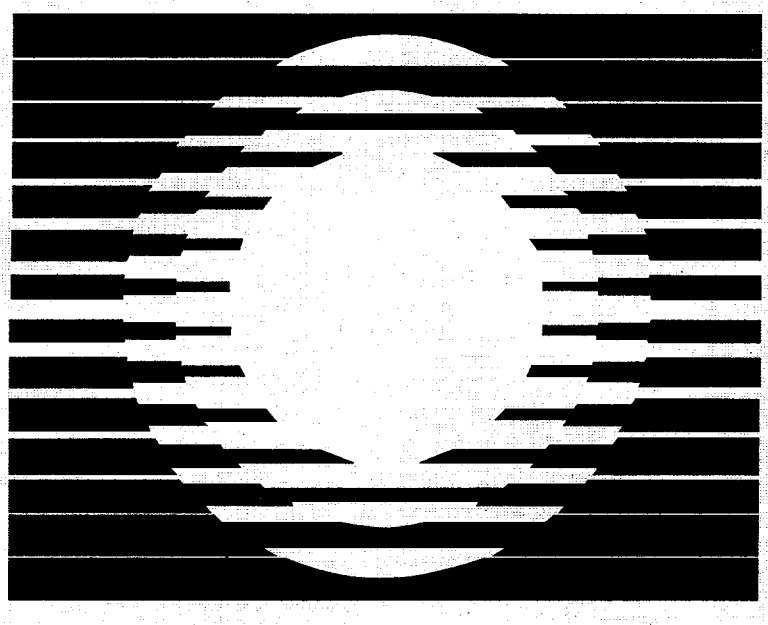


PROCEEDINGS
NINTH NATIONAL CONFERENCE ON UNDERGRADUATE RESEARCH
1995

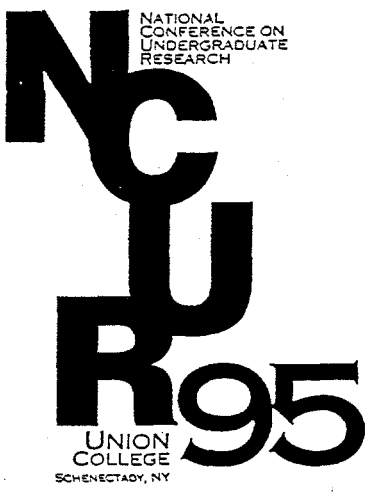
Robert D. Yearout
Editor

Volume II

UNCA
THE UNIVERSITY OF NORTH CAROLINA AT ASHEVILLE

NATIONAL EDITORIAL BOARD NCUR IX

UNCA EDITORIAL BOARD



PROCEEDINGS

NINTH

NATIONAL CONFERENCE

ON

UNDERGRADUATE

RESEARCH

29

VOLUME II

THE UNIVERSITY OF NORTH CAROLINA at ASHEVILLE

JULY , 1995

DISCLAIMER

Foreword for the **Proceedings of the Ninth National Conference on Undergraduate Research**
Union College, April 20-22, 1995

The Ninth National Conference on Undergraduate Research (NCUR 95) was held at Union College in Schenectady, New York. This annual celebration of undergraduate scholarly activity continues to elicit strong nation-wide support and enthusiasm among both students and faculty. Attendance was nearly 1650, which included 1213 student oral and poster presenters. In addition, thirty faculty and administrators gave talks describing their perspectives on undergraduate scholarly activity in symposia organized by the Undergraduate Research Network. Representatives from 280 institutions in 47 states, plus Puerto Rico and Washington, D.C., attended NCUR 95.

There were three plenary speakers at NCUR 95. Dr. Baruch Blumberg, the 1976 Nobel Laureate in Physiology and Medicine and currently Fox Chase Distinguished Scientist at the Fox Chase Cancer Center in Philadelphia, talked about his important discovery of the causes of viral hepatitis and the development of a screening test for the virus in donor blood. Donna Cox, Professor and Associate Director for Technology at the School of Art & Design at the University of Illinois, presented a fascinating video lecture on her work, which is at the interface of computer art and science. Julian Bond, a longtime, articulate champion of civil rights and now Distinguished Adjunct Professor at American University, gave his perspectives on the current state of minority issues in this country.

Perhaps the high point of NCUR 95 was ARTSFEAST, the Friday evening event which featured the work of undergraduates in the performing and visual arts. Over 50 students participated in this event, which included exhibitions of student paintings, drawings, and video art, student compositions by a professional brass quintet.

For the second year in a row, many student papers had to be rejected for presentation at NCUR due to conference size limitations. Thus, submitted papers for presentation at NCUR 95 were put through a careful review process before acceptance. Those of you who have been selected to have your paper appear in these Proceedings have been through yet a second review process. As a consequence, your work has been judged to represent an impressive level of achievement at the undergraduate level. We offer you our congratulations on a job well done, and we hope your work is the precursor of many fine contributions to come from future endeavors.

Tom Werner
Helen Madden
NCUR 95 Site Co-Chairs

Dr. Larry Wilson
Vice Chancellor for Academic Affairs
The University of North Carolina at Asheville

The UNCA Editorial Board has dedicated the Proceedings of the Ninth National Conference on Undergraduate Research to Dr. Larry Wilson, former Vice Chancellor for Academic Affairs at The University of North Carolina at Asheville. In the fall of 1995, Dr. Wilson left UNCA to take on a new role as president of Marietta College in Ohio. His past support of the Proceedings and his commitment to undergraduate research has been greatly appreciated and will be missed.

Dr. Wilson received a B.S. degree in Chemistry from Baker University in Baldwin, Kansas (1958) and a Ph.D. in Chemistry from the University of Kansas (1963). From there, Dr. Wilson embarked on a teaching career at Ohio Wesleyan University where he went on to serve as Acting Provost, Dean of Academic Affairs, and Executive Assistant to the President. During the 1987-88 academic year, which coincided with the first year of the National Conference on Undergraduate Research, Dr. Wilson came to UNCA. Since then, he has served the University as Professor of Chemistry, Vice Chancellor for Academic Affairs, and as Interim Chancellor since 1994.

While a member of the National Editorial Board and a reviewer for the NCUR Proceedings, Dr. Wilson's main support came in the area of finance. By dedicating funds from the UNCA Foundation to send students to the Conference, he also made money available to send the UNCA Editorial Board to review and process papers. This on-site processing made possible the September publication of the Proceedings each year. Without Dr. Wilson's support, the Proceedings would have come out much later in the year resulting in low sales and jeopardizing the Proceedings' ability to sustain itself. With the help of Larry Wilson, the Proceedings has become self supporting through corporate donations and sales as well.

The Proceedings Board wishes Dr. Wilson well in his new position and looks forward to seeing Marietta College participants at NCUR 10!

PREFACE

The 249 papers published in these three volumes are representative of the approximately 1200 papers and posters presented at the celebration of undergraduate research that was NCUR '95'. The faculty, staff and students of Union College are to be commended for making NCUR '95' such an enriching experience. The student authors deserve recognition for their hard work and for their persistence in seeing their research through to its final publications.

Special thanks are extended to Dr. Helen Madden, Ms. Kelly Buttridge and other members of the NCUR '95' Staff. Their help during the days of the Conference was invaluable. Many thanks also to the members of the National Board, who reviewed a large number of papers submitted to the Proceedings under time constraints. Their continued efforts to standardize has raised the Proceedings quality. This project would have not been successful without their assistance.

I would also like to thank the UNCA Editorial Board, who put in many long hours reviewing, checking and organizing papers, and Lisa Reeves, Patricia Roullard and Holly Beveridge, who did all of the cutting and pasting necessary for these volumes. I owe a special debt of gratitude to Katherine Whatley and Lisa Reeves for their support, encouragement and mentorship during the entire process of organizing the Proceedings.

To all the students who submitted papers to the Proceedings go thanks for their participation and congratulations for their achievements. Your enthusiasm and achievement will serve as role models for the next group of student researchers-who we look forward to meeting at the Tenth National Conference on Undergraduate Research at the University of North Carolina Asheville.

Robert D. Yearout
UNC at Asheville
July, 1995

**PROCEEDINGS OF THE
NINTH NATIONAL CONFERENCE ON
UNDERGRADUATE RESEARCH**

Number 1 1995

CONTENTS

VOLUME I

VOLUME II

ENGINEERING AND MATHEMATICS

VOLUME III

BIOLOGICAL SCIENCES

CHEMICAL SCIENCES

Carbon Dioxide Retention in Yucca Mountain

Shawonda Brockington

Department of Chemical Engineering
North Carolina A&T State University
1601 East Market Street
Greensboro, NC 27411-1200

Faculty Advisor:

Dr. Franklin King

The estimation of carbon dioxide retention in the Yucca Mountains is important in order to determine whether the Yucca Mountains represent an acceptable repository for high-level radioactive waste. This study was done in order to provide effective diffusivity data for carbon dioxide. The data could be used to calculate the amount of gas transfer of carbon dioxide through Yucca Mountain tuff layers. All samples studied came from a single drillhole, USW-G4, which is central to the proposed repository site. The samples were taken from the following layers: Upper Nonlithophysal (130.1 m), Middle Nonlithophysal (224.5 m), and Lower Nonlithophysal (243.9 m, 266.9 m, 297.3 m, and 335.2 m).

BACKGROUND

In order to solve the nation's nuclear fuel and high-level radioactive waste disposal problem, the United States is considering the Yucca Mountains as a possible repository site. The Department of Energy hopes to build an underground storage facility for 70,000 metric tons of nuclear waste at Yucca Mountain in Nevada by 2010. If studies of the Yucca Mountains indicate that it is a suitable location for a radioactive waste repository, the Department of Energy expects to apply to the Nuclear Regulatory Commission for authorization to construct the repository and for a license to operate it. The application must contain an examination of the ability of the repository system to comply with the regulations that govern it.

There have been studies on various aspects of the site. In a recent report, the Nuclear Regulatory Commission concluded that the ground water would not be affected by the possible nuclear waste repository. However, the question of diffusion, or gaseous transport of, radionuclides from the repository in the zones accessible to the environment is still to be answered.

Since spent nuclear fuel contains a significant amount of ^{14}C, gaseous releases of ^{14}C in the form of $^{14}CO_2$ are expected. The estimation of diffusive transport of CO_2 through the tuff is important in order to assess whether the Yucca Mountain site is an environmentally safe and acceptable repository site for high-level radioactive waste.

The objective of this research is to study the effect of environmental variables, such as temperature and pressure on the effective diffusivity of carbon dioxide through the samples taken from the Yucca Mountain site.

EXPERIMENTAL PROCEDURES

The procedure for determining the diffusivity of carbon dioxide in the Yucca Mountains involved a study-state counter diffusion method. Using this method, a sample pellet that was cored and cut to appropriate dimensions (2 mm thick. 25.4 mm in diameter) from Yucca Mountain tuff. A sample was then mounted in the diffusion cell (Figure 1) and placed in the tubular furnace. The temperature was kept constant during the diffusivity measurements. Carbon dioxide gas was sent across one face of the pellet, and argon was sent across the other face (Figure 2). The carbon dioxide and argon were supplied by pressurized gas cylinders. Rotameters were used to measure exit flow rates. Sampling tees with septums were placed in the inlet and outlet gas lines (Figure 3). Gas samples were obtained using a syringe and were injected into a computerized Perkin Elmer Gas Chromatography equipped with a Chromatographics-3 software system for analyzing the composition of the exit gas streams.

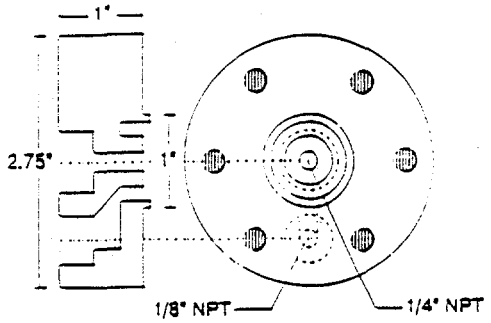


Figure 1: Design of one side of diffusion cell

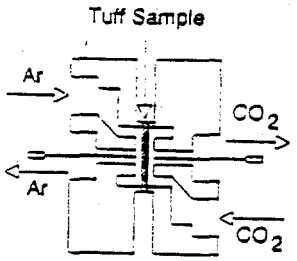


Figure 2: Diffusion cell showing the gas flow direction

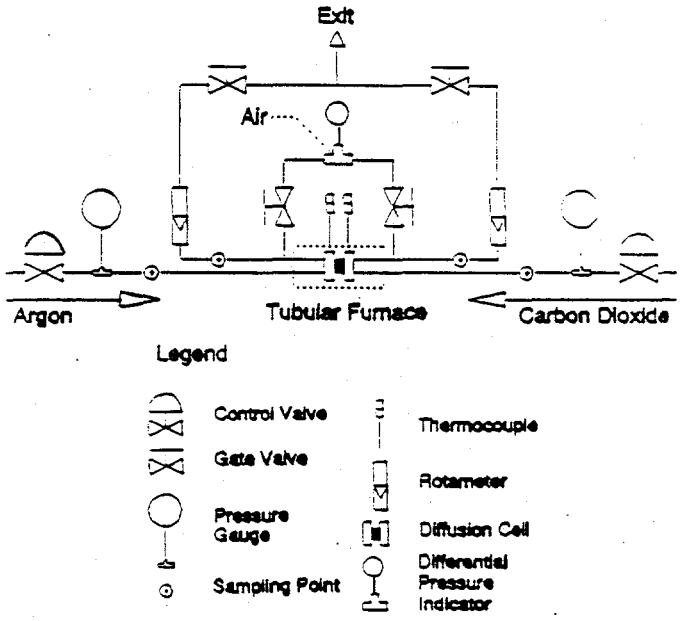


Figure 3: Diagram of the diffusivity-measurement system

The effective diffusivity of carbon dioxide through the samples was calculated using the diffusion cell data and Fick's Law according to the following equation:

$$N_{CO_2} = \frac{PD_e}{RT}\left[\frac{(Y_{CO_2})_{Ar} - (Y_{CO_2})_{CO_2}}{z}\right] \qquad (1)$$

where N_{CO_2} = molar gas flux
D_e = effective diffusivity
P = absolute pressure
T = absolute temperature
R = gas constant
$(Y_{CO_2})_{CO_2}$ = mole fraction of CO_2 in the CO_2 side
$(Y_{CO_2})_{Ar}$ = mole fraction of CO_2 in the Ar side
z = thickness of pellet

The molar flux of the carbon dioxide through the pellet can be determined from the argon rich exit stream and is given by:

$$N_{CO_2} = \frac{4P}{RT}\left[\frac{Q_{Ar}(y_{CO_2})_{Ar}}{d^2}\right] \qquad (2)$$

where d = effective pellet diameter
Q_{Ar} = exit volumetric flow rate of gas mixture in Ar side
$(y_{CO_2})_{Ar}$ = mole fraction of CO_2 in Ar side exit gas stream

Since the arithmetic mean of carbon dioxide concentration across a pellet gives a better representation of the gas concentration on each side of the pellet, the following mole fractions were used in the first equation:

$$(Y_{CO_2})_{Ar} = \frac{0 + (y_{CO_2})_{CO_2}}{2} \qquad (3)$$

$$(Y_{CO_2})_{CO_2} = \frac{1 + (y_{CO_2})_{CO_2}}{2} \qquad (4)$$

where $(y_{CO_2})_{CO_2}$ = mole fraction of CO_2 in the CO_2 side exit stream. Using the four equations described above, it is easy to calculate the effective diffusivity of carbon dioxide from the known experimental data.

Hydrodynamic flow of gases

If there is a difference in absolute pressure across the pellet, a hydrodynamic flow of gases will occur. Because of this gas transfer, equations 1 through 4 can be combined to find the Mass Transfer Coefficient. Equation 1 is then rewritten as follows:

$$N_{CO_2} = \frac{k}{RT}\left[P_{Ar}(Y_{CO_2})_{Ar} - P_{CO_2}(Y_{CO_2})_{CO_2}\right] \qquad (5)$$

where k = mass transfer coefficient
P_{Ar} = absolute pressure on argon side
P_{CO_2} = absolute pressure on carbon dioxide side

This equation is then used to calculate the mass transfer coefficient from the calculated and experimental data.

Table 1: Density and Porosity Data From the Tuff Samples

Subunit Name	Sample ID	Bulk Density (gm/cc)	True Density (gm/cc)	Porosity
Upper Lithophysal	03V 20150	2.053	2.5889	0.207
Middle Nonlithophsal	06V 20157	2.288	2.6170	0.126
Lower Lithophysal	09V 20163	2.314	2.6380	0.123
Lower Nonlithophsal	11V 20168	2.313	2.6340	0.122

Table 2: Calculated Values for Diffusivity and Mass Transfer Coefficient

Subunit Name	Sample ID	Differential Pressure (atm)	Mass Transfer Coeffiencient for Carbon Dioxide (cm^2/s)
Upper Lithophysal	03V 20150	0.0000 1.0207 2.0414 3.0620 4.0827	0.004112 0.033739 0.055837 0.075076 0.099357
Middle Nonlithophysal	06V 20157	0.0000 1.0207 2.0414 3.0620 4.0827	0.001737 0.036962 0.056963 0.073069 0.091333
Lower Lithophysal	09V 20163	0.0000 1.0207 2.0414 3.0620 4.0827	0.001177 0.038344 0.067836 0.090716 0.138287
Lower Nonlithophysal	11V 20168	0.0000 1.0207 2.0414 3.0620 4.0827	0.000632 0.039382 0.054201 0.066823 0.090764

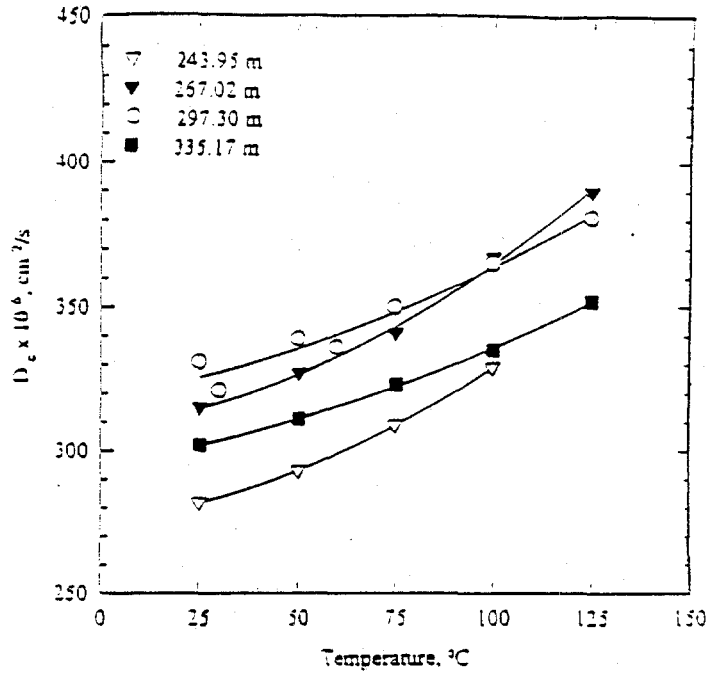


Figure 4: Experimental effective diffusivity data as a function for the sample pellet of lower lithophysal zone tuff

CONCLUSION

The counter diffusion fluxes of the argon and carbon dioxide through the pellet were estimated by measuring the composition of the diffusion cell using a gas chromatograph. In addition, the porosity, specific surface area, and mean pore radius were also determined. Results are seen in Table 1. In general, diffusivity increased with increased porosity, specific surface area, and mean pore radius. The measured data suggest that, effective diffusivity increased with increasing temperature for all samples tested, as illustrated in Figure 4. As pressure differential was increased across the sample, the bulk transfer of carbon dioxide became the controlling mechanism of carbon dioxide transport. Overall results of the study indicate that the mass transfer coefficient of carbon dioxide increases with increasing pressure difference, as described in Table 3.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Dr. Vivian Hampton and Dr. Franklin King for their guidance and support. The author also wishes to thank Ms. Kenyatta T. Davis for her dedication.

REFERENCES

Ali, M.S., S. Ilias, and F.G. King, *Diffusion of Carbon Dioxide Through Yucca Mountain Tuffs*, Presented at the 1992 Annual AIChE Meeting , Miami Beach, Florid, Nov 1-6, (1992).

Annotated Outline for Site Characterization Plans (OGR/B-5), U.S. Department of Energy, Office of Civilian Radioactive Waste Management Report. DOE/RW0142, p.33, April, (1987).

Bish, L., A.E. Ogard, and D.T. Vaniman, *Mineralogy-Petrology and Ground Water Geochemistry of Yucca Mountain Tuffs*, in Scientific Basis for Nuclear Waste Management VII, (1984), G.L. McWay, ed. Elsevier Publishing Company, New York.

Cunningham, R.E., Williams, R.J.J., *Diffusion in Gases and Porous Media*, Plenum Press, New York, (1990).

Nuclear Regulatory Commission, Proposed Amendment to 10 CFR Part 60, *Disposal of High-Level Radioactive Wastes in Geologic Repositories*, U.S. Government Printing Office, Washington D.C., (1984).

Travis, B.J., S.W. Hodson, H.E. Nuttall, T.L. Cook, and R.S. Rundberg, *Numerical Simulation of Flow and Transport in Fractured Tuff*, in Scientific Basis for Nuclear Waste Management, VII, (1984), G.L. McVay, ed., Elsevier Publishing Company, New York.

Wicke, E. and Kallenbach, R., *Counter Diffusion Through Porous Pellets*, Kolloid, Z., p. 135, (1941).

Youngquist, G.R., *Diffusion and Flow of Gases in Porous Solids*, Industrial and Engineering Chemistry, 62(8), August, (1970).

Zimmerman, R. G., F. B. Nimick, and M. P. Board, *Geoengineering Characterization of Welded Tuffs from Laboratory and Field Investigations*, in Scientific Basis For Nuclear Waste Management, VII, (1985), C. M. Jantzen, J. A. Stone, and R. C. Ewing, eds., Materials Research Society, Pittsburgh, PA.

Computer Simulation of Pressure-Volume Behavior in the vicinity of the Triple Point

Yea-Ling Ong

Department of Chemical Engineering
Lafayette College
Easton PA 18042

Research Advisors : Bruce D. Drake, Robert G. Root

Design of multiphase systems requires equations of state to predict physical properties. Conventional equations of state do not predict both solid and gas properties. Also, they require critical point data which are unavailable for large molecules. This work describes initial effort to develop an equation of state based on triple point data. Molecular dynamics is used to estimate isothermal pressure-volume behavior of model substances at temperatures above and below the triple point. The resulting behavior provides insight into the mathematical form of a triple point equation of state.

INTRODUCTION

The Van der Waals, Soave-Redlich-Kwong and Peng Robinson are all established cubic equations of state. The Carnahan-Starling equation is merely a modification of these conventional equations. Although these equations are adequate in predicting the physical behavior of fluids, they fail to model solids and liquids near the triple point where all three phases, solid, liquid and vapor exist in equilibrium. Further, these equations require critical point data which are unavailable for large molecules.

Thus the attempt is to develop an equation of state based on triple point data. Through computer simulation of molecular motion, the isothermal pressure-volume behavior of model substances is traced. The results obtained allow the formulation of a triple point equation of state.

SIMULATION DATA

Molecular dynamics is a method for computer simulation of molecular motion. It allows evaluation of pressure-volume-temperature data at conditions that are not physically realizable.

The intermolecular potential used in our model was the Lennard-Jones or 6-12 potential and has the form

$$U_{LJ}(r) = 4\epsilon[(\sigma/r)^{12}-(\sigma/r)^{6}] \qquad (1)$$

where ϵ is the maximum interaction energy, σ is the diameter of the molecule and r refers to the intermolecular distance. The $(\sigma/r)^{12}$ term refers to the electronic repulsion between the molecules while the $(\sigma/r)^{6}$ term refers to the dipolar attractive force.

At relatively low temperatures whereby σ and ϵ have values which fit the second virial coefficient, Argon has characteristics similar to the Lennard-Jones atoms (Rotenberg, 1967). Hence, the model used in this research was Argon with $\sigma = 3.405$ Å and $\epsilon/k = 119.8$ °K. Using molecular dynamics, as described below, pressure-volume data for all three phases were obtained for several reduced temperatures $T^* = kT/\epsilon$. Pressure-volume data for T^* ranging from 0.63 to 1.35, which encompasses the triple ($T^* = 0.68$) and critical ($T^* = 1.31$) points were obtained. Data were generated by a modified version of Lee's (1988) program for a microcanonical ensemble. An ensemble is a grouping of particles classified based on its constraints. A microcanonical ensemble is composed of a constant number of particles N at a fixed volume V and total energy E. This corresponds to the thermodynamic concept of an isolated system.

In molecular dynamics (MD), Newton's laws of motion are used to predict molecular behavior. By time-averaging the appropriate microscopic quantities (mass, velocities, etc.), we can calculate the desired macroscopic thermodynamic properties (density, pressure, enthalpy, etc.). Molecular dynamics was chosen because of its capability to model quasi-equilibrium states.

Periodic boundary conditions were used to enhance the ability of small systems to simulate the behavior or large systems. Imagine a box containing N number of molecules. Now make several replicas of this basic unit. One can then consider configurations of an infinite system while considering only a limited number of molecules N. As Barker and Henderson (1976) pointed out, the advantage is that surface effects which are usually large for a small N are greatly reduced. Originally pioneered for hard spheres by Alder and Wainwright, molecular dynamics was extended to molecules obeying a Lennard-Jones potential by Verlet.

SIMULATION DETAILS

Initially, the volume, temperature and number of particles were specified. The density was controlled by defining the size of the face-centered cube and the number of particles undergoing motion within it. For all runs, 256 particles were used while the reduced volume, $V^* = V/\sigma^3$ ranged from 0.97 to 600. Each individual particle was randomly assigned a position and an initial velocity. The velocities are corrected so that the ensemble kinetic energy corresponds to the initial temperature. In each cycle, the intermolecular forces acting on each molecule were evaluated using the Lennard-Jones intermolecular potential. A sixth order predictor-corrector method is used to solve the equations of motion. Once again, the velocity of the particles are scaled to correspond to the desired temperature, the intermolecular forces are tabulated and the cycle is repeated until the system reaches equilibrium. Normally, 2400 time-steps are required for equilibration. Thereafter, microscopic properties were amassed and time-averaged to produce the final properties required. Of special interest to this work is the determination of the equilibrium pressure. McQuarrie (1976) derived a relation between pressure and ensemble properties.

$$\frac{P}{kT} = \rho - \frac{\rho^2}{6kT}\int_0^\infty r u'(r) g(r) 4\pi r^2 \, dr \qquad (2)$$

The molecular dynamics implementation of the equation is given by Lee (1988):

$$\frac{PV}{kT} = 1 - \frac{1}{6NkT}\left[\sum_{i<}^{N}\sum_{j}^{N} r_{ij}(t) u'(r_{ij}(t))\right]_{\text{Time-Average}} \qquad (3)$$

where r_{ij} is the intermolecular spacing of molecules i and j. In effect, the pressure is calculated from the average intermolecular force.

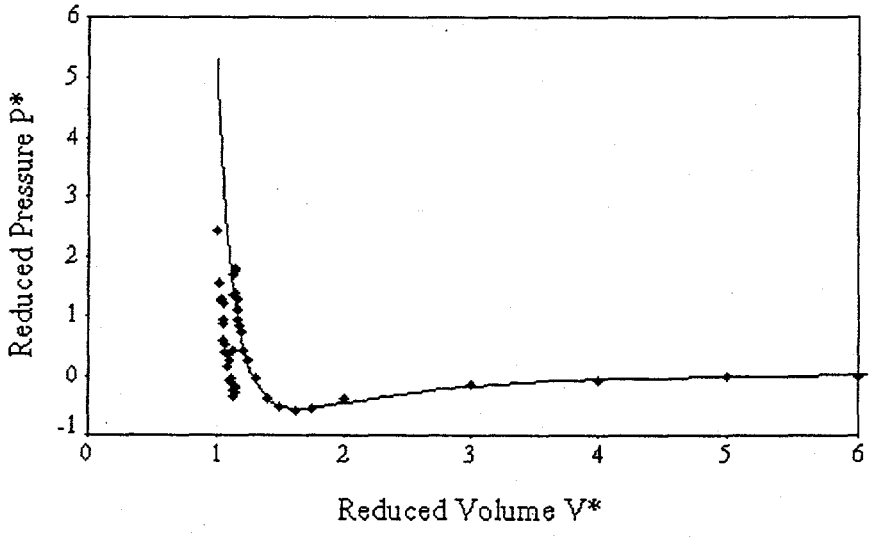


Figure 1. Liquid-Vapor at T* = 0.75

DEVELOPING AN IMPROVED EQUATION

Isothermal vapor-liquid data, obtained through computer simulation, were first modeled using classic equations of state. Figure 1 shows a fit of the Carnahan-Starling equation with a Peng-Robinson type perturbation term for $T^* = 0.75$. It is a typical representation of the pressure-volume behavior at all other temperatures, T^*. As expected, there was good agreement through the liquid-vapor range between the computer simulated data and classic equations of state, the closest agreement being that with the augmented Carnahan-Starling
equation

$$P = \frac{RT}{V}\frac{(1+\eta+\eta^2-\eta^3)}{(1-\eta)^3} + \frac{A}{V^2} + \frac{B}{V^3} \qquad (4)$$

Note that the classic equation fails to model the solid transition ($V^*<1.13$). Modelling the solid phase requires the addition of another term which has not been described in the literature. We treat the solid phase as a perturbation to the vapor-liquid model which exhibits a Van der Waals loop much like that of a vapor-liquid loop. The perturbation should be localized such that as V^* approaches $\pm\infty$, the reduced pressure P^* approaches zero and so that the vapor-liquid fit is not appreciably disturbed. The loop of the solid phase is actually a discontinuity but should be well represented by a steep odd function. Thus, the simplest mathematical form of the added term is V/V^2. The term must be further modified to translate the center of the perturbation away from $V = 0$ and to make the vertical step size at the transition finite.. The correlation with the added term then becomes:

$$P = \frac{RT}{V}\frac{(1+\eta+\eta^2-\eta^3)}{(1-\eta)^3} + \frac{A}{V^2} + \frac{B}{V^3} + \frac{S(V-M)}{(V-M)^2+C} \qquad (5)$$

This approach is an alternative to the direct modelling of the solid liquid transition as proposed by Hsu.

Figure 2 illustrates how the parameters in the solid perturbation term affect the behavior. The solid phase is evident only if we ensure that the slope (S/C) at the midpoint is greater than the decreasing slope of the liquid region. $S/C^{0.5}$ controls the peak to peak (vertical step) height. M is the center between the peaks while C determines the width.

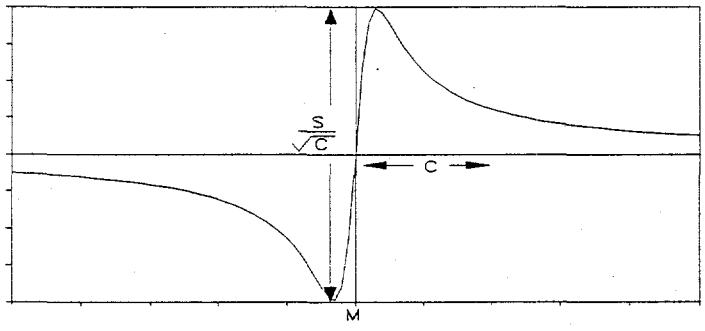


Figure 2 - Effect of parameters on solid perturbation term.

PARAMETER FITTING

For each set of isothermal pressure-volume data, non-linear regression was performed to fit the parameters. A typical fit is shown in Figure 3.
Some of the fitting parameters showed a consistent behavior with changing temperature. To obtain the anticipated thermodynamic behavior, the following trends are expected as the temperature increases. The width of the solid loop, as controlled by C, becomes smaller.

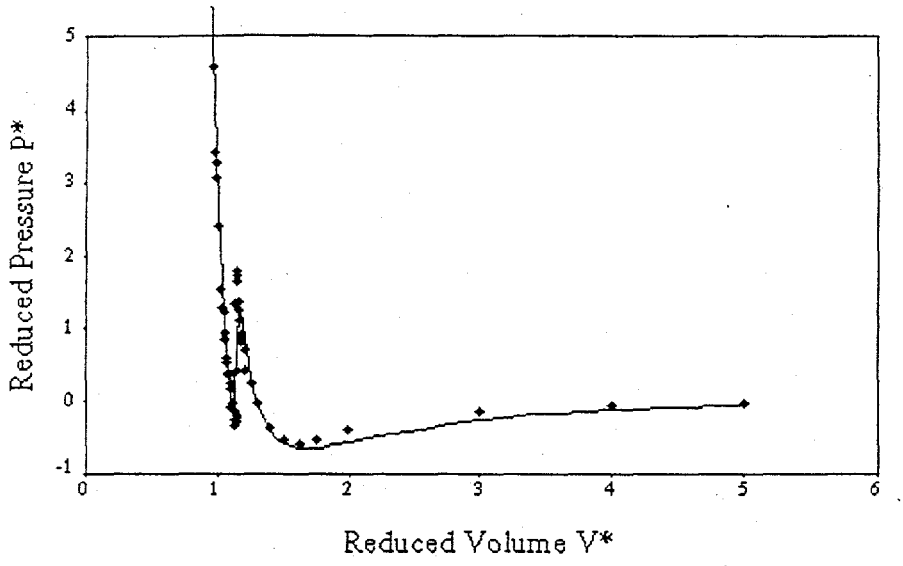


Figure 3

The slope of the liquid reference curve becomes more negative. Thus, S/C must increase to compensate. The center between the peaks is shifted left, toward a lower volume, so M also decreases. The intermolecular separation A is always negative but its magnitude increases with temperature.

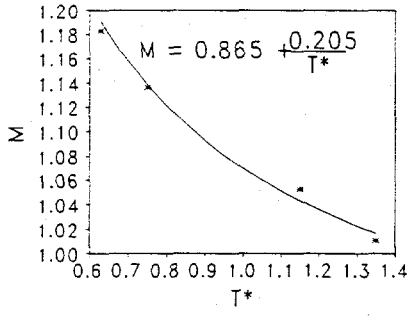


Figure 4

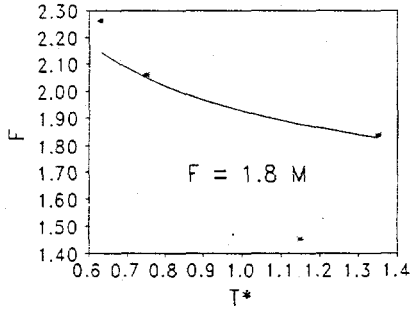


Figure 5

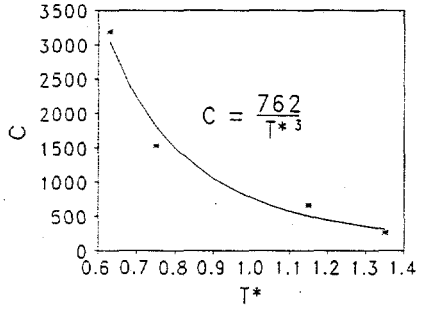


Figure 6

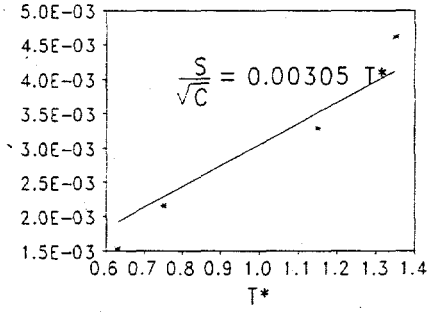


Figure 7

Figures 4 through 7 show the parameters as a function of temperature and their corresponding correlations. While the correct qualitative behavior is observed, the limited number of points prevents us from making any strict deductions about the behavior of each parameter.

CONCLUSIONS

We have shown that for liquid-vapor equilibrium, the experimental data can be correlated with the conventional equation of state. Specifically, the augmented Carnahan-Starling equation was an excellent fit. The addition of the solid loop however, required the addition of another term and the term we formulated gives a good fit. The solid-liquid-vapor correlation is a now a function of six parameters. Further, the effect of T* on these parameters were studied. M and C both decreased consistently with T* while $S/C^{0.5}$ increased with T*. The magnitude of A also increased with T*. F decreases with from T* of 0.63 to 1.15 but then increases. The increase may be an artifact created by fact that the solid loop is at the extreme edge of our data. S and C both fluctuated with temperature. Clearly, more data is required before developing predictive methods for the parameters.

ACKNOWLEDGEMENTS

We gratefully acknowledge support for this research by the Lafayette College EXCEL program.

REFERENCES

Barker, J.A. & Henderson, D., What is "liquid"? Understanding the states of matter, Review of Modern Physics, v 48 no. 4, Oct 1976, pg. 597-603

Hsu, T., Mou, C., Molecular Physics, v75, 1992, pg. 1329

Lee, L.L., 1988, Molecular Thermodynamics of Nonideal Fluids, Butterworth Publishers, Stoneham MA

McQuarrie, D.A., 1976, Statistical Mechanics, Harper & Row Publishers Inc., New York

Rotenberg, A., Journal of Chemical Physics, v47, 1967, pg. 4873

Enzymatic Resolution of Enantiomeric β-Blockers in Non-Aqueous Solvents

Christopher J. Schult

Department of Chemical Engineering
Lafayette College
Easton, Pennsylvania 18042

Faculty Advisor:

Dr. Bruce D. Drake

The Food and Drug Administration is currently moving towards mandating the production of enantiomeric pure pharmaceutical products. β-Blockers are used in the treatment of many cardiac ailments and each chiral form possesses different therapeutic functions. We have demonstrated a reaction that will esterify the secondary chiral alcohol in a stereospecific fashion. The reaction of interest is the conversion of an enantiomeric mixture of a common β-blocker, propranolol, into a chirally pure product using enzymatic esterification. The esterification is performed using a lipase from *Candida cylindracea* in cyclohexane. The selectivity of the lipase is investigated and Michaelis-Menten parameters are presented.

INTRODUCTION

Of the 200 most prescribed drugs, 114 of them contain at least one chiral center and only 12% are marketed and sold in enantiomeric pure form.[1] Effectively each drug dosage may be contaminated with an equal amount of an isomer that does not have any therapeutic effect and could possibly produce unsuspected side effects. For example, one of the most frequently used β-adrenergic blocking agents is the racemate propranolol. The d enantiomer of propranolol is 100 times more potent and is metabolized slower than the l enantiomer. [2] Birth defects caused a few decades ago by the drug thalidomide came about because one form of the compound is a sedative, while its mirror-image is a teratogen. Even when health is not at risk, economy or efficiency can stimulate more attention to producing chirally pure compounds. The market for chiral intermediates is currently $470 million worldwide and is expected to balloon to $2.7 billion by 2000.[1]

The β-blocker family of drugs is a large part of the current pharmaceutical market and many groups have been attempting to develop a method for chiral resolution of this family of compounds. Beinakatti's group has developed a method for the practical synthesis of ±-propranolol and atenolol. [3,4] These methods involve an alteration to the current industrial synthetic production methods for these compounds.

Enzymatic reactions in anhydrous organic solvents possess many advantages over reactions in conventional aqueous solvents. Klibanov et al.[5] present an in-depth review of the advantages of using nonaqueous organic solvent for enzymatic reactions. Many enzymatic reactions are now feasible due to the use of nonaqueous solvents. Previously many of the reactions that enzymes could catalyze were not posible in laboratory settings. These reactions, using water as the solvent, are limited by unfavorable thermodynamic equilibria, side reactions and insolubility of reactants. One common example is lipases. In non-aqueous solvents, lipases catalyze many different reactions (esterifications, transesterifications, aminolysis and oximolysis), however, in aqueous solvents the reactions are suppressed by the hydrolysis reaction. The insolubility of reactants in aqueous solvent is solved by using organic solvents for the reaction.

Previous work by Kirchner et al [6] tested numerous racemic primary, secondary and tertiary alcohols in a asymmetric transesterification reaction using porcine pancreatic lipase. In this study direct resolution of the secondary alcohol ±- propranolol was unsuccessful using a transesterification reaction.

This paper describes a method for the direct resolution of propranolol using a lipase catalyzed esterification reaction in a non-aqueous solvent, cyclohexane. Lipase from *Candida cylindracea* is used to catalyze the esterification of propranolol with stearic acid. The ability of the two enzymes to efficiently resolve the enantiomeric mixture of propranolol is determined. *Candida cylindracea* has been reported to show poor selectivity. [3,4]

Due to the similarity between many of the β-blockers, the reaction that has been

demonstrated with propranolol should work on other β-blockers with a chiral center located at the secondary alcohol.

EXPERIMENTAL SECTION

Materials

Lipase from *Candida cylindrecea* was purchased from Sigma Chemical Company and had a specific activity of 860 units/mg solid. One unit is defined as the amount that will hydrolyze 1.0 μ equivalents of fatty acid from olive oil in one hour at 37 °C and pH 7.2 (30 minute incubation).

(±) - Propranolol as well as the individual isomers were obtained from Sigma Chemical Company. All High Pressure Liquid Chromotography solvents used were obtained from Fisher and were of HPLC grade. The 1 N sodium perchlorate was prepared by dissolving HPLC grade solid sodium perchlorate in HPLC grade water. Two separate grades of Stearic acid were used in the esterification reaction, Grade I (>99%) and Grade II (>90%). The reaction solvent was cyclohexane (USP Grade) with a co-solvent (5%) acetone (USP Grade). The cyclohexane was further purified by recrystallization in our laboratory.

Assays

The analytical method used to determine the amounts of each enantiomer present employed a Hewlett Packard 1050 series HPLC system with UV detector. The UV detector was attached to a HP 3396 Series II Integrator. A Chiracel OD-R column, purchased from Chiral Technologies, was used to separate the enantiomers before the detector. The carrier solvents, 0.5 ml/min, were 1 N Sodium Perchlorate (55%) and acetonitrile (45%). Retention times were 13 and 17 minutes for the R and S formas of propranolol. Helium gas was used to degas the carrier solvents. The UV detector was set at 320 nm. The threshold detection limit of this system was approximately 2 ppm.

Procedures

The esterification reaction was carried out in cyclohexane (85 %) and acetone (15 %). The temperature of the reaction was held constant at 38 ± 0.2 °C in a constant temperature water bath. The reaction was mixed under magnetic stirring during the reaction. Temperature was monitored throughout the reaction with a NIST traceable thermometer (Fisher). The units of enzyme used were kept constant in each of the reactions (2310 units). The concentrations of propranolol and stearic acid were varied using a factorial design method to determine the effects of both the individual and group contributions.

RESULTS AND DISCUSSION

Reaction

The reaction used in this study is an enzyme catalyzed stereospecific esterification of (±)-propranolol with stearic acid. The reaction begins with (R/S) mixture of propranolol and stearic acid, an enzyme (lipase) is added to the reaction mixture yielding the original R-propranolol and a modified form of the S-propranolol (S-propranolol sterate). This can be seen schematically in **Figure 1** .

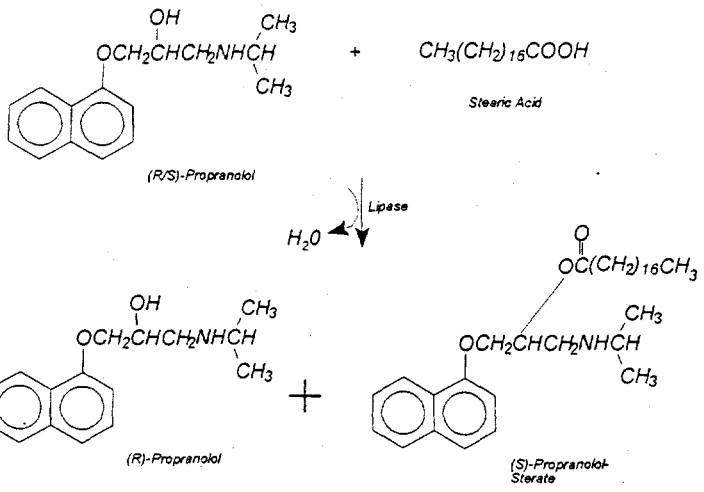


Figure 1. Schematic of enzyme catalyzed esterification reaction.

Kinetics

The kinetics of the lipase catalysis of esterification reactions has been determined to follow Michaelis-Menten first order kinetics [7]. The Michaelis-Menten equation is rearranged to obtain a form of the equation that can be easily graphed to obtain the desired parameters[8] (equation 1).

$$\frac{1}{t}\ln\left[\frac{1}{1-x}\right] = \frac{V_{max}}{K_m} - \frac{C_{propranolol}\,x}{t}\,\frac{1}{K_m} \qquad (1)$$

Table 1. Michaelis-Menten parameters. ^{R} and ^{S} - denote values for R and S forms of propranolol.

Trial	Km^R (mM)	$Vmax^R$ (mM min^{-1})	Km^S (mM)	$Vmax^S$ (mM min^{-1})	%ee R/R+S
1	0.055	8.91E-06	0.065	.00013	62.18
2	0.013	1.36E-07	0.019	9.4E-06	58.59
3	0.054	1.21E-06	0.068	1.8E-05	52.06
4	0.016	9.14E-07	0.009	7.0E-06	69.47

The Michaelis-Menten parameters for each of the trials were determined from a linear regression of the data and graphs of the data can be seen in **Table 1**. The graph of the data for both the S propranolol for each of the trials can be seen in **Figure 3**.

Factorial Design: Determination of Effect of Initial Concentrations Kinetics

The factorial design was used to determine the contributions of the initial propranolol and stearic acid concentrations on the relative rate of reaction and the selectivity of the

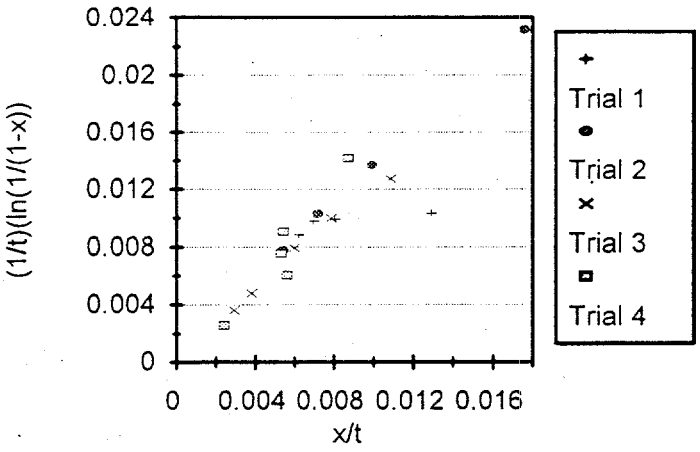


Figure 2. Determination of Michaelis- Menten Parameters for R-propranolol.

enzyme. Figure 4 shows a comparison between the R and S Michaelis-Menten parameters for trial 2. The R and S propranolol lines are virtually identical in the upper region (higher concentration), however, in the lower concentration region the lines deviate. The S-propranolol line changes slope in the lower concentration region. The decrease in slope results in an increase in reaction rate for the S-propranolol form (slope = $1/K_m$). In order to compare the different trials a relative rate relationship was used. The relative rate was the ratio of V_{mS} to V_{mR}. The higher this value the faster the S-propranolol is reacting when compared to the R-propranolol. **Figure 4** shows the comparison of these relative rates. When a high propranolol concentration is present, the amount of stearic acid does not significantly affect the relative rate. However, when the propranolol is present in low concentrations, the concentration of stearic acid present has a large contribution to the relative reaction rate. At low concentrations of stearic acid and propranolol the relative rate is 10 times larger than when stearic acid is present in large concentrations.

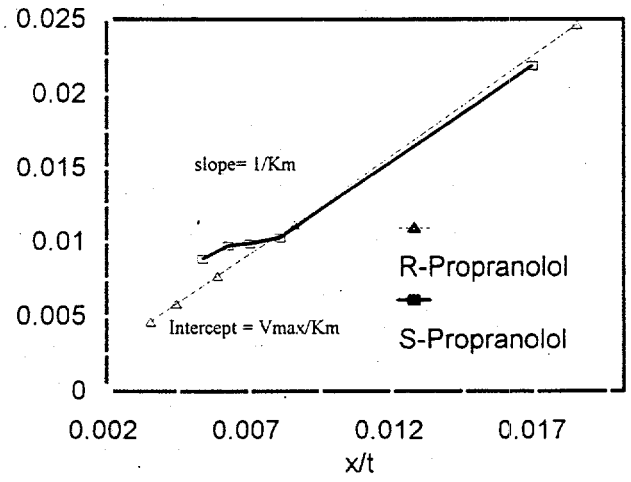


Figure 3. Comparison of R and S propranolol for trial 2.

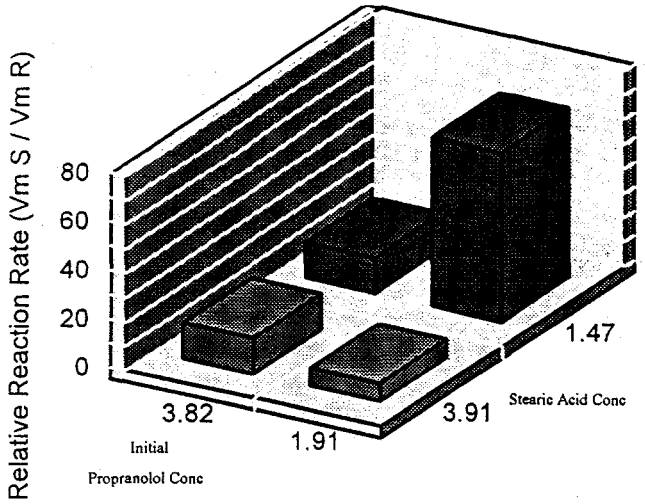


Figure 4. Factorial Design results for determining the contribution of initial concentrations of propranolol and stearic acid on relative reaction rates.

CONCLUSIONS

The proposed enzymatic reaction has been demonstrated and resolution of propranolol is possible by this method. After initial kinetic studies, it can be concluded that the selectivity of the lipase is limited and that other lipases should be investigated. The R-propranolol reaction does fit the Michaelis-Menten model, while the S-propranolol does not due to a change in slope at lower concentrations.

The highest relative reaction rate for the S-propranolol form occurs when the initial concentrations of propranolol and stearic acid are low. According to the initial study, reaction should take place in a CSTR reactor. The flowrates for the CSTR reactor should be such that the feed stream conversion is almost instantaneous. This type of reactor will produce the lowest possible concentrations of each of the reactants and therefore produce a higher relative rate.

Due to the necessity for a low concentration in the reactor, a low energy downstream separation must be used in order to allow for efficient recovery of the products. The proposed solvent, supercritical carbon dioxide, lends to a very favorable separation process. Due to the high dependance of solubility on pressure under supercritical conditions, a simple pressure reduction will allow for the precipitation of each of the products in an energetically favorable fashion. The use of pressure reduction for separation will also allow for particle size and size distribution to be set to the desired parameters by simply adjusting the rates of depressurization. This ability to produce a uniform particle size is important in the pharmaceutical industry.

REFERENCES

1. Chopey, Nicholas. "Chiral is now a bigger byword for drugmakers." Chemical Engineering. Nov. 1991.

2. Ward, J. A. "Chiral Drugs." Chemical and Engineering News. September 19,1994.

3. Bevinakatti, H.S. "Practical Synthesis of Both Enantiomers of Propranolol." J. Org. Chem. 56:1991.

4. Bevinakatti, H.S. "Lipase Catalysis in Organic Solvents." J. Org. Chem. 57:1992. pp 6003-6005.

5. Klibanov, Alexander M. Enzymatic catalysis in anhydrous organic solvents." TIBS. 14:1989.

6. Kirchner, Gerald, Alexander Klibanov. J. Am. Chem. Soc. 107:1985, pp.7072-7076.

7. Zaks and Klibanov. "Enzyme Catalyzed Processes in Organic Solvents." Proc. Natl. Acad. Sci. 82:1985.

8. Fogler, Scott H. Elements of Chemical Reactor Engineering. P T R Prentice Hall. Englewood Cliffs, NJ.

Gas Holdup in a Bubble Column Reactor Using a Porous Plate Distributor: Gas Density Effect

Jum S. Kim

Department of Chemical Engineering
Lafayette College
Easton, PA 18042

Research Advisor: Dr. Sarvani Mikkilineni

The main aim of this study was to investigate the effect of gas density and the liquid properties on gas holdup using a porous plate gas distributor. Several gases (He, N_2, Air, O_2, and Argon) of varying gas density in the range of 0.1769 to 1.7828 kg/m^3 were used to perform the experiment. Different liquids (water, methanol soln., and acetone soln.) were used to study their effect on gas holdup. The Dynamic Gas Disengagement (D.G.D) experiments were also carried out to investigate the hydrodynamics of a bubble column. Gas holdup was found to decrease with increasing gas density and increase with increasing methanol or acetone solution concentration. Dynamic gas disengagement data profile displayed a single slope in bubble flow and two slopes in the churn-turbulent flow representing unimodal and bimodal distribution of bubbles, respectively.

INTRODUCTION

Bubble column reactors are presently employed in the chemical processing industry to bring about chemical transformation and/or physical separation. Their use in the mineral processing industry (as flotation column) appears to be gaining wide acceptability. In addition, biotechnology, food processing and pharmaceutical processes constitute a wide spectrum of chemical industries, where bubble column reactors are used in the processing of products.

Gas holdup (E_g) is the volume fraction of the gas in the column. It is an important parameter since it is indicative of the residence time of the gas and the interfacial area for heat and mass transfer. It is known that liquid coalescing properties affect greatly the coalescence of gas bubbles (Song et al., 1989). Del Pozo et al., (1994) showed that several organic liquids gave much larger gas holdups than water. Large increases in gas holdup could also be obtained by adding either ethanol or pentanol to water. The increase in alcohol solution reduces the surface tension, giving smaller bubbles and hence higher gas holdup values (Song et al., 1989, Fan et al., 1986). The major effect of increased gas density is to stabilize the regime of homogeneous bubble flow and, consequently, to delay the transition to the churn-turbulent flow regime (Krishna et al., 1991).

In order to maximize the mass transfer rate, certain variables need to be investigated. Hydrodynamic parameters, such as gas holdup, interfacial area, dispersion coefficient, bubble size, bubble rise velocity are dependent on the type of gas distributor used, operating conditions, and liquid and gas physical properties.

The primary objective of the present study is to examine the influence of gas density on gas holdup and to develop a model which can be used for industrial reactor scale-up purposes. Although industrial bubble column reactors are often operated at high pressures, it is only relatively recently that experimental data have become available on the influence of increased gas density on gas holdup in bubble columns.

The specific objectives of this research include:

To study the effect of gas density on gas holdup.

To study the effect of liquid surface tension on gas holdup.
To study the effect of liquid concentration on gas holdup.
To identify bubble flow region from the churn-turbulent region and to study hydrodynamic behavior of different gases in these regions.
To develop empirical correlations for gas holdup in terms of superficial gas velocity, liquid concentration, surface tension of liquid, and gas density.

EXPERIMENTAL

Gas holdup was measured in a 0.14 m internal diameter and 3.66 m high plexiglas column. A schematic diagram of the bubble column set up is shown in Figure 1. A polyethylene plate of 70 micron pore size was used as the gas distributor. Gas holdup and D.G.D were measured by using the MCG-1030 analog output gauge. This analog output gauge continuously measures the fluid level in the column. It is made up of a wave guide, a float, a motion device and probe housing. The float contains an embedded magnet and floats freely on the surface of the fluid in the bubble column. The fluid level at any gas velocity is obtained from the digital readout in length units.

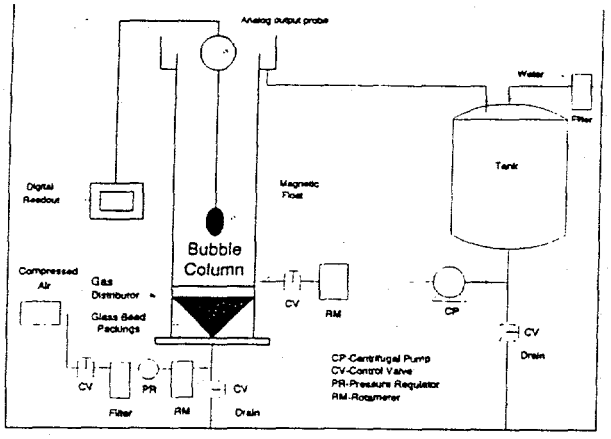


Figure 1: Schematic Diagram for Bubble Column Setup

The dynamic gas disengagement method requires accurate measurement of the decaying surface level of a gas-liquid dispersion in a column upon cessation of gas flow. D.G.D experiments were performed by using the analog output gauge and a stop watch. All experiments were performed batchwise with stationary liquid phase and continuous gas flow through liquid phase in the form of bubbles. The experiment was started by filling the column up to a certain level (above 1 m height) with liquid. The gases (He, N_2, Air, O_2, and Argon) were introduced through a rotameter and a gas sparger into the column. To construct the profile which relates decaying surface level of liquid to time, gas supply to the column was suddenly shut off and the level decrease was recorded (using the digital readout) every 2 seconds until the gas was completely disengaged. The same procedure was repeated over a range of gas velocities to cover both bubble flow and churn-turbulent flow. Three different liquids (water, methanol, and acetone solution) of varying concentration from 0.1% to 1% by volume were used.

RESULTS AND DISCUSSION

The influence of different liquids and their varying concentration on gas holdup is summarized in Figure 2. Gas holdup is dependent on the superficial gas velocity (V_g) in the

column. Figure 2 shows that the methanol and acetone solutions greatly increase the gas holdup. The increase in acetone concentration also increases gas holdup. This is because methanol and acetone solutions have a lower surface tension and produce smaller bubbles compared to water. It is also noticed that the transition to churn-turbulent region is shifted to a higher gas velocity when these organic liquid solutions are used.

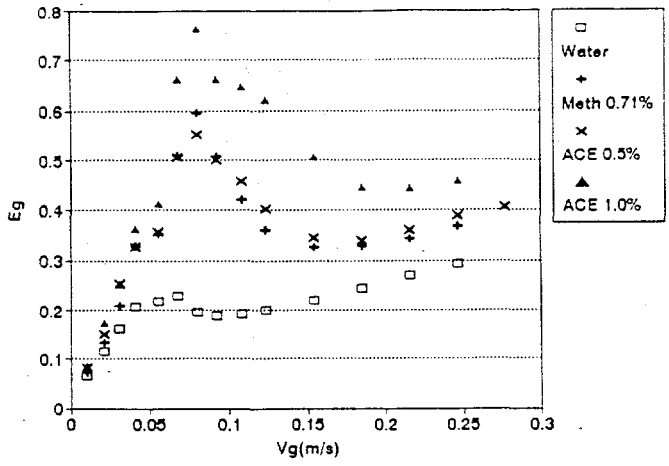


Figure 2: The Effect of Organic Liquids and Their Conc.

The effect of increasing gas density on gas holdup is presented in Figure 3. Gas holdup increases as the gas density decreases. However, it is not observed that the increase in gas density delays the transition to churn-turbulent flow. All four gases have the same transitional gas velocity of 0.065m/s. This conflicts with the results obtained by Krishna et al., (1991) where they found that the increase in gas density delays the transition to churn-turbulent flow.

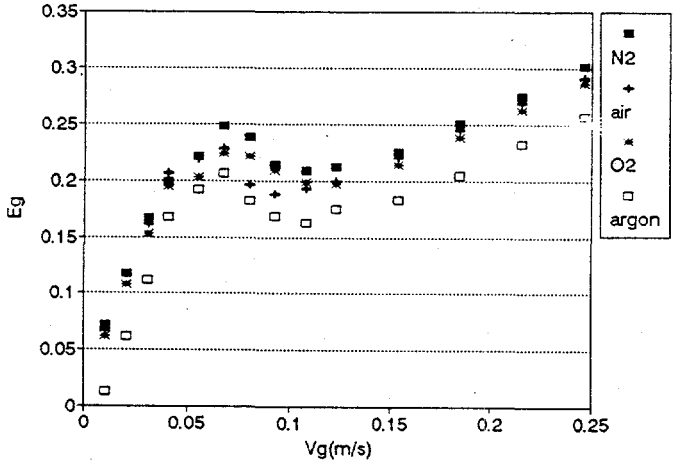


Figure 3: The Effect of Gas Density

Empirical correlations which relate gas holdup to superficial gas velocity, gas density, and liquid concentration, have been developed for both bubble and churn-turbulent flow.

1) Water-Various Gases System

$$E_G = 0.154 + 3.59\ V_g - 0.087\ \rho_g \quad \text{(Bubble Flow)} \qquad (1)$$

$$E_G = 0.717\ V_g^{0.547}\ \rho_g^{-0.502} \quad \text{(Churn-Turbulent)} \qquad (2)$$

2) Air-Various Acetone Solutions

$$E_G = 1.99\ V_g^{0.961}\ C_l^{0.148} \quad \text{(Bubble Flow)} \qquad (3)$$

$$E_G = 0.786\ V_g^{0.461}\ C_l^{0.0547} \quad \text{(Churn-Turbulent)} \qquad (4)$$

where C_l is acetone conc. in vol %, and ρ_g is density of gas.

The dynamic gas disengagement data allows for the analysis of the hydrodynamics of a bubble column both in bubble flow and churn-turbulent flow. The reduction in height is measured every 2 seconds upon termination of gas supply. The D.G.D data profile of air-water system in the churn-turbulent flow is shown in Figure 4.

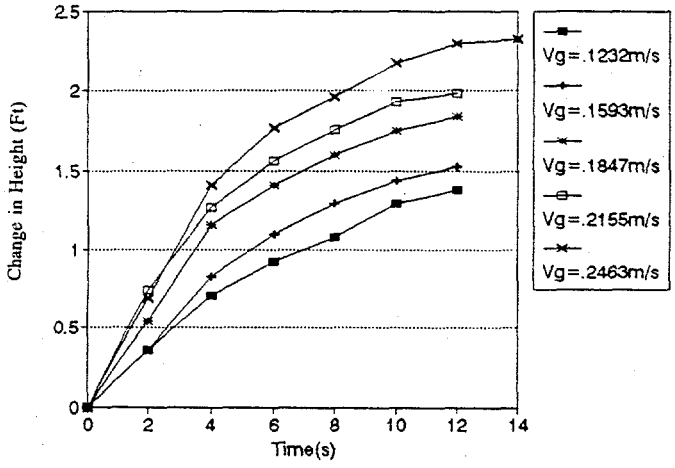


Figure 4: The D.G.D Data Profile (Churn-Turbulent Flow)

In bubble flow, the plots are approximately linear and consist of one slope, indicating a homogeneous bubble size distribution. However as the gas velocity reaches the churn-turbulent flow, the column contains two different bubble sizes distributed in the column, showing a bi-modal distribution. As shown in Figure 4, the line shows a change in slope after 4 seconds have elapsed. The first slope indicates the fast escape of large bubbles. The small bubbles escape after 4 seconds which is represented by the reduced slope. The velocity that represents the beginning of the churn-turbulent flow varies with the liquid used. When water is used as a liquid, this gas velocity is approximately 0.09 m/s, and about 1.5 m/s when the methanol or acetone solutions are used.

The slope of the D.G.D profile of different gases in bubble flow increases as the superficial gas velocity increases and as gas density decreases as depicted in Figure 5. This displays the trend that a lighter gas induces more larger bubbles than a heavier gas which escape faster upon termination of gas supply. Large bubbles decrease the gas holdup because of their small surface area. On the other hand, a lighter gas produces more bubbles than a heavier gas. The total interfacial area in the column thus increases with the use of a lighter gas due to the greater quantity of bubbles produced.

In the churn-turbulent flow, the first slope of the D.G.D profile was divided by the second slope, and this is plotted against gas velocity (Figure 6). A lighter gas produces more small bubbles as compared to a heavier gas, although the fraction of large bubbles remains higher throughout the churn-turbulent flow region. The large fraction of large bubbles lowers the gas holdup; however, the overall gas holdup increases due to the large amount of small bubbles produced.

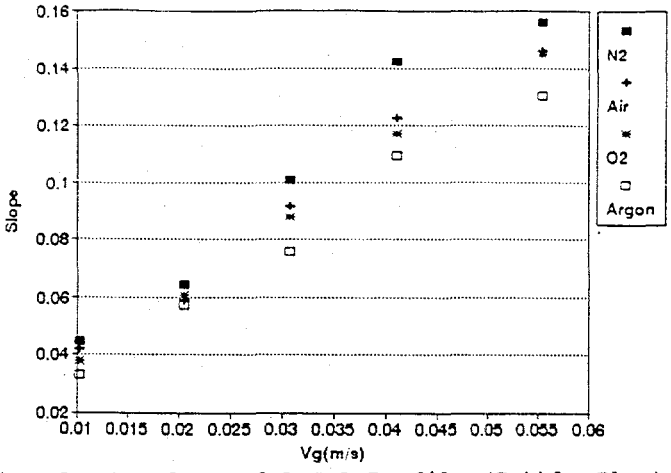
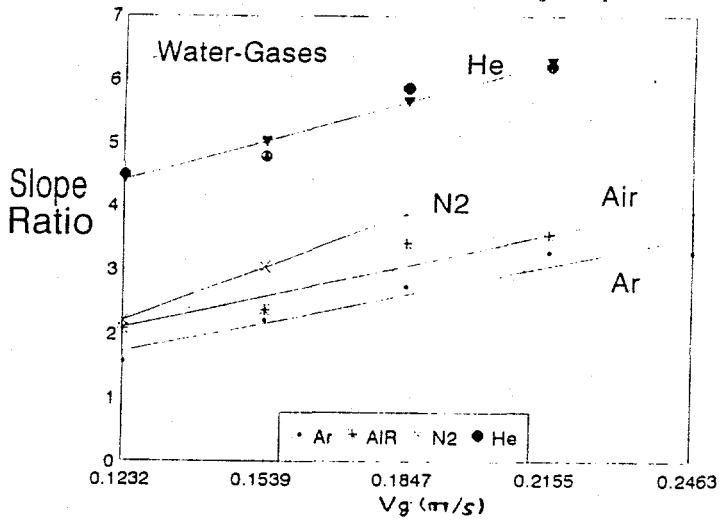


Figure 5: The Slope of D.G.D Profile (Bubble Flow)



Figrue 6: The Slope Ratio of D.G.D Profile (C-Turbulent flow)

CONCLUSIONS

1. A decrease in gas density increases gas holdup.
2. The use of acetone solutions and an increase in its concentration greatly increases the gas holdup.
3. The transition velocity is delayed with the use of acetone or methanol solutions. However, the increase in gas density does not affect the transition velocity.
4. A bi-modal bubble size distribution of churn-turbulent region was clearly verified by the D.G.D experiment.
5. A lighter gas produces more small and large bubbles at the same velocity as compared to a heavier gas in the churn-turbulent flow.

REFERENCES

1 M. Del Pozo, C.L. Briens and G. Wild, The Chemical Engineering Journal, 55 (1994) 1-14.
2 G.H. Song, F. Bavarian, L.S. Fan, R.D. Buttke and L.B. Peck, Can. J. Chem. Eng., 67 (1989) 265.
3 R. Krishna, P.M. Wilkinson and L.L. Van Dierendonck, Chemical Engineering Science, 46 (1991) 2491-2496.
4. L.S. Fan, F. Bavarian, R. Gorowaraa, B.E. Kreischer, R.D. Buttke and L.E. Peck, AIChE Meet., Miami, FL, 1986, paper 41st.

ACCELERATION FEEDBACK CONTROL STRATEGIES FOR ACTIVE MASS DRIVERS

Anmarie E. Belknap

Department of Civil Engineering and Geological Sciences
University of Notre Dame
Notre Dame, IN 46556-0767

Faculty Advisor: B.F. Spencer, Jr.

INTRODUCTION

Recent trends toward building taller, more flexible structures have resulted in designed structures which are much more susceptible to dynamic loadings, such as those caused by wind and earthquakes. Subsequently, researchers have sought to develop active control strategies for structural systems which can minimize the effects of these dynamic loadings. Much of the work done previously in this area has utilized full state feedback control strategies (i.e. the use of displacements and velocities). However, accurate measurement of the displacements and velocities is difficult to achieve directly, particularly during seismic activity, when the foundation of the structure is moving with the ground. Since accelerometers can provide reliable measurements of acceleration at strategic points on the structure quickly and inexpensively, the use of acceleration measurements is ideal during seismic activity.

In this paper, experimental verification of acceleration feedback control strategies is provided for an active mass driver system. Acceleration feedback strategies were found to be effective in previous experiments with both active bracing systems (Spencer, *et al.*, 1993) and an active tendon system (Dyke, *et al.*, 1994b). The experiments in this paper verify that the use of acceleration feedback control strategies with an Active Mass Driver System (AMD) is highly effective in reducing structural responses during seismic activity.

EXPERIMENTAL SETUP

Tests with the active mass driver system were conducted on an earthquake simulator located at the Structural Dynamic and Control/Earthquake Engineering Laboratory at the University of Notre Dame. A three-story, single-bay model building was utilized as the test structure, as seen in Figure 1. The steel building is 62 inches tall, and masses are placed on each floor, totaling 500 lbs.

The Active Mass Driver System was placed on the third floor of the structure for control purposes. The AMD consists of a single hydraulic actuator with steel masses attached at each end of the piston. For this experiment, the moving masses on the AMD weighed 11.5 lb and consisted of the piston, piston rod, and steel disks bolted to the ends of the piston rods. The total mass of the structure, including the non-moving parts of the AMD, weighed 680 lbs. Thus, the moving mass was 1.7% of the structure mass.

Accelerometers positioned on each of the three floors, as well as on the AMD, measured the absolute acceleration of the building and of the AMD. An accelerometer on the base of the structure also measured ground excitation. The displacement of the actuator was obtained by an LVDT (linear variable differential transformer) mounted next to the AMD on the third floor. The accelerations of the three floors, as well as the displacement of the AMD, were used to determine the mathematical model of the system and, thus, control design (see Figure 1).

Implementation of the digital controller was perfomed using the Spectrum Signal Processing Real-Time Digital Signal Processor (DSP). Discussion of the computer system used for control design implementation can be found in Spencer, et al (1994) and Quast, et al. (1994).

SYSTEM IDENTIFICATION

One of the most important aspects of control design is the development of an accurate mathematical model of the system. This was accomplished by measuring input/output relationships of the experimental system and then constructing a mathematical model that matches this behavior as closely as possible. This system identification begins with the collection of input/output data in order to determine the relationships between the accelerations of the floors and the displacement of the AMD. Secondly, a mathematical model is computed according to this data, and finally, the accuracy of the model

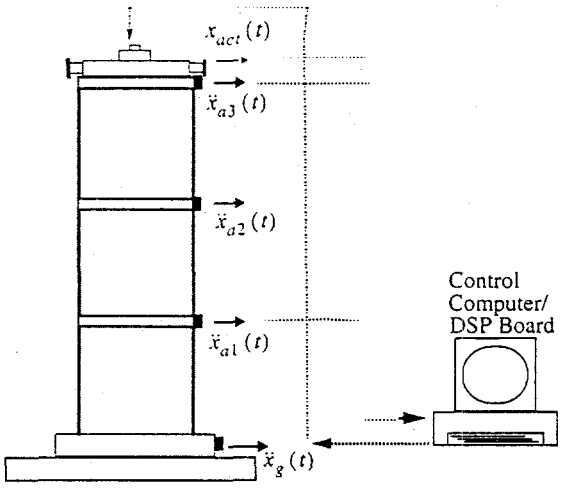


FIGURE 1. Schematic Diagram of Experimental Setup.

is evaluated.

For the AMD experiment, the frequency domain approach was utilized to identify the system. Transfer functions were determined experimentally from each of the system inputs to each of the outputs. Then, these transfer functions are modeled as a ratio of polynomials in s and used to determine a state space representation for the structural system. The block diagram of the identified structural system is shown in Figure 2. The two inputs are ground excitation, $\ddot{x}_g$, and the command signal to the AMD, u. The four measured outputs from the system were the AMD actuator displacement, and the absolute accelerations of the three floors $\ddot{x}_{a1}, \ddot{x}_{a2}, \ddot{x}_{a3}$ of the test structure.

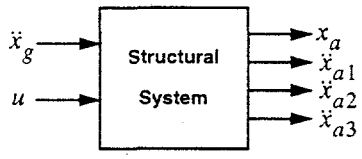


FIGURE 2. System Identification Block Diagram

To determine the transfer functions from the ground acceleration to each of the outputs, a bandlimited white noise acceleration was applied to the structure with the actuator command set to zero. Likewise, transfer functions from the actuator command signal were found by applying a bandlimited white noise to the actuator command while the ground was held fixed at zero. The data were then transferred to a SUN workstation, and experimental transfer functions obtained.

Figures 3 and 4 show representative magnitude and phase plots for the experimentally determined transfer functions. Figure 3 shows the transfer function from the actuator command to the second floor absolute acceleration, keeping the input to the ground at zero. Figure 4 shows the transfer function from the ground acceleration to the third floor acceleration (keeping the input to the control actuator at zero). The peaks of these graphs occur at 5.62 Hz, 16.99 Hz, and 27.99 Hz and correspond to the first three modes of the structural system. The associated damping ratios are given by 0.36%, 0.24%, and 0.15%, respectively.

Next, each of these transfer functions was modeled as a ratio of two polynomials of the Laplace variable s. The mathematical model was then adjusted for consistency with the physics of the physical problem and with each of the other transfer functions.

Since the structural system has several vibrational modes, we determined that only the first three modes of the system were necessary to model for the purposes of control design. Thus, the model of the system needed to be accurate up to approximately 35 Hz. The mathematical model for the transfer functions in Figures 3-5 is overlaid for comparison. The mathematical models for the other transfer functions were similar, in terms of quality, to those depicted here. Each individual transfer function was then assembled in a state space realization which was used as the basis for control design.

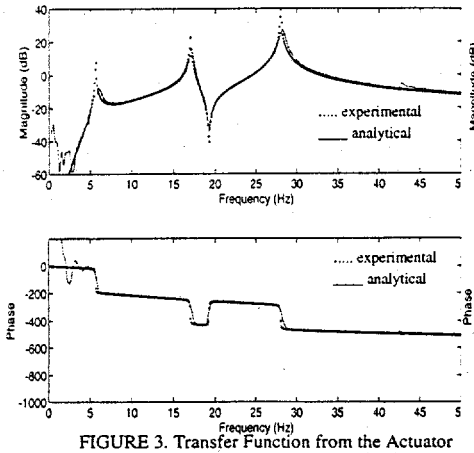


FIGURE 3. Transfer Function from the Actuator Command to the Second Floor Absolute Acceleration

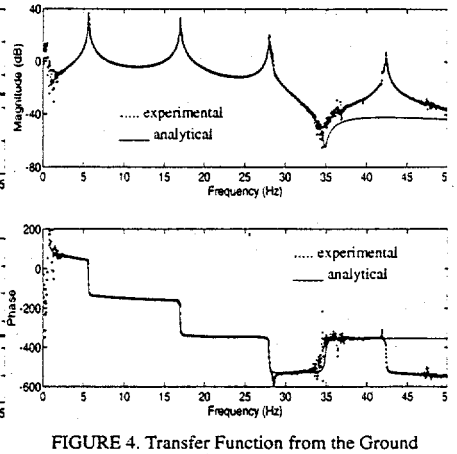


FIGURE 4. Transfer Function from the Ground Acceleration tot he Third Floor Absolute Acceleration.

CONTROL DESIGN

The controller was designed to minimize the absolute accelerations of each of the floors of the structure. Equal weighting was placed on each of the floor accelerations and a smaller weighting was placed on the displacement of the AMD. The model on which the controller was designed was highly accurate below 35 Hz, however, significant modeling error occurred at higher frequencies. If a controller is designed based on these higher frequency results, the system could become unstable, producing disastrous results. Therefore, no significant control effort was applied above 35 Hz. This can be assumed to be acceptable due to the small structural responses above 35 Hz and to the low frequency content of earthquakes.

Several experiments were completed, using the chosen controller, to evaluate the effectiveness of the controller and the acceptability of design. The results of these tests follow. For more detailed information concerning controller design, refer to Dyke, *et al* (1994a).

EXPERIMENTAL RESULTS

Two separate types of tests were run to determine the effectiveness of the controller design. First, a bandlimited white noise ground excitation, from 0-50 Hz, was used to excite the structure and RMS values of structural responses were evaluated. Secondly, a transient excitation which had the characteristics of an earthquake was used to excite the structure, and the ability of the controller to reduce peak acceleration values in the structural response were analyzed.

The experimental uncontrolled and controlled RMS responses for the bandlimited white noise excitation are given in Table 1. An 80-82% reduction was achieved in all RMS absolute accelerations. The experimental controlled and uncontrolled peak responses of the structure due to transient excitation are given in Table 2. A 56% reduction was noted in the third floor absolute acceleration. Figure 5 depicts the ability and speed of the controller to reduce the dynamic responses due to broadband excitation. A visual illustration of the ability of the controller to dynamically reduce the response is given in Figure 6, where an earthquake like excitation is applied.

Table 1.RMS Responses for Bandlimited White Noise Excitation (0-50 Hz)

Control Strategy	$\ddot{x}_{a1}, \frac{\text{in}}{\text{s}^2}$	$\ddot{x}_{a2}, \frac{\text{in}}{\text{s}^2}$	$\ddot{x}_{a3}, \frac{\text{in}}{\text{s}^2}$	x_a, in
Uncontrolled	82.85	103.19	125.42	--
With Controller	16.62	18.18	24.07	0.35
Percent Reduction	79.94	82.38	80.81	--

Table 2. Experimental Peak Responses for Transient Earthquake Excitation

Control Strategy	$\ddot{x}_{a1}, \frac{in}{s^2}$	$\ddot{x}_{a2}, \frac{in}{s^2}$	$\ddot{x}_{a3}, \frac{in}{s^2}$
Uncontrolled	144.36	120.08	171.64
With Controller	82.26	119.19	74.06
Percent Reduction	43.02	0.743	56.85

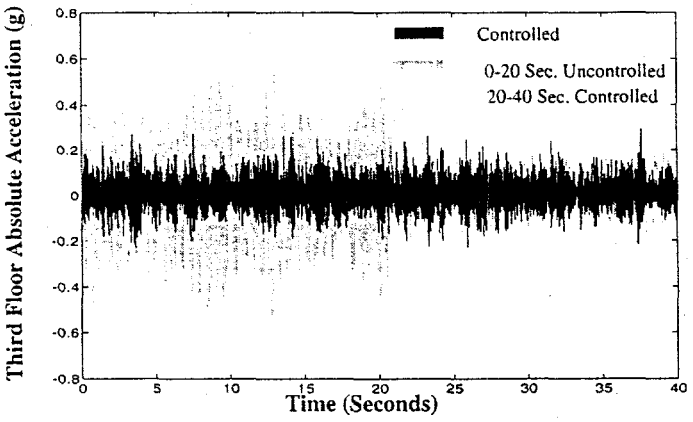


Figure 5.Uncontrolled and Controlled Experimental Third Floor Absolute Acceleration Responses.

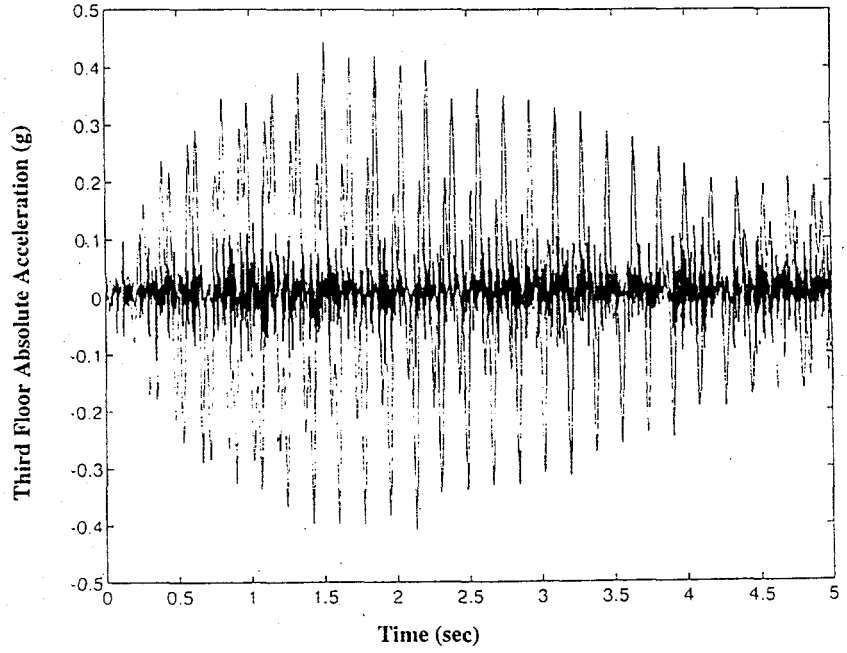


FIGURE 6.Controlled and Uncontrolled Experimental Third Floor Absolute Acceleration Responses.

CONCLUSIONS

Acceleration feedback control strategies have been successfully implemented and verified on a model structural system at the Structural Dynamics and Control/Earthquake Engineering Laboratory at the University of Notre Dame. Using input/output data, an accurate mathematical system was identified and controller was designed. Under broadband excitation, the AMD controller system was able to reduce RMS absolute acceleration responses by 80-82% in each of the three floors of the structure. During a transient excitation, similar to that of an earthquake, the AMD system achieved a 56% reduction in the peak acceleration of the third floor. Based on these results, acceleration feedback control strategies are promising for use with an Active Mass Driver System in effectively reducing the absolute acceleration of a structure due to seismic activity.

REFERENCES

Dyke, S.J., B.F. Spencer, Jr., A.E. Belknap, K.J. Ferrell, P. Quast, and M.K. Sain. 1994. Acceleration Feedback Control Strategies for the Active Mass Driver. *Proc. of the First World Conference on Structural Control,* Univ. of Southern California.

Dyke, S.J., B.F. Spencer Jr., P. Quast, M.K. Sain and D.C. Kaspari Jr. 1994b. "Experimental Verification of Acceleration Feedback Control Strategies for MDOF Structures," Proc. of the 2nd Int. Conf. on Comp. Mech., Athens, Greece, June 13-15.

Quast, P., B.F. Spencer Jr., M.K. Sain and S.J. Dyke. 1994. Microcomputer Implementation of Digital Control Strategies for Structural Response Reduction. *Microcomputers in Civil Engrg.*, (in press)

Spencer Jr., B.F., S.J. Dyke, M.K. Sain and P. Quast. 1993. "Acceleration Feedback Control Strategies for Aseismic Protection." *Proc., American Control Conf.*, pp. 1317-21.

Spencer Jr., B.F., P. Quast., S.J. Dyke and M.K. Sain. 1994. Digital Signal Processing Techniques for Active Structural Control. *Proc. of the 1994 ASCE Struct. Congress*, Atlanta, Georgia.

DYNAMIC ANALYSIS OF A BEAM SUPPORTED BY A WINKLER FOUNDATION SUBJECT TO A MOVING LOAD

Beth T. Kueter

Department of Civil Engineering and Geological Sciences
University of Notre Dame
Notre Dame, IN 46556

Faculty Advisor:

David J. Kirkner

There are several engineering problems which involve a moving load on an elastic or elastic-plastic foundation that may be considered infinite in extent. For example, rutting prediction of flexible pavements and overload of rail-track systems are such problems. Insight into the behavior of these systems can often be achieved through the use of a simplified model, such as a beam supported on a one- or two-parameter foundation. We have developed a finite element program to analyze the steady-state response of an infinite beam resting on an elastic-perfectly plastic or linear strain-hardening Winkler-type foundation subjected to a load moving at constant velocity. Non-dimensional results are presented which characterize the effects of load velocity and foundation plasticity.

INTRODUCTION

The analysis of beams on elastic foundations is a well-explored topic in engineering. Keen interest in the topic is a result of its practical application to modeling systems such as railroad rail-track systems and pavement systems. Beam deflection, bending moment, and contact pressure of a beam on an elastic foundation are just a few of the areas that were investigated in studies by Trochanis et. al. (1987). By applying steady-state assumptions, these problems can be solved using solutions developed for stationary loads, such as in Fryba (1972). In these cases, the foundation is also assumed to be elastic. There are cases, however, where the *inelastic* response of the foundation is such that permanent deflections remain after the load has passed. Examples of such cases are the overload of rail-track systems, resulting in failure of the ballast, and rutting in pavement systems. The beam model could be applied to these cases if a solution for the inelastic response could be attained. However, little research has been done in the area of inelastic foundations. The objective of this work is to formulate a scheme for modeling the results of inelastic foundation behavior when coupled with the dynamic effects of a moving load. A finite element method is employed which approximates the solution of the governing partial differential equations.

PROBLEM STATEMENT

The model consists of an elasto-plastic Timoshenko beam resting on an elasto-plastic Winkler foundation, i.e. a simple bed of springs with stiffness k. A Timoshenko beam was chosen because it was unclear how transverse shear deformation would affect the total deformation of the beam. The Timoshenko beam theory allows for these shear effects, whereas Euler beam theory would not. The beam is subjected to a vertical load, P, which moves at a constant velocity, c, in the x direction from left to right. The beam is assumed to be infinite in length, and is shown in Figure 1.

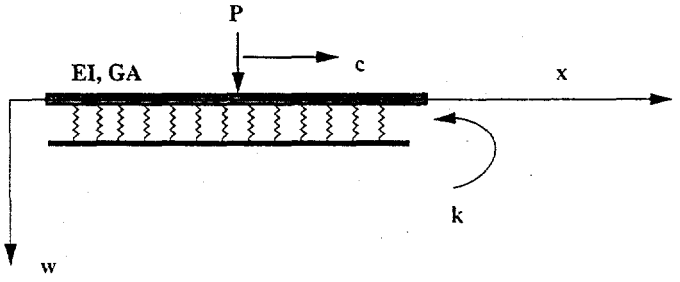


Figure 1: Schematic of beam.

The problem is simplified by assuming that the fixed coordinate system is replaced by a moving coordinate system whose origin is fixed at the point of application of the load, as shown in Figure 2. Also, the system is assumed to have reached steady-state, so that a "snap-shot" of the system would look the same regardless of the position of the load in space.

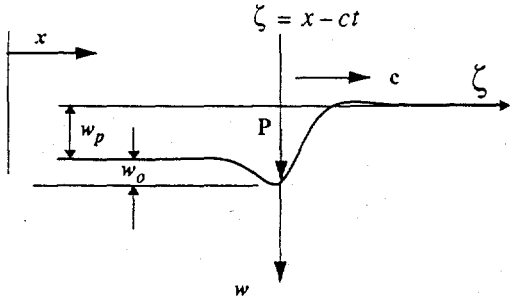


Figure 2: Illustration of deflected beam and moving coordinate system.

SOLUTION METHOD

The use of finite-element theory called for the manipulation of large numbers of sizeable matrices. With the use of finite-element theory, the applied force, P, is related to the vector of unknown displacements and rotations through a global stiffness matrix. The need to manipulate large numbers of these sizeable matrices led to the choice of Matlab (Pro-Matlab..., 1990) as the programming language.

The finite element method involves describing the system as a group of non-linear, algebraic equations. These equations have to be solved iteratively. The algorithm used in this program controls that iterative process. When the values of the solution have converged within a specified tolerance, the iteration stops. After initial test runs had been made, the results were verified by checking the agreement of the maximum displacement with the solution of the governing equations. This check was done for an elastic case and a simple elasto-plastic case (Fryba, 1972). Given a fine enough mesh, these answers agreed quite well.

The first concern with the solution method was the aforementioned assumption of infinite beam and foundation length. Because the problem had been proposed as a simple model of complex systems, one of the goals of the model was that its solution be simple and expedient. Choosing a long length so as to represent an infinite beam was computationally costly. Therefore, boundary matrix elements were developed which, when added to the stiffness matrix of a finite beam, would numerically "extend" the beam to infinity. To do this, a semi-infinite beam (Figure 3) was analyzed.

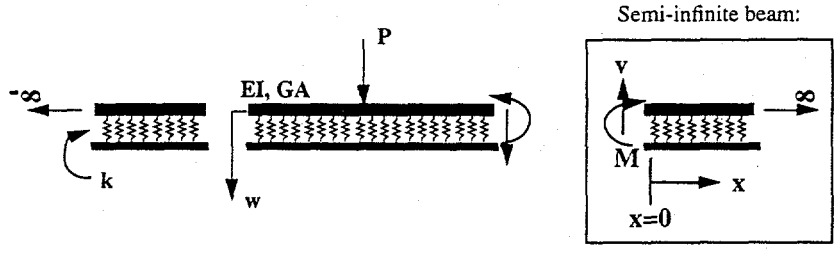


Figure 3: Finite and Semi-infinite Beams

The governing equations of the beam are (Fryba, 1972):

$$GA\,(w''(x) - \theta'(x)) - kw(x) = \rho\ddot{w}$$
$$EI\theta''(x) + GA\,(w(x) - \theta(x)) = 0 \qquad (1)$$

The general solution is obtained as:

$$w(x) = e^{(-\beta_1)x}\{c_1\cos(\beta_2 x) + c_2\sin(\beta_2 x)\}$$
$$\theta(x) = e^{(-\beta_1)x}\{d_1\cos(\beta_2 x) + d_2\sin(\beta_2 x)\} \quad (2)$$

In evaluating these equations, one must look at that finite part of the beam which encompasses all the inelastic deformation. In effect, the infinite beam is sectioned into one finite and two semi-infinite beams. The effects of the semi-infinite beams on the deformation in the finite section are taken into account with a 2x2 "boundary matrix," obtained from the general solution of the governing equations. This matrix adds into the global stiffness matrix. The resulting matrix for the right-hand side of the finite beam is given as equation (3).

$$\begin{bmatrix} V(0) \\ M(0) \end{bmatrix} = \frac{1}{2\beta^2 + \frac{k}{GA}} \begin{bmatrix} 2\beta_1 k & k \\ k & EI(4\beta_1\beta^2) \end{bmatrix} \begin{bmatrix} w(0) \\ \theta(0) \end{bmatrix} \quad (3)$$

where

$$4\beta^4 = \frac{k}{EI} \quad (4) \qquad \beta_1 = \beta\sqrt{1 + \frac{\beta^2 EI}{GA}} \quad (5) \qquad \beta_2 = \beta\sqrt{1 - \frac{\beta^2 EI}{GA}} \quad (6)$$

With minor changes, a similar matrix is used for the left-hand side of the finite beam. When implemented in the program, the boundary matrices yielded the exact answers for maximum displacement that were achieved when the program was run with a much longer, "infinite" beam.

For ease in viewing the results of the parameter study, the results were non-dimensionalized. An approach similar to that found in Fryba was used. Introducing a new independent variable,

$$s = \beta(x - ct) \quad (7)$$

the dimensionless displacement, v(s), is the displacement at a point divided by the displacement under the load, w_o.

$$v(s) = \frac{w(s)}{w_o} \quad (8)$$

The two parameters studied in this survey were velocity of the load and the yield force of the foundation. The velocity is measured by the dimensionless quantity α, defined by:

$$4\alpha^2 = \frac{c^2\rho}{\beta^2 EI} \quad (9)$$

The dimensionless foundation yield force, η_y, is given by

$$\eta_y = 4\frac{q}{kw_o} \quad (10)$$

Finally, the dimensionless load used was

$$P = \|v(s=0)\| = 8 \quad (12)$$

In the results presented here, it was assumed that the beam did not yield. Also, the beam was thin enough that shear forces had little effect on the system. In this case, β_1 and β_2 in equations (5) and (6) equal β.

RESULTS

A series of runs for an elastic beam with shear effects excluded is discussed in this section. To verify that the program yielded accurate results, the experimental value of deflection for a case in which the foundation never yielded at α=0.5 was compared to the theoretical maximum value (Fryba, 1972). The theoretical value was $1/(1-\alpha^2)^{0.5}$, or 1.1547. For the mesh size used (200 elements over a dimensionless length of 15), the model program yielded a value of 1.1549, for an error of 0.017%.

Several cases were studied in an effort to characterize the effects of load velocity and foundation plasticity. Case I studies the effects of foundation yield. It is the case of an elastic beam on which the foundation yield force was gradually lowered from its elastic value of η_y=4.0. Dynamic effects were neglected, with α=0. Thus, η_y became the only parameter in the system. As expected, at lower foundation yield values, which could be caused by fatigue of the ballast material, permanent displacements were recorded, with the depth of the permanent displacement increasing as the yield strength decreased. This pattern can be seen in Figure 4.

Similarly, the effects of speed were studied in Case II. This case holds η_y constant at its elastic value of 4, and varies α. As the velocity increased (Figure 5), the load which had caused no permanent deformation when dynamics were neglected resulted in residual displacements with inertia included.

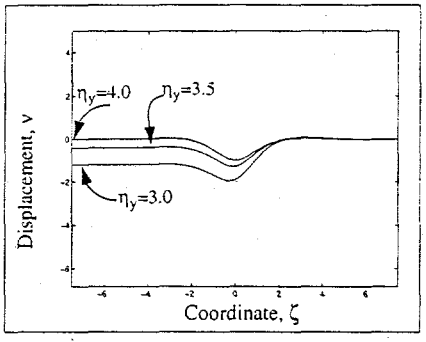


Figure 4: Case I: Deflection Profile for Elastic-Plastic Case Without Dynamic Effects (α=0.0)

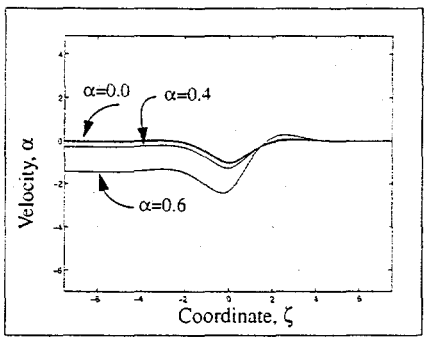


Figure 5: Case II: Deflection Profile for "Elastic" Case (η_y=4.0) With Dynamic Effects

Finally, in Case III, both η_y and α are varied. Plots of alpha at 30% and 50% of critical speed are shown as figures 6 and 7. It should be noted that α=1 represents the critical velocity for the beam (Fryba, 1972). The results of these studies showed that the permanent deflection increases as velocity increases and foundation yield decreases.

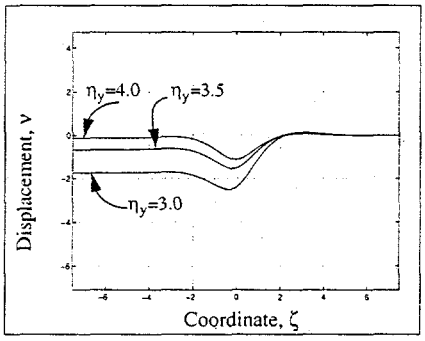


Figure 6: Case III: Deflection Profile for Elastic-Plastic Case With Dynamic Effects of α=0.3

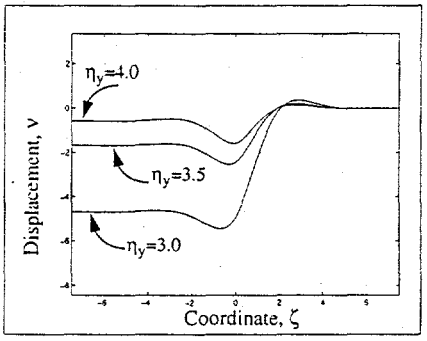


Figure 7: Case III: Deflection Profile for Elastic-Plastic Case With Dynamic Effects of α=0.5

An example of Case III might be the implementation of high-velocity train systems (increasing α) on existing fatigued rail systems (decreasing η_y). Figure 8, which is a comparison of both factors η_y and α, illustrates such an effect.

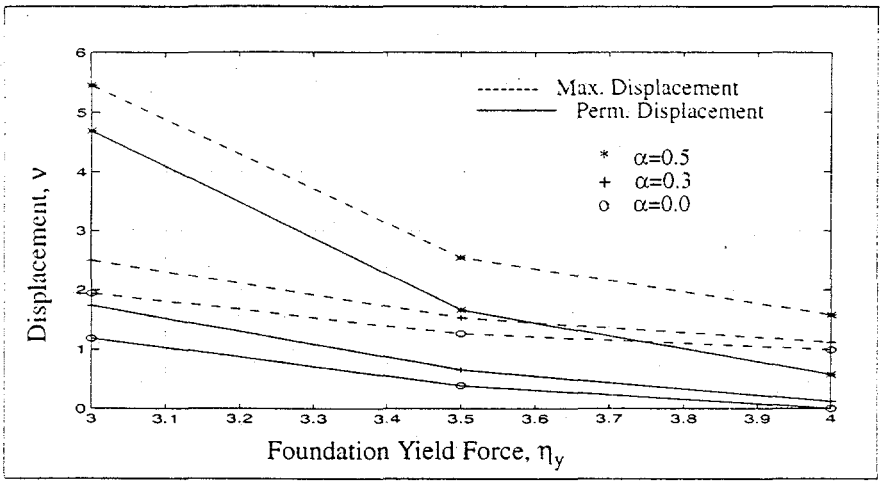


Figure 8: Effects of Yield Force and Velocity on Displacement.

Further study into the beam model will concentrate on adding viscous damping terms to the program and producing more detailed parameter plots. These plots could prove to be an invaluable design tool.

CONCLUSIONS

It is concluded that the simplified, one-dimensional model of a beam resting on a Winkler foundation subject to a moving load can be successfully used to model the deformations in many engineering systems. The following assumptions comprise the basis of the model: the beam and foundation are infinitely long; the load is moving at a constant velocity; the system has reached steady-state. Using these simplifications and special "boundary matrices," a relatively simple program has been created which models inelastic deformations in certain engineering systems.The reliability of the model has been established by showing that the linear-elastic cases and simple inelastic cases converge to calculated theoretical values. Results show the effects of velocity and foundation yield strength on an elastic Euler beam resting on an elasto-plastic foundation. Therefore, further refinements to this model will provide a simple, yet powerful tool which could be used in design to determine the effects of several parameters on the deformation in rail-track and pavement systems.

REFERENCES

Fryba, Ladislav. Vibration of Solids and Structures under Moving Loads. Noordhoof International Publishing. Groningen, The Netherlands. 1972.

Pro-Matlab User's Guide. *The Mathworks, Inc.* 1990.

Trochanis, A. M. et. al. "Unified approach for beams on elastic foundations under moving loads." Journal of Geotechnical Engineering. Vol. 113. August, 1987.

Architecture of an Operating System Specifically Designed for Research in Virtual Reality

Gregory Gerard

Department of Computer Science
Trinity University
715 Stadium Drive
San Antonio, TX 78212

Faculty Advisor:
Gerald Pitts

For the past forty years program design has been severely limited by the amount of power available. Recently this asphyxiation has been relieved by the development of ultra–fast hardware supporting limited excursions into virtual realities, but the infrastructure to take advantage of this renaissance does not yet exist. This research lays the groundwork for future operating environments by re–approaching operating systems, language syntax and semantics, and user interface design by gearing it towards virtual reality applications. This paper will illustrate the effectiveness of microkernel operating system approach to virtual reality development with a special emphasis on distributed computing in heterogeneous environments, shared virtual environments, and applications to conventional, non–virtual reality environments.

INTRODUCTION

Computing hardware capabilities have grown at an astounding rate since the birth of mechanical computation. What has not grown nearly as quickly are the interfaces to these engines. Something interesting, however, has occurred recently — high performance computing on the home desktop.

All aspects of performance computing are appearing on the desktop, ranging from three-dimensional graphics co-processors to *digital signal processing* (DSP) based sound systems. This opens a wide range of possibilities for human-computer interaction, most notably *virtual reality* (VR).

VIRTUAL REALITY

There are as many definitions for virtual reality as there are people discussing the subject. We shall adopt the following definition: an environment in which the level of interaction and absorption is indistinguishable from the real world. Note that this definition does not involve elaborate input devices such as data gloves or output devices like tactile feedback suits and stereoscopic headgear.

Virtual reality is not the substitution of one reality for another, but rather an extension. Consider the following proposition carefully: human sensation of reality is adaptive and is not fixed upon mere physical sensation. A quick check of many college computer labs would confirm this assertion. At any given time, it is not uncommon to find people participating in online games commonly known as *multi-user dungeons* (MUDs), attentions focused fully on the screen watching text scroll past and typing furiously on the keyboard. This is commonly done for hours at a time and in fact it can take some effort to get the attention of one of these people.

These players have simply adapted their realities for a brief period in time in lieu of something more interesting, namely online gaming. With extremely crude hardware, it is possible to achieve virtual reality. By no means shall the assertion be made that this is appropriate for all or many applications, but it makes the point that virtual reality is not a game for those elite with exotic hardware.

Virtual reality then is but a special case of human-computer interaction (HCI). With this in mind and with an eye towards the high-end hardware of popular virtual reality, it is instructive to examine issues of virtual reality and its implementation. Virtual reality is simply the terminus of HCI, where the computer is part of the user's world, invisible when not in use and unobtrusive when being used.

Core to virtual reality is immersion and interaction (Rheingold, 1991). It is

important to maintain a fluid raport between the human and machine. It is only in such an environment that immersion, the fading of external influences, might take place. Because of this many platforms must support the real-time computing constructs, especially if highly detailed simulations are required.

Human-computer interaction is important, but also of great import is human-human interaction. While humans most often interact outside of computer environments, when human-human communication is required inside a computing system, it is imperitive that such an interface be seamless. This sharing of experience should be transparent across heterogeneous platforms and networks. To be appropriate for mass consumption, this must be the case.

It is also important to reduce the modality of environments. For example, as a developer programs, there exist very distinct contexts: editing, running, debugging, etc. This violates the seamlessness virtual reality works so hard to achieve.

PROPOSAL

If virtual reality is merely the logical extension of HCI, then there is every reason to apply it to every part of the computing experience. At no time should the user experience an abrupt shift of thought when switching from reality and computing. It is therefore proposed that a new, wholly different platform of computing be created to support this idea.

It may seem needless to create yet another platform when countless others exist already. Upon further survey, there are several compelling reasons for doing so:

- many operating systems do not support the real-time computing (Silberschatz and Galvin, 1994).
- platforms support primitive networking functionality (Tanenbaum, 1995).
- parallel processing is minimal or non-existent (Tanenbaum, 1995).
- virtual reality is fundamentally object based (Benford, et al., 1994).
- large semantic gap between the operating system and applications.

It is logical to design a platform which would make up for such deficiencies.

DESIGN

Buddha (Gerard, 1995) is an operating system designed from the ground up to correct the errors of its predecessors and go beyond. It is an operating system for the future determined to be small, efficient, and extremely flexible.

It is the designer's belief that many of the touted features in modern operating systems are not necessary (Silberschatz and Galvin, 1994). At the core of this system is a tiny kernel taking care of only the most primitive concepts, building a framework for more robust environments.

Buddha adopts a position similar to that of early RISC designers (Tabak, 1987) — keep it simple, keep it fast. To this end, Buddha pushes many of the responsibilities once assumed by operating systems towards the user level. While this alone is not a unique feature — Mach, Amoeba, and other microkernels have been doing this for some time — Buddha does so in a unique way. First, there is no firm division between operating system code and user code. Second, there is no virtual memory or protection scheme implemented. Third, there is only minimal policy set.

By not setting up an artificial division between operating system code and user code, there is no performance loss caused by switching contexts. Context switching between operating system and user code is an expensive operation and is implemented because of a deep lack of trust on the part of operating system designers. All code under Buddha is trusted as correct. This places a great deal of stress upon language designers by causing them to do a great deal more error checking.

Actually, there are two potential routes a language designer could take. First, implement incredible amounts of error checking at compile time for traditional languages such as C, C++, and Pascal. Second, abandon these low level languages and design languages such as LISP, Scheme, Prolog, and Smalltalk where fatal bugs arise infrequently. These latter languages implement full memory management including garbage collection and are much better suited to analysis.

Because ultra-high level language usage is stressed, rogue pointers cannot develop and elaborate memory management is not necessary. From this, Buddha is able to cleave away overhead which has crippled many implementations (Gerard,

1995).

The only policy set by Buddha is the concept of object. The smallest unit of memory and task management is the object. The term object as used here is exactly the same as that developed in object-oriented design. Objects contain data and operations that can be performed on that data. Since this is implemented at the operating systems level and because these objects are extremely lightweight, they are to be the lingua franca among differing applications. Objects created in one language can be used in another.

This has a profound effect on design of an environment. Objects that the user directly manipulates through GUI gestures correspond directly to low-level system objects; there is no abrupt shift between environments. For example, observe a virtual reality system on current architectures. There is the development environment (typically X-Windows based), there is the operating systems environment (typically UNIX), and there is the actual virtual environment. In Buddha, they are all the same. The semantic gap between environments closes.

CONCLUSION

As of this writing, Buddha exists only as a specification. There is, however, an implementation under construction. It is only when a full implementation is completed that this project's design may be assessed.

REFERENCES

Steve Benford, John Bowers, Lennart E. Fahlen, John Mariani, and Tom Rodden. Supporting Cooperative Work in Virtual Environments. *The Computer Journal*, 37(8):653–668, 1994.

Gregory Gerard. *Buddha: A Distributed Object Operating System*. Thesis, 1995.

Howard Rheingold. *Virtual Reality*. Touchstone, New York, New York, 1991.

Abraham Silberschatz and Peter Galvin. *Operating Systems Concepts*. Addison-Wesley, Reading, Massachusetts, fourth edition, 1994.

Daniel Tabak. *RISC Architecture*. Research Studies Press, Letchworth, Hertforshire, England, 1987.

Andrew S. Tanenbaum. *Distributed Operating Systems*. Addison-Wesley, Reading, Massachusetts, 1995.

Feasibility of Object Oriented, Real Time Animated Rendering for Virtual Reality Applications

Scott McCaskill

Department of Computer Science
Trinity University
715 Stadium Drive
San Antonio, TX 78212

Faculty Advisor:

Dr. Gerald Pitts

INTRODUCTION

Although there exist numerous systems which are capable of real time animation of three dimensional objects, the majority are designed from a standpoint which is unnecessarily specific. Relatively few have been designed with anything more than the physical appearance of the objects in mind--that is, although they may be very good at drawing the objects, they often have no functionality whatsoever where the physical behavior of the objects is concerned. This fact is somewhat surprising, especially considering the often intimate relationship between an object's physical characteristics and the way in which it moves or interacts with other objects in a scene. So even though the method for drawing the individual frames of a bouncing ball may be very different from the way the bouncing motion is calculated, the two are nevertheless equally essential if the goal is to animate a bouncing ball.

This observation is especially notable when the objects to be animated are being used to create some sort of "virtual reality". In such a situation, especially if the viewer is also a participant in the animation (or simulation, more accurately), it is important that the objects have some sort of definable physical behavior. Put another way, although it is certainly possible to create a virtual world in which objects simply pass through one another without ever colliding, such a world would generally not be very useful for the purpose of modeling things in the real world, and would tend to be less interesting in any case.

These observations were the original motivation for doing the research represented by this thesis. The guiding philosophy behind the system described here is that the difference between the domain of the problem (objects in the real world, including their various physical and behavioral characteristics) and the domain of the solution (in this case, the actual implementation of such a system as will be described) should be as small as possible. Specifically, the following goals have been set:

- The system should be designed, from the beginning, with extensibility in mind. This is the most important of the specific goals. In the design of this system, whenever there arose a conflict between this goal and another (usually efficiency), the course dictated by this goal was almost always given precedence
- The system should be portable. The use of an object-oriented design and choice of a widely used (or at least widely available) language mean that making the system portable should pose no significant problems, and there is therefore no compelling reason not to design with portability in mind.
- The system should be efficient. This specification is related to the previous one; the expectation is that the system will be able to run on a number of different platforms, possibly of greatly varying capabilities. In an effort to keep the frame rate tolerably high, each platform will trade speed for increased detail or detail for increased speed, depending on how "fast" the particular machine is.

The goal of this research is the design of a system which is capable of doing real-time animation of three dimensional objects. As the title implies, this system is designed with virtual reality applications in mind (although not exclusively), and from the beginning an object-oriented approach has been used. The language chosen for the implementation is C++.

WHY OBJECT-ORIENTED?

In order to understand why this system is designed the way it is, one must understand the philosophy behind, and capabilities of, an object-oriented design. In procedural languages such as C, a program is specified entirely in terms of functions and data. An object oriented language

augments or replaces (depending on the language) these two constructs with a third construct, the object[1]. Objects are essentially aggregations of functions and data. In C++, for example, a class is really just a structure which may have particular functions associated with it as well as particular data. Although this may not sound very significant, some very powerful ramifications can result from this aggregation. Whereas a typical procedural program is concerned with the flow of data, an object-oriented program is concerned with the interactions of various objects--and this is probably the most important difference between object-oriented design and procedural design, since objects can provide much closer analogues to things in the real world than functions and data by themselves. This feature is especially pertinent in a system such as this, which should be able to model things from the real world. Another important feature which most object-oriented languages provide is inheritance. Using inheritance, it is possible to say that one class "is a kind of" another class. For example, an apple "is a kind of" fruit, a car "is a kind of" vehicle, and a line "is a kind of" drawing primitive. In this manner, inheritance can be used to guarantee that a derived class (a class which inherits from another class) can do whatever its parent class can do, and possibly more. An inheritance hierarchy can have arbitrarily many levels, and some languages, such as C++, also allow multiple inheritance, which is the ability of a class to inherit directly from more than one class (example: a seaplane is both a boat and a plane).

THE DESIGN OF THE SYSTEM

At the very base of the class hierarchy for the system are five fundamental classes: Entity, View, Canvas, Shape, and Control:

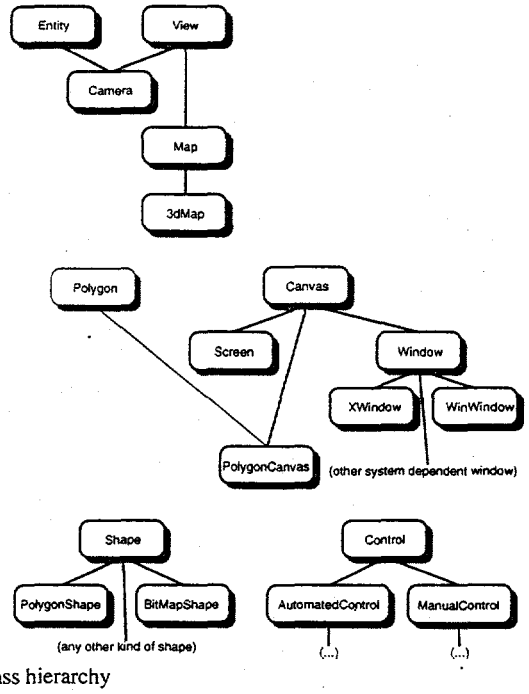


Figure 1: Basic class hierarchy

These five classes are shown here, along with several other classes which may be used in the system. An Entity is anything that can be considered an element of the virtual world--moving, non-moving, visible, invisible, a tree, a polygon based object, a person, or a camera for example. Entity contains only information that would be common to any such object. Entity is not meant to be used directly, though, only derived from. In this way, the Entity class acts as an interface by defining the standard capabilities of that any kind (sub class) of Entity will be able to provide. In this manner, all clients (users) of Entity objects are able to know what these basic capabilities are (although not how they are implemented) regardless of the particular sub class of Entity that is actually being dealt with (since any instance (object) of a sub class can also be

[1] An object, it should be noted, is simply an instance of a class. Classes are to objects what data types are to data and function prototypes are to functions in procedural languages.

regarded as an instance of its parent class(es)). The other four fundamental classes play similar roles for their descendants as well.

An Entity contains information about its location and orientation (which are points specified in the world's coordinate system), a Control object which is responsible for changing the Entity's location and orientation, and a Shape object (which may itself be composed of multiple Shapes) which determines the Entity's physical representation. Additionally, an Entity may be composed of other Entities, each with its own location, orientation, Control, Shape(s), and Entity(ies). The coordinates of any given Entity are specified relative to its parent; and in the case that the Entity is at the top of its "tree" the coordinates are assumed to be world coordinates (there is also a World class, which would function as a sort of parent for an Entity at the top of its tree, but it has not been decided at this time whether it should be derived from Entity).

The ability of an Entity to be composed of multiple other Entities in a tree-like structure is particularly interesting. Following is an example of a partial hierarchy of an Entity which is intended to model a person:

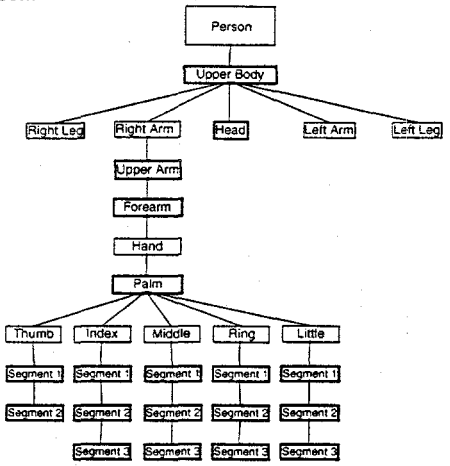


Figure 2: Example Entity hierarchy

In this diagram, each box is an Entity, and dark boxes represent Entities which have a Shape associated with them (remember, there is no requirement that any given Entity have a shape--if it has no Shape, then it will most likely be assumed to be invisible). The particular layout of this tree is significant: the position of any given Entity in the tree is indicative of what it is "attached" to, and therefore how it will move when what it is attached to moves. For example, if the right arm doesn't move, but the upper body does, then relative to the world (which is the view we will most likely be seeing), the right arm still moves. Put another way, if you hold your arm stiffly out in front of yourself and then, without moving the muscles in your arm or shoulder, turn your body so that you are facing what was to your left, then you have not moved your arm relative to your body, but your arm has still moved relative to the rest of the world.

The Shape class is the base class for any type of Shape (like Entity, it is not intended to be used directly, but to be derived from). A Shape is the physical representation of an Entity; it possesses qualities such as size and appearance. Descendants of Shape are not required to use any particular representation (such as wireframe, polygons, scaled bitmaps, etc.), and very few assumptions are made about the actual physical representation of an Entity. Descendants of Shape need only be able to transform themselves (using a single transformation matrix, which may include rotations, translations, scalings, and projections) and draw themselves on a Canvas. In figure 1 above, the example descendants of Shape which are shown are a polygon based shape and a bitmap based shape (a bitmap based shape would consist of several different views of an object from several different angles which can then be scaled depending on how far away the object is, for the effect of a perspective projection). These are only two of the many possibilities. Note also that since a Shape may also consist of multiple "subShapes", any given Shape could easily be a combination of different kinds of Shapes. As in Entity, the coordinates of each "subShape" are assumed to be relative to the "parent" Shape. A Shape must know something about its dimensions (regardless of its particular representation). At the least, it must know the radius of the smallest sphere which will enclose all of itself. Later, when the design of

Shape becomes more advanced, Shapes may be required to know more specific information concerning their dimensions.

The purpose of a View is to specify a way of looking at a world. At the present time, the types of Views which are expected to be the most useful are cameras (which typically, but not necessarily, would implement a perspective projection) and map views. It is the responsibility of a View to determine which Entities in the world are visible, calculate the transformation matrix to transform those Entities however the View sees fit, and then give that transformation matrix to those Entities which are visible and tell them to draw themselves on a particular Canvas. Since homogenous (four dimensional) coordinates are used, it is possible to specify a perspective transformation with a 4 x 4 matrix. In figure 1 above, it should be noted that Camera inherits from both Entity and View. This is not a requirement, but rather an example of how multiple inheritance may sometimes be useful. There will have to be some sort of View in the virtual world, and a camera (a View which implements a perspective projection, in this case) seems like a perfectly good metaphor for that View. Additionally, we find no problems with the idea of a camera having a position in the world and the ability to move around in that world, and so it also seems reasonable that a Camera should be a kind of Entity (an element of the virtual world) as well as a kind of View.

A Canvas is anything that can be drawn on. Typically, this will be either a window (in a graphical user interface) or perhaps the entire screen. However, as with the previous fundamental classes, there can just as easily be other types of Canvases. The Canvas model is as follows: for each frame, all drawing is done to an off-screen page. When the frame is finished being drawn, the Canvas is told to display the next page. The exact implementation of this model varies depending on the particular hardware setup. For example, on an IBM compatible PC with a VGA display, there are basically two possibilities. One is that a video mode which actually supports page flipping[2] is used, and the other is that the drawing is done to an area of system memory which is then rapidly copied to video memory. A Canvas will also be required to implement a z buffer algorithm (this may be integrated into the base class). This is due to the fact that other methods of rendering graphical primitives to the screen (such as the painter's algorithm for drawing polygons, for example) must know ahead of time every kind of primitive that can be drawn. However, this would violate the flexibility offered by encapsulating the physical representation of an object within a Shape, because it would require Canvas to know about every possible kind of Shape. If we require this, then every time we want to add a new kind of Shape to the system, we must make sure that the Canvas knows about it as well (and in detail, at that) so that the Canvas will know how to draw it. This is tantamount to expecting a painter's canvas and paintbrush to "know" what the painter is capable of painting, which of course is quite nonsensical.

In figure 1 above, one particularly interesting possibility is the combination of a Canvas and a Polygon. This would be useful if you wanted to create a virtual world with a television in it, or, even more interesting, you could model a closed-circuit TV system, where Views of various parts of the virtual world (Cameras) could project onto TVs (PolygonCanvases), all of which would be Entities in the virtual world. This type of concise expression is indicative of the expressive power provided by an object-oriented language when it is properly used.

Finally, there is the Control class. Control is the base of a potentially large class hierarchy, since a Control is anything that affects the movement or behavior of an Entity. The metaphor for Control is one of the driver of a car, the captain of a ship, or the pilot of an airplane. Although the potential uses of descendants of Control (since, like the other fundamental classes, Control is meant to be derived from) have not been thoroughly explored at this point, it seems likely that Controls will typically contain such things as velocity and delta orientation for an Entity, as well as routines for getting input from various input devices (keyboard, mouse, joystick, etc.) and/or routines for determining the automated movement of Entities. It is foreseeable that a descendant of Control could be designed to make use of an animation scripting language to determine the automated movement of an Entity. It is also conceivable that, if multiple Entities are allowed to use the same Control, their movements would all be identical (since Control won't necessarily know anything about where an Entity is, only how it is moving), which might make for an interesting effect.

PROBLEMS ENCOUNTERED DURING IMPLEMENTATION

[2] With a VGA display, each pixel on the screen corresponds to a byte in the VGA's memory. Since some modes do not require the entire VGA memory to display one screenful of pixels, page flipping is implemented (in those modes that support it) by simply changing the memory address from which the screen starts displaying.

Although the design has gone through many changes during its evolution, the basic elements are now in place. In developing the system, some of the most important questions which had to be dealt with concerned the ability of the target language to support the flexibility required to implement much of the object-oriented functionality. Ultimately, nearly all of these problems stem from the fact that in C++, as in C, symbols are bound to their types at compile time, rather than at run time, as in SmallTalk or Objective-C. Although this allows the compiled code to be much more efficient, it often comes at the cost of flexibility. This was the primary reason why other languages besides C++ (mainly Objective-C) were considered before any significant implementation work was begun on the system. However, after much research into the C++ language (including considerable experimentation), it has been determined that, if used well, C++ should be adequate for the job. Objective-C, while more flexible and syntactically more simple, was determined to be too slow for this application (it is estimated that this system would be between one and a half and two times slower if implemented in Objective-C instead of C++). Although it is presently too early in the implementation stage of the project to determine exactly how fast the system will be, it is estimated that this will depend more than anything else on the speed of the z buffer which will be used in most (if not all) descendants of Canvas. Although z buffer methods are computationally intensive and can require significant amounts of memory, they will be necessary if the variety of possible Shapes and independence of the five fundamental classes are to be preserved. It is expected, however, that in the future hardware z buffer assistance will become more commonplace (it is in fact already available today, although not necessarily commonplace).

There were a multitude of other design questions as well. For example, in polygon based shapes, how should the relationship between polygons and the vertices which define them be designed? One possibility is to store the vertices in the polygons (as opposed to the shape). This approach would give each polygon it's own set of vertices, and no sharing of vertices would occur. Another approach is to store an array of vertices in the shape, and in each polygon store an array of integers which are indexes into this vertex array. A third option, similar to the second option, is to store the list of vertices in the shape and have an array of pointers to vertices in each polygon. Each approach has its advantages and disadvantages: the first is the most flexible and the slowest, the second is the fastest and least flexible, and the third is somewhat of a compromise between the two. These are the types of questions which must be carefully resolved in order to find the best balance between efficiency and flexibility.

CONCLUSION

This research makes a convincing case for the importance of design in the development of software systems, and has helped show to what extent design directly affects the capabilities of such systems. In this particular case, it has been determined that a system such as the one described herein can be made to have the desired extensibility, portability, and efficiency, even when implemented in a language as strongly typed as C++.

WORKS CONSULTED

Coplien, James. Advanced C++ Programming Styles and Idioms. Addison-Wesley, 1992. ISBN 0-201-54855-0

Cox, Brad J. Object-Oriented Programming - an Evolutionary Approach. 1st ed. Addison-Wesley, 1986. ISBN 0-201-10393-1.

Cunningham, Steve, Nancy Craighill, Martin Fong, Judith R. Brown, eds. Computer Graphics Using Object-Oriented Programming. John Wiley & Sons, 1991.

Foley, James, Andries van Dam, Steven Feiner, John Hughes. Computer Graphics: Principles and Practice. 2nd ed. Addison-Wesley, 1990. ISBN 0-201-12110-7.

Halladay, Steve, Michael Wiebel. Object-Oriented Software Engineering. R&D Publications, Inc., 1993. ISBN 0-13-034489-3.

Hearn, Donald, M. Pauline Baker. Computer Graphics. 2nd ed. Prentice Hall, 1994. ISBN 0-13-161530-0.

Pokorny, Cornel. Computer Graphics: An Object-Oriented Approach To the Art and Science. Franklin, Beedle & Associates Incorporated, 1994. ISBN 0-938661-55-8.

Rumbaugh, James, Michael Blaha, William Premerlani, Frederick Eddy, William Lorensen. Object-Oriented Modeling and Design. New Jersey: Prentice Hall, 1991.

Vince, John. 3-D Computer Animation. Addison-Wesley, 1992. ISBN 0-201-62756-6.

Weiskamp, Keith, Bryan Flamig. The Complete C++ Primer. 2nd ed. Academic Press, Inc., 1992. ISBN 0-12-742686-8.

Image Processing Using Fuzzy Logic

Nakesha L. Gully

Computer Science Department
Xavier University of Louisiana
7235 Palmetto Street
New Orleans, LA 70125

Faculty Advisors:

Marguerite S. Giguette (Xavier University), Antonio M. Lopez, Jr. (Loyola University; Department of Mathematics and Computer Science; New Orleans, LA), and Katherine Henson-Mack (University of Alabama; Department of Computer Science; Tuscaloosa, AL)

This paper presents the use of fuzzy logic in image processing. Specifically, it discusses how image processing and fuzzy logic can be used to detect roads and rivers in aerial photographs. The detection of features in aerial photographs is an important area of research. This detection is used in urban planning, boundary clarification, and in the defense industry. The process of feature detection in this research includes the use of a digitizing scanner, a color editor, histograms, fuzzy logic, and computer programming. Fuzzy logic is used in situations where there is not a definite answer. The use of fuzzy logic is well suited in determining rivers and roads in an aerial photograph since the gray scale value for a road or river is not definitive. Fuzzy logic is used to break down the gray scale of the aerial photographs into a pixel range which represents the range from black to bright white. Once the aerial photograph is scanned digitally, a color editor is used to determine the possible pixel ranges for roads and rivers. A computer program is then written to form histograms of the pixels used in the photograph. These histograms help determine definite pixel ranges which are considered a unique identifier of a river or road.

INTRODUCTION

This research combines two paradigms -- image processing and fuzzy logic. It uses these two paradigms to detect rivers and roads in a region of Alabama. Aerial photographs of varying topographical and geographical areas were used as test data. River and road detection involves manually identifying specific pixels which are observed as part of a river or a road. From there, other homogeneous pixel areas can be identified as part of the river or the road. Fuzzy logic can then be applied to identify non-homogeneous pixels that are also parts of rivers and roads. This paper describes the procedures used to incorporate fuzzy logic in image processing to identify rivers and roads in this test data.

Image Processing

Image processing is used in a number of applications to identify different regions in various types of images (satellite, sensors, photographs). These areas of application include urban planning, defense projects, and political science. Image processing is currently being used in Third World countries for urban planning by observing areas from satellite images for possible spatial expansion [4]. One application of image processing in the defense industry is to send target coordinates from satellite images to an aircraft for strategic action [3]. Image processing can also be used to locate regions in a state and clarify the division of a state's voting district [5].

Image processing research may be divided into the areas of calibration, analysis, and alteration [2]. This paper is on analysis and alteration. Calibration involves producing a "pure record" of the intensity of light to the image. In the analysis stage, the aerial photographs are observed so that rivers and roads can be manually identified by their pixel values. Fuzzy logic is used in the alteration stage by taking the pixel values and breaking down the gray scale to colors ranging from black to bright white. By breaking down the color scale, the image can be observed in its entirety and distinguish any other possible features to detect rivers and roads. The image is further enhanced by histogram scaling or specifically, histogram equalization which adds brightness and darkness to the image.

Fuzzy Logic

Fuzzy logic is a concept that extends binary logic [1]. Binary logic uses a series of one's (1) and zeros (0) to represent the operation of a computer system, where zero represents off and one represents on. Fuzzy logic recognizes a medium between these two extremes. Not only can a situation be considered true or false, but it can also be "mostly true" or "partly false". Thus the name "fuzzy" logic evolved since a "crisp" value often does not exist to represent a given situation. For example, the set of "men" is binary, but the set of "tall men" is fuzzy since the concept of "tall" is relative to a point of comparison. Specification of ranges of values for non-homogeneous points that might be a river or a road is fuzzy, but not binary. Hence, this "fuzzy logic" is used to enhance the images used in this research.

ALGORITHM APPLICATION

In this project, five aerial photographs are taken of two cities in Alabama and scanned on a digitized computer. These two cities, Demopolis and Greensboro, are used because they have varying topographical surfaces. These photographs are digitally scanned to produce gray level images. Since the digital images are represented by pixel values ranging from zero to two-hundred fifty-five, they are easy to manipulate. A user needs only to increase/decrease the numerical value represented in order to intensify or subdue the color in that image. By intensifying the color in an image, one is able to eliminate some of its "fuzziness."

The actual process of analysis begins with experimentation on the digitized images. It uses the XV3 software package. This package, using X-windows, allows the user to distort, change and create new images using various functions. A color editor is used to highlight and darken the images with various colors to observe the changes that occur. The goal at this point in the research is to determine if there is a pixel range which can be considered as the "unique" identifier of a river or road. For example, if pixel range 60 - 75 is a "unique" identifier of a river, then a road should not be detected within these pixel values. To test the program's detection of the most obvious points on the image, rivers and roads are noted manually without the aid of the computer in order to create a control group. After experimenting with the editor, a color-code legend is made for each map showing where the color changes take place.

Even though the editor noted color changes in the river and road areas, there is still no method to determine a river or road without involving interfering factors such as farm land. Therefore, histograms must be created. These bar graphs show the pixel range (0 - 255) and the number of pixels represented at each value. In order to create the histograms, Pascal algorithms are written to read in the pixel values, place them into a matrix, and create the histograms. An output file is created for each map. Figure 1 provides the Pascal program segment that does this.

```
for i := 1 to 557 do
    for j := 1 to 615 do
      begin
          read(infile,image[i,j]);
          hist[image[i,j]] := hist[image[i,j]] + 1;
        end;
for x := 0 to 255 do
    begin
        writeln(outfile,x,' ',hist[x]);
        writeln(x,' ',hist[x]);
    end;
```

Figure 1. Read in the pixel values represented in the image and create the histograms

After the Pascal programs have stored the data, a software package called XMGR is used to view the histograms. This package allows the histograms to be overlapped to compare the different pixel values for each map. The histograms are viewed, and the number of pixels represented for each pixel value (0 - 255) can be seen. Figure 2 displays an example histogram that has had its high values connected by a continuous line.

Most of the histograms have a peak number around one-hundred (100); this number has the highest frequency in the map image. Since one-hundred is "almost" half way between 0 and 255, it signifies that the peak level is in the middle gray area. Had this image been only in

black and white, the gray levels would not be shown. Therefore, the peak value (125) shown above would not be significant. This further emphasizes the importance of fuzzy logic to display the range of color values since "almost" is a relative term.

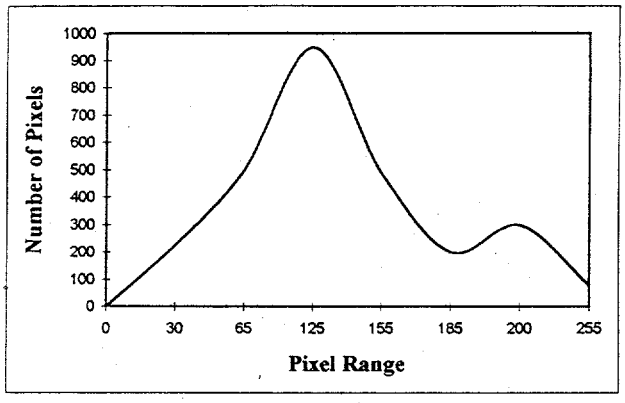


Figure 2: Example histogram that illustrates pixel frequency.

The next step is to determine the fuzzy estimated value (FEV) of each histogram. This value represents the "most typical" gray level of the image and the peak value of the histogram. The algorithm involves taking the minimum value of either the pixel value over the number of gray levels or the frequency of the pixel value divided by the total matrix number. This is so that the FEV is computed by an average of all the gray levels. After the minimum value for each pixel number is taken, the highest number of these is multiplied to the number of gray levels. This value is the FEV. Figure 3 is the Pascal segment that computes the FEV.

```
t = pixel value
maxhist = total number of gray levels (255)
sub_t = frequency of the pixel value
sub_o = total matrix number
_____________________________________

exp_val := 0;
sub_t := 0;
sub_o := 557 * 615 (matrix coordinates)

for t := 0 to maxhist do
 begin
   sub_t := 0;
    for i := 1 to 557 do
          for j := 1 to 615 do
            if t <= image[i,j] then sub_t := sub_t + 1;
    m := min(t/maxhist, sub_t/sub_o);
    if m > exp_val then exp_val := m;
 end;

( REMAP THE GRAY LEVEL )

FEV := exp_val * maxhist;
writeln(FEV);
```

Figure 3. Computation of the Fuzzy Estimated Value (FEV).

After the FEV is found for each map, a standard is calculated for the histograms by finding the average FEV of all maps. This is done by adding the FEV of all five images together and

dividing by five. This standard is used to shift the histograms. That is, the FEV of each histogram is shifted from its original position to match the standard of the five images. Again a Pascal algorithm is used to accomplish this (Figure 4).

```
histtype = array[0..255] of integer;
hist: histtype;
_____________________________________
Rewrite(outfile,'NAME = demo11hist.pgm');

  sumhist := 0;
  For x := 0 to 19 do
      begin
        sumhist := sumhist + hist[x];
        hist[19] := sumhist;
      end;

  For j := 0 to 236 do
      begin
        writeln(outfile,j,hist[j+19]);
        writeln(j,' ',hist[j+19]);
      end;

Close(outfile);
```

Figure 4. Pascal program segment used to shift the histograms.

The histograms are shifted so that the colors in the images use the same color scale. This provides the same basic scale to compare the pixel range of all the images. Furthermore, by shifting the scale, the new pixel range can be compared with the old pixel range to see if the same river or road can be detected in the new pixel range. In most cases, the pixel range is shortened. For example, one map was shifted nineteen places to the left making the pixel range no longer 0 to 255, but 0 to 236. The first nineteen pixels are added to the 20th pixel value so that the entire histogram can still be seen on the computer screen. The histograms are viewed on the XMGR software with the histograms of the original and new being overlapped to show the actual shift.

By shifting the histograms to the same position, the gray levels of the five images are the same. Originally, when the photographs were taken, some came out darker or lighter in certain areas. Therefore, some rivers and roads were darker or lighter than others. This made it difficult to determine a number range to designate a river or a road. When the gray levels of the images are identical, the identification is made easier. Figure 5 is a Pascal segment that writes the output files of shifted histograms.

Additional output files are created displaying the new pixel values of the image. The pixel values have changed to match the shift in the histogram. For example, the histogram that has been shifted nineteen places now contains pixel values with a nineteen point difference. This means that if the original value is 255, it is now 236. If the original value is 219, it is now 200 and so on. In this case, the image is made darker due to the decrease in the value of the pixel number. The images now match the color of the standard (average). Most images are made a little brighter due to the standard. A number range can be determined for the rivers and roads now that each image is almost the same color. However, it is also important to eliminate any interfering factors in the image such as the farm land. This is to make sure that it is not confused with the rivers and roads.

To accomplish this, the color editor on the XV3 is used to try to highlight the same areas in the color scale chosen. This time, however, the pixel values are displayed as slightly higher or lower than before. An application is used to detect lines and eliminate the farm areas. In most cases, when farm land terrain is photographed, the nature of the land is such that the farm areas' nearby pixel values change more than that of a river or a road. Therefore, the pixels that represent farm land do not form a line, but the pixels for rivers and roads are more likely to form a line. Using the formula $y = ax + b$, the algorithm determines whether connecting pixels form a line. If they do not, those pixels are eliminated from the image thus eliminating farm land. As a result, based on the comparison of the original colored areas to

their complimentary (shifted) pixel values, an area is designated as either a river or a road.

```
For i := 1 to 557 do
 for j := 1 to 615 do
  begin
    if image[i,j] < 19 then
          image[i,j] := image[i,j]
   else
     image[i,j] := image[i,j] - 19;
  end;

writeln;writeln;

For i := 1 to 557 do
 begin
   for j := 1 to 615 do
   begin
          if image[i,j] mod 5 = 0 then
                begin
                   write(outfile,image[i,j]:4);
                   writeln(outfile);
        end
     else
     write(outfile,image[i,j]:4);
   end;
  writeln;
  writeln(outfile);
 end;

Close(outfile);
```

Figure 5. Pascal program segment that writes the shifted histograms.

CONCLUSION

This research was performed using multiple paradigms: fuzzy logic, procedural Pascal algorithms, mathematical manipulation of histograms and visual displays. By combining these multi-paradigms, software can be used to detect definitive pixel ranges that identify rivers and roads in an aerial photograph. The research can be expanded to include detection based on the geometric patterns of the features in the photograph. This process would involve the use of mathematical models based on these features.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation grant CDA-930071 and the Center for Undergraduate Research in Multi-Paradigm Design at Loyola University in New Orleans, LA, the National Science Foundation grant CDA-9322010 at the University of Alabama, Tuscaloosa, and the GE Fund's Faculty for the Future Program.

REFERENCES

1. Barron, Janet J., "Putting Fuzzy Logic into Focus", Byte, April 1993, p. 111-118.
2. Berry, Richard, "Image Processing in Astronomy", Sky and Telescope, April 1994, p. 30-36.
3. Hughes, David, "Imaging Advances Leading to Real Time Targeting", Aviation Week & Space Technology, Vol. 140, 1994, p. 46,63-64.
4. Moller-Jensen, L., "Knowledge-based Classification of Urban Area Using Texture and Context Information in Landsat-TM Imagery", Photogrammetric Engineering and Remote Sensing, Vol. 56, 1990, p. 899-904.
5. American Society of Photogrammetry and Remote Sensing. (1983) ACSM-ASP Convention Technical Papers 49th Annual Meeting. Washington, D.C. March 13-18.

An Implementation of Objects and Object-Oriented Programming Using PROLOG

Richard A. Butler and William J. Dean

Computer Science Department
Xavier University of Louisiana
New Orleans, LA 70125

Computer Science Department
Tulane University
New Orleans, LA 70118

Faculty Advisors:

Antonio M. Lopez, Jr. (Loyola University; Department of Mathematics and Computer Science; New Orleans, LA 70118) and Marguerite S. Giguette (Xavier University)

Object-oriented programming (OOP) has proven to be an effective programming paradigm. Because of its simplification of entities and their operations, OOP has been utilized in various design applications, in particular, database representations. However, the retrieval of data from large databases seems to be tedious for database designs implemented in the programming language Smalltalk. This is due in part to Smalltalk's large amount of support and slow searching strategies. On the other hand, PROLOG, with its multiple indexing techniques and search strategies, may increase performance in such retrievals. A prototype system that combines an object-oriented database design with a logic-oriented implementation results in a faster and a more widely applicable programming system. This paper presents such a prototype in PROPS, a PROLOG Object Programming System.

INTRODUCTION

In the past decade, object-oriented programming has taken center stage as a leading programming paradigm. The use of object-oriented techniques has proven effective in a number of applications, due to its simplifying of entities and their operations [1]. In the object-oriented paradigm all entities, both concrete and abstract, are represented as objects. Each object has a set of operations called methods which change, access, or utilize the data and attributes of that particular object.

Object-oriented methods are being used in databases to represent data items and their associated attributes as objects [2]. Smalltalk is a programming language that is object-oriented in nature, and it is useful in describing and implementing a database. However, large databases are problematic for the slow searching methods used by Smalltalk. On the other hand, PROLOG (PROgramming in LOGic) is a declarative, logic-oriented programming language, consisting of facts, rules and queries [5]. It provides the programmer with fast searching techniques due to its internal method of indices. In addition, PROLOG has a connection with relational databases [4]. These facts have spurred the current research and development of the PROLOG Object Programming System (PROPS). This paper presents detailed information about PROPS and its database capabilities.

PROPS: PROLOG OBJECT PROGRAMMING SYSTEM

The inspiration for the development of PROPS stems from a system developed by Zaniolo [7]. His paper emphasizes the major concepts of the object-oriented programming paradigm and how these concepts can be represented in PROLOG. Zaniolo's ideas and explanations were further expanded, and PROPS was developed (Figure 1).

PROPS consists of representations of objects, classes, attributes, methods and values. The concept of using a separate logical engine of rules to represent object instantiation and hierarchies was drawn from Zaniolo's research and coupled with a frame-based representation of objects and their components. Frames structure how an object relates to other objects through attributes, or slots [6]. The following examples described in this section are taken from a simple bank account program, which was written in Smalltalk, and converted into PROLOG code. The program will be discussed in detail in a later section.

In PROPS, the programmer defines objects, classes and instances using the predicates *is_class(<obj_name>)* and *is_instance(<obj_instance>, <obj_name>)*. Attribute variables for a specific class are defined using the predicate *has(<obj_name>, <attribute>)*, and values are assigned to attribute variables using *has_value(<obj_name>, <attribute>, <value>)*.

```
:- dynamic has_value/3.
:- dynamic has/2.
:- dynamic is_instance/2.
:- dynamic is_class/1.
:- dynamic kind_of/2.

has(Name,Val) :-
is_instance(Name,Obj),
has(Obj,Val).

has(Obj,Val) :-
kind_of(Obj,Parent),
has(Parent,Val).

value(Name,Var,Val) :-
(has_value(Name,Var,Val),!);
(is_instance(Name,Obj),
value(Obj,Var,Val),!).

value(Obj,Var,Val) :-
(has_value(Obj,Var,Val),!);
(kind_of(Obj,Parent),
value(Parent,Var,Val),!).

has_method(Name,Method) :-
(is_method(Name,Method),!);
(is_instance(Name,Obj),
has_method(Obj,Method),!).

has_method(Obj,Method) :-
(is_method(Obj,Method),!);
(kind_of(Obj,Parent),
has_method(Parent,Method),!)
.

do_method(Obj,Method,Param
eter):-
has_method(Obj,Method),
method(Obj,Method,Parameter
).
```

Figure 1. The PROPS Logical Engine

Methods are declared and defined using the *is_method* and *method* predicates, respectively:

is_method(<class_name>, <method_name>).

method(<obj_or_class_name>, <method_name>, <parameter>).

The *kind_of(<subclass>, <superclass>)* predicate provides the means by which the PROPS system traverses class hierarchies and implements inheritance of attributes and methods. For example, the PROLOG fact in the knowledge base *kind_of(savings_acct, bank_acct)* exhibits a hierarchical structure. Because *savings_acct* is a subclass of *bank_acct*, it inherits the same attributes of *name*, *address* and *balance*. Figure 2 illustrates an example of inheritance.

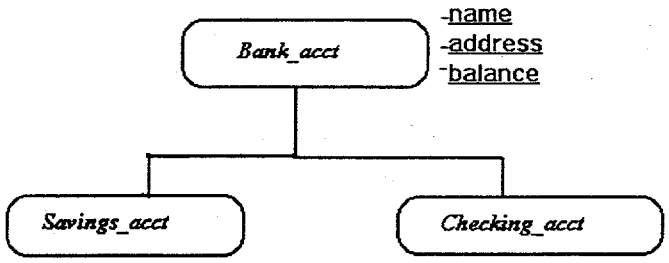


Figure 2. Inheritance Example

Inheritance of instance variables is implemented with the *has* predicate. In addition to the *has* predicates entered into the knowledge base by the user, the engine contains two *has* rules that recursively search through superclasses and classes of instances to find the closest parent for which the *has(<name_or_obj>, <value>)* statement is declared. Upon finding an ancestral object that has the instance variable, the *has* predicate succeeds. For example, the user could query with the statement:

?-has(savings_acct, balance).

The engine would first check if the above predicate exists in the knowledge base. If not, the PROLOG interpreter would utilize the engine's *has* rules to check if the object or class in question is an instance or subclass of a class that has that particular variable declared. This search proceeds recursively up the hierarchical path until either the top class is reached or a class is found to have the variable in question. If the predicate is not found in the knowledge base, the *has* rule is executed. The engine finds that *savings_acct* is a subclass of *bank_acct*. Due to inheritance, the system prints *yes* as the answer. This strategy is also used to allow the inheritance of default values for instance variables.

For the implementation of method inheritance, an altered approach was required. Using the above methods of inheritance causes the system to perform the called method, upon whichever class the method is defined, even if the user sends the message to a subclass of the class owning the method. This approach is carried out by using the *do_method* rule:

do_method(<obj> , <method>, < parameter>) :-

has_method(<obj>, <method>), method(<obj>,<method>,<parameter>).

It is assumed that both the message sender and the receiving object know the parameters of the message and their order. Upon receiving this query or interpreting this message in another

method, the engine will first determine whether the receiving object has that method. It searches for the *is_method(<obj>, <method>)* predicate that corresponds with the specifications in the message. If it does not find the predicate, the engine checks through parent classes, using the same method as the *has* and *value* predicates. If it is eventually determined that the receiver has or inherits the method, the engine uses the *method* predicate to finally send the message to the receiving object.

The implementation of the logical engine is quite simple in nature, and it allows for rapid prototyping. This paper provides a more detailed example of the bank account system utilizing the PROPS engine.

BANK ACCOUNTS: A PROPS EXAMPLE

In Figures 3a and 3b, sample knowledge bases are given for two classes: *bank_acct* and *savings_acct*. This example was selected from [3] as a typical object-oriented problem. Each object is represented by its class declaration, a set of instance variables, inheritance predicates and a set of method declarations and definitions.

```
is_class(bank_acct).
 has(bank_acct,name).
 has(bank_acct,address).
 has(bank_acct,balance).

is_method(bank_acct, new).
method(bank_acct,new,Id) :-
 asserta(is_instance(Id,bank_acct)),
 do_method(Id,initialize,_).

is_method(bank_acct,open).
method(bank_acct,open,Id) :-
 do_method(bank_acct,new,Id).

is_method(bank_acct,initialize).
method(Bank_acct,initialize,_) :-
 write('Give the owner of the acct:'),
 read(Name), nl,
 asserta(has_value(Bank_acct,name,Name)),
 write('Give the address of the owner:'),
 read(Address), nl,
 asserta(has_value(Bank_acct,address,Address)),
 asserta(has_value(Bank_acct,balance,0)),
 asserta(has_value(Bank_acct,int_rate,12/100)).

is_method(bank_acct,close).
method(Bank_acct,close,_) :-
 has_value(Bank_acct,balance,What),
 do_method(Bank_acct,withdraw,What),
 retractall(is_instance(Bank_acct,_)).
 retractall(has_value(Bank_acct,_,_)).

 is_method(bank_acct,deposit).
 method(Bank_acct,deposit,Amt) :-
   has_value(Bank_acct,balance,Current),
   New_amt is Amt + Current,
   retract(has_value(Bank_acct,balance,Current)),

asserta(has_value(Bank_acct,balance,New_amt)).

 is_method(bank_acct,withdraw).
 method(Bank_acct,withdraw,Amt) :-
   has_value(Bank_acct,balance,Current),
   ((Amt > Current, write('That is an overdraw'));
   (Amt =< Current,
   New_amt is Current - Amt,
   retract(has_value(Bank_acct,balance,Current)),

asserta(has_value(Bank_acct,balance,New_amt)))).

is_method(bank_acct,name).
method(Bank_acct,name,Name) :-
 has_value(Bank_acct,name,Name).
```

Figure 3a. The class *bank_acct*

```
is_class(savings_acct).
kind_of(savings_acct, bank_acct).

has(savings_acct, int_rate).

is_method(savings_acct, new).
is_method(savings_acct, int_rate).
is_method(savings_acct, change_rate).
is_method(savings_acct, add_monthly_int).

method(savings_acct, new, Id) :-
asserta(is_instance(Id, savings_acct)),
do_method(Id,initialize, _).

method(Savings_acct, int_rate, Int_rate) :-
has_value(Savings_acct,int_rate, Int_rate).
```

Figure 3b. The class *savings_acct*

In these figures, the methods allow the user to create new instances and change values through queries, instead of requiring access to the actual code file. The changing of instance variable values and creation of new instances are implemented through *retract* and *assert* predicates.

Figure 4 is a log of some of the queries used to test the system. As can be seen, all methods on the savings account 015646897 are performed correctly. The savings account is able to respond to messages invoking it to run methods owned by its parent *bank_acct* class. In addition, the bank account 123 is not able to perform methods owned solely by the

savings_acct class. These sample predicates confirm the functionality of the specified object-oriented goals.

```
?- do_method(savings_acct, open, 015646897).
Give the owner of the acct: william_dean.
Give the address of the owner: hillary_st.

?- do_method(015646897, name, What).
What = william_dean

?- do_method(015646897, int_rate, What).
What = 12/100

?- do_method(015646897, deposit, 1500).

?- value(015646897, balance, What).
What = 1500

?- do_method(015646897, withdraw, 2000).
That is an overdraw

?- do_method(015646897, withdraw, 300).

?- value(015646897, balance, What).
What = 1200

?- do_method(015646897, change_rate, 10/100).

?- do_method(015646897, int_rate, What).
What = 10/100

Give the owner of the acct: john_smith.
Give the address of the owner: generic_st.

?- do_method(123, int_rate, What).
no
```

Figure 4. A run log testing in PROPS

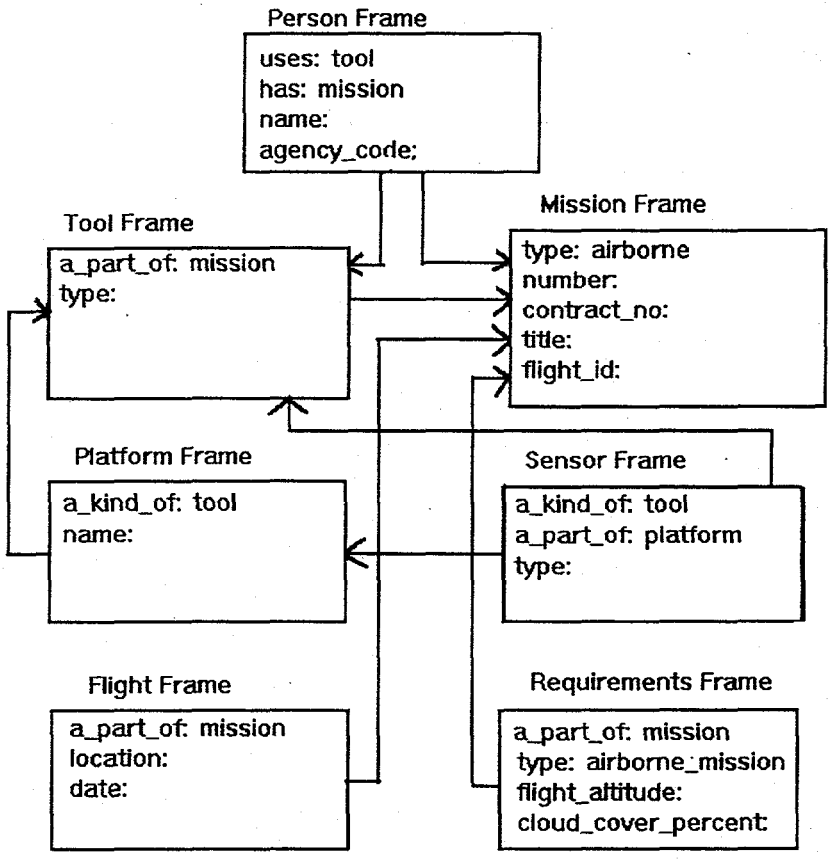


Figure 5a. Frame-Based Representation of a TIMS and CAMS Metadata Design

A PROPS IMPLEMENTATION OF A SIMS DESIGN

Over the years, Thermal Infrared Multispectral Scanner (TIMS) and Calibrated Airborne Multispectral Scanner (CAMS) data sets have been collected by the Stennis Space Center. These data sets form a database of remotely sensed data that will hopefully be incorporated into the Earth Observing System Data and Information System (EOSDIS). Here, the PROPS

system is used to develop an object-oriented representation of the TIMS and CAMS databases for a Stennis Information Management System (SIMS). This implementation is actually an evolution of a frame-based design developed by Saacks-Giguette and Lopez [6].

Figures 5a and 5b illustrate a database representation in PROPS and its frame-based representation developed in [6]. Each object has its own attributes and instance variables. While no methods have been established in [6], the metadata design is still useful because of the way it represents the objects in the database. PROPS would then be a useful tool because of its rapid prototyping and ease of use. It allows easy maintenance of the object-oriented database while utilizing the multi-indexing techniques and search methods of PROLOG as a relational database.

```
is_class(person).
  has(person, mission).
  has(person, name).
  has(person, agency_code).

is_class(mission).
 has(mission, type).
 has(mission, number).
 has(mission, contract_no).
 has(mission, title).
 has(mission, flight_id).
 has(mission, requirements).

is_class(tool).
  has(tool, type).
  has(tool, owner_mission).

is_class(sensor).
kind_of(sensor, tool).
  has_value(sensor, type,
sensor).
  has(sensor, sensor_type).
  has(sensor, owner_platform).

is_class(platform).
kind_of(platform, tool).
  has_value(platform, type,
platform).
  has(platform,
platform_name).

is_class(flight).
  has(flight, location).
  has(flight, date).
  has(flight, owner_mission).

is_class(requirements).
  has(requirements, type).
  has(requirements, altitude).
  has(requirements,
cloud_cover_pct).
```

Figure 5b. PROPS Representation of a TIMS and CAMS Metadata Design

FURTHER RESEARCH AND CONCLUSIONS

Further research is being performed to advance the PROPS system. One new idea includes developing a super method that can be utilized by objects. Another topic for further research is using Structured Object Method (SOM) diagrams for the design of object-oriented knowledge bases for the PROPS system. The SOM diagrams were developed originally by [2] as a way to design knowledge base and database coupling using an object-oriented approach. SOM is similar to frame-based design, which was used to develop PROPS.

The object-oriented paradigm has proven to be very useful in a number of important applications because of its ability to unify and simplify the description of entities and their protocols. However, the implementation of large databases has been hindered by slow searching methods. To overcome the impedance, object-oriented programming has been combined with logic programming to form PROPS, a PROLOG Object Programming System.

ACKNOWLEDGMENTS

This research was partially supported by the NSF grant CDA-9322553 at Loyola University in New Orleans, LA and the NSF grant HRD-9252702 at Xavier University of LA.

REFERENCES

1. Goldberg, A., D. Robson, D. Ingalls. Smalltalk-80: The Language and Its Implementation. Addison-Wesley Reading, MA. 1983.
2. Higa, K., M. Morrison, J. Morrison, O. R. Liu Sheng. "Object Oriented Methodology for Knowledge Base/Database Coupling". Communications of the ACM. Vol. 35, No. 6. 1992. 99-112.
3. Lalonde, W. Discovering Smalltalk. Benjamin/Cummings Publishing Company, Inc. Redwood City, CA. 1994.
4. Lucas, R. Database Applications Using PROLOG. Ellis, Horwood Limited, Chichester, England. 1988.
5. Rowe, N. C. Artificial Intelligence through PROLOG. Prentice Hall, Inc. Englewood Cliffs, NJ. 1988.
6. Saacks-Giguette, M., A. Lopez, Jr. "A Frame-Based Design for the TIMS and CAMS Metadata for a Stennis Information Management System." J. Systems Software. Vol. 20. 1993. 87-92.
7. Zaniolo, C. "Object-Oriented Programming In PROLOG." IEEE International Symposium on Logic Programming. Atlantic City, NJ. 1984. 265-269.

A Parallel Application for Delaunay Triangulation on the Intel Paragon

Francis Y. Fong

Department of Computer Science
Winston-Salem State University
601 Martin Luther King, Jr. Drive
Winston-Salem, NC 27110
e-mail: ffong@152.12.5.2

Faculty Advisor:
Fenglien Lee
e-mail: flee@uncecs.edu

ABSTRACT

Given a point set S with N points on the plane, the two-dimensional Delaunay triangulation (2-d DT) is a triangulation of S such that there is no any other point of S locates inside the circumcircle of each triangle. The 2-d DT has many practical applications on sciences and technologies, such as Unstructured Mesh Generation on Computational Fluid Dynamics (CFD), Geometric Modeling on Computer-Aided Design (CAD), and Voronoi Diagram on Computational Geometry. In this research, we develop a parallel algorithm for constructing the 2-d DT and implement it to a parallel C program on the Intel Paragon. We also evaluate the performance of Paragon by using different sizes of S and different numbers of CPU nodes. The time complexity of the best sequential algorithm for 2-d DT is $O(N \log N)$. Our parallel algorithm improves it to constant or $O(1)$ time by using N nodes [1][4].

INTRODUCTION

Delaunay Triangulation

Given a set S of N points in the plane, a two-dimensional Delaunay Triangulation is the set of straight line segments that joins all the points in S without intersecting each other, such that the convex hull of S is divided into triangles, and the circumcircle of the vertexes of each triangle in DT contains no other point of S [5]. The convex hull of S is the smallest convex polygon that is defined by joining some of the points in S, such that the interior of the polygon includes all the points in S. We call any edge in the DT as a Delaunay edge and any triangle in the DT as a Delaunay triangle. Chazelle solved DT in $O(N)$-time by a linear array [2]; MacKenzie and Stout used a hypercube machine and obtained the time complexity $O(\log N \log \log N)$ [3]; Wang and Chin used a CREW PRAM to solve it in $O(\log N)$-time [6]. Each of the above model has $O(N)$ processors.

The DT has the following two important properties:

Property 1: The edge between each pair of points of the nearest-neighbor belongs to DT. For a point set S, the all nearest-neighbors problem is to find, for each point $p \in S$, a point $q \in S - \{p\}$ such that the Euclidean distance between p and q is minimum [3].

Property 2: If AB is a Delaunay edge in the triangulation, then if C is the site for which $\cos\angle ACB$ is minimum among all points on the same side of line AB on which C lies, then ΔABC is a Delaunay triangle in the triangulation.

To construct a DT, we first distribute the input points to all nodes evenly. Then according to Property 1, we have to find all nearest-neighbor edges. Next, we can find Delaunay triangles by applying Property 2 to each nearest-neighbor edge. Continue to apply Property 2 to all new edges generated to find more new Delaunay triangles, until there is no any new edge found.

Intel Paragon

The Intel Paragon system is a Multiple Instruction stream Multiple Data stream (MIMD) machine with shared memory. The Paragon OSF/1 operating system is run on Intel supercomputers. An Intel supercomputer has a large number of nodes (processors) that work

concurrently on the parts of a problem. The nodes are connected by a high-speed interconnect network in a mesh configuration, and a number of I/O interfaces to communicate with outside world. The node interconnect network is shown in Figure 1.

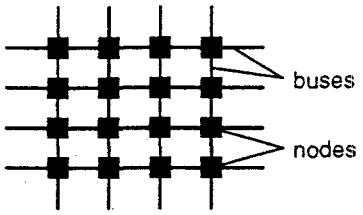


Figure 1: Node interconnect network in the Intel Paragon.

Each node is essentially a separate computer, with one or more i860 XP processors and 16M bytes or more memory. Each node can independently run multiple processes using OSF/1 scheduling mechanisms. The nodes can also work together on a parallel application and exchange data by passing messages. Each node interfaces with the node interconnect network and extracts only messages addressed to itself. Messages addressed to other nodes are passed without interrupting the node processor. Some nodes equipped with I/O connection communicate with other nodes over the network. The access of I/O facilities is transparent; processes on nodes without I/O hardware access the facilities using standard OSF/1 system calls.

The Paragon OSF/1 operating system is an extended version of OSF/1 with enhancements to support parallel processing. OSF/1 from Open Software Foundation is a version of the UNIX operating system that supports most industry standard. Paragon OSF/1 system provides a single system image across multiple nodes. All nodes share a single file system and have equal access to I/O devices. Process identifiers (PIDs) are also unique throughout the system; any node can kill a specified process no matter which node the process is running on [7].

DATA DECOMPOSITION

Data decomposition is the most important concept in our algorithm. We assume a set S of N points in the plane sorted by x and y is provided as input, and sorting is not part of this procedure. We also assume that the input is already in the shared memory of the Intel Paragon. We first divide S into P subsets, for P is the number of nodes used. After the decomposition, each node handles one subset. There is an overlap area with the width of δ between each subset, such that, each subset S_P has approximately N/P points. We assume that δ is the length of the longest edge in the triangulation, and each $x_{N/P}$ is the x-coordinate of the N/P th point from a subset. We also find the *first* and *last* points of each subset. Note, when N nodes are used, P is equal to N; therefore, each S_P has approximately N/N or 1 point, and also $last - first \approx 0$. For example, in Figure 2, 8 points are divided into 2 subsets, so N/P is 4. In the figure, point 4 is the N/P th point of subset S_0, and points 1 and 6 are its *first* and *last* points respectively. Subset S_1 has point 8 as the N/P th point, while its *first* and *last* points are points 5 and 8, respectively.

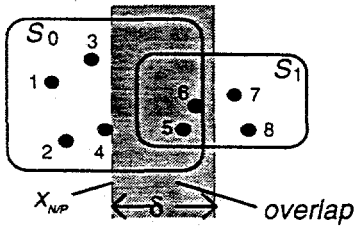


Figure 2: Eight points are divided into S_0 and S_1.

and there are 2 points in the overlap area.

Procedure *DATA_DECOMPOSITION*;
Input: The (x, y) coordinates of each point in S sorted by x and y;
Output: The P subsets of S;
BEGIN
(1) Assign points $[n \times (N/P) + 1]$ through $[(n + 1) \times (N/P)]$ to node n;
(* The first point to the N/P th point of S_n are assigned to node n. *)
(2) Let $x_{N/P}$ be the x-coordinate of the N/P th point of each node;
(3) Let p be $N/P + 1$ th point, and x_p be the x-coordinate of p;
(4) WHILE $|x_p - x_{N/P}| < \delta$ DO
(* Find the points in the overlap area. *)
(4.1) Assign p to the node n;
(4.2) Let p be the next point;
ENDDO;
END *DATA_DECOMPOSITION*.

Lemma 1: The time complexity for data decomposition of the set of N points in the plane by an Intel Paragon with N nodes is $O(1)$.

Proof: In this procedure, obviously, steps (2) and (3) need constant time. When the size of P is very close to N, N/P becomes a constant; therefore, step (1) needs only constant time, too. Since the number of points in the overlap area is the total points in the subset subtracting N/P, and we mentioned before that the total points in a subset to approximately N/P, so the number of points in the overlap area is also a constant. Thus, step (4) only needs constant time. Hence, the time complexity of *DATA_DECOMPOSITION* is constant.

ALL NEAREST NEIGHBORS

After each node received a subset S_P of S, we then find the nearest-neighbors for all points in S. Each node computes the nearest-neighbor, for each point in its S_P, and the edge between the pair of nearest-neighbor points. Figure 3 shows the edges of all nearest-neighbors of a point set.

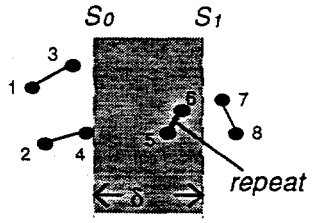


Figure 3: The edges of all nearest-neighbors of a point set.
Edge $E_{5,6}$ is repeated by S_0 and S_1.

Algorithm *NEAREST_NEIGHBOR*;
Input: The (x, y) coordinates of each point in each S_P;
Output: The nearest-neighbor of each point in each S_P and the edge between them;
BEGIN
FOR *first* $\le i \le$ *last* DO IN PARALLEL
(1) FOR *first* $\le j \le$ *last* DO IN PARALLEL
Compute the length of edge $E_{i,j}$, $L_{i,j}$, between points p_i and p_j by the following equation:

$$L_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad (1)$$

(* where (x_i, y_i) and (x_j, y_j) are coordinates of p_i and p_j respectively. *)
ENDFOR;

(2) Find the shortest edge, say $E_{i,k}$, among all E_{ij} edges in step (1), where $k \neq i$ and $p_k \in S_P$.

(3) IF $L_{i,k} \leq \delta$ THEN

Compute the $E_{i,k}$ by the following equation:

$$E_{i,k} = (x - x_i)(y_j - y_i) - (y - y_i)(x_j - x_i) = ax + by + c = 0 \quad (2)$$

(* where a, b, and c are coefficients of the edge equation. *)

ENDIF;

ENDFOR;

END *NEAREST_NEIGHBOR*.

Theorem 1: The time complexity for finding all the nearest-neighbors of a set of N points in the plane by an Intel Paragon with N nodes is $O(1)$.

Proof: In this algorithm, since $last - first \approx 0$, steps (1), (2), (3) and the outer FOR loop all have the same time complexity $O(1)$. Hence, the time complexity of *NEAREST_NEIGHBOR* is constant.

DELAUNAY TRIANGULATION

To construct a two-dimensional Delaunay Triangulation, we first call procedure *DATA_DECOMPOSITION* to distribute the input data. Then, we call algorithm *NEAREST_NEIGHBOR* to find all nearest-neighbor edges. Finally, we find Delaunay triangles by applying Property 2 to each nearest-neighbor edge. Continue to apply Property 2 to all new edges generated, until no more new edges are found. A Delaunay Triangulation of a point set is shown in Figure 4.

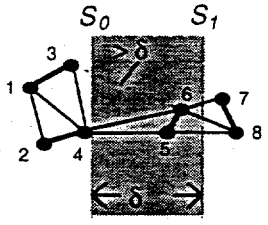


Figure 4: A Delaunay Triangulation of a point set.
Edge $E_{3,6}$ is longer than δ, so it is not in DT.

Algorithm *TRIANGULATING*;
Input: The sorted point set S;
Output: The DT of S;
BEGIN
(1) Call procedure *DATA_DECOMPOSITION* to distribute the point set S;
(2) Call algorithm *NEAREST_NEIGHBOR* to find all nearest-neighbor edges of S;
(3) FOR $first \leq i,\ j \leq last$ DO IN PARALLEL
(3.1) FOR $first \leq k \leq last$ DO IN PARALLEL
(3.1.1) Compute the distance $d1$ and $d2$ for edges $E_{i,k}$ and $E_{j,k}$ by using Equation 2.
(3.1.2) IF ($d1 < \delta$) AND ($d2 < \delta$) THEN
(3.1.2.1) Compute *cosine* for $\angle p_i p_k p_j$ by the following equation:

$$cosine = \frac{d1^2 + d2^2 + dist(E_{i,j})^2}{2 \cdot d1 \cdot d2} \quad (3)$$

(3.1.2.2) To determine p_k is on which side of $E_{i,j}$, by the following equation:

$$plane = ax_k + bx_k + c \quad (4)$$

(* where a, b, and c are the coefficient of $E_{i,j}$. *)
(* if *plane* > 0 then p_k is on the positive half plane of $E_{i,j}$;
else if *plane* < 0 then p_k is on the negative half plane of $E_{i,j}$;
else if *plane* = 0 then p_k is on $E_{i,j}$ *)
(3.1.2.3) IF *cosine* has the smallest value for all p_k so far THEN
Update the value of *min1*, *min2*, *top1*, and *top2*.
(* where *top1* and *top2* are points (k), such that *min1* and *min2* are the current minimum cosine value for positive and negative half plane, respectively. *)
ENDIF;
ENDFOR;
(3.2) Compute coefficients for edges $E_{i,top1}$, $E_{i,top2}$, $E_{j,top1}$, and $E_{j,top2}$, if they exist.
(* these are the new Delaunay edges. *)
ENDFOR;
(4) FOR *first* ≤ $i, j \leq$ *last* DO IN PARALLEL
Output $E_{i,j}$;
ENDFOR;
END *TRIANGULATING*.

Theorem 2: The time complexity for constructing the Delaunay Triangulation of N points in the plane by an Intel Paragon with N nodes is $O(1)$.

Proof: In this algorithm, we have proved that steps (1) and (2) need constant time. Similarly, since *last* – *first* ≈ 0, steps (3) and (4) both need constant time, too. Hence the time complexity of *TRIANGULATING* is $O(1)$.

CONCLUSION

In this paper, we present a very efficient parallel algorithm for two-dimensional Delaunay Triangulation on the Intel Paragon. Our algorithm has the time complexity of $O(1)$, which is the best so far. The only open problem is how to triangulate when four or more points are co-circular.

REFERENCES

[1] G. Akl and K. A. Lyons, *Parallel Computational Geometry*, Prentice-Hall, Inc., 1993.

[2] B. Chazelle, "Computational Geometry on a Systolic Chip," *IEEE Transactions on Computers*, Vol. C-33, No. 9, Sept. 1984, pp774-785.

[3] D. MacKenzie and Q. F. Stout, "Asymptotically Efficient Hypercube Algorithms for Computational Geometry," *Proceedings of the Third Symposium on the Frontiers of Massively Parallel Computation*, October 1990, pp8-11.

[4] P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*, Revised Edition, Springer-Verlag, 1988.

[5] Saxena, P. C. P. Bhatt, and V. C. Prasad, "Efficient VLSI Parallel for Delaunay Triangulation on Orthogonal Tree Network in Two and Three Dimensions," *IEEE Transactions on Computers*, Vol. 39, No. 3, March 1990, pp400-404.

[6] F. Wang and Y. H. Tsin, "An $O(\log N)$ Time Parallel Algorithm for Triangulating a Set of Points in the Plane," *Information Processing Letters*, Vol. 25, 1987, pp55-60.

[7] *Paragon User's Guide*, Intel Supercomputer System Division, 1994.

Using a Neural Network in Predicting Stock Market Return

Jovina Roman
Department of Computer Science
Xavier University of Louisiana
7325 Palmetto St., New Orleans, LA 70125-1145
e-mail: roman@drexel.xula.edu

Faculty Advisor
Akhtar Jameel

In the past many analytical methods have proven ineffective in trying to model the changes in a stock market. A stock market's objective is to maximize a company's return taking into account the random market changes. Can a neural network predict the changes in a stock market? It can be shown how historical stock data in training a neural network can be used to predict future stock return using a Backpropagation neural network. Design and architecture of these networks is entirely empirical. The input-output mapping learned by the network after training is used to predict future stock return in various stock markets. A correlation in stock return variations in five prominent stock markets was also investigated. Stock values for Canada, Hongkong, Japan, UK and USA during 1989 are used to test these hypotheses. A prediction accuracy in the range of 25-75% (55% average) was displayed by the network to predict December 1989 stock return variations based on training data from January through November 1989 on the five stock markets.

INTRODUCTION

This paper presents use of Backpropagation neural networks in predicting random changes in stock market return. The focus is to evaluate two basic stock market prediction hypotheses: (1) use of recent historical stock data to forecast future changes in the stock prices, and (2) use trends in one stock market to forecast changes in other international stock exchanges. A Backpropagation network will be trained with one year stock data from January through November 1989 to determine the best learning rate, number of iterations, size of the network, and number of input samples which will minimize the overall error. This trained network will then be used to predict the stock prices for December 1989. In addition, stock data from each of the five countries: Canada, Japan, Hongkong, UK or USA will be used to predict the stock returns during the same period for the other four countries.

A neural network is a computer program that recognizes patterns and is designed to take a pattern of data and generalize from it. An essential feature of this technology is that it improves its performance on a particular task by trial and error, more preferably "it learns". There are no set rules or sequence of steps to follow in generalizing patterns of data. The network is designed to learn a nonlinear mapping between the input and output data. Generalization is used to predict the possible outcome for a particular task. This process is broken down into two phases known as the *training phase* (learning) and the *testing phase* (prediction).

The stock market return prediction problem can be approached using neural networks because no algorithmic methods exist to describe its variations. Stock prices fluctuate daily resulting in a nonlinear pattern of data. One of the three outcomes may occur in the stock price: rise, fall, or remain the same, which will occur is uncertain. There are several ratios that determine stock price fluctuation, most important is the dividend ratio. According to Weston and Brigham (1993) "the value of the stock today is calculated as the present value of an indefinite stream of dividends." The stock value at present can be determined based on the present value of dividends. The growth in dividends results in a growth in Earning Per Share which means that return on stock will be greater. The changes in dividends and stocks are directly related, consequently the dividend ratio is useful in determining whether stock return will rise or fall.

BACKGROUND

Regression models have been traditionally used to model the changes in the stock market. According to Mendenhall and Beaver (1994) "the objective of a multiple regression analysis is to relate a response variable 'y' to a set of predictor variables." Multiple regression

analysis is the process of finding the least-squares prediction equation, testing the adequacy of the model, and conducting tests about estimating the values of the model parameters. However, these models can predict linear patterns only. The stock market returns change in a nonlinear pattern such that neural networks are more appropriate to model these changes.

Refenes, Zapranis, and Francis (1994) compared regression models with a Backpropagation network using the same stock data. In comparison with regression models Backpropagation proved to be a better predictor. The results showed that the MSE (Mean Squared Error) for the neural network was lower than the MLR (Multiple Linear Regression) model. The MSE for the neural network was 0.044 and the MSE for the MLR model was 0.138 such that the neural net proved to be more effective in learning the training data than the MLR. For the out-of-sample test data the neural network MSE was 0.066 which is also lower than the MLR MSE of 0.128. According to Refenes, Zapranis, and Francis (1994) "neural networks are capable of making better prediction in capturing the structural relationship between a stock's performance and its determinant factors more accurately than MLR models."

Kryzanowski, Galler, and Wright (1993) using Boltzmann machine trained an artificial neural network with 149 test cases of positive (rise in the stock price) and negative (fall in the stock price) returns for the years 1987-1989 and compared this to training the network with positive, neutral (unchanged stock price), and negative returns for the same 149 test cases for the years 1987-1989. The network predicted 72% correct results with positive and negative returns and 46% correct results with positive, neutral, and negative returns. These experiments will attempt to overcome this 72% accuracy using a Backpropagation network.

EXPERIMENTS

Method

Standard supervised Backpropagation neural network learning methodology was followed in these experiments. A subset of available stock market return data was used to construct training samples for the network. Training of a Backpropagation network involves obtaining optimal values for the learning rate, the momentum of learning, estimating the number of hidden layers and the number of nodes in each layer. The overall error is tracked until a minima is obtained by altering the fore-mentioned parameters. A trained network can be used in predicting future stock market returns and to study correlation between variations in other stock market returns.

Network

A Backpropagation neural network implemented by Yoh-Han Pao (1989) was used in the experiments. This network employs a two-pass weighted learning algorithm known as the *generalized delta rule* (Rumelhart, McClelland, the PDP Research Group, 1988). In a forward pass through the network, an error is detected, this error is then propagated backward through the network while weights are adjusted to reduce the overall error. This iterative process that the network goes through in reducing the overall error is known as *gradient descent*.

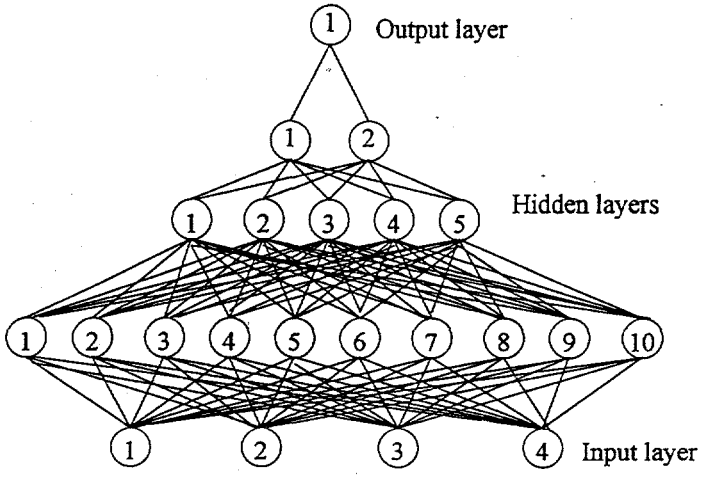


Figure 1: A Backpropagation Network.

These networks are highly suited for the stock market return prediction problem. Backpropagation and other neural networks have been used by Widrow, Rumelhart, and Lehr (1994), Kryzanowski, Galler, and Wright (1993) and Refenes, Zapranis, and Francis (1994) for financial data analysis. This particular network was chosen based on its performance in Refenes, Zapranis, and Francis (1994). Moreover, they have been successfully incorporated into similar speech and character pattern recognition mapping problems.

Data

Historical stock market return data from January through November 1989 for the following five countries: Canada, Japan, Hongkong, UK and USA was used. The stock market return data for each of these countries were substituted by a number between 1 and 0 in the event of rising, falling, and remaining the same . If the return fell the change was represented by 0.1, if it had risen the change was represented by 0.9, and if it remained the same the change was represented by 0.5.

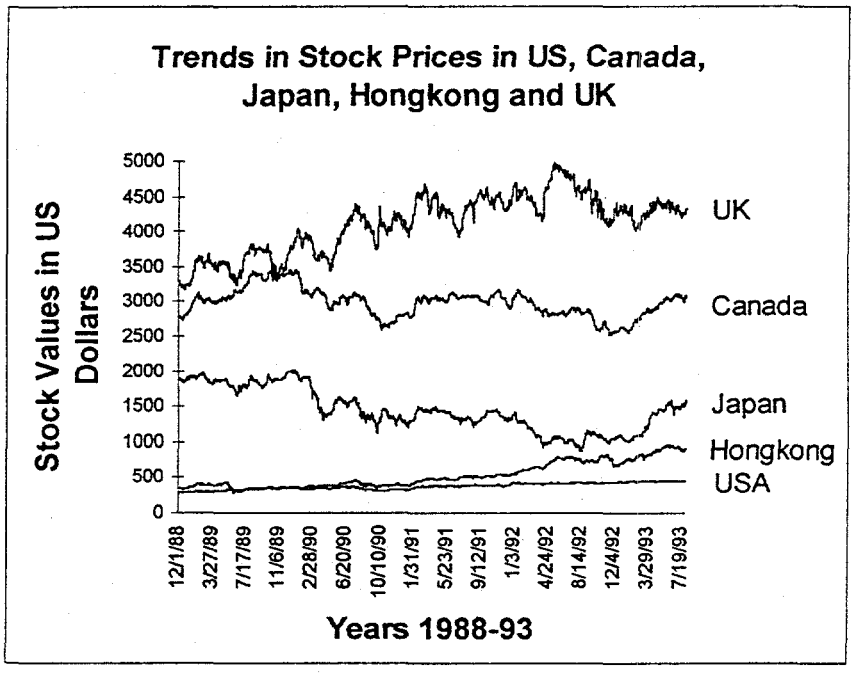


Figure 2: Canada, Hongkong, Japan, UK and USA Stock Exchange Index 1/1/1989 - 12/31/92

Table 1: Prediction Accuracy of a Backpropagation Network Using Historical Stock Data

Sl. No.	Number of Iterations	Training data country (01/89-11/89)	Testing data country (12/89)	Overall error obtained	Accuracy Percentage for 12/89
1	13,000	USA	USA	0.047484	75.0%
2	25,000	USA	USA	0.047465	75.0%
3	50,000	USA	USA	0.047459	75.0%
4	100,000	USA	USA	0.047454	75.0%
5	100,000	Canada	Canada	0.073869	50.0%
6	100,000	Japan	Japan	0.067802	25.0%
7	100,000	Hongkong	Hongkong	0.062071	50.0%
8	100,000	UK	UK	0.051128	75.0%

After tabulating the trends in data an interval of a week (Monday to Tuesday, Tuesday to Wednesday,..., and Thursday to Friday) was chosen resulting in four inputs and one output (Friday to Monday). The training file is set up with 48 weeks data to be trained for January

through November 1989. The testfile is set up with 4 weeks to be tested on for December 1989. Hence, the four countries have four training and testing files for the year 1989. An error file is produced at the end of each run. This file holds the overall error after each iteration. A trained file is created which contains trained data, the total error, number of iterations, inputs, outputs, learning rate, momentum rate, number of inputs, number of hidden layers, and the number of neurons in each layer. A file containing the weights for each neuron is also created.

This trained file will then be used to predict the stock prices for December 1989. In addition, stock data from one of the five countries: Canada, Japan, Hongkong, UK or USA, will be used to predict the stock returns during the same period (January to November 1989) for the other four countries. These experiments are repeated for each of the five countries. The graph in Figure 2 shows the nonlinear trends in the stock market returns for each of these five countries from 1989-92.

Table 2: Prediction Correlation Between the Five International Stock Markets

Sl. No.	Number of Iterations	Training data country (01/89-11/89)	Testing data country (01/89-11/89)	Overall error obtained	Accuracy Percentage (01/89-11/89)
1	100,000	USA	Canada	0.047454	52.00
2	100,000	USA	Hongkong	0.047454	56.00
3	100,000	USA	Japan	0.047454	44.00
4	100,000	USA	UK	0.047454	42.00
5	100,000	Canada	Japan	0.073869	43.75
6	100,000	Canada	Hongkong	0.073869	62.50
7	100,000	Canada	UK	0.073869	60.00
8	100,000	Canada	USA	0.073869	66.67
9	100,000	Japan	Canada	0.067802	45.83
10	100,000	Japan	Hongkong	0.067802	45.83
11	100,000	Japan	UK	0.067802	52.08
12	100,000	Japan	USA	0.067802	58.33
13	100,000	Hongkong	Canada	0.062071	58.33
14	100,000	Hongkong	Japan	0.062071	52.08
15	100,000	Hongkong	UK	0.062071	66.67
16	100,000	Hongkong	USA	0.062071	66.67
17	100,000	UK	Canada	0.051128	64.58
18	100,000	UK	Hongkong	0.051128	62.50
19	100,000	UK	Japan	0.051128	54.17
20	100,000	UK	USA	0.051128	54.17

Results

The results of these experiments are indicated in Tables 1 and 2. For each of these experiments the following parameters were observed to produce the minimum overall error: Inputs 4, Outputs 1, Number of Inputs 48, Hidden layers 3, Neurons in first hidden layer 10, Neurons in second hidden layer 5, Neurons in third hidden layer 2, Learning rate 0.7, Momentum rate 0.9, Target total error 0.001, and Target individual error 0.0001

ANALYSIS OF RESULTS

Kryzanowski, Galler, and Wright (1993) used a Boltzmann Machine on stock market data from 120 companies over a 6 year period 1984-1989 and obtained 71.7% accuracy. In comparison a 73% accuracy of prediction was observed for the USA stock market data from January through November 1989.

Apart from the results indicated in the Tables 1 and 2, a few other variations on the network architecture were implemented. For example, adding another output neuron (two-output network) for 50,000 iterations produced an error of 0.216164. Although the accuracy of prediction was 66.67% the overall error is very high. A larger network with 20, 10, and 5 neurons in the hidden layers also produced a higher error of 0.048040 for 15,000 iterations. A larger learning rate of 0.9 also produced a higher overall error of 0.049217 for 15,000 iterations.

An overall prediction accuracy range of 25-75% (average 55%) was achieved during these experiments in Table 1. The prediction accuracy, with the exception of Japan for stock market returns was 50% or higher. In Table 2 and Figure 2 the prediction accuracy are higher for stock markets which change in the same direction and low for the markets that change in opposite directions. For example, accuracy of predicting US stock market from the Canadian stock market is higher, and the accuracy of predicting Japanese stock market from any other market are typically lower.

CONCLUSION

A prediction accuracy in the range of 25-75% (55% average) was displayed by the Backpropagation network when used to predict December 1989 stock variations based on the training information from January 1989 through November 1989 on five different international stock markets. Backpropagation network is also a good tool to model correlation between these international stock exchanges as seen in Table 2 and Figure 2.

The results of this networks accuracy have proven quite effective. Hence when investing, using backpropagation neural networks that have been trained on stock market returns in previous years, would be a wise choice.

The Backpropagation network is one way of approaching the stock market return problem however, other networks may be tried for prediction. Other networks such as recurrent networks which are better known for time-series problems may also be tried for the stock market return problem. Future work will involve using all four years data 1989-93 for experiments instead of the only one year used at this time. The prediction parameters used can also be looked at in terms of using the stock market return to be trained with the change in dividends mentioned above. The chosen interval of a weeks return could be extended to monthly or yearly intervals.

ACKNOWLEDGEMENT

This research was partially supported by the GE Fund's Faculty for the Future Program.

REFERENCES

Kryzanowski, Galler and Wright, Financial Analysts Journal, Using Artificial Neural Networks to Pick Stocks, July-August 1993. pp. 21-27.

Mendenhall, Beaver, 1994, Introduction to Probability And Statistics, Ninth Edition, International Thomson Publishing.

Pao, 1989, Adaptive Pattern Recognition and Neural Networks, Addison Wesley Publishing Company, Inc.

Refenes, Zapranis, Francis,. Journal of Neural Networks, Stock Performance Modeling Using Neural Networks: A Comparative Study With Regression Models, Vol. 7, No. 2, 1994. pp. 375-388.

Rumelhart, McClelland, the PDP Research Group, 1988, Parallel Distributed Processing Volume1: Foundations, The Massachusetts Institute of Technology.

Weston, Brigham, 1993, Essentials of Managerial Finance, The Dryden Press.

Widrow, Rumelhart, Lehr,. Journal of Communications of the ACM, Neural Networks: Applications in Industry, Business and Science, Vol. 37, No. 3, 1994. pp. 93-105.

Quadratic Feedback Shift Register Sequences

Diana I. Simen

Department of Computer Science
Mount Union College
Alliance, OH 44601

Faculty Advisor:

Dr. Agnes H. Chan

Department of Computer Science
Northeastern University
161 Cullinane Hall
Boston, MA 02115

INTRODUCTION

Random sequences are used in encrypting and decrypting messages. The messages to be encrypted are binary strings. In order to encrypt them, they are XORed with a binary random sequence known only by the sender and receiver of the message. Upon receiving the ciphertext, the receiver will decrypt the message by again XORing it with the same random sequence.

Truly random sequences are hard to generate and are expensive to use, therefore pseudo-random sequences are used instead. There are various types of pseudo-random sequences that can be generated. In many cases, these sequences, which are the "keys" to the encrypted messages, can be determined by "enemy parties" or eavesdroppers after knowing some portion of the key sequence. It would be convenient to have methods of generating pseudo-random sequences that appear to be random to the enemies. Pseudo-random binary sequences are being studied, in the hopes of finding such "good" key sequences.

This paper assumes a basic knowledge of cryptographic theory. For additional information, the references listed at the end of the paper can be referred to.

FEEDBACK SHIFT REGISTER SEQUENCES AND RANDOMNESS PROPERTIES

A *feedback shift register* (FSR) is a simple deterministic device that outputs a sequence $\{a_n\}$. An FSR is shown in figure 1, where $a_{i+r+1} = f(a_{i+1}, a_{i+2}, \ldots, a_{i+r})$. The r-variable function f is called the feedback function.

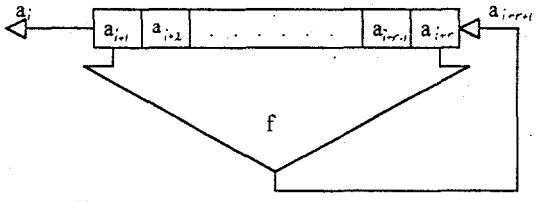


Fig. 1. Feedback Shift Register (FSR)

At any time instant, the content of the registers $(a_{i+1}, \ldots a_{i+r})$ is called a state of the FSR. An FSR is a simple finite state machine. Since there are only finitely many states, a state has to eventually repeat, therefore sequences generated by FSR's are periodic with

period $p \leq 2^r$. Here r is the degree of the sequence which is defined as the number of the highest index plus one.

Since truly random sequences are hard to find, pseudo-random sequences generated by FSR's are used instead. These sequences must have certain characteristics and satisfy certain properties in order to be useful for cryptography. However, not every FSR sequence can be used as a pseudo-random sequence. A sequence is considered to be "random"-looking if it satisfies three randomness properties, as defined by Golomb in [2].

Balance Property

In every period, the number of 1's is nearly equal to the number of 0's; that is in a binary sequence $\{a_n\} = \{a_0, a_1, a_2, ...\}$, the difference between the number of 1's and 0's does not exceed 1.

Run Property

We define the number of consecutive 0's and 1's in a sequence as a *run*. For example, in the following sequence of period 15 there are eight runs, four of length 1, two of length 2, one of length 3, and one of length 4.

000111101011001

The second randomness property states that in every period, half the runs have length one, one-fourth have length two, one-eight have length three, etc., as long as the number of runs exceeds 1 [2]. Also, the *total* number of runs of 1's equals the *total* number of runs of 0's, because the runs always alternate.

Autocorrelation

For a binary sequence $\{a_n\} = \{a_0, a_1, a_2, ...\}$, the *autocorrelation function* A(τ) is defined as

$$A(\tau) = \frac{1}{p} \sum_{n=1}^{p} a_n a_{n+\tau},$$

where τ is the phase shift of the sequence.
The third randomness property requires that the autocorrelation function is two-valued.

$$pA(\tau) = \sum_{n=1}^{p} a_n a_{n+\tau} = \begin{cases} \text{p if } \tau = 0, \\ \\ \text{N if } 0 < \tau < \text{p, where N is an integer.} \end{cases}$$

QUADRATIC FEEDBACK FUNCTIONS

Pseudo-random sequences are called *linear feedback shift register sequences* if they are generated by an FSR with the feedback function f given by

$$f(x_1, ..., x_r) = c_1x_1 + c_2x_2 + ... + c_rx_r.$$

The coefficients $c_1, c_2, ..., c_r$ are 0's or 1's. Golomb proved in [2] that they satisfy all the randomness properties. These sequences are not practical for cryptographic purposes, because they can be easily determined after observing 2r terms of the encrypted message.

By introducing quadratic terms, the sequences will become more complex, and will have slightly different properties. One general form of a quadratic feedback function is given by

$$f(x_0, ..., x_{n-1}) = a_0 + a_1x_1 + a_2x_2 + a_{1,2}x_1x_2 + a_3x_3 + a_{1,3}x_1x_3 + ... + a_{n-2,n-1}x_{n-2}x_{n-1},$$

where the coefficients a_0, a_1, a_2, $a_{1,2}$, ... are 0's or 1's. It is shown in [2] that a_0 has to be always 1, otherwise the period of the sequence will be reduced. The period of the sequences will be $p \le 2^r - 1$, where r is the degree of the sequence. If this equality is satisfied, then the period is at its maximum and the sequence is considered to generate a *full cycle*. In this study, only the sequences that generated full cycles are considered.

Another class of quadratic feedback shift register sequences can be obtained by introducing a constant term to the feedback function, that is

$$x_n = \sum_{i=0}^{n-1} a_i x_i + \sum_{0 \le i < j}^{n-1} a_{ij} x_i x_j + 1,$$

where all additions are mod 2. In other words, each of the terms generated using the feedback function will be complemented recursively, such that each additional term will be dependent upon the complement of the previous one. The sequences generated this way are a more complex variation of the QFSR sequences.

CORRELATION RESULTS

As one part of this project, programs were written that computed autocorrelation functions for sequences of degrees less than 7 that generated full cycles. Here are some of the results:

The autocorrelation function of the quadratic sequences is not two-valued. However, as in the following example of the autocorrelation function of a sequence of degree 4, the values are symmetrical.

τ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$A(\tau)$	15	-1	-1	-1	-9	3	-1	3	3	-1	3	-9	-1	-1	-1

Furthermore, the highest peak, 15, occurs when τ=0. If we considered the values as being a cycle, then the 15 is surrounded by -1's.

The number of quadratic sequences of degree less than 7 that generate a full cycle is enumerated in the next table.

Table 1. Number of QFSR sequences that generate full cycle (with or without constant term)

Degree	1	2	3	4	5	6	7
# of seq.	1	1	2	16	128	1,952	64,056

Since the number of sequences is so large, and significant patterns of their occurrence were not observed, we started narrowing down our choices by restricting the sequences to several patterns. The patterns that the sequences were narrowed down to were of the form $x_0 + x_i + x_j + x_k x_t$, $x_0 + x_i + x_j + x_i x_j$, and $x_0 + x_i x_j$. The following examples have the mentioned patterns, respectively.

$$a_0 + a_1 x_1 + a_3 x_3 + a_{2,3} x_2 x_3$$
$$a_0 + a_1 x_1 + a_3 x_3 + a_{1,3} x_1 x_3,$$
$$a_0 + a_{1,3} x_1 x_3.$$

Restricting the sequences to these patterns, we obtain the results listed in the next table. Only sequences with degrees less than seven were analyzed at this time because of limited CPU time, so the results for the sequences of degrees eight through twelve were taken from [1].

Table 2. QFSR sequences without constant term restricted to several patterns

degree	4	5	6	7	8	9	10	11	12
$x_0+x_i+x_j+x_k x_t$	6	8	16	14	30	8	14	6	6
$x_0+x_i+x_j+x_i x_j$	2	2	6	2	4	0	0	0	2
$x_0+x_i x_j$	0	0	0	0					

Observation 1: From table 2, QFSR sequences of pattern $x_0 + x_i x_j$ did not generate full cycles in sequences of degrees less than seven. Even though sequences of higher degrees could not be checked, it is reasonable to assume that the mentioned pattern will not occur for any other such sequences.

The autocorrelation functions of the quadratic FSR sequences with constant term have the same properties as those without constant term. None of the autocorrelation functions of the FSR sequences with constant term are two-valued. However, comparing the FSR sequences with constant term to those without constant term, the following observations can be made:

Observation 2: Linear FSR sequences give a full cycle whether the constant term is included or not. This can be written as,

$$f(a_1, a_2, \ldots, a_n) <- f(a_1, a_2, \ldots, a_n) + 1.$$

Checking the same patterns considered for quadratic FSR sequences, we get the results summarized in the following table.

Table 3. QFSR sequences with constant term restricted to several patterns

degree	4	5	6	7
# of sequences	16	128	1,952	64,056
$x_0+x_i+x_j+x_kx_l$	4	6	10	20
$x_0+x_i+x_j+x_ix_j$	0	0	0	0
$x_0+x_ix_j$	2	2	6	2

Observation 3: In table 3, it can be seen that none of the QFSR sequences with constant term that generate full cycle are of the form $x_0 + x_i + x_j + x_ix_j$; however, the pattern $x_0 + x_ix_j$ does occur for this type of sequences.

Observation 4: For any degree n, the number of sequences with constant term that generate full cycle is equal to that of sequences without constant term, even though the sequences themselves may differ.

CONCLUSION

We have found no patterns of occurrence of quadratic FSR sequences useful for cryptographic purposes. Further investigation is needed to find viable sequences, but at least we have narrowed down the search somewhat. Finding patterns is necessary for developing a technique of generating pseudo-random sequences that can be used in modern communications systems. Until then, current cryptographic technology must suffice.

REFERENCES

[1] Chan, Agnes H., Games, R. A., and Rushanan Joe. *Quadratic M-Sequences.* to appear in Proceeding of Cambridge Software Encryption Workshop, Cambridge England, Dec. 1993.

[2] Golomb, Solomon W., *Shift Register Sequences.* Aegean Park Press. Laguna Hills, CA 1982.

Virtual Fire Propagation in Heterogeneous Media

Scott Walnum

Department of Computer Science
Trinity University
715 Stadium Dr.
San Antonio, TX 78212

Faculty Advisor:

Gerald Pitts

Fire propagation modeled in a virtual environment is an innovative, effective method for training fire fighters. For this project, an algorithm was designed and implemented for flame propagation through heterogeneous materials. The algorithm is intended to be used in a larger project for virtual reality fire fighter training. Previous models for fire propagation are not real–time, and have little graphic capabilities. This algorithm, VAFP (Virtual Algorithm for Fire Propagation), is implemented in a prototype software demonstration, provides a realistic, real time graphical display. The virtual environment space is defined in which materials with different burning characteristics (such as flammability, burn rate, and fuel content) can be arranged as desired. Flame activation can be placed anywhere in this space and once a material catches on fire, the flame spreads according to the VAFP algorithm. This paper will present a detailed description of the VAFP algorithm, as well as coding techniques, material definitions, and graphical displays. Additionally, model validation, future expansion, and other possible applications and benefits of this simulation model will be provided in the paper.

INTRODUCTION

Interest in virtual reality has grown enormously in the past few years, due in large part to the media. It has been portrayed as being either a great benefactor or the nemesis of humankind. However, the actual use of virtual environments has in reality been fairly limited. The gaming industry has made large contributions towards bringing virtual reality to the average consumer. Like many other new technologies, such as television, entertainment has been responsible for the rapid development and acceptance of virtual reality. Also, like television, virtual environments have many other applications beyond entertainment.

Virtual environments are particularly well suited for training and simulation. A user can be immersed in an environment and exposed to situations which may be too expensive or too dangerous for more conventional training methods. Perhaps the most advanced development of such systems has been in flight simulation. This application offers an effective, affordable solution for training pilots.

The author is involved in the development of a system to be used for the training of fire fighters using virtual environments. There are many advantages to using virtual fire simulation. The trainees could be placed in situations within the environment which would be too dangerous simply for training purposes. However, with the flexibility of virtual reality, the trainees could be taught to recognize areas which could be hazardous in a real fire. It

could also give them knowledge of how fires spread and expose them to the fire's progression from beginning to end. Because the hardware necessary for such an application is becoming more affordable, this method of training is now cost effective. Another reason this is desirable is because it decreases the number of actual training fires, an important component of current fire fighting training which has come under attack by environmentalists.

METHODOLOGY

The emphasis thus far has been on the development of a prototype fire model that is intended to simulate fire propagation through heterogeneous materials. The Virtual Algorithm for Fire Propagation (VAFP) is applied to any fire within the virtual universe and propagates as an actual fire would, according to the properties of the materials which it burns. This prototype algorithm is detailed below.

A voxel approach was used, where a universe is defined as a three dimensional grid with each cubic square within it defining one space (see Figure 1). Each space is either material or air. The materials have several characteristics which are important to VAFP. A material is defined by identifying variables for fuel content, flame resistance, flame growth and fuel consumption. Fuel content is self–explanatory, but flame resistance refers to the amount of fire which must "enter" the space before it will ignite. Once ignited, flame growth is the rate at which a burning material will increase its flame size. The rate at which fuel is depleted is a function of the variable for fuel consumption and the size of the flame. Because each space is an independent material, they can be arranged anywhere within the universe and objects can be constructed using these different materials.

The controlled structure for the VAFP, begins by creating a material queue (a linked list containing all material spaces which are burning). It selects the first node on the queue, then loops through all of the adjoining spaces surrounding the node (see Figure 2). If the adjoining space is also material, then the flame component of that space is increased and checked to see if the fire component has exceeded the flame resistance, then placing the space on the material queue. This is analogous to increasing the temperature of the space. Otherwise the adjoining space is air, in which case flame is transferred into it and it is placed on the flame queue. Since there is air available to the original node, then its flame size will grow according to its variable for flame growth while its fuel content will lessen in relation to the size of the flame and its fuel consumption variable. FC = FC – fc * fs, where FC is fuel content, fc is the fuel consumption variable, and fs is the size of the flame.

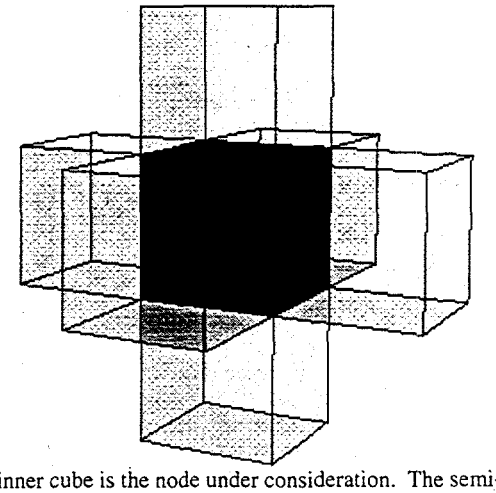


Figure 2. The dark inner cube is the node under consideration. The semi– transparent cubes are the adjoining spaces. These are located within a universe such as that in Figure 1. Each node representing one unit.

In the situation where none of the adjoining nodes are air, then the node is surrounded by material. Without any oxygen, the flame within the material will not increase and fuel will not be consumed. The space can be thought of as being heated beyond the ignition point, but it cannot burn until at least one of the nodes surrounding it is consumed. Once this happens, then the material will be touching air and will therefore be able to burn.

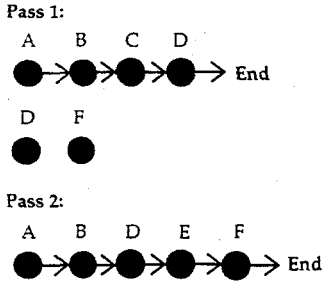
This process then repeats, checking every node that has been placed on the material queue. Nodes are placed on the queue once they have reached the ignition point, and they are removed as soon as all of their fuel is consumed, at which time they are treated as air spaces (see Figure 3a).

Essentially the same procedure is then used for processing the flame queue. The flame queue is composed of spaces which are air and also have flame in them. Like the material queue, the nodes are taken off of the queue and all of the adjoining nodes are checked. If the adjoining node is material, then some of the flame will transfer into the material according to its flame resistance. If the material ignites, it is placed on the material queue. In spaces which are air, the flame will spread into them according to the relative direction of the space plus a small random factor. This causes flame to transfer up more quickly than down or sideways, and the randomness gives the flame a jagged, more natural appearance. A sample screen illustrating how materials can be arranged and how the fire spreads is given (see Figure 3b).

PROBLEMS AND RECOMMENDATIONS

Although this method is fairly effective, it is based primarily on assumptions about flame propagation which need to be validated. The flame spread is essentially a very linear algorithm, but a true flame may spread differently. Further research is being conducted into replacing these equations with more accurate representations of flame spread. Once this research is concluded, the results can easily be incorporated within the prototype thus far developed.

Another major problem which this voxel approach has encountered is the computational explosion as the size of the fire grows. We are developing a new approach, which could be incorporated into further versions of the program. The new process would include tracking only the edges of the flame instead of every node within the flame, which could drastically reduce the numer of computations necessary (see Figure 5). Also being investigated is the possibility of moving away from the present voxel approach. This may



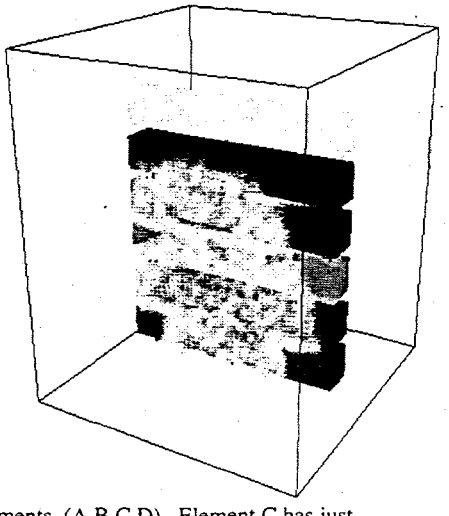


Figure 3a. In pass 1, the queue has four elements, (A,B,C,D). Element C has just consumed all of its fuel, while nodes E and F have just ignited. The next pass through the queue will reveal the queue strucutre in pass 2. C has been removed with E and F added.

Figure 3b This is an example of a universe with six materials.

eventually be necessary because of the freedom of material manipulation, but would require a much more sophisticated algorithm to handle the fire.

There are more factors that act on a fire which also require consideration. An important and very visible characteristic is smoke production. Different materials make different consistencies and volumes of smoke. Also, room temperature and heat flow within the air is very important. Both smoke and heat tend to cause a "hot layer" close to the ceiling, making the depth of that layer very important in a simulation. Both of these are dependent upon air flow. A fire's actions will vary drastically depending upon the amount of oxygen it is receiving. Because this model is primarily intended for simulating indoor environments, the effects of structures like vents, windows and doors on air flow will be important to fire propagation as well as to smoke and heat accumulation. In a fire, opening a door may have drastic and dangerous consequences.

The project could be put to other uses than fire fighter training. A major application of virtual reality is architectural walk throughs. This program could be applied here by giving the architect the ability to design and implement a fire safety system. Fire escapes, accessibility for fire fighters, as well as testing of sprinklers and other fire control devices could all be tested under many different situations. Fires could be started anywhere and their progress and effects observed. How long before smoke detectors would be activated? What might the damage amount to? What are the best combinations of building materials to retard fire propagation. Questions like these could be answered effectively and convincingly, giving the builder an idea how much money to spend on fire protection.

CONCLUSIONS

Virtual reality is an incredible method for exhibiting large amounts of information in an intuitive fashion. It is ideally suited for training applications such as this. As technology continues to increase, and hardware becomes faster and more widely available, applications like this may be used routinely for training. Television's potential as an educational tool has long been overshadowed by its use for entertainment. Is this the fate of virtual environments? Hopefully more applications which are intended for educational purposes will be developed and the great ability of virtual environments to transmit information will be exploited towards the goal of education and not simply entertainment.

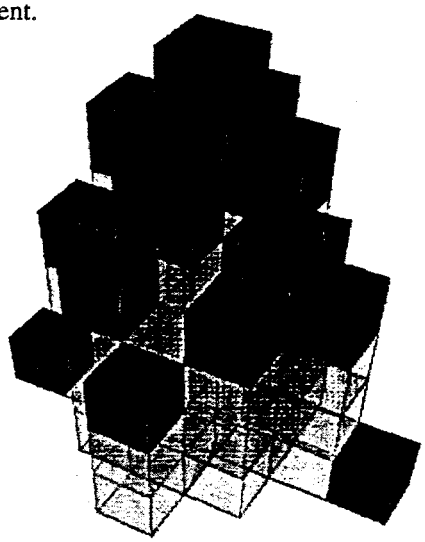


Figure 5. One method for improving the efficiency of VAFP would be to calculate the propagation of the flame only at the edges of the flame. Here is an example of a flame, the edge nodes are solid. The current algorithm computes propagation for both the clear and solid nodes. Therefore, eliminating the calculations for the clear nodes could give a large improvement on procesing time.

Development and Fabrication of Microelectromechanical Optical Switches

Jack C. Barr, Westley D. Cox, and John J. Wolfgang

Microelectronic Systems Research Center
Department of Electrical and Compu ter Engineering
West Virginia University
Morgantown, WV 26506-6104

Advisors:
Lawrence A. Hornak, Ph.D.
and Kolin Brown

A variety of microelectromechanical (MEM) cantilever beam devices suitable for optical beam deflection studies have been successfully designed and fabricated on silicon substrates using standard thin film microfabrication techniques. Emphasizing process simplicity over minimization of beam deflection voltage, processes for fabrication of electrostatically deflectable cantilever beams have been developed in which the cantilever deflection cavity is established in the silicon wafer through control of simple isotropic or anisotropic etches. Preliminary measurements of isotropically etched devices with chrome/oxide cantilevers indicate deflection magnitudes of approximately 6 μm with an applied voltage of 75 Volts. The family of cantilever devices and microfabrication processes developed will serve as a foundation both for integration of MEMs technology into existing microfabrication courses and subsequent research merging MEMs and photonic technologies.

INTRODUCTION

Micro-Electro-Mechanical Systems (MEMS) technology is currently employed in many applications including resonant gate transistors, electromechanical switches, integrated accelerometers, deflectable mirror arrays and neural probes (Younse, 1993; Petersen, 1982). MEMS technology takes advantage of the unique mechanical properties of silicon and other thin film materials used in conventional silicon integrated circuit fabrication to manufacture with high yield a multiplicity of miniature mechanical devices, potentially integrated directly with active electronic control and signal processing circuitry. Recent research efforts seeking to merge MEMS technology with integrated optics adds photonic capabilities (e.g., optical switching, displays, tunable filters, etc.) to the electronic and mechanical functionality of traditional MEMS.

Among the most fundamental microelectromechanical system, the cantilever beam also provides the most basic means of electrostatically controlling the deflection of an optical path which is at the heart of any optical switching function. This paper describes work completed towards developing a set of processes for the microfabrication of MEM cantilever devices. The process knowledge as well as the family of MEMS devices developed are intended to serve as building blocks both for subsequent merging of current research in integrated optics with MEMS capabilities and for development of a microfabrication course drawing upon MEMS technology as an instructional tool. Following the next section in which the basic construction and theory of micro-cantilever structures is introduced, the developed microfabrication processes and devices are described. Preliminary results from isotropically etched cantilever devices are then presented with the final section concluding with a discussion of future directions for the research.

CANTILEVER MEMS

In its simplest form, a microelectromechanical cantilever beam system consists of a thin film beam of length l extending over an open well region in the silicon substrate of depth d. Figure 1 illustrates two such cantilever structures differing only in the etch process with which the well is formed. To enable electrostatic deflection of the beam, the beam is composed of both an insulating silicon dioxide layer of thickness t and a metal thin film. The metal beam layer then acts as one electrode and the doped silicon substrate as the second electrode with the silicon dioxide layer acting as an interelectrode insulator. The electrostatic force arising from a voltage applied to the electrodes will cause the beam to deflect. For an ideal structure, the threshold voltage, V_{th}, for deflection is given by (Petersen, 1978)

$$V_{th} \cong \sqrt{\frac{3Ed^3t^3}{10\varepsilon_o l^4}} \qquad (1)$$

where E is the Young's modulus of oxide (7.2×10^{10} Pa) and ε_o is the permittivity of free space. Electrostatically controlled microelectromechanical cantilever beams have been fabricated previously utilizing a highly doped silicon epitaxial layer as an etch stop for formation of the cantilever beam deflection cavity (Petersen, 1978) and demonstrated as deflectable mirrors for display applications. Favoring reduced fabrication complexity over moderately higher deflection voltages when used in a research or instructional setting, the microfabrication of MEMs cantilever beams on silicon is reported here without definition of the beam deflection well depth using highly doped etch stop layers but rather with control of the etch times of the etchants which form the well.

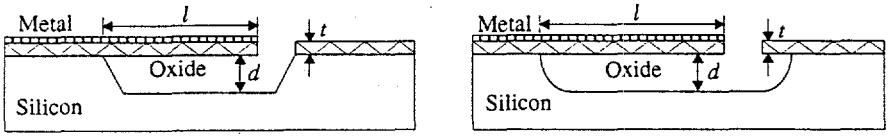


Figure 1. A basic metal/oxide beam cantilever beam structure formed by anisotropic (a), and isotropic (b) etching.

DESIGN AND FABRICATION

Masks were designed to microfabricate three different geometries of cantilever structures using the same mask set. The rectangular, T-beam, and pivoting beam mask geometries are illustrated in Figure 2 with the two mask layers of the process indicated. Basic rectangular beam cantilevers were emphasized on the mask set with individually addressable 1x10 linear arrays of cantilevers arrayed with lengths varying from 10 to 100 μm and widths varying from 5 to 25 μm. T-beam devices of different sizes whose free end was enlarged relative to the attached portion of the beam were similarly arrayed. The pivoting beam devices shown in Figure 2c have an axis of rotation in the middle of a rectangular beam and can be deflected in two directions about that axis. More speculative than the basic rectangular and T-beam structures, the pivoting beam structure had only three different sizes designed in order to allow more area for the basic structures. With the set of cantilever geometries defined in the mask set, the structure of the cantilevers themselves was further varied through experimentation with the oxide thickness, metallization, and beam well etch depth used in the microfabrication process.

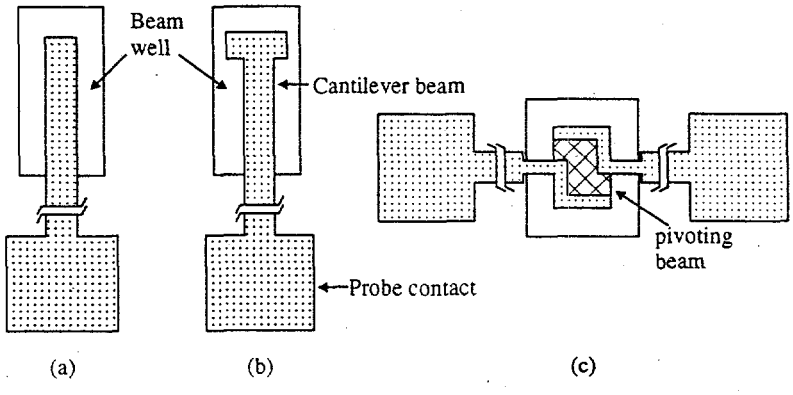


Figure 2. Basic mask geometries for the rectangular (a), T-beam (b), and pivoting beam (c) cantilever MEMS.

A schematic of the entire process flow is shown in Figure 3. The beginning substrates for the process were (100) oriented, n-type arsenic doped 50 mm diameter silicon wafers. Following chemical surface cleaning and an H_2O:HF (100:1) dip to remove native oxide, the wafers were thermally oxidized at 1050°C in a wet O_2 environment. A diagram of the resulting structure is shown in Figure 3a. The oxidation time was varied in order to produce oxide thicknesses ranging from 2000 to 4500 Å. Following oxidation and cooling to room temperature, the wafers were prepared for application of photoresist by vapor treatment with hexamethyldisilazane (HMDS) in a standard vapor priming process at 150°C in order to promote adhesion of the photoresist film. A 0.92 µm thick film of AZ4110 photoresist was then applied by spin casting to the primed wafers and the wafers prebaked for 1 minute on a 90°C hotplate. The pattern of the first process mask which defines both the oxide cantilever beam and well was transferred to the photoresist film by 310 nm wavelength ultraviolet contact lithography (35 second exposure at 20 mW/cm2) and subsequent photoresist development in a 4:1 (H2O:AZ400K) developer solution for 90 seconds. Each wafer was then hard baked in a 120°C convection oven for 30 minutes. The photoresist pattern was then transferred to the thermal oxide layer by wet etching in a 10:1 buffered oxide etchant bath at room temperature, see Figure 3b. The time required to completely etch the oxide varied from 4.1 to 8.9 minutes for the range of oxide thicknesses used. With the removal of the photoresist film by reactive ion etching in an oxygen plasma, the first mask level of the cantilever MEMS process was completed.

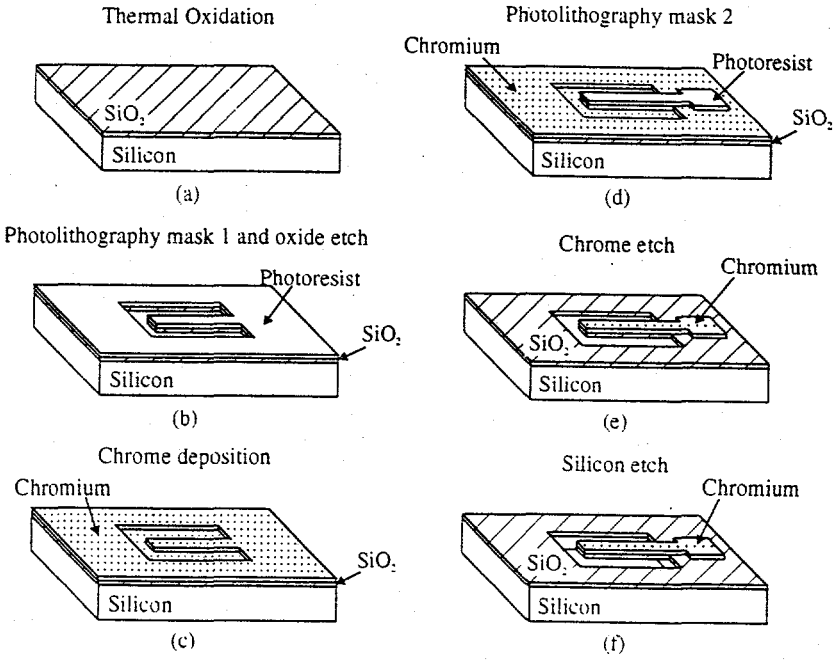


Figure 3. Process flow for cantilever MEMS fabrication.

In preparation for wafer metallization, the wafers were once again chemically cleaned and then dried at 400°C prior to immediate loading into a dc magnetron sputtering system for deposition of the metal film. Extensive experimentation was done with aluminum, copper, and chrome metallization, however chrome was ultimately selected for its superior adhesion and chemical resistance to both the anisotropic and isotropic wet chemical etches. Chrome films ranging from 100 to 750 Å in thickness were sputter deposited, shown in Figure 3c, and then immediately coated with photoresist as described above in order to promote good adhesion of the resist to the chrome layer. The metallization pattern of the second mask was then aligned with respect to the first mask features on the wafer and transferred to the photoresist with the same lithographic process as the first mask with the exception that a development time of 50 seconds was used. Next, the chrome not covered by photoresist was etched away by placing the wafers in a room temperature chromium etchant solution for 15 seconds to 1 minute depending on film thickness with the resulting structure diagrammed in Figure 3d. Following the chrome

metallization etch, the resist was again stripped with an oxygen plasma. Additional cleaning of the wafers using an acetone and ultrasonic bath was found to be necessary at this point due to what was believed to be precipitate particles formed on the wafers from the chrome etch solution. As the process was developed, it was found that proceeding directly to etching of the silicon substrate to create the deflection well and release the cantilever beam resulted in large variability in the degree of built in vertical deflection which the beams would exhibit upon release. The built in deflection of a cantilever beam without an applied voltage is illustrated in Figure 4a. It was experimentally determined that annealing of the patterned chrome metallization at 500°C for 30 minutes in a nitrogen environment resulted in less built in deflection magnitude as well as more uniformity in this deflection across a given wafer. While more controlled, the built in deflection was still significant. For rectangular cantilevers 5x90 μm in size, this zero bias deflection was measured to be as much as 60 μm.

Following cooling of the wafer to room temperature after anneal, the silicon wafers were etched to define the deflection wells and release the cantilever beams, see Figure 3e. The previously patterned oxide layer served as the etch mask for this process. During process development, both an anisotropic <100> silicon etchant (ethylene diamene pyrocatecol) as well as isotropic silicon etchant (25:3:10 HNO_3:HF:$HC_2H_3O_2$) were employed to release cantilevers. To date, all tested MEMS, including those whose testing is reported here, have been fabricated using the isotropic etch process. For this process, wafers were immersed with mild agitation in a room temperature bath of the isotropic etchant for 60 seconds. Following completion of the etch, the wafers were given a final bath rinse in deionized water. Due to the potential for damage to the beams, the wafers were not blown dry with nitrogen as is usually the case, but instead, after allowing excess water to run off the wafers, were placed in a 125°C convection oven to completely dry prior to cleaving and testing. The photomicrograph shown in Figure 5a shows a complete 1x10 array of chrome/oxide cantilever beams, 5x30 μm in size, fabricated with the described process.

CANTILEVER MEMS TEST RESULTS

Upon completion of processing, the wafers were cleaved with a diamond scribe and prepared for electrical probing. A specialized probe setup was created due to the low optical resolution of standard electronic probe station microscopes. Using a Nikon microscope with an enlarged area x-y-z sample translator, a test setup was constructed which enabled clear viewing of the smallest of the cantilevers while also allowing contact of a probe to the chrome electrode of the device and wafer back side during translation of the sample under the microscope. The microscope was fitted with a CCD camera which was input to a VCR in order to make a video record of cantilever device operation. The test setup also consisted of a voltage supply and digital multimeter for applying and measuring the voltage across the device under test. To date, numerous cantilever MEMS fabricated on a wafer with an oxide thickness of 3200 Å, a chrome thickness of 200 Å, and a well depth of 15 μm have been tested due to the overall better device performance seen in this most recent wafer sample. Basic rectangular cantilever devices as small as five μm wide and 40 μm long and as large as 10 μm wide and 100 μm long have exhibited deflection with an applied voltage in the range of 50 to 75 volts. This large deflection voltage is consistent with both the built in beam deflection described above and the relatively deep deflection wells of these devices compared to previous work described earlier using etch stop layers to define well depth. With an isotropic etch, the well depth is dictated by the time it takes for the etch to progress laterally and release the cantilever beam structure. At the same time the etch is proceeding laterally, it is also etching vertically at approximately the same rate. As shown by the isotropic etch lines depicted in Figure 4b, the etch depth directly under the center of the beam will therefore lag the depth elsewhere in the well by approximately half the beam width. The resulting peak in the silicon on the well floor directly under each beam is clearly evident in

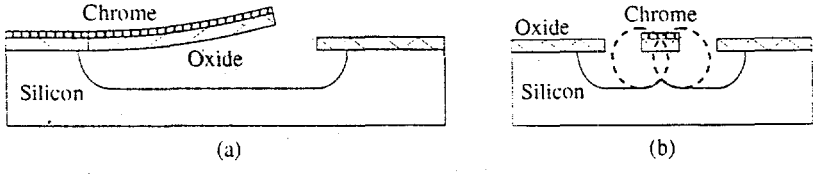


Figure 4. Illustration of (a) released cantilever built in deflection, and (b) the formation of a peak under the beam during the isotropic etch process.

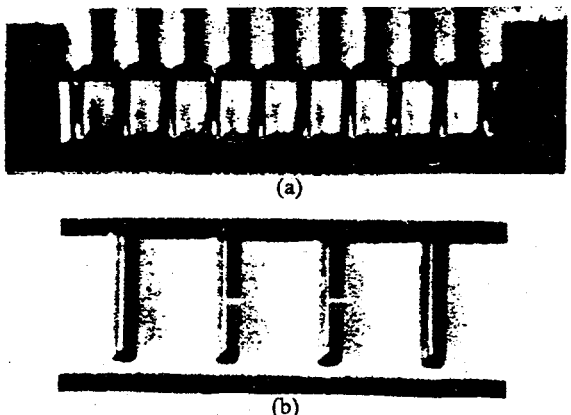


Figure 5. (a) Completed 1x10 array of chrome/oxide cantilever beams released with an isotropic silicon etch. Each rectangular beam is 5x30 μm. (b) Etch peaks directly under the cantilever beams arising from the isotropic etch.

the photomicrograph of Figure 5b. Given the built in deflection of the cantilevers (Figure 4a) which increases the distance d in Figure 1, the presence of these peaks directly under the cantilever structures may have prevented the necessity for much larger deflection voltages. This is a consequence of the higher electric field that would be expected to exist from the reduced gap between the beam and isotropic etch peak. While the short 60-second etch used was well suited for the release of these narrow width cantilever beams, the etch time was not adequate for the lateral component of the isotropic etch to release the wider T-beam and pivoting structures. Since the etch is the final process step and both oxide and metal layers which serve as etch masks are still in place, the same wafer from which test results have been obtained on the narrow released cantilevers can now be placed back in the etch bath and the etch allowed to progress until the larger structures are released. Although the longer etch time necessary for release of these wider beams will result in deeper wells, the correspondingly higher etch peak under the wide cantilever beams is expected to help in maintaining lower deflection voltages in these devices.

SUMMARY

Processes for fabrication of electrostatically deflectable rectangular, T-beam, and pivoting cantilever beams were developed based upon a simple two-mask process with timed wet etches as a means to control the deflection well depth. Measurements of isotropically etched devices with chrome/oxide cantilevers indicate deflection magnitudes of approximately 6 μm with an applied voltage of 75 Volts. Ongoing efforts focus on reducing the built in deflection of the beams thereby reducing the deflection voltage. This will be achieved through both experimentation with chrome and oxide film thicknesses as well as investigation of bimetal films. Work will also extend to evaluation of the more complicated T-beam and pivoting beams once testing of the narrow beams for the deflection characteristics and effectiveness as an optical switch is completed.

ACKNOWLEDGMENTS

We would like to thank the Microelectronic Systems Research Center for providing us with the Silicon substrates, and for allowing us to use all of the materials and equipment housed in the MSRC laboratories. This work was also partially supported by the West Virginia University Department of Electrical and Computer Engineering under a senior design project grant.

REFERENCES

Younse, Jack,. IEEE Spectrum, Mirrors on a Chip, Vol 30, Nov. 93. pg. 27-31.

Petersen, Kurt.. Proceedings of the IEEE, Silicon as a Mechanical Material, Vol 70, May 1982. pg. 420-456.

Petersen, Kurt,. IEEE Transactions on Electron Devices, Dynamic Micromechanics on Silicon: Techniques and Devices, Vol ED-25, Oct 1978. pg 1241-1250.

Petersen, Kurt.. Applied Physics Letters, Micromechanical light modulator array fabricated on silicon, Vol 31, Oct 1977. pg. 521-523.

DEVELOPMENT OF A METAL-SEMICONDUCTOR-METAL PHOTODIODE FOR USE IN A MICROWAVE PHASED ARRAY ANTENNA SYSTEM

Neil B. Davids

Department of Electrical Engineering and Computer Science
United States Military Academy
West Point, NY 10996

Faculty Advisor

MAJ Bradford Tousley

INTRODUCTION

The use of photodiodes in communications systems permits high frequency (microwave) transmission of communications signals. The use of photodiodes in phased array antennas will permit accurately timed signal transmission through the use of lasers. For this use, the photodetector needs have the ability to generate a large photocurrent, high sensitivity, and low dark current. The device bandwidth is critical to the use of microwave signals in the antenna to obtain accurate signal transmission. The phased array antenna will permit directional radiation of microwave signals in order to communicate on the move and to be able to perform the role of rangefinding without being hampered by smoke and fog like today's laser based systems.

METAL-SEMICONDUCTOR-METAL BASICS

The photodiodes under study were Metal-Semiconductor-Metal (MSM) Low Temperature Molecular Beam Epitaxy (LTMBE) grown GaAs semiconductor photodiodes. This type of semiconductor was chosen for its ease of fabrication, high sensitivity, and low dark current.[1] GaAs grown over a rang of temperatures (250°C - 500 °C) was processed into large area MSM photodetectors with high sensitivity and low dark current.[1] The short carrier lifetime of LTMBE GaAs may enable the use of photodiode driven microwaves in the antennas. To optimize the bandwidth and responsivity, the transit time must be reduced as much as possible.

After testing diodes grown at different temperatures and finger lengths, we found that the best growth temperature was 350 °C, and the best finger and gap widths 6.9 μm and of 2.9 μm, respectively, provided low dark current and a large photocurrent.[1] A large number of deep levels, or traps, occur because excess As forms antisite defects and Ga vacancies during the material growth. As a direct bandgap material, GaAs has a high quantum efficiency and optical responsivity, which is critical for efficient photodiode operation. A graph showing the different temperatures and their respective photocurrent and dark current can be seen in Figure 1.

Optical Power = 1 mW Bias Voltage = 5.0 V

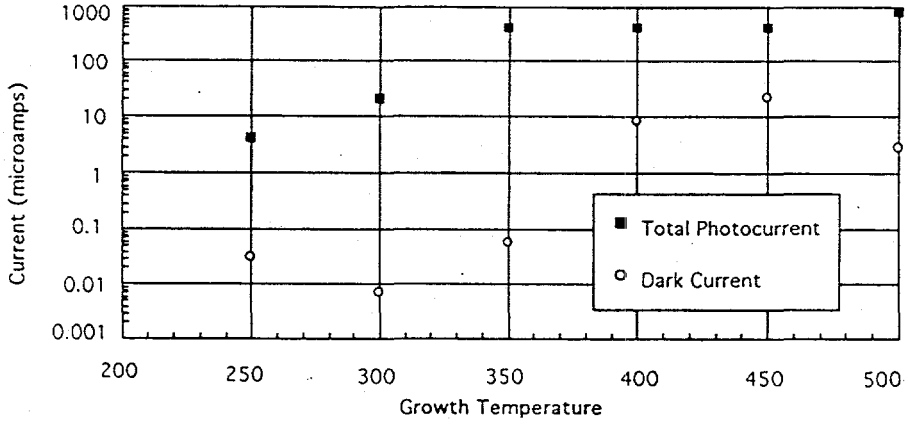


Figure 1: Current Vs. Growth Temperature for LTMBE GaAs .

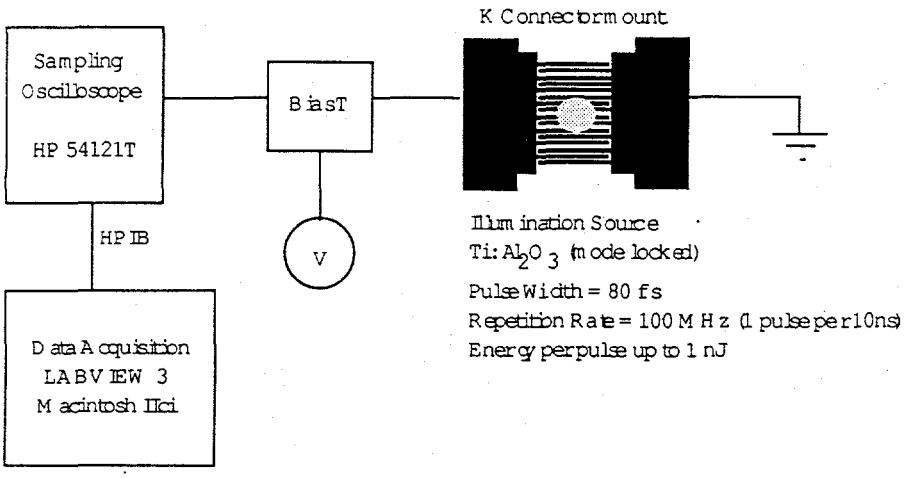


Figure 3: Test Layout

WAVELENGTH TEST

The wavelength was varied from 772 to 850 nm, at approximately 20 nm increments. The voltage was maintained at 6 volts, with the power maintained between 2 and 3 mW. These changes in wavelength do not have any effect on the performance of the diode, at least not in the range of wavelengths tested. The absorption curve for LTMBE GaAs is linear with respect to wavelength, especially over the relatively short range tested here. Testing at wavelengths around 1μm will most likely yield different results than we obtained.

BIAS TEST

The bias voltage was varied across the diode from 5 V to 30 V. The power was maintained at 4 mW and the wavelength at 814 nm. As the voltage was increased, several changes in the diode response was noted. The risetime decreased with respect to an increasing voltage, from 39.1 ps to 33.2 ps over a range of 25 V. Falltimes also decreased as the voltage was increased. For instance, the rise time at 5 V was 240.1 ps, while the risetime at 30 V was 39.6 ps, a reduction of one-sixth. This is to be expected, since the increased bias will increase the speed at which the carriers are swept across the electrodes. The faster the rise and falltimes, the more quickly the diode responds, the better the performance. As the voltage increased, however, the falltime decreases at a lower rate. This is most likely due to velocity overshoot, where the electrons going into the conduction band change momentum and move from the lowest point in the conduction band. Since the mass and the energy stays the same, the velocity must decrease to account for the change in momentum. At 30 V there is also a considerable amount of resonance in the waveform. This is most likely due to the inductor-like properties of the bond wire connecting the diode to the circuit at high bias voltages. The diode performance can be shown in Figure 4.

POWER TEST

Next, we varied the power of the incident beam from 1 to 20 mW, at a fixed bias of two volts, and a wavelength of 790 nm. Increasing the power has an effect of increasing the falltime of the response of the diode. This is what we expected, since the increased power introduces more photons, creating more carriers. Increasing the power, however, did increase the falltime, which means that the diode takes slightly longer to go back to the equilibrium state. This is due to the higher number of carrier pairs that must recombine. Another reason for this phenomenon is the concept of space charge shielding. This shielding is a result of the large number of carriers which
create a bias in and of themselves. This causes the carriers to recombine more slowly, since the bias prevents some carriers from combining immediately. At 20 mW, the diode was able to switch over 25 % of the bias voltage, which means the diode is conducting 25 % of the total current possible, which indicates a highly efficient photodiode. This only occurs when the power is increased to large levels, therefore power creates more carriers , a higher current, and therefore a higher voltage. The waveforms can be seen in Figure 5.

The MSM diodes tested were constructed in an interdigitated electrode pattern, as shown in Figure 2. This pattern permits more surface area between the two contacts than a simple face to face construction, resulting in larger photocurrent generation and a lower device capacitance by an order of 4 compared to conventional p-i-n diodes. This lower capacitance lowers the RC time constant so that the device is not limited by it. The finger length and spacing affects the transit time by the distance traveled by the carriers generated, and this is a factor we want to reduce. However, we want to balance this with the increased capacitances that closer fingers would create. The surface area incident to the laser is also an important parameter to keep in mind, so enough carriers are generated to conduct the photocurrent.

The function of the diode is relatively straightforward. A laser pulses photons onto the GaAs. The incident photons give their energy to electrons located in the valence band of the GaAs. The electrons are transferred to the conduction band, and the holes are left in the valence band. Once the electrons are in the conduction band, the bias applied across the electrodes sweep them across the gap, while the holes are swept in the opposite direction. This creates a current across the gap, which causes the voltage across the gap to short out. The holes and electrons then recombine after a short time.

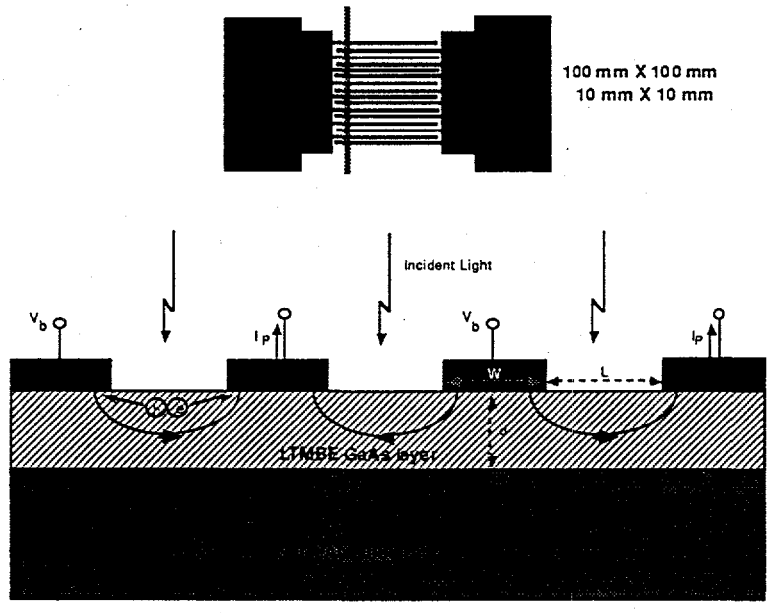


Figure 2: MSM Diode Construction

TESTING EQUIPMENT

The impulse response technique was used[2] with a mode-locked Ti-Al_2O_3 laser impulse source, tunable in near infrared, with an optimal pulse width of 80 fs and a repetition rate of 100 MHz. We were able to obtain up to 100 mW of average usable optical power. The test set up is shown in Figure 3.

It is not possible to measure the pulse width of the beam with conventional electronics due to its extremely short temporal duration. To determine the pulse width, we used an autocorrelator and a spectrometer. A 20 GHz HP 54120 digitizing sampling oscilloscope was used to collect the impulse response, and the devices were connected with a K-connector through a bias T. The response measured was the amount of signal voltage shorted out over the gap. The data was downloaded through Labview III on to a Macintosh computer for storage and analysis. We conducted parametric study by changing the wavelength of the laser, the optical power, and the voltage bias on the diode; and then analyzed the waveforms.

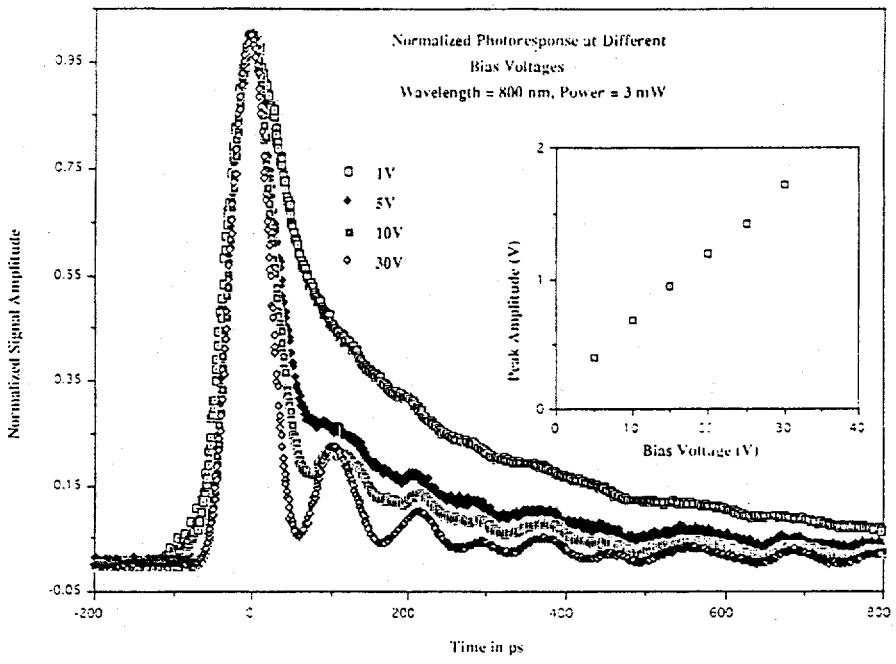


Figure 4: Different Photoresponses at Different Bias Voltages

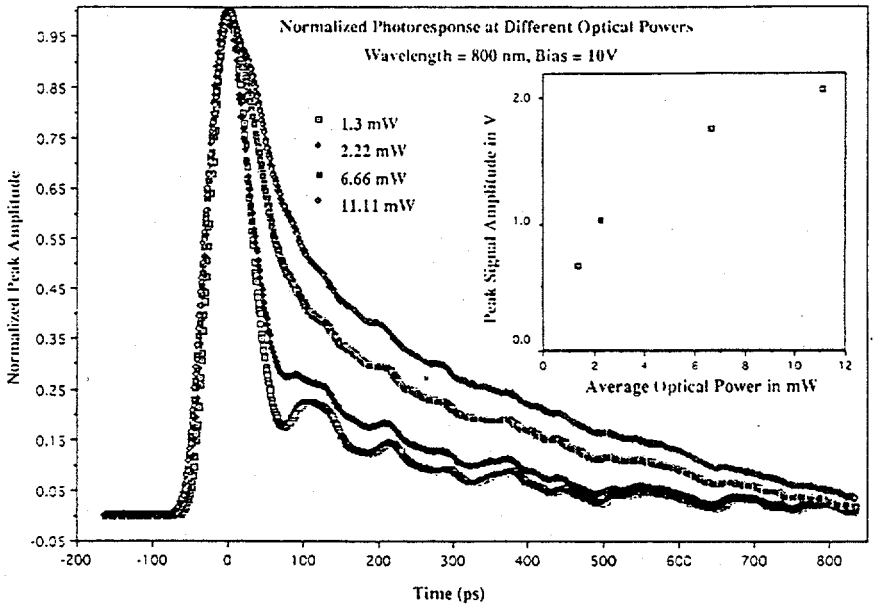


Figure 5: Photoresponses at Different Optical Powers

OPTIMIZATION

After looking at all the data collected, the next step was to find out which combinations of bias voltage and optical power would achieve the best bandwidth response of the diode. The equation used to find the bandwidth was

$$BW = \frac{0.35}{\Delta\tau} \quad (1)$$

where τ is the FWHM of the respective waveform. Using this equation, the bandwidth ranged from 4.88 GHz at 8 V and 3 mW to 6.20 GHz at 30 V and 4 mW. These results show that a high bias, coupled with a relatively low power, gives the best bandwidth. The lower power prevents the effects of space charge screening, and the high bias keeps the sum of the hole and electron velocities high.

Another point of interest is the responsivity of the diode. We used the definition of mV per pJ to obtain our results. The voltage used was the peak voltage shorted across the gap. The power was taken as the average power divided by the repetition rate of the laser (100 MHz), and then reduced by a factor of 0.25 to account for reflection losses, shadowing, and Fresnel losses. We found that at the high bias low power test was again the most successful. At 30 V and 4 mW, the responsivity of the diode was 129.1 mV/pJ. This is a better response than any photodiode grown at the normal 600 °C temperature. This shows the advantage of using LTMBE GaAs in diodes that need a good response time, plus the added benefit of low dark current.

Another advantage of MSM diodes are that they have a small amount of gain in the current generated. This is a result of the holes traveling more slowly than the electrons they are generated with. Since the electrons travel faster across the gap, there is a charge mismatch once they reach the electrode. To compensate for this, the electrode at the far end of the gap supplies more electrons to recombine with the holes, which creates more current.

FUTURE TESTS

Some other types of experiments that could be conducted in the future include a test with other types of lasers and a threshold voltage test. Other types of lasers will give us more information about the wavelength response of the photodiodes. Another test that is planned is taking larger diodes and conducting a spectrum analysis of the response of the diode, in addition to the time domain analysis. Spectral analysis will give us the ability to see the response at different harmonics that occur at larger frequencies. This is possible due to the high responsivity of the detector keeps the signals at the harmonics to be at high enough levels to be studied.

CONCLUSIONS

. By using diodes constructed with smaller finger widths and gaps on the order of 1 μm, the potential for greater response times exists. The diodes can be optimally designed to respond to certain voltages and powers to give the best response time for a specific pupose. This is necessary in order to make these diodes compatible with certain systems. There is an important need for diodes of this type, especially in the use of phased array antennas. Other possibilities exist for diodes with the same characteristics of low dark current and large bandwidth. Testing and analysis, along with the construction of more diodes of different sizes, will give the promise of implementing these diodes into a variety of different devices.

REFERENCES

[1] B. Nabet , "Intermediate Temperature Molecular Beam Epitaxy Growth for Design of Large Area Metal-Semiconductor-Metal Photodetectors".

[2] P. Battacharya, Semiconductor Optoelectronic Devices, Englewood Cliffs, New Jersey: Prentice Hall, 1994.

[3] R. Pierret, "Semiconductor Fundamentals", Reading, Massachusettes: Addison-Wesley Publishing Company, 1988.

[4] Ito, Masanori and Wada., "Low Dark Current GaAs Metal-Semiconductor-Metal (MSM) Photodiodes Using WSi_X Contacts", *IEEE Photonics Journal of Quantum Electronics* Vol Qe-22, No. 7 July 1986.

[5] J Burm, K. Litvin, W. Schaff, and L. Eastman, "Optimization of High-Speed Metal-Semiconductor-Metal Photodiodes". *IEEE Photonics Technology Letters* Vol. 6 No. 6, June 1994.

Negative Output Resistance Voltage Source for Device Testing

Chadwin D. Young

Department of Electrical and Computer Engineering
University of Texas at Austin
Microelectronics Research Center
Austin, TX 78712

Faculty Advisor:

Dr. Dean P. Neikirk

Negative differential resistance (NDR) is achievable in modern semiconducting devices. The NDR region of the current-voltage (I-V) characteristics of these devices is characterized by a negative slope in the I-V curves. This NDR region is attractive in the design of high speed oscillator circuit because it provides the possibility of introducing a negative resistance element in the feedback path of such oscillators. Unfortunately, this attractive feature can make it difficult to simply measure the I-V characteristic in the NDR region. Conventional I-V measurement set-ups involve injecting a current into the device under test (DUT) and measuring the resulting potential produced across the device. One way to conceptualize such a measurement is to think of a load line sweeping across the I-V characteristic. Such load lines display a negative slope. In order to measure an NDR region, a load line with a positive slope is required. In this presentation, the design and construction of a negative output resistance voltage source (NORVS) is described. The NORVS can be used to produce a load line with a positive slope. The application of this source to the measurement of the I-V characteristic of a double barrier resonant tunneling diode (DBRTD) is discussed.

INTRODUCTION

Semiconductors play an integral role in our lives today. These devices are diverse in composition and design. One device of particular interest is the Double Barrier Resonant Tunneling Diode (DBRTD). The concept of the DBRTD was proposed by Leo Esaki in 1957. This device can be used in memory applications or in microwave oscillator applications depending on how large the hysteresis is in the current-voltage (I-V) curve.

Theoretically, showing this hysteresis is quite easy. By simply representing the I-V curve in the form of a mathematical equation, one can easily graph the equation. Experimentally, showing the I-V diagram is difficult because it is hard to determine the area of the curve that contains the hysteresis when using a load line.

Some research has been done to try to alleve this problem. One technique is to use a positively sloped load line instead of a negatively sloped load line. The load line determines the ideal operation point for a device. The load line determines for a given voltage how much current passes through the device under test.

To create this positively sloped load line, a negative impedance converter is required. This is a simple operational amplifier circuit that can serve as a voltage source and negative series resistance.

In this document, a possible solution to experimentally showing the I-V diagram will be explored with some supporting data to verify this possibility.

BACKGROUND

The unique I-V characteristic of the DBRTD is due to a quantum mechanical process called tunneling. So, to prepare the reader of this document for later material, the tunneling process for the DBRTD will be described to better understand its I-V curve.

Tunneling is a process that cannot be described by classical mechanics. The tunneling process is best described by the potential well problem with finite barriers and the Schrödinger Wave Equation. Tunneling is the quantum mechanical process of an electron "tunneling" through a barrier of finite height and thickness.

This "tunneling" is the result of the normalized wave equation evaluated between $-\infty$ and ∞. The original wave equation is determined by the finite potential well and the properties of the Schödinger Wave Equation. Also, the answer to the normalized wave equation is one (1) because nothing can be outside the evaluated range. This probability means that an electron can be found somewhere along the wave propagation defined by the wave equation at any given time. Unlike an infinite potential well where there is no possibility for the wave equation to extend outside the infinite barriers, a potential well with barriers of finite height and thickness does not require that the wave stay between the barriers. Therefore, the wave extends beyond the finite barriers. Since the wave extends beyond the barriers, an assumption can be made. This assumption is that the distance the wave extends can be $-\infty$ and ∞ since the wave distance

is much greater in relation to the barrier thickness. This allows a particle to exist outside the barrier due to the fact that there is a possibility that an electron can reside on the wave propagation outside the finite potential well.

Three things are necessary for tunneling to occur. First, a reservoir of electrons with energy to tunnel is required. Next, energy states must be available for electrons to reside after tunneling. Lastly, the actual chance for tunneling to occur must be great (i.e. a thin barrier). For an optimum chance, a small distance in angstroms is a necessity. To acquire this small distance, a process called Molecular Beam Epitaxy (MBE) is used. MBE is the act of growing thin layers of crystalline material on a substrate. The thin layers produce the barriers of the finite potential well.

From this MBE process, the double barrier resonant tunneling diode (DBRTD) can be fabricated (Figure 1). This device displays a unique I-V curve that is hard to represent experimentally. Although this is the case, an understanding of what the curve is supposed to look like helps when finding a way for experimental representation. From the potential well problem, one knows that there are specific energy levels in the well. The greatest chance for the tunneling process occurs at these energy levels.

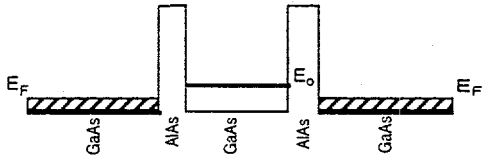


Figure 1. The Double Barrier Resonant Tunneling Diode Structure

To approach an energy level within the well, the voltage across the DBRTD must increase. Increasing the voltage causes movement in the conduction band of the device. From semiconductor physics, the conduction band contains the reservoir of electrons and an ample supply of empty states. As the conduction band shifts due to increased voltage, the electrons approach an energy level within the well. As electrons approach the energy level, the chance for electrons to tunnel is greater. Therefore, increased movement of electrons means increased current. This process will cause an almost positive linear progression in the I-V graph (Figure 2).

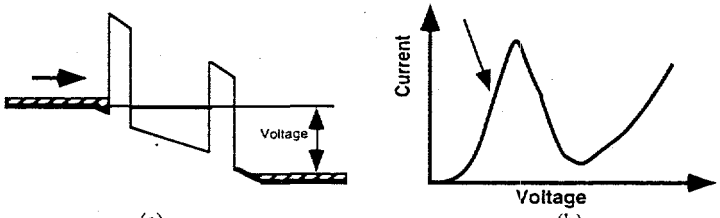


Figure 2. An increased voltage bends the conduction band (a) allowing electrons to tunnel through the quantum energy state creating an increase in current (b).

As the voltage continues to increase, the reservoir of electrons moves past the energy level of the well. Thus, the chance for tunneling diminishes which means that the current also reduces. This process will cause an almost negative linear progression in the I-V graph (Figure 3). This is known as the negative resistance or hysteresis portion of the curve.

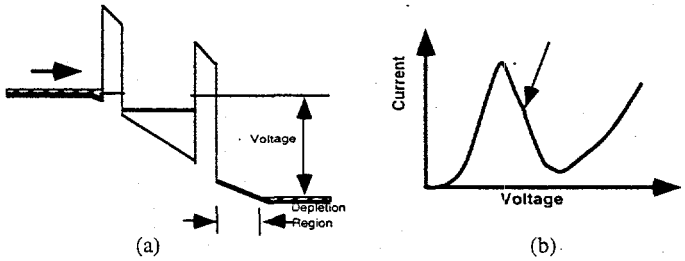


Figure 3. Continuing to increase the voltage creates more band bending (a) which moves past the tunneling state; therefore, decreasing the current (b) because electrons cannot tunnel.

Continuing to increase the voltage will eventually cause the conduction band to become level with the first barrier of the well. This allows electrons to fall into the well and tunnel

through the second barrier at the specific energy level. So, the current will again increase, and the I-V graph will also increase (Figure 4).

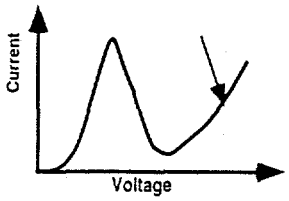


Figure 4. Continued increasing of voltage bends conduction band enough to be level with the first barrier. Here electrons "roll" into the well and tunnel at the first opportune energy level which increases current flow again.

PROBLEM

The major problem with experimental representation for the DBRTD is the negatively sloped load line. Although this type of load line gives researchers problems when trying to depict I-V curves of the DBRTDs, it does have some commercial applications. Two of these applications are oscillators and memory devices.

But, the primary purpose for load lines are the ability to find the operating conditions for given currents and voltages. The problem with negatively sloped load line occurs when the line intersects an I-V curve of a DBRTD in more than one place (Figure 5). When this happens, hysteresis in the curve cannot be determined.

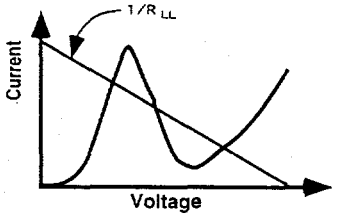


Figure 5. Example of a load line across a DBRTD.

A solution to the problem is to generate a positively sloped load line. This can be done by constructing a negative output resistance (NOR) circuit.

PROJECT

First of all, a circuit representation of the negative load line is needed. From the circuit diagram below (Figure 6), the equation (1) for the load line is:

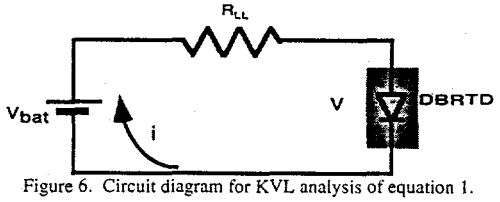


Figure 6. Circuit diagram for KVL analysis of equation 1.

$$v = V_{bat} + iR_{LL} \qquad (1)$$

After solving for the current, the slope can be read from the equation. So, $-1/R_{LL}$ is the slope with V_{bat}/R_{LL} as the y-intercept.

RESEARCH CONDUCTED

A NOR can create a negative resistance that can cancel the negative slope to change the signs and create a positive slope. The NOR circuit used is given below.

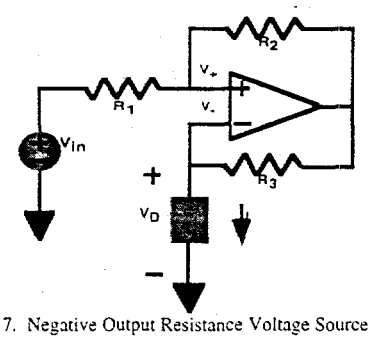


Figure 7. Negative Output Resistance Voltage Source Circuit

In order for the NOR to work properly, research was done to determine the ideal conditions for operational amplifier (op amp) optimum performance. In an ideal op amp, the input has high resistance while the output has low resistance. This means that there is very little input current going into the op amp. Because of this, v_+ and v_- can be assumed to be almost equal. Another researched criterion for the NOR is that the voltage across R_3 needs to be slightly larger than the voltage across R_2. A larger voltage across R_3 will insure that a negative overall voltage is applied across the device under test.

After a lengthy derivation to satify ideal op amp parameters and NOR criterion, the circuit equation (2) for the NOR is given below.

$$v_- = \left(\frac{1 + \frac{R_3}{R_L}}{1 + \frac{R_2}{R_1}} \right) v_+ + \left(\frac{\frac{R_2}{R_1}}{1 + \frac{R_2}{R_1}} \right) v_d \qquad (2)$$

Since v_+ and v_- are almost equal, the ratio of R_2:R_1 and R_3:R_L must be small.

The negative voltage will produce the necessary sign change if the NOR is plugged into the circuit (Figure 8) in place of R_{LL} and create the positive load line.

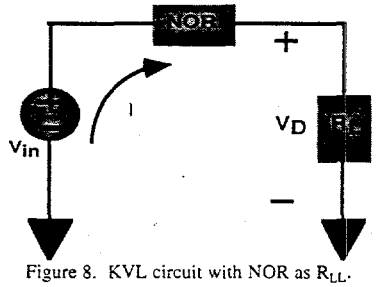


Figure 8. KVL circuit with NOR as R_{LL}.

Applying Kirchoff's Voltage Law (KVL) will generate the following equations.

$$v_d = vin + \left(\frac{R_1 R_2}{R_3} \right) I \qquad (3)$$

$$slope = -\left(\frac{R_1 R_2}{R_3} \right) \qquad (4)$$

Further research showed that the position and slope of the load line can be adjusted by varying v_{in} and R_1, respectively. The purpose for varying R_1 and v_{in} while staying within the op amp parameters is to move the load line about the graph. As v_{in} is varied, the intercept along the x-axis has the ability to move. As R_1 in varied, the positive slope of the load line changes.

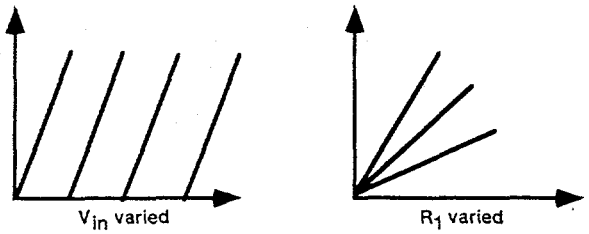


Figure 9. As v_{in} is varied, x-axis translation is achieved (a). As R_1 is varied, positive slope changes are achieved (b).

Applying the NOR to the DBRTD can demonstrate the advantages of using the NOR circuit to determine the negative resistance I-V portion of the device under test (Figure 10). The determination of this part of the curve is crucial to the development of devices because theoretical results can be validated by experimental results.

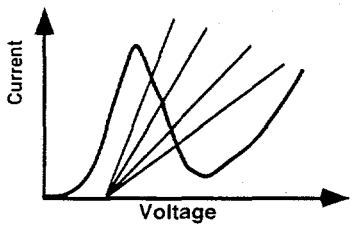


Figure 10. Application for the changing of V_{in} and R_1 to determine hysteresis.

Future research includes using two 8-bit digital to analog converters to regulate the changing of v_{in} and R_1. This would be done using a computer which will increase efficiency in testing because there will be less manual adjustments made by graduate students! Also, very unique I-V curves could be applied to the NOR for experimental verification (Figure 11).

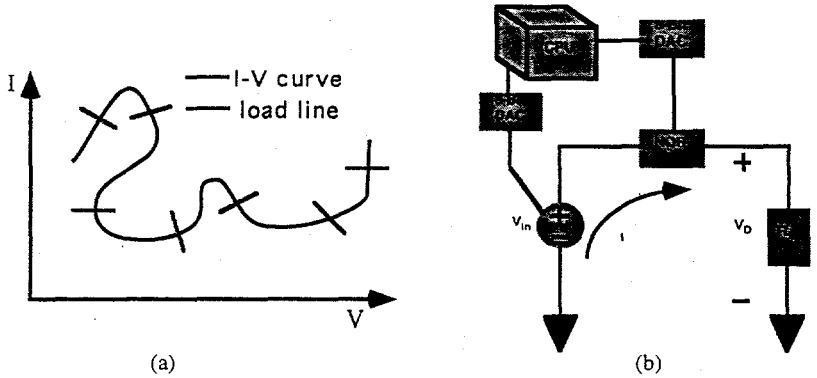


Figure 11. Possible I-V curve (a) needing new NOR technology (b).

CONCLUSION

The implementation of an NOR configured circuit to experimentally measure the NDR region of the DBRTD is only the beginning of breakthroughs in developing a way to look at the negative resistance portion(s) of an I-V curve. This would greatly help in designing new devices or improving old ones. The positive slope would help researchers determine how much hysteresis there is in a potential device that could have some commercial value.

REFERENCES

Goossens, J. G. ;. B., Stephen; Yu, Zhiping; and Dutton, R. W. (1994). An Automatic Biasing Scheme for Tracing Arbitrarily Shaped I-V Curves. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 13, 310-313.

Martin, A. D., Lerch, M. L. F., Simmonds, P. E., & Eaves, L. (1994). Observation of Intrinsic Tristability in a Resonant Tunneling Structure. App. Phys. Lett., 64(10), 1248-1250.

Sze, S. M. (1990). High-Speed Semiconductor Devices. New York: John Wiley & Sons, Inc.

Yngvesson, S. (1991). Microwave Semiconductor Devices. Boston: Kluwer Academic Publishers.

Optical-To-Optical Modulator Based on a Multiple Quantum Well Device

Jordan J. Fisher

United States Military Academy
Department of Electrical Engineering and Computer Science
West Point, NY 10996

Faculty Advisor:

Barry L. Shoop

INTRODUCTION

The large bandwidth and inherent parallelism of optics has led to extensive development of optical devices and systems over the past two decades. Most current applications only use optics for information communication or to perform a single processing function before the optical signal is converted into an electrical signal. As optical computing and other optical functions develop, the need to convert to an electrical signal will diminish. One problem with creating an all optical system is that different applications use laser sources that operate at different wavelengths. Long distant fiber optic systems usually operate at wavelengths greater than 1.2 μm to take advantage of minimum dispersion and attenuation of current fiber systems; whereas, most semiconductor driven optical systems operate at much shorter wavelengths. If complete optical systems are to be developed, it is therefore likely that the carrier wavelength will have to be changed at least once. This will create the need for high speed, energy efficient wavelength converters. This paper proposes one method of optical-to-optical modulation using a multiple quantum well (MQW) device.

THEORY OF MULTIPLE QUANTUM WELL DEVICES

Multiple quantum well devices are semiconductor devices that have been developed primarily over the last decade. Semiconductor technology over this period has advanced rapidly, especially in the area of electrooptic devices. While advances in technology and economics have maintained silicon as a competitive industry-wide standard for semiconductor devices, there has also been significant research leading to many improvements in the realm of III-V semiconductors, named for their atomic structures and their location on the Periodic Table of the Elements. These devices are popular for their low power consumption, high speed, and from the perspective of electrooptic devices, their higher internal quantum efficiency. Furthermore, III-V semiconductors enable scientists to design the bandgap energy of the semiconductor according to their needs. For these reasons many novel device structures based on III-V semiconductors, including multiple quantum well devices, have been achieved leading to continued advances in this field.

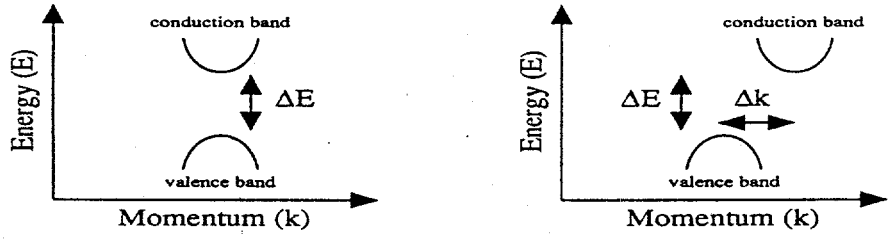


Figure 1. Energy versus momentum diagrams. (a) direct bandgap material and (b) indirect bandgap material.

III - V semiconductors are an important class of device because they are *direct bandgap* materials in contrast to silicon devices, which are *indirect bandgap* materials. Direct bandgap materials have their valence band and conduction band corresponding to the same momentum, $\bar{k}$. Indirect bandgap materials, however, do not. This difference is particularly significant. For an

electron in an indirect bandgap semiconductor to transition from one band to another, it requires a photon for excitation, and a phonon to change its momentum. This required phonon makes electron-hole pair excitation more difficult because it requires more energy and time. Figure 1 shows the relationship between momentum and energy for both indirect and direct bandgap materials.

Probably the most studied III - V semiconductors are gallium arsenide (GaAs) and its derivatives. These materials can be "doped" to be p-type or n-type like any silicon device. The most common MQW's are III - V devices, typically made with GaAs and aluminum gallium arsenide (AlGaAs), that are designed with regions consisting of very thin alternating layers of GaAs and AlGaAs. These layers of GaAs and AlGaAs have different energy band levels which create quantum confinement of the electron-hole pairs in the presence of an applied field. The layers of GaAs, which have a lower energy bandgap, become "wells," while the AlGaAs layers, with a larger bandgap, become "barriers."

As an electric field is applied across the terminals of the MQW, two effects occur. Since the semiconductor layers are extremely thin, the electrons and holes behave as particles. The electrons and holes move in opposite directions due to the Coulombic force of the field, but are confined by the quantum well and prevented from tunneling. Second, the electrons and holes move to lower energies as shown by the tilting of the energy bands in Figure 2(b). The resulting quantum confinement causes discrete energy levels of the electron-hole pairs which results in very strong absorption peaks at the edges of these steps. These absorption peaks are called *excitons*.

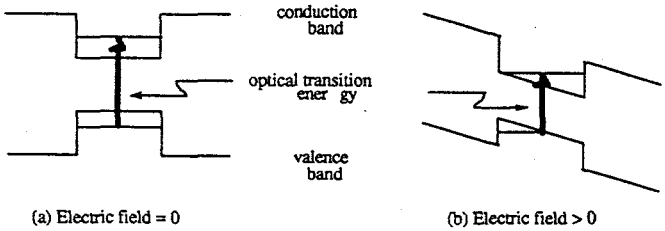


Figure 2. Effects of an applied electric field (E) on MQWs. (a) E = 0, (b) E > 0. Here, the length of the bold arrow represents the energy required to transition from the valence band to the conduction band.

Since the energy of light is directly proportional to its frequency, the effect of quantum confinement causes the optical transition energy to decrease and the absorption peaks to move to lower frequencies (corresponding to longer wavelengths) of light. The shifting of the absorption peaks with an applied field is the underlying principle in the Quantum Confined Stark Effect (QCSC) [3]. A standard measure of the optical absorption of a device is its responsivity (S) measured in amperes per watt. Responsivity relates the electric current flowing in the device to the incident optical power. Figure 3 shows the effects of an applied field on the responsivity of an MQW, and demonstrates how optical modulation can be achieved using a MQW device.

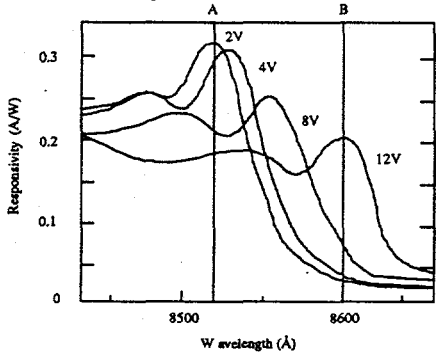


Figure 3. Responsivity versus wavelength for specific applied voltages. Notice that the absorption peaks shift to longer wavelengths as greater fields are applied.

The most common structure of an MQW modulator is the Self Electrooptic-Effect Device (SEED). Here the MQW is placed within the intrinsic region of a reverse-biased *p-i-n* structure.

This provides several advantages. First, large fields can be applied across the device with virtually no current flow. With higher applied fields, larger changes in absorption are achieved, and the device may be controlled through relatively small changes in voltage. Furthermore, the device is more energy efficient since reverse biased diodes have very low leakage current [2].

The most popular type of SEEDs are reflective type devices. This type of modulator has multiple quarter-wave stacks grown at the rear of the device. By allowing the light to pass through the quantum well twice, the reflective modulator has increased energy efficiency because the light has twice the opportunity to interact with the semiconductor. When a coherent light source is incident on the device, a phtotocurrent I_c is produced as with a standard photodiode. This can be regulated by either the wavelength of the incoming light, the intensity of the light, or by the voltage across the SEED.

There is a feedback loop created between the external bias circuit and the SEED. Increasing the amount of photocurrent created by the SEED decreases the voltage across the SEED. This in turn decreases the absorption of the MQW and thus decreases the photocurrent. This feedback loop can be described in the following two equations: $I = S(V,\lambda)P_{inc}$ and $I = I(V)$ where S is the responsivity of the SEED, V is the voltage across the modulator, and P_{inc} is the incident optical power. Using these two equations, a first order analysis of the device modeling the *p-i-n* structure as a capacitor results in:

$$C\frac{dv}{dt} = \left[\frac{dI_c}{dV} - P_{inc}\frac{dS}{dV}\right]v \qquad (1)$$

where C is the capacitance of the MQW, and v is the change in voltage about the equilibrium point [3]. This equation is a classic first-order differential equation which has exponentials as its solutions. If the bracketed term is positive, the feedback is positive and the system is unstable and grows exponentially. In Figure 3, point A represents an unstable operating point because $dS/dV < 0$, and point B represents a stable operating point because $dS/dV > 0$.

Figure 4 shows the circuit configuration of an optical-to-optical modulator based on a MQW device. When the SEED is operated at a wavelength corresponding to point A in Figure 3, any increase in the field across the device will lower its responsivity (absorption), and thus increase its reflection. The field across the device can be controlled using a reverse-biased photodiode and a voltage source. When there is no light incident on the photodiode, it acts an open circuit and there is no field across the SEED. When light is incident on the photodiode, current flows, the field across the SEED is no longer zero, and its absorption decreases. Using this technique, information on one laser beam of wavelength λ_1, incident on the photodiode, can be modulated onto a second laser beam of wavelength λ_2 incident on the MQW device. By applying a continuous wave beam incident on the MQW, the reflected beam will be modulated according to the beam incident on the photodiode. This mode of operation relies on the optical bistability of the SEED, so the information to be transferred should be limited to discrete power levels.

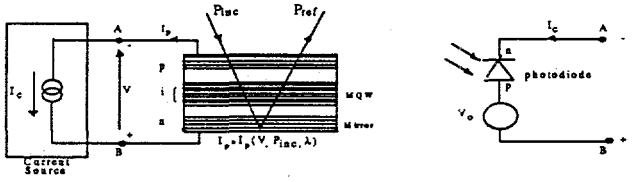


Figure 4. Optical-to-Optical Modulator

Besides bistability, the MQW devices have other useful modes of operation. One of these is self-linearized modulation and optical level shifting. When the device is operated at a wavelength corresponding to point B in Figure 3, the governing equation is:

$$P_a = \frac{h\upsilon}{\eta e} I_c \qquad (2)$$

where P_a is the absorbed optical power, and I is the current through the device, h is Planck's constant, υ is the optical frequency, and η is the internal quantum efficiency. This equation shows that the absorbed optical power is directly proportional to the current flowing through the series circuit. In order to make the current dependent only on some external circuit and independent of the internal voltage, the device must be biased with a constant current source [1].

This *linear* relationship has led to many possibilities. Noninterferometric optical subtraction can be obtained using this effect. Since $P_{ref} = P_{inc} - P_a$, and $P_a = h\omega/e\, I_c$, where I_c

is directly proportional to the power incident on the photodiode, the beam reflected by the MQW is directly proportional to the difference between the beam incident on the photodiode and the beam incident on the MQW [4]-[5].

Using the same circuit design in Figure 4, a linear, inverting optical-to-optical modulator can also be implemented. Information modulated on one laser beam of wavelength λ_1, incident on the photodiode, can be modulated onto a second laser beam of wavelength λ_2, which is incident on the MQW device. By applying a continuous wave beam of constant intensity incident on the MQW, the reflected beam will be modulated inversely proportionally to the beam incident on the photodiode. Information at λ_1 could be modulated onto λ_2, again effectively creating a wavelength converter. If the light incident on the photodiode were incoherent, an incoherent to coherent converter is possible. This would thus enable information from an LED source to be converted to a coherent source.

EXPERIMENTAL DESIGN AND SETUP

Figure 5 shows the experimental setup for verification of the optical-to-optical modulator. The design follows closely from the theory previously developed. Here, the SEED must be carefully reverse biased by a DC voltage and a photocurrent from an external photodiode.

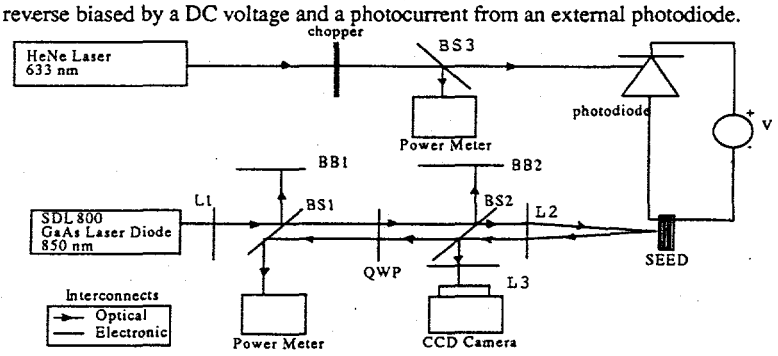


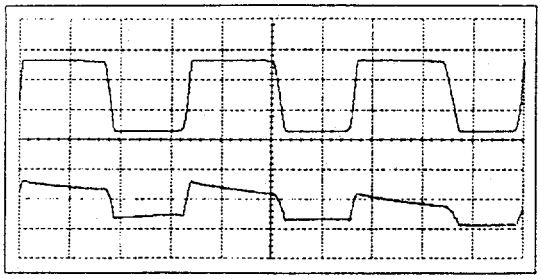
Figure 5. Experimental Setup (BB-Beam Block, QWP-λ/4 plate, BS1-Polarizing Beam Splitter, BS2-50/50 Beam Splitter, BS3-50/50 Beam Splitter, L1-20x Microscope Objective, L2- 10x Microscope Objective, L3-Lens)

The experimental setup was rather simple. The coherent source used to generate the continuous wave beam incident on the SEED was a single mode GaAs laser diode operating at 850 nm wavelength. It was collimated using a 20x microscope objective, and focused onto a 40μm x 80μm SEED using a 10x microscope objective. Two beam splitters were used in this set up. The first was a polarization beam splitter, and the second was a standard 50/50 beam splitter. The polarization sensitive beam splitter transmits vertically polarized light and reflects horizontally polarized light. A quarter wave (λ/4) plate was placed after the polarization beam splitter in order to give the vertically polarized beam a 45° polarization rotation each pass. This ensures that after the beam is reflected off the SEED, it is horizontally polarized so that it is reflected by the beam splitter and projected on to the power meter. The 50/50 beam splitter was used along with a lens and a CCD camera to observe the beam on the device for alignment purposes. The photodiode was connected externally to the SEED chip and the incident beam was completely separate from the SEED optical path. A HeNe laser at 633 nm was used as the beam incident on the photodiode. This beam was modulated using a mechanical chopper, which sequentially blocked and transmitted the beam at a set frequency up to 3 kHz.

EXPERIMENTAL RESULTS

Figure 6 shows measured results using the experimental setup shown in Figure 5. This data was measured using an oscilloscope which was connected to the two power meters. The top trace shows the input square-wave pattern from the modulated 633 nm HeNe laser while the lower trace shows the corresponding output square wave pattern from the 850 nm laser. It is clear from this data that the information from the HeNe was modulated on the 850 nm laser in a non-inverting manner. This result was expected since both the laser diode and the SEED heavy-

hole exciton were at 850 nm. The SEED was therefore operated in a bistable mode with the modulation caused by a change in electric field across the device.



Top Trace: Input Modulating Signal		
	Wavelength:	633 nm
	Channel 1:	5.0 volts/div.
		1.0 ms/div.
Bottom Trace: Output Modulated Signal		
	Wavelength:	850 nm.
	Channel 2:	0.5 volts/div.
		1.0 ms/div.

Figure 6. Experimental Results

While both lasers were operated with equivalent output power (1.3 μW), the modulated beams incident on the power meters showed approximately an order of magnitude difference. This difference can be explained by the combination of losses inherent in the optical path of the beam incident to the SEED, scattering losses, and the absorption of the SEED itself. Tracing optical losses through the lower path of Figure 5, a 50% loss each time the beam travels through a beam splitter is incurred, and a 7% Fresnel loss for each lens and the λ/4 plate. This total loss is 90.65% through the optical path accounting for the majority of the loss.

CONCLUSIONS

A new method of translating information from a light source of one wavelength to a laser of another wavelength has been presented in this paper. Such a device serves an important function, since different optical signal processing and communication processes often use different wavelengths of light. The proposed method offers several distinct advantages. First, it is very power efficient since there is little current flow. Second, it has the potential or being very fast. The speed of the MQW modulator is only limited by its carrier sweep out time. SEEDs as small as 10μm X 10μm have been reported in the literature with switching speeds as fast as 33 psec indicating performance well into the 10 GHz range. Finally, since most SEEDs are GaAs based devices like many other semiconductor devices, this design may be monolithically integrated on a chip with other electronic and electrooptic computing devices.

ACKNOWLEDGMENTS

This research was supported by the Army Research Office.

BIBLIOGRAPHY/REFERENCES

[1] Miller, D.A.B. "Novel Analog Self-Electrooptic-Effect Devices." *IEEE Journal of Quantum Electronics*, vol. 29, no. 2., pp. 678-697, February 1993.

[2] Miller, D.A.B., D.S. Chemlas, T.H. Wood, C.A. Burrus, A.C. Gassard, and W. Wiegman. "The Quantum Well Self-Electrooptic Effect Device: Optoelectronic Bistability and Oscillation, and Self-Linearized Modulation." *IEEE Journal of Quantum Electronics*, vol. 21, no. 9., pp. 1462-1476, September 1985.

[3] Miller, D.A.B., D.S. Chemlas, T.C. Damen, A.C.Gossard, W.Wiegmann, T.H. Wood, and C.A. Burrus. "Bandedge electro-absorption in quantum well structures." *Physics Review Letters*, vol. 53, pp. 2173-2177, 1984.

[4] Shoop, B.L. "Optical Oversampled Analog-to-Digital Conversion," Ph.D. Thesis, Stanford University, June 1992.

[5] Shoop, B.L., B. Pezeshki, J.W. Goodman, and J.S. Harris. "Noninterferometric Optical Subtraction Using Reflection-Electroabsorption Modulators." *Optics Letters*, vol. 17, no. 1, pp. 58-60, January 1992.

Radiometric Calibration of the Landsat Thematic Mapper (TM)

Ron Morfitt

Electrical Engineering Department
South Dakota State University
Brookings, SD 57007

Faculty Advisor:

Dr. Dennis Helder

An extensive project to study radiometric calibration of Landsat Thematic Mapper (TM) data is underway to determine variation in detector response as a function of time. This project is crucial for multi-scene mosaics and quantitative studies ranging from climate to vegetation growth changes. The study is done by extracting and analyzing calibration data from archives located at EROS Data Center, the national archive for Landsat imagery. Initial results for Landsat-4 single orbits indicate system response changes over time. Certain portions of extracted data offer better calibration information than others. By analyzing both day and night data, scene dependent abnormalities were found and their effects isolated from the calibration information. Results from this study will be employed to produce more accurate Landsat imagery.

INTRODUCTION

Current radiometric calibration techniques for the Landsat TM instrument assume that the internal calibrator lamps remain constant with respect to time [1-3]. Combining multiple scenes taken over a period of time, as in mosaics, require accurate radiometric correction so that scene edges cannot be seen. Accurately calibrated imagery is needed in quantitative studies (i.e., climate and vegetation changes over time) to determine small changes through the seasons [4]. Information to determine radiometric corrections are in a calibration file included with scene data archived at the United States Geological Survey's EROS Data Center located at Sioux Falls, South Dakota. This calibration file was initially intended to make calibration easier and imagery more radiometrically correct.

The Landsat TM is a satellite imaging instrument that acquires images in six spectral bands (blue, green, red, near infrared (IR), mid IR, and far IR wavelengths) with a spatial resolution of about 30 meters [1]. Each spectral band, except band 6 which is not examined in this initial research, contain 16 detectors. During image acquisition, a mirror is tilted back and forth east and west as the satellite travels its north-south orbit. Each of the 16 detectors per spectral band, acquire one scan line during the mirror's east to west travel (forward scan) and one scan line during the mirror's west to east travel (reverse scan). Thus 16 lines are obtained simultaneously during each scan. When the mirror reverses direction, a flag shaped shutter blocks the earth from the detectors view.

Two events occur during this shutter obscuration period. First the shutter stops virtually all light from reaching the detectors, this is when dark current restoration occurs. Dark current restoration is the process of setting the response

of the detectors, when there is no light reaching the detectors, to two or three digital numbers (DN). The second event occurs approximately in the middle of shutter obscuration, and consists of applying a known radiance level onto the detectors. Three calibration lamps are used to produce this known radiance level. The shutter obscuration period is stored in the calibration file, and includes information for all spectral bands and all detectors.

Development of a time series of calibration data will determine how the calibration lamps, detectors, and other components, change with respect to time. The objective of this research is to more completely characterize the Landsat TM instrument so that imagery can be optimally radiometrically corrected. This research will determine how the calibration information is best utilized to calibrate TM images, and will also create a better understanding of the TM instrument. Through better understanding and the development of a time series of calibration data, better radiometrically corrected images can be produced. The steps taken to accomplish this objective are first to extract the calibration data, then to reduce the calibration information into a time series, and finally to perform an analysis of the reduced data to characterize the instrument more completely.

PROCEDURE

To more completely characterize the TM instrument, three different time scales are being studied: A single contiguous orbit (both daytime and nighttime scenes), and intermediate length outgassing interval, and the entire lifetime of the instrument.

The first step in this research was to search the archives at the EROS Data Center to determine the availability of the desired scenes. For the early-in-life single orbit data set, 20 contiguous scenes for daytime (November 24, 1982, path 41, rows 18-37) and 14 contiguous scenes for nighttime (December 18, 1982, path 106, rows 214-226) were chosen.

The next step was to extract the calibration data from the scene data, this was accomplished through software[5] developed in South Dakota State University's Electrical Engineering department. There are three levels of archived data in the Thematic Mapper Calibration Archive System (TMACS) which this software creates. Level 0 is the raw data directly obtained from EROS Data Center's archives, both scene and calibration data is retained at this level. Level 1 contains all of the calibration information in an unreduced form. Level 2 is the reduced calibration data, which yields detector response as a function of time. This detector response time series has never before been examined.

The final step in this research is to analyze the data. The commercial matrix manipulation package, Matlab, was used as a visualization tool. Matlab also allows for statistical analysis to be easily performed.

RESULTS

Visualization of the data shows many interesting anomalies (i.e., scan correlated shift, odd vs. even detector response, bright target saturation, scan direction differences and apparent gain change with respect to time), however, only two of these will be discussed here. These two anomalies are scan direction differences and apparent gain change with time.

Figure 1 shows a forward scan (top) and a reverse scan (bottom) during shutter obscuration. These plots are of detector response in DN verses sample number. A small portion of image data can be seen in these plots at the beginning (left) and the end (right) of the shutter obscuration period. During shutter

obscuration in a forward scan, dark current restoration occurs before the lamp pulse. Conversely, in a reverse scan, the lamp pulse occurs before dark current restoration.

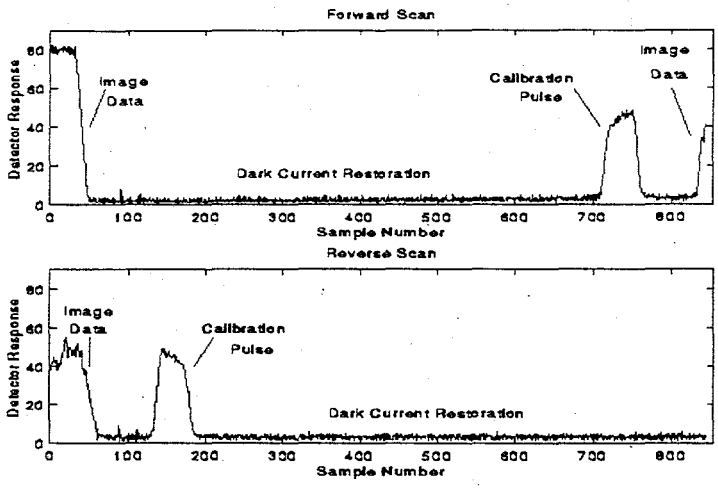


Figure 1: Forward and Reverse Scan Shutter Obscuration Period

Scan direction differences are shown in Figure 2. Notice there are large jumps in pulse values for the reverse scan (bottom) as compared to pulse values for the forward scan (top). The detector response variations for forward and reverse scans is the main difference in these plots. Forward scans are much more stable than reverse scans. Table 1 shows the band averaged standard deviations for forward, reverse and both scan directions. This corroborates the observation that forward scans are more stable than reverse scans. Figure 1 explains why this is true: reverse scan lamp pulses occur closer to the image data which corrupts the calibration pulse detector response on reverse scans. This is justified by examining data from night scenes, which show little difference between variances from forward scans to reverse scans. The significance of these results suggest that only forward scan calibration information should be used when radiometric calibration is performed. The detectors and calibration lamp pulses are quite stable over a short period of time so using only forward scan calibration data to calibrate two consecutive scan lines (forward and reverse) cause no significant errors in image quality.

Table 1: Band Averaged Forward/Reverse Standard Deviation Differences

	Scan Direction		
Band	Both	Forward	Reverse
1	1.046	0.716	1.178
2	0.821	0.526	0.630
3	0.690	0.571	0.735
4	1.004	0.690	0.772
5	0.485	0.297	0.242
7	1.015	1.127	0.392

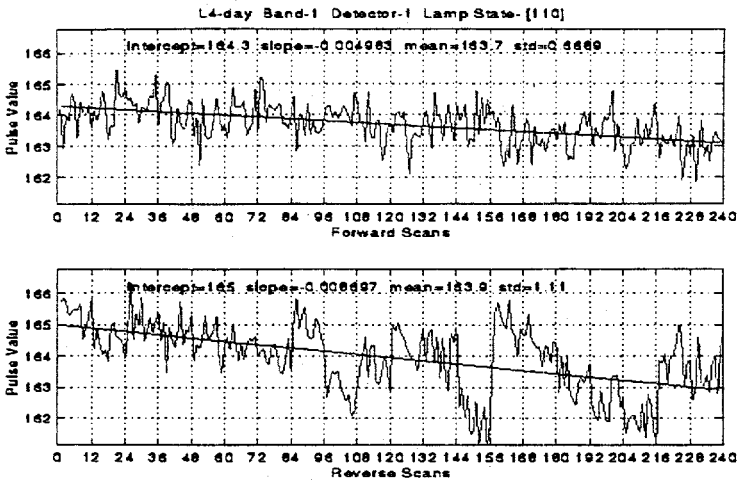


Figure 2: Forward and Reverse Scan Pulse Time Series

Apparent gain changes are also shown in Figure 2 as an observable decay with respect to time. A linear regression line fit to the calibration pulse value data which shows a negative slope. Using higher order polynomials in a least squares fit were shown to be statistically insignificant. This decrease in detector response with respect to time is consistent within each spectral band. Spectral bands one through four exhibit a significant slope while bands five and seven, show little slope (Figure 3). These differences can be attributed to the fact that spectral bands one through four are in the primary focal plane (not cooled) while spectral bands five through seven are in the cooled focal plane because they detect wavelengths in the IR which are dependent on temperature variations. Three possible reasons for this apparent gain change have been hypothesized: calibration lamps dim as time passes, detector response decrease with time, and temperature variations

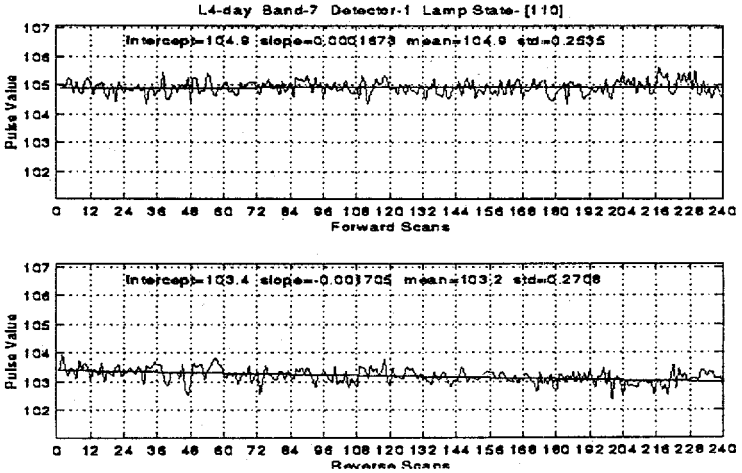


Figure 3: Cold Focal Plane Forward and Reverse Scan Pulse Time Series

affect detector response. Although the reason for this apparent decrease in detector response cannot be determined, Goddard Space Flight Center scientists are currently finding a temperature increase in the primary focal plane which is strongly correlated to the change in detector response [6]. Further analysis will be conducted.

CONCLUSION

A system has been developed to perform comprehensive analysis of Landsat Thematic Mapper radiometry. Initial results indicate a difference in forward and reverse calibration scans, with more variation in reverse scans than in forward scans. This difference is attributed to the fact that reverse scan calibration lamp pulse response is affected by scene data because the pulses occur soon after the shutter obscuration begins. The other major result is an apparent decrease in detector response with respect to time in the primary focal plane. This is associated with an increase in temperature of the primary focal plane. Quantitative analysis of these effects allow determination of an optimal method for radiometrically calibrating Landsat Thematic Mapper imagery. Results of this research have been implemented at the EROS Data Center to calibrate raw image data with a significant improvement in calibration [7].

REFERENCES

[1] J. L. Barker, et. al., "Radiometric calibration and Processing Procedure for Reflective Bands on Landsat-4 Protoflight Thematic Mapper", Landsat-4 Scientific Characterization Early Results: NASA Conf. Pub. 2355. Greenbelt, MD (available NTIS, Acc. #N85-20508 [Vol. II]) pp. 47-86, 1985a.

[2] J. L. Barker, et. al., "Characterization of Radiometric Calibration of Landsat-4 Reflective Bands", Landsat-4 Scientific Characterization Early Results: NASA Conf. Pub. 2355 (available NTIS, Acc. #N85-20508 [Vol. II]) pp. 374-474, 1985b.

[3] J. L. Barker, "Prelaunch Absolute Radiometric Calibration of the Reflective Bands of the Landsat-4 Protoflight Thematic Mapper. Landsat-4 Science Characterization Early Results: NASA Conf. Pub. 2355. Greenbelt, MD (available NTIS Acc. #N85-20508 [Vol. II]) pp. 277-371, 1985.

[4] D. Helder, Personal Communication, South Dakota State University, 1994-1995.

[5] Thematic Mapper Calibration Archive System, Image Processing Lab, Electrical Engineering Department, South Dakota State University, 1994-Present.

[6] J. L. Barker, Personal Communication, Goddard Space Flight Center, 1994-1995.

[7] W. Boncyk, Personal Communication, EROS Data Center, 1994-1995.

Selective Etching of InGaAs over InAlAs Using CH_4/Ar Reactive Ion Etching

Frank D. Bannon, III

Department of Electrical Engineering
Electronic Materials and Processing Research Laboratory
The Pennsylvania State University
University Park, PA 16802.

Faculty Advisor: Theresa S. Mayer

Reactive ion etching (RIE) of Indium Gallium Arsenide (InGaAs) and Indium Aluminum Arsenide (InAlAs) using a methane/argon (CH_4/Ar) mixture is described. The selectivity of the etchant is controlled by varying the CH_4/Ar ratio. Nearly ideal selectivity of InGaAs over InAlAs is demonstrated for CH_4/Ar ratios greater than 1:2. A reduction in the selectivity can be achieved by decreasing the concentration of CH_4. The etch rate is constant with etch time for all ratios of CH_4/Ar.

BACKGROUND

Reactive ion etching (RIE) of III-V compound semiconductor devices is becoming increasingly popular because of the control it provides in terms of material selectivity and pattern definition. In particular, devices fabricated using RIE typically have edge profiles that are nearly vertical for all crystallographic orientations [1]. Such anisotropic edge profiles are especially important when defining features with separations of one micron or less. Presently, RIE is being used to fabricate such high performance III-V devices as heterojunction bipolar transistors (HBT's) and high electron mobility transistors (HEMT's) [2,3]. As the demand for these devices continues to increase, additional research is required to more fully characterize existing etchants and develop more robust etchants.

Fabrication of InP-based HBT's and HEMT's is accomplished most easily if the etchants are selective – etching InGaAs more readily than InAlAs. Several groups have demonstrated that methane-based chemistries can be used to achieve a very high selectivity [2,3,4] and nearly vertical sidewall profiles. During the etch process, however, a polymer is often deposited on the sample and the chamber volume. In the most severe case, this polymer inhibits etching of the semiconductor sample [5]. In order to reduce the polymer deposition, hydrogen (H_2) is often mixed with CH_4. It is believed that H_2 serves to sputter the polymer, therefore reducing its rate of deposition [6]. Unfortunately, the CH_4/H_2 chemistry is only effective when process pressures are maintained below 30 mTorr during the etch [2]. These relatively low process pressures preclude the use of many commercially available RIE systems. In this work, argon (Ar) was investigated as an alternative to H_2 in order to sustain etching at more typical process pressures in the range of 150 mTorr.

EXPERIMENTAL PROCEDURE

The etch rate and the selectivity depend on three process parameters: pressure, power density, and CH_4/Ar ratio. In order to determine if the CH_4/Ar system is a viable alternative to CH_4/H_2, the pressure and power density were fixed and the CH_4/Ar ratio was varied from 1:3 to 3:2. Moreover, at a given CH_4/Ar ratio, the etch time was varied from 5 to 25 minutes. The pressure and power densities used in this work were 150 mTorr and 1.0 W/cm^2, respectively. This power density is comparable to those used in prior studies of CH_4/H_2 [6].

Two samples were needed to investigate the effect of CH_4/Ar ratio on the etch rate of InGaAs and InAlAs. The first consisted of a 1 μm thick layer of InAlAs doped 5×10^{17}/cm^3 while the second consisted of layer of InGaAs with the same thickness and doping. In both cases, these layers were grown epitaxially on an InP substrate in a Varian GEN II molecular beam epitaxy (MBE) system. Following the growth, the samples were masked with photoresist using a standard lithography process to define the features to be etched.

At each CH_4/Ar ratio, a small piece of the InAlAs and the InGaAs samples were etched simultaneously in a planar diode Plasma Technology RIE 80. Prior to loading the samples into the RIE, the chamber was prepared by etching in an O_2 plasma at 100 W and 1 Torr for 5 minutes (necessary to guarantee good run-to-run reproducability). Next, the samples were placed in the RIE, and a low power O_2 etch was performed to remove any photoresist residue that remained on the surface of the sample following the photolithography. Before the process gases were introduced, the chamber was evacuated using a Balzer 170 turbo pump. Once a base pressure of 10^{-5} Torr was attained, the CH_4 and Ar mixture was added to the RIE chamber through a mass flow controller. A constant pressure of 150 mTorr was maintained with a throttle valve placed between the chamber and the turbo pump. The etch began when a bias corresponding to a power density of 1.0 W/cm^2 (electrode area of this system is 240 cm^2) was applied to the sample. Following the etch, the etch depth was measured using stylus profilometry.

RESULTS

The dependence of the InGaAs and InAlAs etch rates as a function of CH_4/Ar ratio are shown in Fig. 1. As is evident from these data, the etch rate of the InGaAs increases from 270 to 390 Å/min as the CH_4 content is increased from 25 to 60%. In contrast, the InAlAs etch rate decreases with increasing CH_4 content. Etching of InAlAs ceases entirely at methane concentrations greater than 50% due to the growth of thin layer of material that cannot be removed with CH_4/Ar. This material can be removed easily by exposing the sample to a low power O_2 plasma for approximately 5 minutes. These results demonstrate that the selectivity of the etch can be adjusted by varying the CH_4/Ar ratio. For low concentrations of CH_4, the selectivity is approximately 2.5 while nearly ideal selectivity can be achieved for CH_4 concentrations above 50%. Moreover, the etch rates measured for InGaAs are comparable to those obtained using CH_4/H_2 at much lower pressures [2].

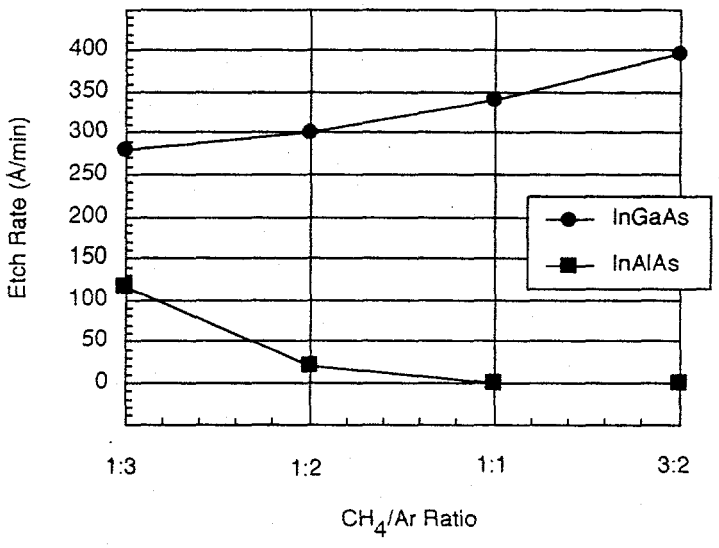


Figure 1. Etch rate as a function of CH_4/Ar ratio for InGaAs and InAlAs. The samples were etched for 5 minutes at a pressure of 150 mTorr and a power density of 1.0 W/cm^2.

The etch rate is shown as a function of time in Fig. 2 for several CH_4/Ar ratios. Because InAlAs cannot be removed as the CH_4 concentration is increased, only the 1:3 ratio is given in this figure. With the exception of the 3:2 ratio, all etch rates are nearly constant with time up to a maximum etch time of 25 minutes. Moreover, inspection of the edge profiles on a sample etched to a depth of 1 μm using a CH_4/Ar ratio of 3:2 revealed nearly vertical sidewalls.

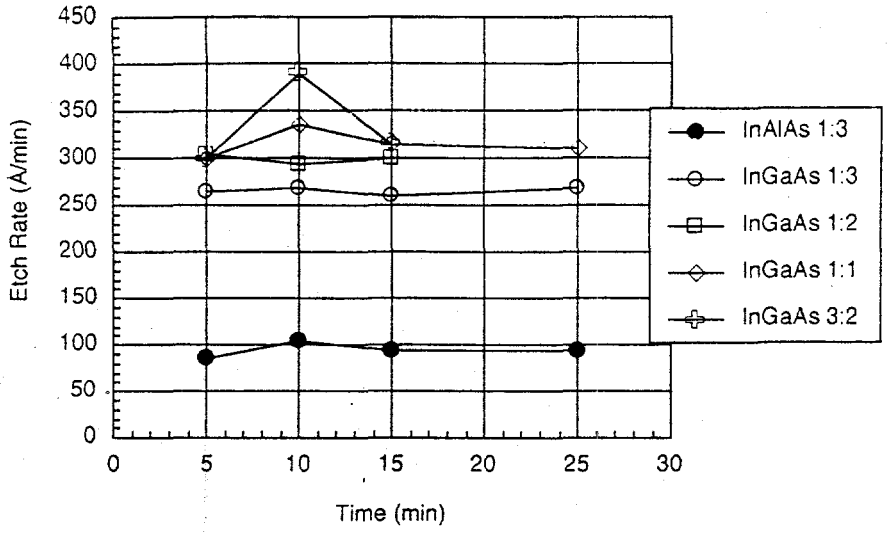


Figure 2. Etch rate as a function of time for CH_4/Ar ratios varying from 1:3 to 3:2. The samples were etched at a pressure of 150 mTorr and a power density of 1.0 W/cm^2.

CONCLUSIONS

We have studied the selective dry etching of InGaAs and InAlAs using CH_4/Ar mixtures. The results demonstrate that etch rates comparable to those obtained using CH_4/H_2 can be achieved at pressures in the range of 150 mTorr. Moreover, the selectivity of the etchant can be controlled by varying the CH_4/Ar ratio. Nearly ideal selectivity of InGaAs over InAlAs can be attained for CH_4/Ar ratios greater than 1:2.

ACKNOWLEDGMENTS

The authors would like to acknowledge the technical support of David Sarge and Brian Luther who assisted in equipment modification. The MBE materials were provided by Wen-Yen Hwang. Helpful discussions with Robert Davis are greatly appreciated. Frank Bannon was supported by a National Science Foundation Undergraduate Research Experience Grant No. ECS-920246.

REFERENCES

[1] Williams, R., *Modern GaAs Processing Methods*, 2nd edition, Artech House, 1990, pp.173-195.

[2] Arnot, H., R. W. Glew, G. Schiavini, L. J. Rigby, A. Piccirillo, *Applied Physics Letters*, Vol. 62, June 1993, pp. 3189- 3191.

[3] Pearton, S. J., W. S. Hobson, *Applied Physics Letters*, Vol. 56, May 1990, pp. 2186-2188.

[4] Pearton, S. J., W. S. Hobson, F. A Baiocchi, A. B. Emerson, and K. S. Jones, *Journal of Vacuum Science and Technology*, Vol. B8, No. 1, Jan/Feb 1990, pp. 57–67.

[5] Katzk, A, *Indium Phosphide and Related Materials, Processing, Technology, and Devices*, Artech House, 1992, pp. 277-306.

[6] Law, V. J., G. Ingrai, M. Tewardt, *Semiconductor Science and Technology*, Vol. 6, May 1991, pp. 411.

Messy Data Assist

Renee M. Barger

Departments of Management and Computer Science
University of North Carolina at Asheville
Asheville, NC 28804-3299

Robert Yearout, George Yates and Charles Massey

Data collected by practicing industrial engineers and ergonomicists frequently do not conform to the assumptions of normality, i.e., that error terms are independently and identically distributed. Error terms within groups also must have a mean equal to zero and equal variances (δ^2). It is very common for field data not to meet these conditions for normality. When this heterogeneous condition occurs, the analyst must depart from the more traditional (parametric) analysis and use non-parametric statistical methods. An analysis recommended by Miliken and Johnson (1984) is Satterwaite's Approximation. This method allows pairwise comparisons of data groups with unequal δ^2. It also is not sensitive to unequal group samples. When making multiple comparisons another common analytical error is failure to consider reduced reliability of the stated confidence level. Satterwaite's Approximation with an associated confidence level analysis is not available in any current software package. This research developed a software application that provides engineers, ergonomicists and managers a tool that determines statistically significant differences between heterogeneous groups and provides the analysts an accurate family confidence level.

INTRODUCTION

For many experiments, the investigator may frequently want to compare the effects of different treatments or, equivalently, the means of the treatment populations. For example, the investigator wants to test the following kinds of hypotheses (Milliken and Johnson, 1984):

$$H_{01}: \sum c_i \mu_i = a \qquad (1)$$

for some given set of coefficients $c_1, c_2, \ldots, c_t$ and for some given constant a:

$$H_{02}: \mu_1 = \mu_2 = \cdots \mu_t \qquad (2)$$

and

$$H_{03}: \mu_i = \mu_{i'} \qquad (3)$$

for some i = i'

Unfortunately, much of the data collected by practicing industrial engineers and ergonomicists frequently do not conform to the assumption of normality, i.e., that error terms are independently and identically distributed. Error terms within groups also must have a mean equal to zero and equal variances (δ^2) for homogeneity. Often, both of these assumptions are violated. Thus the validity of a parametric test, which depends for its validity on the crucial assumption that the investigator is sampling randomly from a distribution belonging to a particular family, may be inappropriate (Sprent, 1989). Additional complications will arise when observations within sample groups are small and unequal. The t-test is considered sufficiently robust to handle non-normal distributions, although its reliability for conditions where

unequal variances and unequal sample size is questionable. It is generally accepted that the issue of unequal variance is a more critical concern than the distribution of the data. When these conditions are present, the investigator must depart from the more traditional (parametric) techniques and use non-parametric (distribution free) techniques. The main advantage of non-parametric methods is that more precise inferences can be made when this "heterogeneous" condition exists. However, the principle disadvantage of using non-parametric techniques is the loss of efficiency and the waste of much useful information (Miller and Freund, 1977).

It may be beneficial to assume that observations come from a particular distribution family and most parametric tests are quite robust. However, the investigator should always test the data to insure that the assumed parameters are indeed true. Tests for homogeneity of variances such as Harley's F-Max, Bartlett's, Box's and Levene's are most appropriate. The Levene's Test is superior when analyzing non-normal and skewed data. It also is an excellent performer for small and unequal data sets (Milliken and Johnson, 1984).

When one of these tests for homogeneity reveals that the data is heterogeneous and some non-parametric method is required, Milliken and Johnson (1984) recommend Satterwaite's Approximation. This technique is appropriate for conditions where there are independent samples with unequal variances (heterogeneous data) and unequal sample sizes. Though the distribution of error terms may or may not be strictly normal, in most ergonomic or industrial application cases the conditions can be assumed to approximate normal. This method will allow for a good approximation by estimating the degrees of freedom v for the case of unequal variances. The procedure is illustrated by the following equation 4:

$$v = \frac{(\sum c_i^2 \hat{\sigma}_i^2 / n_i)^2}{(\sum_i [\frac{c_i^4 \hat{\sigma}_i^4}{n_i^2 (n_i - 1)}])} \tag{4}$$

Summarizing, one rejects the hypothesis:

$$H_0: \sum c_i \mu_i = a \tag{5}$$

if

$$|t_c| = \frac{|\sum c_i \mu_i - a|}{\sqrt{\frac{\sum c_i^2 \hat{\sigma}_i^2}{n_i}}} > t_{\frac{\alpha}{2}, v} \tag{6}$$

The t-test for the stated hypothesis can then be made using the adjusted degrees of freedom and critical t. By using this method the analyst not only retained much of the original information but the test is insensitive to unequal sample sizes. Yearout (1987) demonstrated this method of determining differences between heterogeneous groups using a simple Turbo Pascal (1984) program. Although this program was indeed quite simple, it required the investigator to have access to a Pascal compiler and to be Pascal literate. Thus the program had limited utility and was not user friendly.

Another common analytical error made when making multiple pair-wise comparisons is failure to consider reduced reliability of the stated confidence level. As a result, the investigator may be in error when concluding that there are significant differences between groups. Bonferroni proposed a procedure that will determine the appropriate minimum significance level to obtain a desired confidence level (Neter, Wasserman and Kutner, 1990). The procedure to determine the confidence of any particular comparison (Snedecor and Cochran, 1982) is determined by equation 7.

$$CONFIDENCE\ LEVEL = 100(1-k\alpha) \quad (7)$$

Where the number of k intervals are calculated by equation 8:

$$k = \binom{I}{2} = \frac{I(I-1)}{2} \quad (8)$$

The purpose of this research was to develop a software application that would provide the industrial engineer or ergonomicist with a tool to determine statistically significant differences between heterogeneous groups and to provide an accurate family confidence level.

METHOD

The objective was to design the program simple enough for a technician to use and yet provide sufficient technical data for the Statistician. This stand alone application was developed for IBM-PCs using DOS operation systems. The first step was to re-engineer Satterwaite's Approximation module (Yearout 1987). The creation of a module which performed Bonferroni's Method to determine appropriate confidence intervals for necessary adjustment to the Confidence Level was the next significant feature. The option for the user to select the significance level for the table look-up and comparison for significant difference was important since the results being reported would be in a form of "originally selected" and an "adjusted confidence level". This interaction would allow the user to obtain information that was useful for their specific problem. For the convenience of the user the student-t table would be resident with the program so the program would search the table, compare the data and report the results. Finally, the option is given to the user to print a report.

The program routes the data through the process wherever it is needed and performs the necessary calculations. The Program's Data Analysis Flow is shown in Figure 1.

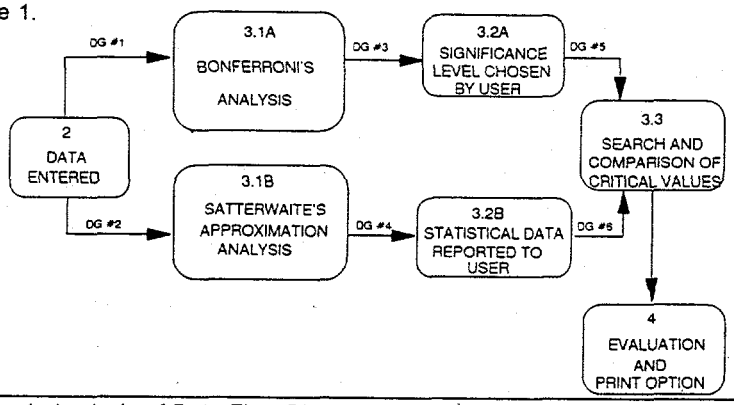


Figure 1: Analysis of Data Flow Diagram

Legend

DG#1
Number of Comparisons
Desired Confidence Level

DG#2
Number in Sample Groups 1/2
Mean of Groups 1/2
Standard Deviations of Groups 1/2

DG#3
Calculated Family Confidence Level
Calculated Bonferroni's Interval

Data Groups

DG#4
Calculated Degrees of Freedom
Calculated Linear Combination
Calculated Combination Variance
Calculated Combination Standard Error
Correction Factor
Calculated T-Value

DG#5
User's Choice of Significance Level

DG#6
Calculated Degrees of Freedom
Calculated T-Value

The ease of use with this program is that a minimal amount of data is required to be entered by the user as demonstrated by Figure 2.

DATA ENTRY

Count the number of groups you will be comparing for this study.
Enter the TOTAl number of groups: **4**

Give the first group of this comparison a name of 8 or less characters: **MALES A**
The second group name with 8 or less characters: **FEMALESA**

Enter the number of samples in MALES A: **8**
Enter the number of samples in FEMALESA : **8**

Enter the mean of sample MALES A : **73.74**
Enter the mean of sample FEMALESA : **87.50**

Standard deviation for MALES A: **4.43**
Standard deviation for FEMALESA : **5.61**

Figure 2: Data Entry Screen

TESTING UTILITY

Data from several different published sources was used to evaluate the performance of the program. Several different minor implementation errors were detected and corrected. An illustration of the execution of the current program is contained in the following example analysis.

DESCRIPTION OF PROGRAM EXECUTION

The program design provides information screens and clear instructions of what is expected from the user. After entering the data such as displayed in Figure 2 the user will then be prompted to state the desired confidence level such as 95%. At that point in the program Satterwaite's Approximation and Bonferroni's intervals are executed. The calculated t value and appropriate family confidence level are then displayed.

The user is then prompted to select the now adjusted Bonferroni Family confidence level or to select another alpha (α)that would insure the originally selected 95% confidence level. The final result with a comparison to the critical t value is displayed on the final screen. The final screen also provides the user with the option to print a report which is a record of all pertinent analysis data for that run, including data entered for analysis, critical statistical values calculated in the analysis, the choice the user made in selecting a final alpha and the final evaluation report.

CONCLUSIONS

The program does provide the investigator with a valuable tool for both heterogeneous and homogeneous data. Its main strengths include: the application is self-contained for both calculations and critical t value tables; and structured to allow the investigator the option to either choose a lesser family confidence level or adjust the α and raise the family confidence interval. In addition a detailed analysis is provided in screen display or complete analysis printout. The application's principle weakness is that the investigator is required to reenter data for each paired comparison. The next stage of their research, the authors intend to restructure the program to allow all data to be entered at the start of the program.

REFERENCES

Miller, I. and Freund, J., 1977, Probability and Statistics for Engineers, Second Edition, Englewood Cliffs:Prentice-Hall,pp. 270-272.

Milliken, G. and Johnson, D., 1984, Analysis of Messy Data, Volume I:Designed Experiments, Belmont:Lifetime Learning Publications, pp. 18-25.

Neter, J., Wasserman, W. and Kutner, M., 1990, Applied Linear Statistical Models, 3rd Edition, Boston:Irwin, pp. 160-161 and 734-735.

Sprent, P., 1989, Applied Nonparametric Statistical Methods, New York:Chapman and Hall, pp. 1-4.

Snedecor and Cochran, 1982, Statistical Methods, Seventh Edition, Ames:The Iowa State University Press, pp. 116 and 166.

Sprague, K., Yearout, R. and Davis, S., 1994, "The Effects of Continuous Working Days on Preferred leisure Noise Levels, Proceedings of the 2nd Annual Mid-Atlantic Human Factors Conference, Washington, pp. 137-138.

Turbo Pascal, Version 3.0, Second Edition, 1985, Scotts Valley:Borland International Inc.

Yearout, R., 1987, Task Lighting for Visual Display Unit Workstations, Kansas State University, Manhattan, Annex 1 Appendix G.

Pilot Fatigue and Instrumental Disputes

Dara Strickland and Maranda McBride

Department of Industrial Engineering
North Carolina Agricultural and Technical State University
419 McNair Hall
Greensboro, North Carolina 27411

Faculty Advisor:
Barbara Pioro

The authors conducted a series of interviews with pilots employed by major airlines and found that many would agree that when it comes to aircraft design, often designers get too caught up in equipping the aircraft with the most recent technology and forget about the most important aspect of safe transport which is the pilot. Highly automated transports have simply focused on the design of the automated cockpit and the question has been raised about the pilot-computer interface when it comes to air-safety (Hughes and Dornheim, 1995). While focusing on the automation aspects of the cockpit, more attention needs to be focused on the human aspect of the cockpit. Two things that need to be considered are the physical and mental state of the pilot as well as their understanding of the new instrument panels.

BACKGROUND

General fatigue is a diffused sensation, which is accompanied by feelings of indolence and disinclination for any kind of activity. One major symptom of fatigue is a general sensation of weariness. Some of the different kinds of fatigue that can occur are 1) visual fatigue, arising from over-tiring the eyes, 2) general body fatigue: physical overloading of the entire organism, 3) mental fatigue, induced by mental or intellectual work, 4) nervous fatigue, caused by over-stressing one part of the psychomotor system, as in skilled work, 5) monotony of either occupation or surroundings, 6) chronic fatigue, an accumulation of long-term effects, and 7) circadian or nyctemeral fatigue, part of the day-night rhythm, and initiating a period of sleep (Grandjean, 1988). Of these seven kinds of fatigue the ones most noted in pilots are visual, monotony, chronic, and circadian fatigue. The proper use of the instruments in the cockpit can reduce fatigue. However, the pilots are not taught in flight school techniques to make their flights more comfortable and how to reduce fatigue during flights.

Most of the pilots interviewed did not see any problem dealing with the placement of the instruments within the cockpit. When asked what they would change to make their flights more comfortable many replied the schedules under which they fly. It is important to address the concerns of the pilots and not merely what new technology can be incorporated into the cockpits. In many cases, eliminating or reducing fatigue proves to be a major factor in the improvement of safety and maintenance during flight. There have been many documented cases related to pilot fatigue that have resulted in physical injuries and still others that have resulted in fatalities.

FATIGUE

Visual Fatigue and Fatigue Due to Monotony

One of the most common causes of visual fatigue seen in pilots is the glare of the sun. Pilots find that a quality pair of sunglasses are necessary during daytime flights to eliminate the eye strain caused by the sun.

The constant monotony and boredom that the pilots encounter between flights causes them to become more fatigued prior to the take-off of their next scheduled flight. When questioned, many pilots respond that it is not as simple as time in the air. What also must be considered is how long the pilot has been on duty. Many complain of the idle hours they spend waiting for their next flights. In most cases they have no where to go and nothing to do. During this time, they tend to become weary yet there is no time to get any restful sleep. Many pilots are not provided meals during their time on duty and are unable to get decent meals which also affects their physical and mental states. When forced to fly under

these very uncomfortable conditions it is difficult to remain alert to the ever changing weather and aircraft conditions.

Many of the pilots interviewed stated that trip scheduling is a serious problem in the aircraft industry. It is believed that it is better for pilots to get as much time in between flights as possible in order to allow them to rest. Most pilots tend to disagree. When interviewed, several of the pilots stated that the three hour breaks are stressful and inconvenient. They prefer to have shorter break time; and therefore, fewer hours on duty.

Chronic Fatigue

Another common area of concern is the required rest time for pilots. It is required for pilots to be allowed at least eight consecutive hours per day for rest. However, this does not allow for the time to get to a hotel, prepare for bed, actually fall asleep, and then get up the next day and prepare for another duty day. In many circumstances, this time is to be used to get a good meal before another rigorous day of scheduled flights because it is not certain if there will be another opportunity to eat.

In one case study provided by the National Transportation Safety Board in Washington, DC documented an accident that occurred in January, 1983 where a pilot had requested a day off from his seven day schedule because he was tired. His request was refused and the aircraft ran off the departure end of the runway during landing (NTSB, 1983). The National Transportation Safety Board determined that one of the probable causes of the accident which resulted in no injuries or fatalities was that the distance and airspeed were misjudged by the pilot due to factors of fatigue(NTSB, 1983).

It seems to be fairly common for pilots to fly even though they know that they are in poor physical and mental flying condition. In an incident on May, 1984 where an aircraft crashed after takeoff the pilot stated that she mismanaged the fuel system and that she "probably took off on fumes." She went on to say that she should have told the company for which she flew that she was too tired to make the flight (NTSB, 1984). Another incident that occurred in May, 1988 witnesses stated that the pilot only had two hours of sleep prior to the departure of the aircraft. Company personnel noted that the pilot looked "extremely tired" and was "really dragging" prior to the last leg of the return flight. One crew member was killed on this flight (NTSB, 1988).

Out of twenty-three reports by the National Transportation Safety Board on Pilot Fatigue as Cause/Factor in aircraft accidents, fourteen of those accidents were due to fatigue factors known by the airline. Furthermore, out of these fourteen cases, twelve resulted in either fatal or serious injuries among the crew and passengers aboard the flights. The scheduling of the pilot really needs to be considered when making improvements for better air safety.

Circadian Fatigue

Circadian fatigue often occurs when a pilot is forced to fly both daytime and nighttime flights. The body's natural time clock is altered; and thus, rest is difficult to accomplish. This is especially common when the pilot has to fly at night and must sleep during the day when he/she is not accustomed to doing so.

One documented case that resulted in two crew members and one passenger deaths the pilot had been working a varied schedule of some nighttime and some daytime hours. The fact that he had had an adequate period for crew rest prior to reporting to work does not subtract from the fact that fatigue was one of the many factors resulting in this disaster (NTSB, 1985).

Survey Results

The interviews and surveys were used to determine the percentages of fatigue. The pilots experienced visual, monotony, chronic, and circadian fatigue.

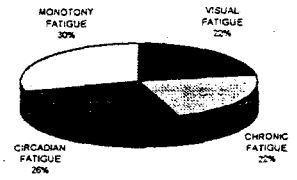


Figure 1: Breakdown of the Major Types of Fatigue Experienced During Flight

INSTRUMENT KNOWLEDGE

The other major area of concern deals with the pilots understanding of the new technology incorporated in the new cockpit designs. One instance is that the pilots are not aware of the ergonomic design of the seats in the cockpit. There have also been several instances where pilots either do not understand what the automation is doing or do not get adequate feedback from the newly automated systems.

The pilots are not instructed that the ergonomic design of the cockpit's seats are for their benefit. Many of the pilots interviewed discussed the discomfort of the current seats. However, these complaints about the seats are because the seat design is not being utilized to its fullest potential by the pilots. Many of the pilots suggest making the seats comfortable by inserting pillows under their buttocks and behind the lumbar portion of their spine; however, this will cause extreme discomfort when flying to long destinations. The current cushion of the cockpit seats satisfies the American National Standard Institute, which is to have a seat cushion thickness of 1.5 to 2 inches thick (Sanders, 1993). This thickness is important because if the seat is "cushy" the pilot's posture is restricted, blood circulation is reduced, skin temperature increases, and pain may result (Sanders, 1993). The seat is designed with a slight concavity with a radius of 45 cm to support the pilot's lumbar (Sanders, 1993).

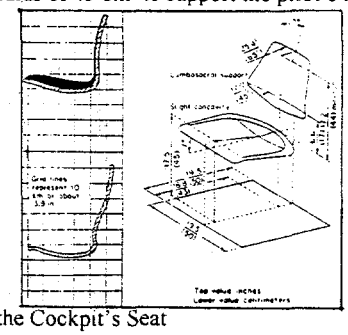


Figure 2: Illustration of the Cockpit's Seat

It has been demonstrated through surveys and research that many of the experienced pilots of the automated aircraft are occasionally surprised by what the systems are doing especially in non-routine flights when time is of the essence. David D. Woods and Nadine B. Sarter, human factors researchers at Ohio State University, believe that pilots do not always have a good "mental mode" of what the computerized flight systems are doing during the entire flight. This "mode confusion" occurs often in vertical navigation. One example of this is when an auto-flight system shifts from a vertical climb mode to altitude capture and then to altitude hold during leveling off all very rapidly (Hughes and Dornheim, 1995). The modes switch so quickly that often the pilot monitoring the systems becomes confused and does not know how to react. Earl L. Weiner, a human factors researcher at the University of Miami who has studied automated cockpits extensively for NASA Ames, says that some glass cockpits have clumsy automation that creates bottlenecks during high-workload periods when pilots are least able to deal with them (Hughes and Dornheim, 1995).

It is because of this that accident investigators would not agree that glass-cockpit designs always improve upon the older systems. One incident that illustrates this is the Kegworth accident. In this accident, the crew reacted inappropriately to engine vibration, noise and shuddering without identifying indications of engine parameters on the CRT display. As a result, they throttled back engine No. 2 and the vibration ceased even though No. 1 had been surging and flames had shot out the back of it. Engine No. 1 then suffered a loss of thrust on final approach leading to a crash in which forty-seven people died. The Air Accident Investigation Branch reported that the new glass-cockpit engine displays should have been tested on with line pilots to see if the displays were effective in communicating information to the crew. Kenneth Smart, chief inspector of air accidents in Great Britain believes the way to make sure new designs are really an improvement is more a factor of scientific approach to testing the new cockpit layouts than anything else (Hughes and Dornheim, 1995).

The occasional collapse in the communication between pilots and automated systems is occurring as computers are becoming more complex and as they are given more authority in the cockpit. However, human factors researchers do not think that glass-cockpit aircraft are

too automated, but they believe that how the automation is implemented is important (Hughes and Dornheim, 1995).

Sarter noted that in glass-cockpit aircraft the auto-pilot systems are becoming more complex and have more control over the aircraft, but the feedback to the pilot is not changing. In some of the cases, the necessary data is available but the pilot does not know how to access it. The question that arises is what good is a data-rich environment if it is more difficult to choose what is important among that data (Hughes and Dornheim, 1995).

CONCLUSION

Perhaps it is necessary to step away from implementing the wonderful technology that is presently available or will soon be available and deal with the human aspect of cockpit design. By looking at flight from a pilot's point of view, improvements may be made to make their job a little easier and less risky. By implementing environments meant to reduce fatigue and educating the pilots about the new systems that have been and will be implemented, perhaps flight can be made more comfortable and safe. It is important to educate the pilots on the proper way to utilize the instrumentation and the equipment to enhance comfort and aircraft control. It is also necessary for pilots to be more in touch with their physical and mental states, and avoid making flights whenever possible if they are too fatigued to do so safely. It must not be forgotten that they are responsible for the lives and well-being of not only themselves but the other passenger aboard the flight.

ACKNOWLEDGMENTS

The authors would like to extend special thanks and appreciation to Mr. Glenn Shellhouse and Ms. Marcie Murdoch as well as the other pilots of Continental Airlines Inc. and USAIR who assisted us in our interviews, information gathering, and surveys.

REFERENCES

Grandjean, Etienne. "11. Fatigue." Fitting the Task to the Man. 1st ed. New York: Taylor and Francis, 1988.

Hughes, David. "Automated Cockpits Special Report, Part II: Studies Highlight Automation 'Surprises'", Aviation Weekly and Space Technology, Vol. 146, No. 6, 6 Feb. 1995: pg. 48-49.

Hughes, David and Michael A. Dornheim. "Automated Cockpits Special Report, Part I: Accidents Direct Focus on cockpit Automation", Aviation Weekly and Space Technology, Vol. 142, No. 5, 30 Jan. 1995: pg. 52-54.

National Transportation Safety Board. Accidents Involving Part 121/135 Operators-Pilot Fatigue Cause/Factor. Washington, D.C., 1983-1993.

Sanders, Mark S. and Ernest J. McCormick. Human Factors in Engineering and Design. 7th ed. New York: McGraw-Hill, 1993.

An Analysis of Silicon Based Solar Cells: The Effects of High Temperature on Power Output

Isaac Garaway

Department of Mechanical Engineering
College of Engineering and Technology
Northern Arizona University
Flagstaff, AZ. 86011

Faculty Advisor:

Donald R. Garrett

By mathematical analysis and experimentation, the use of three types of silicon-based photovoltaic cells in high temperature applications was evaluated in order to determine which is best suited to a high temperature operating environment. An initial mathematical analysis of the temperature effect was followed by experimentally controlled observations of amorphous, polycrystalline, and single crystal silicon solar cells as they underwent different temperature loads. Observations produced graphic representations of each type of cell's efficiency as a function of temperature. Both approaches, mathematical and experimental, revealed a decrease in power as a function of a rise in temperature. While the mathematical solution provided only a general response for all types of silicon solar cells, the experimental process provided specific responses for each cell type. Amorphous, the cheapest of the three, was found to be the least affected by temperature (-.3%/power/Celsius). Single crystal and polycrystalline both responded similarly (-.5%/power/Celsius). The mutually supporting findings from the experimentation and the mathematical analysis form a base to support the study's conclusions: silicon solar cells respond very negatively to higher temperatures and thus are poorly suited for use in hot climates. Of the three types, amorphous would be the best choice for hot climates. Although less efficient, its resistance to power loss combined with lower cost make it the most economical.

BACKGROUND

Silicon solar cells are by far the most widely used solar cell today. Approximately 70% of all solar cells being used are one of these three types-

- Amorphous- This solar cell has had the most popularity in terms of providing cheap photovoltaic energy. While only 8% efficient, its lower production costs make it the cheapest of the three.
- Polycrystalline- This type of solar cell is produced from multi-crystal silicon wafers. Presently, efficiencies of about 12% are available.
- Single crystal- This solar cell is manufactured from a single crystal silicon wafer. The lack of crystal grain structure that inhibits electron flow in the multi-crystal cell allow this cell to achieve considerably higher efficiencies then that of the polycrystalline. The present industry standard is about 16% efficient.

Solar cells in general are very temperature dependent. As would be expected this dependence is most noticeable in high temperature climates. Although the phenomenon of temperature dependence has been recognized since the 1970's, there has not been much literature published on the topic. Furthermore, all the literature that is available looks at this dependence with a very broad focus. None has looked at the specific responses of the different silicon solar cells types. Due to the ever increasing number of silicon-based solar cell products, it has become increasingly important to examine the cell types for temperature dependence characteristics. Changes at high temperatures would have far ranging implications in terms of choice of cell type in a particular setting or appplication.

MATHEMATICAL ANALYSIS

Semiconductors and semiconductor devices such as solar cells are extremely temperature dependent. This dependency is due to the exponential increase of the saturation current density with higher temperature. The principal cause of this variation with temperature is the variation of the intrinsic charge carrier concentration with temperature. The increase in saturation current density across the PN junction causes an exponential increase in the forward biased current,

resulting in an overall decrease in power output. In solar cells where the objective is power output, this effect is quite detrimental.

In more simplified terms - the saturation current density, or what I have called the "recombination current", which is the current created by mobile electrons recombining with vacancies before they reach the outside load, increases as a function of the rise in temperature. As would be expected, the increase in current density is quite detrimental; electrons that would have created power at the outside load never do because they recombine before they reach it.

We can go on and model the saturation current density with the following function:

$$J_s = \frac{q \cdot D_p}{L_p} \frac{n_i^2}{N_D} + \frac{q \cdot D_n \cdot n_i^2}{L_n \cdot N_A} \qquad (1)$$

Where:

- q - electronic charge
- D_p - minority carrier diffusion constant in the p-region
- D_n - minority carrier diffusion constant in the n-region
- L_p - minority carrier diffusion lengths for the p-region
- L_n - minority carrier diffusion lengths for the n-region
- N_A - impurity concentrations for the p-region
- N_D - impurity concentrations for the n-region
- n_i - intrinsic carrier concentration

All of the above variables, except for n_i, are physical constants or material characteristics, thus they are constant for each type of solar cell. The intrinsic carrier concentration, n_i, the focus of this study, is the variable which holds the temperature dependence of the saturartion current, and can be defined by Equation 2.

$$n_i^2 \propto T^3 \exp\left(-\frac{E_g(T)}{Tk}\right) \qquad (2)$$

Where:

$E_g(T)$ - energy gap at temperature T in degrees Kelvin

Figure 1 provides a graphic representation of the effects of temperature on various types of photovoltaic materials.

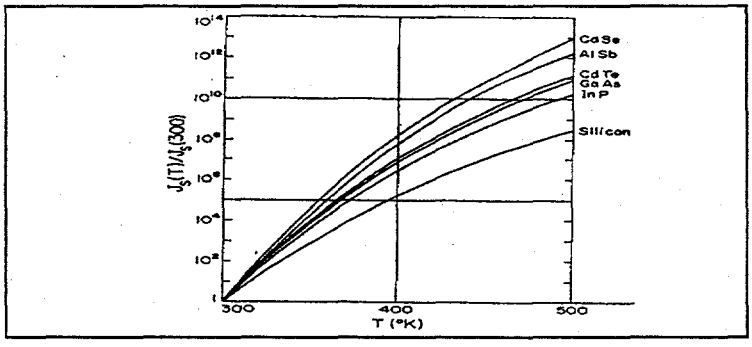


Figure 1- Saturation current density as a function of temperature for various types of photovoltaic materials.

As can be seen the saturation current does exponentially increase with temperature. However, in the temperature range of application concern, 300K - 400K, this curve can be approximated quite accurately with a linear slope. This then leads us to expect a linear decrease in power as a function of the rise of temperature in the range of 300K-400K. This mathematical solution however, provides only a general response for all silicon-based solar cells. In order to

compare each type of silicon solar cell on the basis of its ability to operate under high temperatures, it is necessary to determine the specific response of each type of cell. As mentioned previously, a detailed analysies of this effect on silicon solar cell types is not available. For this reason, experimental methods were incorporated into this study.

EXPERIMENTAL PROCEDURES

The experimentation for this project was simple in terms of the hardware and procedures used. It essentially involved a three-step process which incorporated a 1 watt solar cell of each silicon cell type.
The experiment can be outlined as follows:

- Load each individual cell type with a temperature gradient produced by a 1000 watt sun lamp irradiating the cell surface.
- Measure surface temperature and maximum power output iteratively with a computerized data collection system; METRABYTE hardware was used in conjunction with LABTECH software.
- Compile, analyze, and graphically model the individual curves using computer software; EXCEL and MATHCAD were used.

EXPERIMENTAL RESULTS

Figure 2 represents the compiled data for the specific power response of each silicon solar cell type to an increasing temperature.

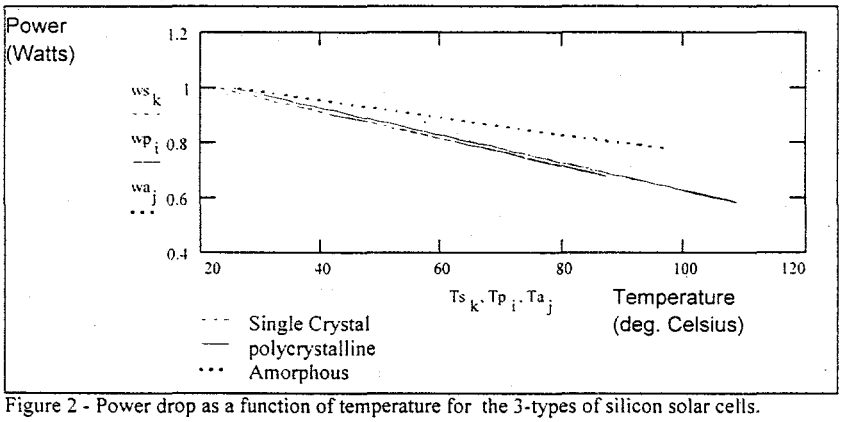


Figure 2 - Power drop as a function of temperature for the 3-types of silicon solar cells.

The slopes for each of these curves were found to be:

Silicon Cell Type	Slope
Amorphous	-.3 % power/Celsius
Polycrystalline	-.5% power/Celsius
Single Crystal	-.5% power/Celsius

OBSERVATIONS

From observing the changes seen in Figure 2, the following remarks can be made:

- Amorphous, the cheapest of the three solar cells, is the least affected by temerature.
- Single crystal and Polycrystalline both have the same temperature response.
- In all cases, the temperature response is significant and quite detrimental.

DISCUSSION

From the results of this research, it is clear that silicon solar cells respond negatively to higher temperatures. While a relationship between temperature response and a recombination current effect has been determined, it remains to be established whether temperature response can be solely attributed to the recombination current effect. While there are other possible causes of a temperature response, currently there are no viable means of analyzing them or for that matter of even verifying their connection to the temperature phenomona. In short, there are no absolutes when discussing some of the phenomena that occur within solar cells.

A number of interesting observations were made throughout this research. In theory, one would expect amorphous to respond differently than polycrystalline and single crystal, simply because amorphous has a slightly different photovoltaic effect. This difference in photovoltaic effect is largely a function of a construction process that differs from that used for both polycrystalline and single crystal. At the beginning of this study however, it was unknown whether the differences in amorphous would result in a difference in temperature response. As seen in the experimental results, it did. Surprisingly however, not only did it have a different response than the polycrystalline and single crystal (both of which carry a public image of greater viability), it was found to be less prone to respond negatively to an increase in temperature.

During the experimentation phase, an additional observation made was that each solar cell had to be loaded down with its maximum power resistor. If another load was applied, I found that the solar cells responded quite irregularly to temperature. In all cases, an increasing temperature caused a far more negative response when loaded down with a load different than that of the maximum power load. From this, one can see how important it is to maintain photovoltaics at their maximum power load, especially when the temperature effect is considered.

CONCLUSIONS

In this study, experimental results were supported by mathematical analysis. From the results, the following conclusions were reached:

- Silicon solar cells in general respond very negatively to higher temperature. For this reason, in high temperature climates a non-photovoltaic energy system would probably be most cost-effective. Passive solar is one of the better possibilities for a non-photovoltaic alternate energy system because unlike photovoltaics, it benefits from sun-generated temperture rises.
- If electrical generation is needed, amorphous cells would be the best choice for hot climates. While it is agreed that amorphous is less efficient, it is cheaper, so you can buy as much or more power for your dollar than you could if purchasing polycrystalline or single crystal cells. When the temperature effect is factored in to this economic equation, the benefit is clear.

RECOMMENDATION

As more and more people make use of photovoltaics, temperature effect is going to become an increasingly more important problem. It is thus important that ongoing research be carried out to explore means of circumventing or eliminating the effect.

ACKNOWLEDGMENTS

I would like to thank Dr. Donald Garrett and Dr. Gila Garaway for their patience, advice and encouragement throughout this project. Also, many thanks to Dr. Richard Neville who took the time to walk through the mathematics with me. Finally, to Yeshua Hamashiach without whom I would not have finished.

REFERENCES

Fahrenbruch, A. and Bube, R. ,1983, Fundamentals of Solar Cells-Photovoltaic Solar Energy Conversion, Academic Press Inc., London .

Neville, R., 1978, Solar Energy Conversions: The Solar Cell, Elsevier Scientific Publishing Company Inc., Amsterdam.

Autoignition Characteristics of Methanol

Michael Bowman

Department of Mechanical Engineering
Union College
Schenectady , NY 12308

Faculty Advisor:

Richard Wilk

A complete understanding of the ignition properties of alcohol fuels is essential if they are to be considered as alternative fuels. An important concern for a viable fuel is the autoignition tendency. An experimental study was conducted to investigate the autoignition process for methanol. A constant volume reactor was used to determine ignition limits and autoignition temperatures. Experimental conditions which were explored included temperatures in the range of 653- 793 K, equivalence ratios of 0.3 - 17.0, and 1 atm pressure. The process was monitored by following the measured temperature and pressure histories to determine the occurrence of autoignition. Ignition delays were extracted from the pressure profiles and were fit to an Arrhenius type expression from which overall activation energies of 52 - 58 kcal/ mol were determined. Also, the data were used to develop ignition diagrams which distinguish the regions of ignition and non-ignition, as well as show the effects of equivalence ratio on minimum autoignition temperature.

INTRODUCTION

In considering the future of energy conversion via the combustion process the finite supply of petroleum cannot be overlooked. Hence, the feasibility of alternative fuels and/or alternatives methods must be investigated. Methanol has been suggested as an attractive alternative to conventional fuels because of potential reductions in pollutant emissions and because methanol can be synthesized from a wide variety of feedstocks, including natural gas, petroleum, oil shale, coal, wood, and even garbage (Bowman 1975). In order to utilize alcohol as a fuel, a complete understanding of the combustion process is necessary.

An important part of the combustion process is ignition. Ignition is the initiation of combustion. When a fuel/oxidizer mixture ignites spontaneously this is known as autoignition. Autoignition in a spark ignition engine can cause engine knock. In the cylinder of the engine the fuel-air mixture is compressed to high pressures and temperatures at which time a spark ignites the mixture producing a flame front which travels through the combustion chamber. If the conditions are right the unburned end gass can autoignite producing a sharp pressure increase followed by rapid pressure oscillations in the combustion chamber. Engine knock causes an overall reduction in the power developed by the engine and limits compression ratio. Engine knock also can cause possible engine component failure due to the extreme pressure increases, decreases in engine performance because of ill-timed pressure increases, localized overheating, and decreased engine life.

The lowest possible temperature at which autoignition occurs is called the autoignition temperature (AIT). Many previous studies have been devoted to determining these temperatures for various fuels. Depending on the method of determination, the conditions, and a host of other variables, differences in AIT values of over 100°C have been reported for the same fuel. Taking methanol in air as an example, Setchkin (1954) reported an AIT of 425°C, Mullins (1955) found an AIT of

574°C, and Daeyup et al. (1993) approximated an AIT of 577°C ±10°C. This discrepancy leads to the motivation for more complete data.

The important temperature regime for autoignition is 600 - 900 K. In a practical situation, fuels spend a fair amount of time in this temperature range, therefore complete knowledge of a particular fuel's ignition stability is vital. Extensive studies on autoignition have been conducted in the range of 1000 - 2000 K using shock tubes and rapid compression machines. The data obtained from such systems is applicable to specific systems with comparable operating conditions, whereas data from the lower temperatures can be used to determine the minimum AIT and other limiting conditions.

Autoignition is also a safety concern. Knowing the conditions which lead to autoignition is important for the safe handling and storage of fuels.

EXPERIMENT

The experimental system consists of a cylindrical 2010 cm^3 stainless steel mixing vessel and a spherical glass reaction vessel. A heated mixing vessel, equipped with a stirring mechanism and a liquid injection port, is used to ensure a homogenous methanol/oxidizer mixture. The reaction vessel is a spherical 1000 cm^3 Pyrex chamber with three flexible stainless steel to glass fittings. This vessel is enclosed in a cylindrical tube ceramic heater which is used to heat the reaction vessel to the desired

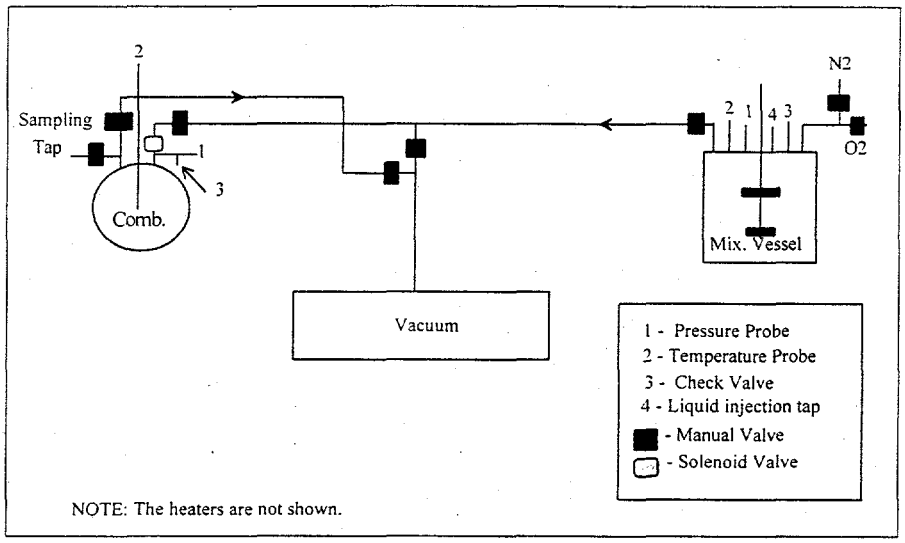


Figure 1 - Schematic of Experimental Facility.

temperatures. The bottom end cap of the heater was fabricated with a three inch view port for observation. The mixing and testing vessels are connected via a transfer line equipped with a solenoid value for rapid transfer of the mixture. The transfer line is heated to avoid condensation in transit. The system is designed to be unidirectional so that the products of the reaction will not flow through the same lines as the fresh mixture. Figure 1 is a representation of the complete testing facility.

In testing the methanol the goals were to determine AIT's at different mixture ratios and to obtain ignition delay data. To accomplish this, 350 μL of methanol (CH_3OH) was mixed with varied amounts of O_2 and N_2. The system was heated to a desired temperature and the mixture was transferred to the testing vessel. The temperature and pressure profiles were monitored real time and observed for the occurrence of autoignition. Autoignition is determined by the onset of rapid pressure rise causing a spike in both the pressure and temperature traces. The time delay was then extracted from the pressure trace by measuring the time from admission of the mixture into the reaction vessel to the onset of the pressure rise.

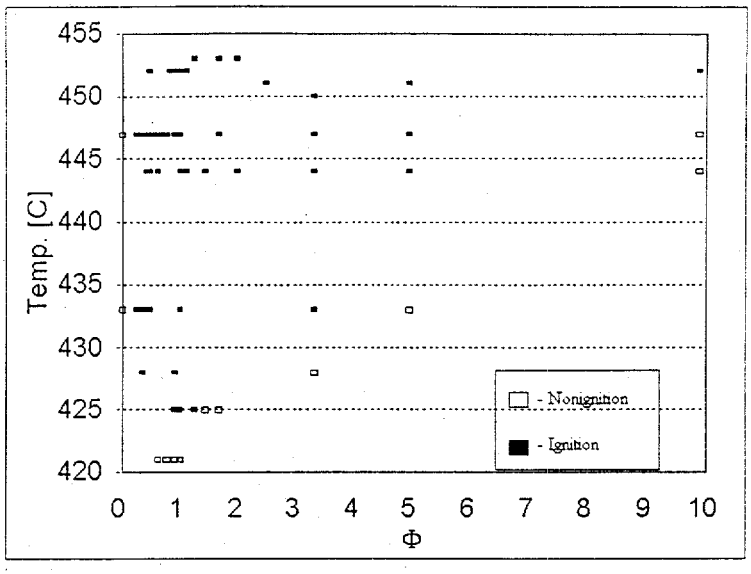


Figure 2 - Autoignition Map for Methanol-Oxygen-Nitrogen Mixtures at P_o=1atm.

RESULTS

The map of the ignition/non-ignition region was determined by plotting the autoignition temperature against the equivalence ratio, shown in Figure 2. Equivalence ratio (Φ) is defined as the actual fuel to oxidizer ratio divided by the stoichiometric fuel to oxidizer ratio. The minimum AIT for methanol was approximated to be 423°C occurring at a Φ of 1.0. This matches up well with previous studies. Setchkin (1954) reported an AIT of 425°C using a similar facility. The map generated also shows how the autoignition temperatures increase for both lean and rich conditions.

The ignition delay data was correlated using an Arrhenius type equation of the form:

$$\tau = A \exp\left(\frac{E}{RT}\right) \qquad (1)$$

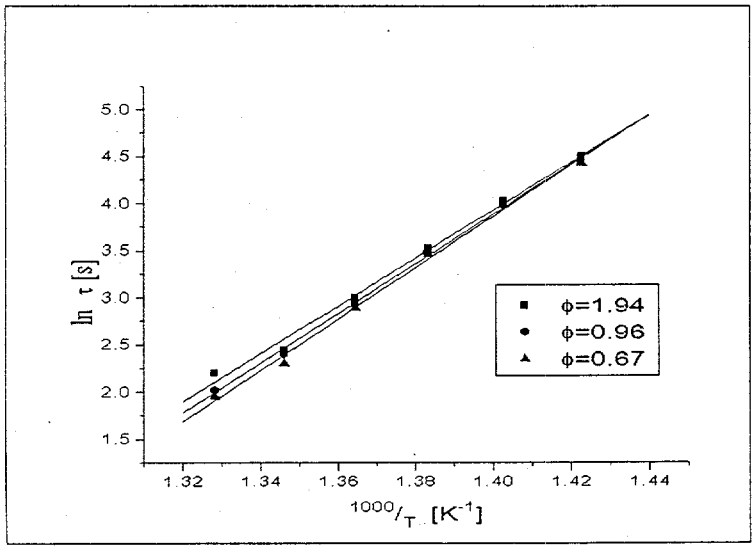


Figure 3 - Linearized Plot Showing the Effect of Temperature on Ignition Delay.

where τ is the ignition delay time, E is the global activation energy, and A is a preexponential constant. Plotting the natural log of the ignition delay, $\ln(\tau)$, against the inverse of the temperature, using constant equivalence ratio data, the global activation energy and preexponential constants can be determined, shown in Figure 3. Table 1 shows the global activation energies for methanol at different equlivance ratios and different initial concentrations of methanol. The average value is 53.9 kcal/mol. This value is comparable with the global activation energy of 56.2 kcal/mol obtained by Nichols (1980). The data that was acquired was checked for reproducibility by running the same mixture multiple times and checking for consistency. Also occasionally random points were rerun to check the validity. All time delay data presented proved to be reproducible with ± 0.1 sec.

	150 μL of CH_3OH			250 μL of CH_3OH			350 μL of CH_3OH		
	ϕ = .59	ϕ = .93	ϕ = 2.2	ϕ = .67	ϕ =1.01	ϕ =2.01	ϕ = .67	ϕ = .96	ϕ =1.94
E	58.4	54.8	53	53.7	52.4	53.2	54.8	53.4	52
Note that units of E = kcal/mol.									

Table 1 - Global Activation Energies for Different Equivalence Ratios and different Initial Concentrations of Methanol.

FURTHER CONSIDERATIONS

Future developments will be to add fuel and oxidizer concentration factors to an empirical fit, i.e.,

$$\tau = A \exp(\frac{E}{RT})[F]^a[O_x]^b \qquad (2)$$

where the exponents a and b represent the dependence of the ignition delay time on initial fuel and oxidizer concentrations. This experimental study does not consider how initial concentration of fuel and oxidizer indenpendently affect AIT values, only equivalence ratio.

Another area that will be explored is the effect of the vessel surface to volume ratio, S/V, on the AIT values. Considering the discrepancies in AIT values pointed out in the introduction the one constraint that has not been accounted for is the vessel S/V. By understanding how the AIT's are affected by the S/V ratio it may be possible to extrapolate the results for AIT values determined to other combustion systems.

ACKNOWLEDGMENTS

I would like to thank New York State Energy Research and Development Authority (NYSEDRA) for the funding of this project and Dr. Richard Wilk for his assistance, as well as the rest of the Union College mechanical engineering staff.

REFERENCES

Bowman, C.T., *A Shock-Tube Investigation of the High-Temperature Oxidation of Methanol.* Combustion and Flame , 25, 343-354, 1975.

Daeyup I., Hochgreb S., Keck J., *Autoignition of Alcohols and Ethers in a Rapid Compression Machine.* SAE Paper 932755, 1993.

Mullins, B. P., *Spontaneous Ignition of Liquid Fuels.* Aeronautical Research and Development North Atlantic Treaty Organization. Butterworths Scientific Publications, London, 1955.

Nichols, R., *Investigation of the Octane Rating and Autoignition Temperature of Methanol-Gasoline Blends.* SAE Paper 800258, 1980.

Setcchkin, N., *Self-Ignition of Combustible Liquids.* Journal of Research of the National Bureau of Standards, Paper # 2516, Vol. 53, 1954.

DEVELOPMENT OF A LIQUID CRYSTAL POLYMER FOR USE AS A BIOLOGICAL MEMBRANE

Eric A. Kutcher

Department of Material Science and Engineering
Cornell University
Bard Hall
Ithaca NY 14850

Faculty Advisor:

Christopher K. Ober

ABSTRACT

Based on the structure of a biological membrane, an artificial membrane was designed. Using the chemical reaction between a dibromide and a tertiary diamine, 1,4-dimethylpiperazine, an ionene was developed with the desired characteristics of the biological membrane. The 1,4-dimethylpiperazine was commercially available, and the dibromide was synthesized. The synthesis of the dibromide occurs in two steps. Using dihydroxymethyl stilbene (DHMS) as the starting material, a diol was synthesized via a second order nucleophilic substitution reaction. The second step in the dibromide synthesis involved another substitution reaction in which the hydroxyl groups at either end of the diol were replaced by bromine. These reactions were monitored using thin layer chromatography (TLC). Techniques such as recrystallization and flash column chromatography[1] were employed for purification of these two steps and the purity was again tested by TLC. Upon completion of each step, Nuclear Magnetic Resonance (NMR) was used to determine the structure of the product. In addition to NMR, Differential Scanning Calorimetry (DSC) was used to look for a glass transition temperature as well as to determine a melting/ degradation point, and finally, FT-IR was employed to determine water absorption of the final polymer.

INTRODUCTION

The purpose of this research was to design an artificial biological membrane that can be used in procedures such as dialysis. The polymer under study is believed to contain many characteristics similar to those of a biological membrane. This specific class of polymer synthesized in this research is known as an ionene. An ionene is a polymer with a quaternary ammonium salt backbone.[2] This quaternary ammonium salt results in a very ordered polymer, and it is this high degree of order that leads to the belief that this ionene can be used as a synthetic biological membrane. Before any definitive conclusion can be made as to the polymer's usefulness as a biological membrane, its ion transport characteristics as well as its single layer thickness must be determined.

The synthesis of this ionene was broken down into three major stages. The first stage was the synthesis of a diol (Fig. 1). This diol was then converted through a substitution mechanism to a dibromide (Fig. 3). Finally, this dibromo compound was polymerized by a reaction with a diamine. The end result of these three stages was a liquid crystalline polymer referred to below as the ionene. Outlined below are the procedures for each stage of the ionene synthesis, and the discussions about these procedures, as well as the results produced thus far. Also, analysis of both the various stages and characteristics of the ionene will be discussed.

EXPERIMENTAL METHODS

There were several stages necessary for the synthesis of the ionene. The synthesis of an ionene involved the reaction between two molecules: a dibromide and a diamine. In this experiment, the diamine used was 1,4-dimethylpiperazine. This diamine was commercially available; hence there was no time spent researching its synthesis. More than ninety percent of the research time involved the synthesis of the dibromo compound. This synthesis involved several stages. The first step towards obtaining the dibromide was synthesizing the dihydroxymethyl stilbene (DHMS). Fortunately there was enough DHMS synthesized from previous work on this ionene that no time was spent creating more.

Synthesis:

Beginning with the DHMS as the starting material, a second order nucleophilic substitution reaction was employed resulting in a diol. The diol referred to throughout this paper has the structure of the compound on the right in the Figure 4.

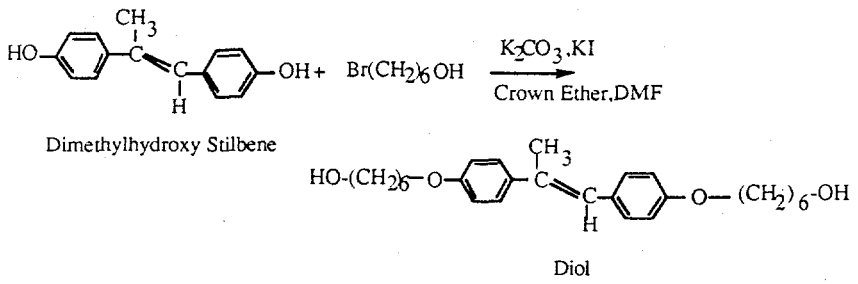


Fig 1: This is the diol reaction in the first stage of the ionene synthesis.

The procedure for this reaction went as follows. First, potassium carbonate (K_2CO_3) was added to the DHMS in a 2:1 molar ratio. To this mixture of solids, potassium iodide (KI) and 18-crown-6 (crown ether) were added in catalytic quantities. This entire mixture was then dissolved in as little DMF as possible and finally heated and stirred at 60°C for approximately 15-20 minutes. Once all the large particles of the K_2CO_3 were broken down to fine particles, the 6-bromo-1-hexanol was added to the solution, again in a 2:1 molar ratio to the DHMS. The temperature of this reaction was raised to approximately 110°C and was stirred at this temperature for approximately 24 hours.

The potassium carbonate was the key to this reaction. The hydrogen ions on either end of the DHMS were removed by the K_2CO_3, leaving behind an anion at either end. (See Figure 2.)

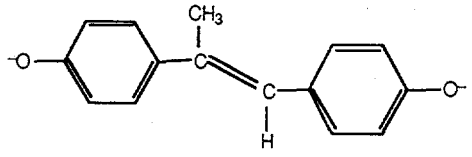


Figure 2: This is the ion created by the reaction of the DHMS and the K_2CO_3

This anion was converted via the second order nucleophilic substitution reaction (S_N2) mechanism to the diol. During this S_N2 reaction, the anions at either end of the molecule in Fig. 2 attacked by the sixth carbon of the 6-bromo-1-hexanol, i.e., the carbon attached to the bromine. At the same time that the bond formed between the carbon and the oxygen ion, a bond broke between the bromine and the carbon. Bromine served as a leaving group for this reaction, enabling the formation of the bond between the anion and the carbon. The result of this reaction can be seen in Figure 1.

Conversion:

Using the previously synthesized diol, the next stage of the ionene synthesis was performed. The synthesis of the dibromide involved a substitution reaction in which the hydroxyl groups at either end of the diol were replaced by bromine ions. The dibromide referred to throughout this paper was the compound on the right in the diagram below.

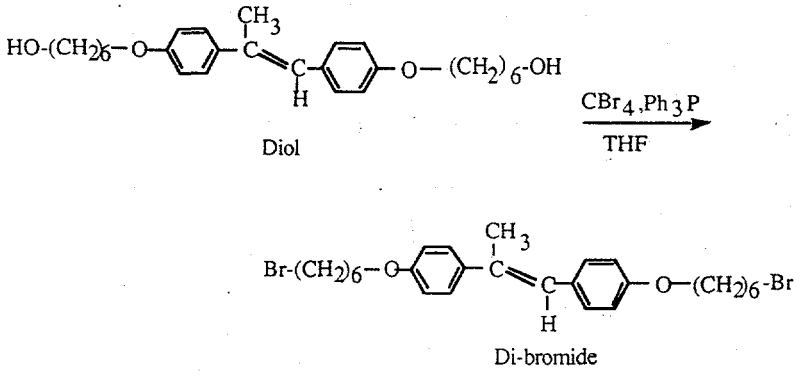


Figure 3: This is the second reaction in the synthesis of the ionene.

The procedure for this reaction went as follows. CBr_4 was added to the diol in a 3.3:1 molar ratio. This mixture was dissolved in as little tetrahydrofuran (THF) as possible and then cooled to 0°C in an ice bath. While the solution was being cooled, triphenylphosphine (Ph_3P) was dissolved in THF and then added dropwise to the cooled solution. This Ph_3P was added in a 3:1 molar ratio to the diol. This dropwise addition was carried out in a nitrogen environment. Once all the Ph_3P had been added, the resulting mixture was stirred in the dark for approximately fifteen hours. Again, this reaction was monitored and terminated using the results of the TLC.

The workup of this reaction, one of the most difficult steps in the ionene synthesis. First, it must be carried out immediately following the termination of the reaction. The first step in this workup was to filter off any product. The solvent that remained after this filtration was rotary distilled off so that only the desired product remains. The solid remaining after the rotary distillation was a yellow gel. Hexane was added to this gel and a white powder precipitated out. Approximately thirty minutes was required for completion of this step. This step of the workup was complete when all of the yellow had turned white. The hexane was filtered off, leaving dibromide and triphenylphosphine oxide behind in the powder. The separation of these two compounds was done using column chromatography with an eluant solvent of 7:1 hexane to ethyl acetate.

Polymerization:

The final stage in this synthesis was the actual polymerization. In this polymerization, a diamine was reacted with the dibromide to form an ionene. The diamine, 1,4-dimethylpiperazine, was added in a 1:1 molar ratio to the monomer (dibromide). A solvent made from a 2:1 ratio of DMF to methanol was added to the mixture, and the resulting solution was heated to 60°C and stirred for two weeks. The compound on the bottom of the figure below was the repeating unit of the ionene researched.

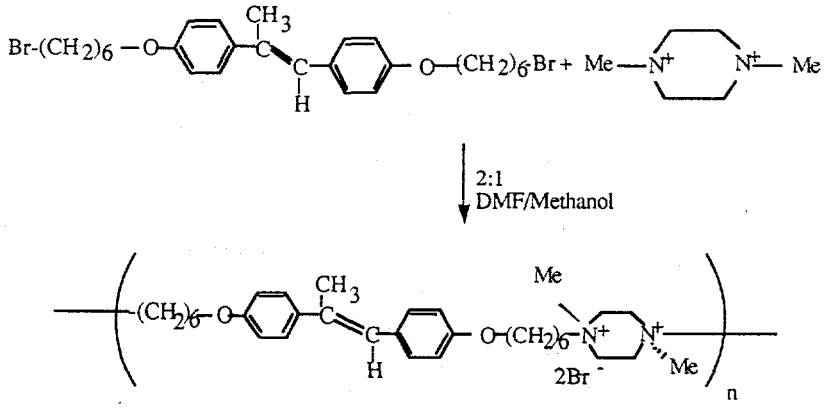


Figure 4: This is a schematic of the final polymerization reaction.

RESULTS AND DISCUSSION

Experimentally, it was found to be critical to the success of the diol reaction that the reaction be permitted to stir in the absence of the 6-bromo-1-hexanol for approximately 20 minutes. The initial stirring at 60°C enabled the K_2CO_3 to attack the hydrogen atoms and thus free up the anions for the nucleophilic substitution. If this stirring and heating were not allowed prior the addition of the $Br(CH_2)_6OH$, the yields were significantly lower as a result of the competing reactions between the K_2CO_3 and $Br(CH_2)_6OH$ with the starting material. Also it was observed experimentally that the highest yields of the diol occurred with a concentrated reaction solution. In a dilute solution there was an increased monosubstituted product which proved unuseful for this synthesis.

This reaction should have had a much higher yield. One possible cause for the low yield of diol might have been a low purity starting material. Because of the difficulty in producing very high purity DHMS, development of the ionene with the use of another starting material, 4,4-biphenol, is under investigation This starting material has already produced significantly higher yields in this diol reaction (approximately 65%).

There were several very important features of the dibromide reaction. First, this reaction had to be carried out in the dark. Bromine was a very reactive element, and light was a source of energy capable of causing the bromine to react. If carried out in the light side reactions occurred, thereby causing a decrease in the yield of desired product and adding to the difficulty in purification. Also, heat was another source of energy capable of causing side reactions with the bromine. It is for this reason that the ideal temperature for this reaction was below room temperature and close to 0°C.

The addition of the Ph_3P to the reaction solution caused an exothermic reaction. The slower the Ph_3P was added the less heat that was given off. It was for this reason that the Ph_3P was added dropwise.

After running several dibromide reactions, it was also noted that the reaction was best carried out in a very concentrated environment. The products of this reaction are mainly of two types: disubstituted and monosubstituted. Since only the disubstituted product was useful, the amount of monobromide produced needed to be limited. If the reaction was concentrated, the dibromide was the dominant product with almost no monobromide resulting.

One of the most important steps in the synthesis of the dibromide was the addition of hexane to the yellow gel in the workup of this reaction. The hexane was useful because it acted to help in the purification of the dibromide product. The diol, CBr_4 and the Ph_3P were all soluble in the hexane, thus any starting material remaining after the reaction was terminated, was extracted from the desired product when the hexane was decanted. After the hexane was decanted, only two compounds remain in the powder. One was the desired dibromide, and the other is triphenylphosphine oxide. The triphenylphosphine oxide was very polar, and thus was easy to separate from the more non-polar dibromide via column chromatography.

Before discussing the synthesis of the ionene, it is important to understand what actually an ionene is: a polymer containing a quaternary ammonium salt in its backbone.

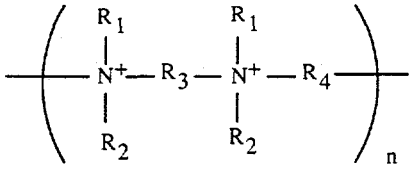


Figure 5: This is a schematic of the general structure of an ionene.

The structure in Fig.5 is a generic repeating unit of an ionene. In this repeating unit there are actually two quaternary ammonium salts (the two nitrogens each having four bonds). Many ionenes are synthesized from a reaction between a dibromo compound and a diamine. In this experiment, the dibromide was that of Fig. 3, and the diamine was 1,4-dimethylpiperazine. The structure of the diamine can be seen in Fig. 4.

For this reaction to be successful, it was essential that the molar ratio of the diamine and the dibromide be 1:1. Any slight alteration in this ratio would cause a very low molecular weight polymer, i.e. very few connected repeating units. A change in this ratio provide the opportunity for side reactions to occur . If the reaction was carried out with the 1:1 molar ratio, the yield of the polymerization was near 100%.

The structure of the ionene was examined in various fashions. First, the structure was checked with the aid of NMR spectroscopy. The results of this shows that the structure obtained was identical to the structure desired and described above. After the structure of the ionene was determined, the melting point and glass transition temperatures were investigated. With the use of the DSC, the melting point was found to be 280.6°C. At this temperature, the polymer did not actually melt, but rather degraded, losing its ordered structure. It was also noted that no glass transition temperature existed between room temperature and 350°C. Taking a close look at the structure of the polymer in Fig. 4, it should be noted that the counterions balancing the charge on the nitrogen ions were bromine. One concern over the use of bromine counterions was their tendency to attract water molecules. This water absorption would make the molecule unstable when exposed to open air. Using FT-IR, this concern was realized. However, this problem was remedied by the change of the counterions to BF_4.[2,4]

CONCLUSIONS

Now that the ionene has been synthesized, it is very important to look at its structure. The whole idea behind this ionene was that it could be used in biological systems, and must therefore have a similar structure to that of a biological membrane.

One of the most recognizable features of a biological membrane is its lipid bilayer. If the ionene is aligned in such a manner as shown in Figure 6, a bilayer very similar to that of a cell is obtained.

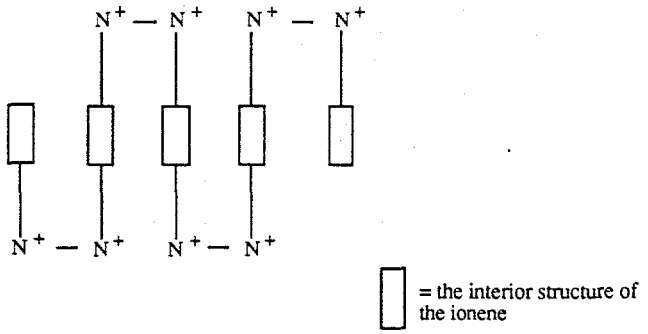


Figure 6: The structure of the ionene when organized as the biological membrane.

There are two characteristics of a cell membrane which need to be incorporated into this artificial membrane. The first characteristic is the membrane thickness. Most cell membranes are on the order of 75 angstroms thick.[3] While it is not possible to convert this polymer into such a thin layer, there are artificial membranes whose functions do not require such thin species. One such application of a thicker membrane used today is the artificial membrane used in dialysis. It is very possible that this ionene could be used in similar applications.

The other characteristic of any living membrane is an active transport pump such as the sodium potassium pump. While it is not possible to incorporate any such pump into this membrane, if the membrane is permeable to certain ions, it is possible to get the same result as with an active pump. A cell is stimulated when it reaches what is known as its threshold potential. Threshold is caused by a depolarization of the cell which is caused as a result of an influx of sodium ions.[5] If the membrane is permeable to certain ions (K,Na,Cl), it is not necessary for such a pump to exist, and the cell can be stimulated by passive transport. While permeability to these ions is necessary, permeability to water is also necessary to aid in the set up the concentration gradients. These gradients will be the driving force necessary for this passive diffusion to take place. Although water permeability is necessary, in order for dialysis to work, water absorption by the membrane is also required. While the changing of the counterions did decrease water absorption, it did not eliminate it, hence giving this ionene both water absorption and transport properties. This only provides more evidence in favor of the use of the ionene as a biological membrane. If the ionene has these permeability properties, different methods will be investigated to created a surface membrane from the polymer.

While the synthesis has been successful, the individual stages will be examined in the future to see if yields can be improved. Already, the yield of the diol reaction has been increased by changing the starting material. The synthesis of the ionene with the 4,4'-biphenol as the starting material is now in progress. The dibromide reaction using the 4,4'-biphenol is under investigation. Hopefully, with these increased yields in the diol reaction and a successful dibromide reaction, an acceptable procedure can be obtained and a surface membrane developed.

REFERENCES

[1] W. Clark Still, Michael Kahn, and Abhijit Mitra, *Rapid Chromatographic Techniques for Preparative Separations with Moderated Resolution,,* Department of Chemistry, Columbia University, New York, 1978.

[2] Gerd C. Hochberg and Rolf C. Schulz, *Synthesis and Properties of Aromatic Ionenes,* Institute of Organic Chemistry, University of Mainz, J.J. Becher-Weg, D-6500 Mainz, Germany, 1993

[3] Arthur C. Giese, Cell Physiology, W.B. Saunders Company, Philadelphia PA, 1968.

[4] Wolfgang H. Meyer and Ligia Domiguez, *Structure and Properties of Ionene Polymers,* Max-Planck Institute for Polymer Research, Mainz, FRG

[5] Roger Eckert, David Randall, Animal Physiology: Mechanisms and Adaptations, 3rd edition, W.H. Freeman and Company, New York, 1988.

J.G. Wang, W.H. Meyer and G. Wenger, *On the Polymerization of N,N,N',N'-Tetramethyldiaminoalkanes with Dibromoalkanes- an in Situ NMR Study,* Max-Planck Institute for Polymer Research, Ackermannweg 10, D-55021 Mainz/FRG.

EFFECT OF ZIRCONIA ON Ni-Al-O METAL-CERAMIC COMPOSITES

Yelena Shapiro

Department of Materials Science and Engineering
Cornell University
Bard Hall
Ithaca, NY 14853-1501

Advisors:

Professor Stephen Sass and Ersan Ustundag

ABSTRACT

Metal-ceramic composites are important potential candidates for future high temperature structural applications. The state of zirconia (ZrO_2) dopant and its influence on the residual stresses in metal-ceramic composites in the Ni-Al-O system were studied using *in situ* high temperature and high resolution X-ray diffraction techniques. Composites were obtained by hot pressing mixtures of powders of zirconyl nitrate and spinel ($NiAl_2O_4$), and then performing partial reduction reactions. ZrO_2 begins crystallizing in the range of 700 - 800°C; in the powder form ZrO_2 is present in the mixed tetragonal-cubic state; upon hot pressing ZrO_2 particles transform into the monoclinic structure and remain in this form during the partial reduction reaction. The initial spinel phase is under a compression stress of 1.07 GPa in the unreduced sample due to the mismatch in the thermal expansion coefficients of $NiAl_2O_4$ and ZrO_2. Nickel particles in the reduced specimen are under a hydrostatic tensile stress of 1.48 GPa due to the mismatch in the thermal expansion coefficients between Ni and alumina.

INTRODUCTION

A metal-ceramic composite is a material that can be designed to obtain an optimum combination of the low density and high strength of a ceramic, and the toughness of a metal. The nickel-alumina composite in the present work is being studied with the hope of using it in jet engines, possibly as a substitute for the currently used Superalloys. For such an application the composite must withstand high temperatures and stresses. The Ni-Al-O metal-ceramic system can give materials with a high Young's modulus (approximately three times higher than that of Superalloys) and it is expected to have as well improved oxidation resistance, strength, fracture toughness, creep and fatigue resistance.

The composite microstructure of Ni particles embedded inside the Al_2O_3 matrix is obtained by partial reduction of the spinel compound $NiAl_2O_4$. Al_2O_3 is one of the most stable oxides, providing excellent oxidation resistance and good high temperature mechanical properties. Ni was chosen because of its usefulness in demanding environments; for example, in Superalloys [1]. Thus, the resultant composite microstructure may be able to satisfy the good oxidation resistance and high temperature mechanical properties requirement of a successful high temperature structural material. It potentially offers a solution to the problem of brittleness of the ceramic component.

The schematic phase diagram in Figure 1 based on the work of Elrefaie and Smeltzer [2] shows the transformation of single phase $NiAl_2O_4$ to two-phase mixture of Ni and Al_2O_3. The reduction reaction is performed in the atmosphere of constant temperature and pressure, by decreasing the oxygen activity level. During the reaction the oxygen is being taken out of the spinel phase, and the resulting two-phase mixture is expected to contain ~20 vol. % Ni and ~ 80 vol. % Al_2O_3.

The sample of pure $NiAl_2O_4$ shown in Figure 2 had a problem of cracking at the grain boundaries during the reduction reaction at high temperatures (such as 1300 °C). The reason for such behavior is not yet completely understood; plausible explanations for such cracking can be stresses generated at the grain boundaries during the reduction reaction caused by the movement of large oxygen ions, and the large volume decrease as a result of exiting oxygen. The cracking problem was solved by doping the system with zirconia (ZrO_2). The uniform microstructure formed with addition of the dopant to the system is illustrated in Figure 3. Zirconia was chosen mainly because it is a very good oxygen conductor; since oxygen ions move relatively rapid in ZrO_2. It is important to determine the atomic structure of the zirconia that is present in order to understand how it prevents the composite from cracking. It is also necessary to measure residual stresses created in spinel as a result of the addition of ZrO_2, as well as the stresses in Ni particles embedded in the Al_2O_3 matrix. The work described here includes the determination of the structure of the ZrO_2 during the various processing steps and the stresses that are present in the final two phase mixture.

EXPERIMENTAL PROCEDURE

The initial step in the specimen preparation procedure was to dissolve zirconyl nitrate ($ZrO(NO_3)_2 \cdot 2H_2O$) in distilled water. (The samples that were used in this study contained 10 weight percent zirconia.) Then spinel powder was added and the resultant solution was mixed using Teflon balls for approximately 6 hours. The next step was to evaporate the water and crush the mixture. Dry powder sieved to obtain a particle size of ~ 6 μm was hot pressed using alumina dies at 1600 °C, with an applied pressure of 25 -30 MPa. The final step was to perform a partial reduction reaction to obtain the two phase composite microstructure.

The analysis of the structure of zirconia was done using the X-ray diffraction technique. A Scintag diffractometer with Cu-Kα radiation was used applying the powder diffraction method, where the wavelength remains constant, and the angle between the incident X-ray beam and the detector (2θ) is varied. Each particle of the powder (or hot pressed platelet) is a tiny crystal, or assemblage of smaller crystals, oriented at random with respect to the incident beam. The mass of powder is equivalent to a single crystal rotated about all possible axes. Each diffraction peak on the produced scan is made up of a large number of small spots, each from a separate crystal particle, with the spots lying so close together that they appear as a continuous line. The scan was then superimposed with the known calculated peaks (found from the computer database) of the compounds involved in the study to find the best fit and identify the composition of the sample, as well as the phases of the various components.

The intensities of the peaks were found using the structure factor. The atomic scattering factor describes the strength of scattering of a given atom in a given direction. It is defined as the ratio of the amplitude of the wave scattered by an atom to the amplitude of the wave scattered by one electron. The scattering by all the atoms in the unit cell is given by the structure factor, F_{hkl}, which describes how the atom arrangement affects the scattered beam, and has the form,

$$F_{hkl} = \Sigma f_n * \exp[2\pi i(hu_n + kv_n + lw_n)] \qquad (1)$$

In this equation, f_n is the atomic scattering factor of the n^{th} atom in the unit cell; h,k,l are Miller indices of the reflecting plane; and u,v,w are the fractional coordinates of the n^{th} atom in the unit cell. Thus, the intensity of any *hkl* reflection can be calculated from a knowledge of the atomic positions, since the intensity of the beam diffracted by all atoms of the unit cell in a direction predicted by Bragg law is proportional to $|F_{hkl}|^2$. [3]

EXPERIMENTAL RESULTS

High Temperature X-ray Diffraction Study of Spinel Powder Mixture

The goal of this experiment is to examine the structures of ZrO_2 simulating hot press conditions without applying pressure with alumina dies. Zirconyl nitrate is amorphous in powder form at room temperature, thus it is expected to crystallize in the mixture of spinel at higher temperatures. The powder was heated only up to 1300°C in air due to the limitations of the diffractometer. (Heating to higher temperatures would require using a vacuum chamber instead of the air atmosphere. Since the hot pressing was performed in air, it was important to keep that condition, limiting the temperature range.) The powder was heated at the rate of 22°C/ minute from room temperature up to 500°C, where the scan was made after holding for 5 minutes. Then scans were made in 100°C intervals (using the same heating rate) up to 1300°C. The powder was held for 15 minutes at 1300 °C before making a scan. The last step of the experiment was to cool down at the rate of 100 °C/ minute to 40 °C (approximately room temperature), where the final scan was taken. Figure 4 shows superimposed scans of one run. The scan in Figure 5 is for the powder cooled down to 40 °C, where all the peaks were shifted to the correct position of the spinel peaks, assuming that spinel is in a stress-free state in the powder form thus its lattice parameter can be used as a reference for the specimens containing zirconia.

The results of this experiment were checked by a second run. The conclusions drawn from the two runs were in agreement.

X-ray Diffraction Analysis of Unreduced and Reduced Specimens

The unreduced specimen was hot pressed using the conditions:
T = 1600 °C, P ~ 25 -30 MPa, t = 4 hours. It was then polished using 600 grit paper, 1200 grit paper and 1 μm diamond paste on a Tex-Met surface. The specimen was analyzed in the diffractometer to identify the phases of zirconia present . Figure 6 shows the scan that compares various phases of zirconia to that present in the unreduced specimen.

Three of the specimens were reduced under different conditions.

1. 1300 °C with $P_{O2} = 10^{-12}$ atm for 4 hours, giving a completely reduced microstructure.
2. 1200 °C with $P_{O2} = 10^{-13}$ atm for 16 hours, giving a partially reduced microstructure.
3. 1100 °C with $P_{O2} = 10^{-16}$ atm for 24 hours, giving a partially reduced microstructure.

All three specimens were analyzed similarly to the unreduced spinel for the phases present as a result of different reduction conditions; special attention was paid to the state of zirconia as compared to the initial powder and unreduced platelet.

X-ray Diffraction Method for Residual Stresses Measurements on Spinel in the Unreduced Platelet and Nickel in the Platelet Reduced at 1300 °C

Residual stresses can be measured by comparing the precise lattice parameters for stressed and unstressed materials. Precise lattice parameters were obtained by using standard silicon powder to align the specimens accurately in the diffractometer. The powder used, Si-640B, from the National Bureau of Standards, is a standard used for sample / machine alignment. The internal correction file is made by running a scan of silicon on a zero-background holder. Then all the following scans are corrected by the computer to match precisely the silicon peaks, thus shifting the entire scan to a better aligned position.

Another important method in the accurate determination of lattice parameters was the plot of lattice parameter vs. Nelson-Riley function $[(\cos\theta)^2 / \sin\theta + (\cos\theta)^2/\theta]$ [3]. That is necessary to avoid the errors introduced by the misalignment of the specimen in the X-ray diffractometer. The experimental data is extrapolated to $\theta = 90°$, where all the errors are minimized. The computer draws a least square fit through the data and the lattice parameter is taken where the value of the Nelson-Riley function corresponds to $\theta = 90°$.

For the unreduced specimen it was important to check the lattice parameter of spinel in accordance with the value obtained from the pure spinel powder. The powder used throughout all the experiments does not have a precise stochiometric relationship between nickel and alumina. The pure spinel powder was used under the assumption that in this form the spinel is in a stress-free state. Figure 7 shows the plots of the lattice parameter of $NiAl_2O_4$ for the powder and the unreduced specimen vs. the Nelson-Riley function.

The lattice parameter of Ni obtained from the reduced platelet was compared with the literature value to determine the presence of stresses due to a mismatch in thermal expansion coefficients between Ni and Al_2O_3.

DISCUSSION

Figure 4 shows that zirconia begins to crystallize in the range of 700 - 800 °C. The emerging peaks are of the tetragonal (or cubic) form of ZrO_2. The resultant microstructure at 40 °C contains zirconia in tetragonal and cubic states (with perhaps a texture or epitaxy) as is seen in Figure 5. Monoclinic zirconia is present in negligible amounts. According to the literature zirconia should be in the cubic state at high temperature (~ 2000 °C) and monoclinic at room temperature [4]. In the present study, the zirconia particles are very small, and they are likely constrained by being embedded in the spinel phase. It seems reasonable to assume that at the beginning of cooling ZrO_2 is in the cubic state and then it partially transforms to tetragonal, while some particles retain their cubic structure. Since the zirconia particles are very small in the powder, perhaps the elastic constraint imposed by the spinel phase prevents them from transforming to the monoclinic structure.

Figure 6 shows that when the powder is hot pressed, the applied stresses apparently cause the tetragonal zirconia to shear into its monoclinic state. The unreduced hot pressed specimen contains spinel and mostly monoclinic ZrO_2. (The tetragonal form of ZrO_2 is present in small amounts).

Zirconia remains in the monoclinic state throughout all the reduction reactions. X-ray diffraction also showed the strong presence of nickel in all three cases. In the sample reduced at 1300 °C, the matrix is α-Al_2O_3, which is hexagonal in structure. In the partially reduced samples (1200 °C and 1100 °C) the structure of the matrix cannot be identified unambiguously. Previous observations [5] showed that a large amount of nickel was present in the matrix phase, which was called "defect spinel". The X-ray diffraction analysis shows that θ-Al_2O_3 (monoclinic phase) gives a good match to the peaks of the matrix.

The residual stress measurements lead to the conclusion that spinel in the unreduced specimens is under hydrostatic compression. The following measurements on the lattice parameter illustrate this point.

Lattice parameter of stress free spinel: $a_o = 8.0463$ Å

Lattice parameter of spinel inside the platelet: $a = 8.0315$ Å

Using Hooke's law for hydrostatic loading with E, the Young's Modulus, and υ, Poissons ratio; the hydrostatic stress P, was calculated with E = 258 GPa, $\nu = 0.279$

$$P = \frac{E(a - ao)}{3(1 - \nu)ao} \qquad (2)$$

P = +1.07 GPa

Spinel is in compression because it has a smaller thermal expansion coefficient than ZrO_2.

Similar calculations were done for the nickel particles in the reduced platelet at 1300 °C. Nickel appears to be in hydrostatic tension. The calculation below supports that claim:

Lattice parameter of nickel [4]: $a_o = 3.5238$ Å

Lattice parameter of nickel in the reduced platelet: $a = 3.5336$ Å

Again, using the expression for hydrostatic loading, eqn. (2), where E=200GPa and $\nu = 0.312$, P = -1.48 GPa.

The thermal stress results on cooling from the reduction temperature of 1300°C to room temperature due to the thermal coefficient mismatch between Ni and Al_2O_3. Nickel is in tension because it has a larger thermal expansion coefficient than Al_2O_3.

In the future residual stresses of the specimens reduced at 1100 °C and 1200 °C need to be measured. The relative amounts of zirconia (monoclinic and tetragonal) in the powder should be

In the future residual stresses of the specimens reduced at 1100 °C and 1200 °C need to be measured. The relative amounts of zirconia (monoclinic and tetragonal) in the powder should be studied. The X-ray determination of the structures of ZrO_2 should be confirmed with electron microscopy observations. In addition, the kinetics of the reduction reactions of samples containing ZrO_2 need to be investigated.

CONCLUSIONS

Zirconia transforms from an initially amorphous state into tetragonal and cubic crystalline structures upon heating, then shears into a monoclinic state upon hot pressing and remains in that form all throughout the reduction reactions performed under various conditions. Zirconia contributes to the residual stresses in the unreduced spinel structure, resulting in the presence of a compressive stress. The Ni in the reduced two phase microstructure is in hydrostatic tension with respect to the alumina.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research under Grant N00014-92-J-1526. The use of the X-ray Diffraction and Materials Preparation Facilities of the Materials Science Center at Cornell University is gratefully acknowledged. The guidance and supervision of Professor Stephen Sass and Ersan Ustundag is greatly appreciated. Also, special thanks for financial support from GE Faculty for the Future program.

REFERENCES

1 M. F. Ashby and D.R. H. Jones, *Engineering Materials I*, International Series on Materials Science and Technology, vol. 34, Pergamon Press, London, 1986.

2 F. A. Elrefaie and W.W. Smeltzer, *J. Electonchem. Soc.*, **128** (1981) 2237-2242.

3 B. D. Cullity, *Elements of X-ray Diffraction*, 2 ed., Addison-Wesley Publishing Co, Inc., Reading, MA, 1978.

4 *Powder Diffraction File Alphabetical Indexes Inorganic Phases Sets 1 - 43*, International Centre for Diffraction Data, Newtown Square Corporate Campus, 12 Campus Boulevard, Newtown Square, Pennsylvania, 1993.

5 E. Ustundag, P. Ret, R. Subramanian, R. Dieckmann and S.L. Sass, *In Situ Metal-Ceramic Microstructures by Partial Reduction Reactions in the Ni-Al-OSystem and the Role of ZrO_2*.. Cornell Materials Science Center, August 1994.

FIGURES

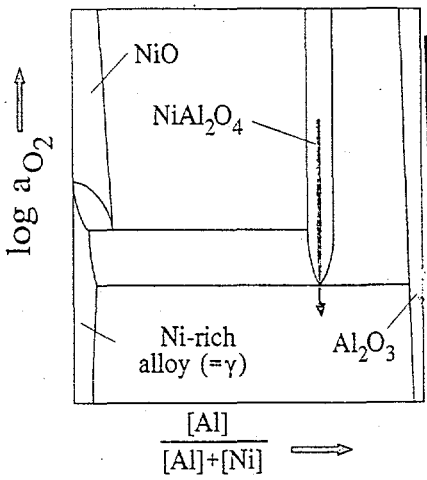


Fig.1 Schematic log a_{O2} vs composition phase diagram of the system Ni-Al-O at constant temperature

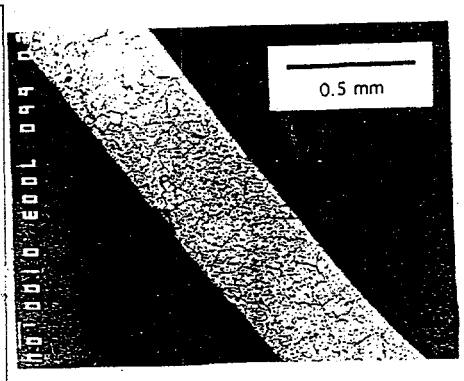


Fig. 2 Cracking due to reduction of pure $NiAl_2O_4$ at 1300 °C at an oxygen activity of 10^{-10}

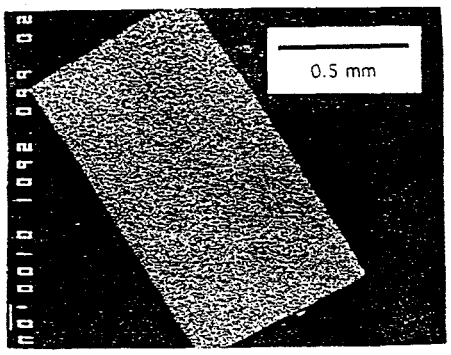


Fig. 3 $NiAl_2O_4$ with 10 wt. % ZrO_2 reduced at 1300 °C at an oxygen activity of 10^{-12}

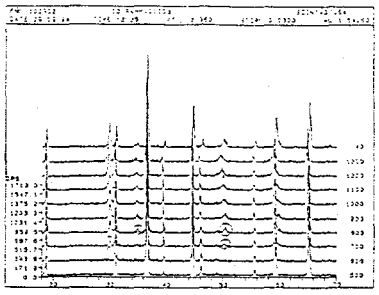


Fig. 4 Superimposed X-ray diffraction scans of high temperature experiment

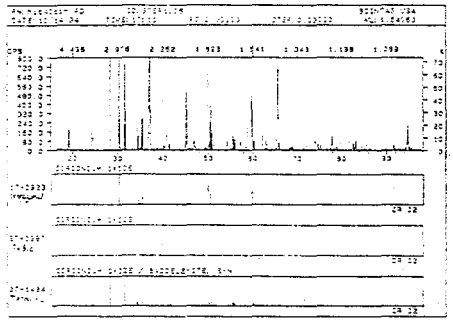


Fig. 5 X-ray diffraction scan of the powder cooled down to 40 °C, and the possible for the state of zirconia in it

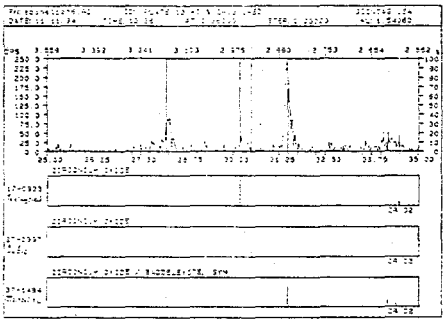


Fig.6 X-ray diffraction scan of the unreduced platelet of $NiAl_2O_4$ containing 10 wt. % ZrO_2

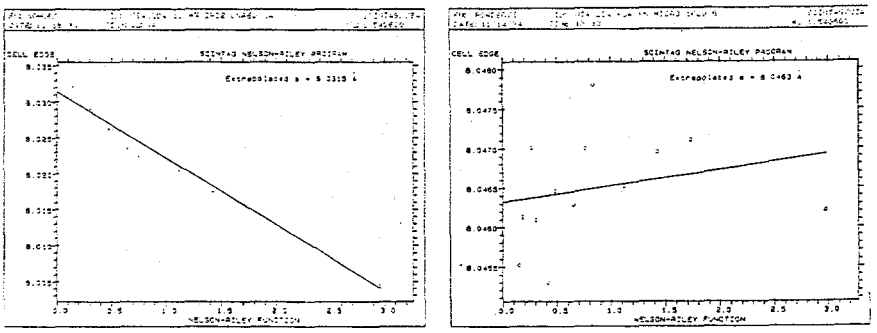


Fig. 7 Right -- Nelson-Riley function vs lattice parameter plot in the powder of pure $NiAl_2O_4$; left -- Nelson-Riley function vs lattice parameter plot of $NiAl_2O_4$ in the unreduced platelet

Healing of Flaws During *In Situ* Formation of Metal-Ceramic Composites by Partial Reduction Reactions in the Ni-Al-O System

Marshall L. Stocker

Department of Materials Science and Engineering
Cornell University
Ithaca, NY 14853

Advisors:

Professor Stephen Sass and Ersan Ustundag

ABSTRACT

Partial reduction reactions in the Ni-Al-O system, starting with the spinel compound $NiAl_2O_4$, are used to form metal-ceramic microstructures *in situ*. Large volume changes associated with the reduction reaction were calculated to generate large stresses. The ability of the compressive stresses to heal microstructural flaws was examined. Experiments showed that the reduction process decreases the initial microstructure porosity of the spinel phase by ~ 9 %. An analysis of crack healing was inconclusive. During this study, it was necessary to develop polishing procedures that avoided the introduction of porosity due to grain pull-out.

INTRODUCTION

The 'concept' airplanes of tomorrow have already been modeled and designed. Yet, what keeps these amazingly fast and efficient airplanes on the drawing table? Simply, current materials used for gas turbine engine blades cannot withstand the additional stresses and higher temperatures that the airplanes of tomorrow will require. Therefore, by developing a material superior to currently used Superalloys the performance levels of gas-turbine engines will increase. In turn, such an advancement in material properties will help move these 'concept' planes off the drawing table and into the air.

The primary requirement of a new gas turbine engine material is having the ability to operate at increased temperature levels while retaining the fracture and strength characteristics of the Superalloys currently used. Ceramics can withstand more strenuous thermal conditions than Superalloys. Yet, ceramics are very brittle (low fracture toughness). Therefore, in the present study, a metal-ceramic composite was examined, which exploits the ability to combine unique characteristics of a ceramic and metal to optimize the mechanical properties appropriate to gas turbine engine requirements. The ceramic microstructure of the composite provides a higher creep resistance, higher melting point, lighter weight, and better corrosion resistance than Superalloys. The metal constituent will give the composite ductility and crack bridging which will increase fracture toughness over ceramics.

Metal-Ceramic Composite Materials

Specifically, the metal-ceramic composite system that was examined was nickel + alumina ($Ni+Al_2O_3$). This system exploits *in-situ* processing wherein the nickel-aluminate ($NiAl_2O_4$) is reduced in a single step to create a $Ni+Al_2O_3$ composite. To synthesize this composite, NiO and Al_2O_3 are sintered to create the nickel aluminate. 5-10 weight percent zirconia is added after sintering to prevent cracking during reduction. This mixture is then milled to less than 6 micron powder and hot pressed. After hot pressing, a pellet with porosity of 1 to 15 percent is obtained. In this form, the spinel is then placed in a reduction furnace where the oxygen is removed from the spinel and the ceramic-metal composite of nickel + alumina is formed. Two distinct microstructures of this composite have been identified(Ustundag et al,1). If the spinel is reduced at temperatures near 1300°C, the nickel takes on the shape of equiaxed spheres within an alumina matrix. Whereas a reduction carried out at 1100°C yields rod-like nickel particles within a metastable 'defect' spinel matrix (Ustundag et al,1).

Stress Generation During Reduciton Reaction

During the reduction reaction, the formation of nickel and alumina from the original spinel causes a volume decrease which is calculated to be 19%. It is believed that this volume decrease superimposes macro-stresses on the microstructure. Examining a theoretical sphere of the composite yields several meaningful observations. It is important to remember that the spinel is reduced from its outer surface towards the inside. This is due to the nature of the spinel's outer

layers reacting in the reducing atmosphere before the oxygen in the middle can be transported to the surface for reduction. Thus, in the sphere, as the outside 'shrinks' due to the volume change from reduction, the spinel that has yet to be reduced undergoes compression. Computing the theoretical stresses within a sphere due to the volume change during reduction yields stresses with a maximum of 7 GPa for a sphere of 2mm radius. There is a variance in this stress which is dependent upon the thickness of the reduced layer versus that of the middle unreduced portion. This calculation assumes the system to be elastic without relaxation which implies that these calculated stresses are the theoretical upper limit possible due to a 19% volume change and that the actual stresses are smaller.

The realization that residual stresses may exist was brought about by the observation of cracking in the spinel (without ZrO_2) microstructure after reduction (Fig. 1). The likely cause of these cracks was the large stresses created by the volume change. This cracked microstructure, therefore, provides direct evidence for the existence of stresses due to the volume change, particularly since many of the cracks are radial.

While these stresses may have initially ruined the microstructure by inducing cracks, the possibility that they could be harnessed for useful application created even greater interest. One use for the large stresses created by volume change during reduction is the healing of flaws which are present within the microstructure prior to reduction. For example, an initially porous spinel may have its pores removed by the compressive stresses.

The ability to heal flaws would be of major significance in the attempt to improve manufacturing capability. For example, mass manufacturing of this composite could be very cheap and efficient if it was not necessary to minimize the flaws in the initial spinel phase. Strict tolerances such as initially low spinel porosity would create notable expenses. Yet, if the reduction reaction could be used to remove initial flaws the Ni + Al_2O_3 composite would be much cheaper to produce.

EXPERIMENTAL PROCEDURE

To determine whether the healing of flaws occurs when the spinel phase is reduced, two specimens were examined for effects due to the existence of stresses created by the volume change. One specimen appeared to be thoroughly cracked throughout (Fig. 2). The other specimen had a high initial porosity of ~12% (Fig. 3). Near perfect cubes were then cut to 2.10 ± 0.1mm edge lengths. While the theoretical stress calculations used spherical geometry, the cube geometry was chosen for the specimens since it was easier to produce than spheres.

A specimen from each of the two spinel microstructures was prepared for scanning electron microscopy (SEM) observation and porosity measurements. During this preparation, a number of polishing problems, primarily artifacts, were encountered. These artifacts and the avenues which were pursued to get rid of them are discussed in the Results and Discussion section. To establish accurate porosity measurements it was necessary to eliminate polishing artifacts. The approach taken to leave only intrinsic porosity was to continue polishing until all further polishing did not cause grain pull-out and did not reduce porosity. After spinel SEM specimens were prepared, micrographs of both the porous and cracked samples were taken. These images were then digitized to measure percent porosity. Porosity levels were obtained by establishing the area fraction of porosity within the digitized image. It is important to note that the area fraction of porosity was assumed to be the same as the volume percent porosity.

Once cracked and porous cubic spinel specimens were generated, the cubes were carefully measured and documented. The reduction reactions were performed in controlled atmosphere furnaces using a partial pressure of oxygen at a level below the stability limit of the spinel phase (Elrefaie and Smeltzer, 2). The two temperatures examined were 1100° C and 1300° C. The time each specimen was exposed to the reducing atmosphere was also varied. For 1100° C, specimens were exposed to the reducing atmosphere for 2, 8, 12, 16, and 24 hrs. For 1300° C, specimens were exposed to the reducing atmosphere for 0.5, 1, and 2 hrs.

Following reduction, the specimens were measured for the percent change in volume. Then each was polished to half of its reduced height to allow for observation of any spinel that might remain in the center of the cube due to incomplete reduction. As was done for the spinel specimen, the reduced specimens were polished until porosity levels no longer decreased and grain pull-out was minimized. Scanning electron microscopy imaging was performed. Porosity measurements were also made and compared to those from the initial spinel (Table 1).

RESULTS AND DISCUSSION

Examining the dimension measurements of the cubic specimens before and after reduction establishes an observed volume shrinkage of ~5% (Appendix). This is 14% less than the calculated percent decrease in volume. To better understand the stresses generated by the volume change, calculations were performed on a spherical geometry which undergoes a 5% volume decrease. As expected, these calculations show that a 5% volume decrease yields maximum stresses significantly lower than those created by a 19% volume decrease. Considering the difference in calculated versus observed volume change, the question arises as to where the additional 14% goes.

Effect of Polishing

Polishing of metal-ceramic and spinel microstructures brings up possibilities of artifacts. The most difficult problem was that of avoiding grain pull-out. Spinel and composite microstructures were initially polished using 600, 800, and 1200 grit silicon carbide paper, successively, on a rotating wet-polisher. Fine polishing was accomplished using 6-micron and 1-micron diamond paste with extender on a long-hair Ceripol cloth. Upon closer examination of the samples, both the spinel and composite samples were excessively porous due to grain pull-out artifacts. The brittle spinel appeared to have entire grains pulled out while the reduced specimens tended to have the softer nickel particles pulled out. To obtain accurate porosity measurements, it is important to limit grain pull-out.

These pull-out artifacts greatly affected porosity measurements since the computer image analysis program is not able to determine which pores are artifacts of the polishing and which are intrinsic. Thus, several polishing techniques were examined for their ability to limit grain pull-out. To determine the extent of the pull-out and the overall amount of porosity present in the specimen, optical microscopy was used to examine the specimens.

The first change in polishing procedure was to apply the diamond pastes with extender to a Tex-Met® short hair nap cloth. The change to short nap had some beneficial effect, but grain pull-out was still greatly affecting observed porosity levels. Then, 1/4 micron diamond paste with extender was used on a Tex-Met® cloth. This change in the size of the diamond grains had no beneficial effect on reducing grain pull-out. Polishing for an extended length of time using the diamond pastes was also investigated. Up to 1 hour of 1 micron diamond polishing did not significantly limit artifacts.

Many additional polishing materials were also evaluated for their ability to limit grain pull-out. 0.05 micron alumina was applied using both a laboratory polisher and a wet-rotary polisher. Using alumina was very destructive to the microstructure and grain pull-out was more extensive than previously observed.

In an attempt to rid the specimens of the damage from the alumina polishing, 1200 silicon carbide paper was used on a wet-rotary polisher. Examination after this procedure showed grain pull-out and porosity to be at a low level. Yet, grain pull-out was still extensive enough to affect the measured porosity levels. Silicon carbide polishing was further examined by increasing the length of time that the sample was polished using 1200 silicon carbide paper. By increasing the time, grain pull-out levels continued to decrease. Eventually, grain pull-out reached a level where additional time using 1200 SiC paper did not reduce grain pull-out. At this point, it appeared as though strictly using SiC paper on a wet-rotary polisher was yielding the best results in terms of limiting grain pull-out. Likewise, it was concluded that diamond paste polishing and alumina polishing were more destructive than beneficial. Diamond polishing using 1 micron paste with and without extender on Tex-Met® cloth was then reexamined. The diamond paste again had no significant effect upon the grain pull-out levels. This procedure seemed to affect the sample by rounding its corners.

To determine if 1200 (6 micron) SiC polishing was going to yield the least artificial porosity (pull-out), 1 micron SiC paper was used on a wet-rotary polisher. This approach appears to be the best at limiting grain pull-out. After preparing specimens using only SiC paper, porosity levels appear to be representative of intrinsic porosity levels with limited pull-out. Following an extensive examination of various polishing techniques, the polishing procedure for both the reduced and spinel specimens was established as follows: 600, 800, 1200 grit and 1 micron SiC paper, successively, on a wet-rotary polisher.

Results of New Polishing Technique

The initial spinel samples that had previously been polished using the original diamond paste procedure were repolished using the current procedure which uses only SiC. Further examination of the spinel showed that the initially porous sample was somewhat less porous, ~ 3% less. After grain pull-out was minimized, this sample was still relatively porous at levels of 12-14% porosity. The initially cracked sample was no longer cracked after the new polishing procedure was applied. The spinel which was originally thought to be cracked in 3 dimensions appears to be cracked only on the surface by diamond paste polishing procedures. Therefore, any examination of crack healing by the compressive stresses created by a volume change would be based on artifacts and difficult to confirm experimentally. Further experimentation is needed to confirm the healing of cracks during the reduction of spinel.

Effect of Volume Change

While the analysis of crack healing was significantly affected by polishing artifacts, the effect of the volume change on porosity levels is more substantial. With grain pull-out at a minimum, the measured porosity during the reduction reaction is given in Table 1. It should be understood that porosity measurements using a digitization process are affected by contrast and image brightness. The extent of variation due to contrast and brightness levels of the digitized

images causes measured porosity levels of a single region to vary by as much as 2%. Porosity levels are nearly the same throughout the material, with variations between different regions of a single specimen to the extent of ±1%. With this in mind, the observed difference in the amount of porosity before and after reduction is significant, so that the minor variations created by the digitization process cannot explain the substantial decrease in porosity which is observed. Porosity of the initially 12 - 14% porous specimen has been measured to decrease to between 2 to 7% (Table 1). The experiments performed are not extensive enough to identify a relationship between reduction conditions and the amount by which the porosity decreased.

Further Questions

Questions that remain include identifying the mechanisms by which porosity is reduced during the reduction reaction. With the examination of crack healing being inconclusive, the stresses which remove pores have yet to be examined for their ability to heal cracks. Perhaps one of the most intriguing questions is "Why does the porosity decrease while the specimen volume does not decrease by the theoretical amount?" If the porosity of the composite decreases and the volume change is calculated to be 19% shrinkage, then the observed volume change should be greater than 19%. It is also important to consider the mechanisms by which the composite relaxes when it is exposed to the calculated high compressive stresses. Plastic deformation and reduction of porosity are two possibilities for the relaxation of the stresses due to the volume change during reduction. The demonstration that such stresses may exist and can beneficially affect microstructure (i.e. by reducing porosity) may ultimately allow the economical production of metal-ceramic composites.

TABLE 1
PERCENT POROSITY

Reduction Conditions	Before Reduction(%)	After Reduction(%)	Δporosity
1300°C			
0.5 hr	13.0	6.0	-7.0%
1.0 hr	13.0	4.2	-8.8%
1.0 hr (2^{nd} run)	13.0	6.3	-6.75%
2.0 hrs	13.0	3.3	-9.75%
Reduction Conditions	**Before Reduction(%)**	**After Reduction(%)**	**Δporosity**
1100°C			
12 hrs	13.0	2.0	-11%
16 hrs	13.0	2.3	-10.7%

SUMMARY

The porosity of the Ni + Al_2O_3 composite produced by reduction was shown to decrease during the reduction reaction. One explanation for this phenomenon is the presence of large compressive stresses generated by a volume decrease during the reduction reaction. The examination of crack healing as a possible result of the compressive stresses yielded no conclusive results. The work also pointed out the necessity of identifying the presence of artifacts generated by the polishing techniques used for ceramics and metal-ceramic composites.

ACKNOWLEDGEMENTS

The guidance of Professor Stephen Sass and Ersan Ustundag was extremely beneficial throughout this project. This research was supported by the Office of Naval Research under grant N0014-92-J-1526 and the Air Force Office of Scientific Research under grant F49620-931-0235. Assistance in polishing procedures provided by Bud Addis was appreciated. Scanning electron microscopy guidance by John Hunt is also gratefully acknowledged.

APPENDIX

Observed Percent Volume Change

C=cracked specimen
P=porous specimen Volume(mm^3):

Conditions	Cinitial	Cfinal	ΔVc	Pinitial	Pfinal	ΔVc
1100°C, 12hrs log a_{O_2}=-16.45	9.952	9.722	-2.3%	9.845	9.283	-5.7%
1100°C, 8 hrs log a_{O_2}=-16.45	10.21	8.947	-12.4%		*	

Conditions	Cinitial	Cfinal	ΔVc	Pinitial	Pfinal	ΔVc
1100°C, 16 hrs log a_{O_2}=-16.45	9.985	9.793	-1.9%	9.712	9.215	-5.1%
1100°C, 24 hrs log a_{O_2}=-16.45			*	9.532	9.143	-4.1%
1300°C, 1 hr log a_{O_2}=-12			*	9.208	8.720	-5.3%
1300°C, 1 hr (2nd run) log a_{O_2}=-12	10.12	10.03	-1.3%	9.433	8.890	-5.8%
1300°C, 2 hrs log a_{O_2}=-12	10.27	9.949	-3.1%	9.697	9.106	-6.1%
1300°C, 0.5 hr log a_{O_2}=-12			*	9.481	8.965	-5.4%

*Note: specimens were lost in reduction furnace and were unretrievable

LIST OF REFERENCES

1. E. Ustundag et al., "*In Situ* Formation of Metal-Ceramic Composites and Ductile Phase Toughened Ceramics by Reduction Reactions," Minerals, Metals & Mat. Soc., (1994), 97- 113.

2. F.A. Elrefaie and W.M. Smeltzer, "Thermodynamics of Ni-Al-O System Between 900 and 1400 K," J. Electrochem. Soc., 128 (1981), 2237-2242.

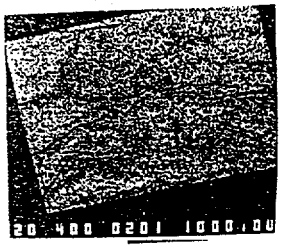


Figure 1 - Nickel + alumina without zirconia after reduction

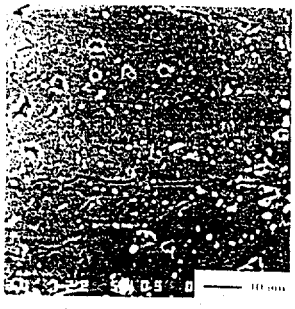


Figure 2 - Cracked specimen

Figure 3 - 13 % porous specimen

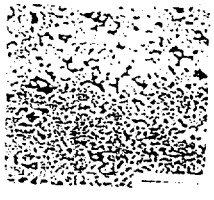


Figure 4 - Spinel grain pull-out

Figure 5 - Spinel with faceted grain pull-out

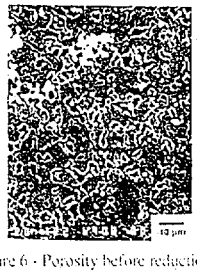


Figure 6 - Porosity before reduction - 9.48%

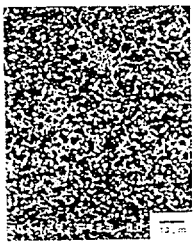


Figure 7 - Porosity after reduction - 2

In Search of Stresses Due to the Volume Change Associated with Partial Reduction of $NiAl_2O_4$

Jessika Trancik

Department of Materials Science and Engineering
Cornell University
Bard Hall
Ithaca, NY 14850

Faculty Advisors:

Professor Stephen L. Sass

ABSTRACT

The Ni-Al-O system has shown evidence of a decrease in volume associated with the reduction of $NiAl_2O_4$. This volume change is believed to have the potential of inducing compressive stresses in an adjacently bonded ceramic object, thereby increasing its resistance to fracture. To measure such stresses, reduction reactions and X-ray diffraction experiments were performed, with a variety of specimens which contained Al_2O_3 within or around the $NiAl_2O_4$ phase. The X-ray measurements detected only tension within the Al_2O_3 phase, instead of the expected compression. Relaxation processes seemed to have decreased the magnitude of the induced stresses. These processes may be related to the addition of 5% and 10% ZrO_2 to the original $NiAl_2O_4$, in order to prevent cracking upon reduction. It was suggested that by careful control of the relaxation processes, it should be possible to produce appreciable compressive stresses in ceramic components.

INTRODUCTION

A successful high temperature structural material, to be used inside jet turbine engines, must have low density, good oxidation resistance, high strength, high fracture toughness, and good creep and fatigue resistance. Most importantly it must exhibit these properties at very high temperatures. The development and analysis of a nickel-aluminate composite microstructure which may satisfy this oxidation resistance and high temperature mechanical properties requirement has led to some significant discoveries[1].

The reduction of $NiAl_2O_4$ at 1300°C proceeds by the reaction $NiAl_2O_4 \Rightarrow Ni + Al_2O_3 + 1/2O_2$. Cracking was observed in pure spinel upon reduction. It was found that the addition of between 2.5 and 10 wt% ZrO_2 prevents cracking. A large volume decrease, calculated to be approximately 19%, is expected during the reduction reaction. This change in the specimen volume should occur upon complete reduction. The volume decrease between the original spinel and the Ni and alumina microstructure, however, was measured experimentally to be only 5% [1]. Mechanics calculations predict the existence of high compressive stresses produced in a material bonded to the reduced composite phase, reaching a range of 10-20 GPa, for a volume change of either 19% or 5%.

Calculations of stresses produced in the nickel and alumina two-phase mixture, which would induce compressive stresses in an adjacently bonded ceramic, were done assuming that at a specific instance the reduction reaction stops with a disk sample which is unreduced in the center and reduced toward the edge, and assuming that the sample is a perfect sphere and the material is purely elastic. None of these factors were maintained exactly, however, because at temperatures of 1100°C and above, the material exhibits some plasticity, the samples are cubes, and the reduction reaction is an ongoing one. The compressive stresses calculated for the inner unreduced region, in the range of 10-20 GPa, are therefore believed to be an upper limit. If this inner, unreduced region were instead Al_2O_3, the interface between the composite and Al_2O_3 could not survive such high stresses. For this reason, relaxation of stresses is believed to occur. With the magnitude of stresses calculated, however, even as an upper limit, some stress is expected to remain after relaxation.

Originally, ZrO_2 was added to the spinel because it was observed to prevent cracking [1]. Samples with no ZrO_2 were found to crack upon reduction (Fig. 1), whereas samples with 2.5%, 5% and 10% ZrO_2 did not crack. The ZrO_2 was therefore believed to alleviate the "micro-stresses" induced in the $NiAl_2O_4$ upon reduction. More recently, it was suggested that ZrO_2 has the ability to alleviate residual stresses of high magnitude. The exact mechanisms of relaxation are currently under investigation.

After the reduction reaction, upon cooling, the two phase mixture of $Ni + Al_2O_3$ incurs internal stresses due to a thermal expansion coefficient mismatch between Ni and Al_2O_3. Ni, with a larger thermal expansion coefficient, contracts more than Al_2O_3. This difference places the Ni under a large hydrostatic tensile stress and the surrounding Al_2O_3 in

a state of compression. These stresses, termed "micro-stresses", differ from the "macro-stresses" due to the volume change caused by the reduction reaction of $NiAl_2O_4$.

The thermal stresses are somewhat analogous to what is expected to occur on a larger scale when the spinel is reduced. In principle, it should be possible to use the volume decrease obtained during reduction to induce compressive stresses in a ceramic object to which the spinel is bonded. If these stresses could be controlled and applied, they could prove to be significant for increasing the fracture toughness of ceramics by a process similar to the strengthening of tempered glass [3]. Hence the intrinsic brittleness of ceramics could be avoided. There are many potential commercial applications for a high-strength ceramic with a high fracture toughness.

The purpose of the present research is to determine whether the volume change associated with a reduction reaction can generate large magnitude stresses that can be used to place a ceramic object in compression. The project was performed by bonding the spinel phase, $NiAl_2O_4$, to either an alumina tube or a rod, and then reducing the spinel to form either a two-phase mixture of $Ni+Al_2O_3$ or Ni+ "defect spinel" [1]. X-ray diffraction techniques were used to check for the presence of stresses in the Al_2O_3.

THERMODYNAMIC BACKGROUND

In order to study the reduction reactions controlling metal-ceramic microstructures in the Ni-Al-O system, knowledge of the effect of the oxygen partial pressure on the equilibrium phases that exist in this system is necessary. Detailed phase diagrams for the Ni-Al-O system at 940°C and 1000°C were obtained by Elrefaie and Smeltzer [4] . A schematic phase diagram for this system, based on their work, is shown in Fig. 2 for 1200°C. It can be seen on this diagram that a two-phase mixture of Ni-rich alloy and Al_2O_3 is expected to form when $NiAl_2O_4$ is reduced by lowering the oxygen partial pressure in the vicinity of the specimen. Üstündag et. al. [1] have shown that the actual phases produced depend upon the reduction temperature.

A metastable ceramic phase forms when the sample is reduced at 1100°C for 24 hours, which has been termed "defect spinel". Nearly pure rod-like Ni particles are found inside this matrix. Equiaxed Ni particles formed in the equilibrium phase Al_2O_3, upon reduction at 1300°C for 4 hours.

EXPERIMENTAL PROCEDURE

Powders of NiO (99%, grain diameter < 20μm) and α-Al_2O_3 (grain diameter <6μm) were mixed together using a ball-mill, until a homogeneous mixture was obtained. This mixture was then reacted in air at 1600°C for 4 days to yield $NiAl_2O_4$, which was crushed to a powder of less than 6 microns in grain diameter and sieved. For the addition of ZrO_2, a water soluble compound, zirconyl nitrate ($ZrO(NO_3)_2 \cdot 2H_2O$), was added in appropriate amounts to $NiAl_2O_4$ powder and then wet ball-milled in distilled water using Teflon balls. This procedure allows for the uniform distribution of the dopant within the polycrystalline spinel. The dried mixture was then compacted in an alumina die with the addition of extra pieces of alumina as needed by the specific sample geometry. The die was placed under 25-30 MPa pressure in a uniaxial hot press for 4 hours in air at 1600°C. The die was later cut so as to form the correct sample shape (Fig. 3).

Great care was taken in order to avoid cracking or otherwise damaging the sample upon cutting and polishing the die to obtain the desired shape. Since the strength of the interface was not known, separation of the Al_2O_3 and $NiAl_2O_4$ phases had to be carefully avoided. Samples were precisely mounted and held in place by a vacuum holder, then were cut and polished using a Conrad diamond wheel. The uniformity of the sample height was carefully checked and the diameter measured. Several specimens of each configuration were cut and then given different reduction treatments, with one piece remaining unreduced for comparison.

As can be seen in Fig. 3, the samples are of four different configurations, consisting of a disk with an Al_2O_3 outer layer and a spinel inner disk (Fig. 3a), a disk with an outer layer of spinel and an inner alumina disk(Fig. 3b), two layers of alumina with a layer of spinel in the middle (Fig. 3c), and two layers of spinel with a middle layer of alumina (Fig. 3d). Three different reduction conditions were used in order to determine the dependence of the stress generation on the reducing temperature. Samples were reduced at oxygen activities four orders of magnitude below the stability limit of the spinel at 1100°C for 24 hours, 1200°C for 8 hours, or 1300°C for 4 hours.

After reduction, specimens were mounted and analyzed using X-ray diffraction (XRD) for residual stress measurements as well as phase identification. The residual stresses in the external Al_2O_3 were calculated from the measured induced strain, assuming elastic behavior. The strain was found using XRD, by measuring the change in atomic plane spacing, of planes parallel to the surface and those at an acute angle to the surface. From this strain, the stress in the sample was calculated. Flat mounting proved to be crucial for correct analysis since a small change in the sample height could incorrectly indicate the presence of a

residual stress. For the samples which had formed a two-phase mixture of Al_2O_3 and Ni after reduction, lead tape was placed over the surface of these areas in order to measure the stresses present in the external Al_2O_3 , blocking a contribution from the Al_2O_3 obtained from the reduction of $NiAl_2O_4$.

Great care had to be taken in order to avoid misalignment of the X-ray equipment as well. Small deviations in machine alignment were minimized, since these could also produce the appearance of a non-existent stress. Two different methods were used to align the X-ray equipment: a stress-free iron powder was analyzed at the sample height and the machine was aligned until the peak positions indicated the presence of no stress, and a position sensitive detector was used to measure the precise height of the sample which was then adjusted accordingly.

EXPERIMENTAL RESULTS AND DISCUSSION

Through careful cutting and polishing, the sample interfaces remained intact. There seemed to be some diffusion of spinel into the alumina, observed by a change in the color of alumina at the interface, which was expected to yield fairly strong bonding at the two-phase interface.

After reduction, however, certain samples cracked and separated at the interface. The inner disk of Ni and Al_2O_3 in the disk shaped sample (Fig. 3a) fell through after reduction at 1300°C for 4 hours, separating from the outer layer of Al_2O_3. It was believed that this was due to the magnitude of tensile stresses at the interface. This observation is evidence for the occurrence of a volume decrease associated with reduction. The same disk-shaped samples reduced at 1200°C and 1100°C both held at the interface, although the 1200°C sample, reduced for 8 hours, showed some cracking and a groove at the interface. These results seemed to indicate that the magnitude of the tensile stresses, within the reduced region, increases with increasing reduction temperature. Since the reduction goes to completion at 1300°C and is expected to cause the largest volume change of the three temperatures, the tensile stresses at the interface will also be the highest, possibly causing the observed separation described above. It should be noted that a 1% decrease in diameter was measured for the sample configuration depicted by Fig. 3a, (Üstündag, private communication), indicating that an approximately 3% decrease in volume occurred due to reduction.

The two different disk-shaped samples (as shown in Fig. 3) were analyzed at TEC (Knoxville, TN), by measuring the plane spacing at different angles to the sample surfaces. The other sample geometries will be analyzed in the near future once their correct alignment can be achieved.

Stress measurements made on the outer alumina layer of sample configuration Fig.3a, showed the unreduced outside edge to be under a tensile stress of $+99.2 \pm 38.9$ MPa. The large estimated error is due to low counting statistics [2]. At the center of the alumina layer, the stress was measured to be $+90.2 \pm 13$ MPa, and at the interface between spinel and alumina the measurement showed a $+33.9 \pm 18.5$ MPa stress. The same sample geometry reduced at 1100°C showed a $+50.1 \pm 24.1$ MPa stress at the outside edge of the alumina, and a stress of $+70.6 \pm 13.4$ MPa at the interface between the defect spinel and the alumina.

The expected results for this disk-shaped sample were that the outside alumina ring would be under low stresses before reduction and under a fairly high residual, compressive stress after reduction, due to the large volume decrease of the inner spinel disk. The spinel will contract during reduction, thereby pulling in the outer layer of alumina. This compressive stress in alumina was expected to show its highest value at the interface. Thus the measured results were unexpected. They have evoked some uncertainty due to large margin of error, up to 50% in some cases. With such microscopic measurements, high precision is essential. The measurements may be unreliable due to reasons which will be addressed in the following paragraphs.

Unexpected results were also obtained from the disk-shaped sample with an outer spinel ring and an inner alumina disk (Fig. 3b). At the center of the Al_2O_3 disk, in the unreduced specimen, the stress was measured at $+61.6 \pm 24$ MPa, and at the interface, a stress of $+123.7 \pm 36.5$ MPa was measured. In the sample reduced at 1300°C, very high compressive stresses were expected to form in the inner alumina disk. The measurements showed, instead, a decrease in tensile stresses at the center of the alumina disk, to $+51 \pm 15.9$ MPa and a stress of $+42.3 \pm 18.8$ MPa at the interface. This is a trend in the right direction, but the low magnitude of the trend was unexpected. It must be noted that all of the residual stress measurements are studying stresses within 10 microns of the surface, due to the low penetration of the X-rays. Thus it is possible that the stresses deep within the specimen are different in magnitude and sign from those close to the surface. Work is in progress to check this possibility.

The results of the residual stress measurements by X-ray diffraction, performed at TEC, were not what had been expected. When going from unreduced spinel to the reduced two-phase mixture, there was clear shifting of characteristic peaks of Al_2O_3 in the correct direction, i.e. from high tensile stresses to low tensile stresses, but the shifts did not indicate an initially unstressed material becoming compressed.

There are several ways in which errors could have been introduced into the measurements of surface residual stresses. One primary source is, as mentioned before, sample height (i.e. the displacement of the sample surface from the diffractometer axis [2]). Slight deviations in sample height can distort stress measurements. Similarly, improper machine alignment, during the radial motion of the X-ray counter, can produce error. Even if both the sample and the machine are aligned correctly, problems may have arisen due to the method of measurement. With the X-ray diffraction technique used, there is an assumption made that the interplanar spacing varies as the incident angle of the X-ray beam increases, relative to the value of interplanar spacing of planes parallel to the surface. This assumption does not provide for the presence of a hydrostatic stress, which would change the interplanar spacing equally at all incident angles.

To measure this possible change in the original interplanar spacing of alumina and the interplanar spacing of the alumina in a reduced sample, for the planes parallel to the surface, a different technique is currently being used. This method applies a standard silicon powder (from the National Bureau of Standards) as an internal calibration, to remove the error due to the sample height mis-alignment. The well-known characteristic peaks of silicon clearly show any deviations in expected position, due to specimen height errors, which can then be applied as a correction factor to the peaks of Al_2O_3. The method involves lightly sprinkling silicon powder over the surface of the sample to be measured by X-ray diffraction. The plane spacing parallel to the surface of the Al_2O_3 in an unreduced sample is measured and compared to the plane spacing parallel to the surface in a reduced sample. From this comparison, a hydrostatic stress can be detected and measured.

In principle, there must be induced stresses in Al_2O_3 bonded to the Ni and Al_2O_3 microstructure, in the specimens discussed above, since the specimen diameter and volume decreased during reduction. These stresses have not yet been detected. Initial stress measurements done at TEC, showed somewhat inconsistent results and low magnitudes.

The validity of the mechanics calculation, predicting the existence of high compressive stresses in the range of 10-20 GPa corresponding to a 5-19% volume change of the reduced composite phase, is currently being tested. The plastic relaxation processes of the material may be efficient enough to relax all high-magnitude residual stresses. It is possible, however, that the best method of measuring these residual stresses has not yet been found. The mechanics calculations are sufficiently convincing to warrant further research into the subject.

It may be possible to decrease the amount of ZrO_2 so that cracking is avoided, yet the accumulated compressive stresses are not entirely relaxed. For this to occur, the tensile stresses in the reduced layer must be below the fracture stress value, yet must be large enough to produce an appreciable hoop stress within the outer ring in Fig. 3b.

With the presence of a material in tension producing compressive stresses in another material, the stresses at the interface between the two materials will be high. The interface can perhaps be strengthened, however, thus allowing the magnitude of the induced stresses to increase without causing separation of the adjacently bonded materials. Development of the interface strength is crucial for potential commercial applications. Longer hot pressing times at 1600°C and increased loading pressure could cause the interface to become more graded, less abrupt, and thus stronger. This graded behavior was already seen to a small, extent and could possibly be improved. More hot-pressing runs will be performed in order to identify and control these significant parameters.

SUMMARY

X-ray measurements were used to check for compressive stresses produced due to the volume decrease of spinel during reduction, in an adjacently bonded ceramic. The measurements thus far have not succeeded in detecting high magnitude compressive stresses. The addition of ZrO_2 may have released the build-up of these stresses. It is suggested that careful control of the amount of ZrO_2, in future work, may allow for the production of compressive stresses in the adjacent ceramic, while avoiding separation of the interface or cracking in the spinel, upon reduction.

A ceramic strengthened by surface compressive stresses would preserve its high strength, low density, good oxidation resistance, and good creep and fatigue resistance, while gaining a high fracture toughness as well as a high impact resistance.

ACKNOWLEDGMENTS

The guidance of Professor Stephen Sass and Ersan Üstündag was invaluable for the performance of this project. This research was supported by the Office of Naval Research under Grant N0014-92-J-1526 and the Air Force Office of Scientific Research under Contract F49620-931-0235. The use of the X-ray Diffraction and Electron Microscopy Facilities at Cornell University, funded by the National Science Foundation under Grant No. DMR-91-21564, is greatly appreciated. The assistance of Maura Weathers with X-ray measurements is gratefully acknowledged. Measurements of the residual stresses parallel to the specimen surfaces were performed by TEC, Knoxville, TN.

REFERENCES
1. E. Üstündag, R. Subramanian, R. Dieckmann and S.L. Sass, In Situ Formation of Metal-Ceramic Microstructures., (1994).
2. B. D. Cullity, Elements of X-ray Diffraction., Addison-Wesley, (1978).
3. J. Zarzycki, Glasses and the Vitreous State, Cambridge University Press, (1991).
4. F. A. Elrefaie and W.W. Smeltzer, J. Electrochem. Soc., 128, 2237-42 (1981).
5. Engineering Materials Handbook, v. 4 - Ceramics and Glasses, ASM, Materials Park, OH (1991).
6. M. F. Ashby and D. R. H. Jones, Engineering Materials I, International Series on Materials Science and Technology, 34, Pergamon Press (1986).

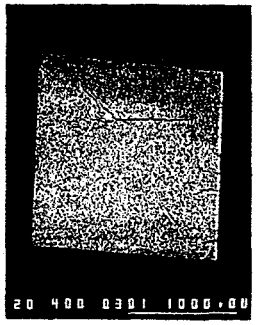


Fig. 1: Cracking in a $NiAl_2O_4$ sample reduced at 1300° C for 1 hour, log aO_2 = -10.54.

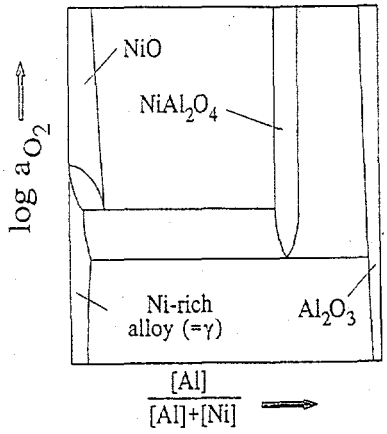


Fig. 2: Schematic log aO_2 vs composition phase diagram at constant temperature for the Ni-Al-O system.

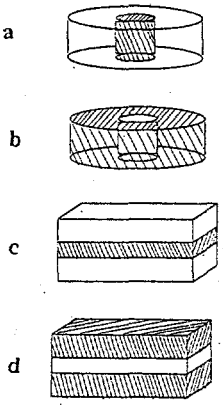


Fig. 3: Specimen configurations.

Testing of a Two-Axis Laser Doppler Velocimeter in the Laboratory and Under Field Conditions in the Chesapeake Bay

Eric L. Luft

Department of Mechanical Engineering
University of Maryland
College Park, MD 20742

Advisor:
Lawrence P. Sanford

Horn Point Environmental Lab
Center for Environmental and Estuarine Studies
University of Maryland
PO Box 775
Cambridge, MD 21613

ABSTRACT

Recent developments have made the laser doppler velocimeter (LDV) technology available in field-ready packaging that should allow an LDV to be used in the field and in the presence of viable organisms (Figure 1). Prior work done with similar instruments (Agrawal and Aubrey[2], and Agrawal *et al.*[3]) suggests that this LDV can be used to measure turbulence in boundary layer flows as well and to measure very slow flows. This instrument was tested, both in the lab and in the Chesapeake Bay, with the goals of measuring turbulent flows on a small scale and resolving very slow flows. The LDV was used with the probe immersed in a water flow and with the probe "looking" through a plexiglass window "into" a flow. Although one of the two measurement axes failed to function properly, velocity measurements from the remaining axis show that the instrument has promise. Turbulent velocity fluctuations were measured and their power spectra were examined. The velocity power spectra agreed well with the theory of isotropic turbulence. In other testing, very slow flows were resolved with good precision. Success in these two areas is vital to understanding the patterns of flow and mixing that define an aquatic environment.

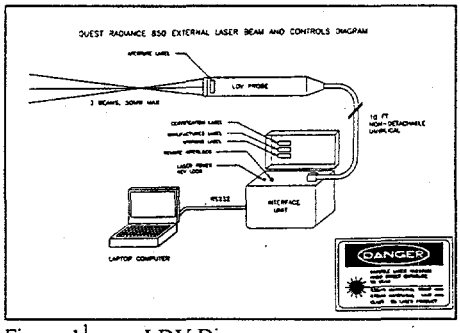


Figure 1[1] LDV Diagram

FIELD TESTING IN THE CHESAPEAKE BAY

Field testing of the LDV took place July 29 and the actual profile was done in the northern Bay just off Pooles Island. The LDV was mounted on a current-oriented profiling rig that would always be oriented into the current flow (Figure 2). Even though only one axis of the LDV was functioning,

as long as the LDV's working axis and the profiling rig's longitudinal axis were parallel, the one functioning axis would "see" the full current strength. In order to provide another velocity time series for comparison, a Marsh-McBirney (MMB) flow meter was mounted next to the LDV's remote probe on the profiling rig (Figure 3). The MMB was not dependent on orientation in order to measure the magnitude of the current velocity,

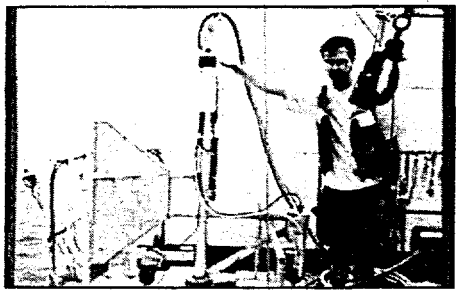


Figure 2
Profiling rig with orientation fin on left

Figure 3
LDV remote probe on left
MMB on right

because it had two functioning measurement axes. Velocity time series were recorded at three depths, with the variability taken to be one standard deviation from the mean (Table 1).

Table 1: Mean and Standard Deviations of Flow Velocities

Depth (m)	LDV mean flow (cm/s)	LDV std	MMB mean flow (cm/s)	MMB std	LDV > MMB %
1.6	46.4	2.6	40.0	2.1	16
3.8	32.1	3.4	26.2	1.6	22
7.4	12.5	2.3	9.3	1.8	34

The time series with the best signal strength (Figure 4) was recorded with the profiling rig resting on the bottom of the bay at a depth of 7.4 meters. Signal strength was dependent on

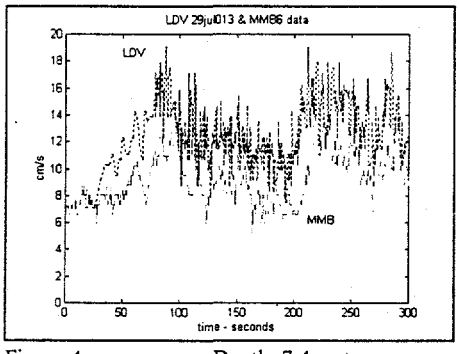


Figure 4 Depth: 7.4 meters

the strength and frequency of the backscatter from suspended solids in the water. Time series were also recorded near the surface and at a middle depth, but the signal was not consistently strong in either of those two cases. This was probably because of settling of the suspended solids in the water. Closer to the bottom, the quantity of suspended

solids was greater and the backscatter was stronger. The velocity time series recorded by the LDV and the Marsh-McBirney at 7.4 meters follow each other closely (see figure 5). The LDV velocities averaged about 24% higher, which follows a trend observed in other testing,

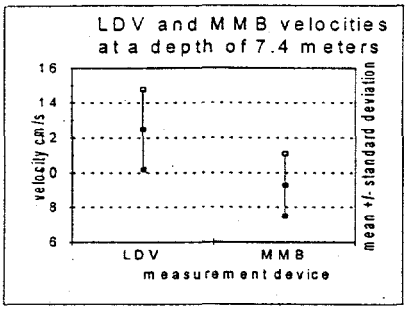


Figure 5 Velocities at 7.4 meters

where the LDV measured velocities that were 20% higher than the velocities given by the calibration of test equipment.

The data record taken on the bottom of the bay was dense enough for spectral analysis to be performed successfully. The theory of isotropic turbulence predicts that the shape of the turbulent energy spectrum will be given by

$$E(k)=A_2\varepsilon^{2/3}k^{-5/3} \qquad (1)$$

where A_2 is an empirical constant that varies with the number of velocity components under consideration, ε is the dissipation rate, and the wave number k takes on values in the inertial subrange[4]. This distribution plotted on a log-log scale would be represented by a line of slope -5/3. The power spectral density (PSD) plot (Figure 6) shows an inertial subrange in the turbulence which manifests itself as a line of slope -5/3 most visible in the plot's low to middle frequency range.

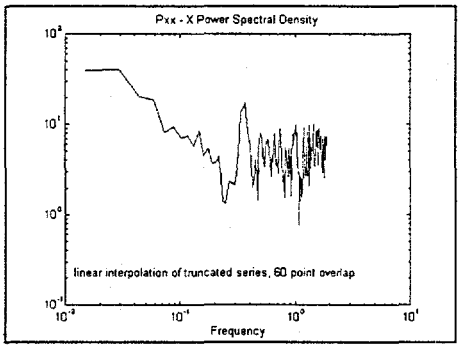


Figure 6 PSD plot of Pooles Island data

This is the anticipated response for an accurate measure of turbulence in a boundary layer such as the bottom of the Bay. Two kinds of noise manifest themselves in the PSD plot. Laboratory tests showed that motion of the fiberoptic cable that connects the remote probe to the interface unit would distort the signal. Therefore, it was not unexpected that oscillatory motion of the fiberoptic connecting cable would introduce periodic noise into the signal. The cable was exposed to current and wave motion, and so the resulting oscillation of the cable would be expected to introduce periodic noise into the signal. The especially strong peak at 0.3 Hz is significant because it fulfills that expectation. A frequency of 0.3 Hz implies a period of 3 seconds for the oscillation of the cable. There is also noticeable high-frequency noise, probably aliased from frequencies above the effective sampling rate. The effective sampling rate was low in part because of problems with LDV's beam strength. The absolute level of the spectrum is somewhat uncertain due to the calibration uncertainties.

RESOLUTION OF VERY SLOW FLOWS

The LDV's ability to resolve very slow flows was tested in a small flume tank (Figure 7). This flume stood about three feet high and used an electric motor and a mixing blade to pull the water through PVC piping and along a rectangular plexiglass channel. The flume's

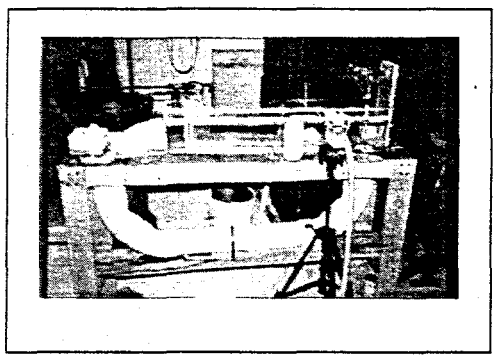


Figure 7 Small flume, shown with LDV remote probe on tripod, right

open section had a cross section measuring 5.0 inches by 5.0 inches. The water volume was sufficiently small that it was easy to seed the small flume with reflective particles. The resulting increase in backscatterance was evident in a marked increase in the strength of the signal. This increased backscattance made up for the loss in signal strength due to the LDV malfunction. Prior to working with the LDV, a series of time trials with dye spots and a stopwatch which showed that at the motor's maximum RPM the flow velocity of the flume was about 2 cm/s. A limited profile of the tank was done with the LDV, measuring velocities at several depths, and varying the resolution of the LDV for the greatest possible precision.

Two more profiles were done with lower resolution which allowed for a denser time series. One such profile was done approximately one inch from the entry grating of the flume, and the other approximately one inch from the exit grating (about 15 inches downstream). These profiles consisted of one or more time series taken at depths ranging from 0.5 inches to 4.0 inches. The mean flow velocities varied from 2.2 ± 0.1 cm/s with a bin size of 0.01 cm/s to 0.5 ± 0.03 cm/s with a bin size of 0.002 cm/s.

Figures 8a and 8b are PSD plots that illustrate turbulence at comparable points in the profiles done near to and far from the entry grating. The plots show a transfer of energy from high frequency to low frequency due to turbulent decay that removes energy from large scales to small scales and in this case, from low frequencies to high frequencies in the flume. Figure 8a is the PSD plot for data taken at a height of 3.5 inches near the flume grating. Figure 8b is the same type of plot for data taken at the same height, but farther from the entry grating. A peak is discernable at about 2 Hz in Figure 8a, whereas in figure 8b the plot drops significantly in that range. In general, the energy densities will be higher near the entry grating, because energy has just been added to the flow via the mixing blade that moves the water through the flume. Because of viscous effects at the boundary layer and within the

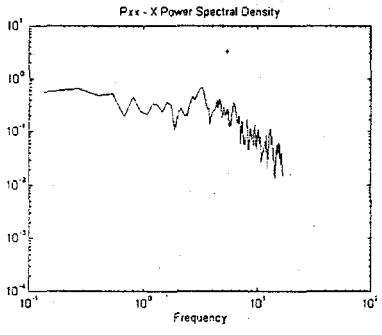


Figure 8a (near)

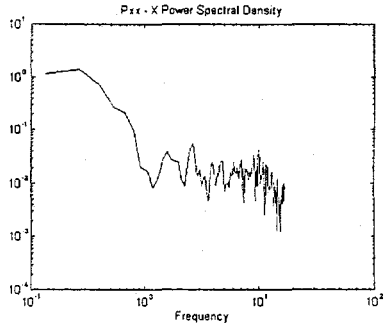


Figure 8b (far)

fluid, a control mass will lose flow energy to heat as it travels from the entry grating to the exit grating.

A purely viscous energy loss would be evidenced by a PSD plot showing uniformly lower values at all frequencies. However, the PSD at higher frequencies (e.g. 1×10^{1} Hz) drops significantly less than the PSD at lower frequencies (e.g. 2×10^{0} Hz). This indicates that energy has been transferred from the middle to the higher frequency ranges, replacing power dissipated in the high range by viscous effects. This is because of the decay, or cascade, of turbulence from the middle to high frequency ranges. It is an anticipated result, indicating successful application of the LDV to analysis of turbulent flow.

CONCLUSION

The LDV worked well in the field, giving a result consistent with the theory of isotropic turbulence. However, it was very dependent on sufficient suspended sediment in the waters where it was being tested. It was possible to achieve sufficient seeding to generate a strong signal in the laboratory by artificial means. Combined with the LDV's success in resolving very slow flows, this suggests that if the problem of signal strength can be solved, it would be possible to use the LDV to measure very slow flows in the presence of viable organisms. For example, the LDV might be used to help quantify the volumetric rate of flow of a filter-feeding organism by measuring the flow velocity of the exhaust stream. In addition, success in the area of turbulence measurement can provide data for use in calculations such as the one outlined by T.F. Gross *et al.*[5], which allows the shear stress at the bottom ocean boundary to be determined from the velocity power spectrum of the turbulence. This type of calculation would allow correlations to be made between shear stress at the bottom boundary layer and mixing in a marine environment.

With proper repairs and adjustments to the beam, sensor and fiberoptic connector alignments, the problems of beam strength and signal strength should be solved. This will have the effect of making the LDV much less dependent on having a high concentration of suspended solids in the water where it is being used. However, the consistently high velocities represent a discrepancy that merits further investigation.

ACKNOWLEDGEMENTS

The authors wish to express appreciation to Charles Pottsmith, whose advice and long-distance trouble-shooting were an invaluable assistance in testing the LDV, and to Philip Luft, who contributed both proofreading skills and word-processing resources towards the completion of this paper.

NOTES

1. Quest Radiance-850TM Two-Axis Laser Doppler Velocimeter Operator's Manual; Quest Integrated, Inc.; April 1993; Appendix I. Laser Safety.

2. Agrawal and Aubrey, "Velocity Observations Above a Rippled Bed Using Laser Doppler Veolcimetry," Journal of Geophysical Research, 97: 20,249 - 20,259, 1992.

3. Agrawal *et al.*, "Enhanced dissipation of kinetic energy beneath surface waves," Nature, 359:219-220, 1992.

4. Sanford, Turbulence and Mixing in Ecosystem Studies, draft, November 1993, pg. 5-7.

5. Gross *et al.*, "Bottom boundary layer spectral dissipation estimates in the presence of wave motions," Continental Shelf Research, 14(10/11):1239-1256, 1994.

In Vitro Flow Experiments with the Union College Vortex Blood Pump

Charles B. Howarth

Department of Mechanical Engineering
Union College
Schenectady, New York 12308

Faculty Advisor: J.R. Shanebrook

ABSTRACT

The Union College Vortex Blood Pump was designed to utilize a pusher-plate and hemofoil inlet valve to establish an internal vortex that simulates the natural and intrinsic self-cleansing of the human heart in order to prevent thromboembolism. The blood pump test system consists of an anthropometrically correct pumping chamber, mock circulatory loop and mechanical drive system. The circulatory loop is used to simulate the resistive and capacitive nature of the human circulatory system while the mechanical drive system can reproduce the everyday demands that are placed on the human heart by varying the stroke volume, pump frequency, and the systolic/diastolic ratio. The objective of this project is to provide experimental evidence as to the effectiveness of the hemofoil inlet valve in establishing a favorable vortical flow pattern. Results to be presented include flow visualization with pliolite tracer particles in water to simulate the flow of blood inside the pump chamber and pressure data to provide evidence of biologically correct test conditions.

INTRODUCTION

As of 1991, the U.S. population was 252 million and one in five suffered from cardiovascular disease. In that same year, 42.7% of all deaths were due to cardiovascular disease [1]. Although heart transplants are becoming increasingly more successful, about 85% of those people waiting for a donor heart die before a heart is available [2]. For this reason many researchers have been working on creating a mechanical device to replace the function of the failing heart.

Although several devices exist that conceptually perform the work of the heart, recipients of these devices often suffer from complications such as pulmonary embolisms and cerebral vascular incidents. Surprisingly, many of these devices perform well in in-vivo testing with calves presumably because of the rete mirabile which is present in the anatomy of the calf and possibly filters blood clots preventing cerebral vascular incidents [3].

Etiology of artificial heart recipients has determined the origin of thrombus formations to be within the pumping chambers of the artificial hearts and on the inlet and outlet valves [4]. Researchers believe that stagnant regions inside the pump are the cause of thrombus formation [5]. When flow rates are increased to compensate for the clotting problem, hemolysis can be the result. On the other hand, large doses of anti-coagulants, such as heparin, reduce the clotting problems significantly but severe internal bleeding may result. The purpose of this paper is to provide evidence as to how a hemofoil inlet valve, when used with a pusher-plate blood pump, can induce a vortical flow pattern within the pump that provides continuous fluid cleansing action over the internal pump surfaces.

METHOD

A pusher-plate blood pump has been designed to provide the self-cleansing flow pattern intrinsic to the human heart. An airfoil-shaped inlet valve, known as a hemofoil valve, sheds a starting vortex which enters into the cylindrical, Plexiglas pump housing and exits through a 29 mm Bjork-Shiley tilting disk valve placed laterally to the inlet valve. The pump has a chamber diameter of 97 mm, a thickness of 32 mm and a maximum stroke volume of 70 mL.

The pump is driven by a motor-cam apparatus, incorporating a four-bar mechanism, which allows for various systolic / diastolic ratios and stroke volumes. A mock circulatory system, shown in Figure 1, contains a flow capacitor to simulate the resilience of the human arteries and a flow resistor, which is a specially designed hose clamp, that can be tightened to provide

for different systemic pressures at the pump outlet. The flow capacitor is a sealed column of water and air located downstream of the pump. As the pressure increases at the outlet of the pump, the air is compressed simulating the expansion of the human arteries. As the pressure at the outlet decreases, the capacitor releases its stored energy analogous to when the human arteries relax during diastole. Pressure taps, located at the inlet and outlet of the pump, are attached to a Schaevitz pressure transducer which outputs a voltage to an oscilloscope. By looking at the pressure trace, the resistance clamp can be adjusted to achieve physiological pressure conditions.

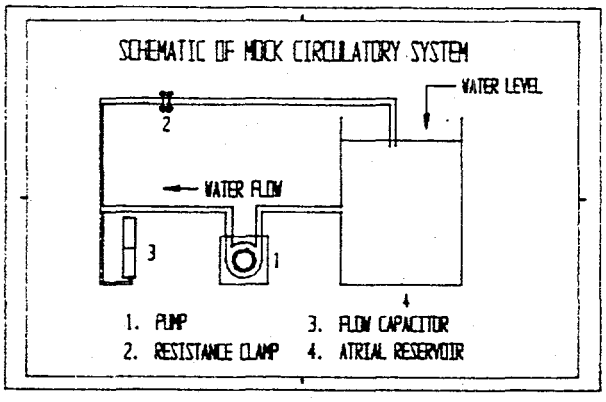


Figure 1: Schematic of mock circulatory system used with the Union College Vortex Blood Pump.

The design of the mechanical drive system and mock circulatory system allow for several variables to be changed in order to evaluate the performance of the pump under different physiological conditions. The variables include cam type, inlet valve, volumetric flow rate, resistance and capacitance.

Pressure measurements are necessary at the inlet and the outlet of the pump in order to assure biologically correct test conditions. Schaevitz pressure transducers are interfaced with the system via pressure taps located at both the inlet and the outlet to the pump. The pressure transducers are connected to an oscilloscope and the wave forms are translated into two coordinate points of voltage and time using a computer. The voltages are converted to the corresponding pressures, using the calibration curves that were developed for each pressure transducer, and a plot of pressure vs. time can be constructed.

Figure 2 shows the pressure traces for the human aorta, left atrium, and left ventricle. The atrial and aortic pressure traces are analogous to the inlet and outlet pressure traces for the

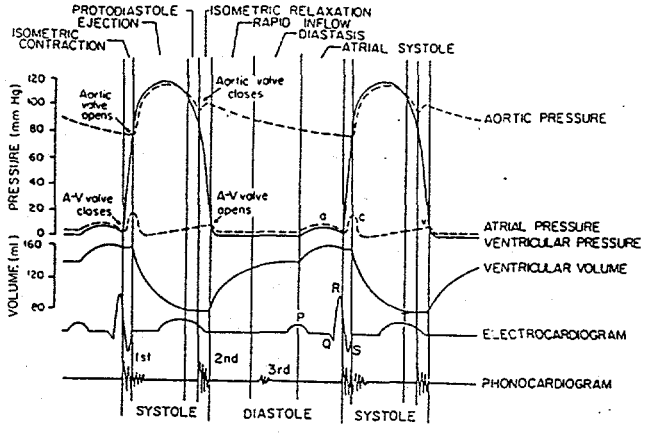


Figure 2: Events in the human cardiac cycle. The pressure traces shown are for the left atrium, ventricle and the aorta [6].

Union College Vortex Blood Pump since the pressure taps in the mock circulatory system are located at the outlet and inlet to the pump rather than within the pump chamber. In order to get a pressure trace analogous to the ventricular pressure trace, a pressure tap would have to be placed within the pump chamber itself which would disturb the flow pattern within the pump chamber. The aortic pressure (analogous to the outlet pressure from the pump) ranges from approximately 80 to 120 mmHg in the human body. As can be seen in Figure 3, the maximum outlet pressure from the Union College Vortex Blood Pump has been adjusted to be approximately 120 mmHg by adjusting the flow resistor. The test conditions for figures 3 and 4 are a pulse of 70 beats/minute and a systolic/diastolic ratio of 0.72 with the flow capacitor half full.

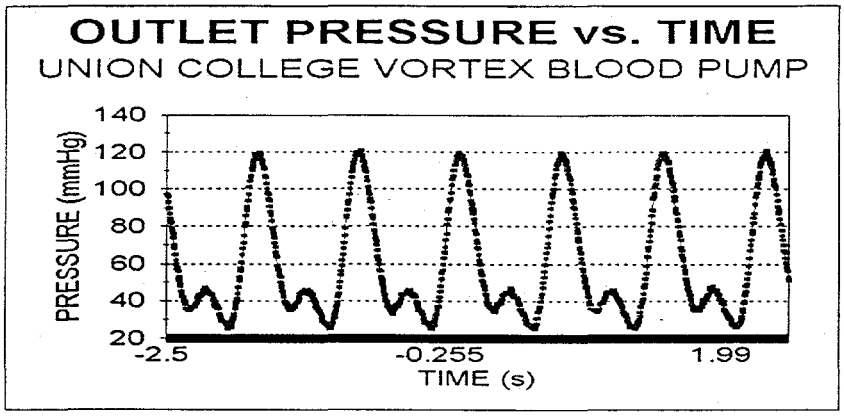


Figure 3: The outlet pressure trace for the Union College Vortex Blood Pump.

According to Figure 2, the atrial pressure in the human body varies between 0 and approximately 15 mmHg. The atrial pressure is analogous to the inlet pressure to the pump. In Figure 2, the initial rolling peak in the pressure trace is due to the slight contraction of the atrium. The spike in pressure is due to the delay in closure of the human mitral valve when the ventricle begins to contract [7]. Figure 4 shows the inlet pressure trace for the Union College Vortex Blood Pump. The minimum and maximum values for the pressure at the inlet are biologically correct.

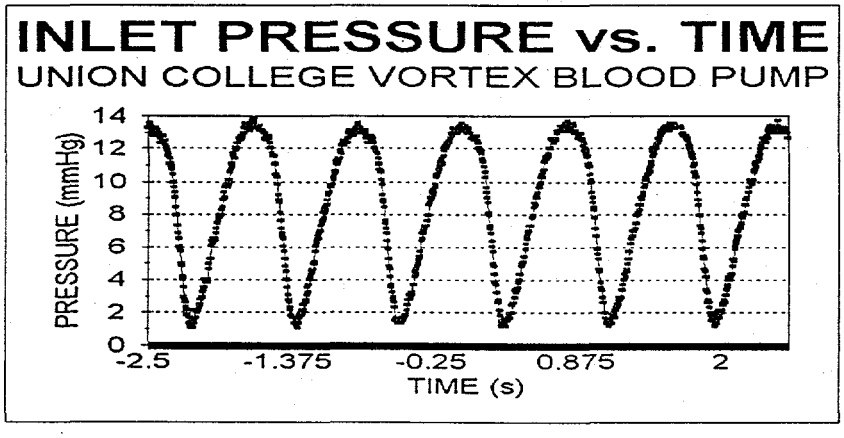


Figure 4: Inlet pressure trace for the Union College Vortex Blood Pump.

Flow visualization has been achieved using pliolite particles, manufactured by Goodyear Chemicals, as tracer particles and water as the test fluid. A fiber-optic lamp and slit lighting technique have been used to illuminate the flow plane. Results presented in the following

figures are with the system configured at a pumping frequency of 60 beats/minute, stroke length of 2.8 cm, theoretical stroke volume of 66.4 cm^3, and a systolic/diastolic ratio of 0.72.

RESULTS

Figures 5 through 7 show the presence of a strong vortex during both systole and diastole. Figure 5 shows the Union College Vortex Blood Pump at end-diastole. (End-diastole is defined to be when the pusher plate is all the way at the end of its stroke, out of the pump.) A strong vortex is shed from the trailing edge of the hemofoil inlet valve. The persistence of the vortical motion is demonstrated in Figure 6 where the pump is shown to be at beginning-systole. (Beginning-systole is defined to be when the pusher plate is beginning to moving into the pumping chamber.) Figure 7 shows the Union College Vortex Blood Pump at mid-systole, where the pusher plate is about half-way through its pumping stroke. By examining these photographs it can be seen that the vortex continues to grow as it moves from the trailing edge of the hemofoil inlet valve, cleansing the internal surfaces of the pump chamber, until it finally leaves through the outlet valve.

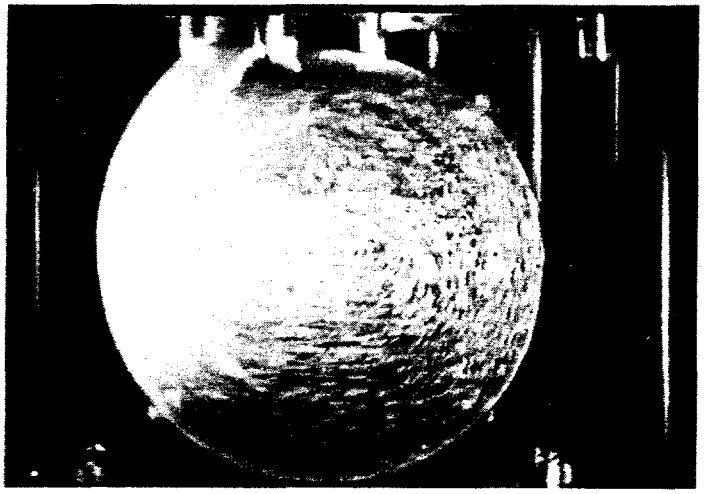


Figure 5 Union College Vortex Blood Pump at End-Diastole.

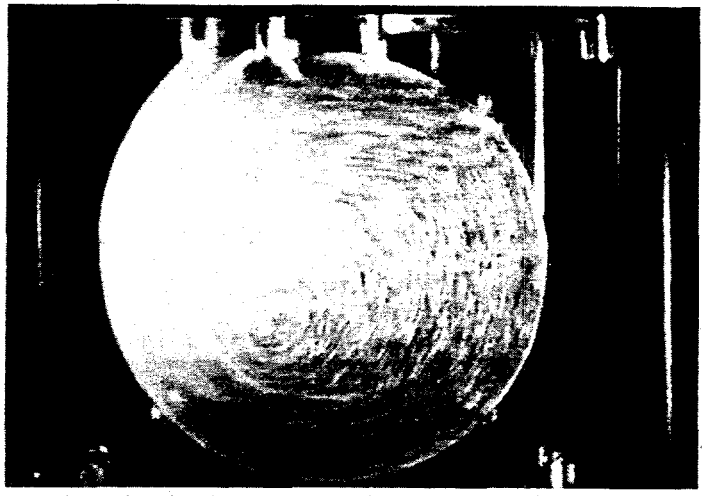


Figure 6 Union College Vortex Blood Pump at Beginning-Systole.

Due to the forward motion of the pusher plate, the chamber vortex is forced to expand towards the outer walls. This helps to increase the flow velocities in this region. It is evident

by examining the flow pattern illustrated by figures 5 through 7 that a thorough cleansing has been accomplished without the formation of stagnant regions.

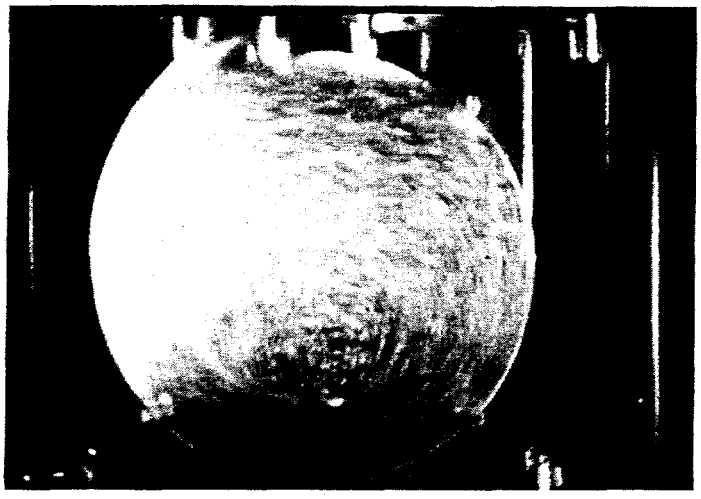


Figure 7 Union College Vortex Blood Pump at Mid-Systole.

CONCLUSIONS

The Union College Vortex Blood Pump was designed to utilize a hemofoil inlet valve in conjunction with a pusher plate and circular pumping chamber to create a vortical flow pattern in order to eliminate stagnant regions within the pumping chamber. The flow visualization results indicate the presence of relatively high velocities throughout the chamber. Future work will include flow visualization results for different system configurations.

REFERENCES

[1] Division of Research Administration, American Heart Association. "Research Facts - Update 1994."

[2] O'Connor, L. "Engineering a Replacement for the Human Heart," Mechanical Engineering, July, 1991, Vol. 113, No. 7, pp. 36-43.

[3] Ward, R.A., Wellhausen, S.R., Dobbins, J.J., Johnson, G.S., and De Vries, W.C. "Thromboembolic and Infectious Complications of Total Artificial Heart Implantation," Blood in Contact with Natural and Artificial Surfaces, Vol. 516 of the Annals of the New York Academy of Sciences, December 28, 1987, pp. 638-640.

[4] Lelkes, P.I. and Samet, M.M. "Endothelialization of the Luminal Sac in Artificial Cardiac Prosthesis: A Challenge for Both Biologists and Engineers," Journal of Biomechanical Engineering, Vol.113, May, 1991, pp. 132-134.

[5] Pennington, D.G., Swartz, M.T., McBride, L.R., Reedy, J.E., and Miller, L.W. "Bridging to Cardiac Transplantation with Circulatory Assist Devices," Artificial Heart 3, (Akutsu and Koyanagi), Springer-Verlag, 1990, pp. 293-301.

[6] Guyton, A.C. Textbook of Medical Physiology. Philadelphia: W.B. Sanders Co. 1966, Appendix.

[7] Berne, Robert M., M.D. Cardiovascular Physiology. C.U. Molly &Co. St. Louis, 1972, p. 76.

Wind Tunnel Investigation of Steady Flow Through Coronary Artery Bypass Grafts

Tricia L. Nelson

Department of Mechanical Engineering
Union College
Schenectady, NY 12308

Faculty Advisor:

J. R. Shanebrook

A common treatment for coronary artery disease is bypass graft surgery. Unfortunately, the long term patency of coronary artery grafts is not significant due to blood clotting caused by adverse flow patterns found at and near graft junctions. A wind tunnel model was designed to model the steady flow of blood in the distal, or downstream of the arterial occlusion, section of the coronary artery bypass graft. Qualitative knowledge was gained specific to the fluid dynamics of flow in the graft. Based on this knowledge, experiments were conducted to improve the fluid dynamics of blood flow in bypass grafts. Specifically, testing with a prosthetic half-delta wing inserted in the graft has been performed so as to alter the existing non-ideal flow patterns and obtain a more desirable flow. The ultimate goal is to improve the patency of bypass grafts as a form of treatment for coronary artery disease.

BACKGROUND

Cardiovascular disease, such as coronary artery disease, accounts for nearly 50% of all deaths in the United States each year (American Heart Association, 1993). Due to this statistic, it is essential that successful forms of treatment be developed. Procedures such as balloon and laser angioplasty, directional atherectomy, and coronary artery stenting are common treatments alternative to bypass graft surgery. Unfortunately, with these procedures, the rate of restenosis,

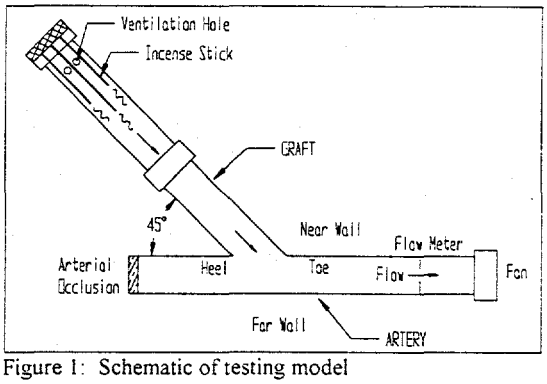


Figure 1: Schematic of testing model

or re-clogging of the artery, remains between 25 and 45 percent (Block, 1991).

Coronary artery bypass surgery, performed approximately 400,000 times each year in the U.S., is the common, last resort procedure used to treat severe cases of coronary artery disease (Ubell, 1994). However, 26% of all coronary artery grafts fail within one year, post-operatively, due to blood clotting, or thrombus formation, in and around the graft junction (Rittgers, 1991). Illustrated in figure 1, a Plexiglas™ cylindrical wind tunnel model (31.75 mm inner diameter, 6.35 mm wall thickness) has been constructed to simulate steady coronary artery blood flow through the distal graft junction. This portion of the graft was chosen for modeling because it has been shown by past research to be the primary site of graft failure due to thrombus formation.

The flow patterns associated with the illustrated distal graft model have been investigated through flow visualization techniques. This method has provided understanding and explanation of the adverse flow patterns, such as stagnation, flow separation, and multiple

vortices, which promote blood clotting at the graft junction. Blood clotting is prone to occur in regions experiencing these flow patterns due to the slower flow of blood. Figure 2 shows the typical areas in which thrombus formation tends to occur. In these regions of slower flow, the blood tends to aggregate and form clots which adhere to the wall area and grow or flow downstream and become lodged somewhere else.

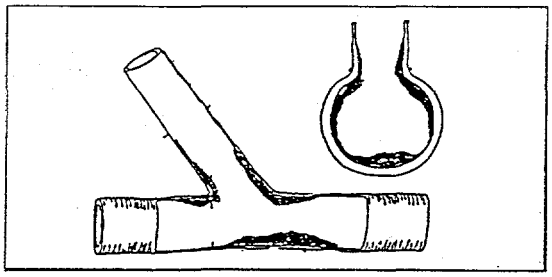


Figure 2: Areas at distal end of graft prone to thrombus formation (Sotturai, 1989)

EXPERIMENTATION

Fluid flow through the distal model of figure 1 consisted of air and smoke. Three to five incense sticks placed axially in the graft section were burned. The resulting smoke was drawn from the graft into the artery section by the flow of room air. This flow was created by a fan which drew the room air through the tunnel at an approximate velocity of 6.8 cm/s (Reynolds number = 140). The thin end plate located upstream of the graft junction represented a 100% occlusion in the artery. Flow visualization results of the flow patterns found in the distal model were obtained through the illumination and video recording of various cross-sectional views of the flow.

Based upon studies of coronary steady flow modeling performed by Nerem (1983) and

Figure 3: Top view of coronary flow through distal graft model (Re=140)

Mark (1982), a steady flow Reynolds number of 140 was chosen for this experiment. Figure 3 is a top view photograph of coronary artery flow through the wind tunnel. The resulting flow patterns indicate that stagnation occurs at the far wall of the artery and flow separation occurs at the toe of the junction. A slow moving recirculatory zone is also discovered near the heel of the junction. Figure 4 shows an axial view of flow in the model at approximately 5 cm downstream of the junction. The double vortex flow pattern formed is the result of the flow stagnation and separation that occurs at the junction. This multiple vortex flow is not desirable because the vortices that form impact one another and create further stagnate regions along the vessel walls, thus promoting thrombus formation downstream.

PROPOSED SOLUTION: THE HALF-DELTA WING

From an engineering standpoint, an elimination of the adverse flow patterns that develop at

Figure 4: Axial view of coronary flow through distal graft model (Re=140)

graft junctions must be accomplished in order to improve the patency of bypass grafts as a form of treatment for coronary artery disease. In addition to the elimination or decrease of the stagnation and separation zones at the junction, it is desirable to eliminate the axial double vortex pattern and produce a single vortex that will evenly cleanse the entire axial wall. The current proposed method of achieving a single axial vortex flow was introduced by Professor J. R. Shanebrook of Union College. It involves the use of a half delta wing, as a prosthetic device, inserted at a certain angle and position within the graft as shown in figure 5. As fluid flows over a delta wing as shown, a vortex will spin off the trailing edge. The hope is to develop a wing at a certain angle and position within the graft such that the vortex formed will disturb the existing non-ideal flow patterns and create a more desirable flow.

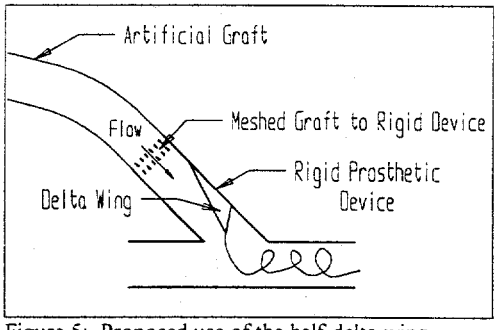


Figure 5: Proposed use of the half-delta wing

There are four variables to consider in the testing of fluid flow with the half delta wing. The first variable is simply the wing--its size, shape, and material. The wing should be as small as possible and constructed of a very smooth material. Both of these factors are important in minimizing the amount of resistance imposed on the flow due to the wing. The second variable to consider is the wing's axial location in the graft. Figure 6 illustrates an axial view of the four slots, each 3.25 inches in length, which have been created along the graft section of the wind tunnel. These permit the insertion of the wing and the testing of its effects at various axial locations. The third variable, which also involves the slots created, is the wing's circumferential position within the graft. The four different slots, described by degrees, allow the testing of different angular positions. Figure 6 illustrates a wing in the 0° slot. The wing is held within the graft by a rod, which is enclosed by a plexiglas housing block that fits in the slots. This holding device allows control over the final variable which is the wing's angle of attack. The rod is inserted through a small bearing, which is located in the plexiglas housing. The bearing allows the wing to be rotated within the graft. Thus, the angle of attack of the wing can be varied. The measurement of this angle was performed with a calibrated potentiometer connected to the end of the rod.

These four variables of wing type, axial location, angular location, and angle of attack allow

a great amount of freedom for testing. Although this provides great promise in the use of a delta wing as a possible solution, the many variables involved create almost an infinite amount of testing possibilities.

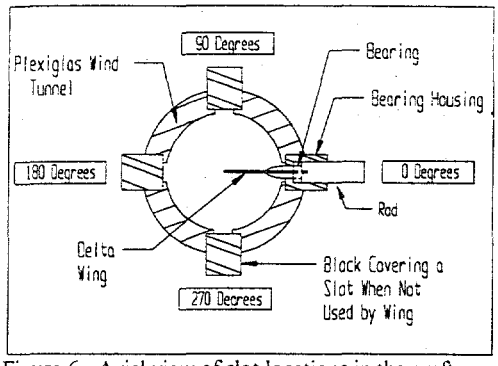


Figure 6: Axial view of slot locations in the graft

FLOW VISUALIZATION RESULTS

Figure 7 illustrates the geometry of the wing used to produce the following documented results. The wing was located in the 90° slot, at an angle of attack of 10° counterclockwise (with point of reference looking down upon this slot), and the rod located 8.5 cm upstream of the graft junction. A top view of the resulting flow at a Reynolds number of 140 is shown in figure 8. This view does not reveal any significant improvements in the adverse flow patterns mentioned. With the use of the wing, separation at the toe appears somewhat diminished but the recirculatory zone at the heel appears larger.

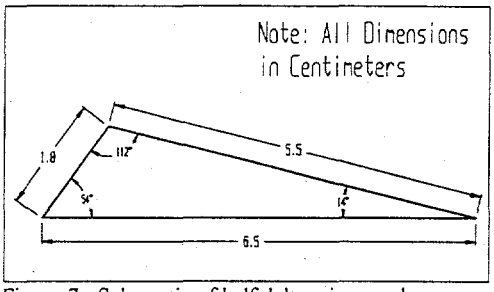


Figure 7: Schematic of half delta wing used.

The more significant result obtained with the wing is observed in the axial view, approximately 5 cm downstream of the graft junction. As stated in reference to figure 4, it is desired to eliminate the double vortex pattern occurring in the axial view and create a single vortex that would evenly cleanse the arterial wall. Figure 9 shows a photograph of the axial flow pattern resulting from the use of a wing. Although a clear, single vortex has not been achieved, a definite dominate vortex occupies a majority of the arterial cross section, thus cleansing a larger portion of the artery wall and reducing the probability of thrombus formation.

CONCLUSION

Thrombus formation is a leading cause of the large percentage of graft failures. In particular, the distal junction of the bypass graft appears to be the primary site for problems due to the adverse flow patterns formed as a result of graft geometry. As observed from a top view of the model, thrombus formation is prone to occur at the stagnate region of the far wall, the slow moving recirculatory zone at the heel, and the separation region at the toe of the junction. The axial view reveals other stagnate regions of flow found along the arterial wall downstream of the graft junction.

The use of the half delta wing is a promising solution to the problem of graft failure. Although the adverse flow patterns observed in the top view of the distal model were not eliminated, the large reduction of the double vortex in the axial view is a significant result. The

dominate vortex, resulting from the use of a single wing, evenly cleanses a larger area of the arterial wall and thus reduces the probability of blood clotting downstream of the graft junction. It is anticipated that further experimentation with various wings will produce more significant results that may lead the way to improved bypass graft patency.

Figure 8: Top view of flow in distal model as a result of one wing

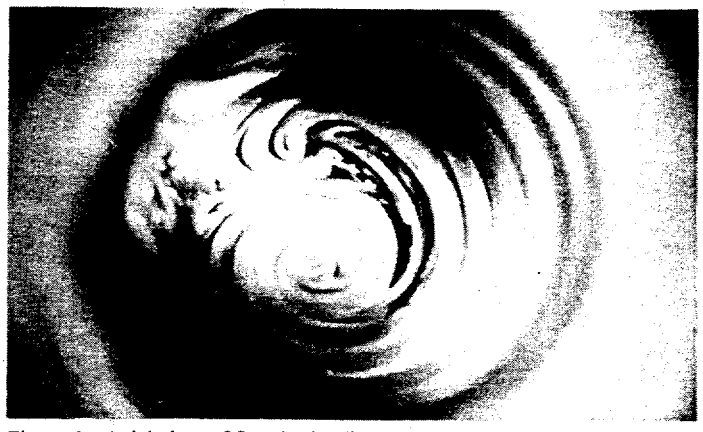


Figure 9: Axial view of flow in the distal model as a result of one wing

REFERENCES

American Heart Association, "Research Facts - Update 1994." 1993. pg. 1.

Block, Peter C., "Coronary-Artery Stents and Other Endoluminal Devices." The New England Journal of Medicine, Vol 324, Jan. 3, 1991, pg. 52.

Mark, F.F., et al., "Quasisteadiness of Flow in a Human Coronary Artery Model." Mechanics of the Coronary Circulation, The American Society of Mechanical Engineers, New York, 1982, pg. 71.

Nerem, R.M., et al., "Coronary Artery Geometry and its Fluid Mechanical Implications." Fluid Dynamics as a Localizing Factor for Atherosclerosis, Springer-Verlag Berlin Heidelberg, 1983, pg. 56.

Rittgers, S.E., et al., "The Effect of Angle and Flow Rate Upon Hemodynamics in Distal Vascular Graft Anastomoses: An In Vitro Model Study." Transactions of the ASME, Vol 113, Nov. 1991, pg. 458.

Sottiurai, V.S., et al., "Distal Anastomotic Intimal Hyperplasia: Histopathologic Character and Biogenesis." Annals of Vascular Surgery, Vol 1, 1989, pg. 27.

Ubell, Earl, "How Necessary is Heart Surgery?" Sunday Times Union, PARADE, March 13, 1994, pg. 4.

Cake-Cutting, Hugo Steinhaus Style: Beyond $N=3$

Carolyn E. Custer

Department of Mathematics
Union College
Schenectady, NY 12308

Faculty Advisor:

Alan D. Taylor

INTRODUCTION

A fair division procedure is a scheme that allocates portions of a whole to n people in as equitable a manner as possible. The problem of fair division made its first recorded appearance as early as the Old Testament. King Solomon was presented with two women claiming motherhood of the same baby. King Solomon offered the solution of cutting the child into two halves and letting each woman take a half. One woman immediately offered to give up her claim on the baby and honor the second woman's claim without dispute, thus proving the validity of her original claim to the King. Although this was a trick and King Solomon never intended to actually dismember the baby, this was in essence a fair-division solution. Similar examples have been recorded since this one, but not until the 1940s, during World War II, did fair division become a topic of modern research. Out of the 1940s come such methods as Lone-Divider, Last-Diminisher, the Moving-Knife, and others designed to resolve issues of fair division.

Fair-division procedures allow for the equitable distribution of continuous (divisible) and discrete (indivisible) goods to the satisfaction of all n people (or players, as they are often described in a game-type scenario) involved. The properties of proportionality and envy-freeness are goals in creating these procedures. Proportionality is the condition met when for n players, the whole is divided into n equal portions of size $1/n$. Envy-freeness is the condition satisfied when every player involved believes that he or she has received at least as valuable a portion as all the other portions (according to his or her own personal criteria) and thus does not begrudge anyone else his or her portion. Proportionality is much easier to obtain than envy-freeness, and in this paper we will deal with fair allocation (proportional schemes) only, except in the case of the original Divide-and-Choose procedure to be described momentarily.

The usefulness of these theories and schemes can be found today in problems involving estates, divorce settlements, economic issues, and national and international politics, but for the sake of explanation, the widely accepted cake-cutting illustration will be used throughout.

The oldest method of solving the fair division problem is the Divide-and-Choose method. This applies when the object in question, or group of objects, must be divided between two players ($n=2$). In the example of the cake, one person is randomly chosen (perhaps by flipping a coin) and cuts the cake into two pieces. This person is the Divider. The other player is then the Chooser, who is allowed to pick which portion he/she believes to be the more valuable and take that. The Divider takes the piece left. This is the procedure. The strategy, however, is the whole point of the game. It is the strategy that allows you to end up with a satisfactory portion, even if all other players in the game are plotting against you. In order for both players to be satisfied with the outcome, they must each believe that both pieces are of size 1/2 (proportionality) or each must think he/she has received the more valuable of the two. The Divider, knowing he or she will be left with the second piece of cake, finds it in his or her best interest to cut the cake such that both pieces are equivalent in value in his/her perception. Then, the Chooser will take the piece that seems most valuable to him/her. Proportionality is

upheld as both players believe they have received at least half of the cake, and since neither believes the other could have received a more valuable piece, the procedure is also envy-free. We must realize that both players will be operating with differing value systems, so that each player may regard a piece differently than the other player. It is intuitively easier to speak in terms of size, so for the sake of simplicity, we use size as the value system in our models. For example, the larger a piece of cake is perceived to be, the more valuable it is in the eye of the beholder.

This field of study was first brought into modern academe by Hugo Steinhaus, while in Poland during the second World War, through his studies into the extension of the Two Person (n=2) Divide-and-Choose method to the Three Person (n=3) Lone-Divider method. Steinhaus was the mathematician who first began working on fair-division procedures. His research, in conjunction with colleagues Stefan Banach and Bronislaw Knaster, began the modern study of fair division procedures as we know it. Already having the age-old custom of Divide-and-Choose method for two players, Steinhaus developed a similar, and proportional, solution for three players that has become the Lone-Divider method as described in Kuhn (1967), and Brams and Taylor (1994) from which the following explanation of Steinhaus' method is paraphrased.

> Imagine we have three players in the game of dividing a cake. We have named them Bob, Carol, and Ted. Bob is chosen randomly as the Divider and proceeds to cut the cake into three pieces; each, in his eyes, making up 1/3 of the original cake. In order to determine the order of choosing such that the method is fair and everyone is satisfied with the final distribution, Carol and Ted should each determine which pieces are acceptable (at least 1/3 of the whole in her or his estimation). We invariably assume that every player will always find at least one piece acceptable.
>
> Case 1: At least one of the players (let's assume this is Carol) approves of at least two of the three pieces. Ted, finding at least one piece acceptable, would choose his piece first, then Carol has at least one of the pieces she liked still available and she takes one. Bob then accepts the last piece as he approved of all three pieces at the start. This is a proportional fair division.
>
> Case 2: Otherwise, both Carol and Ted approve of at most one of the pieces. In this case, we know that there exists at least one piece that only Bob finds acceptable. This piece is then his. The other two pieces are reassembled into one, and Carol and Ted redivide as per divide-and-choose. Since neither Carol nor Ted approves of the piece Bob has, they believe they have more than 2/3 of the cake between them. Simple arithmetic illustrates the idea that if Carol and Ted each take half of what they view as at least two-thirds of the whole, then the result is at least one-third of the whole.
>
> $$(1/2) \times (2/3) = 1/3$$

In the search for a general solution, Steinhaus offered his n=3 Lone-Divider procedure to his colleagues Stefan Banach and Bronislaw Knaster, asking them to extend his scheme to a Lone-Divider for four players, n=4, and perhaps eventually to a general Lone-Divider for any n. Instead, Banach and Knaster branched out on their own and devised the Last-Diminisher procedure that resolves the problem in the general case of n people. That which follows is a brief sketch of this Last-Diminisher scheme for any number of players or partners(n=3) from Steinhaus (1948).

> The partners being ranged A, B, C, ... , N, A cuts from the cake an arbitrary part. B now has the right, but is not obliged, to diminish the slice cut off. Whatever he does, C has the right (without obligation) to diminish still the already

> diminished (or not diminished) slice, and so on up to N. The rule obliges the "last diminisher" to take as his part the slice he was the last to touch. This partner being thus disposed of, the remaining n-1 persons start the same game with the remainder of the cake. After the number of participants has been reduced to two, they apply the classical rule for halving the remainder.

In the years following these developments, more schemes appeared, including the "Moving-Knife" version of Last-Diminisher that first appeared in 1961. Due to Lester E. Dubins and Edwin H. Spanier, it was later modified by A. K. Austin (1982) for another Moving-Knife analogue to the Divide-and-Choose. Harold W. Kuhn, author of "On Games of Fair Division" (1967), also began with Steinhaus' accomplishment as a foundation for his general theory and also veered off in another slightly different direction with the final product. Kuhn's final product is strongly based on an old combinatorial theorem of Konig.

Since Hugo Steinhaus' n=3 scheme was never actually used directly as a foundation for extension, Steven J. Brams and Alan D. Taylor decided to take that approach and work out a similar Lone-Divider scheme for n=4. This procedure appears in the book Fair Division: From Cake-Cutting to Dispute Resolution, by Brams and Taylor (1995), and occupies two and a quarter pages of text.

We improve on this in what follows by presenting a version of the n=4 case that requires only half a page of text (Section 3 - N=4 Lone-Divider Scheme). This is extended to n=5 in Custer (1994). We also down-size these results to obtain another version for n=3 (Section 2 - N=3 Lone-Divider Scheme). Our n=3 version is neither shorter nor simpler than Steinhaus', but our approach yields an n=4 version that is not much more complicated than the n=3 scheme. Although the actual sequence of events began with the development of the n=5 scheme, then the subsequent down scaling to the n=4 and n=3 schemes, the most logical presentation is to build from the bottom up beginning with n=3 instead. The following two sections include the most recent refinements of the Lone-Divider procedure.

N=3 LONE-DIVIDER SCHEME

For the sake of clarity, several assumptions must be made regarding these schemes and the rules of the "game." First, we regard the cake as a symbol of whatever goods we might be dividing and assume that any shavings or pieces from the cake can be reassembled and the cake re-cut as if new. Second, we seek procedures wherein, if there are n players in the game, the cake is originally cut into n pieces. Third, we assume that, given such an initial partition, every player involved in the game must approve of at least one of the pieces of cake, or there would be no point in the player's participation. When a player approves of a piece of cake, that means the player values the piece at no less than $1/n$ of the original cake. Rejection of a piece of cake requires that the player values it as less than $1/n$ of the original cake. Fourth, the fact that Bob is always the Divider in our scenarios is not a loss of generality, but again for simplicity and clarity just a name given to that role in the game. (In practice, the Divider might be randomly chosen at the outset.) Finally, since Bob is the Divider, we assume that he cuts the cake such that each piece is exactly $1/n$ in size in his estimation and that he approves equally of all n shares.

First we will explore the most recent Lone-Divider scheme n=3:

Say there are three players among whom a cake is to be divided; Bob, Carol, and Ted. Bob begins by cutting the cake into three pieces he considers to be equal in value: A, B, and C.

Picture now the three pieces of cake. To Bob, they are each exactly 1/3 of the original cake, so he approves of all three. To Carol and Ted, the three pieces together make up the whole cake, but each piece may not be exactly 1/3

in their respective estimations. The following cases cover all possible combinations of approvals limited only by the rules set forth above.

Case 1: There exists at least one piece of cake that is approved of by Bob alone or by Bob and one other player.

Case 1.1: There exists a piece of cake (say, piece 3) that Bob alone approves of.

Then Bob would take this piece and Carol and Ted would carry out the Divide-and-Choose scheme. Neither Carol nor Ted will resist Bob's receiving this piece because they rejected piece 3 believing it to be less than 1/3. Hence pieces 1 and 2 together must make up more than 2/3 in both Carol and Ted's estimations. These two pieces are reassembled, and Carol and Ted carry out the Divide-and-Choose method from Chapter 1. As stated in Chapter 1, half of more than 2/3 is more than 1/3, so Carol and Ted each feel they are receiving more than 1/3 of the original cake.

Case 1.2: There exists at least one piece (say, piece 2) that is approved of only by Bob and one other player (perhaps Carol).

In this case, Carol would take that piece while Bob and Ted Divide-and-Choose. Since Bob approves of all three, he sees pieces 1 and 3 as making up exactly 2/3 of the cake and is willing to take half of that. Since Ted rejected piece 2 as being less than 1/3, then he must see pieces 1 and 3 together as making up more than 2/3, and so is also willing to take half of that.

Case 2: Otherwise, each piece is approved of by all three players, and any order of choosing could be used.

N = 4 LONE-DIVIDER SCHEME

Our n=4 version of the Steinhaus Lone-Divider scheme is not much more complex than our n=3 version.

Say there are four players among whom a cake is to be divided; Bob, Carol, Ted, and Alice. Bob begins by cutting the cake into four pieces he considers to be equal in value: A, B, C, and D.

Picture now four pieces of cake. To Bob, they are each exactly 1/4 of the original cake, so he approves of all four of them. To Carol, Ted, and Alice, the four pieces together make up the whole cake, but each piece may not be exactly 1/4 in their respective estimations. The following cases cover all possible combinations of approvals limited only by the rules set forth in Section 2.

Case 1: There exists at least one piece of cake that is approved of by Bob alone or by Bob and one other player.

Case 1.1: There exists a piece of cake (say, piece 4) that Bob alone approves of.

Then Bob would take this piece and Carol, Ted, and Alice would carry out the n=3 version of Steinhaus' Lone-Divider scheme. Regardless of where the approvals of Carol, Ted and Alice fall over pieces 1, 2, and 3, we know that they rejected piece 4 believing it to be less than 1/4. Hence pieces 1, 2, and 3 together must make up more than 3/4 in the eyes of Carol, Ted, and Alice. These three pieces are reassembled, and Carol, Ted, and Alice carry out n=3 Lone-Divider method from Chapter 2. Similar to the arithmetic in Chapter 1, 1/3 of more than 3/4 is more than 1/4, so Carol, Ted, and Alice each feel they are receiving more than 1/4 of the original cake.

Case 1.2: There exists at least one piece (say, piece 2) that is approved by only by Bob and one other player (perhaps Carol).

In this case, Carol would take that piece while Bob, Ted, and Alice perform the n=3 version of Steinhaus' Lone-Divider procedure. Ted and Alice allow this because they both believed that piece 2 was less than 1/4 of the cake anyway. Now, Alice, Ted, and Bob reassemble pieces 1, 3, and 4 and follow the n=3 Lone-Divider procedure. Since Bob approves of all four, he sees pieces 1, 3, and 4 as making up exactly 3/4 of the cake and is willing to take 1/3 of that. Since Ted and Alice each rejected piece 2 as being less than 1/4, then they must see pieces 1, 3, and 4 together as making up more than 3/4, and so are also willing to take 1/3 of that.

Case 2: Each piece is approved of by Bob and at least two others.

Picture the four pieces of cake that are to be chosen. We know that Bob approves of all four, and that each piece must also have two approvals other than Bob or we are back in Case 1. Pictorially we can imagine four pieces of cake each with a B, for Bob, and two slots for other approvals under each of them.

(1)	(2)	(3)	(4)
B	B	B	B
C	C	C	T
T	A	T	A

Since there are four pieces, we know that there are eight approvals to fill and three players (Carol, Ted, and Alice) to fill them. This forces at least one player (let's say Carol) to approve of at least three pieces, and at least one other player, Ted, to approve of at least two. Otherwise, we'd have the three players approving of $2 + 2 + 2 = 6 < 8$ or $4 + 1 + 1 = 6 < 8$. Alice must still approve of at least one and she will choose her piece first. This leaves at least one piece that Ted approves of, and he chooses it. Carol must then be left with at least one of her approved pieces, and so takes hers. Bob then takes the last piece. In this manner, each player is satisfied that he or she receives a piece that he or she registered approval of and so is satisfied that he or she received at least 1/4 of the original cake.

CONCLUSION

This paper presents some work done following the tradition begun by Hugo Steinhaus, but not continued until now. There is much more yet to accomplish in this specific area of the field of Fair Division, however, and I see great potential in the hands and minds of others. I am confident that ahead, beyond any specific cases for six, seven, or more players, still lies the general case for any n following the same form and structure we have used in this paper.

REFERENCES

Austin, A. K. (1982). "Sharing a Cake." *Mathematical Gazette* 6, no. 437 (October): 212-215.

Brams, S. J., and A. D. Taylor: Fair Division: From Cake-Cutting to Dispute Resolution, forthcoming from Cambridge University Press (1995).

Custer, Carolyn E. (1994). "Cake-Cutting, Hugo Steinhaus Style: Beyond N=3." Union College, Department of Mathematics, Senior Thesis.

Dubins, L. E., and E. H. Spanier (1961). "How to Cut a Cake Fairly." *American Mathematical Monthly* 68: 1-17.

Kuhn, Harold (1967). "On Games of Fair Division." In Martin Shubik (ed.), *Essays in Mathematical Economics*. Princeton, NJ: Princeton University Press, 1967, pp. 29-37.

Steinhaus, Hugo (1948). "The Problem of Fair Division." *Econometrica* 16, no. 1 (January): 101-104.

COMPUTER MODELING OF HE II IN A MICROCHANNEL ARRAY

Francisco J. Diez-Garias
Atsushi Uchida

Department of Science and Mathematics
Parks College of Saint Louis University
Cahokia, Il 62206

Faculty Advisor: Dr. Charles E. Monfort III
Department of Math and Science
Parks College of Saint Louis University
E-mail: monfort@pxa.slu.edu

Helium II has a number of unique physical properties. Of greatest interest are its transport properties in a microchannel array. The NASA/University Joint Venture (JOVE) program is sponsoring a study to establish helium II's heat transfer characteristics. In this study, the equations for helium flowing through a microchannel array are investigated. The model was based on the Landau two-fluid model. Both the ideal and the Gorter-Mellink states have been included in a computer model written in Fortran and linked to a software package, HEPAK, that contains the thermophysical properties of Helium-4 from 0.8 to 1500K. An attempt is made in this model to explain the physical limitation of heat transport in He II, resulting from the existence of a critical velocity in the superfluid component, above which the He II begins to exhibit viscous flow and non ideal heat transport. This theoretical model will then be compared to the experimental data obtained at Marshall Space Flight Center in Huntsville, Alabama. Finally, similar tests will be conducted in a microgravity environment to complete the study.

Helium II has proven to offer great advantages especially in space as a coolant for space instruments which need temperatures down to 2K. Hofmann et al. showed that a circulating system could be built in space using only a thermomechanical pump. The system is greatly improved if we use the Helium II (He II), since it has a thermal conductivity of the order of 10^7 times greater than He I. The thermo-mechanical effect in He II, also known as the fountain effect, occurs when helium flows from one vessel to another through a narrow capillar or a microchannel. The temperature will decrease where the helium flows out, and it will rise in the other vessel. This is due to the fact that the flow going through the microchannel is mainly superfluid which does not transfer any heat. Therefore, the heat left on the vessel has to be distributed over less amount of helium with the consequent rise in temperature. The opposite effect occurs on the other vessel. One of the applications of this phenomena is as a phase separator, since a porous plug can separate vapor from superfluid helium. The recent successful operation of a porous plug He II phase separator in the space infrared telescope IRAS has increased the studies on the subject.

Landau was the first one to describe He II as a two-fluid model. The two-fluid model defines He II as composed of two fluids: normal fluid, which is described by having a velocity v_n, density ρ_n, viscosity η_n, and specific entropy S_n, and the superfluid, which has a velocity v_s, density ρ_s, no viscosity and no entropy. Notice that v is the total velocity and S is the total entropy.
The momentum density can be written as:

$$\rho v = \rho_n v_n + \rho_s v_s \tag{1}$$

A very interesting case is the counterflow, that is when the net mass flow is zero.

$$0 = \rho_n v_n + \rho_s v_s \tag{2}$$

The helium is static, although the superfluid and normal fluid can flow in opposite direction.

Two different flow regimes are possible when describing flow through a microchannel. These are the ideal and the Gorter-Mellink flow.

Ideal State

It has been proven experimentally that below a critical heat flux Q_c, the helium obeys a heat transport relationship that has a temperature gradient, ΔT, proportional to the heat flux, q,

$$q \propto \Delta T \tag{3}$$

Gorter-Mellink State

Above the critical heat flux Q_c, He II enters another regime where the temperature gradient is now proportional to a higher power of the heat flux,

$$q^3 \propto \Delta T \tag{4}$$

For the ideal state, the Landau's equation can be applied. Which are the Navier-Stokes equations modified for the two components of He II,

$$\rho_n \frac{\partial \vec{v}_n}{\partial t} + \rho_n (\vec{v}_n \cdot \vec{\nabla}) \vec{v}_n = \eta_n \nabla^2 \vec{v}_n - \frac{\rho_n}{\rho} \vec{\nabla} P - \rho_s S \vec{\nabla} T \tag{5}$$

$$\rho_n \frac{\partial \vec{v}_s}{\partial t} + \rho_s (\vec{v}_s \cdot \vec{\nabla}) \vec{v}_s = -\frac{\rho_s}{\rho} \vec{\nabla} P + \rho_s S \vec{\nabla} T \tag{6}$$

Assuming steady state conditions ($\partial / \partial t = 0$), that the flow is parallel ($\partial v / \partial x = 0$) and that ΔT is very small compared to T, the London equation can be obtained,

$$\nabla P = \rho S \nabla T \tag{7}$$

and from (5) and (6) an expression can be derived for the normal velocity,

$$q = \rho S T v_n \tag{8}$$

Also, using Darcy's law, $v_n = -\frac{k\Delta P}{\eta l}$ (9)

where $k = D^2 / 32$ for laminar flow in a circular microchannel, l the length and D the diameter of the microchannel, and equation (6), an expression can be derived for the pressure difference,

$$\Delta P = \frac{32\,\eta l q}{D^2 \rho S T} \tag{10}$$

and for the temperature difference, ΔT from (5) and (8),

$$\Delta T_L = \frac{32\eta l q}{(D\rho S)^2 T} \tag{11}$$

Gorter-Mellink State:

In the presence of superfluid turbulence the Navier-Stokes equations are no longer valid to define the dynamic state of the helium, since they do not include the local interaction between the quantized vortex lines and the normal fluid.

A correction factor F_{sn}, that will account for the friction forces needs to be added on the right hand side of the Navier-Stokes equations,

$$F_{SN} = \rho_s \rho_n A V_{ns}^{\ 3} \tag{12}$$

where A is the cross-sectional area of the microchannel and the relative velocity is

$$V_{ns} = \frac{q}{\rho_s S T} \tag{13}$$

Thus, ΔT can be expressed in terms of the heat flux as follow,

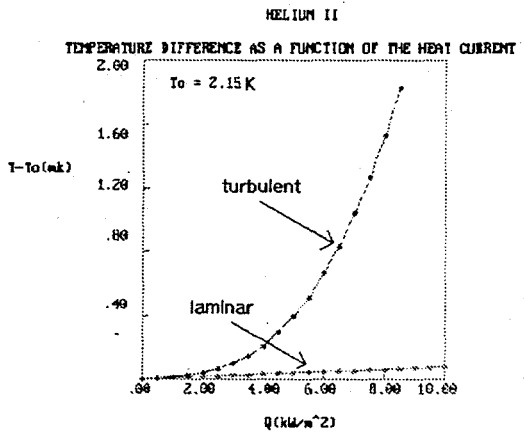
$$\Delta T = \Delta T_L + \frac{l F_{SN}}{\rho_S S} \tag{14}$$

and thus,

$$\Delta T = \frac{32\eta l q}{(D\rho S)^2 T} + \frac{\rho_n l A}{S}\left(\frac{q}{\rho_S S T}\right)^3 \tag{15}$$

A FORTRAN program was written and equations (10) to (15) were inserted into it. The output of the program will provide a graph of ∇T vs heat

flux for any bath temperature between 0.8 and 2.167 K, which is only for He II. The only inputs the user needs to provide are the diameter and length of the microchannel as well as the initial pressure. Then the program will call HEPAK, which is a FORTRAN program written by V. Arp and R. McCarty (see works referenced 'A new range equation of state for helium' by V. Arp and R. McCarty), that provides the thermodynamics properties of He II. Finally, it will plot the graphs both for the laminar and turbulent states. The figure below shows one of many different possibilities that this program offers to the designer of a microchannel, since the user can change the bath temperature as well as the length and diameter of the microchannel. The lower line describes the laminar flow, and the upper line shows the turbulent flow.



Heat flux vs ΔT for a bath temperature of 2.15K. The length of the channel is 1 mm and the diameter is $10*10^{-6}$ m.

Even though the program is in a very early stage, it appears to agree with some preliminary graphs we have from NASA/ Marshall Space Flight Center at Huntsville, Alabama. We expect the error to be less than 10 percent, but more specific data will be needed. The program should prove to be very helpful for those trying to design a microchannel.

In future plans, it will be expanded and hopefully compared with better data from NASA. They have been working on a new microchannel array configuration that if successful will provide very valuable data to this research.

WORKS REFERENCED

H. Nakai, M. Murakami, "Flow phenomena of superfluid helium through a porous plug phase separator." Cryogenics, Vol. 27, Aug., 1987.

Hofmann, "Critical velocity of superfluid helium flow in fine pore filters." Cryogenics, Vol. 30, March, 1990.

U. Schotte, "He II phase separator with slits and porous plugs for space cryogenics." Cryogenics, Oct., 1984.

A. J. Mord and H. A. Snyder, "Self-driven cooling loop for a large superconducting magnet in space." Cryogenics, Vol 32, No 2, 1992

Van Sciver, "Helium Cryogenics." Plenum Press, New York, 1986.

R. D. McCarty, V. D. Arp, "A new range equation of state for helium." Cryogenics, Vol . 35, 1990.

Evaluating Sums of Powers

Anjanette Finnegan

Department of Mathematics
University of St. Thomas
2115 Summit Avenue
St. Paul, MN 55105

Faculty Advisor:

Dr. Peter Yang

INTRODUCTION

Finding the formula for the sums of integer powers of n consecutive integers, i.e., $S_{k,n} = 1^k + 2^k + \cdots + n^k$, is one of the most basic and classical problems in mathematics. In 1631, Johann Faulhaber of Ulm completed his work, *Academia Algebrae*, in which he offered formulas for the sums up to $\sum n^{17}$ [1]. In 1654, Pascal derived a formula for the sums of powers using binomial expansion and his work on the Arithmetical Triangle [2]. Then, in 1713, Jakob Bernoulli offered the following identity relating the sums of powers to the Bernoulli numbers in his book on probability, *Ars Conjectandi*:

$$S_{k,n} = 1^k + 2^k + \cdots + n^k = \sum_{i=0}^{k} \binom{k}{i} (-1)^i B_i \frac{n^{k+1-i}}{k+1-i}. \qquad (1)$$

Here, k represents the positive integers $1,2,3,\cdots$ and $B_0, B_1, B_2, \cdots$ represent the Bernoulli numbers. Bernoulli never offered a proof of (1); he simply wrote down the formulas for the first ten values of k and then claimed that the general pattern was apparent [3]! The classical proofs of Bernoulli's identity (1) are based on either the "Euler-Macluarin Summation Formula" [4] or on the Bernoulli polynomials [5]. In 1755, Euler published a proof of Bernoulli's results based on the calculus of finite differences. In honor of Jakob Bernoulli, he named the coefficients of n for $k = 2,4,6\cdots$ the *Bernoulli numbers*. It is interesting to note that Faulhaber had come up with these numbers, albeit a different method, almost a hundred years before Bernoulli. Then, in 1834, Jacobi applied the Euler-Maclaurin summation formula to the sums of powers and rediscovered the work of Faulhaber [2]. He provided proofs of several of Faulhaber's assertions on the sums of powers [1].

Since then, there has been renewed interest in the sums of powers problem, as can be attested by the number of articles on this topic found in well known mathematics journals. Knuth, in [1], has provided a computation-based approach by which Faulhaber may have discovered his results, and he has shown that similar results occur when the sums are expressed in terms of central factorial powers rather than in terms of ordinary powers.

Owens [6] and Scott [7] have provided proofs indicating that the sum of powers $S_{k,n}$ can be represented as a polynomial in n of degree $k+1$ with constant term equal to zero. Hence, we may write

$$S_{k,n} = a_{k+1}n^{k+1} + a_k n^k + \cdots + a_1 n^1. \qquad (2)$$

In addition, they have developed recursive formulas for the sums of powers by relating the coefficients of (2) and the coefficients of $S_{k+1,n}$, where

$$S_{k+1,n} = b_{k+2}n^{k+2} + b_{k+1}n^{k+1} + \cdots + b_1 n^1. \qquad (3)$$

Thus, they have developed algorithms for obtaining $S_{k+1,n}$ from $S_{k,n}$. The coefficients of $S_{k+1,n}$ are related to those of $S_{k,n}$ by the following formula:

$$b_{j+1} = \frac{k+1}{j+1} a_j \ , \quad \text{for } 1 \le j \le k+1$$

and

$$b_1 = 1 - (b_2 + b_3 + \cdots + b_{k+1}) . \tag{4}$$

In [6], Scott has also developed a system of equations for evaluating the coefficients of $S_{k,n}$. He proceeded in the following manner:

Knowing that $S_{k,n}$ is a polynomial in n of degree $k+1$ with a constant term of zero, he wrote that

$$1^k + 2^k + \cdots + n^k = S_{k,n} = a_1 n + a_2 n^2 + \cdots + a_{k+1} n^{k+1} \tag{5}$$

and

$$\begin{aligned} &1^k + 2^k + \cdots + n^k + (n+1)^k = S_{k,n+1} \\ &= a_1(n+1) + a_2(n+1)^2 + \cdots + a_{k+1}(n+1)^{k+1} . \end{aligned} \tag{6}$$

Subtracting (5) from (6), he obtained

$$(n+1)^k = a_1((n+1) - n) + a_2((n+1)^2 - n^2) + \cdots + a_{k+1}((n+1)^{k+1} - n^{k+1}) . \tag{7}$$

Comparing the coefficients of n^j on each side of (7) he found that

$$\binom{k}{j} = \binom{j+1}{j} a_{j+1} + \binom{j+2}{j} a_{j+2} + \cdots + \binom{k+1}{j} a_{k+1} ,$$

for $0 \le j \le k$. This system of linear equations may be expressed in matrix form as

$$\begin{pmatrix} \binom{k+1}{k} & 0 & 0 & 0 & \cdots & 0 \\ \binom{k+1}{k-1} & \binom{k}{k-1} & 0 & 0 & \cdots & 0 \\ \binom{k+1}{k-2} & \binom{k}{k-2} & \binom{k-1}{k-2} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \binom{k+1}{0} & \binom{k}{0} & \binom{k-1}{0} & \cdots & \cdots & \binom{1}{0} \end{pmatrix} \begin{pmatrix} a_{k+1} \\ a_k \\ a_{k-1} \\ \vdots \\ \vdots \\ a_1 \end{pmatrix} = \begin{pmatrix} \binom{k}{k} \\ \binom{k}{k-1} \\ \binom{k}{k-2} \\ \vdots \\ \binom{k}{0} \end{pmatrix} . \tag{8}$$

The matrix may then be solved to find the coefficients of $S_{k,n}$ in (5).

STATEMENT OF PROBLEM

In this paper, we shall present proofs relating to several different works on the sums of powers. We will provide two alternative methods for deriving the relation (4) between the coefficients of $S_{k,n}$ in (2) and the coefficients of $S_{k+1,n}$ in (3).

METHOD 1

The system of linear equations (8) developed by Scott may be used to develop an algorithm for finding the coefficients of $S_{k+1,n}$ when the coefficients of $S_{k,n}$ are known. The proof follows:

The same method used above in (5) - (8) may be applied to

$$1^{k+1} + 2^{k+1} + \cdots + n^{k+1} = S_{k+1,n} = b_1 n + b_2 n^2 + \cdots + b_{k+2} n^{k+2} \tag{9}$$

to find the coefficient matrix of $S_{k+1,n}$. It turns out to be the following system of $(k \times 2)$ equations:

$$\begin{pmatrix} \binom{k+2}{k+1} & 0 & 0 & 0 & \cdots & 0 \\ \binom{k+2}{k} & \binom{k+1}{k} & 0 & 0 & \cdots & 0 \\ \binom{k+2}{k-1} & \binom{k+1}{k-1} & \binom{k}{k-1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \binom{k+2}{0} & \binom{k+1}{0} & \binom{k}{0} & \cdots & \cdots & \binom{1}{0} \end{pmatrix} \begin{pmatrix} b_{k+2} \\ b_{k+1} \\ b_k \\ \vdots \\ b_1 \end{pmatrix} = \begin{pmatrix} \binom{k+1}{k+1} \\ \binom{k+1}{k} \\ \binom{k+1}{k-1} \\ \vdots \\ \binom{k+1}{0} \end{pmatrix} \quad (10)$$

Multiply both sides of (10) by

$$A = \begin{pmatrix} \frac{k+1}{k+1} & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{k}{k+1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{k-1}{k+1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \frac{k-2}{k+1} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{k+1} & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \quad (11)$$

Eliminating the $(k-2)^{th}$ row and column, (10) becomes

$$\begin{pmatrix} \frac{k+1}{k+1}\binom{k+2}{k+1} & 0 & 0 & \cdots & 0 \\ \frac{k}{k+1}\binom{k+2}{k} & \frac{k}{k+1}\binom{k+1}{k} & 0 & \cdots & 0 \\ \frac{k-1}{k+1}\binom{k+2}{k-1} & \frac{k-1}{k+1}\binom{k+1}{k-1} & \frac{k-1}{k+1}\binom{k}{k-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{k+1}\binom{k+2}{1} & \frac{1}{k+1}\binom{k+1}{1} & \frac{1}{k+1}\binom{k}{1} & \cdots & \frac{1}{k+1}\binom{2}{1} \end{pmatrix} \begin{pmatrix} b_{k+2} \\ b_{k+1} \\ b_k \\ \vdots \\ b_2 \end{pmatrix} = \begin{pmatrix} \binom{k}{k} \\ \binom{k}{k-1} \\ \binom{k}{k-2} \\ \vdots \\ \binom{k}{0} \end{pmatrix} \quad (12)$$

The coefficient matrix of the system of equations (12) can be written as

$$\begin{pmatrix} \binom{k+1}{k} & 0 & 0 & \cdots & 0 \\ \binom{k+1}{k-1} & \binom{k}{k-1} & 0 & \cdots & 0 \\ \binom{k+1}{k-2} & \binom{k}{k-2} & \binom{k-1}{k-2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \binom{k+1}{0} & \binom{k}{0} & \binom{k-1}{0} & \cdots & \binom{1}{0} \end{pmatrix} \begin{pmatrix} \frac{k+2}{k+1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{k+1}{k+1} & 0 & \cdots & 0 \\ 0 & 0 & \frac{k}{k+1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{2}{k+1} \end{pmatrix} \begin{pmatrix} b_{k+2} \\ b_{k+1} \\ b_k \\ \cdot \\ \cdot \\ \cdot \\ b_2 \end{pmatrix} = \begin{pmatrix} \binom{k}{k} \\ \binom{k}{k-1} \\ \binom{k}{k-2} \\ \vdots \\ \binom{k}{0} \end{pmatrix} \quad (13)$$

Comparing (8) and (13) we can see that

$$\begin{pmatrix} \frac{k+2}{k+1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{k+1}{k+1} & 0 & \cdots & 0 \\ 0 & 0 & \frac{k}{k+1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{2}{k+1} \end{pmatrix} \begin{pmatrix} b_{k+2} \\ b_{k+1} \\ b_k \\ \cdot \\ \cdot \\ b_2 \end{pmatrix} = \begin{pmatrix} a_{k+1} \\ a_k \\ a_{k-1} \\ \cdot \\ \cdot \\ a_1 \end{pmatrix} \tag{14}$$

Hence, for $1 \le j \le k+1$,

$$b_{j+1} = \frac{k+1}{j+1} a_j .$$

To find b_1, let us proceed in the following manner:
We know that

$$\sum n^{k+1} = b_{k+2} n^{k+2} + b_{k+1} n^{k+1} + \cdots + b_1 n .$$

Let $n=1$. Then

$$1 = b_{k+2} + b_{k+1} + \cdots + b_1$$

and, therefore,

$$b_1 = 1 - \sum_{j=1}^{k+1} b_{j+1} = 1 - \sum_{j=1}^{k+1} \left(\frac{k+1}{j+1} \right) a_j .$$

This completes the proof.

METHOD 2

Donald E. Knuth, in [1], presented Jacobi's proof of Faulhaber's formulas for even exponents and developed a recurrence between the coefficients of Faulhabers polynomials for sums of powers and the Bernoulli numbers. The ideas used to develop this recurrence can also be used to derive the relation between the coefficients of

$$S_{k,n} = a_{k+1} n^{k+1} + a_k n^k + \cdots + a_1 n^1$$

and the coefficients of

$$S_{k+1,n} = b_{k+2} n^{k+2} + b_{k+1} n^{k+1} + \cdots + b_1 n^1 .$$

Because the sums of powers can be expressed in terms of Bernoulli polynomials,

$$\sum n^k = \frac{1}{k+1} \left[B_{k+1}(n+1) - B_{k+1}(0) \right], \tag{15}$$

and

$$B_k'(x) = k B_{k-1}(x), \tag{16}$$

it can easily be shown that the derivative of $\sum n^k$ with respect to n is the following:

$$\left(\sum n^k \right)' = (k+1) \sum n^k + B_{k+1} . \tag{17}$$

Given the above identity, let us derive a relation between the coefficients of $S_{k,n}$ and $S_{k+1,n}$.

Let

$$S_{k,n} = \sum n^k = a_{k+1}n^{k+1} + a_k n^k + \cdots + a_1 n^1$$

and

$$S_{k+1,n} = \sum n^{k+1} = b_{k+2}n^{k+2} + b_{k+1}n^{k+1} + \cdots + b_1 n^1.$$

Then

$$\left(\sum n^{k+1}\right)' = (k+2)b_{k+2}n^{k+1} + (k+1)b_{k+1}n^k + (k)b_k n^{k-1} + \cdots + b_1 \tag{18}$$

$$= (k+1)\sum n^k + B_{k+1}$$

$$= (k+1)\left(a_{k+1}n^{k+1} + a_k n^k + a_{k-1}n^{k-1} + \cdots + a_1 n^1\right) + B_{k+1}. \tag{19}$$

Comparing the coefficients of the jth power of n in (18) and (19), we obtain

$$(j+1)b_{j+1} = (k+1)a_j, \quad \text{for } j = 1,2,\cdots,k+1.$$

Also, we can see that $b_1 = B_{k+1}$. Thus the relation between the coefficients of $S_{k,n}$ and $S_{k+1,n}$ is as follows:

$$b_1 = B_{k+1},$$

$$b_{j+1} = \frac{(k+1)}{(j+1)}a_j, \quad j = 1,2,\cdots,k+1.$$

This completes the proof.

SUMMARY

We presented two different methods for deriving the relation (5) between the coefficients of $S_{k,n}$ and the coefficients of $S_{k+1,n}$.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my faculty research advisor, Dr. Peter Yang, for his patience, guidance, time, and persistence given to me during my research project. I would also like to thank him for sharing his research and research methods with me. I am also very thankful for the opportunity and funding provided to me by the Young Scholar's Program at the University of St. Thomas.

REFERENCES

[1]. D.E. Knuth, *Johann Faulhaber and Sums of Powers*, Mathematics of Computation, 61: 277-294, 1993.

[2]. A.W.F. Edwards, *Sums of Powers of Integers: a Little of the History*, Mathematical Gazette, 66: 22-28, 1982.

[3]. Jeffrey Nunemacher, *On the Sum of Consecutive Kth Powers*, Mathematics Magazine, 60: 237-238, 1987.

[4]. M. Spivak, *Calculus*, Publish or Perish, 1980.

[5]. H. Edwards, *Fermat's Last Theorem*, Springer-Verlag, 1977.

[6]. Robert W. Owens, *Sums of Powers of Integers*, Mathematics Magazine, 65: 38-40, 1992.

[7]. S.H. Scott, *Sums of Powers of Natural Numbers*, Mathematical Gazette, Dec. 1980, 231-239.

An Empirical Investigation of Outdoor Advertising of Alcohol and Tobacco Targeting Blacks

Jeff Foerster

Department of Marketing
Seattle University
Broadway and Madison
Seattle, WA 98122

Faculty Advisor:

Dr. Sharon Galbraith

INTRODUCTION

A question that has been brewing and burning for some time is whether blacks are more frequently targeted with advertising for alcohol and tobacco products than other population groups . This topic has surfaced many times in classroom discussions and has become a favorite subject for critics. As a result the companies who sell these products have come under close scrutiny concerning promotional and advertising methods.

One of the specific criticisms is that the use of billboards as an advertising medium for alcohol and tobacco products has been disproportionately large in communities with higher than average concentrations of blacks (Becker, 1992). This becomes a significant claim with the correspondingly higher rates of alcoholism, smoking and smoking-related diseases occurring in black communities. Black adults smoke at a rate of 34% compared to 28% for whites and 27% for hispanics (Spratlen, 1991). These health concerns, along with the high incidence of these products in outdoor advertising (Davis, 1987), lean to the question of causality. Billboards are a very effective media and is the reason why they were studied in this research. Billboards act as a reinforcement cue for behavior because an individual may see a billboard message many times (King and Tinkham, 1989/1990). In addition, blacks have been shown to respond positively with behavior to this type of stimulus (Djata, 1987).

Two studies, conducted in St. Louis and Baltimore, have suggested that blacks are more heavily targeted than whites with alcohol and tobacco advertising. The city of St. Louis found twice as many such billboards in black neighborhoods (60%) than in white neighborhoods. Baltimore topped this with a 70% rate of advertising of alcohol and tobacco in the black neighborhoods (Becker, 1992).

The alcohol and tobacco industries are very aware of the criticisms made against them and has responded both reactively and proactively to strengthen their image. For example, in 1991 R.J. Reynolds was ready to launch a new brand of cigarettes targeted specifically to blacks called "Uptown" when public outrage lead R.J.R. to abandon its project. Another example of the tobacco companies' efforts to maintain good relations with the black community is the fact that they have been making donations to fund local community events and help support black organizations such as the NAACP. Additionally, some alcohol companies have sponsored public service messages regarding responsible use of alcohol and warning against underage use.

THIS STUDY

The purpose of this study is to compare the billboard advertising of alcohol and tobacco products in a black community of Seattle with that in a similar non-black community. Based on previous studies and journalistic criticism, one would expect to find a higher incidence of alcohol and tobacco billboard advertising in the black community. Therefore, the hypothesis being tested is: There are significantly more

find a higher incidence of alcohol and tobacco billboard advertising in the black community. Therefore, the hypothesis being tested is: There are significantly more billboard advertisements for alcohol and tobacco products in a black community than in a non-black community of Seattle.

RESEARCH METHODOLOGY

Demographic information was obtained through census data for the Seattle area. Four tracts of land were chosen for the purpose of this study: two adjacent tracts in South Seattle along Rainier Avenue S. and two adjacent tracts in North Seattle along Aurora Avenue N.. The Rainier Area has a higher than average percentage of blacks for King County (36% vs. 5%), while Aurora Avenue has approximately the same population statistics regarding blacks as all of King County (85% white, 2.5% black).

Billboard readings were taken at approximately 30 day intervals because billboards are commonly leased for 30 day periods. For each billboard along the main arterials in the two communities five pieces of information were gathered:

* Product being advertised
* Brand Name
* Slogan/Advertising Theme
* Direction Billboard Facing
* Graphics/Visuals

Data was collected for six consecutive months on the last day of each month, with the first reading being taken on October 31, 1994.

FINDINGS

The Rainier Avenue Area contained 10 billboards. With the six readings, there was a possible occurrence of 60. Along Aurora Avenue there were 17 billboards, which provided a possible occurrence of 102. The billboard readings were tabulated and assembled into information showing the incidence of alcohol and tobacco advertising in the two areas (see tables 1 and 2). These calculations showed that 38.3% of the billboards in the black community (23 of a possible 60) displayed alcohol or Tobacco products (see table 1) while only 24.5% of the those in the white community did so (25 of a possible 102) (see table 2).

A z-test for the difference between the two billboard population proportions was performed, with the calculated z equal to 1.86. The one-tailed probability is equal to .0314, meaning that if there were the same number of alcohol and tobacco billboards between the two areas, there is only a 3.14% probability that one would get the results shown here. Therefore it was concluded that there is a statistically significant difference between the two areas, and that the Rainier Avenue, or black neighborhood, has more alcohol and tobacco advertisements on billboards than the white, or Aurora Avenue, area over the six month period that was studied.

An interesting note concerning the data was the predominance of tobacco products. These advertisements occurred over three times as often as did any ads for alcoholic products. Another interesting finding was that during the holiday season (reading date of December 31, 1994) the incidence of alcohol and tobacco ads in the Rainier, or black, area doubled to 60%, while those in the Aurora, or white, area actually decreased from the previous month.

CAVEATS

A few issues regarding these findings should be noted out of concern for accuracy. First, the findings from this study are not meant to be representative of any communities other than the ones sampled in the city of Seattle. Second, in pure numbers there were more billboards displaying alcohol and/or tobacco ads in the Aurora area (25) than in the Rainier area (23). This would seem to contradict the results if it had not been for two things: billboard percentage and population size. The percentage of alcohol and tobacco billboards was compared, not the absolute numbers,

for a more meaningful comparison. Also, the population in the Aurora area is greater than that of the Rainier area, and this would lead one to expect more billboards and therefore more ads. The third of these issues involves the inclusion of a billboard in the Rainier area. The billboard is located just outside the Northwest corner of the census tract, but was included because it could clearly be seen while still inside the boundary of the area. Lastly, an advertisement from Miller Brewing Co. promoting responsible drinking was counted as an alcoholic ad because beer was displayed prominently and depicted a social setting.

CONCLUSION

In conclusion, the results of this study show that the Rainier Avenue area with a greater than average concentration of blacks has a higher percentage of billboards advertising alcohol and tobacco products than the Aurora Avenue area with a low percentage of blacks and population statistics close to those of King County. Therefore, it can be concluded that alcohol and tobacco manufacturers are targeting inner-city blacks.

As a result of all the attention this subject has been receiving a number of things have been happening. In the black communities themselves there is a proactive campaign to combat the alcohol and tobacco advertising using billboards as a tool. Community groups have rented space to display ads centered around prevention and education of the harmful consequences drinking and smoking (Skolnick, 1993). Another trend is an activity known as "white-washing" in which participants, armed with paint and brushes, cover any ads they find objectionable.

Additionally the industries are increasingly active themselves. The billboard industry has increased the amount of space given to non-profit oranizations for public service messages. Also, alcohol and tobacco industries have been funding community events in black neighborhoods and giving corporate gifts to black organizations, as noted earlier.

In regards to the white-washing trend and surrounding controversy, former NAACP Executive Director Benjamin Hooks has said he sees nothing wrong with how the industries have conducted themselves, and that such actions and criticisms actually insult the intelligence of blacks by implying they are unable to make their own purchase decisions. His comments seem somewhat suspect when we realize the NAACP had received a substantial grant to help fund their annual conference.

The issue of Alcohol and tobacco advertising in black neighborhoods is not fading away, rather, both sides are firmly entrenched and holding their ground. Therefore, the future of this issue is in the hands of those it affects and those who effect it; the black community and the alcohol and tobacco industries, as well as concerned citizens outside these groups who recognize and support the well-being of the black community.

Observations on Five Dates of Alcohol and Tobacco Advertising
On Ten Billboards on Rainier Avenue (Black Neighborhood)

Date	# / % of Alcohol Ads	# / % of Tobacco Ads	# / % of Total Alcohol & Tobacco Ads
Oct. 31	0/0%	2/20%	2/20%
Nov. 30	1/10%	2/20%	3/30%
Dec. 31	3/30%	3/30%	6/60%
Jan. 31	1/10%	3/30%	4/40%
Feb. 28	2/20%	3/30%	5/50%
Mar. 31	0/0%	3/30%	3/30%
Total/Mean	7/11.6%	16/26.7%	23/38.3%
* Percentages are rounded			

Observations on Five Dates of Alcohol and Tobacco Advertising
On Seventeen Billboards on Aurora Avenue (White Neighborhood)

Date	# / % of Alcohol Ads	# / % of Tobacco Ads	# / % of Total Alcohol & Tobacco Ads
Oct. 31	1/5.9%	4/23.5%	5/29.4%
Nov. 30	1/5.9%	4/23.5%	5/29.4%
Dec. 31	1/5.9%	3/17.6%	4/23.5%
Jan. 31	1/5.9%	3/17.6%	4/23.5%
Feb. 28	0/0%	4/23.5%	4/23.5%
Mar. 31	1/5.9%	2/11.8%	3/17.6%
Total/Mean	5/4.9%	20/19.6%	25/24.5%

REFERENCES

Becker, Jeffrey G., Brown Jr, Jesse W., Castellano, Joseph P., Johnson, Elaine M., and Pomeroy, Henry J.. "Distilling the Truth About Alcohol Ads", Business and Society Review, Fall 1992, n83, pp12-17.

Djata[sic] "The Marketing of Vices to Black Consumers", Business and Society. Review, Summer 1992, 62, pp47-49.

King, Karen Whitehill and Tinkham, Spencer F. "The Learning and Retention of Outdoor Advertising". Journal of Advertising Research, (December/January 1989/1990). 29, pp47-51.

Skolnick, Andrew A., "National Medical Association Unveils Billboard Campaign to Promote Health in Black Communities", Journal of the American Medical Association, September 8, 1993, v270 n10, pp1166-1167.

Spratlen, Thaddeus H., "The Controversy Over Targeting Black Consumers in Cigarrette Advertising: Racial/Ethnic and Ethical Issues" Proceedings, Academy of Marketing Science, Robert L. King, Ed., Vol XIV Developments in Marketing Science, (Fifteenth Annual Conference, Fort Lauderdale, FL, May 29-June 1, 1991).

INVERSION OF AN INTEGRAL TRANSFORM FOR SU(p,1)

DEBORAH ALTERMAN, D. MATTHEW HILL, SEJUNG KIM, AND ANGELA M. MINNICK
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109–1003
Department of Mathematics, University of California, Los Angeles, CA 90024–1555
Department of Mathematics, Cornell University, Ithaca, NY 14853–7901
Department of Mathematics, University of Iowa, Iowa City, IA 52242–1466
FACULTY ADVISOR: LISA A. MANTINI
Department of Mathematics, Oklahoma State University, Stillwater, OK 74078–0613

INTRODUCTION

Integral transforms such as the Fourier, Laplace, and Radon transforms have been studied for more than 200 years. They are of interest both for their intrinsic beauty and for their numerous applications to differential equations, x-ray tomography, and other subjects throughout science and engineering (see [AS]). In the theory of representations of Lie groups, integral transforms often give intertwining operators between two different realizations of a given representation. One example of this type is the Penrose transform of mathematical physics, which intertwines the actions of the group SU(2, 2) on spaces of solutions of massless field equations such as Maxwell's equations (see [W]).

In this paper we consider an integral transform similar to a Penrose or Radon transform (see [H]) which arose in studying the representation theory of the group SU(p, q) [M]. Here we consider the special case for which $q = 1$. The transform maps a space of functions in $p+1$ complex variables to a space of holomorphic functions on the unit disk in the complex plane. The problem we wish to solve is to determine the image of the transform, that is, to somehow characterize which holomorphic functions will occur in the image, and to find an inversion formula for the transform.

PRELIMINARIES

We begin with some notation and useful lemmas about complex vector spaces. By $z \in \mathbb{C}^n$ we mean an ordered n-tuple $z = (z_1, \ldots, z_n)$ of complex numbers with $z_j = x_j + iy_j$ for each j. The norm of $z \in \mathbb{C}^n$ is then given by $|z|^2 = |z_1|^2 + \cdots + |z_n|^2 = x_1^2 + y_1^2 + \cdots + x_n^2 + y_n^2$. We normalize Lebesgue measure on $\mathbb{C}^n$ by letting $dm(z) = (1/\pi^n)\, dx_1\, dy_1 \cdots dx_n\, dy_n$, so that $\int_{\mathbb{C}^n} e^{-|z|^2}\, dm(z) = 1$, as follows from Lemma 1.

LEMMA 1. *If j and k are nonnegative integers, then*

$$\int_{\mathbb{C}} z^k \bar{z}^j e^{-|z|^2}\, dm(z) = \begin{cases} 0, & j \neq k, \\ k!, & j = k. \end{cases}$$

PROOF. Use polar coordinates. ∎

A differentiable function $f(z)$ of n complex variables $z = (z_1, \ldots, z_n)$ is *holomorphic* in z_j if $\partial f/\partial \bar{z}_j = 0$ and *antiholomorphic* in z_j if $\partial f/\partial z_j = 0$. We know that a holomorphic function f may be expressed as the sum of a convergent power series in z and that an antiholomorphic function f is the sum of a convergent power series in $\bar{z}$ (see [R]).

We consider the hermitian form of signature $(p, 1)$ on $\mathbb{C}^{p+1}$ given by

$$h(z, w) = z_1\bar{w}_1 + \cdots + z_p\bar{w}_p - z_{p+1}\bar{w}_{p+1} = {}^*w\, I_{p,1}\, z, \tag{1}$$

where $I_{p,1}$ is the diagonal matrix with p +1's and one −1 on the diagonal. The group SU($p, 1$) is the group of linear transformations of $\mathbb{C}^{p+1}$ preserving this form. We say that a subspace V of $\mathbb{C}^{p+1}$ is positive (resp. negative) if $h|V$ is positive (resp. negative) definite. First we wish to parameterize the negative lines. Any complex line V in $\mathbb{C}^{p+1}$ has a single non-zero vector $z = (z_1, \ldots, z_{p+1})$ as a basis. If $z_{p+1} \neq 0$, we multiply through by z_{p+1}^{-1} to obtain a basis vector of the form $(\zeta_1, \ldots, \zeta_p, 1)$, where now clearly a different choice of $\zeta \in \mathbb{C}^p$ gives a different line. We let

$$V(\zeta) = \{(\zeta_1 w, \ldots, \zeta_p w, w) \mid w \in \mathbb{C}\}. \tag{2}$$

LEMMA 2. *Let $\mathbb{C}^{p+1}$, h, and $V(\zeta)$ be as described above. Then a line in $\mathbb{C}^{p+1}$ is negative if and only if it is of the form $V(\zeta)$, for ζ satisfying $|\zeta| < 1$.*

PROOF. Any negative line must have $z_{p+1} \neq 0$, by (1), so is of the form $V(\zeta)$ for some ζ. Fix $\zeta \in \mathbb{C}^p$, and consider the point $(w\zeta_1, \ldots, w\zeta_p, w) \in V(\zeta)$. Then (1) implies that

$$h((w\zeta_1, \ldots, w\zeta_p, w), (w\zeta_1, \ldots, w\zeta_p, w)) = |w\zeta_1|^2 + \cdots + |w\zeta_p|^2 - |w|^2 = |w|^2(|\zeta|^2 - 1).$$

Thus the restriction of h to $V(\zeta)$ is now negative definite if and only if $|\zeta|^2 - 1 < 0$. ∎

We denote by $\mathcal{D}_p$ the set of all $\zeta \in \mathbb{C}^p$ satisfying $|\zeta| < 1$. This is the open ball of radius one in $\mathbb{C}^p$, usually called the unit disk when $p = 1$. If $\mathcal{M}_-$ denotes the space of negative lines in $\mathbb{C}^{p+1}$, then Lemma 2 shows that $\mathcal{M}_-$ is parameterized by $\mathcal{D}_p$.

A parameterization of the maximal positive subspaces of $\mathbb{C}^{p+1}$ involves analysis that is beyond the scope of this work. So we restrict this discussion to the case $p = 1$. Any subspace U of $\mathbb{C}^2$ has a basis vector (z_1, z_2), or, equivalently, $(1, \eta)$ if $z_1 \neq 0$. We let

$$U(\eta) = \{(w, \eta w) \mid w \in \mathbb{C}\}. \tag{3}$$

LEMMA 3. *A line in $\mathbb{C}^2$ is positive if and only if it is of the form $U(\eta)$ for $|\eta| < 1$.*

Thus we see that the unit disk $\mathcal{D}_1$ parameterizes $\mathcal{M}_+$, the space of positive lines in $\mathbb{C}^2$. We use these spaces in the definition of our integral transform.

THE FUNCTION SPACE $\mathcal{H}$

Our goal in this section is to define the space of functions on which our integral transform will act. For simplicity of notation we restrict to the case $p = q = 1$.

First we define an inner product on functions f and g on $\mathbb{C}^2$ as

$$\langle\!\langle f, g \rangle\!\rangle = \int_{\mathbb{C}^2} f(z)\, \overline{g(z)}\, e^{-|z|^2}\, dm(z).$$

For example, let f_{ij} denote the monomial $f_{ij}(z) = z_1^i\, \bar{z}_2^j$, where i and j are non-negative integers. Then Lemma 1 shows that

$$\langle\!\langle f_{ij}, f_{kl} \rangle\!\rangle = \begin{cases} 0, & i \neq k \text{ or } j \neq l, \\ i!\, j!, & i = k \text{ and } j = l. \end{cases} \tag{4}$$

We may now define our space of functions $\mathcal{H}$ so that

$$\mathcal{H} = \left\{ f(z) = f(z_1, z_2) \;\middle|\; \frac{\partial f}{\partial \bar{z}_1} = 0,\ \frac{\partial f}{\partial z_2} = 0,\ \langle\!\langle f, f \rangle\!\rangle < \infty \right\}.$$

The space $\mathcal{H}$ is sometimes called the Bargmann-Segal-Fock space, first studied in [B]. From (4), we note that any polynomial in z_1 and $\bar{z}_2$ will be in $\mathcal{H}$. The other functions f in $\mathcal{H}$ are expressible as power series in z_1 and $\bar{z}_2$, $f(z_1, z_2) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{jk}\, z_1^j\, \bar{z}_2^k$, where the condition $\langle\!\langle f, f \rangle\!\rangle < \infty$ controls the growth of the coefficients a_{jk}.

We are also interested in the subsets of $\mathcal{H}$ given for each integer $m \in \mathbb{Z}$ by

$$\mathcal{H}_m = \{f \in \mathcal{H} \mid f(e^{-i\theta} z) = e^{im\theta} f(z) \text{ for all } z \in \mathbb{C}^2, \theta \in \mathbb{R}\}. \tag{5}$$

We refer to m as the weight of $\mathcal{H}_m$.

LEMMA 4. *The power series expansion of any $f \in \mathcal{H}_m$ has the form*

$$f(z_1, z_2) = \sum_{j=\max\{0,-m\}}^{\infty} c_j\, z_1^j\, \bar{z}_2^{j+m} \tag{6}$$

where the coefficients c_j satisfy

$$\langle\!\langle f, f \rangle\!\rangle = \sum_{j=\max\{0,-m\}}^{\infty} j!\,(j+m)!\, |c_j|^2 < \infty. \tag{7}$$

PROOF. Any monomial f_{jk} satisfies $f_{jk}(e^{-i\theta} z) = (e^{-i\theta} z_1)^j\, \overline{(e^{-i\theta} z_2)}^k = e^{-ij\theta}\, z_1^j\, e^{ik\theta}\, \bar{z}_2^k = e^{i\theta(k-j)}\, f_{jk}(z)$. Thus a monomial $f_{jk} \in \mathcal{H}_m$ if and only if $k - j = m$, or $k = j + m$. Equation (6) follows. Then (7) follows by applying Lemma 1. ∎

Any $f \in \mathcal{H}$ may be expressed uniquely as a sum of functions in $\mathcal{H}_m$, by grouping the terms in the power series for f according to their "homogeneity degree" $k-j$. This says that $\mathcal{H}$ is the direct sum of the subspaces $\mathcal{H}_m$, written as $\mathcal{H} = \bigoplus_{m\in\mathbb{Z}} \mathcal{H}_m$. The group $\mathrm{SU}(p,1)$ acts irreducibly on each of these spaces.

CONSTRUCTION OF THE INTEGRAL TRANSFORM

In this section we give a simplified version of the construction of our integral transform in the case when $p = 1$ and $m \geq 0$. See [M] or [O] for more details.

Like a Radon or Penrose transform, our transform is constructed by choosing $f \in \mathcal{H}_m$, restricting f to a negative line $V(\zeta) \in \mathcal{M}_-$, where $\zeta \in \mathcal{D}_1$, and then integrating. From (2) we may think of $f|_{V(\zeta)}$ as a function of $w \in \mathbb{C}$ where $f|_{V(\zeta)}(w) = f(\zeta w, w)$. For example, if $f_{jk}(z) = z_1^j\, \bar{z}_2^k$, as before, Lemma 1 shows that

$$\int_{\mathbb{C}} f_{jk}(\zeta w, w)\, w^n\, e^{-|w|^2}\, dm(w) = \begin{cases} 0, & n \neq k-j, \\ k!\,\zeta^j, & n = k-j. \end{cases} \tag{8}$$

Thus Lemma 4 and (8) motivate us for each $m \geq 0$ to define

$$(\Phi_m f)(\zeta) = \int_{\mathbb{C}} f(\zeta w, w)\, w^m\, e^{-|w|^2}\, dm(w), \tag{9}$$

for $f \in \mathcal{H}_m$. The transform Φ_m maps a function $f \in \mathcal{H}_m$ into a function $\Phi_m f$ on the parameter space $\mathcal{D}_1$ for $\mathcal{M}_-$, the unit disk in the complex plane.

LEMMA 5 [M]. *Let $m \geq 0$, and let $f \in \mathcal{H}_m$. Then the integral in (9) converges for each $\zeta \in \mathcal{D}_1$. Furthermore, if f has a power series expansion as in (6), then the image $\Phi_m(f)$ satisfies*

$$(\Phi_m f)(\zeta) = \sum_{j=0}^{\infty} c_j\, (j+m)!\, \zeta^j. \tag{10}$$

This series converges for $|\zeta| < 1$ and so determines a holomorphic function of $\zeta \in \mathcal{D}_1$.

We may let $\mathcal{O}(\mathcal{D}_1)$ denote the space of all holomorphic functions on the unit disk $\mathcal{D}_1$. The main result concerning the transform Φ_m is now the following.

THEOREM 1 [M]. *The transform $\Phi_m : \mathcal{H}_m \to \mathcal{O}(\mathcal{D}_1)$ as defined in equation (9) is well-defined and injective for $m \geq 0$.*

INVERSION OF THE TRANSFORM

In this section we develop an inversion formula for our integral transform in the case when $p = 1$ and $m \geq 0$. We then use this formula to characterize the image of the transform. First we present a useful integration lemma.

LEMMA 6. *If j and k are non-negative integers, then*

$$\int_{\mathcal{D}_1} \zeta^j\, \bar{\zeta}^k\, dm(\zeta) = \begin{cases} 0, & j \neq k, \\ \dfrac{1}{j+1}, & j = k. \end{cases}$$

PROOF. Use polar coordinates. ∎

THEOREM 2. *If $g \in \Phi_m(\mathcal{H}_m)$, then $\Phi_m^{-1}(g)$ is given by*

$$(\Phi_m^{-1} g)(z_1, z_2) = \begin{cases} \bar{z}_2^m \displaystyle\int_{\mathcal{D}_1} g(\zeta z_1 \bar{z}_2)\, \zeta^{m-1}\, e^{\bar{\zeta}}\, dm(\zeta), & \text{if } m > 0 \\ g(0) + \displaystyle\int_{\mathcal{D}_1} [g(\zeta z_1 \bar{z}_2) - g(0)]\, \zeta^{-1}\, e^{\bar{\zeta}}\, dm(\zeta), & \text{if } m = 0. \end{cases}$$

PROOF. Assume that g has power series expansion $\sum_{j=0}^{\infty} a_j\, \zeta^j$. We first consider the case $m > 0$. Using power series implies that

$$\begin{aligned} \bar{z}_2^m \int_{\mathcal{D}_1} g(\zeta z_1 \bar{z}_2)\, \zeta^{m-1}\, e^{\bar{\zeta}}\, dm(\zeta) &= \sum_{j=0}^{\infty} a_j\, z_1^j\, \bar{z}_2^{j+m} \int_{\mathcal{D}_1} \zeta^{j+m-1}\, e^{\bar{\zeta}}\, dm(\zeta) \\ &= \sum_{j=0}^{\infty} a_j\, z_1^j \bar{z}_2^{j+m} \int_{\mathcal{D}_1} \zeta^{j+m-1} \sum_{k=0}^{\infty} \frac{\bar{\zeta}^k}{k!}\, dm(\zeta), \end{aligned}$$

since the series converge uniformly on compact subsets of $\mathcal{D}_1$. Now using Lemma 6 it follows that

$$\bar{z}_2^m \int_{\mathcal{D}_1} g(\zeta z_1 \bar{z}_2)\, \zeta^{m-1}\, e^{\bar{\zeta}}\, dm(\zeta) = \sum_{j=0}^{\infty} \frac{a_j}{(j+m)!}\, z_1^j\, \bar{z}_2^{j+m}.$$

Lemma 5 implies that the function $f(z_1, z_2)$ so obtained satisfies that $\Phi_m f = g$. Thus this integral formula does invert Φ_m when $m > 0$.

The case $m = 0$ is completely analogous except that we must deal separately with the constant term in the power series for g in order for the integrand to be proper. ∎

We have demonstrated a process for inverting the operator if we already know that a given function is in the image. This creates a need for a criterion that will determine whether or not a given function $g \in \mathcal{O}(\mathcal{D}_1)$ satisfies $g \in \Phi_m(\mathcal{H}_m)$ for any m. This criterion is given in the following theorem.

THEOREM 3. *Let $g \in \mathcal{O}(\mathcal{D}_1)$. Then $g \in \Phi_m(\mathcal{H}_m)$ if and only if $\langle g, g\rangle_m < \infty$ where*

$$\langle g, g\rangle_m = \begin{cases} \dfrac{1}{(m-1)!} \displaystyle\int_{\mathcal{D}_1} |g(\zeta)|^2\, (1-|\zeta|^2)^{m-1}\, dm(\zeta), & m > 0, \\ \displaystyle\sup_{0 \le r < 1} \frac{1}{2\pi} \int_0^{2\pi} |g(re^{i\theta})|^2\, d\theta, & m = 0. \end{cases}$$

Furthermore, if this is the case, then $\langle g, g\rangle_m = \langle\!\langle \Phi_m^{-1} g, \Phi_m^{-1} g \rangle\!\rangle$.

PROOF. Let $g(\zeta) = \sum_{j=0}^{\infty} a_j \zeta^j$. Lemma 5 implies that $\Phi_m^{-1} g$, if it exists, has power series $\sum_{j=0}^{\infty} (a_j/(j+m)!)\, z_1^j \bar{z}_2^{j+m}$. By Lemma 4,

$$\langle\!\langle \Phi_m^{-1} g, \Phi_m^{-1} g \rangle\!\rangle = \sum_{j=0}^{\infty} |a_j|^2\, \frac{j!}{(j+m)!}, \tag{11}$$

and it can be shown (see [O]) that $g \in \Phi_m(\mathcal{H}_m)$ if and only if the sum in (11) is finite. Now we calculate as follows. For $m > 0$,

$$\begin{aligned} \langle g, g\rangle_m &= \frac{1}{(m-1)!} \int_{\mathcal{D}_1} |g(\zeta)|^2\, (1-|\zeta|^2)^{m-1}\, dm(\zeta) \\ &= \frac{1}{(m-1)!} \int_{\mathcal{D}_1} |\sum_{j=0}^{\infty} a_j \zeta^j|^2 (1-|\zeta|^2)^{m-1}\, dm(\zeta) \\ &= \sum_{j=0}^{\infty} |a_j|^2\, \frac{1}{(m-1)!} \int_{\mathcal{D}_1} |\zeta|^{2j} (1-|\zeta|^2)^{m-1}\, dm(\zeta). \end{aligned}$$

By using polar coordinates and then the definition of the beta function, we see that

$$\langle g, g\rangle_m = \sum_{j=0}^{\infty} |a_j|^2\, \frac{j!}{(j+m)!}.$$

Thus the result holds for $m > 0$. The proof for $m = 0$ is similar. ∎

THE NEGATIVE WEIGHT CASE

In this section we extend the results of the previous section to the case when $p = 1$ and the weight m is negative. First we must define the transform for the case $m < 0$. In analogy with (8), we note that

$$\int_{\mathbb{C}} f_{jk}(w, \eta w)\, \bar{w}^n\, e^{-|w|^2}\, dm(w) = \begin{cases} 0, & n \neq j-k, \\ j!\, \bar{\eta}^k, & n = j-k. \end{cases} \tag{12}$$

Thus, as in (9), we define $\Psi_m : \mathcal{H}_m \to \mathcal{O}(\mathcal{D}_1)$ by

$$(\Psi_m f)(\zeta) = \int_{\mathbb{C}} f(w, \bar{\zeta} w)\, \bar{w}^{|m|}\, e^{-|w|^2}\, dm(w), \tag{13}$$

for $f \in \mathcal{H}_m$, when $m < 0$. As before, this transform is well-defined and injective for $m < 0$, and we can calculate the image of a function $f \in \mathcal{H}_m$ in terms of a power series.

THEOREM 4. *Let $g \in \Psi_m(\mathcal{H}_m)$ with $m < 0$. Then $\Psi_m^{-1}(g)$ is given by*

$$(\Psi_m^{-1} g)(z_1, z_2) = z_1^{|m|} \int_{\mathcal{D}_1} g(\bar{\zeta} z_1 \bar{z}_2) \bar{\zeta}^{-(m+1)} e^{\zeta}\, dm(\zeta).$$

Furthermore, if $m < 0$, then $g \in \Psi_m(\mathcal{H}_m)$ if and only if $\langle g, g \rangle_m < \infty$ where

$$\langle g, g \rangle_m = \frac{1}{(|m|-1)!} \int_{\mathcal{D}_1} |g(\zeta)|^2 (1 - |\zeta|^2)^{|m|-1}\, dm(\zeta).$$

PROOF. The proof is completely analogous to the proofs of Theorems 2 and 3. ∎

GENERALIZATION TO HIGHER DIMENSIONS

We close by noting that an analogue of Theorem 2 holds in the case when $p > 1$ and $m \geq 0$. The integral transform Φ_m for $m \geq 0$ is defined on a space $\mathcal{H}$ of functions of $p+1$ complex variables given by

$$\mathcal{H} = \left\{ f(z) = f(z_1, \ldots, z_p, z_{p+1}) \;\middle|\; \frac{\partial f}{\partial \bar{z}_1} = 0, \cdots, \frac{\partial f}{\partial \bar{z}_p} = 0,\ \frac{\partial f}{\partial z_{p+1}} = 0,\ \langle\!\langle f, f \rangle\!\rangle < \infty \right\},$$

where the inner product and the space $\mathcal{H}_m$ are defined analogously as on $\mathbb{C}^2$. The integral transform uses negative lines $V(\zeta)$ for $\zeta \in \mathcal{D}_p$, as in (2), and is given by

$$(\Phi_m f)(\zeta_1, \ldots, \zeta_p) = \int_{\mathbb{C}} f(w\zeta_1, \ldots, w\zeta_p, w)\, w^m\, e^{-|w|^2}\, dm(w), \tag{14}$$

for $f \in \mathcal{H}_m$ and $m \geq 0$. This function $\Phi_m f$ is holomorphic on the disc $\mathcal{D}_p$. The inversion formula, as in Theorem 2, is then given for $g \in \Phi_m(\mathcal{H}_m)$ by

$$(\Phi_m^{-1} g)(z_1, \ldots, z_{p+1})$$

$$= \begin{cases} \bar{z}_{p+1}^m \displaystyle\int_{\mathcal{D}_1} g(t\bar{z}_{p+1} z_1, \ldots, t\bar{z}_{p+1} z_p)\, t^{m-1}\, e^{\bar{t}}\, dm(t), & \text{if } m > 0, \\ g(0, \ldots, 0) + \displaystyle\int_{\mathcal{D}_1} [g(t\bar{z}_{p+1} z_1, \ldots, t\bar{z}_{p+1} z_p) - g(0, \ldots, 0)]\, t^{-1}\, e^{\bar{t}}\, dm(t), & \text{if } m = 0. \end{cases}$$

ACKNOWLEDGMENTS

The authors would like to thank E. Dunne, A. Noell, D. Ullrich, and D.J. Wright for useful discussions concerning this work. During summer 1993 the authors were supported by National Science Foundation Grant DMS–9300570 at the Research Experience for Undergraduates (REU) site at Oklahoma State University, where this research was conducted. Lisa Mantini also gratefully acknowledges support from the American Association of University Women and from NSF Grant DMS–9304580 at the Institute for Advanced Study while this paper was being prepared.

REFERENCES

[AS] L. Andrews and B. Shivamoggi, *Integral Transforms for Engineers and Applied Mathematicians,* Macmillan, New York, 1988.

[B] V. Bargmann, "On a Hilbert space of analytic functions and an associated integral transform I," *Comm. Pure Appl. Math.* 14 (1961), 187-214.

[H] S. Helgason, *The Radon Transform,* Birkhäuser, Boston, 1980.

[M] L. Mantini, "An integral transform in L^2-cohomology for the ladder representations of $U(p,q)$," *J. Funct. Anal.* **60** (1985), 211-242.

[O] M. Oehrtman and L. Mantini, "Power Series and Inversion of an Integral Transform," to appear.

[R] W. Rudin, *Real and Complex Analysis,* McGraw-Hill Book Company, New York, 1974.

[W] R.O. Wells, Jr., "Complex manifolds and mathematical physics," *Trans. Amer. Math. Soc.* **1** (1979), 296–336.

Monitoring of corrosion in steel structures using optical fiber sensors

L.R. McLaughlin

Photonics Technology Laboratory
Department of Electrical Engineering
Lafayette College, Easton, PA 18042-1775

Faculty Advisor:

K.D. Bennett

A prototype optical fiber sensor for monitoring corrosion on large steel structures has been designed and tested with favorable results. The sensor works by pulling a multimode fiber into a tight bend and securing it with a "corrosion fuse." When the fuse corrodes, it eventually breaks and allows the fiber to straighten. The resulting difference in optical intensity emerging from the fiber is measurable using an OTDR or other optical detector. Initial experiments were carried out to determine the effect of bending fibers in a small radius and showed the feasibility of the device. Following, tests were performed on three in-line sensors in a simulated corrosive atmosphere, and showed that this cheap and easily implemented monitoring scheme could be used to infer the presence of corrosion at different locations, and/or the degree of corrosion at a single location.

INTRODUCTION

Corrosion in large steel structures is a major problem to engineers everywhere in the world. In one form or another, corrosion attacks structures and dangerously degrades them. For example, in the United States there are approximately 576,000 bridges, 41% of which are in substandard condition and are considered potentially unsafe (White, Minor, and Derucher, 1992). According to national statistics, a bridge in the United States sags, buckles or collapses every other day. Corrosion is one of the leading causes behind this problem. In an attempt to combat this problem, the objective of this project is to design and fabricate a practical optical fiber sensor to monitor corrosion on large steel structures.

Although there are several forms of corrosion, this work concentrated on the effects of uniform corrosion. Metals corrode from uniform corrosion via electrochemical reactions (i.e. iron to iron oxide), and it is through this process that the greatest volume of material is lost in civil structures (Fontana, 1986). Since it is a uniform effect, it is hoped that a mathematical model can be formulated for comparing the loss in a sample of the material, found in a sensor, to the amount lost in the actual structure, similar to Komp's model for weathering steel (Komp, 1987). Additionally, uniform corrosion is more convenient to monitor since it occurs on exposed surfaces, as opposed to crevice or galvanic corrosion, which occurs in cracks and connections between materials.

Though a variety of methods are presently being used to monitor corrosion in structures, not all are satisfactory. The most obvious and established method is direct inspection, in which a team of inspectors goes to a site and physically checks each member of the structure, looking for signs of corrosion. However, this technique requires much time and manpower, and in many cases the parts of the structure needing inspection are remote and dangerous for the investigators to access.

Another method in use compares the corrosion of a sample material located near the structure with that of the structure itself. Samples are left in place for an extended period of time, usually several months or years, after which they are removed from the field site and taken back to a laboratory. Tests are performed, can indicate the corrosion experienced by the structure.

In addition, there are several indirect measurement methods coming into use. One of these methods is the AC impedance method, in which the resistance of the structural member is measured. As the corrosion in a structural member increases, the resistance of that member also increases (Fontana, 1986). Another method involves the use of a corrosion coulometer (Granata, Wilson, Jantz, Ettelman, and Wildt, 1994). As a sample in close proximity to the structure corrodes, the current produced during the electrochemical reactions causing corrosion is measured by an inexpensive ammeter.

With the variety of methods described above, it may be asked why a corrosion sensor using optical fiber is necessary. For this application, optical fiber sensoHx have numerous advantages over other monitoring methods. The flexibility, negligible weight, and small size of optical fiber are valuable attributes in a structural application. In addition, fiber allows signals to be transmitted over long distances with little loss, making it possible to remotely monitor the sensors. The ability of optical fiber to resist adverse environments is a great advantage, as the sensors will be constantly exposed to the elements. Finally, since optical fiber is immune to electromagnetic interference, outside factors like radio waves and power transmission lines have no effect on the signals inside the fiber. These advantages make optical fiber sensors ideally suited for use on large steel structures such as bridges, transmission towers, and offshore drilling platforms, as well as a number of other applications.

LIGHT LOSS IN BENT OPTICAL FIBERS

The sensor conceived for the present work is based on a simple amplitude modulation of the light propagating in a multimode fiber. According to the principles of geometric optics, all light rays in the fiber obey Snell's law, leading to a critical angle for total internal reflection, or $\theta_c = \cos^{-1}(n_2/n_1)$, where n_1 and n_2 are the indices of refraction of the core and cladding, respectively, and θ_c is measured relative to the fiber axis. Typically rays propagating at small angles with respect to the fiber axis are referred to collectively as low order modes, whereas guided modes propagating at or near the critical angle are called high order modes. Rays travelling at an angle greater than θ_c are refracted at the core-cladding interface rather than totally internally reflected, meaning that a portion of the energy escapes the core, and generally then escapes the fiber altogether.

When guided light reaches a bend in the fiber, geometry implies that some of the rays will no longer strike the core-cladding boundary within the critical angle, leading to a light loss. The actual amount of energy which departs from the core is dictated by the Fresnel equations, and depends of the angle of incidence at the boundary, the indices of refraction of the core and cladding, the polarization state of the ray, and the degree and direction of the curvature of the fiber at the point of incidence (Badar, Maclean, Gazey, Miller, and Ghafoori-Shiraz, 1989). Mode conversion also takes place at a bend in addition to light loss. That is, light which was originally propagating as a low order mode may reflect from the boundary so as to become a high order mode on the other side of the bend. Likewise, high order modes may be converted to low order status after a bend. The exact nature of mode coupling in bent fibers is complicated, and despite many theories, is not yet entirely well understood.

In addition to geometric optics, light propagation in ordinary optical fibers may be described using wave optics. In this view, modes are given as solutions to the Helmholtz wave equation formulated in cylindrical coordinates with boundary conditions applied at the core-cladding

interface. Furthermore, the total number of modes which may possibly propagate in a fiber can be approximated by

$$N_{max} \approx \frac{2\alpha}{\alpha+2}\left(\frac{\pi a}{\tau}\right)^{2}(n_1^2-n_2^2),$$

where a is the radius of the fiber, τ is the wavelength of the light source, and α is the refractive index profile parameter, having a value of infinity for step index fibers and two for typical graded index fibers (Gloge, 1972). For step index fibers, equation (1) may be simplified to Nmax ≈ $V^2/2$, where V is the normalized frequency of the fiber, given by

$$V = \frac{2\pi}{\tau} a\sqrt{n_1^2-n_2^2}$$

This approximated expression for Nmax has been confirmed using a direct numerical method which involves counting the zero crossings of the J Bessel functions and computing the propagation constants associated with each mode (Bennett and Chen).

A theory put forward by Gloge (1972) suggests that as the fiber is bent to a radius of curvature R, the effective number of modes which will survive the bend is given by

$$N_{eff} = N_{max}\left[1 - \frac{(\alpha+2)\,n_1^2}{\alpha\,(n_1^2-n_2^2)}\left[\frac{2a}{R}+\left(\frac{3\tau}{4\pi n_2 R}\right)^{2/3}\right]\right].$$

A plot of this equation appears in Figure 2, where the fiber parameters correspond to the Spectran ACU MC100C 100/140 graded index fiber used for our sensor. The figure shows that few modes are lost from the fiber when the bending radius is large, but that below a radius of curvature of around 5 cm, the losses begin to accumulate, and for radii in the range of 1 cm, the loss of modes is acute.

What is not shown in Figure 2 is the amount of power which is lost as a fiber is bent. that is, knowing the number of modes is quite different from knowing the amount of power as a function of bend in a fiber. Although for some applications it can be assumed that each mode carries the same amount of power, in most cases this approximation is incorrect, sometimes significantly so. In fact, each mode carries a discrete amount of power which depends on launch conditions and fiber properties. If that amount of power were known, it would be possible to calculate the losses incurred by a device. Unfortunately, it is ordinarily quite difficult to determine the modal power distribution in a fiber.

A method for determining the approximate distribution of the power among the propagating modes was devised separately based on the equations above, and was reported by Bennett and Gencel (1995). The results show it can generally be assumed that the lower order modes carry more power than the higher order modes, and that reasonable accuracy results when a Gaussian power distribution is assumed. In that case, the normalized power versus radius of curvature plot takes on the same general shape as in Figure 2, with a greater slope at small values of R. Thus is should be able to construct a sensing device based on monitoring two or more states of fiber curvature, and expect a reasonably large change in output between the states.

In order to observe the power loss due to bending in the fiber, a series of experiments was conducted on both step and graded index fibers, unjacketed, again as reported by Bennett and Gencel (1995). Laser light was launched into a 100 meter spool of fiber, and came to numerical aperture equilibrium after some 50 meters (Keiser, 1991). Near the far end, the fiber was bent tightly around a stepped rod of known diameters ranging from 2 to 20 mm. Light emerging from the fiber was collected by a photodetector and read with a voltmeter.

The results for the fiber used in our corrosion sensor are plotted in Figure 3 as normalized power versus diameter of the bend in millimeters. Both a half loop and a full loop bend were investigated, as it was hypothesized that a full loop would incur a larger loss in the fiber, and a greater loss would make for easier monitoring of the sensor. (It is interesting to note that neither Gloge's theory, nor indeed any theory of which we are aware accounts for the number of degrees of bend, though both intuition and experiment confirm its importance in determining loss.) The normalized value of 1.0 on the plot corresponds to straight fiber with no loss in intensity. As expected, a smaller bending diameter led to a greater loss in intensity; at a bend of 3 mm, a loss of roughly 10% of the intensity is detected. The fibers tended to break at a diameter of 2.5 mm. We also see that bending the fiber into a full loop resulted in a greater loss than did a fiber bent into a half loop at the same diameter.

SENSOR DEVELOPMENT

Based on the experimental results above, a sensor with the general design shown in Figure 4 was proposed. The fiber is turned into a tight loop or bend with a diameter of a few millimeters and held by a "corrosion fuse." The fuse is metallic ring or wire, or in later designs a strip, that is made of a corrodible material. As discussed above, when light reaches the bend, a decrease in intensity is observed due to refraction. As discussed above, when light reaches the bend, a decrease in intensity is observed due to refraction. After being exposed to the environment for a period of time, corrosion begins to attack the fuse, eventually degrading it catastrophically: then the fuse breaks, the fiber will relax from the bent position and straighten, and the resulting increase in intensity is detected. The amount of corrosion that it takes to cause the fuse to break will be equivalent to a certain amount of corrosion in the structure being monitored. We note that the fuse need not be made of the same material as the structure, since corrosion rates for materials of different geometry are different. That is, regardless of the fuse material and geometry, studies must be conducted to correlate the rate of corrosion in the sensor to that in the bulk material under investigation. Most likely then, the fuse should be made to corrode faster than the actual structure, so that ample warning time is given.

After our sensor concept was proven experimentally, a number of designs were considered. Figure 4 show a schematic for the current design of the sensor. The 100/140 graded index fiber is fed into a plastic and aluminum enclosure which protects that active elements of the sensor, and left unjacketed within the sensor. A spring is fitted over the fiber and hooked by a pin, to which another spring is connected. This mechanism is then attached to the corrosion fuse via a smaller pin. The fuse, which here is a short length of carbon steel approximately 3 mils thick, is fed to the outside and fixed to the top of the sensor. As the upper spring is pulled up into compression, the fiber enclosed by the lower spring is forced into a bend, where the bend radius (and therefore light loss) is adjustable using the retaining blocks. The fuse is the only active part of the sensor that should be exposed to the outside environment; locations where corrosion was not desired were coated with nail polish which effectively protected them. When left exposed, corrosion attacks the fuse and eventually causes it to break. In this case, the spring is no longer held in

compression and releases, which in turn forces the bent fiber out of the blocks and back to the straight position.

We had hoped the sensor would not need to be so complicated. However, initial tests performed to determine fiber resilience showed that after being bent into a small loop for an extended period, the released fiber never fully rebounded to the straight position. We concluded that most polymer coatings needed for fiber protection would behave similarly, leading to the concern that small changes in curvature before and after release would not lead to discernable changes in intensity. It was found that placing the fiber inside a bent spring would ensure that when the fuse corroded through, the fiber would return to the straight position. However, testing with a prototype showed that when the fiber was bent tight enough to detect an intensity change, friction against the retaining blocks occasionally prevented the spring from releasing when the fuse broke. To ensure a positive sensor response, the upper pin and spring mechanism were added. Also, incorporating a full loop bend into the sensor proved to have a number of drawbacks, so although it would lead to a greater signal change, we have settled for less losses in return for a simpler and more reliable sensor design. The sensor prototype was tested using both the photodetector/voltmeter combination and an optical time domain reflectometer (OTDR), and a change in intensity was observed by both.

Once the prototype sensor was tested, a second set of experiments was conceived on the promise that multiple sensors could be used on a single length of fiber. As with any in-line sensor interrogated using OTDR, the concern arises that changes in the modal power distribution due to the first sensor could lead to a misinterpretation of the data at the second sensor. In order to investigate this possibility, the previous experimental system was slightly modified. A second bending post was added to simulate a second sensor on the fiber; both posts had a 4 mm diameter. Readings were taken under several conditions: straight fiber, a bend at point 1, a bend a point 2, and a bend at points 1 and 2 at the same time. We found that there was a noticeable loss after a single bend, as anticipated. However, when the second bend was added, the additional loss was quite small.

On one hand, this result may seem counter-intuitive, as a loss of approximately the same size may be expected after going around a second bend of the same radius as the first. On the other hand, when light reaches a second bend of identical radius, one could reason that all the modes that could be radiated out of the fiber at this bend were already lost at the first bend, leading to a zero loss at the second bend. This explanation is essentially correct, where the small loss observed can be explained by the fact that light is coupled from one mode to another not only at a bend, but along a straight length of fiber as well, reaching a steady or equilibrium state only after hundreds of meters or more (Snyder and Love, 1983).

The effect of modal power coupling was tested with one more experiment. The same setup with two bending rods was used, and the loss was measured as the distance between the two bends varied. It was found that essentially no loss occurs with a separation of less than 5 cm between the bends. With a separation of 5 to 10 cm there is a smaller loss than previously seen, which seems to indicate that it takes at least 10 cm for the light to begin to be redistributed into the modes that were lost around the first bend.

These results led us to believe that our sensor could indeed be used to reliably monitor corrosion at multiple sites. A number of these sensors could be connected in series using a single fiber and interrogated with an OTDR. The fiber could be a continuous length, or several lengths connected together; the only limitation on the latter is that the distance between the fiber bend and the connector be farther than the resolution limit of the OTDR. When straight fiber is monitored with an OTDR, a nearly horizontal line is seen on its plot of backscattered intensity versus distance, as little light is lost in straight fiber. However, when the fiber experiences a sharp bend, as the sensors do, a drop in intensity is seen at the point where the bend is located. With several in-line sensors attached, the output would ideally look like a set of steps going down. When one of the fuses corrodes and the fiber returns to the straight position, the change would be noticeable as a decrease in intensity which disappears.

CORROSION TEST RESULTS WITH MULTIPLE SENSORS

Three sensors of the design shown in Figure 4 were fabricated and tested in a corrosive environment. The sensors were spaced approximately 10 m apart at one end of a 100 m length of fiber. This orientation allowed for the light propagating through the fiber to have the greatest distance possible to achieve equilibrium in the modes. The devices were then monitored by a Photodyne 5400XQ High Resolution OTDR. A corrosive atmosphere was then prepared by adding a salt solution to a heated tank. The sensors were hung inside the tank by wire cradles, and a thermocouple was added to monitor the interior temperature. Figure 5 shows a printout of the OTDR display used as the reference point from which observations would be made. The experiment was checked daily both visually and by the OTDR.

After 36 days at 30°C and little corrosion, the temperature was increased to 44°C and more light corrosives were added. After another 34 days, the third sensor on the fiber link corroded enough to break. Since the sensor was the last in the line, it probably suffered from having an insufficient amount of power scattered into the lossy modes after the previous bends. Therefore, an increase of only 0.1 dB in intensity was detected, which was small but detectable. A week later, the first sensor on the fiber also broke. A more noticeable change was visible due to this break, as is clear from Figure 6. Approximately 0.25 dB of loss was recovered due to the break. The middle sensor of the experiment broke three days after the second sensor, which resulted in nearly 0.5 dB of gain. It is felt that the experiment successfully demonstrates the potential of our sensor concept.

The sensor as presented so far has some drawbacks and complications. One concern is that each sensor only measures the corrosion at one point in time. That is, when a certain amount of corrosion has occurred, the fuse breaks and a change is detected, but the sensor tells nothing of the corrosion that occurs before or after it reacts. One way to address this problem is to vary the thickness of the corrosion fuse between several sensors to indicate differing amounts of corrosion. Several lines of sensors could be deployed together, with each line using a different fuse thickness. An inspector could check each line with an OTDR and determine at least discrete levels of corrosion at various points along the structure. Although this method would no doubt be effective and an improvement over the "digital" scheme described above, it is recognized that continuous monitoring would greatly enhance the usefulness of this sensor.

Another concern is the physical design of the sensor, which will require some modification to assure a smaller, more reliable, and more economical sensor with the longevity required for field testing. Furthermore, a quantitative analysis of the corrosion in the fuse needs to be performed to standardize the results obtained in future experiments. To achieve this, tests are being planned to monitor the corrosion rates of fuses of different thicknesses, geometry, and composition, and compare them to published rates. Such a correlation will greatly improve the usefulness of these sensors, especially as they are eventually tested on structures in the field.

CONCLUSIONS

In this paper we have demonstrated a simple and cost effective fiber optic sensor which could be used to monitor corrosion in a sample material at a remote location. The greater part of this work

consisted of background tests proving the feasibility of this type of corrosion monitoring device. Data collected from bending experiments proved that noticeable losses can be detected when a multimode fiber is bent, and determined useful bending diameters. Additional tests showed that the detection of loss is also feasible when multiple in-line sensors are interrogated by an OTDR, though some uncertainty remains as to what modal power coupling takes place at different radii bends, and how it would affect an analog measurement of the degree of corrosion. The experiments helped in the construction of several prototype sensor designs, most of which were tested both mechanically and in an environmental tank containing a corrosive atmosphere. The results obtained to date all seem to point to the practicality of this corrosion sensor.

ACKNOWLEDGEMENTS

This work was supported in part by the ATLSS Center, an NSF-Engineering Research Center at Lehigh University. The authors are grateful to J. Kayser who offered a great deal of advice on corrosion in structures and also assisted in preparing the corrosion test, and R. Granata who advised on the effects of corrosion, especially in the area of remote monitoring techniques. We would also like to thank A. Gencel for invaluable collaboration, and the staff of the Lafayette College Engineering Machine Shop for assistance in manufacturing the sensors.

REFERENCES

Badar, A.H., Maclean, T.S.M., Gazey, B.K., Miller, J.F., and Ghafoori-Shiraz, H.,"Radiation from circular bends in multimode and single-mode optical fibres," IEE Proceedings, Vol. 136, pt.J, June 1989.

Bennett, K.D. and Chen, R.Z.,"Numerical Determination of the Number of Allowable Modes in Weakly Guiding Optical Fibers," to be published.

Bennett, K.D. and Gencel, A.U.,"Investigation of modal power distribution in multimode fibers used in multiple in-line sensors," Proc. 1995 N. American Conference on Smart Structures and Materials, Vol. 2444 (San Diego, CA), February 1995.

Fontana, M.G.,Corrosion Engineering, McGraw-Hill Book Co., New York, 1986, p. 39-40.

Gloge, D., "Bending loss in multimode fibers with graded and ungraded core index," Applied Optics, Vol. 11, pp. 2506-2512, November 1972.

Granata, R., Wilson, J.C., Jantz, J., Ettelman, A., and Wildt, R., "Service Application of the Corrosion Coulometer," Eighth Annual Report to the National Science Foundation, February 15, 1994.

Keiser, G., Optical Fiber Communications, McGraw-Hill Book Co., New York, 1991, p. 203.

Komp, M.E., "Atmospheric Corrosion Ratings of Weathering Steel - Calculations and Significance," Materials Performance, Vol. 26, No. 7, July 1987, pp. 42-44.

Snyder, A.W. and Love, J.D., Optical Waveguide Theory, Chapman and Hall, London, 1983.

Snyder, A.W. and Love, J.D., "Reflection at a Curved Dielectric Interface - Electromagnetic Tunneling," IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-23, No.1, 1975.

White, K.R., Minor, J. and Derucher, K.N., Bridge Maintenance, Inspection and Evaluation, Marcel Dekker, Inc., New York, 1992, p. 1.

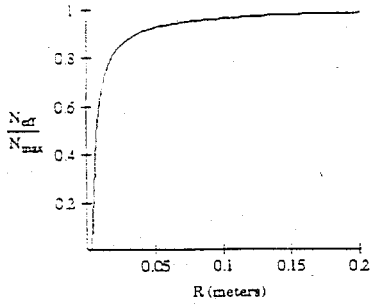


Figure 1. The number of modes which survive a bend of radius R is given by Neff. Neff is normalized by N_{max}, the total number of propagation modes possible in the fiber.

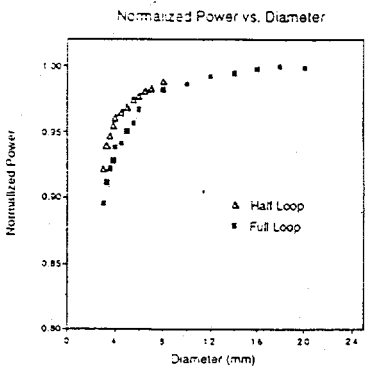


Figure 2. Normalized power received at the fiber output end versus the bending rod diameter for a 180° and a 360° bend.

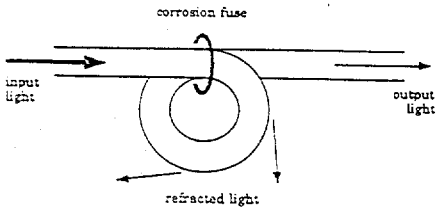


Figure 3. Basic concept for the fiber optic corrosion sensor. When corrosion attack breaks the fuse, the fiber straightens, increasing the light at the output.

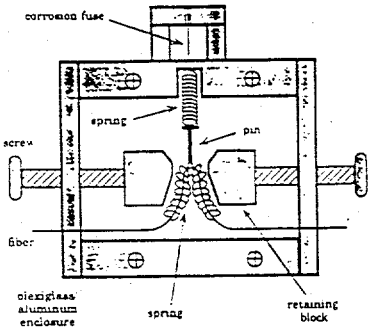


Figure 4. Current design of the fiber optic corrosion monitor.

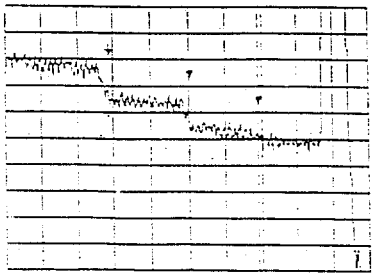


Figure 5. OTDR output from the sensors in the artificial atmosphere tank after being set.

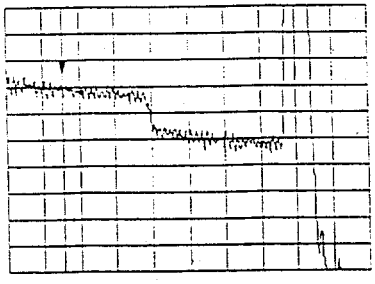


Figure 6. OTDR output showing the intensity gain of 0.2 db after the break at the first sensor in the tank.

A NOVEL SOLUTION OF LINEAR CONGRUENCES

Jeffrey F. Gold

Department of Mathematics
Department of Physics
University of Utah
Salt Lake City, Utah 84112

Faculty Advisor: Don H. Tucker

Department of Mathematics
University of Utah
Salt Lake City, Utah 84112

INTRODUCTION

Although the solutions of linear congruences have been of interest for a very long time, they still remain somewhat pedagogically difficult. Because of the importance of linear congruences in fields such as public-key cryptosystems, new and innovative approaches are needed both to attract interest and to make them more accessible. While the potential for new ideas used in future research is difficult to assess, some use may be found here.

In this paper, the authors make use of the remodulization method developed in [1] as a vehicle to characterize the conditions under which solutions exist and then determine the solution space. The method is more efficient than those cited in the standard references. This novel approach relates the solution space of $cx \equiv a \bmod b$ to the Euler totient function for c rather than that of b, which allows one to develop an alternative and somewhat more efficient approach to the problem of creating enciphering and deciphering keys in public-key cryptosystems.

REMODULIZATION

DEFINITION 1 *If a and b are integers, then*

$$a \bmod b = \{a, a \pm b, a \pm 2b, \ldots\} .$$

The notation $x \equiv a \ mod \ b$, means that x is an element of the set $a \ mod \ b$. The common terminology is to say that x is congruent to a *modulo* b. These sets are also frequently called residue classes since they consist of those integers which, upon division by b, leave a remainder (residue) of a. It is customary to write a as the least non-negative residue.

DEFINITION 2 *If $a_1, a_2, \ldots, a_n, b \in \mathbb{Z}$, then*

$$[a_1, a_2, \ldots, a_n] \bmod b = \{a_1 \bmod b\} \cup \{a_2 \bmod b\} \cup \cdots \cup \{a_n \bmod b\} = \bigcup_{i=1}^{n} \{a_i \bmod b\} .$$

THEOREM 1 *Suppose a, b, and $c \in \mathbb{Z}$ and $c > 0$, then*

$$a \bmod b = [a, a+b, \ldots, a+b(c-1)] \bmod cb .$$

Proof. Write

$$\begin{array}{rcl} a \bmod b & = & \{\ \ldots \quad a-cb, \quad a-(c-1)b, \quad \ldots \quad a-b, \\ & & \quad\ \ a, \qquad a+b, \qquad \ldots \quad a+(c-1)b, \\ & & \quad\ \ a+cb, \quad a+(c+1)b, \quad \ldots \quad a+(2c-1)b, \quad \ldots \ \} \end{array}$$

and upon rewriting the columns,

$$\begin{array}{rcl} a \bmod b & = & \{\ \ldots \quad a-cb, \quad a+b-cb, \quad \ldots \quad a+(c-1)b-cb, \\ & & \quad\ \ a, \qquad a+b, \qquad \ldots \quad a+(c-1)b, \\ & & \quad\ \ a+cb, \quad a+b+cb, \quad \ldots \quad a+(c-1)b+cb, \quad \ldots \ \} \end{array}$$

and forming unions on the extended columns, the result follows.

This process is called remodulization by the factor c.

LINEAR CONGRUENCES

THEOREM 2 *A linear congruence* $cx \equiv a \bmod b$, *where* $gcd(c, b) = 1$, *has as unique solution* $x \equiv a_0 \bmod b$, *where* $a_0 \in \{\frac{a+bk}{c}\}_{k=0}^{c-1}$.

Proof. Suppose one has the linear congruence,

$$cx \equiv a \bmod b \ .$$

where $gcd(c, b) = 1$ and $0 < c < b$. (If c does not satisfy this requirement, then c may be reduced or augmented by some multiple of b so that it satisfies the condition $0 < c < b$.)

Remodulizing $a \bmod b$ by the factor c gives

$$cx \equiv [a, a + b, \ldots, a + b(c - 1)] \bmod cb \ .$$

Because the set $\{a, a + b, \ldots, a + b(c - 1)\}$ forms a complete residue system *modulo* c, there exists an element in this set, call it d, which is divisible by c. Since

$$cx \equiv [a, a + b, \ldots, d, \ldots, a + b(c - 1)] \bmod cb \ ,$$

it is seen that the only solvable linear congruence is

$$cx \equiv d \bmod cb \ .$$

The remaining linear congruences,

$$cx \equiv [a, a + b, \ldots, d - b, d + b, \ldots, a + b(c - 1)] \bmod cb$$

are not solvable, since in each case the factor c is pairwise relatively prime with the residues $\{a, a + b, \ldots, d - b, d + b, \ldots, a + b(c - 1)\}$, and thus does not divide them.

For the solution $cx \equiv d \bmod cb$, however, dividing through by the factor c,

$$\frac{cx}{c} \equiv \frac{d}{c} \bmod \frac{cb}{c}$$

or,

$$x \equiv \frac{d}{c} \bmod b \ .$$

Note that the Euclidean algorithm has not been invoked; all that was necessary to solve this problem was the fact that $gcd(c, b) = 1$. The theorem is illustrated by the following example.

EXAMPLE 1 *Suppose* $12x \equiv 3$ *mod* 7; *this reduces to* $5x \equiv 3$ *mod* 7. *This linear congruence is solvable since* 3 *is divisible by* $gcd(5, 7) = 1$. *Remodulizing* 3 mod 7 *by the factor* 5 *gives*

$$5x \equiv [3, 10, 17, 24, 31] \bmod 5 \cdot 7$$

so that

$$5x \equiv 10 \bmod 35$$

is the only possible solution and, upon dividing all three terms by 5,

$$x \equiv 2 \bmod 7 \ .$$

Note that the remaining linear congruences $5x \equiv [3, 17, 24, 31] \bmod 35$ *do not admit any solutions, since in this example* $gcd(5, 35) = 5$ *does not divide any element in the set* $\{3, 17, 24, 31\}$.

THEOREM 3 *If* $gcd(c, b) = d$ *and* $d|a$, *then the linear congruence* $cx \equiv a \bmod b$, *has* d *distinct (incongruent) solutions modulo* b.

Proof. In the event $gcd(c, b) = d$, then a must be divisible by d, otherwise, the linear congruence will not admit integer solutions. With that in mind, write $c = c_0 d$, $a = a_0 d$, and $b = b_0 d$. If all three terms of the original linear congruence are divided by d,

$$c_0 x \equiv a_0 \bmod b_0 \ .$$

Since $gcd(c_0, b_0) = 1$, the resulting linear congruence has a solution $x \equiv x_0 \ mod \ b_0$. However, the modulus of the original congruence is $b = b_0 d$; therefore, by remodulizing the solution $x_0 \ mod \ b_0$ by the factor d one obtains

$$x \equiv [x_0, x_0 + b_0, \ldots, x_0 + b_0(d-1)] \bmod b_0 d \ .$$

Hence there are d distinct (incongruent) solutions *modulo* b to the linear congruence $cx \equiv a \ mod \ b$ if $gcd(c, b) = d$ and $d|a$. The theorem's utility is demonstrated by the following:

EXAMPLE 2 *Suppose* $6x \equiv 9 \bmod 15$. *Dividing through by the common factor 3,* $2x \equiv 3 \bmod 5$. *This new linear congruence is solvable because 3 is divisible by* $gcd(2, 5) = 1$. *Using the remodulization method,* $2x \equiv [3, 8] \bmod 10$, *where the solution, by inspection, is* $x \equiv 4 \bmod 5$. *Then, remodulizing* $4 \bmod 5$ *by the factor 3, the solutions of the original linear congruence* $6x \equiv 9 \bmod 15$ *are* $x \equiv [4, 9, 14] \bmod 15$.

It is easily seen that the remodulization method is a trial-and-error method; however, after the solution is found, it is unnecessary to carry on any further computations. Another trial-and-error method consists of trying all residues of the complete residue system $[1, 2, \ldots, b] \bmod b$ in the linear congruence $cx \equiv a \bmod b$ until the solution is found. In the case $c \ll b$, there are at most c computations using the remodulization method, compared to b possible computations of the alternate method.

EXAMPLE 3 *Consider the linear congruence* $3x \equiv 5 \bmod 37$. *The remodulization method requires at most 3 steps, compared to 37 possible steps trying solutions of the complete residue system modulo 37. Remodulizing by the factor 3,* $3x \equiv [5, 42, \ldots] \bmod 111$. *By inspection, and requiring only 2 steps, the solution is* $x \equiv 14 \bmod 37$. *Performing the other calculation would have required 14 steps. Of course, simply guessing the solution may sometimes be just as fruitful. Picking an easy example is also helpful.*

A standard method of solving linear congruences involves Euler's phi function [2,3], or totient, denoted by Φ. The totient $\Phi(b)$ enumerates the positive integers less than b which are relatively prime to b. Euler's extension of Fermat's theorem states that

$$c^{\Phi(b)} \equiv 1 \bmod b \ ,$$

if $gcd(c, b) = 1$. Therefore, multiplying the linear congruence $cx \equiv a \bmod b$ through by the factor $c^{(\Phi(b)-1)}$ gives

$$c^{\Phi(b)} x \equiv a \cdot c^{(\Phi(b)-1)} \bmod b \ ,$$

or

$$x \equiv a \cdot c^{(\Phi(b)-1)} \bmod b \ .$$

Thus, finding the solution of the linear congruence $cx \equiv a \bmod b$ requires knowing $\Phi(b)$, or equivalently, the factorization of b.

The remodulization method predicts finding solutions of linear congruences based on the factor c, specifically $\Phi(c)$, rather than the modulus b. In cases dealing with very large integers, and where c is much less than b, or those cases in which the factorization of c is known, it may be more convenient to calculate the totient of c, rather than that of b.

THEOREM 4 *The linear congruence* $cx \equiv a \bmod b$, *where* $gcd(c, b) = 1$, *has as solution*

$$x \equiv \left[\frac{a(1 - b^{\Phi(c)})}{c} \right] \bmod b \ .$$

Proof. Note that the linear congruence $cx \equiv a \bmod b$, where c and b are relatively prime and $0 < c < b$, implies the existence of integers x and y such that $cx - by = a$. Solving this equation instead for y, which is equivalent to the linear congruence $by \equiv -a \bmod c$, shows that the solution, using Euler's theorem, is $y \equiv -a \cdot b^{(\Phi(c)-1)} \ mod \ c$. Substituting this result into $cx - by = a$, $cx = a + by = a + b[-a \cdot b^{(\Phi(c)-1)} \ mod \ c]$. Solving for x,

$$x \equiv \left[\frac{a + b[-a \cdot b^{(\Phi(c)-1)} \ mod \ c]}{c} \right] \bmod b \ .$$

where $-a \cdot b^{(\Phi(c)-1)}$ is augmented by the proper multiple of c to obtain the least non-negative residue *modulo* c.

In the remodulization method, the elements $\{a, a+b, \ldots, a+b(c-1)\}$ are generated by $a+by$, for $y \in \{0,1,2,\ldots,c-1\}$. The $y+1^{st}$ residue in the remodulized form $[a, a+b, \ldots, a+b(c-1)]$ mod bc is the solution, upon division by c.

If one is not interested in finding the least non-negative residue, the solution reduces to

$$x \equiv \left[\frac{a(1-b^{\Phi(c)})}{c}\right] \bmod b \ .$$

Theorem 3 gives the obvious corollary to Theorem 4 in case $gcd(c,b) = d$.

COROLLARY *If $gcd(c,b) = d$ and $d|a$, then the linear congruence $cx \equiv a \bmod b$ has d distinct solutions $x \equiv [x_0, x_0+b_0, \ldots, x_0+b_0(d-1)] \bmod b$, where $a = a_0 d$, $b = b_0 d$, $c = c_0 d$, and*

$$x_0 \equiv \left[\frac{a_0(1-b_0^{\Phi(c_0)})}{c_0}\right] \bmod b_0 \ .$$

REMARK *If one solves the diophantine equation $cx+by = a$; i.e., $cx = a-by = a \bmod b$ formally, then the answer is $x = \frac{a}{c} - \frac{b}{c}y$, but the integer character and information is lost and not easily recovered. In the modular arithmetic format, however, the formula of Theorem 4 (or its corollary by Theorem 3) characterizes the countably infinitely many solutions.*

APPLICATION

In public-key cryptosystems [2,4,5], an enciphering modulus m is created by multiplying two very large primes p and q, say $m = pq$; then one chooses an enciphering exponent e and a deciphering exponent d that satisfy the congruence relation

$$e \cdot d \equiv 1 \bmod \Phi(m) \ ,$$

where $gcd(e, \Phi(m)) = gcd(d, \Phi(m)) = 1$, and $\Phi(m) = (p-1)(q-1)$. By large, it is meant that the primes p and q should have 100 or more digits each. If one chooses the enciphering exponent e to be a prime such that $gcd(e, \Phi(m)) = 1$, then it is unnecessary to calculate $\Phi(\Phi(m))$ for the usual or standard solution

$$d \equiv e^{(\Phi(\Phi(m))-1)} \bmod \Phi(m) \ .$$

Instead, one only needs to calculate the solution

$$d \equiv \left[\frac{1-\Phi(m)^{\Phi(e)}}{e}\right] \bmod \Phi(m) \ ,$$

where $\Phi(e) = e-1$.

It is much easier (and more computationally efficient) to satisfy the condition $gcd(e, \Phi(m)) = 1$ than it is to calculate the prime decomposition of $\Phi(m)$ and its totient $\Phi(\Phi(m))$, even in those cases in which e is not prime but its factorization is known.

EXAMPLE 4 *Suppose $m = 7 \cdot 11 = 77$, then $\Phi(77) = 60$. The problem is to find an enciphering exponent e and a deciphering exponent d which satisfy*

$$e \cdot d \equiv 1 \bmod 60 \ .$$

If one chooses $e = 13$, then d is found by

$$d \equiv \left[\frac{1-60^{\Phi(13)}}{13}\right] \bmod 60 \equiv \left[\frac{1-60^{12}}{13}\right] \bmod 60 \equiv 37 \bmod 60 \ ,$$

whereas $\Phi(\Phi(77)) = \Phi(60) = \Phi(2^2 \cdot 3 \cdot 5) = 16$. Additionally, for $e = 7$, $d = 43$; $e = 11$ gives $d = 11$; $e = 17$ gives $d = 53$; and so on.

This method may not supplant the Euclidean algorithm method. In order to extract a solution from the linear congruence $nx \equiv 1 \bmod m$, the Euclidean algorithm requires at most $\log_2(m)$ iterations, or in the case $n \ll m$, only $1+\log_2(n)$ iterations. According to Bressoud [6], the method described here requires approximately the same number of iterations (perhaps one or two fewer), but since one is dealing with very large integers, i.e., $n \sim 10^{100}$ and $m \sim 10^{200}$, the difference is negligible. Therefore, those who have incorporated the Euclidean algorithm in their computer programs will not likely change to this method. Those just starting may well find this method preferable.

REFERENCES

[1] Jeffrey F. Gold and Don H. Tucker, *Remodulization of Congruences*, Proceedings — National Conference on Undergraduate Research, University of North Carolina Press, Asheville, North Carolina, 1992, Vol. II, 1036–41.

[2] David M. Burton, *Elementary Number Theory*, Second Edition, Wm. C. Brown Publishers, Iowa, 1989, 156–160, 175–179.

[3] Oystein Ore, *Number Theory and Its History*, Dover Publications, Inc., New York, 1988, 109–115.

[4] David M. Bressoud, *Factorization and Primality Testing*, Springer-Verlag New York, Inc., New York, 1989, 43–46.

[5] Kenneth H. Rosen, *Elementary Number Theory and Its Applications*, Third Edition, Addison-Wesley Publishing Company, Massachusetts, 1993, 253–264.

[6] David M. Bressoud. Personal communication.

Jeffrey F. Gold: *gold@math.utah.edu http://www.physics.utah.edu/~jgold/home.html*
Don H. Tucker: *tucker@math.utah.edu*

The Relative Efficiencies of One-Hot and Binary Encodings for FPGA-based Finite State Machines

Vladimir K. Jurista
School of Engineering, Grand Valley State University
301 W. Fulton, Suite 618 Grand Rapids, MI 49504

Faculty Advisor: William A. Chren, Jr.

Experimental comparison of circuit area and critical path length for the one-hot and binary state encoding methods for FPGA-based finite state machines has been performed. Twenty-five machines from the MCNC FSM Benchmark Suite were implemented in three widely-used types of FPGAs. The results indicate that the one-hot method yields significantly smaller and faster implementations only for machines with output pipelining and nine or more states.

INTRODUCTION

The recent availability of register-rich programmable logic devices such as Field Programmable Gate Arrays (FPGAs) has increased the use of "non-binary" methods of state encoding. Alternate schemes such as "k-hot" encoding [1], in which the machine state is represented by exactly k asserted bits in the state register, are being used with increasing frequency. In particular, one-hot encoding (OHE) [2] is especially popular because it is widely believed to allow faster clock speeds when used with register-rich programmable devices such as FPGAs. The justification for the belief is that the simplified state representation allows fewer logic levels in the state feedback logic.

This paper presents the results of research to ascertain the validity of this belief. Such a study is important because of the widespread use of FPGAs for finite state machine (FSM) implementation. A related investigation has been performed into the area-optimal state assignment in Programmable Logic Array (PLA)-based VLSI implementations [3]. However, it is not clear how the results of that study can be applied to FPGAs. Furthermore, that research did not consider the effects of the state encoding on the critical path length. Other related research includes analytical results concerning the *logic*-minimal (viz., non-monolithic and non-programmable) binary state assignment [4], [5].

This study involved the comparison of binary and one-hot encodings of 25 FSMs from the Microelectronics Center of North Carolina (MCNC) FSM Benchmark Suite. The machines were implemented in three types of devices with substantially different architectures. The results indicate that clock speed can be increased by OHE only for "larger" machines (at least nine states) which use output pipelining. Furthermore, a suitably defined "complexity" measure (viz., area-delay product) is smaller for machines using OHE.

FPGA DEVICE ARCHITECTURES AND THE MCNC BENCHMARK SUITE

This section contains descriptions of the benchmark FSMs and the architectures of the FPGAs used to implement them. The FPGAs will be referred to as Device Types 1, 2, and 3.

Programmable Logic Overview

Programmable logic refers to a class of digital electronic devices (called PLDs) whose logic function can be modified by the user. The modifications can be permanent or temporary, in which cases the devices are said to be one-time programmable (OTP) or reprogrammable, respectively.

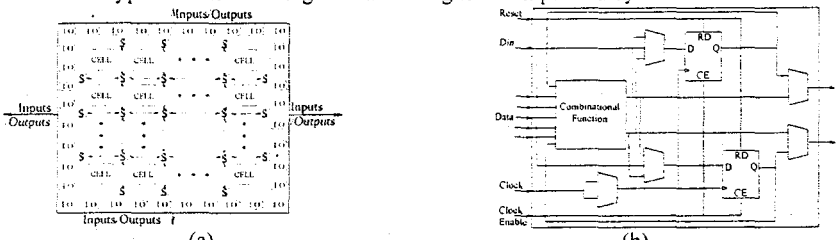
Field Programmable Gate Arrays are a type of PLD which employs a rectangular array of identical, programmable and small logic cells along with programmable wires deployed in a grid pattern. Designs are implemented in the array by using software tools to specify the design (either schematically or textually) and translate the specification to a programming file. The programming file is downloaded to the device so as to both select the function of the logic cells

and activate transistor switches for proper wiring between the cells. Any design (of proper size) can be implemented. Figure 1a shows a typical FPGA architecture.

The logic cells and wiring method vary from vendor to vendor. Some cells, for example those from Xilinx, can implement any logic function with four inputs and one output (registered if needed). Other cells, such as those from Quicklogic, can accommodate up to 14 inputs for a single AND function, likewise with optional registering. Logic cell architectures can range from simple and small AND/Mux structures (e.g., Figure 3b) to relatively complicated and large look-up-table based architectures (e.g., Figure 1b). Strategically located pass transistors or transmission gates affect the programmability. The wiring methods range from "unlimited pure metal" (that is, fast metal interconnections can be placed between any horizontal and vertical wires) to slow, silicon-based (using switch matrices located diagonally with respect to the cells). For this research, three different types of devices with which to implement the FSMs were chosen.

Device Type 1

Device type 1 is shown in Figure 1a. Wiring is accomplished by



(a) (b)

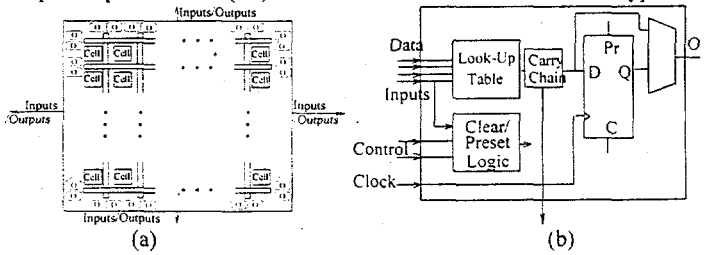
Figure 1: (a) Architecture of Device Type 1, (b) Logic Cell

switch matrices (S), which are switch boxes allowing software reprogrammable connection between lines running horizontally and vertically. The array of logic cells and switch matrices is surrounded by a perimeter of programmable input-output units (I/Os) which contain buffer circuitry for the pins as well as a single D flip-flop for latching data.

The structure of the logic cells consists of a single combinational logic module, two D flip-flops and five multiplexers, as shown in Figure 1b. The logic module employs static RAM to allow reprogrammable implementation of any boolean function with up to five input variables. Alternatively, it can produce the outputs of two independent four variable functions. The results can be registered or passed directly to the cell output. An external data input is also provided and can also be registered. The multiplexers perform the control function. They are configured under software control at initialization and determine, for example, whether the data is to be registered at its source.

Device Type 2

The architecture of device type 2 is shown in Figure 2a, where it can be seen to employ a periphery of input/output elements (Os) whose function is the same as device type 1. The



(a) (b)

Figure 2: (a) Device Type 2 Architecture (b) Logic Cell

wiring method and cells are different, however. There are no switch matrices. Instead, cell interconnection is performed by appropriate selection of dedicated row and column channel wires. Gates are placed by software in locations such that a particular unused path between

output and input can be employed. This method allows higher speed because it avoids the high-resistance S elements in device type 1. The cells are depicted in Figure 2b, where it can be seen that each consists of a four-input look-up table, flip-flop, and high-speed carry and cascade chains. The table computes any four-variable logic function, while the carry and cascade chains offer high-speed local interconnection within neighboring cells. The flip-flop can be programmed to perform D, T, JK, or SR operations.

Device Type 3

Device type 3 has the architecture shown in Figure 3a. The architecture is optimized for

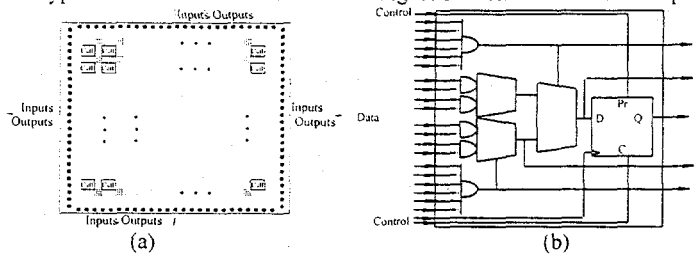


Figure 3: (a) Device Type 3 Architecture (b) Logic Cell

speed because of two features. First, it has one-time programmable wiring resources in which any row wire can be connected to any column wire. This feature is denoted in the figure by the lattice of intersection areas of vertical and horizontal wire bundles. The connection is metal-to-metal and therefore has less resistance and higher speed than switch or gate based methods. The second feature is the simplicity of the logic cells and the input/output elements (denoted by the shaded squares on the periphery of the device). The elements are simpler than for the previous devices because they do not contain latches or flip-flops. The cells are depicted in Figure 3b. Each cell consists of eight AND gates, three 2-to-1 multiplexers and a D flip-flop, configured so that it can perform any boolean function of three input variables, or alternatively, wide-but-shallow functions such as a 14-input AND gate. The delay is small due to the cell's simplicity.

MCNC FSM Benchmark Suite

The MCNC FSM Benchmark Suite is a collection of finite state machines which is widely used to compare the performance of logic synthesis tools. It consists of 53 machines with from 4 to 218 states and with inputs and outputs as large as 27 and 56, respectively. A set of 25 of these machines has been selected for the research (listed in Table 1).

METHODOLOGY

Output pipelining is essential to a full understanding of the results. Figure 4a depicts the

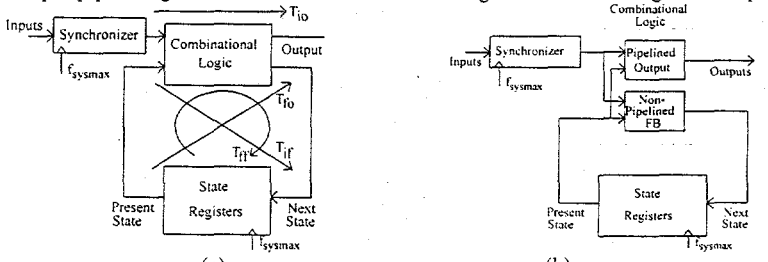


Fig. 4: (a) State Machine Structure (b) Output Pipelined Machine Structure

canonical form of an FSM with an input synchronizer. Four timing parameters are indicated in the figure: T_{io}, defined as the propagation delay from some synchronizer output to some machine output; T_{if}, the propagation delay from some synchronizer output to some state register input; T_{fo}, the propagation delay from some state register output to some machine output, and T_{ff}, the delay from some state register output to some state register input.

The maximum of all possible values of these timing parameters would ordinarily determine the maximum clock frequency f_{sysmax}. However, *for certain types of machines*, the clock speed can be increased by the use of output pipelining, which effectively decouples the output propagation times T_{fo} and T_{io} from the state feedback times T_{ff} and T_{if}. The pipelining is depicted in Figure 4b, where it can be seen that the method requires the partitioning of the Combinational Logic shown in Figure 4a into a pipelined output component and a non-pipelined feedback (FB) component. Such a decomposition allows the machine clock period to be determined only by the maximum of the propagation times T_{ff} and T_{if}. This allows a significant boost in the clock frequency. Of course, the output pipelining will introduce latency in the output signals but in many applications (e.g., digital signal processing), this does not represent a significant problem. In the FPGAs used in this research, the pipelining can be done without increasing the number of logic cells because each cell contains at least one D flip-flop.

In order to assess the relative complexities of the OHE and binary state encoding methods, 25 FSMs were chosen from the MCNC Benchmark Suite and encoded in ABEL HDL. The selected machines are listed in Table 1, where it can be seen that the selected machines have

FSM Name	Inputs	Outputs	Products	States	FSM Name	Inputs	Outputs	Products	States
bbara	4	2	60	10	lion	2	1	11	4
bbtas	2	2	24	6	lion9	2	1	25	9
beecount	3	4	28	7	mc	3	5	10	4
dk14	3	5	56	7	opus	5	6	22	10
dk15	3	5	32	4	pma	8	8	73	24
dk16	2	3	108	27	s27	4	1	34	6
dk17	2	3	32	8	s510	19	7	77	47
dk27	1	2	14	7	shiftreg	1	1	16	8
ex2	2	2	72	19	tav	4	4	49	4
ex3	2	2	36	10	tma	7	6	44	20
ex4	6	9	21	14	train11	2	1	25	11
ex5	2	2	32	9	train4	2	1	14	4
ex7	2	2	36	10					

Table 1: FSMs Used for This Research

numbers of states that range from 4 to 47, and numbers of inputs and outputs that are as large as 19 and 16, respectively. These machines were selected because their size is typical of the majority of FSMs encountered in practice.

The ABEL code was synthesized among the three device types using *Synario*, a commercial retargeting tool. The resulting designs were placed and routed onto a single device of each type, using the vendor-supplied place-and-route kits. By ensuring that the same device was used for implementing each machine, any data effects of inter-device variability and architectural differences were avoided. For each FSM, the total number of logic cells required, as well as the number of cells in the T_{if}, T_{ff}, T_{fo} and T_{io} paths (see Figure 4a) was determined.

Propagation delay statistics in the place-and-route report files allowed the determination of the critical path delays for both the non-pipelined and pipelined versions of the machines. The non-pipelined critical path delay was found as the maximum of T_{io}, T_{if}, T_{ff}, and T_{fo}. The pipelined critical path delay was determined as the greater of T_{if} and T_{ff}. For device types 1 and 3, these delays were measured in numbers of logic cells. For type 2, the delays were measured in units of nanoseconds because the place-and-route tool did not report delays in units of cells. The FPGA area was also measured in numbers of cells. Pipelining had no effect on area measurements because the cells used for combinational logic contained unused data latches.

DISCUSSION OF RESULTS

The results will be presented in terms of the product of the circuit area and critical path length. This product, henceforth referred to as "complexity", has units of logic cells squared (logic cells-nanoseconds) for device types 1 and 3 (2), and is henceforth referred to as "complexity". It is a version of an area-time product, wherein smaller values imply a superior implementation because they mean smaller real-estate usage and/or faster clock speeds.

The arithmetic differences in complexities between the binary and OHE methods, defined as $\Delta C \equiv C_{BIN} - C_{OHE}$ for both the non-pipelined and pipelined cases are plotted in Figures 5a-c. These graphs show that the transition from negative complexity difference, where the binary

complexity is less than that of OHE, to positive difference, where OHE method yields superior results, occurs between eight and nine states, for all three device types. The plots also show that

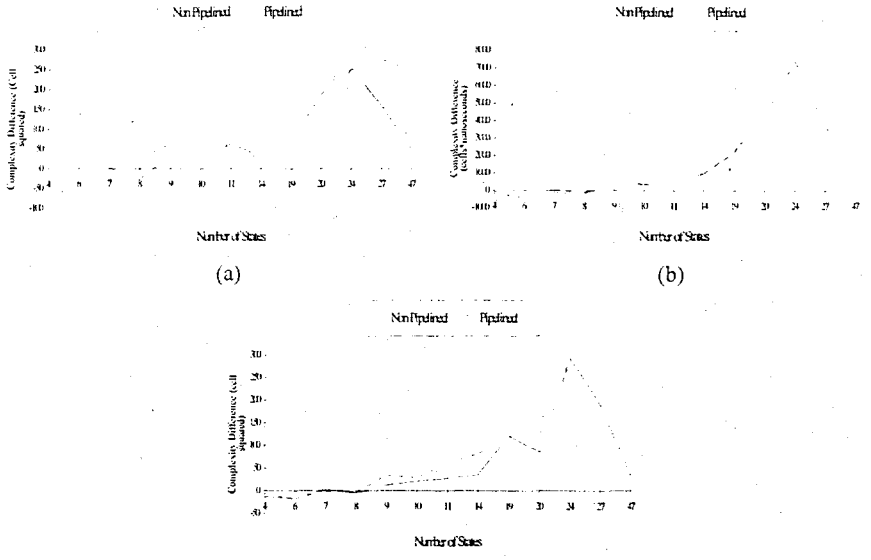


Figure 5: Complexity Difference vs Number of States - Device Types a: 1, b: 2, c: 3

the complexity difference values, compared between pipelined and non-pipelined machines, are considerably different for device types 1 and 3. In most cases for pipelined machines with 9 or more states the binary encoding method has larger complexity than the OHE method.

CONCLUSIONS

From this research three conclusions about the binary and OHE encodings for FPGA-implemented FSMs can be drawn. First, clock speed can be enhanced by OHE only for machines with output pipelining and consisting of at least nine states. Second, the OHE area penalty is incurred only for small machines (viz., less than nine states), regardless of whether output pipelining is used or not. Third, the complexity, defined as the product of the implementation area and the critical path length, of one-hot implementations is smaller than that for binary implementations, for both pipelined and non-pipelined machines. Furthermore, output pipelining enhances the complexity difference between the methods.

ACKNOWLEDGMENT

The author is grateful for the support of the National Aeronautics and Space Administration through grant NAG 8-226 (NASA/University Joint Venture (JOVE) Program) and the Grand Valley State University School of Engineering.

REFERENCES

[1]. D. J. Rosenkrantz, *Half-Hot State Assignments for Finite State Machines*, IEEE Transactions on Computers, vol. 39, no. 5, pp. 700-702, May, 1990.

[2]. Knapp, S.K., *Accelerate FPGA Macros With One-Hot Approach*, The Programmable Gate Array Data Book, Xilinx, Inc., 1992.

[3]. G. De Micheli, R. K. Brayton et. al., *Optimal State Assignment for Finite State Machines*, IEEE Transactions on CAD, vol. CAD-4, no. 3, July 1985, pp. 269-285.

[4]. J. Hartmanis, *On the State Assignment Problem for Sequential Machines 1*, IRE Transactions on Electronic Computers, vol. EC-10, pp. 157-165, June 1961.

[5]. R. E. Stearns and J. Hartmanis, *On the State Assignment Problem for Sequential Machines 2*, IRE Transactions on Electronic Computers, vol. EC-10, pp. 593-603, December 1961.

A.I.D.S. CHOLECYSTOPATHY: A COMPARISON OF ULTRASOUND VS. COMPUTED TOMOGRAPHY IMAGING

Michelle Rhodes and Colleen Ellerick

Department of Diagnostic Ultrasound
Seattle University
Broadway and Madison
Seattle, WA 98122-4460

Faculty Advisor:

Jeff Pope

INTRODUCTION:

With the recognition of AIDS in the USA in 1981 (Rathenberg), a wide range of infectious, inflammatory, and neoplastic diseases for which these patients are at increased risk for have been identified. Many organ systems, especially the biliary system are affected by the HIV virus directly, or indirectly by opportunistic infections. These infections or inflammatory changes of the biliary system can be detected by abdominal ultrasound and computed tomography.

MATERIALS AND METHODS:

Fourteen males and two females(ages 30-51 years, mean age of 39 years) infected with AIDS and presenting with cholecystopathy were retrospectively studied at Harborview Medical Center (Seattle, WA), from a patient population dating back to September 1991. Each patient's medical chart was reviewed to verify the HIV status and exclude any past history of biliary pathology. Other reasons for exclusion included lack of US or CT studies performed within three weeks of each other. The indications for US and CT were clinical and/or biochemical suspicion of hepatobiliary disease.

Anatomical measurements were considered normal or pathological based on the following: the common bile duct was considered dilated if the diameter was 8mm or greater (Cello), using a measuring technique of inner lumen to inner lumen on either sagittal or axial planes. Normal dimensions of the gallbladder were considered to be 4cm or less in the transverse plane. Gallbladder wall thickness was considered abnormal if it measured greater than 3mm on the transverse plane. Striated wall thickening was considered present when the appearance of several irregular, discontinuous, echogenic bands within the gallbladder wall were documented (Rumack). Sludge was considered present when low to mid-level echoes were found within the gallbladder and were found to be gravity dependent after repositioning the patient.

An off-line computer system was used for statistical analysis. We compared the findings of the abdominal ultrasound to the abdominal computed tomography results in documenting gallbladder pathology with the paired t-test, and a one way analysis of variance (ANOVA).

RESULTS:

Sixteen patients were reviewed. Abdominal pain (88%), fever (56%), and nausea and vomiting (31%), were the most

common presenting symptoms. No patients presented with a positive Murphy's sign, and none of the patients had inflammatory bowel disease. Some patients had preexisting abdominal AIDS-related infections or neoplasms that had been documented before the biliary related radiologic exams took place. These infections/neoplasms included cytomegalovirus (n=9), Kaposi's sarcoma (n=2), cryptosporidiosis (n=4), mycobacterium avium-intracellulare (n=7), or non-Hodgkins lymphoma (n=3).The ultrasonographic and computed tomography findings for each of the sixteen patients are summarized in Graph 1 and Table 1.

In this patient population studied, a large and varied spectrum of cholecystopathy was exhibited on US and CT. A one way analysis of variance was used to determine whether US was better than CT in detecting biliary pathology or vice versa. US was found to have a variance of 17.4 while CT was found to have a variance of 8.6. Fourteen study patients had gallbladder wall thickening seen on US, only eight of these were observed by CT. Sludge, polyps, and gallbladder wall striations were undetected on CT, yet evident on US in eight patients. Ultrasound proved to be a more sensitive noninvasive diagnostic modality in detecting evidence of cholecystopathy. It detected a total of thirty five abnormalities of the biliary system versus twenty three by computed tomography. CT proved to be more sensitive in detecting intrahepatic biliary dilatation as it observed three abnormalities while US only found one.

The failure of ultrasound to visualize the biliary system might be attributed to surgical wounds, technique, patient habitus, or intestinal gas in the abdomen, producing acoustic shadowing. Inability of computed tomography to visualize the biliary system could be attributed to poor contrast resolution of the organs. The inability to see polyps or the thick wall of a diseased gallbladder could also be impaired by the lack of spatial resolution. Cases of gallstones were identified on US but not CT, this is likely secondary to limited calcifications.

DISCUSSION:

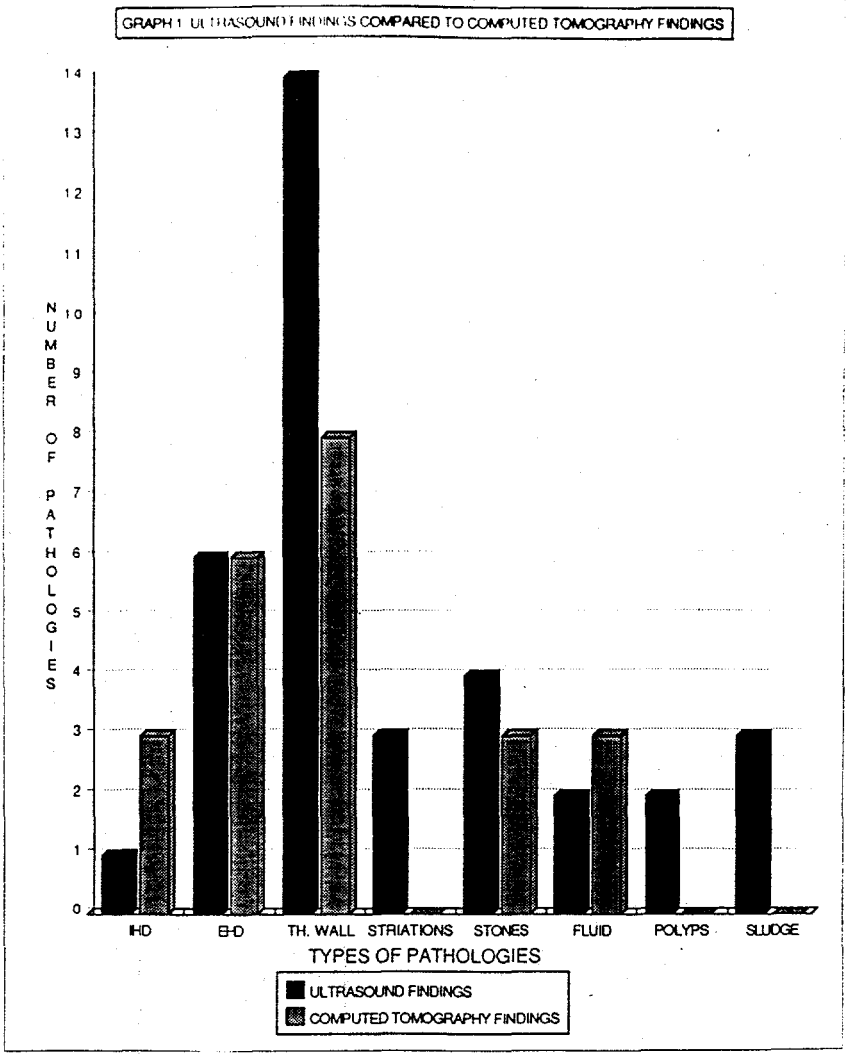
The majority of the radiological features found to be suggestive of AIDS-related cholecystopathy (Graph 1), have been previously described and attributed to AIDS-related infections (McCarty).

Cytomegalovirus has been associated with acute cholangitis in patients with AIDS. The spectrum of CMV in the gallbladder has been less frequently described, but with the increasing number of patients being affected with AIDS, the incidence of AIDS-related cholecystopathy in which CMV is the likely etiology, has grown (Dowsett, Schneiderman, Kavin).

Cryptosporidium Parvum is a coccidial parasite which often causes chronic, life-threatening diarrhea in immunocompromised persons and self-limited diarrhea in immunocompetent persons (Lengerich). It has been suggested that the biliary tree, and in particular the gallbladder, provides a potential reservoir for cryptosporidium. No effective treatment is presently available.

CONCLUSION:

As the AIDS epidemic continues to increase so does our awareness of its diverse manifestations involving the immunocompromised patient. AIDS-related cholecystopathy is becoming more commonly recognized, as our imaging technology improves and our knowledge of AIDS expands. The limited number of patients in this study demonstrated that US was preferable to CT for recognition of significant biliary tract pathology



*IHD-Intrahepatic Dilatation; EHD-Extrahepatic Dilatation; TH-Thickness of Wall

TABLE 1. AIDS CHOLECYSTOPATHY: PATIENT DATA

Number	Age	Symptoms	Ultrasonography Results	Computed Tomography Results	Prior AIDS-related Conditions	Gender
1	45	Pain, Diarrhea,N/V, Wt.Loss	Thick walled GB, Stone	Stone	CMV, MAC	M
2	35	Pain, Fever, Diarrhea, Chills	Thick walled GB, Striations, Fluid	Thick walled GB, Fluid	CMV, MAC	M
3	30	Pain, Fever, N/V, Wt. Loss,Chills	Thick walled GB, EHD	Thick walled GB, IHD and EHD	Microsporidia	M
4	46	Pain, N/V, Diarrhea	Thick walled GB, Fluid, Polyp,EHD	EHD	CS	M
5	30	Pain, N/V	Thick walled GB, EHD	---	MAC	F
6	42	Pain, Fatigue	EHD, Sludge	---	Tuberculosis	M
7	51	----	Thick walled GB, Sludge, Stone	Thick walled GB, Fluid	CS, MAC	F
8	38	Pain, Fever, Dehydration	Thick walled GB, Striations	----	CMV, KS	M
9	38	Pain, Fever, Diarrhea, N/V Wt. Loss	Distended GB, Polyp	Stone	CS, CMV, MAC	M
10	30	Pain, Fever, Diarrhea, Wt. Loss	Thick walled GB, Striations,Stone, Sludge	Thick walled GB	MAC, CMV	M
11	40	Pain, Fever, Dehydration	Thick walled GB, EHD	EHD	CMV	M
12	45	Pain, Fever, Chills, Diarrhea	Thick walled GB, Stone	Thick walled GB, IHD, EHD, Stone	CS, Lymphoma	M
13	34	Pain, Diarrhea, Nausea	Thick walled GB, Mild IHD	Thick walled GB, Mild IHD	CMV, Lymphoma	M
14	39	Pain, Fever	Thick walled GB	----	MAC	M
15	39	Fever	Thick walled GB	Thick walled GB, EHD	CMV, KS, Lymphoma	M
16	42	Pain, Diarrhea, Dehydration	Thick walled GB, EHD	Thick walled GB, EHD, Fluid	CMV	M

*CMV-Cytomegalovirus, MAC- Mycobacterium Avium Complex, CS-Cryptosporidiosis, KS-Kaposi Sarcoma
**EHD-Extrahepatic ductal dilatation, IHD-Intrahepatic ductal dilitation, Wt. Loss-Weight Loss, GB- Gall Bladder, Fluid- Pericholecystic Fluid

among its HIV positive patients. A larger patient population must be studied to draw a definitive decision of the superiority of US to CT when recognizing AIDS cholecystopathy.

REFERENCES:

Agha F, Nostrant T, Abrams G et al: Cytomegalovirus Cholangitis in a Homosexual Man with Acquired Immune Deficiency Syndrome. Am J Gastro 1986;1068 1072.

Benhamou M, Caumes E, Gerosa Y et al: AIDS-Related Cholangiopathy, Critical Analysis of a Prospective Series of 26 Patients. Dig Dis Sci 1993;38:1113-1117.

Blumberg R, Kelsey P, Perrone T et al: Cytomegalovirus and Cryptosporidium Associated Acalculus Gangrenous Cholecystitis. Am J Med 1984;76:1118-1122.

Cello J: Acquired Immunodeficiency Syndrome Cholangiopathy:Spectrum of Disease. Am J Med 1989;86:539-546.

Craig Marveen: Pocket Guide to Ultrasound Measurement.Philadelphia, J.B. Lippincott, 1988;4,5,13.

Dolmatch B, Laing F, Federle M et al: AIDS Related Cholangitis: Radiologic Findings in Nine Patients. Radiology 1987; 163:3 13-316.

Dowsett J, Miller R, Davidson R et al: Sclerosing Cholangitis in Acquired Immunodeficiency Syndrome. Scand J Gastroenterol 1988;23: 1267-1274.

Jacobson M, Cello J, Sande M: Cholestasis and Disseminated Cytomegalovirus Disease in Patients with the Acquired Immunodeficiency Syndrome. Am J Med 1988;84:218-224.

Kavin H, Jonas R, Chowdry L et al: Acalculus Cholecystitis and Cytomegalovirus Infection in the Acquired Immunodeficiency Syndrome. Ann Int Med 1986;104:53,54.

Lengerich I, Addiss D, Marx J et al: Increased Exposure to Cryptosporidiosis Among Dairy Farmers in Wisconsin. J Infect Dis 1993;167:1252-1255.

Margulis S, Honig C, Soave R et al: Biliary Tract Obstruction in the Acquired Immunodeficiency Syndrome. Ann Int Med 1986;105:207-210.

McCarty M, Choudhry A, Helbert M et al: Radiologic Features of AIDS -Related Cholangitis1989;40:582-585.

Rathanberg R, Woelf M, Stoneberner R et al: Survival with the Acquired Immunodeficiency Syndrome, N Engl J Med 1987;317:1297-1302.

Rumack C, Wilson S, Charboneau W: Diagnostic Ultrasound (Vol 1), St. Louis, Mosby,1991;126,131.

Schneiderman D, Cello J, Laing F: Papillary Stenosis and Sclerosing Cholangitis in the Acquired Immunodeficiency Syndrome. Ann Int Med 1987;106:546-549.

DETERMINING PULMONARY CAPILLARY WEDGE PRESSURE WITH PULSED WAVE DOPPLER ECHOCARDIOGRAPHY

Jane M. Nova

Department of Diagnostic Ultrasound
Seattle University
Broadway and Madison
Seattle, WA 98122

Faculty Advisor:

Suzanne Adams

To determine the validity of transthoracic and transesophageal Doppler echocardiography to evaluate pulmonary capillary wedge pressure (PCWP), this author studied intensive care unit (ICU) and operating room (OR) patients. These patients were in normal sinus rhythm or atrial fibrillation. Patients could not be on positive end-expiratory pressure (PEEP) if on a ventilator. Swan-Ganz catheters were utilized to obtain PCWP. The researcher was blinded to the catheter results until Doppler results were analyzed and recorded. Doppler echocardiography was utilized to obtain; 1) time-velocity integrals (TVI) of the "a" and "e" waves of transmitral flow at the level of the mitral valve annulus (end-expiration), 2) TVI of the "x" and "y" waves of pulmonary vein flow, and 3) Q-wave (electrocardiogram) to mitral closure interval and isovolumic relaxation time (IVRT). The correlation was calculated by excluding the outlying data points. The TVI of the "a" and "e" waves of transmitral flow after regression analysis demonstrated an r-value of 0.30. The TVI of the "x" and "y" waves of pulmonary vein flow after regression analysis demonstrated an r-value of -0.086. There were two equations calculated to determine PCWP utilizing Q--mitral closure and IVRT. After regression analysis these equations demonstrated an r-value of -0.079 and -0.077. Correlations demonstrated by the equations derived from the data gathered in this research were r= 0.66 (A/E ROI), r=0.37 (pulmonary vein flow), r=.34 (Q--MC/IVRT), r=.34 (Q--MC/IVRT). The highest correlation obtained was derived from utilizing the first equation by calculating the A/E ROI of transmitral flow. Therefore, the information derived from this equation would be the most accurate to use in the clinical setting. Currently, there is not a standard method to measure PCWP with Doppler echocardiography. *Key words: Doppler echocardiography, pulmonary capillary wedge pressure.*

Pulmonary capillary wedge pressure (PCWP) is measured by direct cannulation of the internal jugular or subclavian vein using a Swan-Ganz catheter. Intensive care units commonly consist of patients presenting with hypotension, hypoxia, and pulmonary congestion, along with multiple cardiovascular abnormalities. Swan-Ganz catheters are currently utilized for prolonged monitoring of these patients.

Normal PCWP should be no greater than 15mmHg. In infarct-related heart failure the PCWP will typically measure in excess of 25mmHg. In hypovolumic shock and septic shock the PCWP is usually decreased, however, septic shock may demonstrate an increased PCWP. Hemodynamic characteristics that can cause an increase in PCWP include: cardiac tamponade, predominant left ventricular (LV) failure, diastolic dysfunction, acute and severe mitral regurgitation, and acute ventricular septal defects. According to Schlant et al, in most clinical situations, PCWP, left atrial pressure, and left ventricular diastolic pressures correlate well. However, in individuals with acute or severe mitral valve disease the PCWP will be higher than left ventricular diastolic pressure, unless there is marked left atrial dilatation, then the PCWP will not demonstrate elevation. Thus, pulmonary capillary wedge pressure is a very important diagnostic tool. These pathological processes demonstrate the importance of having the ability to evaluate and monitor PCWP.

Several non-invasive techniques utilizing Doppler echocardiography have been researched and the outcome has been highly significant for correlation with Swan-Ganz PCWP. A non-invasive technique to determine PCWP would be of great advantage. This would provide the opportunity to utilize PCWP evaluation in the clinical setting more often for patients that do not require hospitalization. This research combines four techniques to determine which method produces the best correlation between Swan-Ganz PCWP and Doppler echocardiographic information.

Methods

The patient population included all intensive care unit (ICU) and operating room (OR) cases who received Swan-Ganz catheters. These patients were in normal sinus rhythm or atrial fibrillation. These patients ranged in age from 45-77 years (mean 61), thirteen being male. All patients were monitored by a Swan-Ganz catheter so PCWP could be measured pre- and post- Doppler interrogation. If on ventilation, no positive end-expiratory pressure (PEEP) could be utilized, due to the affect of PEEP on PCWP and left atrial pressure by producing a variance in these pressures. Consent was obtained by all patients before the Doppler exam was performed. Pulmonary capillary wedge pressure was taken pre- and post- Doppler echocardiography with each patient supine. ICU cases were studied utilizing transthoracic echocardiography with the patient in the left decubitus position with a 2.5 MHz phased-array transducer and a Hewlett-Packard Sonos 1000 ultrasound system. Transesophageal echocardiography was utilized to evaluate OR patients. Intubation and manipulation of the TEE probe was performed by an experienced cardiologist. Each patient was supine and a 5.0 MHz omniplane transducer with a Hewlett-Packard Sonos 1000 ultrasound system was utilized. A respirometer was used to determine end-expiration. Doppler echocardiography was performed by and measurements calculated by experienced echocardiographers. Measurements obtained by Doppler echocardiography included; a) time-velocity integrals (TVI) of the "a" and "e" waves of transmitral flow at the level of the mitral valve annulus (end-expiratory), b) TVI of the "x" and "y" waves of pulmonary vein flow, and c) Q-wave (electrocardiography) to mitral closure interval and isovolumic relaxation time (IVRT). Three cardiac cycles were measured and averaged. The results were calculated using four equations (table 1).

Pulmonary Capillary Wedge Pressure
Equations

Equation One (1)	0.95 + 15.87(A/E ROI)
Equation Two (2)	35 - 0.39x, where x=TVI(x)/TVI(x) + TVI (y)
Equation Three (3)	19.5(Q–MC/A2–MO)+ 3.0
Equation Four (3)	24(Q–MC/A2–MO)

(table 1)

Results

Sixteen patients were studied. Transthoracic echocardiography was performed on four of the sixteen patients. Of these four patients, two were in normal sinus rhythm and two were in atrial fibrillation. They were being treated in ICU for various reasons; one patient was being evaluated for assessment of left ventricular function (atrial fibrillation patient), one was status post coronary artery by-pass graft, two for mitral valve repair revision. Transesophageal echocardiography was performed on eleven patients in the operating room. Of these eleven patients, ten were in normal sinus rhythm and one in atrial fibrillation. Five of these patients were having mitral valve repair performed, six

were having coronary artery by-pass graft (in addition, one of these was having the aortic valve replaced).

The three patients in atrial fibrillation were excluded from the first equation due to the absence of the "a" wave on the mitral valve spectral waveform. The remaining patient's PCWP measured by Doppler echocardiography ranged from 6 mmHg to 33 mmHg. These patient's PCWP measured by Swan-Ganz ranged from 8 mmHg to 22 mmHg. The medical condition of these patients did not account for differences between Doppler measured PCWP and Swan-Ganz measured PCWP. The regression analysis was calculated and graphed. An additional patient was excluded when calculating the correlation due to the wide difference in estimated Doppler PCWP and Swan-Ganz catheter PCWP. The r-value calculated was 0.30. This correlation varied widely from previous literature documented by Stork et al which derived a high correlation of r=0.98.

There were two patients excluded from the second equation. One patients pulmonary vein flow was not measurable. The other was due to the absence of the "x" wave and an augmented "y" wave. The remaining patient's PCWP measured by Doppler echocardiography ranged from 4 mmHg to 22 mmHg. These patient's PCWP measured by Swan-Ganz ranged from 8 mmHg to 30 mmHg. The medical condition of these patients did not account for differences between Doppler measured PCWP and Swan-Ganz measured PCWP.The correlation was calculated excluding two patients who had wide variations in estimated Doppler PCWP compared to the Swan-Ganz catheter PCWP. The r-value calculated was -0.086. This correlation differed greatly from previous research done by Kuecherer et al in which a high correlation of r=-0.88.

There were seven patients excluded from the third equation. These were due to the inability to obtain the proper angle with the TEE probe, therefore, the IVRT could not be measured. The remaining patient's PCWP measured by Doppler echocardiography ranged from 9 mmHg to 21 mmHg. These patient's PCWP measured by Swan-Ganz ranged from 10 mmHg to 30 mmHg. The medical condition of these patients did not account for differences between Doppler measured PCWP and Swan-Ganz measured PCWP. The correlation was calculated and graphed. The r-value demonstrated a low correlation (r= -0.077). This correlation varied greatly from previous literature published by Karp et al, which demonstrated an r-value of 0.93.

There were seven patients excluded from the fourth equation. These were due to the inability to obtain the proper angle with the TEE probe, therefore, the IVRT could not be measured. The remaining patient's PCWP measured by Doppler echocardiography ranged from 7 mmHg to 22 mmHg. These patient's PCWP measured by Swan-Ganz ranged from 10 mmHg to 30 mmHg. The medical condition of these patients did not account for differences between Doppler measured PCWP and Swan-Ganz measured PCWP. The correlation was calculated and an r-value of -0.077 was derived. This correlation varied greatly from previous literature published by Karp et al, which demonstrated an r-value of 0.90.

Discussion

Previous research by Stork et al compared left ventricular filling variables derived by transmitral pulsed Doppler echocardiography and hemodynamic variables as assessed at right heart catheterization (1). This study produced a highly significant correlation (r=.98, p< .001) between PCWP and the A/E ratio of transmitral flow. The patient population included 104 ICU patients with no ventilation or valve disease. All patients were in normal sinus rhythm.

Kuecherer et al compared pulmonary venous flow and mitral inflow measured by transesophageal pulsed Doppler echocardiography to mean left atrial pressure (2). This research utilized a balloon-tipped Swan-Ganz catheter to measure pulmonary capillary wedge pressure. Left atrial pressure was estimated based on the catheter results. This estimate was then used as a measurement of the left ventricular filling pressure. The systolic fraction derived from pulmonary venous flow variables had the highest correlation (r=-.88) with mean left atrial pressure. The ratio of peak early diastolic ("e" wave) to peak late diastolic ("a" wave) mitral flow correlated strongest (r=0.43) of the mitral inflow variables.

Systolic fraction had a higher correlation with the changes in mean left atrial pressure than did the ratio of the A/E waves. The patient population included 47 OR patients studied utilizing only TEE ultrasound with no PEEP and little valve disease. All patients were in normal sinus rhythm.

Karp et al studied the interval from the onset of the QRS complex to mitral valve closure and the interval from aortic valve opening to mitral valve opening (IVRT)(3). This study demonstrated a highly significant correlation between Q-wave to mitral closure and the isovolumic relaxation time to pulmonary capillary wedge pressure (r=0.93, r=0.90). The patient population consisted of patients in normal sinus rhythm or atrial fibrillation, and included valve disease (aortic stenosis, aortic insufficiency, mitral regurgitation, and mitral stenosis).

This research combined the four above equations to determine which one would be most useful in the clinical setting of an echocardiography lab. This research also derived three equations and calculated the correlation of these equations to the Swan-Ganz PCWP. These equations demonstrated correlations of r=0.66 (A/E ROI), r=0.37 (Pulmonary vein flow), and r=0.34 (Q--MC/IVRT).

Limitations

The PCWP was taken pre- and post- Doppler interrogation. This presents a decision as to which PCWP is more accurate to compare to the equations considered. For statistical purposes, the mean value of the PCWPs observed was utilized. This author is concerned due to the fact that some of the PCWP values observed differed significantly from the beginning of the exam to the end of the exam. The cases of wide varying estimated PCWP from catheter PCWP were excluded from the calculations to determine correlation. The highest correlation was related to the equation involving the measurement of the A/E ROI. The limitation of this equation is that all patients must be in normal sinus rhythm. Unfortunately, many patients in echo labs are in atrial fibrillation. The patient sample size in this research was only sixteen. This, in comparison to previous literature, was a small sample. The Swan-Ganz PCWP measurements of these sixteen patients did not cover a broad range (8mmHg to 30mmHg). Most of the Swan-Ganz pulmonary capillary wedge pressures were in the 10mmHg to 18mmHg range. In this research, the high and low pulmonary capillary wedge pressures were excluded when the correlation was determined. If the sample volume of patients were greater and the pulmonary capillary wedge pressures had a greater number in the lower and higher ranges this research may have come to a different conclusion. Therefore, this author believes that this research demonstrates a good preliminary perspective to the role of Doppler echocardiography in determining PCWP, but further research should be performed.

Conclusion

Pulmonary capillary wedge pressure is an important diagnostic tool. The non-invasive nature of Doppler echocardiography would provide a modality to measure PCWP in patients that do not require hospitalization. This research re-created situations demonstrated in prior literature to conclude which method would be most useful in an echocardiography laboratory. The A/E ROI calculated from transmitral flow has demonstrated the highest correlation in both previous literature and this research. This demonstrates a role for Doppler echocardiography in the measurement of PCWP in the clinical setting, however, further research should be performed utilizing a larger number of patients and a wider range of pulmonary capillary wedge pressures.

Special thanks to Dr. Ted Gibbons, Chris Armstrong RDCS/RDMS, Virginia Mason Medical Center, and Rebecca Hartzler.

References:

1. Stork TV, Muller RM, Piske GJ, et al: Noninvasive determination of pulmonary artery wedge pressure: comparative analysis of pulsed doppler echocardiography and right heart catheterization. Crit Care Med 1990; 18:1158-1163.

2. Kuecherer HF, Muhiudeen IA, Kusumoto FM, et al: Estimation of mean left atrial pressure from transesophageal pulsed doppler echocardigraphy of pulmonary venous flow. Circulation 1990; 82:1127-1139.

3. Karp K, Teien D, Eriksson P: Non-invasive estimation of pulmonary capillary wedge pressure at rest and during exercise by electrocardiography, phonocardiography and doppler echocardiography. Acta Med Scand 1988; 224:337-342.

4. Schlant RC, Alexander RW: The Heart, 8th edition, volume one. New York, McGraw-Hill, Inc. 1994; 447-448, 450-458, 472, 541-542,1148.

5. Cheitlin MD, Sokolow M, McIlroy MB: Clinical Cardiology, 6th edition. Conneticut, Appleton & Lange 1993; 120-121.

6. Feigenbaum H: Echocardiography 5th ed., Lea & Febiger, Philadelphia 1994, pp 145-146, 197.

7. Weyman AE: Principles and Practice of Echocardiography 2nd ed., Lea & Febiger, Philadelphia, 1994.

8. Freedman D, Pisani R, Purves R, Adhikari A: Statistics 2nd ed., W.W. Norton & Company, New York, 1991.

AUTOMATED RECOGNITION OF GALAXY MORPHOLOGY USING NEURAL NETWORKS

Matthew L. Nielsen

Department of Astronomy
University of Minnesota
Minneapolis, MN 55455

Faculty advisors:
Stephen C. Odewahn and Roberta M. Humphreys

INTRODUCTION

Two neural network classifiers have been developed which can classify galaxies using optical images as inputs. One classifies galaxies as either elliptical or spiral to an accuracy of 87% when compared with human experts. The other gives a T type in the revised Hubble system with an average residual of 3.3 on the sixteen-step scale when compared with human experts. Hereafter, the first classifier will be referred to as the ES classifier, and the second as the T classifier. Galaxy classification using neural networks has been accomplished previously by S.C. Odewahn (1995) using photometric parameters and radial surface brightness profiles as inputs. The advantage of using raw images is that they contain morphological information which is lost in both the parameters and the profiles. Neural networks have also been used to discriminate between stars and galaxies in large digital sky surveys (Odewahn et al. 1994).

NEURAL NETWORK IMPLEMENTATION

The same basic implementation is used for both neural network classifiers. Biological neurons are simulated on a digital computer. These simulated neurons, called nodes, are arranged in layers with one layer functioning as an input layer and one as an output layer. Another layer, called a "hidden" layer, is sandwiched between these. The connections between neurons are simulated by numbers called "weights" which represent the strength of a connection between two particular nodes. Nodes are only connected to other nodes in adjacent layers. Figure 1 shows the architecture of a typical neural network.

Sample galaxy images are presented to the network's input layer. The software then computes the activity level of each node in the network based on the outputs of nodes in previous layers and on the strengths of connections between nodes. The output node with the highest activation gives the network's guess for the morphological type of the input galaxy.

The networks are trained by presenting many galaxies for which the correct output is known. The back-propogation algorithm, discussed by Rumelhardt and McClelland (1988), is used to modify the weights slightly after each galaxy is processed. In this algorithm, an output error is computed after each iteration by comparing the network's output to the known correct output. The derivative of this error is computed for each weight (connection strength) in the network. The program then modifies each weight by a small amount that is proportional to its error derivative to reduce the output error on subsequent iterations. After many training iterations, the network learns how to accurately classify the galaxies it has trained on and other galaxies as well.

GALAXY DATA

Before any automated classification scheme could be investigated, a large catalog of morphological types on a consistent system was required. A total of 145 RC3 galaxies are present on the NGP plates (O and E bandpasses) of POSS I which has been digitized by the Minnesota Automated Plate Scanner. Of these, 78 have weighted mean revised Hubble types in the RC3. This sample has been enlarged to 440 galaxies by including objects with an isophotal diameter limit of 25″ on the O plate. Two independent sets of visual classifications from the POSS I O print were made by H.G. Corwin and S.C. Odewahn. They have analyzed the T residuals (see Figure 2) derived from the overlap of three catalog combinations: RC3-SCO, RC3-HGC, and SCO-HGC. Impartial regressions were used to transform SCO and HGC types to the RC3 scale. By intercomparing these three independent T catalogs, Corwin and Odewahn computed the mean T error for each catalog as: $\sigma(RC3) = 1.6$, $\sigma(SCO) = 1.9$, and $\sigma(HGC) = 1.6$. These errors were used in assigning relative weights to compute a final catalog of weighted mean types for 440 galaxies with $D(\mu_O = 24.5) \geq 25''$.

THE CLASSIFIERS

The galaxy classifiers are two-layer neural networks (two processing layers plus an input layer). The input layers contain 900 nodes, one for each pixel in 30-by-30 input images. The ES classifier has a hidden layer containing 48 nodes and an output layer containing two nodes, one to indicate "elliptical" and one to indicate "spiral." The node with the highest activation gives the network's classification. The T classifier has a hidden layer containing 72 nodes an output layer containing four nodes representing the continuum of Hubble types. A weighted mean of the output node values is computed for each input image and transformed from this four-step scale to the sixteen-step revised Hubble system. The networks are fully connected meaning that each hidden node is connected to all 900 input nodes and all output nodes.

In order to simplify the classification problem, galaxies with axis ratios of less than 0.5 were eliminated to make the images more homogeneous. Galaxies with isophotal diameters greater than 50″ were also eliminated because they tend to be very bright, causing saturation of the photographic emulsion. The images were scaled to 30 pixels by 30 pixels to prepare them for input to a neural network.

In order to train the networks, the sample galaxies must be binned according to T type, with each bin corresponding to a particular output node. For the ES classifier, galaxies with $T \geq 1$ were called "spiral," and all others "elliptical." For the T classifier, the galaxies were grouped in four bins, each spanning four steps on the sixteen-step Hubble scale. Training and testing sets containing equal numbers of galaxies in each bin were assembled for each of the two classifiers. Galaxies in the training sets were presented repeatedly until the networks learned to classify them correctly. The testing sets were then used to gauge the networks' ability to classify galaxies they had never "seen" before.

RESULTS

Figure 3 shows the accuracy of the ES classifier as a function of training epoch (passes through the training set). The accuracy is gauged by success ratios, the number of correct classifications divided by the total number of galaxies in a particular set. The success ratios for both the training and testing sets increase with more training. It reaches a maximum of 100% on the training data and 87% on the testing data.

Figure 4 shows preliminary results for the T classsifier. Let T_n be the T type given by the neural network classifer for a particular galaxy, and let T_t be the correct T value as determined by the method described in "Galaxy Data." For this plot, galaxies in the testing data were binned according to T_t into eight bins spanning two steps each on the revised Hubble system. T_n and T_t were averaged for all the galaxies within each bin. These average values are the points indicated in Figure 4. The error bars indicate the mean standard deviation in T_n for individual galaxies in the bin. The standard deviation for a particular galaxy is computed by transforming the standard deviation of the output node values to the T scale. The T classifier has not yet been perfected as is obvious

from the plot. The large standard deviations indicate that more than one output node is firing significantly on a particular galaxy, that is, the network is uncertain about its classifications. Also, the points do not form a line of slope unity as they should. The reason can be easily explained. If the output nodes are numbered from one to four, then an early elliptical will cause node 1 to have the largest output, but the other nodes will produce some output as well, skewing the resulting T type to a higher value. The same effect causes the late spirals to be skewed to lower T values, while the intervening types (causing nodes 2 and 3 to fire) are not affected as much because noise in nodes 1 and 4 cancels in the weighted mean. A means of circumventing this problem has not yet been found.

Future work will include (1) enlarging the training set to increase the networks' generalization ability, (2) increasing the size of the T classifier's output layer to eight nodes to increase its accuracy, (3) extending the data sets to all axis ratios, and (4) rotating all images to a constant position angle to increase the sample's homogeneity and thereby its generalization ability.

ACKNOWLEDGEMENTS

I would like to thank Steven Odewahn for providing figures 1 and 2 and for the use of his software in assembling my data sets. I am also indebted to the Minnesota Automated Plate Scanner project for providing the hundreds of galaxy images which made this work possible.

REFERENCES

[1] Odewahn, S.C., *Astronomical Letters and Communications*, Vol. 31, pp. 55-64, 1995
[2] Odewahn, S.C. and Nielsen, M.L., *Vistas in Astronomy*, Vol. 38, pp. 281-286, 1994
[3] McClelland, J.L. and Rumelhardt, D.E., *Explorations in Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1988

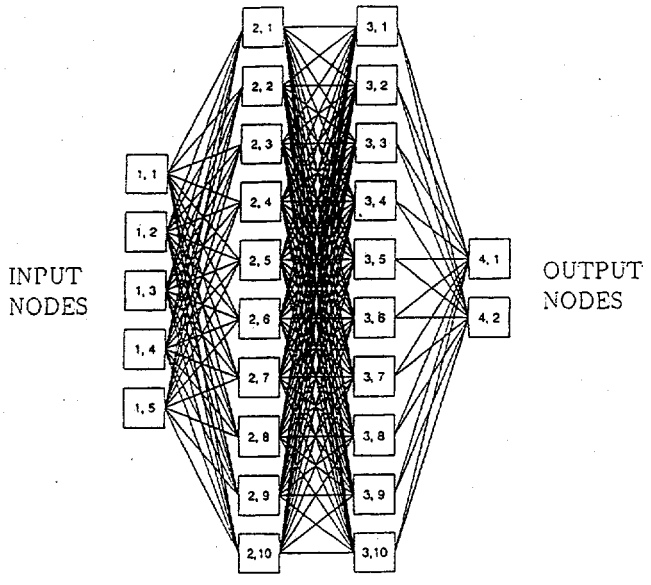


Figure 1: Architecture of a typical neural network.

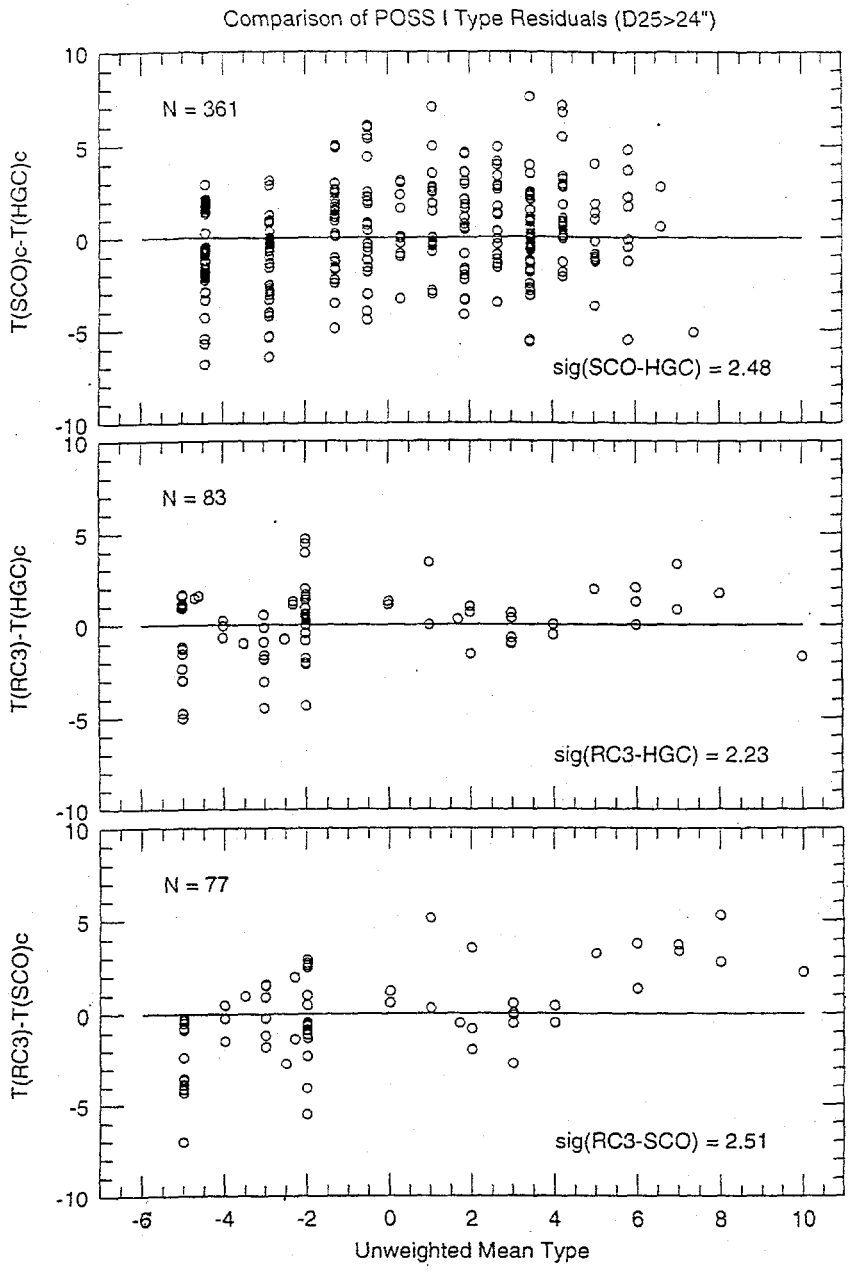


Figure 2: T residuals for three catalog combinations versus T showing the scatter among human experts.

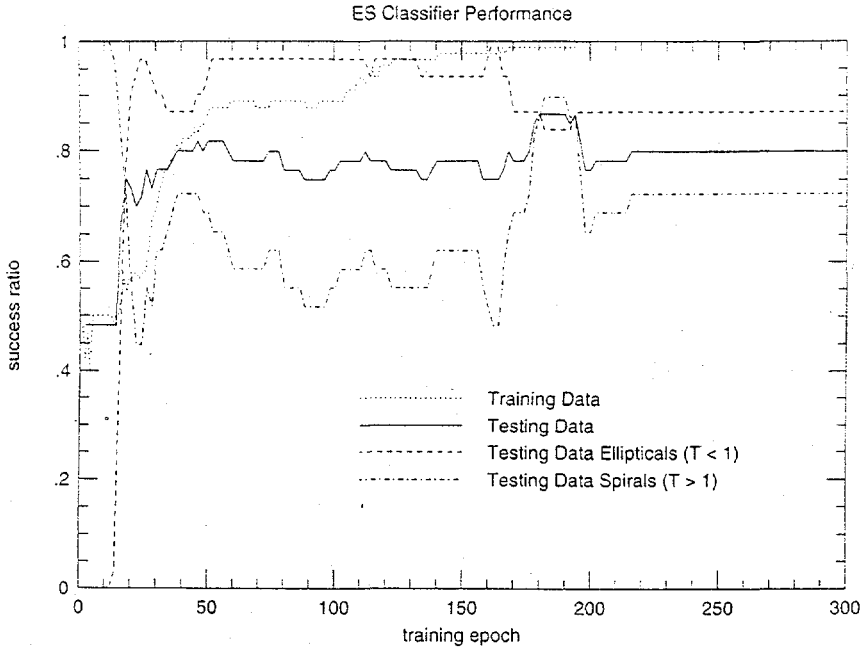


Figure 3: Accuracy of the ES Classifier versus training epoch (passes through the training set). Notice that the success ratio for the testing data reaches a maximum of 87%.

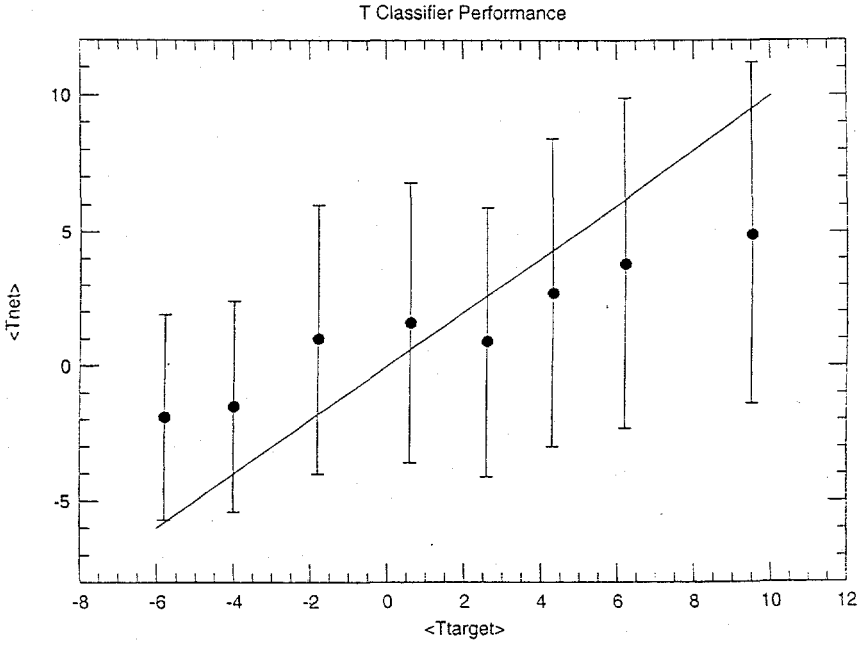


Figure 4: Average T given by the neural network versus that determined by human classifiers for testing data in eight bins spanning the revised Hubble scale. The diagonal line is the ideal line of slope unity. See "Results" for a description of the error bars.

The Disappearance of the Eclipse Minima of the Binary Star SS Lacertae

Tammy L. Clifton

Department of Physics
South Dakota State University
310 Crothers Engineering Hall
Brookings, SD 57007

Faculty Advisor:
Dr. Stephen J. Schiller

SS Lacertae in the open cluster NGC 7209 was once a normal spectroscopic eclipsing binary star system. Photographic observations obtained between 1890 and 1950, which have been published by several astronomers, have clearly identified SS Lacertae as an eclipsing binary system with a known period of 14.4 days and primary and secondary eclipse depths of about 0.5 magnitude. Detailed photometric analysis indicates that eclipse depths decreased gradually and were no longer detectable by 1950. More recent extensive photoelectric photometry have confirmed that the eclipses are now non-existent.

We have compared observed radial velocity data to results of a synthetic light curve program to investigate the change in orbital parameters necessary to account for the observed disappearance of eclipses. In addition, an effort to improve historical photographic photometry is underway that includes applying digitization and image processing techniques to nearly 1700 plates contained in the Harvard College Observatory Archives.

INTRODUCTION

Binary stars are studied for a number of reasons. Through study, one can define certain parameters that we can apply to stars of all types. Information on the aging process of stars and in the dynamical physical interaction that can also be investigated through binaries. The period of the orbit, the semi-major axis of the orbit, the eccentricity of the orbit, the mass of each star, the radius of each component estimated through knowledge of the spectrum and the pair's luminosity, and the spectral type which helps us determine age and temperature all can be determined.

This approach to the study of stellar evolution in binary stars is based upon the membership of these stellar pairs in clusters. In a binary's lifetime one can see birth, life and death as related to other stars. The "life" process in a binary star system begins as two stars orbiting a central point that is common to both, the center of mass. As these stars age, different characteristics become evident. Radii will increase, temperature will decrease, and other changes will take place. At a certain stage, the two stars will become close enough that greater gravitational effects will occur. The two stars will pull towards each other, and mass exchange will take place. After one star has "cannibalized" its companion, the system, of course, has been altered. There is no way to gauge the stellar age of the stars in the pair because they have changed their properties differently than if they evolved as single isolated stars. However, in clusters, using single stars that have evolved normally within the cluster, one can determine the age of the cluster and therefore the binary system that has undergone mass exchange. (In interacting field binaries, we cannot be really determined because the pair lacks single siblings born from the same clouds of dust and gas.) This method allows the mapping of the changes in temperature and luminosity with age for binary stars so that we can trace the evolution that each stellar object must undergo.

In the study of SS Lac, the pair appears to have been a normal eclipsing binary system around the turn of the century. The period was 14.4 days, the eclipse depth magnitudes were about 0.5, and the spectral type for these each of these stars was B9v. The system was a detached system with a semi-major axis of 0.205 AU. The distance is enough that the interaction between the stars was minimal, except for the gravitational effects that determines the characteristics and parameters of a binary system.

BACKGROUND

Several astronomers have studied SS Lacertae. In 1935, Raymond S. Dugan and Francis W. Wright identified SS Lac as an eclipsing system and placed it on a program to extend the history of the variation of period of a considerable number of eclipsing variables. They discovered that the variable was found fainter than normal on ninety good plates taken from the Harvard College Observatory. Twelve plates were taken on one night which cover the descending branch (of the light curve) and apparently indicate a constant minimum brightness lasting at least two hours. They suggested that the entire duration of the eclipse is of the order of one-half day. Through this research, they confirmed that the period is indeed JD 2415900.76 + 14.41629 days, which was further confirmed by Wachmann in 1936. Based upon the assumption that the inclination is 90 degrees, the minimum orbital eccentricity is 0.11.

While studying plates taken from approximately around the same dates that Dugan and Wright used in their study, N. Tashpulatov measured the magnitudes of SS Lac. While he does not mention it, the photometry from his study indicates a change in the system. SS Lac's minima suddenly disappeared according to Tashpulatov. Sometime between October 20, 1949 and August 27, 1951, the minima had completely vanished. Recently, astronomers have reevaluated the Tashpulatov data and suggest that the photometry is unreliable. However, a significant amount of additional photometric evidence leads astronomers to conclude that the system no longer eclipses. Clearly, something dynamic happened to the system.

Currently, research continues in solving the mystery of SS Lacertae's disappearing minima. Most notable of the current researchers are Schiller and Milone (1991), and Lehmann (1990). Photoelectric observations by S. J. Schiller obtained from 1982-1984 confirmed that the predicted light variations of the times of primary minima were not present. A more complete phase coverage that included times of both primary and secondary eclipses, by astronomers M.M. Zakirov and A. A. Azimov during 1984-1989, also showed no variability.

According to Schiller and Milone (1993), the light-curve of Dugan and Wright indicates two components of nearly equal surface brightness and of an early (B9v) spectral type, the system should be easily detected as a double-line spectroscopic binary or a binary system where spectral features from both components may be seen to move on the spectra taken at varying times. However, cross-correlation analysis, a technique that determine velocities of the stars in close binary system, of twelve spectrograms obtained at the Dominion Astrophysical Observatory between 1982-1984 by both Schiller and Milone revealed only a single-line system with radial velocity variations that are not clearly present above the uncertainty in a single observation.

There is no evidence of a change in brightness of the system. Schiller and Milone have preformed a search for evidence of third light in the light curves of Dugan and Wright, Wachmann, and Tashpulatov, but the search has currently proved inconclusive. Their results from digitized data and analysis performed with the Wilson-Devinney program confirm the conclusions of Dugan and Wright. Schiller and Milone have concluded that eclipse conditions have changed radically. They suggest that because of the constancy in the radial velocity, it precludes a normal evolutionary explanation of the vanishing eclipse and imply that changes in period or inclination as a result of gravitational collision with a third body, are the culprits for the disappearing minima.

Thomas Lehmann (1990) also concludes that the minima began disappearing sometime around 1950. His research indicates that there was a continuous change of the amplitude. Based upon the assumption that the observed variable amplitude is due to a continuous change of the inclination i of the orbital plane of the eclipsing binary, the greatest amplitude corresponds with i=90 degrees. If we begin our time when one star moves directly in front of the other and inclination is 90 degrees, then initial time is 1911 ± 3 years. He gives the change in inclination with respect to time as 0.18 ± 0.02 degrees/year. For 1994, the inclination would be 76.86 degrees.

RESEARCH

During the summer of 1994, this project to investigate the disappearance of the

minima of SS Lac began. Using many of the ideas and results from Schiller, Milone, and Lehmann, the radial velocity curves and changes in eclipse depths were analyzed by implementing programs that had not been used previously in analyzing data. Our results are in agreement with previous investigations and indicate that based upon the theoretical and observational data, something drastic has happened to the system recently.

Figure 1

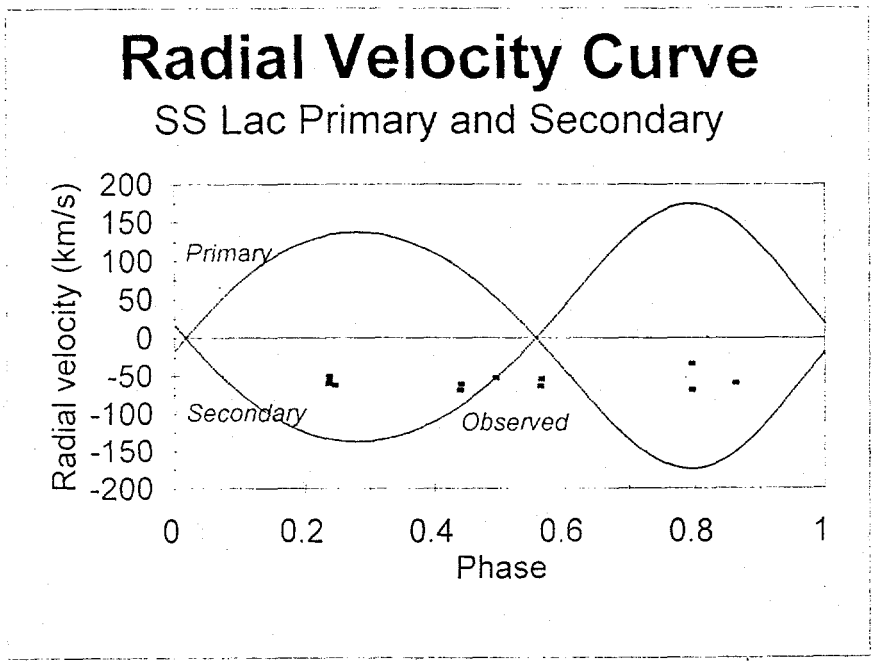
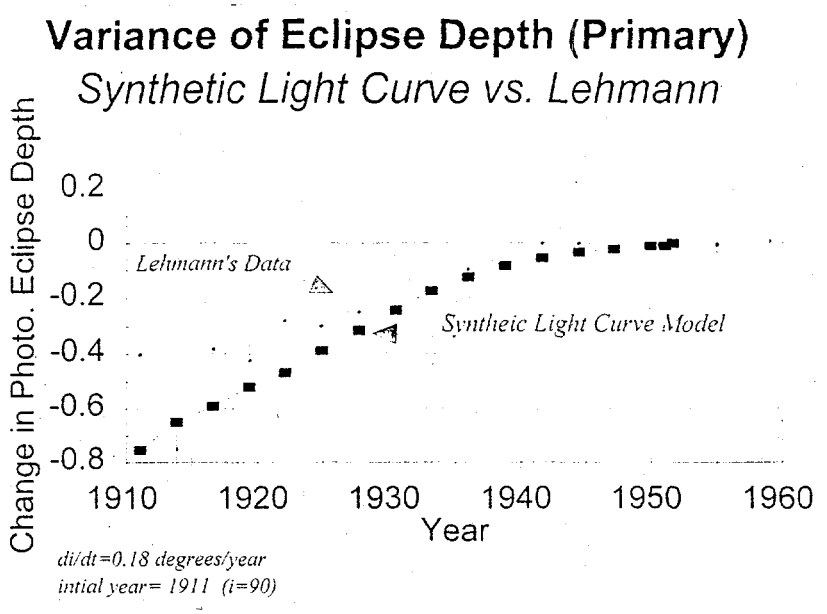


Figure 1 describes the theoretical radial velocity of SS Lac using Lehmann's assumption of change in inclination as compared with observational radial velocity data provided by Schiller in 1983. Lehmann's rate of change in inclination is used to predict the inclination of SS Lac at the time radial velocities were measured. A value of 76.86 degrees results from this calculation. The velocity varies by approximately 150 km/s from 0 km/s, where neither star is coming towards or moving away from earth. This should represent the radial velocity curve of today, however it does not. Today's radial velocity observations show little variation and centers around -60 km/s. This is likely the space velocity of SS Lac's center of mass. The very lack of differentiation in the actual radial velocity data from the theoretical curve illustrates the change in the system we know as SS Lac.

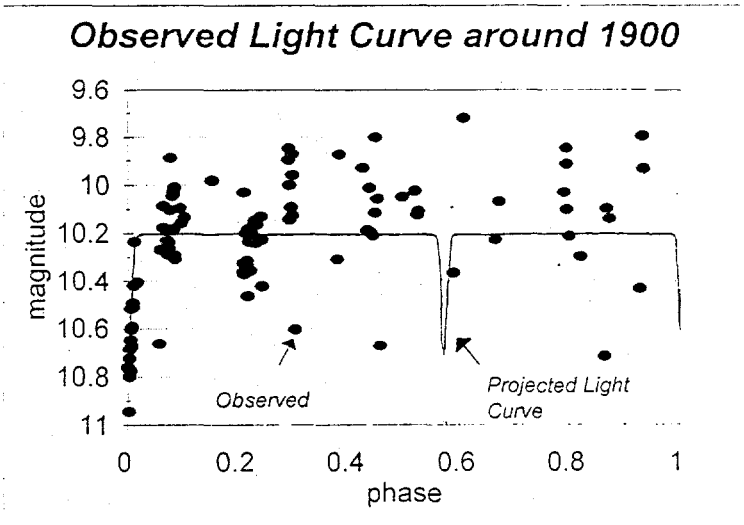
According to Lehmann's information, eclipses had ended around 1950. This calculated using actual photographic data that he had at his disposal. In figure 2, the change in eclipse depth was calculated using Lehmann's methods and a synthetic light curve, a light curve made by inserting parameters such as mass, radii, luminosity, and inclination in the modeling program in the software package *Binary Maker 2.0* are plotted. By the photometric eclipse depth on the synthetic light curve using varying inclinations, it is determined that eclipses ended sometime around the year 1950. The eclipses would end according to Lehmann's analysis about the time Tashpulatov's photometry indicated that it did, assuming apsidal motion is the cause of the orbital change.

Figure 2



Currently, only a fraction of the plates that were obtained from Harvard College Observatory Archives have been analyzed. Initial measurements of the photographic magnitudes of SS Lac are shown in Figure 3. By virtue of the digitization program, several eclipses occurring around 1900 have been observed. Plates from the years 1940-1960 are currently being analyzed.

Figure 3



It is hoped that the disappearance of the eclipses will be observed and the date of the last eclipse will be confirmed. Future investigations will include a search for the third body dramatic event. We hope to solve the disappearance of the eclipse minima of the binary star SS Lacertae.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to the South Dakota Space Grant Consortium and to S. J. Schiller for making this project possible.

REFERENCES

Dugan, R. S., Wright, F. W. "The Period and Photographic Light-Curve of SS Lacertae." *Astronomical Journal*. 1936. Vol. 44, No 1031, pg 150.

Lehmann, T. "The Unusual Eclipsing Variable SS Lacertae." Commission 27 of the *I. A. U. Information Bulletin on Variable Stars*. Number 3610. 23 May 1991. KV ISSN 0374-0676.

Milone, E. F., Stagg, C. R., Schiller, S. J. "Constraints on the Cessation of Eclipses in SS Lacertae and Their Implications for System Evolution." *I. A. U. Information Bulletin on Variable Stars*. 1992. Pgs 479-482.

Schiller, S. J., Milone, E. F., Zakirov, M. M., Azimov, A. A. "The Non-Eclipsing Eclipsing Binary SS Lac." *Bulletin of the American Astronomical Society* , 1991. Vol 23, pg 879.

Tashpulatov, N. "Photographic Observations of SS Lacertae." *Peremennye Zvedt* (Variable Stars), 1965. Vol. 15, pgs, 424-430.

Wachmann, A. A. *Astronomische Nachrichten*. 1936. Vol. 258, pg 361.

Photoluminescence and Photoluminescence Excitation Spectroscopy Of ZnTe/ZnCdTe Multiple Quantum Wells

Susan E. Thysell

Department of Physics and Astronomy
Carleton College
Northfield, MN 55057

Faculty Advisors:

J. J. Song and J. M. Hays

Department of Physics and Center for Laser Research
Oklahoma State University
Stillwater, OK 74078

Photoluminescence (PL) and photoluminescence excitation (PLE) experiments are performed on ZnTe/ZnCdTe multiple quantum wells at temperatures near 10 K. The PL spectra show evidence of quantum confined states, particularly the lower lying heavy hole transition between the n = 1 states and impurity states. The PLE spectra show some structure in the samples due to n = 1 states in the higher lying light hole valence band.

INTRODUCTION

The present experiment uses the methods of photoluminescence (PL) and photoluminescence excitation (PLE) to study the multiple quantum wells (MQWs) of $ZnTe/Zn_{1-x}Cd_xTe$. Some background information about semiconductors is helpful prior to a discussion of MQWs.

A semiconductor is a solid that has a gap between its filled valence band and its empty conduction band. As atoms come together to form molecules in a solid, their atomic energy levels fan out into groups of distinctly spaced degenerate levels, called bands. The energy levels in these bands sometimes overlap, but when they do not overlap there is a band gap -- a range of forbidden energy levels. When the atoms are in the ground state, the electrons stay in the valence band. However, a photon with at least the required band gap energy E_g can excite an electron from the valence band at energy E_v into the conduction band at energy E_c, creating a hole in the valence band (Fig. 1). We are concerned with the two highest valence bands -- one for heavy holes (HH) at low energy near the band edge and one for light holes (LH) at higher energy. These two bands are degenerate at k = 0 in bulk zinc blende semiconductors. If the electron and the hole are bound to one another by the mutual Coulombic attraction, the pair is called an exciton (see Pankove, 1971, for exciton types). When the pair recombines a photon will be emitted that can be detected in PL and PLE studies (Taylor and Zafiratos, 1991).

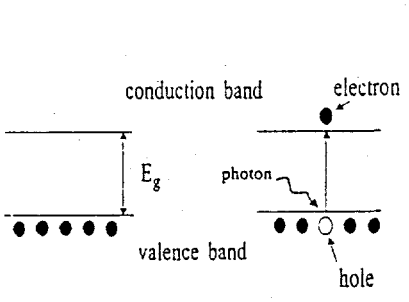


Figure 1: Free electron and hole creation by electron excitation in bulk semiconductor material.

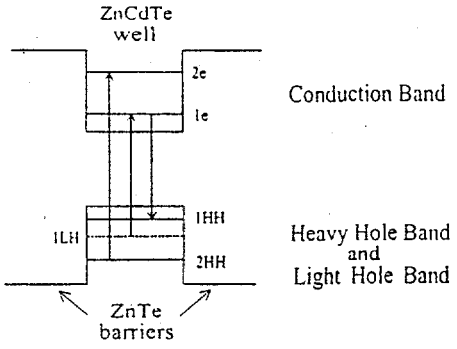


Figure 2: Potential well structure of ZnTe/ZnCdTe MQWs. Arrows show the allowed band transitions for the n = 1 and n = 2 states.

A quantum well is a thin layer of semiconductor material with a smaller band gap than that of the surrounding barrier material. The multiple quantum wells for the present experiment were created by alternating layers of ZnTe and ZnCdTe, forming rectangular wells of lower potential energy in the layers of ZnCdTe (Fig. 2). The electrons and holes in the conduction and valence bands are confined to these potential wells (Saleh and Teich, 1991).

Looking at one dimension in a well of width d, the energy levels E_n can be determined by solving the time-independent Schrödinger equation:

$$E_n = \hbar^2 (n\pi/d)^2 /2m \ , \ n = 1, 2, 3,... \qquad (1)$$

This is an approximation for an infinite well. A smaller well width leads to a larger separation of energy levels within the well. The energy levels in the other two dimensions will still be approximated as continuous bands of energy (Saleh and Teich, 1991). These sub-bands form in both the conduction band and the valence bands.

Band-to-band transitions take place between these energy levels, but only transitions where $\Delta n = 0$ are allowed. Band-to-band absorption occurs when an electron is excited by a photon of energy $h\upsilon > E_g +$ the confinement energy E_n for each band. The electron absorbs this energy and moves from a level in the valence band to an allowed level in the conduction band. Emission occurs when the electrons and holes recombine, emitting a photon. PL allows us to study the lowest possible transition, from the n = 1 electron state in the conduction band to the n = 1 hole state in the valence band (Fig. 2).

It is important to note that ZnCdTe is a semiconducting alloy made up of the II-VI semiconducting materials ZnTe and CdTe. The introduction of Cd lowers the band gap energy of ZnTe, according to the relative concentrations of Zn and Cd. The differing concentrations of Cd also cause a lattice mismatch between the ZnCdTe quantum wells and the ZnCdTe buffer layer, as well as the ZnTe barriers, which causes strain in the material. This strain can cause the LH sub-band to have a smaller band gap (relative to the conduction band) than that of the HH band, allowing a transition from the conduction band in the ZnCdTe quantum well to the LH band in the ZnTe barrier. Figure 3 shows this transition, called Type II, as well as the Type I transition from the conduction band to the HH band. An analysis of PLE studies can tell us whether a change in band energies has occurred. The different band gap energies caused by Cd are useful for one method of creating quantum wells.

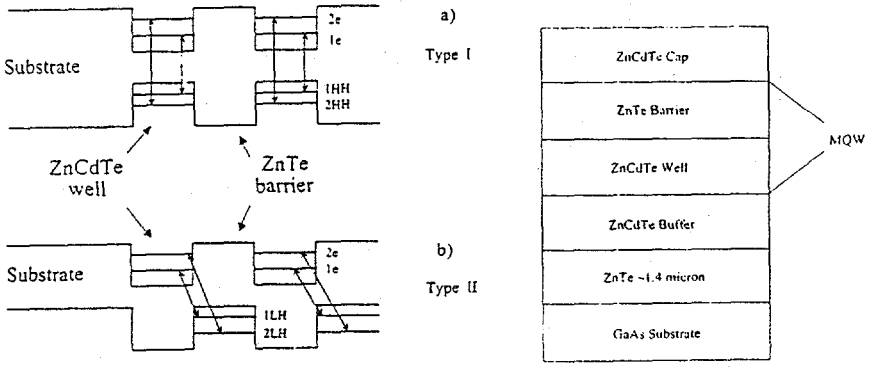


Figure 3: Multiple quantum well structures of a) a Type I quantum well with transitions from the conduction band in the well to the HH band in the well, and b) a Type II quantum well with transitions from the conduction band in the well to the LH band in the barrier.

Figure 4: Schematic diagram of the ZnTe/ZnCdTe MQW sample structure. Only two samples, R821 and R826, have a cap layer.

EXPERIMENT

The ZnTe/ZnCdTe samples used for the present experiment were provided by Sandia National Laboratories. A schematic diagram of their structure is shown in Fig. 4 and the thickness and composition values of each of the seven samples are shown in Table 1.

The samples were mounted with vacuum grease on a copper mount and placed inside a closed-cycle LHe cryostat. The PL data were taken at about 10 K, using an argon laser, a 3/4 m double spectrometer, a photomultiplier tube (PMT) and a photon counter. The laser ran at 5145 Å and 500 mW, but the power was attenuated to 34 mW. The resolution was 20 cm^{-1} for the long scans (5260 Å to 8000 Å) and 5 cm^{-1} for the short scans (5260 Å to 6250 Å).

The PLE data were taken at a temperature of 11 K. A 10 Hz Nd:YAG laser with 8 ns pulse width was used to pump a dye laser using the dye Rhodamine 6G. We used a 1-meter spectrometer, the same type of PMT and a boxcar to record the data. For each sample we set the spectrometer at an appropriate wavelength as determined from the peaks we observed on the PL data. We changed the spectrometer slit widths for each sample. We then scanned the dye laser over a corresponding range of wavelengths, beginning at a wavelength slightly less than that of the spectrometer and moving to shorter wavelengths. See Fig. 5 for a diagram of the PLE apparatus. Similar apparatus was used for the PL studies, except an argon laser replaces the Nd:YAG laser and dye laser and a 3/4 meter double spectrometer was used.

Table 1: Thickness and composition values for the ZnTe/ZnCdTe MQW structures. The composition is given by the Zn percentage.

	R820	R821	R822	R823	R824	R825	R826
ZnCdTe Cap Thickness	--	1730 Å	--	--	--	--	1710 Å
ZnCdTe Cap Composition	--	0.92	--	--	--	--	0.86
# of Periods in MQW	25	25	25	25	25	25	25
ZnTe Barrier Thickness	136 Å	136 Å	136 Å	176 Å	76 Å	95 Å	95 Å
ZnCdTe Well Thickness	51 Å	51 Å	51 Å	66 Å	56 Å	70 Å	70 Å
ZnCdTe Well Composition	0.68	0.68	0.68	0.68	0.68	0.68	0.68
ZnCdTe Buffer Thickness	1.7 μm	1.7 μm	1.7 μm	1.7 μm	1.9 μm	1.9 μm	1.9 μm
ZnCdTe Buffer Composition	0.92	0.92	0.92	0.92	0.86	0.86	0.86

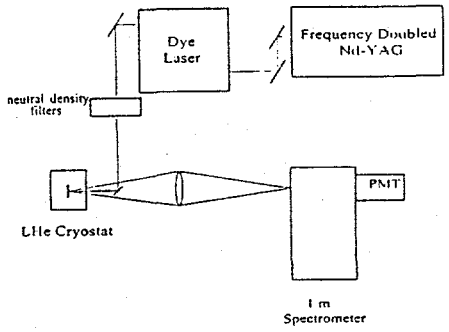
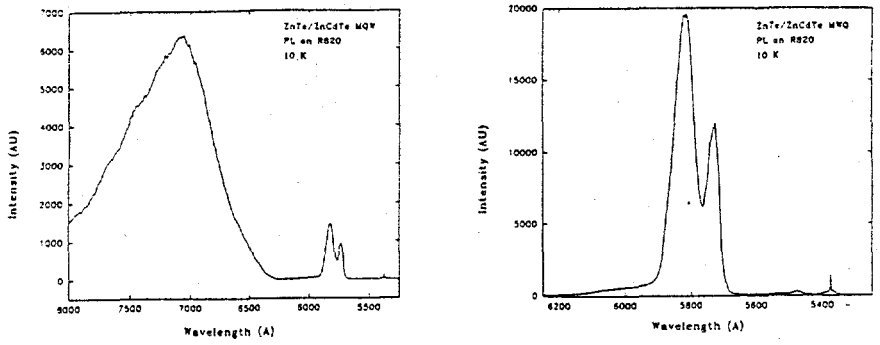


Figure 5: Apparatus for the PLE study, which uses the backscattering configuration. The dye used for the PLE measurements was R6G, which lases in the range 5550 to 5950 Å. The PL apparatus is similar, except an argon laser replaces the Nd:YAG laser and dye laser, and a 3/4 meter double spectrometer is used.

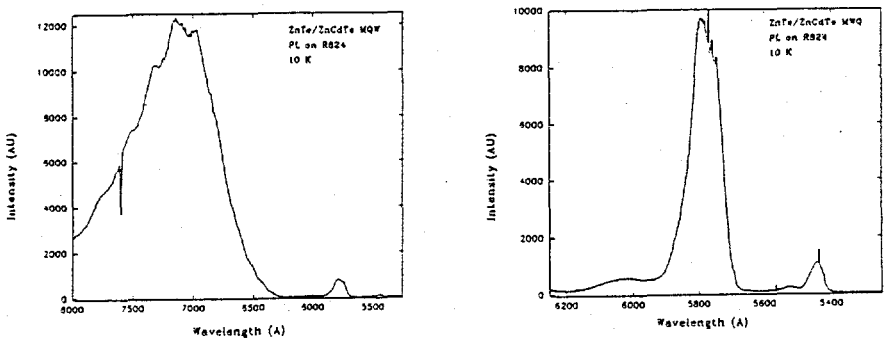
RESULTS AND DISCUSSION

Each long PL run (5260 Å to 8000 Å) showed an intense, broad emission with peak positions between 7100 Å and 7200 Å and lower, narrower peaks between 5700 Å and 5900 Å and around 5400-5500 Å (Fig. 6a) and 7a)). The most well-defined peaks between 5700 Å and 5900 Å, which can be seen more clearly on the shorter runs (Fig. 6b) and 7b)), represent the Type I transition from n = 1 in the conduction band to n = 1 in the HH band. The peak wavelength for each sample was used to determine the spectrometer setting for the PLE runs. Only three of the samples -- R820, R824 and R826 -- showed structure in the PLE studies. Two examples are shown in Figure 8. Calculations based on our data from these samples, taking into account the ZnTe barrier thickness, the ZnCdTe well thickness and composition (percent of Zn in well layer), and the ZnCdTe buffer composition (percent of Zn in buffer layer), allowed the PLE and PL spectra to be fit to expected energy levels. The fits from our calculations (Hwang, 1994) indicate that Type II transitions from n = 1 in the conduction band to n = 1 in the LH band occurred in the three samples mentioned above. The calculations are approximate and the fits are not exact because we did not take the Stokes' shift into account. Our calculations also suggest that the n = 2 state of the HH band, which should be confined, is very close to the band gap of the ZnTe barrier. This may be why we have not observed this n = 2 transition in the PLE data.



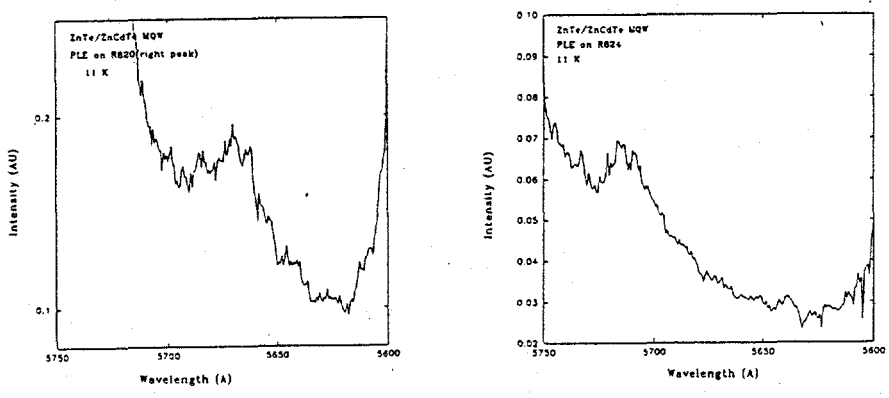
a) b)

Figure 6: PL spectra for sample R820. a) The long range scan from 5260 Å to 8000 Å. Note the large impurity peak at longer wavelengths. b) The short range scan from 5260 Å to 6250 Å, which clearly shows excitonic emission from the quantum wells.



a) b)

Figure 7: PL spectra for sample R824. a) The long range scan from 5260 Å to 8000 Å. The small peak below 5500 Å represents a barrier transition. b) The short range scan from 5260 Å to 6250 Å.



a) b)

Figure 8: PLE spectra for a) sample R820 and b) sample R824. Calculations based upon the PL and PLE suggest that these peaks are due to weakly Type II LH transitions, while the emission shown in the PL spectra are due to Type I HH transitions.

SUMMARY

The present experiment used the methods of PL and PLE to study the MQWs of ZnTe/$Zn_{1-x}Cd_xTe$. The intense, wide peaks in the PL spectra, around 7100 Å, are caused by impurities in the samples, indicating poor sample quality. Calculations based on our PL and PLE data indicate two types of quantum well structures. The well-defined peaks between 5700 Å and 5900 Å on the PL spectra represent the transition in a Type I quantum well from n = 1 in the conduction band to n = 1 in the HH band. The peaks in the PLE spectra from three of the samples -- R820, R824 and R826 -- represent the transition in a Type II quantum well from n = 1 in the conduction band to n = 1 in the LH band. Our calculations also indicate that the n = 2 state of the HH band may be very close to the band gap of the ZnTe barrier, which may explain why we did not observe this transition in the PLE spectra.

ACKNOWLEDGEMENTS

The author thanks John M. Hays for his guidance and illustrations for Figures 2, 3, and 5, Seon-Ju Hwang for his calculations and Scott Rodgers for his explanation of semiconductors.

REFERENCES

Hwang, S. J., unpublished calculations, 1994.

Pankove, J. I., Optical Processes in Semiconductors, Prentice-Hall, Inc., New Jersey, 1971, pp. 12 - 17.

Saleh, B. E.A. and M. C. Teich, Fundamentals of Photonics, John Wiley & Sons, Inc., New York, 1991, pp. 569-581.

Taylor, J. R. and C. D. Zafiratos, Modern Physics for Scientists and Engineers, Prentice-Hall, Inc., New Jersey, 1991, pp. 513-519.

BVRI PHOTOMETRY OF G DWARF STARS TOWARDS THE NORTH GALACTIC POLE

Robert Lanczycki

Center for Galactic Astronomy
Western Connecticut State University, Danbury, CT 06810

Faculty Advisor:

Phillip K. Lu

CCD photometry using broad-band BVRI (B = Blue, V = Visual, R = Red, I = Infrared) system has been obtained for 152 late-F and early-G stars near the North Galactic Pole (NGP) at the National Undergraduate Research Observatory (NURO), Flagstaff, AZ. This study is an extension of a long term and ongoing spectrophotometric study of the South Galactic Pole (SGP), a survey of the F and G stars in a 40 sq-deg area for the K(z)study. All data reduction and analyses were carried out at the Center for Galactic Astronomy (CfGA). Preliminary results are presented here with comparisons between the two galactic poles and data reported by other investigators.

INTRODUCTION

The density distribution, stellar population and chemical abundance in stellar tracing groups between the north and south galactic polar directions provide vital information on the structure and evolution of the Milky Way Galaxy. O,B, A stars and early F stars are too few at a high galactic latitude and too young to be in a state of dynamical equilibrium; giant stars are luminous and can be detected to the great distance but they are such an inhomogeneous group in absolute magnitude, kinematics and chemical composition that stringent selection must be imposed for such study. In contrast, late F and early G stars are unevolved, first generation stars that occur in large numbers and can be seen out to a z-distance of 1 to 2 kpc, suitable for studies in galactic structure. Photometric and spectroscopic data have been obtained for nearly 2000 stars at the Cerro Tololo Inter American Observatory, (La Serena, Chile) by Lu and colleagues for the determination of the velocity dispersion of these stars in study of the galactic gravitational force perpendicular to the galactic plane, simply referred to as the K(z) force.

The North Galactic Pole is part of a much larger project to investigate the galactic gravitational force using the F and G stars at both galactic poles (Lu, 1991, 1994; Philip and Lu, 1989, Lu et al 1992). Using this K(Z) force, Oort (1932, 1960) first suggested that missing mass is a problem in the Galaxy. The amount of local mass density in unit of Mo/pc^3 in the solar neighborhood is often referred as the "Oort Limit." Currently, the missing mass is called Dark Matter, since galactic mass can be matter emitting light from stars, nebulae and clusters as well as undetectable or invisible low luminosity objects such as planets and black holes that do not emit very much visible light.

Stellar population and density distributions are generally assumed to be similar between the north and south galactic poles, but little data were available for analysis and comparison. Philip (1974) found that the distribution of early type Pop. I and II stars were similar at both poles: however, it is not clear that this similarity is true for the older and late F and G-type stars. More recent studies of the velocity dispersion versus metallicity suggest that the metal rich stars, which are generally younger stars concentrated toward the disk seem to have more regular orbits, whereas

metal poor stars away from the disk, have more random motions, thus higher velocity dispersions. Radial velocity measurements of large numbers of stars toward the NGP studied by Latham and colleagues at the Center for Astrophysics, toward the NGP and stars near the SGP studied by Lu and his co-workers at CfGA, have shown that the mean velocity of the two galactic poles has an antisymmetric value [about -11 km/sec at NGP (Latham et al.1991, Latham, 1992) and +13 km/sec at SGP (Lu, 1990, 1994)]. Therefore models (for example, Weinberg, 1991) have proposed that the Milky Way Galaxy is oscillating, resembling a piece of jello vibrating vertically as well as horizontally.

OBSERVATION AND DATA REDUCTION

Analyses using SGP data with the NGP data will be presented in another paper. This study will present the comparison between our data and the data published by Corbally and Garrison (1988, referred as CG). In early studies by Corbally and Garrison, photographic magnitudes and colors were obtained using UBV 3-color direct photographic plates, and thin prism spectra were estimated from a 1.8 degree thin objective-prism at a dispersion of 1300A/mm at H-gamma. Corbally and Garrison provided a list of 192 stars between late F and early G types in an area of 3.7 sq-deg. near Selected Area #57; magnitudes range from 11.1 to 15.1. Subsequently, Garrison and Corbally (1993) published a more detailed classification in the MK system using slit spectra.

This study was carried out at the National Undergraduate Research Observatory (hereafter, NURO), at Anderson Mesa, near Flagstaff, AZ; Western Connecticut State University is a charter member of the consortium. All CCD data were obtained at NURO in May 1993 and 1994, using the 31-inch Lowell reflector with a back-illuminated 512 x 512 Tektronix chip chilled by liquid nitrogen to -111 degrees C. The normal integration time for V, R, or I frame is about 30 seconds to reach 15th magnitude stars but triple time is needed for the B filter in order to achieve the same brightness intensity due to the chip's low quantum efficiency in blue light.

Standard stars in nine selected areas were observed (Landolt, 1973, 1983). Most well calibrated Landolt UBVRI standard stars are about 10th magnitude, with fainter stars available for UBV only. Therefore, the standard deviations show in Figs. 1 and 2 are only for V and B-V. The r.m.s deviations for the standard stars were about 0.02 magnitude for VRI, and 0.04 for B.

All images were transported back to the CfGA either using Exabyte tapes or directly mailed back to the DEC work-station at the CfGA via ftp over Internet. Disk space was a problem during the 1993 run. However, with the currently available 4 Gbytes, all bias and flat field corrections can be performed at one time. Photometric magnitudes were obtained using the aperture photometry task within IRAF (Image Reduction and Analysis Facilities), which generally produces an error of 0.005 magnitude depending on the magnitude and integration time.

RESULTS AND CONCLUSION

Of the 152 stars observed, more than 30 stars have multiple CCD observations separated by about a year. The deviations between two or three observations are generally small with a few exceptions toward fainter stars. The spread can be attributed to variable and binary stars, and some to different integration times. Figs. 3 and 4 show the comparison between our value and CG's photographic magnitudes and B-V colors. The correlation in V magnitude is very good but not so in B-V color. One source of the scatter is probably that the photographic magnitude and CCD magnitudes are not compatible. The lower quantum efficiency of the CCD chip in blue light, which yields a larger error in B, may also contribute the scatter in B-V.

The magnitude and color distributions shown in Figs 5 to 8 support to the selection criteria for the late F and early G stars. Since, these are

preliminary results, observation is still ongoing; other analyses are incomplete, thus, no conclusion about population density and chemical abundance can be provided at this time.

ACKNOWLEDGMENTS

We wish to thank David Griffin, Donald Platt and Laila Weber, for their participation in observations at NURO, and Dr. Dennis Dawson, for his comments in reading this paper. This study has been supported in part by the an Esther Wagner Grant; a Student-Faculty Collaborative Research Grant, School of Arts and Sciences, Western Connecticut State University; an Arizona Space Grant to the Northern Arizona University, and a CSU/AAUP research Grant to Prof. Phillip Lu.

REFERENCES

Corbally,C.J., and Garrison, R.F., 1988, Astron. J.,95,745
Garrison, R.F., and Corbally,C.J., 1993, Astron. J.,106,2301
Latham, D.W., 1992, in the Large Scale Distribution of Gas and Dust in the Galaxy, L. Blitz, ed
Latham,D.W., Mazeh,T., Davis, R.J., Stefanik,R.P., and Abt, H.A. 1991, Astron. J. 101, 625
Lu, P.K., 1991, in The Formation and Evolution of Star Clusters, Janes ed. ASP Conf. Ser. 13, 581
Lu, P.K., 1991, in The Objective-Prism and Other Surveys, Philip and Upgren eds. VVO Contr. 12, pg. 33
Lu, P.K., 1991a, in "Database for Galactic Strucutre", Hauck, Philip and Upgren eds, L. Davis Press, pg.19
Lu, P.K., 1994b, in Hot Stars at the Galactic Halo, Adelman,Upgren and Adelman eds, Cambridge Univ. Press, pg. 124
Lu, P.K., Miller, J., and Platt, D. 1992, Astrophy. J. Suppl. 83, 203
Landolt, A.U., 1973, Astron. J. 78, 959
Landolt, A.U., 1983, Astron. J. 88, 439
Oort, J.H., 1932, Bull. Ast. Inst. Neth. 6, 249
Oort, J.H., 1960, Bull. Ast. Inst. Neth. 15, 45
Philip, A.G.D., 1974, Astrophys. J. 190,573
Philip, A.G.D., and Lu, P.K., 1989, The Gravitational Force Perpendicular to the Galactic Plane, L. Davis Press, Schenectady, N.Y.
Weinberg, M.D., 1991, Astrophys. J. 365

Fig. 1 Deviation in V for Standard Stars

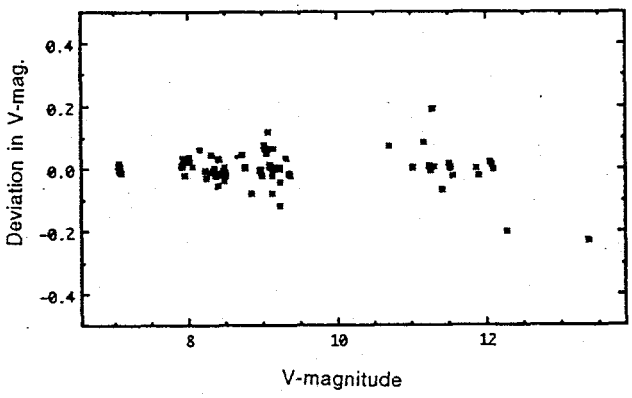


Fig. 2 Deviation in B-V for Standard Stars

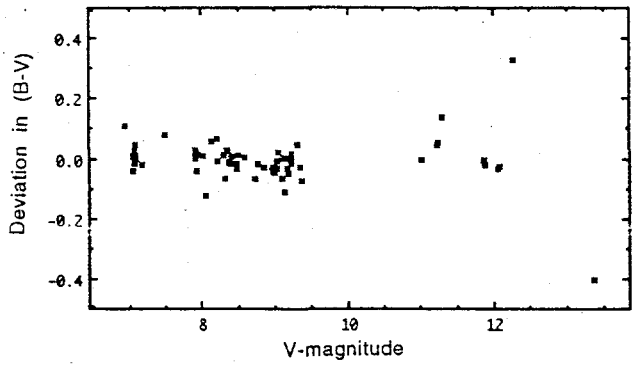


Fig. 3 NURO CCD V-mag. vs Pg V-mag.

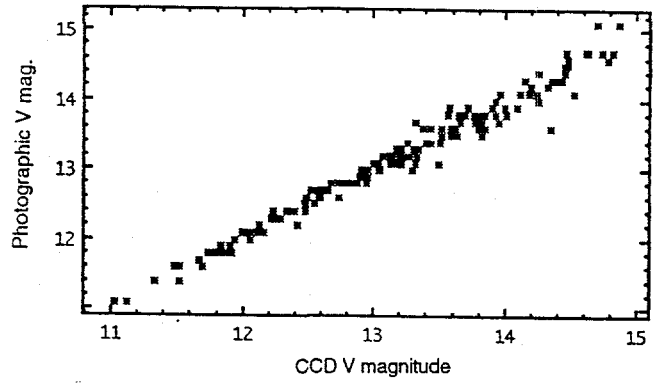


Fig. 4 NURO CCD (B-V) vs Pg (B-V)

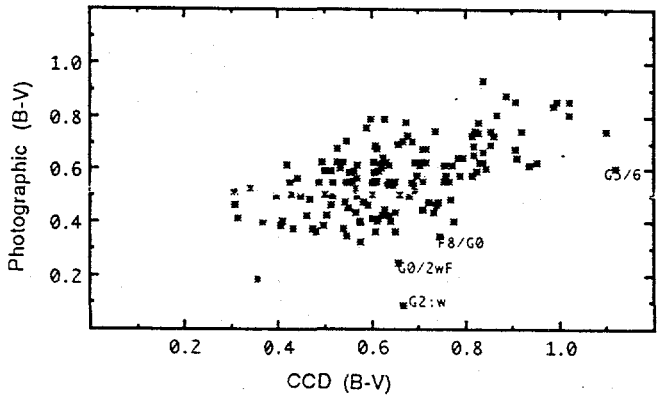


Fig. 5. V-Magnitude Distribution

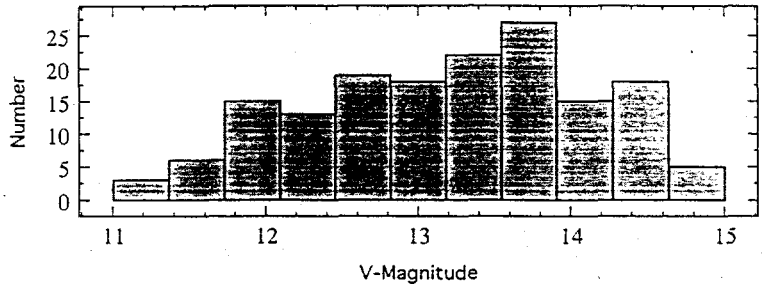


Fig. 6. (B-V) Color Distribution

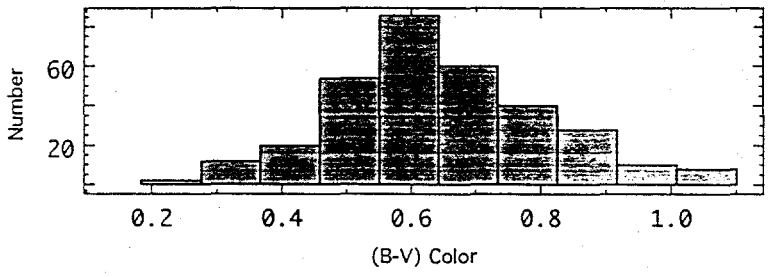


Fig. 7. (V-R) Color Distribution

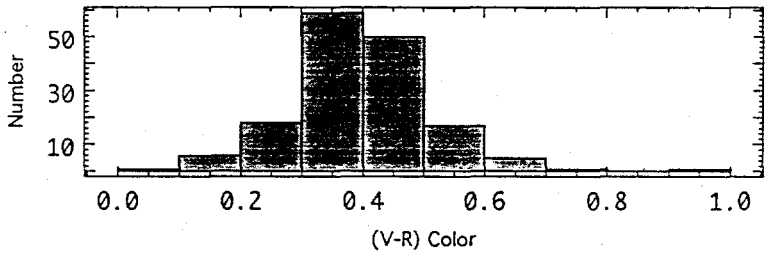
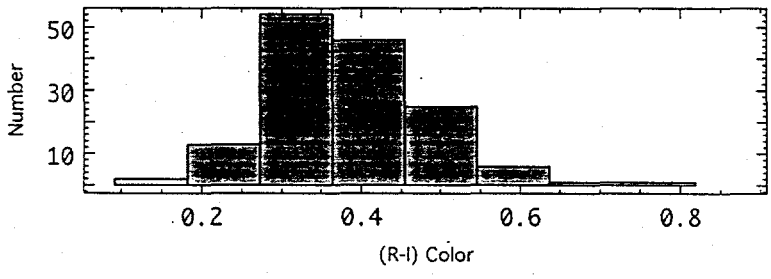


Fig. 8. (R-I) Color Distribution



SPECTROPHOTOMETRIC ANALYSIS OF F AND G-STARS AT THE GALACTIC POLES

Donald Platt

Center for Galactic Astronomy
Western Connecticut State University, Danbury, CT, 06810

Faculty Advisor:

Phillip K. Lu

Slit spectra for the late F and G-stars in the direction of the North and South Galactic Poles have been investigated. Spectral and photometric data in common among the 4 sq-deg. area studied by Corbablly and Garrison (1988,1993), Eriksson (1978) and the 40 sq-deg. area surveyed by Lu, Miller and Platt (1992) were analyzed and compared. Spectral and photometric data of 1500 stars were obtained with the 2D-frutti spectrograph and the CCD image camera with ubvy and H-beta system at the Cerro Tololo InterAmerian Observatory for velocity dispersion studies in the determination of the galactic gravitational force perpendicular to the galactic plane. Spectrophotometric and radial velocity data for these stars are presented and discussed.

INTRODUCTION

The determinations of the total mass density in the solar neighborhood and the distribution of mass in the galactic disk have been studied by Bahcall (1984,b,c) using a joint solution of the Poisson and Boltzmann equations, and by Kuijken and Gilmore (1989, see also Gilmore and Kuijken,1989) using the so-called high altitude potential and high-latitude tracers of G,K-dwarfs. Early investigations by Oort (1932, 1960) led to the first suggestion of missing mass in the galactic disk. The amount of mass per unit volume near the Sun at the galactic plane in unit of Mo/pc^3 is often called the "Oort Limit" in honor of Oort's early attempt.

The recent studies by Bahcall and Gilmore have considerable differences in the results as to how much mass is really missing. Using giant and nearby A and F stars, Bahcall (1984a,b,c) has suggested that about half the total matter density near the Sun is in an unobserved dark matter form. On the other hand, Gilmore and colleagues have suggested there is no significant dark mass associated with the galactic disk. However, a more recently study by Bahcall, Flynn and Gould (1991) has shown that a model with no dark matter is inconsistent with K-giants data at the 87% confidence level, thus further suggested the existence of dark matter.

The problem of determining the local mass density and local column density of the galactic disk is a classical problem in astronomy. The usual approach to this problem is to measure the velocity distribution of a tracer population as function of scale height distance z above the galactic plane. The overall difficulties of selecting tracer stars for the K(z) study have been outlined by King (1989) and others in a workshop held at the Center for Galactic Astronomy (Philip and Lu, 1989).

This project is designed to study a uniformly selected tracer populaion using 1.5 degree thin prism spectra. Candidates from these preselected F and G stars will be further investigated using spectra and radial velocities. Stromgren 4-color ubvy (u = ultraviolet, b=blue, v=violet, y=yellow) and H-beta photometry are used to differentiate high luminosity objects from dwarf main-sequence stars. The limiting magnitude of about 16th magnitude of the stars in this study corresponds to a z-distance of about 1.5 kpc which is necessary for large scale structural study. This magnitude can be obtained with about 5 minutes of integration time using a Charge-Coupled Device (CCD) with the u filter in the Stromgren ubvy system. The u filter is the wavelength region that the CCD is least sensitive to so its integration time is the longest. It requires nearly 30 minutes to obtain a usable 2DF spectra (about 25 cts/pix).

A Catalogue of 3155 stars earlier than G5 stars within a 40 sq-deg area at the SGP is now completed (Lu et al, 1992). Partial radial velocity and photometric results have been reported by Lu and colleagues, (Lu, 1989; Lu, 1990; Lu 1991a,b Platt and Lu, 1993, Lu,1994a,b).

OBSERVATION AND DATA REDUCTION

A thin prism survey centered at the South Galactic Pole (RA= 00h50m, Dec=-27.3) has been completed for a 40 sq-deg area (Lu et al 1992). A total of 3155 candidates of A, F and early G-stars to a limiting mag. of 16.0 were identified. This study was carried out at the Cerro Tololo Inter American Observatory, La Serena, Chile using the CCD Camera/ASCAP (Automated Single Channel Aperture Photometer) with ubvy,H-beta Stromgren filter system and 2D-Frutti spectrograph/ARGUS (Multiple Objects Spectrograph).

At present, a total of 1297 stars now have spectra (or 40%) and 1880 stars have CCD/ASCAP photometry (or 60%) using the 0.9-m and 1.0m telescope. The surge of photometric data was due to the use of a large format 2048 x 2048 CCD on 0.9m during the November 1994 run. With the resolution of about 0.39"/pixel, the quad-mode of the CCD would give about 13.5 arc-minute sky area, which, on average, covered more than 10 F and G stars brighter then 16 magnitude. The normal exposure time is about 2, 1.5, 2, 4 and 8 minutes for Hb,y,b,v and u respectively to reach a 16 mag. star, thus, about 30 minute/field including readout and setup time. The total number of fields per night is about 16 including standard fields. The total number of stars/nt is about 120. A total of 23 nights were allocated to this program on 0.9m and 1-m during October of 1993, November of 1994 to use 2D-Frutti and ASCAP on 1m and large format CCD on 0.9m.

ANALYSIS, RESULT AND CONCLUSION

Spectrophotometric analyses were carried out among the following three databases in photometry and spectroscopy: Corbally and Garrison (1988), Garrison and Cobally (1993); Eriksson (1978) and Lu et al. (1992) plus data obtained after 1992.

1. Photometry

Overall, the comparison between our CCD data on the ubvy,H-beta system and Eriksson's data (Eriksson, 1978) are excellent to good in visual magnitude between photoelectric and photographic data (Fig. 1a and 1b) and good to fair for the (b-y) versus (B-V) colors in the two systems (Fig. 2a and 2b). The scattering is likely attributed to the incompatibility of the two systems, since ubvy, H-beta system is in an intermediate band system (bandpass at about 300A) where UBVRI is a broad band system at about 1000A.

Analyses using CCD data against Corbally and Garrison (1988) photographic data are shown in Figs. 3a and 3b for visual magnitude and b-y versus B-V color. The large scatter in the 3b color plot is obvious in the photographic system, large discordances indicate that their spectral types do not agree with the color indices.

2. Spectroscopy

Spectral classification in the MK system are available in both galactic poles in about a 4 sq-deg area, the correlation between 2DF spectral types versus CG slit spectral types in the SGP are reasonable except for one binary star and the scatter is about 2 to 3 subclasses (Fig. 4a). The correlation between CG's spectral types and CG's photographic (B-V) is shown in Fig. 4b. This plot is similar to the Fig. 9, (b-y) vs 2DF spectral types in Lu et al (1992) except for three large deviations.

Fig. 5a and 5b show the correlations between CG's spectral types vs their photographic B-V and NURO CCD B-V color at the North Galactic Pole.

These two figures suggest that spectral type do not correlate well with color index in either system.

With the current available data only preliminary analyses are presented here. Final analyses and conclusion can not be done between the two Galactic Poles for density distribution and population differencies due to the incompleteness of the data sample. More data to complete this survey will be gathered beginning in the Fall of 1995 with an observing run at CTIO by Professor Lu.

ACKNOWLEDGMENTS

We wish to thank David Griffin, Robert Lanczycki and Laila Weber for their participation in observation at NURO. This study has been partially supported by the Student-Faculty Collaborative Research Grant, School of Arts and Sciences, Western Connecticut State University; Arizona Space Grant to the Northern Arizona University and CSU/AAUP Research Grant to Prof. Phillip Lu.

REFERENCES

Bahcall, J.N., 1984a. Astrophys. J. 276, 156.
Bahcall, J.N., 1984b. ibid, 276, 169.
Bahcall, J.N., 1984c, "Astronomy with Schmidt-Type Telescopes", Capaccioli, ed., Reidel Publ., p241.
Bahcall, J.N., Flynn,C., and Gould, A., 1991, Astrophys. J. 399,234.
Corbally, C.J., and Garrison, R.F., (1988) Astron. J. 95, 161.
Eriksson, P.I., 1978, Uppsula Astr. Obs. Report #11
Gilmore, G., and Kuijken, K. 1989 in "The Gravitational Force Perpendicular to the Galactic Plane", Philip and Lu eds, L. Davis Press,Schenectady, N.Y. pg. 61
King, I.R., 1989, in "The Gravitational Force Perpendicular to the Galactic Plane", Philip and Lu eds, L. Davis Press, Schenectady, N.Y. pg. 147
Kuijken, K. and Gilmore, G. 1989, Mon. Not. Roy. Astron. Soc., 239. 571.
Lu, P. K., 1989, in "The Gravitational Force Perpendicular to the Galactic Plane", Philip and Lu eds, L. Davis Press Schenectady, N.Y. pg. 95
Lu, P. K., 1990, in the ESO/CTIO Workshop on "Bulges of Galaxies", Javis and Terndrup eds, La Serena, Chile pg. 151
Lu, P.K., 1991a, in Symposium on the "Formation and Evolution of Stars Clusters" Janes ed. Publ. Astr. Soc. Pacif.
Lu, P.K., 1991b, In the "Objective-prism and Other Surveys" Philip and Upgren eds. L. Davis Press, Schenectady, N.Y. pg. 33
Lu, P.K., Miller,J. and Platt, D., 1992, Astrophys. J. Suppl. 83,203
Lu, P.K., 1994a, in "Database for Galactic Structure" Hauck, Philip and Upgren eds., L. Davis Press, Schenectady, N.Y. pg. 19
Lu, P.K., 1994b, in "Hot Stars in the Galactic Halo", Adelman and Upgren eds., Cambridge Univ. Press, pg. 124
Oort, J.H., 1932, Bull. Astro. Inst. Neth. 6,.249
Oort, J.H., 1960, Bull. Astro. Inst. Neth. 15, 45.
Philip, A.G.D., and Lu, P. K., 1989, "The Gravitational Force Perpendicular to the Galactic Plane" L. Davis Press, Schenectady, N.Y.
Platt, D., and Lu, P.K., 1993, in the Proceedings of the 7th NURC, pg. 847

Fig. 1a. V vs Eriksson's V (PE)

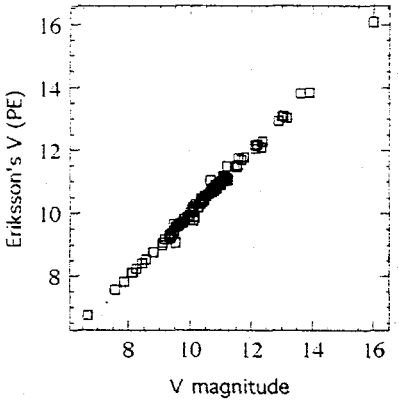


Fig. 1b. V vs Eriksson's V (PG)

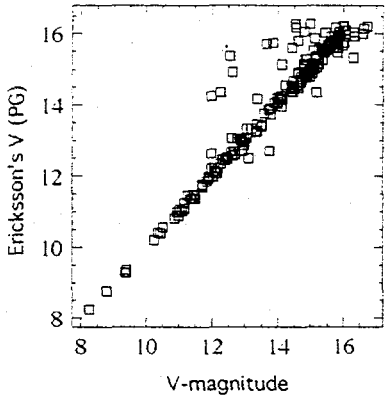


Fig. 2a. b-y vs B-V (Eriksson)

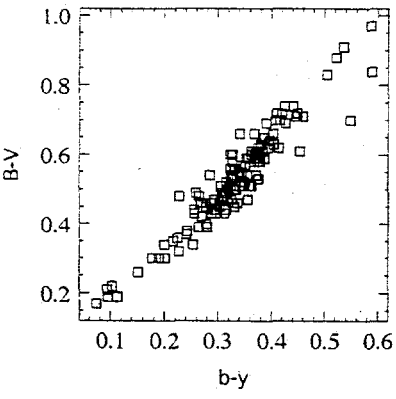


Fig. 2b. b-y vs B-V (Eriksson)

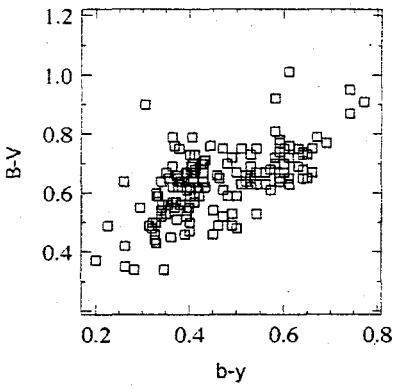


Fig. 3a V vs. V(CG)

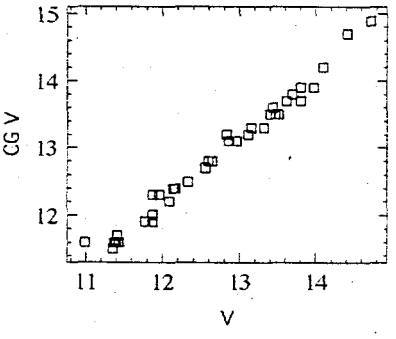


Fig. 3b b-y vs B-V(CG)

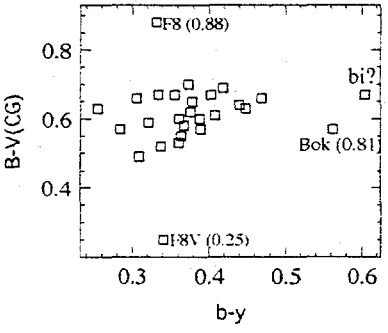


Fig. 4a Spt(2DF) vs Spt(CG)

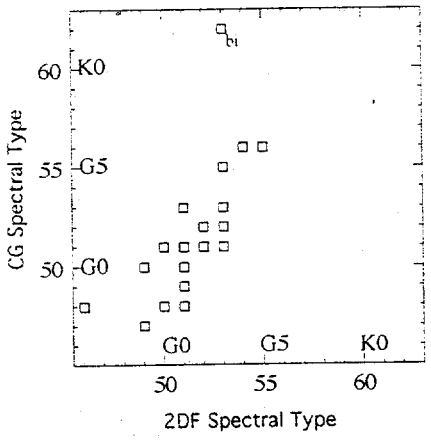


Fig. 4b Spt(CG) vs (B-V)

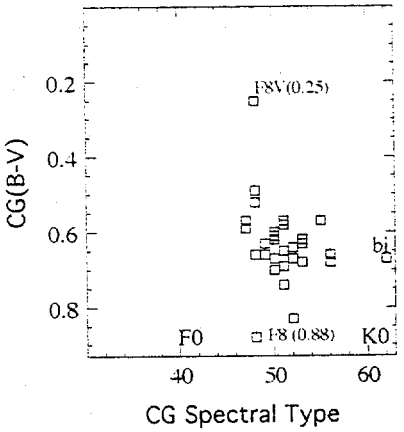


Fig. 5a TP Spt vs Pg (B-V)

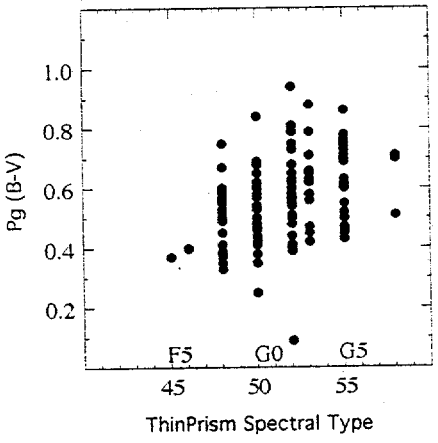
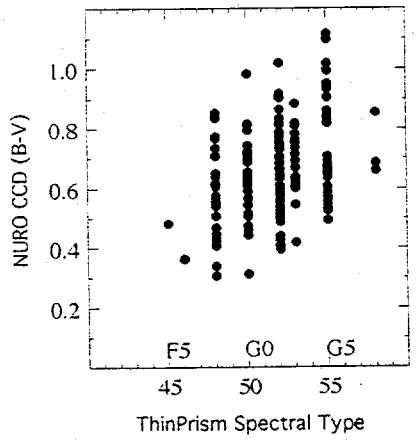


Fig. 5b TP Spt vs CCD (B-V)



AUTOMATIC REAL-TIME PHONEME RECOGNITION

Zac Walton

Department of Physics
Harvard University
Cambridge, MA 02138

NSF Summer Undergraduate Fellow in Sensor Technologies

Faculty Advisor:
J. Van der Spiegel
Center for Sensor Technologies
University of Pennsylvania
Philadelphia, PA 19104

An analog neural network was implemented in hardware for use as a general programmable neural computer. To test the design of the hardware, it was programmed to extract features of the speech waveform for the purpose of phoneme identification. Circuits were designed to detect the following features of the speech waveform: energy onset, energy offset, formant transitions, signal durations, energy amplitude, noise, and voicing. Using a digital computer, these circuits were programmed into the analog neural net. Continuous speech was played at the programmed neural net and the output was monitored by a digital computer. The output was analyzed, and the detection circuits were modified to more accurately extract the desired features. The outputs of the feature detection circuits were then combined to enable the neural net to distinguish individual phonemes.

INTRODUCTION

Despite the extensive efforts of communication companies, computer scientists, and cognitive scientists, there exists no system capable of automatic, real-time, speaker-independent phoneme recognition of continuous speech. The difficulties of this task can be divided into two categories: fundamental ambiguities of the speech waveform and problems related to implementation. Implementation problems will be discussed later in relation to the present system's advantages over previous attempts at phoneme recognition. The fundamental ambiguities of the speech waveform, however, must be understood before any recognition system can be successful.

Fundamental ambiguities and implementation problems

While automatic phoneme recognition has been achieved with varying restraints (extended processing time, limited phoneme set, word isolation, etc.), the conspicuous absence of an automatic dictation system reveals the most obvious ambiguity in the speech waveform. The lack of contextual understanding absolutely prevents any acoustically-based system from distinguishing between different words that have identical pronunciation. Therefore, before computer programmers are able to create a system flexible enough to distinguish the four "there's" in this sentence, they're going to have to limit their recognition scope to what was said, and not what was meant. Furthermore, research has shown that the brain's contextual knowledge permits recognition of greatly degraded speech waveforms, suggesting that in everyday conversation, the acoustical information necessary for the independent identification of each phoneme may not be present (Glass 1986). This phenomenon is evidenced by the fact that when nonsense words are played for a listener, phoneme recognition scores drop dramatically. Thus, any system which does not perform high-level analysis of sentence structure is assured to have recognition accuracy below human rates.

Further fundamental ambiguities in the speech waveform result from the coarticulation of phonemes and speaker dependence. In the most severe cases of coarticulation, the acoustic signal of a phoneme is altered or even deleted to facilitate continuous and efficient motion of the vocal tract. Since the brain is able to perceive coarticulated phonemes by incorporating the characteristics of the preceding and following phonemes, any approach to automatic speech recognition must be able to use previous results to influence the current recognition decision and allow future results to correct the decision. Finally, speaker-dependent phenomena (fundamental frequency, overall tempo, regional accents, etc.) complicate the problem by causing phonetically-identical utterances to have different acoustic signals when produced by different speakers. Since these issues are fundamental to the speech waveform, they must be addressed by any recognition scheme designed to approach human levels of identification accuracy.

Compounding the difficulties associated with the fundamental ambiguities in the speech waveform are problems related to the implementation of a recognition system. In the past, the majority of recognition systems have been software-based and run on digital, serial computers. Even with steadily increasing processing speeds, the enormous amount of computation associated with simulating a neural net or executing a template-matching algorithm prevents serial computers from even approaching real-time results. Furthermore, if such speech recognition software were ever used as the front end of a host serial computer, the host's single processor would be significantly slowed by the additional demands of the recognition software.

A programmable, analog neural computer

To avoid the problems associated with software simulation, a fully-programmable, analog neural computer was built consisting of 1024 neurons. Individual neurons can be connected by an array of programmable switches, and the connection strength can be chosen from a wide range of positive (excitatory) and negative (inhibitory) values. Programmable time constants located on the output of each neuron and throughout the array of switches enable the neural net to store temporal information about a signal. Analog processing and the lack of a system clock enable the hardware to respond to input in real-time, saving hundreds of hours that would be necessary to model the behavior of such complex circuitry. In addition, once the neural net has been programmed for a task, it can be reduced greatly in size and placed inside a serial host machine for use as a front-end input device without requiring any processing time from the host computer. Thus, the architecture of the neural net circumvents the implementation problems associated with software simulations.

Under the present hardware, all speech recognition processing is performed in the analog circuitry of the neural net. However, it is necessary to have a digital computer sample the neurons in order to evaluate the performance of the net. Furthermore, converting a spoken utterance to input for the neural net requires the presence of an artificial cochlea. This additional hardware consists of a bank of sixteen band-pass filters designed to perform a spectral analysis of the input speech waveform.

Extracting features capable of describing phonetic content of the speech waveform

With the described hardware system in place, the task was to determine which features of the speech waveform are important as recognition cues and then to design circuits which respond to such features. Thus, the research project was divided into two sections: researching theories of speech recognition to determine which features to extract and programming the neural net with detection circuits for each of the features. Unfortunately, length restrictions on this paper prevent the enumeration of the entire feature set. The most important conclusion obtained from the research of theories of speech recognition is that there are no true "acoustic invariances" that can be used to identify a phoneme regardless of context. As Liberman observes in his classic article "Perception of the Speech Code," the acoustic signal produced by a speaker can be best understood as a continual transition between phonetic states, as opposed to the discrete expression of invariant acoustical cues. By this reasoning, a recognition system which regards the speech signal as the juxtaposition of independent acoustic elements ignores the most critical aspect of speech production. The fact that the same biological apparatus is used to make adjacent phonemes greatly influences the acoustic qualities of each phoneme. However, the brain is able to filter these acoustical variations and perceive the correct series of phonemes. Liberman explains this phenomenon by attributing the brain's ability to recognize coarticulated speech to its own internal map of the vocal tract. Since the perceiving brain faces the physical restraints of the vocal tract during speech production which lead to coarticulation, it expects the same coarticulation in the speech that it perceives (Liberman 1967). Applying this theory to the problem of automatic speech recognition, the outputs of the feature set will first be combined to give a crude physiological depiction of the characteristics of the vocal tract during the production of the inputted speech. With some knowledge of the time-dependent state of the vocal tract, the neural net should have a better chance of predicting and identifying the altered acoustic signal of coarticulated phonemes. In fact, research with Hidden Markov Models has shown that higher phoneme identification scores are achieved when place of articulation is used as an intermediate step in phoneme recognition (Deng 1992). However, due to time constraints, this research project was limited to the process of extracting features from the acoustic signal; the success of the "recognition by modeled production" technique has yet to be shown.

THE DETECTION CIRCUITS: DIAGRAMS AND PERFORMANCE

After researching theories of human speech recognition, neural net circuits were designed to detect the following features of the speech waveform: energy onset, energy offset, formant transitions, signal durations, energy amplitude, noise, and voicing. Due to length restrictions, only two of the detection circuits will be presented. The following

sections depict the design and performance of the onset/offset detectors and the noise/voicing detectors.

The onset /offset detectors

The onset and offset detectors consist of sixteen neurons, each receiving two inputs from one of the sixteen bands passed by the cochlea. Each neuron receives an excitatory and an inhibitory connection from one band and a time constant is placed on the excitatory connection for the offset and on the inhibitory for the onset (Figure 1). The output of the detector is a spike at the beginning or end of a signal in a single band of the input.

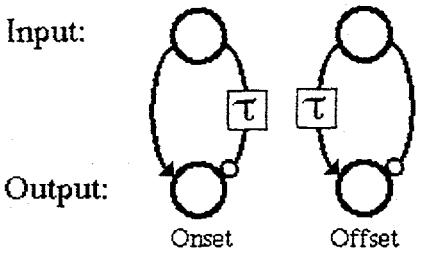


Figure 1:
Onset and offset detector circuit design

Figure 2 shows the output of the two detectors graphed above the input with time on the x-axis and frequency on the y-axis. The row of data labeled "Input" represents the output of the cochlea when played a constant tone followed by the triphone "/g/ /r/ /i/," as in "*greed*."

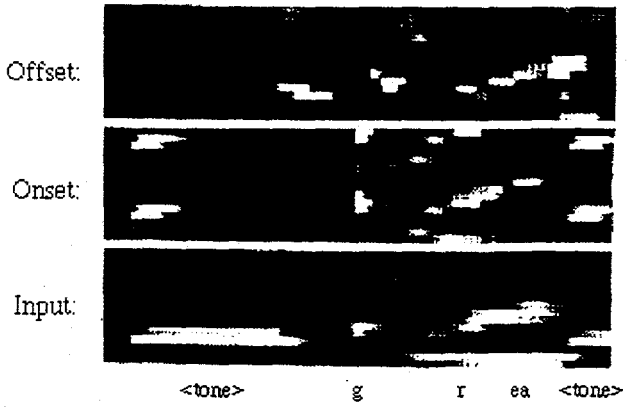


Figure 2:
Onset and offset detector performance

The noise/voicing detector

Two approaches were tested for the task of detecting voicing. The first involved designing a circuit to act as a band pass filter that signaled amplitude modulations in the frequency range associated with voicing. A second, more straightforward approach simply measured the activity in the two lowest bands passed by the cochlea which usually signal the presence of voicing. Despite the first method's sophisticated analysis of the speech signal, it was unable to distinguish the fixed amplitude oscillation of voiced speech from the random amplitude oscillation of noise. Thus the second circuit, which consists of two excitatory connections and an integrating time constant, was selected to detect voicing (Figure 3). To supplement the set of noise detectors described previously, the original speech signal was inhibited with the voicing detector, producing a single detector sensitive to noise.

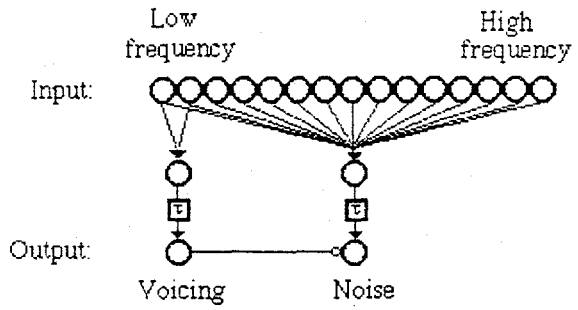


Figure 3:
Voicing/noise detector circuit design

Figure 4 shows the output of both the voicing and noise detectors when played the same phrase: "brush fires are common in the dry underbrush." Unlike the first set of noise detectors, this single noise detector is able to signal all the fricative and plosives, including the relatively weak bursts of the /b/ and /d/.

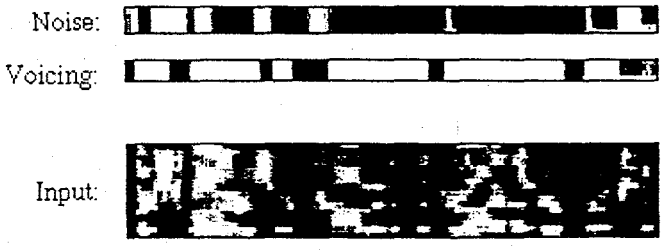


Figure 4:
Voicing/noise detector performance

CONCLUSION

Due to the extreme flexibility of the neural net hardware, the programmed detection circuits were able to signal the entire desired feature set. Problems inherent to the field of analog processing were successfully managed by either the manual adjustment of the circuit's parameters or the adoption of a new approach and a redesign of the detection circuit. Despite occasional disparities between the saturation voltages of the individual neurons, the neural net hardware elements performed predictably and reliably. The only design change that would enable more efficient feature extraction would be the employment of dischargable time constants which would prevent the necessity of multiple duration detectors. Thus, with the critical task of feature extraction completed, the next step towards the goal of phoneme recognition consists of connecting the outputs of the detection circuits either manually or with a back propagation algorithm. If the feature set has been chosen properly, this combination process will lead to the goal of automatic, real-time phoneme recognition.

ACKNOWLEDGMENTS

This research was conducted in support of the doctoral dissertation of Chris Donham. The author would like to thank him for involvement in this fascinating application of the neural net hardware. The author would also like to thank Dr. Paul Mueller and Dr. Jan Van der Spiegel for their advice and opinions on each step of this project. Finally, the author would like to thank Lois Clearfield for providing her time and resources to ensure the success of the SUNFEST program.

REFERENCES

Deng, L., and K. Erler, Structural design of hidden Markov models speech recognizer using multivalued phonetic features: Comparison with segmental speech units, *Journal of the Acoustical Society of America*, 92 (1992) 3058.

Glass, A., and K. Holyoak, *Cognition*, Random House, New York, 2nd ed., 1986, pp. 429-430.

Liberman, A., Perception of the speech code, *Psychological Review*, 74 (1967) 431.

The Measurement of Alkali Doped C_{60} Spectra in the Visible Range

Alfred Badowski
Physics Department
Rensselaer Polytechnic Institute
3005 Hunt II at RPI
Troy, NY 12180

Faculty Advisor:
Laszlo Mihaly
Department of Physics
SUNY Stony Brook
Stony Brook, NY 11794

This project was undertaken in the hopes of extending the frequency range available in the study of fullerenes to the visible and gaining insight into these structures using both infra-red and visible ranges.

Buckminsterfullerene, the object of this study, is a large ball-like molecule composed of 60 carbons. The shape, consisting of 20 hexagonal and 12 pentagonal faces, yields a stable configuration somewhere between graphite and diamond. (see fig. 1) It first came to the scientific community's attention through astrophysical studies of interstellar gas and dust. Initial interest was curbed by the very slow and inefficient process for its production. Recently, however, with the advent of new production techniques and the discovery of its superconducting properties when doped, the molecule has received much attention and publicity.

The ball of the molecule is 7.4 Å in diameter and the molecules reside in a crystal with face-centered cubic structure where the distance between molecular centers is roughly 10 Å. (see fig. 2) [1] Pure Carbon 60 is not a natural conductor because of its lack of electrons in the conduction band. Doping it with alkali metals such as potassium, cesium or rubidium adds the required electrons into the conduction band in order for it to become a semi-conductor. At low enough temperatures the alkali doped C_{60} becomes superconducting and with only two elements it is the simplest superconductor currently available.

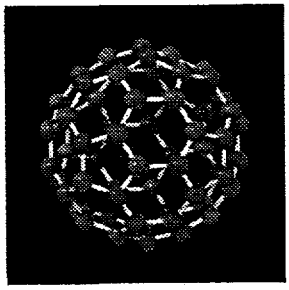


fig. 1

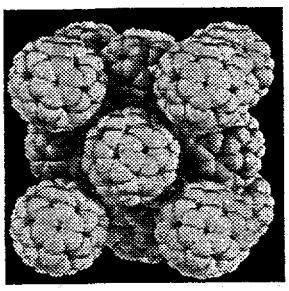


fig. 2

Since the molecules are so large the spaces between them easily accommodate impurities and the structure is easily doped. In the face-centered cubic lattice there are three natural doping sites per molecule, two tetrahedral with four neighboring C_{60} molecules and one octahedral with six neighboring C_{60}'s. In order to have doping higher than A_3C_{60}, where A stands for an alkali metal, the crystal changes structure to body-centered tetragonal for A_4C_{60} and finally to body-centered cubic for doping up to A_6C_{60}.[2] Substances of A_xC_{60} for which x>3 are insulating, however.[3]

There were two systems used in this study. The main instrument used in the visible range study was a modified Perkin-Elmer double pass single beam monochrometer which was integrated with a personal computer (PC). A commercial Bomem Infra-Red Fourier Transform Spectrometer was used to provide additional data in the infra-red.

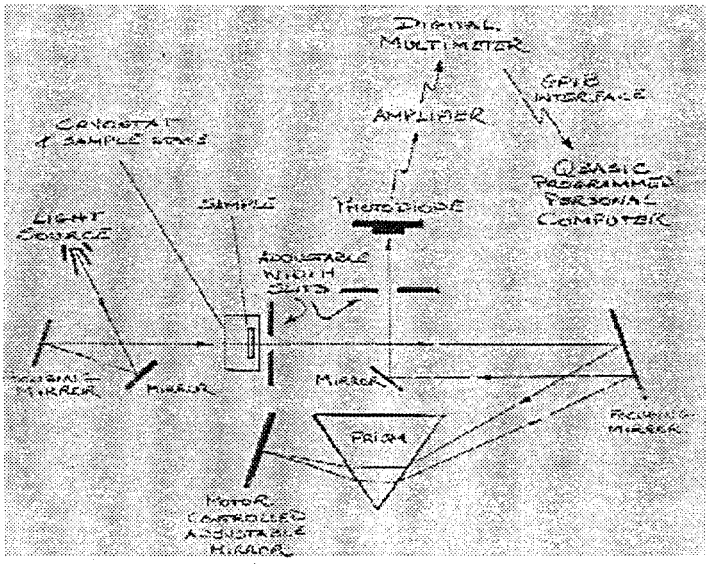


fig. 3

The Perkin-Elmer was automated using a Tandon 3.6deg/step stepping motor driven by input from the parallel port of a PC. The photodiode was connected to a Keithley digital multimeter through a Keithley current amplifier. The multimeter was interfaced with a PC through a National Instruments' General Purpose Interface Bus (GPIB). The data acquisition and instrument control aspects of the system were brought together under the command of a single PC using original code written in Microsoft's Q-BASIC. This enabled the autonomous collection of spectra with minimal supervision. (see fig. 3 for a schematic of the system) The monochrometer was also fitted with a cryostat mount on the sample stage for optional temperature control.

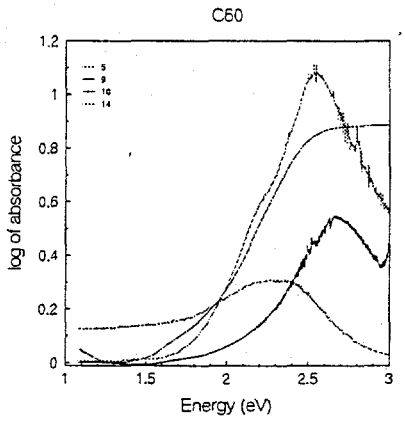


fig. 4

The system was calibrated using standard known light sources such as neon, argon, mercury, and sodium lamps as well as a helium-neon laser and a infra-red diode laser.

Samples were created by evaporating pure C_{60} onto glass slides in a vacuum. Doped samples were prepared by exposing pure C_{60} slides to alkali vapor. Pure samples are brownish-yellow, whereas doped samples appear more brown with stronger doping.

Successive studies of pure C_{60} resulted in better optimization of the instrument due to constant fine tuning of the setup. Figure 4 shows the progress towards an actual C_{60} spectrum for energies between 1 and 3eV for trials 5 through 14. The instrument was more sensitive in the red end of the spectrum primarily due to the photo diode's efficiency profile and the type of light source. Trial fourteen clearly shows the increasing noise at higher energies. For better signal to noise ratio the instrument could be tuned to certain frequency ranges. These ranges were very short, however, and producing one continuous spectrum was difficult.

Study of Rb_1C_{60} revealed that a second hump developed at 1.2eV. This is due to the appearance of the Rb electron in the conduction band.

The use of the Perkin-Elmer along with the commercial Bomem provides

the laboratory with the ability to examine fullerenes in an extended range of energies. Figure 5 shows a compound spectrum produced using both instruments.

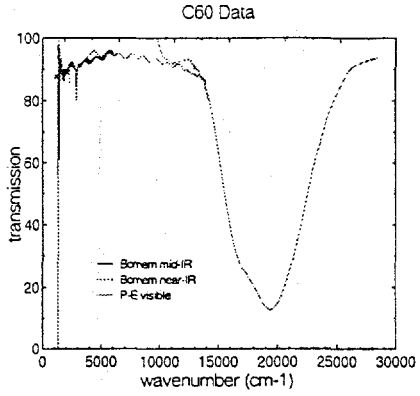


fig. 5

References

1. Huffman, Donald R. Solid C_{60}
PHYSICS TODAY November 1991, p.22

2. Hebard, Arthur F. Superconductivity in Doped Fullerenes
PHYSICS TODAY November 1992, p.26

3. *Ibid., p.27*

FIGURES 1 and 2 courtesy of
Michael C. Martin
Dept. of Physics SUNY Stony Brook

A STUDY OF LEIDENFROST PHENOMENA USING DIGITIZED VIDEO IMAGE CAPTURE AND ANALYSIS

Katarina Plavetic

Physics Department, Guilford College 5800 West Friendly Avenue
Greensboro, NC 27410

Faculty Advisor: Rexford E. Adelberger

When placed on a very hot surface, liquid droplets evaporate slowly while floating on their own vapor film. The motion of vapor lying between the hot surface and the liquid droplets causes vertical and horizontal oscillations of droplets. This motion, called the Leidenfrost Phenomenon after J. G. Leidenfrost, the first person to study this process in 1756, was studied using a CCD video camera and FusionRecorder™ software to analyze the shapes of droplets as they change over time. Some droplets showed distinctive vertical oscillations while others did not. We found that droplets of greater size had a greater number of lobes. The frequencies of oscillations depended upon the number of lobes and the radius of the drops. Our qualitative observations and calculations using a model proposed by Casal and Gouin[1] indicate that drops of greater size and with greater number of lobes had smaller frequencies. These results disagree with observations and estimates by Glasscock and Holter.[2]

[1] Pierre Casal and Henri Gouin, 1994, "Vibrations of Liquid Drops in Film Boiling Phenomena" (translated by G. A. Maugin), *Int. J. Engng. Sci.*, **32**, 10, pp. 1553-1560

[2] Norman J. Holter and Wilford R. Glasscock, 1952, "Vibrations of Evaporating Liquid Drops", *The Journal of The Acoustical Society of America* **24**, 6, pp. 682-686

The Leidenfrost Phenomenon occurs when various liquids are placed on a sufficiently hot surface, floating on a film of their own vapor. The vapor film protects and insulates droplets from touching the hot surface, causing them to evaporate very slowly (on the order of minutes). However, if the surface is cooled, the insulating vapor film is not maintained and droplets contact the hot surface, adhere to the surface and evaporate within seconds.[1]

J. G. Leidenfrost, a German medical doctor, was the first person to study the boiling of small liquid masses and to observe this effect.[2] When performing his experiment, Leidenfrost was interested in the length of time that droplets would last once placed on a hot surface. He noticed that at lower temperatures, droplets would not last very long; they would adhere to the surface and evaporate very quickly. At high temperatures, however, droplets would form the vapor film and take a longer time to evaporate.

Many similar experiments have been performed since Leidenfrost did his pioneering work. Jearl Walker did experiments measuring the Leidenfrost points, the surface temperature at which the droplet evaporation time was the greatest, for various liquids such as vodka, white vinegar and water. He also determined the approximate Leidenfrost points for distilled and tap water.[4] Gottfiried, Lee and Bell did experiments determining the total evaporation times for small droplets of water, carbon tetrachloride, ethanol, benzene and n-octane [5]. In addition, some experiments explored different hot surfaces and their effects on Leidenfrost
Phenomena. Other experiments were conducted exploring the so-called inverse Leidenfrost Phenomenon, where hot metal particles remain suspended by a vapor generated above a cool liquid surface. [6]

Few experiments have been done exploring the shapes of droplets as they

vibrate during the Leidenfrost Phenomenon. Holter and Glasscock were among the first researchers to notice that during Leidenfrost Phenomena, droplets also undergo an unusual type of vibration.[1] They observed droplets in a spheroidal state: the liquid state under the influence of the vapor film, gravity, surface tension and heat.[1] The drops were round in the horizontal plane and flattened spheroids in the vertical plane. They observed that the vibrating droplets are approximately polygons with 2, 3, 4, ... or n sides, appearing to a naked eye in several general shapes: sausage shaped (the basic mode), triangles, squares, pentagons, etc.[1] (see Figure 1)

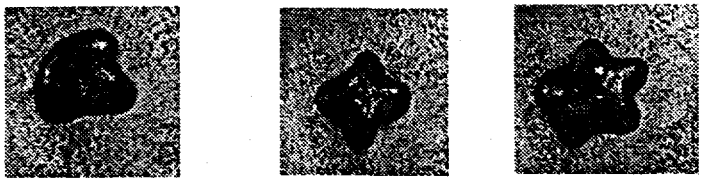


Figure 1: Actual shapes of water droplets as recorded with CCD camera

Piere Casal and Henri Gouin pursued Glasscock's and Holter's research by observing similar properties of Leidenfrost Phenomenon and presenting a computer model of vibrating motions of the drops. Further, using Equation 1, they calculated the frequencies of the vibrations as a function of the number of lobes in droplets.[7]

$$\nu = \frac{1}{2\pi}\sqrt{\frac{(n+1)\,g}{R}} \qquad (1)$$

where n = number of lobes, R = horizontl radius of vibrating droplets, g = gravity constant.

THE EXPERIMENT

This experiment was a continuation of Glasscock and Holter's experiment, including some changes in experimental set up and types of equipment used. The main interest in this research was to explore the changes of shapes of droplets as a function of time as they undergo the Leidenfrost Phenomenon. We wanted to simultaneously examine the motion of the droplets in the plane of the hot surface and perpendicular to the surface. To accomplish this, the experiment and equipment were set up as shown in Figure 2. The mirror was set at approximately 45° to the hot surface so that the recording CCD camera could obtain the images of horizontal and vertical oscillations of the droplets at the same time.

To capture the images of oscillating drops, we used a CCD video movie camera (30 frames per second) with a high speed shutter (1/1000 sec). Those images were digitized using a RasterOps 24TV frame grabber and analyzed on the computer using FusionRecorder™ software. This gave us the opportunity to examine in detail the horizontal and vertical shapes of droplets every 1/10th of a second.

A CCD video camera was used to record the movement of droplets in order to gain digitized images for further computer analysis. CCD stands for Charged Coupled Device. A CCD is a silicone chip which contains tiny cells called photosites or pixels (short for picture element) which convert the light that strikes them into an electric (digitized) signal. The chip is a photosensitive detector.[8]

To obtain color images, photoelements must be alternatively covered with very small colored filters (usually red, green and blue). Information collected from the three different filters is then separated by processing of the video signal

resulting in three distinct images for each filter. The images are then electronically combined into a final color image.[8] A color image can then be visualized on the video or on the computer screen.

FusionRecorder™ software was used to make Quick Time™ movies. The video output of the camcorder was connected to the video input of the Macintosh computer with FusionRecorder™ application. The recording of droplet motions was done with a high speed shutter setting and the FusionRecorder™ frame speed for making Quick Time™ movies was set to take the best number of frames corresponding to the frame speed at which the images were obtained. To avoid frame dropping (when video frames are skipped or not displayed) and to choose the best frame speed, we took images of a running stop watch and made sure that each frame showed a successive time. In addition, the Quick Time™ movies produced from the video tape and those produced from the video image memory were checked to make sure that both used the same frame speed and that no frame dropping occurred for either case. Quick Time™ movies were recorded using a smaller size (3/8 of a normal full screen size) recording window to ensure faster frame rates and gain smoother motion. Quick Time™ movies made it possible to analyze the motion of droplets frame by frame as well as by individual shapes. In addition, the motion could be slowed down or played backward and forward.

EXPERIMENTAL PROCEDURE

Preparing the surface

The hot plate was smoothed with sandpaper, cleaned and greased with Dow Corning high vacuum grease. Sanding of surface was important to reduce any of irregularities, rough scratches or bumps present on the surface that could cause vibrating droplets to break apart. Coating of the surface with grease helped to get droplets in a vibrating mode and decreased the breaking and adhering of water droplets to the metal surface of the hot plate. It was important to cover the surface with grease while it was still cool, because the grease would not adhere when the surface was hot. We found that hot plates with ceramic tops did not produce the Leidenfrost Phenomena.

A small concavity was present at the middle of the hot plate to ensure that drops wouldn't fly off the surface. The hot plate was heated up to a temperature at which water droplets would not adhere to the surface, but would start bouncing and would remain in spheroidal shape.

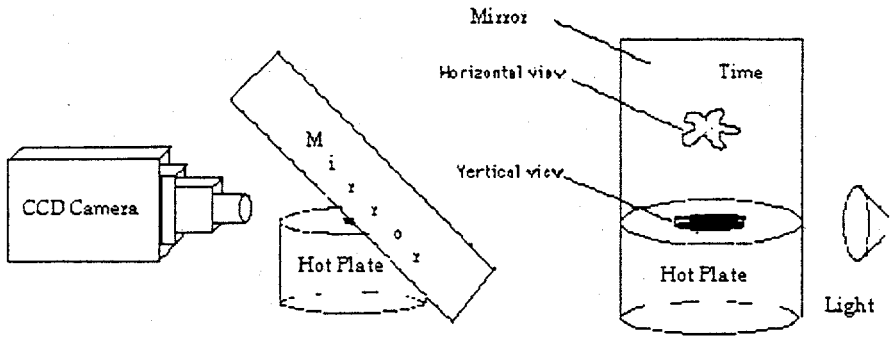
Forming droplets and size

Water used in this experiment was dyed with methylene blue dye to enhance the visibility of droplets during recording with CCD camera and later during analysis. Droplets were formed using a small plastic dropper. The size of droplet was determined by the number of small drops that were added on the surface of hot plate forming a single large drop. Several sets of data were taken for different sizes of droplets. The sizes ranged from 7 drops to 30 drops (~0.4 ml to ~1.5 ml) forming a single oscillating droplet. In a short time (approximately 30 seconds to a minute) following the formation of the large droplet, the droplet would start vibrating with some translational movement.

Recording

The recording was performed in a dark room with two side lights pointing towards the hot plate and droplets. A CCD camera on NTSC (the color-encoding method with 30 frames per second at a resolution of 640x480 adopted by the National Television System Committee in 1953) standard was used for recording of oscillating droplets. The camera was set in macro mode for close up recording (up to five inches away) and the shutter was set on the high speed (1/1000 seconds) for fast moving motion. Final focusing was controlled manually.

To record the droplet from horizontal and vertical view, the experimental set up was built as shown in Figure 2. Droplets were constructed on the hot surface. The evolution of droplet's motion and vibrations was examined and recorded. Then the droplet was removed and the next droplet was constructed.



a) Side view b) Front view

Figure 2: The experimental set up showing a) Side view and b) Front view.

Analysis

Upon recording, the images were digitized using the Rasterops 24TV frame grabber board and FusionRecorder™ applications on Macintosh Computer. Quick Time™ movies were made to analyze the motion of vibrating droplets.

RESULTS PRESENTATION AND DISCUSSION

As observed and described by Glasscock and Holter, we noticed and recorded vibrating droplets in shapes of polygons with two and more lobes. It was noticed that droplets of greater size had greater number of lobes as listed in Table 1.

Table 1: Measured radii and drop sizes with corresponding calculated frequencies and observed numbers of lobes during oscillations.

Drop size (# of drops)*	7	10	15	20	25	30
Radius of droplets (cm)	0.8	1.0 - 1.2	1.5	1.8	2.5	3.0
# of lobes observed	3 and 4	4	5	5 and 6	7	7 and 8
Calculated Frequency (Hz)	12.5	10.6	10.0	9.8	8.9	8.6

* 1 drop = 0.05 ml of liquid

Examining Quick Time™ movies, we observed that droplets ranging in size from 15-20 drops would start their vibrations by at first vibrating in 2 lobe shape, gradually developing into a three lobe shape, followed by four or five lobe shapes in which they would then continue vibrating. This gradual development was not observed with large size droplets (30 drops and more). With large size droplets (30 and more drops) the vibrations were observed on the edge of the droplet while the center mass of the droplet was stationary. Further more, the greater size droplets had less translational motion on the hot surface.

It was observed that during vibrations, droplets often formed a more or less flattened spheroidal shape and then went back into a vibrating mode forming lobe shapes. Return to the spheroidal shape can be explained by the competition between the surface tension trying to keep the spherical drop shape and the vapor

escaping from under the drop lowering the pressure on the edge of the drop and causing oscillations. In addition, it was observed that if the droplets changed the number of lobes, this happened during this transitional - spheroidal stage of evolution. It appeared that the shape of the drops with many lobes was the "unstable shape" with no fixed lobe shape oscillating or rotating for the entire life of the droplet.

Some droplets showed distinctive vertical oscillations while others remained the same height during oscillations. Changes in vertical height of the drop were found to be dependent upon the size of the drops as well. The large drops showed little change in height as the shape changed during vibrations, while the smaller drops showed a more pronounced height change. The small drops were much flatter when the lobe shapes were dominant and would increase in height during spheroidal stage. This supports our contention that the vapor passing along the edge of the drop is causing the lobe shapes. The central part of the large drops was not affected by the lobes appearing along the edge of the drop, but for smaller drops appearance and disappearance of lobes had a greater effect on central part of the drop causing more pronounced height changes. Therefore, the oscillations were not oscillations of the whole drop, but effects along the edge of the drop.

Using the Equation 1, we calculated the frequencies of different size droplets. Those values are listed in Table 1. Notice that as the size of droplets increased the frequencies of oscillations decreased. It is interesting, that the calculated values do not correspond to values observed and estimated by Glasscock and Holter.

During oscillations, tiny drops were ejected at high speed by the large droplets.

ACKNOWLEDGMENTS

The author would like to thank the Jeglinski family for sponsoring this research project, Dr. Rexford Adelberger for guidance, patience and unlimited support, and physics students at Guilford College for support and help.

ENDNOTES

[1] Norman J. Holter and Wilford R. Glasscock, 1952, Vibrations of Evaporating Liquid Drops, *The Journal of The Acoustical Society of America* 24(6), pp. 682-686

[2] Leidenfrost J. G:, De aquae communis nonullis qualitatibus tractatus - On the fixation of water in divers fire. A tract about some qualities of common water (translated into english by Wares C. 1966), *Int. J. Heat Mass Transfer.* 9, pp. 1153-1166

[3] Thimbleby Harold, 1989, The Leidenfrost phenomenon, *Phys. Education* 24, pp. 300-303

[4] Walker Jearl, 1977, The Amateur Scientist, *Scientific American* 237(2), pp. 126-131

[5] Gottfried B. S., Lee C. J. and Bell K. J., 1966, The Leidenfrost phenomenon: film boiling of liquid drops on a flat plate, *Int. J. Heat Mass Transfer.* 9, pp. 1167-1187

[6] Hall R. S., Board S. J., Clare A. J., Duffey R. B., Playle T. S. and Poole D. H., 1969, Inverse Leidenfrost Phenomenon, *Nature* 224, pp266-267

[7] Pierre Casal and Henri Gouin, 1994, Vibrations of Liquid Drops in Film Boiling Phenomena (translated by G. A. Maugin), *Int. J. Engng. Sci.* 32(10), pp. 1553-1560

[8] Christian Buil, CCD Astronomy - Construction and Use of an Astronomical CCD Camera, Willmann - Bell, Inc., Richmond, 1991

Transverse Relaxation Time in DMS Crystals: A Theoretical Study

J. Matthew Kreider

Turner Laboratory
Department of Physics
Goshen College
Goshen, IN 46526

Faculty Advisor:

Carl S. Helrich

A quantum mechanical first-principles calculation of the transverse spin relaxation time, T_2, in Diluted Magnetic Semiconductors (DMS), is presented. The Electron Paramagnetic Resonance (EPR) spectrum for bulk and layered DMS is shown to be a Lorentzian, an experimentally-observed result. For bulk DMS at high densities, T_2 is found to be quadratic in the magnetic ion atomic fraction x, a result that has been confirmed experimentally. A mathematical parameter, ε, is introduced when finding the spectrum, and it is argued that ε is proportional to temperature.

INTRODUCTION

A theoretical study of the transverse spin relaxation time, T_2, in Diluted Magnetic Semiconductor (DMS) crystals, was done, starting from the basic quantum mechanics involved. The theory is developed first for bulk crystals, and is then expanded to deal with layered structures.

DEVELOPMENT OF THEORY

Bulk Crystals

The goal is to iteratively solve the Heisenberg equation,

$$i\hbar\frac{\partial}{\partial t}\hat{\mu}_{xH} = \left[\hat{\mu}_{xH}, \hat{H}\right], \tag{1}$$

for $\hat{\mu}_{xH}$, to use $\hat{\mu}_{xH}$ to predict the EPR lineshape, and to obtain T_2 from the lineshape. $\hat{\mu}_{xH}$ is the magnetic moment operator in the time-dependent Heisenberg representation. The spin Hamiltonian, which contains only total spins on the respective ions, is[1]

$$\hat{H} = \hat{H}^Z + \hat{H}^{Dip} + \hat{H}^J + \hat{H}^{DM} + \hat{H}^I, \tag{2}$$

where $\hat{H}^Z$ is the Zeeman Hamiltonian caused by the external magnetic field, $\hat{H}^{Dip}$ is the dipole Hamiltonian, $\hat{H}^J$ is the isotropic Heisenberg exchange Hamiltonian, $\hat{H}^{DM}$ is the anisotropic Dzyaloshinski-Moriya (DM) exchange Hamiltonian, and $\hat{H}^I$ is the nuclear-spin Hamiltonian. In terms of s-operators we have (in the Schrödinger representation):

$$\hat{H}_S^J = \sum_{i \neq j} \xi_i \xi_j J_{ij} \hat{s}_i \cdot \hat{s}_j \tag{3}$$

$$\hat{H}_S^{DM} = \left(\frac{-i}{2}\right) \sum_{i \neq j} \xi_i \xi_j \begin{Bmatrix} D_{(ij)z}\left(\hat{s}_{i-}\hat{s}_{j+} - \hat{s}_{i+}\hat{s}_{j-}\right) - D_{(ij)-}\left(\hat{s}_{iz}\hat{s}_{j+} - \hat{s}_{i+}\hat{s}_{jz}\right) \\ + D_{(ij)+}\left(\hat{s}_{iz}\hat{s}_{j-} - \hat{s}_{i-}\hat{s}_{jz}\right) \end{Bmatrix} \tag{4}$$

$$\hat{H}_S^Z = \sum_i \xi_i \hat{s}_{iz} \tag{5}$$

$$\hat{H}_S^I = \sum_{n,j} \xi_n \xi_j A \hat{I}_n \cdot \hat{s}_j \tag{6}$$

J_{ij} and D_{ij} are functions of the crystal. A is a constant. ξ_i is the probability of a magnetic site i being occupied.

The magnetic moment operator, in the Schrödinger representation, is

$$\hat{\mu}_{xS} = \sum_i \xi_i \hat{s}_{ix} . \tag{7}$$

After transforming to the interaction representation and nondimensionalizing, the Heisenberg equation becomes

$$i \frac{\partial}{\partial t} \tilde{\mu}_x = \left[\tilde{\mu}_x, \lambda \tilde{H}^P\right], \tag{8}$$

where $\tilde{H}^P$ is the perturbation Hamiltonian in the interaction representation, and λ is a nondimensional constant.

$$\tilde{H}^P = \tilde{H}^J + \tilde{H}^{DM}. \tag{9}$$

$\hat{H}^Z$ was transformed out of the Hamiltonian when the equation was transformed to the interaction representation. Experimental results indicate that the effects of $\hat{H}^{Dip}$ on T_2 are negligible[2], so $\hat{H}^{Dip}$ was not included in the perturbation Hamiltonian. $\hat{H}^I$ is responsible for hyperfine structure at very low magnetic ion densities, but has little effect on regions of high magnetic ion density, so $\hat{H}^I$ was not included in the perturbation Hamiltonian.

The operators were then represented in a 36 dimensional basis of the spin states. The nondimensionalized Heisenberg equation was solved for $\hat{\mu}_{xII}$, to the

second order in the time perturbation, for a bulk DMS. For a bulk DMS, $Z_{1(i)} = 12x$ (on average), where x is the magnetic ion density and $Z_{1(i)}$ is the number of nearest neighbors of the site i.

The theoretical calculation for the EPR lineshape, $I(\omega)$, is[3]

$$I(\omega) = \frac{1}{2\pi} I(\omega) I^*(\omega), \tag{10}$$

where

$$I(\omega) = \int_0^\infty e^{i\omega t} \hat{\mu}_{xH} dt \tag{11}$$

and where $I^*(\omega)$ is the complex conjugate of $I(\omega)$.

Energy from the microwave, which is absorbed by the spins on the ions in the crystal, usually raises a spin state of lowest energy to the next lowest energy level. So the matrix element to consider is $\left\langle \frac{5}{2}, \frac{3}{2} \middle| \hat{\mu}_{xH} \middle| \frac{5}{2}, \frac{5}{2} \right\rangle$.

To ensure that the integral vanishes at $t = \infty$, it is necessary for ω to have a real as well as an imaginary part. We define

$$\omega = \omega_R + i\varepsilon, \tag{12}$$

where ω_R is the real part of ω and ε is the imaginary part of ω. The integral becomes

$$I(\omega) = \int_0^\infty e^{i(\omega_R + i\varepsilon)t} \mu_{xH} dt. \tag{13}$$

$I(\omega)$ is normalized, so that the area under the line is equal to one. $I(\omega)$ is fit with a Lorentzian, and the formula of a Lorentzian is used to solve for T_2.

Layered Structures

In layered structures, all the magnetic sites are occupied, so $x = 1$. $Z_{1(i)} = 12$ for magnetic sites inside an epilayer. $Z_{1(i)} = 8$ for magnetic sites on the border of a magnetic layer. $Z_{1(i)} = 4$ for magnetic sites in a monolayer. So, in general,

$$Z_{1(i)} = 12 f_{12} + 8 f_8 + 4 f_4, \tag{14}$$

where f_{12} is the fraction of magnetic sites that have 12 nearest neighbors; f_8 is the fraction of magnetic sites that have 8 nearest neighbors; and f_4 is the fraction of magnetic sites that have 4 nearest neighbors.

The EPR spectrum and T_2 are found in the same way they are found for the bulk crystal.

DISCUSSION OF RESULTS

For all the graphs, values of $J = 10K$, $D = 0.05J$, and $\lambda = 0.005$ were used. These are accepted values for Mn^{2+} in a Zinc blend crystal.

T_2 did not vary greatly with respect to J or D.

The normalized theoretical EPR spectrum for the bulk crystal is shown in Figure 1.

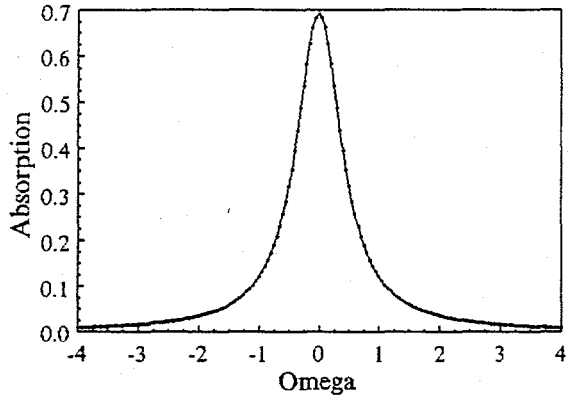


Figure 1: EPR Spectrum (theoretical)

The results confirm that the theoretical EPR spectrum is indeed a Lorentzian. A Lorentzian was obtained for bulk crystals and for layered structures, as well as for the monolayer.

The theory predicts that for bulk DMS crystals at high densities, T_2 is quadratic in x, a result that has been confirmed experimentally. The graph of T_2 vs. x is shown in Figure 2. The points are the experimental data[4] and the curved line is the theoretical fit. ε was chosen arbitrarily.

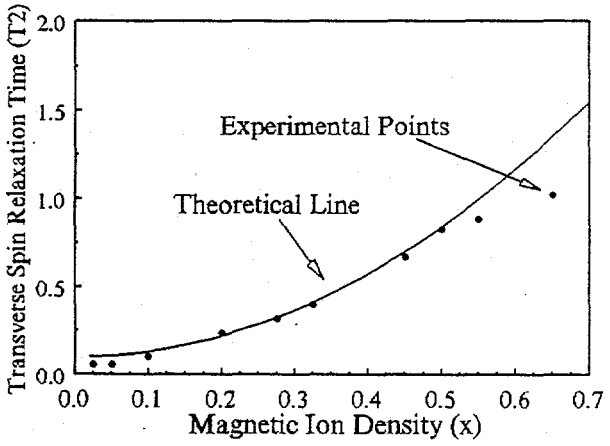


Figure 2: T_2 vs. x (experimental and theoretical)

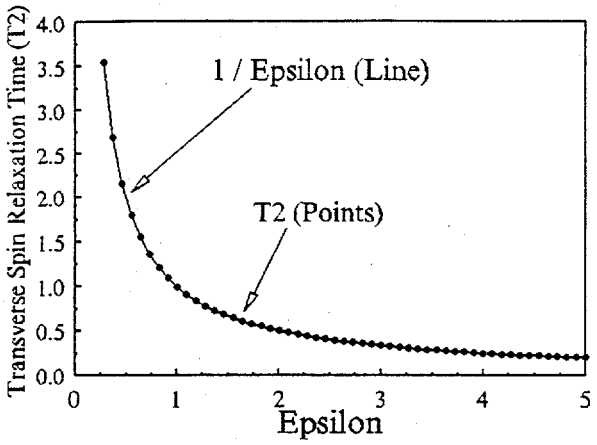


Figure 3: T_2 vs. ε (theoretical)

The graph of T_2 vs. ε, shown in Figure 3, is an almost perfect $T_2 = \frac{1}{\varepsilon}$. Samarth found experimentally[5] that for high temperatures, $T_2 \propto \frac{1}{T}$, which indicates that ε may be a linear function of temperature. Preliminary calculations indicate that ε may be due to vibrations in the crystal lattice. These vibrations can be included in one of the anisotropic contributions to the spin Hamiltonian, i.e. $\hat{H}^{DM}$, using a Born-Oppenheimer approximation[6].

Further work may be done to investigate the effects of adding the nuclear spin interaction to the perturbation Hamiltonian.

REFERENCES

[1] Yang, Gui-Lin. "Electron Paramagnetic Resonance in II-VI Semiconductor Heterostructures". Ph.D. Dissertation, University of Notre Dame, 1993. p 17.

[2] Samarth, Nitin. "Electron Paramagnetic Resonance in $Cd_{1-x}Mn_xTe$, $Cd_{1-x}Mn_xSe$ and $Cd_{1-x}Mn_xS$". Ph.D. Dissertation, Purdue University, 1986. p. 80.

[3] Anderson, P.W. "A Mathematical Model for the Narrowing of Spectral Lines by Exchange or Motion." Journal of Physical Society of Japan vol. 9, no. 3, 1954. p. 321

[4] Barnes, Diana. B.Sc. Thesis, University of Notre Dame, 1992.

[5] Samarth, Nitin. p. 93

[6] Bishop, David M. Group Theory and Chemistry. New York: Dover Publications, 1973. p. 152.

LATTICE ARRANGEMENTS FOR INERT GAS SOLIDS AND HYDROGEN-ARGON MIXTURES USING MONTE CARLO SIMULATIONS

Bruce S. Nguyen

Department of Physics
Jacksonville University
Jacksonville, FL 32211

Research Advisor: T. M. Hakim, Ph.D.

The objective of this ongoing research is twofold: (1) to find the equilibrium lattice spacing in a hydrogen-argon solid mixture, and (2) to study the response of that system to neutron scattering using the second-order self-consistent theory. The first phase will be discussed in this paper. Monte Carlo simulations were used to find the equilibrium lattice constant where the crystal is most stable. The model was first applied to pure inert gas solid monolayers adsorbed on graphite and in the bulk solid, to provide control data. It was then applied to an argon matrix with a dilute concentration of hydrogen. The intricacy of the numerical study, conducted in Turbo C++ on a 486-machine, will be analyzed. The results for the pure inert gas solids and the mixture will be described, and search for the four-digit accuracy in the results will be discussed.

INTRODUCTION

Monte Carlo simulation[1] is used extensively in computer models to study stable systems. It is a statistical approach which tells us the most probable configuration of a system. This project phase used the Monte Carlo simulation to determine the lattice arrangements for inert gas solids—Ar, Ne, Kr, and Xe—and a hydrogen-argon mixture. The objective is to study the dynamics of our hydrogen-argon mixture; determining the lattice arrangement for the mixture is the preliminary work for the next project phase. In order to validate the results of the mixture, the model was first applied to the inert gas solids where their behaviors are known.

The mixtures considered involve hydrogen molecules in an argon matrix and are becoming commonly used in the semiconductor technology in the gaseous phase. While the behavior differs drastically in the solid phase, the latter presents a very interesting system on which computational techniques can be tested. This study is based on the interatomic interactions due to the van der Waals dipolar force[2] which depends on lattice spacing.

MODELS

This project phase is dedicated to determining the equilibrium lattice configuration of a hydrogen-argon mixture. Several models are included here to conduct such study, and archaic models are used to compare and validate the results of our models. The models which are applicable to this phase of the research project include: (1) The Monte Carlo simulation, (2) the lattice arrangements, and (3) the energy models. The Monte Carlo simulation will be used to find the equilibrium configuration of the systems. The lattice arrangements for inert gas solids are considered in both two and three dimensions. The lattice arrangement for the mixture is considered in three dimensions. The energy models are equations where the parameters are determined empirically.

Monte Carlo Simulations

Monte Carlo is well known to be related to gambling, and this simulation similarly is based on a game of chance, with the most probable hits being accepted. The Monte Carlo simulation allows the systems to freely expand and contract, in search of the equilibrium configuration. Similar to a damped oscillator, the Monte Carlo simulation quickly comes to the equilibrium point, but continues to fluctuate about that point. The final equilibrium configuration is the average of all the favorable configurations from energy consideration, which are selectively collected by the following factors: (1) a more stable configuration, or (2) a less stable configuration with a randomly generated number in the range of the weight function[3], given by

$$W = e^{\frac{-|E_2 - E_1|}{K_B T}} \qquad (1)$$

where E_1 is the energy previously accepted, E_2 is the latest energy calculated, k_B is the Boltzman's constant, and T is the ambient temperature.

Lattice Arrangements

The lattice arrangements are definitions of crystal structures which give the location of each spatial lattice points. At each one of those points, there exists an atom or a molecule and the lattice is assumed to be rigid in structure for the current considerations. Three lattices were examined: (1) the triangular lattice monolayer, (2) the face-centered cubic (FCC), and (3) and the hydrogen-argon mixture where the atoms are in the FCC structure and the hydrogen molecules are in the simple cubic (SC) structure. The lattice configurations of the monolayer and the FCC for inert gas crystals are previously determined, but their inclusion is needed to verify the Monte Carlo model. The latter is the system in which the dynamics will be investigated.

The mixture structure is formed by selectively replacing argon atoms in the bulk with hydrogen molecules. In the hydrogen-argon mixture's unit cell, there exist 31 argon atoms and one hydrogen molecule [Figure 1].

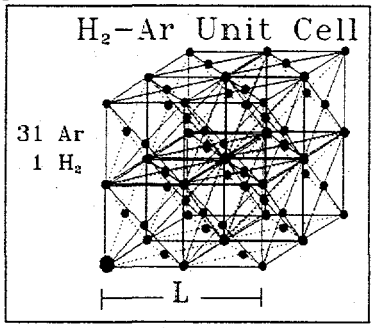


Figure 1. Unit cell of a H_2-Ar mixture. The large circle represents a H_2 molecule, and others represent Ar atoms.

Figure 1

Energy Models

Once the crystal structures are established, the energy of the system lies in its interatomic interaction. The information about interatomic interaction between the atoms lies in the potential energy. There are several models of the potential energy between two atoms. One of the older models is the Lennard-Jones potential[3]. This is an archaic model, but it is included in the research project because much of the older publications use this model.

A more recent model of the potential energy between two atoms is referred to as the Aziz-Chen[4]. In this model:

$$V(r) = \epsilon V(x)$$

$$x = \frac{r}{r_m};\quad V(x) = Ax^{\gamma}e^{-\alpha x} - F(x)\sum_{j=0}^{2}\frac{C_{2j+6}}{x^{2j+6}} \tag{2}$$

$$F(x) = \begin{cases} 1 & x \geq D \\ e^{-\left(\frac{D}{x}-1\right)^2} & x < D \end{cases}$$

where r is the distance between the two atoms, and the other parameters are determined empirically. The Aziz-Chen model is of more interest than the Lennard-Jones model because of its many benefits: (1) the Aziz-Chen model has more empirical parameters, (2) its parameters are determined with more significant digits, and (3) it is a better quantum representation of the inert gas solids in real life since its walls are softer and its well is deeper.

Both the Aziz-Chen and the Lennard-Jones models describe inert gas solids potential interactions. Even though the hydrogen concentration is low in our mixture, the potential interactions between the hydrogen molecules contribute significantly to the total energy of the system. The potential energy model for this type of interaction[5] is given by

$$\Phi_{H_2-H_2}(r) = e^{\alpha-\beta r-\gamma r^2} + F(r)\sum_{i=0}^{2}\frac{C_{2i+6}}{r^{2i+6}} + \frac{C_9}{r^9} \tag{3}$$

$$F(r) = e^{-(1.28\frac{r_M}{r}-1)^2}$$

where r is the vector magnitude from one hydrogen molecule to another, and other parameters are determined empirically.

In the hydrogen-argon mixture, interatomic interaction between hydrogen molecules and argon atoms also have to be considered. Its potential energy model[6] is given by

$$\Phi_{H_2\text{-}Ar}(r) = Ae^{-\beta r} - D_6(r)\frac{C_6}{r^6} - D_8(r)\frac{C_8}{r^8}$$

$$D_n(r) = 1 - e^{-\beta r}\sum_{m=0}^{2n}\frac{(\beta r)^m}{m!} \tag{4}$$

where r is the vector magnitude in between the H_2-Ar pair, and other parameters are determined empirically.

The previous models give the potential interactions, but a dynamic model is needed to determine the lattice spacing as a function of temperature. The self-consistent harmonic free energy, which we refer to as the Gibbs free energy, models the inert gas solids interactions as a function of temperature. We used an alternative form of the equation which is the Einstein approximation[7], given by

$$F_{SCE} = \langle\Phi\rangle_E + 2Nk_BT\ln\left[2\sinh\left(\frac{\beta\hbar\omega_E}{2}\right)\right] - \frac{N}{2}\hbar\omega_E\coth\left(\frac{\beta\hbar\omega_E}{2}\right) \tag{5}$$

where $\langle\Phi\rangle_E$ is the potential energy, N is the number of atoms, k_B is Boltzman's constant, T is the temperature, β is the reciprocal of $k_B T$, $\hbar$ is Planck's constant divided by 2π, and ω_E is the frequency of vibration.

The current interest is on the static equilibrium lattice configuration, but the dynamic model is necessary to observe the lattice arrangements as a function of temperature. Interpolation from the graphs of the frequency, related to the self-consistent harmonic theory, is done for simplicity.

NUMERICAL IMPLEMENTATION

Implementation of the models is conducted by translating the models into abstract computer algorithms. All of the interaction models were implemented for each crystal structure, with and without the dynamical properties. Monte Carlo simulations are used to determine the equilibrium lattice arrangements for each crystal structure. Each crystal structure is expanded and contracted many times, depending on the probability controlled by the Monte Carlo simulation. The expansion and contraction is based on the total system energy, which depends on the lattice spacing and various stresses. Determining the total system energy for each condition and allowing the Monte Carlo simulation to judge whether to keep or drop the configuration will give us the equilibrium lattice configuration of a system. The task then is to determine the system energy of the crystal structures, which is the summation of the interatomic interactions of every atom with every other atom.

In order to determine the lattice spacing and the magnitude of the interactions, a Cartesian coordinate system has to be set up for each crystal structure. The Cartesian arrangements would produce spatial points vectors used to determine the location of each atom or molecule and, consequently, the system energy. The spatial point vectors in the simplest coordinate arrangement for monolayer, FCC, and H2-Ar mixture are R_1, R_2, and R_3, respectively.

$$\vec{R}_1 = \frac{Lx}{2}\hat{i} + \frac{\sqrt{3}}{2}Ly\hat{j} \tag{6}$$

$$\vec{R}_2 = \frac{L}{2}(x\hat{i} + y\hat{j} + z\hat{k}) \tag{7}$$

$$\vec{R}_3 = \frac{L}{4}(x\hat{i} + y\hat{j} + z\hat{k}) \tag{8}$$

The x, y, and z in (6), (7), and (8) represent coordinate indices, and the L represents the lattice constant for that system. Using the spatial point vectors to calculate the energies and summing up the calculations will give us the system energy. In the pure inert gas solid models, only one

energy model is used at a time. In this mixture, the interactions involving the hydrogen molecules have to be considered. The Lennard-Jones and the Aziz-Chen models give the interatomic interactions between the atoms, but parameters are only available for inert gas solids. The H_2-H_2 interaction model and the H_2-Ar interaction model filled the remaining interatomic interactions for the mixture, but this requires the spatial point vectors to differentiate the interaction models. Determining the appropriate energy model is dependent on the atom or molecule at certain spatial lattice points. Differentiating argon atoms from hydrogen molecules can generalize to the following: (1) if each of the coordinate indices for a spatial point vector is divisible by 4, then the vector is pointing to a hydrogen molecule, (2) if the first case is not satisfied and the sum of the coordinate indices for a spatial point vector is divisible by 2, then the spatial point vector is pointing to an argon atom, and (3) finally if neither of the previous cases were satisfied, the spatial point vector is pointing to an invalid point. The latter can be applied to the FCC structure, but not the monolayers. Though not exact, the conditional tests are similar.

Optimization

Processing for this phase of the research project was done on several 486Dx2-66 machines. Data were obtained for the monolayer and the FCC structure, but processing for the mixture took a very long time. A little over two hours was required to process one sample configuration. In order to obtain sufficient accuracy for the mixture, the Monte Carlo simulation must be allowed to search for the equilibrium configuration over many samples. This would take an extraordinary amount of time; therefore, several optimization techniques were implemented.

The first optimization technique implemented was to constrain the spatial point vectors to a spherical shell. Rather than allowing the spatial point vectors to scan in a cubic area, the insignificant interactions at the corners of the cubic were ignored. This optimization technique cuts corners without sacrificing accuracy.

Another optimization technique implemented for the mixture was to use a dynamic linked list. The computer program was allowed to dynamically allocate and deallocate memory to store the calculations. Each of the interactions where processing was not performed for the vector magnitude was calculated and saved in a linked list, and was deallocated once the system energy has been determined for that configuration. An extremely long linked list would actually take longer to search than to calculate. Of course, each type of interaction would have a unique linked list.

The latter was the most important factor in the increase in productivity: For each calculation, all of the interactions that have the same spatial separation are processed. With these optimization schemes applied to the mixture, processing on the same machine now takes under two minutes for one sample configuration as opposed to over two hours. Optimization was more important than mere convenience: It would have been impossible to collect the mixture data on the 486Dx2-66 machine without the optimization schemes.

ANALYSIS AND CONCLUSION

The method that was used to determine the lattice configurations was the Monte Carlo simulation. In search of the equilibrium configuration, the Monte Carlo simulation quickly comes near the equilibrium and tends to fluctuate about that point. Out of 700 sample configurations for the mixture, the Monte Carlo simulation accepted 29 points, most of which are near the equilibrium point.

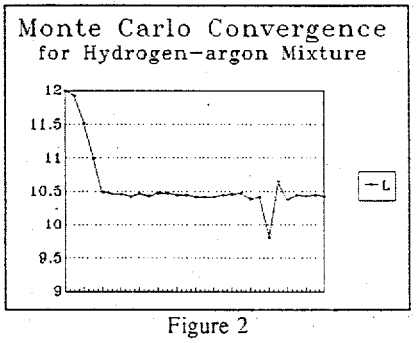


Figure 2. Monte Carlo simulation. The quick convergence is expected, but the quick shift results from hitting a low probability point.

Figure 2

The convergence showed that the Monte Carlo simulation is a good model to use, but the model of the actual system is inconclusive if the convergence is the only factor. Data that have been obtained by other methods reinforced our results.

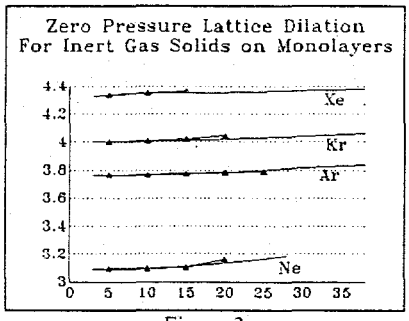


Figure 3

Figure 3. Lattice dilation (K vs. Å) for inert gas solids on monolayers. The solid lines are generated using the Monte Carlo simulation. The triangles represent points that have been calculated using the self-consistent phonon method[8].

Analysis of the Monte Carlo simulation shows that the data consistency increases as the number of sample configurations increases. In the pure inert gas solid models, 10,000 samples were used to get a four-digit consistency. Even though the sample configurations are completely random, the selective scheme of the Monte Carlo model repeatedly produces the same results. The accuracy of the results is not solely dependent on the Monte Carlo simulation, but the energy models play an important role. Many empirical and interpolation parameters were used to determine the state of each configuration; therefore, the accuracy also depends on the parameters to the energy models.

As the data points using the Monte Carlo simulations are within 0.7% of those computed using a very involved self-consistent model, it is clear that the model we used here is functionally useful and practically accurate. Because of this observation, the Monte Carlo model can then be applied to evaluate the lattice constant in the crystalline hydrogen-argon mixture. With relatively few data points, the system seems to settle at 10.45 Å for zero pressure and temperature. More data are needed to confirm the last digit, but its estimate is legitimate. This lattice constant is important because it defines the physical arrangement of the hydrogen-argon crystal structure. This constant will be needed in the next phase of the research project, in which the dynamics of the system will be investigated.

ACKNOWLEDGEMENT

This research was partly funded by a grant from The Research Corporation and a Grant-in-Aid from Sigma Xi, The Scientific Research Society.

REFERENCES

[1]N. Metropolis, *in Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics **21**, 6 (1963).

[2]C. Kittel, *Introduction to Solid State Physics* (John Wiley & Son, New York, 1976).

[3]J. R. Christman, *Fundamental of Solid State Physics* (John Wiley & Son, New York, 1980).

[4]R. A. Aziz, *Inert Gases*, Vol. 34 of *Springer Series in Chemical Physics*, ed. M. L. Klein (Springer-Verlag, New York, 1984), p.5.

[5]I. F. Silvera and V. V. Goldman, Journal of Chemical Physics **69**, 4209 (1978).

[6]R. Leroy and J. Carlay, *Advances in Chemical Physics*, Ed. K. Prawley, (John Wiley & Son, New York, 1980), p.353.

[7]H. R. Glyde and G. H. Keech, Ann. Phys. (N.Y.) **127**, 330 (1980).

[8]S. J. Seeba and T. M. Hakim, *Thermodynamics of Inert Gas Monolayers*, to be published.

Non-degenerate Two-Photon Absorption in ZnSe

Todd E. Buhr and Corey S. Gerving

Department of Physics
United States Military Academy
West Point, NY 10996

Faculty Advisors:
LTC John E. LaSala and CPT Brian Kowalski

Nonlinear absorption mechanisms in polycrystalline ZnSe, a wide gap II-VI semiconductor, are investigated using optical pump-probe techniques. We perform nondegenerate experiments in which a weak 683 nm, 26 ps probe, generated by a hydrogen Raman cell pumped by a 532 nm pulse, arrives at varying times relative to a 532 nm, 40 ps pump beam. By measuring probe absorption versus delay in both parallel and cross polarized configurations, nondegenerate components of the two photon absorption coefficients of ZnSe can be deduced, as can the FCA cross section and free carrier recombination rates. We find that our experimental results compare favorably with theoretical predictions.

INTRODUCTION

Wide gap semiconductors are normally transparent to visible light. When exposed to high intensity light, however, nonlinear effects such as two photon absorption (TPA) and free carrier absorption (FCA) are observed. These nonlinear effects make ZnSe a candidate as an optical power limiting (OPL) material (Hagan, et al, 1988). An OPL material allows high transmittance at low incident intensities, and low transmittance for high incident intensities. ZnSe is also a candidate for a blue-green diode laser, making the free carrier scattering and recombination mechanisms of interest.

Using a pump of intensity greater than 1 GW/cm^2, we create an abundance of free carriers within the polycrystalline ZnSe sample. We then investigate with a weak, nondegenerate probe at varying arrival times relative to the pump in order to gain an understanding of the various mechanisms which contribute to the above properties of ZnSe. The experiment is conducted with the pump and probe both parallel and cross polarized to each other.

THEORY

When the two beams temporally overlap, the probe beam couples with the pump beam through TPA and strong probe absorption is observed. When the probe beam is delayed it experiences FCA by the excess "free" electrons and holes produced through TPA of the pump alone. The governing equations for the development of pump and probe intensities in the medium for the parallel and cross polarized case are as follows (LaSala, et al):

$$\cos\theta \frac{dA^2}{dz} = -\left[\alpha_A + \beta_{2AA}A^2 + 2\beta_{2AB}B^2 + \sigma_A N\right]A^2$$
$$\cos\theta \frac{dB^2}{dz} = -\left[\alpha_B + \beta_{2BB}B^2 + 2\beta_{2AB}A^2 + \sigma_B N\right]B^2 \qquad (1)$$

Table 1: Definition of Terms

	Intensity of pump/probe (GW/cm^2)
2θ	Angle between pump and probe
α_A/α_B	Linear absorption coefficient of the pump/probe[1] (cm^{-1})
β_{2AA}/β_{2BB}	TPA coefficient at the wavelength of the pump/probe[1] (cm/GW)
β_{2AB}	TPA coefficient for a pump photon and a probe photon[1,2] (cm/GW)
σ_A/σ_B	FCA coefficient for the pump/probe[1] (cm^2)
N	Free carrier pair density (cm^{-3})
ω_A/ω_B	Frequency of pump/probe (rad/s)
τ_R	Free carrier recombination time (s)

1. Wavelength dependent
2. Polarization dependent

The first terms within the brackets of the beam equations, α_A and α_B, represent losses in the beam intensities due to linear absorption through the unexcited medium. This loss is proportional to the intensity of the beam. The second terms, β_{2AA} and β_{2BB}, account for losses in the beam intensities due to TPA of the same frequency photon. The loss due to TPA is proportional to the square of the intensity of the beam. The third term, $2\beta_{2AB}$, accounts for loss due to TPA of a pump and probe photon. The 2 is a degeneracy factor; this term is proportional to the product of the beam intensities. At the peak absorption, the terms dealing with TPA are the primary reason for loss of energy in the probe. The final terms, $\sigma_A N$ and $\sigma_B N$, account for loss in beam intensity due to FCA which are proportional to the beam intensity. This is the dominant absorption mechanism in the long delay tail of our absorption data.

The density of free carriers generated by the two beams is modeled as:[2]

$$\frac{\partial N}{\partial t} = \frac{\beta_{2AA}}{2\hbar\omega_A}\left(A^2\right)^2 + \frac{\beta_{2BB}}{2\hbar\omega_B}\left(B^2\right)^2 + 4\frac{\beta_{2AB}}{\left(\hbar\omega_A + \hbar\omega_B\right)}A^2B^2 - \frac{N}{\tau_R} \qquad (2)$$

The first two terms account for an increase in free carriers due to TPA of the pump and probe, respectively. The third term takes into account cooperative TPA of red and green photons. The final term accounts for a loss of free carriers due to recombination. In our model, we assume diffusion is negligible based on the low mobility of carriers in ZnSe.

The value of β_{2AA} has been experimentally determined (Hagan, et. al, 1990(26)) to be 5.8 cm/GW from which the value of β_{2BB} can be calculated using the theoretical expression (Hagan, et. al, 1990 (65))

$$\beta(2\omega) = K\frac{\sqrt{E_p}}{n_o^2 E_g^3}F_2\left[\frac{2\hbar\omega}{E_g}\right] \qquad (3)$$

where F_2 is a function given as

$$F_2(2x) = \frac{(2x-1)^{3/2}}{(2x)^5} \qquad (4)$$

and the other factors are constants for a given medium. For a probe wavelength of 683 nm, we calculate a value $\beta_{2BB} = 6.7$ cm/GW, and further propose that the value of $\beta_{2A\parallel B}$ is between the value of β_{2AA} and β_{2BB}. It is important to note the distinction now made between pump-probe beams being parallel ($\parallel$) and cross polarized ($\perp$) in the subscripts of the β_{2AB}, since these values are dependent upon polarization. The polarization dependency is derived from the relationship between β_2 and the susceptability of the material.[2] This dependancy also predicts that the value of $\beta_{2A\perp B}$ will be on the order of one-third that of $\beta_{2A\parallel B}$.

EXPERIMENT

To test the theory above, we used a basic pump-probe technique in our experiments. (Beam-line diagram shown in Figure 1.) The characteristics of the two beams are shown in Table 2 below.

Table 2: Beam Characteristics

	Pump	Probe
Source	Nd:YAG	Hydrogen Stokes signal
Wavelength	532 nm	683 nm
Duration	40 ps	26 ps
Beam Diameter	1.64 mm	0.33 mm
Nominal Energy	730 μJ	~50 nJ

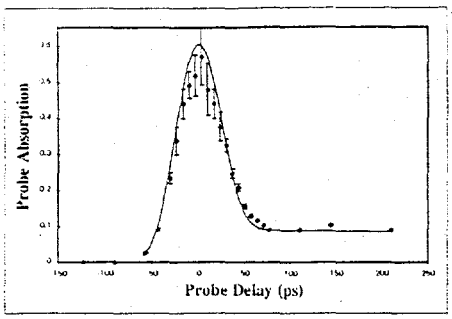


Figure 2(a)

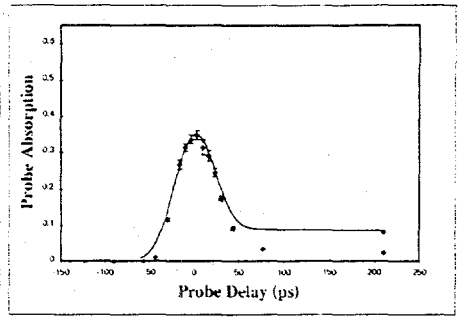


Figure 2(b)

Fig 2: Expiremental and theoretical probe absorption as a function of delay in ZnSe for a) Parallel Polarized and b) Cross Polarized pump-probe beams. The solid dots with error bars are experimental values and the solid curve is a calculation based on the theory of equations (1) and (2).

We fit a numerically computed theoretical data curve to the data for both parallel and cross polarized cases using β_{2AB} and σ_A and σ_B as fitting parameters. As it can be seen, both experimental sets of data fit well with the theoretical curve. Notice in Figure 2(b) that the tail of the curve matches only one outlying data point. This is due to the fact that for a reason unknown to us, the optimal alignment of pump and probe into the ZnSe sample changed and we did not realize it until the end of the experiment. Had we realigned the beams at each point, we are confident that the data would fit the curve more closely. The values for the various parameters used for the plot of the numerical line are listed in Table 3.

Table 3: Parameter Values

Parameter	$A \parallel B$	$A \perp B$
β_{2AA} (green)	5.8 cm/GW	5.8 cm/GW
β_{2BB} (red)	6.7 cm/GW	6.7 cm/GW
β_{2AB}	6.1 ± 0.3 cm/GW	2.4 ± 0.2 cm/GW
σ_A	$2.0 \pm 0.1 \times 10^{-18}$ cm^2	$2.0 \pm 0.1 \times 10^{-18}$ cm^2
σ_B	$4.4 \pm 0.1 \times 10^{-18}$ cm^2	$4.4 \pm 0.1 \times 10^{-18}$ cm^2
τ_R	10 ns	10 ns

CONCLUSIONS

Based on the good agreement between theory and experiment represented in Fig 2, we conclude that the non-degenerate pump-probe interaction in ZnSe is well modeled by equations (1) and (2). This conclusion is further supported by the values of $\beta_{2A}{}_{\perp B}$ and $\beta_{2A}{}_{\parallel B}$ found from fitting the theory to data. The values for $\beta_{2A}{}_{\parallel B}$ lies close to that predicted from equations (3) and (4) when $2\hbar\omega$ is replaced by $\hbar\omega_A + \hbar\omega_B$, a procedure suggested in reference 4 to estimate non-degenerate values of β_2. In addition, $\beta_{2A}{}_{\perp B}$ is close to 1/3 of $\beta_{2A}{}_{\parallel B}$ as predicted in reference 2, a relation that would hold exactly if disperson in ZnSe were negligible in the wavelength range of the experiment (which it is not).

The values of σ_A of and σ_B are comparable to those one would calculate from theory (Seeger, 1991) based on the dominance of the polar-optical phonon scattering mechanism for free carriers, as expected in ZnSe (Ruda, 1992). Based on typical electron and hole mobilities for undoped ZnSe[6] we expect total free carrier absorbtion cross-sections of $\sigma_A = 1.0 \times 10^{-18}$ cm^2 and $\sigma_B = 1.6 \times 10^{-18}$ cm^2. Although our fit values are somewhat larger, they show the same approximate ratio, suggesting the correct wavelength scaling for the assumed scattering mechanism in ZnSe. The large absorption cross-section may indicate that mobility is lower than typically found for high quality ZnSe or additional contributions (such as impurity scattering) are important.

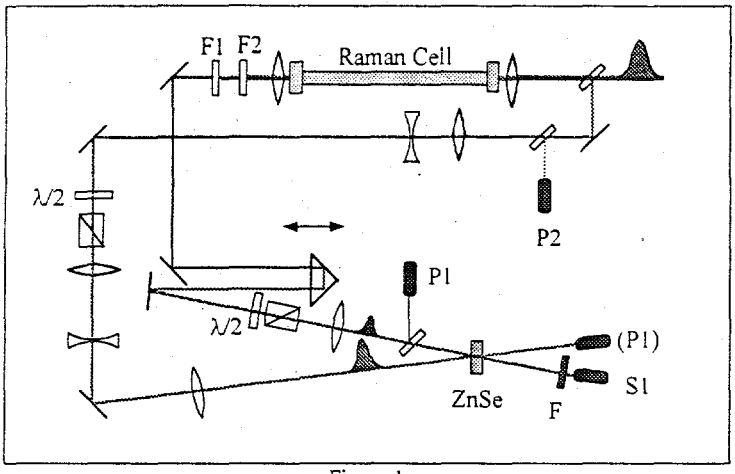


Figure 1

Figure 1: Beam-line Diagram. The light emerging from the frequency doubled Nd:YAG laser output is split using uncoated beam splitters (BS). The probe beam proceeds through the Raman cell (HR) and the retroreflector (RTR) as it makes its way to the sample. After being split off, the pump beam is directed around the table and into the sample. The polarization for the two cases discussed is controlled through the use of the polarizers (P). Detectors P1 and P2 are pyroelectric detectors while S1 is a silicon based photodetector. F is a red glass filter added to stop scattered green pump light from registering on the detector.

Our doubled Nd:YAG was first split into a pump and a weaker probe beam through the use of a wedge type beam splitter. To create a shifted wavelength probe beam with a fixed temporal relation to the pump beam, we used a stimulated Raman process in a two foot hydrogen cell pressurized to 4.2 atm. Using filters to eliminate residual 532 nm and anti-stokes signals from the Raman cell, the remaining stokes (red) beam became our probe. Through the use of the retro-reflector (RTR) mounted on a movable stage, we were able to vary the probe arrival time at our sample. Finally, we ensured the nominal probe energy <50 μJ going into the sample so that the red beam would not produce nonlinear effects of its own.

An important aspect of our experimental setup deals with the use of the detectors P1 and P2. These two detectors helped us window our data to minimize the effects of shot to shot variations in laser energy. The detector at position P2 allowed us to monitor the pump energy for each shot enabling us to eliminate data outside a pre-determined window of pump energy. Detector P1 allowed us to monitor the probe energy just before it entered the sample. Because stimulated Raman scattering is a nonlinear process that starts from noise, we used this detector to normalize the signal at detector S1. The final control we used in the experiment was the employment of polarizers to produce parallel and cross-polarized beams.

RESULTS

The results for the parallel and cross polarized pump-probe study in ZnSe are shown in Figure 2a and 2b, respectively. For each delay setting between ten and twenty data sets were taken. Each data set fell within our predetermined energy window centered at ~730 μJ for the pump pulse. This energy provides an incident peak intensity of ~2.0 GW/cm^2. The transmission of the probe with the pump present was calculated relative to the throughput of the probe with no pump present. The absorption of the probe is defined as follows:

$$ABS \equiv 1 - \frac{\left(\frac{S1}{P1}\right)_{w/\,pump}}{\left(\frac{S1}{P1}\right)_{w/\,o\ pump}} \qquad (5)$$

ACKNOWLEDGEMENTS

The authors would also like to acknowledge the Army Rearch Office as well as the United States Military Academy Photonics Research Center for providing us with the resources necessary to conduct our research.

REFERENCES

Hagan D. J., A. A. Said, M. Sheik-Bahae, E. W. Van Stryland, T. Wei, "Sensitive Measurement of Optical Nonlinearities Using a Single Beam," IEEE JQE 26 (1990) 60+.

Hagan D. J., M. Sheik-Bahae, E. W. Van Stryland, "Disperson and Band-Gap Scaling of the Electronic-Kerr Effect in Solids Associated with Two-Photon Absorption," Physics Review Letter 65 (1990) 96.

Hagan D. J., E. W. Van Stryland, M. J. Soileau, Y. Y. Wu, "Self-protecting Semiconductor Optical Limiters," Optical Letters 13 (1988) 315.

La Sala J. E., G. Benecke, T. Buhr, D. Gardner, C. Gerving, A. Perica, "Nondegnerate Two Photon Absorption Measurements in ZnSe," in preparation.

Ruda H. E., Widegap II-VI compounds for Opto-electronicc Applications (London: Chapman and Hall, 1992), p.202-213.

Seeger K., Semiconductor Physics (Berlin: Springer-Verlag, 1991), p.356.

PHOTOLUMINESCENCE AND LOW-FREQUENCY RAMAN SCATTERING FROM SEMICONDUCTOR NANOCRYSTALS

ERIC E. PEARSON
DEPARTMENT OF PHYSICS
RENSSELAER POLYTECHNIC INSTITUTE
TROY NY 12180-3590

FACULTY ADVISORS:
JOHN SCHROEDER, PH.D. AND JOSEPH HAUS, PH.D

ABSTRACT

Useful information about the electronic band structure, localized states and vibrational properties of CdS_xSe_{1-x} nanocrystals can be determined by analyzing spectra of Raman Scattering, Low Frequency Inelastic Scattering (LOFIS), and Photoluminescence. Relative positions of the LO and 2LO peaks on Raman spectra can be used to determine the concentrations of S and Se in CdS_xSe_{1-x} nanocrystals. LOFIS spectra can be used to determine average diameter of nanocrystals. Photoluminescence spectra can be used to verify the presence of quantum confinement.

By making precise observations of quantum effects and their influence on nanocrystalline structure, it is our hope to contribute to progress toward the synthesis of "pure" nanocrystals without defects or variances. Such crystals would be valuable in the creation of devices that operate by channeling photons instead of electrons. These "optical transistors" would have widespread applications in the field of optical computer technology.

INTRODUCTION

Useful information about the electronic band structure, localized states and vibrational properties can be determined by observing spectra of Raman Scattering, Low Frequency Inelastic Scattering, Absorption, and Photoluminescence. Here, we propose to investigate II-VI semiconductors in the nanocrystalline form of CdS_xSe_{1-x} embedded in a glass filter. In nanocrystal formation, a borosilicate glass sample is injected with ions of Cadmium, Selenium, and Sulfur. Through heat treating, the sample is brought up to its Glass Transition Temperature, at which molecular movement within the sample may occur. This allows the formation of nucleation centers on which nanocrystals may grow. However, such growth only occurs when these centers exceed a certain critical size. This critical size is determined in part by the annealing temperature [1].

There are two variations of this process. One is homogeneous nucleation, in which the growth process, once started, is self-sustaining. In preparing the samples, heterogeneous nucleation was employed, in which the growth process is assisted by the insertion of an impurity.

THEORETICAL BACKGROUND OF LIGHT SCATTERING

In Raman scattering, light is scattered when photons interact with phonons. Phonons are quantized modes of molecular vibration in the crystal lattice. A solid contains *N* lattice sites. On each lattice site there are *m* atoms. Since each atom has 3 degrees of freedom, there are 3m eigenmodes. Of these, 3 are acoustic modes and 3m-3 optic modes. For CdS, m=4, so there are 9 optic modes.

Low Frequency Inelastic Scattering(LOFIS) differs from Raman Scattering in that in LOFIS, vibration occurs at the nanocrystalline level instead of the molecular level. In LOFIS, there are three possible modes of nanocrystal vibration: ellipsoidal, torsional, and breathing[2].

In the photoluminescence of "bulk" materials, light of sufficiently high energy excites an electron from its parent atom's valence band to the conduction band. The presence of an additional electron in the conduction band results in an absence in the valence band. This absence is called a "hole". Eventually, in a process known as non-radiative transition, both the electron and hole "relax" to the extrema of their respective bands. The electron "decays" back to the valence band. This transition releases energy in the form of a photon. It should be noted that the energy of the released photon is equal to the band gap energy.

Photoluminescence in nanocrystalline materials differs from bulk photoluminescence in that the emitted photon energy is greater than the energy of the bulk band gap. This is due to the presence of quantum confinement. In quantum confinement, an energy continuum is divided into discrete energy levels. The end result is a greater change in energy during the electron's return to the valence band, making its emitted photon more energetic. Since the energy of the emitted photon is greater than that of the incident photon, its wavelength is shorter, resulting in a phenomenon known as blue shifting.

EXPERIMENTS AND APPARATUS

Devices used in the experiment were a Spex 1403 0.85m double monochromator whose slit widths were 200μm each, for a resolution of 1.5 cm^{-1}. Laser light was supplied by a Spectra-Physics Ar^+ laser using the laser line at 488 nm. Data for the experiments was processed by a pair of IBM PC AT computers.

RESULTS

The interactions between photons and phonons can be observed on two spectra, shown in Figs. 1 and 2. The relative positions of sulfur longitudinal optical phonon (LO_S) peaks and selenium longitudinal optical phonon (LO_{Se}) peaks on a Raman scattering spectrum can help us compute the concentrations of sulfur and selenium in CdS_xSe_{1-x} [3]:

$$x = (\Delta\omega - 60) * 1.65 + 5 \qquad (1)$$

Here, x represents a percentage. The experiments yield an x-value of 0.9, giving a composition of $CdS_{0.9}Se_{0.1}$. It should be noted that in theory, the Stokes peaks(Fig. 1) and the Antistokes peaks(Fig. 2) should be equidistant from the laser line.

LOFIS spectra were analyzed to determine average nanocrystal diameter. In LOFIS, nanocrystals can be regarded as inelastic bodies that vibrate at distinct frequencies. When photons interact with these nanocrystals, spectral peaks are observed that correspond to these vibrations (Fig 3.). The values of ω at which these peaks occur can be used in the formulae below[4] to calculate 2a, the average nanocrystal diameter:

$$\omega_\tau = 0.82v_t / 2ac \qquad (2)$$

$$\omega_l = 0.9v_l / 2ac \qquad (3)$$

Here, v_t is the transverse velocity of sound (1760 m/s) and v_l is the longitudinal velocity of sound (4250 m/s) An average nanocrystal diameter of 5.08575 nm was calculated. In the Photoluminescence spectrum (Fig. 4), two peaks are present, one of which is identified as the band edge peak. The other peak suggests the existence of quantum confinement within the nanocrystalline structure, causing the blue shifting of light emitted by the sample.

CONCLUSIONS

Spectra of Raman scattering experiments were analyzed to determine the relative concentrations of Sulfur and Selenium in CdS_xSe_{1-x} nanocrystals, yielding a nanocrystal composition of 90% Sulfur and 10% Selenium. Spectral analysis of Low-Frequency Inelastic Scattering (LOFIS) can be used to determine average nanocrystal diameter. In this experiment, an average nanocrystal diameter of 5.08575 nm was observed. The existence of a defect peak in the observed Photoluminescence spectrum verifies the presence of quantum confinement on the nanocrystalline scale.

ACKNOWLEDGEMENTS

The author would like to thank John Schroeder, Joseph Haus, Markus Silvestri, and Mierie Lee for their assistance and support throughout this research project.

REFERENCES

1. S. A. Gurevich et al, *Semiconductors*, Vol. 28(5), May 1994,486–493.
2. B. Champagnon, B. Andrianasolo, E. Duval, *Mat. Sci. Eng.*, B9(1991), 417–420
3. P.D. Persans et al, *Mat. Res. Soc. Symp. Proc.*, Vol. 164, 105.
4. B. Champagnon, B. Andrianasolo, E. Duval, *J. Chem. Phys*, Vol. 94(7), 1 April 1991.
5. X.S. Zhao, J. Schroeder, P.D. Persans, T.G. Bilodeau, *Phys.Rev.*, B 43 12580(1991).
6. J. Schroeder et al, *MRS Symposium* 272, 251(1992)

Fig. 1 **RAMAN SCATTERING (Stokes Peak)**

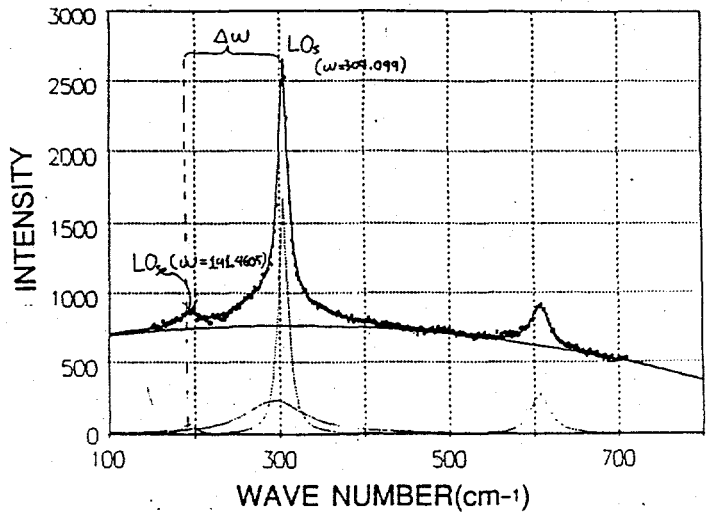


Fig. 2 **RAMAN SCATTERING**
(Antistokes Peak)

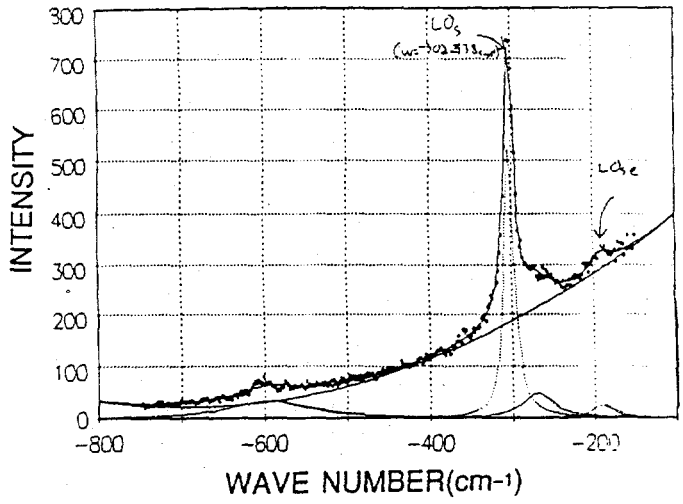


Fig. 3 **LOW-FREQUENCY INELASTIC SCATTERING (LOFIS) SPECTRUM**

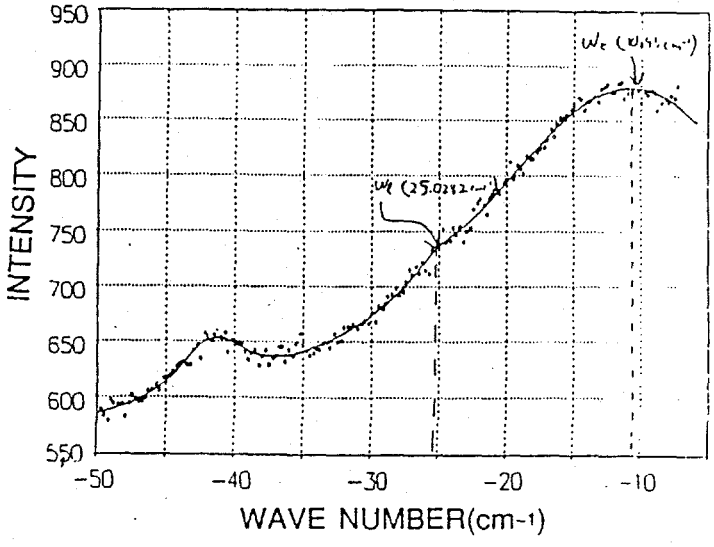
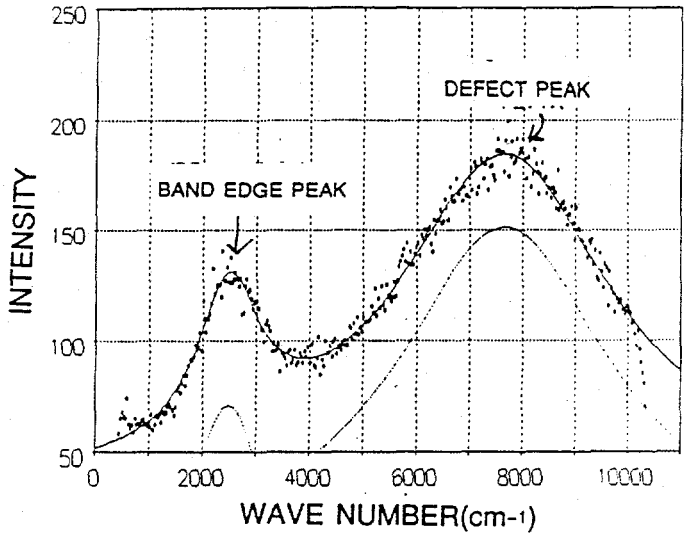


Fig. 4 **PHOTOLUMINESCENCE SPECTRUM**



Thermal Annealing of Organic Thin Films in MIS Devices

Benjamin A. Meier

Department of Physics
Moorhead State University
Moorhead, MN 56563

Vijendra. K. Agarwal

INTRODUCTION

Silicon is used extensively in the semiconductor industry because it has the ability to form a thin native oxide film which is electrically insulating. This native oxide acts as a dielectric spacer between the silicon semiconductor and metal contact in solid state devices. Compound semiconductors such as gallium arsenide (GaAs) do not have native oxides. For these semiconductors to be used in solid state devices, some other insulating material must be deposited on them. Often, other oxides or nitrides are deposited on semiconductors to form insulating barriers. Unfortunately, the techniques used for depositing oxides and nitrides generally require high temperatures and/or electromagnetic fields which can cause defects in the semiconductor.

Research has been conducted on the possible use of organic thin films on semiconductors that lack a native oxide. Recent work on the use of stearic acid ($C_{17}H_{35}COOH$) as a dielectric material in solid state devices has shown promise. It has been demonstrated by Agarwal and Huang [1] that stearic acid films have a comparable dielectric strength to silicon oxide (10^8-10^9 V/m). The advantage of using stearic acid over an oxide or nitride is that stearic acid can be deposited by thermal evaporation on a semiconductor substrate at relatively low temperatures (~68° C.). One of the disadvantages of using stearic acid is that when deposited by thermal evaporation, it forms a fragile film which inherently contains a large number of defects, usually in the form of pinholes and voids. These defects decrease the dielectric strength of the film.

The MIS (metal/insulator/semiconductor) devices created with stearic acid and aluminum on p-type silicon typically do not display the ideal C-V (capacitance-voltage) curve. One possible reason for the unusual C-V curves in our work is the high density of defects in the stearic acid layer. The C-V measurements of MIS devices are generally used to characterize the nature and density of defects on the interface between the insulating layer and the semiconductor substrate. The research presented here focuses on the possibility of thermally annealing the insulating layer in MIS devices. Smith [2] has shown that thermal annealing of stearic acid films in MIM (metal/insulator/metal) devices improves the field strength of the insulating film. Thermal annealing should also reduce the defects in the stearic acid film of MIS devices and consequently improve the C-V characteristics of those devices.

Our studies on the effects of thermal annealing of MIS devices have shown that post-annealing C-V curves are different than pre-annealing C-V curves of the same device. The effects of annealing on MIS devices are inconclusive due to the lack of reproducibility of the C-V curves that stem from difficulties in making contact with the electrode due to the delicate nature of the stearic acid. It can however be concluded that with annealing, the dielectric characteristics of stearic acid in MIS devices are likely to be improved, similar to the results reported by Smith.

THEORETICAL CONSIDERATIONS

The physical structure of an MIS device is shown in Figure 1. A thin layer of an insulating material is sandwiched between a semiconductor substrate and a metal electrode. Electrical connections are made to the MIS device at the metal electrode and with the underside of the semiconductor substrate. The theory pertaining to MIS devices is thoroughly covered in references [3-5], however, only the theory pertaining to p-type semiconductors is discussed here. The ideal MIS device is defined as having the following characteristics:

1. Under a zero bias condition, the work function of the metal, ϕ_m, is the same as the work function of the semiconductor, ϕ_s.
2. The only charges found in the device under any biasing condition are those in the semiconductor and there is an equal number of opposite charges at the metal surface.
3. The insulator is "perfect", i.e., it has no conductance.

When a bias voltage is applied to the MIS device, three conditions can occur. (1) The condition known as accumulation occurs when a negative bias is applied. During accumulation,

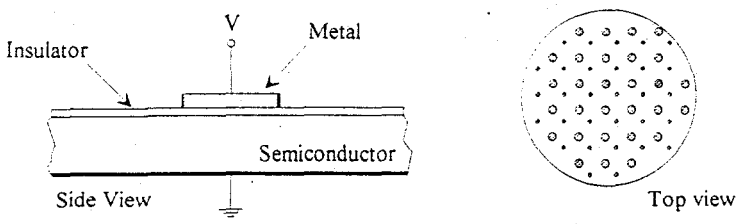


Figure 1: The physical structure of an MIS device.

the majority charge carriers collect at the semiconductor-insulator interface. Accumulation is shown in figure 2a. The Fermi energy of the system is denoted by E_F, the top of the valence band of the semiconductor is at E_V, the conduction band of the semiconductor begins at E_C, and E_i is the intrinsic Fermi level of the semiconductor. (2) Depletion occurs when a small positive bias is applied to the MIS device. The positive bias causes charge to leave the semiconductor-insulator interface and the density of the majority charge carriers is reduced. Depletion is shown in figure 2b. (3) At a higher positive bias, depletion occurs to such an extent that the majority charge carrier in the semiconductor changes (from holes to electrons in p-type semiconductors). This is called inversion. Figure 2c demonstrates the energy band diagram of inversion.

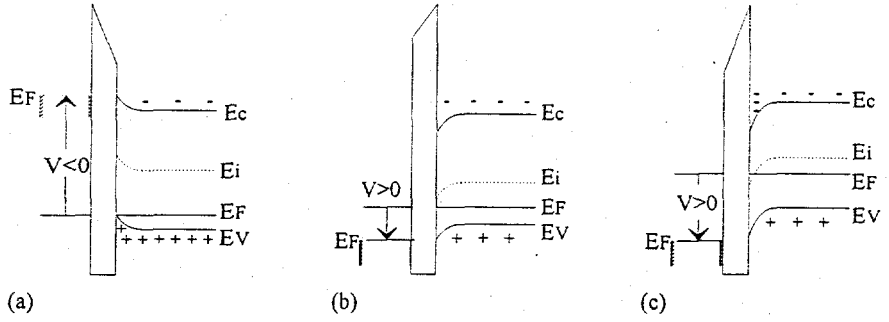


Figure 2: The energy band diagram of an ideal MIS device (p-type semiconductor) under (a) negative bias, (b) small positive bias, (c) large positive bias. Adapted from Sze [3].

The total capacitance of an MIS device can be simplified as the sum of two capacitors in series. The first capacitance, C_i is due to the insulating layer. The other capacitance, C_s is due to a space-charge region in the semiconductor. The semiconductor capacitance arises from the charge separation in the semiconductor and its magnitude depends on the applied bias. The ideal C-V curve of an MIS device with a p-type semiconductor is shown in figure 3. For all frequencies, accumulation occurs at a negative bias and the semiconductor capacitance C_s becomes large causing the total capacitance to approach a maximum, C_i. At a small positive bias,

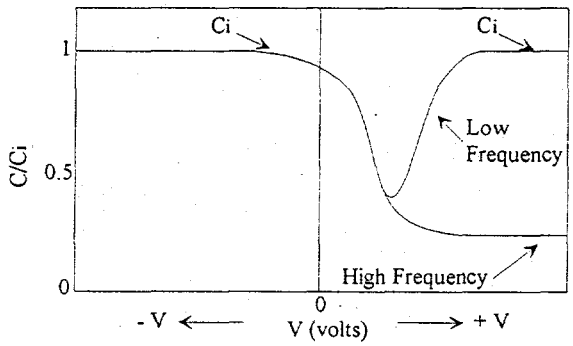


Figure 3: The predicted capacitance with respect to bias voltage of an ideal MIS device (p-type semiconductor). Adapted from Sze [3].

depletion occurs and C_S becomes small which causes the total capacitance to decrease to a minimum value. As the positive bias increases, the C-V response for low and high frequencies is different. At low frequencies (< 1 MHz), inversion occurs and C_s again increases causing the total capacitance to increase. At high frequencies (≥ 1 MHz), the charges in the semiconductor can not keep up with the signal, and the capacitance remains at a minimum value.

EXPERIMENTAL PROCEDURES

The MIS devices were created by thermal evaporation in vacuum. The semiconductor substrates were obtained from a commercial source. The p-type silicon wafers were approximately 2.5 cm in diameter and 0.05 cm thick. The substrates were thoroughly cleaned in an ultrasonic cleaner (acetone bath) for 10 minutes and then rinsed with distilled water. The substrates were then blown dry with a Texwipe MicroDuster® OS (clorodifluoromethane). The stearic acid and aluminum were both deposited onto the silicon substrate by thermal evaporation in a high vacuum. The vacuum system used a pumping stack to reduce the pressure to about 10^{-6} Torr. A liquid nitrogen trap could be used to assist the pumping process and reduce contamination caused by backstreaming oil vapors. The substrate was attached to a substrate holder which was cooled in the vacuum chamber by an array of cooling lines. The pressure of the system was measured with an ion gauge. The deposition rate and the total film thickness were measured with a quartz crystal oscillator. The temperature of the substrate and of the stearic acid source was measured with a T-type thermocouple.

The stearic acid was deposited from a tantalum boat with a perforated cover. The boat was resistively heated to the evaporation temperature of stearic acid (~68° C). The deposition rate of the stearic acid was maintained below 5 Å/s and the final thickness was 1500-2000 Å. After allowing time for the temperature and pressure to decrease, the aluminum was deposited from a tungsten trough. A mask was used to deposit aluminum in an array of small dots on the substrate, each dot creating a separate MIS device (top view in figure 1). The temperature of the substrate was closely monitored while the aluminum was being deposited to avoid re-evaporating the existing stearic acid film. The final aluminum thickness of 2000-4000 Å was achieved with a deposition rate of about 10-50 Å/s in several successive cycles. After the aluminum had been deposited and the system allowed to cool, the sample was removed for C-V testing.

The C-V curves of the MIS devices were measured with a Micromanipulator 410 C-V plotter and recorded with a X-Y recorder. The C-V measurement system is represented by figure 4. The dc bias provided by the C-V plotter was recorded on the X-axis and the device capacitance was recorded on the Y-axis. The stage on which the samples were placed had electrical and water/air connections. An electrical connection to the stage provided the C-V plotter with an ohmic contact to the substrate. A Micromanipulator probe with a tip radius of about 0.5 μm was used to make an electrical connection to the metal electrode.

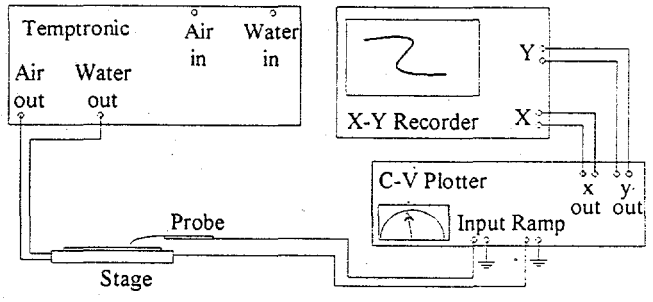


Figure 4: A diagram of the C-V measurement system used to make Capacitance-Voltage measurements on MIS devices.

The programmable temperature controller was used to precisely control the temperature of the sample. The temperature control unit was used to anneal the MIS devices at a temperature near the melting point of stearic acid (~71° C). The temperature of the samples was ramped from about 20° C to 70° C at a rate of 1° C/minute. The temperature was held near 70° C for 15-20 minutes and then ramped back to ~20° C at the same rate. The C-V measurements were made prior to the temperature ramp, while the sample was held at ~70° C, and when the sample had reached thermal equilibrium at ~20° C.

EXPERIMENTAL RESULTS

The C-V measurements made at 1 MHz on MIS devices with stearic acid insulating films usually resulted in a flat response similar to those in figure 5. The flat response does not correspond to the predicted curve for MIS devices measured at 1 MHz. The four curves in figure 5 correspond to four different MIS capacitors on the substrate. The following figures (6 and 7) show the C-V plots at different times during the annealing process for the same four MIS devices. In figure 5, capacitor #1 demonstrates a flat trace, although the trace appears noisy. This was recorded at 20° C prior to annealing.

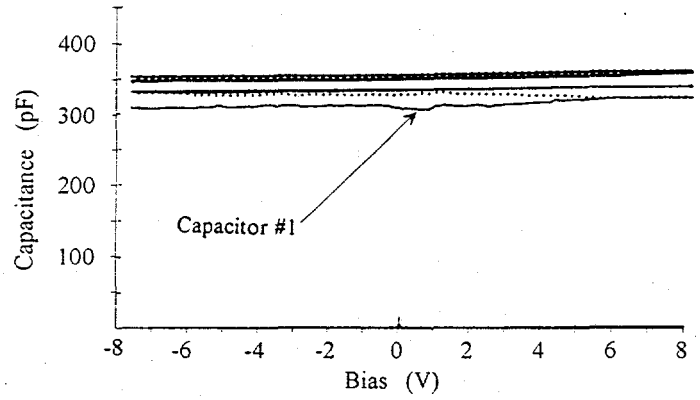


Figure 5: The pre-annealing C-V plots of sample B-15 measured at 1 MHz at ~20° C.

The temperature of the sample was then ramped up to 70° C and the C-V measurements made. Figure 6 shows the C-V curves made at 70° C. The trace for capacitor #1 now shows a small increase in capacitance while the bias changes from negative to positive. Although this C-V curve is still not the expected response, it has changed slightly from the initial curve in figure 5.

Figure 7 shows the post-annealing C-V plot of the same capacitor after the temperature had returned to ~20° C and thermal equilibrium reached. A significant peak in capacitance was observed around the transition from a negative to a positive bias, although this still is not the expected C-V response which was demonstrated in figure 3. The same type of anomalous C-V curve has been demonstrated by prior researchers. An explanation has not been determined for the unusual results.

DISCUSSION OF RESULTS

Capacitance-voltage measurements were performed on more than 100 separate MIS devices during the course of this research. A significant number of the MIS devices produced flat C-V curves, while the remaining MIS devices produced curves that, while not flat, did not resemble the theoretically expected curve. The unusually high number of devices exhibiting unexpected results posed more questions than answers. We are reassessing the experimental procedures and the equipment limitations.

One problem that was identified during the course of this research was with the probe and its contact with the aluminum electrode. The trace of capacitor #1 in figure 5 is not smooth as most C-V plots were, which seemed to suggest that the probe contact with the electrode was not stable. A micromanipulator was used to position the probe tip on the metal electrode. When inconsistencies in the data due to the problems associated with the contact between the probe tip and electrode were encountered, the minimum distance possible to move the probe was calculated to be 15000-20000 Å. This minimum distance is nearly 3 times the combined maximum thickness of the stearic acid and aluminum layers. It is believed that the probe had punched through the soft aluminum and stearic acid films during many of the C-V measurements. While analyzing the data, it was suspected that the flat C-V response in figures 5-7 could simply be the capacitance of the semiconductor. Subsequent measurements on a bare substrate demonstrated the flat response, confirming the suspicion. It is then reasonable that the MIS devices that exhibit flat responses contain such a high density of pinholes and voids that the stearic acid film does not function as an effective insulating barrier between the semiconductor and metal electrode. Capacitor #1 in

figures 5-7 may have initially had a lower density of defects which when annealed showed significant improvement in the C-V behavior.

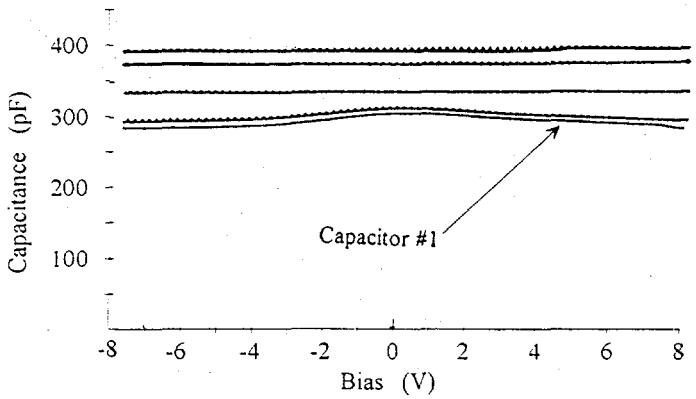


Figure 6: The C-V plots of sample B-15 measured at 1 MHz while held at ~70° C.

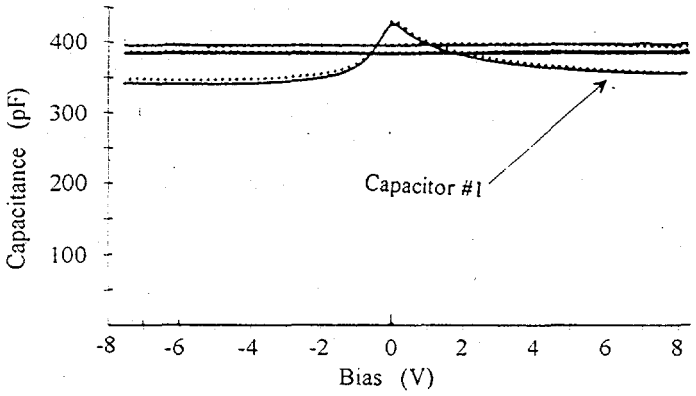


Figure 7: The post-annealing C-V plots of sample B-15 measured at 1 MHz at ~20° C.

Because this preliminary work resulted in changes in the C-V curves of only a few MIS devices, it does not lead to any sound conclusions about the effects of thermal annealing on the C-V measurements of MIS devices, however, it is reasonable to suggest that thermal annealing does have some effect on the insulating layer. Further work would be desirable to resolve the problems associated with the contact between the probe and the metal contact. In order to determine conclusively the effects of annealing on the C-V characteristics of MIS devices, the effects of long term, moderate temperature stress should also be considered.

Thanks are due to Dr. Agarwal for his exceptional guidance and support throughout the duration of this research, to Dr. Buckner for his technical tutoring and suggestions, and to Tim Storsved for his assistance during this research. This research was supported in part by a grant from the National Science Foundation.

REFERENCES

[1] V. K. Agarwal and C. H. Huang, "Self-Healing Process and DC Breakdown in Thermally Grown Stearic Acid Thin Films," Thin Solid Films, 137, 153-160, (1986).

[2] Robert B. Smith, "Effects of Thermal Annealing on the Electrical Breakdown of Thin Fatty Acid Films," M.S. Thesis, Texas Tech. Univ., (1988).

[3] S. M. Sze, Physics of Semiconductor Devices (Wiley-Interscience, New York, 1969).

[4] S. L. Buckner, "Electrical and Optical Properties of Organic Thin Films," Ph.D. Dissertation, Texas Tech. Univ., (1991).

[5] Richard Dalven, Introduction To Applied Solid State Physics (Plenum Press, New York, 1990).

Inclusions in Carbonaceous Chondrite Meteorites and the Formation of the Solar System

K. Amanda Leach

Illinois Mathematics and Science Academy
1500 Sullivan Road
Aurora, IL 60506

Mentors:

Steve Simon and Lawrence Grossman
Department of the Geophysical Sciences
The University of Chicago
5734 South Ellis Avenue
Chicago, IL 60637

BACKGROUND

The solar system is believed to have originated from a great spinning cloud of gas and dust, the solar nebula, at the center of which temperatures were sufficiently hot to evaporate all solids. Some of the material toward the outside cooled to form planets, but most of the material moved toward the center and became the sun. The dust and gas left behind cooled and formed clumps of minerals, preserved as inclusions embedded in certain kinds of carbonaceous chondrite meteorites, a rare type of meteorite. These inclusions are thought to have remained relatively unaltered in the condition in which they condensed or cooled from the solar nebula.

Through observations of the spectrum of the sun, we know its composition, which is roughly the same as that of the primitive solar nebula as the sun contains 99.9% of the matter of the solar system. This knowledge has lead to a thermodynamic model of the gas cloud which predicts which minerals would have condensed first (See Table 1). Many of the minerals in the theoretical condensation sequence are found in inclusions in carbonaceous chondrite meteorites, thereby providing the basis of the proposition that the inclusions represent high temperature condensates from the solar nebula.

Inclusions found in carbonaceous chondrite meteorites are classified into several categories. Some inclusions are fine-grained, where the minerals are present in small grains and are therefore more prone to alteration. Of the coarse-grained inclusions which have been the subject of most studies, there are two types, A and B. Type A inclusions are characterized by a high abundance of the mineral melilite. The subgroups of Type A inclusions are fluffy and compact. Fluffy Type A inclusions are nodular and contorted, with cavities and somewhat heavy alteration due to interaction with the minerals of the matrix of the meteorite. Compact Type A inclusions are approximately spherically shaped with no cavities. The formation of the fluffy type is explained by the aggregation of clumps of grains floating in the solar gas cloud. The second type, Type B inclusions, are characterized by a different mineralogical composition. They also have melilite, but they have much more pyroxene than Type A inclusions. Type B1 inclusions tend to have pyroxene in the core of the inclusion and a mantle of melilite towards the outside while Type B2 inclusions are uniform. These inclusions were probably melted either during or after condensation.

OBJECTIVES

The first objective of the research was to gain a general understanding of the nature of the inclusions and their mineral composition. The second was to compare newly sampled inclusions from the Axtell meteorite to

previously studied meteorite inclusions. The accomplishment of these objectives will lead to 1) a better idea of the types of inclusions that exist, and 2) a more accurate perception of the nature of the solar nebula and it's role in the formation of inclusions.

Table 1: Condensation Sequence

Mineral	Temperature(K)
Corundum	1770
Hibonite	1743
Perovskite	1688
Melilite	1628
Spinel	1501
Metal	1464
Fassaite	1449
Forsterite	1443
Plagioclase	1415

METHODS

The inclusions analyzed were all from the Axtell meteorite. Thin sections of the inclusions were prepared for examination. For identification, each thin section was given a label which includes 'Ax', to identify its source as the Axtell meteorite, and an arbitrary number according to the sequence in which the inclusions were removed from the meteorite. The inclusions from Axtell that were studied include:

Ax-29	Ax-27
Ax-28	Ax-26
Ax-5	Ax-30

A JEOL scanning electron microscope (SEM) was used to create photo mosaics of the inclusions as guides for examination and note-taking. The scanning electron microscope was used in two different ways. In an SEM, a tiny beam of electrons is focused by electromagnets, accelerated at high voltage, and then scanned across the sample. The electrons that bounce back are received and translated into a picture. The first method of analysis is through imaging of the electrons that bounce off the surface of the sample; the image produced shows surface features. The second method is to create an image from backscattered electrons. This method makes use of the fact that heavier elements in a sample absorb and reflect back more electrons than lighter ones, therefore appearing brighter in the image. The minerals in each inclusion were identified using an energy dispersive spectrometer. This instrument allows identification of the X-rays that result from the electron bombardment of the sample. Each element gives off X-rays of a characteristic energy. From the pattern of peaks, the elements present and their proportions can be determined, usually allowing identification of the mineral.

RESULTS

Ax-28 is an unusual, highly irregularly-shaped inclusion. Only a representative portion of the inclusion was photographed because of its shape. It has many hibonite laths occurring with spinel. Ax-29 is a fine-grained inclusion. It includes tiny nodules of spinel enclosed by pyroxene. Other minerals detected by X-ray spectroanalysis of Ax-29 include Al-diopside, fassaite, and anorthite. Ax-27 and Ax-30 are both type A inclusions. However, while Ax-30 is a compact type A, meaning that it is round and mostly melilite, Ax-27 is a fluffy type, meaning that it is nodular and irregular, though still mostly melilite. Ax-27 is 7 mm long in thin section. Ax-30 is 2.2

mm long and features a Wark-Lovering rim layer. Ax-5 and Ax-26 are type B1 inclusions. In the photo mosaic created of Ax-5 the crystals of melilite pointing inward are clearly visible. The mantle of Ax-5 is interesting because of distinct shape the melilite crystals. Ax-26 has a relatively thin melilite mantle. The features of Ax-26 include coarse fassaite crystals separated by fine-grained anorthite.

One petrographic difference between inclusions of Axtell and those of other meteorites is readily apparent. When comparing the B1 inclusions of the Axtell meteorite with those of other meteorites, one can see that the B1 inclusions of Axtell have a melilite mantle which incorporates a noticeable amount of fine-grained spinel, while the B1 inclusions of other meteorites possess spinel almost entirely in the pyroxene core and not the melilite mantle.

CONCLUSION

Comparison of the inclusions studied from Axtell and inclusions of other meteorites does not show large deviations from patterns of mineralogical composition that have been previously observed. The difference noticed in B1 inclusions does not exclude it from the categories of the classification system. It does suggest, however, that further study may reveal further subtleties in conditions of formation that have not been explored.

APPLICABILITY TO OTHER RESEARCH

Information on the formation of these inclusions yields knowledge of the physico-chemical conditions within the solar gas cloud. Along with an accurate representation of this cloud, knowledge gained from this type of research may be applied to the history of the processes that occurred during the formation of the sun and planets.

REFERENCES

Grossman, Lawrence, "Refractory Inclusions in the Allende Meteorite," *Annual Review of Earth and Planetary Sciences*, (1980) Vol 8:559-608.

MacPherson, Glenn J., and Grossman, Lawrence, ""Fluffy" Type A Ca-, Al-rich Inclusions in the Allende Meteorite", *Geochimica et Cosmochimica Acta*, Pergamon Press Ltd., (1984)Vol. 48:29-46.

MacPherson et al. "Refractory Inclusions in the Murchison Meteorite", *Geochimica et Cosmochimica Acta*, Pergamon Press Ltd., (1983)Vol. 47:823-829.

Simon, et al., "Axtell, a New CV3 Chondrite from Texas", *Meteoritics*, (1995) Vol. 30:42-46.

Yoneda, et al. "Chemical Compositions of Refractory Inclusions in Axtell"*Lunar and Planetary Science* (1980), Vol. 26:1541-1542.

Kinetic Studies of the Ionic Current through the M_2 Channel in the Influenza A Virus

Yuqi Qiao

Faculty Adviser : Carl S. Helrich

Turner Laboratory, Department of Physics
Goshen College
Goshen, IN 46526

Currents recorded from single M_2 proteins expressed in artificial phospholipid membranes were studied to determine the dynamics of transitions from various channel states. A self similarity in time, characteristic of chaotic dynamics, was observed in an analysis of the transition rate constants. A fractal dimension of about 1.55 was obtained for transitions of the channel while that of about 1.85 for transitions in the presence of Amantadine. The difference in fractal dimension indicates that amantadine blocks some pathways of channel activity.

INTRODUCTION

The virion lipid envelope of the influenza A virus contains three integral membrane proteins: hemaglutinin (HA), neuraminidase (NA) and a small protein designated as M_2. Within the envelope are the helical viral ribonucleoprotein (RNP) structures. The binding of the virus to an infected cell surface occurs by means of an interaction between HA and cell surface molecules and delivers RNPs into the cytoplasm of the infected cell [reviewed in Lamb, 1989]. In this it is a low intracompartmental pH which causes a conformational change in HA rendering it competent to mediate fusion of the viral envelope with membrane of the infected cell. The small protein M_2 has been identified as an ion channel capable of modulating the pH of intracellular compartments in affected cells [Pinto, *et al.*, 1992].

Amantadine (1-aminoadamantane hydrochloride) is an antiviral drug which specifically inhibits influenza A virus replication. The blockage is thought to occur after the virus is bound to the cell but before uncoating occurs. By expressing M_2 in oocytes of *Xenopus laevis* (African clawed frog) , Pinto and coworkers have found that the M_2 protein forms an ion channel which is blocked by amantadine, regulated by pH, and is selective for ions [Pinto, *et al.*, 1992]

Recently studies by Tosteson and coworkers of the kinetics of single M_2 channel proteins have shown that the M_2 protein forms multisized channels, identified as relative maximum values attained by the current from an event versus amplitude histogram, and that the large conductance states are drastically reduced by the blocker amantadine [Tosteson, *et al.*, 1994].

The present study was of channel dynamics as revealed through the current versus time record obtained form voltage clamp experiments on single channels. We have employed the approach developed by Liebovitch and his coworkers at Columbia University [cf. Liebovitch , 1993; Liebovitch and Czegledy, 1992; Liebovitch and Sullivan, 1987] to study of the dynamics of the individual states identified by Tosteson. These studies indicate that the dynamics of the M_2 protein channel are chaotic, rather than Markovian. The complexity of a chaotic system is indicated by its fractal dimension which is defined by equation (4) in this case. Our studies reveal a definite change in fractal dimension for these transitions under the influence of amantadine form a value of 1.55 to one of about 1.85. This indicates a change in the dynamics that will be discussed later.

THEORY

A chaotic system is dynamically deterministic and inherently nonlinear. Characteristics of such a system are 1) strong dependence on initial conditions, 2) self-similarity over various time scales, and 3) the presence of a strange attractor. In studies of living systems, the initial conditions remain completely unknown. A strange attractor,

which is a surface or closed figure in the phase space of the system, to which the system orbit is attracted, is also very difficult to identify from electrophysiological data. A self-similarity over various time scales is more easily identified from laboratory data. Therefore, the analysis of biological channels exhibiting chaotic dynamics should be based on the premise of self similarity in time and not on the dependence of subsequent state on initial conditions or on the existence of an attractor.

For a given channel state, identified by the magnitude of the resulting channel current, the probability density function for the state α, $P_\alpha(t)$, is defined by writing the probability that the channel is in state α at time t and remains in that state for the next time Δt is given by

$$P\{\textit{channel in state } \alpha \textit{ for } t \to t+\Delta t\} = P_\alpha(t)\Delta t \qquad (1)$$

The probability that the channel closes in the interval Δt is then $P_\alpha(t)k_\alpha(t)\Delta t$, where $k_\alpha(t)$ is the rate coefficient, referred to as the rate constant if it is time independent. Then the number of channels undergoing a transition from state α in time Δt is given by

$$\Delta x_\alpha(t) = NP_\alpha(t)k_\alpha(t)\Delta t = x_\alpha k_\alpha(t)\Delta t \,. \qquad (2)$$

or

$$\frac{d\ln x_\alpha}{dt} = -k_\alpha(t) \qquad (3)$$

If the system is fractal, $k_\alpha(t)$ is self-similar, i.e.

$$k_\alpha(t) = A_\alpha t^{1-D_\alpha} \qquad (4)$$

Where D_α is the fractal dimension. Integrating (3) using (4)

$$x_\alpha = x_0^\alpha \exp\left[\left(\frac{A_\alpha}{D_\alpha - 2}\right)t^{2-D_\alpha}\right]. \qquad (5)$$

With (2) we have

$$P_\alpha(t) = K\exp\left[\left(\frac{A_\alpha}{D_\alpha - 2}\right)t^{2-D_\alpha}\right] \qquad (6)$$

where K is a constant, which we take as unity.

The dwell number, the number of times the channel is in the state α for a time t and then undergoes a transition out of the state in the next interval Δt, we call $N_\alpha(t)$. We define a density, $f_\alpha(t)$, for the dwell number so that $N_\alpha(t) = f_\alpha(t)\Delta t$. Then, from (2) and (6)

$$f_\alpha(t) = \left[A_\alpha t^{1-D_\alpha}\right]\exp\left[\left(\frac{A_\alpha}{D_\alpha - 2}\right)t^{2-D_\alpha}\right], \qquad (7)$$

which, from (6), is clearly $f_\alpha(t) = -\left(dP_\alpha(t)/dt\right)$.

ANALYSIS

Analysis was of data from single M_2 protein channels expressed in artificial phospholipid membranes separating two compartments of a chamber recorded using a standard voltage clamp technique providing sequences of approximately 170 s recorded at 0.25 ms intervals. If we consider, as an approximation over small ranges of t_{eff}, that $k_\alpha(t_{eff})$ is approximately itself a constant, we have from (7) and the definition $N_\alpha(t) = f_\alpha(t)\Delta t$,

$$\frac{d\ln\left[N_\alpha(t_{eff})\right]}{dt_{eff}} = k_\alpha(t_{eff}) \qquad (8)$$

That is, for small intervals of the time t_{eff} the logarithm of dwell number is linear function of t_{eff} with a (negative) slope of $k_\alpha(t_{eff})$. Figure 1 provides an example of such plots. In the analysis a standard time interval between t_{eff} = 2 units and t_{eff} = 4 units was chosen to obtain $k_\alpha(t_{eff})$.

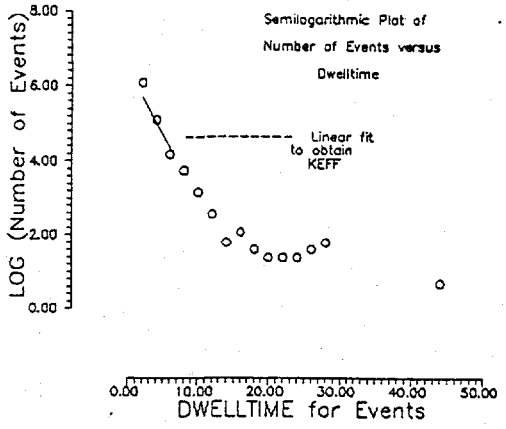


Figure 1 Plot of the natural logarithm of dwell number against the time for an effective time scale of 1.25 ms, i.e. data are sampled every 1.25 ms. The non linearity indicates fractal kinetics.

From (7) it is clear that for fractal channel dynamics the plot of $\ln[k_\alpha(t_{eff})]$ as a function of $\ln(t_{eff})$ is linear. From this plot A_α and D_α may be obtained. Figure 2 provides this plot for four transitions.

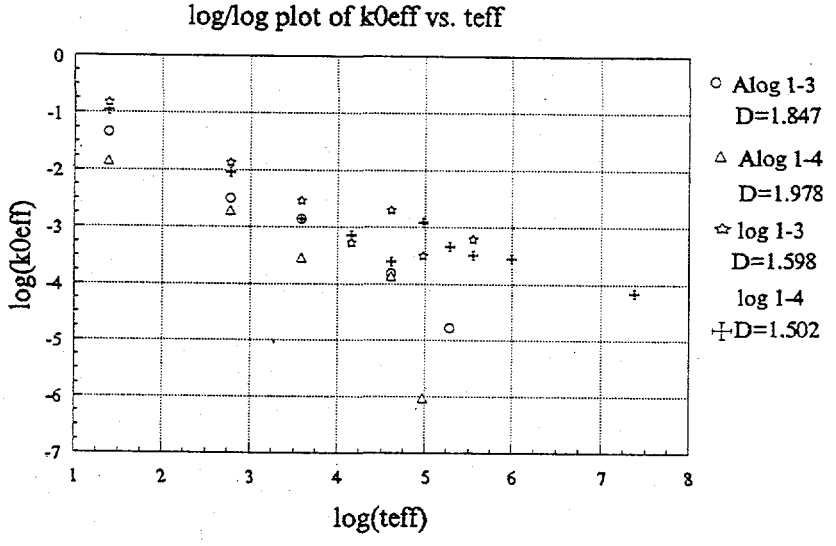


Figure 2 Plots of $\ln[k_\alpha(t_{eff})]$ against $\ln(t_{eff})$ for transitions from the state 1 to the states 3 and 4. The plots with greater slope are those obtained after the channel was bathed in amantadine. The slope of these plots provides the fractal dimension and the position of the line the amplitude of the rate factor.

RESULTS

The rate factors for transitions among the various states are given in the following table. The units for A_α are arbitrary.

Transitions between states	Fractal dimension, D_α	Rate factor amplitude, A_α
1 -- 3	1.589	0.838
1 -- 4	1.502	0.491
1 -- 5	1.552	0.570
2 -- 4	1.673	0.793
2 -- 5	1.504	0.486
3 -- 1	1.670	0.936
4 -- 1	1.505	0.825
4 -- 2	1.604	1.053
A 1 -- 3	1.847	0.935
A 1 -- 4	1.978	0.809
A 2 -- all states	1.808	1.731
A 4 -- all states	1.902	3.890

"A " indicates the presence of amantadine

DISCUSSION

The fractal dimension provides the time variation of the rate coefficient. Liebovitch has pointed out that this time dependence can be interpreted in terms of the paths for transition from the state. Particularly the probability density, $P_\alpha(t)$, may be written as

$$P_\alpha(t) = \int_0^\infty dk\, g_\alpha(k) \exp[-kt]$$

where $g_\alpha(k)$ is the density of the transition, paths k from the state α [Liebovitch and Toth, 1991]. The dependence of $g_\alpha(k)$ on D_α is such that $g_\alpha(k) \rightarrow$ constant as $D_\alpha \rightarrow 2$. A possible explanation for the increase of fractal dimension in the presence of amantadine is, therefore, that amantadine blocks some favorable transition pathways flattening $g_\alpha(k)$. This would be characteristic of an allosteric blocking. Physiological evidence at this time also points toward amantadine as an allosteric blocker [L. Pinto, private communication]

ACKNOWLEDTEMENTS

This work was supported by the Pew Charitable Trust under the auspices of the Pew Science Program in Undergraduate Education. Data were obtained through the laboratory of L.H.Pinto at Northwestern University.

We gratefully acknowledge stimulating and helpful discussion with Professor L.H. Pinto and Dr. Qiang Tu of Northwestern University.

REFERENCES

Lamb, R.A.(1989) The genes and proteins of influenza viruses. In *The Influenza Viruses*, R.M. Krug, ed. (New York: Plenum Publishing Corp.), pp.1-87.

Hay, A.J. (1989). The mechanism of amantadine and rimantadine against influenza viruses. In *Concepts of Viral Pathogenesis III*, A.L. Notkins and M.B.A. Oldstone, eds. (New York: Springer Verlag), pp. 361-367.

Pinto, L.H., Holsinger, L.J., and Lamb, R.A.(1992). Influenza virus M_2 protein has ion channel activity. Cell, *69*, pp.517-528.

Tosteson, M.T., Pinto, L.H., Holsinger, L.J., and Lamb, R.A.(1994). Reconstruction of the influenza virus M_2 ion channel in lipid bilayers. J. Membrane Biol. (in press).

Liebovitch, L.S.(1993). Interpretation of protein structure and dynamics from the statistics of the open and closed times measured in a single ion -channel protein. J.Stat.Phys. 70,329-337.

Liebovitch, L.S. and Sullivan, J.M.(1987). Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons. Biophys.J. 52, pp.979-988.

Liebovitch, L.S. and Czegledy, F.P.(1992). A model of ion channel kinetics based on deterministic, chaotic motion in a potential with two minima. In Annals of Biomedical Engineering. 20, pp.517-531.

Parity Dependence in Level Density of ^{48}Ti

John Quesenberry

Department of Physics
Tennessee Technological University
Cookeville, TN 38505

Faculty Advisor:

J. F. Shriner, Jr.

Angular distributions for the ^{48}Ti(p,p_1) and ^{48}Ti(p,$p_1\gamma$) reactions have been measured at p-wave resonances with incident proton energies in the range E_p = 3.08 - 3.47 MeV. Spins have been unambiguously assigned to 17 of these resonances. By using the observed ratio of $J^\pi = 1/2^-$ to $J^\pi = 3/2^-$ resonances and previous information on the total number of p-wave resonances in this region, we have estimated the total number of $J^\pi = 1/2^-$ resonances. This number is equal within errors to the number of $J^\pi = 1/2^+$ resonances; this agreement supports the hypothesis that the level density does not depend on parity.

INTRODUCTION

In the nucleus, as in any quantum system, discrete eigenstates are a fundamental feature of the system. At low energies (near the ground state), these states are usually relatively simple and well-separated from their neighbors. However as the excitation energy increases, the states on average become closer together. Eventually, the density of states becomes so great that a description of each individual state is not feasible; statistical models then become the most appropriate means of describing the system. A quantity of particular interest in this regard is the level density, the number of levels per energy interval.

The first theoretical work on level densities in nuclei was published by Bethe in 1936 and 1937 [1, 2]. He treated the nucleus as a Fermi gas and obtained the following expression for the level density:

$$\rho(U,J) = \frac{\sqrt{\pi}}{12}\frac{\exp\left(2\sqrt{aU}\right)}{a^{1/4}U^{5/4}}\left(\frac{2J+1}{2\sqrt{2\pi}\sigma^3}\right)\exp\left[-\left(J+\tfrac{1}{2}\right)^2\Big/\left(2\sigma^2\right)\right] \qquad (1)$$

Here ρ is the level density of states with spin J in units of states per MeV, U is the excitation energy in MeV, a is known as the level density parameter, and σ is called the spin cutoff parameter. The parameters a and σ can be determined empirically by comparison with experimental data.

Since Bethe's work, a number of additional refinements have taken place. Gilbert and Cameron [3] modified the low energy form of ρ and included some effects from the nuclear shell model. More recently, phenomenological expressions have been developed by Ignatyuk *et al.* [4] and Kataria *et al.* [5]. Both these models give results similar to Bethe's but with additional correction terms. A common factor in all these models is a dependence on J, the spin of the levels.

Another quantum number applicable to nuclear levels is parity π. The strong nuclear force is believed to conserve parity, and states are identified as $\pi = +1$ or $\pi = -1$ depending on whether their wave functions are even or odd under spatial inversion. An interesting question is whether the level density at a given excitation energy also depends on parity as well as spin. The standard argument has been that by the time the concept of level density is truly meaningful, the energy is high enough and the states sufficiently complicated that parity should not matter. However, data for directly testing this hypothesis is scarce.

To study this question, we performed an experiment to determine the number of $J^\pi = 1/2^-$ levels in the nucleus ^{49}V. A previous measurement [6] of $^{48}Ti[p,p_0]$ and $^{48}Ti(p,p_1)$ in the range of proton energies E_p = 3.08 - 3.86 MeV had determined the number of $J^\pi = 1/2^-$ states in ^{49}V in this energy range and had established the existence of 270 $\ell = 1$ resonances in the p + ^{48}Ti reaction. These $\ell = 1$ resonances can have either $J^\pi = 1/2^-$ or $J^\pi = 3/2^-$, but Li's measurements could often not distinguish the two possibilities. By measuring angular distributions for both $^{48}Ti(p,p_1)$ and $^{48}Ti(p,p_1\gamma)$ for these $\ell = 1$ resonances, we can establish whether they have $J^\pi = 1/2^-$ or $J^\pi = 3/2^-$. A direct comparison of the number of $J^\pi = 1/2^+$ and $J^\pi = 1/2^-$ levels can then be made to examine whether the level density has a parity dependence.

In the next section we discuss the method of identifying the spin by measurement of angular distributions. Then we discuss the experimental arrangement and collection of data. Our results are discussed in the following section, and we summarize our findings in the last section.

ANGULAR DISTRIBUTIONS

An essential aspect of interpreting angular distributions is the coupling of the various angular momenta involved in the reaction. For the p + ^{48}Ti reaction studied here, we must include the spin of the proton, the spin of the ^{48}Ti target nucleus, and the relative orbital angular momentum between them; we shall denote these spins by i, A, and ℓ, respectively. The values i = 1/2 and A = 0 are known; for the p-wave resonances studied here, $\ell = 1$. The possible spins J of a compound nuclear state are those permitted by the vector sum

$$\vec{J} = \vec{i} + \vec{A} + \vec{\ell} \quad . \tag{2}$$

For $\ell = 1$, we find J = 1/2 or J = 3/2 in this reaction. The associated parities are positive for the proton and ^{48}Ti nucleus and $(-1)^\ell$ for the orbital angular momentum; the parity π of the compound state is the product of these three, yielding π = -1 for p-wave resonances. We use the notation J^π to denote spin and parity of the compound state.

In an experiment such as this which studies compound nuclear resonances, the angular distributions can be written as an expansion of Legendre polynomials P_k [7]. The exact form of the expansion is determined both by the various angular momenta in the problem and by the nuclear structure. If a resonance can be considered as isolated, then only even Legendre polynomials appear in the expansion. There are several factors which can limit the complexity of the Legendre expansion, but the operative one in this situation is that $k \le 2J$, where J is the spin of the compound state. So for $J^\pi = 1/2^-$ resonances, the angular distributions $W(\theta)$ for both (p,p_1) and $(p,p_1\gamma)$ reactions can be written

$$W(\theta) = A_0 P_0(\theta) = A_0 \quad , \tag{3}$$

for $J^\pi = 3/2^-$ resonances, the angular distributions can be written

$$W(\theta) = A_0 P_0(\theta) + A_2 P_2(\theta) = A_0\left[1 + a_2 P_2(\theta)\right] . \tag{4}$$

Evaluation of the angular momentum coupling coefficients which contribute to a_2 reveals that $J^\pi = 3/2^-$ resonances can potentially have $a_2 = 0$ for either of the two measurements but not both. Thus if we fit both the (p,p_1) and $(p,p_1\gamma)$ angular distributions using equation (4), a non-zero a_2 in either case is evidence that $J^\pi = 3/2^-$ if we already know that the resonance has $\ell = 1$.

EXPERIMENTAL PROCEDURE

Our experiment was carried out using the 4.0 MeV Van de Graaff accelerator at Triangle Universities Nuclear Laboratory in Durham, NC. Measurements consisted of angular distributions for both $^{48}Ti(p,p_1\gamma)$ and $^{48}Ti(p,p_1)$ in the energy range E_p = 3.08 - 3.47 MeV.

The targets consisted of isotopic ^{48}Ti enriched to 99.1% from its natural abundance of 73.8%. The isotope was supplied in the chemical form TiO_2. It was first reduced by heating in a hydrogen atmosphere and was then evaporated onto carbon backings of thickness 5 $\mu g/cm^2$. Typical thicknesses of Ti were 1 - 3 $\mu g/cm^2$.

For both sets of measurements, the energy of the proton beam was varied until the resonance of interest was located. Then data were collected at the resonance energy for an extended period of time. The overall energy resolution of the system was ≈ 200 - 300 eV FWHM.

The gamma-ray measurements were performed with either 5 or 6 different NaI(Tl) detectors placed at angles ranging from 16° to 160° with respect to the beam direction. All 109 ℓ = 1 resonances previously identified by Li in the energy range E_p = 3.0802 MeV - 3.4246 MeV were studied. Unfortunately, many of the resonances were too close to a neighboring resonance to measure an angular distribution representative of the state of interest. Angular distributions of only 60 states were actually measured, and it was apparent that a significant number of those would probably show the effects of interference with a nearby resonance when analyzed.

After fitting the γ-ray angular distributions and determining the states for which results seemed consistent with an isolated resonance, we measured angular distributions for the $^{48}Ti(p,p_1)$ reaction. These measurements employed 6 different silicon surface barrier detectors located at 90°, 115°, 127°, 137°, 149°, and 165° with respect to the beam. The targets used in this part of the experiment were similar to those used in taking the gamma data. A total of 30 resonances were studied in this run.

ANALYSIS AND RESULTS

The first part of the analysis consisted of extracting the yields at each angle for each resonance. Then both the γ-ray and proton angular distributions were separately fit to equation (4). A fit including the Legendre polynomial P_4 was also performed; a nonzero a_4 coefficient was taken as evidence that there was significant interference from a neighboring resonance, and no further analysis was performed on such states.

If both angular distributions separately gave satisfactory fits, then we examined the two separate fits for internal consistency. A total of only 17 resonances met all criteria. The Legendre coefficients are listed in Table 1 for these 17 resonances. Figure 1 shows typical data and fits for a 1/2$^-$ state, and Figure 2 shows sample data and fits for a 3/2$^-$ state. Note the definite nonzero slope present in Figure 2 which is absent from Figure 1. It is this slope which discriminates a 3/2 state from a 1/2 state.

Of the 17 resonances for which we obtained consistent results, 4 were 1/2$^-$ states and 13 were 3/2$^-$. Assuming that the resonances we were able to fit were representative of the overall sample, we conclude that approximately 23.5% of all p-resonances in this energy range are 1/2$^-$. Since there are a total of 109 known p-wave resonances in this energy range, our data suggests that approximately 26 of these have J^π = 1/2$^-$ and the remaining 83 have J^π = 3/2-. Li [6] had previously established the presence of 34 $J^\pi = 1/2^+$ waves in the same energy range.

To draw any conclusions about whether these numbers are equal requires some estimate of the uncertainties. Unfortunately, such numbers are difficult to obtain accurately, since one really needs to determine the probability of not detecting a level. This probability depends on a number of factors, including the level density itself, the average width, and the experimental resolution. It is most likely that states with very small widths will be missed, since the change in cross section due to their existence is much less apparent than it is for states with large widths. There are statistical tests available which can be applied to a set of

Table 1. Angular distribution coefficients for 17 p-wave resonances in $^{48}Ti(p,p_1)$ and $^{48}Ti(p,p_1\gamma)$

E_p(MeV)	J^π	a_{2p}	$a_{2\gamma}$
3.0802	$1/2^-$	0.039 ± 0.039	-0.030 ± 0.035
3.0873	$3/2^-$	0.298 ± 0.040	-0.577 ± 0.021
3.0933	$3/2^-$	0.223 ± 0.041	-0.689 ± 0.017
3.0975	$1/2^-$	0.030 ± 0.034	0.009 ± 0.035
3.1230	$3/2^-$	0.192 ± 0.039	0.021 ± 0.088
3.1351	$3/2^-$	0.141 ± 0.039	-0.653 ± 0.019
3.1496	$1/2^-$	0.003 ± 0.038	-0.022 ± 0.026
3.1658	$1/2^-$	-0.009 ± 0.038	-0.058 ± 0.028
3.1774	$3/2^-$	0.388 ± 0.037	0.198 ± 0.037
3.1859	$3/2^-$	0.340 ± 0.037	0.114 ± 0.037
3.2244	$3/2^-$	0.021 ± 0.034	-0.581 ± 0.021
3.2399	$3/2^-$	0.218 ± 0.040	-0.047 ± 0.034
3.2455	$3/2^-$	0.469 ± 0.038	-0.247 ± 0.031
3.2804	$3/2^-$	0.362 ± 0.038	0.282 ± 0.036
3.2840	$3/2^-$	0.260 ± 0.040	-0.700 ± 0.030
3.4236	$3/2^-$	0.381 ± 0.041	0.101 ± 0.031
3.4246	$3/2^-$	0.021 ± 0.038	0.581 ± 0.012

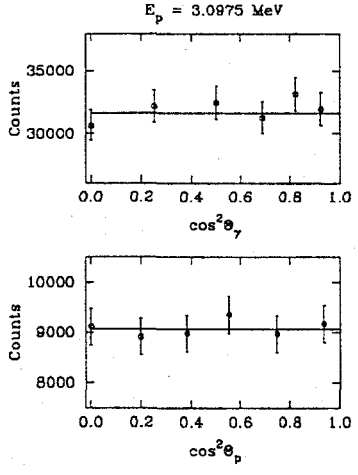


Figure 1. Angular distributions and fits for a typical $1/2^-$ resonance. The upper figure shows the γ-ray distribution, and the lower figure shows the proton distribution.

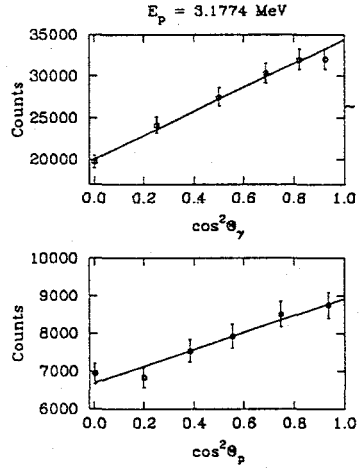


Figure 2. Angular distributions and fits for a typical $3/2^-$ resonance. The upper figure shows the γ-ray distribution, and the lower figure shows the proton distribution.

resonances of the same J^π [8, 9] to estimate the number of missing levels, when applied to the $1/2^+$ levels in ^{49}V, those tests suggest an undercounting of the number of $1/2^+$ levels by about 15%. We estimate the overall uncertainty in the number of $1/2^+$ states to be about 10%. Putting these two factors together gives 39 ± 4 states with $J^\pi = 1/2^+$ in the range E_p = 3.08 - 3.47 MeV. For the p-wave resonances, the uncertainty in J for the majority of levels limits the usefulness of those same tests, but Li *et al.* have estimated that the number of p-waves is actually 20 - 30% higher than they observed. We estimate the uncertainty here to be at least 20%; the larger uncertainty in this case is due to the uncertainty in J for most of these levels. By correcting the number of states by a factor of 1.25 and taking an overall uncertainty of ±20%, we obtain 32 ± 5 states with $J^\pi = 1/2^-$ in the range E_p = 3.08 - 3.47 MeV. The numbers of $1/2^+$ and $1/2^-$ states agree within errors, and we find that this particular measurement is consistent with the hypothesis that the level density is independent of parity.

SUMMARY

We have studied angular distributions of $^{48}Ti(p,p_1\gamma)$ and $^{48}Ti(p,p_1)$ for $\ell = 1$ resonances in the energy range E_p = 3.08 - 3.47 MeV. The initial goal was to establish whether the spins of these $\ell = 1$ resonances was J = 1/2 or J = 3/2. We then planned to compare the number of $J^\pi = 1/2^-$ resonances in this energy range to the number of $J^\pi = 1/2^+$ states to determine if the level density depended on parity for J = 1/2.

The nucleus ^{48}Ti was chosen as a target for this study because many levels had previously been identified in the region of interest. Since the level density is inherently a statistical quantity, studying a large number of levels should provide a better estimate for the level density. Over 100 $\ell = 1$ resonances were studied, but definitive spin assignments could be made for only 17; four of these 17 were J = 1/2 and the remainder were J = 3/2. This leads to an estimate of 26 ± 5 levels with $J^\pi = 1/2^-$ in this 390 keV region, compared to 34 ± 5 levels with $J^\pi = 1/2^+$ in the same region. These data are therefore consistent with no parity dependence for the level density.

In retrospect, ^{48}Ti was not the ideal choice for the target. While a fairly high level density is desirable, as discussed above, the level density in this case proved too high for the angular distribution technique to be consistently successful. Therefore, the uncertainties are significantly larger than one would hope. Additional measurements on a nucleus with a lower level density could provide a more definitive answer on the question of the parity dependence of level densities.

ACKNOWLEDGMENTS

We wish to thank C. R. Westerfeldt and P. M. Wallace of Duke; C. R. Bybee, M. A. Labonte, and G. A. Vavrina of North Carolina State; and B. W. York of Tennessee Tech for their assistance in preparing the accelerator and collecting the data.

REFERENCES

[1] H. A. Bethe, Phys. Rev. **50**, 332 (1936).
[2] H. A. Bethe, Rev. Mod. Phys. **9**, 79 (1937).
[3] A. Gilbert, and A. G. W. Cameron, Can. J. Phys. **43**, 1446 (1965).
[4] A. V. Ignatyuk *et al.*, Sov. J. Nucl. Phys. **21**, 255 (1975).
[5] S. K. Kataria *et al.*, Phys. Rev. C **18**, 549 (1978).
[6] J. Li *et al.*, Phys. Rev. C **44**, 345 (1991).
[7] A. A. Kraus *et al.*, Phys. Rev. **104**, 1667 (1956).
[8] W. A. Watson, III, E. G. Bilpuch, and G. E. Mitchell, Z. Phys. A **300**, 89 (1981).
[9] J. F. Shriner, Jr., and G. E. Mitchell, Z. Phys. A **342**, 53 (1992).

COMPARISON OF LYAPUNOV TIME AND CROSSING TIMES FOR PREDICTION OF ORBITAL INSTABILITIES

Mark C. Lewis

Physics Department
Trinity University
715 Stadium Drive
San Antonio, Texas 78212

Gerald Pitts

Computer Science Department
Trinity University
715 Stadium Drive
San Antonio, Texas 78212

INTRODUCTION

During the 1980's, the fields of non-linear dynamics and chaos came to the forefront of many fields, including celestial mechanics. New computer technology and better algorithms were making long-term simulations of our solar system possible, and many of these simulations pointed to the fact that orbits in our solar system are not stable. There are two basic ways to quantify how stable an orbit is. One way is to measure how long a body stays in the orbit, and the other is to take a measure of the level of chaos the orbit displays. The first can be tested by placing a small body in a nearly circular orbit about the sun; over time the eccentricity of it's orbit will change to the point that its orbit crosses that of a planet. To measure the chaos present in an orbit, we have chosen to use the Lyapunov time, the inverse of the Lyapunov exponent. The Lyapunov time is a measure of how long it takes for two initially adjacent orbits to separate by a given factor. If $d(t)$ is the distance between the two orbits as a function of time then

$$d \div d_0 = e^{\gamma(t-t_0)}, \tag{1}$$

where d_0 is the initial distance, t_0 is the initial time, and γ is the Lyapunov exponent. The value of γ is continuous in phase space, and therefore, all trajectories in a localized region will have similar Lyapunov exponents. It was not until about 1990 that evidence was found for a correlation between the Lyapunov time and the crossing time (Soper, Franklin, and Lecar 1990). Since then, several other studies have been done that confirm this relation to be $T_C \propto T_L^{\,b}$, where T_C, the crossing time, T_L is the Lyapunov time, and b varies slightly for different systems but is usually between 1.7 and 1.9. Lecar, Franklin, & Murison (1992) published numerical results for orbits in our solar system interior to Saturn as well as for satellites in binary systems. For the regions that they studied in our solar system, they found that $b \approx 1.7$. In the binary systems, they worked with orbits with different inclinations. In those systems, the value of b ranged from 1.5 to 2.0 depending on the inclination. Levison & Duncan (1993) took this correlation further out into our own solar system with their results for orbits in the Kuiper Belt. They also found that in regions between the giant planets, the relation $T_C / P_p = 1.4(T_L / P_p)^{1.9}$, where P_p is the period of the nearest interior planet, holds true. When viewing the plots given in both of these papers, one notices a large scatter. This paper presents some preliminary results on work that has been done to reduce this scatter as well a presentation of a methodology adjustment that could increase the precision of these measurements.

MEASUREMENTS

The integrations performed in this research use the mapping method described by Wisdom & Holman (1991). The code was written by Levison & Duncan, and is the same

used in their Kuiper Belt integrations. The system for these simulations was a simplified version of our solar system including only Jupiter, Saturn and the Sun. The crossing time was recorded as the time it took for a minor particle's orbit to cross the orbit of one of the planets, or come within the Hill sphere of one of the planets.[1] The Lyapunov exponent, the inverse of the Lyapunov time, is generally defined as

$$\gamma = \frac{\ln\left(\frac{d(t)}{d_0}\right)}{t - t_0}, \tag{2}$$

where $d(t)$ is the distance between two initially nearby trajectories in phase space. In practice, this is done by creating a shadow particle near each of the particles being integrated at the beginning of the simulation. Because the the value of γ is different in different sections of phase space, it is preferable to keep the shadow particle close to the original. In order to do this, the particles are brought back together at specific times, t_k, during the integration. As described by Wisdom (1983), because the separation between the two particles satisfies a linear differential equation, the absolute distance of separation is irrelevant. We can therefore bring the shadow particle back towards the original along the line between them until they are again separated by a distance d_0. When this method is applied, the Lyapunov exponent at the lth renormalization is calculated by

$$\gamma_l = \frac{\sum_{k=1}^{l} \ln r_k}{t_l - t_0}, \tag{3}$$

where $r_k = d(t_k)/d_0$, and l is the number of times during the integration that the particles are renormalized. This method produces output similar to that shown in Figure 1. If this curve levels out at some value, then it is interpreted as the maximum Lyapunov exponent (or some derivative of that). This is equivalent to (2) as long as the condition

$$d(t) \approx d_0^{1-l} \prod_{k=1}^{l} d(t_k), \tag{4}$$

where d_0 is the distance between the particles at t_0, holds true. If renormalizations are done at constant time intervals, then the number of times the system is renormalized is $l = t_f/t_r$, where t_f is the length of the simulation, and t_r is the time between renormalizations. To see how this varies with d, let us look at some different functions that might approximate the rate at which the particles separate and see how that affects the relation between γ and γ_l. If we assume that the particle trajectories separate at an exponential rate such that $d(t) = d_0 e^{\gamma t}$ and $t_0 = 0$, then $\gamma_l = \gamma$. If we assume linear separation where $d(t) = d_0 + at$ and $t_0 = 0$, then $\gamma_l = l\gamma$. Hence, for linear separation, γ_l grows as $\ln(t)$ instead of going to zero as $\ln(t)/t$. Renormalizing the particles at constant time intervals, one can distinguish between chaotic and non-chaotic orbits though it might be somewhat difficult.

One problem with using a constant renormalization time is that the results have a certain dependence on the length of time between renormalizations. If the renormalization is done too often, then the function $d(t)$ will only display its linear characteristics, and as such, this analysis method will produce output that grows as $\ln(t)$. If it is set too large, then the particles will drift into different areas of phase space, and we lose the benefits of the renormalization process. It is not so obvious that results should be affected by changes in the renormalization time that are near the order of the Lyapunov timescale, but as Figure 1 indicates, this is the case. The first plot in Figure 1 also displays another potential problem with the renormalization time in general, especially with a fixed time interval. It is very difficult to get data for particles that have very short lifetimes. This problem is amplified when trying to collect data for multiple different particles because a renormalization time that works well for one of the particles in the set might be too long for another. Running the two particles in Figure 1 at the same time would require using a renormalization time

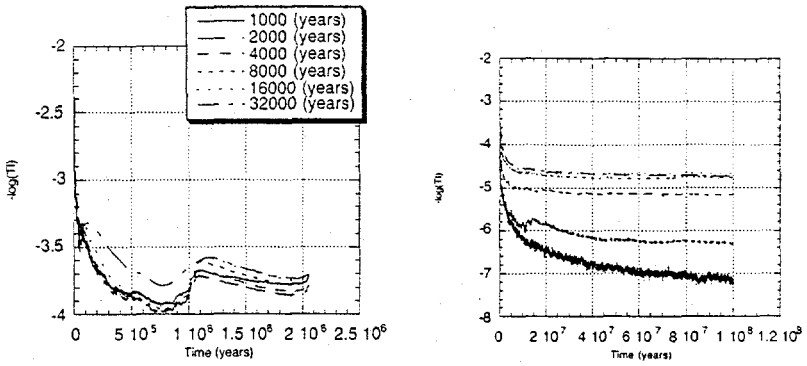


Figure 1. - Plots of the log of γ_i from (3) vs. the run time of the integration for two different orbits outside of Saturn. The different curves in each plot are the output from integrating the same particles but using different renormalization times as indicated by the legend. Notice in the second graph that for short renormalization times, the output increases approximately as $\ln(t)$ instead of converging on $\ln(\gamma)$. This is caused by using a renormalization time that is too short and only allows the particle separation to display linear characteristics.

near 4000-8000 years in order to get reliable data for both of these particles. Because predicting the values of Lyapunov timescales is neither easy nor reliable, generally several runs must be made in advance to find a reliable range of values to use.

One potential way to correct this problem is to have the renormalization triggered by the particles drifting more than a given distance apart instead of by the passage of a fixed time interval. Using such a method, one would calculate the Lyapunov time in the same way as in (3), except now t_r satisfies the equation $d(t_r)=cd_0$, where c is a constant.[i] In this setup, (4) simplifies to $d(t)=d_0c^i$, and as before $\gamma_i=\gamma$. When this method is applied to the linear separation assumptions however, $\gamma_i=(a \ln c)/(d_0c-d_0)$. This could potentially make it impossible to distinguish between chaotic and non-chaotic orbits, because they should both return values that converge on a constant. Though it is not as large a concern with this method as it is when using a constant time interval between renomalizations, there is the

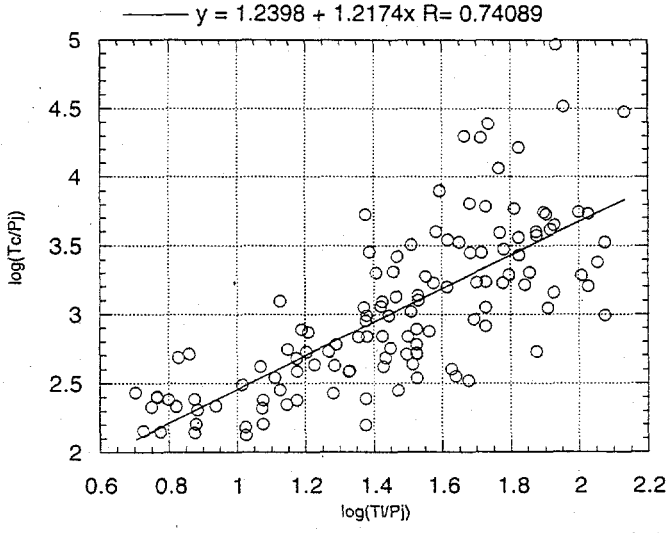


Figure 2. - Plot of the logarithm of the crossing time divided by the period of Jupiter vs. the Lyapunov time divided by the period of Jupiter for valid data on 350 orbits between Jupiter and Saturn.

question of what happens when the renormalization is done too often and the particle separation only displays its linear characteristics. It is interesting to note that the proper choice of c could make this a mute question. By dropping all but the linear terms from the Taylor expansion of the exponential separation case, we get that $d(t_k)=d_0(1+\gamma t_k)$. Substituting this into (4) and then putting that result into (2), we see that $\gamma_l=\gamma/((c-1)\ln c)$. By choosing a value of c such that $(c-1)\ln c$ is equal to one, the output will converge to γ even if the particles are renormalized too often to display their non-linear characteristics. Though it is an interesting fact, we have not been working with this method for a long enough time to test if this is actually an advisable practice. Because the values of c that satisfy this condition are small, .262 and 2.24, we expect that there are many times when this would not be an advisable practice.

In order to insure that the code was running properly, a few simple runs were made to compare the Lyapunov times and crossing times of particles between Jupiter and Saturn. Figure 2 shows a sample of this work. As with previous research, there is a lot of scatter in the data. The reader should notice that the correlation constant for the two sets of data shown is very different from that found by previous research. We do not fully understand the reason for this, but it is possible that this is simply due to the interpretation of the Lyapunov exponent measurements.

REDUCING SCATTER

One of the methods pursued to reduce the scatter in these plots is to analyze small clusters of particles instead of individual particles. By making these clusters small enough that all of the particles in them lie in the same area of phase space, their Lyapunov times should be very similar. Their crossing times however, tend to be much more sensitive to initial conditions, especially for longer-lived particles. If the paths of the particles in phase space are considered, they will be locked into a small volume corresponding to their initial orbit until they happen to cross through a transition area that moves them into another orbit. This behavior of eccentricity jumps has been observed in many previous studies (Gladman & Duncan 1990). Assuming that the median values of the crossing times for an entire cluster will be a more general representation of the crossing time for a given orbit, this will allow us to establish a stronger correlation. Figure 3 shows some of the preliminary results from this effort. There is still quite a bit of scatter in the data, but by comparison with Figure 2 the correlation using the median data points of clusters is better than for individual bodies. It is possible that a more elaborate averaging method using clusters of thousands of particles could lead to even better results, but such a method would be useless as a prediction tool because the computational resources required to calculate the Lyapunov times for thousands of particles would easily outstrip that needed to measure the crossing times.

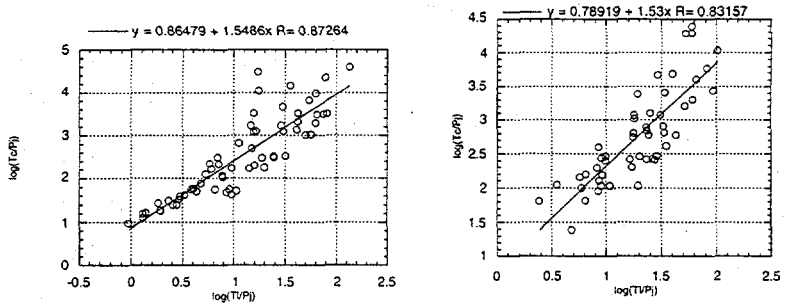


Figure 3. - Plot of the logarithm of the median value of the crossing time for each cluster divided by the period of Jupiter vs. the Lyapunov time or particles in the cluster divided by the period of Jupiter for clusters of 25 and 50 particles in orbits between Jupiter and Saturn. Notice that while there is still scatter, the correlation is stronger than for the plots in Figure 2.

DISCUSSION

There exists a large body of evidence that indicates there is a correlation between the measure of chaos in an orbit and the length of time during which it is stable. This research is aimed at answering three basic questions about this relationship. First, is it universal? Previous work shows that it is present in several different situations; however, there are some examples that have recently been identified of particles in orbits with short Lyapunov times and very long crossing times (Milani & Nobili 1992). Second, is it complete? There are many other factors that could be involved in this relationship such as the period of the main perturber; these other influences need to be isolated. Third, is it accurate? As shown above, there is quite a bit of scatter in the data. We will also consider the possibility that the Lyapunov time is not an accurate measure of the chaos present in an orbit, and that some other method must be applied.

This paper has described some of the preliminary results of this research. To date we have centered most of our efforts on testing the accuracy of the relation. These efforts have met with some progress as we have been able to reduce the scatter in the data to an extent. We have also looked at some of the problems that arise from the current measurement methods, and though we have located some adaptations that look better in theory, we are just beginning to implement these, and as such, definite results on their correctness and usability can not be given at this time.

ACKNOWLEDGEMENTS

I would like to acknowledge H. F. Levison for his work as science advisor on this research, D. Hough as on site advisor, and G. Pitts for providing the computational resources. I also gratefully acknowledge the financial support of the American Astronomical Society (through a NSF Research Experiences for Undergraduates grant).

REFERENCES

Gladman, B., & Duncan,M. J. 1990, *AJ*, 100, 1669
Lecar, M., Franklin, F., & Murison, M. 1992, *AJ*, 104, 1230
Levison, H. F., & Duncan, M. J. 1993, *ApJ*, 406, L35
Milani, A., & Nobili, A. M. 1992, *Nature*, 357, 569
Soper, P., Franklin, F., & Lecar, M. 1990, *Icarus*, 87, 265
Wisdom, J. 1983, *Icarus*, 56, 51
Wisdom, J., & Holman, M. 1991, *AJ*, 102, 1528

---

[i] The Hill sphere is a sphere around the planet with a radius equal to the distance from that planet at which its gravitational force on a body is equal to that of the sun.

[ii] Some preliminary results show that values of c between e and 10e return nearly equivalent results. Of course there are practical reasons for choosing these values. If you are working with particles that have a very short crossing time, it is possible that choosing a small value of c will enable you to get results for more particles. If, on the other hand, all of the particles you are simulating have a very long crossing time, then larger values of c will result in fewer disk writes and could save you disk space. Remember, that if c is chosen such that the Lyapunov exponent varies significantly over a distance cd_0, then the benefit of renormalizing has been lost.

Steady-state analysis of semiconductor laser with feedback loop

Karen Matheny

Physics Depar.nent
Oklahoma State University
Stillwater, Ok 74078

Faculty Advisor: Donna Kay Bandy

Vertical cavity surface emitting lasers (VCSELs) are the newest advance in semiconductor laser technology. They solve the problem of beam divergence associated with the edge-emitting diode lasers by allowing light emission perpendicular to the semiconductor wafer surface. Studies of femtosecond optical pulses make it possible to experimentally examine the dynamics of these diode lasers on a new timescale. Our theoretical model predicts the fundamental behavior of the VCSELs by writing equations of motion for a semiconductor laser in an external cavity. The steady-state analysis indicates that there are two steady-state intensity branches available to this system. Preliminary evidence suggests only the lower branch is stable.

INTRODUCTION

The first commercially available semiconductor lasers were edge-emitting diode lasers. They proved to be of considerable interest because of their potential applications [1-3] to fiber optic communication, laser printing, compact disc players, and many other products. Their use is limited, however, by the nature of their construction (breaking the semiconductor block into chunks) and the necessary testing associated with this product. Also the near flat edges produce a noncircular beam of light that is inconsistent with their potential widespread application. [4]

New construction of the semiconductor laser produces vertical cavity surface emitting lasers (VCSELs) that allow light emission perpendicular to the semiconductor wafer surface and fundamentally resolves the critical problem of high-beam divergence. Low-threshold VCSELs were first produced by AT&T Bell Laboratories and Bell Communications Research. [5,6]

An understanding of the physical processes responsible for the operating characteristics of diode lasers is necessary to improve the performance of these devices. The development of femtosecond optical pulses in the near-infrared wavelength region makes it possible to study these devices on a timescale previously unattainable. To this end, researchers have studied the dynamics of diode laser operation in considerable detail. [7-9]

The semiconductor laser exhibits quasi-independent subsystems that are associated with the characteristic time scales in which electrons in the conduction band, the holes in the valence band, and the lattice itself approach equilibria. This set of quasi-independent subsystems can be modeled by describing the system with additional dynamical variables associated with the temperature or thermal energy of the subsystems. The fundamental equations governing the semiconductor laser first introduced in Ref. [7] are written as follows:

$$\frac{dW}{d\tau} = -\frac{W}{\tau_c} + W(\kappa_e+\kappa_h-\kappa_0)\,, \tag{1a}$$

$$\frac{dn_j}{dt} = -n_j\tau_s - W(\kappa_e+\kappa_h-\kappa_0) + q_j\,, \tag{1b}$$

$$\frac{dE_j}{dt} = -\frac{1}{\tau_j}(E_j - E_0) - \frac{1}{\tau_{cv}}(E_j - E_{j'}) - W(\kappa_e+\kappa_h-\kappa_0)\varepsilon_j + \kappa_f Wh\omega + Q_j - \frac{E_j}{\tau_s}, \quad (1c)$$

$j \neq j'$; $j, j' = e, h,$

where $\kappa_j(n_j,T_j)$ and κ_0 are coefficients of amplification or absorption. κ_f is the free-carrier absorption coefficient and $h\omega$ is the energy of the lasing photon. ε_j is the energy of each particle taken from the conduction band or valence band during the interband transition. Q_j is the carrier energy per unit time transferred by the pumping current, E_0 is the average energy of the electron subsystem within the lattice, τ_c is the lifetime of the photon in the cavity, τ_s is the time of the spontaneous electron-hole recombination, τ_j is the relaxation time of the carriers through the interaction with the photons, and τ_{cv} is the relaxation time of the electron-hole interaction. W is the density of the photons and E_j is the thermal energy of the electron or hole not in the lattice. More information concerning the origin of these equations can be found in Ref. [7].

In this paper we study the steady-state behavior of a semiconductor laser with an external cavity. We briefly discuss the stability of the steady-states and show preliminary results of the temporal trajectories of the laser field intensity.

MODEL CHARACTERISTICS AND STEADY-STATE

VCSELs can be stacked in such a way that laser light from one chip is incident on the next. In order to simulate this, Eqs. (1) are modified to add a feedback loop. See Fig. 1 for a schematic of the geometry. The final set of equations are generated by writing the photon density in terms of the field, incorporating the cavity detuning feature, and making realistic approximations that simplify the equations as follows (see Ref. [7] for details).

$$\frac{dA}{d\tau} = \frac{1}{2}G\Sigma XA - \frac{1}{2}G\Sigma X\alpha B - \frac{1}{2}GA + \frac{1}{2}G\Delta B + \frac{R}{2}GA(\tau - \tau_L), \quad (2a)$$

$$\frac{dB}{d\tau} = \frac{1}{2}G\Sigma XB + \frac{1}{2}G\Sigma X\alpha A - \frac{1}{2}GB - \frac{1}{2}G\Delta A + \frac{R}{2}GB(\tau - \tau_L), \quad (2b)$$

$$\frac{dy_2}{d\tau} = -y_2 + p - X(A^2 + B^2), \quad (2c)$$

$$\frac{dy_3}{d\tau} = -Hy_3 - X(A^2 + B^2)\theta + \eta y_2(A^2 + B^2), \quad (2d)$$

where

$X = y_2 - 1 - \beta y_3$.

$G = \frac{\tau_s}{\tau_c}$, $H = \frac{\tau_s}{\tau_{cv}}$, $\Sigma = \sigma n_0 \tau_c$.

α is the semiconductor enhancement factor, Δ is the detuning between the cavity reference frequency and the laser frequency, and R is the total reflectivity amplitude. β defines the amount of nonlinear interaction with the temperature and η quantifies the free-carrier absorption.

Equations (2) describe the behavior of the laser emitted field - the real and imaginary parts, A and B respectively - the scaled particle density, y_2, and the average energy, y_3.

Setting the derivatives equal to zero and solving for the intensity [$I=(A^2+B^2)$] we find the steady-state solution is a quadratic of the form

$$\eta\, a\, I^2 - (\eta p - \theta a + ab)\, I - ab + (p - 1)\, b = 0, \qquad (3)$$

where

$$a = \frac{1 - R}{\Sigma} \quad \text{and} \quad b = \frac{H}{\beta} .$$

Figure 2 is a typical graph of, H, the scaled rate of spontaneous electron-hole recombination versus the output intensity, I, for a small range of model parameters. We see the steady-state intensity grows monotonically along two branches as H increases; i.e. there are two possible steady-state intensities available to this system. For example, at H=100 there are two values, 26.99 and 213.7, which is shown on the graph. The steady-state graphs, however, are not useful in determining the stability of the steady-state predictions. Only a linear stability analysis calculated around the steady-state can predict the stability of these intensity branches. It is possible to inspect the long-term dynamical behavior of the system and thus obtain an indication of the preferred intensity branch for a given set of initial conditions.

A graph of the temporal evolution of the laser intensity is shown in Fig. 3. The intensity is unstable during the transient phase dynamical solutions to Eqs. (2), but becomes stable at an intensity of approximately 1.5. This value is similar to the lower branch steady-state solution of the quadratic Eq. 3 evaluated under the same conditions.

DISCUSSION

Our preliminary investigation of the semiconductor laser with planewave feedback shows that there are two steady-state intensity branches available to the system and that only one branch is stable over a fairly wide range of the domain H. We can determine the stability of this system through trial and error when the linear stability analysis is not available.

We compare the long-term temporal behavior of the laser to the predicted steady-states and determine that the system described by Eqs. (2) finds the lower intensity branch and resides there stably for values of H greater than approximately 100. There is a discrepancy of approximately 20 percent between the final dynamical intensity value (see Fig. 4) and the lower branch steady-state intensity predictions (see Fig. 2). We attribute this difference to the nonlinearities and phase mismatch of our system and to the fact that the dynamical solution has not arrived to a steady state at the time that the intensity was evaluated in Fig. 4. Phase information is included in unpublished work.

For values of H approximately equal to 60 and with a slightly higher reflectivity than used in Fig. 4, the dynamical behavior is unstable with spiking and periodic waveforms. We believe that both steady-state intensity branches are unstable in this domain and, therefore, the system can reside on neither intensity branch stably-forcing it to oscillate between the two. Future

studies will provide the linear stability analysis, a complete search of parameter space to determine other instabilities (including detuning studies), and additional degrees of freedom by implementing transverse characteristics of the profile. If systems such as VCSELs are to be unstable for any range of parameters, theoretical studies of this system described by the limitations of the model will reveal them.

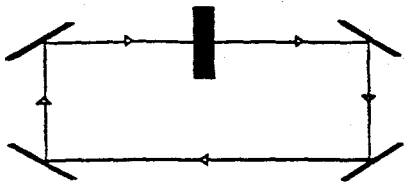


Fig. 1: Semiconductor laser with an external cavity

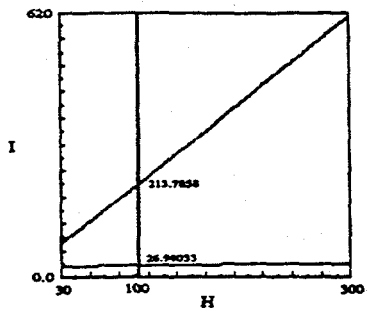


Fig. 2: Steady-State Solutions for H=100, R=0.9333, p=3.0, η=0.1, β=5.0, θ=0.4

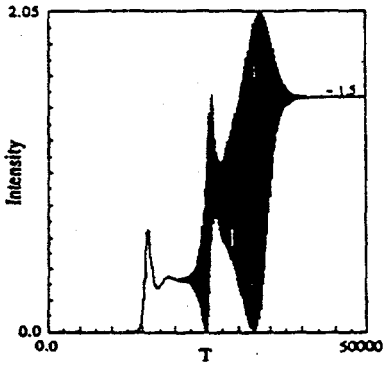


Fig. 3: Dynamics of the semiconductor laser for H=60, G=100, R=0.8, p=1.5, α=4.0, β=0.0

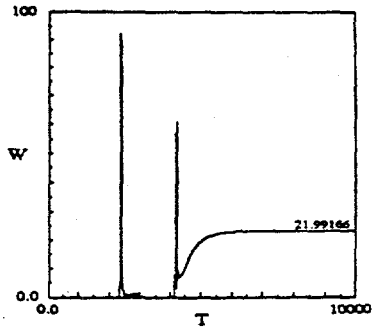


Fig. 4: Temporal Behavior at H=100, G=100, p=3.0, η=0.1, β=5.0, θ=0.4, α=4.0

REFERENCES

1. T. Kimura, J. Lightwave Technol. LT-5, 414(1987).
2. V.W.S. Chan, J. Lightwave Technol. LT-5, 633(1987).
3. M.P. Kesler and Erich P. Ippen, Appl. Phys. Lett. 51 (22), 1765(1987).
4. Jack Jewell and Greg Olbright, Optics and Photonics, March, 8(1994).
5. J.L. Jewell et al., Scientific American, November, 86(1991).
6. J.L. Jewell and G.R. Olbright, Laser Focus World, 28, 217(1992).
7. A.N. Oraevsky, M.M. Clark and D.K. Bandy, Optics Commun. 85, 360(1991).
8. M.S. Stix, M.P. Kesler and E.P. Ippen, Appl. Phys. Lett. 48(25), 1722(1987).
9. For an eclectic collection of papers on VCSELs, see IEEE J. Quantum Electron. 27(1991), and Opt. and Quantum Electron. 2(1992)
10. D.J. Jones and D.K. Bandy, J. Opt. Soc. Am. B 7, 2119(1990)

ACKNOWLEDGEMENT

The Editors and the University of North Carolina at Asheville express their appreciation to:

USDA, Forest Service, Southeastern Forest Experiment Station

Glaxo, Inc.

as the sponsers of the **Proceedings of the Ninth National Conference on Undergraduate Research.**

The Ninth National Conference on Undergraduate Research (NCUR 95) gratefully acknowledges the generous financial support of the following organizations:

AT&T FOUNDATION

THE CAMILLE AND HENRY DREYFUS FOUNDATION, INC.

RESEARCH CORPORATION

GENERAL ELECTRIC FUND

COUNCIL ON UNDERGRADUATE RESEARCH

DR. ROBERT I. PANOFF

SIGMA Xi-SCIENTIFIC RESEARCH SOCIETY

AMERICAN CHEMICAL SOCIETY-EASTERN NEW YORK SECTION

NATIONAL SCIENCE FOUNDATION

UNITED STATES DEPARTMENT OF ENERGY