1. Elevation angle of meteor trajectory (θ) :

$$\theta \approx \arctan\frac{\Delta R}{R_0 \Delta \alpha} \tag{1}$$

2. Absolute velocity ($|v|$):

$$|v| \sim {v_r}\big/{\sin \theta} \tag{2}$$

More details about the technique to characterize a meteor-head echo in *Chau and Woodman* (2004).
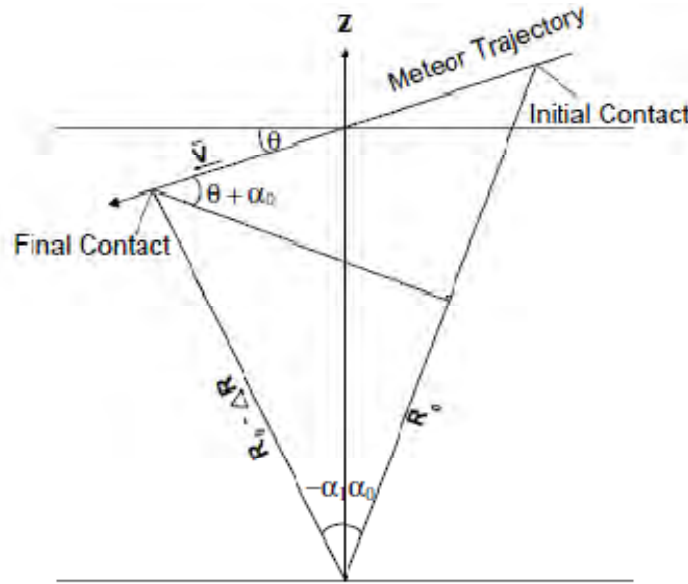

Figure 1: Simplified geometry of a meteor trajectory with respect to the illuminated beam

Meteor Characterization Graphics

Figure 2 shows the main meteor-head parameters: (a) Left, Range versus time plot, (b) Middle, Meteor trajectory in the radar's plane, in addition the radiation pattern of the antenna is showed. Noticed the defined color for the meteor trajectory is related to the value of the SNR, (c) Right, The Pulse-to-pulse Doppler velocity and SNR. For the case of the SNR, is the beam pattern model.

Signatures

To classify the signatures, the information of the panels (a) and (c) in Figures was used. Signature Type 1, are all the events which large changes in SNR (> ~4

dB) and these changes are not correlated with the changes in ranges. Signature Type 2 shows periodic pattern in the SNR. In the easier case (at low radial velocities) this pattern is strongly correlated with the changes in ranges and behaves like ripples. Figure 3 presents an example of Signature Type 1, in the dark square we can clearly see that there is no relation between the changes in power and the changes in range. Figure 4 presents an example of Signature Type 2, as it can be seen inside the dark square, for every change in SNR there is a respective change in range (left).

For this work 38,007 graphs where manually classified. Results showed that for the meteor-head echo campaign of May 2007 (05 and 06 May 2007), a 0.4% of events shows the Signature Type 1, 22.1% the Signature Type 2 and 77.5% presents no signature.
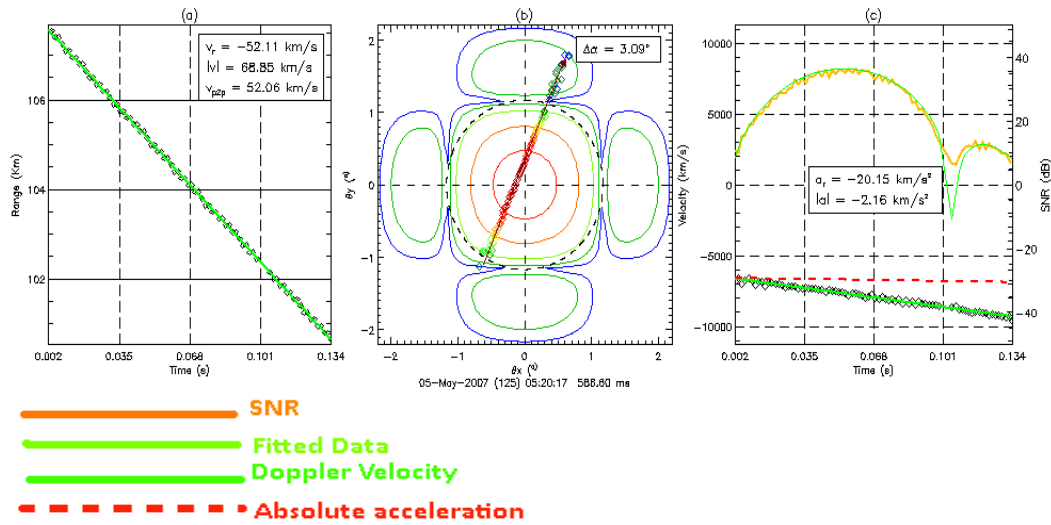
Figure 2: Meteor characterization for most meteor events. (a) Range vs. time, (b) _x vs. _y, and (c) pulse-to-pulse Doppler vs. time. In all three cases there is a fitting line. In panel (c) the SNR is shown a long the meteor trajectory (scale on the right), while the absolute acceleration is given by the slope in dashed lines.
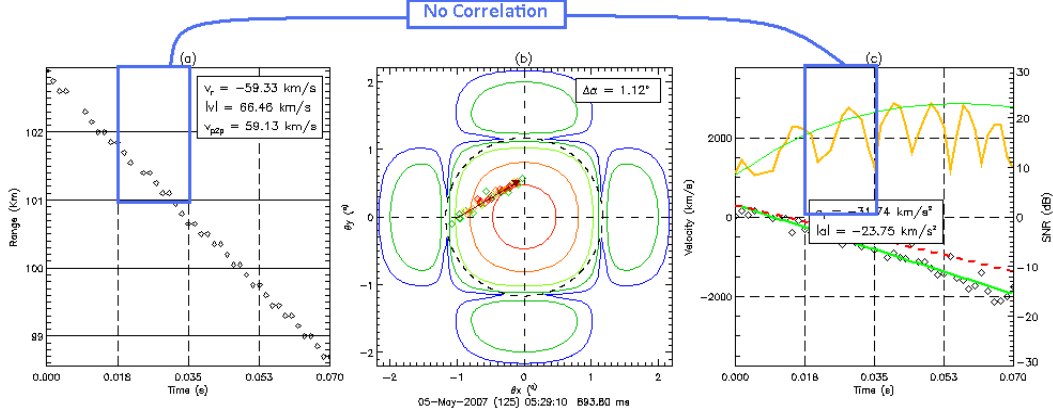
Figure 3: Signature Type 1 example. Notice that there is no relation between the SNR variation and the changes in range, this is possibly cause by fragmentation.
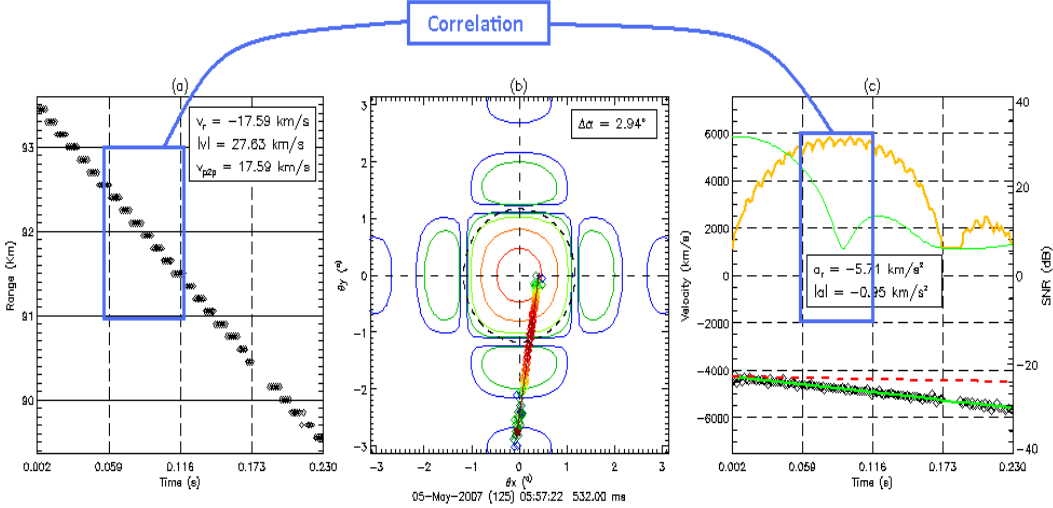

Figure 4: Signature Type 2 example. Notice the correlation between the beat rising in the SNR and the change in range.

Meteor-head echo-Digital receiver Model

The *Meteor-head echo-Digital receiver* model was worked in two schemes. In the first scheme, we model the head echo as a hard target (or copper ball). This definition allows us to compute the amplitude and phase of the received signal using the bistatic radar equation and the distance flew by the electromagnetic wave respectively. Therefore modeling the received signal for different times (defined by the experiment setup: IPP and Sampling Rate), we can follow the time-evolution of the echo because its movement. Additionally, received samples are weighted using the beam pattern model of Jicamarca's antenna. To remark, the meteor's movement is considered as uniform rectilinear motion.

SIGNAL PROCESSING IN THE METEOR-HEAD DATABASE FROM JICAMARCA

In the second scheme, the modeled data sampled at high rate (60 MHz) pass through a Down-Converter (See Figure 5) to produce the resolution required by the experiment. For our current model, only three stages of the Down converter have been modeled: CIC Filter, CFIR Filter and PFIR Filter, and their decimation processes. Notice that the data is modeled in baseband. Figure 6 shows an example of a small radial velocity (i.e. small changes in range). Notice that the ripples (the easiest shape of the Signature Type 2) seem to appear as a natural response of the system "head echo-digital receiver". On the other hand, Figure 7 shows an example of a head with a larger radial velocity. Notice that, although there is no a relationship between the changes in range and the Signature Type 2, there is still a periodic pattern observed in the SNR. Like the radial velocity, the observation of the Signature Type 2 will depend of different parameters (Not show in this report). For example, we expect to see more events with ripples if the sampling rate is larger since the meteor-head echoes will spend more time to travel from one range to the next.
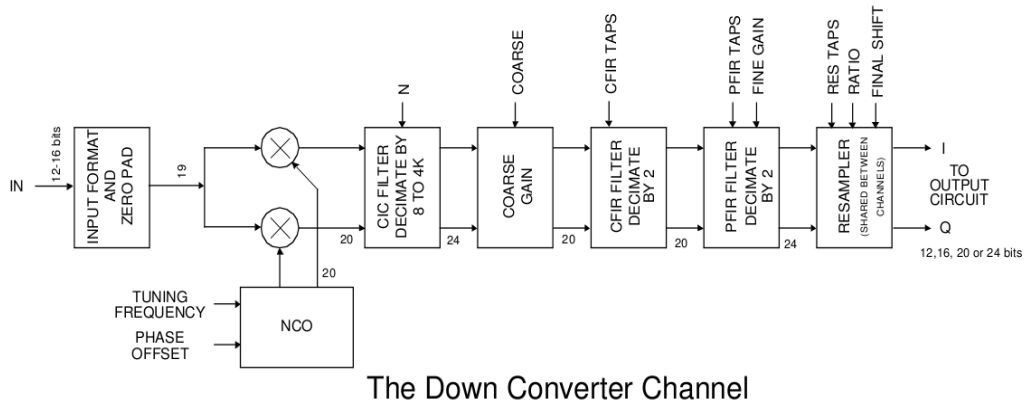

The Down Converter Channel

Figure 5: This flow chart shows the different stages used in the down converter channel. Notice that modeling purposes we have considered some stages as ideal.
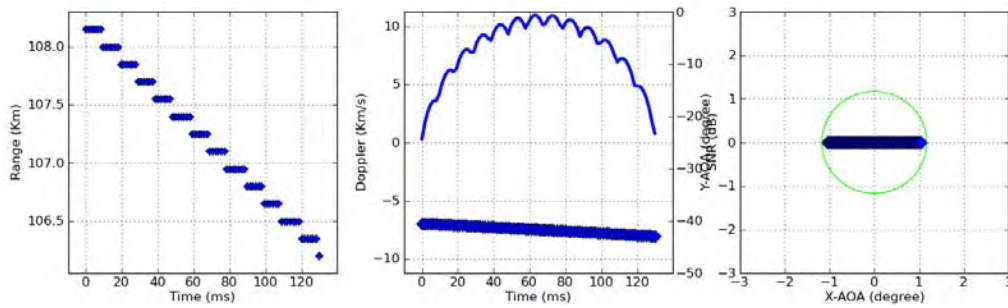

Figure 6: Modeled parameters obtained for an event with a small radial velocity. Notice that meteor is only traveling in the mainlobe.
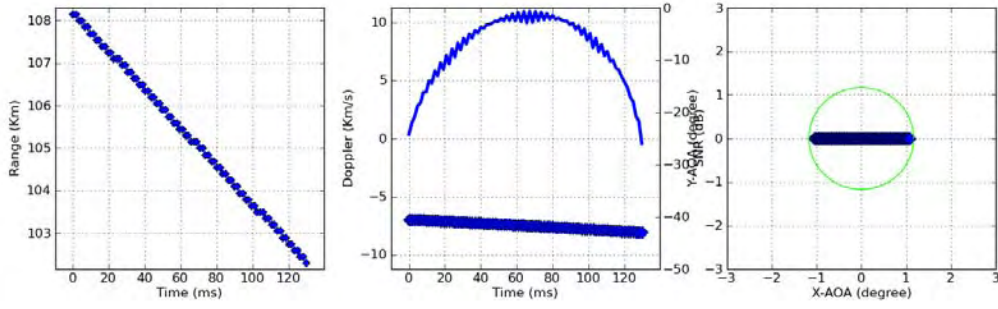
Figure 7: Similar to Figure 6 but showing an event with larger radial velocity. Notice that there is no relationship between the changes in range and the Signature Type 2, but we still see a periodic pattern.

RESULTS

Figure 8 presents an experimental data of Signature Type 2 for the case where the relationship between the SNR and the changes in range are easy to see. It can be noticed the similarities in the shape of the signature in Figure 8 and the model in Figure 6. For both graph it represent the detection of a meteor passing through the main lobe with small radial velocity. Figure 9 shows another example of experimental data of Signature Type 2. In this case the radial velocity is higher than the Figure 8. Notice that although Figures 9 and 7 are not equal, both show the same features, supporting the idea of the systematic issue in the data. The previous examples allow to valid the "head echo-digital receiver" model and additionally, the model will help as tools to look for difficult patterns.
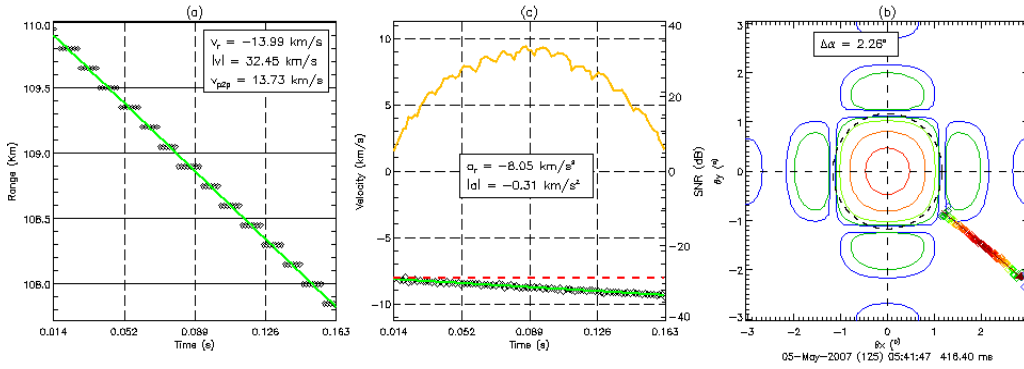

Figure 8: Experimental data of Signature Type 2 for small radial velocity similar to the model in Figure 6.
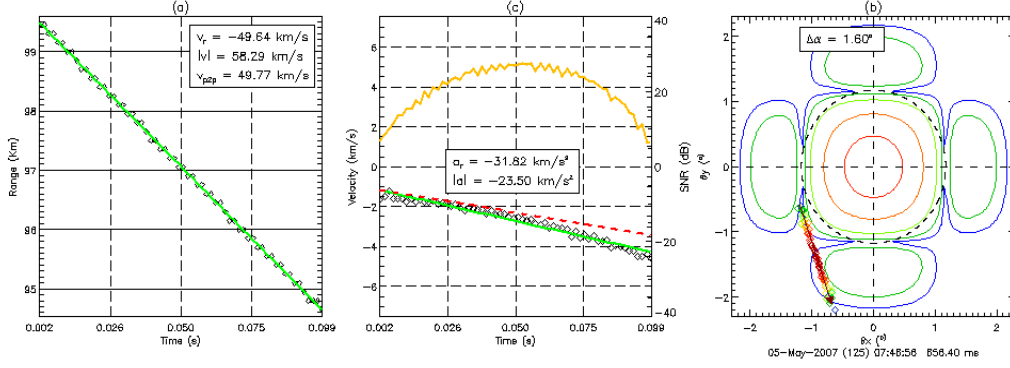
Figure 9: Experimental data of Signature Type 2 with larger radial velocity similar to the model in Figure 7

In order to understand and compare the population showing both signatures and no signatures, histogram plots are presented in the Figure 10. To make a better comparison of the populations of different scales, the histogram presented were normalized. The parameters showed in this Figure are: (a) time, (b) mean SNR, and (c) radial velocity. In the next lines we summary the most important aspects:

1. Both signatures are observed at any time of the experiment (left panel of the Figure 10). This means that the occurrence of both signature are independent of the time. The larger population around ~6 hrs LT is because during this time the radar detects the largest number of meteor and therefore there is more chance of observing both signatures.
2. Both signatures have stronger SNR than the events without any signature (middle panel of the Figure 10). In a conservative approach we can say that SNR from events without signatures are ~5 dB less than events showing one of the two types of signatures.
3. The histograms of the radial velocity (right panel) show similar results in the three cases. The important facts here are: (a) Signature Type 1 and 2 are observed at any radial velocities which mean that both signatures are not exclusive of especial meteor events, (b) Although we expect to see easily the Signature Type 2 as a ripple for low radial velocities, most of the events shows larger radial velocities. This fact is understood since most of the population are detected with large radial velocities and from the model some pattern of the Signature Type 2 at high radial velocities are easy to differentiate (See Figure 9).
4. From the histogram of radial velocity and SNR, we can clearly see why most of the population does not show any signature. Briefly, most of the population shows low SNR and high radial velocities which in terms of signatures mean in most of the events is really hard classify the event as one of the signatures. The model shows at high radial velocities the amplitude of the pattern observed for the Signature Type 2 decrease and it

is more complex. Adding the noise effect, the Signature Type 2 would be impossible of observing. Statistics and modeled results are consistent.
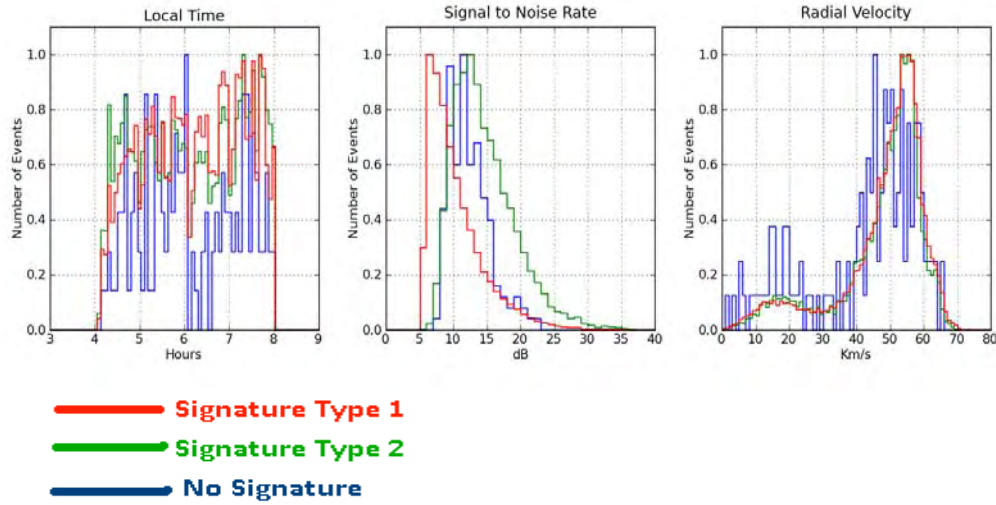

Figure 10: Histogram of the distribution for meteor head echo parameters: (Left) time, (Middle) mean SNR and (Right) radial velocity.

The meteor-head echo-digital receiver model was applied to the best examples of Signature Type 2 in the experimental data. Figures 11, 12 and 13 present some examples from different head echo. Looking at the parameters in all three figures we have the experimental data and the model. Looking at the SNR (Middle panel) and Range (Right panel) in Figure 11 it can be clearly seen the SNR-Range correlation. In Figures 12 the radial velocity is higher but the correlation is still observed. Figure 13 has the highest radial velocity from the last two figures, notice that some SNR-Range relationship still been present in the periodic pattern. Notice the similarities in the shape of the periodic pattern in the all three Figures for both experimental data and its model. In Figures 11 and 13 (Middle panels respectively) we can notice a discrepancies in the Doppler velocity between the experimental data and the model, this is due because the model take the meteor as a target moving at constant velocity, not taking in consideration the deceleration of the meteor. Also notice that since the experimental data was no calibrated there is a change in phase in the meteor trajectory along the radar pattern (Left panel) between the experimental data and the model.
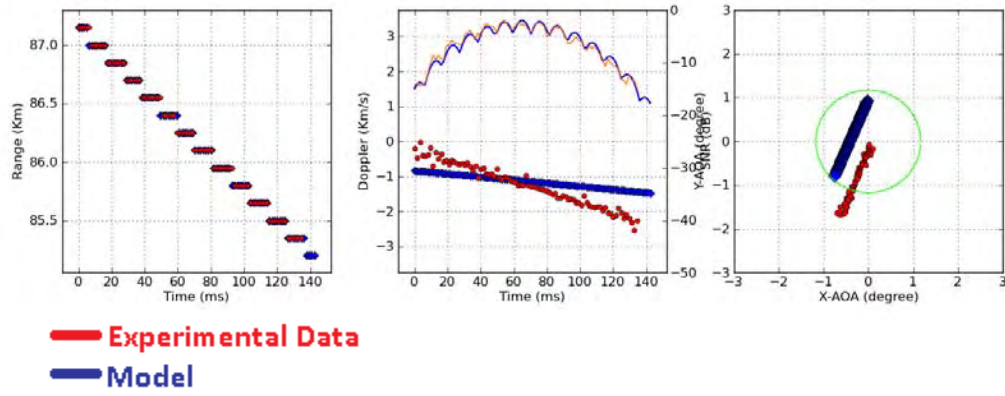
Figure 11: Modeling of experimental data of Signature Type 2. Notice SNR-Range correlation in both experimental data and in the modeling the middle panel.
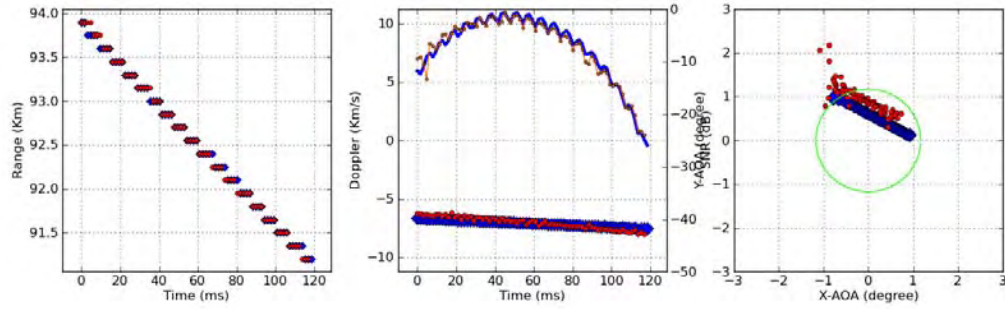

Figure 12: Same modeling for another event of Signature Type 2 with greater radial velocity. As in Figure 11, notice the same correlation in the ripple for both the model and experimental data.
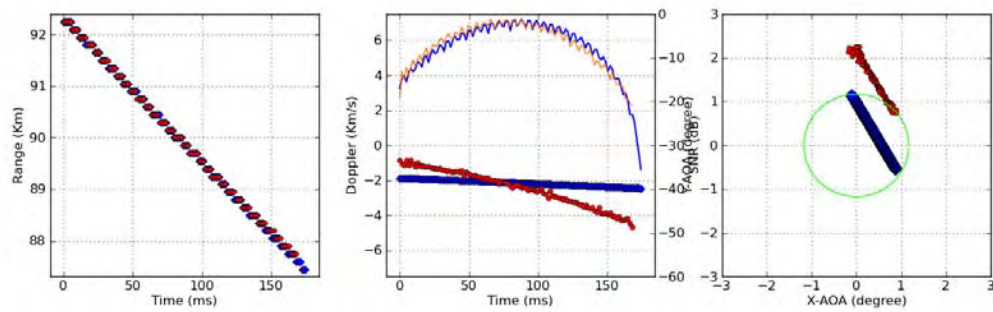

Figure 13: Another Signature Type 2 head-echo but with higher radial velocity than the last two Figures above. Notice some SNR-Range relationship in the periodic pattern.

CONCLUSION

We have shown that Signature Type 2 is present at any time in the datasets analyzed in this report and it is not an isolated event. With low SNR values, this signature is difficult to detect. The inclusion of Signature Type 2 is critical in meteor analysis since ignoring it could lead to misinterpretation and wrong conclusions in meteor physics.

Both signatures allows to define a very simple way to discriminate between a physical (i.e., meteor fragmentation event) and systematic (i.e., because the like-point movement and the digital receiver) result: Signature Type 1 shows changes in SNR larger than ~4 dB (we have defined this value in a very conservative way) and most of the time there is no relationship between the change in SNR and the changes in range. We have seen (not showed in the report) some cases where the SNR ripples and the changes in range hold a relationship but the changes in SNR are larger than ~4 dB. For these cases, we are confident that it is because of a physical result. In addition, the model supports our hypothesis. Signature Type 2 has a periodic pattern (too periodic to think as a result of a natural process) and in the easier case this pattern behaves likes ripples strongly correlated with the changes in range.

The modeled results support the idea that Signature Type 2 is the result of a systematic problem because of the system "head echo-digital receiver". Most of the main features observed in the experimental data are obtained with the model. Figure 11, 12 and 13 are some of these examples. Notice that the authors tried to model the experimental data in the best way, which is not easy, and in some cases, there is a small delay. But the shapes of the examples are in good agreement. The authors are still working in a mathematical model to understand this effect. But we are confident that it is a systematic problem.

REFERENCES

J. J. Sparks, D. Janches, M. J. Nicolls, C. Heinselman, Seasonal and diurnal variability of the meteor flux at high latitudes observed using PFISR doi:10.1016/j.jastp.2008.08.009 (2009).

J. Kero, C. Szasz, A. Pellinen-Wannberg, G. Wannberg, A. Westman, D. Meisel, Three-dimensional radar observation of submillimiter meteoroid fragmentation, doi:10.1029/2007GL032733 (2008).

J. L. Chau and R. F. Woodman, Observations of meteor-head echoes using the Jicamarca 50MHz radar in interferometer mode, SRef-ID: 1680-7324/acp/2004-4-511 (2004).

J. L. Chau, R.F. Woodman and F. Galindo, Sporadic meteor sources as observed by the Jicamarca high-power VHF radar, *Icarus* **188**, pp. 162–174 (2008)

J. L. Chau, R. F. Woodman, and F. R. Galindo, Sporadic meteor sources as observed by the Jicamarca high-power large-aperture VHF radar, *Icarus*, **188**, 162-174 (2007).

M. Oppenheim, G. Sugar, N. Slowey, E. Bass, J. Chau, and S. Close, Remote sensing lower thermosphere wind profiles using non-specular meteor echoes, *Geophys. Res. Lett*., Vol. 36, L09817, doi:10.1029/2009GL037353 (2009).

J. D. Mathews, S. J. Briczinski, A. Malhotra and J. Cross, Extensive meteoroid fragmentation in V/UHF radar meteor observations at Arecibo Observatory, *Geophys. Res. Lett*., 37, L04103, doi:10.1029/2009GL041967 (2010).

VHF/UHF RADAR METEORS AT ARECIBO: THE USUAL SUSPECTS AND SOME BOLIDES

Jonathan Ore,* John D. Mathews,[#] and Akshay Malhotra[+]

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

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

ABSTRACT

Radar meteors are observed at Arecibo Observatory using both the 430 MHz Ultra High Frequency (UHF) and 46.8 MHz Very High Frequency (VHF) radars. These meteor events, viewed using RTI (Range-Time-Intensity) images, often exhibit very complex structures. This complex "head-echo" structure is consistent with multiple close fragments that each produce a head-echo. These head-echoes in turn interfere with the other head-echoes, thus producing the complex echo structure. This interpretation is confirmed by viewing events at both radar frequencies and by noting that trail echoes are often produced at VHF. A wealth of such results is illustrated within this paper. Additionally, some unusual events that are interpreted as being "fossil" radar bolide (fireball) events are discussed.

INTRODUCTION

Radar Meteor Science

An enormous amount of material flux enters the Earth's atmosphere every day; in fact, Mathews et al. predicts it to be between 1.6 x 10^6/2.7 $\times$ 10^6 kg/yr [1]. Some of the particles that come into contact with the atmosphere heat up so much that they produce light through a process known as *ablation*. When this happens, it is traditionally denoted as a *meteor*. A meteor is not to be confused with a meteoroid, which is the actual solid particle undergoing ablation. Meteoroids range greatly in size, varying as small as micrometers up to several kilometers (sometimes called bolides). If a meteoroid manages to survive the descent to Earth's surface and makes impact, it is called a meteorite. Meteorites are not only

---

[#] Faculty Mentor
[+] Mentor

useful as far as amateur enthusiasts are concerned, but they also have the potential to reveal many insights into planetary science.

Since the vast majority of meteoroids are very small, it is very difficult to study them independently. With the inception of RADAR (radio detecting and ranging), smaller meteoroids became much easier to study, but there was still yet a better way for analysis. The process of ablation creates a significant amount of plasma which radars are very adept at picking up. The plasma is formed from ionized particles around the meteoroid and provides a significantly larger target for the radar to detect. The "echo" that the radar observes from the plasma surrounding the meteoroid is called the "head echo" of the meteor and allows for measurement of many vital parameters (altitude, velocity, etc). Other echoes are possible too; anomalous trail echoes (also known as Range-Spread Trail-Echoes RSTE) are formed from trail of plasma left behind the meteor as it travels.

Meteor Modeling

The velocity trajectory and ablation calculations are vital for meteor studies. They help to model how meteoroids interact with the environment, and also aid in comprehension of their own inherent properties (density, size, etc.). The following three differential equations developed by Ceplecha allow for investigation of these properties:

$$\frac{dv}{dt} = -\Gamma A \varrho_d^{-2/3} \varrho m^{-1/3} v^2 \tag{1}$$

$$\frac{dm}{dt} = -\frac{\Lambda A}{2\xi} \varrho_d^{-2/3} \varrho m^{2/3} v^3 \tag{2}$$

$$\frac{dh}{dt} = \frac{l - A/2}{B/2 + h} v \tag{3}$$

$$\frac{A/2 - l}{B/2 + h} = \cos z(t) \tag{4}$$

$$Al + Bh + C = l^2 - h^2 \tag{5}$$

The first three equations represent a solitary non-fragmenting body traveling through the atmosphere. Each variable is defined and measured as follows: $v \equiv$ velocity (m/s); $t \equiv$ time (s); $m \equiv$ mass (kg); $h \equiv$ height (m); $l \equiv$ trajectory distance (m); $\rho \equiv$ air density (kg/m^3); $z \equiv$ inclination trajectory to the perpendicular (m); $\Gamma \equiv$ drag coefficient (unitless); $\Lambda \equiv$ heat transfer coefficient (unitless); $A = Sm^{-2/3}\rho_d^{2/3}$ represents the shape factor; $\rho_d \equiv$ bulk density (kg/m^3); $S \equiv$ head cross-section (m^2), $\xi \equiv$ energy required for ablation of a single unit mass (J). Equation (1) represents the drag-equation, Equation (2) shows the loss of mass due to

ablation, and equation (3) displays the change in height as the meteoroid travels. Equations (4) and (5) are supplements to the previous three differential equations. This model of basic meteoroid travel and behavior helps clarify the complexity of the overall situation – the previous equations describe a very simple setup. It assumes continuous ablation and no fragmentation. If these conditions are not provided, the magnitude of complexity rises rapidly. Fragmentation models are one of the key research areas in meteor studies presently. As we have shown, a great deal of mathematics and physics is required to comprehend meteors' complex behavior. This is a great insight into exactly how challenging the atmosphere is to study [2].

Fragmentation

Fragmentation is a subject of great importance in meteor studies. While still not completely understood, many observations clearly outline its existence. For example, we know that flare trails are a clear sign of fragmentation. Intrapulse fading and interference patterns in RTI images show that the meteoroid fragments are transitioning in and out of phase with each other. The images reported here illustrate a very common phenomenon often observed in optics: Young's Double Slit Experiment. Since the interference patterns we observe are relatively simple, Occam's Razor suggests that the model we use to describe it be relatively simple as well. Our attempt is outlined under Figure 5 in Equations (7) and (8) [7].

Fresnel Zones

Some additional physics are required to understand the following models and graphs. For example, Fresnel Zones are ellipsoidal zones formed by the radar aperture. They define the radiation pattern which results from diffraction of the radio waves from the aperture. Essentially, odd numbered zones will result in constructive interference, while even numbered zones result in destructive interference. Fresnel Zones are extremely important when dealing with radar studies—they are necessary to comprehend the radar equation and for understanding specific meteor characteristics (plasma dispersion and reflection as functions of waves versus plasma frequency) [2]. Fresnel Zones also become very significant when the scatterers we are observing are smaller than the size of the wavelength. This is illustrated by the Fresnel Zone equation below, where λ is the wavelength (m) and n is the Fresnel Zone number:

$$F_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1 + d_2}} \tag{6}$$

The diffraction of waves associated with these zones motivates another very important topic, discussed next.

Wave Diffraction

Diffraction is the natural tendency of an object to bend around an obstacle, as proven by Young's Experiment (results shown in Figure 1). This phenomenon is very evident in the following meteor figures from their interference patterns. Also necessary to understanding these patterns is the radar equation given by (6). To build this equation, we start with a non-directional power density. If we multiply this amount by the gain of the radar, we can obtain the directional power density. Multiplying again by a constant (relating to the radar cross section), σ (m²), we can obtain the reflected power of the radar. In a parallel fashion, we can determine the power received at the receiver. Using the radar aperture area and its efficiency, we can then rearrange our equations to solve for the power received, as displayed by (1). Having an intuitive understanding of the radar equation is essential to comprehending radar meteor studies.

$$P_r = \frac{P_t G_t A_r \sigma F^4}{(4\pi)^2 R_t^2 R_r^2} \tag{7}$$
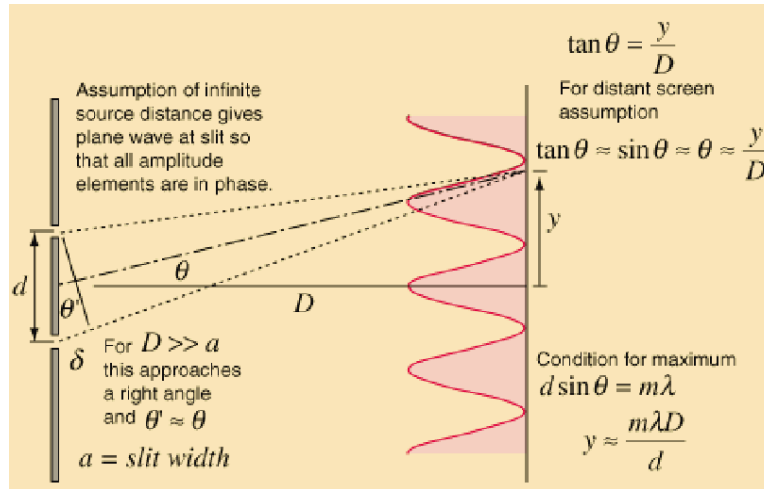

Figure 1: An illustration of Young's Double Slit Experiment.

The purpose of this particular study is to model meteor events, classify them, and also to comprehend the radio science behind their interactions with the atmosphere. Examples include the fragmentation of the meteoroid itself, as well as how it interacts with the atmosphere. Models will be able to demonstrate the fundamental scientific principles occurring and also to reveal still yet unknown processes that have the ability to shed great knowledge and insight into the study of meteors and radio science as a whole.

VHF/UHF RADAR METEORS AT ARECIBO

EXPERIMENTAL SETUP

The data examined in the project comes from the HPLA (High Power Large Aperture) radio telescope at the Arecibo Observatory. The data was obtained in February of 2010 and yielded around 17,000 events. Using data processing programs developed by Mathews *et al* and Briczinski *et al.*, meteors can be observed at high detail and with relative magnitude to the surrounding atmosphere. RTI (Range Time Intensity) plots are used to identify different meteor echoes and trails for classification purposes. The Arecibo Observatory uses two radars (which act in tandem); a UHF at 430MHz and a VHF at 46.8MHz. The UHF is a significantly more sensitive radar, but enjoys a very small beam width (1/6 degree). The VHF has a much wider beam width (1.2 degrees) but is less sensitive. Both radars have an inter pulse period (IPP) of 1ms. The two radars coupled together allow for greater perspective on the meteor events, with UHF sometimes picking up very faint events and the VHF revealing spectacularly long events like bolides. Images of the Arecibo Observatory and its radars are shown below:

Figure 2: Arecibo Observatory located in Puerto Rico.

Four 46.8 MHz Yagi's mounted coaxial to the 430 MHz linefeed

Figure 3: The 430 MHz UHF and 46.8 MHz VHF radars.

EXPERIMENTAL RESULTS

As explained earlier, many meteor images were captured during the radar observations conducted in June 2008 and February 2010. While there are many algorithms available for artificial inspection, personal observation is the source for the following images. This was done so that more atypical events that an automated routine is less likely to pick up were not missed. In addition, modeling of head echo and fragmentation events are shown below. They also include Doppler shifting to enhance their accuracy.
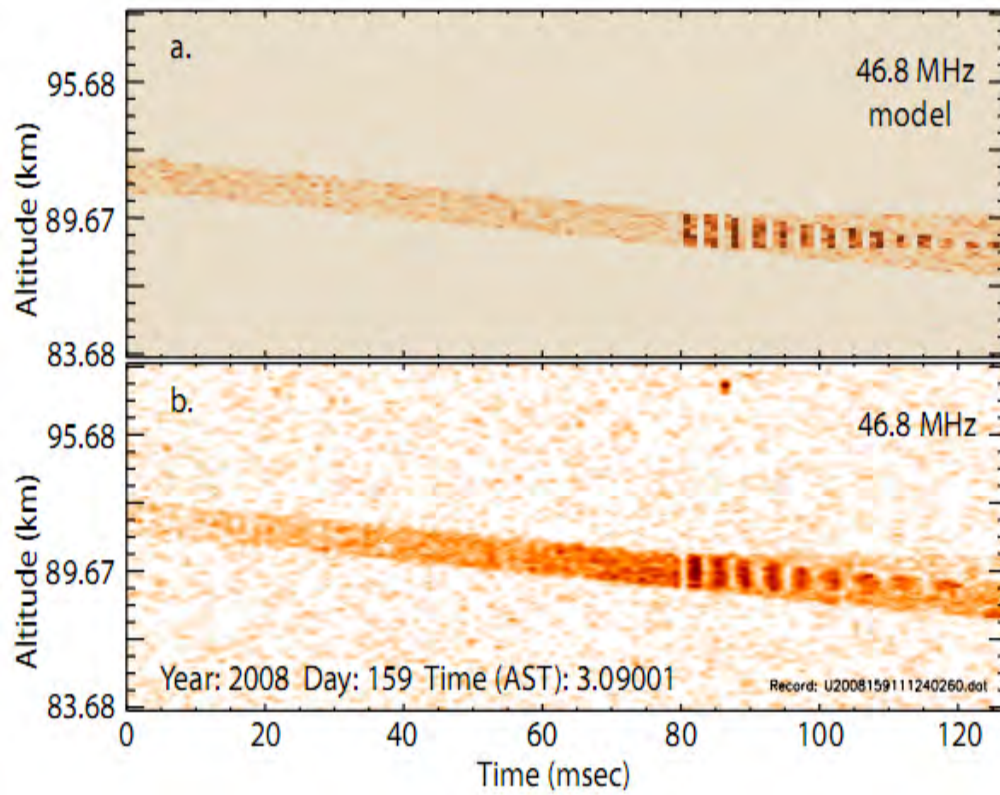
Figure 4: An EMP modeling of a head echo event. Fragmentation becomes very clear after 80ms, when the RTI graph begins to display an interference pattern.
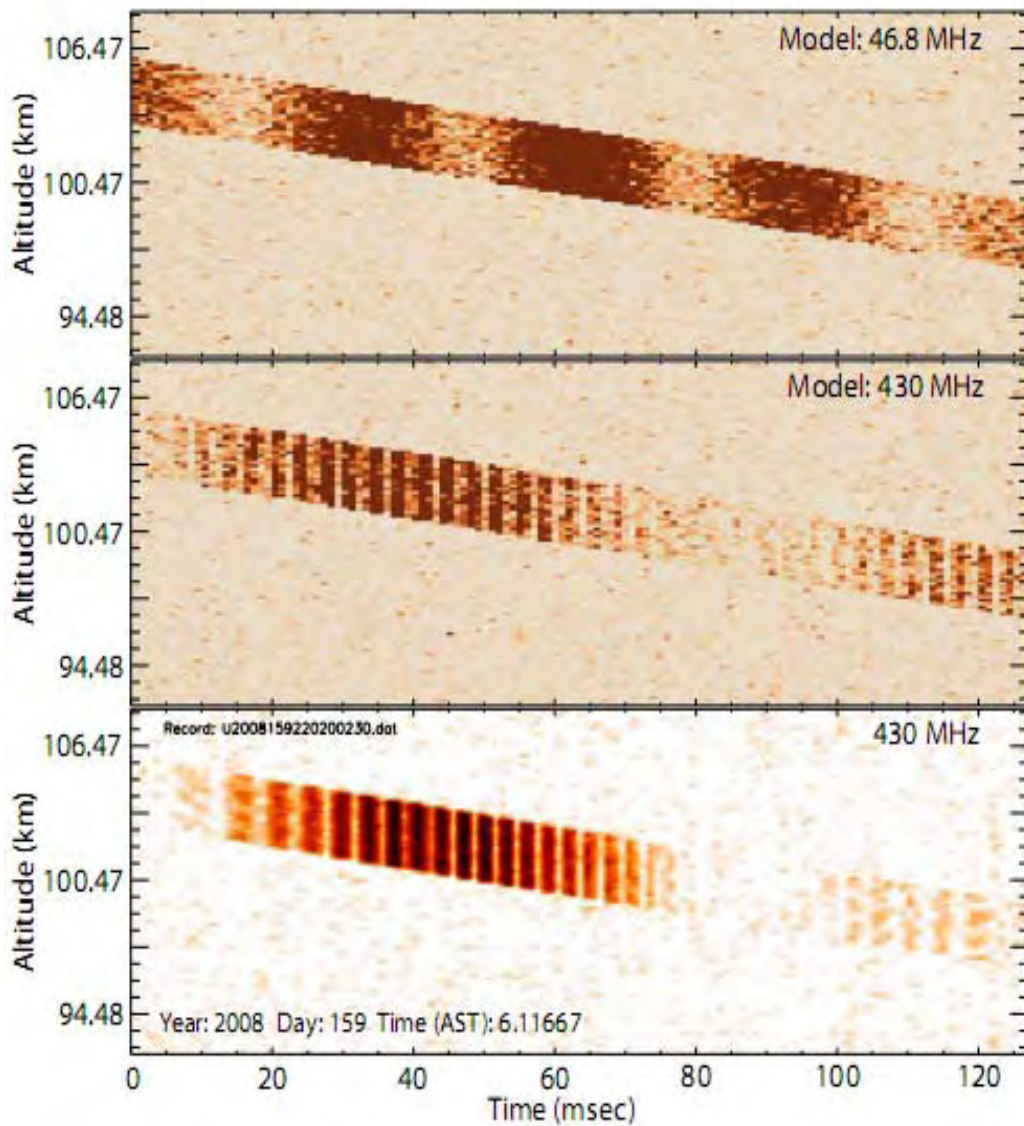
Figure 5: A successful modeling of a fragmentation event using the equations given below. To obtain these equations, we first determine the Doppler Frequency Shift. Multiplying by two accounts for reflection, and removing the clock signal itself through heterodyning leaves us with (8). In reality, the exponential is a sum of sine and cosine (Euler's Formula), which stores the signal's information. By using Quadrature and applying a baseband filter, we can acquire the Doppler effects (taking the derivative of the exponential in (8)) on the signal and obtain the interference pattern displayed in the figure.

VHF/UHF RADAR METEORS AT ARECIBO

$$x_n(t) = A_n \exp\left(\frac{j4\pi R_n(t)}{\lambda}\right) \quad (8)$$

$$R_n(t) = R_n(t_0) - v_n(t - t_0) + d_n(t - t_0)^2 / 2 \quad (9)$$
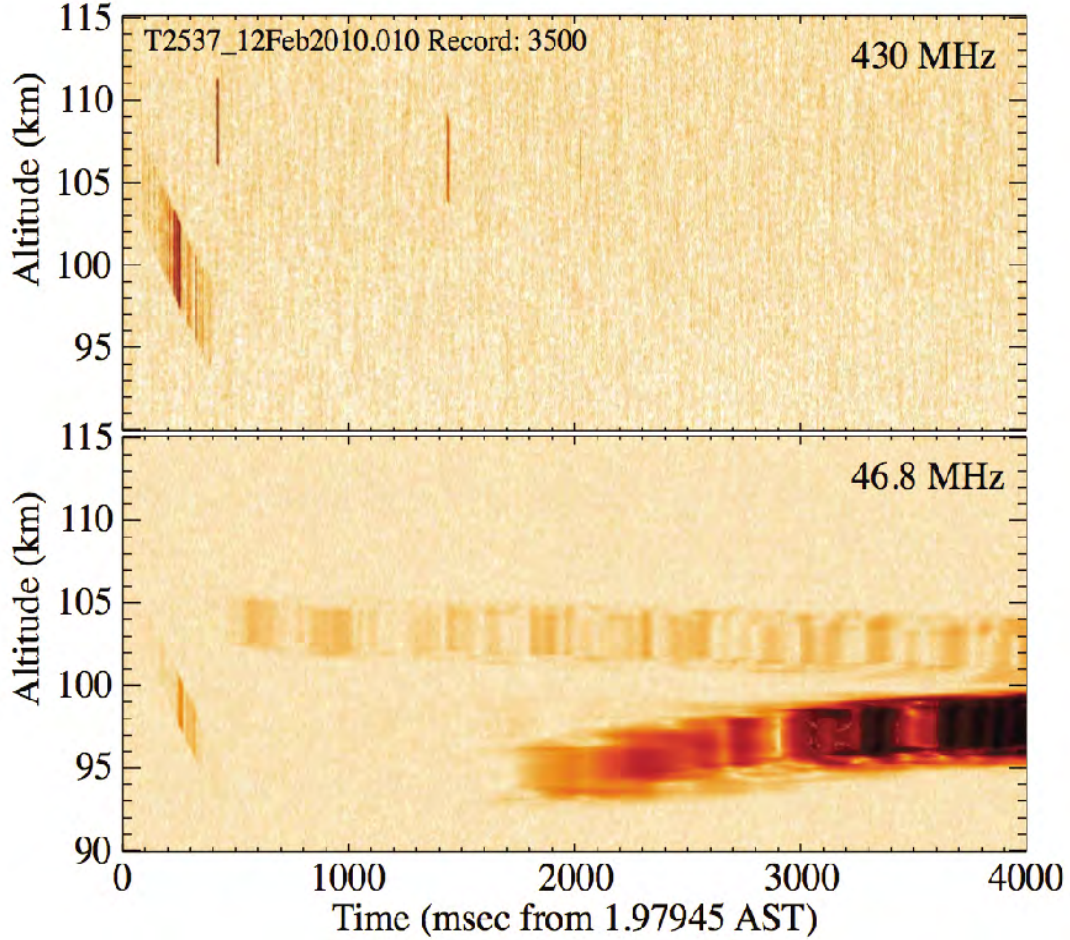

Figure 6: A striking fossil bolide event, caught with very distinctive features in the VHF image (bottom). It is not entirely clear whether the head echo in the UHF image (top) is responsible for creating the trail. The entire event lasts about 10 seconds.
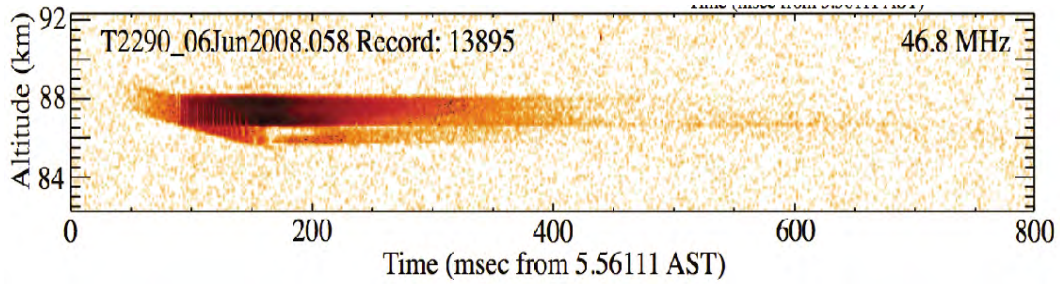
Figure 7: A VHF image which illustrates a highly defined flare and terminal trail resulting from a head echo. There is intrapulse fading between the head echo and trail echo, which gives evidence that the trail plasma is acting as a point scatterer.
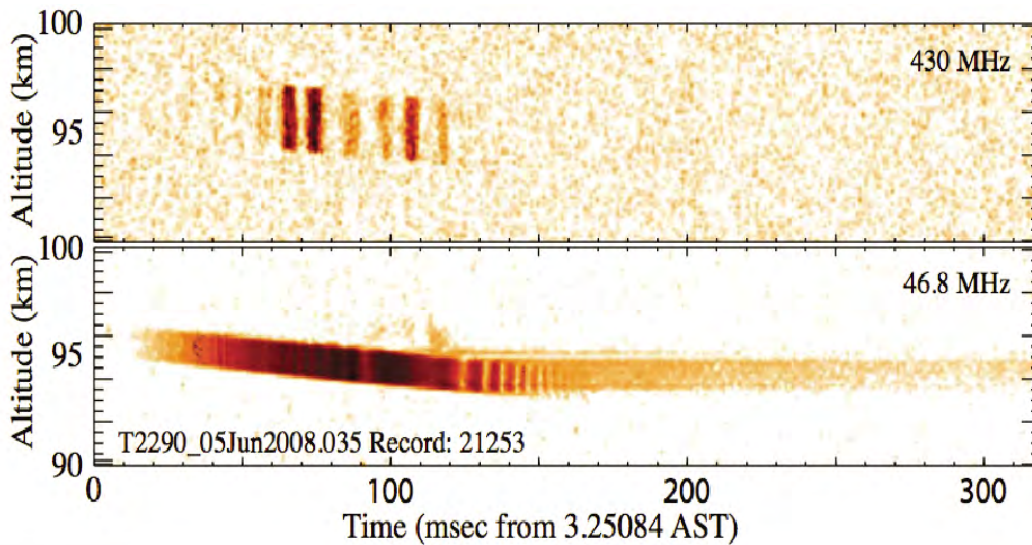

Figure 8: A complex volume fragmentation event which is somewhat common. The trail is only caught at VHF and not UHF - an issue of great interest in radio science.
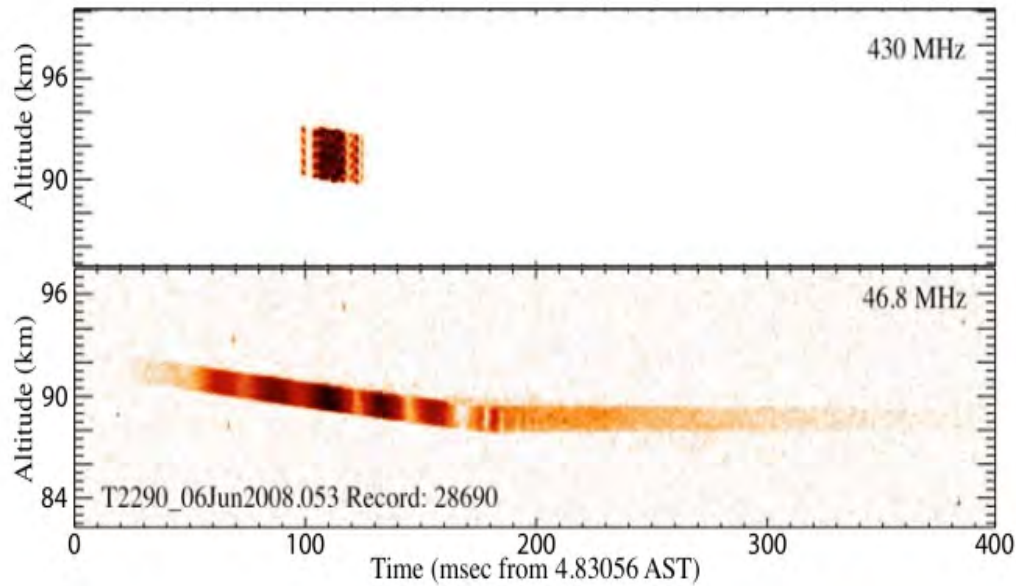
Figure 9: The UHF radar (top) reveals a strong intrapulse fading signature, while the VHF radar displays a terminal flare after about 160ms. Also interesting to note in the VHF radar image is the presence of a LATE (Low-Altitude Trail-Echo) as observed at Jicamarca [3].

DICUSSION:

Since both the UHF and VHF signals are electromagnetic waves, they follow exactly the same behaviors that light does. For example, as discussed earlier, signals that are similar in wavelength can undergo constructive and destructive interference (as displayed by Young's Experiment). In the events above, we can see that the radar scattering targets create interference patterns by acting as the "source" of the echo signals (Huygens-Fresnel Principle). A clear instance of this principle is seen in Figure 5, where the head echo and trail echo create pulse-to-pulse interference around 125 IPP. From the implications of our analysis, it is likely that the meteors are experiencing fragmentation, the specific radio science of which is explained in Mathews [4]. In Figure 3, we show a successful representation of a meteor head echo modeling through Young's Experiment. The model incorporates the important information from the signal in complex form (which allows for Doppler Shift measurement and complete signal information storage) and then presents the interference pattern in the upper part of the graph.

Additionally, events shown here are primarily dominated by fragmentation. In fact, a study by Mathews reports fragmentation for around 90% of models [5]. However, another study performed at RISR (Resolute Bay Incoherent Scatterer Radar) yields classifications of about 48% fragmentation, 32% simple ablation, and 20% differential ablation. Explanations to this significant contrast are given by Mathews and Malhotra in which they acknowledge Arecibo Observatory's

significantly higher sensitivity as compared to RISR. This results in detection of much smaller meteoroids which undergo more frequent fragmentation [6].

CONCLUSION

In all of the figures above, signal interference as a result of fragmentation events are clearly evident. The radar is a very sensitive and powerful instrument - it can detect radar cross-section scatterings of as little size as 10^{-11} m^2. In addition, these images serve as a strong foundation to other possible work. For example, an HF radar is currently at the design stage at Arecibo, which will allow interferometery to be performed. Furthermore, 3D meteor modeling is in close reach as well. These advancements will allow for more detailed studies, especially of fragmentation. Since there are many other theories about the origin of meteor behavior, there is exceptional motivation to locate its exact causes. However, the models and images we have presented here are consistent with meteoroid fragmentation. We have also potentially identified a new class of very intensive radar meteors called "bolides." These meteors have revealed very complex fragmentation patterns (Figure 6), which further demonstrate the need for better models and additional study [7].

Further research is also being conducted on very unique topics, such as Polar Mesosphere Summer Echoes [8], as well as trails echoes that could possible exist in the UHF (Ultra High Frequency) spectrum. In a short while, these and many other useful details such as ablation modeling [9] will hopefully be integrated into radio science to encourage and support current and all future studies.

ACKNOWLEDGMENTS

REFERENCES

[1]   J.D. Mathews et al. "The micrometeorid mass flux into the upper atmosphere: Arecibo results and a comparison with prior estimates," Geophysics Research Letters, vol. 28, pp. 1929-1932, 15 May 2001.

[2]   Zdeněk Ceplecha, "Meteor Phenomena and Bodies," Space Science Reviews vol. 84, pp. 328-471, January 1998.

[3]   Malhotra, A., and J. D. Mathews, "Low-altitude meteor trail echoes," Geophysics Research Letters, vol. 36, 4 pages, 12 November 2009.

[4] J.D. Mathews, "Radio science issues surrounding HF/VHF/UHF radar meteor studies," Journal of Atmospheric and Solar Terrestrial Physics, vol. 66, pp. 285-299, February - March 2004.

[5] J.D. Mathews, "An update on UHF radar meteor observations and associated signal processing techniques at Arecibo Observatory," Journal of Atmospheric and Solar Terrestrial Physics, vol. 65, pp. 1139-1149, October-November 2002.

[6] J.D. Mathews et al., "Extensive meteoroid fragmentation in V/UHF radar meteor observations at Arecibo Observatory," Geophysics Research Letters, vol. 37, 5 pages, 26 February 2010.

[7] Malhotra, A., and J. D. Mathews, "Meteoroid Fragmentation as Revealed in Head- and Trail-echoes Observed with the Arecibo UHF and VHF Radars," July 2010.

[8] M. Rapp and F.-J. L¨ubken, "Polar Mesosphere Summer Echoes (PMSE): review of observations and current understanding, Atmospheric Chemistry and Physics, June-December 2004.

[9] L. Dyrud and D. Janches, "Modeling the meteor head echo using Arecibo radar observations," Journal of Atmospheric and Solar-Terrestrial Physics, vol. 60, pp. 1621-1632, 2008

THE CHARACTERIZATION OF RF ANTENNAS FOR BEAMED POWER

Charles Perkins,* Aseem Singh,[+] and Sven Bilén[#]

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

*Undergraduate Student of
Department of Electrical Engineering
Norfolk State University
Norfolk, VA 23504

ABSTRACT

RF power beaming has been proposed as a viable mechanism for powering devices with low power demands, such as Wireless Sensor Systems (WLSSs). Beamed power greatly reduces the weight and volume of power distribution by eliminating wires, connectors, and support structures. WLSSs currently under development generally utilize battery power; however, the WLSS user community is interested in the beamed power modules as a means of power delivery. This work explored various beamed power system antenna designs and characterized their performance as a function of antenna configuration, distance, and power transferred. The results of this research will help guide the development of beamed power module for the WLSS application and other lower power applications.

INTRODUCTION

Beamed Power

Wireless Sensor Systems (WLSS) currently under development generally utilize battery power; however, the WLSS user community is interested in the beamed power as a means of power delivery. Beamed power offers the possibility of greatly reducing the weight and volume of power distribution by eliminating wires, connectors, and support structures. Some other applications for beamed power include charging cell phones, mp3 players, and other consumer devices. Another application being explored at Penn State is the powering of sensors modules on heat shields [1].

[#] Faculty Mentor
[+] Graduate Mentor

Kurs *et al.* [2] found that two resonant objects with the same resonant frequency tend to exchange energy efficiently, while dissipating relatively little energy in extraneous off-resonant objects. In practical terms, this means that the antennas are critical in a beamed power system. Fabricating antennas that resonate at the same frequency enhances their ability to transfer power efficiently. Singh [1] ran many experiments and came to the conclusion that helical antennas are the most efficient shape for transferring power.

Figure 1 shows a schematic representation of a power beaming system. The power supply and the RF signal generator are connected to the power amplifier. The amplifier is then connected to the transmitting antenna and an RF signal is sent to the receiving antenna. The signal that is received at the receiving antenna is then converted by the power conditioning board. The load is connected to the power conditioning board and uses the power that is transferred in the system.
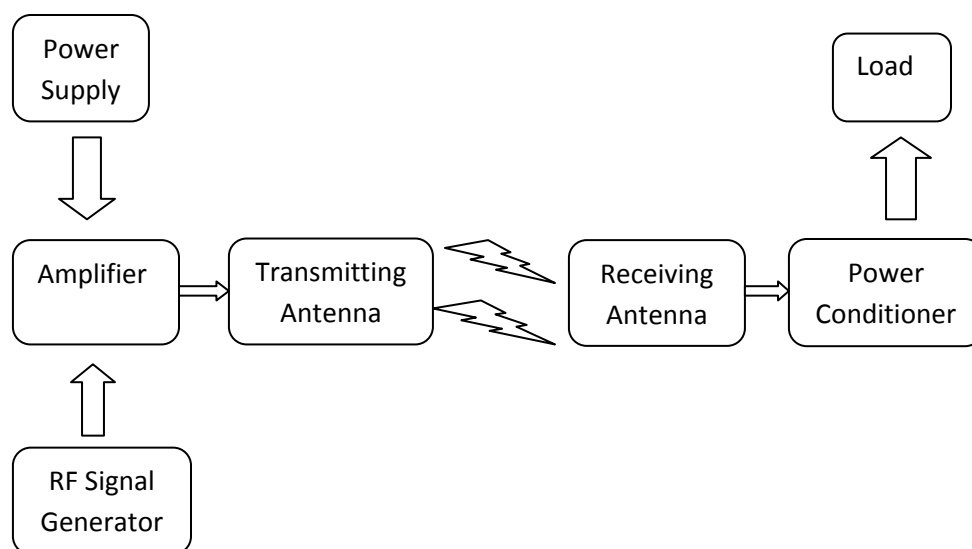

Figure 1: Schematic of a Power Beaming System

This research project explored various beamed power system antenna designs and characterized their performance as a function of antenna configuration, distance, and power transferred. In addition, a new device for use in boosting the low voltages in energy scavenging systems was also evaluated. The results of this research will help guide the development of beamed power module for the WLSS application and other lower power applications.

EXPERIMENT DESCRIPTION
Amplifier Testing

As shown in Figure 1 above, one of the components of the RF power beaming system is the RF amplifier. We tested four amplifiers to determine which had the highest power amplification. Figure 2 provides power delivered into a load resistor of various values. All amplifiers are Mini-Circuits model ZHL-1000-

3W (SMA) with different serial numbers. Table 1 shows that Amplifier N022399 is the best amplifier to use for our system because it gave us the highest power reading into a 50-Ω resistor.
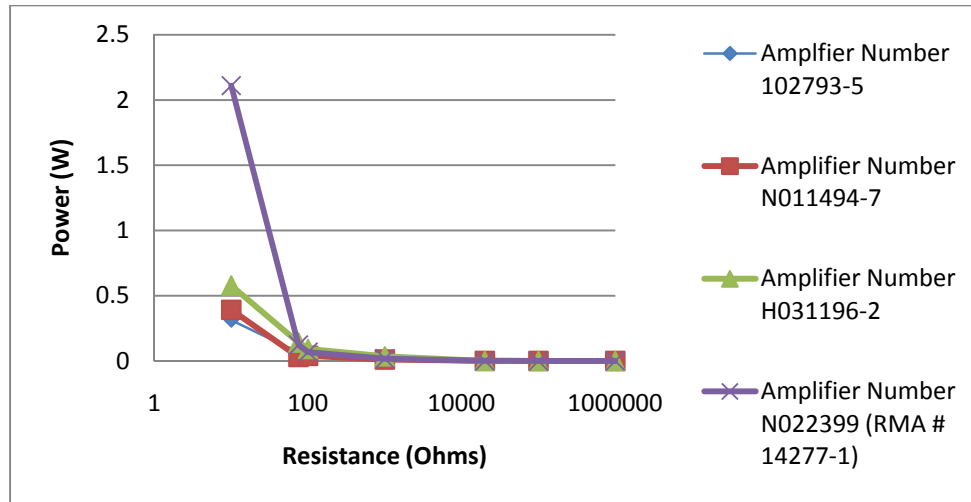

Figure 2: Voltage vs. Resistance reading of different Amplifiers

Table 1: Power Readings from Best Amplifier (N022399)

Resistance (Ω)	V-peak-to-peak (V)	V-max (V)	V-min (V)	Power (W)
10	14.4	6.5	7.8	2.11
50	12.5	6.1	6.4	1.12
75	10.6	4.3	5.3	0.123
100	8.4	3.7	4.6	0.068
1.00E+03	10.2	5.9	4.6	0.017
2.00E+04	13.4	6.5	6.8	0.0011
1.00E+05	12.5	6.5	5.6	0.00021
1.00E+06	10.3	7.2	3.1	0.000052

Antenna Design and Testing

With reference to Figure 1, another critical component in the overall RF power beaming system is the transmitting and receiving antenna. Through testing several different antenna types (Splatch 315 grounded line planar antenna, JJB 916 compact monopole, and the Helical 315 quarter-wave helical antenna), we found that the helical antennas are the best antennas to use because they transfer power the best.

Helical antennas have a number of parameters that affect their resonant frequency. Some of those parameters are the number of turns, spacing between the turns, and the coil diameter. For these experiments, the transmitting antenna had 5 turns with 3 inches of spacing in between each turn and a diameter of 4.5

inches. Total wire length was 70 inches. The transmitting antenna is supported by two wood pieces measuring 15 inches (Figure 3).
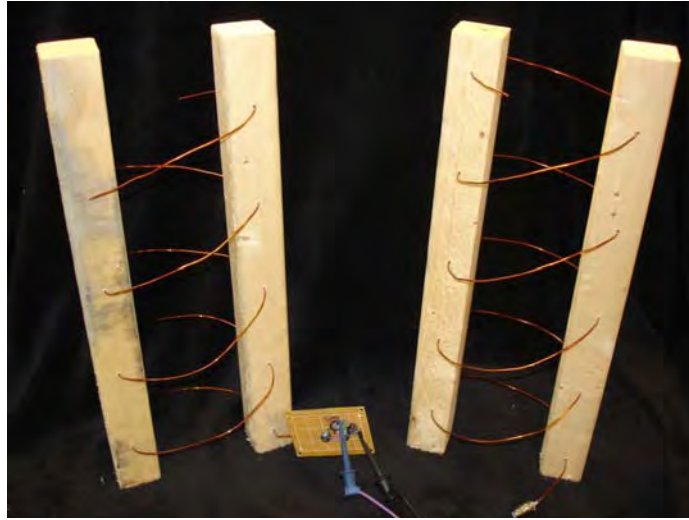

Figure 3: Transmitting and Receiving Antennas

After the transmitting antenna was built, its resonant frequency was measured with a network analyzer. A network analyzer measures the complex impedance (i.e., the magnitude and the phase characteristics) of microwave components such as network amplifiers, components, cables, and antennas. There were three different parameters that we measured with the network analyzer: Standing Wave Ratio (SWR), S_{11}, and complex impedance. SWR measures the standing wave along a transmission line, which results from reflections that occur due to any unmatched loads on a uniform transmission line. The S_{11} component examines the reflected power from the device under test, in this case the antenna. This data was returned as a Log Magnitude in dB and is lowest at the antenna's resonant frequency. The complex impedance consists of the antennas resistive (real) and the reactive (imaginary) impedances, often shown on a Smith Chart. On a Smith Chart, if the reactive component, X, is positive, the device is inductive and if it is negative then the impedance is capacitive. Table 1 provides antenna characteristics as measured by the network analyzer. Some improvements that could have given better data is fine tuning of the antenna by making it even smoother in its construction. Also, making the diameter smaller could result in better transfer of power.

Table 1: Characteristics of the Transmitting Antenna

Freq. of 1st Resonance (MHz)	Freq. of 2nd Resonance (MHz)	Magnitude of 1st Resonance (dB)	Magnitude of 2nd Resonance (dB)	Impedance (Ω)	SWR
127	182	−8.82	−11.05	80.1−j20.95	1.772

RF ANTENNAS FOR BEAMED POWER

Power Conditioning Board

Again with reference to Figure 1, another important component of an RF power beaming system is the power conditioning board. As part of this effort, we examined the possibility of using a Linear LTC3108. This new device is a DC/DC converter that was designed for harvesting and managing energy from very low input voltages such as solar cells or thermoelectric generators. It can work from input voltages as low as 20 mV. Using a small transformer, this device provides power management for wireless sensing. The Power Conditioning Board has a 2.2-V LDO for powering low power processors, while the main output is programmed to power a wireless transmitter. It also contains a storage capacitor that provides power when the input voltage is unavailable. This board is designed to accumulate and manage energy over a long period of time in order to enable short bursts of power to obtain and transmit data. These properties are what interested for our application.

SUMMARY

RF power beaming has been proposed as a viable mechanism for powering devices with low power demands, such as Wireless Sensor Systems (WLSSs). In this work, we explored several of the components needed to transfer power wirelessly. Several amplifiers were tested across a 50-Ω load to determine the one with highest power output. After finding the best amplifier, we fabricated and tested antennas. We found that many parameters have an effect on the power transfer, such as the diameter of the coil, the distance in spacing between the coils, and how many turns are in the helical antenna. With these findings, we conclude that decreasing the diameter and smoothing out the coils of the helical antenna help give better results. This work is still an ongoing as researchers are working to better understand beamed power systems.

ACKNOWLEDGMENTS

REFERENCES

[1] Singh, Aseem, "Wireless Power Transfer for Wireless Sensor Modules," Master Thesis, The Pennsylvania State University, December 2009.
[2] Kurs, Andre, Aristeidis Karalis, Robert Moffatt, J. D Joannopoulos, Peter Fisher, Marin Soljačić, "Wireless Power Transfer via Strongly Coupled Magnetic Resonances," *Science* 6 July 2007, Vol. 317, No. 5834, pp. 83–86.

INTEGRATING PRESSURE MEASUREMENTS INTO WIRELESS SENSOR SYSTEMS

Julie Georgiev* and Sven Bilén[#]
Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

*Undergraduate Student of
Department of Electrical Engineering
Alfred University
Alfred, NY 14802

ABSTRACT

The utility of Wireless Sensor Systems (WLSSs) as an alternative to bulky wired systems has been demonstrated in the laboratory and in space-relevant environments. WLSSs minimize the number of wired connections, which means fewer problems with chafing, sparking, aging, or connector issues, and this translates to lower cost and higher reliability and ease of set up. This paper explores the development of a WLSS capable of multichannel sensor data acquisition and data acquisition with multiple sensor types during a single test. The focus is on the design of a circuit to condition and process pressure sensor inputs. Various pressure sensor signal conditioning circuit designs and hardware configurations were considered. Designs were required to meet existing WLSS functional and performance requirements, and add pressure signal conditioning hardware while not exceeding the existing envelope size constraints.

INTRODUCTION

This project focuses on the Wireless Sensor System (WLSS) for use in space systems under development in the Systems Design Laboratory at The Pennsylvania State University. Wireless sensors have a number of advantages over wired systems including the promise of lower acquisition cost and higher reliability. The number of cables and connectors are minimized in wireless systems, which means fewer chafing, sparking, ageing, and connector issues.[1] A switch to wireless sensors can reduce weight and simplify manufacturing.[2] Fewer wiring and connection failures means greater reliability and less required insulation. Wireless systems would also be an improvement over the non-

---

[#] Faculty Mentor

standardized wired systems, which are often customized to each project making set up and maintenance difficult and costly.[2]

WLSSs have a number of applications in spacecraft and payload systems, including heat shield technology and the development of thermal protection systems. The heat shield protects an entry vehicle from extreme temperatures in space and during atmospheric re-entry. In order to improve heat shield design, collecting data from sensors during re-entry is critical. However, for some scenarios such as landing on Mars, the heat shield may need to be separated from the entry vehicle before landing.[3] This means that, in the case of a wired system, wire cutters are used to separate the wired sensor system contained on the heat shield, which may introduce a "single point mission failure potential."[2] Wireless sensor systems have the potential to resolve these problems and improve spacecraft and payload design.

Figure 1 provides an overview of the WLSS under development, the hardware of which is distributed across a three-board system. The control board contains the ZigBee Telegesis ETRX2 module and the PIC16F886 microcontroller that is responsible for data acquisition and transmission, the power board contains a battery that supplies 3.3 V to the boards, and the signal conditioning board contains the ADG707BRU multiplexor and the AD7680BRM ADC that is the interface between the sensors and the control board.
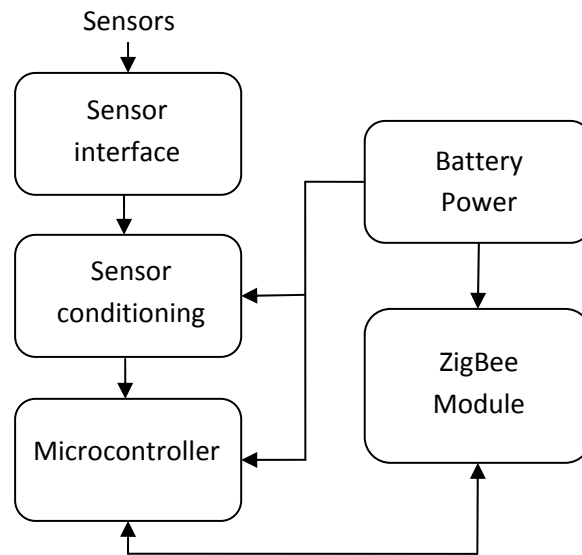

Figure 1: Wireless Sensor System Overview

A recent series of tests (NASA Ames AHF 289, January 2010) demonstrated that the WLSS can be used in testing in an arc jet, which produces a plasma-heating environment that simulates atmospheric reentry. The WLSS used in that test series transmitted single-channel thermocouple data during periods when the test model was out of the plasma flow and recorded data when the model was in the plasma flow. The next step is to move towards multi-channel data acquisition

and data acquisition from multiple sensor types during a single test, including three type-K thermocouples, a Gardon gauge to measure heat flux, and a pressure transducer. The output of the three sensor types to be integrated with the WLSS is shown in Table 1.

Table 1: Sensor Outputs

Sensor	Output
Type K Thermocouple	0–30 mV
Gardon Gauge	0–10 mV
Pressure Transducer	0–5 V

This project focused on the development of the elements necessary for data acquisition from the pressure transducer. The output of the pressure transducer differs from that of the type-K thermocouple in that its voltage output is much larger (volts rather than millivolts). The maximum signal output from the pressure transducer is 5 V, which exceeds the supply voltage of WLSS system of 3.3 V. In addition, since the signal level is several orders of magnitude larger than from the thermocouples, the gain stage used in the current signal conditioning board is not compatible with the pressure transducer. The signal conditioning circuit for the pressure transducer instead will be housed on a daughterboard and be completely separate from the circuit of the signal conditioning board. Due to time constraints, it was not possible to completely redesign the existing signal conditioning board; hence, this compromise was selected.

CIRCUIT DESCRIPTION

The pressure transducer used in testing at NASA Ames is the Kulite ETL-76M-190-15A. For testing at Penn State, a similar but more cost-effective pressure transducer was procured from Omega. The Omega MMA015V5P4A0T4A6 is functionally similar in that it has the same range and output as the Kulite, but has a higher total error band and a smaller temperature compensation range. The specifications of both pressure transducers are provided in Table 2.

Table 2: Comparison of Pressure Transducers

	Kulite ETL-76M-190-15A	Omega MMA015V5P4A0T4A6
Operational Mode	Sealed Gauge/Absolute	Absolute
Pressure Range	15 psia	15 psia
Full Scale Output	0–5 Vdc	0–5 Vdc
Total Error Band	± 0.25%	~ ± 0.36%
Temperature Compensation	−40 to +120 °C	−40 to +105 °C

The signal conditioning circuit that was developed to integrate the pressure transducer into the existing WLSS is shown in Figure 2. To bring the output of the pressure transducer down from its maximum of 5 V to a maximum of 3.3 V, a simple resistance divider (with 13.7 kΩ and 26.7 kΩ resistors) is used. The OPA350EA/250 operational amplifier was chosen to buffer the output and it is also used in the low-pass filter. The filter was designed using FilterPro Desktop from Texas Instruments after specifying a low-pass filter with 1-kHz cutoff, filter order of up to 2, and Sallen–Key filter topology. The internal ADC on the PIC16F886 microcontroller has only 10-bit resolution, whereas 14-bit resolution was required by the end user; thus, an external ADC (AD7680BRM) is used. The ADC used on the signal conditioning board and the one on the pressure sensor signal conditioning circuit interfaces with the PIC microcontroller using SPI protocol.
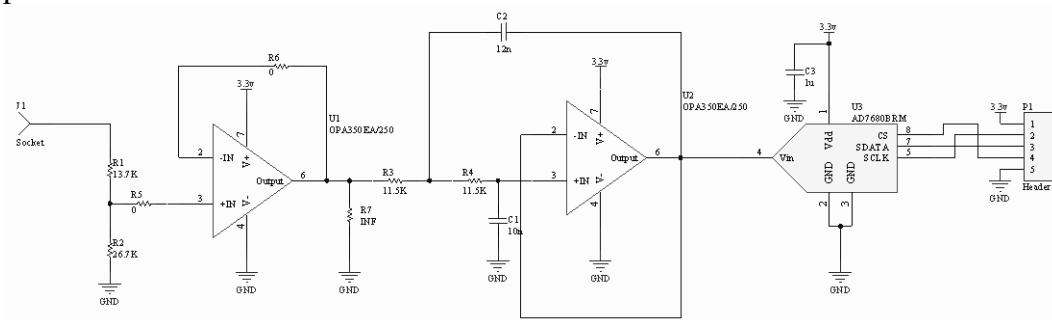

Figure 2: Pressure Sensor Circuit

Once the circuit was designed, the next step was to draw a schematic in the electronics design software Altium. Footprints were made for the op-amp and ADC and from the schematic a PCB design was developed, shown in Figure 3. After routing the board in Altium and generating a Gerber file the board was milled and populated onsite at Penn State.
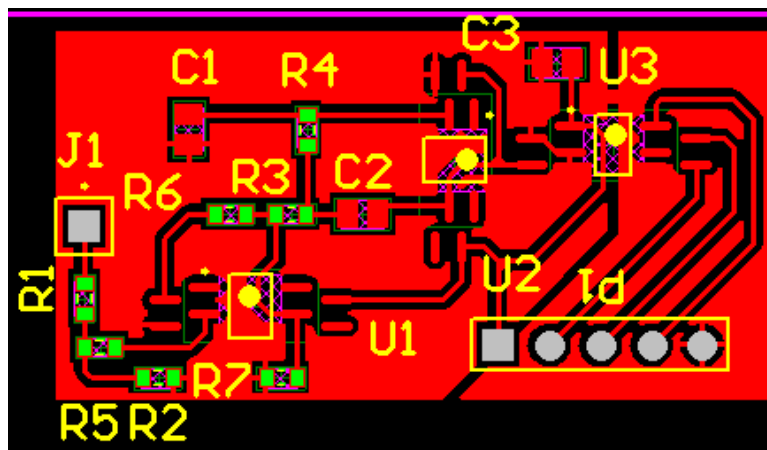

Figure 3: ~720 × 1400 mil PCB Generated in Altium

After the board was completed, it was tested before integration with the WLSS to see if it functioned correctly and met accuracy requirements (Figure 4). The Omega CL8300 DC V/mA Standard high precision voltage source was used in the place of a pressure transducer and the Agilent 34401A 6½-digit multimeter was used in order to obtain end-to-end characterization of the analog signal conditioning and determine if the gain is linear. In all tests the Agilent E3631A triple output DC power supply was used to supply 3.3 Vdc to the circuit. The calculated gain for a correctly functioning circuit is 0.66 V/V.
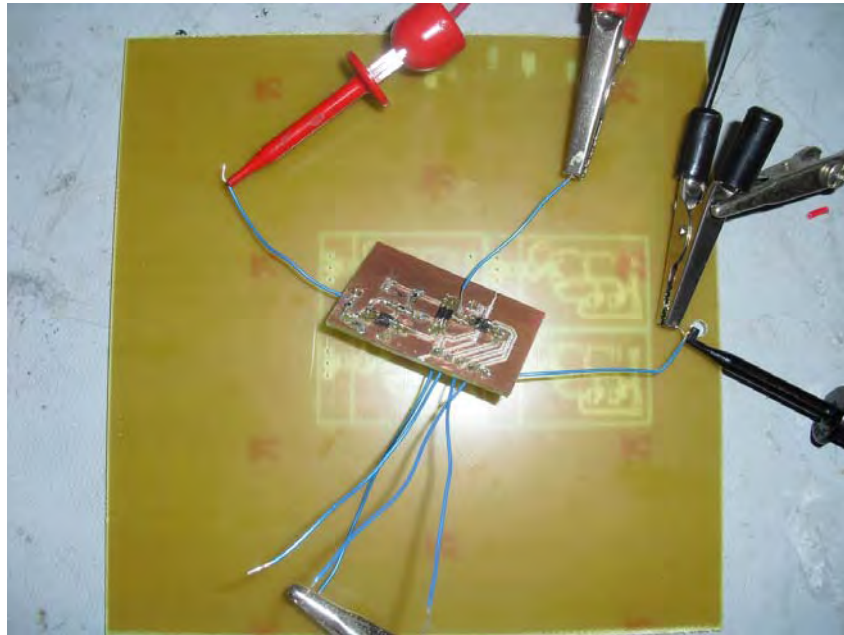

Figure 4: Test Setup of Populated Board for End-to-End testing

Two Agilent 33220A 20-MHz functional arbitrary waveform generators and an Iwatsu–LeCroy LT364 500-MHz oscilloscope were used to verify that data is being output from the ADC. The data out (SDATA) line on the ADC was connected to channel 2 of the oscilloscope, and the output voltage from the analog signal conditioning stage of the circuit was connected to channel 4. The ADC has a minimum requirement for the clock (SCLK) of 250 kHz. The SCLK line was connected to 3.1 V at 300 kHz and channel 1 of the oscilloscope. On the falling edge of the chip select (CS) signal, the ADC samples the analog input. Previous testing where the chip select signal was manually taken from high to low were unsuccessful, so a second waveform generator was used in the second test. The CS line was connected to 3.1 V at 6 kHz and channel 3 of the oscilloscope.

Initial testing with the Omega pressure transducer was also performed to determine the output of the pressure sensor signal conditioning circuit at atmospheric pressure. The Agilent E3631A power supply was used to supply the 24-Vdc excitation voltage to the pressure transducer. The output of the analog

signal conditioning stage was connected to the Agilent 34401A multimeter and the Iwatsu-LeCroy LT364 oscilloscope to monitor noise.

TEST RESULTS

Using a high precision voltage source in lieu of the pressure sensor, preliminary tests confirmed that the resistance divider, op-amp buffer, and low-pass filter worked correctly. With an input voltage of 1 V, the output at each stage was 0.66 V. For measurements taken at an interval of 0.1 V from 0–5 V, the average gain was 0.66 V/V, which matches the calculated gain. This gain curve is shown in Figure 5.
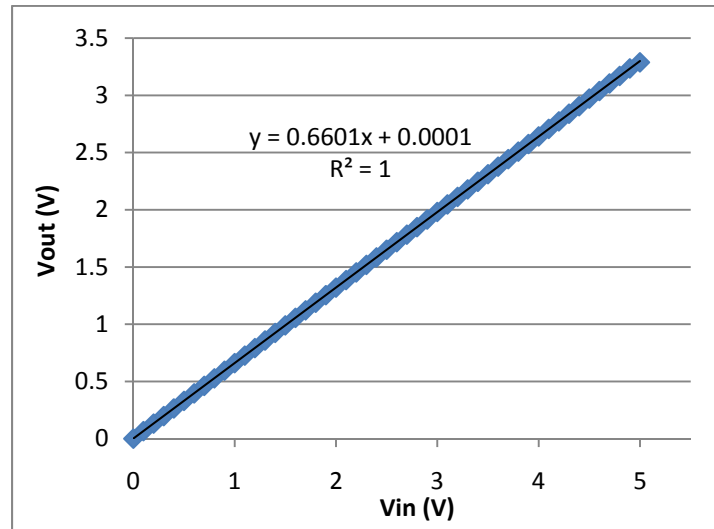

Figure 5: Output Voltage versus Input Voltage for Analog Signal Conditioning

The ADC did output data in the range of 0–5 V and the data captured from the oscilloscope is shown in Figure 6. A higher than expected level of noise was observed, but this is due in part to a complicated experimental setup that involved a lot of crossed wires. The spikes in noise at the rising and falling edges of the SDATA line is due mostly to "cross talk" between the analog and digital sides. It is speculated that after integration with the WLSS most of this issue will be eliminated.
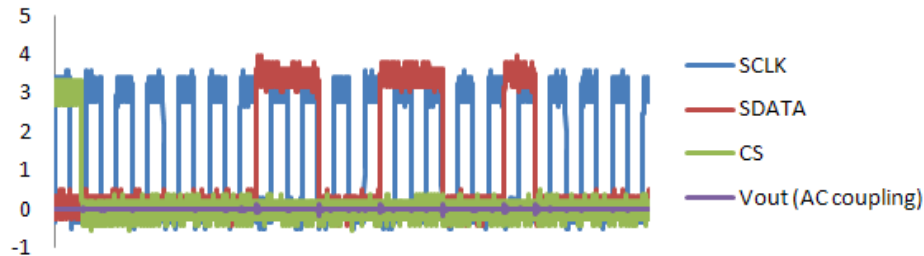

Figure 6: ADC Output

At an input of 0 V, the SDATA line was mostly low; however, noise was present in the least significant bits. The same was true for at an input of 5 V, where the SDATA line was mostly high but with noise in the least significant bits. This suggests an accurate digital output, as the minimum output is represented as all zeroes and the maximum output as all ones. During testing at a high sampling rate a time average of the data is suggested to account for noise.

Preliminary testing with the pressure transducer also suggests that the circuit functions correctly. At atmospheric pressure, approximately 9.62×10^4 pascals or 13.95 psi, the analog stage output a value of 3.087 V. The calibration worksheet included with the pressure transducer yields a linear fit of $y = 0.3337x - 0.0005$ (Figure 7), so at 13.95 psi the pressure transducer should output approximately 4.65 V. With a linear gain of 0.66 V/V, this corresponds to 3.07 V calculated value; therefore, our experimental value of 3.087 is reasonable. Further testing in a vacuum chamber is scheduled to obtain a better understanding of the performance of the pressure sensor signal conditioning circuit.
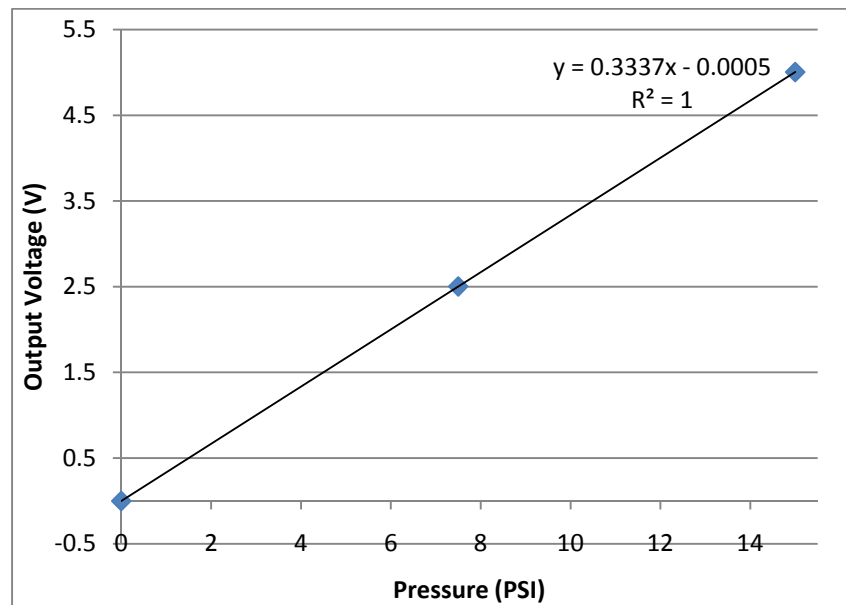

Figure 7: Pressure Sensor Output versus Pressure

CONCLUSION

The objective of this project was to design a circuit to condition and process pressure sensor inputs to increase the technology readiness level of the wireless sensor system currently under development. Many designs were considered and the final PCB design was generated in Altium, milled, populated, and characterized via test. The circuit was shown to operate with a linear gain of 0.66 V/V after end-to-end characterization testing of the analog signal conditioning. The ADC was also verified to output data between 0–5 V. Since the circuit was shown to operate correctly within WLSS functional and performance requirements the next step is to modify the microcontroller code and LabVIEW

interface to interface with the pressure sensor circuit. Future WLSS could include a redesign of the signal conditioning board so that a separate circuit is not required for the pressure transducer.

ACKNOWLEDGMENTS

REFERENCES

[1]   Singh, A., "Wireless Power Transfer for Wireless Sensor Modules," M.S. Thesis paper, The Pennsylvania State University, Dec. 2009.

[2]   Swanson, G., J.J. Schlee , and D. Atkinson, "Multi Nodal Wireless TPS Sensor System Integration," Proceedings of the Sixth International Planetary Probe Workshop, Georgia Institute of Technology, Atlanta, 21–27 June 2008.

[3]   Kot, W. K. R., "Distributed Wireless Sensor Network for Aerospace Platforms," M.S. Thesis paper, The Pennsylvania State University, May 2010.

AN INVENTORY SIZE FEASIBILITY STUDY FOR FAST ALGORITHM ANALYSIS/RE-SYNTHESIS BASED SPEECH ENHANCEMENT

Carnell Hunter,* W. Kenneth Jenkins,[#] Robert K. Nickel,[#] and Chandra Radhakrishnan [+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Virginia Commonwealth University
Richmond, VA 23284

ABSTRACT

The purpose of the project is the development of a fast procedure for (approximate) signal correlation. The correlation procedure is used as a part of a novel single channel speech de-noising procedure. The fundamental way of de-noising a speech signal is through linear convolution, but although effective it is computationally intricate and power consuming. This research aims to reduce the complexity of the process by using segments of a noise corrupted speech signal and running it through a pre-recorded database of clean speech. Which segments are to be selected from the database are determined utilizing the Correlation Based Search procedure. In addition, the recently introduced Modulated Fourier Transform, another argumentatively efficient way of mimicking linear convolution, will be discussed and implemented in this research.

INTRODUCTION

Any arbitrary speech signal can become corrupted with noise due to the environment in which it is generated or by other means. For example, voice recognition in cellular devices can be tainted with other conversations and outside sounds. These conversations and outside sounds are considered to be "noise". This unwanted noise can end up being a problem for voice recognition due to difficultly recognizing what a user is saying. Speech de-noising is beneficial

---

[#] Faculty Mentors
[+] Graduate Mentor

because it helps recover what the user said without the added noise [1]. This project will explore the process of speech signal de-noising.

Speech signal de-noising is not a simple process. Many times in speech processing a long signal may be generated which can be time consuming when de-noising. Linear convolution is the conventional method of de-noising a signal in a computer such as desktop or laptop [2–3]. However, for a cell phone this method can become quite computationally intensive and consume an excess amount of battery life. Understanding the limitations of a cell phone, the aim for this project is to use the recently published Modulated Discrete Fourier Transform (MDFT). The MDFT puts to use the traditional Discrete Fourier Transform and adds an extra modulation. The extra modulation makes the signal easier to use in order to correlate the given speech signal with the filter.

Most researchers, when de-noising, are focused on removing the noise from the speech signal. In this research, the desired segment of the noisy speech signal will be correlated with a database of other clean speech signals. Once the correlation is complete, the part of the database with the highest correlation will replace that segment from the given speech signal. This is speech processing, and these steps are repeated for the entire given speech until it is completely replaced with speech signals from the database [4–6].

In speech de-noising there is always a sacrifice of the original signal. When removing noise from the signal there is almost always some noise left on the signal. Block processing will be used so that there will be no noise left on the signal but due to the use of the database the final signal will not perfectly match the original clean signal.

THEORETICAL BACKGROUND
Correlation Filter Using Fourier Block Processing

In this study, the correlation filter has been implemented through block processing experimenting with *Overlap-Save* and *Overlap-Add* methods [7]. It has been known that most input signals have long lengths (or continuously flowing indefinite lengths) and when it comes to processing the signal, limited computer memory is available. When implementing the correlation filter using DFT on a large number of points, the DFT becomes too large to compute and as a result, a long delay is created. To avoid processing delay, the input signal is segmented into blocks of length L. Then each block can be convolved with a finite impulse response and the DFT can be computed. This procedure is developed as follows.

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

N number of samples in the "long" input signal
M length of impulse response
L length of each individual block
i number of blocks or segments

The equation for the number of segments will be computed:

$$(i-1)*L + M \geq N \tag{3}$$

$$(i-1)*L \geq N - M \tag{4}$$

$$(i-1) \geq (N-M)/L \tag{5}$$

$$i \geq (N-M)/L + 1 \tag{6}$$

$$i = \text{ceil}((N-M)/L + 1) \tag{7}$$

Equations 3–7 indicates how many segments have been created by segmenting input signal $x(n)$. After $x(n)$ convolves with $h(n)$, the output filtered block will have a length of $L + M - 1$. The first well known block processing algorithm is the *Overlap-Save* algorithm where L is selected to be some suitable value with $L > M$. The signal $x[n]$ is divided into blocks of length L and they overlap by $M - 1$ data points. Therefore, the ith block is

$$x_i[n] = x[n + i(L - M + 1)], n = 0,...,L-1. \tag{8}$$

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

$$h_{\text{pad}}[n] = \left| \begin{array}{ll} h[n], & n = 0,..., N\text{-}1 \\ 0, & n = N,..., N_{\text{DFT}}\text{-}1 \end{array} \right. \tag{9}$$

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

$$y_k[n] = \left| \begin{array}{ll} y_{\text{pad}}[n], & n = 0,..., N\text{-}1 \\ 0, & n = N,..., N_{\text{DFT}}\text{-}1 \end{array} \right. \tag{10}$$

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

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

$$s_k[n] = \left[\begin{array}{ll} s[n], & n = 0,..., L\text{-}1 \\ 0, & n = L,..., N_{\text{DFT}}\text{-}1 \end{array} \right. \tag{11}$$

where L= M-N+1. Again, the filter function $h_{\text{pad}}[n]$ is increased with zeros to produce $h_{\text{pad}}[n]$, and then the DFT is computed. Now in each output block

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

Discrete Fourier Transform

In speech processing there are many signals in which computations become difficult and have low efficiencies, but with Discrete Fourier Transforms (DFT) implemented via the FFT algorithm this problem is reduced. It is a useful computation tool that allows analysis of a signal in the frequency domain which requires less power and has a greater efficiency level than in the time domain. The DFT, like its name states, is a tool to be used for discrete-time signals. The DFT evaluates the Fourier Transform at a set a set of equally spaced discrete frequencies to give the samples of frequency used in computation. The defining equations for the DFT and the Inverse Discrete Fourier Transform are as follows:

$$\text{DFT:} \qquad X[k] = \sum_{n=o}^{N-1} x[n]e^{-j2\pi kn/N} \qquad (12)$$

$$\text{IDFT:} \qquad x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j2\pi kn/N} \qquad (13)$$

A great use of the DFT is in computing the convolution. The convolution is a formula that gives a response of a LTI system as a function of both the input signal and impulse response:

$$\text{Convolution:} \qquad y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \qquad (14)$$

In the convolution equation [14] it should be understood that it doesn't matter which of the two sequences are folded and shifted. What does matter is using the DFT reduces these time consuming computations drastically. In the following equation it is shown that converting each signal to its frequency counterpart and taking the product of the two yields the same solution as that equation (14).

$$\text{Convolution with DFT:} \qquad y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = X[k]*H[k] \qquad (15)$$

Another benefit of the DFT is its efficiency when implemented with the Fast Fourier Transform (FFT) algorithm. The FFT is an algorithm which lessens the complexity of the computing DFT from Order (N^2) to Order $(N \log_2 N)$.

Modulated Fourier Transform

Mentioned previously in this paper, the MDFT is beneficial because it does not require zero padding when using *Overlap-Add* FFT block processing. It is a more efficient technique of mimicking linear convolution with the advantage that there is a reduction in the amount of computation being processed, therefore conserving power consumption in devices such as cell phones. The equations that define the MDFT and the inverse MDFT for implementing Right Circular Convolution (RCC) is given below [8]:

$$\text{MDFT: } S[l] = \sum_{n=0}^{N-1} s[n] e^{-j\frac{\pi n}{N}\left(2l+\frac{1}{2}\right)} \tag{16}$$

$$\text{Inverse MDFT: } s[n] = N^{-1} \sum_{l=0}^{N-1} S[l] e^{j\frac{\pi n}{N}\left(2l+\frac{1}{2}\right)} \tag{17}$$

where N is the period of the given sequence. When two sequences of length N are convolved using this method the RCC results obtained are as follows:

$$\text{Right Circular Convolution: } y_r[n] = \sum_{i=0}^{n} h[i]x[n\text{-}i] + e^{j\frac{\pi}{2}} \sum_{i=n+1}^{N-1} h[i]x[n+N-i] \tag{18}$$

If the $-\frac{1}{2}$ in the exponents of equations (16) and (17) are replaced by $+\frac{1}{2}$ the result is Left Circular Convolution (LCC) defined as:

$$\text{Left Circular Convolution: } y_\ell[n] = \sum_{i=0}^{n} h[i]x[n\text{-}i] - e^{j\frac{\pi}{2}} \sum_{i=n+1}^{N-1} h[i]x[n+N-i] \tag{19}$$

While using the MDFT process, the complex results contain real and imaginary values. Since the speech input signals are real only their convolved values must be real values. It was shown in [8] that the MDFT defined by equations [16] and [17] separates the lower half of the convolution vector into the real parts and the upper half into the imaginary parts of $y_r[n]$ (or $y_\ell[n]$) according to:

$$y_r[n] = y[n] + jy[N+n] \tag{20a}$$

$$y_\ell[n] = y[n] - jy[N+n] \tag{20b}$$

Therefore for real valued input signals the MDFT automatically separates the lower and upper half of the output block by implementing block processing with length N sequences and computing only one MDFT defined by either (16) or (17).

EXPERIMENTAL DESCRIPTION

This research on speech processing using Correlation Based Search based was simulated using the algorithmic level language MATLAB. A scenario for Correlation was designed by determined the block length, picking the designated database utterance, and implementing correlation in correlation with speech block processing. Although MATLAB contains its own built in correlation function, due to the way it manipulates the signal it wasn't feasible to use. In this algorithm the speech segments are reversed in the time domain and each database filter is convolved with the speech signal, thus implementing correlation. More efficient ways of correlating these values were tested and before the speech processing could be performed there had to be an understanding of other fundamental algorithms used to manipulate the speech process such as DFT, MDFT, and *Overlap-Add* block processing [9].

MATLAB Implementation of the Discrete Fourier Transform (DFT)

The DFT is used for transforming a signal from its time domain into the frequency domain. In the following MATLAB code, the DFT is implemented. The variable *x* is the signal to be transformed, *N* determines the period for the periodic sequence, and the frequency is sampled at *W* equally spaced frequencies. Compilation of the **for** loops controls the summation given earlier in equation (12).

MATLAB Implementation of the Modulated DFT (MDFT)

MDFT utilizes the standard DFT with an added modulation. In order to modulate the sequence the signal must be multiplied by $e^{j\frac{n\pi}{2N}}$. The advantage of using the MDFT is that it eliminates the need to pad the input signals with zeros. Below is the MATLAB code for the MDFT. The signal is *x* and the filter values are *h*. Again, *N* sets the period for this periodic sequence, where *modfact* is the additional modulator needed to modulate the sequence. The terms *x1* and *h1* are the modulated forms of *x* and *h*. The *cnv* takes the modulated signals and transforms them to their frequency domain using the **fft** command, and then multiplies them in their frequency domain. After the frequency domains of *x1* and *h1* are multiplied, the product is then inversed back to the time domain using ***ifft***, and the *invmodfact* function removes the modulation. These steps are done to perform the computation of the MDFT described previously in equation (16).

```
         Table 1.  MatLab Code for the Discrete Fourier Transform (DFT)

x = [0 1 3 5 7 9];  % signal in time domain
N = 5;              % period
W = exp(-i*((2*pi)/N));

Nn=length(x); % determines the length of the transformed signal

for k=1:Nn;    % this for loop controls the k values, the length of the
                  transformed signal

    yy=0;

    for n=1:N;  % this for loop controls the n values in the summation
        z = (x(n)*(W)^((k-1)*(n-1)));
            yy=yy+z; % constantly adds to yy until summation is complete
    end

    X(k)=yy; % the signal in the frequency domain
    end
```

```
Table 2.  MatLab Code for the Modulated DFT (MDFT)

x = [10*ones(1,16) -10*ones(1,16)];
h = [0:10 10*ones(1,10) 10:-1:0];
N = 32;
modfact = exp(i*((pi)/(2*N))*(0:N-1)); %modulator

x1 = x .* modfact; %modulates signals
h1 = h .* modfact;

cnv = ifft(fft(x1) .* fft(h1));  %convolution

invmodfact = exp(-i*((pi)/(2*N))*(0:N-1)); %demodulates

final = cnv .* invmodfact; %demodulates final signal

X = [real(final) imag(final)]; %separates real and imaginary values
```

Overlap-Add Method

    Overlap-Add block processing allows for long speech inputs to be segmented into blocks during the filtering. The MATLAB code below shows the overlap-add method but uses the MDFT as opposed to the regular DFT. In the code below *input* determines the broken down block size. The variable x is the long input signal and *h* is the filter used. The *start* and *last* terms are used for the broken down blocks and are the beginning and ending values of each block, respectively.

```
                    Table 3.  MatLab Code for Overlap-add Block Processing)

input = 128; %determines the length of x
h = [0:10 10*ones(1,10) 10:-1:0]; %filter
x = randn(1,input); %signal to be filtered
N = length(h) ;
I = length(x);
z = ((I/N)+1)*N; %this determines the length of the filtered signal
Y = zeros(1,z); %sets the size of the filtered signal
modfact = exp(i*((pi)/(2*N))*(0:N-1));
h1 = h .* modfact; %modulates signals
invmodfact = exp(-i*((pi)/(2*N))*(0:N-1)); %demodulates

%%
start = 1;
last = N;

for n = 1:I/N %controls the overlap add sequence

X1 = [x(start:last) .* modfact];
cnv = ifft(fft(X1).*fft(h1));
final = cnv .* invmodfact; %demodulates final signal
Y1 = [real(final) imag(final)]; %separates real and imaginary values

if n ~= I/N %at the last iteration there is no need to add N because it will put us
out of our index
    start = start + N;
    last = last + N;
    X2 = [x(start:last) .* modfact];
    cnv = ifft(fft(X2).*fft(h1));
    final = cnv .* invmodfact; %demodulates final signal
    Y2 = [real(final) imag(final)]; %separates real and imaginary values
    add = Y1(N+1:2*N) + Y2(1:N); %takes the bottom of the preceding circ conv and
adds it to the top of the current circ conv
end

if n == 1 %takes the first values of the first circ conv and places it at the
beginning of the filtered signal
    Y(1:N) = Y1(1:N);
end

if n == I/N %takes the last values of the last circ conv and places it at the end
of the filtered signal
    w = z-N+1;
    Y(w:z) = Y1(N+1:2*N);
end

Y(start:last) = add; %takes the current overlaped signals and places them in the
correct spot in the filtered signal
end
```

The MDFT process occurs as explained previously and produces the convolved output signal values. Afterwards, the subsequent block of *x* is convolved with the filter *h*. The second half of the values from the previous convolution and the first half of the values from the current convolution are then overlapped and added together.

*Correlation Based Search Algorithm**
1. First a speech file called arctic_a0001.wav was read into MATLAB using the *wavread* function and necessary information was found of the file such as duration of the sound file, sampling rate, and bit resolution. Figure 1 shows a

plot of a pre-recorded speech signal in its entirety in MATLAB.
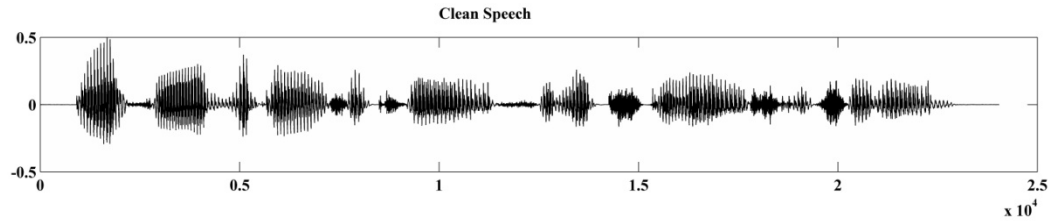
Information of the sound file "arctic_a0001.wav":

Duration = 3.23506 seconds
Sampling rate = 16000 samples/second
Bit resolution = 16 bits/sample

Figure 1: Speech Reference File 1- Arctic_a0001.wav

2. Taking into consideration that MATLAB may not allow for a huge amount of database utterances to be stored in memory, a code was developed to allow for running through the database with ease. In the subsequent code, using *wavread*, the input speech signal is stored into *x*, then down sampled using *downsample* to reduce the number of samples. An *if* statement is utilized, it states that depending on the value of *a* store one of the scripts (`arctic_a000, arctic_a00, arctic_a0`) into *file*. Following the *if* statements it concatenates an entire file name using *strcat* and stores it in *y*. Using *wavread* again the filename *y* is read and then stored into *y*. Then *y* was down sampled and stored in *y1*.The variable *x1* is a segment of the input speech and *x2* is the next speech segment of same length.

3. As mentioned before the built in correlation function caused complication so instead the built in convolution command was used. The code below shows the correlation implementation as well as the method used to find the highest value of the correlation. With *testVect* being a segment of the input speech, if it's flipped first and the convolved with the database utterance using the *conv* function. The flipping of the segment first is what makes this convolution act as a correlation. These values are stored in *corrVal*. Next, *corrVal* is normalized by *normdB* and the normalized value is then stored as *corrVal*. Using *max* returned is the highest value of the correlation, *pkVal*, and the respected index, *loc*. The final value obtained, *corrSpeech*, is the segment with the highest correlation to the input speech segment. Ideally, *corrSpeech* will be used to replace the speech segment.

```
            Table 5 Speech Correlation Function

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This function correlates the testing speech signal with  %
%the speech files stored in the database                   %
% INPUT  :  testVect  = speech segment from testing set    %
%           dbVect    = speech vector from database         %
%                                                          %
% OUTPUT :  corrSpeech  = closest matching speech segment %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% length of test vector
lenTest = length(testVect);

% compute the correlation
corrVal = conv(dbVect,fliplr(testVect));
% normalize
normdB = conv(dbVect .* dbVect,ones(1,lenTest));
corrVal = corrVal./(sqrt(normdB));

% extract the relevant part of the
corrVal = corrVal(lenTest : end-lenTest+1);

% find the peak value and location
[pkVal,loc] = max(abs(corrVal));

% extract the speech segment
corrSpeech = dbVect(loc:loc+lenTest-1);
```

4. After obtaining the replacement segment from the database for each *x1* and *x2*, those replacement segments are stored as *X1* and *X2* respectively. *X1* and *X2* both had to be added together in an overlapping manner with the last half of *X1* being added with the first half of *X2*. They must be added in a manner where if *X1* is reducing *X2* must be increasing which is the purpose of *fact1* and *fact2*. The last half of *X1* and beginning half of *X2* are multiplied by *fact1* and *fact2* and the results are stored into *wtSeg1* and *wtSeg2*. Those values are then added together and stored in their respected locations in the reconstructed speech, *conSpeech*.

```
   Table 6 MatLab Code for Speech Reconstruction

% set the weighting factors
fact1 = (-1/fwdShift)*(1 : fwdShift) +1;
fact2 = (1/fwdShift) * (1 : fwdShift);
%scale values to combine the blocks
    wtSeg1 = X1(SampSize-63 : SampSize) .* fact1;
    wtSeg2 = X2(1 : fwdShift) .* fact2;
%reconstruct speech
conSpeech(start+64: last) = [(wtSeg1 + wtSeg2)];
```

EXPERIMENTAL RESULTS

Experimental Result #1 (SNR vs. Number of Speech Files Searched)

Taking one input signal and running it through the pre-recorded database yielded the following results. Shown in Figure 2 is a graph that depicts the Signal-to-Noise Ratio (SNR) of the number of utterances searched in the library files. The SNR for this test was taken for every 30 utterances of the database.
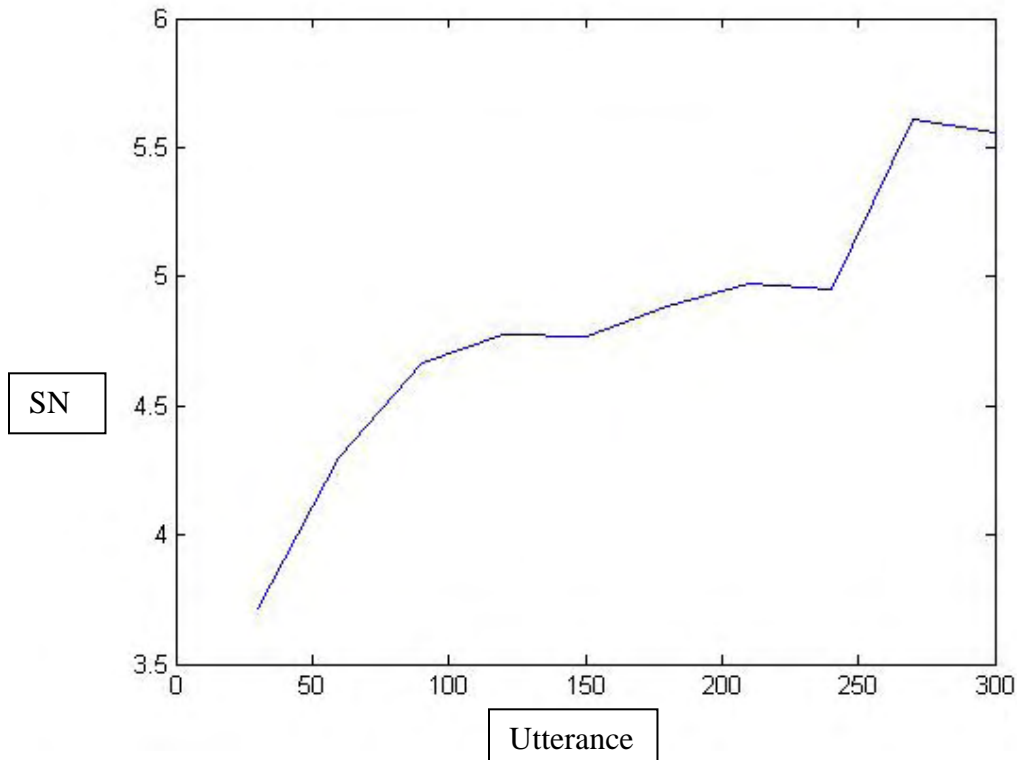

Figure 2: SNR vs Utterances

Experimental Result #2 (Effectiveness of De-noising vs. SNR)

The waves in Figure 3 depicts the input speech signal in the first wave. The following waves show different levels of noise, with each wave decreasing in noise the further down the figure. This figure helps give a visual understanding that with the higher the noise level the further away the signal looks from its original.
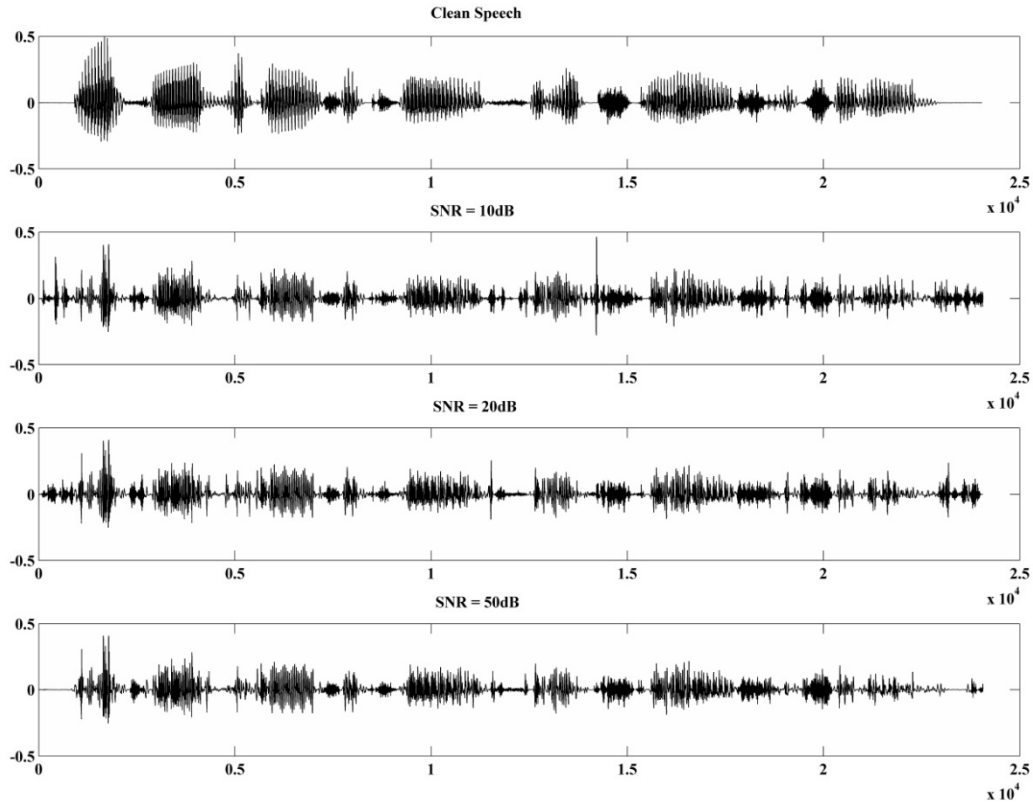
Figure 3: Clean speech with additive Gaussian noise variations

DISCUSSION

From the tests ran for figure 2 the graph was expected to show an increase in SNR that finally reaches a flattened region as the number of utterances are increased. If the utterances 90 to 240 are examined it can be seen that there is a flattening in the SNR graph leading to the conclusion that the SNR was reaching the flattened region. However, there was a significant rise in the SNR at 270 utterances, which negates the assumption that the graph flattening was approaching. This could be due to an insufficient number of utterances being searched or possibly to an inconsistency in the MATLAB code.

For the tests in figure 3, the waves following the clean signal were given a significant amount of noise. The higher the dB the more noise present in the speech signal. In the case with 10 dB, it can be noticed that there is much more noise present on the signal than in 20 dB case. Likewise, if examined closely it is observed that 20 dB has more noise present than in 50 dB. The 50 dB wave is the closest to the original clean speech signal and is contaminated with small amounts of noise.

CONCLUSION

In this research it was shown that various forms of convolution are much more efficient and less time consuming than using the direct linear convolution computation. Due to time constraints the final code of this project did not utilize the *Overlap-Add* method, which would have substantially decreased the amount of time needed to run a sample. However, in the experiments the linear convolution was used, resulting in a large amount of time per run. The results in the SNR experiment were not as expected because it did not possess the flat point which represented where the signal could become no cleaner. In the second experiment, the graph displaying various levels of noise in the processing the contaminated utterance produced expected results that showed a closer similarity to the original signal the higher the SNR was. Future studies will investigate the use of the MDFT to achieve even more computational efficiency.

ACNOWLEDGMENTS

REFERENCES

[1] L. Hakansson, S. Johansson, M. Dahl, P. Sjosten and I. Claesson, *Noise Cancelling Headsets for Speech Communication*," pp. 305-306. Edited by: Gi M. Davis, CRC Press

[2] A. V. Oppenheim and R. W. Schafer, *The Discrete Fourier Transform*, pp. 99, 112-113, Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[3] J. G. Proakis and D. G. Manolakis, "Digital Signal Processing: Principles, Algorithms, and Applications, 4th edition, Prentice Hall, Inc, 2006.

[4] X. Xiao and R. M. Nickel, "Speech enhancement with inventory style speech resynthesis," IEEE Trans. Audio, Speech and Lang. Process., to appear.

[5] X. Xiao, P. Lee, and R. M. Nickel, "Inventory based speech enhancement for speaker dedicated speech communication systems," in Proceedings of the *IEEE Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 3877–3880.

[6] C. Radhakrishnan, S. Raza, R. N. Nickel, and W. K. Jenkins, "Fault Tolerant Fermat Number Transform Domain Adaptive Filters Based on Modulus Replication RNS Architectures," Proceedings of the Forty-Third Annual Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, pp. 1365 - 1369, November 1-4, 2009.

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

[8] C. Radhakrishnan and W. K. Jenkins, "Modified Discrete Fourier Transforms for Fast Convolution and Adaptive Filtering," *Proc. Int. Symp. Circ. Syst.*, pp. 1611-1614, May 30 - June 2, 2010.

[9] Zehra Raza, Robert Nickel, and W. Kenneth Jenkins, "Speech Processing using Matched Filters Based on Transform Domain Processing", *Annual Research Journal - Electrical Engineering Research Experience for Undergraduates*, Department of Electrical Engineering, Penn State University, University Park, PA (Journal/In House Publication), p. 89-105, vol. VII, (2009), ISBN 0-913260-09-6.

ROBUST VIDEO HASHING USING THE THREE DIMENSIONAL DISCRETE COSINE TRANSFORM

Miloud Id Elhaj* and Vishal Monga[#]

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

*Undergraduate student of
Department of Electrical Engineering
Southern Methodist University
Dallas, TX, 75205

ABSTRACT

With the huge amount of information flowing through the internet, one might question the authenticity of the information he or she is getting from the web. Easy manipulation and reproduction of multimedia content further gives rise to the problem of content piracy. An important advance in identifying and securing video content has been that of video hashing (also called video fingerprinting)—where the goal is to extract a short vector called hash from the video which can help reliably identify the video (under distortions) as well as help track it for avoiding/detecting piracy. In this project, I implement a robust hashing algorithm for video data using the 3-D Discrete Cosine Transform (DCT). This function is desired to be robust against incidental processing of the video such as compression, enhancement, and rate changes while being able to detect malicious and significant content changes.

INTRODUCTION

Hashing is the transformation of a string of characters into a usually shorter fixed-length value or key that represents the original string. Hashing is used to index and retrieve items in a database because it is faster to find the item using the shorter hashed key than to find it using the original value. It is also used in many encryption algorithms [1]. In a similar manner and purpose as mentioned earlier, hash functions are used in identifying multimedia applications. Ideally, hash functions can be generated for any video, however, in practice that hash function is unreliable if a bit change between two independently generated hashes

---

[#] Faculty Mentor

of essentially the same video would be regarded as comparing two distinct videos. In addition, if the number of elements representing a video increases, as a consequence the probability of a bit mismatch increases. This repeat addresses this issue by making the hash function more robust to accidental or deliberate changes during video processing. For instance, the original video, its filtered version, its cropped version, its rotated version, its blurred version and its adjusted contrast version should yield relatively close hash values (Using a threshold value)—since their content is the same but their binary forms are very different [2].

Video hashing is being actively researched for applications in search content tracking and secure authentication. Coskun et al. [2] obtain their hash values by applying the 3-D DCT transform to videos, and extract the most significant coefficients in the low frequency spectrum. They also used a randomized version of the 3-D DCT to generate a unique hash vector for a given video. Monga and Mihcak [4] use non-negative matrix factorization (NMF) by obtaining NMF vector and then generate a pseudorandom weight vector $\{ti\}_{i=1}^{M}$. Caspi and Bargeron [5] use Gaussian Pyramid to rescale the individual frames of the input video to produce low resolution images, and then create a threshold using the median gray level producing a binary hash vector for each frame. Fridirich [6] takes the low frequency coefficients in still images and set a threshold to generate a robust hash vector with unique characteristics.

Our implementation in this paper is based on Coskun's approach which takes into account the temporal as well as the spatial features of a video. As a result, if time related changes (i.e., frame rate change, frame dropping) take place in video processing, the hash values after the processing should be closer to that found in the original video. Further, the technique we are presenting can be regarded as more robust than the technique used in Coskun's paper. In our approach we not only considered the coefficients found in the low frequency region, but we extended the selection to middle and high frequency regions and each of the three regions were weighted differently.

ALGORITHMS FOR VIDEO HASHING

DCT Transform

The discrete cosine transform (DCT) helps separate the image or video into parts of differing importance [7]. Using this transform we are able to concentrate the most significant coefficients in a small region of low frequencies as shown in Figure 1. The mathematical representation of the 3-D DCT used in this paper is shown in equation (1).

$$B(i,j,k) = \alpha(i)\alpha(j)\alpha(k) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \sum_{z=0}^{P-1} A(x,y,z) \cos\frac{\pi(2x+1)i}{2M} \cos\frac{\pi(2y+1)j}{2N} \cos\frac{\pi(2z+1)k}{2P} \quad (1)$$

where
$$0 \le i \le M-1, \quad 0 \le j \le N-1, \quad 0 \le k \le P-1$$

$$\alpha(i) = \frac{1}{\sqrt{M}} \,, i = 0; \quad \alpha(i) = \frac{\sqrt{2}}{\sqrt{M}} \,, 1 \le i \le M$$

$$\alpha(j) = \frac{1}{\sqrt{M}} \,, j = 0; \quad \alpha(j) = \frac{\sqrt{2}}{\sqrt{N}} \,, 1 \le j \le N$$

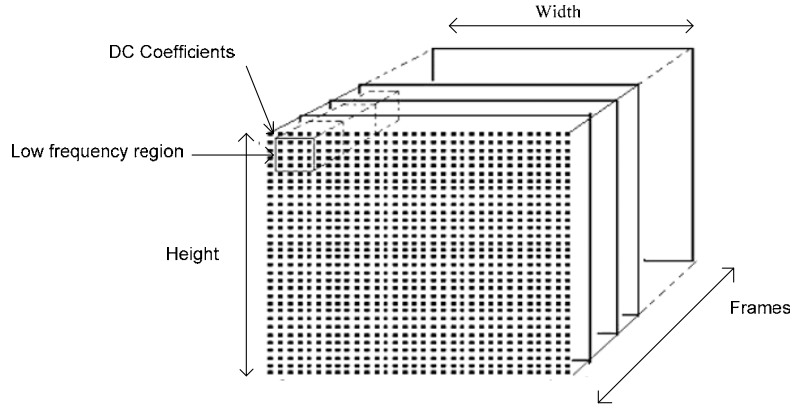$$\alpha(k) = \frac{1}{\sqrt{P}} \,, k = 0; \quad \alpha(k) = \frac{\sqrt{2}}{\sqrt{P}} \,, 1 \le k \le P$$

Figure 1: The low frequency coefficients are found in the low frequency region. (Coskun)

Using this algorithm to generate the hash function would be insufficient when computing a large set of matrices. As depicted in Figure 1, videos are sets of frames (images), when using equation (1) a person needs six nested loops which increase the cost and time needed to compute the results. A faster way of computing the DCT coefficients is by applying the FFT (Fast Fourier Transform) originated from DFT (Discrete Fourier Transform) [8] and has an order of Nlog(N). Here we show the 1D DCT (For simplicity) as a function of DFT in equation (2).

Given the DCT in one dimension as

$$DCT[p] = \alpha(p) \sum_{x=0}^{N-1} a[x] \cos \frac{\pi(2x+1)p}{2N}$$

We can find the DCT in relation to DFT as shown here:

$$DCT[p] = \sum_{x=0}^{N/2-1} a[2x] \cos \frac{\pi(2(2x)+1)p}{2N} + \sum_{x=0}^{N/2-1} a[2x+1] \cos \frac{\pi(2(2x)+3)p}{2N}$$

$$DCT[p] = \sum_{x=0}^{N/2-1} b[x] \cos \frac{\pi(4x+1)p}{2N} + \sum_{x=0}^{N/2-1} b[N-1-x] \cos \frac{\pi(4x+3)p}{2N}$$

where $b[x] = a[2x]$ and $b[N-1-x] = a[2x+1]$.

For $x' = N - 1 - x$

$$\sum_{x=0}^{N/2-1} b[N-1-x] \cos \frac{\pi(4x+3)p}{2N} = \sum_{x'=N/2}^{N-1} b[x'] \cos(2p\pi - \frac{\pi(4x'+1)p}{2N})$$

$$= \sum_{x'=N/2}^{N-1} b[x'] \cos \frac{\pi(4x'+1)p}{2N}$$

By setting $= x'$, and adding the two summations of DCT we get:

$$DCT[p] = \sum_{x=0}^{N-1} b[x] \cos \frac{\pi(4x+1)p}{2N} \quad \text{(a)}$$

Next, we have:

$$DFT[p] = \sum_{x=0}^{N-1} b[x] e^{-j2\pi xp/N} = \sum_{x=0}^{N-1} b[x](\cos \frac{2\pi xp}{N} - j\sin \frac{2\pi xp}{N})$$

Multiplying both sides by $e^{-j\pi p/2N}$ and taking the real part of DFT[p], we get:

$$Re\left[e^{-\frac{j\pi p}{2N}} DFT[p]\right] = \sum_{x=0}^{N-1} b[x] \cos \frac{2\pi xp}{N} \cos \frac{\pi p}{2N} - \sin \frac{2\pi xp}{N} \sin \frac{\pi p}{2N}$$

Using the trigonometric identity $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$
We get:

$$Re\left[e^{-\frac{j\pi p}{2N}} DFT[p]\right] = \sum_{x=0}^{N-1} b[x] \cos \frac{\pi(4x+1)p}{2N} \quad \text{(b)}$$

Form (a) and (b) we see that:

$$DCT[p] = Re\left[e^{-\frac{j\pi p}{2N}} DFT[p]\right] \quad \text{(2)}$$

Euclidean Distance

To calculate the distance between hash vectors of the original videos on one hand and the original with their modified versions on the other hand, we used the Euclidean distance defined in equation (4)

$$X = \sum_{i=0}^{size(hash)-1}(hash1(i) - hash2(i))^2 \quad \text{(4)}$$

Where hash1 and hash2 are hash vectors of videos 1 and 2 respectively that wanted to be compared. Finally, the distance is computed using equation (5).

$$Distance = \sqrt{X} \quad \text{(5)}$$

EXPERIMENTAL SETUP

Read Videos and Preprocessing in MATLAB

We selected a set of 100 arbitrary videos which most of them came from YouTube. All the videos were 10 second or less in duration to accommodate MATLAB's memory space limitation. These videos are then converted to AVI format via a converter. Using the script shown in Appendix A, we were able to read the videos into MATLAB's workspace, and then we created a grayscale version of the original videos. The resulting 100 matrices (one for each video) are processed by applying seven different modifications (Attacks) on them as discussed in Table I. At the end of the modification process we accumulated 800 videos, 100 originals and 700 modified. The last step in the preprocessing was to standardize the size of all 800 videos to (32x32x64), so they could be suitably compared to each other.

Table I. Description of the modifications and their appearance in videos.

Modification Type	Description	Frame Sample
Original Video	The original video from YOU TUBE	
Blur	Applying Average filter to individual frames in a video, the size of the filter matrix was 8×8.	
Noise	Applying Gaussian noise to video frames, the mean and variance values are 0 and 0.01, respectively.	
Crop	Cropping 25% of the original size of the video, the cropped starts at the lower right corner of the video frame.	
Rotate	Rotating each frame a 15 degree clockwise its original position. We also cropped each frame to have the same size as the original video.	

Contrast adjustment	Equalizing the contrast of the original video by applying a histeq() MATLAB function on its frames.	
Rotate, Crop and resize	The frames are rotated 2 degrees clockwise, then the frames are cropped to eliminate the black areas and then the frames are resized to the original frames size.	
Noise and contrast	Applying the same Gaussian noise as discussed earlier and then the contrast of the noisy frames is modified.	

3D DCT Implementation and Processing

As mentioned in the introduction, the 3-D DCT gets the robustness from its ability to take both the temporal and spatial features of a video. We started the 3-D DCT implementation by applying the two dimensional DCT (2-D DCT) function found in MATLAB's library to each individual frame in a video, and with that we found the spatial coefficients of the transform. Then, the results are used temporally to produce a 3-D DCT matrix of size (32x32x64), and that matrix is used to generate the hash values. Next we set a threshold values, this value is determined solely on the observation of its impact on the 3D DCT matrix. The ideal threshold value is the one that minimizes the number of coefficients in a given 3-D DCT matrix, yet the video before and after the threshold should easily be identified as the same video to any individual. The code used to compute the 3-D DCT coefficients is shown in Appendix B.

After running the 3-D DCT function above for a number of times, we concluded that the value 50 was reasonable to proceed to the next step in which we partitioned the reduced 3-D DCT matrix into three regions. By reduction we mean the coefficients with absolute value less than 50 are set to 0, the size of 3-D DCT matrix remains the same (32x32x64). The three regions are region I of low frequency, region II of mid frequency and region III of high frequency. These selected regions make our hash values more robust than that used in Coskun's paper. Figure 2 illustrates how the partitioning is done on the reduced 3-D DCT matrix. The size of these regions was chosen as follows:

Region (1): This low frequency region holds the majority of the significant coefficients. We selected the first 5 rows and columns of the first 8 frames of the 3-D DCT matrix, this region carries 200 coefficients (5x5x8), and it weighs 42/64 of the hash values of a processed video.

Region (2): This region extends from the first frame to the 32nd, and the number of rows and columns extend to 16 and excludes the first 5 rows and

columns. This region carries 7992 coefficients (16x16x32—region (1)), and weighs 14/64 of the hash values.

Region (3): This region takes the remaining 57344 coefficients and contributes 8/64 of the hash values.
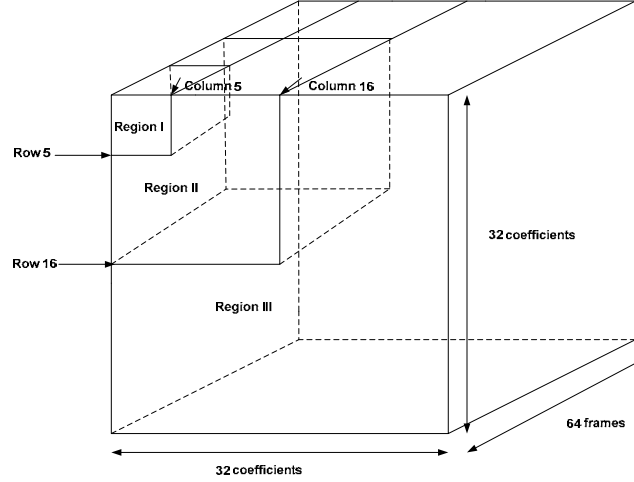

Figure 2: Partitioning of the 3D DCT matrix

Hash Vector and Distance

The three regions outputs three different vector sizes, region I gives a vector of size [1x42], region II gives a vector of size [1x14] and region III gives a vector of size [1x8]. Moreover, these vectors are concatenated to form the hash vector of size [1x64], where the first 42 elements represent region I and so on. We repeated this process until we created a hash vector for each of the 800 videos. Furthermore, we calculated the distance between the hash vectors of the original videos, so the number of distances calculated was 4950 as shown in (c).

$$Number\ of\ Distances\ calculated = \frac{100 \times 99}{2} = 4950 \qquad (c)$$

This is what we use to find the False Alarm probability, the code is provided in Appendix C. To find the probability of Miss-Detection we used the distances found by comparing each original video with its seven distorted versions, and the number of distances calculated was 700 as shown in (d).

$$Number\ of\ Distances\ calculated = 100 \times 7 = 700 \qquad (d)$$

RESULTS AND DISCUSSION

In this section we report the statistical findings after calculating all of the distances. The values of the false alarm and miss-detection shown in the box on each of the graphs indicate that the best value to set a threshold to determine if the videos are the same or not is at that point. We also provide alternative way we

think would make the algorithm even more robust than what we have already accomplished.

1) <u>Blurring</u>: The probability of miss-detection verses false alarm shown in Figure 3 indicates that even with that high amount of blurring shown in Table I., the algorithm proves its robustness against that kind of attack. The value found in this figure reveals that the hash in its best performance would have a probability of false alarm equal to 0.002699, and a probability of miss-detection equal to 0.013and both are very low.
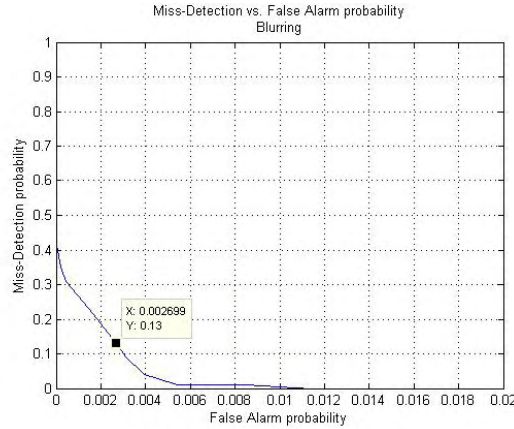

Figure 3: The probability values of the blurring attack, a) the actual size, b) magnified version of (a).

2) <u>Gaussian Noise</u>: Looking at the impact of this attack on the videos we have used (see Table I.), and the results found in Figure 4, we see that the hash can easily identify a video even with that amount of noise and refer it back to its original. Because of this robustness we managed to stay at low probability values, and as it is seen in the figure, the hash at its best gives probability of false alarm and miss-detection in the order of 0.008512 and 0.06 respectively.
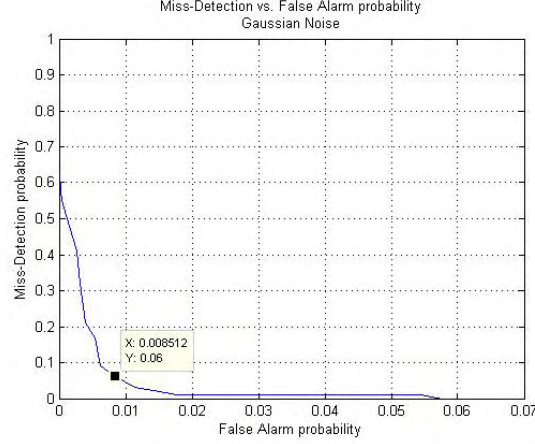
Figure 4: The probability values of the Gaussian noise attack, a) the actual size, b) magnified version of (a).

3) Cropping: In this attack, the hashing algorithm is applied to only 75% of the original video, and Figure 5 shows the resulting probability values. Even though the probability of false alarm and miss-detection 0.1604 and 0.16 respectively are a little high, it is accurate to say that the hash is robust. We say that because a person would uncomfortable watching only 75% of the original frames size in a video, so it is more likely for that person (Pirate) to use smaller value than 25%.
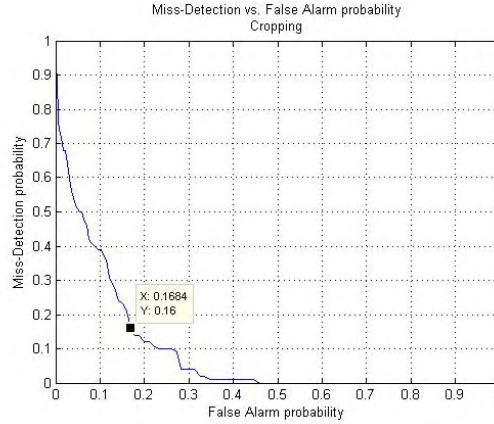

Figure 5: The probability values of the cropping attack

4) Rotating: In a similar manner, we found the values for the probability of false alarm and miss-detection from Figure 6 to be 0.2412 and 0.24 respectively. These values are somewhat high and might cause some problems identifying videos, but we still conceder the algorithm to be robust against this kind of attack. We say that the hash is robust because the video here is rotated 15 degrees from its original position, and usually if someone tries this attack the rotation would be in order of 1 to 5 degrees, and outside this range the video becomes hard to watch.
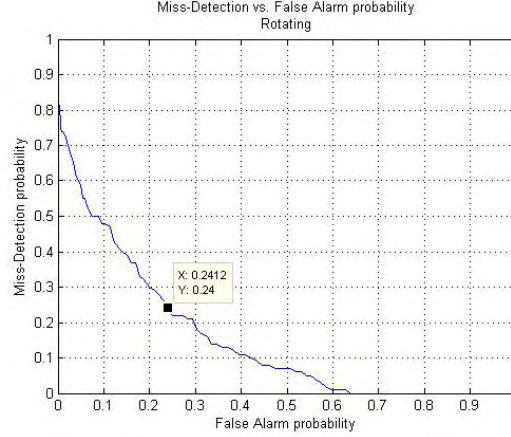
Figure 6: The probability values of the rotating attack

5) <u>Contrast Adjustment</u>: The probability values found in Figure 7 are 0.3029 for false alarm and 0.3 for miss detection. The quality of the videos after the contrast adjustment are overall reasonable, meaning that there are enough details in the videos for someone to actually watch them. Here, we say that the algorithm has a success rate of 70%.
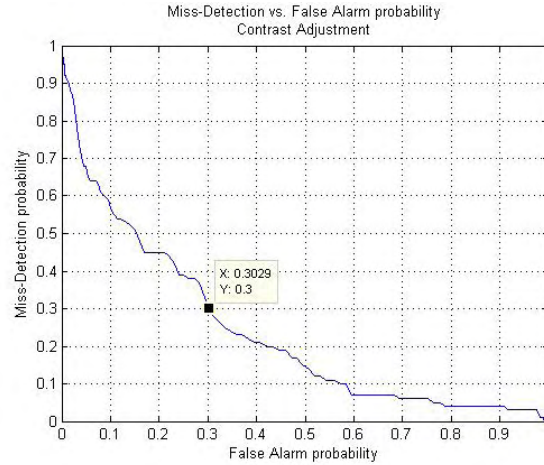
6)

Figure 7: The probability values of the contrast adjustment attack

7) <u>Combination attack 1: Rotating, Cropping and resizing</u>. As the title implies, we combined three different attacks two of which we have already applied the hash algorithm to them but with different parameters as explained in Table I. The probability values of false alarm and miss-detection found in Figure 8 are 0.3463 and 0.3 respectively. From these values we observe that the algorithm becomes vulnerable when multiple attacks are combined together.
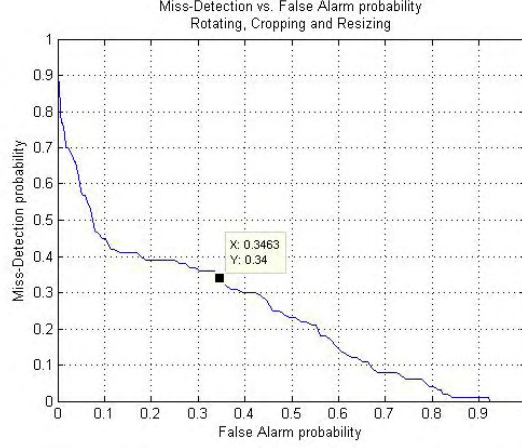
ROBUST VIDEO HASHING USING THE 3-D DCT

Figure 8: The probability values of the combination attack 1

8) <u>Combination attack 2: Gaussian noise and Contrast adjustment</u>. Here we have two attacks combined to further testing the relationship between the hash algorithm and the number of attacks. As a reminder the probability values for false alarm and miss-detection of the noise attack were very low, and values of the contrast attack were relatively high. Looking at Figure 9 we find that the values of false alarm and miss-detection are in between the ones found in each individual attack. The values for which the hash is at its best are 0.1594 and 0.16 for false alarm and miss-detection respectively. This attack proves the fact that the more attacks are combine the weaker the hash algorithm becomes.
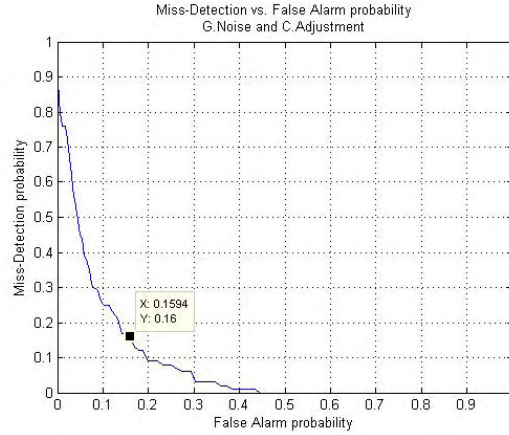
Figure 9: The probability values of the combination attack 2

Finally we evaluated the overall performance of the hash algorithm by combining the results found in all the seven attacks into one graph as shown in Figure 10. The probability of false alarm and miss-detection are 0.2001 and 0.1929, from these values we say that 4701 is our threshold value for the video set we collected, with 20% chance for the hash to fail. In other words, if a distance between two videos is less than or equal to 4701 then the videos are the same,

otherwise they are different. This algorithm would be more robust if we simply increase the size of the hash vector, the larger the hash vector the more robust the algorithm would be. We also can make the stander size (32x32x64) smaller than what it is in this algorithm and keep the hash vector of size [1x64].
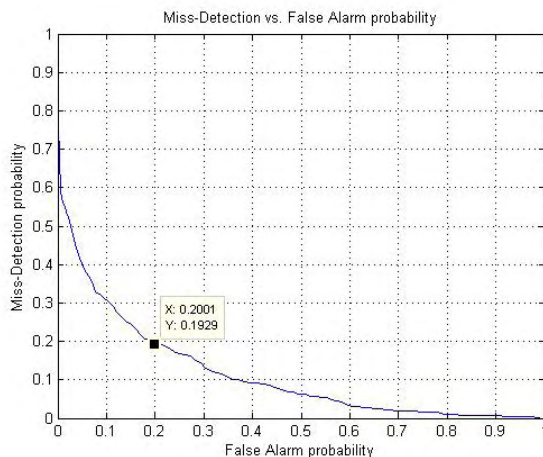

Figure 10: The overall performance of the hash algorithm used in this paper

CONCLUSION

The modifications used in the experiment were representative of real world scenarios, and reasonably the hash function showed high robustness against these distortions with very small probability for failure. In this project we consider largely image processing distortions applied to individual frames. However, temporal distortions such as non-uniform frame dropping are of great interest in video hashing. Because an adversary or a pirate may try to deliberately drop particular frames to beat a hash algorithm, a good direction for future work is to partition the video into random subsets and extract a 3-D DCT based hash for each subset.

ACKNOWLEDGEMENTS

REFERENCES

[1] http://searchsqlserver.techtarget.com/definition/hashing
[2] B. Coskun, B. Sankur, N. Memon, "Spatio-temporal transform based video hashing," IEEE Transactions on Multimedia, 8(6):1190-1208, Dec. 2006
[3] V. Monga and M. K. Mihcak, "Robust image hashing via non-negative matrix factorizations," in ICIP'05,Genoa, Italy, Sep. 2005.

[4] Y. Capi and D. Bargeron, "Sharing video annotations, international conference on image processing," in ICIP'04, Singapore, Oct. 2004.

[5] J. Fridrich and M. Goljan, "Robust hash functions for digital watermarking," in ITCC' 00: Proc.Int. Conf. Information Technology: Coding and Computing, 2000, p. 178.

[6] http://cs.cf.ac.uk/Dave/Multimedia/node231.html

[7] http://fourier.eng.hmc.edu/e161/lectures/dct/node2.html

APPENDIX A: VIDEO PROCESSING

```
%Miloud Id Elhaj
%REU Summer 2010
%Date: 07/06/2010

%This code will read any video format supported by matlab'and then convert
%that video to a grayscale video for further processing in matlab.
clear all; close all; clc
%these global variables are declared so that the functions called
%in this code can access their values

global DCT3; global IDCT3; global hash_region1; global hash_region2;
global hash_region3; global Hash_Vector; global I2; global I;
global frames; global a; global b; global A; global B; global C;

%reading an avi video file
Video = mmreader('Vid1.avi');

%Initializing variables needed to play the video in matlab
nFrames = round(Video.NumberOfFrames);
vidHeight = Video.Height;
vidWidth = Video.Width;

% Preallocate movie structure.
mov(1:nFrames) = ...
    struct('cdata', zeros(vidHeight, vidWidth, 3, 'uint8'),...
           'colormap', []);

% Read one frame at a time.
for k = 1 : nFrames
    mov(k).cdata = read(Video, k);
end

%Declaring the size of universal variables that would process any video
[l_struct, frames]  = size(mov);
[length, width, RGB] = size(mov(1,1).cdata);
Original_Vid         = uint8(zeros(length,width,frames));

%This iteration will convert each frame individually to a grayscale image
for i = 1:frames
Original_Vid(:,:,i) = rgb2gray(mov(1,i).cdata(:,:,:));
end

%resizing the videos so we provent the OUT OF MEMORY message in matlab
Original_Vid = imresize(Original_Vid,[150 200]);
```

```
%Miloud Id Elhaj
%REU Summer 2010
%Date: 07/06/2010
%Functions used to generate the hash vector, the first five functions take
%Original_Vid as their input variable
Blur_Vid(Original_Vid); % blur the video
Noisy_Vid(Original_Vid); % apply guassian noise
Cropped_Vid(Original_Vid); % crop the video
Rotate_Vid(Original_Vid); % rotate the video at a certain
Contrast_Vid(Original_Vid); %adjudt the contrast of a video

Resize_Vid(I); %resize the video to a standerd size
DCT3_Trans(I2); %apply the 3D DCT to the standerdized video
Region1_hash(DCT3); % gives the hash vector from region1
Region2_hash(DCT3); % gives the hash vector from region2
Region3_hash(DCT3); % gives the hash vector from region3
hashVector(hash_region1 , hash_region2 , hash_region3); % this is the
%final hash of size 1x64 and combines the hashes from the three different
%regions.
```

APPENDIX B: 3D DCT COMPUTATION

```
function [DCT3] = DCT3_Trans(I2)
%Miloud Id Elhaj
%REU Summer 2010
%Date: 07/06/2010

% function [DCT3]: takes the standardized video I2 as input
%This function computes the 3d dct coefficients of any grayscale video
%using matlab's 1d dct. The threshold is set on the resulting 3D matrix in
%this case we found that 50 is a good value to reduce the number of coeff.
%yet be able to recognize the features in the video.

global DCT3;

temp1 = zeros(double(size(I2))); %storage variable
temp2 = zeros(double(size(I2))); %storage variable

%applying the 2dct transform on each individual frame
for j = 1:size(I2,3)
temp1(:,:,j) = dct(dct(I2(:,:,j)).').';
end

D = temp1; % D is the 2dct of each individual frame in the video

for i = 1:size(I2,1)

    for k = 1:size(I2,2)
            temp2(i,k,:)= dct(D(i,k,:));
    end
end

DCT3 = temp2; % DCT3 is the 3d dct matrix where most of the %significant
coefficients are in the lower frequency spectrum

%setting the threshold suitable to the application of this function
DCT3(abs(DCT3) < 50) = 0;
```

APPENDIX C: FALSE ALARM PROBABILITY VALUES

```
%Miloud Id Elhaj
%REU Summer 2010
%Date: 07/06/2010

%this script loads hash vectors of the original videos to the workspace,
%and then it calculates the distance between these vectors,and this is
repeated for %different threshold values s.

clear all ; clc

%load the hash vectors
for l = 1:100
    load (['Hash_Vector' num2str(l)]);
end

for s = 1:100:14500
    temp2 = 0;
    temp3 = 0;

    for i = 1:99
        for j = i+1:100
        a = ['Hash_Vector' int2str(i)];
        A = eval(a);
        b = ['Hash_Vector' int2str(j)];
        B = eval(b);

        temp1 = 0;  % initializing the variable to 0

            %this loop with 64 iterations calculates the distance between
            %elements of two vectors

            for k = 1:size(A,2)
                X = (A(k) - B(k))^2;
                temp1 = temp1 + X;
            end

            Dist = sqrt(temp1);

            %checks for the threshold
            if Dist <= s
            temp2 = temp2 + 1;
            end

        temp3 = temp3+1; %outputs the number of times vectors are compared
        end
    end
    disp(num2str(temp2))% outputs the number of distances below the
threshold

end
```

DETERMINING CONVERSION GAIN IN A MARGINAL OSCILLATOR USING NUMERICAL SOLUTIONS

Michael W. Scruggs,[*] Jeffrey L. Schiano,[#] and Thomas J. Tyson[+]

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

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

ABSTRACT

A marginal oscillator uses a nonlinear feedback element to detect small changes in the losses of a resonant system. The key figure of merit for gauging the performance of a marginal oscillator is the conversion gain, which is the sensitivity of the oscillation amplitude to changes in losses. Miller observed discrepancies between the values of conversion gain obtained from theoretical predictions, numerical simulations, and experimental measurements [1]. We show that these discrepancies stem from the protocol used to determine the conversion gain in experiments and simulation studies, and the choice of the integration algorithm used in numerical simulations. By choosing an appropriate ordinary differential equation solver and refining the protocol for determining conversion gain, we demonstrate that the conversion gain estimated from numerical simulations closely matches the value predicted by theory.

INTRODUCTION

Marginal Oscillator

Pound developed the first marginal oscillator in the 1940s as a means for detecting the absorption of radio frequency energy in nuclear magnetic resonance experiments [2, 3]. The marginal oscillator allowed Pound to study molecular structure using nuclear magnetic resonance [4]. Recent applications of marginal oscillators include monitoring mechanical properties of thin-films [5], the measurement of penetration depth in superconductors [6], characterization defects in silicon [7], and ion cyclotron resonance spectroscopy [8].

[#] Faculty Mentor
[+] Graduate Mentor

Figure 1 shows that a marginal oscillator consists of a resonant circuit driven by a memoryless, nonlinear voltage controlled current source. A parallel resistor-capacitor-inductor (RLC) network represents the resonant circuit, where the resistance R accounts for losses within the network. To sustain a steady-state sinusoidal oscillation, the voltage controlled current source must provide power to exactly overcome these losses. Equivalently, the voltage controlled current source must appear as a negative resistance with value $-R$, so that the parallel LC circuit sees an open circuit. The circuit is called a marginal oscillator because the feedback path provides just enough energy to sustain the oscillation. The values of inductance and capacitance determine the frequency of oscillation. By appropriately choosing a nonlinear function $i=G(v)$, small changes in the losses R are revealed by proportionately larger changes in the amplitude of oscillation.
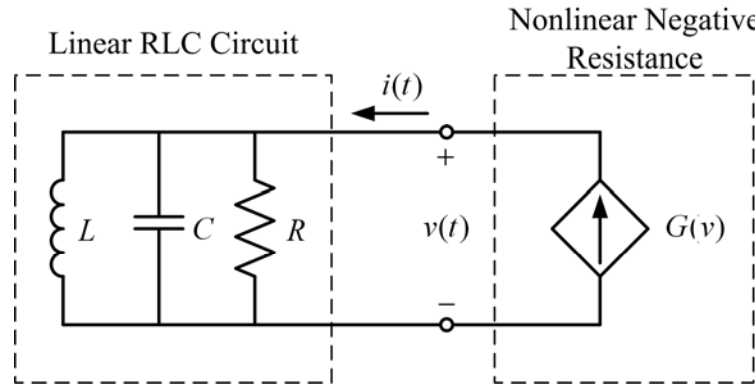

Figure 1: Schematic representation of a marginal oscillator

A second-order nonlinear differential equation for describing the marginal oscillator is obtained in two-steps. First, represent the linear RLC circuit in Figure 1 by the ordinary differential equation

$$\frac{d^2v}{dt^2} + \frac{1}{RC}\frac{dv}{dt} + \frac{1}{LC}v = \frac{1}{C}\frac{di}{dt}. \tag{1}$$

Second, replace the current $i(t)$ by $G(v)$, and define $dG(v)/dv$ as $g(v)$ to obtain

$$\frac{d^2v}{dt^2} + \frac{1}{C}\left(\frac{1}{R} - g(v)\right)\frac{dv}{dt} + \frac{1}{LC}v = 0. \tag{2}$$

Observe that $-g(v) = -di/dv$ represents the conductance looking into the nonlinear negative resistance. When the conductance $-g(v)$ cancels $1/R$, the coefficient of dv/dt vanishes and the marginal oscillation achieves a steady-state sinusoidal oscillation.

Ideally, one could choose the linear function $G(v) = v/R$, so that the resistance $-v/i$ looking into voltage controlled current source is $-R$. Due to component tolerances, the value of $G(v)$ in a practical circuit is either slightly smaller or larger than $1/R$. Correspondingly, the coefficient $1/R - g(v)$ multiplying dv/dt in Equation 1 is either negative or positive, which leads to an increasing, or decreasing oscillation amplitude, respectively.

Rather than attempting to design a circuit where $G(v)$ exactly matches v/R, we implement $G(v)$ as a piecewise linear characteristic, with two symmetrical limiting segments and an amplifying segment as shown in Figure 2. When the magnitude of the voltage v is less than a threshold voltage v_T, the slope g_A is chosen larger than $1/R$, so that the coefficient $1/R - g(v)$ is negative. This leads to an increasing oscillation amplitude. Once the magnitude of oscillation exceeds v_T, the slope decreases to g_B so that $1/R - g(v)$ becomes positive and the oscillation amplitude decreases. It can be shown that the oscillation amplitude converges to a steady-state value following a brief transient response [9]. As the slope of $G(v)$ is not constant, the negative resistance seen looking into the current source is a nonlinear function of the input voltage v.
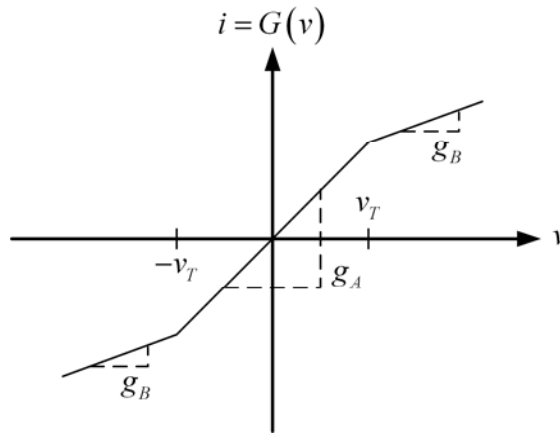

Figure 2: Voltage-current characteristics of the dependent current source

For the purpose of analysis, it is convenient to represent the marginal oscillator as the feedback system in Figure 3. This representation divides the marginal oscillator into a linear feed-forward block representing the linear RLC circuit, and a memoryless, nonlinear feedback block representing the nonlinear negative resistance. The frequency response function of the RLC circuit in the feed-forward block follows directly from Equation 1,

$$H(j\omega) = \frac{V(j\omega)}{I(j\omega)} = \frac{\dfrac{1}{C}j\omega}{(j\omega)^2 + \dfrac{1}{RC}j\omega + \dfrac{1}{LC}}, \tag{3}$$

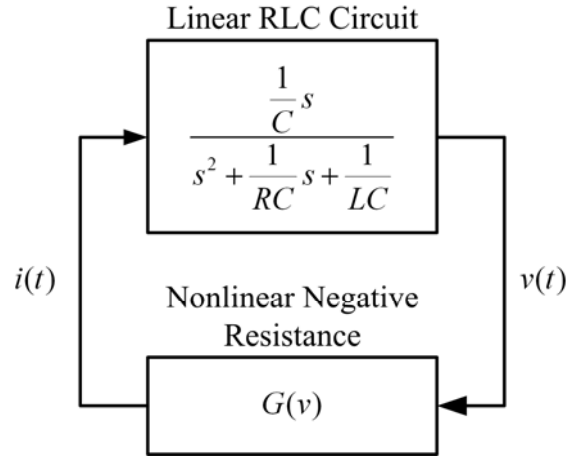and represents the impedance looking into the RLC circuit.

Figure 3: Representation of marginal oscillator as a feedback system

A plot of the frequency response function reveals the oscillation frequency and the bandwidth of the linear RLC circuit. Figure 4 shows the magnitude of $H(j\omega)/R$ as a function of ω. At the resonant frequency

$$\omega_o = \frac{1}{\sqrt{LC}}, \tag{4}$$

the impedance looking into the RLC network is R. The 3-dB bandwidth of the frequency response function is

$$\beta = \frac{1}{RC}. \tag{5}$$

The Q-factor provides a figure-of-merit for a resonant system and is defined as

$$Q = 2\pi \frac{\text{Energy stored}}{\text{Energy dissipated per cycle}}, \tag{6}$$

and for the parallel RLC circuit, the Q-factor is

$$Q = \frac{\omega_o}{\beta} = \omega_o RC. \tag{7}$$

As the electrical losses in the parallel RLC circuit decrease with increasing values of R, the Q-factor increases while the 3-dB bandwidth decreases. In general, a

high Q-factor is desirable because the signal-to-noise ratio of oscillator amplitude measurements is proportional to the square-root of the Q-factor [9].
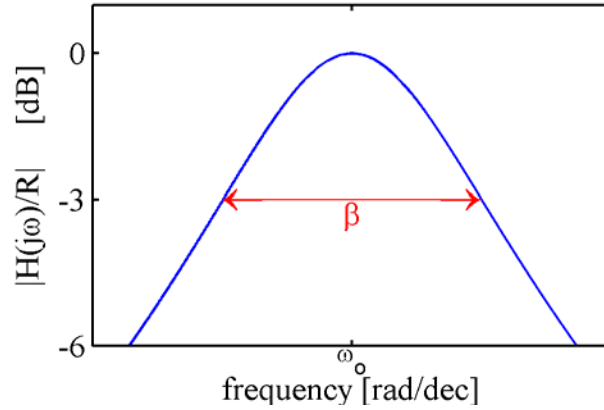

Figure 4: Frequency response function of the RLC circuit

From the standpoint of implementing numerical simulations, it is useful to represent the linear RLC circuit using the all-integrator block diagram in Figure 5. The integrator outputs define the state variables x_1 and x_2, whose initial values are

$$x_1(0) = LC\left[G(v(0)) - \frac{v(0)}{R} - C\dot{v}(0)\right] \qquad (8)$$

$$x_2(0) = Cv(0), \qquad (9)$$

expressed in terms of the voltage $v(t)$.

Conversion Gain

The sensitivity of the oscillation amplitude A of a marginal oscillator with respect to variations in the losses R of the resonant circuit is defined as the conversion gain [9]

$$G_c = \frac{\%\ \text{change in A}}{\%\ \text{change in R}} = \frac{\Delta A/A}{\Delta R/R}. \qquad (10)$$

For a parallel RLC circuit driven by an independent current source at the natural frequency, the conversion gain is unity. In this configuration, a 1% increase in the value of R results in a 1% increase in the amplitude of oscillation. The marginal oscillator achieves conversion gains significantly larger than unity by employing a nonlinear dependent current source $G(v)$. A large conversion gain is desirable as it improves the signal-to-ratio when measuring changes in the resistive losses R of the resonant parallel RLC circuit.
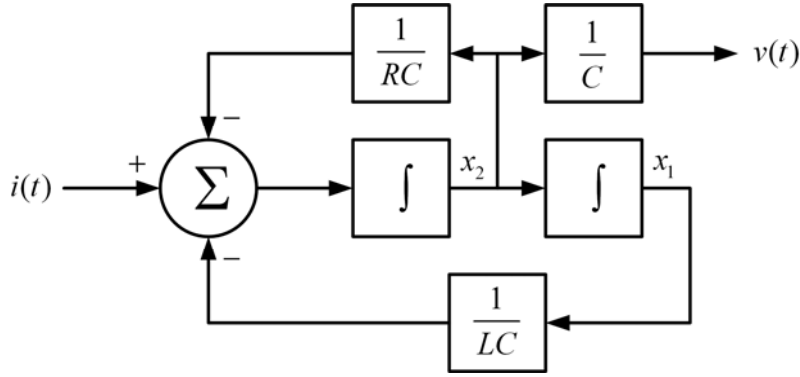
Figure 5: All-integrated block diagram representation of the RLC circuit

For a marginal oscillator using the nonlinear negative resistance described by the voltage-current characteristics in Figure 2, Viswanathan showed that the conversion gain is [9]

$$G_c = \frac{\pi}{2\left(1 - g_B/g_A\right)g_A R \sin\theta},$$ (11)

where the angle $0 \le \theta \le \pi/2$ satisfies

$$\sin 2\theta + 2\theta = \frac{\pi}{1 - g_B/g_A}\left\{\frac{1}{g_A R} - \frac{g_B}{g_A}\right\}.$$ (12)

The oscillation amplitude,

$$v_T = A\sin\theta,$$ (13)

is set by the solution θ to Equation 12, and the threshold voltage v_T defined in Figure 2.

In a typical application, the nominal loss value R is known, and the user specifies the desired amplitude A of oscillation and the conversion gain G_c. The user then chooses g_A so that $g_A R > 1$, and solves Equations 11 through 13 for θ, g_B, and v_T. Combining equations 11 and 12 to eliminate g_B, provides an expression for finding θ in terms of known quantities

$$\sin 2\theta\left[1 + 2G_c\left(g_A R - 1\right)\right] + 2\theta = \pi.$$ (14)

Using the solution for θ in Equation 11, gives an expression for g_B

$$g_B = g_A - \frac{\pi}{2G_C R \sin 2\theta}.$$ (15)

Finally, the threshold voltage v_T is found by solving Equation 13.

Previous Results

Miller compared the conversion gain predicted by Equations 11 through 13 against that observed in numerical simulations and experiments [1]. Figure 6 shows the conversion gains observed by Miller as function of the ratio g_B/g_A. As the value of the ratio g_B/g_A approaches unity, the difference between the conversion gain observed in experiment and that predicted by theory and simulation widens. It was initially thought that this discrepancy results from the fact that the function $G(v)$ used in the experiments has a continuous derivative at the threshold voltage. To test this hypothesis, the marginal oscillator was simulated using a function $G(v)$ that closely matches that used in the experiments. This modification did not significantly change the simulation results.

Hypothesis

The aim of the present study is to determine if the discrepancy between the theoretical predictions and numerical simulations presented by Miller are due to the selection of the ordinary differential equation solver, and the protocol for determining the conversion gain. It is well known that the integration algorithm used to numerically solve an ordinary differential equitation has a significant impact on the accuracy and computation time of the simulation [10]. The particular integration algorithm employed by Miller may not accurately capture the dynamic behavior of the marginal oscillator. In addition, Miller's protocol for determining conversion gain may be flawed. Miller fixed the parameters g_A, g_B, and v_T in the function $G(v)$, and then simulated the marginal oscillation for two different values of electrical losses, R and $R + \Delta R$. Given the steady-state amplitudes observed in both simulations, Miller determined the conversion gain using Equation 10. This protocol may fail if the value of ΔR is too large, as Viswanathan's results show that the oscillation amplitude A is a highly nonlinear function of the electrical losses R.
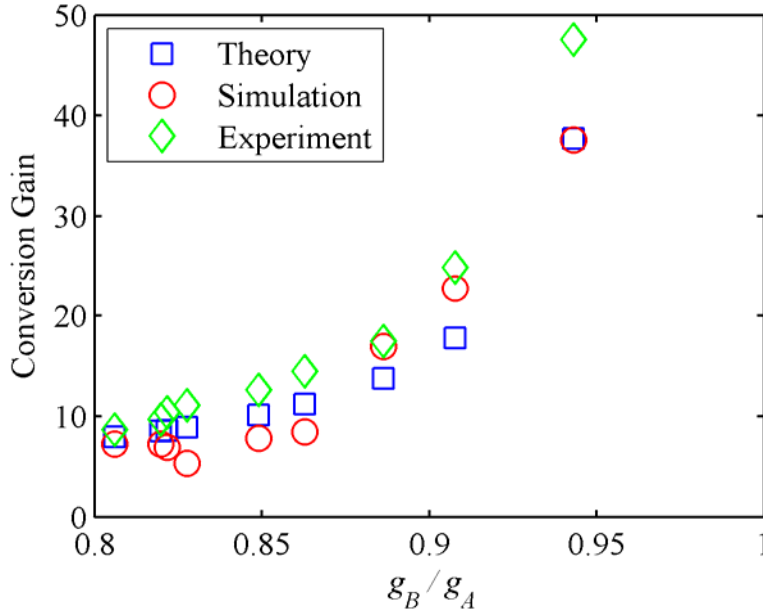
Figure 6: Conversion gain obtained from theoretical predictions, numerical simulations, and experimental measurements [1]

METHODS

Marginal Oscillator Test Bed

The simulations in this study use the parameters in Table 1 that were obtained from a marginal oscillator test bed. The inductance of the coil was measured using a vector network analyzer, whose source frequency matched that of the marginal oscillator. The resonant frequency and Q-factor of the RLC circuit was also measured using the vector network analyzer. Given the measured inductance and observed resonant frequency, the capacitance C was estimated using Equation 4. Finally, the resistance R is obtained from the estimated capacitance, measured Q-factor, and observed resonant frequency using Equation 7.

Table 1: Marginal oscillator parameters

Parameter	R	L	C	Q
Value	64.7 kΩ	26.8 μH	102 pF	127

Determining Conversion Gain from Simulations

Following Miller, conversion gain is estimated by expressing Equation 10 as

$$G_c = \frac{\partial A}{\partial R}\frac{R}{A} \approx \frac{(A_1 - A_0)}{(R_1 - R_0)}\frac{R_0}{A_0},\tag{16}$$

where A_1 and $A_0 < A_1$ represent the steady-state oscillation amplitudes obtained using resistances R_1 and $R_0 < R_1$, respectively, in the parallel RLC circuit [1]. Setting $R_1 - R_0 = \Delta R$ and $A_1 - A_0 = \Delta A$,

$$G_c \approx \frac{\Delta A}{\Delta R} \frac{R_0}{A_0}. \qquad (17)$$

The latter approximation for conversion gain holds as long as the change in loss ΔR is sufficiently small so that the amplitude A is a linear function of R over the interval $[R_0, R_1]$.

Equations 11 through 13 provide a way of determining the oscillation amplitude A as function of resistive losses R for fixed values of g_A, g_B, and v_T. Using the nominal loss resistance of 64.7 kΩ observed in the test bed marginal oscillator, we chose g_A so that $g_A R = 1.1$. We then determined three sets of values for g_B and v_T that yield conversion gains of 10, 25, and 50, respectively, at an oscillation amplitude of 250 mV. For the three fixed parameter sets g_A, g_B, and v_T, the amplitude of oscillation is calculated using Equations 12 and 13 as the resistance R was varied from 62 kΩ to 64.7 kΩ. Figure 7 shows the variation in oscillation amplitude with respect to resistance for the three sets of parameters. The solid, dash-dotted, and dotted curves correspond to fixed values of g_A, g_B, and v_T that yield conversions gains of 10, 25, and 50, respectively, for the nominal loss resistance of 64.7 kΩ. At the nominal value of resistance, all three parameter sets yield an oscillation-amplitude of 250 mV. As the value of resistance decreases, the oscillation amplitude also decreases. The three curves show that as conversion gain increases, the variation in oscillation amplitude with respect to resistance becomes nonlinear. The implication is that as ΔR increases, the value of conversion gain predicted by Equation 17 becomes less accurate.

Simulations

The behavior of the marginal oscillator is simulated using Simulink, the graphical user interface to the computer-aided design tool MATLAB. Within the Simulink environment, we represent the marginal oscillator using the all-integrator block diagram in Figure 8. An alternative implementation in MATLAB, using function calls to a macro that computes the derivatives of the state variables, is undesirable as it leads to significantly larger simulation times.

Figure 7: Variation of amplitude with respect to resistance for fixed values of g_A, g_B, and v_T

Figure 8: Representation of the marginal oscillator in Simulink

Integration Algorithms

MATLAB allows the user to select one of several integration algorithms [10]. The selection of a particular algorithm is determined by two factors. The

first factor is whether or not the ordinary differential equation to be solved is stiff. An ordinary differential equation is said to be stiff if the ratio of the largest and smallest time constant of the solution is large, or, if the properties of the ordinary differential equation change with respect to time. As an example, Equation 1, the ordinary differential equation describing the linear RLC circuit is non stiff. On the other hand, Equation 2, the ordinary differential equation describing the marginal oscillator, is stiff because the coefficient of *dv/dt* varies with respect to *v(t)*. The second factor determining the selection of an integration algorithm is the trade-off between simulation time and accuracy. One-step solvers determine the solution at particular time point using only the last solution point in time, whereas multi-step solvers use solutions obtained at several earlier time points to generate a more accurate solution at the expense of increased simulation time. Table 2 categorizes four integration algorithms, ODE45, ODE23s, ODE113, and ODE15s, as being either one-step or multi-step solvers, and as being suitable for either stiff, or nonstiff ordinary differential equations.

Table 2: Integration Algorithms and Properties

	Nonstiff	Stiff
One-Step Solver	ODE45	ODE23s
Multistep Solver	ODE113	ODE15s

Another classification of integration algorithms is whether they are fixed-step or variable-step solvers. Fixed-step solvers use the same time interval between solution points. In contrast, variable-step solvers vary the time interval between solutions points. When the solution changes rapidly with respect to time, the distance between adjacent time points is decreased. Similarly, when the solution fluctuates slowly with respect to time, the variable-step solver increases the time interval between solution points. Variable-step solvers provide two advantages. They decrease the simulation time by reducing the number of solution points in time intervals where the solution is changing slowly with respect to time, and as result, they also minimize the storage requirements by providing less solution points. The integration algorithms listed in Table 2 may be configured to execute as either fixed-step or variable-step solvers.

The accuracy and speed of the integration algorithms in Table 2 are also determined by two parameters known as the absolute and relative tolerance. During each time step, the integration algorithm estimates the simulation error, and if the error exceeds a specified tolerance, the algorithm automatically performs additional iterations to reduce the error. The relative tolerance measures the error relative to the size of each state variable, while the absolute tolerance represents the acceptable error as the value of the measured state approaches zero.

In this study we use the ODE15s integration algorithm because the ordinary differential equation describing the marginal oscillator is stiff, and we want to increase the accuracy of the solution by using a multi-step solver. In order to reduce the simulation time and reduce the memory required for storing solutions,

the ODE15s integration algorithm is executed as variable-step solver. The values of the absolute and relative tolerances are chosen by performing simulations to investigate the tradeoff between accuracy and simulation time.

RESULTS

Table 3 shows the results of simulation experiments used to select the value of absolute and relative tolerances. These simulations use the marginal oscillator parameters specified in Table 1. The target conversion gain and amplitude of oscillation are 10 and 250 mV, respectively. The value of g_A is chosen so that $g_A R$ is 1.03. The behavior of the marginal oscillator is simulated over a 10 ms window using the variable-step integration algorithm ODE15s. The duration of the window is chosen sufficiently large so that the oscillation amplitude achieves a steady-state value. In each trail the initial conditions are $v(0) = 0.9 \, v_T$ and $dv(0)/dt = 0$. Table 3 shows the simulation time and steady-state oscillation amplitude for five different sets of absolute and relative tolerances.

The results in Table 3 show that as the tolerance values are reduced, the steady-state amplitude approaches the target value of 250 mV while the simulation time increases. Tolerance values smaller than those used in Trial 3 yield negligible improvements in the steady-state amplitude at the expense of considerably longer simulation times. In subsequent simulation studies, the tolerance values were fixed to those used in Trial 4.

Table 3: Results using ODE15s for different tolerances

Trial	Absolute Tolerance	Relative Tolerance	Simulation Time [s]	Amp [V]
1	1 e-15	1 e-09	12.2	0.21177
2	1 e-16	1 e-10	18.2	0.24151
3	1 e-17	1 e-11	36.9	0.24965
4	1 e-18	1 e-12	62.3	0.24988
5	1 e-19	1 e-13	91.9	0.24996

Conversion Gain

In order to determine if the size of ΔR affects the conversion gain in simulation studies, the conversion gain was determined as a function the ratio g_B/g_A for three different values of ΔR. In these studies the value of R was chosen to yield a Q-factor of 100 and the value of $g_A R$ is set to 1.03. In the experimental marginal oscillator, a calibration circuit is used to vary the Q-factor of the linear RLC circuit, and hence the value of R. In the simulation studies, we assume that the calibration circuit changes the value of ΔR so that the Q-factor of the RLC circuit is either 99.0, 99.5, or 99.9 when the calibration circuit is turned on. When the calibration circuit is turned off, the Q-factor is 100. The smallest value of ΔR corresponds to the Q-factor of 99.9. The results in Figure 9 show that when the value of Q-factor approaches the nominal value of 100, the difference between the

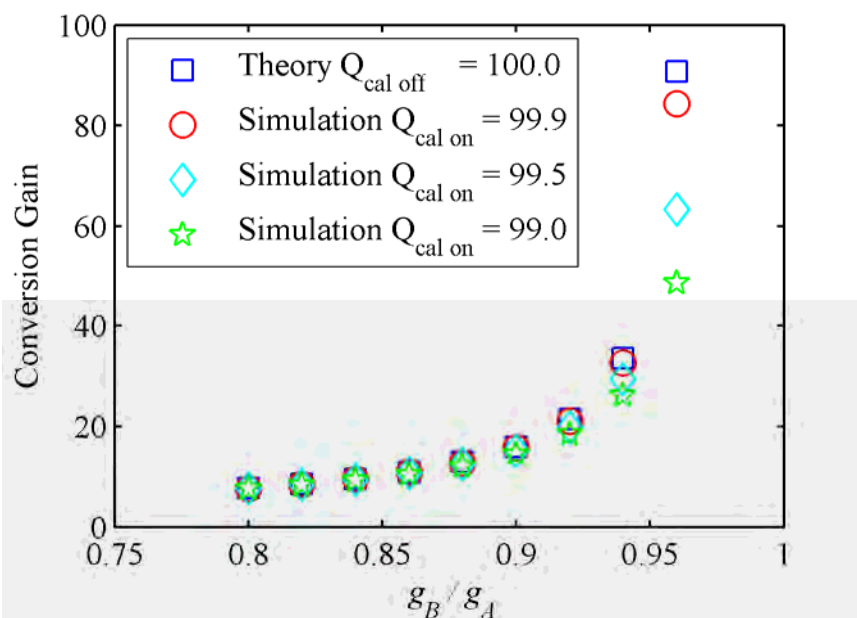conversion gain predicted by theory and that measured in simulations becomes negligible.

Figure 9: Simulated G_c versus g_B/g_A for different Q-factors

DISCUSSION

The results of this study show that the conversion gain predicted by Viswanathan's equations 11 through 13 closely match the values obtained in simulation studies. The discrepancy between the conversion gain predicted from theory and that obtained in simulation results presented by Miller likely result for two reasons. First, this earlier study employed the solver ODE113, which is not tailored for stiff systems. Second, the previous study did not account for the fact that the value of ΔR must be small in order to obtain an accurate estimate of conversion gain using Equation 17. In fact, the earlier used values of $Q_{cal\ on}$ and $Q_{cal\ off}$ of approximately 99.0 and 100, respectively. Based on the results in Figure 9, we should expect a large difference between the predicted and measured conversion gains.

CONCLUSION

The goal of this paper was to test the hypothesis that the discrepancy between the theoretical predictions and numerical simulations presented by Miller are due to the selection of the ordinary differential equation solver, and the protocol for determining the conversion gain. The results of this study support this hypothesis, and verify that the theoretical predictions of Viswanathan are correct. Additional work is needed to see if the conversion gain measured in the experimental system

will also agree with the theoretical predictions when difference in the Q-factor between calibration states is reduced to less than one percent.

ACNOWLEDGMENTS

REFERENCES

[1] D. Miller, J.L. Schiano, and T.J. Tyson, "Comparison of Predicted, Simulated, and Experimentally Measured Conversion Gain in a Robinson Marginal Oscillator," *NSF EE REU PENN STATE Annual Research Journal,* **7**, Edited by S. Bilén and J. Mitchell (2009).

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

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

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

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

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

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

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

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

[10] L.F. Shampine and M.W. Reichelt, "The MATLAB ODE Suite," *SIAM Journal on Scientific Computing,* **18** (1): 1–22 (1997).

MEASURING THE THERMOELECTRIC POWER OF NANOMATERIALS ON A MICROMACHINED WORKBENCH

Matthew Chang*, Srinivas Tadigadapa,[#] and Prasoon Joshi[+]

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

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

ABSTRACT

A micromachined workbench consisting of partially released platforms with independent built-in thermocouples and heaters has been fabricated to study the thermoelectric power (thermopower) of nanomaterials. Nanomaterials are suspended across adjacent platforms. Joule heating of one platform creates a temperature gradient across the nanomaterial which in turn generates a measurable thermopower across the nanomaterial. The nanomaterial's temperature gradient is determined from the built-in thermocouples. The workbench has been characterized for thermal time constant, inter-platform thermal crosstalk, and electrical leakage in an evacuated closed-cycle refrigerator (CCR). Thermopower measurements have been performed on gold nanowires from 100 K to room temperature. The results of those measurements, continuing problems, and an evaluation of the workbench are included in this paper.

INTRODUCTION
Background and Motivation

Thermoelectric devices are small, inexpensive devices which function on the intrinsic material property of thermopower. Despite being used for cooling and, in a few cases, converting heat to electricity, thermoelectric devices have not become widespread because of their low efficiencies [1]. The recent development of novel nanomaterials, such as nanowires, has opened up new avenues for thermopower generation. These nanomaterials have been predicted to demonstrate significantly improved efficiencies over their bulk counterparts [2].

[#] Faculty Mentor
[+] Graduate Mentor

Studying the intrinsic thermoelectric properties of nanomaterials has been a challenge. Previous studies have presented inconsistent results [3] primarily due to a lack of statistically reliable data. Data collection is complicated by the high experimental complexity of yielding testable samples and fabricating reliable test workbenches. Nanomaterials synthesized in the lab must be transferred to and positioned on the workbench. The positioning often introduces post-processing steps which threaten sample yield and condition. Some have resorted to depositing the nanomaterials first and then fabricating the workbench around the nanomaterial. This not only exposes the nanomaterial to processing conditions (i.e. temperatures and chemicals), but also limits the study to one type of nanomaterial. In workbenches that employ thermistors such as platinum resistance thermometers (PRTs), the act of measuring temperature disturbs the actual temperature itself.

A Micromachined Thermoelectric Workbench

A thermoelectric workbench has been fabricated to produce a high yield of test samples in a manner that minimizes post-processing, yet allows accurate measurements on a diverse set of nanomaterials. As shown in Figure 1, the workbench consists of a mesh of platforms with independent built-in gold heaters and gold-nickel thermocouples fabricated beneath a nickel layer. All active components are electrically isolated by a PECVD dielectric (silicon nitride in this case). To provide thermal isolation, a trench is etched between platforms, which are partially released by XeF_2 etching (inset of Figure 1). Electrical contact to the active components is established via contact pads on the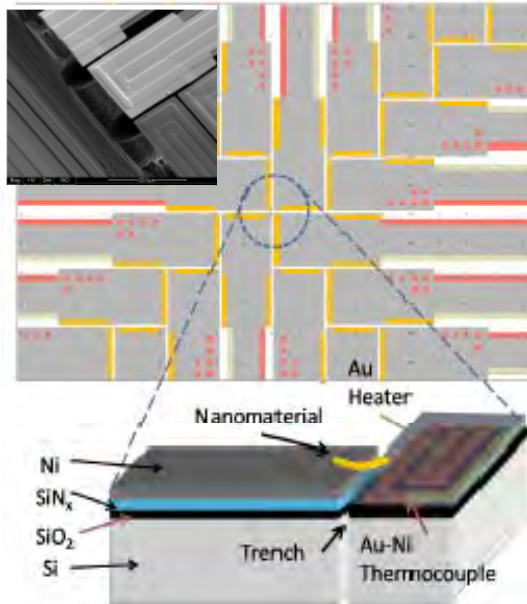 outer edges of the workbench. To perform thermopower measurements, the nanomaterial is deposited so that it is suspended between two adjacent platforms (see Figure 1). One heater is joule heated to create a temperature gradient across the nanomaterial while independent thermocouples measure the temperature gradient without disturbing the temperature. The thermovoltage is measured via the top nickel layer.

Figure 1: Diagram of the thermoelectric workbench. Nanomaterial is suspended across two platforms. The inset shows the partially released platforms. Picture taken from [4].

The workbench has been characterized to investigate the workbench thermal physics. To quantify thermocouple response, thermocouple thermovoltage was

measured as a function of power supplied to the heater. The temperature signal of the heater can be modeled by the heat equation [2]

$$\rho C_p \frac{\partial}{\partial t} T(x,t) - \kappa \nabla^2 T(x,t) = Q = \left(\frac{I}{A}\right)^2 \left(\frac{1}{\sigma}\right) \tag{1}$$

$$\text{with} \quad I = I_0 \sin(\omega t) \tag{2}$$

where ρ is mass density, κ is thermal conductivity, C_p is specific heat capacity, σ is electrical conductivity, and A is cross-sectional area (all of the heater). The supplied heat, Q, is equal to the joule heat generated by the drive current given by Equation (2). The solution to Equation (1) has components that are twice the frequency of the drive current [2]. Therefore, thermocouple response, which is directly related to the temperature signal, should be measured at 2ω.

Workbench characterization also included the thermal time constant, which is a measure of how quickly the temperature signal can respond to a change in the drive signal. By using lumped RC modeling in the thermal domain, it can be shown that the heater's temperature signal has a frequency response which behaves like a lowpass filter [5]. By measuring either the thermocouple signal or the nanomaterial's thermopower signal as a function of frequency, a Bode plot can be constructed. The thermal time constant is the inverse of the 3-dB frequency of the Bode plot. Actual time constants are dependent on device geometry and material. It was also important to characterize thermal crosstalk and electrical leakage between and within platforms. Despite efforts to thermally isolate the platforms, heat can still propagate through the workbench substrate, reducing thermal gradients between platforms. Because an electrical signal is used to drive the heater, electrical leakage from the heater to the nanomaterial or thermocouple may corrupt the actual temperature signals that are being measured.

Thermopower measurements on gold nanowires were performed on the workbench through a range of temperatures. These experiments, which have been performed before [1][3], allow a first-hand evaluation of how well the fabricated workbench accomplished its objectives.

Determination of the Seebeck Coefficient

The basic principles of thermometry can be used to calculate the Seebeck coefficient of the nanomaterial under test. The schematic diagram shown in Figure 2 shows the general test setup being used for thermopower measurements. The different stages and connections from the measurement equipment to the sample introduce local temperature differences (indicated by T_1, T_2, and T_3) which induce Seebeck emfs. Because no material is perfectly homogeneous, these intermediate stages will introduce a thermovoltage in the circuit. Thus, the test setup must be calibrated before accurate measurements can be claimed. For an analytical model, we assume that the material is perfectly homogeneous and that the temperature change across the nanomaterial (ΔT) is small enough so that the

Seebeck coefficient can be considered constant. Then, the Seebeck coefficient of the nanomaterial can be determined from the measured thermovoltages and the standard Seebeck coefficients of gold and nickel in a method similar to that in [6].
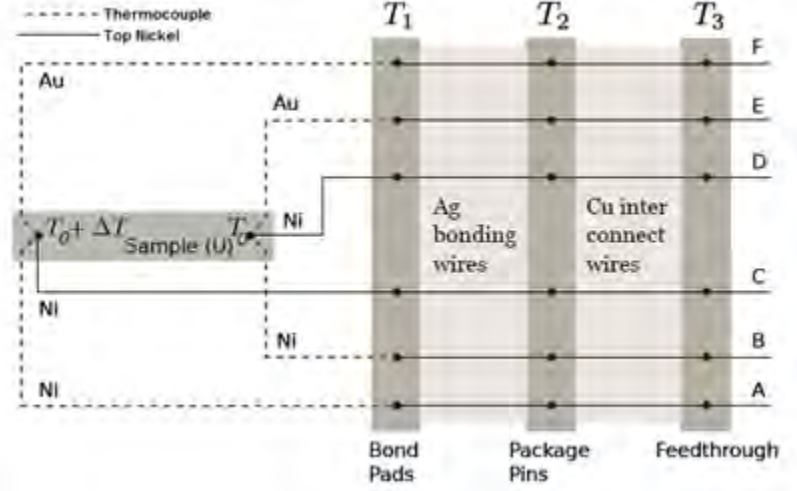

Figure 2: A schematic diagram of the general setup. The dashed line represents thermocouple connections while the solid line represents top nickel connections. T_1, T_2, and T_3 show that each stage of the test setup from workbench to measurement equipment introduce local temperature gradients.

From Figure 2, the following measured thermovoltages can be written as

$$V_{DC} = \int_{T_1}^{T_0} S_{Ni}(T)dT + \int_{T_0}^{T_0+\Delta T} S_U(T)dT + \int_{T_0+\Delta T}^{T_1} S_{Ni}(T)dT = \int_{T_0}^{T_0+\Delta T} (S_U(T) - S_{Ni}(T))dT$$

$$= (S_U(T_0) - S_{Ni}(T_0))\Delta T \tag{3}$$

$$V_{EB} = \int_{T_1}^{T_0} S_{Au}(T)dT + \int_{T_0}^{T_1} S_{Ni}(T)dT = \int_{T_1}^{T_0}(S_{Au}(T) - S_{Ni}(T))dT \tag{4}$$

$$V_{FA} = \int_{T_1}^{T_0+\Delta T} S_{Au}(T)dT + \int_{T_0+\Delta T}^{T_1} S_{Ni}(T)dT = \int_{T_1}^{T_0+\Delta T}(S_{Au}(T) - S_{Ni}(T))dT \tag{5}$$

where S_{Ni}, S_{Au}, and S_U are the Seebeck coefficients of nickel, gold, and the nanomaterial, respectively. We have ignored the Seebeck emfs generated from the temperature intervals T_3 to T_1 because they will be cancelled out due to the material's homogeneity. By manipulating these three thermovoltages, the Seebeck coefficient of the nanomaterial can be determined to be

$$S_U(T_0) = S_{Ni}(T_0) + \frac{V_{DC}}{V_{FA} - V_{EB}}(S_{Au}(T_0) - S_{Ni}(T_0)) \tag{6}$$

which is only dependent on the well-known Seebeck coefficients of nickel and gold, and the three measured thermovoltages.

EXPERIMENT DESCRIPTION

All measurements were performed in vacuum (~200 mT) in a Cryodyne closed-cycle refrigerator (CCR) with a Lakeshore temperature controller unless stated otherwise.

Workbench Characterization

To measure thermocouple response and verify workbench functionality, an 8.6–Hz sine wave was driven through the heater by an HP8165A Signal Generator at varying amplitudes. 8.6 Hz was used because it was around a peak in the gain of the thermal frequency response of the workbench (see thermal time constant results). An EG&G Model 5302 Lock-in amplifier was used to lock onto the resulting thermocouple signal, which as mentioned earlier, was at 2ω or 17.2 Hz. The HP8165A provided the reference signal for the lock-in amplifier. A Keithley 2000 multimeter measured the heater current and an Agilent 54624A Oscilloscope was used to observe the applied waveform across the heater to catch any clipping due to the HP8165A's power limit. **Error! Reference source not found.** shows a schematic of the measurement setup used for all characterization tests except DC electrical leakage.
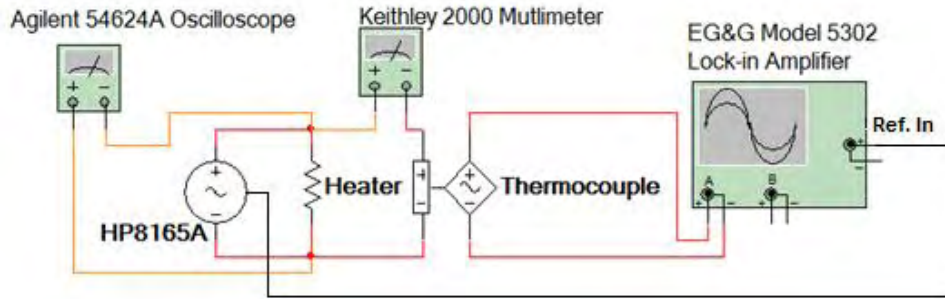

Figure 3: Schematic diagram of the measurement system used for characterization

To measure the thermal time constant, the experimental setup in Figure 3 was used. The drive signal amplitude was held constant while the frequency of the signal was swept. A Bode plot of the 2ω thermocouple signal allowed easy determination of the thermal time constant by locating the 3–dB cutoff frequency.

AC thermal crosstalk was performed with the experimental setup in Figure 3. The thermocouple signals of a cluster of platforms were measured while only one platform was heated, as shown in Figure 4. The amplitude of the drive signal from

the HP8165A is swept while the frequency was held at 8.6 Hz. The 2ω thermocouple signals indicated how much thermal crosstalk there was between platforms.

DC electrical leakage tests were performed with a Keithley 2400 Sourcemeter. A drive voltage is applied across two terminals from the same platform that are supposed to be an open-circuit (i.e. heater and thermocouple) and any leakage current through the PECVD nitride is measured. All possible pairs of heater, thermocouple, and top nickel were tested for electrical leakage before every thermopower measurement.

Figure 4: Measuring thermal crosstalk. Platform Q3Y3 is heated while the thermocouple signals of all platforms were measured.

The nomenclature of the workbench is also worth mentioning. On the workbench, platforms are organized in an x–y coordinate system with each platform sitting on an axis. Each platform is identified by which quadrant it is in and how far up the specified axis it is located. For example, platform Q3Y3 is located in the third quadrant, and is the third platform on the y-axis.
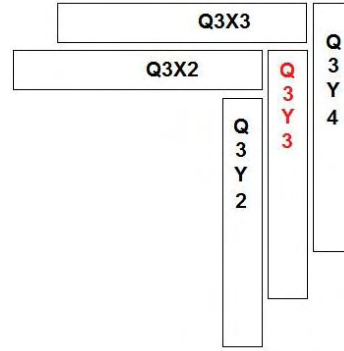
Thermopower Measurements

To perform thermopower measurements, the nanomaterial must make ohmic contact with the top nickel. Gold nanowires suspended in iso-propyl alcohol were spun onto the workbench. The alcohol was evaporated, depositing the nanowires on the workbench. Although the workbench was designed to avoid post-processing, electron beam induced deposition (EBID) of platinum was used to clamp the nanowires to the top nickel layer to ensure ohmic contact, as shown in Figure 5. DC I-V measurements on the nanowires were always performed to ensure ohmic contact prior to thermopower measurements. Occasionally it was necessary to breakdown any oxides which were introducing contact problems with ±5 V across the nanowire, with a low current compliance to prevent damage to the device.
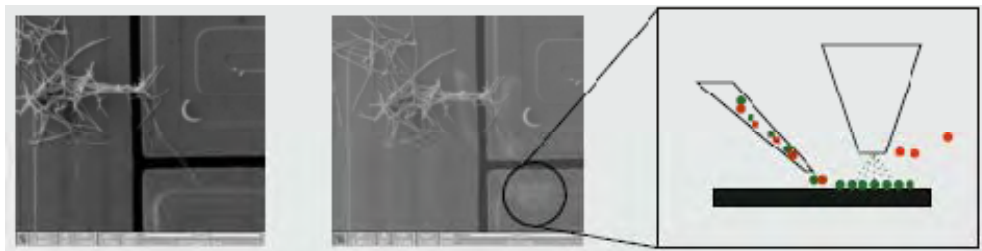

Figure 5: SEM images of deposited nanowire before (left) and after (middle) EBID of platinum to ensure ohmic contact. The right hand picture shows a diagram of the EBID process. Picture taken from [4].

Thermopower measurements were performed with a similar configuration as the characterization measurements (Figure 3) except with the thermocouple replaced by the top nickel contact pads (the nanomaterial). The nanomaterial's thermopower was measured at 2ω. The HP8165A is operated at 8.6 Hz in 50-Ω and 1-kΩ mode to achieve the widest range of heater power possible. To protect the device from electrostatic discharge (ESD), the device is held at ground with a breaking box until all the instrumentation had been configured for measurement.

EXPERIMENTAL RESULTS

The following results come from three samples: a single nanowire, a nanowire clump (called the nanoclump), and an empty workbench (no nanowire). The measurements were performed on both released and unreleased platforms.

Deposition

Nanowires were successfully deposited onto the workbench surface. Figure 6 shows two samples: a nanowire clump (Figure 6a) and a single nanowire (Figure 6b). Figure 6a also shows a single nanowire suspended across the upper right and lower right platforms, but that sample was lost to ESD shortly after being placed in the CCR. In both samples, an EBID process was used to anchor the nanowires to the nickel.
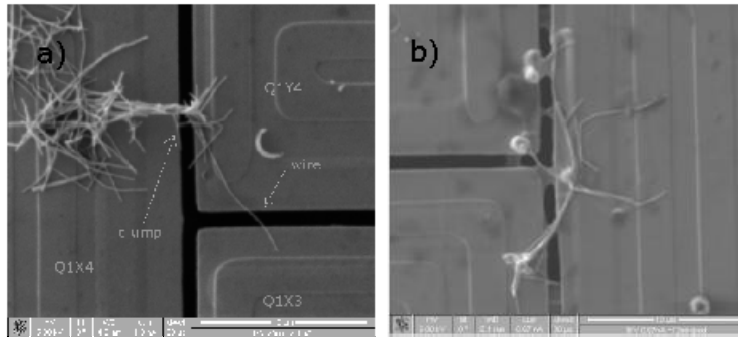

Figure 6: SEM images of nanowires that occupied test sites on different workbenches. a) A nanowire clump bridging the left and upper right platform and a single nanowire bridging the upper right and lower right platform. The single nanowire was lost to ESD. b) A single nanowire bridging three platforms. The nanowire has been anchored to the substrate by platinum clamps.

The anchor is only shown in Figure 6b. DC I-V measurements were used to verify ohmic contact. Nanowires began in the 1–10 kΩ range, but processing and other unknown events increased the resistance to as high as 100 kΩ.

Thermal Time Constant

Figure 7 shows the thermal time constant frequency response plot of three thermocouples on the nanoclump sample up to 256 Hz. The measurements were

performed in vacuum (200 mT air) and at 293 K. The nanoclump was bridging across platforms Q1Y4 and Q1X4 (see Figure 6a). Q1Y4 was the heated platform, but signals could be seen from all thermocouples. The individual platform signals were normalized from thermocouple 2ω signals of 3.16, 3.20, and 3.34 µV for platforms Q1Y4, Q1X3, and Q1X4, respectively. A line is drawn from 0.707 gain to graphically locate the 3–dB frequency. This particular sample showed 3–dB cutoff frequencies of roughly 20 Hz, or a time constant of 50 ms. Table 1 shows the thermal time constant of the single nanowire sample (Figure 6b) as a function of temperature.
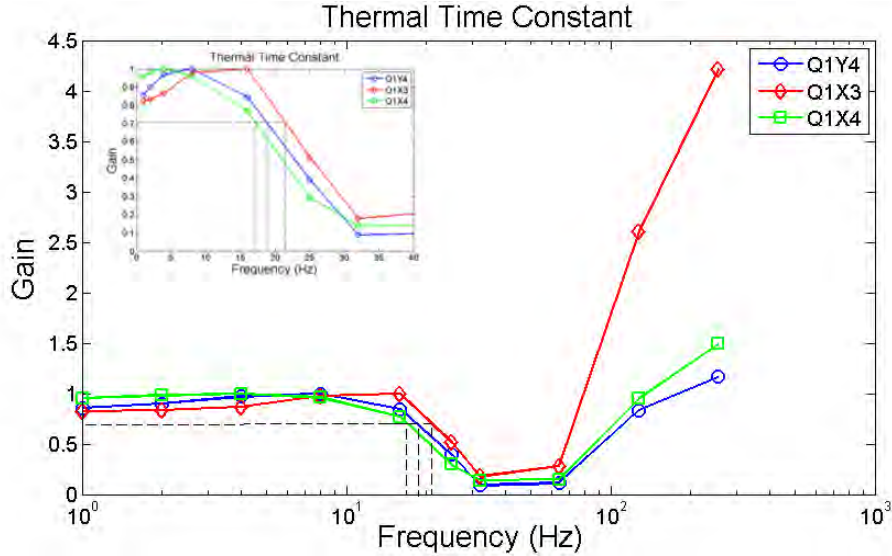

Figure 7: Thermal time constant measurement of three thermocouples in the nanoclump sample at 293 K and 200 mT (air). Gains were normalized from thermocouple 2ω signals of 3.16, 3.20, and 3.34 µV for platforms Q1Y4, Q1X3, and Q1X4, respectively. Time constants were roughly 50 seconds. The high gain at high frequencies is predicted to be a result of noise. The inset zooms in on the lowpass portion of the curve.

Table 1: Thermal time constant measurements on a single nanowire from 125 K to 293 K. The nanowire was located on platform Q4Y3.

Temperature (K)	3–dB Frequency (Hz)	Time Constant (ms)
125	21	47.62
150	21	47.62
175	19	52.63
200	20	50.00
225	19	52.63
250	18	55.56
293	19	52.63
Average	19.57	51.24
Std. Deviation	1.13	2.95

Electrical Leakage

   Electrical leakage of the PECVD nitride was measured on all test sample platforms prior to measurement. To quantify the leakage, a Keithley 4200 Semiconductor Characterization System was used to measure the precise leakage currents/voltages permeating through the PECVD nitride. Leakage was measured through the thermocouple when 10 and 15 V were applied to the heater. A radial distribution of platforms was measured from two empty workbenches (one for 10 V and another for 15 V) and the results for the 15–V leakage measurements are shown in Table 2. No significant leakage was found in the 10–V case.

Table 2: Electrical leakage from the heater to thermocouple with 15 V applied to the heater.

Platform	P_{HTR} (mW)	< I_{LEAK}>	I_{LEAK} σ	<V_{LEAK}>	V_{LEAK} σ
Q1Y1	443	-60.53 nA	1.00 pA	18.90 mV	12.06 mV
Q1Y7	237	10-100 nA	N/A	---	---
Q2Y6	226	-64.71 pA	34.08 pA	---	---
Q2Y2	518	0.35 µA	17.37 nA	Broken Heater	N/A
Q3Y3	324	-31.59 nA	6.03 nA	10.65 mV	3.55 mV
Q3Y5	254	0.97 µA	238.89 nA	1.79 V	144.00 mV
Q4Y4	261	-110.94 nA	74.50 nA	> 10.00 V	Total Breakdown
Q4Y7	229	-104.80 pA	45.82 pA	---	---

Thermopower Measurements

   Thermopower measurements performed on the nanoclump are reported here. Platform Q1Y4 was heated while thermocouple responses in platforms Q1Y4, Q1X3, and Q1X4 were measured (see Figure 6a). The nanoclump bridged platforms Q1Y4 and Q1X4, so the difference in thermocouple readings between these two platforms ($\Delta V_{TC} = V_{TC,HOT} - V_{TC,COLD}$) produced a measure of temperature difference; absolute temperature is unknown without calibration. The nanoclump was measured under several conditions: vacuum (8 mT air), 780 Torr of N_2, vacuum after three cycles of XeF_2 etching, and vacuum after five cycles of XeF_2 etching. Each etching cycle was 60 seconds at 2 Torr of XeF_2. The former two measurements were performed on an unreleased workbench, whereas the latter two measurements were performed on workbenches released by the XeF_2 etching. The different etching cycles provided different levels of release, allowing a study on the effect of releasing the workbench. Figure 8 shows the measurements performed on an unreleased workbench at 293 K in vacuum while Figure 9 repeats the measurement after 5 cycles of XeF_2 etching.

   A summary of the thermopower and thermocouple 2ω signals under the different conditions are presented in Table 3. The slope of the thermopower vs. ΔV_{TC} graph should be proportional to the Seebeck coefficient of the nanowire and top nickel combined. These slopes are presented for the different conditions in Table 4.

Table 3: Summary of the thermopower and thermocouple response measurements through all measurement conditions.

Platform	Slopes (μV/W)			
	8 mT	780 Torr N_2	3 Cycles XeF_2	5 Cycles XeF_2
Q1Y4	30.0	28.5	28.9	28.6
Q1X3	30.2	29.0	29.0	29.0
Q1X4	32.1	33.1	37.9	38.7
Thermopower	16.5	16.2	17.6	40.0

Table 4: $V_{\text{Thermovoltage}}/\Delta V_{\text{TC}}$ for the different measurement conditions. If the thermocouples perform similarly, this dimensionless number is proportional to the nanowire and top nickel's combined Seebeck coefficient.

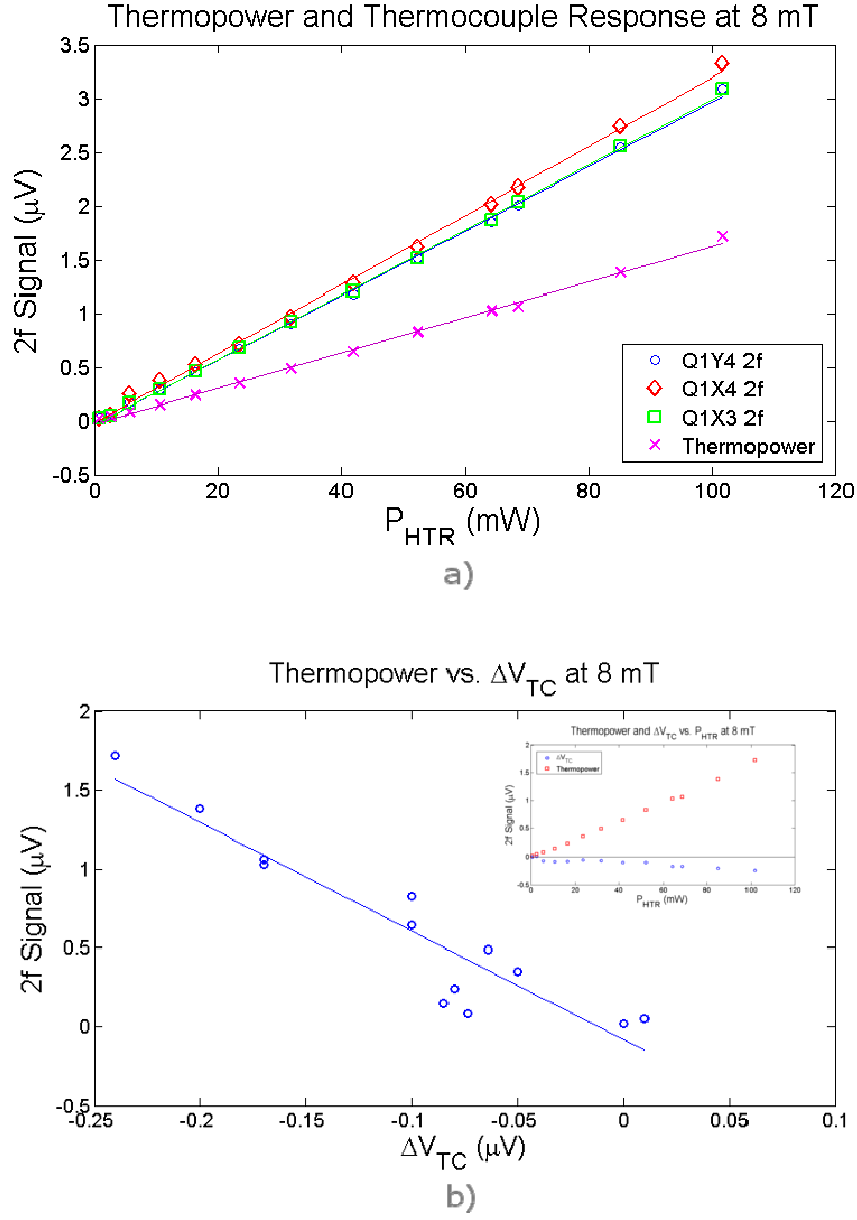| Measurement Condition	$V_{\text{Thermopower}}/\Delta V_{\text{TC}}$
8 mT (air)	7.86
780 Torr N_2	3.52
After 3 cycles of XeF_2	1.93
After 5 cycles of XeF_2	3.96

Figure 8: Thermopower and thermocouple measurements of the nanoclump on an unreleased workbench at 293 K in vacuum. (a) Thermopower and thermocouple response vs. heater power. (b) Thermopower vs. ΔV_{TC} where $\Delta V_{TC} = V_{TC,HOT} - V_{TC,COLD}$. The inset shows thermopower and ΔV_{TC} vs. heater power.
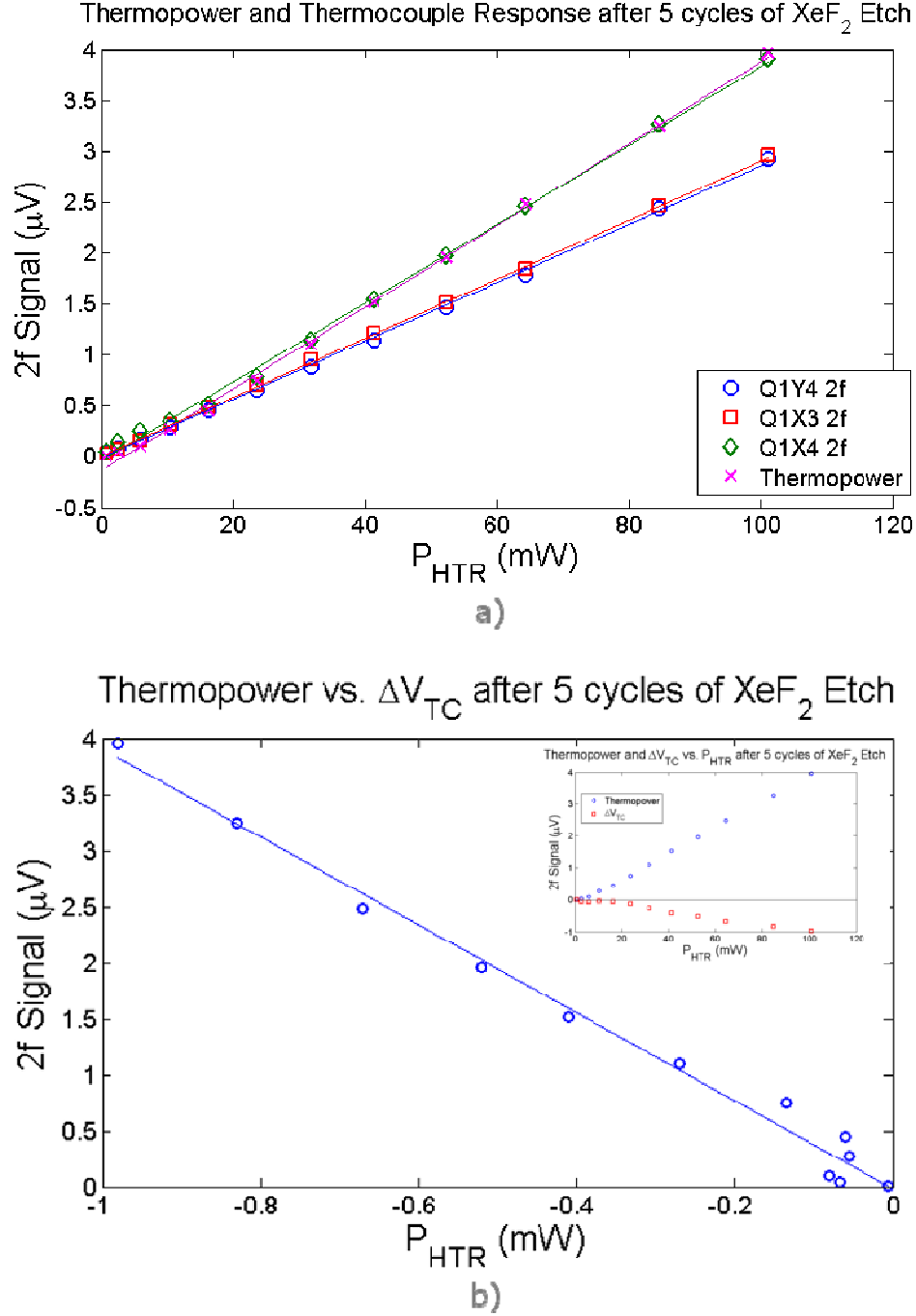
Thermopower and Thermocouple Response after 5 cycles of XeF$_2$ Etch

a)

Thermopower vs. ΔV_{TC} after 5 cycles of XeF$_2$ Etch

b)

Figure 9: The same measurements as in Figure 8 after 5 cycles of XeF$_2$ etching. (a) Thermpower and thermocouple vs. heater power. The thermopower has doubled. (b) Thermopower vs. ΔV_{TC}. There is a strong linear relationship thermopower and ΔV_{TC}. Inset shows thermopower and ΔV_{TC} vs. heater power.

DISCUSSION

Deposition

   The deposition of the nanowires was successful. However, the need to clamp the nanowire to the nickel layer with platinum anchors shows the frailty of the ohmic contact between workbench and nanowire. Additionally, the changing resistance of the nanomaterial highlights the contact issues that continue to exist despite clamping. Because measurements often span many hours, unsteady conditions are unacceptable for a measurement workbench. The formation of an oxide on the nickel layer may be behind the contact issues [2]. This hypothesis is supported by the recovery of ohmic contact after applying ±5 V across the nanowire. To make the workbench a viable option for high yield accurate measurements, the formation of the native oxide must be dealt with. The deposition process itself can be further improved. To achieve a high yield of testable samples, yet prevent the formation of clumps and aggregates, a balanced nanowire concentration in the nanowire solution is required.

Thermal Time Constant

   The lowpass behavior of the temperature signal's frequency response was seen in nearly all thermal time constant measurements. As shown in **Figure 7** the frequency response unexpectedly showed a slight attenuation at low frequencies before peaking at around 8 Hz. As a practical system, extra poles and complex frequency responses are expected. However, the slight attenuation in frequency response was not seen in similar measurements by [3]. At high frequencies, the frequency response diverged. Parasitics will induce unwanted signals at high frequencies, but these should not appear at 2ω. Another possible source of the high frequency gain is environmental noise which is coupling with the system via exposed wires. Despite the quarks in the frequency response, a clear attenuation is seen in all samples at around 20 Hz corresponding to a time constant of 50 ms.

   The lumped RC model is temperature dependent through variables such as thermal conductivity [5], but no clear temperature dependence was seen in the thermal time constant from 125 K to 293 K as shown in Table 1. Because of the thermal time constant measurement's repeatability and relative independence from temperature, the thermal time constant can be used in the future to verify the presence of a nanomaterial and/or functionality of the workbench.

Thermal Crosstalk and Electrical Leakage

   Thermal crosstalk could be seen during thermal time constant measurements as well as thermocouple measurements. Referencing Figure 7, thermal time constant measurements were seen in all thermocouples despite the fact that only one heater was heated (Q1Y4). In fact, the thermocouples showed nearly identical magnitude responses, indicating that the etched trench was not providing any thermal isolation (at this point, the platforms had not been released).

   The severity of the thermal crosstalk can also be seen in the thermocouple response measurements of Figure 8a and Figure 9a. In these measurements, only

platform Q1Y4 was heated, but the neighboring thermocouples were picking up as much of the temperature signal as the heated platform's thermocouple. After 5 cycles of XeF_2 etching to release the platforms, the difference between thermocouple signals (ΔV_{TC}) grew from about -2.1 μV/W to -10.1 μV/W indicating that the release increased the thermal isolation. However, it was not the heated platform but rather the colder platform that grew hotter as indicated by the larger negative ΔV_{TC}. This reversal will be addressed in the discussion on thermopower measurements. Releasing the platform improved thermal isolation but the temperature gradient was still significantly reduced. Simulations have shown that the current 5 μm trench allows a significant amount of heat to conduct through the workbench substrate. The next step to increase thermal isolation is to deepen the trench between platforms.

The results of the electrical leakage test showed that while 10 V across the PECVD nitride does not cause significant leakage, 15 V readily breaks down the dielectric. As a result, any measurements performed on the workbench should stay at a maximum of about 10 V to ensure no damage to the workbench. Nevertheless it was occasionally found that 10 V was enough to generate breakdown; electrical leakage should be checked for prior to any measurement.

Thermopower Measurements

A clear thermopower signal was measured from the nickel contact pads of the workbench as seen in Figure 8 and Figure 9. As expected, the thermopower was linearly related to heater power as well as ΔV_{TC}. This reaffirms that the thermopower is indeed a result of a temperature gradient across the platforms. On an unreleased workbench, the produced thermopower per unit heater power was 16.5 μV/W with a ΔV_{TC} per unit heater power of -2.1 μV/W. A negative ΔV_{TC} indicates that the heated platform is actually producing a *smaller* thermocouple signal (is cooler) than the non-heated platform. After releasing the workbench with 3 cycles of XeF_2 etching, the thermopower per unit heater power increased slightly to 17.6 μV/W while ΔV_{TC} per unit heater power increased to -9.1 μV/W. An additional 2 cycles of XeF_2 etching increased the thermopower per unit heater power significantly up to 40 μV/W while ΔV_{TC} per unit heater power only increased to -10.1 μV/W. The disproportionate increase in thermopower with only a slight increase in ΔV_{TC} as well as the negative ΔV_{TC} points to several potential problems involving the workbench and the measurement process.

The results bring into question the reliability of the built-in thermocouples. It is possible that the temperature at the thermocouple is not representative of the surface temperature that is sensed by the nanomaterial. To test this, thermopower measurements were repeated in 780 Torr of N_2 to see if convective cooling would change the thermopower while maintaining the same ΔV_{TC}. 780 Torr of N_2 did not appear to generate enough convective cooling but more work in this manner can be done to see if thermocouple temperature follows surface temperature. The lack of reproducibility of thermocouples is another potential flaw. While gold-nickel

junctions are a standard in thermometry, thin-film junctions have different properties than bulk junctions [2]. A micromachined workbench with thin-film junctions may have different levels of sensitivity. This would explain why the heated platform appeared colder than the non-heated platform. The sensitivity of different thermocouples can be checked by heating different heaters and taking advantage of thermal crosstalk to see how different thermocouples respond. Thermocouple sensitivity must be controlled in order to effectively use the analytical model to determine Seebeck coefficient.

Another potential problem with the workbench is the distributed heater. In the current workbench design, the heater is a uniform resistor, so the entire length of the heater is heated equally when current runs through it. Some heat may be isolated at the platform where trenches and releases have been performed, but this may not be true farther up the conducting lines to the contact pads where a large amount of heat may be dissipated. This heat can contribute to thermal crosstalk, but more importantly, it may heat the cold junction of the thermocouple causing the thermocouple to produce a smaller signal than expected. This would also explain why the heated platform appears cooler than it should be. To address this problem, the heater should be redesigned as a lumped heater so that the majority of the supplied power to the heater goes to the nanomaterial.

One difficulty of thermopower measurements comes from the fact that every material shows a thermopower. The current workbench and associated setup has not been calibrated to see how much of the thermopower is coming from the (non-ideal) test setup versus how much is actually coming from the nanomaterial. In the worst case, there is no temperature gradient across the nanomaterial (due to thermal dissipation and crosstalk), and the measured thermopower is purely due to the external setup.

The inconsistencies and problems in the thermopower measurements can be summed up in the seemingly random pseudo-Seebeck coefficient of the gold nanoclump and top nickel combination that was measured under the different conditions in Table 4. Before the coefficients can be considered Seebeck coefficients, the thermocouple signal must first be calibrated and converted into a temperature difference and the 2ω signal from the lock-in amplifier must be converted into an actual thermocouple signal.

CONCLUSION

A micromachined workbench was fabricated to provide high yield and reliable measurements on the thermopower of nanomaterials. The temperature gradient across the nanomaterial can be measured by independent thermocouples without affecting with temperature itself. Workbench characterization showed that the thermal time constant of the workbench was about 50 ms. The measurement is repeatable, and can be used as a verification test in the future. A significant amount of thermal crosstalk was found between platforms which may be a consequence of the distributed heater and/or insufficient thermal isolation. Electrical leakage through the PECVD nitride was not a major issue if the applied

voltages were held under 10 V. Thermopower measurements performed on a gold nanoclump showed that the workbench successfully heats and measures thermopower/thermocouple responses. General trends match what is expected from thermopower measurements. However, improvements need to be made. The heater should be redesigned to be lumped to help contain the supplied power to the nanomaterial. Increased thermal isolation can be achieved by etching a deeper trench between platforms. Thermocouples must also be made reproducible, and their measurements confirmed somehow. A method of calibrating the thermocouples must also be found.

REFERENCES

[1] DiSalvo, Francis J. "Thermoelectric Cooling and Power Generation". <u>Science</u>. Vol 285. July 30, 1999: 703-706.

[2] Duarte, Nicolas, "Thermopower Measurement of Gold Nanowire Systems Using a Micromachined Workbench". PhD Thesis, The Pennsylvania State University (2008).

[3] Shi, L. et al., "Measuring Thermal and Thermoelectric Properties of One-Dimensional Nanostructures Using a Microfabricated Device". <u>Journal of Heat Transfer.</u> Vol. 125. October 2003: 881-888.

[4] Joshi, Prasoon. et al., "A Micromachined Workbench for Measuring Thermopower of Nanomaterials". Poster presented at the Hilton Head 2010 Solid-State Sensors, Actuators, and Microsystems Workshop. June 6-12, 2010; Hilton Head, SC.

[5] Senturia, Stephen D. "Microsystem Design". Kluwer Academic Publishers, 2001.

[6] Sumanesekera, G.U., Grigorian, L., Eklund, P.C., "Low-temperature thermoelectrical power measurements using analogue subtraction". <u>Measurement Science and Technology</u>. Vol 11. March 2000: 273- 277.

COMPARATIVE ANALYSIS OF SILICON CANTILEVER AND MICROMACHINED QUARTZ GRAVIMETRIC SENSORS

Emmanuel Carpena,* Ping Kao,[+] and Srinivas Tadigadapa[#]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Universidad del Turabo
Gurabo, PR 00778-3030

ABSTRACT

Force Microscope (FM) technique and Quartz Crystal Microbalance (QCM) technique were implemented in order to detect gases. FM technique used a cantilever structure coated gold on one of the sides and functionalized by two types of alkanethiols for detecting target gas. It results the change of surface energy and gives the bend of cantilever. The signal from cantilever can be detected by laser beam deflection. Thickness shear mode quartz crystal microbalances (QCM) with micromachined technology exhibits a sensitivity of 0.14 ngcm^{-2}Hz^{-1}. QCM technique implements the same fuctionalization strategy as in the FM technique and the response is given in change in resonant frequency. Both of FM and QCM were functionalized with hexadecanethiol (HD) and 16-mercaptohexadecanoic acid (MCH) and resulted in hydrophobic and hydrophilic surfaces. For the hydrophilic surface a change in 14(mV) was detected and 2.6(mV) for the hydrophobic surface. While in the QCM technique produces a frequency change around the 1,777 Hz for a hydrophilic surface and gives a surface coverage of 241 ngcm^{-2}. In the hydrophobic surface frequency change around 797 Hz and gives surface coverage of 108 ngcm^{-2}. In both techniques the response in hydrophilic experiments is larger than the response in the hydrophobic one. . A relationship between pure mass loading and position sensitive detector voltage change is reported here. This project contributes in the development and understanding the relationship between FM and high frequency resonator (56 MHz) for same adsorbed molecules. It will help to improve the

[#] Faculty Mentor
[+] Graduate Mentor

design of detecting very small amount specific gases that can guide directly into the application of sensing system.

INTRODUCTION

The requirement of cheap, simple and reliable bioanalytical technology always remains a challenge. Such technologies normally consist of four main components: 1) transducer (the change of physical signals into more convenient ones); 2) biochemical recognition, 3) signal conditioning, 4) signal interpretation algorithms [2]. Many different techniques are designed and used for the element of transducer and have their own merits and demerits. Advancements in Microelectomehcanical System (MEMS) have facilitated the journey of cheap simple and reliable sensors development. They normally consist of high sensitivity mass detector, typically, a cantilever structure or a piezoelectric resonator. Recent advances in micro and nanoscale fabrication techniques have show that it is possible to measure mass into the 10^{-21} gram (zeptograms) range[6,7], allowing the possibility of single molecule detection. The general idea is that physical, chemical or biological interaction with MEMS structures can affect the characteristic of the transducer in such way that the resulting change can be measure by any means: electronically, optical and others. These lead to cantilever for high sensitivity to applied forces or high resonance frequencies for faster response and QCM for high sensitivity to pure mass loading or viscoelastic damping. The comparative analysis of cantilever and micromachined quartz gravimetric sensor to designed surface reaction gives the fundamental understanding of transducer component.

Force Microscope (FM) Technique

Force Microscope is one of the techniques implemented in this research. Since the arrival of the Atomic Force Microscope probes in 1986 [5], more in depth studies on cantilevers structures have driven them from a simple probe into more sophisticated applications [2].

In this work cantilevers are implemented as chemical sensor. As commonly defined, a cantilever is a long structure that is sported in only one of the sides, while the other side is suspended (see Figure 1-A). Studies say that cantilevers tend to bend when they are exposed to mechanical surface stress if they are prepared with different opposite faces [3]. Knowing that if they are opposite, if one of them expand the other compress and vice versa, is the principle why the cantilevers in these sensors are functionalized by a bio-affinity layer in only one of the sides with a substance that exhibits a reaction to the targeted molecules. The force of chemical interaction generates the stress needed on the surface of the structure to bend the cantilever. [2] This happened because surfaces usually tend to change as a result of surface energy induced by adsorptive processes. This bending action can be read by optical reflection of a laser that is targeted in the tip of the structure.

The surface stress in terms of deflection can be calculated by a formula stated in Raiteri et al. 2001, and can be related to mass in the following formula:

$$\Delta z \cong \left(\frac{3(1-v)}{E^2} \right) \Delta(\Delta\sigma_1 - \Delta\sigma_2) \tag{1}$$

where t is thickness of the beam and E and v are mechanical parameters of its materials: Yung modulus and Poisson ration, respectively.[8]
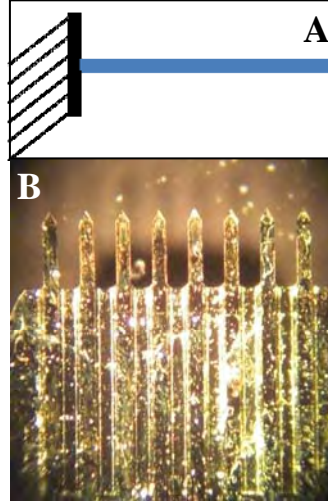

Figure 1: **A)** A schematic image of the cantilever structure. It can be observed that is clamped in one of the ends and suspended in the other one. **B)** Actual cantilever array structure used in the experiments of this research. Photo was taken through a microscope.
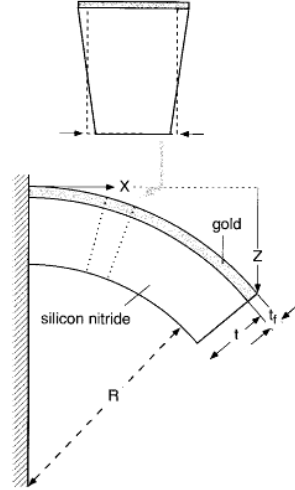
Figure 2: A cantilever picture with the top of the structure coated with thin layer of gold. It can be observe the curvature of bending. This Image is from a paper by Hans-Jürgen Butt in 1995, named "A Sensitive Method to Measure Changes in the Surface Stress of Solids." This is the same structure used in this research.

Quartz-Crystal Microbalance (QCM)

A shear mode quartz crystal microbalance (QCM) typically consists of a slab of thin, single-crystal piezoelectric quartz with two large lateral dimensions metal electrodes and has been widely used since their simple linear relationship to change in the resonance frequency to added mass was elucidated by Sauerbrey in 1959 modified in eqn. (2).[1] If attached material is rigid, sufficiently thin compared to the quartz crystal, the relationship between resonance shift and the mass loading (Δm) is given by eqn. (2):

$$\Delta f = \left(\frac{f_0}{A \rho_q t_q} \right) \Delta m = - \left(\frac{2 f_0^2}{A \sqrt{\rho_q \mu_q}} \right) \Delta m = -S_f \frac{\Delta m}{A} \tag{2}$$

where f_0 is the fundamental resonant frequency with no attached mass, μ_q is the shear modulus of the quartz (2.947×10^{10} N m^{-2} for AT-cut quartz), ρ_q is the density of quartz (2.648×10^3 kg.m^{-3}), t_q is the thickness of the crystal, and A is the area of the electrode on the quartz crystal. A very important aspect of eqn. (2) is that the sensitivity of QCM will be easily enhanced if increasing resonant frequency with decreasing the thickness of the crystal. [5]

The most commonly available commercial QCM consists of a ~330–150 μm thick and ~25 mm diameter disk made from an AT-cut quartz crystal assembled in a single reaction cell, has a shear resonance frequency in the 5–10 MHz range, and exhibits a sensitivity of ~17 – 4 ngcm^{-2}Hz^{-1}. Although the few resonators with even higher frequencies are commercially available the fragility of large area thin

resonators has precluded their extensive use in comparison to the 5–10 MHz resonators. Motivated by the improvement of sensitivity and resolution to mass loading and the size of array via miniaturization of quartz resonator, we have designed and fabricated 29 µm thick quartz resonator arrays operating at fundamental resonance frequency of ~56 MHz.
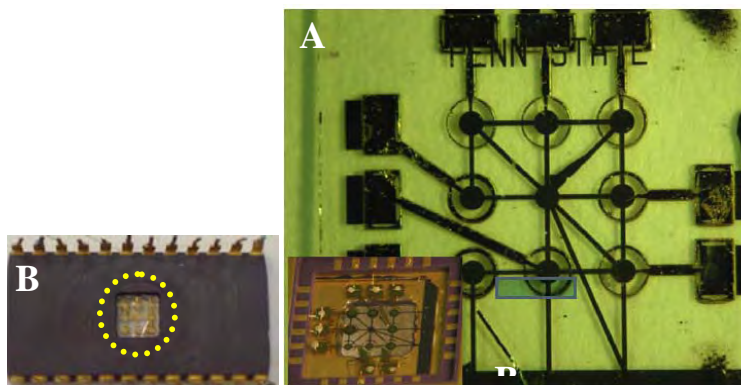

Figure 3: A) Photo of the fabricated quartz resonators array that was used in this research. B) 24 pins device package showing the hole that expose the resonator array showed in A.

In this research, we present the same functionalize surfaces for two different transducers, discussed in the FM technique and the QCM technique. The comparative analysis of adsorbed molecules between silicon cantilever and micromachined quartz gravimetric sensors will improve the design of detecting very small amount specific gasses that can guide directly into the application of sensing system.

EXPERIMENTAL SECTIONS
Materials
Hexadecanethiol (HD) (95%), 16-Mercaptohexadecanoic acid (MCH) (99%) and were obtained from Sigma Aldrich and Ethanol (Pharmco, A.C.S./U.S.P. grade) was used as received. The 6% acetic acid (AA) (J.T. Baker, A.C.S reagent) was diluted by 18 MΩ-cm DI water (Millipore Milli-Q system; Barnstead international). The 1 m M alkanethiol solutions were prepared by dissolving the HD and MCH in ethanol without additional purification.

Surface Functionalization
The cantilevers covered on one side with a 5nm Cr/30nm au layer were inserted in alkanethiol solutions for 18 hours at room temperature. This was followed by rinsing with pure ethanol and drying with pure nitrogen.

Top gold electrodes of the micromachined QCM array were cleaned by 3 cycles of exposure to UV ozone, each 15 minutes long, followed by thorough rinsing with ethanol and immersion in ethanol for 1 hour before functionalization.

Self-assembled monolayer of HD and MCH was obtained by immersion of a freshly cleaned QCM top electrodes into a alkanethiol solution for over 18 hours at room temperature.

Force Microscope (FM)

For the FM technique an optical setup is been used. The laser is guided and aimed at the tip of the cantilever structure (see Figure 4-A). The reflected light is then guided to a position sensitive detector (PSD) that is based on a quadrant photodiode of four cells. Each cell is coupled to the input of a separate transimpedance amplifier were the output voltage of each cell are proportional to the amount of light in each of the cell. The PS is calibrated mounted on a moving XYZ micro- positioner in order to calibrate the sensor before the experiment start. The cantilever is clamped inside a sealed chamber that contains an inlet and outlet plugs (see Figure 4-B and 4-C) through were the gas were flowed. The chamber contains an antireflective window in order to target the tip of the cantilever and read the reflected beam while maintaining the environment sealed inside the chamber. In order to vibrate the cantilever and have a better resonant peak, the chamber has a clamped piezo-stack actuator driven with a signal generator. The FM experiment can be conducted with or without the actuator. Frequency spectrum, deflection and room temperature can be measured through a data Acquisition connected to a user interface/Virtual Instrument (VI) created on LabView by Sangmin Jeon from the The Nanoscale Science & Devices Group. This software also, saves all the data in text files for latter data analysis.

The experimental procedure of the FM technique consist of after setting up the whole system into an optimal state, heat is applied to a diluted solution of acetic acid (6%) and using nitrogen as the carrier gas to transport the suspended gas or vapor from the elementary flask to the gas chamber. A cooling coil is hooked up to the gas line in order to maintain the temperature of vapor gas entering the chamber. The vapor gas was controlled by temperature of acetic acid solution and pressure.

Quartz Crystal Microbalance (QCM)

For the QCM technique a packaged ceramic chip that contain several quartz crystals resonators is hooked up to an Agilent 4395A Impedance Analyzer that will sweep frequency while measuring impedance. This device is mounted also on a special sealed gas chamber that contains an inlet and outlet ports (see Figure 5) that fixed the path of the gas. The whole setup is driven by a VI that manages to control a National Instruments RF multiplexer module in order to record the data with multiples resonators at the same time. The VI records all the data points of each sweep.
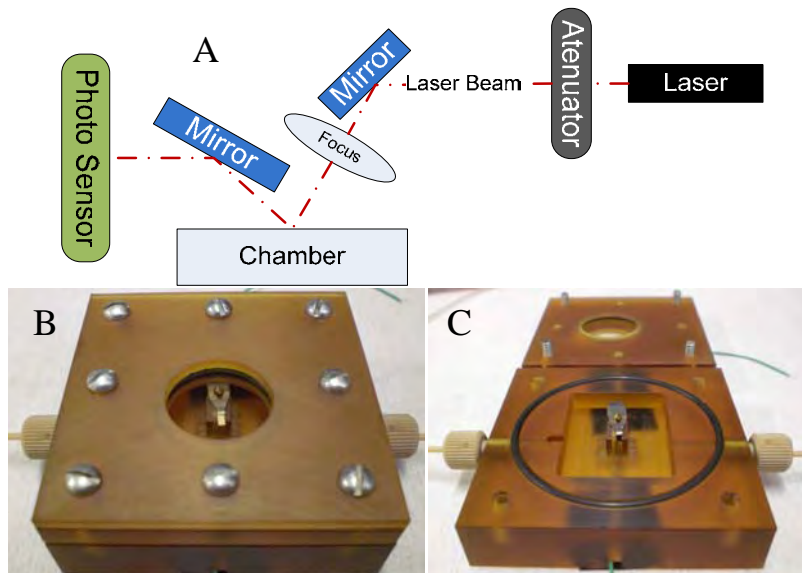
Figure 4: A) In house optical setup of the FM Technique. Basically a laser is shot and passed through an attenuator that attenuates the amount of light in order to have fewer reflections in the photo sensor and therefore have less noise. Then is aimed and focus at the cantilever that sits on the sealed chamber. Another mirror captures and redirects it straight at to the photo sensor. **B)** Photograph of the in house designed gas chamber showing the window and the port outlets were the gas tubes are connected. **C)** Photograph of the chamber completely open. It can be observe the clamping mechanism that holds the cantilever in place.
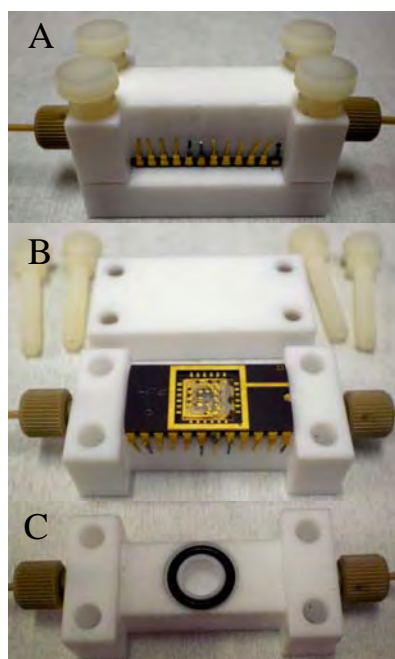
Figure 5: A) QCM gas chamber. It consist of sealed chamber that hold the device inside with four screws**. B)** Picture of the chamber open and showing the device. **C)** Picture of the O ring that sealed the chamber with the device

The experimental procedures for the QCM experiment are as follow: before starting the experiment the machine has to be calibrated with the fixture compensation of open, short and load. The gas flow experimental procedure is the same as the FM technique. A nonlinear regression was used to fit the phase angle data to a Lorentzian function and obtained the resonance frequency.

EXPERIMENTAL RESULTS
FM Technique

For this technique two experiments are presented. One with a hydrophilic surface and the other with an hydrophobic surface. In both of the experiment resonance frequency, deflection and room temperature were measured. The actuator was on in both of the experiments. AF cantilever experiment obvious room temperature responses were observed in the PSD voltage of the cantilever, with and without the actuator as can be seen in Figure 6. The experimental procedure consists of first flowing pure nitrogen through the chamber, around 30m to 1 hour, in order to measure a base line and use this initial data to compensate the temperature response. 30 minutes before the next stage start pre-heating the acetic acid up to a temperature of 50ºC. Then flow the acetic acid for

about 1hour to 1.5 hours. Then purge nitrogen again for round 30 minutes to 1 hour.

<u>Hydrophilic Surface Experiment</u>

For FM the techniques with a hydrophilic functionalized cantilever surface the experiment run as it should. Only a laser beam interruption by accident produced a huge peak in the middle of flowing acetic acid. At the beginning of the experiment it room temperature response were observed (See Figure 6). Resonant Frequency data do not show distinguishable response to any variable (See Figure 7).
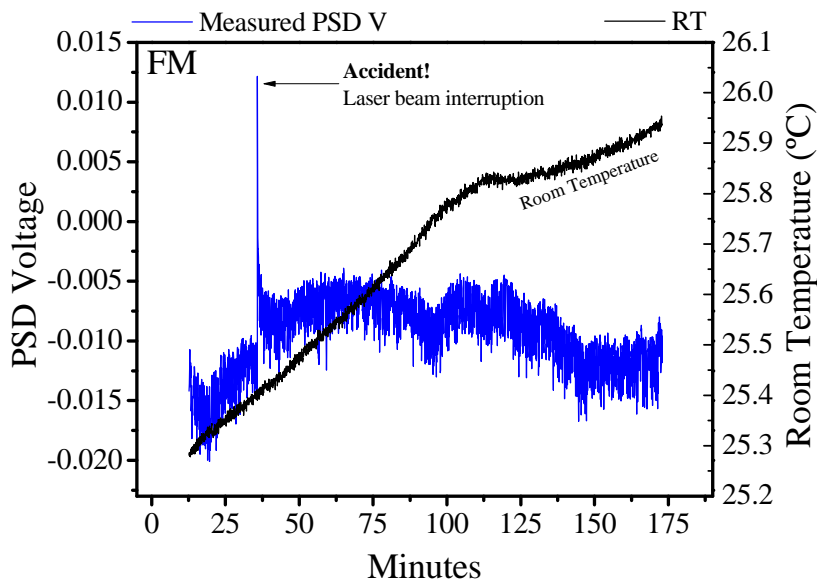

Figure 6: The real time measurement of PSD voltage and room temperature for hydrophilic surface
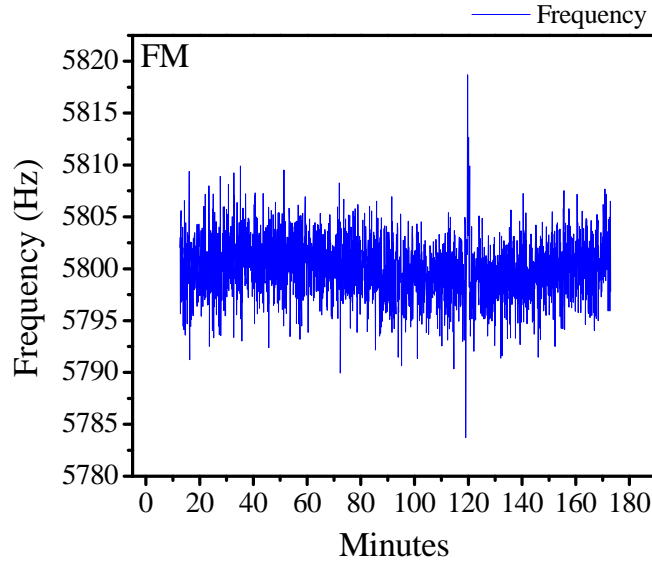
Figure 7: The real time measurement of resonance frequency of cantilever for hydrophilic surface

After compensation of temperature influence, the PSD voltage of sensor plotted as a function of time (Figure 8) shows the decreasing of acetic acid vapor, shown in Figure 8.
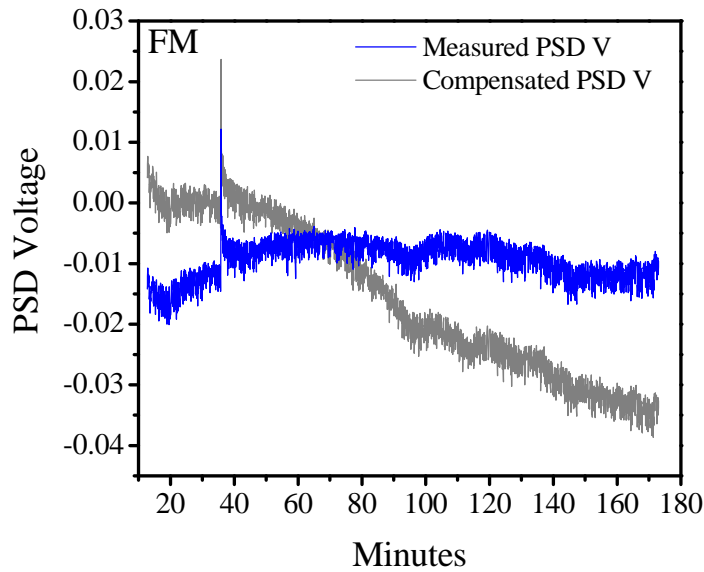
Figure 8 Comparison of the Measured PSD voltage and the Compensated PSD voltage plotted as a function of time.

In Figure 9 the arrows indicate the exact moment were the flow was change. The results from this experiment are summarized in table 1.
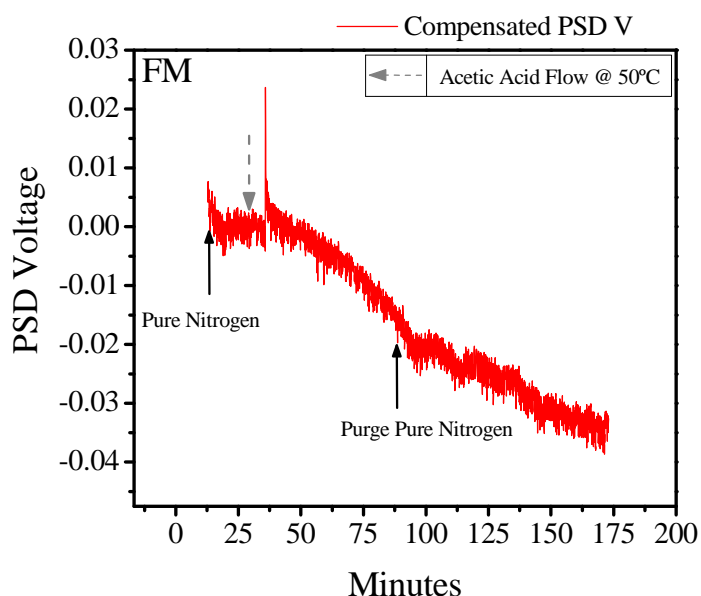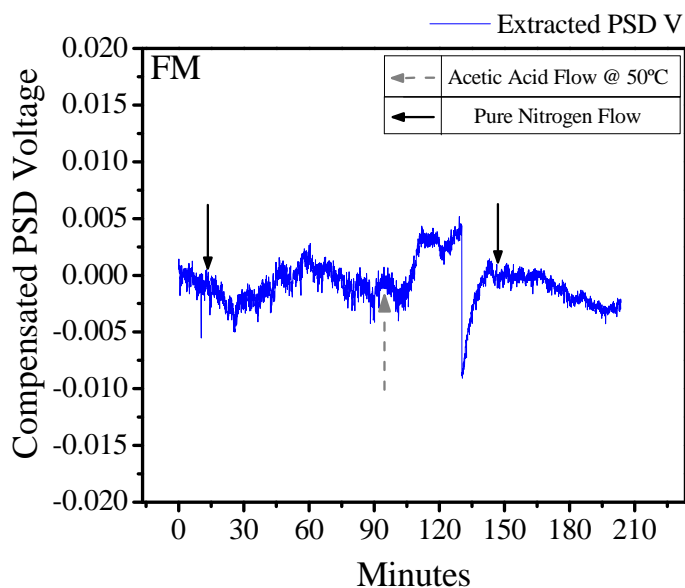
Figure 9: The PSD voltage of sensors plotted as a function of time for exposure to acetic acid. The cantilever was functionalized by 16-mercaptohexadecanoic acid.

Table 1: Final Table FM Hydrophilic Experiment

Air	Acetic Acid	Nitrogen
PSD Voltage (mV)	**ΔPSD Voltage (mV)**	**ΔPSD Voltage (mV)**
0.19 ± .092	-14.00 ±0.63	-34.32± 0.90

Hydrophobic Surface Experiment

   For the AF technique with a hydrophobic functionalized cantilever structure at the beginning was having the similar responses as the hydrophilic, even in the positive temperature coefficient. Almost at the finish of the second stage of the experiment atmospheric condition changes very fast changing the room temperature very drastically and therefore changing the deflection response, in other words bending the cantilever. Two compensation graphs were make in order to fix the data and compensate the huge temperature change. The resulting graph can be observed in Figure 10. The strange huge inverse peak appears because two different compensation fittings were made in order to obtain this graph. The results from this experiment are summarized in Table 2.

Figure 10: The PSD voltage of sensors plotted as a function of time for exposure to acetic acid. The cantilever was functionalized by hexadecanethiol.

Table 2: Final Table FM Hydrophobic Experiment

Air	Acetic Acid	Nitrogen Purge
PSD Voltage (mV)	**ΔPSD Voltage (mV)**	**ΔPSD Voltage (mV)**
0.81 ± 0.30	2.65 ± 0.21	-2.12 ± 0.13

QCM Technique

The same experiment as the FM technique was reproduce using QCM sensors functionalized with two different surfaces, hydrophilic and hydrophobic. In both of the experiments, the distinguishable responses were obtained.

Hydrophilic Surface Experiment

QCM technique with a hydrophilic surface produces frequency change around the 1777 Hz. This can be observed in Figure 11. All the data for this experiment is summarized in table 3

Table 3: Final Table QCM Hydrophilic Surface

Air	Acetic Acid	Nitrogen Purge
F_0	**ΔF(Hz)**	**ΔF(Hz)**
56353526 ± 2.7	-1777 ± 14.8	-247 ± 1.0

Hydrophobic Surface Experiment

The same exact experiment was conducted using a hydrophobic surface. A frequency change of around 797 Hz were obtained (See Figure 11). All the data for this experiment is summarized in table 4.

Table 4: Final Table QCM Hydrophobic Surface

Air	Acetic Acid	Nitrogen Purge
F_0 (Hz)	ΔF(Hz)	ΔF(Hz)
56371282 ± 4.8	-796 ± 1.3	- 247± 4.3
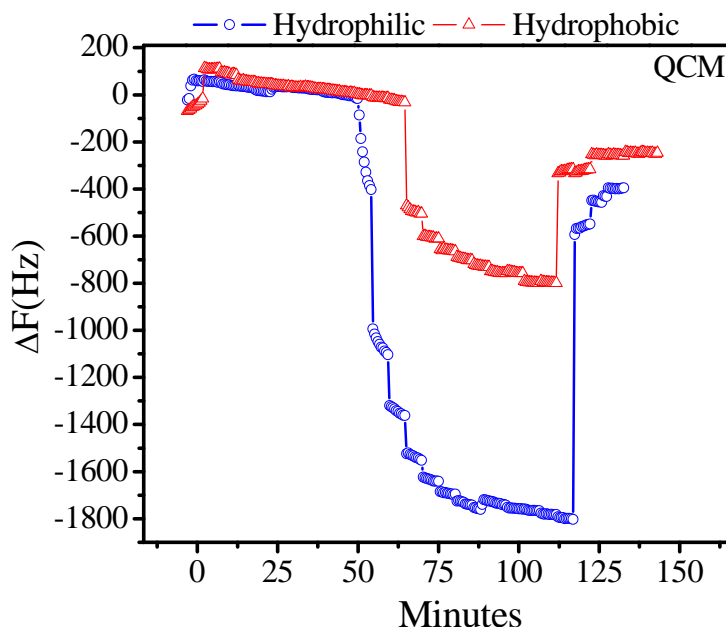
Figure 11: The frequency shift of QCM functionalized with HD and MCH plotted as a function of time for exposure acetic acid

DISCUSSION

Two fundamental experiments were conducted on both techniques, hydrophilic functionalized surfaces and hydrophobic functionalized surfaces. The experiments were performed with or without the actuator. It was believe that with the actuator, what is called a dynamic response which means that the best response was going to be in Resonant Frequency changing. And in the other hand without the actuator is what is called a static response that means the deflection is going to be the best response. Our results show that with the actuator on, a better response in deflection was obtained. Frequency graphs are very noisy and it barely changes (See Figure 7).

QCM technique is a very straight forward technique. The way the experiment is configured does not have notable room temperature problems on the raw data. QCM sensor produce a very large frequency change which can be defined a very

sensitive method. Sensitivity calculation was performed based on the eq. 1. In comparison with a previous work by Okahata et al, our numbers are similar to them and summarized in table 5.

Table 5: Comparison Surface coverage of acetic acid on hydrophobic surface

	Our Values	Okahata Values
F(MHz)	56.36	9 (7th overtone)
sensitivity(ngHz^{-1}cm^{-2})	0.14	0.78
ΔF (Hz)	1777	250
mass loading (ng/cm^2)	241	195

In both techniques hydrophilic functionalization have a bigger response that the hydrophobic one. This result can be seen in Table 6 and the relationship is observed in Figure 12. The other alkanethiol functionalized surface can use applied for obtaining more data points. It will help to predict more precise relationship between surface loading and surface stress. In principle, the surface stress can be calculated according to cantilever structure. It will be finished after the calibration experiment of surface stress.

Table 6: The summary of PSD voltage and frequency shift for hydrophilic and hydrophobic surface

	Hydrophilic Response	**Hydrophobic Response**
FM	-14.00(ΔPSD mV)	2.65 (ΔPSD mV)
QCM	-1777 (ΔHz)	-797 (ΔHz)
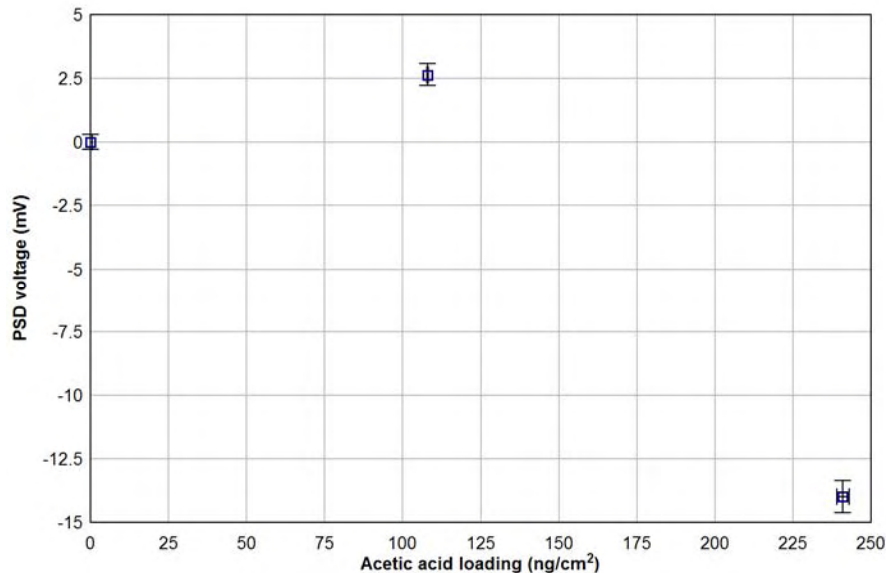
Figure 12: The surface loading versus PSD voltage

CONCLUSION

A lot of experiments were successfully performed within two months. The preliminary results of FM and QCM shows a good linear relationship between surface mass loading and PSD voltage change which can be further covered to surface stress. However, the further experiments are required for concluding a solid relationship. More experiment of more alkanethiol functionalized surface are proposed in the future. The theoretical model of discussing pure surface mass loading and surface stress need to be further analyzed based on these results. These results show a good start of bridging force microscopy and QCM in the fundamental understanding and will help to improve the design of transducer for sensing applications.

ACKNOWLEDGMENTS

REFERENCES

[1] A. Janshoffm, H.J. Galla, C. Steinem, *Pizoelectric Mass-Sensing Devices as Biosensors – An Alternative to Optical Biosensor?,* Angewandte Chemie, Volume 39, Issue 32, Weinheim Fed. Rep. of Germany: WILEY-VCH Verlag GmbH, 2000, Pages: 4004-403.

[2] N. V. Lavrik, M. J. Sepaniak, P. G. Datskos, *Cantilever transducers as a platform for chemical and biological sensors,* 75, Rev. Sci. Instrum., 2004, 2229

[3] H. J. Butt, *A Sensitive Method to Measure Changes in the Surface Stress of Solids*, Volume 180, Issue 1, Journal of Colloid and Interface Science, 1996, Pages 251-260.

[4] P. Kao, S. Tadigadapa, *Micromachined quartz resonator based infrared detector array,* Volume 149, Issue 2, Sensors and Actuators A: Physical, 2009, Pages 189-192.

[5] G.Binnig, F. Quate, and Ch. Gerber, *Atomic Force Miscroscope,* Volume 56, Issue 9, Physical Reveiew Letter, 1986, Page 930-935

[6] B. Ilic, H. G. Craighead, S. Krylov, W. Senaratne, C. Ober, and P. Neuzil, "Attogram detection using nanoelectromechanical oscillators," Journal of Applied Physics, vol. 95, pp. 3694-3703, 2004.

[7] K. L. Ekinci, Y. T. Yang, and M. L. Roukes, "Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems," Journal of Applied Physics, vol. 95, pp. 2682-2689, 2004.

[8] R. Raiteri, M. Grattarola, H.J. Butt, P. Skladal, "Micromechanical cantilever-based biosensors", Journal of Sensors and Actuators, vol. B 79, pp.115-1126, 2001.

NONLINEAR OPTICAL LIMITING IN FIBER ARRAY WITH L34

Seonwoo Lee,* Iam-Choon Khoo,[#] Zhiwen Liu,[#] Junbin Huang,[+] Chuan Yang[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
Cornell University
Ithaca, NY 14853

ABSTRACT

In this project we have tested a fiber array sample. These kinds of samples have been made before, but this has gold nanoparticles suspended in the nonlinear liquid within the sample. The nanoparticles should enhance the nonlinear effects. We found that at low intensity, the transmission through the sample is 84.41%, and at higher intensity, it is 6.55%. Also, to test this sample, we did type I second harmonic generation, and achieved an efficiency of 30% in the process.

INTRODUCTION

Many optical sensors today, whether it be our eyes or our man made devices, are very sensitive to the input intensity. If the light intensity is too high, then the sensor could be damaged or destroyed. A device that can let low intensity light through but attenuate high intensity light is desired and can be achieved with nonlinear optics.[1] This project is to investigate whether a new device developed at Penn State University can achieve this function. We hope that this kind of device can be used to protect optical sensors from intense lasers, and yet allow viewing of scenery-low light level.

Currently, there are protective devices that shield sensors from femtoseconds-picoseconds-nanosecond pulsed lasers, and but ones that work against nanoseconds and longer [microseconds] pulsed lasers are still being sought after.[1,2]

This project combines nonlinear optics properties of liquids with total internal reflection to achieve nonlinear optical limiting.

[#] Faculty Mentor
[+] Graduate Mentor
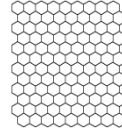
The sample is set up as follows.

Figure 1: Overall Capillary Array Structure

The fiber array is set up similar to a honeycomb. There are many hollow fibers that are packed very tightly together. The diameter of a hollow fiber is about 10 microns. The sample is then filled with a nonlinear liquid.
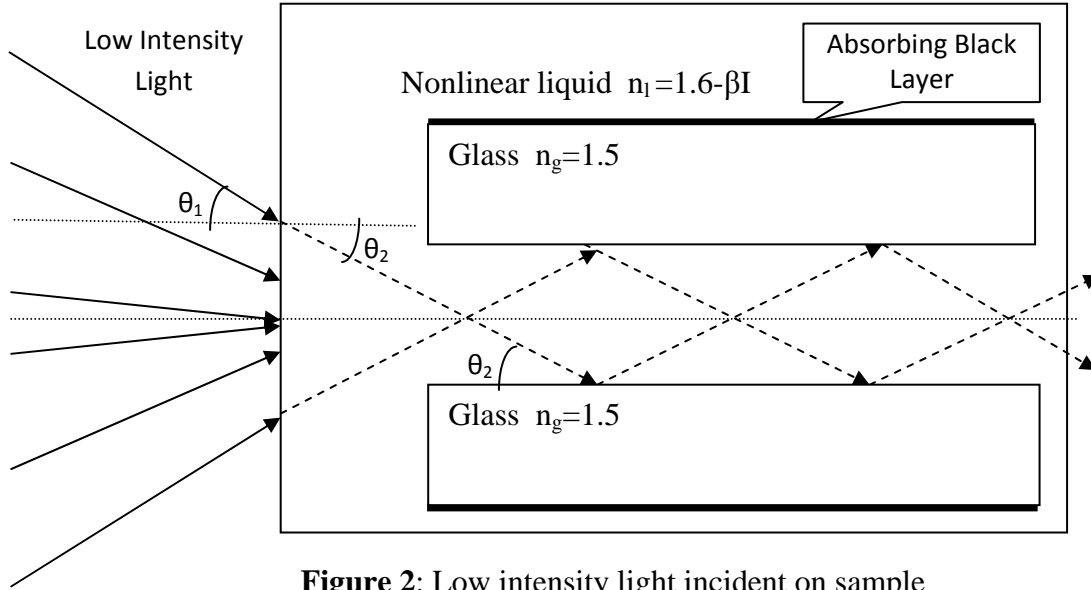

Figure 2: Low intensity light incident on sample

The sample consists of a fiber array structure, similar to that in optical fibers, filled with L34, a nonlinear liquid. Light incident on this sample should be selectively attenuated.

For a simple theoretical model, this sample is taking advantage of the phenomenon known as total internal reflection. If the refractive index of the core of a fiber structure is larger than its surroundings, and the incident light's angle θ is small enough, total internal reflection occurs—all of the incident light is kept inside the fiber structure. The derivation for this angle is as follows, given by Snell's law.

$$\sin(\theta_{cr}) = n_{liquid} \, \sin(\theta_2)$$ (1)

$$n_{liquid} \, \cos(\theta_2) = n_{glass}$$ (2)

$$\theta_2 = \arccos(\frac{n_{glass}}{n_{Liquid}})$$ (3)

$$\theta_{cr} = \arcsin(\sqrt{n_{liquid}^2 - n_{glass}^2}$$ (4)

Only the rays from the outer envelope of the low intensity light have been shown to pass through the fiber structure in Figure 2. However, all the rays that are incident with an angle smaller than the critical angle actually propagate through the sample. This sample serves as an angular filter, blocking light that is incident at too steep an angle.

The refractive index of this nonlinear liquid depends on the incident light intensity. In reality, all materials have this property. However, for most materials the beta value is so small that the nonlinear properties can be ignored. In this experiment, the beta value is relatively large, so we expect to see nonlinear effects.

As the incident light intensity increases, the refractive index of the nonlinear liquid n_l will decrease, which decreases the angle required to cause internal reflection, causing more of the light to be absorbed by the sample.
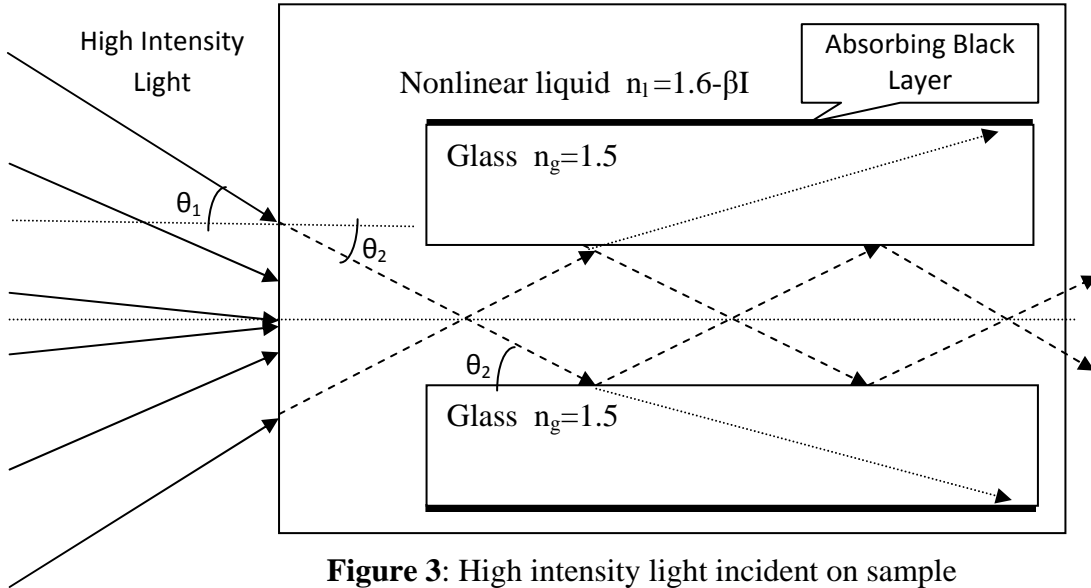

Figure 3: High intensity light incident on sample

In this situation, the high intensity light causes a smaller angle required for total internal reflection. The rays from the outer envelope are not totally reflected, escape the core, and get absorbed by the black layer. The lines with the

longer dashes show the inner part of the light undergoes total internal reflection and appears on the other side of the sample. All the light inside this smaller envelope reaches the other side without diverging.

The higher the intensity of the light, less light propagates to the other side of the sample without being absorbed. The increase in intensity serves to decrease the critical angle required for unattenuated propagation through the sample. If this sample is placed in front of an optical sensor, it should serve to protect them from high intensity light but still allow them to detect low intensity light.

This sample also has gold nanoparticles suspended in the L34 liquid, which enhances the nonlinear properties of the material.[3,4] There have been previous experiments with this fiber array structure before, but not with gold nanoparticles suspended in the liquid.[3]
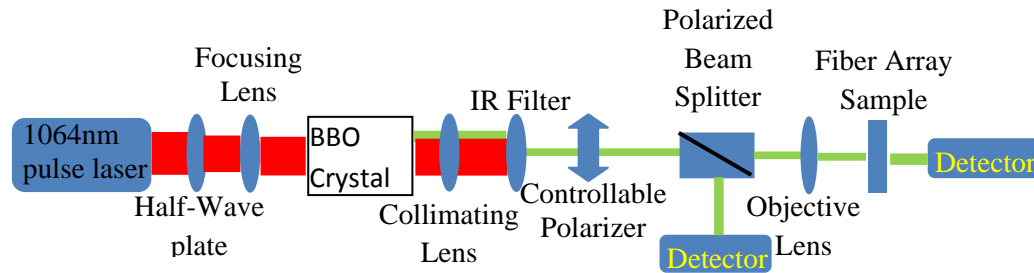
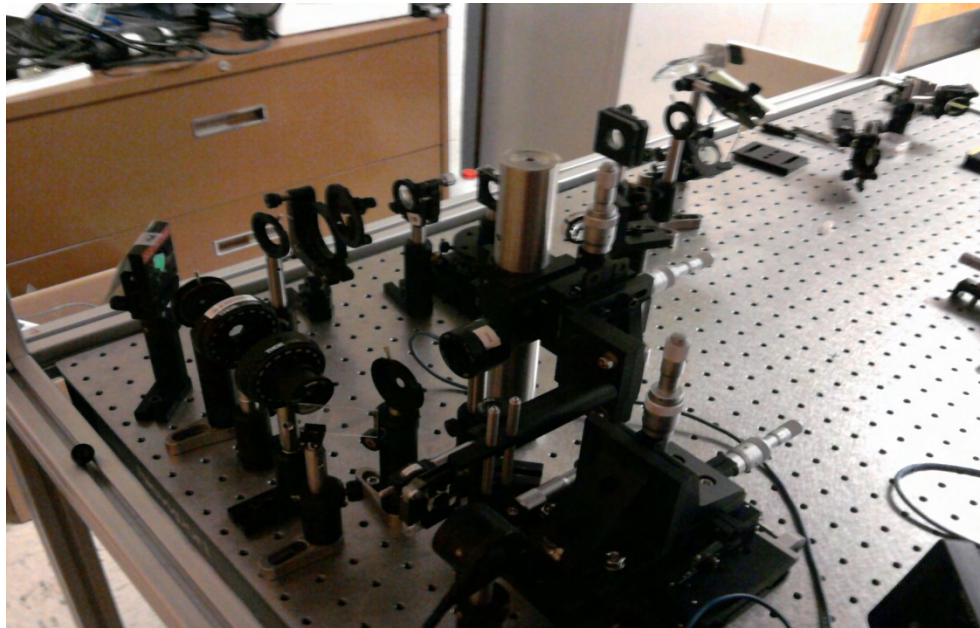## EXPERIMENT DESCRIPTION

Figure 4: Experiment Setup

Figure 5: Experiment Setup

NONLINEAR OPTICAL LIMITING IN FIBER ARRAY WITH L34
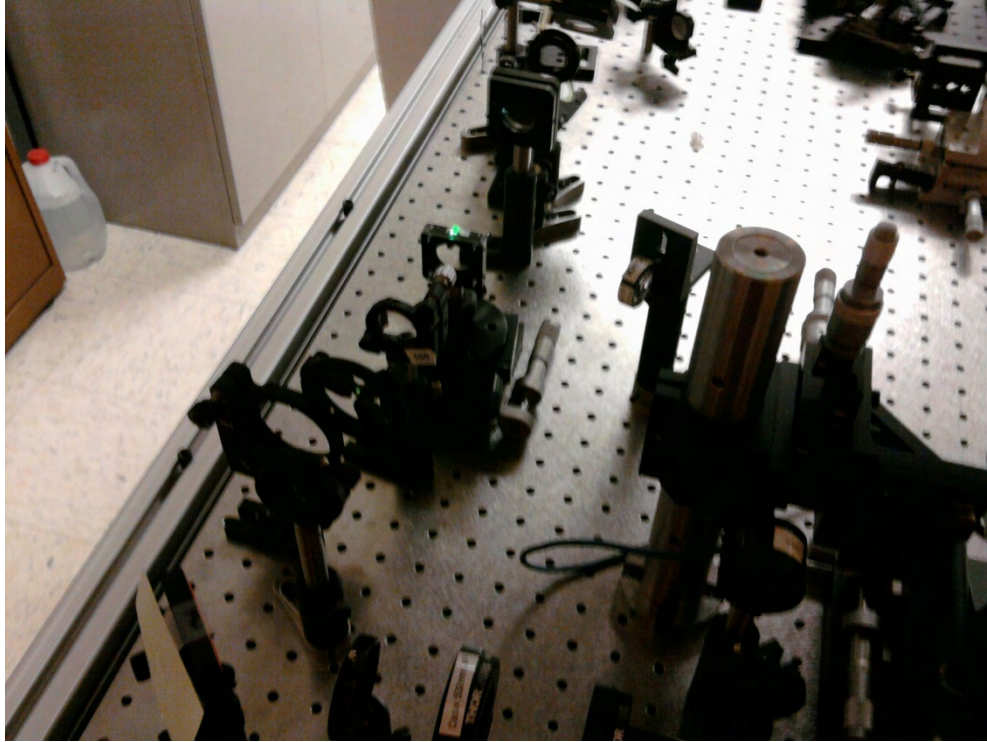
Figure 6: Experiment Setup

We used a 1064 nm nanosecond pulse laser and a BBO crystal. The BBO crystal is used to generate a 532 nm pulse laser by doubling the frequency of a 1064 nm pulse laser. The pulse laser produces a much higher peak intensity than a continuous wave laser, which helps make the nonlinear properties of the sample more apparent. Light at 532 nm (green) is required because we are interested in its behavior in the 500 nm to 600 nm range. We used type I second harmonic generation to produce green light. However, second harmonic generation has limited efficiency, so we filtered out the remaining infrared light so we could test the sample with pure green light. We can see the green light in the crystal in Figure 6.

The controllable half-wave plate and polarized beam splitter control the power incident on the fiber array. The half-wave plate linearly regulates the polarization of the light along some axis that is not parallel to either the x or y axes. The polarized beam splitter completely transmits light parallel to the x axis and completely reflects light parallel to the y axis. By adjusting the polarization of the input beam, the power of the light that is transmitted by the beam splitter is controlled. We first collected data without any sample for calibration – because the power dumped before the sample is directly proportional to the power incident on the final detector, this data can be used to determine how much power is incident into the sample.

The objective lens is used to focus the light as tightly as possible to produce the intensity that is required for nonlinear effect to become apparent.

The input power is varied, and the input and output power is compared to see if the liquid in the fiber array is in fact nonlinear and if this device would be feasible in protecting optical devices from high intensity light and yet allow them to detect low intensity light.

We also ran the experiment with just the nonlinear liquid.

EXPERIMENTAL RESULTS

The experiment necessitated a pulsed laser. We used second harmonic generation to take an infrared pulsed laser and use it as a green one. This process of second harmonic generation relies on phase matching, and is not perfect. We were able to achieve 30% efficiency – there was 50 mW of infrared light incident on the BBO crystal, and 15 mW of green light observed after the infrared filter. This process requires careful alignment of all the components of the setup. Also, the polarization of the light must be aligned with the polarization axis of the BBO crystal, so rotating the light's polarization axis was critical in obtaining maximum power.

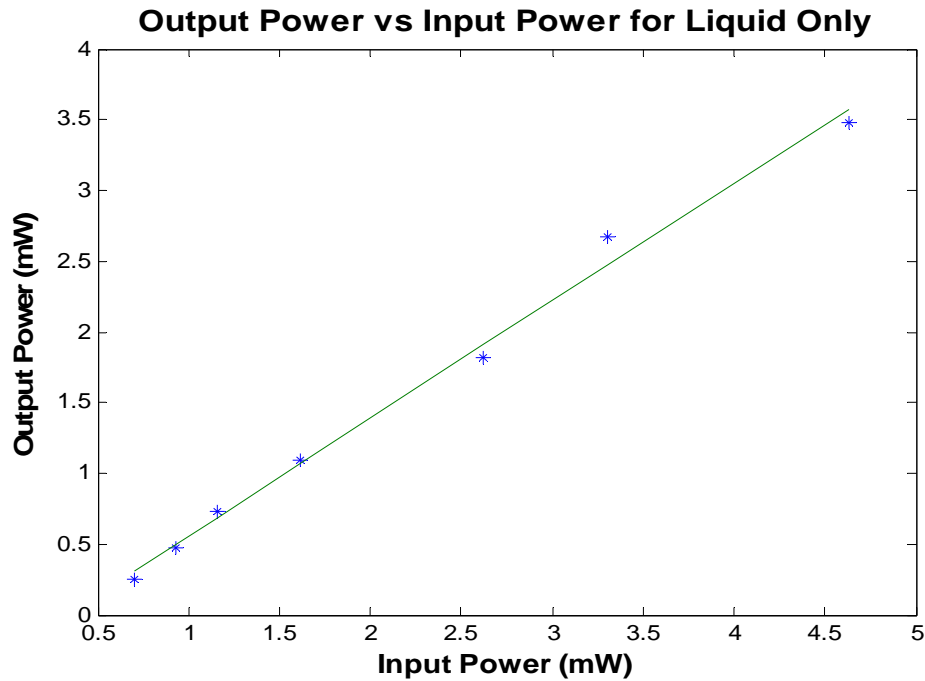We were able to focus the light into mostly one fiber.

Output Power vs Input Power for Liquid Only

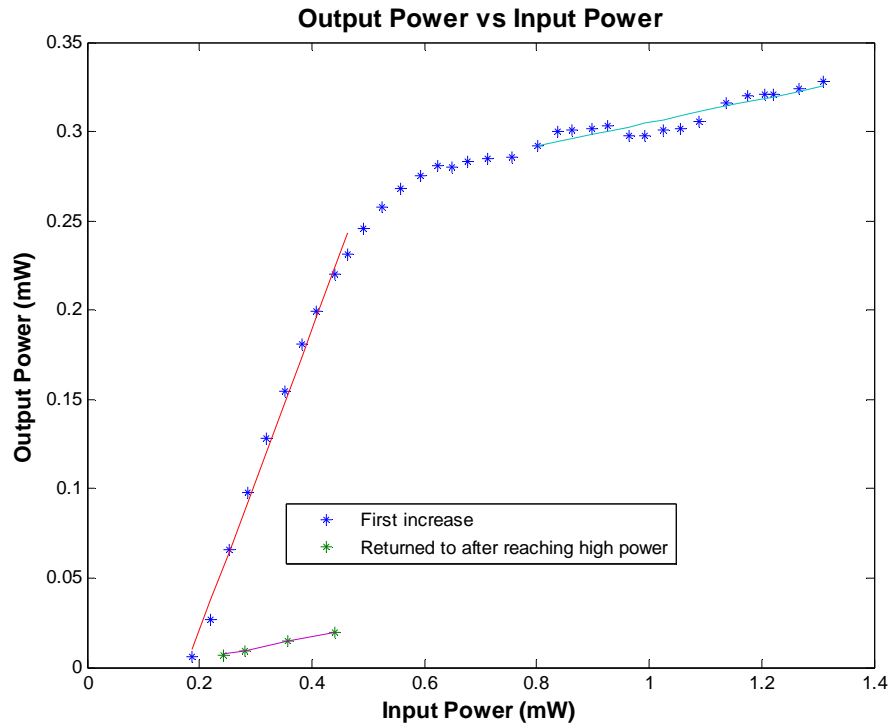**Figure 7:** Output power vs. input power for liquid only

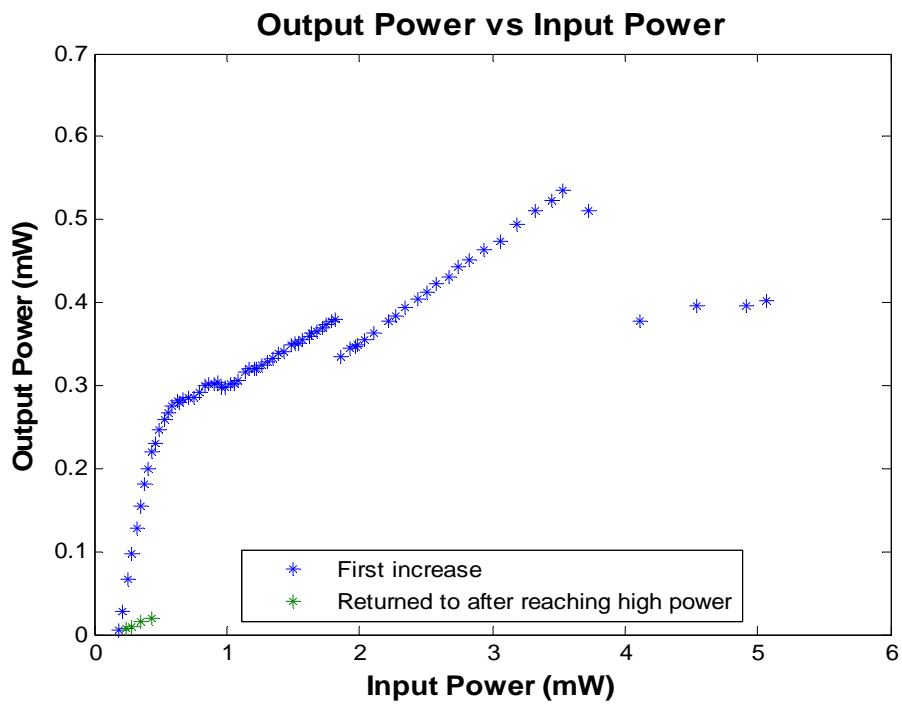**Figure 8:** Graph zoomed in on low input power region

Figure 9: Full data set

DISCUSSION

We obtained excellent efficiency in the type I second harmonic generation. This was essential for producing enough power to observe nonlinear effects in the sample.

We can see from Figure 7 that while the liquid only sample does absorb some of the input power, it does not exhibit nonlinear limiting effects. Without a fiber array structure to serve as an angular filter, the liquid does not accomplish our goal.

From Figure 8 we can see that nonlinear limiting effects do occur in the fiber array sample—there is a much higher increase in the output power from low incident power than from high incident power. The nonlinear effects start becoming apparent around 0.5 mW.

We also noticed some unexpected behavior of the sample. We measured the output at low incident power. Afterwards, we measured the output at high incident power, and when we measured the output at low incident power again, we found that the output was much lower than before, as seen in Figure 7.

We believe this is because the laser heats up the L34 liquid and causes it to boil, leaving bubbles. These bubbles will scatter the light, causing the number of rays of light that undergo total internal reflection to decrease dramatically. Once the incident power is lowered again, the light is limited by about an order of magnitude more than before. Hence, this process of increasing the incident power is not immediately reversible. We believe the fibers should be filled with liquid if they are given enough time.

The transmission at low intensities is 84.41%, and at high intensities it is 6.55%. After reaching high intensity and returning to low intensity, the transmission is 6.44%, which is very similar to that of high intensity. We believe that once bubble formation occurs, it dominates the behavior of the sample.

CONCLUSION

In this experiment we have found that the fiber array sample does in fact exhibit nonlinear limiting abilities. This could be used for protecting optical sensors. The user should be aware that the liquid in a fiber could vaporize at high intensity, causing limiting at low incident power. This may not become apparent in practice because the light would usually be incident into many fibers, not just one as we have done. However, if this does present itself as a problem, the simple solution would be to move the sample to a new location—the light would be incident into a new cluster of fibers, which would not have bubbles inside.

REFERENCES

1. I. C. Khoo, "Nonlinear Organic Liquid Cored Fiber Array for All- Optical Switching and Sensor Protection against Short Pulsed Lasers," IEEE J. Selected Topics in Quantum Electronics JSTQE 14, no. 3, pp 946 – 951 (2008)
2. Iam Choon Khoo,* Andres Diaz, and Jianwu Ding. "Nonlinear-absorbing fiber array for large-dynamic-range optical limiting application against intense short laser pulses". Journal of the Optical Society of America. June 2004. 1234-1240.
3. I. C. Khoo, A. Diaz, S. Kubo, J. Liou, Mike Stinger, T. Mallouk and J. H. Park, "Nano-dispersed Organic Liquid and Liquid Crystals for All-Time-Scales Optical Switching and Tunable Negative- and Zero- Index Materials," Molecular Crystals Liquid Crystals 485:1, pp. 934-944 (2008).
4. I.C. Khoo, S. Webster, S. Kubo, W. J. Youngblood, J. Liou, A. Diaz, T. E. Mallouk, P. Lin, D. Peceli, L. A. Padilha, D. J. Hagan, and E. W. Van Stryland, "Synthesis and characterization of the multi-photon absorption and excited-state properties of 4-propyl 4'-butyl diphenyl acetylene," J. Mater. Chem. 19, 7525-7531 (2009)

ELECTRICAL CHARACTERIZATION OF CONDUCTING MATERIALS AT MICROWAVE FREQUENCIES

Cameron Patterson,* Michael Lanagan,[#] and Steve Perini[+]

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

*Undergraduate Student of
Department of Electrical & Computer Engineering
University of Alabama
Tuscaloosa, AL 35487

ABSTRACT

This research involves the electrical characterization of thin film conducting materials using a network analyzer (HP 8510). Gold microstrips and coplanar waveguides were fabricated using a simple sputter coater. These transmission lines were connected to the network analyzer in order to transmit microwave frequency signals and monitor the lines' performance. Microwave loss of microstrips was determined from the scattering parameters S_{11} and S_{21}, which are related to the reflected and transmitted power, respectively. The thickness dependent microwave loss of gold microstrip lines was characterized for films between 25 and 200 nm. The loss range spanned an order of magnitude, decreasing with increased conductor thickness. Gold transmission lines were used as a test case in preparation for characterizing graphene, an extremely conductive material which is relatively new to researchers.

INTRODUCTION

Graphene

Graphene is a brand new type of electrical conductor. Ideally, it is a single-atom thick (often called 2-dimensional) layer of carbon, but it can be conductive with up to ten layers as well.[1] Not only is graphene believed to be far more conductive than any metal we have yet encountered, but it can also be engineered to have energy gaps similar to those of semiconductors.[1] These characteristics open up a host of possibilities for extremely conductive surfaces and lines,

---

[#] Faculty Mentor
[+] Staff Mentor

transistor logic applications, transparent electronics, and continuing miniaturization.

The ultimate goal of this research is to electrically characterize graphene on the macroscale level to improve the production process. Table 1 below compares published DC conductivity values.

Table 1: Electrical Conductivities of Conductors at Room Temperature[6][7]

Material	Conductivity ($x10^7$ Sm^{-1})
Graphene	10
Silver	6.18
Copper	5.84
Gold	4.43

Microstrip

AC conduction loss of microwave transmission lines can be directly correlated with the DC conductivity of the constituent metal. The primary device used in this project was the microstrip, which is a type of a transmission line that consists of a dielectric substrate with a thin, narrow metal film mounted on one side and a thin metal ground plane covering the entire opposite side. This line transmits electric signals in the microwave frequency range, and when monitored properly can be used to characterize the metal being used. A microstrip has a characteristic impedance Z_0 that will affect its interaction with a circuit. For the purposes of this experiment, a Z_0 value of 50 Ω was necessary to achieve

Figure 1: A microstrip consists of a dielectric substrate has a thin conducting strip of a particular width on one side, and a thin conducting "ground plane" that covers the entire bottom side. (Image from Reference 4.)

maximum power transfer across the microstrip. This impedance could only be achieved if the microstrip was designed with the correct width, which is dependent on the dielectric constant and height of the substrate.[2]
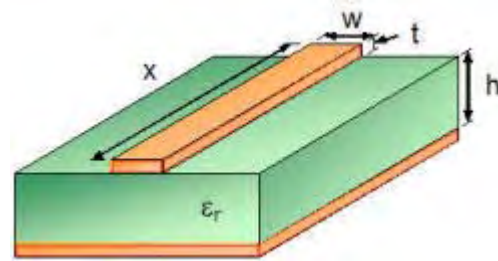
Coplanar Waveguide

Another type of transmission line that can be used to characterize a conductor is a coplanar waveguide (CPW). This fixture is very similar to a microstrip in that it can transmit microwave frequency electrical signals. Instead of having a single ground plane covering the entire bottom side of the substrate, a coplanar waveguide has two ground planes on the same side of the substrate as the center conducting strip.
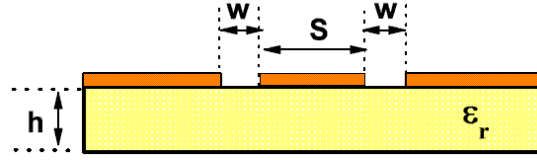
Figure 2: A coplanar waveguide consists of a dielectric substrate with three conducting planes: One center conductor of a particular width and two equally sized ground planes, one on each side of the center conductor. (Image from Reference 5.)

Scattering Parameters

Scattering parameters (or S-parameters) describe the behavior of a network when stimulated by electrical signals. Our research involved a two port network, and the most important S-parameters were the magnitude of the coefficients S_{11} and S_{21}. The coefficient S_{mn} indicates the magnitude of the signal sent from port n, as seen at port m. In other words, we were concerned with the magnitude of the signal that was reflected by the device under test (DUT) and the magnitude that was transmitted. This is the equation that allowed us to determine the loss:

$$S_{21}^2 + S_{11}^2 + \alpha_f = 1$$

S_{21} and S_{11} are in the linear scale and α_f is the linear loss fraction across the transmission line.

Skin Effect

The skin effect is a phenomenon that causes only a narrow depth of a conducting line to be used in the actual transmission of electric signals. This means that the current density is higher closer to the surface. The skin depth δ is the point at which the current drops to $1/e$ (36.8%) of its original value.[2] This depth can be determined by the equation:

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \quad [3]$$

In this equation, f is the frequency of the signal being transmitted, μ is the absolute permeability of the conductor, and σ is the conductivity of the conductor. The skin depth of gold at 10 GHz is approximately 790 nm. All of the sputtered gold films in this study had thicknesses less than the skin depth.

EXPERIMENT DESCRIPTION

This research had two main focuses. On the one hand, we were attempting to derive an accurate film thickness dependency for gold microstrips and CPWs. Very thin films of gold were used (on the order of 10^{-7} m) in the transmission lines being tested. The loss across a transmission line can be represented as a

function of the frequency of the signal, the conductivity of the metal, and the thickness of the metal. By testing gold transmission lines of different thicknesses at the same frequency, we sought to determine an accurate relationship between the thickness of the film and the loss per unit length of the transmission line. This would be useful in determining an optimal thickness that allows for a tolerable loss and low material cost.

In addition to finding thickness dependency for gold transmission lines, we sought to determine the best procedure to characterize conducting materials. Both microstrips and CPWs were tested in an attempt to decide a simple yet accurate means of characterizing conductors. This was very important for our plans to characterize graphene. A time-efficient and dependable method is always important in research.

The first part of the experiment required fabrication of the transmission lines to be tested. A sputter coater machine was used to create gold microstrips and CPWs. This machine is designed to sputter atoms of a desired material onto a target to create thin metal films. Since very precise dimensions were required for the transmission lines in order to ensure the correct characteristic impedance, a mask was designed and laser-cut to be placed on top of the substrate for the sputtering process.

Upon completion of the fabrication process, we used a four point probe to measure the sheet resistances of the gold films, which are listed in Table 2 along with their corresponding resistivities and conductivities. These sheet resistance values allow for the calculation of a film's conductivity, and subsequently a microstrip's expected loss. The program TXLine, a multi-purpose transmission line calculator, can be used to design microstrips and CPWs and project an estimated loss from the film's DC conductivity. The expected loss calculations are listed in Table 3.

Table 2: Sheet Resistances, Resistivities, and Conductivities of Different Microstrips

Film Thickness (nm)	Sheet Resistance ($\Omega/\square$)	Resistivity ($\Omega \cdot m$)	Conductivity (S/m)
25	4.80	$1.20 \cdot 10^{-7}$	$8.33 \cdot 10^{6}$
50	3.15	$1.58 \cdot 10^{-7}$	$6.35 \cdot 10^{6}$
100	0.99	$9.9 \cdot 10^{-8}$	$1.01 \cdot 10^{7}$
200	0.48	$9.6 \cdot 10^{-8}$	$1.04 \cdot 10^{7}$

Generally, the sheet resistance is expected to decrease with thickness, as the material would be able to support a greater number of charge carriers. Conductivity is a material parameter and should be independent of the conductor thickness. For the results shown in Table 2 the conductivity decreases with decreasing film thickness, indicating that the material is changing for the thinnest layers. All conductivities measured were lower than the published value shown in Table 1.

Figure 3: The microstrip is fastened into an adapter that allows the network analyzer to transmit electric signals across it.

When characterizing a conductor used on a CPW, one must use a very precise machine known as a probe station. This machine is designed to interact with CPWs with center conductor widths on the order of 1 mm. It uses two separate probes with three prongs each; one probe is positioned at each side of the waveguide to make contact with all three conductive plates on both ends. Additionally, a microscope is in place to observe the small DUT and ensure that a proper connection is being made. Before testing, an extensive and precise physical calibration of the probe station must be performed to secure a successful connection.

The most important machine used throughout the research process was the network analyzer, which can produce signals over a specified frequency sweep and provide the magnitude and phase of the scattering parameters. These parameters were very important to the experiment because they were used to determine the amount of loss over a particular transmission line or device. When testing a CPW, the network analyzer was connected through coaxial cables to the probe station in order to transmit microwave frequency signals. Prior to testing, the analyzer had to be calibrated in order to differentiate between loss along the coaxial cables and loss across the DUT. The loss per distance of the gold DUT was then compared with calculated loss values for gold, as well as loss values for transmission lines of different thicknesses in an effort to find the correct thickness dependence.

The process for testing a microstrip is very similar to that used for CPWs. The only difference is that a much simpler adapter is needed to connect the DUT to the network analyzer. No microscope or delicate probes were needed for our purposes. However, the network analyzer still needed to be calibrated before use to ensure that the loss reading did not include the loss across the coaxial cables connected to the network analyzer.

EXPERIMENTAL RESULTS

Many inconclusive tests were conducted before we achieved useful results. Gold microstrips fabricated on PET (polyethylene terephthalate) substrates proved to have substantial amounts of loss. Additionally, the linear S_{11} value was often

quite high, indicating an unwanted impedance mismatch. Failure to gather reliable data led to fabricating CPWs and microstrips on alumina substrates.

The dimensions for the CPWs to be tested called for very precisely sized sputtering masks. Microstrip masks required a laser-cutting machine to fashion a rectangular opening of the proper dimensions in a very thin piece of cardstock. Occasionally, this mask would not lay perfectly flat on the substrate and an unwanted effect called shadowing would occur. This allowed a very thin layer of gold, in addition to that sputtered within the proper dimensions, to fall on the outside of the desired microstrip dimensions in an area that should be covered by the mask. This phenomenon, though undesirable, was tolerated as it did not cause an unreasonably high impedance mismatch when the microstrip was tested with the network analyzer. However, this type of mask was unacceptable for CPWs, which required more precision due to the proximity of the conducting planes. If a shadowing effect was too large on a CPW, it could create a short circuit between the center conductor and the two ground planes. Therefore, tape was used as a more protective mask. After the adhesive was applied to the alumina substrate, the appropriate shapes were etched using the laser cutter. Upon completion of the sputtering process, the rest of the tape would be removed completely, leaving only the appropriate gold films. This method, however, proved to be unsuccessful as the thin film of gold was easily torn when the adhesive was removed. When connected to the probe station, our CPWs failed to produce reliable results. Furthermore, the three-point probes used were so delicate that one was accidentally broken in the testing process and needed to be replaced.

Finally, we decided to use alumina substrates to fabricate microstrips using a simple cardstock mask that would be applied during the sputtering process. Due to limited time, we were only able to successfully test four different microstrips, though these four proved useful in determining a thickness relationship. Microstrips were made with film thicknesses of 25 nm, 50 nm, 100 nm, and 200 nm, and the following data were recorded when the microstrip was tested at 10 GHz.

Table 3: Calculated and Measured Values Corresponding to Film Thickness

Film Thickness (nm)	S₁₁ Linear Magnitude	S₂₁ Linear Magnitude	Linear Loss Fraction	Log. Loss (dB/cm)	Calculated Log. Loss (dB/cm) using TXLine	Calculated Linear Loss Fraction using TXLine
25	0.107	0.046	0.986	14.7	5.97	0.826
50	0.089	0.118	0.978	13.1	3.96	0.686
100	0.098	0.278	0.913	8.35	1.28	0.312
200	0.139	0.597	0.624	3.35	0.645	0.172
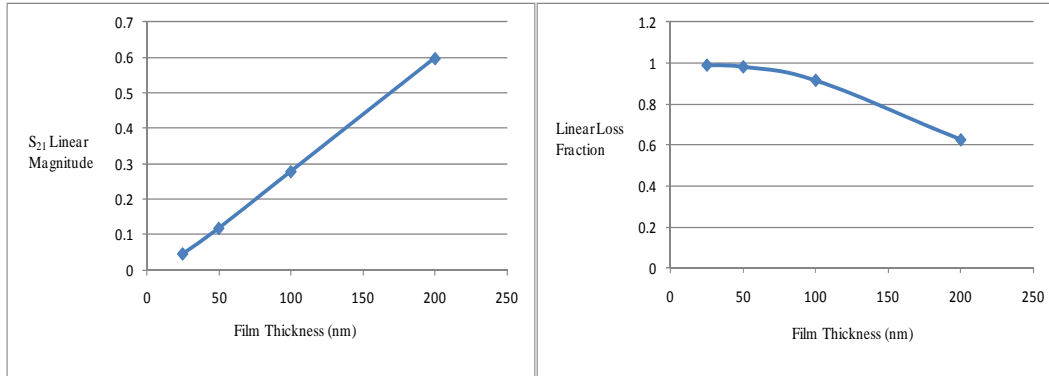
Figure 4: Graph showing the relationship between gold thickness and the linear magnitude of the S_{21}

Figure 5: Graph showing the relationship between gold thickness and the linear fraction of the loss across the microstrip.

DISCUSSION

These data show a very clear film thickness dependency. Though the calculated loss values are significantly lower than the measured values, both sets of data follow the same trend. The most likely source of error is a direct result of the imprecise method of fabrication. Sputtering gold onto a surface often leaves a very uneven film that hinders electrical current.

Furthermore, though our deposition method limits us to extremely thin films of gold, we can estimate the behavior of the complete range of thicknesses using the data we have collected. Figure 4 shows that the S_{21} linear magnitude appears to increase linearly with the film thickness, indicating an increased signal transmission. This is in keeping with our expectations since the films are far thinner than the skin depth of gold at 10 GHz. However, once the film reaches the skin depth, the S_{21} parameter would have to gradually plateau and never reach 1, as that would indicate that 100% of the signal is being transmitted.

If we were to see the entire waveform of the linear loss fraction (seen in Figure 5), it would look similar to that of a negative cubic root function. It begins

at a plateau close to 1, as the vast majority of the signal is being lost across the thin conductor. At a certain point, the loss begins to fall more drastically than a linear rate, but then it would once again plateau close to 0 as the loss percentage reaches a minimum.

CONCLUSION

From the tests performed on gold microstrips on alumina substrates, we can conclude that skin depth must be taken into account when fabricating thin transmission lines. Even the thickest of the lines we tested had a loss fraction of over 50%. However, if one desires a specific maximum tolerated loss for a transmission line, he can analyze the relationship between the film thickness and loss fraction in order to minimize the amount of metal used.

Additionally, it seems that for the purposes of our experiments, microstrips were far more successful than CPWs. Being limited to the use of a sputter coater to fabricate these transmission lines, microstrips were the only feasible option. For more advanced fabrication techniques though, CPWs may prove just as valuable as microstrips. Graphene transmission lines are often fabricated using a very precise technique called photolithography, which is capable of creating very small gaps between conducting planes on a CPW without compromising the quality of the transmission line. Despite this ability, microstrips are significantly easier to test as the adapting fixture is much simpler and quicker to use.

Finally, it is useful to perform theoretical analysis of a graphene microstrip for the purpose of comparing it with gold microstrips. Using the DC conductivity of graphene (10^8 S/m), the 3.35 Å separation between graphene layers,[8] and TXLine, we find that the loss across a 10 layer graphene microstrip would be approximately 4.16 dB/cm, assuming graphene acts as a metal conductor and exhibits the same skin effect properties. Recalling that TXLine calculated a loss of 5.97 dB/cm for a 25 nm gold microstrip (Table 3), it is astounding to note that a 3 nm thick graphene microstrip would have lower attenuation per unit length. Furthermore, it's important to note that graphene does not have a skin effect like other conductors,[9] and therefore would not be affected by the current-limiting effects of a thin metal film. Graphene exhibits ballistic conduction, a phenomenon in which charges do not scatter very easily, allowing them to flow unimpeded for a relatively long distance. This characteristic of graphene is important to its extremely high conductivity despite its extremely thin films. This research has developed an effective method for electrically characterizing conductors that can be used with graphene in the future.

ACKNOWLEDGMENTS

REFERENCES

[1] A. K. Geim & K. S. Novoselov, "The rise of graphene," *Nature Materials*, Vol. 6, pp. 183-191, Mar. 2007.

[2] D. M. Pozar, *Microwave Engineering*, 2nd ed. New York: John Wiley and Sons, Inc.

[3] S. Ramo, J. Whinnery, and T. Van Duzer, *Fields and Waves in Communications Electronics*, 2nd ed. New York: John Wiley and Sons, Inc.

[4] http://www.rfcafe.com/references/electrical/images/microstrip.jpg. Accessed: July 10, 2010.

[5] http://bama.ua.edu/~tmewes/Java/CPW/cpw.GIF. Accessed: July 10, 2010.

[6] D. R. Lide, *Handbook of Chemistry and Physics*, 89th ed. Boca Raton: CRC Press.

[7] The A to Z of Materials, "Graphene: The Best Electrical Conductor Known to Man," *The A to Z of Materials*, March 25, 2008. [Online]. Available: http://www.azom.com/news.asp?newsID=11679. [Accessed: July 10, 2010].

[8] A. R. West, *Basic Solid State Chemistry*, 2nd ed. Chichester: John Wiley and Sons, Inc.

[9] D. Jeon, K. J. Lee, M. Kim, D. C. Kim, H. Chung, Y. Woo, and S. Seo, "Radio Frequency Electrical Characteristics of Single Layer Graphene," *Japanese Journal of Applied Physics*, Vol. 48, Sep. 2009.

AUTHOR INDEX
(* Indicates REU Student Author)