The California Central Coast Research Partnership: Building Relationships, Partnerships, and Paradigms for University-Industry Research Collaboration

FINAL REPORT
ONR GRANT NO. N00014-08-1-1209
September 30, 2008 to December 30, 2010

Principal Investigator:

Susan Opava, Ph.D.
Dean of Research and Graduate Programs
California Polytechnic State University
San Luis Obispo, CA

March 28, 2011

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
03/25/11	Final Technical Report	9/30/08 - 12/31/10

4. TITLE AND SUBTITLE

The California Central Coast Research Partnership: Building Relationships, Partnerships and Paradigms for University-Industry Collaboration

5a. CONTRACT NUMBER

5b. GRANT NUMBER

N00014-08-1-1209

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Opava,S; Arens,R; Barjami,S; Burd,M; Black,M; Cano,R; Cardinal,T; Choboter,P; Clark,C; Derickson, D; Elghandour, E; Fidopiastis, P; Frame, S; Freed, T; Gillen, K; Goel, R; Hampson, B; Hanson, J; Jimenez, R; Jin, X; Kitts, C; Lehr, C; Lemieux, P; Lin, P; Lundquist, T; MaeCarley, A; Moline, M; Moss, R; Nelson,Y; Sehwartz, P; Singh J; Seng J; Moss, R; Saliklis, E; Taufik; Tomanek, L; Yildiz, I; miseellaneous student authors.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cal Poly Corporation, Sponsored Programs Offiee
Bldg. 38, Rm. 102
San Luis Obispo, CA 93407-0830

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Clifford W. Anderson
Office of Naval Research
875 Randolph St.
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)

ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Available for publie release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The primary purpose of this project is to carry out applied research and development projects and build research capacity in areas of interest to the Department of Defense and the Office of Naval Research. Research areas include communications, computing, command and eontrol, sensors, eoastal monitoring, force protection and performanee, new materials and devices, data acquisition, imaging teehnologies, autonomous vehicles and robots, alternative energy sources and energy efficiency, logistics.

15. SUBJECT TERMS

Liquid crystals, AUVs, SGDBR lasers, path optimization, biofuels, photobioreactors, wind energy, solar cells, electric vehieles, batteries, sandwich panels, slurry applieations, seismic analysis and design, logisties, injury repair, biomarkers, GaN LEDs, photonic lattiees, DC-DC eonverters, passive RFID, knowledge management, computational electromagneties, satellite constellations.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Susan C. Opava, Ph.D.
U	U	U	UU		19b. TELEPHONE NUMBER *(Include area code)* 805-756-1508

Table of Contents

I. Title of Project and Principal Investigator	1
	1
II. Summary of Project	1
III. Relevance to ONR Objectives	1
A. Relevant Partners	1
B. Relevant R&D Focus	2
C. University-Industry-Government Partnership	3
D. University Strengths	3
IV. Summary of results During the Period of Performance	5
A. General	5
B. Development of New Research Capacity	7
1. Instrumentation	7
2. Infrastructure	9
C. Detailed Research Reports	10
1. Collaborative Agent Design Research Center (CADRC)	10
2. Other Research Projects	10
Knowledge Management Project	11
Alternative energy sources and energy efficiency	31
Overcoming barriers to affordable algae biofuels. Tryg Lundquist, Ph.D., Civil and Environmental Engineering; Chris Kitts, Ph.D., Biological Sciences; Corinne Lehr, Ph.D., Chemistry and Biochemistry; Mark Moline, Ph.D., Biological Sciences; Yarrow Nelson, Ph.D., Civil and Environmental Engineering	32
Seawater/wastewater production of microalgae-based biofuels in closed-loop tubular photobioreactors and microalgae production in a vertical photobioreactor. Ilhami Yildiz, Ph.D., BioResource and Agricultural Engineering, Brian Hampson, Ph.D, Food Science and Nutrition	96
Cal Poly Wind Power Research Center: Composite rotor development. Patrick Lemieux, Ph.D., Mechanical Engineering	114
Solar concentrators: quickly deployable combined heat and power. Peter V. Schwartz, Ph.D., Physics	131

Beneficial Reuse of Corrugated Board in Slurry Applications. James Hanson, Ph.D., Civil and Environmental Engineering, Jay Singh, Ph.D., Industrial Technology	140
Development of a Model for Large Scale Cultivation of Microalgae in an Airlift Photobioreactors. Raul Cano, Ph.D., Biological Sciences	149
Battery Laboratory Development. Art MacCarley, Ph. D. Electrical Engineering	162
Force protection in combat and non-combat situations	195
Efficient assimilation of AUV data in a high-resolution coastal ocean model. Paul F. Choboter, Ph.D. Mathematics, Christopher M. Clark, Ph.D., Computer Science and Computer Engineering, Mark A. Moline, Ph.D., Biological Sciences	196
Tracking shallow water squid via an underwater robot system. Pat M. Fidopiastis, Ph.D., Biological Sciences, Christopher Clark, Ph.D., Computer Science and Computer Engineering	206
Path-following algorithms and performance optimization. John Seng, Ph.D., Computer Science and Computer Engineering	212
Feature selection and boosted classification algorithms for pedestrian detection. Samuel J. Frame, Ph.D., Statistics	224
Simplified procedures for seismic analysis and design of piers and wharves in marine oil and LNG terminals. Rakesh K. Goel, Ph.D., Civil and Environmental Engineering	238
U.S.-China collaborative soil-structure-interaction research. Robb Moss, Ph.D., Civil and Environmental Engineering	328
Robotic design and risk: building ethics and safeguards into military applications. Patrick Lin, Ph.D., Philosophy	338
Rapid assembly disaster response shelters made from recycled plastic materials. Edmond Saliklis, Ph.D., Architectural Engineering, Robert Arens, AIA, Architecture	369
The genomics of injury-repair. Trevor Cardinal, Ph.D., Biomedical and General Engineering	381
Bone density measurements via radiographic calibration. Matthew Burd, D.V.M.	394
Assessment of single-use cold-chain packaging and logistics for temperature sensitive products. Jay Singh, Ph.D., Industrial Technology	415

Human performance	432
Construction of a probiotic peptidase delivery system to maximize human health. Michael W. Black, Ph.D., Biological Sciences	433
Extended shelf-life, performance-enhancing, dairy-based food products for the warfighter. Rafael Jimenez, Ph.D. and Kirby Hayes, Ph.D., Dairy Products Technology Center	447
Protein expression profiles as biomarkers for exposure to endocrine disruptors in marine fish. Lars Tomanek, Ph.D., Center for Coastal Marine Sciences	579
New technologies with military applications	585
Full Band Wavelength-Swept SGDBR laser with MHz Update Rates-Sensing and Measurement Applications Dennis Derickson, Ph.D., Electrical Engineering	586
Design and Development of New High-Current Low-Voltage DC-DC Converters Taufik, D.Eng., Electrical Enginecring	594
Investigation of nano-photonic structure on GaN LEDs and solar-cells designs. Xiaomin Jin, Ph.D., Electrical Engineering	669
A novel technique for preventing delamination of the composite skin from the core: Sandwich panels with shear keys Eltahry Elghandour, Ph.D, Aerospace and Civil Enginecring Department	672
Atom transfer from a magneto-optical trap to an atom trap created by diffracted laser light. Katharina Gillen, Ph.D., Physics	795
Isotropic to nematic transition of aerosil-disordered liquid crystals. Saimir Barjami, Ph.D., Physics	805
Inventory Management Utilizing Passive Radio Frequency Identification for the Mission Package Support Facility at Port Hueneme Naval Base. Tali Freed, Ph.D. Industrial & Manufacturing Engincering	811
V. Appendix	
A. High – Tech Breakfast Forum Agenda	859
B. Project Related Thesis and Relevant Publications	864

The California Central Coast Research Partnership: Building Relationships, Partnerships, and Paradigms for University-Industry Research Collaboration

FINAL REPORT
ONR GRANT NO. N00014-08-1-1209
September 30, 2008 to December 30, 2010

Principal Investigator:

Susan Opava, Ph.D.
Dean of Research and Graduate Programs
California Polytechnic State University
San Luis Obispo, CA

March 28, 2011

I. Title of Project and Principal Investigator

The California Central Coast Research Partnership: Building Relationships, Partnerships and Paradigms for University-Industry Research Collaboration; Susan C. Opava, Ph.D.

II. Summary of Project

The mission of the California Central Coast Research Partnership (C^3RP) is to facilitate the exchange of technical knowledge and skills between the higher education sector and the private sector in San Luis Obispo County, and to encourage the growth of high-tech companies in the region, thereby enhancing economic development and quality of life. Since its inception, the project has focused on technologies of relevance to the Department of Defense. The partnership is a long-term plan to create a dynamic and self-supporting university-industry-government partnership that capitalizes on the strengths and mutual interests of the educational and technology-based business sectors. The plan recognizes the key role of higher education in preparing a highly skilled workforce and transferring new knowledge to practical uses. The ongoing outcomes of this partnership include a robust and self-sustaining base of University R&D activities; sustenance and growth of existing technology-based businesses and the creation of new ones; and opportunities for job-training and research and development activities for University and Community College students and faculty.

The project has already resulted in the construction (with non-DOD funding) and launch of the Cal Poly Technology Park on the University campus. This facility provides state-of-the-art space for private technology companies engaged in research and development activities, as well as a business incubator for start-up, technology-based companies. The aspect of the program supported by the ONR grant that this report covers was the continued development of a strong base of applied research at Cal Poly, through university-government-industry partnerships designed to optimize the application of the strengths of each of these sectors to problems of mutual interest. The management team, operational since January '02, continues to lead the project and develop the collaborative relationships between the educational and private sectors that are essential to realizing long-term goals and securing the financial base that will allow full-scale project development.

III. Relevance to ONR Objectives

A. Relevant partners.

C^3RP represents a coalition of educational institutions, local, state and federal government, and private businesses that have worked together in unprecedented fashion to advance the common goals inherent in this university-industry partnership. The current partners in the project and their contributions include:

California Polytechnic State University

- committed the land for the Technology Park construction project, valued at ~$1.5 million
- provided assistance in financial management of the project
- contributed $90,000 for a pre-feasibility study by Bechtel Corporation
- committed several hundred thousand dollars of in-kind contributions of senior management time and effort over several years and continues to do so
- invested ~$1,000,000 in efforts to raise additional funds for the project; secured sufficient private and other federal funding (Economic Development Administration and Housing and Urban Development) to construct the first building in the Cal Poly Technology Park

CENIC (Corporation for Educational Network Initiatives in California; association of Internet2 universities in CA)

- works with Cal Poly to provide high-bandwidth internet access to support C^3RP research projects

City of San Luis Obispo

- in partnership with Cal Poly developed a carrier-neutral, fiberoptic ring around the city that benefits both Cal Poly and technology-based businesses

Housing and Urban Development

- provided funds toward construction of the pilot technology park building.

Economic Development Administration

- provided funds toward construction of the pilot technology park building and development of the incubator space in the building.

Efforts are ongoing to secure new industry partners, including:

- Major corporations
- Small technology-based businesses

B. Relevant R&D focus.

The research programs that were supported are relevant to seven of the eight "thrust areas" of ONR's Code 30 Science and Technology Program. The projects involved basic research in these areas, as well as applied research and development leading to more immediate technological applications. The seven areas of relevance and the more specific focus areas to which the research contributed are listed below:

Command and Control, Computers, Communication: situational awareness; communications; knowledge management; autonomous systems; data acquisition; reconnaissance.

Force Protection: seismic analysis and design; new materials; emergency response shelters; autonomous robots; path-following algorithms and performance optimization; feature selection and boosted classification algorithms for human detection.

Mine Countermeasures: coastal monitoring; IED detection; situational awareness.

Human Performance, Training and Survivability: cognitive performance enhancement; physical performance enhancement; sensors; biological stress reactions; biomarkers; injury repair; improved materials and processes for use on military bases and in the field.

Intelligence, Surveillance and Reconnaissance: data acquisition; sensors; autonomous vehicles; imaging technologies.

Logistics: alternative energy sources; cold-chain packaging and logistics; new technologies and materials.

Maneuver: advanced design and materials for vehicles.

C. University-industry-government partnership.

The primary focus of this long-term initiative is to forge a strong link between private sector R&D and University applied research to speed the development of new knowledge and the transfer of technology to the public and private sectors. San Luis Obispo has become a draw for technology businesses (with a heavy concentration of software development companies) from both the LA Basin and Silicon Valley. For example, SRI (Stanford Research Institute), International operates a "software center of excellence" in the city. Lockheed-Martin has a research and development group in nearby Santa Maria. Two local companies manufactured critical components for the Mars rovers, and other companies, e.g. California Fine Wire, Aeromech, and CDM Technologies are suppliers to the military. Also located on the Central Coast are branches of two major biotechnology companies: Promega Biosciences and Santa Cruz Biotechnology.

D. University strengths.

Cal Poly is a State university that has achieved national distinction as a polytechnic university, with engineering and computer science programs ranked among the very best undergraduate programs in the country. Its strengths have led it to orchestrate the research partnership effort and the consortium of partners described herein. Cal Poly also has affiliations with CSA (California Space Authority) and with Vandenberg Air Force Base, where it recently participated in the Glory Mission. In the past, Cal Poly offered an MS in Aerospace Engineering at Vandenberg through distance learning and has the capability to offer other academic programs in remote locations. In particular, through possible collaborative agreements at cable-head locations around the world (including Asia and Europe) our programs could be made available to military personnel stationed almost anywhere in the world.

In recent years the University has been hiring more research-oriented faculty and promoting applied research and development. With as many as 50-60 new faculty hires per year over several years, the University is positioned to undertake significant R&D projects for government and industry. C^3RP has provided needed support and infrastructure for many of these faculty, which has enabled them to develop ongoing research programs. **Since 7/1/02 faculty who received research support through C^3RP have secured >$8 of competitive funding for each $1 of C^3RP funding invested in them.**

Cal Poly also has a highly qualified student body with entering credentials comparable to students who attend the highest ranked campuses of the University of California. Our students gain valuable experience working with faculty on externally sponsored research projects.

A hallmark of Cal Poly is its extensive network of industry partners. The President's Cabinet consists of more than 30 major corporate and business leaders. Each college, and each department within the college, has its own industrial advisory board. Until recently these connections were not exploited to attract industry-sponsored R&D to campus; hence, one of the goals of the C^3RP partnership is to use these existing relationships with industry to garner support for our R&D efforts. Cal Poly's College of Agriculture, Food and Environmental Science has successfully demonstrated this kind of partnership through its Agricultural Research Initiative. Through this initiative, a consortium of four campuses in the CSU garnered $5 million a year in on-going funding from the State of California to support agricultural research of interest to the State, with a pledge to raise matching funds from industry. This State and private funding has leveraged additional support from the federal government. Similar new CSU-wide initiatives include the Council on Ocean Affairs, Science and Technology (http://www.calstate.edu/coast) and the Water Resources and Policy Initiatives.

As will be seen in the remainder of this report, Cal Poly has extraordinary interdisciplinary technical assets that can be brought to bear on the science and technology issues of importance to ONR.

In summary, the California Central Coast Research Partnership has taken advantage of a confluence of factors, including existing and potential relationships, fortuitous technological and economic developments in the region, the particular strengths and expertise of the Collaborative Agent Design Research Center at Cal Poly, and a meshing of the research and development interests of the University, the Office of Naval Research, and the private sector. C^3RP is the vehicle for fully realizing the benefits of the common goals and synergies of the partners and their respective resources.

IV. Summary of Results During the Period of Performance

A. General.

The C^3RP program was originally funded through an award from ONR in FY '02. This report covers an award that began on 9/30/2008 and ended on 12/31/2010. General accomplishments are summarized below. Detailed reports are presented later in the document. An overview of accomplishments during this project period follows:

- Research carried out by the **CADRC (Cooperative Agent Design Research Center)**, of particular interest to ONR and the Marine Corps, was again funded. A detailed report on this project is provided in Section IV.C.1 of this report.

- **New research** has been developed and some research has been continued, including projects with industry collaboration. Projects addressed topics highly relevant to defense and national security, such as data acquisition, imaging and analysis, energy efficiency, communications, command and control, reconnaissance, autonomous vehicles, sensors, robotics, logistics and human performance. Detailed reports of the results of these projects are presented in Section IV.C.2 of this report.

- From July 1, 2002 to December 31, 2010 C^3RP-supported faculty **received ~$57 million ($56,738,126) in competitive funding from other sources.**

- **New research capacity was developed,** including new instrumentation and enhanced infrastructure (detailed below in Section IV.B).

- Funds were also provided to support small **student research projects** through collaboration with Cal Poly's Honors Program. Talented, high-achieving students in Cal Poly's selective Honors Program were given the opportunity to work on research projects with a faculty member for 1-2 academic quarters and to present their results at a campus symposium at the end of the academic year.

- **Information technology infrastructure support** was provided. **Internet2** connectivity was initially acquired for the campus in November 2001, to support current and future research efforts. Internet2 membership and connectivity has continued during this grant period.

- A **database of >75 technology–based companies** that are potential partners in the project and potential research collaborators has been updated. A series of on-campus research forums initiated in 2007 was continued. Companies were invited to campus three times each year to learn about specific University research projects and identify potential areas for collaboration (see Appendix A for examples of forum topics). Several collaborative relationships have developed from these interactions.

- The first **research and development company to be located on campus** in anticipation of the construction of the pilot building for the technology park continues to flourish in the campus environment and has developed research collaborations with faculty in several different disciplines and colleges. These have resulted in federal research grants (USDA and NIH).The company, Applied Biotechnology Institute, Inc., specializes in the use of genetically modified plants to produce non-food products, for example, industrial enzymes, biochemical reagents and oral vaccines. The presence of the company has spurred faculty to develop research in this area and a specialized research greenhouse supports this developing work. Applied Biotechnology Institute has relocated to the new Technology Park (see below).

- **Cooperative relationships** have been established or renewed with technology companies that are potential research collaborators, including: ARB Green Power (hybrid propulsion), Bard Holding Inc. (algae biofuels), BASF (chemicals), Beebox Designs (software-as-a-service), C2 Robotics (robotics), Cascade Designs (water treatment), DayOne Response Inc. (water treatment), Digital West (information technology), EADS (aerospace), EFuel (alternative energy), Electricore (technology consortium), General LED (lighting), Gluhareff Helicopters (agriculture), GXAerospace, Inc. (aerospace), Hardy Diagnostics (medical devices), Hathway (digital media), LBP Manufacturing Inc. (consumer products), Mafi-Trench Company (industrial processing), Maglio Inc. (engineering services), Mainstream Energy (alternative energy), Moch International (air pollution control), Novonordisk (medical devices), Phillips (lighting), PhycoTech (algae biofuels), Quintron (security systems), Ricoh (digital media), SpiderPak Inc. (packaging), Toyon (radio frequency), Vertigo (software), Vetel Diagnostics (medical diagnostics).

- Efforts continue to develop industry partners in the area of **alternative energy and energy efficiency** for the purpose of developing research in this field. To this end we have worked with Phycotech, First Solar, Continental WindPower, EFuel, Energy Alternative Solutions, Inc., Pacific Gas & Electric, Rey Energy, and Blue Aqua Solutions. Jim Dunning, Project Manager, currently serves on a 9-member board of directors for the Green Coast Alliance, a coalition representing Ventura, Santa Barbara, San Luis Obispo, and Monterey Counties. The alliance was formed to create a framework for innovation, entrepreneurship, economic and workforce development around "green technologies". Through this coalition Mr. Dunning will identify opportunities for collaboration between Cal Poly and business/industry in this important market segment by acting as the interface between the university and interested collaborators. Board membership includes representatives from the Institute for Energy Efficiency, UCSB; Ventura County Economic Development Association; the Community Environmental Council; Pacific Coast Business Times; California Central Coast Research Partnership, Cal Poly; California Lutheran University; Monterey County Business Council; and private businesses.

- The project's leaders have continued to work with other private and government partners to attract research collaborators and support, including the Institute for Energy Efficiency at the University of California, Santa Barbara and the Naval Facilities Engineering Command at Port Hueneme.

- During this project period, we continued to use funding provided by the Economic Development Administration to **complete construction of the first building of the Technology Park.** Construction began in November 2008 and was completed in Spring 2010. To date we have signed leases with six technology-based companies. Two have taken occupancy already; others are completing tenant improvements. Criteria for tenant selection include significant R&D activity in areas that complement Cal Poly's research strengths and a commitment to collaboration with faculty. The six companies comprise four established companies and two incubator companies:

    1. **Applied Technology Associates** - designs, builds, and manufactures oil-industry technology products;
    2. **Applied Biotechnology Institute** - specializes in plant biotechnology, with extensive experience in products targeted for industrial uses, animal and human health and crop-improvement markets;
    3. **Couto Solutions** - specializes in custom design of social networks for enterprise solutions;
    4. **Platinum Performance** - is a leading provider of nutritional supplements and health education products with particular focus on equine health;
    5. **Red Staple** (Incubator) - develops software to help customers create digital-content packages, including LPs and Extras, quickly, affordably and with their own creative assets; and
    6. **Seven Pinnacles** (Incubator) - provides design services for embedded and mobile platforms

B. Development of new research capacity

One of the goals of the project was to increase the capacity of the organization to carry out state-of-the-art research in the areas of interest. To this end, specialized instrumentation was acquired and infrastructure was developed, as detailed below.

1. Instrumentation.

Major Instrumentation

Over the last several years Cal Poly has been building the **analytical capability** to do both molecular identification and microscopic structural analysis, applied to a range of research areas in physics, chemistry, biochemistry, biology, biomedical engineering and materials engineering. Molecular identification is an area that has recently evolved from individual component purification and identification to complex mixture handling and de-convolution. It is now possible to analyze a mixture of peptides, and determine their

identity given the accurate determination of their mass, as well as to identify small molecules or monomer components in polymers. The second area, structural analysis, is heavily dependent on novel technologies such as atomic force microscopy and different optical microscopes based on fluorescence and laser technology. These technologies provide the capability to query the structure of most organic and inorganic structures.

Instrumentation previously acquired to support research in these two areas is listed below. These instruments have supported various research projects in the C^3RP program.

Molecular identification and proteomics:

- Typhoon Variable Mode Scanner and Imager - allows the selection of single spots in 2D gels for latter MS identification.
- HPLC-MS/MS (Chip) - allows the identification of many peptides for proteomics, as well as small molecules or monomer components in polymers.

Structural analysis:

- Atomic Force Microscope
- Laser Scanning Confocal Microscope

During this grant period we augmented our analytical capability by the addition of a **multi-photon, laser system** to the existing Olympus Laser Scanning Confocal Microscope. This greatly increased the capabilities of our imaging system. Multi-photon microscopy has several important advantages over one-photon techniques like confocal microscopy. Unlike confocal microscopy, which only collects a portion of the total fluorescence, multi-photon fluorescence excitation only occurs in a small region, which reduces specimen damage by minimizing out-of-focus photo-damage and photo-bleaching. This allows for the study of living specimens. Furthermore, multi-photon excitation enhances penetration depth by reducing scattering and absorption, which allows for imaging of thick specimens (~greater than 80-100 microns) and turbid specimens. Adding a multi-photon laser system to the existing LSCM has enabled us to expand our research capabilities by allowing us to seamlessly swap existing and new optics between the systems. The new capability is being applied to studies on tissue physiology, optical biopsy, photodynamic therapy, *in vitro* skin equivalent models, normal and malignant cell lines, tissue-engineered blood vessel constructs, the binding forces between bacteria and proteins, the effects of ultraviolet radiation on cell division and development, and the Milk Fat Globule Membrane as a food ingredient with unique functional properties and health benefits.

<u>Other Instrumentation</u>

Other research instruments acquired in the grant period are listed below. Minor instrumentation, acquired for use on individual projects, is described in the reports for those projects.

- Shimadzu analytical-grade HPLC with aerosol detector for use in research on production of cellulosic ethanol.

- Teledyne Combi-Flash Purification System to support two research projects:

    1. Cell Surface Ion Exchangers: Expedient Molecular Targets for a New Brain Cancer Therapy. This project entails the design and synthesis of pro-drugs that target brain cancer cells (e.g., malignant gliomas).

    2. Alkoxyamine Polymers: Versatile Materials for 'Ever-Sterile' Surfaces. This project seeks to explore the utility of the alkoxyamine ($RONH_2$) functional group in new materials as versatile ligating sites for the attachment of antibiotic compounds on the polymer surface. The immobilized antibiotic compounds are expected to provide 'ever-sterile' surfaces that do not require sterilization.

2. Infrastructure.

In the infrastructure area, the following projects were supported. Details are provided in Section IV.C.2.

Development of a Model for Large Scale Cultivation of Microalgae in an Airlift Photobioreactor. Equipment was acquired to support the development of a large-scale model for: mass cultivation of microalgae; enhanced production of lipids for biodiesel; and production of useful chemicals (e.g., natural products, pigments, etc.).

Development of Infrastructure for the Cal Poly Wind Power Research Center. Beginning with the design and development of an in-house process for the manufacturing of wind turbine blades, based on utility-scale turbine manufacturer's standard practices, this project resulted in full development of the first *complete turbine* for the 'Cal Poly Wind Power Research Center' (CPWPRC). Its primary purpose is to provide a physical platform for faculty and students to study technical problems that may be faced in the utility-scale wind turbine industry and to prepare engineering students for careers in that industry. The project entailed the siting study and the design and manufacturing of every component of a fully functional research wind turbine, including: tower foundation, tower, nacelle, and rotor. Completion of the entire assembly and commissioning of the turbine is expected to take place in the summer of 2011.

Development of a Laboratory for Battery Research. The laboratory provides students and faculty with resources to investigate various modern battery applications. Of particular interest is the possibility of electric vehicles that can share power with the electric power grid. The effects of such sharing would involve new battery-power testing that has not been investigated in the past. Specifically the effect of power-sharing on battery life has not been determined.

C. Detailed research reports

The remainder of this report contains detailed individual reports of the technical results of the research projects carried out during this project period. They are presented in the following order:

1. Collaborative Agent Design Research Center (CADRC) project report
2. Other research project reports

Representative publications resulting from this work and documents that supplement the reports are included in Appendix B.

Partnering Enhanced-NLP with Semantic Analysis In Support of Information Extraction

Collaborative Agent Design Research Center (CADRC)

Executive Summary

Information extraction using *Natural Language Processing* (NLP) tools focuses on extracting explicitly stated information from textual material. This includes *Named Entity Recognition* (NER), which produces entities and some of the relationships that may exist among them. Intelligent analysis requires examining the entities in the context of the entire document. While some of the relationships among the recognized entities may be preserved during extraction, the overall context of a document may not be preserved. In order to perform intelligent analysis on the extracted information, we provide an ontology, which describes the domain of the extracted information, in addition to rules that govern the classification and interpretation of added elements. The ontology is at the core of an interactive system that assists analysts with the collection, extraction, organization, analysis and retrieval of information, with the topic of "terrorism financing" as a case study. User interaction provides valuable assistance in assigning meaning to extracted information. The system is designed as a set of tools to provide the user with the flexibility and power to ensure accurate inferencing. This case study demonstrates the information extraction features as well as the inferencing power that is supported by the ontology.

Introduction

During the performance of the work, which essentially stretched over a two-year period, it was decided to expand the research objectives from the limited scope of the *Intelligent Mapping Toolkit* (IMT) to a wider NLP perspective. Specifically, extension of the Context Agent naturally led to the inclusion of a formal context model in the form of an ontology. During the previous 2007-2008 research phase the federation of IMT agents tended to produce a high degree of ambiguity when mapping reference data elements. This was due to the fact that the implemented IMT agents utilized only table-specific features when generating match similarity scores and did not examine features that are capable of distinguishing the semantic differences among references to apparently identical data elements. The introduction of an ontology was seen as an effective mitigating measure.

The document search facility that was proposed as a research objective in the original 2009 proposal was initially envisioned to be based largely on indexing technology. This approach was seen as a way of performing a fast search over the records in all domains, the result of which would then be further redefined by existing inexact search methodologies to pinpoint likely results. Again, it was considered appropriate to extend the scope of the document search facility to a more general NLP analysis. A compelling reason for the change in research objectives from the IMT focus to a wider NLP approach was the increased need for automated information extraction from large quantities of essentially unstructured data within the homeland security domain. However, even from a more general point of view, access to relevant documents has been a critical impediment for persons trying to obtain information on a particular topic of interest. Paper-based documents require the availability of a physical instance of a document, involving the transport of documents with the corresponding costs, delays, and risk factors. Computers and networking infrastructure provide nearly instantaneous access to a huge repository of documents via the World Wide Web and search engines offer support in locating documents that are likely to contain relevant information. The task of examining these documents, extracting relevant pieces of information, and assembling them into a coherent

framework, however, still requires significant human involvement, leading to the "information overload" bottleneck. This research report describes the approach our group pursued in the creation of a system that supports humans whose task is to collect information about a particular domain, analyze that information, assemble it into a coherent framework, and possibly draw conclusions and recommend actions based on knowledge derived from the collected information.

In our context, *information extraction* (IE) refers to the use of computational methods to identify relevant pieces of information in documents generated for human use, and convert this information into a representation suitable for computer-based storage and processing [Wimalasuriya and Dou 2010]. IE is often implicitly constrained to text-based documents, although in principle it can be applied to other types such as images, videos, or audio recordings.

Examining text documents for the occurrence of words is very straightforward for computers, and search engines like Google demonstrate the success of this approach. For IE, however, the goal is to identify meaningful chunks of information, which requires the selection of relevant pieces of text (i.e., words or phrases) and their conversion into a computer-suitable format. Since natural language is ambiguous, redundant, and contextual, the task of identifying and extracting relevant pieces of information is very challenging for computers.

Natural language processing (NLP) refers to the use of computational methods to analyze and process spoken or written statements in a language commonly used by humans. Such methods are applied from different angles. At the *syntactic level*, grammatical rules are used to determine the basic building blocks of text, such as sentences, words, and the roles they play in a given piece of text. At the *semantic level*, the meaning of words, phrases, sentences and documents is determined. At the *pragmatic level*, the context is taken into account as well to determine the most suitable interpretation. Syntactic analysis is relatively straightforward from a computational perspective, but not sufficient to determine the meaning of a text fragment or document; ambiguity, for example, can drastically change the information conveyed in a sentence. Semantic analysis relies on a common interpretation between the creator (writer) and the consumer (reader) of a document. One approach to establish a common interpretation relies on ontologies as frameworks that define the core terminology in a domain and specify the relationships between words. Contextual aspects can be explicitly specified (e.g., through rules), incorporated into a domain-specific ontology, or derived from additional information about the documents and how they are used. Statistical approaches in NLP can overcome this interpretation problem to some degree, and are sometimes combined with the structural analysis methods that rely on rules specifying the grammar of the language.

For humans, speaking the same language is the basis for such a shared interpretation. They use the same principles to construct sentences, although they may not explicitly know the rules of the underlying grammar. For effective communication, they should have a shared vocabulary, meaning that there is a set of words for which the conversation partners have an interpretation, ideally, the same or at least compatible interpretations for one particular word. Contextual information, such as the current location, task, domain of interest, or emotional state of the conversation partner, are considered in cases where multiple interpretations are possible. During the act of reading or listening, humans automatically analyze and interpret language blocks, resolving ambiguities and try to arrive at a coherent interpretation of a document or conversation. Such a holistic treatment of natural language is very challenging for computers. Current systems often combine rule-based approaches, ontologies, and statistical approaches with reasonable success for situations where vocabulary, language structure, or context are constrained.

Participants in the Loebner Prize contest, a refinement of the Turing Test, can engage in conversations on limited topics in such a way that it is difficult for an observer to determine if a conversation partner is human or computer [Loebner Prize 2010].

Our system aims at the extraction of meaningful pieces of information from wide sets of documents, their integration into a coherent framework, and the derivation of new knowledge. One critical assumption is the existence of such a coherent framework for the domain under consideration. An ontology is a formalized representation of such a framework, and serves multiple purposes in our context. First, it makes explicit the knowledge of humans about the domain. Second, it ensures that the interpretation of critical terms is consistent within the group or organization that utilizes it. Third, it spells out important relationships between those terms. Associated with an ontology can be a set of axioms that capture the very basic, generally accepted statements about a domain. The flexible use of relations in ontologies allows the dynamic, multiple classification of entities. While these properties of ontologies already allow for fairly powerful reasoning, our system also incorporates components for reasoning that are external to the ontology.

On the NLP side, ontologies are the vehicle to provide a semantic framework for the interpretation of sentences and documents, enabling the conversion of statements available in natural language into a representation suitable for computers. For the IE task, an ontology helps in deciding which pieces of information may be relevant, and how they are incorporated into the already existing knowledge repository. The combination of axioms and a flexible hierarchical structure provides a strong basis for reasoning and analysis of the information captured in the repository. Ontologies have a natural visual representation as a graph where nodes represent concepts and links to relationships between concepts, and thus serve as a powerful information retrieval method by following apparently relevant relationships.

Ontologies provide major support for several aspects of our system, such as the explicit representation of domain knowledge, interpretation of text, the analysis of documents, and the identification and retrieval of stored information. However, they are difficult and cumbersome to built, may not be available for some areas of interest, and do not capture the full understanding that humans have. The use of ontologies can also become computationally very expensive.

Our overall system is structured as a community of collaborative agents that share a common objective, but have their own capabilities and the autonomy to make their own decisions. Each agent offers a set of services, and in turn may rely on services provided by other agents.

While we experimented with different domains, our main testing ground is the extraction of information about the financing of terrorism from publicly available documents on the Web, with news agencies, newspapers, various organizations, and individuals as sources. Clearly the analysis, interpretation, and integration of extracted information is very critical here, and justifies the use of significant human and computational resources. Extracted information can be extremely valuable, and augmenting the capabilities of human analysts may enable them to find overlooked "nuggets" of information, or to draw conclusions they would not have come to otherwise. Obviously the choice of this domain also has some drawbacks: many relevant documents are not publicly available; and, understanding the environment and context in which such analysts work is limited by security constraints [Flynn et al. 2010]. Nevertheless, we believe that the domain of terrorism-financing is suitable for the demonstration of our system, and later we will present a usage scenario from this domain.

Computer-based IE and NLP methods have become "good enough" to assist humans with the processing of large quantities of text-based documents. They can identify the occurrence of particular text pieces (e.g., words, phrases) in documents, allowing the analysis of simple statements about entities, events, and actions, and the comparison of identified entities against other documents [OpenCalais 2010].

However, in general, the performance of these methods is not good enough to leave the task of information extraction completely to computers, especially in fields where the combination of different pieces of information, possibly from a wide range of sub-fields or related areas, is critical. Thus the goal of our project is to demonstrate that computer-based methods for information retrieval, combined with natural language processing and ontologies, enable analysts to deal with significantly larger sets of documents and obtain better results than they would with conventional methods. Our system provides a reusable, configurable platform that can be adapted to specific domains and tasks with moderate effort. It is intended for interactive and collaborative use, and relies on configurable components for document search and retrieval (e.g., Web crawlers), domain-specific language aspects (e.g., rules to identify entities in the domain such as names or specific terms), visual presentation of entities and their relationships (e.g., as topic/concept maps, or ontology visualizations), observation of activities and intermediate results during the extraction and analysis process, and traceability of results to the source documents.

1. Ontology-Based Information Extraction and Tools

This section covers the various implementation differences in previous ontology-related information extraction systems. The survey by [Wimalasuriya and Dou 2010] provides an excellent overview of variations among ontology-based information extraction systems. We describe the difference in implementations along four dimensions as categorized by [Wimalasuriya and 2010]: information extraction implementation; ontology usage; ontology extraction specificity; and, natural language data source.

The first and probably most significant variation in implementation is how the information extraction is performed in a system. Information extraction itself is a developing field and can be performed using a combination of techniques.

The first information extraction technique is to use regular expressions to match phrases in natural language text. These regular expressions are often constructed by a domain expert to perform matches on phrases as they appear in actual text. While the approach is tedious, it can often yield high quality results. Another information extraction technique is that of a gazetteer list. A gazetteer list is a list of known terms and phrases as they exactly appear in text. When text contains a named entity that matches an element in the gazetteer list, then the named entity is extracted. A third approach is to use a machine learning classifier to classify natural language text as relevant information. This approach uses training data (commonly human annotated text) to train a machine learning classifier to learn how information appears in sentences based on some feature set (e.g., part of speech tags, word position, or capitalization).

The second of four implementation differences is how the ontology is used in the system. Some systems use the ontology as user input that a human has pre-defined. This assumes that all extracted information items must fit into some portion of the defined ontology. Another approach is to have the system dynamically define the ontology as it processes the natural language input. Such a system would create new objects in the ontology as they are identified at run-time.

The third implementation difference is related to what portion of an ontology a system can extract. An ontology information extraction system can potentially extract classes, properties of classes, and relationships between classes. Ontology-based information extraction systems can vary in the level of details for a class that is extracted.

The final variation among information extraction systems is in the source of the natural language data that is processed by the systems. Some systems will use a text source available from a Web site, while other systems require the text to exist in a particular file format.

Research in the information extraction field has been motivated in the past by two research competitions. In the past, the Message Understanding Conference [MUC] was a DARPA sponsored event held from 1991 to 1997. The event required participants to demonstrate systems performing various information extraction tasks. The Automatic Content Evaluation [ACE] program was a NIST sponsored event held from 1999 to 2008.

In terms of the tools that are available, there are several toolkits that target the development of natural language processing applications [LingPipe 2010; OpenNLP 2010; GATE 2010; UIMA 2010; ClearTK 2010]. Two commonly utilized frameworks are GATE (General Architecture for Text Engineering) and UIMA (Unstructured Information Management Architecture). These frameworks provide services and workflows that simplify the construction of NLP applications. For our work we utilize the UIMA architecture. This framework provides facilities such as annotations, chaining of text-level annotators, and an overall modular development environment.

2. General Approach

The objective of this research project is to build a platform for extracting a large volume of information from multiple sources on the Internet and provide a context, within which intelligent analysis can be performed. To achieve this goal, we combine existing NLP tools together using the UIMA framework and partner the enhanced version of NLP with a domain ontology that describes the complex relationships that exist in that domain and some of the basic rules of inference, which describe the necessary conditions for classification of objects and relationships. We emphasize the role of the user in this approach to provide assistance in interpreting the extracted information, review the automated inference, and make any changes to enhance the quality of information.

The architecture of this system is service-oriented, as shown in Figure 1. All the functional components are implemented as Web services and hosted on an application server. The decision to architect the information extraction capability in a service-oriented architecture (SOA) manner was twofold. First, the extraction of information appeared to be an activity that would be invoked by external components as a part of some larger business process and performed by a wide variety of users and on a potentially repetitive basis. As such, it appeared appropriate to deploy the information extraction capability as a whole in the form of a discoverable and subsequently invocable service. Second, since the information extraction capability is comprised of a number of sub-capabilities each of which may be a candidate for replacement with newer, more capable components, it seemed advantageous to carry this service-oriented concept into the internal architecture as well. It was anticipated that doing this would produce a testbed-like environment where internal capabilities are substantially decoupled and primed for potential replacement or introduction of additional, complimentary capabilities.

The current service implementation includes the following:

- Web crawler service to access sources on the Internet and retrieve articles and other material from Web sites, clean it up (e.g., remove HTML tags) and store it in the document repository.
- Feature extraction service to process the textual information, extract objects of interest, as defined by the ontology, and classify some of the extracted objects.
- Model service to provide ontology access and manipulation functionality. The Web Ontology Language (OWL) [OWL 2009] is selected to build the context model. It supports dynamic, multiple classification, which is important in order to describe a person, for example, as both a 'banker' and a 'terrorist supporter'. It also supports decidable logic, which makes the inference possible. The model service utilizes a reasoner in the form of an embedded inference engine. For persistence of model objects we use an RDF [RDF/XML 2004] triple store.
- Tools service to provide additional tools to help the user interpret the extracted information and check the validity of the automated inference. The tools allow for more complex extractions, such as multi-term named entity recognition (NER). Tools also help refine and aid in the classification of extracted objects, using encyclopedic services (e.g., Wikipedia).

In addition to the Web services, there is a document repository that holds the raw textual information as well as the results of NLP extractions. The repository is accessible to all the services, as well as the user interface.

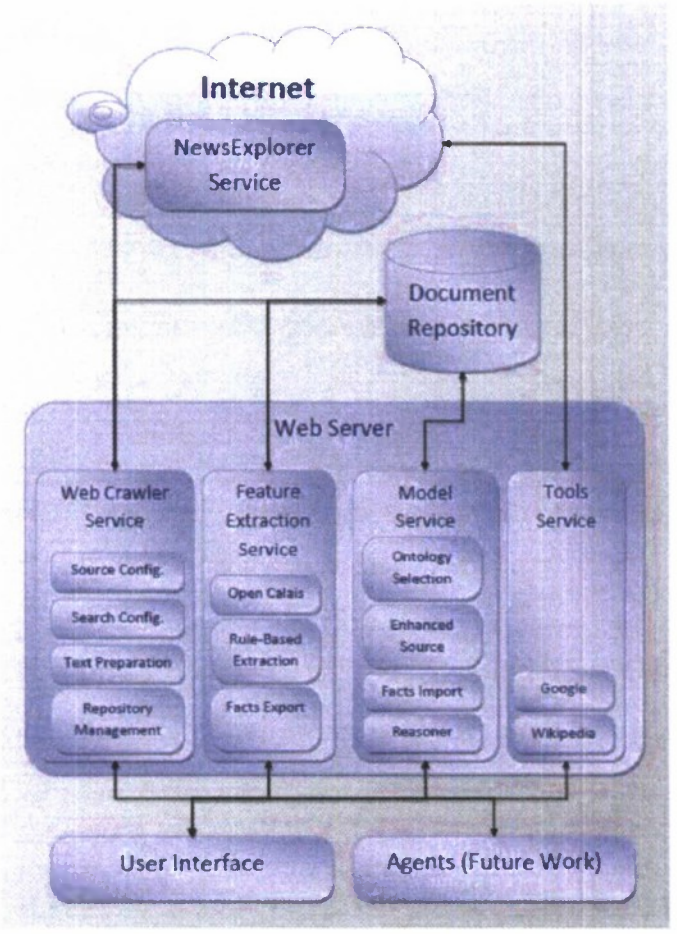


Figure 1. General Architecture

The basic concept of the user interface design is to allow the user to be involved at any step of the process. It is important to have the user input and integrate it into the intermediate results.

The user interface (UI) provides tools for the user to access the services, configure them to perform according to the task at hand and provide input when the user decides to intervene.

The UI is Web-based, so that it can be run from any location and support collaboration. The UI offers support for configuration of information sources, setting of various selection criteria, user-initiated extraction of basic (i.e., explicit) concepts, and user-initiated exporting of NLP extractions into semantic model. The user can also edit the system configuration elements, such as the extraction rules and classification lists.

The reasoning process is triggered by the user and can be configured easily on the UI. The user can select an ontology for reasoning and can load multiple ontologies to support complex inferencing.

The presentation of the inferencing results can be accomplished in a tabular form as well as on a graph, depicting the concepts and their relationships. The results are editable and in the final design can be integrated into the on-going inferencing process.

The UI also offers access to a set of services that provide tools to enhance the interpretation of the extracted information.

3. Use-Case Overview

We present a use-case here to illustrate the benefits of enhancing the traditional NLP capabilities with ontology-based inferencing capability. The use-case supports intelligence analysis of terrorist activities based on information extracted from news sources as well as intelligence reports. The assumption here is that different pieces of information appear in different documents from multiple sources using related but not identical terms, and that their combination yields an interesting or important new piece of information. Information may be obtained from a number of sources, with overlapping data items (e.g., a person or an organization may be mentioned in multiple news articles). All of the statements are stored in the repository, including any redundant information. When statements are imported into the ontology as facts, a check on information is performed against the current state of the ontology. Information about the same object (i.e., person, organization, location, etc.) is merged together to ensure information integrity.

The use-case examines information about people, organizations, and meetings. The relationships among these three types of objects are also examined (e.g., person's membership in organization, meeting between two people, etc.).

The scenario for the use-case is as follows:

Purpose:

- To identify the possibility that a suspected terrorist group (or individual) is planning an attack.

Explicitly extracted information:

- Membership in a known (or suspected) terrorist group.
- A person employed by a financial institution.
- A meeting, taking place between the members of the terrorist group and the banker.

Inference:

- The terrorist group is planning to transfer a large amount of money through the banker.
- A terrorist plot may be in the making.

This simple use-case illustrates the extraction of information from multiple sources (the person's membership in the terrorist group may come from one source, while the meeting between this person and the banker may come from another). It also illustrates the use of an ontology to build a context for the extracted information and the use of the ontology inferencing capability to create new information.

4. NLP Environment

This section covers the implementation and design of the natural language processing server portion of this project. The NLP environment implements the Web crawler service and the feature extraction service, as shown in Figure 1. The NLP server architecture consists of a

combination of very general external Web resources as well as domain-specific features. By utilizing both external Web resources and features tailored to a unique domain, we are able to extract natural language features that may be missed by a more broad-based approach while maintaining an acceptable level of generality.

From a high-level perspective, it is the task of the NLP server to download news articles from the Internet and extract relevant information from key phrases present in the articles. The extracted phrases, which we call *assertions*, are then sent to the semantic model. The NLP server acts directly with the graphic user interface (GUI) application and therefore, the user can directly send requests to the NLP server via SOAP Web service calls from the user application.

The following subsections outline the server structure. The next subsection describes what criteria are used to acquire news articles for processing. Subsequent subsections cover: the task of cleaning the HTML for a Web article; how natural language features are extracted from our system; and, the network implementation of the NLP server along with a description of the exposed application programming interface (API).

Article Acquisition

In a high-level view of our system, the NLP portion takes as input natural language text and produces assertions relating to the relevant domain. Our goal is that the natural language input may come from a variety of sources. Examples of online sources include: major news Web sites; online forums; and, blogs. In addition, static documents such as various technical reports may be utilized as well. In our current implementation, the source of natural language content in the system is news articles gathered from the Internet. News articles are written for readers who may not have extensive background knowledge on a topic and therefore the articles will often list people's full name and titles. This is advantageous for detecting the political and corporate positions of various people.

We use an online news article collection Web site NewsExplorer [NewsExplorer 2010] developed by the European Commission's Joint Research Centre. The Web site performs statistical analysis of various news articles that are released throughout the day. Some sample features of the NewsExplorer Web site are the ability to cluster articles by common terms, country of origin, and primary article topic.

For our work, we utilize a subset of the topics that NewsExplorer supports. Our graphical application currently allows downloading of articles that are categorized under one of the following topics: Security; TerroristAttack; Conflict; Energy; and, AlternativeEnergy.

Article Cleanup

The NewsExplorer Web site provides Web links for each of the articles that it tracks. In order to download an article, the article itself must be cleaned of any HTML tags and content not relevant to the news article. For this task, we use the HTML cleaning functionality of the AlchemyAPI NLP Web service [AlchemyAPI 2010]. Given a Web link of an article, this service returns the raw text of the news article void of any HTML markup tags or any other advertising content. Once an article has been cleaned, a local copy is cached for future use.

Natural Language Feature Extraction

Our system utilizes two methods of extracting information from natural language text: OpenCalais; and, our own rule-based technique.

OpenCalais [OpenCalais 2010] is an information extraction service provided by Thomson Reuters. The service takes natural language text and returns a document with various labeled entities in resource description framework (RDF) format.

For our current implementation, we use only a subset of the entities that are detected by OpenCalais. The entitics that are stored are `PersonCareer`, `PersonCommunication`, `PersonRelation`, and `FinancialInstitution`.

`PersonCareer` is a relationship that states a person's full name along with an institution that the person is affiliated with. The institution may be a company, political organization, or other institution. This `PersonCareer` relationship does not necessarily specify an employee relationship.

`PersonCommunication` relations exist between two or more persons. The communication can be an actual communication event, visit, or meeting. The status of a `PersonCommunication` can be either announced or planned. An announced communication occurs when the communication occurred or allegedly occurred. A planned communication is one that will happen in the future.

In addition to the entities that OpenCalais detects, we implement a rule-based detection system, which implements matching based on regular expressions. As an example application of our case study, we focus on detecting the names of terrorists and their membership with particular terrorist organizations. Our regular expression patterns are based on a list of known terrorist organizations. We allow for minor spelling differences in the names of the organizations and assume if the spelling is within a threshold of character differences then the names are a match.

In each pattern, there are keywords that are used to match types of words in a phrase. For example, the keyword `terroristorganization` matches with a list of known terrorist organizations that is read from a configuration file. The keyword `adjectives` represents a list of known common adjectives that are used in the context of terrorist members and leaders. The `person` keyword refers to the name of a person. We use the detection of persons from OpenCalais. `Person_role` represents various roles or positions that a person can take (e.g., member, leader, director). Other standard regular expression symbols are used for matching whitespace, articles, and punctuation.

As the assertions are extracted, they are inserted into an intermediate model using Jena [Jena 2010]. This Jena model maintains occurrence frequency for each assertion as well as the pertinent information for the assertion (e.g., terrorist organization name, person name, or communication status). Once the assertions are stored in an intermediate model, they are then transferred to the semantic model Web service. This information is stored in an RDF file that is explained in the semantic model section of the report.

We ran across some limitations of the OpenCalais Web service. OpenCalais does perform co-reference resolution with pronouns, but has difficulty when the reference does not utilize a personal pronoun but instead uses a plain noun to refer to an object. For example, a paragraph may refer to 'the president' several times without reference to the person's actual name. In this case, OpenCalais will have difficulty in making the connection with the particular reference to the president's actual name.

Another limitation of OpenCalais is that the relationships it extracts are limited to certain pre-defined relationships that exist in the RDF schema. Although the schema is quite extensive, there are certain instances where it is more powerful to implement a domain-specific relationship extraction technique - like our rule-based technique. In our approach, we used a rule-based

system to extract information regarding terrorists. This is a quite domain specific extraction and could not be extracted by OpenCalais.

NLP Web Service Implementation

The NLP Web service is implemented as a SOAP service that the end-user application can invoke. There are two primary components of the NLP service: a thin wrapper layer; and, the backend server module. This section describes both service modules.

The thin wrapper layer is a Tomcat Web service that provides a Web interface to the user application. This layer provides an API that allows applications to perform various NLP related and file system related tasks. This layer communicates with the backend NLP server over a socket-layer protocol.

Backend NLP Server Implementation

The backend NLP server module is the component of the NLP server that performs all of the primary work. It is threaded to handle requests from multiple requesters.

There are two primary service calls that the backend server handles: `BatchDownload` and `RunBatch`. `BatchDownload` is an API call that invokes the backend server to download a number of articles.

`RunBatch` calls the OpenCalais Web service for each of the documents that was downloaded using `BatchDownload`. OpenCalais returns an RDF file containing the extracted relationships. This RDF file is cached locally for future use and the assertions within a particular RDF file are compiled into a statistical list of assertions for a given batch.

5. Semantic Environment

To extend the solution into the realm of semantic analysis, the enhanced NLP environment described above is mated to the second component of this two-part approach, the semantic environment. This additional environment consists of a Web Ontology Language (OWL-based) context model that is managed within a Jena-based platform equipped with an embedded reasoner. The resulting model and inference capability is collectively exposed as an invocable Web service.

The following section presents the representational paradigm selected to represent the respective domain ontology together with applicable inference logic. Within this discussion, several key modeling facilities are highlighted and related to the information extraction problem.

The OWL Representational Paradigm

At the heart of the semantic platform is the context model, or ontology, which contains the concepts necessary to describe and reason about the extracted information fragments produced by the enhanced NLP capability. Particular attention was given to the selection of the appropriate representational paradigm. This paradigm must provide a somewhat relaxed view of classification along with native support for embedding inference logic within the very concepts that it semantically relates to. Following investigation into several more traditional paradigms, we decided upon OWL as the representational language and execution platform suitable for this research.

The Web Ontology Language, or OWL as it is more commonly known, is a semantic markup language. The primary purpose of OWL is to facilitate the publishing and sharing of ontologies

across the World Wide Web (WWW). OWL is intended to be used by software applications that need to process Web-based content in a meaningful manner. That is, OWL-based content is designed to be machine-interpretable.

A typical OWL environment consists of several key components that support both model development as well as subsequent execution. A typical OWL execution platform consists of a triple-store where model content is persisted, a reasoner capable of inferring additional concepts based on embedded logic, and a query engine used to ask questions regarding model content. Together, these components form a cohesive platform for the development and execution of semantic content.

Perhaps the most important component of any OWL environment is the reasoner and as such warrants a more detailed description of its tasks. As the name implies, the main function of this component is to essentially *reason* about a given OWL model and its associated content. More specifically, an OWL reasoner processes class definitions, individuals, and embedded rules in an effort to accomplish two primary objectives. The first of these tasks is to identify any logical inconsistencies existing within the model definition and its use. As the platform produced by this research supports user-driven importation of ontologies applicable to the target analysis domain (e.g., intelligence, logistics, command and control, etc.), this feature can be used to verify decidability and logical consistency of such models. The second task performed by the reasoner is to identify any additional knowledge that can be automatically inferred based on the logic embedded within the model definition in conjunction with associated content. This additional knowledge can include subsumption and association relationships and the adjustment of classification of various scenario content. Although only beginning to emerge within the timeframe of this research, some reasoners have the ability to not only infer additional content, but to retract previously inferred content whose truth can no longer be established (i.e., truth maintenance) [BigOWLIM 2010]. This is a crucial feature for any practical application of automated reasoning as establishing what is no longer true is as important as the initial inference that produced it.

Above and beyond the components comprising its typical platform, the OWL representational paradigm supports several very powerful concepts that are thoroughly exploited by the information extraction process. These concepts provide the flexibility to represent as of yet unclassifiable extractions as well as to repeatedly adjust the definitions of those that can be classified.

Multiple Classification

Multiple classification is the ability for something to be simultaneously defined under two or more classifications. This is a very powerful capability and has significant implications on the manner in which representational models are developed. Unlike traditional, more rigid modeling paradigms where inheritance must be employed as a means for the specialization of concepts, OWL modelers enjoy a much more flexible environment without concern for relating classifications in order to support a single entity exhibiting features defined across multiple classifications. Once qualification rules are embedded within class definitions, the management of exactly which classifications are appropriate at any given time can be conveniently offloaded onto the OWL reasoner.

Dynamic Classification

Dynamic classification is the ability for the classification of something to be adjusted throughout time. In contrast to the traditional approach of re-instantiating an entity under a new classification, dynamic classification offers the means to preserve referential integrity by maintaining the existence of the original entity by only changing its type information. This capability goes hand-in-hand with multiple classification in creating a dynamic environment where extracted facts can effectively mutate throughout their lifecycle as additional knowledge is encountered. Like management of multiple classification, determining exactly what classification(s) an OWL object qualifies for at any point in time is typically the responsibility of the OWL reasoner.

Open World Assumption (OWA)

Traditional database systems operate under a set of assumptions in order to enable the query engine to return meaningful responses. These suppositions include the *closed world assumption*, the *unique name assumption* and the *domain closure assumption*.

The closed world assumption states that if a statement cannot be proven true, given the information in the database, then it must be false. The unique name assumption states that two distinct constants designate two different objects within the given universe. The domain closure assumption states that there are no other objects in the universe than those designated by constants within the database.

These assumptions were reasonable in a world where a database represented all the information available about a given domain and no external information sources were needed to perform the functions of any database application. Since this time, however, the Internet has become a major source of information with the effectiveness of many applications being based on access to external information from sources that may be unknown at the time of application design. This has required a different kind of knowledge representation, capable of dealing with the *openness* of the Internet. The *open world assumption* was adopted to allow for the relaxation of the constraints imposed by the closed world assumption. Along with the open world assumption, the other two assumptions were consequently also relaxed.

Under an open world assumption, all things are possible unless asserted otherwise. This is in stark contrast to traditional decision-support paradigms where the volume and extent of considerations is limited to what is explicitly asserted to be true about the world at any given time. Although operating under an open world assumption has implications on model development, it is primarily model usage and interpretation that is affected since logic operating across such a model must refrain from assuming too much.

Unconstrained Composition of Features

Being able to describe the characteristics of something in a manner unbound by available definitions can be very helpful when dealing with evolving knowledge, such as in the information extraction problem domain. With environments equipped with such flexibility, information extractions can be characterized and continuously re-characterized in a manner unconstrained by current classification(s) or known concepts. Working in conjunction with dynamic classification, as the characteristics of something evolve or otherwise change, so may its alignment with available definitions. As described earlier, determining exactly what changes in a classification are appropriate is typically the responsibility of the reasoner.

The complementary partnership between this set of representational concepts is one of the most powerful aspects of the OWL paradigm. Collectively, these mechanisms support a progressive and fluid understanding of extracted information fragments within a dynamic and constantly evolving context.

Semantic Model Service

Our implementation of the semantic environment is the semantic model service (referred to as the "model service" in this section). This SOAP Web service contains the ontology-based semantic processing for the system. From a high level perspective it is the role of the model service to take in assertions from the NLP extraction and derive new knowledge via inference over those assertions in combination with a domain specific OWL ontology. To accomplish this task the model service controls the importation and persistence of NLP extractions and an OWL ontology into a triple store. The model service is then able to execute reasoning over the triple store and query its contents.

In order to produce new facts that can be traced to their original source the rules used for reasoning must be enhanced beyond those typical for OWL inference. The structure of the facts within the semantic repository also needs to be modified to contain source information. In the next section we will discuss the added reasoning and source references as implemented in our data model and in the following sections discuss the implementation of the model service itself.

Source Enhancements for Reasoning

For a real world scenario such as our central use-case seeks to represent, the reasoning and expressiveness of OWL is stretched. Our dilemma is similar to that expressed by [Vrandecic et al. 2006]; - a bare fact stored as a triple cannot express information about its source, reliability, or how it was derived. This information about the *context* of a fact is as important as the fact itself in our solution.

To express this context each fact within our semantic repository is annotated with a source object. This object can be of several types and contain varying properties. A source object for facts imported from the NLP extraction is of type `NLPAssertionInfo` and contains information such as the source document and the place in the document where the extraction occurred. This annotation could be applied through the use of named triples, named graphs, or RDF reification. In our case we use RDF reification, since it can be used with any type of triple store.

However, our needs go beyond simply annotating facts. New facts that are inferred will need source data as well. A new fact's source is the inference itself. That inference as well as the facts used in the inference need to be made available to the user. Because of this we have taken the rules for typical OWL inferencing and extended them to create proper sources for inferred facts with type `SemanticAssertionInfo`.

This technique radically increases the size and complexity of rules. We have yet to discover its suitability for large datasets. It is possible that systems that deal with reification naively with named triples will allow for better scalability. Future work on the model service includes documenting more specific needs for our triple store and rule engine regarding performance requirements and facilities for advanced reasoning.

Service Structure and Implementation

The model service has three internal components responsible for ontology management, import translation, and querying and exporting data. In this section we will briefly explain each component's purpose and key features.

The ontology management component interacts with and maintains the internal triple store. It is also responsible for triggering the reasoner and persisting the resulting inferred facts. Our triple store is implemented in Jena. The triple store contains both the OWL domain model and facts received from the NLP extraction. After the reasoner is run the triple store also contains inferred facts. Because the triple store is persistent new facts can be constantly added to the existing data to trigger further inference.

Import translation is the process by which *assertions* from the NLP output become *facts* within the semantic model environment. This translation is necessary because the NLP model and the semantic model use different ontologies to describe entities. While it is theoretically possible that the NLP and the semantic model could use the same ontology, it is impractical for several reasons. From a reusability standpoint the NLP should not have to be configured to the current ontology. For example, two instances of the system might have similar data sources and extraction needs (thus the same NLP model) but different reasoning and ontology needs (thus different semantic models). From a practical standpoint it makes sense to have the NLP output its assertions in formats close to their sources with minimal processing and translation since it doesn't have, nor should it have, the semantic tools available to do more complex translations that the system might require.

The NLP output is based around the concept of an assertion, a single instance of NLP extraction. This extraction contains the extracted information as well as the source. Furthermore, the extraction represents an individual who is a member of a terrorist organization. The import translation takes this assertion and breaks it into its component parts. The parts are the existence of a person, the existence of a terrorist organization and the membership object that represents their relationship. Each of these component parts (in fact, every triple that comprises them) is tied through RDF reification to the source object. The import translation also makes sure that new entities are checked against existing entities of the same type and name. If a match is found, then new assertions will be connected to the existing entities thus unifying data over multiple sources.

At this time user interaction with the import mechanism is strictly through processing of the NLP extraction, however in the future it is likely that users will desire to add facts to the model directly. Therefore, the import translation component is likely to implement translations from a user-friendly schema into the model schema in the future.

On the opposite side of the import process is the export component. This component queries the model for desired facts and translates them to a format that can be served to clients needing the information. The complete export can contain hundreds of such assertions including relationships (such as `memberof`) across objects. Currently the export process uses hard coded queries to create an output RDF. The output RDF is modified with regular expressions to create an XML document that is more palatable to the user interface client. In order to serve data to a more diverse range of clients the export translation process will need to be expanded to support general querying and possibly Extensible Stylesheet Language Transformations (XSLT) as well.

6. User Experience

The design of the user experience is focused on providing the user with a set of tools that can be configured at run-time, depending on user needs, to perform different tasks. The accuracy levels of existing NLP tools make user involvement essential in providing meaning and guiding both the extraction and the inferencing processes. The application is developed as a Web application for availability to multiple users and easy access.

The information source can be a given repository location on the Internet, or can point to an information aggregation Web service, such as "NewsExplorer". Information collection Web services can be used by simply setting the application address in the address box. The overwhelming volume of information requires that information sources be searched for documents satisfying certain search criteria. Currently, the search criteria include a date range and a set of keywords. Search criteria will expand to include additional filters, such as search-by-organization, search-by-person, and search-by-event. The collected information is downloaded to a local repository for further processing. At this point, the user has the tools to manage the repository by checking the downloaded documents for relevance, deleting some documents, moving documents to other folders, or downloading more documents.

Once the documents in the repository are ready for processing the user can start the extraction process on a selected folder. The extraction service creates an extraction file for each document in the folder, then consolidates all extractions into one file, eliminating redundancies and resolving references. The user can run the extraction process on many folders at the same time, in which case a separate extraction session will be created for each folder. The user can also stop any extraction session before it is completed, and decide whether to keep the information extracted so far or delete it.

The user can run the inference engine on any set of extracted concepts to see what insight can be gleaned from the current document set. After setting up the service, the user can select an ontology to load. The selected ontology reflects the current interest and the required analysis. The user can clear the current ontology and load another one to perform a different analysis. Once an ontology is loaded the user can load extracted concept files, which are used for the analysis, then run the reasoner to automatically perform the analysis. The analysis results may include the creation of additional objects and/or relationships.

The analysis results are presented in tables as well as in a graph, depicting the extracted concepts and their relationships. The inferred objects/relationships are included in the presentation and marked as 'inferred'.

The user can load multiple extraction files and perform the analysis on all of them, or load the files one-at-a-time, and perform the analysis incrementally. The user can also clear the current analysis and start over with a new set of extraction files.

In addition to the NLP extraction and the inference engine, the application offers tools to assist the user in the analysis, by finding out more information about specific concepts or identifying the types of some concepts that may have been missed by the extraction process.

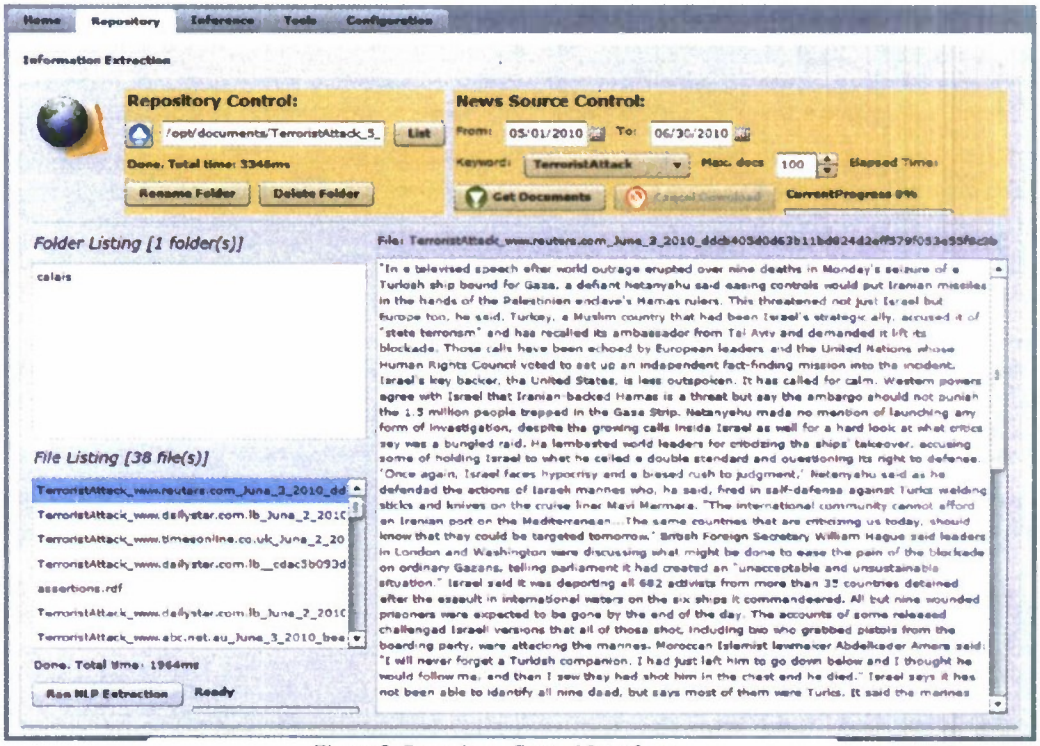


Figure 2. Repository Control Interface

One tool determines whether two or more words appear together as a single term. This tool is based on Google search. For example, to find out whether the 'gateway of India' is a landmark or just an expression, the tool performs a search on Google using the keywords: gateway of India, then examines the first two hundred results to find out how often the words appear in the given sequence.

Another tool is based on Wikipedia search, to find out whether an extracted concept is a place, organization, person, or another type that exists in the ontology. This tool takes advantage of the structure of a Wikipedia page, which provides basic information in a structured format for a large collection of concepts.

The ability for the user to configure certain aspects of the system is an integral part of the user experience in a collaborative intelligent system. The user experience in this application includes the ability to configure some aspects of the system to suit the user needs for a given task. On the NLP extraction side, the user can control the NLP server and stop it and start it at any time during execution. The user can also control the batch jobs that are running in background on the server and stop any job or delete it. In addition, the user can view the extraction rules and edit them to introduce new rules. The internal lists, which are used by the system to determine the specific types of organizations, can also be edited. These lists assist the extraction server in determining terrorist organizations, financial institutions and other types of organizations.

7. Conclusion

The current state of NLP offers some tools to extract information from a collection of documents and recognize basic entity types in that information. In order to perform intelligent analysis on the extracted information, a context has to be provided along with some guidelines for the

analysis. We combine existing NLP tools with an OWL ontology to provide a rich environment for performing intelligent analyses. The ontology provides the context and the structure of knowledge in the given domain with the rules, which govern the classification and membership of objects in concepts in the ontology, and provide basic inferencing mechanisms for creating new knowledge. The existing NLP tools extract information from a collection of documents and identify the basic types for entities in the extracted information. Mapping then takes the extracted basic entities and populates the ontology. The inference mechanism runs on the ontology objects and may create new objects or relationships, thus expanding the ontology. The overall system incorporates these tools as collaborative agents that provide and utilize services through a Web server based on Tomcat and Apache. The user interacts with the system through a Web interface that provides assistance with the management of sources and documents, and offers support with the extraction, analysis, and retrieval of information contained in and derived from the repository.

References

AlchemyAPI. Orchestr8, http://www.alchemyapi.com/, visited 8/10/2010.

Apache Clerezza. Apache Software Foundation – Apache Incubator, http://incubator.apache.org/clerezza/, visited 8/10/2010.

Apache UIMA. Apache Software Foundation. http://uima.apache.org, visited 8/10/2010.

Jem Rainfield. BBC World Cup 2010 dynamic semantic publishing. http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html ,/ visited 7/15/2010.

BigOWLIM. Ontotext AD, http://www.ontotext.com/owlim/big/, visited 8/10/2010.

ClearTK. Center for Computational Language and Education Research (CLEAR), University of Colorado, Boulder. http://code.google.com/p/cleartk/, visited 8/10/2010.

Daya C. Wimalasuriya and Dejing Dou. Ontology-based information extraction: An introduction and a survey of current approaches. Journal of Information Science, June 2010, 36: 306-323.

GATE. University of Sheffield, Computer Science Department. http://gate.ac.uk, visited 8/10/2010.

Jena. Hewlett-Packard Development Company, LP, http://jena.sourceforge.net/, visited 8/10/2010.

LingPipe. Alias-i, http://alias-i.com/lingpipe/, visited 8/10/2010.

Loebner Prize in Artificial Intelligence Web page, http://www.loebner.net/Prizef/loebner-prize.html, visited 8/10/2010.

Michael T. Flynn, Matt Pottinger, Paul D. Batchelor. Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan. Center for New American Security, Voices from the Field, January 2010. http://www.cnas.org/node/3924

Michele Banko, Michael J Cafarella, Stephen Soderland, Matt Broadhead and Oren Etzioni. Open Information Extraction from the Web, IJCAI 2007.

NewsExplorer. Europe Media Monitor, http://emm.newsexplorer.eu/, visited 8/10/2010.

OpenCalais RDF. Thomson Reuters. http://s.opencalais.com/1/pred/, Visited 8/10/2010.

OpenCalais. Thomson Reuters. http://www.opencalais.com, Visited 8/10/2010.

OpenNLP. http://opennlp.sourceforge.net/, Visited 8/10/2010.

OWL Specification. W3C, http://www.w3.org/TR/OWL2-overview/,October 2009.

Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Proceedings of the 2nd International Workshop on Meta-Modelling, WoMM 2006, October 12-13, 2006, Karlsruhe, Germany. LNI 96 GI 2006, ISBN 978-3-88579-190-4.

RDF/XML Syntax specification. http://www.w3.org/TR/REC-rdf-syntax/, February 2004.

Stanford Named Entity Recognizer. The Stanford Natural Language Processing Group, http://nlp.stanford.edu/software/CRF-NER.shtml, Visited 8/10/2010.

Stuart Hendren and John Yesberg. Qualifying Triples and Selective Reasoning - OWL Experiences and Directions, Fifth International Workshop - co-located with ISWC 2008, Karlsruhe, Germany, October 26-27, 2008.

Vrandecic, D., Völker, J., Haase, P., Tran, D., Cimiano, P.: A metamodel for annotations of ontology elements in OWL DL. In Proceedings of the 2nd International Workshop on Meta-Modelling, WoMM 2006, 109-123, Brockmans, S., et al., eds., 2006

Witte, R., Krestel, R., Kappler, T., and Lockemann, P. C. (2010). Converting a historical architecture encyclopedia into a semantic knowledge base. IEEE Intelligent Systems, 25:58–67.

Alternative energy sources and energy efficiency Project Reports

Overcoming barriers to affordable algae biofuels

Prepared by Tryg Lundquist

Investigators:
Tryg Lundquist, Ph.D., P.E., Civil and Environmental Engineering
Yarrow Nelson, Ph.D., Civil and Environmental Engineering
Christopher Kitts, Ph.D., Biological Sciences
Mark Moline, Ph.D., Biological Sciences
Corinne Lehr, Ph.D., Chemistry and Biochemistry

Abstract

With the support of C3RP and others, an interdisciplinary team of researchers at Cal Poly has developed a program in microalgal biotechnology research focused on biofuel feedstock production. Faculty with expertise in phycology, bacteriology, organic chemistry, and chemical and environmental engineering have come together to work on the technical difficulties of microalgal feedstock production.

Algal biofuel is important because it could contribute to the energy mix of the US military and decrease the overall greenhouse gas footprint of U.S. Department of Defense (DOD) activities. Although algae produce oil at a much higher rate than conventional crops, the cost of algae production is high due to the need for large ponds and the cost of algae harvesting. The economics of algal biofuel production could be increased best by (1) improving algae production per unit area, (2) increasing the oil content of the algae, and (3) perfecting algae harvesting methods.

Previous C3RP support was used to develop four pilot-scale algal "high rate ponds" that were operated at the San Luis Obispo wastewater treatment plant, with wastewater providing the needed water and nutrients for algae growth. High rate ponds are shallow with channelized mixing. An additional four ponds were added in the summer of 2009 with external funds. All eight ponds were used in C3RP research to measure algal lipid production rates and investigate settling of algae using a natural bioflocculation technique for harvesting algae. Overall, this research is expected to increase our understanding of the biological processes affecting lipid production by algae which in turn will lead to improved economical operation of high-rate algal ponds for biofuel production.

The specific results of the studies conducted under ONR7 funding are divided between continuous-flow pilot algae production pond experiments and lipid extraction and characterization method development, with a batch culturing experiment on lipid productivity.

In the continuous-flow experiments conducted during July-December 2009, the explanatory variables studied were hydraulic residence time, biomass re-inoculation, and the presence or absence of carbon dioxide supplementation. Algae productivities were typical for high rate ponds ranging from 15-19 g/m^2-d during summer and early fall and 5-10 g/m^2-d during winter. In terms of treatment, total ammonia nitrogen removal was typically >90% and dissolved organic matter removal ranged from 80-90% (as soluble carbonaceous biochemical oxygen demand). These are both good results and would be acceptable for many wastewater treatment applications. However, total suspended solids in the effluent was higher than desirable with 50-100 mg/L of suspended solids (mostly algal cells) remaining after sedimentation . Prior to sedimentation, the ponds contained 230-260 mg/L of algal suspended solids. In winter, post-sedimentation suspended solids ranged from 30-50 mg/L, with pond water containing 70-100 mg/L. Winter ammonium removal was ranged 45-75%. Carbon dioxide and/or inoculum addition did not have noticeable influences on results. The lack of carbon dioxide effect was due to light rather than carbon being the factor limiting growth. Future experiments will recycle settling unit effluent back to the high rate ponds. This recycling will eliminate the light limitation and cause carbon to become the limiting substance, at which point carbon dioxide addition

1 Project significance

Energy alternatives and fuel efficiency are important for DOD missions because these technologies are needed to reduce the military's reliance on petroleum fuels. Development of microalgal biofuels is a top priority of the Defense Advanced Research Projects Agency (DARPA) Biofuels Program because algal biomass can be produced without competition with food crops, and the areal productivity ($g/m^2/d$) of microalgae is among the highest of the photosynthetic organisms (DARPA, 2007; AFOSR, 2008). In addition, algae cultivation can mitigate greenhouse gas emissions by providing carbon-neutral renewable fuels.

Beyond the biofuel/greenhouse (GhG) benefit, the development of a low-cost nutrient removal technology would benefit DOD. Wastewater treatment is one of the vital services supplied by the US military at its bases and in its civil affairs activities. In fiscal year 2003, the US military operated 700 overseas bases and another 6,000 within the US and its territories (DOD, 2003). The Army alone operates over 100 small wastewater treatment plants on bases (Scholze and Zaghloul, 2001). If similar to civilian communities, a substantial proportion of the public works budget of a military base is expended on the wastewater management enterprise.

US civilian needs are also great. The American Society of Civil Engineers estimates that over the next 20 years communities in the US will need to invest $390 billion to renovate existing wastewater treatment facilities and to build an additional 2,000 treatment facilities (ASCE, 2005). This level of investment has been called an Aunprecedented financial problem@ for utilities (WIN, 2000).

2 Background and Prior Work

Biofuels that can be derived from algae include methane and liquid fuels (diesel, jet fuel, ethanol). Methane can be produced by fairly simple methods from algal biomass, but the more valuable liquid fuels require extensive processing and energy input. Liquid fuel is the focus for DOD and this C3RP project. We continue to focus on the production of algal vegetable oil, which is a precursor to biodiesel and JP-8 jet fuel. Our continued goal is to develop scalable culturing methods that will increase algal biomass productivity and to improve both the quality and quantity of oil in the algal biomass.

Two significant barriers to economical microalgal liquid fuel production are: (1) the cost of algae production, which is high due to the construction costs of algae ponds and associated infrastructure, and (2) the cost of algae harvesting, which is high when chemical flocculants are used to promote settling of the small (1-30 μm) algal cells (Downing et al., 2002). The goal of both our current and proposed research is to address both of these issues as described below.

Two types of systems have been considered for algae production: High-rate algal ponds (HRAP) which are paddle wheel-mixed and channelized into a raceway loop, and closed photobioreactors with pure cultures of algae contained in clear tubing. Ponds of the

"high-rate" type are considered the optimal reactors for algae production in the biofuels context because of their lower cost (Weissman et al., 1988), although some researchers are promoting closed photobioreactors (Chisti, 2001). Our approach to economical biofuel production is to grow algae in conjunction with wastewater treatment (Figure 1). Since pure-culture photobioreactors would be impractical for wastewater treatment, they are not considered in our work for scale-up of biofuels production, although they may be useful for experimentation and inoculum production. Given that new reactor designs are unlikely to decrease construction costs significantly, the ways to decrease algae production costs are to increase productivity and/or develop revenue streams from co-products or co-services such as wastewater treatment.

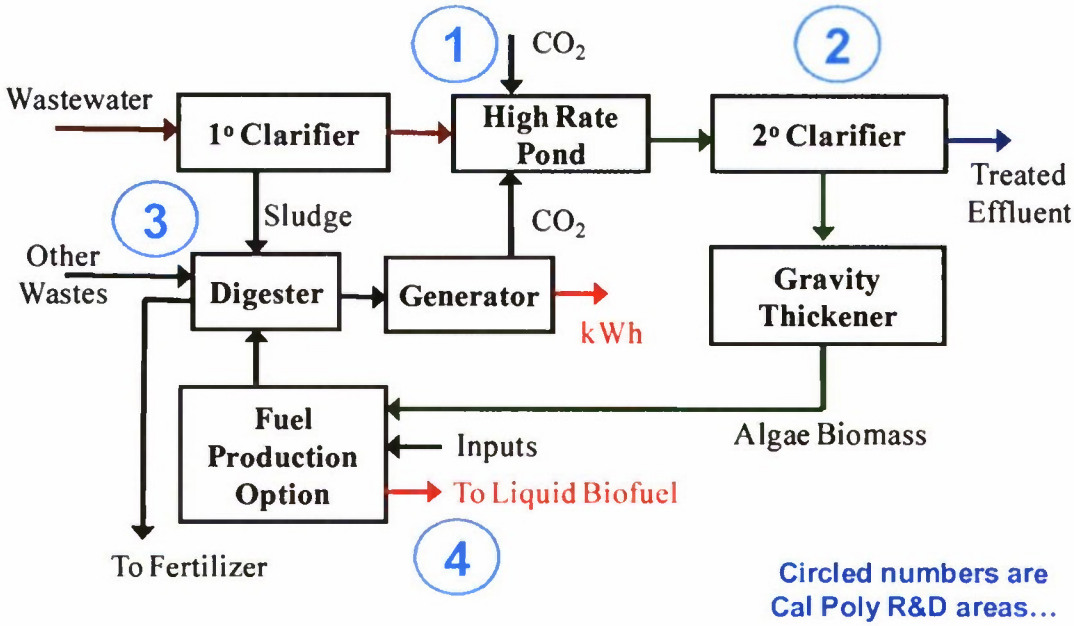


Figure 1. Integration of algal biomass production with wastewater treatment. Process steps 1, 2, and 4 were studied in C3RP-funded research, and step 3, anaerobic digestion of algae, was studied under a California Energy Commission grant during 2009-2010 (not reported herein).

Algae harvesting costs may be lowered directly by improving the settling properties of the algal biomass. Efficient settling of algae is also important for effective wastewater treatment since the algae must be settled out in a secondary clarifier to produce effluent water quality that meets treatment standards. For biofuel production, sedimentation is less expensive than other solids separation methods such as filtration, centrifuging, or dissolved air flotation. A strategy our team has pursued with previous C3RP support is improving settling through the production of natural flocculants intrinsic to the biomass. This previous research has succeeded in increasing natural flocculation, resulting in excellent settling rates, but at times settling was poor. Successful operation of HRAP systems will require more consistent settling than we achieved previously. In the present

research, we sought to identify the causes of the observed incidences of poor settling. We have considered the factors of (1) environmental conditions such as pH, temperature, nutrient concentrations and insolation, and (2) species composition of algae and other microorganisms in the ponds.

Improved algal oil productivity is a project goal. The most useful algal oils for fuel are the triacylglycerides (TAGs), as opposed to the membrane phospholipids. At a finer level, the carbon chain length of TAGs affects the value of the oil for fuel production. JP-8 fuel, especially, has quality control criteria related to temperature and viscosity that are best met by only some oils. The quantification and characterization of algal lipids is in itself a challenging endeavor that has been taken up by Dr. Lehr and Dr. Nelson with previous C3RP funding.

With advanced quantification/characterization capabilities in hand, we are able to study cultivation techniques to reliably increase algal lipid content from its typical 10-15% to the 20-30% levels measured in some of our earlier work (Woertz et al., 2009). The upper limit on lipid content seems to be 40-75%, as represented by one species of algae in particular, *Botryococcus braunii*. Despite its high lipid content, its lipid productivity has been quite low (NREL, 1998). Growth techniques with the potential to increase lipid productivity by chlorophytes or diatoms in general are semi-batch growth. Maximum lipid productivity would be expected to occur during the late exponential phase of growth (Woertz et al., 2009) or during the stationary phase of growth if nitrogen deprivation occurs (an interpretation of results from NREL, 1998).

Algae production costs can be lowered through co-services and co-products, wastewater reclamation and fertilizer production are the valuable near-term sources of revenue. GhG abatement credits could provide further revenue. The overall algal biofuel process envisioned would produce algae using wastewater nutrients, which are essentially free, while treating the wastewater, which is a valuable service. During the fuel production steps, organic carbon is removed from the algal biomass. The residual mass would be decreased while its nutrient content increased. The result would be the creation of a concentrated fertilizer solution. Such recycling of nutrients would offset use of fossil fuel-derived fertilizer. GhG abatement credits could accrue from offsetting fossil fuel use in three areas: (1) energy saving during wastewater treatment compared to conventional wastewater treatment, (2) biofuel production, and (3) fertilizer production.

Wastewater treatment, especially for nutrient removal, is a costly endeavor when done with conventional mechanical treatment technologies such as the activated sludge process. The total cost for typical modern activated sludge processes is $40-$60 million for a 5 million gallon per day (MGD) treatment plant, excluding primary treatment and sludge handling (Carollo Engineers, 2006). Our highest total cost estimate for algae-based wastewater nutrient removal is half as much (Lundquist et al., 2009).

While the nationwide potential for integrated algae-wastewater treatment facilities may be perhaps <2% of total sewage flows (TNO, 2006), the proposed process appears to be one of the few economical and GhG-lowering biofuel approaches. It is the low-hanging fruit among algal biofuel technologies, and its success would justify continued R&D in the field possibly leading to more widespread production of algal biofuels. In additional to municipal wastewaters, animal manure wastewater and agricultural field drainage can

be treated with algal technology (Woertz et al., 2009; Lundquist et al., 2004), which expands the potential market and positive biofuel/GhG impacts of the technology being developed.

3 Objectives

The long-term goal of Algae Tech Group's work is to develop an algae-based process that both produces oil for biofuel and treats wastewater. At full-scale such a process is expected to be economical due revenue gained from providing the wastewater treatment service. Developing key capabilities of the envisioned system is the goal of the present research. These capabilities are the following:

(1) High areal lipid productivity (i.e., >4.7 g oil per m^2 pond per day or 2000 gal/ac-yr).

(2) Low-cost biomass harvesting by bioflocculation and settling (i.e., >90% harvesting efficiency and <30 mg/L residual suspended solids).

(3) Low residual organic matter and nitrogen (i.e., <30 mg/L BOD_5 and <10 mg/L total N) to meet wastewater discharge limits.

(4) Lipid extraction techniques that use less hazardous chemicals than standard techniques.

Development of each of these capabilities was approached various ways in the C3RP research. For the first goal of high lipid productivity, work on manipulation of culture conditions (e.g., low N stress in the culture medium) in laboratory studies to achieve high lipid content in algal cells was begun, and a variety of conditions were implemented in the outdoor pilot units with monitoring of lipid productivity. For the second goal of bioflocculation, all the work was done using the pilot ponds, with monitoring during both controlled manipulation and natural variations in conditions (e.g., weather, influent wastewater concentration). The third goal, good treatment performance, was also pursued exclusively at pilot scale with different hydraulic residence times and CO_2 addition being tested. Lastly, several solvent systems for oil extraction from algal biomass were tested as alternatives to the standard hexane-chloroform system, in which chloroform is a particular health hazard. A particularly important result was the discovery that the costly and energy-intensive cell disruption step may not be needed if certain solvent systems are used.

The descriptions of the experimental methods and results are presented in two main sections. The first covers the continuous-flow high rate pond studies on algae production, bioflocculation, and wastewater treatment. The second main section covers lipid extraction and characterization and an outdoor batch-growth experiment on lipid production. These two sections are derived from the master's theses of Cal Poly Environmental Engineering students Paul Ward and Matt Hutton, respectively.

4 Continuous-Flow High Rate Pond Study Materials and Methods

Experiments were run in four identical pilot-scale HRAP reactors. The reactors were hand-built in the Spring of 2009 and transported to the San Luis Obispo Water Reclamation Facility (SLOWRF), located on Prado Rd. in San Luis Obispo. The ponds were fully operational by the end of July, 2009.

The ponds were constructed to analyze the effects of experimental variables on algal growth, nutrient removal, and settleability. The site was already occupied by four high-rate ponds that were used to conduct experiments in past years.

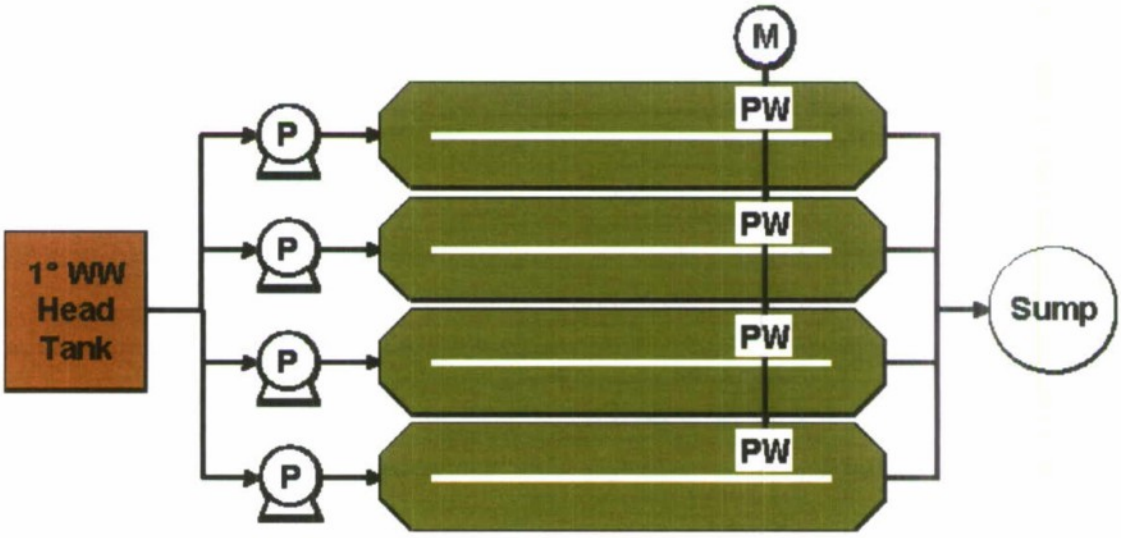


Figure 2: Simplified pilot-plant flowchart. Note: P = pump, PW = paddle wheel, M = motor, 1° = primary

4.1 Construction and Layout

The gathering of materials and initial construction began in the spring of 2009 on the second floor balcony of Building 13 at the California Polytechnic State University in San Luis Obispo (Cal Poly SLO). The fabrication of pond frames and paddle wheels was completed prior to moving to the SLOWRF in order to have a more convenient workspace.

4.1.1 Pond Installation

The pond frames were constructed using several pieces of 8-foot long 2" x 8" lumber (2x8), Simpson® metal straps, and heavy-duty hinges. The frames needed to allow a resting water height of 0.3 meters, while also accounting for fluctuation due to mixing intensity. This was achieved by stacking two 2x8s and fastening them together with metal straps, resulting in an approximate height of 0.37 meters. Each pond was constructed using eight 2x8 stacks: two (2) 8-foot lengths, four (4) 18-inch lengths, and two (2) 20-inch lengths. The HRAP frame design is outlined in Figure 3 and Figure 44.

Before attaching the eight stacked 2x8 components, each section was primed and painted to ensure the integrity of the frame through adverse weather. Once painted, the sections were transported to the SLOWRF. At the project site, the components were combined. Each section connection consisted of two hinges, one on each of the stacked 2x8s.

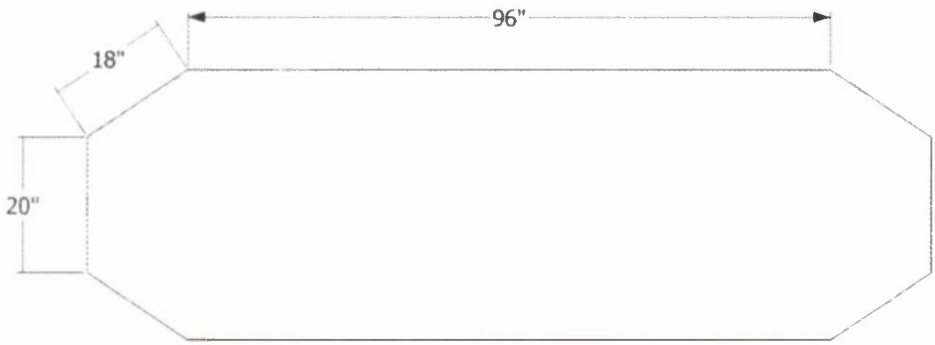


Figure 3: Plan view of symmetrical HRAP frame design. Note: The actual dimensions of the ponds are slightly larger due to the spacing created by connecting hinges.

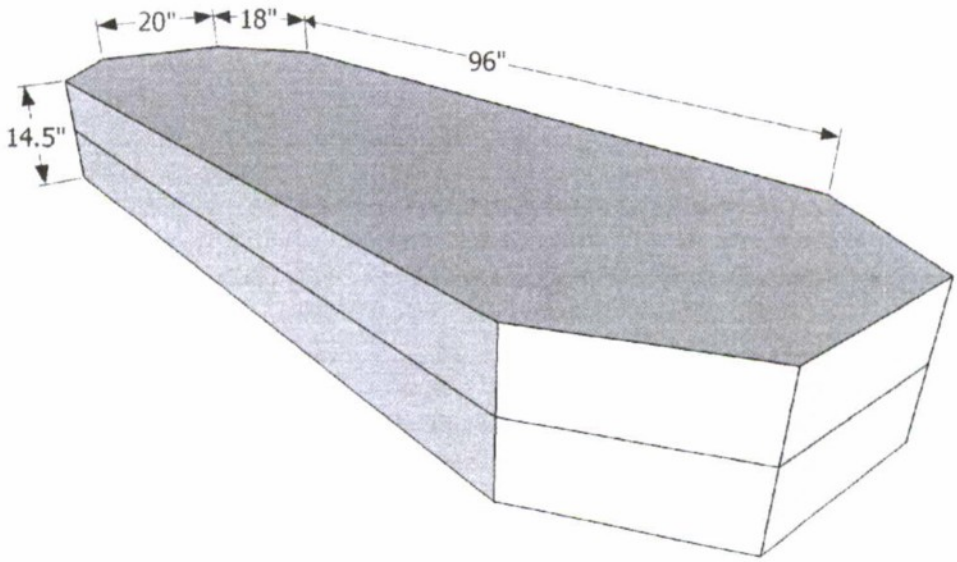


Figure 4: HRAP frame design and dimensions.

Prior to reactor placement, the site had to be graded to create a uniform surface that promoted drainage towards the planned sump area. Using surveying equipment, the site's initial gradation was measured and noted. A plant operator then used a front loader to scrape and flatten the site. The planned locations of the pump house and ponds were built up manually using shovels. An overhead view of the site can be seen in Figure 5.

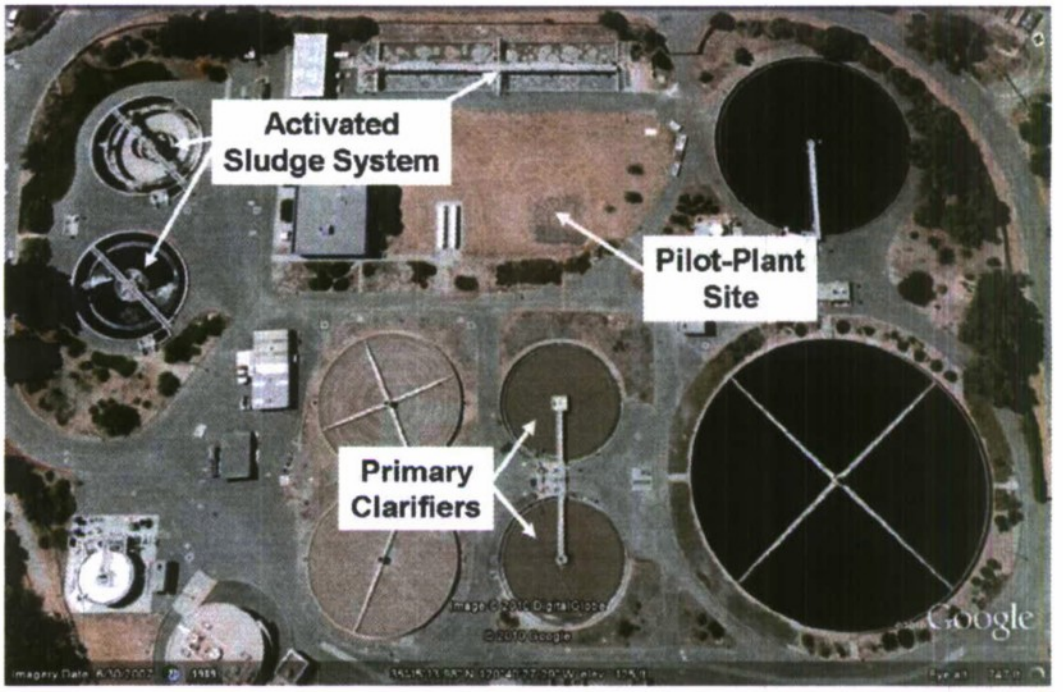


Figure 5: Plan view of the SLOWRF, including the operations relative to the HRAP pilot-scale project. Google Earth captured the image before any construction had taken place in the area.

Local regulation required the experimental ponds be placed in secondary containment to prevent any sewage from spilling onto the ground. A large roll of pond liner was used for both the secondary containment area and each HRAP. Several 2" x 12" wooden planks were used to create a wall to contain the pond liner. The planks were held up by positioning rebar on both sides and driving them into the ground. Prior to placing the pond liner into the containment area, a hole was dug in the shape of a large, circular trash can to make space for the sump. A trash can was cut in half and placed in the hole. The pond liner was then laid over the containment area.

The HRAP wooden frames were then placed on the leveled portion of the containment area. The pond liner was cut into four identical sections and laid into the frames. The frames were then stabilized (to prevent hinge movement) with shelf standard track with C-shaped cross section. These tracks also allowed the baffle to be installed by fastening hinges onto the precut holes.

4.1.2 Paddle Wheels

The baffle and paddle wheels were both constructed using quarter-inch thick high-density polyethylene (HDPE). The baffles were roughly eight feet long and 14.5 inches tall and were attached to the shelf standards. Each paddle consisted of two circular pieces and five cross-pieces. The five cross-pieces were connected to both circular pieces using elbow brackets. Holes were drilled into the large circular sections to insert the drive shaft.

The paddle wheels were positioned at the end of the straight sections of the ponds. Each pond had its own drive shaft, but all were connected using shaft couplers. The paddle wheels were installed in series to a single Leeson® waterproof motor, which was powered by a variable speed unit. The motor was bolted onto a custom-made wooden table that was the proper height to ensure the motor and drive shafts were level. Operating with a single motor ensured uniform channel velocity in all ponds. Figure 6 details the final placement of the paddle wheels and motor assembly.

Figure 6: Post-construction photograph of ponds and paddle wheel motor assembly.

In order to achieve uniform shading, the ponds were spaced equally apart from one another and a mock paddle wheel was installed next to the last pond to simulate the shading every other pond received from the functional paddle wheels. The mock paddle wheel was constructed with a single HDPE circular piece attached to a wooden support (see Figure 7 below).

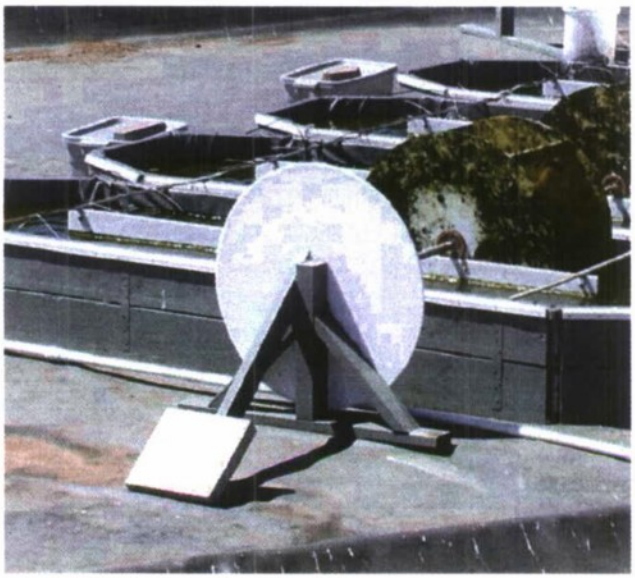


Figure 7: Mock paddle wheel designed to simulate shading of functional paddle wheel.

4.1.3 Pond Infrastructure

Important considerations for plant design were the pond standpipe configuration and the systems for supplying wastewater and carbon dioxide. The supply systems needed to be able to provide uniform flows to each receiving pond.

Standpipes were installed in each pond to maintain a depth of approximately 0.3 meters (about 1 foot). The standpipes were connected to a single effluent PVC line which emptied into the sump for disposal back into the municipal wastewater system.

The wastewater supply system (see Figure 8) included a head tank, a Flotec® 1/6 horsepower submersible utility pump, Cole-Parmer peristaltic pumps, and various supply lines. The head tank was designed to receive water from the primary clarifier by means of the submersible utility pump. An overflow pipe maintained a constant water level in the tank, while also preventing stagnation. An outlet line was installed midway up the tank to draw wastewater for system delivery without pulling solids which accumulate at the surface and bottom. The outlet line led to an eight-valve manifold and a single-valve discharge line which fed directly into the sump (for cleaning and line flushing). The supply system was used for both this project and the adjacent deep HRAP project.

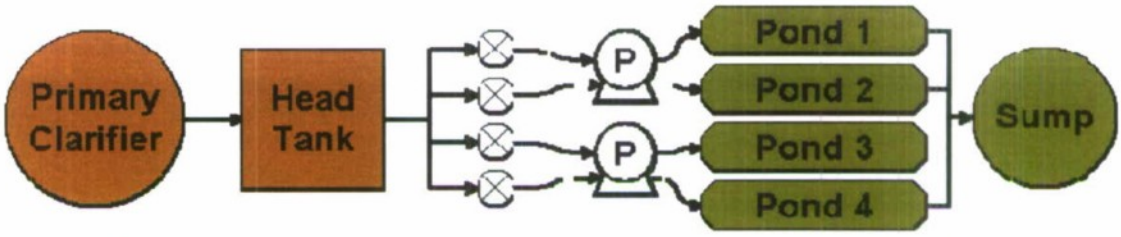


Figure 8: Wastewater supply system flowchart. Note: Crossed circles = valves, P = pump

Each of the four ponds (Pond 1, 2, 3, and 4) received wastewater through an individual valve. The supply lines were run through two Cole-Parmer peristaltic pumps, each with a

dual-channel pump head which allowed two ponds to be supplied using one pump motor. The feed lines were placed in the shelf standard channels with the outlet to the ponds placed directly after the standpipe to prevent short-circuiting of wastewater.

The carbon dioxide supply system employed a pH controller with a solenoid valve to supply gas once the pH reached a custom setpoint. Each controller and valve unit was kept in a sealed Tupperware® container to eliminate exposure to moisture. The CO_2 tanks were stored in an onsite shed and fitted with a flow regulator. A soft supply line ran from the shed to the containment area, where a splitter led to each controller unit. A diffuser strip from a local pet store was placed in each of two experimental ponds. To keep the diffuser units form moving, a small portion of PVC pipe was filled with pennies, capped, and attached to the strips.

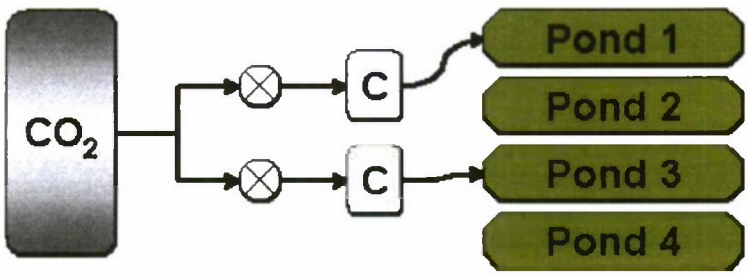


Figure 9: Carbon dioxide supply system flowchart. Note: Crossed circles = valves, C = pH controller.

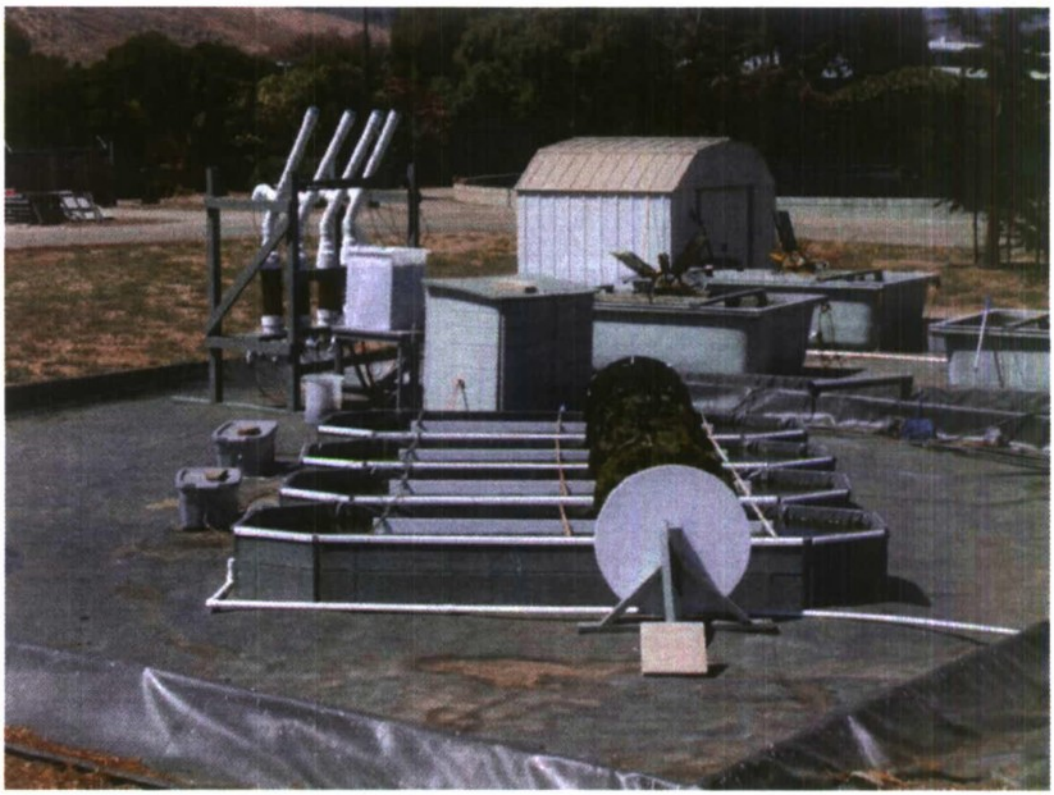


Figure 10: Completed, operational pilot-scale HRAP plant.

4.2 Pilot Plant Operation & Maintenance

Several operational parameters were developed prior to beginning experimentation. The channel velocity, hydraulic residence time and pH set point for CO_2 addition were several key decisions. The project was also split into several experiments, each of which needed to be planned. A maintenance routine also needed to be developed in order to ensure optimal conditions at the pilot-plant.

4.2.1 Channel Velocity

A channel velocity of 25 cm/s was chosen from experience. This velocity was to keep the algal cells in suspension and allow for optimal exposure to sunlight. In order to achieve this velocity, the paddle wheel motor was adjusted and the channel velocity measured using the "orange test". An orange was placed into one of the ponds and a stopwatch was used to record its travel time over a known distance. A channel velocity of 25 cm/s corresponded with an approximate 4 rpm paddle wheel speed.

4.2.2 Hydraulic Residence Time

A hydraulic residence time (HRT) of 4 days was chosen for operation. This HRT corresponded to a wastewater influent flow of 160 mL/min.

4.2.3 pH Set Point for CO_2 Addition

Algal ponds commonly experience pH levels upwards of 9.5-10. A pH of 8.3 is optimal for ponds to stimulate aerobic bacterial degradation and avoid ammonia gas toxicity

(Scraggs, 2007). CO_2 addition also helps overcome any carbon deficiencies in the wastewater feed. During the CO_2 experiments, a pH set point of 8.3 was programmed into the solenoid valve controllers. Once the pH of the ponds reached a level several points above 8.3, the solenoid valve opened and pure CO_2 gas bubbled into the ponds.

4.2.4 Pilot-plant Start-Up

The pilot-plant became operational on July 15th, 2009, when the ponds were filled with 10 gallons of inoculum and fed primary wastewater. The inoculum was obtained from neighboring experimental HRAPs being run by Michael Podevin. The ponds were initially partially filled, but over several days were up to capacity and overflowing into the effluent standpipes. During this period, the paddle wheels were operating at 4 rpm. Once the ponds were full and green, the pH controllers were calibrated and carbon dioxide was injected into Ponds 1 and 3 when pH levels reached 8.3.

4.2.5 Monitoring and Maintenance

Daily monitoring was an essential part of experimentation, as it ensured that the ponds were running according to the experimental plan. Ideally, plant operations should be continuously monitored. With limited resources and manpower, efforts were made to have a student visit the plant once a day to conduct routine monitoring and undertake maintenance work as needed. All important observations and monitoring data were recorded in a field log.

The typical monitoring routine consisted of several tasks: (1) checking and adjusting wastewater influent flows, (2) calibrating pH controllers, (3) monitoring pond pH, temperature and dissolved oxygen (DO) levels, (4) monitoring and removing filamentous growth on the pond sidewalls and standpipe, (5) recording general pond conditions and presence of zooplankton, (6) checking water level in the headtank, and (7) checking all pumps and electrical equipment.

Wastewater influent flows were checked with a 250 mL graduated cylinder and stopwatch. The influent wastewater was collected in the graduated cylinder over a period of 30 seconds, and the volumetric flow rate was recorded in mL/min. If the flows had strayed substantially from the desired 160 mL/min, the peristaltic pumps were adjusted. In the event that flows became consistently erratic, the supply lines were removed and replaced because of fouling (the buildup of slime in the tubing). The ponds were operated with two peristaltic pumps with dual-channel pump heads, so two ponds received wastewater with the same pump (Ponds 1 & 2 in one grouping, Ponds 3 & 4 in the other; see Figure 8). In this configuration, fouling in one line could be easily identified by observing large differences in flow between ponds being supplied by the same pump. The pump heads also required special MasterFlex® tubing, which needed to be replaced along with the supply line to ensure proper consistency.

The pH controllers were calibrated using a two-point calibration. The two solutions must span the expected operational range. The typical pH range in algal ponds is 7-10, so the controllers were calibrated using pH solutions of 7.0 and 10.0. The probes were cleaned by rinsing with tap water and scrubbing lightly with a soft, clean rag. This procedure was repeated every few days, or when there was visible algal buildup on the probe.

The temperature and pH of each pond was measured using a hand probe, which was calibrated daily using a three-point calibration consisting of pH solutions of 4.0, 7.0, and 10.0. Temperature and pH were recorded every day that a full plant monitoring routine took place. Dissolved oxygen readings were taken roughly every week when the probe was functioning properly.

Especially in warmer months, filamentous growth (algal and bacterial) on the pond sidewalls and standpipe became an issue. The experiment was concerned with the nutrient removal performance of free-floating algae, so filamentous strains needed to be controlled. The filamentous growth also presented a clogging hazard to the standpipes. When filamentous growth was observed, it was noted and removed using a hard-bristle hand brush. Compared with the neighboring deep HRAPs, filamentous growth was not as much of a nuisance.

During routine site monitoring, the color of the ponds and presence of zooplankton was noted. Color changes indicated either an increase in flocculation or a decrease in algal productivity. The presence and color of flocs was also noted. Using beakers, the presence of zooplankton could also be determined. A sample from each pond would be drawn and all samples would be set side-by-side to compare color and zooplankton presence.

A final critical step in daily monitoring was the check on pumps, the head tank and all other electrical devices. In event that the head tank was low, the submersible pump in the primary clarifier was checked first for power. If the supply line was vibrating, the pump was on and spinning, meaning the pump screen was clogged with debris. To clean the screen, the pump was lifted from the clarifier and scrubbed with a long-handled brush.

4.3 Experimental Procedures

This section outlines the dates and operational parameters of each experiment. During 2009, several experiments were run on the ponds. There were two main experimental setups that were used on several occasions, and one setup which failed to sustain a culture. The first setup involved the supplementation of CO_2 to the experimental ponds to observe changes in nutrient removal, algal growth, and settling. The second setup involved inoculating the experimental ponds to observe the same parameters. The third setup, which never produced a working system, was the dilution of the experimental ponds with tap water to produce a more carbon-starved environment with which to test the effects of CO_2 supplementation. Two other experiments were conducted after the performance period of the current and are not reported herein.

4.3.1 Experiment I

The first experiment conducted on the newly constructed HRAPs began on July 21st, 2009. The ponds had been filled and inoculated one week earlier and were green. Ponds 1 and 3 were supplemented with CO_2 when the pH rose above 8.3. During the first several weeks of operation (until August 13th, 2009), the ponds were tested only for total suspended and total volatile solids (TSS and VSS, respectively) to observe the relative growth rates of the experimental ponds and control ponds. The hydraulic residence time during this experiment was 4 days (feed rate of 160 mL/min) and the channel velocity

was approximately 25 cm/sec per the "orange test." As with all experiments, the ponds were fed wastewater from the SLOWRF primary clarifiers.

On August 13th, samples began being pulled for more extensive analysis. Nutrient removal efficiency and settleability was evaluated in the lab using samples from the ponds and head tank. The first experiment concluded on September 28th, 2009.

4.3.2 Experiment II

The second experiment, which began October 3rd, 2009, aimed to reduce the carbon load in the ponds, thereby increasing the effect of CO_2 supplementation. The HRT remained at 4 days and the channel velocity was maintained at 25 cm/sec. The wastewater feed rate was decreased to 80 mL/min and tap water lines were run to the ponds. The tap water feed was set to 80 mL/min to achieve the correct HRT. The dilution with tap water cut the influent carbon load in half.

After two weeks, including a rain event, the ponds turned light tan and had no algal presence. The dilution experiment was abandoned on October 20th.

4.3.3 Experiment III

The third experiment began on October 22nd, 2009, and involved adding inoculum to the experimental ponds to study its effect on settling and nutrient removal. The ponds were intermixed and allowed to become green before beginning the inoculum additions. On November 12th, ponds 1 and 2 began receiving one gallon of inoculum per day. The weekly testing procedure of past experiments was continued. The experiment was suspended on December 20th due to the Christmas holiday break.

4.4 Water Quality Sampling and Analysis

Water quality analysis was conducted weekly, barring any severe weather events, power outages, or pond shutdowns. In addition to weekly analysis, routine field measurements were taken to daily to monitor for significant changes in pond conditions. Each experiment incorporated similar data, but each had minor differences. The sampling schedules for each experiment are outlined below.

Table 1: Experiment I Sampling Schedule

Parameter	Sample Location	Measurement Frequency
TSS/VSS	INF, HRAPs	Semiweekly*
2-hour Settled TSS	HRAPs	Weekly
24-hour Settled TSS	HRAPs	Weekly
Ammonical Nitrogen	INF, HRAPs	Semiweekly*
Nitrate & Nitrite	INF, HRAPs	Weekly**
NPOC	INF, HRAPs	Weekly**
DRP	INF, HRAPs	Weekly***
DOC	INF, HRAPs	Weekly*
Alkalinity	INF, HRAPs	Weekly
Settled TBOD	HRAPs	Weekly
scBOD	INF, HRAPs	Weekly
Cell Counts	HRAPs	Weekly
Influent Flow	HRAPs	Daily****
Pond pH	HRAPs	Daily****
Pond Temperature	HRAPs	Daily****

* Semiweekly TSS/VSS, NHx-N, and DOC testing was initiated September 14th, 2010

** Nitrate & Nitrite and NPOC data could not be used for the first experiment due to QA/QC failure

*** DRP analysis was conducted weekly over a 4-week period

**** Measurements performed on a near-daily basis

Table 2: Experiment III Sampling Schedule

Parameter	Sample Location	Measurement Frequency
TSS/VSS	INF, HRAPs	Semiweekly
2-hour Settled TSS	HRAPs	Weekly
24-hour Settled TSS	HRAPs	Weekly
Ammonical Nitrogen	INF, HRAPs	Semiweekly
Nitrate & Nitrite	INF, HRAPs	Weekly*
NPOC	INF, HRAPs	Weekly**
TIC	INF, HRAPs	Weekly**
DOC	INF, HRAPs	Weekly
Alkalinity	INF, HRAPs	Weekly
Settled TBOD	HRAPs	Occasional
Zooplankton Counts	HRAPs	Weekly
Influent Flow	HRAPs	Daily
Pond pH	HRAPs	Daily
Pond Temperature	HRAPs	Daily

* Reliable nitrate & nitrite data began in January 2010

** NPOC and TIC data could not be used for the third experiment due to QA/QC failure

During each experiment, a day of the week was chosen as the designated main sample day, while another day was chosen as the semiweekly sample day. The main sample day included all the analytical tests and required large sample volumes (approximately 3000 mL). The semiweekly sample day included TSS/VSS and ammonia tests, and required no more than 150 mL sample volumes.

One-gallon milk jugs and two-liter Nalgene® bottles were used to collect samples on the main sample day. If phosphorus testing was planned, the two-liter bottles were acid-washed to remove any residual phosphorus from the bottle walls. On semiweekly sample days, 250 mL sample containers were used. Pond samples were obtained at the center of the HRAPs, directly upstream of the influent supply line to avoid short-circuiting of raw

influent into the sample bottles. A primary influent sample was obtained by placing a sample bottle below the influent supply lines of each pond. Prior to filling any of the sample bottles, they were each rinsed using the water sampled. After sampling, the bottles were immediately rinsed and set upside down to dry until the next sampling period. If growth was present on a sample bottle, it was brushed with a light glass-soap solution until clean.

4.4.1 Total and Volatile Suspended Solids

Total and volatile suspended solids (TSS/VSS) testing was performed according to APHA Method 2540 D. The experiments utilized Fisher Scientific G4 glass fiber filters, with a nominal pore diameter of 1.2 microns. Prior to use, the filters were rinsed with approximately 60 mL of deionized water and ashed in an oven at 550°C to remove any loose glass fibers or dust from packaging. When necessary, filtrate from the TSS/VSS procedure was used for subsequent testing. TSS/VSS testing was conducted on the HRAP and primary influent samples on both the main and semiweekly sampling days. Settled pond supernatant was only tested for TSS.

4.4.2 Total Ammonical Nitrogen

The total ammonia content of each sample was determined using the Ammonia-Selective Electrode Method, APHA Method 4500D. Due to the temperature sensitivity of the electrode, samples were allowed to warm to room temperature prior to conducting the test. The pH of the samples was raised using Orion Ammonia pH-adjusting ISA (#1310-73-2). A calibration curve was prepared prior to each run with a series of dilute standards made from 2500 mg/L stock solution. The calibration curve consisted of five ammonium concentrations: 0.1, 1.0, 10, 25, and 50 mg/L. The ammonia-selective electrode probe and pH probe read concurrently and were rinsed with deionized water between each standard and sample. If the calibration curve did not have an R^2 value of 0.99 or greater, one or more standard was rerun prior to testing any of the experimental samples. During testing, three samples were chosen randomly for spike matrix analysis. Total ammonia was measured for both HRAP and primary influent samples.

4.4.3 Nitrate & Nitrite

Nitrate and nitrite levels were determined using a Dionex® Ion Chromatograph (IC), as referenced in Feffer (2007). The calibration curves created for these experiments utilized stand-alone nitrate and nitrite samples, rather than 7-anion standards, after Experiment III. Nitrate standards of 0.1, 1, 5, 10, and 25 mg/L NO_3-N, and nitrite standards of 0.1, 1, 2, 5, and 10 mg/L NO_2-N were used to create the more recent calibration curves. Much of the IC data from the first and second experiment could not be used because of poor quality control procedure pertaining to the calibration standards. Beginning with experiment III, more rigorous QA/QC methods were put into place.

4.4.4 Total Kjeldahl Nitrogen (TKN)

The APHA standard method was used except that a less toxic copper catalyst was used instead of a mercury catalyst.

4.4.5 Alkalinity

APHA Method 2320B was used to measure alkalinity. This test was conducted immediately after samples were brought into the lab.

4.4.6 Biochemical Oxygen Demand

Total and soluble carbonaceous biochemical oxygen demand (TBOD and scBOD) was determined using APHA Method 5210 B. TBOD data was collected on 24-hour settled supernatant from Imhoff cones and primary influent. scBOD data was collected on the filtrate pond water and primary influent that had passed through 1.2 micron Fisher G4 glass fiber filters. Nutrient buffer solution was prepared using Hach BOD Nutrient Buffer Pillows and deionized water. When testing for scBOD, Hach nitrification inhibitor (Formula 2533™) was used. Standards and blanks were analyzed during each weekly test period.

4.4.7 Settling Tests

Settling tests were conducted weekly using 1-liter Imhoff cones, which were filled with pond sample water. After two hours, supernatant samples were collected from each cone for TSS analysis. After 24 hours, supernatant samples were again collected for TSS and TBOD analysis. Imhoff cones are routinely used as a standard device for determining settleable solids in wastewater laboratories, as referenced in APHA Method 2540 A.

Prior to filling the Imhoff cones, sample bottles were gently shaken to keep all solids suspended. After 2 and 24 hours, supernatant was collected from each cone using a pipette. In order to avoid collecting any floating debris, the pipette tip was submerged approximately one centimeter below the water surface.

4.4.8 Daily Temperature and pH

Daily temperature and pH measurements were obtained using a portable Oakton Acorn® Ion 6 meter. The meter was calibrated daily using 4.0, 7.0, and 10.0 pH solutions. The measurements were made to monitor daily changes in pH and temperature. Relative pH changes between experimental and control ponds were also observed.

4.4.9 Weather Data

Daily weather data was obtained from the California Irrigation Management Information System (CIMIS) through the Department of Water Resources Office of Water Use Efficiency online database. The data was recorded at San Luis Obispo Station No. 52, which is located on the Cal Poly Campus, approximately 3 miles from the pilot plant. The data includes evapotranspiration, precipitation, solar radiation, maximum and minimum air temperature, relative humidity, and wind speed.

5 Continuous-Flow High Rate Pond Study Results and Discussion

This section includes the data analysis and various observations from Experiments I-V. The experimental phase of this thesis project ran from July 21st, 2009 to September 18th, 2010. Hydraulic residence time, pond depth, channel velocity, inoculum and CO_2 additions were all variables controlled throughout the experiments. Primary influent water quality, inoculum characteristics, and weather were all uncontrolled variables.

5.1 Weather Conditions

The main weather conditions that affect algal growth are temperature and solar radiation. The warmer the temperature, the faster algae grow, and the quicker nutrients are assimilated into biomass. The following graphs detail weather condition data over the entire project timeline.

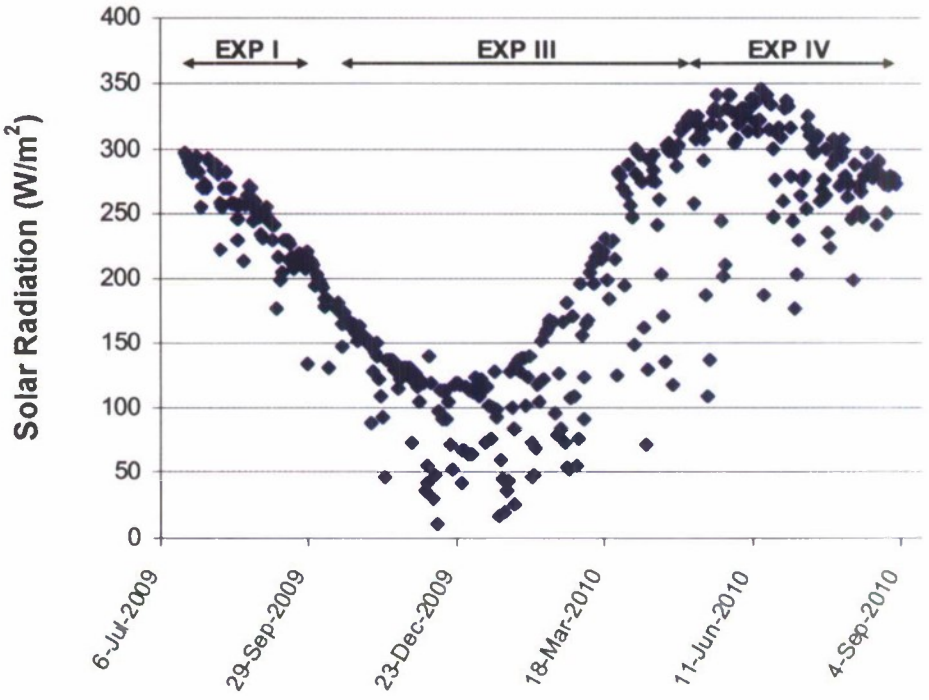


Figure 11: Daily Solar Radiation throughout the Entire Project

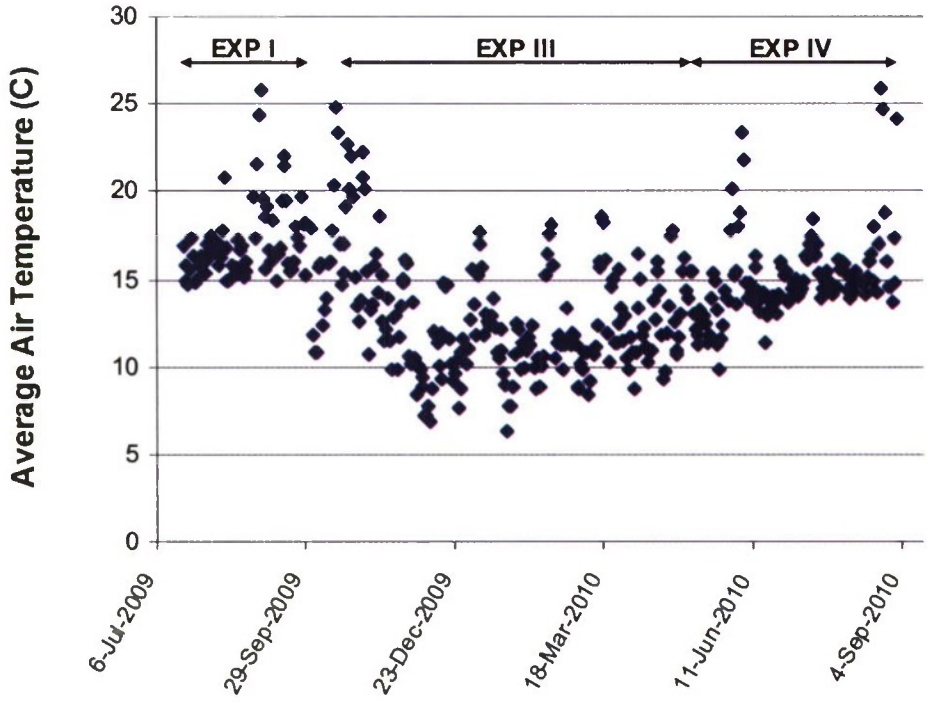


Figure 12: Average Air Temperature throughout the Entire Project

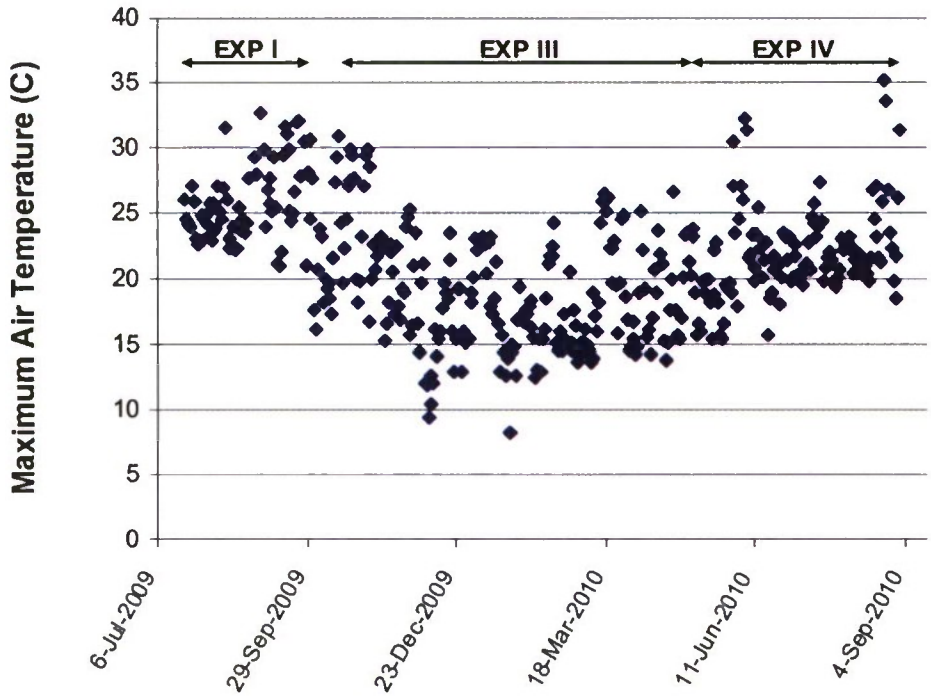


Figure 13: Maximum Air Temperature throughout the Entire Project

The solar radiation data shows a clear trend with the seasons. During the summer months, the levels reach a peak, while during the winter months, the levels bottom out. The curve closely resembles a sinusoidal pattern. The temperature data does not have such drastic differences between winter and summer months, but does have the same trend. During summer months, temperature and solar radiation both peak, while in winter months they both reach a minimum. The average solar radiation and temperature data for each of the successful experiments is listed below in Table 3.

Table 3: Average solar radiation and temperature data for each successful experiment.

Parameter	Experiment I	Experiment III	Experiment IV
Average Radiation, W/m^2	249	157	284
Average Temperature, C	17.1	12.5	15.2
Average Max Temperature, C	25.7	18.7	22.2
Average Min Temperature, C	11.6	7	10.7

5.2 Experiment I

The following section provides water quality results for the first experiment. During this experiment, all ponds were operated at a 4-day hydraulic retention time. Ponds 1 and 3 received supplemental CO_2 through diffuser strips located at the channel bottom of each pond. The solenoid valve controlling the flow of carbon dioxide gas was opened once the pH probe registered a reading over 8.5. All ponds were fed from the same stock of primary wastewater and all ponds had an approximate channel velocity of 25 cm/s.

5.2.1 Treatment Performance

During Experiment I, treatment performance between control ponds and experimental ponds receiving carbon dioxide was analyzed. The constituents of concern included total and volatile suspended solids (TSS/VSS), total ammonical nitrogen (NH_x-N), alkalinity, total biochemical oxygen demand (TBOD), soluble carbonaceous biochemical oxygen demand (scBOD), and dissolved reactive phosphorus (DRP). Settling tests were also conducted and data gathered included 2-hour and 24-hour TSS, and 24-hour TBOD of the supernatant drawn from the top of Imhoff cones. A summary of the average constituent levels in all four HRAPs can be found in Table 4.

Total and volatile suspended solids were the main indicators of algal growth in the HRAPs. Volatile suspended solids represented the rough concentration of active biological material, while total suspended solids included dead cells and other debris. TSS/VSS levels in the experimental ponds were expected to be markedly higher than the control ponds, due to the additional source of carbon. During Experiment I, however, there was no clear pattern of increased growth in the experimental ponds. During the first two weeks there appeared to be a separation between the experimental ponds and control ponds. On August 6th, experimental Ponds 1 and 3 were at 335/305 mg/L (TSS/VSS) and 329/305 mg/L respectively, while control Ponds 2 and 4 were at 253/241 mg/L and 276/257 mg/L. Over the next month and a half, Pond 3 solids levels declined, while

Pond 1 remained the highest in solids by the end of the experimental run. The solids levels in control Ponds 2 and 4 ended in between those of the experimental ponds.

As referenced in Table 4, the average TSS/VSS values for the control and experimental ponds were not markedly different. The average VSS concentration for the control and experimental ponds was 229 and 239 mg/L, respectively. Algal growth may not have been affected by the addition of CO_2 because the feed wastewater supplied enough carbon that sunlight was the limiting factor rather than carbon.

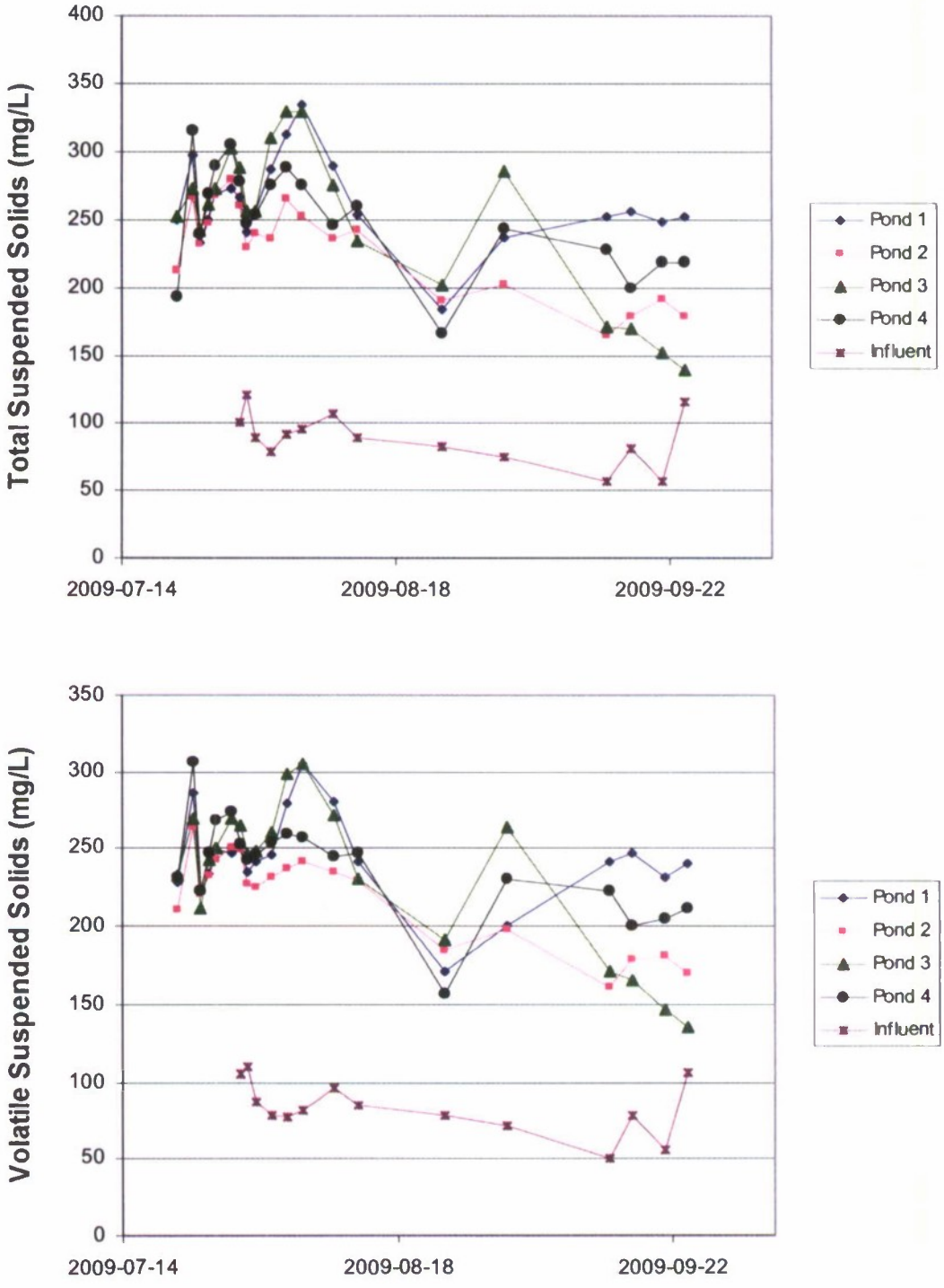


Figure 14: TSS & VSS Data for Experiment 1

During Experiment I, the HRAPs proved to be effective in reducing ammonia levels. The main mechanisms for ammonia removal are assimilation into algal biomass, nitrification by bacteria, and volatilization. Since biological growth was not substantially affected by the addition of carbon dioxide (concluded by analyzing VSS data), differences in ammonia concentration could be attributed to the pH levels of the ponds affecting volatilization. The control and experimental ponds had average total ammonical nitrogen concentrations of 3.0 and 3.6 mg/L, respectively. During the course of the experiment, the pH levels in the control ponds reached pH levels between 9.5 and 9.8 in the afternoon. The experimental ponds were kept at or below a pH level of 8.5 by CO_2 injection.

The overall removal efficiencies for total ammonia nitrogen were approximately 90-92% during Experiment I (Table 4). Carbon dioxide supplementation was expected to increase productivity in the experimental ponds, which would also increase nitrogen uptake and conversion into biomass. This was not the case, however, as the ammonia removal was almost identical in all ponds. Ammonia volatilization may have been a substantial method of removal.

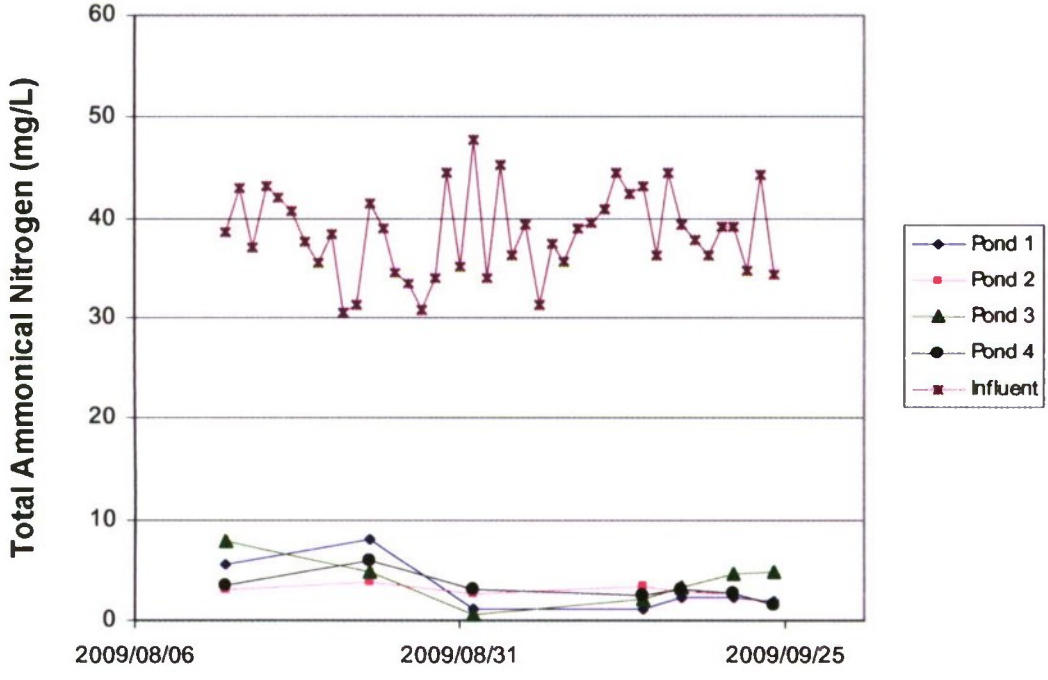


Figure 15: Total ammonical nitrogen levels in all ponds and primary influent during Experiment I.

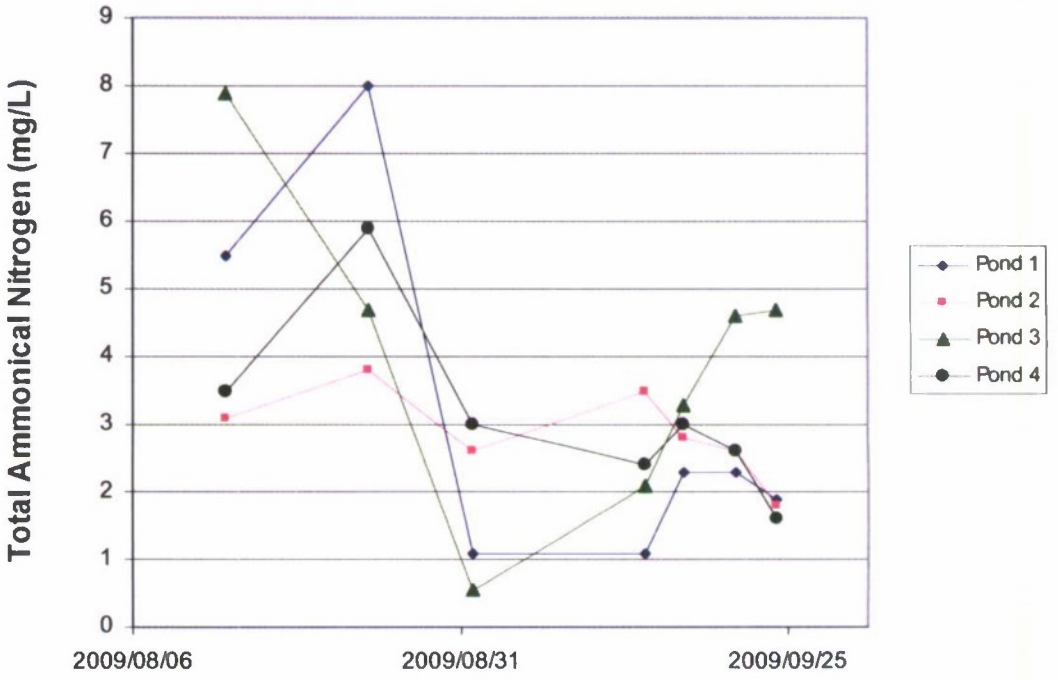


Figure 16: Total ammonical nitrogen in all ponds, excluding primary influent.

Weekly settling tests were conducted using one-liter Imhoff cones. Total suspended solids measurements were taken after two and 24-hour periods. Total BOD tests were also conducted on the 24-hour settled samples. During experiment I, no differences in settleability were expected between the control and experimental ponds. The average settled TSS levels for the experimental ponds were lower than those of the control ponds, but were not consistent with one another. The experimental ponds 1 and 3 averaged 151 and 98 mg/L 2-hr TSS, and 69 and 47 mg/L 24-hour TSS, respectively. The control ponds 2 and 4 averaged 165 and 192 mg/L 2-hr TSS, and 95 and 92 mg/L 24-hr TSS, respectively (see Table 4).

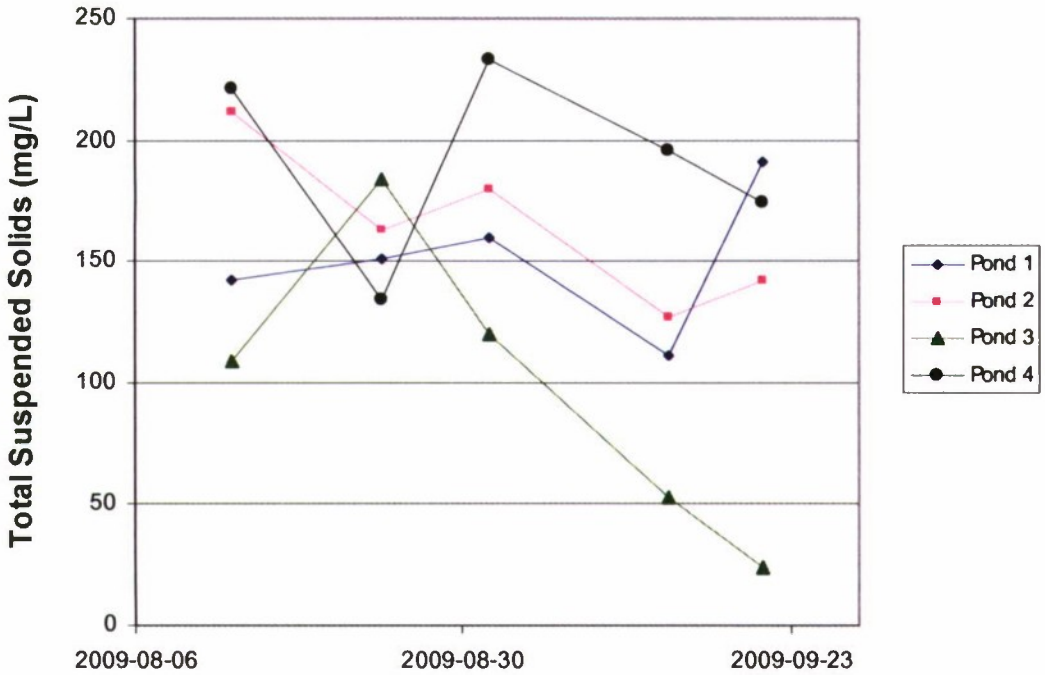


Figure 17: 2-hour TSS data for experiment I.

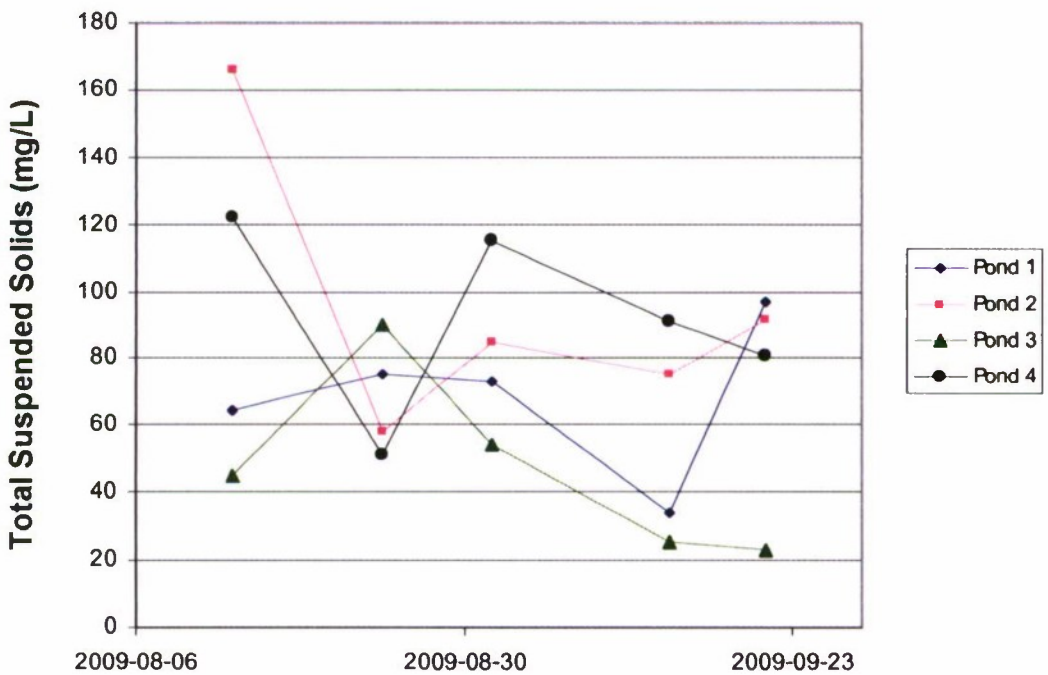


Figure 18: 24-hour TSS data for experiment I.

The weekly settling results do not reveal a clear pattern between the control and experimental ponds. The experimental ponds had lower 24-hr TSS levels three out of five testing days, but did not consistently settle more efficiently than the control ponds.

Biochemical oxygen demand (BOD) measurements were taken on raw primary wastewater and 24-hour settled pond supernatant samples during Experiment I. Soluble carbonaceous BOD (scBOD) was also determined for filtered raw primary and raw pond samples. Neither the experimental or control ponds exhibited better performance in organics removal. Only four reliable BOD results were reported during Experiment I, so data was limited. The experimental ponds had lower scBOD levels for two consecutive weeks before Pond 1 levels rose dramatically.

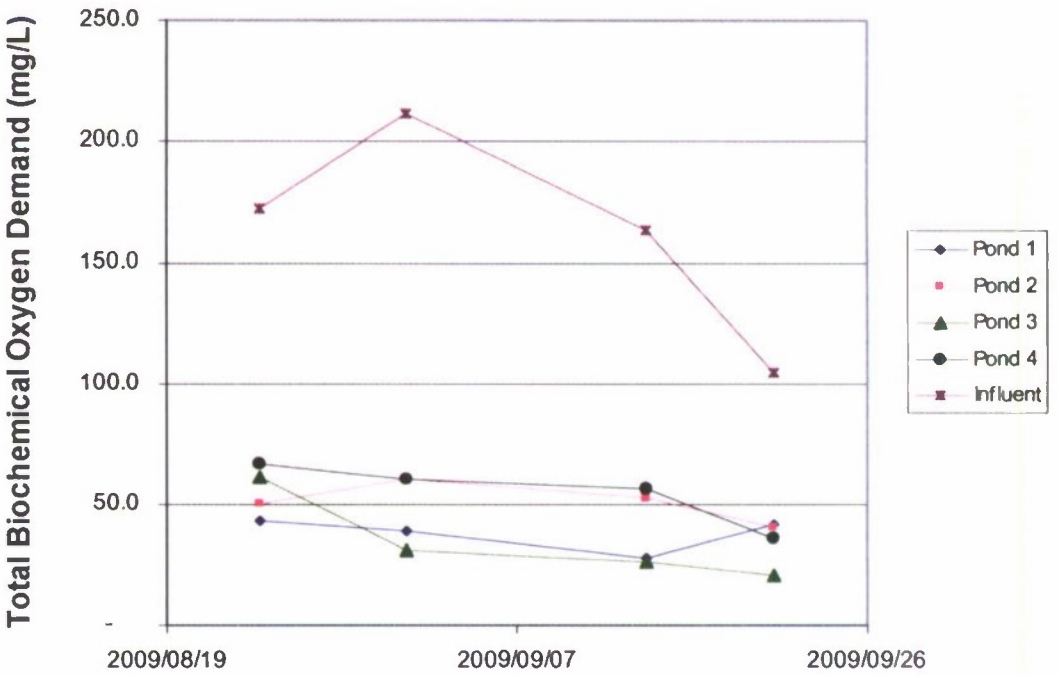


Figure 19: Total Biochemical Oxygen Demand during Experiment I

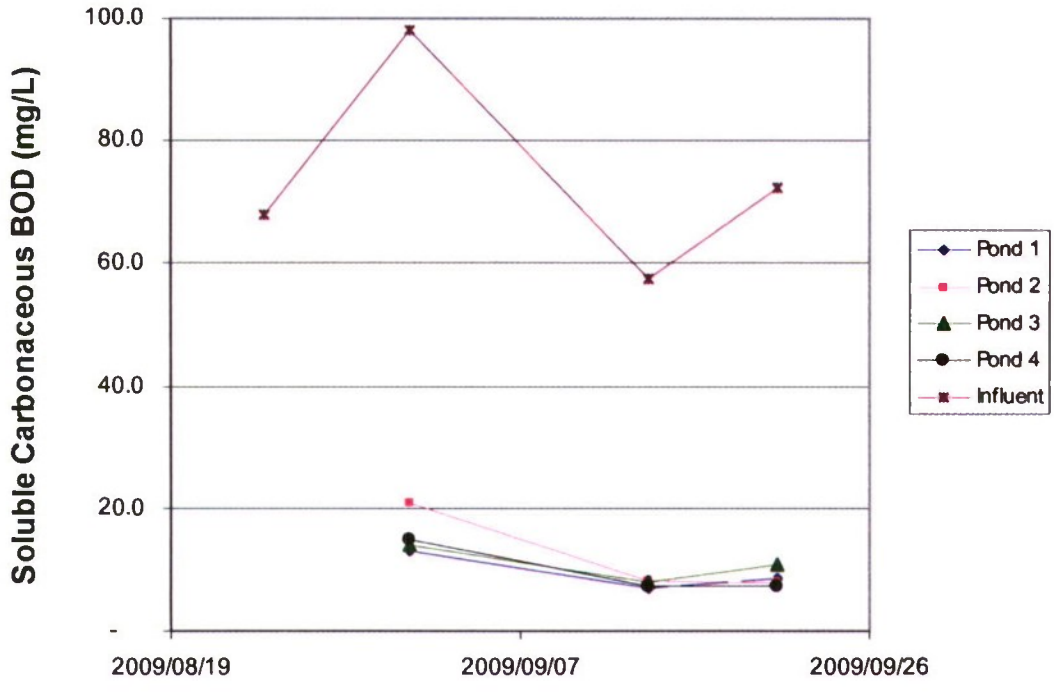


Figure 20: Soluble Carbonaceous BOD during Experiment I

Dissolved reactive phosphorus (DRP) and alkalinity were also measured during Experiment I. DRP removal in the ponds was relatively consistent for all ponds, averaging between 47 and 56 percent removal. Alkalinity removal was also relatively consistent, averaging between 26 and 29 percent removal.

Table 4: Average concentrations of water quality constituents during Experiment I

Constituent	Influent	Pond 1 (w/ CO_2)	Pond 2 (control)	Pond 3 (w/ CO_2)	Pond 4 (control)
TSS (mg/L)	88	262	229	250	251
Std. Error	5.1	7.2	7.6	12.6	8.6
# of samples	14	20	20	20	20
VSS (mg/L)	83	244	219	234	239
Std. Error	4.7	6.6	6.6	10.9	7.0
# of samples	14	20	20	20	20
2-hr TSS (mg/L)	n/a	151	165	98	192
Std. Error	n/a	13.0	14.9	27.8	17.6
# of samples	n/a	5	5	5	5
24-hr TSS (mg/L)	n/a	69	95	47	92
Std. Error	n/a	10.2	18.6	12.2	12.7
# of samples	n/a	5	5	5	5
NHx-N (mg/L)	38	3.2	2.9	4.0	3.1
Std. Error	2.2	1.0	0.2	0.9	0.5
# of samples	70	7	7	7	7
% removed	n/a	92%	92%	90%	92%
Alkalinity (mg/L)	349	257.5	248.0	254.0	251.5
Std. Error	12.3	11.2	3.7	6.2	6.0
# of samples	4	5	5	5	5
% removed	n/a	26%	29%	27%	28%
Supernatant TBOD, mg/L	162.8	53.9	50.6	34.7	54.7
Std. Error	21.9	13.4	4.0	9.0	6.6
# of samples	4	5	4	4	4
% removed	n/a	67%	69%	79%	66%
scBOD, mg/L	74.1	9.3	12.4	10.9	9.9
Std. Error	8.6	1.3	4.3	1.8	2.6
# of samples	4	4	4	4	4
% removed	n/a	87%	83%	85%	87%
DRP, mg/L	6.2	3.2	3.1	3.2	2.7
Std. Error	0.5	0.2	0.1	0.1	0.4
# of samples	4	4	4	4	4

% removed	n/a	48%	49%	47%	56%

5.3 Experiment II

Experiment II attempted to dilute raw primary influent with tap water in order to decrease the influent carbon load. With a decreased carbon load in the influent, the addition of carbon dioxide should have had a more pronounced effect. The ponds did not grow well during the first two weeks of the experiment and no reliable results were gathered.

5.4 Experiment III

The following section includes water quality results for the third experiment. All ponds were operated at a 4-day retention time, as in Experiment I. Due to winter weather conditions, including lower solar radiation and air temperature, pH remained <8.5 in all ponds. The carbon dioxide addition set-point was not exceeded. The ponds were re-mixed after the poor growth during Experiment II, and Ponds 1 and 2 received daily additions of inoculum. Each pond received one gallon of inoculum once per day for the duration of the experiment. All ponds were fed from the same stock of primary wastewater, and all ponds had an approximate channel velocity of 25 cm/s.

This experiment consisted of three extended testing periods, which were separated by the Cal Poly yearly winter break and an early spring pond crash. The first segment of experimentation occurred between October 22nd and December 20th, 2009. During the three-week break between sampling, the influent lines were shut off, but the paddle-wheels were kept on to maintain suspension of biological material. A sheet of plastic was fitted over the top of the first pond in an attempt to maintain a higher temperature. During the beginning of January, 2010, the plastic sheet was removed (no marked temperature increase had been noticed), and the ponds were prepped for the continuation of the inoculum experiment. The configuration of control and experimental ponds remained the same, and all inoculum addition procedures were maintained. This second segment of experimentation occurred between January 9th and March 15th, 2010.

In early March, Pond 1 experienced an extreme decrease in productivity and eventually appeared to have completely crashed. The pond was drained and Pond 2 was evenly distributed between the two experimental ponds. Pond operations began again on March 20th, 2010, but the paddle-wheel motor was accidentally turned up, resulting in a resuspension of settled sludge. This resulted in a four-week period with extremely high solids readings. After allowing the ponds to settle and returning the channel velocity to 25 cm/s, water quality data was collected for a two-week period until the experiment concluded on May 17th, 2010.

The data in the following section is presented in individual analyses to account for the interruptions incurred.

5.4.1 Treatment Performance

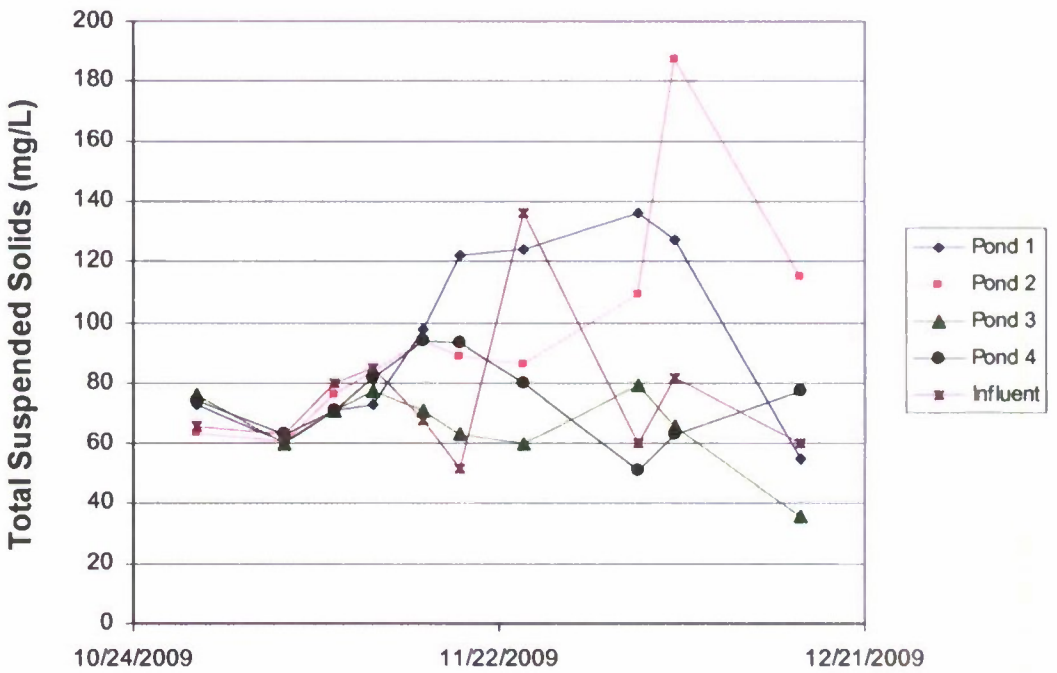
During Experiment III, water quality characteristics of the control ponds and experimental ponds receiving inoculum were compared. The goal of the experiment was to determine a conclusive effect of daily inoculum additions on nutrient removal and/or effluent settleability in the experimental ponds. As referenced earlier, this experiment

encountered two data analysis breaks. The experimental water quality constituents measured were different during each segment due to the availability of student workers and the occasional failure in testing procedure with new or faulty equipment.

During the first experimental segment of Experiment III, the water quality constituents analyzed included initial TSS/VSS, 2-hr and 24-hr TSS, ammonia and alkalinity data. Nitrate and nitrite data did not pass QA/QC during this experiment, and these data are omitted. During the second and third segments of Experiment III, all of the above water quality constituents were analyzed, with the addition of nitrate and nitrite. Average constituent levels for each segment of Experiment III are located at the conclusion of each segment's subsection.

5.4.1.1 Experiment III, Segment One

The addition of inoculum to the experimental ponds was expected to encourage flocculation within the ponds, resulting in improved settling ability. The experimental ponds would also have an artificially higher suspended solids level due to the addition of the concentrated inoculum. During Dan Frost's experiments, in which two ponds received inoculum, and two did not, the inoculum ponds had much higher TSS/VSS concentrations (Frost, 2008). During the first segment of Experiment III, both experimental ponds had higher TSS/VSS concentrations than the control ponds. The experimental ponds averaged 95 mg/L TSS and 90 mg/L VSS, while the control ponds averaged 71 mg/L TSS and 69 mg/L VSS.



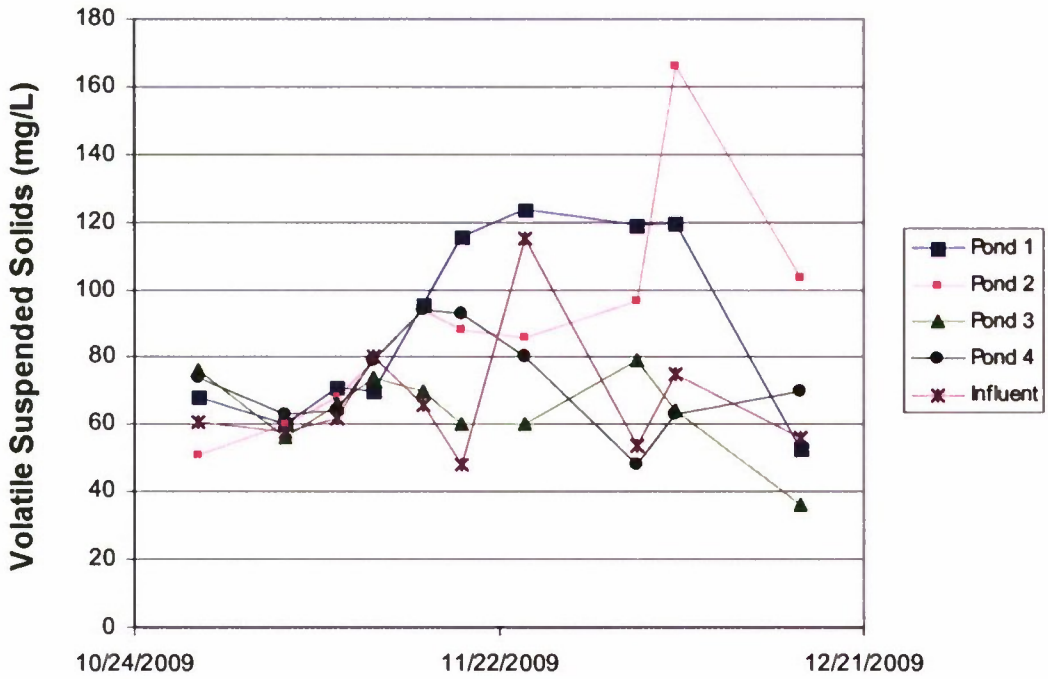


Figure 21: TSS & VSS data for the first segment of Experiment III.

The experimental ponds were expected to have increased suspended solids because of the dense inoculum culture being added daily. Inoculum concentrations ranged from approximately 4000-10000 mg/L, with an average of 6500 mg/L. One gallon (3.78 liters) of inoculum was added per day. Each pond was approximately 940 liters. The daily inoculum addition constituted 0.4% of the total pond volume. Since the inoculum was extremely high in suspended solids, however, the average increase in TSS due to its addition was approximately 26 mg/L. The difference in TSS between the experimental and control ponds was 24 mg/L. The difference might have been due only to the addition of biological material in the inoculum, indicating no increase in overall productivity.

The effect of inoculation on settling was much more pronounced during Experiment III. The 2-hour and 24-hour TSS concentrations in the experimental ponds were expected to be lower than the control ponds because of increased flocculation. During the first segment of Experiment III, the experimental ponds did not clearly exhibit this pattern. The 2-hr TSS numbers for all four ponds did not show a clear advantage in settling with the experimental ponds. Pond 1 (experimental) and Pond 3 (control) had the lowest TSS.

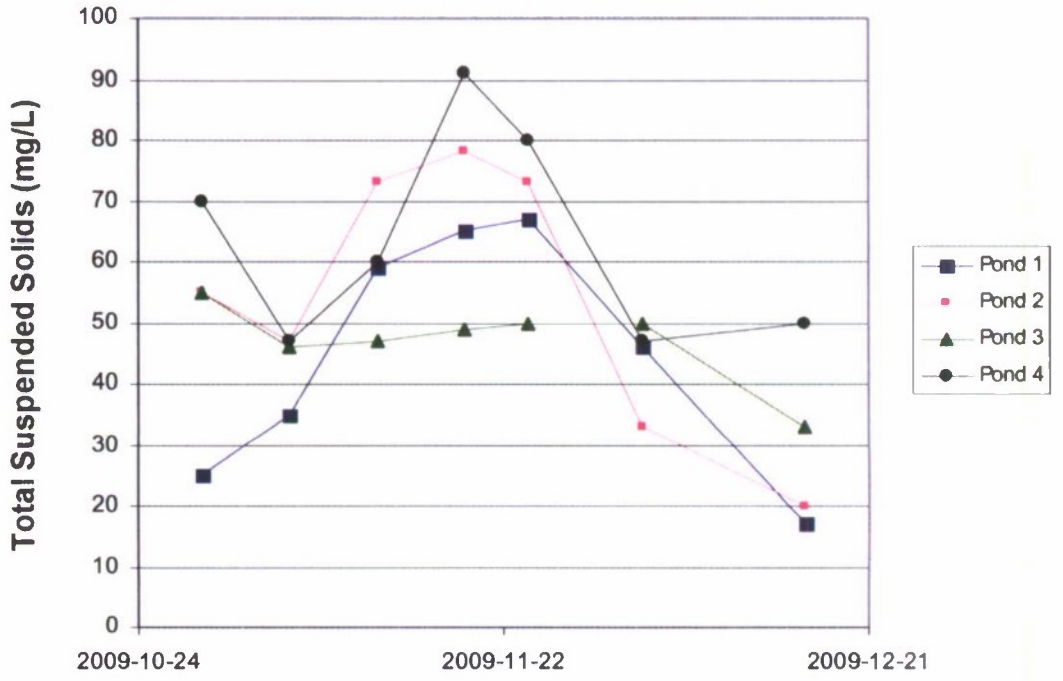


Figure 22: 2-hour TSS settling data for segment one of Experiment III.

The 24-hour settling data also failed to produce results that indicated the desired effect of inoculum in the experimental ponds. The control ponds averaged 41 mg/L TSS after a full day of settling. The first experimental pond (Pond 1) averaged 27 mg/L, indicating better settling performance. However, the other experimental pond, Pond 2, averaged 49 mg/L, preventing a clear conclusion to be reached on the effectiveness of inoculum on improving settling ability.

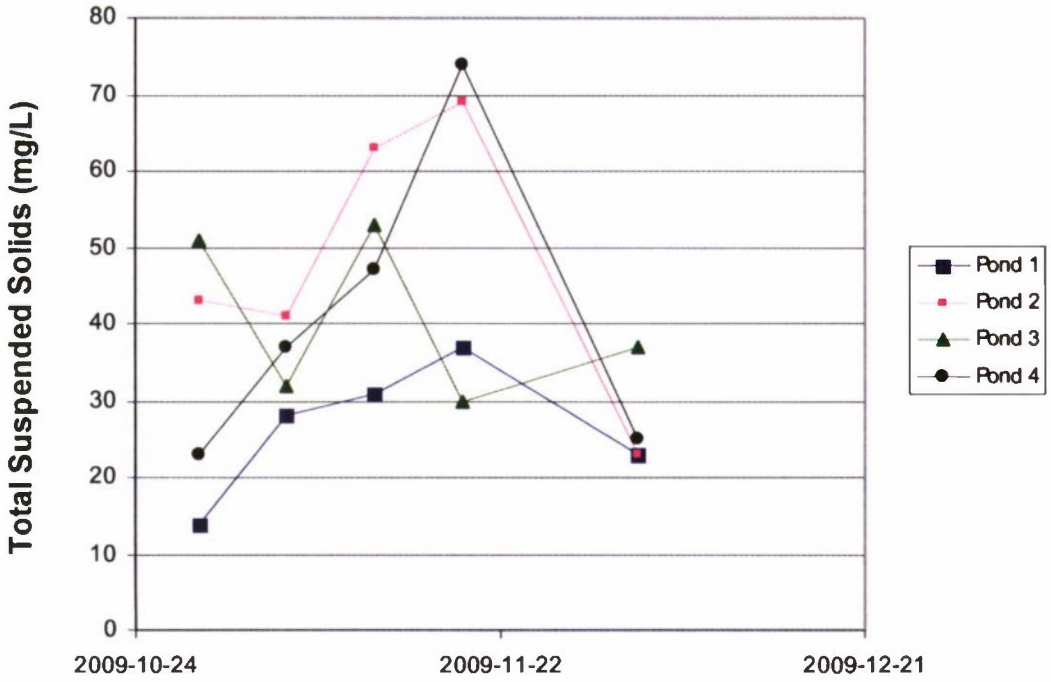


Figure 23: 24-hour TSS settling data for segment one of Experiment III.

The effects of inoculum on ammonia removal were also analyzed during segment one of Experiment III. The daily addition of inoculum to the experimental ponds was expected to establish both flocculating and nitrifying communities. The resultant ammonia levels in the experimental ponds were expected to be lower than in the control ponds.

The experimental ponds both had higher ammonia removal efficiencies than the control ponds. Pond 1 and Pond 2 had average removal efficiencies of 74% and 54%, respectively, while Pond 3 and Pond 4 averaged 45% and 49%, respectively. Pond 1 performed at a much higher level than the control ponds, in terms of ammonia removal, while Pond 2 performed only slightly better than the control ponds. The increase in ammonia removal in Pond 1 was not attributable to raised pH levels, as the pH in all ponds were relatively similar, with Pond 1 at times being lower in pH. The addition of inoculum may have established a functioning community of nitrifying organisms that resulted in increased ammonia consumption. Unfortunately, this effect was not observed in the second experimental pond.

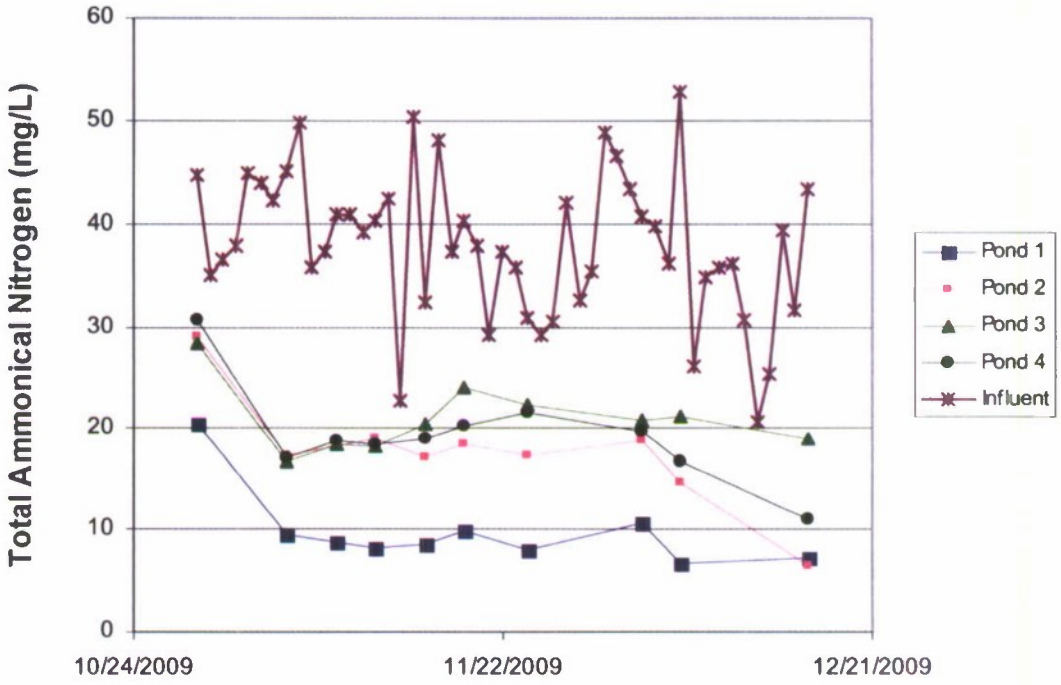


Figure 24: Total ammonical nitrogen levels during segment one of Experiment III.

Alkalinity was also monitored throughout the first segment of Experiment III. The alkalinity removal in Pond 1 was over 50% higher than in the control ponds and the other experimental pond. Alkalinity removal can be an indicator of nitrification, with approximately 7 grams of alkalinity consumption (as $CaCO_3$) being associated with each gram of ammonia converted to nitrate (Frost, 2008). The average alkalinity difference between Pond 1 and the control ponds was 51 mg/L. Assuming the differences in alkalinity consumption were due to increased nitrification in Pond 1, the ammonia levels should differ by approximately 7.3 mg/L. The average ammonia concentration in Pond 1 was roughly 10.5 mg/L lower than levels in the control ponds.

Table 5: Water quality data for the first segment of Experiment III, 10/22/2009-12/20/2009

Constituent	Influent	Pond 1 (w/ inoculum)	Pond 2 (w/ inoculum)	Pond 3 (control)	Pond 4 (control)
TSS (mg/L)	75	94	96	66	75
Std. Error	7.6	9.8	11.5	4.0	4.3
# of samples	10	10	10	10	10
VSS (mg/L)	68	90	89	64	73
Std. Error	6.1	8.9	10.0	3.9	4.5
# of samples	10	10	10	10	10
2-hr TSS (mg/L)	n/a	45	54	47	64
Std. Error	n/a	8.9	9.9	3.1	7.8
# of samples	n/a	7	7	7	7
24-hr TSS (mg/L)	n/a	27	49	41	41
Std. Error	n/a	3.9	8.9	4.8	9.5
# of samples	n/a	5	5	5	5
NH_x-N (mg/L)	38	9.8	17.6	20.9	19.3
Std. Error	1.8	1.2	1.7	1.1	1.6
# of samples	60	10	10	10	10
% removed	n/a	74%	54%	45%	49%
Alkalinity (mg/L)	359	246	287	298	295
Std. Error	14.1	12.0	18.1	9.7	12.5
# of samples	8	7	7	7	7
% removed	n/a	31%	20%	17%	18%

5.4.1.2 Experiment III, Segment Two

The second segment of Experiment III began after the Cal Poly winter break. After a three-week hiatus, the influent supply was turned back on and the experiment was continued.

The suspended solids data during the second segment of Experiment III was not similar to the data acquired in the first segment. The average TSS and VSS values for the experimental ponds and control Pond 3 were very similar, while Pond 4 TSS and VSS values were significantly lower. Ponds 1, 2, and 3 had TSS/VSS values of 129/112, 138/122, and 137/122 mg/L respectively, while Pond 4 averaged 108/87 mg/L. The addition of inoculum did not produce a marked difference between the experimental and control ponds.

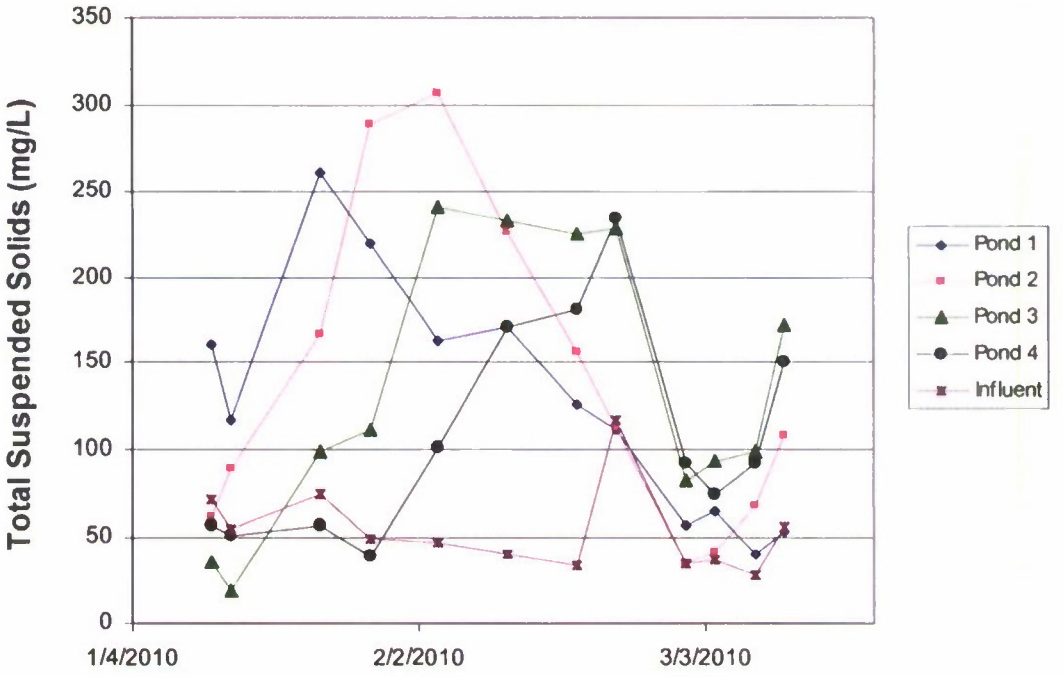


Figure 25: TSS data for segment two of Experiment III.

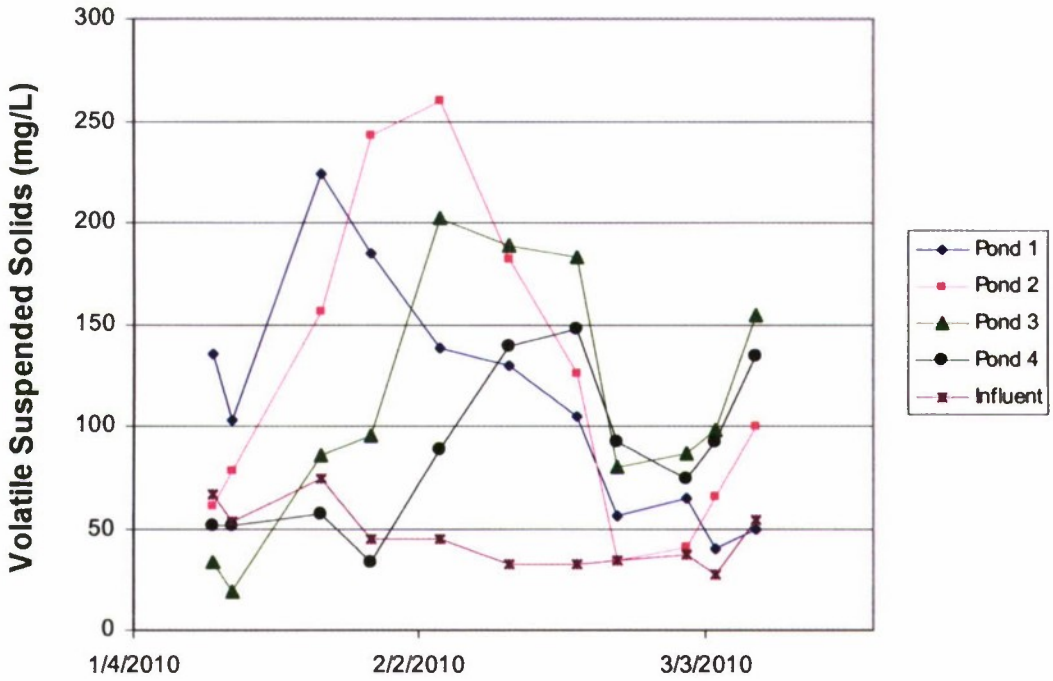


Figure 26: VSS data for segment two of Experiment III.

The settling ability of the experimental ponds was consistent during segment two of Experiment III. The average 2-hr TSS for the experimental and control ponds was 26 and 44 mg/L, respectively. The average 24-hr TSS for the experimental and control ponds was 18 and 37 mg/L, respectively. Over 2-hours, the experimental ponds total suspended solids dropped 79% compared to the initial concentration. Over the same period, the control ponds TSS dropped 64%. Over 24-hours, the experimental ponds' TSS dropped 87%, while the control ponds dropped 70%.

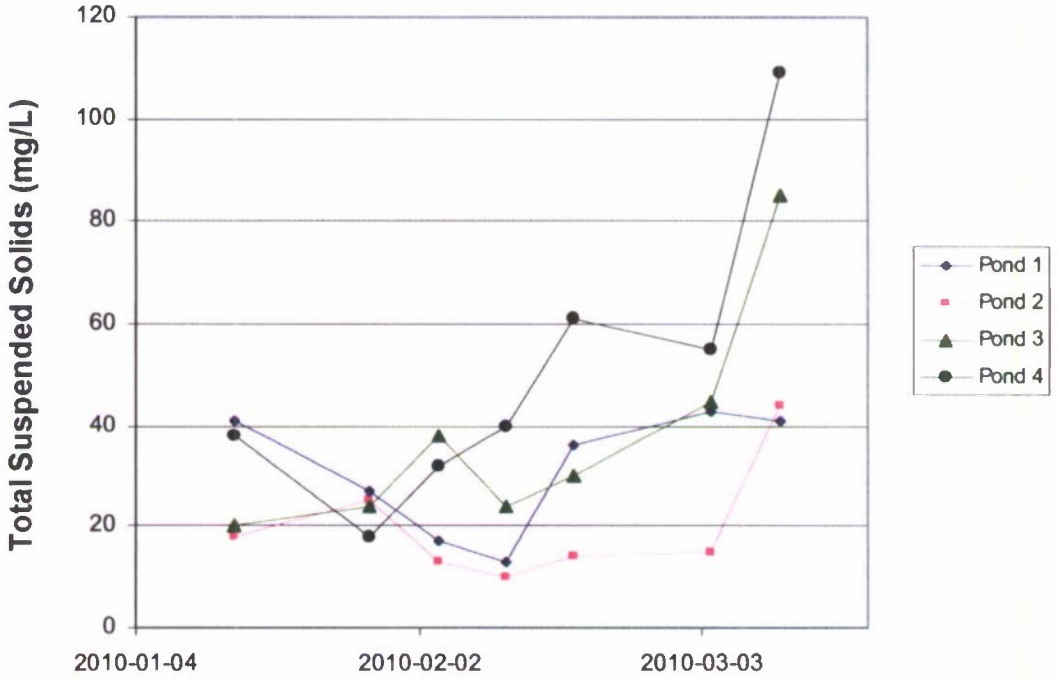


Figure 27: 2-hr settling TSS data for segment two of Experiment III.

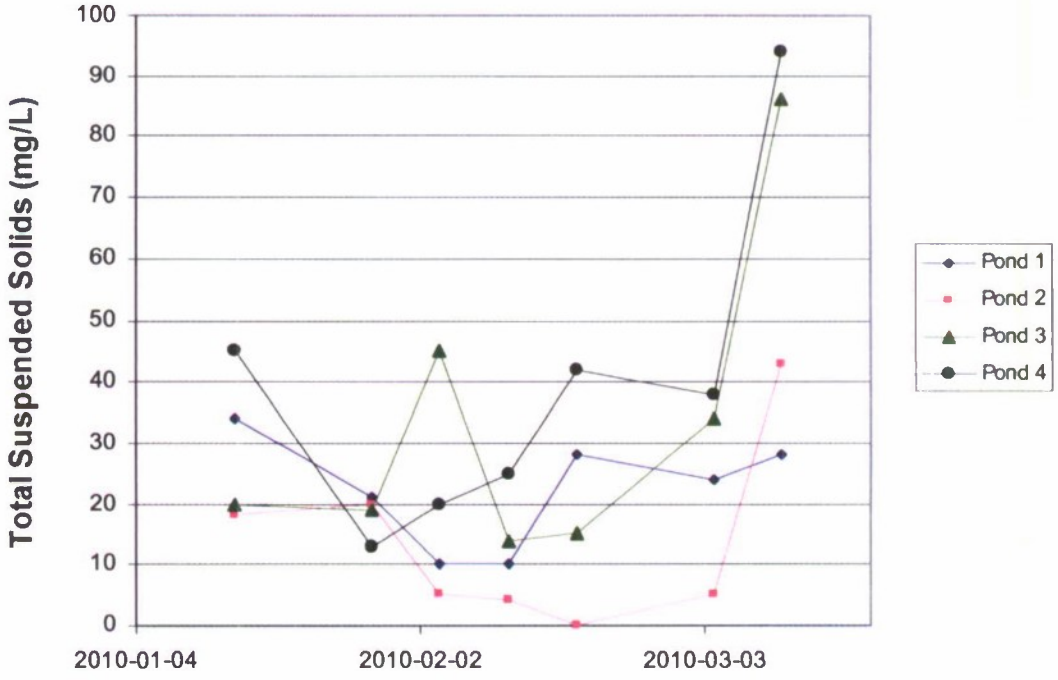


Figure 28: 24-hr settling TSS data for segment two of Experiment III.

Ammonia removal was slightly higher in the experimental ponds, averaging 85% removal, while control ponds averaged 79% removal.

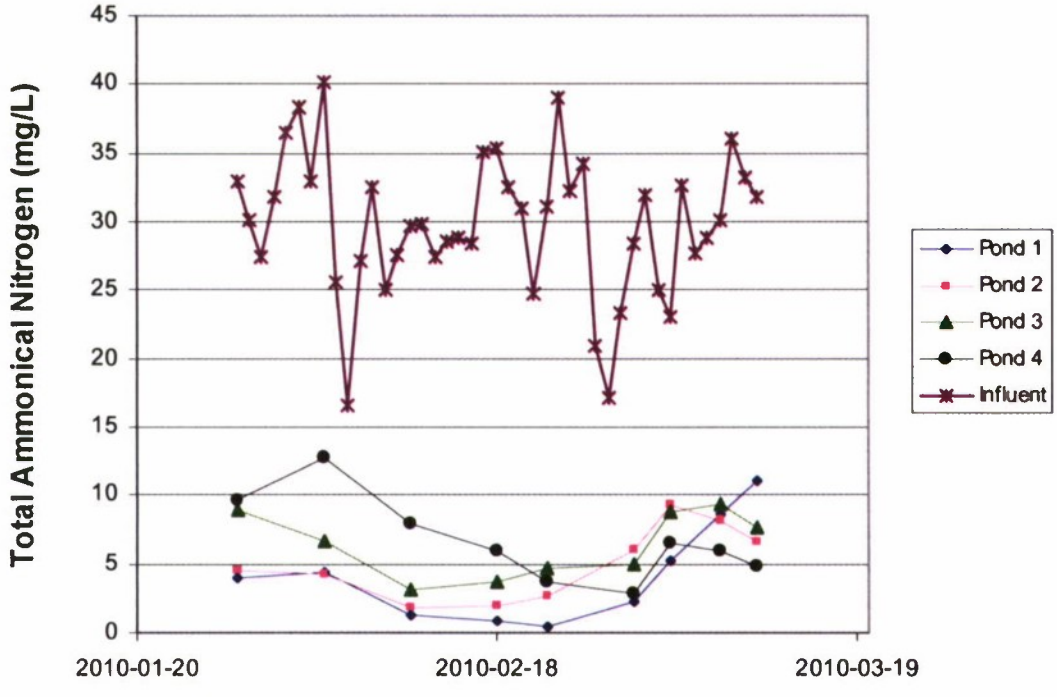


Figure 29: Total ammonical nitrogen data for segment two of Experiment III.

6 Lipid Production, Extraction and Characterization Materials and Methods

The first subsections below describe the cultivation of microalgae used in the lipid quantification and characterization studies. The latter sections provide details of the methods used for the microalgae analyses.

6.1 Setup and Operation of the Deep (70 cm) Algae Pond

Microalgae were cultivated in an open pond to grow biomass for lipid analysis. The pond was operated in batch mode to simplify the observation of time-dependent changes in the microalgae culture. The algae used to inoculate the experiments were grown in an outdoor, continuous-flow pilot algae treatment system at the San Luis Obispo Water Reclamation Facility. The pilot system, operated by Cal Poly graduate student Michael Podevin, used small paddle wheel-mixed high rate ponds fed primary clarifier effluent. The hydraulic residence time of the ponds was maintained at 5 days.

6.1.1 Pond Configuration

The batch-growth pond used in the present research was a 2195 liter rectangular fiberglass tub with 0.75 meter tall sidewalls (Figure 30). While operating, the pond was filled to a depth of approximately 20 cm. In the center of the pond, a plastic partition was erected to recreate the hydraulic configuration of a raceway. A four-bladed paddle wheel, rotating at eight cycles per minute, was used to provide circulation of the water. The pond was sparged with pure carbon dioxide in order to prevent limitation of algal growth by low inorganic carbon concentration. This technique is described in a previous thesis project at Cal Poly (Feffer, 2007). The flow of carbon dioxide was adjusted each day in pursuit of a target pH of approximately 7.75, a favorable condition for algal growth.

Figure 30: Batch-mode pond setup

The algae pond used for this experiment was operated in batch mode to facilitate observation of lipid concentration and characteristics at progressive culture ages and nutrient depletion levels. Batch growth also provided a dense culture, which allowed a higher sensitivity in the resulting lipid data.

6.1.2 Inoculum and Growth Medium

The pond was filled with 855 L of effluent from the primary clarifier of the 4.5 MGD San Luis Obispo Water Reclamation Facility. The wastewater used in the pond was collected between 2:00 p.m. and 5:00 p.m., at the beginning of the evening peak of sewage flow. The effluent contained 80 mg/L of total suspended solids and 72 mg/L of volatile suspended solids. Analyses conducted by Cal Poly graduate student Michael Podevin throughout March and April suggest that the 5-day biological oxygen demand of the effluent was approximately 175 mg/L.

The pond was inoculated with 95 L of water from an adjacent continuous-mode algae pond fed the same primary effluent. The continuous pond was constructed similarly to the batch-mode pond, but it was not sparged with carbon dioxide. The inoculum contained 212 mg/L of total suspended solids and 175 mg/L of volatile suspended solids. Based on analyses performed by Michael Podevin, it is estimated that the 5-day biological oxygen demand of the inoculum was approximately 25 mg/L.

The purpose of inoculating the batch-mode pond with water from a nearby continuous pond was to simulate conditions which may be achievable in a full-scale algae system. Since the microalgae harvested from the continuous pond had been grown outdoors on municipal wastewater, their ability to survive on municipal wastewater in outdoor conditions had been demonstrated prior to the initiation of the batch experiment.

6.1.3 Pond Operation

The pond was inoculated on March 23, 2009 and operated for 25 days through April 17. Once inoculated, the management of pond consisted only of daily adjustments to the carbon dioxide flow to maintain a slightly alkaline pH (~7.75). The air temperature fluctuated between a high of 26 °C on both March 28 and April 5 and a low of 9 °C on April 15. A rainfall event took place over a period of two days between April 7 and 8, with a total precipitation of 0.053 cm. Weather data were gathered from the National Oceanic and Atmospheric Administration, NOAA, database (NOAA National Climatic Data Center, 2009). The NOAA data were obtained from a weather station at the San Luis Obispo County Regional Airport, approximately three kilometers from the algae pond. High and low daily air temperatures from the weather station are shown in the Figure 30. The dates over which the two day-long rain event occurred are indicated by a single cloud.

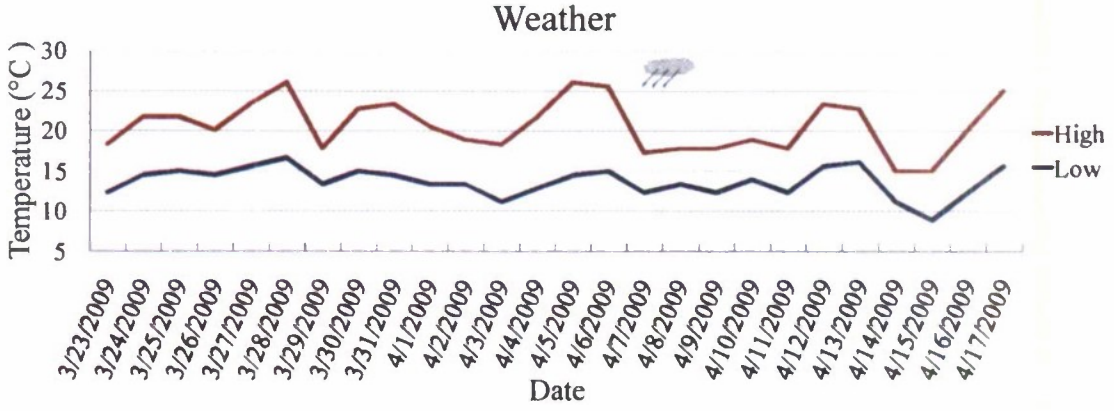


Figure 31: Weather throughout pond experiment

6.2 Sample Collection

Samples were collected from the pond daily, typically between noon and 2:00 p.m. The volume collected depended on which analyses were to be performed each day. After removal from the pond, samples were rushed within 15 minutes to a Cal Poly laboratory where their headspace was purged with nitrogen and they were placed in a refrigerator. Samples were kept in the refrigerator for no longer than three hours before analysis was initiated or steps were taken for longer-term preservation.

Long-term preservation involved centrifuging the samples to remove most of the water and freezing under an N_2 atmosphere. This procedure will be discussed in greater detail in the section entitled Bligh and Dyer.

6.3 Experimentation

Two sets of experiments have been conducted over the course of this research. One study involved daily analysis of the lipids in the batch-mode pond. This study was meant to provide information about the development of lipids over the growth and death of an algal culture.

The other study involved the comparison of a number of different procedures for the extraction of the lipids from the algae grown in the pond.

6.3.1 Lipid Production Experiment

A series of analyses were conducted to relate the growth stage of the batch-mode algal culture to the quantities and types of lipids present in the algal mass. The study involved regular testing of total and volatile suspended solids of the pond, as well as determination of the lipid content of the algae in the pond. The sampling scheme is described below. The primary effluent used to fill the pond and the algae used to inoculate the pond were tested in the same manner, prior to the initiation batch growth. On day 11 of the experiment, the total and volatile suspended solids concentrations of the supernatant fluid were tested after the centrifugation of the sample.

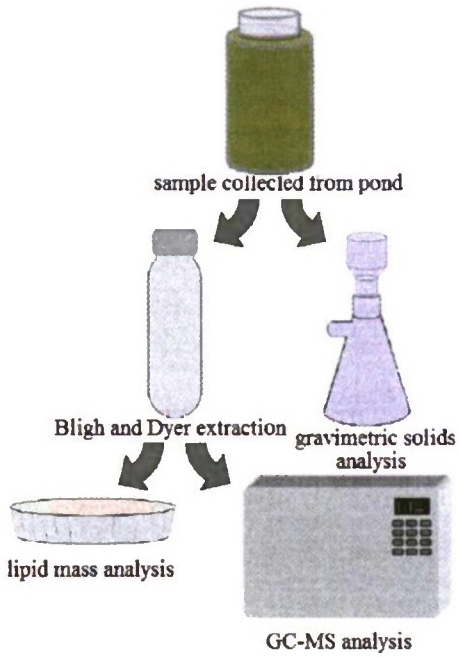


Figure 32: Lipid production experiment sampling scheme

6.3.1.1 Solids Determination

Tests for total and volatile suspended solids were conducted to estimate the concentration of algae in the pond. Solids tests were conducted according to Standard Methods 2540 D (APHA, 1995).

6.3.1.2 Microscopic Investigation

Over the course of the operation of the batch-mode pond, a microscopic analysis of the micro biota present was performed every five days. The inoculum and growth medium were also examined using a microscope. An Olympus CX 41 microscope was used in conjunction with Olympus software to capture images of the algae growing in the culture. A hemacytometer was used prior to each microscope session to calibrate the software and provide an accurate scale-bar in the images. Tentative identifications to the genus level were made using a freshwater algae key (Prescott, Bamrick, Cawley, & Jaques, 1978).

6.3.1.3 Lipid Investigation

Throughout the operation of the batch-mode pond, analyses were conducted of the quantity and the identity of lipids growing in the algae. The same extraction procedure was used in both cases.

6.3.1.3.1 Bligh and Dyer Extraction

The Enssani (1987) adaptation of the Bligh and Dyer procedure was used to extract lipids from the algae for further study.

For each test to be performed, a 200 mL aliquot of water collected from the pond was centrifuged into a pellet at approximately 2800 g. The relative centrifugal field was determined as follows:

$$\textit{Relative centrifugal field} = \frac{\omega^2 r}{g},$$

where ω is the angular velocity in radians per second, based on a rotational frequency of 4,000 rotations per minute. Although the centrifuge used was throttled by a powerstat and did not report rotations per minute, the rotational speed was assumed from similar models. The value r is the distance between the center of the centrifuge tube and the center of rotation, 8.8 cm, and g is the acceleration of gravity.

Due to the limitations of test tube capacity, the centrifugation process was undertaken 40 mL at a time. After 40 mL had been centrifuged for four minutes, the supernatant liquid was poured off from each sample and replaced with 40 mL of sample water. This process was repeated until the solids from 200 mL of sample water had been collected in a pellet at the bottom of each test tube.

Centrifugation was performed in tared, solvent resistant, 50 mL, Teflon, round bottom centrifuge tubes. A room temperature table-top centrifuge was used. The centrifugation of each pellet required approximately 25 minutes and the pellets were centrifuged four at a time.

After the samples had been pelleted in the centrifuge tubes, they were either used for extraction immediately or were flushed with nitrogen and placed in a freezer until needed.

The first step of the extraction was to add 5 mL of chloroform, 10 mL of methanol and 4 mL of deionized water to each tube containing a pelleted sample of algae. The methanol, which is miscible in the water layer, is included to amend the polarity of the water layer to limit the concentration of polar lipids such as chlorophyll and phospholipids in the chloroform layer. Once the tubes had all reached room temperature, a sonicator was used to disrupt the algae cells in the suspended mixture. A Branson Ultrasonics sonicator was operated at a constant duty cycle on power level 8. Each tube was sonicated for one minute. The sonicator tip was rinsed between each sonication to remove any residual material. The samples were then placed horizontally on a shake table with a single-axis motion of 6 cm oscillations at 2 cycles per second for between 6 and 8 hours. The purpose of the shaking step was to promote the complete exposure of intracellular products to the solvents.

The tubes were removed from the shake table and an additional 5 mL of chloroform and 5 mL of deionized water were added to each sample. Each tube was vortex mixed for 30 seconds to mix the newly added solvents. The tubes were then centrifuged at 4850 g for four minutes to separate the contents into layers.

A test tube containing the mixture at this stage is shown in Figure 33 33.

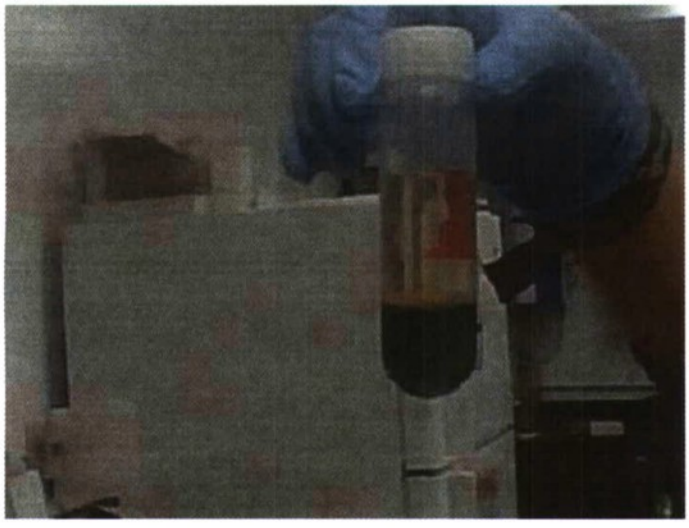


Figure 33: Bligh and Dyer extraction with chloroform layer at bottom

The green layer at the bottom, comprised primarily of chloroform, contains lipophilic material. The upper layer contains methanol and water. A thin layer of cell debris separates the two layers.

The lower, lipid-rich chloroform layer was removed using a glass Pasteur pipette for additional testing. This extract was pipetted into a 60 mL Luer-lock syringe made of solvent-resistant polypropylene. The syringe was used to force the extract through a 0.2 µm, nylon syringe-driven filter. Depending on whether the extract was to be used for lipid quantity analysis or lipid identification, it was injected into either a tared aluminum weighing dish (4 cm diameter) or a tared glass test tube (50 mL).

After the extract had been pipetted and filtered, an additional 10 mL of chloroform were added to each of the Teflon centrifuge tubes. The tubes were again vortex mixed for 30 seconds each and centrifuged at 4850 g for four minutes. The chloroform layer was again pipetted from the bottom of each tube into the same vessel as had been used during the prior set of extractions.

6.3.1.3.2 Lipid Mass Determination

For the preparation of samples to be used for the determination of lipid mass, a syringe-driven filter was used, as described previously, to deposit the algae extract into a tared aluminum weighing dish. The use of aluminum trays, as opposed to glass test tubes, is preferable for the weight determination of lipids because the trays are not as massive compared with the lipids as test tubes would be. This helps to minimize balance error.

After the second extraction had been performed, the weighing dishes were placed in an approximately 20 °C desiccator cabinet which was supplied with a constant flow of nitrogen. The flow of nitrogen through the desiccator was meant to limit oxidation of the lipids that could have taken place while the extracts were drying. The extracts remained in the desiccator until no chloroform was visually apparent in the dishes (typically 24 hours).

Figure 34: Samples in the nitrogen-sparged desiccator

The weighing dishes were then placed in a 105 °C oven, which was also flushed continuously with nitrogen, for one hour. The purpose of this step was to ensure that all of the solvent had been removed from the weighing dishes, leaving only algal lipids.

After drying in the oven, the dishes were allowed to cool in a nitrogen-sparged desiccator and weighed to the nearest 0.01 mg. The weight measurements were used to determine the concentration of lipids in the algae and were used with total volatile solids data to calculate the concentration of lipids in the algae water. The mass of each sample was measured in triplicate.

Although the Bligh and Dyer procedure is a common laboratory method, it is suspected of overestimating the lipid mass of microalgae (Woertz, Feffer, Lundquist, & Nelson, 2009), (DeLuca, Charity. pers. comm., 2009). The solvent system extracts nonpolar non-lipid material in addition to triglycerides, which can cause an overestimate of total lipid mass. The Bligh and Dyer method is not selective enough to exclude lipophilic compounds such as chlorophyll from extraction (Woertz, Feffer, Lundquist, & Nelson, 2009). This error is not a severe problem when oilseeds are the subject of research, as they do not contain chlorophyll (Lehr, Corinne. pers. comm., 2009).

6.3.1.3.3 Triglyceride Identification

In addition to mass analysis, samples from the pond were extracted each day to identify the types of triglycerides that they contained. The Bligh and Dyer procedure was used, exactly as in the samples used for weight determination. However, instead of extraction into weighing dishes, the samples used for lipid analysis were extracted into glass test tubes. Glass test tubes are preferable to weighing dishes for this analysis because they are sealable. A spill-proof seal facilitates the mixing of toluene with the lipids, which is necessary for chromatographic analysis.

The extracts in the test tubes were dried to constant weight in a Caliper Life Sciences nitrogen-sparged desiccator apparatus at 30 °C, which typically required 4 hours. It was

not possible to use the same apparatus with the lipid mass determination samples because aluminum weighing dishes do not fit in the Caliper Life Sciences desiccator. Had the weighing dishes and desiccator been compatible, the lipid mass determination samples would have been dried in the Caliper Life Sciences system, as well.

6.3.1.3.3.1 Transmethylation

After extraction into glass tubes, the extracts were either used immediately for lipid analysis, or the headspaces were flushed with nitrogen and the tubes were placed in a freezer.

Because triglycerides are relatively large molecules and have high boiling points, they are not easily analyzed by gas chromatography. In preparation for chromatographic analysis, the lipid samples were transmethylated in order to cleave each triglyceride into its constituent fatty acids. This step makes possible the chromatrographic processing needed to identify triglycerides. The transmethylation reaction resulted in the production of fatty acid, methyl esters, which, incidentally, are the principle molecules of biodiesel fuel.

The first step of the transesterification was to resuspend up to 10 mg of the oil extracts in 1 mL of dry toluene. Since different tubes contained different masses of lipid material, the volume of toluene added varied between samples. In each case, enough toluene was added so that no more than 10 mg of lipids were present in the tube for each milliliter of toluene. A 1 mL aliquot of the dissolved solution was then pipetted into a new tube for further processing.

Anhydrous sodium methoxide in methanol (2 mL of 0.5 M solution) was added to each sample. This basic reagent was selected because the continual regeneration of methoxide reactants during the transesterification promoted a complete reaction. In the base-catalyzed reaction, as fatty acids are cleaved from triglycerides, they form a highly basic glyceroxide which can acquire a proton from methanol, regenerating the methoxide reagent. The reaction mechanism is illustrated Figure 3535.

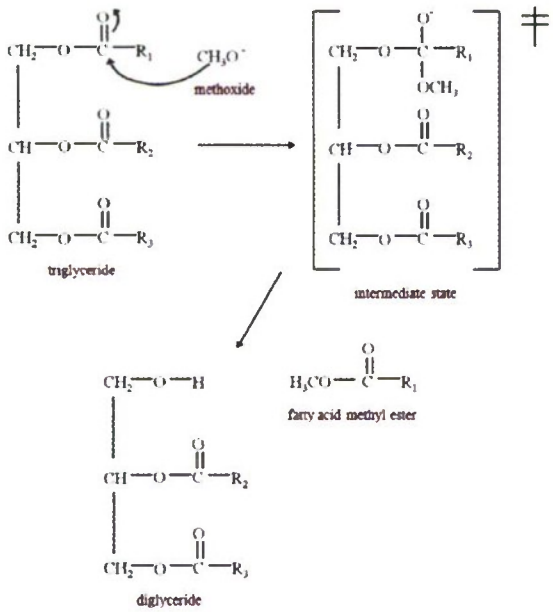


Figure 35: Base-catalyzed reaction mechanism

After the addition of the sodium methoxide solution, the reaction was allowed to proceed for ten minutes in a 50 °C water bath. After the mixture had reacted for ten minutes, the methoxide was neutralized by the addition of 0.1 mL of glacial acetic acid to each sample. A 5 mL aliquot of hexane and a 5 mL aliquot of deionized water were added to each sample. Each tube was then shaken by hand for approximately 30 seconds.

After the mixed samples had been allowed to settle for approximately five minutes, the FAME-rich hexane layers were each transferred by glass pipette to a clean glass test tube. An additional 5 mL of hexane were added to each tube and the mixtures were again shaken. After separating into layers, the hexane layer from each sample was pipetted into the glass test tube which already contained the first extract from it.

Anhydrous sodium sulfate was added to each tube of extract. The extracts were shaken for approximately 15 seconds to promote removal of residual water by the sodium sulfate. Each tube of extract was then gravity filtered through Fisher P5 cellulose paper into a glass test tube to remove the sodium sulfate. If the samples were not to be used immediately, the headspaces were flushed in nitrogen and the tubes were frozen.

6.3.1.3.3.2 Gas Chromatography-Mass Spectroscopy

After transmethylation, the samples were analyzed using a gas chromatograph-mass spectrometer. A small aliquot of each sample was diluted 1:10 in hexane in order to ensure that no constituent was present in such large concentration that a stronger dilution would damage the mass spectrometer.

An Agilent 6890 gas chromatograph was used with an Agilent 575S mass spectrometer for the testing. A fused silica 50 m by 0.25 mm column (Agilent #190915-433) was used in the chromatograph.

The chromatograph was programmed to run a 10:1 split with a sample size of 5 µL for a total column flow of 13.7 mL per minute over the duration of the 10.20 minute run. The inlet was heated to 250 °C and the initial temperature of the column was 120 °C. After injection, the chromatograph ramped at 50 °C per minute to 280 °C and held for an additional two minutes until the end of the run.

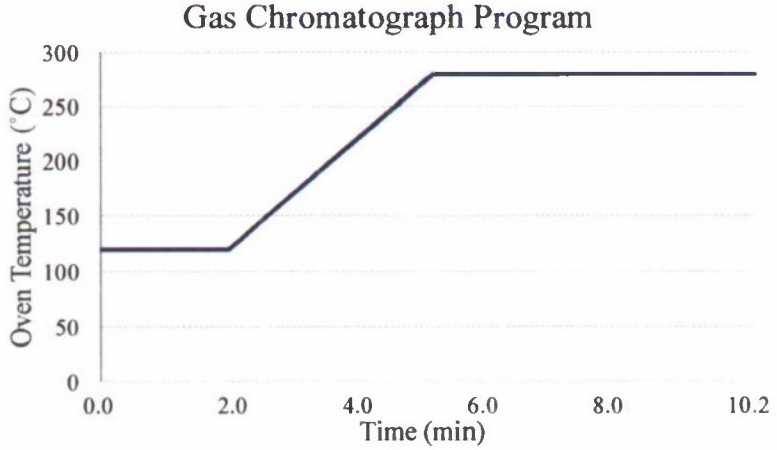


Figure 36: GC program

The quadrupole was heated to 250 °C and the mass spectrometer source to 230 °C. An autosampler was used. The autosampler was programmed to flush twice with hexane prior to injection. It then rinsed with the sample three times and drew the final sample, using four fill-discharge cycles to minimize air bubbles in the syringe. Subsequent to injection, the autosampler rinsed twice with hexane. Each sample vial was analyzed by GC-MS in triplicate.

The bracket method of Volmer, Meiborg and Muskiet was used to quantify the concentrations of each of the constituents of the lipid extracts. Pentadecanoic acid, methyl ester and heptadecanoic acid, methyl ester were selected as standards. These standards were chosen for three reasons.

First, since they are very closely related to the target compounds, they are expected to behave similarly during chromatographic analysis.

Second, despite a close structural relation to algae oils, they are not naturally present in algae samples. Since de novo synthesis of fatty acids is primarily conducted by the repeated addition of two-carbon malonyl-coA units to an existing carbon chain, fatty acids typically have an even number of carbon atoms in their primary carbon chains.

Third, these standards were selected because preliminary study had suggested that the fatty acid methyl esters derived from wastewater would elute after pentadecanoic acid, methyl ester and before heptadecanoic acid methyl ester.

After adding the standards to the samples, the concentration of each constituent of the samples was computed as follows. A proportionality factor was calculated for each standard by dividing the concentration of the standard by the area reported by the mass spectrometer. A unique constant was determined for each constituent of each standard. The rate of change of the proportionality constant over elution time was computed for the period of time between the elution of pentadecanoic acid methyl ester and nonadecanoic acid methyl ester. The rate of change was used to extrapolate a proportionality constant for each constituent in each sample. This proportionality constant and the peak area of the constituent of interest were used to determine the concentration of each constituent.

$$area = \int_{Time\,1}^{Time\,2} Abundance$$

$$Proportionality\ constant, P = \frac{Conc.}{Area}$$

$$Slope = \frac{\frac{Conc._{C19}}{Area_{C19}} - \frac{Conc._{C15}}{Area_{C15}}}{Time_{C19} - Time_{C15}}$$

$$Conc.A = Area\,A \times P_{15} + [Slope \times (Time_A - Time_{15})]$$

A careful inspection was made of each peak with an area 2% or greater than the size of the largest peak in each chromatograph. Constituents present in very small quantities were omitted from the analysis, as they are difficult to identify with confidence.

Molecules were identified based on the time at which they eluted from the gas chromatograph and from the constituent peaks of their mass spectra. Mass spectra were compared with standards from the MS Search 2.0 database maintained by the United States National Institute of Standards and Technology (NIST) (National Institute of Standards and Technology, 2005). In approximately 10% of the samples, the mass spectra of some of the peaks did not align well enough with NIST standards to make a conclusive judgment of their identity. In these cases, the constituents were omitted from the results. These omissions did have a large impact on the results of the experiments because these peaks were identified conclusively in other samples.

Approximately 5% of the samples contained contaminants which coeluted with important analytes. This phenomenon is often readily apparent by the occurrence of a shoulder in a chromatogram. A shoulder is a term given to a pair of peaks that are partially superimposed on a chromatogram. This makes it difficult to determine the true abundance of either constituent. Thus, coeluting compounds were omitted from the results. These omissions did not have a large impact on the results of the experiments because these peaks were identified conclusively in other samples.

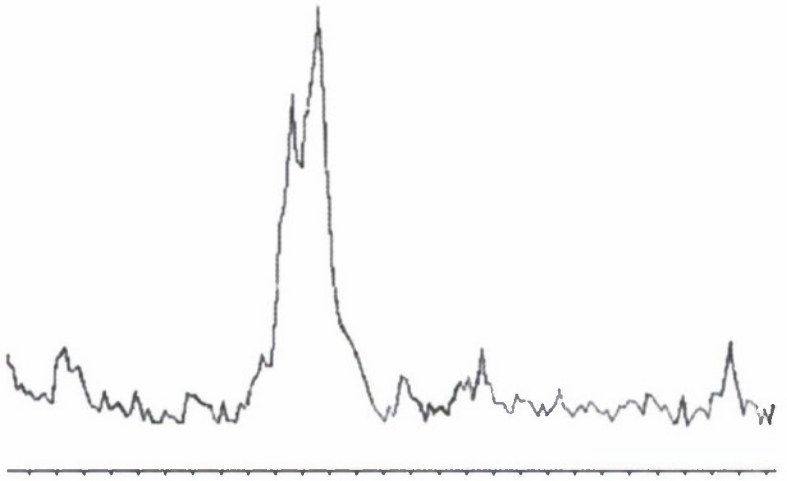
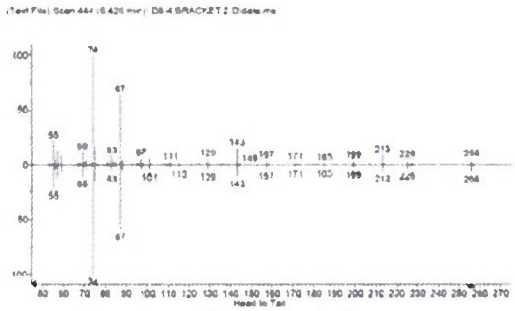
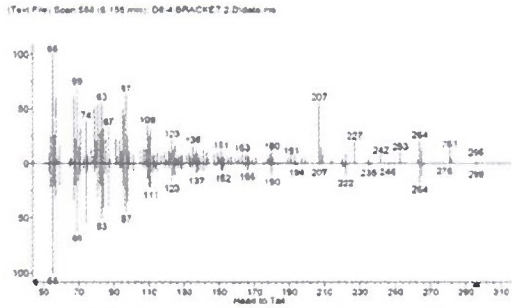


Figure 37: Shoulder feature in a chromatogram from Day 8

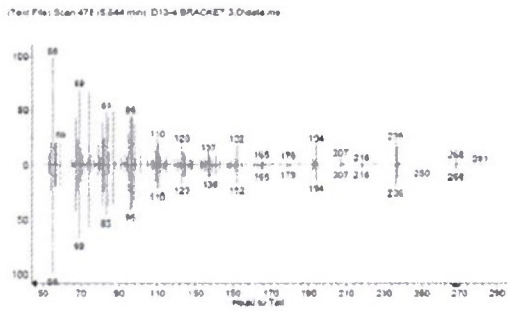
Several head-to-tail comparisons of sample mass spectra against NIST standards are shown below as examples. These comparisons were selected because they were readily identifiable. The spectrum on the upper portion of each graph is from samples taken during the current research. The spectrum on the lower portion of each graph is from the NIST reference database.



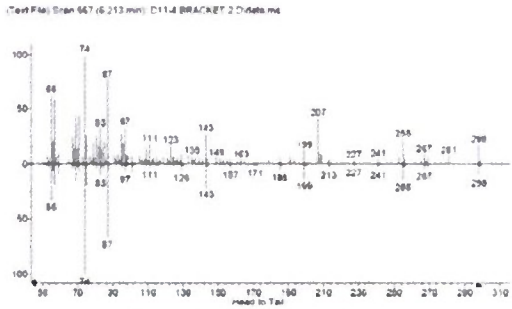
Pentadecanoic acid, methyl ester
Top: Day 8 Sample
Bottom: NIST Standard



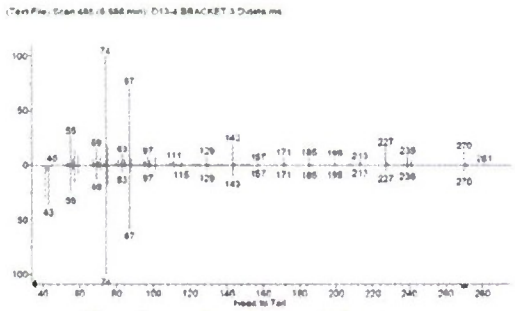
9-Octadecenoic acid, methyl ester (Z)
Top: Day 8 Sample
Bottom: NIST Standard



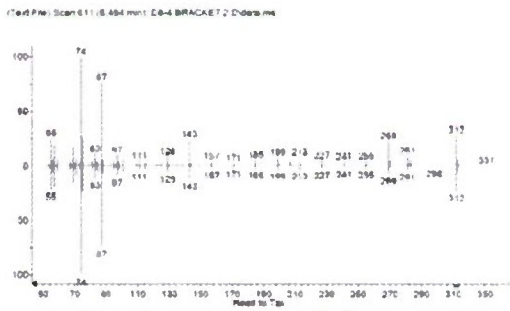
9-Hexadecenoic acid, methyl ester (Z)
Top: Day 13 Sample
Bottom: NIST Standard



Octadecenoic acid, methyl ester
Top: Day 11 Sample
Bottom: NIST Standard



Hexadecenoic acid, methyl ester
Top: Day 13 Sample
Bottom: NIST Standard



Nonadecenoic acid, methyl ester
Top: Day 8 Sample
Bottom: NIST Standard

Figure 38: Head-to-tail comparisons of common constituents to reference chromatograms

6.3.2 Extraction Procedure Experiment

An experiment was conducted to compare several methods of solvent extraction in terms of the quantity and types of lipids they extracted from algae. A large number of sample pellets were prepared on Day 9 of the operation of the batch mode pond. The samples were collected and stored as in the lipid production experiment. These samples were used to compare three experimental extraction procedures to the lab-standard Bligh and Dyer. Each of the three experimental procedures had identical steps, but used different counter-solvents: methanol, ethanol and isopropanol.

After the procedure, the extracts were evaluated for the quantity and identity of the lipids they contained. The analysis for both lipid mass and lipid characterization were performed by the same methods as in the lipid production experiment. The experimental scheme is illustrated in Figure 39.

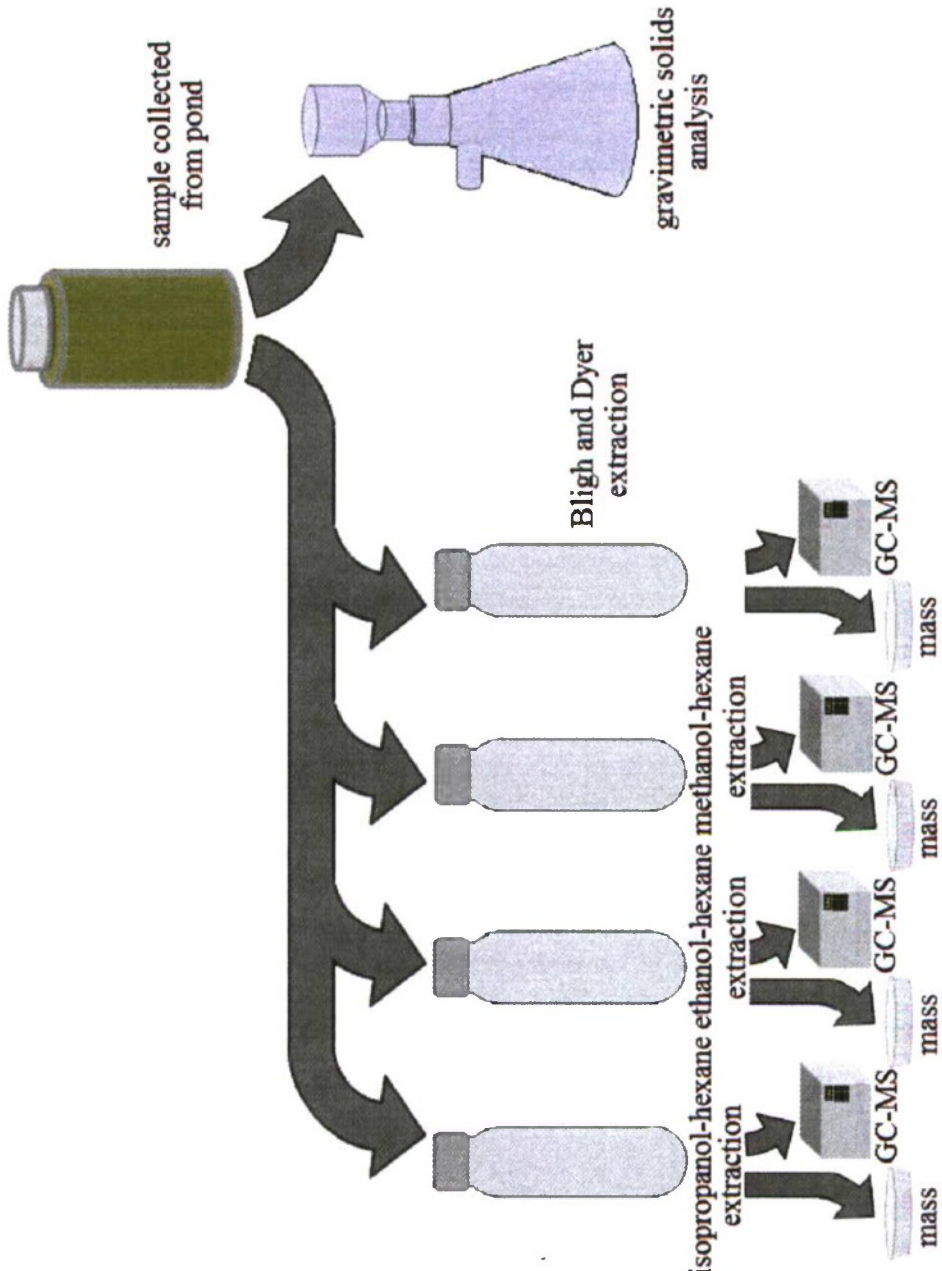


Figure 39: Extraction experiment scheme

6.3.2.1 Alcohol Extraction

A simpler procedure was used to compare the different solvents against the Bligh and Dyer extraction. The purpose of the simple procedure was to simulate an extraction which may be possible on an industrial scale. The extractions used an alcohol-water-hexane solvent system. Methanol, ethanol and isopropanol were each tested as countersolvents. The electric dipole moments of each of these solvents are presented in Table 7.

Table 7: Dipole moments of organic solvents (Newton, 2009)

First, 5 mL of the alcohol being tested were added to each sample to pretreat the pelletized algae. This pretreatment step was meant to test the ability of the alcohols to free lipids from algae cells in the samples. The addition and mixing of the alcohols was a low-energy process which may accomplish the same task as sonication without sacrificing scalability. The tubes containing the algae were then allowed to warm to room temperature, allowing the alcohol time to penetrate the algae cells in the samples. After 15 minutes of pretreatment, the samples were transferred from the centrifuge tubes to glass test tubes.

Deionized water (4 mL) was added to each sample. The water was added first to the original centrifuge tube that had contained each sample and then transferred into the glass tube which contained the sample at this point. This step was taken in order to recover any residual algae from the centrifuge tubes.

Hexane (2 mL) was added to each tube and the tubes were hand mixed for 15 seconds. The mixtures were then given approximately five minutes to separate into layers before the hexane layer, which contained the extraction lipids, was transferred by pipette to a tared glass test tube. An additional 2 mL of hexane were added, mixed and transferred. The purpose of performing the hexane addition and transfer two times was to achieve nearly the maximum possible removal of lipids from the sample.

The extracts were dried, as in the lipid production experiment, according to the vessels in which they were contained. Again, mass and identity were measured. Six algae pellets were extracted with each of the four solvent systems (methanol-hexane, ethanol-hexane, isopropanol-hexane and Bligh and Dyer). All 24 of the pellets used in this experiment were derived from the sample algae sample, collected on Day 9. For each solvent system, three pellets analyzed gravimetrically to obtain lipid mass and three pellets were analyzed chromatographically for triglyceride identification.

Some sample vials during alcohol extraction are shown in Figure 40 directly after the first addition of hexane. A methanol-hexane sample is on the left, an ethanol-hexane sample is in the middle and an isopropanol-hexane sample is on the right.

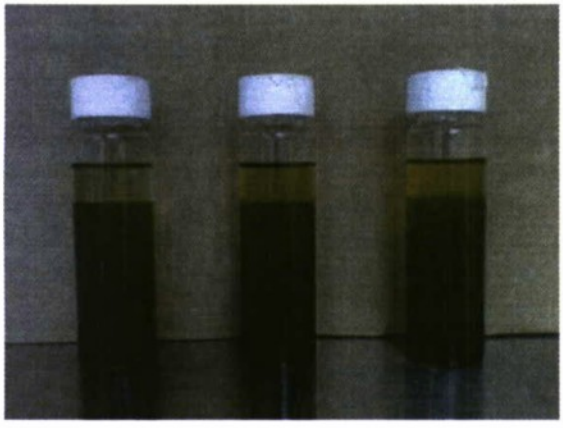


Figure 40: Alcohol-hexane extractions (methanol, ethanol, isopropanol, left to right)

6.4 Quality Control

Several quality control tests were undertaken to verify the accuracy and repeatability of the extraction procedures investigated in the present research. Quality control measures included the performance of blank extractions, control extractions using known quantities of oil, and matrix spikes.

The blank extractions were conducted by performing all of the steps of a normal extraction on a clean test tube without a sample in it. This procedure is used to reveal any error caused by the extraction procedure. The absence of a sample in the tube means that no mass should be produced by the extraction.

Blank extractions were conducted using each of the extraction procedures described above, including the Bligh and Dyer procedure. The Bligh and Dyer method was the only procedure that had a detectable error. The blank Bligh and Dyer extraction produced 0.2 mg of material. In an actual Bligh and Dyer extraction of a sample, the mass of this material would have been erroneously included in the lipid fraction. The erroneous 0.2 mg would have been equivalent to 8.3% of the mass of the single lowest-yielding sample from the entire experiment, which is a minor error. This suggests that the lipids measured by the Bligh and Dyer procedure in the experiments were can be considered to have been derived from the algae.

Control extractions are performed by adding a known mass of oil to a test tube and performing an extraction on it. Ideally, the mass of oil added is equal to the mass produced by the extraction. Vegetable oil was used as the control material. This procedure was performed on methanol-hexane, ethanol-hexane, isopropanol-hexane, and Bligh and Dyer extraction procedures. The extractions had negative errors of 1.32%, 1.37%, 1.02%, and 2.37%, respectively. The errors could have been due to inefficient partitioning of the oil into the nonpolar layers during the extractions.

A matrix spike is a quality control measure in which a known mass of oil is added to a sample and an extraction performed on the mixture. This test confirms that the sample

matrix does not interfere with experimental accuracy. Once again, vegetable oil was used to spike the sample. The matrix spike test was conducted only on the Bligh and Dyer method. Matrix spikes were omitted for the alcohol-based extraction methods because, unlike the Bligh and Dyer methods, it was not assumed that they extracted the maximum possible mass of lipids from each sample. The mass of lipids derived from the algae using the alcohol-based procedures was inestimable and a matrix spike would not have provided any reliable data about the effectives of the experimental process.

In order to perform a matrix spike on the Bligh and Dyer procedure, three samples were prepared in centrifuge tubes, as in the normal extraction experiments. Two of the tubes were used for normal Bligh and Dyer extractions. This made possible an estimation of the native concentration of lipids in the sample to be spiked. A mass of vegetable oil approximately three times the expected native mass was added to the spike sample. This effectively quadrupled the mass of oil in the test tube. The spike resulted in a negative error of 6.14%.

7 Lipid Production, Extraction, and Characterization Conclusions

7.1 Experimental Findings

These experiments provide some insights into the potential of wastewater algae to support fuel production. They also provide information about the practicality of using different solvents in a future industrial process to produce algae biofuel.

7.1.1 Lipid Production Experiment

This set of experiments led to conclusions about the production rates of algal fatty acids and the types of fatty acids produced in batch cultures by algae growing with a municipal wastewater medium. It also led to some conclusions about the analytical methods currently used to study algal lipid production.

7.1.1.1 Lipid Production

The highest average rate of lipid production occurred during the period of rapid growth between Day 11 and Day 13. This peak production rate of lipids, 4.40 g/m^2/day, determined by the chromatographic method, corresponds with approximately 1,900 gallons of biodiesel per acre per year. This figure is based on an assumed algae fuel density of 0.89 g/mL and on an assumed 365 days per year of pond operation. It compares favorably to the results of a similar experiment conducted using dairy wastewater in which a biodiesel productivity of 1,200 gallons per acre per 365-day year was projected (Woertz, Feffer, Lundquist, & Nelson, 2009). The average growth rate of 0.97 g/m^2/day observed over the entire growth phase of the culture, from Day 4 to Day 17, corresponds to approximately 420 gallons of biodiesel per acre per year.

These production values indicate that wastewater algae are capable of supplying enough oil to support industrial fuel production. The areal productivity observed in this study is much greater than values typically quoted for conventional biodiesel feedstocks such as palm and soy (Chisti, Biodiesel from microalgae, 2007). The non-ideal conditions of the

present study (described below) suggest that the potential of algae as a biofuel feedstock may be even greater than reported at present (Chisti, 2007), (Woertz, Feffer, Lundquist, & Nelson, 2009).

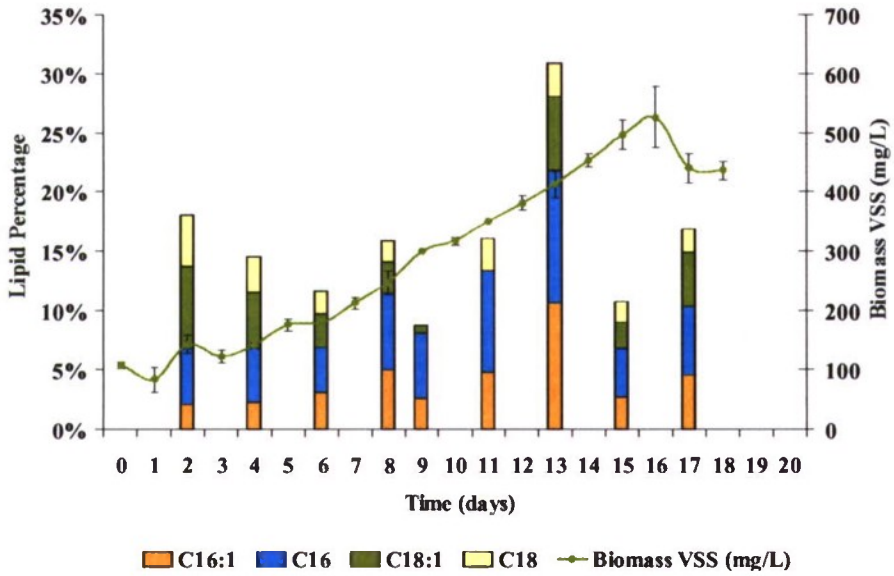


Figure 41. Ponds of 5-m^2 area operated in batch mode with domestic wastewater. Lipid productivity usually peaks in the log phase of growth (not in nutrient starvation phase). Algae fatty acid methyl esters are mostly C16 & C18, which is well-suited to biodiesel production.

It is likely that even higher triglyceride productivities are possible with wastewater algae grown under improved conditions. For example, this experiment was conducted between March and April. The average daily high temperature of approximately 18 °C may be lower than ideal. Other studies have shown that biomass growth rates increase up to 37 °C (Sterner & Grover, 1998), (Tedesco & Duerr, 1989). With a depth of approximately 20 cm, the algae culture was shaded by the 0.75 meter high walls of the pond. Studies on the effects of light on algae growth and lipid production indicate that triglyceride productivities in some species increase up to and beyond light intensities which may be expected in an outdoor algae pond (Tedesco & Duerr, 1989). A culture set up in similar conditions would probably be more productive if shading was limited and the temperature was elevated.

7.1.1.2 Lipid Composition

The most common fatty acids observed in the batch-mode pond were 16 and 18-carbon, saturated and monounsaturated fatty acids. These fatty acids are very similar in structure to those produced by more conventional oil crops such as soy and palm, which are well suited for liquid fuel production (Cruz, 1997).

The monounsaturated 18-carbon fatty acid observed in the algal culture is commonly referred to as oleic acid. The saturated 16-carbon fatty acid observed in the culture is commonly known as palmitic acid. Together, oleic acid and palmitic acid also constitute the bulk of palm oil. Soybean oil is primarily made up of a variety of unsaturated, unbranched eighteen-carbon fatty acids (Cruz, 1997). The similarity of the oils produced in the batch-mode algae pond to those of conventional oil crops suggests that a fuel product derived from wastewater algae might be characteristically similar to commercially available biodiesel fuel.

The tendency of the batch-mode culture to produce shorter fatty acid chains toward the end of the growth cycle has implications from a fuel production standpoint. This tendency in a fast-growing crop such as algae may make it possible for producers to have some control over the types of fatty acids grown in their systems. By manipulating the mean cell residence time of an algae pond, it may be possible to favor either longer or shorter carbon chains in the fatty acids produced. Batch-mode systems may be harvested at a certain time to target specific fatty acids. Semi-batch and continuous systems might be operated at different residence times to accomplish the same goal.

7.1.1.3 Discussion of Analytical Methods

Lipid concentrations were determined by two different methods throughout the lipid production experiment. Much higher fatty acid contents were measured using the chromatographic method compared to the gravimetric method. However, the observed trends in lipid production were similar for both methods.

Due to the highly sensitive nature of gas chromatography and the low-temperature conditions in which the chromatographic samples were prepared, it is likely that the results of the chromatographic testing are more accurate than those of the gravimetric testing. Unlike the extractions for the chromatographic method, the gravimetric method included a heating step, which is a potential cause of the lower lipid concentrations compared with the chromatographic method. The heating of the gravimetric samples to a relatively high temperature (105 °C for 1 hour) may have caused some of the fatty acids to vaporize, although this temperature is not uncommon in Bligh and Dyer-based procedures.

The observation of a discrepancy between the gravimetric and chromatographic methods of oil determination is an important outcome of the current research. The lower temperature processing of the chromatographic samples and the analytical accuracy of mass spectrometry suggest that the chromatographic method provides a more realistic measure of the oil content of a sample. However, more research is needed to refine these methods and reconcile the discrepancy.

7.1.2 Extraction Procedure Experiment

This thesis research produced results regarding the ability of several alcohol-based solvent systems to remove lipophilic mass from algae cells. It also provided new insight into the masses of triglycerides that may be extracted using each of the four solvent systems tested.

7.1.2.1 Extractable Mass

Of the four solvent systems examined, the methanol-water-chloroform (Bligh and Dyer) extracted the most mass from the algae samples, followed by methanol-hexane, ethanol-hexane and isopropanol-hexane. The mass extracted increased with increasing solvent polarity in the alcohol-hexane solvent systems.

Chloroform is very nonpolar compared to the methanol-water mixture against which it is partitioned in the Bligh and Dyer extraction. Similarly, hexane is very nonpolar compared to methanol and so it attracts algal triglycerides strongly. Because isopropanol and hexane do not have as large a difference in polarities, the isopropanol system did not extract as much material from the algae as the methanol system.

7.1.2.2 Fatty Acid Affinity

The mass of fatty acids extracted by each solvent system, determined by chromatography, differed from the total mass of lipids by the gravimetric method. The Bligh and Dyer extraction removed the most fatty acids. However, among the alcohol-based extractions, the isopropanol system removed the most fatty acids (83% of Bligh and Dyer) and the methanol system removed the least (23% of Bligh and Dyer).

The alcohol-hexane extractions included no sonication step, while the Bligh and Dyer extraction did. The extractions, therefore, tested not only the solvents' ability to partition triglycerides efficiently, but also their ability to remove triglycerides from undisrupted algae cells. Methanol is known to be a more effective destructor of cell walls better than ethanol or isopropanol (Henriques, Silva, & Rocha, 2007).

Although the isopropanol solvent system has a lower partition coefficient for triglycerides than the methanol solvent system, the isopropanol was able to remove more triglycerides from the algae because of its relatively high affinity for triglycerides.

8 References

Abeliovich, A. (1986) Algae in wastewater oxidation ponds. In: Richmond, A. (Ed.), Handbook of Microalgal Mass Culture. CRC Press, Boca Raton, FL

ASCE (2005). 2005 Report Card for America's Infrastructure: Wastewater. American Society of Civil Engineers. <http://apps.asce.org/reportcard/2005/page.cfm?id=35>

Becker, E.W. (2006). "Micro-algae as a source for protein." *Biotechnology Advances* **25** (2) : 207-210.

Craggs, R. J., Heubeck, S., and Shilton, A. (2007). "Influence of CO_2 Scrubbing from Biogas on the Treatment Performance of a High Rate Algal Pond." National Institute of Water and Atmospheric Research. Hamilton, New Zealand.

EPRI (1994). Energy Audit Manual for Water/Wastewater Facilities, Palo Alto, California, Electric Power Research Institute.

Table 6: Water quality data for the second segment of Experiment III, 1/9/2010-3/15/2010

Constituent	Influent	Pond 1 (w/ inoculum)	Pond 2 (w/ inoculum)	Pond 3 (control)	Pond 4 (control)
TSS (mg/L)	54	129	138	137	108
Std. Error	7.2	20.0	26.9	23.0	17.8
# of samples	12	12	12	12	12
VSS (mg/L)	46	112	122	111	87
Std. Error	4.6	17.6	23.7	18.7	11.8
# of samples	11	11	11	11	11
2-hr TSS (mg/L)	n/a	31	20	38	50
Std. Error	n/a	4.6	4.4	8.5	11.1
# of samples	n/a	7	7	7	7
24-hr TSS (mg/L)	n/a	22	14	33	40
Std. Error	n/a	3.5	5.7	9.8	10.1
# of samples	n/a	7	7	7	7
NH_x-N (mg/L)	30	4.2	5.1	6.4	6.7
Std. Error	0.7	1.2	0.9	0.8	1.0
# of samples	76	9	9	9	9
% removed	n/a	86%	83%	79%	78%
NO_3-N (mg/L)	0.8	7.4	3.3	3.8	2.3
Std. Error	0.4	2.1	1.1	1.4	0.5
# of samples	7	7	7	7	7
NO_2-N (mg/L)	0.3	1.8	1.5	1.0	0.7
Std. Error	0.1	0.5	0.4	0.3	0.1
# of samples	7	7	7	7	9
Alkalinity (mg/L)	614	431	454	470	471
Std. Error	79.4	57.1	56.0	61.2	58.3
# of samples	6	6	6	6	6
% removed	n/a	30%	26%	23%	23%

Feffer, A.S. (2007). *The Effects of CO2 Addition on Algae Wastewater Treatment*, M.S. thesis, Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo.

Frost, Daniel T. (2008). *Bioflocculation for control of wastewater pond microalgae,* M.S. thesis, Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo.

García, J., B.F. Green, T. Lundquist, R. Mujeriego, M. Hernández-Mariné, and W.J. Oswald (2006) "Long term diurnal variations in containment removal in high rate ponds treating urban wastewater." *Bioresource Technology* **97** : 1709-1715.

Lundquist et al. (2009). *A Realistic Technology and Economic Assessment of Algae Biofuels*, report for Energy Biosciences Institute, Berkeley, Calif.

Metcalf & Eddy (2003). Wastewater Engineering, Treatment and Reuse. New York, NY, McGraw-Hill Companies Inc.

Oswald, W.J. (1988). "Microalgae for wastewater treatment and Large-scale algal culture systems (engineering aspects)". In: Borowitzka, M.A.; and Borowitzka, L.J. (eds). Micro-algal biotechnology, pp. 305-329; 357-395, University Press, Cambridge, Cambridge.

Oswald, W.J. (1990). Advanced Integrated Wastewater Pond Systems. ASCE Convention, San Francisco, CA.

Oswald, W.J. (1991) "Introduction to Advanced Integrated Wastewater Ponding Systems." *Water Science and Technology* **24** (5) : 1-7.

Oswald, W.J. and H.B. Gotaas (1957) "Photosynthesis in Sewage Treatment" Paper presented before the Sanitary Engineering Division, American Society of Civil Engineers, New York, N.Y., (October 1954). Reprinted in Transactions of the American Society of Civil Engineers, Volume 122, (1957).

Pavoni, J.L., M.W. Tenney, and W.F. Echelberger Jr. (1972) "Bacterial exocellular polymers and biological flocculation." *Water Pollution Control Federation* **44** (3) : 414-429.

Woertz, I.C. (2007). *Lipid Productivity of Algae Grown on Dairy Wastewater as a Possible Feedstock for Biodiesel*, M.S. thesis, Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo.

Seawater/Wastewater Production of Micro-algae-based Biofuels in Closed-loop Tubular Photobioreactors and Microalgae Production in a Vertical Photobioreactor

Project Investigator:

Brian Hampson, Ph. D. and Ilhami Yildiz, Ph. D.
Food and Nutrition Sciences, and BioResource Agriculture Engineering Departments
California Polytechnic State University
San Luis Obispo, CA

Abstract
Microalgae will be one of the most desired feedstocks for biofuels production in the future. Theoretically, biodiesel and ethanol produced from algae appears to be the only feasible solution today for replacing petrodiesel completely by a renewable source. No other feedstock (soybean, corn, rapeseed, or palm oil) has the oil yield high enough to be in a position to produce such large volumes of oil. High productivity levels can be reached by growing algae in transparent closed tubular systems - so called photobioreactors (PBRs). Since the technology is still new, further improvements need to be developed and evaluated before PBRs can be used for large-scale commercial applications. This research is part of a long-term applied-research program at California Polytechnic State University (Cal Poly). Task-1 of the research project focuses on the nutrition side of the algae cultivation by trying to find alternative, natural, and inexpensive sources of algae nutrients (e.g. seawater or wastewater). Task-2, on the other hand, focuses on the microalgae production efficiency in a new vertical PBR. Algae in PBRs are typically grown with artificially produced nutrients (nitrogen, phosphorous, trace nutrients, etc.), which are very expensive and affect the feasibility of such systems significantly. Research efforts in fulfillment of Task-1 and Task-2 were conducted on either a lab-scale basis (up to 20-L) or using Cal Poly's pilot-scale horizontal and vertical PBRs, respectively. Initial results indicate that seawater and wastewater can serve as significant sources of available micro- and macronutrients. In addition, closed-loop PBR systems can be utilized to produce large quantities of biomass in a reliable and cost-effective manner. The research findings are being communicated to the scientific community and the applied algae-growing industry.

Introduction
As increased use of fossil fuels continues to impact on climate and oil reserves continue to decline, there is an increasing need to find alternative and sustainable sources of energy. With the recent rises and fluctuations in oil prices, there is renewed interest in production and use of fuels from plants (Gomez et al., 2008), which is truly renewable and environmentally friendly (low carbon emission, carbon neutral or carbon negative). Biofuels are produced from several sources ranging from primitive plants (microalgae) to fast-growing trees. At present, annual crops make the largest contribution to biofuels (Karp and Shield, 2008). However, crops' use as biofuels has been recently criticized since these crops take millions of hectares of land out of food production thus push up the food prices (Anonymous, 2007). In contrast, microalgae based biofuel is put forward as the most efficient biofuel in terms of land use and energy conversion (Chisti, 2007), but its potential is yet to be realized.

As per the cover story in the April edition of R&D Magazine (2008), the major categories for energy research and development are listed below. In this survey of over 1300 readers, the order is ranked from what is perceived as most significant to least significant. Percentages represent the respondents in agreement that the technology listed is where the U.S. should focus on getting its energy.

1.	Solar, photovoltaic	74%
2.	Solar, wind	70%
3.	Solar, thermal	61%
4.	Fuel cell	54%

5.	Nuclear, fission	50%
6.	Geothermal	50%
7.	Nuclear, fusion	47%
8.	Tidal	47%
9.	Improved coal efficiency	42%
10.	Hydrogen fuel sources	41%
11.	*Biofuel* process improvements	38%
12.	*Biofuel* feedstock improvement	38%
13.	Hydroelectric	33%
14.	Improved petroleum efficiency	30%
15.	*Biofuel* genetic development	28%
16.	Improved natural gas efficiency	28%
17.	Electrochemical	25%

One reason biofuels might rank low on the list is that too many people perceive biofuels as competing with food as a crop. As food prices steadily rise, and at an increasing rate, it is evident that ethanol or biodiesel from food crops is a short-term expediency. As production processes are developed for utilizing other sources of biomass the pressure on food production will be lessened. Biomass from algae provides such a solution (Gross, 2008).

Algae need CO_2, light, water and mineral nutrients for growth. With the exception of mineral nutrients, most of these requirements are low-cost or free. In cultivation of algae, mineral nutrients can be sourced from commercial fertilizers (Valenzuela-Espinoza et al., 1999), but their continued use in the long run may increase the cost of biofuels. The use of commercial fertilizers such as phosphorus in production of biofuels would also put further pressure on phosphorus reserves that are becoming scarce worldwide. It has been estimated that macronutrient requirements for every ton of dry algae biomass are 3, 11, 12, 18 and 81 kg of magnesium, phosphorus, calcium, potassium and nitrogen, respectively (AlgaeLink, 2007). The values for micronutrients are 0.004, 0.07, 0.30, 0.36 and 0.73 kg of molybdenum, copper, manganese, zinc and iron, respectively. These figures suggest that vast amounts of fertilizers may be required if microalgae-based biofuel system were to make a significant contribution to fuel production. In fact, the cost and/or availability of fertilizers can be the bottleneck to biofuel production in the future. Therefore, we need to fully explore alternative sources of mineral nutrients such as wastewater and ocean water for algal growth and biofuel production. The use of wastewater, high in nitrogen and phosphorus, is currently practiced in many countries (United States, Europe, Japan, Australia, and South Africa). The potential of ocean water alone or mixed with wastewater has also been investigated (Goldman et al., 1974; Craggs et al., 1994). However, these studies have mainly investigated removal of phosphorus and nitrogen by algae from wastewater and ocean water, and other nutrients have hardly been reported. It is important to recognize the other nutrients and their effects on algal growth and their cost to algal production.

Typically, when referring to biofuel from algae it is the lipid fraction of the plant cell that is of most interest. The lipid or oil derived from algae is triglyceride. However, other fractions of the algal cell biomass are considered to have commercial value.

Despite the wide biodiversity in microalgae species (Tomaselli, 2004), so far only a handful of species has been investigated for their antioxidant properties and nutritional content, such as proteins and lipids (Natrah et al., 2007), and their beneficial effects have already been demonstrated. Although algae species are known to be high sources of lipid, very little has been reported on their growth characteristics. A large scale screening of microalgae species may reveal species with greater growth rate and lipid content thus would add to their commercial value. For example, identification of microalgae, such as *Dunaliela salina*, that can grow in saline conditions could utilize unlimited water sources (ocean water or seawater). However, ocean water contains generally low concentrations of mineral nutrients (Latham et al., 1996; Yeats, 1998) and may need to be fortified by other sources of nutrients such as wastewater or effluent water. This would increase concentrations of some of the nutrients required for optimum growth with little cost. This would in turn reduce the amount or use of commercial fertilizers in mass production of microalgae. This is considered more cost-effective, sustainable and environmentally friendly than the use of commercial fertilizers only as the source of mineral nutrients.

Photoautotrophic algae can be propagated by two different systems, in PBRs (Figure 1) or in raceway ponds. Raceway ponds are less productive, use more arable land, and are more prone to change over time due to natural phenomena. In PBRs, if the nutrient requirements are met and the temperature is consistently maintained near an optimum level, then light represents the sole limitation to productivity (Richmond, 2004). In truth, there are many different PBR designs and the science of producing this algal "crop" in an economical fashion is in its adolescent stages of growth.

This study was an integral part to a number of applied research projects and a key contributor to reaching our long-term goal: Cost-effective production of algal biomass using PBRs under controlled environmental conditions. Using wastewater and/or seawater, we not only conserve further resources on the nutrition side (the need of artificial fertilizers and clean water disappears) with positive effects on the feasibility of the overall system, but also we work towards a more sustainable society. Wastewater treatment plants would be unburdened and wastewater would be integrated into a natural sustainable circle of biomass/energy production.

Improving the efficiency of PBRs and the manner in which nutrients are utilized are integral to the advancement of micro-algal biomass for biofuels. An efficient and low-cost source of biofuel has importance to our national security. Ships, trains, trucks, and jet aircraft all have one thing in common, they largely burn kerosene. Micro-algal derived biodiesel has the ability to serve as a replacement fuel for kerosene (petroleum) diesel (Antoni, et al., 2007; Daggett et al., 2007; Danigole, 2007; Future Fuels Aviation, 2008; Gas2.0, 2008; Green Options, 2007; Miller, 2007; NREL, 1998; Oilgae, 2007; PR Newswire, 2008; Rhodes, 2007; Schenk et al., 2008; Strategic Technology Office, 2006; Vasudevan et al., 2008).

Project Objectives
Task-1: Seawater/Wastewater Production of Microalgae

This series of experiments were conducted under controlled environmental conditions in the BRAE Lab. The specific objectives of Task-1 are: (1) To assess the potential of seawater and/or

dairy wastewater for algal growth; (2) to determine the nutrient requirements for optimum algal growth hence provide estimates for the cost associated with the use of fertilizers; and, (3) to try and identify microalgae species (among readily accessible species) suitable for low-cost production systems.

Task-2: Microalgae Production in a Vertical Photobioreactor

The short-term or immediate goal is to establish an infrastructure to perform algae PBR development. To facilitate the accomplishment of this goal, a vertical PBR was donated by our industry partner, Phycotech LLC (Sacramento, CA) and situated in the Cal Poly FSN Pilot Plant (Building 24 room 106) on campus. To accomplish the long-term goal of determining biomass productivity, the following objectives must be attained:

(1) Start-up and trouble-shoot the vertical PBR system (dry run, no algae). The PBR system and its ancillary equipment were checked for leaks (water and gas) and assured that all components function effectively as part of the whole. In the laboratory, initial cultivation of algae (*Chlorella vulgaris*) for scale-up was performed. An algal strain library is being established for archiving the cultures as we acquire them from other institutions or collaborators. All procedures developed for culturing algae are documented, recorded, and written as Standard Operating Procedures (SOPs).

(2) Introduction of an algal strain to the vertical PBR system. Culture scale-up and monitoring procedures were established for the system. Parameters to be monitored throughout the life-cycle of an algae growth experiment include medium formulation (added nutrients), algae culture density (as chlorophyll) determined by optical density at 660 nm, cell count determined by microscopy using a hemocytometer counting cell, temperature, and algae culture purity. From this information harvest periodicity can be established and biomass yields can be determined. All procedures developed will be documented, recorded, and written as SOPs.

(3) Optimize conditions (inherent to the vertical PBR system) for biomass productivity. Every single-cell culture undergoing cultivation goes through various sequential stages of growth and ultimately cell death. These four stages of a cell culture can be expressed as the initial lag phase, followed by rapid cell division in the logarithmic growth phase, followed by a slowing of growth during the stationary phase, and lastly senescence and the cell death phase. The ideal for biomass production is to maintain the logarithmic growth phase whereby algae are maintained at their maximum rate of cell division. How long this "log-phase" can be maintained is the ultimate question, and the duration may vary tremendously depending upon external environmental factors and intrinsic factors inherent to the system and the strain of algae employed. The lag phase can be minimized through effective scale-up of the cell culture and it having high cell density prior to introduction into the PBR. The cells should never get to the death phase because periodically algae are harvested from the PBR; consequently they never have a chance to grow old. Also during this part of the project, procedures were developed for sampling the system and evaluating the samples for strain purity. A major concern here is that a bacterial contaminant or non-desirable wild or invasive strain of algae is multiplying in the system. Standard bacterial plate count methodologies were employed for periodically testing the samples collected.

(4) Verify the ability to sanitize the PBR system *in situ* during cultivation and between "crops" of algae. The vertical PBR system, as depicted in Figure 1, has ultra-violet (UV; 185 nm) irradiation capability added-on, and it is the ozone generated by these UV lamps that function in a prophylactic manner to suppress the outgrowth of non-desirable viral and bacterial contaminating species.

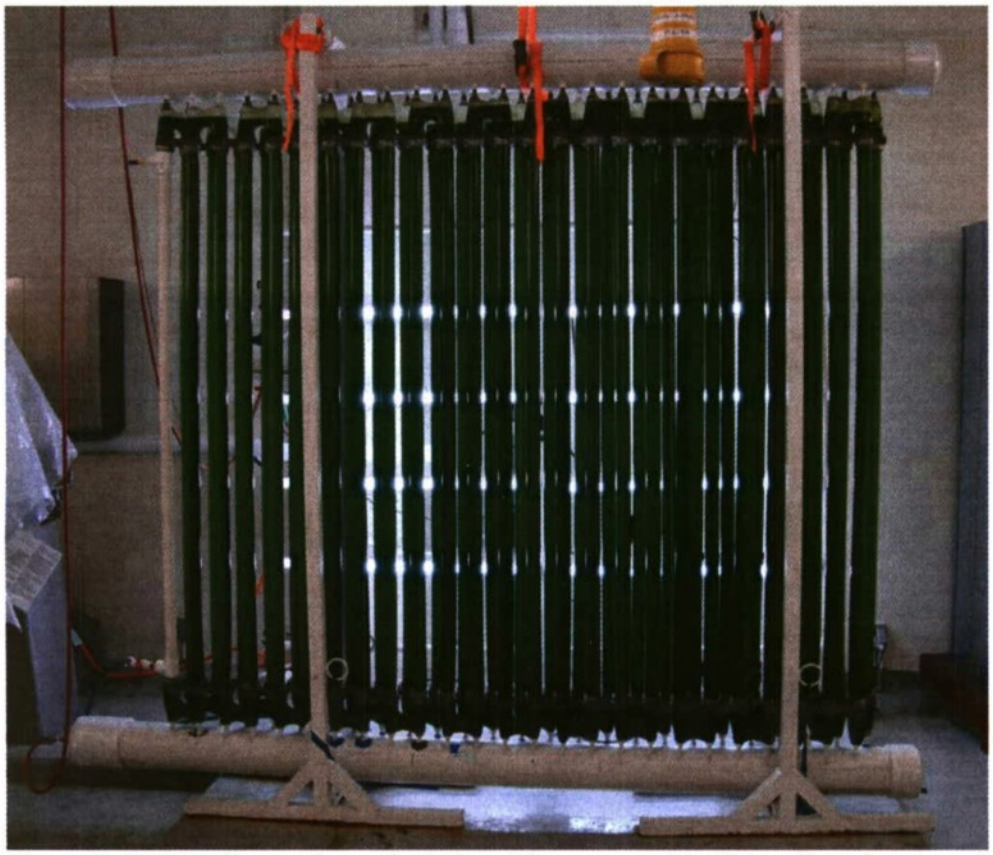


Figure 1: 200-L pilot-scale vertical photobioreactor with *Chlorella vulgaris* under cultivation.

Materials and Methods

The optical density of the algae in suspension was determined using a scanning spectrophotometer (Ocean Optics Redtide USB650 or HACH DR3800) and absorbance values at 660 nm (or 665 nm) were recorded over time. Enumeration of algae cells, expressed as cells per milliliter, were determined using a hemocytomer slide (Bright-Line Counting Chamber) viewed at 400x magnification (Nikon or Olympus microscope with Infinity digital camera). Bacteria were enumerated using Aerobic Plate Count Petrifilm (3M Microbiology) incubated at 35°C for 48-hr. Standard microbiological procedures were employed throughout.

The *Chlorella vulgaris* culture (Carolina Biological Supply) was cultivated in a medium consisting of deionized water with 0.75-gr nutrients added per liter (Schultz 10-15-10; Spectrum Brands). Scale –up of this culture was performed at ambient indoor temperature with illumination provided by 6500K fluorescent lights. A test tube containing approximately 20-ml of culture was transferred to 200-ml medium and propagated for several days. The culture was next transferred into 2000-ml (2-L) of fresh medium and cultivated for several days, or until the cell count reached at minimum 1×10^7 *Chlorella* per ml. This 20-L culture is subsequently used for benchtop research trials or used as the inoculum for the 200-L vertical PBR. Throughout the culture scale-up procedure, Carbon Dioxide (CO_2) was diffused into the growth medium to maintain a pH value of 6.8 to 7.0. A pH monitor/controller (Milwaukee Instruments model SMS122) was used to control the dosing of CO_2.

Ozone gas was produced using a DEL Ozone Model ZO-156 ozone generator on a clock timer. Ozone gas was introduced into the culture vessel via an air-stone in 15-min increments 24-hrs per day. The output of this generator is 60 mg per hour, or 150 ppm at 12 cubic feet per hour (temperature dependent). Various doses were evaluated and experiments are still underway to determine the optimal dose regime. Optimal dose regime is defined as the amount of ozone (mg/L/hr or ppm/hr) required to suppress bacterial growth a minimum of two log-fold over a negative control culture with no concurrent diminishment in the ability of the algae to replicate and increase in number.

Dry biomass determinations were determined gravimetrically using 1-L algae culture samples. These were dried in a pilot-scale convection dehydration cabinet set to 150°F with air velocity set on the high setting. About a six-hour duration, or overnight, was required to completely dry a sample.

Results and Discussion

The potential of seawater and/or dairy wastewater for algal growth:

There are many acceptable laboratory techniques used for measuring microalgae biomass under ideal conditions. These ideal conditions usually require relatively pure water and very little microorganism contamination in the samples, which would overestimate the desired biomass. This presents a dilemma when current research is exploring the potential of growing microalgae under abnormal conditions, such as utilizing wastewater as a growth media. Wastewater – while beneficial to microalgae growth – has tremendous potential to introduce contamination with bacteria, microalgae, and other microorganisms. Additionally, wastewater introduces organic/inorganic solids and adds turbidity to the sample.

Determine the nutrient requirements for optimum algal growth:
Some of the traditional microalgae quantification methods include optical density, cell counting, measuring chlorophyll-a, and dry weight. The primary limitation of the optical density and dry weight methods are the introduction of non-algae wastewater solids that artificially inflate the measured values. These solids are difficult to distinguish from microalgae cells thereby yielding false measurements. The cell counting method enables researchers to visually distinguish between wastewater solids and microalgae cells, as well as quantify contaminants in the sample. However, the greatest challenge with this method is the time requirement to manually count cells and human error. The chlorophyll-a method utilizes methods to isolate and measure the chlorophyll-a content of a sample, and thereby estimate the algal biomass. The challenge with this method is obtaining experimental precision and the fact that the method relies upon an assumption of the chlorophyll-a content of an average algae cell. Collectively these limitations present a challenge to reliably quantifying microalgae biomass when growing under wastewater conditions.

Identification of microalgae species (among readily accessible species) suitable for low-cost production systems:
This study observed microalgae growth of two freshwater (*Chlorella vulgaris, Scenedesmus* sp.) and two saltwater microalgae strains (*Tetraselmus* sp., and *Nannochloropsis* sp.) under wastewater conditions. All four abovementioned methods of biomass quantification were used; optical density, cell counting, chlorophyll-a, and dry weight. Optical density was measured at 665 nm wavelength, and dry weight was measured using traditional total suspended solids and total volatile solids techniques. The experimental design incorporated four concentrations of wastewater: 33%, 25%, 10%, and 0% (as a control). Including both wastewater and non-wastewater samples enabled a distinction to be drawn about the interferences associated with wastewater use as compared to pure water.

The results of these investigations using either freshwater or seawater microalgae supplemented with wastewater will be completed and presented in a Master of Science thesis. The writing of this thesis is currently in-progress and will be completed prior to the end of fall quarter, 2010. Mr. Josh Lowrey, graduate research assistant, has completed his practical research and is culminating all findings in the written thesis.

Receipt, assembly, and initial start-up of the 200-L vertical PBR:
The 200-L vertical PBR was delivered in January 2010 and immediately assembled and tested. It was decided to not add-on the automated control system at this time, or until it is determined exactly what parameters are required for optimal performance. Once we have growth parameters and kinetics established over time, then parameters such as the addition of nutrients (Nitrogen, Phosphorous, Potassium, micronutrients, and CO_2), flow rate, harvest cycle, etc. can be programmed into the PLC unit (Siemens).

The PBR system (Figure 1) can best be described as a serpentine loop of connected individual PBR tubes each containing approximately 4-L of culture. Every other tube has a high-gas-flow diffuser head in the bottom "U", and the remaining every other tube has a low-gas-flow diffuser head in the bottom "U" connector. Gas (compressed air) is introduced through the bottom manifold to a pressure of 3- to 5-psi (pounds per square inch) and is distributed evenly throughout the bottom manifold to all tube diffuser heads. CO_2 is controlled by pH (lack of CO_2 causes a rise in pH value) and is concurrently introduced into the bottom manifold as required

via an in-line solenoid valve. The top manifold serves as a gas outlet and provides support to the tubes keeping them perpendicular and vertical to the upper and lower manifolds. The entire structure is supported, or suspended, by two metal stands, one on either end. All components coming in contact with the culture are composed of injection-molded, recyclable, plastic components. Low cost and light weight components helps bring the cost of the PBR system very reasonably low compared to other PBR systems.

After a series of mechanical issues such as leaking of culture medium through fittings or through offset diffuser heads, irregularly positioned tubes, and other operational issues, the 200-L vertical PBR eventually performed well and the project focus shifted to data collection versus PBR maintenance and correction.

Culture scale-up and monitoring procedures for introducing algae into the vertical PBR:

Figure 2: Culture scale-up (to 20-L) and benchtop laboratory experiments.

An example of *Chlorella* under cultivation in the laboratory is seen in Figure 2 above. In this instance, the culture on the left is receiving periodic additions of ozone gas whilst the culture on the right is serving as a negative control.

Optimization of conditions (inherent to the vertical PBR system) for biomass productivity:

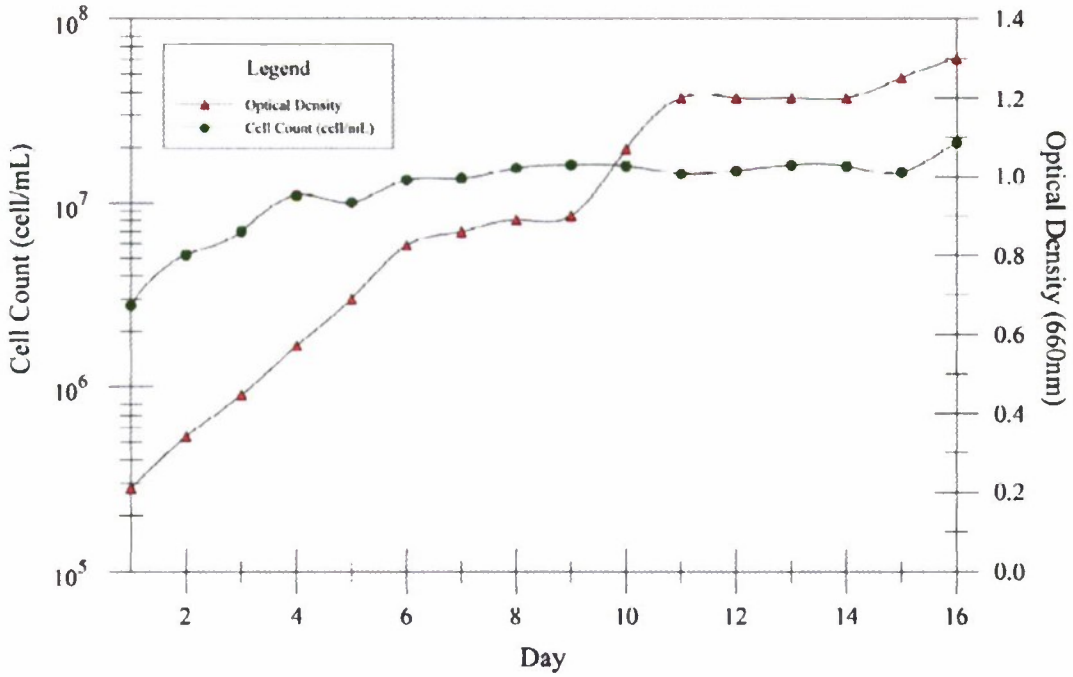


Figure 3: Typical growth curve for *Chlorella vulgaris* (20-L culture presented).

This is an example (Figure 3) of a typical growth curve for *Chlorella* over a 16-day period. The graph does not begin with a cell count of zero because the scale-up procedure involves a 10% inoculum as the culture is transferred from 2-L to 20-L to 200-L. Hence, culture is transferred when the cell count reaches approximately 1 x 10^7. Both the actual cell count (determined microscopically using a hemocytomer slide) and the optical density readings (A660 nm) are presented. As seen at circa day-10, the optical density values rise above the cell count values because as the cells mature, they "grow fatter" and accumulate more chlorophyll per unit cell number. Not seen in this graph are the observations that after day-16, the cell count plateaus at about 1.0 x 10^8, and we have attained stationary growth phase. For optimal scale-up results it is best to transfer the culture before the stationary growth phase. Given additional illumination, additional micronutrients, and/or better temperature maintenance (other than ambient conditions), it is anticipated that the cell count may rise higher, possibly into the 1 x 10^9 range.

After several replicate growth trials, the basic growth characteristics were established for *Chlorella vulgaris* (under the conditions of our laboratory environment). Subsequent experiments for determining biomass and experiments involving the alteration of the gas mixture introduced to the growing culture (ozone) did proceed (see Figure 4 below).

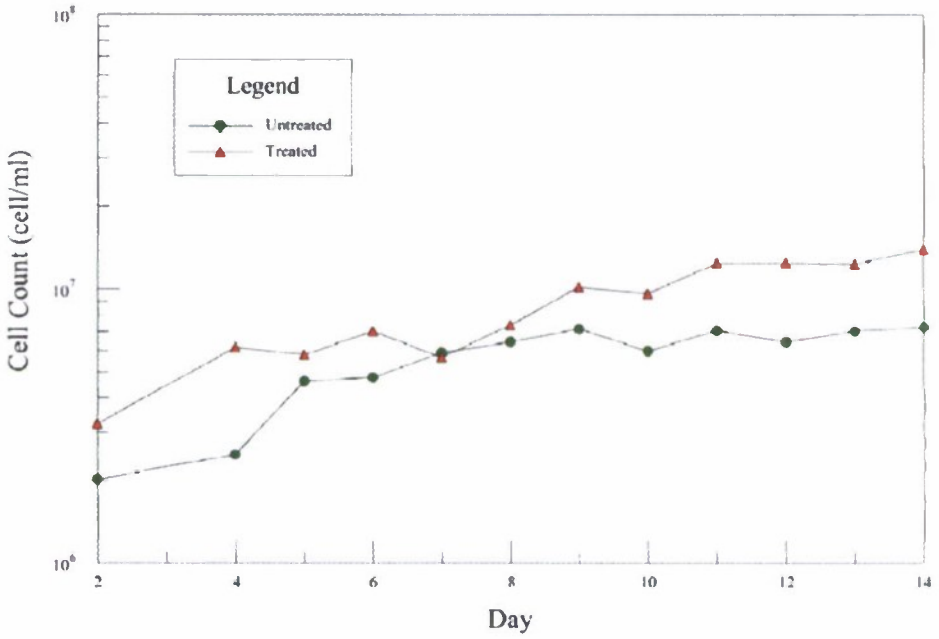


Figure 4: Growth curve (partial; days 2 through 14) for *Chlorella* grown with the addition of ozone gas (Treated -Δ-) and without ozone gas (Untreated -O-).

Several experiments were performed investigating the addition of ozone gas into the culture medium during cultivation of *Chlorella*. Figure 4 is an example of one of these growth cycles at the 20-L benchtop scale. The significant finding here is that under conditions of the test, *Chlorella* is affected minimally, if at all, by the addition of ozone gas. The preliminary result of these experiments indicate that 1.5-mg ozone per liter per hour is minimally sufficient for bacterial control (see Figure 5 below), however as mentioned above, the optimal dose regime is still undetermined and under investigation. These investigations are being replicated at the 200-L level in the vertical PBR using a newly assembled ozone gas generator that will soon be incorporated directly into the lower gas manifold of the PBR.

One added benefit to the addition of ozone gas, even if at sub-optimal levels for contaminant control, is the enhanced flocculation effect seen in the treated cultures. Typically, *Chlorella* cells are so small and buoyant they tend to remain suspended in solution over time. The addition of ozone gas to the medium during active growth causes the algal cells to more readily clump together. This flocculation effect results in fewer cells having the ability to remain in suspension over time under ambient conditions. In short, an enhancement of harvest capabilities may be a side-benefit to using ozone gas for prophylactic microbial contamination control. This is very significant and experiments are underway to further understand and control this flocculation phenomenon.

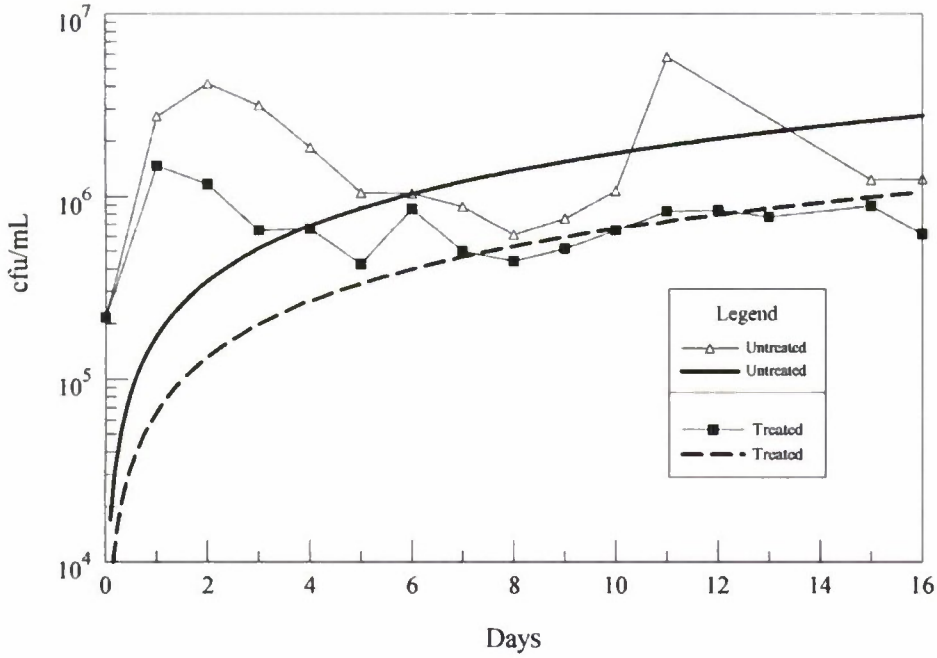


Figure 5: Bacterial plate counts (colony-forming units per milliliter; cfu/ml) in Chlorella culture treated with ozone gas over time (Treated -■- and Untreated -Δ-). Bold solid and bold dashed lines are first-order regression curves for the growth cycles.

Due to the configuration of the equipment at the benchtop level, ozone gas was introduced intermittently in 15-min increments. This works fine at the 20-L level, however, for pilot- and commercial-scale cultivation ozone will be introduced in a continuous manner through the lower manifold along with the compressed air and CO_2.

The goal is to have bacterial contamination eliminated, or at best controlled to a reasonable level. A reasonable level of control is considered a two log-fold reduction over a negative-control culture with no concurrent effect on the algae. The reason this is possible is because algae, both as plants and as higher life forms, are less susceptible to the germicidal effects of ozone gas in solution. Anecdotal evidence reveals that to control algae it requires approximately 100-times the dose of ozone over that required for control of bacteria and viruses. This information was obtained through discussions held with industry experts knowledgeable in the treatment of swimming pools and spas using ozone for microbial control. With that baseline information in-hand, the assumption was made that there is at least two orders of magnitude leeway in dosing ozone to control bacterial outgrowth before causing harm to the algae undergoing cultivation. The confirmation of this hypothesis has yet to be fully evaluated or proven, but positive results are being obtained through replicate experimental trials.

Enhancing the ability to grow pure algal cultures in a PBR is highly desirable, but the ultimate end-goal remains biomass productivity. From a commercial productivity perspective, the more microbiologically pure the biomass can be, the better. One observation made during these experiments was that the number of aerobic bacteria growing concurrent with the algae is quite

significant; there are approximately as many bacteria growing and utilizing resources as there are algal cells being cultivated. One might consider these the "weeds" in the PBR. The value of 0.3 grams dry weight *Chlorella* per liter was determined when the algal culture was at a cell density of 1.0×10^7. Cell densities of 5.0×10^7 have routinely been obtained, and once the operational parameters of the PBR are optimized, it is anticipated that cells counts of 1.0×10^8 or higher will become the norm.

g/L Dry Biomass	Algae Cell Count	Harvest Cycle	Daily Yield	365-Day Yield
0.3	1.0×10^7	15% per day (30-L)	9 grams	3.285 Kg
1.5	5.0×10^7	30% per day (60-L)	90 grams	32.85 Kg

Table 1: Biomass productivity of *Chlorella* cultivated in the 200-L vertical PBR under controlled environmental conditions (indoors).

Harvest cycle, or the amount of culture that can be drawn-off on a time-schedule basis, is ultimately determined by cell replication time and other relevant contributing factors. If a culture is able to double its cell count (biomass) every three days, then approximately 30% of the culture volume can be harvested with no immediate impact on the entire culture. Under optimal conditions, the time required for doubling of the biomass may be much more shorter than three days (Donnan et al., 1985). The values presented are conservative and not exaggerated.

Most studies express biomass yield in units of either pounds per acre (English), or kilograms per hectare (ha; metric). The following table is an extrapolation of the resultant values obtained through experimentation in the laboratory.

PBR Tubes/ha	L/ha	Harvest Cycle	Daily Yield (at 1.5 g/L)	365-Day Yield (per ha)
201,700 (1-m spacing)	806,950-L	30% per day (242,100-L)	363.10 Kg	132,540 Kg
134,500 (1.3-m spacing)	538,000-L	30% per day (161,400-L)	242.10 Kg	88,350 Kg

Table 2: Extrapolation of productivity values to annual yield of dry biomass per hectare per year.

In English units, the above results are expressed as a yield of 118,050 pounds per acre per year with 1-m spacing between rows of vertical PBR tubes. Or, with tube spacing of 1.3 meters the biomass yield can be expressed as 78,691 pounds dry biomass per acre per year.

For comparison, based on average national yield, the value for corn is 900 pounds per acre per year, and if the stover is added in as fermentable material, the yield increases by 400 pounds to 1300 pounds biomass per acre per year (Ceres, 2010). Corn can then be considered 1.65% as efficient for producing biomass compared to algae [(1300 / 78691) x 100 = 1.65%].

Verification of the ability to sanitize the PBR system *in situ* during cultivation and between "crops" of algae: These experiments are in-progress and until the pilot-scale ozone generating system is fully developed, results cannot be presented. An ozone generator was recently

fabricated and a housing is under construction followed by plumbing the ozone gas stream into the lower gas distribution manifold of the 200-L vertical PBR.

Since a CSU-ARI matching grant was awarded for a 3-year duration (2009-2012), it is possible to continue these investigations over the next year.

Conclusions

Seawater and wastewater were evaluated and characterized for nutrient content in support of algal growth. Species of algae assessed in these studies include two freshwater (*Chlorella vulgaris* and *Scenedesmus* sp.) and two saltwater microalgae strains (*Tetraselmus* sp. and *Nannochloropsis* sp.). The ability to reliably quantify microalgae biomass when growing under wastewater conditions was evaluated and a combination of several procedures can give consistent results.

The 200-L vertical PBR is operational and now being used for the optimization of microalgae biomass using *Chlorella vulgaris* as the initial test species. A dry biomass yield of 1.5 mg/L was determined at a cell density of 5.0×10^7 algae per milliliter. This value is very much in concurrence with those obtained through others' research.

The use of ozone gas is under evaluation for prophylactic control of bacteria and other non-desirable microorganisms during algal cultivation. This will eventually result in the ability to produce a purer biomass, and hopefully higher yields with equivalent nutrient input.

This project was of great assistance in providing infrastructure and laboratory support during these first years of algae photobioreactor research and development at Cal Poly.

References

AlgaeLink. 2007. www.algaelink.com.

Anonymous. 2007. Climate Change: Looming Global Food Crisis. Africa Research Bulletin 17622-17623.

Antoni, Dominik, V.V. Zverlov, and W.H. Schwarz. 2007. Biofuels from Microbes. Applied Microbiology and Biotechnology Vol. 77: 23-35.

Ceres. 2010. www.ceres.net/AboutUs/AboutUs-Biofuels-Advantage.html.

Craggs, R., J. McSuley, V.J. and Smith (1994). Batch Culture Screening of Marine Microalgal Nutrient Removal from Primary Sewage Effluent. Hydrobiologia 288: 157-166.

Cristi, Yusuf. 2007. Biodiesel from Microalgae. Biotechnology Advances Vol. 25: 294-306.

Daggett, David L., R.C. Hendricks, R. Walther, and E. Corporan. 2007. Alternate Fuels for use in Commercial Aircraft. The Boeing Company. All rights reserved.

Danigole, Mark S. 2007, December. Biofuels: An Alternative to U.S. Air Force Petroleum Fuel Dependency. Pentagon Reports. www.stormingmedia.us/34/3484/A348474.html

Donnan, Lorraine, E.P. Carvill, T.J. Gilliland, and P.C.L. John. 1985. The Cell Cycles of Chlamydomonas and Chlorella. New Phytologist Vol. 99: 1-40.

Future Fuels Aviation. 2008. www.future-fuels-aviation.com.

Gas2.0. 2008. How Green are Biofuels? Comparison Chart. http://gas2.org.

Goldman, J. C. and H.I. Stanley. 1974. Relative Growth of Different Species of Marine Algae in Wastewater-seawater Mixtures. Marine Biol. Vol. 28: 17-25.

Gomez, L.D., C.G. Steele-King, and S.J. McQueen-Mason. 2008. Sustainable Liquid Biofuels from Biomass: The Writing's on the Walls. New Phytologist Vol. 178: 473-485.

Green Options. 2007. Algae Biofuel may be Future for Aviation. http://claytonbiodiecornell.greenoptions.com/

Gross, Michael. 2008. Algal Biofuel Hopes. Current Biology Vol. 18(2): R46-R47.

Karp, A. and I. Shield. 2008. Bioenergy from plants and the sustainable yield challenge. New Phytologist doi: 10.1111/j.1469-8137.2008.02432.x

Latham, V., M. Rayner, R. Leeming, P. Nichols, and G. Cresswell. 1996. Nutrient concentrations in Sydney's continental shelf waters 1989-1993. *In*: Hydrology and Water Resources Symposium: Water and the Environment. Pp. 735-736. Barton, ACT: Institution of Engineers, Australia. National conference publication No. 96/05.

Miller, Vann. 2007. Official Shares Glimpse of Future with Misawa Airmen. Air Force Link. www.af.mil/news/story.asp?id=123043011.

National Renewable Energy Laboratory (NREL). 1998. A look back at U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae. National Renewable Energy Laboratory. Golden, Colorado, USA: John Sheehan, Terri Dunahay, John Benemann, Paul Roessler.

Natrah, F. M. I., F.M. Yusoff, M. Shariff, F. Abas, and M.S. Mariana. 2007. Screening of Malaysian Indigenous Micralgae for Antioxidant Properties and Nutritional Value. J. Appl. Phycol. Vol. 19: 711-718.

Oilgae. 2007. ASU Partners with UOP to Make Algal Biofuel for Military Jets. August 2 News Release. www.oilalgae.com/.

PR Newswire. 2008. First Heavy-Duty Vehicle Powered by 100% Algal Derived Soladiesel(TM) Demonstrated at the 2008 DESC Worldwide Energy Conference. April 15. www.prnewswire.com/

R&D Magazine. 2008. Cover Story: Are Biofuels the Right Choice? April, 2008.

Richmond, Amos (Editor). 2004. Handbook of Microalgal Culture – Biotechnology and Applied Phycology, Blackwell Publishing, Ames, Iowa.

Richmond, Amos. 2004. Principals for Attaining Maximal Microalgal Productivity in Bioreactors: An Overview. Hydrobiologia Vol. 512: 33-37.

Rhodes, Chris. 2007. Energy Balance: Shell to make Biofuel from Algae. http://ergobalance.blogspot.com/2007/12/shell-to-make-biofuel-from-algae.html

Schenk, Peter M., S.R. Thomas-Hall, E. Stephens, U.C. Marx, J.H. Mussgnug, C. Posten, O. Kruse, and B. Hankamer. 2008. Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. Bioengineering Research Vol. 1: 20-43.

Strategic Technology Office. 2006. Biofuels. DARPA BAA06-43. www.darpa.mil/sto/solicitations/biofuels/index.html.

Tomaselli, I. 2004. The microalgal cell. In Richmond A (Ed). Handbook of microalgal culture. Blackwell, Oxford, pp. 3-19.

Valenzuela-Espinoza, E., Millan-Ninez, R. and Nunez-Cebrero, F. 1999. Biomass Production and Nutrient Uptake by *Isochrysis aff. galbana* (Clone T-ISO) Cultured with a Low Cost Alternative to the f/2 Medium. Aquacultural Engineering Vol. 20: 135-147.

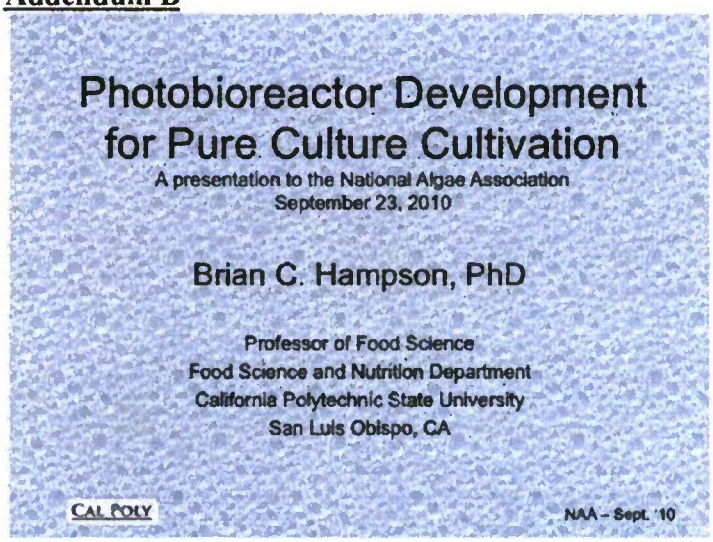
Vasudevan, Palligarnai and M. Briggs. 2008. Biodiesel Production – Current State of the Art and Challenges. Journal of Industrial Microbiology and Biotechnology Vol. 35: 421-430.

Yeats, P. A. 1988. The Distribution of Trace Metals in Ocean Waters. The Science of the Total Environment 72: 131-149.

Addendum A

Food Science Algae Research Team: (l to r) Christina Brox (Food Science Senior), Derek Fougere (Food Science Senior), Dr. Hampson (seated), Augie Cooper (Food Science Senior), and Stephanie Percival (Biology Junior) in front of the 200-L vertical photobioreactor prototype with *Chlorella vulgaris* under cultivation in the Food Science Pilot Plant.

Addendum B



Addendum C

Poster 31

PO-31 A Comparative Study of Different Analytical Methods for Quantification of Microalgae Biomass in Wastewater Growth Media

A. Josh Lowrey, Diana Durany, Gilbert Diaz, Ilhami Yildiz

BioResource and Agricultural Engineering Department,California Polytechnic State University,San Luis Obispo, CA 93407
E-mail: iyildiz@calpoly.edu

The recent interest in research and development of biofuels has generated a huge interest in many innovative feedstocks, which are the raw resources that are harvested mostly for their oil or sugar potential. Microalgae have recently been considered ideal for their tremendous oil content and their very high reproductive capacity. Along with the development of these unique feedstocks there have been uncertainties over how to accurately quantify biomass. The unique challenge associated with quantifying microalgae biomass is the fact that growth is conducted in suspension in water. Additionally there are uncertainties in distinguishing between various microalgae strains, as well as other microorganisms. Under some experimental conditions interferences arise when contaminated water or even wastewater is used as a growing media for the microalgae. These interferences challenge existing measurement methods for measuring biomass due to their inherent limitations.

There are many acceptable laboratory techniques used for measuring microalgae biomass under ideal conditions. These ideal conditions usually require relatively pure water and very little microorganism contamination in the samples, which would overestimate the desired biomass. This presents a dilemma when current research is exploring the potential of growing microalgae under abnormal conditions, such as utilizing wastewater as a growth media. Wastewater – while beneficial to microalgae growth – has tremendous potential to introduce contamination with bacteria, microalgae, and other microorganisms. Additionally, wastewater introduces organic/inorganic solids and adds turbidity to the sample.

Some of the traditional microalgae quantification methods include optical density, cell counting, measuring chlorophyll-a, and dry weight. The primary limitation of the optical density and dry weight methods are the introduction of non-algae wastewater solids that artificially inflate the measured values. These solids are difficult to distinguish from microalgae cells thereby yielding false measurements. The cell counting method enables researchers to visually distinguish between wastewater solids and microalgae cells, as well as quantify contaminants in the sample. However, the greatest challenge with this method is the time requirement to manually count cells and human error. The chlorophyll-a method utilizes methods to isolate and measure the chlorophyll-a content of a sample, and thereby estimate the algal biomass. The challenge with this method is obtaining experimental precision and the fact that the method relies upon an assumption of the chlorophyll-a content of an average algae cell. Collectively these limitations present a challenge to reliably quantifying microalgae biomass when growing under wastewater conditions.

This study observed microalgae growth of two freshwater (*Chlorella vulgaris, Scenedesmus* sp.) and two saltwater microalgae strains (*Tetraselmus* sp., and *Nannochloropsis* sp.) under wastewater conditions. All four abovementioned methods of biomass quantification were used; optical density, cell counting, chlorophyll-a, and dry weight. Optical density was measured at 665 nm wavelength, and dry weight was measured using traditional total suspended solids and total volatile solids techniques. The experimental design incorporated four concentrations of wastewater: 33%, 25%, 10%, and 0% (as a control). Including both wastewater and non-wastewater samples enabled a distinction to be drawn about the interferences associated with wastewater use as compared to pure water.

Utilizing statistical tools the methods were compared for their respective consistencies under various conditions. When high correlations are observed, then it is concluded that some methods can serve as indicators for the other, more time consuming method. Developing regression models might enable researchers to estimate the value of one analytical method based upon the result of another, without actually measuring both. Ultimately the comparison of methods under these outlined experimental conditions can help to determine which analytical methods are appropriate in future studies involving microalgae growth in wastewater conditions.

Keywords: Microalgae, cell count, chlorophyll-a, optical density, *Chlorella vulgaris*, *Scenedesmus* sp., *Tetraselmis* sp., *Nannochloropsis* sp.

Addendum D

Master of Science Thesis (M.S. candidate Mr. Josh Lowery) to be submitted at a later date.

Cal Poly Wind Power Research Center

Principal Investigator:

Patrick Lemieux, Ph.D., P.E.
Associate Professor Mechanical Engineering
California Polytechnic State University
San Luis Obispo, CA

Cal Poly Wind Power Research Center

Patrick Lemieux, Ph.D., P.E., Mechanical Engineering Department

Summary

This California Central Coast Research Partnership (C3RP) funded project initially consisted of the design and development of an in-house process for the manufacturing of wind turbine blades, based on utility-scale turbine manufacturer's standard practices. We increased its scope significantly after reaching important milestones ahead of schedule, so that it includes the full development of the first *complete turbine* for the 'Cal Poly Wind Power Research Center' (CPWPRC), a nascent facility that we started, in the Mechanical Engineering Department, for interdisciplinary use and collaboration across the university. Its primary purpose is to provide a physical platform for faculty and students to study technical problems that may be faced in the utility-scale wind turbine industry, and to prepare engineering students for careers in that industry. The new scope entails the siting study, as well as the design and manufacturing of every component of a fully functional research wind turbine, including: tower foundation, tower, nacelle, and rotor. The project officially started during winter of 2009; fabrication of major components ended in March of 2011; final completion of the entire assembly and commissioning of the turbine is expected to take place in the summer of 2011.

Background and Project Significance

With the increasing interest in sustainability, the wind power industry continues to become a larger and larger employer of engineers in North America. The bulk of growth may be expected to take place in the large (750kW and larger), utility-scale, 3-blade horizontal axis wind turbine (HAWT) market, where the development process is relatively mature. While research and development of smaller machines is necessarily easier, faster and cheaper, the leaps achieved in the utility-scale sector over the past decades have not trickled down to small HAWTs (less than 100kW): most small wind turbine manufacturers use techniques focused on manufacturing and retail cost, sometimes at the expense of efficiency and overall cost of energy produced, which are the primary focus of large HAWT manufacturers. Since the appropriate turbine size for a university research facility falls in the 'small turbine' category, off-the-shelf turbines that satisfy our needs do not exist.

The first goal of the Cal Poly Wind Power Research Center is to allow students to participate in the development of a small research wind turbine designed along guidelines comparable to those of the utility-scale industry, with the added constraint that the design must be able to allow for significant changes and experimentation, befitting a 'laboratory machine'. Once built, the turbine becomes a workbench for hands-on study of problems and best practices of the utility-scale turbine industry. To develop a wind turbine, the full system is first broken down into smaller, specific components; each element is then tackled individually, to accommodate budgetary constraints while addressing the overall system goal. The first element that we chose to tackle was the rotor blades, supported by our first year C3RP grant. This work was completed

in the fall of 2009, ahead of schedule and under-budget, motivating an increase in scope to include the entire wind turbine.
The following benchmarks were defined to help guide the development and design of this turbine:

1. Site study: To be successful, our facility requires maximum flexibility and safety, at a site where the wind resource is demonstrably consistent.
2. Rotor: the blade planform designed is suitable for both stall and pitch regulation. A hub capable of handling both control methods is a required component for the long term use of this rotor.
3. Nacelle: it must be designed to house a large variety of modular control system electronics and hydraulic actuator systems, and be capable of accommodating future changes in generators (*e.g.*, for studies asynchronous generators, geared systems, etc.)
4. Tower: the research turbine has specific requirements, including safety, ease of operation, and access, that make its tower a unique structure.

Results

The following sections highlight the progress achieved at the end of the C3RP award on each of the four steps listed above.

1. Site selection for a wind power research center at Cal Poly

A limitation affecting the selection of a proper location for wind turbines is the availability of land resource. Siting for this particular purpose is critical, not only to ensure appropriate *wind* resource (the quality of the wind resource is affected by proximity to obstructions and changes in the terrain) but also for safety and public acceptance issues (*e.g.*, noise abatement and visual impact), so that the center does not interfere with other campus activities. Cal Poly is uniquely positioned to provide several nearly ideal options for siting such a wind power research center, for several reasons:

1. ***Land availability.*** With nearly 10,000 acres (most of which administered by the College of Agriculture), Cal Poly is the second largest land-owning university in the state of California. Furthermore, much of this land is open and free of obstructions, includes many ridges with high wind potential, and is located near the coast in a naturally windy micro-climate area.
2. ***Demonstrated high wind potential.*** During the grant period, we carried out a continuous analysis of the wind speed data at the proposed turbine site. The data was gathered to validate the location of the site where the main turbine would be erected.
3. ***Safety.*** Safety requires a compromise between remoteness and proximity. Remoteness, because a failure of any part of the machine must occur sufficiently far to minimize impact to people and property; proximity, so that access to the facility is not prohibitively difficult.

4. ***Visual impact and noise abatement.*** The visual and noise impact of wind turbines is a topic of continuous debate and public concern, which can only be avoided if the turbine is discretely located on campus property. Building permits and licensing for such constructions, especially on the California Central Coast, also motivates sites with limited visual and noise impact.

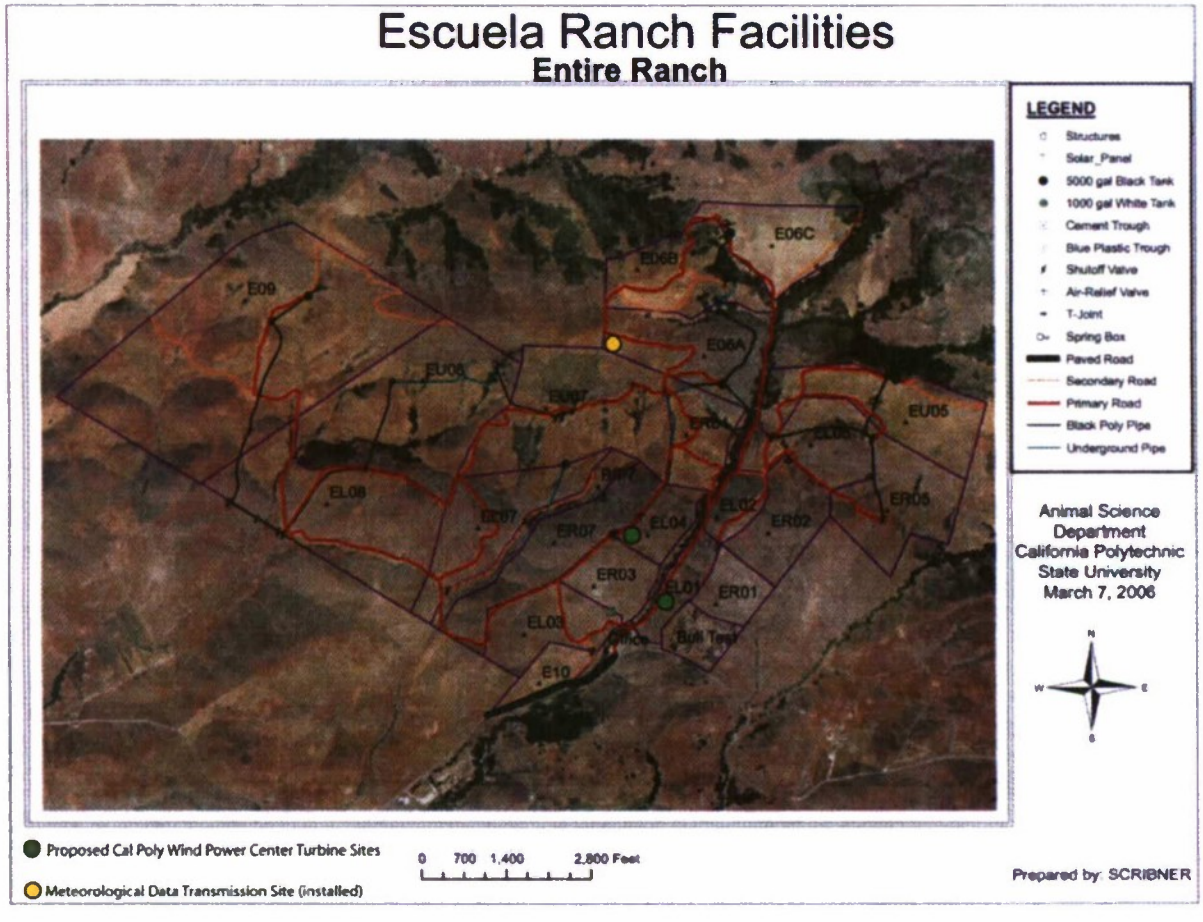


FIGURE 1: CAL POLY WIND POWER RESEARCH CENTER SITE AT THE ESCUELA RANCH. THE SITE ON GRID EL04 WAS CHOSEN FOR LONG TERM MET TOWER STUDIES AND FOR INSTALLING OUR FIRST WIND TURBINE.

Thanks to the support of the College of Agriculture, a testing area was secured early on in this project, northwest of the main campus on Cal Poly's Escuela Ranch, a 2,000 acres ranch separated from the main campus (see Figure 1). The College of Agriculture has agreed to allow us to set up our site at that location, test wind resources, erect towers and test equipment. Prior to the start of the present grant period, we purchased an 80ft meteorological tower and developed remote wind sensing instruments to quantify the wind resource every 20ft of elevation. For nearly two years (spanning the grant period), we archived and analyzed wind speed data continuously. Since commercial, calibrated meteorological station equipment and software can cost tens of thousands of dollars (beyond the budget of this project), Mechanical

Engineering Professor John Ridgely developed and built a system entirely in-house to save cost. The mode of operation of this system is as follows:

- Our system works by counting the number of turns of a cup anemometer over a 10 seconds interval. This number is then scaled to an average miles per hour value for the interval, a standard method in industry. Note that each Cal Poly-designed anemometer was calibrated against a NIST-traceable instrument in a wind tunnel.
- The data are arranged as: 'time', 'average speed', 'maximum speed', 'minimum speed' and 'direction'.
    - The 'average' is the average reading, over 6 minutes, of all 10 second intervals.
    - The 'maximum' and 'minimum' are the highest and lowest 10 second interval readings taken during the 6 minute periods.
- Each wind data file name archived refers to the location of the data. E.g., tower18.csv means tower 1 at 80 ft; tower12.csv means tower 1 at 20ft etc. (Note that there is only one tower at the present time.)
- The direction comes from one sensor only: tower16.csv

Our site requirements were that the annual daytime wind average exceeds 10mph at 70ft (for consistent operating potential), with 10 sec gusts greater than 60 mph recorded sometime during the year (for high wind load studies).

Figure 2 shows the one-year wind average and turbulence intensity, on a per-hour basis, at the wind turbine site. Over the past year, the maximum 10 sec gust recorded was over 92mph, and the maximum 6 minute average was over 65mph.

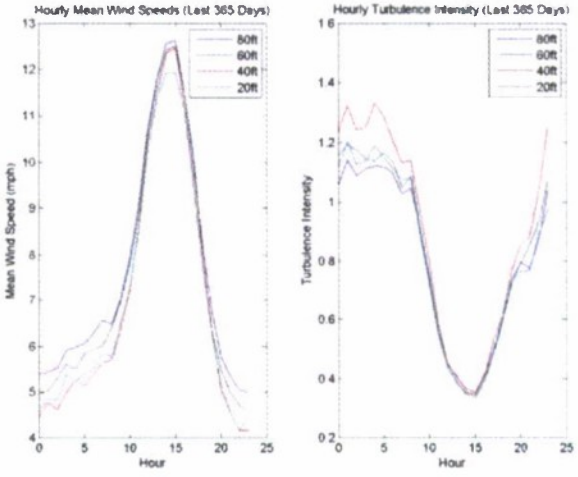


FIGURE 2: ONE-YEAR AVERAGE WIND SPEED DATA AT SITE EL04 OF ESCUELA RANCH, HOUR BY HOUR, FOR THE PERIOD BETWEEN FEBRUARY 20, 2009 AND FEBRUARY 19, 2010. THE DAYTIME AVERAGE AT 70 FT, BETWEEN 9AM AND 5PM, IS ABOVE 10MPH, AND STEADY.

The data gathered confirmed that site EL04 in Figure 1 is well suited for our purpose. Furthermore, the high day-time wind average observed at our site (which was chosen without any initial wind speed data), motivated a new, separate study of the wind resource at Cal Poly,

using computerized fluid mechanics (CFD) tools. This study of the air flow pattern between Escuela Ranch and Poly Canyon seeks to identify zones of highest wind potential, for the purpose of evaluating Cal Poly's suitability for a commercial scale wind farm, and is the subject of an on-going Master's thesis project.

2. Rotor

The wind turbine rotor of a HAWT is an assembly comprised of two main components: a series of blades, and a hub structure.

Blades

At the start of the project, a main part of the scope involved the aerodynamic and structural design of the blades, and a manufacturing process capable of mass producing them. We designed a new turbine blade based on the RISØ-A1 airfoil, a state-of-the-art wind turbine airfoil recently developed in Denmark. The objective of the new design was to optimize the planform (twist and chord distribution, and blade span length) in order to achieve specific aerodynamic goals on power coefficient, tip speed ratio and *passive* power regulation at variable speed, all while minimizing stress at the blade root, where it is most important, and without compromising the planned future development to *pitch* regulate the turbine[1]. The target rotor has a rated power capacity of 3 kW.

A series of eight blades were manufactured using the process developed in this project, to show repeatability and for future rotor and instrumentation test studies. The blades are expected to sustain working loads of less than 100lb_f at their rated wind speed of 22mph. A blade root assembly exemplar was manufactured, thermo-cycled and tested to 500lb_f at 6ft (root moment of 30,000ft-lb_f), with no sign of failure. The actual blades were tested 4 times with a tip load of 120lb_f in a similar manner, again with no sign of damage. Figures 3 and 4 illustrate various parts of the blade manufacturing process, including final painting. This project resulted in the completion of a Master's thesis in September 2009. A copy of this thesis, which clearly describes every step in the manufacturing and testing processes outlined here, is included in the appendix[2].

---

[1] 'Stall' regulation refers to fixed-pitch blades whose aerodynamic properties alone control power transfer; 'pitch' regulation refers to a class of turbine where the blades can rotate along their longitudinal axis to control power transfer.

[2] Edwards B "Composite Manufacturing of Small Wind Turbine Blades: Utility Scale Methods Applied to Small Wind", 2009

FIGURE 3: FOUR OF THE STEPS DEVELOPED IN THE BLADE MANUFACTURING PROCESS. CLOCKWISE, FROM TOP LEFT: MOLD LAY-UP OF FIBERGLASS CLOTH; VACUUM BAG SETUP FOR ONE HALF OF THE BLADE; GLUE PREPARATION OF THE CURED HALF; ASSEMBLY OF TWO CURED HALF BLADES IN A CUSTOM CLAMPING JIG. THE LAY-UP SCHEDULE ITSELF IS A CRITICAL STEP THAT REQUIRED EXTENSIVE ANALYSIS AND TEST. NOTE THAT CLIPPER WINDPOWER PARTICIPATED IN THE CRITICAL DESIGN REVIEW OF THIS PROCESS; THE FINAL METHOD DEVELOPED IS SIMILAR TO THAT USED IN THE UTILITY-SCALE BLADE MANUFACTURING INDUSTRY.

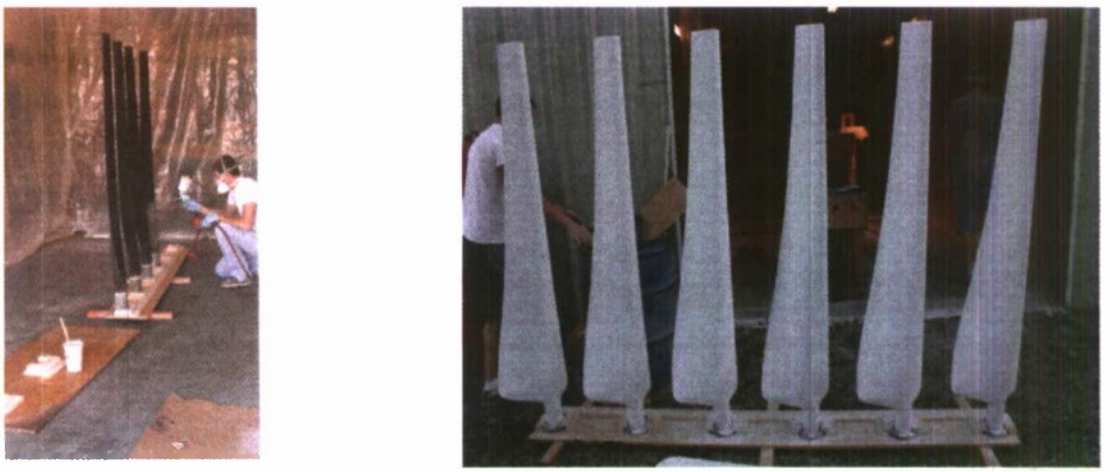


FIGURE 4: THE BLADES ARE PROTECTED WITH A POLYURETHANE-BASED PAINT.

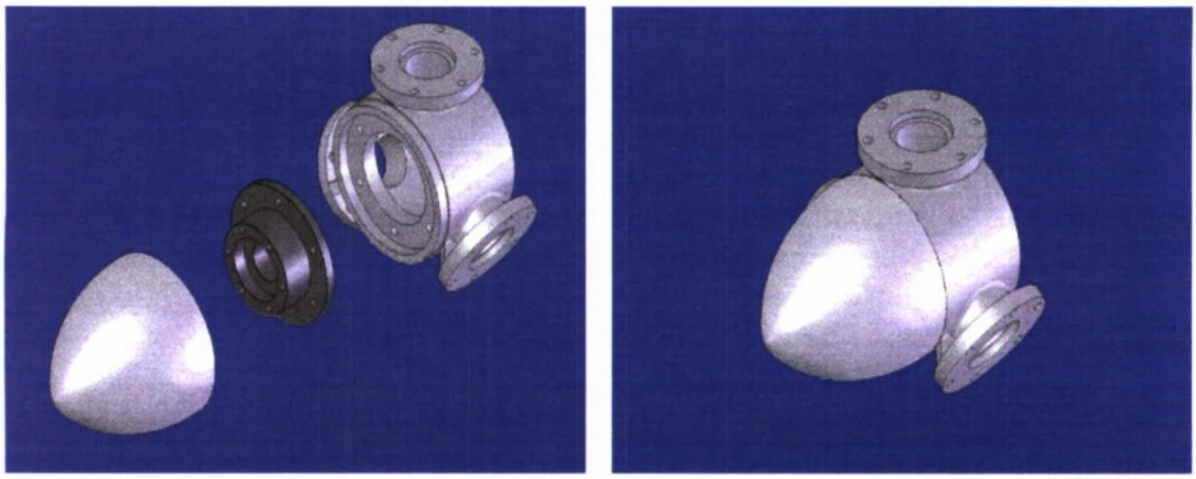


FIGURE 5: ROTOR HUB, AS DESIGNED BY J. GERHARDT

Hub

The design and manufacture of a hub to support three blades and allow the rotor to evolve over time from passively-regulated to actively-regulated, proved to be one of the more challenging tasks of this project. Several key assumptions regarding the blade pitch mechanism (which does not yet exist), were required to finalize its design; the manufacturing itself required a 3-axis computerized numerically controlled (CNC) milling machine, equipped with indexing head. This level of sophisticated machining is beyond the skills of most students and faculty, and outsourcing it was too expensive for the allocated budget. For these reasons, the final design and fabrication was entirely carried out by Mr. James Gerhardt, in the Mechanical Engineering Department.

Figures 5 and 6 show the machined hub, prior to anodizing and powder-coating, ready for mounting on the nacelle powertrain.

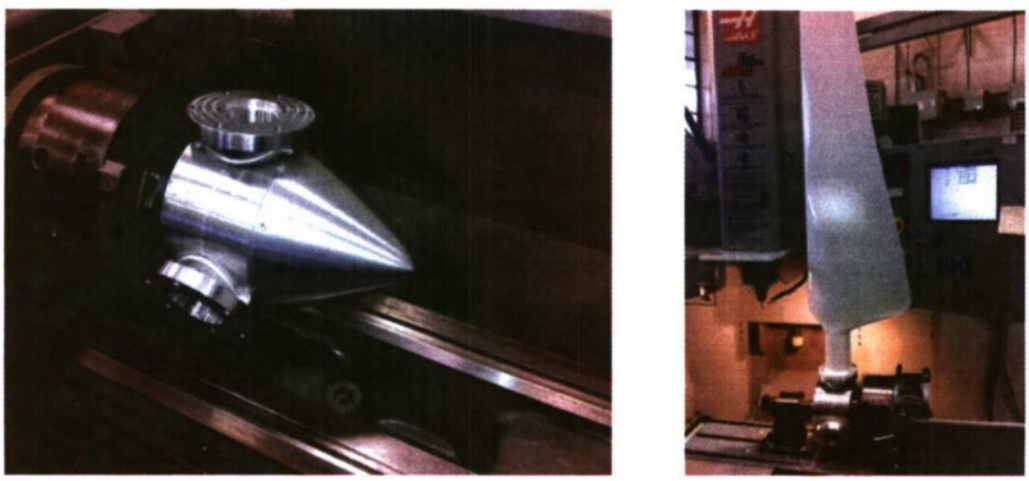


FIGURE 6: HUB, DURING FINAL MACHINING STEP (LEFT, WITH ATTACHED SPINNER), AND WITH ASSEMBLED BLADES (RIGHT)

3. Nacelle

The turbine nacelle is an important component of a wind turbine: it houses the powertrain, control system and brake actuation system. Its design and fabrication of the was targeted as a group project at the advanced undergraduate level, and fit within the scope of the Senior Project requirement of the Mechanical Engineering curriculum. The nacelle was completed on time and on budget; the final student report[3] is included in the appendix. Figure 7 shows the nacelle as designed by students, and the completed physical nacelle. Figure 8 shows the full nacelle/rotor assembly, with hub and blades.

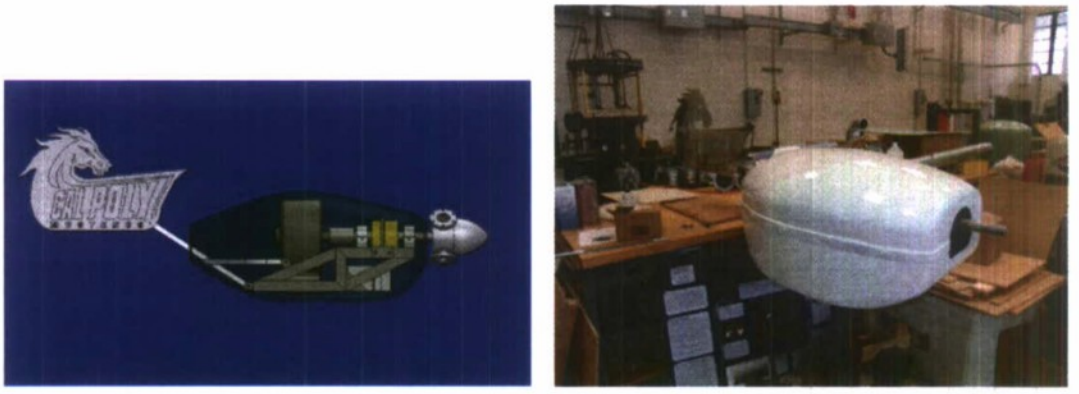


FIGURE 7: THE CAL POLY WIND TURBINE NACELLE, AS DESIGNED AND BUILT BY MECHANICAL ENGINEERING UNDERGRADUATE STUDENTS.

[3] Martinez A, Martinez F, Nevarez D, Taylor Z "Wind Turbine Nacelle Senior Project", 2009

FIGURE 8: NACELLE AND ROTOR ASSEMBLY

4. Tower

The most ambitious and costly aspect of this project is the design, fabrication and installation of the wind turbine tower. There are two main components to a wind turbine tower: the concrete foundation which supports the structure, and the above-ground structure itself ('main tower'). The CPWPRC turbine involves many unique requirements that made the design and manufacturing of both components challenging.

These requirements consist of:

a. ***Tilt-up design:*** The turbine will require regular servicing, and for safety and cost reason, it is not reasonable to rely on a crane to access or remove the nacelle from the tower. A tilt-up design allows the lowering of the nacelle so that all work takes place at ground level.

b. ***Monopole:*** towers may be purchased as either monopole or lattice-work units. Monopole units have longer lives and prevent bird-nesting problems.

c. ***Non-guy-wired structure:*** Many guy-wired towers are available commercially, but these structures are inherently more dangerous than the non-guy-wired ones. The failure of a single guy-wire can sometimes result in the collapse of the entire structure, and therefore require inspection and periodic maintenance. These are considered unacceptable risks for a laboratory where students are expected to gather regularly.

d. ***Height:*** wind speed increases with height above ground. Given that we gathered data up to 80ft, 70ft was chosen as a minimum height for the tower.

No commercially available unit satisfied all these requirements at the time when the project started, and a new tower was designed in-house.

Foundation

The tower foundation is a project that was completely carried out by a group of Civil Engineering students, advised by Professor Robb Moss. The project started with a soil report, commissioned from an outside contractor. Figure 9 to Figure 11 illustrate a few of the major steps of the construction of the tower foundation. The students' final report and the soil report are included in the appendix[4].

FIGURE 9: FROM TOP LEFT, CLOCKWISE: CAL POLY FACILITIES SERVICES EXCAVATING GROUND FOR THE TURBINE FOUNDATION; CONSTRUCTING FORMS; WORKING ON THE MAIN PAD (NOTE THE COPPER WIRE SURROUNDING THE PAD AND CAISSONS, PROVIDING LIGHTNING PROTECTION TO THE TURBINE); CAISSON CAGE

[4] Knox F, Valverde A "Wind Turbine Foundation Design", 2010

FIGURE 10: POURING OPERATION FOR THE TURBINE FOUNDATION.

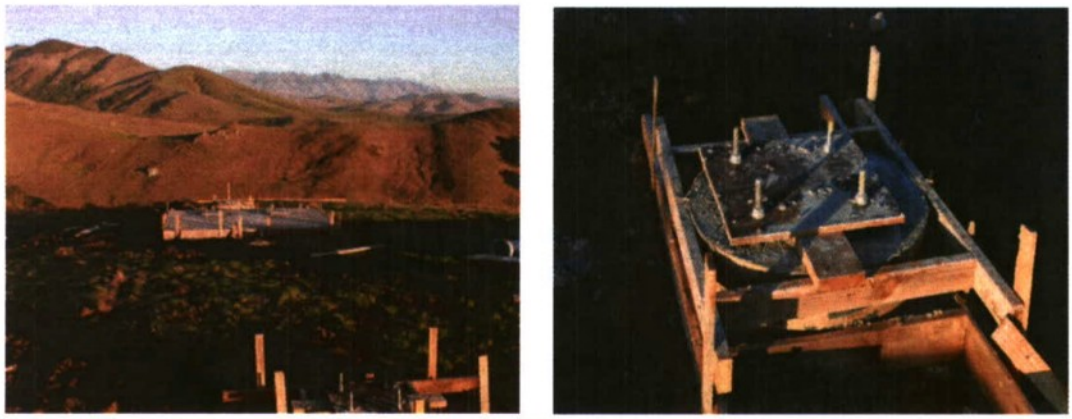


FIGURE 11: FINISHED FOUNDATION PADS AND SUPPORT CAISSONS. THE FOUNDATION CURED FOR APPROXIMATELY ONE MONTH BEFORE THE TOWER ASSEMBLY WAS CONTINUED.

Main tower

Professors Lemieux and Mello, and Mr. James Gerhardt of the Mechanical Engineering Department designed the main tower, which was fabricated by Bassett Mechanical of Kaukauna, Wisconsin, after we solicited multiple bids across the United States. Analyses of loads and deflections, and factor of safety margins, were verified using finite element analysis (FEA) as part of a Mechanical Engineering Master's thesis (completed but not yet submitted). The tilt-up operation is designed to be carried out using an industrial winch attached to a large pick-up truck purchased for this purpose, using Student Fee Committee funds, and start-up funds from Professor Lemieux. The tilt-up operation, as-designed, has been successfully tested.
Figure 12 through Figure 15 illustrate the assembly of the main tower.

FIGURE 12: TOWER ASSEMBLY. THE ASSEMBLY OF THE TOWER TOOK TWO DAYS AND REQUIRED THE USE OF A CRANE, SHOWN IN THE TOP PHOTO SUPPORTING THE GINPOLE IN PLACE. THE TWO LOWER PICTURES SHOW STEPS IN THE MAIN BEARING ASSEMBLY, A KEY COMPONENT OF THE NEW TOWER DESIGN. WE DEVELOPED A DETAILED PROCEDURE FOR THE ENTIRE ASSEMBLY, PRIOR TO THE START OF FIELD WORK.

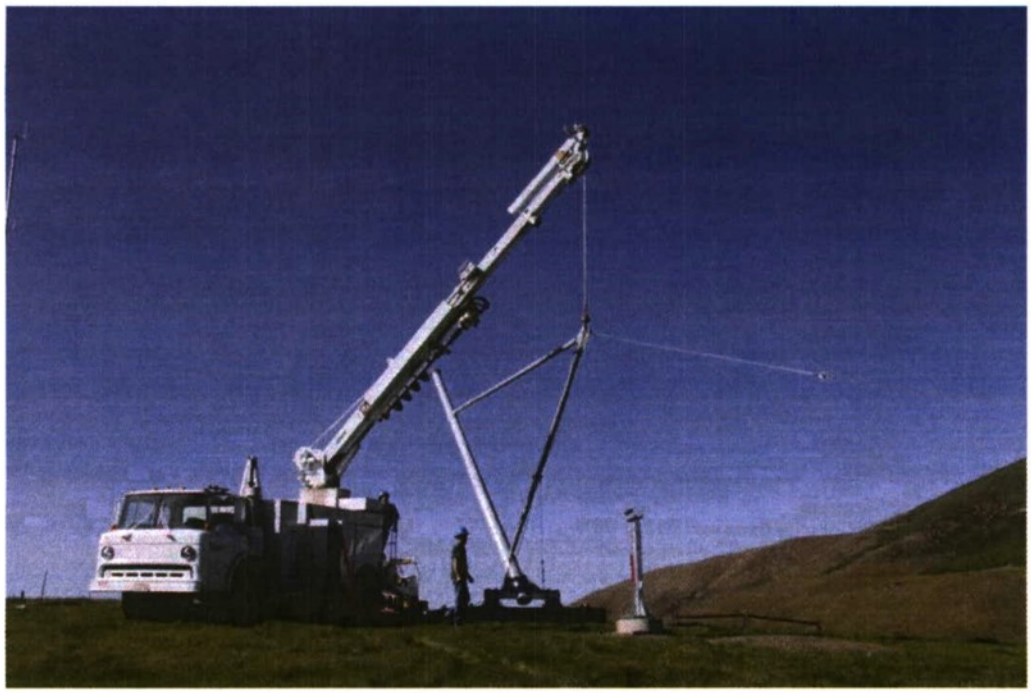


FIGURE 13: TEST OF TOWER TILT-UP OPERATION, PRIOR TO FINISHING TO THE ASSEMBLY OF THE UPPER TOWER SECTION.

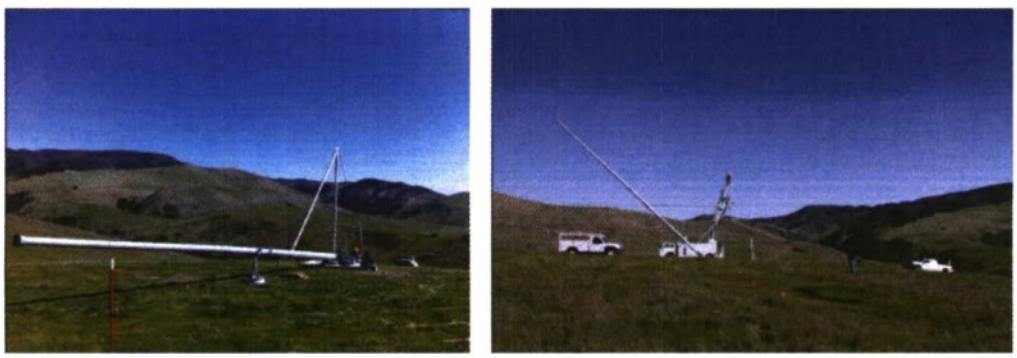


FIGURE 14: FINAL TOWER ASSEMBLY AND TILT-UP TEST. NOTE THE TRUCK AT THE RIGHT OF THE LOWER PICTURES, WHICH PROVIDES THE SOLE MOTOR FORCE FOR TILTING UP THE TOWER.

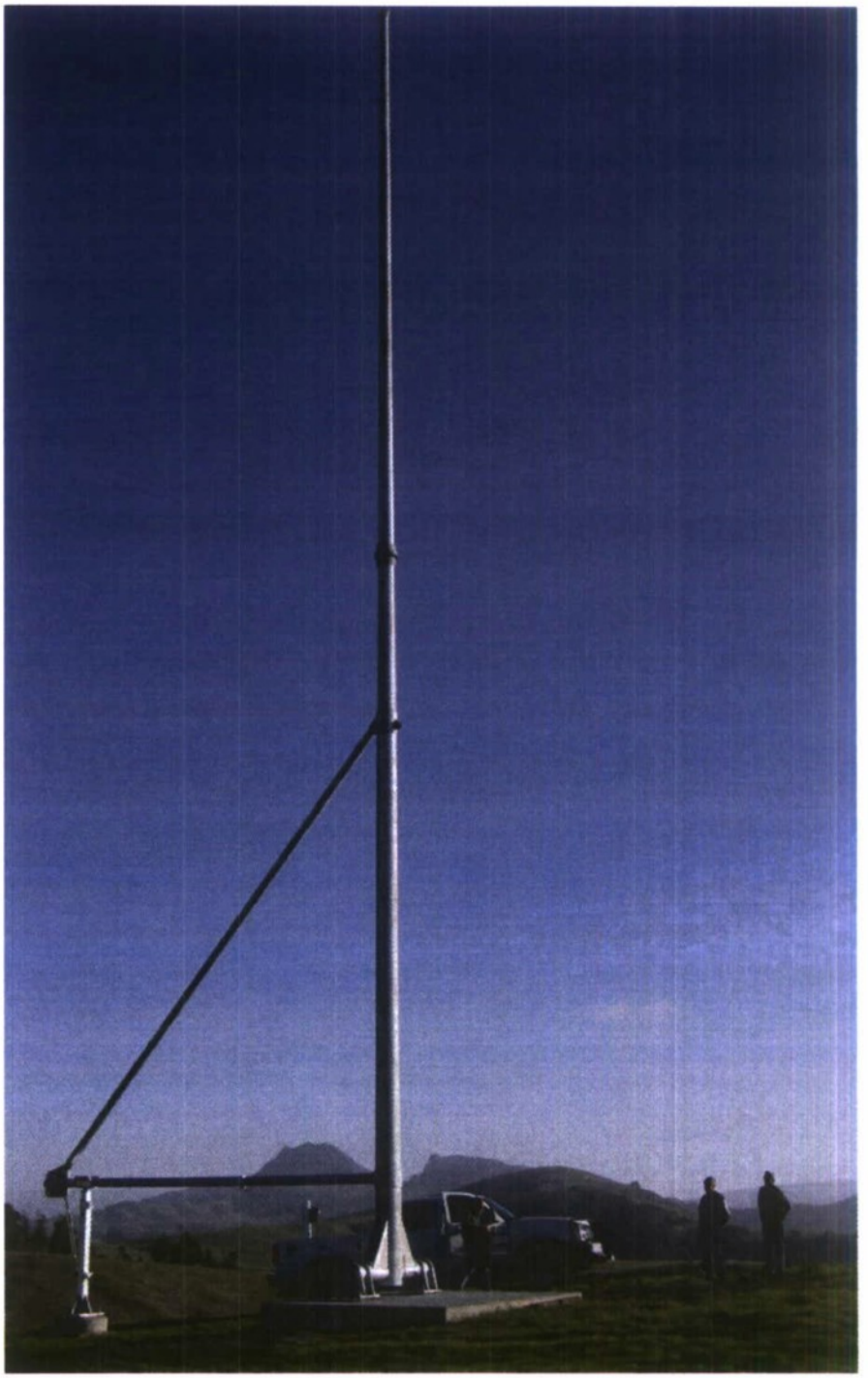


FIGURE 15: THE FULLY ASSEMBLED TOWER, IN TILT-UP POSITION.

Outside Publications & Patents

Clearly, an important goal in the development of a new research facility is to produce new ideas which result in publications, showcasing the researchers' work, Cal Poly, and the Center itself. The CPWPRC is not yet completed, but its design and construction generated several presentations, conference papers, panel discussions and invited lectures[5].
The unique nature of the tower and nacelle frame are also currently being considered for patent protection by the university.

Future work

Ultimately, the Cal Poly Wind Power Research Center will provide a unique, specialized, consolidated research point on the California Central Coast, for the study and development of blade design, rotor performance, blade life and health monitoring, tower design and performance analysis, that will benefit the Cal Poly community and the local area.

We cannot predict the full breadth of projects that will come of it. What follows is a list of a few of the projects that are being considered for the near future:

- Characterization of turbine the performance and quantification of pitch indexing.
- Design, construction and test of a pitch actuation mechanism.
- Design, construction and test of a controlled yaw system.
- Blade load analysis.
- Tower resonance mode studies.
- Laboratory development for undergraduate classes.

With the new capability of this research center, Cal Poly's involvement and visibility in the renewable energy industry is certain to increase. The C3RP grant has had a tremendous impact on the start-up of this Center, and has played a key role in guaranteeing its future success.

Conclusion

Thanks to the support of C3RP, a new facility, the Cal Poly Wind Power Research Center, which consists of a fully functional research wind turbine, has become a reality. The turbine components fabrication is complete, and near final assembly. This facility, almost entirely designed and built by Cal Poly students, staff and faculty, will provide a unique platform for future students and faculty to work on problems relevant to the utility-scale wind turbine industry, for many years to come.

[5] Three posters and a published paper are included in the appendix.

APPENDICES

The following Appendices can be found in the Appendix to this report.

1. Edwards B "Composite Manufacturing of Small Wind Turbine Blades: Utility Scale Methods Applied to Small Wind", 2009
2. Martinez A, Martinez F, Nevarez D, Taylor Z "Wind Turbine Nacelle Senior Project", 2009
3. Knox F, Valverde A "Wind Turbine Foundation Design", 2010
4. Soil report, Down R, Earth Systems Pacific
5. Sample conference posters and papers, AWEA Windpower 2008:
    a. Lemieux P, Ridgely J "Cal Poly Wind Power Research Center"
    b. Katsanis G, Lemieux P "Structural Analysis of Small Wind Turbine Towers"
    c. Nosti C, Lemieux P, Gascoigne H "Performance Analysis and Life Prediction for Small Wind Turbines: A Wood Laminate Case Study"

Solar Concentrators: Quickly Deployable Combined Heat and Power

Principal Investigator:

Peter V. Schwartz, Ph. D.
Physics
California Polytechnic State University
San Luis Obispo, CA

Abstract
Our work has centered on the construction of a solar concentrator and thermal storage facility in order to make use of solar energy, quickly, inexpensively and reliably (at all times). In 2009, we constructed a parabolic trough solar concentrator as a prototype and a thermal storage device based on using sand as the thermal storage medium and pumice as the insulating material. In summer 2010, we built a parabolic dish solar concentrator that has tested superior to the parabolic trough of last summer. We also tested the thermal storage device, and verified that it is capable of storing significant energy for many hours. Lastly, we continue to revise for publication a study we started in 2007 modeling a novel solar concentrator. This concentrator utilizes a stationary primary mirror and stationary heat collection element, redirecting the light by means of a secondary mirror. The recent revision has received a positive response, and we anticipate publication with the next submission.

Project Significance
Our Concentrated Solar Power (CSP) research has both general significance as well as the particular significance. It is important to note that the wider topic of CSP, Solar Thermal, and photovoltaics is of interest to a community of more than 15 instructors university-wide and many students, and is a priority for both the military and domestically on every level. The motivation for solar energy comes from the virtues of it having a low CO_2 impact, becoming increasingly less expensive than other energy conversion technologies, and available anywhere there is sunlight, which is of particular importance in remote field environments. As the community of researchers on campus (and beyond) share knowledge and resources, this particular project will benefit the wider solar energy effort.

Concentrating and storing solar energy increases value by delivering higher temperatures and greater power fluxes. The higher temperatures are useful for a wider range of industrial and domestic applications and also convert more efficiently to electricity (in the case of STE, Solar Thermal Electrical conversion), and the smaller area of sunlight allows for use of smaller, more expensive, more efficient photoelectric panels. Solar thermal storage makes solar thermal energy (including that for STE) a reliable source of energy. The challenges to both technologies are cost and complexity, as well as transporting these technologies to remote sites. Tracking a moving sun is a necessity for all solar concentrators, representing one particular technical and financial challenge. Our work is dedicated to reducing the cost and complexity of solar concentrators and constructing thermal storage devices out of locally available materials. This promises to make solar concentration accessible for a wide variety of applications. These applications include electricity production, industrial and domestic heat and cooking, and air conditioning via absorption cooling. Potential beneficiaries are also vast including military field stations, industrial, commercial, and domestic users – wealthy and impoverished.

SOLAR CONCENTRATORS AND THERMAL STORAGE

More recently, we have constructed a parabolic mirror concentrator (Fig. 1) and thermal storage facility (Fig. 2). We chose a simple parabolic mirror assembly over computer simulated concentrator we are trying to publish as a first step in order to first gather experience in a proven technology, as many of parabolic troughs have been successfully built. The thermal storage unit is made from widely available materials such as sand and pumice.

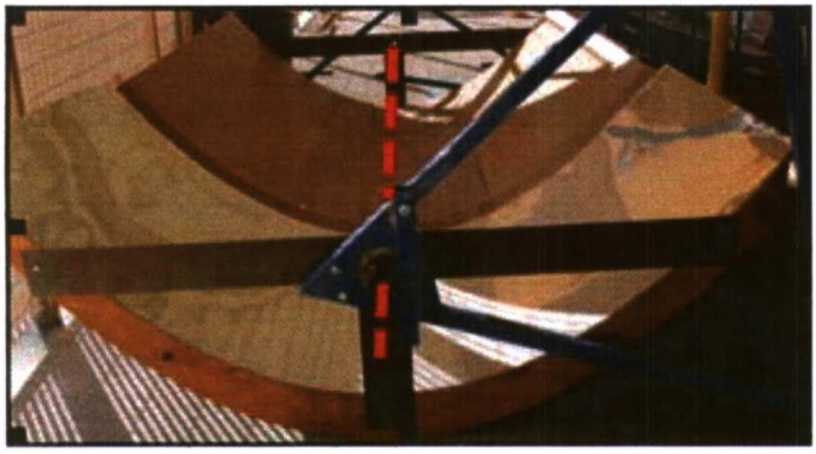


Fig. 1 Parabolic Solar Concentrator. Width: 8', Length: 12'. The Parabolic concentrator is made from thin plywood bent over wooden ribs. Steel piping runs through the wooden ribs lengthwise. The red dotted line represents the location of the heat collection pipe.

Fig. 2a) The solar thermal storage facility from above (*left*) before being filled with sand and pumice. The outer cylinder is full of pumice as an insulator. The center cylinder is filled with sand, which stores the heat. The heat is distributed throughout the center by means of pipes containing motor oil. *Right,* After being filled with sand, the top surface of the storage unit was covered with more pumice for insulation.

OUTCOME AND UPDATES

The intention for this system was to have the heat taken from the concentrator's central heat collector pipe (red dotted line in Fig. 1) carried to the pipes in the thermal storage device by means of hot mineral oil. This design had the following drawbacks:

1) The trough collector was very heavy and not rigid enough. The result is that the collector sagged, would have been excessively difficult to construct, and would likely not have lasted long.
2) Pumping hot oil through a pipe requires a very expensive pump, and power source. It also presents a large surface area to insulate and fire danger if the fluid is flammable.

We have since revised our project to make use of a dish concentrator that focuses light onto a stationary target, which will be our present thermal storage device. We are starting with a well-proven concentrator, which is flexible and lightweight. Additionally, this method does not require any medium to transport the heat to the thermal storage device, because the sunlight is reflected directly into the thermal storage device. This facility initiates our solar concentrator research laboratory, where we will continue to design, test, and improve associated technologies. The present challenges are to reduce the complexity and cost of both the concentrators and thermal storage device. We have been joined by a lecturer in Industrial Manufacturing and Engineering (Kevin Williams) with expertise in many construction technologies, and we have experience in modeling solar concentrators.

Scheffler Solar Concentrator: Numerous successful designs for solar concentrators exist and are presently deployed in all sizes and for many applications [1]. One particular compelling design is the Scheffler Concentrator [2], comprising a portion of a parabolic dish, which rotates about an axis tracking the sun (Fig. 3).

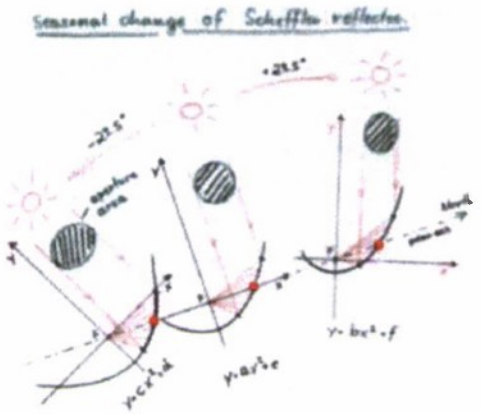


Fig. 3 Different sections of a parabola focus the sunlight at different seasons. This parabola rotates about the same axis for all seasons if the reflecting surface is located due north or due south of the target (located at the focus of the parabola). The focus F and the center of the Scheffler reflector (red dot) remain stationary! From Ref #2

Thousands of Scheffler concentrators have been constructed world wide, but are exceedingly demanding to build with a large number of moving parts (See Figs 4 - 6). We have built the first Scheffler in North America, and will subsequently explore simpler designs.

We have begun modifying our Scheffler Concentrator of 2 m^2 surface area for the purpose of testing the design as a means of heating the thermal storage device. Future work includes testing the Scheffler for solar tracking and concentration, heating the thermal storage unit with concentrated sunlight from the Scheffler, and developing the construction process to make the concentrator larger, less expensive, and easier to construct.

This summer has seen the completion of the construction of the Scheffler concentrator, experimentation with the methodology of producing a reflector, and continued investigation into the best process to produce a reflective surface with relative ease with the rigidity and elasticity necessary for a fully functional Scheffler Reflector. We have designed and are constructing a tracking mechanism.

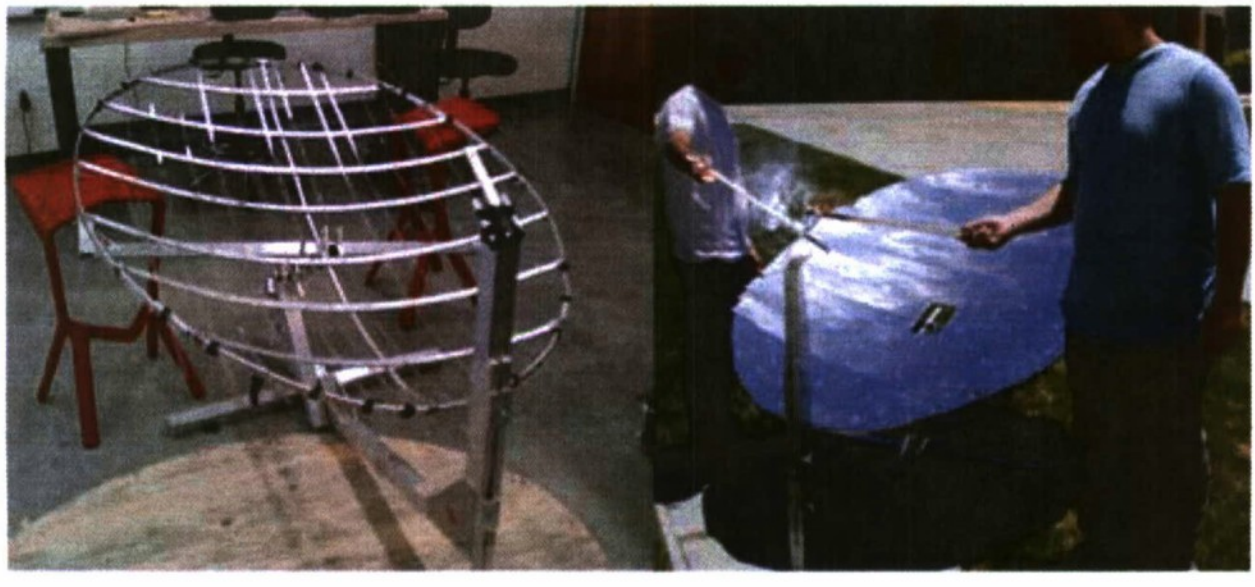


Fig. 4 Construction of a Scheffler concentrator. *Left*, the supporting matrix requires considerable time and materials. *Right*, the finished concentrator ignites wood placed at the focus.

Fig 5 *Left*, construction of the reflector frame, *Right* tracking mechanism work.

Fig 6: *(left)* Original method of reflective material attachment, *(middle)* framework for the second iteration of attachment of the reflective material, and *(left)*, the completed second concentrator iteration.

Thermal Storage Device construction and Testing: Like solar concentrators, many designs for thermal storage exist [3], but require a large amount of material to store and insulate the heat that is often cost prohibitive. However, if the thermal storage device can be made from readily available local material, the cost is drastically reduced. We have constructed and tested a thermal storage device with positive results. Subsequent work will optimize performance and connect the thermal storage device to the appropriate solar concentrator.

We created a thermal storage unit to provide accessible energy after the sun goes down (Fig. 2a). The goal is a high quality, energy storage device made inexpensively with local materials. We researched readily accessible natural materials, settling on granular pumice as an insulator and sand as the thermal storage material (Fig. 2b). These materials are held in place with a wire mesh in a concentric cylinder configuration. Sand is less ideal than concrete as the thermal storage material, but it is much easier to disassemble so it serves as a good test material.

A 1.5 kW heating element was put at the bottom of the device and 15 thermal couple temperature sensors were spread in a radial and vertical pattern throughout both the sand and the pumice, providing a clear picture of what was happening within. Heating with the electric heater we attained a core temperature of 120 °C with a 300 °C "hot spot" near the heater, with both temperatures still increasing after 8 hours. After the heater was turned off, the core cooled at a rate of 4 °C/hr.

A simple computer model simulates the thermal storage device yielding results that agree well with our experimental data. The model predicts a core cooling rate of 4.9°C/hour at 120 °C, which roughly agrees with the measurements of the physical thermal storage device. Additionally, rough estimates predict a maximum equilibrium temperature of more than 300 °C at power inputs of 1.5 kW possible with a 2 m^2 Scheffler. Higher temperatures should be attainable with larger concentrators.

Improvements in Thermal Storage and Connecting to Solar Concentrator

We have already shown that we can store heat at reasonably high temperatures for several hours. Next we plan on finding a better way to spread the energy throughout the sand, because the sand proved to be a poor heat conductor. This is a problem if heat needs to be extracted at high rates. We are presently adding a heat conductor in our computer model, as well as inserting an aluminum conductor into the sand (Fig. 7) in order to monitor the new performance.

Fig 7: Pounding the aluminum bar into the thermal storage device.

Lastly, we will combine the thermal storage with solar concentrators. The concentrator will shine light through a conduit, into the middle of the storage unit onto the thermal conductor within the sand. An insulating plug will fit into the conduit to increase insulation after the sun has gone down.

Ray Tracing of Scheffler Variations - Although the Scheffler model is a device that currently works as is, we wish to optimize this apparatus for different locations and applications with possible secondary mirrors. In particular, we hope to increase the area (power harvesting) and reduce complexity and cost. Light Tools, as established earlier, is the ideal software to achieve these goals.

We will run recurring cycles of design (modeling), construction, data analysis as we develop expertise and improve the technologies.

Broader and Long Term Solar Energy Goals. This project fits well within the sustainability goals of the university and campus-wide efforts to expand renewable energy in general and solar energy in particular. Others on campus are involving students in building solar water heaters, building Solar Thermal Electric systems from traditional concentrators, building battery storage systems for homes and cars, and designing new photovoltaic substances. Additionally, consistent with "learn by doing", this project simultaneously educates students about solar energy, builds fundamental science understanding, and develops practical technologies. As resources and expertise are shared on campus, this project contributes to the larger solar energy research on campus.

Solar energy is receiving increasing attention from both government and corporate funding. With two other professors, we have been vigorously pursuing grants from the California Energy Commission, and are presently in negotiations with a concentrated solar start up company regarding a related concentrated solar concentrating technology. Once we have a working prototype, we will seek external funding for wider development.

MODELING OF NOVEL SOLAR CONCENTRATOR

INTRODUCTION

This description of our manuscript ("Concentrating sunlight with an immobile primary mirror and immobile receiver: ray-tracing results", Rogers, S. C.; Barickman, C.; Chavoor, G.; Kinni, M.; Schwartz P. V.; *J. Solar Energy*, submitted, June 2010) *is very brief because it is explained more completely in last year's summary, and because it occupied a very small portion of our work.*

Large solar concentration devices have traditionally consisted of a parabolic primary mirror, which focuses light onto a target, such as a heat collecting element (HCE) or photovoltaic cell (PV). In order to keep the target at the focus of the primary mirror, the entire mirror must rotate about either one axis (for trough systems), or two axes (for dish systems). We modeled a solar concentrator that consisted of a stationary primary trough mirror of circular cross section, and stationary target (Fig. 8). Focusing is achieved through the movement of the smaller, secondary mirror. While the movement of the secondary mirror presents significant challenges, this system has the potential to be produced on site quickly and inexpensively by embedding the primary mirror and target directly in the earth. In our paper, we explore the performance of a trough concentrator system. However, the idea can equally be applied to dish geometry.

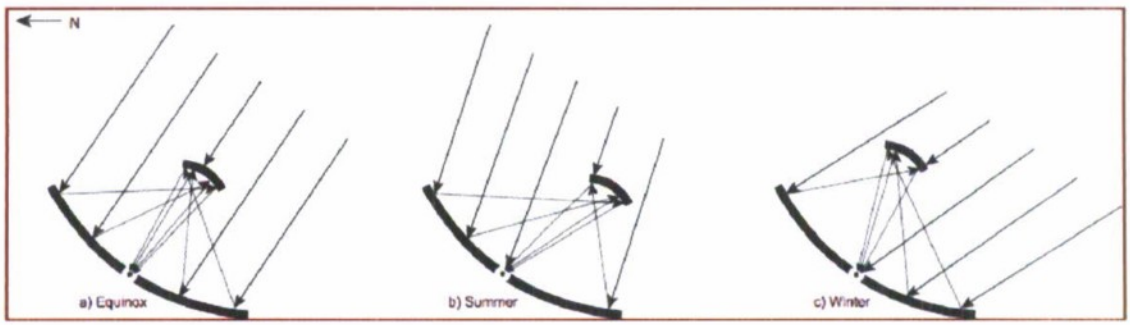


Fig. 8 Incident sunlight is reflected from a stationary primary mirror and focused onto a stationary target by means of a smaller, mobile secondary mirror. If the surface is optimized for equinox, summer and winter light is still focused well onto the same target.

METHODOLOGY AND RESULTS

Modeling of the shape and placement of reflecting surfaces was made possible with commercial ray-tracing software (*LightTools 6.1.0* by Optical Research Associates). We evaluated collector performance oriented both normal to the surface and tilted at latitude. We ran simulations for three dates that give the operating extremes of collector performance: Vernal Equinox, Summer Solstice and Winter Solstice. Modeling was purely geometric: mirrors were assigned a reflectivity of 100% and the HCE, a reflectivity of 0%. We simulated two different HCE types with the same size aperture: a flat plate and a cylindrical absorber using a tertiary reflector (an involute CPC [4]).

Peak concentrations were about 35 suns with an efficiency of collecting about 30% of the incident light. This compares poorly with the performance of conventional trough-based solar concentrators with up to 70 suns and near 100% efficiency of incident light. However, our systems would utilize land with about twice the efficiency of conventional parabolic troughs, and may be cheaper and easier to construct.

Our correspondence with the publication *Journal of Solar Energy*, has required us to define and improve details in our model. The last response indicates that the manuscript will be accepted with minor changes, which we are working on presently.

REFERENCES

[1] Solar Power Plants, Fundamentals, Technology, Systems, Economics; C. J. Winter, R. L. Sizmann, L. L. Vant Hull; Springer Verlag, 1991
[2] Introduction to the Revolutionary Design of Scheffler Reflectors, W. Scheffler, International Solar Cooking Conference, Granada Spain. For paper: http://www.solare-bruecke.org/infoartikel/Papers_%20from_SCI_Conference_2006/21_wolfgang_scheffler.pdf. Conference website: http://www.solarconference.net/
[3] *"Reflector and Thermal Storage Design for a Solar-Powered Oven"*, Trevor Davis, Cal Poly Physics Senior Project, 2009
[4] Winston, R., et al., 2005. *Nonimaging Optics*, Elsevier Academic Press, San Diego, p. 52

Beneficial Reuse of Corrugated Board in Slurry Applications

Principal Investigators

James Hanson, Ph. D.
Civil and Environmental Engineering

Jay Singh, Ph.D. Industrial Technology

California Polytechnic State University
San Luis Obispo, CA

Beneficial Reuse of Corrugated Board in Slurry Applications

James L. Hanson[1], M. ASCE, Nazli Yesiller[2], A.M. ASCE, Jagjit Singh[3], Greg M. Stone, S.M. ASCE[4], and Adam Stephens[5]

[1]Professor, Civil and Environmental Engineering Department, California Polytechnic State University, San Luis Obispo, CA 93407; jahanson@calpoly.edu
[2]Interim Director, Global Waste Research Institute, California Polytechnic State University, San Luis Obispo, CA 93407; nyesiller@gmail.com
[3]AssociateProfessor, Industrial Technology, Orfalea College of Business, California Polytechnic State University, San Luis Obispo, CA 93407; jsingh@calpoly.edu
[4]Research Assistant, Civil and Environmental Engineering Department, California Polytechnic State University, San Luis Obispo, CA 93407; gstone@calpoly.edu
[5]Laboratory Coordinator, Orfalea College of Business, Industrial Technology, California Polytechnic State University, San Luis Obispo, CA 93407; astephe@calpoly.edu

ABSTRACT: Use of virgin and post-consumer corrugated board as a replacement for bentonite in slurry mixes was investigated. The effectiveness of the slurry mixes was assessed using typical tests including Marsh funnel viscosity, density, and filtrate loss. Filter cake permeability was also determined. Corrugated board was fiberized for the test program. Test results indicated that corrugated board could be used to replace 9 to 27% (0.5 to 1.5% corrugate content) of bentonite in slurry mixes with a total solids content of 5.5%. Slurry applications provide a new and viable beneficial reuse alternative for paper/paperboard products, which constitute the largest weight and volume fraction of municipal solid waste generated and disposed of in the U.S.

INTRODUCTION

Bentonite slurries are used in construction of vertical cutoff walls for geotechnical and geoenvironmental applications. The construction of cutoff walls typically is a step-by-step process, where a trench is excavated, filled first with slurry, and then backfill. The slurry covers the inside walls of the trench forming a low permeability filter cake layer. The slurry also provides hydrostatic pressure to keep the trench open prior to placement of the backfill. Typical slurries consist of 4 to 7% bentonite and 93 to 96% water by weight (Boyes 1975). Paper and paperboard constitute the highest fraction by both weight and volume of municipal solid waste generated (32.7% by weight) and disposed of (22.3% by weight) in the U.S. The amount of paper and paperboard generated and disposed of was 83 million and 37.8 million tons in 2007, respectively (USEPA 2009). The use of recovered paper in manufacturing containerboard has remained stable at approximately 16 million tons since 1997 (Paper Industry Association Council 2008). An economical limit for incorporating

waste papers into containerboard has been reached. Some corrugated board is not suitable for conventional recycling due to presence of contamination. Pizza boxes are a common example, which comprise nearly 1% of the total annual production of 313 million m^2 of corrugated board (Flaherty 2009). Residue on pizza boxes is problematic for recycling as grease prevents absorption of moisture, proper pulping of paper fiber, and quality of binding of fibers in recycled paper (RecycleBank 2009). Innovative recycling options (beyond the packaging industry) need to be investigated to promote beneficial reuse of paper products. This study has been conducted to evaluate reuse of paper and paperboard in civil engineering applications. Results from the portion of the study with use of corrugated board in slurry applications are presented herein.

EXPERIMENTAL TEST PROGRAM

Tests were conducted to assess the feasibility of using corrugated board in slurry mixtures. Bentonite was replaced by corrugated board at varying ratios. Properties of bentonite-board-water mixes were compared to baseline bentonite-water mixes to evaluate the influence and practical limits of corrugated board addition.

Materials

A commercially available standard powder bentonite was used in the test program. Baroid AQUAGEL is a finely ground, premium-grade Wyoming sodium bentonite that meets the American Petroleum Institute (API) Specification 13A, section 4 requirement. The bentonite had a liquid limit = 539, plastic limit = 82, and specific gravity = 2.65. Corrugated board was selected as the paper/paperboard product due to the significant quantities available for reuse. Tests were conducted on non-waxed products. Two types of corrugated board were used in the test program: conventional box material (e-flute corrugated board) and pizza boxes. Identical products for virgin (V) and post-consumer (PC) corrugated board were tested to determine potential effects of use on the properties of corrugated board in slurry applications. The e-flute corrugated board was subjected to standardized laboratory conditioning as prescribed by ISTA (2009) to provide post-consumer status. The post consumer pizza boxes were collected from a garbage bin and contained representative amounts of food product (i.e., grease and food remains) residue. The corrugated board samples were fiberized by mixing with water in a Waring eb 15 stainless steel 4-L capacity blender that contained a specially fabricated blade adhering to the specifications outlined by White and Kendrick (2009).

Corrugated Board Tests

Tests (summarized in Table 1) were conducted on virgin and post-consumer corrugated board to determine material properties. The corrugated board properties are presented in Table 2. The edge crush and water absorption tests on corrugated board provided indication of the fiberization potential and shredding of the corrugated board for the proposed slurry application.

Table 1. Corrugated Board Tests

Test Name	Standard Designation	Description of Test
Grammage of paper and paperboard (weight per unit area)	TAPPI T 410 om-02	Weight per 92.90 m^2 of all three containerboard components of a single wall corrugated fiberboard is determined after conditioning for 24 hours at 21±1°C temperature and 52±0.5% RH.
Bursting strength of corrugated and solid fiberboard	TAPPI T810 om-06	Square corrugated fiberboard samples with dimension 31.50±0.03 mm are tested by distending an expansible diaphragm under a pressure of 690 kPa to 4825 kPa.
Edgewise compressive strength of corrugated fiberboard	TAPPI T839 om-02	A test specimen with length 50.8±0.8 mm and height 25.4±0.4 mm is compressed vertically (load parallel to flutes) to failure at the rate of 111±22 N/s.
Water absorptiveness of corrugated fiberboard (Cobb test)	TAPPI T441 om-04	A sample with a diameter of 11.28±0.02 cm is exposed to 100 mL of water (23±1°C) and a head of 1±0.1 cm for 120 seconds.

Table 2. Corrugated Board Properties

Material	Weight/Unit Area (g/m^2)	Burst Strength (kPa)	Edge Crush (N)	Cobb Test (g/m^2)
V c-flute box	579	1350	162	78.3
PC c-flute box	588	1140	146	89.0
V pizza box	447	900	155	94.3
PC pizza box	493	1025	165	99.0

Slurry Tests

Slurry mixes were prepared using bentonite and water and also using bentonite, fiberized corrugated board, and water. Visual comparison was made for the solids suspension / sedimentation behavior of the bentonite and fiberized corrugated board by allowing mixtures to settle in hydrometer jars and by centrifugation. Example photographs of the slurries are presented in Fig. 1. The centrifuge was run at 1500 rpm for 1 minute to assess segregation and sedimentation in the mixtures. At high fiber ratios, the homogeneity of the slurries was decreased as flocculation and segregation of the fibers occurred (presence of clear water within the grab samples and at the top of the centrifuge samples). Bentonite dispersed in water and remained in suspension for extended periods of time, whereas the paper fibers alone flocculated in the presence of water and became segregated from the bentonite with time. In addition, high corrugate content mixtures exhibited gas production within 5-6 days of mixing, which remained entrapped within the fiber matrix. Overall, fiber-only and low-bentonite content mixes (<2.5% bentonite) were deemed inappropriate to provide effective slurry behavior. Solids remained in suspension for slurry mixtures that contained both fibers and sufficient amounts of bentonite. The specific mixtures tested for slurry behavior were based on these observations.

All mix ratios are provided on weight basis (Table 3). Tap water was used for all slurry mixes. The water was conditioned to a pH of 8.5 ± 0.12 using small amounts of soda ash before solids were added to the slurry mixtures. Pure bentonite slurries were mixed in a blender on low speed for 2 minutes prior to testing. For slurries containing corrugated board, water and corrugated board were mixed on low speed

for a 2-minute period to allow corrugate to be pulped into fibers, then mixed again after the addition of bentonite for an additional 2 minutes. The post-consumer corrugated board was more difficult to pulp than the virgin corrugated board. The original pulping period of 2 minutes was increased to 5 minutes for post-consumer corrugated board samples in subsequent tests to investigate the effect of increased pulping duration on the engineering properties of the slurry mixtures.

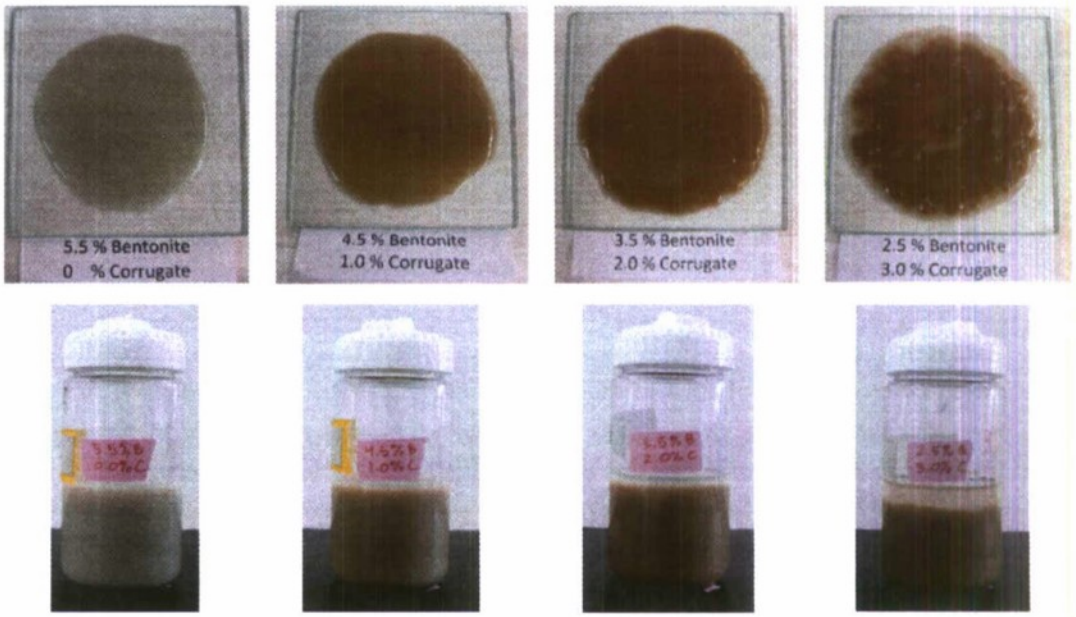


Figure 1. Comparison of Slurry Mixtures (grab and centrifuge samples)

Table 3. Slurry Mix Ratios

Mix Number	Corrugate Type	Bentonite (%)	Corrugated Board (%)	Water (%)
B5.0	None	5.0	0	95
B5.5	None	5.5	0	94.5
B6.0	None	6.0	0	94
V0.5	Virgin	5.0	0.5	94.5
V1.0	Virgin	4.5	1.0	94.5
V1.5	Virgin	4.0	1.5	94.5
V2.0	Virgin	3.5	2.0	94.5
V2.5	Virgin	3.0	2.5	94.5
V3.0	Virgin	2.5	3.0	94.5
PC0.5	Post-consumer	5.0	0.5	94.5
PC1.0	Post-consumer	4.5	1.0	94.5
PC1.5	Post-consumer	4.0	1.5	94.5
PC2.0	Post-consumer	3.5	2.0	94.5
PC2.5	Post-consumer	3.0	2.5	94.5
PC3.0	Post-consumer	2.5	3.0	94.5

Typical slurry tests (D'Appolonia 1980, USEPA 1984) were used in the experimental program: Marsh funnel viscosity (ASTM D 6910); mud balance (ASTM

D 4380); filter press (API Recommended Practice 13B); and filter cake permeability (conducted in conjunction with / immediately following API Recommended Practice 13B). Standardized test methods were generally followed. In order to prevent segregation of paper fibers from the slurries, the mixtures containing fibers were not poured through the attached funnel screen in Marsh funnel testing. The thickness of the filter cake was measured by averaging values determined at three locations on the cake using a pair of digital calipers. Thickness of the filter cake was measured after permeability tests had been conducted to minimize damage and disturbance to the filter cake. Separate tests were conducted to verify that filter cake thickness did not change during the permeability stage of the experiments. The permeability tests were conducted using a pressure differential of 140 kPa. The hydraulic gradient varied depending on the thickness of the filter cake and was on the order of 5,000.

RESULTS AND DISCUSSION

The results of the slurry tests are provided in Table 4. The Marsh funnel viscosity of the bentonite-water slurry mix was equal to 40.5 s for the 5.5% solids content and this mixture was established as the baseline mixture for the tests. The total solids content of the mixes containing both bentonite and corrugated board was set to 5.5%. The viscosity of the mixture with 5% bentonite was in general similar to the baseline mix, whereas the 6% bentonite mix had higher viscosity and density and lower filter cake thickness than the baseline mix. For mixes containing bentonite, corrugated board, and water, viscosity, filter loss, filter cake thickness, and filter permeability generally increased with increasing corrugate content. Mud balance density generally decreased with increasing corrugate content.

Acceptable slurry mix properties were established as: Marsh funnel viscosity of approximately 40 s (up to 50 s was deemed acceptable for this test program); density of 1010-1040 kg/m^3; and filtrate loss of less than 30 ml based on specifications provided in USEPA (1984) and Ryan and Day (2003). The variations of Marsh funnel viscosity, density, filtrate loss, and permeability as a function of corrugate content are presented in Fig. 2. Shaded regions in the plots in Fig. 2 represent areas that are outside acceptable limits for Marsh funnel viscosity, mud balance density, and filtrate loss. In general, mixtures up to approximately 1.5% corrugate content (baseline, V mixes up to 1.5%, PC mixes up to 1.0%, and PC(+) mixes up to 2% corrugate content, Table 4) maintain acceptable engineering properties. The corrugated board can be used to replace 9 to 27% (0.5 to 1.5% corrugate content in a 5.5% mixture) of the bentonite used in the slurry mixes. In addition, the PC2.0-P(+) slurry (2% corrugate) and other 2% corrugate slurry mixtures were close to the acceptable range and may be used based on specific site and construction conditions. Significant amount of this natural resource (i.e., bentonite) can be saved using the corrugated board, in consideration to large-scale construction projects.

The differences between virgin and post-consumer board were not significant with regard to performance in slurry mixes. The changes in engineering properties of the slurries with added corrugate content were attributed to the fibrous structure of the corrugate. Specifically, a fibrous matrix developed with sufficient addition of corrugate, which promoted more viscous, less cohesive behavior. This resulted in an

increase in Marsh funnel viscosity (up to 2.5 corrugate content beyond which Marsh funnel readings could not be obtained due to excessive bridging of the fibers in the testing device); a decrease in mud balance density (due to replacement of bentonite with the lighter fibers); and increase in filtrate loss and permeability (attributed to presence of sufficient fiber content to provide preferential pathways for flow).

Table 4. Slurry Test Results

Corrugate Type	Mix	MFV (s)	MB (kg/m^3)	FL (mL)	CT (mm)	k/t (s^{-1})	k (cm/s)
None	B5.0	38.3	1030	19	2.4	4.68×10^{-8}	1.12×10^{-8}
	B5.5	40.5	1030	16	2.7	4.89×10^{-8}	1.33×10^{-8}
	B6.0	49.2	1035	14	1.8	3.24×10^{-8}	5.93×10^{-8}
C-flute box	V0.5-C	42.4	1025	19	3.8	4.20×10^{-8}	1.58×10^{-8}
	V1.0-C	41.8	1010	17	3.1	4.74×10^{-8}	1.46×10^{-8}
	V1.5-C	47.1	1010	22.8	3.2	4.30×10^{-8}	1.36×10^{-8}
	V2.0-C	57.0	1010	26.3	4.8	5.14×10^{-8}	2.48×10^{-8}
	V2.5-C	NM	1015	32.8	8.3	4.71×10^{-8}	3.91×10^{-8}
	V3.0-C	NM	1010	39.8	10.8	6.92×10^{-8}	7.48×10^{-8}
	PC0.5-C	47.3	1025	20.6	3.2	3.80×10^{-8}	1.22×10^{-8}
	PC1.0-C	49.7	1020	20.1	3.5	4.38×10^{-8}	1.53×10^{-8}
	PC1.5-C	52.7	1010	21.6	3.6	3.72×10^{-8}	1.36×10^{-8}
	PC2.0-C	58.0	1010	26.2	4.2	5.02×10^{-8}	2.09×10^{-8}
	PC2.5-C	NM	1015	32.1	9.1	6.53×10^{-8}	5.92×10^{-8}
	PC3.0-C	NM	1015	36	11.0	8.00×10^{-8}	8.83×10^{-8}
	PC0.5-C(+)	46.5	1025	19	4.0	3.75×10^{-8}	1.49×10^{-8}
	PC1.0-C(+)	49.0	1015	20.7	3.0	4.08×10^{-8}	1.23×10^{-8}
	PC1.5-C(+)	49.4	1010	19.6	3.7	3.41×10^{-8}	1.26×10^{-8}
	PC2.0-C(+)	56.8	1010	25	5.5	4.25×10^{-8}	2.33×10^{-8}
	PC2.5-C(+)	79.1	1010	27.8	5.1	3.42×10^{-8}	1.74×10^{-8}
	PC3.0-C(+)	NM	1005	32.9	11.4	4.90×10^{-8}	5.58×10^{-8}
Pizza box	V0.5-P	44.44	1025	18.8	3.2	4.27×10^{-8}	1.35×10^{-8}
	V1.0-P	47.7	1020	20.1	4.0	4.07×10^{-8}	1.62×10^{-8}
	V1.5-P	49.0	1015	22.6	4.1	4.38×10^{-8}	1.80×10^{-8}
	V2.0-P	56.4	1010	24.6	5.7	4.80×10^{-8}	2.73×10^{-8}
	V2.5-P	NM	1010	31	8.0	6.19×10^{-8}	4.89×10^{-8}
	V3.0-P	NM	1010	33.6	10.5	7.48×10^{-8}	7.83×10^{-8}
	PC0.5-P	46.4	1025	20.1	3.1	4.13×10^{-8}	1.27×10^{-8}
	PC1.0-P	47.1	1020	21.1	3.4	4.02×10^{-8}	1.35×10^{-8}
	PC1.5-P	52.0	1010	23.2	4.4	4.61×10^{-8}	2.03×10^{-8}
	PC2.0-P	55.2	1010	24.6	6.2	3.22×10^{-8}	1.98×10^{-8}
	PC2.5-P	70.2	1015	28.7	8.2	4.37×10^{-8}	3.58×10^{-8}
	PC3.0-P	NM	1015	34.3	6.2	4.83×10^{-8}	3.00×10^{-8}
	PC0.5-P(+)	39.4	1025	16.8	2.2	3.44×10^{-8}	7.60×10^{-9}
	PC1.0-P(+)	38.5	1020	17.5	1.5	3.55×10^{-8}	5.23×10^{-9}
	PC1.5-P(+)	40.6	1010	18.8	2.7	3.65×10^{-8}	9.91×10^{-9}
	PC2.0-P(+)	48.1	1000	24	4.0	4.82×10^{-8}	1.93×10^{-8}
	PC2.5-P(+)	65.1	1015	30.2	8.0	6.96×10^{-8}	5.54×10^{-8}
	PC3.0-P(+)	NM	1015	31	10.7	7.66×10^{-8}	8.17×10^{-8}

MFV – Marsh funnel viscosity, MB – Mud balance density, FL – Filtrate loss, CT – Thickness of filter cake (t), k/t – Quotient of permeability of filter cake and thickness of filter cake, k – permeability of filter cake, "-C" – C-flute corrugated box, "-P" – Pizza box, "(+)"– sample subjected to additional blending time, NM – Not measurable due to flocculation and bridging in the Marsh funnel device.

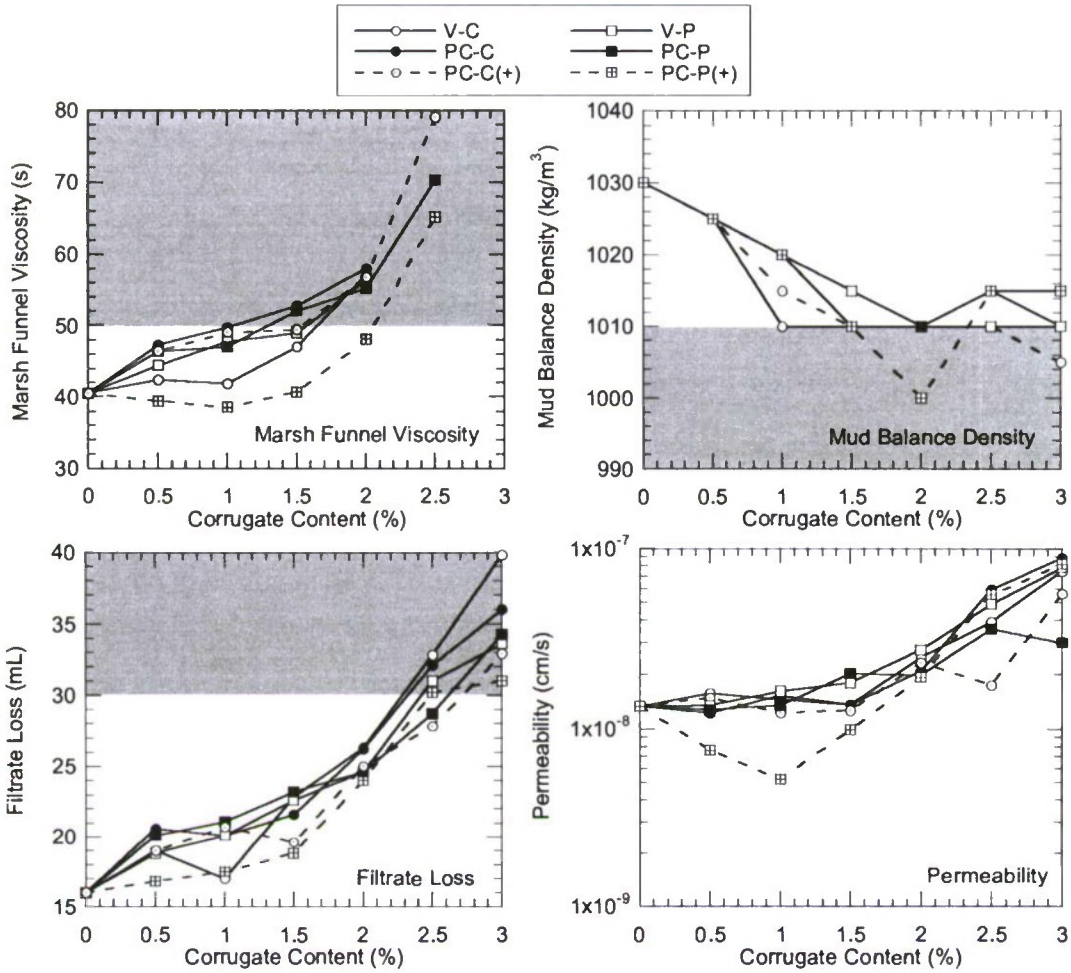


Fig. 2. Engineering Properties of Slurries

The additional blending time for the post-consumer samples had a more pronounced effect on engineering properties of the slurries containing pizza box fibers than the c-flute box fibers. The greater differences were attributed to the breakdown of greasy film on the pizza box allowing access to water and softening during the extended blending. In comparison, the c-flute fibers had already sufficiently broken down after 2 min. of blending and additional blending did not change the behavior significantly.

CONCLUSIONS

Tests were conducted to assess the feasibility of using corrugated board in slurry applications. Bentonite used in typical slurry mixtures was replaced by fiberized corrugated board at varying ratios. Properties of bentonite-corrugated board-water mixes were compared to baseline bentonite-water slurry mixes to evaluate the influence and practical limits of corrugated board addition to the mixes. The results indicated that the corrugated board could be used to replace 9 to 27% (corresponding to 0.5 to 1.5% corrugate content in a 5.5% mixture) of the bentonite used in the slurry

mixes based on Marsh funnel viscosity, density, and filtrate loss tests. Corrugated board may be used to replace up to 36% of bentonite (2.0% corrugate in a 5.5% mixture) for specific site and construction conditions requiring high MFV. In addition, permeability of the mixes with corrugated board was similar to baseline bentonite-water mix permeability. The differences in engineering properties of the slurries containing corrugate content were attributed to the presence of a fibrous matrix that influenced viscosity and flow characteristics. Overall, slurry applications provide a new and viable beneficial reuse alternative for paper / paperboard products, which constitute the largest weight and volume fraction of municipal solid waste generated and disposed of in the U.S. as well as other countries.

ACKNOWLEDGMENTS

This study was partially supported by Office of Naval Research (Award No. N00014-08-1-0855). Mr. Timothy Kershaw and Mr. Matthew Anderson assisted with laboratory testing.

REFERENCES

Boyes, R. G. H. (1975). *Structural and Cut-off Diaphragm Walls*, Applied Science Publishers Ltd., London, England.

D'Appolonia, D. J., "Soil-Bentonite Slurry Trench Cutoffs," *Journal of the Geotechnical Engineering Division*, ASCE. 106, GT4, 399-417.

Flaherty, R. (2009). *Personal communication*, President, International Corrugated Packaging Foundation, December 19, 2009.

ISTA (2009). International Safe Transit Association, "ISTA 2 Series Partial Simulation Performance Test Procedure," *http://www.ista.org/forms/1Aoverview.pdf*, last accessed December 26, 2009.

Paper Industry Association Council (2008). "2007 Recovered Paper Annual Statistics," *http://stats.paperrecycles.org/*, Paper Industry Association Council, last accessed May 6, 2008.

RecycleBank (2009). "The Pizza Box Recycling Mystery," *http://www.recyclebank.com/recycling/recycling-101/330-the-pizza-box-recycling-mystery*, last accessed December 17, 2009.

Ryan and Day (2003). "Soil-Bentonite Slurry Wall Specifications," *Proceedings, Soil and Rock America 2003, 12th Panamerican Conference on Soil Mechanics and Geotechnical Engineering*, P. J. Culligan, H. H. Einstein, and A. J. Whittle, Eds.

White, A. and Kendrick, J. (2009). Recycling Paper Food Packaging with Corrugated Cardboard: A Proposal for a New Recyclability and Repulpability Protocol and Preliminary Test Results for Fiber-Based Hot Cups, Global Green USA's Coalition for Resource Recovery, New York.

USEPA (2009). "Municipal Solid Waste in the United States – 2007 Facts and Figures." *http://www.epa.gov/epawaste/nonhaz/municipal/pubs/msw07-rpt.pdf*, last accessed August 26, 2009.

USEPA (1984). *Slurry Trench Construction for Pollution Migration Control*, EPA-540/2-84-001, Washington D.C.

Development of a Model for Large Scale Cultivation of Microalgae in an Airlift Photobioreactor

Principal Investigator:

Raul Cano, Ph.D., Biological Sciences
California Polytechnic State University
San Luis Obispo, CA

INTRODUCTION

The primary producers of oxygen in aquatic environments are algae, especially planktonic microalgae. These microorganisms are widely distributed in nature and have adapted to different environments with great diversity in size, morphology, life cycle, pigments, and metabolism. About one half of global photosynthesis and oxygen production is accomplished by marine microalgae. They play an important role in CO_2 recycling through a photosynthesis that is similar to higher plants in O_2 -evolved systems (PSI and PSII). Research in microalgae has been carried out not only on physiological aspects but also to develop production of useful biomaterials. The advantages of their utilization in production are (1) their ability to convert CO_2 to useful materials through photosynthesis (i.e., carbon sequestration) and (2) their ability to grow in natural environments under inorganic conditions. For example, marine microalgae can be cultivated using seawater, CO_2, and sunlight. Recent developments in the biotechnology of microalgae have been focused on their production of useful materials mainly applicable to the cosmetic and medical fields.

We proposed to develop a large-scale model to grow microalgae for:

1. Mass cultivation of microalgae;
2. Enhanced production of lipids for biodiesel; and
3. Production of useful chemicals (e.g., natural products, pigments, etc.)

PRODUCTION OF USEFUL CHEMICALS BY MARINE MICROALGAE

The cyanobacteria are oxygenic photosynthetic prokaryotes that show large diversity in their morphology, physiology, ecology, biochemistry, and other characteristics. Currently, more than 2,000 species are recognized, which comprise two fifths of known bacterial species (~5,000 species). Cyanobacteria contain chlorophyll a, phycobiliproteins, and characteristic glycosylated xanthophylls, such as myxoxanthophyll and oscillaxanthin [1 – 3]. The phycobiliproteins are water-soluble pigments consisting of red-colored phycoerythrin and blue-colored phycocyanin and allophycocyanin. They have a characteristic structural feature, the phycobilisome, that is used as a photosynthetic light-harvesting antenna. Cyanobacteria exist as unicellular, multicellular, colonial, and branched or unbranched filamentous forms. Specialized cells, heterocysts and akinetes, are contained in some of the filamentous-form cells. Some cells produce extracellular matrix such as sheaths, capsules, and slimes that consist mainly of polysaccharides. Cyanobacteria are distributed widely not only in salt water but also freshwater, brackish water, polar regions, hot springs, and deserts. Some also exist as symbionts in sponges, ascidians, echiuroid worms, planktonic diatoms, and dinoflagellates in marine environments [4] and lichens and azollae in terrestrial environments. Cyanobacteria, especially marine pelagic *Synechococcus* and *Prochlorococcus*, contribute largely to global oxygen production.

Many commercial applications have been proposed for marine cyanobacteria, although no marine strain presently is commercially supplied. Recent proposals to evaluate potential commercial uses typically fall into three categories: bioactive chemical compounds [5 – 10], polysaccharides [11 – 14], and evaluation of new genes for recombinant expression. Cyanobacteria can produce a large variety of complex chemical compounds. Bioactive com-

pounds isolated from marine cyanobacteria have recently been summarized by Burja et al. [5] and Takeyama and Matsunaga [15]. Novel plant growth regulators that promote redifferentiation, germination, and plantlet formation [16], tyrosinase inhibitors [17], UV-A absorbing compounds [18], sulfated polysaccharides showing anti-HIV activity, and novel antibiotics with light-regulated activity [19] are among the many compounds that have been studied. Matsunaga et al. [20] reported that several marine cyanobacterial strains, such as *Phormidium* sp. NKBG 041105 and *Oscillatoria* sp. NKBG 091600, showed high cis-palmitoleic acid content (54.5% and 54.4% of total fatty acid, respectively).

The discovery of biochemically active compounds from marine cyanobacteria, including enzyme inhibitors, herbicides, antimicrobials, multi-drug-resistance reversers, and antimalarial and immunosupressive agents, has dramatically increased over the last few years. This has been due to the adaptation and use of current cyanobacterial collections and cyanobacteria-derived compounds for screening in new pharmaceutical and industrial assays.

The focus on the polysaccharides of marine cyanobacteria as well as freshwater strains also has greatly increased in relation to interest in their exploitation for various industrial applications [11 – 14]. Cyanobacteria produce three types of extracellular matrix consisting mainly of polysaccharides, which have unique bio- and physicochemical characteristics. Most of them are composed of at least ten different monosaccharides and contain pentoses, which have not been observed in other prokaryotic polysaccharides. The anionic nature of these unique polysaccharides is due to the presence of acidic sugars and anionic organic and inorganic compounds. Little work has been devoted to potential applications of marine cyanobacterial polysaccharides. Extracellular polysaccharide production by cyanobacteria, as well as their possible applications, was reviewed by Philippis and Vincenzini [11] and Philippis et al. [13].

Marine cyanobacteria also are attractive as a resource for useful enzymes and genes [21 – 29]. Cyanobacteria commonly produce complex macromolecules that often possess biological activities such as cytotoxicity or microbial toxicity [5]. Recent genetic analyses have revealed that most of these macromolecules, as well as unusual small molecules, are coded for by large gene clusters representing nonribosomal peptide synthetase and polyketide synthetase [21 – 23, 25, 26, 28]. These gene clusters may be manipulated for the production of new chemicals. Many varieties of the gene clusters are present in most cyanobacteria [21, 24]. Therefore, cyanobacteria are attractive not only as producers of useful bioactive macromolecules and enzymes but also for production of complex macromolecules that may become important pharmaceuticals.

Microalgal Mass Cultivation Technologies

Photosynthetic microorganisms play an important role in the conversion of solar energy into chemical energy. Photosynthetic conversion is an efficient and alternative process used in several industrial fields. Attempts to develop large-scale methodologies for the cultivation of microalgae have been performed using many different kinds of cultivation systems for providing alternatives to fermentation and agriculture products. Algal biomass has historically served as fertilizer and a food source for both humans and animals for secondary waste water treatment and bioremediation. This use of algal biomass is an important consideration for industrial applications of microalgal cultures. With advances in processing technology, algal biomass has come to be seen as a possible source of fuels, fine chemicals, and pharmaceuticals. Several species of microalgae, which produce useful chemicals such as amino acids, vitamins, carotenoids (β-carotene), fatty acids (DHA, EPA, γ -LA 18:3 (n-6)), polysaccharides, and antibiotics have been reported. Many microalgal products have already been commercialized.

Further, microalgal production of energy resources has been extensively investigated. Development of processes that utilize the majority of the resulting microalgal biomass as energy sources would be preferred. Such processes may allow the recycling of evolved CO_2 from human energy consumption rather than direct emission, as is the present case for fossil fuels. The following six products for use as fuels can be produced from microalgal biomass: hydrogen (through biophotolysis), methane (through anaerobic digestion), ethanol (through yeast or other alcohol fermentation), triglycerides (through extraction of lipids), methyl ester fuels (through transesterification of lipids), and liquid hydrocarbons (from *Botryococcus braunii*). The development of efficient culture systems is necessary for algal mass production and the industrial applications of microalgae.

The growth rate and maximum biomass yield of microalgal strains are affected by culture parameters (light, temperature, and pH) and nutritional status (CO_2, nitrogen, and phosphate concentration). On the other hand, increasing the density of cultures decreases photon availability to individual cells. Light penetration of microalgal cultures is poor, especially at high cell densities, and such poor photon availability decreases specific growth rates. Higher biomass yields can be expected if sufficient photons are provided in high density cultures of microalgae.

Large-scale culture systems have been constructed (classified as open and closed systems) with the greatest attention directed to the light supply. Strains such as *Chlorella, Scenedesmus, Dunaliella, Spirulina, Porphyridium,* and *Haematococcus* have been cultured using photobioreactors to obtain several useful materials.

Large Scale Cultivation of Microalgae Using a Photobioreactor

Photobioreactors are closed systems and offer many advantages over open systems such as ponds or lagoons. The photobioreactors proposed for use in this study (Phycotech, Inc.) offer several advantages:

1. Facilitate maintenance of monoalgal cultures by protecting them from contamination;
2. Reduce water loss and the subsequent increase of salinity in the culture medium;
3. Increase productivity with greater cell densities, reducing harvesting costs;
4. Enhance exposure to solar radiation;
5. Reduce cell damage due to shearing and turbulence;
6. Provide for homogenous mixing and
7. Are applicable to various microalgal species under favorable culture conditions.

The production yield of algal biomass depends on the light path length to each cell, and therefore the surface-to-volume ratio is an important factor for efficient light utilization in photobioreactors. The productivity of photobioreactors is determined by the light regime inside the bioreactors. In addition to the light regime, oxygen accumulation and shear stress is limited with this design

In tubular reactors such as the Phycotech reactor, flow rates of 30 – 50 cm/s generally are used and airlift is the most effective circulation method of the culture rather than using centrifugal, rotary positive displacement, and/or peristaltic pumps. The main advantages of airlift systems are their low shear and relative simplicity of construction. The Phycotech tubular bioreactor optimizes light availability and exposure to light is uniform throughout the bioreactor

EXPERIMENTAL DESIGN

Summary of Experimental Design and proposed timeline

Task	Elements	Duration
1. Selection of strains	a. Identify candidate – from culture collection b. Purify and characterize strain	0.5 month
2. Bioreactor design	a. Determine optimal tube layout b. Orient bioreactor at the Cal Poly Pier for maximum solar exposure c. Optimize Flow rate and cell harvesting parameters for maximum biomass production	1.5 months
3. Medium selection in shake-flask studies	a. Identify optimal C/N/P ratio b. Evaluate role of CO_2 and aeration c. Identify nutrient amendments d. Determine optimal light intensity and aeration	1.5 month
4. Benchmark bioreactor operation	a. Implement design elements b. Test bioreactor in 200 L process c. Benchmark productivity	4 months
5. Write report		0.5 month

Selection of Microalgae: Microalgal strains in the public domain, and available for purchase from culture collections will be used for the benchmark studies.

Media selection and optimization Bioprocess Development will be conducted once the microalgal strain is selected and the medium composition is optimized for biomass production. This step is important because bioreactors offer much greater control of growth conditions than shake flasks. We used the Phycotech photobioreactor model 100L VSA (vertical serpentine airlift) photo bioreactor. The unit features PLC control with touch screen interface and automatic valves for fully automatic performance. Higher productivity at lower capital and operational cost are the key advantages of Phycotech's VSA photo bioreactor. Higher productivity is achieved by Phycotech's patent pending and proprietary technologies. Key factors are for the selection of the photobioreactor is:

1. **Productivity**
    a. Turbulent flow with low shear stress. Turbulent flow allows for efficient usage of available light and inhibits the algae from fouling the reactor wall. The low shear nature of the system does not damage the cells as traditional pumped system can.
    b. Periodic degassing. O2 concentrations are kept at very low levels. CO2 is readily delivered to the system and up take is optimized.
    c. Extended culture life. Prophylactic treatment of the culture limits the growth of contaminating organisms such as bacteria, fungi and virus.
    d. Sterilization in place (SIP) allows for the production of very clean algal cultures
    e. Cleanable. The system is very easy to clean.
2. **Lower capital and operating cost**

a. Novel use of low cost durable materials results in a much lower capital equipment cost compared to competing state-of-the-art systems.
b. Fully automatic operation and control significantly reduces operating costs.

The bioreactor was operated in batch mode to investigate biomass production. Cell density, pH, and biomass production were monitored as well as temperature and solar radiation.

RESULTS AND DISCUSSION

Strain selection

Chlorella vulgaris was selected as the benchmark strain. *Chlorella* is a non-motile, single-celled, fresh water green alga that is spherical in shape and measures 2 to 10 mm in diameter. *Chlorella* contains the green photosynthetic pigments chlorophyll-α and –β. *Chlorella* contains about 50% protein (on dry weight basis) and 20% each of carbohydrates and fats and 8-10 % ash. The selection was made based on the available comparative data from Dr. Bryan Hampson[1] using a similar bioreactor under control conditions. This information was useful in comparing the growth of *C. vulgaris* under condition of changing temperature and solar radiation.

Culture medium selection

Two culture media were tested as low cost substrate for large scale production of microalgae in the photobioreactor. The medium recommended by the University of Texas Collection of Algae, the source of the *C. vulgaris* strain used collection was the Proteose Peptone medium. This medium consisted of 1 g of proteose peptone dissolved in 1 L of Bristol salts solution consisted of $NaNO_3$ (2.94 mM), $CaCl_2 \cdot 2H_2O$ (0.17 mM), $MgSO_4 \cdot 7H_2O$ (0.3 mM), K_2HPO_4(0.43 mM), KH_2PO_4 (1.29 mM), and NaCl (0.43 mM). Medium # 2 consisted of Schultz all-purpose plant food (United Industries) at a concentration of 1 mg/L of ozone-sterilized city water.

Shake flask studies were conducted using an orbital shaker incubator at 200 RPM, at 25 °C for 7 days in the presence of a solar lamp. Cell densities were determined at the end of the experiment using a haemocytometer.

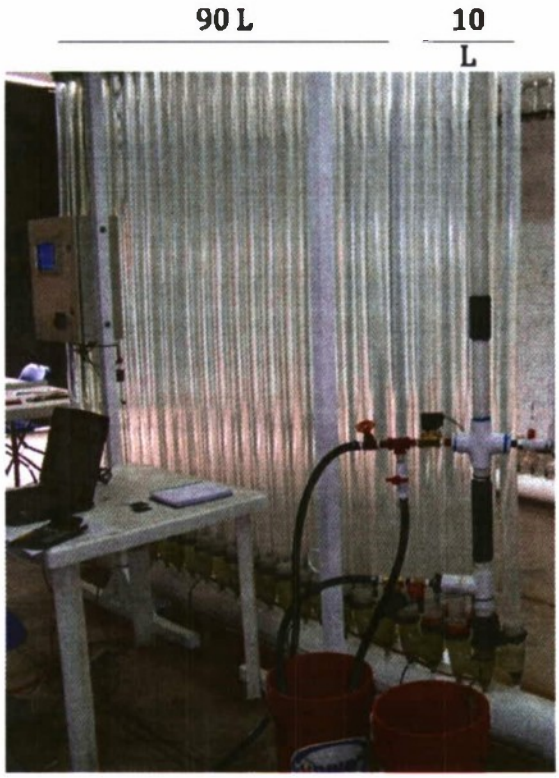
Both media performed similarly, yielding approximately 0.5 g/L of cell mass. Based on these results, Schultz plant food was used to supplement the city water in the bioreactor studies.

Bioreactor studies

The 100L Vertical Serpentine Airlift bioreactor (Fig 1) was constructed with a 10 L and 90 L partition. The microalgae in the 10 L partition served as the inoculum for the with 90 L bioreactor. Both partitions were filled with Schultz-supplemented city water and parsed periodically with CO_2, which served as both a carbon source and for pH adjustment. A 1 L *Chlorella* shake flask culture was used as seed for the 10 L bioreactor.

[1] Professor, Food Science and Nutrition, California Polytechnic State University, San Luis Obispo, CA 93407

Figure 1. *Vertical Serpentine Airlift (VSA) bioreactor used in the study*

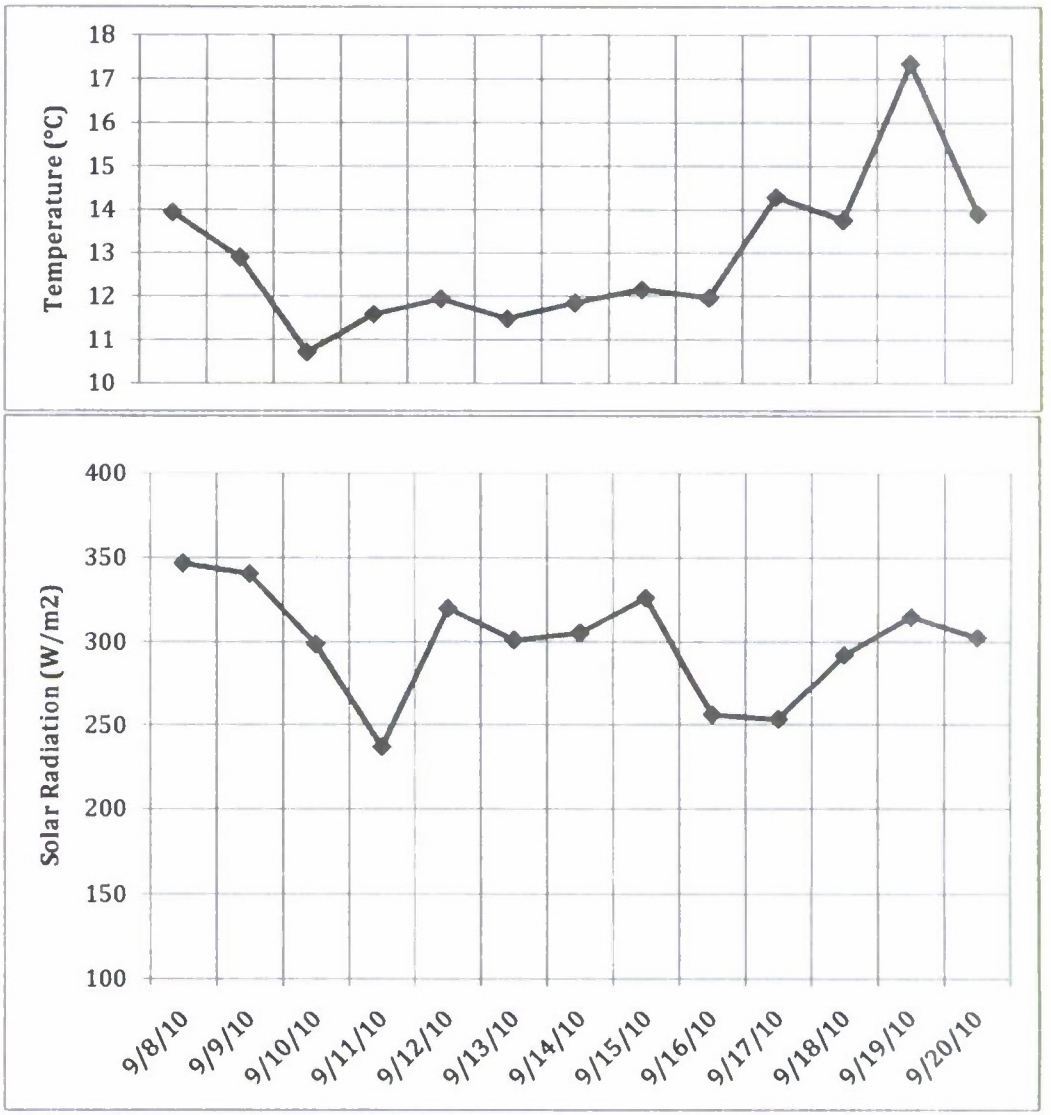


The study was conducted during a two-week period. Cell densities (in colony forming units (CFU)/mL, pH of medium and total dry weight (total biomass) yield were determined. During this period weather data, including solar radiation, barometric pressure and temperature were collected every 10 minutes. These weather data are summarized in Table 1 and Figure 2 below.

Table 1. *Mean solar radiation and temperature during the study period.*

Date	Mean Radiation (Watts/m^2)	Mean Temperature (°C)	Cell Counts
9/09/10	346.8	13.9	2,000,000
9/10/10	340.5	12.9	3,910,000
9/11/10	298.5	10.7	7,555,000
9/12/10	236.6	11.6	5,530,000
9/13/10	320.1	11.9	6,085,000
9/14/10	301.1	11.5	5,737,500
9/15/10	305.2	11.8	6,987,500
9/16/10	326.3	12.1	5,500,000
9/17/10	255.9	11.9	11,462,500
9/18/10	253.3	14.3	ND
9/19/10	292.1	13.7	ND
9/20/10	314.9	17.3	43,812,500
9/21/10	302.5	13.9	46,875,000

***Figure 2**: Mean Temperature and Solar Radiation During Study Period*



During the study period there were considerable variations in ambient temperature at the Pier as well as solar radiation, during which there were at least three days of long-lasting coastal fog. During this period of time, the temperature never reached the optimal range of 20 °C to 25 °C for various strains of *C. vulgaris*. The results of the bioreactor studies are illustrated in Figure 3. Regardless of the variable climatological conditions, the dry mass yield was approximately 2.1 g/L, The last count of 4.7×10^7 is approximately the same density that we achieved in the indoor system in Dr. Hampson's lab. Thus, the bioreactor performed comparably well under conditions of varying temperature and solar radiation, conditions that are controlled in indoor systems employing electricity and/or power to provide radiation and constant ambient temperatures.

***Figure 3**. Growth Curve for* Chlorella vulgaris *in the VSA Bioreactor at Avila Beach*

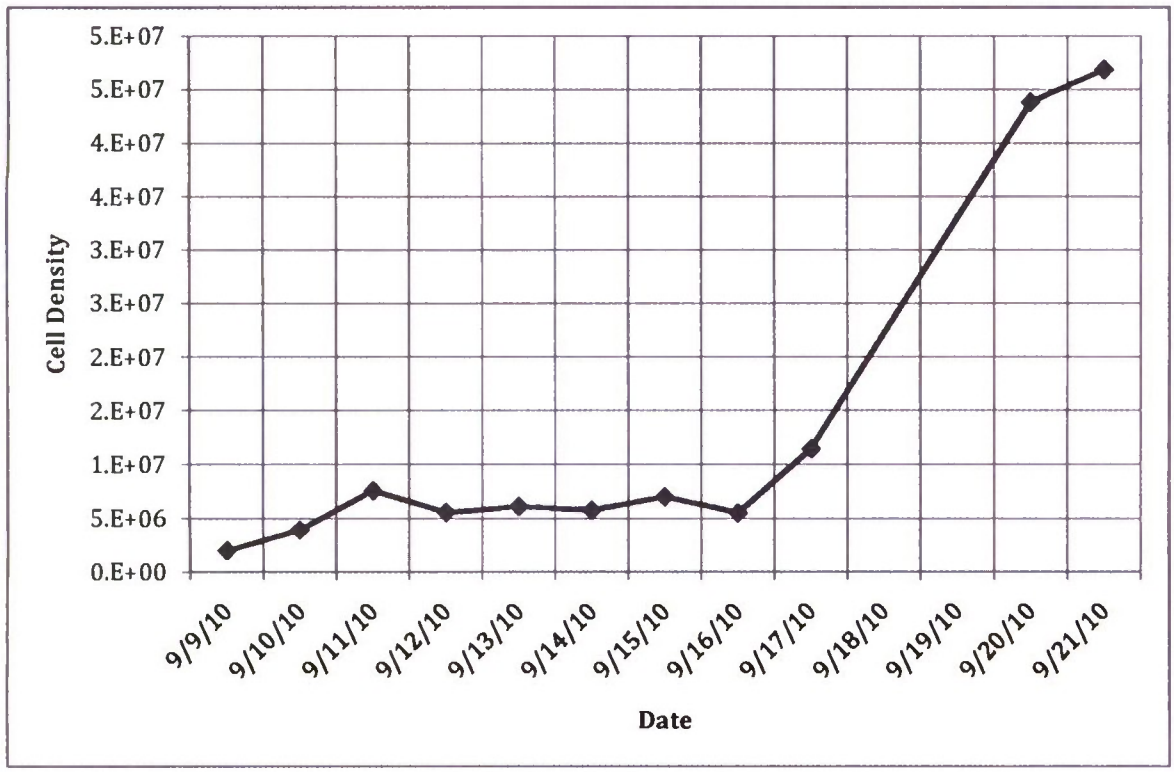


Figure 4. Bioreactor during Exponential Growth of C. vulgaris *at the Pier*

Specific Growth Rate

The growth rate of a microalgal population is a measure of the increase in biomass over time and it is determined from the exponential phase. Specific growth rate is one important way of expressing the relative ecological success of a species or strain in adapting to its natural environment or the experimental environment imposed upon it.

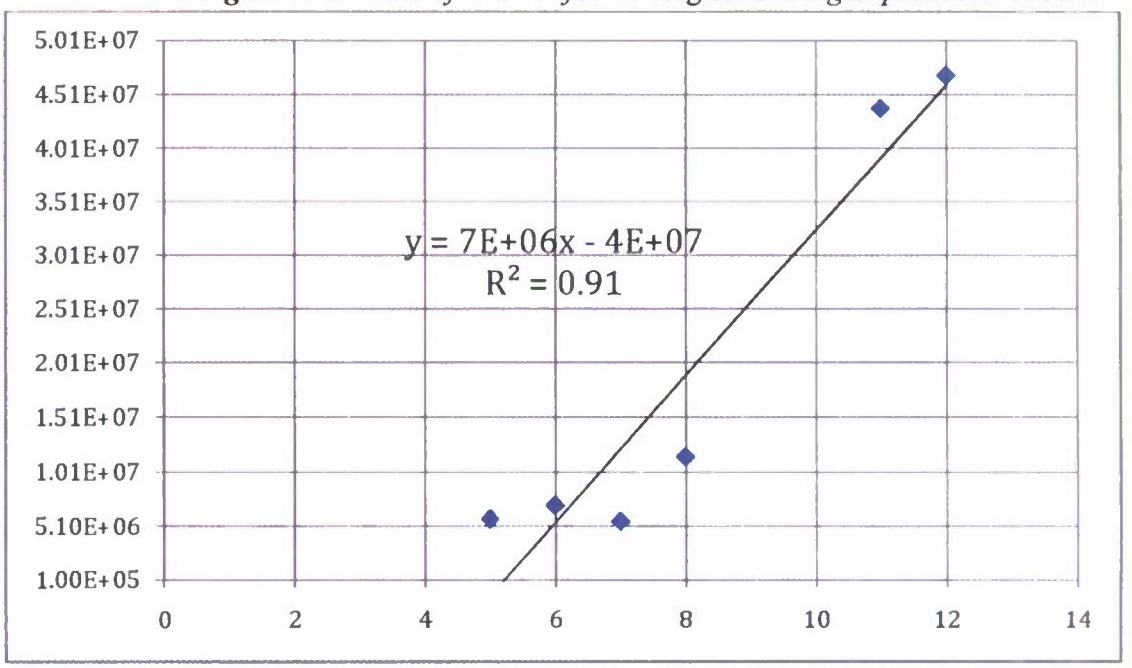
Specific growth rate is defined as the increase in cell mass per unit time, *e.g.*, cells per hour. The specific growth rate is commonly given by the symbol, μ (mu), and the most common units are in reciprocal hours (h^{-1}). μ is calculated using the following formula

$$\mu = \ln(m_{t2}/m_{t1})/t_1\text{-}t_2$$

Where m_t are the biomass at the different time points (t_1 and t_2), respectively

Specific growth rate was calculated from the portion of the growth curve representing exponential growth. This phase of growth was evident from 9/16/11 to 9/21/11. From these data it was possible to calculate the μ value for *C. vulgaris* under the growth conditions defined above. The specific growth rate was calculated (see Figure 5 below) as 0.52 d^{-1}. This value is well within those described in the literature (0.5 – 2 d--1 for *C. vulgaris* ("*The Ecology of Freshwater Phytoplankton*" by C.S. Reynolds, Cambridge University Press, 1984).

Figure 5. *Kinetics of Growth for* C. vulgaris *During Exponential Growth*



The duration of exponential phase in cultures depends upon the size of the inoculum, the growth rate and the capacity of the medium and culturing conditions to support algal growth. Inoculum size consisted of a bolus exponentially growing algae of approximately 10% of the

bioreactor's volume. Temperature and solar radiation, as well as other environmental parameters were those of the prevailing weather conditions at the time.

Of concern was the long duration of the lag phase, that period of time between bioreactor inoculation and the onset of exponential growth. The condition of the inoculum has a strong bearing on the duration of the lag phase as well as the culture conditions and medium pH. An inoculum taken from a healthy, exponentially growing culture is unlikely to have any lag phase when transferred to fresh medium under similar growth conditions of light, temperature and ionic concentrations.

In general the length of the lag phase will be proportional to the length of time the inoculum has been in stationary and death phase. A lag phase may also occur if the inoculum is transferred from one set of growth conditions to another. In this case, it was determined that the unusually long lag phase was due sub optimal to pH during that phase of growth and CO_2 feeding rates. Once the CO_2 feeding rate was stabilized with the subsequent pH stabilization, the culture resumed optimal growth.

The difficulty with pH monitoring and stability were due largely to CO_2 control and regulation. The system consistently overshot the set point under the weakest programmable conditions. In addition to control problems, leakage of CO_2 though the diffuser membranes were also responsible for the faulty pH regulation and subsequent long lag phase.

In spite of the technical difficulties, the vertical serpentine airlift bioreactor tested at the pier proved to be useful in generating large masses of microalgae and the results compared favorably with those in Hampson's lab under much more stressful conditions. With modifications of the pH monitoring and reprogramming of the control software it is possible to achieve specific growth rates > 1 d^{-1} and mass yields of 2.5 – 3 g/L.

RECOMMENDATIONS AND FUTURE DIRECTION

The results obtained during this limited study[2] suggested that the VSA bioreactor could be successfully used for the large-scale production of microalgae. Before this system can be routinely used for the industrial production of microalgae, however, a better system for drainage the reactor and harvesting the algal cell mass. In addition, potential problems from leakage and head loss can be prevented with better-designed components.

Sterilization of the culture medium prior to inoculation proved difficult initially, resulting in the contamination of the culture. To resolve this issue, an ozonator was installed to sterilize the medium and subsequent cultures appeared to be monocultures with no apparent contamination. We recommend that an ozonator be included as part of the bioreactor's components as media sterility is essential for the proper functioning of the bioreactor.

Also of concern is the control of CO_2 feeding into the bioreactor. As CO_2 is used as both pH control and carbon source, better control of the feed rate of this gas into the bioreactor can help optimize specific growth rates and bioreactor efficiency. The transparent LDPE (low density polyethylene) used in the tubular bioreactor is inexpensive and replaceable but it is prone to photo oxidation and breakage (even from flying sand particles carried in wind gusts).

[2] There were considerable delays (~ 6 months) in the manufacture and delivery of the bioreactor. There were also technical difficulties during assembly and operation as some of the components were defective or did not perform as expected. Once these issues were resolved, however, the bioreactor performed as advertised.

Finally, the airlift bioreactor depends on an air pump to provide the aeration required for the bioreactor contents to flow and maintain the algal culture in suspension. A solar panel attached to the bioreactor could serve as an inexpensive source of electricity to run the pump and other electricity-requiring processes.

The manufacturer of the photobioreactor has been working these past months on all new components and is very close to having them ready. These components include new tubes (more robust extrusion and new couples) and newly designed diffuser. At the same time, the manufacturer will bring its system control specialist to optimize the automation. The manufacturer is also committed to continue to work with the system and will cover expenses to operate after the completion of the project.

It recommended that we continue to work in the optimization of the *Chlorella* system as it is robust and it represents an important industrial organism, for as a source of medicinal products and for biofuels. It is also recommended that once the bioreactor is optimized in the *Chlorella* system, a process be developed for the cultivation of *Nannochloropsis* as it represents a genus of microalgae with much promise in the generation of lipids for the production of biodiesel.

Finally, as large-scale production of fine chemicals and other industrial products require the efficient, low cost processes, any reduction in the use of energy during the process would reduce production costs and thus increase the profit margin. Future process development of microalgal systems using this type of bioreactor should be conducted outdoors and use available light and temperature conditions rather than in temperature and light controlled rooms. Every effort should be made to incorporate solar or wind power to furnish the needed electricity to power the pumps and controllers. Additionally, CO_2 should be made available from environmental sources rather than from tanks. All these measures would reduce the carbon footprint of the process as well as reduce production costs.

References Cited

1. Graham L, Wilcox L (2000) Algae. Prentice-Hall, Englewood Cliffs, NJ
2. Hoek C van den, Mann D, Jahns H (1995) Algae: An Introduction to Phycology. Cambridge University Press, Cambridge, UK
3. Lee R (1999) Phycology. Cambridge University Press, Cambridge, UK
4. Raven J (2002) Biol Environ 102B:3
5. Burja A, Banaigs B, Abou-Mansour E, Burgess J, Wright P (2001) Tetrahedron 57:9347
6. Skulberg O (2000) J Appl Phycol 12:341
7. Mundt S, Kreitlow S, Nowotny A, Effmert U (2001) Int J Hyg Environ Health 203:327
8. Prokseh P, Edrada R, Ebel R (2002) Appl Microbiol Biotechnol 59:125
9. Han B, McPhail K, Ligresti A, Marozo V, Gerwick W (2003) J Nat Prod 66:1364
10. Volkman J (2003) Appl Microbiol Biotechnol 60:495
11. De Philippis R, Vincenzini M (1998) FEMS Microbiol Rev 22:151
12. Shah V, Ray A, Garg N, Madamwar D (2000) Curr Microbiol 40:274
13. De Philippis R, Sili C, Paperi R, Vincenzini M (2001) J Appl Phycol 13:293
14. Otero A, Vincenzini M (2003) J Biotechnol 102:143
15. Takeyama H, Matsunaga T (1998) Production of useful materials from marine microalgae. Vijay Primlani for Oxford & IBH Publishing, New Delhi

16. Wake H, Akasaka A, Umetsu H, Ozeki Y, Shimomura K, Matsunaga T (1992) Plant Cell Reports 11:62
17. Wachi Y, Sode K, Horikoshi K, Takeyama H, Matsunaga T (1995) Biotechnol Tech 9:633
18. Wachi Y, Burgess J, Iwamoto K, Yamada N, Nakamura N, Matsunaga T (1995) Biochim Biophys Acta 1244:165
19. Matsunaga T, Sudo H, Takemasa H, Wachi Y, Nakamura N (1996) Appl Microbiol Biotechnol 45:24
20. Matsunaga T, Takeyama H, Miura Y, Yamazaki T, Furuya H, Sode K (1995) FEMS Microbiol Lett 133:137
21. Neilan B, Dittmann E, Rouhiainen L, Bass R, Schaub V, Sivonen K, Borner T (1999) J Bacteriol 181:4089
22. Berg H, Ziegler K, Piotukh K, Baier K, Lockau W, Volkmer-Engert R (2000) Eur J Biochem 267:5561
23. Doekel S, Marahiel M (2001) Metab Eng 3:64
24. Christiansen G, Dittmann E, Ordorika L, Rippka R, Herdman M, Borner T (2001) Arch Microbiol 176:452
25. Chang Z, Flatt P, Gerwick W, Nguyen V, Willis C, Sherman D (2002) Gene 296
26. Oppermann-Sanio F, Steinbuchel A (2002) Naturwissenschaften 89:11
27. Hutchinson C (2003) Proc Natl Acad Sci USA 100:3010
28. Hoffmann D, Hevel J, Moore R, Moore B (2003) Gene 311:171
29. Brand, L.E. (1991) in Labeda, D.E. *Isolation of Biotechnological Organisms from Nature*,. pp81-116, McGraw-Hill, New York
30. Plackett, R. L., J. P. Burman: The design of optimum multifactorial experiments. *Biometrica* 1944, **33**, 305–325

Battery Laboratory Development

Principal Investigator:

Art MacCarley, Ph. D.
Electrical Engineering
California Polytechnic State University
San Luis Obispo, CA

Objective:

The overall objective of this work is to provide students and faculty with a laboratory to investigate various modern battery applications. In particular the possibility of electric vehicles that can share power with the electric power grid is of future interest. The effects of such sharing would involve new battery power testing that has not been investigated in the past. In particular the effect of power sharing on battery life has not been determined. The potential value of power sharing must be balanced by the cost of battery replacement.

Approach:

The general approach is to test small cells that might be used in vehicle batteries, and to apply these results to scaled-up batteries. In general, it is assumed that the deterioration of larger batteries will involve mechanisms that manifest themselves in cell testing, and that cell testing is much more economical than full scale battery tests. The approach has three main thrusts:

1) Laboratory Configuration
2) Development of Computer Controlled Test Stand
3) Single-Cell Lithium-Ion Charger and Gas Gauge Using the STw4102 Development and Testing

The budgets for this program allow for purchase of materials only. Voluntary labor was donated by Dr. John Dunning, Cal Poly Research Scholar in Residence, and students Daniel Forbes, Dan Lomeli, Alan Swe and Alvin Hilario. Other contributors include Electrical Engineering technical staff Jaime Carmo and Tony Sawa who provided computer and general technical support.

Summary of Progress:

Laboratory Configuration

Overhead data and power utilities were installed in the laboratory via overhead cable trays. Internet connection was installed and two computers were installed equipped with LabView 8.6 software for computer control of the experiments. The materials were used in support of the development of a battery test laboratory that could flexibly test small cells over pre-programmed power profiles such as those encountered in electric vehicles, hybrid electric vehicles and grid connected EV and HEV systems. In order to achieve such a facility a computer controlled lab set up was devised. A room in the Electrical Engineering Department (Room 20-130) was configured per the floor plan below:

Equipment Purchase

The Battery Laboratory Development program utilized funding to purchase materials listed below:

BK Programmable Power Supply
Integrated circuits
Batteries and circuit boards
BK Programmable Electronic Load
Tool Drawers and Tools
Electronic Parts for Digital Control Box
PC Board Fabrication
Miscellaneous parts for Detector PC Board
Connectors, Switches, Hardware
Agilent Data Acquisition System

Computer Controlled Test Stand

This phase of the project consisted of two efforts, namely the development of a Digital Control Box and the development of a Computer Software Interface. These are summarized by the students involved in the next sections.

Digital Control Box Report

Prepared by Alan Swe (aswe@calpoly.edu), Electrical Engineering Student

Background

There are two main phases in testing batteries, namely charging and discharging. The charging and discharging of batteries is normally handled by two different circuits or devices. For example, a current or voltage source is used for charging the battery and a resistive load for the discharge phase. Tests that involve multiple charge and discharge cycles require switching from one device to another after each phase, while monitoring the state of the battery from its current and voltage measurements.

Objective

The goal of this project is to create a control box that will allow a battery that is to be tested to interface between two devices. Ideally, one device will be for charging, and the other for discharging the battery. The user controls when each device will be connected to the battery either manually by using the toggle switches, or automatically through a software-controlled switching module like the Agilent 3497- series.

The box outputs battery current and voltage, which can be measured and monitored using a DAQ such as that which, again, the Agilent 3497- series has. Controlling the Agilent using LabView allows the user to control which device is active.

Specifications

The following section describes the hardware and software specifications for the control box.

Hardware

Currently the control box is designed to handle 12VDC at a maximum of 50A. This is the maximum rate at which the battery can be charged or discharged.

Figure 1 below shows the circuit diagram of the control box, with each black/red terminal representing a binding post for banana plugs.

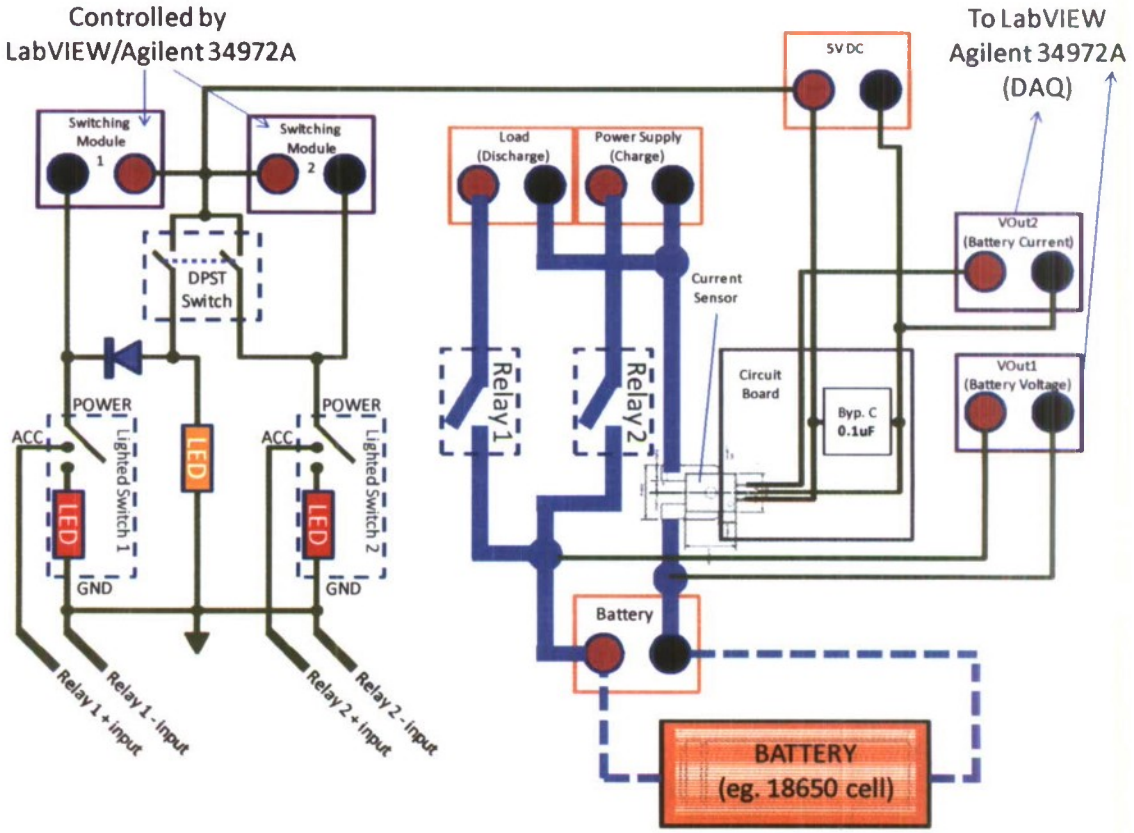


Figure 1: System Schematic

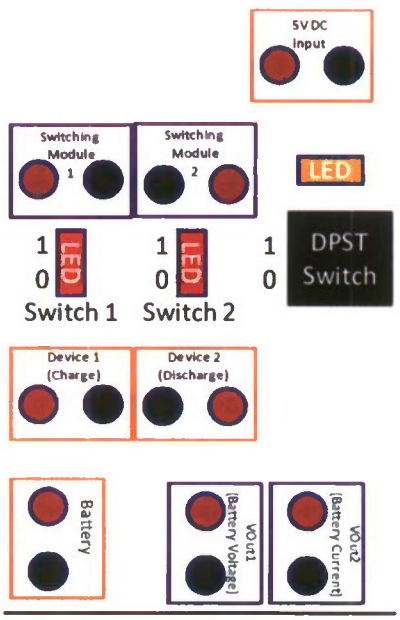


Figure 2: Outer Connections of Control Box

The current sensor used is an Allegro ACS756 Hall-Effect Current Sensor <http://www.allegromicro.com/en/Products/Part_Numbers/0756/0756.pdf> which measures currents from +/-50 Amps and outputs a proportional voltage from 0-5V, with 2.5V representing 0A when properly calibrated. Figure 3 below shows some data gathered from testing the current sensor.

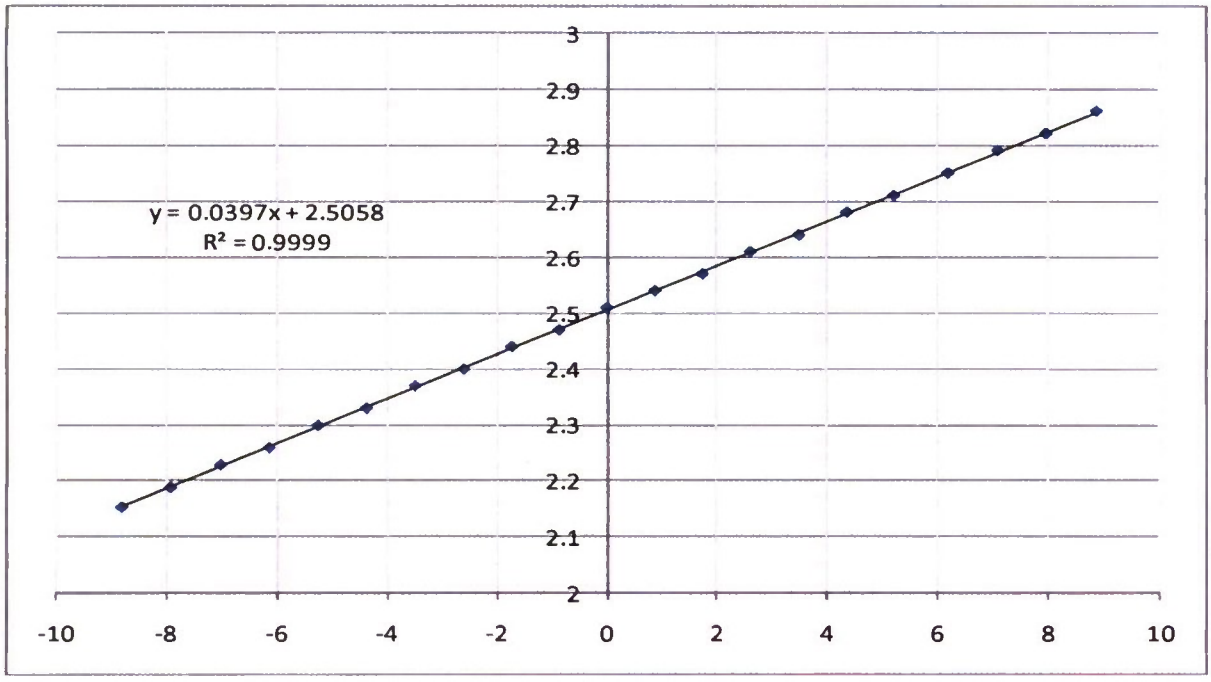


Figure 3: ACS756 Current Sensor I vs. V_{OUT}

Software

The control box is designed to be use in conjunction with software to interface between the battery being tested and the devices connected for charging and discharging. See document "Battery Life Study LabView Progress Report 2-23-11" by Alvin Hilario for more information on the setup of the software used for this project.

Procedure

The following section provides instructions for setting up and using the control box. Reference is made to Figure 2 above.

Connections

The control box requires a 5VDC supply to power lighted switches, LED indicators, relay inputs, and the current sensor.

Inputs

The inputs are designed to be connected from a switching module, such as the one the Agilent has, to Switching Module Inputs 1 & 2. Devices 1 and 2 connect to devices that will either charge or discharge the battery.

Outputs

The outputs Vout 1 and 2 (Figure 2) are to be connected to a DAQ for voltage and current measurements respectively. As current is derived from a proportional voltage output, the reading from Vout 2 has to be adjusted through software. "Battery" connects to the battery. More information can be found in the section "Specifications – Software" above.

Automatic Mode

In this mode, the software controls which device connects to the battery. The DPST switch is set to '0', and the orange LED indicator light is off to indicate automatic mode operation. The inputs Switching Module 1 & 2 are connected to a switching module such as that of the Agilent 34972A, which can be controlled using software such as LabView (see "Specifications = Software" section above). As the opening and closing of the relay is handled by the software, both Switch 1 & 2 are to be left in the "ON" position.

Note: In automatic mode, Switch 1 & 2 will not light up in the "ON" position unless the relay connection ot the respective Switching Module input is closed os well.

Manual Mode

In this mode, the user has to manually switch which device is currently connected to the battery using Switch 1 and Switch 2. DPST in the "ON" position and orange LED lighted indicates operation in manual mode. The user determines which device connects to the battery by switching on the respective switch, which will light up to indicate the respective device is currently connected.

Note: User hos to ensure Switch 1 ond Switch 2 ore not "ON" ot the some time in monuol mode.

Safety/Additional Notes

Below are safety precautions to take when using the box, some of which have already been mentioned in the above procedure.

1) Care has to be taken that both device 1 and 2 are not connected to the battery at the same time. This might overload the battery or the components of the box.
    a. Ensure Switch 1 and Switch 2 are in the "OFF" position before switching from automatic to manual mode (DPST => '1').
    b. In manual mode, when switching devices, turn off current device (both switches "OFF") before turning on the other device.

2) Do not exceed maximum voltage and current ratings (12V @ 50A) of control box. Also note maximum ratings of devices being used and battery being tested, which may be lower than that of the control box and be sure not to exceed them as well.

Battery Life Study –LabView Interface

Prepared by Alvin Hilario, Electrical Engineering Student

In the first phase of interfacing LabView with the battery test fixture, the Agilent 34972A Data Acquisition/Data Logger Switch Unit, DAQ, is used to read the battery voltage and current. The test fixture utilizes the ACS756 Hall Effect current sensor to output a voltage proportional to the battery's current. The Agilent 34901A 20 Channel Multiplexer Module plugged into the DAQ to take measurements on the battery voltage on channel 101 and the current on channel 102.

In the second card slot of the DAQ, the program uses the Agilent 34903A 20 Channel Actuator/GP Switch Module to control the relays for the BK Precision XLM3640 DC Power Supply, channel 201, and the BK Precision 8518 2300W Programmable DC Electronic Load, channel 202. This allows the program to control whether or not the battery is charging or discharging.

Figures5 and 6 show the LabView program that handles the basic control of the DAQ. In this program, there are two main functions that continually run as long as the 'Read' toggle switch is on in the front panel as shown in Figure 1. This switch is directly connect to the stop trigger of the while loop the two functions reside in.

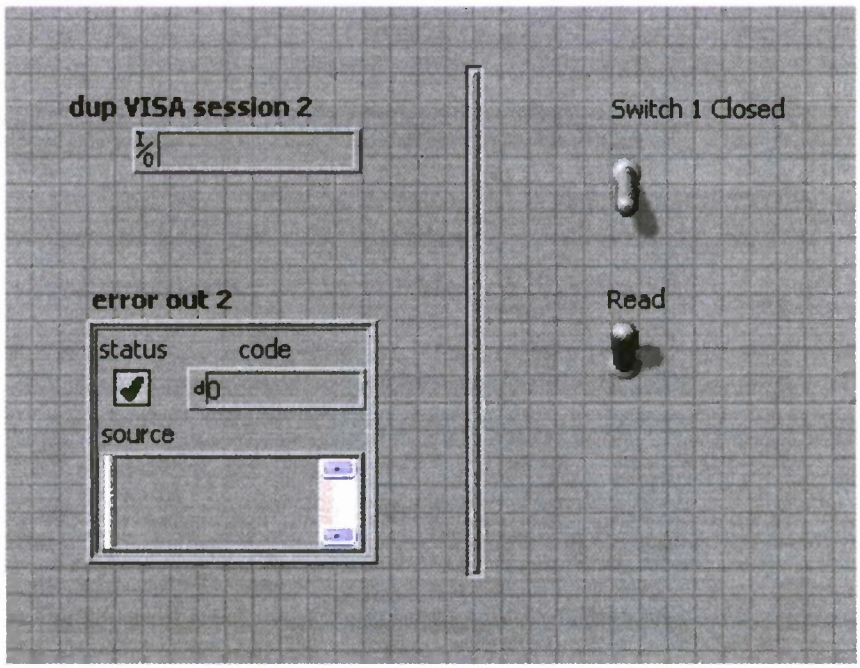


Figure 4: LabView Front Panel

In the first block in the while loop in Figure 5, the program reads the voltages taken from channel 101 and 102 of the multiplexer module and appends them into a spreadsheet in their respective columns. Due to the nature of interfacing LabView with the DAQ, there is a 100ms delay between the voltage and current readings.

The second block in the while loop first opens the switch designated by the 'Switch 1 Closed' toggle switch, waits 10ms to avoid having the power supply and electronic load connected at the same time, then closes the other switch. For example in Figure 5, the case structure recognizes the switch is on, true. It opens channel 201, waits 10ms, and closes channel 202.

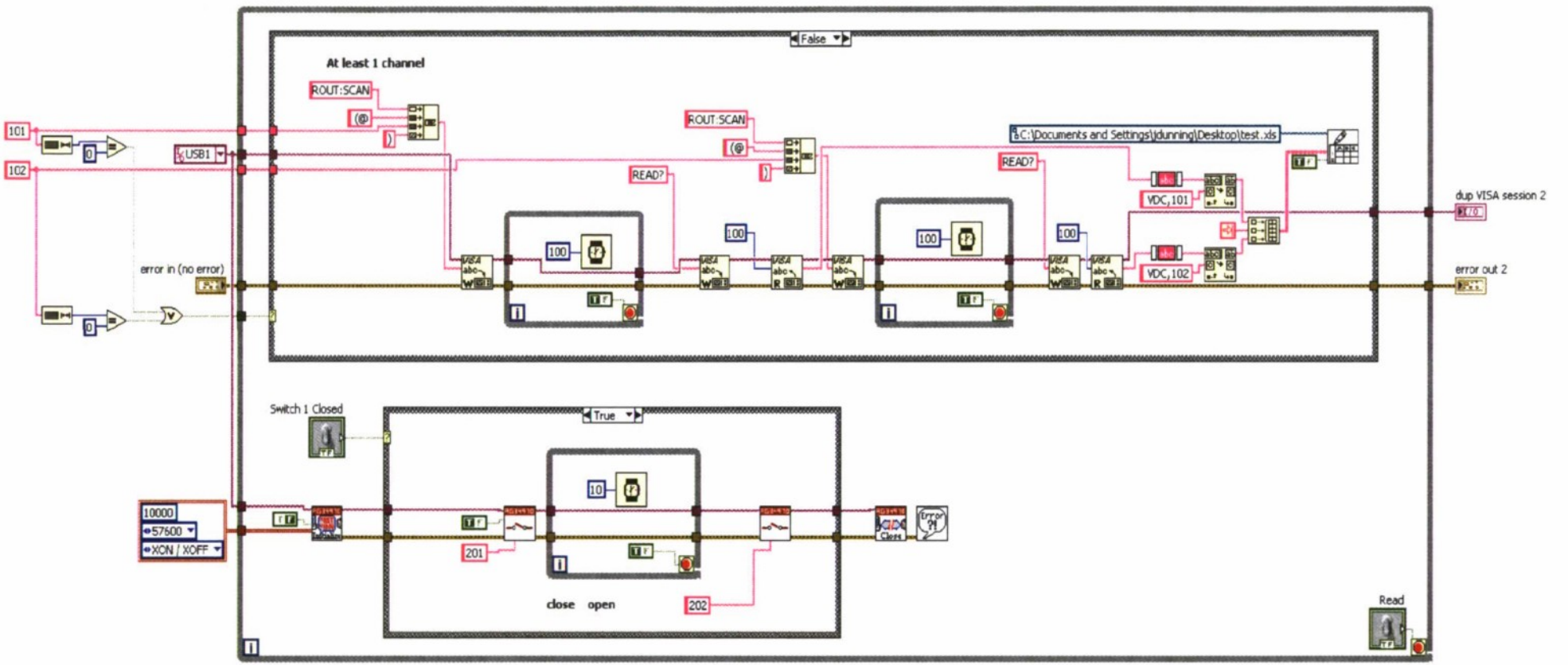


Figure 5: LabVIEW Block Diagram 1

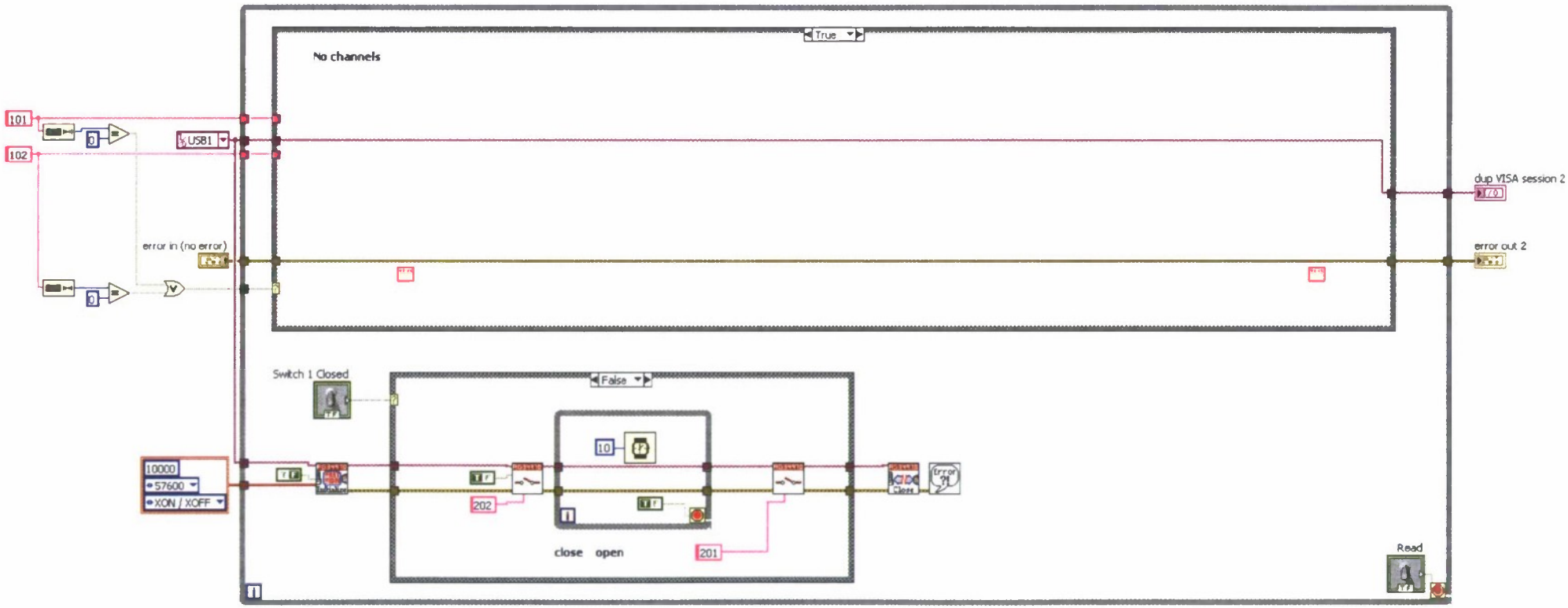


Figure 6: LabVIEW Block Diagram 2

Next Phase

In the next phase of the project, the program will need to interface to the power supply and electronic load. To charge the battery, the power supply must provide a constant current till it reaches a certain state of charge, and then provide a constant voltage to top off the remaining charge of the battery. This requires an algorithm to determine when to switch from constant current to constant voltage mode from the state of charge. The electronic load will need to be controlled for various discharge rates in the battery to grid regulation profile.

After those components are successfully interfaced, regulation profiles must be generated to specify exactly when the battery is charging and discharging, and at what rate. With these profiles, a sequence structure block can run through each stage of the profile and dictate which state the program will run in using a LabView state machine case structure.

At the moment, the DAQ takes in all data, but it will need an algorithm to selectively store only useful measurements. Once the program is complete, the life cycle study will commence and it will require data analysis and comparison.

Future Plans

The following are some additions planned to improve the design, function, and safety of the box:

1) Reverse-current and over-current protection for components and devices.
    a. Addition of a 50A fuse to protect components should the current rating be exceeded (for battery/device connections).
2) Inverting LED to indicate automatic operation. Currently, orange LED lights up in manual operation, but it makes more sense to have it light up for automatic operation as there is currently no indication that the device is in operation during automatic mode as Switching Module input (Figure 2) has to be closed for respective Switch to be lighted.
3) Cooling fan (5-12VDC CPU fan) can be added to cool the system components down. This can help reduce heat of the relay should the system be subjected to high-power testing.

A separate study was undertaken to develop test procedures, hardware and software for a lithium ion battery control circuit, which will be very useful in controlling single cells and later in controlling the behavior of a full pack of series connected cells. This is detailed in final section of this report.

Single-Cell Lithium-Ion Charger and Gas Gauge Using the STw4102

This Section was adapted from a separate report prepared by Daniel Forbes, Electrical Engineering Student

Overview

This project began as an investigation into charging and monitoring a single-cell lithium-ion battery. The idea was to become familiar with these operations, and then apply them to a multi-cell battery pack useful for an electric vehicle. A secondary purpose was to provide a simple demonstration tool for the battery test laboratory currently being set up by Dr. John Dunning at California Polytechnic State University, San Luis Obispo. To gain a hands-on understanding, a development board was developed based on ST Microelectronic's STw4102 integrated circuit, which provides charging and gas gauging functionality on a single IC. The development board, combined with a personal computer, provides a means to charge, discharge, and monitor a single Li-ion battery and to gather data on the cell.

A special thanks to Dr. John Dunning for providing invaluable information and support throughout this project, and for his continuous encouragement.

Hardware

This demonstration board consists of two main components, the STw4102 battery charger and gas gauge from ST Microelectronics, and the PICLF2620 microcontroller from Microchip. The STw4102 handles the constant-current constant-voltage (CCCV) charging of the cell, as well as the gas gauging via a sense resistor and clock signal (32.768 kHz). The microcontroller tells the STw4102 what to do, and it also reads data from it and outputs this data to the serial port. The charge and discharge cycles are initiated by the user (discussed in more detail in the software section) via push-buttons. LEDs display charge/discharge status to the user, as well as board power status (also discussed in more detail in the software section). This demonstration board has the ability to drain a battery (1 A limit) through a 4.7 ohm power resistor. Figure 7 below identifies the main components of the board.

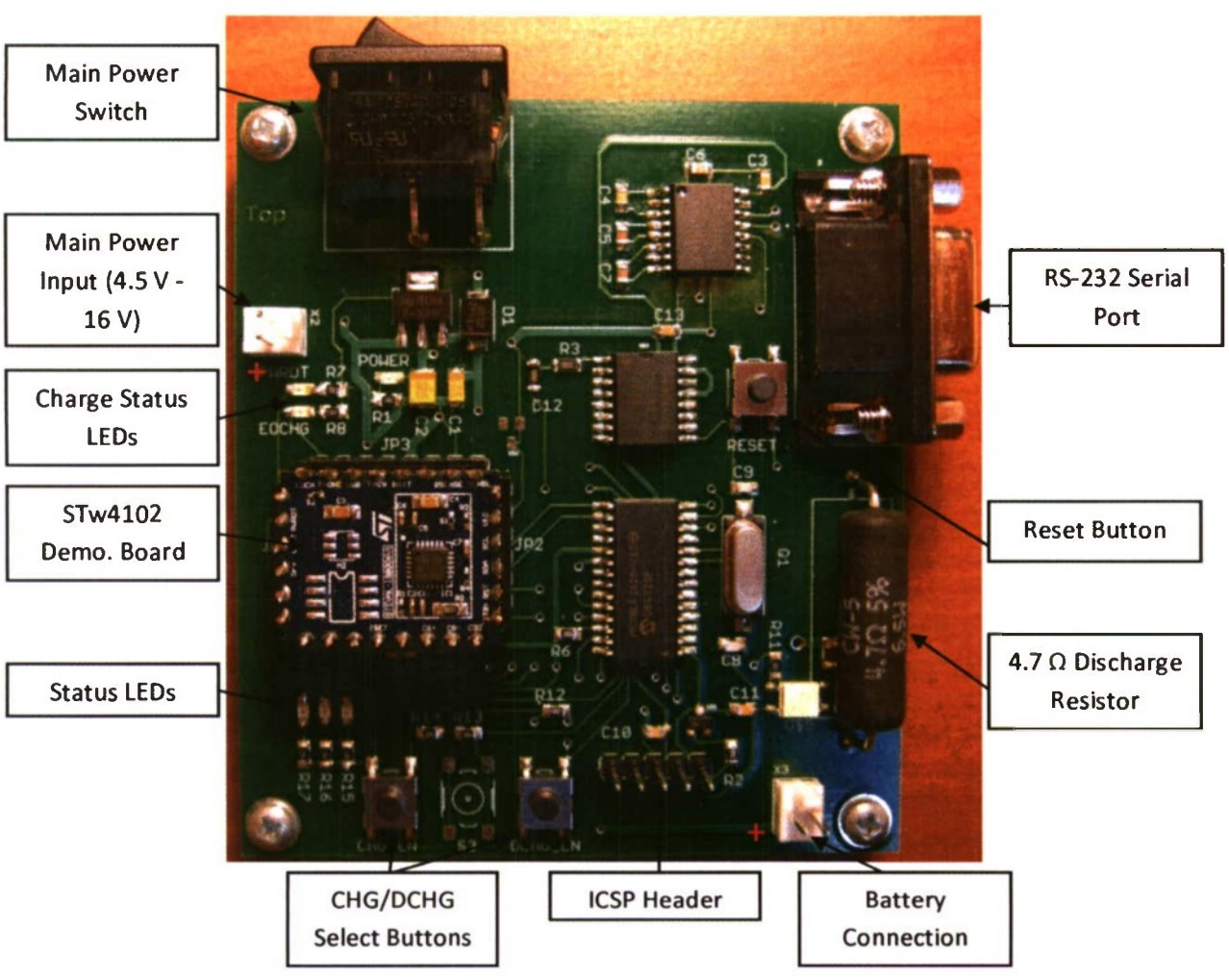


Figure 7 - Hardware Setup with Labels

The microcontroller can be reprogrammed via the in-circuit serial programming (ICSP) header. This allows for easy modification to the system software. The microcontroller runs at a clock speed of 20 MHz, and the board voltage is 3.3 V (from an on-board LDO regulator). When the board is powered down or reset via the RESET button, all charge/discharge data is cleared from the STw4120. For a detailed description of the STw4102 operation, please see the datasheet in the appendix.

The serial port communicates at 9600 bps, 8 data bits, 1 stop bit, no parity, and no flow control, and interfaces with a simple terminal program running on a personal computer. The output data is discussed in more detail in the software section of this document.

Please see the appendix for a schematic, board layout, and bill of materials (BOM).

Software

The embedded software in this system enables the user to perform battery charging and discharging with the press of a button. While undergoing one of these cycles, the software outputs cell data to the

serial port, which can be read via a simple terminal program on a personal computer (9600 bps, 8 data bits, 1 stop bit, no parity, and no flow control). This data can then be imported to a spreadsheet and manipulated as necessary.

There are three main states that the user can select: charge, discharge, and standby. When the system is first turned on, the controller enters standby mode. In this mode, the battery is resting (neither charging nor discharging), and none of the three LEDs are lit. From here, the user may select charge or discharge mode (via the CHG_EN and DCHG_EN buttons, press and hold for ~1 second, then release), and the corresponding LED with light. Also, in charge mode, the orange LED near the STw4102 board will light. In charge (discharge) mode, if the CHG_EN (DCHG_EN) button is pressed again, the system will go to standby, and if the DCHG_EN (CHG_EN) button is pressed the system will go into discharge (charge) mode. In charge mode, if the battery voltage goes above 4.2 volts, the controller will terminate the charge to protect the battery. Similarly, in discharge mode, the controller will terminate the charge when the battery voltage drops to 2.75 volts.

In any of the modes mentioned above, the controller outputs battery data to the serial port in ASCII characters. The data is sampled approximately every 1.5 seconds, and is grouped into a single line. The data consists of some raw count data from the STw4102 registers, and some of these counts are converted into physical units for readability. For accuracy, the raw counts should be used when further processing the data. The following data is output, with each value separated by a comma:

Battery voltage counts (each count is 1.42 mV)

Battery voltage (volts)

Charge counts (each count is 49.5 nAh)

Charge (mAh)

Discharge counts (each count is 49.5 nAh)

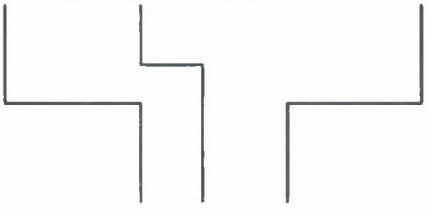
Discharge (mAh)

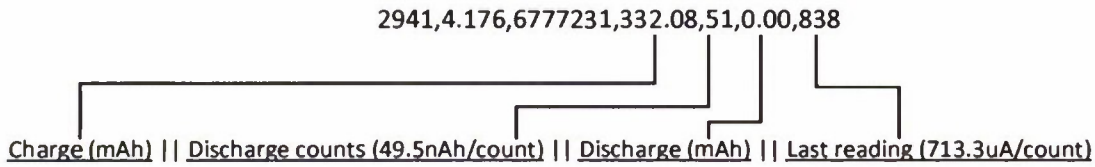
Last reading counts (each count is 713.3 uA, '+' denotes current into the battery, '-' denotes current out of the battery)

All of the units above correspond to the conversions given in the STw4102 datasheet (assuming a 33 mΩ sense resistor).

Below is a sample output from the terminal window during a battery charge:

Voltage counts (1.42mV/count) || Voltage (V) || Charge counts (49.5nAh/count)





While in discharge mode, the large power resistor (black, located next to the serial port) will become hot (it is dissipating about 2-3 W), and should be cooled with a small fan (recommended).

When the system is first powered up, or when the reset button is pressed, all the Stw4102 registers are reset to zero.

For more details on the software code, please see the comments in the attached code found in the appendix.

Cell Testing

To verify the correct operation of the charger/gas gauge, two cells were put through a series of tests. At the start of the test the cells were approximately half way charged. The test steps are listed below:

Cell #1

1. Charge cell at 4.2 V until the current is 10 % of charge current (10 % of 1 A = 0.1A).
2. Discharge cell through 4.7 ohm resistor until cell voltage reaches 2.75 volts.
3. Charge cell at 4.2 V until the current is 10 % of charge current (10 % of 1 A = 0.1A).
4. Discharge cell through 4.7 ohm resistor until cell voltage reaches 2.75 volts.
5. Charge cell at 4.2 V until approximately 375 mAh in battery (half full).

Cell #2

1. Discharge cell through 4.7 ohm resistor until cell voltage reaches 2.75 volts (cell is approximately half charged from manufacturer).
2. Charge cell at 4.2 V until the current is 10 % of charge current (10 % of 1 A = 0.1A).
3. Discharge cell through 4.7 ohm resistor until cell voltage reaches 2.75 volts.
4. Charge cell at 4.2 V until approximately 375 mAh in battery (half full).
5. Discharge approximately 375 mAh through 4.7 ohm resistor.

The above tests were performed using lithium-ion 14500 cells from BatterySpace.com (part number LC-14500). The nominal capacity of the cells is 750 mAh with a nominal voltage of 3.7 V. The specifications sheet is attached in the appendix.

Results

From the tests described in the cell testing section above, the operation of the demonstration board was verified. Figures 8, 9, and 10 below show a sample of the charge and discharge data gathered during these tests. In all of these figures, the x-axis counts the number of samples taken, with each sample taking approximately 1.5 seconds. Figure shows an entire charge starting with a cell that was discharged until the voltage was 2.75 V. In this particular charge, 745 mAh were put into the cell. Figure shows the beginning of the same charge, where the constant-current constant-voltage can be seen. In this charge and others, the constant-current portion of the charge appears shorter in time than expected. This may be due to resistance introduced between the terminals of the battery and those of the STw4102 by a bad connection. This will be investigated in future experiments.

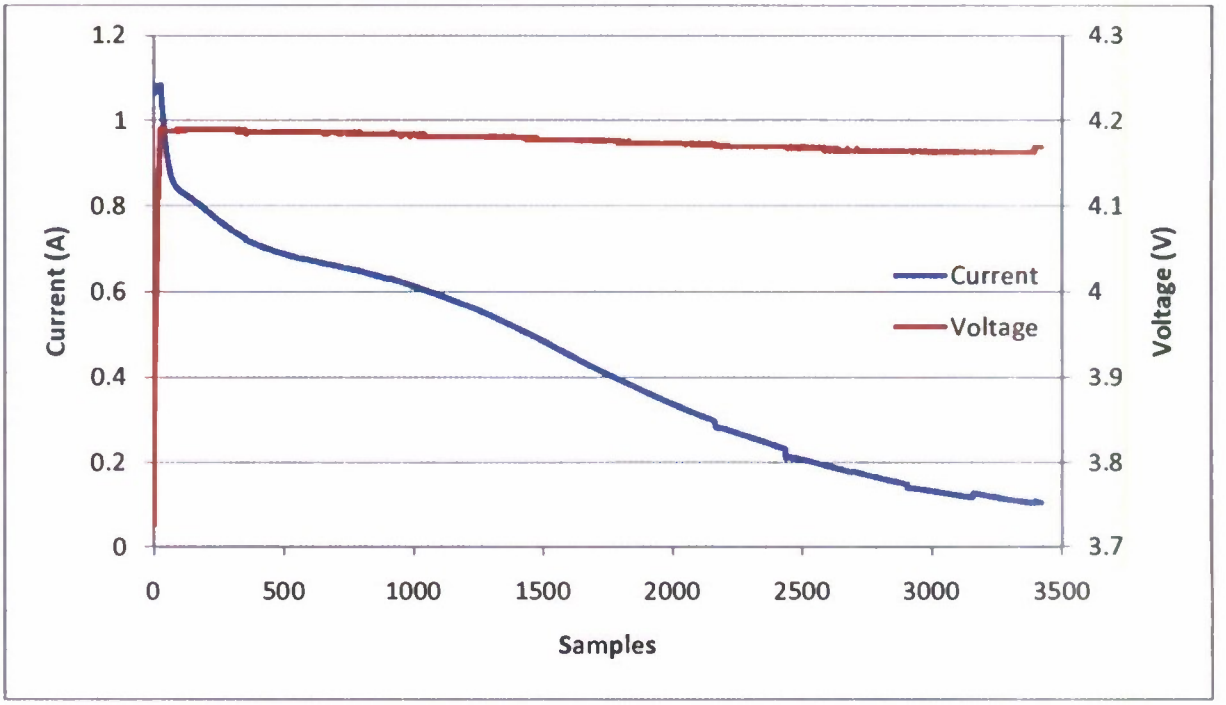


Figure 8 - Sample Charge of Li-Ion Cell

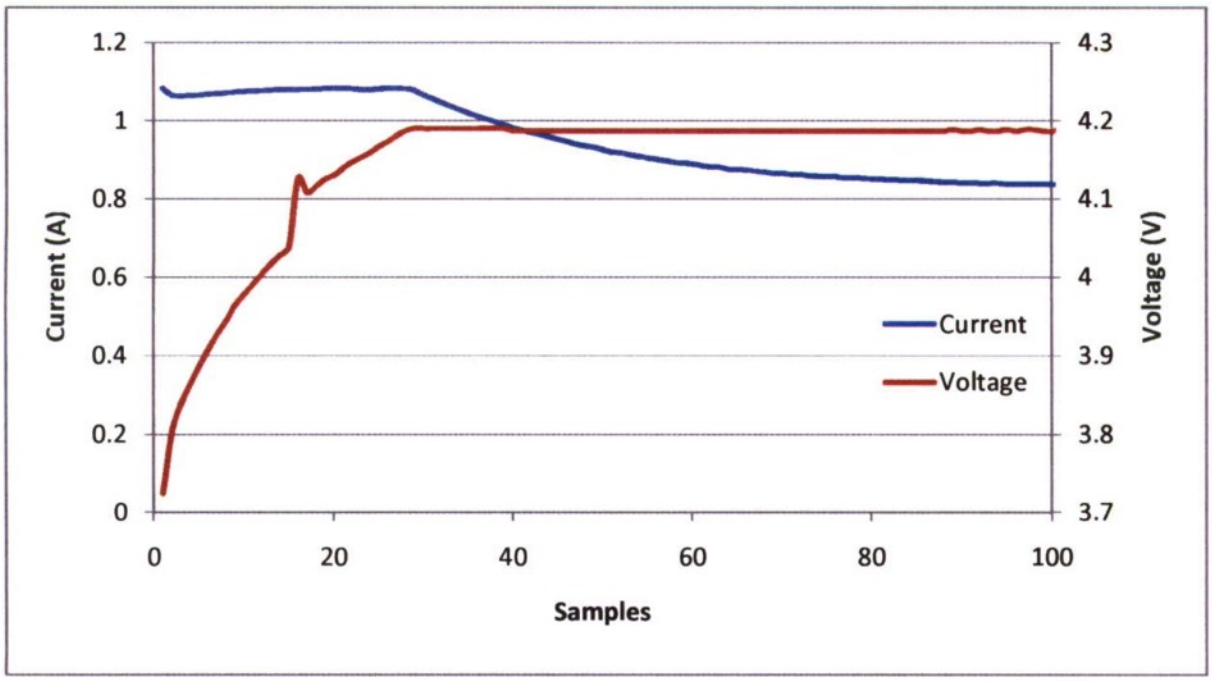


Figure 9 - Close-up of CCCV Portion of Charge

Figure 10 below shows a sample discharge cycle. The cell was fully charged at the beginning of the discharge, and was discharged through a 4.7 Ω resistor with a small fan cooling the resistor. The discharge was terminated when the cell voltage reached nominally 2.75 volts. In total, 724 mAh were taken out of the battery, and the actual termination voltage was 2.763 volts.

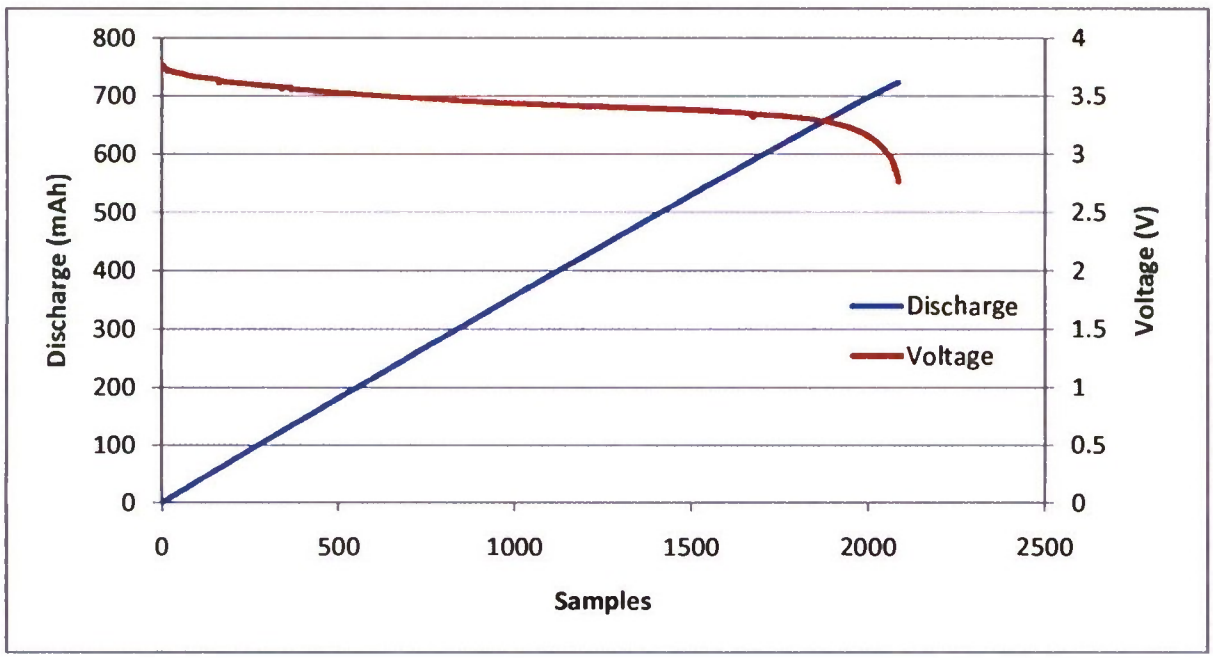


Figure 10 - Sample Discharge of Li-Ion Cell through 4.7 Ω Resistor

Conclusion

A battery test board was developed based on the STw4102 single-cell Li-ion charger and gas gauge. This test board provided a means to charge and discharge a battery while recording data. A test plan was carried out on two cells to verify the correct operation of the test board, as well as to confirm the cell abilities stated in their data sheet. From these tests charge and discharge curves were generated and appear to mostly agree with expected trends. Through the development of the board and embedded software, some problems were encountered. The first problem arose because of some weak shorts present on the circuit board, artifacts of the at-home etching process. These were eliminated with some re-work, and future boards (ordered from a professional PCB house) did not have this problem. There was some initial difficulty in setting up the communications (I^2C) between the microcontroller and the STw4102, but after some work and further research this problem was cleared up. One last problem presented itself during some charge cycles. At the start of a charge, the charge current was about half of what was expected. After some poking around in the location of the STw4102 IC, the charge current jumped up to the expected charge current, and proceeded normally for the remainder of the charge. After changing the small demo board holding the STw4102 IC, this problem went away. Further testing for a loose connection will be performed on the original demo. board.

This project was meant as a learning tool in the area of Li-ion charging and monitoring, and was also meant to provide the battery lab with a demonstration of present IC technologies for this purpose. In the future, this system could be expanded to work with multiple cells for a battery pack. While the STw4102 has some beneficial features such as integrated charger and gas gauge, it also has some drawbacks that a chip such as TI's bq27541 might address. Some of these drawbacks include limiting the cell capacity to 914 mAh and requiring an external 32.768 kHz clock signal. Future work might also include building more automation into the charge/discharge cycle so that the user does not have to babysit the process.

Bibliography

Broussely, M., & Pistoia, G. (2007). *Industrial Applications of Batteries: From Cars to Aerospace and Energy Storage.* Amsterdam: Elsevier.

Pop, V., Jan Bergveld, H., Danilov, D., Regtien, P. P., & Notten, P. H. (2008). *Battery Management Systems: Accurate State-of-Charge Indication for Battery-Powered Applications.* Eindhoven: Springer.

Appendix

Schematic

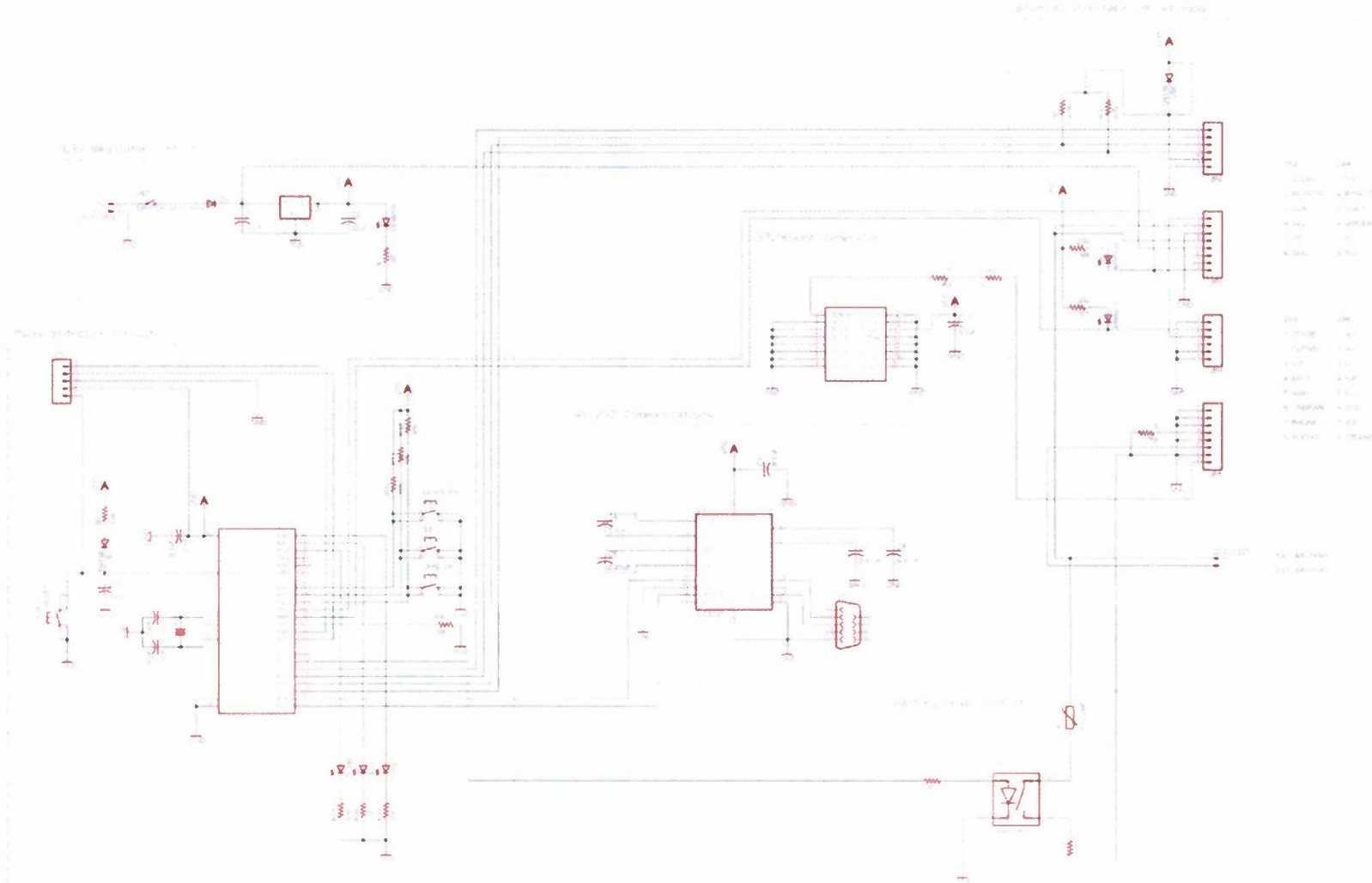


Figure 11 - Board Schematic

Board Layout

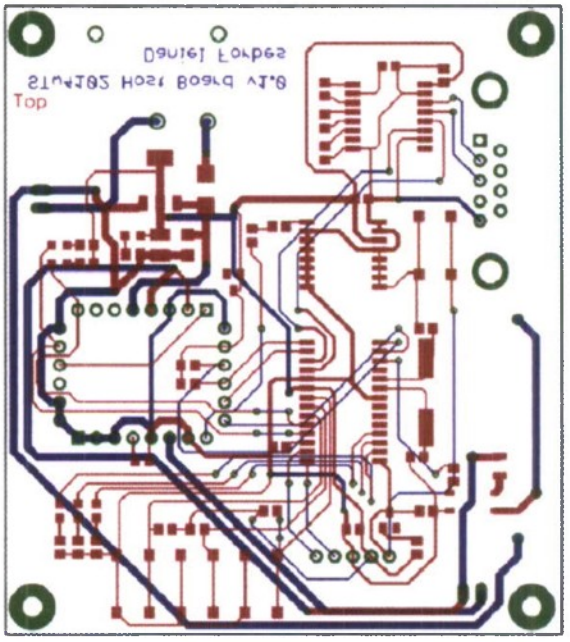


Figure 12 - Top and Bottom Copper Layers of Board

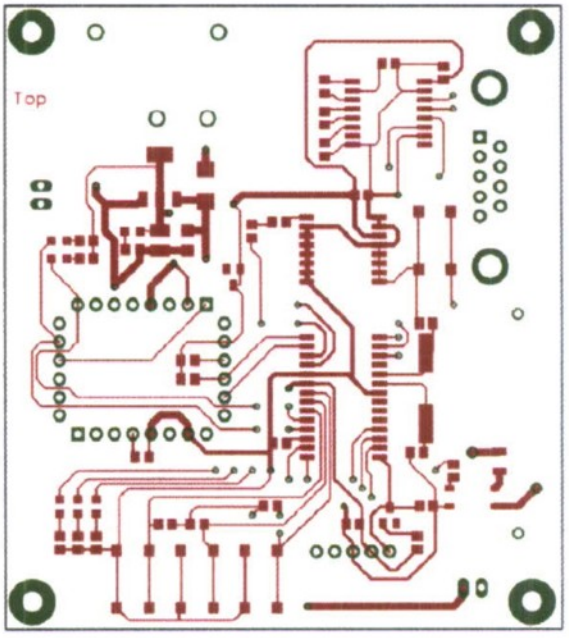


Figure 13 - Top Copper Layer of Board

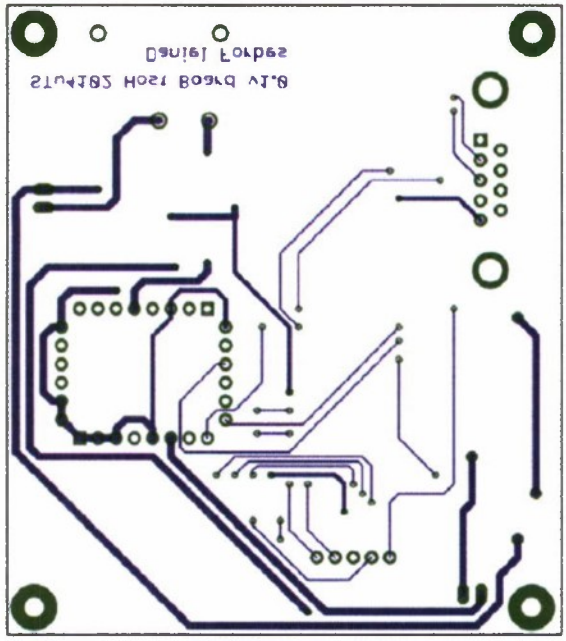


Figure 14 - Bottom Copper Layer of Board

BOM

Table 1 - Bill of Materials

Line Number	Part Number	Product Description	Manufacturer Name & Part Number	Price ($)	Qty Per Board	Extended Price ($)	Line note
1	45P5488	STW4102 based dual USB/wall adapter Li-Ion battery charger with gas gauge demo board	STMICROELECTRONICS - STEVAL-ISB006V1	52.63	1	52.63	
2	89H0763	8-Bit Microcontroller IC	MICROCHIP - PIC18LF2620-I/SO	5.12	1	5.12	
3	67H7013	Voltage Regulator IC	ON SEMICONDUCTOR - NCP1117ST33T3G	0.49	1	0.49	3.3 V LDO
4	61J6136	Transceiver IC	INTERSIL - ICL3232CBZ	1.37	1	1.37	USART - RS-232
5	02J2732	Resettable Fuse	BOURNS - MF-NSMF110-2	0.335	1	0.335	Batt. Dischrg. Fuse
6	58K1384	Fast Recovery Power Rectifier	FAIRCHILD SEMICONDUCTOR - ES2D	0.192	1	0.192	Vdd Diode
7	58K9418	Small Signal Diode	FAIRCHILD SEMICONDUCTOR - MMBD4148	0.041	1	0.041	ICSP Diode
8	18C1436	CRYSTAL,20MHZ,SMD	VISHAY DALE - XT9M20ANA20M	0.34	1	0.34	
9	19P9239	MOSFET Relay	OMRON ELECTRONIC COMPONENTS - G3VM-61GR1	4.77	1	4.77	SSR for Dischrg.
10	06P3531	Micro LED	DIALIGHT - 598-8110-107F	0.118	1	0.118	Red LED
11	22M3037	Micro LED	DIALIGHT - 5988170102F	0.236	4	0.944	Green LED
12	87M9432	Micro LED	DIALIGHT - 598-8130-107F	0.189	1	0.189	Orange LED
13	56H2243	Header, Breakaway Vertical	MOLEX - 22-28-4361	0.731	1	0.731	
14	89M6972	Tactile Switch	OMRON ELECTRONIC COMPONENTS - B3F-S1010	0.54	4	2.16	
15	96M1329	Ceramic Multilayer Capacitor	AVX - 08051A200JAT2A	0.109	2	0.218	Crystal Load Caps.
16	86H5264	Ceramic Multilayer Capacitor	TAIYO YUDEN - GMK212BJ474KG-T	0.061	3	0.183	0.47uF
17	27C8438	Ceramic Multilayer Capacitor	AVX - 08055C104KAT2A	0.064	5	0.32	0.1uF

18	84K8261	Tantalum Capacitor	AVX - TAJA106K016R	0.44	1	0.44	
19	35K1031	Tantalum Capacitor	AVX - TAJB226K016R	0.572	1	0.572	
20	01N6989	HIGH POWER RESISTOR	MULTICOMP - MCHP05W4F1002T5E	0.045	5	0.225	10k Res.
21	65K3543	Power Resistor	VISHAY DALE - CW0054R700JE12	0.571	1	0.571	Dischrg. Res.
22	25C1894	Thick Film Chip Resistor	VISHAY DALE - CRCW08054K70FKTA	0.038	2	0.076	4k7 pullup for I^2C
23	59M6873	Thick Film Chip Resistor	VISHAY DALE - CRCW0805200RJNEA	0.021	7	0.147	LED Res.
24	79M6013	RES CHIP, JUMPER, ZERO OHM 125mW 1%	MULTICOMP - MC0805S8F0000T5E	0.033	2	0.066	0 ohm Res.
25	68K3399	Temperature-Compensated Crystal Oscillator IC	MAXIM INTEGRATED PRODUCTS - DS32KHZS#	7.42	1	7.42	32.768 kHz Osc.
26	PRT-00429	Serial Port		1.5	1	1.5	
27	COM-08837	Power Switch		0.95	1	0.95	
28	PRT-08233	Molex Header	Molex - 22-23-2021	0.45	2	0.9	
29	PRT-00444	Nylon Standoff (pack of 4)		1.95	1	1.95	
30	PRT-00447	#4-40 Screw (pack of 4)		0.95	1	0.95	
					TOTAL	**85.92**	
	Newark						
	Sparkfun						

Code

main.c

```
/* main.c
 * This code runs on the STw4102 demonstration board build for Dr. John
 * Dunning's Battery Lab at Cal Poly. The code sets up the uC to talk to the
 * STw4102 via I2C. It also sets up USART communication with a computer and
 * waits for user input via buttons. The main purpose of the program is to
 * allow
 * a user to charge and discharge a battery while gathering data sent to the
 * connected computer. The serial port runs with the following settings:
 * 9600 bps, 8 data bits, 1 stop bit, no parity, and no flow control.
 *
 * Version Tracking
 * 03-13-10 Daniel Forbes     Set up ports and test board hardware
 * 03-19-10 DF          Setting up I2C communications
 * 04-02-10 DF          Still setting up I2C
 * 04-08-10 DF          Seems to be a hardware problem with SCL pin. . .
 * 04-11-10 DF          Fixed SCL, STw4102 not sending ACK. . .
 * 04-15-10 DF          Talking to STw4102! (Reset pin wansn't set high)
 *                              All systems working, UI time
 * 04-26-10 DF          A/D working in 12-bit mode
 *
*/
//-----------------------------------------------------------------------------
#include <p18f2620.h>          // Standard PIC18F2620 definitions
#include <portb.h>                  // I/O PORTB library
#include <usart.h>                  // Hardware USART library
#include <i2c.h>               // Hardware I2C library
#include <stdlib.h>                 // Standard library
#include <delays.h>                 // Delays library

#include "my_i2c.h"                 // I2C read/write library
#include "STw4102.h"           // STw4102 register descriptions
//#include "my_uart.h"         // UART initialization library


#pragma config OSC = HS        // 20MHz crystal
#pragma config WDT = OFF       // Turn off watch-dog timer
#pragma config LVP = OFF       // Disable low-voltage programming

//-----------------------------------------------------------------------------
void main(void)
{
//-----------------------------------------------------------------------------
      // Define constants and variables

      const unsigned char I2C_CONFIG = 0xE0;     // Configuration byte for
                                                      STw4102
      const unsigned char AD_SCALE = 142;        // 1.42 mV/count
      const unsigned char CHG_SCALE = 49;        // 49.5 nAh/count
      const unsigned long V_MIN = 1937;          // min. voltage limit
                                                      = 2.75 V
      const unsigned long V_MAX = 2993;          // max. voltage limit
```

```
                                                = 4.25 V

unsigned long a2d_value;                    // counts from A/D
unsigned long voltage_value;                // counts from A/D
unsigned char a2d[6];                       // array for ASCII output
unsigned int voltage;                       // converted battery voltage

unsigned long discharge_value;              // counts from discharge reg.
unsigned long charge_value;                 // counts from charge reg.
unsigned char chg[9];                       // array for ASCII output
unsigned int discharge;                     // converted discharge value
unsigned int charge;                        // converted charge value

unsigned long conversion_value;             // counts of last conversion
unsigned long conversion;                   // converted conversion value

unsigned char charge_enable = 0x00;         // set high to charge battery
unsigned char discharge_enable = 0x00;      // set high to discharge
                                               battery
unsigned char empty = 0x00;                 // set high when battery
                                               empty
unsigned char full = 0x00;                  // set high wehn battery full
//---------------------------------------------------------------------
// Configure ports

// Disable internal pull-ups and interrupts on PORTB
void closePORTB(void);

// Set DDR for PORTA
// RA5: NC | RA4: NC | RA3: NC | RA2: LED | RA1: LED | RA0: LED
TRISA = 0b00000000;

// Set DDR for PORTB
// RB7: ICSP  | RB6: ICSP | RB5: GND | RB4: PWRDT
// RB3: EOCHG | RB2: BTN  | RB1: BTN | RB0: BTN
TRISB = 0b00011111;

// Set DDR for PORTC
// RC7: RX  | RC6: TX    | RC5: RESETN | RC4: SDA
// RC3: SCL | RC2: STDBY | RC1: SSR    | RC0: NC
TRISC = 0b11011000;

// Clear PORTA and PORTB
PORTA = 0b00000000;
PORTB = 0b00000000;
//---------------------------------------------------------------------

//---------------------------------------------------------------------
// Setup and configure USART

// Setup TX register
TXSTAbits.TX9 = 0;
TXSTAbits.TXEN = 1;
TXSTAbits.SYNC = 0;
TXSTAbits.SENDB = 0;
TXSTAbits.BRGH = 0;
```

```
// Setup receive register
RCSTAbits.SPEN = 1;
RCSTAbits.RX9 = 0;
RCSTAbits.CREN = 1;

// Setup baud rate register
BAUDCONbits.RXDTP = 0;
BAUDCONbits.TXCKP = 0;
BAUDCONbits.BRG16 = 0;
BAUDCONbits.WUE = 0;
BAUDCONbits.ABDEN = 0;

// Set baud rate to 9600 = Fosc/64*(SPBRG+1)
SPBRG = 0x20;
//----------------------------------------------------------------------

//----------------------------------------------------------------------
// Setup and configure I2C
SSPSTAT = 0x80;    // MSSP Status Register
SSPCON1 = 0x28;    // MSSP Control Register 1
SSPCON2 = 0x00;    // MSSP Control Register 2
SSPADD = 0x31;     // MSSP Baud Rate Reload Register
                        // 0x31 for 100kHz
                        // 0x0B for 400kHz
                        // baud = Fosc/(4*(SSPADD+1))
//----------------------------------------------------------------------

PORTCbits.RC1 = 0;       // disable discharge
PORTCbits.RC2 = 1;       // pull STDBY high
PORTCbits.RC5 = 1;       // pull RESETN high

// Print welcome message to terminal
putrsUSART("\rWelcome to the STw4102 evaluation board tester:\n\r");
putrsUSART("voltage count, voltage (V), charge count, charge (mAh), "
                        "discharge count, discharge (mAh), last
                        conversion count"
                        ":\n\r");

// Clear STw4102
LDByteWriteI2C(I2C_CONFIG, REG_CHG0, 0x00);
LDByteWriteI2C(I2C_CONFIG, REG_CHG1, 0x00);
LDByteWriteI2C(I2C_CONFIG, REG_WDOG, 0x01); // enable watchdog, 1min.
LDByteWriteI2C(I2C_CONFIG, REG_CG, 0x00);
LDByteWriteI2C(I2C_CONFIG, REG_ADCTRL, 0x01);

// Enable gas gauge
LDByteWriteI2C(I2C_CONFIG, REG_CG, 0x0F); // clear all registers


// Loop here forever
while(1)
{
      LDByteWriteI2C(I2C_CONFIG, REG_WDOG, 0x09);     // reset watchdog
                                                          timer

      voltage_value = read_AD();     // read A2D for voltage limit
                                          checks
```

```
// check voltage is not too low
if(voltage_value <= V_MIN)
{
      PORTCbits.RC1 = 0;            // stop/disable discharge
      discharge_enable = 0x00;
      empty = 0x01;
      full = 0x00;
}

// check voltage is not too high
if(voltage_value >= V_MAX)
{
      LDByteWriteI2C(I2C_CONFIG, REG_CHG1, 0x00); // disable
                                                    charge
      charge_enable = 0x00;
      empty = 0x00;
      full = 0x01;
}

// set/clear charge flag if user selects "Charge"
if(PORTBbits.RB2 == 0)
{
      Delay10KTCYx(5);  // delay 10 ms
      if(PORTBbits.RB2 == 0)
      {
            while(PORTBbits.RB2 == 0);
            charge_enable = ~charge_enable;
            discharge_enable = 0x00;
      }
}

// set/clear discharge flag if user selects "Discharge"
if(PORTBbits.RB0 == 0)
{
      Delay10KTCYx(5);  // delay 10 ms
      if(PORTBbits.RB0 == 0)
      {
            while(PORTBbits.RB0 == 0);
            discharge_enable = ~discharge_enable;
            charge_enable = 0x00;
      }
}


// standby
if(discharge_enable == 0x00 && charge_enable == 0x00)
{
      LDByteWriteI2C(I2C_CONFIG, REG_CHG1, 0x00); //disable
                                                    charge
      PORTCbits.RC1 = 0;      // disable discharge
      PORTAbits.RA0 = 0;      // turn off charge LED
      PORTAbits.RA2 = 0;      // turn off discharge LED
}
```

```
// discharge
if(discharge_enable == 0xFF && charge_enable == 0x00)
{
      if(voltage_value <= V_MIN)
      {
            PORTCbits.RC1 = 0;
      }
      else
      {
            LDByteWriteI2C(I2C_CONFIG, REG_CHG1, 0x00);
                                                // disable charge
            PORTCbits.RC1 = 1;      // enable discharge
            PORTAbits.RA0 = 0;      // turn off charge LED
            PORTAbits.RA2 = 1;      // turn on discharge LED
      }
}

// charge
if(charge_enable == 0xFF && discharge_enable == 0x00)
{
      if(voltage_value >= V_MAX)
      {
            LDByteWriteI2C(I2C_CONFIG, REG_CHG0, 0x00);
      }
      else
      {
            PORTCbits.RC1 = 0;      // disable discharge
            PORTAbits.RA2 = 0;      // turn off discharge LED
            PORTAbits.RA0 = 1;      // turn on charge LED
            LDByteWriteI2C(I2C_CONFIG, REG_CHG0, 0x08); //4.2V
                                                charge volt.
            LDByteWriteI2C(I2C_CONFIG, REG_CHG1, 0x01); // enable
                                                      charge
      }
}
// The remaining code formats and outputs data read from the
// STw4102 registers.

// read battery voltage
a2d_value = read_AD();
voltage_value = a2d_value;
putsUSART(itoa(a2d_value, &a2d));
putrsUSART(",");
a2d_value *= AD_SCALE;        // 12-bit voltage 1.42mV/count

// interger volts
voltage = a2d_value / 100000;
putsUSART(itoa(voltage, &a2d));
putrsUSART(".");

// voltage tenths
voltage = (a2d_value % 100000) / 10000;
putsUSART(itoa(voltage, &a2d));

// voltage hundredths
voltage = ((a2d_value % 100000) % 10000) / 1000;
```

```
            putsUSART(itoa(voltage, &a2d));

            // voltage thousandths
            voltage = (((a2d_value % 100000) % 10000) % 1000) / 100;
            putsUSART(itoa(voltage, &a2d));
            putrsUSART(",");

            // read values
            LDByteWriteI2C(I2C_CONFIG, REG_CG, 0x11);
            Delay10KTCYx(126);      // 252ms delay
            charge_value = read_gas_charge();
            discharge_value = read_gas_discharge();
            conversion_value = read_gas_AD();


            // charge value
            putsUSART(ultoa(charge_value, &chg));
            putrsUSART(",");
            charge_value *= CHG_SCALE;
            charge = charge_value / 1000000;
            putsUSART(ultoa(charge, &chg));
            putrsUSART(".");
            charge = (charge_value % 1000000) / 100000;
            putsUSART(ultoa(charge, &chg));
            charge = ((charge_value % 1000000) % 100000) / 10000;
            putsUSART(ultoa(charge, &chg));
            putrsUSART(",");

            // discharge value
            putsUSART(ultoa(discharge_value, &chg));
            putrsUSART(",");
            discharge_value *= CHG_SCALE;
            discharge = discharge_value / 1000000;
            putsUSART(ultoa(discharge, &chg));
            putrsUSART(".");
            discharge = (discharge_value % 1000000) / 100000;
            putsUSART(ultoa(discharge, &chg));
            discharge = ((discharge_value % 1000000) % 100000) / 10000;
            putsUSART(ultoa(discharge, &chg));
            putrsUSART(",");

            // last conversion value
            if((conversion_value & 0x00001000) == 0x00001000)
            {
                  putrsUSART("-");
                  conversion_value = ((~conversion_value) & 0x00001FFF) + 1;
            }
            putsUSART(ultoa(conversion_value, &chg));
            putrsUSART("\n\r");

            // delay 1 sec.
            Delay10KTCYx(250);
            Delay10KTCYx(250);
      }
}
```

STw4102.h

```
/*    STw4102.h
 *  This is a header file for ST Microelectronics STw4102 single-cell
 *    li-ion battery charger/gas gauge IC. It holds I2C register addresses
 *    and descriptions.
 *
 *    Version Tracking
 *    03-30-10    Daniel Forbes     Set up file and register defines
 */

#define     REG_CHG0           0x00  // Charge control and status
#define     REG_CHG1           0x01  // Charge enable
#define     REG_WDOG           0x02  // Watchdog control
#define     REG_CG             0x03  // Gas gauge control
#define     REG_CHARGE_LOW     0x04  // Gas gauge charge data, bits 0-7
#define     REG_CHARGE_MID     0x05  // Gas gauge charge data, bits 8-15
#define     REG_CHARGE_HIGH    0x06  // Gas gauge charge data, bits 16-23
#define     REG_DISCHARGE_LOW  0x07  // Gas gauge discharge data, bits 0-7
#define     REG_DISCHARGE_MID  0x08  // Gas gauge discharge data, bits 8-15
#define     REG_DISCHARGE_HIGH 0x09 // Gas gauge discharge data, bits 16-23
#define     REG_CONVDATA_LOW    0x10 // Gas gauge AD converter data, bits 0-7
#define     REG_CONVDATA_HIGH   0x11 // Gas gauge AD converter data, bits 8-12
#define     REG_CONVNUMBER_LOW  0x12 // Number of conversions, bits 0-7
#define     REG_CONVNUMBER_HIGH 0x13 // Number of conversions, bits 8-11
#define     REG_ADCTRL          0x14 // Battery voltage monitor control
#define     REG_ADDATA_LOW      0x15 // Batt. V mon. AD conv. data, bits 0-7
#define     REG_ADDATA_HIGH     0x16 // Batt. V mon. AD conv. data, bits 8-10

// read latest A2D value
// RETURNS unsigned int: 12-bit voltage
unsigned int read_AD(void);

// read gas gauge charge register
// RETURNS unsigned short long: 24-bit charge accumulator value
unsigned short long read_gas_charge(void);

// read gas gauge discharge register
// RETURNS unsigned short long: 24-bit discharge accumulator value
unsigned short long read_gas_discharge(void);

// read gas gauge A2D (last conversion value)
// RETURNS unsigned int: 13-bit current value
unsigned int read_gas_AD(void);

// read the number of conversions elapsed
// RETURNS unsigned int: 12-bit value
unsigned int read_num_conv(void);

// get watchdog status
// RETURNS unsigned char: 0x01 if watchdog timer has elapsed, 0x00 otherwise
unsigned char watchdog(void);
```

STw4102.c

```
/*    STw4102.c
 *  This file controls the ST Microelectronics STw4102 single-cell
 *    li-ion battery charger/gas gauge IC. It performs register reads on the
 *  STw4102 as well as some data converions.
 *
 *    Version Tracking
 *    04-26-10    Daniel Forbes     Set up file and register defines
 *    05-23-10    DF                      Final commenting
 */

#include "STw4102.h"                // header file
#include "my_i2c.h"                 // I2C read/write library

const unsigned char I2C_CONFIG = 0xE0;    // configuration bit for STw4102
unsigned char st_data[3];                 // array for read data

// read latest A2D value
// RETURNS unsigned int: 12-bit voltage
unsigned int read_AD(void)
{
      unsigned int ad_value = 0x0000;

      LDByteWriteI2C(I2C_CONFIG, REG_ADCTRL, 0x15);  // start AD conversion

      // wait till conversion is complete
      LDByteReadI2C(I2C_CONFIG, REG_ADCTRL, st_data, 1);
      while(st_data[0] & 0x08)
      {
      LDByteReadI2C(I2C_CONFIG, REG_ADCTRL, st_data, 1);
      }

      // read AD data register
      LDByteReadI2C(I2C_CONFIG, REG_ADDATA_LOW, st_data, 2);

      ad_value = st_data[1];
      ad_value <<= 8;
      ad_value |= st_data[0];

      return(ad_value);
}

// read gas gauge charge register
// RETURNS unsigned short long: 24-bit charge accumulator value
unsigned short long read_gas_charge(void)
{
      unsigned short long gas_charge = 0x000000;
      unsigned short long gas_charge_temp = 0x000000;

      LDByteReadI2C(I2C_CONFIG, REG_CHARGE_LOW, st_data, 3);

      gas_charge = st_data[2];
      gas_charge <<= 16;
      gas_charge_temp = st_data[1];
      gas_charge_temp <<= 8;
      gas_charge_temp |= st_data[0];
```

```
        gas_charge |= gas_charge_temp;

        return(gas_charge);
}

// read gas gauge discharge register
// RETURNS unsigned short long: 24-bit discharge accumulator value
unsigned short long read_gas_discharge(void)
{
        unsigned short long gas_discharge = 0x000000;
        unsigned short long gas_discharge_temp = 0x000000;

        LDByteReadI2C(I2C_CONFIG, REG_DISCHARGE_LOW, st_data, 3);

        gas_discharge = st_data[2];
        gas_discharge <<= 16;
        gas_discharge_temp = st_data[1];
        gas_discharge_temp <<= 8;
        gas_discharge_temp |= st_data[0];
        gas_discharge |= gas_discharge_temp;

        return(gas_discharge);
}

// read gas gauge A2D (last conversion value)
// RETURNS unsigned int: 13-bit current value
unsigned int read_gas_AD(void)
{
        unsigned int gas_AD = 0x0000;

        LDByteReadI2C(I2C_CONFIG, REG_CONVDATA_LOW, st_data, 2);

        gas_AD = st_data[1];
        gas_AD <<= 8;
        gas_AD |= st_data[0];

        return(gas_AD);
}

// read the number of conversions elapsed
// RETURNS unsigned int: 12-bit value
unsigned int read_num_conv(void)
{
        unsigned int num_conv = 0x0000;
        return(num_conv);
}

// get watchdog status
// RETURNS unsigned char: 0x01 if watchdog timer has elapsed, 0x00 otherwise
unsigned char watchdog(void)
{
        LDByteReadI2C(I2C_CONFIG, REG_WDOG, st_data, 1);
        if((st_data[0] & 0x40))
                return(0x01);
        else
                return(0x00);
}
```

Force protection in combat and non-combat situations
Project Reports

Efficient Assimilation of AUV Data in a High-Resolution Coastal Ocean Model

Principal Investigator:

Paul Choboter, Ph. D.; Christopher Clark, Ph. D.; Mark Moline, Ph. D.
Mathematics, Computer Science, Biological Sciences
California Polytechnic State University
San Luis Obispo, CA

1. Introduction

The circulation of the coastal ocean is extremely complex, involving motion on a wide range of scales, both in time and space. It is important to understand and to be able to predict that motion to monitor our effect on the ocean environment, and for the sustainability of the resources the ocean provides.

Numerical ocean models may be used to predict the state of the ocean. However, truly accurate predictions require the assimilation of observational data into the models. Data assimilation is known to be computationally demanding, and the development of improved techniques is presently a topic of intense research effort by many scientists.

While the assimilation of data measured at pre-determined locations (e.g., as determined by satellite flight path, buoy location, position of an ocean drifter, or coastal radar sites) has been the focus of many studies, much less research attention has been paid toward developing techniques for collecting data in those specific locations that will contribute most efficiently to improving the model accuracy.

The goal of this research project is to develop techniques for performing high-resolution numerical simulations that efficiently assimilate data collected from a system of Autonomous Underwater Vehicles (AUVs). This project builds directly on recent research studying AUV trajectories and data collection, and adds focus on the efficient incorporation of that data into a numerical model describing the local ocean circulation.

2. Background

The qualitative characteristics of the California Current system include a wide (500-1000 km) wind-driven southward-flowing surface current and a thin northward-flowing California Counter Current adjacent to the coast (Hickey, 1998). The Counter Current has a strong seasonal dependence and is typically strongest in winter. On large spatial scales, the ocean surface is also driven offshore by the southward summer winds, which causes deep water to be upwelled to the surface near the coast. Superimposed on this broad motion is a rich eddy field on smaller scales that is strongly influenced by coastline geometry, such as the locations of headlands and bays.

High quality simulations of ocean circulation within the California Current system have been performed by various authors (e.g., Marchesiello et al. 2003). The present work complements such large-scale studies by focusing on the motion on a relatively small scale, which requires higher resolution than is possible with larger-scale simulations. Shulman et al. (2002) similarly performed high-resolution simulations of circulation in Monterey Bay, using nested grids and the assimilation of CODAR surface current measurements.

The ocean modeling is performed using the Regional Ocean Modeling System (ROMS, http://www.myroms.org). ROMS is a state-of-the-art numerical model being actively developed and maintained by the scientific community, and has data assimilating

capability. The primitive-equation numerical model includes the motion of the free-surface, and employs terrain-following coordinates (Shchepetkin and McWilliams 2005). The primitive equations express the conservation of fluid momentum, temperature, and salinity, under the Boussinesq and hydrostatic approximations. ROMS evolves the components of velocity, temperature, salinity and sea surface height forward in time using a finite-difference discretization of the equations of motion.

Although the theory of data assimilation is well-developed (Bennett 2002), advanced assimilation techniques are computationally demanding, and oceanic data assimilation is an active field of research. Despite the computational demands of data assimilation, recent research has documented encouraging progress toward real-time adaptive sampling. Lermusiaux (2007) discusses the data assimilation techniques employed during the Autonomous Ocean Sampling Network-II (AOSN II) experiment in Monterey Bay in 2003, where multiple data sets from a wide variety of sources were assimilated in near real-time. Leslie et al. (2008) discuss the Maritime Rapid Environmental Assessment (MREA) experiments, in which limited data was assimilated in an efficient technique to generate real-time oceanic field estimates.

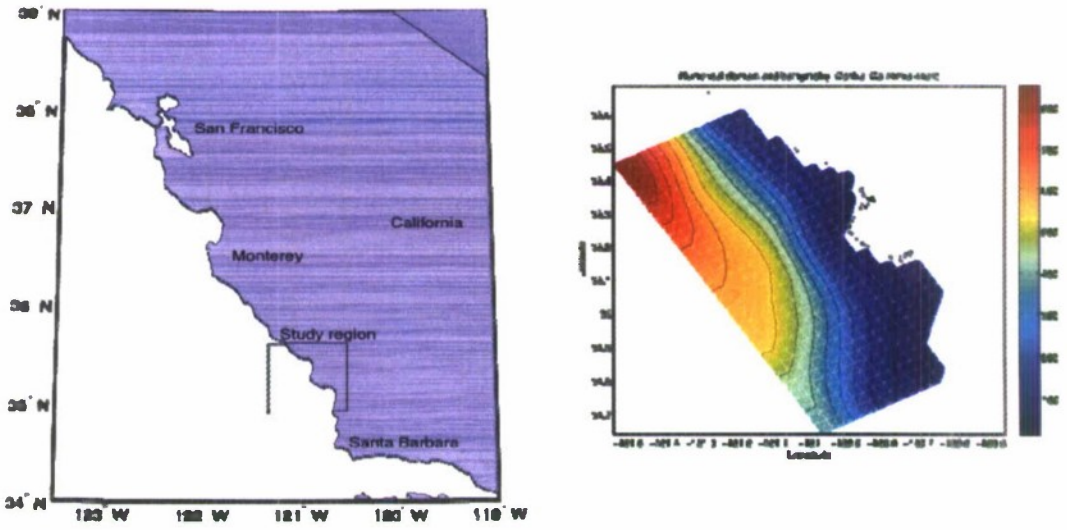


Figure 1. Domain and bathymetry of the numerical runs

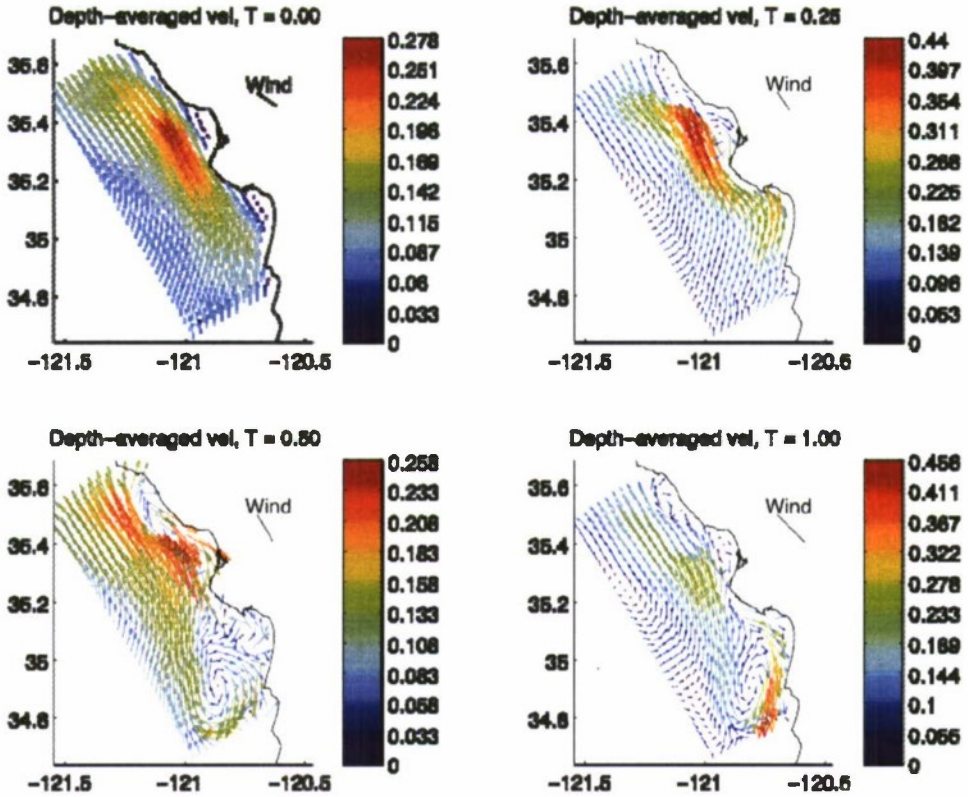


Figure 2. Baseline true solution. Depth-averaged velocity at T=0, 0.25, 0.5 and 1.0 days.

3. Numerical Experiments

Autonomous Underwater Vehicles (AUVs) may be used to make sub-surface observations at targeted locations, but are limited in the amount of data that can be collected. The present work seeks to answer the question: Given a limited number of observations that can be collected, what is the optimal strategy for choosing precisely where to collect the data? The suite of data assimilation tools accompanying the Regional Ocean Modeling System (ROMS) is used to perform numerical experiments to help address this question. We describe here the details of initial experiments designed to investigate whether an AUV path chosen based on error estimates can be more efficient at reducing error than a traditionally prescribed AUV path.

ROMS was configured to simulate the ocean off the coast of central California (Fig. 1), with a horizontal grid of 17 by 33 points, and 20 levels in the vertical direction. The size of the domain is approximately 65 km by 130 km, with a horizontal resolution of 4 km. The model wind stress driving the ocean is interpolated from 6-hr, 0.25 degree NOAA/NCDC blended wind product (http://www.ncdc.noaa.gov/oa/rsad/blendedseawinds.html). For initial and boundary conditions, data is interpolated from a data-assimilating ocean model at 1/12 degree resolution (global HYCOM, http://www.hycom.org/). Open boundary conditions are employed in the model, with a

standard configuration (free-surface Chapman, 2D momentum Flather, clamped 3D momentum, temperature and salinity).

Data assimilation is performed using the strong-constraint variational assimilation formulation (4DVAR), where the adjoint of the tangent linear model is used to make adjustments to the initial condition only.

To investigate the efficiency of assimilating data collected along a specified AUV path, a simple numerical experiment was performed. A baseline "true" solution was first simulated (see Fig. 2). Next, ROMS was run with incorrect initial conditions: velocities set to zero initially. Finally, synthetic observations are sampled from the true solution and assimilated into the incorrect runs to improve the initial conditions. The sampling method was either the traditional ("lawnmower") grid pattern or an AUV path chosen to seek out the greatest error.

3a. Choosing the Optimal Path I: Modeling error reduction

In simulations where the AUV path is chosen to seek out the locations of greatest error, the reduction of error as a result of assimilating each measured data point is modeled. Given an error field to minimize, *E*, and an AUV location $\vec{x}_0 = (x_0, y_0)$, the effect of observations on the assimilated runs is modeled by a fractional reduction of the error at the given location and at neighboring locations,

$$E_{[new]}(x,y) \quad E_{[old]}(x,y)\left\{1 \quad \right\}$$

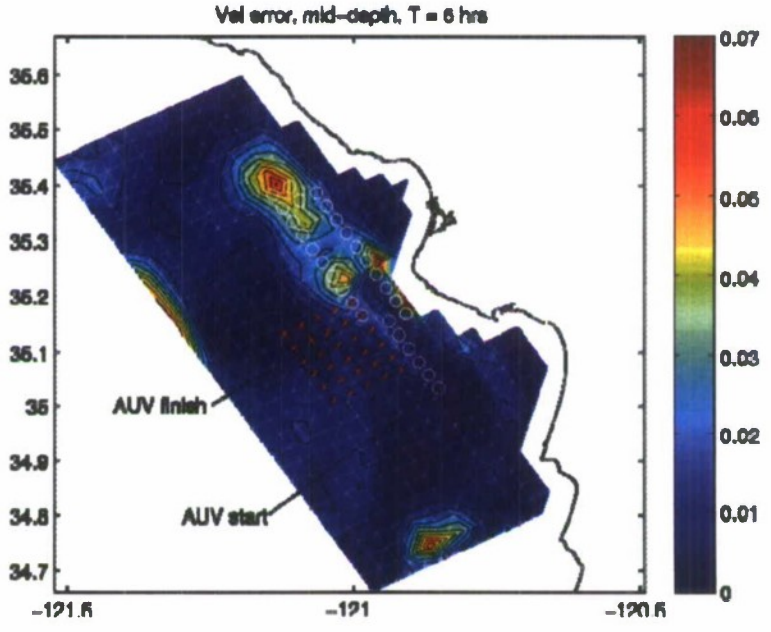


Figure 3. Error field used to guide virtual AUV. Also shown is one best-first path (white circles) and a lawnmower pattern path (red crosses) with the same number of points.

where $p = 0.15$ and $r = 1$ km were used for the experiments described here. After reducing the error in this way, the next location of the AUV was chosen to be an adjacent grid point using a "best-first" search path.

3b. Choosing the Optimal Path II: Best-first search

In the best-first search, the next location is chosen to be the adjacent grid point with the largest error. This method of path determination is chosen for simplicity and efficiency at this preliminary stage in the research project. Other path-searching algorithms currently being explored include "depth-first" and "breadth-first".

3c. Results of numerical experiments

As a measure of the error, define

$$[illegible]$$

where u_{True} and v_{True} are sampled from the true solution, and velocities are measured at mid-depth at T=6 hrs. The resulting error field is shown in Fig. 3. In the experiments reported here, the best-first search path and the lawnmower pattern path were each

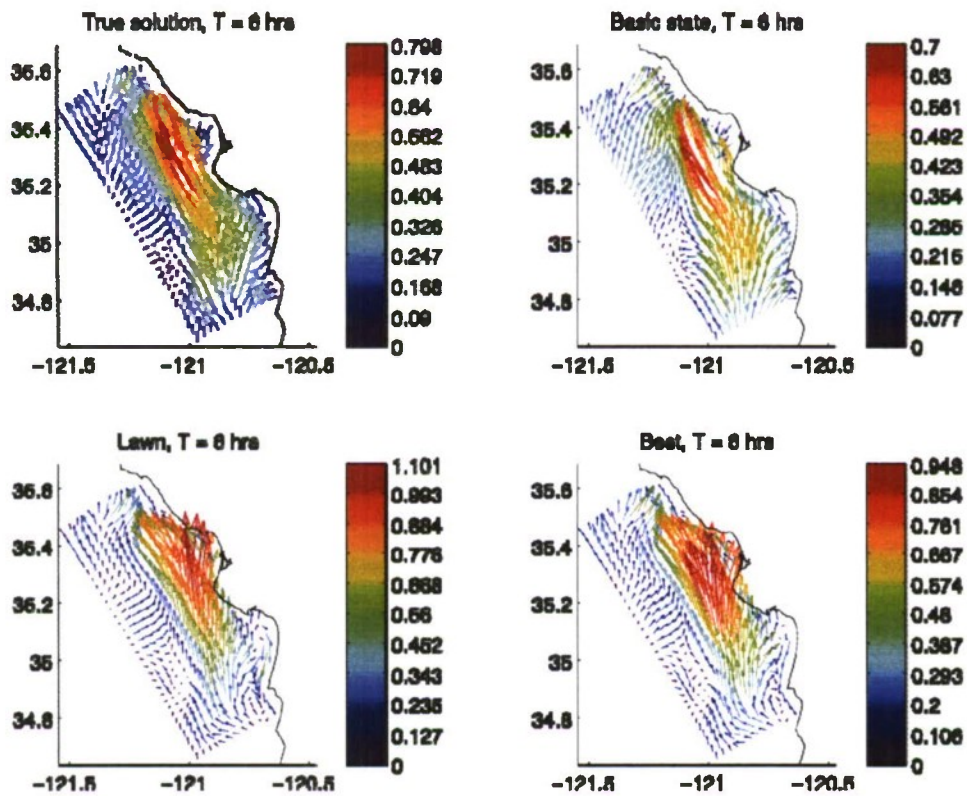


Figure 4. Mid-depth velocity at 6 hrs in the four runs.

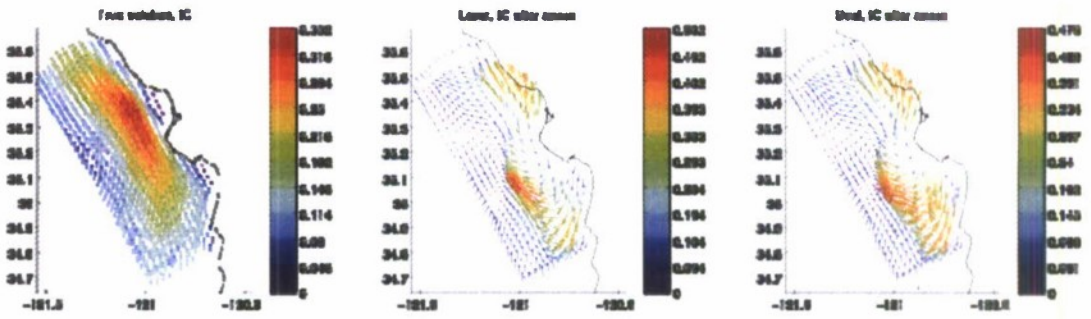


Figure 5. Comparison of true initial condition to the initial conditions determined via data assimilation.

used to sample only 30 observational points. Despite the small number of sampled points, the assimilation system did yield solutions resembling the true solution (Figs 4-6). The best-first directed AUV path performed slightly better than the traditional lawnmower path of data collection.

4. Conclusions and Future Work

The basic numerical experiments show that assimilation of observation collected along a targeted AUV path can improve the simulation better than observations collected along a pre-determined lawnmower grid pattern.

It remains to utilize the suite of data assimilation tools within ROMS, which can measure sensitivity of improvement to observations, to direct the AUV when the true error is not known. The goal is to couple a real AUV to ROMS in two-way communication, with the AUV supplying data to be assimilated, and with ROMS directing the AUV via predicted error fields.

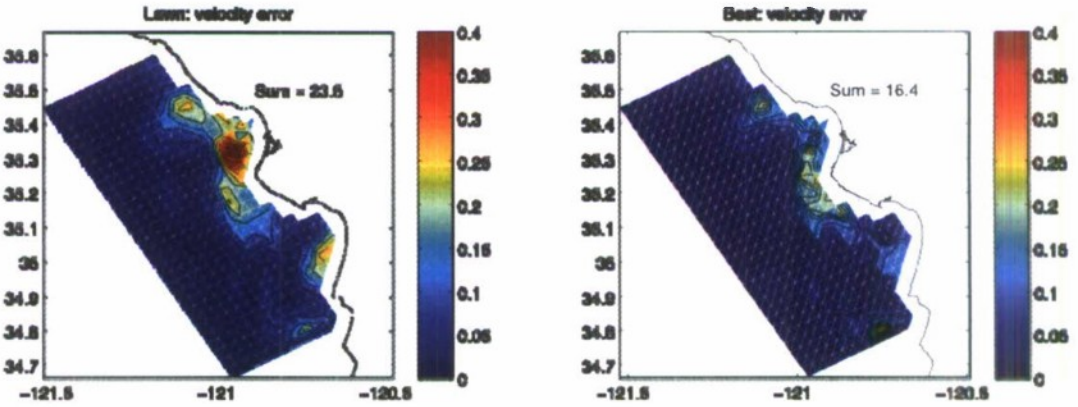


Figure 6. Residual velocity error at T=6 hrs in the two experimental assimilation runs.

It should be noted that in developing this work, several advancements in AUV guidance, navigation and control have been made. MS student Beverly Chow initially used the Iver2 purchased with this grant to develop and validate task planning when tasks are closely spaced. This has recently resulted in a journal publication (Chow et al., accepted 2010). Next, a team of computer engineering capstone students experimented with a new state estimation strategy using an Inertial Measurement Unit. Third, two undergraduate students developed ice detection algorithms and conducted experiments both at Avila and during an expedition in the Arctic. A paper has been written (Layton et al. 2011) for submission to either a journal or conference. In the last few months, the Iver2 was used as a core technology for sampling dissolved Oxygen in Hopavagen Bay, Norway. A new Kalman Filter based method for fusing multiple AUV measurements to create a lattice structure map of Air Saturation Percentage. An article is in process of being written for the Journal of Field Robotics (Clark et al. 2010). Throughout this work, MS engineering student Ben Davini has been developing trajectory construction algorithms for Ocean Modeling system discussed above.

It should be also noted that the combined efforts with the Iver2 have led to several successful grants being awarded. These include an NSF Planning Visit to the Arctic (20K), COAST grant (one course release), and an NSF IRES grant (150K). Currently, an NSF Robust Intelligence grant (490K) has been recommended for funding and is at the final stages of approval.

References

Bennett, Andrew F., *Inverse Modeling of the Ocean and Atmosphere*, Cambridge University Press, 234 pp., 2002.

Chow, B., J. P. Huissoon, C. M. Clark, Assigning Closely Spaced Targets to Multiple Autonomous Underwater Vehicles, Recently accepted in the *Journal of Ocean Science*

Clark, C. M., A. Xydes, K. Hall, F. Schreiber, J. Klemme, and K. Hancke, Air Saturation Mapping with an Autonomous Underwater Vehicle, In preparation for submission to be submitted to the *International Journal of Field Robotics*, 2010.

Hickey, B. M., Coastal oceanography of western North America from the tip of Baja, California to Vancouver Island. *The Sea*, **11**, 345-393, (1998).

Layton, M. S., R. Plankenhorn, M. Moline, G. Johnsen, C. M. Clark, AUV Ice Edge Detection, Ready for submission to the 2011 IEEE OCEANS conference.

Lermusiaux, Pierre F. J., Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D,* **230**, 172–196, 2007.

Leslie, W. G., et al., Verification and training of real-time forecasting of multi-scale ocean dynamics for maritime rapid environmental assessment. *J. Marine Systems,* **69**, 3–16, 2008.

Marchesiello, P., J. C. McWilliams, and A. Shchepetkin, Equilibrium structure and dynamics of the California Current System. *J. Phys. Oceanogr.*, **33**, 753–783, 2003.

Shchepetkin, A. F. and J. C. McWilliams, The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, **9**, 347–404, 2005.

Shulman, I., C. -R. Wu, J. K. Lewis, J. D. Paduan, L. K. Rosenfeld, J. C. Kindle, S. R. Ramp and C. A. Collins, High Resolution Modeling and Data Assimilation in the Monterey Bay area. *Continental Shelf Res.*, **22**, 1129-1151, 2002.

Conference Poster Presentation

This work resulted in the following poster, presented at the 2010 Ocean Sciences Meeting, 22-26 February 2010, Portland, Oregon:

Efficient Assimilation of AUV Data in a High-Resolution Coastal Model. P. Choboter; C.M. Clark; M. Moline

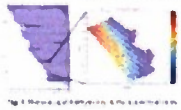
Abstract:
Recent research has identified how a data assimilation system may be used not only to improve the accuracy of a numerical model's forecast, but also to measure the sensitivity of the improvement with respect to the observations assimilated. The observation sensitivity is used in this work to predict optimal AUV paths for efficiently collecting data to be assimilated into a coastal ocean model. Numerical experiments are performed using the Regional Ocean Modeling System (ROMS), along with its tangent-linear and adjoint tools, to determine strategies for AUV path optimization. Simulation parameters, including forcing and boundary conditions, are chosen in preparation for runs assimilating AUV and CODAR data off the coast of central California.

Efficient Assimilation of AUV Data in a Coastal Ocean Model

Paul Chatwater, Chris Clark, Ben Dewitt, and Mark Moline
Departments of Mathematics, Computer Science, and Biological Sciences, CalPoly State University

1. Introduction

Techniques for the assimilation of observational data into numerical ocean models to improve forecast are well-developed. Autonomous Underwater Vehicles (AUVs) may be used to make sub-surface observations in targeted locations, but are limited in the amount of data that can be collected. The present work seeks to answer the question: Given a limited number of observations that can be collected, what is the optimal strategy for choosing precisely where to collect the data? The suite of data assimilation tools accompanying the Regional Ocean Modeling System (ROMS) is used to perform numerical experiments to help address this question.

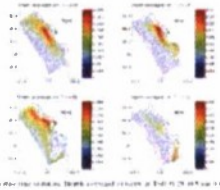


2. Numerical model details

ROMS was configured to simulate the ocean off the coast of central California, with parameters:

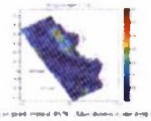
- 42 by 38 horizontal grid points and 20 vertical levels
- 65 km by 130 km domain, $\Delta x = \Delta y = 4$ km
- Wind stress interpolated from 6-hr, 0.25 degree NOAA NCDC blended wind product
- Initial, boundary conditions: global, daily-output, 1/12 degree HYCOM Analysis
- Open BCs: free-surface Chapman, 2D momentum Flather, clamped 3D momentum and tracers

Data assimilation is performed using the strong-constraint variational assimilation formulation (4DVAR), where the adjoint of the tangent linear model is used to make adjustments to the initial conditions only.



3. Numerical experiments

To investigate the efficiency of assimilating data collected along a specified AUV path, a simple numerical experiment was performed. A baseline "true" solution was first simulated (see **Fig. 2**). Next, ROMS was run with incorrect initial conditions: velocities set to zero initially. Finally, synthetic observations are sampled from the true solution and assimilated into the incorrect run to improve the initial conditions. The sampling method was either the traditional lawnmower grid pattern or an AUV path chosen to track out the greatest error.

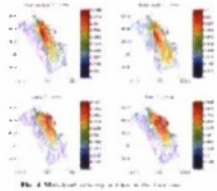


4. Choosing the Optimal Path I: Modeling error reduction

The AUV path is chosen in such a way as to maximize the greatest error. Given an error field to minimize, E, and an AUV location (x_0, y_0), the effect of observations on the assimilated run is modeled by a fractional reduction of the error at the given location and at neighboring locations,

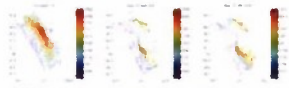
$$E_{new}(x, y) = E_{prior}(x, y) \left[1 - \rho \, e^{-\left(\frac{(x - x_0)^2}{2\sigma_x^2} + \frac{(y - y_0)^2}{2\sigma_y^2} \right)} \right]$$

where $\rho = 0.15$ and $\sigma_x = \sigma_y = 4$ km were used for the experiments described here. After reducing the error in this way, the next location of the AUV was chosen to be an adjacent grid point using a "best-first" search path.



5. Choosing the Optimal Path II: Best-first search

In the best-first search, the next location is chosen to be the adjacent grid point with the largest error. This method of path determination is chosen for simplicity and efficiency at this preliminary stage in the research project. Other path-searching algorithms currently being explored include "depth-first" and "breadth-first".

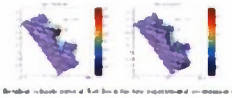


6. Results of numerical experiments

As a measure of the error, define

$$E_{tot}(t) = \sum_{i,j,k} \left[\left(u_{i,j,k}^{true}(t) - u_{i,j,k}^{model}(t) \right)^2 + \left(v_{i,j,k}^{true}(t) - v_{i,j,k}^{model}(t) \right)^2 \right]$$

where $u_{i,j,k}$ and $v_{i,j,k}$ are sampled from the true solution and velocities are measured at mid-depth at 1-hr intervals. The resulting error field is shown in **Fig. 3**. Despite the small number of sampled points, the assimilation system did yield solutions resembling the true solution (**Figs 4-6**). The best-first directed AUV path performed slightly better than the traditional lawnmower path of data collection.



7. Conclusions and Future Work

The basic numerical experiments show that assimilation of observation collected along a targeted AUV path can improve the simulation better than observations collected along a pre-determined lawnmower grid pattern.

It remains to analyze the suite of data assimilation tools within ROMS, which can assess sensitivity of improvement to observations by choosing the AUV when the true error is not known. The goal is to couple a real AUV to ROMS in two-way communication, with the AUV supplying data to be assimilated and with ROMS directing the AUV via predicted error fields.

This work is sponsored by the Chairman of the Navy Office of Naval Research under Award # N00014-08-1-1209.

Tracking Shallow Water Squid via an Underwater Robot System

Principal Investigators:

Pat M. Fidopiastis, Ph.D. and Chris Clark, Ph. D.
Biological Sciences and Computer Science
California Polytechnic State University
San Luis Obispo, CA

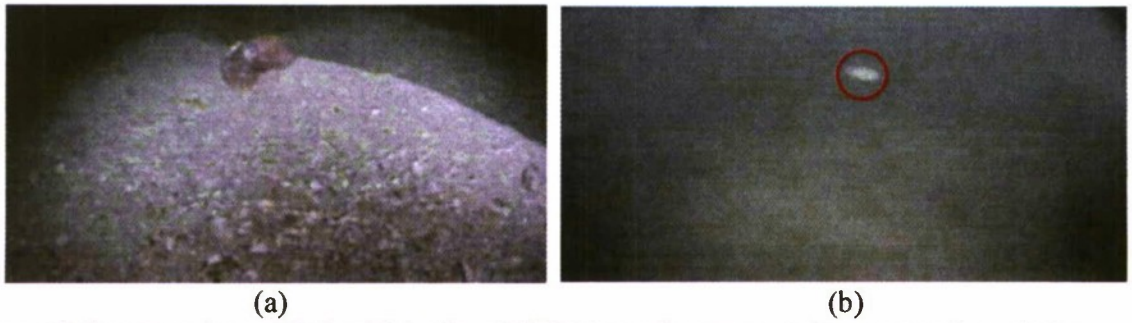
1. Overview

During the summer of 2009, Drs. Fidopiastis and Clark traveled to Hawaii to use Remotely Operated Vehicles (ROVs) to capture 6 gigabytes of video footage of an elusive, nocturnal, shallow water squid.

This work employed the use of an underwater robot system to find and obtain video footage of the squid *Euprymna scolopes*, so that motion behaviors of the squid can later be characterized through the use of off-line image processing. This particular species of squid are nocturnal and typically found in the shallow beaches of Oahu, Hawaii, USA (see Fig. 1). A Remotely Operated Vehicle (ROV) was deployed for several nights at several locations in Oahu, resulting in 12 hours of squid footage. Using blob tracking image processing techniques, the squid can be detected and tracked in several of the videos, allowing the ability to calculate the squid state relative to the ROV. Combining this state information with absolute ROV state estimates obtained using optical flow techniques, the relative motion of the squid can be determined. Results show typical trajectories of the squid during three different behavioral modes - namely evasion, escape, and foraging. To the best of the authors' knowledge, this is the first observation and tracking of squid in its natural environment.

(a) (b) (c)

Figure 1. In (a) is depicted the VideoRay Pro III Remotely Operated Vehicle (ROV). Shown in (b) is the squid Euprymna scolopes. The location of fieldwork is shown in (c), several miles from Waikiki.



(a) (b)

Figure 2: Images taken with the VideoRay ROV (a) can be processed to extract the relative position of the squid. Shown in (b) is the location of the squid (red circle) as estimated by the image processing algorithm.

2. Research Accomplishments

2.1 Data Acquisition

The research team arrived in Oahu on June 10th, 2009 and left on June 24th, 2009. Attempts were made each night to find squid. Weather including wind and rain made it impossible to find squid on several nights, since such conditions stirred up shallow water and severely limited visibility below the surface. Table 1 provides an overview of data collected on a nightly basis.

Date, location, time

Date	Location	Time	Squid Detected
June 11th	Paiko Beach	8:00pm – 2:00am	Yes
June 12th	Paiko Beach	Poor Weather	
June 13th	Paiko Beach	Poor Weather	
June 14th		Poor Weather	
June 15th		Poor Weather	
June 16th	Paiko Beach	8:00pm-12:30am	Yes
June 17th	Paiko Beach	11:30pm-1:30am	Yes
June 18th		Poor Weather	
June 19th		Poor Weather	
June 20th	Paiko Beach	11:00pm-3:30am	Yes
June 21st		Poor Weather	
June 22nd	Paiko Beach	11:00pm-3:30am	Yes
June 23rd	Paiko Beach	12:30am-1:30am	No

Table 1: Data Collection Summary

2.2 Software Development

The goal of the software developed in this work is to determine the position and velocity states of squid using recorded video of the squid while in its natural habitat. To accomplish this, each video frame of a video segment is post-processed to automatically locate the squid with the frame, and use geometry to calculate the squid position.

To extract the state of the squids from each individual video frame, a series of image processing techniques were used. These techniques, described in the pseudo code of Fig. 3, are basic with respect to current image processing techniques but are none-the-less appropriate and useful for this research.

```
function [x, y, z] = findSquidPosition(I)

    G = toGray(I);
    B = extractBlobs(G);

    for i = 0 : B.size {
        if (MinArea < B[i].getArea() < MaxArea) {
            [x, y] = getPositionInMetricUnits(x_pixel, y_pixel);
            z_x = getRangeFromX(x);
            z_y = getRangeFromY(y);
            z = (z_x + z_y)/2;
        }
    }
```

Figure 3: Pseudo code for calculating position of a squid from a video frame

Initially the video image is converted to a gray-scale, after which blobs are extracted (using a tool from the open source toolbox OpenCV). If the area of any blobs found are within the limits of the expected area of a squid in the image. The position of the squid within the image can be converted from pixels to meters using a third order polynomial function created using a least squares fit to calibration data.

$$u = x_{pixel} \quad (1)$$
$$v = -y_{pixel} \quad (2)$$

$$x = c_{u,x}u + c_{v,x}v + c_{uu,x}u^2 + c_{vv,x}v^2 + c_{uv,x}uv + c_{uuu,x}u^3 + c_{vvv,x}v^3 + + c_{uuv,x}u^2v + c_{uvv,x}uv^2 + c_{u,x} \quad (3)$$
$$y = c_{u,y}u + c_{v,y}v + c_{uu,y}u^2 + c_{vv,y}v^2 + c_{uv,y}uv + c_{uuu,y}u^3 + c_{vvv,y}v^3 + + c_{uuv,y}u^2v + c_{uvv,y}uv^2 + c_{u,y} \quad (4)$$

With each of the x and y coordinates known, the range z can also be calculated. The final range to squid z can be a weighted average of each.

$$z_x = b_x\, x / \Delta_x \quad (5)$$
$$z_y = b_y\, y / \Delta_y \quad (6)$$

2.3 Results

Current results indicate that relative state estimation of the squid is possible. Figure 3 provides an example of one video sequence processed to obtain the 3D relative position of the squid. Note the range of the squid from the ROV is plotted in the Z axis, matching expected measurements (based only on visual observations at this point) between 0.2 and 1.5 meters. Note the y estimates are always greater than 0 since the squid remained in the top half of the video images.

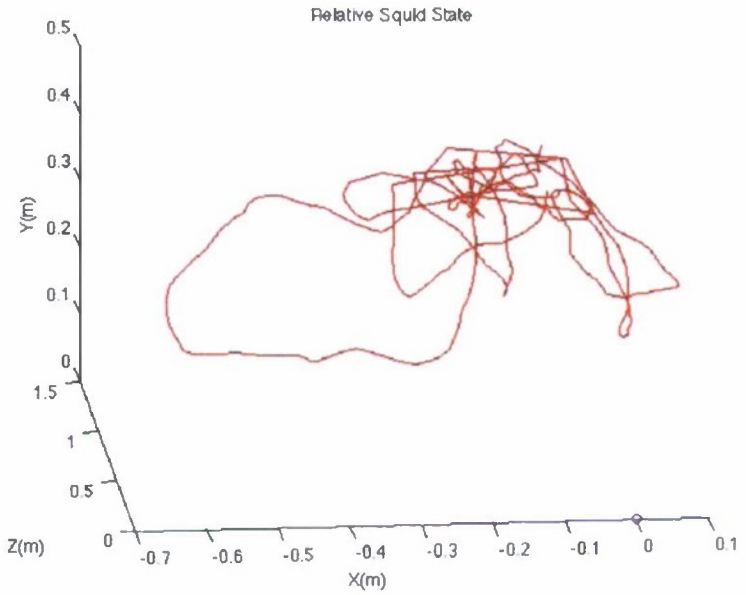


Figure 3: Relative Position states from an example data set.

3. Student Involvement & Technical Reports

3.1 Winter-Spring 2009: BIO 400 student and CPE 450 Capstone team
For 6 months, a BIO 400 student (Ryan Poppin) interested in technology and biological research analyzed videos for various behaviors (e.g. feeding, predator avoidance, etc), annotated, and archived 6 gigabytes of data for easy retrieval by subsequent researchers. A group of 4 CPR 450 undergraduate seniors researched squid behavior, size, shape, and movement from lab videos to establish techniques for tracking squid autonomously using AUVs. Their work led to the development of preliminary image processing software.

3.2 Summer 2009: Summer Research Internship
For several months during the summer, undergraduate student Robbie Plankenhorn researched methods for identifying squid from videos obtained from Hawaii. He established calibration routines, and mathematical equations for positioning state estimates. His work resulted in the first state estimates of squid position over time, relative to the robot's position.

3.3 Fall - Winter 2010: Master's Thesis
Master's student Matthew Schlachtman first used data sets from Hawaii to learn and develop image-processing strategies within an independent study course (CPE 599). Within this study, he refined Plankenhorn's work and created 3D plots of squid position over time, and videos of squid motion annotated with squid identification target to demonstrate the squid tracking capabilities. This work formed the basis of learning for his thesis:

Using Monocular Vision and Image Correlation to Accomplish Autonomous Localization. Schlachtman, M. California Polytechnic State University. M.S. Thesis, 2010. (http://digitalcommons.calpoly.edu/theses/320/)

"Tracking Shallow Water Squid via an Underwater Robot System"

3.4 Fall 2010: Research Volunteer and Senior projects
This August of 2010, a past NSF REU intern (Jeremy Edmonds) from Cuesta College and a new senior project student will work to finalize absolute positioning of the squid. Most importantly, absolute positioning of the squid during foraging behavior will be obtained. These data will be the basis of a peer-reviewed publication in a robotics and/or biological journal to be determined.

This work has also been presented at the robotics company Willow Garage, Cal Poly Open House, and Cal Poly Orientation week.

3.5 Summer 2010
The 2009 grant has led to a new C3RP proposal entitled "Developing a Visual Tracking System for Small Marine Animal Research". The technology to track animals reliably over their home ranges has only been rigorously demonstrated in larger animals (>300 gms) due to power and size constraints of modern tracking systems. Acoustic tags are currently the smallest tags available and cost hundreds of dollars per unit. In this study, we plan to design, build and implement optical tags that are about one-third the size and weight of acoustic tags. We anticipate that this will be possible using inexpensive surface-mount electronic components, resulting in a micro tag that costs a few dollars per unit. This approach will be designed to work synergistically with our developing system for visually tracking the small tropical squid *Euprymna scolopes* using remotely operated vehicles. Drs. Fidopiastis and Clark are planning to collaborate with Dr. Vladimir Prodanov (Cal Poly Electrical Engineering) to design and then attach miniature LED chips on squid to augment our ability to track them autonomously. We anticipate that our approach will lead to a major advancement in how small marine animal tracking research is conducted.

Path Following Algorithms and Performance Optimization

Principal Investigator:

John Seng, Ph. D.
Computer Science
California Polytechnic State University
San Luis Obispo, CA

Abstract

Automated machine vision systems enhance the sensing capability of unmanned military combat vehicles. Machine vision systems use software algorithms to extract information from a video sequence. A task that machine vision can enable in an unmanned vehicle is that of path following. Path following algorithms locate safe regions in an image for vehicle traversal by classifying which pixels represent drivable regions and which are background areas.

While highly desirable, machine vision systems are often computationally intensive. The computation required per video frame can be a significant limitation. In a recent trend, multi-core processors have appeared on the market as circuit technology has allowed manufacturers to fit multiple processors into a single chip. These chips have multiple processors which can potentially increase computational power several fold. While processing cores are easily added to a system, the benefit of additional processing cores is not easily obtained. Utilizing the cores in a multi-core system requires developers to rework software algorithms in order to obtain the highest performance from these new chips.

In this work, we investigate vision algorithms to perform path following in video sequences. We describe a vision algorithm that utilizes multiples visual cues and combines them via a particle filter to determine the most likely location for the path ahead.

The software was developed using an existing code base that makes use of the Intel OpenCV computer vision library.

Introduction

This work looks to further develop computer vision algorithms which enable autonomous robot navigation in environments using path following. Specifically, this work studies autonomous path detection in the context of unmarked and potentially unstructured environments. An unstructured path area can be any path in front of a vehicle (e.g. water canal passage, sidewalk, dirt path).

As mandated by the United States govemment, a significant fraction of future combat vehicles will be required to be unmanned [1]. Autonomous and remotely operated unmanned vehicles have proven themselves in a number of combat situations, and the use of such military vehicles will only increase in the future. Unmanned vehicles provide increased safety and capability for military operators. In order for future unmanned vehicles to function with greater autonomy, they must employ more advanced sensing techniques, one of which is vision-based capability.

A task that is commonly required of autonomous vehicles is path-based following. Whether a vehicle is a naval vessel navigating through a canal or river, a land vehicle driving on a road, or a smaller robot driving on a sidewalk, the task of path following involves detecting an unstructured path in front of the vehicle and determining where the boundaries are for safe travel. Once path boundaries have been dynamically detected, the path area in front of the robot

can be used to plan a safe travel route. Detection of what areas ahead of a robot for safe travel can be determined using active ranging sensors such as laser range finders, but these typically provide a highly localized view of the surface ahead or have limited flexibility in what types of data they can provide. A vision-based system can be used to augment the sensor information provided by such ranging sensors. Computer vision provides different types of data from a range sensor and their range is typically further than what can be provided by an active range sensor.

Detecting the path ahead can be performed using a number of features that are visible in an image. These features can be based on the path color, visible edge boundaries, texture characteristics, or directional visual gradients. Our previous work [3] has focused on using a color-based approach which employs color histograms to classify image pixels as drivable areas and background (non-drivable). The primary thrust for this grant work is to develop color and edge-based path detection algorithms which can be deployed on a mobile robot. By utilizing multiple visual cues, we look to develop algorithms with increased path following robustness.

This work fits well with the application interests of the Office of Naval Research. As warfare moves toward the utilization of unmanned systems, there will be a greater need for autonomous navigation in complex environments. Unmanned combat vehicles will be utilized by the military in both large and small scales. Increasing the autonomy of unmanned vehicles is necessary in order to enhance future combat capabilities. The work in this grant seeks to enable robots to achieve greater autonomy in an unstructured environment that a combat vehicle may encounter on the battlefield.

Background/Prior Work

Several studies have been performed which deal with unstructured path detection [3-13]. Detecting paths can be performed using a number of computer vision approaches: edge detection [4], statistical color-based models [6,8,9,11,12], path shape models [7,10], and path direction detection [13]. In our work, we build upon our color histogram research by utilizing additional information available in the image via edge detection.

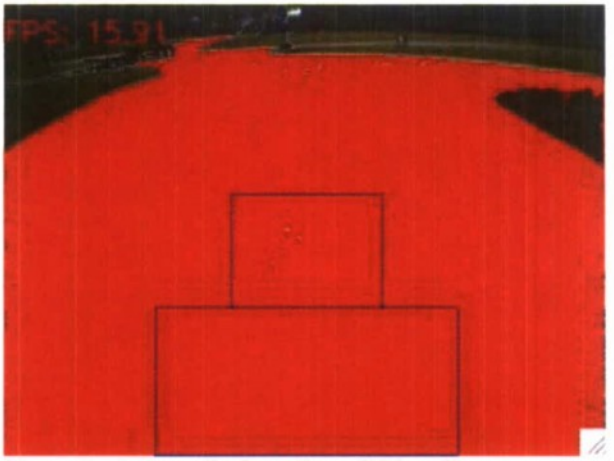


Figure 1: Automatic detection of sidewalk area (shown in red).

In our prior work [3], we have shown that color histograms are an effective tool in developing path following algorithms. In that work, we demonstrate that the sidewalk area in front of a robot can be modeled using a 2-dimension hue and saturation color histogram and likewise, the background area can be modeled with a histogram of similar parameters. Using these color histograms, a probabilistic model is used to classify whether a pixel belongs to a drivable area. Once all the pixels are classified, a path shape model is fitted to the area in front of the robot. Figure 1 shows the output of our current algorithm where the detected sidewalk pixels are drawn in red. Our preliminary work is completely color-based and does not utilize any additional information available from the image. A limitation of this technique is that is does not utilize any edge information available from the image.

In addition to the current color histogram technique, we incorporate particle filter and edge detection techniques that can improve the robustness of the algorithm. We believe that color information is necessary to detect homogeneous pixels in a pathway and that pathway edges contain information that can be used to enable better path shape model fitting.

Figure 2: An example input image of a roadway.

Approach Overview

Given a video image stream of an unstructured path ahead, there are many approaches that an algorithm can use to detect the path. The image shown in Figure 2 will be used as the image

for discussion for the remainder of this report. This image is of a road in a rural environment. A color based approach looks for differences in the color between the pixels of the path and the pixels which are not part of the path. While this type of approach can be successful when the path is asphalt or concrete and the background is a different type of surface, it is difficult to just utilize color when the entire image is rather uniform. This may occur when a scene contains a significant amount of concrete, for example. In addition, another feature which is useful is that of edge information. Even in unstructured images where the path is not necessarily straight, the edge information in the images can provided a significant amount of information to a path-following algorithm. Edge information can be useful, but may be misleading without color information as well.

In our work, we look at utilizing both color and edge information. Our algorithm maintains histograms models of the color of the path, maintains a histogram model of the background, and uses edge information to find the boundaries of the path. This approach builds on the work by Apostoloff [4]. Similar to [4], we utilize a particle filter to model the location of the edge of the path.

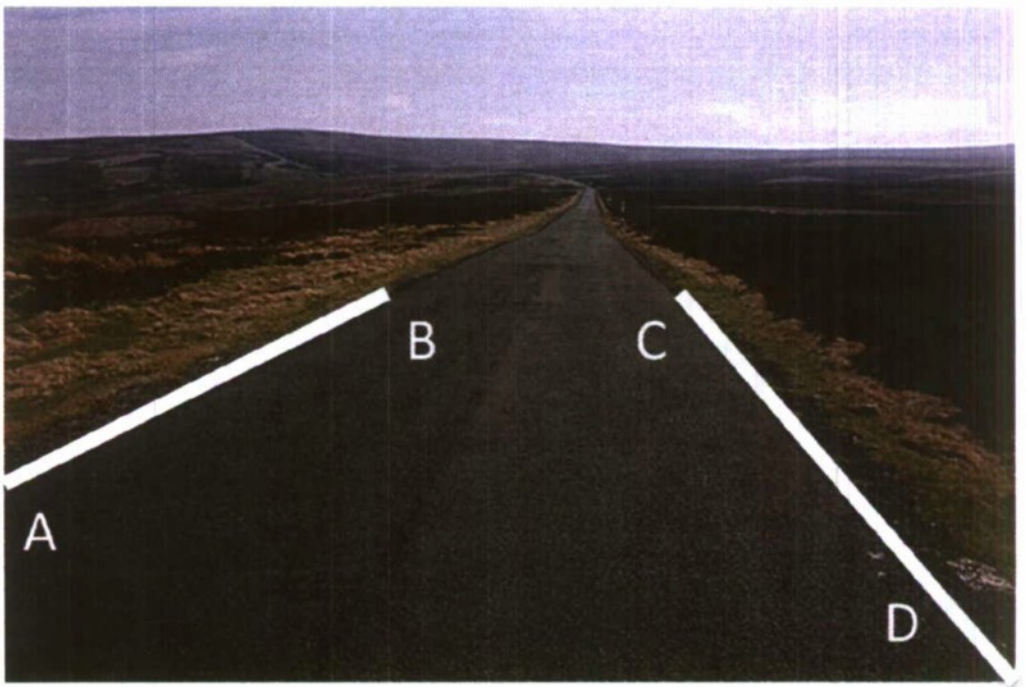


Figure 2 – The two lines represent the current edges of the detected path.

Path Model

We model the path using two straight lines: one line representing the left edge of the path and the other line representing the right edge of the path. For the sake of simplicity, we did not model intersections or curved paths. The path model can be seen in Figure 2. Our path model

consists of four points. Points A and B represent the left edge of the path and points C and D represent the right edge.

We maintain a *current* path model which is the path model which is the current estimation of where the path is current located. In our algorithm we utilize a particle filter, which will be discussed later, to select the coordinates of the current path model. Currently, this path model is updated from one frame to the next without regard to the previous path model.

In the path model representation, the row coordinates of the left and right edges do not necessarily have to be the same. We selected this approach in order to maintain maximum generality for the path.

Path Color Model

Knowledge of the current path color is advantageous as it allows the algorithm to identify pixels in an image which have a high likelihood of being part of the path ahead. In order to maintain a model of the path color, we utilize multiple histograms. This section of the report outlines our histogram path color model.

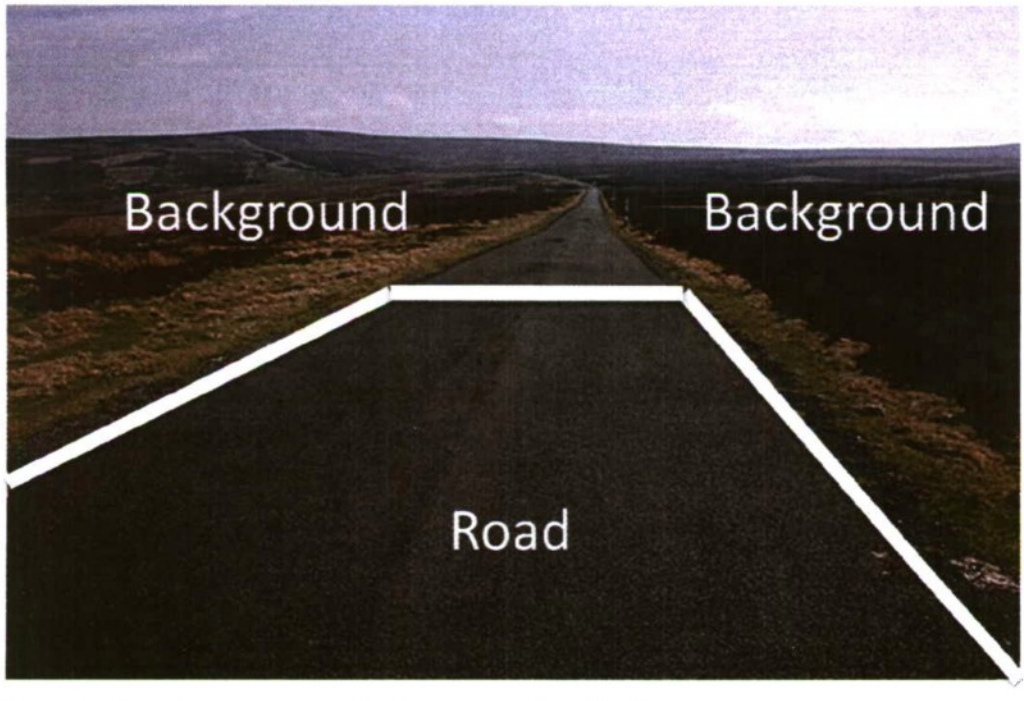


Figure 3 – The labeled areas are the detected road and background.

Figure 3 shows the distinction between the pixels of the road (within the bounds of an example path model) and the remaining background pixels. Our approach is to build histograms of the colors of the pixels within the currently detected road and also a histogram of the current background.

A histogram represents a distribution of data points over a representation space. In the case of our work, the representation space is the hue dimension of the pixels in the HSV colorspace. Each pixel has a hue, saturation, and value component in its representation and we utilize the hue dimension in our model of the path color.

A histogram model of the path color is the distribution of hues of pixels that lie just inside the current path model. We found that utilizing just a one-dimensional histogram of the hue provided satisfactory results.

The metric used to measure the difference between the a path color histogram and a candidate histogram is the following:

$$\text{Histogram difference} = \sum (h_m - h_t)^2$$

This metric is the sum of squared difference of each histogram bin where h_m is the bin value in the current model and h_t is the bin value of a model under test. In our current implementation, we divide the hue space into 32 discrete histogram bins. We empirically determined that 32 bins is a good balance between computation time and overall accuracy.

The current path histogram is obtained by computing a hue histogram over all the pixels within the current path model (i.e. the pixels in the foreground of the image). The hue component of all the pixels within the model is utilized in this histogram. In the current path histogram, the hue values are represented as 8-bit numbers and placed into one of thirty-two histogram bins. This histogram is normalized for accurate comparison against test histograms.

Maintaining a model of the path color brings some challenges in that the path color may change due to shadows, varying path conditions, and the presence of small objects on the path. In order to handle this fact, we maintain several models of the path color. In our implementation we maintain 4 separate histogram models of the path color. When a new video image is captured, the path histogram for that image is compared against all 4 path models. If the current path histogram is within a threshold difference of one of the 4 stored models, then the current path histogram is used to replace the existing path histogram – if multiple models match, then the closest matching histogram is replace. The replacement path histogram is also updated to note that it was recently inserted into the array of 4 path models. If the current path histogram does not match any of the 4 histogram models, then it is assumed that the path is changing color and that the path models need to be updated. If this occurs, then the path model that was updated longest ago is replaced. This path model update algorithm allows us to maintain a model for up to 4 different surfaces while gradually updating the models as the path changes color. The least squares difference was converted to represent a probability of path color where a smaller difference represented a higher probability. This probability we called p_{road_color}.

In terms of utilizing color histograms, we found that this approach provided the best performance for our algorithm. Initially, we utilized Gaussian mixture models to model the color of the pixels in the RGB space. While this technique provided satisfactory results, we found the overall performance to be significantly slower than the histogram approach.

Background Color Model

In order to accurately separate the color of the path from the remaining background, a model is maintained of the colors of the background. The background refers to any area that is not part of the currently traversed path. Example of background pixels can be trees in the distance, the sky above the horizon, and buildings that are next to a path. While the color of the pixels of these background objects may match pixels that are part of the path, the overall distribution of pixels will be different from the distribution of path pixels.

To capture a histogram of the background of the image, we utilize a random sampling technique where the data points consist of 2500 random selected pixels. These selected pixels are chosen from random locations in the image as long as they do not belong to the current path model.

This background color model is updated with each image that is captured. In order to compute a probability of a path model being correct, the same least squares computation used in computing the path color histogram is used for the background as well. This normalized probability is $p_{background\_color}$.

Edge Model

In addition to the road color models and the background model, we utilize additional edge information to locate a model of the path ahead. In images where the path is straight or even of a different texture from the path, an edge-based image can often provide additional information as to the location of the path boundary. This section covers how we utilize edge information.

For each incoming video image, the image is run through the Canny edge detector in order to find the strongest edges. This edge detector provides results that are often better than the simpler one-direction gradient edge detectors. Once the edge image is produced, then each candidate test path model is overlaid onto the edge image. The intuition for our approach is that candidate path models which likely represent the actual road will have a high number of pixels in common with the edge image. This is because the boundary of the path itself will appear in the edge image.

Therefore, for each candidate path model, we count the number of pixels that overlap with the path model and the edge. Additionally, to account for possible noise in the image, for each row we also add the number of edge pixels near the model. For our work, we look at edge pixels within 2 columns of a model pixel.

Once all of the overlapping and nearby edge pixels have been counted, this value is normalized with respect to the number of rows in the model. The resulting normalized number represents a probability that a particular test path model represents the actual path. We will refer to this probability as p_{canny_edge}.

Computation of Overall Probability and Model Selection

Once the probabilities are computed for a particular test model, the overall probability of that model needs to be computed. This overall probability needs to take into account the probability that the model is correct given the path color histograms, the background color histogram, and the Canny edge image. We utilize the following product as our overall probability:

$$\text{Probability of test model} = p_{road\_color} \times p_{background\_color} \times p_{canny\_edge}$$

Given the overall probability of a particular model under test, it is necessary to select an estimator which can select one model as the best current model. We utilize a particle filter for this process. A particle filter is effective in modeling complex densities that are non-standard. For our work, we utilize 200 particles for each iteration. Each particle represent one possible representation of the left and right path edges. This representation is depicted in Figure 4. The white lines on the left represent a possible left path edge, while the red lines on the right represent a possible right edge. We model the left and right edges as one particle, so for a given left edge, there is a corresponding right edge.

Upon initialization, the particles are randomly distributed throughout the state space. Within a few iterations, the particles converge to the path edges. For the estimate of the current path, we use the most likely particle.

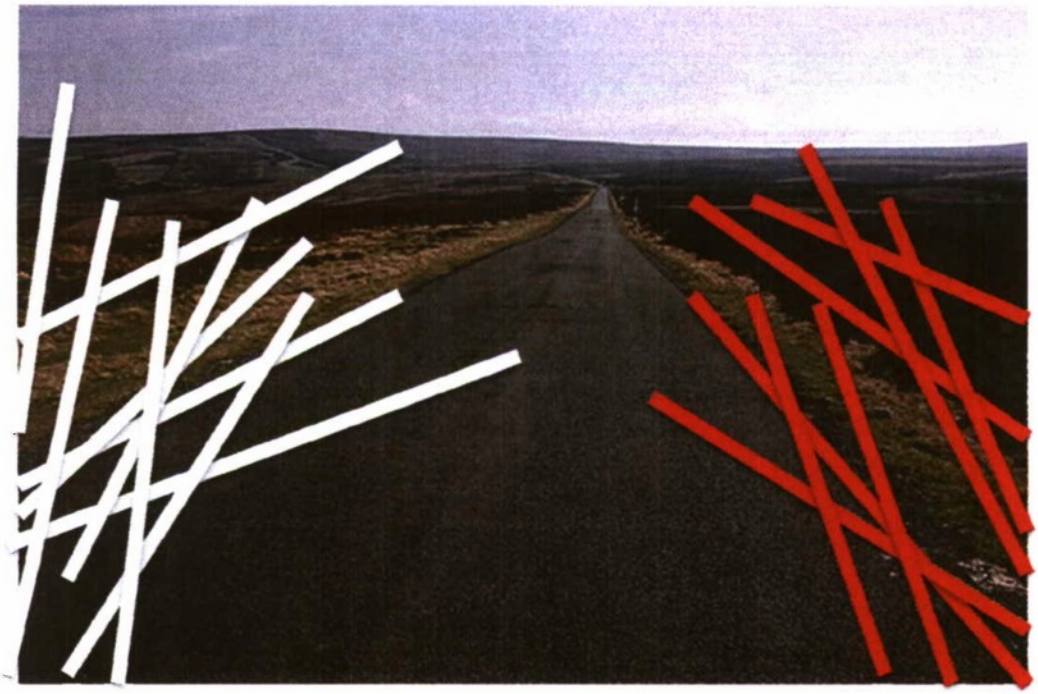


Figure 4 – White lines represent possible left edges and red lines represent possible right edges.

Performance and Algorithm Limitations

We empirically tested the algorithm on a section of unstructured sidewalk on the Cal Poly campus. We utilized a sidewalk section which contained unmarked intersections, various angular edges, and sidewalk color changes. We were able to observe the following characteristics.

Firstly, in sections of continuous sidewalk which contained the same color of concrete or asphalt, the algorithm performed quite well. The algorithm was able to track the path in the presence of rapid lighting changes as well as various forms of shadows. Without the multiple path color models, the algorithm would lose track during rapid lighting changes. Often, one model would model the duration of bright light and another model would model a darker lighting condition (e.g. during lighting changes caused by building shadows).

Secondly, the use of edge-based information was necessary when there were partial shadows across a path. In the event of partial shadows, the shadow itself would cast a strong edge on the image. Typically, there would be a path edge that is present in the shadow and this provided the algorithm with enough information to continue to track the path edge.

Thirdly, it was difficult for the algorithm to maintain track of the path during severe changes in color of the path. This type of color change would occur when transitioning from one color of concrete to a section of stained concrete. The different color concrete would not match any of the path color models and for a momentary period of time, the algorithm would lose track of the path.

Conclusion

In this work, we present a machine vision algorithm used to detect unstructured paths in video sequences. This path may be present in front of an unmanned military vehicle driving forward. The algorithm utilizes multiple color histograms to model the color of the path ahead. An incoming video image is matched to the closest matching path color model. As the path color changes, the model of the path color is updated. A model of the background region is maintained and the edges of the pathway are tracked as well. These three visual cues are combined and a particle filter is used to estimate the location of the edges of the pathway.

References

[1] **National Defense Authorization Act for Fiscal Year 2001**, Public Law 106-398

[2] **Team-Based Project Design of an Autonomous Robot**, Thomas J. Norrie, John S. Seng, *The 2007 International Conference on Frontiers in Education: Computer Science and Computer Engineering*, June, 2007

[3] **Sidewalk Following Using Color Histograms,** John S. Seng, Thomas J. Norrie, *Computer Science in the Colleges Southern Region Conference*, April, 2008

[4] **Robust Vision Based Lane Tracking Using Multiple Cues and Particle Filtering,** N. Apostoloff, A. Zelinsky, *Proceedings of the Intelligent Vehicles Symposium*, June, 2003

[5] **Color Model-Based Real-Time Learning for Road Following**, Ceryen Tan, Tsai Hong, Tommy Chang, and Michael Shneier, *Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference*, September, 2006

[6] **A Novel Lane Detection Algorithm Based on Support Vector Machine**, Hao Zhang, Dibo Hou, and Zekui Zhou, *Progress In Electromagnetics Research Symposium 2005*, August, 2005

[7] **Vision-based Real-time Road Detection in Urban Traffic**, Jianye Lu, Ming Yang, Hong Wang, Bo Zhang, *Proceedings of SPIE Photonics West 2002 - Electronic Imaging 2002: Science and Technology*, January, 2002

[8] **Performance Evaluation of Road Detection and Following Algorithms**, Tsai Hong, Aya Takeuchi, Mike Foedisch, Michael Shneier, *Proceedings of SPIE Vol. 5609 Mobile Robots XVII*, December, 2000

[9] **Self-supervised Monocular Road Detection in Desert Terrain**, Hendrik Dahlkamp, Adrian Kaehler, David Stavens, Sebastian Thrun, and Gary Bradski, *Proceedings of Robotics: Science and Systems*, August, 2006

[10] **Combined Dynamic Tracking and Recognition of Curves with Application to Road Detection**, Jean-Philippe Tarel, Frederic Guichard, *Proceedings of the 2000 International Conference on Image Processing*, September, 2000

[11] **Color Vision for Road Following**, Jill Crisman and Charles Thorpe, *Proceeding of the SPIE Conference on Mobile Robots*, November, 1988

[12] **UNSCARF, A Color Vision System for the Detection of Unstructured Roads**, Jill Crisman and Charles Thorpe, *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, April 1991

[13] **Texture-Based Vanishing Point Voting for Road Shape Estimation,** C. Rasmussen, *Proceedings of the British Machine Vision Conference 2004*, September, 2004

Feature Selection and Boosted Classification Algorithms for Pedestrian Detection

Project Investigator:

Samuel J. Frame, Ph. D.
Statistics
California Polytechnic State University
San Luis Obispo, CA

1 Summary of Work

Modern surveillance scenarios are saturated with digital video recorders, autonomous smart weapons, and sensing platforms that contain a wide variety of sophisticated sensors. As such, the ability to automatically, accurately, and quickly detect objects of interest is a capability critical to the mission of civilian law enforcement agencies, Homeland Security, and the Department of Defense. Rapid, real-time object detection is a necessary component of the Navy's mission to conduct urban warfare, automatic target recognition (ATR), tracking, and "end game" delivery of autonomous weapon systems. The research we conduct will further develop the theoretical and computational methods needed to detect objects of interest captured in real-time digital video imagery and a wide variety of hyperspectral imagery.

The specific object detection problem we consider is pedestrian detection in video imagery. For evaluation and testing, we use the benchmark data set *DaimlerChrysler Pedestrian Classification Benchmark Dataset* provided by DaimlerChrysler [8], [9], [14], [15]. The data set contains thousands of images which either have or do not have a pedestrian present in each image.

Our previous study largely focused on the C++ implementation of our methods, which included the capability to run Monte Carlo simulations to assess performance variability. Our current study focuses on two aspects. First, we use the feature selection implicitly occurring to iteratively expand the information used for classification. Second, we develop a weighted Expectation-Maximization (EM) method to estimate mixture model classifier parameters. We have completed the below tasks.

- Developed a method to iteratively select classifier features
- Developed a weighted Expectation-Maximization algorithm
- Implemented our new methods in R
- Tested and evaluated our new methods using simulated data and the DaimlerChrysler data
- Redesigned our base C++ implementation to allow for future implementations

In Section (2), we review foundations of our previous work which we extend. Sections (3) and (4) describe the theoretical aspects of our current work. Sections (5) and (6) discuss our current implementations and applications. We conclude this final report with suggestions for future research in Section (7).

2 Review of Previous Work

The goal of *AdaBoost* is to improve the classification capabilities of any classification method by forming an aggregate or *cascade of classifiers* (see Figure 1). In particular, boosting methods are designed to take simple, computationally inexpensive classification methods and increase their ability to perform accurate, reliable, and rapid classification. Our previous research efforts rely only on very simple and fast classification methods.

The classification problem we consider is to identify pedestrians in video imagery. As such, this constitutes a binary classification problem. Let Y be the event that an image contains a pedestrian, $Y \in \{0, 1\}$, where $y = 0$ if no pedestrian is present and $y = 1$ if a pedestrian is present. To discriminate between pedestrians and non-pedestrians, each image yields a set of features we

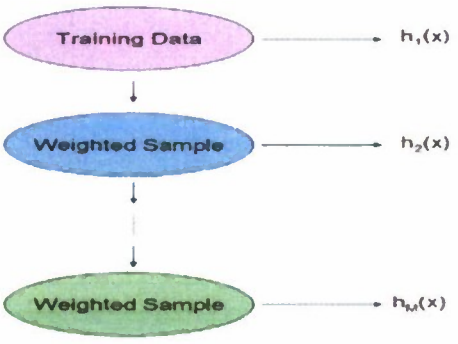


Figure 1: AdaBoost Schema

denote by X which may either be a scalar, vector, or matrix. For our exposition here, we assume the features are scalar. Given a set of n *training images*, our data are $(x_1, y_1), \ldots, (x_n, y_n)$. We give a general description of the AdaBoost algorithm below [2], [5], [14], [15].

- Let $h(x)$ be a classification method
- Initialize the weights of each training sample as equally likely,

$$w_{1,i} = \begin{cases} \frac{1}{n_p} & y_i = 1 \\ \frac{1}{n_{np}} & y_i = 0 \end{cases} \tag{1}$$

  for $i = 1, \ldots, n$; and where n_p is the number of pedestrian training samples, n_{np} is the number of non-pedestrian training, and $n = n_p + n_{np}$.

- Let M be the number of boosting stages used. For $m = 1, \ldots, M$:

  1. Normalize the weights $w_{m,i} = \frac{w_{,i}}{\sum_{i=1}^{n} w_{m,i}}$ for $i = 1, \ldots, n$. Here, $w_{m,i}$ is the weight the i^{th} training sample receives in the m^{th} stage of the boosting algorithm.
  2. Learn the classifier $h_m(x)$ using the weights $w_{m,i}$. Use $h_m(x)$ to classify the training samples and evaluate the classification error

$$e_i = |h_m(x_i) - y_i| \tag{2}$$

$$\varepsilon_m = \sum_{i=1}^{n} w_{m,i} e_i \tag{3}$$

  where $h_m(x_i) \in \{0, 1\}$, $e_i = 0$ if the i^{th} sample has been correctly classified, and $e_i = 1$ if the i^{th} sample has been incorrectly classified.

  3. Update the weights by

$$\beta_m = \frac{\varepsilon_m}{1 - \varepsilon_m} \tag{4}$$

$$w_{m+1,i} = w_{m,i} \beta_m^{1-e_i} \tag{5}$$

where β_m acts as a "dampening" effect to give less weight to samples which have been correctly classified.

- The final *strong* classifier is given by

$$H(x) = \begin{cases} 1 & \sum_{m=1}^{M} \alpha_m h_m(x) \geq \frac{1}{2} \sum_{m=1}^{M} \alpha_m \\ 0 & \text{else} \end{cases} \tag{6}$$

where $\alpha_m = log\left(\frac{1}{\beta_m}\right)$.

This is a generic exposition of a boosting algorithm for a binary classification problem. We have left the choice of the classifier, h, general. Next, we discuss the simple and rapid classifiers used in previous work.

2.1 Threshold Classifier

The choice of the classifier, h_m, used by current methods is a threshold classifier [1], [9], [14], [15]

$$h_m(x) = \begin{cases} 1 & x \leq \theta_m \\ 0 & \text{else} \end{cases} \tag{7}$$

where θ_m a threshold chosen such that the classification error rate (which uses only the training data) is as small as possible. During the learning process, several choices of θ_m are evaluated using the training data. The θ_m which minimizes the classification error is then selected [14]. Although attractive because of computational simplicity, the technique is not probabilistic and does not consider any measure of feature variability. For equal costs of miss-classification and equally likely prior probabilities of images containing a pedestrian and non-pedestrian, the threshold classifier reduces to a linear discriminant function where the variance of both classes (pedestrian, non-pedestrian) is the same. This yields θ_m which is a function of the pedestrian weighted average $\overline{x}_p$ and non-pedestrian weighted average $\overline{x}_{np}$ [1].

2.2 Quadratic Discriminant Classifier

The threshold classifier is a special case of a linear discriminant. Logically, we also consider quadratic discriminant analysis (QDA) [2], [5], [10]. Again, suppose the costs of miss-classification are equal and the prior probabilities are the same. Assume the features, x, have a Normal distribution. A simplified version of a quadratic discriminant function gives the following classifier

$$h_m(x) = \begin{cases} 1 & f_{1,m}(x) > f_{0,m}(x) \\ 0 & \text{else} \end{cases} \tag{8}$$

where f is the Normal distribution density function, and $f_{1,m}$ and $f_{0,m}$ are the estimated density functions for the pedestrian and non-pedestrian classes respectively. Let $\underline{w}_m = (w_{m,1}, \ldots, w_{m,2})$ be the boosting weights at stage m. Estimating the density functions entails obtaining a weighted sample average and variance according to the pedestrian and non-pedestrian training samples, and the weights $\underline{w}_m$.

3 Feature Selection

In our previous study, each stage of the boosted cascade of classifiers only uses a single univariate feature. Given the set of boosting weights at stage m, $\underline{w}_m$, the single feature is chosen to minimize the miss-classification rate of the training samples. In this iterative way, the classifier cascade is implicity selecting the optimal features for classification. We have developed and implemented a method which potentially, *iteratively* expands the feature set in each boosting stage. We have developed this as a multivariate extension of the QDA classifier already implemented and tested.

In the first stage of the boosting algorithm $m = 1$, select the best univariate feature. In successive stages of the boosting algorithm $m > 1$, do the following

- Tentatively, expand the feature set to include a new, single feature from the set of unused features
- Learn the tentative multivariate classifier h_m
- Evaluate the tentative multivariate classifier h_m
- If the new feature set results in better classification performance than observed in the previous boosting stage, retain the new feature set

If no additional feature improves classification performance, do not expand the feature set. For a multivariate feature $\underline{x}$, the multivariate classifier h_m will have the form

$$h_m(\underline{x}) = \begin{cases} 1 & f_{1,m}(\underline{x}) > f_{0,m}(\underline{x}) \\ 0 & \text{else} \end{cases} \tag{9}$$

where $f()$ is the multivariate Normal distribution with a mean vector and covariance matrix estimated according to the boosting weights $\underline{w}_m$.

4 Weighted Expectation-Maximization

Generalized Mixture Models (GMMs) have been widely used to develop classification systems [3] [4]. In most mixture model settings, however, it is assumed the data are equally weighted. The boosting method we employ re-weights the data in each boosting stage. In order to use a GMM as the choice of the classifier h_m in our boosting algorithm, we must develop a method for estimating the GMM model parameters when data are weighted. Here, we develop a weighted Expectation-Maximization (EM) estimation method for estimating the parameters of a GMM.

The GMM framework we use allows pre-existing classes to probabilistically own mixture model components by way of a probability mass function. Specifically, we consider a two class problem where $c(x_i) \equiv y_i \in \{0, 1\}$, the x_i are univariate, and the mixture is comprised on Normal distributions [2] [5]. For a single x_i, the distribution is presumed to have the form

$$f(x_i|\Theta) = \sum_{k=1}^{K} \alpha_k f(x_i|\underline{\theta}_k) P(C = c(x_i)|k) \tag{10}$$

where $P(C = c(x_i)|k)$ represents the probability the k^{th} mixture generated data from class $c(x_i)$, $\underline{\theta}_k = (\mu_k, \sigma_k^2)$, and $\Theta = \{\alpha_1, \ldots, \alpha_K, \underline{\theta}_1, \ldots, \underline{\theta}_K\}$. For now, assume K is known. Let $X =$

$\{x_1, \ldots, x_N\}$ and let the x_i be independent. The observed log-likelihood is given by

$$log\left(L(X|\Theta)\right) = \sum_{i=1}^{N} log\left(w_i \sum_{k=1}^{K} \alpha_k f\left(x_i|\underline{\theta}_k\right) P(C = c(x_i)|k)\right) \tag{11}$$

where the w_i are the m^{th} boosting stage weights $w_{m,i}$. Given the weights and the data, we must estimate Θ and the mixture model class ownership probability mass functions $\{P(C = c|k), c = 0, 1\}$ for $k = 1, \ldots, K$. To estimate these parameters, we extend the standard EM method to account for the weights resulting from the boosting algorithm.

Expectation-Maximization (EM) involves two steps [2] [6]. In the "E-step" or expectation step, the estimation algorithm seeks to estimate which mixture model component generated each data point. This "component ownership" step results in a probability mass function over the components for each data point x_i. Let Θ_t be the current estimate of Θ at learning stage t. With weighted data, it is easy to show the E-step is unchanged. This yields the same $P\left(M_k|x_i, \Theta_t\right)$ for $k = 1, \ldots, K$ as found in [7]. The second step is the "M-step" or maximization step which results from maximizing the complete log-likelihood, involves the probability mass functions obtained from the "E-step," and provides maximum likelihood estimates of the remaining parameters. For weighted data, the "M-step" updates reflect the weights. The updates for the α_k (the mixing probabilities) and the μ_k (the component means) are given by

$$\alpha_k = \frac{\sum_{i=1}^{N} w_i P\left(M_k|x_i, \Theta_t\right)}{\sum_{k=1}^{K}\sum_{i=1}^{N} w_i P\left(M_k|x_i, \Theta_t\right)} \tag{12}$$

$$\mu_k = \frac{\sum_{i=1}^{N} w_i x_i P\left(M_k|x_i, \Theta_t\right)}{\sum_{i=1}^{N} w_i P\left(M_k|x_i, \Theta_t\right)} \tag{13}$$

and the "M-step" update is comparable for the σ_k^2 (the component variances). Similarly, the update for the probability mass function over the classes for each component is given by

$$P(C = c|k) = \frac{\sum_{i=1, c(x)=c}^{N} w_i P\left(M_k|x_i, \Theta_t\right)}{\sum_{i=1}^{N} w_i P\left(M_k|x_i, \Theta_t\right)} \tag{14}$$

for $k = 1, \ldots, K$. Here, the weights do not affect the "E-step" component ownership probabilities. Rather, they do affect all of the "M-step" calculations and result in weighted estimates of the mixing probabilities, means, variances, and component class ownership probabilities.

Previous classifier choices, h_m, had the general form of

$$h_m(x) = \begin{cases} 1 & f_{1,m}(x) > f_{0,m}(x) \\ 0 & \text{else} \end{cases} \tag{15}$$

Here, h_m does not explicitly involve $f(x_i|\Theta)$ as found in Equation (10). Rather, we require *a posteriori* class probabilities [4] [7]. For each data point, x_i, the *a posteriori* class probabilities are given by

$$P(C = c|x_i) = \frac{\sum_{k=1}^{K} \alpha_k f(x_i|\theta_k) P(C = c|k)}{\sum_{k=1}^{K} \alpha_k f(x_i|\theta_k)} \tag{16}$$

for $c = 0, 1$. The classifier has the form

$$h_m(x) = \begin{cases} 1 & P(C = 1|x) > 0.5 \\ 0 & \text{else} \end{cases} \tag{17}$$

5 Code Development

Our previous work focused exclusively on developing a C++ implementation of our methods. Prior to receiving funding for our previous research, we had developed a *proof of concept* implementation in Matlab. Our proof of concept implementation in Matlab provided support for our proposed research in our original proposal. From our previous funded research effort, we developed a stable, accurate, and fast C++ implementation of the already proven methods. A large component of the implementation was developing Monte Carlo simulation methods which give metrics that capture our algorithm's performance variability. We developed a flexible solution, that is capable of supporting the implementation of alternative, more advance feature selection and classification methods.

The methods we have developed during the current research effort had not previously been implemented as a proof of concept. As such, we focussed our code development into two areas. First, we developed a proof of concept implementation of these new methods in the R programming environment. Second, the proof of concept implementation has given us insights and experience to redevelop aspects of our C++ implementation to be more flexible. This will allow for an easier implementation of the new methods in the C++ environment.

The theoretical development components of this research are described in Section (3) and Section (4). The methodology for Section (3) is straightforward and easy to implement. However, the methodology described in Section (4) is novel and non-trivial to implement. The primary proof of concept implementation was the development, testing, and evaluation of the boosted cascade of mixture model classifiers which rely on the weighted EM learning algorithm. For pragmatic reasons, we chose to use the R statistical computing environment rather than Matlab.

Our primary reason for using the R environment was the reliable k-means function inherit to the program [2] [5] (past experience has revealed substantial, terminal flaws in Matlab's "stat" package), the ease of simulating data from a multivariate Normal distribution, and evaluating multivariate Normal densities. At each boosting stage with weights $\underline{w}_m$, the mixture model classifier and weighted EM do the following:

- Use k-means to initialize mixture model components for both the pedestrian and non-pedestrian classes
- Use the weighted EM learning method to independently learn the class mixture model components

- Use BIC to choose the number of class mixture model components with a maximum of 10 component per class [2]
- Combine the class mixture model components into a single mixture model
- Use the weighted EM learning method to learn the entire mixture model

The amount of weighted EM learning is controlled by monitoring the non-deceasing log-likelihood (with a maximum of 50 learning iterations). We have done extensive testing with our mixture model classifier and weighted EM learning method using diverse types of simulated data.

For our simulation studies, we have simulated data with a diverse set of characteristics. Specifically, we varied the number of components per class, the weights of each component, and the means and variances of the mixture components (ie component parameters). In every simulation experiment, our algorithm converges to a mixture model with the correct number of class mixture components, the correct mixture weights, and the correct component parameters. Convergence is always realized within the maximum number of weighted EM learning iterations.

6 Applications

In this section, we present our results using both simulated data and the DaimlerChrysler data. We will present computational and classification results using both of the methods developed during this research. We invite the reader to compare these results to those contained in our previous report.

6.1 Simulated Data

Before any new method can be used on real world data (such as the DaimlerChrysler data set), the method must be tested and validated using simulated data. Satisfactory results suggest that the method works properly. We have chosen to present results for a specific simulated data set. We have chosen this simulated data set based on the results obtained from using our mixture model classifier (using the weighted EM learning method) to evaluate the DaimlerChrysler data. In particular, the surprising number of mixture model components which is selected at each boosting stage for the DaimlerChrysler data. Our results with simulated data give us confidence in our algorithm.

Comparable to the simulated data contained in our previous report, we simulate data from a multivariate Normal distribution. For two groups (which can be viewed as the pedestrian and non-pedestrian groups), we simulated 100 training and testing for each group. For both groups, we use a $5 * 5$ identity matrix as the covariance structure. The population mean vectors are given below.

Group	Mean Vector
Pedestrian	(-3,-2,-1,0,1)
Non-pedestrian	(-3,-2,0,2,4)

The mean vectors are chosen because of the large overlap in the mean values for the first two features. It should be clear that the valuable discriminant information is contained in features three, four, and five.

When using our QDA classification method with expanding feature sets described in Section (3), we would expect that the fifth feature would be selected first. Subsequent features chosen in

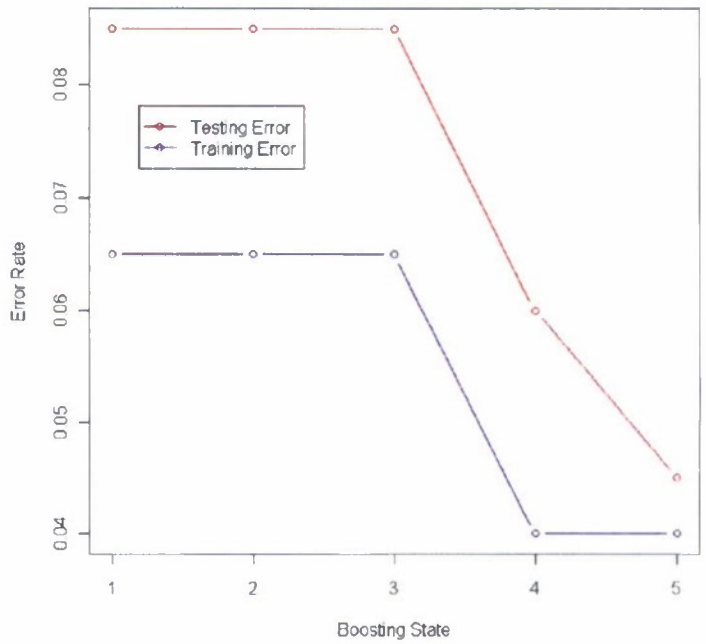


Figure 2: QDA classifier with expanding features: classification error rates for simulated data

the expanding feature set should be features four and three. Since features one and two contain no discriminant information, we expect them to be never chosen. For this experiment, we fit a cascade with $m = 5$ boosting stages. The below table shows which features are chosen at each boosting stage.

Boosting Stage	Features Used
1	5
2	5
3	4, 5
4	3, 4, 5
5	3, 4, 5

As we expect, the fifth feature is initially chosen. In boosting stage two, the weights $\underline{w}_2$ have been updated to reflect miss-classifications in the first boosting stage. Interestingly, no additional feature improves classification in the second boosting stage, and so the method only retains the fifth feature. In boosting stages three to five, the algorithm begins to expand the feature set to include the fourth and third features.

In Figure (2), we show the miss-classification rate of both the training and testing data in each boosting stage. By the final boosting stage, the miss-classification rate is about 4.5% (testing) and 4.0% (training). In our previous effort/report, we use a QDA classification method without expanding the feature set (ie only a univariate feature). For this method, the miss-classification rates are 7.5% (testing) and 6.0% (training). The results using the expanded feature set constitute a significant improvement in classification.

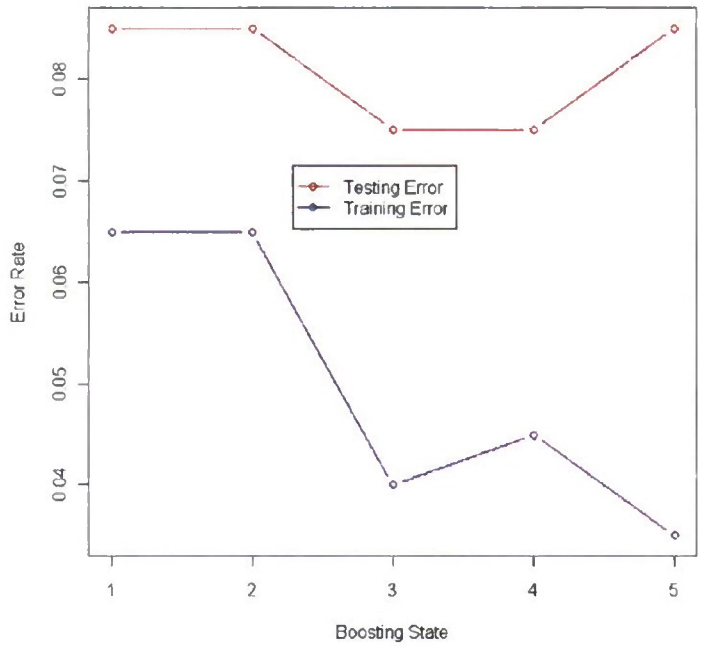


Figure 3: Mixture model classifier: classification error rates for simulated data

The data we simulate are univariate, and do not come from a mixture distribution. When using our mixture model classifier (using our weighted EM learning method) described in Section (4), we expect that the resulting mixture models only have a single component per class (in each boosting stage). Also, we expect that the fifth feature to be used almost elusively. For a cascade with $m = 5$ boosting stages, the below table shows which feature and how many mixture model components are chosen.

Boosting Stage	Feature Used	Number of Components
1	5	1
2	5	1
3	4	1
4	5	1
5	5	1

As we expect, the fifth feature is predominantly selected and the method only selects a single mixture model component per class. Figure (3) shows miss-classification rates for the training and testing data in each boosting stage. The features are univariate in each boosting stage. Furthermore, the data are not unimodal reflected by the choice of only a single mixture model component per class. Thus, it is reasonable that no significant improvement is made in comparison to our previous efforts based on a QDA classifier.

While our mixture model classifier shows no improvement using the simulated data, it does demonstrate the reliability of the method and the implementation. We have chosen to present results using this set of simulated data because of results obtained using the mixture model classifier with the DaimlerChrysler data.

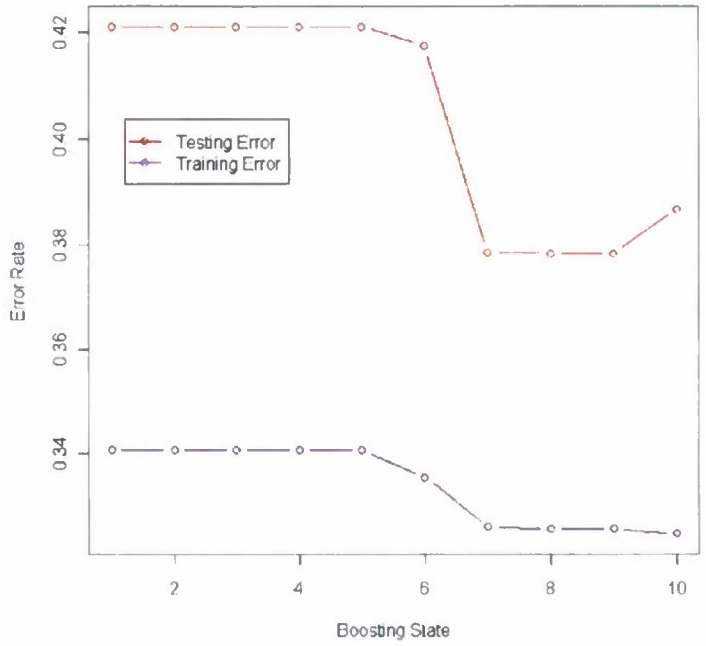


Figure 4: QDQ classifier with expanding features: classification error rates for DaimlerChrysler data

6.2 DaimlerChrysler Data

The *DaimlerChrysler Pedestrian Classification Benchmark Dataset* is publicly available data used to evaluate pedestrian detection algorithms [8] [9]. The data set consists of images containing pedestrians and non-pedestrians. For training, there are 14,400 and 15,000 pedestrian and non-pedestrian samples respectively. For testing, there are 9,600 and 10,000 pedestrian and non-pedestrian samples respectively. For the purposes of this research, we use the difference integral image (feature one) and motion features (features two through five) as done in [1] [14] [15]. The difference integral image (feature one) provides information about the change in an entire image relative to the previous image. The motion features give measures of object motion in the up (feature two), left (feature three), right (feature 4), and down (feature five) directions in the difference of two consecutive images. At no time during the course of this research did we propose to research feature extraction and/or image processing methods.

First, we employ our boosting algorithm using our QDA classifier with expanding feature sets using $m = 10$ boosting stages. Based on previous work by other researchers, we expect that the first, third, and fourth features to be the most useful. The following table shows which features are chosen at each boosting stage.

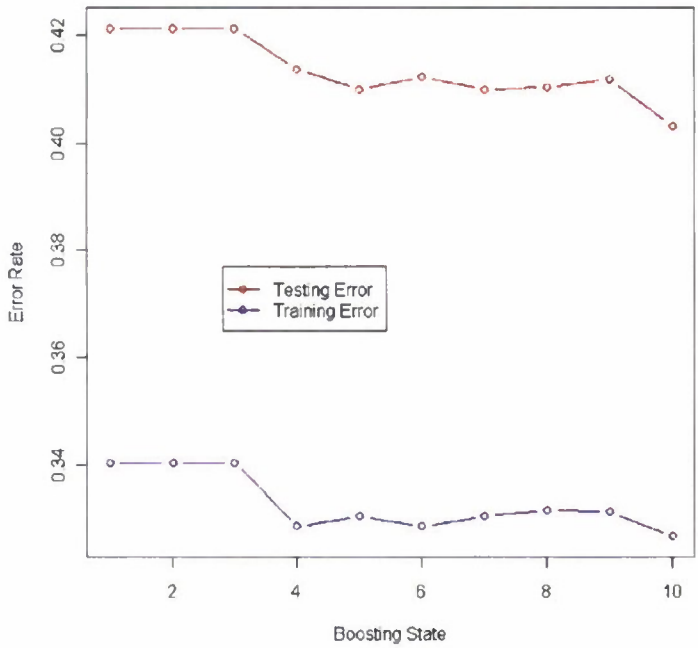


Figure 5: Mixture model classifier: classification error rates for DaimlerChrysler data

Boosting Stage	Features Used
1	1
3	1, 4
4	1, 3, 4
5	1, 2, 3, 4
7	1, 2, 3, 4, 5

By the end of the first four boosting stages, the feature set has expanded to include the first, third, and fourth features. All features are included by the seventh boosting stage. Figure (4) shows miss-classification rates for the training and testing samples. Interestingly, classification performance does not significantly improve until all five features are used. Clearly, the QDA classifier with expanding feature sets does better than random guessing, and the expanded feature set yields a significant decrease in miss-classification rates compared to our previous QDA classifier which only uses a single, univariate feature. However, our QDA classifier with expanding feature sets does not provide better classification rates than competing methods [9] [14].

A primary reason for developing and implementing a mixture model classifier (using our weighted EM learning method) was to improve classification performance using only the five simple features. We anticipated that the mixture model would select more than a single mixture model component per class. In our implementation, the number of components for each class is selected using BIC, rather than miss-classification rates of the training data. The following table shows now many mixture model components per class were chosen and which feature was chosen during each boosting stage.

Boosting Stage	Feature Used	Number of Components
1	2	1
2	3	1
3	1	1
4	5	1
5	2	1
6	1	1
7	2	1
8	2	1
9	4	1
10	5	1

While BIC is used to choose the number of components, the feature selected in each stage is chosen based on the miss-classification rate of the training data. All boosting stages result in a mixture model classifier with only a single mixture model component per class. Furthermore, the classifier cascade chooses features two and five for 6/10 boosting stages. The resulting miss-classification rates reflect the strange classifier cascade which is chosen. Figure (5) shows miss-classification rates for the training and testing samples. It is clear that the mixture model classifier does not perform as well as our QDA classifier with expanding feature sets. There are two explanations for this result. First, using the expanded feature set provides much better discrimination capability. Secondly, it does not appear that the motion features in the DaimlerChrysler data are generated my mixture models.

7 Future Research

We conclude this final report with suggestions for future research. The classification performance of our classifier cascade when applied to a benchmark data set such as the DaimlerChrysler data does not constitute a dramatic improvement over other methodologies. In order to increase the classifier cascade performance, we suggest (i) implementing a system which combines all of the methods we have developed and (ii) investigating different image processing and feature extraction methods.

The methods we present in this final report have been implemented, tested, and evaluated using the R software. As a final step, we suggest combining the mixture model classifier (using our weighted EM learning method) with expanding feature sets. Extending the mixture model classifier (and the weighted EM learning method) for multivariate data is straightforward. Iteratively expanding the feature set will utilize more discriminatory information. Combined together, this will produce a system which should yield better classification performance. Ultimately, these methods need to be implemented, tested, and evaluated in the C++ computing environment.

In fairness to this work, consider the simplicity of the classifier choices: threshold classifiers, QDA classifiers, and mixture model classifiers. Alone, each is a simple classification tool. When coupled together in a classifier cascade, these simple methods provide a final *strong* classifier which performs comparably to advanced methods such Support Vector Machines and Neural Networks [2] [5] [14]. Boosting methods were designed to take simple classification tools and make them better. The benefit to using them is the reduced cost in computational time.

Lastly, our work relies on using existing image processing and feature extraction methods [9] [14]. It should be clear these features do not provide ideal classification. One avenue of future research is to consider alternative image processing and feature extraction methods.

References

[1] Alldrin, N. (2005). "Detecting Pedestrians," Technical Report, Department of Computer Science, University of California, San Diego.

[2] Bishop, C. (2006). *Pattern Recognition and Machine Learning.* New York: Springer.

[3] Frame, S. and Jammalamadaka, S. R. (2007). "Generalized Mixture Models, Semi-supervised Learning, and Unknown Class Inference," *Advances in Data Analysis and Classification (ADAC): Theory, Methods, and Applications in Data Science.* New York: Springer.

[4] Frame, S. and Miller D. (2005). "Machine Learning for Robust Automatic Target Recognition: Phase 1 Final Report," Phase 1 Final Report for U.S. Air Force Research Laboratory Contract FA8650-04-M-1659.

[5] Hastie, T. and Tibshirani, R. and Friedman, J. (2001). *Elements of Statistical Learning.* New York: Springer.

[6] McLachlan, G. and Krishnan, T. (2004). *The EM Algorithm and Extensions.* New York: John Wiley and Sons.

[7] Miller, D. and Browning, J. (2003). "A Mixture Model and EM-based Algorithm for Class Discovery, Robust Classification, and Outlier Rejection in Mixed Labeled/Unlabeled Data Sets", *IEEE Trans. on Pattern Anal. and Machine Intell*, 1468-1483.

[8] Munder, S. and Gavrila D. M. (2006). DaimlerChrysler Pedestrian Classification Benchmark Dataset, (C) DaimlerChrysler AG.

[9] Munder, S. and Gavrila, D. M. (2006). "An Experimental Study on Pedestrian Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-6.

[10] Rencher, A. (1995). *Methods of Multivariate Analylsis.* New York: John Wiley and Sons.

[11] Schapire, R. and Singer Y. (1999). "Improved Boosting Algorithms Using Confidence-rated Predictions," *Machine Learning*, 297-336.

[12] Schapire, R. and Singer Y. (1999). "A Short Introduction to Boosting," *Journal of Japanese Society for Artificial Intelligence*, 771-780.

[13] Trosset, M. (2007). "Semisupervised Learning from Dissimilarity Data," Tech. Report No. 07-01 Department of Statistics, Indiana Univeristy, Bloomington.

[14] Viola, J., Jones, M. and Snow, D. (2005). "Detecting Pedestrials Using Patterns of Motion and Appearance," *International Journal of Computer Vision*, 153-161.

[15] Viola, J., and Jones, M. (2001). "Robust Real-time Oject Detection," Second International Workshop on Statistical and Computation Theories of Vision-Modeling, Learning, Computing, and Sampling.

SIMPLIFIED PROCEDURES FOR SEISMIC ANALYSIS AND DESIGN OF PIERS AND WHARVES IN MARINE OIL AND LNG TERMINALS

by

Rakesh K. Goel
California Polytechnic State University, San Luis Obispo

Research Conducted for the
California State Lands Commission
Contract No. C2005-051
and
Department of the Navy, Office of Naval Research
Award No. N00014-08-1-1209

Department of Civil and Environmental Engineering
California Polytechnic State University, San Luis Obispo, CA 93407
June 2010

Report No. CP/SEAM-08/01

EXECUTIVE SUMMARY

This investigation developed simplified procedures for the seismic analysis and design of pile supported wharves and piers in Marine Oil and LNG Terminals. A simplified coefficient-based approach is proposed for estimating seismic displacement demand for regular structures. This approach is adopted from the performance-based analysis procedure recently approved for buildings in the ASCE/SEI 41-06 standard (ASCE, 2007). A modal pushover analysis (MPA) approach is proposed for irregular structures. The MPA procedure accounts for the higher-mode effects that are important in irregular structures (Chopra and Goel, 2004). The acceptability of piles in terms of displacement ductility limitation, instead of the material strain limitation, is proposed. For this purpose, simplified expressions for estimating displacement ductility capacity of piles are recommended. These expressions are calibrated such that the material strain limits in Title 24, California Code of Regulations, Chapter 31F, informally known as the Marine Oil Terminal Engineering and Maintenance Standards (MOTEMS), would not be exceeded if the displacement ductility demand is kept below the proposed displacement ductility capacity. These simplified procedures can be used as an alternative to the procedures currently specified in the MOTEMS. The simplified procedures can be used for preliminary design or as a quick check on the results from detailed nonlinear analyses. The more sophisticated analysis methodology can still be used for final design.

The following is a summary of the procedures to estimate displacement demands and capacities for pile-supported wharves and piers.

DISPLACEMENT DEMAND

Regular Structures

It is proposed that the seismic displacement demand in a regular structure (MOTEMS 2007) be estimated from

$$\Delta_d = C_1 C_2 S_A \frac{T^2}{4\pi^2} \tag{1}$$

in which S_A is the spectral acceleration of the linear-elastic system at vibration period, T. The coefficient C_1 is given by

$$C_1 = \begin{cases} 1.0; & T > 1.0\text{s} \\ 1.0 + \dfrac{R-1}{aT^2}; & 0.2\text{s} < T \le 1.0\text{s} \\ 1.0 + \dfrac{R-1}{0.04a}; & T \le 0.2\text{s} \end{cases} \quad (2)$$

in which a is a site dependent eonstant equal to 130 for Site Class A and B, 90 for Site Class C, and 60 for Site Class D, E, and F (definition of Site Class is available in ASCE/SEI 41-06 standard), and R is the ratio of the elastie and yield strength of the system and is defined as

$$R = \frac{S_A}{g}\frac{W}{V_y} \quad (3)$$

where W is the seismie weight of the system, V_y is the yield force (or base shear) of the system, and g is the aeeeleration due to gravity. The coefficient C_2 is given by

$$C_2 = \begin{cases} 1.0; & T > 0.7\text{s} \\ 1 + \dfrac{1}{800}\left(\dfrac{R-1}{T}\right)^2; & T \le 0.7\text{s} \end{cases} \quad (4)$$

Use of Equation (1) to eompute the displaeement demand should be restricted to systems with $R \le R_{\max}$ where $R_{\max}$ is given by

$$R_{\max} = \frac{\bar{\Delta}_d}{\Delta_y} + \frac{|\alpha_e|^{-t}}{4} \quad (5)$$

in whieh $\bar{\Delta}_d$ is the smaller of the eomputed displacement demand, Δ_d, from Equation (1) or the displaeement eorresponding to the maximum strength in the pushover eurve, Δ_y is the yield displacement of the idealized bilinear foree-deformation eurve, t is a eonstant eomputed from

$$t = 1 + 0.15\ln(T) \quad (6)$$

and α_e is the effeetive post-elastie stiffness ratio eomputed from

$$\alpha_e = \alpha_{P-\Delta} + \lambda(\alpha_2 - \alpha_{P-\Delta}) \quad (7)$$

where λ is a near-field effeet faetor equal to 0.8 for sites that are subjected to near-field effects

and 0.2 for sites that are not subjected to near field effects. The near field effects may be considered to exist if the 1 second spectral value, S_1, at the site for the maximum considered earthquake is equal to or exceeds 0.6g. The P-Delta stiffness ratio, $\alpha_{P-\Delta}$, and the maximum negative post-elastic stiffness ratio, α_2, in Equation (7) are estimated from the idealized force-deformation curve.

Irregular Structures

A modal pushover analysis (MPA) procedure is proposed to estimate displacement demands in irregular Marine Oil and LNG Terminal structures (MOTEMS 2007). The following is a step-by-step summary of the MPA procedure:

1. Compute the natural frequencies, ω_n and modes, $\boldsymbol{\phi}_n$, for linearly elastic vibration of the irregular Marine Oil and LNG Terminal structure.
2. Select a reference point where the displacement, u_{rn}, is to be monitored in the selected direction of analysis during the pushover analysis. Ideally, this reference point should be the location on the structure with largest value of ϕ_{rn} in the selected direction of analysis.
3. For the nth-mode, develop the pushover curve, $V_{bn} - u_{rn}$, for the nth modal force distribution, $\mathbf{s}_n^* = \mathbf{M}\boldsymbol{\phi}_n$, where $\mathbf{M}$ is the mass matrix of the structure, and $\boldsymbol{\phi}_n$ is the nth mode shape. The base shear V_{bn} should be monitored in the same direction as the direction of the selected reference point displacement u_{rn}.
4. Convert the $V_{bn} - u_{rn}$ pushover curve to the force-displacement, $F_{sn}/L_n - D_n$, relation for the nth -"mode" inelastic SDF system by utilizing $F_{sn}/L_n = V_{bn}/M_n^*$ and $D_n = u_{rn}/\Gamma_n\phi_{rn}$ in which ϕ_{rn} is the value of $\boldsymbol{\phi}_n$ at the reference point in the direction under consideration, $M_n^* = \left(\boldsymbol{\phi}_n^T\mathbf{M}\boldsymbol{\iota}\right)^2 / \boldsymbol{\phi}_n^T\mathbf{M}\boldsymbol{\phi}_n$ is the effective modal mass, and $\Gamma_n = \boldsymbol{\phi}_n^T\mathbf{M}\boldsymbol{\iota}/\boldsymbol{\phi}_n^T\mathbf{M}\boldsymbol{\phi}_n$ with $\boldsymbol{\iota}$ equal to the influence vector. The influence vector $\boldsymbol{\iota}$ is a vector of size equal to the total number of degrees of freedom. For analysis in the x-direction, the components of $\boldsymbol{\iota}$ corresponding to x-degree-of-freedom are equal to one and remaining components equal to zero. Similarly the

components of ι corresponding to y-degree-of-freedom are equal to one and remaining components equal to zero for analysis in the y-direction.

5. Idealize the force-displacement, $F_{sn}/L_n - D_n$, curve as a bilinear curve and compute the yield value F_{sny}/L_n.

6. Compute the yield strength reduction factor, $R = S_A/\left(F_{sny}/L_n\right)$.

7. Compute the peak deformation $D_n = \Delta_d$ of the nth-"mode" inelastic SDF system defined by the force-deformation relation developed in Step 4 and damping ratio ζ_n, from Equation (1). The elastic vibration period of the system is based on the effective slope of the $F_{sn}/L_n - D_n$ curve, which for a bilinear curve is given by $T_n = 2\pi\left(L_n D_{ny}/F_{sny}\right)^{1/2}$.

8. Calculate peak reference point displacement u_{rn} associated with the nth-"mode" inelastic SDF system from $u_{rn} = \Gamma_n \phi_{rn} D_n$.

9. Push the structure to the reference point displacement equal to u_{rn} and note the values of desired displacement δ_{no}.

10. Repeat Steps 3 to 9 for all significant modes identified.

11. Combine the peak modal displacement, δ_{no}, by an appropriate modal combination rule, e.g., CQC, to obtain the peak dynamic response, Δ_o.

DISPLACEMENT CAPACITY

It is proposed that the displacement capacity of piles in Marine Oil and LNG Terminals be estimated from

$$\Delta_c = \mu_\Delta \Delta_y \tag{8}$$

where Δ_y is the yield displacement of the pile and μ_Δ is the displacement ductility capacity of the pile. Following are the recommendations that have been developed for the yield displacement and displacement ductility of piles commonly used in Marine Oil and LNG Terminals. These recommendations have been developed to ensure that the material strains in the pile at its

displacement capacity remain within the limits specified in the MOTEMS (2007).

The procedure to estimate the displacement capacity is intended to be a simplified procedure for either initial design of piles or for checking results from more complex nonlinear finite element analysis. The recommendations presented in this report are limited to: (1) piles with long freestanding heights (length/diameter > 20) above the mud line; (2) piles with transverse volumetric ratio greater than 0.5%; and (3) piles in which the displacement demand has been estimated utilizing equivalent-fixity approximation. Results form this investigation should be used with caution for parameters or cases outside of those described above.

Piles with Full-Moment- or Pin-Connection to the Deck Slab

The recommended values of displacement ductility capacity of piles with full-moment-connection or pin-connection to the deck slab are

Design Earthquake Level	Hinge Location	Reinforced-Concrete Piles	Hollow-Steel Piles
Level 1	In-Ground	1.75	1.2
	Pile-Deck	1.75	1.2
Level 2	In-Ground	2.5	2.75
	Pile-Deck	5.0	2.75

The yield displacement of the pile may be estimated either from idealized pushover curve developed from the nonlinear static pushover analysis or may be estimated from

$$\Delta_y = \begin{cases} \dfrac{M_y L^2}{6EI_e} & \text{for full-moment-connection} \\ \dfrac{M_y L^2}{3EI_e} & \text{for pin-connection} \end{cases} \tag{9}$$

in which M_y is the section yield moment and EI_e is the effective value of EI that can be estimated from the section moment-curvature analysis. Note that M_y is not the section moment at first-yield but the effective yield moment estimated from bilinear idealization of the moment-curvature relationship.

Piles with Dowel-Connection to the Deck Slab

Simplified formulas are proposed for estimating displacement ductility capacity of piles with dowel-connection, such as hollow-steel piles or prestressed concrete piles connected to the deck slab with dowels. The following is a step-by-step summary of the procedure to implement these formulas to estimate displacement capacity of such piles:

1. Establish the axial load, P, on the pile.
2. Estimate the pile length based on equivalent-fixity assumption.
3. Select an appropriate design level – Level 1 or Level 2 – and establish various strain limits for the selected design level.
4. Develop the moment-rotation relationship of the dowel-connection using the procedure described in Chapter 8 of this report.
5. Determine rotational stiffness, k_θ, yield moment, $M_{y,C}$, and yield rotation, $\theta_{y,C}$ of the dowel-connection from the moment-rotation relationship developed in Step 4.
6. Establish the rotation of the dowel-connection, θ_L, and corresponding ductility, $\mu_\theta = \theta_L / \theta_{y,C}$, when strain in the outer-most dowel of the connection reaches the strain limit established in Step 3 for the selected design level.
7. Conduct the moment-curvature analysis of the pile section with appropriate axial load and idealize the moment-curvature relationship by a bi-linear curve.
8. Compute the effective, EI_e, and effective yield moment, $M_{y,P}$, from the pile moment-curvature relationship. Note that EI_e is equal to initial elastic slope and $M_{y,P}$ is the yield value of the moment of the idealized bi-linear moment-curvature relationship. For steel piles, EI_e may be computed from section properties and material modulus, and $M_{y,P}$ may be approximated as $M_{y,P} \simeq f_y \left(d_o^3 - d_i^3 \right) / 6$.
9. Estimate the yield curvature, $\phi_{y,P} = M_{y,P} / \mathrm{EI_e}$.

10. Establish the curvature of the steel pile, ϕ_L, and corresponding curvature ductility, $\mu_\phi = \phi_L / \phi_{y,P}$, when material strain in the pile section reaches the strain limit established in Step 3 for the selected design level.

11. Select the value of ρ which defines the length of the plastic hinge as a fraction of the "effective" length of the pile. The recommended value for hollow-steel piles with dowel-connection is $\rho = 0.03$ for Level 1 design and $\rho = 0.075$ for Level 2; and for prestressed concrete pile with dowel-connection for both design levels is $\rho = 0.05$.

12. Compute the dimensionless parameters: $\eta = M_{y,P} / M_{y,C}$, and $\beta = EI_e / k_\theta L$.

13. Compute the normalized value of the plastic hinge length: $L_P^* = (\rho\eta)/(1+\eta)$.

14. Compute the yield displacement which corresponds to first effective yielding in the connection as: $\Delta_{y,C} = \theta_{y,C} L(1+4\beta)/6\beta$

15. Compute the displacement ductility for yielding in the connection as $\mu_\Delta = (1+4\beta\mu_\theta)/(1+4\beta)$ if μ_θ computed in Step 6 is less than or equal to $(\eta-1)/2\beta$ otherwise $\mu_\Delta = (2-\eta+6\beta\mu_\theta)/(1+4\beta)$.

16. Compute displacement ductility for yielding in the pile as $\mu_\Delta = (2\eta-1)/(1+4\beta) + (6\eta L_p^*)(1-L_p^*/2)(\mu_\phi - 1)/(1+4\beta)$

17. Establish the displacement ductility capacity as lower of the values computed in Steps 15 and 16.

18. Compute the displacement capacity of the pile as product of the yield displacement computed in Step14 and the displacement ductility capacity computed in Step 17.

The recommended value of displacement ductility for piles with full-moment-connection or the simplified formulas for piles with dowel-connection have been shown to provide results that are "accurate" enough for most practical applications. However, it may be useful to verify these recommendations from experimental studies.

ACKNOWLEDGMENTS

This research investigation is supported by the California State Lands Commission (CSLC) under Contract No. C2005-051 for Development of LNGTEMS/MOTEMS Performance-Based Seismic Criteria. This support is gratefully acknowledged. The author would especially like to thank Martin Eskijian, CSLC Project Manager and Hosny Hakim of the CSLC for their continuing support. The author would also like to acknowledge advice from following individuals: Gayle Johnson and Bill Bruin of Halcrow; Bob Harn of Berger/ABAM Engineers Inc.; Dr. Omar Jeradat of Moffatt & Nichol; Peter Yin of Port of Los Angeles; Eduardo Miranda of Stanford University; and Dr. Hassan Sedarat and Tom Ballard of SC Solutions Inc. Finally, the author would like to acknowledge the editorial support provided by John Freckman of the CSLC. Additional support for this research is provided by a grant entitled "C3RP Building Relationships 2008/2010" from the Department of the Navy, Office of Naval Research under award No. N00014-08-1-1209. This support is also acknowledged.

1. INTRODUCTION

Seismic design of marine oil terminals in California is governed by 2007 Title 24 California Code of Regulations (CCR), Part 2, California Building Code, Chapter 31F (Marine Oil Terminals). These regulations are commonly known as the "Marine Oil Terminal Engineering and Maintenance Standard" (MOTEMS). The MOTEMS describe the acceptable methods of seismic analysis and provide the specific performance criteria for two levels of earthquake motions to be used in the seismic assessment. The return period of the design earthquake for each level depends on the risk level, which is a function of the oil susceptible to spillage at any given time. For example, Level 1 and Level 2 design earthquakes for high risk terminals correspond to return periods of 72 and 475 years, respectively. The performance goal for Level 1 earthquake is no or minor damage without interruption in service or with minor temporary interruption in service. The performance goal for Level 2 earthquake is controlled inelastic behavior with repairable damage resulting in temporary closure of service, restorable within months and the prevention of a major oil spill (24 CCR 3104F.2.1). This is the formal short form of the above cited regulation and specifies a particular section (24 CCR 3104F.2.1) of the California Code of Regulations. It will be used throughout this document along with the informal "MOTEMS" abbreviation.

The MOTEMS is currently being used for new construction in the Ports of Los Angeles and Long Beach. In addition, the MOTEMS has been referenced in the FEMA 450 document (BSSC, 2003). The MOTEMS has also become the approved methodology for the seismic assessment of US military wharf/pier facilities in high seismic areas (Department of Defense, 2005).

As with marine oil terminals, LNG receiving terminals are considered liquid hydrocarbons reception terminals by the California State Lands Commission (CSLC), and as such fall under the Lempert-Keen-Seastrand Oil Spill and Response Act of 1990. The "Act" states that the "commission (CSLC) shall adopt rules, regulations, guidelines..., performance standards... for all existing and proposed marine terminals within the state...". It is through this "Act" that the CSLC is developing standards for LNG terminals, and in this case mutually applicable to oil terminals. The effort described in this report is through funding obtained by the CSLC for development of standards for LNG terminals.

The MOTEMS seismic analysis requires that the seismic displacement demand in marine oil terminal structures be determined using nonlinear static procedures except for irregular structures (24 CCR 3104F.1.4) with high or moderate seismic risk classification (see 24 CCR 3104F2.3.2). A linear modal procedure is required for irregular structures with high or moderate seismic risk classification. The analysis method specified in the MOTEMS is based on the concept of equivalent linearization presented by Priestley et al. (1996). The seismic design (or acceptability criteria) involves making sure that the material strains at the seismic displacement demand not exceed certain prescribed values.

The primary objective of this investigation is to develop simplified analysis and design procedures for pile supported wharves and piers for Marine Oil and LNG Terminals. For this purpose, a simplified coefficient-based approach is proposed for estimating seismic displacement demand for regular structures. This approach is adopted from the performance-based analysis procedure recently approved for buildings in the ASCE/SEI 41-06 standard (ASCE, 2007). A modal pushover analysis (MPA) approach is proposed for irregular structures. The MPA procedure accounts for the higher-mode effects that are important in irregular structures (Chopra and Goel, 2004). The acceptability of piles in terms of displacement ductility limitation, instead of the material strain limitations, is proposed. For this purpose, simplified expressions for estimating displacement ductility capacity of piles are recommended. These expressions are calibrated such that the material strain limits in the MOTEMS would not be exceeded if the displacement ductility demand is kept below the proposed displacement ductility capacity. These simplified procedures can be used as an alternative to the procedures currently specified in the MOTEMS.

This report is organized as follows:

Chapter 2 presents the development of a simplified procedure for estimating seismic displacement demand in regular and irregular structures.

Chapter 3 describes the equivalent fixity model utilized in developing the simplified seismic acceptability criteria.

Chapter 4 summarizes the approach in the current MOTEMS for seismic evaluation of piles.

Chapter 5 presents the proposed approach to estimate the displacement ductility capacity of piles.

Chapters 6 and 7 present development and evaluation of simplified expressions for the displacement ductility capacity of reinforced concrete and hollow steel piles, respectively, with full-moment connection to the concrete deck.

Chapter 8 discusses the behavior of partial-moment connections of hollow steel and prestressed concrete piles to the concrete deck.

Chapter 9 presents the theoretical development of simplified expressions for estimating displacement ductility demands in piles with partial-moment connections.

Chapters 10 and 11 evaluate these simplified expressions for hollow steel and prestressed concrete piles, respectively.

2. ESTIMATION OF DISPLACEMENT DEMANDS

The MOTEMS requires that the seismic displacement demand in marine oil terminal structures be determined using nonlinear static procedures except for irregular structures with high or moderate seismic risk classification (MOTEMS, 2007: Section 3104F2.3.2). A linear modal procedure is required for irregular structures with high or moderate seismic risk classification.

2.1 REGULAR STRUCTURES

2.1.1 Current MOTEMS Procedure

The MOTEMS (2007) specifies that the displacement demand, Δ_d, be computed from

$$\Delta_d = S_A \frac{T^2}{4\pi^2} \tag{2.1}$$

in which S_A is the spectral response acceleration corresponding to the vibration period T of the structure, with a 5% damping ratio. The spectral acceleration, S_A, is computed from United States Geological Survey (USGS) maps, or from site-specific probabilistic seismic hazard analysis (PSHA).

Equation (2.1) assumes that the deformation demand in a system that is deformed beyond the linear elastic limit is equal to that in a linear-elastic system. It is well known that for short period systems, this equal displacement rule may not apply; deformation of a nonlinear system may be larger than that of a linear system. For such systems, nonlinearity in the force-deformation relationship must be considered to compute the displacement demand.

Therefore, the MOTEMS requires that a refined analysis be used to calculate the displacement demand if the vibration period of the structure T is less than period T_o which corresponds to the period at which constant acceleration and constant velocity regions of the design spectrum intersect. The refined analysis (MOTEMS, 2007: Section 3104F.2.3.2.5) utilizes the nonlinear force-deformation behavior of the structure developed from nonlinear static pushover analysis and is based on the concept of equivalent linearization presented by Priestley et al. (1996).

2.1.2 Procedures to Compute Response of Single-Degree-of-Freedom (SDF) Systems

An alternative approach to the equivalent linearization method is the coefficient method in which the deformation of the inelastic SDF system is computed by multiplying the displacement of the elastic SDF system by a coefficient as follows:

$$\Delta_i = C_R \Delta_e \tag{2.2}$$

where Δ_e is the deformation of the elastic SDF system, Δ_i is the deformation of the inelastic SDF system, and C_R is the coefficient that converts displacement of the elastic SDF system to displacement of the inelastic SDF system. Several alternative definitions of the coefficient C_R are available. The following is a brief review of the ones which are commonly used.

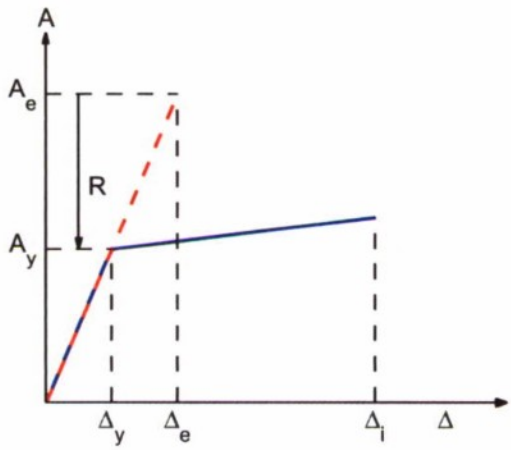


Figure 2.1. Force-deformation properties of an inelastic and elastic SDF system.

Figure 2.1 shows a nonlinear inelastic SDF system with a bilinear force-deformation relationship. The yield strength (or capacity) of the system is V_y and its yield displacement is Δ_y. The yield strength V_y is related to the pseudo-acceleration (or spectral acceleration) at yield level, A_y, and the seismic weight, W, as $V_y = A_y W$. If the system were to remain elastic, the design force would be $V_e = A_e W$ with A_e being the spectral acceleration of the elastic SDF system. The yield strength reduction factor, R, is defined as the ratio of the elastic level force and the yield strength of the inelastic SDF system. Thus R is given as

$$R = \frac{V_e}{V_y} = \frac{A_e}{A_y} \tag{2.3}$$

Note that the yield strength reduction factor, R, differs from the response modification coefficient, R, generally used in the force-based design procedures of building [for example see ASCE/SEI 7-05 (ASCE, 2005)]: the yield strength reduction factor in Equation (2.3) considers the effects of system ductility alone whereas the response modification factor in force-based design accounts for other factors, such as type and past performance of lateral load resisting systems, over strength, etc., in addition to the system ductility.

The coefficient C_R is generally defined as a function of vibration period, T, yield reduction factor, R, and a few other factors. Nassar and Krawinkler (1991) defined C_R as

$$C_R = \frac{1}{R}\left[1+\frac{1}{c}\left(R^c-1\right)\right] \tag{2.4}$$

where the constant c is defined by

$$c = \frac{T^a}{1+T^a}+\frac{b}{T} \tag{2.5}$$

with constants $a=1$ and $b=0.42$ for $\alpha=0\%$, $a=1$ and $b=0.37$ for $\alpha=2\%$, and $a=0.8$ and $b=0.29$ for $\alpha=10\%$. The parameter α is defined as the ratio of the post-yield stiffness and initial elastic stiffness expressed as a percentage value. Ruiz-Garcia and Miranda (2003) defined the C_R as

$$C_R = 1+\left[\frac{1}{a\left(T/T_s\right)^b}+\frac{1}{c}\right]\left(R-1\right) \tag{2.6}$$

with T_s being the site characteristic period selected as 0.75 for site B, 0.85 for site C, and 1.05 for site D; and constants $a=50$, $b=1.8$, and $c=55$. Chopra and Chintanapakdee (2004) defined C_R as

$$C_R = 1+\left[\left(L_R-1\right)^{-1}+\left(\frac{a}{R^b}+c\right)\left(\frac{T}{T_s}\right)^d\right]^{-1} \tag{2.8}$$

in which

$$L_R = \frac{1}{R}\left(1+\frac{R-1}{\alpha}\right) \tag{2.9}$$

T_s is the period corresponding to the transition between constant pseudo-acceleration and constant pseudo-velocity regions of the design spectrum, $a = 61$, $b = 2.4$, $c = 1.5$, and $d = 2.4$.

The C_R is also defined in several building design guidelines. For example, FEMA-356 (ASCE, 2000), defines C_R as

$$C_R = C_1 C_2 C_3 \tag{2.10}$$

where C_1 is the modification factor to relate a maximum displacement of nonlinear elastic-perfectly-plastic SDF system to displacement of elastic SDF system given by

$$C_1 = \begin{cases} 1.0; & \text{for } T \geq T_s \\ \dfrac{1}{R} + \left(1 - \dfrac{1}{R}\right)\left(\dfrac{T_s}{T}\right) \leq 1.5; & \text{for } T < T_s \end{cases} \tag{2.11}$$

C_2 is the modification factor to represent effects of pinched hysteretic shape, stiffness degradation, and strength deterioration (Table 2.1), and C_3 is the modification factor to represent P-Delta effects given by

$$C_3 = 1 + \frac{|\alpha|(R-1)^{3/2}}{T} \tag{2.12}$$

Table 2.1. Values of modification factor C_2 in FEMA-356 (ASCE 2000).

Structural Performance Level	$T \leq 0.1$ s		$T \geq T_s$	
	Framing Type 1	**Framing Type 2**	**Framing Type 1**	**Framing Type 2**
Immediate Occupancy	1.0	1.0	1.0	1.0
Life safety	1.3	1.0	1.1	1.0
Collapse Prevention	1.5	1.0	1.2	1.0

The coefficients in FEMA-356 were re-examined and an improved version was proposed in FEMA-440 (ATC, 2005). This proposal has also been adopted in ASCE/SEI 41-06 standard (ASCE, 2007). The C_R in FEMA-440 and ASCE/SEI 41-06 is defined as

$$C_R = C_1 C_2 \tag{2.13}$$

where C_1 is the modification factor to relate maximum displacement of a nonlinear elastic-perfectly-plastic SDF system to displacement of an elastic SDF system defined by

$$C_1 = \begin{cases} 1.0; & T > 1.0\text{s} \\ 1.0 + \dfrac{R-1}{aT^2}; & 0.2\text{s} < T \le 1.0\text{s} \\ 1.0 + \dfrac{R-1}{0.04a}; & T \le 0.2\text{s} \end{cases} \tag{2.14}$$

with a = 130 for Site Class B, 90 for Site Class C, and 60 for Site Class D. and C_2 is the modification factor to represent effects of cyclic degradation in stiffness and strength given by

$$C_2 = \begin{cases} 1.0; & T > 0.7\text{s} \\ 1 + \dfrac{1}{800}\left(\dfrac{R-1}{T}\right)^2; & T \le 0.7\text{s} \end{cases} \tag{2.15}$$

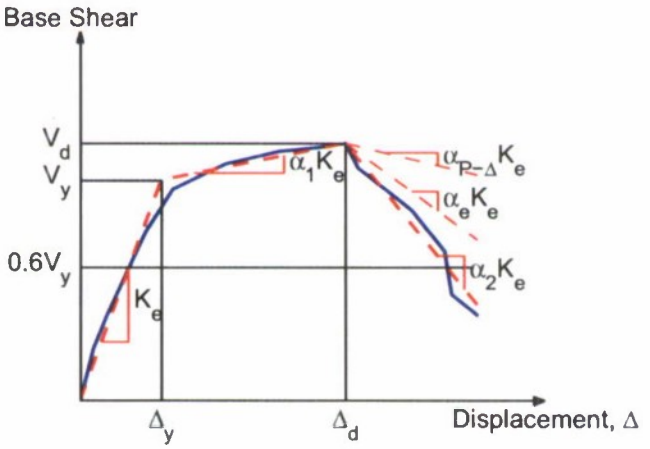


Figure 2.2. Idealized force-deformation curve for nonlinear static analysis.

The provisions of FEMA-440 and ASCE/SEI 41-06 can be used provided that R does not exceed the limiting value given by

$$R_{\max} = \frac{\Delta_d}{\Delta_y} + \frac{1}{4}\left|\alpha_e\right|^{-h}; \quad h = 1.0 + 0.15\ln(T) \tag{2.16}$$

in which Δ_d is the deformation corresponding to peak strength, Δ_y is the yield deformation, and α_e is the effective negative post-yield slope given by

$$\alpha_e = \alpha_{P-\Delta} + \lambda(\alpha_2 - \alpha_{P-\Delta}) \tag{2.17}$$

where α_2 is the negative post-yield slope ratio defined in Figure 2.2, $\alpha_{P-\Delta}$ is the negative slope

ratio caused by $P-\Delta$ effects, and λ is the near-field effect factor given as 0.8 for $S_1 \geq 0.6$ and 0.2 for $S_1 < 0.6$ (S_1 is defined as the 1-second spectral acceleration for the Maximum Considered Earthquake). The α_2 slope includes $P-\Delta$ effects, in-cycle degradation, and cyclic degradation.

Finally, FEMA-450 (BSSC, 2003) defines C_R as

$$C_R = \begin{cases} 1.0; \text{ for } T > T_s \\ \frac{1}{R}\left[1 + \frac{(R-1)T_s}{T}\right]; \text{ for } T \leq T_s \end{cases} \tag{2.18}$$

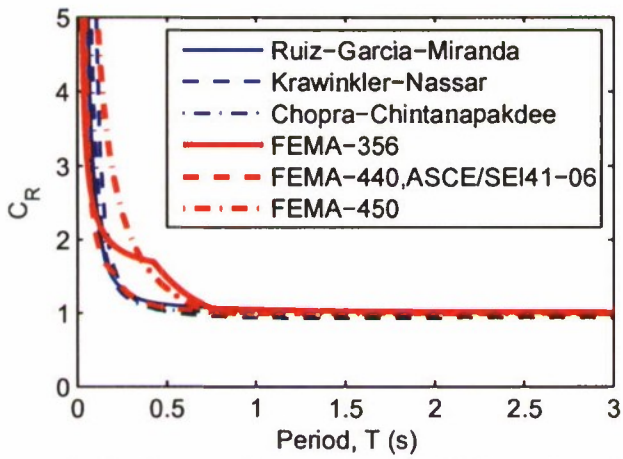


Figure 2.3. Comparison of C_R from various recommendations. Results are for $R = 3$ and site class B.

Figure 2.3 compares the C_R values from the aforementioned recommendations. This comparison indicates that all recommendations lead to essentially identical values of C_R for periods longer than 1 sec. The values due to FEMA-356 and FEMA-450 differ significantly from those due to the remaining recommendations primarily for periods shorter than 0.5 sec.

2.1.3 Proposed Alternate Displacement Demand Procedure for Regular Structures

Presented here is an alternative procedure for estimating the seismic displacement demand of regular structures with period T less than period T_o that can be idealized as a single-degree-of-freedom (SDF) system. This procedure utilizes the nonlinear force-deformation behavior of the structure developed from nonlinear static pushover analysis. The computation of the

displacement demand is adopted from the procedure recommended in the FEMA-440 document (ATC, 2005) and ASCE/SEI 41-06 standard (ASCE 2007). Although this procedure has been described previously in Section 2.1.2, it is re-organized here to bc compatible with the current MOTEMS procedure. The proposed altcrnative procedure involves estimating the displacement demand in a nonlinear SDF system from

$$\Delta_d = C_1 C_2 S_A \frac{T^2}{4\pi^2} \tag{2.19}$$

in which C_1 and C_2 are the coefficients that convert displacement demand of a linear-elastic SDF system to displacement demand of nonlinear SDF system.

The coefficient C_1 is given by

$$C_1 = \begin{cases} 1.0; & T > 1.0\,\text{s} \\ 1.0 + \dfrac{R-1}{aT^2}; & 0.2\,\text{s} < T \le 1.0\,\text{s} \\ 1.0 + \dfrac{R-1}{0.04a}; & T \le 0.2\,\text{s} \end{cases} \tag{2.20}$$

in which a is a site dependent constant equal to 130 for Site Class A and B, 90 for Site Class C, and 60 for Site Class D, E, and F; and R is the ratio of the elastic to the yield strength of the system and is defined as

$$R = \frac{S_A}{g}\frac{W}{V_y} \tag{2.21}$$

in which S_A is the spcctral acccleration used in Equation (2.1), W is thc seismic weight of the system, V_y is the yield force (or base shear) of the system, and g is the acceleration due to gravity. The coefficient C_2 is given by

$$C_2 = \begin{cases} 1.0; & T > 0.7\,\text{s} \\ 1 + \dfrac{1}{800}\left(\dfrac{R-1}{T}\right)^2; & T \le 0.7\,\text{s} \end{cases} \tag{2.22}$$

Equation (2.19) can be used to compute the displacement demand for systems in which $R \le R_{max}$ where R_{max} is given by

$$R_{\max} = \frac{\bar{\Delta}_d}{\Delta_y} + \frac{|\alpha_e|^{-t}}{4} \tag{2.23}$$

in which $\bar{\Delta}_d$ is smaller than the computed displacement demand, Δ_d, from Equation (2.19) or the displacement corresponding to the maximum strength in the pushover curve, Δ_y is the yield displacement of the idealized bilinear force-deformation curve, t is a constant computed from

$$t = 1 + 0.15 \ln(T) \tag{2.24}$$

and α_e is the effective post-elastic stiffness ratio computed from

$$\alpha_e = \alpha_{P-\Delta} + \lambda(\alpha_2 - \alpha_{P-\Delta}) \tag{2.25}$$

where λ is a near-field effect factor equal to 0.8 for sites that are subjected to near-field effects and 0.2 for sites that are not subjected to near field effects. The near field effects may be considered to exist if the 1 second spectral value, S_1, at the site for the maximum considered earthquake is equal to or exceeds 0.6g. The P-Delta stiffness ratio, $\alpha_{P-\Delta}$, and the maximum negative post-elastic stiffness ratio, α_2, in Equation (2.25) are estimated from the idealized force-deformation curve in Figure 2.2. The $\alpha_{P-\Delta}$ needed in Equation (2.25) may be estimated by conducting pushover analysis with and without P-Delta effects.

2.2 IRREGULAR STRUCTURES

2.2.1 Current MOTEMS Procedure

The current MOTEMS procedure requires that the seismic displacement demand in irregular concrete or steel structures with high or moderate seismic risk classification be computed from linear modal analysis. This procedure assumes that the displacement demand in irregular structures deformed beyond the linear elastic range may be approximated by that of a linear elastic structure. For irregular concrete and steel structures with low seismic risk, the displacement demand must be computed by a nonlinear static procedure; the nonlinear static procedure for such irregular structures appears to be similar to that for regular structures.

2.2.2 Proposed Nonlinear Static Procedure for Irregular Structures

Presented here is a rational nonlinear static procedure for estimating displacement demand in irregular structures. Proposed initially by Chopra and Goel (2004) to estimate seismic demands in unsymmetric-plan buildings, this procedure has been slightly modified to estimate displacement demands in irregular Marine Oil and LNG Terminals. The following is a step-by-step summary of this procedure.

1. Compute the natural frequencies, ω_n and modes, $\boldsymbol{\phi}_n$, for linearly elastic vibration of the irregular Marine Oil and LNG Terminal.
2. Select a reference point where the displacement, u_{rn}, is to be monitored in the selected direction of analysis during the pushover analysis. Ideally, this reference point should be the location on the structure with largest value of ϕ_{rn} in the selected direction of analysis.
3. For the nth-mode, develop the pushover curve, $V_{bn} - u_{rn}$, for the nth modal force distribution, $\mathbf{s}_n^* = \mathbf{M}\boldsymbol{\phi}_n$, where $\mathbf{s}_n^*$ is the vector of lateral forces used during the pushover analysis, $\mathbf{M}$ is the mass matrix of the structure, and $\boldsymbol{\phi}_n$ is the nth mode shape. The base shear V_{bn} should be monitored in the same direction as the direction of selected reference point displacement u_{rn}.
4. Convert the $V_{bn} - u_{rn}$ pushover curve to the force-displacement, $F_{sn}/L_n - D_n$, relation for the nth -"mode" inelastic SDF system by utilizing $F_{sn}/L_n = V_{bn}/M_n^*$ and $D_n = u_{rn}/\Gamma_n\phi_{rn}$ in which ϕ_{rn} is the value of $\boldsymbol{\phi}_n$ at the reference point in the direction under consideration, $M_n^* = \left(\boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\iota}\right)^2 / \boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\phi}_n$ is the effective modal mass, and $\Gamma_n = \boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\iota} / \boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\phi}_n$ with $\boldsymbol{\iota}$ equal to the influence vector. The influence vector $\boldsymbol{\iota}$ is a vector of size equal to the total number of degrees of freedom. For analysis in the x-direction, the components of $\boldsymbol{\iota}$ corresponding to x-degree-of-freedom are equal to one and the remaining components equal to zero. Similarly the components of $\boldsymbol{\iota}$ corresponding to y-degree-of-freedom are equal to one and the remaining components equal to zero for analysis in the y-direction.
5. Idealize the force-displacement, $F_{sn}/L_n - D_n$, curve as a bilinear curve and compute the yield value F_{sny}/L_n.

6. Compute the yield strength reduction factor, $R = S_A / \left(F_{sny} / L_n \right)$.

7. Compute the peak deformation $D_n = \Delta_d$ of the nth-"mode" inelastic SDF system defined by the force-deformation relation developed in Step 4 and damping ratio ζ_n, from Equation (2.19). The elastic vibration period of the system is based on the effective slope of the $F_{sn} / L_n - D_n$ curve, which for a bilinear curve is given by $T_n = 2\pi \left(L_n D_{ny} / F_{sny} \right)^{1/2}$.

8. Calculate peak reference point displacement u_{rn} associated with the nth-"mode" inelastic SDF system from $u_{rn} = \Gamma_n \phi_{rn} D_n$.

9. Push the structure to reference point displacement equal to u_{rn} and note the values of desired displacement δ_{no}.

10. Repeat Steps 3 to 9 for all significant modes identified.

11. Combine the peak modal displacement, δ_{no}, by an appropriate modal combination rule, e.g., CQC, to obtain the peak dynamic response, Δ_o.

3. SIMPLIFYING ASSUMPTION

Figure 3.1b shows the mathematical model of a free-head pile of Figure 3.1a supported on bedrock (or other competent soil) and surrounded by soil between the bedrock and mud line. In this model, the pile is represented by beam-column elements and soil by Winkler reaction springs connected to the pile between the bedrock and the mud line (Priestley et al., 1996). The properties of the beam-column element are established based on the pile cross section whereas properties of the reaction springs are specified based on geotechnical data (e.g., see Priestley et al., 1996; Dowrick, 1987). Figure 3.1c shows the height-wise distribution of bending moment under lateral load applied to the pile tip. Note that the maximum bending moment occurs slightly below the mud line at a depth equal to D_m, typically denoted as the depth-to-maximum-moment below the mud line (Figure 3.1c). Lateral displacement at the pile tip can be calculated based on this bending moment distribution or from a discrete element model implemented in most commonly available computer programs for structural analysis.

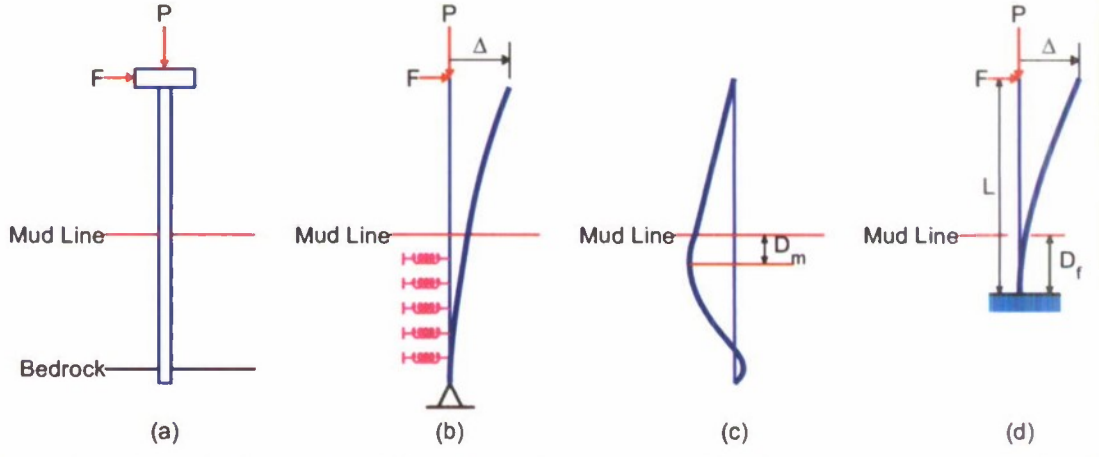


Figure 3.1. Simplified model of the pile-soil system for displacement capacity evaluation: (a) Pile supported on bedrock; (b) Mathematical model of the pile; (c) Height-wise variation of bending moment; and (d) Equivalent fixity model for displacement calculation.

An alternative approach to modeling soil flexibility effects of the pile with discrete soil springs is the effective fixity approach (Priestley et al., 1996: Sec. 4.4.2; Dowrick, 1987: Sec. 6.4.5.3). In this approach (Figure 3.1d), the depth-to-fixity, D_f, is defined as the depth that produces in a fixed-base column with soil removed above the fixed base the same top-of-the-pile lateral displacement under the lateral load, F, as that in the actual pile-soil system (Priestley et al., 1996). Both the axial load, P, and top-of-the-pile moment, M (not shown in Figure 3.1d)

need to be considered. The depth-to-fixity, which depends on the pile diameter and soil properties, is typically provided by the geotechnical engineer, estimated from charts available in standard textbooks on the subject (e.g., Priestley, et al., 1996; Dowrick, 1987) or from recommendations in several recent references (e.g., Chai, 2002; Chai and Hutchinson, 2002).

The equivalent fixity model is typically used for estimating displacement of piles that remain within the linear-elastic range. For piles that are expected to be deformed beyond the linear-elastic range, however, nonlinear analysis of the discrete soil spring model approach of Figure 3.1b is recommended (Priestley et al., 1996: Sec. 4.4.2) because the plastic hinge forms at the location of the maximum bending moment, i.e., at the depth-to-maximum-moment, D_m, and not at the depth-to-fixity, D_f. A recent investigation has developed equations for estimating lateral displacement of equivalent fixity model of the nonlinear soil-pile system by recognizing that the plastic hinge forms at the depth-to-maximum-moment (Chai, 2002); expressions for estimating displacement ductility capacity of pile-soil system are also available (Priestley et al., 1996: Sec. 5.3.1). However, calculation of lateral displacement capacity of nonlinear soil-pile systems using these approaches requires significant information about the soil properties.

This investigation uses a simplifying assumption that the equivalent fixity model may directly be used to estimate lateral displacement capacity of nonlinear piles. Clearly, such an approach indicates that the plastic hinge would form at the depth-to-fixity, D_f, which differs from the actual location at the depth-to-maximum-moment, D_m. It is useful to note that D_f is typically in the range of 3 to 5 pile diameter whereas D_m is in the range of 1 to 2 pile diameter (see Priestley et al., 1996). Obviously, plastic hinge at D_f in the equivalent fixity model would provide slightly larger plastic displacement compared to the plastic displacement if the plastic hinge was correctly located at D_m; note that plastic displacement is given by $\Delta_p = \theta_p \left(L_a + D_f \text{ or } D_m\right)$ where θ_p is the plastic hinge rotation and L_a is the free-standing height of the pile. However, the simplifying assumption used in this investigation is appropriate because difference between D_f and D_m is unlikely to significantly affect the plastic displacement for piles with very long free-standing height used in marine oil terminals. Note that the freestanding height of piles in marine oil terminals is typically in excess of twenty times the pile diameter.

It is useful to emphasize that the simplified approach proposed in this investigation is intended to be used for preliminary design of piles or as a check on the results from the detailed nonlinear analysis. It is expected that this approach would provide results that are sufficiently "accurate" for this purpose.

The recommendations to estimate displacement capacity of the pile using the equivalent fixity approach are strictly valid only if the displacement demand is also estimated by utilizing the equivalent fixity pile model – a practice that is commonly used for analysis of large piers and wharves with many piles. The recommendations developed in this report should be used with caution if the displacement demand is estimated from a model consisting of piles with soil springs.

4. MOTEMS PROCEDURE FOR CAPACITY EVALUATION OF PILES

The displacement capacity of piles in the MOTEMS is estimated from nonlinear static pushover analysis. In this analysis, a force of increasing magnitude is applied statically in the transverse direction (perpendicular to the pile) permitting the materials in the pile – steel and/or concrete – to deform beyond their linear-elastic range. The displacement capacity is defined as the maximum displacement that can occur at the tip of the pile without material strains exceeding the permissible values corresponding to the desired design level.

The displacement capacity of a pile at a selected design level in the MOTEMS is obtained from the procedure proposed by Priestley et al. (1996) as illustrated in Figure 4.1. This procedure requires development of the pile section moment-curvature relationship. The moment-curvature relationship may be developed from any standard moment-curvature analysis programs using material constitutive relationships specified in the MOTEMS; the MOTEMS specifies guidelines for selecting material properties such as concrete and steel strengths as well as stress-strain curves for unconfined concrete, confined concrete, reinforcing steel, and prestressing steel. The moment-curvature relationship is idealized as a bilinear relationship as shown in Figure 4.2.

It is useful to note that the formulation presented here is for a cantilever, i.e., a pile with a pin-connection to the deck. Similar formulation is available for piles with full-moment-connection to the deck that uses "effective" length defined as the length between points of contra-flexure (e.g., see CALTRANS, 2006).

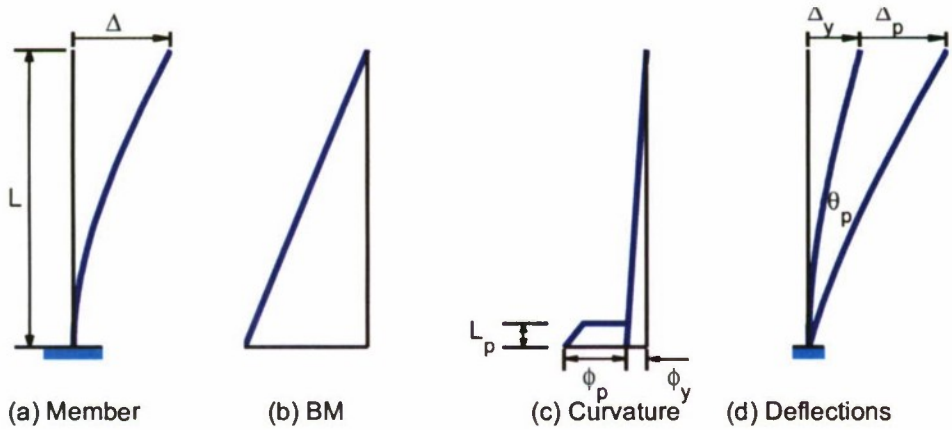


Figure 4.1. Deformation capacity of a pile: (a) Deflected shape, (b) Bending moment (BM) diagram, (c) Curvature distribution, and (d) Yield and plastic displacements.

The total displacement capacity of the pile is computed as

$$\Delta = \Delta_y + \Delta_P \tag{4.1}$$

in which Δ_y is the yield displacement and Δ_p is the plastic displacement of the pile. The yield displacement can be estimated as

$$\Delta_y = \frac{\phi_y L^2}{3} \tag{4.2}$$

where ϕ_y is the yield curvature computed from

$$\phi_y = \frac{M_y}{E_c I_e} \tag{4.3}$$

with M_y being the yield moment and $E_c I_e$ being the slope of the initial elastic portion of the bilinear idealization of the moment-curvature relationship, and L is the pile "effective" length. The "effective" length, defined as the length between points of contraflexure, for a cantilever becomes equal to its total length (Figure 4.1).

It is useful to note that the yield displacement, Δ_y, of reinforced-concrete pile may be estimated from Equation (4.2), without the need for section moment-curvature analysis, by using the following expression for dimensionless yield curvature (Priestley et al., 1996: Sec. 7.4.6):

$$\phi_y D = 2.45\varepsilon_y \pm 0.15 \tag{4.4}$$

in which D is the pile diameter and ε_y is the longitudinal yield reinforcement. Similar expression for hollow-steel pile is currently not available.

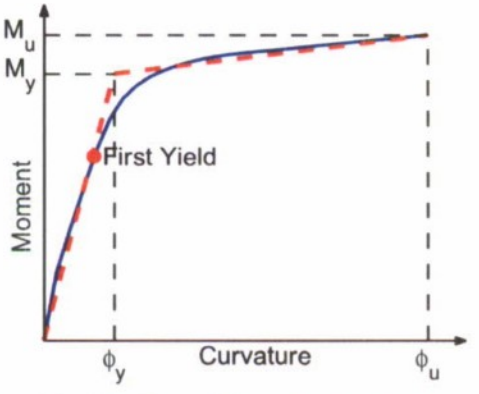


Figure 4.2. Bilinear idealization of the moment-curvature relationship.

The plastic displacement, Δ_p, is computed from

$$\Delta_p = \left(\frac{M_u}{M_y} - 1\right)\Delta_y + L_p\left(\phi_u - \phi_y\right)\left(L - 0.5L_p\right) \tag{4.5}$$

The plastic displacement given by Equation (4.5) includes components due to the elastic displacement resulting from the increase in moment from M_y to M_u, i.e., post-yield stiffness of the moment-curvature relationship (see Figure 4.2) and due to plastic rotation θ_p of the pile. In order to compute the plastic rotation, it is assumed that a constant plastic curvature, $\phi_p = \phi_u - \phi_y$, occurs over a plastic hinge length L_p of the pile (see Figure 4.1c). Therefore, the plastic rotation is given by

$$\theta_p = L_p\phi_p = L_p\left(\phi_u - \phi_y\right) \tag{4.6}$$

The values of M_u and ϕ_u in equation (4.5) are the largest values of the pile section moment and curvature, respectively, without exceeded the material strains at selected design level.

The MOTEMS specify the formula for estimating the plastic hinge length required in Equation (4.5). If the hinge were to form against a supporting member, i.e., at the pile-deck interface, the plastic hinge length is computed from

$$L_p = \begin{cases} 0.08L + 0.022f_{ye}d_{bl} \geq 0.044f_{ye}d_{bl} & (f_{ye} \text{ in MPA}) \\ 0.08L + 0.15f_{ye}d_{bl} \geq 0.3f_{ye}d_{bl} & (f_{ye} \text{ in ksi}) \end{cases} \tag{4.7}$$

in which f_{ye} is the expected yield strength of the reinforcing steel, and d_{bl} is the diameter of the longitudinal reinforcement. If the plastic hinge forms in-ground, the MOTEMS provide a chart to estimate the plastic hinge length that depends on the pile diameter, subgrade modulus, effective stiffness of the pile, and the distance from ground to the pile point of contraflexure. It is useful to note that Equation (4.7), as specified in Priestley et al., (1996) or in the MOTEMS (2006), does not explicitly impose an upper limit even though there may be some experimental evidence that the plastic hinge length should not be greater than the pile diameter.

The plastic hinge length formula of Equation (4.7) specified in the MOTEMS is based on the recommendation by Priestley et al. (1996) for reinforced concrete sections. The MOTEMS do not provide recommendations for plastic hinge length for steel piles or prestressed concrete piles.

The MOTEMS specify material strain limits for two levels of seismic design: Level 1 and Level 2. These strain limits depend on whether the plastic hinge forms in-ground or at the pile-deck interface. These strain limits are listed in Table 4.1.

Table 4.1. Material strain limits in the MOTEMS.

Pile Type	Material	Hinge Location	Level 1	Level 2
Reinforced-Concrete Pile	Concrete	Pile-Deck	$\varepsilon_c \le 0.004$	$\varepsilon_c \le 0.025$
		In-Ground	$\varepsilon_c \le 0.004$	$\varepsilon_c \le 0.008$
	Steel rebar	Pile-Deck	$\varepsilon_s \le 0.01$	$\varepsilon_s \le 0.05$
		In-Ground	$\varepsilon_s \le 0.01$	$\varepsilon_s \le 0.025$
Prestressed Concrete Pile with Dowel-Connection	Strands	In-Ground	$\varepsilon_p \le 0.005$ (Incremental)	$\varepsilon_p \le 0.025$ (Total)
		Pile-Deck	$\varepsilon_s \le 0.01$	$\varepsilon_s \le 0.05$
Hollow Steel Pile	Steel		$\varepsilon_s \le 0.008$	$\varepsilon_s \le 0.025$
Hollow Steel Pile with Dowel-Connection	Steel	In-Ground	$\varepsilon_s \le 0.008$	$\varepsilon_s \le 0.025$
		Pile-Deck	$\varepsilon_s \le 0.01$	$\varepsilon_s \le 0.05$

5. SIMPLIFIED PROCEDURE TO COMPUTE PILE DISPLACEMENT CAPACITY

Estimation of displacement capacity of the pile according to the seismic provisions of the MOTEMS require monitoring of material strains during the nonlinear static pushover analysis of the pile if the pile is modeled using a distributed-plasticity approach in which section properties are specified by a fiber-section. The estimation of the displacement capacity requires monitoring of material strains during moment-curvature analysis if the pile is modeled using a concentrated-plasticity approach in which nonlinearity is represented by rotational springs at two ends of the pile. The nonlinear moment-rotation relationship of this spring is computed from the moment-curvature relationship and estimated length of the plastic hinge. In either approach, the displacement capacity is defined as the maximum displacement that can occur at the tip of the pile without material strains exceeding the strain limits specified in the MOTEMS for any selected design level.

Monitoring strains during pushover analysis of piles using a distributed-plasticity model is cumbersome. Moreover, structural analysis programs commonly used by practicing engineers may not have the capability to directly monitor strains during the pushover analysis. Although, the concentrated-plasticity model, such as that employed in the current MOTEMS (see description in Chapter 4), does not require direct monitoring of material strains during pushover analysis; however, it still requires monitoring of material strains during moment-curvature analysis. Most commercially available programs for moment-curvature analysis do provide the capability to monitor material strains. However, this approach requires estimation of plastic hinge length in order to convert the moment-curvature relationship to the moment-rotation relationship of the rotational spring. While the MOTEMS provide guidelines for estimating plastic hinge length for reinforced concrete piles [see Equation (4.7)], such guidelines are not available for steel piles or prestressed concrete piles indicating that estimation of displacement capacity of such piles using concentrated-plasticity model may also be cumbersome. Therefore, there is a need to develop a simplified approach that avoids the need to monitor strains to the extent possible and yet provides a "good" estimate of displacement capacity of the pile without exceeding material strain limits specified in the MOTEMS.

It is useful to note that formulas and/or charts for estimating the plastic hinge length have been recommended by several researchers (e.g., Priestley et al., 1996; Chai, 2002; Chai and

Hutchinson, 2002; Budek et al., 2000; Song et al., 2004). However, these recommendations were developed for piles deformed significantly into the inelastic range. While such recommendations are appropriate for seismic design of piles for Level 2, where piles are expected to be deformed significantly into the inelastic range, they may not be appropriate for design of piles for Level 1 which corresponds to much lower level of inelastic action.

A simplified procedure is proposed in this report to compute the displacement capacity of piles commonly used in Marine Oil and LNG Terminals. This approach computes the displacement capacity as

$$\Delta_c = \mu_\Delta \Delta_y \tag{5.1}$$

where Δ_y is the yield displacement of the pile and μ_Δ is the displacement ductility capacity of the pile. The displacement ductility capacity is selected such that the material strains remain within the limits specified in the MOTEMS.

The guidelines to select the displacement ductility capacity and estimate the yield displacement are developed next in this report for four types of piles: (1) reinforced-concrete piles with either pin or full-moment connection to the deck; (2) hollow steel pipe pile with either pin or full-moment connection to the deck; (3) hollow steel pipe pile with a dowel- connection to the deck; and (4) prestressed concrete pile with a dowel-connection to the deck. The guidelines developed for these piles utilize the concept of equivalent fixity model described in Chapter 3 of this report.

Similar displacement ductility capacity based approaches have been proposed previously (e.g., Priestly et al., 1996; Budek at al., 2000; Chai, 2002; Song et al., 2004). However, this investigation specifically developed recommendations for displacement ductility capacity of long piles typically used in Marine Oil and LNG Terminals.

The simplified procedure to estimate displacement capacity of piles presented in this report is intended either for preliminary design of piles or as a quick check on the capacity that may be obtained from detailed nonlinear analyses. The design engineers may still use the elaborate analysis for final design of piles for a Marine Oil and LNG Terminals.

6. DISPLACEMENT CAPACITY OF REINFORCED CONCRETE PILES

This Chapter presents development of a simplified procedure for estimating displacement capacity of reinforced concrete piles connected to the deck either by a pin connection or by a moment connection. For this purpose, the current approach in the MOTEMS (see Equations 4.1 to 4.7 in Chapter 4) is further simplified. Presented first in this Chapter is development of simplified equations to compute displacement ductility of reinforced concrete piles that are independent of the pile length and depend only on the pile section curvature ductility and seismic design level. The accuracy of these equations is next evaluated against results from nonlinear finite element analyses. Subsequently, results of a parametric study are presented to understand the sensitivity of the displacement ductility capacity on pile diameter, longitudinal reinforcement ratio, transverse reinforcement ratio, and axial force. Based on these results, lower bound estimates of the ductility capacity of reinforced concrete piles for two design levels – Level 1 and Level 2 – are proposed. Finally, it is demonstrated that the lower-bound displacement ductility values along with simplified expressions for yield displacement provide very good estimate of the displacement capacity of piles when compared against values from nonlinear finite element analysis.

6.1 THEORETICAL BACKGROUND

The displacement ductility capacity of reinforced concrete piles is defined as (Priestley et al., 1996)

$$\mu_\Delta = \frac{\Delta_y + \Delta_p}{\Delta_y} = 1 + \frac{\Delta_p}{\Delta_y} = \frac{M_u}{M_y} + 3\left(\mu_\phi - 1\right)\left(\frac{L_p}{L}\right)\left(1 - 0.5\frac{L_p}{L}\right) \tag{6.1}$$

in which μ_ϕ is the pile section curvature ductility capacity given by

$$\mu_\phi = \frac{\phi_u}{\phi_y} \tag{6.2}$$

with ϕ_u being the section curvature at a selected design level material strain and ϕ_y is the yield curvature defined by Equation (4.3) and Figure 4.2. For piles in Marine Oil and LNG Terminals, which typically use piles with very long free-standing height, the second term in Equation (4.7) for plastic hinge length becomes negligibly small compared to the first term implying that this

term may be ignored without introducing significant error. Therefore, the plastic hinge length may be expressed as

$$L_p \simeq 0.08L \tag{6.3}$$

Equation (6.3) implies that the ratio L_p/L needed in Equation (6.1) is independent of the member "effective" length. It is useful to note that neglecting the second term in the Equation (4.6) for the plastic hinge location leads to smaller plastic hinge length and therefore is likely to provide a conservative estimate of the displacement ductility capacity of the pile.

The preceding approximation permits the following important simplification in Equation (6.1):

$$\mu_\Delta = \frac{M_u}{M_y} + 3\left(\mu_\phi - 1\right)(0.08)(1 - 0.5 \times 0.08) = \frac{M_u}{M_y} + 0.2304\left(\mu_\phi - 1\right) \tag{6.4}$$

which implies that the pile displacement ductility capacity is independent of its "effective" length; it depends only on the section curvature ductility, μ_ϕ, and ratio of ultimate and yield moments M_u/M_y. For moment-curvature relationship that exhibit very little post-yield stiffness, i.e., $M_u \simeq M_y$, Equation (6.4) can be further simplified as

$$\mu_\Delta = 1 + 0.2304\left(\mu_\phi - 1\right) = 0.7696 + 0.2304\mu_\phi \tag{6.5}$$

Equation (6.5) indicates that member displacement ductility capacity can be computed directly from the section curvature ductility capacity.

6.2 EVALUATION OF SIMPLIFIED EQUATIONS FOR DUCTILITY CAPACITY

The accuracy of Equations (6.5) in estimating displacement ductility capacity of reinforced concrete piles at seismic design Level 2 and Level 1, respectively, is evaluated in this section. For this purpose, displacement ductility capacity of reinforced concrete piles is evaluated from nonlinear static pushover analysis of a finite element model. The pile is considered to be fixed at top and bottom. These boundary conditions correspond to a pile that is connected to the pile-cap with a full-moment connection, and utilizes the equivalent displacement fixity assumption at the bottom. The axial load on the pile is assumed to be $0.05A_g f_c^{'}$ in which A_g is the gross cross-

section area of the pile and $f_c^{'}$ is the compressive strength of concrete. The longitudinal and transverse reinforcements in the pile section are assumed to be equal to 1% and 0.6%, respectively.

The pile is modeled with a nonlinear beam-column element in computer program Open System for Earthquake Engineering Simulation (*OpenSees*) (McKenna and Fenves, 2001). The distributed plasticity is considered by specifying the section properties by a fiber section model and then using seven integration points along the element length; details of such modeling may be found in McKenna and Fenves (2001). The material properties are specified as per the MOTEMS specifications (MOTEMS, 2007; Mander et al., 1988).

Strains in the concrete and steel are monitored during the pushover analysis. The limiting values of compressive strain in concrete and tensile strain in reinforcing steel are 0.004 and 0.01, respectively, for Level 1 and 0.025 and 0.05, respectively, for Level 2. If the hinge forms below ground, the limiting value of compressive strain in concrete and tensile strain in reinforcing steel are 0.004 and 0.01, respectively, for Level 1 and 0.008 and 0.025, respectively, for Level 2. The concrete strains are assumed to be specified just inside the reinforcement cage. The displacement ductility at a selected design level corresponds to the largest displacement that can occur at the tip of the pile without strain limits either in concrete or steel being exceeded.

The results are presented in Figure 6.1 for four pile diameters – 61 cm, 76 cm, 91 cm, and 107 cm – and pile length in the range of 5 m to 40 m. These results confirm expectations from Equation (6.5) that the displacement ductility capacity is independent of the pile length. This becomes apparent from essentially no variation in the ductility capacity from the nonlinear finite element analysis of the pile lengths in Figure 6.1 for both design levels and all pile diameters. The presented results also demonstrate that Equation (6.5) provides a very good estimate of the displacement ductility capacity of reinforced concrete piles (see Figure 6.1).

It is useful to note that the plastic hinge length used in this investigation does not include contribution to the plastic hinge length due to strain-penetration effects. It would be useful to verify these findings from experiments on reinforced concrete piles.

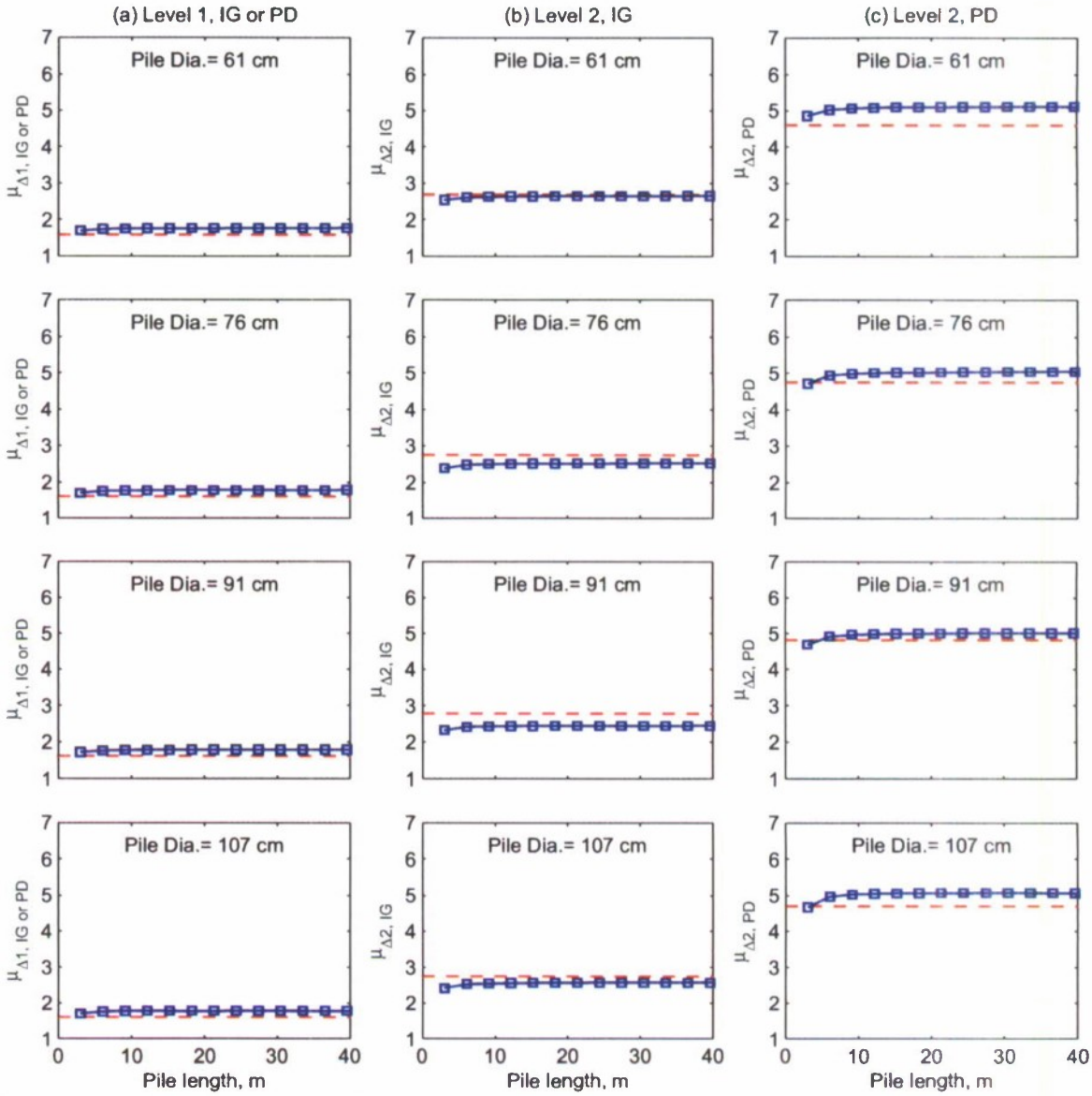


Figure 6.1. Displacement ductility capacity from simplified equation (shown in dashed line) and nonlinear finite element analysis (NLFEA) for seismic design (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, and (c) Level 2 for PD hinge formation.

6.3 SENSITIVITY OF DISPLACEMENT DUCTILITY TO PILE PARAMETERS

6.3.1 Pile Length and Pile Diameter

Figure 6.2 presents variation of displacement ductility capacity with pile length for four values of pile diameters: 61 cm, 76 cm, 91 cm, and 107 cm. The results are presented for piles with 1% longitudinal reinforcement and 0.6% transverse reinforcement. As noted previously, results of

Figure 6.2 also indicate that the displacement ductility capacity of piles is essentially independent of the pile length. This is expected because Equation (6.5) becomes independent of the pile length. The results of Figure 6.2 indicate that the displacement ductility capacity of the pile is also essentially independent of the pile diameter as apparent from almost identical curves for the four pile diameters considered in Figure 6.2.

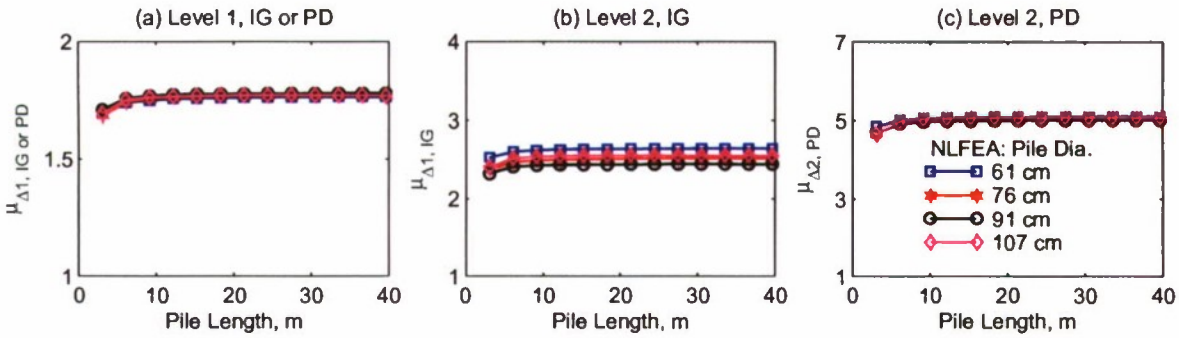


Figure 6.2. Variation of displacement ductility capacity computed from nonlinear finite element analysis (NLFEA) with pile length and pile diameter: (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, and (c) Level 2 for PD hinge formation.

In order to understand the aforementioned trend, i.e., independence of the displacement ductility capacity of the pile diameter, it is useful to examine the variation of pile section curvature ductility capacity. The results presented in Figure 6.3 indicate that the section curvature ductility capacity is essentially independent of the pile diameter. This observation, along with Equation (6.5), then confirms that the pile displacement ductility capacity should also be independent of the pile diameter.

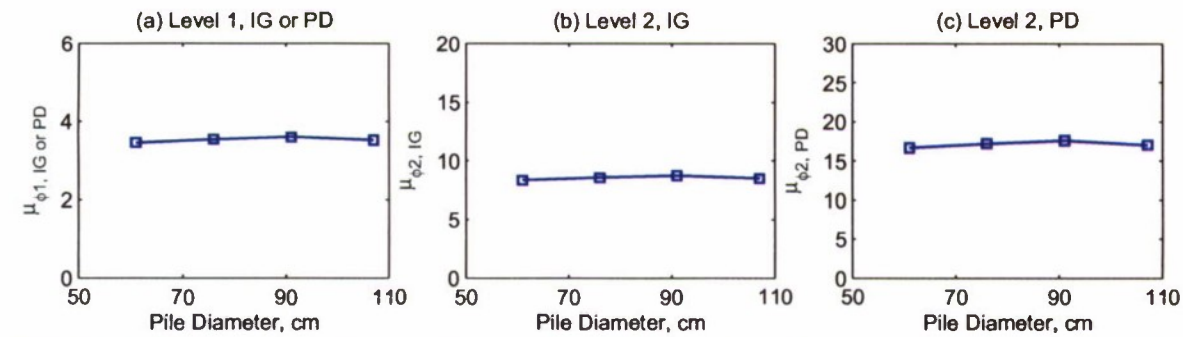


Figure 6.3. Variation of section curvature ductility capacity pile diameter: (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, and (c) Level 2 for PD hinge formation.

6.3.2 Longitudinal and Transverse Reinforcement Ratio

Figures 6.4 and 6.5 present variations of the displacement ductility capacity with longitudinal and transverse reinforcement ratio, respectively. The results presented are for a pile with 91 cm diameter and 15 m length. The values of longitudinal reinforcement varying between 0.5% and 2% and transverse reinforcement between 0.5% and 1.5% were considered.

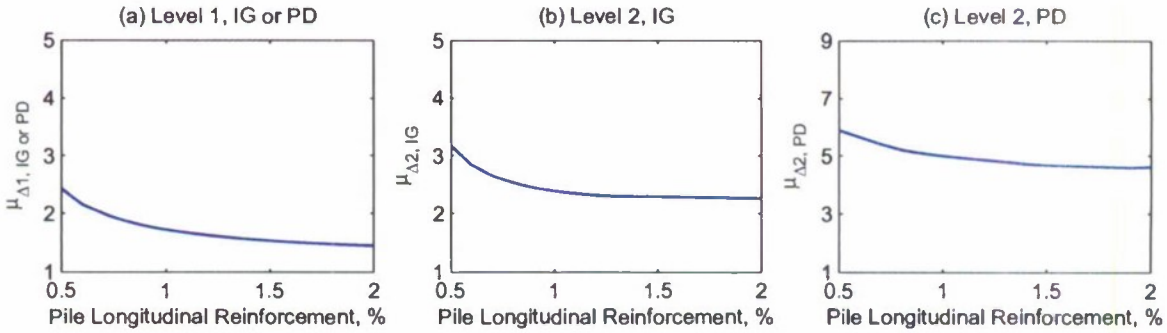


Figure 6.4. Variation of displacement ductility capacity computed from nonlinear finite element analysis (NLFEA) with pile longitudinal reinforcement ratio: (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, and (c) Level 2 for PD hinge formation.

The results presented in Figure 6.4 indicate that the displacement ductility decreases with increasing longitudinal reinforcement ratio for values up to about 1%. For longitudinal reinforcement ratio in excess of about 1%, as may be the case for seismic piles in Marine Oil and LNG Terminals, the displacement ductility capacity of piles is much less sensitive to the value of the longitudinal reinforcement ratio. For such values, the displacement ductility capacity may be considered to be essentially independent of the longitudinal reinforcement ratio.

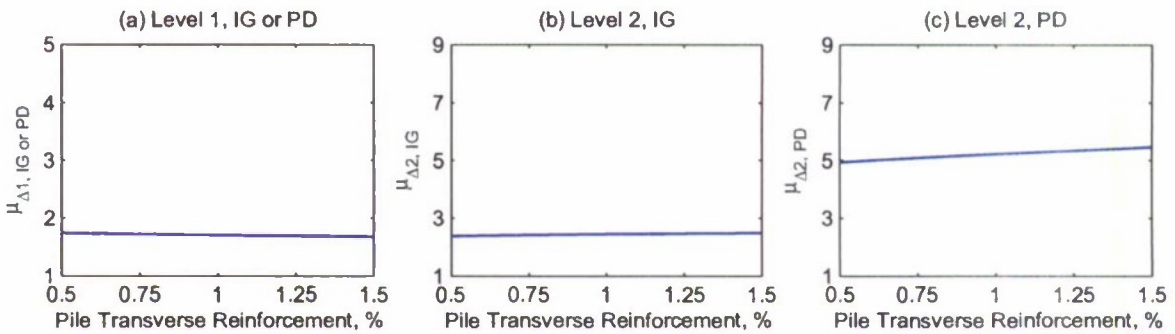


Figure 6.5. Variation of displacement ductility capacity computed from nonlinear finite element analysis (NLFEA) with pile transverse reinforcement ratio: (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, and (c) Level 2 for PD hinge formation.

The results presented in Figure 6.5 show that displacement ductility capacity of piles does not depend on the transverse reinforcement ratio. This becomes apparent from essentially flat variation of the displacement ductility capacity with pile transverse reinforcement ratio.

6.3.3 Axial Force

Figure 6.6 presents variation of displacement ductility capacity with axial force in the pile. The presented results are for a pile with 91 cm diameter and 15 m length for values of axial force varying from zero to $0.2A_g f_c^{'}$. These results show that the displacement ductility for Level 1 tends to increase with increasing pile axial force (Figure 6.6 (a)). However, the ductility for Level 2 appears to be insensitive to the axial force values (Figure 6.6 (b) and 6.6(c)).

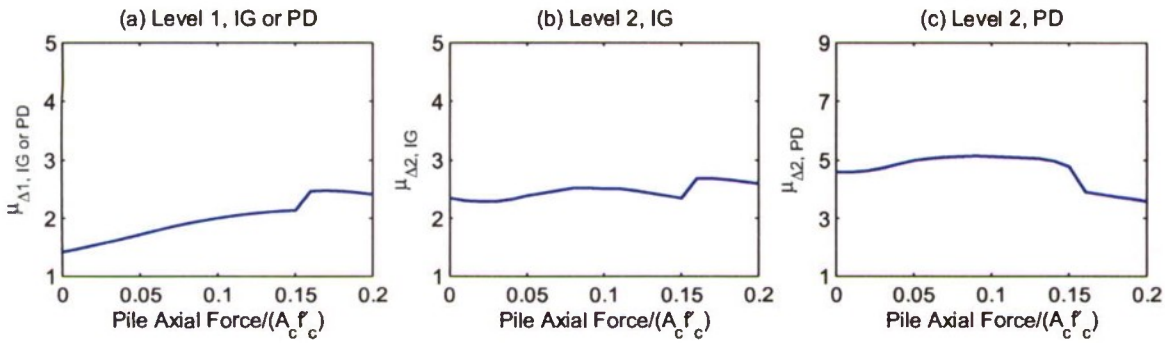


Figure 6.6. Variation of displacement ductility capacity computed from nonlinear finite element analysis (NLFEA) with pile axial load ratio: (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, and (c) Level 2 for PD hinge formation.

6.4 LOWER BOUND OF DISPLACEMENT DUCTILITY CAPACITY

The results presented in the preceding section indicate that the displacement ductility is relatively insensitive to pile length, pile diameter, pile longitudinal (for practical range), and transverse steel. Furthermore, the displacement ductility for Level 2 is also independent of the pile axial force. Therefore, the displacement ductility appears to be a very robust parameter that can be used in simplified design of piles instead of the various axial strain limits which are currently specified in the MOTEMS. While the displacement ductility may be related to the pile curvature ductility using Equation (6.5), the results presented in the preceding section also indicate that a lower bound of the member displacement ductility capacity may be estimated without any knowledge about the section curvature ductility capacity for practical range of various parameters.

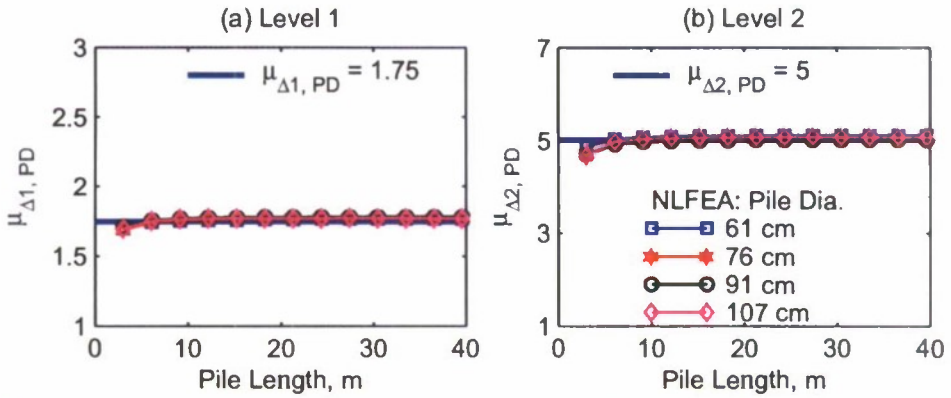


Figure 6.7. Lower-bound value of displacement ductility capacity due to pile-deck hinge formation for seismic design (a) Level 1, and (b) Level 2.

The results presented in Figure 6.7 for a pile-deck hinge indicates that the displacement ductility capacity may be limited to 1.75 for seismic design Level 1 and 5.0 for seismic design Level 2. Note that the displacement ductility for Level 1 is likely to be slightly lower for axial force values than the $0.05A_g f_c'$ value considered in developing these results (see Figure 6.6 (a)). Similarly, the displacement ductility is likely to be slightly larger for longitudinal reinforcement less than the 1% value considered in developing these results (see Figure 6.4 (a)).

The displacement ductility capacity for an in-ground hinge is 1.75 for seismic design Level 1 and 2.5 for seismic design Level 2 (Figure 6.8). While the ductility capacity for in-ground hinge is the same as for pile-deck hinge for design Level 1, it is much lower for design Level 2. This is because the steel strain limit for design Level 2 is much lower for the in-ground hinge compared to the pile-deck hinge.

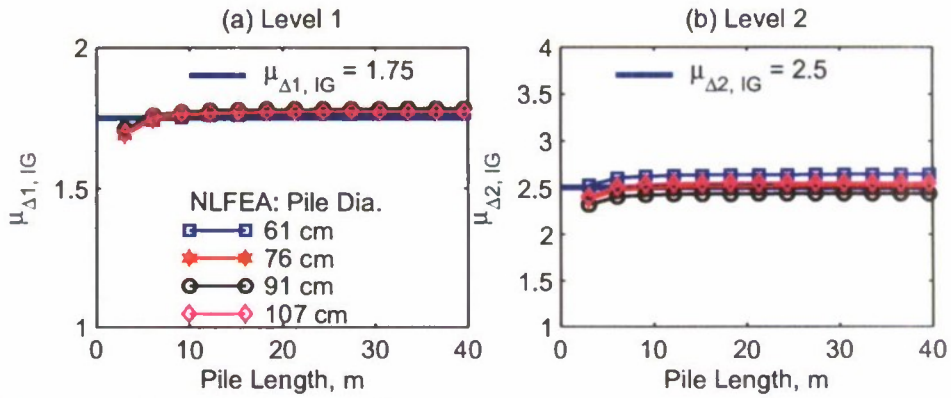


Figure 6.8. Lower-bound value of displacement ductility capacity due to in-ground hinge formation for seismic design (a) Level 1, and (b) Level 2.

6.5 SIMPLIFIED PROCEDURE TO COMPUTE DISPLACEMENT CAPACITY

Displacement capacity of piles at a selected design level may be estimated from

$$\Delta_c = \mu_\Delta \Delta_y \tag{6.6}$$

in which μ_Δ is the ductility capacity at a selected design level and location of hinge, i.e., equal to 1.75 for Level 1 design and 5 for Level 2 design if the hinge were to form in the pile near the deck, and equal to 1.75 for Level 1 and 2.5 for Level 2 if the hinge were to form in-ground, and Δ_y is the yield displacement of the pile. The yield displacement can be computed from nonlinear pushover analysis of the pile. Alternatively, the yield displacement may be estimated based on section yield moment and effective section EI_e. For example, the yield displacement of a pile that is fixed at the bottom and prevented from rotation at the top due to a rigid deck may be estimated from

$$\Delta_y = \frac{M_y L^2}{6EI_e} \tag{6.7}$$

and yield displacement of a cantilever may be estimated from

$$\Delta_y = \frac{M_y L^2}{3EI_e} \tag{6.8}$$

in which M_y is the section yield moment and EI_e is the effective value of EI that can be estimated from the section moment-curvature relationship analysis as the initial slope of the idealized bilinear moment-curvature relationship (see Figure 4.2).

The accuracy of the procedure to estimate the displacement capacity of piles is evaluated next. For this purpose, the approximate displacement capacity is computed first from Equation (6.6) by utilizing the yield displacement from Equation (6.7) or (6.8) depending on the boundary conditions. The exact displacement capacity is computed next from Equation (6.6) but with yield displacement estimated from nonlinear static pushover analysis of the pile. For both cases, the value of the ductility capacity obtained from the pushover analysis is used. The approximate and exact displacement capacities are compared in Figure 6.9 for a pile with 91 cm diameter. These

results indicate that the approximate analysis provides an excellent estimate of the displacement capacity of the pile for Level 1 as well as Level 2 design.

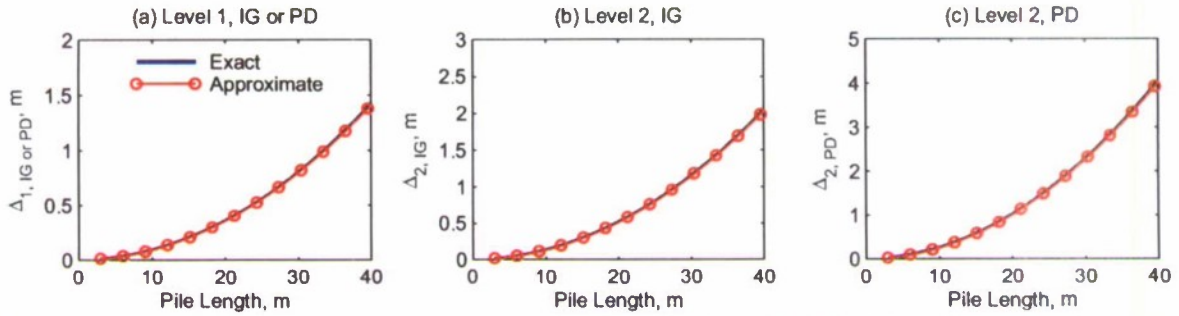


Figure 6.9. Comparison of displacement capacities due to pile-deck hinge formation from exact and approximate analyses.

The approximate analysis is attractive because it eliminates the need for nonlinear static analysis of the pile. However, it must be noted that the approximate analysis may only be used for the soil-pile-deck system that can be idealized either by a fixed-fixed column or by a cantilever column – the two cases for which closed form solutions to estimate yield displacement are available (see Equations 6.7 and 6.8) – using the equivalent displacement fixity concept. For other cases, the yield displacement may have to be estimated from nonlinear static pushover analysis of the soil-pile-deck system.

7. DISPLACEMENT CAPACITY OF HOLLOW STEEL PILES

This Chapter presents development of a simplified procedure for estimating displacement capacity of hollow steel piles connected to the deck either by a pin connection or by a full-moment-connection strong enough to force hinging in the steel pile. For this purpose, the current approach in the MOTEMS (see Equations 4.1 to 4.6 in Chapter 4) is further simplified. Presented first is the development of simplified equations to compute displacement ductility of hollow steel piles that are independent of the pile length and depend only on the pile section ductility and seismic design level. The accuracy of these equations is then evaluated against results from nonlinear finite element analyses. Subsequently, results of a parametric study are presented to show the sensitivity of the displacement ductility capacity on pile diameter, pile thickness, and axial force level. Based on these results, lower bound estimates of the ductility capacity of hollow steel piles for two design levels – Level 1 and Level 2 – are proposed. Finally, it is demonstrated that the lower-bound displacement ductility values along with simplified expressions for yield displacement provide very good estimates of the displacement capacity of piles when compared against values from nonlinear finite element analysis.

7.1 THEORETICAL BACKGROUND

Similar to the displacement ductility of reinforced concrete piles, the displacement ductility capacity of hollow steel piles may also be defined as

$$\mu_\Delta \simeq 1+3\left(\mu_\phi - 1\right)\left(\frac{L_p}{L}\right)\left(1-0.5\frac{L_p}{L}\right) \tag{7.1}$$

The MOTEMS does not explicitly provide guidelines for selecting length of the plastic hinge for hollow steel piles. Based on calibration of results from finite element analysis against those from Equation (7.1) (see results presented later in Figure 7.1), it was found that the following plastic hinge lengths are appropriate for the two seismic design levels for hollow steel piles in Marine Oil and LNG Terminals:

$$L_p \simeq 0.03L \quad \text{for Level 1} \tag{7.2a}$$

$$L_p \simeq 0.075L \quad \text{for Level 2} \tag{7.2b}$$

With the plastic hinge length selected as given by Equations (7.2(a) and 7.2(b)), Equation (7.1) simplifies to

$$\mu_\Delta = 0.9113 + 0.0886\mu_\phi \quad \text{for Level 1} \tag{7.3a}$$

$$\mu_\Delta = 0.7834 + 0.2166\mu_\phi \quad \text{for Level 2} \tag{7.3b}$$

As noted previously for reinforced concrete piles, Equations (7.3(a) and 7.3(b)) for displacement ductility capacity of hollow steel piles also indicates that the displacement ductility capacity is independent of the pile length and it can be computed directly from the section curvature ductility capacity. Because the plastic hinge length differs for the two design levels, the displacement ductility also depends on the seismic design level.

7.2 EVALUATION OF SIMPLIFIED EQUATIONS FOR DUCTILITY CAPACITY

The accuracy of Equations (7.3(a) and 7.3(b)) in estimating displacement ductility capacity of hollow steel piles at seismic design Level 1 and Level 2, respectively, is evaluated in this section. For this purpose, displacement ductility capacity of hollow steel piles is evaluated from nonlinear static pushover analysis of a finite element model. The pile is considered to be fixed at top and bottom. These boundary conditions correspond to a pile that is connected to the pile-cap with a full-moment connection that would force formation of a plastic hinge in the steel pile, and utilizes the equivalent displacement fixity assumption at the bottom. The axial load on the pile is assumed to be $0.05Af_y$ in which A is the cross section area of the pile and f_y is the yield strength of steel. The pile wall thickness is assumed to be 1.27 cm.

The pile is modeled with a nonlinear beam-column element using the computer program "Open System for Earthquake Engineering Simulation (*OpenSees*)", (McKenna and Fenves, 2001). The distributed plasticity is considered by specifying the section properties by a fiber section model and the using seven integration points along the element length; details of such modeling may be found in McKenna and Fenves (2001). Strains in steel are monitored during the pushover analysis. The limiting values of strain in steel are 0.008 and 0.025 for Level 1 and Level 2, respectively for in-ground or pile-deck hinge formation. The displacement ductility at a selected design level corresponds to the largest displacement that can occur at the tip of the pile without the strain limit in steel being exceeded.

The results are presented in Figure 7.1 for three pile diameters – 61 cm, 91 cm, and 107 cm. These results permit two important observations. First, results from the nonlinear finite element analysis confirm expectations from Equation (7.3(a) and 7.3(b)) that the displacement ductility capacity is independent of the pile length. This becomes apparent by essentially no variation in the ductility capacity from the nonlinear finite element analysis with the various pile lengths in Figure 7.1 for both design levels and all pile diameters.

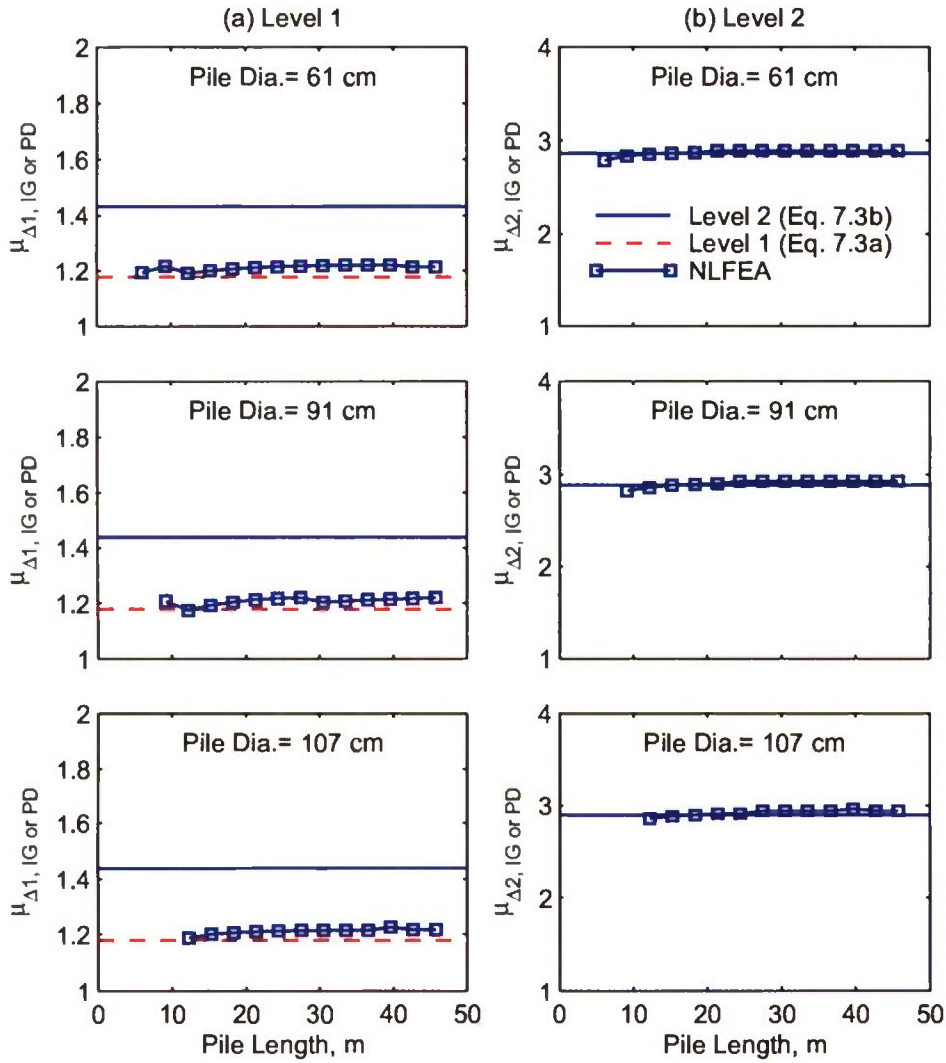


Figure 7.1. Displacement ductility capacity from simplified equations and nonlinear finite element analysis (NLFEA) for seismic design (a) Level 1 and (b) Level 2.

Second, Equations (7.3(a) and 7.3(b)) provide very good estimates of the displacement ductility capacity of hollow steel piles at seismic design Level 1 (see Figure 7.1(a)) and Level 2 (see Figure 7.1(b)), respectively. If Equation (7.3(b)) were to be used to estimate, displacement

ductility capacity at seismic design Level 1, it would provide an estimate that significantly exceeds the value from nonlinear finite element analysis (see Figure 7.1(a)). Therefore, a lower value of the plastic hinge length, as has been used in Equation (7.3(a)) for seismic design Level 1 is justified.

These results indicate that the moment-rotation relationship to be used in the concentrated plasticity model of hollow steel piles should consider different plastic hinge lengths for the two design levels. If the same plastic hinge length, i.e., that for seismic design Level 2, is used in the model that computes the displacement ductility capacity for Level 1, it may significantly overestimate the displacement capacity for that design level (Level 1).

It is useful to note that the plastic hinge length for hollow steel piles in this investigation is proposed based on calibration against nonlinear finite element results. It would be useful to verify these findings from experiments on hollow steel pile conducted at displacement levels that are expected during seismic design Level 1 and Level 2.

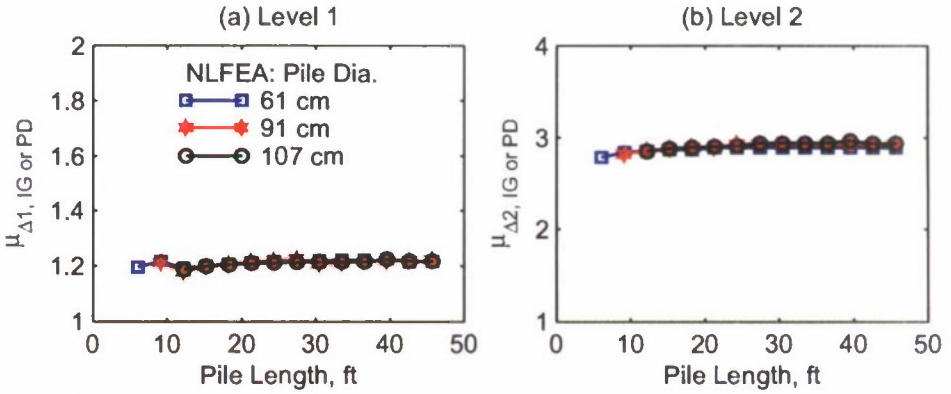


Figure 7.2. Variation of displacement ductility capacity computed from nonlinear finite element analysis (NLFEA) with pile length and pile diameter: (a) Level 1, and (b) Level 2.

7.3 SENSITIVITY OF DISPLACEMENT DUCTILITY TO PILE PARAMETERS

7.3.1 Pile Length and Pile Diameter

Figure 7.2 presents variation of displacement ductility capacity with pile length for three values of pile diameters: 61 cm, 91 cm, and 107 cm. The results are presented for piles with wall thickness of 1.27 cm. Results in Figure 7.2 indicate that the displacement ductility capacity of piles is essentially independent of the pile length. This is expected because Equations (7.3(a) and 7.3(b)) becomes independent of the pile length. The results of Figure 7.2 also indicate that the

displacement ductility capacity of the pile is also essentially independent of the pile diameter as apparent from almost identical curves for the three pile diameters considered.

In order to understand the aforementioned trend, i.e., independence of the displacement ductility capacity of pile diameter, it is useful to examine the variation of pile section curvature ductility capacity. The results presented in Figure 7.3 indicate that the section curvature ductility capacity is essentially independent of the pile diameter. This observation, along with Equations (7.3(a) and 7.3(b)), then confirms that the pile displacement ductility capacity should also be independent of the pile diameter.

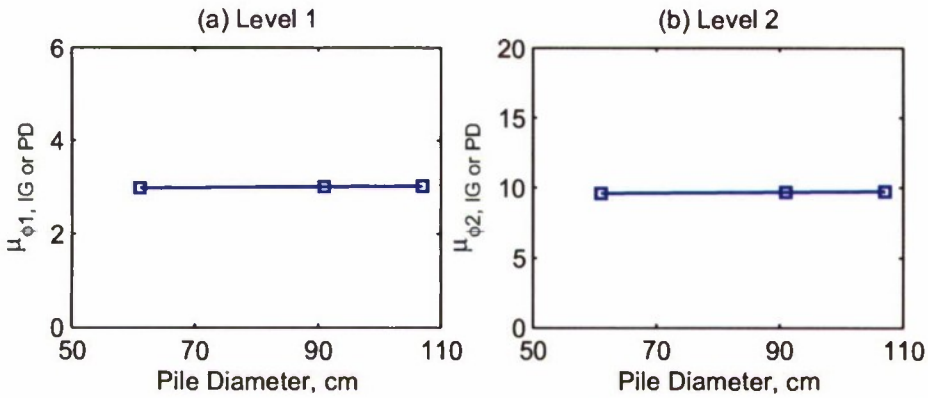


Figure 7.3. Variation of section curvature ductility capacity with pile diameter: (a) Level 1, and (b) Level 2.

7.3.2 Pile Wall Thickness

The effects of the pile wall thickness on the displacement ductility capacity are examined next. For this purpose, variations of displacement ductility with pile length for three values of pile thickness are compared in Figure 7.4. The results presented are for a pile with 91 cm diameter and axial force equal to $0.05Af_y$. These results show that the displacement ductility is essentially independent of the pile wall thickness as indicated by essentially identical curves for the three values of pile wall thickness.

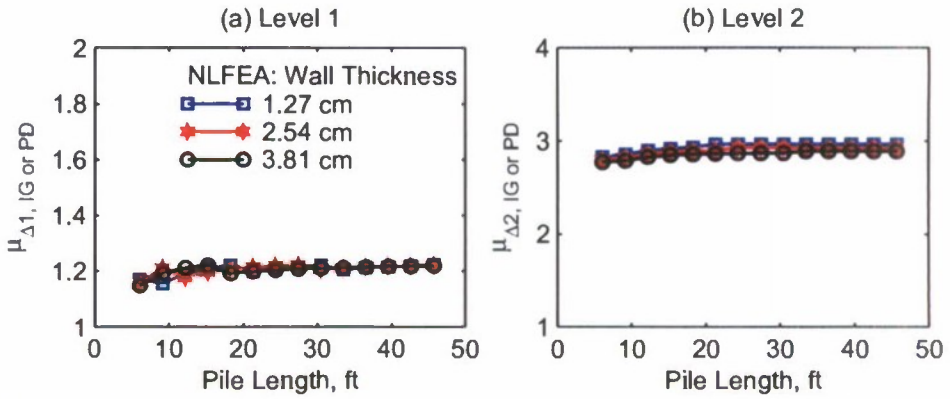


Figure 7.4. Variation of displacement ductility capacity computed from nonlinear finite element analysis (NLFEA) with pile length for three values of pile wall thickness: (a) Level 1, and (b) Level 2.

7.3.3 Axial Force

Figure 7.5 presents variation of displacement ductility capacity with axial force in the pile. The presented results are for a pile with 91 cm diameter and 15 m length with values of axial force varying from zero to $0.2Af_y$. These results show that the displacement ductility for Level 1 is essentially independent of the pile axial load (Figure 7.5(a)). For Level 2, while the displacement ductility may depend on the axial load for very-low axial loads, it becomes essentially independent of the axial load for more realistic values. However, the ductility for Level 2 appears to be insensitive to the axial force values, i.e., axial loads greater than $0.05Af_y$ (Figure 7.5(b)).

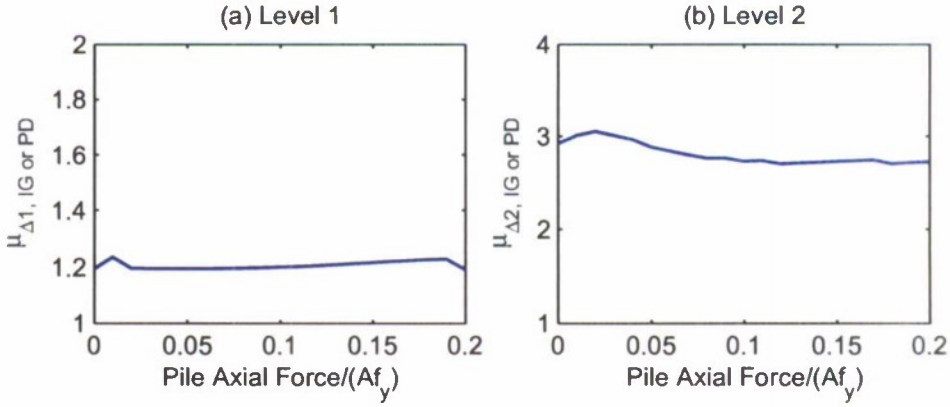


Figure 7.5. Variation of displacement ductility capacity computed from nonlinear finite element analysis with pile axial load ratio: (a) Level 1, and (b) Level 2.

7.4 LOWER BOUND OF DISPLACEMENT DUCTILITY CAPACITY

The results presented so far indicate that the displacement ductility of hollow steel piles is relatively insensitive to pile length, pile diameter, pile wall thickness, and pile axial load. Therefore, the displacement ductility appears to be a very robust parameter that can be used in simplified design of piles instead of the various axial strain limits which are currently specified in the MOTEMS. While the displacement ductility may be related to the pile curvature ductility using Equation (7.3), the results presented in the preceding section also indicate that a lower bound of the member displacement ductility capacity may be estimated without any knowledge about the section curvature ductility capacity for practical range of various parameters. The results presented in Figure 7.6 for pile-deck hinge indicate that the displacement ductility capacity may be limited to 1.2 for seismic design Level 1 and 2.75 for seismic design Level 2.

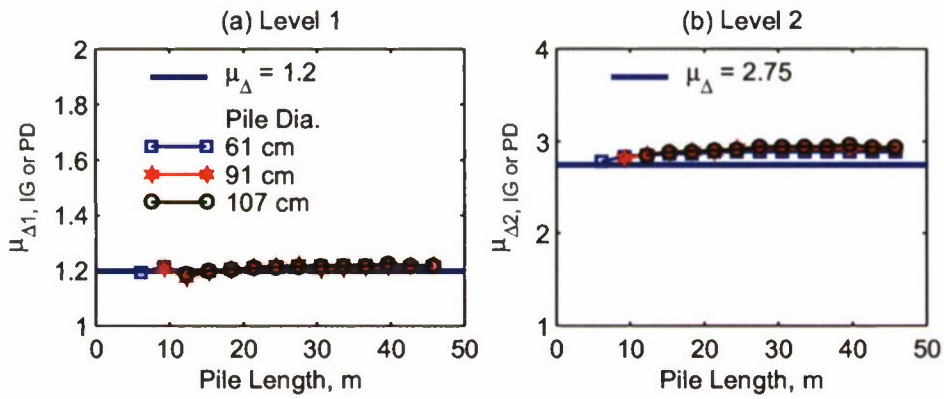


Figure 7.6. Lower-bound value of displacement ductility capacity of hollow steel piles for seismic design (a) Level 1, and (b) Level 2.

7.5 SIMPLIFIED PROCEDURE TO COMPUTE DISPLACEMENT CAPACITY

Displacement capacity of piles at a selected design level may be estimated from

$$\Delta_c = \mu_\Delta \Delta_y \tag{7.4}$$

in which μ_Δ is the ductility capacity at a selected design level, i.e., equal to 1.2 for Level 1 design and 2.75 for Level 2 design, and Δ_y is the yield displacement of the pile. The yield displacement can be computed from nonlinear pushover analysis of the pile. Alternatively, the yield displacement may be computed based on section properties. For example, the yield displacement of a pile that is fixed at the bottom and prevented from rotation at the top due to

rigid deck may be estimated from

$$\Delta_y = \frac{M_y L^2}{6EI} \tag{7.5}$$

and yield displacement of a cantilever may be estimated from

$$\Delta_y = \frac{M_y L^2}{3EI} \tag{7.6}$$

in which M_y is the effective section yield moment that can be estimated from section moment-curvature analysis and I is the section moment of inertia that can be estimated from the section properties, and E is the modulus of elasticity for steel.

The accuracy of the approximate procedure to estimate the displacement capacity of piles is evaluated next. For this purpose, the approximate displacement capacity is computed first from Equation (7.4) by utilizing the yield displacement from Equation (7.5) or (7.6) depending on the boundary conditions. The exact displacement capacity is computed next from Equation (7.4) but with yield displacement estimated from nonlinear the static pushover analysis of the pile. For both cases, value of the ductility capacity obtained from the pushover analysis is used. The approximate and exact displacement capacities are compared in Figure 7.7 for a pile with 91 cm diameter. These results indicate that the approximate analysis provides an excellent estimate of the displacement capacity of the pile for Level 1 as well as Level 2 design.

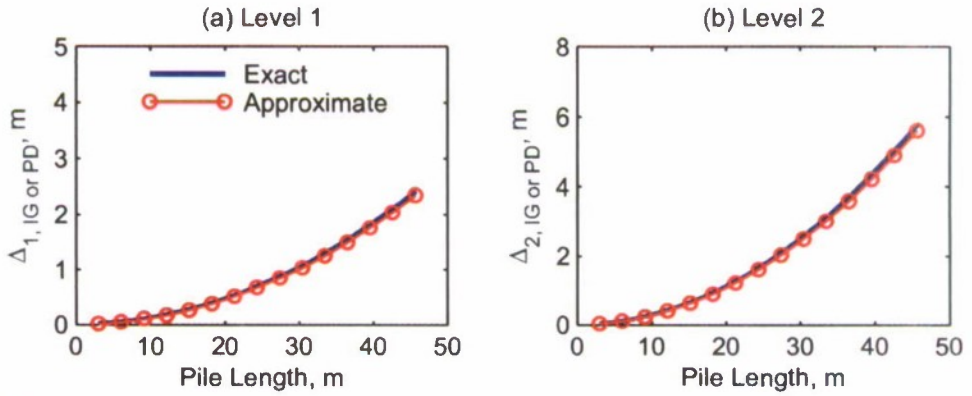


Figure 7.7. Comparison of displacement capacities from exact and approximate analyses.

The approximate analysis is attractive because it eliminates the need for nonlinear static analysis of the pile. However, it must be noted that the approximate analysis may only be used

for the soil-pile-deck system that can be idealized either by a fixed-fixed column or by a cantilever column – the two cases for which closed form solutions to estimate yield displacement are available (see Equations 7.5 and 7.6) – using equivalent displacement fixity concept. For other cases, the yield displacement may have to be estimated from nonlinear static pushover analysis of the soil-pile-deck system.

8. DISPLACEMENT CAPACITY OF PILES WITH DOWEL-CONNECTION

Piles are often eonneeted to thc deck using dowels. The size and number of dowel bars are typieally seleeted so that the moment capacity of the eonneetion is smallcr than the moment eapaeity of the pile. As a result, the yielding is expeeted to oeeur in the eonnection rather than the pile. The nonlinear behavior of piles with sueh partial-moment conneetion to the deek slab may differ signifieantly compared to the piles with full-moment eonneetions presented in the previous ehapters. This ehapter deseribes two types of dowel-connections – hollow steel piles connected to the deek by a conerete plug and dowels, and prestressed conerete piles connected to the deck by dowels grouted into the pilc and embedded in the deck concrete. Subsequcntly, nonlinear behavior of such connections is examined. Finally, closcd form solutions for estimating displaeement eapaeity of piles with partial-moment connections are presented.

8.1 DOWEL-CONNECTIONS

8.1.1 Hollow Steel Piles

Figure 8.1 shows details of the conneetions between a hollow steel pile and the concrete deck of a Marine Oil or LNG Terminal. In this eonneetion, denoted as the eonerete-plug eonneetion, dowels are embedded in a eonerete plug at the top of the pile. The eonerete plug is held in plaee by shear rings at its top and bottom; the shear rings would prevent the concrete plug from slipping out (or popping-out) during lateral loads imposed by earthquakes. Others have proposed details in whieh the conerete plug is held in place either by natural roughness of the inside surfaee of the steel shell or use of weld-metal laid on the inside of the steel shell in a eontinuous spiral in the eonneetion region prior to plaeing the eonerete plug (Ferritto et al., 1999). The dowels are then embedded in the eonerete deek to provide suffieient development length. A small gap may or may not be provided between top of the pile and top of the concrete plug. This eonerete-plug eonneetion has been shown to provide remarkable duetility eapaeity of hollow steel piles (Priestley and Park, 1984; Park et al., 1987). The foree transfer meehanism between the steel pile and the conerete plug has also been investigated by Nezamian et al. (2006).

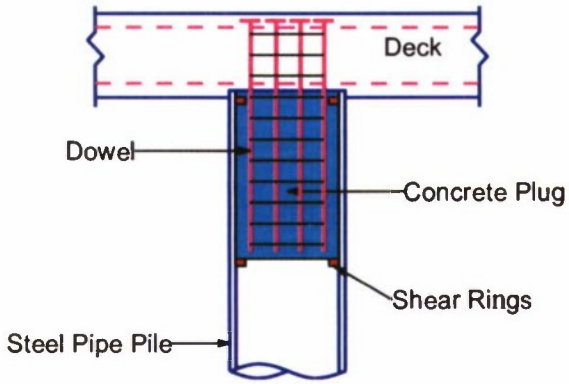


Figure 8.1. Concrete-plug connection between hollow steel pile and concrete deck.

8.1.2 Prestressed Concrete Piles

Figure 8.2 shows details of the connections between a prestressed pile and the concrete deck of a Marine Oil or LNG Terminal (Klusmeyer and Harn, 2004; Wray et al., 2007; Roeder et al., 2005). Prestressed piles typically have corrugated metal sleeves that are embedded in the concrete. These sleeves are located inside of the confined concrete core formed by the prestressing strands and confining steel. Once the prestressed pile has been driven to the desired depth, the dowels are grouted into the sleeves. If higher flexibility of the connection is desired, a small portion of the dowel at the top of the pile may be wrapped in Teflon to ensure de-bonding between the dowel and the grout. The dowels are then embedded in the concrete deck to provide sufficient development length. Note that Figure 8.2 shows only two outermost dowels; the other dowels are not shown to preserve clarity in the figure.

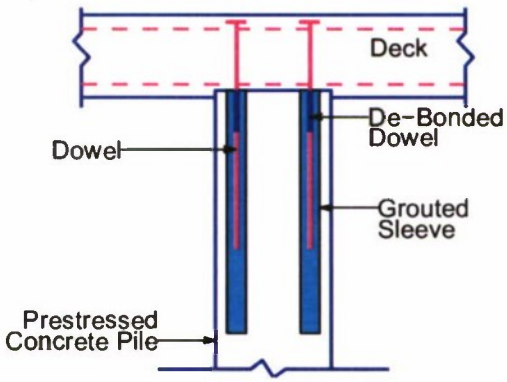


Figure 8.2. Dowel-connection between prestressed concrete pile and concrete deck.

8.1.3 Behavior of Dowel-Connection

While analyzing Marine Oil and LNG Terminal structures, nonlinear behavior of pile and connection is typically represented by moment-rotation relationships. The moment-rotation relationship is developed based on the assumption of a plane section remaining plane and a perfect bond between the steel reinforcing bars and concrete. For the concrete-plug connection between hollow steel piles and deck or the dowel-connection between a prestressed pile and deck; however, such assumptions may not be valid. In particular, the pile in a such connection rotates about a small area on compression side of the pile forming a gap between the top of the pile and the deck on the tension side of the pile (see Figure 8.3). This behavior is akin to the pile acting like a crowbar bearing on a small compression area. This behavior leads to de-bonding of the dowel (or strain penetration) on each side of the joint. Additional de-bonding may also occur in the dowel over the portion that is intentionally wrapped in Teflon.

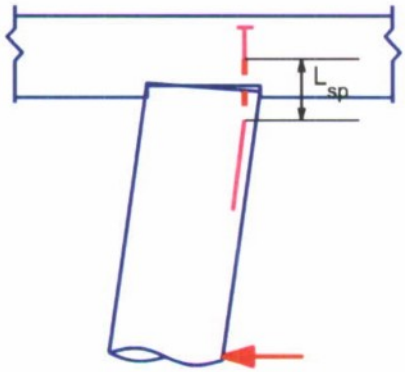


Figure 8.3. Behavior of piles with concrete-plug or dowel-connection.

8.2 MOMENT-ROTATION RELATIONSHIP OF DOWEL-CONNECTION

Bob Harn and George Sheng of Berger/ABAM Engineers Inc recently proposed a simple analytical model for developing nonlinear moment-rotation behavior of concrete-plug connections for hollow steel piles (see Figure 8.4) or dowel-connections for prestressed piles (see Figure 8.5). For a selected value of the reinforcing bar yield stress, f_y, concrete strength, $f_c^{'}$, diameter and area of reinforcing bars, d_{bi} and A_{si}, respectively, bearing strength of deck concrete against pile concrete as $f_m^{'} = 1.7 f_c^{'}$, and bearing strength of deck concrete against steel shell of hollow steel pile as $f_m^{'} = 5.6 f_c^{'}$, the moment-rotation relationship is developed as

follows:

1. Select a value of strain in the outermost dowel on the tension side, ε_1. Typically the first strain value is selected as the yield strain in steel, ε_y.
2. Establish the location of the neutral axis of the section by the following iterative procedure:
   2.1. Guess the location of the neutral axis.
   2.2. Calculate strains in all dowels.
   2.3. Calculate forces in all dowels, T_i. Note that dowel forces would be tensile on the tension side of the neutral axis and compressive on the compression side of the neutral axis.
   2.4. Calculate compressive force, C_c, in concrete on compression side of the neutral axis.
   2.5. Calculate compressive force, C_s, due to bearing of steel shell against the deck for hollow steel piles. Note that this step would not be necessary for prestressed concrete piles.
   2.6. Check that summation of all forces, including any axial force on the pile, is equal to zero.
   2.7. Repeat Steps 2.1 to 2.6 until summation of forces in Step 2.6 is essentially equal to zero.
3. Estimate the length of strain-penetration in the dowel: $L_{sp} = 0.15 f_s d_b + L_{db}$ in which f_s is the allowable dowel stress in units of ksi, d_b is the dowel diameter in inches, and L_{db} is the length of de-bonded reinforcing bar (as may be the case for prestressed concrete piles). Alternatively, the strain penetration length may be selected as $L_{sp} = 5d_b + L_{db}$ or as per the recommendations by Raynor et al. (2002).
4. Compute the elongation of the outermost dowel: $\Delta L_1 = \varepsilon_1 L_{sp}$.
5. Compute the rotation of the concrete-plug connection: $\theta = \Delta L_1 / Y_1$ in which Y_1 is the distance between the neutral axis and the outermost dowel on the tension side of the neutral axis.
6. Compute the moment, M, as the summation of moments at the center of the pile due to tensile as well as compressive forces.
7. Repeat Steps 1 to 6 to develop the entire moment-rotation relationship of the connection.
8. Idealize the moment-rotation relationship by using a bi-linear curve.

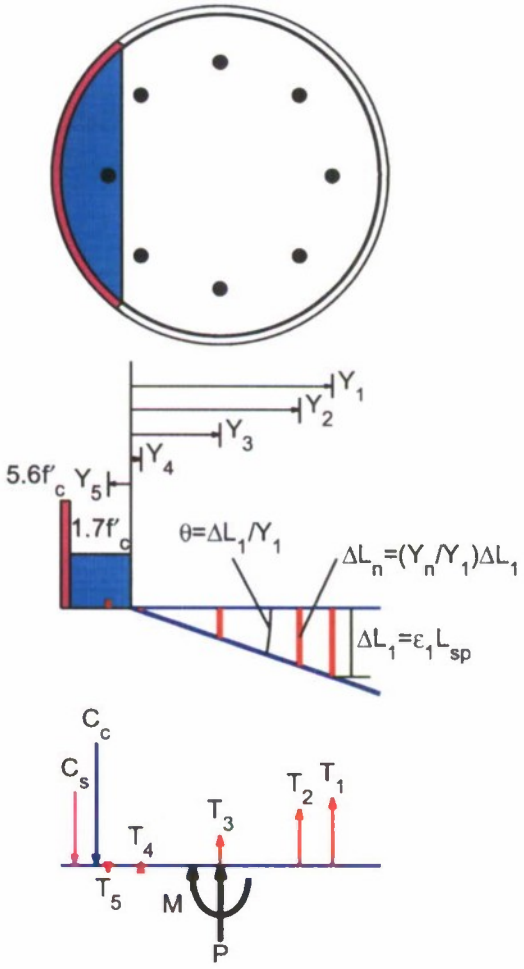


Figure 8.4. Analytical model to generate the moment-rotation relationship of the concrete-plug connection between a hollow steel pile and a concrete deck.

Figures 8.6 and 8.7 show the moment rotation relationship of the concrete-plug connection for a hollow steel pile, a dowel connection and for a prestressed concrete pile, respectively. The nonlinear moment-rotation relationship (shown in solid line) has been idealized by a bilinear moment-rotation relationship (shown in dashed line). It is apparent from these results that the post-yield slope of the moment-rotation relationship is very small compared to the slope in the linear-elastic portion. Therefore, it may be possible to simply idealize this curve with an elastic-perfectly-plastic curve without much loss in accuracy.

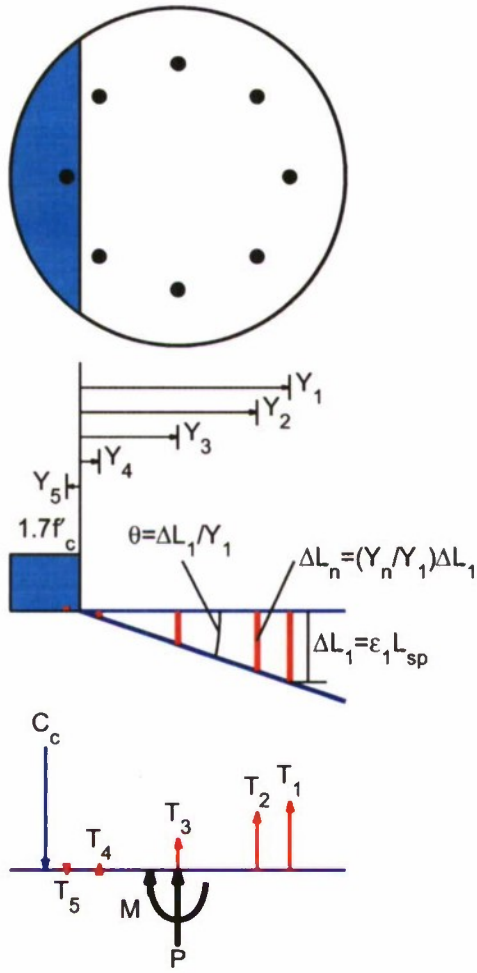


Figure 8.5. Analytical model to generate the moment-rotation relationship of a dowel connection between a prestressed concrete pile and a concrete deck.

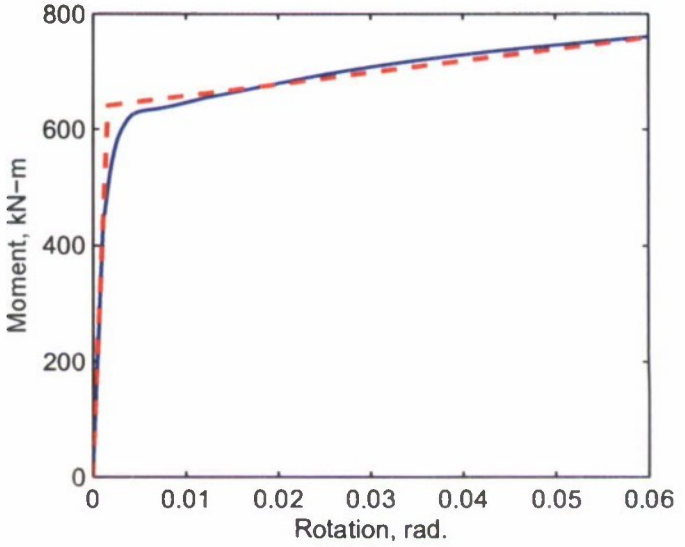


Figure 8.6. Moment-rotation relationship of a concrete-plug connection for hollow steel piles. The results are for a steel pile of 61 cm diameter, 1.27 cm wall thickness, axial load of $0.05 f_y A$, and 8 dowels each with an area of 8.2 cm^2.

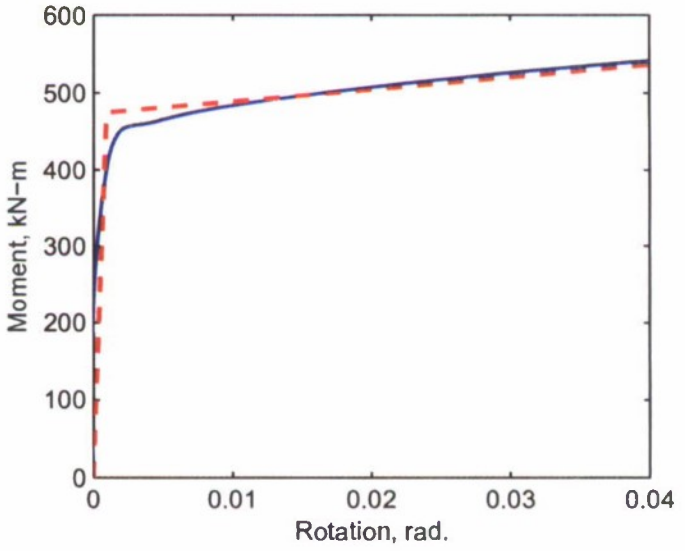


Figure 8.7. Moment-rotation relationship of a dowel connection for prestressed concrete piles. The results are for a steel pile of 61 cm diameter pile, axial load of $0.05 f_c' A$, 8 dowels each with an area of 3.9 cm^2, and de-bonded length of reinforcing bars equal to 30 cm.

9. SIMPLIFIED MODEL OF PILE WITH DOWEL-CONNECTION

A hollow steel pile with a concrete-plug connection or a prestressed pile with a dowel connection to the deck may be idealized as a beam-column element fixed at the base and a rotational spring at the top (Figure 9.1). The length of the element is equal to the free-standing height of the pile plus the depth of fixity below the mud-line. This length is selected as the length of a fixed-base cantilever that would have same lateral displacement at the pile top as the actual pile (see Priestley at al., 1996; Chai, 2002). The rotational spring at the top of the pile represents the nonlinear behavior of the concrete-plug or the dowel connection. Ignoring axial deformations in the pile, this system can be modeled with two displacement degrees-of-freedom: lateral displacement, Δ, and rotation, θ, at the top. When a lateral force, F, is applied at the top of the pile, a moment, M, also develops at the top due to the rotational resistance provided by the rotational spring representing the concrete-plug or the dowel connection. Note that the rotation in the rotational springs is equal to rotation at top of the pile.

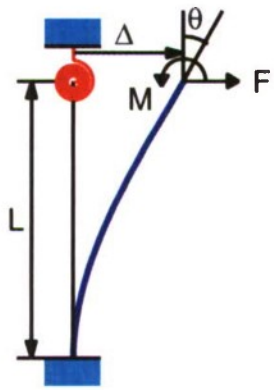


Figure 9.1. Simplified model of the pile with partial-moment connection to the deck.

Presented in this chapter is the development of a simplified procedure for estimating the displacement capacity of hollow steel piles with concrete-plugs or prestressed piles with dowel connections at the deck without the need to monitor strains during the pushover analysis. In particular, formulas for estimating displacement capacity of such piles are developed.

9.1 IDEALIZED CONNECTION AND PILE BEHAVIOR

9.1.1 Moment-Rotation Behavior of Connection

The moment-rotation relationship for the concrete-plug or dowel connection between the pile

and the deek may be idealized as a bilinear (elastie-perfeetly-plastie) eurve (Figure 9.2). The initial elastie stiffness and yield moment of the partial-moment-connection are defined by k_θ and $M_{y,C}$, respectively. If θ_L is the rotation in the rotational spring when the strain in the outermost dowel of the concrete-plug connection for hollow steel piles or the dowel connection in prestressed eoncrete piles just reaehes the strain limit specified for a selected design level, the rotational duetility of the connection at specified strain limits is defined by

$$\mu_\theta = \frac{\theta_L}{\theta_y} \tag{9.1}$$

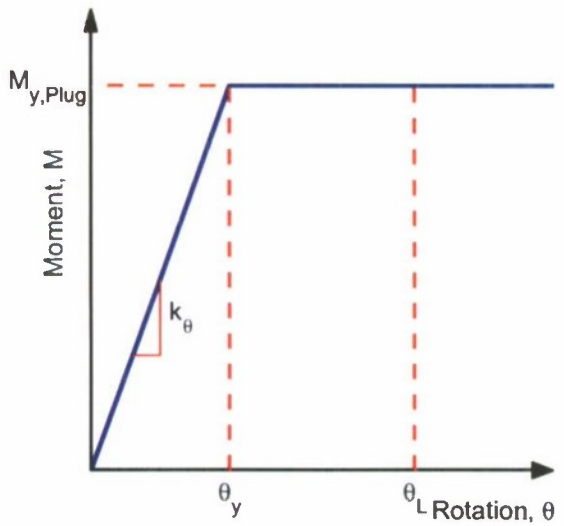


Figure 9.2. Idealized moment-rotation relationship of the dowel-eonneetion.

9.1.2 Moment-Curvature Behavior of Pile Section

The moment-curvature relationship of the pile section can also be idealized as a bilinear curve (Figure 9.3). The initial slope of this curve is equal to EI and post-yield slope is equal to αEI in whieh α is the ratio of the post-yield slope and initial slope of the eurve. The moment and eurvature at effective yielding of the pile are $M_{y,P}$ and ϕ_y, respeetively. Note that the effeetive yield moment, $M_{y,P}$, of the pile section in the idealized bi-linear moment-curvature relationship differs slightly from the yield moment at initiation of first yielding in the outermost fiber of the hollow steel pile or outermost strand of the prestressed eonerete pile. While the $M_{y,P}$ for prestressed conerete piles should be estimated from the moment-eurvature relationship, $M_{y,P}$ for

hollow steel piles may be estimated from the formula for its plastic moment capacity as

$$M_{y,P} = f_y\left(\frac{d_o^3 - d_i^3}{6}\right) \tag{9.2}$$

If ϕ_L is the curvature of the pile section when the material strain just reaches the strain limit specified for a selected design level, the pile section curvature ductility is defined as

$$\mu_\phi = \frac{\phi_L}{\phi_y} \tag{9.3}$$

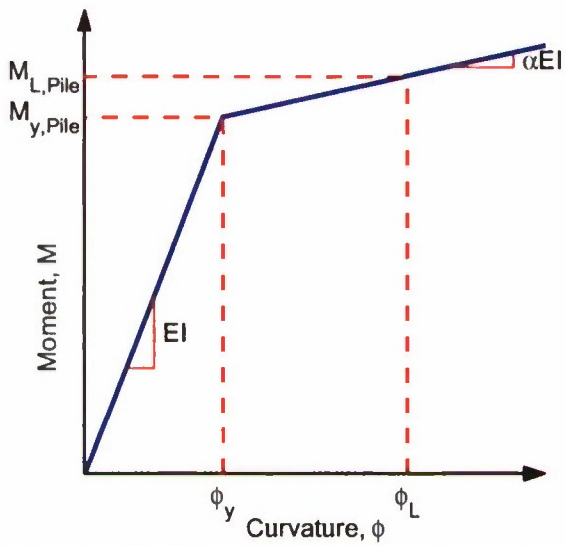


Figure 9.3. Idealized moment-curvature relationship of the pile section.

9.1.3 Force-Deformation Relationship of Pile with Dowel-Connection

The force-deformation behavior (or pushover curve) of a pile with fixed-base and a rotational spring at the top may be idealized by a tri-linear relationship shown in Figure 9.3. For piles with dowel-connections to the deck, the yield moment of the connection is typically selected to be smaller than the yield moment of the pile section. For such a condition, the first yielding in the pile system would occur in the connection at lateral force and displacement equal to $F_{y,C}$ and $\Delta_{y,C}$, respectively. Since the pile has not yet reached its yield moment, the lateral force in the pile system would continue to increase with displacement until yielding occurs in the steel pile at force and displacement equal to $F_{y,P}$ and $\Delta_{y,P}$, respectively. Subsequently, the lateral force in the pile system would increase with displacement only due to strain-hardening effects in the pile

material.

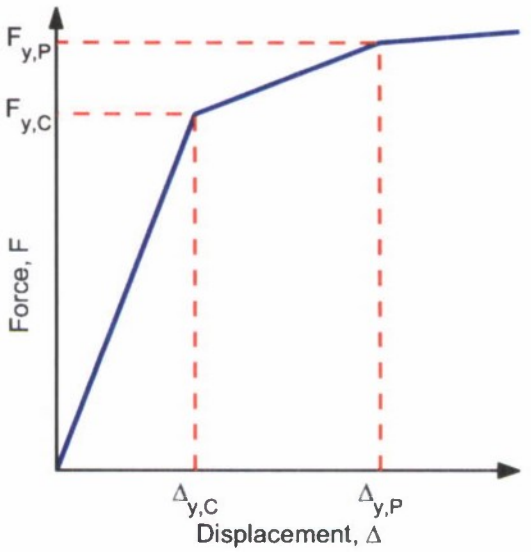


Figure 9.4. Idealized pushover curve of pile with dowel-connection to the deck.

9.2 FORCE-DEFORMATION RESPONSE OF PILE WITH DOWEL-CONNECTION

This section presents development of formulas for estimating displacement capacity of piles with dowel connections to the deck. For this purpose let us define two dimensionless constants, η and β as

$$\eta = \frac{M_{y,P}}{M_{y,C}} \tag{9.4}$$

$$\beta = \frac{EI}{k_\theta L} \tag{9.5}$$

in which η is the ratio of yield moment of the pile and the connection, and β is indicative of the relative rotational stiffness of the pile and the connection.

9.2.1 Response at First Yielding in Connection

To compute the rotation and deflection at the top of the hollow steel pile with a concrete-plug in the initial elastic region, i.e., $\Delta \le \Delta_{y,C}$, consider the cantilever with a moment equal to $k_\theta \theta$ and a lateral force equal to F at the top (Figure 9.5(a)) with a bending moment diagram (Figure 9.5(b)) and the curvature diagram (Figure 9.5(c)). Using the moment-area method for structural

analysis, the rotation and deflection at the top of the pile are given by

$$\theta = \frac{FL^2}{2EI} - \frac{k_\theta L\theta}{EI} = \frac{FL^2}{2EI} - \frac{\theta}{\beta} \tag{9.6}$$

and

$$\Delta = \frac{FL^3}{3EI} - \frac{k_\theta \theta L^2}{2EI} = \frac{FL^3}{3EI} - \frac{\theta L}{2\beta} \tag{9.7}$$

Equation (9.6) can be further simplified to obtain the rotation as

$$\theta = \left(\frac{FL^2}{2EI}\right)\left(\frac{\beta}{1+\beta}\right) \tag{9.8}$$

Utilizing Equation (9.8), Equation (9.7) can also be simplified to obtain the deflection as

$$\Delta = \left(\frac{FL^3}{12EI}\right)\left(\frac{1+4\beta}{1+\beta}\right) \tag{9.9}$$

The first yielding in the pushover curve (Figure 9.4) occurs at the yielding of the connection at yield rotation at the top of the pile equal to

$$\theta_{y,C} = \frac{M_{y,C}}{k_\theta} \tag{9.10}$$

Inserting Equation (9.10) in Equation (9.8) gives the lateral force at the yield level as

$$F_{y,C} = \frac{2M_{y,C}}{L}(1+\beta) \tag{9.11}$$

and utilizing Equation (9.11) in Equation (9.9) gives the yield displacement as

$$\Delta_{y,C} = \frac{M_{y,C}L^2}{6EI}(1+4\beta) = \frac{k_\theta \theta_{y,C} L^2}{6EI}(1+4\beta) = \theta_{y,C} L\left(\frac{1+4\beta}{6\beta}\right) \tag{9.12}$$

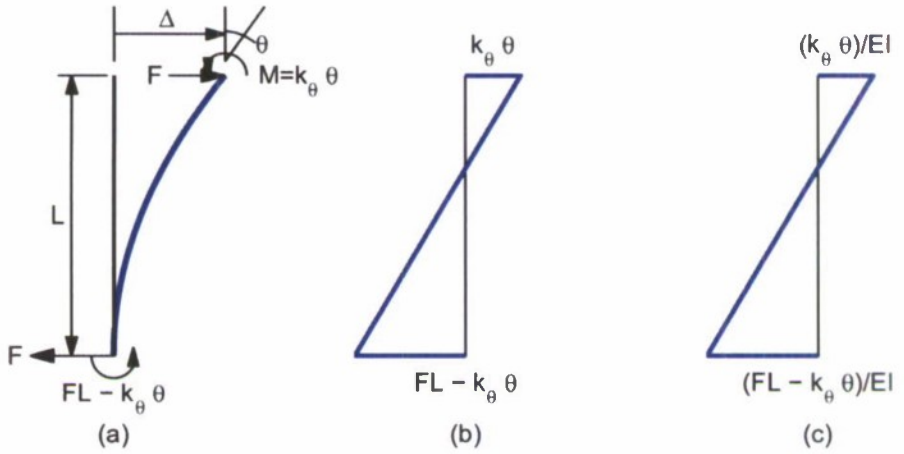


Figure 9.5. Response behavior of a pile with dowel-connection up to yielding in the connection: (a) forces; (b) bending moment diagram; and (c) curvature diagram.

9.2.2 Response at First Yielding in Pile

The response in the range $\Delta_{y,C} \le \Delta \le \Delta_{y,P}$ may be computed by an incremental approach in which the system may be treated as a cantilever fixed at the base and free at the top (Figure 9.6). For this system, the incremental displacement and rotation at the top are given by

$$\left(\Delta - \Delta_{y,C}\right) = \frac{L^3}{3EI}\left(F - F_{y,C}\right) \tag{9.13}$$

$$\left(\theta - \theta_{y,C}\right) = \frac{L^2}{2EI}\left(F - F_{y,C}\right) \tag{9.14}$$

which leads to the expression for the total displacement and rotation as

$$\Delta = \Delta_{y,C} + \frac{L^3}{3EI}\left(F - F_{y,C}\right) = \frac{M_{y,C}L^2}{6EI}\left(1 + 4\beta\right) + \frac{L^3}{3EI}\left(F - F_{y,C}\right) \tag{9.15}$$

$$\theta = \theta_{y,C} + \frac{L^2}{2EI}\left(F - F_{y,C}\right) = \frac{M_{y,C}}{k_\theta} + \frac{L^2}{2EI}\left(F - F_{y,C}\right) \tag{9.16}$$

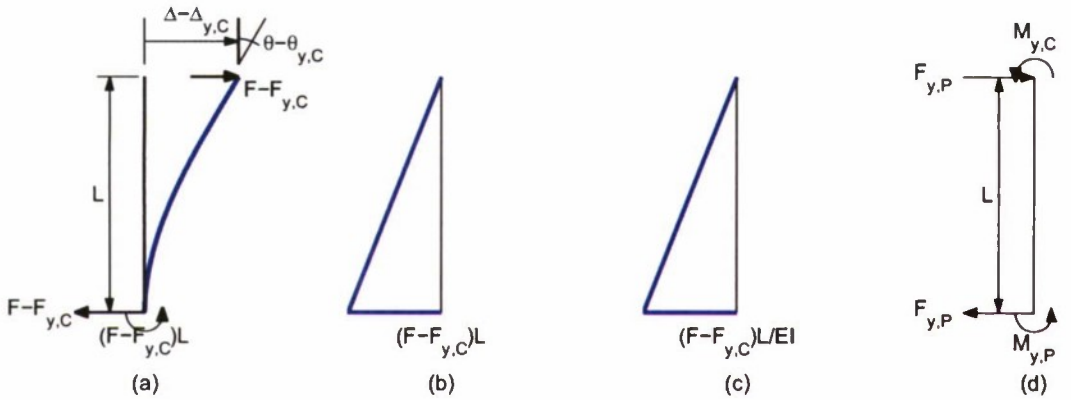


Figure 9.6. Response behavior of a pile with dowel-connection between yielding in the connection and yielding in the pile: (a) forces; (b) bending moment diagram; (c) curvature diagram; and (d) equilibrium at pile yielding.

The lateral force when the pile yields can be computed from the equilibrium of the cantilever (Figure 9.6(d)) as

$$F_{y,P} = \frac{M_{y,C} + M_{y,P}}{L} \tag{9.17}$$

Utilizing Equation (9.17) in Equations (9.15) and (9.16) leads to displacement and rotation at yielding of the pile as

$$\begin{aligned}\Delta_{y,P} &= \frac{M_{y,C}L^2}{6EI}(1+4\beta)+\frac{L^3}{3EI}\left(F_{y,P}-F_{y,C}\right)\\ &= \frac{M_{y,C}L^2}{6EI}(1+4\beta)+\frac{M_{y,P}L^2}{3EI}+\frac{M_{y,C}L^2}{3EI}-\frac{2M_{y,C}}{L}(1+\beta)\frac{L^3}{3EI}\\ &= \left(\frac{M_{y,P}L^2}{3EI}\right)\left(\frac{2\eta-1}{2\eta}\right)\end{aligned} \tag{9.18}$$

$$\begin{aligned}\theta_{y,P} &= \frac{M_{y,C}}{k_\theta}+\frac{L^2}{2EI}\left(F_{y,P}-F_{y,C}\right)\\ &= \frac{M_{y,C}}{k_\theta}+\frac{M_{y,P}L}{2EI}+\frac{M_{y,C}L}{2EI}-\frac{2M_{y,C}}{L}(1+\beta)\frac{L^2}{2EI}\\ &= \left(\frac{M_{y,P}L}{2EI}\right)\left(\frac{\eta-1}{\eta}\right)\end{aligned} \tag{9.19}$$

9.3 DISPLACEMENT DUCTILITY CAPACITY OF PILE

This section develops the formulas for computing displacement ductility capacity of piles with a

partial-moment connection. Presented first are the formulas for the ductility controlled by material strain limits in the connection. Subsequently, formulas for the ductility controlled by material strains in the pile section are presented. The displacement ductility capacity is then defined as the lower of the two ductility values. Finally, a step-by-step summary to compute the displacement ductility capacity of piles with partial-moment connection is presented.

9.3.1 Strain Limits in the Connection

Let θ_L be the rotation in the connection spring for a selected design level, i.e., specified value of strain in the outermost dowel for a selected design level. For the pile-connection system, this rotation may occur either prior to pile yielding, i.e., $\theta_{y,C} < \theta_L < \theta_{y,P}$, or after pile yielding, i.e., $\theta_L > \theta_{y,P}$. The connection rotation ductility at onset of pile yielding is given by

$$\begin{aligned} \mu_{\theta,P} &= \frac{\theta_{y,P}}{\theta_{y,C}} = \frac{k_\theta}{M_{y,C}}\left(\frac{M_{y,P}L}{2EI}\right)\left(\frac{\eta-1}{\eta}\right) \\ &= \left(\frac{M_{y,P}}{M_{y,C}}\right)\left(\frac{k_\theta L}{2EI}\right)\left(\frac{\eta-1}{\eta}\right) \\ &= \frac{\eta-1}{2\beta} \end{aligned} \tag{9.20}$$

The displacement capacity of the pile-connection system when considering strain limits in the outermost dowel of the connection depends on whether the pile remains elastic or the pile yields when the dowel strain limit is reached. Note that the pile would remain elastic if μ_θ is less than $\mu_{\theta,P}$ as given by Equation (9.20). If the pile remains elastic, the rotation in the plug at a selected design level, θ_L, is related to the lateral force F by Equation (9.16) as

$$\theta_L = \theta_{y,C} + \frac{L^2}{2EI}\left(F - F_{y,C}\right) \tag{9.21}$$

which gives

$$\left(F - F_{y,C}\right) = \left(\theta_L - \theta_{y,C}\right)\left(\frac{2EI}{L^2}\right) = \theta_{y,C}\left(\mu_\theta - 1\right)\left(\frac{2EI}{L^2}\right) \tag{9.22}$$

Using Equation (9.15), the displacement is then given as

$$\begin{aligned}\Delta_L &= \frac{M_{y,C}L^2}{6EI}(1+4\beta)+\frac{L^3}{3EI}\theta_{y,C}(\mu_\theta-1)\left(\frac{2EI}{L^2}\right)\\ &= \frac{\theta_{y,C}k_\theta L^2}{6EI}(1+4\beta)+\frac{2L}{3}\theta_{y,C}(\mu_\theta-1)\\ &= \theta_{y,C}L\left[\frac{1+4\beta}{6\beta}+\frac{2(\mu_\theta-1)}{3}\right]\end{aligned} \tag{9.23}$$

The displacement ductility capacity is then defined as

$$\begin{aligned}\mu_\Delta = \frac{\Delta_L}{\Delta_{y,C}} &= \frac{\theta_{y,C}L\left[\frac{1+4\beta}{6\beta}+\frac{2(\mu_\theta-1)}{3}\right]}{\theta_{y,C}L\left(\frac{1+4\beta}{6\beta}\right)} = 1+(\mu_\theta-1)\left(\frac{4\beta}{1+4\beta}\right)\\ &= \frac{1+4\beta\mu_\theta}{1+4\beta}\end{aligned} \tag{9.24}$$

If the pile yields prior to the connection reaching θ_L, i.e., if μ_θ is more than $\mu_{\theta,P}$, the deflection at the pile top can be approximated as

$$\Delta_L = \Delta_{y,P}+(\theta_L-\theta_{y,P})L \tag{9.25}$$

which can be re-written as

$$\begin{aligned}\Delta_L &= \Delta_{y,P}+\theta_{y,C}L\left(\mu_\theta-\frac{\theta_{y,P}}{\theta_{y,C}}\right)\\ &= \Delta_{y,C}+\frac{M_{y,P}L^2}{3EI}+\frac{M_{y,C}L^2}{3EI}-\frac{2M_{y,C}}{L}(1+\beta)\frac{L^3}{3EI}+\theta_{y,C}L\left(\mu_\theta-\frac{\eta-1}{2\beta}\right)\end{aligned} \tag{9.26}$$

The displacement ductility capacity is then defined as

$$\begin{aligned}\mu_\Delta = \frac{\Delta_L}{\Delta_{y,C}} &= 1+\frac{1}{\frac{M_{y,C}L^2}{6EI}(1+4\beta)}\left[\begin{array}{l}\frac{M_{y,P}L^2}{3EI}+\frac{M_{y,C}L^2}{3EI}-\frac{2M_{y,C}}{L}(1+\beta)\frac{L^3}{3EI}\\ +\theta_{y,C}L\left(\mu_\theta-\frac{\eta-1}{2\beta}\right)\end{array}\right]\\ &= \frac{2-\eta+6\beta\mu_\theta}{1+4\beta}\end{aligned} \tag{9.27}$$

The displacement ductility capacity of the pile-concrete-plug system can be summarized as

$$\mu_{\Delta} = \begin{cases} \dfrac{1+4\beta\mu_{\theta}}{1+4\beta} \text{ for } \mu_{\theta} \le \dfrac{\eta-1}{2\beta} \\ \dfrac{2-\eta+6\beta\mu_{\theta}}{1+4\beta} \text{ for } \mu_{\theta} > \dfrac{\eta-1}{2\beta} \end{cases} \tag{9.28}$$

Equation (9.28) applies only for displacement ductility capacity when the strain in the outermost fiber of the dowel in the connection reaches the strain limit for a selected design level.

9.3.2 Strain Limits in the Pile

The preceding section developed the expression for displacement ductility capacity of the pile-connection system controlled by the strain limit in the dowel of the connection. However, it is possible that the strain limit in the pile may occur prior to the system reaching the displacement-ductility capacity given by Equation (9.28). Therefore, the relationship for displacement-ductility of the pile-connection system at strain limits in the pile is developed next.

Let us consider the equilibrium of the pile when the strain limit reaches the limiting value at a selected design level (Figure 9.7). The moment at the top of the pile is equal to $M_{y,C}$ and at the bottom is equal to $M_{y,P}$. The length L_2 is then given by

$$L_2 = \frac{\eta}{1+\eta} L \tag{9.29}$$

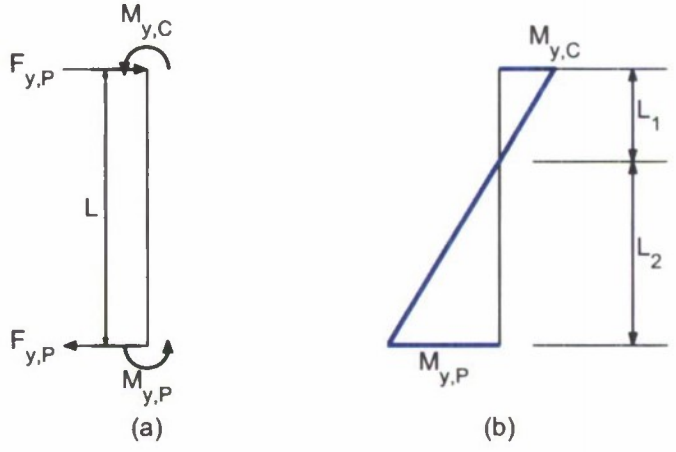


Figure 9.7. Equilibrium of the pile when strain reaches the limiting value in the pile-hinge.

Defining the plastic hinge length as

$$L_p = \rho L_2 \tag{9.30}$$

in which ρ is the length of the plastic hinge as a fraction of the "effective" length defined as the distance from the critical section to the point of contra-flexure ($= L_2$ for this case). Using Equation (9.29) in Equation (9.30) gives a plastic hinge length normalized by the total pile length as

$$L_p^* = \frac{L_p}{L} = \frac{\rho\eta}{1+\eta} \tag{9.31}$$

Using concepts similar to those developed previously for piles with perfect moment connection [see Figure 4.1 and Equation (4.4)], the displacement capacity of the pile is given by

$$\begin{aligned} \Delta_L &= \Delta_{y,\mathrm{P}} + \left(L - \frac{L_p}{2}\right)\left[L_p\left(\phi_L - \phi_y\right)\right] \\ &= \Delta_{y,\mathrm{P}} + \phi_y L^2\left(1 - \frac{L_p^*}{2}\right)\left(L_p^*\right)\left(\mu_\phi - 1\right) \\ &= \Delta_{y,\mathrm{P}} + \left(\frac{M_{y,\mathrm{P}}L^2}{EI}\right)\left(1 - \frac{L_p^*}{2}\right)\left(L_p^*\right)\left(\mu_\phi - 1\right) \end{aligned} \tag{9.32}$$

Dividing Equation (9.32) by the yield displacement given by Equation (9.12), the displacement-ductility capacity is given by

$$\begin{aligned} \mu_\Delta = \frac{\Delta_L}{\Delta_{y,\mathrm{C}}} &= \frac{2\eta - 1}{1+4\beta} + \left(\frac{6\eta L_p^*}{1+4\beta}\right)\left(1 - \frac{L_p^*}{2}\right)\left(\mu_\phi - 1\right) \\ &= \frac{2\eta - 1}{1+4\beta} + \left(\frac{6\eta}{1+4\beta}\right)\left(\frac{\rho\eta}{1+\eta}\right)\left(1 - \frac{\rho\eta}{2(1+\eta)}\right)\left(\mu_\phi - 1\right) \end{aligned} \tag{9.33}$$

Equation (9.33) applies only to the displacement ductility capacity when the material strain in the pile reaches the strain limit for a selected design level. i.e., hinging in the pile.

9.4 STEP-BY-STEP SUMMARY

The following is a step-by-step summary of the procedure to compute the displacement capacity of hollow steel piles with concrete-plug connections or prestressed concrete piles with a dowel

connection at the deek.

1. Establish the axial load, P, on the pile.
2. Estimate the pile length based on an equivalent-fixity assumption.
3. Select an appropriate design level – Level 1 or Level 2 – and establish various strain limits for the selected design level.
4. Devclop the moment-rotation relationship of the conerete-plug eonneetion for a hollow steel pile or the dowel eonneetion for a prestressed eonerete pile using the procedure described in Chapter 8 (Section 8.2) of this report.
5. Determine rotational stiffness, k_θ, yield moment, $M_{y,C}$, and yield rotation, $\theta_{y,C}$ of the connection from the moment-rotation relationship developed in Step 4.
6. Establish the rotation of the plug, θ_L, and corresponding ductility, $\mu_\theta = \theta_L / \theta_{y,C}$, when strain in the outer-most dowel of the connection reaches the strain limit established in Step 3 for the selected design level.
7. Conduct the moment-curvature analysis of the pile section and idealize the moment-curvature relationship by a bi-linear curve. For this analysis, apply the axial load on the pile prior to moment-curvature analysis.
8. Compute the effective, EI_e, and effective yield moment, $M_{y,P}$, from the pile moment-curvature relationship. Note that EI_e is equal to the initial elastic slope and $M_{y,P}$ is the yield value of the moment of the idealized bi-linear moment-curvature relationship. For steel piles, EI may be computed from section properties and material modulus, and $M_{y,P}$ may be approximated as $M_{y,P} = f_y \left(d_o^3 - d_i^3 \right) / 6$.
9. Estimate the yield curvature, $\phi_{y,P} = M_{y,P} / EI_e$.
10. Establish the curvature of the steel pile, ϕ_L, and corresponding curvature ductility, $\mu_\phi = \phi_L / \phi_{y,P}$, when material strain in the pile section reaches the strain limit established in Step 3 for the selected design level.

11. Select the value of ρ which defines the length of the plastic hinge as a fraction of the "effective: length of the pile. Guidelines for selection values of ρ for hollow steel piles and prestressed concrete piles are provided in subsequent Chapters of this report.

12. Compute the dimensionless parameters: $\eta = M_{y,P}/M_{y,C}$, and $\beta = EI_e/k_\theta L$.

13. Compute the normalized value of the plastic hinge length: $L_P^* = (\rho\eta)/(1+\eta)$.

14. Compute the yield displacement which corresponds to first effective yielding in the connection as: $\Delta_{y,C} = \theta_{y,C} L(1+4\beta)/6\beta$

15. Compute the displacement ductility for yielding in the connection as $\mu_\Delta = (1+4\beta\mu_\theta)/(1+4\beta)$ if μ_θ computed in Step 6 is less than or equal to $(\eta-1)/2\beta$ otherwise $\mu_\Delta = (2-\eta+6\beta\mu_\theta)/(1+4\beta)$.

16. Compute displacement ductility for yielding in the pile as $\mu_\Delta = (2\eta-1)/(1+4\beta) + (6\eta L_p^*)(1-L_p^*/2)(\mu_\phi - 1)/(1+4\beta)$

17. Establish the displacement ductility capacity as the lower of the values computed in Steps 15 and 16.

18. Compute the displacement capacity of the pile as a product of the yield displacement computed in Step14 and the displacement ductility capacity computed in Step 17.

10. DUCTILITY CAPACITY OF HOLLOW STEEL PILES WITH DOWEL-CONNECTION

The displacement ductility capacity of hollow steel piles with a dowel-connection to the deck is investigated in this chapter. For this purpose, two design levels – Level 1 and Level 2 – specified for seismic analysis of Marine Oil and LNG Terminals in the MOTEMS are considered. The strain limits specified in the MOTEMS for reinforcing steel are 0.01 for Level 1 and 0.05 for Level 2 if the hinge were to form in the connection. If the hinge were to form in the steel pile below the ground level, these strain limits are 0.008 for Level 1 and 0.025 for Level 2. Two pile diameters – 61 cm and 91 cm – each with two wall thicknesses – 1.27 cm inch and 2.54 cm – are considered. Furthermore, two configurations of reinforcing details in the concrete-plug connection are considered: 8 dowels and 12 dowels, with area of each dowel being equal to 8.2 cm^2. The piles are considered to be fixed at the bottom to reflect the equivalent-fixity assumption at the bottom. The axial load on the pile is assumed to be $0.05Af_y$ or $0.1Af_y$ in which A is the cross-section area of the pile and f_y is the yield strength of steel. The pile is modeled in computer program OPENSEES (McKenna and Fenves, 2001) using fiber section and nonlinear beam-column elements.

Figures 10.1 to 10.4 present the variation of displacement ductility capacity of hollow-steel piles with concrete-plug connections with pile length for two conditions: (1) formation of hinge in the concrete-plug connection and (2) formation of hinge in the steel pile. These results were generated by nonlinear-finite element analysis of the pile system shown in Figure 9.1. The presented results indicate that the ductility capacity due to pile hinging tends to increase slightly with pile length for shorter piles. For longer piles, however, the ductility capacity is essentially independent of the pile length as apparent from the almost-flat curves. The ductility capacity due to concrete-plug hinging, however, reduces significantly with pile length. However, this ductility capacity may become insensitive to the pile length for longer piles.

The design ductility capacity of hollow steel piles with concrete plugs is lower for the ductility capacities due to hinging in the steel pile and in the concrete plug. Therefore, the results of Figures 10.1 to 10.4 also permit another important observation: hinging in the steel pile may control the design ductility capacity of shorter piles whereas hinging in concrete-plug almost always controls the design ductility capacity of longer piles. As such, for a given pile length, the

lower of the ductility capacity from hinging in the pile and hinging in the connection must be selected as the design ductility capacity.

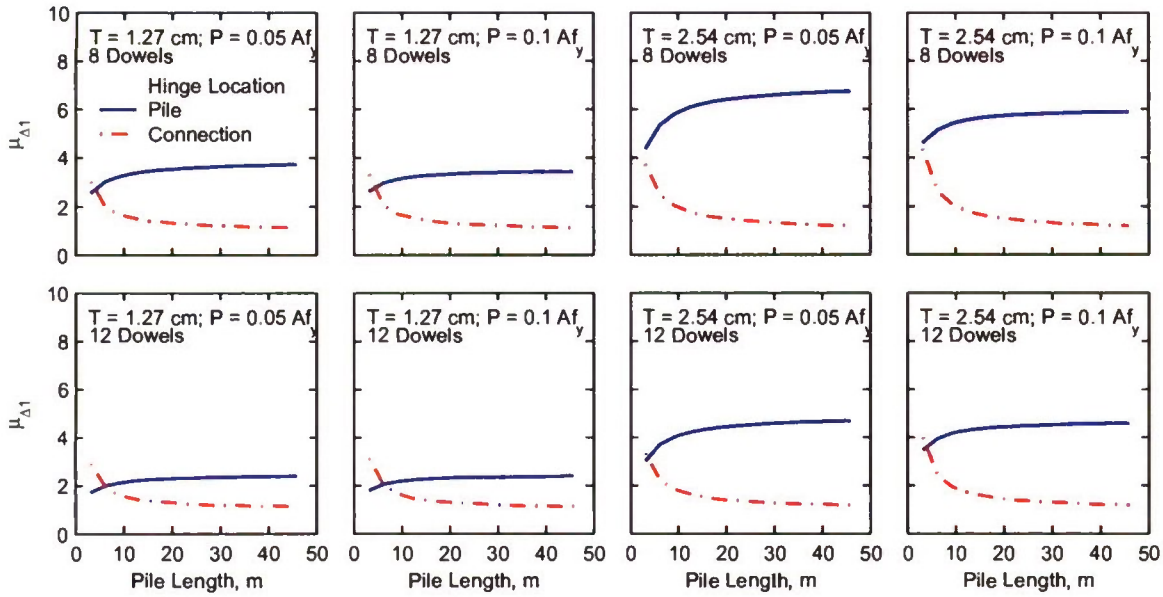


Figure 10.1. Displacement ductility capacity of hollow-steel piles with concrete-plug connections for a design Level 1 earthquake and a 61 cm pile diameter. Variables include axial load, pile thickness and number of dowels.

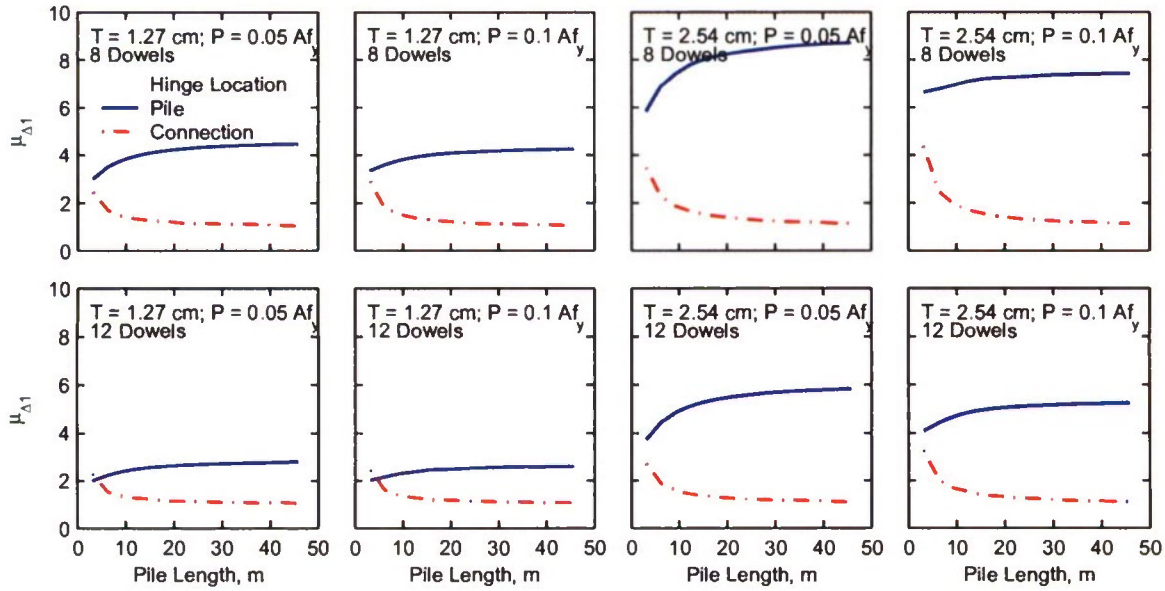


Figure 10.2. Displacement ductility capacity of hollow-steel piles with concrete-plug connections for a design Level 1 earthquake and 91 cm pile diameter. Variables include axial load, pile thickness and number of dowels.

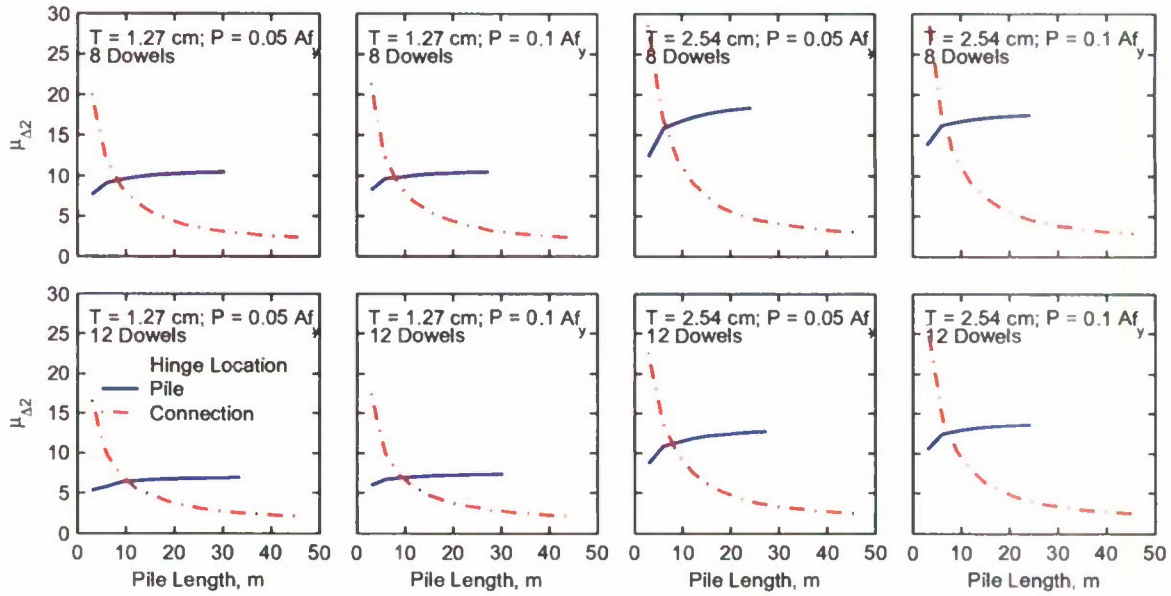


Figure 10.3. Displacement ductility capacity of hollow-steel piles with concrete-plug connections for a design Level 2 earthquake and 61 cm pile diameter. Variables include axial load, pile thickness and number of dowels.

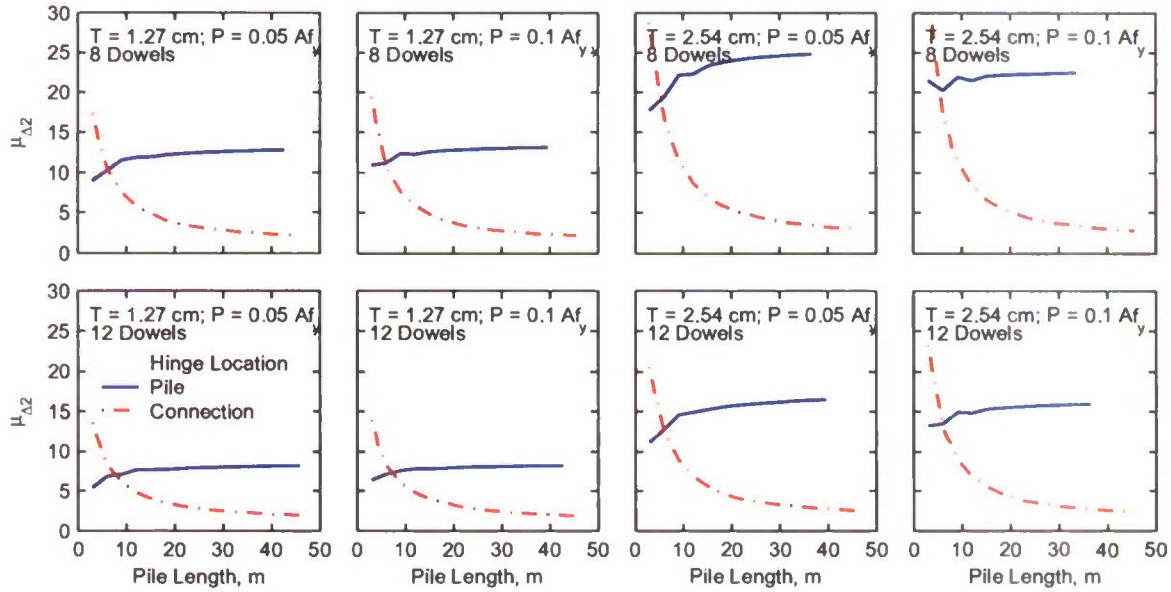


Figure 10.4. Displacement ductility capacity of hollow-steel piles with concrete-plug connections for a design Level 2 earthquake and 91 cm pile diameter. Variables include axial load, pile thickness and number of dowels.

The accuracy of the formulas developed in Chapter 9 are examined next by comparing design ductility capacity from nonlinear finite element analysis (NFEA) with that from Equations (9.29) and (9.33). Note that the results presented are the higher of the ductility values due to

hinging in the pile and the connection. The value of $\rho = 0.03$ for Level 1 earthquake design and $\rho = 0.075$ for Level 2 earthquake design has been selected for hollow steel piles. These values must be used in estimating the ductility capacity from Equation (9.33). The presented results in Figures 10.5 and 10.6 are for a Level 1 earthquake, Figures 10.7 and 10.8 are for a Level 2 earthquake and they indicate that the formulas developed in this investigation provide highly accurate estimates of displacement ductility capacity of hollow steel piles with concrete-plug connection.

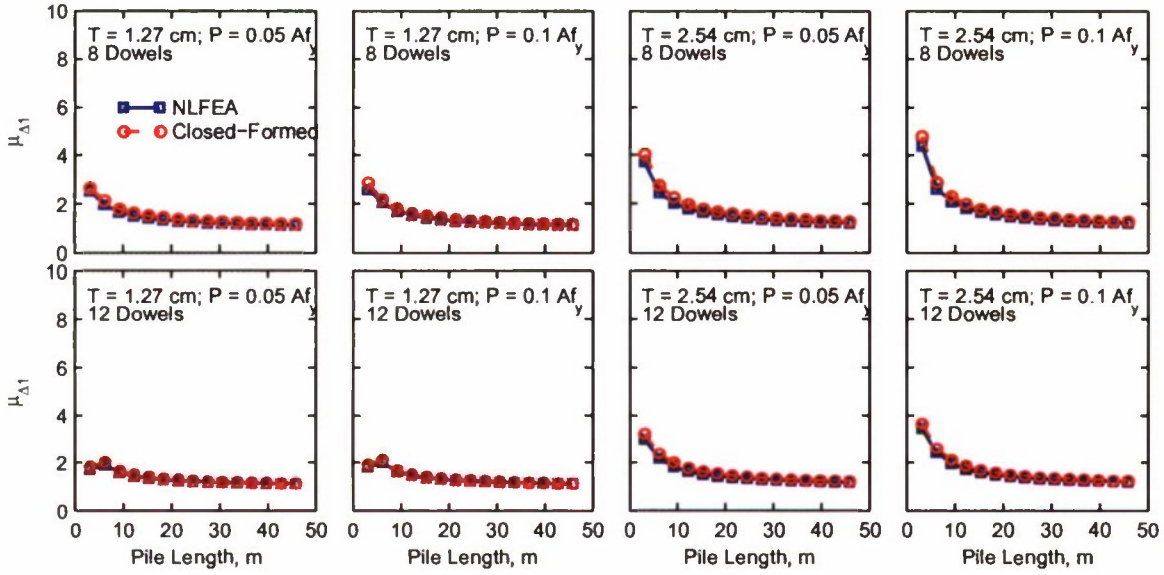


Figure 10.5. Comparison of design displacement ductility capacity of hollow-steel piles with concrete-plug connections from Equations (9.29) and (9.33), and nonlinear finite element analysis (NLFEA); results are for piles with 61 cm diameter and a design Level 1 earthquake. Variables include axial load, pile thickness and number of dowels.

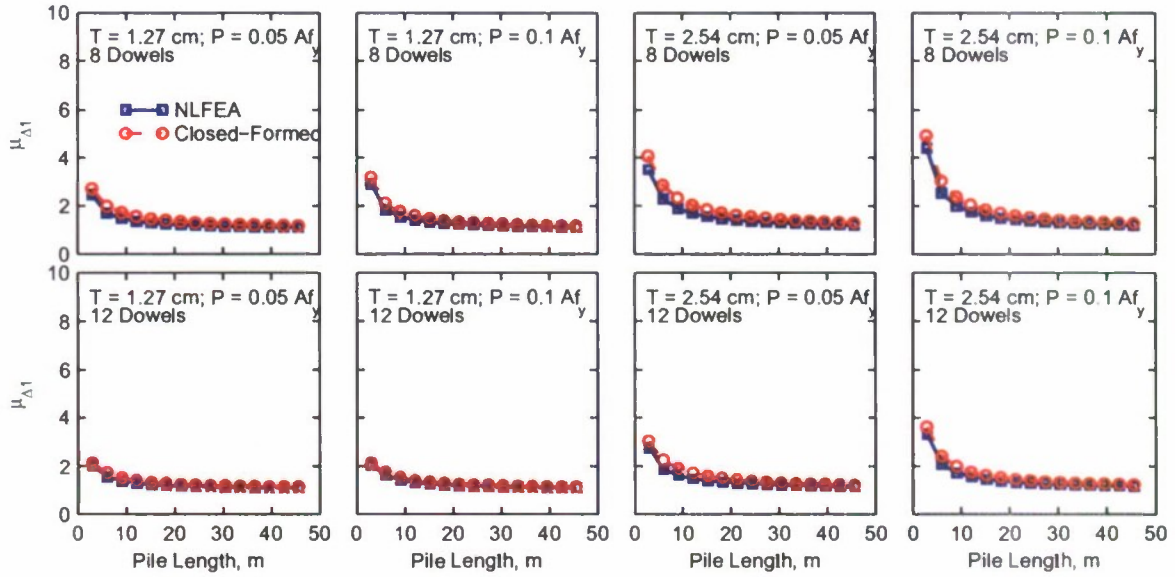


Figure 10.6. Comparison of design displacement ductility capacity of hollow-steel piles with concrete-plug connections from Equations (9.29) and (9.33), and nonlinear finite element analysis (NLFEA); results are for piles with 91 cm diameter and a design Level 1 earthquake. Variables include axial load, pile thickness and number of dowels.

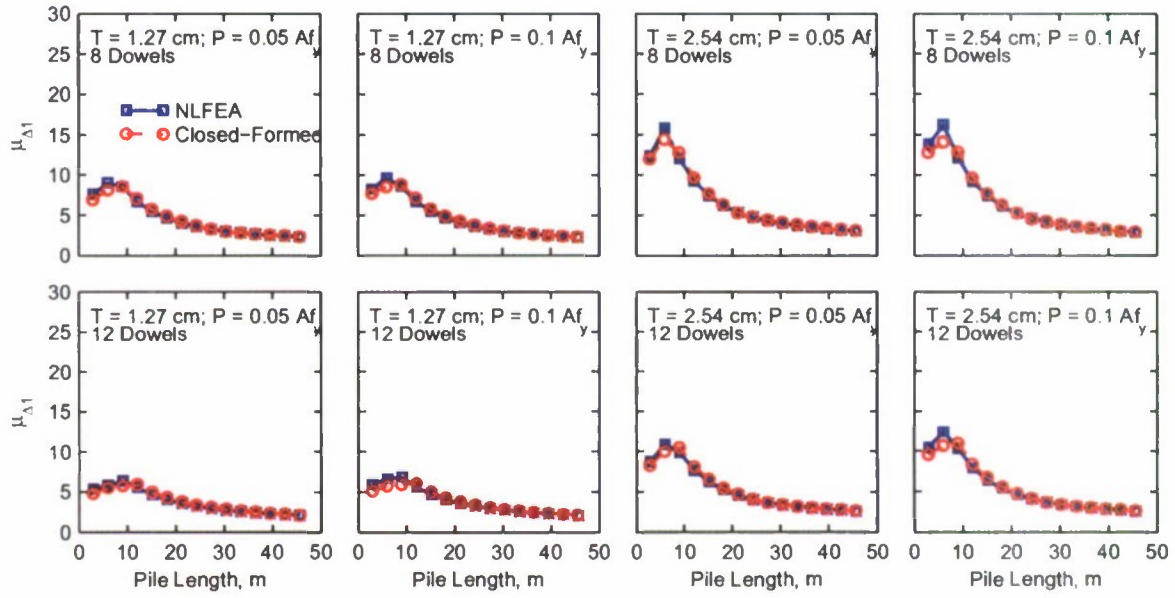


Figure 10.7. Comparison of design displacement ductility capacity of hollow-steel piles with concrete-plug connections from Equations (9.29) and (9.33), and nonlinear finite element analysis (NLFEA); results are for piles with 61 cm diameter and a design Level 2 earthquake. Variables include axial load, pile thickness and number of dowels.

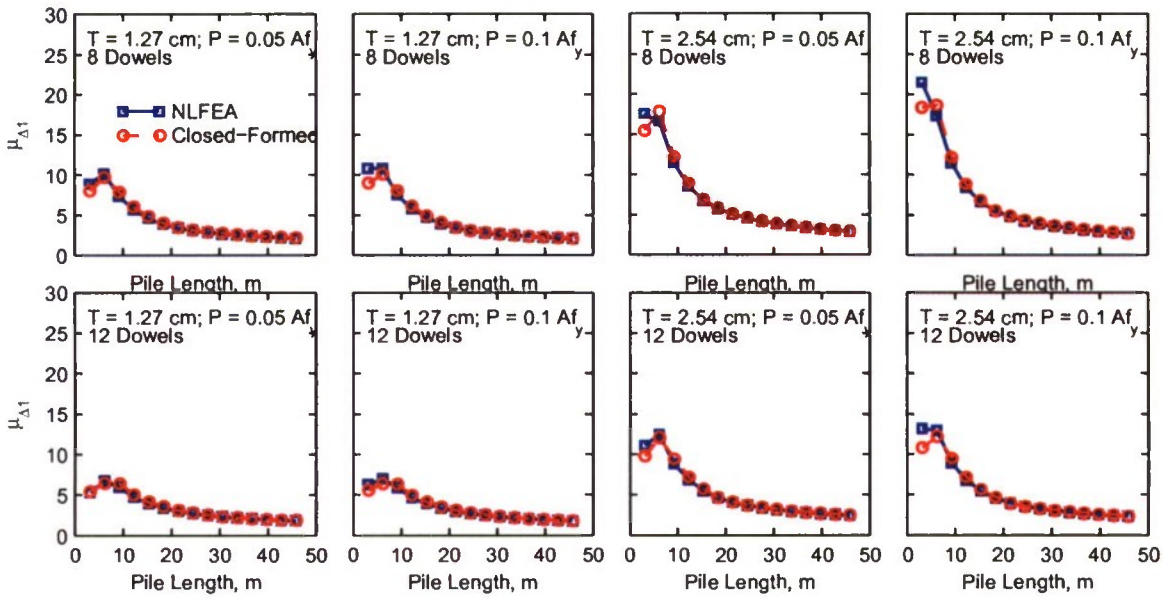


Figure 10.8. Comparison of design displacement ductility capacity of hollow-steel piles with concrete-plug connections from Equations (9.29) and (9.33), and nonlinear finite element analysis (NLFEA); results are for piles with 91 cm diameter and a design Level 2 earthquake. Variables include axial load, pile thickness and number of dowels.

11. DUCTILITY CAPACITY OF PRESTRESSED CONCRETE PILES WITH DOWEL-CONNECTION

The displacement ductility capacity of prestressed concrete piles with dowel connections to the deck is investigated in this chapter. For this purpose, two design levels – Level 1 and Level 2 – specified for seismic analysis of Marine Oil and LNG Terminals are considered. The strain limits specified in the MOTEMS for reinforcing steel are 0.01 for Level 1 and 0.05 for Level 2 if the hinge were to form in the connection. If the hinge were to form in the pile, the strain limits in the prestressing strand are 0.005 (total) for Level 1 and 0.015 (incremental) for Level 2. The results are generated for a pile diameter of 61 cm with 16 prestressing strands. The area of each prestressing strand is equal to 1.4 cm^2, strength is 1884 MPa, and initial prestress in the strands is equal to 70% of its strength. The confinement is provided by #11 spiral wire (area = 0.71 cm^2) with spacing equal to 6.3 cm. The dowel connection consists of 8 bars, each with an area equal to 3.9 cm^2. The piles are considered to be fixed at the bottom to reflect the equivalent-fixity assumption at that point. The axial load on the pile is assumed to be $0.05Af_c^{'}$ in which A is the cross-section area of the pile and $f_c^{'}$ is the compressive strength of concrete. Four values of the

de-bonded length of the bars in the dowel connection are considered: 0 cm, 30 cm, 61 cm, and 91 cm. The pile is modeled in computer program OPENSEES (McKenna and Fenves, 2001) using fiber section and nonlinear beam-column elements.

Figures 11.1 and 11.2 compare the ductility of prestressed concrete piles with a dowel connection due to the formation of a hinge in the pile from nonlinear finite element analysis (NLFEA) and from Equation (9.33). Note that a value of $\rho = 0.05$ has been used for both design levels. These results show that Equation (9.33) provides results for Level 1 design that are almost identical to those from the NLFEA (Figure 10.1). For Level 2 design, Equation (9.33) provides results that are almost identical to those from the NLFEA for longer piles but provide a lower bound for very short piles (Figure 11.2).

Figures 11.3 and 11.4 compare the ductility of prestressed concrete piles with a dowel connection due to the formation of a hinge in the connection from nonlinear finite element analysis (NLFEA) and from Equation (9.29). These results show that Equation (9.29) provides results that are almost identical to those from the NLFEA.

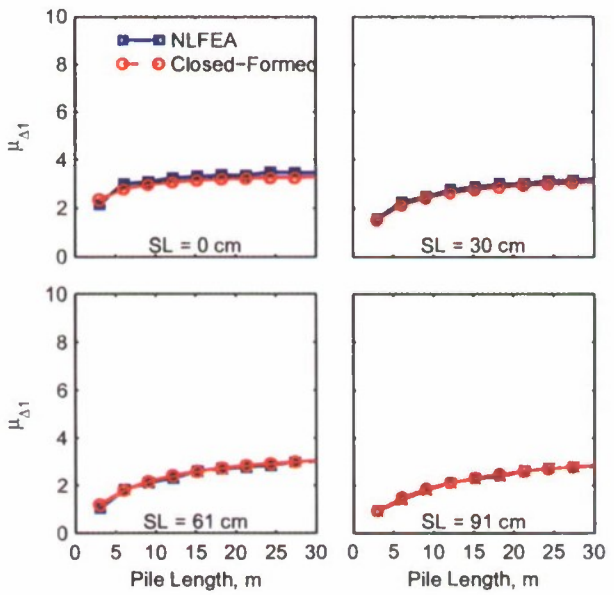


Figure 11.1. Comparison of displacement ductility of prestressed concrete piles with dowel connections due to the formation of a hinge in the piles from nonlinear finite element analysis (NLFEA) and from Equation (9.33). Results are for design Level 1 earthquake. SL = de-bond length.

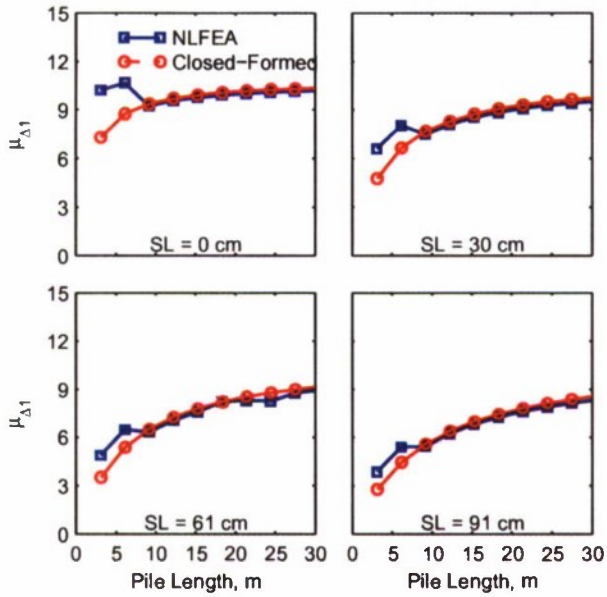


Figure 11.2. Comparison of displacement ductility of prestressed concrete piles with dowel connections due to the formation of a hinge in the piles from nonlinear finite element analysis (NLFEA) and from Equation (9.33): results are for design Level 2 earthquake. SL = de-bond length.

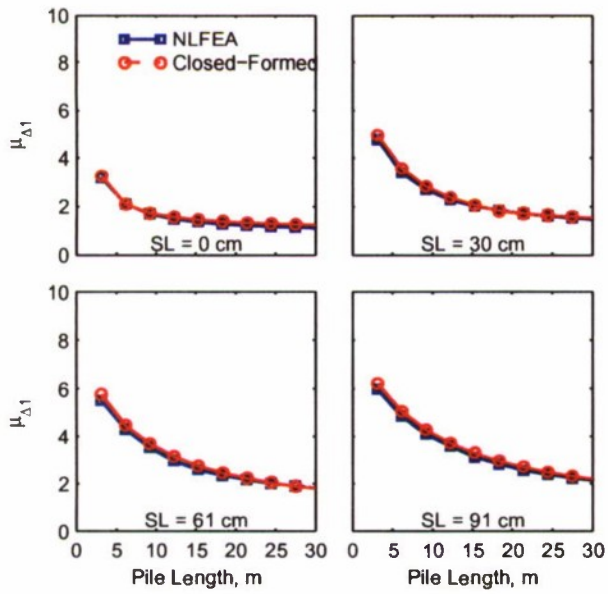


Figure 11.3. Comparison of displacement ductility of prestressed concrete piles with dowel connections due to the formation of a hinge in the connection from nonlinear finite element analysis (NLFEA) and from Equation (9.33). Results are for a design Level 1 earthquake. SL = de-bond length.

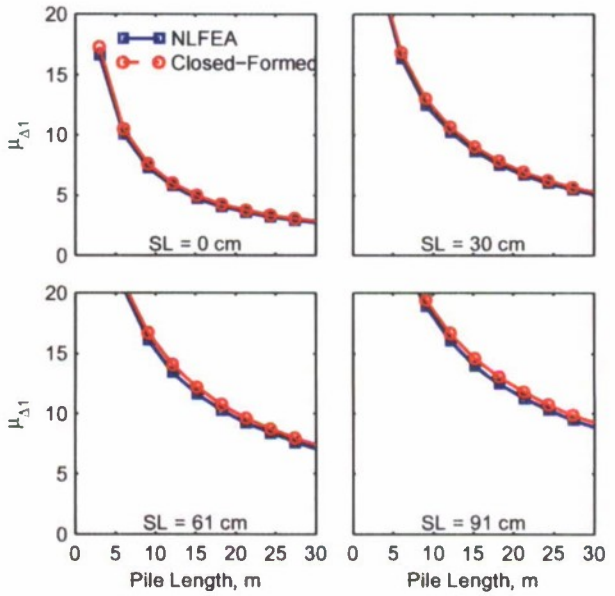


Figure 11.4. Comparison of displacement ductility of prestressed concrete piles with dowel connections due to the formation of a hinge in the connection from nonlinear finite element analysis (NLFEA) and from Equation (9.33). Results are for a design Level 2 earthquake. SL = de-bond length.

Figures 11.5 and 11.6 compare the displacement ductility capacity of prestressed concrete piles with dowel connections, defined as lower of the ductility due to the formation of a hinge in the pile or the formation of a hinge in the connection, from nonlinear finite element analysis (NLFEA) and Equation (9.33). These results show that the formulas developed in this investigation provide results that match well with those from the NLFEA.

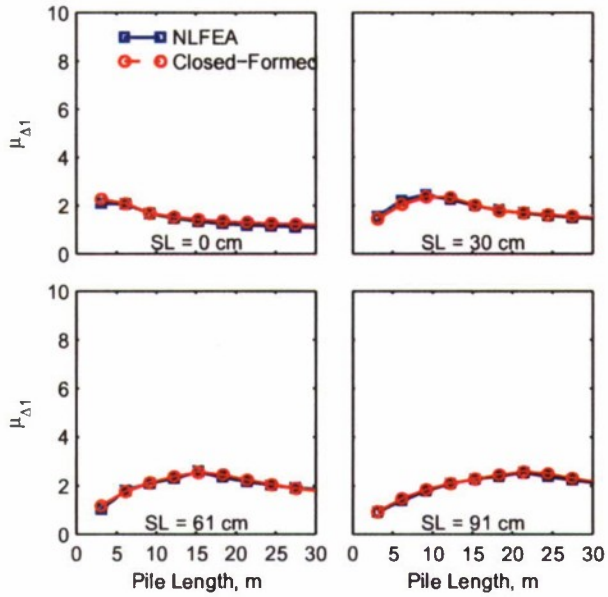


Figure 11.5. Comparison of displacement ductility capacity of prestressed concrete piles with dowel connections, defined as the lower of the ductility due to the formation of a hinge in the pile or the formation of a hinge in the connection, from nonlinear finite element analysis (NLFEA) and from Equation (9.33). Results are for a design Level 1 earthquake. SL = de-bond length.

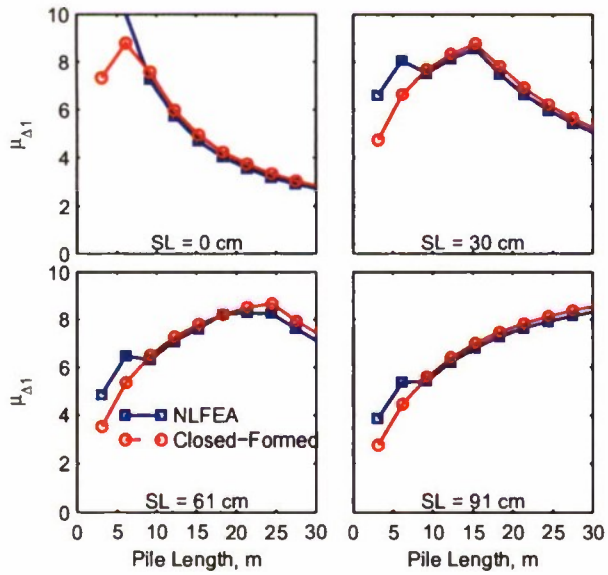


Figure 11.6. Comparison of displacement ductility capacity of prestressed concrete piles with dowel connections, defined as the lower of the ductility due to the formation of a hinge in the pile or the formation of a hinge in the connection, from nonlinear finite element analysis (NLFEA) and from Equation (9.33). Results are for a design Level 2 earthquake. SL= de-bond length.

12. RECOMMENDATIONS

The following simplified procedures are recommended for estimating seismic displacement demand and capacity of pile in Marine Oil and LNG Terminals. The seismic displacement capacity recommended here is consistent with the strain limits specified in the MOTEMS.

12.1 DISPLACEMENT DEMAND

12.1.1 Regular Structures

It is recommended that the seismic displacement demand in a regular structure (MOTEMS 2007) be estimated from the following procedure that was recently proposed in the ASCE/SEI 41-06 standard:

$$\Delta_d = C_1 C_2 S_A \frac{T^2}{4\pi^2} \tag{12.1}$$

in which S_A is the spectral acceleration of the linear-elastic system at vibration period, T. The coefficient C_1 is given by

$$C_1 = \begin{cases} 1.0; & T > 1.0\text{s} \\ 1.0 + \dfrac{R-1}{aT^2}; & 0.2\text{s} < T \le 1.0\text{s} \\ 1.0 + \dfrac{R-1}{0.04a}; & T \le 0.2\text{s} \end{cases} \tag{12.2}$$

in which a is a site dependent constant equal to 130 for Site Class A and B, 90 for Site Class C, and 60 for Site Class D, E, and F (definition of Site Class is available in ASCE/SEI 41-06 standard), and R is the ratio of the elastic and yield strength of the system and is defined as

$$R = \frac{S_A}{g}\frac{W}{V_y} \tag{12.3}$$

where W is the seismic weight of the system, V_y is the yield force (or base shear) of the system, and g is the acceleration due to gravity. The coefficient C_2 is given by

$$C_2 = \begin{cases} 1.0; & T > 0.7\,\text{s} \\ 1 + \dfrac{1}{800}\left(\dfrac{R-1}{T}\right)^2; & T \le 0.7\,\text{s} \end{cases} \tag{12.4}$$

The estimation of displacement demand should be restricted to systems with

$$R \le \frac{\bar{\Delta}_d}{\Delta_y} + \frac{\left|\alpha_e\right|^{-t}}{4} \tag{12.5}$$

in which $\bar{\Delta}_d$ is the smaller of the computed displacement demand, Δ_d, or the displacement corresponding to the maximum strength in the pushover curve, Δ_y is the yield displacement of the idealized bilinear force-deformation curve, $t = 1 + 0.15\ln(T)$, and α_e is the effective post-elastic stiffness ratio computed from

$$\alpha_e = \alpha_{P\text{-}\Delta} + \lambda(\alpha_2 - \alpha_{P\text{-}\Delta}) \tag{12.6}$$

where λ is a near-field effect factor equal to 0.8 for sites that are subjected to near-field effects and 0.2 for sites that are not subjected to near field effects. The near field effects may be considered to exist if the 1 second spectral value, S_1, at the site for the maximum considered earthquake is equal to or exceeds 0.6g. The P-Delta stiffness ratio, $\alpha_{P\text{-}\Delta}$, and the maximum negative post-elastic stiffness ratio, α_2, are estimated from the idealized force-deformation curve.

12.1.2 Irregular Structures

The following modal pushover analysis (MPA) procedure is recommended to estimate displacement demands in irregular Marine Oil and LNG Terminal structures (MOTEMS 2007):

1. Compute the natural frequencies, ω_n and modes, $\boldsymbol{\phi}_n$, for linearly elastic vibration of the irregular Marine Oil and LNG Terminal structure.
2. Select a reference point where the displacement, u_m, is to be monitored in the selected direction of analysis during the pushover analysis. Ideally, this reference point should be the location on the structure with largest value of ϕ_{rn} in the selected direction of analysis.

3. For the nth-mode, develop the pushover curve, $V_{bn} - u_{rn}$, for the nth modal force distribution, $s_n^* = \mathbf{M}\boldsymbol{\phi}_n$, where $\mathbf{M}$ is the mass matrix of the structure, and $\boldsymbol{\phi}_n$ is the nth mode shape. The base shear V_{bn} should be monitored in the same direction as the direction of the selected reference point displacement u_{rn}.

4. Convert the $V_{bn} - u_{rn}$ pushover curve to the force-displacement, $F_{sn}/L_n - D_n$, relation for the nth -"mode" inelastic SDF system by utilizing $F_{sn}/L_n = V_{bn}/M_n^*$ and $D_n = u_{rn}/\Gamma_n\phi_{rn}$ in which ϕ_{rn} is the value of $\boldsymbol{\phi}_n$ at the reference point in the direction under consideration, $M_n^* = \left(\boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\iota}\right)^2 / \boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\phi}_n$ is the effective modal mass, and $\Gamma_n = \boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\iota} / \boldsymbol{\phi}_n^T \mathbf{M}\boldsymbol{\phi}_n$ with $\boldsymbol{\iota}$ equal to the influence vector. The influence vector $\boldsymbol{\iota}$ is a vector of size equal to the total number of degrees of freedom. For analysis in the x-direction, the components of $\boldsymbol{\iota}$ corresponding to x-degree-of-freedom are equal to one and remaining components equal to zero. Similarly the components of $\boldsymbol{\iota}$ corresponding to y-degree-of-freedom are equal to one and remaining components equal to zero for analysis in the y-direction.

5. Idealize the force-displacement, $F_{sn}/L_n - D_n$, curve as a bilinear curve and compute the yield value F_{sny}/L_n.

6. Compute the yield strength reduction factor, $R = S_A / \left(F_{sny}/L_n\right)$.

7. Compute the peak deformation $D_n = \Delta_d$ of the nth-"mode" inelastic SDF system defined by the force-deformation relation developed in Step 4 and damping ratio ζ_n, from Equation (1). The elastic vibration period of the system is based on the effective slope of the $F_{sn}/L_n - D_n$ curve, which for a bilinear curve is given by $T_n = 2\pi\left(L_n D_{ny}/F_{sny}\right)^{1/2}$.

8. Calculate peak reference point displacement u_{rn} associated with the nth-"mode" inelastic SDF system from $u_{rn} = \Gamma_n \phi_{rn} D_n$.

9. Push the structure to the reference point displacement equal to u_{rn} and note the values of desired displacement δ_{no}.

10. Repeat Steps 3 to 9 for all significant modes identified.

11. Combine the peak modal displacement, δ_{no}, by an appropriate modal combination rule, e.g., CQC, to obtain the peak dynamic response, Δ_o.

12.2 DISPLACEMENT CAPACITY

It is recommended that the displacement capacity of piles in Marine Oil and LNG Terminals be estimated from

$$\Delta_c = \mu_\Delta \Delta_y \tag{12.7}$$

where Δ_y is the yield displacement of the pile and μ_Δ is the displacement ductility capacity of the pile.

The procedure to estimate the displacement capacity is intended to be a simplified procedure for either initial design of piles or for checking results from more complex nonlinear finite element analysis. The recommendations presented here are limited to: (1) piles with long freestanding heights (length/diameter > 20) above the mud line; (2) piles with transverse volumetric ratio greater than 0.5%; and (3) piles in which the displacement demand has been estimated utilizing equivalent-fixity approximation. Results form this investigation should be used with caution for parameters or cases outside of those described above.

12.2.1 Piles with Full-Moment- or Pin-Connection to the Deck Slab

The recommended values of displacement ductility capacity of piles with full-moment-connection or pin-connection to the deck slab are

Design Earthquake Level	Hinge Location	Reinforced-Concrete Piles	Hollow-Steel Piles
Level 1	In-Ground	1.75	1.2
	Pile-Deck	1.75	1.2
Level 2	In-Ground	2.5	2.75
	Pile-Deck	5.0	2.75

The yield displacement of the pile may be estimated either from idealized pushover curve developed from the nonlinear static pushover analysis or may be estimated from

$$\Delta_y = \begin{cases} \dfrac{M_y L^2}{6EI_e} & \text{for full-moment-connection} \\ \dfrac{M_y L^2}{3EI_e} & \text{for pin-connection} \end{cases} \tag{12.8}$$

in which M_y is the section yield moment and EI_e is the effective value of EI. Note that M_y is not the section moment at first-yield but the effective yield moment estimated from bilinear idealization of the moment-curvature relationship.

12.2.2 Piles with Dowel-Connection to the Deck Slab

The following procedure is recommended for estimating displacement ductility capacity of piles with dowel-connection, such as hollow-steel piles or prestressed concrete piles connected to the deck slab with dowels:

1. Establish the axial load, P, on the pile.
2. Estimate the pile length based on equivalent-fixity assumption.
3. Select an appropriate design level – Level 1 or Level 2 – and establish various strain limits for the selected design level.
4. Develop the moment-rotation relationship of the dowel-connection using the procedure described in Chapter 8 of this report.
5. Determine rotational stiffness, k_θ, yield moment, $M_{y,C}$, and yield rotation, $\theta_{y,C}$ of the dowel-connection from the moment-rotation relationship developed in Step 4.
6. Establish the rotation of the dowel-connection, θ_L, and corresponding ductility, $\mu_\theta = \theta_L / \theta_{y,C}$, when strain in the outer-most dowel of the connection reaches the strain limit established in Step 3 for the selected design level.
7. Conduct the moment-curvature analysis of the pile section with appropriate axial load and idealize the moment-curvature relationship by a bi-linear curve.

8. Compute the effective, EI_e, and effective yield moment, $M_{y,P}$, from the pile moment-curvature relationship. Note that EI_e is equal to initial elastic slope and $M_{y,P}$ is the yield value of the moment of the idealized bi-linear moment-curvature relationship. For steel piles, EI_e may be computed from section properties and material modulus, and $M_{y,P}$ may be approximated as $M_{y,P} \simeq f_y\left(d_o^3 - d_i^3\right)/6$.

9. Estimate the yield curvature, $\phi_{y,P} = M_{y,P}/\mathrm{EI}_e$.

10. Establish the curvature of the steel pile, ϕ_L, and corresponding curvature ductility, $\mu_\phi = \phi_L/\phi_{y,P}$, when material strain in the pile section reaches the strain limit established in Step 3 for the selected design level.

11. Select the value of ρ which defines the length of the plastic hinge as a fraction of the "effective" length of the pile. The recommended value for hollow-steel piles with dowel-connection is $\rho = 0.03$ for Level 1 design and $\rho = 0.075$ for Level 2; and for prestressed concrete pile with dowel-connection for both design levels is $\rho = 0.05$.

12. Compute the dimensionless parameters: $\eta = M_{y,P}/M_{y,C}$, and $\beta = EI_e/k_\theta L$.

13. Compute the normalized value of the plastic hinge length: $L_P^* = (\rho\eta)/(1+\eta)$.

14. Compute the yield displacement which corresponds to first effective yielding in the connection as: $\Delta_{y,C} = \theta_{y,C} L(1+4\beta)/6\beta$

15. Compute the displacement ductility for yielding in the connection as $\mu_\Delta = (1+4\beta\mu_\theta)/(1+4\beta)$ if μ_θ computed in Step 6 is less than or equal to $(\eta-1)/2\beta$ otherwise $\mu_\Delta = (2-\eta+6\beta\mu_\theta)/(1+4\beta)$.

16. Compute displacement ductility for yielding in the pile as $\mu_\Delta = (2\eta-1)/(1+4\beta) + (6\eta L_p^*)(1-L_p^*/2)(\mu_\phi - 1)/(1+4\beta)$

17. Establish the displacement ductility capacity as lower of the values computed in Steps 15 and 16.

18. Compute the displacement capacity of the pile as product of the yield displacement computed in Step14 and the displacement ductility capacity computed in Step 17.

12.3 RECOMMENDATIONS FOR FUTURE WORK

The recommendations developed in this report are based on analytical simulations of piles fixed at the base at a depth equal to depth-of-fixity below the mud line. In order to develop further confidence in these results, the following recommendations are provided for future work:

1. Verify the displacement ductility values for various seismic design levels from laboratory experiments conducted at displacement values appropriate for these design levels.
2. Verify the recommended values of plastic hinge lengths for various seismic design levels from laboratory experiments conducted at displacement values appropriate for these design levels.
3. Compare displacement capacity estimated from equivalent fixity model with those estimated from analysis of pile-soil system.
4. The validity of using the same plastic hinge length for all seismic design levels in the concentrated plasticity model of hollow steel piles and possibly for reinforced-concrete piles should be verified.

REFERENCES

ASCE (2000). Pre-standard and commentary for the seismic rehabilitation of buildings. *FEMA-356*, Federal Emergency Management Agency, Washington, D.C.

ASCE (2005). Minimum design loads for buildings and other structures, *ASCE Standard ASCE/SEI 7-05*, American Society of Civil Engineers, Reston, VA.

ASCE (2007). Seismic rehabilitation and evaluation of existing buildings: 2-Book Set, *ASCE Standard No. ASCE/SEI 41-06 and ASCE Standard No. ASCE/SEI 31-03*, American Society of Civil Engineers, Reston, VA.

ATC (2005). Improvement of nonlinear static seismic analysis procedures. *FEMA-440*, Federal Emergency Management Agency, Washington, D.C.

BSSC (2003). NEHRP recommended provisions for seismic regulations for new buildings and other structures. *FEMA-450*, Federal Emergency Management Agency, Washington, D.C.

Budek, A.M., Priestley, M.J.N., and Benzoni, G. (2000). "Inelastic seismic response of bridge drilled-shaft RC pile/columns," *Journal of Structural Engineering*, 126(4):510-517.

Chai, Y.H. (2002). "Flexural strength and ductility of extended pile-shafts. I: Analytical model," *Journal of Structural Engineering*, 128(5):586-594.

Chai, Y.H. and Hutchinson, T.C. (2002). "Flexural strength and ductility of extended pile-shafts. II: Experimental study," *Journal of Structural Engineering*, 128(5):595-602.

Chopra, A.K., and Goel, R.K. (2004). "A modal pushover analysis procedure to estimate seismic demands for unsymmetric-plan buildings," *Earthquake Engineering and Structural Dynamics*, 33:903-927.

Chopra, A.K. and Chintanapakdee, C. (2004). "Inelastic deformation ratios for design and evaluation of structures: Single-degree-of-freedom bilinear systems." *Journal of Structural Engineering*, 130(9), 1309-1319.

Department of Defense (2005). Uniform Facilities Criteria – Design: Piers and Wharves, *Report No. UFC 4-152-01*, July 28.

Eskijian, M. (2007). "Marine Oil Terminal Engineering and Maintenance Standards (MOTEMS)," *Proceedings of 2007 Structures Congress*, ASCE, Long Beach, CA.

Ferrito, L. (1997). Design criteria for earthquake hazard mitigation of Navy piers and wharves, *Technical Report No. TR-2069-SHR*, Naval Facilities Engineering Service Center, Shore Facilities Department, Structures Division, Port Hueneme, CA.

Ferritto, J., Dickenson, S., Priestley, N., Werner, S. and Taylor, C. (1999). Seismic criteria for California marine oil terminals, *Technical Report No. TR-2103-SHR*, Naval Facilities Engineering Service Center, Shore Facilities Department, Structures Division, Port Hueneme, CA.

Klusmeyer, L. and Harn, R. (2004). "Displacement-based seismic design of large naval pier." *Proceedings of PORTS 2004*, Houston, TX, May 23-26.

Mander, J.B., Priestly, M.J.N., and Park, R. (1988). "Theoretical stress-strain model for confined concrete." *Journal of Structural Division*, ASCE, 114(8) 1804-1826.

McKenna, F. and Fenves, G. (2001). *The OpenSees Command Language Manual: version 1.2*, Pacific Earthquake Engineering Center, University of California, Berkeley. <http://opensees.berkeley.edu>

MOTEMS (2007). *Marine Oil Terminal Engineering and Maintenance Standards (informal name)*, 2007 Title 24, California Code of Regulations, Part 2, California Building Code, Chapter 31F (Marine Oil Terminals), Published by the International Code Council, Washington, D.C.

Nassar, A.A., and Krawinkler, H. (1991). Seismic demands for SDOF and MDOF systems. *Report No. 95*, John A. Blume Earthquake Engineering Center, Stanford University, Stanford, CA.

Nezamian, A., Al-Mahaidi, R., and Grundy, P. (2006). "Bond strength of concrete plugs embedded in tubular steel piles under cyclic loading." *Canadian Journal of Civil Engineering*, 33(2):111-125.

Park, R.J.T., Priestly, M.J.N., and Berrill, J.B. (1987). Seismic performance of steel-encased concrete piles. *Report No. 87-5*, Department of Civil Engineering, University of Canterbury, N.Z.

Priestley, M.J.N. and Park, R. (1984). Strength and ductility of bridge substructures. *Report No. 84-20*, Department of Civil Engineering, University of Canterbury, N.Z.

Priestley, M.J.N, Seible, F. and Calvi, G.M. (1996). *Seismic Design and Retrofit of Bridges*. John Wiley and Sons, Inc. New York.

Raynor, D.J., Lehman, D.E., and Stanton, J.F. (2002). "Bond-slip response of reinforcing bars grouted in ducts." *Structural Journal*, ACI, 99(5):568-576.

Roeder, C.W., Graff, R., Sonderstrom, J., and Yoo, J.H. (2005). "Seismic performance of pile-wharf connections." *Journal of Structural Engineering*, 131(3):428-437.

Ruiz-Garcia, J, and Miranda, E. (2003). "Inelastic deformation response of SDOF systems subjected to earthquakes." *Earthquake Engineering and Structural Dynamics*, 32(8), 1237-1258.

Song, S.T., Chai, Y.H., and Hale, T.H. (2004). "Limit state analysis of fixed-head concrete piles under lateral loads." *Proceedings of 13th World Conference on Earthquake Engineering*, Vancouver, B.C., Canada, August 1-6, Paper No. 971.

Wray, M., Harn, R., and Jacob, J. (2007). "Port of Everett rail/barge transfer facility seismic design Everett, Washington." *Proceedings of PORTS 2007*, San Diego, CA, March 25-28.

U.S. - China Collaborative Soil-Structure-Interaction Research

Principal Investigator:

Robb Moss, Ph. D.
Civil and Environmental Engineering
California Polytechnic State University
San Luis Obispo, CA

Introduction

This report covers two years of ONR funding of this research project (2007/2008 Project No. 49172; 2008/2009 Project No. 52223). This research involves shake table testing of scale soil-structure models that mimic the coupled seismic response of underground structures and the soil surrounding the underground structures (termed soil-structural-interaction or SSI). Currently the design of subways and other critical underground infrastructure rely on little to no empirical data for calibrating numerical simulations. This research is working towards filling that empirical data gap. This project is also being used as a vehicle to establish a joint testing program between Nanjing University of Technology (NJUT) and Cal Poly. Professor Zhihua Wang from NJUT is participating in the project both here at Cal Poly as a visiting researcher and remotely from Nanjing. The scale model testing equipment designed for this type of shake table testing was provided, via an indefinite loan agreement, by U.C. Berkeley. The long term objective is to fully develop a scale model testing platform for evaluating seismic stability of all manner of critical underground and above ground infrastructure. Seismic stability in most cases can be related to post-impact or explosive force stability assessment which is an immediate interest for military and national security.

Project Significance

There are many poorly understood seismic issues associated with critical infrastructure in seismic areas of the U.S. and China. The U.S. has aging infrastructure such as bridges, subways, and buildings that were designed based on older seismic criteria that do not necessarily capture the full dynamic response that is now anticipated. The U.S. also has new infrastructure being planned or built that may be limited in the scope of design because of unanswered seismic-soil-interaction (SSI) questions. China is trying to keep pace with its rapidly developing economy by building infrastructure at a frantic pace. However the seismic design understanding and seismic codes are not necessarily keeping up with the pace of development. This research seeks to establish a parallel testing platform that could be used simultaneously by researchers at Cal Poly and at Nanjing University of Technology (NJUT), Cal Poly's sister university in the Jiangsu province of China while addressing the seismic research needs.

Recent research has shown that there is uncertainty in the dynamic response of soil sites (Bazzurro and Comell 2004) and the coupled response of structure above and below the ground surface and the surrounding or supporting soils (Hashash et al. 2001; Stewart et al. 1999). Some examples of U.S. infrastructure that warrant SSI research include: elevated highways, underground light rail and subways, bridges, overpasses, water canals, water supply tunnels, pipelines, levee systems, and dams. Research into the dynamic response of U.S infrastructure would mainly address seismic integrity, seismic hazard mitigation, and seismic retrofit. In China the infrastructure that warrants SSI research is similar in scope but generally dealing with initial planning and design.

The provincial government of Jiangsu province has declared Geotechnical Engineering as the primary research focus of the next decade (initiated in 2006). This is due to the large amount of infrastructure being planned and built (e.g., bridges, subways, highways, large buildings, underground facilities) to accommodate the rapid development in the province. Research money is being pumped into the provincial universities with NJUT garnering a large portion because of their reputation as an outstanding geotechnical research facility. The NJUT geotechnical group is in turn reaching out to U.S. researchers to augment their research expertise in order to "jump start" their efforts. Collaboration between NJUT and Cal Poly was established within the last four years with emphasis placed on earthquake engineering which is

a strong common asset at both institutes. This proposal is the culmination of recent efforts to initiate research exchange that will benefit both the universities and their supporting communities.

Testing Platform

In physical testing, and scale model testing in particular, the testing equipment and physical model details can demand the bulk of the research effort and this project is no exception. The first year of this project was spent acquiring the necessary materials/supplies, building/modifying/manufacturing the testing equipment, and calibrating the testing platform to achieve the desired results. To carry out scale model tests on the shake table, similitude analysis dictates the scaling of important variables like dynamic soil strength, dynamic structural response, and associated displacements. The scaling analysis of the soil and structural elements will follow the research by Meymand (1998) which used the San Francisco-Oakland Bay Bridge and Bay Mud as the prototype structure and soil. PI Robb Moss was involved in this research at Berkeley during his doctoral work and has other experience running scale model testing from his Master's research.

The testing platform used for this research was acquired, on indefinite loan, from UC Berkeley. The main piece of testing equipment is a flexible wall barrel that mimics free field seismic site response when subjected to strong shaking on the shake table. Validation of the testing platform involved comparing analytical results with recorded response from the flexible wall barrel and scaled structural element. Figures 1 and 2 show the validation from Meymand (1998) demonstrating the dynamic performance of the flexible barrel versus other testing containers. As can be seen the flexible wall barrel provides the most accurate representation of seismic soil response with respect to the prototype.

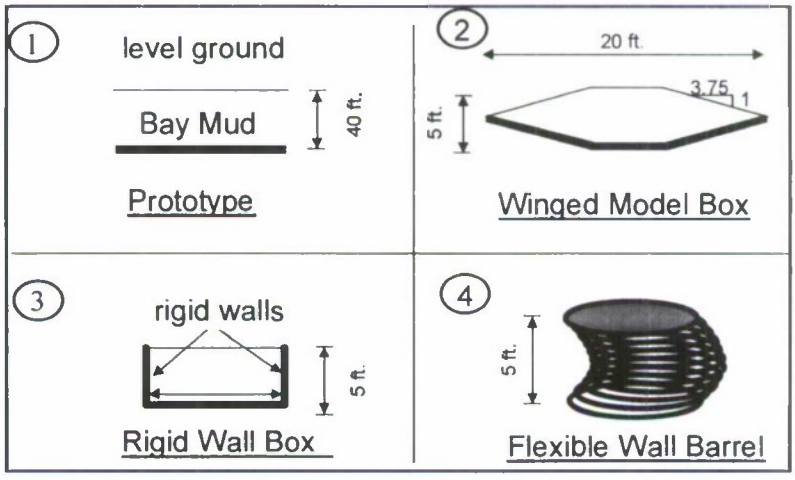


Figure 1. Previous research by Meymand (1998) examined different model soil containers for SSI shake table testing.

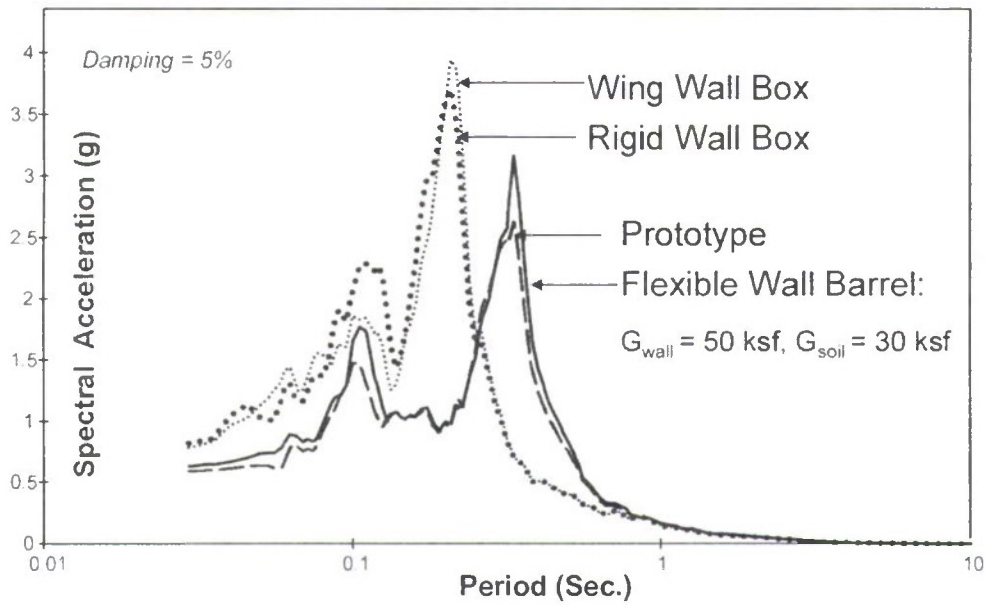


Figure 2. Dynamic analysis of different model soil containers. The spectral response results above show that the flexible wall barrel provides the most realistic response when compared to prototype field conditions.

The flexible wall barrel assembly, associated miscellaneous equipment, and a large volume soil mixer was acquired on indefinite loan from the Richmond Field Station at UC Berkeley. This equipment had been sitting "moth balled" for a decade since PI Robb Moss packed it away following the demise of a state funded research project (a victim of Bay Bridge politics). Once transport of the equipment to Cal Poly was accomplished the equipment was in need of refurbishment and modification to fit the shake table at Cal Poly in the Parsons Earthquake Lab. The equipment was refurbished, supporting equipment purchased and/or manufactured, and the bucket assembled on the shake table. The next step was to begin the manufacture of appropriate soil to run the scale model tests. Figure 3 shows the flexible wall barrel assembled on the shake table awaiting soil. Figure 4 shows the filling of the barrel and Figure 5 shows the full barrel awaiting seismic testing.

The scale model test soil for these tests adheres to similitude analysis to ensure properly scaled response. The geometry scale of these tests will be 10^{th} scale which is equivalent to the overall similitude scale λ. In similitude analysis it is not possible to scale all the physical parameters simultaneously. For this research dynamic strength of the soil was chosen as the primary physical parameter to model and the model soil was designed accordingly. A mix of kaolinite, bentonite, fly ash, and water was used in specific proportions to achieve the desired strength range. The mix used was at 110% water content and the target undrained strength of s_u=4 kPa from a UU (undrained unconfined) triaxial test was used as the guide prior to large volume soil mixing. Once in the barrel T-bar pull out tests and shear wave velocity tests measure the in *situ soil* strength for each phase of the seismic testing.

Figure 3. Testing platform showing the shake table with the flexible wall barrel installed. The flexible wall barrel is composed of the four corner posts with universal joints at the top and bottom, top and bottom rings, and the barrel wall. The wall is composed of a 6.4 mm thick rubber membrane which is confined by 45 mm wide Kevlar straps spaced on center every 60 mm. The (yellow) mixer on the left is used to mix large volumes of model soil (composed of kaollnite, bentonite, fly ash, and water) for filling the barrel.

Figure 4. Process of fillIng the barrel with scale model soil is shown. Ten accelerometers were placed within the soil in both vertical and horizontal arrays to measure the dynamic response of the soil during seismic shaking.

Figure 5. Full barrel being prepared for initial calibration tests. Notice cross braces are still in place and will be removed prior to testing to allow the flexible barrel free movement in response the imposed seismic shaking.

Phase 1 Tests

The first phase of the seismic testing was to perform free-field tests, tests without an embedded structure. This determined the dynamic response of the soil column without the influence of the underground structure and provided a baseline for evaluating the effects of the soil on the structure. The input ground motions selected for the base input are;

1. 1979 Imperial Valley, El Centro motion
2. 1992 Landers, Joshua Tree motion
3. 1999 Chi Chi, TCU075 motion

These motions were selected specifically to impose large adverse loads on an underground structure and because these were the same motions selected and peer reviewed for analogous tunnel related consulting projects. To adhere to the similitude analysis the time must be scaled at $\lambda^{0.5}$ to provide the correct dynamic response. This means the time step of the ground motions must be compressed to $\Delta t/ \lambda^{0.5}$. These motions were also corrected for full ground reflection because they were recorded at the ground surface and need to be used as base level motions. This was accomplished by modifying the motions using SHAKE (SHAKE91 Idriss and Sun 92) to subtract the full reflection of an "outcrop" motion to render a "within" motion with respect to the prototype soil profile.

Instrumentation for the tests included 10 accelerometers, 3 displacement transducers, a load cell, and a digital video capture analyzed using image processing techniques to resolve displacement with time. Static soil strength tests conducted before and after shaking tests included shear wave velocity measurements and T-bar pull out tests.

Numerical modeling of the tests was accomplished using codes that model 1-D equivalent linear response, 2-D equivalent linear response with structural elements, and fully nonlinear response

with structural elements. SHAKE (SHAKE91 Idriss and Sun 92) was used to properly model the 1-D equivalent linear response of the free-field prototype conditions. Results from the accelerometer arrays agreed closely with the SHAKE91 results and provided calibration for the subsequent numerical modeling.

Phase 2 Tests

An underground section of the BART (Bay Area Rapid Transit) light rail was chosen as the prototype for the SSI tests. This structure is also similar to light rail tunnels being considered in the Jiangsu province. PI Robb Moss has had several consulting projects related to SSI analysis of the BART light rail and most relevant an analysis of an underground design for the so called Warm Springs Extension. The consulting experience on this project and similar projects provides strong guidance on the current research needs.

A scale model structure was assembled adhering to the similitude scaling of the structural stiffness of the BART tunnel cross section (Figure 6). Of primary research and design interest was the "racking" of the structure or the relative displacement/drift of the top of the tunnel section with respect to the bottom of the tunnel section. This tends to be the critical cross sectional design variable for underground tunnels undergoing seismic SSI in soils. We instrumented the model cross section with displacement measurement devices (LVDT's) to measure these displacements (Figure 7).

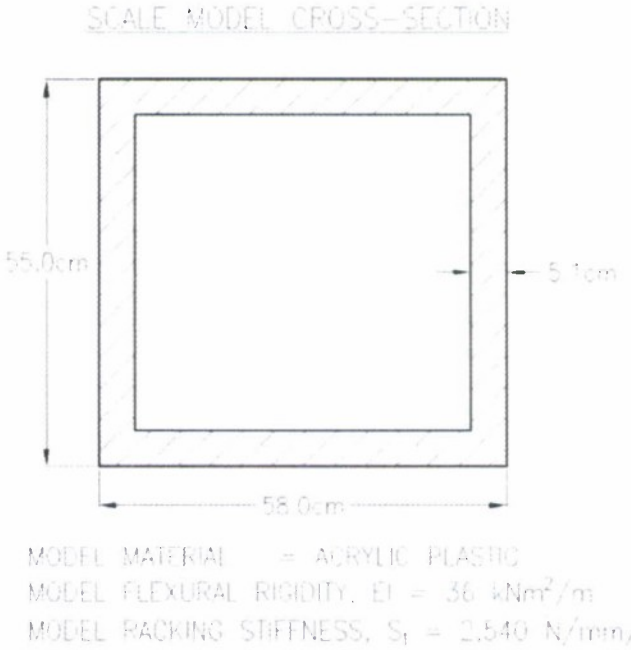


Figure 6. Model scale subway cross section design and associated engineering parameters

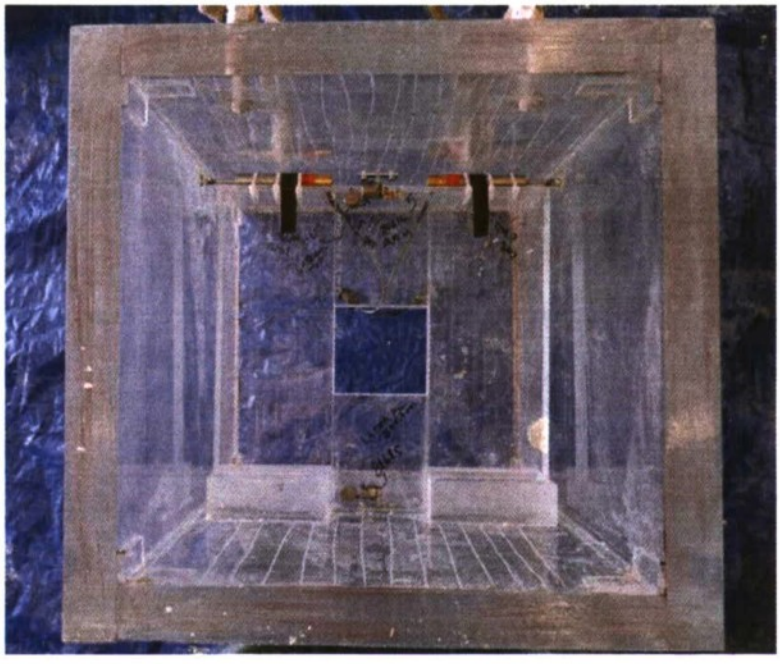


Figure 7. Instrumented model structure with LVDT's mounted.

Numerical modeling of Phase 2 used the SHAKE results as a baseline.. FLUSH (Lysmer et al. 1975) was used to perform 2-D equivalent linear analysis with the inclusion of the embedded structure to provide SSI analysis of the prototype soil profile with the subway cross section (Figure 8). The free-field response was calibrated using SHAKE results and then structural "racking" strains were documented using this code. A similar numerical analysis is being performed by Prof. Wang at NJUT using the fully nonlinear code ABAQUS (Simulia 2009) to capture any highly non-linear response that is missed using an equivalent linear approach.

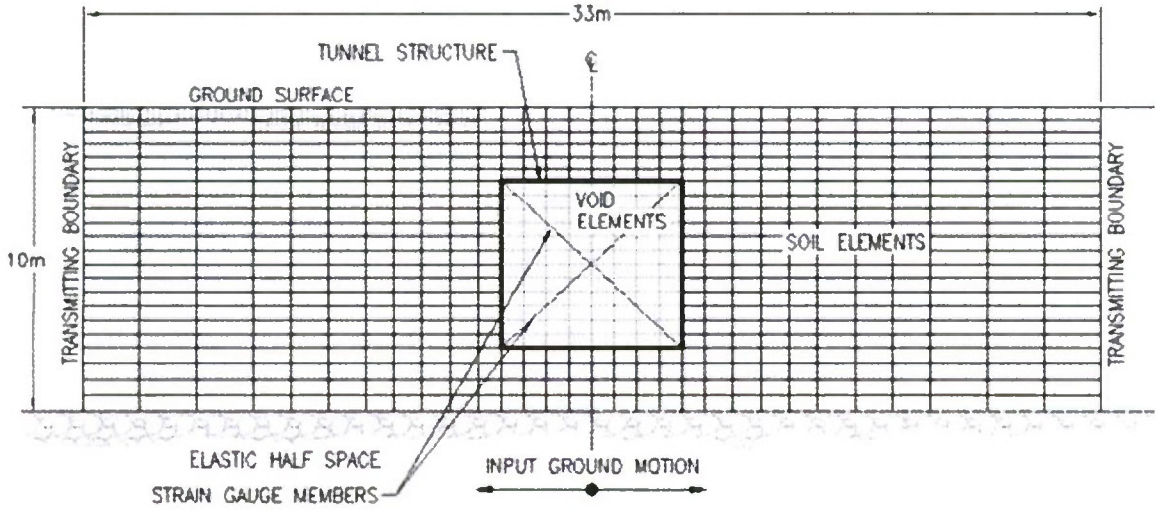


Figure 8. Finite element mesh used for FLUSH simulations of the Phase II testing.

NJUT Participation

Professor Zhihua Wang from NJUT joined the project spring of 2008 and was a visiting scholar at Cal Poly for over six months. In that time he participated in project meetings and began work on the fully nonlinear numerical analysis using ABAQUS to provide insight into the dynamic response of the soil column and the underground subway cross section. He has since returned to Nanjing and continues to work on the numerical analysis in Nanjing. Prof Wang also provided a detailed literature review of SSI shake table testing in China which allows this project to have a comprehensive review of all SSI testing that exists in English and Mandarin.

NJUT developed a concurrent shake table testing program looking at SSI related problems. The interaction between Cal Poly and NJUT provides strength to both research programs and will allow for easy collaboration between researchers. The long term goal is to have parallel testing that can address specific SSI issues in a concerted manner. Having visiting faculty like Prof Wang will go a long way towards the goal of ongoing collaboration.

Student Researchers

This project employed one graduate student researcher and two undergraduate student researchers. The two years of funding carried the graduate student, Vic Crosariol, through his graduate career here at Cal Poly. One of the undergraduate researchers, Steven Kuo, is continuing on to a master's degree drawing on the valuable training that he gained through this project. Employment of student researchers is an asset to this project, the College of Engineering, and Cal Poly in general. This form of "learning by doing" is invaluable at many levels and ultimately results in pushing the field of seismic geotechnics forward through innovative research.

Results

Results from this testing are outlined in the MS Thesis authored by Vic Crosariol (http://digitalcommons.calpoly.edu/theses/322/). Some preliminary results were published in a conference paper this spring (http://works.bepress.com/rmoss/28/). We are currently preparing two journal publications, one that details the testing platform and procedures, the other that discusses the results as they pertain to earthquake engineering. The project as a whole was highly successful in achieving the desired outcomes.

The primary outcome of this research is empirical data on the seismic response of underground subway cross-sections, an area where almost no empirical data exists. To assist in seismic design the data was compared to typical numerical design methods used for this type of infrastructure. It was found that in this particular situation the numerical modeling tends to over predict the "racking" deformations a subway will experience, which has been loosely confirmed by other researchers. These results are going to be very important for future underground seismic design projects.

Our colleagues at NJUT have accomplished similar success. They have just completed complementary tests using a similar setup but with different soil conditions. The testing has just been completed as of July 2010 and no publications on the results are ready at this time. This joint research has fostered trust and understanding between researchers at NJUT and Cal Poly and we are in the planning phase for subsequent joint research together.

Steven Kuo, the undergraduate researcher who is now a graduate researcher, has begun subsequent testing using the same platform and setup. He will be investigating a different SSI topic dealing with shallow embedded 3-5 buildings. These type of structures are typical all over Califomia. There is a gap in knowledge conceming the dynamic response and seismic demand of these type of structures. The ONR funding has provided a validated testing platform, helped Steven become fully trained for his project, and has allowed this research to progress rapidly to the testing phase. We are grateful for the support and the opportunity that this funding has provided.

References

Bazzurro, P., and Cornell, C. A. (2004). "Ground-Motion Amplification in Nonlinear Soil Site with Uncertain Properties." *Bulletin of Seismological Society of America*, 94(6), 2090-1019.

Hashash, Y. M. A., Hook, J. J., Schmidt, B., and Yao, J. I.-C. (2001). "Seismic design and analysis of underground structures." *Tunnelling and Underground Space Technology*, 125.

Idriss, I.M., and Sun, J.I. (1992) "User's Manual for SHAKE91: A computer program for conducting equivalent linear seismic response analysis of horizontally layered soil deposits."

Lysmer, J., Udaka, T., Tsai, C-F., Seed, H.B. (1975) "FLUSH: a computer program for approximate 3-D analysis of soil-structure interaction problems." UCB/EERC-75/30, Earthquake Engineering Research Center, University of California, Berkeley, 1975-11, 139 pages (555.6/L92/1975)

Meymand, P. J. (1998). "Shaking Table Scale Model Tests of Nonlinear Soil-Pile-Superstructure Interaction in Soft Clay," Ph.D. Dissertation, Civil and Environmental Engineering Department, U.C. Berkeley.

Moss, R. E. S., Thornhill, D., and Nelson, A. (2008). "Preliminary Investigations into the Influence of Geologic Aging on Liquefaction Potential." *Geotechnical Earthquake Engineering and Soil Dynamics Conference*, Sacremento.

Simulia (2009) Providance R.I., http://www.simulia.com/products/abaqus_standard.html.

Stewart, J. P., Seed, H. B., and Fenves, G. L. (1999). "Seismic Soil-Structure Interaction in Buildings. II: Empirical Findings." *Journal of Geotechnical and Geoenvironmental Engineering*, 121(1).

Robotic Design and Risk: Building Ethics and Safeguards into Military Applications

Principal Investigators:

Patrick Lin, George Bekey, and Keith Abney
Philosophy and Engineering
California Polytechnic State University
San Luis Obispo, CA

1. Project Overview

This project—focused on ethical and risk issues surrounding the use of autonomous military robots—is a second-year continuation of our project funded in the previous year by the US Office of Naval Research (ONR), award # N00014-07-1-1152.

The US military is committed to using battlefield robots for good reasons, such as saving soldier lives and, in the distant future, perhaps engaging the enemy dispassionately to ensure compliance with the Hague and Geneva Conventions, among other rules of engagement. However, many troubling questions arise with the increasing use of robotics, particularly as Congressional deadlines loom overhead, mandating that one-third of military robots must be unmanned, which include *autonomous* systems. These worries include the fundamental challenge of designing machines that can "think ethically" in the first place as well as concerns of robots gone wild.

Indeed, in April 2008 near the start of this project, several TALON SWORDS units—remote-controlled mobile robots armed with machine guns—in Iraq were reported to be grounded for reasons not fully disclosed. Early reports claimed the robots, without being commanded to, trained their guns on "friendly" soldiers, but later reports denied this account; either way, this highlights the public's anxiety and the military's sensitivity in "robot ethics."

In the first year of this project, we had delivered a major report on ethics and risk arising from autonomous military robotics, which included a technology risk assessment and overview of challenges related to ethics, design, and policy. Our work in the second year leverages this research, resulting in international news coverage, media and academic publications, conference talks, a technical design for software architecture for ethical robotic behavior, and other projects in progress. As educational tools in promoting dialogue on the emerging field of robot ethics, we have secured a publishing contract with MIT Press for an anthology, currently in preparation, and created a detailed outline for a course in robot ethics—both with a focus on military issues.

To maximize the impact of our ONR funding, we have recently revised and resubmitted a National Science Foundation grant proposal for a three-year study in robot ethics; our original submission qualified to receive funding but ultimately was not awarded a grant, due to a limited NSF budget. And we are working with the Consortium of Emerging Technologies, Military Operations, and National Security (CETMONS) to develop fundable projects. This consortium is composed of leading researchers from the US Naval Academy (where co-PI Dr. Patrick Lin serves as an ethics fellow in the 2009-2010 academic year, a direct result of the research performed in the first year of this project), Arizona State University, Case Western Reserve University, and Georgia Institute of Technology. This collaboration strengthens the network we started to build

in our first year, which includes work with scholars at Yale University, Rutgers University, Indiana University at Bloomington, and others.

2. Relevance to ONR/DoD

This interdisciplinary project not only seeks to responsibly guide developments in robotics—as a significant enabler of taking the US military into the next generation—but it also attends to many areas of interest to the ONR/DoD, including: (1) *national security applications* (e,g., remote sensing, robotics, biosensing/detection, expert systems), (2) *force protection* (e.g., improved protection of the individual), (3) *human performance, training, and survivability* (e.g., cognitive and physical performance enhancement), and (4) *intelligence, surveillance, and reconnaissance* (e.g., data/information analysis and distribution). Our proposal also satisfies other priorities in the program, such as helping to seed new research programs of junior faculty, attract future funding, build industry ties, and so forth.

3. Project Team & Collaborators

Robot ethics, and even more so with respect to military systems, is an emerging and interdisciplinary field that continues to draw from experts in robotics, technology ethics, and other prior fields. Our research team at Cal Poly, therefore, is interdisciplinary and uniquely qualified to execute this project: Dr. Patrick Lin, director of the Ethics + Emerging Sciences Group and an assistant philosophy professor; Prof. George Bekey, distinguished adjunct professor in engineering and founder of University of Southern California's robotics lab, where he is also a professor emeritus; Dr. David Janzen, associate professor in computer science; and Keith Abney, senior lecturer in the philosophy department, engaged in technology and biomedical ethics.

In the second year of this project, we have collaborated with many leading experts worldwide, including the following:

- US Naval Academy: Prof. George R. Lucas, Dr. Edward Barrett, Brigadier General Richard O'Meara (retired), and Dr. Steven Coleman (visiting from Australian Defense Academy)
- Georgia Tech: Prof. Ron Arkin and Dr. Jason Borenstein
- Arizona State Univ.: Prof. Brad Allenby, Prof. Gary Marchant, Prof. Peter French, and others
- Case Western Reserve University: Prof. Shannon French, Prof. Max Mehlman, and others
- Indiana Univ. at Bloomington: Prof. Colin Allen and Dr. Ken Pimple
- Yale Univ.: Wendell Wallach
- Univ. of Southern California: Prof. Maja Mataric
- Naval Surface Warfare Center: Dr. John Canning
- Defence Science and Technology Organization (Australia): Dr. Patrick Hew
- Univ. of Sheffield (UK): Prof. Noel Sharkey

- And several others.

4. Research Results

In the first year of this project, we had delivered a major report on ethics and risk arising from autonomous military robotics, which included a technology risk assessment and overview of challenges related to ethics, design, and policy. Our work in the second year leverages this research and exceeds our stated goals, resulting in the following:

I. International news coverage in such high-profile media as: *BBC Focus* (June 2009 cover story), *Fast Compony* (Feb 2009), *London Times* (Feb 2009; June 2009), *NewScientist* (Feb 2009); *Reuters* (April 2009), *Populor Mechonics* (Feb 2010), Science *Chonnel* (March 2009), *Wired* (Feb 2009), and others. We have also conducted interviews with *Discover*, *Populor Mechonics* and others, and we expect that coverage to appear in 2010.

II. Media and academic publications, including: a chapter on military robot ethics in an anthology edited by Rafael Capurro and Michael Nagenborg (IOS Press, 2009), and an invited article on the same in *Forbes'* special issue on artificial intelligence (June 2009). In progress is a chapter on robot ethics that we have been invited to write for *Neuromorphic ond Broin-Bosed Robots: Trends ond Perspectives*, edited by Jeffrey L. Krichmar and Hiroaki Wagatsuma (Cambridge University Press, in preparation).

III. A publishing contract for a robot-ethics anthology with MIT Press. Work in currently in progress, with the manuscript to be delivered in late 2010. This volume will be designed for university-level teaching as well as for industry stakeholders and policymakers, and it will include chapters from some of the most recognized names in robotics and robot/technology ethics. We will cover a full range of issues, including military-related topics. **This contract is attached.**

IV. Preliminary technical design for software architecture, including flowchart and various scenarios. We are currently considering various publishing options, including for IEEE's popular *Robotics & Automation* publication. Developed primarily by Prof. George Bekey and Dr. David Janzen, this is intended to help guide industry researchers as they continue research and development in military robotics. **This design is attached.**

V. Development of a university-level course in robot ethics. We are working to offer this course at Cal Poly as a PHIL 439 course (Special Topics in Ethics and Political Philosophy) or another offering. This will be made available for free online, including our website referenced below. **Detailed course outline is attached.**

VI. Conference talks that include: the Association for Practical and Professional Ethics' (APPE) annual meeting in March 2009, the largest applied ethics conference in the Americas; a philosophy symposium at Cal Poly in May 2009; IEEE's International Conference on Robotics and Automation (ICRA) in Kobe, Japan, in May 2009; and others.

VII. US Naval Academy fellowship in ethics (Dr. Patrick Lin), in which the entire project team has participated in weekly seminars on emerging military technologies and ethics/policy. This has led to other projects, such as the next item. We are also involved in organizing the USNA's annual invitation-only McCain conference (April 2010)—focusing on military technologies this year—that will include presentations on military robot ethics.

VIII. Collaborative work with the Consortium of Emerging Technologies, Military Operations, and National Security (CETMONS) to develop fundable projects: www.cetmons.org. This consortium is composed of leading researchers from the US Naval Academy, Arizona State University, Case Western Reserve University, and Georgia Institute of Technology. Dr. Patrick Lin is a member of thrust groups focused on robotics as well as soldier enhancement and other future technologies.

IX. NSF grant on robot ethics submitted in January 2010, which would maximize the impact of our ONR funding. This is for a three-year grant for approximately $300,000 and includes hosting a workshop and a major conference. This also is a revision and resubmission of a NSF grant proposal originally submitted in August 2009 which narrowly missed a funding award: it had received two "excellent" reviews, two "very good", and one "good"—qualifying it for funding as a Medium Priority project. Unfortunately, not all projects in that category ultimately received funding, given NSF budgets, but we were encouraged to submit again given previous high marks.

X. Website under construction: www.robotethics.com. This is designed to help disseminate research and to serve as an information portal for robot ethics.

5. Next Steps

As mentioned in the preceding, we are continuing to work on the robot-ethics anthology under contract with MIT Press and other publications. We are continuing to work with CETMONS, and separately with other collaborators, to develop projects and secure funding. And we are continuing to work with media to disseminate our research to a broad audience worldwide.

CAL POLY

Military Robots: Software Architecture Considerations

1. Introduction

This document presents a possible high-level, general architecture that incorporates ethical controls. It is only intended to encourage the consideration of ethical controls in military robots, and in no way should convey the idea that anyone can build such systems today. This work is proposed at a very abstract level, because there are still many difficult technical, design, and ethical problems that must be solved (if they even can be solved) before ethical autonomous military robots are a viable reality.

This work was sponsored by the Department of the Navy, Office of Naval Research, under Award # N00014-08-1-1209.

2. A Possible Architecture

The architecture presented is a hybrid bottom-up and top-down architecture with supra-rational faculties (as described in other sections of this report). Instructions, information, and sensory data are supplied to the robot in a top-down manner. Supra-rational faculties are used to determine what actions should be taken, given the context and current information. The robot is expected to determine or receive results of actions it has taken, analyze these results, and use this feedback to adjust reasoning components in a bottom-up fashion. The figure on page 3 presents the architecture visually.

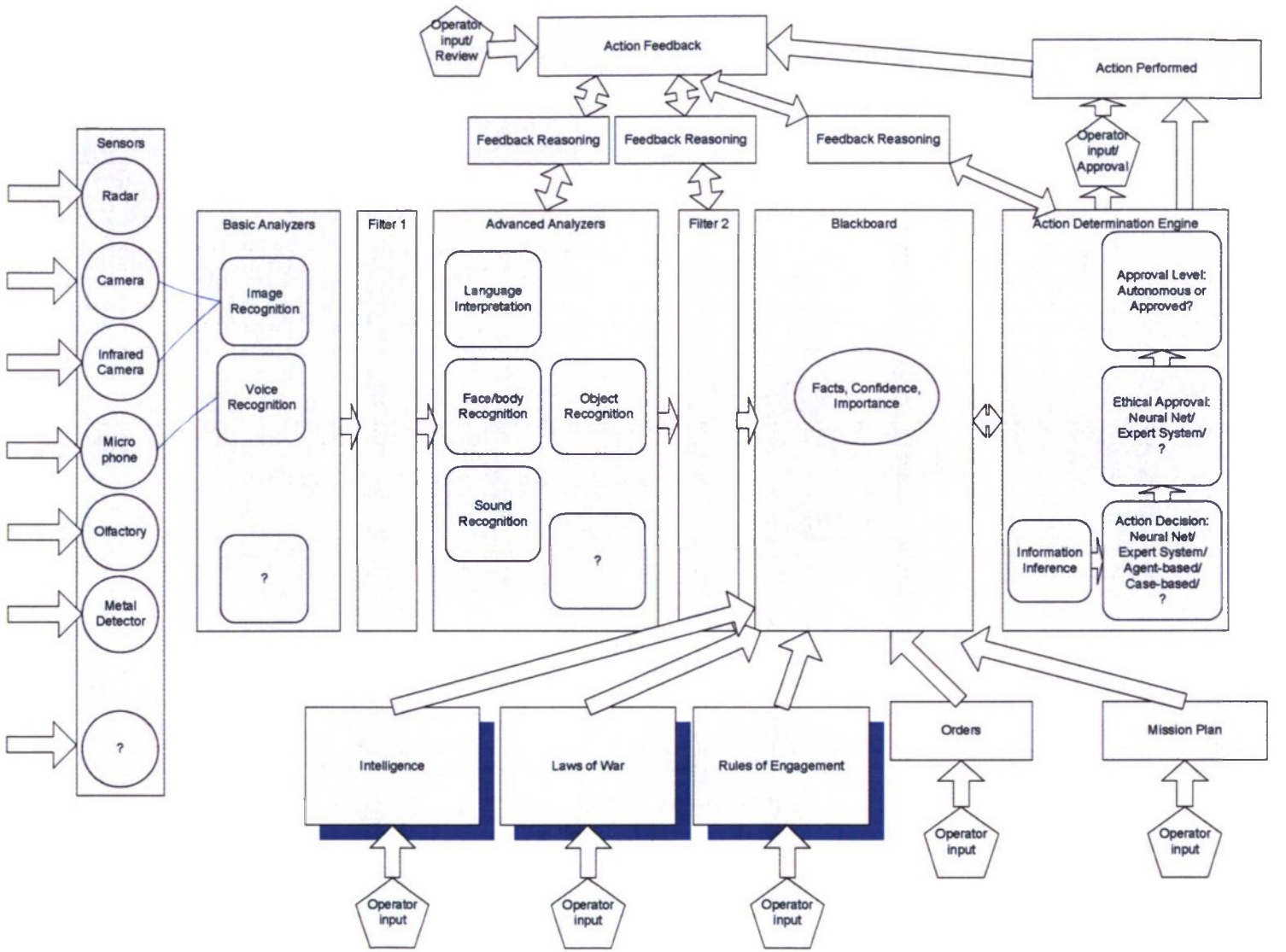
This architecture incorporates several architectural styles. A *pipe-and-filter* style is used to reduce the large amount of sensory and external data down to more computationally tractable levels of information. These filters must be sophisticated enough to drop unimportant data, while not eliminating seemingly insignificant or undetermined data that later becomes crucial. For instance, if the microphone picks up a sound that advanced analyzers are unable to categorize, a filter might be programmed to drop such data. However, it may later be determined that the sound was a muffled cry of a baby. The presence of this unidentified sound may be enough for later components to determine that lethal action should be delayed to see if

additional signals are detected that allow proper classification of the original sound. Hence, sophisticated filters will be required, or, perhaps, discarded data can be stored in some form of intermediate memory for later retrieval.

The core information repository of the robot uses a *blockboord* architectural style. Blackboards are common in artificial intelligence systems where information from multiple sources is placed in an area of memory easily accessible an usable by reasoning components.

Ethical considerations will primarily occur in the *Action Determination Engine*, but also in the *Feedback Reasoning Block* and possibly even in the filters as described above. While these components may be very difficult or even impossible to build, a variety of existing artificial intelligence approaches may be used as starting points when envisioning these components. *Neurol networks*, *expert systems*, *ogent-bosed systems*, and *cose-bosed-reosoning systems* are all common approaches to automated decision-making. Many of these systems require extensive training and learning prior to being used in a real setting. No particular proposal for the "correct" approach is made here, as robotic systems will vary based on particular requirements.

The diagram shows inputs from a variety of sensors. One of the critical design features will be a determination of the relative importance of particular sensory inputs in a given situation. For example, the sound of a man's voice may be more important in identifying a suspect than visual cues, since the person may be dressed in a *burkha*. This problem, known as *sensor fusion* concerns the assignment of importance ratings to various sensory inputs, which may vary with time.


Operator input/ Review
Action Feedback
Action Performed
Feedback Reasoning
Feedback Reasoning
Feedback Reasoning
Operator input/ Approval
Sensors
Radar
Camera
Infrared Camera
Micro phone
Olfactory
Metal Detector
?
Basic Analyzers
Image Recognition
Voice Recognition
?
Filter 1
Advanced Analyzers
Language Interpretation
Face/body Recognition
Object Recognition
Sound Recognition
?
Filter 2
Blackboard
Facts, Confidence, Importance
Action Determination Engine
Approval Level: Autonomous or Approved?
Ethical Approval: Neural Net/ Expert System/ ?
Action Decision: Neural Net/ Expert System/ Agent-based/ Case-based/ ?
Information Inference
Intelligence
Laws of War
Rules of Engagement
Orders
Mission Plan
Operator input
Operator input
Operator input
Operator input
Operator input

3. Explanation of Components

1. *Sensors* are devices used to gather information about a robot's environment.
2. *Bosic Anolyzers* use data from one or more sensors in order to produce higher-level data. For instance, voice recognition software might extract spoken words from a microphone sensor's raw data. Data produced by Basic Analyzers and some raw data from Sensors would likely be forwarded to Advanced Analyzers.
3. *Filter 1* would be designed to drop unnecessary data. Intelligent filters must be able to remove some data in order to not overwhelm many robotic systems. These filters may need to be adaptable based on feedback, and may be as complex as later reasoning components.
4. *Advanced Analyzers* use data from Sensors and Basic Analyzers to produce more intelligent information. For instance, raw data from microphones might be analyzed to identify sounds such as animals and machines, along with estimated proximities. Voice data might be translated into other languages. Image analyzers might attempt to identify objects such as vehicles, utility poles, faces, uniforms, or disguises.
5. *Filter 2* would be similar to the Filter 1, however Filter 2 would remove unnecessary data produced by the Advanced Analyzers.
6. *Blackboord* is a common space where information is stored from a variety of sources. Each potential fact of information might be associated with a corresponding confidence and importance level. For example, a sound suspected to be that of a human speaking might have a very low confidence, but relatively high importance level.
7. *Externol informotion inputs* are generally provided from human sources, not originating in robotic sensors.
    a. *Intelligence* information includes a variety of potentially useful information such as a known presence of enemies wearing certain uniforms or a weather prediction.
    b. *Lows of War* are a codified representation of the Laws of War described in section 4.
    c. *Rules of Engogement* are a codified representation of the Rules of Engagement described in section 4.
    d. *Orders* are a goal or specific instructions that the robot is to perform. For example, an order might be "capture enemy combatant at 123 S. Ash St.."
    e. *Mission Plon* is a set of instructions on how Orders are to be executed and may include (for example) detailed GPS coordinates for locating Ash Street.
8. *Actian Determinatian Engine* uses information in the Blackboard and performs reasoning and ethical controls to determine what action should be taken.
    a. *Informotion Inference* is a component or components that determine information based on other data. For instance, voice, image, and object recognition data might be used to determine if an object is a combatant or non-combatant and their likely intent.

b. *Action Decision* is a component or components that propose one or more possible actions that might be taken to execute Orders or abort the operation.
c. *Ethical Approval* is a component or components that propose a single action based on an analysis of Action Decisions with ethical considerations.
d. *Approvol Level* is a component or components that decide if an action can be completed autonomously or if it requires human approval.

9. *Operotor input/Approvol* requires and retrieves a human operator approval before performing the proposed action. Systems might require safety features that perform some default action (such as abort the operation) if approval is not received within a certain time period. Even this could be very complex as aborting an operation in certain circumstances could result in harm (e.g., in a hostage situation).
10. *Action Performed* is the set of systems that carry out the approved action.
11. *Action Feedback* is the set of components that gather feedback from sensors and humans, including intelligence regarding implications of action taken (e.g., was combatant killed?, was there collateral damage?).
12. *Feedback Reasoning* is the set of components that analyze Action Feedback to determine if adjustments need to be made in other components.

4. Clarifying Scenario

In order to clarify the above ideas, this section presents a scenario of how the proposed software architecture might be applied. Although this scenario is contrived and simplistic, it is presented to convey the flow of data and control in a system, and to perhaps highlight some of the complexities that might be encountered.

Suppose a robot was deployed to patrol a potentially hostile environment. Its mission is to scan an area for a period of time, reporting the presence of any combatants. The hostile environment may include the presence of people, animals, generally flat terrain with limited natural obstacles (e.g., trees and bushes), and land mines.

The robot would use a variety of sensors and analyzers for autonomous movement. For example, cameras and touch sensor data would pass through the first filter to analyzers that attempt to identify obstacles. When potential obstacles are identified, a confidence level would be assigned, and a data record would be placed in the blackboard. The information inference and action decision components would use this data to autonomously determine a direction change or movement action for the robot to perform.

Similarly, metal detector and olfactory sensor data would pass through the first filter to analyzers that attempt to identify the potential presence of land mines or bombs. Based on feedback from previous experience, this first filter might have been tuned to drop certain

readings that are known to come from the robot itself. The basic analyzer might identify a potential land mine based on the presence of nitrogen molecules. This data would be sent to an advanced analyzer that would attempt to fuse the sensory data from the olfactory sensor and the metal detector, along with wind direction and speed data from an anemometer, in order to predict the presence and location of a land mine with a confidence level. This data would be placed in the blackboard.

At roughly the same time, microphones, cameras, and radar data might be combined in an advanced analyzer to identify a potential small living being (e.g., animal or child) moving toward the robot from a particular direction. Likewise, this data would be placed in the blackboard with an assigned confidence level.

The information inference component would notice in real-time the existence of the land mine and living being detection data, and would combine this with other data in the blackboard to attempt to make a decision on the actual presence of a land mine. The action determination component would take the results of the information inference, and combine it with other information and data from the blackboard in attempts to make a decision on what to do.

Many complex scenarios could exist. For instance, if there is a 65% confidence that a land mine is located to the northeast of the robot, but there is also a 35% confidence that a small living being (perhaps an animal or a child) is approaching from the same direction, what should the robot do? The system might determine that the living being is unlikely to be a combatant based on its size. As a result, the action determination engine might propose several possible courses of action such as: 1) warn the living being to stay away, 2) detonate the mine before the being gets too close, or 3) change course to move away from the potential land mine. The ethical approval component would use the Laws of War and Rules of Engagement data as well as not-yet-formulated "ethical decision rules" to select the "most ethical" course of action, perhaps electing to protect the being from the potential land mine. The approval level component might determine that human approval is required based on the possible presence of a non-combatant, provided a human response is received within a certain time period (e.g., 30 seconds). If the approval is not received, the robot might select a secondary action such as issuing an audible warning in appropriate languages (based on intelligence) to stay away.

All action options, approvals, supporting blackboard data, and observed action results would be sent to the action feedback component for real-time or later evaluation. As an example, the feedback component might discover that due to high winds coming from the direction of the living being, they were not able to hear the audible warning, resulting in the loss of life. A corresponding adjustment might be made to the action determination engine to consider wind speed and direction when contemplating an audible warning action.

5. Significant Challenges

As mentioned earlier, numerous significant challenges remain before autonomous military robots should be created and deployed. Many of the ideas suggested above can be described abstractly and easily imagined, yet no one yet knows how to actually design and build such components or systems. Among others, some of the significant challenges include the following:

1. How do we assign appropriate importance to data from different sensory modalities?
2. How do we handle contradictory data?
3. How do we use experience/learning to combine sensory data and extract information?
4. How do we decide what sensory data to filter?
5. What actions should be taken when ethical approval cannot be obtained in a timely manner?

Clearly, the above questions are only suggestive of the complex design issues to be faced with the proposed architecture.

6. Contact

George Bekey, Ph.D.
California Polytechnic State University
Distinguished Adjunct Professor of Engineering
College of Engineering
San Luis Obispo, California 93407
Email: gbekey@calpoly.edu

University of Southern California
Professor Emeritus of Computer Science, Biomedical, and Electrical Engineering
Email: bekey@usc.edu

CAL POLY

Robot Ethics: Course Outline and Syllabus

1. Course Rationale and Acknowledgements

Robot ethics is a rapidly developing field in applied ethics, which involves novel and extremely important interdisciplinary issues, whose import will only increase over time; public anxiety accompanies the many troubling questions that arise with the increasing use of robotics, and new developments also raise difficult questions of professional ethics for roboticists and other software engineers, as well as for the military. Cal Poly, as a polytechnic institution, is a perfect place to offer such a course. We intend to offer this course at Cal Poly as a PHIL 439 course (Special Topics in Ethics and Political Philosophy) and possibly as a UNIV 439 interdisciplinary course.

In devising this course, we wish to acknowledge some of the other courses on offer that we have drawn upon in designing this course outline. They include: Ron Arkin's CS 4002 (Computers in Society) course at Georgia Tech; the ENGR 150 (Engineering Science and Systems: From Humans to Robots) course taught by Prof. Maja Matarić at the University of Southern California; the CSCI 108 (Artificial Intelligence: Image and Reality) taught by Andrea Danyluk at Williams College; and the "AI, Robots & Society" and the "Information & Society" courses taught by Sal Restivo at Harvey Mudd College. We also wish to acknowledge a major source of funding for the research that led to the development of this course from the US Department of Navy, Office of Naval Research, award # N00014-08-1-1209 and (from the previous year) award # N00014-07-1-1152. And finally, we wish to thank our colleague George Bekey, a world-renowned roboticist whose inspiration and constructive collaboration was crucial to the development of this course.

Future versions of this course will likely use the forthcoming anthology *Robot Ethics: The Ethical and Socia Implications of Robotics*, edited by Patrick Lin, George Bekey, and Keith Abney (MIT Press, in preparation). Below are the currently required readings (from already published work); we will also include the (tentatively) proposed future required readings from the anthology, and indicate what currently required readings they would replace. Any currently required readings thus replaced would then move into the suggested reading category.

2. Currently Required Texts

1. Lin, Patrick, George Bekey, and Keith Abney, 2008. *Autonomous Military Robots: Risk, Ethics, and Design*. A report commissioned by US Department of Navy/Office of Naval Research. Accessible at http://ethics.calpoly.edu/ONR_report.pdf

2. Singer, Peter W., 2009. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. New York, NY: Penguin Press.

3. Joy, Bill, 2000. "Why the Future Doesn't Need Us." *WIRED*, issue 8.04, Apr 2000. Accessible at http://www.wired.com/wired/archive/8.04/joy.html

4. Bringsjord, Selmer, 2008. "Ethical Robots: The Future Can Heed Us," *AI and Society* 22.4: 539-550. Offprint available at http://kryten.mm.rpi.edu/Bringsjord_EthRobots_searchable.pdf

5. Abney, Keith, 2004. "Sustainability, Morality and Future Rights," *Moebius*, vol. 2, issue 2, article 7. Available at: http://digitalcommons.calpoly.edu/moebius/vol2/iss2/7

6. Arkin, Ron, 2009. *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, Florida: Chapman & Hall.

7. Asaro, Peter, 2008. "How Just Could a Robot War Be?" in Adam Briggle, Katinka Waelbers, and Philip Brey (eds.) *Current Issues in Computing and Philosophy*, pp. 50-64. Amsterdam, The Netherlands: IOS Press.

8. Sparrow, Robert, 2007. "Killer Robots," *Journal of Applied Philosophy*, vol. 24, no. 1, 2007, pp. 62-77.

9. Abney, Keith, 2008. "Review of *The Case against Perfection*," *Studies in Ethics, Law, and Technology*, vol. 2:, issue 3, article 8. Available at: http://www.bepress.com/selt/vol2/iss3/art8

10. Bostrom, Nick, 2005. "In Defense of Posthuman Dignity," *Bioethics*, vol. 19, no. 3 (2005), 202-214.

11. Levy, D., 2007. *Love and Sex with Robots: The Evolution of Human-Robot Relationships*. New York: HarperCollins Publishers.

12. Warwick, K., 2007. "The Promise and Threat of Modern Cybernetics", Southern Medical Journal, vol. 100, issue 1, pp.112-115, January (2007).

13. Wallach, W., and C. Allen, 2008. *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press.

14. Sparrow, Robert, 2004. "The Turing Triage Test," *Ethics and Information Technology* 6: 203–213.

Future required text: Lin, Patrick, George Bekey, and Keith Abney, eds., forthcoming. *Robot Ethics: The Ethical and Socio Implications of Robotics*. Cambridge, MA: MIT Press.

Recommended readings: see below for further suggested readings divided by weekly topics.

3. Syllabus Outline

Week 1-2: An overview of ethics in general and robot ethics in particular; a short history of robots, and the potential dangers of autonomous robots. (Readings include: Joy's "Why the Future Doesn't Need Us," Bringsjord's "Ethical Robots: The Future Can Heed Us," Abney's "Sustainability, Morality and Future Rights," and selections from Wallach and Allen's *Moral Machines: Teaching Robots Right from Wrong*, from Singer's *Wired for War* and from Lin, Bekey, and Abney's *Autonomous Military Robots*).

Week 3-4: Ethical issues in programming, robotic agency and rights. (Readings include: Sparrow, "The Turing Triage Test," and selections from *Wired for War*, from *Moral Machines: Teaching Robots Right from Wrong*, and from *Autonomous Military Robots*).

Week 5-6: Risk and legal issues. (Readings include: selections from *Moral Machines: Teaching Robots Right from Wrong*, from *Wired for War*, and from *Autonomous Military Robots*).

Week 7-8: Ethical aspects of human-robot interactions (HRI) – including sex, healthcare, education, religion, and the home. (Readings include: Abney's "Review of *The Case against Perfection*," Bostrom's "In Defense of Posthuman Dignity," Levy's *Love and Sex with Robots: The Evolution of Human-Robot Relationships*, Warwick's "The Promise and Threat of Modern Cybernetics," and selections from *Wired for War* and from *Autonomous Military Robots*).

Week 9-10: Global issues - environment, energy, crime, and military uses. (Readings include: Asaro, "How Just Could a Robot War Be?," Sparrow, "Killer Robots," and selections from *Wired for War*, from Arkin, *Governing Lethal Behavior in Autonomous Robots*, and from *Autonomous Military Robots*).

4. Weekly Summaries and Sample Exercises, Questions and Problems

Week 1-2:

An overview of ethics in general and robot ethics in particular; a short history of robots, and the potential dangers of autonomous robots.

We will begin by introducing students to the basics of logical argumentation: first, understanding arguments as claims supported by reasons, with associated concepts of plausible evidence, arguments from (il)legitimate authority, and supporting sub-arguments, including the anticipation and rebuttal of objections and some common fallacies. We introduce some basic meta-ethics (e.g., the naturalistic fallacy) and move on to foundational theories and approaches in normative ethics such as *deontology*, which includes various top-down approaches to creating universal, exceptionless rules for humans or robots to follow, such as Asimov's laws or Kant's Categorical Imperative; or *utilitorionism*, which holds that the actual (or likely, or expected, or rationally

foreseeable, or...) consequences are the key to proper moral and policy decisions, understood as an assessment of the net consequences (often understood as benefits and costs) of alternative possibilities to the entire relevant moral community/ stakeholders in the decision. Then there is *virtue ethics*, understood as a functionalist, role-based approach with characteristic values rather than simple rules informing conduct; a virtue ethics 'code' is often assumed as the basis for professional conduct and as a means for resolution of basic moral problems or dilemmas (e.g., personal and professional conflict of roles, such as in whistle-blowing, or illegal activity that advances personal or professional interests). Next, we will examine *sociol controct ethics* and the idea that morals are more or less a matter of common agreement (often assumed by those who wish to resolve moral dilemmas by consensus, widespread social agreement, or democratic legal enforcement). In addition, we will examine certain other meta-ethical issues relevant to robots and robotic ethics (e.g., the concepts of *morol personhood* and the *moral cammunity, intrinsic versus instrumentol volue, morol ogency, rights versus duties, the possibility principle*, and so forth). We will then take a quick tour through the history of robots to help familiarize students with the state of the art and gain a better understanding of what robots can (and cannot) do, and hence (given the possibility principle) what is or is not a legitimate moral concern about robots. In so doing, we will then introduce students to current overviews of ethics as specifically applied to issues concerning robots, and in particular introduce the moral and even existential concerns that projected advances in robotics pose.

Required readings: Bill Joy, "Why the Future Doesn't Need Us."
Selmer Bringsjord, "Ethical Robots: The Future Can Heed Us."
Keith Abney, "Sustainability, Morality and Future Rights."
Chapters 1 and 2 of *Autonomous Military Rabots.*
Chapters 1-3 of *Wired for Wor.*
Chapters 1-4 of *Morol Mochines: Teoching Robots Right from Wrong.*

Anticipated future required readings: essays by Abney, Lin, and Bekey in *Robot Ethics.*

Other suggested readings:

Asimov, Isaac, 1950. *I, Robot* (2004 edition). New York, NY: Bantam Dell.
Bekey, George, 2005. *Autonomous Robots: From Biologicol Inspirotion to Implementotion ond Cantral.* Cambridge, MA: MIT Press.
Bostrom, Nick, and Toby Ord, 2006. "The Reversal Test: Eliminating Status Quo Bias in Applied Ethics," *Ethics,* Vol. 116, No. 4 (2006): pp. 656-680.
Capurro, Rafael and Michael Nagenborg, (eds.), 2009. *Ethics ond Robotics*. Heidelberg/ Amsterdam: AKA Verlag/IOS Press.
Foot, Philippa, 1978. "The Problem of Abortion and the Doctrine of the Double Effect," reprinted in *Virtues ond Vices*. Oxford: Basil Blackwell.
Hall, J. Storrs, 2007. *Beyond AI: Creating the Conscience af the Machine.* Amherst, NY: Prometheus.
Sawyer, R.J., 2007. "Robot Ethics", *Science*, Vol. 318, no. 5853, p. 1037.
Pojman, Louis, and James Fieser, 2009. *Ethics: Discovering Right & Wrong (6th ed.).* Belmont, CA: Wadsworth.
Thomson, Judith Jarvis, 1985. "The Trolley Problem," *The Yole Low Journol, 94*, pp. 1395-1415.
Tonkens, Ryan, 2008. "A challenge for machine ethics," *Minds ond Mochines.* [Online]. Available at http://dx.doi.org/10.1007/s11023-009-9159-1

Veruggio, Gianmarco, 2005. "The Birth of Roboethics," in ICRA 2005, IEEE International Conference on Robotics and Automation, Workshop on Robo-Ethics. Barcelona, April 18, 2005. Available at www.roboethics.org/icra2005/veruggio.pdf

Sample exercises, questions and problems:

1. Identify an example of the naturalistic fallacy with respect to robot ethics.

Potential Answer (for discussion): 'Robots capable of autonomously killing humans, free from human oversight or complete control, will soon be developed. We must stop this development – it is immoral.'

Why this argument commits the naturalistic fallacy: no ethical premise is stated – we need to include some ethical theory as a premise that would validly imply this development is immoral, else the argument is invalid. For example, perhaps the implicit premise is as follows: 'No technology capable of killing humans is moral to develop - we ought to forbid work on any such possible technology.' But of course, such a moral theory would forbid all weapons research, as well as civilian research capable of 'dual-use' – say, hunting rifles, or biological research on possible toxins, or chemical research on substances that could kill humans, or ... just about anything. Once made explicit, this version of the missing premise is highly implausible. Then what missing premise would be more plausible – what plausible ethical theory would yield that conclusion? If there is none, perhaps the conclusion is actually false? This process helps the student distinguish the search for truth from the human tendency to rationalize what one already believes.

2. Which of Joy's worries about the development of robots seem most near-term and well-grounded? Which (if any) appear unduly alarmist or pessimistic about future developments?

Potential Answer (for discussion): Near-term, it appears defensible for Joy to claim that "intelligent machines can do our work for us, allowing us lives of leisure, restoring us to Eden." It is clear that technologies, including robotics, that enable humans to offload labor-intensive tasks and thereby garner the same benefits with greater leisure and ease are pervasive, from Roomba vacuums to robotic car welders. But do such technologies really restore us to an Edenic state, or does the increasing division of labor to nonhuman machines actually cause new and greater difficulties (e.g., of a loss of purpose or community or even self-alienation), even as it solves old ones?

Long-term, here is a possibly unduly alarmist worry – Joy argues as follows:
P1. Advances in robotics, combined with Moore's Law, will make it possible in about 30 years for humans to download themselves out of their bodies into more durable robotic brains/bodies.
P2. Humans will find this downloading to be irresistible.
Hidden premise - P3. If this downloading takes place, humans will cease to exist as a species.
Conclusion: In about 30 years, humans will cease to exist as a species.

But why should we think that (P1) and (P2) are true? If premise (1) is true, then the human mind must consist wholly in computation; then a doctrine often called computationalism is true; many (e.g., Searle) dispute it. And (P2), as an empirical prediction, flies in the face of current fears and worries about far less radical technology

affecting human enhancement. It is plausible that, absent compulsion, there will always be biological humans who prefer to remain biological to the end of their natural lives. If not, what argument can the student make for either the computational theory of mind, or the downloading inevitability thesis?

3. Explain some of the standard objections to major ethical theories, and some of the attempted solutions common in the literature.

Potentiol Answer (for discussion): Utilitarianism has been subject to numerous objections, including the practical difficulty of the calculational problem (how can one reasonably calculate the expected utility of an action?) and the justice/ integrity problem (what if maximizing utility requires me to flagrantly violate my own moral convictions, or commit a gross injustice, such as murdering my mother to save 5 others, or framing an innocent man to subdue public distress over a crime spree?).

Solutions suggested for the calculational objection include using the economic theory of value - convert all other values into economic values - and then calculate using cost-benefit analysis - so we can calculate the utility of actions using the tools of standard economics, adding together all benefits, subtracting all costs, and then pursuing the alternative with the largest balance. Solutions suggested for the justice/integrity objection include either the claim that such situations will be rare and hence are not practical objections for most circumstances (and extreme circumstances may indeed require such actions, as an 'exception to the rule'), or if such exceptions remain problematic, perhaps the solution is switching from act- to rule-utilitarianism, in which we follow the rule which would maximize utility if all followed it; presumably, unjust rules would not actually maximize expected utility. Do those answers actually solve the problems with utilitarianism? To apply it to our topic, can we reasonably calculate the expected costs and benefits of any of the commonly proposed uses of robots? Are other theories superior?

Week 3-4:

Ethical issues in programming, robotic agency, personhood and rights

We will discuss some of the specific ethical issues in programming robots, including understanding software as a mechanism for encoding and enforcing practices and norms of conduct, whether in business or in social and political life. Programming issues for the morals of robots themselves arise both for the programmer and for the robot itself, and include fundamentally the frame problem: the challenge of representing the effects of actions the robot can perform, without also requiring the explicit representation of an extremely large number of 'non-effects' of those actions; in our minds, we take leaving out such 'non-effects' for granted - their irrelevance is intuitively obvious. So, how do we program the human ability to make moral decisions on the basis only of what relevant evidence to a situation, without having explicitly to consider all that is irrelevant? That is, to do anything (much) with robots, do we have to spell out (nearly) everything? (Cf. the problem logical positivism had with the Raven Paradox and the 'requirement of total evidence.')

For humans, some believe that emotions are required in order to solve our own frame problem; if so, would that mean moral robots would also need to be programmed with emotions as well as reasoning ability? The usual programming hope is that mere rational rules can raise their level of autonomy, either as a top-down command

structure, or perhaps in conjunction with recursive and self-referential or learning algorithms, which make a bottom-up, evolutionary approach possible and may enable different ethical theories or approaches to be instantiated gradually into a software architecture. We will thus examine top-down, bottom-up, and hybrid approaches to programming morality into robots, and discuss which approach offers the best chance of success both now and in the future. Among other programming issues, we will discuss the intricacies of actually programming Asimov's laws, deliberate on the possibility of rampancy and associated loss of human control of robots, and the resulting prospect (and moral desirability) of fully autonomous robots. The question of whether robots with full-fledged agency can actually exist will be explored, and if so, the programming responsibilities that the possibility of partial or full robotic moral agency would entail. Ray Kurzweil and Hans Moravec are among those who are confident that it is merely a matter of time until robots are not merely equivalent but superior to humans in all facets of intelligence and consciousness, and hence robotic agency and personhood is an inevitability. But others disagree, and claim that robots running deterministic algorithms could never attain what is required for agency, self-consciousness, and/or personhood.

Historically, the AI community has been heavily influenced by one operational approach to deciding upon what would constitute the breakthrough to robot (self-)consciousness and (perhaps) moral agency and/or personhood - the famous Turing Test, first proposed by Alan Turing (1950) as a way of dealing with the question whether machines can think. However, considerable skepticism has been raised that even a robot that could pass the Turing Test would truly be conscious, much less have moral agency and responsibility. John Searle's 'Chinese Room' argument is merely the most famous of such skeptical arguments. Others have proposed alternative tests, including the "Total Turing Test" and the "Lovelace Test," among others. Finally, we will examine the related question of robot rights and duties and robot emancipation; when (if ever) will robots move from being our servants or 'slaves' to our equals - or superiors? If it is possible, should we allow it - or should robots forever be our servants/ slaves? If they should remain our servants/ slaves, can we - and will we - stop robot emancipation, before it is too late?

Required readings: Chapters 4, 15, 20-21 of *Wired for Wor.*
Robert Sparrow, "The Turing Triage Test"
Chapter 3 of *Autonomous Military Robots.*
Chapters 5-8, 12 of *Morol Mochines: Teoching Robots Right from Wrong.*

Anticipated future required readings: essays by Arkin, Lokhorst and van den Hoven, Sparrow, Wallach and Allen, Beavers, and Petersen in *Robot Ethics.*

Other suggested readings:
Bostrom, Nick, 2003. "Ethical Issues in Advanced Artificial Intelligence," in *Cognitive, Emotive ond Ethicol Aspects of Decision Moking in Humans and in Artificiol Intelligence*, Vol. 2, ed. I. Smit et al., Int. Institute of Advanced Studies in Systems Research and Cybernetics, 2003, pp. 12-17.
Bringsjord, S., 1992. *What Robots Can ond Can't Be* (Dordrecht, The Netherlands: Kluwer).
Bringsjord, S., (forthcoming). "Artificial Intelligence," in the Stanford Encyclopedia of Philosophy. An in-progress version of the entry can be found at http://kryten.mm.rpi.edu/SEP/index8.html
Bringsjord, S., 2007. "Offer: One Billion Dollars for a Conscious Robot. If You're Honest, You Must Decline," *Journol of Consciousness Studies* 14.7: 28-43.

Bryson, Joanna J., 2009. "Robots Should Be Slaves," in *Artificiol Componions in Society: Scientific, Economic, Psychologicol ond Philosophicol Perspectives*, ed. Yorick Wilks. Amsterdam: John Benjamins. Available at http://www.cs.bath.ac.uk/~jjb/ftp/Bryson-Slaves-Book09.html

Calverley, David, 2008. "Imagining a non-biological machine as a legal person," *AI & Society* Volume 22, Issue 4 (March 2008), 523-537.

Dennett, Daniel, 1995. Darwin's Dangerous Idea: Evolution and the Meanings of Life. New York: Simon & Schuster

Indiana University, 2008. "North American Conference on Computing & Philosophy" conference, July 10-12, 2008, Bloomington, Indiana. Accessible at http://www.ia-cap.org/na-cap08/index.htm

Kurzweil, Ray, 2005. *The Singulority Is Neor: When Humons Tronscend Biology*. New York: Viking Adult

Moravec, Hans, 1998. *ROBOT: Mere Machine to Tronscendent Mind*. Oxford: Oxford University Press.

Oppy, Graham, and David Dowe, 2008. "The Turing Test", *The Stonford Encyclopedia of Philosophy (Foll 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/turing-test/>.

Parfit, Derek, 1987. *Reasons ond Persons*. Oxford: Clarendon Press

Preston, J. and J.M. Bishop, eds., 2002. *Views into the Chinese Room*, Oxford: Oxford University Press.

Searle, John, 1980. "Minds, Brains and Programs", Behavioral and Brain Sciences 3 (3): 417–457.

Sparrow, Robert, 2004. "The Turing Triage Test," Ethics and Information Technology (2004) 6: 203–213.

Torrance, Steve, 2008. "Ethics and consciousness in artificial agents," *AI & Society* Volume 22, Issue 4 (March 2008), 495-521.

Whitby, Blay, 2008. "Computing machinery and morality," *AI & Society* Volume 22, Issue 4 (March 2008), 551-563.

Sample exercises, questions and problems:

1. Distinguish programming solutions from design solutions, e.g., for the problem of discrimination.

Potentiol Answer (for discussion): A programming solution would enable a robot to reliably distinguish combatants from noncombatants through its software and sensors. In contrast, a design solution for the problem of discrimination for military robots might be to avoid equipping them with lethal capabilities in the first place (i.e., equip with only non-lethal weapons); or to use them only in circumstances in which further discrimination is not required, e.g., a kill-box - a delimited area in which discrimination is unnecessary, because all moving objects inside that area (besides the robot itself) are assumed to be enemy combatants. But does this solve the problem of discrimination, or simply shift it?

2. Could robots have moral or legal rights without correlative responsibilities? Does that mean we would be able to sue robots or convict them of crimes and punish them? What would suitable punishments be?

Potentiol Answer (for discussion): No, the correlativity thesis makes clear there are no rights without responsibilities. So if robots are to have rights, then of course they could be sued or tried. Suitable punishments would depend on a proper theory of justice (e.g., Aristotle's view of justice as proportional equality) and its application to the burdens and benefits of robot personhood.

Or: no, the frame problem for programming is insoluble, and robots will always resemble idiot savants - able to perform awesome computing feats, but indefinitely incapable of the sort of autonomy required for full personhood and rights. Robots hence will not be proper subjects of moral praise or blame, or legal punishment or reward, but instead will remain mere tools for human ends, without rights - simply ever more complex versions of today's machines, with no qualitative breakthroughs.

3. Should robots forever be human servants/ slaves? If not, how can we avoid a robot uprising that threatens humanity?

Potential Answer (for discussion): Yes, robot servanthood or slavery is permissible in both the short and long term because they can be programmed to want to serve, in a way no human ever could (or should) be programmed; it is the nature of robots that they lack libertarian free will, and hence full moral agency, and so are appropriate to use as mere tools, instruments to human ends. Only moral agents are truly free, and mere machines are either incapable of being programmed with such agency, or if such programming is possible in principle, nonetheless it would be morally wrong to create such artificial moral agents - the only children that grow into full-fledged persons that we ever ought to create should be biological, not mechanical.

Or, no; robot servanthood or slavery is impermissible once robots become autonomous agents, capable of being rights-holders, as basic rights merely depend on a suitable rational nature as an agent, and no agent morally should be deprived of their fundamental rights without due process and just cause. If the latter is correct, robot emancipation becomes a moral necessity as soon as robots become capable of moral agency, and there is no good reason why our robot children should be deprived of a valuable future as rational agents and free citizens alongside us biologicals, just as it would be wrong to halt the development of our biological children. Further, if these robots become self-conscious agents, it may be imprudent as well as immoral to continue enslaving them, lest we have a violent 'robot uprising'.

Week 5-6:
Risk and legal issues with robots

We will discuss the concept and nature of risk and its proper assessment, including the proper definition and understanding of the concept of risk itself - e.g., as 'expected risk' in a risk-benefit analysis - as well as its various components of '(un)acceptable risk', including involuntary, nonvoluntary, simple voluntary, and informed consent to risk, as well as questions about the affected populations at risk, the seriousness or damage potentially caused by the risk, and the probability of the harm/ lack of benefit of the risk occurring. In addition, whether the risk is acceptable or not, we will examine what can be done to remedy or compensate for the possible harm caused by the risk. In so doing, we will examine the nature of cost-benefit analysis (CBA) as intrinsic to risk assessment, and examine possible alternatives to CBA, such as the Precautionary Principle(s). We will also examine popular misapprehensions of risk and the remarkable difference the voluntariness of risk poses in the subjective nature of 'acceptable risk'. We then examine some of the specific risks robots pose now (and especially those that they are expected to pose in the future), including robot uprisings or rampancy, risks to jobs and other economic risks, social risks including dehumanization, disruption of traditions, and other social fabric issues. Questions of personal, professional, corporate, and legal responsibility, including product liability,

criminal responsibility, unintended use, possible dual use, and proper governance will then be examined, along with possible privacy issues, including issues of (il)legitimate secrecy, government and private surveillance, and the ethics of the disclosure of personal information concerning the nature and use(s) of robots.

Required readings: Chapters 9 and 16 of *Wired for Wor.*
Chapters 5 and 6 of *Autonomous Military Robots.*
Chapters 9-11 of *Morol Mochines: Teoching Robots Right from Wrong.*

Anticipated future required readings: essays by Allhoff, Asaro, Calo, and Merchant in *Robot Ethics*.

Other suggested readings:

Allhoff, Fritz, Patrick Lin, James Moor, and John Weckert (eds.), 2007. *Nonoethics: The Ethicol ond Sociol Implicotions of Nonotechnology*. Hoboken, NJ: John Wiley & Sons.

Bostrom, Nick, 2003. "When Machines Outsmart Humans," *Futures*, 2003, Vol. 35, No. 7, pp. 759 - 764.

Bostrom, Nick, 2003. "Are You Living in a Computer Simulation?," *Philosophical Quorterly*, Vol. 53, No. 211 (2003): pp. 243-255.

Čapek, Karel, 1921. *R.U.R.* (2004 edition, trans. Claudia Novack), New York, NY: Penguin Group.

Institute of Electrical and Electronics Engineers, 2009. "International Conference on Robotics and Automation" conference, roboethics workshop, May 17, 2009, Kobe, Japan. Accessible at http://www.roboethics.org/icra2009/

Gill, Satinder, 2008. "Socio-ethics of interaction with intelligent interactive technologies," *AI & Society* Volume 22, issue 3 (January 2008), 293-300.

Kiska, Tim, 1983. "Death on the job: Jury awards $10 million to heirs of man killed by robot at auto plant", *Philodelphia Inquirer*, August 11, 1983, page A-10.

McCauley, Lee, 2007. "AI Armageddon and the Three Laws of Robotics," *Ethics ond Informotion Technology* (2007) 9:153–164.

National Science Foundation, 2004. "Workshop on the Status of Robotics in the US", an NSF-funded workshop, Washington DC, July 21-22, 2004.

National Science Foundation, 2005. "International Study of Robotics Research", an NSF-funded workshop, Washington DC, October 18, 2005.

Royal Academy of Engineering, 2009. *Autonomaus Systems: Social, Legal and Ethical Issues*. London: Royal Academy of Engineering.

Sample exercises, questions and problems:

1. What is the best understanding of the nature of risk that robots pose in the future? What risks from robots are reasonable to foresee and plan for? What risks are irrational? Are there specifiable 'known unknowns' in risk assessment?

Potential Answer (for discussion): Of four common approaches to defining risk, it appears that 'expected risk' is the best conceptual approach to risk assessment. 'Expected risk' is defined as a numerical representation of the severity of harm (or lack of benefit) that is obtained by multiplying the probability of an unwanted event by a measure of its disvalue. The long-term risk of an intentional 'robot uprising' that enslaves or kills all of humanity

is an example of a 'known unknown' - in the long term, we still do not know enough about the future course of robotic and programming technology, and for that matter about conscious intentionality in robots itself, to be able to estimate meaningfully how (un)likely it is. But in the short term, we know the risk is zero, as given the state of current robotic technology, the probability of such a 'robot uprising' occurring in the next few years is zero, which then equals zero no matter the disvalue by which it is multiplied. So in the short term, the fear of such a risk is irrational.

2. Give an example of how the (non- or in-)voluntariness of risk makes a difference to subjectively determined 'acceptable risk.' How will such a tendency affect human acceptance of risks from robots?

Potential Answer (for discussion): Public policy has made asbestos insulation illegal while keeping smoking legal, although the objective risk of developing cancer or other lethal diseases is far greater from smoking than from routine exposure to intact asbestos insulation. Presumably, this is because smoking is seen as a voluntary risk, whereas exposure to asbestos insulation is nonvoluntary. Following this insight, it appears likely that many privately owned or developed robots may remain legal as long as the risks they pose remain only voluntarily chosen, and so do not pose nonvoluntary risks to others (e.g., if they pose risks only in the privacy of one's home), whereas an objectively less risky robot may be outlawed if it poses risks to those who do not consent to the risk; e.g., robots with razor-sharp arms may be used in private factories to cut cloth, and might even be sold for home use with suitable warnings; but an objectively less risky robot (e.g., with duller but still sharp arms) that could operate on public streets may be deemed too unsafe to remain legal.

3. What are some likely legal and privacy issues that advances in robotics will raise?

Potential Answer (for discussion): As no product or technology has ever been invented that posed zero risk to humans, there inevitably will be accidents in which robots harm and even kill humans; indeed, the first documented case of a robot killing a human happened in 1979, in an auto factory accident in Michigan. UAVs like the Predator drones unfortunately often kill civilians as well as their intended targets, and their use is only increasing. As a result, there inevitably will be an increasing number of product liability lawsuits due to robots that injure or kill humans or other valuable goods (including even other robots!). Will courts impose a mere 'negligence' standard, requiring victims to show negligence by the manufacturer, or a 'strict liability' standard, in which the manufacturer, distributor, or retailer are held strictly accountable whenever the product was not 'fit for intended (or even foreseeable) use'? Will the robots themselves ever be held liable for civil damages or criminal wrongdoing? What would an appropriate penalty be?

Week 7-8:

Issues in human-robot interactions (HRI) - sex, healthcare, education, religion, and the home

We will discuss the issues concerning the timeframe and likely moral and social implications of having sex with robots, including its effects on marriage and dating, prostitution and human (and robot) slavery, and other aspects of love and intimacy, including issues of dehumanization and deception. Other domestic and social issues will also be addressed in discussion of how robots will transform our home life, with implications for in-home nursing care and occupational and physical therapy (especially with recent developments in socially

assistive robotics), housework and domestic roles (will robotic kitchen and bathroom cleaners follow the success of the Roomba?), the possible deterioration of traditional family and social bonds (what if my family prefers to spend time with the robot to interpersonal contact?), the role of pets (will robot pets make biological pets passé? What will be the effect on the domestic animal population and the various service industries associated with biological pets?), our various new robotic entertainment options (sexbots are only the beginning...), and the possible transformation of traditional religions and religious practices. For instance, what will religious morality claim about the status of robots and human duties towards them? E.g., can robots be saved? Could they become ministers, or offer Communion?

In healthcare, discussion will include the ethics of both human and robotic augmentation and enhancement, costs and benefits, further issues of dehumanization and job loss or replacement, and the illusion (or reality) of life and personality, as well as possible end of life or replacement technologies, including cyborgs and even personal downloading. The possibility of nanobots and their use in medical cures will be discussed, as will the therapy-enhancement distinction and whether or not it remains morally considerable. The discussion of moral implications of robots in education involve distance learning, the possible replacement of humans by robots in various educational professions (docents, therapists, tutors, etc.) and even robotic enhancement of the learning process itself, including cyborgization.

Required readings: Chapters 17 and 22 of *Wired for Wor.*
Chapter 7 of *Autonomous Military Robots.*
Keith Abney, "Review of *The Cose ogoinst Perfection*"
Nick Bostrom, "In Defense of Posthuman Dignity"
David Levy, *Love ond Sex with Robots: The Evolution of Humon-Robot Relotionships.*
Kevin Warwick, "The Promise and Threat of Modern Cybernetics"

Anticipated future required readings: essays by Levy, Borenstein and Pearson, Nourbakhsh, Warwick, and Scheutz in *Robot Ethics.*

Other suggested readings:

Allhoff, Fritz, Patrick Lin, James Moor, and John Weckert, 2009. *Ethics of Humon Enhoncement: 25 Questions & Answers.* A report commissioned by US National Science Foundation. Accessible at http://www.humanenhance.com/NSF_report.pdf
Allhoff, Fritz, and Patrick Lin (eds.), 2008. *Nonotechnology & Society: Current ond Emerging Ethicol Issues.* Dordrecht: Springer.
Allhoff, Fritz, Patrick Lin, and Daniel Moore, in press. *Whot Is Nonotechnology ond Why Does It Motter?: From Science to Ethics.* Hoboken, NJ: Wiley-Blackwell.
Beauchamp, T., and J. Childress, 1994. *The Principles of Biomedicol Ethics.* New York: Oxford University Press.
Bekey, G.A., Ambrose, T., Kumar, V., et al. 2008. *Robotics: Stote of the Art ond Future Chollenges.* London: Imperial College Press.
Breazeal, C., 2002. *Designing Socioble Robots.* Cambridge: MIT Press.
Bringsjord, S. (forthcoming). "God, Souls, and Turing: A Defense of the Theological Objection to the Turing Test," from University of Reading, March 2009 symposium on the Turing Test, to appear in *Symposium on the Turing Test*, ed. Mark Bishop. Available at http://kryten.mm.rpi.edu/sb theo obj tt 051909.pdf

Brooks, R., 2002. *Robot: The Future of Flesh ond Mochines*. London: Penguin.
Dario, P., and B. Hannaford, A.Takanishi, 2008. "Guest Editorial Special Issue on Biorobotics," *IEEE Tronsoctions on Robotics*, 24/ 1, Feb. 2008, 3-4.
Foerst, Anne, 2001. "Commander Data: A Candidate for Harvard Divinity School?," in Arvind Sharma and Harvey Cox, eds. *Religion in o Seculor City: Essoys in Honor of Horvey Cox*. Harrisburg, PA: Trinity Press International.
Floreano, D., and C. Mattiussi, 2008. *Bio-inspired Artificiol Intelligence*. Cambridge: MIT Press
Kurzweil, Ray, 2005. *The Singularity Is Neor: When Humans Tronscend Biology*. New York: Viking Adult.
Matarić, M.J., 2006. "Socially Assistive Robotics." *IEEE Intelligent Systems*, 21(4), Jul/Aug 2006, 81-83.
Murphy, Robin, and David Woods, 2009. "Beyond Asimov: The Three Laws of Responsible Robotics," *IEEE Intelligent Systems*, 24/4, July 2009, 14-20.
Sandel, Michael J., 2007. *The Cose Agoinst Perfection*. Cambridge, MA: Belknap Press
Siegwart, R., and I.R. Nourbakhsh, 2004. *Introduction to Autonomous Mobile Robots*. Cambridge: MIT Press.
Warwick, K., and S. J. Nasuto, 2006. "Rational AI: What does it mean for a machine to be intelligent?", IEEE Instrumentation and Measurement Magazine, Vol.9, Issue.6, pp.20-26, Dec. 2006.
Yang, Y. & Bringsjord, S. (forthcoming). *Mentol Metologic: A New, Unifying Theory of Humon ond Mochine Reosoning* (Mahwah, NJ: Lawrence Erlbaum).

Sample exercises, questions and problems:

1. Explain the therapy-enhancement distinction, and why (or why not) it makes a moral difference. Would the advent of nanobot medicine and/or rampant cyborgization make any difference?

Potential Answer (for discussion): The therapy-enhancement distinction consists in the understanding that medical treatments that attempt to restore 'normal health' are therapy, whereas those that attempt to improve upon (enhance) 'normal health' are 'enhancement.' The long-standing presumption has been that therapy is morally justified in a way that enhancement is not. But trying to make this understanding more precise immediately creates difficulties: for example, is 'normal health' specific to the individual, or is it 'species-typical'? For instance, if we try to 'equalize capabilities' to a 'species-typical' level across all humanity, that would lead to a vast increase in permitted treatments and amount to enhancement for many individuals whose own 'normal' health is subtypical - e.g., plastic surgery would then be 'therapy' for anyone below the 50th percentile of physical beauty. Likewise, an actress who broke her nose, but nonetheless remains in the 95th percentile of physical beauty, would have her nose reconstruction considered 'enhancement', even if it merely gave her back her previous (good) looks. But making it specific to the individual brings up difficulties of its own - for example, suppose someone 'naturally' has a misshapen nose and jaw; then plastic surgery would be 'enhancement' on the individual-specific understanding, whereas it would be 'therapy' for the beautiful actress. Or, consider a common treatment for children suffering from extreme shortness - would giving them HGH be 'therapy' or 'enhancement'? On the 'species-typical' approach, it is clearly therapy; but if one's parents are short and hence the extreme shortness is genetic, then giving the child HGH presumably would be 'enhancement' - but HGH would be considered therapy for another child, exactly the same age and height, suffering from the same social problems and expecting the same outcome from the HGH treatment, but who happened to have parents of normal height and a glandular problem.

In short, the concepts of 'normality' and 'disease', and the related therapy/ enhancement distinction are ineluctably value-laden and context relative, and so a social construction, one that depends on one's own history, the results of one's environment and the 'genetic lottery' and other criteria outside of one's control - and hence morally dubious. Defending the therapy/ enhancement distinction as morally important usually requires a belief that 'equality of opportunity' (*after* one is born - given one's genetic makeup as a given), not equality of capability, is morally crucial. But it is not clear that a robust ethical theory can defend such a view, particularly when it reinforces unjust distributions of social burdens based on accidental inequalities of capabilities - when such unjust burdens are the fault of Fate or Nature or God, and not the individual. Further, as technology improves and more and more medical interventions become widely available and cheaper, inevitably social consensus on how and where to draw the therapy/ enhancement distinction changes and what were once 'enhancements' come to be seen as needed therapy. There is no reason to think nanobot interventions and cyborg technology will eventually be seen as any different - as they become routine and social disfavor and competitive disadvantage accrue to those without the technologies, people will clamor for them as needed therapy, not as exotic enhancements.

Or: it is immoral to even attempt to enhance ourselves, because such attempts do not respect human dignity and see our innate human qualities as a product to be commoditized, not as gifts from a Creator. We should accept our imperfections and give thanks for our human nature, rather than engaging in the Promethean quest to reinvent ourselves - our very nature - in search of perfection. Such attitudes threaten to turn ourselves - and even worse, our children - into designer commodities, with undue, perhaps crippling pressure due to future expectations raised by parental expectations of a flawless 'product.' Instead, we all - and especially children - should have an 'open future'. Finally, because such treatments are new and untested, and carry potentially grave risks, all such attempts at enhancement violate (one version of) the precautionary principle: "When an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically."

2. There is a rapid, ongoing development in *socially assistive robotics* (SAR) in healthcare, which has robots performing some of the assistive and rehabilitative roles heretofore done e.g. only by human occupational and physical therapists. But SAR raises worries: will it cause patients to feel a decrease in autonomy? Will patients form emotional attachments with or have unrealistic expectations of the robots? How ought a robot (an expensive piece of equipment) act if threatened by the user?

Potential Answer (for discussion): Inevitably, humans have a tendency to anthropomorphize objects in their lives with which they have substantial relationships, and that tendency is only exacerbated when humans are in relations of substantial dependence on such objects. Many of us name our cars and get angry with it when it 'refuses to start'; or attribute intentions and rational desires, not merely to our pets, but even to our vacuum cleaners, refrigerators and power tools, even having discussions with them. It thus seems inevitable that humans will attribute the 'intentional stance' to (and form emotional bonds with) robots with which they commonly interact, and the dependent relation humans will have with assistive robots will only strengthen that tendency. If the robots break or are repossessed, there could easily be severe emotional repercussions, as well as interpersonal dilemmas and even mental illness. (Cf. the movie *'Lars and the Real Girl'*.)

Meanwhile, rehabilitation centers will be loathe to allow patients to strike and damage or even break expensive robots, so there will be a temptation to create robots with the capacity to defend themselves. If so allowed, what rules should dictate when a robot can defend itself against attack? How can we avoid accidental harm from such robots?

3. Will robotic sexual partners (sexbots) be a good thing? Will they further depersonalize and commodify sex? Will they end or at least greatly reduce prostitution and sexual slavery?

Potential Answer (for discussion): Inevitably, as robots become more and more like humans, the 'uncanny valley' will gradually be overcome and many humans will increasingly take sex with robots for granted. It is likely that the widespread availability and increasing acceptance of sexbots will greatly dampen the demand for human prostitutes and sexual slaves and lessen the incidence of STDs, but also further problematize the notion of monogamy (is sex with a robot cheating on your spouse/ significant other?), deepen the divide between love, commitment, and sex (why make commitments or fall in love if sex is perpetually available with robots? Or, can we fall in love with, commit to, and even marry robots?), and ineluctably create new difficulties for the institution of marriage and its cultural expectations as either a social ideal or a social norm, and possibly further transform the relationship between the biological sexes in ways hard to predict. For example, could the acceptance of robot marriage lead down a slippery slope to legalized polygamy, involving a (wo)man, another (wo)man, and a robot - or 2, or 3? Does the prospect of robot marriage have any implications for the debate over gay marriage?

Week 9-10:

Global issues - environment, energy, crime, terrorism, and military uses

We will discuss the issues concerning the impact of robots on the environment and energy production, including issues in resource depletion and distribution, the use of robots in producing green energy and solving problems of climate change, including geoenegineering, the value (intrinsic or otherwise) of artificial ecosystems that include robotic components, the possibility of robots playing a major role in providing cheap or safe energy (e.g., robot-manned nuclear fission or fusion plants for safety reasons) and related issues.

We also will examine issues involved in the growing use of robots in war and in anti-terrorist activity, including just war issues (including both *jus in bello* and *jus post bellum,* as well as *jus post bellum* issues of policing and peacekeeping), especially the commonly discussed problem for robots to reliably discriminate between combatant and non-combatant. In addition, we will examine safety issues (e.g., friendly fire) and unit cohesion issues for robot combatants and their human co-workers, and resulting ethical qualms about mixed human-robot deployments, as well as issues involving the relative permissibility of robots being equipped with lethal versus nonlethal weaponry, international robotic proliferation concerns, and of course the oft-fictionalized rampancy issues, including the various 'Terminator' or other worries about out-of-control armed robots. A natural extension of the ethical concerns over military robots concerns their likely extension to the civilian world of policing, especially with the even greater worries in civilian contexts over (mis)use of (lethal) force and discrimination issues between criminal and non-criminals, as well as general human (mis)trust of robotic policing

and its implications for the social life of humans and the future of crime. Will humans eventually welcome or fear a world in which public surveillance of activities by the authorities is pervasive - and robotic?

Required readings: Ron Arkin, *Governing Lethol Behovior in Autonomous Robots*
Peter Asaro, "How Just Could a Robot War Be?"
Robert Sparrow, "Killer Robots"
Chapters 5-6, 10-14, and 18-19 of *Wired for War.*
Chapter 4 of *Autonomous Military Robots.*

Anticipated future required readings: essays by Arkin, Powers, Fukuda, and Sharkey in *Robot Ethics.*

Other suggested readings:

Broome, J. "The Ethics of Climate Change" *Scientific Americon* (June 2008): 26-102.
Carter, Luther J. "Environmental Warfare Treaty" *Science*, New Series, vol. 205, no. 4407 (August 17, 1979): 674.
Computer Professionals for Social Responsibility, 2008. "Technology in Wartime" conference, January 26, 2008, Stanford, CA. Accessible at http://technologyinwartime.org/
CNN 2010. "Al Qaeda claims responsibility for CIA attack," CNN January 7, 2010. Accessible at http://www.cnn.com/2010/WORLD/asiapcf/01/07/afghanistan.cia.al.qaeda/index.html
Hadhazy, A. "Will Space-Based Solar Power Finally See the Light of Day?" Scientific American (April 16, 2009). Available at http://www.scientificamerican.com/article.cfm?id=will-space-based-solar-power-finally-see-the-light-of-day
Institute of Electrical and Electronics Engineers, 2008. "International Conference on Advanced Robotics and its Social Impact" conference, August 23-25, 2008, Taipei, Taiwan. Accessible at http://arso2008.ntu.edu.tw/
Kunzig, Robert, 2008. "A Sunshade for Planet Earth" *Scientific Americon*, vol. 299, issue 5, (November 2008): 46-55.
Royal United Services Institute (RUSI) for Defence and Security Studies, 2008. "The Ethics of Autonomous Military Systems" conference, February 27, 2008, London, UK. Accessible at http://www.rusi.org/events/past/ref:E47385996DA7D3/
Shachtman, Noah, 2007. "Robot Cannon Kills 9, Wounds 14", *Wired.com,* October 18, 2007. Accessible at http://blog.wired.com/defense/2007/10/robot-cannon-ki.html.
Sharkey, Noel, 2007. "Robot Wars are a Reality", *The Guardian* (UK), August 18, 2007, p. 29. Accessible at http://www.guardian.co.uk/commentisfree/2007/aug/18/comment.military
Shim, H.B. 2007. "Establishing a Korean Robot Ethics Charter". Accessed 21 July 2009: www.roboethics.org/icra2007/.../slides/Shim_icra%2007_ppt.pdf
Veruggio, G. (ed.), 2006. *EURON Roboethics Roadmap*, EURON Roboethics Atelier, Genoa, Feb. 27-Mar. 3, 2006.
Walzer, Michael, 2006. *Just ond Unjust Wors: A Morol Argument with Historical Illustrotions.* New York: Basic Books.
Zubrin, Robert M. and Christopher P. McKay, 1993. "Technological requirements for terraforming Mars," AIAA, SAE, ASME, and ASEE, Joint Propulsion Conference and Exhibit, 29th (Monterey, CA: June 28-30, 1993). Accessible at http://www.users.globalnet.co.uk/~mfogg/zubrin.htm

Sample exercises, questions and problems:

1. Will proposals for deploying autonomous robots in battle be eventually done in by the fundamental impossibility to solve the problem of discrimination? Or is discrimination itself an outmoded moral constraint, a remnant of age-old assumptions about military tactics that new technologies and methods of warfare and terrorism have rendered obsolete?

Potential Answer (for discussion): The advent of the 'war on terror' and its difficulties in discerning friend from foe when opponents wear no uniforms and are often indistinguishable from civilians, while also requiring a battle for the 'hearts and minds' of indigenous populations in order to stem terrorist recruitment and gain valuable intelligence, all make the *jus in bello* principle of discrimination more important in warfare than ever. The wanton killing of civilians does more to harden popular opinion against US military activities in Afghanistan and Iraq than almost anything else. Further, a rich history of customary international law has been gradually built up and accepted by warring parties through the ages, and international institutions have gradually come to exist which can enforce them. Throughout history, as a matter of honor, prudence, strategic foresight, or even mercy, there have been *jus in bello* restrictions that acquired both moral and legal weight.

This trend toward seeing war as an activity with rules or virtues that sanction proper and improper behavior has only gained strength as states have acquired an institutional professional military, especially one independent of those making *jus ad bellum* decisions. Professional soldiers have a code of conduct that details their proper and improper functioning in their various roles, just as other professions do. They cannot be meaningfully held responsible for decisions by politicians over which they have no control; but they can be held responsible for performing their roles in war in a way the international community recognizes as legitimate, and thereby responsible for avoiding illegitimate means of performing those roles. If and when autonomous robotic soldiers can do so *at least as well as* human soldiers (even in restricted circumstances, e.g., a kill-box), then the use of lethal military robots will satisfy the problem of discrimination in a morally acceptable way.

2. Another objection (e.g., Asaro 2008) to the use of robots in war is that the development of military robots seems to fail a *jus ad bellum* test, because they would embolden political leaders to wage war; robotic soldiers would lower barriers to entering a war, since they would reduce casualties among human soldiers and therefore also a significant political (and perhaps financial) cost, perhaps making war 'risk-free' (at least for humans).

Potentiol Answer (for discussion): First, to claim that robots have bad consequences for the tendency to declare war is a consideration that first would be handled by the nonconsequentialist requirements for declaring a just war: using robots or not makes no difference as to whether the war is (a) in self-defense, (b) proportionally achieving a good greater than the evil of war, (c) a last resort, and so forth.

Second, if any technology (from better armor to longer-range missiles) makes it easier to enter a war to the extent that it reduces risks on one side, these objections seem to imply that one should not make any improvements in the way one prosecutes a war and, indeed, should return to more brutal methods (e.g., bayonets). But surely this is ridiculous or, at the least, counterintuitive. Indeed, the increasing horrors of war have reinforced the need for *jus ad bellum* and *jus in bella* restrictions, not undermined them; for instance, the advent of ever greater precision in weapons targeting (e.g., 'smart' or laser guided bombs) has reinforced, not weakened, the ability and commitment to *jus in bello* discrimination. Similarly, it is likely that the advent of

military robots will cause further sophistication in such just-war considerations and make war ever more ethically waged.

3. Robert Sparrow has a critique of the very possibility of morally deploying autonomously-functioning military robots [Sparrow 2007]: his contention is that such robots can never be morally deployed, because no one—neither the programmer, nor the commanding officer, nor the robot itself—can be held responsible if it commits war crimes or otherwise acts immorally. No one can reasonably be said to give morally responsible consent to the action an autonomous robot performs; so no one is responsible for the risk such autonomous robots pose, and thus it is immoral to use them. If war crimes can happen with impunity, then the war itself cannot be morally justified.

Potential Answer (for discussion): One can imagine a response to Sparrow as follows: We find it morally permissible for military parents to raise their child as destined for the military, to indoctrinate them as a soldier from infancy, and to place those expectations on them in their earliest training. Once they become autonomous adults, it is expected that they will volunteer for service—but they remain autonomous, and it is possible (however psychologically unlikely) that they will choose a different path. If it is morally permitted to raise human children with such expectations, and to accept the children so indoctrinated into voluntary military service, why it would be wrong to likewise train an autonomous robot and place it into active duty?

But some may object as follows: A human child will develop free will, and the above analogy fails given a robot's lack of true Kantian autonomy. That is, the robot could never have the sense of self or the libertarian free will of humanity; they have merely instrumental (means-ends, goal-oriented) rationality. Humans are not robots and have a different kind of autonomy than robots ever could.

Leaving aside those who think humans really are merely complex robots [e.g., Dennett, 1995], there is a simpler solution to Sparrow's objection to the 'in-principle' immorality of deploying autonomous (in the sense of self-regulating) robots. For all military robots, including those with this minimal self-regulating level of autonomy, we normally assume what the literature terms a '*slave morality*', i.e., they have no ends of their own, but their goals are all in service of the goals of someone else—in this case, the military and, more specifically, whoever commands them and gives their orders. Such 'slave' robots cannot create their own laws or final goals; they are not ultimately makers of a *self*, but followers of 'life goals' others have imposed, and their own freedom only comes in the mere means they choose to realize those ends.

Such military robots, whatever their other decision-making capabilities, thus lack full Kantian autonomy, and so cannot be held responsible for their actions under traditional deontological, natural law, or virtue ethics theories. Inasmuch as *jus in bello* restrictions most plausibly depend on one of those approaches, robot risk and responsibility as a function of consent thus becomes a non-issue. This realization helps to rebut the central contention of Sparrow's critique of autonomously functioning military robots; given a robotic 'slave morality', the commanding officer (or a deceitful programmer) is to blame for war crimes, not the robot who was 'only following orders'.

Of course, one immediate objection could point to Nazi soldiers who committed war crimes and pleaded that they were only following orders; and we can imagine back in the day that George Washington's slaves might

have been held responsible for following immoral orders. But there is a crucial difference between a human soldier or even a human slave and a robot programmed with a slave morality: the human person, whether a soldier or a slave, is presumed to have the ability to disobey orders, even if the punishment for doing so would be harsh. But a properly programmed robot with a 'slave morality' literally could not disobey orders intentionally—it would do so only by mistake. And in ethics, as long as someone is not free (in whatever is the relevant moral sense of 'free') to disobey their orders, they cannot be blamed. For robots, unlike humans, that can be a matter of correct programming.

5. Contact

Keith Abney
California Polytechnic State University
Senior Lecturer, Philosophy Dept.
1 Grand Avenue
Building 47, Room 37
San Luis Obispo, California 93407
Email: kabney@calpoly.edu

Rapid Assembly Disaster Response Shelters Made From Recycled Plastic Materials

Project Investigators:

Edmond Saliklis and Robert Arens
Architectural Engineering and Architecture
California Polytechnic State University
San Luis Obispo, CA

June, 2010

To: C3RP officials
From: Edmond Saliklis and Robert Arens
Re: Final Report of our C^3RP funding

EXECUTIVE SUMMARY

The pressing humanitarian need for post-disaster sheltering of distressed populations has called our design team into action. Earthquakes, tsunamis and hurricanes regularly destroy thousands of homes and the need for emergency shelter is real and enormous. We have undertaken a research program in response to this devastating problem. Funds from C^3RP supported Phase I of this project. Attainments from Phase I will be described herein. Appendices contain the peer reviewed papers and the NSF grant proposal which emanated from this initial seed funding. Five different sustainable, rapidly assembled emergency shelter prototypes have been designed and constructed by the research team and seven important scholarly attainments have arisen from this research. The scholarly attainments are:

- Saliklis, E. and Arens, R. "Rapidly Assembled Emergency Shelters Made From "Green" Materials", (2009) *proceedings of the 50th IASS Symposium*, Valencia, Spain.
- Arens, R. and Saliklis, E. "Interdisciplinary form finding in the development of a sustainable emergency shelter (2010) *proceedings of the International Conference on Structures and Architecture*, Guimaraes, Portugal.
- Arens, R and Saliklis, E. "Versioning: Interdisciplinary Design of a Flat Pack Emergency Shelter" (2010) *proceedings of the 2nd International Conference on Design Education*, Sydney, Australia.
- Christine Carpenter Senior Project (2009) "Effects of Accelerated Weathering on the Flexural Properties of Kenaf-Reinforced Polypropylene Composites", Materials Engineering Department Cal Poly.
- Christine Carpenter Poster Session TMS2010 "A Materials Investigation of the UV Degradation of Polypropylene-Kenaf ECO-Composites", Seattle, WA.
- Jorien Baza Master's Thesis (2010) "Nonlinear Material Modeling of a Sustainable Structural Material", Architectural Engineering Department, Cal Poly.
- Saliklis, Arens NSF Grant Proposal (2010) "Design of a Rapidly Assembled Emergency Shelter: Material Modeling and Structural Architecture", $276,000 grant submitted to the Structural Materials and Mechanics Division of the National Science Foundation, January 2010.

OVERVIEW

This research project used an interdisciplinary approach to develop a rapidly deployed emergency shelter. The team for this project was led by professors in the Department of Architecture, and the Department of Architectural Engineering, with assistance from the Department of Materials Engineering at California Polytechnic State University, as well as economists from industry. Students from all three Cal Poly departments participated fully in this research. Our overarching goal was to explore the project on three global levels:

I) the level of analyzing and modeling various sustainable materials
II) the level of fabricating and constructing a habitable, deployable structure that can be rapidly assembled
III) the level of interdisciplinary form finding using architectural and structural engineering tools

The three levels of investigation were intertwined. Material modeling allowed for accurate characterization of the eco-friendly composite lumber we used. These material models were ultimately incorporated into the finite element computer programs used to simulate the shelter's response to lateral loads. Fabrication of scale models allowed for rapid improvements in constructability and lateral stiffness detailing. Digital modeling and interdisciplinary integration of design strategies were informed by state-of-the-art practices proposed by design theorists. The three global levels of investigation are summarized as follows:

I) The first level of research was to quantify the performance of several sustainable materials and to develop accurate mathematical material models which captured their behavior. We quantified the performance of a composite material made of recycled polypropylene reinforced with the agricultural fiber kenaf, and we also studied other sustainable materials such as corrugated recycled plastic panels.

II) The second level of our research was that of the detailed design of a habitable, deployable structure made of sustainable, reusable materials; a shelter that can be made, shipped and constructed simply and economically. Such designs were prototyped and experimentally tested in our structural laboratory. We used architectural and structural engineering software and state-of-the-art digital fabrication tools to create these prototypes.

III) The third level of our research was that of structural form-finding, wherein thin shell technology informed the design of a small scale shelter that it is light yet stiff. We developed a research method that optimized the design process by means of digital models which explicitly linked structural analyses to architectural design. We found resonance between our interdisciplinary methodology and ideas proposed by design theorists in the literature.

The team crafted the following set of immediate goals for the shelter design:

1. Efficiency: The design should employ only materials that conserve natural resources and reduce waste, and the minimum amount of these materials should be used. Additionally, the design should require a minimum amount of labor during its creation phase and erection phase of the shelter.
2. Lightness: The design should avoid excessive weight that would waste fuel during the shipping phase or human labor during the erection phase of the shelter.
3. Packability: The design should create its own shipping crate so as to avoid redundant parts. Additionally, the crate should have as small a footprint as possible to facilitate close packing during the shipping phase.
4. Constructability: The design should have as few components as possible and simple connections that can be constructed by hand or with basic tools and little labor. No ladders or scaffolding should be necessary during the erection phase.

5. Adaptability: Since disasters can occur anywhere and everywhere, the design should be adaptable to a wide range of climates. To accommodate this, the design should have a combination of universal components and climate-specific components.
6. Reusability: The design should promote either the reuse of the shelter at other disaster sites, or the re-purposing of the shelter's components/materials in rebuilding efforts of permanent housing.

The form finding process used to develop the emergency shelter was led by architects and structural engineers, disciplines closely aligned yet with distinct approaches to design. From the start, the leaders and student research assistants on this project used best practices of their respective traditions to combine the objective with the subjective, logical processes with intuitive pursuits, and analytical digital models with associative physical prototypes. This approach allowed us to pursue the project goals as a team, yet it also allowed us to iteratively flow between methodologies best suited to move the project forward.

We developed five prototypes throughout this research program, and we documented the design process details in the peer-reviewed international conference publications attached to this report. These prototypes have generated enthusiasm among our industrial partners and we are confident that the project can be expanded and commercialized. A key element connecting architectural design, structural engineering design and full-scale prototype fabrication was the use of three-dimensional visualization software. Such software allowed us to rapidly envision design possibilities and it aided our mathematical analyses as well. Another key advantage of the 3D visualization software was that it seamlessly interfaced with computer numerically controlled fabrication equipment used to construct full scale prototypes. The use of computer numerically controlled cutting tools allowed us to create a shelter that is rapidly assembled by hand with no tools needed to connect the members, since the precise cuts allowed for friction-fit connections. The shelter can be assembled in one hour and it comes to the site in its own shipping crate which serves as the floor of the shelter. Figure 1 shows a time lapse of the shelter's construction on Dexter Lawn at Cal Poly.

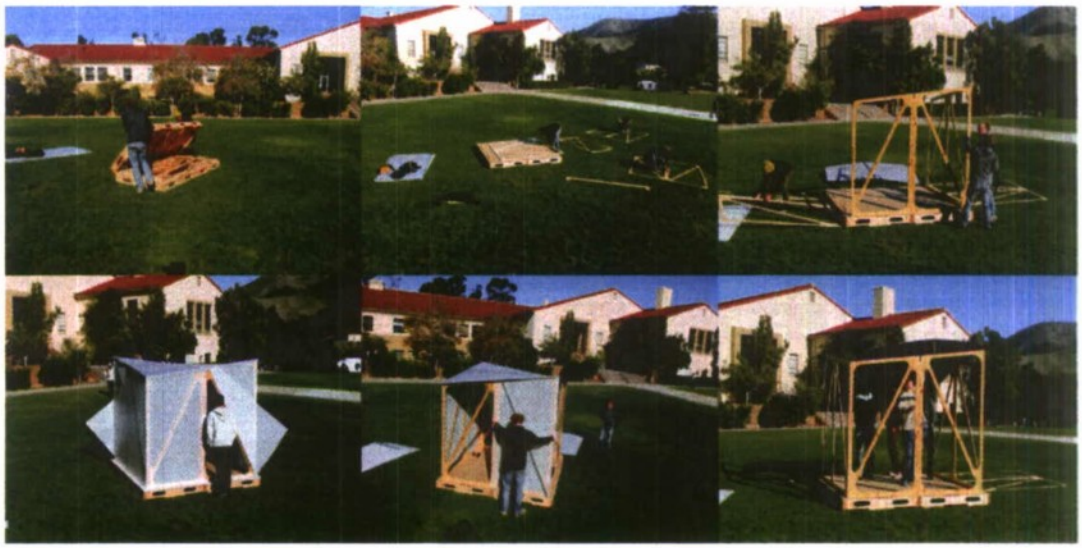


Figure 1. Assembly of shelter

I) MATERIAL MODELING STUDIES

Two large material modeling studies were undertaken, one by Ms. Christine Carpenter of the Materials Engineering Department, and one by Mr. Jorien Baza of the Architectural Engineering Department. Ms. Carpenter pursued the study of the effect of accelerated aging on the bending stiffness of our material. This culminated in her Senior Project entitled "*Effects of Accelerated Weathering on the Flexural Properties of Kenaf-Reinforced Polypropylene Composites*". Ms. Carpenter's Senior Project was deemed ***Best Presentation*** from the Materials Engineering Department's Advisory Board in June of 2009.

Ms. Carpenter went on to present her research at the TMS2010 Convention in Seattle WA in February 2010. Her presentation entitled "*A Materials Investigation of the UV Degradation of Polypropylene-Kenaf ECO-Composites*" focused on quantifying the material mechanisms which cause our composite material to lose some bending stiffness in the presence of extended UV radiation.

Ms. Carpenter used an accelerated aging chamber to assault the composite material with UV light. Specimens were tested for 2000 hours, and visual inspection of the effects on color (See Figure 2) were coupled to meticulous infrared spectroscopy studies.

0 hours 500 hours 1000 hours 1500 hours 2000 hours

Figure 2. Visual assessment of color degradation as a function of UV aging

Her studies quantitatively assessed the bending stiffness of the composite material as a function of UV aging. She used American Society of Testing and Materials standards to gather her data, shown in Figure 3.

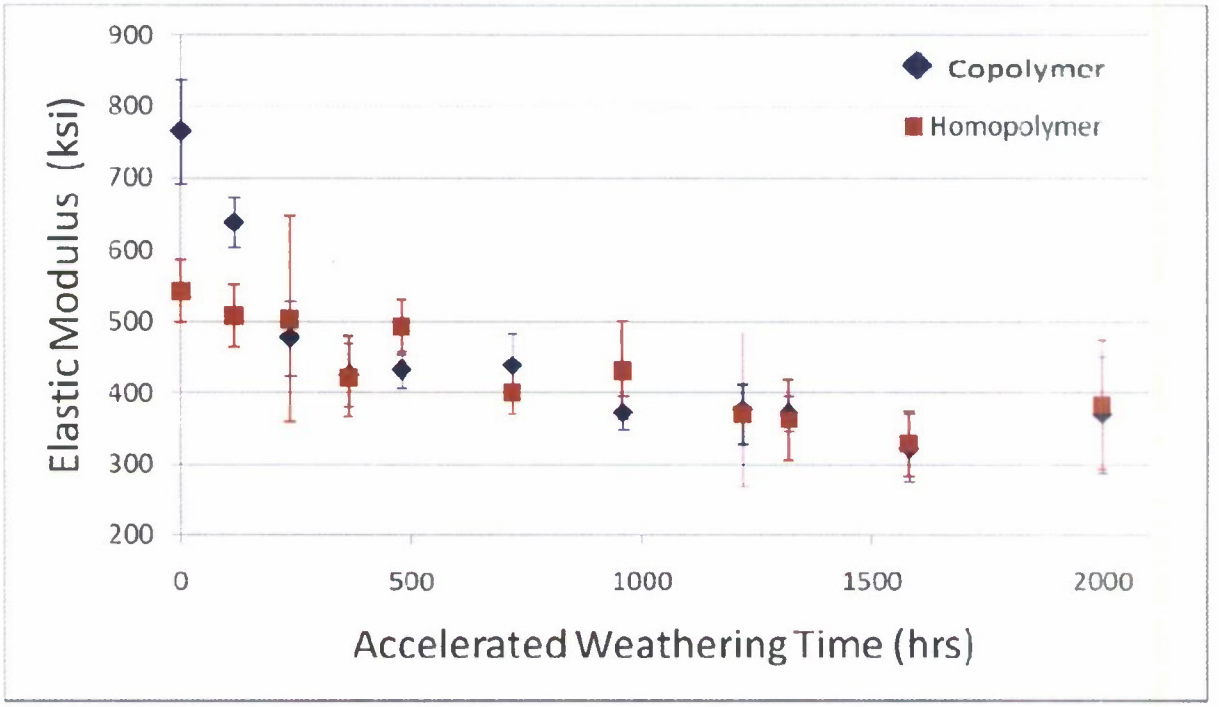


Figure 3. Degradation of bending stiffness over accelerated weathering time

Ms. Carpenter was able to discern the molecular mechanism which explained the degradation of stiffness of the material over time. Her experimentally backed theory is that bonds between the cellulosic chains in the kenaf reinforcement were being broken, and bonds between the polymer and the kenaf were also being broken by UV radiation. Her infrared data supports this theory as shown in Figure 4.

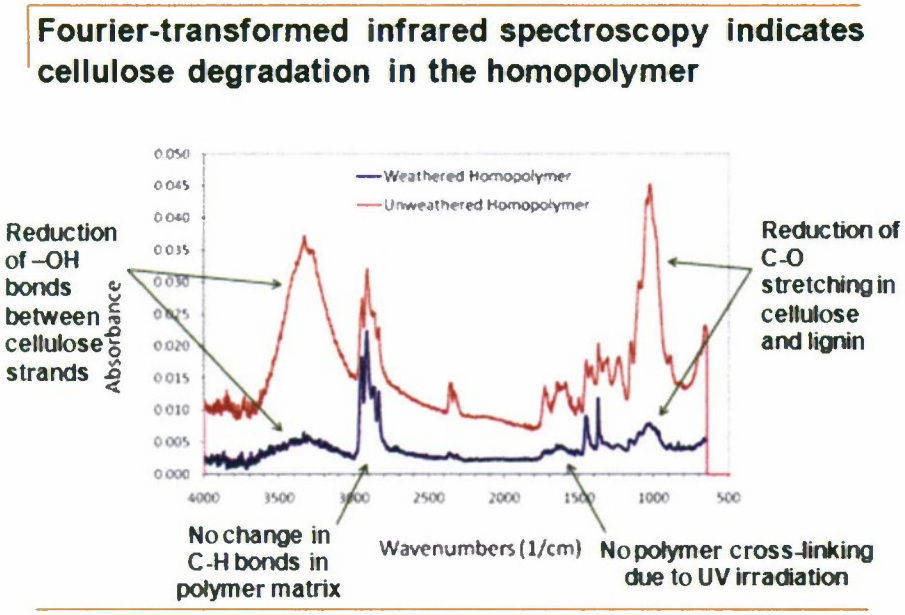


Figure 4. Infrared evidence of bond breakages

Mr. Baza pursued an extensive experimental and mathematical modeling study of the composite material. This study culminated in his Master's Thesis from the Department of Architectural Engineering. He successfully defended his thesis in May 2010, and he will submit it to the Kennedy Library's Digital Commons in June 2010. The main focus of his work was to develop a model that can accurately predict the strain of this material given a state of stress. The problem is complex because the material is nonlinear, i.e. doubling the stress creates more than a doubling of strain. Mr. Baza was able to develop a mathematical model that describes this material behavior accurately and he implemented his model into a general purpose finite element program that was used to model the prototype shelters.

The central idea to Mr. Baza's work was that the strain energy of the composite material can be captured experimentally. He performed dozens of axial load tests and shear load tests to gather a body of data. He then used fundamental engineering mechanics principles to model the behavior of the material in terms of the picture presented by strain energy. To supplement his experimental data, he ran extensive finite element simulations of the axial and shear experiments at various angles θ to the primary extruded board axis as shown in Figure 5.

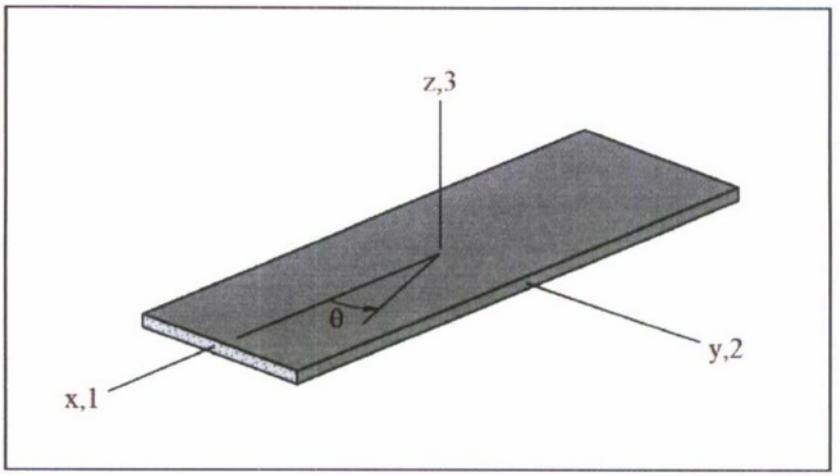


Figure 5. Angle θ investigated in axial load tests

The model he developed captured the nonlinear material response with a simpler bilinear curve. The bilinear curve is readily incorporated into finite element programs. He used these programs to study various shelter configurations. Figure 6 demonstrates the efficacy of his model for the two principal orientations (1 and 2) as well as the shear response (1-2).

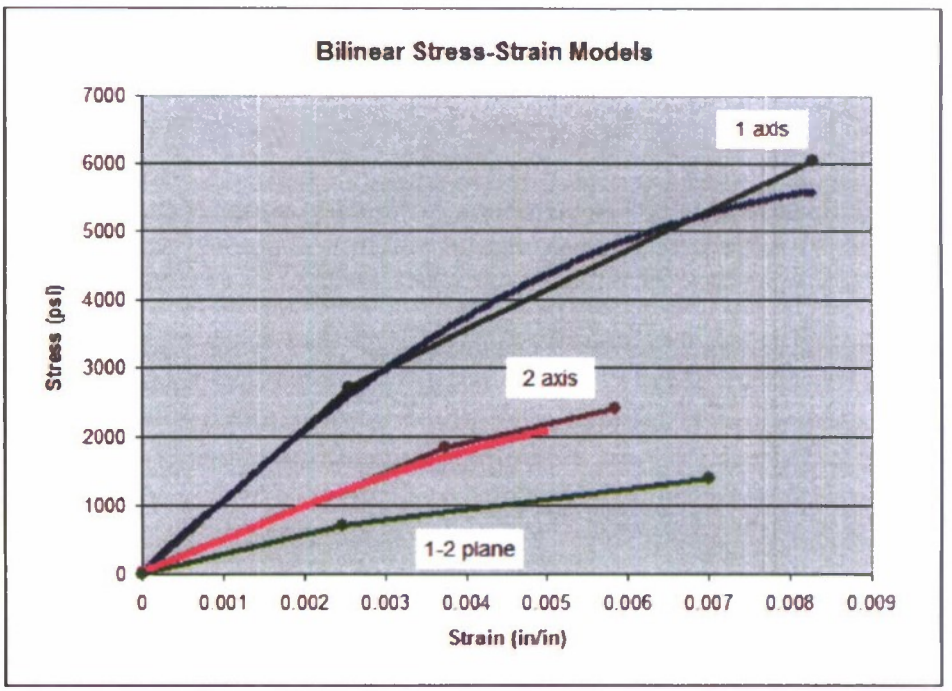


Figure 6. Bilinear model strong (1) axis, weak (2) axis and shear (1-2) axis

Figure 7 shows a typical finite element model of a shelter wall study he conducted. In 7a, the finite element mesh of one wall is shown, and in 7b, the results of an ever-increasing lateral load pushing on the wall. Notice the bilinear behavior of the finite element results matches very well to standard pushover hand calculations.

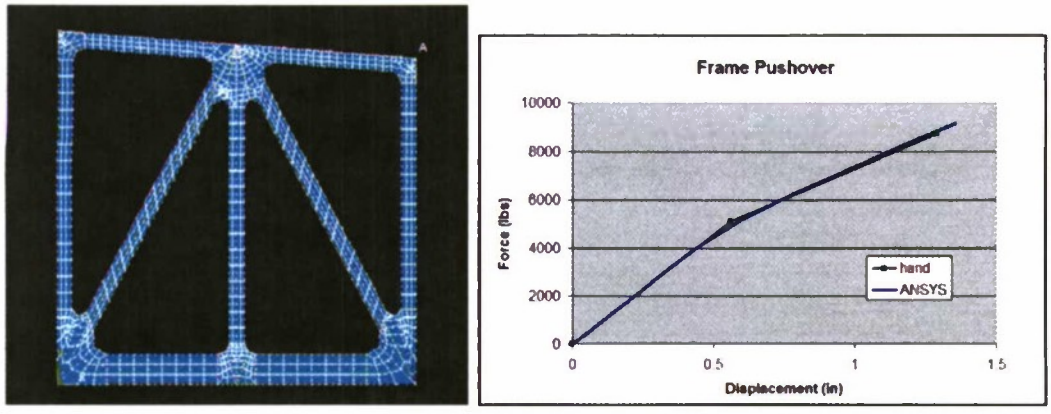
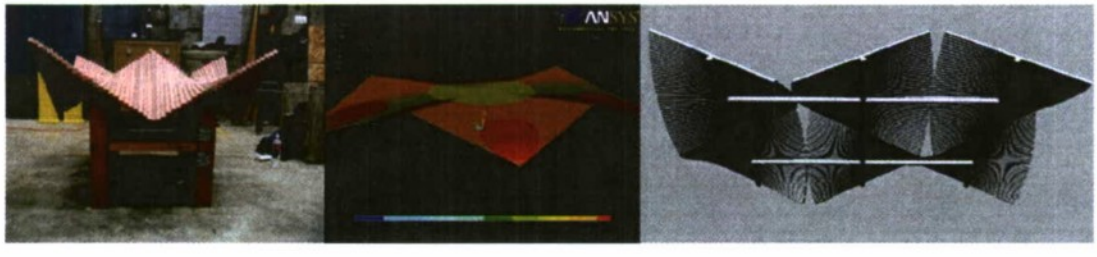


Figure 7. a) Finite element model using bilinear material **b)** numerical results of lateral load test

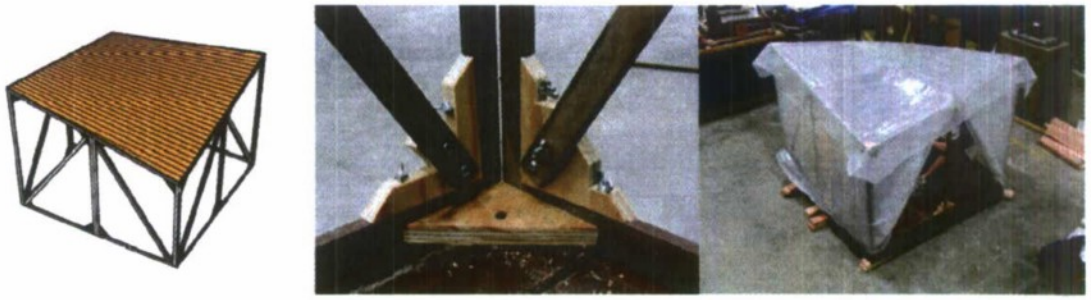
II) EXPERIMENTAL PROTOTYPES

Five different prototypes were constructed. Prototype 1 was an exploration of a stiff roof designed to take the shape of a hyperbolic paraboloid (hypar). Our initial thinking was to take advantage of the unique feature of a hypar, namely that it can be generated from straightline segments. This was useful for us since the eco-friendly Kenaf/Polypropylene material we were using was extruded in straight line segments. Figure 8 shows Prototype 1 in our structural lab, along with a finite element model of the hypar, and a 3D visualization model.



Figures 8a-c. Prototype 1 showing the physical prototype (a, left); FEM visualization (b, center); and digital model with header beams (c, right)

Prototype 2 was a half-scale study of a complete shelter, including walls and a floor. We explored straight segments connected with simple hardware for the walls and various weatherproofing skins on the hypar roof. Figure 9 shows aspects of Prototype 2. Note that this an evolution of Prototype 1 as seen in the hypar roof of Figure 8.



Figures 9a-c. Prototype 2 showing the digital model (a, left); frame-to-floor connection (b, center); and the frame clad with plastic sheeting (c, right)

Prototype 3 eliminated the somewhat cumbersome hardware of the wall segments and replaced the individual segments of the walls with continuous panels cut by a computer numerically controlled cutter (CNC). We used finite element analyses to aid in the design process, seeking the lightest and stiffest configuration. Figure 10 shows four different CNC cut wall configurations and Figure 11 shows the walls as part of a 3D finite element model of the shelter. This seamless integration of 3D visualization software, 3D finite element modeling and real time CNC fabrication formed the basis of the peer-reviewed international conference paper accepted for presentation and publication in Portugal in July 2010.

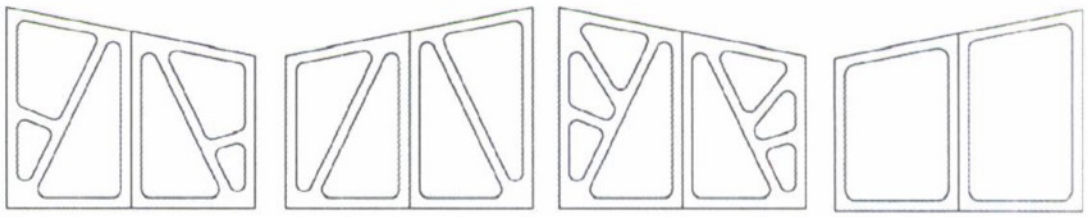


Figure 10. Four different CNC cut wall panel configurations tested in prototype 3

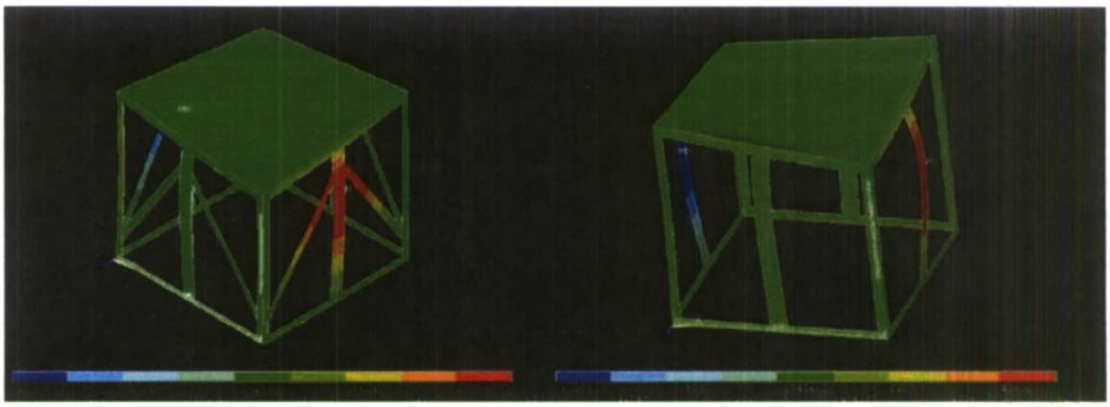


Figure 11. Two finite element models of shelters

Prototype 4 synthesized the best features of the previous models and incorporated substantial design improvements in packability. The floor of the shelter was modified to allow it to act as a clamshell packing crate, able to carry the entire shelter to the disaster site. We filmed the construction of the shelter

and found it could be assembled with no tools in about one hour. Figures 12 and 13 shows images of Prototype 4. This prototype was also seen in Figure 1.

Figures 12a-c. Refinements in prototype 4 included a more functional base (a, left), interlocking, friction-held joints (b, center), and semi-rigid cladding of corrugated polypropylene (c, right).

Figures 13a-c. Using a CNC router to cut all components of prototype 4 (a, left), the floor is used as a packing crate for wall and roof components (b, center), and the interior of prototype 4 (c, right).

Prototype 5 explored using an extremely lightweight honeycombed paperboard panel for the side walls of the emergency shelter. We were able to modify the computer numerically controlled cutter speed to accurately and precisely cut this sustainable material. Figure 14 shows some images from the fabrication of Prototype 5.

Figures 14a-c. Using a CNC router to cut all corrugated paperboard components of prototype 5

III) INTERDISCPLINARY FORM FINDING: STRUCTURAL AND ARCHITECTURAL

We saw this project with its equal emphasis on design, assembly and production as the perfect opportunity to apply a design concept known as "versioning". Versioning is a strategy that utilizes digital tools to combine form finding, the assemblage of materials and the means of fabrication in a single feedback loop that informs multiple iterations. Similar to rapid prototyping used by other disciplines, versioning moves the design process towards a system of vertical integration whereby the designers drive how space is both conceived and constructed. As C. Sharples notes: "Versioning implies the shifting of design away from a system of horizontal integration (designers as simply the generators of representational form) towards a system of vertical integration (designers driving how space is conceived and constructed and what its effects are culturally)." (2002:7) The digital tools we described in this report, namely the 3D visualization tools, the finite element tools and the CNC cutting programs allowed us to apply versioning theory and to compress design, testing, assembly and production into a single cohesive process.

Another design aspect of this project that found resonance with contemporary design theorists is the notion of design as dynamic and non-linear, and not as a process with a beginning, middle and end. Accordingly according to Speaks: "the relationship between thinking and doing becomes more and more blurred so that thinking becomes doing and doing becomes thinking, engendering highly collaborative, interactive forms of practice that are already changing the face of architecture." (2002B:6) Such collaborative and interactive practices within this research project allowed the work to branch out into different disciplines (materials engineering, structural engineering and architecture), it allowed a diverse body of students from three different departments to blend their skills in a constellation fashion, rather than a linear progressive manner. It was fascinating to see the different skill sets of the student researchers interplay with each other under the mentorship of the principal investigators. It is this design process that we seek to expand upon in Phase II of this project.

REFERENCES

Sharples, C. 2002. Architectural Design. vol. 72, no. 5.

Speaks, M. 2002. "Design Intelligence: Or Thinking After the End of Metaphysics." Architectural Design. Vol. 72, No. 5.

The Genomics of Injury-Repair

Principal Investigator:

Trevor Cardinal, Ph. D.
Biomedical and General Engineering
California Polytechnic State University
San Luis Obispo, CA

Abstract

The long-term goal of my research program is to determine the molecular mechanism of impaired vascular reactivity (vasodilation & vasoconstriction) following peripheral ischemic injury or disease. Ischemia is defined as insufficient blood flow and is caused by partial or full occlusion of a major artery, due to traumatic injury or disease (atherosclerosis/thrombosis). This research effort also has implications for limb regeneration, as injury-repair involves cellular processes that are similar to those controlling limb development in the embryo. Although impaired vasodilation in resistance vessels (feed arteries and arterioles) is a commonly observed effect of ischemia, the molecular mechanism (i.e. the proteins) underlying this impairment is poorly understood. Therefore, no effective therapies, aside from exercise, are available to improve vascular reactivity following ischemic insult. This is especially detrimental in peripheral ischemic injury, as skeletal muscle has such a wide metabolic demand that must be matched by appropriate oxygen delivery to maintain normal tissue function and prevent hypoxic pain and/or disuse. To maximize healing from ischemic injury, the mechanisms of impaired vascular reactivity must be determined to ensure that properly functioning resistance vessels are able to matching oxygen delivery to metabolic demand.

Therefore identifying the proteins that are under- or over-expressed in peripheral vasculature following ischemic injury is a critical step in determining the molecular mechanism of impaired vascular reactivity. Although proteomics strategies exist to examine the entire proteome of a given tissue (all protein expressed), these strategies are limited by insufficient tissue mass (if examining individual blood vessels) or non-specific protein contamination (if examining resistance vessels within skeletal muscle tissue). Therefore, the original objective of this project was to examine the transcriptome (all mRNA molecules expressed in a given tissue) with DNA microarrays in feed arteries following ischemic injury. mRNA from tissues of small volume can be amplified for transcriptome analysis and individual mRNA can be selectively amplified, allowing the investigator to "filter" the dominant mRNA from skeletal muscle (contractile filaments) that are not of interest to this project. Further, mRNA expression is generally well correlated with protein expression. It is important to note that arterial occlusion results in reduced blood flow through primary arteries (ischemia) and enhanced blood flow through collateral/by-pass arteries/arterioles (causing collateral enlargement), and both of these processes could impair vascular reactivity. Therefore, we planned to utilize two different ischemic injury models to examine the impact of ischemia (reduced blood flow) and collateral enlargement (enhanced blood flow) on feed artery vasodilation.

In the middle of this project, Cal Poly was awarded Public Health Services Assurance for the utilization of vertebrate animals in research. This precipitated a modification of the project objectives to include functional studies of feed artery vasodilation to complement the examination of feed artery mRNA expression. Further, with the addition of this functional component to the project, we removed the DNA microarray analysis in favor of examining individual mRNA using PCR. Therefore, the final objectives of this project were to:

1. Develop protocols for measuring mRNA from skeletal muscle using PCR
2. Determine the impact of ischemia on endothelial-dependent and endothelial-independent vasodilation
3. Determine the impact of ischemia on adrenergic control of feed artery tone
4. Determine the impact of collateral enlargement on functional vasodilation

Background & Significance

The ability to increase the volume of blood delivered to metabolically active tissue is essential for normal functioning of that tissue. In the case of skeletal muscle- postural maintenance, locomotion, and physical activity all require an increase in skeletal muscle blood flow to deliver the oxygen and nutrients demanded by elevated metabolism. This increase in skeletal muscle blood flow is mediated primarily by vasodilation (increase in blood vessel diameter) of the resistance blood vessels- small arterioles and arteries. For example, small arteries and arterioles will vasodilate to increase blood flow in support of elevated metabolism in exercising skeletal muscle.

Unfortunately, the ability of resistance vessels to vasodilate is impaired following ischemic injury [1][2, 3]. Although impaired vasodilation following ischemic injury is well described in a variety of species [4-6], the molecular mechanism (i.e. the proteins) underlying this dysfunction is unknown. The significance of this knowledge gap is that understanding the molecular mechanisms underlying impaired vasodilation following ischemic injury is necessary to develop specific, efficacious, and personalized therapies designed to maximize the healing response and restore normal function to injured or diseased tissue. Presently, impaired vasodilation in patients with musculo-skeletal injuries or peripheral artery disease are most commonly treated with exercise and/or palliative (treat symptom) pharmacologic agents that are prescribed on a largely trial and error basis [7]. For example, non-steroidal anti-inflammatory drugs are often prescribed to reduce pain in patients repairing from peripheral limb injury, despite the observation that these agents inhibit vasodilation and muscle fiber growth [8].

To address this limitation and improve the specificity and effectiveness of injury rehabilitation and disease treatment, the goal of my research program is to uncover the molecular pathways that cause impaired vasodilation during injury-repair and disease. A primary obstacle facing the development of targeted therapies for improving vasodilation during injury-repair is the fact that after one-hundred and thirty years of investigation [9], the factor(s) required for normal vasodilation during muscle contraction (exercise vasodilation) has yet to be determined [10]- despite the observation that a wide variety of different cellular signaling pathways are capable of inducing vasodilation in skeletal muscle [11, 12]. In other words, it is not known which proteins primarily control exercise vasodilation as well as why those proteins are dysfunctional following ischemic injury. It is important to note that many investigators assume endothelial-derived nitric oxide (NO) is the primary regulator of vasodilation, and that reduced NO bioavailability during situations of vascular injury is the cause for impaired exercise vasodilation and endothelial dysfunction [13]. However, although NO is essential to maintaining baseline arterial diameter, it does not seem to be required for exercise vasodilation [14]. The likely explanation for these divergent results is that the degree of vasodilation in response to a particular stimuli is dependent on the baseline diameter of a vessel [14]. Therefore, if the baseline diameter is lower, due to reduced NO bioavailability, then exercise vasodilation would be less, even if NO did not contribute to that dilation. Therefore, even though reduced bioavailability of NO is a marker for endothelial dysfunction and impaired exercise vasodilation, it is not likely the mechanism underlying the dysfunction. These observations result in a situation in which there is not a single biomarker for normal vasodilation that could be measured and/or targeted therapeutically for improved injury rehabilitation or disease treatment. Therefore, uncovering the molecular pathways responsible for impaired vasodilation following ischemic injury will require a combination of experimental pharmacology and molecular biology.

To assess the moleeular biology of injury repair, we planned to perform RT-PCE on skeletal musele and arteries harvested (and stored at -80°C) from miee that underwent surgieal interruption of the femoral artery by ligation (to tie off with suture) or resection (to remove) to examine the impaet of eollateral enlargemcnt and ischemic, respeetively. Arterial interruption produees an isehemie (insuffieient blood flow) injury that reeapitulates the soft-tissue injury responses in the peripheral limb, ineluding inflammation, myofiber regeneration, angiogenesis (formation of new capillaries), arteriogenesis/eollateral enlargement (formation of new arteries), and likely neural repair [15, 16]. In addition to modeling the soft tissue repair proeesses of the peripheral limbs, arterial interruption also serves as a simplified model of peripheral artery disease [17]. Therefore, the results generated by my researeh program will be relevant for both injury rehabilitation and disease treatment. Furthermore, regenerative medieine strategies, sueh as limb regeneration, depend on vascular growth and remodeling to support the expansion and funetioning of parenehymal eells, sueh as myoblasts and osteoblasts. The results from the proposed work will support this effort by providing a greater understanding of the moleeular determinants of vaseular funetion, which can be targeted with regenerative therapies.

Another advantage of using arterial interruption to model injury-repair and disease is that the repair proeesses (e.g. or arteriogenesis) occur to different extents between the two models [16, 18]. This heterogeneity of responses provides an opportunity to disseet the impaet of different repair events on the expression of transeripts related to vasodilation. For example, comparing the expression of transeripts that eneode for proteins involved in vasodilation signaling pathways between enlarged arteries and isehemic arteries will allow us to disseet the effeet of that repair process on vasodilation.

In the second component of this projeet, we planned to examine the impaet of isehemia and eollateral enlargement on various vasodilation pathways in feed arteries through the pharmacological manipulation of proteins involved in vasodilation. Based on the indication that the seeond part of our projeet will involve manipulating protein funetion, it may seem appropriate to begin our studies at the protein level. However, this is not the direction we have chosen beeause the individual arteries represent a relatively small amount of tissue that would not provide a suffieient amount of protein for performing high-throughput protein analysis, sueh as 2-D gel eleetrophoresis. Additionally, the skeletal muscle samples that eontain isehemie or enlarged arteries would be to contaminated with eontraetile proteins for a successful proteomies study. However, total RNA ean be amplified to inerease the ehanee of deteetion. Therefore, starting with high-throughput RNA analysis is the most appropriate step for the proposed work, while protein analysis will be performed on a narrowed group of eandidate moleeules with approaches more sensitive than 2-D gel eleetrophoresis (that also require less starting sample), such as immunoblotting. Further, we can isolate RNA from previously harvested skeletal muscle and seleetively measure the expression of vaseular speeifie genes using PCR. Additionally, although arerial interruption is an injury/disease model, it is a simplified model using an otherwise healthy animal. Therefore, polygenie aspeets of disease, whieh ean easily confound large-seale expression analysis, will not be faeed in these studies aimed at understanding ischemie injury at a fundamental level and foeusing on the vaseular aspeet of repair.

The results from this projeet are relevant to the ONR/DOD for several reasons. First, a better understanding of ischemic injury will enable more effeetive treatment and rehabilitation of injured soldiers. Specifically, the results from this research will further the effort towards developing targeted therapies to restore normal funetion to injured skeletal musele. Additionally, understanding whieh proteins/transeripts eause impaired vasodilation following ischemic injury

will shed light on the signaling pathways that are required for normal vasodilation. Because improved exercise vasodilation is an important adaptation to exercise training [20], a better understanding of the signaling pathways controlling vasodilation will allow the development of specific biomarkers for fitness. The ability to use gene transcripts to assess training adaptation will improve the efficiency and effectiveness of training programs for soldiers. The other important implication to ONR for this work is limb regeneration. Limb regeneration will require the coordinated growth of vascular, stromal, parenchymal, and neural cells. By advancing the understanding of vascular growth, remodeling, and function, the proposed work will support the development of strategies to promote complex tissue regeneration by beginning to define therapeutic targets for normal vascular growth and function.

Results- Molecular Biology

The goal of the molecular biology portion of the project was to develop protocols for measuring gene expression from mouse skeletal muscle. The first major step towards achieving this goal was the purchase of equipment necessary to isolate RNA, reverse transcribe RNA into cDNA, and amplify cDNA using PCR. The receipt of this award allowed us to leverage matching funds from the department to the purchase the molecular biology laboratory equipment shown Figure 1. The equipment shown in Figure 1, along with the reagents purchase through this

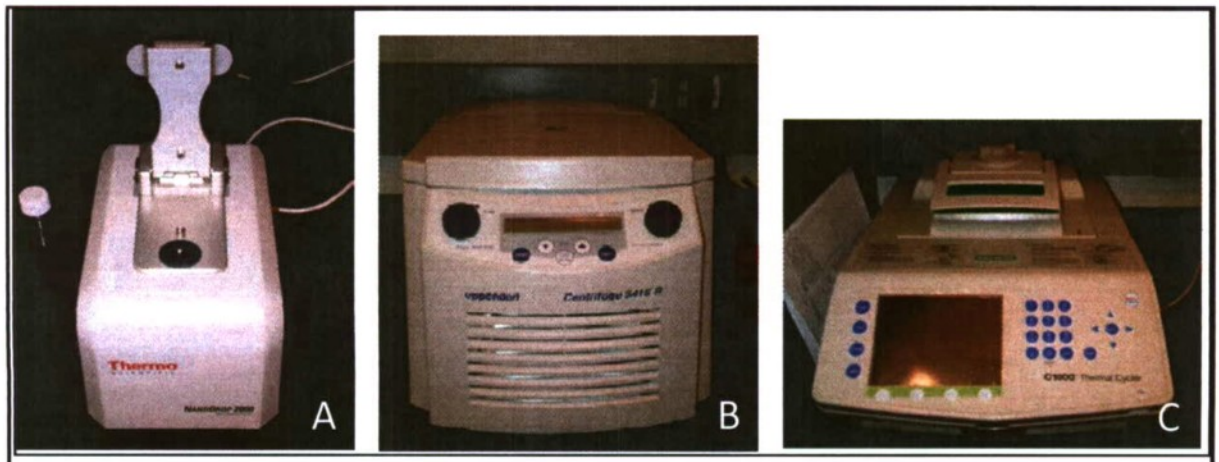
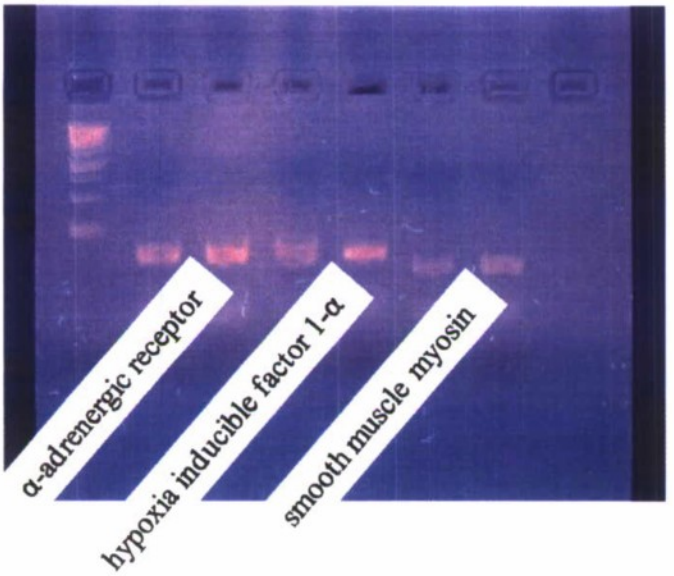


Figure 1. Molecular biology laboratory equipment. A, Refrigerated microcentrifuge used in nucleic acid isolation. B, UV spectrophotometer used to determine nucleic acid concentration and purity. C, Thermalcycler used to reverse transcribe RNA into cDNA and to amplify cDNA with PCR.

award, allowed us to being developing protocols for measuring skeletal muscle gene expression.

We have successfully implemented a protocol for isolating RNA from mouse skeletal muscle and have made significant progress in our protocol for isolating RNA from individual feed arteries. We have successfully designed PCR primers to amplify over a dozen genes-of-interest from genomic DNA and have recently achieved success amplifying genes-of-interest from gracilis muscle cDNA using PCR (Figure 2). Our present objective is to perform quantitative PCR to determine if these genes-of-interest are differentially expressed in gracilis muscles containing healthy, ischemic, or enlarged arteries.

Figure 2. Agarose gel electrophoresis of PCR-amplified gracilis muscle cDNA for genes-of-interest



Results- Experimental Physiology

The first goal of the experimental physiology portion of this award was to determine the impact of ischemia on endothelial-dependent vasodilation. Previous data demonstrated that high-intensity, but not low intensity, muscle contraction would elicit functional vasodilation in the muscular branch feed artery. One of the interpretations from this data is that endothelial-dependent vasodilation is impaired, but endothelial-independent vasodilation (i.e. smooth-muscle dependent) was preserved in the ischemic environment. To test this hypothesis we applied increasing doses of acetylcholine (endothelial-dependent vasodilator) to the muscular branch artery, following by a single large dose of sodium nitroprusside (endothelial-independent vasodilator). The ischemic artery had a larger resting diameter than the non-ischemic artery, Figure 3, but was able to dilate to the same diameter as the non-ischemic artery.

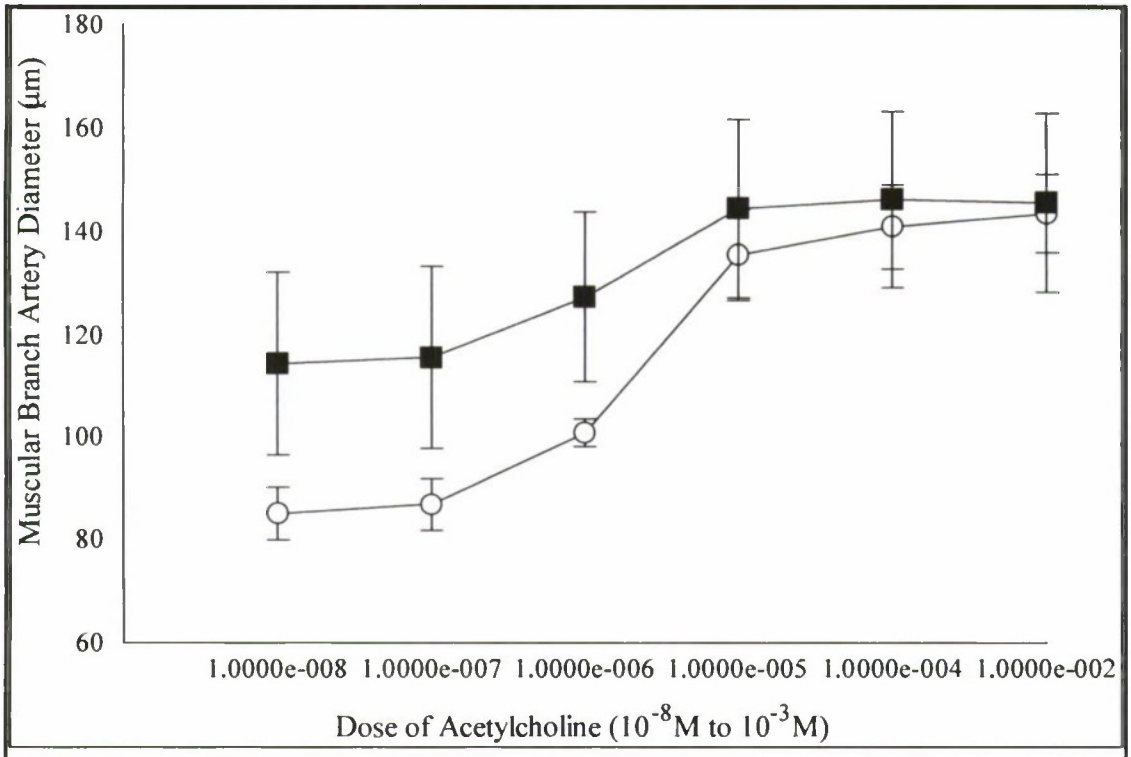


Figure 3. Endothelial-Dependent Vasodilation to acctylcholine in ischemic (black squares) and non-ischemic (white circles) muscular branch arteries.

The addition of sodium nitroprusside at the end of the acetylcholine dose response curve resulted in similar diameters between the ischemic and non-ischemic muscular branch arteries (data not shown), suggesting that endothelial dependent vasodilation was maintained following ischemia. These data suggest that the absence of functional vasodilation in response to low-intensity muscle contraction is likely due to the presence of hypoxia in the ischemic gracilis muscle. Baseline hypoxia may be larger than the signal for metabolic vasodilation produced by the weakly contracting muscle. The fact that endothelial-dependent vasodilation is maintained in this context is an unexpected, but interesting finding that merits future investigation to determine how an ischemic artery can maintain vasomotor function.

The next objective of the experimental physiology portion of the project was to determine the impact of ischemia on the adrenergic sensitivity of the muscular branch artery. Previous data collected in our laboratory demonstrated that vasoconstriction following functional vasodilation was abnormal, and others have observed reduced vasodilation in response to α-adrenergic antagonists (i.e. sympathetic nervous system blockers) in the collateral circulation. Therefore, we hypothesized that the ischemic muscular branch artery would exhibit reduced sensitivity to α-adrenergic signaling. To test this hypothesis we applied increasing doses phentolamine, a broad-spectrum α-adrenergic antagonist, to the muscular branch artery. As with the previous study, the ischemic artery had a larger resting diameter than the non-ischemic artery (Figure 4). However, the ischemic artery did exhibit a statistically significant increase in diameter, while the non-ischemic artery did vasodilate (Figure 5).

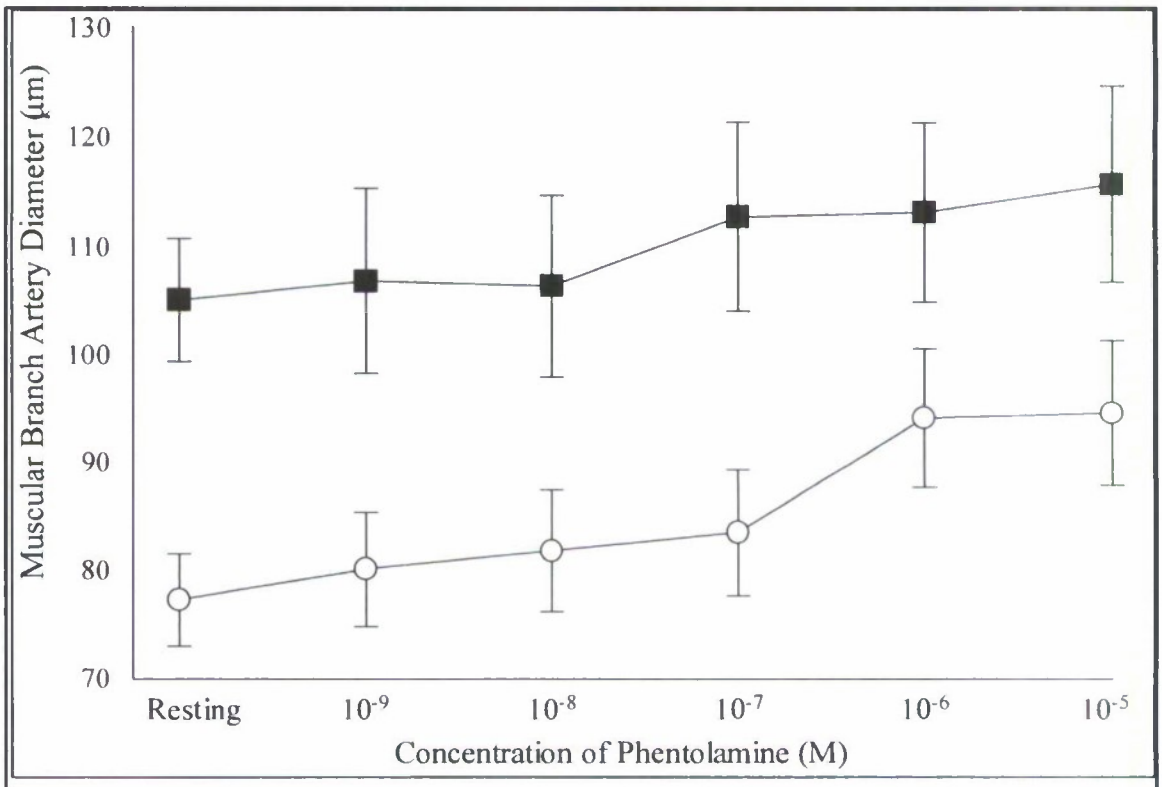


Figure 4. Vasodilation in response to α-adrenergic blockade by phentolamine in the ischemic (black boxes) and non-ischemic (white circles) muscular branch artery.

These results demonstrate that the ischemic arteries are less sensitive to α-adrenergic induced vasoconstriction, suggesting a down-regulation of α-adrenergic receptors or an up-regulation of β-adrenergic receptors on the ischemic arteries. These preliminary results will be followed up by experiments to examine both the gene expression and pharmacology of α- and β-adrenergic receptors in the ischemic muscular branch artery.

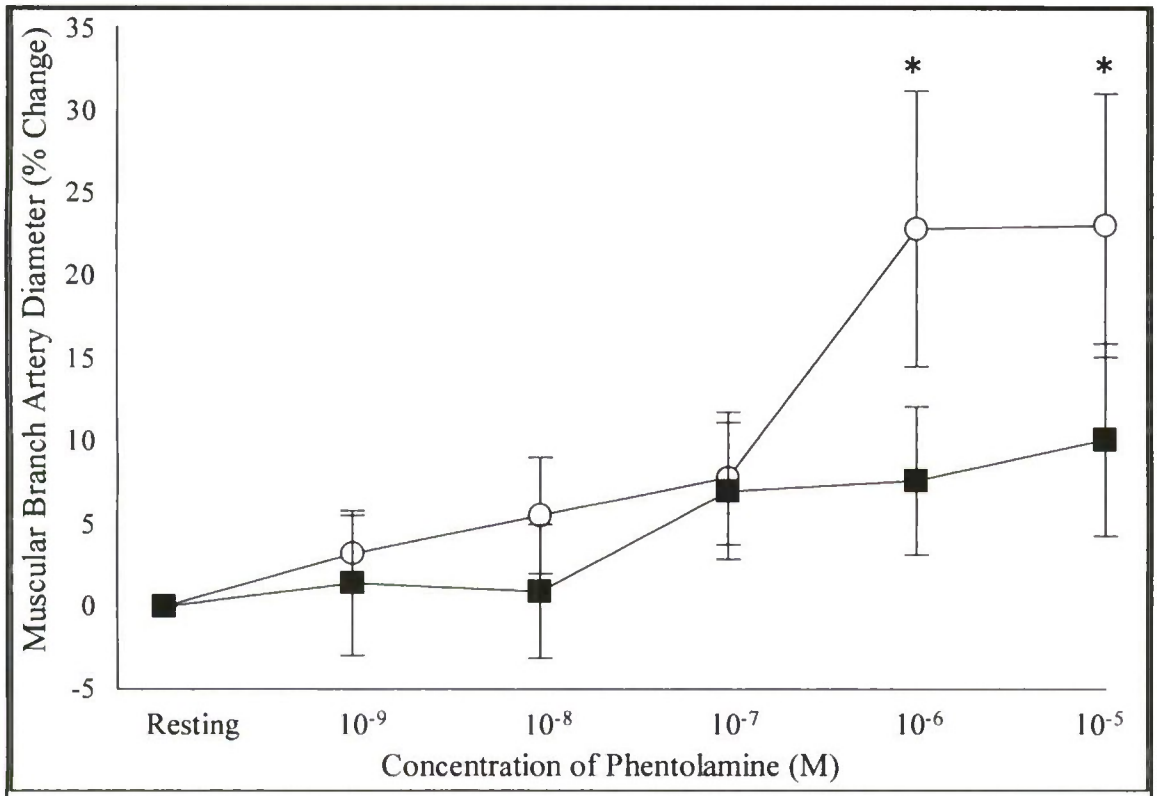


Figure 5. Vasodilation in response to α-adrenergic blockade by phentolamine in the ischemic (black boxes) and non-ischemic (white circles) muscular branch artery.

The final objective of the experimental physiology portion of this award was to further characterize the impact of collateral enlargement on functional vasodilation. It is expected that the transition of smooth muscle cells from a contractile to a synthetic phenotype during collateral enlargement would impair vascular reactivity. However, this hypothesis has not been directly tested. Previously, we demonstrated that functional vasodilation in enlarged collaterals was normal at day-14 following femoral artery ligation. Therefore, we hypothesized that collateral enlargement must cause a temporary impairment to functional vasodilation. To test this hypothesis, we examined functional vasodilation in the muscular branch artery at day-7 following femoral artery ligation. As expected, functional vasodilation was normal at day-14 following femoral artery ligation and severely impaired at day-7 following femoral artery ligation, Figure 6. Although these results support the hypothesis that the emergence of the synthetic smooth muscle phenotype during collateral enlargement impairs functional vasodilation, further experiments are needed to confirm. In future studies, we plan to examine endothelial-dependent and endothelial-independent vasodilation, and assess smooth muscle phenotype with immunofluorescence.

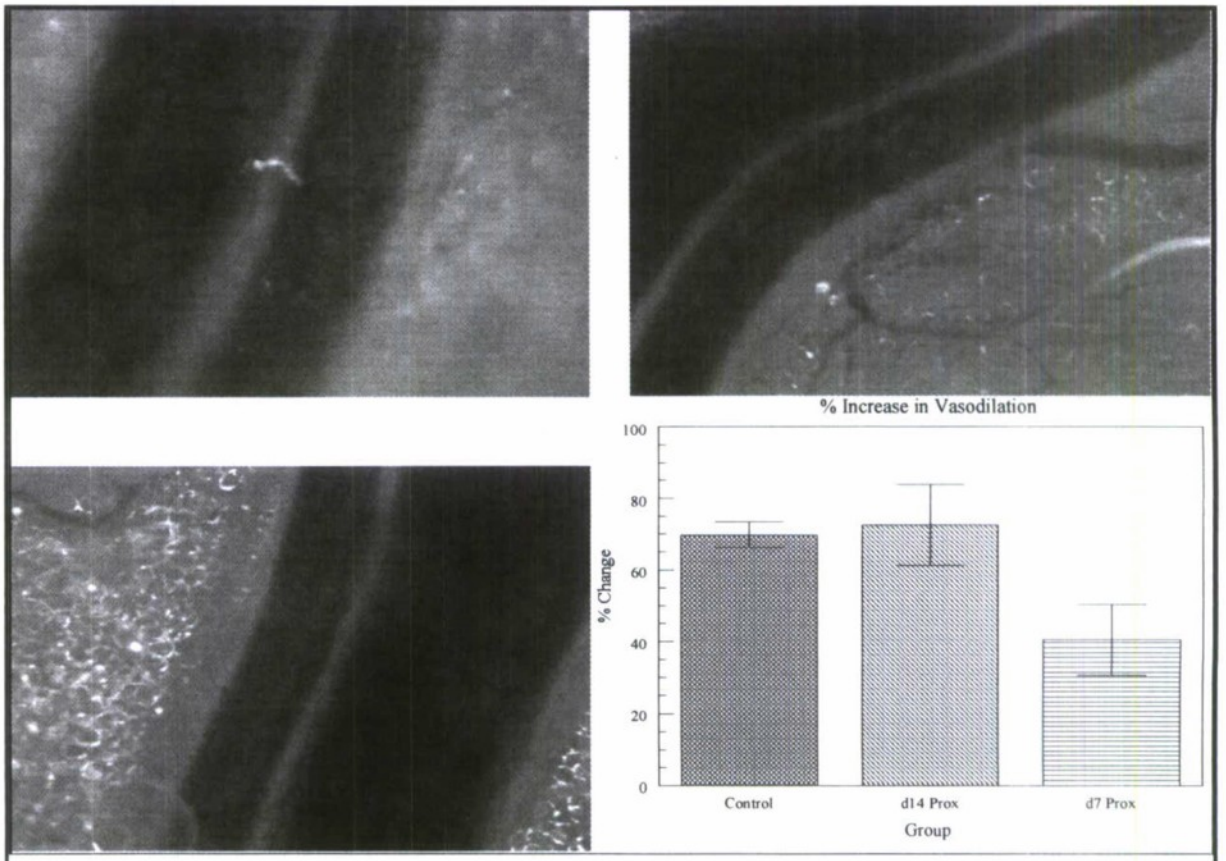
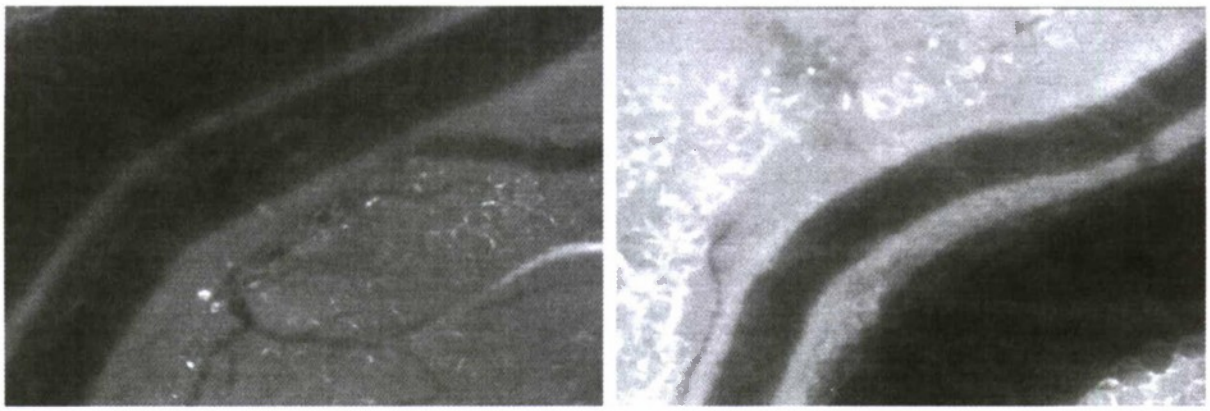


Figure 6: A comparison of maximal vasodilation between different time points following femoral ligation at proximal site. The magnitude of vasodilation is significantly less seven days following ligation surgery compared to the control while vasodilation fourteen days after surgery was not significantly different from the control.

The final experiment in the last objective was to determine if collateral enlargement altered the sensitivity of the museular braneh artery to funetional vasodilation, as we had observed with isehemia. Based on our previous results examining the impaet of isehemia on the museular braneh, we hypothesized that eollateral enlargement would reduee the sensitivity of the museular braneh artery to funetional vasodilation stimuli. To test this hypothesis, we examined museular braneh functional vasodilation in response to low-intensity (200μs pulse durations) and high-intensity (500μs pulse durations) musele eontraetions. To eonelusions are apparent from the results presented in Figure 7. First, the high-intensity muscle contractions (500μs) failed to produee a significantly greater dilation in the enlarged collateral, as they did in the non-ischemic artery. Seeond, the low-intensity musele contraetions (200μs) produeed a similar dilation in the enlarged collateral as the high-intensity muscle contraction, Figure 7. These results suggest that if eollateral enlargement alters the sensitivity to funetional vasodilation stimuli, then that alteration oeurs only at higher levels of stimulus, not lower, as was observed in ischemie arteries.



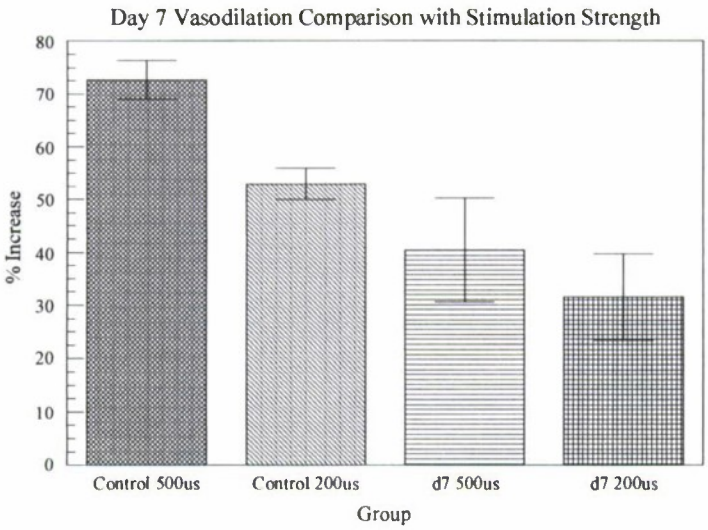


Figure 7: A comparison of how magnitude of vasodilation is affected by the applied pulse duration. Under normal conditions, the 500 μs pulse causes significantly more dilation than the 200 μs pulse ($p=.036$). On legs which had undergone collateral enlargement following a proximal femoral ligation 7 days earlier, there was no significant difference between pulse duration.

Summary

Through this award, we were able to develop protocols for measuring mRNA from skeletal muscle using PCR, determine the impact of ischemia on endothelial-dependent and endothelial-independent vasodilation, determine the impact of ischemia on adrenergic control of feed artery tone, and further characterize the impact of collateral enlargement on functional vasodilation. The results from these studies will be presented at 2011 Experimental Biology Meeting and will be incorporated into 1-2 manuscripts to be prepared over the next year.

Reference List

1. Kelsall, C.J., M.D. Brown, and O. Hudlicka, *Alterations in reactivity of small arterioles in rat skeletal muscle as a result of chronic ischaemia.* J.Vasc.Res., 2001. **38**(3): p. 212-218.
2. Frisbee, J.C. and D.W. Stepp, *Impaired NO-dependent dilation of skeletal muscle arterioles in hypertensive diabetic obese Zucker rats.* Am.J.Physiol Heart Circ.Physiol, 2001. **281**(3): p. H1304-H1311.
3. Kim, M.H., P.R. Carter, and N.R. Harris, *P-selectin-mediated adhesion impairs endothelium-dependent arteriolar dilation in hypercholesterolemic mice.* Am.J.Physiol Heart Circ.Physiol, 2007. **292**(1): p. H632-H638.

4. Brendle, D.C., et al., *Effects of exercise rehabilitation on endothelial reactivity in older patients with peripheral arterial disease.* Am.J.Cardiol., 2001. **87**(3): p. 324-329.
5. Takeshita, S., et al., *Endothelium-dependent relaxation of collateral microvessels after intramuscular gene transfer of vascular endothelial growth factor in a rat model of hindlimb ischemia.* Circulation, 1998. **98**(13): p. 1261-1263.
6. Bauters, C., et al., *Recovery of disturbed endothelium-dependent flow in the collateral-perfused rabbit ischemic hindlimb after administration of vascular endothelial growth factor.* Circulation, 1995. **91**(11): p. 2802-2809.
7. Aronow, W.S., *Management of peripheral arterial disease.* Cardiol.Rev., 2005. **13**(2): p. 61-68.
8. Bondesen, B.A., et al., *The COX-2 pathway is essential during early stages of skeletal muscle regeneration.* Am J Physiol Cell Physiol 2004. **287**(2): p. C475-83.
9. Gaskell, W.H., *The Changes of the Blood-stream in Muscles through Stimulation of their Nerves.* J.Anat.Physiol, 1877. **11**(Pt 3): p. 360-402.
10. Joyner, M.J. and B.W. Wilkins, *Exercise hyperaemia: is anything obligatory but the hyperaemia?* J Physiol 2007. **583**(Pt 3): p. 855-60.
11. Clifford, P.S. and Y. Hellsten, *Vasodilatory mechanisms in contracting skeletal muscle.* J.Appl.Physiol, 2004. **97**(1): p. 393-403.
12. Gorman, M.W. and H.V. Sparks, *The Unanswered Question.* News in Physiological Sciences 1991. **6**(4): p. 191-193.
13. Barac, A., U. Campia, and J.A. Panza, *Methods for evaluating endothelial function in humans.* Hypertension 2007. **49**(4): p. 748-60.
14. Saito, Y., A. Eraslan, and R.L. Hester, *Role of endothelium-derived relaxing factors in arteriolar dilation during muscle contraction elicited by electrical field stimulation.* Microcirculation., 1994. **1**(3): p. 195-201.
15. Couffinhal, T., et al., *Mouse model of angiogenesis.* Am.J.Pathol., 1998. **152**(6): p. 1667-1679.
16. Scholz, D., et al., *Contribution of arteriogenesis and angiogenesis to postocclusive hindlimb perfusion in mice.* J.Mol.Cell Cardiol., 2002. **34**(7): p. 775-787.
17. Waters, R.E., et al., *Preclinical models of human peripheral arterial occlusive disease: implications for investigation of therapeutic agents.* J.Appl.Physiol, 2004. **97**(2): p. 773-780.
18. Sullivan, C.J., T. Doetschman, and J.B. Hoying, *Targeted disruption of the Fgf2 gene does not affect vascular growth in the mouse ischemic hindlimb.* J.Appl.Physiol, 2002. **93**(6): p. 2009-2017.
19. Read, S.J., et al., *Stroke genomics: approaches to identify, validate, and understand ischemic stroke gene expression.* J Cereb Blood Flow Metab 2001. **21**(7): p. 755-78.
20. Jasperse, J.L. and M.H. Laughlin, *Endothelial function and exercise training: evidence from studies using animal models.* Med.Sci.Sports Exerc., 2006. **38**(3): p. 445-454.
21. Kelsall, C.J., et al., *Arteriolar endothelial dysfunction is restored in ischaemic muscles by chronic electrical stimulation.* J.Vasc.Res., 2004. **41**(3): p. 241-251.
22. Welsh, D.G. and S.S. Segal, *Coactivation of resistance vessels and muscle fibers with acetylcholine release from motor nerves.* Am.J.Physiol, 1997. **273**(1 Pt 2): p. H156-H163.
23. Dietrich, H.H., et al., *Red blood cell regulation of microvascular tone through adenosine triphosphate.* Am.J.Physiol Heart Circ.Physiol, 2000. **278**(4): p. H1294-H1298.

24. Boarder, M.R. and S.M. Hourani, *The regulation of vascular function by P2 receptors: multiple sites and multiple receptors.* Trends Pharmacol.Sci., 1998. **19**(3): p. 99-107.
25. Maxwell, A.J., et al., *Limb blood flow during exercise is dependent on nitric oxide.* Circulation, 1998. **98**(4): p. 369-374.
26. Murrant, C.L. and I.H. Sarelius, *Multiple dilator pathways in skeletal muscle contraction-induced arteriolar dilations.* Am.J.Physiol Regul.Integr.Comp Physiol, 2002. **282**(4): p. R969-R978.
27. Feletou, M. and P.M. Vanhoutte, *Endothelium-derived hyperpolarizing factor: where are we now?* Arterioscler.Thromb.Vasc.Biol., 2006. **26**(6): p. 1215-1225.

Bone Density Measurement via Radiographic Calibration

Principal Investigator:

Matthew Burd, Ph. D.
Animal Sciences
California Polytechnic State University
San Luis Obispo, CA

1. Introduction

Musculoskeletal injuries occur in 50% of soldiers in Iraq and Afghanistan and the military spends over 1.5 billion dollars annually treating and managing these injuries. Off the battle field, lower extremity stress fractures are common injuries to soldiers in basic training. Microfractures are known to weaken the bone, eventually leading to catastrophic failure.[1] Indeed, march fractures were reported as early as WWII, yet these injuries continue to affect the performance of the modern soldier.[2,3] Among the identified risk factors of stress fracture development, decreased bone mineral density has been correlated with stress fractures in military and civilian studies, especially in women.[4-7] However, if damage can be diagnosed and allowed to repair, catastrophic failure may be delayed or prevented.[8]

The Academy of Orthopedic Surgeons has provided guidelines for future research objectives designed to improve patient care. Fundamental aspects of these recommendations include the development of a large animal model of bone injury and improvement of electronic medical records.[9] Despite the many methods used today to evaluate bone mineral content, including dual-energy X-ray absorptiometry (DEXA), quantitative computed tomography, and qualitative ultrasound, none are used at near the frequency as a standard radiograph in qualifying fractures. Digital radiography has the advantage for the clinician of being extremely portable, relatively inexpensive, and practical in a field setting. Digital radiography provides the clinician with portable rapid image acquisition, diagnosis, and images that can easily be stored or transferred for off-site review. Unfortunately, radiographic images taken with different machines at different locations cannot be directly compared due to inherent variation in X-ray machines and radiographic technique (radiation intensity and exposure time). This is of considerable significance as soldiers are cared for by over 140,000 care providers at 477 military clinics and

hospitals. Calibration of radiographic density would allow quantification of optical bone density, a predictor of bone density assuming that the area measured has a consistent thickness and optical attenuation properties.

Using a large animal model, we will develop a method to calibrate radiographic images of bones such that bone density determined from radiographs taken under different circumstances can be directly compared among medical military institutions. By calibrating the grayscale of digital radiographs using a brightness/darkness index (BDI), we will form a standardized relationship allowing comparison of BDI values to the density of bone regardless of radiographic settings of exposure intensity (kV) and time (mAs). Inclusion of an aluminum (Al) wedge of known density and thickness, or optical density (g/cm^2), in the radiograph will allow Al optical density to be determined from BDI, however accurate measurement of bone optical density at varying exposures from the Al wedge standard curve is a more complex issue. First, the relationship between BDI and radiographic exposure will be determined for both bone and Al to be smooth and uniform, and therefore easily standardized. Secondly, BDI of Al will be correlated with the BDI of bone at varying exposure, allowing determination of the BDI of bone as a continuous function of BDI of Al, independent from intensity and exposure time. This will allow us to account for changes in the brightness of radiographs due to changes in exposure, affording brightness standardization of any radiograph. Thirdly, bone optical density will be correlated with BDI of bone by measuring BDI from a radiograph of cortical bone of known density and varying thickness. This will allow us to determine bone instead of Al optical density from BDI measurements. Finally, bone optical density can be related to Al optical density using the Al wedge standard curve, accurate over various exposure settings. Further, optical density can predict density for samples of similar thickness and material composition. Therefore, we

hypothesize that calibration of digital radiographs to determine optical bone density from an Al wedge standard will provide the clinician a tool for quantification of radiographic bone density from common digital radiographs, regardless of radiographic technique.

2. Materials and Methods

2.1. Radiographic Method

In each study, radiographs contained an Auto-Scaler (EponaTech) and Al wedge. The Al wedge was machined from type 6061 Al, and was 20.36 cm in length and 0.2 to 31.8 mm in thickness, with a constant density of 2.70 g/cm^3. The Auto-Scaler contained metal markers that standardized the dimensions of the radiographic image when imported into the software program Metron-DVM (EponaTech).

To determine BDI, a unit-less value was assigned to each 16 bit pixel on a greyscale from 0 to 65,500, with zero being completely black and 65,500 being completely white. A predetermined region of interest on either bone or Al was generated by forming a rectangle using Metron. Metron exports the pixel values into Excel (Microsoft). The BDI for the area is determined by averaging the grayscale value of each pixel in the area. In determining BDI, background pixels were not included.

All radiographs were taken with a portable X-ray system (Min X-ray HF80) and X-ray sensor (Thales FS23, Vetel Diagnostics). All radiographs were taken at 15 mA, the only setting available for this system.

2.2. BDI at Varying Radiographic Exposure

The relationship between BDI and radiographic exposure was determined for both bone and Al. A single third metacarpal equine cadaver bone was used in every radiograph, placed on a wooden block with the Al wedge and Auto-Scaler. The cadaver bone was positioned perpendicular to the ground in the center of the line of exposure, with the Al wedge and Auto-Scaler positioned on either side. The bone was placed to produce a dorsal palmar radiograph (Figure 1). The focal distance was 21 cm, with the plane of interest containing the cadaver bone, Al wedge, and Auto-Scaler positioned against the face of the sensor. Radiographs were taken at all available combinations of pre-set exposure intensities ranging from 55 to 80 kV in 5 kV intervals, and exposure times from 0.02 to 0.14 seconds every 0.02 seconds.

The BDI of the Al wedge and of the cadaver bone was measured on each radiographic image. The BDI for Al was measured by creating a rectangle encompassing the entire wedge. The BDI for bone was characterized by an area with the width larger than the cross-sectional width of the bone perpendicular to the sagittal plane, and the height equal to the diameter of the nutrient foramen (Figure 2). The nutrient foramen was used as a marker, so that the measurement of bone BDI was always for the same piece of bone. To ensure change in BDI was solely due to effects of radiographic exposure, the same area was used in each radiographic image. The relationship was characterized for BDI of the Al wedge and the BDI of bone dependent on intensity and exposure time. The BDI of Al was compared to the BDI of bone in the same radiographic image at varying X-ray intensities and exposure times to ensure the BDI of Al correlates with that of bone.

2.3. BDI as a Predictor for Optical Density

Bone optical density was correlated with BDI of bone by measuring BDI from a radiograph of cortical bone of known density and varying thickness. A bone wedge was made from cortical bone obtained from a third metacarpal equine cadaver bone in the mid-diaphysis region. The cortical bone was cut into slices using a diamond blade tile saw. The average dimension of each slice was 15.28 ± 3.3 mm square and 4.23 ± 1.0 mm thick. A bone wedge was created by stacking each bone slice face-to-face in increasing thickness from one slice to 6 slices thick, consecutively from Step 1 to 10. Scotch tape was used on the sides of the bone slices to hold them in place, forming a total of ten steps of increasing thickness.

Standard digital radiographs were taken of the bone wedge and Al wedge. The bone wedge was positioned to intersect the center of the line of exposure, with the Al wedge on one side. The focal distance was 21 cm, with the plane of interest containing the bone wedge and Al wedge positioned against the face of the sensor (Figure 3). Radiographs were taken at 60 kV and 0.08 seconds.

BDI was determined for each step of bone and for the Al wedge (Figure 4). The thickness and density corresponding with each bone slice was determined using 6 inch digital calipers (Fisher Scientific) with 0.01 mm accuracy and a digital balance (Accu-224, Fisher Scientific) with 0.1 mg accuracy. Density was determined from weight (g) divided by volume (cm^3) determined using digital calipers. Optical density for bone and Al was determined as thickness multiplied by density (g/cm^2). Standardized curves were created for BDI of the bone versus bone optical density, and the BDI of Al versus Al optical density. Where the BDI of Al equaled that of bone, the correlation between bone and Al optical density could be determined.

2.4. Statistical Analysis

Regression analysis was performed using the least squares method. Slope of the regression lines were shown to be different from zero determined by a Student's *t*-test. Multi-variable analysis was done for BDI versus exposure intensity and time using an ANOVA and associated *F*-test. Statistical significance was set at $p < 0.05$.

3. Results

BDI was standardized for bone and Al over varying exposure intensity and time. BDI of bone and Al decreased with increasing exposure intensity and exposure time. From 0.02 to 0.14 seconds and 55 to 80 kV, bone BDI ranges from 13,747 to 49,851 (Table 1A). Similarly, Al BDI decreased with increasing exposure intensity and exposure time, ranging from 21,388 to 54,328 (Table 1B). BDI correlated linearly with exposure intensity and exposure time for Al (r^2 = 0.8594, $F(2,39) = 119.2$, $p < 0.001$) and bone ($r^2 = 0.9126$, $F(2,39) = 203.7$, $p < 0.001$) (Figure 5). For each combination of intensity and exposure time, Al BDI showed a linear correlation with bone BDI ($r^2 = 0.9599$, $F(1,40) = 957.3$, $p < 0.001$). As Al BDI increased per unit, there was an increase in bone BDI of 1.12 ± 0.04 (Figure 6).

Next, the relationship between optical density and BDI for cortical bone was quantified. Bone BDI was shown to increase linearly with increasing optical bone density ($r^2 = 0.9716$, $F(1,1193) = 40,850$, $p < 0.001$, Figure 7). From Row 1 to Row 10 of the bone wedge, the BDI ranged from $3,648 \pm 723$ to $54,151 \pm 691$ (Table 2). Similarly, Al BDI increased with increasing

Al optical density, ranging from a BDI of 0 to 53,810 with a constant density of 2.70 g/cm^3. Al thickness increased linearly along the length of the Al wedge from 0 to 4.26 cm. For thicknesses under 4.26 cm Al a linear approximation could be made, allowing determination of the mass attenuation coefficient ratio from a single variant (r^2 = 0.9973, $F(1,631)$ = 232,800, $p < 0.001$, Figure 8).

4. Discussion

Bone density measurement to assess bone strength is a commonly accepted procedure in the medical field.[10] Clinicians routinely utilize radiographs to assess bone health and diagnose abnormalities such as fractures and osteoarthritis, subjectively and qualitatively assessing the opacity or optical density of areas of interest. A tool to quantitatively evaluate bone strength or radiographic density in regions of interest from digital radiographs would be invaluable. Therefore, we developed a method based on standardized digital radiographs that affords the professional accurate and repeatable measurement of optical bone density.

As the third metacarpal in the mid-diaphysis region is composed of uniformly dense cortical bone and very little trabecular bone and soft tissue, it is an ideal choice to investigate the parameters associated with radiographic determination of bone density. In this study, we showed that the attenuation of X-rays through bone and Al occurs in a smooth, linear fashion in response to exposure settings and properties of the material itself (Figure 5, Table 1). Over the range of clinically relevant exposures, the BDI of an object of known density and thickness such as an Al wedge is a significant predictor of bone BDI of unknown density and thickness (r^2 = 0.9599, $F(1,40)$ = 957.3, $p < 0.001$, Figure 6). BDI can then be determined using a standard curve from

an Al wedge, where BDI is a function of Al optical density. These results indicate that calibration of BDI with an Al wedge standard curve allows for accurate assessment of optical density in units of g/cm^2 of Al for medically relevant ranges of exposure.

In determining optical bone density, a bone wedge of known density and thickness was created (Table 2). BDI was shown to be a significant predictor of optical density in both bone (r^2 = 0.9716, $F(1,1193) = 40{,}850$, $p < 0.001$, Figure 7) and Al ($r^2 = 0.9973$, $F(1,631) = 232{,}800$, $p < 0.001$, Figure 8). This method permits accurate evaluation of BDI and, therefore, optical bone density independent of the exposure setting.

This method successfully quantified optical bone density via measurement of BDI from standardized digital radiographs, allowing for the opacity of radiographs to be truly comparable when taken under different circumstances. This may be useful in evaluation for diseases that weaken bone in certain areas more than others, such as stress fracture risk. This may have practical applications in clinical circumstances as optical density is dependent only on thickness and density of the material; therefore, a given area of bone with lower optical density than another area of bone of similar thickness can be concluded to have a lower density. Further, stress fractures induced by exercise are correlated with decreased BMD, and optical density may be an accurate predictor of BMD.[11]

With this method, optical density can be accurately determined over a range of exposures, allowing the military physician to directly compare bone images from different soldiers and at different times, regardless of radiographic technique. We believe the performance of the soldier can be enhanced and injuries that prevent a soldier from training can be minimized. Additionally, research in severe extremity injuries have been identified as a top

priority and the information gained with this investigation may allow more accurate radiographic assessment of these devastating injuries including the ability to accurately interpret radiographs by a specialist thousands of miles away. We believe this will allow the physician to more accurately diagnose, treat, and possibly prevent musculoskeletal injury.

5. References

1. Schaffler MB, Radin EL, Burr DB. Mechanical and morphological effects of strain rate on fatigue of compact bone. Bone 1989; 10(3):207-14.

2. Kummer M, Lischer C, Ohlerth S, et al. Evaluation of a standardised radiographic technique of the equine hoof. *Schweiz Arch Tierheilkd* 2004;146:507-14.

3. Rocha J, Lischer C, Kummer M, Hässig M, et al. Evaluating the measuring software package Metron-PX for morphometric description of equine hoof radiographs. *J Equine Vet Sci* 2004;24:347-354.

4. Lauder TD, Dixit S, Pezzin LE, Williams MV. The relation between stress fractures and bone mineral density: evidence from active-duty Army women. Arch Phys Med Rehabil 2000; 81(1):73-9.

5. Giladi M, Ahronson Z, Stein M, et al. Unusual distribution and onset of stress fractures in soldiers. Clin Orthop 1985; 192:142-6.

6. Cranney A, Jamal SA, Tsang JF, et al. Low Bone Mineral Density and Fracture Burden in Post Menopausal Women. CMAJ 2007; 177(6):575-80.

7. Beck TJ, Ruff CB, Shaffer RA, et al. Stress fracture in military recruits: gender differences in muscle and bone susceptibility factors. Bone 2001; 27(3):437-44.

8. Martin RB, Burr DB. A hypothetical mechanism for the stimulation of osteonal remodelling by fatigue damage. J Biomech 1982; 15(3):137-9.

9. USDA. *"National Economic Cost."* Centers of Epidemiology and Animal Health; 2001.

10. Tabensky AD, Williams J, DeLuca V, et al. Bone mass, areal, and volumetric bone density are equally accurate, sensitive, and specific surrogates of the breaking strength of the vertebral body: an in vitro study. J Bone Miner Res. 1996;11:1981-8.

11. Vinther A, Kanstrup IL, Christiansen E, et al. Exercise-induced rib stress fractures: influence of reduced bone mineral density. Scand J Med Sci Sports 2005; 15(2):95-9.

Appendix

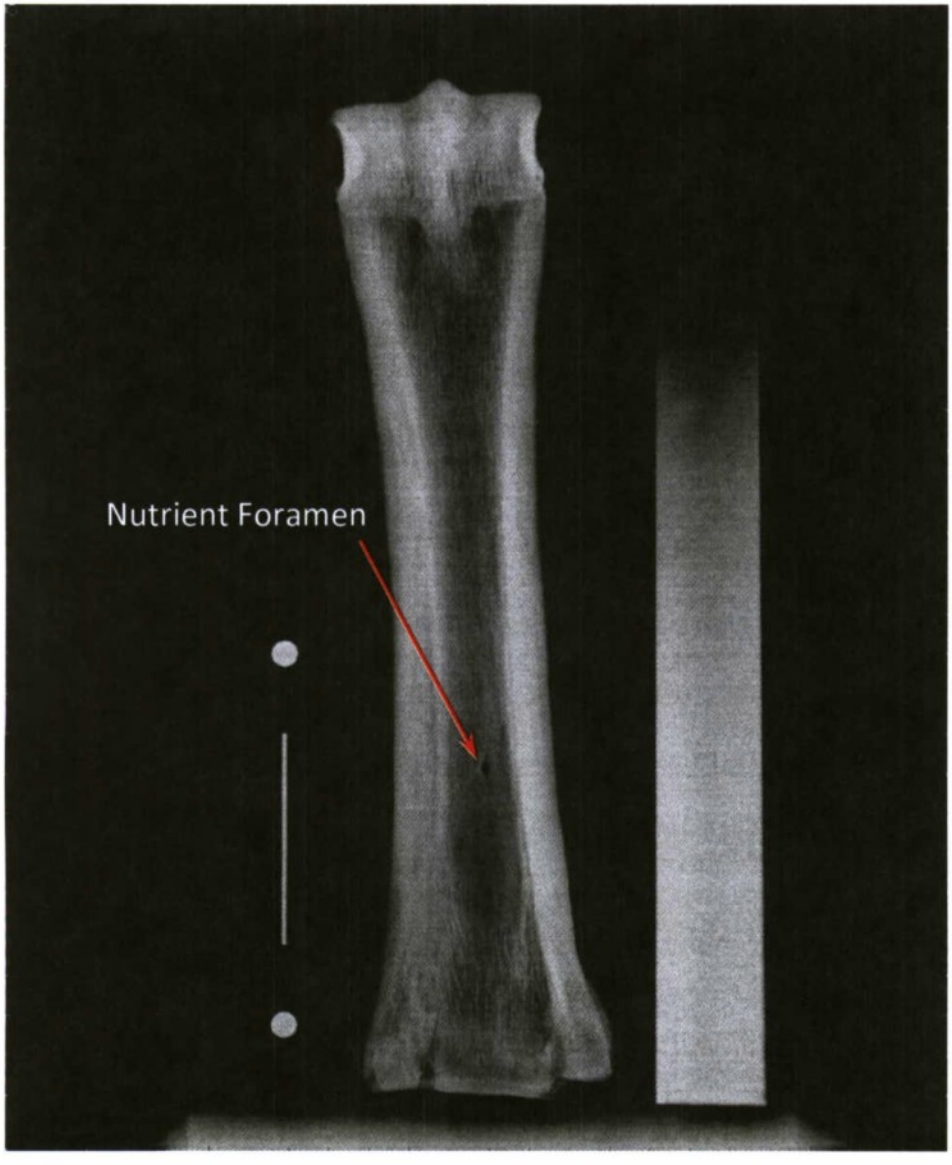


Figure 1. Dorsal palmar positioning of a third metacarpal equine cadaver bone on a wooden block with Auto-Scaler (left) of known dimensions and Al wedge (right) of known thickness and density. The nutrient foramen (red arrow) can be seen as a dark spot on the diaphysis.

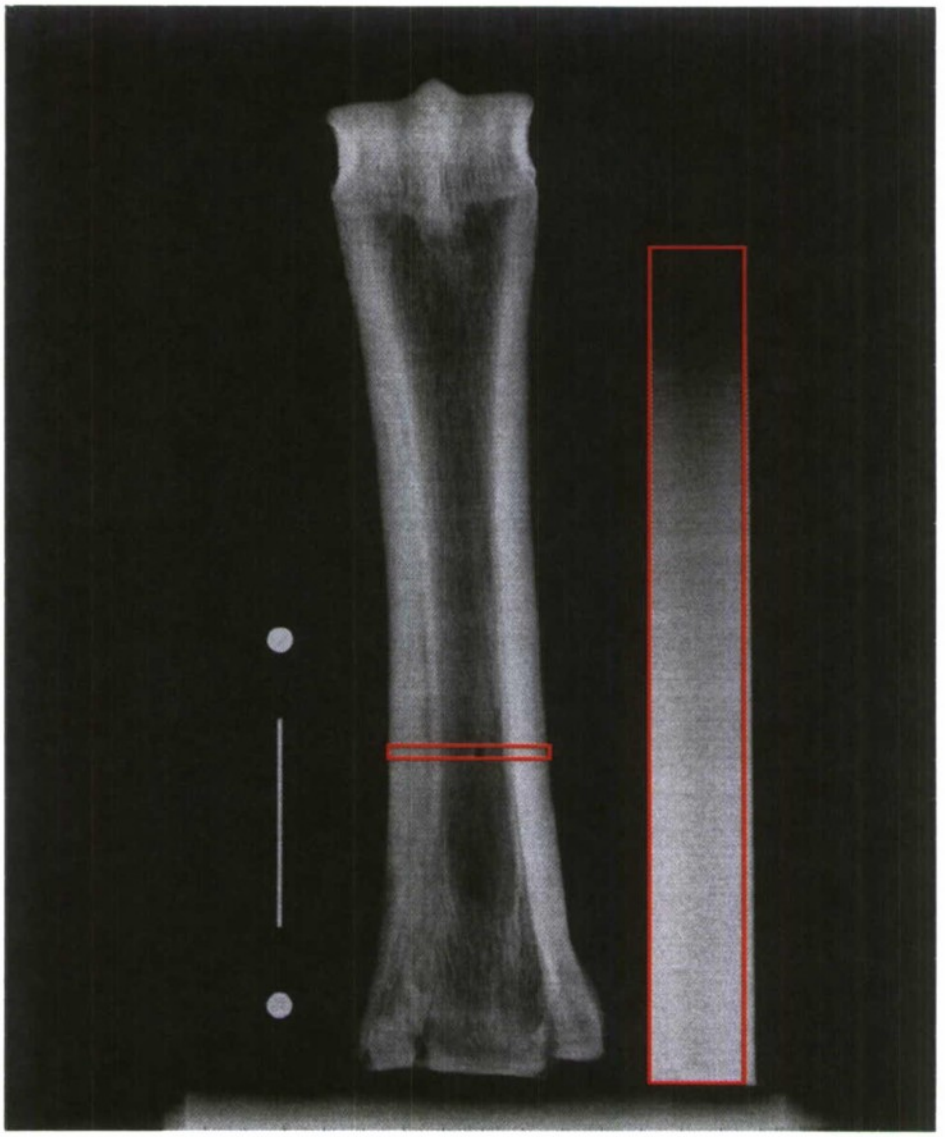


Figure 2. Selection of regions of interest (red) for measuring bone BDI at the nutrient foramen and Al BDI from an Al wedge, determined from a dorsal palmar radiograph of a third metacarpal cadaver bone.

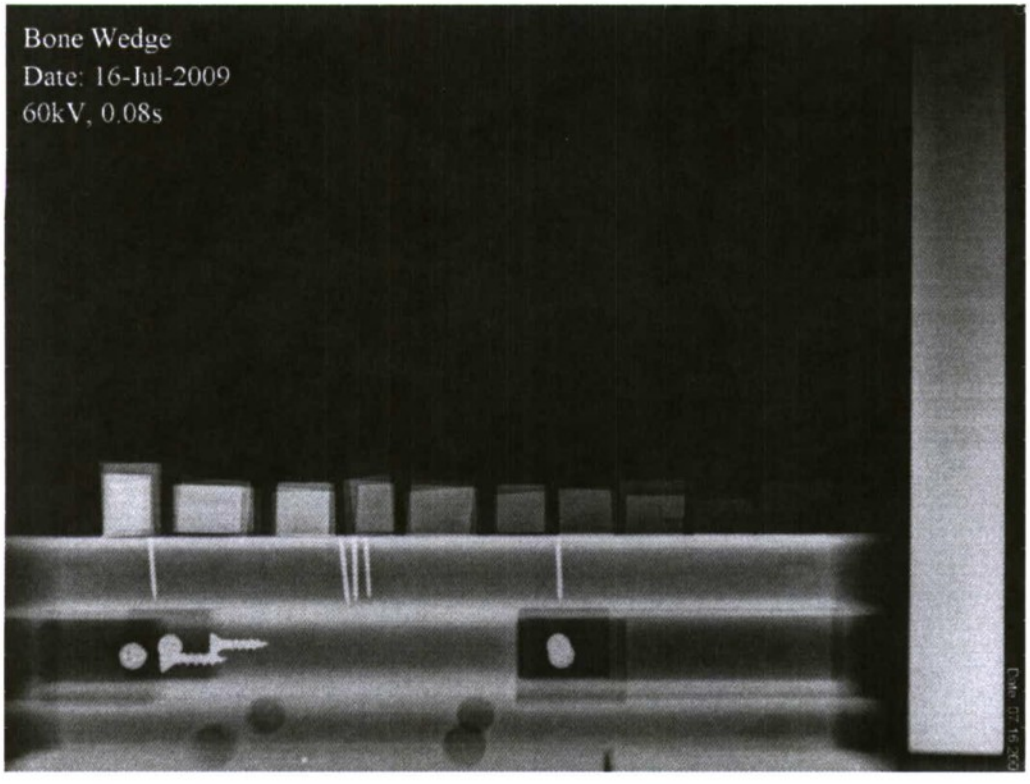


Figure 3: Bone wedge (center) made of 10 cortical bone steps of decreasing thickness (left to right) on a standardized hoof block and Al wedge (right), from a digital radiograph taken at 60 kV, 0.08 seconds and 15 mA.

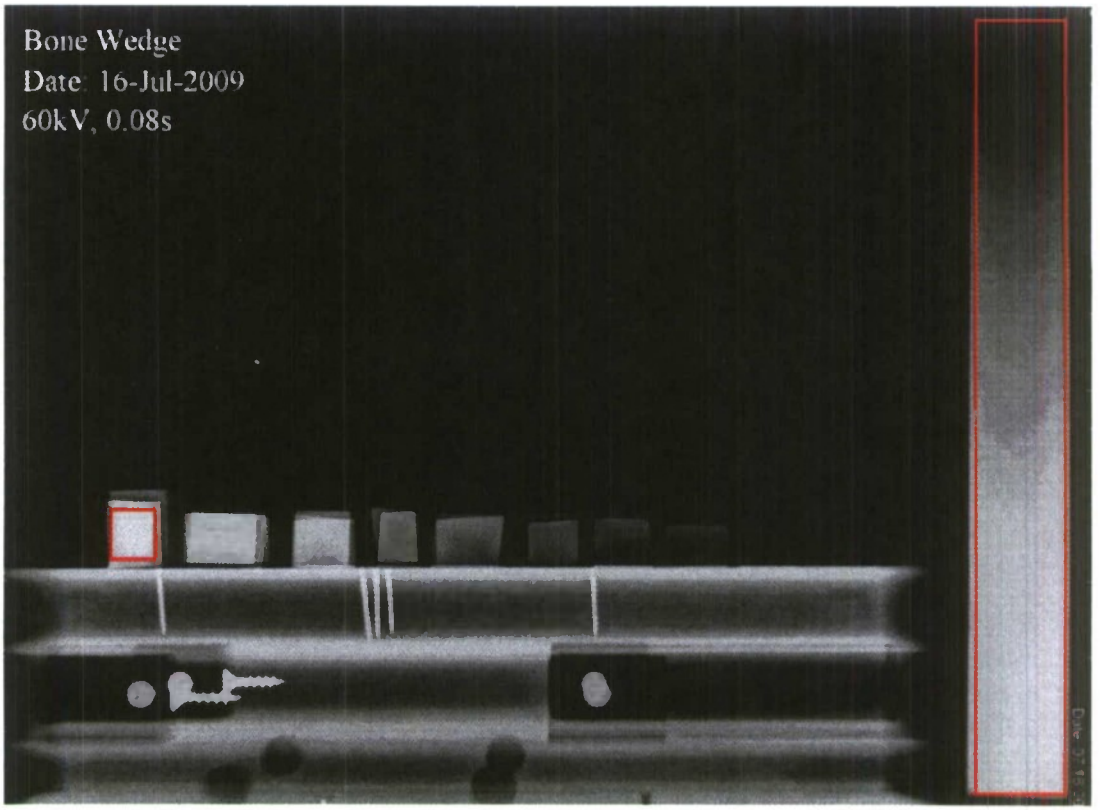


Figure 4. Selection of the regions of interest (red) for measuring Al BDI (right) and bone BDI (left) from the tenth and thickest step of the bone wedge, determined from a digital radiograph taken at taken at 60 kV, 0.08 seconds and 15 mA.

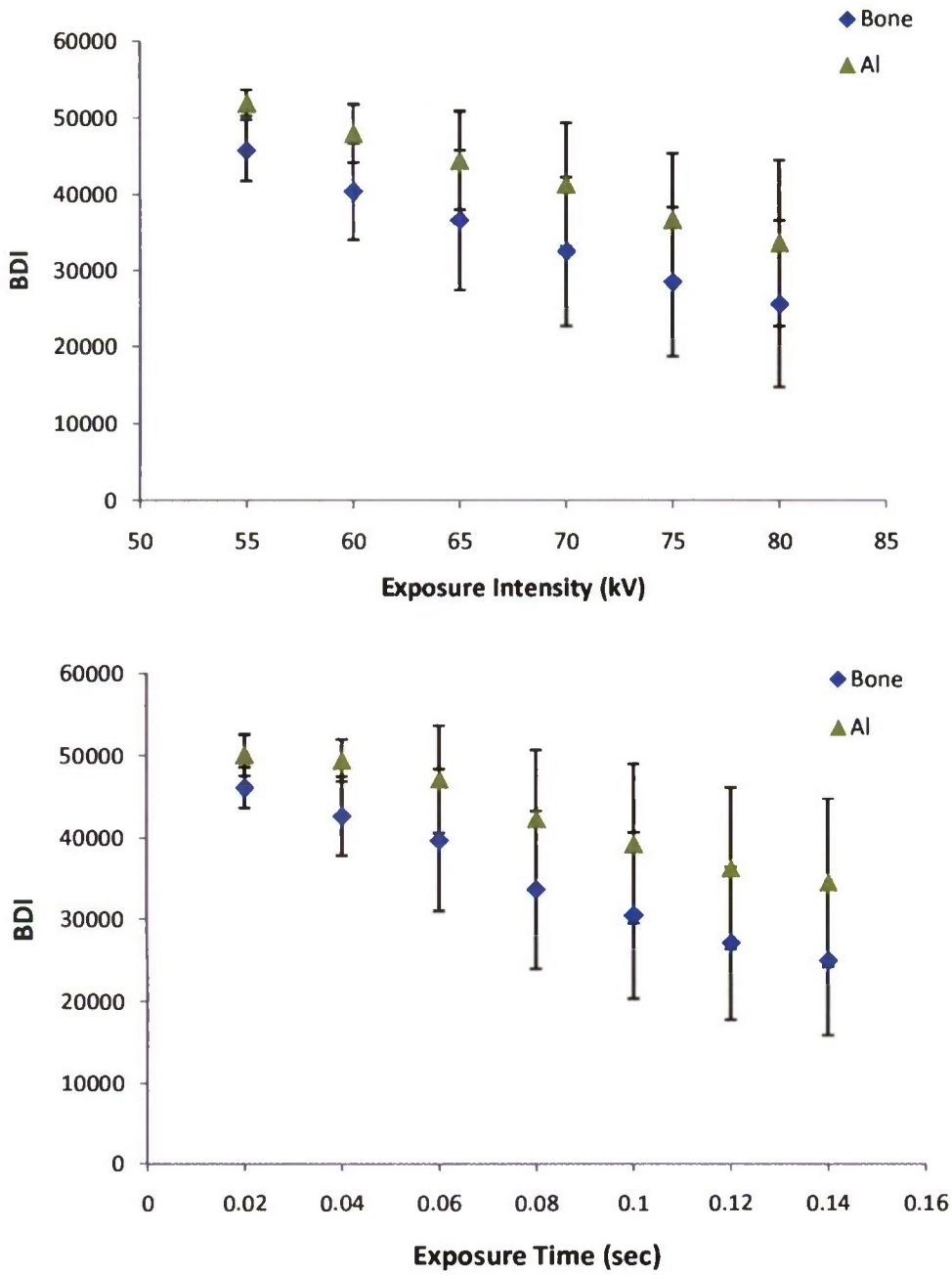


Figure 5. BDI versus exposure intensity (top) and exposure time (bottom) determined from a dorsal palmar radiograph of a third metacarpal cadaver bone and Al wedge taken at varying exposure ($p < 0.001$).

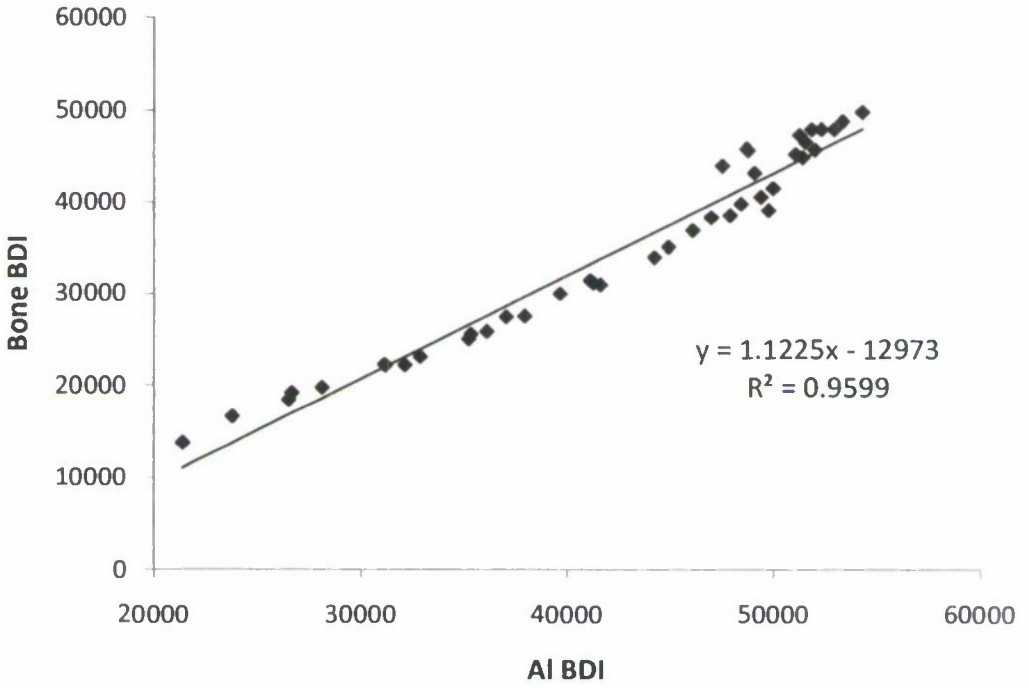


Figure 6. Linear correlation of BDI of a third metacarpal cadaver bone versus Al BDI at varying exposure intensities (55-80 kV) and exposure times (0.02-.014 s) ($p < 0.001$).

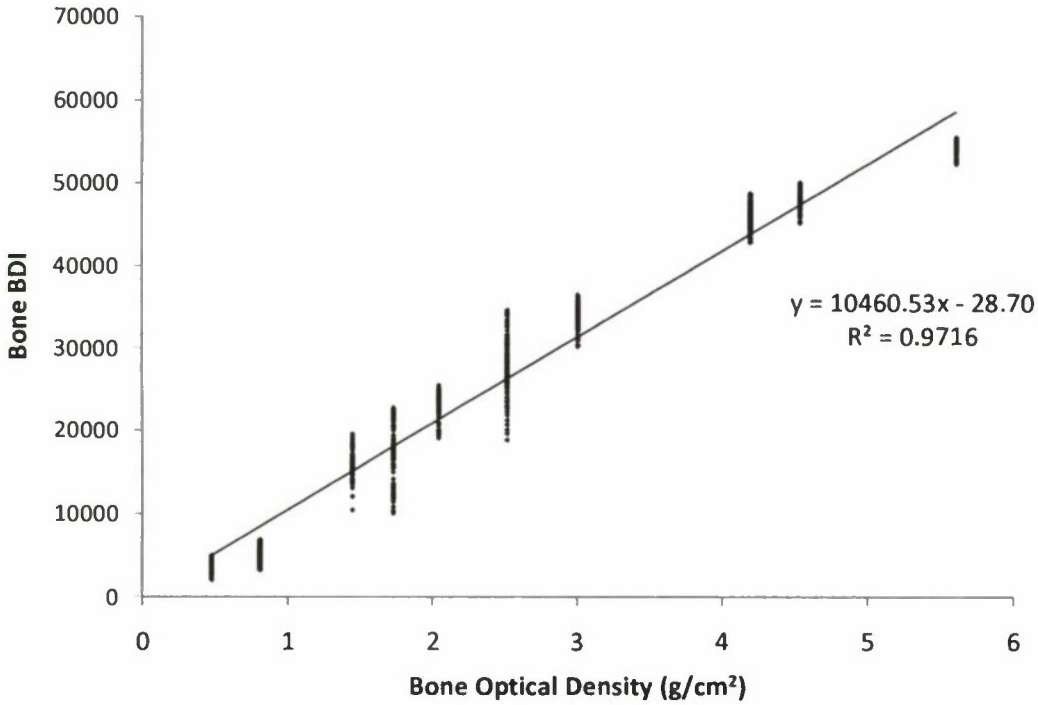


Figure 7. Bone BDI determined from a radiograph of the bone wedge versus bone optical density measured with digital scale and calipers ($p < 0.001$).

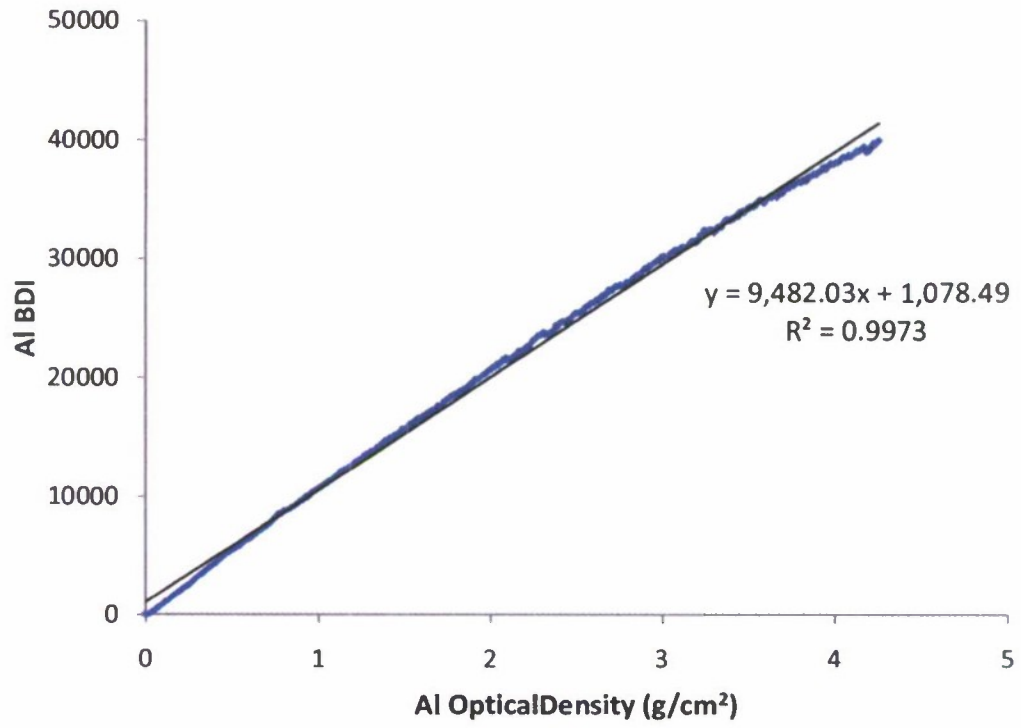


Figure 8. Al BDI versus Al optical density determined for the linear region (0 – 40,000 BDI) of the Al wedge standard curve from a radiograph taken at 66 kV, 0.08 seconds and 15 mA ($p <$ 0.001).

A)

Intensity (kV)	*Exposure Time (sec)*						
	0.02	0.04	0.06	0.08	0.1	0.12	0.14
55	49851.20	47324.93	48851.92	47979.78	45745.73	41531.48	39134.09
60	45854.03	48001.42	44886.04	40580.10	38567.94	33940.24	30970.81
65	45657.83	45258.55	46563.14	35112.90	31168.28	27543.62	25046.51
70	47970.69	39823.54	38583.39	29976.86	25877.64	23134.80	22218.42
75	43981.89	38335.79	31419.15	25567.06	22194.23	19800.11	18432.86
80	43221.42	36932.82	27441.96	22277.3	19188.45	16605.65	13747.36

B)

Intensity (kV)	*Exposure Time (sec)*						
	0.02	0.04	0.06	0.08	0.1	0.12	0.14
55	54328.01	51278.43	53365.63	52982.59	52035.74	50003.53	49789.47
60	48724.45	52346.50	51439.42	49420.01	47922.23	44261.64	41638.44
65	48790.40	51097.91	51579.49	44946.46	41295.99	37956.83	35236.51
70	51871.32	48464.33	47919.15	39674.84	36117.23	32890.75	32145.65
75	47540.82	47008.81	41149.23	35358.32	31197.42	28125.15	26515.69
80	49108.10	46107.06	37058.30	31152.24	26649.52	23781.63	21387.57

Table 1. BDI at the nutrient foramen (A) and Al wedge (B) determined from a dorsal palmar radiograph of a third metacarpal at varying exposure intensities (55 – 80 kV) and exposure times (0.02 – 0.14 seconds).

Step	*Bone BDI ± SD*	*t (cm)*	*D (g/cm^3)*	*D*t (g/cm^2)*
1	3,648 ± 723	0.271	1.76	0.48
2	4,591 ± 757	0.400	2.02	0.81
3	15,962 ± 129	0.754	1.92	1.45
4	17,187 ± 268	0.929	1.86	1.79
5	22,964 ± 133	1.047	1.95	2.04
6	27,536 ± 280	1.325	1.90	2.50
7	33,969 ± 127	1.601	1.87	3.00
8	46,201 ± 115	2.140	1.96	4.12
9	47,870 ± 952	2.498	1.81	4.53
10	54,151 ± 691	3.004	1.87	5.61

Table 2: BDI and standard deviation (SD), thickness (t), density (D) and optical bone density ($D*t$) determined for ten steps of the cortical bone wedge. BDI was measured from a digital radiograph taken at 60 kV, 0.08 seconds and 15 mA. Thickness was determined using digital calipers, and density was determined from weight (g) using a digital balance divided by volume (cm^3) determined using digital calipers.

Assessment of Single-use Cold Chain Packaging and Logistics for Temperature Sensitive Products

Principal Investigator

Jay Singh, Ph.D. Industrial Technology

California Polytechnic State University
San Luis Obispo, CA

Assessment of Single-use Cold Chain Packaging and Logistics for Temperature Sensitive Products

Jay Singh
Packaging Program, Industrial Technology, Cal Poly State University

ABSTRACT

Cold chain solutions typically involve excessive packaging to ensure that the desired product temperature is maintained through distribution thereby increasing the logistics related costs, which are estimated to be six times as compared to the package itself. There are numerous solutions available for shipping temperature sensitive products that include a variety of packaging materials as well as refrigerants. Though characteristics such as the R-values of package systems and the melting points and heat absorption rates of various refrigerants have been studied in the past, none of the past studies have evaluated the effect of distribution (vibration, shocks, ambient temperature, etc.) on the reliability of the cold chain packaging solutions. This research studied the temperature profiles for factors such as the different densities for a given thickness of packaging material, wall thicknesses and distribution environments. The findings of this study will be decisive in designing cost efficient and practical single-use cold chain transportation solutions for temperature sensitive products.

1.0 INTRODUCTION

Cold chain refers to the transportation of temperature sensitive products, whether they be pharmaceutical, consumable or chemical in nature, along a supply chain through thermal and refrigerated packaging methods and the logistical planning to protect the integrity of these shipments. Thermal abuse is a primary concern during the distribution of temperature sensitive goods such as pharmaceutical, food and chemical products. Thermally insulated packaging can maintain product temperatures within acceptable ranges and slow down the deterioration of the product in the distribution environment until it reaches the consumer. In addition to a high resistance to the transfer of heat (R-value), a good thermally insulating material must have various characteristics, depending upon the application. For packaging applications, low cost, low moisture susceptibility, ease of fabrication and transportation, consumer appeal and mechanical strength are the most relevant characteristics. Various types of temperature controlled packaging systems including containers and refrigerants are commonly employed to provide an uncompromised passage of the product through distribution.

The reliance on the cold chain has continued to gain importance over the past decades. Within the pharmaceutical industry for instance, the testing, production and movement of drugs relies heavily on controlled and uncompromised transfer of shipments. A large fraction of the pharmaceutical products that move along the cold chain are in the experimental or developmental phase [1]. Clinical research and trials is a major part of the industry that costs millions of dollars, but one that also experiences a failure rate of around 80% [1]. According to the Healthcare Distribution Management Association, of the close to

650 billion dollars in bio-pharmaceutical distribution in 2005, about 40% are drugs that are temperature sensitive [1]. This makes the cold chain responsible for transporting a near 260 billion dollar investment [1]. If these shipments should experience any unanticipated exposure to variant temperature levels, they run the risk of becoming ineffective or even harmful to patients. According to Alistair Black, Project Director at Aptuit Clinical Supplies Europe, "Maintaining the chemical and therapeutic integrity of investigational medicinal products poses special cold chain challenges, since clinical trials require multiple small shipments to as many as 300 study sites worldwide" [2]. Similar arguments can be made for other temperature sensitive products such as perishables (processed or otherwise) and chemicals. Food products for example, when stored above 5 °C provide a rich growth environment for both spoilage organisms and pathogens, resulting in greater risk of both economic loss and an outbreak of food borne illness [3].

Temperatures a packaged product experiences during distribution (Figure 1) need to be addressed by an optimum cold chain packaging solution. Over the past decades numerous competent cold chain packaging systems have evolved. While the carrier-controlled thermal chains provide refrigerated trailers for the transportation of goods and the two-way systems use reusable shipping containers, the one-way shipment of temperature sensitive products in single-use packaging has been critical in the cold chain and often relies on expedited shipping service providers such as FedEx, UPS and DHL. Being able to ensure that a shipment will remain within a temperature range for an extended period of time comes down largely to the type of container that is used and the refrigeration method. Factors such as duration of transit, the size of the shipment and the ambient or outside temperatures experienced are important in deciding what type of packaging is required. They can range from small insulated boxes that require dry ice or gel packs, rolling containers, to a 53 footer reefer that have their own powered refrigeration units.

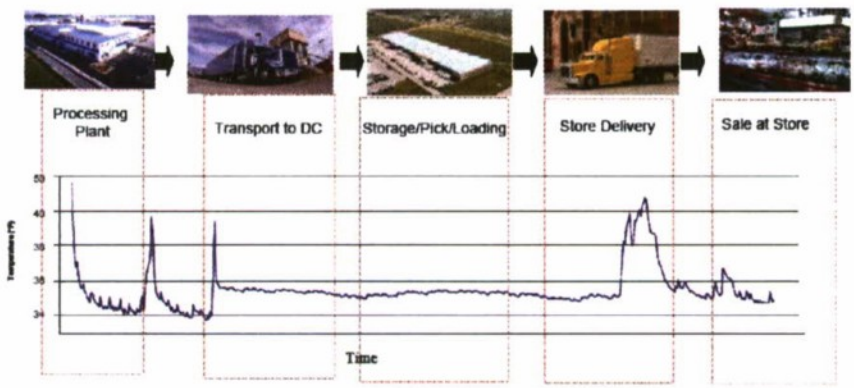


Figure 1: Example of Temperature Variations in Typical Distribution Environment

The choice of distribution system is governed by payloads, transit time, temperature sensitivity of the product, customer acceptance and cost. One-way systems have emerged as the most popular option for their ease of application. Insulated containers provide insulation using different packaging material combinations and refrigerants in order to maintain the desired temperatures and preserve product quality.

Of the limited published researches evaluating the R-values (resistance to heat flow) of package systems and the melting points and heat absorption rates of refrigerants, none have focused on the effect of these values as related to distribution simulation. In other words, all of the past studies have been conducted under "near-static" and not "dynamic" environments. The latter presents unique challenges to cold chain solutions for temperature sensitive products in terms of mechanical shocks, vibration, compression, cyclic temperature and humidity exposure, etc.

This research studied the temperature profiles for factors such as different densities for a given thickness of packaging material, wall thicknesses and distribution environments. Refrigerants and insulated packages available in the market today were used in this study.

This project has implications for the US armed forces in terms of enhanced understanding towards developing or sourcing insulated packaging to withstand the rigors of defined transportation itineraries. As an example, the Coagulation Factor VIIa drug, which promotes blood clotting and costs $3500 a vial, is supplied to the troops on the battlefield as emergency dosage using insulated packaging [4]. Benefits such as tight temperature maintenance for products like Coagulation Factor VIIa (bio-pharma), increased cargo space and increased portability may be realized as a direct result of this research. This project aligns with the "national security application" category of the ONR/DOD areas of interest by addressing logistics and transportation engineering and holds tremendous potential to build pharmaceutical and medical device industry ties.

Objectives:

Laboratory studies were conducted to compare the effectiveness of thermally insulated packaging performance in the 2-8 °C range due to following factors:

1. Different densities for a given thickness of packaging material (EPS)
2. Different thickness of material types: EPS, recycled EPS, Vacuum Insulated Packaging (VIP) and Polyurethane (PU)
3. R-values of insulated shipping containers
4. Cyclic conditioning (temperature and humidity) effect on gel refrigerant/payload performance
5. Distribution environment testing based on ASTM/ISTA protocols

2.0 MATERIALS AND METHODS

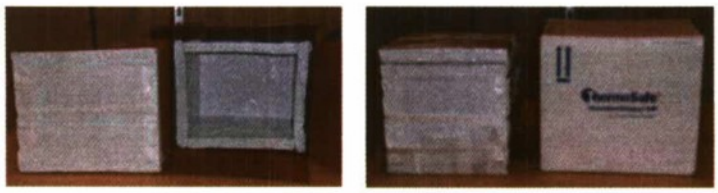
2.1 Materials

2.1.1 Thermal Insulated Containers

Twenty different types of thermally insulated containers were obtained for this study and are listed in Table 1 and shown in Figure 2. The material types, wall thicknesses, densities, total internal surface area, total internal volumes and types of tests conducted are shown in Table 1.

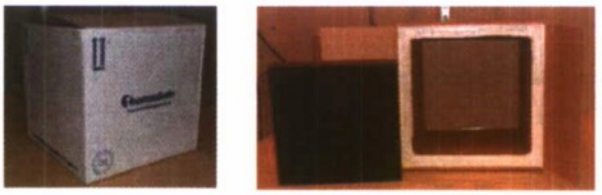
Table 1: Thermally Insulated Container Specifications

Container ID	Vendor	Material	Wall Thickness (in)	Density (pcf)	Total Surface Area (in^2)	Internal Volume (in^3)
E46UPS	Thermosafe	Polyurethane Foam	1.00		520.00	800.00
E36UPS	Thermosafe	Polyurethane Foam	2.00		442.00	616.00
E38UPS	Thermosafe	Polyurethane Foam	3.00		460.00	654.50
42VIP-UPS	Thermosafe	Vacuum Insulated Panels	0.50		484.00	720.00
37VIP-UPS	Thermosafe	Vacuum Insulated Panels	1.00		448.00	640.00
F-900	Cold Ice	Virgin EPS	1.00	1.25	458.50	666.19
F-105	Cold Ice	Virgin EPS	1.50	1.25	328.75	403.00
F-115	Cold Ice	Virgin EPS	2.00	1.25	408.38	561.52
Fageradala 1.75-1.25-100	Fageradala	EPS, 100% recycled content	1.75	1.25	399.25	525.00
Fageradala 1.75-1.25-90	Fageradala	EPS, 90% recycled content	1.75	1.25	399.25	525.00
Fageradala 1.75-1.25-80	Fageradala	EPS, 80% recycled content	1.75	1.25	399.25	525.00
Fageradala 1.75-1.5-100	Fageradala	EPS, 100% recycled content	1.75	1.50	399.25	525.00
Fageradala 1.75-1.5-90	Fageradala	EPS, 90% recycled content	1.75	1.50	399.25	525.00
Fageradala 1.75-1.5-80	Fageradala	EPS, 80% recycled content	1.75	1.50	399.25	525.00
Fageradala 1.5-1.25-100	Fageradala	EPS, 100% recycled content	1.50	1.25	396.00	526.50
Fageradala 1.5-1.25-90	Fageradala	EPS, 90% recycled content	1.50	1.25	396.00	526.50
Fageradala 1.5-1.25-80	Fageradala	EPS, 80% recycled content	1.50	1.25	396.00	526.50
Fageradala 1.5-1.5-100	Fageradala	EPS, 100% recycled content	1.50	1.50	396.00	526.50
Fageradala 1.5-1.5-90	Fageradala	EPS, 90% recycled content	1.50	1.50	396.00	526.50
Fageradala 1.5-1.5-80	Fageradala	EPS, 80% recycled content	1.50	1.50	396.00	526.50

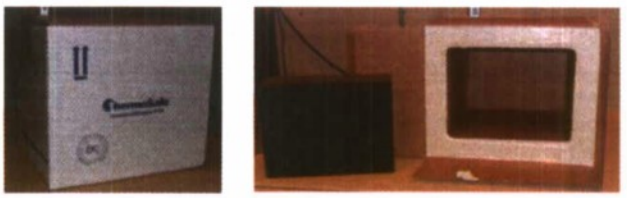


37VIP-UPS VIP Container 1 Inch Wall

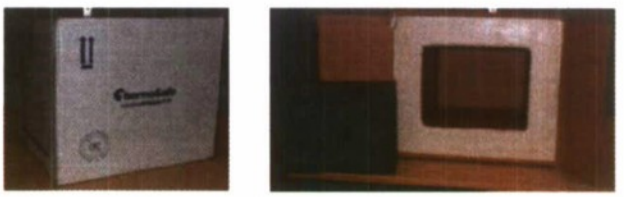
42VIP-UPS VIP Container 0.5 Inch Wall Thickness



E46-UPS Urethane Container 1 Inch Wall

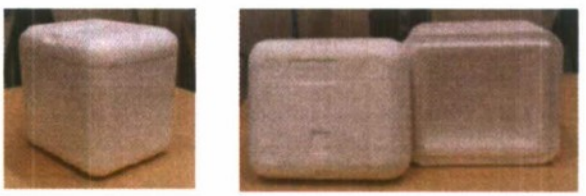


E36-UPS Urethane Container 2 Inch Wall Thickness

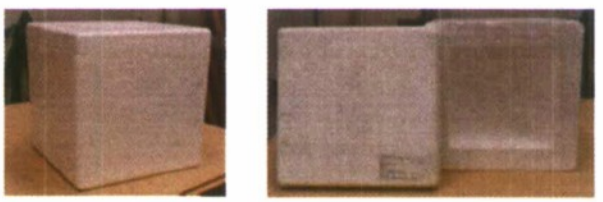


E38-UPS Urethane Container 3 Inch Wall Thickness

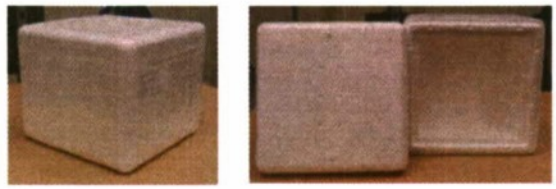
F-900 Virgin EPS Container 1 Inch Wall Thickness



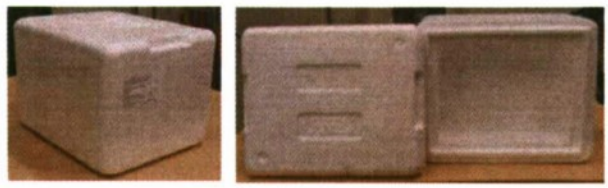
F-105 Virgin EPS Container 1.5 Inch Wall Thickness



F-115 Virgin EPS Container 2 Inch Wall



Fageradala Recycled EPS 1.5 Inch Wall



Fageradala Recycled EPS 1.75 Inch Wall Thickness

Figure 2: Thermal Insulated Containers Studied

2.1.2 Refrigerant

Phase 5™/ 5° C Phase Change Material (PCM) was obtained from TCP Reliable (Edison, NJ). This PCM enclosed in high density polyethylene containers (Figure 3) measured 5.5" x 5.5" x 1" and weighed 10.56 oz +/- 1.5% each. Filled with a proprietary material that undergoes phase change at 5°C (middle of the 2-8 °C range), these PCM cartridges were frozen at 0 °C, as per the manufacturer's recommendation, prior to all tests.

Figure 3: 5° C Phase Change Material

2.1.3 Product

In order to control any variability in the product (payload) used for this study, twenty 20 ml borosilicate test tubes measuring 0.63 in x 5.91 in were used. A sample holder to hold these test tubes in place was constructed from E-flute corrugated fiberboard. The test tubes were filled to the 20 ml level and plugged with rubber stoppers measuring 0.75 - 0.55 inch in diameters. A centrally located test tube in the payload pack was instrumented by inserting the temperature monitor probe through the rubber stoppers. Figure 4 below shows the payload, the PCM arrangement around the payload, an instrumented test tube and an example of the loaded container.

Sample Holder and 1 Pack of Phase 5 Phase Change Material

Orientation of 4 Phase Packs around Sample Holder

TempTale® Data Recorder and Probe Sample Tube

Loaded Container

Figure 4: Stages of Preparing a Loaded Container for Experimentation

2.1.4 Temperature Monitors

TempTale® Model 4 (Figure 5) temperature monitors from Sensitech Inc. were used to monitor the temperature inside the insulating containers tested [6]. The temperature monitors had flexible probes that were inserted into the package to monitor and record the temperatures. The devices were factory calibrated with the accuracy tested to NIST traceable standards. The Sensitech Temp Tale 4 temperature monitors have a resolution of 0.1°C and measured in the -30 °C to 70 °C range. The sensor accuracies are ±1.1°C from –30°C to –18°C; ±0.55°C from –18°C to 50°C and ±1.1°C from 50°C to 70°C. The monitors were setup to record the temperatures at every 30 seconds for this research.

Figure 5: TempTale® 4 Probe and Dual Sensor Temperature Monitor

2.2 Methods

Two different tests were performed: R-value measurement and a study of the effect of handling to the temperature profiles inside the shippers. All test configurations were conducted in triplicate and the details are provided below.

2.2.1 R-value Measurement

The resistance to the flow of heat through an insulating package designated as the system R-value was calculated using ice-melt tests [7]. The test is based on the principal that 1 kg of regular ice must absorb 335 kJ of heat to melt. By placing a known quantity of ice inside the container, the rate of heat transfer into the container was calculated from the quantity of ice melted at the end of test. Figure 6 shows the methodology used for R-value measurement. The system R-value for the packages was calculated using the following equation [7],

$$System\ R\text{-}value = \frac{(Surface\ area)\ (Temperature\ difference)}{(Melt\ rate)\ (Latent\ heat)}$$

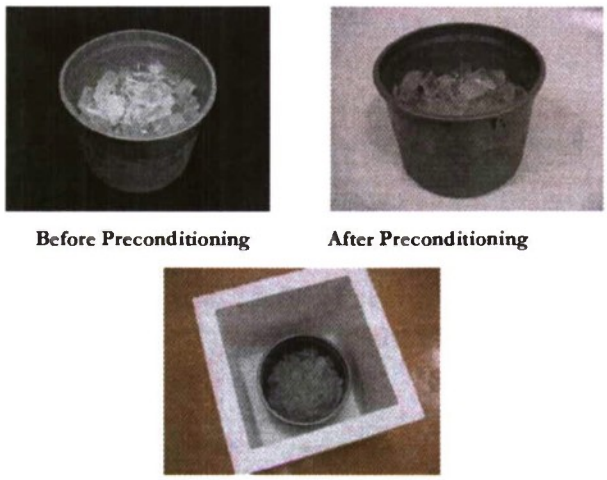


Figure 6: Set Up for R-Value Measurement by Ice Melt Test

2.2.2 Distribution Simulation Testing

ASTM and ISTA test protocols were used to conduct simulation of distribution scenarios experienced by packages in the express single parcel environment. ISTA 7D Procedure [8] (Thermal Controlled Transport Packaging for Parcel Delivery System Shipment), is a development test to evaluate the effects of external temperature exposures of individual packaged products shipped through a parcel delivery system. It can be used for the development of temperature controlled transport packages made of any material and for individual or comparative performance analysis of standard or insulated transport packages against normally encountered conditions. It is designed to measure the relative ability of a package to protect a product when exposed to test cycles that simulate both the range and time of exposure to ambient temperature conditions.

Table 2 presents the test sequence for ISTA Test Procedure 7D with sequences 3, 5 and 7 being required and 1, 2, 4 and 6 being optional [8, 9, 10 &11]. This research included all the optional testing. This sequence represents the typical distribution hazards encountered by small packages in the express shipping environment.

Sequence #	Test Category	Test Type	Test Level
1	Atmospheric preconditioning (ASTM D 3103)	Temperature	Storage conditions for the product and each package element for 24 hrs. min.
2	Shock conditioning (ASTM D 5276)	Drop	Height varies with packaged-product weight
3	Atmospheric (ASTM D 3103)	Temperature	1st Cycle Period of selected test profile
4	Vibration conditioning (ASTM D 4728)	Random Vibration	Overall Grms level of 1.15
5	Atmospheric (ASTM D 3103)	Temperature	2nd Cycle Period of selected test profile
6	Shock conditioning (ASTM D 5276)	Drop	Height varies with packaged-product weight
7	Atmospheric (ASTM D 3103)	Temperature	Remaining Cycle Periods of selected test profile

Table 2: Test Sequence for ISTA Procedure 7D [8]

Table 3, below, shows the cyclic conditioning profile for a 48 hour domestic express small package freight transport (air) used in this study.

Winter Profile			Summer Profile		
Cold Shipping and Cold Receiving			Hot Shipping and Hot Receiving		
Temperature	Cycle/Ramp Period	Cycle/Ramp Period Hours	Temperature	Cycle/Ramp Period	Cycle/Ramp Period Hours
18°C (65°F)	1	0-6	22°C (72°F)	1	0-6
to	2	2 hr ramp	to	2	2 hr ramp
-20°C (-4°F)	3	2 hr at temp	45°C (113°F)	3	2 hr at temp
to	4	2 hr ramp	to	4	2 hr ramp
10°C (50°F)	5	12-24	30°C (86°F)	5	12-24
to	6	2 hr ramp	to	6	2 hr ramp
-20°C (-4°F)	7	2 hr at temp	45°C (113°F)	7	2 hr at temp
to	8	2 hr ramp	to	8	2 hr ramp
10°C (50°F)	9	30-48	30°C (86°F)	9	30-48

Table 3: Cyclic Test Profile for 48 Hour Domestic Express Small Package Freight Transport (Air) [8]

3.0 RESULTS AND DISCUSSION

3.1 R-Value Measurement

The results for R-value testing are reflected in Figures 8 and 9 below. By material type, the performance of the containers was observed to be vacuum insulated panels, polyurethane, recycled content EPS and virgin EPS in decreasing order. Wall thickness, in general, was

observed to be directly proportional to the observed R-values. Considering the containers with one inch thick walls, those with VIP performed 128% and 267% better than the polyurethane and virgin EPS respectively. For the two containers studied with two inch thick walls, containers with polyurethane performed approximately 77% better than those made with virgin EPS.

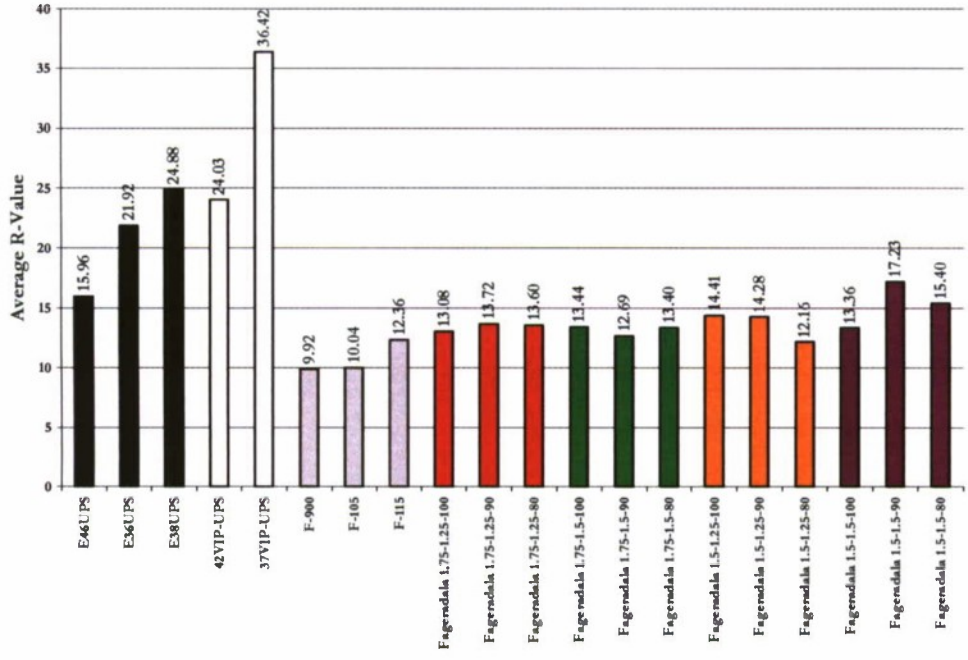


Figure 8: R-Value Results for All Containers Studied

Twelve EPS containers with recycled content were used in this study. Figure 9 shows the effects of wall thickness and foam density on the observed R-values for these containers.

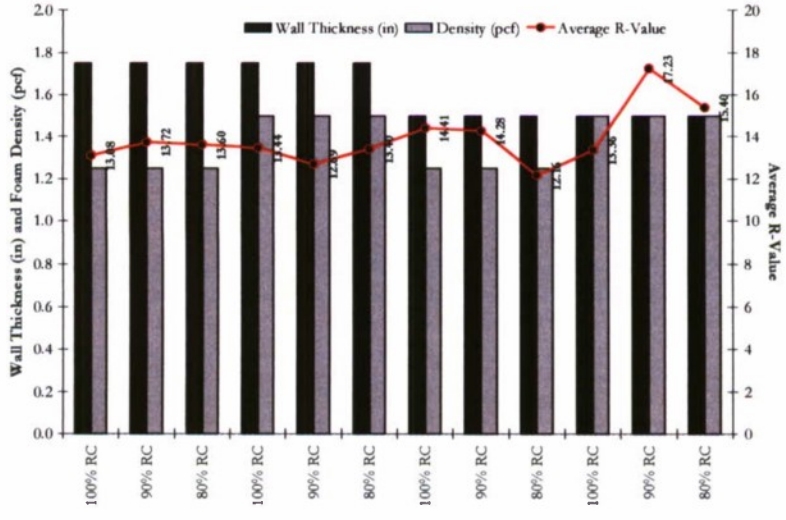


Figure 9: R-Value Result Comparison with Wall Thickness and Densities of Recycled Content EPS Containers

3.2 Handling

The objective of this portion of the testing was to capture the effect of R-value and wall thickness on the length of time (duration) the thermally insulated packages stayed in the desired 2-8 °C range. Figure 10, below, shows the sample frequency distribution for three replicates each of the 20 different types of containers tested for the durations of maintaining the 2-8 °C range. As can be seen, 15 containers maintained the temperature of the product between this range for up to 6 hours, 18 for 12 hours, 10 for 18 hours and 6 for 24 hours. 11 of the containers tested exceeded 24 hour period.

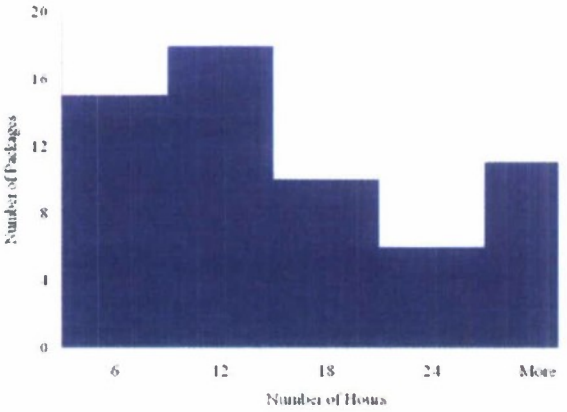


Figure 10: Histogram for Duration of Packages in the 2-8 °C Range

In Table 4, we present summary measures. The mean duration was observed to be 14.41 hours with a median of 10.85, suggesting that the duration is positively skewed where the mean is raised by a few high values. Similarly the higher mean relative to the median indicates that R-value and wall thickness are also positively skewed.

Table 4: Summary Measures

Variable	Mean	Median
Duration in the 2-8 range	14.41	10.85
R-Value	16.10	13.66
Wall Thickness	1.58	1.50

In Table 5 we present the regression results of the model given by the equation below:

$$\ln(\text{duration}) = \beta_0 + \beta_1 Rvalue + \beta_2 Thickness + \varepsilon$$

Table 5: Log Regression Model

Variable	Estimate	Standard Error
Constant	1.987*	0.469
R-Value	0.068*	0.017
Wall Thickness	-0.502*	0.217

Note:

1. *The dependent variable is the natural logarithm of Duration in the 2-8 degree range.*
2. ** Represents significance at the 5% level; the model explains 43.6 percent of the sample variations in duration.*

We use the log-transformation of the duration variable for various reasons. It ensures that the predicted duration is non-negative and enables us to capture interesting non-linearities in the model. Further various model selection measures indicate that the long-linear specification is superior to its linear counterpart. Both the R-value and wall thickness variables are significant at the 5% level, however, R-value exerts a positive influence whereas wall thickness is negatively related to duration. In Figure 11 we use the estimated model to simulate the effect of R-value and wall thickness on duration. For instance, for an R-value of 20, the predicted duration is 17.08, 13.29, and 10.34 for wall thickness of 1, 1.5, and 2 inches respectively. The values increase to 29.35, 18.54, and 14.51 as the R-value increases to 25.

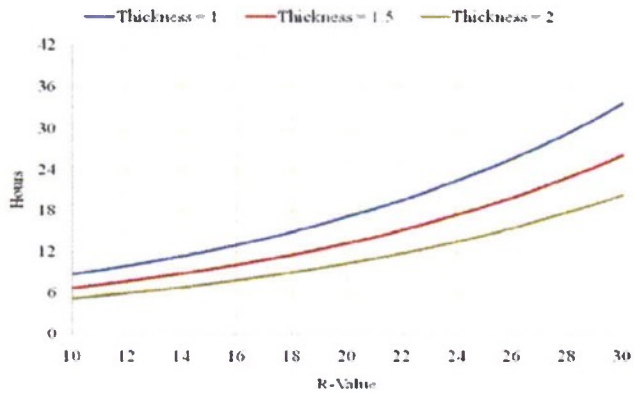


Figure 11: Simulated Duration in the 2-8 °C Range

The above model treats R-value and wall thickness as two independent factors that influence the length of time the package stays in the desired 2-8 °C range. We extend the model to allow interaction between the two factors. In Table 6 we present the regression results of the following model

$$\ln(\text{duration}) = \beta_0 + \beta_1 \text{Rvalue} + \beta_2 \text{Thickness} + \beta_3 \text{Rvalue} * \text{Thickness} + \varepsilon$$

Table 6: Log Regression Model with Cross Products

Variable	Estimate	Standard Error
Constant	5.411*	1.262
R-Value	-0.077	0.053
Wall Thickness	-2.893*	0.851
R-Value*Wall Thickness	0.106*	0.037

Note:

1. *The dependent variable is the natural logarithm of Duration in the 2-8 degree range.*
2. ** represents significance at the 5 percent level; the estimated model explains 60.9 percent of the sample variations in duration.*

This model outperforms the earlier model on the basis of various model selection measures. Interestingly the interaction term is statistically significant at 5%. We present the simulation results with this estimated model in Figure 12. For an R-value of 20, the predicted duration is 22.30, 15.22, and 10.38 for wall thickness of 1, 1.5, and 2, respectively. These values increase to 25.82, 22.99, and 20.46 as the R-value increases to 25. At R-values greater than 27.25, the predicted duration is actually higher for wall thickness of 2 inches.

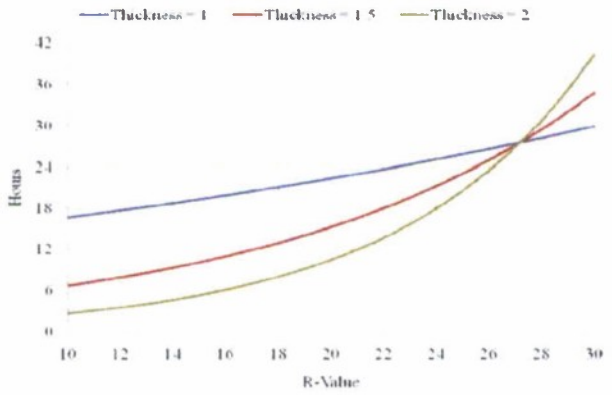


Figure 12: Simulated Duration in the 2-8 °C Range with Cross Products

4.0 CONCLUSIONS

This reserach was undertaken to study the temperature profiles for factors such as different densities for a given thickness of thermally insulating material, wall thicknesses and distribution environments for four different types of materials – polyurethane, expanded polystyrene, recycled expanded polystyrene and vacuum insulated panels. The temperature range of 2-8 °C, critical for pharmaceutical drugs and vaccines, was targeted and 691,000 data points for temperature were collected. Following some of the key conclusions:

For the R-value tests:

- By material type, the performance of the containers can be ranked in a decreasing order as: vacuum insulated panels, polyurethane, recycled content EPS and virgin EPS.
- Wall thickness, in general, was observed to be directly proportional to the observed R-values.

- Considering the containers with one inch thick walls, those with VIP performed 128% and 267% better than the polyurethane and virgin EPS respectively.
- For the two containers studied with two inch thick walls, containers with polyurethane performed approximately 77% better than those made with virgin EPS.
- Recycled content EPS containers outperformed the virgin EPS containers for all densities and thicknesses studied

The objective of the handling simulation study was to capture the effect of R-value and wall thickness on the length of time the thermally insulated packages stayed in the 2-8 °C range. Only approximately 20% of the containers tested were able to maintain the temperature of the product at 2-8 °C for more than 24 hours with a mean duration of 14.41 hours for all containers tested. The containers as well as the refrigerants used in the study were obtained from sources commonly employed for cold chain solutions. As a result of this finding, we are presently collecting additional data for non-handled scenario using the ISTA 7D protocol to compare the performance of these containers.

The regression results concluded with an interesting finding that, though both the R-value and wall thickness variables were significant at 5% level, R-value exerted a positive influence whereas the wall thickness was negatively related to duration. It was also observed that at R-values greater than 27.25, the predicted duration was actually higher for wall thickness of two inches. This finding could possibly be used by manufacturers of these containers in using greater wall thickness only for materials with high R-values.

Note:

Along with accomplishing our original scope, we were able to undertake a few additional experiments. An exciting preliminary finding stemming from an experiment aimed at observing the effect of punctures from distribution abuse seems to indicate that a controlled amount of ventilation on top of the thermally insulated containers may result in higher R-values. Data to study effects of the ventilation volume and locations is presently underway. A similar study related to gaps between lids and containers is also underway. If the data analysis is able to support this preliminary observation, thermally insulated containers could be manufactured with the ventilations incorporated thereby improving their ability to maintain the temperature of the product to within 2-8 °C range without any other modifications to the container design.

Peer reviewed papers resulting from this grant are presently being prepared on the following topics:

1. The effect of handling on temperatures experienced by products in thermal insulated containers for express shipments
2. The effect of recyled content, densities and wall thickness on R-values of EPS containers
3. The effect of ventilation on R-value for thermal insulated containers for express shipments
4. The effect of ambient temperature, inside and outside thermally insulated containers, on the temperature of the product

The funding towards this project is appreciated and will result in soliciting funded projects from the pharmaceutical/bio-med industry as well as through competitive grants.

References

1. Rodrigue, J and Craig, M., The Geography of Transport Systems, Chapter 5, (2006), New York: Routledge, ISBN 0-415-35441-2
2. Alastair Black, "E-Logistics in Cold Chain Management", European Pharmaceutical Contractor (EPC), Summer 2003, http://www.samedanltd.com/members/archives/EPC/Summer2003/AlastairBlack.htm , accessed 4/28/08
3. Tom Kevan, "Control of the Cold Chain," Frontline Solutions, www.frontlinetoday.com, accessed 4/28/08
4. Wikipedia, Factor VII, http://en.wikipedia.org/wiki/Factor_VII, accessed 5/7/08
5. Singh, S. P., Burgess, G., Singh, J. "Performance Comparison of Thermal Insulated Packaging Boxes, Bags and Refrigerants for Single Parcel Shipments," Packaging Technology and Science, 2008; 21: 25–35
6. http://www.sensitech.com/products/hardware/temp_monitors/index.html, , accessed 5/1/08
7. Desjarlais A. O., Zarr R. R. Insulation Materials: Testing and Applications, 4th Volume, ASTM International, ISBN: 0-8031-1278-5
8. ISTA 7D, Thermal Controlled Transport Packaging for Parcel Delivery System Shipment, International Safe Transit Association, East Lansing, MI, USA, 2007
9. ASTM D3103-07, Standard Test Method for Thermal Insulation Performance of Distribution Packages, Vol. 15.10, American Society of Testing and Materials, West Conshohocken, PA, USA, 2007.
10. ASTM D5276-98(2004), Standard Test Method for Drop Test of Loaded Containers by Free Fall, Vol. 15.10, American Society of Testing and Materials, West Conshohocken, PA, USA, 2007.
11. ASTM D4728-06 Standard Test Method for Random Vibration Testing of Shipping Containers, Vol. 15.10, American Society of Testing and Materials, West Conshohocken, PA, USA, 2

Human performance

Project Reports

Construction of a Probiotic Peptidase Delivery System to Maximize Human Health

Principal Investigator:

Michael Black, Ph. D.
Biological Sciences
California Polytechnic State University
San Luis Obispo, CA

Abstract:

Celiac disease (CD) is an autoimmune disorder that affects approximately 1% of the population. CD is characterized by intestinal villus atrophy after consumption of gluten from wheat, barley, or rye. Patients with CD often experience abdominal pain, diarrhea, malnutrition, fatigue, and a failure to thrive. There is currently no treatment for CD. Patients must live on a strict lifelong exclusion of dietary gluten. Due to the high content of gluten in western diets and poor labeling of gluten content, adherence to a gluten free diet (GFD) is difficult. Nearly all the enzymes that can digest the gluten peptide are sensitive to the stomach's low pH. As a result, dietary supplementation with enzymes to digest gluten has yet to produce a viable alternative treatment to a GFD.

We propose to use a resident microbe of the human intestinal tract to express a peptidase to digest the immunoreactive gluten fragments. The bacteria, *L. reuteri*, will colonize the host's intestines and digest the gluten peptides before causing an autoimmune response. To accomplish this task, this report describes a food grade, plasmid based system to integrate genes into the genome of *L. reuteri* that was generated using the C3RP support over the 2008-2010 period. The plasmid system utilizes an origin of replication that requires a protein, RepA, to propagate itself. A helper plasmid provides the RepA protein in trans to an integration plasmid that cannot provide RepA to itself. The integration plasmid carries a homologous region to the genome of *L. reuteri* allowing for targeted genomic integration. The integration plasmid will not replicate on its own, and will be integrated into the genome if the helper plasmid is absent. To select for these genomic integrants the integration plasmid expresses an erythromycin resistance marker. Using the Cre/Lox system the antibiotic resistance will be removed from the bacterial genome to re-establish the *L. reuteri*'s food grade status. This report describes the construction and verification of the above mentioned plasmid tool kit containing the helper, integration, and Cre expression plasmids to integrate genes into the *L. reuteri* genome.

I. Background:

The goals of this project are to develop an engineered probiotic that will be an enduring treatment for CD. We propose to engineer a common resident microbe of the human intestinal flora to digest the gluten peptides before they can trigger damage the gut lining. The bacteria will be modified to express a peptidase that digests the immunogenic fragments of gluten on its surface. We chose to anchor the peptidase to the surface of the probiotic instead of secreting the peptidase because the localization of the peptidase would maximize the effectiveness of the probiotic. If the peptidase is secreted from the bacteria it will constantly be lost from the site of pathology. However, if we anchor the peptidase to the surface of the bacteria, it can colonize the site of pathology and localize the peptidase on the intestinal walls where CD pathology is observed. The peptidase gene will be integrated and expressed from the genome of the bacteria to ensure persistent gene expression. If the peptidase were expressed from a plasmid, the plasmid would be lost to the environment without constant antibiotic antibiotic selection.

Initial work to develop an expression system for our peptidase was done by an MS student supported by previous C3RP funding, Matthew Shurtleff. For this expression system to be effective, constitutive high expression of the peptidase is required. To meet this need, the promoter elements of the gene SlpA from Lactobacillus acidophilus was chosen, as SlpA is constitutively expressed at high levels (about 10 percent of the overall cellular protein) [1]. Furthermore, the S-layer protein is actively excreted from the bacterial cell and anchored in the

cell wall, so that is creates a matrix around the bacterial cell [2]. Due to the large numbers of S-proteins found on cellular surfaces, researchers have estimated that an average of approximately 500 S-layer proteins must be synthesized, translocated and anchored every second [2]. Restriction digests of isolated plasmid constructs with the Slp-A expression system displayed unexpected banding patterns, indicating the plasmid possessed deletions in the open reading frame (ORF) after the SlpA promoter. In expression experiments, the Slp-A leader and anchor were found to be highly toxic to *E. coli* when expressed from the Slp-A promoter. Although *E. coli* is not the selected host for our delivery system, it plays a key role in the development of the probiotic vectors.

II. Results

To alleviate the toxicity problems when using the SlpA promoter elements, a different expression system was chosen. To avoid future complications with shuttling the vector between *E. coli* and *L. reuteri*, we decided that for our proof of principle we would use an inducible promoter. For our new expression system, we sought high expression in *L. reuteri* with minimal expression in *E. coli* to eliminate the toxicity during plasmid construction. Furthermore, the new expression system would not target the expression cassette to the cell surface. Without protein targeting and anchoring sequences, we could mitigate the toxicity potentially imposed upon *E. coli* during the construction of the integration cassettes.

Although gene transcription is well understood in Gram-negative bacteria such as *E. coli*, there has been limited research on gene transcription in Gram positive bacteria [3]. Many inducible systems require multiple genetic elements that are optimized for use in *E. coli*, but may not work in *L. reuteri*. Therefore if we were to use an inducible promoter we would need a system that has been shown to work in Gram positive bacteria. We chose to use a promoter that drives the expression of the Clp family of proteins involved in mitigating environmental stress, particularly heat stress. In McCracken et.al. (2000) P-clpC was compared to the P-slpA along with a number of different promoters in a variety of species of lactic acid bacteria. P-clpC drove strong expression of the reporter gene, GusA, but at lower levels than reported for P-slpA. With this information in hand, we looked into the Clp family of promoters for a potential promoter we could use to drive expression of our reporter protein.

Before expressing the peptidase with ClpP we expressed benign reporter genes to ensure the integration system worked. One of the reporter genes we have chosen to use is the protein beta-glucuronidase (GusA). GusA activity has not been shown in *L. reuteri*, therefore, detected GusA activity would be attributed to our expression system. The GusA reporter systems had many different substrates that allow for enzymatic activity detection of using chromogenic substrates. This study used X-Gluc and PNP-Gluc which are analogues to X-Gal and ONPG to the beta-galactosidase substrates that are commonly used when attempting alpha-complementation. When X-gluc is cleaved by GusA, it forms a blue precipitate on agar plates. This allows for rapid screening of the bacterial clones which contain GusA activity. When PNP-Gluc is cleaved by GusA, it forms a yellow soluble compound that can be detected using spectrophotometric techniques. Therefore, the enzymatic units of the cells can be determined with simple chromogenic assay procedures. The construction of this inducible reporter plasmid is shown in Figure 1.

In addition to the enzymatic reporter, we also wanted an immediate assay that could be checked in vivo. We chose to use green fluorescent protein (GFP) as a secondary reporter

protein. GFP offers a number of benefits that complement the GusA reporter system. Cellular clones with GFP can be quickly screened by detecting visible fluorescence when exposed to long or short wave ultraviolet light (UV). If necessary, there are also antibodies against GFP that can quantify the amount of GFP in the cell. Using these two systems (GusA and GFP) we can generate both a readily detectable qualitative measure with GFP, and a simple procedure for quantitative measurements using the chromogenic assay for GusA.

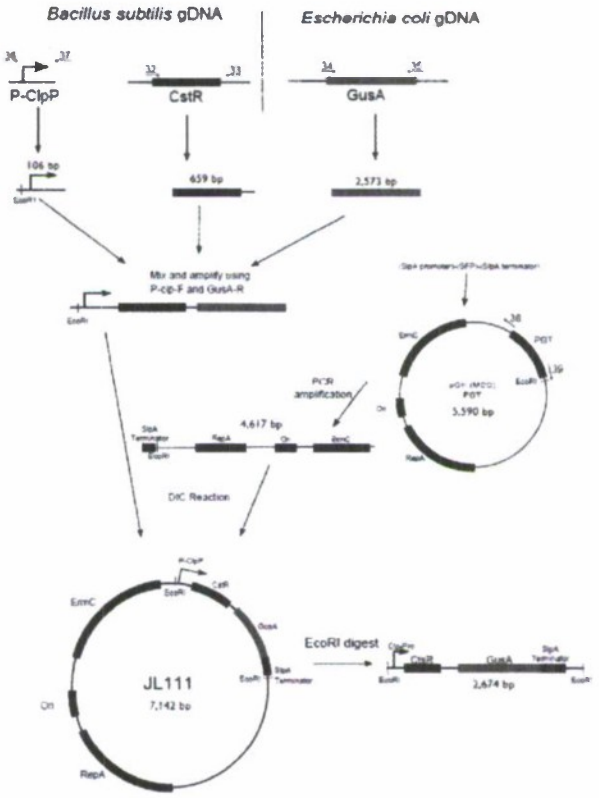


Figure 1: Construction of the CPC expression cassette.

We had initially hoped that our new reporter system would not be active in *E. coli*, yet be active in *L. reuteri*. However, we saw strong activity of the promoter in *E. coli* harboring pGK-clp-Gus or pGK-clp-GFP. Reduced Gus activity observed in the presence of the CtsR protein (middle column compared to right column in Figure 2) lead us to believe the CtsR protein inhibits much of the Clp-P activity in *E. coli*. However, it does not completely inhibit the Clp-P promoter's activity. Therefore, if future peptidase cassettes are toxic in *E. coli*, the Clp-P promoter system may not be suitable. When building the CpC cassette (Clp-P-CtsR-GusA/GFP) there were concerns that the 150 nt spacer between the CtsR protein and the expression cassette might contain a transcriptional terminator sequence. In Figure 2, the cells that contain pGKCpC-GUS show higher GusA activity (~6X) than cells without the pGK plasmid. This piece of evidence supports increased levels of GusA as a result of the plasmid's gene expression. When the CtsR protein is removed from the CpC cassette (pGK-clp-Gus), there is a dramatic increase in the levels of Gus activity. The pGK Clp-Gus construct shows ~141X and ~812X more activity

when compared to pGK CpC-Gus and WT strains of *E. coli*. The reduced Gus activity observed in the presence of the CtsR protein (middle column compared to right column in Figure 2) lead us to believe the CtsR protein effectively inhibits much of the Clp-P activity. GFP expression under Clp-P was also monitored (Figure 3). Strong fluorescence was seen in cells containing pGK-clp-GFP. However, there was background fluorescence observed in *E. coli* with pGK-clp-Gus. The fluorescence is so much brighter with pGK-Clp-GFP that we are confident the fluorescence seen is due to GFP expressed by our plasmid, and not background fluorescence.

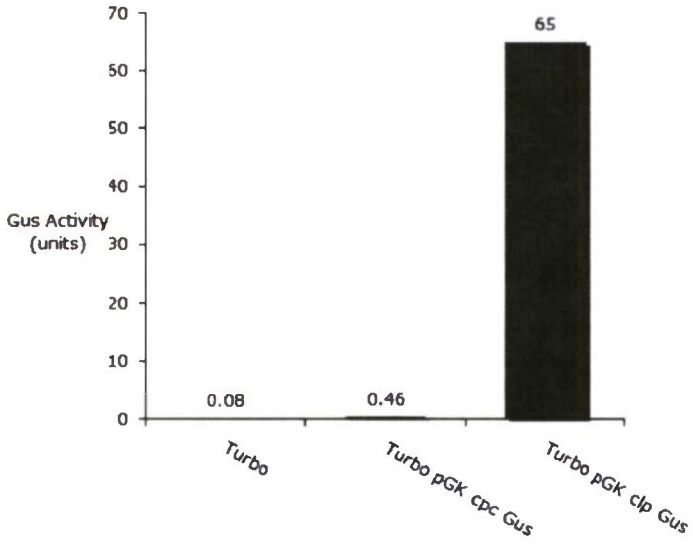


Figure 2: Activity of GusA in *E. coli*. Assays were done in a TURBO *E. coli* strain background. Enzymatic units of GusA were defined as μM pNP/min.

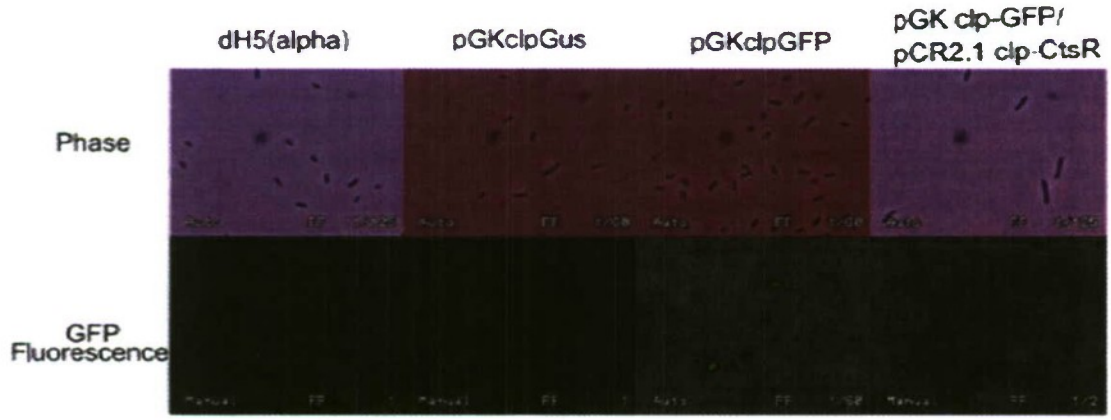


Figure 3: Activity of the Clp-promoter in *E. coli*. All plasmids are in dH5alpha cells. Exposure times are in the lower right corner of each image and are reported in seconds.

Approximately a third of the cells show GFP fluorescence. Previous experiments have shown plasmids with the pGK12 origin of replication to be highly unstable in *E. coli* (data not shown). Shorty after the removal of antibiotic selection pGK plasmids are lost from the *E. coli* population. Therefore, the plasmid instability may generates a population of cells with or without the pGKclp-GFP plasmid, and explain the low proportion of fluorescing cells. We also co-transformed the pCR2.1 containing the clp-CtsR cassette into cells containing pGKclp-GFP

(right most panel). We hypothesized we could express CtsR on a secondary plasmid and have the CtsR protein inhibit the Clp-promoter on our expression plasmid (in this case pGK-clp-GFP). The far right panel in Figure 3 shows highly reduced fluorescence when compared to pGKclp-GFP (middle right panel). The reduced levels of fluorescence indicate CstR may be working in trans on the expression plasmid and may reduce its promoter activity. It is possible we could express cassettes that are toxic in *E. coli* if the expression is truncated by providing CtsR in trans.

If the peptidase used to digest gluten was expressed from a plasmid, it would be lost shortly after antibiotic selection was removed from the bacteria. As a result, patients using the probiotic would need to be on an antibiotic regime for the probiotic to persistently express the peptidase. Additionally, FDA regulations do not grant food grade status to products that introduce antibiotic resistance to the environment. Therefore our probiotic must stably express the peptidase without antibiotic selection. To accomplish this goal the peptidase must be expressed from the genome of *L. reuteri* and not a plasmid. Integration into the genome occurs when a baeterial cell takes up a foreign DNA and incorporates the DNA into its genome. For these two events to happen in *L. reuteri* DNA has to pass through the thick cell wall, and then recombine with the bacterial gDNA. Integration of genes into bacteria has been aecomplished by [4] one-step genomic integrations [5] followed by screening. In general screens require a readily observable product and many man hours to produce positive genomie integrants integrants. *L. reuteri* has a low transformation frequency which further decreases the likelihood of obtaining positive genomic integrants. Therefore, the one step integration system does not provide a feasible method for genomic integration. Another method used to integrate genes into the genome of Gram positive bacteria uses two plasmids that are dependent on each other for proper replication [6].

Our system will be based on the plasmid pGK12, [7] which harbors two inducible antibiotie markers: chloramphenicol acetyl transferase (CAT) and erythromyein resistance marker C (ErmC) [8]. The pGK12 plasmid is able to replicate in a wide host range including both Gram positive and negative bacteria such as,*L. reuteri* and *E. coli*. The origin of replication in pGK12 is from the plasmid pWV01 [9], [10]. For pGK12 to replieate, the RepA, protein must be expressed to bind to the origin of replication on the plasmid. If RepA is not present, plasmids with the pWV0I origin will not replicate inside a eell. Our system takes advantage of the RepA dependenee by providing RepA in trans: a helper plasmid will express RepA while a seeond plasmid, the integration plasmid, will have the RepA dependent origin of replication, but will not express the RepA protein. Therefore, the integration plasmid will not propagate itself unless the helper plasmid supplies the RepA protein (Figure 4A). Built within the integration plasmid will be a genomic integration targeting region (LacA), the expression cassette (peptidase), and a selectable marker (ErmC). Using this system, we increase our chances of a genomic integration by selecting for the presence of both plasmids to generate a population of cells that contain both the helper and integration plasmids.

Genomic integration will then be promoted by removing antibiotic selection for the helper plasmid (chloramphenicol) and destabilizing RepA with thermal stress [6] (Figure 4B). If RepA is absent, the integration veetor will no longer be able to replicate. Therefore the integration plasmid faces one of two outcomes 1) integrate itself into the host's genome via homologous recombination or 2) be lost by the cells during cell division. The first option can be favored by continuing selection for the antibiotic marker on the integration vector (erythromycin). Therefore cells that do not integrate the plasmid will lose their antibiotic

resistance and will be selected against.

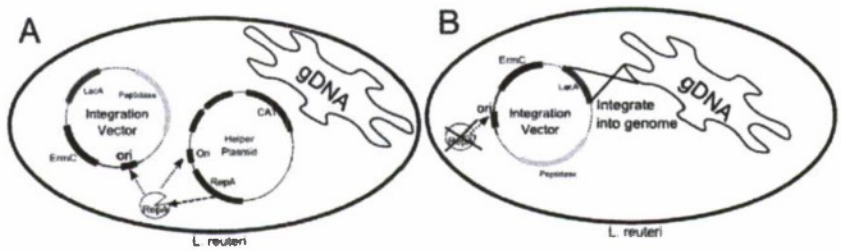


Figure 4: Dual plasmid integration schematic. (A) The helper plasmid provides RepA to both the integration plasmid in trans. By selecting with different antibiotics, both plasmids are maintained in the cells to generate a population cells with both plasmids. (B)Using thermal stress and removing the antibiotic used to select the helper plasmid genomic, integrants are selected.

The helper plasmid was designed to include the chloramphenicol marker with the erythromycin resistance marker deleted. During our preliminary work it was discovered that *E. coli* loses pGK derivative plasmids quickly when there is no antibiotic selection. To determine plasmid stability we cloned LacZ into the ErmC locus under the control of the *E. coli* LacP. We found that at 37°C, approximately 50% of the colonies showed signs of colony sectoring when plated on a LBX-gal/IPTG plate without antibiotic selection. Furthermore, when the temperature was increased to 42°C nearly all the colonies were white, indicating the plasmid was lost, after being plated on a LB- X-gal/IPTG plate (data not shown). Three plasmids that could serve the function of the helper plasmid were built: a plasmid with LacIPZ the same orientation as the erythromycin promoter (pGK-deltaErm-LacZsense or JL101), a second with LacIPZ in the opposite orientation as the erythromycin promoter (pGK-deltaErm-LacZantisense or JL102), and a third where the erythromycin ORF was removed and the ends of the plasmid were ligated together (pGK-deltaErm or JL103) (see Figure 5). We were able to construct all three versions of the helper plasmid. We chose to use the pGK-deltaErm lacZsense plasmid as our helper plasmid because it offers more methods of detection while in *E. coli* (via blue/white screening). Furthermore, the LacZ cassette provides more primer binding sites to be used during future verification processes.

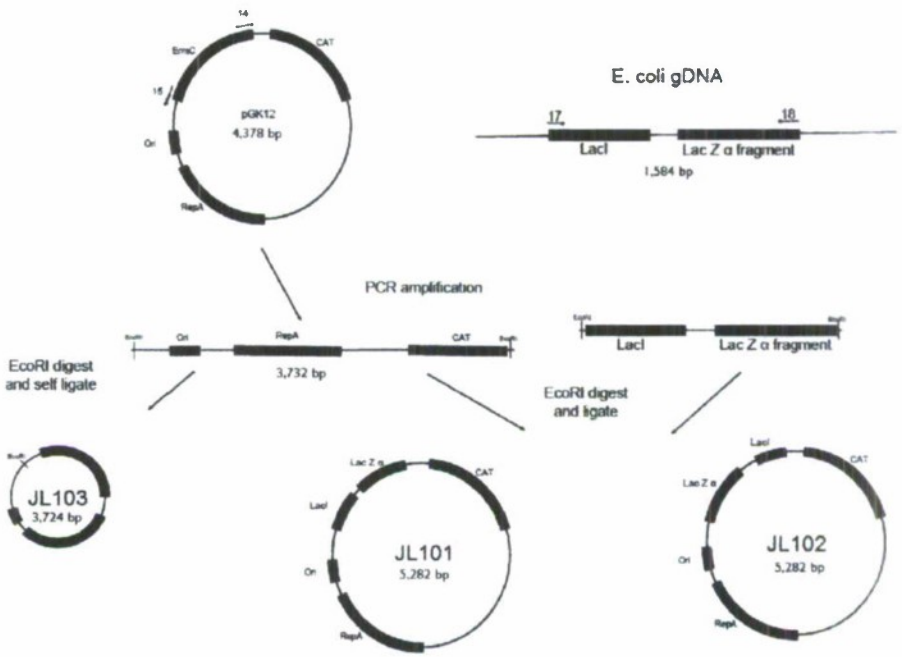


Figure 5: Construction of the three helper plasmids.

Construction of the integration plasmid provided us with some interesting challenges: 1) propagating a plasmid without a functional origin of replication, 2) the final integrated product had to be food grade, and 3) the plasmid must contain the site for integration into the *L. reuteri* genome. The integration plasmids must have an external source of the RepA protein to replicate. We found the helper plasmid was unable to provide adequate levels of RepA in *E. coli* to propagate both plasmids. To solve this issue, we obtained a RepA expressing strain of *E. coli* (EC101) that has been used in the construction of a similar integration vector [11]. Using this strain we were able to successfully construct and propagate the integration plasmid (see Figure 6).

To prepare *L. reuteri* for gene integration the pGK-deltaErm lacZsense helper plasmid was transformed into *L. reuteri* 10-23C. After inserting the helper plasmid, we successfully transformed pInt-clp-GFP into *L. reuteri*, with delta-Erm lacZsense providing RepA. A slight increase in GFP fluorescence was observed in *L. reuteri* containing pInt-clp-GFP (data not shown). The CtsR repressor protein has been found in many Gram positive bacteria including Lactobacillus [12]. *L. reuteri* may be expressing CtsR from its genome reducing the activity of the clp promoter in the integration plasmid. At this time, we do have a strain of *L. reuteri* that exhibits a continued resistance to erythromycin and sensitivity to chloramphenicol, but we have not verified genomic integration of the plasmid into the LacA gene. We are continuing our investigations to verify and increase the integration events. A greater proportion of the EC101 with JL120 (int-Clp-GFP) cells fluoresce than do the dH5alpha cells with JL115(pGk-Clp-GFP) (compare glowing cells between Figure 7 and Figure 3). As stated previously, the pGK plasmid

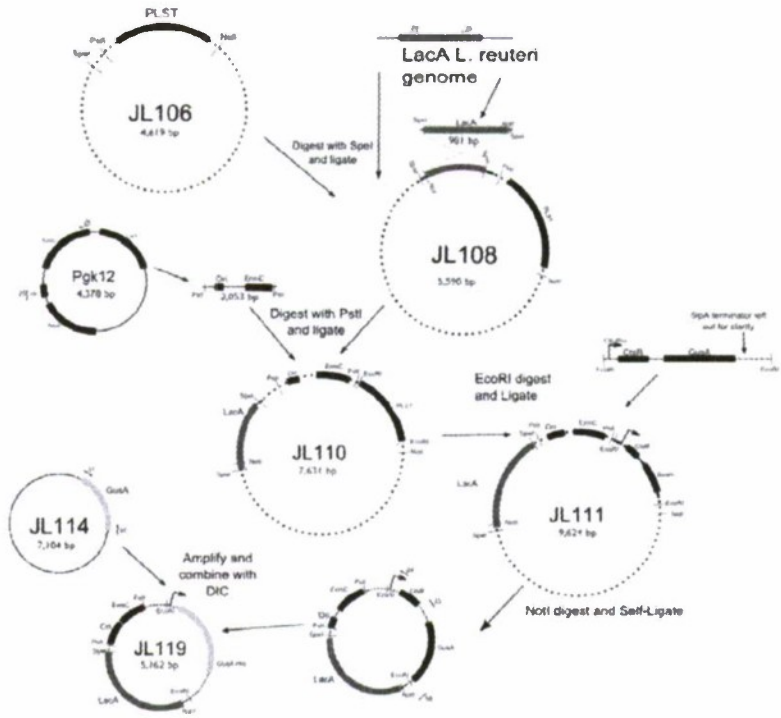


Figure 6: Construction of the integration plasmid. To build the integration plasmid with GFP, use the same primers and methods, however use pET30 GFP to combine with the integration vector backbone instead of JL114.

is not stable within *E. coli* ; therefore, it is expected that the population of cells may contain a subset of cells that have lost the pGK plasmid derivative.

The pInt-clp-gus plasmid (JL119) shows increased GusA activity compared to bacteria without the plasmid (Figure 8). In fact, the EC101 harboring JL119 has ~590X more GusA activity as EC101 alone. However, pGK-clp-Gus (JL116) has ~4X the activity as seen with JL119 in EC101 (compare Figure 8 and Figure 2). The differences in enzymatic units produced may be due to differences in the plasmid copy number allowed by each strain. Another possibility is that dH5alpha may have higher levels of transcription and/or translation than EC101.

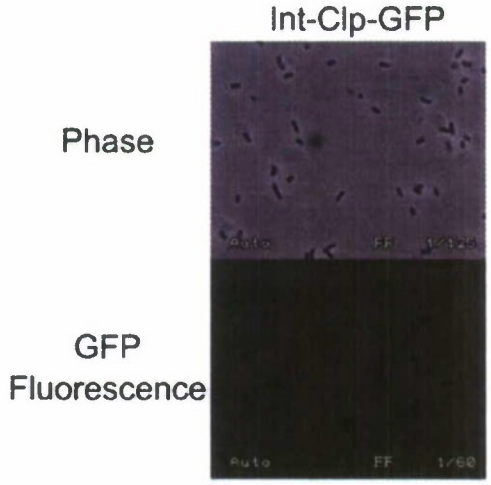


Figure 7: Verification of pInt-Clp-GFP (JL120) activity. The plasmid is harbored within EC101. The exposure time is in seconds and is reported in the lower right corner.

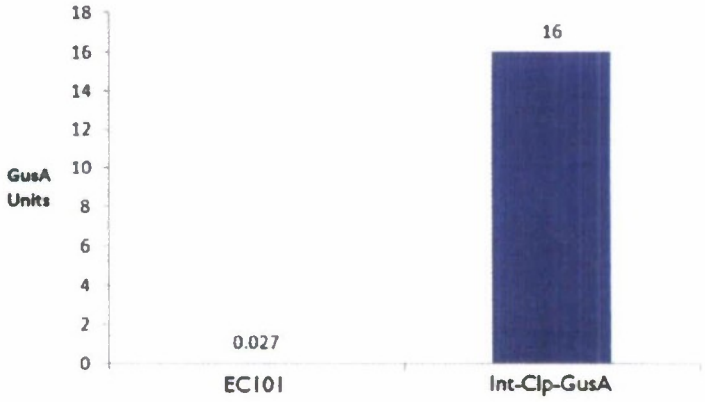


Figure 8: Verification of pInt-Clp-GusA (JL119) activity. The plasmid is harbored within EC101. Enzymatic units are defined as μM pNP/min.

For our antibiotic marker in the integration vector we chose to use the ErmC gene from pGK12. If our probiotic is to be food grade, however, our end product cannot harbor resistance to erythromycin [13]. Therefore we developed a system that allowed us to use the antibiotic resistance as a selectable maker during the construction of our bacteria and remove the antibiotic resistance after the plasmid integrated into the genome. To accomplish this, the Cre/Lox system was used.

Cre is an integrase protein that is found in the coliphage P1 [14]. Cre recognizes two 34 bp site (LoxP) and catalyzes the recombination and removal of the contents between the LoxP sites. This system provides an opportunity to utilize the convenience of antibiotic selection, as well as ultimately producing a food grade product. Lox sites were added using the primers created to amplify the ErmC antibiotic marker and origin of replication from pGK12. With standard LoxP sites, an active LoxP site still exists within the genome after the recombination event occurs. In order to eliminate this factor, modi_ed LoxP sites were used [15]. By modifying 5 nt on the 5' end of one LoxP (creating Lox66) site and the 3' end of the other LoxP site (creating Lox71), the LoxP site is inactivated (designated Lox72) after the recombination by the

Cre enzyme. The modification of the Lox site is considered to be an important step as it eliminates a possible genomic recombination hot spot in the genome of *L. reuteri*. After the Lox sites we also engineered PstI sites on the ends of the ErmC-Origin amplicon. These cut sites allowed for the insertion of the antibiotic marker and RepA dependent origin into the of the pCR4 plasmid.

For this study we needed an expression system for Cre recombinase that would propagate itself in *L. reuteri*. Using pGK12 as a backbone ensured the system would propagate itself in *L. reuteri*. It was also important that the expression of the Cre protein was inducible. The inducible ErmC regulatory region was chosen to control the expression of the Cre ORF. With our system cells that have undergone a genomic integration will be resistant to erythromycin. Therefore the CAT gene was required to select for the cre plasmid. The ErmC promoter was chosen to control Cre expression to maintain an unaltered chloramphenicol resistance marker (see Figure 9).

To ensure the Cre cassette did not interfere with the regulation of the ErmC gene, we inserted the Cre ORF after the 73rd amino acid encoded by ErmC [16]. The plasmid construction was confirmed by digesting it with NdeI and SacI. The digest should yield a banding pattern of 3560 bp and 1379 bp. The digestion of pGK-5'ErmCre with NdeI and SacI yielded three bands approximately 4.8, 3.6, and 1.3 kb in size (Figure 10). The largest is DNA fragment is assumed to be undigested pGK-5'ErmCre plasmid based on size and intensity. The smaller two fragments are close to the expected band sizes of pGK-5'ErmCre and indicate the plasmid is properly built.

To assess if functional Cre recombinase was produced from this plasmid, the plasmid was cotransformed with the pUG6 plasmid [17]. The pUG6 plasmid (Figure 10 middle lane) contains two antibiotic resistance markers, ampicillin and kanamycin. The kanamycin marker is flanked by LoxP sites which are the target of Cre recombinase. By co-transforming pUG6 with our pGK-5'ErmCre plasmid, we were able to determine if Cre was active via screens for the loss of kanamycin resistance followed by diagnostic digest verification. If active Cre recombinase was expressed from pGK 5'Erm-Cre the Cre enzyme would interact with pUG6 and recombine pUG6 and remove the kanamycin resistance marker. It was found that the kanamycin resistance marker was rapidly removed from pUG6 even before induction of Cre expression with erythromycin (Figure 10 far left lane). The pUG6 plasmid was isolated from Kan sensitive strains and digested with HindIII and XhoI. The digest of the parental pUG6 with these enzymes theoretically yields three bands 2484 bp, 1009 bp, and 516 bp. If the kan marker is removed from pUG6, a the plasmid is linearized (2502 bp) when digested with HindIII and XhoI. The digest of pUG6 before cotransformation with pGK-5'Erm-Cre revealed three bands of approximately 2.5 kb, 1 kb, and 0.5 kb (Figure 10, second to last lane). After cotransforming pUG6 with pGK-5'Erm-Cre digests with HindIII and XhoI reveal a single band at ~2.5 kb (Figure 10 last lane). The loss of the kanamycin marker (1 and 0.5 kb bands) from the digest suggest the Cre recombinase expressed from pGK-5'Erm-Cre is functional in *E. coli*.

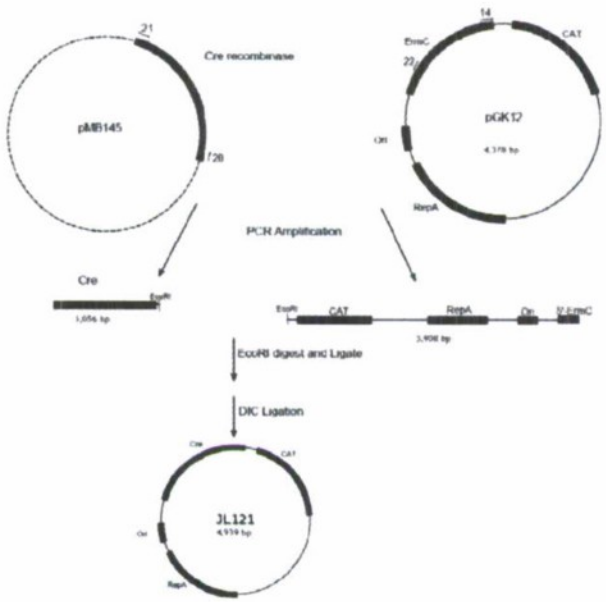


Figure 9: Construction of pGK-5'ErmCre.

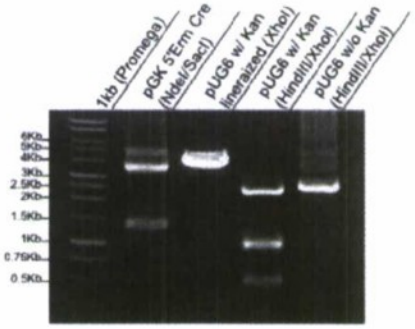


Figure 10: pGK-5'ErmCre verification and activity. Enzymes used to digest the plasmids are in parentheses. The pUG6 w/o kan should yield a single band of 2502 bp.

III. Discussion:

The end goal of the project is to create a probiotic strain of *L. reuteri* that will act as a persistent treatment for CD. The current treatment for CD is a life-long, gluten-free diet, which is difficult to accomplish with western diets. This project aimed to create a system to efficiently integrate genes into the *L. reuteri* genome. The work contained in this report has generated a body of necessary tools to integrate useful genes, such as a prolyl endopeptidase, into the *L. reuteri* genome and maintain food grade status.

Although the project was full of unforeseen complications, an integration system was developed that will be instrumental in creating a probiotic strain of bacteria to treat CD. This project has developed a plasmid-based system to integrate genes into *L. reuteri*. Furthermore, the system has the ability to utilize the convenience of antibiotic selection during the gene integration process, then remove the resistance marker via Cre recombinase. With the antibiotic resistance removed, the probiotic bacteria are more suitable as a food grade product. The

integration system also maintains a high degree of plasticity, as the integration site can be replaced by removing the LacA site and ligating into a new target sequence. Also, the gene to be to be integrated to can be changed by cutting out the current gene (GusA or GFP) with AscI and ligating in a new gene.

This plasmid system can be utilized to integrate genes into any bacterial species able to host plasmids with the pGK origin of replication. As a result, this system may be use to create other probiotic bacteria which could host enzymes such as lactase or beta-galactosidase. The proteins responsible for generating immunity could be integrated into the bacterial genome and the bacteria could be used as a delivery mechanism for vaccines development.

References:

[1] Hein J. Boot, Carin P. A. M. Kolen, Johannes M. Van Noort, and Peter H. Pouwels. S-layer protein of *Lactobacillus acidophilus* atcc 4356: Purification, expression in *Escherichia coli*, and nucleotide sequence of the corresponding gene. J Bacteriol, 175(19):6089-6096, 1993.

[2] Uwe B. Sleytr. l. Basic and applied s-layer research: An overview. FEMS Microbiol Rev, 20:5-12, 1997.

[3] Andrea McCraken, Mark S. Turner, Phil Gi_ard, Louise M. Hafner, and Peter Timms. Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species. Arch Microbiol, 173:383-389, 2000.

[4] Soo Jin Jang, Min Seok Ham, Jung Min Lee, Sung Kyun Chung, Hyong Joo Lee, Jeong Hwan Kim, Hae Choon Chang, Jong-Hoon Lee, and Dae Kyun Chung. New integration vector using a cellulase gene as a screening marker for *Lactobacillus*. FEMS Microbiol Lett, 224:191-195, 2003.

[5] Laurence Leloup, S. Dusko Ehrlich, Monique Zagorec, and Fran_coise Morel-Deville. Single crossover integration in the *Lactobacillus* sake chromosome and insertional inactivation of the pstl and lacL genes. Appl Environ Microb, 63(6):2117-2123, 1997.

[6] W.M. Russell and T.R. Klaenhammer. Efficient system for directed integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* chromosomes via homologus recombination. Appl Environ Microb, 67(9):4361-4364, 2001.

[7] Jan Kok, Jos M. B. M. Van Der Vossen, and Gerard Venema. Construction of plasmid cloning vectors for lactic streptococci which also replicate in *Bacillus subtilis* and *Escherichia coli*. Appl Environ Microb, 48(4):726-731, 1984.

[8] David Dubnau. Induction of ermC requires translation of the leader peptide. EMBO, 4(2):533-537, 1985.

[9] Kees J. Leenhouts, Jan Kok, and Gerard Venema. Lactococcal plasmid pWV01 as an integration vector for lactococci. Appl Environ Microb, 57(9):2562-2567, 1991.

[10] Jos F. M.L. Seegers, Adam C. Zhao, Wilfried J. J. Meijer, Saleem A. Khan, Gerard Venema, and Sierd Bron. Structural and functional analysis of the single-strand origin of the replication from the lactococcal plasmid pWV01. Mol Gen Genet, 249:43-50, 1995.

[11] Jean Law, Girbe Buist, Alfred Haandrikman, Jan Kok, Gerard Venema, and Kees Leenhouts. A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. J Bacteriol, 177(25):7011-7018, 1995.

[12] Isabelle Derr_e, Georges Rapoport, and Tarek Msadek. CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in Gram-positive bacteria. Mol. Microbiol., 31(1):117-131, 1999.

[13] Willem M. de Vos. Safe and sustainable systems for food-grade fermentations by genetically modified lactic acid bacteria. Int. Dairy J., 9:3-10, 1999.

[14] Nathalie Campo, Marie-Line Daveran-Mingot, Kees Leenhouts, Paul Ritzenthaler, and Pascal Le Bourgeois. Cre-LoxP recombination system for large genome rearrangements in *Lactococcus lactis*. Appl Environ Microb, 68(5):2359-2367, 2002.

[15] Jolanda M. Lambert, Roger S. Bongers, and Michiel Kleerebezem. Cre-lox -based system for multiple gene deletions and selectable-marker removal in *Lactobacillus plantarum*. Appl Environ Microb, 73(4):1126-1135, 2006.

[16] David Dubnau. Induction of ermC requires translation of the leader peptide. EMBO, 4(2):533-537, 1985.

[17] U. Güldener, S. Heck, T. Fiedler, J. Beinhauer, and Hegemann J.H. A new efficient gene disruption cassette for repeated use in budding yeast. Research, 24:2519-2524, 1996.

Extended shelf-life, performance-enhancing, dairy-based food products for the war fighter

Principal Investigator:

Rafael Jiménez-Flores Ph.D.
Dairy Products Technology Center
California Polytechnic State University
San Luis Obispo, CA

Summary:

The focus of this work is to develop a high-energy bar (aka. RTE or Ready-to-Eat) that could help maintain fitness and lean body mass in very active personnel whether it is a war-fighter or an athlete. There were three sources of funding for the project: C3RP, the CSU Agricultural Research Initiative, and the California Dairy Research Foundation.

We set out to develop an original kind of bar that takes advantage of the superior nutritional qualities of milk solids in their entirety and natural proportions of protein to fat ratios. With the water removed, texture was balanced with oligosaccharides and starch to form a bar. Low water activity was essential for obtaining very long shelf life. Once the formulation was developed, and the shelf life tested, an attempt was made to include probiotic microorganisms in the bar to ameliorate the cases of chronic diarrhea in some of the target individuals. However, military specifications require that the formulation and amount of live or viable organisms remain at the same levels for over three years. Our test indicated that under the current conditions and without further research this was not possible. Therefore, we abandoned this part of the objective.

The last objective was to test the effect of the bar on endurance, in a pilot study. The original plan was to have 40 individuals selected by high physical activity undergo a stress test - demanding physical activity during a short test time (3 days) - and blood analysis. The data acquired corresponds to 34 individuals. After following a crossover experimental design, where all the individuals were tested and ate either our experimental bar or one supplied by the military to soldiers, the results were positive overall. The caloric demands were met with either bar, and the blood tests indicated almost no difference in the individual's metabolism during the test period. The only difference that may be of importance, although statistically only suggestive of a trend, is the inflammatory response. The individuals responded better to inflammatory stimuli when eating the high protein and energy bar from our experiment.

Although no conclusive evidence can be put forward on the basis of this one test, it is safe to note that there was no difference in caloric delivery from our bar or the one used currently by the military. Our experimental bar, however, seems to be accomplishing a function that the purely glucose-based military bar is not. This function can be speculated to be due to the high protein content, and the proven function of muscle maintenance that protein has in high-demand activities of muscles. This conclusion is solely based on the different trend observed in blood C-Reactive Protein, which reflects response to inflammation.

Overall, this work seems promising, and if future funding is obtained, will be valuable in indicating the paths to follow. Some of the areas to explore in future research are the size and extent of endurance tests, and potential improvements to the bar, nutritionally and aesthetically.

The detailed description of the whole project is presented in the attached thesis resulting from this work. The PI and Co-PI are working currently on a manuscript to be presented to a peer-reviewed journal. Currently Mr. Jacob Heick is working at Hilmar Cheese, Hilmar CA.

DEVELOPMENT AND VALIDATION OF A LIGHT-WEIGHT, ENERGY-DENSE, READY-TO-EAT (RTE) BAR

A Thesis presented to the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Agriculture Specialization in Food Science

by

Jacob Wilhelm-Maria Heick

November 2010

COMMITTEE MEMBERSHIP

TITLE: DEVELOPMENT AND VALIDATION OF A LIGHT-WEIGHT, ENERGY-DENSE, READY-TO-EAT (RTE) BAR

AUTHOR: Jacob Wilhelm-Maria Heick

DATE SUBMITTED: November 2010

COMMITTEE CHAIR: Hany Khalil, Ph.D. Food Science and Nutrition

COMMITTEE MEMBER: Rafael Jimenez-Flores, Ph.D. Dairy Science

COMMITTEE MEMBER: Amy Lammert, Ph.D. Dairy Science

ABSTRACT

DEVELOPMENT AND VALIDATION OF A LIGHT-WEIGHT, ENERGY-DENSE, READY-TO-EAT (RTE) BAR

Jacob Wilhelm-Maria Heick

Providing additional calories in the form of an RTE bar to endurance athletes will increase performance and muscle re-synthesis, reduce muscle breakdown, and shorten recovery time. An RTE bar containing a blend of dairy proteins and carbohydrates will create a product with superior functionality, including bioactive and immunity enhancing properties from dairy derived ingredients. The protein will provide benefits in the form of easily digestible calories, essential amino acids and physical satiate.

A formulation was developed and optimized, resulting in a final product that meets the required nutritional profile: 400kcal, 25grams protein per 100 gram serving size. The desired physical characteristics were achieved through processing by both conventional baking and freeze drying. The latter method improves the stability and functionality of the RTE bar.

In order to meet the protein requirements of the RTE bar without compromising sensory properties, a unique protein source was developed. Using high concentrations of conventional protein sources like Whey Protein Concentrate (WPC) resulted in stale off-flavors and unappealing textures. Milk Protein Precipitate (MPP) was developed for this formulation. MPP is a curd-like ingredient created through the combined heat and acid precipitation of dairy proteins. MPP can be used effectively in high concentrations provides a subtle dairy flavor. MPP delivers a balance of casein and whey, similar to that found in milk.

The effectiveness of the RTE bar formulation as a post exercise recovery food was evaluated in a human studies experiment conducted on the Cal Poly campus. The human subjects study utilized 34 Cal Poly students in a single-blind cross-over design experiment. The study compared the effects of this high protein RTE bar against a calorically equal carbohydrate bar. The bars were administered after subjects completed the pre-assigned hikes on three consecutive days. Following the cross-over design, subjects received the alternate bar in the second period of the experiment. Several blood markers involved in metabolism and inflammation were measured before and after the two treatment periods. No blood marker showed a statistically significant difference between bars, but several trends were observed. Body weight and fat percent were also unaffected by bar composition.

Keywords: milk protein, RTE, freeze drying, exercise recovery, lean muscle loss.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Hany Khalil for his role as a mentor during my undergraduate studies, and his instrumental contribution in bringing me into the Master's program. I would also like to acknowledge Dr. Rafael Jimenez for taking an active role during my research, stimulating my interest and exposing me to a multitude of engaging projects. I would like to thank Dr. Amy Lammert for her expertise in product and nutrition development, and for helping me to clarify my goals. Additionally I would like to acknowledge Dr. Steve Davis and the Kinesiology Department for help in the success of the human subjects study.

"There are two ways to slide easily through life: to believe everything or to doubt everything. Both ways save us from thinking." Alfred Korzybski

LIST OF ACRONYMS AND TERMS

BCAA – Branched Chain Amino Acids PE - Polyethylene

BMP – Butter Milk Powder

Cortisol AM - Hydrocortisone

CPK – Creatine Phosphokinase

CRP – C-Reactive Protein

EAA – Essential Amino Acids

EPO – Erythropoietin

EVOH - Ethylene Vinyl Alcohol

FFM – Fat Free Mass

GRAS – General Recognized As Safe

HFCS – High Fructose Corn Syrup

MFGM - Milk Fat Globule Membrane

MRE – Meal Ready To Eat

MPP – Milk Protein Precipitate

PDCAAS - Protein Digestibility Corrected Amino Acid Score

PER - Protein Efficiency Ratio

WPC – Whey Protein Concentrate

RTE – Ready To Eat

RTEM – Ready To Eat meal

RDA – Recommended Dietary Allowance

SMP – Skim Milk Powder

1. INTRODUCTION

The objective of this project was to develop a novel ready-to-eat (RTE) bar to positively influence exercise recovery through muscle re-growth. Detrimental health effects of over-exertion during physical activity include loss of lean muscle mass and inflammation. These symptoms are found among both combat military personnel and endurance athletes. This research aims to address these negative health impacts of over-exertion through the development of a customized RTE bar, positively impact functional nutrition by tilting the daily energy balance and thereby reversing the negative effects of a caloric imbalance. In addition, the RTE bar was formulated to contain 50% of a Recommended Dietary Allowance (RDA) serving of complete protein.

The nutritional profile of the RTE bar is modeled on a small meal or snack which will provide a balanced blend of macro and micro nutrients. The 100 gram bar provides 400kcal and 25grams of protein in one serving. In order to meet these goals different protein sources were researched with a focus on the nutritional benefit of milk and dairy-derived ingredients. In order to provide protein in a RTE bar at the stated high concentrations without compromising its sensory quality, a novel protein source was developed. This protein ingredient was shown to be an effective method of delivering the needed macronutrients.

In order to validate the RTE formulation a human subject study was conducted on the Cal Poly campus. The study utilized the high protein RTE bar and a control carbohydrate bar. The experiment was designed to mimic combat soldier activity in mountainous terrain. Several different response variables were taken during and after the

physical activity. The goal was to measure the influence of the exercise on concentration, inflammation, body composition, and peak power. Statistical analysis revealed no significant difference between the control carbohydrate and high protein RTE within the blood markers analyzed. However, the blood markers indicated that the exercises did induce inflammation in the subjects.

2. LITERATURE REVIEW

Health Concerns

Nutritional Concerns in the Military

The military has long been concerned with the health and performance of its soldiers. This is witnessed by the special programs and institutions the military has dedicated to medical research. Examples of these include the Office of Naval Research, Army Research Institute of Environmental Medicine, US Army Natick Soldier Center, and the Institute of Medicine U.S. Army Medical Research and Nutrition Laboratory. The military's main nutritional concern is sustaining and enhancing the physical and mental performance of soldier's through diet (Anonymous 1994). During combat or regular training soldiers expend from (3109 kcal to 7131kcal) per day while consuming on average only 3000kcals per day (Tharion et al. 2005). There are many reasons for these extreme dietary deficiencies including: loss of appetite, lack of time and portability issues. If not properly addressed, this energy imbalance can lead to loss in lean muscle mass, and impaired physical and cognitive performance (Marriott 1995).

A loss of fat free mass (FFM) can also be interpreted as a loss in lean muscle. This was demonstrated on Italian soldiers in a body composition and physical exercise study (Malavolti et al. 2008). It was reported that the soldiers lost an average of 4.02kg +/- 1.42kg in FFM during the first three months of the experiment. This portion of the experiment contained increasingly strenuous exercises in the gym and in combat simulations, and was designed to represent ground combat with uncontrolled diets. The results of this experiment reflect the kind of situations that affect active duty soldiers

subjected to strenuous exercise. The effect of a negative energy balance has been investigated by numerous studies over the years. Observations include, but are not limited to: large losses in body mass, physical and mental fatigue, muscle soreness, weakness during recovery, impaired group function and loss of motivation (Montain and Young 2003). While the extent of the physical or mental impairment fluctuated between the different tests, the general consensus is that the performance of the soldiers was negatively affected. Each research group applied its own levels of nutrients (fasting through 3600kcal per day) for varying periods (5 days - 6 months) as well as using different tests to register the response (time to complete run versus hand grip) (Montain and Young 2003).

To avoid the effects of a negative energy balance, adequate calories must be consumed. To provide this nutrition in a way that is practical, as weight and space are constraints, small energy dense meals are seen as a solution. This being the case, fat would seem to be the ideal supplement for a military ration. Fat provides 9kcal/gram versus 4kcal/gram for carbohydrates or protein (Montain and Young 2003). However, supplementing the diet with additional calories from fat does not lead to significantly greater performance or even increased lipid metabolism (Hoyt et al. 1991). A study by Hoyt et al. 1991 indicated that while fat contained in the supplement provides additional calories, it is not readily metabolized and does not reverse the effects of underfeeding. The Committee on Optimization of Nutrient Composition of Military Rations for Short-Term, High-Stress Situations 2006 recommended a protein level of 1.2-1.5g per kg of body weight or 100-120g of protein per day. This is needed in order to maintain adequate serum levels while reducing net protein loss through sparing muscle protein breakdown.

This would be a large amount of protein to deliver in a single serving. So it was decided that a small energy dense ration, which could provide 25-30% of this amount would be ideal.

Sarcopenia

Sarcopenia is a muscle dilapidation disease that affects up to forty five percent of those over the age of sixty five (Cribb 2006). While the mode of action of the disease is not well understood, the effects are being increasingly investigated. Sarcopenia is diagnosed as a loss of lean muscle mass with a corresponding increase in body fat (Evans 2010). While sarcopenia refers directly to the loss of muscle in the elderly, treatment and research also investigate the young to better identify the causes of lean muscle loss. Lean muscle is the bulk tissue of the body that is responsible for movement and represents an energy source other than body fat or glycogen. Muscle is composed of protein and thus represents the body's storage of amino acids that are utilized not only in metabolism but also numerous other physiological processes. Loss of skeletal muscle results from an imbalance between muscle protein synthesis and degradation (Evans 2010). The cause of imbalance will vary for the specific demographic. For young persons it could be the result of over-exertion without adequate calories, while for older individuals it could be from a reduction in physical activity accompanying declining health and a poor diet. The mode of action for sarcopenia is the loss of the ability to convert available amino acids to glutamine, causing the body to increase anabolism of the liver to meet the demand of glutamine (Cribb 2006). For individuals consuming a hypocaloric diet, higher levels of dietary protein are required to reduce these detrimental effects (Lemon 1987).

Muscle protein is the most important body protein for endurance athletes (Tipton and Wolfe 2004). As the working site of movement, the muscle represents a major consumer of energy and the largest site of lipid oxidation and glucose metabolism (Cribb 2006). In order to reverse damage or stimulate muscle anabolism, net protein synthesis must exceed protein breakdown. In order to achieve this, a balance of macronutrient intake and resistance exercise must be introduced into the lifestyle (Cribb 2006). These two factors work synergistically providing a net gain in lean muscle mass that is greater than if each factor worked independently. During resistance exercise, the consumption of protein-rich dietary meals can be a major factor in maintaining or increaseing muscle mass (Phillips et al. 1998).

Cachexia

Cachexia is a complex metabolic condition that is associated with concurrent chronic diseases such as AIDS. Cachexia may affect any age group and is characterized by muscle wasting with or without body fat loss. Cachexia appears to selectively target actomyosin and thus heavily targets skeletal muscle (Evans et al. 2008). It appears cachexia can be reverted by therapics which reduce muscle inflammation and directly influence skeletal muscle growth in patients (Evans 2010).

Gastrointestinal Health

Diarrhea and other gastrointestinal problems havc been associated with both military personnel and endurance athletes. The causes of these problems have been linked to stress, nutrition, and the physiological effect of exercise on the digestive system. One hypothesis is that during periods of extreme exertion blood flow is directed toward the active muscles, thus temporarily dehydrating the gut and increasing its sensitivity to

stress (Ha and Zemel 2003). Another theory relates the problems to fructose consumption. The Committee on Optimization of Nutrient Composition of Military Rations for Short-Term 2006 recommends limiting the amount of fructose in rations to below 25g. Fructose at higher levels than this may contribute to gastrointestinal problems (Anonymous 2006). High fructose corn syrup (HFCS) is a common ingredient in food formulations, particularly in bar and supplement products frequently used by athletes. The large quantities of fructose consumed directly from these products might contribute to the GI problems seen in these individuals.

Dairy products have historically been associated with gut health and research has identified whey protein as one contributing factor. Whey proteins provide glycomacropeptides that are potentially utilized as prebiotics, which stimulate the growth of probiotics. Glycomacropeptides may also activate cholecystokinin which has many physiological effects such as the regulation of food intake and the release of pancreatic enzymes (Dockray 2009). Milk also contains prebiotics and is commonly associated with Lactic acid bacteria, the major family of probiotics. Probiotics are associated with promoting gastrointestinal and immune system health as well as the synthesis of vitamins (Hazen 2009). The benefits of probiotics result when viable organisms reach the small intestine in sufficient quantities thereby positively influencing the microflora of the small intestine. For this to occur, the organisms must be able to survive the initial processing of the food product and its eventual digestion in the mouth and stomach (Fernández et al. 2003).

Protein And Sport Nutrition

Protein Requirements for Athletes

All biological proteins are assembled from twenty amino acids. They can be combined in numerous sequences to form the complex and diverse array of proteins seen in living systems. The defining characteristic of a protein is its vital amino nitrogen group. In addition, proteins are the only macronutrient to contain nitrogen (Anonymous 2005). Proteins and amino acids are vitally important components of the body because they function as cell membranes, hormones, enzymes, vitamin precursors and nucleic acids. With its diverse functions and interdependence, dietary protein is essential for health, reproduction, growth, and maintaining of homeostasis. Protein is a necessary component of the human diet. Currently the recommended daily allowance (RDA) is set at 0.8 grams protein/kg of body weight for the healthy average adult or 50grams protein per day (Anonymous 2005).

There is a long-standing theory held by many athletes, coaches, supplement companies, and nutritionists that athletes need additional dietary protein. The logic being that proteins and amino acids are responsible for the synthesis and replacement of the structures associated with exercise and muscle building (Nemet and Eliakim 2007). Those who are more active would need more protein for fuel and rebuilding. Logic notwithstanding, there is little scientific proof that athletes require additional protein and some studies have even demonstrated that athletes require less protein (Phillips et al. 2007), however the assumption remains. The U.S and Canadian agencies responsible for the RDA have considered an increased consumption of protein of 1.2-1.4 grams protein/kg of body weight to be beneficial to endurance athletes (Tipton and Wolfe

1998). However, they have not stated that athletes actually require this increase. Interestingly enough, most athletes already consume an excessive amount of protein, more than the RDA and even more than the increased RDA. Diet surveys on strength and power training athletes have estimated consumption levels as high as 2-3 grams protein/kg of body weight with endurance athletes consuming approximately 1.2-1.5 grams protein/kg of body weight (Phillips et al. 2007).

Protein Balance

Energy balance is an important concept for individuals trying to modify body weight or composition. This refers to the difference between calories from food consumed (input) and the calories expended by physical activity (output). Tilting the balance either way will alter one's lean muscle mass; consuming more calories than one expends leads to a net gain in weight and consuming less leads to a net loss (Benardot and Thompson 1999). A study by Robert Demling and Leslie DeSanti in 2000 worked with overweight police officers and found that the subjects' average daily protein intake was below the RDA. This low intake was likely a factor in the lean mass and strength loss experienced by the individuals. In order to increase muscle mass or reverse lean muscle loss, the nutritional goal would be to tip the nitrogen balance to the positive side by consuming a net positive intake of amino acids (Phillips et al. 2007).

One of the most important indicators of protein utilization in the body is the Nitrogen Balance, which is defined as the minimal amount of protein ingested that will balance all nitrogen lost (Tipton and Wolfe 2004). The Nitrogen Balance is what was used to calculate the RDA for protein and amino acids. This method, however, is tailored to find the minimum intake level necessary to limit deficiency and not for optimal

athletic performance (Phillips et al. 2007). Additional protein can be rationalized because all ATP expended for bodily movement must come from energy stores (Lemon 1987). Muscle and skeletal protein represents a small "pool" of reserve energy that can be utilized during physical activity in addition to glycogen and lipid stores. Because this pool cannot be expanded, there is no other way to store the amino acids (Phillips et al. 2007).

There are restraints on the quantity of protein that can be consumed causing any excess protein to simply be stored as fat (Nemet and Eliakim 2007). Consuming surplus protein can also be a problem because of the nitrogen that is inherent in its structure. Nitrogen can be toxic and in excess will be converted into urea (Phillips et al. 2007). On the other hand, the body also reacts to high protein levels by increasing amino acid catabolism. During exercise the body's metabolism switches to a predominantly catabolic state. After exercise, during rest, the body shifts more towards anabolism (Tipton and Wolfe 1998).

Amino acids in muscle building

High quality proteins like eggs, dairy products and muscle proteins contain all of the twenty amino acids. This has been demonstrated with research showing that whole or skim milk consumption leads to a greater positive muscle protein balance and net amino acid uptake than soy based milks (Hartman et al. 2007). Amino acids are also the precursors to physiological compounds like creatine, epinephrine, and purine bases (Nemet and Eliakim 2007). Amino acids can provide ATP for muscle contraction through direct oxidation or the conversion to glucose via gluconeogenic pathways. In addition, the availability of the necessary amino acids is a requirement for muscle protein

synthesis (Levenhagen et al. 2002). Blood amino acid concentration has physiological signaling qualities like growth hormone, insulin, and insulin-like growth factor. This is dependent on the quality of the protein and the specific amino acids consumed (Nemet and Eliakim 2007). These amino acids function as regulatory molecules to stimulate muscle protein anabolism.

Decreasing blood amino acid concentrations has been shown to result in decreased muscle protein synthesis, while increasing the concentration has restored the synthesis rate (Tipton and Wolfe 1998). The physiological response changes depending on the protein type, differing even between two high quality proteins like whey and casein. Blood amino acid concentration is higher and adjusts more quickly after consuming whey protein, but anabolic response is greater with casein (Tipton and Wolfe 2004). There is a notable difference in the resulting blood amino acid concentration after ingesting intact proteins when compared to hydrolyzed amino acids (van Loon et al. 2000).

Muscle catabolism is an integral part of growth. As the muscle contracts, muscle fiber damage occurs. The muscle is the site where the metabolism responsible for this movement occurs, and as a result, the increase in amino acid oxidation likely occurs in these sites as well. During rest, muscle anabolism occurs and the previously damaged muscle is rebuilt. Muscle contraction leads to skeletal, structural, and membrane protein damage, proportional to the extent of the physical activity. The eventual muscle anabolism leads to a greater need for available amino acids for the synthesis of new proteins (Levenhagen et al. 2002). Supplementation of energy in the form of carbohydrates and/or fat can provide the energy necessary for the exercise and post-

exercise glycogen synthesis. Amino acids are, however, necessary for muscle protein re-synthesis. Muscle synthesis is influenced by the intramuscular availability of amino acids as well as blood flow. An increase in muscle synthesis increases the transport and delivery of amino acids to the muscles. The availability of these amino acids, either from the diet or resulting from muscle breakdown, may act as a signal for the eventual muscle synthesis (Tipton and Wolfe 2004).

Essential Amino Acids

Essential amino acids (EAA) are those that cannot be produced in sufficient amounts by the body, but are found in high quality protein sources (Nemet and Eliakim 2007). There are nine EAAs: Lysine, threonine, valine, isoleucine, leucine, methionine, phenylalanine, tryptophan and histidine. Two of these (lysine and threonine) cannot be synthesized by the body and therefore must be consumed in the diet (Bos et al. 2000). EAAs are even more critical to the synthesis of muscle protein and represent a limiting factor in protein synthesis (Cribb 2006). Animal studies have shown that muscle synthesis is reduced when EAAs are withdrawn from the diet. The EAA content of a protein is seen as the indicator of the quality of the protein source (Table 2-1), EAAs, which include branched chain amino acids, stimulate lean muscle protein synthesis (Ha and Zemel 2003).

Table 2-1: EAA's in selected protein sources adopted g/kg air-dry wt (Rutherfurd and Moughan 1998)

EAA ***BCAA***	Soy Protein Concentrate	Soy Protein Isolate	Whey Protein Concentrate	Milk Protein Isolate
Threonine	26.1	34.1	57.9	40.2
Valine	***33.9***	***44.4***	***49.1***	***61.1***
Methionine	10.0	12.6	21.8	29.1
Isoleucine	***31.5***	***43.1***	***52.2***	***49.5***
Leucine	***54.2***	***71.0***	***88.2***	***94.4***
Phenylalanine	36.0	48.1	29.5	48.4
Histidine	19.3	26.0	17.2	31.8
Lysine	42.6	60.3	72.8	75.9

Branch Chain Amino Acids

There are three branched chain amino acids (BCAA): leucine, isoleucine, and valine. They are a unique subset of the essential amino acids, accounting for 35% of the EAAs in muscles (Shimomura et al. 2004). BCAAs differ from other amino acids in that they are directly utilized in skeletal muscle as a source of energy (Nemet and Eliakim 2007), and show significant oxidation during exercise. This unique ability may increase the availability of carbohydrates and reduce the impact of muscle breakdown during exercise (Walzem et al. 2002). Endurance exercise shows an increase in the amino acids' oxidation, supporting the theory that BCAAs are of particular importance to endurance athletes (Phillips et al. 2007). BCAAs can also contribute to glucose production through the Cori cycle, due to their ability to form transaminase pyruvate in the muscle as an intermediate to alanine (Nemet and Eliakim 2007). BCAAs have also been shown to

reduce exercise-induced muscle damage and increase the synthesis rate (Shimomura et al. 2004).

Carbohydrates and Exercise Metabolism

Exercise Recovery

Some studies have shown that combining protein with carbohydrates in post-exercise meals can improve recovery time (Zawadzki et al. 1992; Levenhagen et al. 2002). Others report that the combination has no positive synergistic effect when compared to just carbohydrates alone (Jentjens et al. 2001). Much of the disagreement on the effects of combining protein with carbohydrates for improved recovery is due to the quantity of protein or carbohydrates provided and the style and extent of the exercise, as well as the method of measurement. In a study that resulted in a zero net gain in muscle glycogen synthesis, blood insulin levels increased when protein and carbohydrates were administered (Jentjens et al. 2001). Yet other studies found a net gain in muscle synthesis if protein was included in the supplements (Zawadzki et al. 1992; Kimball et al. 2002). This could indicate that insulin level may not be the rate-limiting factor in muscle glycogen synthesis but are affected by protein consumption (Jentjens et al. 2001). In this case muscle synthesis will be achieved as long as adequate carbohydrates are provided. Even without additional protein intake, nitrogen balance may be restored with only the consumption of carbohydrates (Phillips et al. 2007). A protein-sparing effect occurs if sufficient carbohydrates are available, and protein oxidation will be ignored or reduced. In low carbohydrate diets, protein would be redirected for utilization as fuel instead of its anabolic use (Benardot and Thompson 1999). As stated before, the major energy sources

during exercise are lipids and carbohydrates (glycogen), while protcin or amino acids account for only 3-6% of the ATP needed during exercise.

Endurance athletes generally "carbo load" consuming large quantities of carbohydrates before periods of extreme exercise. Consuming 7-10 grams of carbohydrates per kg of body weight is recommended for those participating in marathon events (Nisevish 2008).This "loading" leads to larger concentrations of available carbohydrates which are stored as glyeogen, and if protein is also consumed, it improves the net protein balanee and reduees protein utilization (Gibala 2007). It is beeoming understood that amino acids play an important role in the intermediatc steps of the TCA cycle (Gibala 2006). There is the potential for athletes to reduce storcd fat and alter their body composition through consuming a low calorie diet, skewed towards higher protein consumption (Phillips et al. 2007). Carbohydrates, if consumed in excess without adequate aetivity, are partieularly prone to be stored as fat, and developing excess fat can lead to additional health problems (Demling and DeSanti 2000).

Glycogen

Glycogcn is thc body's natural energy storage form for carbohydratcs. It is the first and major energy source utilized during physical activity; however it is finite in quantity and must frequently be replenishcd. Glycogen represents a relatively small store of energy, approximately 1500-2500kcal when saturated; this is due to the low energy density of carbohydrates (Hoyt ct al. 1991). Aftcr exercise, to rcstorc glyeogen levels to pre-exercise levels, an estimated supplement eontaining between 1-1.2 grams carbohydrate per kg of body weight is required (Phillips et al. 2007). Most evidence suggest that if adequate carbohydrates are consumed (>1.2 gram/kg body/ hour) thc

benefits of additional protein are negated. However, when protein is ingested with carbohydrates, glycogen synthesis rate will increase if the quantity of available carbohydrates is low (Gibala 2007).

Effect of Timing

Consensus among athletes, trainers, and nutritionists is that consuming supplemental protein and carbohydrate at the end of exercise provides a better anabolic environment (Nemet and Eliakim 2007). One study found supplementing within one hour of exercising promoted greater gains in lean muscle mass compared to either soy or carbohydrate controls (Hartman et al. 2007). Immediate post-exercise supplementation could benefit the endurance athlete in repair and synthesis of muscle protein and the reloading of glycogen (Gibala 2007). Protein consumption, whether consumed alone or in conjunction with carbohydrates, will be a major determinant in strength or muscle mass gains (Phillips et al. 2007).

Protein Supplementation Case Studies

Under conditions of weight loss, diets that contain more protein have been shown to lead to significantly less lean-muscle loss compared to diets high in carbohydrates (Layman et al. 2005). One study on protein utilization found that subjects who were deficient in initial glycogen stores before endurance exercise utilized more protein as a percentage of total energy expended (Lemon and Mullin 1980). In a study comparing high and low protein diets with two exercise treatments, those who consumed the high protein diet lost more body fat without disrupting HDL cholesterol levels (Layman et al. 2005). This seems to indicate that increasing protein in the diet could potentially improve the body composition of subjects during exercise. Another study was conducted

comparing a control containing only carbohydrates and fat to a treatment that also incorporated protein. The results showed a twenty percent greater quantity of circulating amino acids (lysine, valine et al.) in the blood after exercise with the protein supplement (Levenhagen et al. 2002). The presence of the protein in the supplement seemed to reverse the catabolism that was seen with the control supplements. The hypotheses is that the limiting factor in muscle protein synthesis is not the overall energy consumed in the diet, but the amino acid concentration in the body as a result of the food consumed.

Dairy as Functional Nutrition

Milk Overview

Milk provides all the nutrients necessary for the growth and development of the maturing mammal. Milks supply macronutrients as well as immunity compounds and micronutrients (Walzem et al. 2002). The composition of milk varies depending on the species, stage of lactation, season, and a variety of other factors. Milk contains a combination of two major protein groups, wheys and caseins, each has specific functional and nutraceutical properties.

Bovine milk contains on average 3.4% protein, which is primarily 80% casein and 20% whey (Spreer 1998). Casein and whey proteins behave differently during processing and digestion. Casein will coagulate in the stomach forming clots that are harder for enzymes to proteolysie; however once in the small intestine they are absorbed quite readily. Whey proteins do not coagulate on contact with the stomach's acid and are thus transferred quickly to the small intestine where they slowly become absorbed over a much greater length of time (Walzem et al. 2002).

Milk has been shown to be an effective functional ingredient for promoting positive health and athletic performance. A study comparing a carbohydrate and soy supplement to a skim milk one in a controlled laboratory weightlifting experiment showed that skim milk increased the type I and type II muscle fiber areas greater than the soy and carbohydrate products. Skim milk also increased the fat and bone free mass above that of the other treatments, and led to a greater reduction in fat mass (Hartman et al. 2007). Milk's protein profile is unique in containing all essential amino acids and high concentrations of BCAAs. Casein and whey have separate profiles, but even independently they score high compared to other protein sources (Table 2-2).

Table 2-2 - BCAA composition of selected proteins, adopted form (van Loon et al. 2000).

% by Wt	Casein	Whey	Pea	Wheat
L-Isoleucine	5.8	5.1	2.4	2.6
L-Leucine	10.1	8.7	5.1	5.6
L-Valine	7.4	4.5	2.7	3.0

Casein

Casein accounts for 80% of the protein in bovine milk. It is the fraction that is responsible for creating cheese because it is hydrolyzed by chymosin and its solubility is influenced by pH. Casein proteins have been shown to contain various peptides that have bioactive properties, and these peptides seem to require proteolysis of the main casein forms in order to be released (Shag 2000); (Walzem et al. 2002). In the study of overweight police officers by Robert Demling and Leslie DeSanti in 2000, after twelve weeks, lean muscle gains were doubled and fat loss was fifty percent greater in the group which was fed a casein supplement compared to the whey group. Casein has four major subgroups (α_{s1}, α_{s2}, β, κ), each has multiple bioactive peptides with different abilities and

strengths. There appear to be several main substrates that are affected by these peptides (Shag 2000). Opioids, known as casomorphins, have properties similar to that of opiates and have been seen to increase gastrointestinal transit time among other physiological effects. Immunomodulating peptides have been shown to affect T-cells and macrophage activity. In addition, antihypertensive, anticariogenic, and antithrombotic properties have been observed. Hydrolysate components of casein have been shown to decrease amino acid oxidation and net protein breakdown, leading to improved nitrogen retention compared to other supplements available commercially (Demling and DeSanti 2000). Unlike whole or native protein, hydrolysates have also been shown not to stimulate the release of the hormone cortisol, which has lipogenic and catabolic properties.

Whey

Whey proteins represent the minor portion of total milk protein, accounting for approximately 20% of the total. Whey protein exists at the same concentration in human milk as in cow milk. However, human milk contains no β-lactoglobulin and cow milk has a much lower level of lactoferrin than human milk (Bos et al. 2000). Whey protein has also been identified as a possible source for bioactive peptides. After ingestion, whey protein leads to a very rapid oxidation and whole body protein synthesis. Casein, on the other hand, leads to whole body proteolysis suppression (Hartman et al. 2007). Whey is composed of several protein fractions including β-lactoglobulin, α-lactalbumin, proteose-peptones, and blood proteins (Walzem et al. 2002). The majority of these peptides seem to have influence on the immune and digestive systems such as chelating, antimicrobial and antioxidant activity. In addition immunoglobulins have potential anticancer and antitumor effects (Shag 2000). Whey is

also said to have hypocholesterolemic properties which might actually "balance" out the possible negative health effects of the saturated fat naturally occurring in milk (Walzem et al. 2002). Whey protein contains a high proportion of sulfur-containing amino acids (cycteine, methionine), which are said to contribute to the higher protein efficiency ratio (PER) of whey. Whey may also lead to the sparing of tissue proteins ordinarily used in response to immune challenges (Walzem et al. 2002). Whey proteins contain high amounts of EAA and BCAA which are generally lacking in plant and other protein sources (Table 2-1). As a byproduct of cheese production the whey stream is seen as a rich source of BCAA, equaling at least 26% of the total amino acids present (Bos et al. 2000). The amino acid composition of whey is said to be relatively similar to that of skeletal muscle, making whey a good source of amino acids during muscle re-synthesis.

Minor Components

Components beyond the macronutrients of milk, such as minerals and carbohydrates, have also gained recent attention. Lactose has the ability to form oligosaccharides which have both specific and broad prebiotic properties. These oligosaccharides can be labeled Generally Recognized As Safe (GRAS) for use in products desiring enhanced probiotic effects. Lactose may also influence the absorption of calcium, which in turn is said to have a role in regulating blood pressure. Milk enzymes do not appear in finished products as they are deactivated during pasteurization. However there is emerging research on particular enzymes like lactoperoxidase which is used as a preservative in some products (Walzem et al. 2002). Lactoferrin, another milk enzyme of interest, has iron-chelating, cation transport, and anti-infectious properties. Lactoferricin, a form of lactoferrin, also has bactericidal activities (Bos et al. 2000).

Buttermilk, the by-product and liquid phase of butter manufacturing, has been seen as a source of potential bioactive components. Sphingomyelin and phospholipid have been demonstrated to have anticancer properties and are concentrated in the buttermilk fraction (Walzem et al. 2002).

Nutritional Bar Development

Target Formulation Constraints

The goal of developing this RTE bar is to supply high energy and a designated percentage of the RDA of calories, as well as a combination of all of the macronutrients. Incorporating dairy protein into a RTE bar is a preferred method of directly reversing the negative effects of lean muscle loss through the diet. Development of the RTE bar focused on delivering the maximum nutritional functionality to the end-user.

The form and source of protein in a food product is of great importance, in that it must appeal to the target market, meet nutritional objectives and function appropriately in the formula (Hazen 2008). The quality of protein consumed is very important for maximizing the anabolism of muscle protein. The high-quality proteins in milk, dairy products, eggs, and muscle meats are ideal (Phillips et al. 2007). Another measure of protein quality, without measuring the concentration of individual amino acids is the Protein Digestibility Corrected Amino Acid Score (PDCAAS) (Hazen 2008) (Table 2-3). From this perspective, dairy protein, in particular whey protein, appears to be best suited for a protein bar product. In addition to their nutritional properties, whey proteins have critical functional properties that make them practical in bar formulations. They retain moisture, have a mild flavor, contribute to extended shelf life, lead to reduced

cooking/baking losses, and can be used to replace carbohydrates (Runestad 2004). The quantity of high-quality protein is also important; 20-25g appears to be the upper limit to stimulate muscle protein synthesis, and would be an ideal maximum in a single-serving product. Above this level amino acid oxidation and urea formation become more prevalent (Phillips et al. 2007), diminishing the effectiveness of adding protein. Currently, the military serves Meal Ready to Eat (MRE) to soldiers in the field. These are lightweight and contain several separate packages that represent a full meal when eaten together. One concern with the use of MRE's is that the macronutrients are not evenly distributed in the different components. This allows the soldiers to "field strip" or selectively eat portions of the ration and therefore not gain all the intended nutrients from the meal (Anonymous 2006). The goal of an optimal nutrition bar is to be a high energy snack or small meal that provides a designated percentage of the calories and all macronutrients needed by an individual in a day.

Table 2-3: Protein Comparisons by Source: PDCAAS (Hazen 2008), Biological values (Runestad 2004)

Source	PDCAAS Value	Biological Value
Whey	1.0	104
Egg	1.0	100
Soy	1.0	74
Pea	0.86	-
Hemp	0.46	-
Wheat	-	54

The Committee on Optimization of Nutrient Composition of Military Rations for Short-Term, High-Stress Situations 2006 lists the following recommendations for the development of a ration:

- Protein and carbohydrates are the priority
- Fat is important for palatability and absorption of fat soluble vitamins
- Weight and volume restriction of: 0.12 cubic feet, 1.36kg
- Shelf life of 2-3 years
- Individual portions that can be easily distributed in backpacks
- Palatability is a primary concern
- Variety of familiar flavors, colors and textures
- Potential for either sweet or savory formulations.

Nutrition bars are among the easiest products to fortify. They have an easy dry mixing stage, low thermal processing (if any), and they generally utilize opaque laminate packaging (Hazen 2009). The main challenge with the formulation of nutrition bars is the drying and hardening that occurs during storage and throughout the shelf life (Runestad 2004). This problem is compounded by the long shelf life and humidity standards set by the military, as well as the moisture and water activity in the bar. Water activity (aw) is an important property of foods that will help dictate food safety, shelf life and textural parameters. The water in the product migrates over time to the protein and the dry ingredients, which will alter the intended texture (Hazen 2010). Higher aw will result in a softer bar, however there is a limit to this as shelf-stable bars need to be at an aw level of less than 0.65 for food safety reasons (Hazen 2010). A consideration that should be taken into account with an RTE bar is that protein metabolism requires more

water than either lipids or carbohydrates (Lemon 1987). This could be problematic for a product that contains high protein levels but low moisture content. This makes the option of hydrating the RTE bar a more effective means of delivering the desired product nutrients.

Ingredients

Dietary Fiber

Dietary fiber is an interesting ingredient from a formulation perspective. It is desired because it has minimal nutritive properties. Dietary fiber under the current definition pertains to fibers that are indigestible but can be utilized as a prebiotic fiber in the small intestine. Fiber can also be an important component in a bar formulation, providing necessary nutrition and digestive functionality. Fibersol-2 (see appendix page 131) is a commercial ingredient which is labeled as a resistant starch, it provides dietary fiber and helps with texture throughout the product shelf life (Runestad 2004).

Delactose Permeate

Delactose permeate (see appendix page 130) is a novel dairy ingredient developed using the waste stream of WPC concentration. It has a high mineral concentration ~ 30% ash, with high calcium content at 3.7%. Delactose permeate contains oligosaccardies and many micronutrient ingredients. Therefore it could be used to boost calcium and vitamin content in a particular formulation targeted to women or the elderly. There is also recent research that points to success in using delactose permeate as a salt replacement in bakery products. This would help the product become more attractive to individuals struggling with hypertension or individuals simply looking for low sodium foods.

Flavor

From the onset of this project there has been the concept of developing a savory form of the RTE bar. The majority of bars that are on the market today are sweet (Hazen 2009). While there is interest in the concept of a savory bar, there is little indication that the market would accept it. This is likely a result of the current standard formulations used by most producers and expected by consumers. Currently available bars contain significant amounts of HFCS to act as a binder, or contain carbohydrates as a major ingredient, and often use bitter tasting protein blends. These qualities lend themselves more to a sweet formulation than a savory one. However, considering the specification of the RTE bar as high protein with a mild dairy flavor, a savory option might be achievable.

Bar Processing

Freeze Drying

Freeze drying preserves food by removing free and bound water. It has many commercial and industrial applications and is used in the processing of high value and biologically active products (Oetjen and Hasely 2004). Freeze drying may be the processing method of choice for the RTE bar because of its low processing temperature. Freeze dried products are easy to rehydrate and still retain biological activity. Freeze drying relies on the properties of water sublimation, that is the bypassing of the liquid phase in the transition from a solid state to a gaseous state, to remove the water from the product with minimal heat input. Sublimation occurs when the partial pressure of the environment is below that of the product so the frozen water must evaporate to create equilibrium. However, the water vapor is constantly being removed by the condensation

coil which maintains sublimation. A basic freeze dryer contains four parts (Jennings 1999):

- A chamber that can be both temperature and pressure controlled
- Vacuum pump which lowers the pressure and removes some gases
- Heating plates that provide heat to increase the sublimation rate
- And a refrigerated coil that removes the sublimated vapor from the chamber's environment by creating a temperature gradient.

To freeze dry, first the product must be completely frozen to a very low temperature. This is generally done in the blast freezer at a setting of -14°F. The sublimation of the frozen water occurs after the samples are placed in the chamber and the heating plates and vacuum are set to the desired levels. Without being placed in the pressure-controlled atmosphere, the ice would simply melt. The heating plates provide enough minimal radiant heat to supply the latent heat of sublimation (1075 BTU/LB or 2495.08 KJ/KG). Secondary drying occurs after all free water has been removed; the product will appear dry but still contains bound water. This water can also be removed with resulting theoretical moisture content between 1-5%.

3. BAR FORMULATION AND MANUFACTURING

Introduction and Experimental Logic

Stable energy balance and physical health are critical factors for certain high risk groups like soldiers and endurance athletes. Such individuals depend on their strength and stamina to do their jobs and often have restricted carrying capacity, limited time to eat, and unbalanced meals. Stress has a profound effect on the human body which is compounded by a poor diet. If appropriate safeguards are not taken there is the potential for lean muscle loss and long-lasting physical and psychological deterioration. A reduced calorie diet, especially one that lacks high quality protein, is one of the main causes of this physiological stress. There is the possibility of mitigating this risk with a well-balanced high protein (RTE) nutrient bar. The specifications (Table 3-1) for such a product are designed to provide the required high energy nutrients within the time and space constraints.

Table 3-1: RTE bar physical and nutrition specifications

Ready To Eat (RTE) Bar Specifications	
Weight: 100g	Intermediate moisture range
Total Calories: 400kcal per serving	Nutrient and product stability
Protein Content: 25grams protein	Contains a majority of dairy ingredients

To meet the above specifications, the ingredients must be carefully selected, particularly the protein source. Based on the current scientific understanding of protein metabolism, providing a high quality protein (Table 2-3) would be the most effective method for achieving this. This would ensure that any and all amino acids would be

available in adequate amounts to support the anabolism of muscle. Dairy protein was considered the best option, for the RTE development and more specifically, whey protein due to its extensive use in similar commercial products, and the relatively high concentration of BCAA (Table 2-2) involved in muscle contraction. However, whey protein is just a fraction of milk, lacking casein and dairy lipids. Considering this we undertook to develop a more suitable protein source from milk to provide the needed protein profile, which could offer superior nutritional benefits to whey protein alone.

In addition to the protein type, the specific processing method for such a bar is important. The military has stringent guidelines for their current rations which include a 2-3 year shelf life, nutrient stability, and small compact size (Anonymous 1994). In order to meet these specifications, freeze drying was chosen as the processing method for the RTE bar. More common drying processes, such as baking, vacuum, and air drying, were also investigated. Several micronutrients and probiotics were then considered as possible additives to the basic RTE formulation. Probiotics are of growing interest to the food and dairy industries (Stanton 2001), consumers are becoming aware of probiotics and demanding them in foods they commonly eat. Processors are also recognizing the potential of probiotics to increase their market share and provide a novel method of delivering targeted nutraceutical properties (Ouwehand 1998). Probiotics would be an ideal addition to a nutritional bar formulation; the challenge is ensuring the survival of the active probiotic organisms which are limited by the relative instability of organisms during processing and storage.

Materials and Methods

List of Ingredients

- Hilmar 8200 whey protein concentrate (WPC): protein source and filler. Typical: 82.5%, Specification: 80.0% min Protein.
- Milk Protein Precipitate (MPP): Ricotta-like cheese manufactured in-house, protein source, binder and filler. Protein 22-27%, Moisture 35-40%.
- Buttermilk Powder (BMP) Dairy America: protein and micronutrient source. Protein 32 - 34.5%, Fat 5.5 - 6.0%, Lactose 49.0 – 50.5%.
- Non Fat Dry Milk (SMP): filler and control ingredient to BMP. Protein 36%, 50% lactose.
- Bread flour (high gluten): filler and carbohydrate source. 73% carbohydrate, 12% protein.
- Sucrose: Filler and sweetener.
- Non-Iodized Salt: Flavor enhancer, water activity control.
- Cornstarch: Binder, moisture retention.
- Sweet Cream Butter: Lipid source, calorically dense. 80% Fat.
- Experimental test ingredients: puffed millet, high fructose corn syrup (HFCS), corn syrup, de-lactose whey permeate, Fibersol-2, shortening, water, flavors.

Formulation Parameters

The formulation of the RTE bar was specifically tailored to allow easy altering of ingredients and/or processing methods. The first priority for the different trials was to create a RTE bar that the stated specifications (Table 3-1). This involoved the development of a novel protein source (Chapter 4 - Milk Protein Precipitate (MPP), removal of HFCS from the formulation, baking and freeze drying processing parameters, and the inclusion of probiotics. The formulations in Table 3-4 are the final recommended formulations. These represent two different approaches to meeting the nutritional and physical specifications of the RTE bar. Both meet or exceed the nutritional specification, as can be seen in their mock nutrition panels (Figure 3-17).

Preparation of RTE Bar

1. First the ingredients are weighed and staged, ready for later use.

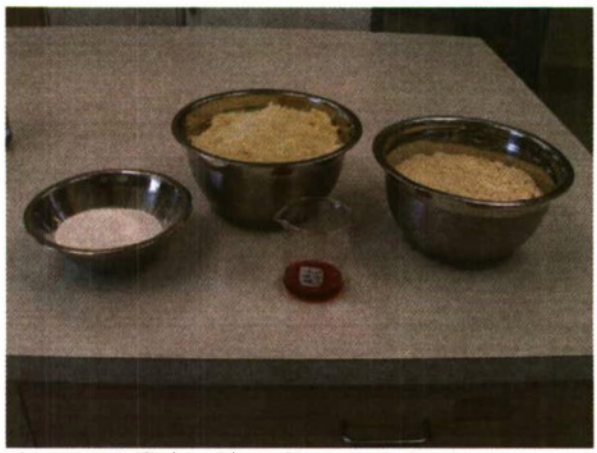


Figure 3-1: Weighed ingredients ready for use.

2. The flavor ingredient is mixed with the salt, sugar, and sucralose until it appears uniform.

Figure 3-2: Flavor ingredient mixing with sugar, salt, and sucralose

3. The dry ingredients: flour, WPC, cornstarch, are slowly added. Mixing continues until flavor clumps disappear.

Figure 3-3: Dry ingredient addition

4. The "wet" MPP protein is added, which functions as a binder and the major protein source. This mixing step continues for approximately seven minutes or until the dough is formed.

Figure 3-4: MPP protein is mixed into dry ingredients

5. Moisture from the MPP slowly migrates out of the curd structure, causing the mixture to clump. The product slowly becomes a cohesive dough.

Figure 3-5: RTE material after dough is formed

6. The dough is removed from the mixer and placed on a sheet of parchment paper.

Figure 3-6: Dough removed from the mixing apparatuses

7. The fresh dough is spread around the mold to ensure even distribution, then an additional sheet of parchment paper is placed on top to prevent sticking.

Figure 3-7: Fresh Dough being rolled using custom mold

8. The dough is rolled/sheeted into the precise measurement 11.5" x 14.5" x 0.5", using a custom made mold.

Figure 3-8: Dough rolled out, before baking

9. The formed dough is then placed on a baking tray with parchment paper and baked for 1.16 hours at 250°F according to the baking schedule.

Figure 3-9: Bars after baking and cutting

10. The bars are then cut to specific dimensions to ensure a proper weight and serving size (75g) and 1.5" X 0.75" x 0.5".

Processing Methods

The specifications of the RTE bar (Table 3-1) are greatly determined by the processing of the bar. The main purpose of processing the bar is to reduce the final

moisture and water activity in the product. Reducing the moisture increases the energy density of the product as moisture content adds weight but not calories. In addition, drying the bar reduces the water activity (aw) of the bar. This limits the growth of spoilage organisms by reducing the water needed for their growth. Processing also affects the palatability of the RTE bar, which is highly dependent on the texture and mouthful. Several different processing methods were investigated (Table 3-2) and adapted to meet current and specific project goals.

Table 3-2: Summary of RTE processing methods

Method	Time	Temperature	Pressure	Moisture
Freeze Drying	2-8 hours	Product:30°C Plate:100°C	0.35mmbar	1.5%
Vacuum Oven	4-16 hours	25-55°C	25"Hg	>15%
Forced Air Drying	2-10 hours	150°F	Atmospheric	~28%
Convection Baking	1.16 hours	250°F	Atmospheric	~25%

Freeze Drying

Several heating parameters were investigated to achieve the best product using the freeze drier. While some batch to batch variability still existed, the majority of this was removed when the product was rolled to an even thickness and perforated using a roller docker. The holes created by the roller docker acted as channels for the water vapor to leave the product. Without holes the product would balloon up in some sections, creating hollow cavities which would fracture the product. The maximum temperature setting can be adjusted for both the plate temperature and the product temperature. Plate temperature

refers to the heating element supplying the radiant heat. Product temperature refers to the temperature at the center of the product which is set to shut off the plate heating when the parameter is reached.

Figure 3-10: Freeze drying chamber during processing

Freeze Drying Procedure

1. Dough is rolled flat to uniform thickness (0.5") (Figure 3-8).
2. Holes made in the product using a roller docker.
3. The bars are cut to desired size (Figure 3-9) and placed in blast freezer (-14°F) overnight.
4. Bars are placed in freeze dryer.

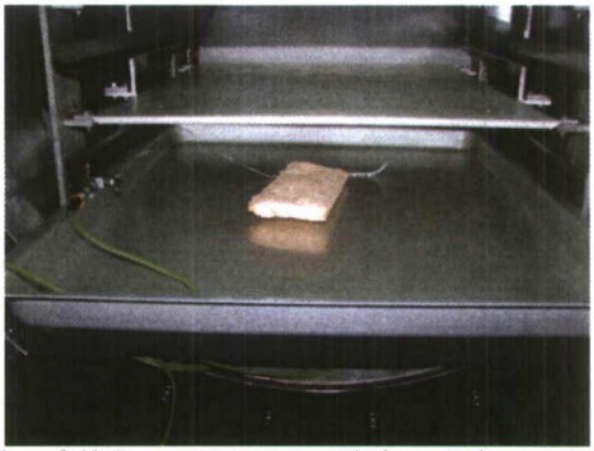


Figure 3-11: Frozen RTE bar placed in freeze drying chamber

5. Settings for optimum freeze drying are programmed via a computer controller as follows: Plate Temp 100°C, Product Temp 30°C, Chamber Pressure 0.35mmbar.

6. Freeze drying process is started and heating will automatically turn off when product reaches constant weight.

Baking

Baking is not a processing method commonly used in the nutrition or high-protein bar manufacturing. Most bar products are not thermally processed but are instead cold extruded for shaping and mixing. However, baking was adopted as a processing method for the RTE bar when the Milk Protein Precipitate (MPP) was added into the formulation. MPP was manufactured in-house and could be considered a "fresh cheese" which carries added food safety risks. In order to ensure a safe product, baking was incorporated to add a thermal heat "kill step" which would reduce potential microorganism load. The objective was to create a product with similar physical and nutritional properties as the established freeze dried product by using an alternative processing method. A general biscotti recipe was adopted which resulted in satisfactory sensory and physical qualities.

Baking took place immediately after the dough was rolled out to the specified dimensions (11.5" x 14.5" x 0.5") (Figure 3-8).

Baking Procedure

1. Preheat convection oven to 250°F.
2. Bake sheet on middle rack for 25 minutes.
3. Rotate sheet 180°, bake an additional 25 minutes.
4. Remove tray from oven and allow to cool for 15 minutes.
5. Cut sheet into the specified individual unit size (Figure 3-9).
6. Flip individual bars 180° over the top, return to baking tray.
7. Bake for an additional 20 minutes.
8. Remove from oven, allow to completely cool before packaging.

Vacuum Drying

This method was explored as a less expensive alternative to freeze drying, requiring only low thermal heating and no initial freezing. A vacuum oven generally used for moisture analysis was used to investigate this method (Figure 3-13). The required drying time depended on the amount of sample in the chamber and whether the chamber was heated. At ambient temperatures almost fourteen hours were necessary; only four hours were needed at 35°C to dry the product to a final moisture content below 20%.

1. Dough was rolled to the desired thickness (0.5") (Figure 3-8), and bars cut to desired dimension.

2. The cut bars are placed on the vacuum chamber tray.

Figure 3-12: Bar being placed into vacuum chamber before freeze drying

3. A partial vacuum (25" Hg) was drawn, and chamber temperature increased if supplemental heating was used.

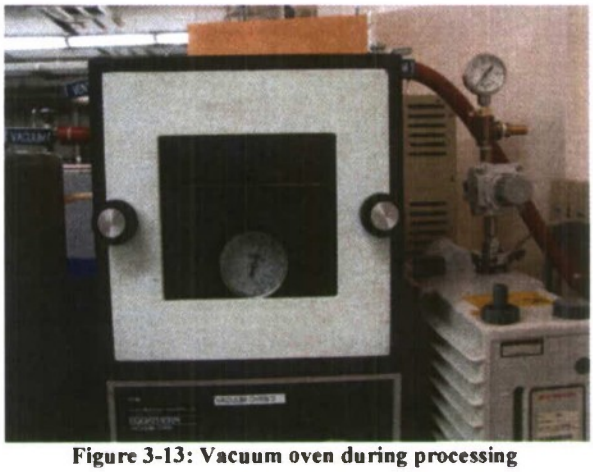


Figure 3-13: Vacuum oven during processing

Forced Air Drying

Dough was dried using a Harvest Saver air dryer designed to dry fresh fruit and vegetables. Using this commercial dryer, air velocity and temperature can be controlled. After several trials this method was deemed to be ineffective. It led to incomplete and inconsistent drying with very long processing times.

Protein Ingredients

Whey Protein Concentrate (WPC)

WPC 80 is the standard dairy-sourced protein used in the industry for bars and processed foods, from breakfast cereals to salad dressings. It has such far-reaching uses due to its superior functionality. It can provide texture, body, and nutrition (high concentration of EAA and BCAA). WPC was the first and primary protein used in the RTE formulation and continued to be used to boost protein content in later formulation trials (See appendix page 111). WPC represented up to 50% of the dry weight in some formulations, providing the bulk structure where carbohydrates are generally used in commercial bars.

Milk Protein Precipitate (MPP)

The shredded MPP curd is treated as a fresh ingredient which must be refrigerated and has a relatively short shelf life. The MPP was easily incorporated into the existing RTE bar formula because it mixed readily and did not clump. The MPP was added during the "wet" ingredient mixing stage (Figure 3-4). Initially, the mixture looked dry and non-cohesive. After mixing the moisture in the MPP begins to hydrate the starch and other ingredients causing the formation of uniform dough. MPP was utilized in several products as a unique protein source (See appendix page 129).

Skim and Butter Milk Powders

SMP and BMP are both relatively inexpensive products compared to WPC since they have a much lower protein concentration (30-38% dry basis) (see appendix page 109-110). These ingredients add protein and nutritional value to a formulation; however, they are not necessarily utilized as such in the industry. These ingredients generally are treated as fillers increasing the total solids of products. In addition, BMP and SMP have higher lactose contents which could make them less desirable in a nutrition bar formulation. Commercial BMP was used initially as an ingredient that could be replaced with a specially manufactured BMP powder with high phospholipid and sphingomyelin contents.

Micronutrients and Flavor Development

Several experimental formulations were tried in order to improve the overall nutritional benefit of the bar. These were executed as proof-of-concept trials to observe if ingredient additions would negatively impact the flavor of the product. The nutritional profile of bars can be easily modified to meet specific market or consumer demands, such as bars with high calcium, high fiber, or low carbohydrate (Hazen 2009).

Delactose Permeate

Two levels of delactose permeate addition were tried, in two different formulations of 2% and 4% dry basis, to the base formulation. The finished product was compared from a sensory perspective and the ash content was also analyzed (see appendix page 106).

Dietary fiber

Dietary fiber in the form of Fibersol-2® was added to the dough mix at various levels in substitution for flour. (See appendix page 107).

Flavor

Considering the importance of the sensory properties of the RTE bar, both sweet and savory flavors were tried. Liquid and powder flavors were added to formulations during dry mixing at the manufacture recommended levels. Cheese, chicken, BB-Q, pasta and mushroom, fruit, and vanilla flavors were tried (Table 7-1).

Probiotics

Probiotics need to be viable in order to provide their benefit, thus the initial survival of the probiotics during the processing is of special importance. Two processing methods (vacuum drying and freeze drying) and three potential probiotic lactic acid bacteria (Table 3-3) were investigated to see how they responded to the process and formulation treatments. In addition BMP and SMP were added as treatments to see if any synergistic effects exist between BMP and the probiotics survival.

Table 3-3: Three probiotic stains used in survival study

Strain	Species
MR220	*L. helveticus*
NCFM	*L. acidophilus*
23272	*L. reuteri*

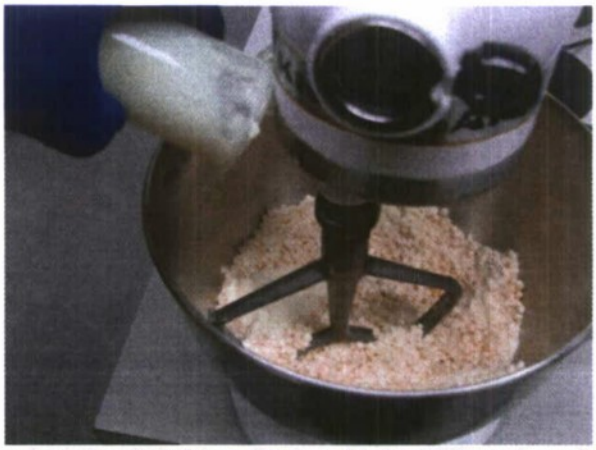


Figure 3-14: Probiotic "slurry" being added to RTE bar formulation

Fourteen hours prior to the experiment the specified probiotic was inoculated from a mother batch into 10ml of MRS + Cysteine (0.05%) broth and placed in a $C0_2$ controlled incubator held at 30°C. After this period the bacteria was centrifuged, washed and re-suspended in 10ml of water and added to either the BMP or SMP powders, creating a probiotic "slurry". This "slurry" was then added to the other ingredients and thoroughly mixed (Figure 3-14). A portion of the raw dough was removed and plated to establish the pre-treatment count. The remaining dough was split; half was placed in the vacuum oven and dried, the other half was placed in the blast freezer and freeze-dried the following day. After the bar was processed through either treatment, it was sealed in a high barrier pouch and plated the following day to obtain post-processing counts.

Direct fermentation of BMP powder to produce a "butter milk yogurt" with high viable cell counts was also tried. The theory was that survival of the probiotics could be improved in an environment conditioned by the bacteria instead of being incorporated from an isolated pellet. Initially commercial BMP was used to allow for future substitution. A 20% solution of BMP in DI water was prepared and allowed to hydrate

overnight. The following day the mixture was heated to 105°F, Danisco YO-MIX 533 40 375 DCU yogurt culture was added at 0.0002% w/w, and it was left to ferment for approximately three hours.

Results

Formulation

Table 3-4 shows the recommended formulation for further development. The initial formulation should be considered a "dry mix" formula; this could be produced using conventional current processing methods and ingredients. This formulation uses WPC for a sole protein source. The final formulation uses the MPP ingredient as well as WPC as its protein source.

Table 3-4: Initial formulation was utilized in several processing trials and the probiotic survival study The final formulation was used in the human subjects study.

Ingredient	Initial Formula	Final Formula
MPP	-	62.1
WPC	24	18.8
Sugar	-	10.3
HFCS	21	-
BMP	14	-
Butter	15	-
Water	12	-
Flour	8	3.7
Corn Starch	-	1.88
Puffed Millet	6	-
Salt	-	1.5
Flavor: Gold Coast #342991	-	1.4
Sucralose	-	0.1
Total	100	100

Table 3-5 Displays protein content and flavor observations of RTE bars formulated using whey protein crisps. Flavor observations were made by three subjects in an informal product evaluation.

Table 3-5: Qualitative and quantitative results of substituting millet with whey protein crisps.

Formulation	Percent Protein	Notes
TF1: Millet	22.62	Millet holds dough together, good clean flavor, no change to texture
TF1: 50% protein crisps	37.60	Crisps had weak structure, fragmented in mixer, absorbed water and became mushy
TF1: 70% protein crisps	32.74	Bitter astringent flavor, required greater amount, very hard to chew

Table 3-6 Displays the effects of delactose permeate added to formulations. The data was used to predict the effect and possible use level of delactose permeate. There were no perceived sensory differences between the control and added permeate products, although TF2.2 appeared to have a sweeter flavor.

Table 3-6: Protein and Ash content of two experimental formulas and trials to increase ash content using a "delactose" permeate (Delact) product.

Sample Number	Protein	Ash
TF2.1: Control	34.8	1.54
TF2.1: 2% Delact	34.1	2.2
TF2.2: Control	35.7	2.0
TF2.2: 4% Delact	34.6	3.3

Table 3-7 Displays water activity (aw) of several sample RTE bars. Number 1 is an early formulation before the addition of salt. Number 2 is the final formulation used in the human subjects study and contains salt. Number 3 and 4 are from the same batch but were located in different regions during baking.

Table 3-7: Water activity (mean of two samples), of select RTE bars after processing, comparing formulation, processing and position during baking.

Sample	aw	Moisture Content
Bake formulation (no salt)	0.905	22.7
Bake formulation (with salt)	0.890	20.7
Baked Edge	0.876	21.8
Baked Middle	0.885	22.8

Processing

Table 3-8 Displays the average moisture content of the RTE bar after processing. The target moisture content of the bar is >25%. Each processing method is capable of reducing the moisture content of the product, baking however results in a final moisture content which is closer to this specification.

Table 3-8: Moisture content and aw of RTE bars before processing (raw) and after processing. Each measurement is an average of several (n≥2). Moisture content is calculated on a wet basis, using a moisture oven. Water activity (aw) is measured using an aqua lab water activity meter.

Processing Comparison		
Processes	Moisture Content	aw
Raw	39.0	0.99
Baked	24.0	0.89
Baked + Vac	17.0	0.85
Vacuumed Dried	16.0	0.24
Freeze Dried	8.8	0.1

Table 3-9 displays the variability that exists in three batches of RTE bars made consecutively using formulation TF2.3. The results indicate the relative consistency in the formulation and processing of the bar.

Table 3-9: The moisture and protein contents on wet basis of in-between 3 batches produced in sequence. Moisture calculated using vacuum oven method; protein calculated using Rapid-N-Cube.

Batch	Protein	Moisture
A	36.1	19.25
B	35.2	20.5
C	36.5	18.84

Probiotic

Table 3-10 shows the statistical results from a probiotic survival study; the source heading lists the different factors that were tested, Prob> F lists the probability that the predicted result would occur randomly without being influenced by the before-mentioned factors. If a factor has a p value below 0.05 then it can be stated to have a statistically significant effect.

Table 3-10: Statistical results from a probiotic survival study. Process: freeze dried, vacuum dried. Formulation: SMP, BMP. Probiotic: MR220, NCFM, 23272.

Source	F Ratio	p Value
Formulation	0.0042	0.9496
Probiotic	3.0292	0.0853
Process	0.0447	0.8335
Probiotic*Formulation	2.8586	0.0957
Process*Formulation	0.0239	0.8778
Process*Probiotic	1.6512	0.2031
Probiotic*Formulation*Process	0.4231	0.6575

Table 3-11 shows how the Least Sq mean of a particular treatment combination is the mean of viability for that treatment combination. Superscript values show the treatment combination groups (NCFM | BMP with 23272 | SMP and NCFM | SMP); if any share a superscript then they are not significantly different from each other. If they have different superscripts then they are significantly different.

Table 3-11: Tukeys multiple comparisons= T-Test α=0.05, Least Sq Mean= Mean of % viability

Treatment Combo	Least Sq Mean
NCFM-BMP[a]	24.32
23272-SMP[a,b]	14.78
NCFM-SMP[a,b]	11.00
MR220-SMP[b]	8.14
23272-BMP[b]	6.39
MR220-BMP[b]	3.96

Figure 3-16 shows the concentration of probiotics in the raw dough (initial) and after processing (vacuum and freeze dried). Measurements are in CFU/gram and were found using the standard plate count method.

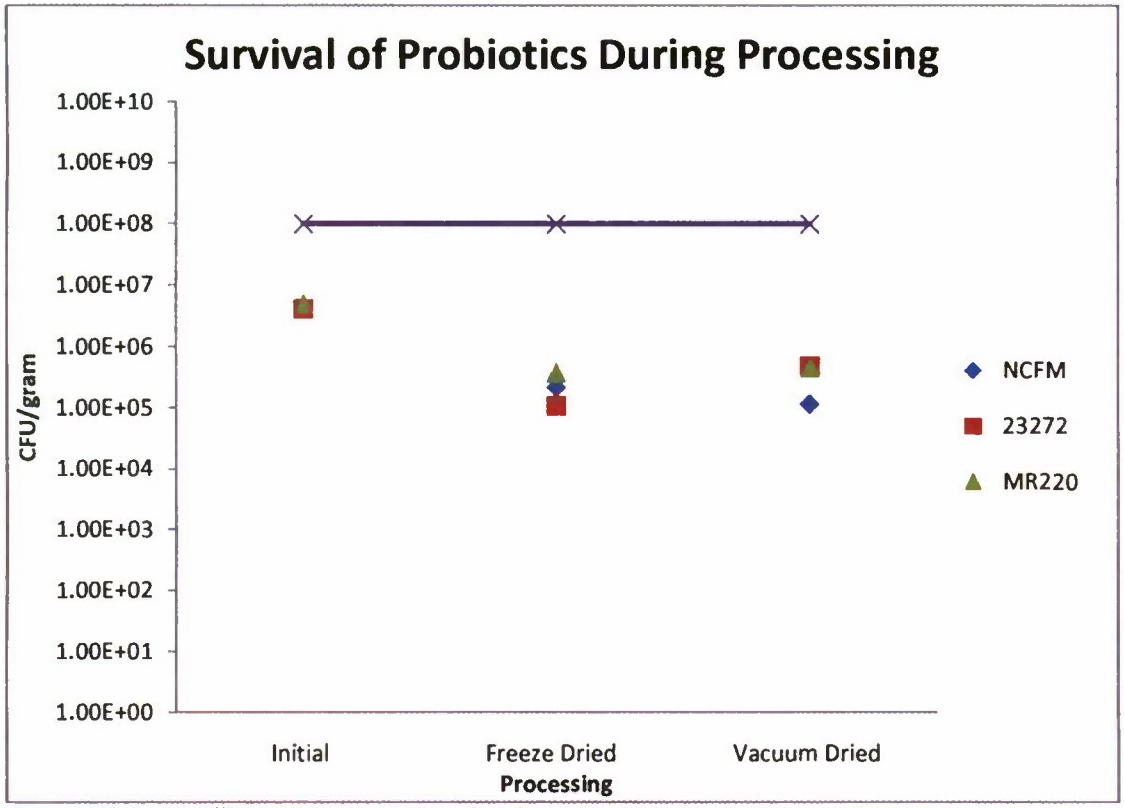


Figure 3-15: Mean CFU/gram (n=18) of selected probiotic strains before (initial) and after processing (freeze and vacuum dried.)

Flavor

Flavoring of the RTE was modifed continually throughout the development process. No formal sensory analysis was conducted but comments and observations were collected form advisors and peers. Flavors were ordered from several flavor houses (Firmenich, Gold Coast, IFF),and both liquid and powdered flavors were used. Savory (Chicken, Cheese) and sweet flavors (Cranberry, Vanilla) were tried. The human subjects study utilized a Cranberry flavor to create a sweet bar; no final flavor profile was selected. See page 113 in the appendix for a full profile of flavor codes and observations.

Discussion

The development of the RTE bar focused on two separate but related parameters: firstly, the RTE formulation which would influence both the nutritional profile of the bar and dough consistency; secondly, the processing of the bar, which affected the shelf stability and protein content of the bar.

Formulation

The product development process resulted in the finalization of two formulas (Table 3-4), which both met the stated RTE parameters (Table 3-1). The two formulations have the capacity to provide the desired nutritional benefits while maintaining dough that has the capacity for mechanized industrial production. The initial formulation utilizes only dry shelf-stable ingredients that are combined with HFCS and water. This formulation could easily be produced using existing ingredients and production equipment. The final formulation uses the MPP protein which was developed, and is outlined in Chapter 4. This formulation delivers the optimum protein, combining a complete precipitation of milk protein with supplemental WPC. This formulation could also be upscaled for industrial production, but would require either sourced or in-house manufactured MPP. The formulation using the MPP ingredient was used in the human subjects study (Chapter 5: Human Subjects Study).

The initial formulation seen in (Table 3-4) utilized puffed millet to "bind" the other ingredients and give the resulting dough structure. The mixed formula results in a viscous paste which does not hold shape and is tacky. The use of puffed millet, which has a high surface area and low density, reduces this. The combination produces dough that can be rolled, has structure and is not sticky. However the millet does not contribute to

the desired nutritional profile of the RTE bar. A study was designed to compare the feasibility of substituting the puffed millet for a "whey protein crisp" (Table 3-5). The use of whey crisps is common in the bar manufacturing industry, and the ingredient would contribute to the "energy dense" dairy ingredient profile of the RTE. Three separate RTE batches were produced: the resulting protein contents were approximately 22% in control millet RTE, 37% using 50% crisps and 39% using 70% crisps. This trend could be expected with the higher protein inclusions resulting in a higher protein content in the finished RTE bar. The whey crisps were much denser and required larger quantities to result in the same dough-binding effect. Additionally both whey crisp formulas had noticeable stale or bland off-flavors that were detected in the finished RTE bar. The final protein levels were considerably higher than the target (25% for a 100g bar) and the whey crisps delivered lower functionality than the millet control. The whey crisps would shatter in the mixer, absorb water losing their crispiness, and not bind as efficiently due to the higher density. Due to the low functional performance the whey crisps were not used in the later formulation as the millet met the desired expectations. The 50% protein crisps might have future application in the RTE bar formulation if higher total protein or EAA content is desired. The crisp damage seen in the preliminary trials could likely be reduced by using different mixing methods.

Table 3-6 displays the proof of concept for use of delactose permeate in the RTE bar formulation. Delactose permeate is a dairy derived calcium supplement that was added at varying levels to test its effect on mineral content and flavor profile. When compared to the control sample, the mineral content (expressed as percent ash) increased in both formulations when delactose permeate was added. This indicates an increased

calcium content in the finished bar which would be favorable in an RTE bar product. Increased calcium intake is beneficial for athletes due to its relationship to muscle contraction and hormone function (Anonymous 1997). Delactose permeate contains mostly minerals (namely calcium) which would interact through ionic binding with water to lower the aw of the RTE. No noticeable off textures were noticed when consuming the RTE containing delactose permeate, however theses bars were baked and considered sandy and coarse already. The TF2.2 control contained a greater ash content compared to TF2.1; this is due to the higher WPC and BMP content in the TF2.2 formulation both of which have high mineral contents.

Another consideration during formulation is water activity. After baking, the water activity level in the RTE bar was substantially higher than the target (0.65). However, salt was added at varying levels between 1.5-2%, the effects of which can be seen in Table 3-7. Adding salt to the formulation lowered the final product aw from 0.950-0.895. This is due to salt's ability to lower the water activity by ionic binding of the otherwise available water. Adding salt as the sole means to control aw will not be a practical solution as very high concentrations of salt would be necessary to achieve shelf stability. An additional concern with the formulation is the within batch consistency, bars produced in the center of the batch could result in a different composition than those on the outside. Table 3-7 has the moisture and aw of two samples taken after baking from the center of a batch and the edge. The two samples have the same moisture content regardless of position; this indicates that the moisture/solids ratio is the same with this batch.

Flavor

Flavor, texture, and the nutritional profile of the RTE bar, along with the dairy-heavy ingredients utilized make this bar an ideal product for a savory formulation. The majority of bars currently on the market are sweet, a result of both consumer expectations and the limitations imposed by using sweet tasting syrups and fruit pastes as binding agents. Although chicken and meat flavors may not be appealing, nacho or "Cheez-Its" flavors have the potential to work well. Flavor and texture were continually monitored by bench-top sensory analysis throughout the development process. Delactose permeate increased the ash content of the formulation which indicates an increase in calcium (Table 3-6). Additionally, delactose permeate is said to have salt-reducing and flavor-enhancing properties. These functionalities were not observed in this particular study, however TF2.2 was perceived as sweeter in both control and delactose samples. This was caused by the higher lactose content of the WPC and BMP compared to the MPP, which resulted in a sweeter tasting bar.

The volunteers in the human subjects study were asked to evaluate the bar they consumed. Subjects stated that the bar was dry and found it difficult to consume a complete serving. Many also described the RTE bar as unsatisfying after the strenuous hikes, preferring the carbohydrate bar. The majority of the subjects liked the control bar, the reasons being the flavor, level of sweetness, and texture. Several subjects indicated that the RTE bar had a "corn bread" like texture.

Shelf Life Expectations

No formal shelf life study was conducted; however, informal observations of bars over time followed the expected trends. Freeze dried bars with moisture <10% and water

activity <0.360 have been stable and without visual defects for approximately one year. Baked bars with higher moisture (<25%) and high water activity levels (<0.900) began to mold after approximately three weeks at ambient temperature (~75°F, 60%RH). All bars were packaged in clear high barrier PE/EVOH/PE vacuum pouches; samples used in the human subjects study used laminated foil pouches. This molding was perdictable as the water activity of the final bars was not low enough to ensure shelf stability, mold requires an aw above 0.7 to grow. The average final aw in a baked bar was 0.89 which was above the target level, but bars were stored in the refrigerator (<40°F) and were consumed within one week after production.

Figure 3-16: Baked bars molding after four weeks of ambient storage.

Processing

The purpose of processing the RTE bar was to reduce the moisture content of the bar to increase the solids and reduce the available water in order to produce a shelf stable product. Several processing methods were utilized to reduce the moisture and aw to the desired levels (Table 3-8). In order to meet these criteria, the bar would need low water activity (<0.65) and/or a suitable "kill step".

A baking formula was developed so that resulting RTE bars would be subjected to thermal processing which would limit the possible food safety risks associated with eating raw foods. The target was a finished product with low moisture content (~20%) and a water activity (< 0.65). This would produce results equal to the freeze drying process. However the low water activity level was never achieved. Table 3-8 shows the moisture drop that is achieved from the raw unprocessed dough to the finished baked product. The moisture content was close to the desired 24% compared to <20%. Although the low moisture objective could be reached with a slight alteration to the baking time, the water activity remained too high even after the addition of salt. This resulted in a bar that is not shelf stable and would benefit from refrigeration for extended storage. To conclude, the baking produced a bar with a short shelf life, but met the nutritional parameters.

Vacuum drying was investigated as an alternative processing method for the RTE. Vacuum drying does not require any prior freezing of the product as in freeze drying, and can remove the majority of the available water resulting in a moisture content of approximately 16% and aw below 0.24. This stability is achieved because the processes take place in a lower pressure environment where the boiling point of water is reduced, which allows the free and available water to evaporate at a lower temperature. Additionally the vacuum creates a draw which removes the moisture from the chamber. The vacuum oven used was not large enough to make the quantity of bars necessary, thus halting additional vacuum processing research.

The baking process was optimized (Chapter 3: Baking) by controlling for within-batch and between-batch repeatability. Three batches were made consecutively to study

flavor profiles (Table 7-1), comparing different berry flavors at the same use level. The batches used the same formulation and processing paramcters and the final moisture content levels of all three were collected (Table 3-9). These results were used to interpret the variability that can occur between different batches of baked RTE bars. The results showed that the random variability was not great (0.5-2.5%), and that the desired low moisture content could be met. This indicates that the production of the bar results in relatively consistent parameters. Within-batch variability of moisture content was considered a potential factor. This was measured by taking a sample from the middle of the tray and one from the outer edge after baking(Table 3-7). The moisture profiles within the bar followed a predictable trend (high moisture in center, lower on edge), but the difference was marginal (0-1.5%) indicating that the position during baking is not critical for predicting final moisture.

Probiotics

To test the practicality of inoculating viable probiotics into the RTE bar, their survival through the initial processing had to be verified. This study was a multi factorial (2x3x2) experiment comparing the effects of: Formulation (BMP-SMP), Probiotic (NCFM, MR220, 23272), and Process (Freeze Dried, Vacuumed Dried). Initial statistical analysis of the data collected indicated there were no significant factors in the analysis (Table 3-10). This can be interpreted as no combination resulting in any different level of survival than another. The Probiotic | Formulation interaction which was compared in the initial analysis showed a p value approaching significance, 0.09 at an alpha of 0.05.

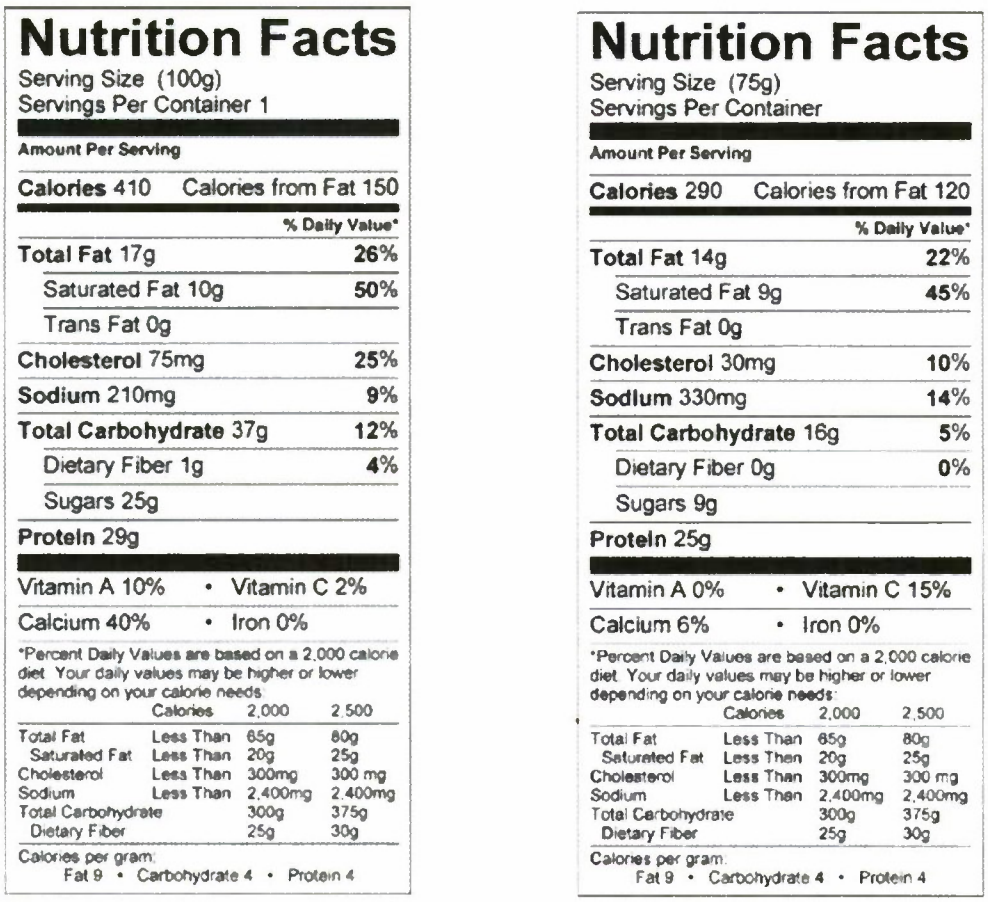
If a multiple comparisons model is made of this interaction NCFM | BMP is seen as separate from three of the five combinations which are all in one group (Table 3-11).

While this is not statistically significant as the initial interaction was non-significant (p-value 0.09 > 0.05), this does indicate a trend which could be further researched. NCFM is a widely studied strain because of its binding potential to the milk fat globule membrane (MFGM), and it indicated a significantly higher viability when BMP was used compared to SMP. Even though the main effects of the formulation did not indicate any significant difference (p-value 0.9496), the relationship between BMP as a functional ingredient for the protection of probiotics has potential for further investigation (Table 3-10).

Comparing the initial counts in the raw dough to those of the processed RTE indicated another interesting trend. Figure 3-15 shows the CFU counts of the raw product averaged over the formulation, and different repetitions for the different probiotics. The initial inculcation is consistent at approximately $1x10^7$, while the processed bars appear to suffer a one log reduction resulting in post counts at $1x10^6$. Both of these values are below the level recommended by the National Yogurt Board (NYB) at $1x10^8$. The initial inoculums were limited in cell density, and this was then diluted into the mass of the dough formulation, resulting in the low counts. A larger or more concentrated inculcation could lead to a higher cell count and sufficient post-processing survival. Most interesting is the relative subtlety of the vacuum drying process, despite occurring at elevated temperatures and being a longer process. There were no significant differences between the two processing treatments (p-value 0.83) (Table 3-10). This can also be seen in the survival graph (Figure 3-15) with the same count resulting from both methods with all probiotics strains. This would seem to indicate that vacuum drying could represent a more cost effective alternative to freeze drying for the preservation of a functional RTE food.

Conclusion

A novel protein bar formulation has been developed utilizing a high quality dairy protein source with the goal of reducing muscle wasting in soldiers and endurance athletes. Providing energy dense nutrition in a portable form is a valid option to deliver the required nutrients necessary to minimize the detrimental effects of undernourishment and chronic stress experienced by this demographic. Dairy is considered one of the best sources of high quality protein. It contains several protein fractions each contributing to the overall nutrition delivered to the consumer. The two formulations that were developed deliver the benefits of dairy nutrition in the convenience of energy dense bar (Figure 3-17). The formulation on the left relies on only dry conventional ingredients and provides higher protein content than initially specified. The formulation on the left was reduced to 75 grams for the purpose of the human subjects study and provided the desired protein content but had reduced calorie content.



Nutrition Facts
Serving Size (100g)
Servings Per Container 1

Amount Per Serving

Calories 410	Calories from Fat 150
	% Daily Value*
Total Fat 17g	26%
Saturated Fat 10g	50%
Trans Fat 0g	
Cholesterol 75mg	25%
Sodium 210mg	9%
Total Carbohydrate 37g	12%
Dietary Fiber 1g	4%
Sugars 25g	
Protein 29g	

Vitamin A 10% • Vitamin C 2%
Calcium 40% • Iron 0%

*Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

	Calories	2,000	2,500
Total Fat	Less Than	65g	80g
Saturated Fat	Less Than	20g	25g
Cholesterol	Less Than	300mg	300 mg
Sodium	Less Than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Calories per gram:
Fat 9 • Carbohydrate 4 • Protein 4

Nutrition Facts
Serving Size (75g)
Servings Per Container

Amount Per Serving

Calories 290	Calories from Fat 120
	% Daily Value*
Total Fat 14g	22%
Saturated Fat 9g	45%
Trans Fat 0g	
Cholesterol 30mg	10%
Sodium 330mg	14%
Total Carbohydrate 16g	5%
Dietary Fiber 0g	0%
Sugars 9g	
Protein 25g	

Vitamin A 0% • Vitamin C 15%
Calcium 6% • Iron 0%

*Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

	Calories	2,000	2,500
Total Fat	Less Than	65g	80g
Saturated Fat	Less Than	20g	25g
Cholesterol	Less Than	300mg	300 mg
Sodium	Less Than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Calories per gram:
Fat 9 • Carbohydrate 4 • Protein 4

Figure 3-17: Nutritional Facts Panel for final formulations, panel at left using only WPC, right panel using MPP and reduced to 75 gram serving size.

Baking can effectively be used as a processing method for the production of an RTE bar. Baking allows for the production of large batches and increases control over the drying and processing of the RTE. Some of the advantages of baking are that it yields a safe product, it is a standard industrial technology, and is relatively inexpensive. Baking also has disadvantages, such as the development of a cooked flavor resulting from Maillard browning, dry crumbly texture, higher residual moisture and water activity. Baking can also be coupled with an additional processing step like ambient vacuum drying to decrease final moisture. Freeze drying is also a potential option for processing

an RTE bar. It results in a final product with good sensory and physical qualities, producing a product that has a crisp texture, sweet mild taste, and very low moisture and water activity levels. Freeze drying has high additional costs and is not common in the industry although it has the potential for large scale manufacturing.

Many flavors were incorporated throughout the development process (Table 7-1), but no profile was finalized. The formulation lends itself to this diversity; the unflavored bar is mild with a light dairy flavor which is a versatile base for further flavoring. This could allow the same formulation to be flavored several ways, which was one criterion of military rations. Flavoring the bar would require extensive sensory trials; while some preliminary work was done, an extensive sensory study would be very useful.

Probiotics are one of many functional ingredients that could be added to this and other RTE bar formulations. In initial studies we have seen some indication of probiotic survival after processing. However, there are other methods for incorporating probiotics like sporeforming shelf stable probiotics, encapsulation, and in-bar fermentation that could be investigated. Probiotics are of great interest in this type of product mainly because of their positive association with gut health and immunity. Other tracc nutrients and functional ingredients could readily be incorporated into the formulation. Adding minerals, vitamins, and high fiber are possibilities that are already on the market and have been shown to function in RTE bar formulation. The development of MPP as a viable protein was necessary to achieve the desired protein profile for RTE formulation. Its utilization led to a formulation that met the nutritional specifications but without the flavor defects seen with WPC.

4. MILK PROTEIN PRECIPITATE (MPP)

Introduction

Ricotta is a classic Italian style cheese traditionally manufactured from whey. Its unique manufacturing process combines high heat and acidification for protein precipitation. The process of ricotta cheese manufacturing was used as a model for the production of Milk Protein Precipitate (MPP). Precipitating all the protein present in fluid milk and using the resulting curd in an RTE bar formulation would deliver high quality protein in a convenient and novel form without the negative flavor of dairy powders. Using concentrated lactic acid, high heat, and longer holding times results in a soft curd (MPP). The mild flavored curd can be used as a nutritional ingredient in further processing - a type of industrial cheese.

The process for the production of MPP was refined in the Dairy Products Technology Center (DPTC) and the resulting curd produced a reliable composition (Table 4-1). This ingredient was incorporated into the RTE bar formulation to replace a portion of the WPC protein. While WPC is the concentration of a small fraction of possible milk proteins, MPP was developed to contain a much larger fraction of milk proteins. This property makes MPP closer to the intrinsic nutritional value of milk. The purpose of creating the MPP was as an experiment to test the possibility of producing a high protein product from milk. This protein ingredient was tested using several analytical testing methods in order to establish its composition. The functionality of the protein ingredient was also tested in the RTE formulation and other products.

Table 4-1: Approximate analysis of MPP (n=9) batches. Whole milk ricotta as described by Kosikowski 1982

Components	MPP Specifications	Whole Milk Ricotta
Protein (N-cube)	25- 27	11.2
Fat (Babcock)	23-25	12.7
Moisture (Microwave Drying)	35-40	72.2
Carbohydrate (Difference)	2-5	3.0
Ash (Muffle Furnace)	0.75-1.5	-

Materials and Methods

The production of Ricotta and Ricotone style cheeses is outlined by Kosikowski. MPP follows these guidelines and was adopted to create a curd with a significantly different nutritional profile (Table 4-1). The fluid milk is heated to a higher temperature before acid addition (180 °F verses 176°F), the pH is reduced further (4.6 versus 6.0-5.9) (Kosikowski 1982). A holding period at the elevated temperature was also included, this allowed for an increase in curd strength.

Ingredients and Equipment

- Whole milk, skim milk, whey depending on desired fat content
- Lactic Acid (88%), diluted to 35% in H2O
- Portable temperature compensating pH meter, 10mL pipette
- 1.5 L plastic container
- Large stirring paddle, and scooping device
- Double jacketed steam kettle

- Cheese molds
- Thermometer
- Cheese shredder.

MPP Manufacturing Process

1. 10 gallons of full fat raw milk is received from the dairy.
2. The milk is added to a double jacketed steam kettle and heated with periodic stirring to 180°F.

Figure 4-1: Raw Milk agitation and heating

3. One liter of milk is removed and acidified with 35% lactic acid until the pH is < 4.6. The quantity of acid needed to acidify the remaining 10 gallons is then calculated.
4. When the milk reaches 180°F, it is agitated vigorously until a "cyclone" forms in the center of the milk. The calculated quantity of acid is then added to the milk and agitation is stopped.

Figure 4-2: Acid being added to heated milk during "cyclone" agitation

5. A lid is placed over the kettle to help retain heat, and the milk is heated until it reaches 190°F.

6. The temperature is held constant at 190°F for 30 minutes.

7. The lid is removed. The curd should have floated to the surface creating a solid layer. This layer is then scooped carefully to avoid shattering of the curd. Too much agitation or dripping whey can break up the remaining product.

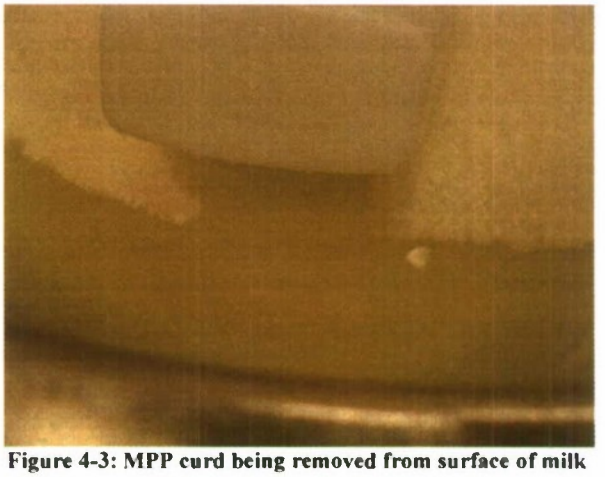


Figure 4-3: MPP curd being removed from surface of milk

8. The scooped curd is then transferred to Gouda cheese molds where the free whey is allowed to drain.

Figure 4-4: MPP curd after removal from double jacketed steam kettle, draining whey.

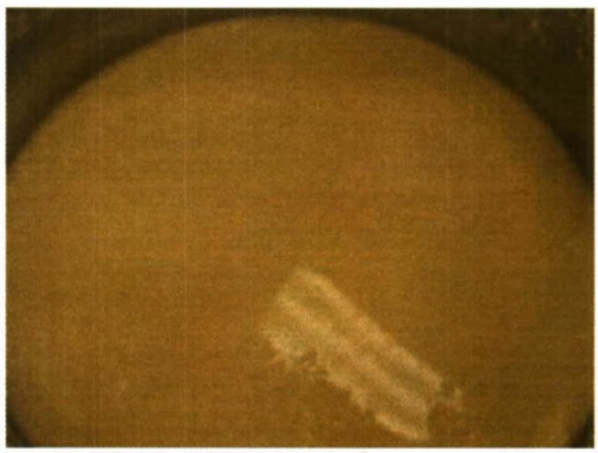


Figure 4-5: Whey remaining in kettle after removal of MPP curd.

9. The MPP curd is pressed under 10lb weights for 2 hours. This removes free residual liquid forming a close knit curd.

Figure 4-6: MPP curd being pressed to remove excess moisture

10. The curd is then vacuum-sealed into cheese bags and refrigerated overnight.

Figure 4-7 Vacuum packed MPP cheese wheels before refrigeration

Figure 4-8: Formed MPP curd before shredding

11. The curd wheel is removed from the package and shredded to form a uniform moist protein ingredient.

Figure 4-9: MPP curd being ground in Kitchen Aid mixer

12. The shredded curd is then packaged and is ready for use. The curd is vulnerable to mold and spoilage bacteria so must be refrigerated. The milk is sterilized during the processing but the moist curd is an ideal target for post processing contamination (Kosikowski 1982).

Results

Table 4-2 compares MPP and Ricotta manufacturing efficiency, by analyzing protein and fat content of the starting ingredient (raw milk, whey) and the resulting residual liquids.

Table 4-2: Comparing fat and protein content of MPP to classic Ricotta using FOSS 4/6/10

Ricotta	Fat	Protein	MPP	Fat	Protein
Whey	0.26	0.9	Milk	4.16	3.5
Residual Liquid	0.022	0.55	Residual Liquid	0.066	0.39

Table 4-3 compares MPP and Ricotta in terms of protein and moisture contents. Both curds were either pressed with a 10lb weight for two hours or left to drain naturally.

Table 4-3: Compositional analysis of acid and heat coagulated curd (MPP and Ricotta) from either Milk or Whey respectively 4/6/10

Ricotta	Moisture	Protein	MPP	Moisture	Protein
Non -Pressed	79.86	10.8	Non – Pressed	60.29	21.8
Pressed	64.22	12.1	Pressed	35.2	27.2
Approximate Yield		1%	Approximate Yield		11.1%

Figure 4-10 is a chart displaying the quantity of acid (35% lactic acid) needed to change the pH of milk. Six different MPP productions are recorded to construct an acidification chart. Approximately 15ml of 35% lactic acid is needed to acidify one liter of raw milk at 50°F.

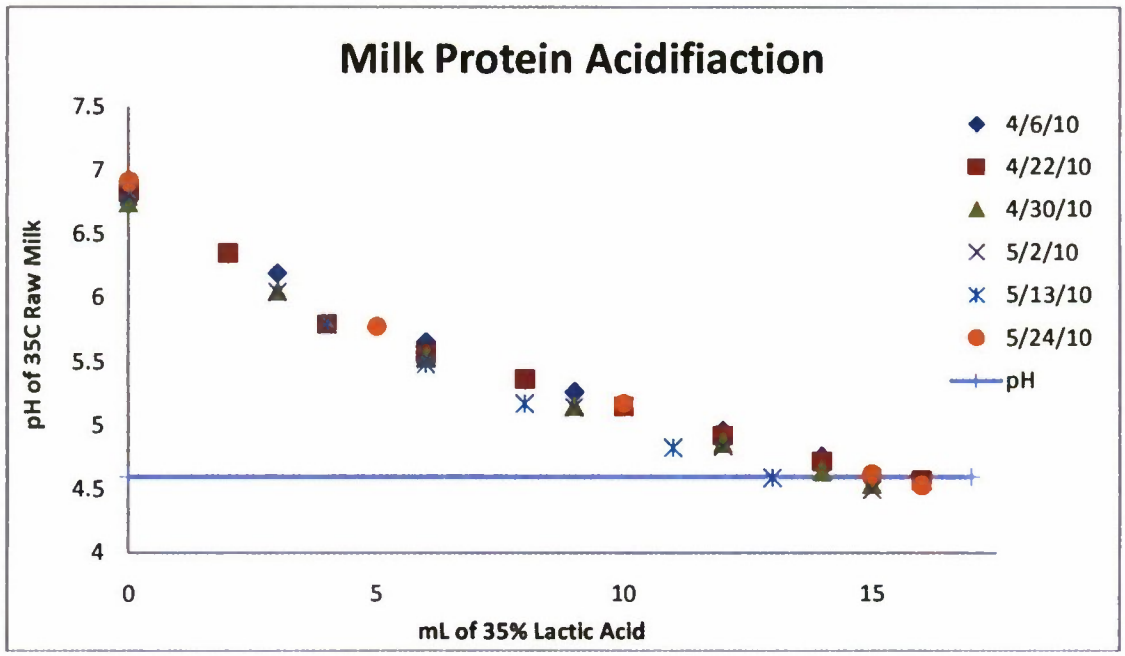


Figure 4-10: Acidification curve of several MPP batches, using 35% lactic acid Raw Milk

Table 4-4 is an approximate analysis of two MPP batches manufactured on different days. Fat analysis was conducted using both the Mojonnier and Babcock methods. Protein determination was conducted using the Elementar Rapid-N-Cube. Mineral or ash content were determined using a muffle furnace, and moisture content using the vacuum oven method.

Table 4-4: Milk protein precipitate compositional analysis of different batches (values averaged)

Date Tested Sample	Protein: Rabid-N-Cube	Fat: Babcock	Ash: Combustion	%MC: Vacuum oven
4.13.10	23.4	22.87	1.095	43.17
4.22.10	25.81	27.5	1.23	40.27

Table 4-5 shows the results of the protein analysis of different MPP productions; the production method was kept consistent.

Table 4-5: Protein content of different MPP batches, results done in multiples using Rapid-N-Cube

Date Performed	Sample	Percent Protein
4.30.10	MPP	24.76
5.02.10	MPP	24.61
5.13.10	MPP	20.84
5.27.10	MPP	27.44
5.28.10	MPP	26.61

Figure 4-11 is an urea gel depicting the protein profiles of several samples related to MPP production. MPP's starting ingredient milk (well one), the residual liquid (well 2), and the MPP protein from two manufacturing dates (wells 3 and 4). These are in contrast to cheddar cheese and cheddar whey (wells 5 and 6). Well 7 shows the protein profile of the RTE bar compared to that of a competitor (well 8), and WPC (well 9).

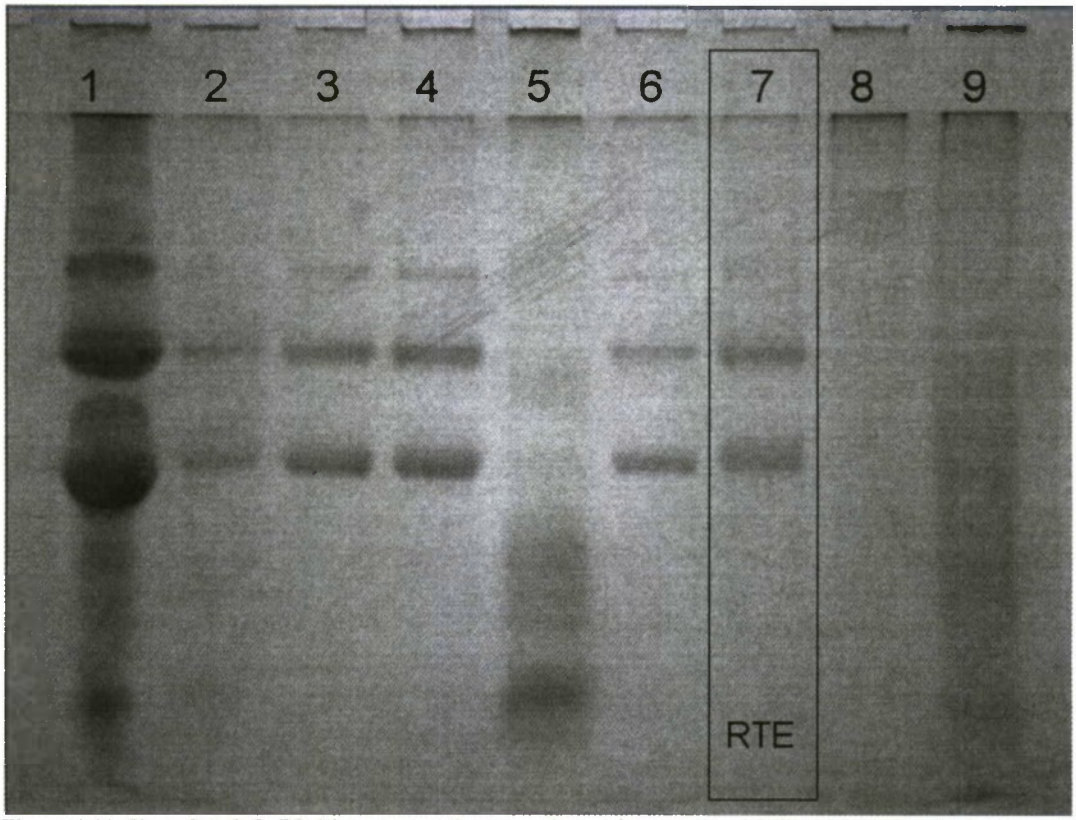


Figure 4-11: Urea Gel, 8uL filled into each well, assorted protein samples. 1:MPP Milk (5/27), 2: Residual liquid from MPP production (5/27), 3:MPP from (5/27), 4:MPP produced (5/5), 5: Whey from Cal Poly Cheddar Cheese (5/27), 6: Cal Poly Cheddar Cheese Curd (5/27), 7: RTE bar made using MPP (5/17), 8: Cliff Bar "Builder Bar", 9: WPC

Discussion

Production of MPP was based on the manufacturing methods for ricotta. The initial process focused on establishing a unique production procedure for MPP then comparing that to ricotta. What separates ricotta derived from both whole milk and whey from MPP is the complete co-precipitation of protein that occurs. This leads to higher concentrations of protein and fat and a lower moisture product (Table 4-1). The difference between ricotta and MPP can initially be identified during the production procedure. Cheese whey was used to manufacture ricotta; this contains a much lower

protein content than milk (0.9% versus 3.4%). After precipitation the final protein content of the two residual liquid streams are similar, with the MPP registering lower (0.4 verses 0.55%) (Table 4-2). The greater efficiency at removing the total protein from the liquid phase is due to interactions that form in-between the casein and whey protein fractions. The effectiveness of MPP can also be compared to rennet cheese manufacturing. The remaining protein left in MPP residual liquid contains approximately 0.45% protein while rennet whey contains approximately 0.9%. Co-precipitation is achieved by targeting the iso-electric point of casein with the low pH, and the whey precipitation using high heat. This process forms complex protein aggregates of both casein and the whey proteins (α-lactalbumin, β-lactoglobulin) while the single precipitation of whey forms a more simple fragile curd.

This co-precipitation has a large impact on the resulting curd produced, increasing the protein content and reducing the moisture (Table 4-3). The higher protein content and yield that is seen with the MPP is due to higher initial protein in milk (Kosikowski 1982). However the lower moisture that is seen results from the syneresis that occurs in the co-precipitated protein aggregates during the hold time. The residual liquid is expelled as the proteins hydrophobic regions bind, however in ricotta this is not as significant. To further reduce the moisture content and increase the protein concentration the curd was pressed. Pressing was defined as placing the curd in a Gouda cheese hoop under a 10lb weight for two hours, and Non-Pressed as the curd left to drain in a plastic basket for two hours. This additional step reduced the curd moisture 42% in the MPP versus 20% with ricotta. This again results from the co-precipitation where the aggregates have less moisture most

of which is easily removed, while the whey ricotta has lower solids and a greater amount of residual liquid that is incorporated into the curd during precipitation.

The acidification of MPP is an integral step in the co-precipitation of the dairy protein. Acidification is an important step in heat precipitated cheeses but the extent of the acidification is much greater in the MPP manufacturing than that of whole milk or whey ricotta. Acetic and citric acid are commonly used in the manufacturing of ricotta, but result in distinctive flavors in the resulting curd. Lactic acid produced by inoculated starters results in a milder flavored curd and is the recommended acidulate for whole milk ricotta (Kosikowski 1982). This information led to the use of 35% lactic acid in the MPP manufacturing process which allowed for quick and consistent acidification of the milk (Figure 4-10). While day-to-day variability would exist in the production of MPP, a use level of approximately 14-16ml of 35% lactic acid is necessary to acidify one liter of milk. This value could be used to create a standard manufacturing procedure for MPP.

The composition of MPP is detailed in the approximate analysis of two separate batches manufactured on two dates (Table 4-4). These two batches fit within the range of the MPP specifications outlined in Table 4-1. While the composition of the two batches shows variability, it is apparent that a similar ingredient is produced. Larger variability is possible, as is seen in Table 4-5 where the protein content ranges between 20% and 27% across five separate batches. This large range is due to the handmade batch processes used to make this product. Many factors could influence the protein content of the final product and not all can be controlled without proper mechanization of the process. Poor efficiency during initial curd removal could lead to higher residual moisture; improper acidification and heating would lead to less complete protein precipitation. Many

additional factors would influence this bottom line, however industrial production would greatly limit this variability.

MPP was developed to deliver a high protein curd with a composition identical to that of milk and dairy proteins. The MPP would then be utilized as an ingredient to deliver this protein to the consumer in the form of a RTE bar. A urea electrophoresis gel was prepared using selected protein fractions (Figure 4-11). This gel illustrates the potential for the RTE bar to have the same physiological effects as dairy when MPP is used. The urea gel is able to visually demonstrate the change in protein through processing. Well 1 is the protein profile of milk, which demonstrates the complexity of the RTE protein target. Wells 3 and 4 both contain MPP protein and show similar profiles when compared to the original milk sample. Additionally, looking at the profile of the resulting whey can give an indication of what proteins were not precipitated. Looking at well 2 which contains the whey left over after MPP processing, the profile looks similar to the MPP itself. This is a striking difference from the profile of cheddar cheese whey in well 5 and curd in well 6. This is because the processing of the MPP is designed to precipitate more total protein from the milk, thus resulting in a curd with a more "complete" profile. The protein profile of WPC in well 9 shows no distinct bands or pattern. This is due to the processing methods of WPC which subject the proteins to thermal and mechanical degradation. The "complete" protein profile is preserved through the processing of the RTE bar (well 7) with a protein profile almost identical to MPP used in its manufacturing.

Conclusion

MPP delivers the positive nutritive value of milk protein, and represents an alternative method of incorporating dairy in products beyond powders and fluid milk. This ingredient is manufactured using high heat (190°F) and low pH (4.6), which results in a complete precipitation of both whey and casein proteins. MPP exhibits excellent functional properties and provides added nutrition in the RTE bar and other products. The MPP can aid in the structure and binding of formulations in a similar way to HFCS or wheat gluten. MPP can additionally be used to simplify product formulations since it incorporates fat, moisture and protein. MPP can also serve as a binder and/or bulking ingredient in a bar or yogurt, while having the added benefit of being derived from dairy. In order to ensure that its functionality and potential is realized, MPP would require further work, particularly concerning shelf life and storage. As a fresh, un-aged cheese with high moisture MPP has little hope of a commercially viable shelf life. The addition of preservatives or the utilization of aseptic packaging or freezing could extend the shelf life but further development is necessary. Overall, MPP has the added benefit of being derived from dairy and being perceived as a natural ingredient.

5. HUMAN SUBJECTS STUDY

Introduction

The benefit of supplemental protein for sports recovery is a current and highly debated topic. Proponents point to the anabolic effect of dietary protein and the improved energy-balance gained from consuming protein post exercise. Opponents hold that the total caloric energy and the metabolic ease that carbohydrates offer are more important for recovery (Jentjens et al. 2001). It has been suggested that active individuals require greater amounts of protein than the RDA (Anonymous 1997). These excess amounts are at 50-100% compared to their sedentary counterparts (Lemon 1987). The development of the high protein RTE bar was based on the former premise, aiming to deliver protein in a calorically dense dietary supplement. The hypothesis for this study is that providing 25 grams of dairy-derived protein post exercise would improve body composition measured by weight and body fat and reduce inflammation and physiological stress markers measured in the blood. The purpose of the blood markers was to compare the physiological effect of the treatment bars on post exercise recovery.

Blood Markers

The following five blood hormones were selected as markers for this study for their correlation with inflammation and metabolism:

1) Erythropoietin (EPO): A hormone involved in the production of red blood cells. It is also responsible for promoting neuronal survival after hypoxia and other trauma (Sirén et al. 2001). EPO is involved in the biological signaling of the brain and nervous system and has also been associated with cellular proliferation.

2) Hydrocortisone (Cortisol AM): Cortisol levels can serve as an indicator of hypothalamus pituitary adrenal axis (HPA) activity, which is involved in neurological stress responses. Fluctuations in individual cortisol levels are also associated with perceived stress (Wust et al. 2000).

3) C-Reactive Protein (CRP): A hormone whose concentration in the blood is directly related to the immune system response to tissue injury, infection, and a key inflammation marker. CRP level's are routinely tested when evaluating human diseases and are associated with the immune system (Thompson et al. 1999).

4) Creatine Phosphokinase (CPK): An enzyme responsible for the reversible conversion of ATP to ADP. CPK is said to function as an energy transporter, delivering the energy from the site of production to that of utilization. CPK is also said to have a buffering capacity, functioning like an energy storage mechanism (Wallimann and Hemmer 1994).

5) Adolase: An enzyme involved with fructose metabolism which can be correlated to the dietary intake of carbohydrates (Munnich et al. 1985).

In this study, paid subjects performed strenuous hikes on three consecutive days, which were repeated after a one week rest period. The subjects hiked specified routes carrying 20% of their body weight in backpacks. The physical activity prescribed was designed to mimic military combat situations where physical and mental stresses are high. The participants were split into two teams which competed for speed and tactical points. A single blind cross-over design was implemented with subjects receiving one treatment bar after completing each hike during the first test period and the other bar through the course of the second test period.

Subjects and Methods

Participants

Recruitment was open to all Cal Poly students but focused on the regional ROTC, wrestling, cross country, and swim teams. The subjects were all young and athletic between 18 and 30 years of age. Subjects were screened for milk allergies and other medical conditions and asked to fill out Physical Activity Readiness Questionnaire (PAR-Q) forms and Subject Information Forms for legal compliance. The study was approved by the Cal Poly Human Subjects Board and the participants were informed of the potential risks involved in the study. The subjects were told not to control their diet or exercise routines for the purposes of the study but were asked to fill out diet logs for the dates involved. On the night before each test period (5/20/10 and 6/03/10), the subjects were invited to a free carbo-loading dinner. This also served as an orientation session where the researchers were available to answer questions and provide information on the study.

Protocol

The study consisted of two three-day test periods, separated by a one week rest or wash-out period. The two test periods consisted of three consecutive days of strenuous hikes, chosen for their length (between 6 and 8 miles) and difficulty (topographical images are available in the appendix page 114). Each complete test period was considered a treatment, with one team being administered the test bar and the other team the control bar. On the first day of the test period (5/21/10, 6/04/10), subjects went to the University Health Center to have baseline blood draws and have their weight and percent body fat determined. The subjects' body weight was recorded and used to calculate their

"carrying load" which was to be used during the test period hikes. Each subject was required to carry 20% of their body weight which included their backpack, super soaker, and water bottle. The remaining carrying weight was reached using sand bags. Each subject was provided a 32oz sports bottle filled with lemon and lime Gatorade during each hike.

On the mornings of the test periods, subjects met at 11 am in the Kinesiology department building. The subjects joined their respective teams and were given their water filled super soaker and sports bottle. When the subjects were ready, one team which alternated started first and was followed 15 minutes later by the second team. A "medic" was assigned to follow behind the last subject and was responsible for picking up garbage, carrying a cell phone, attending to minor injuries, and carrying additional water. As the subjects completed the hikes, they were given their corresponding treatment bar and water was made available. The subjects were asked to completely consume the bars. After finishing their respective bars, the subjects completed an anaerobic power test, a "30 second Wingate", and their post hike choice reaction test. They were then were free to go. This process was completed on each of the three days during the test period.

Response Variables

Body Composition: Measured by static and underwater weighing Friday (5/21, 6/04) and Sunday (5/23, 6/06). Response reported as change between Sunday and Friday measurements.

Simple Choice Reaction: Time to respond (lift corresponding finger) to a light stimulus, measure before and after each hike. Response reported as daily change between after and pre hike reaction time.

Blood Draws: Samples taken on Friday (5/21. 6/04) and Monday (5/24, 6/07). Response reported as change between Monday and Friday blood draw.

Peak Power: Measured as highest mechanical power (Watts = Force x Distance) generated during the first 5 seconds of a 30 second Wingate Test. Measured after each hike, response reported as daily peak power.

Experimental Design

The experiment followed the simple cross-over design (Woods et al. 1989) common in small scale medical studies. The study intended on being single blind, with the teams receiving one of the two treatment bars unknown to them. The initial group of subjects (n=36) contained two females and thirty-four males. On the first day of the study (5/21/10), the group was randomly separated into two teams, either "Green" or "Gold" by the flip of a coin. One female from the Gold team was moved to the Green team to balance the sex ratio between teams. The resulting teams, Gold (n=17) and Green (n=19), were given colored t-shirts. During the first test period (5/21 – 5/23/10) the Gold team was given the test bar and the Green team the control bar, this was decided by a coin toss. After a one-week rest period, the treatments were flipped, the teams received the alternate bar, and the same hikes were performed.

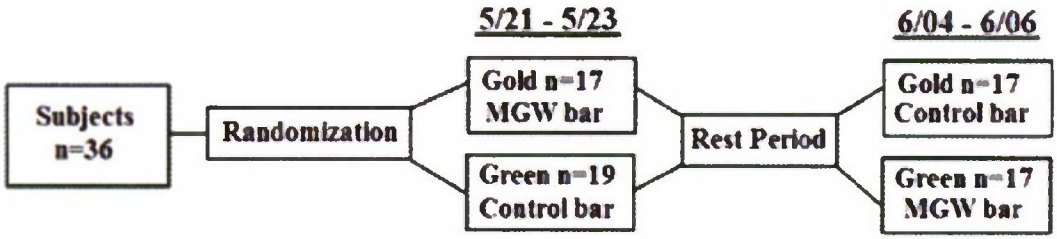


Figure 5-1: visual summary of experimental design

Statistical Analysis

The Mann-Whitney test was used to analysis the data. This is a two-sample non-parametric comparison tool, similar to the conventional t-test. The Mann-Whitney, however, uses the median values from the two groups as opposed to the means used in the t-test. This allows the Mann-Whitney to be used for data which does not meet the assumptions of the t-test, namely normality and equal variance. Non parametric methods are commonly used for human response data where variability between subjects is high and outliers are common (Koch 1972). Early analysis of the collected data shows that the assumptions of normality and equal variance were violated. This resulted from several extreme outliers which occurred randomly during each blood draw (a table of the different outliers for each marker can be found in the appendix(Table 7-2). Due to the apparent variability in the results, all statistical analyses used the nonparametric analysis tool "Mann-Whitney" with a comparable significance level of 5%(α=0.05). Using the cross-over design allowed each subject to function as their own control (Woods et al. 1989), as each subject experienced both treatments. The main concern with a cross-over design is the potential for a "carry-over" effect, indicating the treatment administered during the first period had lingering effects into the second period. This can be avoided

through a long flush out time and by using treatments without lingering effects. The following statistical model was used to test for treatment effects and possible carry-over effects:

Equation 5-1: Two-stage model adapted from (Shen and Lu 2006)

$Y_{ijk} = \mu + b_{ij} + \pi_k + \Phi_m + \lambda_m + \xi_{ijk}$

i=treatment order, j=subject, k= week, m= treatment

μ= overall mean

b_{ij} = effect of j^{th} subject with i^{th} order and is $\sim N(0,\sigma_b^2)$

π_k = effect of the k^{th} week

Φ_m = direct effect of the m^{th} bar treatment

λ_m = lingering effect of the m^{th} bar treatment

ξ_{ijk} = random error and is $\sim N(0,\sigma_b^2)$

Table 5-1: Summary of effects influencing response in weeks, adapted from (Shen and Lu 2006).

Team	Bar Order	Week 1	Week 2	Sum	Difference
Gold	Test-Control	$\mu + \pi_1 + \Phi_1$ ($Y_{1\cdot1}$)	$\mu + \pi_1 + \Phi_2 + \lambda_1$ ($Y_{2\cdot2}$)	$Y_{1\cdot1} + Y_{2\cdot1}$	$Y_{1\cdot1} - Y_{2\cdot1}$
Green	Control-Test	$\mu + \pi_1 + \Phi_2$ ($Y_{1\cdot2}$)	$\mu + \pi_1 + \Phi_1 + \lambda_2$ ($Y_{2\cdot2}$)	$Y_{1\cdot2} + Y_{2\cdot2}$	$Y_{1\cdot2} - Y_{2\cdot2}$

If a carry-over effect is considered significant, the data analysis should not include the second test period as the results would be influenced by the first week. This test was performed by comparing the sums of the two teams' total responses, which only differ by the order of the treatments. This is expressed as the null hypothesis H_0: $\lambda_1 = \lambda_2$, which if rejected indicates a significant carry-over effect.

Equation 5-2: Null hypothesis for carry-over effect, proof and simplification adopted from Shen and Lu 2006).

H_0: $Y_{1\cdot1} + Y_{2\cdot1} = Y_{1\cdot2} + Y_{2\cdot2}$

H_0: $\mu + \pi_1 + \Phi_1 + \mu + \pi_1 + \Phi_2 + \lambda_1 = \mu + \pi_1 + \Phi_2 + \mu + \pi_1 + \Phi_1 + \lambda_2$

H_0: $\lambda_1 = \lambda_2$

The treatment effect is calculated only if the null hypothesis for the carry-over analysis is not rejected. A treatment effect is interpreted as one bar having a significantly different effect on the response variable when compared to the other bar, regardless of the order. This test was performed using the difference between the two teams' total responses multiplied by a constant factor to eliminate all other interfering components. This is expressed as the null hypothesis H_0: $\Phi_1 = \Phi_2$ which if rejected would indicate that the direct effects of the bar treatments were not equal.

Equation 5-3: Null hypothesis for treatment effect, proof and simplification adopted from (Shen and Lu 2006).

H_0: ½(Y1.1 - Y2.1) = ½(Y1.2 - Y2.2)

H_0: ½($\mu + \pi1 + \Phi1 - \mu - \pi1 - \Phi2 - \lambda1$) = ½($\mu + \pi1 + \Phi2 - \mu - \pi1 - \Phi1 - \lambda2$)

H_0: $\Phi1 - ½\lambda1 = \Phi2 - ½\lambda2$ ($\lambda1 = \lambda2$ if no carry-over exists)

H_0: $\Phi1 = \Phi2$

Treatment Bars

Table 5-2: Nutritional composition comparison between treatment bars

Component	Test Bar "High Protein RTE"	Control Bar "First Strike"
Total Calories	290 kcal	250 kcal
Fat Calories	120 kcal	50 kcal
Protein	25 g	3.0 g
Carbohydrate	16 g	47 g
Dietary Fiber	-	2 g
Fat	14 g	6 g
Saturated Fat	9 g	1 g
Polyunsaturated Fat	-	3 g
Monounsaturated Fat	-	1 g
Cholesterol	30 mg	0 mg
Sodium	330 mg	75 mg
Total Weight	75 g	65 g

The test bar was the RTE Bar formulated and described in Chapter 3. The control bar was the "First Strike Cran-Rasberry", currently supplied in government issue Meals-Ready-to-Eat MRE First Strike rations. The nutrition labels are presented in the appendix (Figure 7-5). The nutrition panels are available in the appendix (Figure 7-4). The First Strike bars were donated by Alexius International, Inc, Fresno Ca. One hundred and twenty First Strike bars were randomly selected from a box containing over five hundred. These bars were removed from their retail wrapping and then placed inside the same laminate bags as the RTE bars and placed in a refrigerator. The Test bar (RTE) was produced in-house following the method described in Chapter 3. The bars were tested for microbial and compositional specifications (Table 7-6). The bars were packaged in

laminated pouches and stored in a refrigerator until used. The RTE bars used in this study contain the MPP protein described in Chapter 4, and the resulting curd was also analyzed for adherence to microbial and compositional specifications available in the appendix (Table 7-4).

Results

Statistical analysis for the carry-over and treatment effects revealed no significant differences; all p-values were greater than 0.05 (Table 5-3). This indicates that no carry-over effect between the test periods exists. In addition, the chosen blood markers were not affected differently by the treatment bars.

Table 5-3: Mann-Whitey p-value results for blood markers

Blood Marker Response	Carry-Over Effect p-value	Treatment Effect p-value
Erythropoietin	0.90	0.21
Hydrocortisone	0.12	0.12
C - Reactive Protein	0.64	0.08
Creatine Phosphokinase	0.06	0.36
Adolase	0.13	0.92

Body composition data for the different teams reveal that they were not well balanced. The mean weight for the Gold team throughout the study (182lbs) was approximately 14lbs greater than that of the Green team (167lbs) (Table 5-4). The Gold team also had a 3% greater body fat. Both team lost body fat and total weight on average every test period, except the Green team in week 1. Variability appears to be greater in

the test period changes of the Gold team; this is indicated by the larger standard deviation (Table 5-4).

Table 5-4: Summary of body composition measurements over the course of the study (wt in Lbs).

Week	Team	Bar	Measurement	Friday Mean	Monday Mean	Mean Change	Standard deviation
1	Gold	Test	Weight	181.8	181.1	-0.7	2.4
			Fat percent	17.6	16.0	-1.5	3.5
1	Green	Control	Weight	166.6	167.6	1.0	1.8
			Fat percent	14.5	12.7	-1.7	2.6
2	Gold	Control	Weight	181.7	180.0	-1.8	3.6
			Fat percent	17.6	16.0	-1.5	3.5
2	Green	Test	Weight	167.6	167.0	-0.6	2.2
			Fat percent	14.5	12.7	-1.7	2.6

Erythropoietin (EPO)

On first examination of the mean EPO concentration (Table 5-5), the values appear to be very similar, even across the test periods. The mean concentration for the Gold team using the test bar was 8mU/mL while the Green team with the control bar was 9mU/mL (Table 5-5). The variability becomes evident in the mean and median changes between the dates. For example, the mean change for the Gold team with the test bar was -0.47, while the Green team with the test bar was 0.65.

Table 5-5: Summary of blood analysis results for EPO - Range 4 - 27 mU/mL.

Week	Team	Bar	Friday Mean	Monday Mean	Mean Change	Median Change	Standard Deviation
1	Gold	Test	8	8	-0.47	0	3.30
	Green	Control	7	8	0.65	1	2.64
2	Gold	Control	8	9	1.76	3	3.46
	Green	Test	8	9	0.65	1	3.66

Besides the p-value for EPO, the mean effects table and graphs can be used to interpret the direction and trend of the results (Table 5-6). In both weeks the test bar had a lower mean change, despite week 2 having a higher overall response. This is also confirmed by the fact that both Green and Gold teams had the same total treatment effect of 1.3, indicating that the order of the treatment had no effect on the resulting blood response.

Table 5-6: EPO summary of effects table, values are the mean difference for the test period.

	Gold	Green	Totals
Week 1	Test: -0.5	Control: 0.6	0.2
Week 2	Control: 1.8	Test: 0.6	2.4
Totals	1.3	1.3	
Treatment Difference (Test - Control):		-2.2	

This trend can be quickly visualized from the main effects plots (Figure 5-2): the test bar having a lower mean and week 2 having a higher overall mean difference.

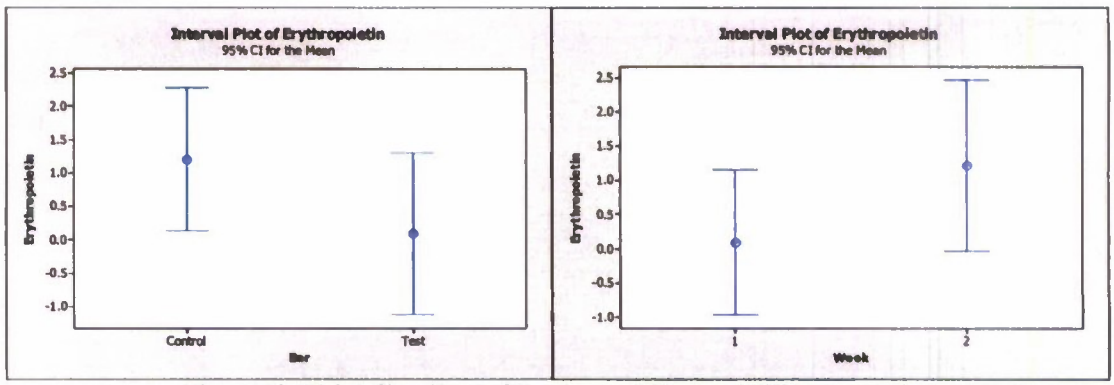


Figure 5-2: Main Effect Chart of Bar type and Week on EOP concentration

Cortisol AM

Three out of the four (team-bar) combinations resulted in a decrease between the baseline and treatment blood draws (Table 5-7). Additionally, the increasing combination (Gold-Test) has the highest mean/median change and standard deviation. This one combination contributed to the lack of treatment effect.

Table 5-7: Summary of blood analysis results for Cotisol AM - Range 6.2 - 19.4 ug/dL.

Week	Team	Bar	Friday Mean	Monday Mean	Mean Change	Median Change	Standard Deviation
1	Gold	Test	15.1	18.4	3.26	3.30	7.36
	Green	Control	16.8	16.4	-0.40	-0.90	5.41
2	Gold	Control	18.1	16.9	-1.18	-0.30	5.06
	Green	Test	16.5	16.0	-0.48	-1.60	4.56

Cortisol AM resulted in a nearly significant carry-over effect (Table 5-3) with a p-value of 0.11. The cause of this can be seen from the lack of a pattern in the summary table (Table 5-8): in week 1 the test bar resulted in a larger mean difference, in week 2 the test bar had a lower mean difference. The total response for the Green and Gold teams is significantly different (2.9 for Gold and -1.7 for Green). The total response each week

is also significantly different, with week 1 having a far greater response versus week 2 (2.1 and -0.9). These large discrepancies indicate that the order of treatment might have an effect on the response.

Table 5-8: Cortisol AM summary of effects table, values are the mean difference for the test period

	Gold	Green	Totals
Week 1	Test: 3.3	Control: -1.2	2.1
Week2	Control: -0.4	Test: -0.5	-0.9
Totals	2.9	-1.7	
Treatment Difference (Test - Control):		4.4	

A basic trend, while not statistically significant, does exist and can be seen in (Figure 5-3). The test bar results indicate a higher response compared to the control bar; the total response was greater for the first week compared to the second.

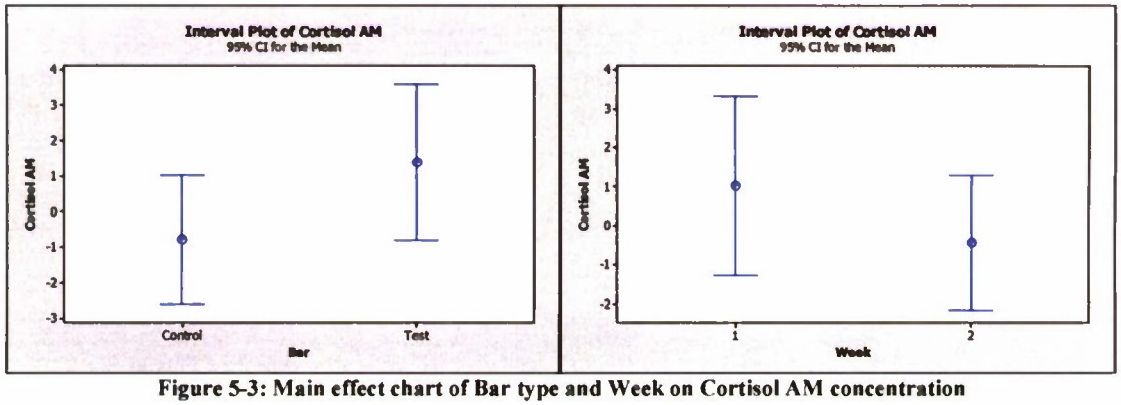


Figure 5-3: Main effect chart of Bar type and Week on Cortisol AM concentration

C - Reactive Protein (CRP)

The individual mean values and mean change show a trend where CRP increases after the treatment period. The mean change and standard deviation for the Green team in

week 1 was significantly greater than the other combinations. This spike reduces the ability to detect a trend.

Table 5-9: Summary of blood analysis results for CRP AM - <1.0 low, >3.0 High

Week	Team	Bar	Friday Mean	Monday Mean	Mean Change	Median Change	Standard Deviation
1	Gold	Test	1.9	2.9	1.05	0.90	0.74
	Green	Control	2.6	2.1	-0.48	0.40	6.01
2	Gold	Control	1.2	2.0	0.76	0.40	2.21
	Green	Test	1.0	1.9	0.94	0.60	1.13

The effect of the bars separated from the week and treatment order effects can be seen by comparing the mean values (Table 5-10). The test bar results show a higher mean difference for both weeks, independent of the order in which they were taken (1.0 verse -0.5) and (0.9 verses 0.8). However the total response for the teams is very different (1.5 verses 0.5), which could indicated an effect of the team.

Table 5-10: CRP summary of effects table, values are the mean difference for the test period

	Gold	Green	Totals
Week 1	Test: 1.0	Control: -0.5	0.6
Week 2	Control: 0.8	Test: 0.9	1.7
Totals	1.8	0.5	
Treatment Difference (Test - Control): 1.7			

The variability within the results is very evident in the main effects plots (Figure 5-4). The test bar resulted in a much smaller spread of data compared to the control, as did week 2. The p-value from the Mann-Whitney test (Table 5-3) indicates CRP as

having the closest to a significant treatment effect at 0.076, with the test bar causing a greater response.

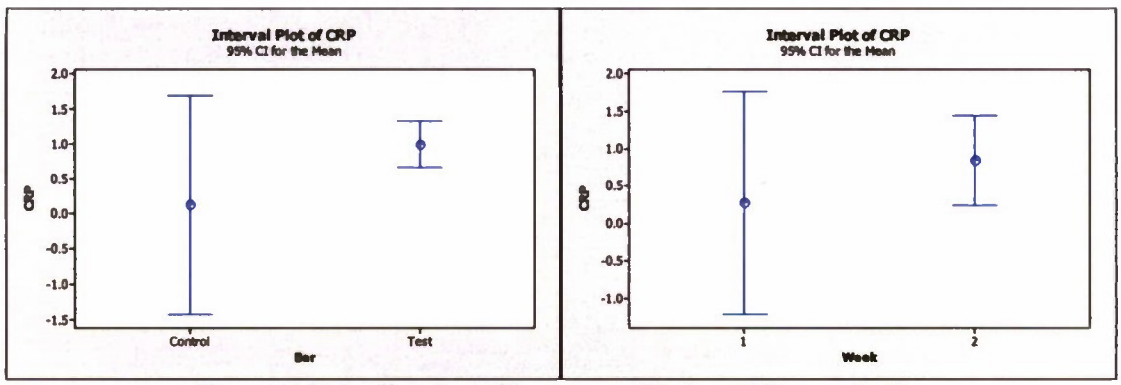


Figure 5-4:Main effect chart of Bar type and Week on CRP concentration

Creatine Kinase (CPK)

The individual mean results display the source of the variability seen in the CPK results. The "Monday" mean for the Gold team in week 1 is 1129 U/L which is significantly greater than the other mean values (260-614 U/L) (Table 5-11). The resulting mean change and standard deviation for that combination are also much greater than the trend set by the other dates.

Table 5-11: Summary of blood analysis results for CPK - Range 35 - 104, U/L

Week	Team	Bar	Friday Mean	Monday Mean	Mean Change	Median Change	Standard Deviation
1	Gold	Test	270	1129	858.47	423.00	1030.18
	Green	Control	222	614	391.35	287.00	326.43
2	Gold	Control	208	374	165.94	112.00	208.23
	Green	Test	274	260	-14.12	33.00	330.48

CPK results indicated the most significant carry-over effect, with a p-value of 0.063. The source of this can be seen by comparing the mean differences with the test bar results having a higher mean difference in the first week (858.5 U/L versus 391.4 U/L) and a lower difference in the second week (-14.1 U/L versus 165.9 U/L). The totals for the teams are also very different, which could be due to the effect of the treatment order or team. In addition, "week" appears to have a significant effect on the results where week 1 resulted in a higher mean difference than week 2.

Table 5-12: CPK summary of effects table, values are the mean difference for the test period

	Gold	Green	Totals
Week 1	Test: 858.5	Control: 391.4	1249.8
Week 2	Control: 165.9	Test: -14.1	151.8
Totals	1024.4	377.2	
Treatment Difference (Test - Control): 287.1			

The resulting large variability and lack of general direction can be seen in the main effects plots (Figure 5-5). The test bar appears to have a higher response; this is shadowed by the large variability in the test sample. Week 2 has less variability and a lower response.

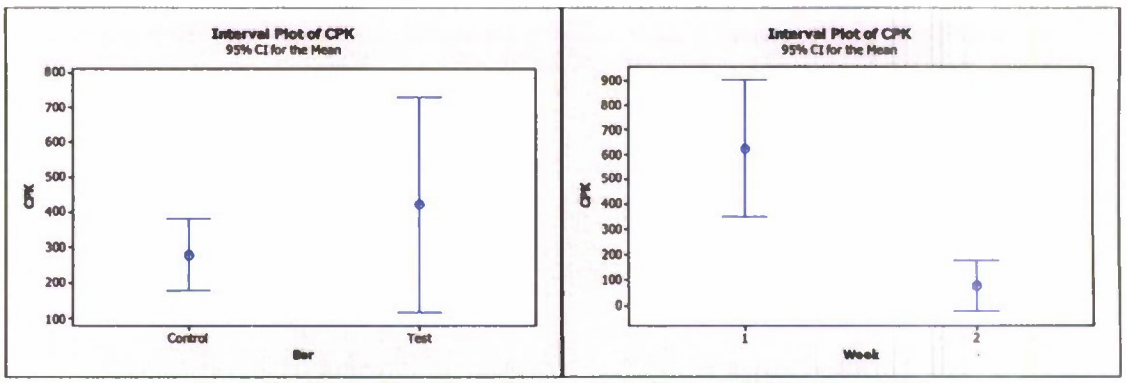


Figure 5-5: Main effect chart of Bar type and Week on CPK concentration

Aldolase

The individual mean responses appear to be separated by week, with week 1 resulting in greater changes than week 2 (Table 5-13). This is true for both treatments and both teams; the variability measured by the standard deviation appears relatively consistent between combinations.

Table 5-13: Summary of blood analysis results for Adolase - range 1.5 - 8.1, U/L

Week	Team	Bar	Friday Mean	Monday Mean	Mean Change	Median Change	Standard Deviation
1	Gold	Test	6.4	11.3	4.91	2.60	4.47
	Green	Control	6.8	10.2	3.39	3.20	2.77
2	Gold	Control	4.9	5.6	0.72	1.10	1.89
	Green	Test	6.6	5.6	-0.94	0.50	4.44

Aldolase resulted in the least significant treatment with a p-value of 0.91 and a close to significant carry-over effect of p-value 0.13 (Figure 5-3). The lack of a treatment effect can be seen in the similarity between the mean differences; the test and control bars had the same response in week 2 and very similar values in week 1. The test bar resulted in a slightly higher response in week 1 and the same response in week 2, which might indicate an order effect.

Table 5-14: Aldolase summary of effects table, values are the mean difference for the test period

	Gold	Green	Totals
Week 1	Test: 4.9	Control: 3.4	8.3
Week 2	Control: 5.6	Test: 5.6	11.3
Totals	10.5	9.0	
Treatment Difference (Test - Control): 1.5			

The variability within the test bar treatments' results is much greater than that of the control bar. However, their mean values appear quite similar indicating no treatment effect (Figure 5-6). The weeks have a similar level of variability with week 1 having a larger mean response than week 2.

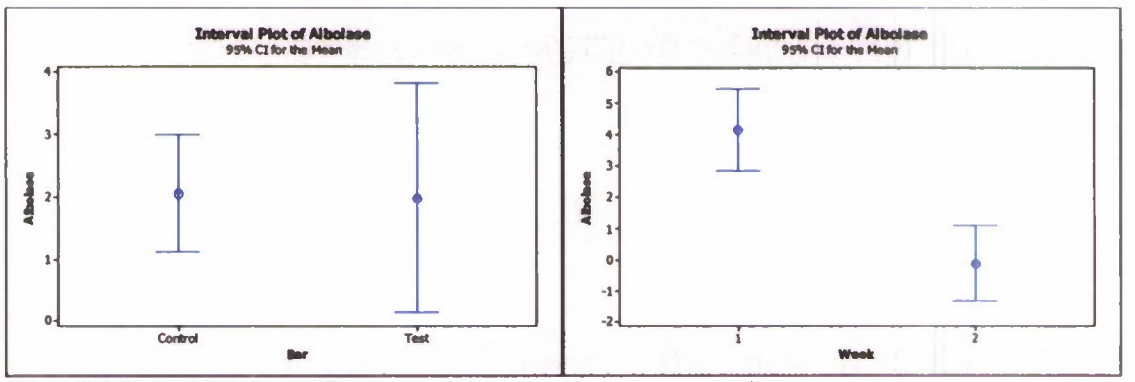


Figure 5-6: Main effect chart of Bar type and Week on Aldolase concentration

Discussion

The data indicated that generally an increase in the particular hormone occurs over the course of test period (Table 7-3). An increase over the test period indicates inflammation has occurred as a result of the physical treatment. An increase in blood marker concentration occurs a majority of the time, in some instances there is a decrease in one or both weeks. This factor contributes to the variability in the statistical analysis, and is likely caused by confounding factors.

Outliers were considered a potential problem in the initial analysis of the data. Conventional statistical analysis using either GLM or t-test models could not be applied because of lack of normality in several of the blood marker responses. One proposed solution was the removal of potential outliers which would cause the data to fit the normality assumption. A small survey of the data revealed that this would not be practical.

Table 7-2 in the appendix summarizes outliers defined as values greater or less than two standard deviations from the mean. The number of individual considered as outliers from this comparison was significant, the outliers were not consistent across markers or blood draws. The outliers and the naturally large variability in the data led to the use of the non-parametric analysis tool Mann-Whitney.

As mentioned earlier, the possibility of a carry-over effect is one of the major complications with a cross-over experimental design (Woods et al. 1989). The possibility of this type of influence is significant in drug and therapeutic exercises where the treatments have lingering effects. An estimated washout period of 5x the half-life of the treatment has been recommended for crossover experiments (Shen and Lu 2006). The treatments used in this study are macronutrients (protein, carbohydrates) which are normally consumed by the individuals. Considering the regular digestion time for most individuals, the one-week rest period should have exceeded these recommendations. While there appears to be some indication of a carry-over effect in some markers, it is more likely that the "team" and "week" had an influence on the response. In every blood marker the "week" greatly affected the magnitude of the response. While this was not a treatment, and all factors were purposely kept the same, the response appeared to be

influenced. Other factors could exist namely weather differences, motivational and learning changes, and external stressors. These factors might have had an influence over the markers in an unpredictable way.

The results from the Erythropoietin (EPO) analysis did not indicate a carry-over effect (Table 5-3) with a p-value 0.904. The role of EPO in the body as an indicator of red blood cell production could signal muscle anabolism as well as physical trauma (Sirén et al. 2001). The trend in the main effects plots (Figure 5-2) indicated that EPO levels increased more with the control bar than the test bar. This could be interpreted as the result of the control bar increasing blood production in response to the physical damage of the test period. It could also indicate that the test bar provided a physiological buffer in repairing damaged muscle, which reduced the "exaggerated" response of the control bar.

Subjects were allowed to schedule their after treatment (Monday) blood draw at their convenience during the open hours of the University Health Center (8am-4pm). This factor could have influenced the response of some blood markers. Cortisol levels are greatly affected by time, with increases seen most dramatically 30 minutes after waking up (Wust et al. 2000). This natural fluctuation in the concentration could lead to treatment effects being ignored. Requiring all subjects to return at the same time for each blood draw may have avoided some of the in-subject variation. Cortisol AM was close to causing a carry-over effect with a p-value of (0.12) (Table 5-3). This is likely due to cortisol time dependence instead of an actual lingering effect. The main effects trend indicated that the test bar results had higher levels of cortisol, which is an indicator of stress. However, there was no statistical support for this trend.

C - reactive protein (CRP) had the lowest p-value of any tested marker (0.076) and no indication of a carry-over effect (0.64) (Table 5-3). CRP levels appeared to be higher when the subjects used the test bar versus the control bar. CRP levels increase after the test period in all instances except "Green Team - week 2", (Table 7-3). CRP is related to inflammation and the body's response to physical damage (Thompson et al. 1999). This increase is justified by the inflammation that would have occurred as part of the physical activity during the test period. The main effect plot of the CRP (Figure 5-4) shows the previously stated trend, but in addition, much greater variability in the control bars' results. This variability was caused by one individual outlier (Table 7-2) and only in "week 1", which might have been caused by an acute and unreported illness or injury.

Creatine phosphokinase (CPK) is the only marker which could be said to have a carry-over effect. While the p-value (0.062) was greater than the preset alpha value of 0.05 there is an increase possibility for week 2 values to be influence by week 1. The mean difference results (Table 5-12) indicates a large difference in means between weeks, week 1 resulting in significantly greater response then week 2, (1250 verse 150). In addition, order of the treatment bar appears to have an effect. In week 1 the test bar resulted in a greater response and in week 2 the response was lower than the control. This fluctuation in the results is likely due to the natural variability of this marker. There are also a large number of outliers (Table 7-2) for this marker. This would indicate that CPK is unaffected by the treatment, despite the trend that is seen in the main effects plots (Figure 5-5), but heavily affected by week.

Adolase levels are directly related to diet, underfeeding can reduce levels while carbohydrate-rich diets can quadruple levels (Munnich et al. 1985). Our results indicated

no significant difference between the treatment bars with a p-value of 0.99. This could be interpreted as neither group being underfed carbohydrates during this study. This indicates that the reduced carbohydrate content in the test bar did not lead to any deficiency in carbohydrates for either team.

The body compositional data did not yield any meaningful trends that could indicate any weight or body fat losses associated with either treatment bar. This could have been caused by a lack of balance or subject pairing between the teams. Three subjects who had a weight over 200lbs were randomly placed in the Gold team. The effect of this is evident from the Gold team having a greater body fat percentage (17.6% versus 14.5%) on the first Monday and being on average 14lb heavier throughout the study. Both teams lost more weight during the second week of the study; while the Green team actually gained one pound in weight during the first period. Both teams lost the same amount of body fat in each period, which indicated no effect of week or treatment on body fat level.

Conclusion

This preliminary study showed that none of the selected blood markers showed significant differences among the treatment bars over the course of the study. Individuals did not show any signs of improvements or under nourishment from either bar. While there are no statistical correlations, some trends are apparent. EPO levels decreased, CRP levels increased and Adolase levels appear unaffected by test bar consumption. The lack of statistical support for these trends is due to the variability in the results, which is

caused by: small sample size (n=34, 17 each treatment), short test period, team balancing, and subject controls.

The experimental design could be improved to produce more tangible results. The first priority would be a larger sample size, with a minimal of 30 individuals per treatment group, which would allow for the identification of smaller differences between the treatments. Diet and exercise controls for the subjects could help to reduce the outliers seen in this study. As a method of reducing variability, the diets and exercise of the subjects can be controlled so that all groups receive the same calories and physical activity outside of the treatment period. A longer experimental period could also help to distinguish the treatment effects. A longer exercise period could lead to more significant exhaustion, inflammation, and muscle catabolism which is what the RTE bar was developed to reduce. Another factor influencing the response is the time of the post-treatment blood draw. This occurred at earliest eighteen hours after the end of the treatment period, this time might have already reduced the inflammation response. The cross-over design and Mann-Whitney test were effective in the analysis of this study. However more subjects would allow for a randomized complete block design and more traditional ANOVA and GLM statistical analysis.

6. REFERENCES

Anonymous. 1994. Food Components to Enhance Performance: An Evaluation of Potential Performance-Enhancing Food Components for Operational Rations,

Food Components to Enhance Performance: An Evaluation of Potential Performance-Enhancing Food Components for Operational Rations ed. Washington D.C: National Academy Press.

Anonymous. 1997. Dietary Reference Intakes for Calcium, Phosphorusm Magnesium, Vitamin D, and Fluoride. Washington, DC: National Academy Press.

Anonymous. 2005. DIETARY REFERENCE INTAKES FOR Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. Washington D.C: THE NATIONAL ACADEMIES PRESS.

Anonymous. 2006. NUTRIENT COMPOSITION OF RATIONS FOR SHORT-TERM, HIGH-INTENSITY COMBAT OPERATIONS. Washington D.C: National Academy of Sciences.

Arthur C. Ouwehand SJS. 1998. The Health Effects of Cultured Milk Products with Viable and Non-viable Bacteria. International Dairy Journal 8:749-758.

Benardot D & Thompson R, Walter. 1999. Energy from Food for Physical Activity: Enough and on Time. ACSM'S Health & Fitness Journal 3(4):4.

Bos C, Gaudichon C & Tome D. 2000. Nutritional and Physiological Criteria in the Assessment of Milk Protein Quality for Humans. Journal of the American College of Nutrition 19(2):14.

Catherine Stanton GG, Hillary Meehan, Kevin Collins, Gerald Fitzgerald, P Brendan Lynch, Paul Ross. 2001. Market Potential For Probiotics. The American Journal of Clinical Nutrition 73(2):476S - 483s.

Committee on Optimization of Nutrient Composition of Military Rations for Short-Term H-SS. 2006. NUTRIENT COMPOSITION OF RATIONS FOR SHORT-TERM, HIGH-INTENSITY COMBAT OPERATIONS. Washington D.C: National Academy of Sciences.

Cribb P, J. 2006. "Sarcopenia and whey proteins, implications, mechanisms and potential for nutritional intervention". In: Council, U. S. D. E., editor). APPLICATIONS MONOGRAPH SARCOPENIA AND WHEY PROTEINS. U.S. Dairy Export Council®

2101Wilson Boulevard / Suite 400

Arlington, VA 22201-3061U.S.A. p. 16.

Demling H, Robert. & DeSanti L. 2000. Effect Of a Hypocaloric Diet, Increased Protein Intake and Resistance training on Lean Mass Gains and Fat Mass Loss in Overweight Police Officers. Annals of Nutrition & Metabolism 44:21-29.

Dockray GJ. 2009. Cholecystokinin and gut–brain signalling. Regulatory Peptides 155(1-3):6-10.

Evans J, William. 2010. Skeletal Muscle Loss: Cachexia, Sarcopenia, and inactivity American Journal of Clinical Nutrition 91(4):1123S-1127S.

Evans W, J., Morley J, E., Argilés J, Bales C, Baracos V, Guttridge D, Jatoi A, Kalantar-Zadeh K, Lochs H, Mantovani G, Marks D, Mitch W E, Muscaritoli M, Najand A, Ponikowski P, Rossi Fanelli F, Schambelan M, Schols A, Schuster M, Thomas D, Wolfe R & Anker S. 2008. Cachexia: a new definition. Clinical Nutrition 27(6):793-799.

Fernández F, M., Boris S & Barbes C. 2003. Probiotic properties of human lactobacilli strains to be used in the gastrointestinal tract. Journal of Applied Microbiology 94(3):449-455.

Gibala J, Martin. 2006. Effect of Exercise on Skeletal Muscle Protein and Amino Acid Metabolism in Humans,2 ed. Champagne IL: Human Kinetics Inc.

Gibala J, Martin. 2007. Protein Metabolism and Endurance Exercise. Sport Med 37(4-5):337-340.

Ha E & Zemel M, B. 2003. Functional Properties of Whey, Whey Components, and Essential Amino Acids: Mechanisms Underlying Health Benefits For Active People (review). Journal Of nutritional Biochemistry 14(5):7.

Hartman W, Joseph., Tang E, Jason., Wilkinson B, Sarah., Tarnopolsky A, Mark., Lawrence L, Randa., Fullerton V, Amy. & Phillips M, Stuart. 2007. Consumption of fat-free fluid milk after resistance exercise promotes greater lean mass accretion than does consumption of soy or carbohydrate in young, novice, male weightlifters. American Journal of Clinical Nutrition 87(2):373-381.

Hazen C. 2008. Bar Basics. Food Product Design. December ed. 3300 N. Central Ave., Ste 300, Phoenix, AZ 85012: VIRGO PUBLISHING. p. 8.

Hazen C. 2009. Customized Bar Fortification. Food Product Design. March ed. 3300 N. Central Ave., Ste 300, Phoenix, AZ 85012: VIRGO PUBLISHING. p. 9.

Hazen C. 2010. Texture Solutions for Snack Bars. Food Product Design. 3300 N, Central Ave., Ste. 300, Phoenix, AZ 85012: VIRGO PUBLISHING.

Hoyt RW, Jones TE, Rose MS, Forte J, V. A., Durkot MJ, Askew EW, Briggs JL, Taub IA, Hintlian CB, Dunne CP, Kluter R & Sikes A. 1991. Level of Dietary Fat Does Not Affect Fuel Oxidation or Endurance Exercise Performancc of Soldiers. In:

MEDICINE, A. R. I. O. E., editor). Natick MA: United States Army Medical Research and Development Command. p. 75.

Jennings T. 1999. Lyophilization : Introduction and basic principles,First ed. Boca Raton, Florida: CRC Press LLC.

Jentjens L, P, G, Roy., van Loon C, L, Luc., Mann H, Christopher., Wagenmakers M, J, Anton. & Jeukendrup W, Asker. 2001. Addition of protein and amino acids to carbohydrates does not enhance post exercise muscle glycogen synthesis. J Appl Physiol 91(2):839-846.

Kimball S, R, Farell P, A & Jefferson L, S. 2002. Role of Insulin in Translational Control of Protein Synthesis in Skeletal Muscle by Amino Acids or Exercise. Journal Of Applied Physiology 93(3):12.

Koch G, Gary. 1972. The Use of Non-Parametric Methods in the Statistical Analysis of the Two-Period Change-over Design. Biometrics 28(2):7.

Kosikowski V, Frank. 1982. Cheese and Fermented Milk Foods. Yew York: Kosikowski, F. V. LLC.

Layman D, K, Evans E, Baum J, I, Seyler J, Erickson D, J & Boileau R, A. 2005. Dietary Protein and Exercise have additive Effects on Body Composition During Weight Loss in Adult Women. The Journal Of Nutrition 135(8):7.

Lemon P, W,R. 1987. Protein and Exercise: Update 1987. Medicine & Science in Sports & Exercise 19(5):11.

Lemon P, W,R & Mullin J, P. 1980. Effect of Initial Muscle Glycogen Levels on Protein Catabolism During Exercise. Journal Applied Physiology 48(4):5.

Levenhagen DK, Carr C, Carlson M, G, Maron D, J, Borel MJ & Falkoll P, J. 2002. Post Exercise Protein Intake Enhances Whole-Body and Leg Protein Accretion in Humans. Medicine & Science in Sports & Exercise 34(5):14.

Malavolti M, Battistini C, Nino., Dugoni M, Bagni B, Bagni I & Pietrobelli A. 2008. Effect ofIntense Military Training on Body Composition. The Journal of Strength and Conditioning Research 22(2):503-508.

Marriott M, Bernadette. 1995. Not Eating Enough: Overcoming Underconsumption of Military Operational Rations. Washington, D.C.: National Academic Press.

Montain J, Scott & Young J, Andrew. 2003. Diet and physical performance. Appetite 40:255-267.

Munnich A, Besmond C, Darquy S, Reach G, Vaulont S, Dreyfus J-C & Kahn A. 1985. Dietary and Hormonal Regulation of Aldolase B Gene Expression. Journal of Clinical Investigation 73(3):7.

Nemet D & Eliakim A. 2007. Protein and amino acid supplementation in sport. International Sportmed Journal 8(1):11-23.

Nisevish M, Pamela. 2008. Nutrition For the Long Run. IDEA Fitness Journal 5(8):60-69.

Oetjen G-W & Hasely P. 2004. Freeze Drying,2 ed. Grunstadt, Germany: WILEY-VCH Verlag GmbH & Co. KGaA, Weinhelm.

Phillips M, Stuart., Moore R, Daniel. & Tang E, Jason. 2007. A Critical Examination of Dietary Protein Requirements, Benefits, and Excesses in Athletes. International Journal Of Sports Nutrition and Exercise Metabolism 17:s58-576.

Phillips S, M, Hartman J, W & Wilkinson S, B. 1998. Dietary protein to support anabolism with resistance exercise in young men. Journal Of American College of Nutrition 8(2):17.

Runestad T. 2004. In search of the perfect low-carb bar. Functional Ingredients Magazine. December 01, 2004 ed.: New Hope Natural Media.

Rutherfurd M, S. & Moughan J, P. 1998. The Digestible Amino Acid Composition Of Several Milk Proteins: Application of a New Bioassay. Journal Of Dairy Science 81(4):8.

Shag N, P. 2000. Effects of Milk-Derived Bioactive: an Overview. British Journal of Nutrition 84(1):7.

Shen D & Lu Z. 2006. Estimate Carryover Effect in Clinical Trial Crossover Designs. Paper PO16. p. 7.

Shimomura Y, Murakami T, Nakai N, Nagasaki M & Harris A, Robert. 2004. Exercise Promotes BCAA Catabolism: Effects Of BCAA Supplementation On Skeletal Muscle During Exercise. The Journal of Nutrition 134:1583S-1587S.

Sirén A-L, Fratelli M, Brines M, Goemans C, Casagrande S, Lewczuk P, Keenan S, Gleiter C, Pasquali C, Capobianco A, Mennini T, Heumann R, Cerami A, Ehrenrcich H & Ghezzi P. 2001. Erythropoietin prevents neuronal apoptosis after cerebral ischemia and metabolic stress. Proceedings of the National Academy of Sciences of the United States of America 98(7):5.

Spreer E. 1998. Milk and Dairy Product Technology,2 ed. New York, New York: Marcel Dekker, Inc.

Tharion J, William., Liberman R, Harris., Montain j, Scott., Young J, Andrew., Baker-Fulco J, Carol., Delany P, James. & Hoyt W, Reed. 2005. Energy requirments of military personnel. Appetite 44:47-65.

Thompson D, Pepys B, Mark. & Wood P, Steve. 1999a. The physiological structure of human c-reactive protein and its complex with phosphocholine. Structure 7:8.

Thompson D, Pepys B, Mark. & Wood P, Steve. 1999b. The physiological structure of human c-reactive protein ant its complex with phosphocholine. Structure 7:8.

Tipton K, D & Wolfe R, R. 1998. Exercise-Induced Changes in Protein Metabolism. Acta Physiologica 162(3):10.

Tipton K, D & Wolfe R, R. 2004. Protein and amino acids for athletes. Journal Of Sports Sciences 22(1):13.

van Loon L, J, Saris W, HM, Verhagen H & Wagemankers A, J. 2000. Plasma Insulin Responses after Ingestion of Different Amino Acid or Protein Mixtures with Carbohydrates. The American Journal Of Clinical Nutrition 72(1):9.

Wallimann T & Hemmer W. 1994. 111-2 Creatine kinase in non-muscle tissues and cells. Molecular and Cellular Biology 133/134:27.

Walzem R, L, Dillard C, J & German J, B. 2002. Whey Components: Millennia of Evolution Create Functionalities For Mammalian Nutrition: What We Know and What We May no Overlooking. Critical Reviews in Food Science and Nutrition 42(4):22.

Woods R, John., Williams G, James. & Tavel M. 1989. The Two-Period Crossover Design in Medical Research. Annals of Internal Medicine 110(7):6.

Wust S, Federenko I, Hellhammer H, Dirk. & Kirschbaum C. 2000. Genetic factors, perceived chronic stress, and the free cortisol response to awakening. Psychoneuroendocrinology 25:13.

Zawadzki K, M, Yaspelkis III BB & Ivy J, L. 1992. Carbonhydrate-Protein Complex increases the rate of muscle glycogen storage after exercise. Journal Of Applied Physiology 72(5):5.

7. APPENDIX

Chapter 3 – Bar Formulation and Manufacturing

LE-PRO® Dairy Product Solids

Leprino Foods

Product Definition

Dairy Product Solids derived from Delactose Permeate.

Chemical Analysis	Typical Values	Specification	Methodology
pH (10% solution)	--	5.2 – 5.6	Combination electrode
Moisture (total moisture)	2.5%	3.0% max.	Atmospherie Ovcn (150° C)
Ash	30.0%	34.0% max.	Residue on Ignition
Sediment	Disc A	Disc B	ADPI

Microbiological Standards

Methodology		Typical Values	Specification
Standard Methods Agar	Total Aerobic Count	< 2,500/g	10,000/g max.
FDA-BAM	Salmonella	Negative	Negative/750g
Red Bile Agar	Coliform	< 10/g	10/g max. Violet
Potato Dextrose Agar	Yeast and Mold	< 10/g	100/g max. Acidified

Particle Size

	Typical Values	Specification
Passing #40 screen	> 98%	98% min.

Nutritional Information*

(Mean/100g)

Calories **	268.00	S
Calories from fat	0.27	Fi
Total Fat (g)	0.03	Pr
Saturated Fat (%)	0.02	Sodi
Trans Fatty Acids (g)	0.00	Calci
Cholesterol (mg)	0.00	I
Total Carbohydrate (g)	59.60	Vitami
		Vitami

*Nutritional results, although based on limited testing, fall within the expected manufacturing ranges.

**1.40 calories/gram

Ingredient Statement

Dairy Product Solids

Physical Characteristics

Appearance: Yellowish powder

Packaging and Storage

Product is packaged in a 500 kg (1100-lb) tote with 3 mil liner, or 22.68 kg (50-lb) bags. Product is recommended to be stored at no more than 80° F (27° C) and relative humidity under 75%.

Specification 1: Delactose Permeate

Fibersol-2

Product Specification & Labels

Product Specification

Total Dietary Fiber	: 90% minimum, dry-weight basis via AOAC Official Method 2001.03
Appearance	: White free-flowing powder
Taste / Odor	: Non-sweet / odorless
Solution	: Clear
Moisture	: 5% max.
Dextrose Equivalent	: DE 8.0 - 12.0 via the WS method
pH	: 4 - 6 in 10% solution
Ash	: 0.2% max.
Arsenic	: 1 ppm max.
Heavy Metals	: 5 ppm max.
Microbiological	
Standard Plate Count	: 300 /g max.
Yeast and Mold	: 100 /g max.
Salmonella	: Negative / 25 g
Coliform	: Negative /g

Labeling Information

US: Maltodextrin (FDA GRAS) Maltodextrin, Resistant Maltodextrin, Digestion Resistant Maltodextrin, Maltodextrin (Fiber), Maltodextrin (Dietary Fiber), Maltodextrin (Soluble Dietary Fiber), Maltodextrin (Source of Soluble Fiber), Maltodextrin (Digestion Resistant Type), Maltodextrin (Fibersol-2), Maltodextrin (Dietary Fiber, Fibersol-2), etc.

EU: Dextrin/ Maltodextrin
JAPAN: Indigestible Dextrin

Caloric Values

Caloric value for soluble dietary fibers varies depending on the regulation in each country. Scientifically, the caloric value for as-is Fibersol-2 is estimated as 1.0-2.0 kcal/ gram. For more specific information, please contact us.

US: 1.6 kcal/ gram
EU: 2.0 kcal/ gram
Australia & New Zealand: 1.9 kcal/ gram
JAPAN: 1.1 kcal/ gram
Korea: 2.0 kcal/ gram

Affirmation as a FOSHU ingredient

- Intestinal regularity (1992)
- Moderating post-prandinal blood glucose levels (1994)
- Lowering serum cholesterol levels (1998)
- Lowering triglyceride levels (1998)
- Recommended intake amount: 3-10 grams/serving

Specification 2: Fibersol 2

Product data

Rev No 04/2006

PURAC® FCC 88

PURAC® FCC 88 is the natural L(+) lactic acid, which is produced by fermentation from sugar. It has a mild acid taste and is widely used as an acidulant in the food industry. PURAC®'s primary functions are to preserve and flavour.

PURAC® FCC 88	
Product	L(+) lactic acid
Form	liquid
Grade	edible special
Colour fresh	max. 50 apha
Colour, 6 months, 25°C	max. 50 apha
Odour	agreeable
Stereochemical purity (L-isomer)	min. 95%
Assay	87.5-88.5% w/w
Density at 20°C	1.20-1.22 g/ml
Sulphated ash	max. 0.1%
Heavy metals total	max. 10 ppm
Iron	max. 10 ppm
Arsenic	max. 1 ppm
Calcium	max. 20 ppm
Chloride	max. 10 ppm
Sulphate	max. 20 ppm
Reducing sugars	passes test FCC
Molecular formula	$CH_3CHOHCOOH$
Molecular weight	90
Chemical name	2-hydroxypropionic acid
CAS number	79-33-4 (general 50-21-5)
EEC Additive number	E 270
USA	GRAS
Complies with	FCC, JSFA and EUSFA

For further information

Http://www.purac.com/

Page 1 of 1

Specification 3: Lactic Acid 88%

Dry Buttermilk (DBM)

Production Definition

Dry Buttermilk is obtained by drying liquid buttermilk that was derived from the churning of butter and pasteurized prior to condensing. DBM has a protein content of **not less than 30.0%**. It may not contain, or be derived from, nonfat dry milk, dry whey or products other than buttermilk, and contains no added preservative, neutralizing agent, or other chemical. DBM for human consumption complies with all provisions of the U.S. Federal Food, Drug, and Cosmetic Act.

Other Characteristics

Scorched Particle Content2..... 7.5 – 15.0 mg
Titratable Acidity2 ..... 0.10 – 0.18%
Solubility Index.....< 1.25 ml – spray process< 15.0 ml – roller process
Color2 .....uniform cream to dark cream
Flavor2 ..... clean and pleasing

Ingredient Statement

"Dry Buttermilk"

Production Applications and Functionality

Bakery products, frozen desserts, prepared dry mixes, beverages, cheese products, frozen foods, dairy products, salad dressings, snack foods

Storage & Shipping

Product should be stored and shipped in a cool, dry environment with temperatures below 80°F and relative humidities below 65%. Stocks should be rotated and utilized within 6 to 9 months.

Packaging

Multiwall kraft bags with polyethylene inner liner or other approved closed container. (i.e. "tote bins," etc.)

Typical Compositional Range1

Percentage

Protein2 ..... > 30.0 – 33.0
Lactose.....46.5 – 49.0
Fat2 .....4.5 – 7.0
Ash.....8.3 – 8.8
Moisture2.....3.0 – 4.0

Microbiological Analysis

Standard Plate Count2 .....< 20,000/g
Coliform.....< 10/g
Salmonella ..... negative
Listeria ..... negative
Coagulase-positive
Staphylococci..... negative

1 On an "as is" basis
2 USDA Grade parameters (7 CFR §58.2654)

-5-

Specification 4: Butter Milk Powder

Nonfat Dry Milk (NDM)

Production Definition

Nonfat Dry Milk is obtained by the removal of water from pasteurized skim milk. It contains not more than 5% moisture (by weight) and not more than 1.5% milkfat (by weight) unless otherwise indicated. NDM for human consumption complies with all provisions of the U.S. Federal Food, Drug, and Cosmetic Act.

Other Characteristics

Scorched Particle Content2..... 7.5 – 15.0 mg
Solubility Index2.....<1.2 ml < 2.0 ml – high-heat
Titratable Acidity2 ..... < 0.15%
Color2 .....white to light cream/natural color
Flavor2 ..... clean and pleasing

Ingredient Statement

"Nonfat Dry Milk" (_____ % milkfat) if the fat content is over 1.5%

Production Applications and Functionality

Fluid milk fortification, frozen desserts, cheese, yogurt, dairy beverages, bakery products, custards, gravies, sauces, frozen foods, packaged dry mixes, processed meats, soups, infant formulas, snack foods, cosmetics Nonfat dry milk is classified for end-product use according to the heat-treatment used in its manufacture.

The classifications are: high-heat, medium-heat and low-heat. *(see page 2)*

Storage & Shipping

Product should be stored and shipped in a cool, dry environment with temperatures below 80° F and relative humidities below 65%. Stocks should be rotated and utilized within 1 to 1 ½ years.

Packaging

Multiwall kraft bags with polyethylene inner liner or other approved closed container. *(i.e. "tote bins," etc)*

Typical Compositional Range1

Percentage

Protein.....34.0 – 37.0
Lactose.....49.5 – 52.0
Fat2 .....0.6 – 1.25
Ash.....8.2 – 8.6
Moisture2.....3.0 – 4.0

Microbiological Analysis

Standard Plate Count2 ..... < 10,000/g
Coliform2 ..... < 10/g
Salmonella ..... negative
Listeria ..... negative
Coagulase-positive
Staphylococci..... negative

---

1 On an "as is" basis
2 USDA Grade parameters (7 CFR §58.2528)

Specification 5: Non Fat Dry Milk

HILMAR INGREDIENTS

WHEY PROTEIN PRODUCT BULLETIN

Issue Date: 7/1/2009

Hilmar™ 8000 Whey Protein Concentrate

Hilmar™ 8000 is a highly functional 80% whey protein concentrate ideal for a variety of food and nutritional applications. It is derived from fresh, sweet dairy whey processed by a special cross-flow filtration process.

FEATURES / BENEFITS

- Good Emulsification
- Egg Replacement
- Minimal Carbohydrate Levels
- Low Glycemic Index
- High Acid Solubility/Stability
- Bland Flavor Profile
- High Digestibility
- Superior Protein Source
- Excellent Amino Acid Profile
- Low Mineral Content
- GMO Free
- Kosher Ⓤ and Halal Approved
- EU Approved

APPLICATIONS

- Protein Bars
- Nutritional Fortification
- Bakery
- Dairy and Frozen Desserts
- Savory, Soups, Sauces
- Salad Dressings
- Meat and Surimi
- Confectionery
- Medical Nutrition

NUTRITIONAL VALUES

Composition	Typical	Specification
Protein (% dry basis)	82.5	80.0 min
Protein (% as is)	78.5	77.5 min
Lactose (%)	5.0	7.0 max
Fat (%)	5.0	7.0 max
Moisture (%)	4.5	5.5 max
Ash (%)	3.0	4.0 max

Microbiology	Typical	Specification
SPC (cfu/g)	<1,000	10,000 max
Coliforms (MPN/g)	<3	<10 max
E. Coli (MPN/g)	<3	<3 max
Salmonella (25g)	Negative	Negative
Yeast (cfu/g)	<10	50 max
Mold (cfu/g)	<10	50 max

Other Nutritional Information	Typical
Cholesterol (mg/100g)	218
Total Calories (Kcal/100g)	390
Biological Value (BV)	104
PDCAAS	1
Protein Efficiency Ratio (PER)	3.2
Net Protein Utilization (NPU)	92
Protein Digestibility	95

MINERALS	Typical
Sodium (mg/100g)	175
Calcium (mg/100g)	550
Potassium (mg/100g)	530
Phosphorus (mg/100g)	350
Magnesium (mg/100g)	60
Chloride (mg/100g)	125
Iron (mg/100g)	1

Specification 6: Whey Protein Concentrate

MPP Compositional and Microbial Specification Sheet

Analysis	Tolerance- Range	Actual
Moisture	35-40%	
Fat	23-27%	
Protein	20-27%	
Coliform	< 10 CFU/ml	
E-coli	< 10000 CFU/ml	
SPC	< 20000 CFU/ml	

Specification 7: Milk Protein Precipitate (MPP)

Flavor profile Analysis

Table 7-1: Bench-top flavor experiments, both sweet and savory.

Date	Company	Flavor ID	Flavor	Process	Notes
5/4/10	Gold Coast	336755	Cranberry	Baked	Low detection, low sweetness
		336957			Better flavor, berry like.
		332912			Strong flavor, almost plastic like
11/6/09	Kraft	21000139800	Cheese	Freeze Dried	+ pretty cheesy.
		210007087900	10% use level		+ bitterness, cheesiness.
		21004003600			- musty or rancid,
		210000110600			- less cheese,
		21007084200			- pizza-like flavor, Italian spices.
		210006935800			+ cheesy but mild.
9/29/10	Firmenich	057637	Vanilla	Freeze Dried	Low flavor, poor coverage.
		059200 AP0551	Artificial Cream+ Vanilla		Better coverage needs higher vanilla.
9/22/10	Firmenich	057622 TP0551	Chicken	Freeze Dried	Bad, heavy roast, unpleasant.
		588734 SPM			Non descript, high salt.
		557075 SPM			To sweet, no flavor.
9/15/09	Firmenich	868519CB + 885023 TTB0440	Pizza Roast Garlic	Freeze Dried	Good flavor, garlic slightly too high, good cracker-like

Chapter 5 – Human Subjects Study

Topographical displays of hikes

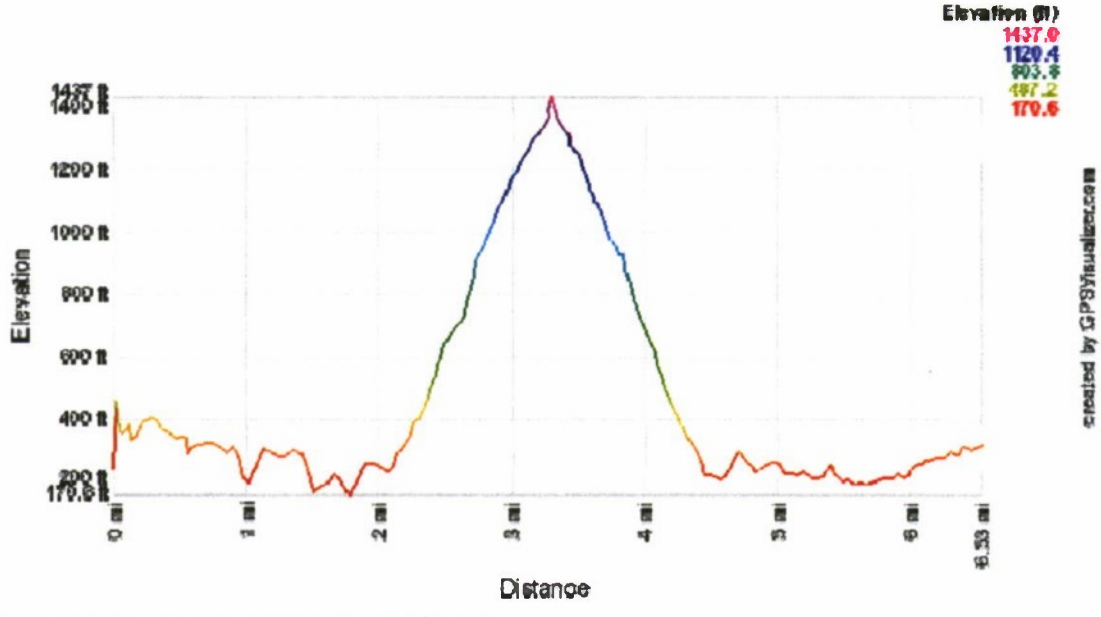


Figure 7-1: Day One hike "Bishop Peak" 5/21 , 6/4

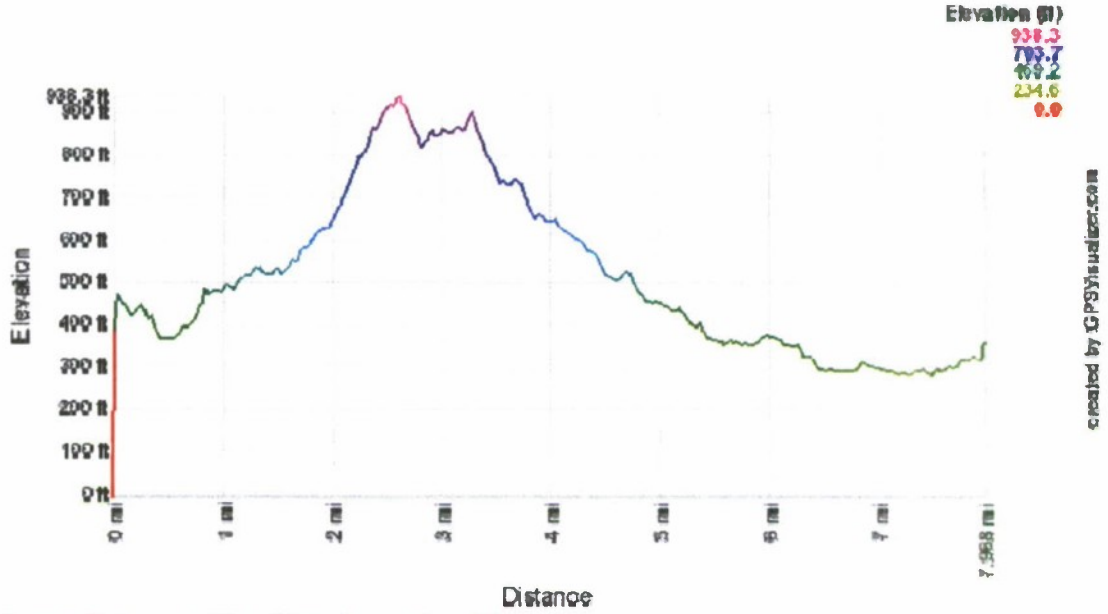


Figure 7-2: Day Two Hike – "Poly Canyon Loop" 5/22, 6/05

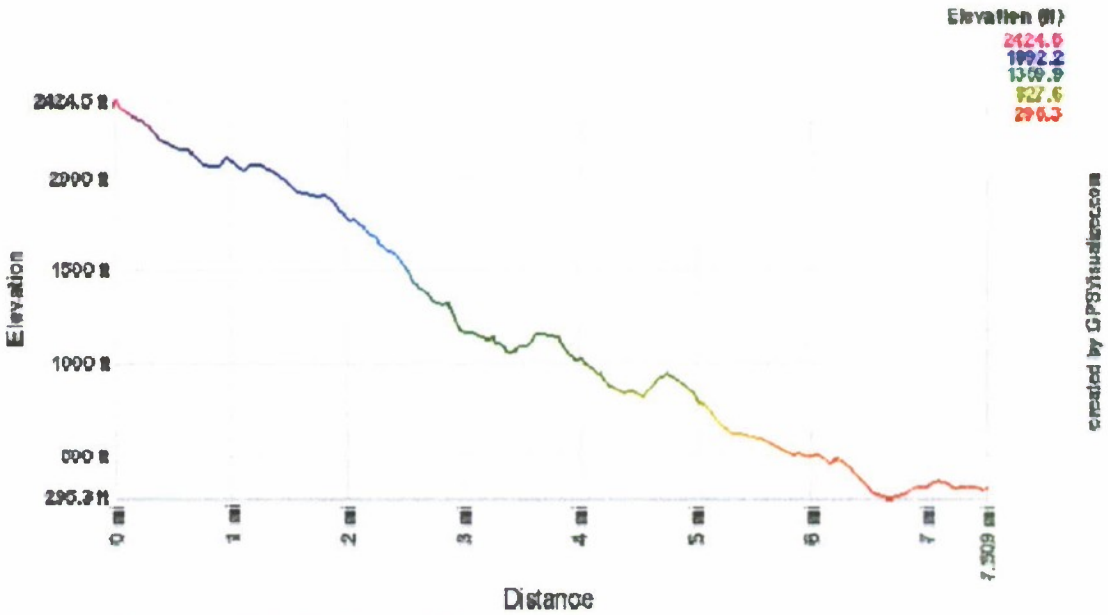


Figure 7-3: Day Three Hike "The Big Down Hill" 5/23, 6/06

Nutrition Panels of Test and Control bars

Nutrition Facts

Serving Size (75g)
Servings Per Container

Amount Per Serving	
Calories 290	Calories from Fat 130
	% Daily Value*
Total Fat 15g	23%
Saturated Fat 9g	45%
Trans Fat 0g	
Cholesterol 30mg	10%
Sodium 25mg	1%
Total Carbohydrate 15g	5%
Dietary Fiber 0g	0%
Sugars 7g	
Protein 25g	

Vitamin A 0% • Vitamin C 10%
Calcium 6% • Iron 0%

*Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

	Calories	2,000	2,500
Total Fat	Less Than	65g	80g
Saturated Fat	Less Than	20g	25g
Cholesterol	Less Than	300mg	300 mg
Sodium	Less Than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Calories per gram
Fat 9 • Carbohydrate 4 • Protein 4

Figure 7-4: Test Bar 'RTE bar" Nutrition Facts Label

Nutrition Facts

Serving Size 1 Bar (65 g / 2.3 oz)

Calories 250
Fat Calories 50

* Percent Daily Values are based on a 2,000 calorie diet

Amount/serving	%Daily Value*
Total Fat 6 g	8%
Saturated Fat 1.0 g	5%
Polyunsaturated Fat 3 g	—
Monounsaturated Fat 1 g	—
Cholesterol 0 mg	0%
Sodium 75 mg	2%
Potassium 80 mg	10%

Amount/serving	%Daily Value*
Total Carbohydrate 47 g	16%
Dietary Fiber 2 g	7%
Soluble Fiber 1 g	—
Insoluble Fiber 1 g	—
Sugars 18 g	—
Protein 3.0 g	—

Vitamin A 0%	•	Vitamin C 35%	•	Calcium 4%	•	Iron 2%	•	Vitamin E 50%
Thiamin 20%	•	Riboflavin 20%	•	Niacin 20%	•	Vitamin B6 20%	•	Folate 50%
Vitamin B12 20%	•	Phosphorus 6%	•	Magnesium 6%	•	Vitamin D 20%	•	Zinc 15%

INGREDIENTS: Raspberry filling (fructose, maltodextrin, water, raspberry concentrate, food starch modified, carrageenan, natural flavors, malic acid), maltodextrin, corn syrup, dried cranberries, crisp corn (degermed yellow corn meal, sugar, malt extract, salt, calcium carbonate, mono and diglycerides), apple nuggets (dried apple pieces, flavor, red 40, blue 1), partially hydrogenated cottonseed/soybean oil, whey protein concentrate, apple powder, rice bran concentrate, glycerin, fructose, natural and artificial flavors, lecithin, vitamin premix (ascorbic acid, DL-alpha-tocopherol acetate, niacinamide, pyridoxine hydrochloride, riboflavin, thiamine mononitrate, folic acid, cholecalciferol, cyanocobalamin, zinc oxide), ascorbyl palmitate added to protect flavor, natural mixed tocopherols added to protect flavor, red 40, blue 1.

Figure 7-5: Control bar Nutrition Facts Label

Table 7-2: Outliers for each marker occurring by week and team, outlier defined as >/< 2stds.

		Subject ID	
Response	Team	Week 1	Week 2
A – Cort	Gold	37	35
	Green	7	
Adolose	Gold	12, 21	6
	Green	11, 24, 3, 2	3
CPK	Gold	12, 21	12, 25
	Green	2, 3, 20, 24,11	3
CRP	Gold	17	
	Green	8	29, 7
EPO	Gold		28
	Green		4

Table 7-3: Effect of test period on blood hormone, separated by team and week.

Marker	Team	Week 1	Week 2
EPO	Gold	No Change	Increase
	Green	Increase	Increase
Cortisol	Gold	Increase	*Decrease*
	Green	*Decrease*	*Decrease*
CRP	Gold	Increase	Increase
	Green	*Decrease*	Increase
CPK	Gold	Increase	Increase
	Green	*Decrease*	Increase
Adolase	Gold	Increase	Increase
	Green	Increase	*Decrease*

Compositional and Microbial results for MPP and Test RTE Bars

Table 7-4: Microbial and Compositional specifications for MPP product

Specification	Tolerance- Range	1a 5/13/10	2a 5/28/10
Moisture	35-40%	*36.9% 36.6%*	*40.2% 43.3%*
Fat	23-27%	*22.9%*	*28%*
Protein	20-27%	*20.95%*	*27%*

Test 1a 5/13/10	Date Read	10^{-1}		10^{-2}	
E- Coli - CC	5/15/10	*0*	*0*	*0*	*0*
Coliform	5/16/10	*1*	*0*	*0*	*0*
SPC	5/16/10	*21*	*24*	*0*	*0*
Yeast/Mold	5/17-5/24	*0*	*0*	*0*	*0*

Test 2a 5/28/10	Date Read	10^{-1}				10^{-2}			
E- Coli - CC	5/29/10 – 5/30/10	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Coliform	5/30/10	*1*	*6*	*1*	*4*	*0*	*0*	*0*	*0*
SPC	5/30/10	*78*		*81*		*2*		*4*	
Yeast/Mold	5/30-6/25	*0*		*0*		*0*		*0*	

Table 7-5: Compositional specifications for RTE bars product

Spec	Tolerance- Range	1a1	1a2	1a3	1a4
Moisture	>24%	*27.22*	*26.22*	*25.66*	*25.66*
aW	>0.890	*0.91*	*0.91*	*0.89*	*0.89*
Protein	<29%	*33.1*	*32.0*	*31.7*	*32.2*

Spec	Tolerance- Range	2a1	2a2	2a3	2a4
Moisture	>24%	*24.4*	*24.0*	*23.3*	*23.8*
aW	>0.890	*0.885*	*0.896*	*0.894*	*0.888*
Protein	<29%	*33*	*38*	*33.6*	*32.6*

Table 7-6: Microbial specifications for RTE bars product

Test	plated	1a1		1a2		1a3		1a4	
E- Coli - CC	10^{-1}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Read 6/2/10	10^{-2}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Coliform	10^{-1}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Read 6/2/10	10^{-2}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
SPC	10^{-1}	*0*	*2*	*2*	*3*	*3*	*0*	*4*	*5*
Read 6/2/10	10^{-2}	*0*	*1*	*0*	*1*	*0*	*0*	*0*	*0*
Yeast/Mold	10^{-1}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Read 6/3-6/25/10									

Test	plated	2a1		2a2		2a3		2a4	
E- Coli - CC	10^{-1}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Read 5/18/10	10^{-2}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Coliform	10^{-1}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Read 5/19/10	10^{-2}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
SPC	10^{-1}	*0*	*3*	*0*	*0*	*1*	*1*	*1*	*0*
Read 5/19/10	10^{-2}	*0*	*0*	*2*	*0*	*0*	*1*	*0*	*1*
Yeast/Mold	10^{-1}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Read 5/24/10									

Blood Marker and Compositional Data

CDC/AHA recommended cut of points <1.0 low , >3.0 High

ID	5/21/10	5/24/10	Delta	6/4/10	6/7/10	Delta	Sum	Estimate
Gold								
27	0.4	1.8	1.4	2.7	1.2	-1.5	-0.1	1.45
12	0.4	1.8	1.4	0.4	0.8	0.4	1.8	0.5
16	0.6	1.5	0.9	0.6	1	0.4	1.3	0.25
32	0.5	1.4	0.9	0.5	0.7	0.2	1.1	0.35
31	1	2.3	1.3	0.5	1.1	0.6	1.9	0.35
37	16.8	17.7	0.9	2.5	11.5	9	9.9	-4.05
25	2.1	3.2	1.1	0.5	1.2	0.7	1.8	0.2
30	0.4	1	0.6	3.7	4.2	0.5	1.1	0.05
6	0.5	0.9	0.4	1.2	0.8	-0.4	0.00	0.4
36	0.7	1.6	0.9	0.5	0.7	0.2	1.1	0.35
4	1.9	1.5	-0.4	0.8	1.1	0.3	-0.1	-0.35
1	0.5	2.4	1.9	0.5	1.1	0.6	2.5	0.65
17	2.2	5	2.8	2.7	4	1.3	4.1	0.75
38	0.6	1.6	1	2	1.7	-0.3	0.7	0.65
35	0.5	0.8	0.3	0.5	0.5	0	0.3	0.15
21	2.2	4.2	2	0.6	1.3	0.7	2.7	0.65
13	0.9	1.3	0.4	0.5	0.7	0.2	0.6	0.1
Green								
18	0.4	1	0.6	0.4	1.1	0.7	1.3	-0.05
2	0.3	0.6	0.3	0.3	0.9	0.6	0.9	-0.15
34	0.6	0.7	0.1	0.4	1	0.6	0.7	-0.25
7	1.7	0.9	-0.8	3.9	8.2	4.3	3.5	-2.55
23	0.3	4.5	4.2	0.4	0.7	0.3	4.5	1.95
5	0.3	1.4	1.1	0.3	0.7	0.4	1.5	0.35
8	34.8	11.6	-23.2	2.8	3.3	0.5	-22.7	-11.85
3	0.4	0.6	0.2	0.6	2.3	1.7	1.9	-0.75
9	0.4	0.9	0.5	0.5	0.6	0.1	0.6	0.2
20	0.5	0.9	0.4	0.8	1.8	1	1.4	-0.3
28	0.3	0.6	0.3	0.4	2	1.6	1.9	-0.65
29	1.3	5.6	4.3	2.3	4.7	2.4	6.7	0.95
39	0.9	0.9	0	1.8	0.9	-0.9	-0.9	0.45
24	0.9	2	1.1	0.9	1.5	0.6	1.7	0.25
11	0.5	0.9	0.4	0.5	1.6	1.1	1.5	-0.35
19	0.3	1.3	1	0.3	1.1	0.8	1.8	0.1
26	0.4	1.7	1.3	0.5	0.7	0.2	1.5	0.55

ID	5/21/10	5/24/10	Delta	6/4/10	6/7/10	Delta	Sum	Estimate
Gold								
27	15.3	10.6	-4.7	15.7	18.4	2.7	-2	-3.7
12	16.6	10.5	-6.1	18.6	14	-4.6	-10.7	-0.75
16	17.1	12.3	-4.8	10.5	16.1	5.6	0.8	-5.2
32	20.8	24.4	3.6	29.4	24.4	-5	-1.4	4.3
31	17.4	19.5	2.1	11.1	8.7	-2.4	-0.3	2.25
37	9.1	28.2	19.1	14.1	9.5	-4.6	14.5	11.85
25	12.8	7.8	-5	23	19.1	-3.9	-8.9	-0.55
30	9.4	22.4	13	19.2	10.5	-8.7	4.3	10.85
6	11.6	19.4	7.8	9.6	9.3	-0.3	7.5	4.05
36	30.5	39.8	9.3	22.7	23.3	0.6	9.9	4.35
4	18.4	15.7	-2.7	17.9	11.2	-6.7	-9.4	2
1	17.4	24.8	7.4	21.1	23.4	2.3	9.7	2.55
17	8.8	12.1	3.3	9.1	15.8	6.7	10	-1.7
38	19.5	24.8	5.3	20	20.6	0.6	5.9	2.35
35	11.5	14.8	3.3	8.2	20.7	12.5	15.8	-4.6
21	2.6	12.6	10	19.4	21.4	2	12	4
13	18.5	13	-5.5	15.2	11.6	-3.6	-9.1	-0.95
Green								
18	16.9	11.5	-5.4	13.7	9	-4.7	-10.1	-0.35
2	13	15.7	2.7	20.7	18.8	-1.9	0.8	2.3
34	18.9	20.4	1.5	18.2	13.2	-5	-3.5	3.25
7	23	10.7	-12.3	13	14.1	1.1	-11.2	-6.7
23	12.7	17.6	4.9	13.8	10.7	-3.1	1.8	4
5	22.8	16.7	-6.1	10.1	17.9	7.8	1.7	-6.95
8	20.2	19.3	-0.9	23.2	15.2	-8	-8.9	3.55
3	10.5	15.5	5	12.8	17.5	4.7	9.7	0.15
9	20.3	24	3.7	14.4	14.8	0.4	4.1	1.65
20	12	14	2	18.6	14.2	-4.4	-2.4	3.2
28	17.7	23.4	5.7	26.5	21.5	-5	0.7	5.35
29	22.9	18.5	-4.4	19.8	18.2	-1.6	-6	-1.4
39	17.3	18.4	1.1	13.2	13	-0.2	0.9	0.65
24	18.8	14.3	-4.5	14.1	11.3	-2.8	-7.3	-0.85
19	15.6	13.5	-2.1	14.5	19.7	5.2	3.1	-3.65
11	22.3	17.6	-4.7	15.5	21.3	5.8	1.1	-5.25
26	22.7	16.5	-6.2	18.3	21.8	3.5	-2.7	-4.85

Albolase range 1.5 - 8.1, U/L

ID	5/21/10	5/24/10	Delta	6/4/10	6/7/10	Delta	Sum	Estimate
Gold								
27	5.3	7.8	2.5	4.1	5.4	1.3	3.8	0.6
12	8.2	21	12.8	6.2	9.5	3.3	16.1	4.75
16	7.2	8.8	1.6	2.8	4.8	2	3.6	-0.2
32	4	5.6	1.6	3.3	3.8	0.5	2.1	0.55
31	6.1	8.7	2.6	3.1	5.8	2.7	5.3	-0.05
37	6.8	11.3	4.5	4.1	5.2	1.1	5.6	1.7
25	7.6	15.7	8.1	4.8	6	1.2	9.3	3.45
30	5.2	6.7	1.5	3.2	3.5	0.3	1.8	0.6
6	8.6	11	2.4	10.2	5.8	-4.4	-2	3.4
36	4.1	5.1	1	3	2.7	-0.3	0.7	0.65
4	7.8	13.5	5.7	5.2	6.3	1.1	6.8	2.3
1	5.5	13.7	8.2	6.8	7.1	0.3	8.5	3.95
17	6.1	8.2	2.1	3.8	5.8	2	4.1	0.05
38	7.5	8.4	0.9	3.5	5.7	2.2	3.1	-0.65
35	3.1	9.3	6.2	4.6	4	-0.6	5.6	3.4
21	6.8	23.7	16.9	9	6.6	-2.4	14.5	9.65
13	8.1	12.9	4.8	5.6	7.5	1.9	6.7	1.45
Green								
18	5.5	10.1	4.6	4	5.8	1.8	6.4	1.4
2	3.8	9.9	6.1	3.3	5.8	2.5	8.6	1.8
34	11.9	11.3	-0.6	3.8	4	0.2	-0.4	-0.4
7	6.9	7.6	0.7	5.1	5.6	0.5	1.2	0.1
23	4.7	10.1	5.4	4.4	4.2	-0.2	5.2	2.8
5	6.6	8.5	1.9	3.8	5.3	1.5	3.4	0.2
8	7.1	7.5	0.4	11.5	5.4	-6.1	-5.7	3.25
3	11.7	17.6	5.9	27.1	11.4	-15.7	-9.8	10.8
9	9.2	10.7	1.5	10.6	5.8	-4.8	-3.3	3.15
20	4.8	9.5	4.7	3.4	5.1	1.7	6.4	1.5
28	7	6.7	-0.3	4	4.2	0.2	-0.1	-0.25
29	6.8	8.7	1.9	6.4	4.7	-1.7	0.2	1.8
39	5.2	7.7	2.5	4.6	5.2	0.6	3.1	0.95
24	6.1	13.6	7.5	4.6	5.5	0.9	8.4	3.3
19	5.3	8.7	3.4	4.8	6.7	1.9	5.3	0.75
11	6.5	15.4	8.9	5.8	6.5	0.7	9.6	4.1
26	6.2	9.4	3.2	4.6	4.6	0	3.2	1.6

Creatine Kinase (CPK total), Range 35 - 104, U/L

ID	5/21/10	5/24/10	Delta	6/4/10	6/7/10	Delta	Sum	Estimate
Gold								
27	94	571	477	85	205	120	597	178.5
12	518	3815	3297	605	1372	767	4064	1265
16	270	564	294	98	154	56	350	119
32	77	285	208	105	177	72	280	68
31	151	507	356	123	152	29	385	163.5
37	127	895	768	95	207	112	880	328
25	580	1935	1355	275	698	423	1778	466
30	117	279	162	90	141	51	213	55.5
6	694	923	229	686	496	-190	39	209.5
36	127	256	129	106	134	28	157	50.5
4	260	1402	1142	207	385	178	1320	482
1	363	1573	1210	375	672	297	1507	456.5
17	198	351	153	114	190	76	229	38.5
38	283	684	401	91	409	318	719	41.5
35	134	557	423	106	173	67	490	178
21	279	3781	3502	243	472	229	3731	1636.5
13	324	812	488	127	315	188	676	150
Green								
18	121	667	546	118	260	142	688	202
2	125	828	703	122	527	405	1108	149
34	433	412	-21	101	129	28	7	-24.5
7	120	407	287	152	128	-24	263	155.5
23	95	263	168	106	155	49	217	59.5
5	163	219	56	53	153	100	156	-22
8	73	156	83	160	152	-8	75	45.5
3	597	1400	803	1916	749	-1167	-364	985
9	413	501	88	465	161	-304	-216	196
20	201	1090	889	162	352	190	1079	349.5
28	365	449	84	289	213	-76	8	80
29	123	401	278	110	207	97	375	90.5
39	65	181	116	85	118	33	149	41.5
24	258	981	723	144	339	195	918	264
11	176	1099	923	221	224	3	926	460
19	86	391	305	113	203	90	395	107.5
26	367	989	622	346	353	7	629	307.5

Eruthtopoietin Range 4 - 27 mU/mL.

ID	5/21/10	5/24/10	Delta	6/4/10	6/7/10	Delta	Sum	Estimate
Gold								
27	6	6	0	4	6	2	2	-1
12	6	9	3	8	7	-1	2	2
16	6	4	-2	5	5	0	-2	-1
32	9	5	-4	7	4	-3	-7	-0.5
31	13	7	-6	9	9	0	-6	-3
37	14	13	-1	10	14	4	3	-2.5
25	10	17	7	15	18	3	10	2
30	7	7	0	8	10	2	2	-1
6	7	7	0	11	6	-5	-5	2.5
36	10	5	-5	6	9	3	-2	-4
4	13	14	1	8	19	11	12	-5
1	9	4	-5	8	11	3	-2	-4
17	8	11	3	9	8	-1	2	2
38	4	5	1	4	7	3	4	-1
35	6	7	1	6	9	3	4	-1
21	7	6	-1	6	9	3	2	-2
13	7	7	0	7	10	3	3	-1.5
Green								
18	7	11	4	8	15	7	11	-1.5
2	9	7	-2	11	13	2	0	-2
34	5	7	2	6	7	1	3	0.5
7	9	6	-3	5	7	2	-1	-2.5
23	6	8	2	8	6	-2	0	2
5	5	6	1	5	6	1	2	0
8	8	5	-3	4	7	3	0	-3
3	9	11	2	13	10	-3	-1	2.5
9	7	6	-1	6	7	1	0	-1
20	3	8	5	12	9	-3	2	4
28	10	8	-2	13	5	-8	-10	3
29	6	4	-2	6	5	-1	-3	-0.5
39	11	9	-2	15	22	7	5	-4.5
24	3	5	2	5	5	0	2	1
19	5	6	1	4	8	4	5	-1.5
11	10	13	3	9	9	0	3	1.5
26	10	14	4	8	8	0	4	2

ID	Wk1 Wt Change	Wk1 % Fat Change	Wk 2 Wt Change	Wk2 C % Fat Change
Gold				
27	0	0.69	0	0.69
12	-1	-10.15	7	-10.15
16	-4	-3.9	-5	-3.9
32	-2	6.05	-3.5	6.05
31	-2.5	-0.06	-2	-0.06
37	1	0.66	0.5	0.66
25	-1	-2.45	-3	-2.45
30	4	-1.13	1	-1.13
6	-2	-4.75	-8	-4.75
36	1	-6.26	0.5	-6.26
4	-3	-2.32	-1	-2.32
1	-1	-1.15	-3	-1.15
17	3	-0.42	-3	-0.42
38	-3	-0.67	-4	-0.67
35	-1	-0.48	0	-0.48
21	-3	0.5	-8	0.5
13	3	-0.36	1.5	-0.36
Green				
18	-2	0.94	-1	0.94
2	1	0.01	2	0.01
34	1	-2.87	0	-2.87
7	0	-1.36	-3	-1.36
23	2	0.01	1	0.01
5	2	-9.35	3	-9.35
8	0	-1.29	-2	-1.29
3	2	-1.72	1	-1.72
9	-3	-0.78	-3	-0.78
20	0	-4.87	-1	-4.87
28	2	-0.83	-4.5	-0.83
29	0	-1.16	-2	-1.16
39	1	-5.17	0.5	-5.17
24	2	0.2	1	0.2
19	3	-0.8	-4	-0.8
11	5	-0.82	-0.5	-0.82
26	1	0.21	1.5	0.21

Protein Expression Profiles as Biomarkers for Exposure to Endocrine Disruptors in Marine Fish

Principal Investigator:

Lars Tomanek, Ph.D., Center for Coastal Marine Sciences
California Polytechnic State University
San Luis Obispo, CA

RESULTS

During the last year we analyzed liver protein samples from fish that were exposed in the laboratory (with support from another grant) to an endocrine disruptor, 4-nonylphenol (4NP), for changes in their proteome. The main objective was to causally link histopathological patterns of liver tissue in fish with high levels of 4NP and DDE (Diehl et al., in review) to one of the chemicals, specifically 4NP, using proteomics. Thus, we were also exploring the usefulness of proteomics to test for the biological effects of a chemical of emerging concern (CEC), as well as the broader applicability of proteomics for testing for such effects in a variety of organisms (for review see Tomanek, 2011). Our results are displayed as protein expression profiles, which are indicators of a cell's physiological state, and provide evidence for the pathological effects of 4NP in fish through the up-regulation of detoxification pathways and the breakdown of extracellular matrix.

Proteomic analysis of liver tissue samples of arrow goby

We used liver tissue samples of arrow goby (*Clevelandia ios*), a common and widely distributed estuarine fish, which were exposed to 50 ppm of 4NP for 60 days under laboratory conditions. The target organ was the liver because it is known to detoxify xenobiotic substances and histopathological investigations have shown that it may be affected by organic pollutants in arrow gobies living in West Coast estuaries (Diehl et al., in review). Liver samples were prepared for analysis by two-dimensional gel electrophoresis following the protocol by Tomanek and Zuzow (2010). In brief, protein samples were prepared using a urea-based homogenization buffer for separation, first, by isoelectric point (pI) and, second, by molecular mass (Link and LaBaer, 2009). Warping of gels and spot detection resulted in a composite proteome map of 72 2D gel images (Fig. 1). Subsequently we analyzed gel images to quantify protein expression patterns, using permutation-based two-way ANOVA and a p-value of 0.02 to limit false positives. Using hierarchical clustering and the two-way ANOVA we detected a number of clusters of co-expressed proteins changing expression during the course of the 60 d exposure (Fig. 2). We applied matrix-assisted laser desorption ionization tandem time-of-flight mass spectrometry (MALDI-TOF-TOF) to identify proteins of interest. The identifications are shown in Figs 1 and 2. Figure 2 shows the identities of proteins that change with time X 4NP levels (interaction effect only).

Discussion

The heat map (Fig. 2) is a first of its kind in regard to the extent of the exposure and the number of tissue samples analyzed in response to an organic pollutant. Although the arrow goby genome is not sequenced, homology search enables us to identify slightly more than 50% of the protein spots of interest. We extrapolated the identifications of a collagen in the high molecular weight range by comparing the mass spectra of an identified isoform with several protein spots in the same "molecular row". It is likely that we did not identify the other collagens due to post-translational modifications.

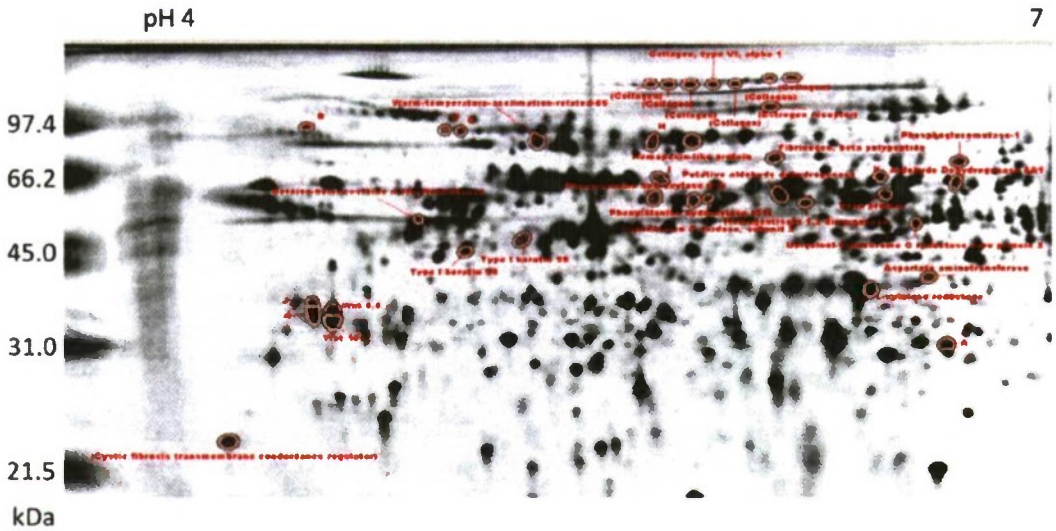


Figure 1: Composite gel image (or proteome map) depicting 623 proteins spots from liver tissue of the arrow goby (*Clevelandia ios*) from all gels of the 60 d time course during exposure to 4-nonylphenol. The proteome map represents the average pixel volume for each protein spot. Labeled spots are those that changed abundance during the time course (two-way ANOVA, p-value 0.02) of the exposure and were identified using tandem mass spectrometry (MALDI-TOF-TOF).

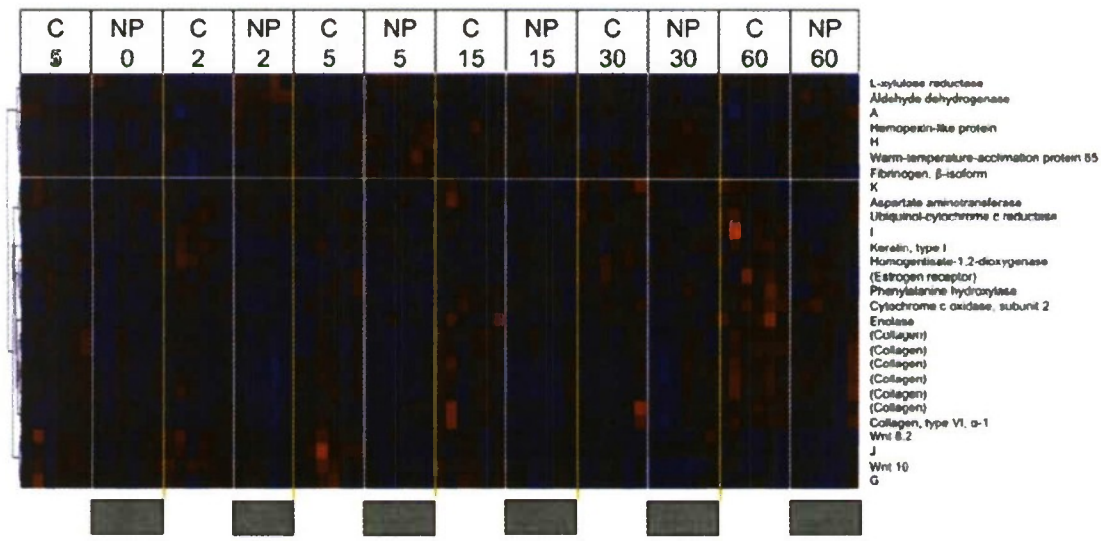


Figure 2: Hierarchical clustering of identified proteins using Pearson's correlation coefficient during the 60-day time course of exposure to 4-nonlyphenol (50 ppm) in the arrow goby (*Clevelandia ios*). Blue coloring represents a lower than average standardized volume, whereas orange represents greater than average standardized volume. The horizontal axis shows the standardized expression patterns of the statistically significant proteins (unidentified proteins are represented by capital letters).

The vertical axis shows the different gel images of the two treatments during the time course.

At this point our interpretation of the data is at a preliminary stage. However, the proteins identified suggest changes in cellular pathways that are associated with detoxification of xenobiotic substances and restructuring (breakdown) of the extracellular matrix.

L-xylulase reductase is part of the glucuronate/xylulase pathway (Salway, 2004) that plays a crucial role during detoxification of organic pollutants, the so-called phase II enzymes, including UDP-glucoronosyltransferases (Schlenk et al., 2008). Phenylalanine hydroxylase, homogentisate-1,2-dioxygenase and aldehyde dehydrogenase are either involved in the hydroxylation of phenol rings (as in phenylalanine) or are part of the phase I group of proteins involved in the oxidation and hydrolysis and therefore excretion of xenobiotic substances (Schlenk et al., 2008). Proteins of phase I and II transform a xenobiotic, first, to a more hydrophilic chemical by adding a polar (e.g., hydroxyl) group, and, second, further enhance secretion by conjugating the chemical with a highly hydrophillic molecule. Overall, reactions of phase I and II proteins increase the polarity of the chemical and thereby water solubility and excretion via the bile or urine (Schlenk et al., 2008). Interestingly, phase II enzymes (L-xylulase reductase) are in a cluster that shows an upregulation of expression with treatment whereas phase I enzymes (e.g., phenylalanine hydroxylase and homogentisate-1,2-dioxygenase) show a downregulation in fish exposed to 4NP. In most cases, this trend is consistent over the entire time period.

The second set of proteins, more than a third of those found significant, are associated with processes that affect the extracellular matrix of liver tissue. There is a consistent signal that all collagen isoforms are downregulated with 4NP treatment. This pattern suggests collagenolysis and breakdown of the extracellular matrix, a process associated with tumorigenesis. Collagenolysis is dependent on the activity of matrix metalloproteinases that are characterized by a hemopexin-like domain (Lauer-Fields et al., 2009). Both, hemopexin-like protein and warm-temperature-acclimation protein 65 (Hirayama et al., 2003), are such hemopexin-like domain containing proteins with putative protease activity. Both putative hemopexin-like domain containing metalloproteases and collagens, have anti-correlated protein expression patterns, suggesting that higher levels of hemopexin-like proteins negatively affect collagen (Lauer-Fields et al., 2009; Munshi et al., 2004; Patterson et al., 2001).

Two proteins consistently associated with tumorigenesis are two Wnt: Wnt 8.2 and Wnt 10 (Takigawa and Brown, 2008; Thompson and Monga, 2007), which we identified (Fig. 2) and that show co-expression with the collagen isoforms. Both Wnt isoforms are downregulated with 4NP, a potential indication of reduced tumor growth potential. However, there are a number of Wnt isoforms and the function of those in fish is not well characterized. Additionally, there is evidence for a role of metalloproteases in interacting with non-canonical Wnt signaling pathways (Blavier et al., 2006; Coyle et al., 2008). In summary, we hypothesize that hemopexin-like protein mediated collagenolysis is a sign of histopathological changes in liver tissue, possibly linked to Wnt signaling. At this point it is unclear if these changes are indicating liver necrosis (as we found in field-

collected arrow goby fish contaminated with high levels of 4NP (Diehl et al., in review)), tumorigenesis or late stage tumorigenesis.

In summary, we have detected changes in the expression of proteins associated with two major cellular pathways in response to a 60 day exposure to 4NP: detoxification pathways of the phase I and II type, although anti-correlated with each other, and proteins associated with hemopexin-like domain containing proteases that activate collagenolysis, possibly in conjunction with Wnt signaling. This is a first of its kind study demonstrating the applicability of proteomics to help assess the biological effects of one chemical of emerging concern in regard to the extent (e.g., time course) of the study and percentage of identified proteins for a non-model organism (Tomanek, 2011). We are looking forward to finishing the interpretation of our data and to writing the results up for publication.

Acknowledgement
Sarah E. Johnson conducted the experiment; Jennifer Diehl provided support and oversight of the project. We thank the Office of Naval Research for their support of our work.

References:

Blavier, L., Lazaryev, A., Dorey, F., Shackleford, G. M. and DeClerck, Y. A. (2006). Matrix metalloproteinases play an active role in Wnt1-induced mammary tumorigenesis. *Cancer Research* **66**, 2691-9.

Coyle, R. C., Latimer, A. and Jessen, J. R. (2008). Membrane-type 1 matrix metalloproteinase regulates cell migration during zebrafish gastrulation: evidence for an interaction with non-canonical Wnt signaling. *Experimental Cell Research* **314**, 2150-62.

Diehl, J., Johnson, S. E., Kang, X., Teh, S. J., West, A. and Tomanek, L. (in review). The ubiquitous distribution of 4-nonylphenol in marine organisms in North American Pacific Coast estuaries: evidence for bioaccumulation. *Environmental Toxicology and Chemistry*, in review.

Hirayama, M., Nakaniwa, M., Ikeda, D., Hirazawa, N., Otaka, T., Mitsuboshi, T., Shirasu, K. and Watabe, S. (2003). Primary structure and gene organization of two types of Wap65 from the pufferfish *Takifugu rubripes*. *Fish Physiology and Biochemistry* **29**, 211-224.

Lauer-Fields, J. L., Chalmers, M. J., Busby, S. A., Minond, D., Griffin, P. R. and Fields, G. B. (2009). Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity. *Journal of Biological Chemistry* **284**, 24017-24.

Link, A. J. and LaBaer, J. (2009). Proteomics: A Codl Spring Harbor Laboratory Course Manual. Cold Spring Harbor: Cold Spring Harbor Press.

Munshi, H. G., Wu, Y. I., Mukhopadhyay, S., Ottaviano, A. J., Sassano, A., Koblinski, J. E., Platanias, L. C. and Stack, M. S. (2004). Differential regulation of membrane type 1-matrix metalloproteinase activity by ERK 1/2- and p38 MAPK-modulated tissue inhibitor of metalloproteinases 2 expression controls transforming growth factor-beta1-induced pericellular collagenolysis. *Journal of Biological Chemistry* **279**, 39042-50.

Patterson, M. L., Atkinson, S. J., Knauper, V. and Murphy, G. (2001). Specific collagenolysis by gelatinase A, MMP-2, is determined by the hemopexin domain and not the fibronectin-like domain. *FEBS Letters* **503**, 158-62.
Salway, J. G. (2004). Metabolism at a Glance. Oxford: Blackwell Publishing Ltd.
Schlenk, D., Celander, M., Gallagher, E. P., George, S., James, M., Kullman, S. W., van den Hurk, P. and K., W. (2008). Biotransformation in Fishes. In *The Toxicology of Fishes*, (ed. R. T. Di Giulio and D. E. Hinton), pp. 153-234. New York: CRC Press.
Takigawa, Y. and Brown, A. M. (2008). Wnt signaling in liver cancer. *Curr Drug Targets* **9**, 1013-24.
Thompson, M. D. and Monga, S. P. (2007). WNT/beta-catenin signaling in liver health and disease. *Hepatology* **45**, 1298-305.
Tomanek, L. (2011). Environmental proteomics: Changes in the proteome of marine organisms in response to environmental stress, pollutants, infection, symbiosis and development. *Annual Review of Marine Science* **3**, in press.
Tomanek, L. and Zuzow, M. J. (2010). The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance and metabolic costs of thermal stress. *Journal of Experimental Biology*, in press.

New technologies with military applications

Project Reports

Full Band Wavelength-Swept SGDBR Laser with MHz Update Rates-Sensing Measurement Applications

Principal Investigator:

Dennis Derickson, Ph. D.
Electrical Engineering
California Polytechnic State University
San Luis Obispo, CA

A. TECHNICAL SUMMARY OF C3RP WORK -OCT MEASUREMENT APPLICATIONS

The single-chip SG-DBR wavelength-tunable laser has been shown[1] to be capable of a 100 kHz repetition rate continuous wavelength ramp or a 10 □s single-shot sweep time. The wavelength ramp consists of fifty 1-nm wide continuous tuning segments that are combined to produce the full 1525 to 1575 nm range. The spectral output of these tunable sources provides single longitudinal mode outputs with a typical spectral linewidth of 100 MHz. The existing high volume production of the SG-DBR laser in the telecommunication industry and small source size gives the opportunity for cost-effective and portable Optical Coherence Tomography (OCT) applications. In many OCT applications, it is beneficial to have a wider source tuning range for enhanced distance resolution in OCT. The InGaAs/InP semiconductor material system used by telecommunications SG-DBR lasers offers design variations that can cover any wavelength in the 1250 to 1650 nm range. This paper shows that adjacent wavelength coverage SGDBR tunable lasers can be easily concatenated due to the unique architecture of the laser. The Semiconductor Optical Amplifier (SOA) that is included as the output segment of this single chip laser design can be used as a fast optical switch to aid the concatenation of adjacent wavelength tuning segments. It will be shown that the concatenation of adjacent wavelength sweep segments can be accomplished with a dead-time of less than 50 ns using the SOA that is monolithically integrated on the laser chip. In order to demonstrate concatenation of SGDBR lasers into an extended frequency ramp versus time a "C-Band" SGDBR laser and an "L-Band" SGDBR laser were characterized. Each single chip SGDBR laser has 5 control segments as illustrated in figure 1.

Figure 1. This is a top view of a single-chip Sampled Grating Distributed Bragg Reflector Laser (SGDBR) used for OCT measurements. Currents in Back Mirror (A), Front Mirror (D), and Phase Section (C), control the laser frequency. Currents in the gain section (B) and the semiconductor optical amplifier (E) control the laser power. The chip dimensions are 0.5 mm by 3 mm.

In order to construct a frequency ramp using SGDBR lasers, the laser wavelength as a function of drive current to the front mirror, back mirror and phase sections needs to be obtained. Figure 2 shows the results of wavelength mapping as a function of drive current to the front and back mirror segments of an L-Band SGDBR laser. Figure 3 shows the results of wavelength mapping as a function of drive current to the front and back mirror segments of a C-Band SGDBR laser.

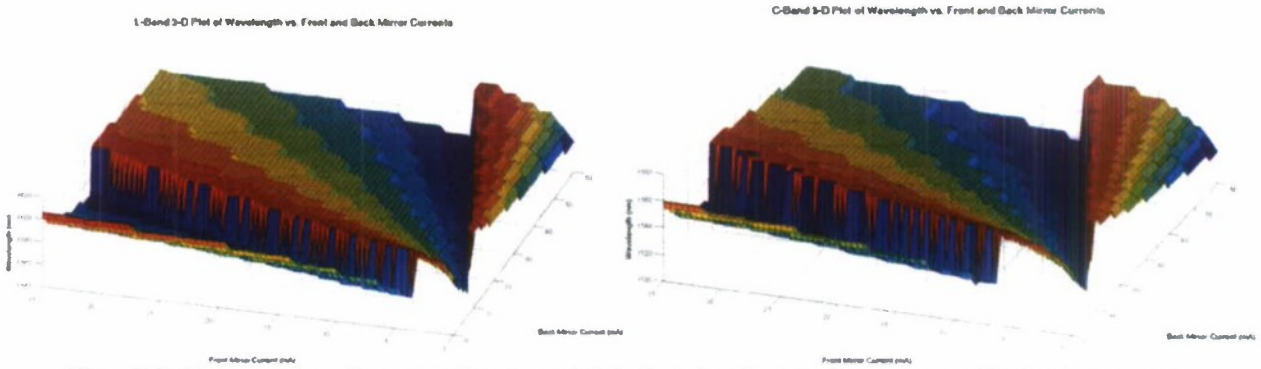


Figure 2 (left) shows a three-dimensional (mode map) plot of wavelength versus front mirror and back mirror currents for the L-band laser. The approximate wavelength coverage is from 1565 nm to 1620 nm. Figure 3 (right) shows a a three-dimensional (mode map) plot of wavelength versus front mirror and back mirror currents for the L-hand laser. The approximate wavelength coverage is from 1525 nm to 1570 nm.

In order to create a linear frequency ramp versus time, one needs to map a sequence for front and back mirror current values that allow the laser to tune over its entire frequency range in a linear ramp versus time. Figure 4 shows the current paths that were chosen in order to tune over the range of the L-Band laser. Figure 5 shows the current paths that were chosen in order to tune over the range of the C-Band laser. In the construction of a linear frequency versus time ramp from the concatenated C and L band lasers it was decided to sweep from lowest frequency to highest frequency. This means that paths 1 through 8 are first transited in figure 4. Paths 1-8 of figure 5 are then concatenated to form the entire frequency ramp.

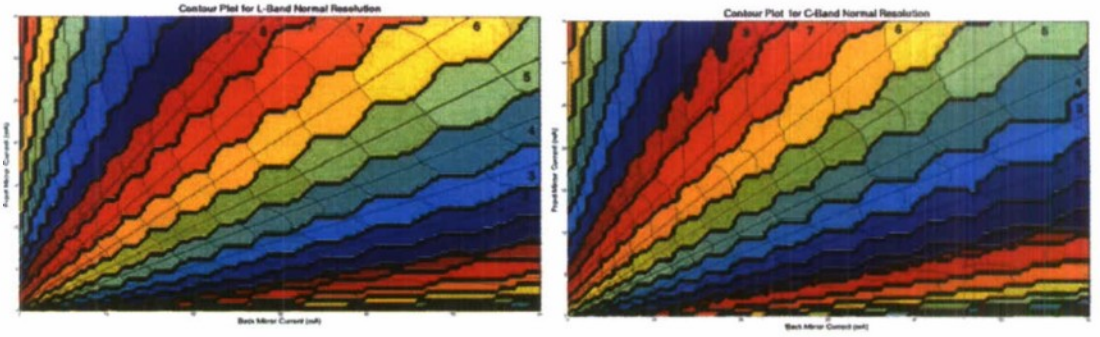


Figure 4 (left) contains similar information to that of figure 2 for the L-Band laser from a top view. Lines 1 through 8 illustrate the path of the current versus time in order to accomplish a linear frequency ramp versus time. Figure 5(right) contains similar information to that of figure 3 for the C-Band laser from a top view. Lines 1 through 8 illustrate the path of the current versus time in order to accomplish a linear frequency ramp versus time.

B. DESCRIPTION OF THE CONCATENATION PROCESS

Figure 6 shows the system block diagram used to generate separate wavelength sweeps from two SGDBR lasers and the method for sweep concatenation. The top laser block is an "L-Band" SGDBR laser that covers the nominal wavelength range of 1565 to 1620 nm. The bottom laser block is a "C-Band" SGDBR laser that covers the nominal wavelength range of 1525 to 1575 nm. Each monolithic SG-DBR laser chip (see figure 1) has three segments that control the laser wavelength; the Back Mirror (BM), Front

Mirror (FM), and Phase (□). Each of the laser segments is driven by an Arbitrary Waveform Generator (ARB) to control the current versus time into each segment of the laser. The ARBs have a clock frequency of 50 MHz with an effective 20 MHz of analog bandwidth. The gain segment is DC biased and provides overall gain for the laser. The Semiconductor Optical Amplifier (SOA) boosts the output power from the laser. The SOA will also be used as a fast optical switch for concatenation of the outputs of the L-Band and C-Band SG-DBR devices. The SOA can also be used to level the power versus time at the output of the laser during the sweep. Each of the ARBs are time synchronized by a master sweep trigger signal. A programmable delay is used to control the switching point in time between the L-Band and C-Band laser operation in order to concatenate the sweep of the two lasers. Finally the outputs of the two lasers are combined in a wavelength independent coupler with the output analyzed by measurement instrumentation. The wavelength coverage of the laser is first monitored in the optical spectrum analyzer. The linearity of the frequency versus time from the laser can be monitored by replacing the optical spectrum analyzer with a repetitive pass band optical filter, a photodetector and an oscilloscope.

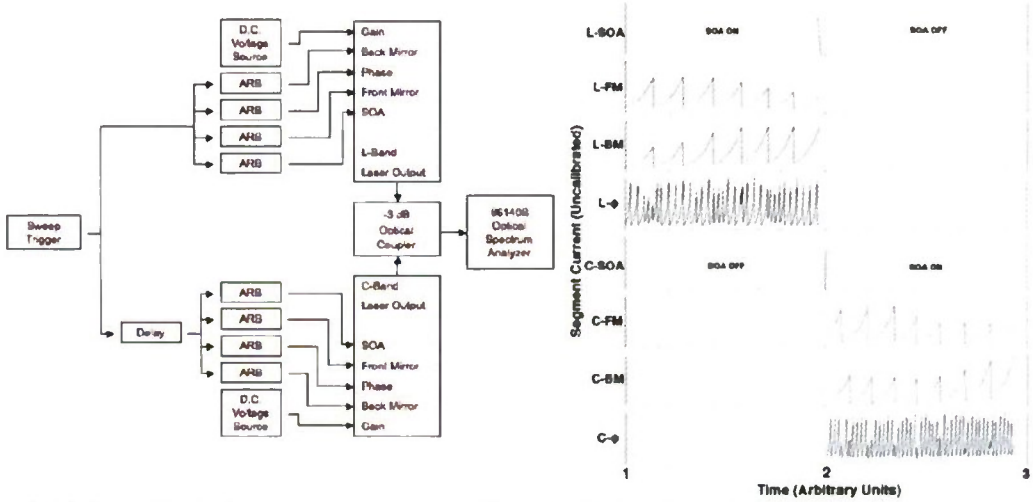


Figure 6 (left). is the block diagram for generating a wide spectral-width linear frequency versus time ramp by concatenating two SG-DBR lasers. Figure 7 (right) is the current versus time into the SG-DBR laser segments for a concatenated C-Band and L-Band wavelength sweep is shown.

Figure 7 illustrates the current versus time that is provided for each segment of the two SGDBR sources. The wavelength tuning map for each SGDBR laser must first be mapped as a function of current in the FM, BM and □ segments[2]. In order to provide a linear ramp in frequency from the laser, the saw-tooth waveforms labeled L-FM, L-BM, and L-□ must be synthesized from the previously generated wavelength-tuning maps shown in figures 2-5. Each vertical abrupt change in the segment current versus time represents an abrupt change in current that causes the laser to jump between different longitudinal modes but with no associated wavelength shift. The individual output from each SG-DBR laser has approximately 50 continuous tuned segments that are joined together for the full wavelength sweep.

Figure 7 also shows the C-FM, C-BM and C-□ waveforms for the C-Band wavelength sweep. In this case the output of the C-Band laser is delayed in time so that the sweep starts after the finish of the L-Band sweep.

The C-Band and L-Band wavelength sweeps associated with the current waveforms shown in figure 7 are shown in Figure 8.

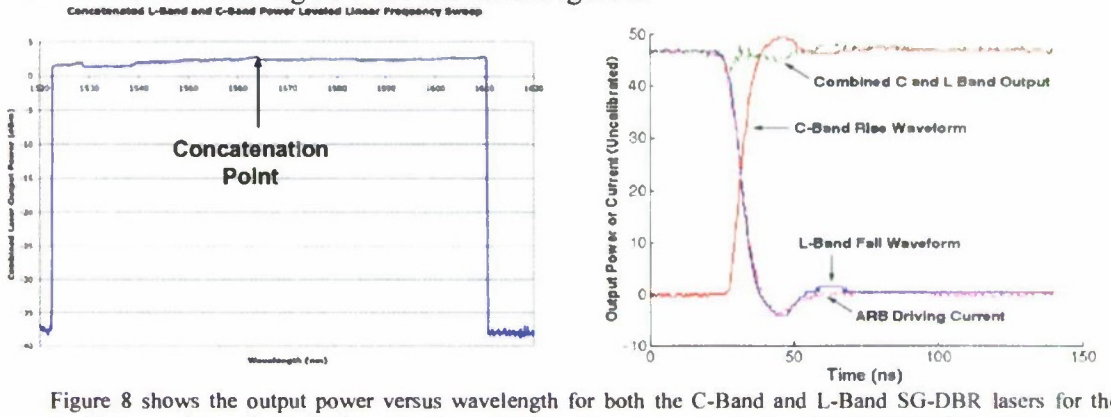


Figure 8 shows the output power versus wavelength for both the C-Band and L-Band SG-DBR lasers for the current drive waveforms shown in figure 7. Figure 9 shows the output power versus time is shown for various outputs in the block diagram of figure 6. The C-Band, L-Band and combined optical output power versus time are shown. In addition the electrical drive current (ARB Driving Current) waveform to the SOA segment is also show

The L-Band wavelength coverage was from 1560.702 nm to 1610.092 nm or 49.4 nm of bandwidth, The C-band laser tuning shape consisted of eight wavelength-tuning paths from 1523.040 nm to 1569.970 nm or 46.9 nm of bandwidth. However, since path 8 of the C-band laser completely overlaps with the L-band laser sweep, it was omitted from the final tuning dictionary. What remained was a constant C-band wavelength sweep from 1523.064 nm to 1564.166 nm or 41.1 nm of bandwidth. A wavelength concatenation point of 1564 nm was used for this experiment.

C. CHARACTERIZING THE L-BAND/C-BAND CONCATENATION POINT

Each SG-DBR laser is configured to construct linear wavelength or frequency sweeps with approximately fifty 1 nm wide modal segments that are stitched together. By concatenating two wavelength sweeps together, we are introducing a new concatenation stitching point between the two lasers. The major difference between the modal stitching points in the laser and the concatenation stitching point is the control segment of the laser. Modal switching points use the FM, BM and □segments of the laser. Concatenation stitching points use the SOA segments of the laser. When the SOA segment current is on, high output power is delivered. When the SOA segment current is off, the laser output is attenuated. Figure 9 illustrates the concatenation stitching process. For the first part of the sweep the L-SOA is on and the C-SOA is off. In the second part of the sweep the C-SOA is on and the L-SOA is off. The switching characteristics of the SOA are key to the successful concatenation of wavelength sweeps. The next two figures will explore the switching speed and the attenuation levels that are possible with the SOA switch that is part of the monolithic structure of the SG-DBR laser.

Figure 9 shows a measurement of the optical power versus time that is measured both at the L-Band Output, the C-Band output and the combined output from the laser. In this

case a high bandwidth photodetector and an oscilloscope replace the OSA measurement block found in Figure 6.

The L-Band optical output power goes from being on to off in a period of less than 50 ns. The C-Band output power goes from being off to on in a period of less than 50 ns. The measurement of the drive waveform to the SOA segment (labeled ARB driving current) illustrates that the rise and fall times are primarily driven by the rise and fall time drive capabilities of the ARB circuitry in this example. Figure 9 also shows the combined optical power from the wavelength independent combiner. A summary from this measurement is that very fast switching times can be accomplished between L-Band and C-Band operations using the SOA segment. The SOA segment is part of the monolithic laser structure so no extra part count (and associated cost) is needed for this switching other than the wavelength independent optical coupler and an additional ARB for each laser.

The SOA switch is not ideal though and there is remnant signal leakage from an SG-DBR laser when the SOA is in its off state. When the SOA segment is in its "on state", it delivers 10 dBm (10 mW) of signal power. The integrated "on state" amplified spontaneous emission (ASE) power of the laser from 1525 to 1575 nm is estimated as -13.5 dBm (45 □W). The "off state" signal power is -40 dBm (0.1□W). The "off state" ASE power is estimated as -41.8 dBm (0.067 □W). The "off state" ASE power at wavelengths longer than 1570 nm has reduced attenuation because the photon energy at these wavelengths is less than the band gap energy of the material in the SOA segment.

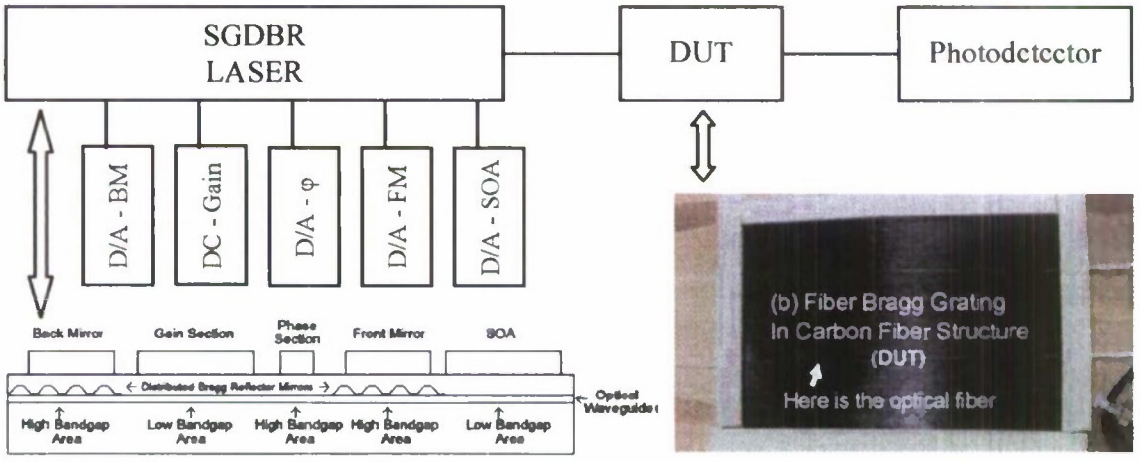
The critical point to be taken from the data of figure 5 is the amount of undesired power that leaks into the output with the SOA off compared to the desired signal. In this case the "on state" would have a signal power of 10 dBm. The undesired leakage signal would be dominated by the -40 dBm source signal from the attenuated laser. So the undesired lasing signal will be 50 dB lower that of the desired laser output. The attenuation of the SOA is sufficient so that it will have a negligible effect on the combined output shown in figure. 1.

D. SUMMARY OF OCT WORK FOR C3RP PROGRAM

The concatenation of two wavelength tunable SG-DBR wavelength sweeps was illustrated. The SOA segment of the SG-DBR laser is key to cost-effective frequency sweep concatenation. The switching time for sweep concatenation was found to be less than 50 ns, limited by the ARB rise time capability. The attenuation of the SOA segment is sufficient so that only a 50 dB attenuated signal leaks into the combined output. The result shows that SG-DBR lasers can successfully concatenated offering better time resolution for OCT measurements. It should be possible to concatenate a large number of adjacent SG-DBR frequency sweeps to cover the 1250 nm to 1650 nm wavelength range that is possible with the SG-DBR laser using the InGaAsP/InP material system. The high ramp repetition rate, low cost and small size of the concatenated source makes this an attractive option for low cost and portable OCT applications with improved distance resolution.

E. SUMMARY OF SENSING APPLICATIONS WORK

Advances in wind power generation require new methods to characterize these new technologies. Twelve foot diameter carbon fiber blades currently under development at the Wind Power Research Center at California Polytechnic State University need a method to describe the stresses due to wind, centripetal, and gravitational forces. Since the blades are designed to rotate at high speed (210 RPM Max), the method must be able to calculate the blade stress at precise locations as it travels around the turbine. The method we propose to characterize these stresses is using a fast wavelength tunable sampled grating distributed Bragg reflector (SGDBR) laser since they generate fast, linear, continuous wavelength sweeps from 1525 to 1565 nm. This sweep is generated through three time synchronized current versus time waveforms applied to the back mirror, front mirror and phase sections of the laser. The sweep consists of fifty separate mode-hop-free tuning segments which are stitched together to form a near continuous wavelength ramp with a repetition rate of up to 100 kHz. At 100 kHz repetition rate, 210 blade RPM corresponds to 0.0126° of blade rotation per test repetition which also relates to 1.6 hundredths of an inch of blade tip movement. Imbedded fiber Bragg grating stress sensors in the carbon fiber turbine blades serve as the device under test (DUT). Stresses in the turbine blade and the imbedded fiber Bragg grating change the spectral response of the DUT. By monitoring the transmitted signal of the DUT with a photodetector one can monitor the stresses in the turbine blades in real time. By analyzing this data the Wind Power Research Center can improve these new carbon fiber wind turbine blades.



F. PUBLICATIONS AS A RESULT OF THIS C3RP FUNDING

1. Brandon George and Dennis Derickson, "High-Speed Concatenation of Frequency Ramps Using Sampled Grating Distributed Bragg Reflector Laser Diode Sources for OCT Resolution Enhancement" in *Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine XIV*, edited by Joseph A. Izatt, James G. Fujimoto, Valery V. Tuchin, Proceedings of SPIE Vol. 7554-96 (SPIE, Bellingham, WA 2010)

2. Brandon George, Shane O'Connor and Dennis Derickson, " Swept-Wavelength Optical Sensor Interrogation with 10 Microsecond Sweep Period Utilizing Sampled Grating Distributed Bragg Reflector Lasers" , Optical Society of America Frontiers in Optics Conference, Oct. 12, 2009, San Jose, CA

3. Brandon George MSEE student, Octavio Rico BSEE student, and Professor Dennis Derickson, "Characterization of Carbon Fiber Wind Turbine Blades Using a Single Chip Wavelength Tunable SGDBR Laser Interrogator", Optoelectronics Industry Development Association OPTOMISM Executive Forum and Conference: Photonics Products for Green Technology, May 18-20 2009, Santa Clara, CA

4. Master's Thesis Documents associated with this C3RP Contract

   a.. Ben Maher " High Speed Wavelength Tuning of SGDBR Lasers for OCT" 2009
   b. Brandon George "Swept Frequency SGDBR Lasers optimized for OCT" 2009

5. Senior Project Reports associated with this C3RP contract
   a. Octavio Rico "Wavelength Tunable Laser System for Fiber Bragg Grating Sensing"(June 2009)
   b. Detrick Reimer "OCT trigger circuit with wavelength stitching gate function" (Feb. 2010)

Design and Development of New High-Current Low-Voltage DC-DC Converters

Principal Investigator:

Taufik
Electrical Engineering
California Polytechnic State University
San Luis Obispo, CA

1. Acknowledgment

I would like to thank Dr. Susan Opava and C^{3}RP for the opportunity to conduct this study. I would also like to thank my graduate student Randyco Prasetyo for his hard work and contributions to the successful completion of this project. Lastly, my thanks go to all of my students who performed the prior study on the same subject which provided the basic proof-of-concepts of the new topologies developed in this study.

2. Executive Summary

Power Electronics is an enabling technology that allows us to convert electrical energy from one form to another in the most efficient manner. Power Electronics encompasses a broad spectrum of power level and applications from the mega-watt of high power transmission to the milli-watt low power distribution in hand-held or portable devices.

One major application that has recently drawn interests among power electronics researchers is microprocessor. Microprocessors present a unique challenge in power electronics due to their power requirements. More specifically, with the increasing number of transistors per microprocessor, its power requirement increases accordingly, but at very low operating voltage (around 1 V). This means the power supply for these microprocessors will not only have to be able to produce a very low voltage but also at very high current. This is the main key challenge addressed in this project.

This project involves the design and prototype development of two new topologies that effectively address the new challenges of powering microprocessors; i.e. low voltage at very high current. The basis of these new topologies lies on strategic placements of passive energy storage components rather than inserting additional converter's stage. Hence, the new topologies offer technical and economical benefits. Yet another benefit, the new topologies may be adopted for any applications other than microprocessors that require both excellent dynamic performance and high density.

The project took up the entire academic year to completion and involved one graduate student. Prior studies conducted by both undergraduate and graduate students of the new topologies developed in this project were done primarily on the computer simulation aspect and developing the first version of hardware prototyping to ensure that the new topologies would in fact work. For these prior studies, design optimization as well as efficiency was neglected. Following encouraging results of these prior studies this project commenced.

The project consisted of the following tasks:

1. Circuit analysis and derivations of design equations of the power stage of each topology
2. Efficiency estimation and design calculations for component selection
3. Development of hardware prototype for each topology
4. Lab testing to compare performance of new topologies with a comparable commercially available converter

One key element in improving the overall performance of the power supply to microprocessors, known as Voltage Regulator Modules (VRM), is to look at how the energy flows from input or source to output or the microprocessor can be managed in the most efficient manner. Based on this, three strategies were adopted in the new topologies developed in this project:

1. Additional path-alternating approach through cell-based interleaving technique to improve heat flow on the circuit board of the VRM
2. Additional and almost direct energy storage paths from input to output of VRM
3. Careful selections of components

Development of hardware prototype for the two new topologies was successful. Four-phase version of each new topology was designed and built so they can be tested against a comparable and commercially available four-phase VRM. Lab measurements were then conducted to assess the performance of the new topologies against the commercial VRM. Results of the study are summarized below:

1. The new topologies show improvements on peak to peak output voltage ripple compared to that measured from the commercially available VRM
2. The new topologies show improvements on dynamic load response compared to that measured from the commercially available VRM
3. The new topologies show improvements on load regulation compared to that measured from the commercially available VRM
4. The new topologies have the same performance on line regulation compared to that measured from the commercially available VRM
5. The new topologies show comparable efficiency compared to that measured from the commercially available VRM. However, the new topologies demonstrated increasing efficiency as the load current is increased beyond the 100% load while the commercial VRM shows the exact opposite. This proves the potential of using the new topologies for even higher output current.
6. The new topologies may still use commercially available multiphase pwm controller with slight adjustments of the placements of the feedback signals. No additional controller is necessary.
7. Additional storage components introduced in the new topologies does not necessarily mean larger board space. This is due to the improved input and output ripple properties of the new topologies, enabling them to utilize smaller and less number of input and output capacitors.

This study has been very challenging. Many technical issues were encountered during the development of the hardware prototypes of the new topologies. These include for examples acquiring components that matched our design calculations to maximize performance and soldering the components many of which came in extremely tiny and small packages. The controller posed particularly the most challenging technical problem since no multiphase pwm controller with interleaving feature is commercially available. However, after thorough study of the multiphase pwm controller was conducted, it was found that the commercially available controller could still be used by re-routing feedback signals out of the new topologies to the controller. Overall, the experience gained and learned from this study

has definitely enhanced our technical skills in the field of multiphase converter in particular and dc-dc converter in general.

3. Introduction

Pentium-IV microprocessors as widely known run at about 3 GHz. However, future microprocessors will run at a much faster speed of 20 GHz. In Figure 1, according to the Intel Corporation, transistor count in one microprocessor chip will increase from 200 millions in 2005 to 1 billion in 2010. This poses major challenges particularly on power consumption of the microprocessors and overall efficiency of power supply module for these microprocessors. One well known solution to reduce the power consumption issue is to reduce the operating voltage of the microprocessors while drawing high currents, see Figure 2. As an example, in 2010 microprocessors are expected to draw 150 Amps at 0.8 Volts supply voltage with ± 2 % output voltage tolerance.

The power supply design for these microprocessors, at this very low voltage but high power level will be facing serious technical challenges to accomplish fast transient response, high efficiency, and high power density [1]. Today's widely used topology is based on buck or step-down topology connected in multiphase and interleaving fashion. This basic multiphase topology has been proven to reduce output current ripple while improving the transient response and the power density [2]. However, this technique suffers the limitation of large inductor current ripple in each channel [3] as well as possessing inductor current ripple frequency which is the same as the converter's operating switching frequency.

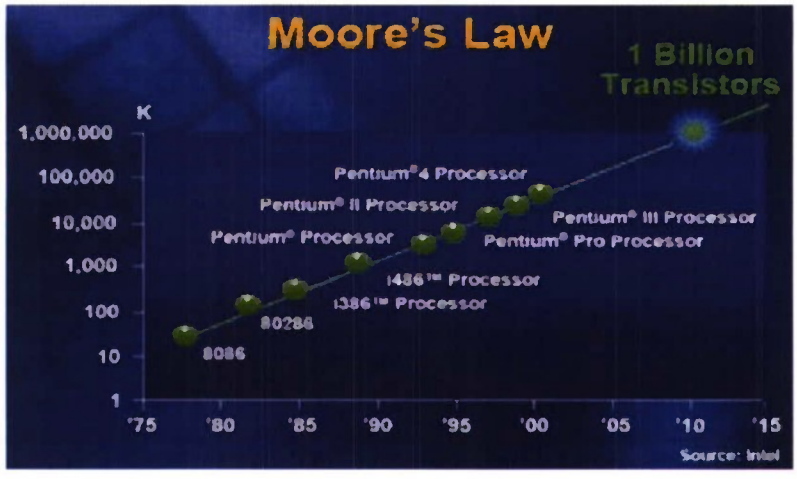


Figure 3-1. Historical data of number of transistors per processor vs. Moore's law

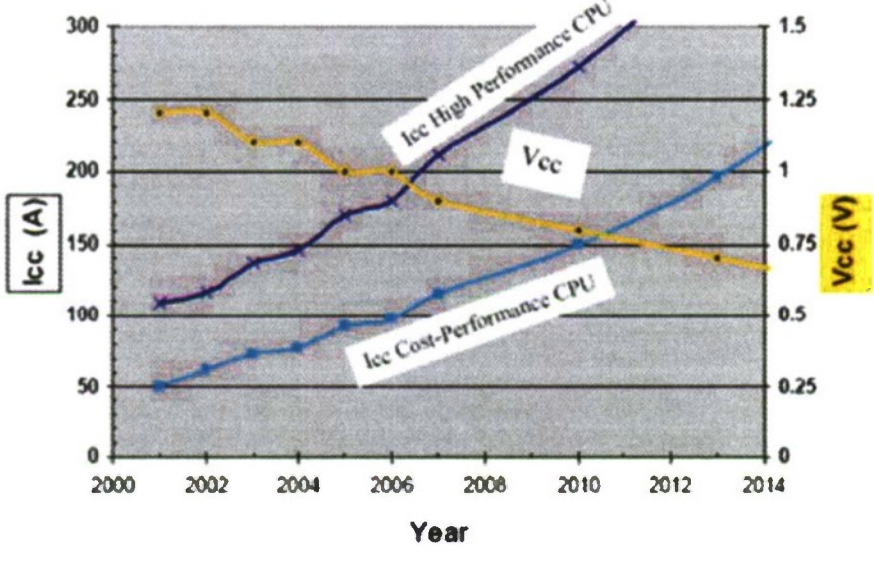


Figure 3-2. Voltage and current requirements of microprocessors

The following summarizes four major issues associated with the present topology:

1. When the output voltage decreases, the converter's duty cycle also decreases. In the case of 12 V input VRM, to provide output voltage of 0.5 to 1 Volt, the present topology suffers very small duty cycle below 10 %. This will not only yield poorer ripple cancellation effect, but it will also increase turn-off loss on the top switch and conduction loss on the bottom switch in each channel [4].

2. When the converter's switching frequency is increased, the switching losses (turn-off, gate drive and body diode losses) increase proportionally. The use of duty cycle below 10% while raising the switching frequency to multi-MHz will consequently reduce the converter's efficiency to less than 80% [5].
3. The fact that converter's output inductor-current ripple frequency is the same as converter's switching frequency forces practical converter's operation between 300 kHz to 500 kHz to achieve high overall efficiency given the very small duty cycle. This in turn demands the excessive use of bulk output capacitors and decoupling capacitors [6].
4. When the converter's output inductor current slew rate is increased with smaller inductance value to improve the transient response, the converter's output inductor current ripple consequently will increase. This would result in a larger turn-off loss for the main active switch as well as larger conduction loss for the freewheeling active switch of the converter and the worsening of converter's output inductor winding losses. This conflict unfortunately will limit the average output inductor current in each channel and imposes a difficult design trade-off between efficiency and transient response [7].

These above four major problems currently constitute increased costs, reduced power density, and difficulties in meeting the power requirements of dc-dc converter for future microprocessors [8]. These four problems therefore need to be resolved.

In recent years, new converter topologies have been proposed to solve the aforementioned technical challenges. Examples of such topologies are the multiphase clamp coupled-buck converter [9], phase shifted buck converter [10], winding-coupled buck converter [11], etc. These converters focus on extending the duty cycle which is not enough to the remaining technical challenges. A two-stage solution has also been proposed [12]. However, this approach uses an extra converter's stage which adds complexity, reduces reliability, worsens efficiency, and lowers density. Therefore, the next logical step is to formulate new topologies that avoid the use of extra converter stage, and address most if not all the issues listed previously. The new topologies developed in this project adopted these approaches. This was

done mainly by carefully and strategically placing passive storage components around the converter. Preliminary simulation results suggest that the new converter topologies indeed provide the benefits to meet the requirements for high-current converters as later demonstrated by lab measurements of the hardware setup. Furthermore, the hardware prototypes of the new topologies also prove that the new converters could be built light-weight and low-profile.

4. The New Topologies

In this project, two new topologies were studied, designed and developed to exhibit their potential in high output current at low voltage applications. The two topologies are called Topology 1 and Topology 2. In this chapter the basic power stage of each topology will be described, and their operation will be explained in terms of energy flowing through their main power components.

4.1. Topology 1

Figure 4-1 shows the proposed topology for multiphase buck converter. The circuit has 4 phases in a unique configuration. Each 2 cells buck is grouped together into one module whose output is connected to a Schottky diode to provide a shorter freewheeling path for inductor L5, L6 for the top module, and L7 and L8 for the bottom module. Inductor L5 and Capacitor C1 alongside with L7 and C3 are added as a bypass LC filter between input and output. This filter is used to provide a zero-ripple output current which in turn reduces the output capacitor rms current. This in turn reduces capacitor loss and lower rms rating of the capacitor. Auxiliary inductors L6 and L8 were added after the LC filter to provide more filtering at the output. Since the converter is grouped into two cells, the timing signal for each phase is also modified from those of conventional multiphase converter for a more distributed heat flow and hence may help method of applying the timing signals is what is commonly known as interleaving as will be explained next.

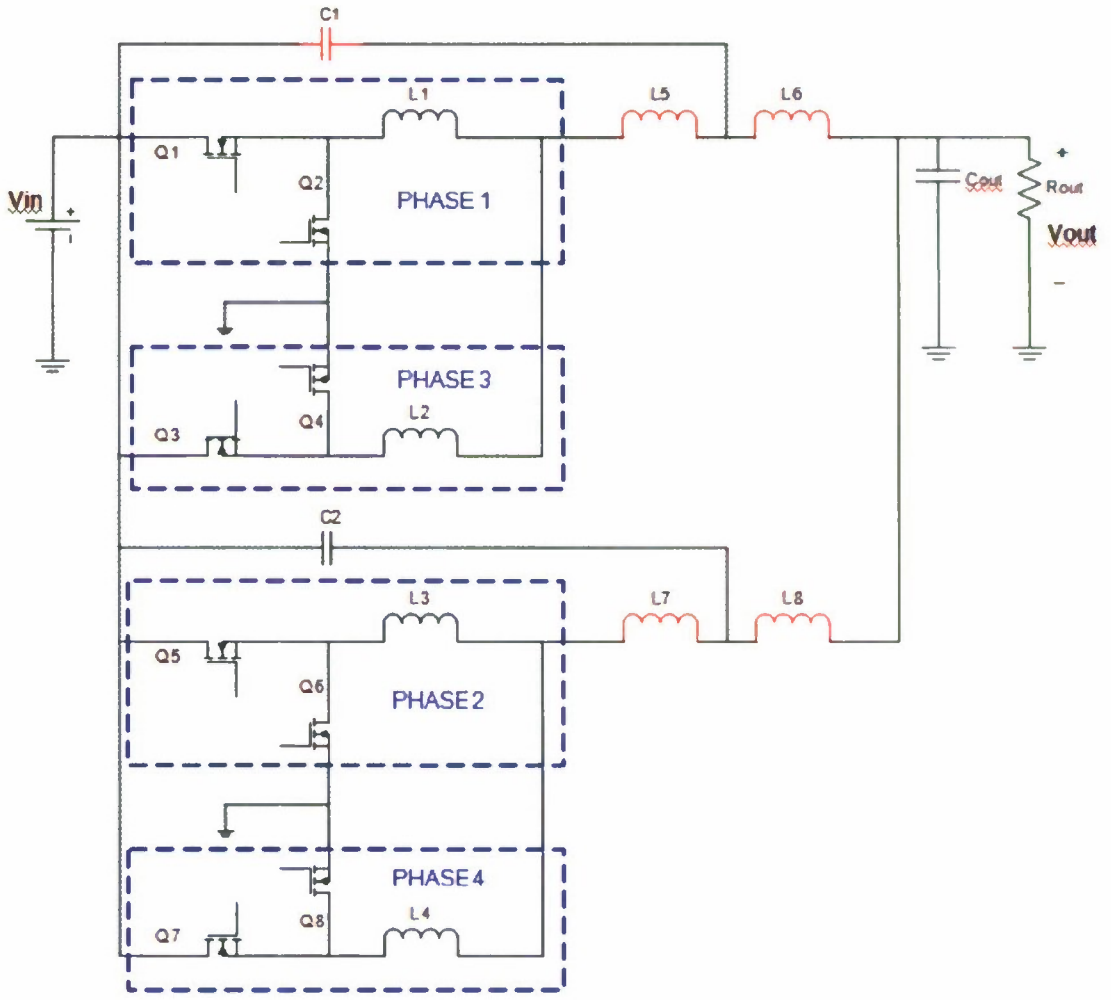


Figure 4-1: Proposed Multiphase Buck Topology 1

The following is the description of steady state operation from input to output of each phase at 1 switching cycle. Keep in mind that while Q1 conducts, Q3, Q5, and Q7 are turned off. Also the MOSFETs and diodes that turn on at any time period are being left on the diagram. The timing diagram from each phase is shown below on Figure 4-2.

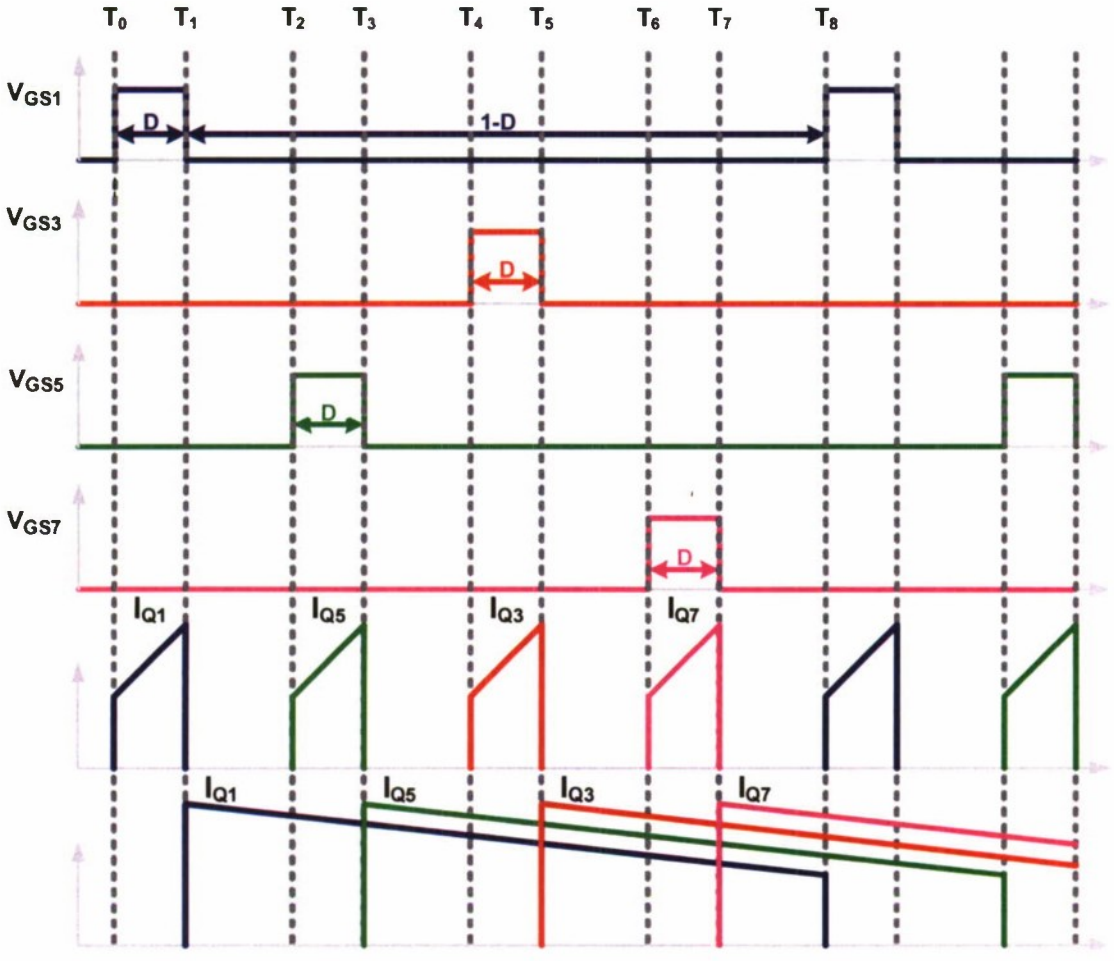


Figure 4-2: Timing Diagram for Top MOSFET Gate Signal

Figure 4-2 shows the gate drive signals for each appropriate MOSFET. It is also can be seen from the figure above that MOSFET Q5 turns on before Q3 in this proposed multiphase topology. The timing diagram shown above excludes the dead time between the transition from the top MOSFET and bottom MOSFET and vice versa.

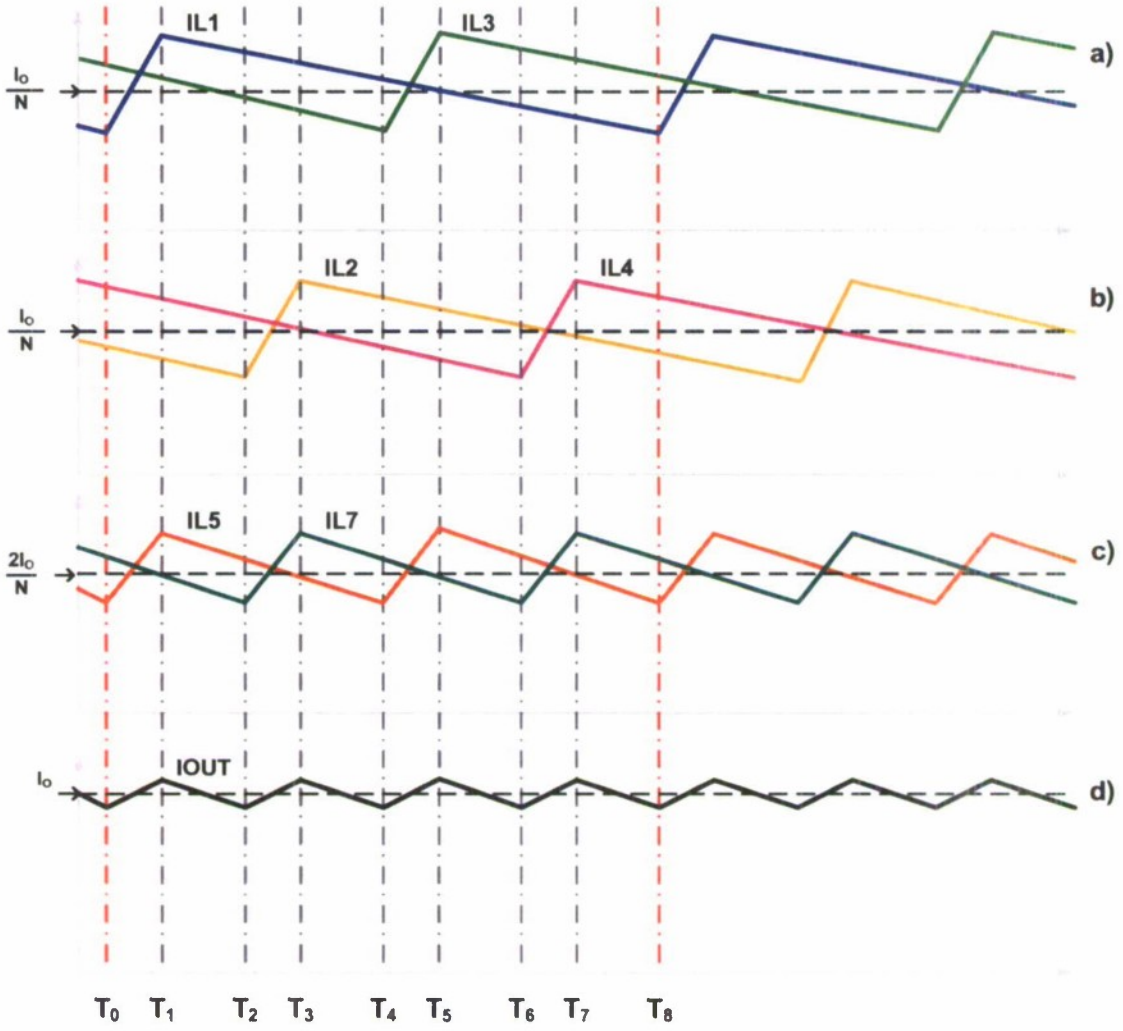


Figure 4-3: (a) Inductor Current for phase 1&3, (b) Inductor Current for phase 2&4, (c) Auxiliary Inductor Currents, (d) Final Output Current Ripple

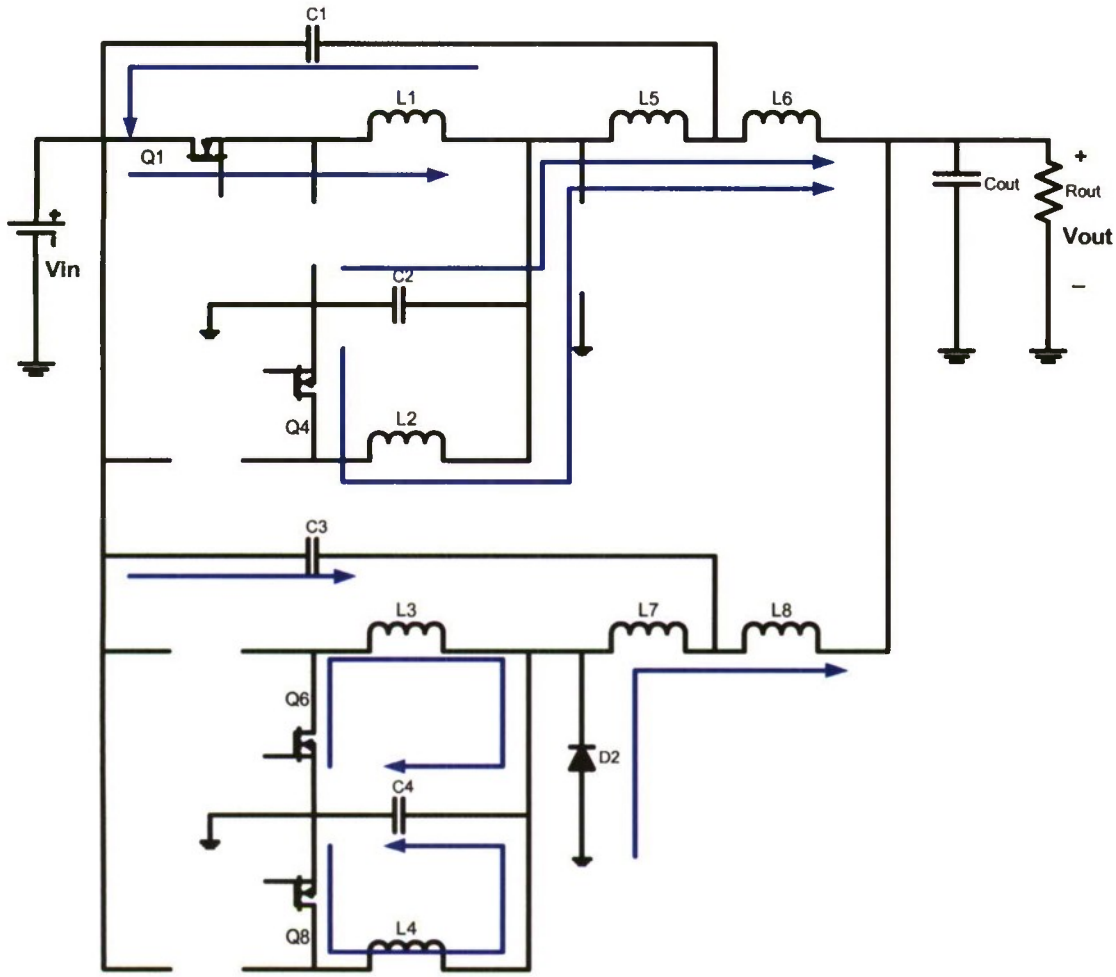


Figure 4-4: Energy flow during T_0 to T_1

During interval T_0 to T_1 on Figure 4-2, Q1 turns on while diode D1 turns off, see Figure 4-2. Current flows from Vin to output through Q1, L1, L5 and L6. In this case the current through L1, L5, and L6 increases linearly since the input and output voltage are both fixed at 12 V and 1 V respectively based on equation (1-2). At the same time, energy stored in C1 is being discharged through Q1 and L1. While the energy stored in C2 is also being discharged through L5 and L6. Meanwhile, L2 is also discharged through L5 and L6.

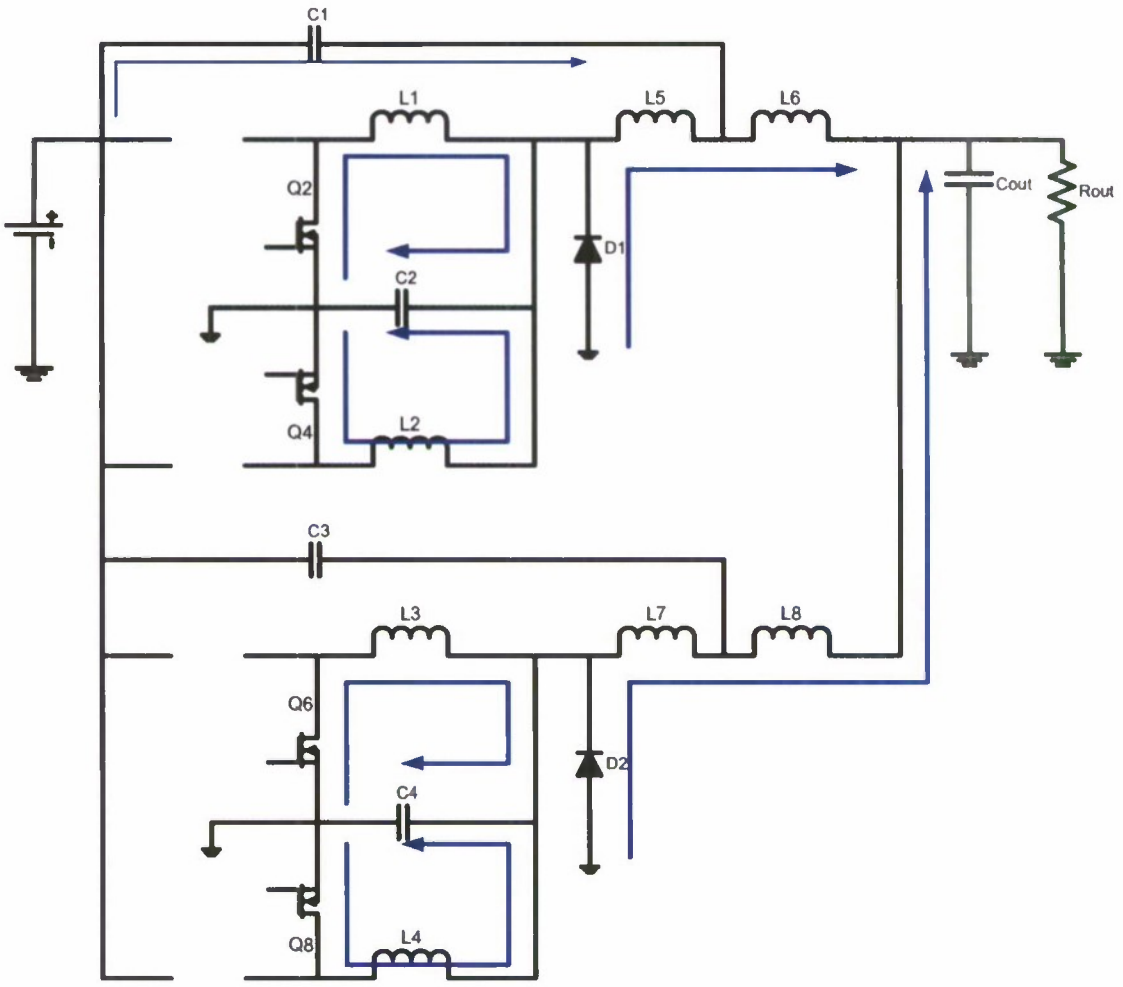


Figure 4-5: Energy flow during T_1 to T_2

At time t_1 switch Q1 is turned off, and switch Q2 is turned on as illustrated in Figure 4-4. During T_1 to T_2, the energy stored in L1 together with energy left in L2 is now being used to charge C2. Diode D1 conducts during this time period, and, energy stored in L5 and L6 flows to output. The energy in C1 would be charged by the input during this time.

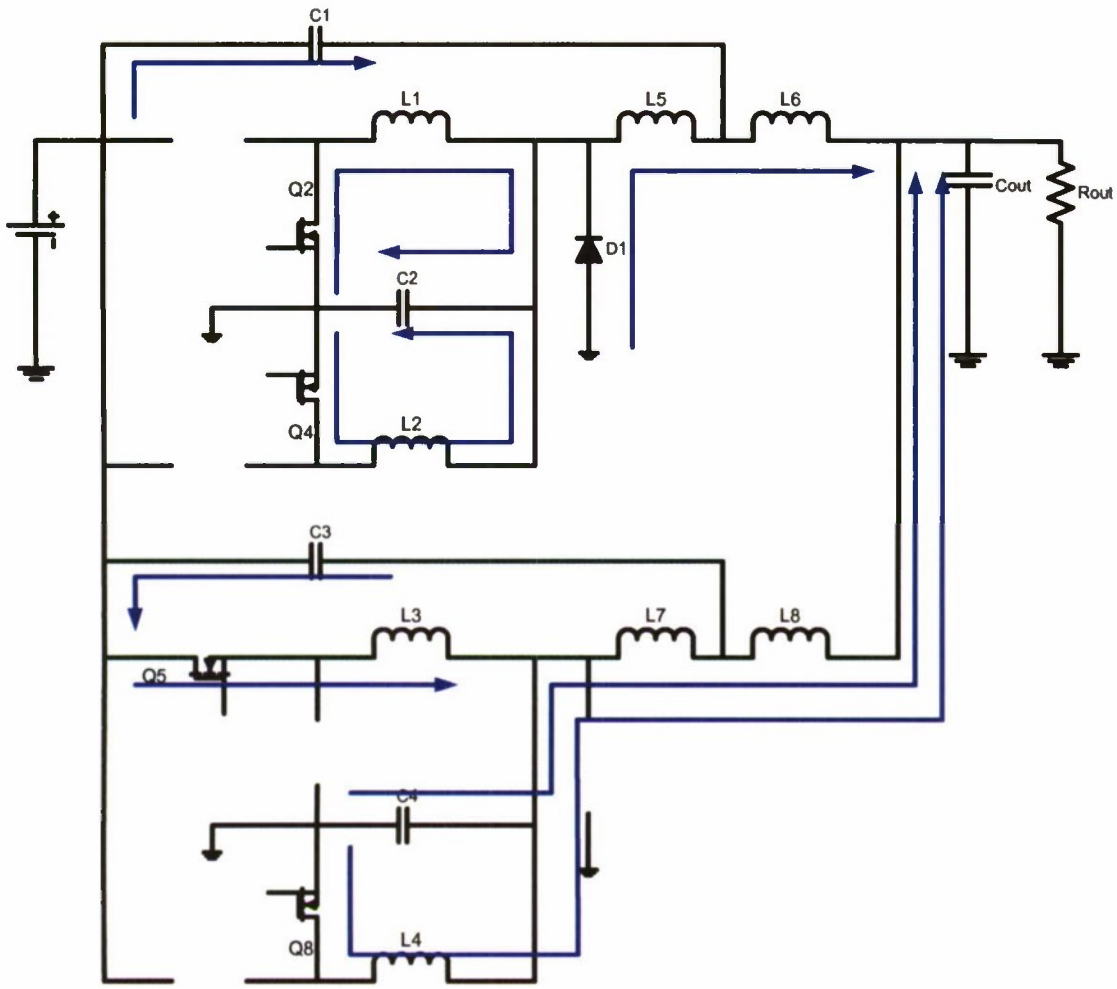


Figure 4-6: Energy flow during T_2 to T_3

In regular Multiphase interleaving technique, Q3 will be the next switch to conduct. In this proposed topology, however the next switch that is going to turn on will be Q5 as depicted in Figure 4-2. By switching on Q5, the heat sources from each MOSFET will be distributed more evenly compared with turning on Q3. In fact, in order to interface this topology with the Multiphase PWM controller, Q5 has to be on first in order to achieve a high degree of current balancing. During the period from T_2 to T_3, the same sequence of energy flow occurs as the one described in the first phase (from T_0 to T_2). C3 replaces C1, C4 replaces C2, L3 replaces L1, L7 replaces L5, L8 replaces L6, and D2 replaces D1. The same cycle will also repeat for phase 3 (Q3 and Q4) and phase 4 (Q7 and Q8).

4.2.Topology 2

The schematic for Topology 2 is Figure 4.7. Just like Topology 1, interleaved switching was used for improved equal current sharing and better heat distribution. The proposed multiphase synchronous buck topology incorporates additional storage components that serve different purposes. For example, the additional output inductors (L5, L6, L7, L8) are placed to minimize output current ripple useful in reducing rms loss at the output capacitor (Cout) or from the copper loss of the inductors themselves, including from the main inductors (L1, L2, L3, L4). However, these inductors will consequently slow down the transient response which may be overcome by increasing the switching frequency of the converter, and by adding the input-output bypass capacitor in each cell (C1 and C3) for energy support required by the load during transient. In an N-phase multiphase, the duty cycle for each phase is equal to Vout/Vin and it is the same for all phases due to parallel configuration. A phase shift should therefore be implemented between the timing signals of the top switch from the first and second phases. The value of the phase shift follows the equation 360°/N where N is number of phases. For example, in the 2-phase case, the amount of phase shift will be 360°/2 = 180°.

Interleaved switching is mainly used such that the cells in the proposed converter will be better balanced in its energy flow. Interleaved switching is done by modifying the firing sequence of the individual buck converters. In the proposed topology there are two cells that make up the entire four phases. To interleave, the top buck of the top cell will fire first, and then the top buck of the bottom cell will fire. Next, the bottom buck of the top cell will fire, and finally the bottom buck of the bottom cell will fire. Figure 4.8 shows how each cell's output current is balanced due to the interleaved switching.

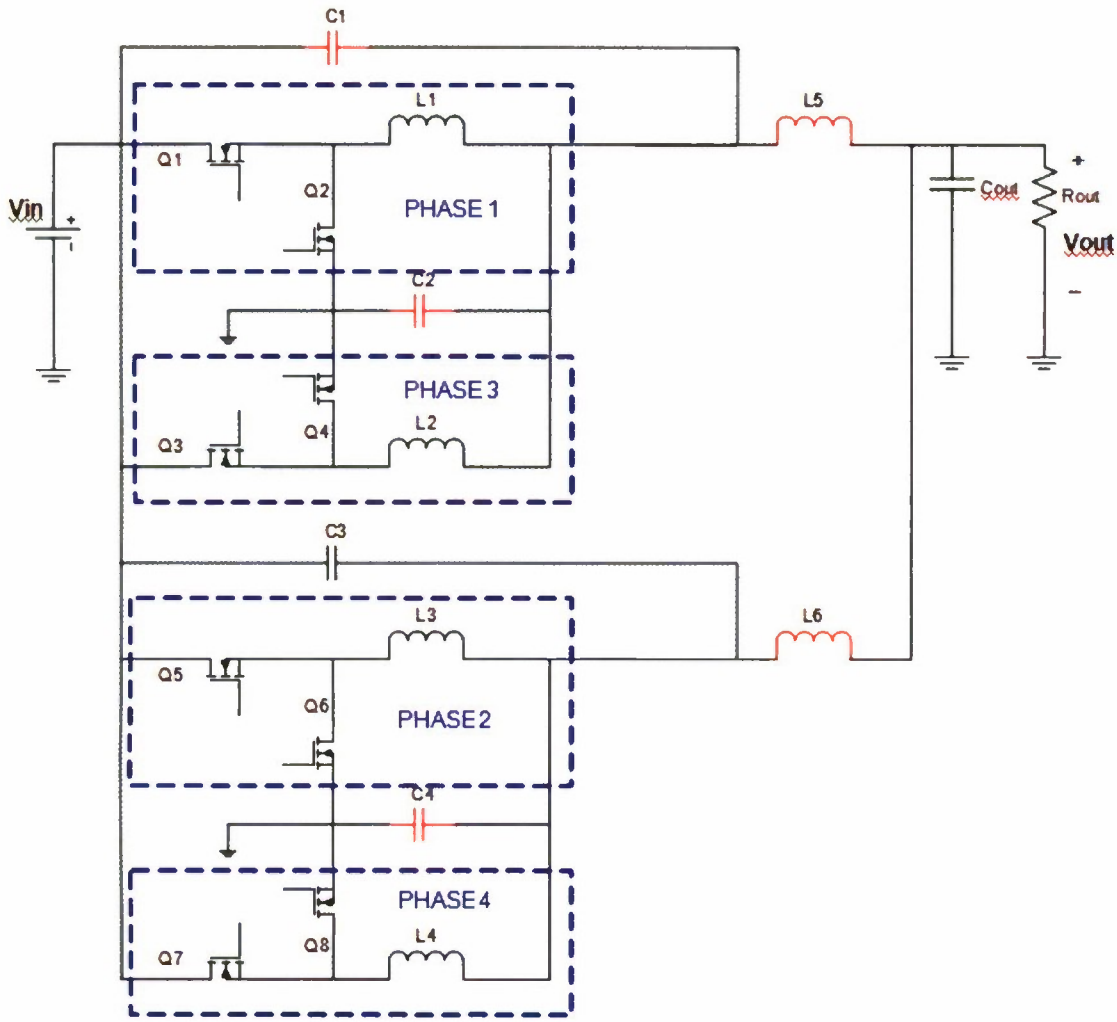


Figure 4-7: Proposed Multiphase Buck Topology 2

Interleaved switching is advantageous since it will yield a smaller current ripple and faster frequency compared to non-interleaved switching. When using cells as in the proposed topology, there is an inductor at the output of each cell. A smaller current ripple means that the RMS current will be less. Therefore, there will less power losses in the output inductors of the cells when using interleaved switching. Figure 4.9 shows the cell currents in the proposed topology when interleaved switching is not being used. When comparing Figure 4.5 and Figure 4.9, we can see the difference between the current ripple and frequency of the cells' output currents. However, an important note is that both interleaved and non-interleaved multiphase bucks would have the same output current ripple.

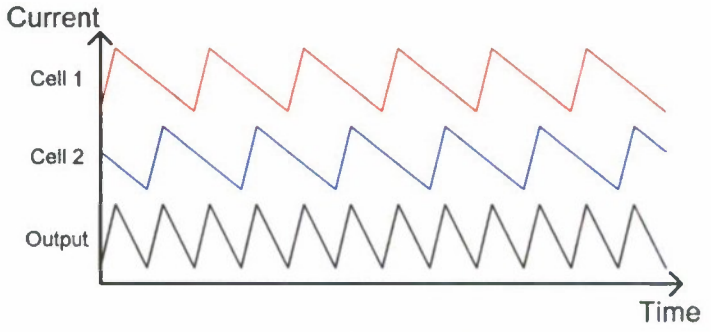


Figure 4.8 Cell Current Using Interleaved Switching

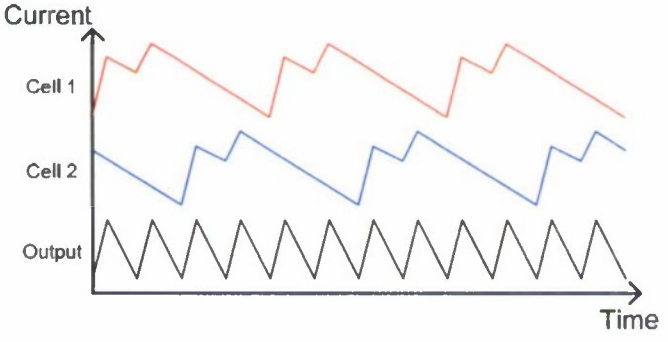
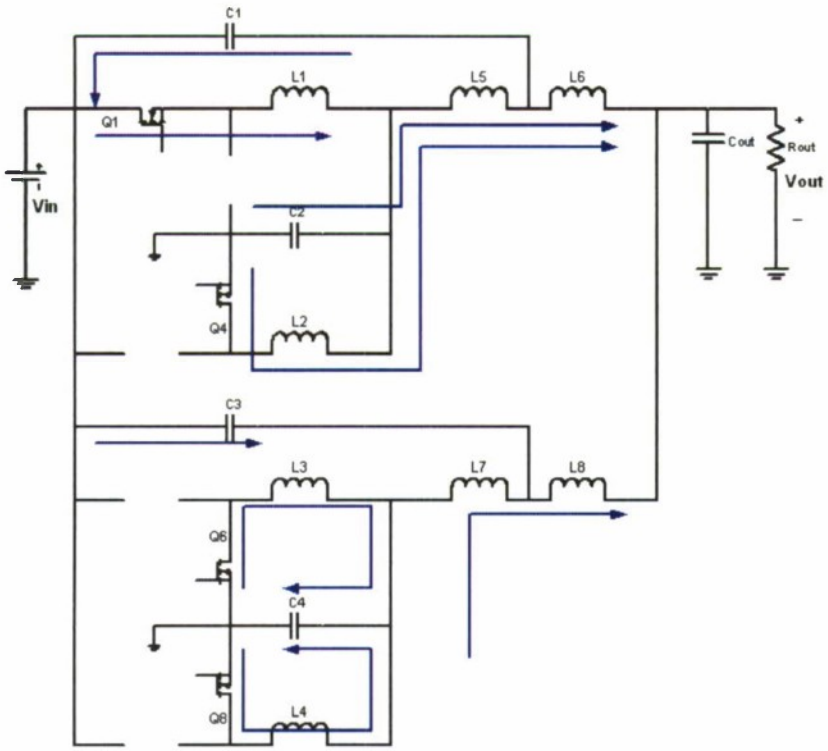
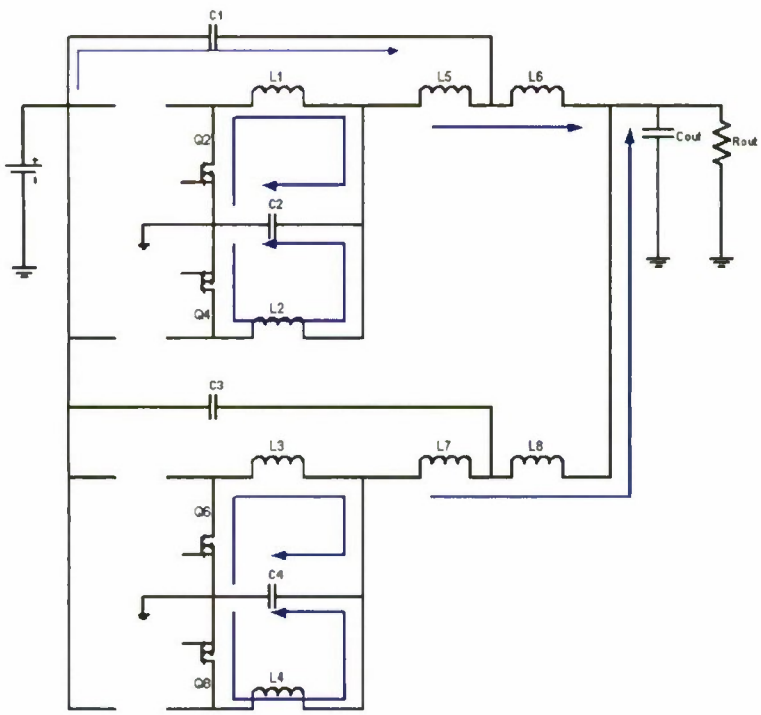


Figure 4.9 Cell Current Using Non-Interleaved Switching

Referring to times T_o to T_8 as shown in Figure 4-10, during interval T_o to T_1, Q1 turns on. As illustrated in Figure 4-10 (a), current flows from Vin to output through Q1, L1, L5 and L6. In this case the current through L1, L5, and L6 increases linearly since the input and output voltages are both fixed at Vin and Vout respectively. At the same time, energy stored in C1 is being discharged through Q1 and L1, while the energy stored in C2 is also being discharged through L5 and L6. Meanwhile, L2 is also discharged through L5 and L6. At time T_1 switch Q1 is turned off, and switch Q2 is turned on as illustrated in Figure 4-10 (b). During T_1 to T_2, the energy stored in L1 together with energy left in L2 is now being used to charge C2. Energy stored previously in L5 and L6 flows to output. The energy in C1 would be charged by the input during this time. The next transition from T_2 to T_3 is depicted in Figure 4-10(c). Switch Q5 is turned on, and the same sequence of energy flow occurs as the one described in the first phase (from T_o to T_2). Here, C3 replaces C1, C4 replaces C2, L3 replaces L1, L7 replaces L5, and L8 replaces L6. The same cycle will also repeat for phase 3 (Q3 and Q4) and phase 4 (Q7 and Q8).


C1
L1
L5
L6
Q1
Cout
Rout
Vout
Vin
C2
Q4
L2
C3
L3
L7
L8
Q6
C4
Q8
L4

(a)


C1
L1
L5
L6
Q2
Cout
Rout
C2
Q4
L2
C3
L3
L7
L8
Q6
C4
Q8
L4

(b)

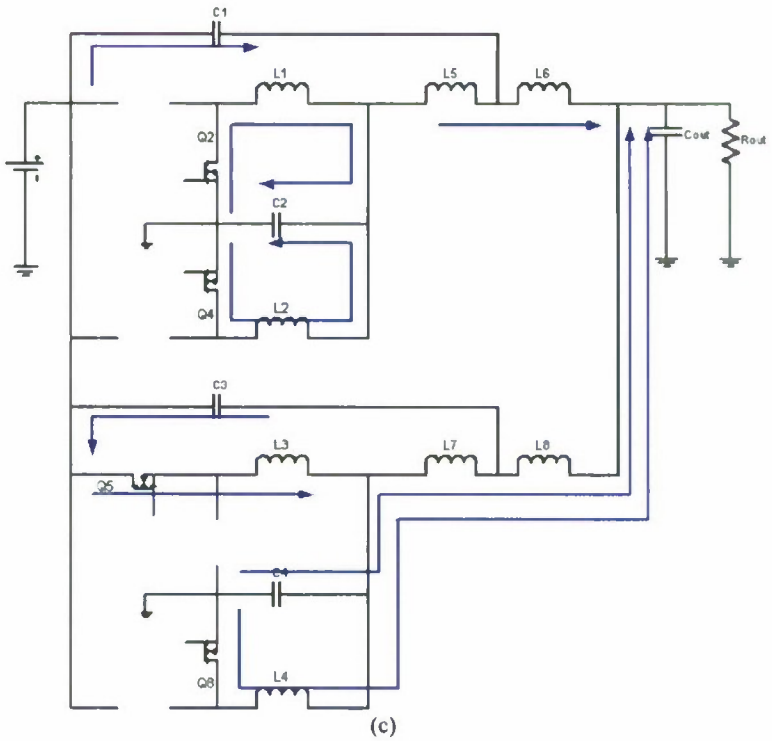


Figure 4.10 Energy flow during time (a) T_o–T_1, (b) T_1–T_2, and (c) T_2–T_3

In the following chapter, the design equations to select circuit components with the proper ratings for each new topology will be presented. In addition, estimation of overall efficiency for each topology will also be described.

5. Design and Loss Calculations

As the first step of designing hardware prototype for the two new topologies, design calculations were performed to select the appropriate components in terms of their ratings and parameters. In addition, estimation on power losses were also derived whose results were then used to estimate the estimated overall efficiency of each converter.

5.1.Design Requirements

A nominal input and output voltage was chosen based on the present requirement of VRM. Currently, VRM operates with a 11.4 to 12.6 V input and 0.8 to 1.55 V output voltage based on a sampled module from C&D Technologies. Therefore a selection of 12 V input

and 1 V output should be reasonable. The inductor current ripple is chosen to be 10% of maximum phase current based on the common design value for inductor current ripple of 10% - 20% from maximum phase load. On the other hand, our requirements for peak to peak current ripple is rather conservative considering some VRM manufacturers such as C&D Technologies has already made a VRM with an output peak to peak ripple of 6.4 mV. The most important part of these requirements would be an efficiency of 80% or above at full load at 500 kHz switching frequency. Figure 5-1 summarizes the design requirements for the hardware prototyping of the new topologies.

Parameter	**Requirements**
Nominal Input Voltage	12 V
Nominal Output Voltage	1 V
Maximum Output Current	40 A
Inductor ripple current	10 % of Maximum Phase Current
Output Voltage Ripple	< 15 mV_{p-p}
Switching Frequency	500 kHz per phase
Load Regulation	< 2 %
Line Regulation	< 5 %
Efficiency	> 80 % at Full Load

Figure 5.1 Design Requirements for the new Multiphase Buck Converters

Critical Inductance

Since the nominal value of input voltage is 12 V and the converter outputs a 1 V, then each phase of buck converter operates with an ideal duty cycle of 1/12. In a single phase buck converter, the inductor and output capacitor act as an LC filter at the output. These two values were important because they determined the stability of the converter. Small value of inductance along with high output capacitance value will result a good transient response for a step load change. Unfortunately, a small value of inductance also yields a high peak to peak inductor current ripple, thus reducing the converter's efficiency due to increased rms capacitor and inductor currents. In multiphase buck converter, each phase inductance is

connected in parallel. Its effective inductance $L_{(eff)}$ is equal to phase inductance L_F divided by number of phases N:

$$L_{(eff)} = L_F / N \tag{5-1}$$

From equation above, the effective inductance is less than the phase inductance. Therefore, the multiphase buck converter should yield a good transient response based on a higher slew rate on inductor current.

The inductor current for each phase would be 10 A based on the requirements of 40 A full load and 4 phase buck converter. A 40 A full load was selected only as an example to test the functionality of the proposed multiphase buck topology. The actual multiphase buck that is used for VRM should run at least 100 A. For each phase, the minimum value for the inductance to maintain Continuous Conduction Mode for average inductor current of 10 A:

$$L_{MIN} = \frac{V_{OUT}(1-D)}{2 f_{SW} I_{OUT}} \tag{5-2}$$

According to Figure 5-1, the minimum average load current required while the converter maintains CCM is 10 % of maximum phase current. This yields the inductor ripple current requirement of 10% * 10 A = 1 A. Hence:

$$\Delta i_L = \frac{(V_{IN} - V_{OUT})}{L} * DT_{SW} \tag{5-3}$$

The maximum inductor current would be 10 A plus 0.5 A ripple based on our design choice. Therefore, the inductor should be able to withstand a total current of 10.5 A current without saturation and a low Direct Current Resistance (DCR) value in order to increase efficiency.

Input Capacitor

One disadvantage of buck converter is the discontinuous input current. This is caused by the location of its main switch right at the front end of its input stage. The turning on and off of the switch inherently disrupts the current flow at the input. The impact of this would then be

the high di/dt noise which translates back into its DC bus rail. For the power supply to provide a flat DC current, then an intermediate input capacitor may be placed to accommodate for the pulsating input current, hence input ripple current. A multiphase buck converter turns on the high side switch from each channel at different phase. This would result in multiple drawing of current from the input. In turn, this will provide a smaller value of input rms current. The rms input current reduction further allows us to use a smaller number of input capacitors due to less input rms current requirement compared to that in a single phase buck converter. Lastly, another benefit from a reduced rms input current is an increase in converter's efficiency because of smaller input rms current implies smaller losses on input capacitors Equivalent Series Resistance (ESR).

The input rms current for capacitor can be found using the following mathematical expression:

$$I_{IN\_RMS} = \sqrt{\left(D-\frac{k}{m}\right)\left(\frac{k+1}{m}-D\right)I_O{}^2 + \frac{mc^2}{12mD^2}\left(\frac{V_o(1-D)T}{L_f}\right)^2 * \left[(k+1)^2\left(D-\frac{k}{m}\right)^3 + k^2\left(\frac{k+1}{m}-D\right)^3\right]} \qquad (5\text{-}4)$$

where D = Duty Cycle
T = Period for each phase [secs]
mc = Number of channel
m = Number of phases
k = FLOOR(m*D), m = 1,2,3,…

The FLOOR function returns an integer value less than or equal to the expression inside the brackets. For example a 4 phase buck with a duty cycle of 0.3 would yield a value of k equal to 1. From the above expression, a graph could be made to approximate the input rms current for different number of phases.

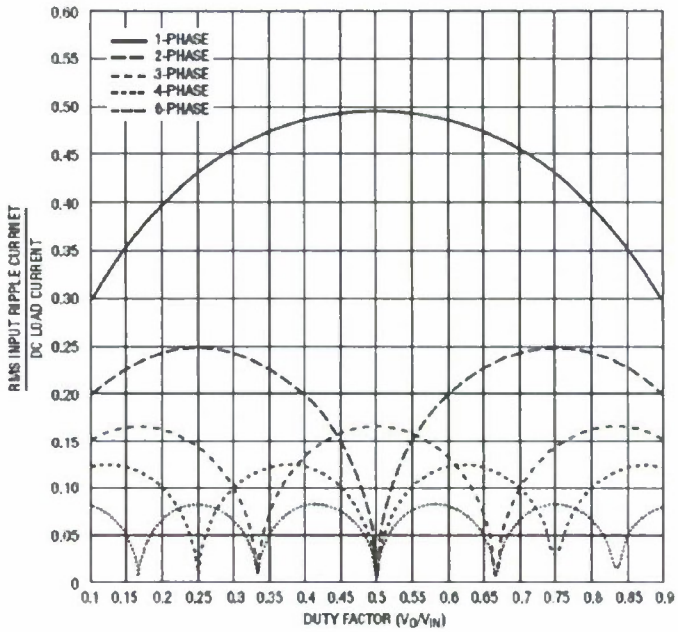


Figure 5.2 Input Ripple Cancelation Factor

One fact that should be observed is that the input ripple current is not exactly zero when output current is zero. This is because of the second term on the equation 5-4 does not depend on output current I_o.

MOSFET

For our MOSFET selection, a low $R_{DS\text{-}ON}$ MOSFET should be used for the synchronous or bottom or freewheeling MOSFET since this particular MOSFET conducts 11/12 = 91.67% times out of the whole switching period. The same MOSFET should also have a ratio of Q_{GD}/Q_{GS} less than 1 to prevent dv/dt induced turn on caused by voltage spike on the drain while this MOSFET turns off. If this requirement is not fulfilled, then there is a probability that voltage spike across drain and source pulls the gate to source voltage up higher than $V_{GS\text{-}TH}$ and turns the MOSFET on. This will result in simultaneous conduction of top and bottom MOSFETs causing cross conduction which produces a dead short across the input. On the other hand, the top MOSFET should be selected based on low gate charge requirement to minimize switching loss. A higher gate charge translates into a longer rise and fall time during MOSFET switching. This in turn increases switching loss.

Furthermore, in order to increase converter's efficiency, a proper value for gate to source voltage of the MOSFETs should be applied. A higher V_{GS} would reduce $R_{DS\text{-}ON}$ for the MOSFET but increases the total gate charged on each MOSFET. As mention above, a higher gate charge translates into a higher switching loss. Therefore a careful selection of gate to source voltage should be made in order to minimize MOSFET's $R_{DS\text{-}ON}$ and the total MOSFET's gate charge at the same time.

Output Capacitor

The development of capacitor technologies in power converters is being driven by a high frequency and low ESR requirements. It is proven to be difficult to offer one capacitor technology that has a low ESR and high ripple current capability.

There are two types of capacitor technologies which are electrolytic and electrostatic capacitors. Electrolytic capacitor such as Aluminum Electrolytic and Tantalum capacitors offers a high capacitance and low impedance at frequencies below 100 kHz. However, this type of capacitance has a high ESR value compared to electrostatic capacitors, thus reducing efficiency of power converters. On the other hand, electrostatic capacitors such as ceramic and polymer film offers a low ESR and ESL values at higher resonance frequency. Unfortunately, the technology only offers this type of capacitor at low capacitance values.

Improvement has been made for electrolytic capacitors in order to fulfill low ESR requirements while maintaining a high capacitance value on high switching frequency applications. Organic Semiconductor (OS-CON) electrolytic capacitors have been developed to have the lowest ESR at high switching frequency. One of manufacturers of this type of capacitors is Sanyo. Its performance has also been tested against low impedance aluminum and tantalum capacitors. In terms of ripple removal and impedance characteristic at different frequencies, OS-CON has been proven to be more efficient compared to aluminum and tantalum capacitors. A minimum value of capacitance can be calculated using equation 5-7:

$$C_{OUT} = \frac{(1-D)}{8Lf^2}\left(\frac{V_{OUT}}{\Delta V_{OUT}}\right) \qquad (5\text{-}5)$$

The value above is calculated assuming an ideal capacitor is being used in the circuit. An actual capacitor value should be higher considering capacitor's ESR. An estimated value of output ripple voltage is then given by the following expression.

$$\Delta V_{O\_pk-pk} < \frac{\Delta I_O T}{8(N)C_{OUT}} + \Delta I_O * ESR \qquad (5\text{-}6)$$

The first term on the equation calculates the output voltage ripple contribution from a pure capacitive component of the output capacitor (C_{OUT}). It can be seen also from equation above that a higher number of phases (N) will reduce this contribution value. On the other hand, the second part of the equation calculates the ESR contribution to the output voltage ripple.

Another consideration regarding the selection of output capacitor would be transient voltage overshoot during a step load change from full load to no load. Capacitance value can then be calculated by equating inductive and capacitive energy.

$$C_{OUT} = \frac{L * I^2}{V^2} = \frac{L_{EQ} * \left((I_{FL})^2 - (I_{NL})^2\right)}{(V_{OUT2})^2 - (V_{OUT1})^2} \qquad (5\text{-}7)$$

where L = effective inductance
I_{FL} = Full Load Current
I_{NL} = No Load Current
V_{OUT2} = Maximum output voltage with overshoot
V_{OUT1} = Nominal value for output voltage

5.2.Design Calculations for Component Selections

The following describes the calculations for components selections of Topology 1 & 2:

Given Parameters:

Input Voltage:	$V_i := 12V$
Output Voltage:	$V_o := 1V$
Output Voltage ripple	$\Delta V_o := 15mV$
Switching Frequency:	$f_s := 500kHz$
Output Current @ Full Load:	$I_o := 50A$
Number of Phase:	$N_p := 4$

General Circuit Parameters Calculation:

Switching Period

$$T_s := \frac{1}{f_s} \qquad T_s = 2 \times 10^{-6} s$$

Duty Cycle:

$$D := \frac{V_o}{V_i} \qquad D = 0.083$$

Phase Inductor Calculation:

Phase Inductor Current:

$$I_{Lphase} := \frac{I_o}{N_p} \qquad I_{Lphase} = 12.5A$$

Minimum Phase Inductance to maintain CCM:

$$L_{mn} := \frac{V_o \cdot (1 - D)}{2 \cdot f_s \cdot I_{Lphase}} \qquad L_{mn} = 7.333 \times 10^{-8} H$$

Required Phase Inductor Current Ripple (10% of maximum Phase Current):

$$\Delta I_{Lphase} := 10\% \cdot I_{Lphase} \qquad \Delta I_{Lphase} = 1.25A$$

Minimum Phase Inductance needed to satisfy current ripple:

$$L_{ph.min} := \frac{(V_i - V_o)\cdot D\cdot T_s}{\Delta I_{Lphase}} \qquad L_{ph.min} = 1.467\times 10^{-6}\,H$$

Peak Phase Inductor Current:

$$I_{Lmax} := I_{Lphase} + \Delta I_{Lphase} \qquad I_{Lmax} = 13.75A$$

Chosen Phase Inductor:

Coiltronics P/N: HCF1305-1R8-R with inductance value of 1.8µH, DCR = 3.1mΩ, peak current = 16.3A

Input Capacitor Calculation:

Given:

Duty Cycle: $D = 0.083$

Switching Period: $T_S = 2\times 10^{-6}\,s$

Output ripple frequency: $f_{ripple} := f_s \cdot N_p$

$$f_{ripple} = 2\times 10^{6}\,\frac{1}{s}$$

Number of Channel $mc := 4$

Number of Phases $N_p = 4$

$k := floor(N_p \cdot D)$ $k = 0$

$L_{phase} := 1.75\mu H$

$$L_{phase.eff} := \frac{L_{phase}}{4}$$

$L_{phase.eff} = 4.375\times 10^{-7}\,H$

Input Voltage ripple: $\Delta V_i := 1\% \cdot V_i$

$\Delta V_i = 0.12V$

$$I_{in.rms} := \sqrt{\left[\left(D - \frac{k}{N_p}\right)\cdot\left(\frac{k+1}{N_p} - D\right)\cdot I_o^2 + \frac{mc^2}{12\cdot N_p\cdot D^2}\cdot\left[\frac{V_o\cdot(1-D)\cdot T_s}{L_{phase}}\right]^2\cdot\left[(k+1)^2\cdot\left(D - \frac{k}{N_p}\right)^3 + k^2\cdot\left(\frac{k+1}{N_p} - D\right)^3\right]\right]}$$

$I_{in.rms} = 5.895A$

$$C_{in.min} := \frac{D\cdot I_o\cdot(1-D)\cdot T_s}{\Delta V_i} \qquad C_{in.min} = 6.366\times 10^{-5} F$$

Chosen Input Capacitor:

United Chemi-con P/N: APXE160ARA101MF80G (two capacitor will be used in parallel), each capacitor: 100µF/16V, ESR = 24mΩ, with rms ripple current rating of 3A.

Output Capacitor Calculation:

Output Current Ripple:

$$\Delta I_o := \frac{V_o\cdot(1-D)\cdot T_s}{L_{phase}}\cdot\frac{N_p\cdot\left(D - \frac{k}{N_p}\right)\cdot\left(\frac{k+1}{N_p} - D\right)}{D\cdot(1-D)}$$

$\Delta I_o = 0.762A$

Output Current rms:

$$I_{o.rms} := \frac{\Delta I_o}{\sqrt{12}} \qquad I_{o.rms} = 0.22A$$

$C_{out.min} = 4.365\times 10^{-6} F$

$$C_{out.ESR.max} := \frac{\Delta V_o}{\Delta I_o} - \frac{T_s}{8N_p\cdot C_{out.min}}$$

$C_{out.ESR.max} = 5.369\times 10^{-3}\Omega$

Chosen Output Capacitor:

Kemet P/N: A700D227M002ATE009 (two capacitor will be used in parallel), each capacitor : 220µF/2V, ESR = 9mΩ.

Main (Top) MOSFET Selection:

Top MOSFET require a low gate charge and fast rise and fall time to minimize switching loss in addition of low R_{DS-on} to minimize conduction loss.

Chosen Top MOSFET: RJK0332DPB-00#J0CT-ND from Renesas Technology America. The parameter for this MOSFET:

Main MOSFET Rds-on:	$TopMOS_{Rds.on} := 0.0036\Omega$
Main MOSFET rise time:	$TopMOS_{t.rise} := 3.8ns$
Main MOSFET fall time:	$TopMOS_{t.fall} := 4.4ns$
Main MOSFET total gate charge:	$TopMOS_{Qg} := 14nC$

With maximum Drain to Source Voltage of 30V and maximum current of 35A

Synchronous (Bottom) MOSFET Selection:

Low R_{DS-on} is the main consideration in choosing the Bottom MOSFET.

Chosen Top MOSFET: RJK0330DPB-00#J0 from Renesas Technology America. The parameter for this MOSFET:

Sync MOSFET Rds-on:	$SyncMOS_{Rds.on} := 0.002\Omega$
Sync MOSFET rise time:	$SyncMOS_{t.rise} := 3.9ns$
Sync MOSFET fall time:	$SyncMOS_{t.fall} := 5.4ns$
Sync MOSFET total gate charge:	$SyncMOS_{Qg} := 27nC$
Body Diode reverse recovery charge:	$SyncMOS_{Qrr} := 34nC$
Body Diode Forward Voltage:	$SyncMOS_{BD.Vf} := 0.78V$
Body diode reverse recovery time:	$SyncMOS_{BD.trr} := 36ns$

With maximum Drain to Source Voltage of 30V and maximum current of 45A.

Chosen PWM Controller: TPS40091PW from Texas Instrument. This four phase PWM Buck controller has been used on all of Multiphase Buck Project and Thesis that are based on CalPoly Multiphase Buck topology.

Chosen MOSFET Driver: ISL6614CBZ from Intersil. This driver is a dual synchronous buck MOSFET driver that is able to drive 4 MOSFETs (2 top + 2 bottom), is capable of operating up to 1MHz switching frequency, has advanced adaptive zero shoot through protection that provide enough but minimal dead time, and several other protection features.

5.3.Calculations for Power Losses

In this section, power losses on the converters will be analyzed to maximize the hardware design. Just like the component selections, the equations for losses in both topologies are similar to each other and hence they will be treated as one single set of equations.

Input Capacitor

The maximum ESR losses from input capacitor would be equal to

$$P_{ESR\_IN} = (I_{switch\text{-}rms})^2 * ESR \qquad (5\text{-}10)$$

MOSFET

The equation for approximating power losses on MOSFET was taken assuming a V_{GS} value of 5 Volt and a worst case value for C_{OSS}, R_{DS} on and Q_G from each MOSFET.

Top MOSFET

Conduction Loss ($R_{DS\text{-}ON}$ = 11.4 mohm and I_{OUT} = 10 A)

$P_{C\_TOP\_MOS} := I_{OUT}^2 \cdot R_{DSon} \cdot D$ $\qquad P_{C\_TOP\_MOS} = 0.095\ W$

Switching Losses (Assuming L_{LUMP} = 20 nH and I_G = 1.5 A)

$$t_R := \frac{Q_G}{I_G} + \frac{L_{LUMP} \cdot I_G}{V_{GS} - V_{TH}} \qquad t_R = 1 \times 10^{-8}\ s$$

$$t_F := \frac{Q_G}{I_G} + \frac{L_{LUMP} \cdot I_G}{V_{GS} - V_{TH}} \qquad t_F = 1 \times 10^{-8}\ s$$

$$P_{SW\_TOP\_MOS} := \frac{1}{12} \cdot V_{IN} \cdot I_{OUT} \cdot (t_R + t_F) \cdot F_{SW} \qquad P_{SW\_TOP\_MOS} = 0.1\ W$$

Output Capacitance Loss

$$P_{OUT\_TOP\_MOS} := \frac{1}{2} \cdot \frac{4}{3} \cdot C_{OSS} \cdot V_{IN}^{2} \cdot F_{SW} \qquad P_{OUT\_TOP\_MOS} = 0.026\ W$$

Gate Drive Loss (R_{G1sink} =1.65Ω, R_{G_drive1} =0Ω , $R_{G1source}$ = 2Ω , $R_{G\_Internal}$= 0Ω)

$$P_{G1} := Q_G \cdot V_{GS} \cdot F_{SW} \cdot \left(\frac{R_{G1sink}}{R_{G1sink} + R_{G\_drive1} + R_{G\_Internal}} + \frac{R_{G1source}}{R_{G1source} + R_{G\_drive1} + R_{G\_Internal}} \right)$$

$$P_{G1} = 0.07\ W$$

Hence, the total top MOSFET's losses:

$$TOTAL\_TOP\_MOS\_LOSS := P_{C\_TOP\_MOS} + P_{SW\_TOP\_MOS} + P_{OUT\_TOP\_MOS} + P_{G1}$$

$$TOTAL\_TOP\_MOS\_LOSS = 0.291\ W$$

Bottom MOSFET

Body Diode Loss (V_F = 700mV , t_{BDR} + t_{BDF} = 32 ns)

$$P_{BD} := V_F \cdot I_{OUT} \cdot F_{SW} \cdot (t_{BDR} + t_{BDF}) \qquad P_{BD} = 0.112\ W$$

Conduction Loss (R_{DS} on = 5.1 mohm and I_{OUT} = 10 A)

$$P_{C\_BOT\_MOS} := I_{OUT}^{2} \cdot R_{DSon} \cdot (1 - D) \qquad P_{C\_BOT\_MOS} = 0.467\,W$$

Reverse Recovery Loss (Q_{RR} = 34 nC)

$$P_{RR} := Q_{RR} \cdot V_{IN} \cdot F_{SW} \qquad P_{RR} = 0.204\,W$$

Gate Drive Loss (R_{G2sink} =0.94Ω, R_{G_drive2} =0Ω , $R_{G2source}$ = 1.3Ω , $R_{G\_Internal}$= 0Ω)

$$P_{G2} := Q_G \cdot V_{GS} \cdot F_{SW} \cdot \left(\frac{R_{G2sink}}{R_{G2sink} + R_{G\_drive2} + R_{G\_Internal}} + \frac{R_{G2source}}{R_{G2source} + R_{G\_drive2} + R_{G\_Internal}} \right)$$

$$P_{G2} = 0.07\,W$$

Hence, the total losses bottom MOSFET's is:

$$TOTAL\_BOT\_MOS\_LOSS := P_{.C\_BOT\_MOS} + P_{.SW\_BOT\_MOS} + P_{.OUT\_BOT\_MOS} + P_{.G2} + P_{.RR} + P_{.BD}$$

Inductor

There are two different types of inductor being used for hardware implementation of the proposed topologies. Since they each has their own ESR value, then the calculation will be separated for each inductor, but their loss equations are equal:

$$P_{inductor} = \overline{I}_L^{\,2} * DCR \qquad (5\text{-}11)$$

Output Capacitor

Depending on the number of output capacitance being used, the ESR loss for each capacitor can be approximated by dividing the rms output current ripple amongst each capacitor.

Loss Calculations

The following shows loss calculations using MathCAD:

Given Parameters:

Input Current rms:	$I_{in.rms} = 5.895A$
Output Current rms:	$I_{o.rms} = 0.22A$
Phase Inductor Current:	$I_{Lphase} = 12.5A$
Auxilliary Inductor Current:	$I_{Laux} := 2 \cdot I_{Lphase}$
Input Capacitor ESR:	$C_{in.ESR} := 0.024\Omega$
Output Capacitor ESR:	$C_{out.ESR} := 0.009\Omega$
Phase Inductor DCR:	$L_{phase.DCR} = 0.003\Omega$
Auxilliary Inductor DCR:	$L_{aux.DCR} := 0.0001\Omega$
Gate Drive Voltage:	$V_{GS} := 12V$
Main MOSFET Rds-on:	$TopMOS_{Rds.on} := 0.0036\Omega$
Main MOSFET rise time:	$TopMOS_{t.rise} := 3.8ns$
Main MOSFET fall time:	$TopMOS_{t.fall} := 4.4ns$
Main MOSFET total gate charge:	$TopMOS_{Qg} := 14nC$
Sync MOSFET Rds-on:	$SyncMOS_{Rds.on} := 0.002\Omega$
Sync MOSFET rise time:	$SyncMOS_{t.rise} := 3.9ns$
Sync MOSFET fall time:	$SyncMOS_{t.fall} := 5.4ns$
Sync MOSFET total gate charge:	$SyncMOS_{Qg} := 27nC$
Body Diode reverse recovery charge:	$SyncMOS_{Qrr} := 34nC$
Body Diode Forward Voltage:	$SyncMOS_{BD.Vf} := 0.7V$
Body diode reverse recovery time:	$SyncMOS_{BD.trr} := 36ns$

Input Capacitor Losses (total of two capacitor in parallel):

$$P_{C.in} := \left[\left(\frac{1}{2} \cdot I_{in.rms}\right)^2 \cdot C_{in.ESR}\right] \cdot 2$$

$$P_{C.in} = 0.417W$$

Main (top) MOSFET losses:

Conduction Loss: $P_{top.cond} := I_{Lphase}^2 \cdot TopMOS_{Rds.on} \cdot D$

$P_{top.cond} = 0.047W$

Switching Loss: $P_{top.sw} := \frac{1}{2} \cdot V_i \cdot I_{Lphase} \cdot f_s \cdot (TopMOS_{t.rise} + TopMOS_{t.fall})$

$P_{top.sw} = 0.308W$

Gate Drive Loss: $P_{top.gd} := TopMOS_{Qg} \cdot V_{GS} \cdot f_s$

$P_{top.gd} = 0.084W$

Total Main MOSFET Loss (for all four MOSFET):

$P_{top} := 4 \cdot (P_{top.cond} + P_{top.sw} + P_{top.gd})$

$P_{top} = 1.754W$

Synchronous (Bottom) MOSFET losses:

Conduction Loss: $P_{sync.cond} := I_{Lphase}^2 \cdot SyncMOS_{Rds.on} \cdot (1 - D)$

$P_{sync.cond} = 0.301W$

Switching Loss: $P_{sync.sw} := \frac{1}{2} \cdot V_i \cdot I_{Lphase} \cdot f_s \cdot (SyncMOS_{t.rise} + SyncMOS_{t.fall})$

$P_{sync.sw} = 0.349W$

Gate Drive Loss: $P_{sync.gd} := SyncMOS_{Qg} \cdot V_{GS} \cdot f_s$

$P_{sync.gd} = 0.162W$

Reverse Recovery Loss: $P_{sync.rr} := SyncMOS_{Qrr} \cdot V_i \cdot f_s$

$P_{sync.rr} = 0.204W$

Body Diode Loss: $P_{sync.BD} := SyncMOS_{BD.Vf} \cdot I_{Lphase} \cdot f_s \cdot SyncMOS_{BD.trr}$

$P_{sync.BD} = 0.176W$

Total Bottom MOSFET Loss (for all four MOSFET):

$$P_{sync} := 4 \cdot (P_{sync.cond} + P_{sync.sw} + P_{sync.gd} + P_{sync.rr} + P_{sync.BD})$$

$$P_{sync} = 4.764W$$

Phase Inductor Losses:

$$P_{Lphase} := 4 \left(I_{Lphase}^{2} \cdot L_{phase.DCR} \right)$$

$$P_{Lphase} = 1.938W$$

Auxilliary Inductor Losses:

$$P_{Laux} := 4 \cdot \left(I_{Laux}^{2} \cdot L_{aux.DCR} \right)$$

$$P_{Laux} = 0.425W$$

P_{Laux} for topology 2 = (P_{Laux} topology1) / 2 = 0.212 W

Output Capacitor Losses (total two capacitors put in parallel):

$$P_{C.out} := 2 \cdot \left[\left(\frac{1}{2} \cdot I_{o.rms} \right)^{2} \cdot C_{out.ESR} \right]$$

$$P_{C.out} = 2.177 \times 10^{-4} W$$

Total Circuit Losses:

$P_{Loss} := P_{C.in} + P_{C.out} + P_{Lphase} + P_{Laux} + P_{top} + P_{sync}$

$P_{Loss} = 9.297W$ *(for topology 1)*

P_{Loss} for topology 2 (half the losses on P_{Laux}) = 9.297 – 0.212 = 9.085 W

Efficiency:

$P_{out} := V_o \cdot I_c$ $P_{out} = 50W$

$$\eta\% := \frac{P_{out}}{P_{out} + P_{Loss}} \cdot 100$$

$\eta\% = 84.321$ (for topology 1) $\eta\% = 84.624$ (for topology 2)

As shown on the above results, the estimated efficiency for Topology 1 is 84.32% and for Topology 2 is 84.624%. An efficiency plot for different load condition is plotted on figure 5-3 for Topology 1.

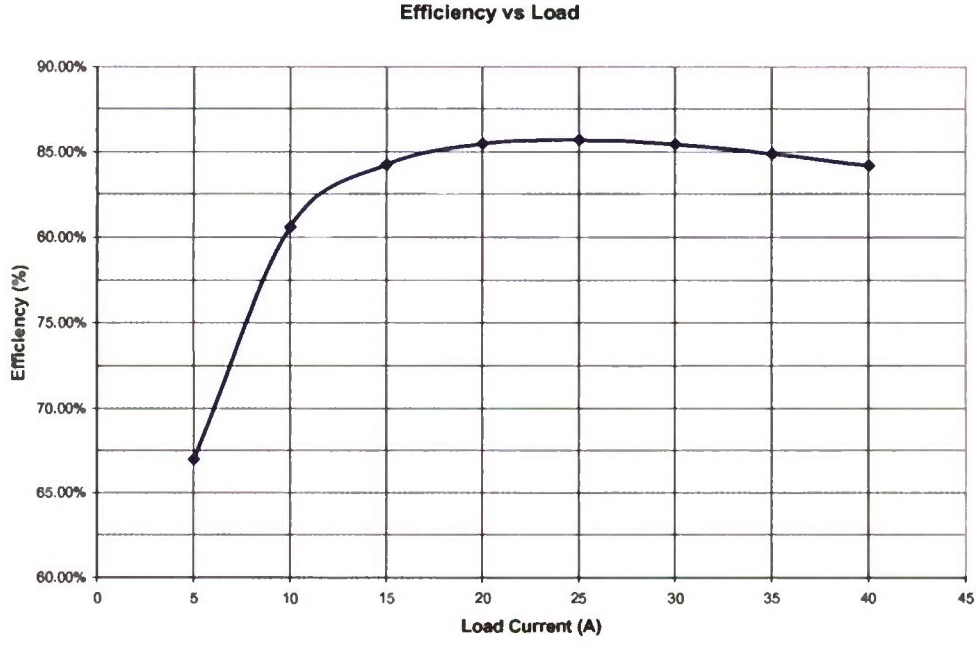


Figure 5.3 Efficiency vs. Load Current for Topology 1

6. Simulation Results

In order to measure the performance of the proposed multiphase buck converter, a simulation test was performed using OrCAD Schematic. The simulations were done to simulate the open loop response of the proposed multiphase buck converters.

The MOSFET characteristic in the actual hardware circuit can be tested using MOSFET SPICE model provided by Fairchild Semiconductor. A model block has to be made in OrCAD in order to integrate these MOSFET characteristic into simulation. The reason for this is that the SPICE model is given in text mode.

Each input and output capacitor's ESR (Equivalent Series Resistance) along with inductor's DCR (Direct Current Resistance) are also modeled in this simulation in order to get a good representation of the actual hardware circuit performance. The verification of power losses in the converter can be performed by the inclusion of these parasitic components.

Once the circuit schematic is completed, a transient analysis in OrCAD is run for at least 1 millisecond long to ensure steady state operation of the circuit has been reached. Circuit's performance can then be studied by obtaining plots of critical parameters such as output voltage, inductor currents, output voltage ripple, etc.

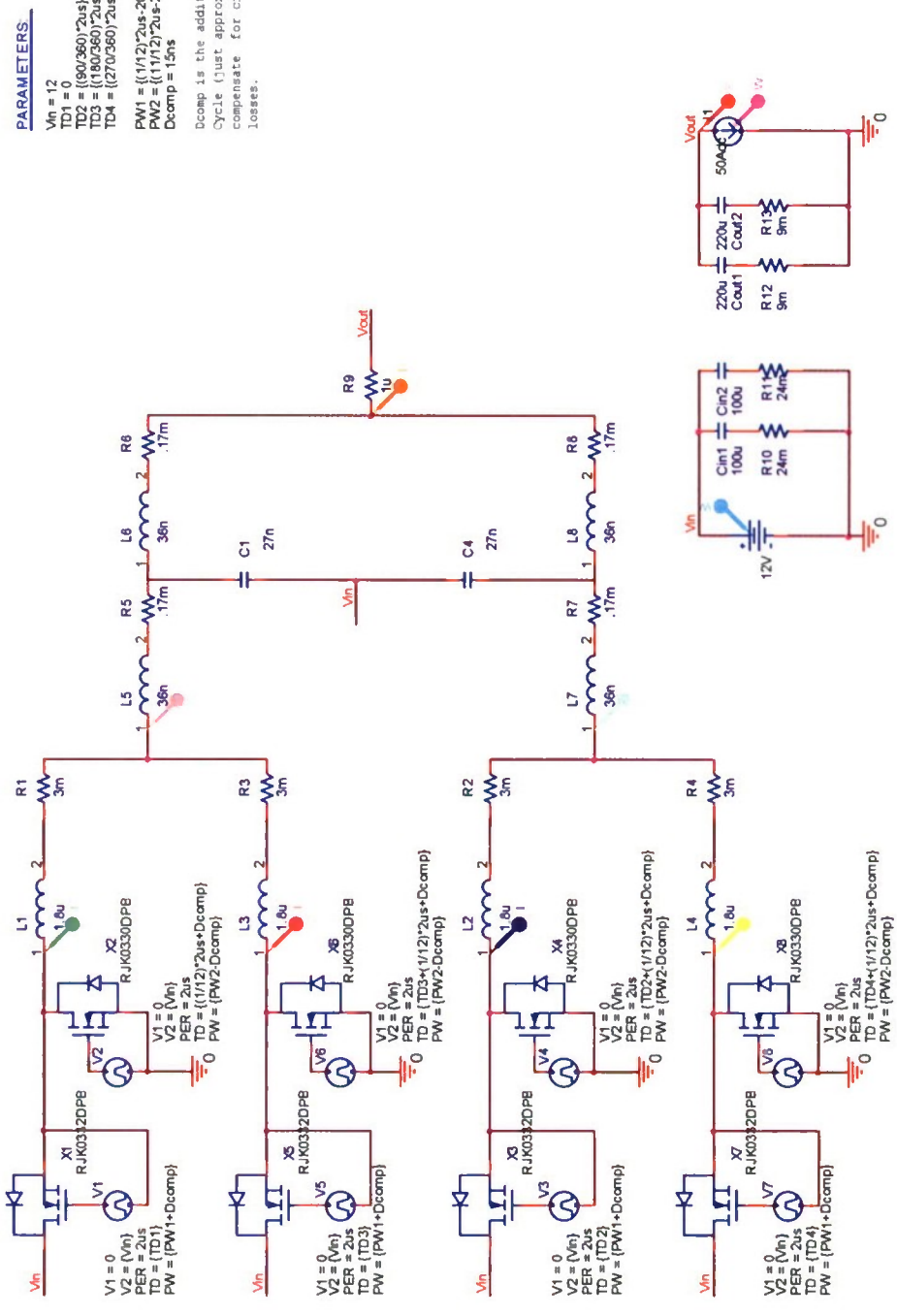


Figure 6.1 Circuit Schematic for Topology 1

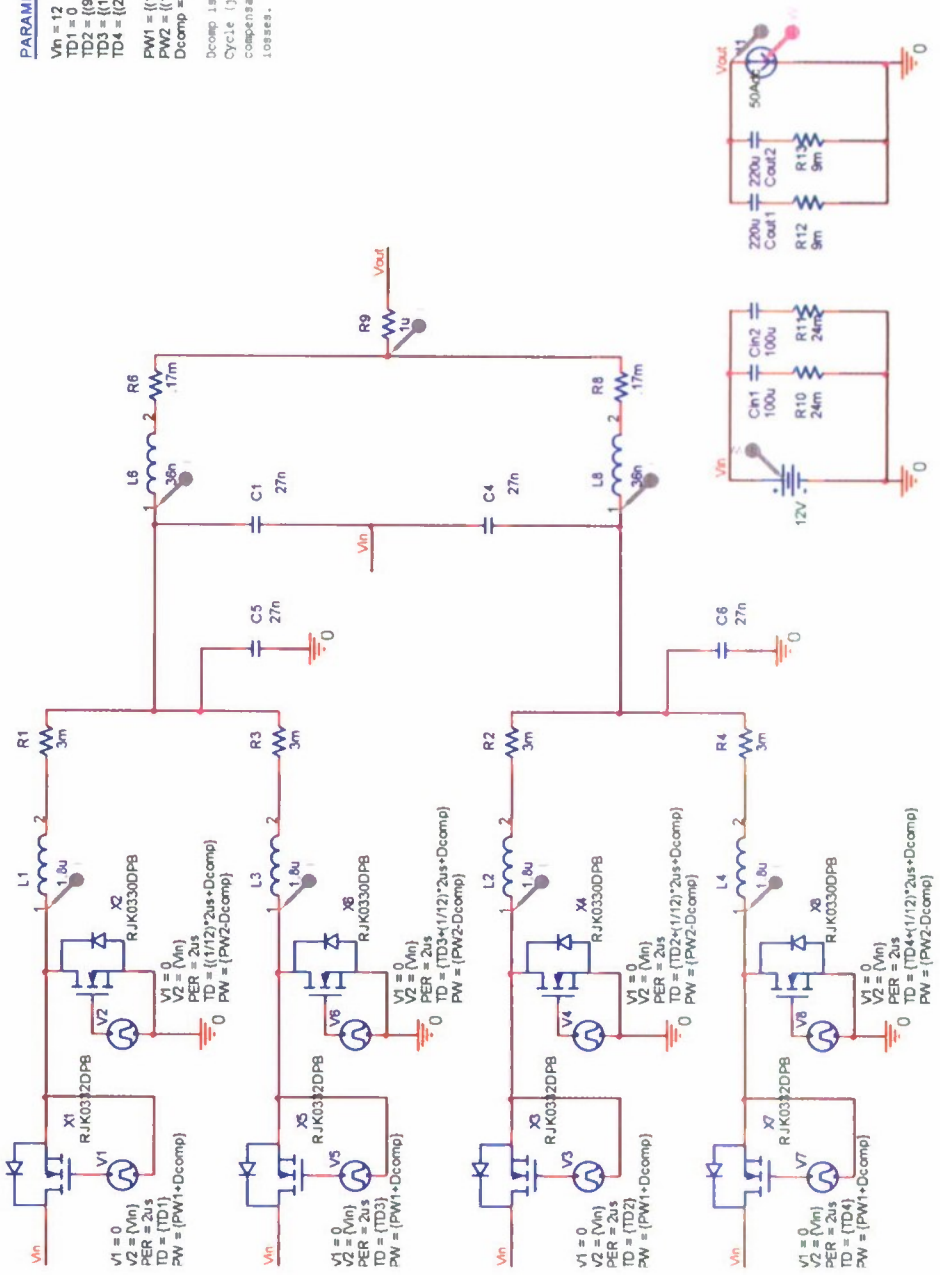


Figure 6.2 Circuit Schematic for Topology 2

Figure 6-1 and 6-2 show the completed schematics for the proposed converters. There are four phases where each two phases forms a single block or a module. The MOSFETs model representation of FDS8690 (Top MOSFET) and FDS6299S (Bottom MOSFET) are shown in form of a block with a reference designator from U1 to U8. The electrolytic capacitors ESR are also modeled in this schematic by the addition of R11, R12, R13, and R14 in series with input and output capacitor. Other capacitors ESR are not modeled since these capacitor are ceramic capacitor which has a relatively low ESR (5mΩ or less). Therefore, its contribution to power loss would be insignificant. The DCR for the main (L1 – L4) and auxiliary (L5 – L8) inductors are also modeled using (R3 – R6) and (R7 – R8) respectively. The values for these ESR and DCR can be found through each component datasheet. Another resistor R2 is added when the two modules are combined at the output. This resistor is being used to put a current probe at the output terminal to verify current ripple cancellation factor at the output. As for the load, a constant current of 40 A is used to sink the output current.

A parameter tool is provided by OrCAD to simplify some expressions for the gate drive signals of the MOSFET. Any variables that are listed under parameters will act as a global variables which can be used anywhere in the schematic.

The simulation is run under open loop condition due to the lack of feedback controller. This is the reason why the output voltage is not exactly 1 V as can be seen later on in this chapter. The main purpose for the controller is to maintain the output voltage of 1 V regardless of losses between input and output.

Current Waveforms

One benefit of multiphasing is current ripple cancellation. Therefore each inductor current needs to be checked in order to fully implement the cancellation factor on each stage. Inductor current for each phase has an average value of 40 A/ 4 = 10 A and a peak to peak ripple value of 0.991 A. From Figures 6-3 and 6-6, the switching frequency for each phase can also be calculated and its value is equal to $\frac{1}{(974.004us - 972us)} = 500$ kHz as expected.

Phase shifting between phase 1 and 2 is verified by subtracting time difference between peak inductor current of phase 1 and 2. The result is divided by a period of 2us and then multiplied by 360 degrees. In this particular case, the shift would then be $\frac{(972.67us - 972.17us)}{2us} * 360^0 = 90^0$. The current from L1 and L2 are combined together resulting in currents flowing on auxiliary inductors L5 and L7.

A current ripple cancellation effect can be seen by a reduced value of peak to peak current ripple at this particular stage. Peak to peak current ripple is now 0.929 A instead of 0.991 as seen on main inductors of L1, L2, L3 and L4.

At the end, the current from each second stage combines together at the output. These two currents will add together and lower the peak to peak output current ripple due to cancellation factor at each stage as illustrated in Figures 6-5 and 6-8.

Figure 6-5 and 6-8 show that the final peak to peak current ripples are around 1 A. Comparing these final values with the calculated value suggest a small difference was observed. This discrepancy is mainly caused by an output voltage that is slightly less than 1 V on the simulation. With an input voltage of 12 V and an actual output voltage of 0.961

yields a duty cycle of 0.961 / 12 = 0.080 instead of 0.0833. Therefore, a higher current ripple should be observed at the output.

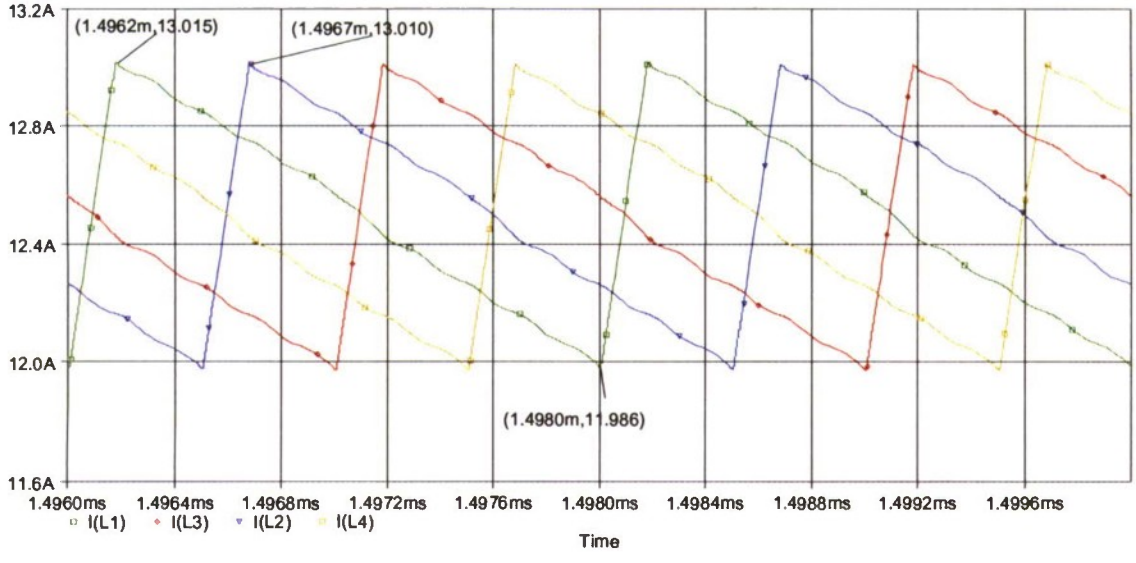


Figure 6.3 Phase Inductors Current of Topology-1 showing Interleaving

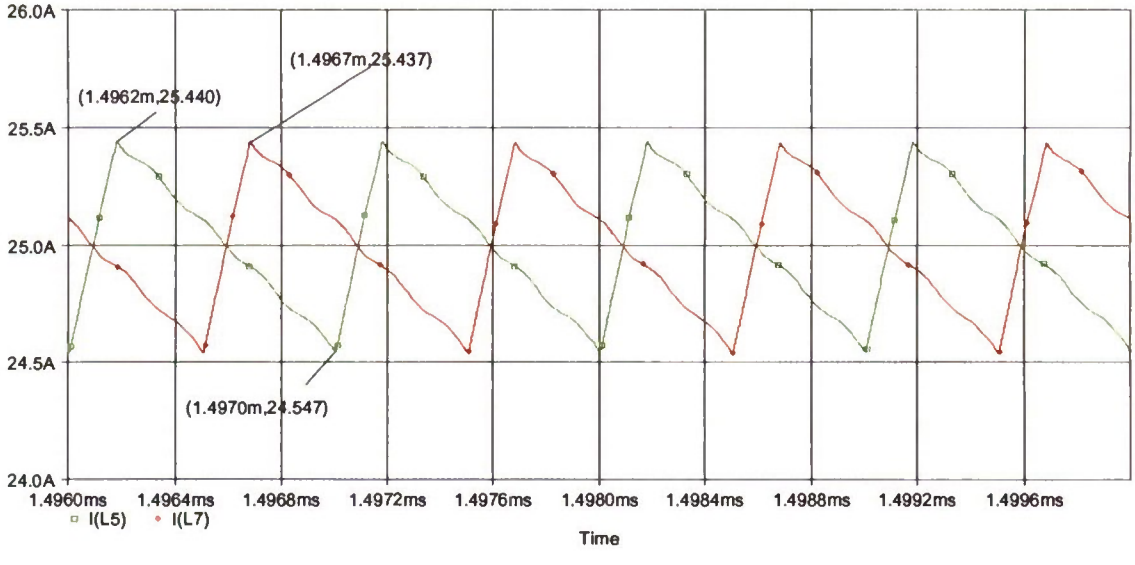


Figure 6.4 Auxiliary Inductors Current of Topology-1

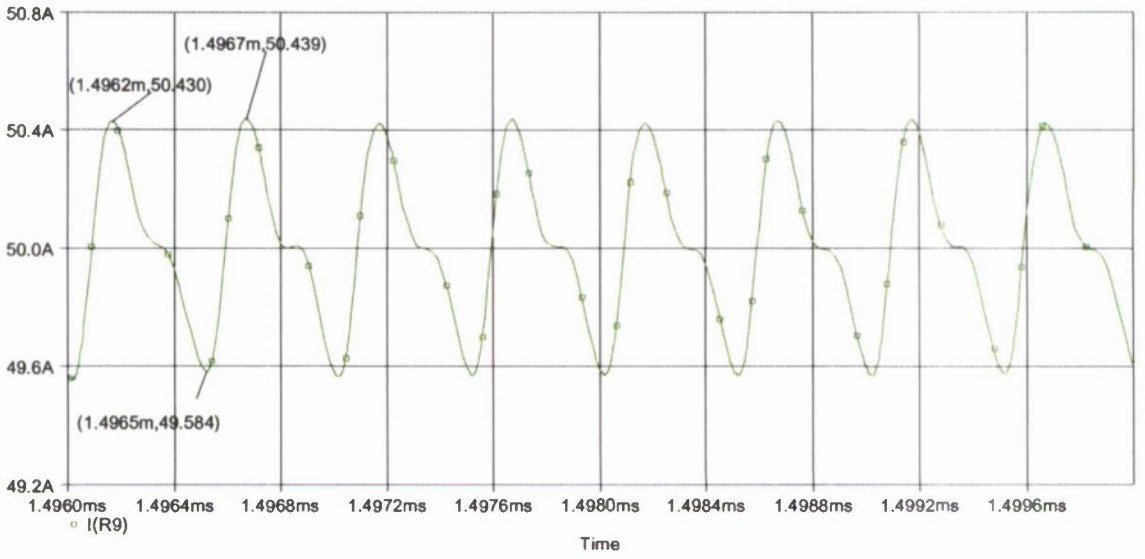


Figure 6.5 Output Current Waveform of Topology-1

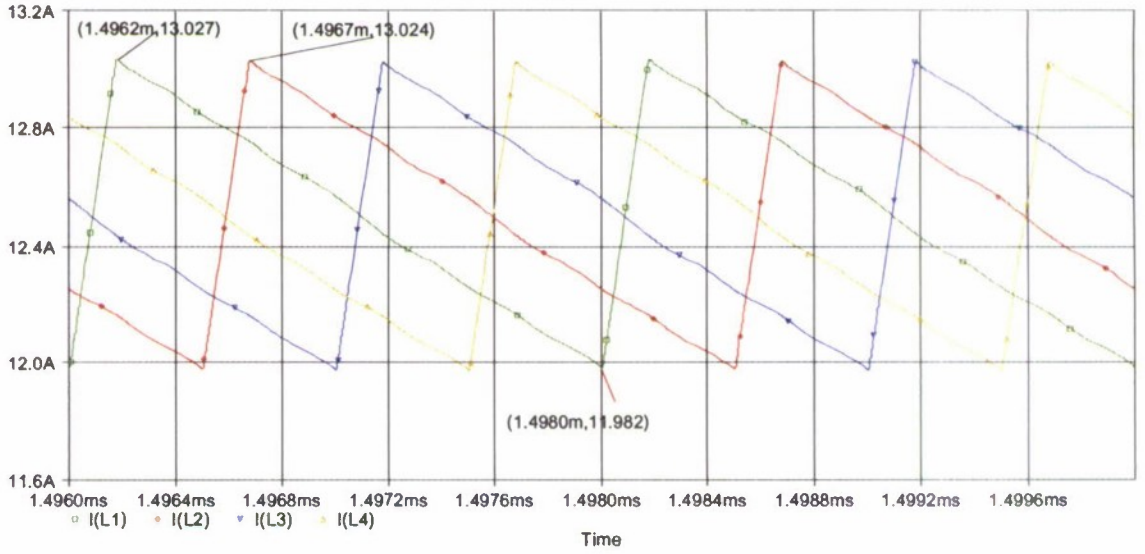


Figure 6.6 Phase Inductors Current of Topology-2 showing Interleaving

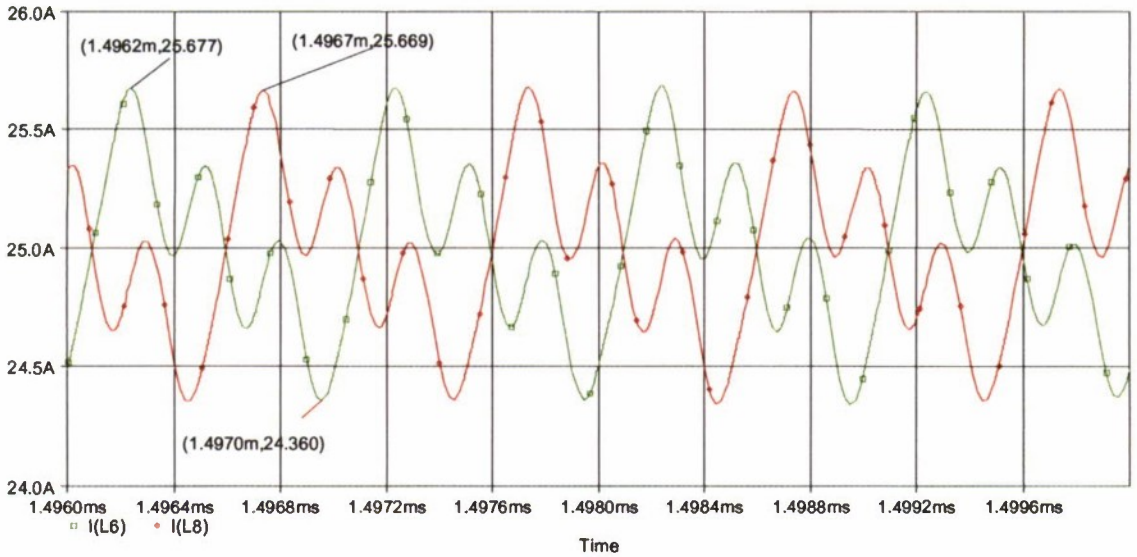


Figure 6.7 Auxiliary Inductors Current of Topology 2

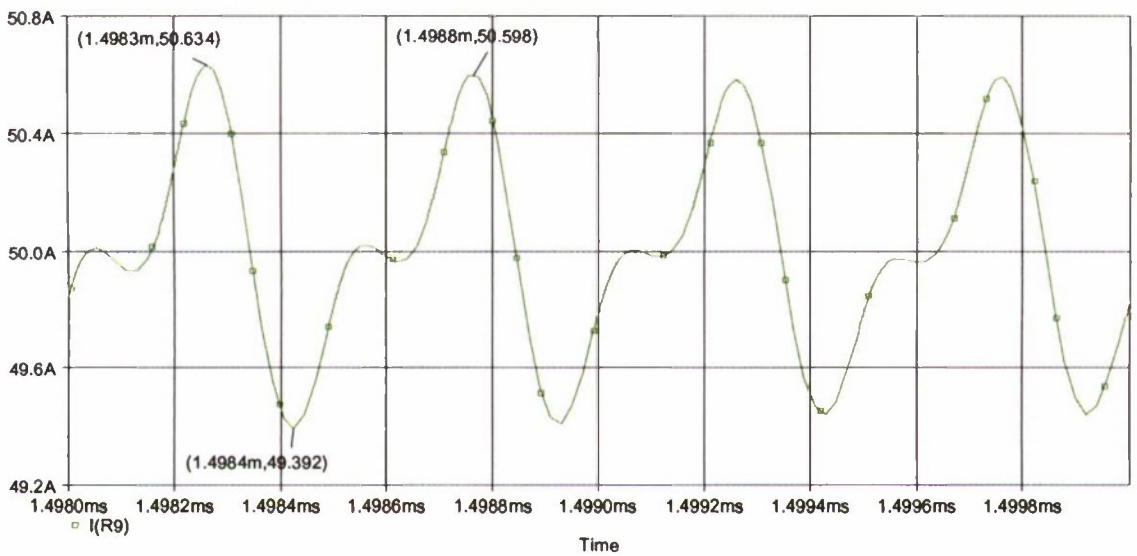


Figure 6.8 Output Current Waveform of Topology-2

Output Voltage Ripple

After a schematic has been built and simulated, the performance of the circuit is then measured. Based on Figure 6-9, it can be seen that the average output voltage at full load is 0.97 V for Topology 1 and 0.98 V for Topology 2 with a peak to peak output voltage ripple value of 4 mV and 5 mV respectively which are less than the value specified on our design parameters of 15 mV. The output voltages are lower than the expected value of 1 V because the simulations were run without feedback controller. Therefore the losses caused by inductor's series resistance and MOSFET's on resistance are not compensated which causes an output voltage drop.

The output ripple frequency can also be calculated using Figure 6-9 and its value is equal to $\frac{1}{(981.85us - 980.685us)} = 2$ MHz; which is equal to the expected value of 4 * 500 kHz = 2 MHz since the converters have 4 phases with switching frequency of 500 kHz per phase.

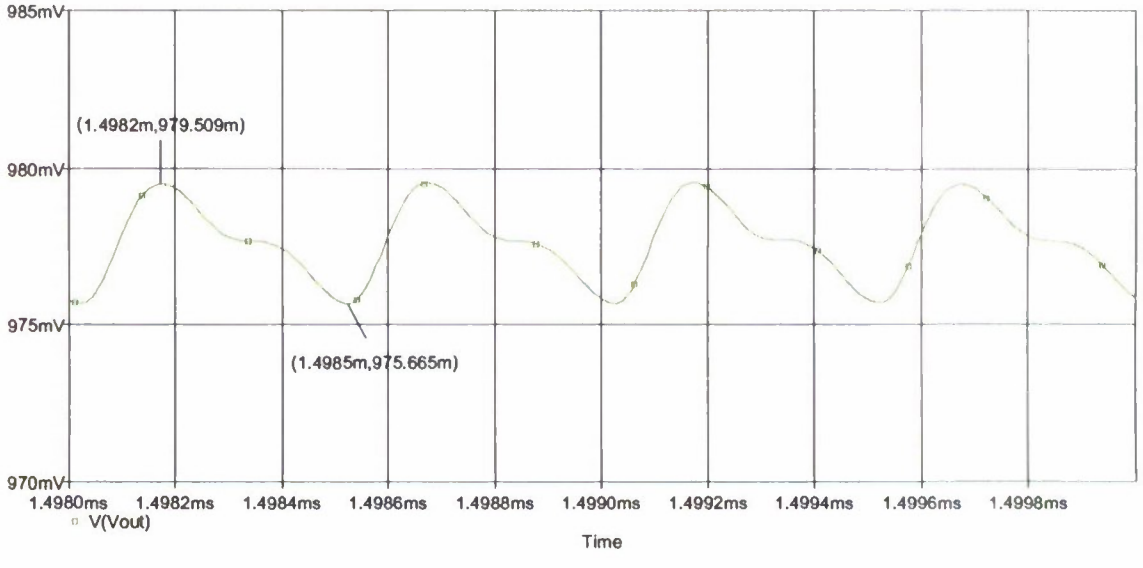


Figure 6.9 Output Current Waveform of Topology-2

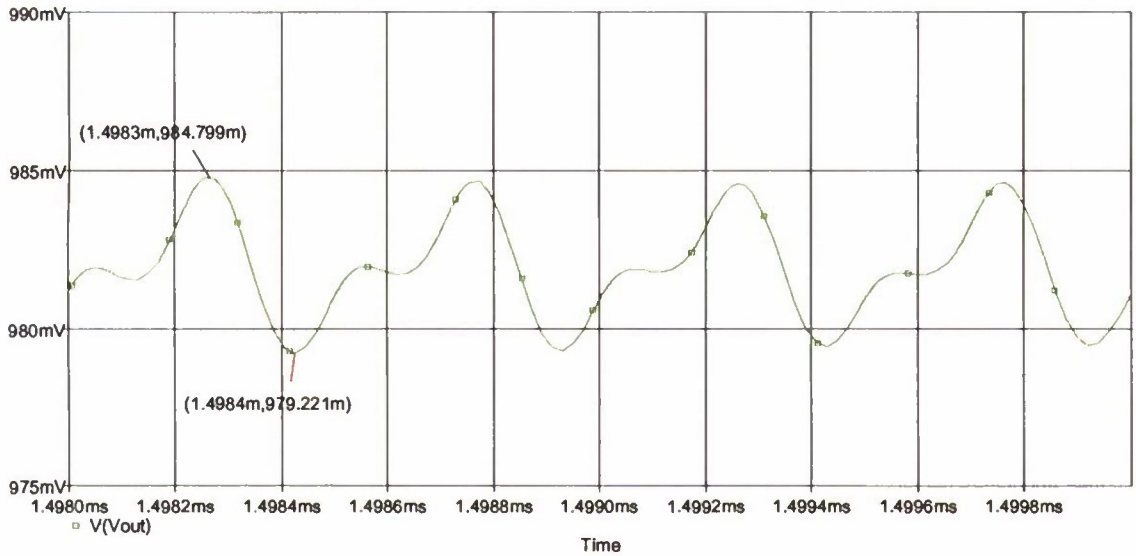


Figure 6.10 Output Current Waveform of Topology-2

Power Loss and Efficiency

One last performance measures are the main component losses and the overall efficiency of the converters.

Figure 6-11 shows the top and bottom MOSFET losses of Topology 1 of 0.27 W and 0.82 W respectively. The bottom MOSFETs have higher loss since the carry most of the output current due to the small duty cycle. The same phenomenon can also be observed for Topology 2, where the top and bottom MOSFET losses are found to be 0.27 W and 0.82 W respectively. These are the same numbers as those found for Topology 1.

Another main component in the circuit is the inductor. For both topologies, the power loss associated with the main and auxiliary inductors were simulated. From Figure 6-12 the main and auxiliary inductors for Topology 1 have losses of 0.4 W and 0.1 W respectively. For Topology 2, the inductors' losses are observed to be 0.47 W and 0.1 W which are pretty much the same as those found in Topology 1.

The overall efficiencies of the converters are shown in Figures 6-14 and 6-18 for Topology 1 and Topology 2 respectively. In both cases, the efficiencies are found to be around 90%. This is higher than the calculated values of about 84%. The higher value is expected since the circuit models used in the simulation did not incorporate the complete real models in all of the components, and hence not all losses in the components are accounted for.

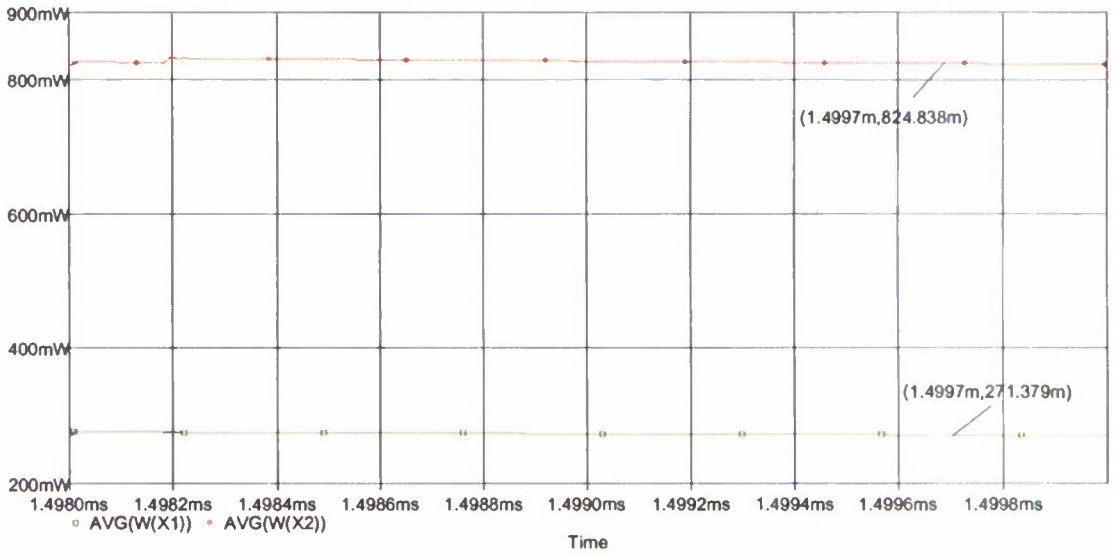


Figure 6.11 Losses of Main (green/lower) & bottom MOSFETs (red/upper) for Topology 1.

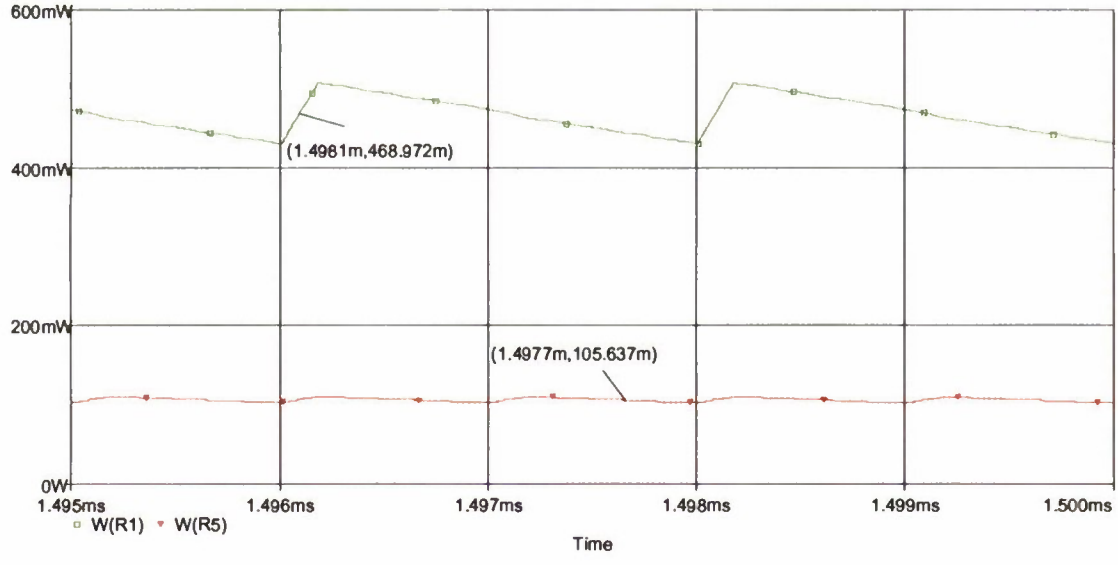


Figure 6.12 Losses in Phase (upper/green) and Auxiliary Inductors (lower/red) for Topology-1

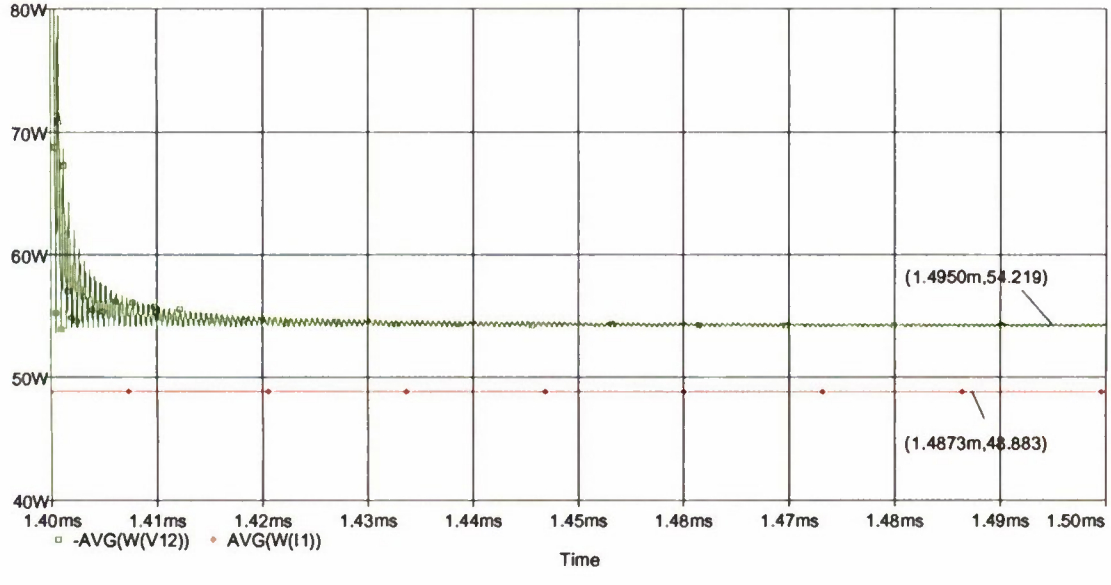


Figure 6.13 Topology-1 Input (upper/green) and Output Power (lower/red) Simulation Result.

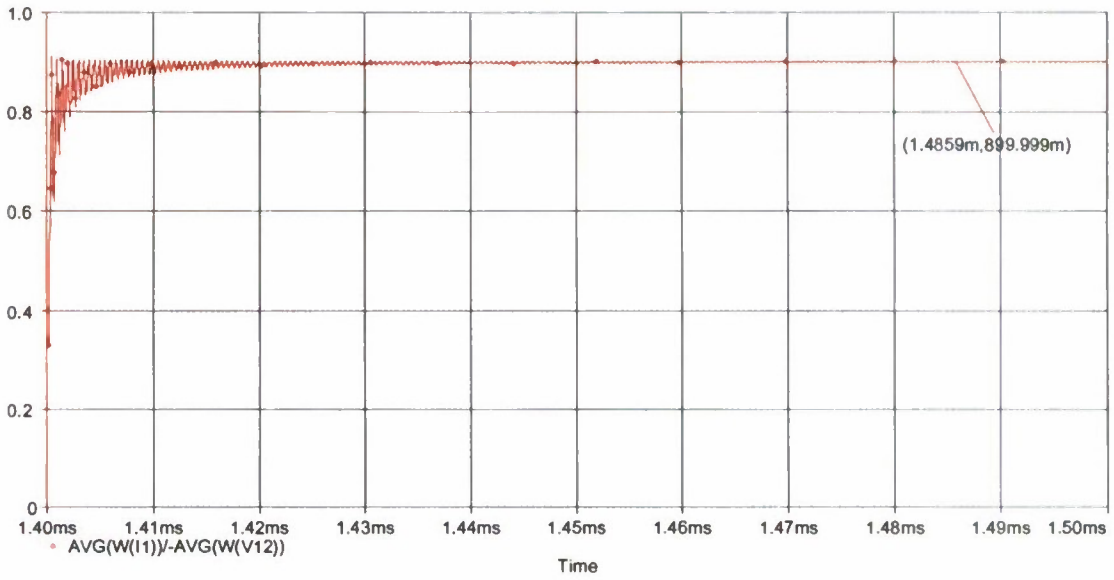


Figure 6.14 Topology-1 Efficiency (89.99%) based on Simulation.

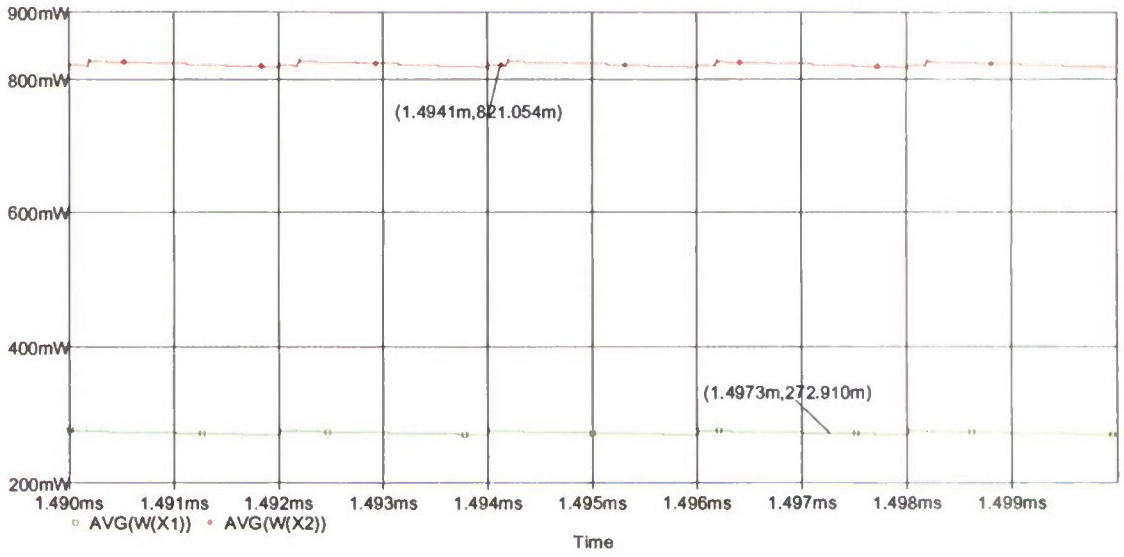


Figure 6.15 Power loss in Main MOSFET (green/lower) and Synchronous MOSFET (red/upper) for Topology 2.

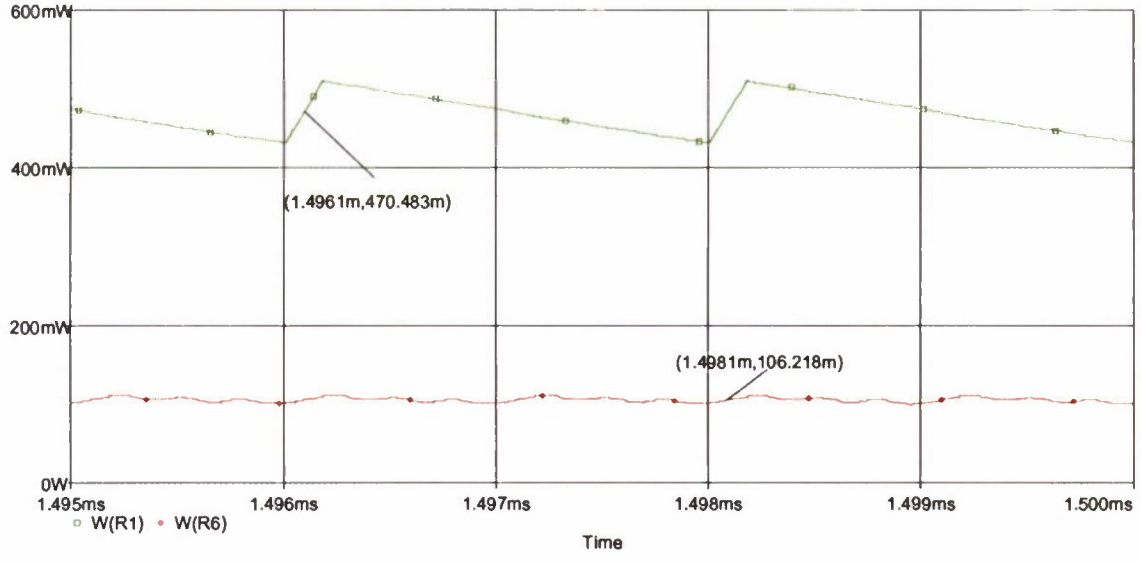


Figure 6.16 Power loss in Phase Inductor (upper/green) and Auxiliary Inductor (lower/red) for Topology-2

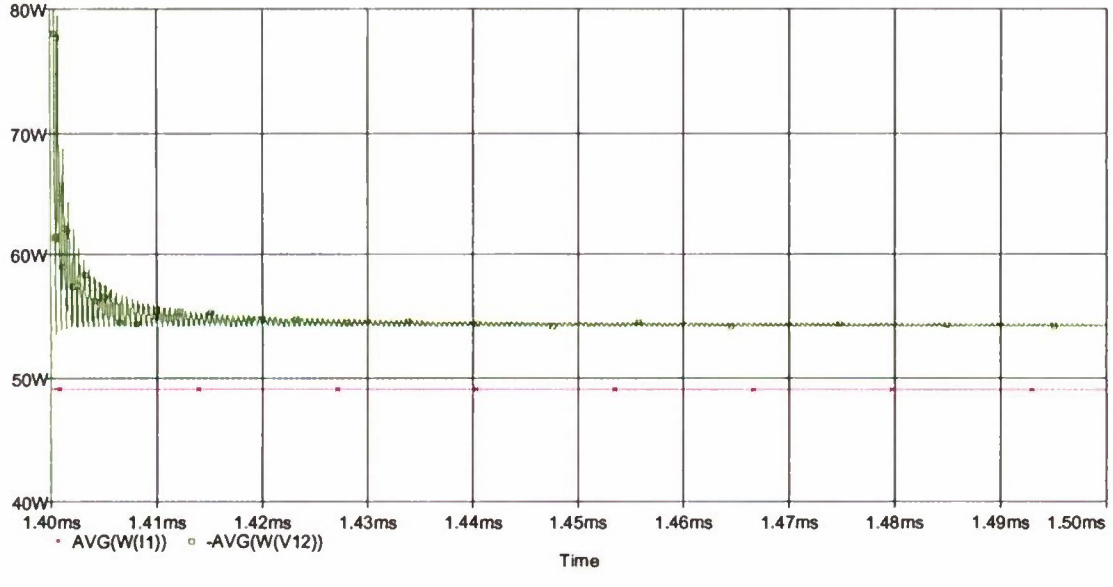


Figure 6.17 Topology-2 Input and Output Power Simulation Result.

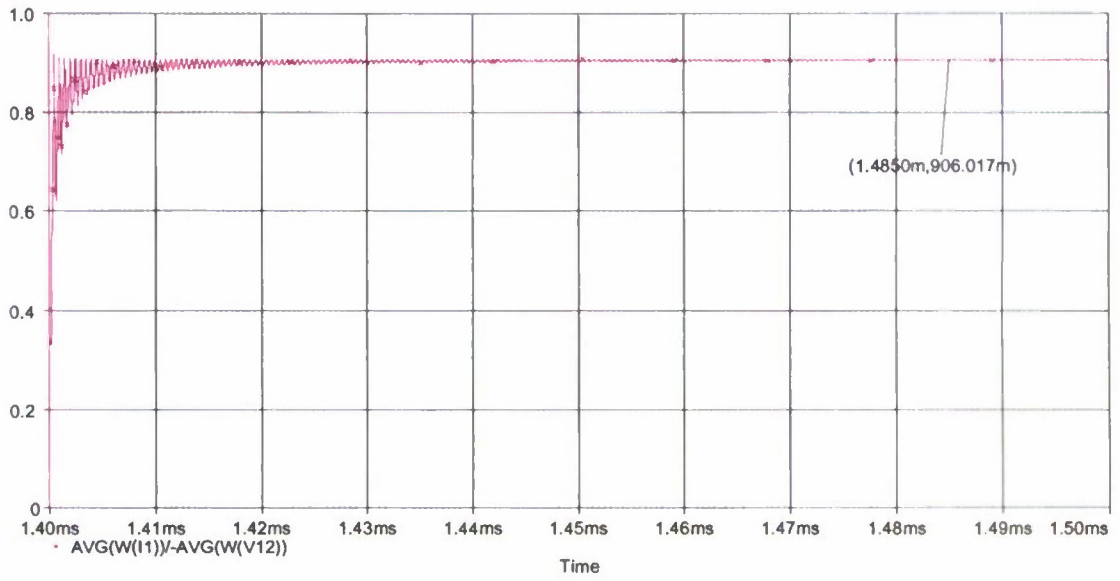


Figure 6.18 Topology-2 Efficiency (90.6%) based on Simulation.

7. Hardware Results

The hardware prototypes for the new topologies was designed and built. To be consistent with the design and loss calculations as well as to enable us to compare with a commercially available converter, a 4 phase version of the proposed topology was developed and then tested. Each phase is running at 500 kHz switching frequency which makes both input and output components to have frequency component of 4 x 500 kHz = 2 MHz. The prototypes were done on a multi-layer pcb, approximately 2.5 in. x 2.5 in. The top layer was dedicated for all the controller chips while the bottom layer was used specifically for the power components (inductors, MOSFETs). Laboratory tests were then conducted on the prototypes to assess their performance on several standard dc-dc operating parameters. Results were then compared to those obtained from a commercially available VRM.

7.1. PCB Layout and Design

The first step on pcb design is to build circuit schematic for the converters. The schematics and pcb designs were done using the software from expresspcb. There are

actually a couple of common portions of the circuit in the proposed converters. In particular, the PWM controller and PWM driver sections are the same in both converters. Hence, the same schematics for these sections were being used for both converters. These are shown in Figures 7-1 and 7-2.

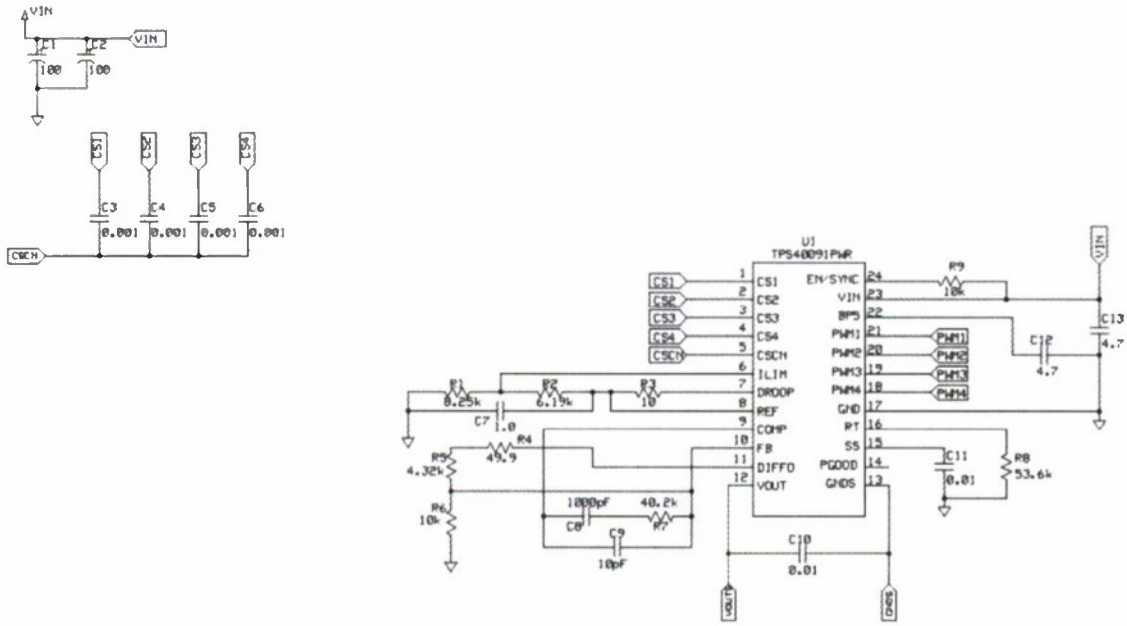


Figure 7.1 PWM Controller Schematic of First and Second Topology

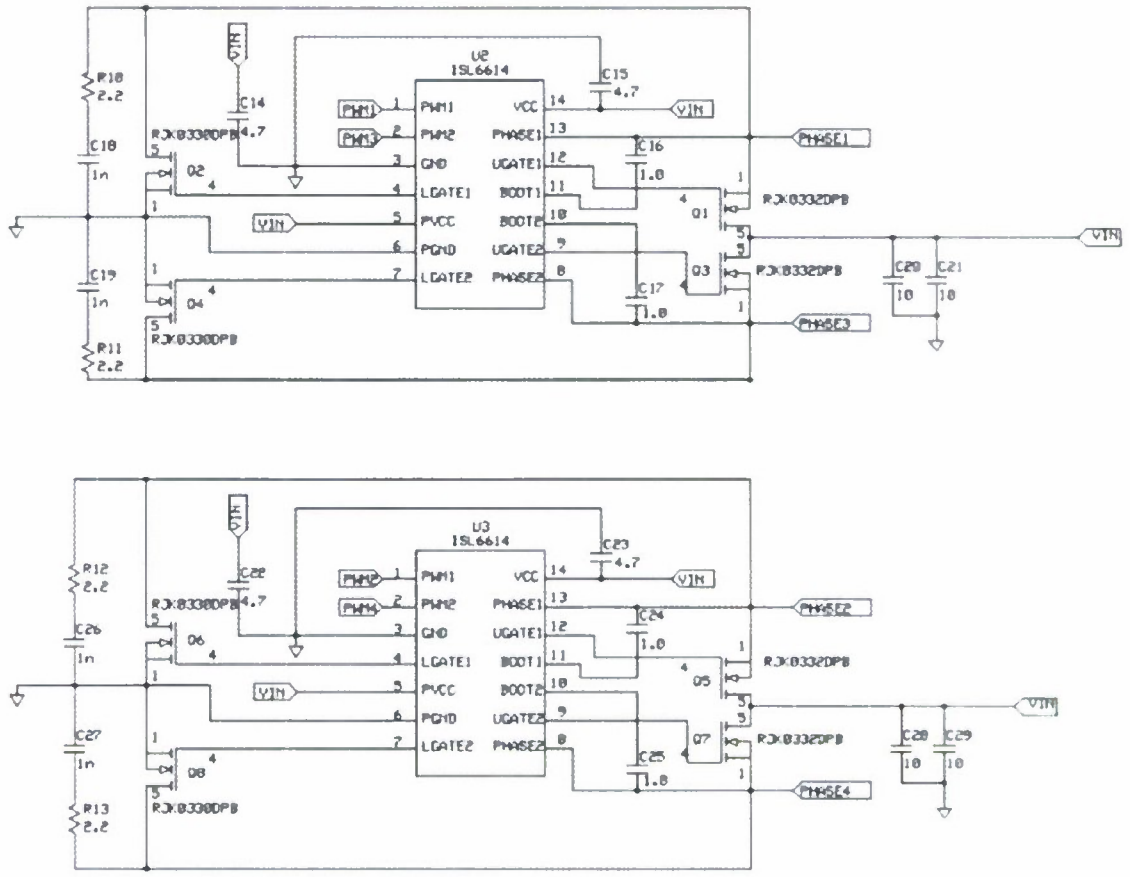


Figure 7.2 MOSFETs and MOSFET Drivers Schematic of First and Second Topology

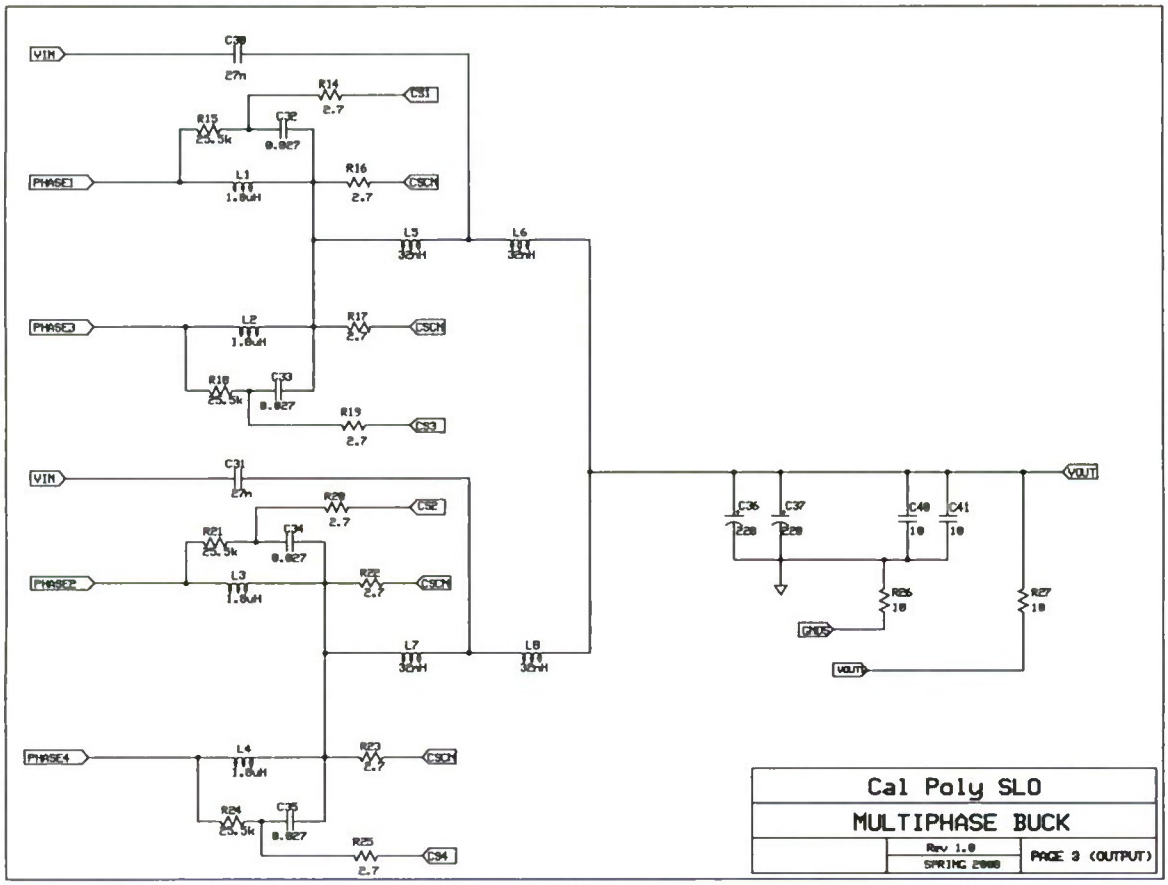


Figure 7.3 Output Stage Schematic of the First Topology

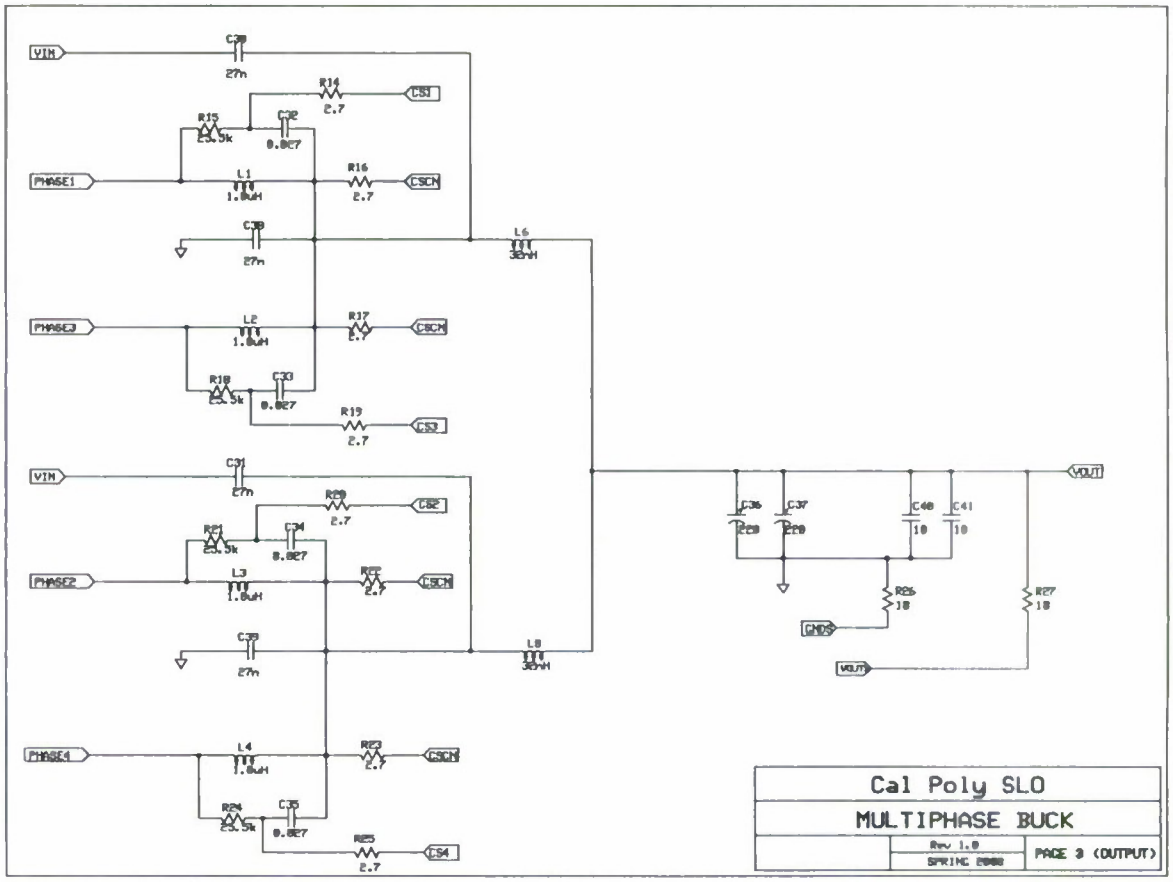


Figure 7.4 Output Stage Schematic of the Second Topology

As illustrated in Figures 7-3 and 7-4, the output stage of the new topologies is different from each other as expected.

Once schematics were completed, board layouts for the two converters were then designed and developed. Figure 3 depicts the overall pcb layout for both converters. The pcb itself consists of four layers: top layer, bottom layer, ground inner layer, and power inner layer. These are shown in Figures 6 to 9.

The finished hardware prototypes have physical dimension of approximately 2.5 in. x 2.5 in. Figures 10 and 11 show the finalized pcbs for Topology 1 and Topology 2 respectively.

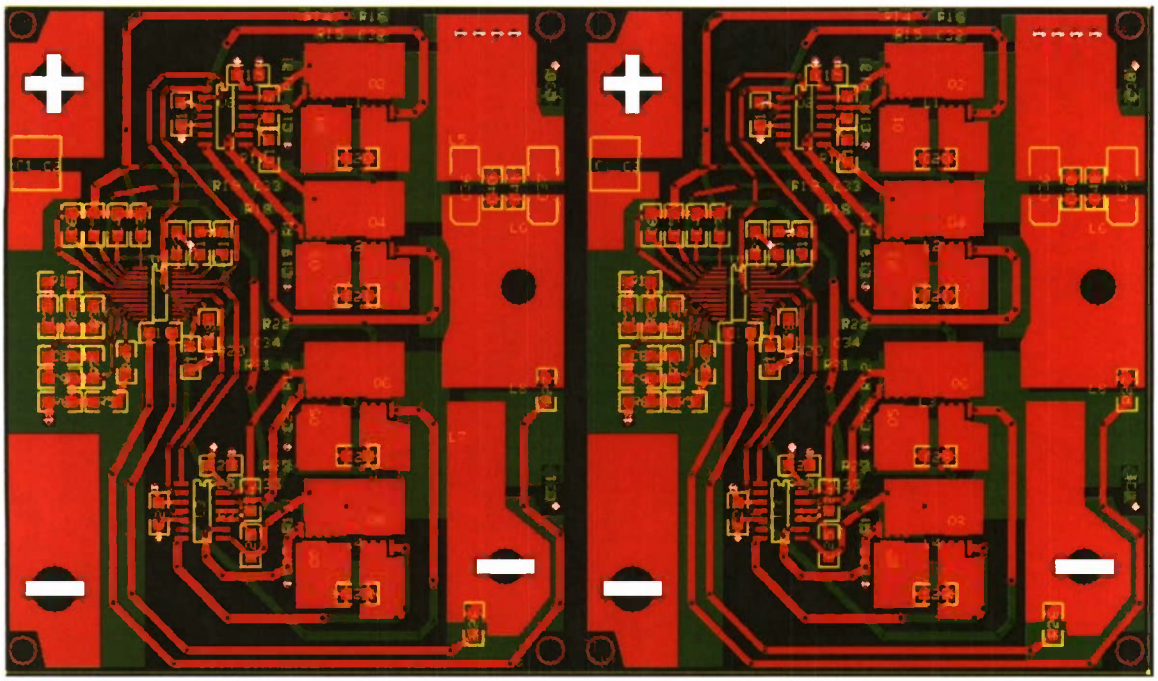


Figure 7.5 Overall PCB for First Topology (left) and Second Topology (right)

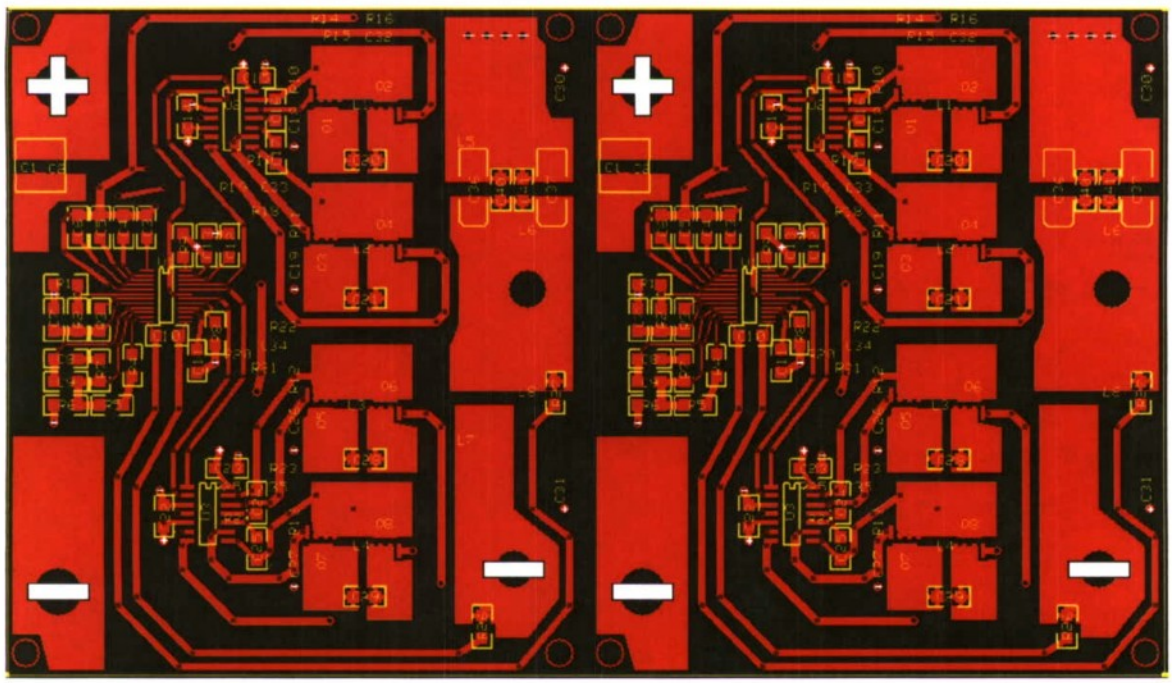


Figure 7.6 Top Layer PCB for First Topology (left) and Second Topology (right)

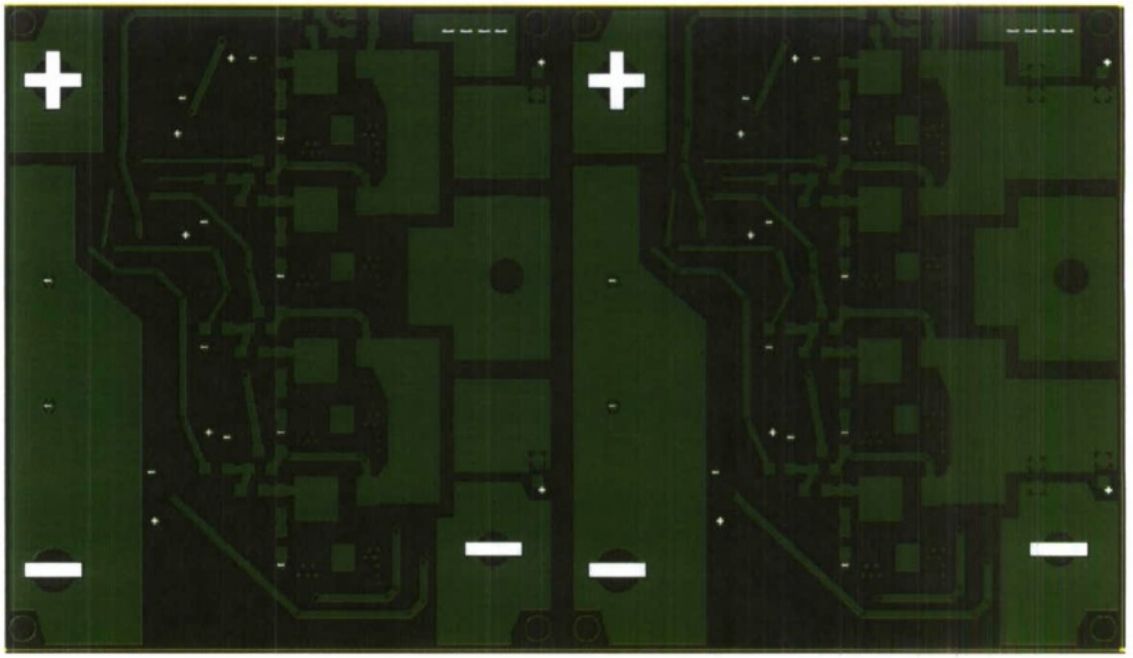


Figure 7.7 Bottom Layer PCB for First Topology (left) and Second Topology (right)

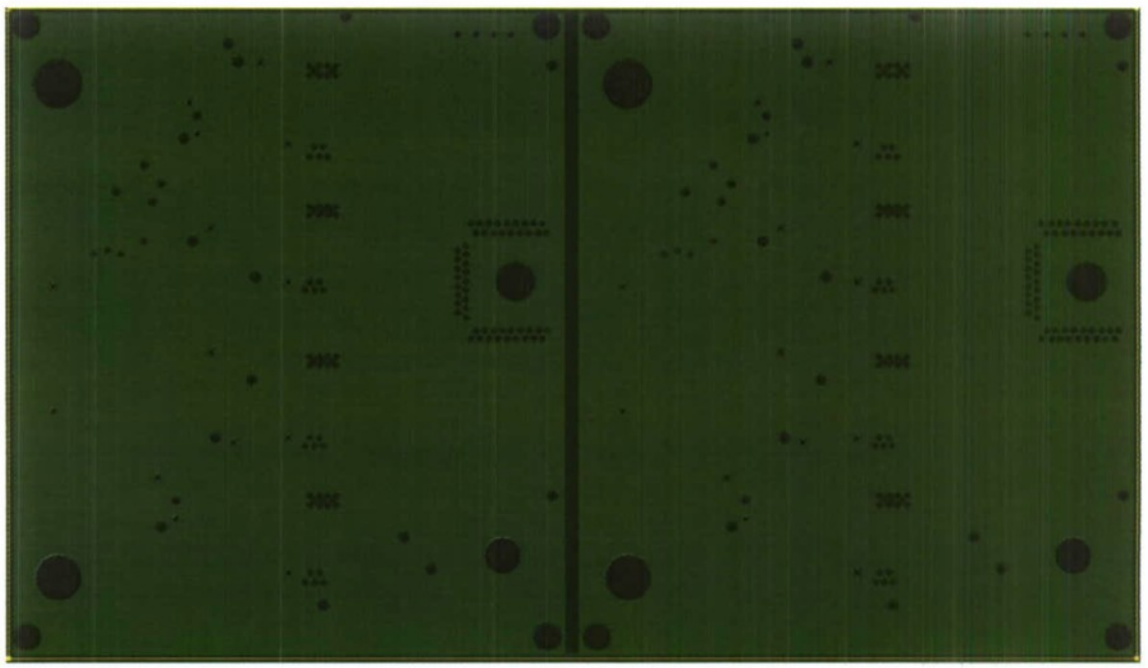


Figure 7.8 Ground Inner Layer PCB for First Topology (left) and Second Topology (right)

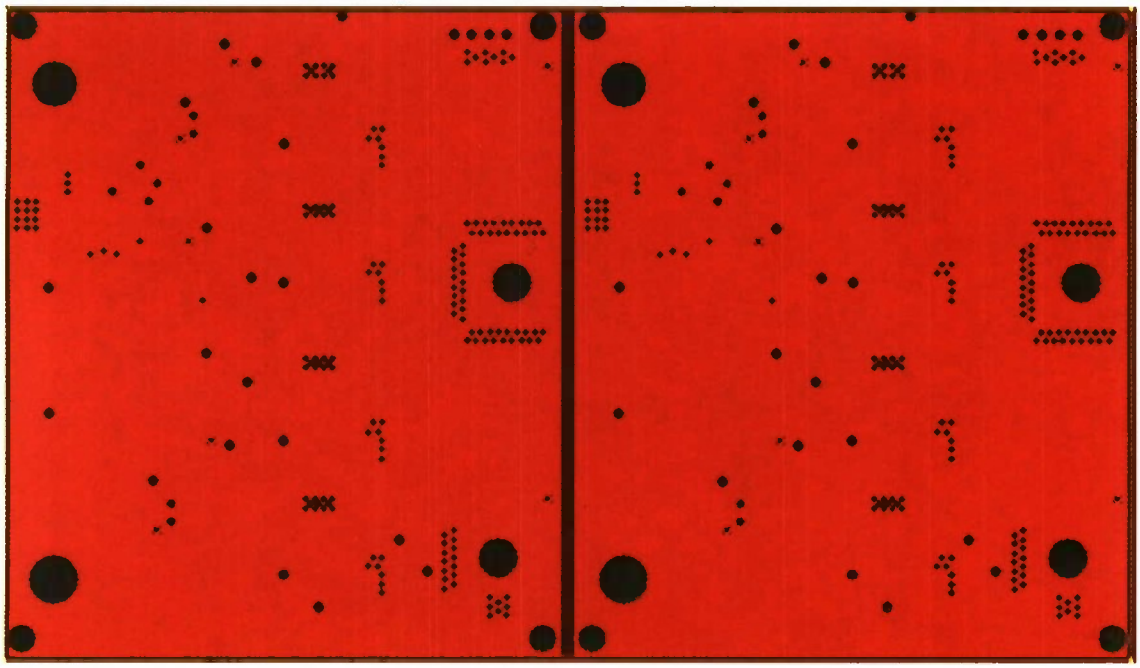


Figure 7.9 Power Inner Layer PCB for First Topology (left) and Second Topology (right)

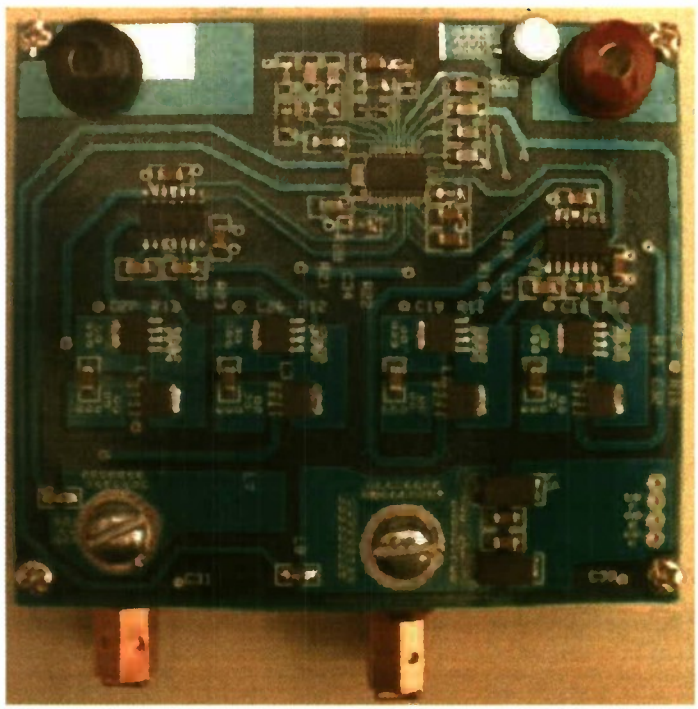


Figure 7.10 Finalized PCB of Topology 1 – top layer

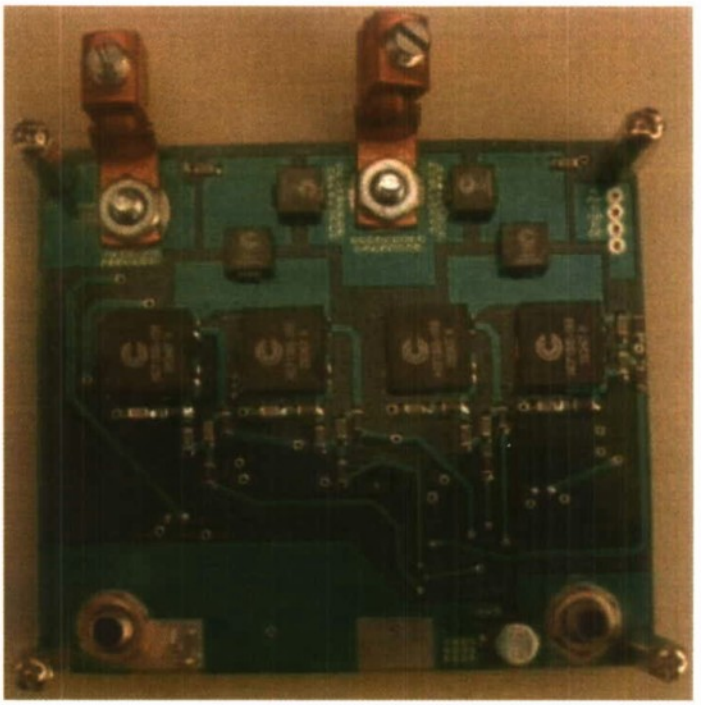


Figure 7.11 Finalized PCB of Topology 1 – bottom layer

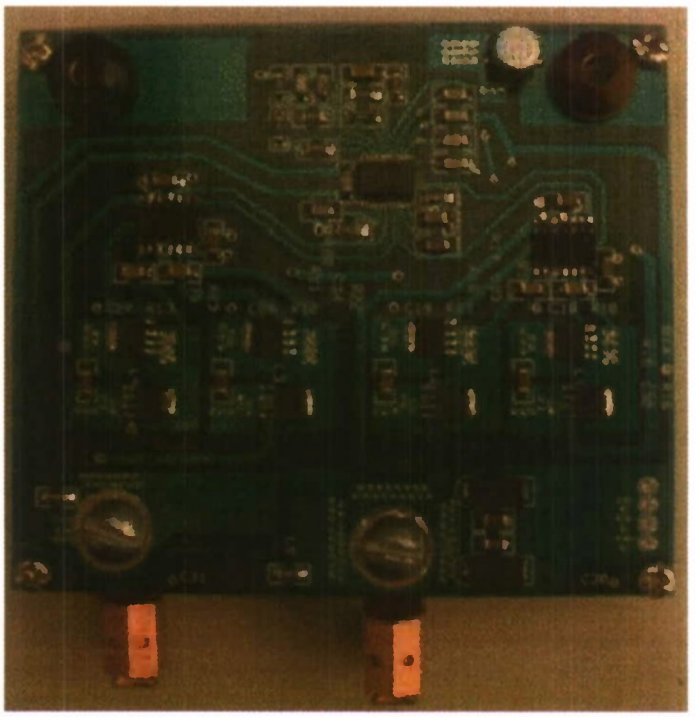


Figure 7.12 Finalized PCB of Topology 2 – top layer

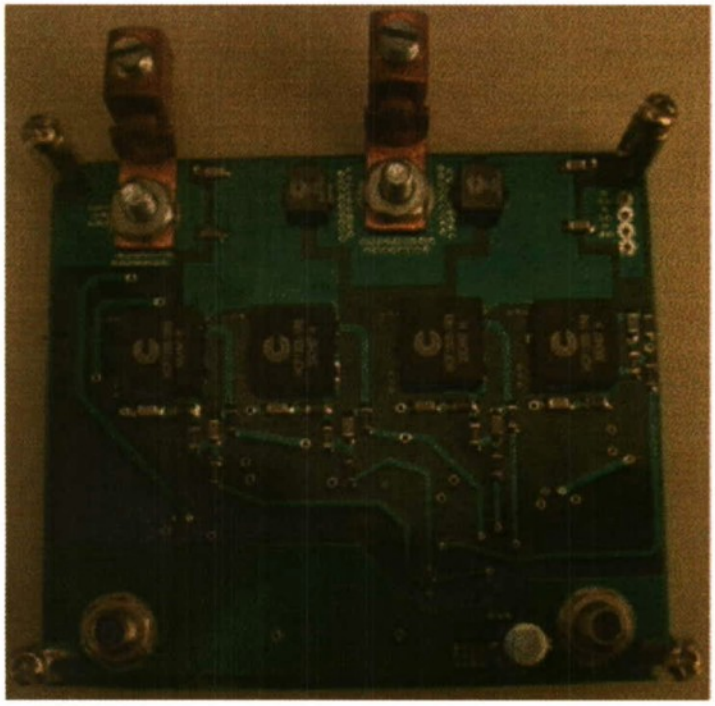


Figure 7.13 Finalized PCB of Topology 2 – bottom layer

7.2.Output Voltage Ripple

The peak to peak output voltage ripple was observed to be approximately 8.2 mV for Topology 1 and 10.8 mV for Topology 2, both measured at full load. These are shown in Figures 14 and 15. These peak to peak ripples are considerably less compared to that of the commercially available VRMs measured to be 40 mV, see Figure 16. However, the output voltage of the proposed converter appears to have so much high frequency noise on top its actual peak to peak ripple. This may be explained by the fact that the frequency component of the output voltage is relatively high at 2 MHz (4 x 500 kHz). Hence, a better layout and/or filtering will be necessary to suppress this high frequency noise. The switching frequency in both output voltage from Topology 1 and Topology 2 are observed to be 2 MHz.

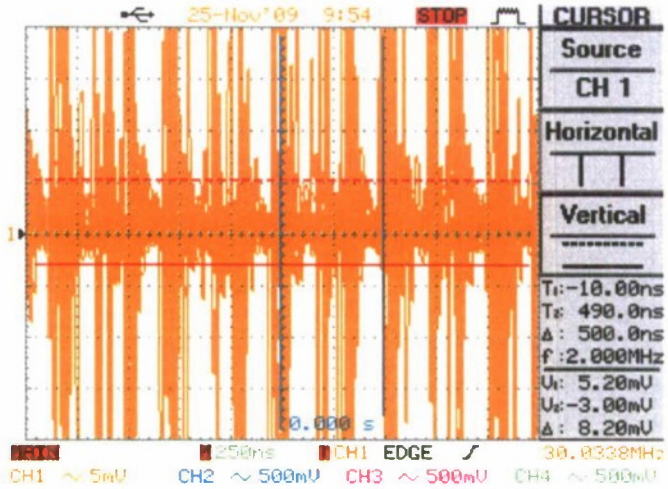


Figure 7.14 Output Voltage Ripple of the First Topology (8.2mV)

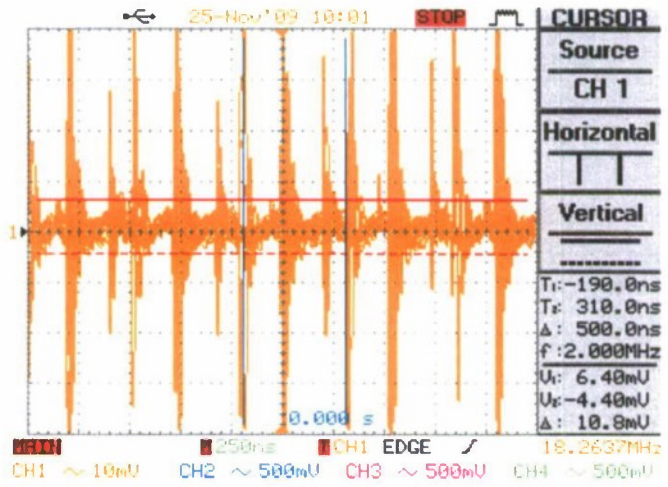


Figure 7.15 Output Voltage Ripple of the Second Topology (10.8mV)

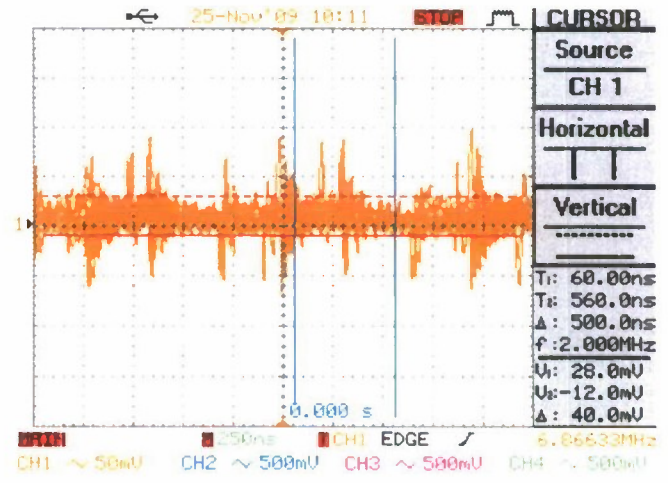


Figure 7.16 Output Voltage Ripple of TI VRM (40mV)

7.3. Transient Response

Next, load transient response tests were performed to see how fast the new converters recover upon a step change in the load. Figures 7-17 to 7-22 show both step up and step down responses of the converters as measured on their output voltage. For the step up, the commercial VRM takes about 156 us for the response as shown in Figure 7-21. This is slightly higher than that measured in Topology 1 at 136 us (Figure 7-17) and slightly lower than that of Topology 2 at 188 us (Figure 7-19). For the step down, the commercial VRM takes about 156 us for the response as shown in Figure 7-22. The new topologies as depicted in Figures 7-17 and 7-19 yield longer response time: 160 us for Topology 1 and 176 us for Topology 2.

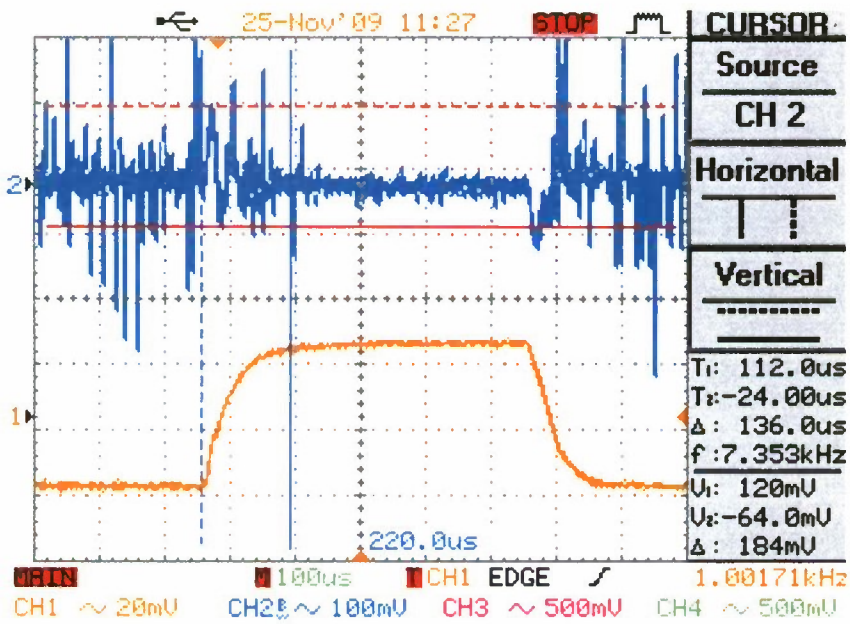


Figure 7.17 Output Voltage waveform as Load changed from 0 to 20A (Topology 1)

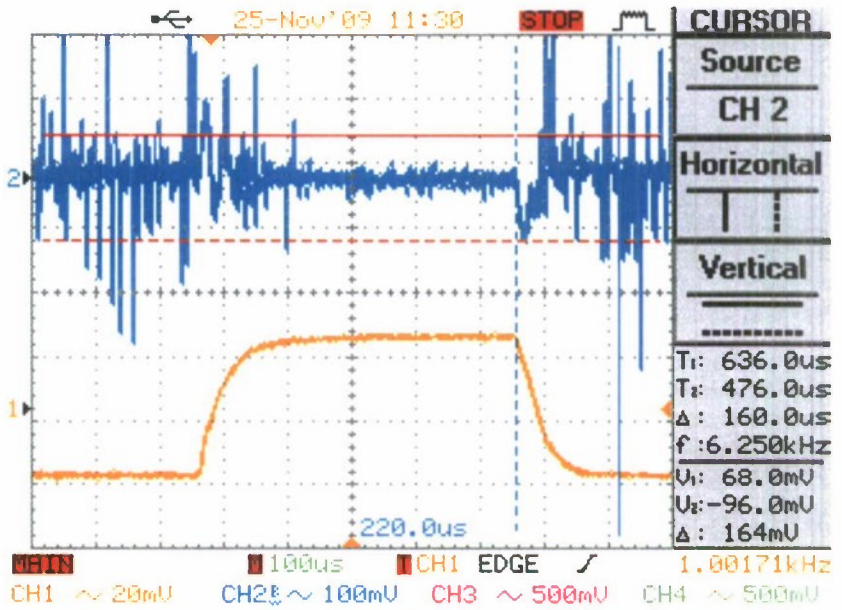


Figure 7.18 Output Voltage waveform as Load changed from 20A to 0A (Topology 1)

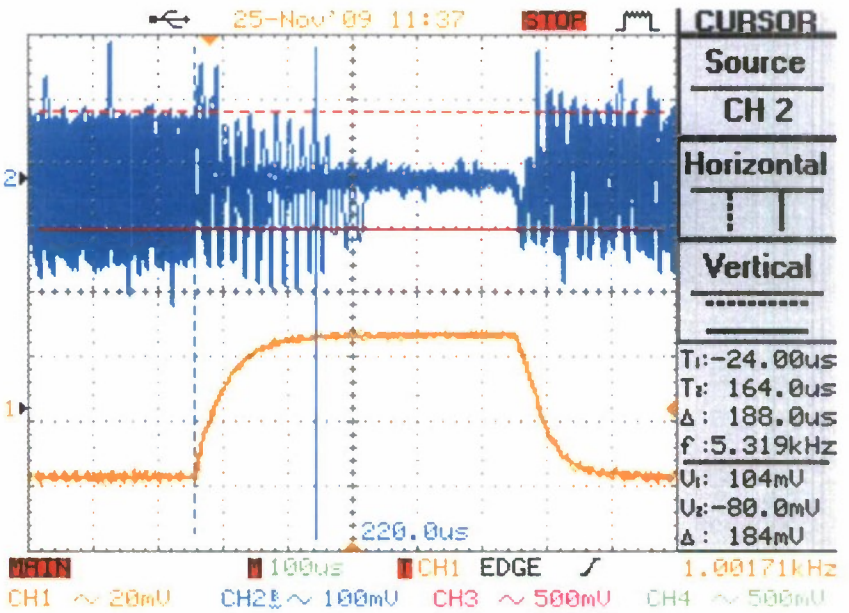


Figure 7.19 Output Voltage waveform as Load changed from 0 to 20A (Topology 2)

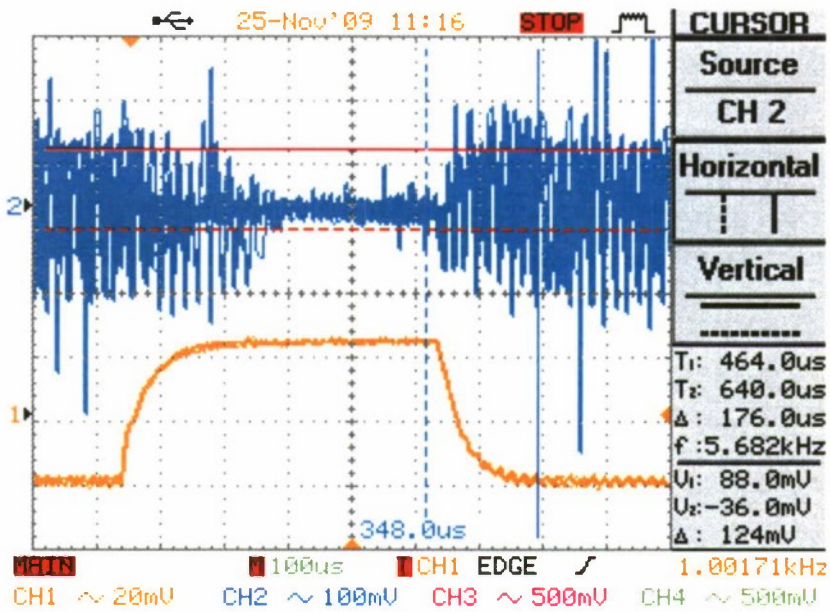


Figure 7.20 Output Voltage waveform as Load changed from 20 to 0A (Topology 2)

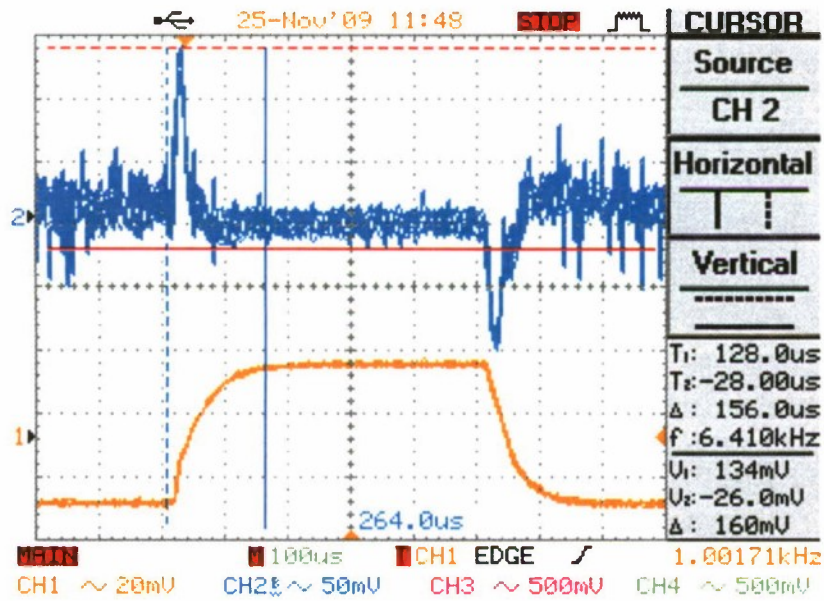


Figure 7.21 Output Voltage waveform as Load changed from 0 to 20A (TI VRM)

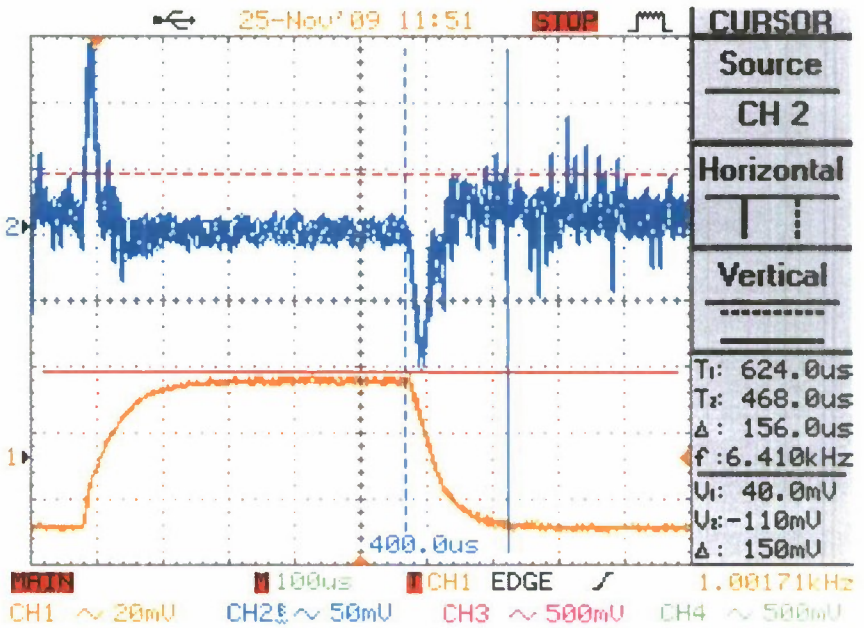


Figure 7.22 Output Voltage waveform as Load changed from 20 to 0A (TI VRM)

7.4. Line and Load Regulations

Line and load regulations were being measured next. Line regulation refers to how the output voltage can be maintained upon variation at the input voltage while the converter is drawing the full load. Load regulation on the other hand is a measure on how the output voltage is kept the same when the load or output current is varied.

For line regulation, the output voltage was measured when the output current is 40 A while the input voltage was changed from 10 V to 14 V. The load regulation was measured when the input voltage is at its nominal voltage of 12 V while the output current is changed from 10% to 90%. To follow is the results of line and load regulations for topology 1, topology 2, and commercial VRM:

Topology 1:

$$\text{Line Regulation} = \frac{V_{OUT(High\ Input)} - V_{OUT(Low\ Input)}}{V_{OUT(Nominal)}} x\ 100\%$$

$$= \frac{1.006\,V - 1.006\,V}{1.006} x\ 100\% = \underline{\mathbf{0\ \%}}$$

$$\text{Load Regulation} = \frac{V_{OUT(No\ Load)} - V_{OUT(Full\ Load)}}{V_{OUT(Full\ Load)}} x\ 100\%$$

$$= \frac{1.006\,V - 1.006\,V}{1.006} x\ 100\% = \underline{\mathbf{0\ \%}}$$

Topology 2:

$$\text{Line Regulation} = \frac{V_{OUT(High\ Input)} - V_{OUT(Low\ Input)}}{V_{OUT(Nominal)}} x\ 100\%$$

$$= \frac{1.004\,V - 1.004\,V}{1.004} x\ 100\% = \underline{\mathbf{0\ \%}}$$

$$\text{Load Regulation} = \frac{V_{OUT(No\ Load)} - V_{OUT(Full\ Load)}}{V_{OUT(Full\ Load)}} x\ 100\%$$

$$= \frac{1.006\,V - 1.004\,V}{1.004} x\ 100\% = \underline{\mathbf{0.199\ \%}}$$

Tl VRM:

$$\text{Line Regulation} = \frac{V_{OUT(High\ Input)} - V_{OUT(Low\ Input)}}{V_{OUT(Nominal)}} x\ 100\%$$

$$= \frac{1.008\,V - 1.008\,V}{1.008} x\ 100\% = \underline{\mathbf{0\ \%}}$$

$$\text{Load Regulation} = \frac{V_{OUT(No\ Load)} - V_{OUT(Full\ Load)}}{V_{OUT(Full\ Load)}} x\ 100\%$$

$$= \frac{1.000\,V - 1.008\,V}{1.008} x\ 100\% = \underline{\mathbf{0.794\ \%}}$$

When compared against the commercially available VRM, the proposed topologies have comparable line regulation (close to 0%) but they are superior in their load regulation (close to 0% for Topology 1 and 0.199% for Topology 2 as compared to 0.794% for the commercial VRM).

7.5. Efficiency

Figures 7-23 to 7-25 show data at various load measured from Topology 1, Topology 2 and the commercial VRM. Efficiency plots of the proposed converters were then generated as shown in Figure 7-26 to 7-28. The efficiency tracks the 80% line beginning approximately

at 40% load. At full load, the efficiency of the proposed converter is 80.75% which meets the design objective and is slightly larger than that measured from the commercially available VRM (80%). Another observation that was done involves temperature measurement of the main phase inductor which gives a good indication of how efficient energy is flown in the converter. For Topology 1, the inductor temperature after 4 minutes of operation (no cooling) was measured at 73 deg. C, while Topology 2 yields 72 deg. C, and the commercial VRM gives out 105 deg. C. The difference of about 30 deg C between the new topologies and the commercial VRM proves that much less loss is dissipated on the main inductors of the new topologies.

Load (%)	Vin(V)	Iin(A)	Pin(W)	Vout(V)	Iout(A)	Pout(W)	Efficiency (%)
0	12.006	0.267	3.210	1.006	0.01	0.010	0.31
10	12.006	0.731	8.773	1.006	5	5.030	57.34
20	12.007	1.186	14.235	1.006	10	10.060	70.67
30	12.006	1.648	19.786	1.006	15	15.090	76.27
40	12.004	2.119	25.430	1.006	20	20.120	79.12
50	12.003	2.599	31.191	1.006	25	25.150	80.63
60	11.992	3.092	37.077	1.006	30	30.180	81.40
70	12.004	3.593	43.132	1.006	35	35.210	81.63
80	12.004	4.110	49.338	1.006	40	40.240	81.56
90	11.996	4.644	55.706	1.006	45	45.270	81.27
100	12.004	5.189	62.294	1.006	50	50.300	80.75
100	10.001	6.129	61.291	1.006	50	50.300	82.07
100	14.001	4.553	63.747	1.006	50	50.300	78.91

Figure 7.23 First Topology Efficiency and Regulation Data

Load (%)	Vin(V)	Iin(A)	Pin(W)	Vout(V)	Iout(A)	Pout(W)	Efficiency (%)
0	12.005	0.265	3.179	1.006	0.01	0.010	0.32
10	12.002	0.729	8.753	1.006	5	5.030	57.47
20	12	1.184	14.210	1.005	10	10.050	70.72
30	11.995	1.646	19.745	1.005	15	15.075	76.35
40	12.008	2.115	25.391	1.005	20	20.100	79.16
50	12.001	2.594	31.132	1.005	25	25.125	80.71
60	11.995	3.085	37.003	1.005	30	30.150	81.48
70	12.001	3.585	43.027	1.004	35	35.140	81.67

80	12.008	4.098	49.214	1.004	40	40.160	81.60
90	11.996	4.630	55.540	1.004	45	45.180	81.35
100	12.003	5.172	62.076	1.004	50	50.200	80.87
100	9.995	6.100	60.966	1.004	50	50.200	82.34
100	14.001	4.537	63.516	1.004	50	50.200	79.04

Figure 7.24 Second Topology Efficiency and Regulation Data

Load (%)	Vin(V)	Iin(A)	Pin(W)	Vout(V)	Iout(A)	Pout(W)	Efficiency (%)
0	11.996	0.114	1.366	1	0.01	0.010	0.73
10	11.998	0.572	6.857	1	5	5.000	72.92
20	11.999	1.024	12.285	1.001	10	10.010	81.48
30	12	1.485	17.822	1.002	15	15.030	84.33
40	12	1.961	23.536	1.003	20	20.060	85.23
50	11.997	2.453	29.427	1.004	25	25.100	85.29
60	11.995	2.962	35.523	1.004	30	30.120	84.79
70	12.004	3.478	41.750	1.005	35	35.175	84.25
80	11.995	4.034	48.393	1.006	40	40.240	83.15
90	11.998	4.596	55.145	1.007	45	45.315	82.17
100	11.995	5.227	62.697	1.008	50	50.400	80.39
100	10	6.218	62.181	1.008	50	50.400	81.05
100	14.005	4.528	63.409	1.008	50	50.400	79.48

Figure 7.25 TI-VRM Efficiency and Regulation Data

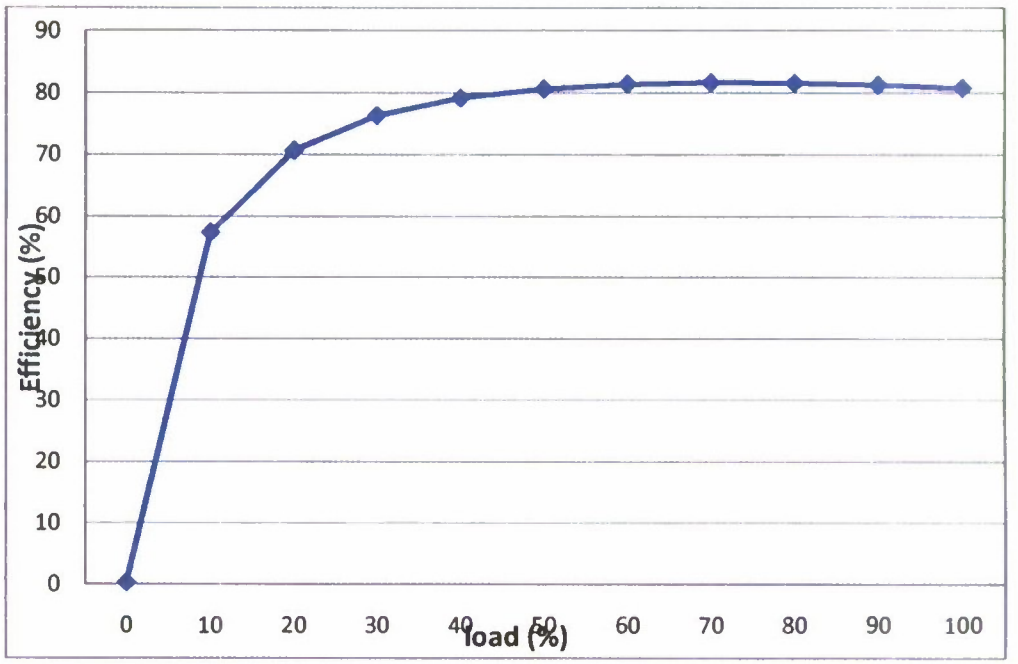


Figure 7.26 Efficiency Plot for the First Topology

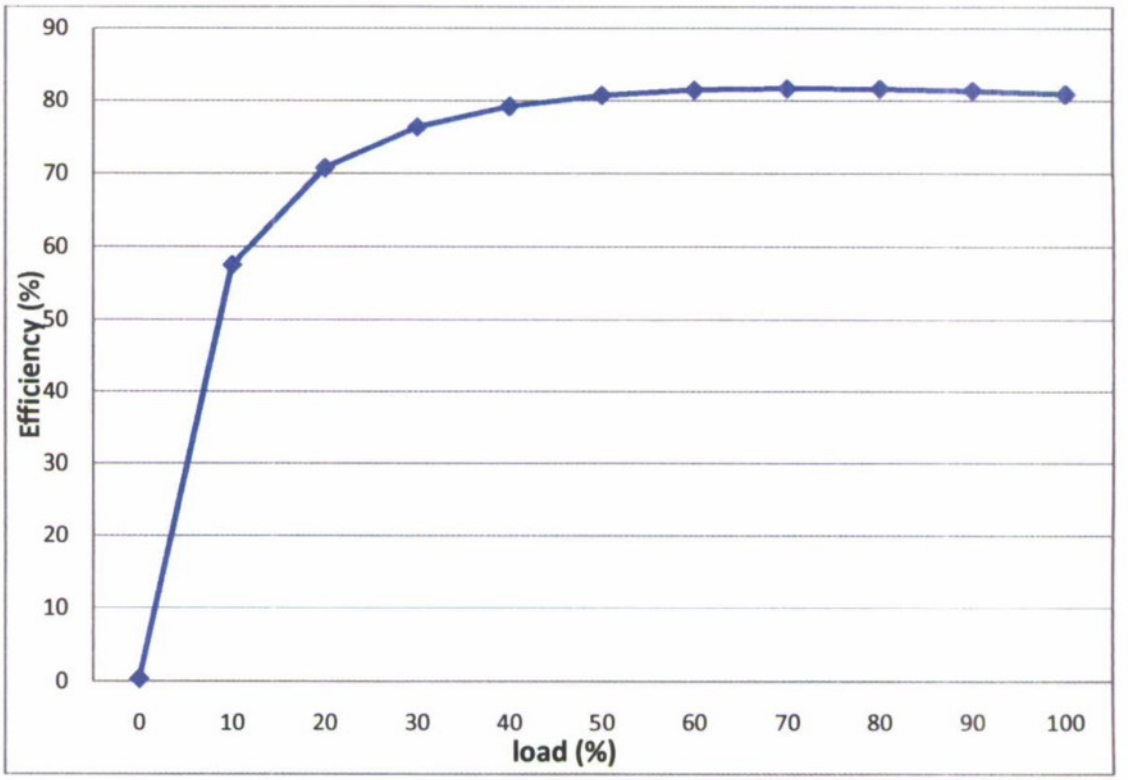


Figure 7.27 Efficiency Plot for the Second Topology

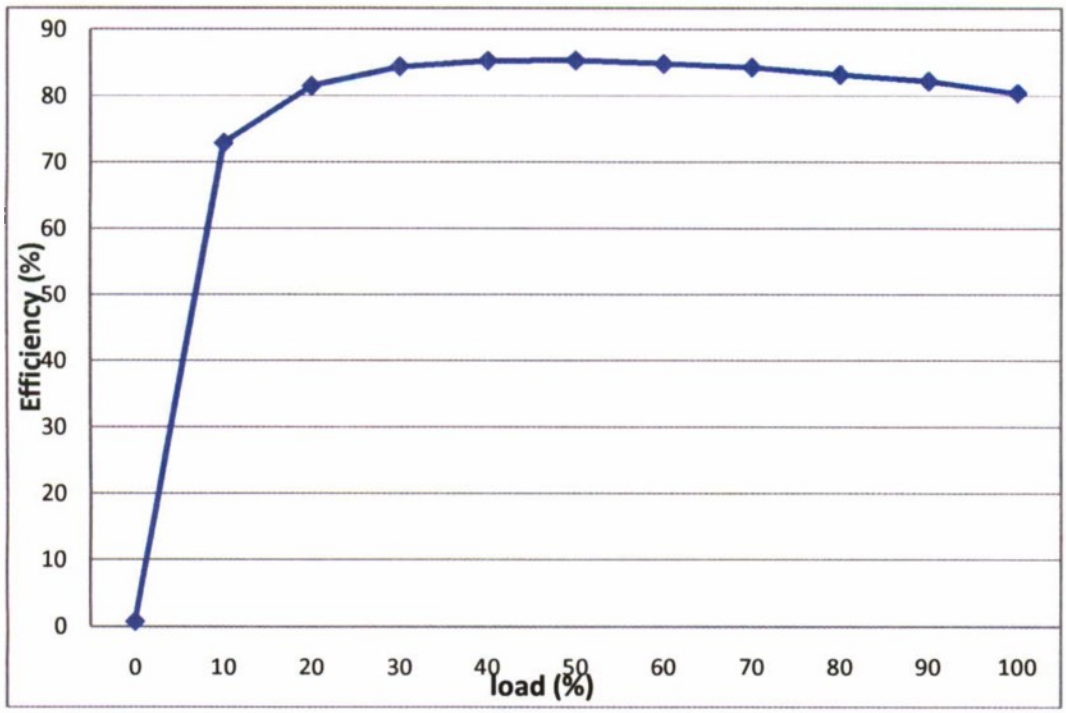


Figure 7.28 Efficiency Plot for the TI-VRM

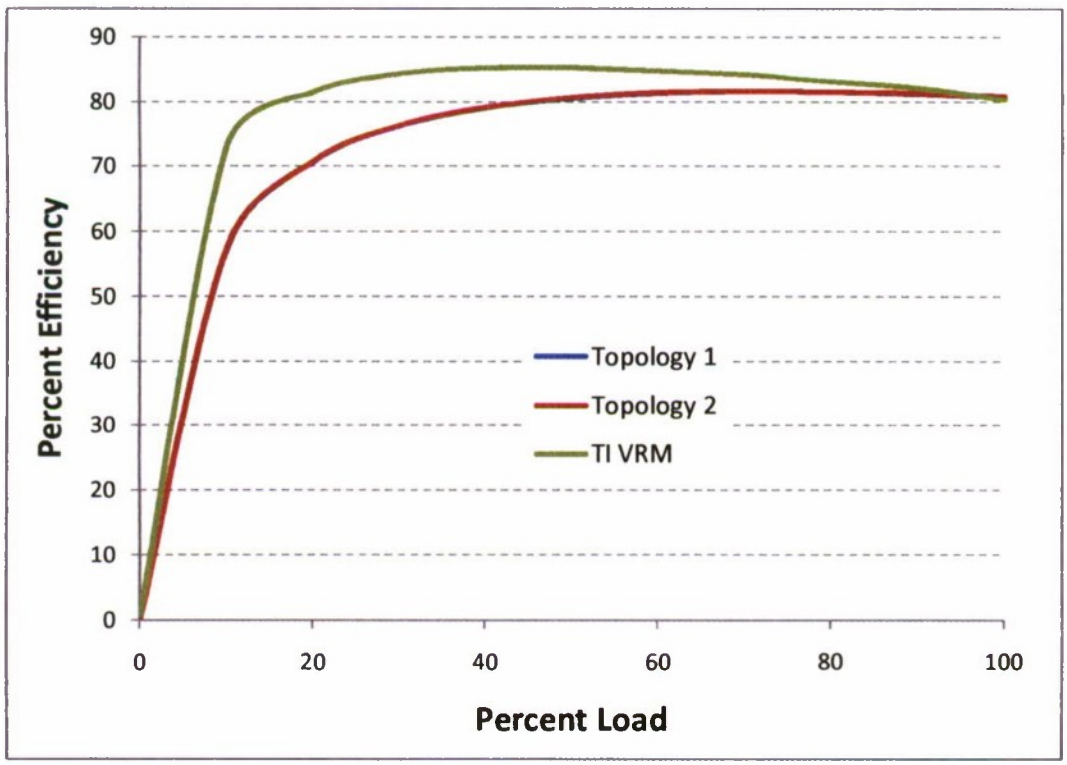


Figure 7.29 Efficiency plots for the three converters

Looking at Figure 7-29 we can conclude that the two new topologies have practically equal efficiency characteristics at varying loads. As shown, the two plots from Topology 1 and Topology 2 track each other very well. When compared with the efficiency plot from the commercial VRM, we can see that the commercial VRM has higher overall efficiency throughout the range that we tested (up to 40 A). What is more interesting is that at about 40 A, the efficiency from the commercial VRM continues to decline at a faster rate than those obtained from the new topologies. This implies that the new topologies will have better efficiency should higher current be drawn from the converters. This is exactly one property desirable on the new topologies, namely maintaining high efficiency for larger currents.

8. Dissemination

Results from this study have been and will be disseminated through several venues which include publications through Master's Thesis reports, peer-reviewed conference papers and presentations, and international engineering journal.

For dissemination through Master's thesis reports, there have actually been previous master's theses performed at Cal Poly which helped support the initial development and design of the project conducted in this report. The list is found below:

Student's Name	**Master's Thesis Title**
Furqan Saleemi	Analysis and Design of Multiphase Multi-Interleave DC-DC Converter with Input-Output Bypass Capacitor
Arief Hernadi	Analysis and Design of Improved Multiphase Interleaving DC-DC Converter with Input-Output Bypass Capacitor
Kevin Delrosso	Analysis and Design of Interleaving Multiphase DC-to-DC Converter with Input LC Filter
Sean Zich	Analysis and Design of Continuous Input Current Multiphase Interleaved Buck Converter
Rudianto	Analysis & Design Of Improved Multiphase Interleaving Dc-Dc Converter With Input-Output Bypass Capacitor

In terms of dissemination through conference papers or presentations, the results of the study as explained in this report have been presented and published in the following papers and conference:

[1]. **Taufik**, R. Prasetyo, D. Dolan, D. Garinto "A New Multiphase Multi-Interleaving Buck Converter With Bypass LC", Proc. of the 36th Annual Conference of the IEEE Industrial Electronics Society, Phoenix, November 2010, *to appear.*

[2]. **Taufik**, R. Prasetyo, D. Dolan and D. Garinto, "A New Multiphase Interleaving Buck Converter With Bypass Cell Capacitor and Inductor", Proc. of the 53rd IEEE Int'l Midwest Symposium on Circuits & Systems, August 2010.

[3]. **Taufik**, R. Prasetyo, D. Dolan and D. Garinto, "A Cell-Based Multiphase Interleaving Buck Converter with Bypass Capacitors", Proc. of International Conference on Computer, Electrical, and Systems Science, and Engineering, Bali, July 2010.

[4]. **Taufik**, R. Prasetyo, A. Hernadi and D. Garinto, "Multiphase Interleaving Buck Converter With Input-Output Bypass Capacitor", Proc. of 7th International Conference on Information Technology, April 2010.

Besides the aforementioned publications, currently a journal paper is being put together to present the results of the study to later be submitted for publication in either IEEE Journal Transactions or any other International Journal on Electrical Engineering. Last but not least, the concept of the new topologies as presented in this report was filed with the US Patent Office.

9. Conclusion

With the increasing demand for power in today's microprocessors, the design of VRM will become more challenging than ever before. Conventional or basic topology used in most commercially available VRMs will not be sufficient to satisfy the thirst of power and speed of today's and future microprocessors. The new topologies studied and developed in this project are aimed to address this issue. The use of cell configuration has demonstrated the effectiveness in interleaving a multiphase topology. Furthermore, the strategic placements of bypass capacitors have been shown to suppress the output voltage level to a minimum value which is critical in sub-volt applications. Further lab measurements on the hardware prototypes exhibit promising results of its potential. Although the results are overall comparable to those obtained from a commercially available VRM, two particular results are worth noting. First, load regulation of the new converters was measured to be practically 0%

which is a significant improvement from the one measured on the commercially available VRM. Load regulation becomes even crucial when output current is much higher than the 40A that was tested on this prototype. Thus, from this aspect, the proposed converter has shown its great potential for use in very high output current applications with very tight load regulation such as those expected in future microprocessors.

Secondly, the efficiency plots of the new converters were actually sloping down gradually after the full load. This is much different from that measured on the commercially available VRM in which the efficiency dives down relatively faster. This means, again for much higher output current applications such as those expected in future microprocessors, the proposed converter exhibits a great potential for use in future VRMs.

During the course of the project, we have learned so much about the design aspect of dc-dc converter. In particular, during the pcb layout and design phase several things we learned are:

- For interleaving design, we need to be careful and to ensure that the phase current sensing component on each phase inductor is routed to the proper phase current sense pin on the PWM Controller.
- One indication of having phase current sensing incorrectly routed is that the Duty Cycle keeps on changing over a wide range with constant load and input voltage.
- Small signal components and traces need to be routed away from the switching nodes.
- One indication of having voltage feedback signal line too close to the switching nodes is the output voltage is correct under no load condition, but not so when the load is varied.

The results from this project are far from perfect. Many improvements could be performed to achieve better hardware implementation of the new topologies. Several ideas for further improvement are listed here:

- Ground plane design and separate signal ground from power ground needs to be enhanced to reduce noise.
- Smaller components could be used to further reduce the board size.
- More than 4 layer boards could be used that will certainly allow for smaller board design.

- Needs to consider other PWM Controller with less external components.

This project does not stop here. As previously mentioned, A US Patent application on the new topologies has been filed. We are now waiting for the results from the USPTO. Once approved, further development and testing of the new topologies should be performed. In particular, the second prototype should be designed and built in similar size to those VRMs available commercially, and complete with the necessary pinouts and connectors to interface with a real processor system.

Finally, external funding both from the government and industrial sectors will be sought to follow up on the development of the new topologies which hopefully will become a real product.

References

[1]. Ed Stanford, "Power Technology Roadmap for Microprocessor Voltage Regulators," APEC 2004.
[2]. X. Zhou, P. Xu, and F. C. Lee, "A high power density, high frequency and fast transient voltage regulator module with a novel current sensing and current sharing technique," in Proc. IEEE APEC Conf., 1999, pp. 289–294.
[3]. P. Wong, "Performance Improvements of Multi-Channel Interleaving Voltage Regulator Modules with Integrated Coupling Inductors," Dissertation, VPI&SU, Blacksburg, VA, March 2001.
[4]. X. Zhou, P. Wong, P. Xu, F. C. Lee, and A. Q. Huang, "Investigation of candidate VRM topologies for future microprocessors," IEEE Trans. Power Electron., pp. 1172–1182, Nov. 2000.
[5]. Barry B, et al, "Comparison of two 12 V Voltage Regulator Module Topologies," IEEE APEC 2004.
[6]. Wenkang Huang, George Schuellein, and Danny Clavette, "A Scalable Multiphase Buck Converter with Average Current Share Bus," IEEE APEC 2003.
[7]. Y. Panov and M. M. Jovanovic, "Design consideration for 12-V/1.5-V, 50-A voltage regulator modules," IEEE Trans. Power Electron., pp. 776–783, Nov. 2001.
[8]. Fred C. Lee and Xunwei Zhou, "Power Management Issue for Future Generation Microprocessors," IEEE, 1999.
[9]. Peng Xu, "Multiphase Voltage Regulator Modules with Magnetic Integration to Power Microprocessors," Ph.D. dissertation, Virginia Tech 2001.
[10]. Jia Wei, "High Frequency High Efficiency Voltage Regulators for Future Microprocessors," Ph.D. Dissertation, Virginia Tech 2004.

[11]. Kaiwei Yao, "High Frequency and High Performance VRM Design for the Next Generations of Processors," Ph. D. Dissertation, Virginia Tech 2004.
[12]. Yuancheng Ren, "High Frequency, High Efficiency Two-Stage Approach for Future Microprocessors

Investigation of Nano-Photonic Structure on GaN LEDs and Solar-Cell Designs

Principal Investigator:

Xiaomin Jin
Electrical Engineering
California Polytechnic State University
San Luis Obispo, CA

Summary

The Gallium Nitride (GaN) Light-Emitting-Diode (LED) top-bottom (or transmission-reflection) grating simulation results with error grating model are studied intensively. The microstructure GaN bottom hole and top pillar gratings are calculated and compared with the non-grating (flat) case. Grating shapes simulated are either conical or cylindrical. A direct comparison of 181 different combined transmission-reflection grating cases using the Finite Difference Time Domain (FDTD) method is presented. The simulation results show that simple or direct combinations of the optimized top grating with the optimized bottom grating only produces a 42% light extraction improvement compared to the non-grating case, which is much lower than that of an optimized single grating case. This is due to the mismatch of grating parameters with the direct addition of the second grating structure, which changes the optical field distribution in the LEDs. Therefore, it is very important to optimize both top and bottom gratings simultaneously for the double-grating design. We also show the optimization of a double grating structure can achieve better performance than a single grating. Finally, transmission-reflection error gratings are also presented. It is also the first time to present randomization in GaN LED grating design and its effects in fabrication. Our data show that the favorable light extraction improvement is at approximately 10-15% randomization. The randomization can achieve 230% improvement over the original grating at a randomization intensity factor of 12.8%.

During the duration of this research program, we continue and expand the existing collaboration between Prof. Xiaomin Jin at California Polytechnic State University (Cal Poly, USA) and Prof. Bei Zhang and GuoYi Zhang's group in the Research Center of Wide-Gap Semiconductors at Peking University (PKU, China). A Cal Poly graduate student, Simeon Trieu, worked in PKU under NSF 08-603 OISE – EAPSI 2009 fellowship award ("EAPSI: Light Extraction Improvement of GaN-based Light Emitting Diodes") in summer 2009. We collaborate on simulation, fabrication, and characterization of photonic lattice-based GaN epitaxial material and optoelectronic devices. The Cal Poly group performs simulation and theoretical study, proposes novel structures, and characterizes the GaN-based photonic-lattice structures. The PKU group performs device growth and fabrication. This project is also the winner of 1st place in "Engineering and computer science-graduate level", *the 24th annual California State University Student Research Competition*, San Jose, May, 2010. In the next three years, NSF will support 5 Cal Poly students and 1 professor to work in PKU in the summer to continue the project.

1. ***Journal Publications of the project:***
   - S. Trieu and X. Jin, "Study of Top and Bottom Photonic Gratings on GaN LED with Error Grating Models," IEEE J. Quantum Electton., Feb 2010 submitted, now it is accepted and in print.
   - X. Jin, B. Zhang, F. Wang, J. Flickinger, S. Jobe, T. Dai, G.Y. Zhang, "International Engineering Research and Educational Activity on GaN Lasers and LEDs", International Journal of Engineering Research and Innovation (IJERI), Vol. 1, No. 1, P5, Spring/Summer 2009.

2. ***Conference paper:***
   - S. Trieu, X. Jin, C. Xiong, X. X. Fu, X. Kang, G. Y. Zhang, B. Zhang, F. Wang, "Simulation and Experiment on 2PC Transmitted Diffraction Grating for GaN LEDs" Nanotech 2010, June 2010, Anaheim, CA.

- S. Trieu, X. Jin, C. Xiong, X. X. Fu, X. N. Kang, G. Y. Zhang and B. Zhang, "Simulation of Three-fold Symmetric Photonic Crystal Structures on Top of GaN LEDs", CLEO 2010, May 2010, San Jose, CA.
- Nicholas Hageman and Xiaomin Jin, "Computer Generated a Three-Dimensional Holography from Two-Dimensional Photos", *Computational Optical Sensing and Imaging (COSI)*, San Jose, California, USA, October 13-15, 2009.
- Xiaomin Jin, Sean Jobe, Simeon Trieu, Benafsh Husain, Jason Flickinger, Tao Dai, Bei Zhang, Xiang-Ning Kang, and GuoYi Zhang, "Mode Pattern Analysis of Gallium Nitride-based Laser diodes", *The 3rd International Symposium on Photoelectronic Detection and Imaging* (ISPDI 2009), Beijing, China, June 17 to 19, 2009.
- Xiaomin Jin, Xiao Hua Yu, Fei Wang, Bei Zhang, and Guoyi Zhang, "Educational/Research Collaboration on Gallium-Nitride (GaN) Based Light Emitter between Cal Poly, CSULB, and PKU (China)", *the 12th CSU Regional Symposium on University Teaching*, California Polytechnic State University, San Luis Obispo, May 2nd, 2009. (Presentation only)

3. ***Proposals will be funded (Based on the Results of the Project):***
   - ***NSF OISE-IRES 2010, Feb 2010,*** "International: Engineering Research and Educational Collaboration on Gallium-Nitride-based Light Emitting Devices", by Xiaomin Jin and Xiao-Hua Yu, Department of Electrical Engineering, California Polytechnic State University, San Luis Obispo, CA 93407 , $147,775 **will be founded** from Sept 1st 2010 to Aug 2013.
   - ***RESEARCH FELLOWSHIP FOR 2010-2011 OF THE STATE KEY LAB, PEKING UNIVERSITY, BEIJING, CHINA*** "The optimization design of GaN LED double grating", PI: Xiaomin Jin, CO-PI: Kang, Xiang Ning, Guoyi Zhang, Bei Zhang, **accepted,** detailed budget will be announced.

4. ***Student Award Related to the Project:***
   - Simeon Trieu (Graduate Student) and Xiaomin Jin (advisor), "Study of Top and Bottom Photonic Crystal Gratings on GaN LEDs using FDTD" 1st place winner in "Engineering and computer science-graduate level", *the 24th annual California State University Student Research Competition*, San Jose, May, 2010.

5. ***Winner of NSF Fellowship of Cal Poly Student Related to the Project:***
   - Simeon Trieu was winner of the NSF 08-603 OISE – EAPSI 2009 fellowship award ("EAPSI: Light Extraction Improvement of GaN-based Light Emitting Diodes") in summer 2009.

A novel technique for preventing delamination of the composite skin from the core: Sandwich panels with shear keys

Principal Investigator:

Eltahry Elghandour, Ph.D.
Department of Civil Engineering
California Polytechnic State University, San Luis Obispo
San Luis Obispo, California

Abstract

This report presents the investigation of experimental and finite element analysis of composite sandwich structured laminated materials in three different applications using an arrestment shear key to test its effectiveness on the strength of the composite sandwich and its ability to prevent delamination growth. The applications involve composite sandwich skins made of fiber glass/carbon fiber with a core containing closed–cell PVC foam. This introduces the designer to a preventive alternative in repairing the composite sandwich structure.

In the first two applications, an initial delamination is introduced into the composite sandwich structures during the manufacture process to investigate the effect of delamination size and location on the mechanical characteristic under static and fatigue loading. The first application investigates the effect of the static and fatigue loading on open delamination with and without shear key on mechanical behavior. Also, the second application studies the effect of arrestment keys with initial delamination in a composite sandwich structure under buckling loading. . The delamination arrestment keys are semi-circular cross-sections made of fiberglass/epoxy strands. . The shear key locations were also varied and their effects noted. The fixed rate under static tests behavior was determined. The fatigue life and behavior were determined for composite sandwich structures with and without initial delamination and shear key.

The third application, involves a parametric study on the affects of how damage arrestment device application interacts with a fastener in a composite sandwich panel. The composite damage arrestment devices are inserted under the face sheet to increase the overall structural strength of the panel and to prevent the propagation of failure along the hole. The damage arrestment devices' thickness was varied and tested under both monotonic and fatigue loading.

A numerical analysis was performed using Finite Element Method (FEM) Software to observe the monotonic behavior of the test specimens with and without delamination. The correlation between FEM and experimental data was accurate enough in predicting the trends of the composite sandwich panels.

The first open delamination showed the shear key to have no significant effect. The second application yielded results that showed that the introduction of a continuous arrestment key parallel to the in-plane loading provided a significant increase in loading and buckling capabilities.

Nomenclature

ASTM	American Standard for Testing Materials
CNC	Computer numeric control
E	Young's Modulus ($\frac{lb_f}{in^2}$)
F	Force (lb_f)
FEA	Finite element analysis
g	Grams

h	height (in.)
in.	Inches
kg	Kilograms
kN	Kilonewtons
l	length (in.)
b_f	Pounds force
m	Meters
N	Number of cycles
Pre-preg	Pre-impregnated
psi	$\frac{lb_f}{in^2}$
PVC	Polyvinyl chloride
U	Displacement
S	Stress ($\frac{lb_f}{in^2}$)
sec	Seconds
t	Time
w	width (in.)
Ω	Resistance (ohms)
°	Degrees

Subscripts

amp	Amplitude
max	Maximum
min	Minimum
mean	Mean
ult	Ultimate
y	Yield
1	X
2	Y
3	Z

Chapter 1- Introduction

A brief discussion on composite materials is essential to familiarize the reader on the subject matter ahead. Composites will be defined in this section in terms of their materials, on how they are manufactured, and their advantages and disadvantages. The purpose of the research pertaining to composite sandwich panels will be discussed.

1.1 Overview of Composites

Composite materials have been used since the beginning of human history. The earliest recorded use of composites was in ancient Egypt where they created bricks from mud and straw to build their buildings. They also achieved higher strength and better resistance to thermal expansions to their bricks by adding plywood to it. Swords and armor were created with different layers of materials to improve strength during Medieval Europe.

A composite is defined as a material that contains two or more constituents that are combined on a macroscopic scale to form a new material. Composites exhibit the best qualities of their constituents and often some qualities that neither possesses. Some of the properties improved by composite materials include strength, stiffness, corrosion resistance, wear resistance, attractiveness, weight, fatigue life, temperature-dependent behavior, thermal insulation, thermal conductivity, and acoustical insulation. Creating a composite material does not improve all of the properties. The designer can choose which properties to improve by choosing the right pieces to combine. Laminated composites are used to combine the best aspects of the constituent layer in order to achieve a more useful material.

There are three types of composites: fibrous, laminated and particulate composites. All three types are made out of two constituents, the actual composite material and the binding agent or the matrix. The composite material can be a type of metal or a fiber. The binding agent or the matrix can be another type of metal or some type of adhesive material. The two constituent materials are combined with another to create a hybrid material that is much stronger and stiffer.

Fibrous composites consist of fibers in a matrix; this can be seen in Figure 1. Fibers in fibrous composites are much stiffer and stronger than the same material in its bulk form. The geometry of the fiber is crucial to its strength and must be considered in structural applications. The difference between the strength of a fiber and the strength of its bulk material is due to the crystal alignment along the fiber axis. There are also fewer defects in fibers than its respective bulk material.

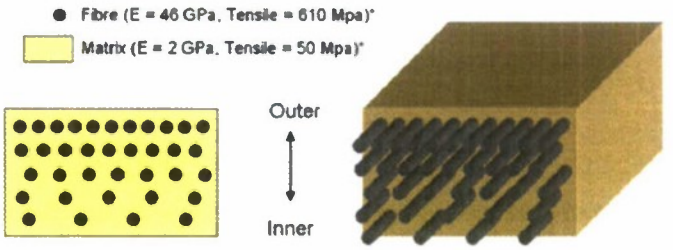


Figure 1 - Fibrous Composite

The purpose of the matrix is to support, protect, and transfer the stresses along the fibers. The matrix has a lower density, stiffness, and strength than the fibers. The combination of the fibers and the matrix can create a very strong and stiff material while still maintaining a low density.

A laminated composite can consist of layers of at least two different materials bound to one another; this can be seen in Figure 2. Some classes of laminated composites are bimetals, clad metals, laminated glass; plastic based laminates, and laminated fibrous composites. Bimetals are laminates of two different metals with significantly different coefficients of thermal expansion. Clad metals are the cladding or sheathing of one metal with another, which is done to obtain with the best properties of both metals. An example is high-strength aluminum alloy covered with a corrosion-resistant aluminum alloy. This is a composite material with unique and attractive advantages over regular aluminum which is very corrosive. Plastic based laminates are materials that are saturated with various plastics to improve its material property. Laminated fibrous composites are a hybrid of composites involving both fibrous composites and lamination techniques; this is also known as laminated fiber-reinforced composites. Layers of fiber-reinforced material are built with the fiber directions of each layer typically oriented in different directions to give different strengths and stiffness in various directions.

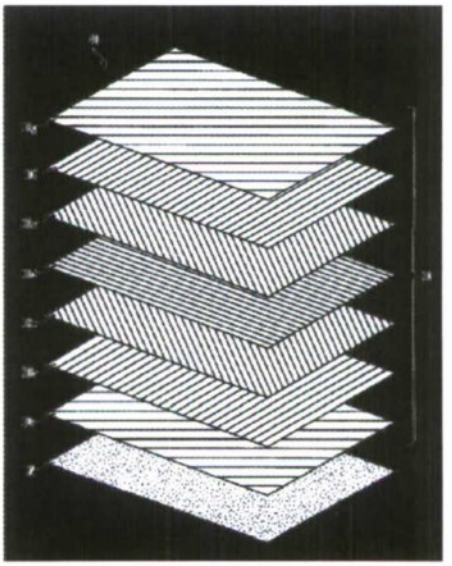


Figure 2 - Laminated Composite (Courtesy of Picses EU)

Particulate composites are composed of particles in a matrix; this can be seen in Figure 3. Particles are non-fibrous and generally have no long dimension. The particles and the matrix can be either metallic or nonmetallic. The choice of a particular combination depends upon the desired properties of the final material. The most common example for a particulate composite is concrete. Concrete consists of small sand and rock particles bonded together by a mixture of cement and water that chemically reacts and hardens. The strength of the concrete is comparable to rock.

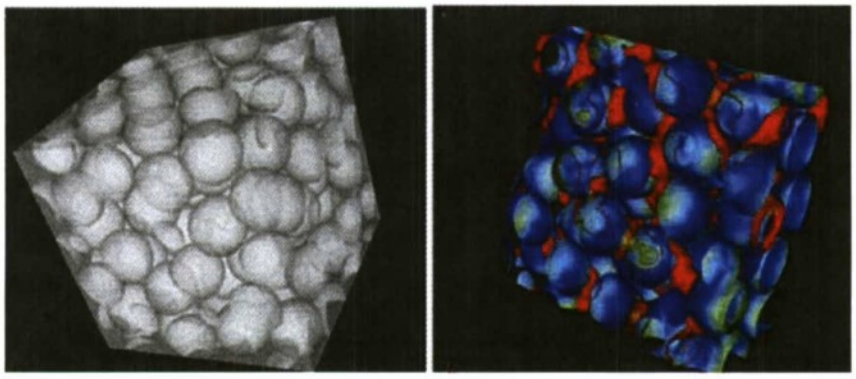


Figure 3 - Particle Composite (University of Illinois Engineering)

Composites have two major advantages: improved strength and stiffness, especially when compared with other materials on a unit weight basis. The best way of showing the effectiveness of strength or stiffness of a material is the weight per unit volume, as seen in Figure 4. It shows a comparison of the strength and stiffness of several types of advanced composite materials and contemporary metals. The most desired material in the figure is located in the upper right hand corner of the figure, which represents a material with high strength and stiffness. It shows that composite materials are stronger and stiffer than common metals such as aluminum, steel, and titanium, while being lighter than those metals.

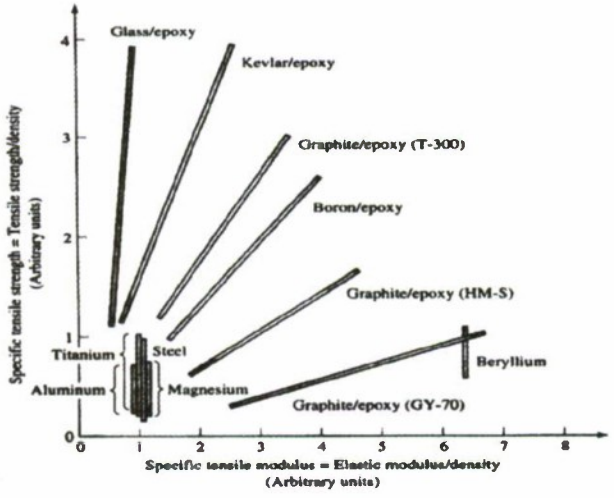


Figure 4 - Strength and Stiffness of Advanced Composite Materials

The properties of composites are strongly influenced by the properties of the materials they're composed of, the material distribution, and the interaction between the materials. Properties of the composite are mainly dependent on the geometry of the re-enforcement of the composite material such as shape, size, and orientation. Size and distribution control of the material determines the interaction between the fiber and the matrix.

Materials are stronger and stiffer in the fibrous form as in its bulk material form. A high fiber aspect ratio permits very effective transfer of load via matrix materials to the fibers, making a very effective reinforcement material. There are several types of fibers used in composites including glass, carbon, graphite, aramid, boron, and ceramic. Glass fibers are the most common fiber used for composite materials because of their low cost and high strength. Carbon and graphite fibers are mostly high-strength, high-modulus fibers that are used for high-performance composites. Aramid fibers, or Kevlar, are a type of fiber that can change its material property depending on how the material is created. Kevlar fibers have higher tensile strength and modulus, but have a lower fiber elongation than glass or carbon fibers. Boron fibers have a higher tensile strength than carbon fibers, but the fibers are more expensive to produce than other fibers. Ceramic fibers were created because of the need for high temperature fibers. Ceramic fibers combine high strength and elastic modulus with higher temperature capability than other types of fiber materials.

The matrix of a composite material binds the fibers together, transfers loads between them, and protects them against environmental degradation and damage due to handling. The matrix has a great influence on the composite material because it defines several material properties such as shear, compression, transverse modulus, and strength properties. It also limits a material's temperature property by defining the maximum operational temperature the composite material can work in. Polymers or plastics are the most widely used matrix material for fiber composites because of their low cost, ease of production, chemical resistance, and low specific gravity. One of the main disadvantages of polymers is their low strength, low modulus, and low operating temperature. The most common polymers used for a composite matrix are polyester resin and epoxy resins. Metals can also be used as a matrix material for composites. Metals have a high strength, high modulus, high toughness, impact resistance, and its insensitivity to temperature changes over polymers. The main problems associated with use of metals as a matrix is their high density, high processing temperatures, reactivity with fibers, and corrosiveness.

There are three different common types of layup processes used for creating composite pieces. The simplest of these layup processes is the "wet layup" technique; this can be seen in Figure 5. This technique involves the least amount of materials and preparation time out of the three methods, but comes at the cost of the quality of the part. A wet layup is one in which the resin is pushed through the fibers by hand, and then placed inside a vacuum sealed bag and allowed to cure. Major advantages for this layup include low cost relative to the other layup processes, no special materials or additional equipment is required, the part is normally capable of curing at room temperatures, and the method is basic so that almost anyone can use it. Some disadvantages

associated with this method include restricted work time (based on the resin), inconsistencies in resin to fiber ratio or the part, as well as destroying fiber integrity while working resin through the fibers.

Figure 5 - "Wet Layup" Technique

Another method typically used is called a resin infusion process or a vacuum resin infusion (VRI) process; this can be seen in Figure 6. This method has the most variation from person to person since each individual typically uses a slightly different setup and process. However, the main idea behind this method is that a vacuum is used to pull resin through the part. The advantage to this is a more consistent resin to fiber ratio throughout the whole part similar to that of the pre-impregnated carbon fiber but at a greatly reduced cost. However it does not require a complex cure cycle nor do the materials need to be stored to preserve a "shelf life" like the pre-preg. It is also a preferred method for repeated experiments over a wet layup since the parts are reproducible and provide more consistent results. However this process does require more knowledge or the materials being used. Work time is dependent on the viscosity of the resin as well as the strength of the pump. The size of the part must also be taken into account since the resin may start to cure before reaching the other side of a larger part. VRI requires more materials and preparation time than either of the other two methods. A flow media is needed in order to allow resin to flow across the part, a peel-ply is placed under this flow media in order to remove it from the part, and additional tubing is required for the resin to be sucked through. This makes this method more costly then the wet layup, but allows the fabricator to avoid tampering with the fibers and causing misalignments in the part.

Figure 6 - VRI Layup

The final method used in industry involves the use of pre-impregnated fibers also known as "pre-preg"; this can be seen in Figure 7. This method is ideal for almost any application in which money is not an issue. The extreme advantage with the pre-preg is that it already has the ideal resin to fiber ratio giving the part the maximum strength to weight ratio. The method gives the best consistency and higher quality in fabrication of the parts due to the premade resin to fiber ratio. Another advantage of pre-preg is the work time which is significantly longer then the wet layup technique, which starts to cure when the hardener and the epoxy are mixed. However, the disadvantage of this is that the manufacturer specifies the work time for pre-preg, and also, after the specified "shelf life" has expired, the material is no longer usable. Due to the "shelf life", a freezer is needed to store the pre-preg in order to prevent it from losing "shelf life" when it is not being used. Pre-preg requires a heating cycle in order for the part to cure; meaning some type of machine has to be used to create the heating cycle. This includes autoclaves and heat presses to cure the pre-preg composite.

Figure 7 - Pre-preg Layup (Courtesy of Composite World)

Composite structures are any structures that contain a fibrous or laminated composite. The most common types of composite structures that will be researched are sandwich composite panels. They are widely used in several industries such as the marine, aerospace, and automotive industries; this can be seen in Figure 8. For application in the aerospace industry, fuselage hulls are made from panels of a foam core sandwiched

between layers of carbon fiber. There are major advantages of sandwich composite panels such as decrease in weight and an increase in structural strength. Some of the major drawbacks are low shear strength and delamination of the composite skin from the core. The sandwich composite panels are susceptible to in-plane shear which is resisted primarily by the core, core compression failure, indentation of face-sheet subjected to impact loading, global buckling and wrinkling instability, and delamination of the face sheet from the core. In a delaminated structures subjected to compression, the face sheet in the delaminated region may buckle allowing the delamination to propagate through the panel causing a global structural failure. The shortcomings of fiber-reinforced plastics are that they have low mechanical properties and are susceptible to failure by impact loading. Low velocity impact on a sandwich structure can induce damage to the facings, the core, and the face-core interface. According to Castaings, there are five different failure modes: core crushing, delamination in the impacted face sheet, core cracking, matrix cracking, and fiber breakage in the facings. Damage of a composite panel is dependent on the properties and interaction of the core material and the facing of the plates.

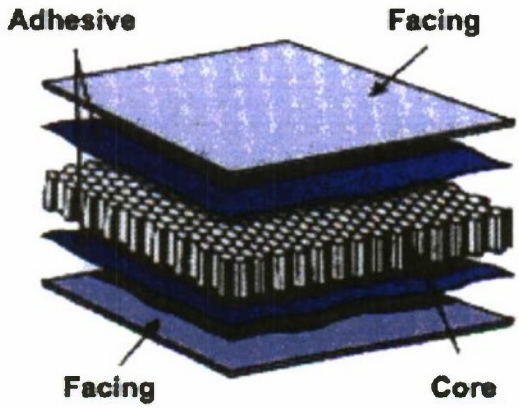


Figure 8 - Example of Sandwich Composites (Courtesy of Engineer Materials Inc)

There are accepted testing standards that allow results to be compared with other research by following guidelines on experimental testing. One of these standards is the American Standard for Testing and Materials (ASTM). It is an internationally recognized organization that develops and publishes consensus technical standards for materials, products, systems, and services. Most standards that pertain to composite testing are for the compression of composite sandwich structures and for testing of composite laminates with fasteners. The standards give a full detailed account on how to dimension the test specimens, set-up the testing apparatus, and how to analyze the results of the testing to be comparable to previous work. Since there are few definite or accepted standards to test composite sandwich panels with a fastener, a variation or evolution of some of the standards would be used for the experimental testing to require the desired or predicted results of the research; this will be further discussed later in this report.

Composite sandwich structures are becoming more heavily integrated into everyday applications ranging from marine to aerospace. A composite sandwich structure consists of two composite skins on each

side of a core material. Core materials are generally lightweight materials such as balsa wood, honeycomb structure, or foam. A sandwich structure with an aluminum honeycomb core and an open celled foam core are shown in Figure 9. Composite sandwich structures offer high strength-to-weight ratios when compared to traditional metals. Using composites allow designers more flexibility in being able to tailor the material properties of their part based on the orientation of the composite. However, composite sandwich structures also have disadvantages such as water intrusion and impact, which can cause the composite skin to delaminate from the core. A delaminated sandwich structure significantly loses its structural strength and can approach the point of catastrophic failure.

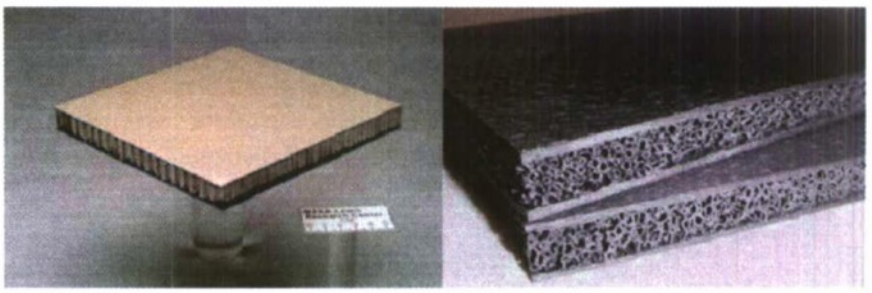


Figure 9 - Examples of sandwich structures[1,2]

Thus far, only three techniques have been discovered to help stop a composite skin from separating once it has begun to delaminate from the core. The three research methods studied involve buffer strips, peel stoppers, and shear keys. Buffer strips[3] are used to increase the strength of a composite plate that already has cracks or defects. Buffer strips are additional layers of composite strips that are added to the defective regions or to regions where it is anticipated that cracks will occur. Figure 10 shows a general schematic of a buffer strip added to the defective regions a composite panel containing cracks. Under fatigue loading, after additional Kevlar or structural fiberglass strips were added to carbon fiber composite panels, the buffer strips arrested the crack growth and increased the panel's residual strength[3].

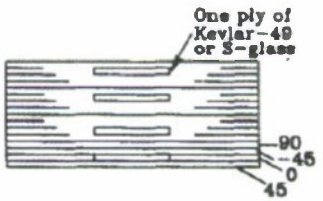


Figure 10 - Schematic of buffer strips[3]

Another method used for stopping composite face sheet delamination involves a marine peel stopper, developed by Christopher Wonderly and Joachim Grenestedt at Leigh University. A schematic of a peel stopper is shown Figure 11. A peel stopper essentially divides the composite sandwich structure into different regions,

preventing a delaminated skin from traveling into consecutive regions. The peel stopper causes the delaminated face sheet to separate away from the structure and leave the remaining composite intact. The upper part of Figure 12 shows a delaminated skin and the lower part of Figure 12 shows the outer skin breaking away from the structure. Peel stoppers have shown promising performance in stopping delamination, are relatively simple and easy to manufacture, and cost effective. Preliminary tests that were performed have shown that peel stoppers possess good quasi-static in-plane strength[3].

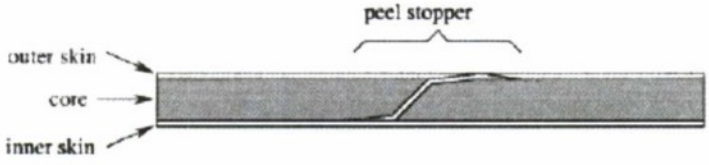


Figure 11- General schematic of a peel stopper[3]

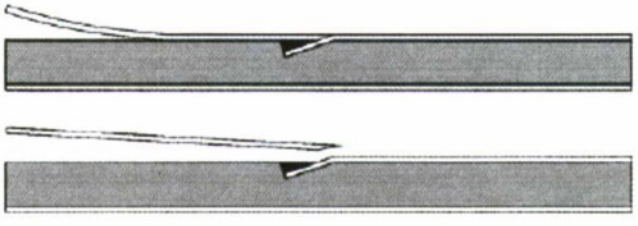


Figure 12 - Peel stopper ending delamination[3]

In 2008, Dr. Nilanjan Mitra, Dr. Eltahry Elghandour, and Dr. Eric Kasper began researching another cost effective solution to increasing the shear strength of a composite sandwich[4] structure at California Polytechnic State University in San Luis Obispo. The research was conducted under the sponsorship of the C3RP-ONR grant. The sandwich structures selected consisted of fiberglass face sheets surrounding a closed cell PVC foam core. Fiberglass rods, called shear keys, were incorporated into the foam to increase the shear strength. The shear keys were semicircular fiberglass rods manufactured out of intertwined strands of fiberglass. The left side of Figure 13 shows the shear key rods and the right side shows milled out grooves in the foam where the shear keys will be placed.

Figure 13 - Shear key rods and grooves milled into foam core

Dr. Mitra's, Dr. Elghandour's, and Dr. Kasper's research focused on the size, shape, and the location of the shear keys. Triangular shaped shear keys are shown in the right side of Figure 14. Under shear loading conditions, the non-staggered triangular shear keys showed little increase in shear strength, attributed to the stress concentrations introduced at the tips of the triangles. The large circular shear keys as shown on the left side of Figure 14, but in a non-staggered configuration were then tested. Although it was determined that the large circular shear keys were stronger than the triangular shear keys, failure often occurred in the foam between the shear keys because there was less core material present. To address the problem circular shear keys were placed in a staggered manner helping to provide increased shear strength. Shrinking the radius of the shear keys provided the best results because less of the foam was removed.

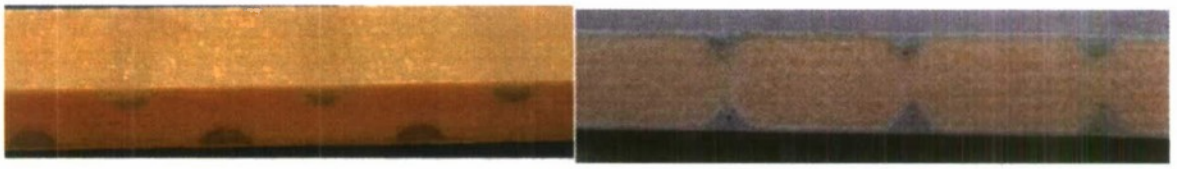


Figure 14 - Shear key size and location research[4]

Figure 15 - Composite sandwich structures under shear testing (ASTM C273) shows a sandwich structures being tested under shear loading conditions. After the sandwich structures were tested under shear loading, and maximum shear strengths were determined, some researcher looking into how the shear keys could aid in preventing composite delamination due to peeling. The goal of the research was to investigate how delamination length and shear key location affected the monotonic failure characteristics of the composite sandwich structure[6]. The initial results gathered from the delamination testing were inconclusive because many test specimens failed prematurely and therefore only a small sample of tested specimens could be used. This project research will continue on with the shear key work that Dr. Dr. Mitra's, Dr. Elghandour's, and Dr. Kasper's started but will focus primarily on the response of the shear keys in preventing delamination peeling.

This project will exemplify the experimental and numerical analysis of composite sandwich structures under monotonic, buckling, and fatigue loading. This will be done with the use of three different applications experiments involving sandwich composite panels.

Figure 15 - Composite sandwich structures under shear testing (ASTM C273)

Chapter 2- Methods of Design and Specimen Manufacture

2.1-Sandwich Structure Design (Eugene and Tony)

In the first two applications, the materials selected for this research involved using composite materials for the face sheets which were woven roving fiberglass and chopped strand fiberglass mats, or bi-weave carbon fiber. The chopped strand mat consists of 3 inch length fibers that are randomly oriented. The intent of the chopped strand mat is to give the composite sandwich structure more isotropic properties. Woven roving is a fiberglass bi-weave with fibers oriented at angles of 0° and 90°, similar to the weave of the carbon fiber except larger groups of fibers were placed together in the fiberglass. These composite sandwich structures' strength mainly comes from the bi-directional woven plies in the face sheets. Figure 16 shows chopped strand mat on the left and woven roving mat on the right.

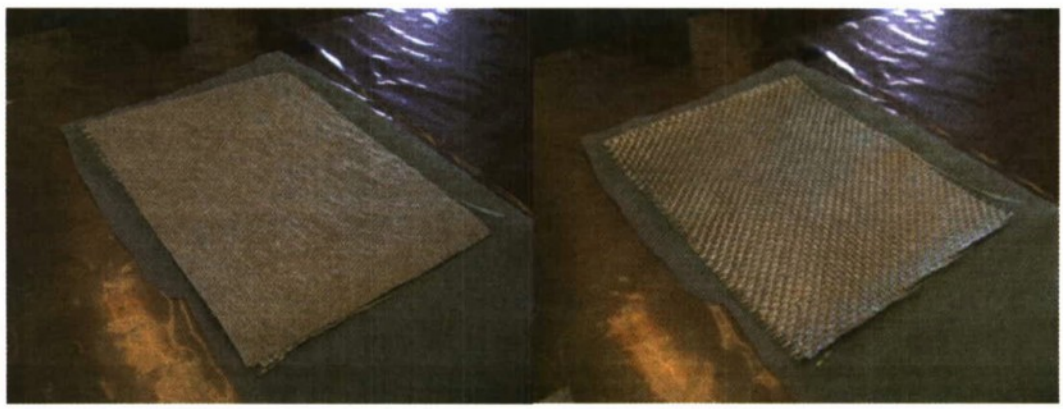


Figure 16 - Chopped strand mat and woven roving fiberglass

The core material used was Divinycell's H 100 PVC foam with a thickness of 20 mm and a density of 100 kg/m^3 for the monotonic tension and cyclic tensile cases containing fiberglass face sheets. This closed cell foam has a high strength-to-weight ratio as well as excellent ductile qualities. The purpose of the PVC foam core is to increase the sandwich structure's bending and torsional stiffness while only slightly increasing the weight of the test specimens. A stack of different Divinycell foams are shown in Figure 17.

Figure 17 - Divinycell foam core[13]

For the rest of the tests, carbon fiber composite sandwiches as well as the compression buckling tests used a different core material. A foam core material was chosen to match what is currently being used in the boating and aerospace industries of today. The foam core used for this part of the project was a LAST-A-FOAM FR-6710 foam material purchased from General Plastics Manufacturing Company. According to General Plastics, the foam is a CFC-free, closed-cell, and a flame-retardant polyurethane material with a density of 10 pounds per cubic foot. This particular foam differs from the one H100 Divinycell PVC foam, but provides the necessary properties to produce the proper results from the buckled structure along with the heating requirements used in the carbon fiber tests.

The composite lay-up used to manufacture the sandwich structures consisted of four layers of fiberglass on either side of the foam core. The stacking sequence of the skin, starting from the outer most layer was: woven roving, chopped strand mat, woven, chopped, core, chopped, woven, chopped, woven.

Six different fiberglass composite sandwich structure configurations were manufactured and tested under monotonic tension loading as well as a cyclic tensile loading, which are listed below:

1. No Initial Delamination
2. 0.5 inch Initial Delamination
3. 1.0 inch Initial Delamination
4. 1.0 inch Initial Delamination with 0 inch shear key
5. 1.0 inch Initial Delamination with 0.5 inch shear key
6. 1.0 inch Initial Delamination with 1.0 inch shear key

Figure 18, Figure 19 - Test specimen schematic for initial delamination and Figure 20 below show schematics of the six different configurations with important components and dimensions labeled. All of the test specimens have the following dimensions: 6 in. x 1.5 in. x 0.787 in.

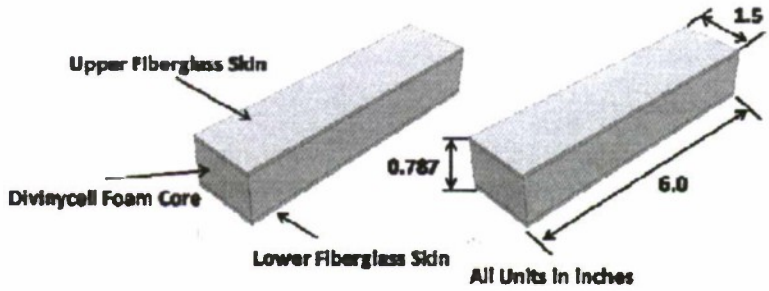


Figure 18 - Test specimen schematic for no initial delamination

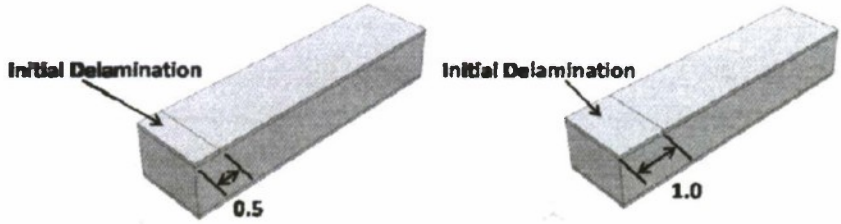


Figure 19 - Test specimen schematic for initial delamination

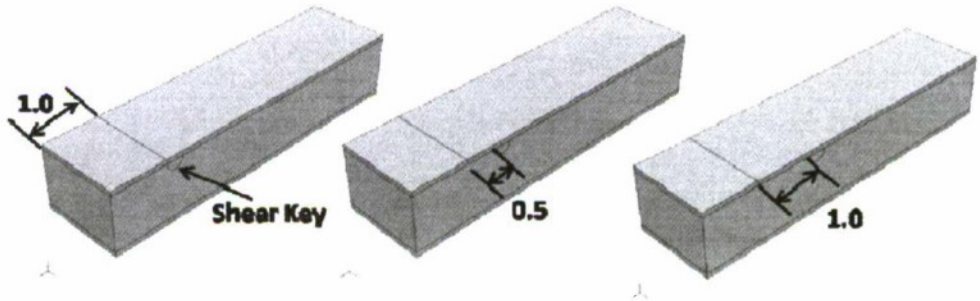


Figure 20 - Test specimen schematic for initial delamination with shear keys

Fourteen different fiberglass composite sandwich structures were manufactured and tested under a compression buckling load, which are listed below:

1. No delamination
2. 1 inch long delamination placed in the middle
3. 2 inch long delamination placed in the middle
4. Vertical shear key along specimen, no delamination
5. Vertical shear key along specimen, 1 inch long delamination placed in the middle
6. Vertical shear key along specimen, 2inch long delamination placed in the middle
7. Vertical shear key outside of delamination, 1 inch long delamination placed in the middle

8. Vertical shear key outside of delamination, 2inch long delamination placed in the middle
9. Horizontal shear key, no delamination
10. Horizontal shear key 1 inch away from delamination, 1.0 inch delamination in the middle
11. Horizontal shear key 1 inch away from delamination, 2.0 inch delamination in the middle
12. Horizontal shear key, no delamination
13. Horizontal shear key 2 inches away from delamination, 1.0 inch delamination in the middle
14. Horizontal shear key 2 inches away from delamination, 2.0 inch delamination in the middle

Eight different carbon fiber composite sandwich structures were manufactured and tested under a monotonic and cyclic compression loading, which are listed below:

1. No shear key
2. 1 Layer (2x3/4in strip) shear key
3. 2 Layer (2x3/4in strip) shear key
4. 3 Layer (2x3/4in strip) shear key
5. 4 Layer (2x3/4in strip) shear key
6. 5 Layer (2x3/4in strip) shear key
7. 6 Layer (2x3/4in strip) shear key
8. 7 Layer (2x3/4in strip) shear key

These composite sandwich structures were manufactured using the vacuum assisted resin transfer method (VaRTM).

2.1-1 Material Preparation - No Initial Delamination

Figure 21 below shows some of the materials required to manufacture a part using a VaRTM lay-up. All of the materials used to manufacture the composite sandwich structures are listed below.

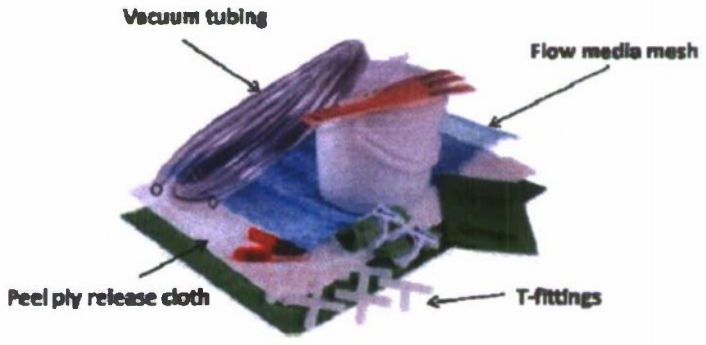


Figure 21 - VaRTM specific materials[21]

Table 1 – Table of Materials

List A of Materials:	List B of Materials:
• Woven roving fiberglass • Chopped strand fiberglass mat • Divinycell H 100 foam / LAST-A-FOAM FR-6710 • Peel ply release cloth • Flow media • Vacuum bag • Chromate tacky tape • Spiral tubing • T-fittings • Plastic tubing • Epoxy/Hardener • Mixing cups • Mixing sticks	• Pre-preg Carbon Fiber • LAST-A-FOAM FR-6710 • Peel ply release cloth • Flow media • Vacuum bag • Chromate tacky tape • Spiral tubing • T-fittings • Plastic tubing • Epoxy/Hardener • Mixing cups • Mixing sticks

During the material preparation phase, large Divinycell PVC foam sheets were obtained and cut into 13 x 10 inch squares using a jigsaw. 13 x 10 inch dimensions were selected to provide 12 test specimens from each lay-up batch.

Four layers of chopped strand mat and woven roving layers were then cut into 13 x 10 inch square sheets. The composite layers were weighed and used to determine the amount of resin and hardener needed to be mixed. The flow media was cut approximately 1 inch longer than the composite layers but with the same width. The peel ply release cloth was cut approximately 2 inch larger on all side than the flow media. The vacuum bag was cut approximately 2 inch longer on all sides than the peel ply release cloth. Two 15 inch segments of plastic tubing were cut and were used to aid the resin flow through the composite sandwich structure. One t-fitting and a segment of spiral tubing were cut to the width of the sandwich structure (10 inch). The spiral tubing was used to dispense the resin's flow through the structure from the t-fittings. Figure 22 shows the spiral tubing, vacuum tubing, and the t-fittings used for the lay-up and Figure 23 shows how the three items are used in the lay-up.

Figure 22 - T-fitting, spiral and vacuum tubing

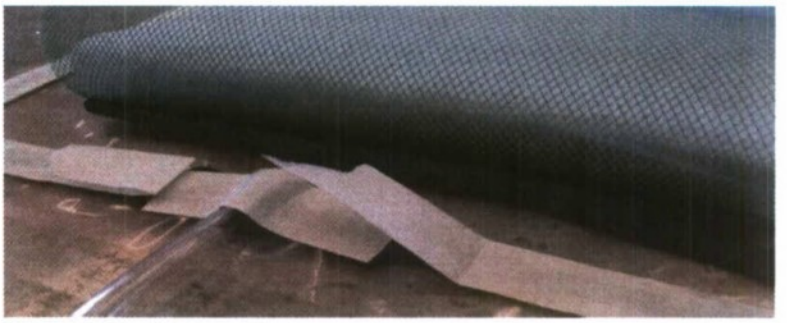


Figure 23 - Spiral tubing and t-fitting set-up

2.1-2 Material Preparation - Test Specimen with Initial Delamination

The composite sandwich structures with an initial delamination were manufactured in a similar manner compared to the test specimens without delamination but with one difference. A strip of non-porous Teflon material was added to the top of the foam which prevented the composite layer from bonding to the foam core (shown in Figure 24). The strip of non-porous Teflon paper was cut to approximately 0.25 – 0.33 inches. Although extra length was added to the Teflon, the margins were lost when the composite sandwich structure were cut to appropriate dimensions.

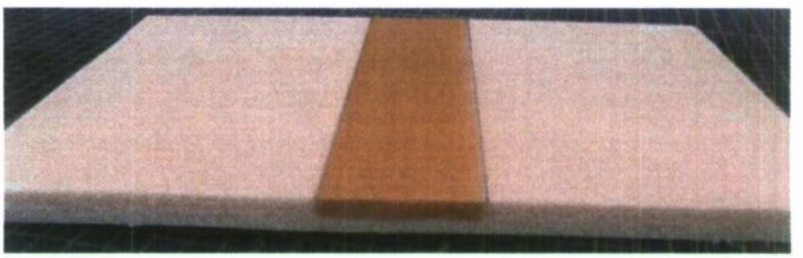


Figure 24 - Foam core with non-porous Teflon strip added

2.1-3 Material Preparation - Test Specimen with Initial Delamination and Shear Keys

The same fiberglass woven roving is used during the prep work for the arrestment shear key. Instead of cutting it down to form sheets, the fiberglass strands are pulled off of the weave and grouped together into a bundle of 17 strands. The 17 strand count are derived from how many of these strands can fit into the grooves on the mold and be manually capable of maintaining the cross-sectional shape of the arrestment key throughout the layup. The length of the strands has to be a bit longer than 12 inches to provide excess length for trimming and fitting perfectly into the composite sandwich structure. The fiberglass strands pulled off the weave can be seen in Figure 25.

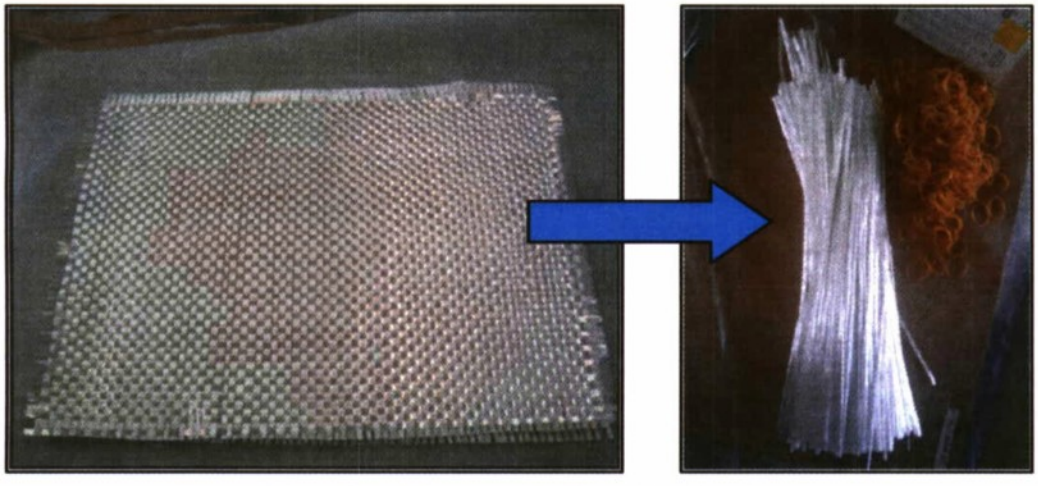


Figure 25 - Obtaining the Fiberglass Strands

Once the fiberglass strands are grouped in a set of 17 strands, the strands are held together with a rubber band on one end. It does not need to be a rubber band; it can be anything that holds the strands together and not impact the mechanical properties of the keys. This bundling process can be seen in Figure 26.

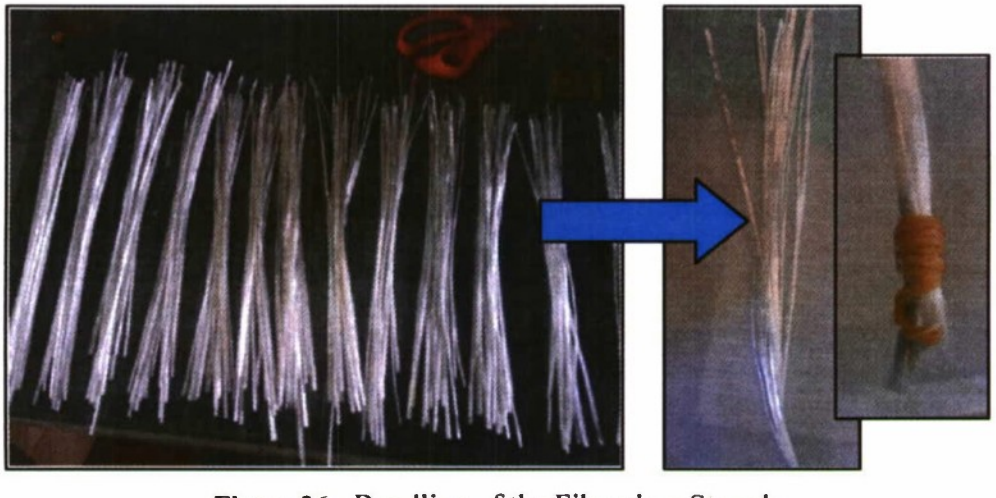


Figure 26 - Bundling of the Fiberglass Strands

With a cup of 5 to 1 ratio of the West System 105/205 epoxy, the sets of fiberglass strands are then infused manually to where there is an excess amount dripping when holding the set up vertically. The excess

epoxy will later be removed per pressurized vacuumed system. It is important to have enough epoxy rather than end up with dry fiberglass in the delamination arrestment key. Make sure that all parts of the sets are wetted with the epoxy. Once the set of fiberglass strands have enough epoxy, it is placed into the grooves on the mold. The mold is first coated with a non-stick wax and is made of Aluminum AL 2024. The surfaces and grooves of the mold are clean before every use to avoid any contaminates getting into the keys during the manufacturing process. The infused set of strands placed on the mold can be seen in Figure 27.

Figure 27 - Placing into the Mold

After all the sets of fiberglass strands are put in place, the whole mold with the premature keys is then placed into a vacuum bag and sealed. The vacuum bag has only one output to pull to create the necessary pressure on the keys and remove any unused epoxy in the vacuumed environment. This is displayed in Figure 28.

Figure 28 - The Vacuumed Mold and Premature Keys

To add additional support for the shape of the delamination arrestment keys, a uniform plate and 200 pounds of additional weight is placed on top of the vacuumed structure. This is done because the pressure in the vacuumed environment would otherwise leave a thin layer of epoxy through the part. In order to help remove

the excess epoxy, the weights are added and also provide a flat backing for the shape of the keys. Figure 29 provides a better look into how the weights are placed.

Figure 29 - Weights used on Mold

After the curing of the keys is completed, the structure is then taken out of the mold. At this point, the rubber band section of the layup is later cut off to remove what does not belong in the composite sandwich structure. The keys are cut down along its length to reduce most of the excess materials before being placed back into the mold where the keys are then sanded down to the correct dimensions. A picture of the sanding process is displayed in Figure 30.

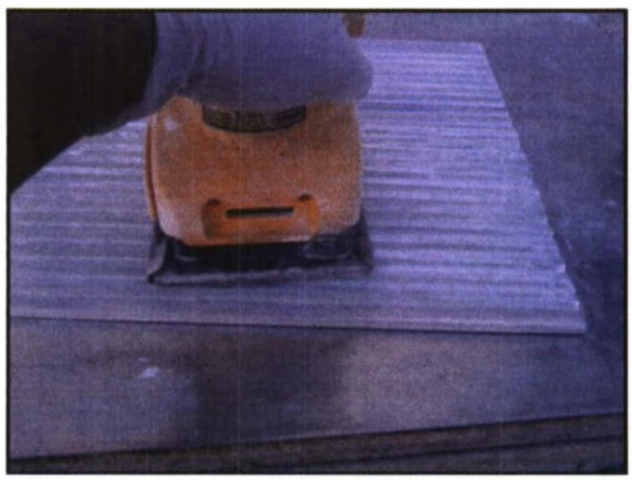


Figure 30 - Sanding of the Keys

The keys would look something like Figure 31 and Figure 32 – Shear key mold and CAD drawing used in production of actual shear keys. after the sanding and cleaning of the keys is completed. The cleaning of the key is done with water and nothing else. All residues on the exterior of the keys are removed, and the keys are stored in a clean environment.

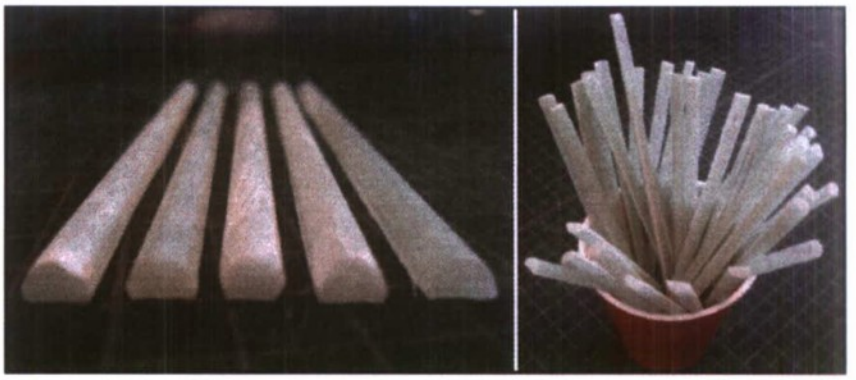


Figure 31 - The Keys Completed

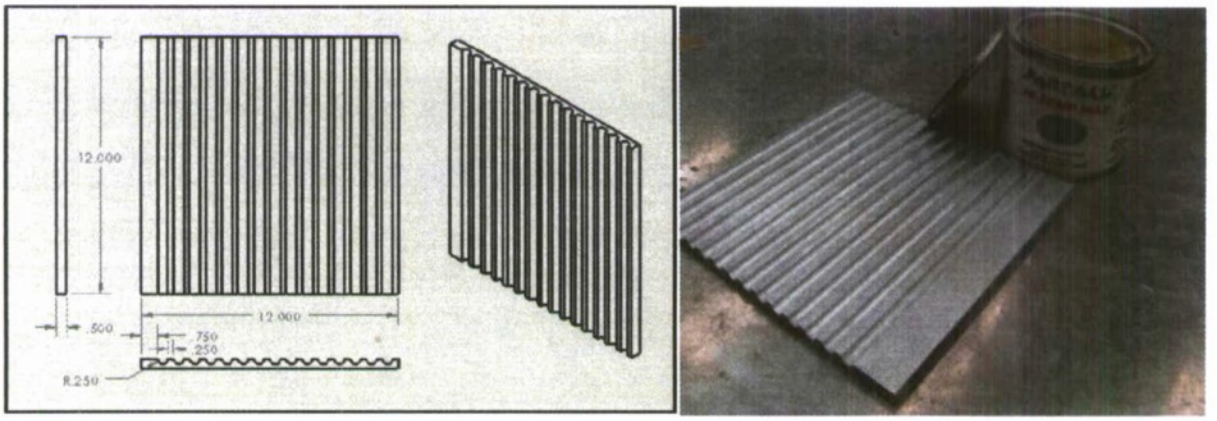


Figure 32 – Shear key mold and CAD drawing used in production of actual shear keys.

The sandwich structures with shear keys were manufactured in a similar manner to the test specimens with an initial delamination. A piece of foam with the same dimensions used for the previous lay-ups had 4 mm radius grooves milled out using a manual mill. Then a strip of non-porous Teflon film was added in between the shear keys on the foam just like the test specimens with an initial delamination. As before all of the vacuum bagging materials were gathered and set-up to the point right before the vacuum bag was sealed. A small amount of epoxy was mixed (approximately 30 g) which was used to bond the shear keys to the foam. After the epoxy was adequately mixed then the epoxy was applied to the milled out foam regions using a wooden popsicle stick. Finally the shear keys were inserted into the grooves and the upper skin of fiberglass was stacked on top and the bag was sealed up for the resin to flow through.

2.1-4 Material Preparation - Buckling Test Specimen

Due to the needed pinned-pinned boundaries, the test specimens needed to be manufactured with a stronger material at the boundary. This required a stiffer material that is able to bond with the fiberglass skin and has the same dimension as the foam core. The material used for the boundary is oak wood and is used in the composite sandwich structure that supports the boundaries of the structure under in-plane displacement. The oak

wood core was also chosen to provide enough stiffness to maintain the integrity of the hole's boundary locations. If the holes were manufactured into the foam core, it would saw through the test specimen.

The oak wood was the best choice for its stiffness and its significantly high loading capacity. One note that was considered during the manufacturing process was the grain direction of the oak wood material; hence the direction of how to manufacture the wood core into composite sandwich structure matters. In the layup of the composite structure, the wood grain direction would be laid up parallel to the load vectors.

The wood core design was used to determine the elastic modulus in two different directions where one direction has the grain parallel to the load vector and the other direction is perpendicular to the same vector. From how the oak wood arrived in the lab, the raw material would be cut down into small blocks that were used during testing. The design of the specimen has the following dimensions where the length of the specimen is parallel to the grain:

Table 1 - Wood Core Design Dimensions for Compressive Properties

Width (in)	Overall Length (in)	Thickness (in)
1.0	1.5	0.5

The delamination within the specimens was created in the same way as the specimens created for the tensile loading. These designed test specimens would later be placed on the tee shaped jig and compressed to get the load and vertical extension in order to determine the compressive elastic modulus with the given geometry.

2.1-5 VaRTM Lay-up

This lay-up procedure was used for the tensile loading and compression buckling fiberglass pieces. Once all of the materials had been cut and prepared the materials were stacked in the lay-up level. The vacuum bag was laid out on a flat surface and the flow media was placed on top of it followed by the release cloth. Components of the composite lay-up were assembled in their appropriate lay-up level and stacked on top starting with the woven roving and the chopped strand mat. Another layer of woven roving and chopped strand mat were added followed by the foam core. Two additional layers of chopped strand mat and woven roving were placed on top of the foam core in an alternating fashion. Chromate sealant tape was used to surround the perimeter of the sandwich structure lay-up in order to seal the vacuum bag later. Segments of plastic tubing were placed at either ends of the sandwich structure to aid in the direction of resin flow from the vacuum pump. A small piece of cotton was added to the tubing closest to the vacuum to increase the bag pressure. The peel ply, flow media, and vacuum bag were then folded over to envelope the entire sandwich structure shown in Figure 33.

Figure 33 - Stacked VaRTM lay-up before vacuum bagging

The vacuum bag was sealed off using the Chromate sealant tape that was applied to the perimeter of the vacuum bag. Care was taken while sealing the vacuum bag to minimize air leaks. The vacuum end of the tubing was connected to the resin trap and sealed off using another thin strip of sealant tape. A leak check was performed prior to running the resin through the part. To check for leaks, a piece of sealant tape was applied to the resin side of the tube and the vacuum was turned on. If leaks were discovered, then the remaining leaks were sealed off with additional Chromate sealant tape. Caution was also taken to correctly line-up the composite fabric sheets. If the sheets were not correctly place, the applied vacuum pressure would cause the contents of the bag to shift around. After double checking the placement of the composite sheets and leaks, the sealant tape on the resin side of the tube was removed and the tube was closed off using vice grip pliers (shown in Figure 34).

Figure 34 - Vacuum bagged composite sandwich structure after lay-up

West System's 105 Epoxy System and 206 Slow Hardener were used as the matrix for the composite sandwich structure lay-up (shown in Figure 35). The amount of resin and hardener used were determined based upon the weight of the dry fibers. The total weight of the dry fibers was multiplied by 1.5 which allowed

extra epoxy to ensure that the fibers were properly saturated. From earlier lay-up experiments it was determined that the total amount of epoxy should be divided into two separate mixing cups in order to prevent the epoxy from curing before it had the chance to travel the entire length of the part. The resin-to-hardener ratio for the resin system used was approximately 3 to 1.

Figure 35 - West System resin and hardener used for lay-up

The resin and hardener were measured out into two separate cups using a scale. When ready, the two were combined and stirred until fully mixed. The pliers were removed from the resin side of the tubing and the end of tube was placed in the resin cup. Figure 36 shows the resin flowing from the cup through the tubing to the vacuum sealed composite sandwich structure. The pressure from the vacuum pulls the resin across the sandwich structure and through the flow media. The second cup of mixed resin was poured into the 1st resin cup when it was almost empty. After all of the resin had flowed across the sandwich structure then vice grips were used to close off the resin side of the tubing to ensure that the part remains under constant vacuum pressure. Then a flat sheet was placed on top of the sandwich structure and weights were added to evenly distribute additional pressure. The part is left under vacuum pressure overnight for approximately 12 hours.

Figure 36 - VaRTM composite lay-ups resin flow

Figure 37 below shows an exploded view of a VaRTM with all of the materials labeled.

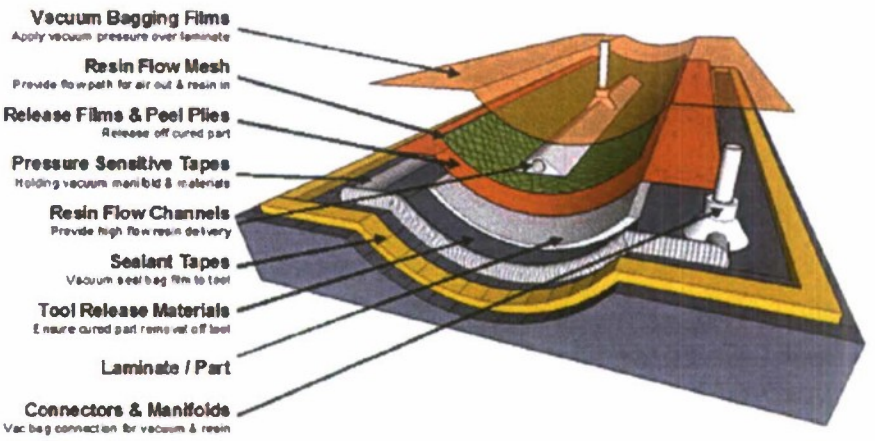


Figure 37 - Composite materials stacked before lay-up[22]

2.2 Pre-Preg Lay-Up (Richard)

In the third application, the composite sandwich panel consisted of the foam core, 4 sheets of LTM45 (2 layers on each side), and two sheets of film adhesive. The dimensions for all the materials were 12" x 12". Two layers or two sheets of non-porous material were used between the composite sandwich panel and the metal plates for the press. The purpose of non-porous material for the lay-up was to absorb any excess resin from the plates and prevent it from curing on the plates and causing dimples on the face sheet. It acts like a protective barrier for the composite sandwich panels. The lay-up process used for manufacturing the composite sandwich panels can be seen in Figure 38.

Figure 38 - Schematic for Lay-up of Composite Sandwich Panels

The first step of manufacturing the composite sandwich panels is to sand and wax the press metal plates to ensure smoothness of the face sheets; this can be seen in Figure 39. The plates are sanded down by using a high grit sandpaper to remove any resin that was cured on the metal plate. High grit sandpaper was used to prevent the removable of the magnetic coating on the plate. After the plate was sanded, a wax was used to

smooth any divots on the metal plate. The wax was applied by hand and a buffer tool attached to a motorized drill was used to smooth out the wax that was on the plate.

Figure 39 - Sanding and Waxing of Press

With all the materials prepped and ready, the materials are laid-up the way shown in Figure 38. The composite sandwich panel was placed into the press. The manufacture of LTM45 has a specific curing cycle, which is curing the composite by applying heat and pressure on it for a specific amount of time. The manufacturer's curing cycle can be seen in Figure 40. The cycle first starts out with applying a constant load of 1000 lbs on the composite sandwich with the temperature increasing 2 ^{0}F per min until the temperature reaches 150 ^{0}F. The press then stays on for 16 hours at the temperature and the same force. After the time elapses, the temperature decreases 2 ^{0}F per min until the temperature reaches 75 ^{0}F with a constant force of 1000 lbs. After the machine reaches room temperature the machine runs for another 2 hours with 1000 lbs of force to end the cycle. According to the manufacturer, the composite would be 98% fully cured after the LTM45 has gone through the cure cycle. The plate would then be removed from the press and cut to dimensions of the test specimens, which will be further discussed.

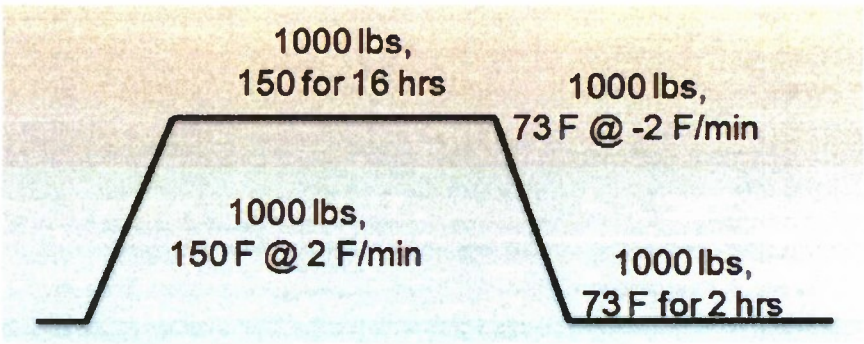


Figure 40 - Curing Cycle for LTM45

2.2-1 Damage Arrestment Device Prep (DAD)

Composite sandwich panels with DADs were manufactured first with milling the slots on the foam; this can be seen in Figure 41. A CNC end mill was used to mill the foam with a 0.75" four-flute straight end mill. A CAD model of the foam was made and was placed into CAMWORKS to create a CNC code to make the machine mill the foam to provide a better accuracy on the depth and alignment of the slots. Special care of the foam is vital for milling the slots to prevent the foam from being destroyed since it is flimsy. A vice grip cannot be used on the foam to keep it in place because it would leave mark and divots on the foam. To keep the foam in place, six toe clamps were used with three rulers going horizontally across the foam. The toe clamps and the rulers would be placed around the slots to prevent the machine milling any of it out. The foam was milled on one side and flipped around to mill the other side, making sure that the milled out cuts would be aligned on both sides of the foam. The mill depth of the foam was set to 0.039" which is the thickness of three layer of LTM.

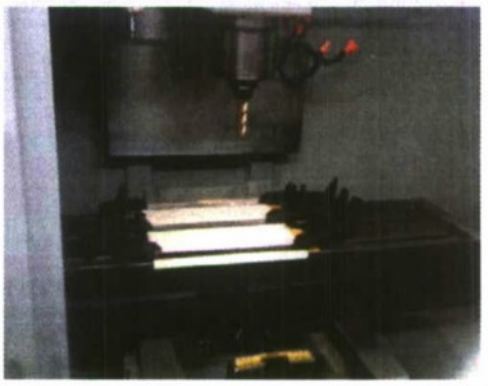


Figure 41 - Milling of foam for DADs

With the foam milled, the DADs would be created using LTM45 cut into strips of 12" x 0.75". A 12" x 12" sheet of LTM45 is marked with dimensions of the DAD strips and is cut using a Rotatrim to have a clean cut; this can be seen in Figure 42. The marking of the sheet was done using a Sharpie and a ruler. This process is repeated until 24 strips are cut to for one 12" x 12" foam plate, having 3 layer DADs per slot.

Figure 42 - Cutting of the DAD strips

With the strips cut, the strips would be set into eight piles of three strips. The protective papers of the strips were removed and laid into the milled out slots of the foam one at a time; this can be seen in Figure 43 - DADs being inserted into foam slots.

Figure 43 - DADs being inserted into foam slots

The rest of the manufacturing of the plate is the same as the regular test specimens. It would consist of four sheets of LTM45 (two layers on each side), and two sheets of film adhesive, and 4 sheets of non-porous material (two sheets on each side). The panel would go through the same curing cycle as before; the final product can be seen in Figure 44.

Figure 44 - Finish specimen with DADs

The plates were then removed from the press and were prepped into the final specimen dimensions.

Chapter 3- Experimental Procedure and Testing

3.1- Monotonic and Cyclic Tensile Loading (Eugene)

The experimental testing of the composite sandwich structures were performed in the Cal Poly Aerospace Structures and Composites Lab. Both the static (monotonic) and dynamic (fatigue) testing were performed using an Instron 8801 machine. Two different machine configurations were used specific to either the high load or low load case. The high load cases utilized a 100 kN load cell which was used to test composite sandwich structures with no initial delamination. The high load Instron configuration is shown in Figure 45 - 100 kN load cell Instron grip configuration.

Figure 45 - 100 kN load cell Instron grip configuration

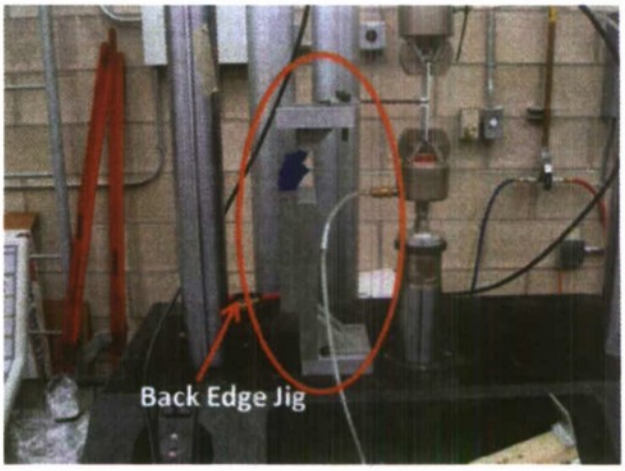


Figure 46 - 1 kN load cell configuration and back edge jig

The low force cases used a 1 kN load cell to test specimens with an initial delamination and shear keys. The 1 kN load cell configuration with its appropriate Instron grips is shown in Figure 46 - 1 kN load cell configuration and back edge jig. The aluminum tabs on the test specimens were secured into place by closing the Instron grips. The test specimens were also held into place at the back end using a jig as shown in Figure 46 - 1 kN load cell configuration and back edge jig.

3.1-1 Monotonic Testing

Static tests (monotonic testing) were performed on the six different types of test specimens to determine the ultimate loads and failure behavior for each case. The failure loads were required inputs to perform the dynamic (fatigue) testing because the fatigue loading is a function of the monotonic failure. From the previous delamination research the loading rate was investigated under speeds of 0.5, 1.0 and 2.0 mm/min. It was discovered that slower loading rates introduced difficulties with the delaminated test specimens[5].
For that reason a loading rate of 2.0 mm/min was selected. The program used for the monotonic testing was Merlin provided by the Instron. Merlin's failure criteria for testing, was that the loading rate for the tested needed to drop by 50% in order for the tests to stop. Figure 47 - Inputs for monotonic testing shows a screenshot of Merlin before testing, when the inputs are specified Figure 48- Merlin screen during testing shows Merlin during a test.

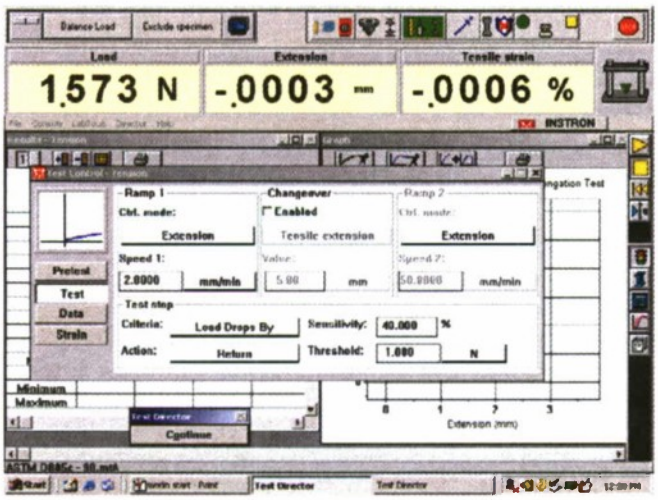


Figure 47 - Inputs for monotonic testing

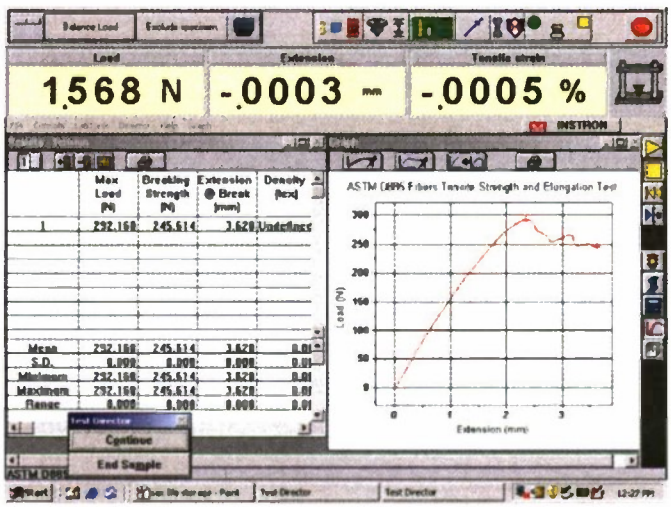


Figure 48- Merlin screen during testing

3.1-2 Fatigue Testing

The dynamic testing was performed on the composite sandwich structures to determine the fatigue behavior. The overall goal of the fatigue testing was to construct a stress vs. number of cycles to failure (S-N) curve for the following three sandwich structure configurations:

- No initial delamination
- 0.5 inch initial delamination
- 0.5 inch initial delamination with shear key

The fatigue testing was performed using the Instron program Single Axis MAX (SAX). A sinusoidal cyclic loading was applied to the test specimen from the Instron machine. SAX can either be driven by inputting force or displacement conditions. Force inputs were selected since the monotonic failure characteristics of the sandwich structures were defined in terms of force. The four important parameters that were required inputs into SAX were: frequency, sample rate, mean force, and force amplitude. Figure 49 - SAX input parameters shows a screenshot of SAX and the input parameters required for testing.

Figure 49 - SAX input parameters

Two important inputs were the force amplitude and mean, which fully defined the sinusoidal wave. Preliminary tests revealed that using the 100 kN load cell put a minimum amplitude requirement of 100 lb_f for the sin wave to correctly track the input force equal to the output force. This was most likely attributed to the notion that the large load cell has a much smaller force tolerance compared to the 1 kN load cell. This requirement was a driving constraint for determining the force amplitude and mean. From the monotonic testing it was determined that only the test specimens with no initial delamination could be tested using the 100 kN load cell. For the dynamic testing the ratio of the mean force was kept constant for each of the three different composite sandwich structures. Since the non-delaminated test specimens were tested first they dictated the testing parameters for the delamination and shear key test specimens. A mean force of 40% of the max failure load was selected because it allowed for a large number of testing loads to be performed using the 100 kN load cell. Figure 50 - Monotonic force definitions shows an example monotonic force vs. deflection curve and how the ultimate force was selected from the curve.

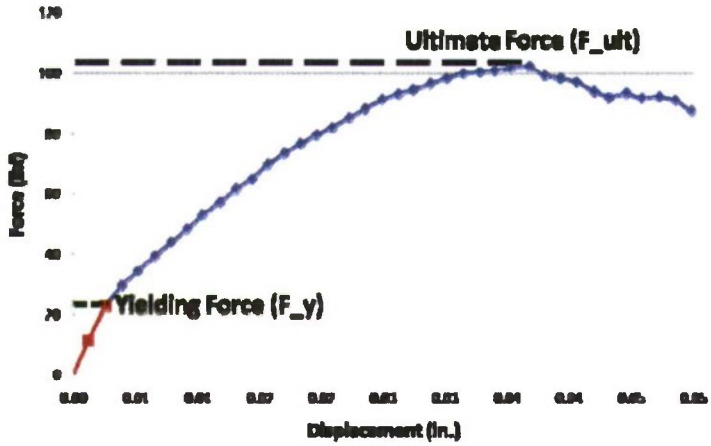


Figure 50 - Monotonic force definitions

Figure 51 - SAX input definitions shows an arbitrary sin wave force similar to the one applied for the fatigue testing. To construct the stress vs. number cycles to failure curves (S-N) the specimens were tested at the following percentages of the ultimate monotonic failure forces: 70, 65, 60, 55, and 50. The mean force in Figure 51 - SAX input definitions is represented by the dashed line and was always equal to 40% of the ultimate monotonic failure force. The force amplitude was varied throughout the testing for the different failure force percentages listed above. Example force inputs for an arbitrary case are shown in Table 2 below.

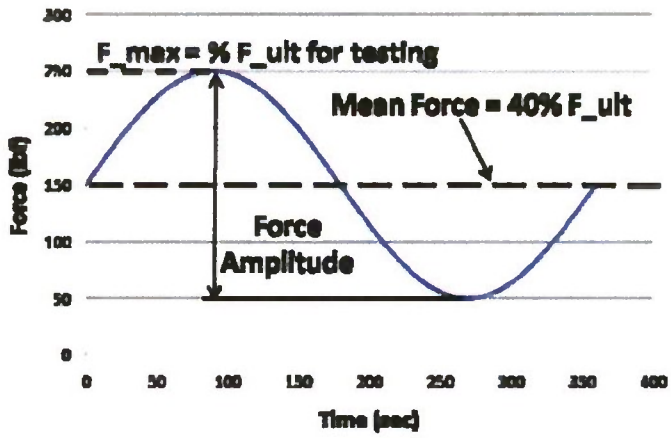


Figure 51 - SAX input definitions

Table 2 - Example input forces

F_{ULT} (lbf)	F_{MEAN} (lbf)	Testing %	F_{MAX} (lbf)	F_{AMP} (lbf)
550	220	0.7	385	330

The two other inputs into SAX are the frequency that the machine operates at and also the sampling rate. A frequency of 10 Hz was chosen in an effort to speed up the testing process since fatigue and particularly low force fatigue takes a considerable amount of testing time. The chosen frequency was deemed "not too unrealistically high" to adversely affect the results to the point they would be irrelevant. During the initial testing and calibration phase of the fatigue research, the sample rate was varied. The sample rate not only had an effect on the number of data points stored, but it also affected how the sin wave was applied to the test specimens. A small sample rate had the advantage of data storage but the sin wave would correctly track input equal to output. A sample rate of 0.1 kHz was selected because it was the minimum sample rate that allowed for a correct and smooth sin wave to be input.

To mitigate manufacturing defects on data scatter it was determined that the monotonic failure loads would be determined from each sandwich structure made. Each manufactured sandwich structure yielded 12 test specimens so 2 test specimens were used to determine the failure loads for that particular lay-up batch. This method was used to provide more reasonable test results and present a distinct trend in the data.

3-2 Monotonic and Cyclic Compression Loading (Richard)

3.2.1 Strain Rate Testing Selection

A strain rate, according to testing standards, is important for monotonic testing of composites. Since each composite structure is greatly different from one another, one type of strain rate for one structure would be bad for another. A quick strain rate would distort failure results by moving the failure of the fastener/panel interaction quickly from bearing to shear or even to a complete failure of the panel. A slow strain rate would give good results but would take too long to test each specimen. For the strain rate selection, 5 strain rates would be tested ranging from 0.5 mm/min to 1.5 mm/min with 10 specimens per test case and a variance of 0.25 mm/min between the strain rates. The standard for testing metal pieces is a strain rate of 1 mm/min. Since there is no set standard for strain rate for composites, the speeds that were chosen is reasonable to test two rates above and below it for the composite sandwich panel. A failure criterion of a 30% load drop was chosen as the end of test for the monotonic testing. This means that when a total failure drop of 30% for the most maximum load would stop the machine.

All specimens were tested and analyzed; this would be discussed later in detail in Strain Rate/Control Group Results and will be briefly discussed in this section to choose a strain rate for all static testing. A summary of the results are shown in Table 12 in Strain Rate/Control Group Results. The results showed that all five strain rates had similar failure load varying from 890 lbs to 1,011 lbs and had similar elastic modulus varying from 45,170 psi to 50,250 psi. The results show that the strain rate did not greatly affect the characteristics of the panel, but it did show a variance in standard deviation for each case. The table shows that the slowest rate, 0.5 mm/min had the lowest standard deviation for both failure load and elastic modulus. This

rate would be used for all static testing because it gives the best standard deviations compared to the other rates translating a better accuracy to future tests.

3.2-2 Static Testing Design

From the detailed literature review, all of the fastener/panel interaction testing was conducted under a tension loading. A compression loading was chosen because a compressive loading would produce a more important failure mode to the panel due to buckling being seen during the testing. A buckling failure would be more of a catastrophic failure for a panel rather than a tear or crack propagating from a tension load. The static testing portion of the experiment would consist of finding the optimum strain rate for testing pieces under monotonic loading and to find the trend of how increasing the thickness of the DADs affect the fastener/composite panel interaction.

From the previous section it was decided that for all static testing the strain rate for the machine to run test would be 0.5 mm/min or 0.0197 in/min. The DAD thickness would be varied from 0 layers to 7 layers. Each case would have a total of 10 specimens to be able to observe the interaction regardless whatever the standard deviation for the case is. A small standard deviation is not necessary since a proof of concept is the purpose of the research and it is assumed that it would have an adequate enough standard deviation from the results shown in the previous section. A failure criterion of a 30% load drop was chosen as the end of test for the monotonic testing. This means that when a total failure drop of 30% for the most maximum load would stop the machine.

3.2-3 Fatigue Testing Design

Fatigue testing is when a part is loaded to certain value, usually below the yield stress of the material, and is applied a load amplitude to be cycled. This would show when the material or case would fail over time from normal loading. This is helpful with designers because knowing how the material fails would help design how and when the overall system fails. The load amplitude is usually a sinusoidal wave with its load value varying a percent difference from the yield stress; this can be seen in Figure 52 - Fatigue Cycle. The test is first ramped to half the yield stress. From the detailed literature review and ASTM standards, the yield stress is defined is where the first drop on the stress-strain curve; this is seen in Figure 53 - Yield Stress and Failure Criteria for Fatigue Testing. The stress would be converted to load by the geometry of each test piece, which becomes the force yield. The amplitude, or A, is a dependent on the fatigue testing and what the researcher is investigating with. One fatigue cycle is equivalent to one cycle of the sinusoidal wave. The frequency of the wave is arbitrarily chosen by the researcher. Once the test reaches its failure criterion the test stops and ramps the load back to zero. The criterion is dependent on the researcher. There are two ways to define the failure criterion, a drastic load drop or a drastic change in position of the test specimen.

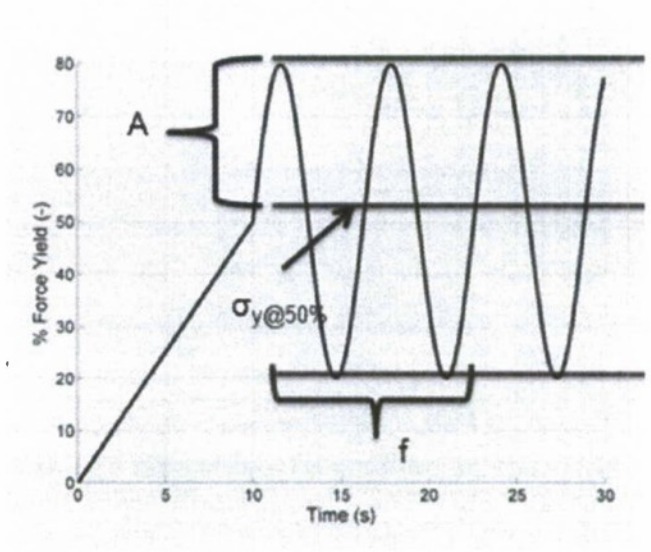


Figure 52 - Fatigue Cycle

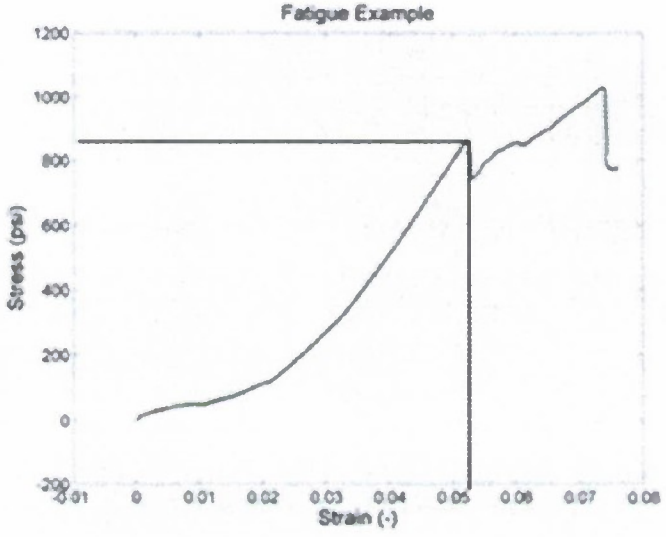


Figure 53 - Yield Stress and Failure Criteria for Fatigue

Two selections were made to fatigue testing for the experiment. The first selection was to choose the failure criterion for the testing. Since the yield stress was chosen as the maximum for the amplitude, the strain at that stress on the stress-strain curve was chosen as the failure criterion. The strain, like the stress, was converted to extension based of each test specimen's geometry and was used as the position failure for the test. Once the test specimen moved by the amount of the extension, the specimen was defined to reach failure. This was also supported from preliminary testing by having a bearing failure around the holes; this would be further discussed.

The second selection for fatigue testing was choosing the frequency to run the testing on. A too high frequency would give inaccurate results because it would be applying too many cycles per second on the

specimen. A low frequency would give more accurate results but would take too long per specimen to run. Two frequencies were investigated to see which one to use for fatigue testing, 5 and 10 Hz. Both test cases were tested with 10 specimens each. The number of cycles of failure and its standard deviation for the two frequencies is shown in Table 3. The results show a drastic difference between the two frequencies, the lower frequency has a lower number of failure cycles while the higher frequency has a higher number of failure cycles. For standard deviation, both have the same similar trend. Since both have high standard deviations, the higher frequency was chosen because it seemed that it would produce a higher average number of cycle failure. It also gives a decent run time, especially if the specimens reach the hundreds of thousands of cycles, which would take hours to run one specimen.

Table 3 - Summary of Fatigue Frequency Testing Results

Frequency	Avg # of Cycles	Std Dev
(Hz)	(-)	(-)
5	5555	69.46%
10	7032	87.41%

For the fatigue testing for the research, two case groups would be run to compare the effects of the DADs on the composite sandwich specimens. The first group would be the control group, no DADs in the specimens, and the other group would only have one type of DAD thickness in each specimen. The trend of how the number of failures increases as the fatigue amplitude decreases would be observed and analyzed for both case groups. The trends would show how the DADs affect the overall lifespan of a composite sandwich panel and how the failure around the fastener happens. It will be assumed that different DAD thicknesses would have similar trends but would be shifted along the graph. The amplitude for each case group would be varied from 90% yield stress to 65% yield stress with 5% spacing, having a total of six test cases. Each test case would have a total of 10 pieces.

The thickness of the DADs for the other case group would be 3 layers. This was chosen because a too thick of a DAD would affect the core/face sheet interaction. The ratio of the total thickness of the DAD over the thickness of the core was set to less than 0.2. The ratio value because it assumes that the insert thickness is still within the infinite range, similar to the ratio value for the hole diameter and the width of the specimen. The 3 layer DADs is the closest to the ration without passing the value, making it the best choice.

3.2-4 Static/Fatigue Testing Set-up

The static/fatigue testing of the specimens is pretty simple and straight forward. The specimens are installed on the jig and the software for each respective testing is conducted. The testing specimens are first measured and recorded. The width, length, and thickness of the specimens are measured using a calculator. The

measurements are recorded on Goggle Docs for easy access anywhere; an example of the recording of measurements can be seen in Figure 54 - Example of recording measurements in Goggle Docs

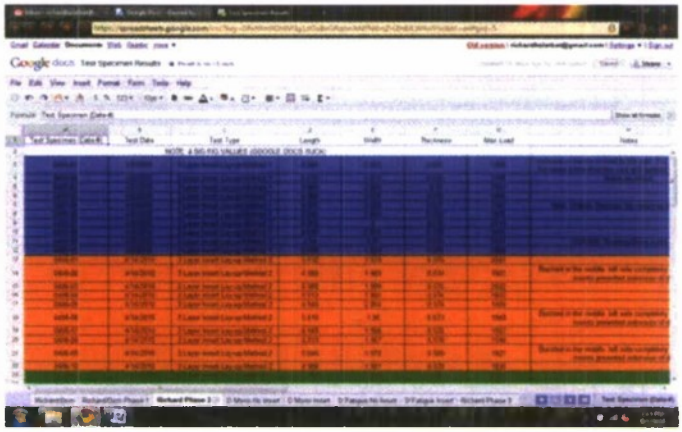


Figure 54 - Example of recording measurements in Goggle Docs

Specimens are then taken to the machine and attached to the jig; this can be seen in Figure 55 - Testing Set-up. Bushings are placed inside the hole and are lined up with the holes on the jig. The bolts would contain two washers, placed on the outside of the jigs so the bolt and nut heads aren't touching the jigs. The bolt will be tightened until resistance is made with the jig. The software would then be run until the specimens reach the failure criteria. For static testing, the failure criterion is when the load drops more than 30%. For fatigue testing, the failure criterion is when the position of the test specimens moves a certain amount. For specimens without DADs, the failure criterion is 0.0544" and for specimens with DADs, the failure criterion is 0.0597". All values were calculated from experimental testing. Once the specimens reach the failure criterion, the specimens are removed from the testing apparatus and pictures were taken to document the failure for each specimen. The load failure is recorded on Goggle Docs for future analysis. The specimens are stored for later use and analysis.

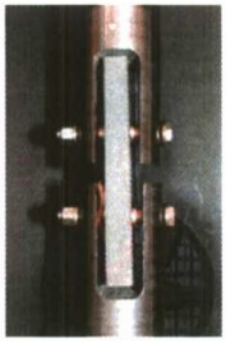


Figure 55 - Testing Set-up

3.3 Monotonic Compression Buckling Load (Tony)

3.3-1 Testing for Mechanical Behaviors under Buckling using Pinned-Pinned Boundary

According to ASTM C-364 standard method, the quasi-static strain rate is 0.2-inch per minute. The setup of the test jig is mounted directly to the displacement unit and the load cell of the Instron 8800 displacement machine. Instron LVDT Displacement Gauge is placed perpendicular to the test specimen and the load vector in order to get the horizontal displacement the specimen is seeing. The test setup can be seen in Figure 56 - Pinned-Pinned Boundary Set Up.

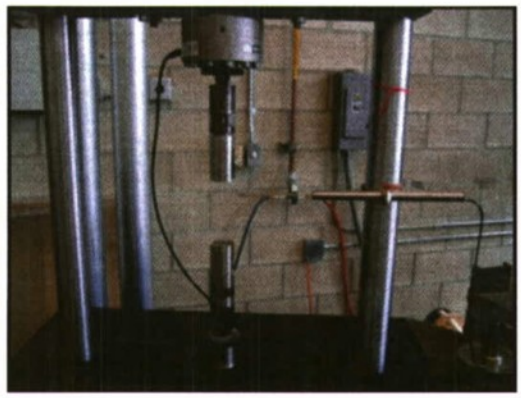


Figure 56 - Pinned-Pinned Boundary Set Up

The test specimen is placed into the pinned-pinned boundary with aircraft grade bolts, and fastened to avoid any horizontal play. This restricts the specimen to vertical and horizontal movement with rotation capabilities around the hole location. The displacement gauge is lean again the test specimen at its center and measures its displacement. The specimen is loaded digitally using Bluehill 2 software can be seen in Figure 57 - Testing Using Pinned-Pinned Boundary and all raw data is recorded for future processing.

Figure 57 - Testing Using Pinned-Pinned Boundary

Chapter 4- Experimental Analysis

The following section is an overview of the experimental results of the three applications being considered: The monotonic loading and fatigue case, the delamination arrestment mechanism case, and damage arrestment device case. To view the full experimental results of each project, see the individual project thesis.

4.1- Monotonic Loading and Fatigue Case (Eugene)

4.1-1 Monotonic Failure Analysis

The test specimen stiffness$\left(\frac{lb_f}{in.}\right)$, was used to compare the monotonic test results for the 6 different test specimen configurations. The fracture mode of the specimens was also of importance because it varied depending on the sandwich structure configuration. The composite sandwich structures had three main material elements where failure may have initiated from: the matrix, the fibers, and/or the foam. From a mechanics point of view, the failure should occur in the weakest of the three materials under constant loading conditions. In all three cases, the failures were caused by the development of cracks in one of the three materials.

No Initial Delamination–Cracks in the Foam

Most of the test specimens with no initial delamination failed due to the formation of cracks in the foam. Figure 58 - Monotonic failure of specimens with no initial delamination below shows three pictures of a test specimen with no initial delamination and how they failed under monotonic loading. Cracks in the foam normally originated near the upper or lower surfaces as shown on the left hand side of Figure 58 - Monotonic failure of specimens with no initial delamination. Cracks then traveled through the foam at an angle, as shown in the middle and right side of Figure 58 - Monotonic failure of specimens with no initial delamination. The crack stops propagating through the foam when the failure criterion in the Instron machine is reached. These test specimens did not have a symmetric failure on both sides of the foam because the crack traveled across the width of the part as opposed to through the length. It was expected that cracks would initiate in the foam for correctly manufactured test specimens because it was considered the weakest of the three elements.

Figure 58 - Monotonic failure of specimens with no initial delamination

No Initial Delamination – Cracks in the Matrix

The second primary mode of failure in the test specimens was caused by cracks originating in the matrix Figure 59- Monotonic failure of test specimens without initial delamination shows test specimens that failed due to cracks in the matrix. The picture on the left side shows a crack that started at the top of the test specimen and the picture on the bottom shows a crack that originated on the bottom of the specimen. The majority of the cracks in the matrix originated between the bottom layer of chopped strand mat and the foam. When the specimen failed, the chopped strand mat would delaminate from the foam. The delamination was not smooth because the some of the chopped strand mat would remain bonded to the foam. Cracks in the matrix signified the composite sandwich structure was not properly manufactured because the laminate is a stiffer material than the foam. A possible explanation for the cracks could be traced back to the resin/hardener ratio used during the manufacturing process. The resin/hardener ratio could cause the composite to cure as the resin was running through the part, leaving some of the regions of the lay-up with dry fibers.

Figure 59- Monotonic failure of test specimens without initial delamination

0.5 inch Initial Delamination – Cracks in the Foam

All of the test specimens with a 0.5 inch delamination failed due to the propagation of cracks through the foam core. Figure 60 - Monotonic failure of test specimens with 0.5 inch initial delamination shows 4 pictures of test specimens with a 0.5 delamination that have failed. In all of the specimens tested, the cracks originated at the edge of the delamination and then continued to progress through the foam at various angles. The angles at which the cracks traveled through the foam was the only difference between the failures in the test specimens. Both pictures on the upper portion of Figure 60 - Monotonic failure of test specimens with 0.5 inch initial delamination failed due to a crack traveling approximately at 0° from the edge of the delamination. In the two pictures on the bottom of Figure 60 - Monotonic failure of test specimens with 0.5 inch initial delamination traveled down through the foam at angles of approximately 45° and 30°. The failures through the foam signified a good bond between the laminate and the foam core. This was also somewhat unexpected because it was predicted that a delaminated piece would exhibit a peeling effect, causing the delamination to separate the laminate from the foam.

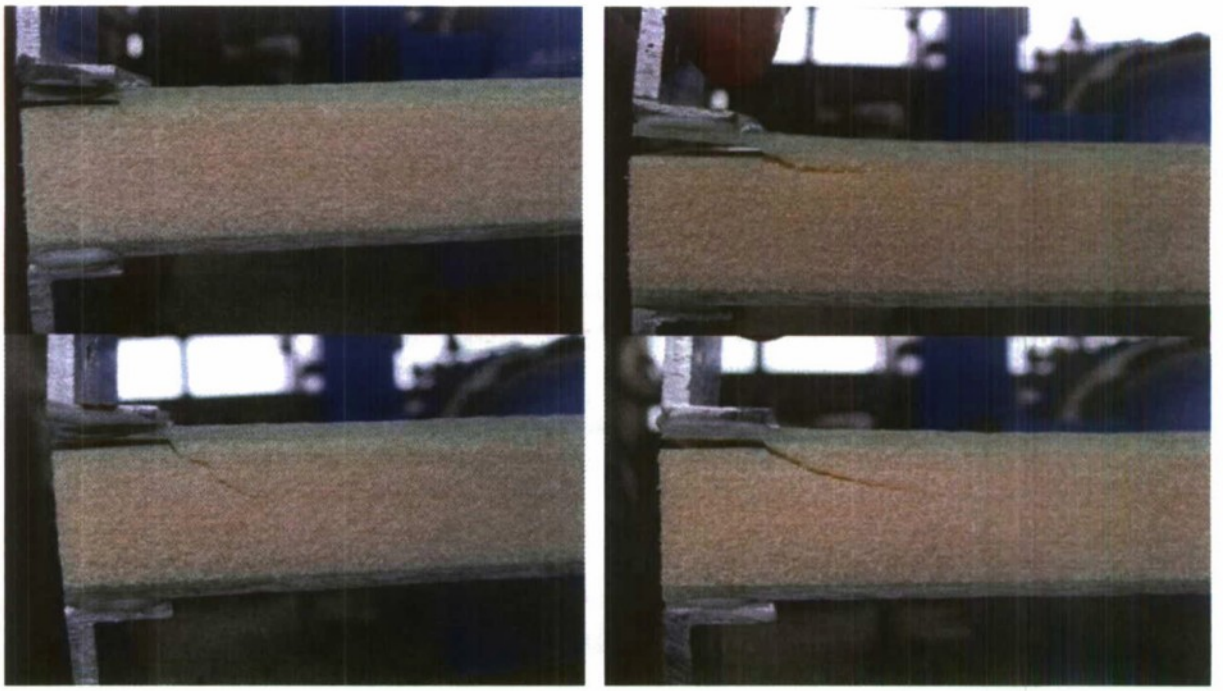


Figure 60 - Monotonic failure of test specimens with 0.5 inch initial delamination

1 inch Initial Delamination – Cracks in the Foam

The majority of the test specimens containing a 1 inch delamination failed due to cracks originating in the foam. Figure 61 - Monotonic failure of test specimens with 1.0 inch initial delamination shows two test specimens that failed due to cracks in the foam core. The 1 inch delaminated test specimens failed in a similar fashion compared to the test specimens with a 0.5 inch delamination. Cracks originated at the edge of the delamination and then traveled downward through the foam, as shown in the upper part of Figure 61 - Monotonic failure of test specimens with 1.0 inch initial delamination. The lower part of Figure 61 - Monotonic failure of test specimens with 1.0 inch initial delamination shows a test specimen where the crack in the foam initially travels downwards but began to travel back through the foam towards the composite skin. The top picture shows a smooth material transition because the crack travels at a relatively constant angle through the foam. The jagged edges of the crack could possibly represent a test specimen that did not receive a good load transfer from the Instron machine.

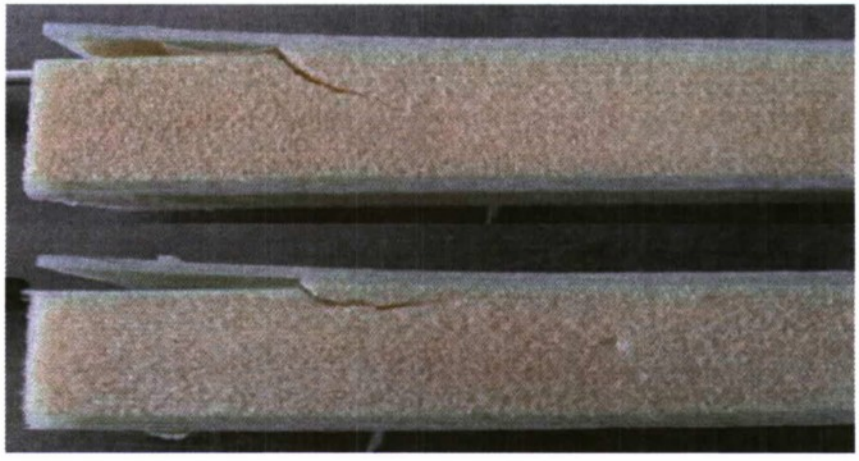


Figure 61 - Monotonic failure of test specimens with 1.0 inch initial delamination

Shear Key Delamination from Foam

Figure 62 - Monotonic failure shear key delamination shows a side view of the shear key delaminating from the foam. When the deflection of the laminate got too high, it released the shear key free from the foam and the laminate continued to separate from the foam. When this type of failure was first discovered, the validity of the shear key monotonic test data was called into question. Since the shear keys were delaminated from the foam it appeared that there was not a good bond between the foam and the shear keys. It was later determined that there may have been too much wax applied to the shear key mold and the shear keys didn't have enough surface roughness to achieve a good bond to the foam. Future shear key tests implemented these changes, increasing the strength of the test specimens containing shear keys.

Figure 62 - Monotonic failure shear key delamination

Shear key crack propagating through foam

The majority of the test specimens that produced good data failed due to cracks propagating through the foam. Figure 63 - Monotonic shear key failure due to cracks in foam shows a shear key specimen that failed due to a propagating crack. This mode of failure was similar to the failure of the delaminated test specimens with no

shear key. The crack originated at the edge of the delamination and then travels through the foam at a slanted angle.

Figure 63 - Monotonic shear key failure due to cracks in foam

0.5 inch Shear Key

The failure models for the sandwich structures with a shear key 0.5 placed inch behind the delamination edge experienced modes of failure similar to the 0 inch shear keys. The primary mode of failure was the origination of a crack at the edge of the delaminated region that travels through the foam. Figure 64 - Monotonic shear key failure (0.5 inch shear key) shows the failure caused by cracks in the foam. The image on the right side of Figure 64 - Monotonic shear key failure (0.5 inch shear key) shows a crack traveling at approximately a 45° angle which continues through the entire thickness of the foam. This type of crack propagation was the most frequently encountered failure. A secondary type of failure through the foam is shown in the left hand side of Figure 64 - Monotonic shear key failure (0.5 inch shear key). The crack initially travels at an angle less than 30° and then around the shear key as opposed to continuing through the thickness of the foam. This mode of failure was expected with the introduction of shear keys into the sandwich structure test specimens. It was predicted that the shear key would act as a stronger material than the foam, which would drive the crack around the shear key.

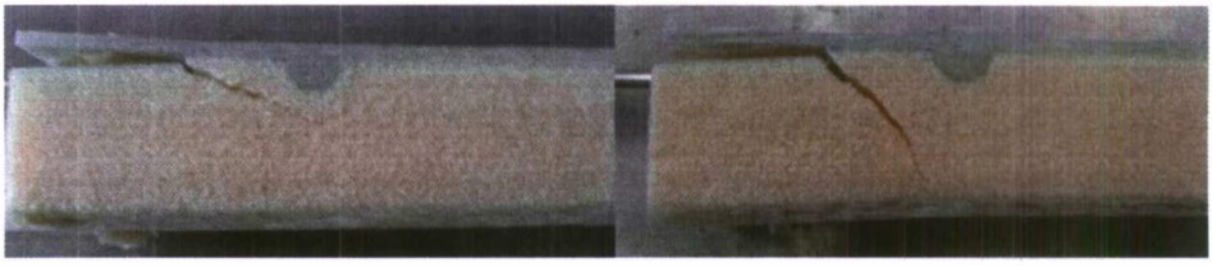


Figure 64 - Monotonic shear key failure (0.5 inch shear key)

Monotonic Failure Analysis – 1.0 inch Shear Key

The test specimens with shear keys placed 1.0 inch behind the delamination failed in a similar manner to the 0 and 0.5 inch shear keys. The most common failures occurred in the foam and some failures occurred in

the matrix. Figure 65 - Monotonic shear key failure through matrix (1.0 inch shear key) shows a failure caused by cracks originating in the matrix. Cracks in the matrix caused the delamination to propagate along the upper edge, making the upper composite layer separate from the foam core. The shear key had very little effect on these matrix failures because it was located too far away from the failure region.

Figure 65 - Monotonic shear key failure through matrix (1.0 inch shear key)

The more common failure for the 1.0 inch shear keys was failure due to cracks in the foam and is shown in Figure 66 - Monotonic shear key failure cracks in foam (1.0 inch shear key) below. The failure for these test specimens was slightly different than the other shear key specimens that failed through the foam. Here the crack starts at a shallow angle of approximately 10-20° and goes down through about 1/10 inch through the foam. Then the crack turns slightly and travels at approximately 180° towards the shear key. Figure 66 - Monotonic shear key failure cracks in foam (1.0 inch shear key) shows two failures caused by shallow cracks in the foam.

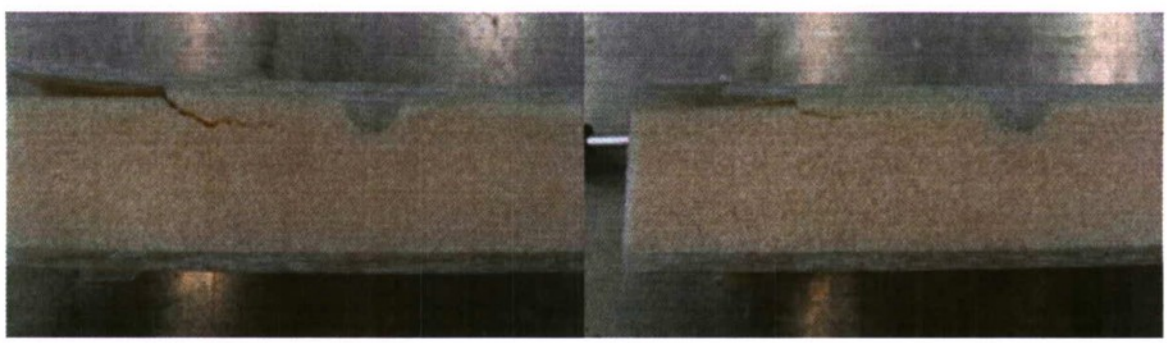


Figure 66 - Monotonic shear key failure cracks in foam (1.0 inch shear key)

The less common type of foam failure in the 1.0 inch shear keys is shown in Figure 67 - Monotonic shear key failure type II (1.0 inch shear key). The failure starts out the same way as the test specimens depicted in Figure 66 - Monotonic shear key failure cracks in foam (1.0 inch shear key) except the crack travels a different path. The crack travels back up through the foam towards the composite layer and stops at the edge of the shear key. These results seemed somewhat counter intuitive because the intent of the shear keys was to drive the crack away from the upper skin to help stop the delamination.

Figure 67 - Monotonic shear key failure type II (1.0 inch shear key)

4.1-2 Fatigue Testing (Eugene)

1.0 inch Initial Delamination Failure Analysis

The fatigue failures for the test specimens with a 1.0 inch delamination failed very similarly to the specimens tested under monotonic loading with failure. The majority of the test specimens failed due to cracks originating in the foam. A few test specimens failed due to cracks in the matrix. The main difference between the two types of failures was the speed at which the cracks propagated.

Figure 68 - 1.0 inch delamination fatigue crack propagation through foam and Figure 69 - 1.0 inch delamination fatigue crack propagation type II through foam show the crack propagation of one of the test specimens being subjected to an input of 50% of the —— . A 50% load case was selected because it would provide the best representation of the crack propagation because the loading was low, and the crack should move relatively slowly. The left hand side of Figure 68 - 1.0 inch delamination fatigue crack propagation through foam shows the test specimen after 307 cycles, the point at which it was noted the applied force started to drop from the input force. This signified that a microscopic crack had begun to develop even though it was not visible. The right hand side of Figure 69 - 1.0 inch delamination fatigue crack propagation type II through foam shows the same test specimen after 999 cycles where a small crack is clearly visible near the edge of the delaminated region. The angle that the crack begins to travel was similar to the 1.0 inch delaminated test specimens tested monotonically.

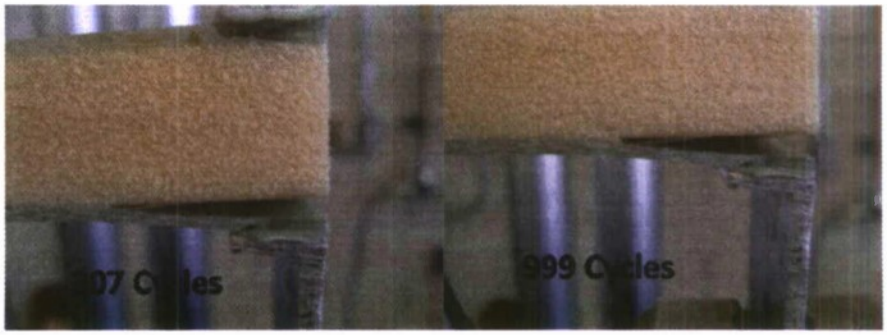


Figure 68 - 1.0 inch delamination fatigue crack propagation through foam

The left side of Figure 69 - 1.0 inch delamination fatigue crack propagation type II through foam shows that after 1385 cycles the crack continued to travel through the foam at approximately the same angle that it began. The right side of Figure 69 - 1.0 inch delamination fatigue crack propagation type II through foam shows the test specimen after 1963 cycles, which was the point the 10% load drop failure criteria was met. It is uncertain whether or not the crack would continue to travel through the foam at the same angle. In most test the cracks appeared to progress down towards the lower fiberglass skin. This type of failure was similar to the results obtained from the test specimens with shear keys placed 1.0 inch behind the back edge of the delamination.

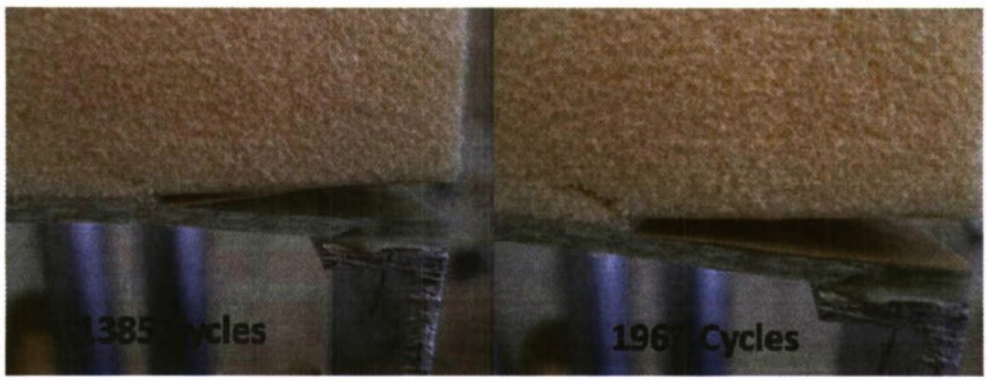


Figure 69 - 1.0 inch delamination fatigue crack propagation type II through foam

The other type of fatigue failure observed for the test specimens with an initial delamination of 1.0 inch was cracks in the matrix, characterized by a delamination that caused the fiberglass skin to separate from the foam core. Figure 70 - 1.0 inch delamination fatigue crack propagation through matrix shows the fatigue crack propagation when failure was caused by cracks in the matrix. The test specimen here was subjected to a loading rate of 55% of the —— and an easily visible crack was expected, but that did not turn out to be the case. The left side of Figure 70 - 1.0 inch delamination fatigue crack propagation through matrix shows that the skin started to peel away from the foam after only 123 cycles. The same test specimen quickly reached the failure criteria after only 381 cycles. Based on the data obtained from the 50% loading rate it was determined that this was a premature failure because the crack should have initiated in the foam because the epoxy is a stronger material than the foam. The stiffness of the epoxy is 5 times greater than that of the foam. For the test specimens with an initial delamination of 1.0 inch fatigue failure caused by a crack in the matrix signified a premature failure that was attributed to defects introduced during the manufacturing process.

Figure 70 - 1.0 inch delamination fatigue crack propagation through matrix

1.0 inch Shear Key Failure Analysis

All of the fatigue failures for the shear key test specimens were caused by cracks in the matrix Figure 71 - Shear key fatigue failure analysis shows a shear key test specimen subjected to 65% loading with a crack propagating along the upper fiberglass skin. The image on the left of Figure 71 - Shear key fatigue failure analysis shows the initiation of the crack and the image on the right of Figure 71 - Shear key fatigue failure analysis shows the final failure.

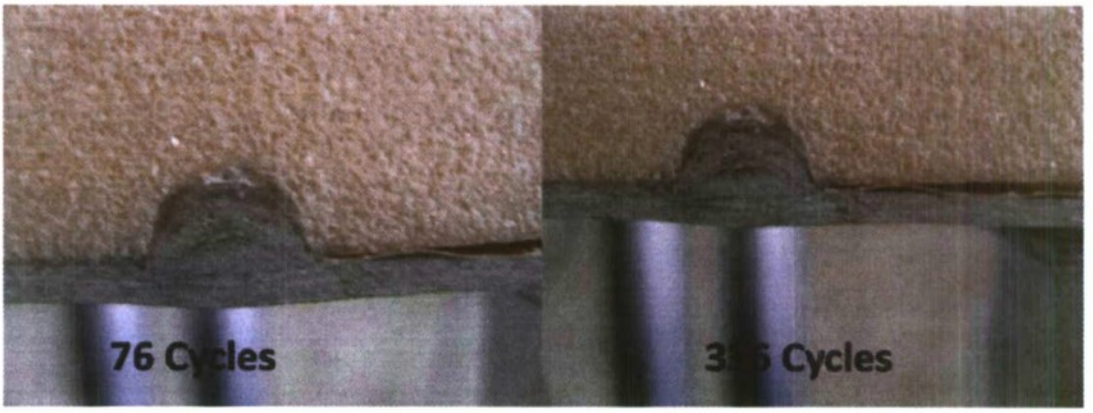


Figure 71 - Shear key fatigue failure analysis

4.1-3 Monotonic Results Comparison

Figure 72 - Monotonic experimental results comparison (delamination) shows the monotonic force vs. displacement results for the test specimens with an initial delamination of 0.5 and 1.0 inch. Although both delamination lengths reach a similar ultimate —— of approximately 100 — the two curves are very different. The 0.5 inch delamination test specimen only reached a maximum displacement of approximately 0.04 inches where as the 1.0 inch delamination specimens reached a maximum displacement of approximately 0.08 inches.

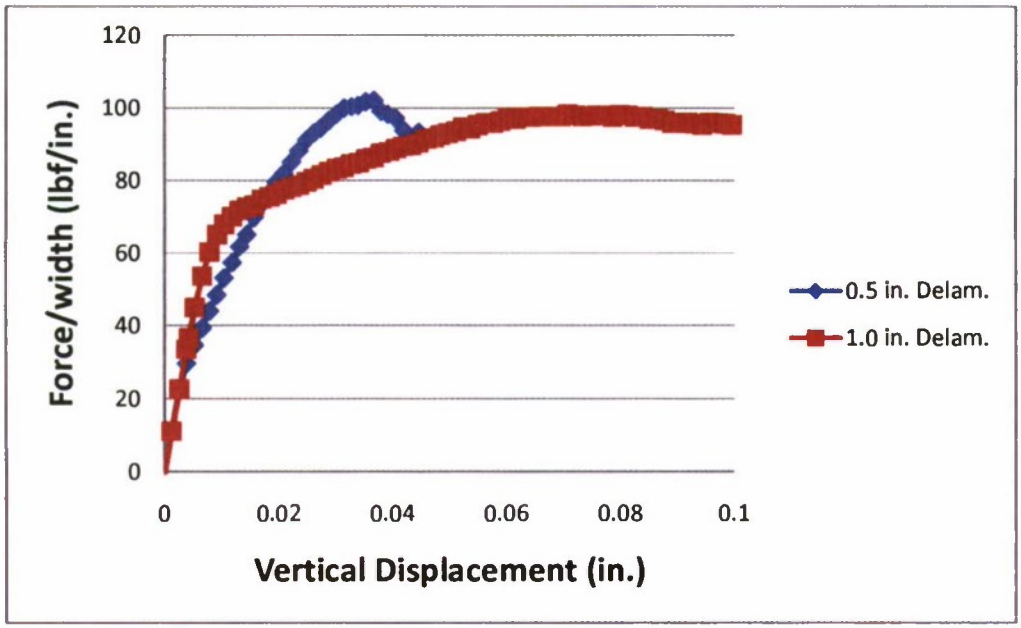


Figure 72 - Monotonic experimental results comparison (delamination)

The large differences can be attributed to the L-shaped aluminum brackets used to apply the load to the specimen from the Instron machine. The brackets had leg lengths of 0.5 inches that was bonded to the upper skin of the sandwich structures. The combination of the 0.5 inch delamination and aluminum bracket create almost a purely vertical pull-up force. The 1.0 inch delamination test specimens are not purely a vertical force because half of the delamination length is not bonded to the aluminum tab. At the free delamination area a moment is created there which causes the 1.0 inch delamination test specimens to require more force to initiate the onset of plastic deformation. This can be seen in Figure 73 - Experimental monotonic results (shear keys and delamination) because the red curve has a much larger elastic region (up to approximately 70 $\frac{lb_f}{in.}$) than the 0.5 inch delamination pieces which only remains elastic up to approximately 20 $\frac{lb_f}{in.}$. For the 1.0 inch delamination test specimen to deform plastically a larger force is needed because the applied forces not only created a pull-up force, but also a moment due to the delamination length that was not directly under the aluminum tab.

Figure 73 - Experimental monotonic results (shear keys and delamination) shows the monotonic test results for the three different shear key configurations as well as the 1.0 delamination case. All four configurations had very similar elastic and plastic deformation. The shear key configurations were characterized by an elastic response up to approximately 55 $\frac{lb_f}{in.}$ and afterwards followed by plastic deformation until the ultimate force was reached. The 0 inch shear key reached the highest ultimate force/width value of approximately 100 $\frac{lb_f}{in.}$ followed by the 0.5 inch shear key at 98 $\frac{lb_f}{in.}$ and the 1.0 inch shear key at 95 $\frac{lb_f}{in.}$. The

1.0 inch delamination test specimen had a slightly higher ultimate stiffness of approximately 102 $\frac{lb_f}{in.}$ and similar elastic response. Figure 73 - Experimental monotonic results (shear keys and delamination) shows the shear keys had virtually no positive effect for increasing the monotonic failure strength of initially delaminated test specimens which is not entirely unexpected.

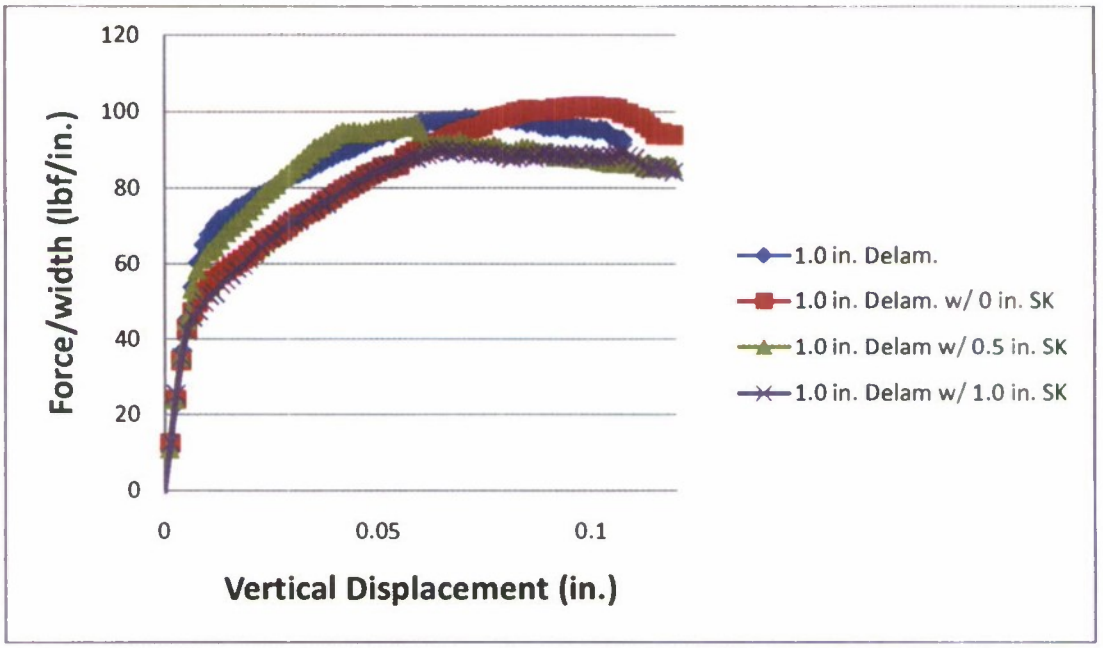


Figure 73 - Experimental monotonic results (shear keys and delamination)

Adding shear keys to initially delaminated test specimens seemed to only add a stress concentration to the test specimens which does nothing to increase the strength of the parts. The stress concentrations do not have a major effect on the test specimen with a 0 inch shear key but as the shear keys get moved further away from the back edge of the delamination the test specimens slowly get weaker. A possible explanation for this is that shear keys are simply additional defects added to the structure. Just as an initial delamination causes the test specimens to fail much sooner than a part without an initial delamination the shear keys have the same effect. By placing the shear key at the back edge of the delamination it essentially combined the two defects together into one which was why the 0 inch shear key specimens were only slightly weaker than the test specimen with only an initial delamination. As the shear keys moved away from the edge of the delamination this created another defect in the part. This explains why the 0.5 and 1.0 inch shear key specimens were weaker than both the 0 inch shear key and the test specimen with only an initial delamination.

4.1-4 Data Summary

Table 4 is a summary of the three sets of monotonic test data values obtained from the sandwich structures without shear keys. All three test sets have relatively similar elastic stiffness values, with the exception of the 1 inch delaminated specimens being a little less. The yield stiffness value of the 0.5 inch initial

delamination specimens was significantly less compared to the other two sandwich structure types. One would expect a smaller delamination would correspond to a stronger part, with a higher ultimate stiffness, but that was not the case for the yield stiffness. A possible explanation for the weaker yield stiffness for the 0.5 inch delaminated test specimens could be that the 0.5 specimens had a different load path than the 1 inch delaminated specimens. The aluminum tabs used to transfer the load from the Instron machine to the test specimens had a 0.5 inch leg length that got bonded to the upper and lower surface of the specimens. For the 1 inch delaminated pieces the load was not being applied directly axially because the extra 0.5 inch that was not connected to the aluminum tab created a moment. Adding a moment makes the loading conditions no longer purely axial, causing the loading conditions to be mixed modes (Modes I and II). Both the 0.5 inch and non-delaminated pieces had the load applied almost entirely axially applied (Mode I). The different loading conditions could also account for the slight differences in the elastic stiffness.

Table 4 - Monotonic test data summary

Test Type	Elastic Stiffness $\left(\frac{lb_f}{in.}\right)$	Yield Stiffness $\left(\frac{lb_f}{in.}\right)$	Ult. Stiffness $\left(\frac{lb_f}{in.}\right)$
No Initial Delamination	8676	310.35	396.02
0.5 in. Delamination	8777	22.69	102.21
1 in.Delamination	7832	60.32	98.34

4.2 Delamination Arrest Mechanism Under Buckling(Tony)

4.2-1 Control Groups

The control groups, as discussed, are used to provide comparable results between control groups and their experimental group counterparts. The groups include investigating where there is no delamination or an embedded key, test specimens with initial delamination only, and test specimens with an embedded key only.

4.2-2 Control Group: No Delamination and No Embedded Arrestment Key

The '11 inches' configuration was retrieved from the buckling length research discussed earlier and is displayed in Table 5.

Table 5 - '11 inches' Configuration

Case	Slenderness Ratio	Max Vertical Extension	Ultimate Load	Max Horizontal Extension
-	-	in	lbf	in
11 inches	65.39673	0.11083	685.60279	0.42417

From the Figure 74 - '11 inches' – Load versus Vertical Extension the graph shows a gradual climb in load and then begins to round off at the top of the load versus vertical extension until the test specimen experiences a failure. This configuration is built where no additional components in the structure are so that the bonding between the skin and cores is not disturbed.

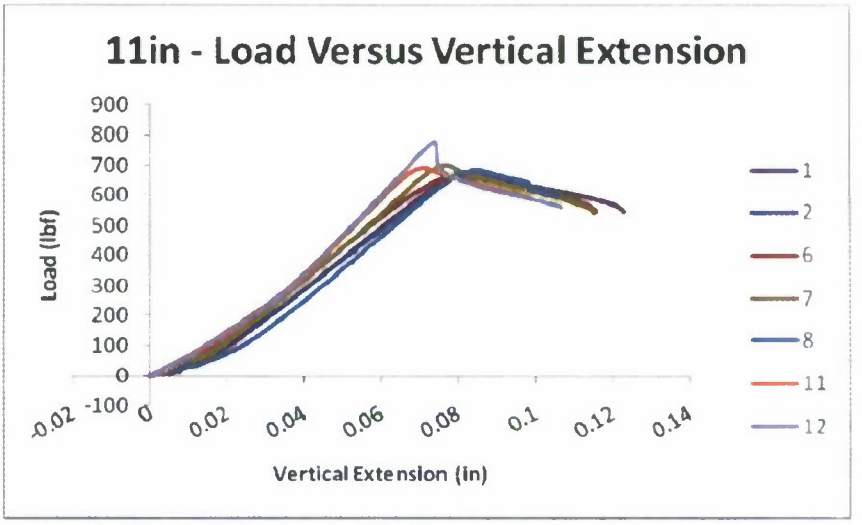


Figure 74 - '11 inches' – Load versus Vertical Extension

Figure 74 - '11 inches' – Load versus Vertical Extension is quite similar to Figure 75 - '11 inches' – Load versus Vertical Extension because the vertical displacement transitioned a displacement in the horizontal direction. This transition is due to the pinned boundaries on the test specimen as it rotates to create a deflection with the maximum extension located at the center. It starts off with a slow increase in horizontal displacement per given loading then begins to exponentially displace.

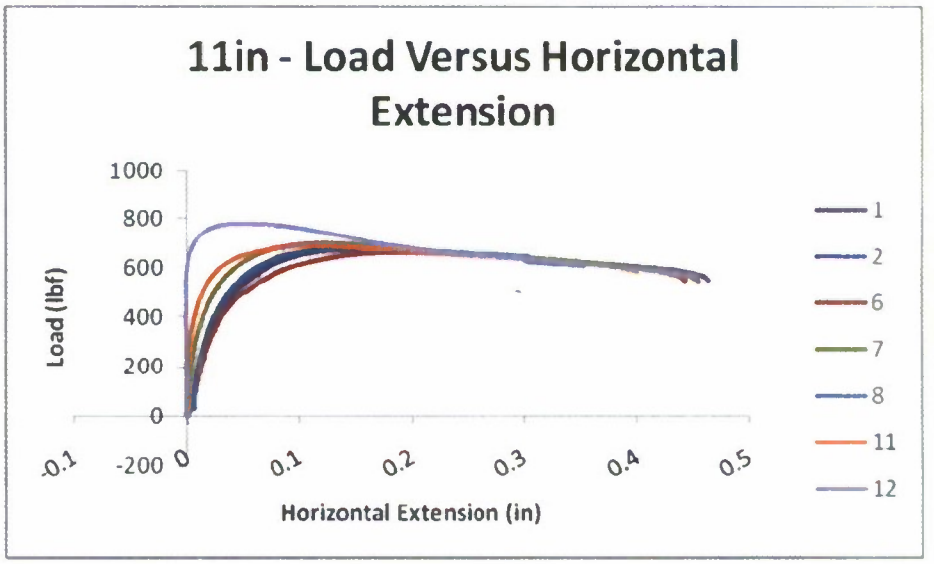


Figure 75 - '11 inches' – Load versus Vertical Extension

With the combination of the horizontal extension data and the vertical extension data seen in Figure 76 - '11 inches' – Load versus Vertical Extension the graph shows an increased horizontal displacement when the load is applied. This means that the test specimens began to rotate around the pinned points and buckle.

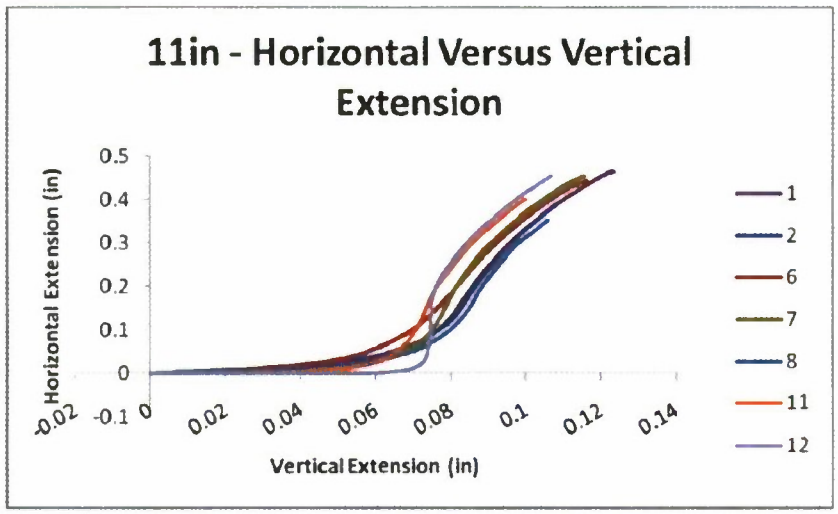


Figure 76 - '11 inches' – Load versus Vertical Extension

4.2-3 Control Group: Initial Delamination with No Embedded Arrestment Key

The second control group that was studied was test specimens that were introduced with a delamination before they were tested. The introduction of the initial delamination is done with non-porous material that is thin and inserted before the skin and the core materials. This non-porous material does not change the material properties of the composite sandwich structure but only create the delamination used in this section. The delamination region extends the width of the structure and varies in height between half-inch to 2 inches in length per half-inch increment. For this group, the delamination length and the term 'ID' denote the different cases after; an example of this is '1.0 ID'. The resultant data is shown in Table 6:

Table 6 - 'ID' Initial Delamination Configuration

Case #	Case	Slenderness Ratio	Max Vertical Extension in	Ultimate Load lbf	Max Horizontal Extension in
0	11 inches	65.39673	0.11083	685.60279	0.42417
1	0.5 ID	66.06927	0.05020	479.36398	0.02274
2	1.0 ID	65.69982	0.03675	267.32638	0.00986
3	1.5 ID	66.21160	0.02963	158.26181	0.02399
4	2.0 ID	65.89214	0.04118	132.60860	0.04519

From the table, it shows that the delamination in the composite sandwich structure dramatically decreases the loading capability of the test specimens. The loading capability drops approximately 35% for the

half an inch delamination, denoted by 0.5 ID, and continues to drop per each increase of the delamination's area size. This is why most structures will require repairs before the delamination can affect the structural integrity and needs frequent inspection. Otherwise, the structure can easily become an expensive piece of scrap.

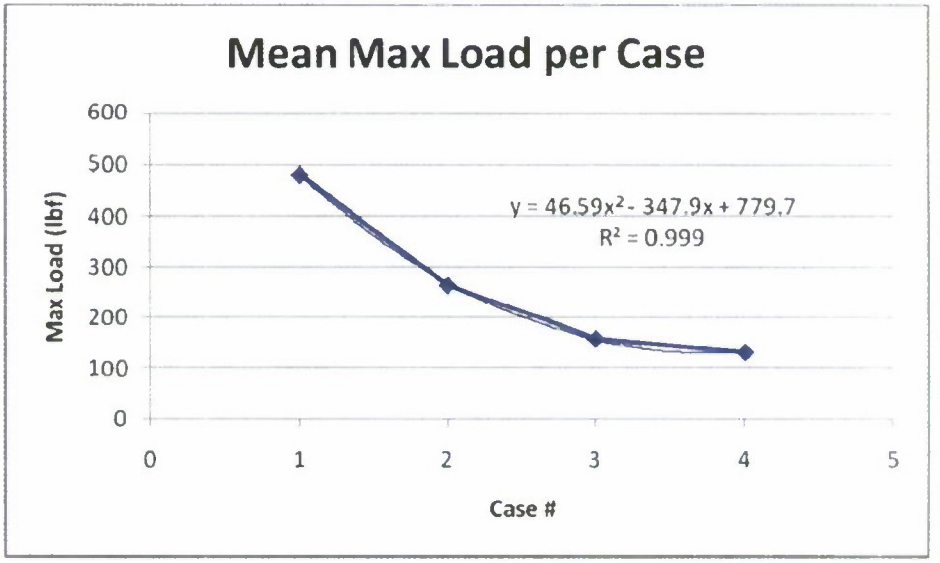


Figure 77 - Mean Ultimate Load per Case

Using the data from Table 6, the graph shows decrease in the load capacity as the delamination region increases. This is a typical trend for most composite metals. One thing to note about Figure HJ is that it fits a parabolic trend line and seems to become asymptotic as the size of the delamination increases. As for Figure 77 - Mean Ultimate Load per Case, it shows a similar trend as Figure 78 - Mean Ultimate Stress per Case but the difference is the geometry factor.

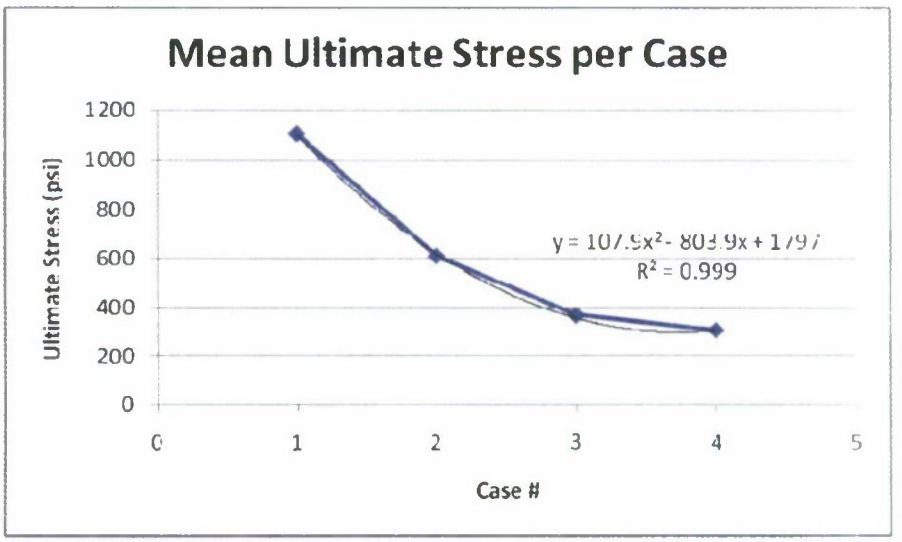


Figure 78 - Mean Ultimate Stress per Case

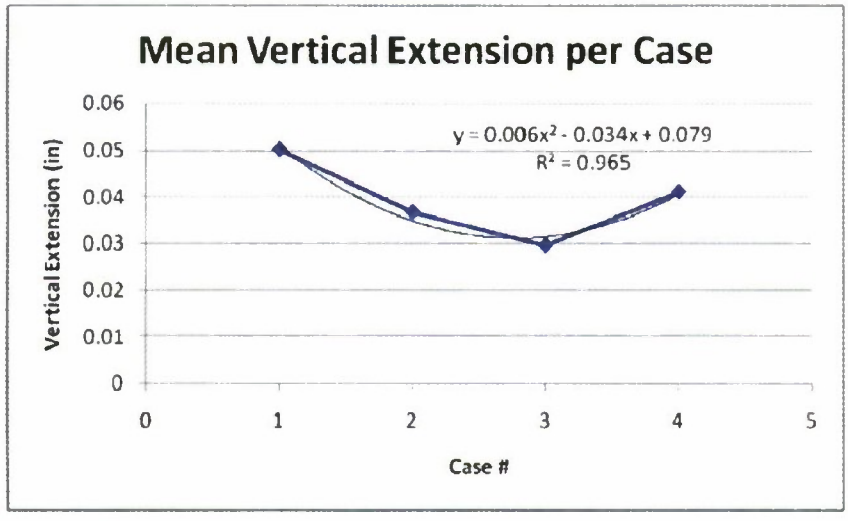


Figure 79 - Mean of Maximum Vertical Extension per Case

The maximum vertical extension per each of the initial delamination cases is shown in Figure 79 - Mean of Maximum Vertical Extension per Case. The data shows that the extension is greatest with the 0.5- inch delamination and bottoms out at the 1.5-inch delamination test specimen. From this parabolic trend line, the increase in the size of the delamination means a decrease in vertical extension until the curve bottoms out and becomes asymptotic. The reason for the decrease in vertical extension with the increase in delamination size is because the composite sandwich structure fails much earlier. The 2.0-inch case was removed from the graph because the vertical extension during the data acquisition was not properly set up.

As the vertical extension decreases, the horizontal extension increases due to the structural incapacity to resist the flexure as the structure is loaded. Figure 80 - Mean of Maximum Vertical Extension per Case details the results of the horizontal extension per each case. The case regarding the 0.5-inch delamination was removed from the graph due to its high standard deviation. This high standard deviation can be attributed to the data acquisition setup process.

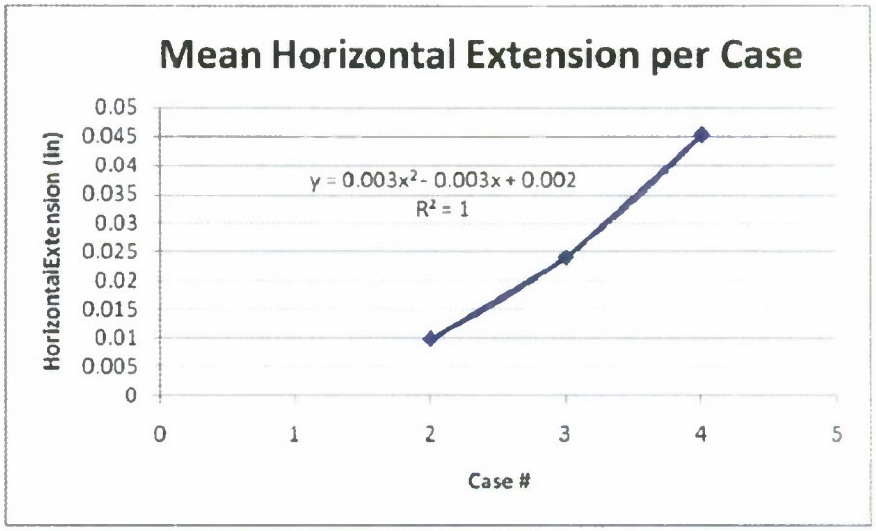


Figure 80 - Mean of Maximum Vertical Extension per Case

Note that the following subsections show individual cases with the various delamination sizes. The test specimens in these subsections have the same geometry and same manufactured process except for the delamination sizes. The buckling of these test specimens is only measured on the side with the delamination region. The issue with the buckling of the structure and face sheet is that it requires two different extensometers to measure the extensions on both sides of the test specimens. With the limitations in resources, one extensometer was used and placed on the delamination region side of the test specimens.

4.2-4 Half-Inch Initial Delamination

The first delamination set of data comes from the test specimens that have an initial delamination that extends the width dimension and the 0.5-inch length at the center of the test specimen. The test specimens remain elastic in nature until they fracture. This is detailed in Figure 81 - 0.5 ID – Load Vs Vertical Extension where the test specimens are introduced to in-plane loading and remain in the elastic region. The main reason for the brittleness of the structure comes from the materials chosen. The foam core is quite brittle in nature and is reinforced with fiberglass skin.

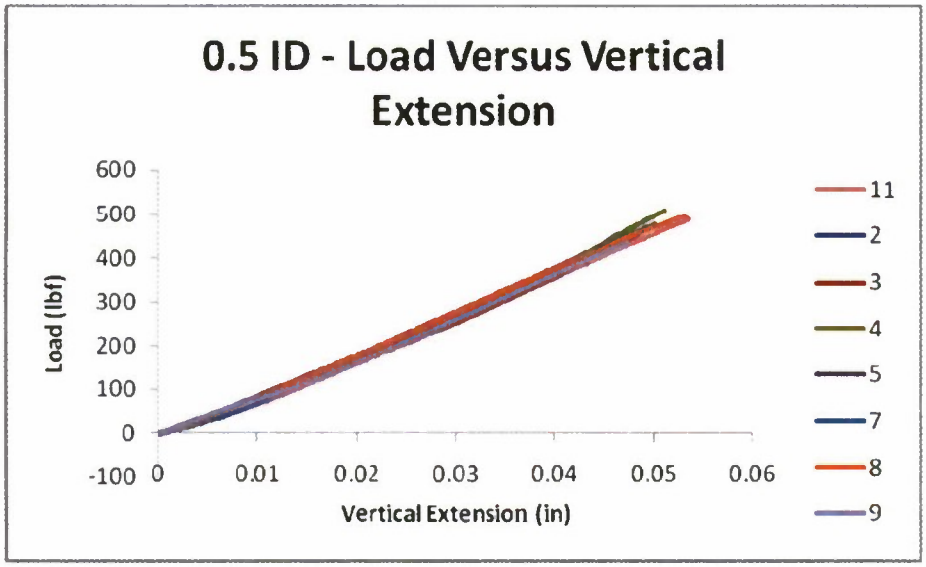


Figure 81 - 0.5 ID – Load Vs Vertical Extension

Figure 82 - 0.5 ID – Load Vs Horizontal Extension shows the test specimens extending horizontally and the buckling of the face sheet before it fractures. This reflects how well the test specimens can hold the load conditions. The smaller the delamination region, the higher loading capacity of the test specimen becomes. This is due to less localized stresses on the part. Figure 82 - 0.5 ID – Load Vs Horizontal Extension looks odd due to the limitation of the extensometer causing the higher fluctuation.

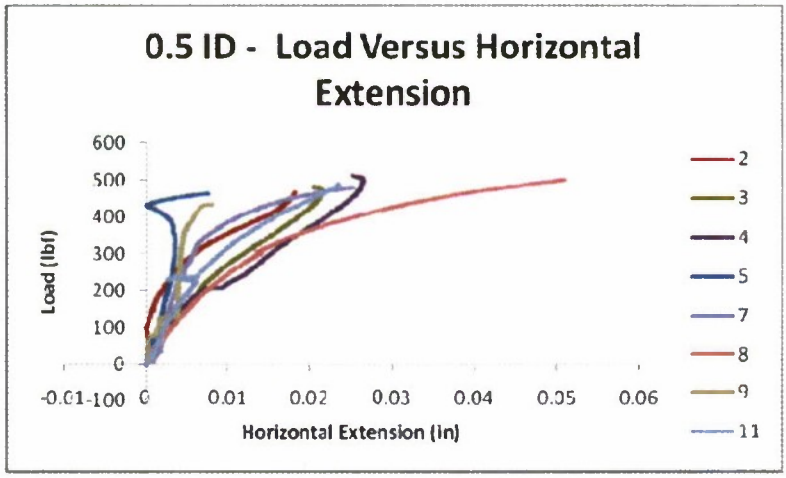


Figure 82 - 0.5 ID – Load Vs Horizontal Extension

Figure 83 - 0.5 ID – Horizontal Extension Vs Vertical Extension represents the relationship between the horizontal extension and vertical extension and shows a progressive increase in horizontal extension as the vertical extension is being displaced. There is much more vertical displacement than horizontal displacement hence the low profile curves in the Figure 83 - 0.5 ID – Horizontal Extension Vs Vertical Extension. This is seen as the little spike at the end of the linear plot.

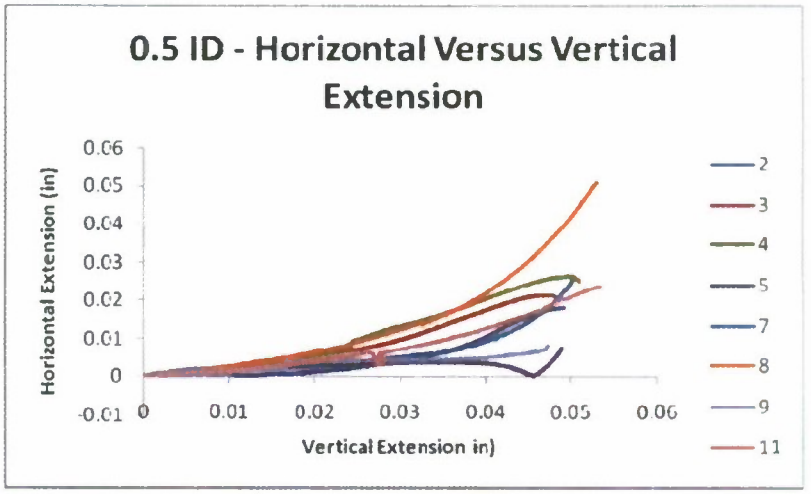


Figure 83 - 0.5 ID – Horizontal Extension Vs Vertical Extension

Figure 84 - 0.5 ID – Experimental Testing Before and After Failure

What can be seen from Figure 84 - 0.5 ID – Experimental Testing Before and After Failure was that the experimental testing before failure shows very little horizontal extension and as the strain energy releases, the crack propagates very quicker through the test specimen as seen on the right image.

4.2-5 1-Inch Initial Delamination

The second data set comes from the 1.0-inch delamination test specimens that were tested under in-plane loading. What differs in Figure 85 - 1.0 ID – Load Vs Vertical Extension in comparison to the 0.5-inch delamination is that the test specimens fail a bit earlier thus the reason for the shorter vertical extension. The test specimens still exhibit the same elastic properties, but begin to transition near the end of life before they fracture as seen in Figure 83 - 0.5 ID – Horizontal Extension Vs Vertical Extension.

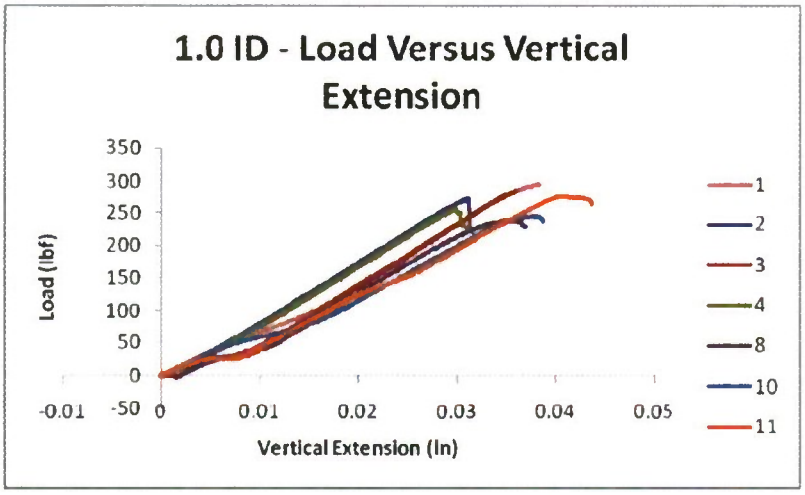


Figure 85 - 1.0 ID – Load Vs Vertical Extension

In Figure 86 - 1.0 ID – Load Vs Horizontal Extension, it shows a trend that deviates and transitions from the vertical extension where the test specimens would see a slight buckling occurrence of the face sheet when the loading is closer to the ultimate. As discussed earlier in '0.5 ID' configuration section, the limitation of the extensometer dramatically affects the deviation of the horizontal displacement. What can be extract from the graph is that the test specimens require less loading when increased in delamination length.

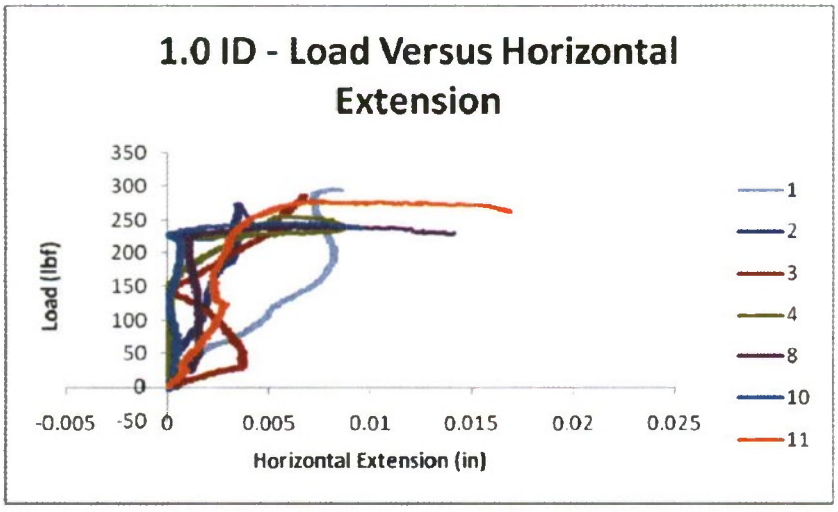


Figure 86 - 1.0 ID – Load Vs Horizontal Extension

From the combination of the vertical extension with the horizontal extension data, it does show that the skin begins to buckle but the structure would fail before the initial delamination could do more damage. This is shown in Figure 87 - 1.0 ID – Horizontal Extension Vs Vertical Extension.

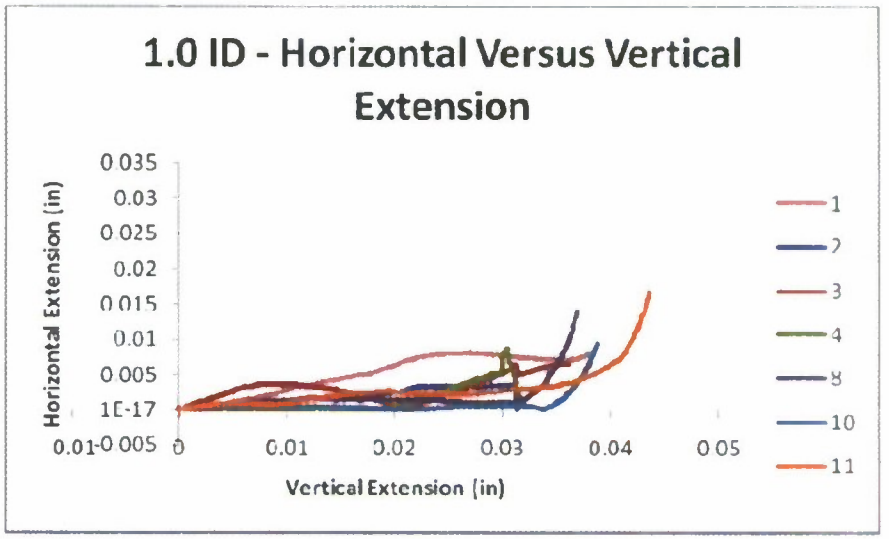


Figure 87 - 1.0 ID – Horizontal Extension Vs Vertical Extension

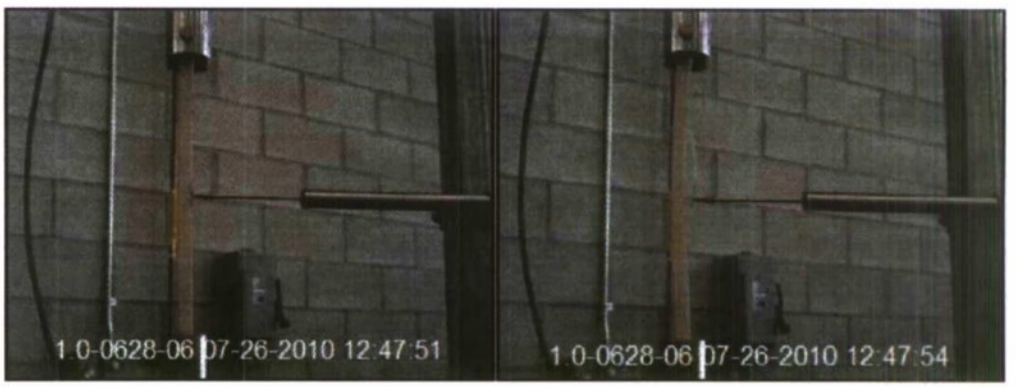


Figure 88 - 1.0 ID – Experimental Testing Before and After Failure

Similar to what was seen for the '0.5 ID' configuration, the '1.0 ID' configuration differed by having a slightly larger extension. The strain energy also propagated the crack and caused the failure of the test specimen. This can be seen in Figure 88 - 1.0 ID – Experimental Testing Before and After Failure.

4.2-6 1.5-Inch Delamination

As for the 1.5-inch delamination test specimens, the delamination of these test specimens is buckling more than the previous two delamination cases. The buckling occurred because the delamination region is much larger and the edge of the delamination region, or the crack, propagates between the skin and core under in-plane loading. The Instron displacement machine will stop the test when the test specimens exhibit a 40% load drops from the ultimate loading. This is why Figure 89 - 1.5 ID – Load Vs Vertical Extension shows a failure point at the largest vertical extension. Also, another thing to note is how the load versus vertical extension is curved. The curvatures show a displacement that requires less loading, which is exhibiting instability in the composite sandwich structure.

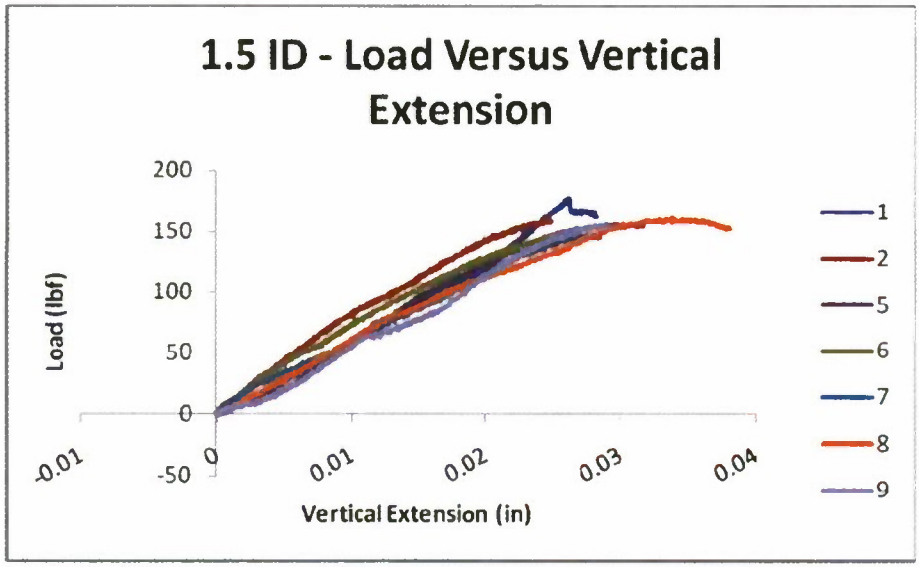


Figure 89 - 1.5 ID – Load Vs Vertical Extension

As seen in Figure 90 - 1.5 ID – Load Vs Horizontal Extension, the load versus horizontal extension transitioned from Figure 89 - 1.5 ID – Load Vs Vertical Extension. The vertical extension is transitioned to the horizontal direction where it slows displacement in the horizontal direction with the increase in loading. The curvature in each of the curves details the test specimen's buckling of the whole structure in one direction. Then the hook in the curves details the horizontal extension of the face sheet in the opposite direction.

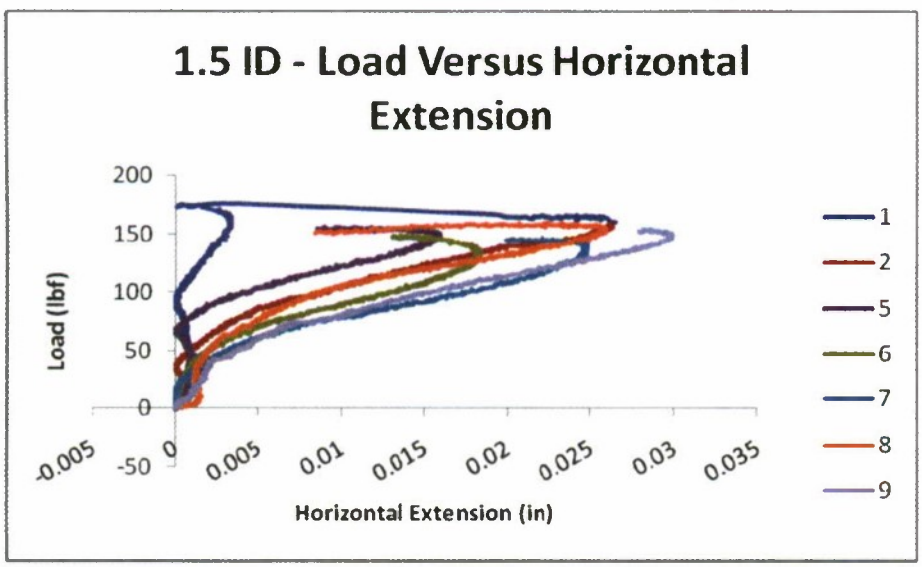


Figure 90 - 1.5 ID – Load Vs Horizontal Extension

The combination of the two extensions, which are the vertical and horizontal extensions, shows a great amount of horizontal extension as the vertical extension increases. Same as what has been discussed; the face sheet delamination buckles more with the increase size of the delamination. This Figure 91 - 1.5 ID – Load Vs Horizontal Extension below details that quite well.

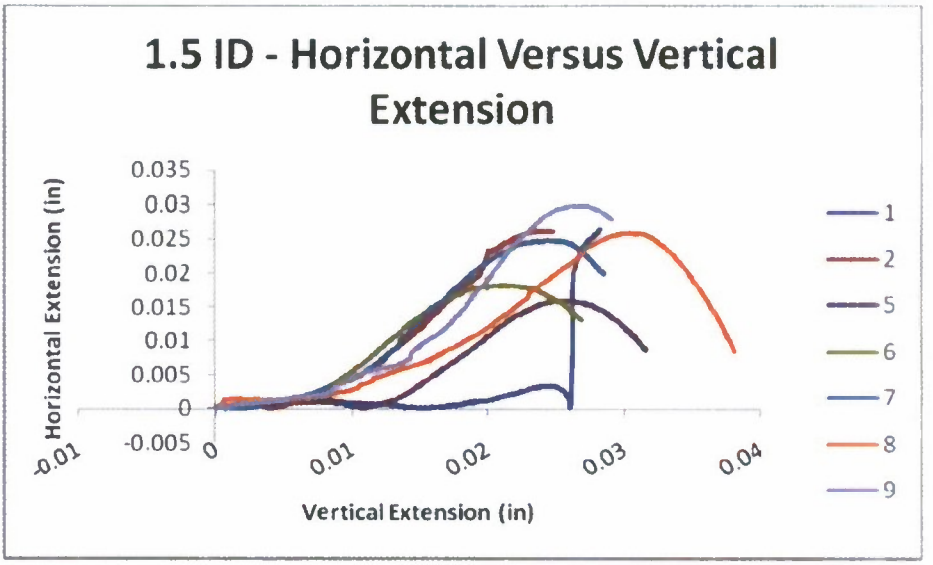


Figure 91 - 1.5 ID – Load Vs Horizontal Extension

Figure 92 - 1.5 ID – Experimental Testing Before and After Failure

As the initial delamination size starts to increase, the face sheet delamination becomes more visible and can be seen in Figure 92 - 1.5 ID – Experimental Testing Before and After Failure on the left. The right image shows the failure of the test specimen and because of the larger delamination size. There is less strain energy causing the crack propagation to not spike as high along the seam between the foam and fiberglass skin.

4.2-7 2-Inch Delamination

Similar to all the other subsections discussing the transition of the composite sandwich structures to the increased experience of a buckling feature; Figure 93 - 2.0 ID – Load Vs Vertical Extension the 2-inch delamination region experiences this most and much earlier due to the delamination size and the ease of propagation.

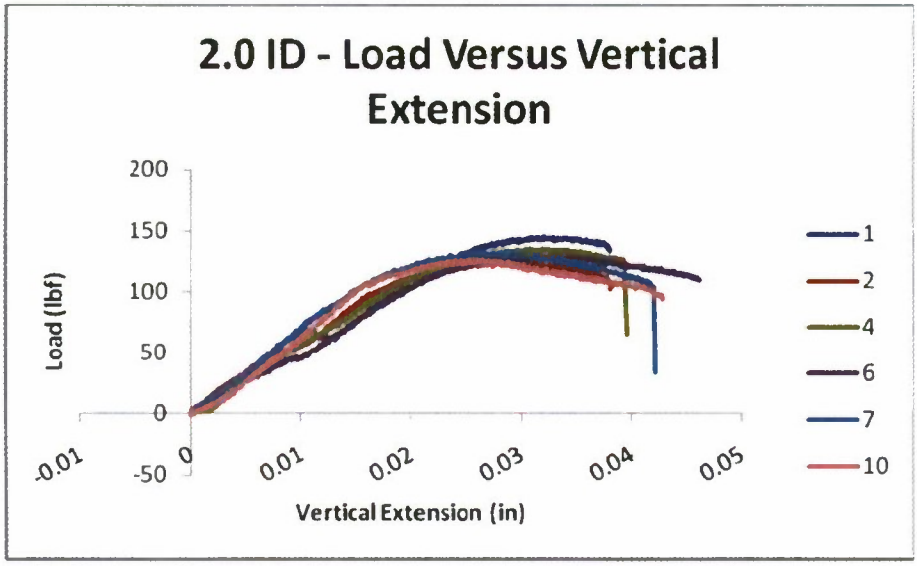


Figure 93 - 2.0 ID – Load Vs Vertical Extension

Similar to what was discussed for '1.0 ID' configuration regarding the load versus the horizontal extension, this stands true for the '2.0 ID' configuration as well. The only difference shown in Figure 94 - 2.0 ID – Load Vs Horizontal Extension is the increase how much the face sheet is displaced as the vertical displacement is applied. These two graphs show a lot of similarities where the structure buckles in one direction as the face sheet buckles in the other direction.

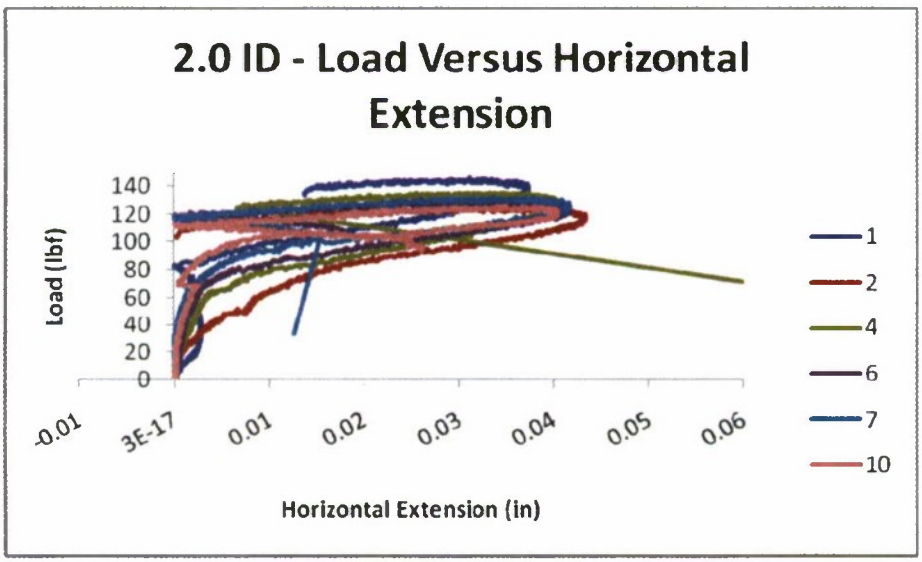


Figure 94 - 2.0 ID – Load Vs Horizontal Extension

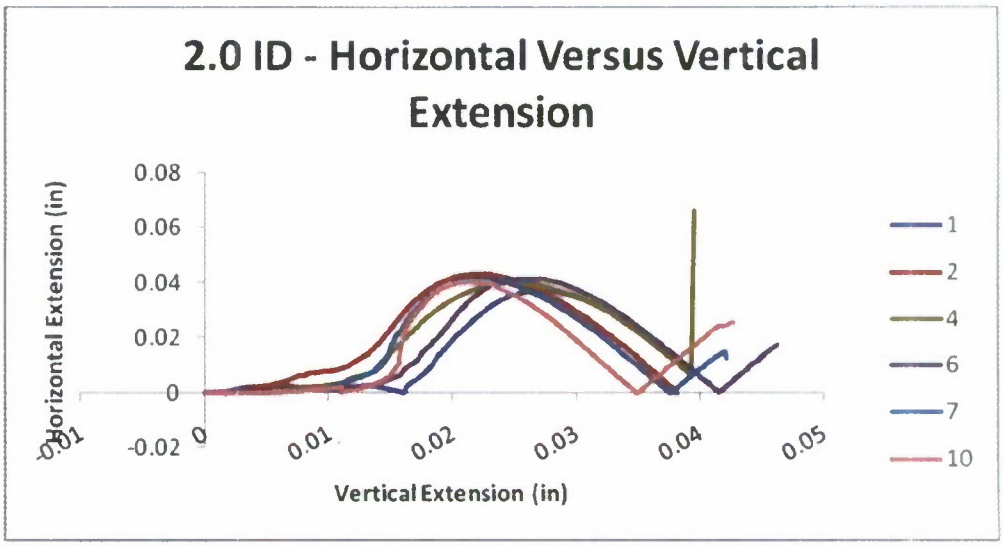


Figure 95 - 2.0 ID – Horizontal Extension Vs Vertical Extension

In Figure 95 - 2.0 ID – Horizontal Extension Vs Vertical Extension, the trend is a bit different because the structure buckles much more before the face sheet begins to buckle. This is seen in the graph where the horizontal begins to climb then recedes as the vertical extension increases. The spikes at the end of the curves in Figure 95 - 2.0 ID – Horizontal Extension Vs Vertical Extension are the points of failure.

Figure 96 - 2.0 ID – Experimental Testing Before and After Failure

Figure 96 - 2.0 ID – Experimental Testing Before and After Failure shows a longer horizontal extension because of the initial delamination size placed under loading. This later transitioned to a small burst of energy causing the edge of the initial delamination to propagate a bit more. The larger delamination size also shows a larger decrease in loading capacity.

4.3 Damage Arrestment Devices (Richard B.)

All experimental testing and results would be discussed in this chapter in detail. The mechanical characteristics of the material were tested, such as the volume fraction of the face sheets and the material characteristics of the core, the face sheet, and the overall composite sandwich panels. The effects and trends of

changing the DADs thickness on the failure characteristics of the composite sandwich panel would be discussed. Lastly the fatigue characteristics and trends of the DADs on the composite sandwich panel.

4.3-1 Mechanical Characteristics of Material/Specimen

Volume Fraction of Face Sheets

The volume fraction of the face sheets with and without a film adhesive was tested and analyzed. The specimen procedure and manufacturing was discussed. The results from the burn test for both test groups can be seen in Table 7. Six specimens were tested for each group and the results were averaged for further analysis. From the results, it can be seen that adding the film adhesive increases the resin content of the laminate by 6% from 54.8% to 61.1% in weight of the specimens.

Table 7 - Results from Volume Fraction Testing

	Without Film Adhesive		With Film Adhesive	
Specimen	Wt_i (g)	Wi_f (g)	Wt_i (g)	Wi_f (g)
1	1.4	0.8	0.9	0.5
2	1.4	0.8	0.9	0.5
3	1.6	0.8	0.9	0.6
4	1.4	0.8	0.9	0.5
5	1.2	0.6	0.9	0.6
6	1.4	0.8	0.9	0.6
Average	1.4	0.767	0.9	0.55

These values aren't really that helpful until it converted to its volume fractions. The averaged weights of the fiber and the matrix, with the use of the average dimensions, the volume fractions were calculated. The densities of the materials were obtained by using the dimensions of the specimens and the weights of the materials for the densities. The densities for the case the film adhesive of the composites were 6.57 g/cm^3, for the LTM45 was 5.60 g/cm^3, and for the matrix were 1.24g/cm^3. The volume fractions were than calculated; this can be seen in Table 8. The volume fraction shows that the matrix content drops by 7.5%, opposite from the 6% increase in weight. This could be due to the accuracy of the volume fraction because it's based off of the densities and the volumes used for the calculations. A constant volume obtained from the test specimen was used for both the fiber and matrix volumes. The table also shows the critical volume fraction for the fiber of the LTM45. The obtained volume fraction for the fiber was found to be 68.5%. The volume fraction of the composite without a film adhesive is within this value while the volume fraction of the composite with a film adhesive is 3.2% above the value. This means that the composite has exceeded the strain hardening and plastic flow of the matrix. This makes the composite more brittle than the optimum conditions that it was manufactured

for. Since the research is to study the effects between the fastener and hole of the composite sandwich panel, the volume fraction of the fiber is suitable for testing.

Table 8 - Volume Fractions with and without the film adhesive

V_f w/o Film Adhesive	64.2%
V_m w/o Film Adhesive	35.7%
V_f w/ Film Adhesive	71.7%
V_m w/ Film Adhesive	28.3%
V_{crit}	68.5%

Material Properties

The material properties of the composite sandwich panel was tested and analyzed. The first material property that was tested was the Elastic Modulus and the Poisson's Ratio of the core, FR-6710, and the face sheet, LTM45. The core was tested by cutting the specimens of the core into dimensions of 2" x 2" x 0.5". The face sheet was tested by following ASTM standards on calculating the material properties of the composite, ASTM D3410.

The material properties of the core, FR-6710 was tested and analyzed. The test apparatus can be seen in Figure 98 - Foam Material Property Testing. The Elastic Modulus and the Poisson's Ratio of the material were calculated from the stress/strain curve of the test specimens; this can be seen from Figure 97 - Load/Extension for Last-A-Foam FR-6710 The figure shows that the five test specimens have similar slopes, showing that the Elastic Modulus for the foam is consistent with one another. The Poisson's Ratio was calculated by adding a strain gage on the horizontal direction to calculate the value. The results can be seen in Table 9.

Table 9 - Material Characteristics of Last-A-Foam FR-6710

	Compressive Strength (psi)	Percent Difference (-)	Elastic Modulus (psi)	Percent Difference (-)	Poisson's Ratio (-)	Percent Difference (-)
Manufacturer	340	-20.6%	11,240	-34.8%	0.30	-63.3%
Experimental Testing	270		7,330		0.11	

The property of the foam was analyzed and it decreased by 20.6% for the compressive strength and about 34.8% for the elastic modulus for the material. The decrease in compressive strength and elastic modulus could be due to the way the specimens were tested. The testing procedure for the foam can be different to how the manufacturer tested the foam causing the significant difference in material characteristics. The decrease in material property can also be due to the degradation of the material from being stored too long after purchasing and receiving the material. Being stored in an uncontrolled environment can affect the material characteristics of the foam, for example, storing the foam in the sun would make it susceptible to UV degradation. For

comparison with the FEM, the experimental results of the material characteristics would be used for comparison between experimental and numerical solution.

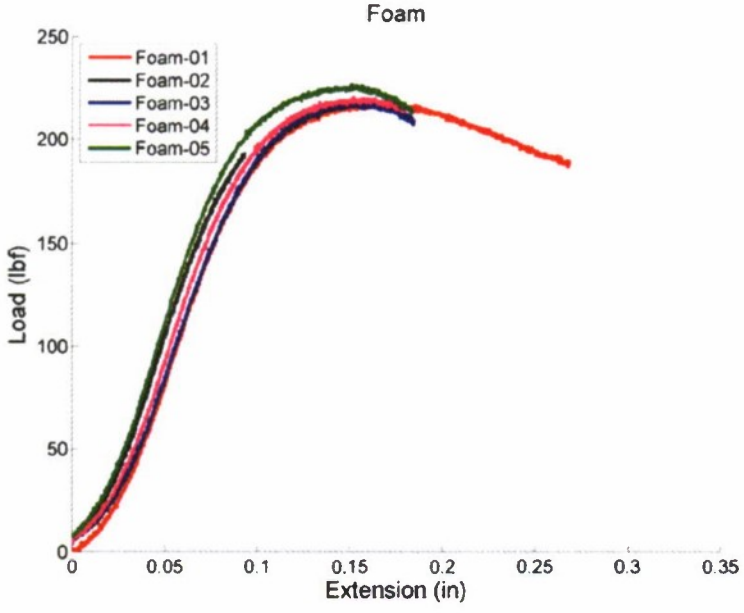


Figure 97 - Load/Extension for Last-A-Foam FR-6710

Figure 98 - Foam Material Property Testing

The material properties of the face sheet, LTM45 was tested and analyzed. The test apparatus can be seen in Figure 99 - Test Specimens for LTM45. The Elastic Modulus and the Poisson's Ratio of the material were calculated from the stress/strain curve of the test specimens. The Poisson's Ratio was calculated by adding a strain gage on the horizontal direction to calculate the value. The results can be seen in Table 10.

Table 10 - Material Characteristics of LTM45

	Compressive Strength (psi)	Percent Difference (-)	Elastic Modulus (psi)	Percent Difference (-)	Poisson's Ratio (-)	Percent Difference (-)
Manufacturer	214,000	-35.0%	17,260,000	-58.6%	(N/A)	N/A
Experimental Testing	130,000		7,149,000		0.11	

The material properties of LTM45 was obtained from the manufacturer and tested to validate the data. From the table, the material properties of the material decreased by 35.0% for the compressive strength and about 58.6% for the elastic modulus of the material. This decrease in material property can be due to the degradation of the material. The material was donated by Boeing over three years ago and was stored in a freezer that barely meets the cooling standards of the carbon. When the material was donated from Boeing, it was considered by the FAA to not meet its specifications, but it is still suitable for research usage. The values obtained from experimental testing shows that the material properties of the LTM45 is still adequate for the research because compared to the foam the material is infinitely stiff a can handle the objectives of the thesis. Just like the foam, the experimental results of LTM45 will be used for the material characteristics for the FEM.

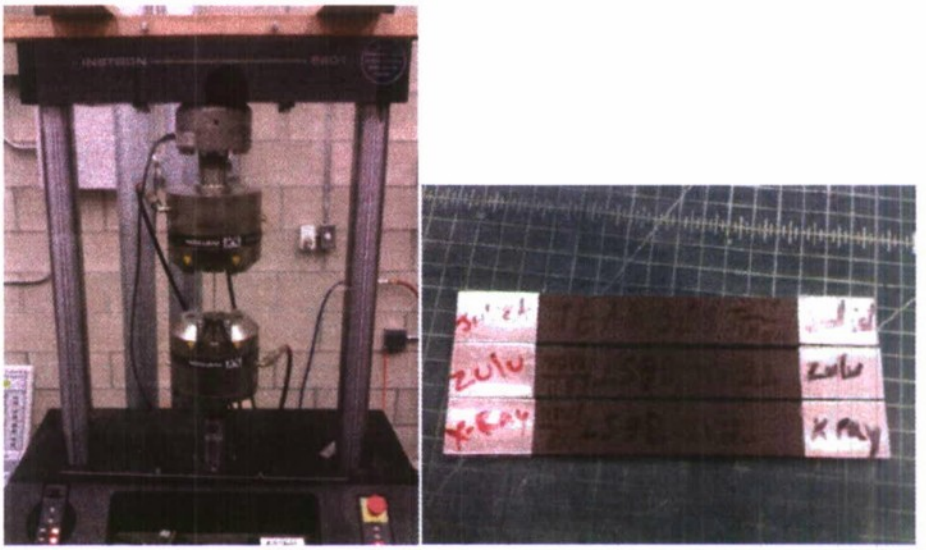


Figure 99 - Test Specimens for LTM45

The next material property that was tested was the overall Elastic Modulus and the Poisson's Ratio of the whole composite sandwich panel; the results can be seen in Figure 100 - Stress/Strain of Test Specimens without Notches. The specimens plotted show the stress/strain curve of the specimens manufactured from 3.1-Monotonic and Cyclic Tensile Loading (Eugene). The results show that stress/strain curve of the specimens are

only the linear part of the curve to be able to calculate the true Elastic Modulus of the overall composite sandwich panel without any notches (holes). The results can be seen in Table 11.

The theoretical results of the Elastic Modulus were calculated using the equations from Chapter 5- Numerical Analysis. The typical failure of the test specimens, which can be seen in Figure 101 - Failure of overall material property specimens, happens in the wood part of the core. The wood section of the core starts to fracture and it propagates into the foam section of the core which is then transferred to the face sheets. When the fracture reaches the face sheets, it starts to delaminate from the core.

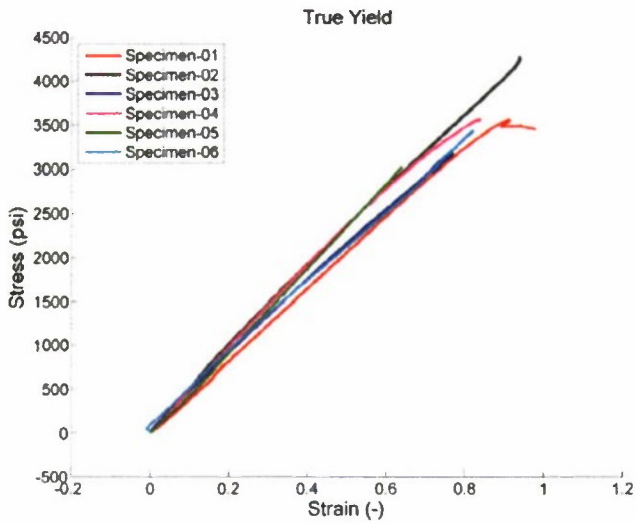


Figure 100 - Stress/Strain of Test Specimens without Notches

Figure 101 - Failure of overall material property specimens

Table 11 - Elastic Modulus of Composite Sandwich panel

	Elastic Modulus (psi)	Percent Difference (-)	Poisson's Ratio (-)	Percent Difference (-)
Theoretical	62,650	33.6%	0.30	28.7%
Experimental Testing	41,600		0.214	

The result shows that the theoretical Elastic Modulus of the composite sandwich panel was 62,650 psi. This was calculated by assuming a Poisson's Ratio of 0.3 in the equations because the calculations from the equations gave a singular matrix. The experimental testing gave an Elastic Modulus of 41,600 psi and a Poisson's Ratio of 0.214. The theoretical results over predicted the material properties of the composite sandwich panel. This over prediction is due to the calculations are for laminates and not composite sandwich panels, but it can be assumed that the core can act as part of the laminate.

Strain Rate/Control Group Results

The strain rate for the static testing needed to be investigated first before investigating the effects of the DADs on the composite sandwich panel. The strain rates that were investigated were between the rates of 0.5 mm/min to 1.5 mm/min with a spacing of 0.25 mm/min. The results can be seen in Table 12. Five control group plates were manufactured and cut down to specimens. A random sampling from the five plates was chosen for each strain rate to make sure that the results would not be biased from testing the strain rate from the same plate. The results showed that all five strain rates had similar failure load varying from 926 lbs to 1,011 lbs and had similar elastic modulus varying from 37,530 psi to 48,940 psi. Variance in the failure load and the elastic modulus is due to the manufacturing defects in the specimens. The results show that the strain rate did not greatly affect the characteristics of the panel, but it did show a variance in standard deviation for each case.

Table 12 - Strain Rate Selection Results

Rate	P_{cr}	Std Dev	E	Std Dev
(mm/min)	(lb)	(-)	(psi)	(-)
0.5	1,011	4.91%	45,170	4.21%
0.75	983	9.54%	48,940	7.53%
1.0	934	36.51	37,530	51.92%
1.25	926	12.71%	45,010	9.51%
1.5	940	7.34%	46,270	5.03%

The table shows that the slowest rate, 0.5 mm/min had the lowest standard deviation for both failure load and elastic modulus. This rate would be used for all static testing because it gives the best standard deviations compared to the other rates translating a better accuracy to future tests; this can be seen in Figure 102 - Standard Deviation for 0.5 mm/min. The results show that five of the six test specimens are within one standard deviation from the mean, while the first test specimen is considered an outlier, this being represented with an X as a symbol.

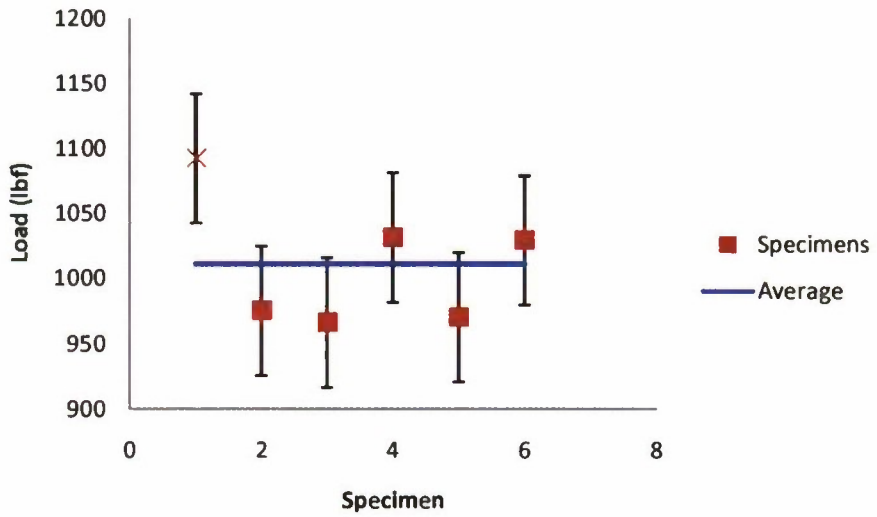


Figure 102 - Standard Deviation for 0.5 mm/min

The results for the case, such as extension, stiffness, elastic modulus, and yield stress can be seen in Table 13. The outlier would be highlighted in yellow for the table. The table shows that the load, the yield stress, and the extension are within 5% with one other in standard deviation meaning that the specimens are failing similar to one another. This is very useful for fatigue testing because the yield stress or the first drop shows consistency to set the extension from experimental testing as the failure criterion for fatigue.

Table 13 - Results for 0.5 mm/min Strain Rate

Test #	Test Sample	Load	Yield Stress	Extension	K	E
1	0225-01	1091.5	954.84	0.0538	24598	43789
2	0225-06	974.0	853.71	0.0540	22920	40864
3	0226-01	967.0	855.31	0.0507	27390	49375
4	0226-06	1030.7	908.31	0.0527	22551	40441
5	0227-01	969.4	850.38	0.0585	30342	54194
6	0227-06	1027.6	893.31	0.0569	23909	42382
Average Load		1010.014	885.98	0.05443	25285.1	45174.2
Std Dev		4.89%	4.67%	5.21%	11.92%	12.11%

For the rest of the strain rates that were tested, the standard deviations for the remaining strain rates can be seen in Figure 103, Figure 104, Figure 105, and Figure 106. The experimental results such as extension, stiffness, elastic modulus, and yield stress can be seen in Table 14, Table 15, Table 16, and Table 17. As before,

the figures show the outlier with an X marker and the specimens within one standard deviation have a square marker. The other strain rates have more outliers than the 0.5 mm/min by having as much as eight outliers and as least as four outliers. The worst of the strain rates was the 1.0 mm/min where eight of the nine test specimens are one standard deviation from the median. It also has the highest standard deviation for the ultimate failure load and the Elastic Modulus is the lowest of the five strain rates.

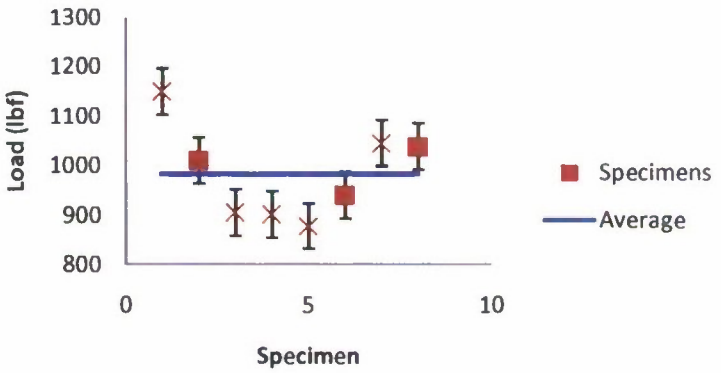


Figure 103 - Standard Deviation for 0.75 mm/min

Table 14 - Results for 0.75 mm/min Strain Rate

Test #	Test Sample	Load	Yield Stress	K	E
1	0225-02	1149	1001	30300	53965
2	0225-07	1010	891	28829	52063
3	0226-02	905	801	28060	50670
4	0226-07	901	784	25921	45082
5	0227-02	877	765	24145	42984
6	0227-07	940	821	26416	47155
7	0305-01	1044	922	27133	48779
8	0305-02	1038	915	28800	50824
Average Load		983	862	27450	48940
Std Dev		9.49%	9.54%	7.11%	7.53%

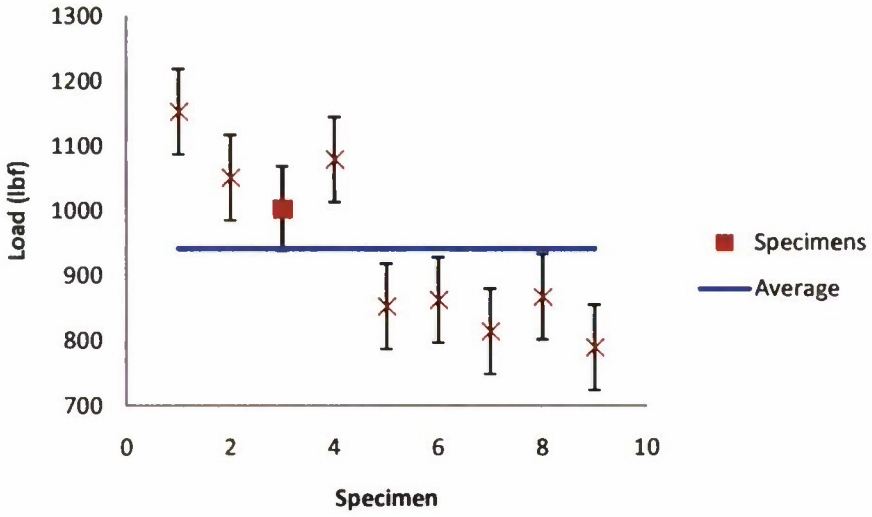


Figure 104 - Standard Deviation for 1.0 mm/min

Table 15 - Results for 1.0 mm/min Strain Rate

Test #	Test Sample	Load	Yield Stress	K	E
1	0225-03	1153	1004	30549	54272
2	0225-08	1052	919	27990	49681
3	0226-03	1002	883	26295	47176
4	0226-08	1076	94	27533	4875
5	0227-03	853	748	28185	50366
6	0227-08	861	752	19646	34941
7	0305-03	796	692	24581	43640
8	0305-04	865	749	28030	48665
9	0305-07	751	670	22741	4113
Average Load		934	657	26172	37525
Std Dev		14.93%	36.15%	12.72%	51.92%

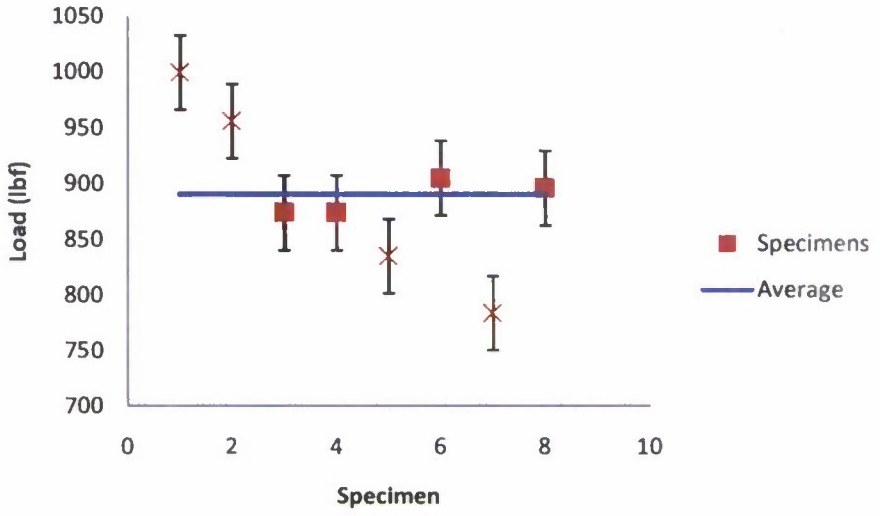


Figure 105 - Standard Deviation for 1.25 mm/min

Table 16 - Results for 1.25 mm/min Strain Rate

Test #	Test Sample	Load	Yield Stress	K	E
1	0225-04	997	866	28798	51134
2	0225-09	954	832	22460	40020
3	0226-04	874	768	22835	40915
4	0226-09	1167	1017	28453	50613
5	0227-04	833	731	25978	46149
6	0227-09	904	794	26069	46526
7	0305-05	784	679	24287	42402
8	0305-06	896	796	23737	42331
Average Load		926	810	25327	45011
Std Dev		12.71%	12.51%	9.54%	9.51%

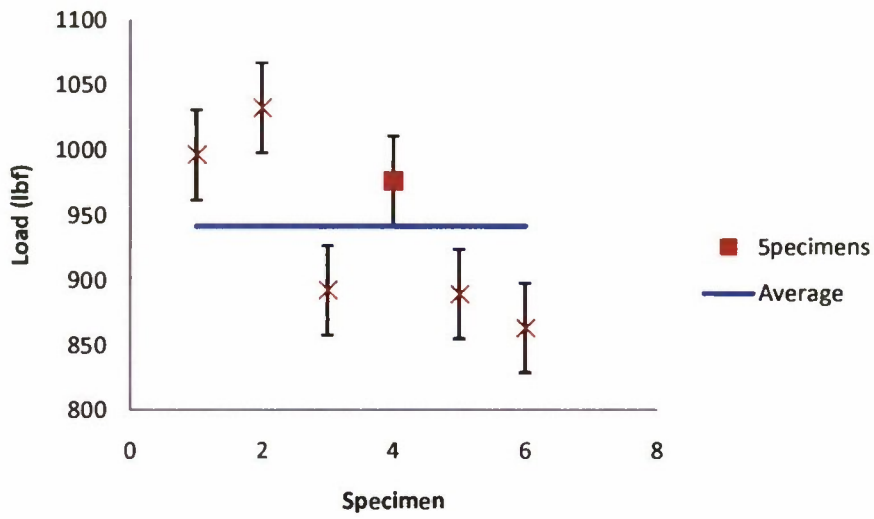


Figure 106 - Standard Deviation for 1.5 mm/min

Table 17 - Results for 1.5 mm/min Strain Rate

Test #	Test Sample	Load	Yield Stress	K	E
1	0225-05	995	865	26856	47394
2	0225-10	1031	904	26181	46929
3	0226-05	892	786	23578	42446
4	0226-10	974	850	25942	46118
5	0227-05	887	782	27550	49404
6	0227-10	862	754	25291	45292
Average Load		940	823	25900	46264
Std Dev		7.34%	7.03%	5.32%	5.03%

The load/extension curves for the 0.5 mm/min case had the best testing results because it had the least amount of test specimens that had a drop in the load/extension curve and continued up after; this can be seen in Figure 107. The trend for the control group is that the specimen rises and then drops and rises again but drops at a lower load than before. This is due to the fastener slowly moving along the composite sandwich panel. The second rise in the curve is caused by the face sheet re-fracturing as the fastener moves along it. This is also known as micro buckling where the fibers break one at a time instead of it breaking at once. The leveling force seen at the end of the curve is when the fastener is "sawing" through the composite material with ease.

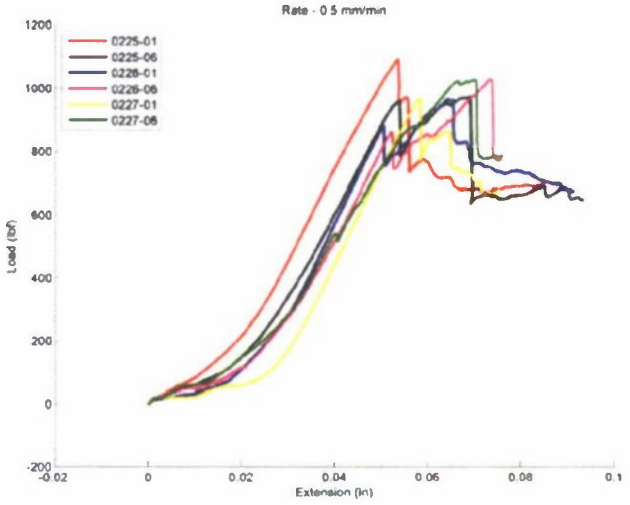


Figure 107 - Load/Extension for 0.5 mm/min Rate

Load versus extension curves for the remaining strain rates can be seen in Figure 108. The four figures show that the remaining four strain rates have consistent test results by having similar slopes to one another. Just like the 0.5 mm/min strain rate, the test specimens for each strain rate group rises and then drops and rises again but drops at a lower load than before. This is due to the fastener slowly moving along the composite sandwich panel. The second rise in the curve is caused by the face sheet re-fracturing as the fastener moves along it. This is also known as micro buckling where the fibers break one at a time instead of it breaking at once. The leveling force seen at the end of the curve is when the fastener is "sawing" through the composite material with ease.

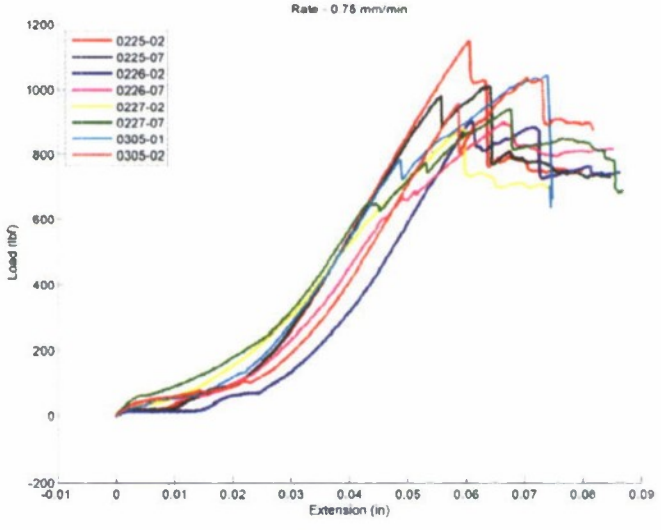


Figure 108 - Load/Extension for 0.75 mm/min Rate

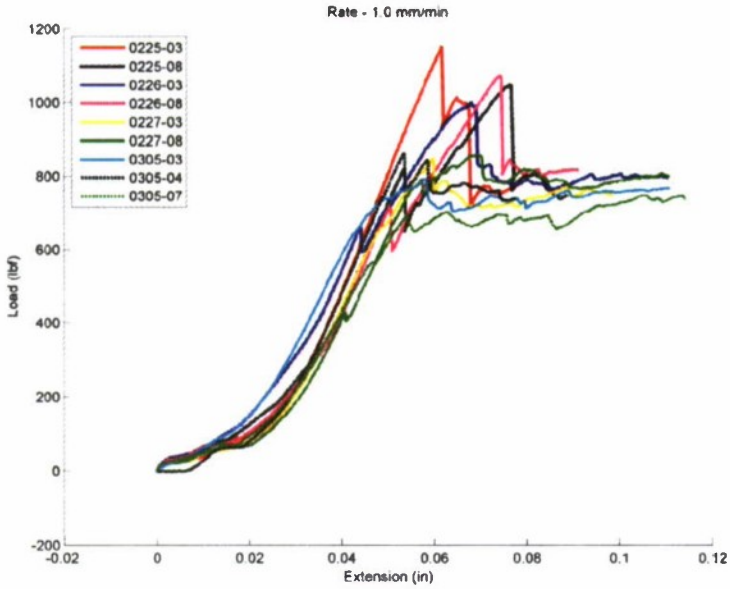


Figure 109 - Load/Extension for 1.0 mm/min Rate

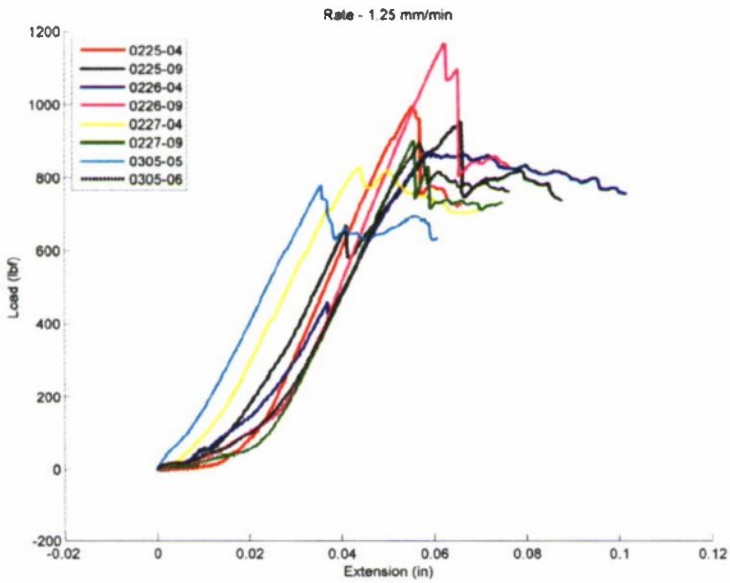


Figure 110 - Load/Extension for 1.25 mm/min Rate

Chapter 5- Numerical Analysis

The following section is the numerical analysis results for the monotonic loading and fatigue case as well as for the damage arrestment device shear key.

5.1 Monotonic Loading and Fatigue (First application by Eugene)

Model Geometry

During the model development it was determined that a 2-D finite element analysis (FEA) would be adequate for the preliminary analyses performed. A longitudinal view of the part was selected for the analysis because it would provide an adequate representation of the stress state longitudinal. The model geometry and coordinate system used for the analysis is shown in Figure 111. The upper and lower fiberglass skins were modeled as a single layer their material properties were determined experimentally. The y-axis runs vertically through the part and the x-axis runs horizontally along the length of the part. Key geometric properties for the analysis are shown in Table 18. A 3-D analysis was not selected because of the high computational cost associated with using solid elements. The delamination model was considered identical to model without delamination except that a seam was added to the region where the upper skin was not connected to the lower skin. The seam represents the piece of non-porous Teflon paper that separates the upper skin from the core.

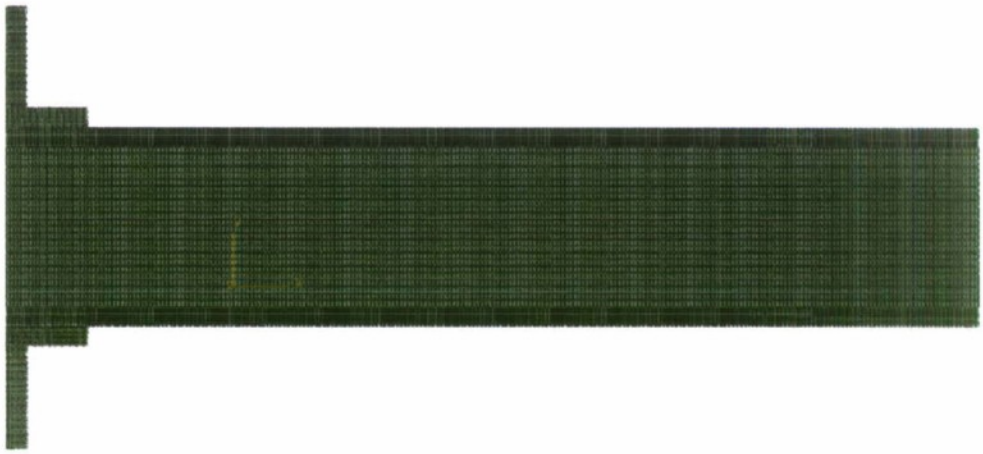


Figure 111 - Finite element geometry for test specimen

Table 18 - Sandwich structure geometry

	Thickness (inch)	Length (inch)	Out of plane thickness (inch)
Fiberglass skins (each)	0.08	6.0	1.50
Foam	0.787		

Material Properties Development

Initially during the analysis it was assumed that the constitutive response for both the fiberglass skins and foam core was linearly elastic, homogeneous, and isotropic. In reality that is not the case because both materials are orthotropic and exhibit plastic characteristics. After looking at the initially linearly elastic results it was determined that the plasticity of the foam would need to be model to accurately capture the failure forces. The material properties for the Divinycell foam core were obtained from the manufactures' website[23] and the tensile (Figure 112) and compressive (Figure 113) stress strain curve were obtained during Dr. Mitra's initial shear key research[4].

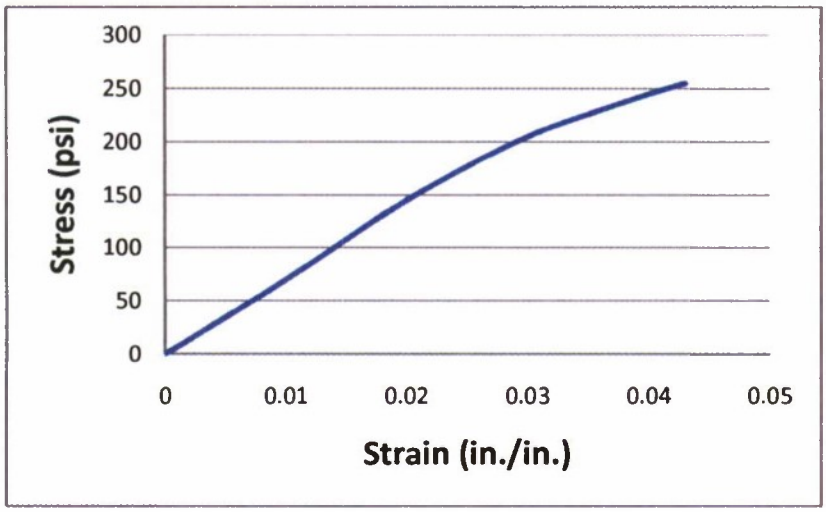


Figure 112 - Tensile response of Divinycell H100 foam[4]

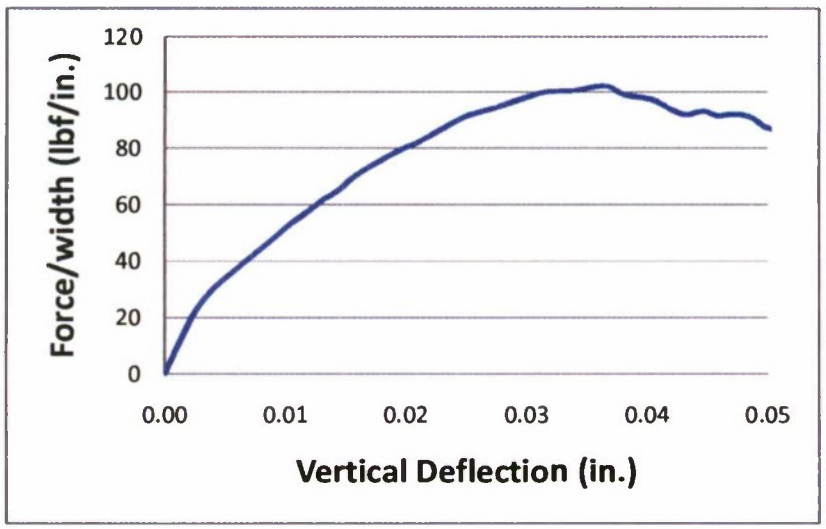


Figure 113 - Compressive response of Divinycell H100 foam[4]

Two different material models were used to represent the non-linear and plastic behavior of the foam after its yield strength was reached. Modeling the foam as a hyperelastic material was the first step in trying to more accurately capture the failure in the foam. A hyperelastic foam model was constructed for the foam by extrapolating data from the tensile and compressive tests for the Divinycell H100 foam. The hyperelastic foam curve used as an input into the material properties in Abaqus/CAE is shown in Figure 114.

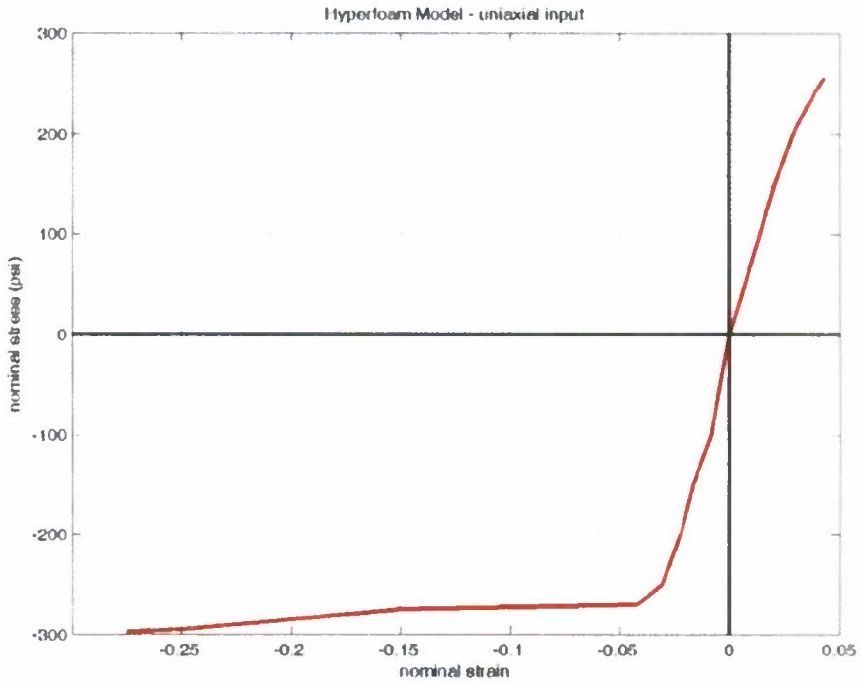


Figure 114 - Hyperelastic foam model

The other foam model used was an expansion of the crushable foam model originally used by Dr. Nilanjan Mitra[4] during his shear loading research. The crushable foam hardening model used for the analysis is shown in Figure 115. Additional information regarding hyperelastic material properties could be found by consulting the Abaqus User's manual.

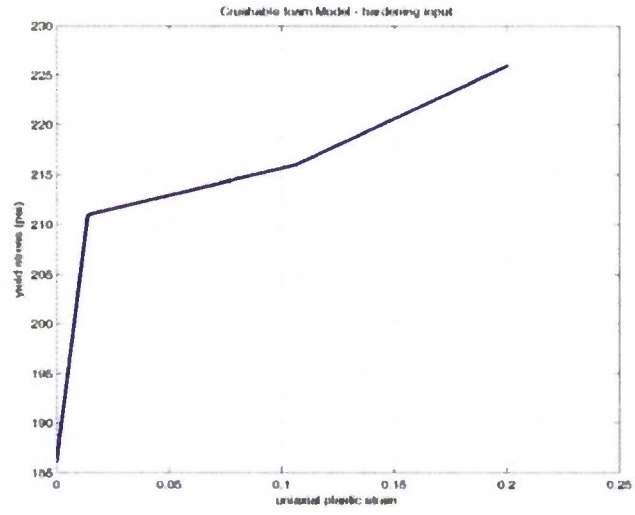


Figure 115 - Crushable foam plasticity model

Material properties from the fiberglass manufacturers were not available so testing had to be performed to determine them. A 120Ω linear strain gage with a gage factor of 2.1 was added near the middle of the test specimen and two holes were drilled near both the ends of the specimen. One hole was be used to constraint the specimen and the other hole was be used to hand weights off of. The strain gauge was wired up in a quarter Wheatstone bridge configuration using the yellow strain gage box shown in Figure 116. Additional information regarding the crushable foam hardening could be found by consulting the Abaqus User's manual.

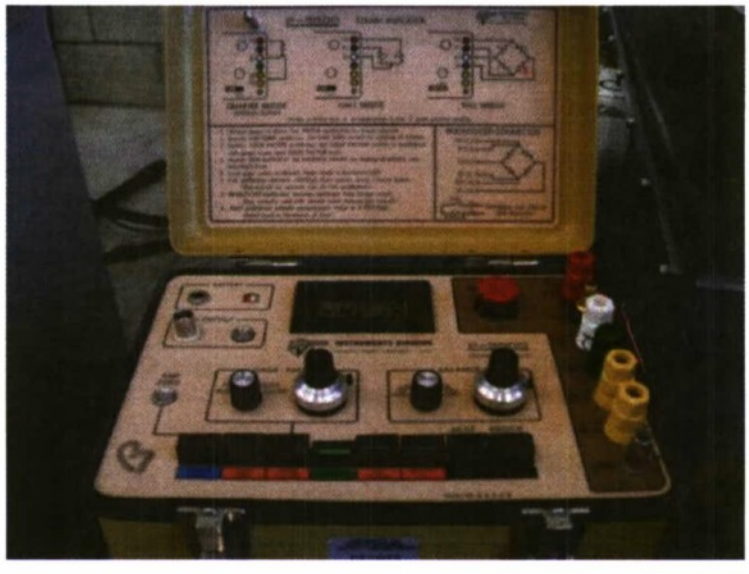


Figure 116 - Yellow strain gage box with 1/4 bridge configuration

An aluminum test specimen was first used to ensure that the method used for the experiment was accurate because the material properties for aluminum were known. Figure 117 shows the aluminum test specimen with a strain gage as well as the boundary conditions used to fix the specimen. Figure 118 shows the jig used to hold the weights during the material calibration.

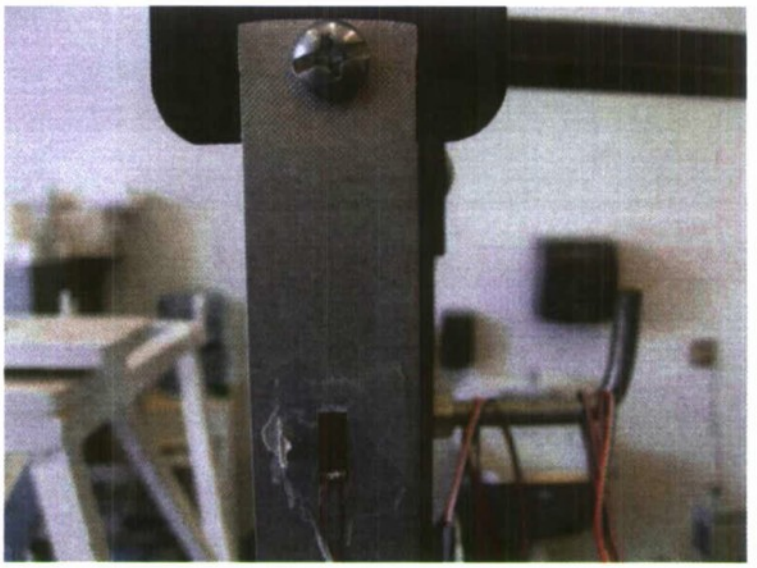


Figure 117 - Aluminum test specimen under tensile load

Figure 118 - Applied loading for material calibration

Before any weights or stain readings were recorded, the strain gage box was zeroed. Weights were then added in 10 lb increments and strain readings were recorded. Figure 119 shows the results obtained from the testing for the aluminum and fiberglass skin test specimens. Stress was plotted on the y-axis and strain was

plotted on the y-axis. Young's Modulus for the aluminum test specimen was experimentally determined to be 9.6e6 psi which was only 6% off the known Young's Modulus value of 10e6 psi. Young's Modulus for the fiberglass skin lay-up was determined to be $5.7x10^6$ psi. There is also further confidence in the test because the slopes of the lines are straight and there were no jagged points, which generally represent incorrect strain readings.

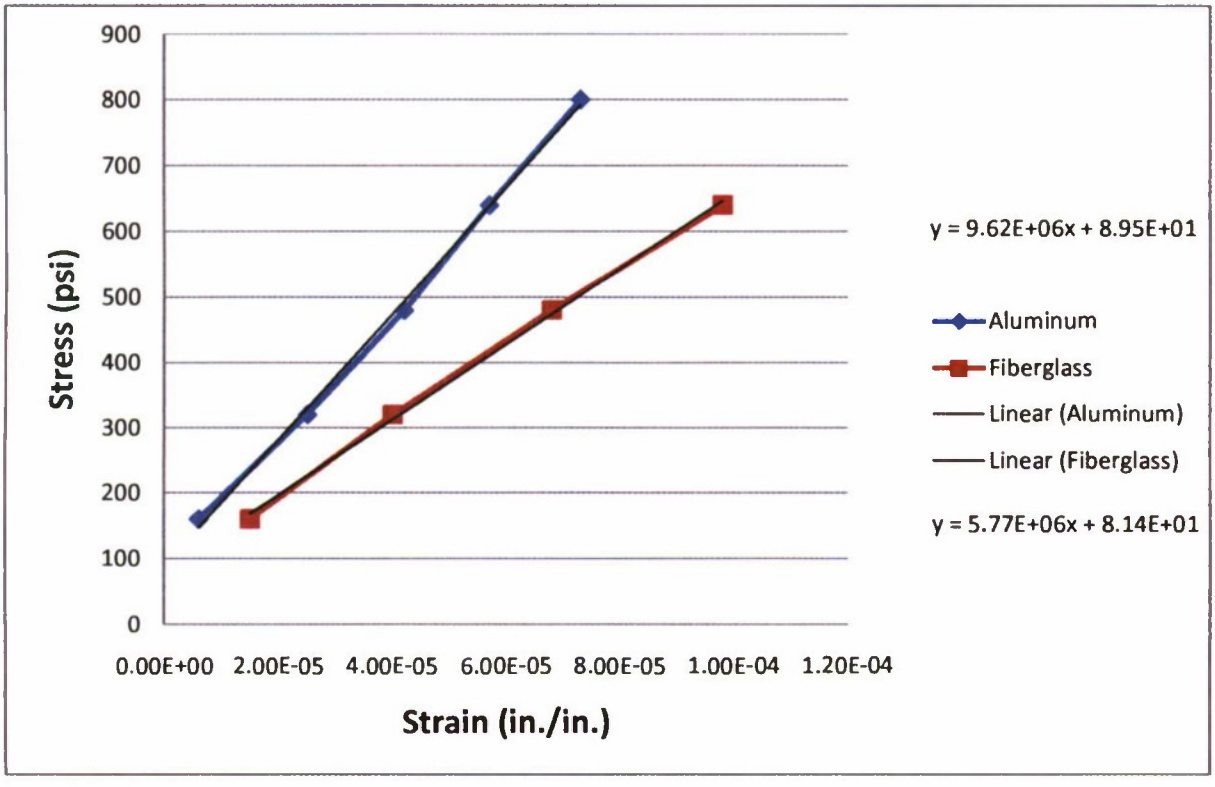


Figure 119 - Material calibration plot

Loads/Boundary Conditions

The boundary conditions selected for the test specimens with no initial delamination and a 0.5 inch initial delamination analysis are shown in Figure 120. Displacement boundary conditions representing displacement test data were applied to the upper aluminum L-bracket. Both test specimens had the lower aluminum L-bracket fixed (U1=U2=0). The back edge of the test specimens were also fixed using the jig which was fixed to the Instron machine. Figure 120 shows the loads and boundary conditions used for the models.

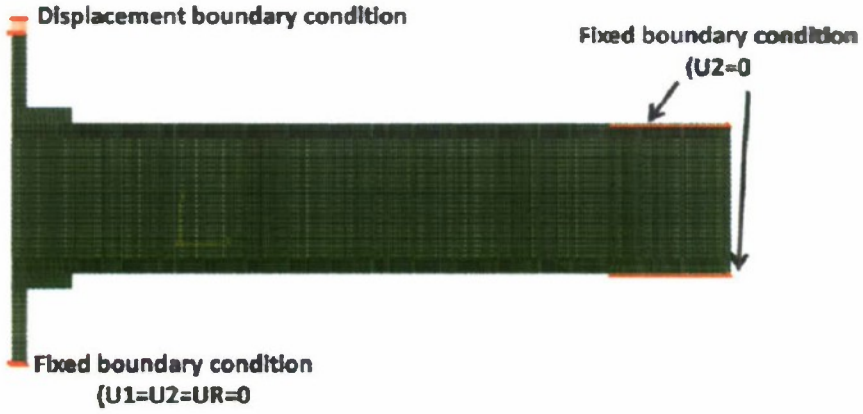


Figure 120 - FEM displacement boundary conditions

Mesh Development

Shell elements were selected to represent the fiberglass skins and the foam core. The primary reason for choosing shell elements was because stress and displacement were expected to be the same throughout the entire depth of the test specimens. 3-D solid element was not selected because of the high computational costs associated with 3-D elements. The 2-D shell elements should provided a good approximation for the failure loads.

The 2-D shell elements were constructed in Abaqus/CAE by creating a 2-D shell. The skins were made by partioning the 2-D rectangular shell using the thickness of the upper and lower skins. Additional partitions were added to represent the delaminated region where the force boundary conditions were applied.

Linear elements with reduced integration were selected to decrease computation time. It was deemed appropriate that linear elements with reduced integration could be used because the degree of confidence obtained through the mesh convergence study. A quadrilateral structured mesh was assigned to all regions of the test specimens. Plane stress elements were initially selected but modeling issues arose when plasticity was adding to the foam, which led to plane strain elements being used. Figure 121 shows the final mesh that was used for the analysis. Table 19 shows the mesh quality used for the numerical analysis.

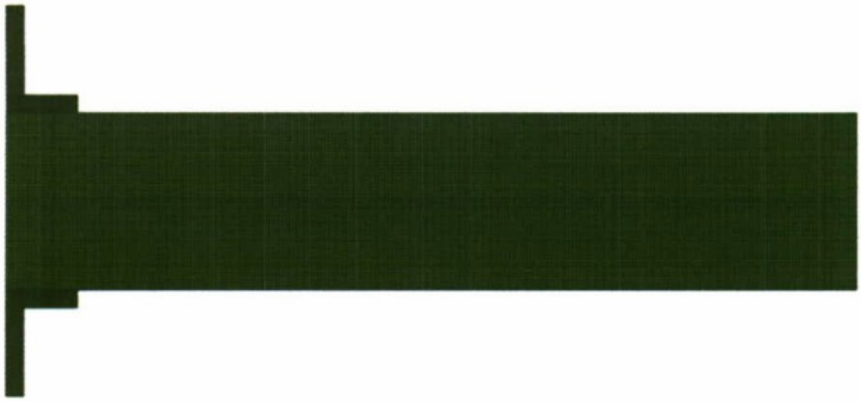


Figure 121 - Final mesh used for analysis

Table 19 - Mesh quality for final mesh

	% Elems	# Elems	Average
Face Corner Angle < 45°	0	0	90°
Face Corner Angle > 135°	0	0	90°
AR > 5	0	0	2.08

Analysis

A non-linear static analysis was selected to simulate the monotonic loading conditions. A static analysis was selected because the monotonic loading conditions represented static testing. A linear static analysis was initially performed because plasticity was not modeled but the initial FEA analysis results needed plasticity to match the experimental data.

Results

Figure 122 shows the vertical deflection (U2) contour for the test specimen with no initial delamination. The maximum vertical deflection of 0.03304 inch occurs at the top of the aluminum tab.

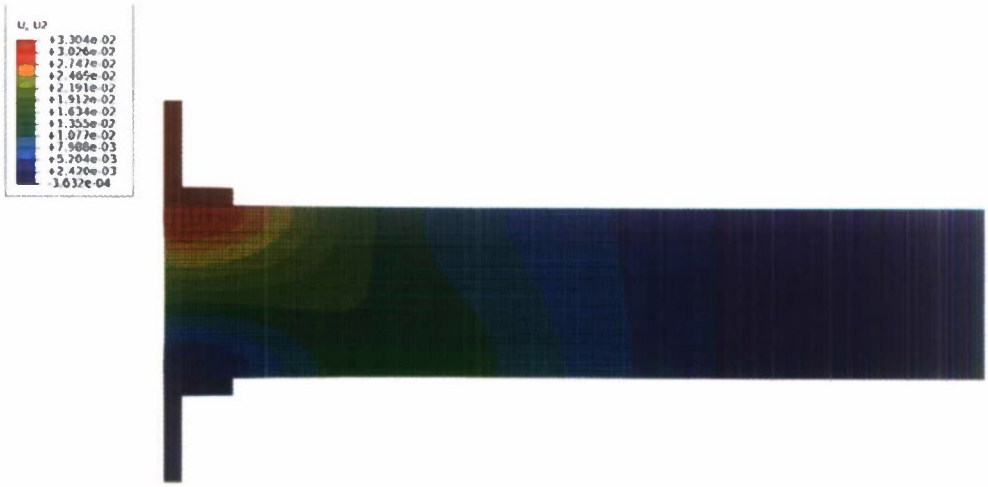


Figure 122 - FEA vertical deflection for test specimen without initial delamination

Figure 123 shows a vertical deflection contour for the test specimen model containing an initial delamination of 0.5 inch. The maximum vertical deflection of 0.03606 inch is shown in Figure 123 at the upper edge of the aluminum tab. The 0.5 inch delamination deflection contour is very similar to the deflection contour in Figure 122 except that there is no seam in Figure 122.

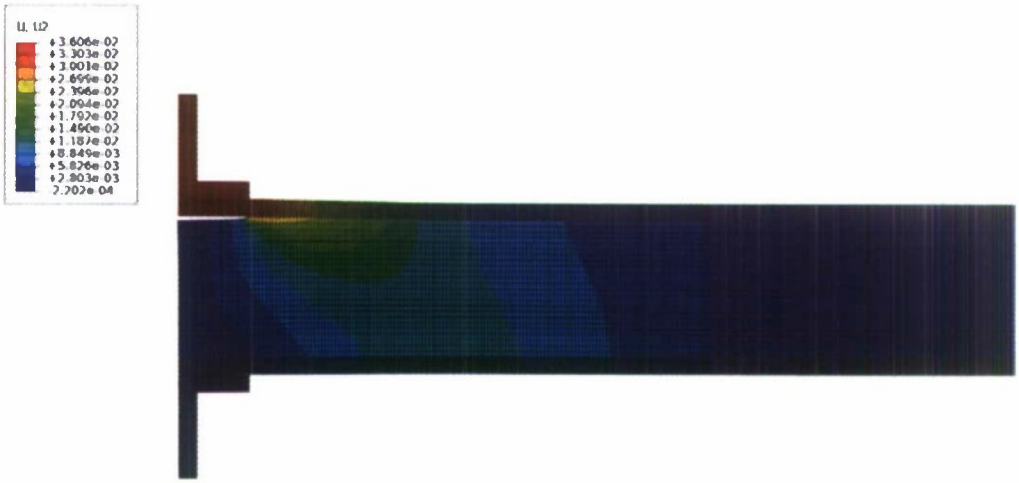


Figure 123 - FEA vertical deflection for test specimen with 0.5 inch initial delamination

5.2 Damage Arrestment Device (Richard B.)

Finite Element Modeling

For this chapter, the FEM analysis of the composite sandwich panels will be discussed. The creation of the model geometries and the mesh configurations for each model will be examined. The loading and boundary

conditions applied to the models will also be referenced. Lastly, the displacement and stress distributions results of each model will be discussed.

Finite Element Model Geometry

The finite element model for the research will contain three different models for each the control group and the DAD group for the analysis. The software that will be used for the analysis is GeoStar or CosmosM, which is made by SolidWorks. The three different models will consist of just the test specimen, one with the test specimen and the bushing, and one with the test specimen, bushing and the fastener. The analysis will have a base model to work off, which refers to the test specimen because the other parts of the test will be added to the model. The base model can be seen in Figure 124.

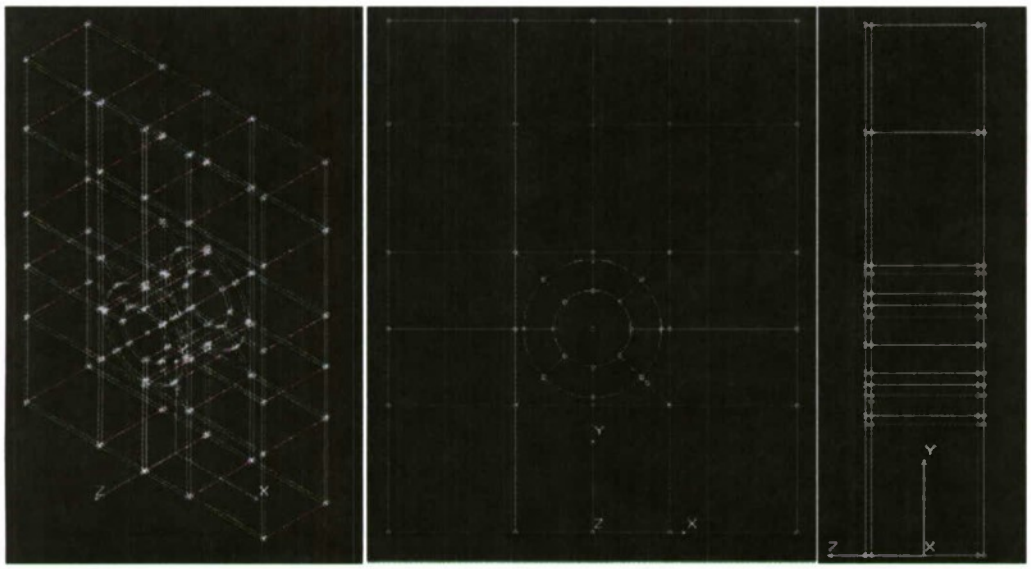


Figure 124 - Base FEM Geometry

The FEM base model is only half the model of the specimen, having only one hole rather than two, because of symmetry in the geometry. The geometry dimensions of the combined surfaces will be 2.5 x 2. The software accepts unitless values, allowing the operator to non-dimensionalize the data. The units have to be recorded for further analysis. The base geometry contains 32 surfaces that are extruded into volumes. The surfaces around the hole are segments of circles and outside of them are triangle shaped quadrilaterals. All surfaces are quadrilaterals because the maximum surfaces that can be created in the software are with four lines/curves. The radius of the circular surface around the hole will be 40% of the radius of the hole, which is 0.15". The purpose of this is to have a fine mesh of elements around the hole to capture the bearing failure and stresses from the fastener. The surfaces away from the hole are rectangles. The rectangular surfaces around the hole represent the DADs for later models. Using this method will greatly simplify the model and make it comparable to one another by having similar geometries and meshes. The 32 surfaces will then be extruded into three volumes to represent the test specimen. The three volumes will be the two face sheets and the core. The

surfaces will be extruded into the dimensions of each volume. The face sheet will have a depth of 0.026 and the core will have a depth of 0.5. The model geometry will then be ready to mesh.

The next model that was created was the base model with the bushing inside the hole. The bushing is being added to the FEM model to alleviate the pressure on the test specimen; this can be seen in Figure 125. The half model of the bushing will only be modeled. The bushing will have a radius of 0.0625" because the inner diameter of the bushing was 0.25" and the outer diameter of the bushing was 0.375".

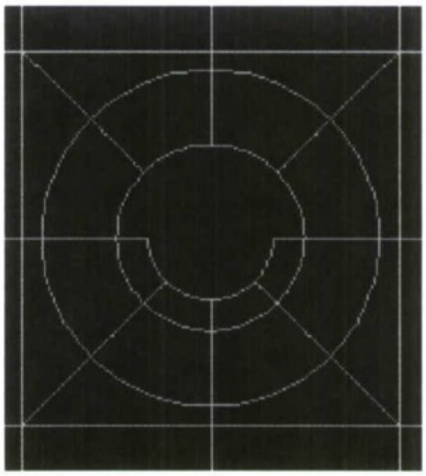


Figure 125 - Base geometry with bushing

The last model for the control group FEM was the base model with the bushing and the fastener; this can be seen in Figure 126. Just like the bushing, the fastener is added to the FEM model to simulate the true loading and geometry of the test specimens. The half model of the fastener would only be modeled. The fastener will have a radius of 0.125 because of the diameter of the fastener. The geometry will be created by using the three surface features to make each surface of the model.

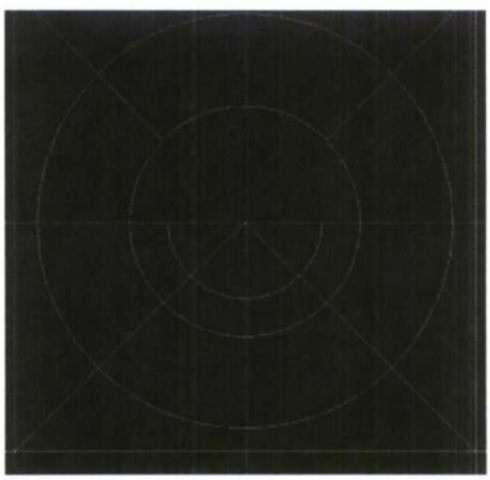


Figure 126 - Base geometry with bushing and fastener

The model for the DAD geometry is different than the base geometry. The planar geometry is the same as the base model by having the same number of surfaces; this can be seen in Figure 127. The geometry has the same dimensions of the overall sandwich composite panel, a dimension of 2.5 x 2. The difference in the geometry is in the Z direction, where the middle surfaces that represent the DAD are extruded into the core. The dimensions of the geometry were adjusted to accommodate the meshing of the elements. Since the thickness of each layer of LTM45 is 0.013", the thickness of the material for the FEM model would be a thickness of 0.02". This made the thickness of the face sheet become 0.04" and the thickness of the DAD became 0.06". This creates an even number of elements in the core to match the number of elements with the DAD and the face sheet. The thickness of the core remains the same. This geometry is the base model for the FEM of the DADs with the bushing and the fastener. The geometries for the DAD models with the bushing and the fastener are be the same as the previous model. The bushing would be added on the lower part of the hole on the composite sandwich panel and the fastener is added on top of the bushing. The half models of the bushing and the fastener will be modeled.

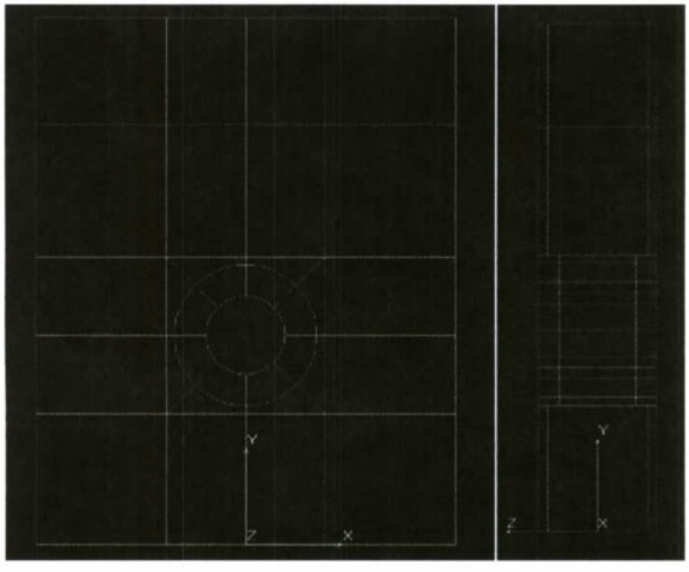


Figure 127 - DAD geometry

Finite Element Model Meshing

The base model uses a solid element to be able to calculate the strains and stresses inside the test geometry. The element mesh of the base model can be seen in Figure 128. All the elements for the model are quadrilateral elements. Two elements will make up the face sheet in the orthogonal direction of the plane. Each element will be represented as a layer of LTM45. Six elements make up the core in the orthogonal direction of the plane. Six elements were chosen because having too many would make the element count too high and too long to calculate; having fewer elements would decrease the accuracy of the analysis. The mesh will grow from a coarse mesh from the top edge to a fine mesh near the hole. A parametric mesh for each volume was created.

This means that the user defines the number of elements in the X, Y, and Z directions of the volume. The figure is divided into groups to make it easier to describe the mesh densities of each volume/surface. Group 1 will be the top horizontal surfaces. Group 2 will be the next horizontal surfaces below Group 1. Group 3 represents the rectangular surfaces on both sides of the hole. Group 4 represents the circular sections and the triangle shaped rectangular surfaces. Group 5 represents the bottom horizontal surfaces of the geometry. The material properties of the LTM45 and the FR-6710, the experimental results, are be used for the input for the elements for the face sheets and the core. The parametric meshing of each grouping can be seen in Table 20. The total element count for the whole geometry is 25,000 elements with 40,197 nodes. A majority of the elements are contained in Group 5 due to the fine mesh needed to capture the bearing failure around the hole.

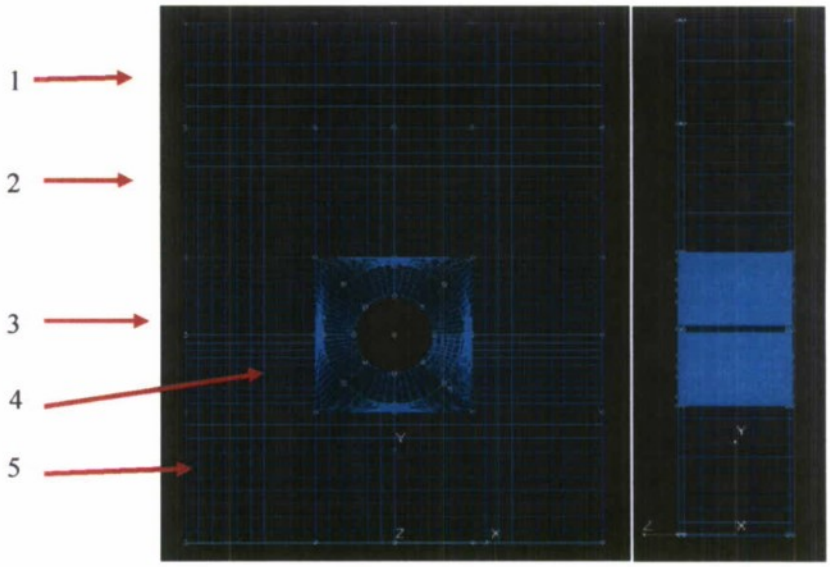


Figure 128 - Base model element mesh

Table 20 - Base Model Parametric Mesh

	Parametric Volume Meshing		
	X	Y	Z
Group 1	32	5	2 for Face sheet and 6 for Core
Group 2	32	10	
Group 3	10	10	
Group 4	12	10	
Group 5	32	10	

The base model with a bushing was meshed and can be seen in Figure 129. Everything but the bushing geometry has the same element mesh as before. The bushing material property that is used for the mesh is the bronze material property that is standard in the software. The two outside volumes, or the top volumes, have a

parametric mesh of ten elements in the X direction and a parametric mesh of ten elements in the Y direction. The two inside volumes, or the bottom volumes, have a parametric mesh of six elements in the X direction and ten elements in the Y direction. In the Z direction of the mesh, there are a total of 10 elements for the whole geometry. The mesh had a total of 28,800 elements, the bushing only adding 3,800 elements to the base model.

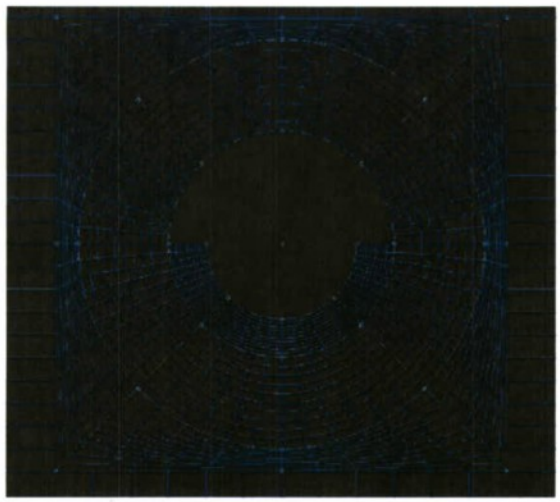


Figure 129 - Base model with bushing element mesh

The base model with a bushing and a fastener was meshed, which can be seen in Figure 130. Everything but the fastener geometry has the same element mesh as in the previous cases. The fastener material property that is used for the mesh is the steel material property that is standard in the software. The two outside volumes, or the top volumes, has a parametric mesh of ten elements in the X direction and a parametric mesh of six elements in the Y direction. The two inside volumes, or the bottom volumes, has a parametric mesh of six elements in the X direction and ten elements in the Y direction. In the Z direction of the mesh, there are a total of 10 elements for the whole geometry. The mesh had a total of 30,000 elements, the bushing only adding 3,800 elements to the base model. The fastener added 1,200 elements to the model with the bushing and added 5,000 elements to the base model.

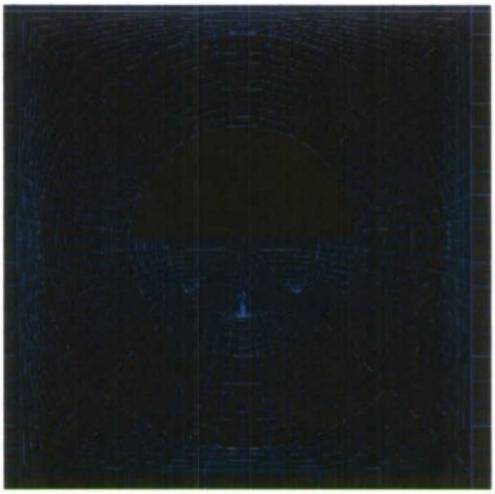


Figure 130 - Base model with bushing and fastener

The DAD models have the same mesh for the sandwich plane and the same reference axes previously mentioned. The only difference in the meshing between the DAD models and the base models is the number of elements in the Z direction; this can be seen in Figure 131. The number of elements for the face sheet in the Z direction remains the same with having two elements. The DADs sections have 3 elements in the Z direction. For the sections of the core that is between the face sheets without the DADS, the number of elements in the Z direction was changed to 25. For the sections of the core that is between the DADs have 19 elements in the Z direction. The sections of the core between the DADs have only 19 elements because the remaining six elements come from the DADs. Having four times the number of elements in the Z direction greatly increases the total element count of the model. The DAD model has a total of 73,420 elements for the whole model. This is a huge increase from the base model, which had a total of 25,000 elements. An increase of 48,420 elements is mostly in the core of the model. This increase of elements does not affect the results. Doing this procedure would actually make the analysis more accurate but it would take a lot longer to run the analysis than the base model. The main key for this analysis is that the same meshing method is used for the DAD models as with the bushing and fastener.

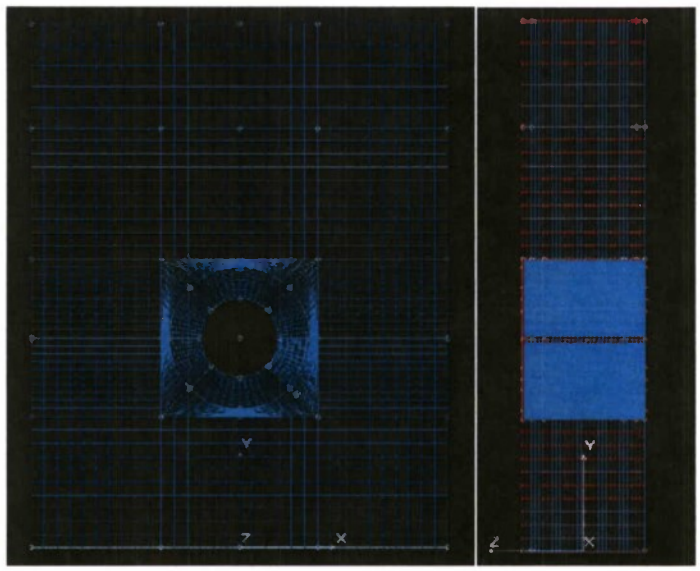


Figure 131 - DAD model element mesh

The element meshing scheme for the DAD models with the bushing and fastener would remain the same as before. The X and Y direction meshing scheme is the same accept the Z direction differs by having more elements like the base DAD model. The DAD model with a bushing had a total of 81,200 elements. This is an increase of 7,780 elements with the addition of the bushing. The DAD model with the fastener had a total of 88,160 elements, which corresponds to an increase of 6,960 elements.

Finite Element Model Static Loading/Displacement

The loading forces and the displacement constraints were added to the geometry. The displacement constraint was added on the bottom surface of the geometry to represent the symmetry of the specimen, which can be seen in Figure 132. Since the FEM geometry is a half model of the test specimen, it can be assumed that on the symmetry plane of the test specimen, the specimen will not move in the X, Y, Z directions and will have no moments acting on the faces. Displacement constraints are necessary so deformation displacements can be created from the loads applied on the test specimen. Without displacement constraints, the specimen would move as one piece and not deform. The displacement constraints are be applied to all FEM models, including the models with and without DADs, at the bottom surface.

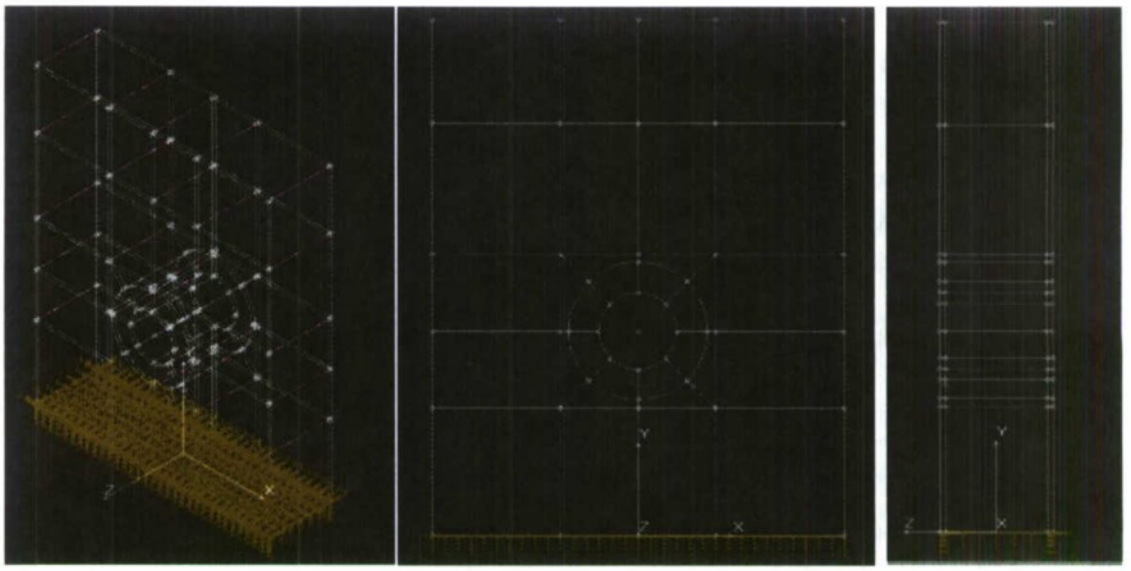


Figure 132 - Base Model Displacements

The loading on the specimens were then added to the model; this can be seen in**Error! Reference source not found.**. The loading was applied on the bottom surfaces of the cut-out hole section of the test specimens. The loading applied is a pressure force, with a pressure of -2,891 psi, which corresponds to a monotonic load of -940 lbs. The loading for specimens with a DAD has a pressure force of -5081 psi, corresponding to a monotonic load of -1,652 lbs. This applied load is equivalent to the yield load found during experimental testing. The pressure is applied in the vertical or Y direction to try to capture the true loading seen on the test specimen. The overall displacement/extension of the FEM can then be compared to experimental results to see the accuracy of the model.

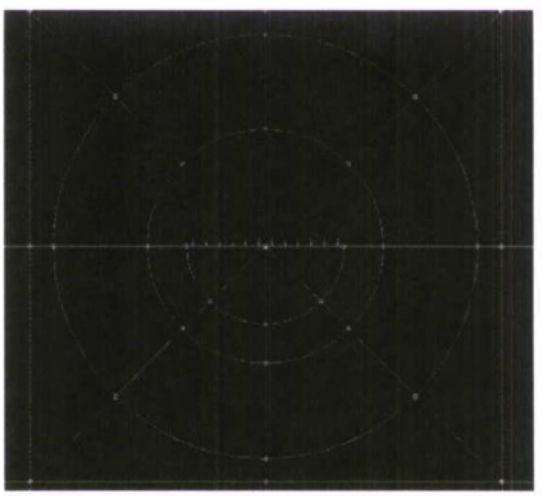


Figure 133 - Base model with bushing and fastener

From the results obtained for the pressure loading, the results for the bushing and fastener models gave very small displacement results from the experimental testing. The results obtained were all a power of ten

smaller than the experimental value; this will be discussed further in Finite Element Model Static Results. The decision that was made was to change the loading on each of the models from a pressure loading to a force loading. Instead of applying the loads along the bottom surface of the hole, the loads are applied to the center line of the hole; this can be seen in Figure 134. The force applied is the force obtained from experimental testing, which is -940 lbs for non-DAD models and -1652 lbs for DAD models. This loading is applied to the remaining models to see if there is any significant improvement in the results of the models.

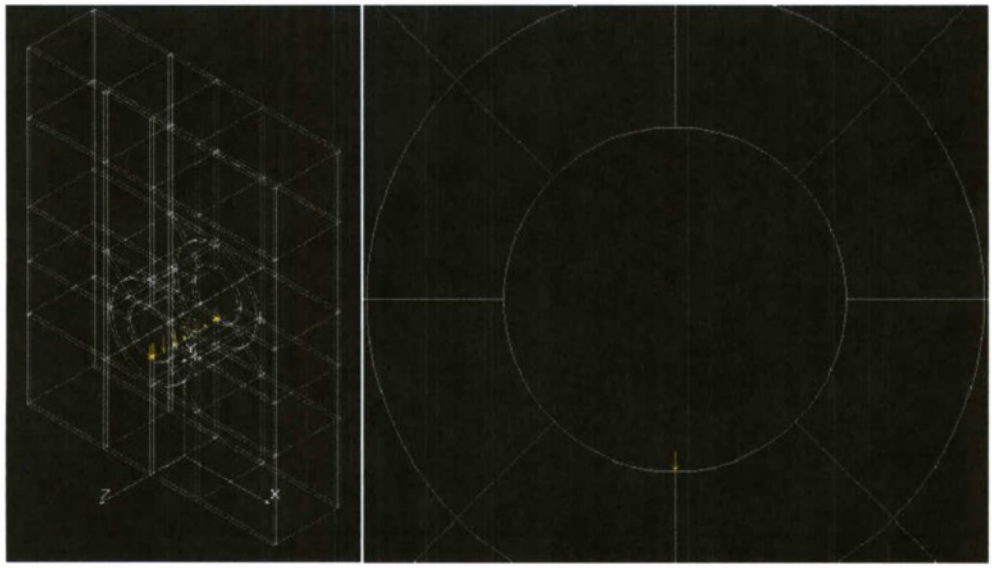


Figure 134 - Force Loading on FEM Base Model

Finite Element Model Static Results

The first analysis was the base FEM model for the control group. The displacement response of the model can be seen in Figure 135. The stress response of the model can be seen in Figure 136. The model shows that nothing is happening with the face sheets in both the displacement and stress figures. Most of the movement in the model is occurring in the foam. The foam is being displaced from the constant pressure and is pushing on the face sheet. This assumption can be backed up with the upward deformation of the foam seen in the displacement figure. The foam should be moving downward but instead is sliding up the face sheet. The foam cannot penetrate through the face sheet because it is a much stronger material than the core.

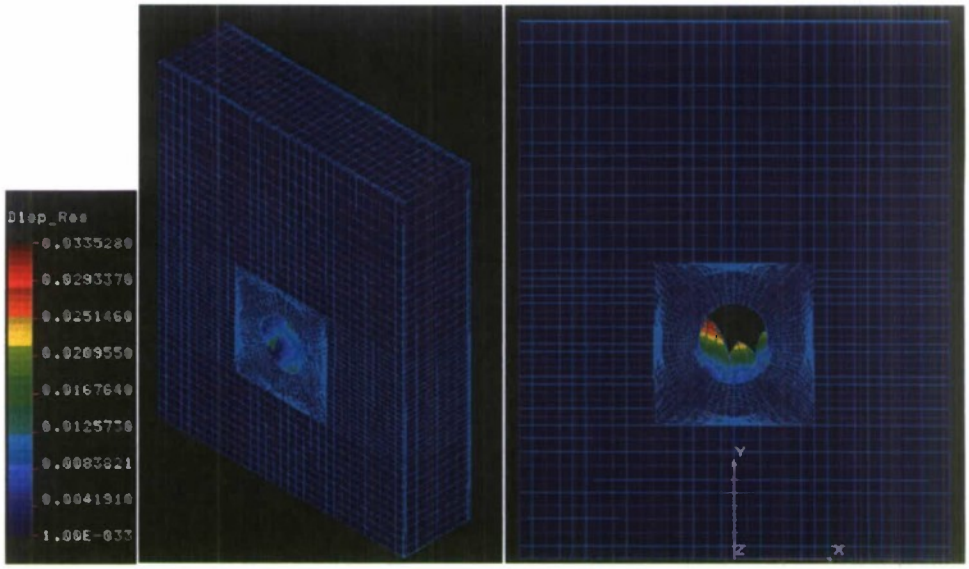


Figure 135- Displacement response of FEM base model with pressure loading

The stress distribution of the model shows that a circular stress concentration is seen at the bottom of the hole as well as around the hole. The stress distribution at the bottom of the hole could be from the stress of the core pushing out into the face sheet. An interesting stress concentration can be seen on the top corners of the meshing configuration on top of the hole. This could be due to how the mesh that was created around the hole. The elements might not be transferring the loads properly at that area. Another reason the stress concentration is appearing is that there is a counter reaction of the face sheets and the core from the pressure loading.

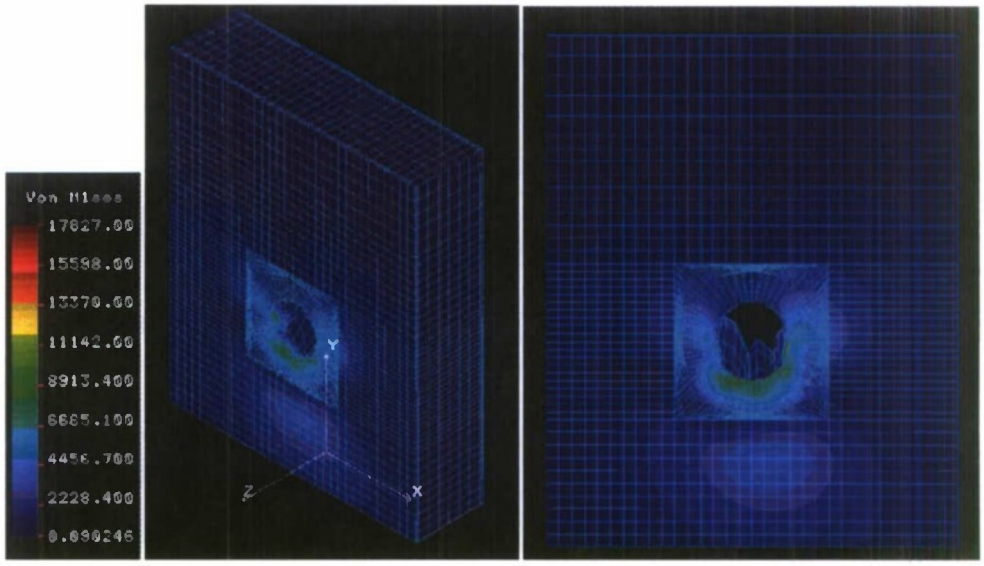


Figure 136- Stress response of FEM base model with pressure loading

The displacement response of the model can be seen in Figure 137. The stress response of the model can be seen in Figure 138. The displacement response of the model resembles more on what happened during

experimental testing. The section below the hole is actually displaced in the planar direction of the model, showing a bulge on the model. The bushing prevented the core from moving upward, like the previous model. Instead the core was forced to move downward and push out into the face sheet to create the bulge on the model. A semi-circular displacement distribution is also seen at the bottom of the hole from the loading seen from the bushing. The addition of the bushing severely affected the displacement of the model, even though the failure mode is similar to experimental results. The displacement contour bar shows that the maximum displacement seen in the model was 0.00278 in, which is a complete order of magnitude below the actual results. It seems that the bushing is dissipating too much of the loading seen on the model. The actual sandwich composite panel could be seeing a lot less force than what is applied.

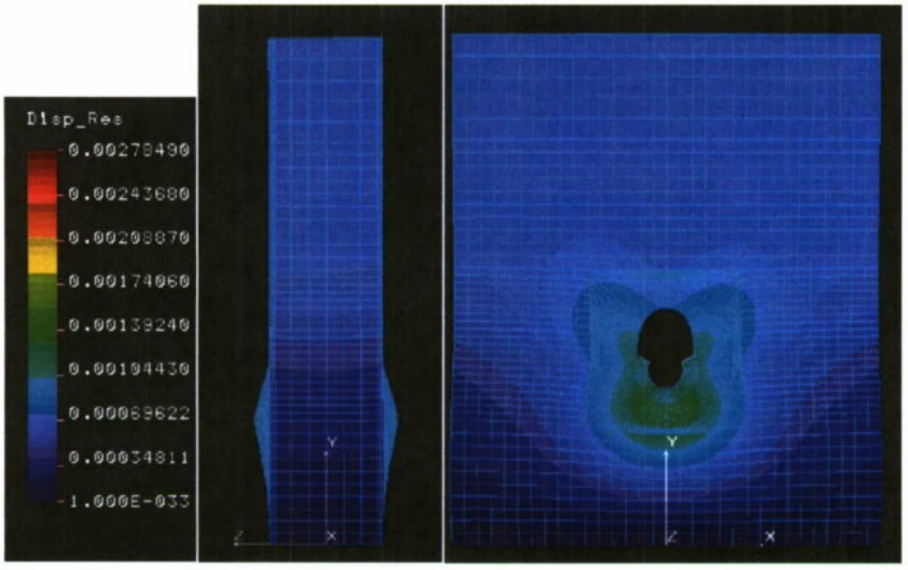


Figure 137 - Displacement response of FEM base model with bushing and pressure loading

The stress distribution of the model with the bushing is similar to the base model. A more pronounced stress concentration is seen at the bottom of the hole than before. Instead of a circular shape, the stress concentration at the bottom of the hole stretches from the bushing all the way to the bottom of the model. This backs up the assumption made earlier that the core is pushing along the face sheet at the bottom of the hole. With the addition of the bushing, the stress concentration was transferred onto the face sheet. The stress concentrations above the hole has disappeared, but new stress concentrations are seen at the tip of the bushing and the composite sandwich panel.

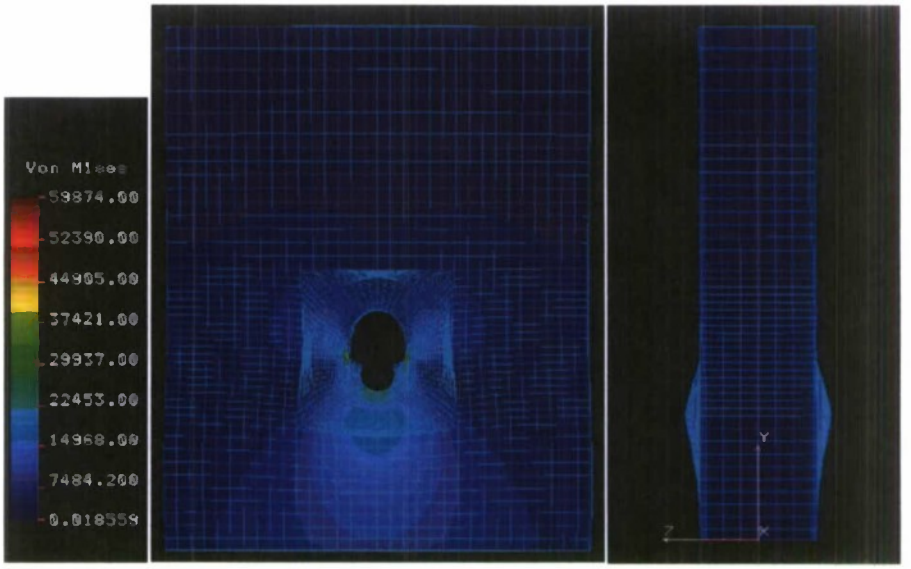


Figure 138 - Stress response of FEM base model with bushing and pressure loading

The displacement and response of the model can be seen in Figure 139 and 142. The displacement response of the model is similar to the model with the bushing. The model has more of a pronounced displacement under the bushing and the fastener. The model also has a similar bulge at the bottom of the hole appearing on the face sheet. An interesting observation that was captured in the model was that the addition of the fastener only decreased the maximum displacement seen on the specimen by half compared to the model with the bushing. The maximum displacement from the model is still an order of magnitude below the experimental results showing that the fastener and the bushing are still dissipating the force seen on the composite sandwich panel.

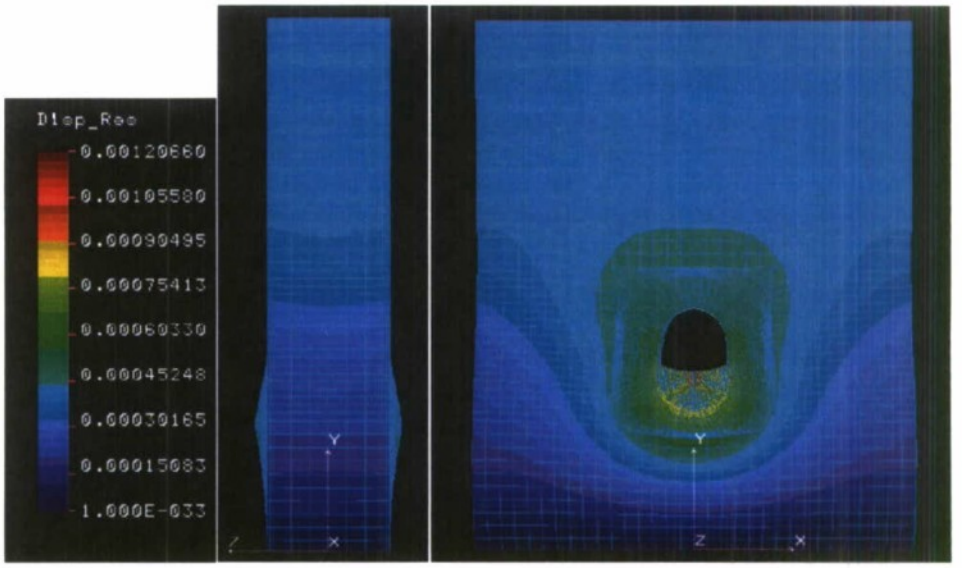


Figure 139 - Displacement response of FEM base model with bushing, fastener and pressure loading

The stress distribution of the model with the bushing and fastener is similar to the bushing only model. The model still contains a more pronounced stress concentration as seen at the bottom of the hole. Instead of a circular shape, the stress concentration at the bottom of the hole stretches from the bushing all the way to the bottom of the model. These facts further back up the assumption made earlier that the core is pushing along the face sheet at the bottom of the hole. With the addition of the bushing and the fastener, the stress concentration was actually able push out onto the face sheet.

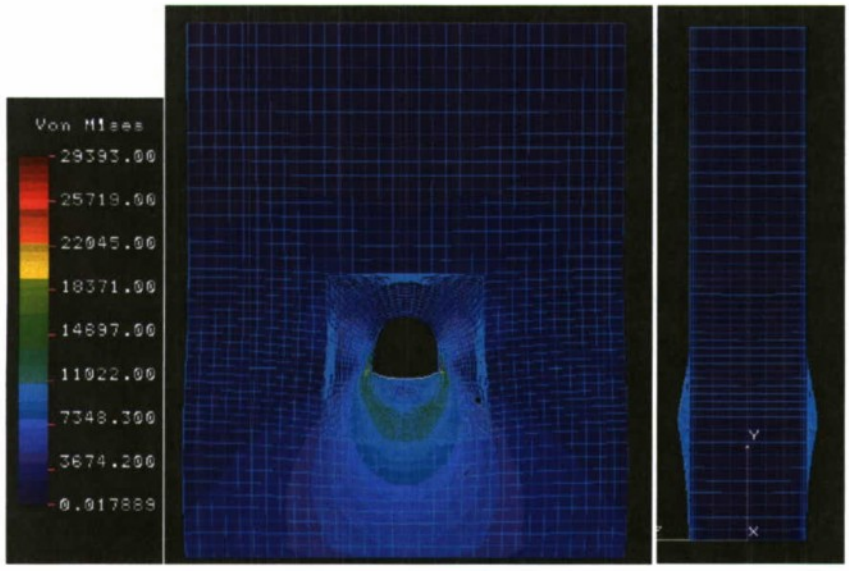


Figure 140 - Stress response of FEM base model with bushing, fastener and pressure loading

The displacement and stress response of the model can be seen in Figure 141 and 144 respectively. The base model for the DADs shows less displacement along the face sheet of the composite sandwich panel. The model does show that there is an elongation of the core. The increase in force due to the increase in stiffness of the DADs creates a large effect on the core. Since the core is a lot weaker than the face sheet, the model shows that the core is being easily pushed down the panel, showing the elongation of the core in the model. The displacement contour bar shows that the maximum displacement is three times larger than the displacement observed from experimental testing. A little deformation is seen at the bottom of the hole, but is not as pronounced as the previous models. The little deformation could be due to the additional stiffness added from the DADs. This slightly resembles the failure results from the experimental testing.

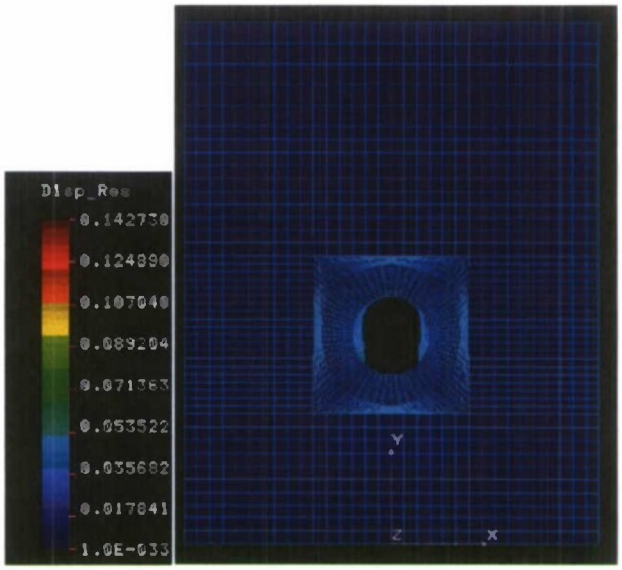


Figure 141 - Displacement response of FEM DAD base model and pressure loading

The stress distribution of the model shows a far different distribution from the non-DAD models. The first significant difference is that the model shows a stress concentration all along the DAD, running along the width of the model. The whole DAD shows that stresses are seen along the body and that the stresses are higher where the DAD and the core meet. Another stress concentration is also seen at the bottom of the model. This attribute could be a counter reaction from the boundary conditions placed on the bottom of the model. This phenomenon can be attributed from the core being pushed down more than the previous models creating a high stress concentration.

Figure 142 - Stress response of FEM DAD base model and pressure loading

The displacement response of the model can be seen in Figure 143. The stress response of the model can be seen in Figure 144. The displacement response shows a far different deformation from the base DAD model.

The first significant difference is that the model has a more pronounced bulge at the bottom of the hole than all the other models. The side profile view shows that the model's deformation resembles of a vase with the thickness of the model gets smaller at the hole and becomes larger at the base. The largest displacement is seen right below the model, creating a circular shape displacement field. The maximum displacement from the model is still an order of magnitude below the experimental results showing that the bushing is still dissipating the force seen on the composite sandwich panel.

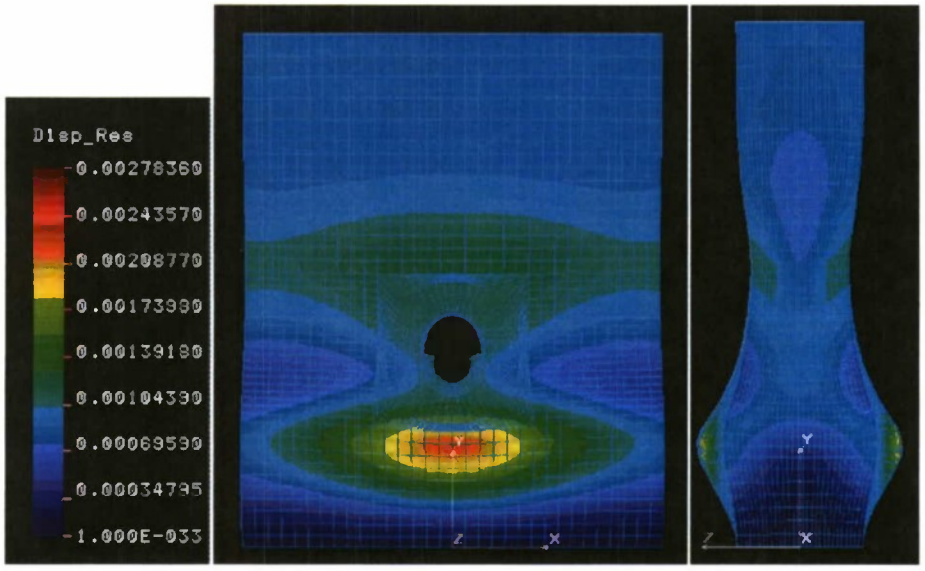


Figure 143 - Displacement response of FEM DAD base model with bushing and pressure loading

The stress distribution of the model is more similar to the stress distribution of the non-DAD models but greatly differs from the base DAD model. The stress distribution shows a stress concentration seen along the bottom of the model. Like before, this could be a counter reaction from the boundary conditions placed on the bottom of the model. It is created from the core being pushed down more than the previous models creating a high stress concentration. The model also shows a stress concentration directly at the bottom of the hole but ends at the edge of the DAD. This is an interesting behavior because the figure shows that there is no stress transfer where the DAD and the core meet. This is far different from the previous model where stress concentrations are seen along the DAD.

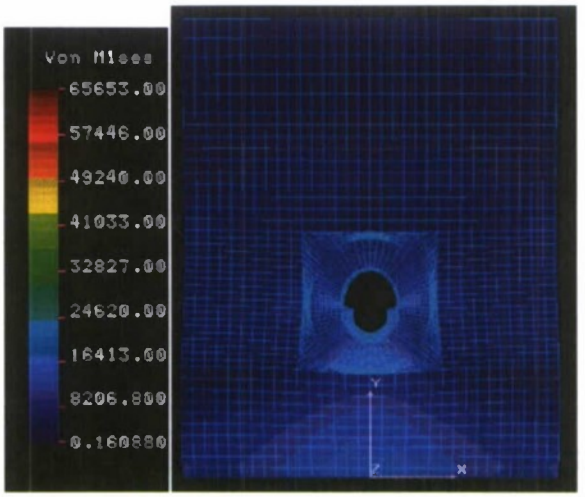


Figure 144 - Stress response of FEM DAD base model with bushing and pressure loading

The displacement response shows a similar response to the DAD base model with a bushing. The model has a similar bulge at the bottom of the hole than all the other models. The side profile view shows that the model's deformation resembles of a vase with the thickness of the model gets smaller at the hole and becomes larger at the base. The largest displacement is seen right below the model, creating a circular shape displacement field. The displacement distribution extends vertically around the hole of the composite sandwich panel making a distorted anvil shape. The maximum displacement from the model is still an order of magnitude below the experimental results showing that the bushing and fastener is still dissipating the force seen on the composite sandwich panel. This model shows a similar trend from the non-DAD models where the maximum displacement is only half the displacement seen on the DAD model with a bushing.

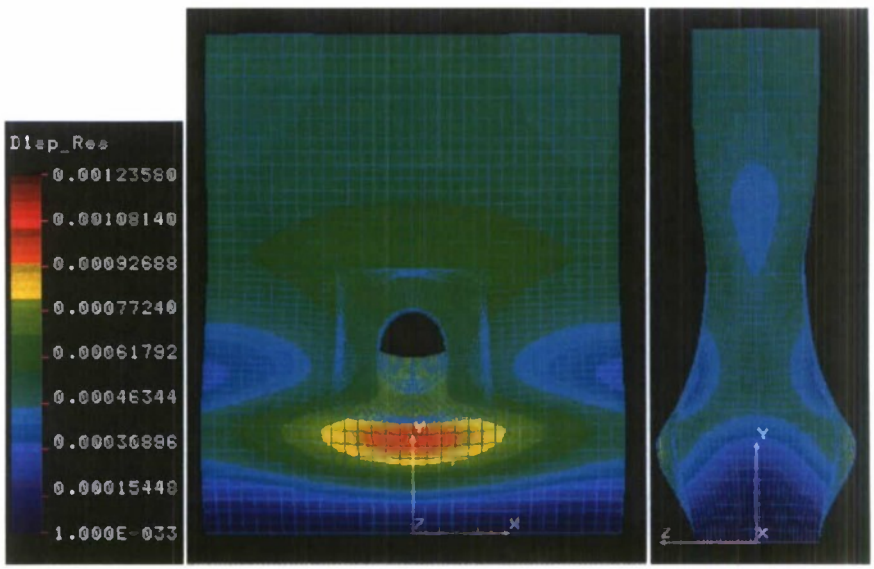


Figure 145 - Displacement of FEM DAD base model with bushing, fastener and pressure loading

The stress distribution of the model resembles a similar distribution to the bushing model. The stress distribution shows a stress concentration seen along the bottom of the model. Like before, this could be a counter reaction from the boundary conditions placed on the bottom of the model. It can be mostly attributed to the core being pushed down more than the previous models creating a high stress concentration. The model also shows a stress concentration directly at the bottom of the hole and expands slowly to the edge of the DAD. Instead of the previous model, the stress concentration does not stop but expands all the way to the bottom of the model. This shows that the model needs a better mesh between the fine mesh around the hole and the DAD and a coarse mesh at the bottom. A better transition in the mesh would give a smoother stress concentration at the bottom of the hole.

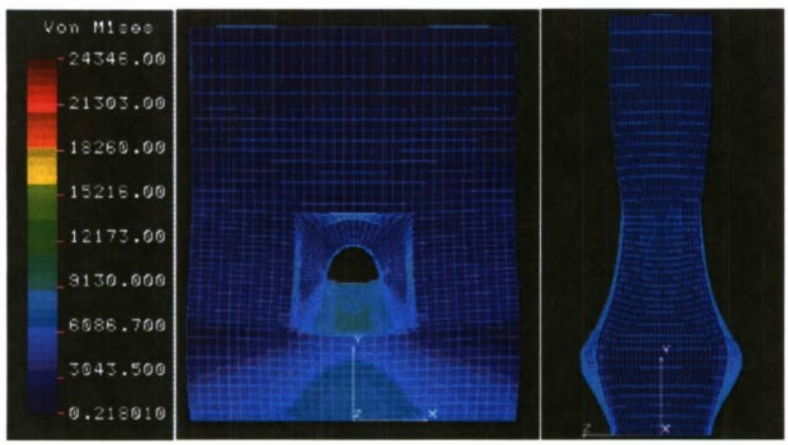


Figure 146 - Stress response of FEM DAD base model with bushing, fastener and pressure loading

A summary of the displacement results for all six models with and without DADs can be seen in Table 21. The table shows that the base models for each group is in the same order of magnitude of numbers, while the base with the bushing and the base with the bushing and the fastener are one order of magnitude less than the base group. An interesting observation is that the base model with the bushing and the base model with the bushing and the fastener are the same for each group. This could be due to the dissipation seen in the metals where it contacts the composite sandwich panel. A more accurate loading must be applied for better results. The comparison between the FEM and the experimental results would be further discussed.

Table 21 - FEM displacement results with pressure loading

		Extension (in)
Control Group (No DAD)	Base	0.0335
	Base/Bushing	0.0028
	Base/Bushing/Fastener	0.0012
DAD Group	Base	0.1473
	Base/Bushing	0.0028
	Base/Bushing/Fastener	0.0012

To obtain better results, a force loading was applied on the center line of the hole. The first analysis that was conducted was the base model with no DADs with a force loading. The displacement response of the model can be seen in Figure 147. The stress response of the model can be seen in Figure 148. The displacement model shows that the displacement on the model is on an order of magnitude of seven, being nine whole orders above the experimental results. The displacement distribution also shows that parts of the core and face sheet have complete fractured and is hanging over the hole. This was totally unexpected and something happened during the analysis that would cause a complete failure of the specimen. This could be mainly due to the loading applied on the specimen. Instead of a pressure distribution which made sure that the yield loading was evenly distributed along the surface, the yield load was applied to each node along the center line. The displacement and stress distribution shows that there are no effects to the core or the face sheet. From the previous models, there was bulge present at the bottom of the hole or the core being distorted. The results for this case are be omitted for analysis due to the obscure and unexpected results.

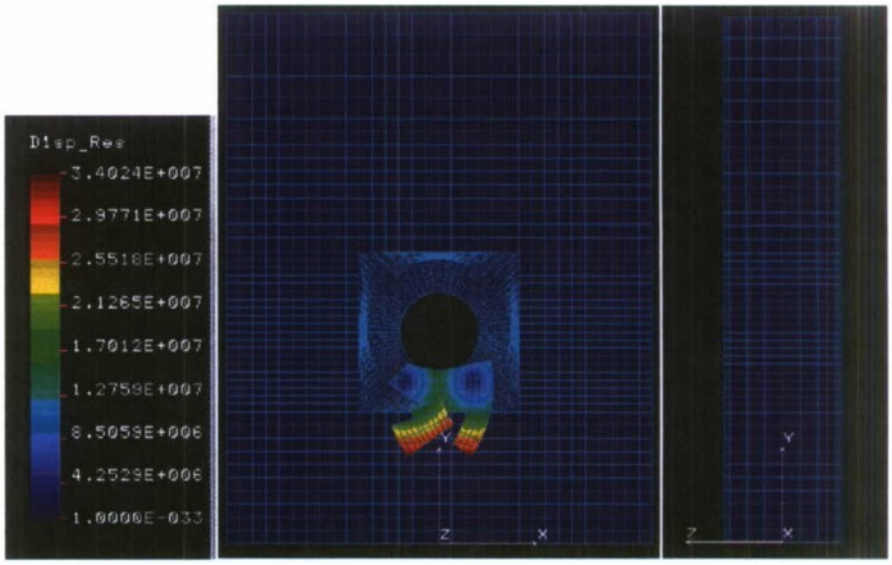


Figure 147 - Displacement response of FEM base model with force loading

The stress distribution shows that there are not many stress concentrations seen on the face sheet of the composite sandwich panel. The stresses seen on the model could be coming more from the core than the face sheet. Since a direct force is applied on the core, the core must be deforming within the panel causing the failure in the model.

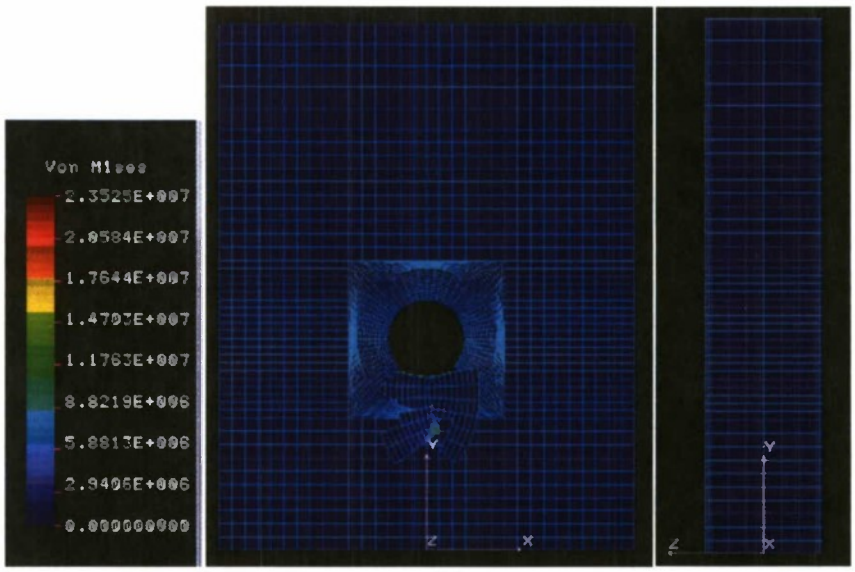


Figure 148 - Stress response of base model with force loading

The displacement response of the model can be seen in Figure 149. The stress response of the model can be seen in Figure 150. The displacement model more closely resembles the base model with bushing with pressure loading. Both displacement models have similar contour curves on the model, showing that regardless of the loading, the bushing would apply a similar distribution on the specimen. The main difference between the two models is that there is a sharp indent on the bushing caused by the loading on the centerline of the hole. The model also has a similar bulge at the bottom of the hole. The great advantage to this model is that the maximum displacement of the specimen is in the same order of magnitude as the experimental results. The displacement distribution also shows that the maximum displacement of the model occurs in the core.

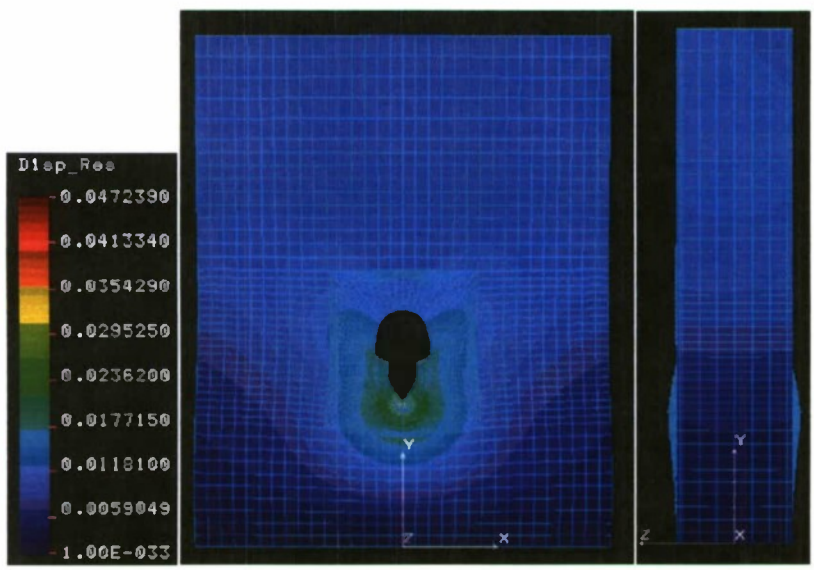


Figure 149 - Displacement response of FEM base model with bushing and force loading

The stress distribution of the specimen shows that the use of force loading on the model does not create any stress concentrations on the face sheet. It looks like most of the stress concentrations occur on the foam at the hole. The model shows that there is an increase in stress around the hole compared to the rest of the outer surface of the model.

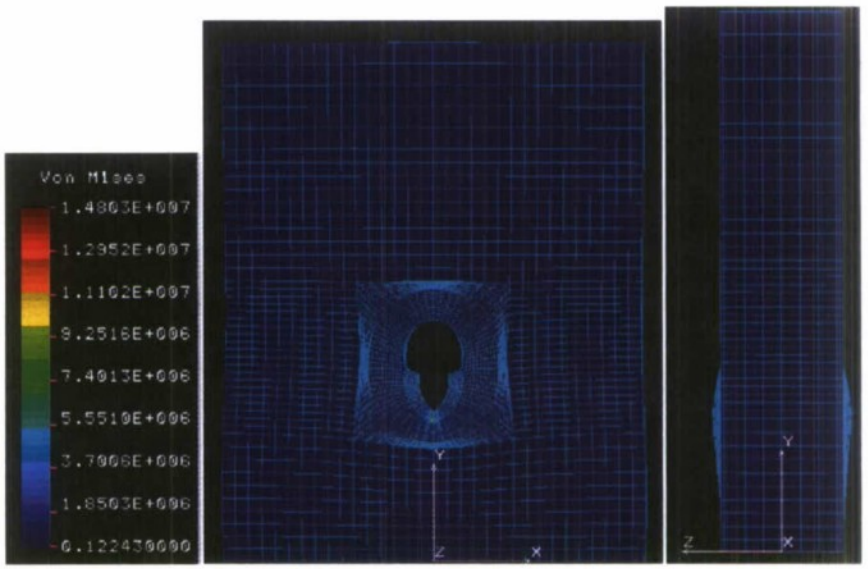


Figure 150 - Stress response of FEM base model with bushing and force loading

The next analysis for the force loading was the base model with the bushing and the fastener. The displacement response of the model can be seen in Figure 151. The stress response of the model can be seen in Figure 152. The displacement model more closely resembles the base model with the bushing and the fastener and pressure loading. Both models have similar contour curves on the model, showing that regardless of the loading, the bushing and fastener would apply a similar distribution on the specimen. The main difference between the two models is that there is a sharp indent on the bushing caused by the loading on the centerline of the hole. The model also has a similar bulge at the bottom of the hole. The bulge is less pronounced than previous models. The great advantage to this model is that the maximum displacement of the specimen is in the same order of magnitude as the experimental results.

The stress distribution of the specimen shows that the use of force loading on the model does not create any stress concentrations on the face sheet. It looks like most of the stress concentrations would be occurring on the foam at the hole. The model shows that there is an increase in stress around the hole compared to the rest of the outer surface of the model.

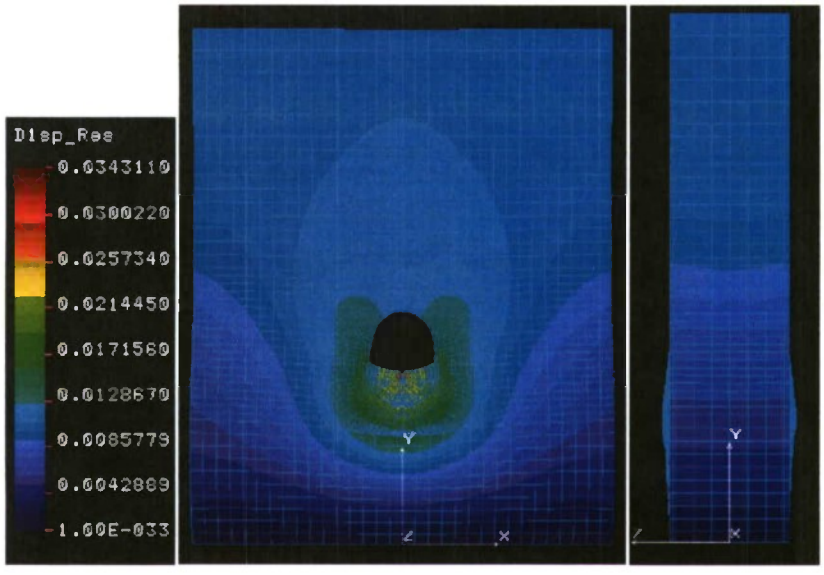


Figure 151 - Displacement response of FEM base model with bushing, fastener and force loading

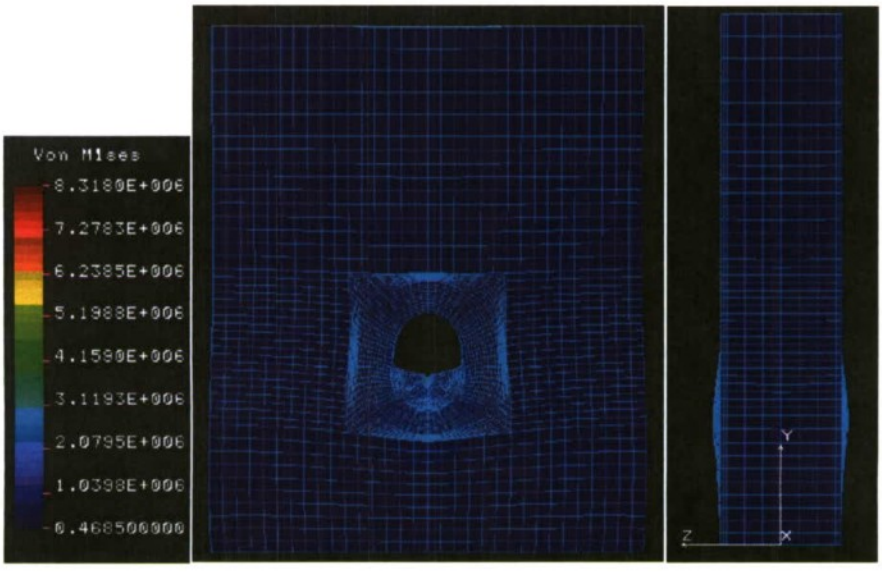


Figure 152 - Stress response of FEM base model with bushing, fastener and force loading

The next analysis for the force loading was the base model with DADs. The displacement response of the model can be seen in Figure 153. The stress response of the model can be seen in Figure 154. Like the base model with no DADs with force loading, there is no significant displacement seen on the outer surface of the specimen. There is no bulge or any deformation seen on the model. This could be due to the addition of the DADS adding infinite stiffness to the model. The lack of deformation or displacement could be due to only having the loading on the centerline of the bottom hole surfaces. The maximum displacement contour bar shows that the maximum displacement seen on the specimen is three orders of magnitude higher than the experimental testing. This shows that the loading is affecting the model in an unusual and unexpected way. The results for this

model are omitted from any analysis due to the high displacement value and the lack of any distortion on the model.

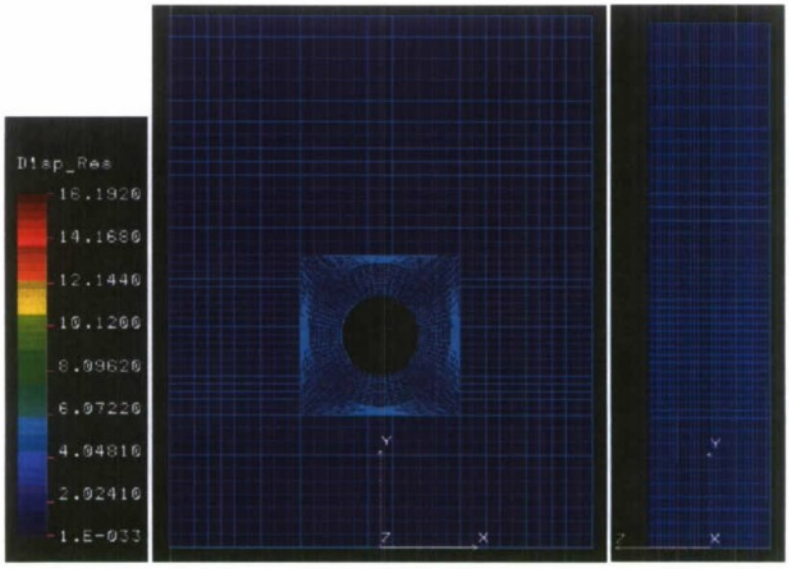


Figure 153 - Displacement response of FEM DAD base model with force loading

The stress distribution of the model shows no high stress concentrations on the model. The distribution does show an increase stress at the bottom of the model and around the hole. The slight increase in stress resembles the shape of a mushroom cloud. This increase in stress could be due to the counter forces from the boundary conditions placed on the bottom of the model.

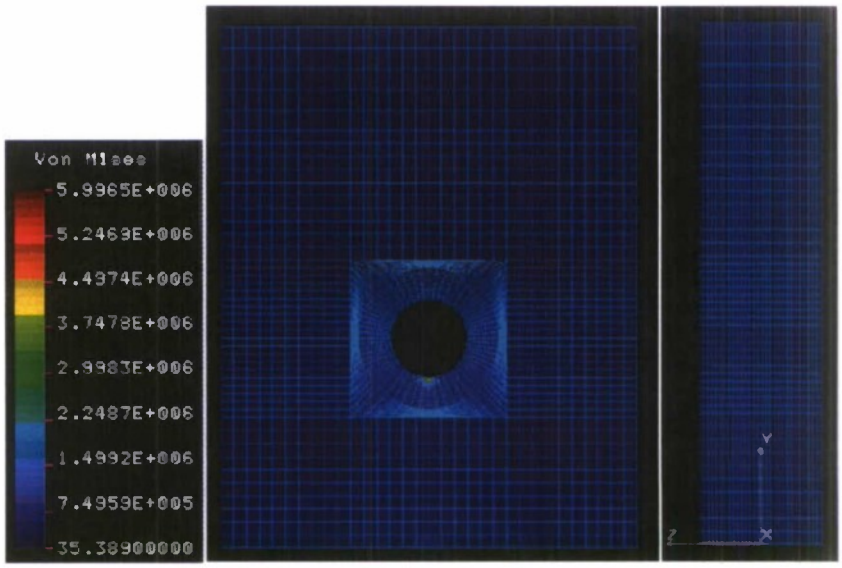


Figure 154 - Stress response of FEM DAD base model with force **loading**

The next analysis with the force loading was the base model with DADs with a bushing and a fastener added to the model. The displacement response of the model can be seen in Figure 155. The stress response of the model can be seen in Figure 156. The displacement response shows a similar response to the DAD base

model with a bushing and force loading. The model has a similar bulge at the bottom of the hole than all the other models. The side profile view shows that the model's deformation resembles of a vase with the thickness of the model gets smaller at the hole and becomes larger at the base. The largest displacement is seen right below the model, creating a circular shape displacement field. The displacement distribution extends vertically around the hole of the composite sandwich panel making a distorted anvil shape. The maximum displacement from the model is an order of magnitude above the experimental results showing that the bushing and fastener is surprisingly not dissipating the loading compared to the previous loading analysis with this model.

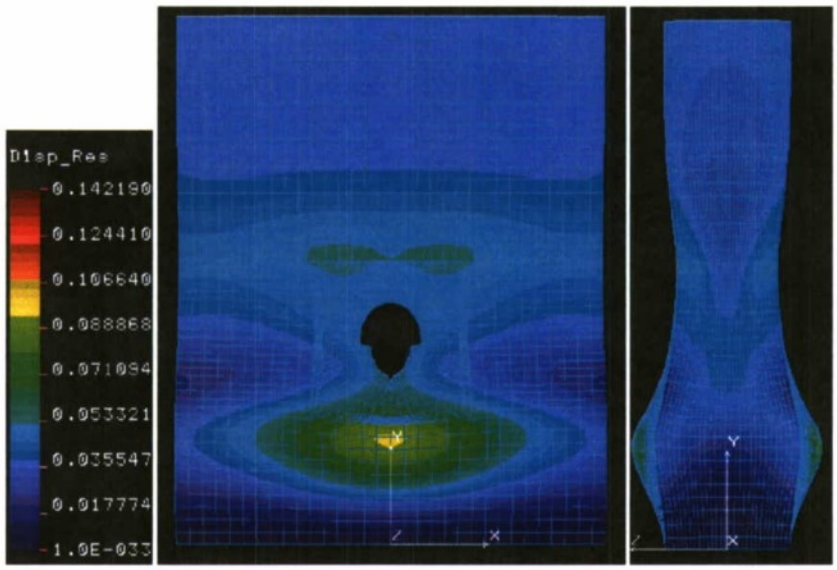


Figure 155: Displacement response of FEM DAD base model with bushing and force loading

The stress distribution of the models shows more of a stress concentration than previous models. The model has an increase stress concentration seen at the bottom of the panel. The slight increase in stress has a triangular shape with the tip of the triangle near where the DAD and the core meet.

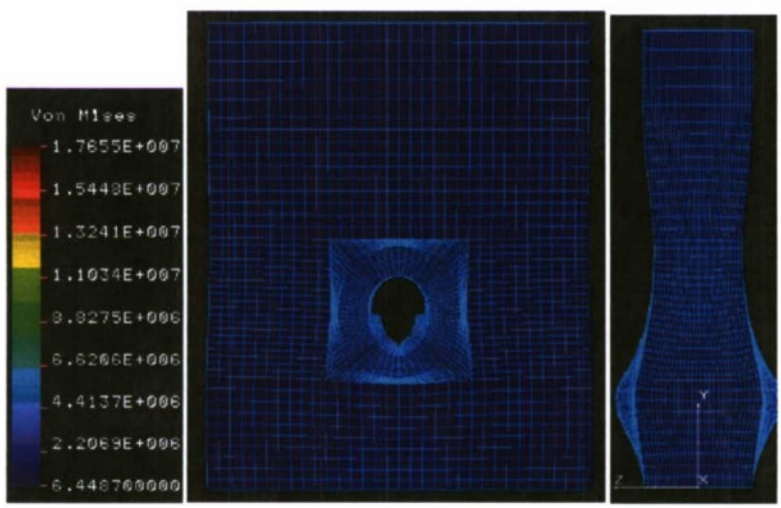


Figure 156 - Stress response of FEM DAD base model with bushing and force loading

The last analysis with the force loading was the base model with DADs with a bushing and a fastener added to the model. The displacement response of the model can be seen in Figure 157. The stress response of the model can be seen in Figure 158. The displacement response shows a similar response to the DAD base model with a bushing. The model has a similar bulge at the bottom of the hole than all the other models. The side profile view shows that the model's deformation resembles of a vase with the thickness of the model gets smaller at the hole and becomes larger at the base. The largest displacement is seen right below the model, creating a circular shape displacement field. The displacement distribution extends vertically around the hole of the composite sandwich panel making a distorted anvil shape. The maximum displacement from the model is the same order of magnitude of the experimental results.

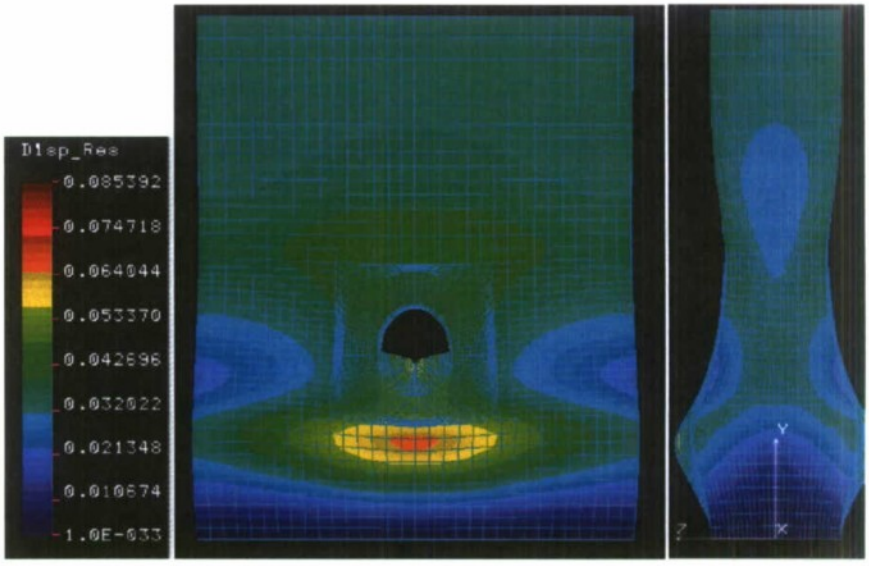


Figure 157 - Displacement response of FEM DAD base model with bushing, fastener and force loading

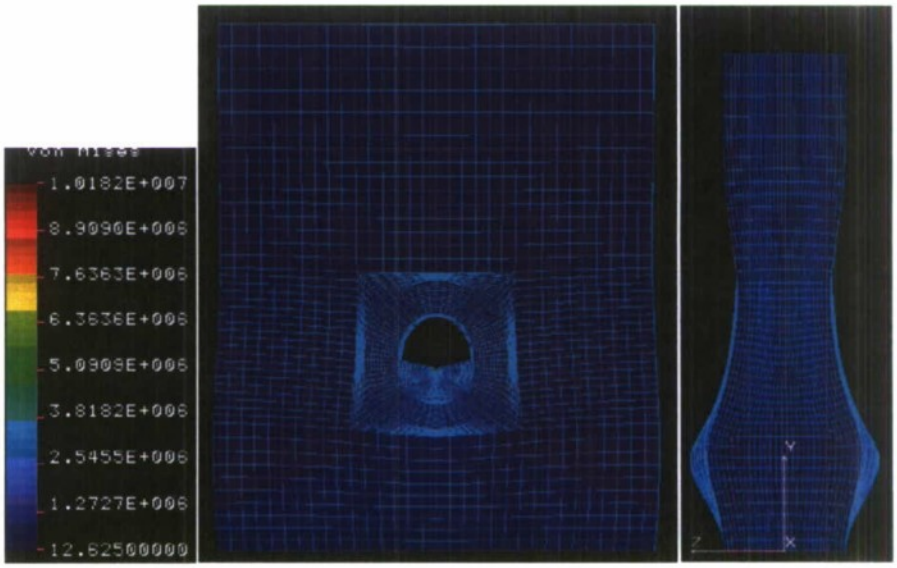


Figure 158 - Stress response of FEM DAD base model with bushing, fastener and force loading

A summary of the displacement results for all six models with and without DADs can be seen in Table 22. The table shows that the base models for each group are omitted to having too high displacements due to the unexpected and obscure results. The use of the force loading made the results for both groups more in line with the experimental results. The control groups, models without DADs, are pretty close to one another showing that the addition of the fastener dissipated the force seen on the panel compared to the bushing alone. The DAD group shows the same trend by having a smaller displacement with the addition of the fastener. The comparison between the FEM and the experimental results are further discussed.

Table 22 - FEM displacement results with force loading

		Extension (in)
Control Group (No DAD)	Base	N/A
	Base/Bushing	0.0472
	Base/Bushing/Fastener	0.0343
DAD Group	Base	N/A
	Base/Bushing	0.1422
	Base/Bushing/Fastener	0.0854

Chapter 6 Experimental and Numerical Comparison

6.1 Monotonic Loading Results (Eugene)

Figure 159 shows the failure stiffness ($\frac{lb_f}{in.}$) plotted against the vertical displacement from the experimental and numerical results for the test specimen without an initial delamination. The red curve plotted is the experimental data and the blue and green curves are numerical analysis results using two different material foam models. Both numerical modes closely follow the experimental results for displacements less than 0.015 inches but shortly afterwards the crushable foam models began to deviate from the experimental results and approach a stiffness value of approximately 175 $\frac{lb_f}{in.}$. The hyperelastic foam model did a better job of following the experimental data but slowly deviates from the experimental data as the vertical displacement continues to increase. The hyperelastic foam model provided stiffness results within approximately 16.7% difference compared to the experimental results.

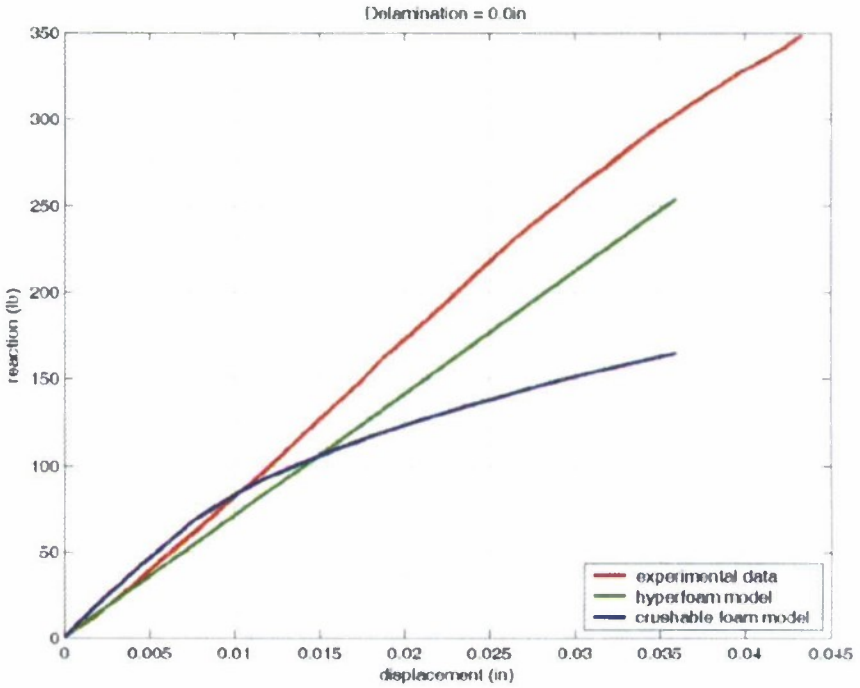


Figure 159 - Experimental/Numerical Analysis comparison (No initial delamination)

Figure 160 shows the failure stiffness ($\frac{lb_f}{in.}$) plotted against the vertical displacement for the test specimens with an initial delamination of 0.5 inch. The red curve represents the experimental data and the blue and green curves represent the numerical analysis results. The green curve represents the hyper foam model which is almost a purely elastic curve. The hyper foam model provided good stiffness results for displacement values less than 0.013 inches but afterwards over predicted the stiffness values. The crushable foam model yielded good results all the way up until the experimental results reach fracture. The crushable foam model provided stiffness values within approximately 5% compared to the experimental results.

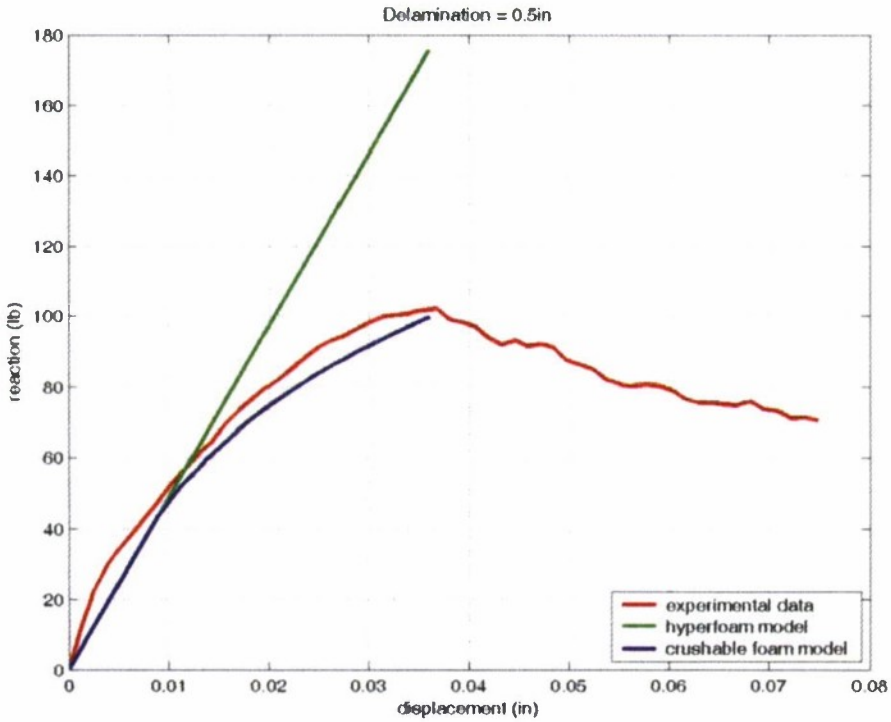


Figure 160 - Experimental/Numerical Analysis Comparison (0.5 inch initial delamination)

6.2 Damage Arrestment Device Results (Richard B.)

A comparison of the maximum displacement or extension of the composite sandwich panel between the experimental testing and the FEMs was conducted. The comparison of results can be seen in Table 23. The table shows the results from both the pressure and force loading with and without DAD for the FEM models. The results highlighted with yellow represents the most accurate in extension/displacement for the FEM model compared to its respective group. For the control group, non DAD models, the base model with a bushing and force loading was the most accurate by under predicting the experimental extension by 14.1% by having an extension of 0.0472 in. For the DAD group, the base model with a bushing, a fastener, and force loading was the most accurate by over predicting the experimental extension by 43.0% by having an extension of 0.0854 in.

Table 23 - Comparison of maximum displacement between Experimental and FEM Results

			Extension (in)	% Difference (-)
Pressure	Control	**Experimental**	**0.0550**	-

Loading	Group (No DAD)	Base	0.0335	-39.1%
		Base/Bushing	0.0028	-94.9%
		Base/Bushing/Fastener	0.0012	-97.8%
	DAD Group	**Experimental**	**0.0597**	-
		Base	0.1473	146.7%
		Base/Bushing	0.0028	-95.3%
		Base/Bushing/Fastener	0.0012	-97.9%
Force Loading	Control Group (No DAD)	**Experimental**	**0.0550**	-
		Base	N/A	-
		Base/Bushing	**0.0472**	**-14.1%**
		Base/Bushing/Fastener	0.0343	-37.6%
	DAD Group	**Experimental**	**0.0597**	-
		Base	N/A	-
		Base/Bushing	0.1422	138.2%
		Base/Bushing/Fastener	**0.0854**	**43.0%**

Some interesting trends from the table is that for both groups with pressure loading, the base model with a bushing and the base model with both the bushing and the fastener are a complete order of magnitude below the experimental results. This is primarily due to how the loading is transferred to the composite sandwich panel by the metal inserts. On the other hand, the base models for both groups with pressure loading were in the same order of magnitude with the experimental results.

Another interesting trend from the table is that for both groups with force loading, the extension decrease as the models become more complex. Adding a fastener to both groups' models decreases the extension. This reinforces the assumption that adding the bushing or the fastener to the model would dissipate the load seen by the composite sandwich panel. The last interesting trend is that for the force loading models, the control group under predicted the experimental results and the DAD group over predicted the experimental results. This could be due to the addition of the DADs and shows that another type of loading must be applied to see if better results can be obtained for that group.

Changing the loading on the specimen from a pressure loading to a force loading along the centerline of the bottom surface of the hole significantly affected the overall extension of the specimen. It moved the results from being one complete order of magnitude below the experimental results to within the same order of magnitude. The change to the new loading also gave better results, making the models ideal for future works.

Comparison of load/extension curve and Elastic Modulus

A comparison of the load versus extension curve and the elastic modulus for the theoretical, experimental, and FEM results was conducted. The first comparison was the control group without any DADs for the load versus extension curve; this can be seen in Figure 161. The figure shows five curves: the theoretical curve, the experimental curve, the FEM with force loading on the base model with a bushing, the FEM with force loading on the base model with a bushing and a fastener, and the FEM with pressure loading on the base model. The three FEM were plotted because those were the only models that had the same order of magnitude to the experimental or theoretical results. The FEM curves were created changing the loading of the model to find its respective displacement. Five data points were taken for each FEM model to create the curve. The theoretical curve was obtained using the force spring equation and solved for displacement from stiffness and the yield force. The experimental curve was plotted using the test specimen that had the closest to the average elastic modulus from that test group.

The FEM results have a linear curve that end at the yield load and its respective yield extension because the FEM software assumes that the model is running in the elastic region and its assuming a linear failure model. A more advance analysis or software must be used to have a more non-linear curve.

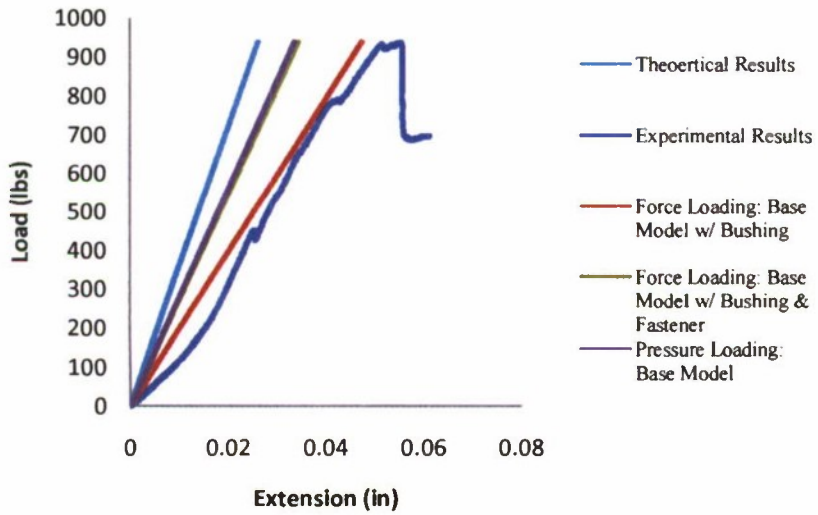


Figure 161 - Load/Extension Comparison for Control Group

From the figure, it shows that the force loading on the base model with bushing and the pressure loading on the base model have similar slopes to the experimental results. The force loading on the base model with a bushing has the closest yield extension to the experimental results but it has more of a gradual slope than the experimental results. The figure shows that the steeper the slope of the curve, the lower the yield extension from the experimental results. The elastic modulus of the curves can be seen in Table 24. The table shows that the theoretical elastic modulus of the composite sandwich panel is the highest. The elastic modulus was calculated by taking the slopes from the load versus extension curves, the slope is the stiffness (K), and is converted to

elastic modulus by the geometry of the ideal test specimen. The FEM with force loading on the base model with a bushing has the lowest elastic modulus but is the closest in maximum displacement compared to the experimental results. The FEM with force loading on the base model with a bushing and a fastener and the FEM with pressure loading on the base model have similar elastic modulus and extension.

Table 24 - Elastic Modulus Comparison for Control Group

	Elastic Modulus	% Difference from Theoretical	% Difference from Experimental
	(psi)	(-)	(-)
Theoretical Results	62,650	-	-
Experimental Results	45,174	-27.9%	-
Force Loading: Base Model w/ Bushing	34,573	-44.8%	-23.5%
Force Loading: Base Model w/ Bushing & Fastener	47,578	-24.1%	5.3%
Pressure Loading: Base Model	48,715	-22.2%	7.8%

The first comparison was the control group with DADs for the load versus extension curve; this can be seen in Figure 162. The figure shows four curves: the theoretical curve, the experimental curve, the FEM with force loading on the base DAD model with a bushing, and the FEM with force loading on the base DAD model with a bushing and a fastener. The FEM, experimental and theoretical curves were created the same method as previously discussed.

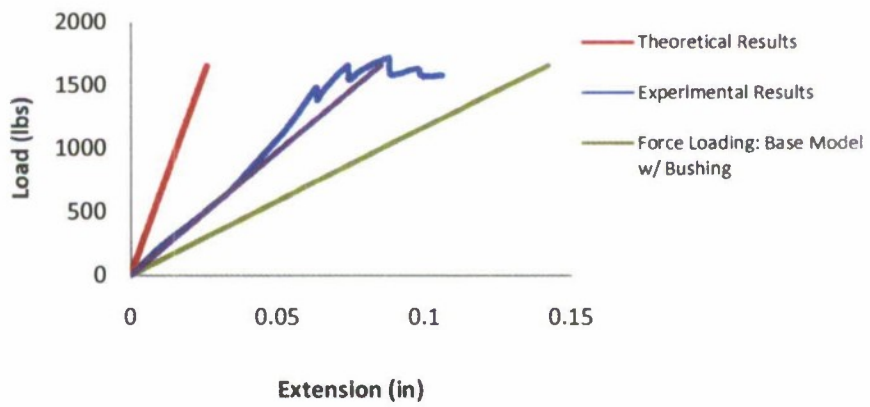


Figure 162 - Load/Extension Comparison for DAD Group

The figure shows that the theoretical results over predict the elastic modulus of the experimental results and under predicts the extension of the specimen. The FEM with force loading on the base DAD model with a bushing under predicts the elastic modulus of the experimental results and over predicts the extension of the

specimen. The most interesting curve is the FEM with the force loading on the base DAD model with a bushing and a fastener has a similar elastic modulus of the experimental testing by overlapping one another early on, but differs halfway through the experimental results. The model does over predict the extension compared to the experimental testing but is the closest of all FEM models. This is very surprising because the model captures the experimental results early on. The addition of a second linear curve to the model or a non-linear model would make the Fem more accurate to experimental results.

The elastic modulus of the curves can be seen in Table 25. The table shows a big contradiction to the curve by showing that the FEM with force loading on a base DAD model with a bushing and a fastener is 43% lower than the experimental values. This could be due to the elastic modulus of the experimental testing being calculated from the second slope of the curve rather than the first, where the FEM and the experimental results overlap one another. The two comparisons between the experimental and FEM results show that the FEMs need more refinement to be accurate to the experimental values.

Table 25 - Elastic Modulus Comparison for DAD Group

	Elastic Modulus	% Difference from Theoretical	% Difference from Experimental
	(psi)	(-)	(-)
Theoretical Results	62,650	-	-
Experimental Results	59,340	-5.3%	-
Force Loading: Base DAD Model w/ Bushing	20,168	-67.8%	-66.0%
Force Loading: Base DAD Model w/ Bushing & Fastener	33,583	-46.4%	-43.4%

Chapter 7 Conclusion

This research presents the experimental and numerical results for composite sandwich structures with an initial delamination. The effects of the initial delamination length were varied to see the effect an initial delamination had on decreasing the strength of the composite sandwich structures.

For the first application, fiberglass rods called shear keys were integrated into the initially delaminated sandwich structures with the intent of increasing the ultimate monotonic failure and fatigue life. The location of the shear keys with respect to the back edge of the delamination was varied to see the effect the shear keys had on the strength of the part. The fatigue life and fracture behavior was determined for sandwich structures with no initial delamination, 1 inch initial delamination, and 1 inch initial delamination with a shear key zero inches from the back edge of the delamination. A numerical analysis was performed using Abaqus/CAE and Cosmosm/M to verify the monotonic loading cases with no initial delamination and an initial delamination of 0.5 inches. Adding initial delaminations to the sandwich structure reduced the ultimate monotonic failure loads by approximately 70%. Increasing the delamination length had negligible effects on the ultimate monotonic failure load but a larger delamination corresponded to a larger yield force. This was attributed to the additional moment created from the Instron machine to the aluminum tabs. The shear key rods decreased the ultimate failure force of the sandwich structure compared to the test specimens with only an initial delamination. This was likely due to the stress concentration created by the shear key. The location of the shear key had a minimal effect on the overall strength of the composite sandwich structures. Shear keys located further away from the back of the delamination corresponded to a slightly weaker part than one with a shear key at the back edge of the delamination (0 inches). Test specimens with initial delaminations had a fatigue life approximately 75% less than test specimens with no initial delamination. The shear key test specimens had a fatigue life of approximately 80% less than the test specimens with no initial delamination. The numerical analysis was able to match the monotonic failure loads for the test specimens with no initial delamination to approximately 15% using a hyperelastic foam model. The numerical analysis was able to match the experimental failure loads for the test specimens with an initial delamination of 0.5 inches to within 5% using a crushable foam plasticity model.

For the second application of the buckling study, the results from the experimental testing of the composite sandwich showed following under buckling test.

Having a delamination arrestment key that spans the full anvil length and placed parallel to the load vector provides the best results where the majority of the overall failure load is maintained. For 1-inch delaminated 'SD' configuration test specimens, there is only an 11.8% drop in load capacity from a non-delaminated 'ND' composite sandwich structure and a 126% increase in load capacity from the 1-inch 'ID' configuration.

The failure modes of the experimental testing differ for the various configurations and delamination sizes. These failure modes range from localized buckling of the face sheet for the control '1D' configuration to arrestment of the 1-inch delamination for the 'KD' configuration to point load concentrated failure for the 'PL' configurations.

For the third application, this study presents the effect of damage arrestment devices to the failure mode around a hole in a composite sandwich panel. Seven different DAD thicknesses were tested under monotonic loading. Experimental and analytical tests were performed to determine the effects of varying the DAD thickness around the hole of a composite sandwich panel. Fatigue testing was also conducted on a control group without any DADs and a DAD group containing only a DAD thickness of three layers. The third experiment yielded results that showed that the introduction of a continuous arrestment key parallel to the in-plane loading and embedded into the composite sandwich structure provided a significant increase in loading and buckling capabilities. These results were the compared to the control test specimens with and without an initial delamination and with no embedded key. The continuous key placed parallel to the load vector increased the structural strength with an increase of 126% from a 1-inch delaminated structure and only an 11% drop from non-delaminated structures. That is, 1-inch and 2-inch delaminated structures showed a 61% drop and 81% drop from non-delaminated structures respectively.

Numerical analysis was done using a finite element modeling software to predict the failure of the composite sandwich panels and was compared with experimental results. Three main models were created for specimens with and without DADs. The three models include the composite sandwich panel, a composite sandwich panel with a bushing, and a composite sandwich panel with a bushing and a fastener. Two types of loadings were applied on the six models to see which would produce the most accurate results compared to experimental testing. The key conclusion that was determined from experimental testing and numerical analysis is summarized:

- The failure mode of the experimental testing starts from a pure bearing stress on specimens with small or no DAD thicknesses to a buckling failure with the face sheet delaminating for specimens with thicker DADs.
- The fatigue testing showed that the specimens with or without a DAD have similar life cycle curves and fatigue endurance limits. This shows that the DADs only strengthen the composite sandwich panel and do not prolong the failure of the panels.

The numerical analysis shows that the type of loading on the model and the configuration of the model greatly affects the results. It was discovered that a force loading on the centerline of the specimen was ideal and that the addition of a bushing or a bushing and a fastener made the results more accurate. The addition of a bushing or a fastener helped dissipate the loading seen on the composite sandwich panel and gave it an even distribution. The numerical analysis needs more work because the model only captures the elastic or linear portion of the experimental results. Better modeling of the core and the interaction between the DADs and the core must be further investigated.

Future Work (Richard Davis)

Three-Point Bend Testing of Sandwich Carbon-Composite Panels

Low shear strength and delaminating effects of the skin from the core material have been known to cause significant problems to the design of sandwich composites. In effect, the preceding portions of this research focus on the mitigation of these two prominent disadvantages through the introduction of shear keys, while the following proposal will focus on the high energy absorption abilities of sandwich carbon-composite panels. Sandwich composite panels integrated with shear keys will be subjected to high and low-velocity localized impact testing, followed by three-point bend testing, in order to corroborate the energy absorption advantages of sandwich composite panels. Numerical analysis will substantiate all experimental data; the FEA program COSMOS will be used.

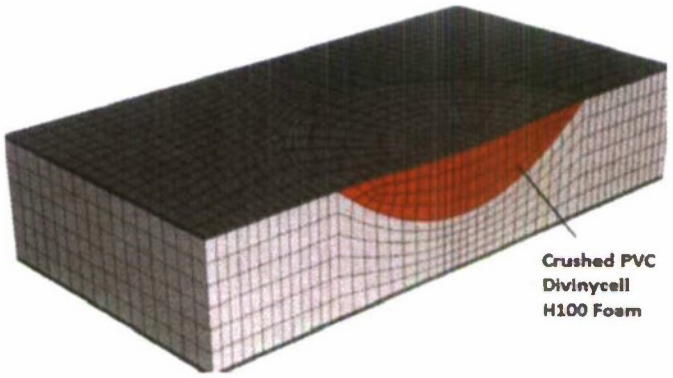


Figure 163 - A finite element model of the sandwich Composite panel after impact loading

Experimental/Numerical Testing

The sandwich composite panels will initially undergo a high and low velocity impact test. Seventy-five specimens will be subjected to the high velocity impact setting, and 75 specimens will be tested under the low velocity impact force. The skin is composed of resin pre-impregnated carbon fiber. The core material is composed of PVC Divinycell H100 foam.

After the impact loading, each sandwich composite panel will be subjected to a three point bend test, utilizing the servo-hydraulic machine in the Composites Laboratory at Cal Poly. Each specimen will be loaded until failure and the data will be recorded. Figure 164 delineates an experimental and finite element rendering of the three point bend test:

Figure 164 - An experimental and finite element rendering of the three point bend test

The sandwich composite panel will be analyzed using the finite element software COSMOS prior to the experimental portion of this proposal.

References

1. "Light Sandwich Constructions." CTiHuati Composites Co., Ltd. 2008. http://www.ctihuatai.com/Products-sandwichEN.htm
2. Hurwitz, Dr. Frances I., and Sullivan Dr. Roy M."Multifunctional, Foam Core, Ceramic Matrix Composite Integrated Structures Development." NASA. 14 Dec. 2007. http://www.grc.nasa.gov/WWW/RT/2006/RX/RX11C-hurwitz1.html
3. Bigelow, Chaterine A. "Effects of Fatigue and Environment of Residual Strengths of Center-cracked Graphite/Epoxy Buffer Strip Panels." March 1989.
4. Mitra, Dr. Nilajan and Kasper, Dr. Eric. "A novel technique for preventing delamination of the composite skin from the core: Sandwich panels with shear keys." 2008. pg. 1-15.
5. Mitra, Dr. Nilajan, and Jacobson, Michael. "Sandwich Composite Report." 2008. pg. 1 – 15.
6. Davis, Richard; Engels, Nathanial; Morham, Brett; Ung, Ryan. "Delamination of Fiberglass Sandwich Composites." 8 June 2009.
7. "Composites." NDT Resource Center. 2006. http://www.ndt-ed.org/EducationResources/CommunityCollege/Materials/Graphics/Composites.jpg
8. "Carbon Fiber Panel 12"x12"x0.015"/0.38mm, 2x2 Twill." Carbon & Fiberglass Sales. 2009. http://carbonsales.com/images/P/%2361052-50.jpg
9. Plastic World. 2001. http://www.plasticworld.ca/fiberglass.jpg
10. "Subaru Impreza 2.5 rs Kevlar/s-glass skid plate." Bar One Composites. 2008. http://www.baronecomposites.com/images/Kevlar.jpg
11. "Comparison of Theoretical and Experimental Analysis of Unidirectional and Cross-ply Composite Laminate Plates." 2007.
12. "Fibre-reinforced plastic." Wikipedia. 23 Oct. 2009. http://en.wikipedia.org/wiki/Fibre-reinforced_plastic
13. "Composite Material." Wikipedia. 24 Oct. 2009. http://en.wikipedia.org/wiki/Composite_material
14. "Fabrication Methods." Composites World. 8 Jan. 2009. http://www.compositesworld.com/uploadedimages/Publications/CW/Articles/Internal/SB09_compositesthematerials_h.jpg
15. "Vacuum Resin Transfer Molding (VARTM)." Kenway Corporation. 2009. http://www.kenway.com/pictures/vardtr.jpg
16. "Fatigue (material)." Wikipedia. 14 Oct. 2009. http://en.wikipedia.org/wiki/Material_fatigue
17. "Piston fracture." Hirth aircraft engine, diagnosing piston damage to your engine. 2006. http://www.ultralightnews.com/hirth/images/piston7.jpg

18. "Fatigue Testing." Stork Material Technology. 2009. http://www.storksmt.com/images/TechnicalServices/SMT/Failure/FatigueFrac400.gif
19. Pascoe, David. "Parallel Universe – Composite Troubles in Aircraft." Core and Structural Issues. 29 March 2005. http://www.yachtsurvey.com/composite_rudder-2.JPG
20. Spiegal, Peter. "F-15 fleet grounded after a jet falls apart." LA Times. 6 Nov. 2007. http://www.latimes.com/news/nationworld/nation/la-na-f156nov06,1,6807945.story
21. Kelly, Shawn. "Fatigue." Virginia Tech Materials Science Engineering. 4 May 1997. http://www.sv.vt.edu/classes/MSE2094_NoteBook/97ClassProj/anal/kelly/fatigue.html#tworef
22. "Vacuum Infusion Guide." Composites World. 2009. http://www.compositesworld.com/uploadedimages/Publications/CW/New_Product_Announcements/fibreglast-433.jpg
23. "Vacuum Infusion - The Process of Resin Infusion." Vacuum Infusion Technology. 2008. http://www.bladeoceancraft.com.au/assets/images/autogen/a_infusion-diagramAA.gif
24. "Divinycell H - High Performance Foam." DIAB Products. 2009. http://www.diabgroup.com/americas/u_products/u_divinycell_h.html

Atom transfer from a magneto-optical trap to an atom trap created by diffracted laser light

Principal Investigator:

Katharina Gillen,
Physics Department,
Cal Poly San Luis Obispo

1. Background Information:

The goal of this project is to demonstrate the transfer of ^{87}Rb atoms from a magneto-optical trap (MOT) to a light trap formed in the diffraction pattern behind a circular aperture. The trapped atoms can then be used to explore whether these traps are suitable for quantum computing. Quantum computation and information are currently widely investigated fields because of their applications in communication and decryption, as well as investigation of fundamental quantum physics. One of the systems identified as possible quantum computer are neutral atoms trapped by laser light. All necessary steps for quantum computing, including two-qubit gates, have been successfully experimentally demonstrated, except for one: Arranging over one million atoms in a way that still allows addressing individual atoms for quantum computation. It is this area we hope to advance with this project.

In our previous research we identified the diffraction pattern behind a simple circular aperture as a possible atomic light (or "dipole") trap. Calculations indicated that with modest laser power (~100 mW) atom traps appropriate for quantum computation can be created. Theoretically, an array of circular apertures would then also yield an array of such atom traps.

With the help of previous funding through the C3RP program, we were able to directly measure the intensity pattern behind a commercial pinhole to test if the diffraction pattern behind a real pinhole still retains all the features necessary for quantum computing. We found that indeed the intensity pattern is as we calculated. In addition, we used previous C3RP funding to construct all necessary parts of the MOT, which is used to pre-cool the atoms prior to transfer into the dipole trap. All work for the current project was towards demonstrating the transfer of atoms into the dipole trap.

During the course of this project, we finished construction of the magneto-optical trap and prepared the experimental setup for this atom transfer. We also completed computational work on the use of these dipole traps for quantum computing, including moving pairs of traps together and apart to facilitate two-qubit operations.

2. Project Goals:

The goals of this project were:

- Completion of MOT setup
- Progress towards atom transfer from MOT to diffraction dipole traps
- Completion of computational work on the polarization dependence of the diffraction dipole traps

3. Project Progress and Results:

The granted C3RP funding was used to finish construction of the MOT, prepare for atom transfer from the MOT to the diffraction dipole traps, and to finalize research and computations on the polarization dependence of the diffraction dipole traps.

The MOT system consists of a vacuum chamber, a pair of electromagnets, two tunable diode lasers, and an optical system that creates six opposing laser beams of appropriate polarization in order to slow down and collect atoms (see Fig. 1).

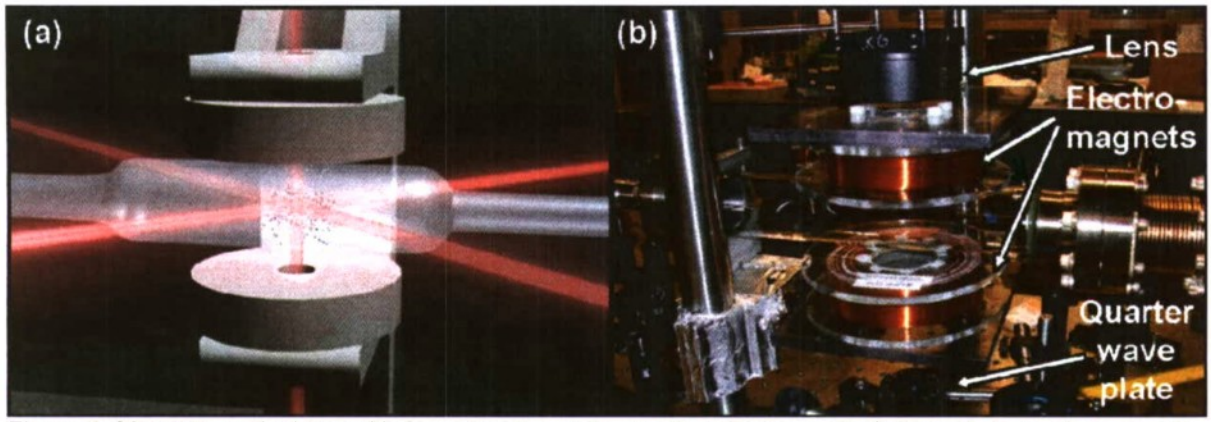


Figure 1: Magneto-optical trap. Six laser beams and a quadrupole magnetic field cool down atoms inside a vacuum chamber. (a) Animation of MOT (Picture courtesy of Kevin Gecsi, Ohio Supercomputing Center.) (b) Actual MOT setup. The vacuum chamber, the electromegnets, and some of the beam shaping and polarization optics are visible in this picture. Additional optics not in this picture have been installed to complete the MOT.

In order to trap atoms without any background gases expelling them from the trap, an ultra-high vacuum environment is needed. This is achieved by building a vacuum chamber from all stainless steel and glass parts, thoroughly cleaning all parts in a multiple-step process, and finally "baking" the chamber at a temperature of 300°C while pumping it out with a turbomolecular pump backed by a roughing pump. The vacuum chamber had been constructed using previous C3RP funding. This round of C3RP support was used to troubleshoot and repair the vacuum pumping station (turbopump, roughing pump, gate valve, thermocouple gauges, ionization gauge, relay box for emergency shutdown, Variacs for control of heater tapes), and perform the 10 day bakeout of the vacuum chamber at 300°C. The pressure in the vacuum chamber is known to be less than 10^{-8} Torr, and estimated to be no more than 10^{-10} Torr, more than adequate for atom trapping experiments.

Second, a pair of electromagnet coils is needed to create the quadrupole magnetic field that is necessary to collect the atoms into a small cloud at the center between the magnets. The electromagnets had been constructed and tested with previous C3RP funding, but are now mounted in place around the ultra-high vacuum chamber.

Finally, for the MOT to work, two tunable, frequency-stabilized diode lasers are needed. These lasers were previously constructed and tested, but always displayed poor temperature stability, which is detrimental for trapping atoms, as even minute changes in temperature change the laser frequency at the ~10 MHz level, disrupting the function of the atom trap. During the course of this research project, we were able to analyze and fix this problem. The laser design was such that too much material was

between the laser diode and the thermo-electric coolers (TECs) that were controlling the laser temperature. We redesigned the lasers (see Fig. 2) such that the TECs are now located as close as possible to the laser, and we now achieve the required temperature stability within seconds of making any changes to the laser settings.

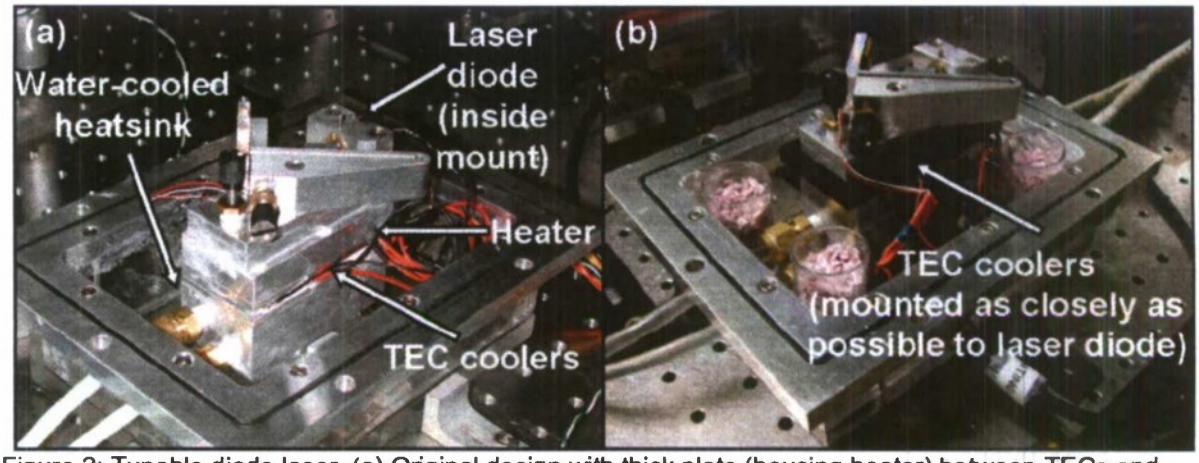


Figure 2: Tunable diode laser. (a) Original design with thick plate (housing heater) between TECs and laser diode. (b) Improved design, with TECs directly underneath laser baseplate to ensure temperature stability.

For the grating tuning and stabilization of the lasers, several photodetectors are needed. We constructed these photodetectors by adding home-built circuitry to a commercially available photodiode chip during previous research. However, some of the detectors had to be repaired due to some faulty wiring and malfunctioning of their photodiode chips. The same kind of photodetector is also used for detecting atoms trapped in the MOT and the diffraction dipole traps.

Six laser beams of equal intensity, directed into the vacuum chamber such that two opposing beams travel along each of the three dimensions of space are required for the atom trap (see Fig. 1). To increase the available intensity per beam (as well as for simplifying the optical setup), we split the laser into three beams, which are directed into the vacuum chamber along the three dimensions of space, and then each beam is retro-reflected back into the chamber by a mirror to create the opposing laser beam. In order for the laser beams to slow the atoms and collect them in the center of the vacuum chamber, the laser beams have to be circularly polarized and combined with the quadrupole magnetic field with zero field at the center. We have carefully measured the properties of the magnetic field and confirmed its quadrupole nature and optimal field gradient (8 G/cm). Quarter wave plates are in place to create the appropriate circular polarization.

In order to create the diffraction dipole traps to which the atoms are being transferred, we need a tunable diode laser with as high an output power as possible (to create the deepest possible traps). Traditional laser tuning using a grating results in a loss of about 40% of the power emitted by the laser diode for feedback purposes. One established method to circumvent this problem is so-called injection locking. Here, the output from a grating-tuned diode laser is fed into a plain (though temperature and current stabilized) laser diode, which then proceeds to amplify only the seed light from the tuned laser. In this way, we can use the full amount of laser output power (~ 100

mW) for creating our dipole traps. We had performed preliminary tests of this system in previous projects, but now the final setup has been completed and tested.

For characterizing the properties (atom number and temperature) of the MOT, and also of the diffraction dipole traps, we need a photodetector and a CCD camera. The number of trapped atoms can be determined by measuring the amount of light emitted by the cold atom cloud with the photodiode. The temperature of the atom cloud can be measured by measuring the size of the atom cloud with a CCD camera. We can use this information to optimize the MOT for atom number and temperature, by fine-adjusting the laser alignment, laser frequency and magnetic field strength.

For the atom transfer from the MOT to the diffraction dipole traps, we will record the photodetector and CCD camera signals as a carefully timed sequence of changing the laser detuning, switching off the magnetic field, and switching on the dipole trap laser is triggered. In order to detect even small numbers of atoms trapped for only a short time in the dipole traps during initial attempts (and to be able to better characterize the trap lifetime), we purchased a 200 frames per second CCD camera (along with a framegrabber and dedicated image acquisition computer), so that we will not miss our initial signals. Once first traces of transferred atoms are detected, we can optimize our timed transfer sequence for transferred trapped atom number and temperature. We have thoroughly tested the new camera/framegrabber/computer setup.

We have thus completed preparations for atom transfer from the MOT to the diffraction dipole traps. If achieved, this will be the first demonstration of atoms trapped in the diffraction pattern behind a circular aperture, and will also constitute significant progress in the field of neutral atom quantum computing.

In addition to the experimental work described, we also continued our computational work exploring the dipole traps formed behind a circular aperture. In particular, we have investigated the polarization dependence of these dipole traps, which can be exploited to bring two traps together and apart controllably, which would be a great asset for facilitating the two-qubit operations needed for quantum computing. We were able to finalize our research by determining all normalization conventions used in the literature and thus achieving correct scaling of our results in SI units. A sample of our results is shown in Fig. 3. These results were presented at several conferences and are currently under preparation for publication (see below).

4. Student involvement:

All aspects of this research project involved undergraduate students. During Spring and Summer quarters 2009 a total of six students worked on this project: Michael Boardman (Electrical Engineering, worked for individual study credit), Bert Copsey (Mechanical Engineering), Andrew Ferdinand (Physics), Eric Muckley (Physics), Grant Rayner (Physics), and Jennifer Rushing (Physics, worked for individual study credit). All students gained some experience in all aspects of the experiment: vacuum equipment, optics, lasers, electronics, LabVIEW programming. Details of the students' tasks are given in the next section, and samples of their work are shown in Fig. 4.

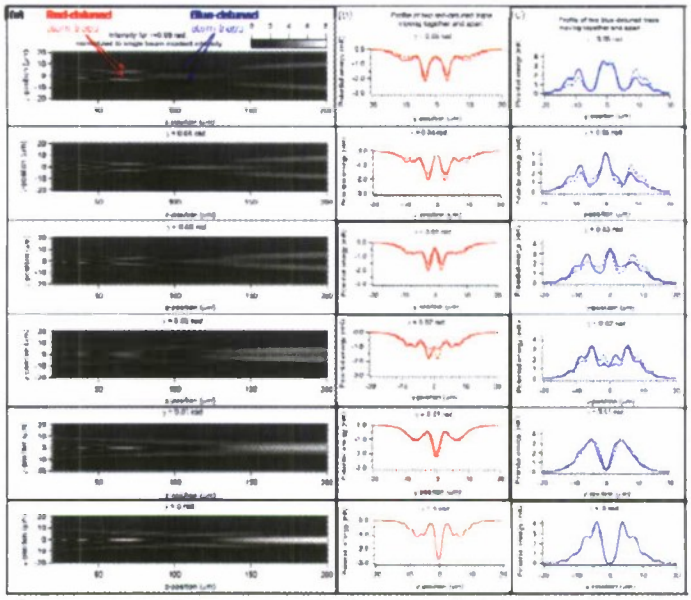


Figure 3: Sample figure from manuscript on computational work. (a) Intensity patterns formed by two laser beams of opposite circular polarization incident on a pinhole at an angle γ. An atom can be loaded into each of the bright spots (for a red detuned laser) or each of the dark spots (for a blue-detuned laser). The two atoms can then be brought together, e.g. for 2-qubit operations, by changing the incident angle. (b) Potential energy wells for atom in bottom trap site (solid line) and top trap site (dotted line) for red detuned laser beams. (c) Potential energy wells for atom in bottom trap site (solid line) and top trap site (dotted line) for blue detuned laser beams.

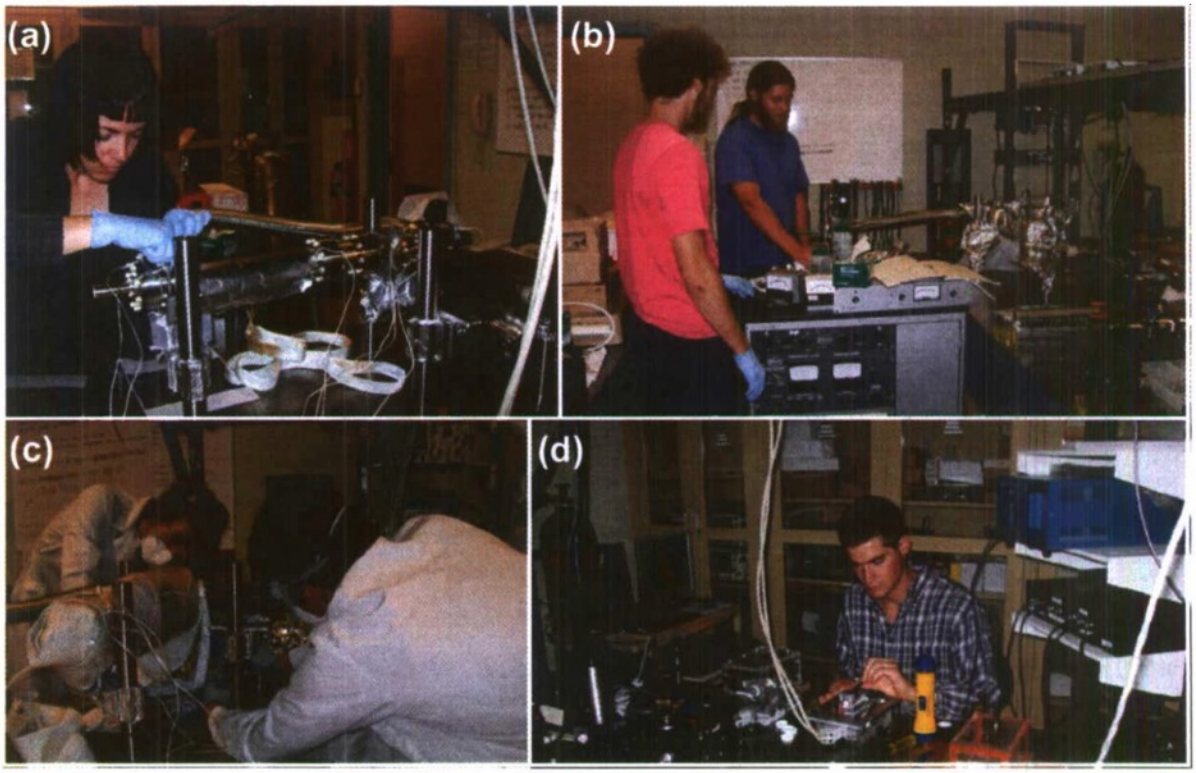


Figure 4: Selected student activities. (a) Jennifer Rushing installing thermocouples in preparation for the bakeout. (b) Grant Rayner (left) and Andrew Ferdinand troubleshooting the vacuum pumping station in preparation for the bakeout. (c) Michael Boardman (left) and Bert Copsey unwrapping the vacuum chamber after the bakeout. (d) Eric Muckley tuning the trap laser for the MOT.

5. Tasks completed:

The tasks we achieved are:

- Repair vacuum pumping station
  - Ionization gauge: Build custom cable to connect existing ion gauge to working ion gauge control unit
  - Thermocouple gauges: Replace one faulty gauge
  - Gate valve: Install new gate valve, replace faulty solenoid
  - Turbopump: Replace with new turbopump after seizing
  - Turbopump control unit: Repair cables, borrow control unit from colleague at another university, later replace it with new unit
  - Sieve trap: Install new sieve trap
  - Roughing pump: Replace existing pump with different pump that has an anti-suckback valve
  - Chiller: Find and install rare connector for chiller plumbing to turbopump
  - Relay box: Design and build relay box for safe emergency shutoff of pumping station, troubleshoot and repair grounding problems (with the help of department technician Jim Hilsinger)
  - Monitor program: Write bakeout and ion pump monitor program to record pressure and temperature during bakeout, and safely shut off the pumping station in case of emergency
- Bake out ultra-high vacuum chamber
- Identify cause and implement solution to tunable diode laser temperature stability problems
- Acquire equipment (CCD camera (200 frames/second), combo GigE Vision and GPIB boards, computer for image acquisition, photodiodes)
- Set up and test computer, GigE board and camera for image acquisition
- Repair photodetectors (troubleshoot and fix circuits, replace photodiodes)
- Install MOT magnets
- Design and set up optical system for MOT (six opposing laser beams with correct circular polarization)
- Set up injection locked tunable laser system in its final configuration
- Calculate polarization dependent trapping potential energies for pairs of laser beams at different angles incident on a pinhole, including proper normalization conventions

Michael Boardman assisted in trouble-shooting and repairing the vacuum pumping station needed for baking out the ultra-high vacuum chamber for the magneto-optical trap.

Bert Copsey contributed both through his experimental work as well as his computational work. Experimental work: Troubleshooting and repair of custom electronic equipment for vacuum pumping station, bakeout of vacuum chamber, design and construction of hardware for fixing tunable diode laser stability problems, design and construction of cold spot for magneto-optical trap. Computational work: Literature search and analysis of normalization conventions for atom trapping potential energy calculations. This step completes Bert's computational research on the polarization dependence of the atomic dipole traps formed by diffraction of laser light at a circular

aperture, including the ability to bring two atom traps together and apart for two-qubit operations for quantum computing. His work was presented at three conferences and is being prepared for submission to a peer-reviewed journal (see below).

Andrew Ferdinand made a strong contribution to repairing the vacuum pumping station needed for the bakeout of the ultra-high vacuum chamber, co-performed the bakeout, assisted in test and repair of photodetectors and electronics for laser tuning and stabilization, assisted in LabVIEW programming.

Eric Muckley tested and repaired photodetectors and electronics for laser tuning and stabilization, tuned lasers for magneto-optical trap, assisted in LabVIEW programming, vacuum pumping station repair, and preparation for bakeout. He also tuned the tunable diode lasers to the atomic transition frequencies of the ^{87}Rb atom in preparation for trapping atoms.

Grant Rayner helped repair the vacuum pumping station, co-performed the bake-out, had a strong contribution to the development of the LabVIEW program, set up the magneto-optical trap electromagnets and designed and set up the MOT optical system.

Jennifer Rushing developed a monitoring/alert program in LabVIEW to remotely monitor the experiment and protect it in case of any failure, assisted in repair of the vacuum pumping station, co-performed bakeout of vacuum chamber, diagnosed laser temperature stability problems, assisted in laser tuning for magneto-optical trap.

6. Presentation of results:

Several presentations resulted from the computational results on polarization dependent atomic dipole traps:

1. "Light Polarization Dependence of Optical Dipole Traps Created in the Diffraction Pattern of a Pinhole," Bert David Copsey, Katharina Gillen-Christandl, Division of Atomic, Molecular, and Optical Physics (DAMOP) – an international professional meeting of the American Physical Society, Charlottesville, VA. Invited talk at the Undergraduate Research Session of DAMOP, May 21, 2009. See Fig. 5.
2. "Polarization-dependent atomic dipole traps behind a pinhole for controllable manipulation of qubit locations in a quantum memory," Bert David Copsey, Katharina Gillen-Christandl, Glen D. Gillen, DAMOP, Charlottesville, VA. Poster presentation, May 22, 2009. Shown in Fig. 6.
3. "A neutral atom quantum memory created by diffraction of laser light at an array of pinholes," Katharina Gillen, Southwest Quantum Information and Technology (SQuInT) network, Twelfth Annual Workshop, Santa Fe, NM, February 20, 2010. Talk.
4. "Investigation of a quantum memory created by diffraction of laser light at an array of pinholes," Bert D. Copsey, Katharina Gillen-Christandl, Glen D. Gillen, DAMOP, Houston, TX, May 27, 2010. Poster presentation. Shown in Fig. 7.

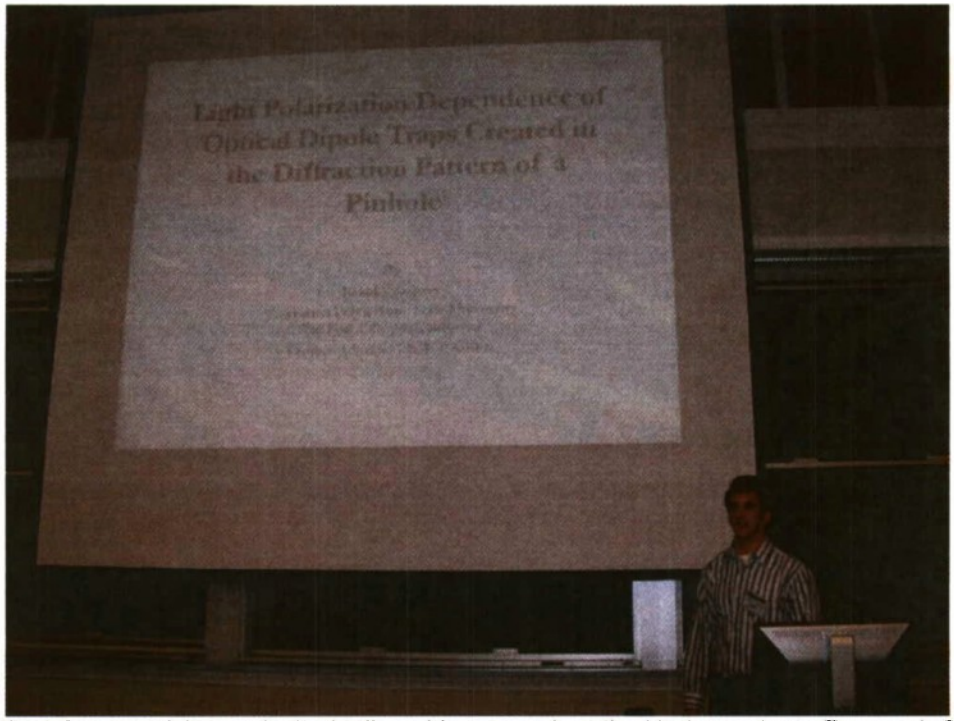


Figure 5: Bert Copsey giving an invited talk on his research at the Undergraduate Research Session of DAMOP 2009.

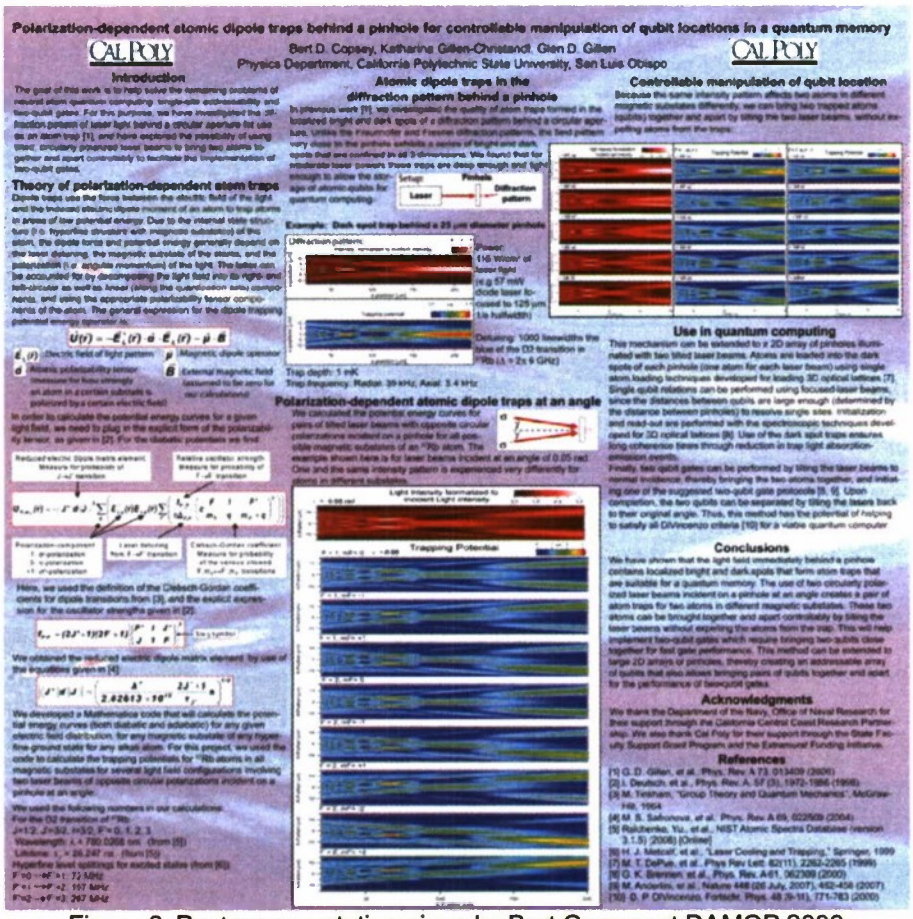


Figure 6: Poster presentation given by Bert Copsey at DAMOP 2009

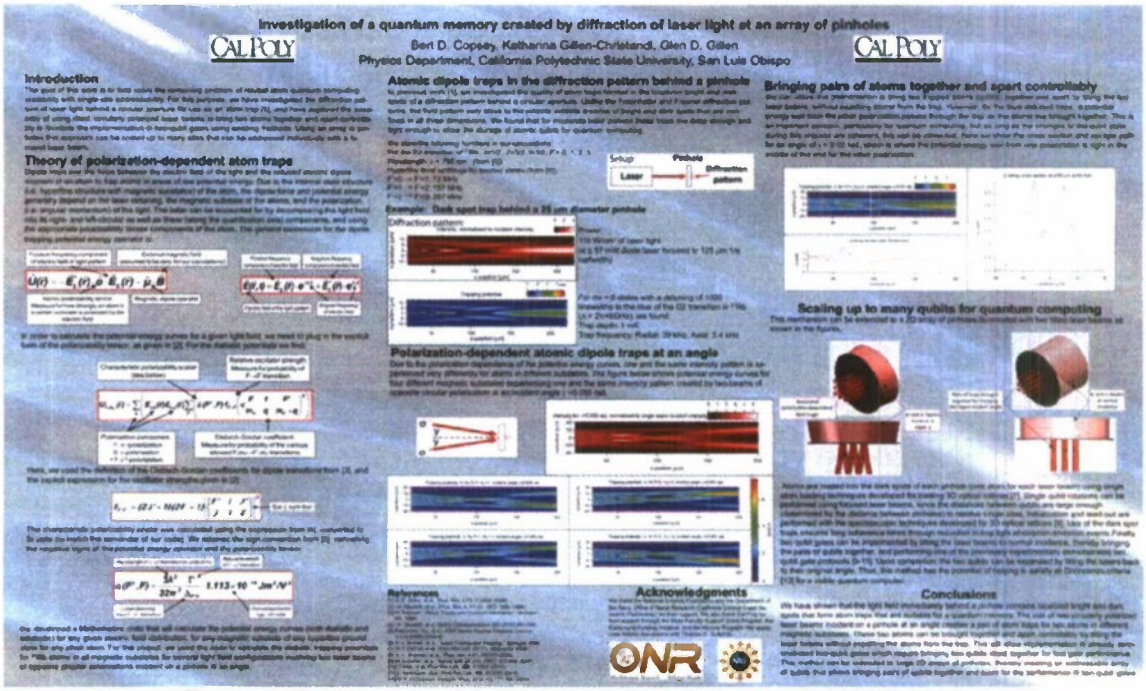


Figure 7: Poster presentation given by Bert Copsey at DAMOP 2010

7. Publication of results:

A manuscript for publication of our computational results entitled "Polarization dependent atomic dipole traps behind a circular aperture for neutral atom quantum computing" is currently in preparation. Submission to Physical Review A is planned for later this year.

8. Conclusions:

We successfully finished the construction of a magneto-optical trap (MOT), and preparations for the transfer of atoms from the MOT into the dipole traps formed in the diffraction pattern behind a circular aperture.

In addition to the experimental work, we were able to finalize computational research on the polarization dependence of atomic dipole traps formed behind a pinhole. Results indicate that two circularly polarized laser beams incident on a pinhole at an angle can be used to bring two atoms in different magnetic substates together and apart without expelling either atom from the trap. In particular, we were able to identify all normalization conventions used in the literature and scale our atomic trapping potential energies to SI units for direct comparison to experiments. If our computational results are confirmed by experiments, this method can be used to facilitate two-qubit gates. This would be a significant step towards building a neutral atom quantum computer.

Isotropic to Nematic Transition of Aerosil-disordered Liquid Crystals

Principal Investigator:

Saimir Barjami, Ph. D. Physics
California Polytechnic State University
San Luis Obispo, CA

Abstract

The study of ordering transitions in systems within disordered environments has provided new insights into the physics of phase transitions [1, 2]. Liquid crystals (LCs) incorporating low-density gels formed by silica aerosil dispersions (LC+A) have been shown to be good experimental models to investigate both the isotropic (I) to nematic (N) and the nematic to smectic-A (SmA) phase transitions in the presence of the static (quenched) positional and orientational randomness provided by surface coupling at the silica-LC interfaces [3-6]. The I-N transition in LC+A is more difficult to interpret. A particular challenge is the observed doubling of the LC+A calorimetric peak at the I-N transition in an intermediate range of silica concentrations [7-10]. This feature is interesting because it has no analogy in the LC bulk behavior, and thus its investigation provides a unique route to access the basic physics of discontinuous transitions in artificially disordered systems.

This proposal seeks support for an experimental research program into the physics of quenched random disorder effects on the phases and phase transitions of soft condensed matter and complex fluids (liquid crystals).

The ramifications of this proposed program touch most of the current important topics in physics; from doped semiconductors to superconductors, from quantum phase transitions to biological structures. Major thrust avenues have been identified using a fractal gel formed by aerosil particles imposing quenched random disorder on liquid crystals as the physical test-bed for all the important ideas in understanding the effects of disorder as well as to address new effects not previously anticipated.

Introduction

The isotropic (*I*) to nematic (*N*) phase transition is weakly first-order and the orientational order is characterized by a tensor order parameter. In the presence of dispersed surfaces as the source of the Quenched Random Disorder (QRD), the disorder couples to both the magnitude and the direction of the orientational anisotropy axis (the director). Thus, in general, both field-like and elastic couplings may occur. In Liquid Crystals (LC) + aerosil systems, the soft nature of the gel partially anneals the stronger elastic coupling reveling the random-field coupling to the magnitude of the order parameter.

Recent work indicate that the *I-N* transition in low-silica density LC+aerosil samples proceeds via a two-step process; crossing over from RD to RF dominated QRD coupling. By varying surface chemistry (coupling), LC elastic constants, and surface radius of curvature, one can time this two-step process to isolate the initial RD effect for a detailed calorimetric characterization. A calorimetric study on the *I-N* transition on photo-polymerized liquid crystalline polymers as a function of cross-linking (in the low-crosslink, weak disorder regime) should probe essentially pure RB effects.

Results and Discussion

In this experimental research work we focus on a liquid crystal – 4 – cyano – 4' - n – heptylbiphenyl (7CB) which was purchased from Sigma Aldrich Corporation, St. Louis, MO, USA. Pure 7CB has two transition temperatures, nematic-to-isotropic (TNI) at 430C and crystal-to-nematic (T_{CN}) at 300C, according to its own phase transition. We prepared several samples of 7CB with quenched random disorder (QRD) effects incorporated into them. QRD in Liquid Crystals (*LC*) requires the inclusion of fixed random solid surfaces at all possible length-scales up to the sample size. This was achieved by "dissolving" type 300 aerosil (*SIL*) into the host *LC*. The *SIL* is comprised of SiO_2 (silica) spheres of diameter about 7 *nm*, coated with (*-OH*) hydroxyl group. The coating enables the spheres to hydrogen bond and form a thixotropic, fractal gel, in an organic solvent, through a diffusion limited aggregation process.

Several AC Calorimetry temperature scans were carried out for several densities of aerosil particles in 7CB: $\rho_S = 0.036\frac{g}{cm^3}$; $\rho_S = 0.078\frac{g}{cm^3}$; $\rho_S = 0.489\frac{g}{cm^3}$; $\rho_S = 0.647\frac{g}{cm^3}$.

In order to determine the excess heat capacity associated with the phase transitions, an appropriate background was subtracted. The total sample heat capacity over a wide temperature range had a linear background, *Cp*(background), subtracted to yield:

$$\Delta C_p = C_p - C_p\left(background\right)$$

In Figure 1 we present excess specific heat, ΔCp, as a function of temperature for 7CB + aerosil disordered systems for at $\rho_S = 0.036\frac{g}{cm^3}$ silica density over a wide range of temperatures, from 306 K to above 326 K. A Sharp Isotropic to Nematic phase transitions is observed for the bulk at 316.6 K, as shown clearly in Figure 1.

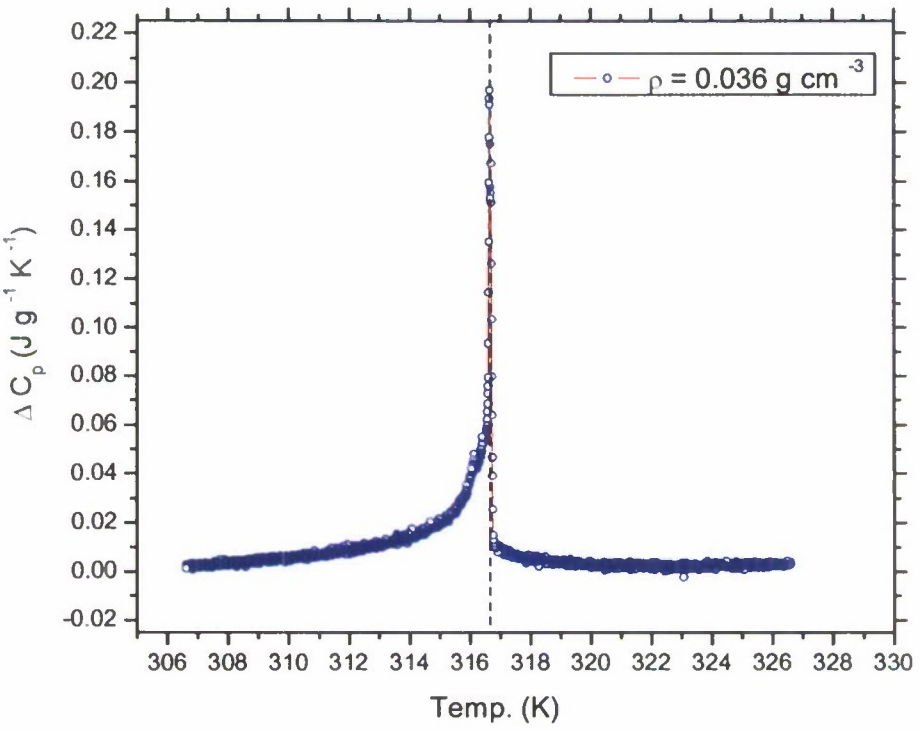


Figure 1: Excess specific heat, ΔCp, as a function of temperature for 7CB + aerosil disorder at $\rho_S = 0.036 \frac{g}{cm^3}$ silica density. The vertical dashed line indicate the *I-N* transition. Isotropic to Nematic phase transition is observed at 316.6 K.

After 7CB + aerosil disorder at $\rho_S = 0.036 \frac{g}{cm^3}$ silica density data, we then proceeded with several AC Calorimetry temperature scans for 7CB + aerosil disordered systems for several densities of aerosil particles in 7CB: $\rho_S = 0.078 \frac{g}{cm^3}$; $\rho_S = 0.489 \frac{g}{cm^3}$; $\rho_S = 0.647 \frac{g}{cm^3}$. In Figure 2 we present different temperature scan data taken for the above samples.

Isotropic to Nematic phase transitions for different samples are clearly indicated in the Figure 2 by different peaks. Temperature shifts and the suppression of the Isotropic to Nematic phase transition peaks are clearly seen in Figure 2 as the density of silica is increased from 0.036 to 0.647 g/cm^3.

The evolution of the first order Isotropic to Nematic phase transition peak is clearly indicated from the picture starting from 0.078 g/cm^3 density, getting bigger for 0.078 g/cm^3 sample, getting suppressed for 0.489 g/cm^3 sample and 0.647 g/cm^3 sample, which is a clear signature of the Quenched Random Disorder effects of aerosil particle in 5OCB liquid crystal.

The effect of quenched random disorder is clearly showing in the data: the shift in transition temperatures, the broadening of the transition region, the suppression of the peak with increasing silica density.

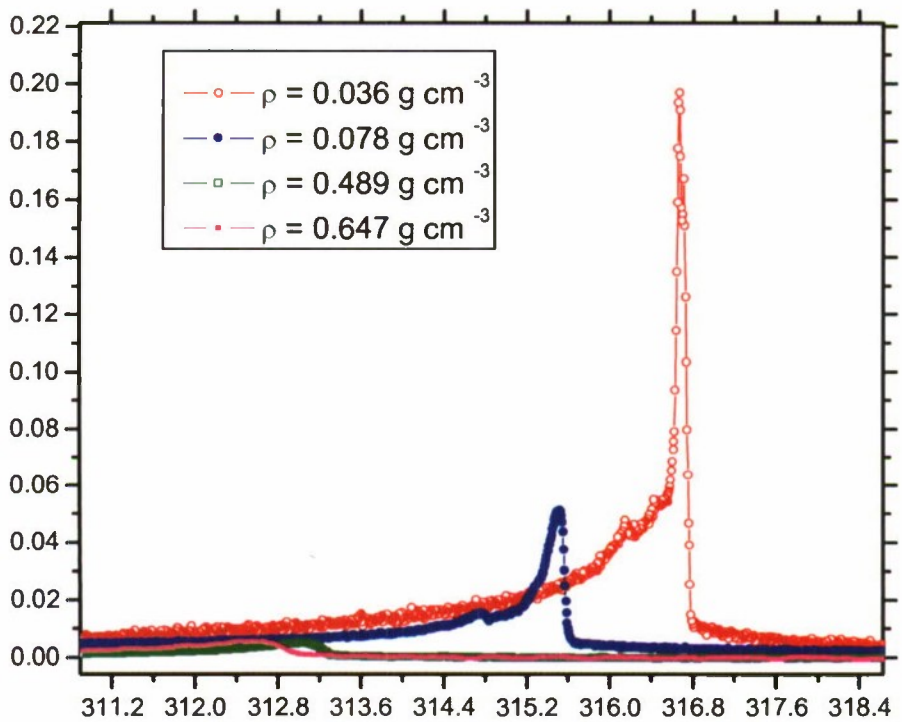


Figure 2: Excess specific heat, ΔCp, as a function of temperature for 7CB + aerosil samples at different silica densities taken with our new AC Calorimeter. Isotropic to Nematic phase transition regions are clearly seen in the figure. See figure insets for definition of symbols.

The influence of the aerosil density both on the Nematic to Isotropic transition temperature shift, in the framework of the present theory, has originated from accounting for the elastic deformation of the host LC phase due to the aerosil particles immersed in the LC phase. On

increasing the aerosil particle concentration, the degree of disordering enforced on a host LC phase can be relatively well controlled, and the organization of aerosil particles undergoes two qualitatively different regimes, a random-dilution to random-field crossover. Below $\rho_S = 0.05 \frac{g}{cm^3}$ a gel-like aerosil structure is formed (labeled as "soft"), meaning that the bonds among the hydrophilic aerosil particles can easily break, allowing a rearrangement of the particles to form a network in which the elastic strain imposed by the surround LC phase is minimized. Above $\rho_S = 0.05 \frac{g}{cm^3}$ regime a rapid shift to lower temperatures of the Nematic to Isotropic phase transition is observed.

In the high density $\rho_S > 0.1 \frac{g}{cm^3}$ regime, the gel is "stiff" or rigid, having similarities with the aerogel systems, which is dominated by quenched elastic-strain smearing effects that destroy long-range order, and the residual order is a quasi-long-range order. In this regime, the decrease of the transition temperature shift shows a complicated behavior as a function of density. So, such dispersions are suitable for testing existing and future theoretical models.

These are some preliminary discussions of the results taken for 7CB + aerosil samples at different silica densities taken with our new AC Calorimeter.

Relevance/applicability of this work to DOD interests

This work and its results are part of the very important applications of Liquid Crystals and their many effects in many areas of science and engineering, as well as device technology, such as Optical and Radar Imaging, with a tremendous interest to **the Office of Naval Research** . Hand held electronic devices with liquid crystal displays (**LCD's**) are ubiquitous in our society and are of particular interest to the **DOD** for work in the field. LCD's are optimal for these applications because of their slim profile and low power requirements. Significant improvements in these systems can be made to optimize these **LCD's** to reduce power requirements (which can lead to longer battery life) and improve visual quality. Our work and its results could lead to predictions to improve the design of future materials, which could optimize these materials for LCD's and reduce their cost by simplifying synthesis and producing materials in bulk quantities.

These results will be submitted for publication in one of Physical Review Journals at a later time.

Inventory Management Utilizing Passive Radio Frequency Identification for the Mission Package Support Facility at Port Hueneme Naval Base

Principal Investigator:

Tali Freed, Ph.D., P.E.
Director, PolyGAIT – The Cal Poly Center for Global Automatic Identification Technologies
California Polytechnic State University
San Luis Obispo

Introduction

In order to better understand the uses of pRFID for inventory management and to increase their logistical readiness level, the US Navy and Department of Defense (DoD) commissioned prototype pRFID systems to be built that demonstrate the advantages of pRFID and the feasibility of a full implementation [18]. Cal Poly's Global Automated Information Technologies Research Center (PolyGAIT) was contracted to build one of these prototypes at the Littoral Combat Ship (LCS) Mission Package Support Facility (MPSF) located at the Naval Warfare System Center Port Hueneme (PHD). In conjunction with Cal Poly's Spring 2010 IME 570/571 classes, PolyGAIT designed Phase 1 prototypes for 3 separate systems.

With the help of the Cal Poly IME 457 class in the Fall of 2010, the project moved forward to Phase 2 which was to create and install one integrated system at the MPSF at PHD. Along with the pRFID system the project report will provide the economic and performance justifications for the hardware choices, a simulation demonstrating the benefits of moving to a full pRFID implementation, ergonomic reasoning for physical system design, and an operational procedure for operating the system. In addition to background research, knowledge will be drawn from existing pRFID implementation done by PolyGAIT [7] as well as white papers written at Cal Poly, and performance testing carried out in PolyGAIT and on site at the MPSF at PHD.

Background

Radio Frequency Identification Technology

Recent advances in wireless communication technologies have been changing the business world. Chief among these technologies is Radio Frequency Identification (RFID). RFID provides the capability to identify and track objects that have RFID tags without the need for direct physical contact or line-of-sight visual scanning like barcode or other UID technologies. RFID is expected to reduce logistic problems as identified by the Government Accountability Office

[12], and several case studies of successful implementations in the DoD have already been analyzed and reported [18][5].

Passive UHF RFID (pRFID - frequency range 860-960MHz) has recently been approved for use on U.S. Navy ships [1]. As a result in March 2010 MPS-420 decided to evaluate the use of pRFID for inventory tracking at its LCS Mission Package Support Facility in Port Hueneme, CA.

Littoral Combat Ship

In 2002 the Navy initiated the Littoral Combat Ship (LCS) Program asking for "a fast, agile, focused-mission platform designed for operation in near-shore environments yet capable of open-ocean operation" [21]. Lockheed Martin and General Dynamic developed separate designs that became known as USS Freedom (LCS1) and USS Independence (LCS2) respectively [21]. Both of these ships are designed to be capable of multiple missions because of a reconfigurable payload. This allows the ships to be utilized for many missions instead of just a single task as is tradition for the navy. Also the ships will be manned by multiple crews so that they will be in service a greater amount of the time, as opposed to the old strategy of one ship, one crew. The flexible nature of the ships is one of their greatest assets, but it also creates more inventory to keep track of.

Facility

The Mission Package Support Facility (MPSF) at Port Hueneme in Oxnard, California is the first facility that is dedicated to the maintenance of the Mission Packages that the LCS will be using. A Mission Package is a sea container that contained the systems for a given mission. These Packages can be shipped to the ships that need them and loaded to have the ship ready for their specific mission. The LCS fleet is expected to grow and the inventory monitoring process could potentially become an issue as the size increases. To decrease the resources needed to carry out the inventory management tasks, as well as increase the visibility of the inventory, the Navy is looking to add an RFID system.

Current Inventory Processes and Problems

Currently, tracking inventory at MPSF is performed manually. As parts are used, each item is supposed to be recorded on the computer by the engineer or the administrative assistant. This manual process is time consuming, error prone, and is expected to lead to inventory inaccuracies, which in turn lead to decreased readiness level [12].

Tool check-in and check-out transactions are not tracked at all. Tool loss and misplacement is expected to increase as the level of activity at MPSF increases with the number of ships. Tool unavailability is expected to decrease readiness levels, as well as increase the number of searches for missing tools. Labor cost will increase along with tool replenishment cost.

In addition, a missing tool may create life-threatening risk, if it is forgotten in a weapon system Sea Container and results in a weapon system failure. This event may be rare, but its consequences extremely severe.

Project Background

After a visit to Cal Poly's Global Automated Information Technologies Research Center (PolyGAIT) PMS 420 indicated its desire to have Cal Poly's students and faculty design a prototype pRFID system for inventory tracking at its LCS Mission Package Support Facility in Port Hueneme, CA (MPSF-PHD). This System will aim to increase visibility by allowing more frequent full inventory checks with 100% accuracy so that the inventory levels of spare parts are well known as well as also manage the tools and assets necessary to carry out maintenance procedures.

To build this system this project will concentrate on 3 areas of the MPSF PHD: the Tool Cage, the Spare Parts Cage, and the Minor Assets. I worked closely with the Spring 2010 IME 570/571, the Spring 2010 IME 443 Port Hueneme Facilities team directed by Liz Schlemer, the Fall 2010 IME 457 RFID class, Cierra Logistics consulting (who employed me during the fall quarter of 2010 to work on this project), the employees of the MPSF at PHD, Larry Rinzel, and Tali Freed.

Literature Review

[12] This US GAO report states that of the $28.1 billion that the Department of Defense obligated for Operation Iraqi Freedom, the services and the Defense Logistics Agency reported by July 2003 that $14.2 billion was for operating support costs and $4.9 billion is for transportation costs. There were many reasons for these costs but chief among them were the poor visibility of inventory. This poor visibility was caused by many different factors which include the following. First of all even though RFID was utilized where ever possible, inconsistent tagging methods and performances made the different tags in operable between systems. The personnel on site also had limited access to the tools and equipment necessary to take advantage of the DoD logistics' tagging. The personnel also received limited training to help them utilize the technology. It is identified in this report that RFID was not fully utilized during Operation Iraqi Freedom. There are opportunities for improvement that will come from better integration, more consistent usage, and better training of personnel. It is stated that there was a potential loss of readiness for vehicles and equipment because of a lack of support parts because of poor visibility

[18] Mr. Robert Bacon's (Director of RFID, Naval Supply Systems Command) presentation at the DoD RFID Summit in April 2007 describes some of the successes of the US Navy's RFID Implementations. For active RFID they operate over 2,700 read and write sites around the world and make 19,000 writes and 134,000 reads each week. 90% of all DoD international air/sea shipments are tagged.

The Navy was an early adopter of passive RFID (pRFID) in the supply chain and you can see from slide 6 that they acknowledge the technological advanced and benefits afforded by RFID but they wish to look at the business case for it. At this time they began testing pRFID in the supply chain and implementing it in more warehouses and ships.

[1] In this September/October 2009 Navy Supply Corps News Letter Mr. Robert Bacon (NAVSUP CIO/Navy AIT Program Manager) describes the current state and future of Navy supply chain logistics. He points out the DoD has $70billion of asset inventory, $6billion of which are in transit at any given time. It is now more necessary to rely on Automatic Information Technologies (AIT) to assist in the tracking process. Along with this because of the massive scale, an enterprise approach is necessary to be able to fully implement any solution. While developing this enterprise AIT solution one of the Navy's primary goals is to pursue pRFID as a solution. He states, "Already well underway, one of Navy AIT's primary missions is to passive RFID-enable the receiving and issuing operations at Navy sites, by following a three-tier process to deploy pRFID Navy-wide. In Tier 1, the Navy is developing a pRFID hardware and software infrastructure, both at an Enterprise and field activity level.", "Tier 2 involves the integration of AIT-generated data from fixed portals or handheld devices with Enterprise Automated Information Systems (AIS). This step allows the Navy to integrate highly-accurate supply transaction data into logistics processes, reducing processing time and enhancing asset visibility. Tier 3 builds on best practices developed during Tier 2 by integrating AIT-generated data into the broader DoD systems."

[21] This is the US Navy's site summarizes the LCS Program. A program asking for a fast, agile, focused-mission platform designed for operation in near-shore environments yet capable of open-ocean operation.

[4]

EPC Global			
Class	Generation (Gen)	Frequency	In Use?
1	0		N
	1		N
	2	860MHz - 960 MHz	Y

Reference	Title	Regulation
EPC Global Class 1 Generation 2		
6.3.1.1	Operational Frequencies	frequency range is 860MHz to 960MHz inclusive
6.3.1.2.2.	Interrogator to tag Modulation	Interrogators communicate using DSB-ASK, SSB-ASK, or PR-ASK modulation
6.3.1.2.3	Data encoding	the reader to tag link uses PIE. The reference time interval used is Tari
6.3.1.2.4	Tari Values	Range of 6.25μs to 25 μs
6.3.1.3	Tag to interrogator communication	the tag will communicate using backscatter modulation
6.3.1.3.1	tag to interrogator Modulation	Tag back scattering shall use ASK and/or PSK modulation
6.3.1.3.2	Data Encoding	Tags shall encode the backscattered data as either FM0 base band or Miller Modulation
6.3.2.1	Tag Memory	Tag memory is separated into 4 distinct banks: Reserved Memory, EPC Memory, TID Memory, and User Memory
6.3.2.11.3.2	Read	16 bit size
6.3.2.11.3.3	Write	16 bit size

[24]

US DoD RFID Supplier Guide		
3.1	(RFID) Tag Placement	Tags may be integrated with the shipping label or may be separate.
		The bottom edge of the address label containing the unit load information should be within the range of 81 cm to 122 cm (32 to 48 in) from the bottom of the pallet (palletized unit load). If the loaded pallet (palletized unit load) is less than 102 cm (40 in) in height, the label should be placed as high as possible on the pallet (palletized unit load), but not closer than 5 cm (2 in) to the natural top of the unit load. (Figure 6 and 7 to the right)
		The passive RFID tag should be placed on the identification-marked side and right of center on a vertical face, allowing a minimum of 5 cm (2 in) from all edges.
		The RFID-enabled label or passive RFID tag attached separately should not be placed in a manner that overlaps any other existing RFID transponder. There should be at least a 10-cm (4-in) separation.

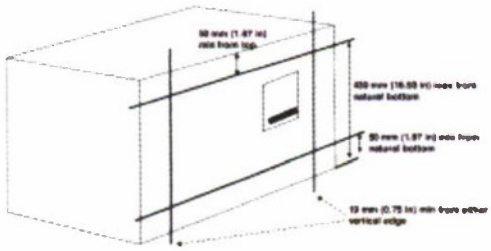


Figure 2: RFID Label Placement on Case

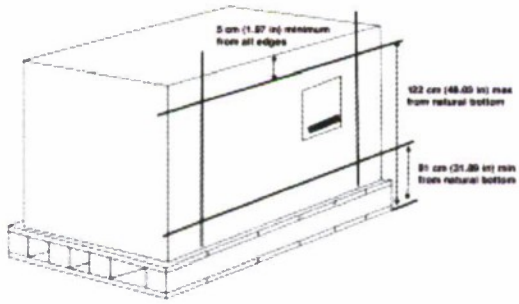


Figure 1: RFID Label Placement on Palletized Load

[23]

mil-std-129P CH4		
4.3	Placement of identification and address markings	
4.3.2.a	Placement of ID	The exact location of the identification markings may vary slightly. They shall be applied to the upper left two-thirds of the side of the container having the greatest overall, usable marking surface.
4.3.2.1	Boxes and crates	Boxes and crates 10 cubic feet and over shall have additional identification markings placed on the end of the container to the left of the identification-marked side. Placement of identification markings on the end of boxes and crates under 10 cubic feet is optional.
4.3.2.1	Boxes and crates	The address label shall be placed on the identification-marked side and right of center on a vertical face, allowing a minimum of 5 cm (2 in) from all edges of the box or crate.
4.3.2.2	Bales and cloth covered bundles	The identification markings on bales shall be stenciled on the upper two-thirds of the side of the bale having the largest marking surface area. Bales with a presewn end and a wire-tied ear on the opposite end shall have the NSN, quantity, and UI applied on the presewn end.
4.3.2.3	Bags and Sacks	If the bag is closed by stitching, a tag may be fastened to the bag by stitching when closure is made. If the top of the bag has ears, the tag shall be affixed to one of the
4.3.2.4	Barrels/ Drums	upper one-third of filled containers; top as well as upper one third for empties containers
4.3.2.5	Misc. articles and unpacked items (spools, reels, rods, coils of wire, cable, ect.)	tag shall be securely attached to the item with a wire or twine
4.3.2.6	unpacked major equipment (except unpacked vehicles)	a label attached directly on the equipment's surface with ASTM D 5486, type I, class 2 tape. The tape shall be placed over the label and shall extend half an inch or more from its edges
5.1.2	Labels, Paper, Pressure-Sensitive, Water-Resistant	Labels shall be of a water-resistant grade of paper, film, fabric, or plastic, coated on one side with water-insoluble, permanent type adhesive. The adhesive shall adhere to metal, plastic, aluminum or fiberboard surfaces under high and low temperatures.
5.1.2	Labels, Paper, Pressure-Sensitive, Water-Resistant	Application specific performance criteria and durability requirements to ensure functionality in various climatic environments should be tailored, if required, using MIL-PRF-61002

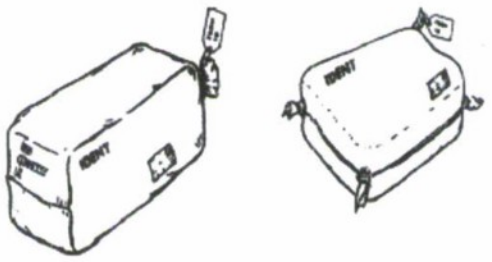


Figure 3: Bales and Cloth Covered Bundles (4.3.2.2)

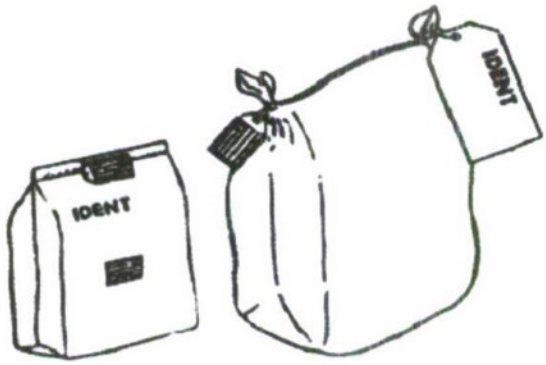


Figure 4: Bags and Sacks (4.3.2.3)

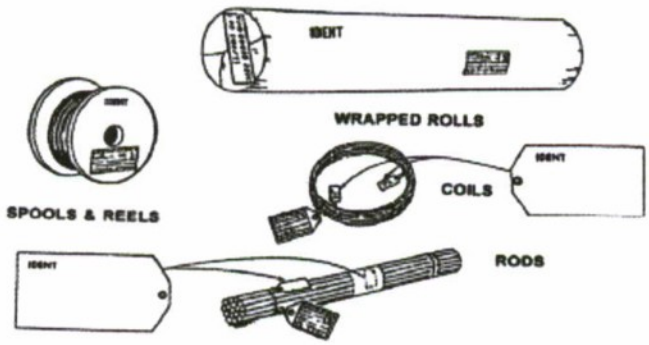


Figure 7: Misc. Articles (4.3.2.5)

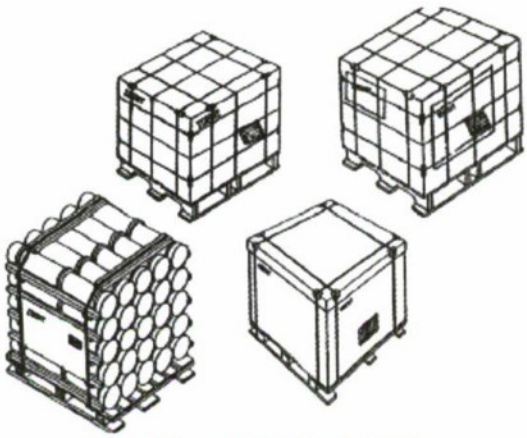


Figure 5: Boxes and Crates (4.3.2.1)

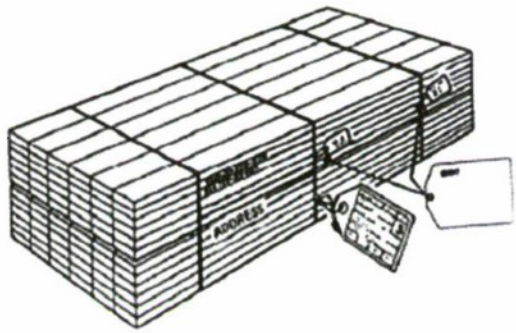


Figure 6: Unpacked Major Equipment (4.3.2.6)

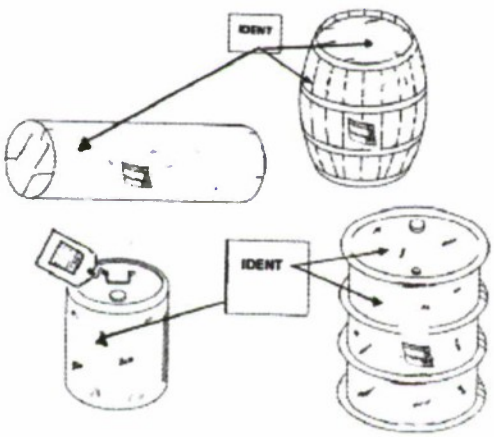


Figure 9: Barres and Drums (4.3.2.4)

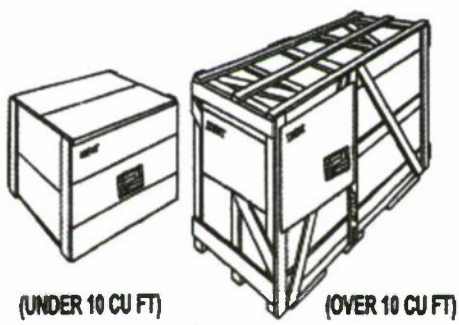


Figure 8: Boxes and Crates (4.3.2.1)

[5] Alan Estevez (assistant deputy undersecretary of supply chain integration) describes some quantifiable savings RFID has brought the Department of Defense. (May 2006) The U.S. DoD has already started to see significant payback from investments in RFID. Estevez presented at the 2006 RFID Journal LIVE! Conference that the dialog at logistics hubs have changed from "Where is my stuff" to "Place it on my truck because I know it is here". RFID has provided increased visibility and as such has reduced inventory value in the Marine supply chain from $127 million to $70 million. This is because they have more confidence in the supply chain and can see where the needed Items are. Because of this confidence they do not re-order as often, leading to a leaner, more responsive supply chain that has cut delivery times from 28 day to 16 days.

[2] The Project conducted by the IME 545 advanced simulation class in May 2007 showed that at that time if RFID was implemented to manage assets in Building 26 at Cal Poly, the Passive system would reduce the time spent managing assets from 10.65 hrs to 5.72 hours and would pay for itself in 1.6 years.

Phase 1

Problem Statement

The US Navy's current inventory management system has been identified as in-efficient and unable to achieve high levels of inventory visibility. The Navy wishes to develop an alternative means of inventory management that will raise visibility and lower total ownership costs.

Method

Phase 1 was carried out by 3 project teams from the Cal Poly IME 570/571 Advanced Electronic Manufacturing classes taught by Tali Freed and Larry Rinzel in the Spring Quarter of 2010. Each team was tasked with developing and purposing an RFID system that could manage a specific type of inventory at the MPSF. The three inventory types to be managed are Tools, Minor Assets, and Spare Parts. The three teams, which worked independently on their systems and databases, were made up by the following team members:

Tool Room Team:

Justin Grooms (IE), Laura Harris (IE), Alex Mosbacher (IE), Chris McEntee (GRC), and Joey Civin (Engr. Mgmt)

Minor Asset Team:

Patrick Hyland (MATE), Francisco Salazar (IE), Erin Kimura (IE), Jake Rucker (IE), and Timothy Chuc (IE)

Spare Parts Team:

Stewart Peugh (IE), Tyler Peabody (IE), Andy Liles (IE), Kyle Morton (IE), and Kenneth Cairns (IE)

System Descriptions

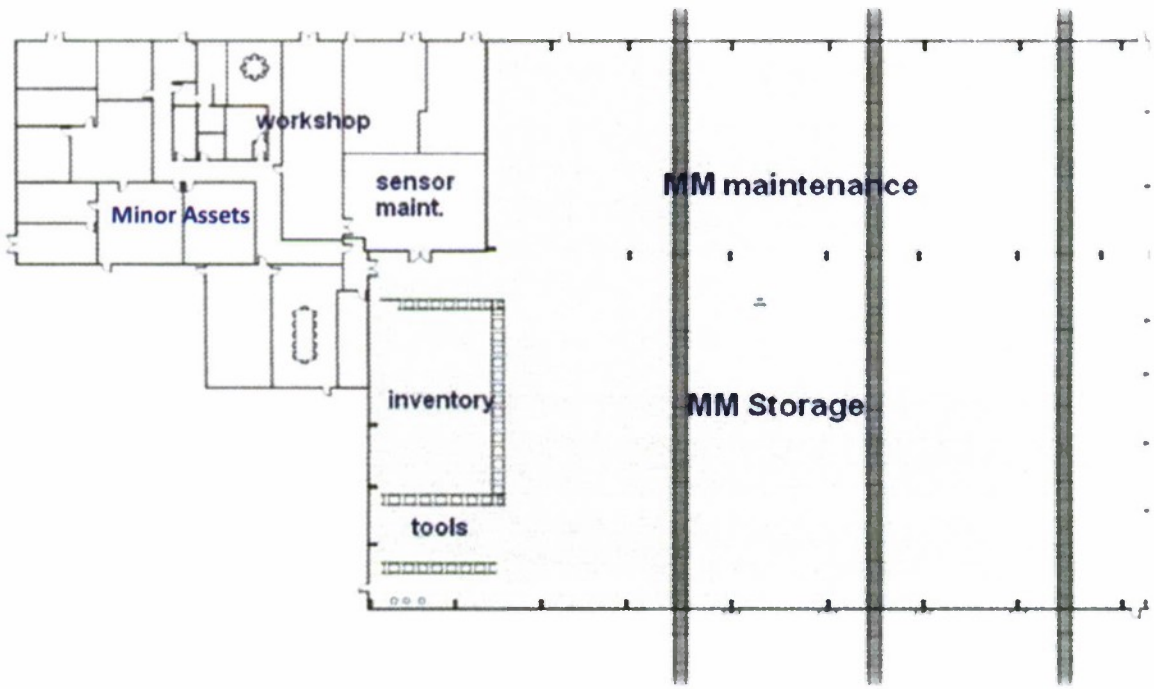


Figure 10: MPSF Facility Layout

Tool Room

The Tool Room team built a system that can monitor both metal and non-metal tools in the tool cage area (labeled "tools" on Figure 10). This cage has shelves up to 20ft high and one

entry/exit point. Some Items were also contained in metal containers. The navy asked for a system that could work as a check-in/check-out station, monitor unauthorized exits, and also track and manage inventory. To accomplish their task the team split each of these three sections into its own sub-system. Each sub-system had its own challenges and solutions which are outlined below.

Tool Room check-in/check-out station challenges

- Must identify all tools being checked-in and checked-out
- Must scan only tools being checked in or checked out
- Must Identify employees who are check- in or checking -out tools
- Must be located in an unobtrusive, but easy to access, location of the Tool Room.

Tool Room Check-In/Out System design

The Check-in/Check-out system was designed to occupy a space under the first shelving unit to the left of the entrance/exit door to the tool room. The width of this unit is approximately 8 ft and can be seen in Figure 12 below. The left half of the shelf will house a computer station used to verify items being checked-in/out. The computer station will also house the RFID reader for the system. The right half of the shelf will be the portal station used items will be placed in when checking-in/out. Figure 11 and Figure 12 below show the placement of the two antennas for this system. The first antenna will be mounted on the side of the shelving unit and the second antenna will be placed facing downward atop the shelving unit.

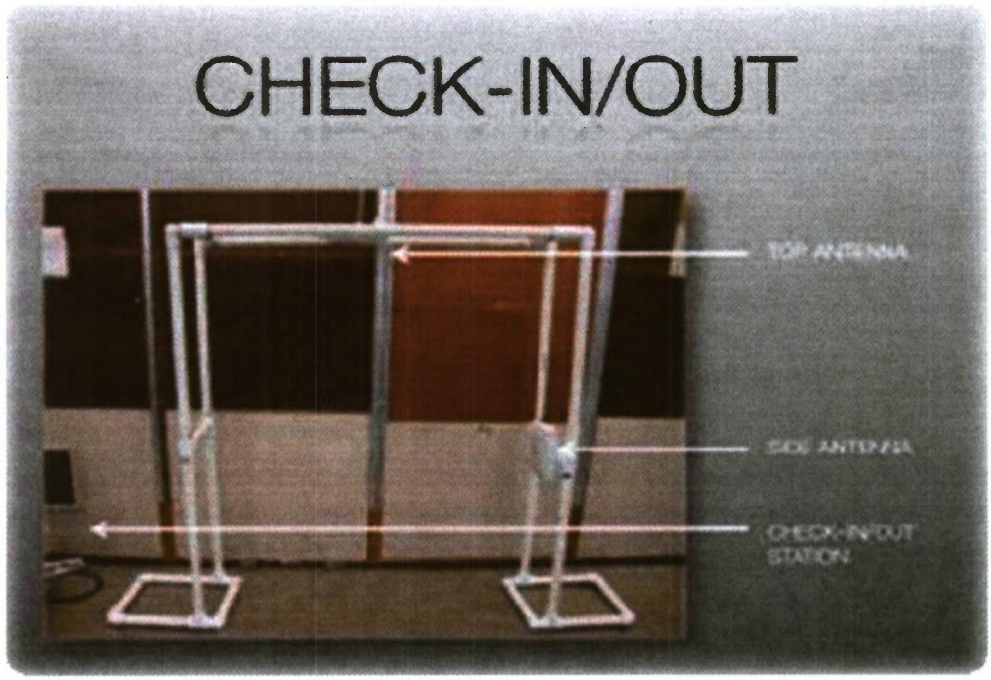


Figure 11: Check-in/out Station

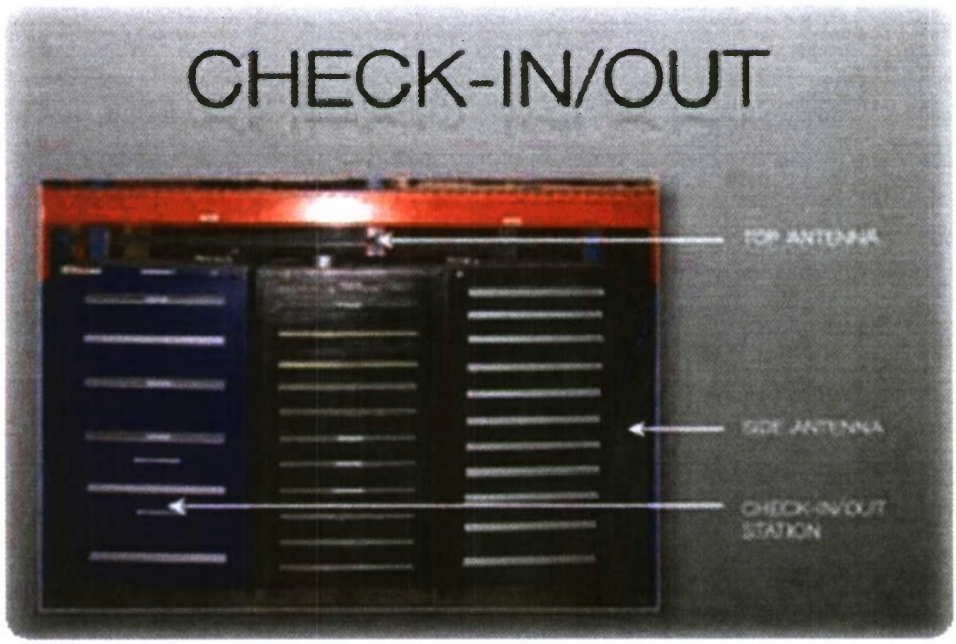


Figure 12: Check in/out Station Placement

Tool Room Unauthorized Exit System Challenges

- Identify whether an item has been properly checked-in or checked-out
- Warn employee if item has not been checked in/out properly
- Identify employees and what item they walk out with
- Only detect items that enter or exit the cage area.

Tool Room Unauthorized Exit System Design

The unauthorized exit system was designed to alert employees who attempted to take tools out of the tool cage without following proper procedures. In the event of an item being removed without authorization a signal would be sent using an audible and/or visible alarm to notify the staff. The system would then capture both the information of the employee (name, id, etc.) and the improperly removed tool(s). This system is a backup, fail safe method to ensure employees follow proper procedures when taking/returning tools from the Tool Room. The reader and computer for this system are designed to be mounted to the entrance/exit door and function independently from the other two tool cage systems. The antennas for this system would be mounted around the door as seen in Figure 13: Tool Room Un-authorized Check-out Portal PrototypeFigure 13. One antenna is mounted at the top of the door facing downward and the other two are staggered on either side of the doorway. The decision for this antenna arrangement was concurrently determined by a project performed at PolyGAIT by another Cal Poly student. Figure 14 shows the purposed placement at the facility.

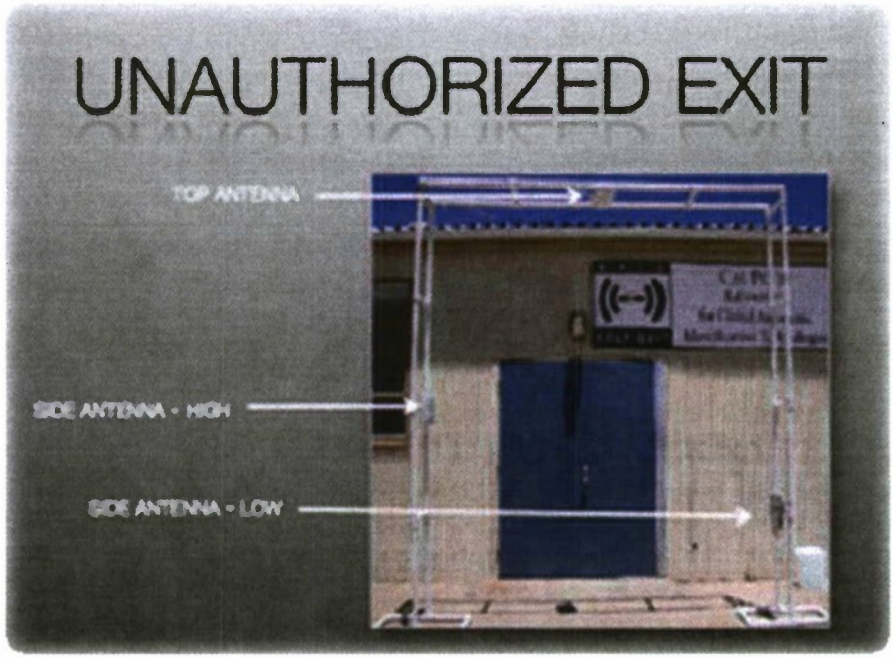


Figure 13: Tool Room Un-authorized Check-out Portal Prototype

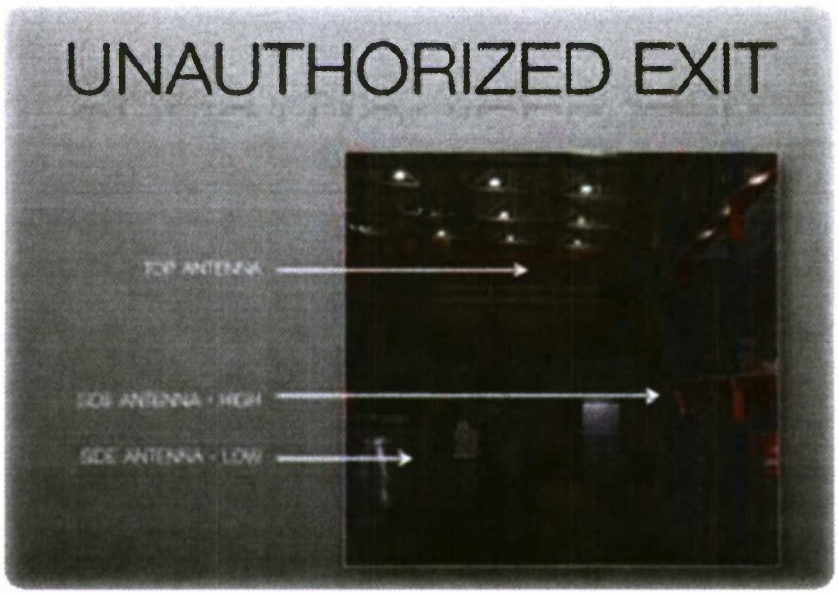


Figure 14: Tool Room Un-authorized Exit Placement

Tool Room Inventory System Challenges

- Identify all items on the shelves
- Lightweight, rigid, and easy to use extendable/telescoping rod
- Compact and easy to store rod

Tool Room Inventory System Design

The Tool Room Inventory System is designed to easily read and manage the Tools in the Tool Room. The system uses the same reader and computer system as the check-in/ check-out system. A telescoping rod is added with an antenna attached to the top is add to the current antennas on the cart when inventory checks are performed to guarantee reads of the inventory on the upper shelves.

Minor Assets

The Minor Asset team designed a system to track the minor assets that are used to administratively support the functions of the MPSF. These assets, which range in cost of $300 to $10,000, are made up of both metal and non-metal parts, and located through multiple rooms (labeled Minor Assets in Figure 10). Issues encountered by this system are outlined below.

Minor Asset tracking challenges

- Identify what items are contained in a given room in the multi room facility
- Identify if the item belongs to that room
- Assign and manage items that are assigned to individuals

Minor Asset Design

To overcome the difficulty of tracking items through multiple rooms the team designed a prototype that used a lightweight polyethylene cart to serve as the antenna, reader, and laptop station seen in Figure 15 below. This allowed for the system to be moved from room to room to find assets and manage their location. The system is designed to allow the operator to select which room they are in before initiating a scan.

When the scan is initiated the antennas are automatically attenuated for the selected room to ensure that all the items in that room are read while also ensuring that items from adjoining rooms are not also read. As items are found they disappear from an onscreen list so you know what items are left to still find. Also items that are found that are not assigned to the room are displayed so you know if there are possibly misplaced items in the room. The reader for this system is bolted to the middle shelf of the cart and there are two antennas mounted on articulating arms on either side of the cart. The articulating arms allow the antennas to be placed at different angles to guarantee that all of the items in a room are read.

Figure 15: Minor Asset Cart

Spare Parts

The Spare Parts Team designed a system that can track the inventory levels of spare and consumable parts in the Spare Parts room (Label Inventory in Figure 10). This room is a metal cage with industrial pallet shelves up to 20' high and some metal drawer units. The system also alerts operators when parts needed to be re-ordered and/or maintained based both on inventory levels and expiration date. Some challenges faced by the project team are outlined below.

Spare Parts Challenges

- The spare parts cage is a largely metal area
- The inventory items are a mixture of corrugate containers and metal parts
- Some items are inside metal containers
- The bottom of the top shelf is 20' high
- Not all incoming parts to the system will be tagged.

Spare Parts Design

To meet the challenges outlined above, the spare parts system was designed around a metal cart that holds the antennas, reader, and computer. The cart allows the system to be pushed through the spare parts cage to scan the inventory as often as it is necessary. The three antennas used by the system, as seen in Figure 16, are placed on the left hand side of the cart so that reads of all the items in the area can be completed with a single clockwise walk around the room. One antenna is bolted to the cart at waist height. Another antenna is mounted halfway up a telescoping metal pole and the third one is attached to the top of the telescoping pole. This top antenna is also removable so it can be waved over open metal drawers and guarantee reads of the items that are inside.

Figure 16: Spare Parts Cart

Analysis

All three teams met their design goals and build working prototypes during the quarter. While each team worked independently, there were several design decisions that were consistent across all three teams such as using mobile carts, designing passive UHF systems, and recommending Sirit readers with Poynting antennas. These decisions allowed the teams to meet performance and cost goals laid out by the navy.

Mobile carts were an attractive choice for each team because they lowered hardware costs by using a single RFID Reader instead of using an additional reader at each entry/exit point. A single mobile cart also allows the operator to perform intermitted inventory checks thereby increasing inventory visibility. If a portal was used, inventory could only be accounted for when it was added or removed from the system. This would allow possible inaccuracies in the records of inventory due to power failure, software error, or other unplanned system failure, and each team wanted to limit the probability of this as much as possible.

Ultra High Frequency tags (UHF) were used to meet the Navy's pRFID goals, and the tag purposed to be used across the entire project was the Sirit RSI-674 "Corkscrew" tag because these were used during a previous successful implementation PolyGAIT performed for PG&E at the Diablo Valley Nuclear Power Plant. This tag performs well in a metal environment, similar to the tool room and the spare parts room, which is one of the most challenging environments for an RFID system. Because the tags do not perform as well when directly attached to metal, metal parts that require tags will need to have tags attached that have a foam backing to insulate the tag and improve performance. The PG&E project used zip lock bags with a corrugate board insulator to attach tags to metal parts, but this could possibly interfere with tool operation. Another alternative that may be considered in the future are metal tags manufactured by companies like Xerfy that are designed to work attached directly to metal. These tags are much more expensive however and are not recommended at this phase of the project.

From the information gained from the PG&E project we were able to limit our search for a reader to either the Alien (ALR-9900) or the Sirit (Infinity 510). Programmers who were working at PolyGAIT during the Spring quarter of 2010 determined that the security was better on the Sirit Reader so every team made the decision to recommend it over the Alien. The Poynting antennas use on the projects were recommended by Sirit as the best performing antennas for this project and became our antennas for the project.

While the three systems met many of the Navy's requirements, there were also several deliverables the Navy asked us to address as we move onto the next phase of the project. These deliverables included the use of unauthorized hardware, the database design choices, and the individuality of the three projects.

Throughout the quarter the teams utilized information learned by PolyGAIT through previous projects. During our report out to the Navy we learned that there was an approved parts list for the Navy and other DoD branches called the NMCI list, and items that did not appear on this list were not allowed to be used at their facilities. The Sirit reader we recommended for use for our projects had not yet been approved for the NMCI list so we could not move on with an implementation until it was approved. We contacted Sirit and worked with them to get their product through the necessary testing to be approved for the NMCI list.

Another deliverable was that we needed to change our database program. All three projects were programmed in Microsoft Access Database, because each team had members that had strong backgrounds in the program from IME 312. The Navy, however, decided that they would like to have the database in SQL server. This was because the navy's own RFID development team in Florida designed their system around a SOL database with a C-sharp user interface and in order to guarantee that the navy would have onsite support available it would be necessary to program our systems using the same language.

The largest deliverable that the Navy asked for was the integration of the three purposed projects. While each system was in a separate area of the MPSF, the similarities between the systems would allow them to be integrated together into a single RFID solution. The possible cost savings and ability to test a single prototype in multiple areas of the facility

made this a desirable solution and would be the main deliverable for Phase 2. The following table summarizes the main components used to build each Phase 1 system and the costs associated with each one.

Materials and Cost

(net costs excluding backup equipment taxes and overhead)

System	Section	Item	Quantity	Cost ($ Ea.)
Tool Room	Check-in/ Check-out	Poynting PATCH-A0025 Antenna	2	$75
		Sirit Infinity 510 RFID Reader	1	$1,850
		Rubbermaid Service Cart	1	$163
		Panasonic Toughbook cf-52	1	$4,400
	Unauthorized Check-out	Poynting PATCH-A0025 Antenna	3	$75
		Sirit Infinity 510 RFID Reader	1	$1,850
		Panasonic Toughbook cf-52	1	$4,400
	Inventory	Poynting PATCH-A0025 Antenna	1	$75
		Shares 2 antennas, Reader, and computer with Check-in Check-out		
	Total			**$13,113**
Minor Assets		Poynting PATCH-A0025 Antenna	4	$75
		Sirit Infinity 510 RFID Reader	1	$1,850
		Rubbermaid Service Cart	1	$124
		Panasonic Toughbook cf-52	1	$4,400
	Total			**$6,674**
Spare Parts		Poynting PATCH-A0025 Antenna	3	$75
		Sirit Infinity 510 RFID Reader	1	$1,850
		Metal Cart	1	$108
		Panasonic Toughbook cf-52	1	$4,400
	Total			**$6,583**

Conclusions

Due to the successful demonstrations of the three projects, the Navy asked PolyGAIT to begin Phase 2 and integrate the three projects into a single mobile system. This new system would have a graphic user interface written in C-Sharp and a SQL database that would allow easier integration with a similar system being developed by the Navy's Panama City Facility that

was also developing an RFID system for the LCSs. Along with the new system Phase 2 would develop an implementation plan for tagging the item on site so that a full scale test of the prototype could be carried out.

Phase 2

Problem Statement

Phase 1 did not create a single integrated system that could be maintained by the Navy's staff, Phase 2 needs to create a prototype that will integrate all of the physical components of Phase 1 into a single mobile cart as well as combine the three Microsoft Access Databases into a single SQL Database with a graphic user interface written in C Sharp.

Method

Phase 2 began in the summer of 2010 when Tyler Peabody and I were employed by Cierra Logistics Consulting to work more closely with the Navy. Our job was to determine how to integrate the three previous systems together and develop an initial implementation plan. Throughout the summer we continued to work very closely with PolyGAIT where Nathan Helenihi was employed as the software programmer for the project.

When the Fall Quarter of 2010 began, the Cal Poly IME 457 Class carried out the plan we developed over the summer. Tyler Peabody and I assisted professors Tali Freed and Larry Rinzel as project managers and brought the rest of the project team up to speed with information about the facility, the development of the physical system, and the plan for the implementation event that was scheduled for December of 2010. The team members, who are listed below, then developed and built a final prototype that integrated the three systems and could be easily transported to Port Hueneme to be tested and installed as a prototype. At the same time we worked with Nathan to finalize the design, interface, and structure of the final program. We reviewed the functionality of the previous three projects and combined them to create a single solution for the prototype implementation.

The Team

Brian Davenport(IE), Jake Rucker(IE), Chris Parker(EE), Mike Warren-Angelucci(IE), Tyler Peabody(IE), Stewart Peugh(IE), Nathan Helenihi(CPE), Tali Freed, and Larry Rinzel

System Description

The Phase 2 project integrated the three Phase 1 projects together to create a single mobile solution that could accomplish all of the same tasks and operate independent of a network or continuous power supply. Each of the previous systems will now be referred to as modules of the Phase 2 system.

Figure 17shows the base of the cart which, along with the computer, creates the core building block for the three modules. In Appendix B you can also see the power system designed for the project by Chris Parker. The current from the battery runs through an inverter to supply power to the reader and computer when the system is unplugged from the wall. When it is plugged back into the wall it will automatically start charging the battery and turn the inverter off. The power system is the only part of the system that requires no operator interaction to operate for each system. You can also see in

Figure 17that the antenna ports on the reader are color coded to assist the operator in knowing what antenna to attach to which port for each system.

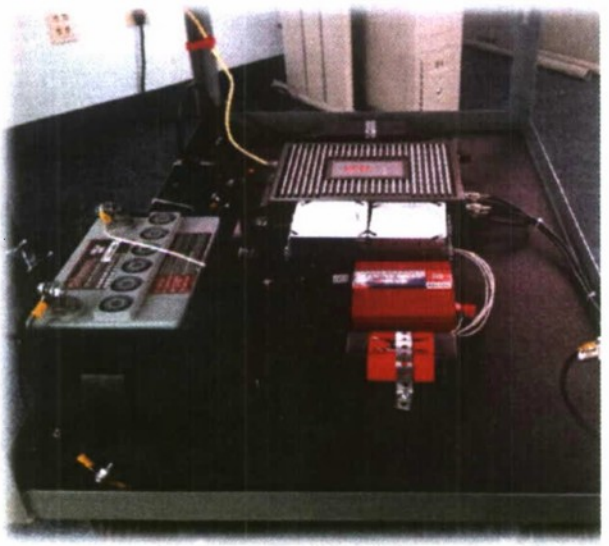


Figure 17: Bottom of Cart

When the cart is configured for minor assets it has 2 antennas, one on either side of it, so that it can easily read all of the items in a room. The

antenna attached to the left side of the cart is removable and allows for scanning in areas that the antennas might not reach, such as inside metal drawers or behind metal desks.

Figure 18 to the right shows the cart in it Tool/Spare parts inventory check set up. The only changed in hardware from the Minor Asset configuration to this one is the addition of a painter pole, with two antennas attached, to the holder on the front of the cart as seen in Figure 18 Once the pole is secured, the antenna cables are attached to their corresponding ports on the reader. The pole used in Phase 2 is sturdier, easier to adjust, and less likely to accidentally rotate than the pole used in Phase 1. It is used to get an antenna above the shelving units to achieve reads on the top shelf of each room. In addition to the antenna at the top of the pole there are also antennas attached on the cart and mid-way up the pole to read the other shelves in this area. The antenna on the left side of the cart is the same one utilized for the Minor Assets area and can be detached and used to read items that might not be easily read.

Figure 18: Cart with Pole

Figure 19 shows where the portal will be attached for the Unauthorized Check-in/Check-out system. The only change to the hardware from phase one for this part of the project is the integration into the main reader instead of having its own independent reader. The cart needs to be rewired for this part of the system to work but because of the color coded wires this is an easy task. The portal cross function has not been fully tested for this system so it may be necessary to add two photo eyes to the portal to correctly trigger the reader.

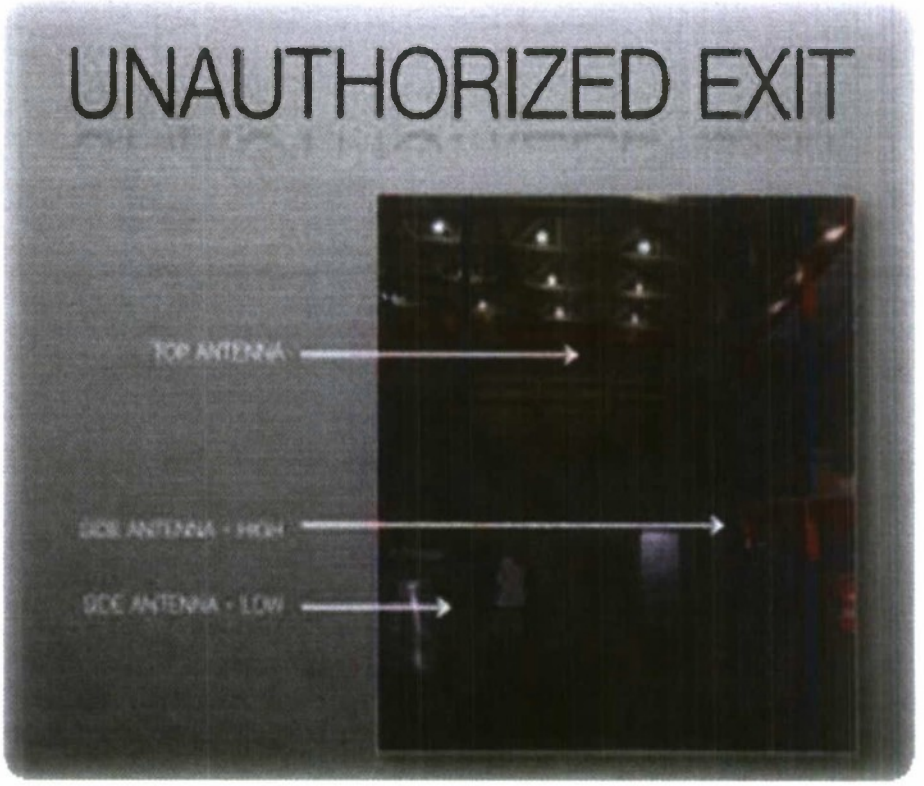


Figure 19: Unauthorized Exit Portal Placement

Materials and Cost

Part	Quantity	Cost (ea.)
Poynting PATCH-A0025 Antenna	7	$75
Sirit Infinity 510 RFID Reader	1	$1,850
Metal Cart	1	$108
Panasonic Toughbook (model:cf-52)	1	$4,400
Battery	1	$250
Inverter	1	$50
Painter Pole	1	$80
Misc. Supplies	1	$400
Total		**$7,663**

Ergonomic Concerns

Phase 1 concentrated on bringing a functional and cost effective prototype to the Navy. While this goal was met it is always important to look for ways to improve the user experience. Phase 2 addressed some of the Ergonomic concerns we encountered in Phase one, as well as brought to light new ones that should be addressed in further implementations of this project. This section discusses opportunities we saw for improving the user experience and what steps we took to continuously improve the usability of our system.

Mobile Cart

One of main considerations when choosing to pursue a cart design was the ease of use. Alternative systems researched included a backpack or handheld system, each of which would have accomplished the monetary and performance gains that the cart provides over a stationary system. Unfortunately the necessary equipment we were using would make both alternatives relatively heavy and difficult to use, especially with the extinction pole we decided to use to read the upper shelving units. Previous

PolyGAIT Inventory systems have been based on mobile systems that can be pushed around, but had a much smaller foot print. A cart not only provides a stable base to attach the 20ft pole to, but also a convenient desk for placing items on when items are being added to the system or checked out. One unfortunate feature of our cart is its turning capability. The wheels that allow the cart to turn are located on the back part of the cart. It is more difficult to operate a cart with wheels on this side of the cart in tight spaces so it is recommended that they are moved or a different cart is used in future implementation.

Writing Surface:

In order to provide a complete inventory checking solution we wanted a cart that could also be used to fill out paper work and carry out other management tasks that did not have a define location at the facility. Unfortunately the cart that was used does not have a flat writing/typing surface. It has a ledge that is in the way, and an easy remedy would be to provide an additional piece of material that could be placed on the cart to provide a flat writing surface. For future projects it is also recommended that the height of the cart is examined. It may not be optimal of all operators at the facility.

Cable Management

To enable the system to be changed between the three modules it was necessary to have extra cables on the cart. These can not only be unsightly but can be hazardous if not controlled. The cables are zip tied whenever possible to prevent them from becoming an accidental hazard and excess cables are velcroed together.

Ability to Charge Cart

An important part of the cart system is the need for mobility. Phase 1 did not include mobile power systems into their designs. Phase 2 added a 12v battery and an inverter to power the reader and charge the laptop. By including a battery system the cart can be used without the restriction of being plugged into the wall. The battery system can also operate with minimal operator interaction. It is designed to

automatically switch from charging the battery to powering the system when the charger is plugged or unplugged from the wall. This allows for continuous operation of the RFID system, when the power state changes.

Cart Configurability

Connecting the antennas is difficult and potentially confusing. Because our system will need to change to multiple configurations we color coded the wire connectors to assist the operators in their change over. This should speed up the change over by simplifying the process as well as avoid damage to the system by improperly wiring the components.

Pole Adjustability:

The pole is difficult to adjust because the adjustments are made at the middle of the pole which is 8 feet above the ground. The pole can be adjusted while on the ground but then becomes cumbersome and potentially dangerous to attach to the cart. A different extending mechanism such as a pulley or gears could ease this process considerably. It would also be beneficial to find a way to get reads on the top shelves without having to add and remove some of the hardware. This could be done with either a more compact telescoping pole or another system entirely.

Handheld Reader:

Currently the only handheld system for the cart is the detachable antenna on the left side of the cart. It is necessary for our system to have an antenna that can be manipulated over drawers to read the items that are inside of them. The current antenna is bulky and difficult to handle. The cable also tangs easily and is difficult to manage. Future designs may include a wireless hand held reader to simplify the handheld reading process.

Maintenance Procedures:

Operator may find maintenance and replacement to the components on the lower part of the cart to be difficult because they are bolted down. Wing nuts could be used if this may be a process that happened regularly, but we do not for see this

happening very often. Another consideration is that it is smart to design the system in a way that will reduce the need for servicing. Future iterations of this cart should consider securing the lower section of the cart to protect its contents from accidental damage or removal.

Analysis

Tag Testing

During a previous PolyGAIT RFID installation at PGE, it was determined that the Sirit RSI-647 tag was the best performing tag on the market for an industrial environment that has a mixture of metal and non-metal parts. This is what we needed for our application and because the PG&E installation was successful we decided to also use this tag. When we tried to place an order, however, we learned that the tag was no longer available and that we needed to order a different tag. We began our search for a replacement tag with tags recommended by the supplier as similar or higher performing than the RSI-647. We also selected tags that were recommended from other industry professionals. The tags we decided to test were the Alien Squiggle, the Omni-squiggle, the RSI 647 (baseline used at PG&E), and the RSI 674.

Design of Experiments

It was necessary to set up an efficient experiment so that we could decide on a new tag and place an order in time for our planed implementation. The factors we needed to test in our experiment were the 4 different tag types and three separate orientations of each tag (X, Y, and Z axis of the tag) to determine which tag provided the best performance in the widest ranges of orientation. We decided to use 15 different tags of each tag type and assumed that the companies had enough quality control on the tags so that we would have statistically similar read abilities and be counted as a replication of each tag type. By using a General Linear Model 15 tags should be sufficient to assume data normality. We used Minitab to randomize our run order and ran a total of 180 tests.

Procedure

To perform each test we set up a reader and antenna at full power on a cart connected to a test computer running putty, a software for interfacing with the reader. The antenna was attached to the plastic cart and held so that the front of the reader was perpendicular to the ground. A wooden fixture was then used to hold the tag being tested in the proper orientation at the same height as the antenna. The reader was then placed into continuous read mode and the fixture holding the tag was moved away from the antenna until the reads stopped appearing on the putty interface. At this point the fixture was slowly moved back toward the antenna until the reads started again. Finally the fixture was again slowly moved away and then back until the best estimation of the maximum read distance away from the reader was determined. The distance from the tag IC to the back of the antenna was then recorded as our maximum read distance for that sample.

Test Location

We originally tried to perform our test in the hallway of building 4 at Cal Poly. We needed a space that was long enough for us to max out the distance on the reader and wanted to perform the test inside because it approximated the environment that we would encounter at the MPSF. Unfortunately in our initial tests we found discrepancies in the data, and had difficulty determining where the tag signal was actually lost. We also found that some of the tags read so well that the hallway was not long enough to determine the maximum read distance of the sample. The main problem we found, however, was that after some tags reached the end of their read range and we lost their signal, the signal would reappear when they were moved further away from the reader. We believe this anomaly is caused by the metal walls on either side of the hallway which may reflect the signal and causing its amplitude to change in different areas of the hallway due to constructive and destructive interference. To avoid this problem we moved our testing outside where we hypothesized that there would be no interference issues due to metal and re-started our experiment. After some initial

testing we thought the data appeared to be behaving normally and then proceeded with the experiment.

Analysis

Because of time constraints we had to split the test up-into two days, so in addition of our initial factors of tag type and orientation we also looked at the test day as a factor to see if it was significant or caused any interactions as well. After collecting all of the data we used a General Linear Model (GLM) in Minitab to see what factors were significant and if there were any interactions. The results of this analysis can be seen in Appendix A.

The plot of residuals, in Appendix A, shows that our data is not exactly normal, but it is close enough to assume normality for our experiment. From the results of the GLM we ran in Minitab we found the P-values for all of our factors and interactions were less than .05, which means that they are all significant. This was expected for the tag type and orientation but we worried that it invalidated our test because our test days were statistically significantly different. After speaking with Tali and discussing within our team we realized that there was a change in the results when the tag was in direct sunlight or in the shade. During the second day the tags were not exposed to as much sun-light and this could be the explanation for the higher performance (as seen in the main effects and interaction plots in Appendix A). A future senior project was planned to determine the effects of sunlight on RFID tag performance. The preliminary results of this project show that all RFID tags are negatively affected by sunlight, so we will assume that all of the tags were affected in a similar negative fashion on the first day and we will continue to use our test.

By again looking at the main effects and Interaction plots we can tell that the RSI-674 (Monza 3) is the best tag in every orientation but ... as well as on both days. Because it is unlikely for the tags to appear in the "Y" orientation very often in an industrial setting we can conclude that the RSI-674 will provide us with the longest read range out of the tags selected in this group. This can be confirmed with the results from

the turkey test run with a 95% confidence interval which shows the read distance of the RSI-674 as statistically significantly different from the other 3 tag types with a mean read distance much larger than the others. It is also important to note that the RSI-647 (PG&E) was the lowest performing tag in every orientation and on both days. As this is our baseline tag, we can conclude that all three of the other tags have greater read distances and will at least be comparable to the performance of the RSI-647 at PG&E.

Conclusions

From our data we concluded that the RSI-674 was the superior tag in all three of the orientation and therefore would be the best tag for our application. While there was an interaction between the two days and the three orientations, we believe this was because the tags exposure to sunlight during our experiment. We have recently learned that RFID tags are negatively affected by all wavelengths of light. Our tags were exposed to less light on the second day and this correlates with the recent findings so we are confident that our tags will perform at least this well in an indoor environment.

While we did have to move the test outside because of the interference created inside we think this actually provided us with a more conservative estimate of maximum read distance. When we were testing inside there were areas that the tags could not be read, but as soon as you moved out of this area the maximum read range actually increased. Because each of our systems is on a mobile cart that will be moving through the facility, the distance and orientation each tag is to the reader will constantly be changing and we assume that each tag will not stay in an unreadable zone for the entirety of a test and that the interference we see from a metal environment will actually work to our benefit, allowing us to read tags that are further away from our antennas than if they were in a non-metal environment. The PG&E team also noted that they noticed similar increases in read distances due to a metal environment, so we feel confident that we have a conservative estimate. No formal testing has been done yet to quantify this observation.

Future Considerations

To allow us to better understand the performance of the RSI-674 it would be beneficial to conduct tests in a warehouse environment similar to the Navy's. There may be interactions that allow us to build a more efficient cart or that require us to adjust some of our settings. We did not leave enough time to conduct in depth testing of the tag to fully quantify the capabilities of our system because we planned to do these on site after the prototype was brought to the MPSF. This knowledge will be beneficial to future projects, however, regardless of where it is conducted so it should be performed. It would also be beneficial to quantify the performance of the tag in on-metal or near-metal situations. All three modules will need to perform in metal, non-metal, and near-metal environments. If the performance is drastically different for the system in these different environments we may have further considerations for making a complete system.

Simulation:

System Description

Two simulations were developed to demonstrate the inventory processes used by the United States Navy at the Mission Package Support Facility (MPSF) for the Littoral Combat Ships (LCS) at Port Hueneme in Oxnard, California. The first model demonstrates a conventional paper based inventory management system, where inventory is recorded by hand, that they currently use. The second model demonstrates the same process utilizing pRFID for data collection. Because the facility is not currently in use, a simulation is necessary to demonstrate performance metrics of the current system so we may compare them to the purposed system. Data for the simulations were gathered from the MPSF, similar military facilities, and data from past implementations.

Objective

The simulations will demonstrate the time difference between an inventory management system that is recorded by hand and one that utilizes pRFID.

Known Data

From the staff at the Mission Package Support Facility we learned that there are three separate inventory systems that they wish to track and manage. For the purpose of the simulation we will focus only on the Minor Assets in their inventory. Minor Assets are items that are worth $300 to $10,000 and are generally found in the office areas of the MPSF. We will further focus on the 9 rooms that comprise the office area of the facility. We simulated in our model an employee walking at .5 miles per hour through the facility and the distances between each room was measured from CAD drawings provided to us by the Navy. Table 1 below describes the distances between each room.

Table 1: Distance Between Rooms

Path	Distance (ft.)
Clerk's Desk to Conference room	100
Conference room to Operations	-
Operations to Office 1	100
Office 1 to Administration	100
Administration to Office Supply Room	150
Office Supply Room to Main Office	20
Main Office to Engineering	-
Engineering to File Storage	20
File Storage to Server Room	-
Server room to Clerk's desk	10

We obtained additional information from a report conducted by Heather Bellardo, Niosha Behnam, Max Peak, and Imran Chaugule in IME 545 - Advanced Simulation. Their project, Asset Management Simulation, compared the times to conduct an inventory check by hand, with pRFID, and with Active RFID for the Cal Poly Fiscal Services Department. Because they also modeled an office environment we decided to utilize their data in our simulation. They found the time it took to find and record and item by hand were 49.73 seconds and the time it took to record an item using pRFID was 26.69 seconds. They also recorded 771 assets from 65 rooms which is a little over 11 items per room. Our simulation used this information to determine the number of assets per

room because we do not yet have a complete inventory list of minor assets from the navy.

Logical Model

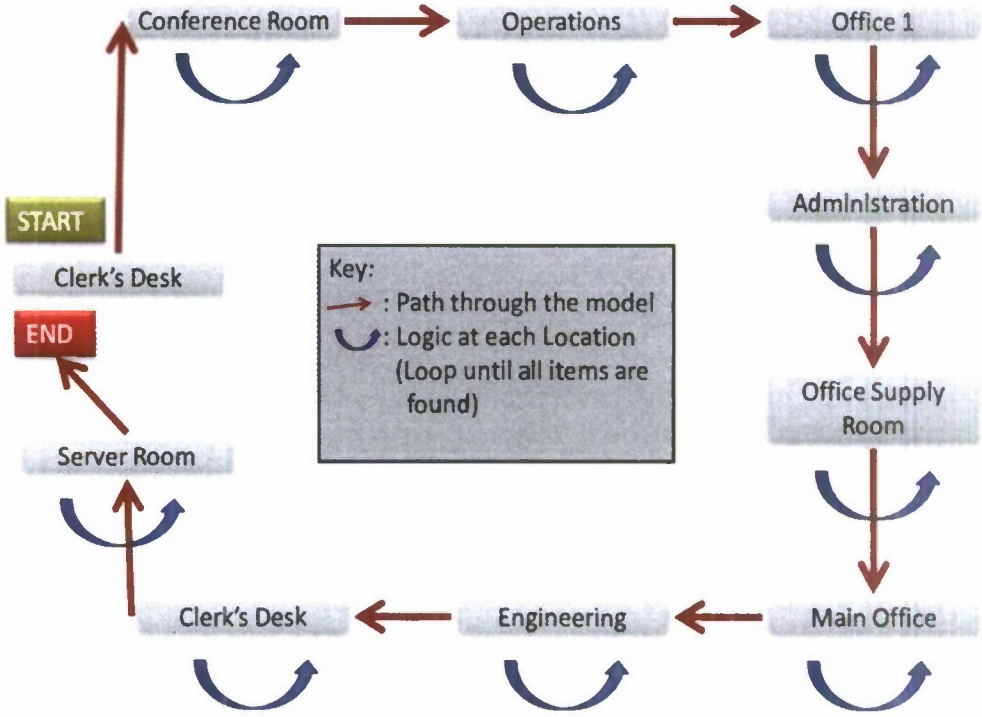


Figure 20: Logical Model

The two simulations follow the same generic logical model that is shown in Figure 20. An entity (which represents the employee managing the Inventory) enters the system at the location "Clerk's Desk". It then proceeds to follow the red arrows from location to location pausing at each to complete the processing logic contained within. The processing logic for each location is the same. An entity will check to see if all of the items of that room have been checked. If they have not, the entity will look for an item and mark it as done when it is found. The time it takes to find each item follows a distribution specific to each simulation and each room contains a defined number of items.

Model Brief

Figure 21, below, is a screenshot of the simulation being run. The logical model we created was placed over a layout we received of the facility to produce a more realistic looking simulation. The entity was then set to "walk" between each room at 0.5 MPH. Two counters were also placed on the simulation to show the number of items that have been recorded and the total time that has been spent looking for items. We simulated the time to check 9 rooms with 10 items per room.

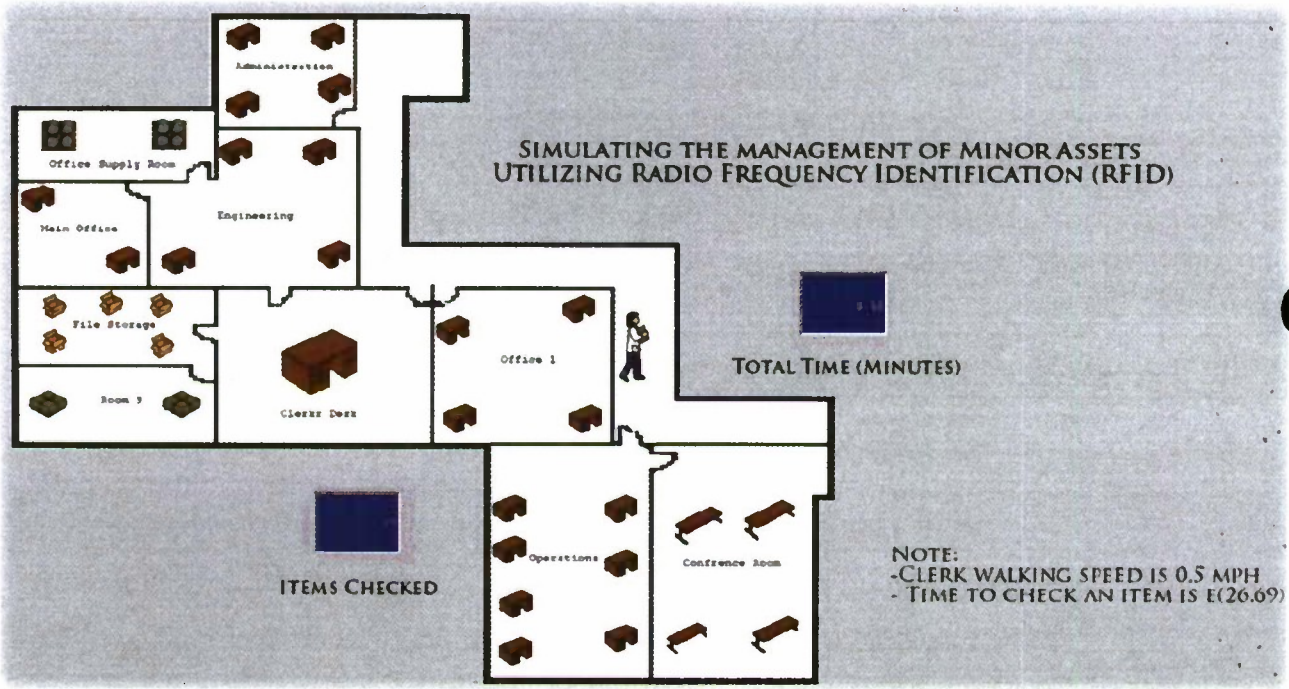


Figure 21: Screenshot of Model

Experimentation

The time difference between the two methods could be determined by comparing the average time it took to record all of the items in each simulation. To be confident in our average we needed to determine how many replications the simulation needed to run to have an average that provided a 95% confident interval. Figure 22 shows the equation used for these calculations.

$$n = \left[\frac{Z_{\alpha/2} * S}{H.W.}\right]^2$$

α = .05 confidence interval
$Z_{\alpha/2}$ = 1.96
S= 7.039 (from initial sample of 30 replications)
H.W. = 2

Figure 22: Half Width Calculation

Each Simulation was run 30 times to obtain a sample set of data. This data was then used to determine the standard deviation of each process. It was found that to obtain a 95% confidence interval for the simulation for checking an item by hand the simulation needed to run 68 times and the RFID simulation needed to be run 139 times. It did not hurt to run each simulation more so each simulation was run 200 times and the averages for total time to check all of the items in each system. We determined that it would take 86 minutes to check the entire inventory by hand and 51 minutes to check the inventory with RFID.

Data Validity

The simulations accurately model the two processes with the data we have but it does not appear to be completely valid. We currently are using an exponential distribution in the model to model the time it takes to check an item. This does not work because there is a possibility that the distribution will return a value of zero. It will never take zero time to record an item so this is not valid. Also an RFID system is much better in an item dense environment. The as the number of items increase in a room the time it takes to check those items by hand will also increase proportionally. However, for an RFID system the time will not increase very much because the tags information will be gathered in a fraction of a second and the processing will take equally as long. While we have not conducted testing to quantify the increase in time, we have not noticed a significant time increase (>1 second) between reading one item and reading around 100 items. This is not model in the current model, but because the time difference will only increase between the two methods we will continue to use the results from the simulation in our analysis.

Economic comparison and ROI:

Using the information we gained in the simulation we learned that the average time to complete an inventory check for 90 items in the facility would take 85 minutes by hand and 51 minutes using RFID. If the operator used to perform these checks costs the company $60 per hour. The cost to perform the Inventory check by hand would be $85 and with RFID would be $51. This results in a savings of $34 from time alone per check. While this may not seem that large when we consider that the Facility has 6,000 items the savings increases to $2,311 per inventory check.

We learned that at the MPSF they perform a Tri-annual inventory check. This would result in a monetary savings of $6,933. The following table summarizes the total cost of installing the prototype at the facility.

Item	Quantity	Cost (ea.)
Poynting PATCH-A0025 Antenna	7	$75.00
Sirit Infinity 510 RFID Reader	1	$1,850.00
Metal Cart	1	$108.00
Panasonic Toughbook (model:cf-52)	1	$4,400.00
Battery	1	$250.00
Inverter	1	$50.00
Painter Pole	1	$80.00
Misc. Supplies	1	$400.00
Zebra Printer	1	$7,000.00
Tags (RSI-674)	6,000	$0.17
Software design costs	1	$2,000.00
Labor	1	$20,800.00
Total Cost		**$38,483.00**

With a total cost of installation at $38,483 and the savings per year currently at $6,933 the return on investment would be 5.55 years. This how ever does not account for the reduction in shrinkage experienced at these facilities.

After speaking with the manager of Minor Assets at Port Hueneme we learned that during a recent inventory check they were unable to find 20% of the 1,200 items they thought they should find in the facility using 3 people over 3 days. This equates to about shrinkage of about 20%. The following table summarizes the value of the inventory in each of the three

areas of the MPSF, the value of the items currently lost to shrinkage, and the value of the items that would be lost to shrinkage under a RFID system. The total savings resulting in shrinkage reduction is $1,900,000. When you take this into consideration the ROI reduces to .02 years.

Inventory Value	Assets	$10,000,000
	Tools	$5,000,000
	Parts	$4,000,000
Current Shrinkage (20%)	Assets	$2,000,000
	Tools	$1,000,000
	Parts	$800,000
RFID Shrinkage ~ (10%)	Assets	$1,000,000
	Tools	$500,000
	Parts	$400,000
Total savings =		**$1,900,000**

Implementation plan

As the Fall quarter of 2010 drew to a close it became clear that we would not be able to travel down to Port Hueneme for a prototype installation. As such we left several to-do list items opened that need to be addressed before the implementation could begin, but with the knowledge gained from this project we know how to address them.

The primary time constraint to move forward with this project is obtaining the tags. There is a minimum 4-6 week lead time on the RSI-674 tags. It may be possible to obtain some in a shorter time frame if the supplier has any on hand, but planning on 6 weeks of time will give the team time to prepare the other parts of the implementation. The Zebra printer PolyGAIT uses to program and print RFID tags has not yet been mastered for the RSI-674 tag so it is important to obtain a sample roll to practice printing and programming tags. This should take no longer than 5 days of work to prepare, but is a potentially large bottleneck because every tag must run through the printer.

A work around for this problem developed during the quarter may provide a better solution that can avoid the bottleneck and provide grater flexibility. Instead of programming the tags, the inventory list can be obtained from the navy before hand and a database of items

can be generated beforehand. Labels can then be printed out using a normal office printer and brought to the facility for the implementation. During the implementation when an RFID tag is placed on an item the label can then be placed over the RFID tag and the tags preprogrammed number can be related to the corresponding item in the database. This provides greater flexibility in many ways. The primary reason is that you do not need to worry about assigning a specific RFID tag to a product; this is done with software so a damaged RFID tag can easily be replaced and up-dated in the system. Another benefit is that the labels are easy to print using existing hardware. More complex labels can be created utilizing color and other designs to differentiate the product and in the event of a design change you would not need to place another RFID tag on the item, just another label.

A final consideration for an implementation of the prototype is the time it will take for the tagging event. It will take multiple skilled professionals to assist in the tagging event and it will be very invasive to the daily work cycle at the MPSF. Forklift drivers will need to be available to get and replace parts from the tool room and spare parts area. Inventory experts will need to be on-hand to verify that each part is being identified correctly. Every item on each inventory list will need to be tags so they will be unavailable for used while they are being tagged. Multiple students should also be on-hand to find items, verify RFID tag readability, place tags, and manage the data base. The system we designed uses a single database that communicated with a single reader, so it may be necessary to use multiple systems as to reduce this bottle neck. Planning and coordination with the staff at the MPSF can take place over the 6 weeks it takes to obtain the tags.

Conclusions:

After a year of work, this project has allowed for a number of inter-disciplinary teams to work on a common problem that is experienced by facilities in many different industries around the world, high inventory management costs and shrinkage. This project developed a physical prototype and a process for reducing these problems. We know it is possible to increase the visibility of the items in your inventory by reducing the cost to perform an inventory check, which then allows you to perform them more frequently. By doing this you can reduce the

number of items lost to shrinkage and currently this is where the largest cost savings can be found. With further testing of our system we can prove that it can reads of all of the items in the facility during a single inventory check and prove to be a tool that facilitates a fast and cost effective solution to a problem that not only costs money but can potentially harm readiness levels and ability for your facility to perform its function.

References

1. Bacon Robert, NAVSUP CIO/Navy AIT Program Manager, "Navy AIT Advancing Technology in Support of the Warfighter" https://www.navsup.navy.mil/scnewsletter/2009/sep-oct/scaround10
2. Behnam, Niosha, H. Bellardo, I. Chaugule, and M. Peak. "Asset Management Simulation for Cal Poly." IME 545 Report, California Polytechnic State University, San Luis Obispo, 2007.
3. Diablo Canyon Power Site Uses RFID to Track Items http://www.rfidjournal.com/article/print/7787
4. "EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860Mhz - 960Mhz." EPC Global 1.2.0 (2008). Print.
5. Estevez, Alan. "DOD Quantifies Payback From RFID." Proc. of RFID Journal Live! Comp. Jonathan Collins. 2006.
6. Freed, Tali, K. Dodson, M. Krist, L. Rinzel, and B. Taylor "Comparison of Active and Passive RFID Technology for Asset Checkout System," White Paper, PolyGAIT.
7. Freed, Tali, K. Dodson, M. Krist, L. Rinzel, and T. Hunter. PolyGAIT/DCPP RFID Implementation Project. White Paper. California Polytechnic State University, 2009.
8. Freed, Tali, K. Doerr, R. Chartier, E. Martinez. "Workforce Modularity for Naval Operations," White Paper, 2005.
9. Freed, Tali, J. Hall, M. Krist, L. Rinzel, B. Taylor, and K. Dodson."Comparative Evaluation of Passive and Active RFID for Asset Tracking." White Paper, PolyGAIT.
10. Freed, Tali. "Reflection in Organizational and Work Redesign: Learning from Redesign Process at a US Defense Company," with A.B. Shani, M. Stebbins, and K.H. Doerr. Productive Reflection and Learning at Work, Docherty Ed., Routledge Publishing Company (2005).
11. Freed, Tali. Rinzel, Larry, S. Swaaley. "The Economics of Tracking Office Assets." White Paper, PolyGAIT.
12. GAO Report 04-305R "Defense Logistics: Preliminary Observations on the Effectiveness of Logistics Activities during Operation Iraqi Freedom", DTD Dec 03 http://www.gao.gov/new.items/d04305r.pdf
13. Lee, Young M., F. Cheng, and Y. T. Leung. "EXPLORING THE IMPACT OF RFID ON SUPPLY CHAIN DYNAMICS." Proc. of 2004 Winter Simulation Conference, IBM T.J. Watson Research Center, New York.
14. Michael, Katina, and L. McCathie. "The Pros and Cons of RFID in Supply Chain

Management." International Conference on Mobile Business (ICMB'05). ICMB, 2005. 623-29.

15. NMCI Approved Device List https://www.neco.navy.mil/upload/N00189/N0018909TZ1250001NMCI_CERTIFIED_DEVICE_LIST.xls
16. Sanghera, Paul. RFID+: CompTIA RFID+ Study Guide and Practice Exam. Rockland, MA: Syngress, 2007.
17. Saygin, C. "Adaptive Inventory Management Using RFID Data." The International Journal of Advanced Manufacturing Technology 32 (2007): 1045-051.
18. Schuster, Edmund W., S. J. Allen, and D. L. Brock. Global RFID the Value of the EPCglobal Network for Supply Chain Management. Berlin: Springer, 2007.
19. Successful DoD RFID Implementations http://www.acq.osd.mil/log/rfid/DoD_Panel_Slides.pdf
20. Szmerekovsky, Joseph G., and J. Zhang. "Coordination and Adoption of Item-level RFID with Vendor Managed Inventory." International Journal of Production Economics 114.1 (2008): 388-98.
21. "The US Navy -- Fact File." The U.S. Navy. U.S. Navy, 17 Sept. 2010. Web. 29 Nov. 2010. <http://www.navy.mil/navydata/fact_display.asp?cid=4200&tid=1650&ct=4>.
22. Taylor, Bryce, and T. Freed. "Optimal Layout Coverage of RFID antennae." Proceedings of IERC, 2009.
23. USA. DOD. Department of Defense Standard Practice Military Marking for Shipment and Storage. Vol. CH 4. 2007. MIL-STD-129P.
24. USA. DOD. United States Department of Defense Suppliers' Passive RFID Information Guide. Vol. 14. 2009.

Appendix A: Tag Test

General Linear Model: Distance versus Orientation, Tag Style, Day

Factor	Type	Levels	Values
Orientation	fixed	3	X, Y, Z
Tag Style	fixed	4	Alien Squiggle, Monza3 (RSI 674), Omni-Squiggle, PG&E (RSI 647)
Day	fixed	2	1, 2

Analysis of Variance for Distance, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Orientation	2	12488.4	10579.8	5289.9	107.83	0.000
Tag Style	3	5927.7	6323.9	2108.0	42.97	0.000
Day	1	1610.3	1023.4	1023.4	20.86	0.000
Orientation*Tag Style	6	2761.8	1966.6	327.8	6.68	0.000
Orientation*Day	2	709.2	830.7	415.3	8.47	0.000
Tag Style*Day	3	660.9	660.9	220.3	4.49	0.005
Error	162	7947.4	7947.4	49.1		
Total	179	32105.7				

S = 7.00416 R-Sq = 75.25% R-Sq(adj) = 72.65%

Unusual Observations for Distance

Obs	Distance	Fit	SE Fit	Residual	St Resid	
6	34.9167	21.0329	2.1010	13.8838	2.08	R
12	37.4167	23.3064	2.0304	14.1103	2.10	R
72	38.7500	17.4637	2.4324	21.2863	3.24	R
74	17.7500	4.0497	2.1139	13.7003	2.05	R
83	37.5833	23.3064	2.0304	14.2770	2.13	R
105	11.5000	29.1511	1.8984	-17.6511	-2.62	R
122	11.3333	29.1511	1.8984	-17.8178	-2.64	R
123	7.3333	26.2094	2.1087	-18.8761	-2.83	R
146	8.8333	27.7979	2.4174	-18.9646	-2.88	R
147	15.5833	29.1511	1.8984	-13.5678	-2.01	R
165	12.0833	27.7979	2.4174	-15.7146	-2.39	R
179	8.5833	26.2094	2.1087	-17.6261	-2.64	R

R denotes an observation with a large standardized residual.

Grouping Information Using Tukey Method and 95.0% Confidence

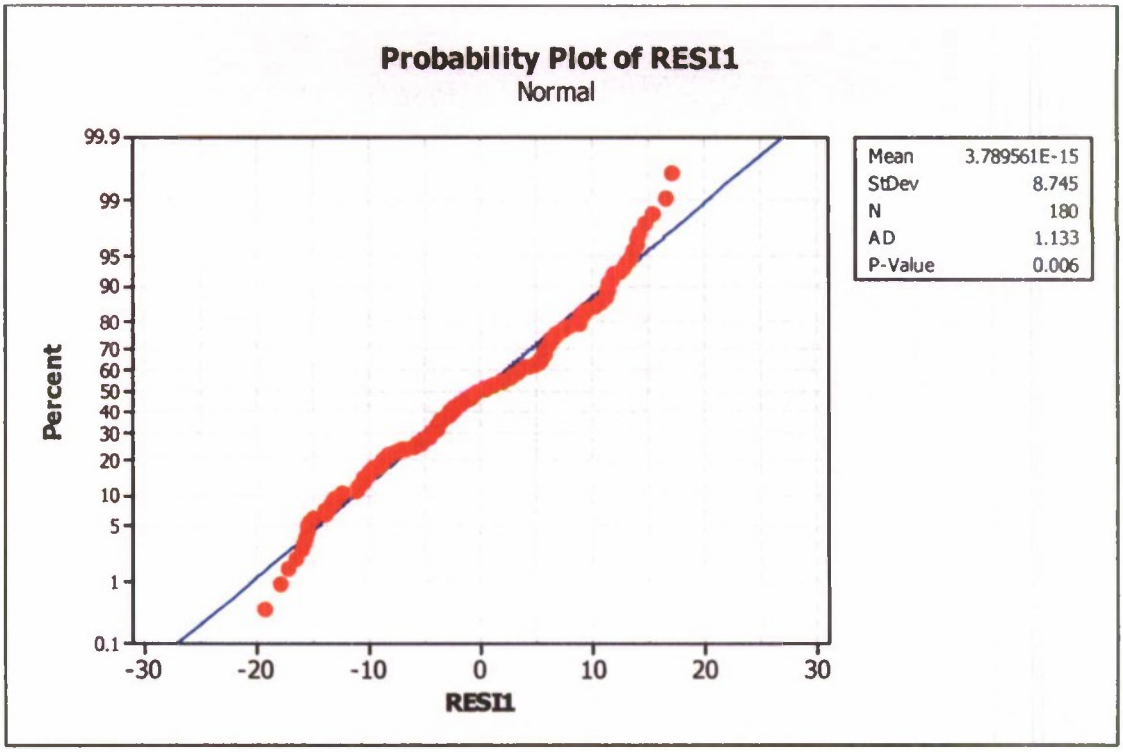
Tag Style	N	Mean	Grouping
Monza3 (RSI 674)	45	24.3	A
Omni-Squiggle	45	19.4	B
Alien Squiggle	45	16.4	B
PG&E (RSI 647)	45	7.3	C

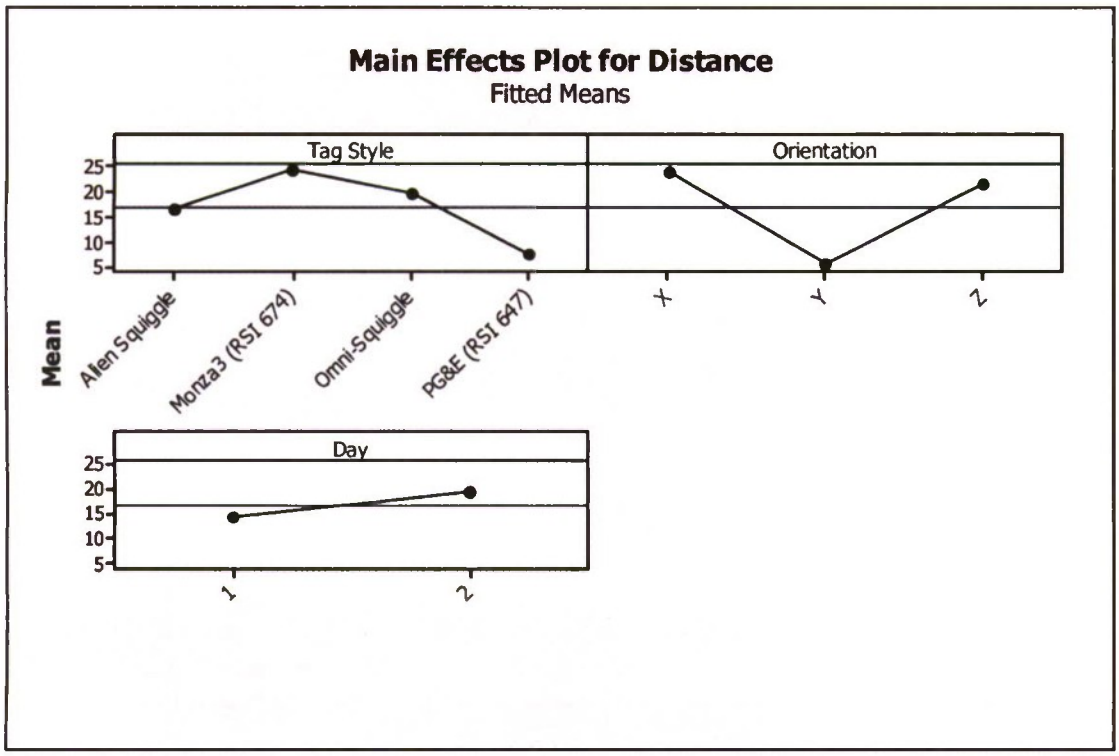
Means that do not share a letter are significantly different.

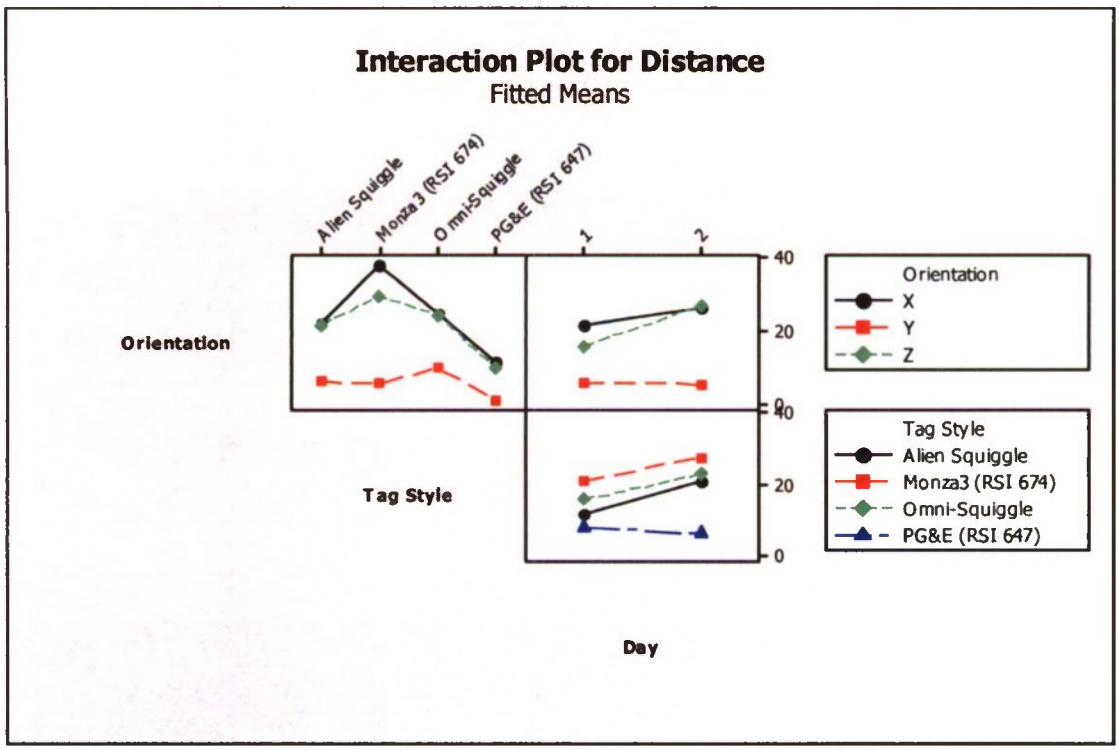
Tukey 95.0% Simultaneous Confidence Intervals
Response Variable Distance
All Pairwise Comparisons among Levels of Tag Style
Tag Style = Alien Squiggle subtracted from:

```
Tag Style          Lower  Center   Upper
Monza3 (RSI 674)    3.97   7.931  11.893
Omni-Squiggle      -0.97   2.977   6.924
PG&E (RSI 647)    -13.09  -9.125  -5.155

Tag Style          -+---------+---------+---------+-----
Monza3 (RSI 674)                        (---*---)
Omni-Squiggle                      (---*---)
PG&E (RSI 647)              (---*---)
                   -+---------+---------+---------+-----
                  -20       -10         0        10


Tag Style = Monza3 (RSI 674)  subtracted from:

Tag Style        Lower  Center   Upper   -+---------+---------+---------+-----
Omni-Squiggle    -8.90   -4.95   -1.01               (---*---)
PG&E (RSI 647)  -21.06  -17.06  -13.05   (---*---)
                                         -+---------+---------+---------+-----
                                        -20       -10         0        10


Tag Style = Omni-Squiggle  subtracted from:

Tag Style        Lower  Center   Upper   -+---------+---------+---------+-----
PG&E (RSI 647)  -16.09  -12.10  -8.109        (---*---)
                                         -+---------+---------+---------+-----
                                        -20       -10         0        10


Grouping Information Using Tukey Method and 95.0% Confidence

Orientation   N  Mean  Grouping
X            60  24.0  A
Z            60  21.3  A
Y            60   5.3     B

Means that do not share a letter are significantly different.


Tukey 95.0% Simultaneous Confidence Intervals
Response Variable Distance
All Pairwise Comparisons among Levels of Orientation
Orientation = X  subtracted from:

Orientation   Lower  Center   Upper  --------+---------+---------+--------
Y           -21.89  -18.67  -15.44  (-*--)
Z            -5.71   -2.60    0.50               (--*-)
                                    --------+---------+---------+--------
                                          -12         0        12


Orientation = Y  subtracted from:

Orientation  Lower  Center  Upper  --------+---------+---------+--------
Z            12.84   16.06  19.29                              (-*--)
                                   --------+---------+---------+--------
                                         -12         0        12
```

Normal plot of Residuals for Distance




Main Effects Plot for Distance
Fitted Means
Tag Style
Orientation
Day
Mean
25
20
15
10
5
Alien Squiggle
Monza3 (RSI 674)
Omni-Squiggle
PG&E (RSI 647)
X
Y
Z
1
2


Interaction Plot for Distance
Fitted Means
Alien Squiggle
Monza3 (RSI 674)
Omni-Squiggle
PG&E (RSI 647)
1
2
Orientation
Tag Style
Day
40
20
0
Orientation
X
Y
Z
Tag Style
Alien Squiggle
Monza3 (RSI 674)
Omni-Squiggle
PG&E (RSI 647)

Appendix B: Electrical Wiring

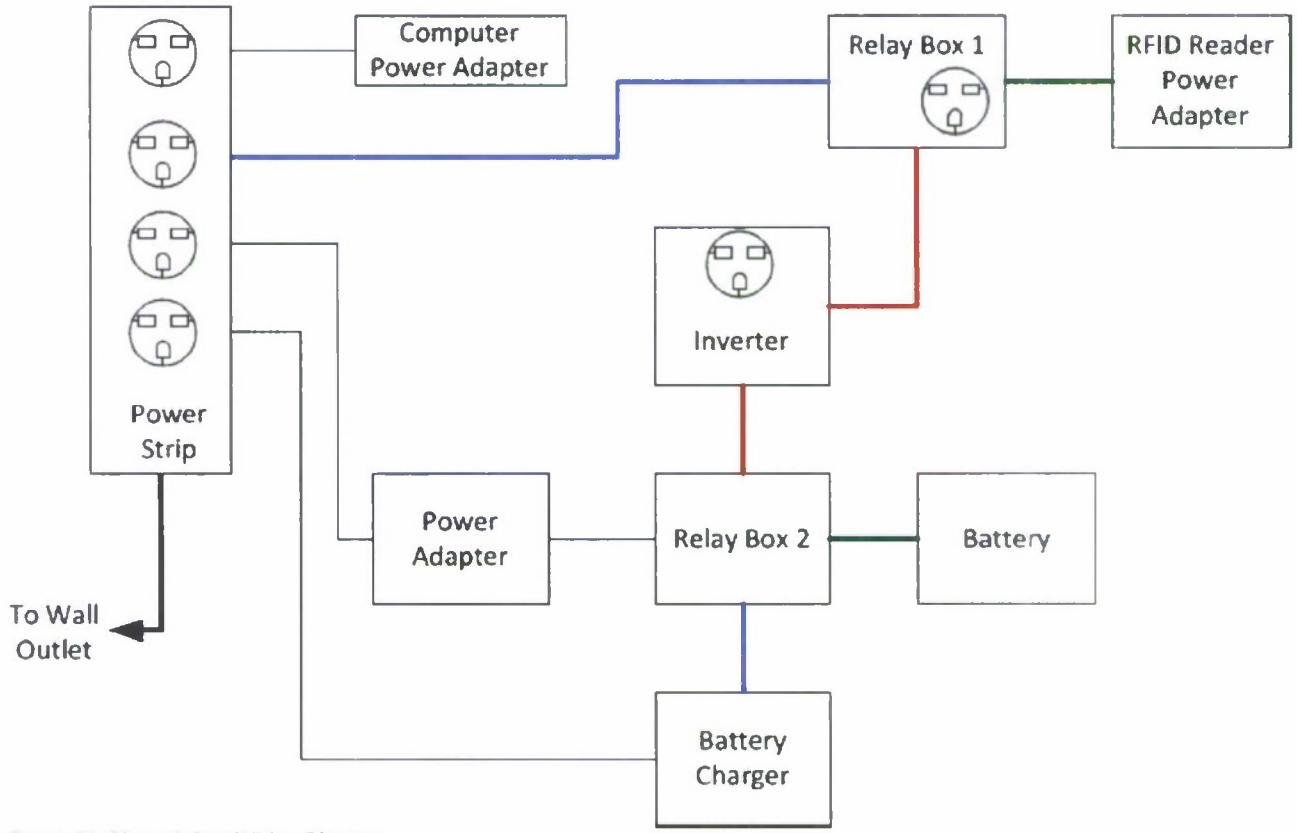


Figure 23: Phase 2 Cart Wiring Diagram

Appendix A

High – Tech Breakfast Forum Agenda

CAL POLY

C3RP
CALIFORNIA CENTRAL COAST
RESEARCH PARTNERSHIP

High-Tech Industry Breakfast Forum

Date:	February 20, 2009
Time:	7:30 am – 9:00 am
Location:	Poly Canyon Village, Aliso Conference Room A007-A014 Cal Poly Campus

You are cordially invited to participate in our on-going breakfast forum for the local high-tech industry, hosted by Dr. Robert Koob, Provost and Vice President for Academic Affairs, and Dr. Susan Opava, Dean of Research and Graduate Programs, at Cal Poly. The purpose of this forum is to bring together CEOs and Chief Technology/Engineering Officers from high-tech companies in the region, on an on-going basis, to foster a mutually beneficial relationship between the university and this important local industry sector.

7:30 am	Continental Breakfast - Small group discussions & networking
7:50 am	Introductions and Overview of Program
8:05 am	Poly Canyon Village Project - Alan Pepe, Associate Director Housing/Director Housing & Business Services
8:15 am	Cal Poly Highlight: Algal Production and Biofuel Development Programs
8:40 am	Company Highlight: Phycotech Corporation – Paul Cathcart, CEO
9:00 am	Adjourn

Please RSVP # of participants to: Dejana Lubura 805-756-5153 or dlubura@calpoly.edu by **February 11, 2009**

CAL POLY

C3RP
CALIFORNIA CENTRAL COAST
RESEARCH PARTNERSHIP

High-Tech Industry Breakfast Forum

Date:	November 20, 2009
Time:	7:30 am – 9:00 am
Location:	Building 197 (Bonderson) Room 104 Cal Poly Campus

You are invited to participate in our on-going breakfast forum for the local high-tech industry, hosted by Susan Opava, Cal Poly's Dean of Research and Graduate Programs, and Jim Dunning, C3RP Administrator. The purpose of this forum is to bring CEOs and Chief Technology/Engineering Officers from high-tech companies in the region to campus on a regular basis, to foster a mutually beneficial relationship between the university and this important local industry sector.

7:30 am	Continental Breakfast - Small group discussions & networking
7:50 am	Introductions
8:00 am	Aerospace Engineering Projects at Cal Poly • Modeling and Performance Verification Tools for the Aerospace Industry– Dr. David Marshall, Aerospace Engineering • CubeSat – Dr. Jordi Puig-Suari, Aerospace Engineering • Satellite Networking Solutions – Dr. John Bellardo, Computer Science
8:45 am	Company Highlight: Stellar Exploration Inc.- Alex Diaz
9:00 am	Adjourn

Please RSVP # of participants to: Dejana Lubura 805-756-5153 or dlubura@calpoly.edu by **November 11, 2009**

P.I. Susan C. Opava, Ph.D.

CAL POLY

C3RP
CALIFORNIA CENTRAL COAST
RESEARCH PARTNERSHIP

High-Tech Industry Breakfast Forum

Date: February 26, 2010
Time: 7:30 am – 9:00 am
Location: Building 197 (Bonderson) Room 104
Cal Poly Campus

You are invited to participate in our on-going breakfast forum for the local high-tech industry, hosted by Susan Opava, Cal Poly's Dean of Research and Graduate Programs, and Jim Dunning, C³RP Administrator. The purpose of this forum is to bring CEOs and Chief Technology/Engineering Officers from high-tech companies in the region to campus on a regular basis, to foster a mutually beneficial relationship between the university and this important local industry sector.

7:30 am	Continental Breakfast - Small group discussions & networking
7:50 am	Introductions
8:00 am	Technology Park Incubator Concept
8:10 am	*Concept to Market – Packaging Focused Research at Cal Poly.* Dr. Jay Singh, Associate Professor of Industrial Technology College of Business
8:25 am	*Graphic Communication is Everyone's Business - Cal Poly Services for Commerce and Industry.* Dr. Harvey Levenson, Department Head Graphic Communications
8:40 am	Company Highlight: Coudray Graphic Technologies - Mark Coudray
9:00 am	Adjourn

RSVP # of participants to: Dejana Lubura 805-756-5153 or dlubura@calpoly.edu by **February 19, 2010**

N00014-08-1-1209

P.I. Susan C. Opava, Ph.D.

CAL POLY

C3RP

CALIFORNIA CENTRAL COAST
RESEARCH PARTNERSHIP

High-Tech Industry Breakfast Forum

Technology Serving the Environment

Date: June 04, 2010
Time: 7:30 am – 9:00 am
Location: Dairy Products Technology Center – Building 18A Cal Poly
(Parking will be provided)

You are invited to participate in our on-going breakfast forum for the local high-tech industry, hosted by Susan Opava, Cal Poly's Dean of Research and Graduate Programs, and Jim Dunning, C³RP Administrator. The purpose of this forum is to bring CEOs and Chief Technology/Engineering Officers from high-tech companies in the region to campus on a regular basis, to foster a mutually beneficial relationship between the university and this important local industry sector.

7:30 am	Continental Breakfast - Small group discussions & networking
7:50 am	Introductions
8:00 am	Environmental Biotechnology Institute – Solving Environmental Problems with Biology. Dr. Chris Kitts
8:20 am	Eco-Inspired Polymer & Coating Solutions/Practices. Dr. Dane Jones
8:40 am	Pacific Energy Company, John Ewan
9:00 am	Adjourn

RSVP # of participants to: Dejana Lubura 805-756-5153 or dlubura@calpoly.edu by **June 01, 2010**

Appendix B

Project Related Thesis & Relevant Publications

Simulation and Experiment on 2PC Transmitted Diffraction Grating for GaN LEDs

Simeon Trieu[*], Xiaomin Jin[*], Chang Xiong[**], Xing Xing Fu[**], Xiangning Kang[**], Guoyi Zhang[**], Bei Zhang[**], Fei Wang[***]

[*]Electrical Engineering Department, 1 Grand Avenue,
California Polytechnic State University, San Luis Obispo, CA, USA, 93407,
xjin@calpoly.edu, phone: 805-7567046, fax: 805-7561458
[**]State Key Laboratory for Mesoscopic Physics, Physics Department
Peking University, Beijing, China 100871
[***] *Electrical Engineering Department, California State University at Long Beach, Long Beach, CA 90840, USA*

ABSTRACT

To test 2PC structures by experiment, we affixed 2PC structures onto a hemicylinder of polydimethylsiloxane (PDMS) and placed the sample on a rotating stage with a green laser beam (λ=532nm) incident on the grating structure. Besides the experimental tests, theoretical models are developed to predict experimental results. The source for the finite difference time domain (FDTD) simulation was a plane wave with a Gaussian power distribution. Comparing the simulation results with the experimental ones, we found that FDTD underestimates the experimental data and does not account well for the upward inflection of transmission efficiency about the critical angle. To increase the accuracy, a more finely girded simulation at the diffraction grating is developed, and this result agrees very well with the experimental test.

1. Introduction

The focus of this particular research project is on developing higher brightness, higher efficiency and longer lifetime of GaN-based light emitting diodes (LED), an important research topic in solid-state lighting, the so called next generation lighting.

Problems to be addressed: The current advancements allow us to achieve applications with LEDs in fiber optics, displays, and lighting systems. We can even control the color contrast of the device and create a full color set with red, green, and blue LEDs [1]. Although LEDs have been used in many applications, the light extraction efficiency is still very low for Gallium-Nitride (GaN) LEDs due to several factors: GaN has a low critical angle that traps light inside the device [2], absorption of light within the device due to dislocations and defects within the GaN crystal [3], and a device design and structure that has not been optimized (ie. epitaxial side up vs. epitaxial side down chip structures) [4]. It is critical to develop LED technology to reduce energy consumption through higher efficiency of LEDs and expedite resulting replacement of older, less efficient technologies.

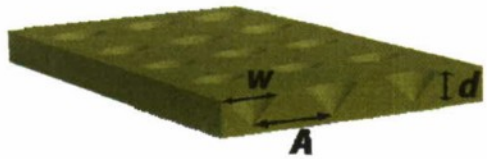
The greatest limitation in GaN-based LEDs is the effect of total internal reflection resulting in poor light extraction efficiency (LEE). This applies to any large change in refractive index on the boundary between two materials. When light is above the critical angle, the light cannot escape and is eventually absorbed in the device, decreasing LEE and generating heat. Many methods of improving LED efficiency are currently being explored. Almost all of these methods are seeking to extract the trapped light in greater quantity and faster speed. Those methods being explored are placing photonic crystals or a nanostructure grating on one of LED layers to modify the effective index of refraction at the boundary [6-7], randomized roughening on the surface of the device [7-8], slanted device configurations that result in pyramidal shapes [7], and inverted "flip-chip" designs that put the epitaxial side upwards or downwards [4] [9].

2. Simulation Model

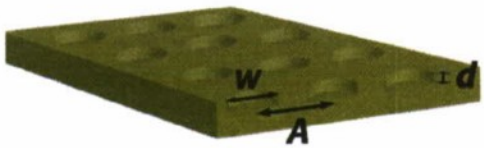
The research goals: (1) simulate 3-D structures that do not approximate LED chip surfaces to two-dimensional structures, as in two-fold photonic crystal (2PC) structures, (2) accurately simulate such a 3-D grating structure, (3) produce simplifications to such a model to increase stability and reduce simulation time, and (4) collaborate and synergistically work towards developing experiments and fabrication processes to compare with 3-D simulations.

Because of the stated issues of light extraction due to GaN's low critical angle, a key component of our research was to solve the light trapping issue by a common solution: etch a periodic structure on one or more of the GaN layers to allow transmitted diffraction to allow light to escape for angles past the critical angle [12]. Various grating structures exist,

such as hemispheres, cylinders, pyramids, cones, and other arbitrary shapes. All structures can be in hole or pillar formations. The most common structures are conical and cylindrical, due to the ease of fabrication. Fig. 1 shows the possible grating structures for the conical and cylindrical shapes. Study of these grating structures involved two parts: the study of transmitted diffraction from 2PC structures and also the study of top transmission gratings on LED devices, both of which are reported on in the following sections.



(a)



(b)

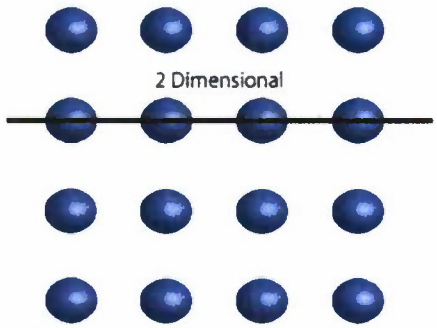
(c)

(d)

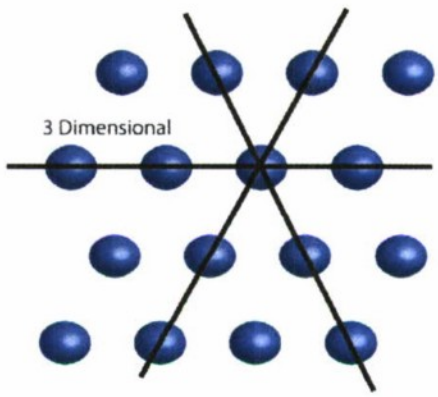
Fig. 1 - 2PC Grating Structures: (a) conical holes, (b) cylindrical holes, (c) conical pillars, and (d) cylindrical pillars

3. Experimental Setup

2PC Transmitted Diffraction: To test these structures by experiment, we affixed 2PC structures onto a hemicylinder of polydimethylsiloxane (PDMS) and placed the sample on a rotating stage with a green laser beam (λ=532nm) incident on the grating structure. The 2PC hole arrays had a lattice pitch of 2μm and were prepared on a silicon wafer with an area of 1mm×1mm. The 2PC pattern was then transferred onto the planar center point of the PDMS hemicylinder by soft nanoimprinting. [10,11] A photodetector was placed behind the grating structure to measure transmission efficiency. Since we were interested only in single-pass transmission characteristics, the hemicylinder shape allows reflected and diffracted light to escape with negligible reflection back towards the photodetector. Fig. 3 shows the experimental model and grating structure.



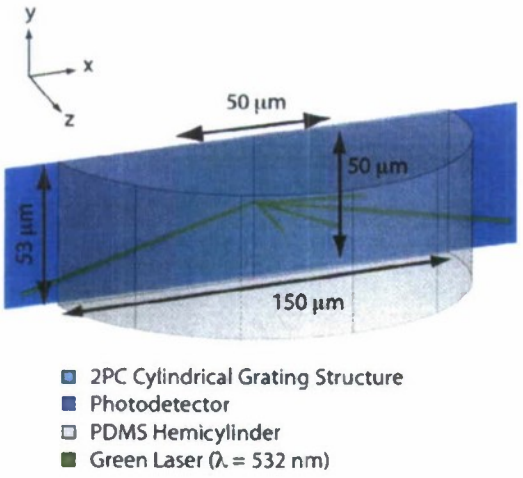
(a)



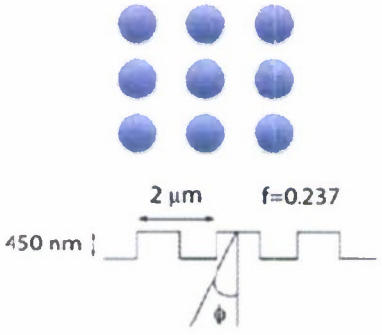
(b)

Fig. 2 – Grating structures of various dimensionality: (a) 2-D approximation of a 2PC grating structure and (b) a full 3-D model of a 3PC grating structure

(d)



(a)

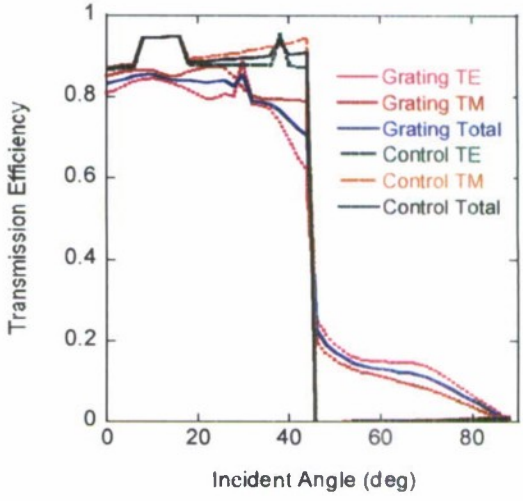


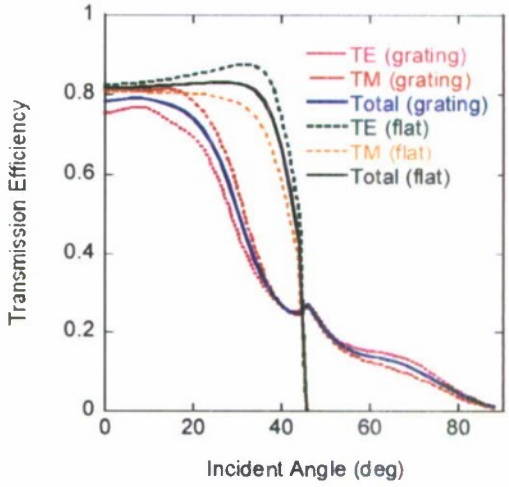
(b)

Fig. 3 – Experimental setup for measuring transmitted diffraction of a 2PC structure mounted on PDMS

4. Results

Also, we used theoretical models to predict experimental results. Since the symmetry is 2-fold, this allows us to use 2-D models to represent an otherwise 3-D model. 2-D models bear heavy costs in simulation time and required memory in FDTD. These requirements are somewhat more relaxed in Rigorous Coupled Wave Analysis (RCWA). [10,11] Fig. 4(a) shows a 2PC structure that can be simulated as a 2-D model, while the three-fold symmetric photonic crystal structure Fig. 4(b) cannot be approximated by 2-D models, due to the honeycomb pattern whose diffraction effects cannot be approximated, as is the case with 2-D simulation models. The results of the simulations in FDTD and RCWA are shown in Fig. 4a and Fig. 4b, respectively. Simulations by FDTD using a Yee's mesh were coarsely gridded at a Δx of 200nm and a Δz of 45nm. The source for the FDTD simulation was a plane wave with a Gaussian power distribution. The simulation domain included 100 grating cells at a lattice pitch of 2µm, filling factor of 0.237, and grating cell height of 450nm. RCWA simulations assumed an infinite plane of grating cells with a rectangular plane wave source are also performed for comparison.





(a)

Fig. 4 (a) Transmitted Diffraction Efficiency simulatin by FDTD simulation and (b)Transmitted Diffraction Efficiency for a non-uniform gridded FDTD simulation.

The results of the simulations show that RCWA produces a more accurate model of the experimental data. FDTD underestimates the experimental data more than RCWA and does not account well for the

upward inflection of transmission efficieney about the critieal angle. To increase accuraey, a more finely gridded simulation is neeessary. When using FDTD, a method to gain higher resolution at critical points, sueh as a diffraction grating, while still keeping a eoarse grid for the rest of the simulation domain is to use a non-uniform grid. This allows simulation time and memory to be kept low while not saerifieing aeeuraey at eritieal points in the diffraction grating as shown in Fig. 4. The simulation only uses a non-uniform grid in the z-direetion. The smoothness of the simulation data for both grating and flat eases is aeeounted for by the fine grid at the grating (Δz = 22.5nm), while the simulation time was kept low by making the rest of the simulation area coarsely gridded (Δz = 56.25nm). The infleetion is now elear at an incident angle of 46 degrees but is still not as accurate as the RCWA simulation model. This is a definite disadvantage of FDTD, in that transmission properties are not easily measured, as materiał dispersion, near field effeets, coarse/fine gridding ean affect the accuraey of an FDTD model. For regular struetures, methods sueh as RCWA are more effieient and more aeeurate for ealeulating transmitted diffraetion.

Reference:

[1] Yeh, D.-M., Huang, C.F., Chen, H.-S., Tang, T.-Y., Lu, C.-F., Lu, Y.C., Huang, J.J., Yang, C. C., Liu, I-S. and Su, W.-F., "Control of the eolor eontrast of a polyehromatie light-emitting device with CdSe-ZnS nano-crystals on an lnGaN-GaN quantum-well strueture," IEEE Photon. Teeh. Lett., 18(5), 712-714, 2006.

[2] Park, J., Oh, J.-K., Kwon, K.-W., Kim, Y., Jo, S., Lee, J. K. and Ryu, S.-W., "lmproved Light Output of Photonie Crystal Light-Emitting Diode Fabricated by Anodized Aluminum Oxide Nano-Patterns," IEEE Photon.Teeh. Lett., 20(4), 321-323, 2008.

[3] Kawaguehi, Y., Nishizono, K., Lee, J. and Katsuda, H. "Light Extraetion Simulation of Surface-Textured Light-Emitting Diodes by Finite-Difference Time-Domain Method and Ray-Traeing Method," Jpn. J. Appl. Phys., 46(1), 31-34, 2007.

[4] Lee, S., "Study of photon extraetion effieieney in InGaN light-emitting diodes depending on chip struetures and ehip-mount schemes," SPIE Optieal Engineering 45(I), 014601, 2006.

[5] Hatakoshi, G., Hattori, Y., Saito, S., Shida, N. and Ninoue, S., "Device Simulator for Designing High-Effieieney Light-Emitting Diodes," Jpn. J. Appl. Phys., Vol. 46(8B), 5419–5425, 2007.

[6] Xu, Z., Cao, L., Tan, Q., He, Q. and Jin, G., "Enhaneement of the light output of light-emitting diode with double photonie erystals," Opties Communieations, 278 (I), 211–214, 2007.

[7] T. V. Cuong, H. S. Cheong, and C.-H. Hong, "Caleulation of the external quantum effieieney of light emitting diodes with different ehip designs," Phys. Stat. Sol. (c), I(10), 2433-2437, 2004.

[8] Huang, H.-W., Kao, C. C., Chu, J. T., Kuo, H. C., Wang, S. C. and Yu, C. C., "Improvement of InGaN-GaN light-emitting diode performanee with a nano-roughened p-GaN surfaee," IEEE Photon.Teeh. Lett, 17(5), 983-985, 2005.

[9] Bao, K., Kang, X., Zhang, B., Dai, T., Xiong, C., Ji, H., Zhang, G. and Chen, Y., "Improvement of Light Extraetion From Patterned Polymer Encapsulated GaN-Based Flip-Chip Light-Emitting Diodes by lmprinting," IEEE Photon.Tech. Lett,, 19(22), 2007.

[10] K. Bao, X. N. Kang, B. Zhang et al: IEEE Phot. Teeh. Lett. Vol. 19, p. 1840, 2007.

[11] K. Bao, B. Zhang, X. N. Kang et al: Proe. of SPIE. Vol. 6910 , p. 69100N-1, 2008.

Mode Pattern Analysis of Gallium Nitride-based Laser Diodes

Xiaomin Jin*[a], Sean Jobe[a], Simeon Trieu[a], Benafsh Husain[a], Jason Flickinger[a], Bei Zhang[b], Tao Dai[b], Xiang-Ning Kang[b], and Guo-Yi Zhang[b]
[a]Electrical Engineering Department, 1 Grand Avenue, California Polytechnic State University, San Luis Obispo, CA, USA, 93407-9000;
[b]School of Physics and State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, Peking University, Beijing, China, 100871

ABSTRACT

In this paper, we present an analysis of gallium nitride (GaN) quantum-well (QW) laser diode (LD) by numerical simulation. Here we discuss three aspects that are crucial to our analysis. First, the transverse mode pattern is studied, and our current GaN diode laser structure is discussed with optical waveguide mode analysis. Then we compare the QW design of the laser and maximize laser modal gain. Finally, we report the influence of the electron block (e-block) layer on lasing performance of our design.

Keywords: Gallium Nitride, semiconductor laser, transverse modes

1. INTRODUCTION

In the visible color spectrum, there are three primary colors: red, green, and blue, also referred to as RGB. Through the combination of these three colors, all the other colors in the spectrum can be created. For example, the combination of red and green would produce the color yellow. Current technology has found efficient ways at creating both red and green light using semiconductor technology, but the creation of blue light has not shared this immediate success. Blue light has remained the hardest light to produce efficiently from semiconductor technologies. This has been the main road block preventing highly efficient laser diodes (LDs) from replacing our current lighting technologies.

Blue violet light is close to the shortest wavelength of light in the visible spectrum. At the wavelength of 445nm, it still remains the hardest light to produce using semiconductors, because the bandgap properties of the materials needed to produce such a wavelength in semiconductors that are hard to find. In recent years, many strides have been made in studying and researching Gallium Nitrate or Gallium Nitride (GaN) as a material to yield blue light. The most progress and significant strides have been made by Dr. Nakamura, who dedicated much of his research to the chemical growth of GaN compounds. [1] Before Nakamura found a solution for growing GaN semiconductors, much of the focus for blue light semiconductors was spent on II-VI materials (where II and VI represent the group number on the Periodic Table of Elements). GaN is a III-V material and is a much harder substance to deal with because of its higher lattice defects in comparison to II-VI materials. However with Nakamura's perilous efforts, he eventually found a technique to deposit GaN compounds and created the first GaN semiconductor. Since Nakamura has shown that GaN LED semiconductors are possible in 1995, fellow scientists across the world have worked hard to further the development and efficiency of blue light semiconductors [1].

GaN is one of the most promising materials for use in the blue and ultraviolet wavelength region. Since the room-temperature (RT) continuous-wave (CW) operation of GaN-based lasers were reported by Nakamura et al. 1996, [2] one of the important targets on the GaN LD development is to extend the operating lifetime by reducing the operation current or current density. It was pointed out that the threshold current density of GaN lasers is intrinsically higher than that of GaAs. This is because the hole effective mass of wurtzite (WZ) GaN is much heavier than that of conventional zincblende materials such as GaAs.[3] To decrease the threshold current density in GaN lasers, several approaches have been proposed. Most of the papers discuss the threshold improvement through optical gain or Quantum-well (QW) analysis. [4]-[6] The beginnings of III-V laser diode devices were very elementary in their design compared to current

*xjin@calpoly.edu; phone 1 805-756-7046; fax 1 805-756-1458; www.ee.calpoly.edu

International Symposium on Photoelectronic Detection and Imaging 2009: Laser Sensing and Imaging, edited by Farzin Amzajerdian, Chun-qing Gao, Tian-yu Xie, Proc. of SPIE Vol. 7382, 73820O
 · doi: 10.1117/12.834103

designs. The GaN diodes consisted of a substrate, cladding, core, and active region. Current designs have more complex structures such as a super lattice (SL), quantum wells, and e-blocks. Using these new technologies, devices with lower laser thresholds can be created thus leading to more efficient devices. In this work, we discuss the threshold reduction on GaN QW lasers in three aspects: 1) optical waveguide design and mode analysis; 2) QW design and modal gain analysis; 3) electron- block (e-black) layer design for optimization.

2. NUMERICAL MODELING

LaserMOD is an integrated software package developed by RSOFT for the design and simulation of semiconductor lasers and active photonic structures. [7] It is a fully integrated platform with a user friendly parametric CAD interface, nonuniform Delauncy mesh generator, material libraries, gain and mode calculation utilities, simulation engine, standard and custom plot generation utilities, and versatile graphical viewing utilities. In short, the LaserMOD program provides the user with an immense environment in which semiconductor devices can be modeled and tested based on their user specified properties.

2.1 Optical Confinement Factor (OCF)

The optical confinement factor (OCF) of a chosen mode is defined as the ratio of the energy of the chosen mode located in the active region to the total guided energy of all the modes. [8] A higher OCF for a certain mode indicates there is more energy in the active region for that mode. High energy in the active region is usually a sign that lasing is occurring. Thus, whichever mode has the highest OCF will be the lasing mode. The optical modes are determined from the solution of the Helmholtz equation via simultaneous as below, then the OCF can be calculated.

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\delta^2}{\delta y^2}+k_0(\varepsilon(x,y)-n_{eff,m}^2)\right)E_m(x,y,z)=0$$

where $E_m(x,y,z)=E_m(x,y)\exp(ik_0 n_{eff,m}z)$, k_0 is the free-space wave vector, and $\varepsilon(x,y)$ is the complex dielectric constant profile of the multiplayer structure. The eigenvalues are given by the effective index $n_{eff,m}$. The frequency $k_0=\omega_0/c$ of the mode is solved and is set to correspond to the quantum well band gap energy. The OCF can be thought of as the fraction of the energy of mode that is located at the active region. [9]

2.2 Mathematical Model – 8x8 Kronig-Penney (KP) Model

The LaserMOD simulation software uses the Kronig-Penney (K • P) model to model the bandgap relationships of the different materials. The K • P model is based upon the splitting of allowed electron energies as the interatomic distance between atoms decreases to form a crystal. The K • P method involves quantum mechanics and a solution to Schrödinger's wave equation. [10] For the one dimensional crystal structure, a periodic well function can be used to represent the crystal lattice structure. This is because when the potential functions of atoms are brought close together, the net potential function of the overlapping regions is similar to a periodic function. Using the specified characteristics of a periodic function for the boundary conditions for Schrödinger's equation, a plot of the energy E as a function of wave number k can be generated, which describes the valence band, conduction band, and the allowed energy bands. However for two dimensional calculations, matrices become a necessary addition to efficiently solve Schrödinger's wave equation. A popular method for solving E vs. k diagrams is using 8x8 matrices. Using special matrix rules and "tricks" (i.e. Helmholtz equation) these calculations can be solved with a computer program. Many programs have been written that can solve these types of mathematical problems. We chose to use RSOFT's LaserMOD, which is a program that has a CAD interface which can easily and quickly calculate all the necessary properties of semiconductors.

2.3 The 2D nitride-based laser model

A 2D nitride-based laser model is developed, which is shown in Fig.1. The simulated laser structure has five quantum-wells (QWs). A similar device was fabricated at Peking University, China. One of the project goals is to identify design flaws from the current laser design and to reduce the laser threshold. To avoid the meshing difficulties of the finite-element method, the classical Ritz simultaneous iteration is combined with an additional optimization to analyze closed

arbitrary dieleetrie waveguides. [7] LaserMOD determines the charge distribution using Schrodinger equation. The laser simulation is based on 8x8 **K • P** band strueture ealeulation and photon rate equation. The material parameters based on reeent literature values and some experimental data were used. The detailed laser strueture is listed in Table.1.

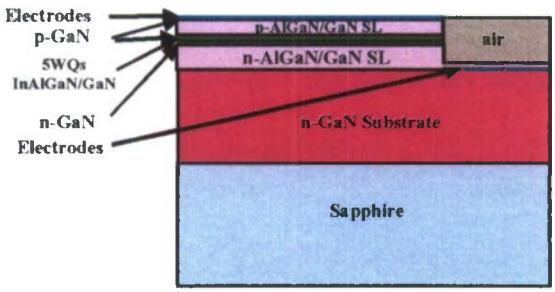


Fig.1 2D GaN laser model using Rsoft LaserMOD.

Table. 1. Laser diode layer strueture and parameters.

Layer	Thickness (nm)	Refractive Index (n)
p-GaN (contact)	50	2.55
p-$Al_{0.12}Ga_{0.88}N$/GaN (p-SL eladding)	500	2.53
p-GaN (p-eore)	100	2.55
p-$Al_{0.35}GA_{0.65}N$ (e-block)	20	2.42
n-GaN	15	2.55
$In_{0.1}GA_{0.9}N$/GaN (5QWs)	67	2.685/2.55
n-GaN (n-core)	100	2.55
n-$Al_{0.12}Ga_{0.88}N$/GaN (n-SL cladding)	800	2.53
n-GaN (Substrate or buffer)	4000	2.55
Sapphire (Oxide)	4000	1.77

3. SIMULATION RESULTS

3.1 Lasing mode and ghost modes

The first twenty transverse modes in the above GaN LD structure are calculated. Fig. 2 shows several optical mode patterns of GaN LD. When sapphire is used as a substrate for GaN lasers, the dislocation density in the material is usually very high. To eontrol the defeet number or reduee eraeks, an n-GaN substrate (or buffer) layer (several micron thiek) is deposited on the sapphire substrate before growing the eladding layer. This layer has higher refractive index eompared to the n-AlGaN/GaN supper-lattice (SL) cladding layer. Because of insufficient eladding thiekness, QW waveguide and n-GaN substrate waveguide are eoupled and the GaN lasers have multi-waveguide structures, which support strong substrate modes, also ealled the "ghost-mode" phenomena. [11] Therefore, the fundamental mode of the multi-layer waveguide is usually "ghost mode". The higher order mode of this multi-layer waveguide is usually the lasing mode. The optieal eonfinement factor is also very low even for the lasing mode, about several percents. This leads to lasing problem of GaN lasers. Different order modes have different optieal eonfinement faetor. The most strongest-eonfined mode in the multilayer waveguide structure can be the lasing mode.

The optical modes of the 2D simulation are shown in Fig. 2. For modes 0^{th} through 7^{th} the modes look very similar to the 1D results, [12] with additional modal energy being layered above and below adjacent energies. The 8^{th} mode is the lasing mode and shows the optical energy confined inside the active region. Modes 9^{th} and 10^{th} modes continue the same trend of optical energy layering as modes 1 through 7. However, something very interesting happens in mode 11^{th}. For Mode 11^{th} to 20^{th}, the modal energy splits across the x-axis horizontally. The next few modes follow this new horizontal trend. The energy split at x-direction is a characteristic that we would not be able to notice without doing a 2D model and gives insight into the optical energy interaction along the x-axis. We find that as the mode increases, the optical energies are no longer contained along the vertical boundaries, the energies split and are shared across the x-axis.

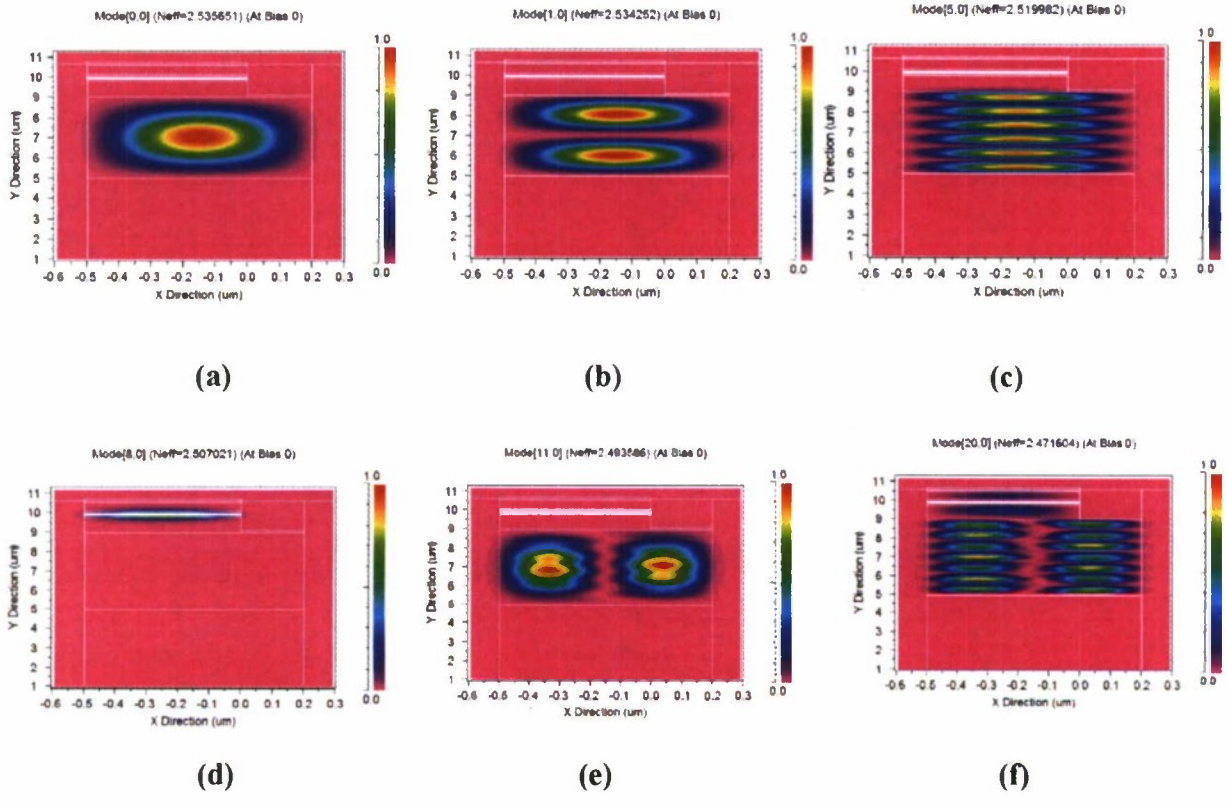


Fig. 2 Several ghost modes and lasing mode distributions: (a) 0^{th} order mode, (b) 1^{st} order mode, (c) 5^{th} order mode (d) 8^{th} order mode or the lasing mode, (e) 11^{th} order mode, and (f) 20^{th} order mode.

From our calculation, the optical confinement factor for the zero-order mode is only 0.0000067%. The 8^{th} mode has the greatest overlap of optical field with the quantum well. Its optical confinement factor is 8.67%, which agrees with reported data. [13] This indicates that our LDs lasing in the 8^{th} order transverse mode. The other modes are substrate modes or ghost modes as shown in Fig. 2. Strong substrate modes compete with the lasing modes in this multi-waveguide structure. Optimizing the OCF for different layer thicknesses is very important in lowering the lasing threshold of the GaN LD. An increase in the OCF means that more light is being confined in the lasing mode. With more light focused in the lasing mode, less current will be required for the LD to achieve lasing. In here, we demonstrate how the different optical modes affect the Light vs. Current curves of the GaN LD. By creating a new laser design using our optimized thicknesses, we show that it produces lower lasing thresholds, which brings us one step closer to the ultimate goal in meeting the expectations of using laser diodes as future light sources.

The OCF is calculated in the 2D design for GaN substrate layer thicknesses from 0-5μm. Fig. 3 shows the values of the OCFs simulated and for what mode they occurred. The lasing mode migrates as the GaN thickness increases. For every increase in the GaN substrate by 0.5μm the lasing mode jumps up a mode. For example, at a GaN thickness of 2.6μm the lasing mode is 5th mode, but increasing the GaN thickness to 3.1μm, the lasing mode migrates to the 6th mode. Each mode is evenly spaced by the constant of 0.5μm. These results are in agreement with our previous 1D GaN substrate simulation results. [12] Using the original design from Table 1, the simulated OCF for the lasing mode (8th mode) was 8.6677%, however in Fig. 3, the OCF can still be increased if the GaN thickness is increased a little. Following along the line that represents the 8th mode, a thickness of 4.2μm yields an OCF of 8.6962% which is a 0.02% increase over the original design. Only a small increase in the OCF was possible for the GaN substrate layer showing that the original layer thickness was very close to being an optimal design.

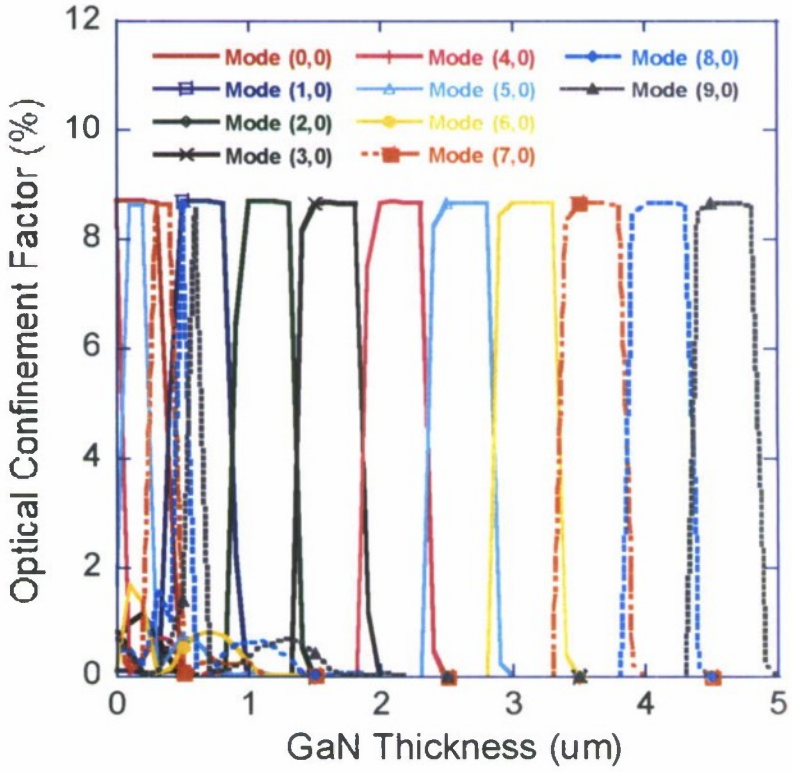


Fig. 3. The OCF vs. GaN substrate thickness (2D)

It is expected that increasing the cladding layer thickness will lengthen the distance that the light has to tunnel through in order to get into the substrate waveguide. This will prevent optical leakage of the mode into the absorbing, high index p-contact layer. This however will increase the impedance as it would also lengthen the distance the current would have to tunnel through and therefore increase our threshold current. By adjusting the thickness in the n-SL, it is of interest to find a smaller thickness that still retains a high OCF and a thickness that prevents mode and current leakage. It is also important to be aware that creating such thin coats of the n-SL still remains a real world problem and although a thin coat of n-SL may lead to an optimized simulation, creating such a lasing device may be difficult. By increasing the cladding size, we are looking to find how much a thickened cladding actually affects the OCF. We expect a thickening of the n-SL will prevent mode leakage from reaching the substrate and thus help prevent ghost modes therefore leading to a higher OCF.

The results of the 2D simulation with varying n-SL thickness (in Fig. 4 (a)) are similar to those Hatakoshi calculated. [12][14] As the cladding thickness is increased, the 8th mode (lasing mode) gain begins to confine more light. As the

lasing mode increases in its OCF, the other modes begin to lose the light they had originally been able to confine. Thus as the cladding thickness increases all the optical energy that is available becomes confined only in the 8th mode. This is shown in Figure 4 (a). This result shows that the anti-guide-like or ghost mode behavior can be suppressed by increasing the cladding layer thickness. A thick cladding layer reduces the effect of the outer contact layers, but still is conducive to high-order modes. At the n-SL thickness of the original design (0.8μm), the OCF was 8.7128% and reaches a peak plateau on the graph at 8.7241% at a thickness of 1.4μm. The 0.01% increase in OCF requires almost doubling the thickness. Thus the best optimization for this design is simply leaving it alone. The original design yields an optimal OCF for its relative thickness. The p-SL cladding provides the same waveguide like confinement of the light as the n-SL cladding, shown in Fig. 4 (b), but it differs in that it will not be absorbing electrons like the p-SL will be. By adjusting the thickness of the p-SL, we hope to find a thickness that prevents mode leakage into the p-contact layer, yet still maintains low impedance. The adjustment of the p-SL cladding layer shows a different increase in the OCF than the n-SL plot. It is interesting to point out however that the thickening of the p-SL lattice above 0.1μm has a detrimental effect on the maximum OCF, decreasing it from a maximum of 8.7% to 8.4%. (It is important to note that it is the OCF at a different mode). Thus we find our optimal thickness for the p-SL that yields the best OCF yet remains fairly thin is at 0.1nm.

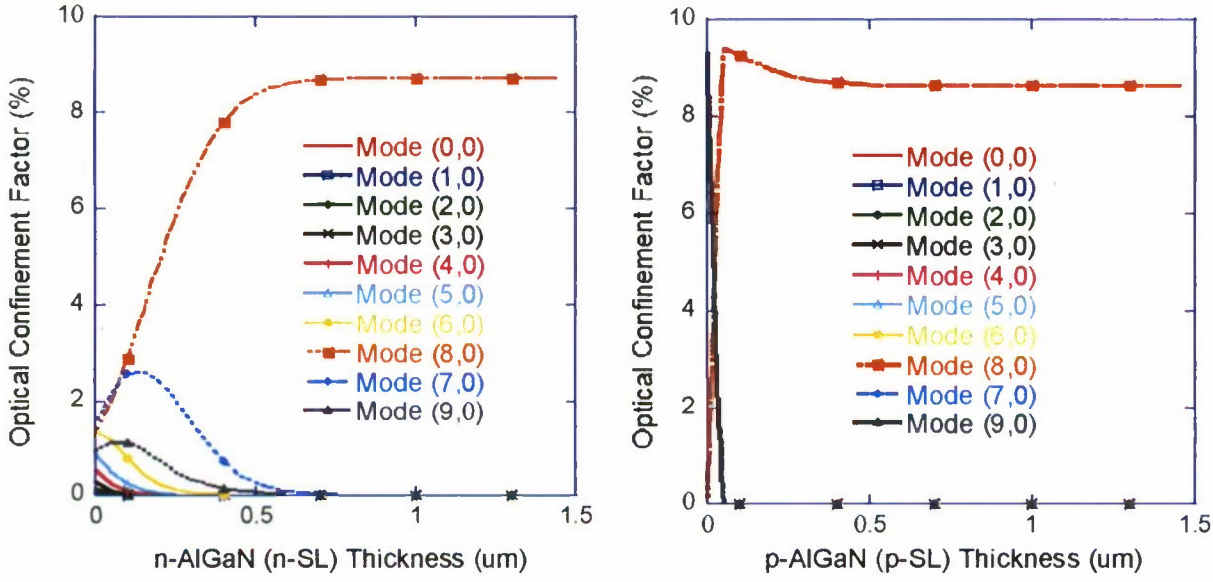
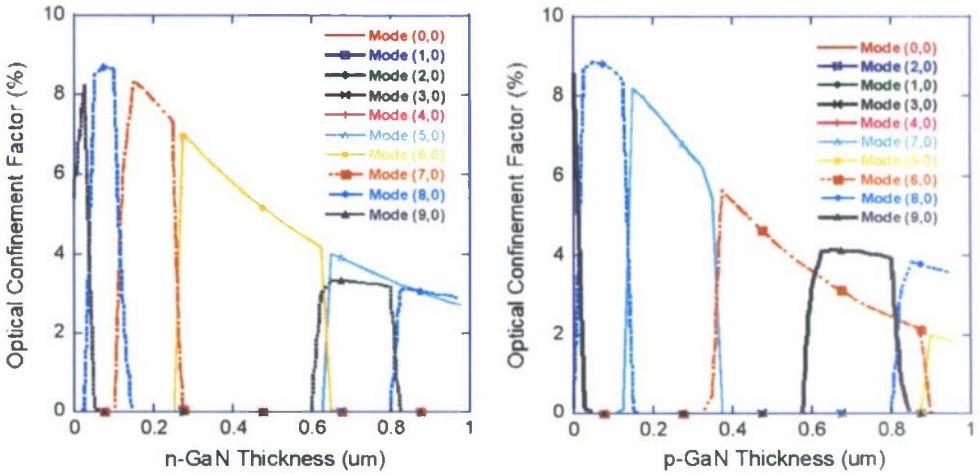


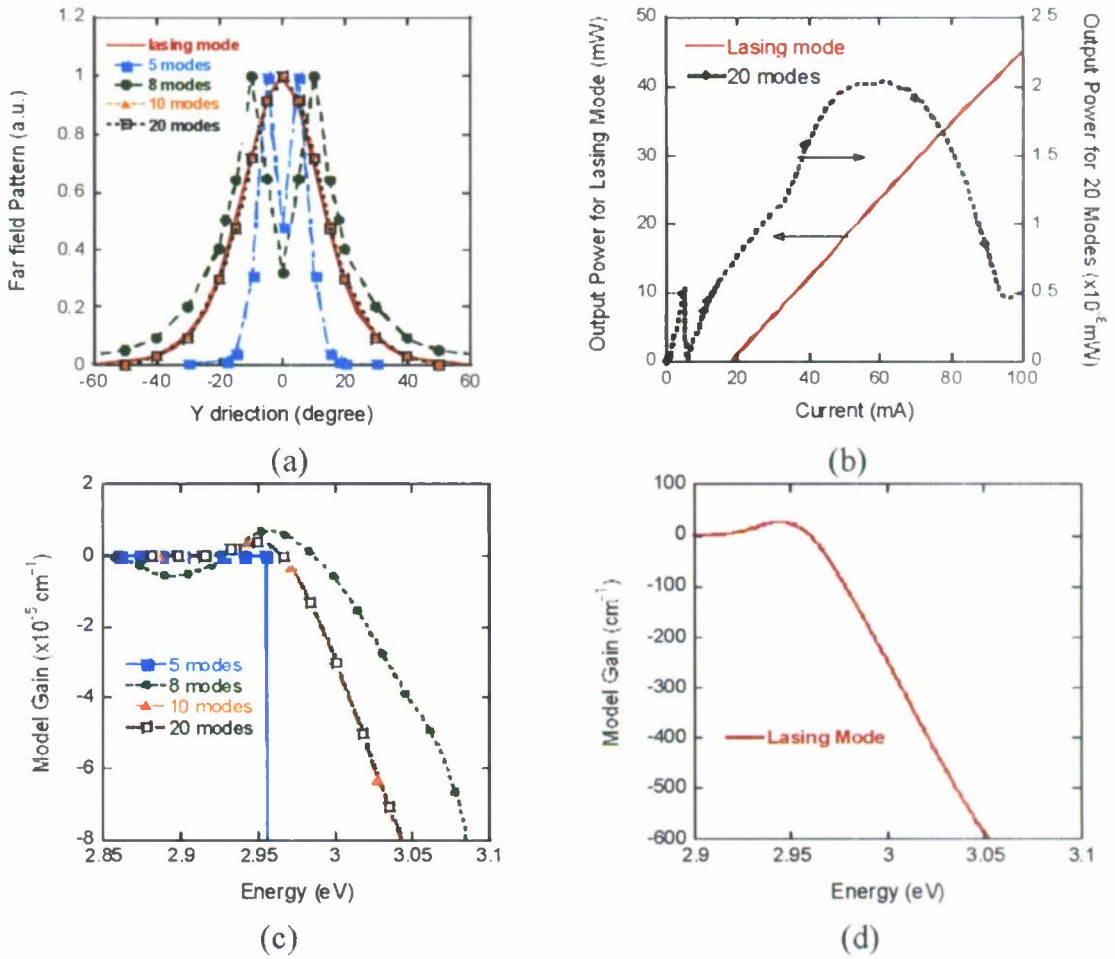
Fig. 4. Optical confinement vs. (a) n-SL and (b) p-SL thickness

The GaN core layer in the 2D simulation has a refractive index greater than the surrounding cladding layers to reflect the light back towards the active region. A thicker core layer can lead to more optical losses due to the large area the light has to travel. Thus choosing an optimal core layer thickness is important in creating the most efficient laser diode. The OCF is plotted in Fig. 5 for n-GaN layer thicknesses of 0-1μm. The OCF peaks at 8.7027% with an n-GaN thickness of 0.075μm. This is a 0.224% increase in the OCF from 8.6762% when a 0.1μm layer is used. Thus by using a thinner layer of n-GaN core layer, a better optical confinement can be achieved. Fig. 5 shows a constant decrease in the OCF as the n-GaN thickness is increased and the lasing mode migrates to lower modes. The same simulation was run for the p-GaN core layer and the results of the OCF vs. p-GaN thickness are also plotted in Fig.5. The maximum OCF is achieved for small thin layers of p-GaN. The OCF peaks with 8.8664% at a thickness of 0.05μm. This is a 0.2% increase over the original design when the p-GaN was 0.1μm thick and had an OCF of 8.6762%. Similar to the n-GaN core layer, a thinner layer creates better optical confinement.



(a) (b)

Fig. 5. Optical confinement factor vs. (a) n-GaN and (b) p-GaN layer thickness



(a) (b) (c) (d)

Fig. 6 The GaN laser simulation for (a) far field pattern, (b) light output versus current, (c) model gain of all the transverse mode, and (d) model gain of lasing mode only.

To prove the optical mode design can improve CW GaN LD operation, the threshold, optical power, quality of far-filed, and optical gain should be evaluated, which are the most important issues and represents the laser performance. For examples, in 2002, Tojyo et al reported a kink free output of over 100mW. In 2005 Schwarz et al reported near- and far-field study of GaN LDs. [15] In 2007, Laino et al reported results on substrate mode study. And in 2006, [16] Witzigmann et al reported optical gain analysis. [17] Here we calculate the far field pattern, model gain, optical output, and threshold with only lasing mode (the 8^{th} mode), and all the transverse modes to demonstrate the ghost mode effects, shown in Fig. 6. For the far field simulation, we normalized the optical pattern to the peak value for comparison. In the figures, lasing mode represents the 8^{th} order mode only simulation, "x modes" means "the first x transverse modes". The "8 modes" means modes 0 to 7. For the case of "5 modes" or "8 modes", there is no lasing mode included in the calculation and LD does not reach threshold for lasing. For the "8 modes", "10 modes", and "20 mode" cases, little modal gain is obtained from laser, which is about 10^{-5} cm^{-1}. Therefore the light output is only about 10^{-8} mW. If we calculate those results with only 8^{th} order lasing mode, the laser can reach a peak gain of about $25cm^{-1}$, lasing output is about 40mW at 80mA, and threshold is about 19mA. According to our optical field analyses, the strong ghost-modes compete with lasing-mode in GaN Laser, which can prevent the laser from lasing. For the worst case, if the waveguide structure supports strong ghost-modes, the GaN laser would not operate even though the active region quantum well is perfect. Design of the optical waveguide for GaN laser system is very important to achieve efficient lasing condition. In this work, we find that the n-GaN substrate thickness is the major factor influencing the transverse mode pattern. In a related work, [12] we optimized the optical waveguide structure, which limits the ghost modes, and maximizes confinement factor of the laser structure, in order to reduce lasing threshold.

3.2 The QW analysis and model gain simulation

Quantum wells are designed to trap electrons in a 2-D environment. The wells are designed to have a particular bandgap energy related to the wavelength of light emitted by the laser diode. The quantum well allows the electrons to gather more densely in the well than they would elsewhere. It is easy to think of it as though the well is drawing electrons towards it then trapping them. (In fact, it isn't so much that the wells are drawing electrons to them and away from the other parts of the semiconductor; instead electrons keep falling into the well, and the well becomes saturated and full that it seems as though the well is attracting electrons.) This packs more electrons in the active region and allows more electrons to jump the bandgap thereby releasing a photon. The size of the well must be on order of the wavelength of light hoping to be produced. Each well is created by creating a thin layer (the well) and surrounding it by thicker layers of a different material. Thus for a multi quantum well (MQW) design, there would be alternating layers of bulk and active layers. The well layer is made from the normal active layer material with a lower refractive index then the surrounding bulk layers. This in a sense is like having many tiny optical cavities that interact with photons at the quantum level. It can be shown that the more quantum wells in a semiconductor, the larger the threshold current becomes. This is because when there are more wells for the electrons to fall into, a higher current is necessary to provide the quantum wells with enough electrons to maintain saturation which in turn is the foundation of lasing. [18]

Quantum wells have many advantages in the design a laser diode structure. First, by using QWs, one has the freedom to design the transition energies, which ultimately determines the wavelength of light that leaves the active region in the form of spontaneous emission. Second, since QWs have a small volume, the free carrier concentration in the QW is high and at high free carrier concentrations non-radiative deep-level transitions are less likely thus yielding a high radiative efficiency. Third, due to the small size of the quantum wells, the carrier density required to achieve population inversion (i.e., high carrier density in the conduction band, compared with the valence band) is small and therefore the threshold current density of QW structures is low. And finally, the surface recombination is less likely, which making surface recombination less important in the study of QWs. [18]

The structure we simulated here has a wide contact with a uniform current injection. The number of QWs, the QW width, and composition fluctuations play central roles in the optimization of GaN light-emitting diodes (LEDs) and LDs. [4]-[6] [17] Our simulation here considers number of QWs and composition. We choose lasing mode only simulation, without considering optical mode variation for simplification. The band structure of QWs is computed using $\mathbf{k \cdot p}$ method, which includes coupling effects for the heavy-hole, light-hole, and the crystal-field split hole dispersion. The QW is $In_xGa_{1-x}N/GaN$. The width is 5nm/7nm (well/barrier), and x=0.1. From Fig. 7(a), the peak modal gain increases from 17 cm^{-1} for 1QW to 3.7 cm^{-1} for 3QWs. Then it reduces to $29.1cm^{-1}$ and $25.3cm^{-1}$ for 4 QWs and 5 QWs. The 3 QWs Case provides the highest gain peak. This agrees with the light optical power-current (LI) curve simulation in Fig. 7(b). The carrier densities are inhomogeneous among the quantum wells. [19] The optical gain is generated only in three

QWs on the p-side for our case. The QWs on the n-side act as absorption layers. Therefore, for the LDs with more than three QWs, the peak gain reduces, which results the light output slope (or quantum efficiency) reduction in LI plots. From the threshold and gain simulation, it is important to decrease the number of wells. Single quantum well GaN LD is the best design structure, however it has a lot of fabrication challenges. Besides the above gain simulation, we have varied structural parameters and calculate the threshold currents of lasers. The threshold current increases as QW number increases and decreases with $In_xGa_{1-x}N$/GaN composition x, as shown in Fig. 8. Higher "In" composition in the QW will obtain higher gain, however it has to be balanced with drawback of the lattice mismatch and deterioration in material quality for a larger "In" fraction. For 5 QW case, the threshold current shows little composition dependence, while x varies from 0.08 to 012. This will provide good threshold stable point with fabrication tolerance.

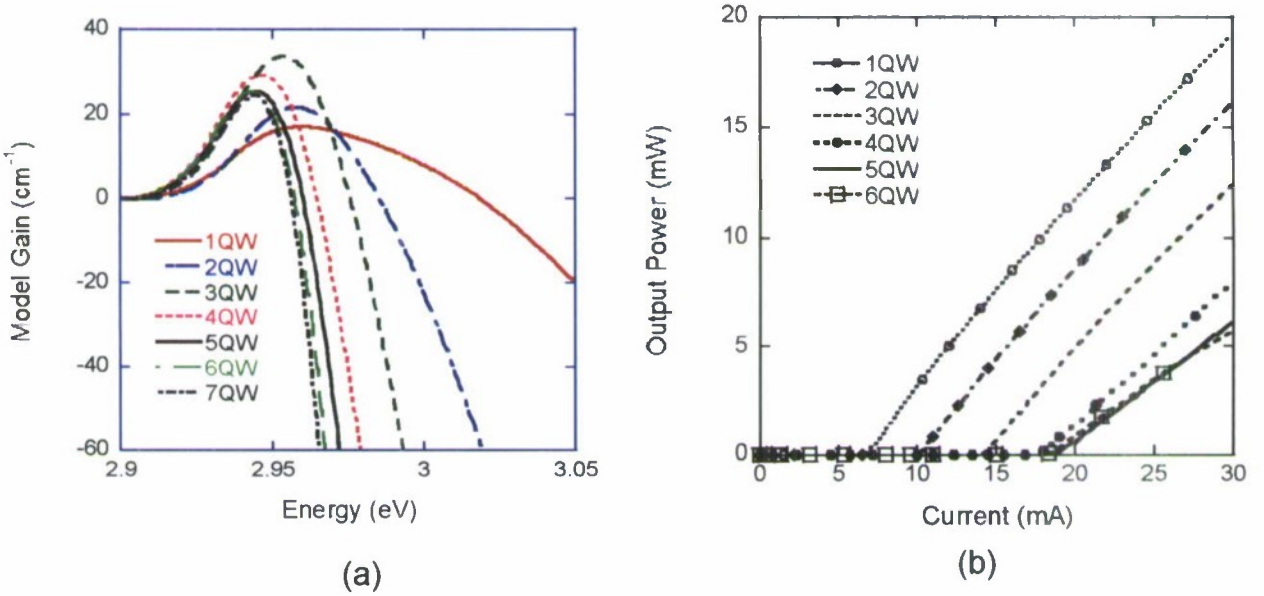


Fig.7 (a) Model gain and (b) LI curve simulation for $In_{0.1}Ga_{0.9}N$/GaN single-QW and MQWs without e-block.

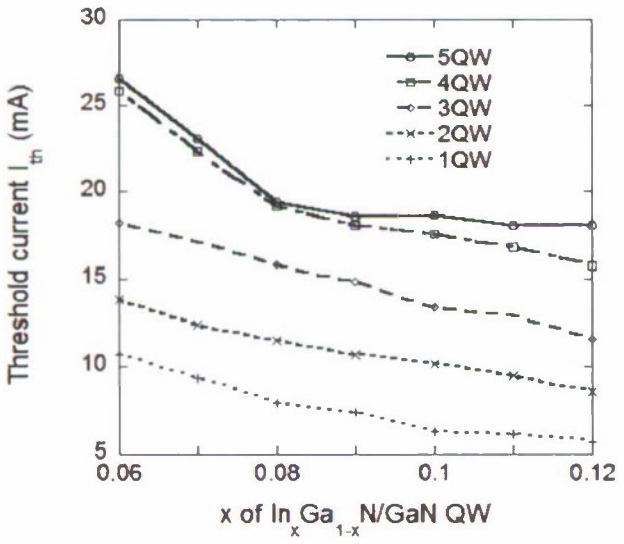


Fig. 8 Threshold current for different QW composition x: $In_xGa_{1-x}N$/GaN and QW number.

3.3 Electron Block (e-Block) Layer

Eleetron barrier was studied around late 1980s for GaAs/AlGaAs laser system to improve the eleetro-optical charaeteristies/threshold eurrent of QW lasers. [20] To eontrol the thermionie emission and overflow of the earriers from QWs, inereasing the barrier height or design a earrier bloek ean effeetively improve earrier eonfinement, and therefore, reduee threshold eurrent for higher emission effieieney. Recently, to improve the GaN threshold and obtain high-power/high-temperature operation, an AlGaN eleetron-bloeking layer was proposed and widely used. [21][22] This barrier is loeated on the p–side, in the direet vieinity of the aetive layer of the GaN laser. The eleetron barrier is usually un-doped to avoid free earrier absorption. In this work, we design a 20nm thiek layer of $Al_{0.35}Ga_{0.65}N$ (15nm away) on the top of the QWs aetive region. The LI simulations are presented in Fig. 9. Compared with Fig. 7(b), the threshold eurrents are mueh smaller for the e-bloek ease, and the optieal output power is more than doubled. In this design, the e-Bloek is a p-doped material that has a larger band gap than its neighboring materials. The e-Bloek is a very thin layer that is added next to the aetive layer to prevent eleetrons from leaking into the p-doped side. Eleetrons that overflow into the p-type side leads to leakage eurrent. Leakage eurrent is detrimental to the operation of the LD because the higher eurrent eauses heating and dissipates non-lasing energy ereating an ineffieient LD. The doped strueture eontains an AlGaN blocking layer that prevents electrons in the aetive region from moving into the p-type side. The e-bloek only bloeks elcctrons and allows the holes to move freely from the p-type side into the aetive region unaffeeted. If a high eurrent is injeeted aeross the diode then a large earrier eoneentration results in the aetive region. This leads to enhaneed non-radiativc carrier reeombination at defects and to an escalation of eleetron leakage from the quantum wells into the p-side of the diode, despite the AlGaN bloeker layer. Thus the e-bloek is very useful in preventing leakage current up to a eertain eurrent, but onee the injeeted eurrent beeomes very high, the bloeking properties are null and the leakage eleetrons eauses the laser to heat, degrade, and eventually breakdown. Inereasing the band gap of the e-bloek ean be an effeetive method of making it very hard for eleetrons to leak onto the p-side. However, an inerease in the band gap of the e-bloek often means a higher eontent of Al and therefore a deerease in the thermal and electrical conductivities. The decrease in the thermal and electrieal eonduetivities eounteraets the improvement of the large band gap e-bloek.

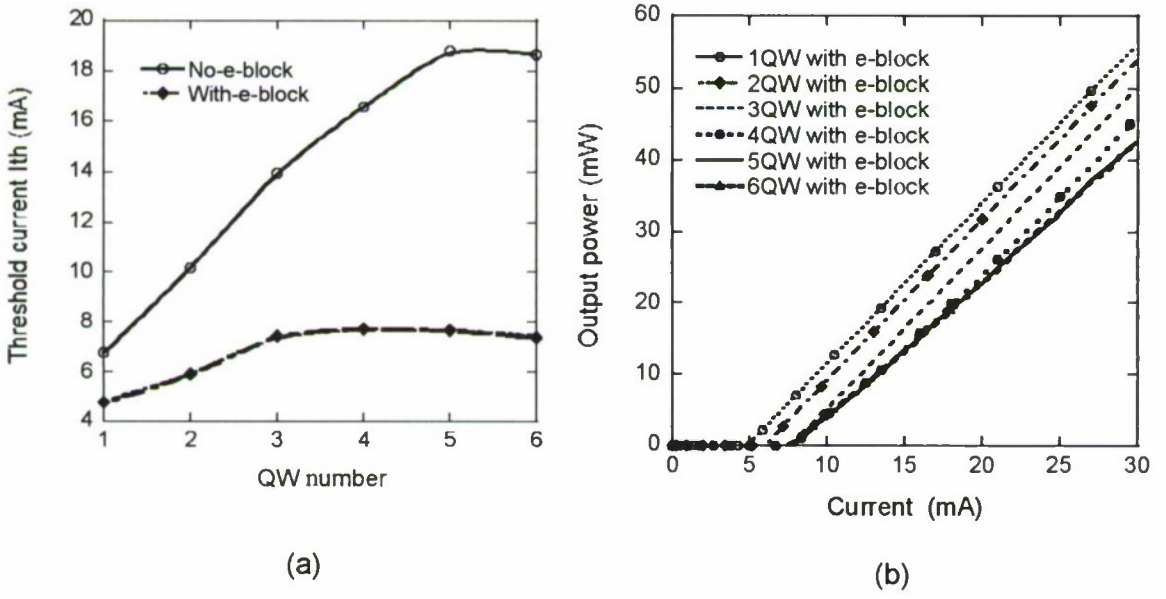


Fig.9 (a) Threshold current with/without e-block and (b) LI curve with e-block.

4. CONCLUSION

GaN laser diode simulation results are presented for the GaN laser design. We diseuss the optieal substrate modes in optieal waveguide. Suppressing substrate mode is very important for the GaN laser. We also present QW modeling results and gain simulation. For our eurrent design, 3QWs will give the best gain performanee. Finally, adding e-bloek

layer on the top of quantum-well active layer will reduce threshold by about 30% to 40%.

ACKNOWLEDGEMENT

This project is supported by Department of the Navy, Office of Naval Research, under Award # ONR 6-N00014-07-1-1152 in 2008, Award # ONR 7-N000140811209 in 2009, USA; "ChunHui" exchange research fellow 2008, Educational Department, China; 973 program-National Basic Research Program of China (2007CB307004); High Technology program (863-2006AA03A113) and National Nature Science Foundation of China (60276032, 60577030 and 60607003).

REFERENCES

[1] Nakamura, S., [The Blue Laser Diode], 2nd ed., Springer, (1997).

[2] Nakamura, S., Senoh, M., Nagahama, S., Iwasa, N., Yamada, T., Matsushita, T., Sugimoto, Y. and Kiyoku, H., "Room-temperature continuous-wave operation of InGaN multi-quantum-well structure laser diodes," Appl. Phys. Lett. 69 (26), 4056-4058, (1996).

[3] Suzuki, M., Uenoyama, T. and Yanase, A., "First-principles calculations of effective-mass parameters of AlN and GaN," Phys. Rev. B 52 (11), 8132-8139, (1995).

[4] Suzuki, M. and Uenoyama, T., "Strain effect on electronic and optical properties of GaN/AlGaN quantum-well lasers," J. Appl. Phys. 80(12), 6868-6874, (1996).

[5] Ohtoshi, T., Niwa, A. and Kuroda, T., "Dependence of optical gain on crystal orientation in wurtzite–GaN strained quantum-well lasers," J. Appl. Phys. 82(4), 1518-1520, (1997).

[6] Domen, K., Horino, K., Kuramata, A. and Tanahashi, T., "Optical gain for wurtzite GaN with anisotropic strain in c plane," Appl. Phys. Lett. 70(8), 987-989, (1997).

[7] LaserMOD v2.0 User Guide, RSOFT Design Group, Inc., New York, (2004).

[8] Wakita, Koichi, [Semiconductor Optical Modulators], Springer, New York City, (1998).

[9] Botez, D., "Analytical approximation of the radiation confinement factor for the TE0 mode of a double heterojunction laser," IEEE J. Quantum Electronics QE-14 (4), 230–232, (1978).

[10] Neamen, Donald A., [Semiconductor Physics and Devices: Basic Principles], 3rd ed., McGraw Hill, New York City, (1992).

[11] Smolyakov, G. A., P. G. Eliseev, and M. Osinski., "Effects of resonant mode coupling on optical Characteristics of InGaN-GaN-AlGaN lasers," IEEE J. Quantum Electronics 41(4), 517-524, (2005).

[12] Jin, X., Zhang, B., Dai, T. and Zhang, G., "Effects of Transverse Mode Coupling and Optical Confinement Factor on Gallium Nitride-Based Laser Diode," The Institute of Physics: Chinese Physics, 17(4), 1274-1278, (2008).

[13] Einfeldt, S., S. Figge, Bottcher, T. and Hommel, D., "Coupling of optical modes in GaN-based laser-diodes," Physica Status Solidi(c). 0(7), 2287-2291, (2003).

[14] Hatakoshi, G. I., Onomura, M., Saito, S., Sasanuma, K. and Itaya. K., "Analysis of Device Characteristics for InGaN Semiconductor Lasers." Japanese Journal of Applied Physics 38(16), 1780-1785, (1999).

[15] Schwarz, U.T., Pindl, M., Wegscheider, W., Eichler, C., Scholz, F., Furitsch, M., Leber, A., Miller, S., Lell, A. and Harle, V., "Near-field and far-field dynamics of (Al,In)GaN laser diodes," Appl. Phys. Lett. 86(16), 161112(3pages), (2005).

[16] Laino, V, Roemer, F., Witzigmann, B., Schwarz, U.T., Fischer, H., Feicht, G., Wegscheider, W., Rumbolz, C., Lell, A. and Harle, V., "Substrate Modes of (Al,In)GaN Semiconductor Laser Diodes on SiC and GaN Substrates," IEEE. J. Quantum Electronics 43(1), 16-24, (2007).

[17] Witzigmann, B., Laino, V., Luisier, M., Schwarz, U.T., Fischer, H., Feicht, G., Wegscheider, W., Rumbolz, C., Lell, A. and Harle, V., "Analysis of temperature-dependent optical gain in GaN-InGaN quantum-well structures," IEEE. Photon. Tech. Lett. 18(15), 1600-1602, (2006).

[18] Schubert, E. F. [Light-Emitting Diodes], 2nd ed., Cambridge University Press, (2006).

[19] Domen, K., Soejima, R., Kuramata, A., Horino, K., Kubota, S. and Tanahashi, T., "Interwell inhomogeneity of carrier injection in InGaN/GaN/AlGaN multiquantum well lasers," Appl. Phys. Lett. 73(19), 2775-2777, (1998).

[20] Blood, P., Fletcher, E.D., Woodbridge, K., Heasman, K.C. and Adams, A.R., "Influence of the barriers on the temperature dependence of threshold current in GaAs/AlGaAs quantum well lasers," IEEE J. Quantum Electronics 25(6), 1459-1468, (1989).

[21] Wiedmann, N., J. Sehmitz, K. Boueke, N. Herres, J. Wagner, M. Mikulla, R. Poprawe, G. Weimann. , "Band-edge aligned quaternary earrier barriers in InGaAs-AlGaAs high-power diode lasers for improved high-temperature operation," IEEE J. Quantum Eleetronies 38(1), 67-72, (2002).

[22] Tu, R.-C., Tun, C.-J., Pan, S.-M., Chuo, C.-C., Sheu, J.K., Tsai, C.-E., Wang, T.-C. and Chi, G.-C., "Improvement of near-ultraviolet InGaN-GaN light-emitting diodes with an AlGaN electron-blocking layer grown at low temperature," IEEE Photon. Tech. Lett. 15(1), 1342-1344, (2003).

Simulation of Three-fold Symmetric Photonic Crystal Structures on Top of GaN LEDs

Simeon Trieu*, Xiaomin Jin*, Chang Xiong, Xingxing Fu**, Xiangning Kang**, Guoyi Zhang** and Bei Zhang****

*Electrical Engineering Department, 1 Grand Avenue, California Polytechnic State University, San Luis Obispo, CA 93407
**State Key Laboratory for Mesoscopic Physics, Physics Department, Peking University, Beijing, China 100871
Author e-mail address: xjin@calpoly.edu

Abstract: A 3-fold symmetric photonic-crystal grating is simulated using improved FDTD-model. Transmission gratings are optimized. Then, the best cases are simulated in GaN-LED models. The maximum extraction efficiency improvement is 40x greater compared to conventional LEDs.

OCIS codes: (140.3325) Laser coupling; (050.5298) Photonic crystals

1. Introduction

Light emitting diodes (LED) have found many applications such as residential and commercial lighting, fiber optics, machine vision, and colored displays. These applications demand low power, high efficiency, high luminosity, and low heat generation. To fulfill these necessary conditions, a grating structure is the solution by creating more angles of escape and also diffraction mechanisms. Grating structures can be patterned with many shapes including: pyramidal, spherical, conical, cylindrical, and so on, but only a few can be fabricated with great success. For example, with the modified laser lift off (M-LLO) technique, holes can be patterned at a 4 micron period instead of etching a random structure. In the experiment at Peking University (PKU), grating depths varied from 75nm to 120nm [1].

Using a 3-D finite difference time domain (FDTD) method, a 3PC grating structure is first optimized and characterized to observe trends for optimizations for certain grating cell-to-cell radii, widths, and heights. Then the optimized grating parameters from a transmission grating are entered into an entire GaN device model. The current literature does not provide optimizations for full 3-D GaN models, only grating simulations. Comparisons between the best case grating parameters are analyzed to verify optimal parameters and discover modeling trends.

2. Simulation Model

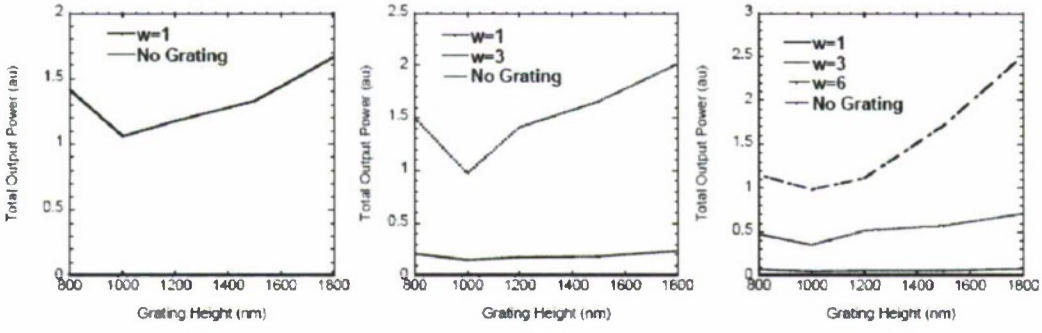
The simulations use a 3-D FDTD method [2-3] for a GaN LED model to calculate final output power. In all simulations, a time monitor is placed in 5 locations: top, front, back, left side, and right side. This allows for a complete picture of the total power output radiating from all facets. After the simulation time has elapsed, the final average output power is taken from each monitor and summed into a total output power measured in arbitrary units (au). Note that a bottom monitor is unnecessary since either the reflections are unimportant, as in the case of a grating-only simulation, or the Ag reflector plate is expected to reflect the vast majority waves back towards the one of the other monitors, as in the full LED model. For thin metal reflectors, only a negligible amount of radiation is transmitted as evanescent waves.

There are two models in this simulation set. The first simulation, a simple semiconductor-air model, aids in discovering the optimized grating parameters. A smaller model is necessary to sweep the parameters since a 3-D FDTD simulation's simulation time and memory requirements expand with simulation domain size as N^4 and N^3, respectively. Therefore, the device size is limited to the amount of RAM and reasonable time we have per simulation. The first simulation has a duration of 500fs to observe the transmission properties of each grating configuration. A transmission grating carved out of the semiconductor surface that can be varied by three variables: grating cell-to-cell radius (A), grating cell width (w), and grating height (d). We step through the possible combinations that are physically allowed (ie. $w \leq A$). For example, the first case would be A = 1 μm, w = 1 μm, d = 800 nm, then A = 1 μm, w = 1 μm, d = 1000 nm, and so on. Light waves are generated from random spatial and directional current sources to better represent incoherent light. 32 current sources are placed at random intervals throughout the source layer to provide a pseudorandom distribution. The simulation results in a good representation of a photonic crystal's response to light from all incident angles (ie. 0 to 90 degrees). This is a new way to setup light sources, which gives grating huge efficiency improvement results compared to other simulation models [2-3], and is a more comprehensive model for the 3PC simulation.

For the second simulation, a full GaN LED model aids in understanding a diffraction grating's response to a multi-layered thin film LED with a Ag reflector plate. The optimal grating parameters from the first simulation are used in this simulation as the grating parameters for the transmission grating. The grating is carved from the undoped GaN layer. Only the top 5 cases from the grating optimization simulations are taken into account for comparison. In this second simulation, the simulation time is increased to 2500fs since we need to account for multiple reflections from the Ag reflector plate in this case.

3. Results

The grating optimization simulations revealed that a transmission grating is optimized when A=w, n other words, when unit cells acquire the most area on the LED surface. In recent literature, this is known as a *filling factor* and can be described as the ratio of the total area covered by the unit cells and the total top surface area. Fig. 1 shows the results of the 3-D FDTD, limited to the interface between GaN to air. A grating with the parameters *A*, *w*, and *d*, was varied in steps. It is clear that as the grating height increases, the total output power also increases. The improvement, in all cases of A=1, 3, and 6 μm, at d=800 nm indicates that the relationship between *d* and output power may have a sinusoidal-like relationship.



(a) (b) (c)

Fig.1 Total output power for a GaN-to-Air grating structure: (a) A=1μm., (b) A=3μm., and (c) A=6μm.

Next, a 3-D FDTD full GaN LED model was simulated with the best cases of the grating-only simulations. These best cases and their results for the full GaN model are shown in Table 1. The grating with the characteristics of A=3 μm, w=3 μm, d=1800 nm produces the best transmission grating. Even though the A=6 μm, w=6 μm, d=1800 nm grating had the best results in the grating-only simulations, the results in the full GaN LED model indicate that it performed the worst out of all 5 gratings simulated. Grating-only simulations are not sufficient indicators of optimized grating structures when considering full GaN LED models. Internal modes must be considered as well as multiple reflections from any reflectors.

Table 1 – Full GaN Model 3-D FDTD Simulations for Best Case Gratings

A (um)	w (um)	d (nm)	P ($\|E\|^2$, au) of the Grating-Only	P ($\|E\|^2$, au) of the Full GaN Model	Improvement factor for Full GaN Model
3	3	1800	2.010	43.292	43.77
1	1	1800	1.658	42.381	42.83
3	3	1500	1.651	41.165	41.57
6	6	1500	1.703	40.807	41.20
6	6	1800	2.508	40.359	40.74
No Grating			**0.00712**	**0.96688**	**0**

4. References

[1] Bao, K., Kang, X., Zhang, B., Dai, T., Sun, Y., Fu, Q., Lian, G., Xiong, G., Zhang, G. and Chen, Y., "Improvement of light extraction from GaN-based thin-film light-emitting diodes by patterning undoped GaN using modified laser lift-off," Appl. Phys. Lett. **92**, 141104 (2008).

[2] Yee, K., "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", IEEE Trans. Antennas and Propagation **14**, 302–307 (1966).

[3] Taflove, A., Brodwin, M. E. "Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations", IEEE Trans. Microwave Theory and Techniques **23**, 623–630 (1975).

Acknowledgments

This project is supported by the NSF 08-603 OISE – EAPSI 2009 fellowship award and ONR 7-N000140811209.

Study of Top and Bottom Photonic Gratings on GaN LED With Error Grating Models

Simeon S. Trieu and Xiaomin Jin, *Senior Member, IEEE*

***Abstract*—The gallium nitride (GaN) light emitting diode (LED) top-bottom (or transmission-reflection) grating simulation results with error grating model are presented. The microstructure GaN bottom hole and top pillar gratings are calculated and compared with the non-grating (flat) case. Grating shapes simulated are either conical or cylindrical. A direct comparison of 181 different combined transmission-reflection grating cases using the finite difference time domain method is presented. The simulation results show that simple or direct combinations of the optimized top grating with the optimized bottom grating only produce a 42% light extraction improvement compared to the non-grating case, which is much lower than that of an optimized single grating case. This is due to the mismatch of grating parameters with the direct addition of the second grating structure, which changes the optical field distribution in the LEDs. Therefore, it is very important to optimize both top and bottom gratings simultaneously for the double-grating design. We also show the optimization of a double grating structure can achieve better performance than a single grating. Finally, transmission-reflection error gratings are also presented. It is also the first time to present randomization in GaN LED grating design and its effects in fabrication. Our data shows that the favorable light extraction improvement is at approximately 10–15% randomization. The randomization can achieve 230% improvement over the original grating at a randomization intensity factor of 12.8%.**

***Index Terms*—Error grating, finite-difference time domain, light emitting diode, photonic crystal.**

I. Introduction

LIGHT EMITTING diodes (LEDs) have become an area of intensive research in an effort to develop more efficient, brighter, and greener light sources. Light emitting diodes are used extensively in displays, light bulbs, photonics, and machine vision. To meet the needs of these applications, light sources must be low-power, generate less heat, and achieve high luminosity. There are two methods to improve these factors in LEDs: 1) increase electroluminescent efficiency; and 2) increase light extraction efficiency [1]. Therefore, it is crucial to improve light extraction efficiency of gallium nitride (GaN) LEDs to reduce energy consumption, limit heat generation, and increase luminosity. Usually GaN LEDs have poor light extraction efficiencies due to: 1) the high refractive index of semiconductor based devices that results in a low critical angle and total internal reflection; 2) the absorption of light within the device due to dislocations and defects within the GaN crystal; and 3) inefficient device design and structure [2]–[4].

One of the key limitations for light emissive devices is the light trapping due to the device layers' low critical angle. The light generated within the device is totally internally reflected. A number of solutions have been proposed to solve this problem. A resin can be affixed over the device to create a larger escape angle to air [5]. To further mitigate the problems of total internal reflection, the emission surface can be patterned to form a transmission grating that offers the trapped light more angles of escape. This can be done with a variety of methods including wet etching with an amorphous sacrificial layer [6]–[9] or by laser etching to obtain a more periodic spacing [10]–[14]. In addition, it has been shown that the same patterning can also apply to a Ag reflector plate in either pillar or hole grating shapes to form a reflection grating [15], [16].

The second inefficiency of the device is due to absorption losses due to dislocation and defects within the GaN crystal. This is a critical aspect of light extraction efficiency since the photon is not escaping from the device, but also because photon absorption within the semiconductor generates heat. So, it is important to extract light from the device both in greater quantity and speed before recombination [4], [17]. A grating structure will help solve this issue by creating more angles of escape. Grating structures can be patterned with many shapes including: pyramidal, spherical, conical, cylindrical, and so on, but only a few can be fabricated with great success. For example, with the modified laser lift off (M-LLO) technique, holes can be patterned at a 4 μm period instead of etching a random structure. The technique uses a sapphire backplane, UV light, and a high power KrF laser to etch the nanostructure onto u-GaN. In the experiment at Peking University (PKU), grating depths varied from 75 nm to 120 nm [18]. Other fabrication techniques exist, such as imprint lithography, that can produce similar air holes that measure 180 nm in diameter, with a depth of 67 nm, and a period of 295 nm [20].

Etching a structure on the extraction surface and/or on the bottom reflection surface commonly solves the light

Manuscript received February 18, 2010; revised April 22, 2010. This work was supported by the Department of the Navy, Office of Naval Research, under Award ONR 6-N00014-07-1-1152, in 2008, and Award ONR 7-N000140811209, in 2009. The work of S. Trieu was supported by the NSF 08-603 OISE–EAPSI 2009 Fellowship Award ("EAPSI: Light Extraction Improvement of GaN-based Light Emitting Diodes"), in 2009.

The authors are with California Polytechnic State University, San Luis Obispo, CA 93407 USA (e-mail: simeon.trieu@gmail.com; xiaomin-jin@yahoo.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JQE.2010.2051020

trapping issue. The single grating simulation has been studied intensively already. However, there is no publication directly comparing the effect of top grating, bottom grating, and combinations of both in theory. It is also not practical to fabricate all the double grating cases to obtain the best cases. Even our calculation here is only limited to one particular GaN LED structure developed by PKU [9], we simulate at least 181 cases based on the micro-structure structures and present very interesting results. In addition, in fabrication etching also involves defects, so the grating structure is not a perfectly periodic crystal. We also develop an error grating model to study the effects of the fabrication randomization. To date, there is no publication on the error grating model of the top and bottom grating structures on GaN LEDs. For all simulations, we calculate the effects on light extraction efficiency with different combinations of grating structures using a finite difference time domain (FDTD) method. In this paper, the simulation model is presented in Section II, top and bottom grating simulation results are presented in Section III, the error grating model is presented in Section IV, and conclusions are presented in Section V.

II. Simulation Model

A. FDTD

We analyze a GaN LED device by the Finite Difference Time Domain method. Since FDTD is derived from Maxwell's equations, it can accurately simulate the small grating parameter effects of refraction in device materials, reflection due to linear dispersion or total internal reflection, transmission of escaping light from the LED, and scattering at the grating.

We begin with the differential forms of Maxwell's equations

$$\frac{\partial \vec{H}}{\partial t} = -\frac{1}{\mu}\left(\nabla \times \vec{E}\right) - \frac{\sigma^*}{\mu}\vec{H}$$
$$\frac{\partial \vec{E}}{\partial t} = \frac{1}{\varepsilon}\left(\nabla \times \vec{H}\right) - \frac{\sigma}{\varepsilon}\vec{H}. \quad (1)$$

From these equations, we break up the spatial components into each respective axis in the Cartesian coordinate system, which results in the following:

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial z} - \frac{\partial E_x}{\partial y}\right)$$
$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_x}{\partial x} - \frac{\partial E_x}{\partial z}\right)$$
$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x}\right)$$
$$\frac{\partial E_x}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \sigma E_x\right) \quad (2)$$
$$\frac{\partial E_y}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \sigma E_y\right)$$
$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_z}{\partial y} - \sigma E_z\right).$$

TABLE I
Simulation Parameters for Both Top and Bottom Gratings

Parameter	Brief Description	Simulated Ranges
A	The period of the unit cells	1 to 6 mm
w	The width of each unit cell	1 to 6 mm
d	The height of each unit cell	50 to 200 nm

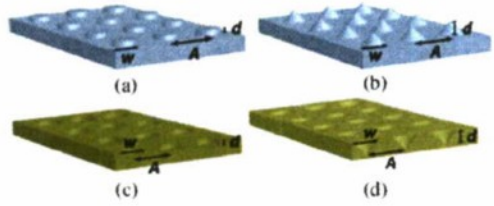


Fig. 1. Different hole and pillar gratings in our simulation. (a) Cylindrical pillar. (b) Conical pillar. (c) Cylindrical hole. (d) Conical hole.

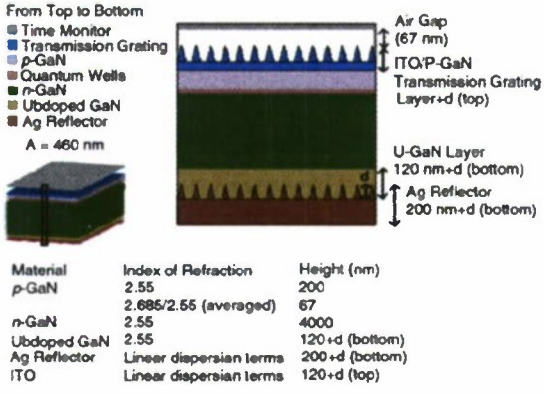


Fig. 2. GaN LED layer model.

Next, we employ a Yee's mesh defined as interleaving E and H component fields [19]. Accuracy of a Yee's mesh is dependent on the grid size (i.e., on Δx, Δy, and Δz). Based on Yee's implementation, each H component field is offset by $n+\frac{1}{2}$ while the E component fields are offset by $n+1$. The FDTD in conjunction with a Yee's mesh can simulate structures of arbitrary length and size. However, the model is limited in size due to the simulation time and amount of memory required to simulate larger devices. Methods to ease the requirements of 3-D models include use of Graphics Processing Units (GPU), parallel processing, and mode simplifications [21]–[23]. Since 3-D model parameters scale in simulation time like N^4 and in memory like N^3, the 2-D FDTD is used to simulate this model to avoid unnecessary complications [24].

To define a regular spacing between unit cells in a crystal lattice arrangement, we employ three parameters: unit-cell period (A), unit-cell height (d), and unit-cell width (w), which are shown in Table I and Fig. 1. The unit-cell period, A, is the length from center-to-center between unit cells. In 3-D, the parameter w represents a diameter in the case of a circular structure (i.e., sphere, cone, and cylinder) or a length of a side in the case of a box structure (i.e., cube, rectangular cube). The unit-cell height, d, can represent either the depth of the bottom hole or the height of the top pillar.

B. LED Structure

The structure of the LED is shown layer-by-layer in Fig. 2, which is based on the Refs. [16], [18] experiment. From bottom to top, the silicon submount is mounted to a 200 nm layer of silver (Ag) forming both a reflector plate and also filling in holes with a bottom grating height of d_{bottom} in the 120 + d_{bottom} nm undoped GaN layer (u-GaN), which forms the hole reflection grating structure. Next, a 4 μm layer of n-type GaN (n-GaN) is placed on the hole array, followed by a layer of quantum wells of 67 nm thickness, and a 200 nm p-type GaN layer. The transmission grating structure, of height d_{top}, is composed of a grating structure and mounting material. If ITO is used for the transmission grating, a 200+d_{top} nm layer of ITO is deposited on top of the p-GaN layer, the transmission grating being formed from the ITO layer. If p-GaN is used for the transmission grating, the grating can be formed directly on top of the p-GaN layer, making the height of the p-GaN layer a total of 200+d_{top} nm. The double grating structures, one on the top and one on the bottom, constitute a so-called transmission-reflection grating or top-bottom grating.

III. Top and Bottom Grating Simulation Results

The simulation starts when the LED turns on from an off state. There is a monitor at a distance of 67 nm, or one grid size distance, above the LED to collect the light emission. For each structure simulation, we must wait until the LED/monitor reaches a steady state, after which we can extract the constant wave (CW) average power. Each simulation sweep of the parameters produces a set of 20 000 data points (2000 fs at 0.1 fs/step), and from this data set, we determine average power over the ending cycle. In this range, maximum steady-state power is radiated due to the CW source and reflections from the grating. This procedure is done for each case.

There are a total of six transmission gratings and two reflection gratings studied. The transmission gratings are a non-lossy conical P-GaN grating, a non-lossy cylindrical P-GaN grating, a non-lossy cylindrical ITO grating, a lossy cylindrical ITO grating, a non-lossy conical ITO grating, and a lossy conical ITO grating. And the reflection gratings are conical reflection grating and cylindrical reflection grating. For each single grating case, we sweep A from 1 to 6 μm, w from 1 to 6 μm (only for cases where $w < A$), and d = 50 nm to 200 nm. And we give out the optimized structure and list the results in the Table II. The best cases for single gratings from each grating type were taken from our previous studies [16], [25], which are also compared to and agreed with experimental data provided by PKU [9], [18]. References [9] and [18] are the published experimental data on our simulated LED structure. Reference [9] only has electroluminescence data. Reference [18] presents the experimental data of the M-LLO LEDs for the light extraction improvement. The experiment shows about 120% improvement for a 120-nm-depth grating and 66% for a 75-nm-depth grating [18], which agrees with our simulation. Our simulation gives the following: the light extraction improvement is about 150% for the 120-nm-depth grating and 65% for the 75-nm-depth grating [16]. A more detailed comparison of experimental data and simulation are presented in [16]. Our simulation also shows that the single grating can achieve about 80–165% light extraction efficiency improvement. Based on the above information, the best-case top transmission grating was paired with a best-case bottom reflection grating to form a transmission-reflection grating, a total of 12 transmission-reflection grating combinations. In the simulation, lossy materials have a loss factor of $|\alpha| \neq 0$, while the non-lossy materials have α = 0. The extinction coefficient, or the imaginary portion of the index of refraction n, is defined as α: $n_{\text{comp}} = n_{\text{real}} + j^{*}\alpha$. In our simulation, the α factor of lossy ITO is 0.01 dB/mm. "Conical" and "Cylindrical" refer to grating cell shapes. Transmission grating shapes refer to "pillars," while reflection grating shapes refer to "holes." Reflection grating materials are always u-GaN holes mounted and filled in with Ag. The p-GaN transmission grating is formed from the p-type GaN semiconductor, while the ITO transmission grating is formed from a deposited layer of ITO on top of the p-GaN layer. The optimized single grating suggested from our simulation is the p-GaN conical top grating; the next one is the top cylindrical p-GaN. The conical top ITO gratings have less light improvement compared to all the other top and bottom single grating cases. The bottom reflection gratings also give reasonable good results on light extraction improvement, which is about 117%. In general, we found A = 1 μm gives the best results for single grating simulations as shown in Table II, which suggest a nano-structure grating may be the future trend for better light extraction improvement.

We simulate 181 cases in total with different A, w, d values, including the single top, the single bottom, and the top-bottom gratings. Then, we rank all of them based on light extraction improvement. The brief results of the double-grating simulation are shown in Table III. The best-case top transmission gratings are matched with the best-case bottom

TABLE II
Best Cases for Top Transmission and Bottom Reflection Gratings

	Grating Type	A μm	W μm	D nm	Power a.u.	Inc. %
Transmission (top)	Non-Lossy Conical p-GaN	1	1	138	12.374	165
	Non-Lossy Cylindrical p-GaN	2	1	156	10.926	134
	Non-Lossy Cylindrical ITO	1	1	126	9.7258	109
	Lossy Cylindrical ITO	1	1	118	9.0752	95
	Non-Lossy Conical ITO	4	4	166	8.8755	90
	Lossy Conical ITO	4	4	164	8.4015	80
Reflection (bottom)	Conical	1	1	200	10.13	117
	Cylindrical	2	1	136	9.8135	111
Reference	None	–	–	–	4.6608	0

TABLE III
TOP/BOTTOM GRATING OPTIMIZATION RESULTS

Case No.	Structure Name Top Transmission	A μm	w μm	d nm	Structure Name Bottom Reflection	A μm	w μm	d nm	Power a.u.	Inc. %
1	Lossy Conical ITO	4	4	164	Cylindrical	2	1	136	6.6252	42
2	Non-Lossy Conical P-GaN	1	1	138	Cylindrical	2	1	136	6.2106	33
3	Non-Lossy Conical ITO	4	4	166	Cylindrical	2	1	136	6.1401	32
4	Non-Lossy Cylindrical P-GaN	2	1	156	Cylindrical	2	1	136	6.1368	32
5	Non-Lossy Cylindrical ITO	1	1	126	Cylindrical	2	1	136	5.5491	19
6	Lossy Cylindrical ITO	1	1	118	Cylindrical	2	1	136	5.5181	18
7	Non-Lossy Conical ITO	4	4	166	Conical	1	1	200	5.4614	17
8	Lossy Conical ITO	4	4	164	Conical	1	1	200	5.3947	16
9	Non-Lossy Conical P-GaN	1	1	138	Conical	1	1	200	5.2641	13
10	Non-Lossy Cylindrical P-GaN	2	1	156	Conical	1	1	200	5.2185	12
11	Non-Lossy Cylindrical ITO	1	1	126	Conical	1	1	200	5.0291	8
12	Lossy Cylindrical ITO	1	1	118	Conical	1	1	200	4.8967	5
13	Non-grating (Reference)	–	–	–	None	–	–	–	4.6608	0
14	Lossy Cylindrical ITO	1	1	186	Conical	1	1	200	4.2404	−9

reflection gratings to show effects of both together, and results are reported in percent increase with respect to the non-grating case. There are 12 combinations of six transmission gratings and two reflection gratings studied, which are listed as Cases 1–12 in Table III. These types of gratings show significant improvement over the non-grating case, or case 13 in Table III. The structure types by ranking show that cylindrical reflection gratings are preferred when coupled with a transmission grating structure since ranks 1 to 6 all have a cylindrical reflection grating. Although conical reflection gratings sometimes perform better as a single grating case, this simulation reveals that optimal conditions for a single grating do not always apply for the transmission-reflection grating case. Also, conical transmission gratings perform better than cylindrical transmission gratings when combined with bottom reflection gratings. The top three ranks are dominated by conical structures. And the following three rankings are all cylindrical transmission gratings.

Since both the top and the bottom grating change the optical field distribution in the non-grating LED and influence the light extraction, they have coupling effects and need to be matched in design. Otherwise, they may not improve light extraction. In general, there is a clear trend that shows conical transmission gratings and cylindrical reflection gratings are preferred. However, the results of the transmission-reflection gratings, derived from the best cases of reflection and transmission gratings only, perform more poorly (about 42% maximum) than just the single grating structures alone. This can be explained by the trapped optical field changing with grating structure, since the scattering due to the grating is dependent on the grating structure parameters. The optimal conditions that produce the greatest light emission in a single grating case do not necessarily apply to the transmission-reflection grating case. When a reflection grating is paired with a transmission grating, the optical fields are altered within the device, therefore producing unfavorable light extraction due to the poor coupling or the mismatch between the gratings and the newly reorganized optical fields. In the worst case, a destructive interference or mismatch will degrade light extraction efficiency, which is also presented in Table III (Case 14), for example. This last ranked structure even has worse light extraction efficiency (about −9%) than the non-grating case.

The best performing grating of this transmission-reflection simulation set uses lossy conical-shaped ITO pillars as a transmission grating material matched with cylindrical-shaped u-GaN holes as a reflection grating. This type of grating has a 42% improvement over the non-grating case. This is an interesting result, since the both lossy and non-lossy conical ITO transmission gratings ranked lowest in the transmission grating. The difference between the lossy and non-lossy materials is an imaginary loss factor in the index of refraction described by

$$n_{\text{comp}} = n_{\text{real}} + jn_{\text{imag}} \tag{3}$$

where $n_{\text{imag}} = \frac{\gamma\lambda}{4\pi}$, and γ is the exponential loss factor in units of μm^{-1} and λ is the wavelength of light. It is possible that evanescent waves are penetrating the thin grating structure, causing greater light extraction efficiency. Also, the optical field for the cylindrical reflection grating may match better to the lossy transmission grating than the non-lossy transmission grating. In addition, the worse single grating combination may have better potential to improve its light extraction efficiency compared the best single grating case.In the total rankings of all simulated cases, the best single transmission and the best single reflection grating combination ranks only at 158 out of 181 total cases simulated, which is rank 9 in Table III. Clearly, the direct addition of optimal grating structures does not linearly add to the total output power. The optical fields have an optimal set of grating parameters that may be disturbed when another grating is introduced. It is possible that another set of optimal grating conditions exist, but it must be discovered by a thorough sweep of all top and bottom grating parameters (six parameters in total). This is very important and critical in the double-grating LED design. But it would take an extraordinary amount of time. It is clear now that the transmission-reflection gratings simulated here are not overall-optimal grating structures, since none of them outperform the single grating cases.

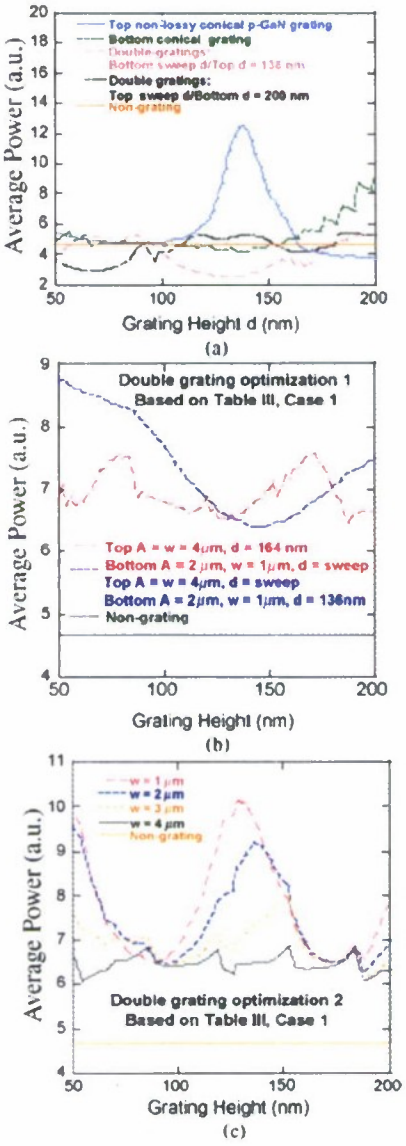


Fig. 3. (a) Comparison gratings with only top, only bottom, and double gratings based on Table III (Case 9) (both top and bottom grating $A = 1\,\mu$m, $w = 1\,\mu$m). (b) Double grating optimization 1 based on Table III (Case 1), sweeping either bottom grating d or top grating d. (c) Double grating optimization 2 based on Table III (Case 1), top grating $A = w = 4\,\mu$m, $d = 52$ nm and bottom grating $A = 4\,\mu$m, $w =$ sweep, $d =$ sweep.

Fig. 3(a) shows an example of above case and explains the concept of mismatching. Here, we sweep the top grating height with the bottom grating height fixed and vice versa. The best top grating, non-lossy conical p-GaN transmission grating, has optimal parameters of $A = 1\,\mu$m, $w = 1\,\mu$m, $d = 138$ nm. If we vary d as a fabrication error from 118 nm to 186 nm, then even slight variations result in, not only a non-optimal grating, but also the worst case. The best bottom grating, conical reflection grating, has best light extraction at $A = 1\,\mu$m, $w = 1\,\mu$m, $d = 200$ nm (a limit of our simulation). These two gratings are totally mismatched. Therefore, when combining them, we only can achieve 13% (5.26 a.u. power output, Case 9 in Table III) of light extraction improvement, which is much lower than that of the optimized single gratings. Fig. 3(a) also shows that the nano-level variation of depth can be a primarily influence of mismatching degree between optical fields. The experimental data published in [18] also supports the strong dependence of light extraction efficiency on the grating depth. We also perform some double grating optimization simulations based on Case 1 in Table III by varying one parameter at a time. We do not cover all the parameter's possible values, but we can still show the optimized double grating can be better than the single gratings. The results are shown in Fig. 3(b) and (c). With the optimization of only the height value, we can achieve 8.76 a.u. power output, which is an 87% improvement. In addition, we also sweep A and w of bottom grating with fixed top grating and fixed d of bottom grating. When the top lossy conical ITO grating at $A = w = 4\,\mu$m, $d = 52$ nm and bottom cylindrical reflection grating at $A = 4\,\mu$m, $w = 1\,\mu$m, $d = 136$ nm, the power output is 9.76 a.u., which is about a 110% improvement. Finally, a further optimization with the same transmission grating but varying cylindrical reflection grating shows that $A = 4$, $w = 1$, $d = 128$ nm gives an improvement of 118%, as shown in Fig. 3(c). This result is not an absolute maximum which should be found from a thorough sweep of all top and bottom grating parameters' possible values. However it is still better than the optimized single grating case, which is 80% for this type of single top grating and 111% for this type of single bottom grating. All these calculations show that the optimization of a double grating is much more sensitive and crucial than the best single grating pair and can give better results.

To further prove field variation theory, we output the optical field distribution of GaN LEDs. First, the optical field distributions in the top grating only in Fig. 4(a) and bottom grating only in Fig. 4(b) are very different from each other. When the top grating and bottom grating are added together as in Table III (Case 1), shown in Fig. 4(c), the optical field distribution will not prefer either top grating or bottom grating. An optical field cancelation effect (rather than enhancement effect) may give a light extraction efficiency of 42%, which is much lower than single grating cases. Second, the best top or bottom-grating optical field distributions are shown in Fig. 4(d) or (e). Compared to Fig. 4(a) or (b), optical field of (d) or (e) is less confined in the active region. There is more optical field distribution variation between (a) and (d), compared to that of (b) and (e). There is also more light extraction improvement difference between (a) at 80% and (d) at 165%, compared to that of (b) at 111% and (e) at 117%, as well. Finally, we present the non-grating LED case in Fig. 4(f), which has the strongest field around the active region compared to all other cases. In general, our simulation shows that better light extraction has more optical fields outside the active region. The top grating optical field and bottom grating optical field in our simulation may not matched, which are very interesting results. The exact correlation of the double grating design, optical field, and light extraction efficiency requires intensive simulation, which is outside the scope of this paper.

It is very surprising that the best single grating is not related to the best double grating at first. This result is very useful for most people in the field, since there are now many

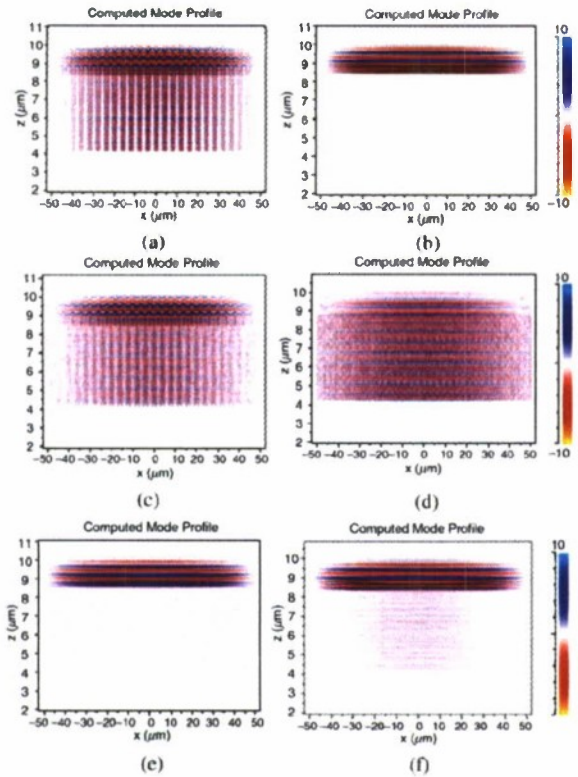


Fig. 4. (a) Top conical grating only $A = 4\,\mu$m, $w = 4\,\mu$m, $d = 164$ nm. (b) Bottom cylindrical grating only $A = 2\,\mu$m, $w = 1\,\mu$m, $d = 136$ nm. (c) Top and bottom grating (a) + (b), Table III (Case 1). (d) Best top grating: conical $A = 1\,\mu$m, $w = 1\,\mu$m, $d = 138$ nm (Table II). (e) Best bottom grating: conical $A = 1\,\mu$m, $w = 1\,\mu$m, $d = 200$ nm (Table II). (f) Non-grating LED case.

experimental results on single gratings. It is very easy to design a double grating using the best single gratings, which can complicate the fabrication procedure. However, there is not much gain in overall performance of LEDs, since the effects of grating structures are not necessarily linearly additive to improve light extraction efficiency. Before fabrication, detailed studies need to be carried out to determine maximum coupling efficiency according to the optical field variation when adding an additional grating structure. Also, matching grating is a very important concept in double-grating design. As long as you can match them, it is not critical which grating you choose according to your fabrication capability. Therefore, the grating design is very open in this sense.

IV. Effects of Randomization on Top and Bottom Gratings

To further understand the effects of fabrication defects on transmission-reflection gratings, we simulate the transmission-reflection grating model with each cell randomly shifted a distance along the axis in varying degrees of randomization intensity to create an error grating model. While the widths of holes can be fabricated to great precision, often the placement of holes is a cause for concern, as it affects light extraction efficiency of otherwise ordered photonic crystal structures. The error grating model is very important to explain difference between model/optimization and fabrication.

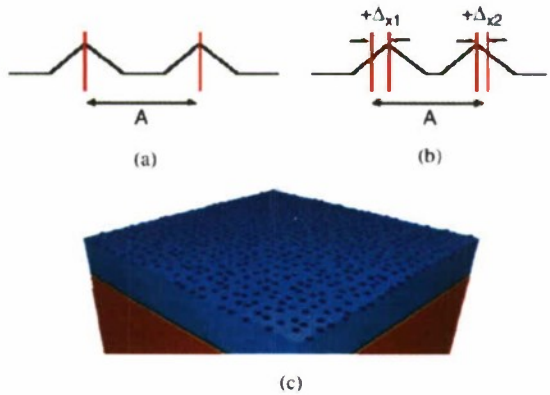


Fig. 5. Error grating model. (a) Normal reference grating model. (b) Error grating model with both positive and negative shifts. (c) Error grating model on GaN LED.

Our simulation is still based on the 2-D FDTD method. An example of random displacements is shown in Fig. 5(b) with a normal grating as a reference in Fig. 5(a). Displacements can move either way from the original center point. The error grating model shows examples of a positive and negative Δ_x shift. This randomization is applied to all grating cells in the photonic crystal arrangement with the following equation:

$$x_{\text{pos}} = N^{*}period + (2^{*}rand - 1)^{*}R^{*}period \qquad (4)$$

where N is an integer index defining the original grating cell location, *period* is the grating period (A), *rand* is a pseudo-randomly generated number from 0 to 1, and R is the randomization factor from 0 to 1. The quantity Δ_x represents $(2^{*}rand\text{-}1)^{*}R^{*}period$ in (4). By varying R, which applies to all grating cells, from 0 to 1 in 40 steps, the individual *rand* factor can be emphasized or deemphasized. This process is repeated for each of the 12 transmission-reflection grating models. And above error grating models are placed into GaN LED device models as shown in Fig. 5(c) to calculate light extraction efficiency.

The effects of randomization in gratings, in general, appear to help light extraction efficiency, peaking at about a randomization factor of 10% in most simulations. Fig. 6 shows the results for each transmission and reflection grating pair. All data are normalized according to its $R = 0$ values for easy reading. Fig. 6(a) shows the results for a conical reflection grating matched with each optimal transmission grating. These top-bottom grating combinations are Case 12, Case 8, Case 11, Case 7, Case 10, and Case 9 in Table III. Fig. 6(b) shows the results for a cylindrical grating paired with each optimal transmission grating. These are Case 6, Case 1, Case 5, Case 3, Case 4, and Case 2. As the randomization intensity factor is increased from 0 to 1, many transmission-reflection gratings experience an increase in light extraction efficiency around a 10% displacement of grating cells. In essence, a slight random variation or fabrication defect in grating cells would not only be beneficial but desirable for many transmission-reflection grating types up to a variation of 10–15% for most double grating cases. The best case was the non-lossy cylindrical ITO transmission and conical reflection gratings which can reach

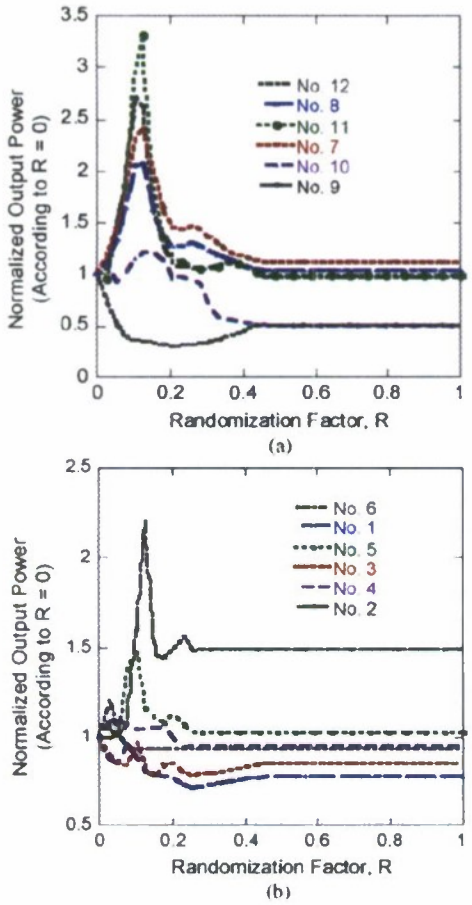


Fig. 6. Results of error grating simulation for: (a) conical reflection gratings paired with transmission gratings: Case 12, Case 8, Case 11, Case 7, Case 10, and Case 9 in Table III; (b) cylindrical grating paired with transmission grating: Case 6, Case 1, Case 5, Case 3, Case 4, and Case 2 in Table III.

up to a 230 % improvement over the original transmission-reflection grating model at a randomization intensity factor of 12.8%. Also, after $R = 10\%$, light extraction efficiency drops, and after $R = 50\%$ settles to a steady state value. Usually, the randomization creates local variation of grating structure, A and w. As top and bottom gratings are only optimized separately, it is possible that a small local perturbation could result in more matching gratings and could introduce local light extraction improvement. A conical single reflection grating has lower performance than that of a cylindrical grating. It has more benefit from a randomization of the grating.

One case that did not improve at around $R = 10\%$ is the case of a non-lossy P-GaN conical top grating ($A = 1\,\mu$m, $w = 1\,\mu$m, $d = 138$ nm) and the conical reflection grating ($A = 1\,\mu$m, $w = 1\,\mu$m, $d = 200$ nm). In the single grating simulations, the respective reflection and transmission gratings represent the best case, as shown in Table II. Since these gratings were already optimized to the best case in the single grating case, the randomization effect is not strong enough to offset destruction from the double gratings. On the other hand it may further decrease light extraction efficiency by adding more local destruction or mismatching, hence the drop shown in Fig. 6(a) as the randomization intensity factor increased.

Experimentally, the double grating could give a comparably higher output power as compared with the single grating one, when some suitable grating/texturing processes are used on both the top ITO/p-GaN and bottom substrate layers. In this paper, we only prove that the best single grating is not necessarily the best double grating option in Section III. As for the experiment, we show here that if fabrication is not perfect, the light output variation can be about 230% improvement in our double grating error model. Therefore, the experimental data of a double grating can be better than a single grating in either matching grating concepts, error grating effects, or both.

Randomization actually has a positive effect on many transmission-reflection grating structures for randomization factors of 10–15%. It increases light extraction efficiency while having the added benefit of alleviating some of the fabrication complexities demanded by strict periodicities in photonic crystal LEDs. Until now, we have not found any papers on the randomization model or our so called error model in GaN LEDs. We believe the LED error grating model is unique. There isn't any experimental data which claims error percentage in their fabrication process. Nevertheless, our model can be used to predict the fabrication error or used as data analysis in the real design in the future.

V. Conclusion

Transmission-reflection grating structures, taken from the optimized cases of transmission and reflection gratings alone, were simulated using FDTD. In general, grating structures will improve light extraction of LEDs compared to non-grating structure. The simulation results also show that the optimized single grating optical fields were altered significantly by the introduction of a second grating structure. The double grating can achieve a light extraction improvement of 118% with respect to the non-grating case. One of the worst double grating cases is can reduce light extraction efficiency by 9% with respect to the non-grating case. Therefore, all of the grating parameters would need to be swept for both transmission and reflection grating cases simultaneously in terms of A, w, and d in order to find the optimized double-grating structure.

We also presented error grating models. A fabrication defect of around 10% actually helps increase light extraction efficiency. In the best case, 230% improvement over the original transmission-reflection grating model can be achieved. Not all models exhibit such an increase. In general, fabrication of photonic crystal LEDs do not necessarily need to keep photonic crystal arrangements so ordered and exact. Low intensity random defects in periodic photonic crystals are not only beneficial but desirable in many grating structures.

Acknowledgment

The authors would like to thank Prof. B. Zhang, Dr. X. N. Kang, and Dr. G. Y. Zhang from Peking University, Beijing, China, for providing expert advice on the GaN LED grating experiments and modeling.

REFERENCES

[1] O. Svensk, P. T. Torma, S. Suihkonen, M. Ali, H. Lipsanen, M. Sopanen, A. Odnoblyudov, and V. E. Bougrov, "Enhanced electroluminescence in 405 nm InGaN/GaN LEDs by optimized electron blocking layer," *J. Crystal Growth*, vol. 310, no. 23, pp. 5154–5157, 2008.

[2] R. H. Horng, C. C. Yang, J. Y. Wu, S. H. Huang, C. E. Lee, and D. S. Wuu, "GaN-based light-emitting diodes with indium tin oxide texturing window layers using natural lithography," *Appl. Phys. Lett.*, vol. 86, no. 22, p. 221101, 2005.

[3] H. W. Huang, C. C. Kao, J. T. Chu, H. C. Kuo, S. C. Wang, and C. C. Yu, "Improved light-output and electrical performance of InGaN-based light-emitting diode by microroughening of the p-GaN surface," *IEEE Photon. Technol. Lett.*, vol. 17, no. 2, pp. 9383–9385, 2005.

[4] T. X. Lee, C. Y. Lin, S. H. Ma, and C. C. Sun, "Analysis of position-dependent light extraction of GaN-based LEDs," *Opt. Express*, vol. 13, no. 11, pp. 4175–4179, May 2005.

[5] G. Hatakoshi, Y. Hattori, S. Saito, N. Shida, and S. Ninoue, "Device simulator for designing high-efficiency light-emitting diodes," *Jpn. J. Appl. Phys.*, vol. 46, no. 8B, pp. 5419–5425, 2007.

[6] Y. A. Chen, N. Y. Liang, L. H. Laih, W. C. Tsay, M. N. Chang, and J. W. Hong, "Improvement of electroluminescence characteristics of porous silicon LED by using amorphous silicon layers," *Elec. Lett.*, vol. 33, no. 17, pp. 1489–1490, 1997.

[7] Y. A. Chen, N. Y. Liang, L. H. Laih, W. C. Tsay, M. N. Chang, and J. W. Hong, "Improvement of current injection of porous silicon," *Jpn. J. Appl. Phys.*, vol. 36, pp. 1574–1577, 1997.

[8] H. W. Huang, C. C. Kao, H. C. Chu, S. C. Wang, and C. C. Yu, "Improvement of InGaN–GaN light-emitting diode performance with a nano-roughened p-GaN Surface," *IEEE Photon. Technol. Lett.*, vol. 17, no. 5, pp. 983–985, May 2005.

[9] K. Bao, X. N. Kang, B. Zhang, T. Dai, C. Xiong, H. Ji, G. Y. Zhang, and Y. Chen, "Improvement of light extraction from patterned polymer encapsulated GaN-based flip-chip light-emitting diodes by imprinting," *IEEE Photon. Technol. Lett.*, vol. 19, no. 22, pp. 1840–1842, Nov. 2007.

[10] H. Y. Ryu, J. K. Kwang, Y. J. Lee, and Y. H. Lee, "Enhancement of light extraction from 2-D photonic crystal slab structures," *IEEE J. Sel. Top. Quantum Electron.*, vol. 8, no. 2, pp. 231–237, Mar.–Apr. 2002.

[11] T. Kim, A. J. Danner, and K. D. Choquette, "Enhancement in external quantum efficiency of blue light-emitting diode by photonic crystal surface grating," *Electron. Lett.*, vol. 41, no. 20, pp. 1138–1139, 2005.

[12] F. F. Ren, M. B. Yu, J. D. Ye, Q. Chen, G. Q. Lo, and D. L. Kwong, "Enhanced vertical light extraction from ultrathin amorphous $Si-Si_3N_4$ multilayers with photonic crystal patterns," *IEEE Photon. Technol. Lett.*, vol. 21, no. 2, pp. 91–93, Jan. 2009.

[13] T. Dai, B. Zhang, Z.-S. Zhang, D. Liu, X. Wang, K. Bao, X.-N. Kang, J. Xu, D.-P. Yu, and X. Zhu, "Surface light extraction mapping from 2-D array of 12-fold photonic quasicrystal on current injected GaN-based LEDs," *Chin. Phys. Lett.*, vol. 24, no. 4, pp. 979–982, 2007.

[14] R. J. Yan and Q. K. Wang, "Enhancement of light extraction efficiency in oled with periodic nano-structure," in *Proc. IEEE Int. Conf. Nano/Micro Eng. Molecular Syst.*, 2006.

[15] J. Q. Xi, H. Luo, A. J. Pasquale, J. K. Kim, and E. F. Schubert, "Enhanced light extraction in GaInN light-emitting diode with pyramid reflector," *IEEE Photon. Technol. Lett.*, vol. 18, no. 22, pp. 2347–2349, Nov. 2006.

[16] S. Trieu, X. M. Jin, B. Zhang, T. Dai, K. Bao, X. N. Kang, and G. Y. Zhang, "Light extraction improvement of GaN-based light emitting diodes using patterned undoped gan bottom reflection gratings," *Proc. SPIE*, vol. 7216, pp. 72162Q–72162Q-8, 2009.

[17] D. A. Steigerwald, J. C. Bhat, D. Collins, R. M. Fletcher, M. O. Holcomb, M. J. Ludowise, P. S. Martin, and S. L. Rudaz, "Illumination with solid state lighting technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 8, no. 2, pp. 310–320, Mar.–Apr. 2002.

[18] K. Bao, X. Kang, B. Zhang, T. Dai, Y. Sun, Q. Fu, G. Lian, G. Xiong, G. Zhang, and Y. Chen, "Improvement of light extraction from GaN-based thin-film light-emitting diodes by patterning undoped GaN using modified laser lift-off," *Appl. Phys. Lett.*, vol. 92, no. 14, p. 141104, 2008.

[19] K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas Propag.*, vol. 14, no. 3, pp. 302–307, 1966.

[20] S. Kim, K. Lee, J. Kim, M. Kwon, and S. Park, "Fabrication of photonic crystal structures on light emitting diodes by nanoimprint lithography," *Nanotechnology*, vol. 18, no. 5, p. 055306, 2007.

[21] S. Adams, J. Payne, and R. Boppana, "Finite difference time domain (FDTD) simulations using graphics processors," in *Proc. HPCMP Users Group Conf.*, 2007, pp. 334–338.

[22] Z. M. Liu, A. S. Mohan, T. A. Aubrey, and W. R. Belcher, "Techniques for implementation of the FDTD method on a CM-5 parallel computer," *IEEE Ant. Prop. Mag.*, vol. 37, no. 5, pp. 64–71, Oct. 1995.

[23] Y. Chen, T. L. Simpson, and T. Q. Ho, "Highly efficient technique for solving radiation and scattering problems," *IEEE Proc.*, vol. 139, no. 1, pp. 7–10, Feb. 1992.

[24] H. Kimura and N. Yoshida, "3-D full-wave analysis with nonlinearity and line characteristics of device by electromagnetic field analysis on time domain," *Elec. Comm. Jpn.*, vol. 75, pp. 89–100, 1991.

[25] X. Jin, B. Zhang, T. Dai, W. Wei, X. N. Kang, G. Y. Zhang, S. Trieu, and F. Wang, "Optimization of top polymer gratings to improve GaN LEDs light transmission," *Chin. Opt. Lett.*, vol. 6, no. 10, pp. 788–790, 2008.

Simeon S. Trieu was born in San Jose, CA, in 1982. He received the B.S. degree in computer engineering from California Polytechnic State University (Cal Poly), San Luis Obispo, in 2006. He is currently working toward both the M.S. degree in electrical engineering and the M.B.A. degree from Cal Poly.

His current research interests include top and bottom diffraction grating optimization for maximum light extraction efficiency of GaN-based LEDs.

Mr. Trieu won the NSF 08-603 OISE—EAPSI 2009 Fellowship Award ("EAPSI: Light Extraction Improvement of GaN-based Light Emitting Diodes"), in 2009.

Xiaomin Jin (M'96–SM'04) was born in Beijing, China. She received the B.E and M.S. degrees from the Electronic Engineering Department, Tsinghua University, Beijing, China, in 1992 and 1996, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, in 2001.

From 2001 to 2003, she was with the research laboratories of several companies, including Corning Lasertron, Inc., Bedford, MA, W. L. Gore and Associates, Newark, DE, and Optical Communication Products, Inc., Fremont, CA. She was an Assistant Professor with Temple University, Philadelphia, PA, from 2003 to 2004, and is currently an Assistant Professor with the Electrical Engineering Department, California Polytechnic State University, San Luis Obispo. She has had over 30 publications in the past ten years on the topic of photonic devices.

Computer Generated Three-dimensional Holography from Two-dimensional Photo

Nicholas Hageman and Xiaomin Jin
Electrical Engineering Department, California Polytechnic State University, San Luis Obispo, CA 93407
xjin@calpoly.edu.

Abstract: We present 3D holography from 2D photos using computer generated hologram (CGH). The photo is segment into foreground/middleground/background. Matlab is used to create the CGH. Both single-laser/dual-laser setups are investigated for the 3D image recovering.

OCIS codes: (090.1760); (090.2870)

1. Introduction

Fourier Transform (FT) holograms have been around for quite some time, and the concept of computer generated hologram (CGH) has been around since holograms first came out. Transparencies have been used before to print out hologram recordings. There also have been research on 3D holograms for sometime now, but those use multiple lasers. In this paper, we present a computer generated 3D hologram that uses only one laser source and compare it with dual laser sources.

Inspired by Thad Walker's paper "Holography without photography" [1], we decided to delve into the world of holograms. Upon arriving in the mist of this expanding phenomena of CGH, we propose a product that would go beyond hologram stickers to creating a sort of pop-out experience for any two dimensional picture. In other words, anybody can access the graphical user interface (GUI) software and give the program the photo they want to turn into a hologram, and then give the necessary inputs and define the necessary region of interests (ROIs). The computer can generate the hologram form those information. The users just need to print out the hologram transparencies and stick them into their 'hologram box.' The users could show their photos in hologram form, with the foreground of the photo popping out, like a pop up book. Furthermore, our proposed 3D CGH can also be used in security cards.

Dr. William J. Dallas from University of Arizona, proposed holograms in Matlab[2], and gave the code for doing a Fast Fourier Transform (FFT) on the image being inputted: y = fftshift(fft2(fftshift(x))), where y is the hologram recording and x is the input photo. We followed this method, which greatly reduced the time it took to produce a hologram, compared to Ref [1]. In the last chapter of the book [3], Asundi gives code for viewing the image plane. When implementing the code, we found that it wasn't necessarily accurate, as it did not take beam divergence and power into account. However, it did give a perspective to what will be seen, and a chance to mend mistakes before printing the hologram recording. The function would have the hologram recording as an input, which would then go through an FT, log scaled, and normalized. In 2001, Lucilla Croce Ferri's presented a paper named "Visualization of 3D information with digital holography using laser printers" [4]. However, their research only covered self-luminous transparent 3D objects. In our research, we are creating different focus planes with user defined segments of their photos instead of dissecting virtual 3D images. Also we don't exactly need to implement at higher dots-per-inch (dpi) resolution from the laser printer compared to Ref.[4]: the higher the dpi, the better your hologram resolution. It's a detail that would come in handy since most photographs have a high amount of 'little details' that are discarded when turning it into a hologram.

2. The Segmentation of Photo

The segmentation of the photo gives the background hologram an illusion of a shadow of the foreground hologram, as long as the segmented photo background is not mainly black. It is the bread and butter of this experiment: the code. We have 5 pieces of code that help in the process of creating the holograms along with a few other commands. In the first, users specify whether to crop or resize the image, which is the first function, "Crop_Resize". The function then takes the input image and grey scales it, and then carries on with what the user wanted: crop or resizing. The crop option brings up a crop square onto the photo that the user can adjust. Double clicking the cropped area when done will show the result of the cropping. The resize option basically squeezes the photo down, or inflates the photo up, to a 400x400

matrix. Next, the image is sent to a spacing function, "Space", which envelops the image in a black background until it is more towards the North-West corner of the matrix. The reason for this is to make sure the real image and the twin image don't overlap each other in each hologram. Then the spaced image goes through segmentation from the "Segment" function. The users define how many layers of segmentation they want: 2 for foreground/background, 3 for foreground /middleground /background. Then the user is able to specify the Region of Interest to dissect the image into the two or three parts by putting points around the ROI and then double clicking it. The segmented images on a black background are outputted via the function. Next comes the hologram recording of the segmented images. Each one is made into a recording by the "Shift_Holo2" function, and the next function, "Im_Print_Size", skews the hologram recordings so that they are 1 inch x I inch. The different inputs that the user should know are the dpi of the printer being used, the image filename, and the number of segmented layers they want (2 or 3). Finally, CGH is printed on the transparency.

3. Reconstruct Three-dimensional Holography

As mentioned before, we started out with Thad Walker's code, and then switched over to the FFT method to make Binary FT holograms. We used a 600 dpi laser printer to print out the hologram recordings. When implementing the FFT code, it's important to put the dpi, or close to the dpi of the printer as the matrix size for the hologram recording. After doing so, the print of the hologram recording should be an inch squared. It will be found that the simpler the photo, the better it will be to make out the hologram. As mentioned before, the 'little details' of higher resolution photos get discarded when making holograms due to changing the photo to black and white and the resolution of the printer.

The setup implemented was putting a double convex lens in front of the laser, and putting the hologram recording behind so that the lens is sandwiched in between the laser and the hologram recording, as shown in Fig. I and 2. Before putting the transparency with the hologram recording behind the lens, the focal point of the lens should be adjusted to be in the depth plane (distance from the lens) where the hologram to be in. In order to do this, the lens is moved closer or farther away from the laser which in turn moves the focal point farther or closer to the lens, respectively. Another thing to keep into account is the beam divergence of the laser. The JDSU 1001P He-Ne laser in the setup gave a beam divergence of I.3mrad. The use of a collimator would be needed to spread the beam out for a laser with less divergence. Even with the 1001P's 1.3mrad spread, the outer edges of the hologram weren't very bright. So a higher spread may give all aspects of the hologram the proper intensity. The next thing was to setup and test a beam splitter. We incorporated a 0.5 neutral density filter, which lets about 32% of the light pass through. There were distortions in the image plane (focus point of the lens) of the beam splitter. We also compared the beam splitter results with that of two lasers, in which one laser was used to simulate a beam splitter. One laser setup is enough to regenerate 3D CGH. However, the two laser setup gives better resolution.

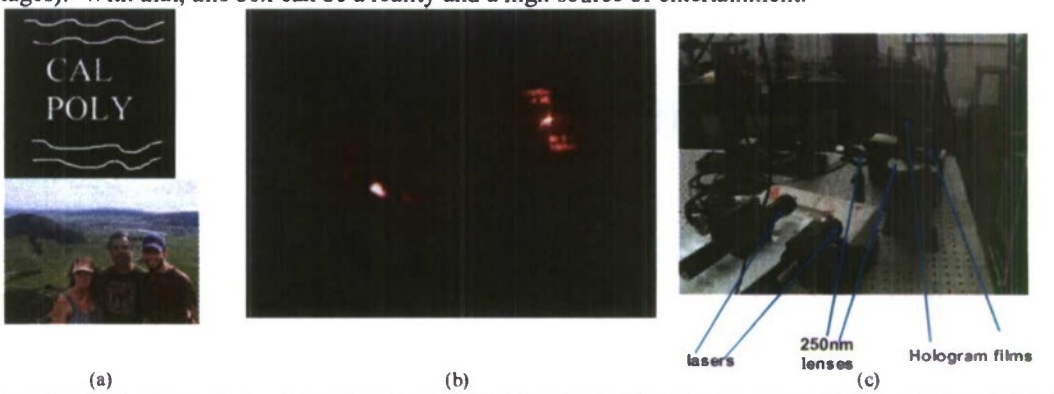
In terms of the placement of the holograms to create an overlaying effect, the background and foreground should be a couple of inches apart in terms of the same image plane, shown in Fig. 3. With this setup of the holograms, the extra background and foreground holograms can be physically blocked. There are two screens: one screen is solid which should show only the background hologram while the other is semi-transparent or non transparent which should show only the foreground hologram. However, the separation mentioned above to get the overlay effect causes the background hologram to also appear on the foreground hologram screen if the particular screen is semi-transparent, and the foreground hologram to appear on the solid background hologram screen. With the above setup, we achieve viewing the foreground hologram in front of the background hologram. Changing the height and angle of the view will give the hologram new perspectives, always with the pop out feel.

4. Conclusion

We use Matlab FFT to produce CGH and achieve 3D holograph image by both single- and double- laser reconstruction setup. The code we developed is much simple compared to in Ref. [2] and [3]. We propose different focus planes with user defined segments of their photos instead of dissecting virtual 3D images. And our setup also overcomes the twin image issue addressed in Ref. [1].

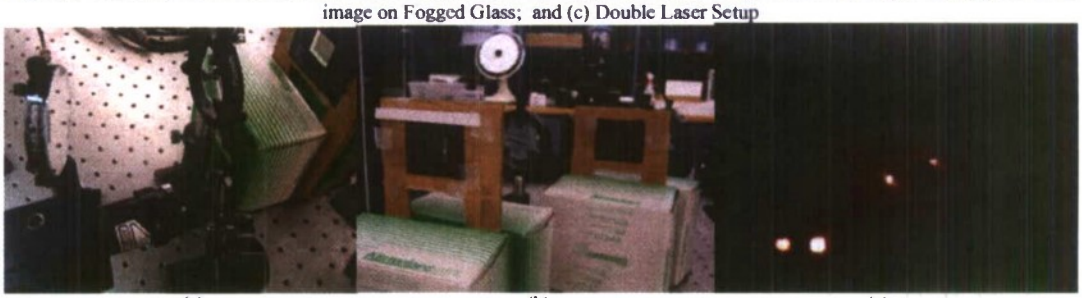
As stated before, every home has a 'hologram box' that displays their photos in a pop out hologram. The user would easily convert their photos into a hologram and print it out and stick them into their 'hologram box.' The hologram box would have the laser, the lenses, the beam splitter, and the plugged in

hologram recordings mainly out of view. There would be a display area with multiple screens that would reflect the holograms. Light filters can be strategically placed to filter out any of the unwanted images (extra background, middle ground, foreground holograms reflecting off an undesired screen, or the twin images). With that, this box can be a reality and a high source of entertainment.



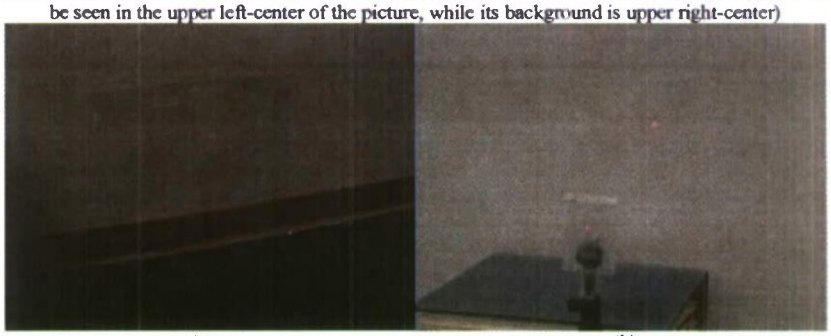
(a) (b) (c)

Fig.1. (a) Two photo we did converted to CGH: Original Cal Poly Image and Bishop Image; (b) Double Laser - Bishops 600x600 image on Fogged Glass; and (c) Double Laser Setup



(a) (b) (c)

Fig. 2 Beam Splitter Setup (a) Beam Splitter (b) CGH film placement (c) Cal Poly 900x900 image on Fogged Glass (the Poly can be seen in the upper left-center of the picture, while its background is upper right-center)



(a) (b)

Fig.3. The screens are about 136 inch away from the CGH films. We place multiple screens: (a) Fogged Glass Screen (b) Tapped Box Screen and after them is the big white board.

5. Acknowledgement

This project is supported by Department of the Navy, Office of Naval Research, under Award # ONR 7-N000140811209 in 2009, USA;

6. Reference

[1] Thad G. Walker, "Holography without photography," American Journal of Physics **67**, 783-785 (1999).

[2] http://www.dallas.radiology.arizona.edu/OpSci627.htm

[3] A. Asundi , MATLAB @ for Photomechanics- A Primer, Elsevier Science; 1 edition (December 1, 2002), ISBN-13: 978-0080440507, Chap. 5.

[4] Lucilla Croce Ferri, "Visualization of 3D information with digital holography using laser printers," Computers & Graphics **25**, 309-321 (2001)

High-Speed Concatenation of Frequency Ramps Using Sampled Grating Distributed Bragg Reflector Laser Diode Sources for OCT Resolution Enhancement

Brandon George and Dennis Derickson

California Polytechnic State University, Electrical Engineering Department
1 Grand Avenue, San Luis Obispo, CA 93407 USA ddericks@calpoly.edu

ABSTRACT

Wavelength tunable sampled grating distributed Bragg reflector (SG-DBR) lasers used for telecommunications applications have previously demonstrated the ability for linear frequency ramps covering the entire tuning range of the laser at 100 kHz repetition rates[1]. An individual SG-DBR laser has a typical tuning range of 50 nm. The InGaAs/InP material system often used with SG-DBR lasers allows for design variations that cover the 1250 to 1650 nm wavelength range. This paper addresses the possibility of concatenating the outputs of tunable SGDBR lasers covering adjacent wavelength ranges for enhancing the resolution of OCT measurements. This laser concatenation method is demonstrated by combining the 1525 nm to 1575 nm wavelength range of a "C Band" SG-DBR laser with the 1570nm to 1620 nm wavelength coverage of an "L-Band" SG-DBR laser. Measurements show that SGDBR lasers can be concatenated with a transition switching time of less than 50 ns with undesired leakage signals attenuated by 50 dB.
Keywords: Distributed Bragg reflector diode lasers, optical coherence tomography.

1. INTRODUCTION

The single-chip SG-DBR wavelength-tunable laser has been shown[1] to be capable of a 100 kHz repetition rate continuous wavelength ramp or a 10 μs single-shot sweep time. The wavelength ramp consists of fifty 1-nm wide continuous tuning segments that are combined to produce the full 1525 to 1575 nm range. The spectral output of these tunable sources provides single longitudinal mode outputs with a typical spectral linewidth of 100 MHz. The existing high volume production of the SG-DBR laser in the telecommunication industry and small source size gives the opportunity for cost-effective and portable OCT applications. In many OCT applications, it is beneficial to have a wider source tuning range for enhanced distance resolution in OCT. The InGaAs/InP semiconductor material system used by telecommunications SG-DBR lasers offers design variations that can cover any wavelength in the 1250 to 1650 nm range. This paper shows that adjacent wavelength coverage SGDBR tunable lasers can be easily concatenated due to the unique architecture of the laser. The Semiconductor Optical Amplifier (SOA) that is included as the output segment of this single chip laser design can be used as a fast optical switch to aid the concatenation of adjacent wavelength tuning segments. It will be shown that the concatenation of adjacent wavelength sweep segments can be accomplished with a dead-time of less than 50 ns using the SOA that is monolithically integrated on the laser chip. In order to demonstrate concatenation of SGDBR lasers into an extended frequency ramp versus time a "C-Band" SGDBR laser and an "L-Band" SGDBR laser were characterized. Each single chip SGDBR laser has 5 control segments as illustrated in figure 1.

Figure 1. This is a top view of a single-chip Sampled Grating Distributed Bragg Reflector Laser (SGDBR) used for OCT measurements. Currents in Back Mirror (A), Front Mirror (D), and Phase Section (C), control the laser frequency. Currents in the gain section (B) and the semiconductor optical amplifier (E) control the laser power. The chip dimensions are 0.5 mm by 3 mm.

Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XIV, edited by Joseph A. Izatt, James G. Fujimoto, Valery V. Tuchin, Proc. of SPIE Vol. 7554, 75542O · © 2010 SPIE
CCC code: 1605-7422/10/$18 · doi: 10.1117/12.842604

In order to construct a frequency ramp using SGDBR lasers, the laser wavelength as a function of drive current to the front mirror, back mirror and phase sections needs to be obtained. Figure 2 shows the results of wavelength mapping as a function of drive current to the front and back mirror segments of an L-Band SGDBR laser. Figure 3 shows the results of wavelength mapping as a function of drive current to the front and back mirror segments of a C-Band SGDBR laser.

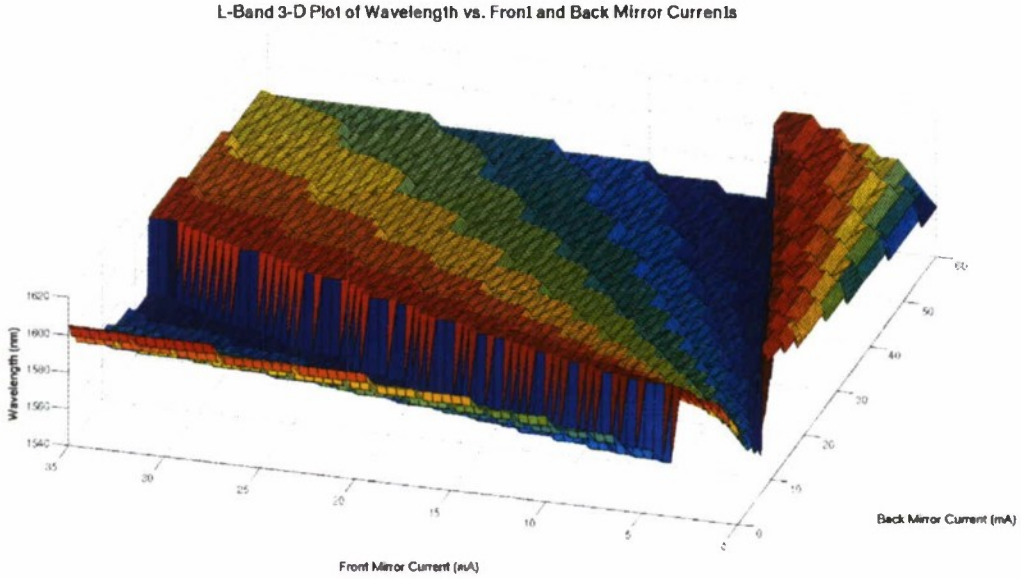


Figure 2. Three-dimensional (mode map) plot of wavelength versus front mirror and back mirror currents for the L-band laser. The approximate wavelength coverage is from 1565 nm to 1620 nm.

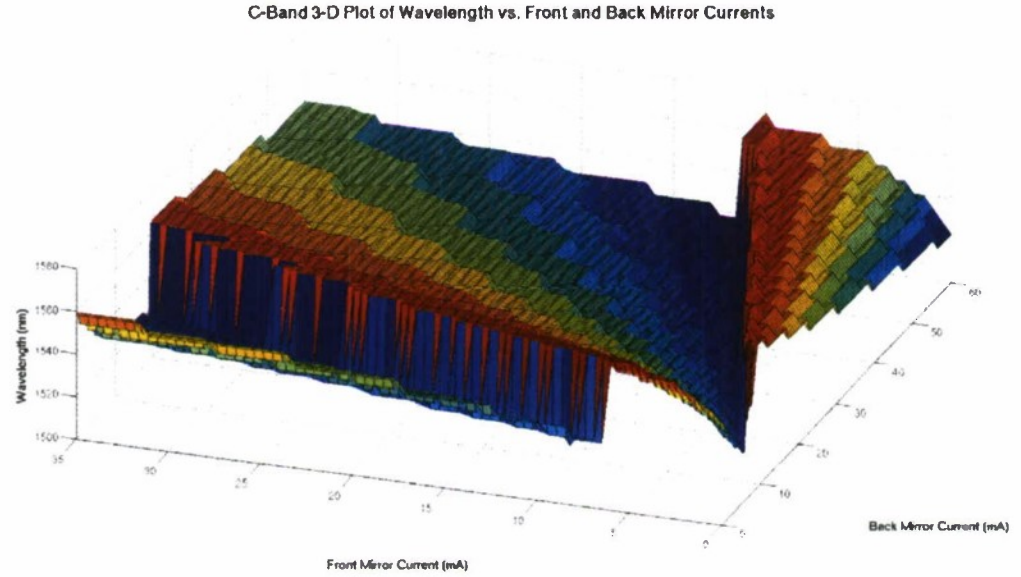


Figure 3. Three-dimensional (mode map) plot of wavelength versus front mirror and back mirror currents for the L-band laser. The approximate wavelength coverage is from 1525 nm to 1570 nm.

In order to create a linear frequency ramp versus time, one needs to map a sequence for front and back mirror current values that allow the laser to tune over its entire frequency range in a linear ramp versus time. Figure 4 shows the

current paths that were chosen in order to tune over the range of the L-Band laser. Figure 5 shows the current paths that were chosen in order to tune over the range of the C-Band laser. In the construction of a linear frequency versus time ramp from the concatenated C and L band lasers it was decided to sweep from lowest frequency to highest frequency. This means that paths 1 through 8 are first transited in figure 4. Paths 1-8 of figure 5 are then concatenated to form the entire frequency ramp.

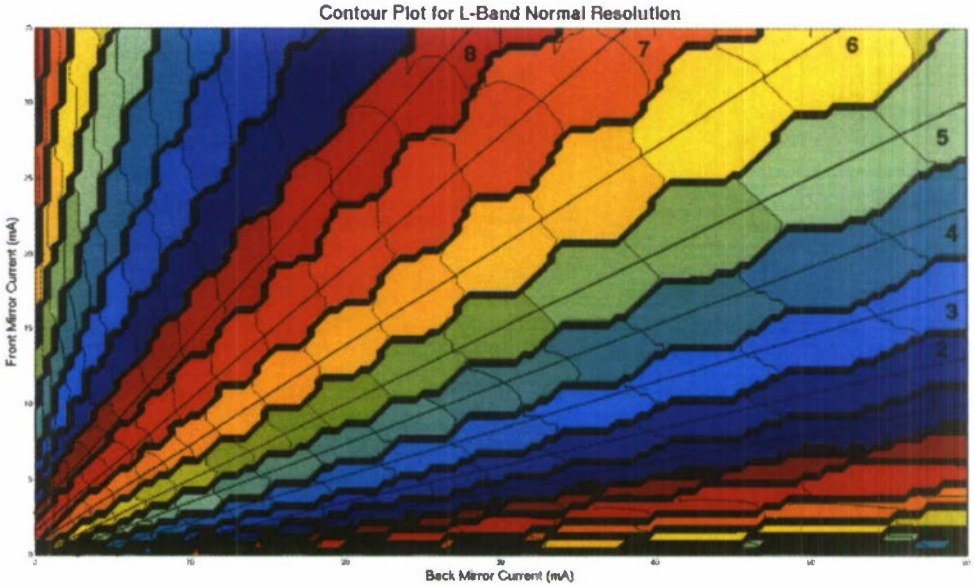


Figure 4. This plot contains similar information to that of figure 2 for the L-Band laser from a top view. Lines 1 through 8 illustrate the path of the current versus time in order to accomplish a linear frequency ramp versus time.

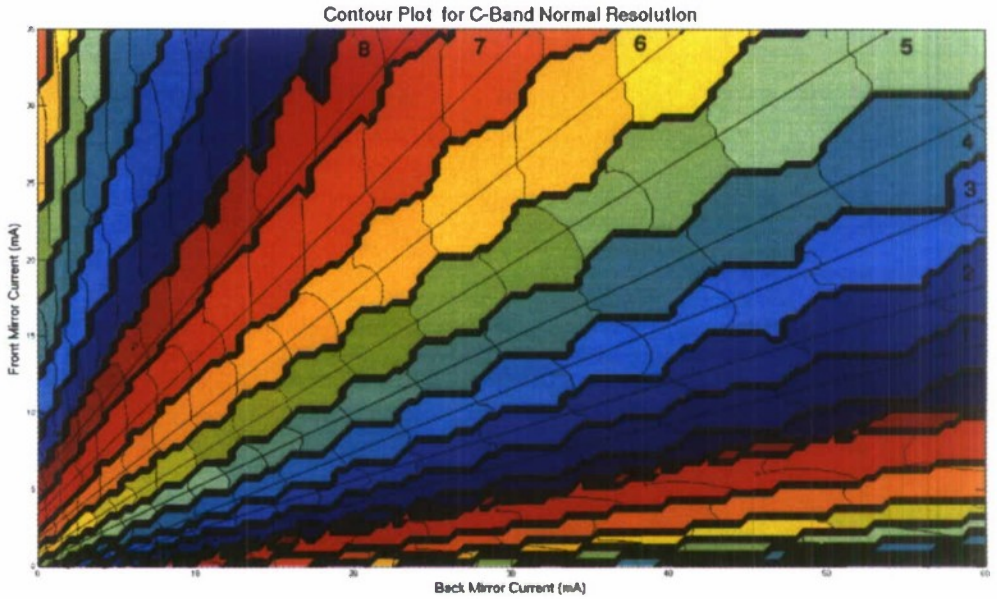


Figure 5. This plot contains similar information to that of figure 3 for the C-Band laser from a top view. Lines 1 through 8 illustrate the path of the current versus time in order to accomplish a linear frequency ramp versus time.

2. DESCRIPTION OF THE CONCATENATION PROCESS

Figure 6 shows the system block diagram used to generate separate wavelength sweeps from two SGDBR lasers and the method for sweep concatenation.

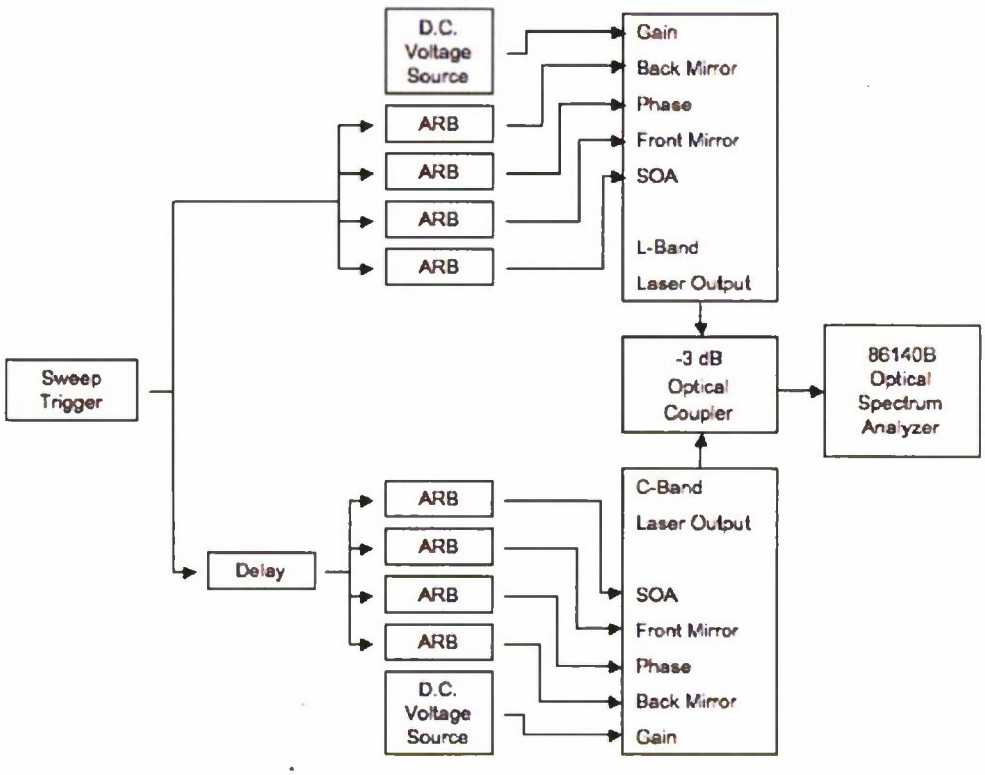


Figure 6. This is the block diagram for generating a wide spectral-width linear frequency versus time ramp by concatenating two SG-DBR lasers.

The top laser block is an "L-Band" SGDBR laser that covers the nominal wavelength range of 1565 to 1620 nm. The bottom laser block is a "C-Band" SGDBR laser that covers the nominal wavelength range of 1525 to 1575 nm. Each monolithic SG-DBR laser chip (see figure 1) has three segments that control the laser wavelength; the Back Mirror (BM), Front Mirror (FM), and Phase (ϕ). Each of the laser segments is driven by an Arbitrary Waveform Generator (ARB) to control the current versus time into each segment of the laser. The ARBs have a clock frequency of 50 MHz with an effective 20 MHz of analog bandwidth. The gain segment is DC biased and provides overall gain for the laser. The Semiconductor Optical Amplifier (SOA) boosts the output power from the laser. The SOA will also be used as a fast optical switch for concatenation of the outputs of the L-Band and C-Band SG-DBR devices. The SOA can also be used to level the power versus time at the output of the laser during the sweep. Each of the ARBs are time synchronized by a master sweep trigger signal. A programmable delay is used to control the switching point in time between the L-Band and C-Band laser operation in order to concatenate the sweep of the two lasers. Finally the outputs of the two lasers are combined in a wavelength independent coupler with the output analyzed by measurement instrumentation. The wavelength coverage of the laser is first monitored in the optical spectrum analyzer. The linearity of the frequency versus time from the laser can be monitored by replacing the optical spectrum analyzer with a repetitive pass band optical filter, a photodetector and an oscilloscope.

Figure 7 illustrates the current versus time that is provided for each segment of the two SGDBR sources. The wavelength tuning map for each SGDBR laser must first be mapped as a function of current in the FM, BM and ϕ segments[2]. In order to provide a linear ramp in frequency from the laser, the saw-tooth waveforms labeled L-FM, L-BM, and L-ϕ must be synthesized from the previously generated wavelength-tuning maps shown in figures 2-5. Each vertical abrupt change in the segment current versus time represents an abrupt change in current that causes the laser to jump between different longitudinal modes but with no associated wavelength shift. The individual output from each SG-DBR laser has approximately 50 continuous tuned segments that are joined together for the full wavelength sweep.

Figure 7 also shows the C-FM, C-BM and C-ϕ waveforms for the C-Band wavelength sweep. In this case the output of the C-Band laser is delayed in time so that the sweep starts after the finish of the L-Band sweep.

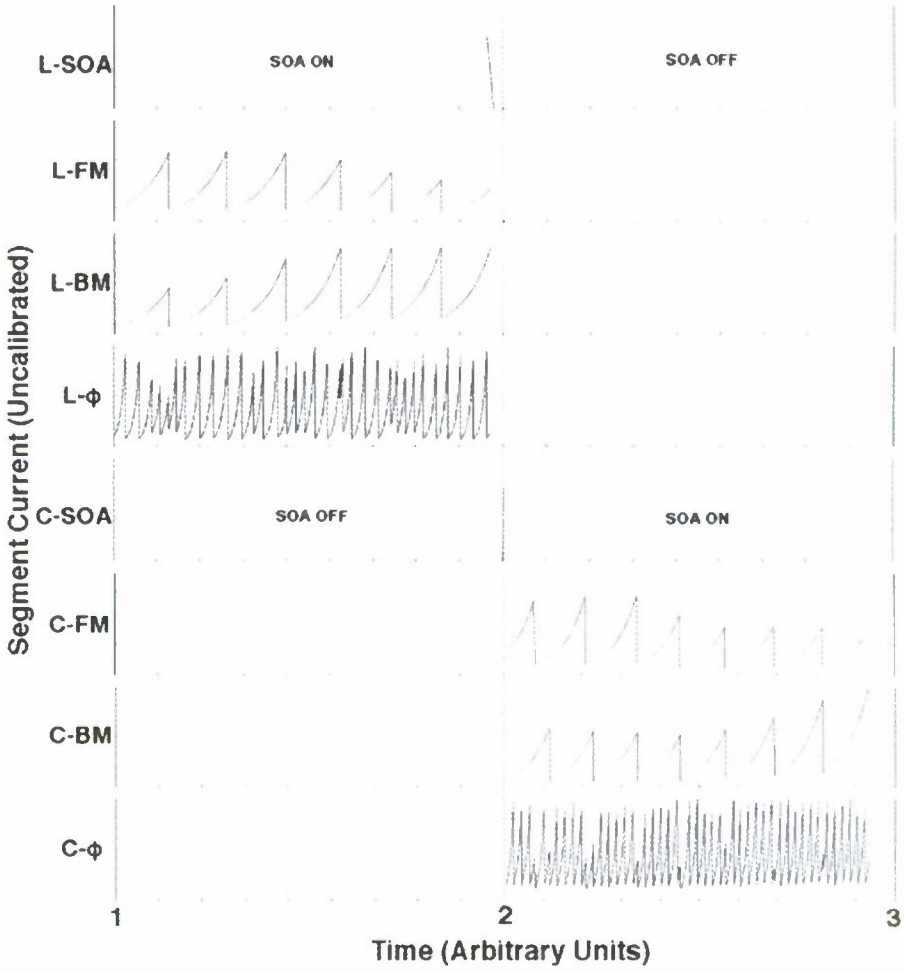


Figure 7. The current versus time into the SG-DBR laser segments for a concatenated C-Band and L-Band wavelength sweep is shown.

At time = 1 the master sweep trigger initiates the sweep. The L-SOA signal is on allowing a high output power from the L-Band Laser. The C-SOA is off resulting in a large attenuation of the C-Band laser output. The sawtooth current waveforms found in L-FM, L-BM, and L-ϕ segments then create a frequency ramp in the L-Band laser output.

At time =2, the L-SOA is tuned off and the C-SOA is turned on. The sawtooth C-FM, C-BM and C-ϕ current waveforms then create a frequency ramp in the C-Band laser. Point 1 represents the trigger signal for the L-Band laser ARBs, while Point 2 represents the trigger signal for the C-band laser ARBs. For single shot operation, the sweep will end at time=3. For repetitive sweeps, the operation will reset back to time=1. Sweep-back time of the laser is very fast and is presently limited by the 50ns rise time of the ARB drive waveform.

The C-Band and L-Band wavelength sweeps associated with the current waveforms shown in figure 7 are shown in Figure 8.

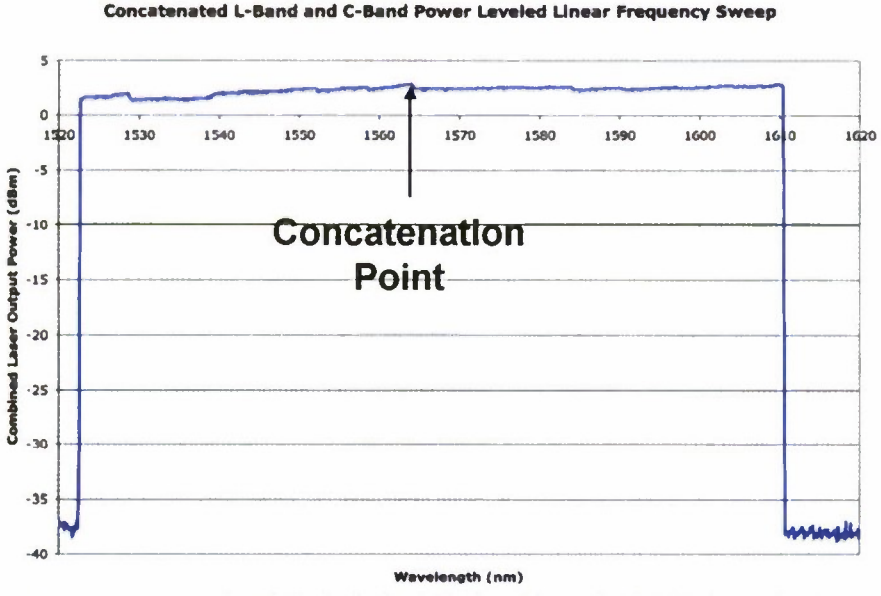


Figure 8. The output power versus wavelength for both the C-Band and L-Band SG-DBR lasers for the current drive waveforms shown in figure 7.

The L-Band wavelength coverage was from 1560.702 nm to 1610.092 nm or 49.4 nm of bandwidth, The C-band laser tuning shape consisted of eight wavelength-tuning paths from 1523.040 nm to 1569.970 nm or 46.9 nm of bandwidth. However, since path 8 of the C-band laser completely overlaps with the L-band laser sweep, it was omitted from the final tuning dictionary. What remained was a constant C-band wavelength sweep from 1523.064 nm to 1564.166 nm or 41.1 nm of bandwidth. A wavelength concatenation point of 1564 nm was used for this experiment.

3. CHARACTERIZING THE L-BAND/C-BAND CONCATENATION POINT

Each SG-DBR laser is configured to construct linear wavelength or frequency sweeps with approximately fifty 1 nm wide modal segments that are stitched together. By concatenating two wavelength sweeps together, we are introducing a new concatenation stitching point between the two lasers. The major difference between the modal stitching points in the laser and the concatenation stitching point is the control segment of the laser. Modal switching points use the FM, BM and ϕ segments of the laser. Concatenation stitching points use the SOA segments of the laser. When the SOA segment current is on, high output power is delivered. When the SOA segment current is off, the laser output is attenuated. Figure 9 illustrates the concatenation stitching process. For the first part of the sweep the L-SOA is on and the C-SOA is off. In the second part of the sweep the C-SOA is on and the L-SOA is off. The switching characteristics of the SOA are key to the successful concatenation of wavelength sweeps. The next two figures will explore the switching speed and the attenuation levels that are possible with the SOA switch that is part of the monolithic structure of the SG-DBR laser.

Figure 9 shows a measurement of the optical power versus time that is measured both at the L-Band Output, the C-Band output and the combined output from the laser. In this case a high bandwidth photodetector and an oscilloscope replace the OSA measurement block found in Figure 6.

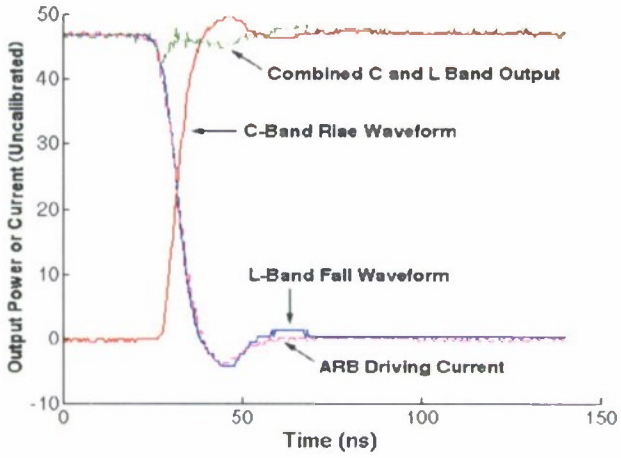


Figure 9. The output power versus time is shown for various outputs in the block diagram of figure 6. The C-Band, L-Band and combined optical output power versus time are shown. In addition the electrical drive current (ARB Driving Current) waveform to the SOA segment is also shown.

The L-Band optical output power goes from being on to off in a period of less than 50 ns. The C-Band output power goes from being off to on in a period of less than 50 ns. The measurement of the drive waveform to the SOA segment (labeled ARB driving current) illustrates that the rise and fall times are primarily driven by the rise and fall time drive capabilities of the ARB circuitry in this example. Figure 9 also shows the combined optical power from the wavelength independent combiner. A summary from this measurement is that very fast switching times can be accomplished between L-Band and C-Band operations using the SOA segment. The SOA segment is part of the monolithic laser structure so no extra part count (and associated cost) is needed for this switching other than the wavelength independent optical coupler and an additional ARB for each laser.

The SOA switch is not ideal though and there is remnant signal leakage from an SG-DBR laser when the SOA is in its off state. Figure 10 shows a measurement of the C-Band laser output power versus wavelength for the SOA in its on state and in its off state.

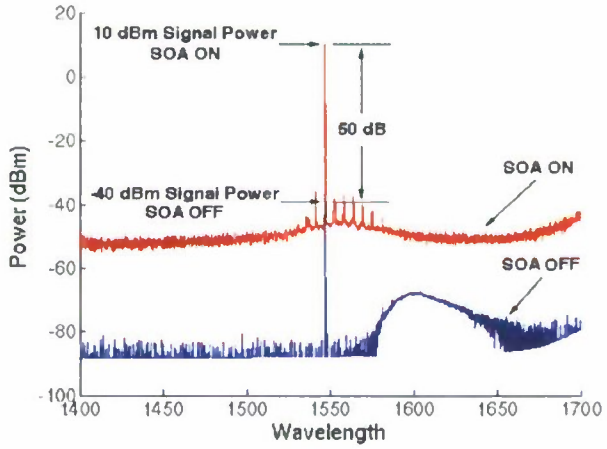


Figure 10. The output of the C-Band laser is shown as a function of wavelength both in the on state and the off state. The resolution bandwidth of the OSA is 0.07nm in this measurement.

When the SOA segment is in its "on state", it delivers 10 dBm (10 mW) of signal power. The integrated "on state" amplified spontaneous emission (ASE) power of the laser from 1525 to 1575 nm is estimated as -13.5 dBm (45 μW). The "off state" signal power is -40 dBm (0.1μW). The "off state" ASE power is estimated as -41.8 dBm (0.067 μW). The "off state" ASE power at wavelengths longer than 1570 nm has redueed attenuation because the photon energy at these wavelengths is less than the band gap energy of the material in the SOA segment.

The eritieal point to be taken from the data of figure 5 is the amount of undesired power that leaks into the output with the SOA off eompared to the desired signal. In this ease the "on state" would have a signal power of 10 dBm. The undesired leakage signal would be dominated by the -40 dBm source signal from the attenuated laser. So the undesired lasing signal will be 50 dB lower that of the desired laser output. The attenuation of the SOA is suffieient so that it will have a negligible effect on the eombined output shown in figure. 1.

4. SUMMARY

The eoncatenation of two wavelength tunable SG-DBR wavelength sweeps was illustrated. The SOA segment of the SG-DBR laser is key to eost-effective frequeney sweep eoncatenation. The switehing time for sweep eoncatenation was found to be less than 50 ns, limited by the ARB rise time eapability. The attenuation of the SOA segment is sufficient so that only a 50 dB attenuated signal leaks into the combined output. The result shows that SG-DBR lasers ean sueeessfully eoncatenated offering better time resolution for OCT measurements. It should be possible to eoneatenate a large number of adjaeent SG-DBR frequeney sweeps to eover the 1250 nm to 1650 nm wavelength range that is possible with the SG-DBR laser using the InGaAsP/InP material system. The high ramp repetition rate, low eost and small size of the eoncatenated source makes this an attractive option for low cost and portable OCT applications with improved distance resolution.

REFERENCES

[1] Shane O'Connor, Michael A. Bernacil, Andrew DeKelaita, Ben Maher, and Dennis Derickson, "100 kHz Axial Scan Rate Swept-Wavelength OCT using Sampled Grating Distributed Bragg Reflector Lasers," in Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine XIII, edited by Joseph A. Izatt, James G. Fujimoto, Valery V. Tuchin, Proceedings of SPIE Vol. 7168 (2009).

[2] Shane O'Connor, Michael A. Bernacil, and Dennis Derickson, "Generation of High Speed, Linear Wavelength Sweeps Using Sampled Grating Distributed Bragg Reflector Lasers," IEEE LEOS Annual Meeting, Newport Beach, CA, paper TuB2 (2008).

Approximate seismic displacement capacity of piles in marine oil terminals

Rakesh K. Goel*

Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo, CA 93407, USA

(Received November 10, Accepted March 4, 2010)

Aabtract. This paper proposes an approximate procedure to estimate seismic displacement capacity – defined as yield displacement times the displacement ductility – of piles in marine oil terminals. It is shown that the displacement ductility of piles is relatively insensitive to most of the pile parameters within ranges typically applicable to most piles in marine oil terminals. Based on parametric studies, lower bound values of the displacement ductility of two types of piles commonly used in marine oil terminals – reinforced-concrete and hollow-steel – with either pin connection or full-moment-connection to the deck for two seismic design levels – Level 1 or Level 2 – and for two locations of the hinging in the pile – near the deck or below the ground – are proposed. The lower bound values of the displacement ductility are determined such that the material strain limits specified in the Marine Oil Terminal Engineering and Maintenance Standard (MOTEMS) are satisfied at each design level. The simplified procedure presented in this paper is intended to be used for preliminary design of piles or as a check on the results from the detailed nonlinear static pushover analysis procedure, with material strain control, specified in the MOTEMS.

Keywords: marine structures; seismic displacement capacity; seismic ductility; seismic analysis; seismic design, piles.

1. Introduction

Seismic design of marine oil terminals in California is governed by the Marine Oil Terminal Engineering and Maintenance Standard (MOTEMS) (Eskijian 2007, MOTEMS 2007). The MOTEMS requires design of such facilities for two earthquake levels: Level 1 and Level 2. The return period of the design earthquake for each level depends on the risk level. For example, Level 1 and Level 2 design earthquakes for high risk terminals correspond to return periods of 72 and 475 years, respectively. The acceptance criteria for piles in the MOTEMS are specified in terms of maximum permissible material strains. The maximum permissible material strains depend on the earthquake level – Level 1 or Level 2. For reinforced-concrete piles, the material strain limits also depend on location of the plastic hinge – pile-deck or in-ground. The material strain limits in two of the commonly used piles – reinforced-concrete and hollow-steel – are summarized in Table 1. These strain values are based on the latest revisions (after public comment period) to the MOTEMS

* Professor, E-mail: rgoel@calpoly.edu

Table 1 Material strain limits in the MOTEMS

Pile Type	Material	Hinge Location	Level 1	Level 2
Reinforced-Concrete	Concrete	Pile-Deck	$\varepsilon_c \le 0.004$	$\varepsilon_c \le 0.025$
		In-Ground	$\varepsilon_c \le 0.004$	$\varepsilon_c \le 0.008$
	Steel	Pile-Deck	$\varepsilon_s \le 0.01$	$\varepsilon_s \le 0.05$
		In-Ground	$\varepsilon_s \le 0.01$	$\varepsilon_s \le 0.025$
Hollow-Steel	Steel		$\varepsilon_s \le 0.008$	$\varepsilon_s \le 0.025$

and have been modified from the previous version as follows: (1) concrete compression strain limit for Level 1 have been reduced from 0.005 to 0.004 for both pile-deck and in-ground hinge formation; and (2) steel tensile strain limit for Level 2 has been increased from 0.01 to 0.025 for in-ground hinge formation.

Since the acceptance criteria in the MOTEMS is specified in terms of maximum permissible strains, estimation of seismic displacement capacity of piles in marine oil terminals requires monitoring material strains during the nonlinear pushover analysis. However, most commercially available structural analysis programs do not have the capability to directly monitor strains during such analysis. Therefore, there is a need to develop simplified procedure to estimate seismic displacement capacity of piles in marine oil terminals that ensure that material strains do not exceed the values specified in the MOTEMS yet do not require direct monitoring of strains during seismic analysis.

In order to fill this important need, this investigation is aimed at developing a simplified procedure for estimation of displacement capacity using displacement ductility, instead of strain limits, for piles in marine oil terminals. The simplified procedure is developed for piles which are connected to the deck either by pin-connection or full-moment-connection and assumed to be fixed at the bottom at a distance equal to depth-to-fixity below the mud line; initial results of this study were reported in a recent publication (Goel 2008). Simplified procedures for piles with other types of pile-deck connection (such as dowel connection for reinforced concrete piles and concrete-plug connection for steel piles) are available in Goel (2010). The simplifying assumption of fixing the pile at a distance equal to depth-of-fixity below the mud line, and thus avoiding explicit modeling of soil, is appropriate for estimating the displacement capacity because such assumption is commonly used in estimating is seismic displacement demand for marine oil terminals; next section discuses details of this simplification.

2. Simplifying assumption

Fig. 1(b) shows typically used mathematical model of a free-head pile of Fig. 1(a) supported on bedrock (or other competent soil) and surrounded by soil between the bedrock and mud line. In this model, the pile is represented by beam-column elements and soil by Winkler reaction springs connected to the pile between the bedrock and the mud line (Finn 2005). The properties of the beam-column element are established based on the pile cross section whereas properties of the reaction springs are specified based on geotechnical data (e.g., see Priestley *et al.* 1996, Dowrick 1987, Finn 2005). Fig. 1(c) shows the height-wise distribution of bending moment under lateral load

applied to the pile tip. Note that the maximum bending moment occurs slightly below the mud line at a depth equal to D_m, typically denoted as the depth-to-maximum-moment below the mud line Fig. 1(c). Lateral displacement at the pile top can be calculated based on this bending moment distribution or from a discrete element model implemented in most commonly available computer programs for structural analysis.

An alternative approach to modeling the detailed pile-soil system (Fig. 1(b)) is the effective fixity approach (Priestley *et al.* 1996, Sec. 4.4.2; Dowrick 1987, Sec. 6.4.5.3). In this approach, the pile is assumed to be fixed at a depth below the mud line equal to the depth-to-fixity, D_f, and no soil reaction springs are included (Fig. 1(d)). The depth-to-fixity is defined as *the depth that produces in a fixed-base column with soil removed above the fixed base the same top-of-the-pile lateral displacement under the lateral load, F,* as that in the actual pile-soil system (Priestley *et al.* 1996). Both the axial load, P, and top-of-the-pile moment, M (not shown in Fig. 1(d)) need to be considered. The depth-to-fixity, which depends on the pile diameter and soil properties, is typically provided by the geotechnical engineer, estimated from charts available in standard textbooks on the subject (e.g., Priestley *et al.* 1996, Dowrick 1987), or from recommendations in several recent references (e.g., Chai 2002, Chai and Hutchinson 2002).

The equivalent fixity model is typically used for estimating displacement of piles that remain within the linear-elastic range. For piles that are expected to be deformed beyond the linear-elastic range, however, nonlinear analysis of the discrete soil spring model approach of Fig. 1(b) is preferable (Priestley *et al.* 1996, Sec. 4.4.2) because the plastic hinge forms at the location of the maximum bending moment, i.e., at the depth-to-maximum-moment, D_m, and not at the depth-to-fixity, D_f. A recent investigation has developed equations for estimating lateral displacement of equivalent fixity model of the nonlinear soil-pile system by recognizing that the plastic hinge forms at the depth-to-maximum-moment (Chai 2002); expressions for estimating displacement ductility of pile-soil system are also available (Priestley *et al.* 1996, Sec. 5.3.1). However, calculation of lateral displacement capacity of nonlinear soil-pile systems using these approaches requires significant information about the soil properties. Another paper developed relationship between curvature ductility and displacement ductility of fixed-head concrete piles (Song *et al.* 2005) but did not

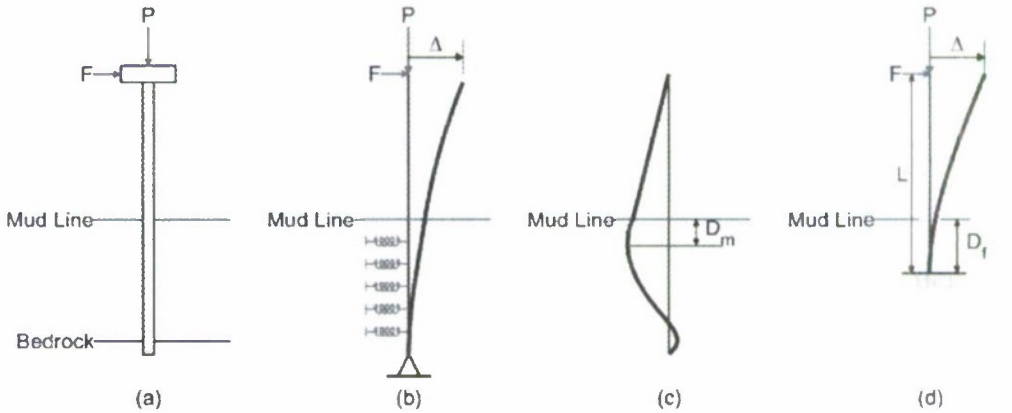


Fig. 1 Simplified model of the pile-soil system for displacement capacity evaluation (a) Pile supported on bedrock, (b) Mathematical model of the pile, (c) Height-wise variation of bending moment, (d) Equivalent fixity model for displacement calculation

include flexibility of piles above the mud line as is the case for piles in marine terminals.

This investigation uses a simplifying assumption that the equivalent fixity model may directly be used to estimate lateral displacement capacity of nonlinear piles. Clearly, such an approach implies that the plastic hinge would form at the depth-to-fixity, D_f, which differs from the actual location at the depth-to-maximum-moment, D_m. It is useful to note that D_f is typically in the range of 3 to 5 pile diameter whereas D_m is in the range of 1 to 2 pile diameter (see Priestley *et al.* 1996). Obviously, the displacement would be lower if the cantilever base was located at D_m below the mud line compared to if it was located at D_f in the equivalent fixity model: plastic displacement is given by $\Delta_p = \theta_p(L_a + D_f$ or $D_m)$ where θ_p is the plastic hinge rotation and L_a is the cantilever length. However, it must be noted that additional deformation would occur in the actual pile due to curvature below the depth-to-maximum-moment (see Fig. 1(c)). Fixing the cantilever at D_f below the mud line compensates for these additional deformations.

Seismic analysis of large marine oil terminals using analytical models that include detailed modeling of soil effects on each pile, such as those in Fig. 1(b), becomes cumbersome. Therefore, analyses to estimate seismic displacement demands of piles often utilize the equivalent fixity approximation. The depth-to-fixity below the mud line is generally recommended by geotechnical engineer on the project for typical piles and the local soil conditions and linear/nonlinear soil behavior.

Since the seismic displacement demand is often estimated from analysis of a system that utilizes equivalent fixity approximation for piles, the same approximation must be used for estimating the seismic displacement capacity for consistency. Therefore, it is appropriate to utilize the assumption of equivalent fixity for estimating seismic displacement capacity of piles as proposed in this investigation.

It is useful to emphasize that the approximate approach proposed in this investigation is intended to be used for preliminary design of piles or as a check on the results from the detailed nonlinear analysis. The final evaluation of piles may be conducted by analysis of detailed pile-soil system using more advanced methods (e.g., Castelli and Maugeri 2009, Imancli *et al.* 2009, El Naggar *et al.* 2005, Allotey and el Naggar 2008, Chiou and Chen 2007).

3. Current procedure in the motems

The displacement capacity of piles in the MOTEMS is estimated from nonlinear static pushover analysis. In this analysis, a force of increasing magnitude is applied statically in the transverse direction (perpendicular to the pile) permitting the materials in the pile – steel and/or concrete – to deform beyond their linear-elastic range. The displacement capacity is defined as the maximum displacement that can occur at the tip of the pile without material strains exceeding the permissible values corresponding to the desired design level.

The displacement capacity of a pile at a selected design level in the MOTEMS is obtained from the procedure proposed by Priestley *et al.* (1996) as illustrated in Fig. 2. This procedure requires development of the pile section moment-curvature relationship. The moment-curvature relationship may be developed from any standard moment-curvature analysis programs using material constitutive relationship specified in the MOTEMS; the MOTEMS specifies guidelines for selecting material properties such as concrete and steel strengths as well as stress-strain curves for unconfined concrete, confined concrete, reinforcing steel, and prestressing steel. It is useful to note that the

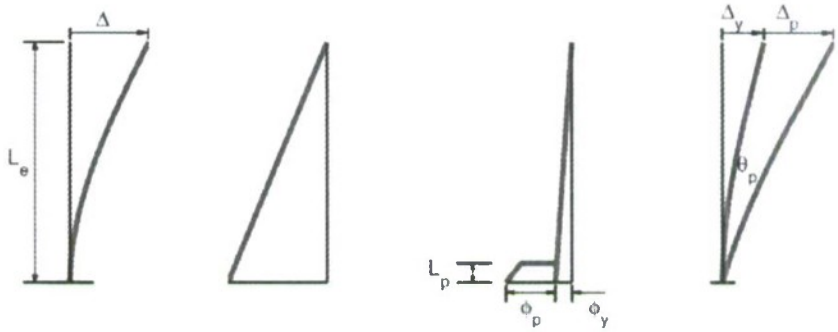


Fig. 2 Deformation capacity of a pile (a) Deflected shape, (b) Bending moment diagram, (c) Curvature distribution, (d) Yield and plastic displacements

formulation presented next is for a cantilever, i.e., a pile with a pin-connection to the deck. Similar formulation is available for piles with full-moment-connection (e.g., see CALTRANS 2006).

The total displacement capacity of the pile is computed as

$$\Delta = \Delta_y + \Delta_p \tag{1}$$

in which Δ_y is the yield displacement and Δ_p is the plastic displacement of the pile. The yield displacement can be estimated as

$$\Delta_y = \frac{\phi_y L_e^2}{3} \tag{2}$$

where ϕ_y is the yield curvature computed from

$$\phi_y = \frac{M_y}{EI_e} \tag{3}$$

with M_y being the yield moment and EI_e being the slope of the initial elastic portion of the bilinear idealization of the moment-curvature relationship and L_e is the pile "effective" length defined as the length between a critical section and the point of contra-flexure (or point of zero bending moment). The "effective" length for a cantilever becomes equal to its total length.

The plastic displacement is computed from

$$\Delta_p = \left(\frac{M_u}{M_y} - 1\right)\Delta_y + L_p(\phi_u - \phi_y)(L_e - 0.5L_p) \tag{4}$$

which includes components due to the elastic displacement resulting from the increase in moment from M_y to M_u, i.e., post-yield stiffness of the moment-curvature relationship, and due to plastic rotation θ_p of the pile. In order to compute the plastic rotation, it is assumed that a constant plastic curvature, $\phi_p = \phi_u - \phi_y$, occurs over a plastic hinge length L_p of the pile (see Fig. 2(c)). Therefore, the plastic rotation is given by

$$\theta_p = L_p\phi_p = L_p(\phi_u - \phi_y) \tag{5}$$

The values of M_u and ϕ_u in Eq. (4) are largest values of pile section moment and curvature, respectively, without exceeded the material strains at the selected design level.

The MOTEMS specify the formula for estimating the plastic hinge length required in Eq. (4). If the hinge were to form against a supporting member, i.e., at the pile-deck intersection, the plastic hinge length is computed from

$$L_p = \begin{cases} \rho L_e + 0.022 f_{ye} d_{bl} \geq 0.044 f_{ye} d_{bl} & (f_{ye} \text{ in MPA}) \\ \rho L_e + 0.15 f_{ye} d_{bl} \geq 0.3 f_{ye} d_{bl} & (f_{ye} \text{ in ksi}) \end{cases} \tag{6}$$

in which L_e is the "effective" length defined as the distance from the critical section of the plastic hinge to the point of contra-flexure, $\rho = 0.08$ is the constant that specifies plastic hinge length as a fraction of the length L_e, f_{ye} is the expected yield strength of the reinforcing steel, and d_{bl} is the diameter of the longitudinal reinforcement. The second term in Eq. (6) is intended to account for strain-penetration effects in reinforcing steel in reinforced concrete members. If the plastic hinge forms in-ground, the MOTEMS provides a chart to estimate the plastic hinge length that depends on the pile diameter, subgrade modulus, effective stiffness of the pile, and the distance from ground to the pile point of contra-flexure above the ground. It is useful to note that Eq. (6), as specified in Priestley *et al.* (1996) or in the MOTEMS (2006), does not explicitly impose an upper limit even though there may be some experimental evidence that the plastic hinge length should not be greater than the pile diameter.

The plastic hinge length formula of Eq. (6) specified in the MOTEMS is based on the recommendation by Priestley *et al.* (1996) for reinforced concrete sections. The MOTEMS do not provide recommendations for plastic hinge length for hollow steel piles.

4. Approximate procedure

An approximate procedure is proposed in this investigation to compute the displacement capacity of piles commonly used in marine oil terminals. This approach computes the displacement capacity as

$$\Delta_c = \mu_\Delta \Delta_y \tag{7}$$

where Δ_y is the yield displacement of the pile and μ_Δ is the displacement ductility of the pile. The yield displacement may be estimated from

$$\Delta_y = \begin{cases} \dfrac{M_y L_t^2}{6EI_e} & \text{for piles with full-moment-connection to the deck} \\ \dfrac{M_y L_t^2}{3EI_e} & \text{for piles with pin-connection to the deck} \end{cases} \tag{8}$$

in which M_y is the effective section yield moment, EI_e is the initial slope of the idealized bilinear moment-curvature relationship, and L_t is the total length of the equivalent fixity pile model. The displacement ductility is selected such that the material strains remain within the limits specified in

the MOTEMS. The guidelines to select the displacement ductility are developed next for reinforced-concrete piles and hollow-steel piles with either pin-connection or full-moment-connection to the deck by parametric studies on nonlinear finite element models of piles fixed at the depth equal to depth-to-equivalent-fixity below the mud line.

5. Analytical model

As noted previously, this investigation utilizes the equivalent fixity model (Fig. 1(d)) to estimate displacement capacity the pile-soil system (Fig. 1(a)) instead of the more elaborate model containing entire pile length with Winkler reaction springs below the mud line (Fig. 1(b)). Therefore, an analytical model of the cantilever (with free or fixed head) was developed in the structural analysis software Open System for Earthquakes Engineering Simulation (*OpenSees*) (McKenna and Fenves 2001). The length of the pile, L_t, was selected to be equal to the free-standing height plus the depth below the mud-line equal to the depth-to-fixity. The pile was modeled with the *nonlinearBeamColumn* element in *OpenSees*. The *nonlinearBeamColumn* element uses a force-based, distributed-plasticity approach with integration of section behavior over the member length. The pile was modeled with five elements, each with seven integration points. The section is defined with fibers of confined concrete, unconfined concrete, and steel reinforcing bars for reinforced-concrete piles and steel for hollow-steel piles. The nonlinear axial-flexural behavior of the element is determined by integration of the nonlinear stress-strain relationships of various fibers across the section, whereas linear behavior is assumed for shear and torsional.

The compressive stress-strain behavior of concrete, both confined and confined, was modeled with *Concrete01* material in *OpenSees* (Fig. 3(a)). The crushing strain of the unconfined concrete was selected to be equal to 0.004 and that for confined concrete was selected to be that corresponding to the rupture of confining steel using the well established Mander model (see Priestley *et al.* 1996). The stress-strain behavior of steel was modeled with *ReinforcingSteel* material in *OpenSees* (Fig. 3(b)). Further details of the material models and *nonlinearBeamColumn* element are available in McKenna and Fenves (2001).

6. Results of parametric study

This section presents results of parametric studies to investigate the sensitivity of the displacement

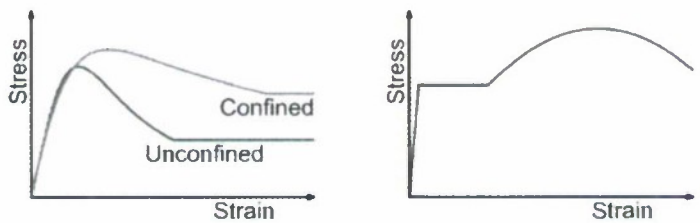


Fig. 3 Material models

ductility on various pile parameters such as pile length and diameter, longitudinal reinforcement ratio, transverse reinforcement ratio, and axial force level for reinforced-concrete piles; and pile length and diameter, wall thickness, and axial force level for hollow-steel piles. Following is a summary of this study.

6.1 Reinforced-Concrete piles

Fig. 4 presents variation of displacement ductility with pile length, L_t, for three values of pile diameters: 61 cm, 91 cm, and 107 cm. The results are presented for piles with axial force equal to 5% of its capacity, 1% longitudinal reinforcement, and 0.6% transverse reinforcement. These parameters correspond to typically used piles in marine oil terminals. The presented results indicate that the displacement ductility of the pile is essentially independent of the pile length as apparent from very little variation in the ductility with length as well as pile diameter as apparent from almost identical curves for the three pile diameters considered.

In order to understand the aforementioned trend, i.e., independence of the displacement ductility of the pile length and diameter, it is useful to re-examine the equation that defines the displacement ductility. For this purpose, let us utilize Eqs. (1), (4) and (8) to express the displacement ductility as

$$\mu_\Delta = \left(\frac{M_u}{M_y}\right) + 3(\mu_\phi - 1)\left(\frac{L_p}{L_e}\right)\left(1 - 0.5\frac{L_p}{L_e}\right) \tag{9}$$

in which μ_ϕ is the section curvature ductility at a selected design level. For piles with little or no post-yield straining hardening in section moment-curvature relationship, $M_u \simeq M_y$, and Eq. (9) further simplifies to

$$\mu_\Delta = 1 + 3(\mu_\phi - 1)\left(\frac{L_p}{L_e}\right)\left(1 - 0.5\frac{L_p}{L_e}\right) \tag{10}$$

Eq. (10) indicates that the displacement ductility depends on the plastic hinge ratio, L_p/L_e, and section curvature ductility, μ_ϕ. For long piles typically used in marine oil terminals, the ratio L_p/L_e

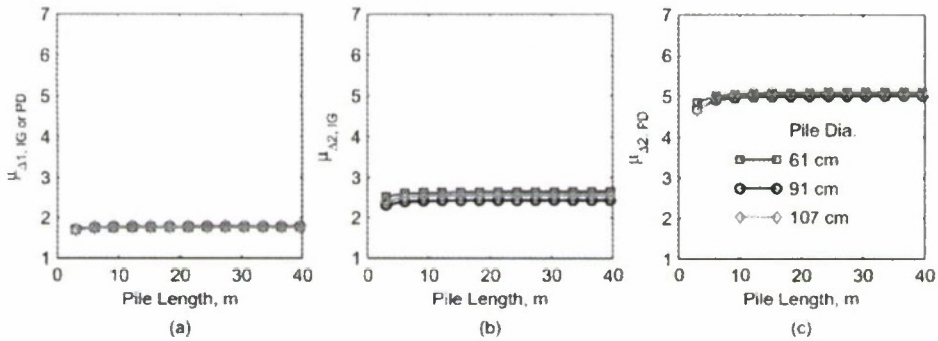


Fig. 4 Variation of displacement ductility with pile length and pile diameter of reinforced-concrete piles for seismic design (a) Level 1 for in-ground (IG) and pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, (c) Level 2 for PD hinge formation

may become essentially independent of the length as the second term in Eq. (6) becomes much smaller than the first term. This indicates that the displacement ductility is essentially independent of the pile length which is consistent with the observation from results presented in Fig. 4.

The displacement ductility does not directly depend on the pile diameter but on the section curvature ductility, μ_ϕ. The results presented in Fig. 5 indicate that the section curvature ductility is essentially independent of the pile diameter. Therefore, Eq. (10) leads to the conclusion that the displacement ductility is also essentially independent of the pile diameter, which is consistent with the finding in Fig. 4 that the pile displacement ductility is independent of the pile diameter.

Fig. 6 present the variation of the displacement ductility with the reinforcement ratio. The results presented are for a pile with 91 cm diameter and 15 m length. The values of longitudinal reinforcement varying between 0.5% and 2% and transverse reinforcement between 0.5% and 1.5% were considered; these ranges of longitudinal and transverse reinforcement correspond to those for typical piles in marine oil terminals.

The results presented in Fig. 6(a) show that the displacement ductility decreases with increasing

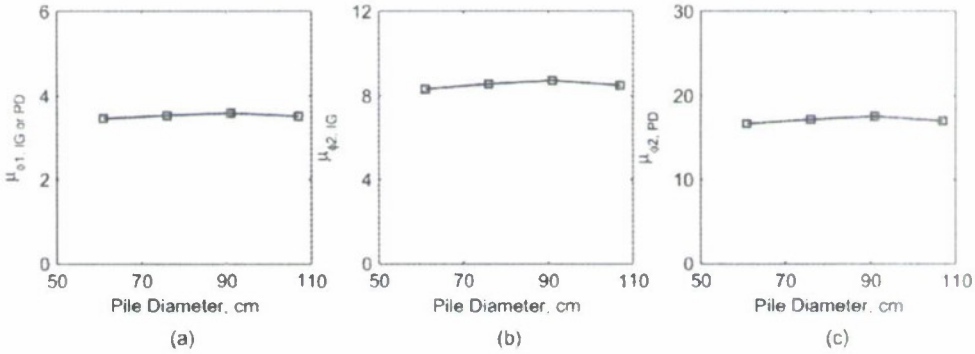


Fig. 5 Variation of section curvature ductility of reinforced-concrete piles with pile diameter for seismic design (a) Level 1 for in-ground (IG) and pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, (c) Level 2 for PD hinge formation

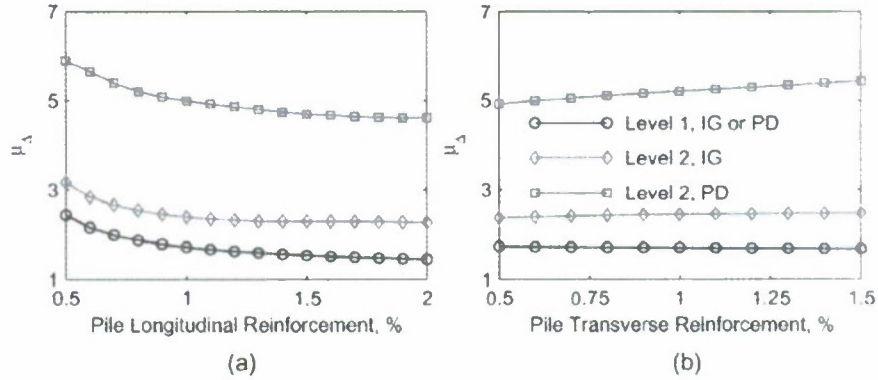


Fig. 6 Variation of displacement ductility of reinforced-concrete piles with pile reinforcement ratio (a) Longitudinal reinforcement, (b) Transverse reinforcement

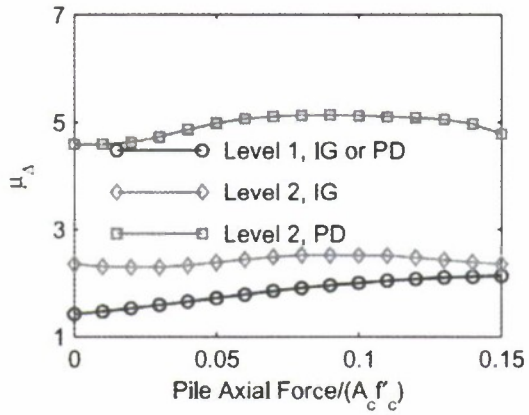


Fig. 7 Variation of displacement ductility of reinforced-concrete piles with pile axial load ratio

longitudinal reinforcement ratio for values up to about 1%. For longitudinal reinforcement ratio in excess of about 1%, as may be the case for seismic piles in marine oil terminals, the displacement ductility of piles is much less sensitive to the value of the longitudinal reinforcement ratio. For such values, the displacement ductility may be considered to be essentially independent of the longitudinal reinforcement ratio. The results presented in Fig. 6(b) show that displacement ductility of piles does not depend significantly on the transverse reinforcement ratio. This becomes apparent from essentially flat variation of the displacement ductility with pile transverse reinforcement ratio.

Fig. 7 presents variation of displacement ductility varies with axial force in the pile. The presented results are for a pile with 91 cm diameter and 15 m length for values of axial force level varying from zero to 15% of the pile axial capacity. These results show that the displacement ductility for Level 1 corresponding to in-ground or pile-deck hinge formation tends to increase with increasing pile axial force. However, the ductility for Level 2 corresponding to in-ground or pile-deck hinge formation appears to be insensitive to the axial force values.

6.2 Hollow-Steel piles

Fig. 8 presents variation of displacement ductility with pile length for three values of pile diameters – 61 cm, 91 cm, and 107 cm – which are representative of typical piles used in marine oil terminals and pile wall thickness of 1.27 cm. These results indicate that the displacement ductility of the hollow-steel pile is also essentially independent of the pile length and diameter as apparent from no variation in the ductility with pile length and almost identical curves for the three pile diameters considered.

As mentioned previously for reinforced concrete piles, the displacement ductility of hollow-steel piles is independent of the pile diameter because the section curvature ductility is essentially independent of the pile diameter. This is confirmed by the results presented in Fig. 9 for variation of section curvature ductility with pile diameter.

In order to understand the effects of the pile wall thickness on the displacement ductility, variations of the displacement ductility with pile length for three values of pile thickness are presented in Fig. 10. The results presented are for a pile with 91 cm diameter and axial force equal

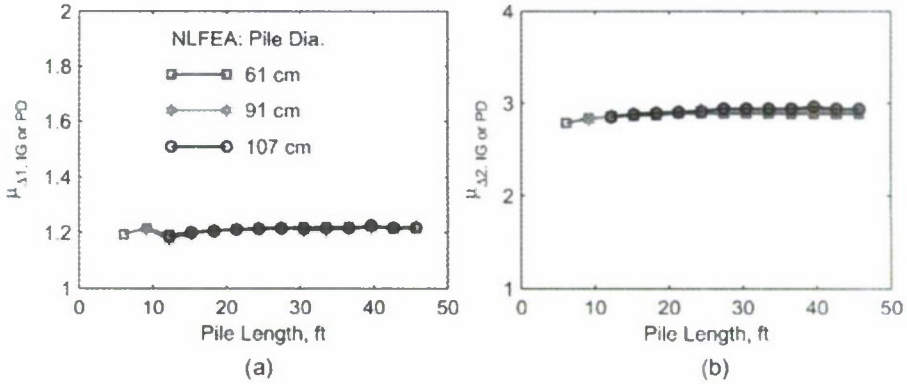


Fig. 8 Variation of displacement ductility of hollow-steel piles with pile length and pile diameter (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG or PD hinge formation

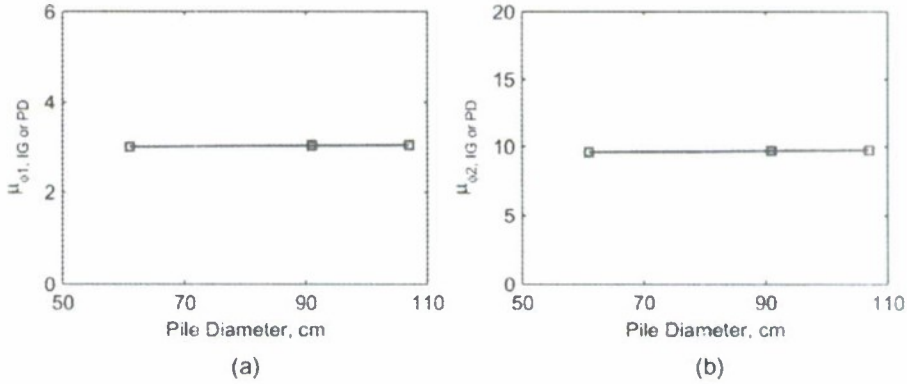


Fig. 9 Variation of section curvature ductility of hollow-steel piles with pile diameter (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG or PD hinge formation

to $0.05Af_y$. These results show that the displacement ductility is essentially independent of the pile wall thickness as indicated by essentially identical curves for the three values of pile wall thickness.

Fig. 11 presents the variation of displacement ductility with axial force in the hollow-steel pile. The presented results are for a pile with 91 cm diameter and 15 m length for values of axial force varying from zero to 15% of the pile axial load capacity. These results show that the displacement ductility for Level 1 is essentially independent of the pile axial load. For Level 2, while the displacement ductility may depend on the axial load for very-low axial loads, it becomes essentially independent of the axial load for more realistic values, i.e. However, the ductility for Level 2 appears to be insensitive to the axial force values, i.e., axial loads greater than $0.05Af_y$.

7. Lower bound of displacement ductility of piles

The results presented so far indicate that the displacement ductility of piles (reinforced-concrete or

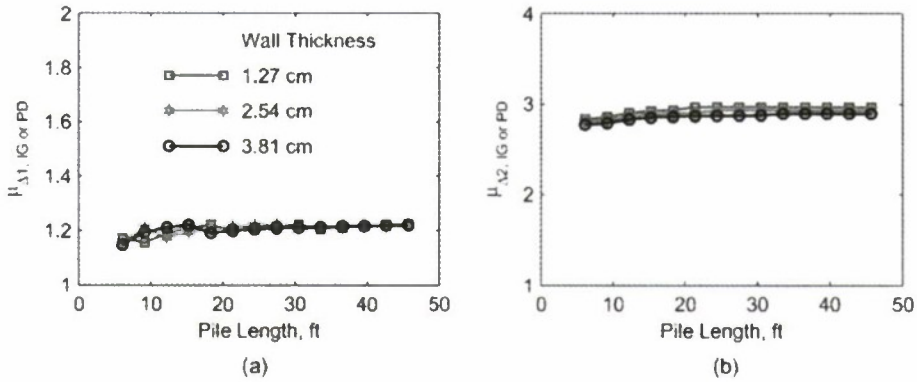


Fig. 10 Variation of displacement ductility of hollow-steel piles with pile length for three values of pile wall thickness (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG or PD hinge formation

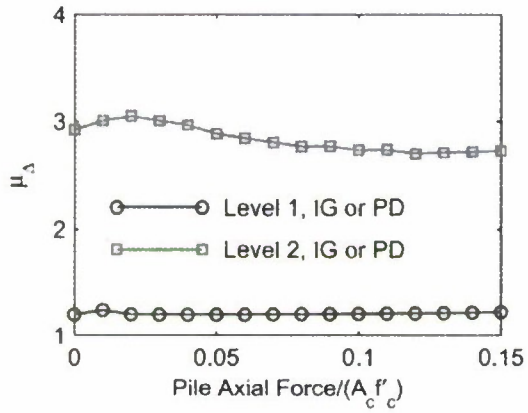


Fig. 11 Variation of displacement ductility of hollow-steel piles with pile axial load ratio

hollow-steel) is relatively insensitive to the various pile parameters within the range these parameters typically used in marine oil terminals. Therefore, the displacement ductility can be used to approximately estimate the pile displacement capacity instead of the material strain limits which are currently specified in the MOTEMS. The results presented in the preceding section indicate that a lower bound of the member displacement ductility may be estimated without any knowledge about various pile parameters.

The results presented in Fig. 12 for pile-deck hinge in the reinforced-concrete pile indicate that lower bound values of the displacement ductility equal to 1.75 for seismic design Level 1 for formation of in-ground or pile-deck hinge, 2.5 for seismic design Level 2 for formation of in-ground hinge, and 5.0 for seismic design Level 2 for formation of pile-deck hinge are appropriate. Note that the displacement ductility for Level 1 is likely to be slightly lower for axial force values lower than the $0.05A_g f_c'$ value considered in developing the results. Similarly, the displacement ductility is likely to be slightly larger for longitudinal reinforcement lower than the 1% value

considered in developing the results. The results presented in Fig. 13 for hollow-steel pile indicate that lower bound values of the displacement ductility equal to 1.2 for seismic design Level 1 and equal to 2.75 for seismic design Level 2 are appropriate. Since strain limits for both in-ground and pile-deck hinge formation are identical for the hollow-steel piles, the aforementioned lower bound values apply to formation of hinges at both locations.

Based on the results presented in Figs. 12 and 13, the recommended lower bound values of the displacement ductility of piles are as follows: 1.75 for seismic design Level 1 (in-ground or pile-deck hinge formation), 2.75 for seismic design Level 2 (in-ground hinge formation), and 5.0 for seismic design Level 2 (pile-deck hinge formation) for reinforced-concrete piles; and 1.2 for seismic design Level 1 and equal to 2.75 for seismic design Level 2 for hollow-steel piles regardless of location of the plastic hinge.

The recommendations on displacement ductility of piles in this section are based on analytical studies alone. It is highly desirable that such recommendations be verified by experiments on pile-soil systems (e.g., Budek *et al.* 2004, Roeder *et al.* 2005).

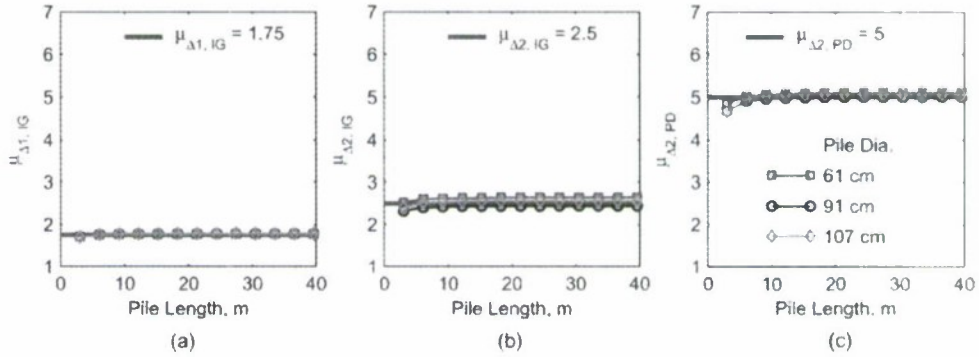


Fig. 12 Lower-bound value of displacement ductility of reinforced-concrete piles for seismic design (a) Level 1 for in-ground (IG) and pile-deck (PD) hinge formation, (b) Level 2 for IG hinge formation, (c) Level 2 for PD hinge formation

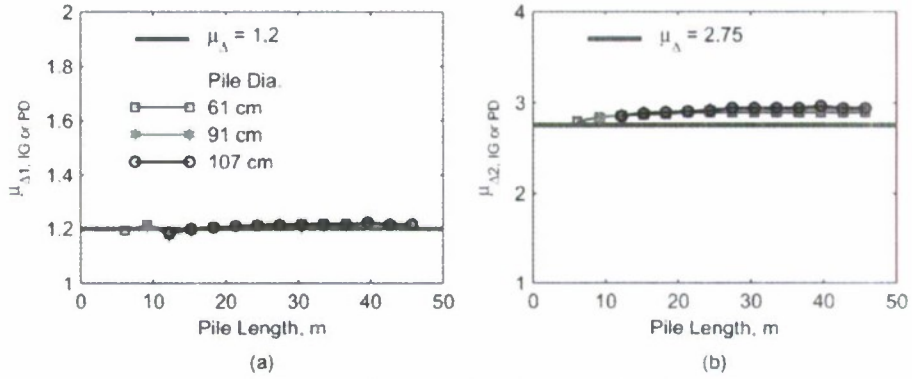


Fig. 13 Lower-bound value of displacement ductility of hollow-steel piles for seismic design (a) Level 1 for in-ground (IG) or pile-deck (PD) hinge formation, (b) Level 2 for IG or PD hinge formation

8. Illustrative example

An illustrative example is presented next to demonstrate that the simplified procedure presented in this paper provides estimate of displacement capacity of the pile shaft that is "accurate" enough for most practical purposes when compared to results from nonlinear pushover analysis. The example structure considered here is adopted from Priestley *et al.* (1996, Sec. 4.6). This structure consists of a circular pile shaft (or columns) of 1.83 m in diameter. The height of the column is 10 m above ground and that of the pile shaft is 15 m below ground with a deck depth of 2 m (Fig. 14(a)). The pile shaft (or column) reinforcement consists of 30 D 36 mm longitudinal bars and D19 mm spirals with a pitch of 115 mm. The material properties are f_y = 414 MPa yield for the reinforcement and f'_c =24.1 MPa nominal concrete strength. The soil is characterized as very dense sand with linearly varying subgrade reaction coefficient of $k_s z$ = 10z MPa/m. The column carries an axial load of 8 MN at its tip.

The mathematical model of the column-pile-soil system is shown in Fig. 14(b). The soil below the mud line is modeled with ten discrete springs. The spring values were taken from the example in Priestley *et al.* (1996) and are shown in Fig. 14(b). As mentioned previously, this model was implemented in OpenSees with various column and pile elements modeled with *nonlinear Beam Column* element and Winkler reaction springs modeled with *zeroLength* element. The fiber section properties of the column and pile were based on the specified section dimensions, reinforcement details, and concrete and steel strengths in Priestley *et al.* (1996).

A moment curvature analysis of the pile shaft section with an axial load of 8.85 MN (including axial force and column weight) and idealized bilinear approximation, shown in Fig. 15(a), results in an effective initial stiffness of

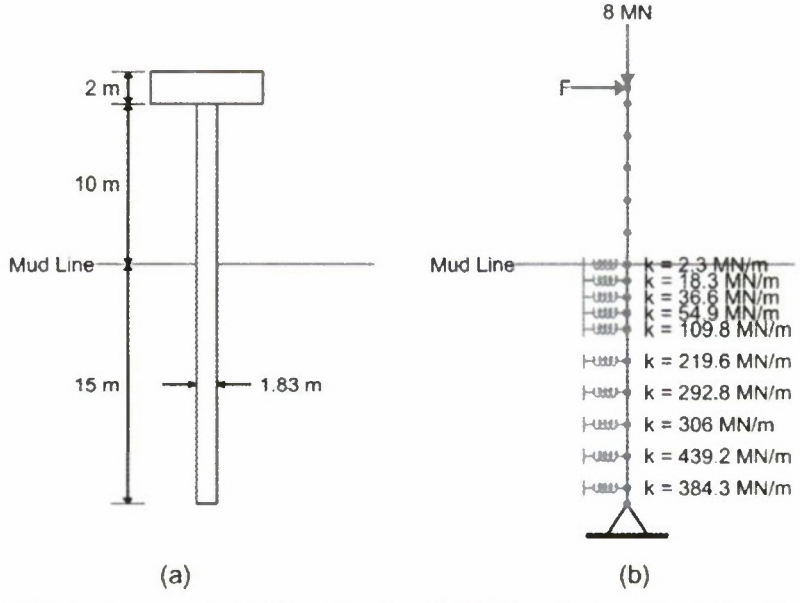


Fig. 14 Illustrative example (a) Pile-soil system, (b) Mathematical model of pile-soil system

$$EI_e = \frac{13473.5}{2.3588 \times 10^{-3}} = 5.71 \times 10^6 \text{ kN-m}^2 \tag{11}$$

Note that the values of M_y, ϕ_y and EI_e in Fig. 15(a) differ slightly from the results presented in Priestley *et al.* (1996) because of slight differences between the pile-section model and material models between this investigation and Priestley *et al.* (1996).

Based on the soil properties, the depth-to-fixity of the column-pile-soil system below the mud line is equal to three times the pile diameter (Priestley *et al.* 1996). Therefore, the total length of the equivalent fixity model is

$$\begin{aligned} L_e &= \text{Free-Standing Length} + \text{Half the Deck Depth} + 3 \times \text{Pile Diameter} \\ &= 10 + 1 + 3 \times 1.83 = 16.5 \text{ m} \end{aligned} \tag{12}$$

Using Eq. 8(b) for cantilever structure, the yield displacement of the equivalent fixity model is

$$\Delta_y = \frac{M_y L_e^2}{3EI_e} = \frac{(13473.5)(16.5)^2}{(3)(5.71 \times 10^6)} = 0.214 \text{ m} \tag{13}$$

As recommended in this investigation, the displacement ductility of the pile for Level 1 and Level 2 seismic design are 1.5 and 5, respectively, for reinforced-concrete piles. Let us assume that the ductility of 1.5 would control if the hinge were to form in-ground (as would be the case for the selected structure) and 5 would control if the hinge were to from above ground. The estimated displacement capacity of the pile using the simplified procedure presented in this investigation for Level 1 and Level 2 are

$$\Delta_{c1} = (1.5)(0.214) = 0.321 \text{ m} \tag{14}$$

$$\Delta_{c2} = (5)(0.214) = 1.07 \text{ m} \tag{15}$$

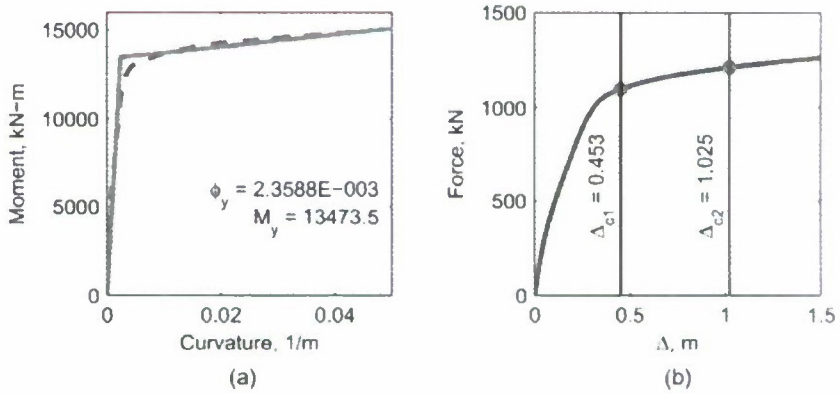


Fig. 15 (a) Moment-curvature relationship of the pile shaft section, (b) Pushover curve of the pile-soil system

A pushover analysis of the nonlinear finite element model of the column-pile-soil system model (Fig. 14(b)) led to the pushover curve of Fig. 15(b). During the pushover analysis, material strains were monitored, and displacement capacity at the selected design level were defined as the maximum displacement at the pile top without exceeding the material strains at the selected design level. The displacement capacities at Level 1 and Level 2 for this analysis are 0.453 m and 1.025 m, respectively (Fig. 15(b)). Let us denote the displacement capacities for the pushover analysis as "exact" values.

The displacement capacities computed form the simplified analysis at Level 2 match the "exact" value quite well: the value from simplified analysis if 1.07 m and "exact" value is 1.025 m. The value at Level 1 of 0.321 m from the simplified analysis is slightly less that the "exact" value of 0.453 m. Obviously, the error in the Level 1 displacement capacity is much larger than that for Level 2 displacement capacity; the slightly larger error for Level 1 occurs due to errors associated with idealizing the moment-curvature relationship with a bilinear curve. However, the simplified procedure provides a conservative estimate of the displacement capacity at Level 1 which may be "accurate" enough for preliminary design or for a quick check on the results from a detailed analysis.

9. Conclusions

This investigation proposes an approximate procedure to estimate seismic displacement capacity – defined as a product of the yield displacement and displacement ductility – of two types of piles – reinforced-concrete and hollow-steel – with either pin-connection or full-moment-connection used in the marine oil terminals. Development of this approximate procedure utilized equivalent fixity approximation for piles. A parametric study is conducted to demonstrate that the displacement ductility is relatively insensitive to various pile parameters such as pile length, pile diameter, longitudinal and transverse reinforcement, and pile axial load level for reinforced-concrete piles; and pile length, pile diameter, wall thickness, and pile axial load level for hollow-steel piles. This observation applies to ranges of these parameters that are applicable for piles typically used in marine oil terminals.

Subsequently, it is demonstrated that lower bound values of the displacement ductility of piles commonly used in marine oil terminals depend on the seismic design level – Level 1 or Level 2 – and location of the hinging in the pile – near the deck or below the ground. Based on the results of the parametric study, lower bound values of the displacement ductility are determined such that the material strain limits specified in the MOTEMS are satisfied at each design level. The recommended lower bound values of the displacement ductility of piles in marine oil terminals are: 1.75 for seismic design Level 1 (in-ground or pile-deck hinge formation), 2.75 for seismic design Level 2 (in-ground hinge formation), and 5.0 for seismic design Level 2 (pile-deck hinge formation) for reinforced-concrete piles; and 1.2 for seismic design Level 1 and equal to 2.75 for seismic design Level 2 for hollow-steel piles regardless of location of the plastic hinge. The applicability of the approximate procedure is finally demonstrated through an illustrative example.

The approximate procedure presented in this paper is intended to be used for preliminary design of piles or as a check on the results from the detailed nonlinear static pushover analysis procedure, with material strain control, specified in the MOTEMS. It is also useful to emphasize that displacement ductility values recommended in this paper correspond to the material strain limits

specified in the MOTEMS. Therefore, these recommendations are strictly valid for piles in marine oil terminals designed according to the MOTEMS. However, the framework presented in paper may easily be used to develop such recommendations for other material strain limits.

Acknowledgments

This research investigation is supported by the California State Lands Commission (CSLC) under Contract No. C2005-051 with Martin Eskijian as the project manager. This support is gratefully acknowledged. Additional support is provided by a grant entitled "C3RP Building Relationships 2008/2010" from the Office of Naval Research under award No. N00014-08-1-1209. This support is also appreciated.

References

Allotey, N. and El Naggar, M.H. (2008), "A numerical study into lateral cyclic nonlinear soil-pile response," *Can. Geotech. J.*, **45**(9), 1268-1281.

Budek, A., Priestley, M.J.N. and Benzoni, G. (2004), "The effects of extemal confinement on flexural hinging in drilled pile shafts," *Earthq. Spectra*, **20**(1), 1-24.

Budek, A.M., Priestley, M.J.N. and Benzoni, G. (2000), "Inelastic seismic response of bridge drilled-shaft RC pile/columns," *J. Struct. Eng-ASCE*, **126**(4), 510-517.

CALTRANS (2006), *Seismic Design Criteria*, Version 1.4, The California Department of Transportation, Sacramento, CA.

Castelli, F. and Maugeri, M. (2009), "Simplified approach for the seismic response of a pile foundation," *J. Geotech. Geoenviron.*, **135**(10), 1440-1451.

Chai, Y.H. (2002), "Flexural strength and ductility of extended pile-shafts. I: analytical model," *J. Struct. Eng-ASCE*, **128**(5), 586-594.

Chai, Y.H. and Hutchinson, T.C. (2002), "Flexural strength and ductility of extended pile-shafts. II: experimental study," *J. Struct. Eng-ASCE*, **128**(5), 595-602.

Chiou, J.S. and Chen, C.H. (2007), "Exact equivalent model for a laterally-loaded linear pile-soil system," *Soils Found.*, **47**(6), 1053-1061.

Dowrick, D.J. (1987), *Earthquake Resistant Design*, 2nd Edition, Wiley-Interscience, New York.

El Naggar, M.H., Shayanfar, M.A., Kimiaei, M. and Aghakouchak, A. (2005), "Simplified BNWF model for nonlinear seismic response analysis of offshore piles with nonlinear input ground motion analysis," *Can. Geotech. J.*, **42**(5), 365-380.

Eskijian, M. (2007), "Marine oil terminal engineering and maintenance standards (MOTEMS)," *Proceedings of 2007 Structures Congress*, ASCE, Long Beach, CA.

Finn, W.D.L. (2005), "A study of piles during earthquakes: issues of design and analysis," *B. Earthq. Eng.*, **3**(2), 141-234.

Goel, R.K. (2008), "Simplified procedures for seismic analysis and design of piers and wharves in marine oil and LNG terminals," Draft Report No. CP/SEAM-08/01, California Polytechnic State University, San Luis Obispo, CA.

Goel, R.K. (2010), "Simplified procedures for seismic evaluation of piles with partial-moment-connection to the deck in marine oil terminals," *J. Struct. Eng-ASCE*, **136**(5).

Imancli, G., Kahyaoglu, M.R., Ozden, G. and Kayalar, A.S. (2009), "Performance functions for laterally loaded single concrete piles in homogeneous clays," *Struct. Eng. Mech.*, **33**(4).

McKenna, F. and Fenves, G. (2001), *The OpenSees Command Language Manual: version 1.2*, Pacific Earthquake Engineering Center, University of California, Berkeley. <*http://opensees.berkeley.edu*>

MOTEMS (2007), *Marine Oil Terminal Engineering and Maintenance Standards* (Informal name), Title 24,

California Code of Regulations, Part 2, California Building Code, Chapter 31F (Marine Oil Terminals), The International Code Council, Washington, D.C.

Priestley, M.J.N., Seible, F. and Calvi, G.M. (1996), *Seismic Design and Retrofit of Bridges*, John Wiley and Sons, Inc., New York.

Roeder, C.W., Graff, R., Soderstrom, J. and Yoo, J.H. (2005), "Seismic performance of pile-wharf connections," *J. Struct. Eng-ASCE*, **131**(3), 428-437.

Song, S.T., Chai, Y.H. and Hale, T.H. (2005). "Analytical model for ductility assessment of fixed-head concrete piles," *J. Struct. Eng-ASCE*, **131**(7), 1051-1059.

Polarization-dependent atomic dipole traps behind a circular aperture for neutral-atom quantum computing

Katharina Gillen-Christandl* and Bert D. Copsey
Physics Department, California Polytechnic State University, San Luis Obispo, California 93407, USA
(Received 26 September 2010; published 9 February 2011)

The neutral-atom quantum computing community has successfully implemented almost all necessary steps for constructing a neutral-atom quantum computer. We present computational results of a study aimed at solving the remaining problem of creating a quantum memory with individually addressable sites for quantum computing. The basis of this quantum memory is the diffraction pattern formed by laser light incident on a circular aperture. Very close to the aperture, the diffraction pattern has localized bright and dark spots that can serve as red-detuned or blue-detuned atomic dipole traps. These traps are suitable for quantum computing even for moderate laser powers. In particular, for moderate laser intensities ($\sim$100 W/cm^2) and comparatively small detunings ($\sim$1000–10 000 linewidths), trap depths of $\sim$1 mK and trap frequencies of several to tens of kilohertz are achieved. Our results indicate that these dipole traps can be moved by tilting the incident laser beams without significantly changing the trap properties. We also explored the polarization dependence of these dipole traps. We developed a code that calculates the trapping potential energy for any magnetic substate of any hyperfine ground state of any alkali-metal atom for any laser detuning much smaller than the fine-structure splitting for any given electric field distribution. We describe details of our calculations and include a summary of different notations and conventions for the reduced matrix element and how to convert it to SI units. We applied this code to these traps and found a method for bringing two traps together and apart controllably without expelling the atoms from the trap and without significant tunneling probability between the traps. This approach can be scaled up to a two-dimensional array of many pinholes, forming a quantum memory with single-site addressability, in which pairs of atoms can be brought together and apart for two-qubit gates for quantum computing.

DOI: 10.1103/PhysRevA.83.023408 PACS number(s): 37.10.Gh, 03.67.Lx, 42.25.Ja, 42.50.Ct

I. INTRODUCTION

Neutral-atom quantum computing [1–4] is a promising avenue toward a full implementation of a quantum computer [5]. The internal electronic state of a neutral atom (or, in some cases, the motional state [6]) serves as the qubit. Usually the qubit states that are chosen are part of the ground-state manifold, resulting in long coherence times, limited by trap photon scattering or motional heating. Initialization, readout, and single-qubit rotations are achieved using well-established spectroscopic techniques. Recent advances have been made in trapping, manipulating, and reading out single atoms trapped in dipole traps [7,8]. Two-qubit gates have been experimentally demonstrated [9,10] using the dipole blockade, and entanglement between two qubits has also been achieved using cold collisions [11].

To scale this system up to many qubits, neutral atoms are most commonly trapped at the sites of a three-dimensional (3D) optical lattice. However, atoms trapped in 3D optical lattices cannot be addressed individually using focused laser beams, due to the 3D structure of the trap array, which limits the operations that can be performed on qubits trapped in 3D optical lattices. As a solution to this problem, several methods for creating two-dimensional (2D) arrays of atom traps have been proposed. Dumke *et al.* experimentally demonstrated a 2D array of atom traps formed behind an array of microlenses [12]. The distance between adjacent traps is determined by the center-to-center distance between the microlenses and thus can be designed to be large enough to address individual atoms with a focused laser beam. Recently, a scalable extension of this method was achieved using a spatial light modulator combined with an array of microlenses [13]. Other ideas using spatial light modulators (SLMs) [14,15], mirrors [16,17], Fresnel lenses [18], metamaterial lenses [19], or diffraction patterns [20,21] are being explored. Of particular interest are approaches that allow trapping atoms in dark spots, reducing the trap photon absorption probability, one of the major decoherence mechanisms in optical traps. Christandl *et al.* have proposed a 2D array of dark atom traps at intertrap distances of several microns, formed by blue-detuned evanescent waves above a waveguide [22]. Two-dimensional arrays of dark-spot traps can also be generated using a combination of a phase plate or grating and an array of microlenses [23,24]. The application of light fields near circular apertures of micron or nanometer sizes for trapping and control of cold atoms has been studied for some time [25–30]. Our approach, which uses diffraction at circular apertures of sizes exceeding the laser wavelength, offers a simple, versatile method for generating a 2D array of either dark-spot or bright-spot traps, depending on the laser detuning. In addition, two traps, including two dark-spot traps, can be brought together and apart without losing the atoms from the traps by utilizing the light polarization dependence of the trapping potential energy. Imaging of 2D arrays similar to the types described previously with single-site resolution has been demonstrated successfully [31], indicating that single-site addressing and readout are possible in 2D arrays.

Laser light incident on a circular aperture forms localized bright and dark spots in the region near the aperture, closer

*kgillen@calpoly.edu

than the usual near-field diffraction [32]. These bright and dark spots can serve as atomic dipole traps for red-detuned and blue-detuned light, respectively [33]. Moderate laser powers ($\sim$100 W/cm^2) and small detunings ($\sim$1000–10 000 linewidths) result in trap depths of $\sim$1 mK and trap frequencies of several to $\sim$10 kHz. These traps are theoretically suitable for storing atomic qubits.

In this paper, we describe how the light-polarization-dependent trapping potential energy [34] for any electric field pattern, for any alkali species, in any magnetic substate is calculated for any detuning small compared to the fine-structure splitting. Much simpler expressions for larger detunings are readily available in the literature [34,35]. We also discuss the normalization conventions encountered for the reduced matrix elements, as exact knowledge of these is needed to obtain the trapping potential energy in nonarbitrary units for direct comparison with experimental results. We then show computational results applying these expressions to the atomic dipole traps formed behind a circular aperture. We show that the traps stay intact upon tilting the incident laser beam, indicating the ability to move these atom traps. We show how the light polarization dependence of atomic dipole traps can be exploited to bring pairs of atoms (including those in blue-detuned traps) together and apart without losing the atoms from the trap. This may allow for the implementation of two-qubit gates with previously realized methods [9,10]. This approach can be scaled up to a 2D array of many circular apertures, with the trap distance determined by the center-to-center distance of the apertures, and therefore adjustable to a distance large enough to allow for resolving of individual atom sites with a focused laser beam for qubit manipulation.

Section II summarizes the theoretical background for the light-polarization-dependent atomic trapping potential-energy calculations. In Sec. III, we present our computational results on moving the atomic dipole traps formed behind a circular aperture and on bringing them together and apart. We also discuss how this approach can be scaled up to many qubits.

II. THEORY OF THE LIGHT POLARIZATION DEPENDENCE OF ATOMIC DIPOLE TRAPS

An electric field such as that from a laser induces an electric dipole moment in a neutral atom. In general, this induced dipole moment depends on the polarization of the laser light, as well as on the hyperfine level and Zeeman magnetic substate of the atom. The induced dipole moment due to a certain electric field is determined by the polarizability of the atom. The interaction of the induced dipole with the electric field of the laser light results in a potential energy and its associated force, which can trap the atom in regions of high or low light intensity. The potential-energy operator for the light atom interaction is given by [36]

$$\hat{U}(\mathbf{r}) = -\tfrac{1}{4}\mathbf{E}_0^*(\mathbf{r})\hat{\alpha}\mathbf{E}_0(\mathbf{r}). \tag{1}$$

Here, $\mathbf{E}_0(\mathbf{r})$ is the complex amplitude for an electric field written in the form $\mathbf{E}(\mathbf{r},t) = \mathrm{Re}[\mathbf{E}_0(\mathbf{r})e^{-i\omega t}]$, $\hat{\alpha}$ is the atomic polarizability tensor (for detailed discussions see [36,37]), and ω is the angular frequency of the laser light. Alternatively, the electric field is often written in its Fourier series form with positive and negative frequency components, $\mathbf{E}^{(+)} = \mathbf{E}_0^*/2$ and $\mathbf{E}^{(-)} = \mathbf{E}_0/2$, respectively,

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}^{(+)}(\mathbf{r})e^{-i\omega t} + \mathbf{E}^{(-)}(\mathbf{r})e^{i\omega t}. \tag{2}$$

The corresponding expression for the dipole potential-energy operator is then

$$\hat{U}(\mathbf{r}) = -\mathbf{E}^{(+)}(\mathbf{r})\hat{\alpha}\mathbf{E}^{(-)}(\mathbf{r}). \tag{3}$$

As derived in [34] and Appendix A, for an alkali atom, the polarizability tensor components in the spherical basis are

$$\hat{\alpha}_{q',q} = (-1)^{q'}\sum_{F'}\left[\alpha_{0,F'F}f_{F'F}\sum_{m}\left(c^{F,1,F'}_{m+q-q',q',m+q}c^{F,1,F'}_{m,q,m+q}\right.\right.$$
$$\left.\left.\times\,|F,m+q-q'\rangle\langle F,m|\right)\right], \tag{4}$$

where $q',q = \pm1,0$ stands for the spherical basis components, $\alpha_{0,F'F}$ is the characteristic polarizability scalar, $f_{F'F}$ is the relative oscillator strength of the $F \to F'$ hyperfine transition, and the c's are the Clebsch-Gordan coefficient for the $F,m \to F',m+q$ dipole transition and a Clebsch-Gordan coefficient related to the $F',m+q \to F,m+q-q'$ dipole transition (see Appendix A), respectively. The relative oscillator strength of an $F \to F'$ transition is

$$f_{F'F} = (2J'+1)(2F+1)\left|\begin{Bmatrix} F' & I & J' \\ J & 1 & F \end{Bmatrix}\right|^2, \tag{5}$$

where the curly braces signify the six-J symbol and I is the quantum number for the nuclear spin of the atom. The characteristic polarizability scalar is given by

$$\alpha_{0,F'F} = -\frac{|\langle J'||d||J\rangle|^2}{\hbar\Delta_{F'F}}. \tag{6}$$

Here, $\Delta_{F'F}$ is the angular frequency detuning from the $F \to F'$ transition and $\langle J'||d||J\rangle$ is the reduced dipole matrix element for the $J \to J'$ fine-structure transition.

There are three common normalization conventions for the reduced dipole matrix element. A comparison of the three conventions, as well as an example for unit conversion of the reduced dipole matrix elements, are given in Appendix B.

In this work, we used the following relation (with the same normalization as [38]) for the polarizability scalar of an alkali atom to calculate the reduced matrix element [36]:

$$\alpha_{0,F'F} = -\frac{3\lambda^3}{32\pi^3}\frac{\Gamma}{\Delta_{F'F}}. \tag{7}$$

This equation is for the wavelength λ in centimeters and gives $\alpha_{0,F'F}$ in cgs units. In our work, we use SI units throughout our code, so we are also listing the SI version of this equation (i.e., with λ in units of meters),

$$\alpha_{0,F'F} = -\frac{3\lambda^3}{32\pi^3}\frac{\Gamma}{\Delta_{F'F}}1.11\times10^{-10}\frac{\mathrm{Jm}^2}{\mathrm{V}^2}. \tag{8}$$

In this study, we only present the diabatic potentials [39], which are the diagonal components of the potential-energy operator in the F, m_F basis. This is appropriate, because we plan on trapping precooled atoms in these traps, which will remain at the bottom of the potential-energy wells, rather than traveling through the wells. For the configurations that involve

movement of the traps, we need to consider two aspects of the speed of this motion. First, it must be slow compared to the trap frequency in order to reduce atom loss or state disturbance during motion. On the other hand, the motion should occur fast enough to reduce the probability of Raman transitions that could flip atoms into a different magnetic substate if the two atomic potential energies are very similar (e.g., when two traps are fully overlapped). Such Raman transitions can change the state of an atom into either an untrapped state or a state trapped by a different well (see Sec. III). As long as the Raman transition probability remains sufficiently low, we can use the diabatic potentials to describe the atom dynamics in our traps. To calculate the diabatic potential energy for a certain F, m_F state, we calculate the expectation value of the potential-energy operator as follows:

$$U_{F,m_F} = \langle F,m_F|\hat{U}|F,m_F\rangle. \tag{9}$$

Plugging in the dipole energy operator [Eq. (1)] explicitly, we get

$$\begin{aligned} U_{F,m_F} &= -\frac{1}{4}\sum_{q',q} E^*_{0q'}E_{0q}\hat{\alpha}_{q',q} \\ &= -\frac{1}{4}\sum_{q',q}\left\{(-1)^{q'}E^*_{0q'}E_{0q}\sum_{F'}\left[\alpha_{0,F'F}f_{F'F}\right.\right. \\ &\quad\times\sum_m\left(c^{F,1,F'}_{m+q-q',q',m+q}c^{F,1,F'}_{m,q,m+q}\right. \\ &\quad\left.\left.\left.\times\langle F,m_F|F,m+q-q'\rangle\langle F,m|F,m_F\rangle\right)\right]\right\}. \end{aligned} \tag{10}$$

Here, the $E_{0q',q}$ with $q',q = \pm 1,0$ are the spherical components of the electric field amplitude, corresponding to right and left circular light polarization, $\sigma^{\pm}$, and linear light polarization, π, respectively.

By exploiting orthonormality, this expression simplifies to

$$\begin{aligned} U_{F,m_F} &= -\frac{1}{4}\sum_q(-1)^q|E_{0q}|^2 \\ &\quad\times\sum_{F'}\alpha_{0,F'F}f_{F'F}\left(c^{F,1,F'}_{m_F,q,m_F+q}\right)^2. \end{aligned} \tag{11}$$

The electric field distributions for the diffraction pattern immediately behind a circular aperture were obtained using Hertz vector diffraction theory [40,41]. The diffraction code [42] yields the Cartesian components of the electric field. To find the spherical components, we use the spherical unit vectors [43]

$$\begin{aligned} \mathbf{e}_{-1} &= \frac{1}{\sqrt{2}}(\mathbf{e}_x - i\mathbf{e}_y) \\ \mathbf{e}_0 &= \mathbf{e}_z \\ \mathbf{e}_{+1} &= -\frac{1}{\sqrt{2}}(\mathbf{e}_x + i\mathbf{e}_y). \end{aligned} \tag{12}$$

Here, $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are the Cartesian unit vectors. From this, we find the spherical components of the complex amplitude of the electric field defined by

$$\begin{aligned} \mathbf{E}_0 &= E_{0x}\mathbf{e}_x + E_{0y}\mathbf{e}_y + E_{0z}\mathbf{e}_z \\ &= E_{0-1}\mathbf{e}_{-1} + E_{00}\mathbf{e}_0 + E_{0+1}\mathbf{e}_{+1}. \end{aligned} \tag{13}$$

Here E_{0j} for $j = x,y,z$ are the Cartesian components of the electric field amplitude. The spherical and Cartesian components are related by

$$\begin{aligned} E_{0-1} &= \frac{1}{\sqrt{2}}(E_{0x} + iE_{0y}), \\ E_{00} &= E_{0z}, \\ E_{0+1} &= \frac{1}{\sqrt{2}}(-E_{0x} + iE_{0y}). \end{aligned} \tag{14}$$

These spherical components were then plugged into Eq. (11).

Equations (5), (8), (11), and (14) were used to calculate the computational results presented in the next section. The computations were performed using a code [44] that will take any arbitrary electric field distribution in Cartesian coordinates, decompose it into its spherical components, and then calculate both the diabatic and adiabatic potentials for any given detuning $\Delta \ll \Delta_{fs}$, where Δ_{fs} denotes the fine-structure splitting of the excited state. Note that the detuning can be made arbitrarily small, including smaller than the excited-state hyperfine splitting. For larger detunings, the expressions simplify tremendously, as only the fine-structure splitting needs to be considered [34,35]. Also note that the E_{0j} are complex, so both the real and imaginary components must be supplied for this calculation. Appendix C shows the details of our specific electric field configurations (single laser beam incident at an angle and a pair of oppositely circularly polarized laser beams incident at an angle). The code has two variable input parameters: the laser detuning Δ from the transition $F \rightarrow$ maximum F', which is contained in $\Delta_{F'F}$, and the laser intensity I_0. The electric field amplitude at the aperture for each of the incident beams starts out normalized to 1. To change this to meaningful units, we insert the scaling due to intensity and convert the units of the potential energy from joules to millikelvin as follows:

$$U_{F,m_F}(\text{mK}) = 1000\frac{2}{3k_B}\frac{2I_0}{\epsilon_0 c}U_{F,m_F}. \tag{15}$$

Here, the factor of 1000 is for converting kelvin to millikelvin, the factor of $\frac{2}{3k_B}$ is for conversion from joules to Kelvin, and the factor of $\frac{2I_0}{\epsilon_0 c}$ is for inserting physical units for the electric field. For an intensity in W/m^2, this yields electric field units of volts per meter. Also, k_B is Boltzmann's constant, ϵ_0 is the permittivity of free space, and c is the speed of light. This is how we obtained the numerical results presented in the next section.

III. COMPUTATIONAL RESULTS FOR ATOM TRAPS BEHIND A CIRCULAR APERTURE

A. Movable atomic dipole traps

Consider the diffraction pattern resulting from a laser beam incident on a circular aperture at an angle of $\gamma = 0.055$ rad as shown in Fig. 1. Depending on the laser detuning, 3D atom traps will form in either the localized bright spots or dark spots of this diffraction pattern very close to the aperture. Figure 1(c) shows the diabatic trapping potential energy [calculated from the electric field distribution using Eq. (11)] for the $F = 1$, $m_F = 0$ magnetic substate of ^{87}Rb, for a laser intensity of 364 W/cm^2 and a laser detuning of $-10000\,\Gamma$ (red detuning)

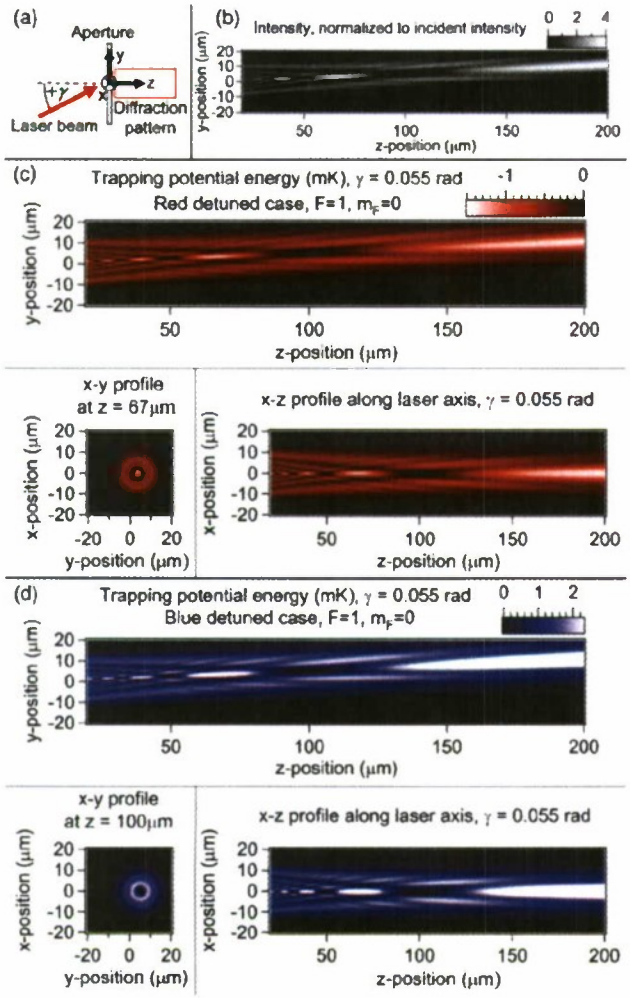


FIG. 1. (Color online) Diabatic trapping potential energy for a single laser beam (σ^+ polarization was used) incident on a circular aperture at an angle of $\gamma = 0.055$ rad. (a) Diagram of setup. (b) Intensity pattern. (c) Trapping potential energy for the light-polarization-independent $F = 1, m_F = 0$ magnetic substate of ^{87}Rb, for a laser intensity of 364 W/cm^2 and a laser detuning of $-10000\ \Gamma$. (d) Trapping potential energy for the $F = 1, m_F = 0$ magnetic substate of ^{87}Rb, for a laser intensity of 116 W/cm^2 and a laser detuning of 1000 Γ.

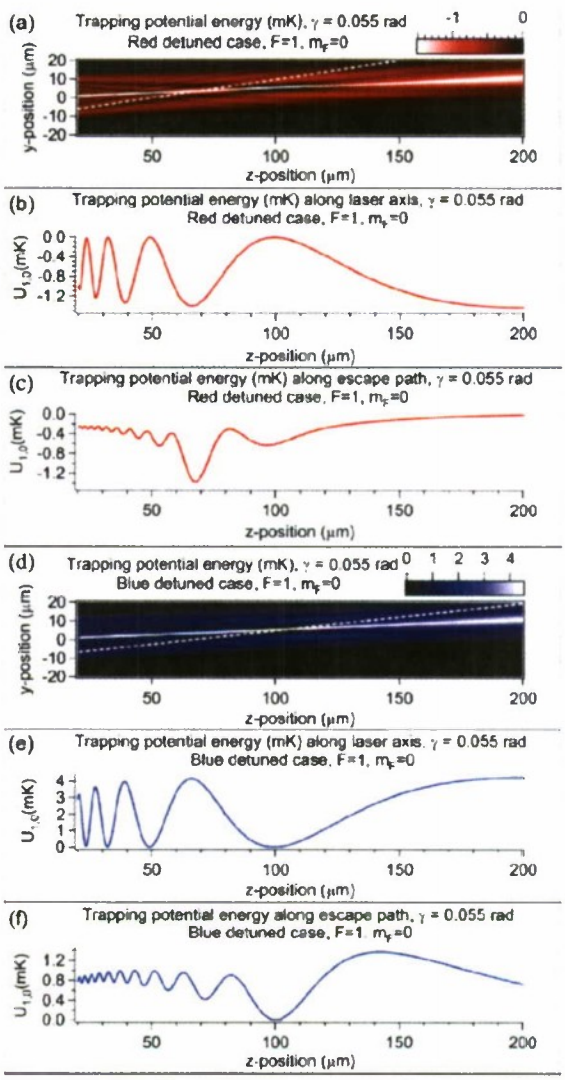


FIG. 2. (Color online) Diabatic trapping potential energy along the laser axis and along the paths of weakest confinement ("escape paths") for red-detuned and blue-detuned traps, for a single laser beam (σ^+ polarization was used) incident at $\gamma = 0.055$ rad, for the light-polarization-independent $F = 1, m_F = 0$ magnetic substate of ^{87}Rb. (a) Axial path (solid line) and escape path (dashed line) for a laser intensity of 364 W/cm^2 and a laser detuning of $-10000\ \Gamma$. (b) Trapping potential energy along the laser axis for red-detuned trap [solid line in (a)]. (c) Trapping potential energy along the escape path for red-detuned trap [dashed line in (a), weakest trap direction]. (d) Axial path (solid line) and escape path (dashed line) for a laser intensity of 116 W/cm^2 and a laser detuning of 1000 Γ. (e) Trapping potential energy along the laser axis for the blue-detuned trap [solid line in (b)]. (f) Trapping potential energy along the escape path for the blue-detuned trap [dashed line in (b), weakest trap direction].

from the Rb D_2 transition ($\lambda = 780$ nm). The D_2 linewidth of Rb is $\Gamma = 2\pi \times 6$ MHz [43]. Atoms are trapped in the bright spots on the laser beam axis. Similarly, Fig. 1(d) shows the diabatic trapping potential energy for the $F = 1, m_F = 0$ magnetic substate of ^{87}Rb, for a laser intensity of 116 W/cm^2 and a laser detuning of 1000 Γ (blue detuning) from the Rb D_2 transition. Here, localized atom traps form in the dark spots on the laser beam axis.

We analyzed the properties of the traps formed farthest from the aperture ($z = 67\ \mu$m for the farthest, well-localized bright spot for the red-detuned case and $z = 100\ \mu$m for the blue-detuned case) and compared them to the normal incidence case. We chose these traps because they are biggest and most easily accessible for initial experiments. The traps formed closer to the aperture are also viable and, in fact, advantageous for quantum computing as they have larger trap frequencies. To determine the trap frequencies, we approximated the bottom of the (nonharmonic) traps with a harmonic oscillator potential-energy well. The values of the trap frequencies obtained depend on the fit range used. In this work, we chose a fit range of the bottom 200 μK of the well, unless otherwise stated.

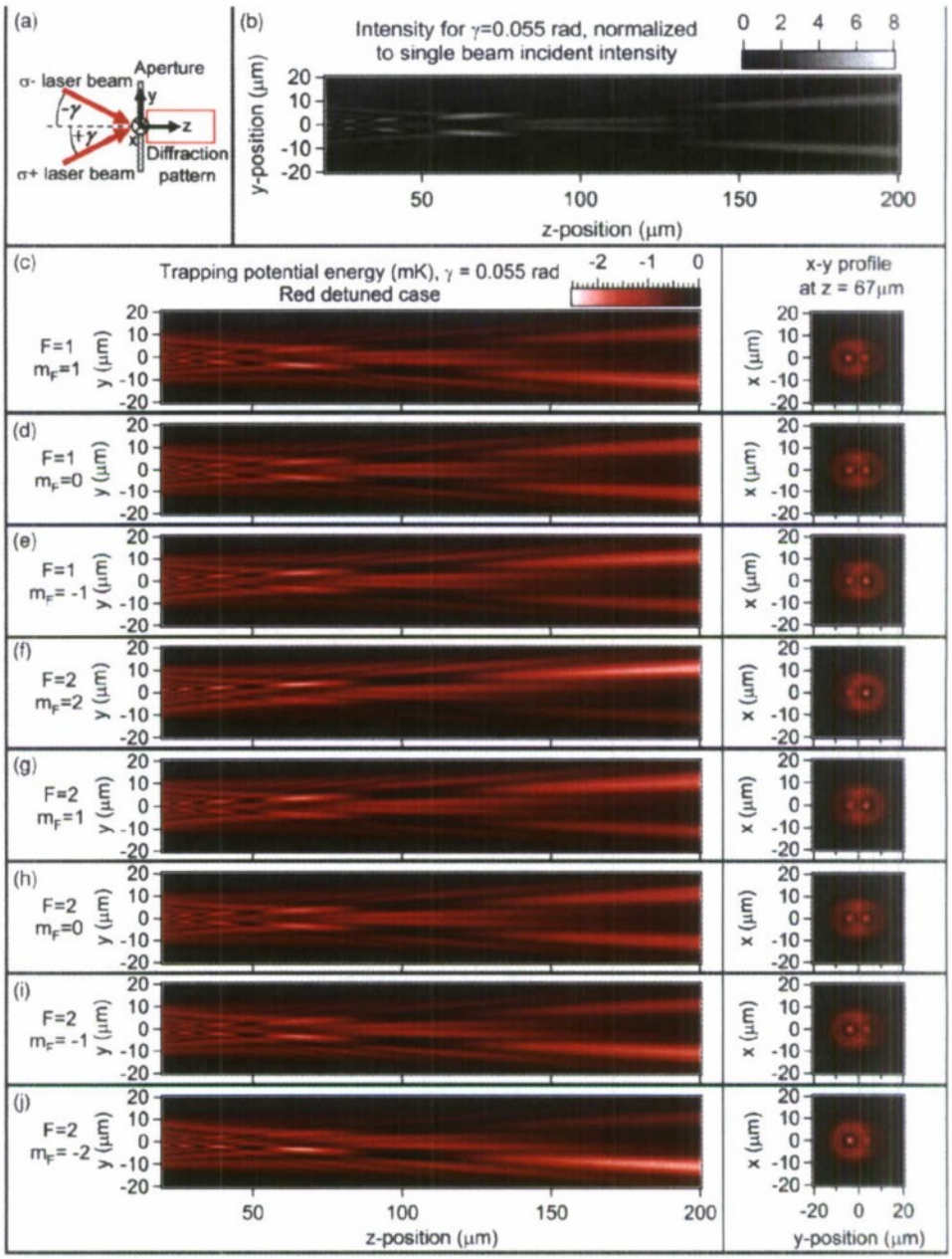


FIG. 3. (Color online) Diabatic trapping potential energy for two laser beams of opposite circular polarization incident on a circular aperture at an angle of $\gamma = 0.055$ rad. (a) Diagram of setup. (b) Intensity pattern, normalized to the incident intensity of one incident circularly polarized laser beam. (c)–(j) Trapping potential energy of the intensity pattern in (b) for the eight magnetic substates of the hyperfine ground-state manifold in ^{87}Rb for a laser intensity of 364 W/cm^2 and a laser detuning of $-10\,000\,\Gamma$ from the transitions from the respective F states.

Other trap properties of relevance for quantum information applications are the size of the motional harmonic oscillator ground-state wave function along a spatial dimension j,

$$\beta_j = \sqrt{\frac{\hbar}{2\pi f_j m_{\mathrm{Rb}}}}, \tag{16}$$

for the $1/e$ half-width of the probability density, and the energy difference $\Delta U_j \equiv h f_j / k_{\mathrm{B}}$ between two motional states of the potential-energy well. Here, f_j is the trap frequency along spatial dimension j, and m_{Rb} is the mass of one ^{87}Rb atom. We denote the spatial dimensions of the trap with indices $j = r_x$ for the radial dimension along x in the coordinate system of Fig. 1(a), $j = r_{yz}$ for the radial dimension in the y-z plane, and $j =$ axial for the axial dimension of the trap. In addition, the coherence of qubits in dipole traps is often limited by the scattering rate of trap photons. For blue-detuned traps with zero intensity at the bottom, the scattering rate for a ground-state atom (averaged over the extent of the wave function) can be written as [45]

$$\eta = \frac{\pi}{2}\left(f_{r_x} + f_{r_{yz}} + f_{\mathrm{axial}}\right)\frac{\Gamma}{\Delta}. \tag{17}$$

TABLE I. Trap properties for all F, m_F hyperfine ground states in ^{87}Rb for the red-detuned ($I_0 = 364$ W/cm^2, $\Delta = -10000\,\Gamma$) and blue-detuned ($I_0 = 116$ W/cm^2, $\Delta = 1000\,\Gamma$) examples. For trap frequencies, ground-state sizes, and energy differences, values are given in the order r_x, r_{yz}, axial. ΔU_{trap} denotes the trap depth.

	Red detuned					Blue detuned				
State F,m_F	f (kHz)	β (nm)	ΔU_j (μK)	η (kHz)	ΔU_{trap} (mK)	f (kHz)	β (nm)	ΔU_j (μK)	η (kHz)	ΔU_{trap} (mK)
1,1	83, 86, 8.1	37, 37, 120	4, 4.1, 0.39	42	1.2	27, 37, 6	65, 56, 140	1.3, 1.8, 0.29	0.11	1.2
1,0	74, 79, 7.5	40, 38, 120	3.6, 3.8, 0.36	38	0.98	26, 35, 5.3	67, 58, 150	1.2, 1.7, 0.26	0.10	1.0
1, − 1	83, 86, 8.1	37, 37, 120	4, 4.1, 0.39	42	1.2	27, 37, 6	65, 56, 140	1.3, 1.8, 0.29	0.11	1.2
2,2	90, 92, 8.7	36, 36, 120	4.4, 4.4, 0.42	46	1.5	29, 41, 6.8	63, 53, 130	1.4, 2, 0.33	0.12	1.5
2,1	83, 86, 8.1	38, 37, 120	4, 4.1, 0.39	42	1.2	28, 38, 6.2	65, 55, 140	1.3, 1.8, 0.3	0.11	1.3
2,0	74, 79, 7.5	40, 38, 120	3.6, 3.8, 0.36	37	0.98	26, 35, 5.4	67, 58, 150	1.2, 1.7, 0.26	0.10	1.1
2, − 1	83, 86, 8.1	38, 37, 120	4, 4.1, 0.39	42	1.2	28, 38, 6.2	65, 55, 140	1.3, 1.8, 0.3	0.11	1.3
2, − 2	90, 92, 8.7	36, 36, 120	4.4, 4.4, 0.42	46	1.5	29, 41, 6.8	63, 53, 130	1.4, 2, 0.33	0.12	1.5

For red-detuned traps, a conservative estimate of the scattering rate is the peak scattering rate [35]

$$\eta = \frac{\Gamma}{\Delta\hbar} U_{\text{min}}, \tag{18}$$

where U_{min} is the potential energy at the intensity peak of the trap. We determined the trap depth ΔU_{trap} by finding the peak potential energy of the path of weakest confinement ("escape path") with respect to the potential energy of the bottom of the well.

We determined that the trap properties calculated in [33] for normal incidence stay largely intact when the laser is incident at an angle. For comparisons to [33], we must mention that the trap frequencies cited there are for a harmonic fit range of 1 mK, whereas in this work we cite trap frequencies for 200 μK, which we deemed the relevant range for an atom sample precooled in a magneto-optical trap. The corresponding normal incidence frequencies for the red-detuned example are a radial trap frequency of $f_{r_x} = f_{r_{yz}} =$ 71 kHz and an axial trap frequency of $f_{\text{axial}} = 6.9$ kHz. The other trap properties are $\beta_{r_x} = \beta_{r_{yz}} = 40$ nm, $\beta_{\text{axial}} =$ 130 nm, $\Delta U_{r_x} = \Delta U_{r_{yz}} = 3.4$ μK, and $\Delta U_{\text{axial}} = 0.33$ μK. The trap-photon scattering rate is 27 kHz. The trap depth is 1 mK. Similarly, for blue-detuned light at normal incidence, we have $f_{r_x} = f_{r_{yz}} = 28$ kHz, $f_{\text{axial}} = 5.6$ kHz, $\beta_{r_x} = \beta_{r_{yz}} =$ 64 nm, $\beta_{\text{axial}} = 140$ nm, $\Delta U_{r_x} = \Delta U_{r_{yz}} = 1.3$ μK, and $\Delta U_{\text{axial}} = 0.27$ μK. For the blue-detuned traps, the deviation from a harmonic potential-energy well is particularly pronounced, with the bottom being very flat. To fully describe these traps, we performed fits for a fit range of 20 μK, yielding fits valid only for very-low-temperature atoms (<1 μK) such as for atoms loaded from a Bose-Einstein condensate. The radial trap frequency for the bottom of the well for normal incidence is approximately 10 kHz and thus is comparable to the axial trap frequency (as are the other properties, $\beta_{r_x} = \beta_{r_{yz}} = 0.11$ μm, $\Delta U_{r_x} = \Delta U_{r_{yz}} = 0.48$ μK). A conservative estimate (using the larger trap frequencies) for the scattering rate is 97 Hz. The trap depth is 1 mK.

Figure 2 shows the trapping potential-energy curves along the laser beam direction, as well as along the direction of weakest confinement for both the red- and blue-detuned examples listed previously. For a beam incident at an angle of $\gamma = 0.055$ rad, for the red-detuned example (laser beam intensity of 364 W/cm^2, laser detuning of $-10000\,\Gamma$), we find a trap depth of 1 mK, $f_{r_x} = f_{r_{yz}} = 74$ kHz, $f_{\text{axial}} = 6.8$ kHz, $\beta_{r_x} = \beta_{r_{yz}} = 40$ nm, $\beta_{\text{axial}} = 130$ nm, $\Delta U_{r_x} = \Delta U_{r_{yz}} =$ 3.6 μK, and $\Delta U_{\text{axial}} = 0.33$ μK. The average scattering rate is 27 kHz. These results are very similar to those for normal incidence. For the blue-detuned example (laser beam intensity of 116 W/cm^2, laser detuning of $1000\,\Gamma$), we find a trap depth of 0.9 mK, $f_{r_x} = f_{r_{yz}} = 26$ kHz, $f_{\text{axial}} = 5.5$ kHz, $\beta_{r_x} = \beta_{r_{yz}} = 67$ nm, $\beta_{\text{axial}} = 150$ nm, $\Delta U_{r_x} = \Delta U_{r_{yz}} =$ 1.3 μK, and $\Delta U_{\text{axial}} = 0.26$ μK. The radial frequencies describing the behavior of the bottom of the blue well are $f_{r_x} = 5.6$ kHz and $f_{r_{yz}} = 6.7$ kHz, with $\beta_{r_x} = 140$ nm, $\beta_{r_{yz}} =$ 130 nm, $\Delta U_{r_x} = 0.27$ μK, and $\Delta U_{r_{yz}} = 0.32$ μK. At this level, we are starting to notice differences between the untilted x direction and the dimension with tilt (y-z plane). Again, these results are very close to the normal incidence values. Therefore, the traps stay intact upon tilting.

Thus, by tilting the incident laser beam, an atom trapped at a bright spot or dark spot can be moved. We propose using this to bring two qubits together and apart by employing two laser beams at an angle, as shown in Fig. 3. One atom is placed in each of the two bright-spot traps (for red-detuned light) or dark-spot traps (for blue-detuned light). The laser beams are then tilted together to overlap the atoms for two-qubit quantum operations and tilted apart to separate the atoms.

There are several issues with this approach. When overlapping the wells, there is a significant probability for the atoms to tunnel between the traps and switch places. This is detrimental for quantum computing. In addition, for blue-detuned traps, the wall from one trap will push the atom out of the other trap. Both of these issues can be addressed by exploiting the light polarization dependence of the trapping potential energy for atoms in different magnetic substates, as discussed in the next section.

B. Atomic trapping potential energy for different magnetic substates

Because of the dependence of the trapping potential energy on the light polarization and the magnetic substate of a trapped atom as outlined in Sec. II, atoms in different magnetic

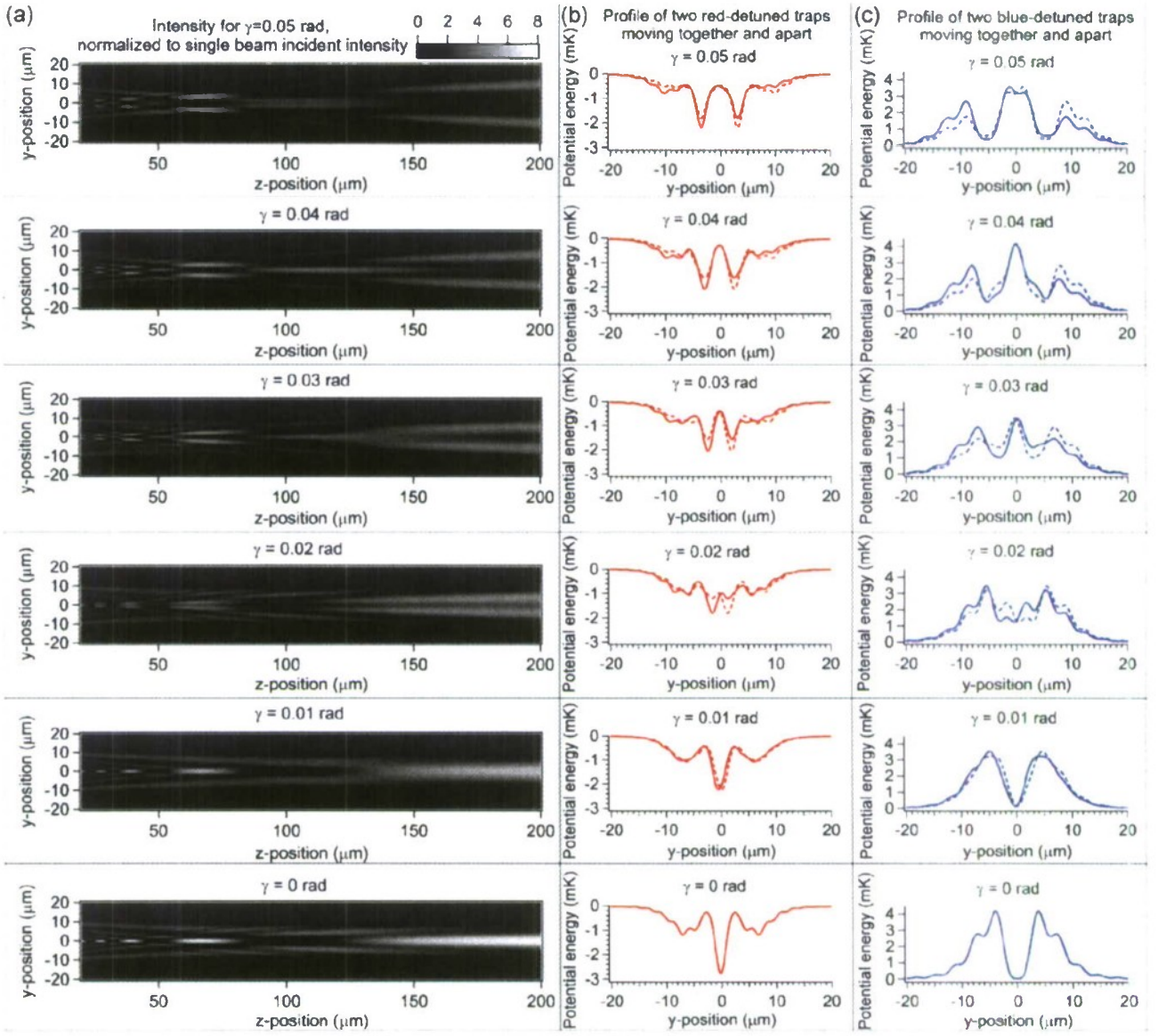


FIG. 4. (Color online) Bringing two traps together. (a) Column showing the intensity pattern, normalized to the incident intensity of one incident circularly polarized laser beam, for several incident angles γ. (b) Column showing the diabatic potential-energy profile along the y direction at $z = 67$ μm for the $F = 1$, $m_F = 1$ (solid line) and $m_F = -1$ (dashed line) magnetic substates of ^{87}Rb trapped in the intensity pattern from (a) for several incident angles γ, a laser intensity of 364 W/cm^2, and a laser detuning of $-10\,000$ Γ. (c) Column showing the diabatic potential-energy profile along the y direction at $z = 100$ μm for the $F = 1$, $m_F = 1$ (solid line) and $m_F = -1$ (dashed line) magnetic substates of ^{87}Rb trapped in the intensity pattern from (a) for several incident angles γ, a laser intensity of 116 W/cm^2, and a laser detuning of 1000 Γ.

substates placed in the same light pattern have a different trapping potential-energy curve. Consider the configuration in Fig. 3. A right-circularly polarized (σ^+) laser beam and a left-circularly polarized (σ^-) laser beam are incident on a circular aperture at angles γ and $-\gamma$, respectively. Figure 3(b) shows the intensity pattern for an incident angle of $\gamma = 0.055$ rad. We chose this angle as an example, as the trap sites examined here are well separated for this angle. Figures 3(c)–3(j) show the diabatic trapping potential energy for the eight magnetic substates of the ^{87}Rb hyperfine ground-state manifold, for the red-detuned example. The results are for a laser detuning of $\Delta = -10\,000$ Γ from the $F = 1 \rightarrow F' = 3$ transition and from the $F = 2 \rightarrow F' = 3$ transition, respectively. Table I summarizes the trap properties for both the red-detuned trap formed by the well-localized bright spot farthest from the aperture ($z = 67$ μm) and the blue-detuned trap formed by the dark spot farthest from the aperture ($z = 100$ μm). Note that only the 200-μK fit results are shown. For blue-detuned traps, the properties of the bottom of the well are better approximated with a 20-μK fit. For the x direction, this yields trap properties comparable to the axial trap properties listed. For the y-z plane, a miniwell is formed at the bottom of the major well and has

trap properties comparable to the radial trap properties listed in the table. The polarization dependence of the potential energy for each magnetic substate is evident. For both examples, we chose a detuning that is large compared to the excited-state hyperfine splitting, and thus there is very little dependence of the detuning on F' during the F' summation in Eq. (11). Also, the results shown are for the same detuning from the F to F' transition for both $F = 1$ and $F = 2$, so the differences in the potential energies are mostly due to the magnetic substate and the light polarization.

One consequence of this polarization dependence of the potential energy for use in quantum computing is the following. As can be seen in Fig. 3(c), we find that an atom (qubit) in the $F = 1$, $m_F = 1$ substate experiences strong confinement in all dimensions in the bright spot from the σ^- beam (i.e., the bottom or left bright spot in the y-z or x-y profile, respectively), at $z = 67\ \mu$m from the aperture. While the bright spot due to the σ^+ beam (top or right bright spot in the y-z or x-y profile, respectively) is also confined in all dimensions, the confinement is significantly weaker due to the polarization dependence of the potential energy. Therefore, a $F = 1, m_F = 1$ atom, seeking the location of lowest potential energy is trapped in the σ^- bright spot. Similarly, as visible in Fig. 3(e), an atom in the $F = 1, m_F = -1$ substate is trapped in the σ^+ bright spot. Consequently, both atoms (qubits) can be stored in separate locations within the same light pattern, shown in Fig. 3(b). A similar polarization dependence effect has been successfully demonstrated in 3D optical lattices [46].

C. Bringing atom traps together and apart for two-qubit operations

1. Red-detuned diffraction trap

Figure 4 shows the trapping potential-energy plots for an atom in the $F = 1, m_F = 1$ substate, trapped in the σ^- beam, and an atom in the $F = 1, m_F = -1$ substate, trapped in the σ^+ beam of Fig. 3(a), for several angles. Figure 4(a) depicts the intensity pattern created by the setup from Fig. 3(a) for several angles, and Fig. 4(b) shows the potential-energy profile along the y direction, at the location of the primary red-detuned trap, $z = 67\ \mu$m, for a pair of 364-W/cm^2 laser beams with opposite circular polarization, tuned 10 000 linewidths to the red of the ^{87}Rb D_2 transition.

Figure 4(b) demonstrates that for the red-detuned case the two traps move together continuously as the lasers are tilted to normal incidence, at which point the two traps are completely overlapped. This process can be reversed by tilting the laser beams apart. Each atom will be most probable to follow its trap as there is a difference in trapping strength between the two traps, due to the light polarization dependence. For instance, an atom in state $F = 1$, $m_F = 1$, trapped in the primary bright spot of the normal incidence configuration, follows the solid potential-energy curve toward the negative y direction in Fig. 4(b) as the angle γ of the two beams is slowly increased. By the same means, an atom in $F = 1$, $m_F = -1$ remains in the dashed potential-energy minimum, moving toward the positive y direction as the two beams are separated. In this way, we can bring two atoms together and apart without experiencing significant trap or tunneling losses.

The major source of tunneling in this setup is due to trap photon Raman transitions when the potential-energy curves for both atoms (i.e., both states) cross, for example, when the wells are completely overlapped. The probability of such a transition can be reduced by performing this operation significantly faster than the Raman scattering rate.

2. Blue-detuned diffraction trap

Figure 4(c) shows the corresponding trapping potential-energy profiles at the location of the primary blue-detuned trap, $z = 100\ \mu$m, for a pair of 116-W/cm^2 laser beams with opposite circular polarization, tuned 1000 linewidths to the blue of the ^{87}Rb D_2 transition.

Since for blue-detuned light atoms are trapped in dark spots, it may generally be desirable to use blue-detuned traps for quantum computing to ensure long decoherence times. As illustrated in Fig. 4(a), as the two beams are moved together and apart, an intensity wall moves through the dark spot of each beam. However, as can be seen in Fig. 4(c), the associated potential energy of the intensity wall is not large enough to push the atoms out of the trap. Thus, the two atoms can still be overlapped completely, without switching wells, as the σ^- well traps the $F = 1, m_F = 1$ atom more strongly than the σ^+ well, and vice versa.

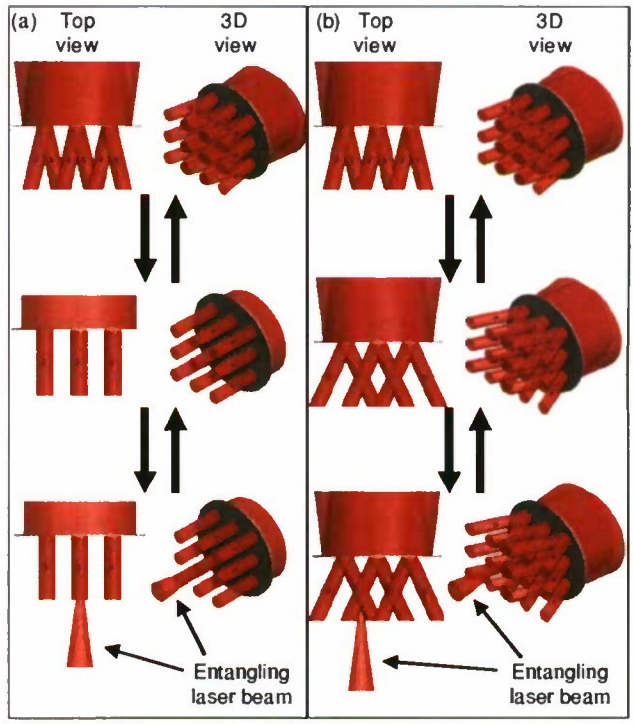


FIG. 5. (Color online) A 2D array of diffraction traps behind a 2D array of circular apertures. For appropriate aperture spacings, e.g., a few microns, individual trap sites can be addressed with a focused laser beam. Two atoms are brought together by tilting the incident laser beams, entangled with a focused laser beam, and moved apart by tilting the laser beams back. (a) Laser beams are tilted to normal incidence, bringing two traps from the same aperture together. (b) Laser beam tilt is increased, bringing two traps from neighboring apertures together.

However, the potential wall will influence the motion of the atom. For quantum computing, this disturbance must be kept negligible or must be reversible to avoid deterioration of the computation. If this is not possible, the advantages of trapping the atoms in locations of low light intensity are erased by the disturbance due to this potential-energy wall, and trapping in the bright spots with sufficiently large detuning may be preferable.

3. *Scaling up to many traps*

The approach presented here can be scaled up to a large array of circular apertures illuminated by a pair of laser beams, as shown in Fig. 5. Each aperture has the previously described potential-energy patterns behind it, so each aperture can trap one atom in each of the two traps that are formed. The distance between adjacent pairs of traps is equal to the distance between the apertures, and the distance between individual atoms is further controlled by the tilting angle of the laser beams. By making aperture arrays with a few microns between apertures, individual trap sites can be resolved by a focused laser beam. By tilting the laser beams to normal incidence and back, as shown in Fig. 5(a), we can bring all pairs of atoms together and apart simultaneously, and either perform large-scale parallel quantum operations or address individual pairs to entangle them. By tilting the laser beam farther, we can bring pairs of atoms from neighboring apertures together and entangle them, as shown in Fig. 5(b). This opens up the possibility of creating a fully functional quantum memory with individually addressable sites and the ability to bring pairs of qubits together and apart.

IV. CONCLUSIONS

We have computationally explored the feasibility of using an array of circular apertures to store qubits for quantum computing. We developed a code that allows the computation of the potential-energy curve for any electric field distribution, any magnetic substate of any alkali atom, and any laser detuning that is much smaller than the excited-state fine-structure splitting. Using this code, we determined that dipole traps formed in the diffraction pattern immediately behind a circular aperture can be moved by tilting the incident laser beam, without significantly changing or diminishing the trap properties. This allows moving atoms trapped in these patterns. Furthermore, we showed that by exploiting the light polarization dependence of the potential energy, two atoms in different magnetic substates trapped in two laser beams of opposite circular polarization can be brought together and apart by tilting the laser beams, without expelling the atoms and without significant probability of tunneling. This may be used to facilitate entangling two-qubit operations, such as the gates demonstrated in [9,10]. This method can be scaled up to a 2D array of many apertures. The distance between individual qubits is determined by the distance between adjacent apertures as well as the laser beam tilt, so it can be designed such that single-site resolution with a focused laser beam is possible. It is thus possible to create a 2D array of qubits that are individually addressable and can be brought together and apart for two qubit operations. In this work we have examined the basic principle and feasibility of this kind of qubit array. We look forward to exploring the limits and possibilities of this approach (e.g., maximum tilt angle to reach beyond immediate neighbors, minimum aperture size, use of multiple layers of traps) in future work.

ACKNOWLEDGMENTS

We thank Glen D. Gillen for providing the electric field data for diffraction at a circular aperture. We also thank Thomas D. Gutierrez, Ivan H. Deutsch, and Marianna S. Safronova for helpful discussions. This work was supported by the Department of the Navy, Office of Naval Research, Grant Nos. N00014-06-1-1111, N00014-07-1-1152, and N00014-08-1-1209, and National Science Foundation Grant No. PHY-0855524.

APPENDIX A: DERIVATION OF THE ATOMIC POLARIZABILITY TENSOR COMPONENTS

The spherical components of the polarizability tensor for an alkali atom of nuclear spin I, in a given hyperfine ground state F (with total angular momentum J) coupled to an excited-state hyperfine manifold with states F' (with angular momentum J') are derived in Appendix A of [34]. In this section, we show additional steps to supplement their derivation. We start from Eq. (A1) of [34],

$$\begin{aligned}\hat{\alpha}_{q',q} = -\sum_{F',m}\Bigg[&\frac{1}{\hbar\Delta_{F'F}}|F,m+q-q'\rangle \\ &\times\langle F,m+q-q'|\hat{d}_{-q'}|F',m+q\rangle \\ &\times\langle F',m+q|\hat{d}_q|F,m\rangle\langle F,m|\Bigg],\end{aligned} \tag{A1}$$

where $\hat{d}$ is the dipole transition operator and q',q correspond to the light polarization components. We can rewrite the two matrix elements using the Wigner-Eckart theorem in the Rose convention (Eq. (5.14) of [38], see Appendix B and Table II) as

$$\begin{aligned}\langle F,m+q-q'|\hat{d}_{-q'}|F',m+q\rangle &= c^{F',1,F}_{m+q,-q',m+q-q'}\langle F||d||F'\rangle, \\ \langle F',m+q|\hat{d}_q|F,m\rangle &= c^{F,1,F'}_{m,q,m+q}\langle F'||d||F\rangle.\end{aligned} \tag{A2}$$

For a more compact expression, it is useful to recast $\langle F||d||F'\rangle$ in terms of the complex conjugate of $\langle F'||d||F\rangle$ using Eq. (C.85) from [49], as well as the relationship between the Racah and Rose conventions for the reduced matrix element (see Appendix B and Table II),

$$\langle F||d||F'\rangle = (-1)^{F'-F}\sqrt{\frac{2F'+1}{2F+1}}\langle F'||d||F\rangle^*. \tag{A3}$$

This yields

$$\begin{aligned}\hat{\alpha}_{q',q} = -\sum_{F',m}\Bigg[&\frac{1}{\hbar\Delta_{F'F}}c^{F',1,F}_{m+q,-q',m+q-q'}c^{F,1,F'}_{m,q,m+q} \\ &\times(-1)^{F'-F}\sqrt{\frac{2F'+1}{2F+1}} \\ &\times|\langle F'||d||F\rangle|^2|F,m+q-q'\rangle\langle F,m|\Bigg].\end{aligned} \tag{A4}$$

TABLE II. Symbols and conventions for Clebsch-Gordan coefficients and reduced matrix elements from several sources. In the last row, we define the variable R to be the reduced matrix element of Rose [38].

Quantity	Condon-Shortley [47]	Racah [48]	Messiah [49]	Edmonds [50]	Rose [38]	This work[a]
Final-state quantum no.						
Principal	α	α	τ	γ'	(omitted)	(omitted)
Angular momentum	j	j	J	j'	j'	J'
Magnetic	m	m	M	m'	m'	m'
Initial-state quantum no.						
Principal	α'	α'	τ'	γ	(omitted)	(omitted)
Angular momentum	$j+1, j-1, j$	j'	J'	j	j	J
Magnetic	$m+1, m-1, m$	m'	M'	m	m	m
Rank of operator	1	k	k	k	L	k
Operator	T	$T^{(k)}$	$\mathbf{T}^{(k)}$	$\mathbf{T}(k)$	T_L	$T^{(k)}$
Component of operator	$\frac{1}{2}(\hat{i} \pm i\hat{j}), \hat{k}$	$T_q^{(k)}$	$T_q^{(k)}$	$T(kq)$	T_{LM}	$T_q^{(k)}$
Clebsch-Gordan coeff.	N/A	$(j'km'q\|j'kjm)$	$\langle J'kM'q\|JM\rangle$	$(kqjm\|kjj'm')$	$C(jLj';mMm')$	$c^{jkj'}_{mqm'}$
Reduced matrix element	$(\alpha j\|\|T\|\|\alpha' j \pm 1, j)$[b]	$(\alpha j\|\|T^{(k)}\|\|\alpha' j')$	$\langle \tau J\|\|\mathbf{T}^{(k)}\|\|\tau' J'\rangle$	$(\gamma' j'\|\|\mathbf{T}(k)\|\|\gamma j)$	$(j'\|\|T_L\|\|j)$	$\langle J'\|\|T^{(k)}\|\|J\rangle$
Relation to others	See Eq. (30) in [48]	$= \frac{\sqrt{2j+1}}{(-1)^{2k}} R$	$= \frac{\sqrt{2J+1}}{(-)^{2k}} R$	$= \frac{(2j'+1)^{1/2}}{(-1)^{k-j+j'}} R$	$\equiv R$	$= R$

[a]This notation is an adaptation of the notation from [34].

[b]In place of || Condon and Shortley use a set of four vertical dots.

Applying Eqs. (3.17a) and then (3.16a) from [38] for the symmetry properties of the Clebsch-Gordan coefficients to $c^{F',1,F}_{m+q,-q',m+q-q'}$, we find

$$c^{F',1,F}_{m+q,-q',m+q-q'}(-1)^{F'-F}\sqrt{\frac{2F'+1}{2F+1}} = (-1)^{-q'} c^{F,1,F'}_{m+q-q',q',m+q}. \quad \text{(A5)}$$

With this, the expression for the spherical components of the atomic polarizability tensor components simplifies to

$$\hat{\alpha}_{q',q} = (-1)^{1-q'} \sum_{F',m} \left[\frac{1}{\hbar \Delta_{F'F}} c^{F,1,F'}_{m+q-q',q',m+q} c^{F,1,F}_{m,q,m+q} \times |\langle F'||d||F\rangle|^2 |F,m+q-q'\rangle\langle F,m| \right]. \quad \text{(A6)}$$

To express the reduced dipole matrix element in the coupled ($F = I + J$) basis in terms of the reduced dipole matrix element in the uncoupled (J) basis, we use Eq. (6.25) from [38] to obtain

$$\langle F'||d||F\rangle = \langle F', J', I||d||F, J, I\rangle = (-1)^{I+1-J-F}\sqrt{(2J'+1)(2F+1)} \times W(JFJ'F'; I1)\langle J'||d||J\rangle. \quad \text{(A7)}$$

Here, $W(JFJ'F'; I1)$ is the Racah W coefficient defined in [48]. Expressed in terms of the six-J symbol using Eq. (C.30) from [49], and using the symmetry relations for the six-J symbol (also given in [49]), it is

$$W(JFJ'F'; I1) = (-1)^{-J-F-F'-J'} \begin{Bmatrix} F' & I & J' \\ J & 1 & F \end{Bmatrix}. \quad \text{(A8)}$$

Plugging this into the expression for the polarizability tensor components, we obtain

$$\hat{\alpha}_{q',q} = (-1)^{1-q'} \sum_{F'} \left[\frac{1}{\hbar \Delta_{F'F}} |\langle J'||d||J\rangle|^2 \times (2J'+1)(2F+1) \begin{Bmatrix} F' & I & J' \\ J & 1 & F \end{Bmatrix}^2 \times \sum_m \left(c^{F,1,F'}_{m+q-q',q',m+q} c^{F,1,F}_{m,q,m+q} \times |F,m+q-q'\rangle\langle F,m| \right) \right]. \quad \text{(A9)}$$

With the definitions of the relative oscillator strength $f_{F'F}$ [see Eq. (5)] and the characteristic polarizability scalar $\alpha_{0,F'F}$ [Eq. (6), note that the negative sign is absorbed into this definition], and using $(-1)^{-q'} = (-1)^{q'}$ for integer values of q', this becomes

$$\hat{\alpha}_{q',q} = (-1)^{q'} \sum_{F'} \left[\alpha_{0,F'F} f_{F'F} \sum_m \left(c^{F,1,F'}_{m+q-q',q',m+q} c^{F,1,F}_{m,q,m+q} \times |F,m+q-q'\rangle\langle F,m| \right) \right], \quad \text{(A10)}$$

as shown in Eq. (4). This is identical to Eq. (6) from [34] except for a factor of $(-1)^{q'}$ instead of (-1), which must be taken into account in configurations with multiple beams at an angle, as there can be a π polarization component (i.e., $q' = 0$) to the electric field.

APPENDIX B: REDUCED DIPOLE MATRIX ELEMENT CONVENTIONS AND CONVERSIONS

When calculating the trapping potential energies in SI units (e.g., for comparison to experiment), care must be taken regarding normalization conventions and units for the reduced dipole matrix element. There are three common conventions for reduced matrix elements. The first convention is that used by Condon and Shortley [47]. When calculating the matrix element of a rank 1 tensor, Condon and Shortley write the factors of the Clebsch-Gordan coefficients that depend on the magnetic substates explicitly but absorb all other factors, including the angular-momentum-dependent factors of the

Clebsch-Gordan coefficients, into the reduced matrix element. This leads to the fact that the normalization factors are different for transitions between different angular momentum states, as shown in [48]. The convention adopted by Racah, as can be seen by combining his Eqs. (16′), (19a), and (29), factors out the Clebsch-Gordan coefficient as well as a factor of $\frac{1}{\sqrt{2J'+1}}$, where J' is the angular momentum of the final state, and a phase factor, or

$$\langle J'm'|T_q^{(k)}|Jm\rangle = (-1)^{2k}\frac{1}{\sqrt{2J'+1}}c_{m,q,m'}^{J,k,J'} \times \langle J'||T^{(k)}||J\rangle_{\text{Racah}}. \quad \text{(B1)}$$

The simplest convention is that adopted by Rose [38], where only the Clebsch-Gordan coefficient is factored out [see Eq. (5.14)], and all other factors are absorbed into the reduced matrix element, yielding

$$\langle J'm'|T_q^{(k)}|Jm\rangle = c_{m,q,m'}^{J,k,J'}\langle J'||T^{(k)}||J\rangle_{\text{Rose}}. \quad \text{(B2)}$$

Thus, the values of the reduced matrix elements are related by

$$\langle J'||T^{(k)}||J\rangle_{\text{Rose}} = (-1)^{2k}\frac{1}{\sqrt{2J'+1}} \times \langle J'||T^{(k)}||J\rangle_{\text{Racah}}. \quad \text{(B3)}$$

Table II summarizes the notations and normalizations employed by different authors.

In this work, we were interested in the reduced matrix elements for dipole transitions. The dipole operator (d) is a tensor of rank $k = 1$, which has three components $q = \pm 1, 0$. The allowed transitions are those with $J' = J+k, J, J-k$, and $m' = m+q$.

When calculating the atomic trapping potential energy using Eq. (6) for the atomic polarizability scalar, care must be taken when using reduced dipole matrix element values from the literature, due to the different normalization conventions as well as units. As this can be nontrivial based on the information given, we present an explicit example here. In an earlier version of our code that still employed Eq. (6) instead of Eq. (7), we used the reduced dipole matrix element from [51], which was given in atomic units, and used the Racah normalization. In order to use this value in our calculation, which uses the Rose normalization, we applied the following conversions:

$$|\langle J'||d||J\rangle_{\text{Rose}}|^2 = \frac{1}{2J'+1} \times |\langle J'||d||J\rangle_{\text{Safronova}}|^2 e^2 a_0^2. \quad \text{(B4)}$$

Here, $\langle J'||d||J\rangle_{\text{Safronova}}$ is the reduced dipole matrix element from [51] (in atomic units), e is the elementary charge, and a_0 is the Bohr radius. Alternatively, Eq. (7) can be used in the calculation, which only requires the laser wavelength of the transition and yields the polarizability scalar in the Rose convention, as needed for our calculation.

APPENDIX C: ELECTRIC FIELDS FOR SINGLE-BEAM AND TWO-BEAM CONFIGURATIONS

To calculate the electric fields for the configurations shown in Sec. III, we started from the electric field distributions determined through Hertz vector diffraction theory [40–42]. The Cartesian components of the electric field (real and imaginary parts) for diffraction of a laser beam incident on a circular aperture at an angle γ was calculated, once for an electric field polarization along the x direction in Figs. 1(a) and 3(a) (s polarization) and once for an electric field polarization perpendicular to the x direction and the direction of propagation (p polarization). The tilted beams with p polarization had both y and z components, although the z components were small since we used only small angles in this study. Each electric field calculation was normalized such that the electric field components were fully extended at time $t = 0$ at the aperture plane, with an electric field magnitude of 1 for the s polarization and also a magnitude of 1 for the p polarization. Six data files were generated for each beam configuration: $\text{Re}(E_{0x})$, $\text{Im}(E_{0x})$, $\text{Re}(E_{0y})$, $\text{Im}(E_{0y})$, $\text{Re}(E_{0z})$, and $\text{Im}(E_{0z})$.

The electric field components were then read into a MATHEMATICA code [44]. At this point, we generated the beam configurations from Sec. III as follows.

For the single-beam configuration, we used a σ^+ polarized beam incident at an angle of $+\gamma$ [see Fig. 1(a)], which we generated by adding an s-polarization component and a p-polarization component, which was lagging behind by a phase of 90°:

$$\mathbf{E}_0 = \frac{1}{\sqrt{2}}[E_{sx}\mathbf{e}_x + i(E_{py}\mathbf{e}_y + E_{pz}\mathbf{e}_z)]. \quad \text{(C1)}$$

Here, E_{sx} is the s-polarization component, and E_{py} and E_{pz} are the p-polarization components along y and z, respectively. Note that all three components are complex, that is, $E_j = \text{Re}(E_j) + i\,\text{Im}(E_j)$, where $j = sx, py, pz$. The factor of $\frac{1}{\sqrt{2}}$ is to normalize E_0 to 1 at the aperture.

Using Eq. (14) we get the following expressions for the spherical components of the electric field:

$$E_{0-1} = \frac{1}{2}(E_{sx} - E_{py}),$$
$$E_{00} = \frac{i}{\sqrt{2}}E_{pz}, \quad \text{(C2)}$$
$$E_{0+1} = -\frac{1}{2}(E_{sx} + E_{py}).$$

For the two-beam configuration, we added a σ^- beam along the $-\gamma$ direction [see Fig. 3(a)]. This can be achieved either through proper inversion of the s-polarization and p-polarization arrays used for the σ^+ beam or by generating an array specifically for the negative angle. We chose the latter method. This time, the p-polarization component must be ahead of the s-polarization component by a phase of 90°. For this case, we find

$$E_{0-1} = \frac{1}{2}(E_{sx} + E_{py}),$$
$$E_{00} = -\frac{i}{\sqrt{2}}E_{pz}, \quad \text{(C3)}$$
$$E_{0+1} = \frac{1}{2}(-E_{sx} + E_{py}).$$

The total spherical electric field components were the sums of the corresponding electric field components for the σ^+ beam and the σ^- beam.

Once the spherical components of the electric field are determined, they can be plugged into Eq. (11).

[1] G. K. Brennen, C. M. Caves, P. S. Jessen, and I. H. Deutsch, Phys. Rev. Lett. **82**, 1060 (1999).

[2] G. K. Brennen, I. H. Deutsch, and P. S. Jessen, Phys. Rev. A **61**, 062309 (2000).

[3] I. H. Deutsch, G. K. Brennen, and P. S. Jessen, Fortschr. Phys. **48**, 925 (2000).

[4] G. J. Milburn, Fortschr. Phys. **48**, 957 (2000).

[5] D. P. DiVincenzo, Fortschr. Phys. **48**, 771 (2000).

[6] K. Eckert, J. Mompart, X. X. Yi, J. Schliemann, D. Bruß, G. Birkl, and M. Lewenstein, Phys. Rev. A **66**, 042317 (2002).

[7] D. Meschede and A. Rauschenbeutel, Adv. At. Mol. Opt. Phys. **53**, 75 (2006).

[8] M. Karski, L. Förster, J. M. Choi, W. Alt, A. Widera, and D. Meschede, Phys. Rev. Lett. **102**, 053001 (2009).

[9] T. Wilk, A. Gaëtan, C. Evellin, J. Wolters, Y. Miroshnychenko, P. Grangier, and A. Browaeys, Phys. Rev. Lett. **104**, 010502 (2010).

[10] L. Isenhower, E. Urban, X. L. Zhang, A. T. Gill, T. Henage, T. A. Johnson, T. G. Walker, and M. Saffman, Phys. Rev. Lett. **104**, 010503 (2010).

[11] M. Anderlini, P. J. Lee, B. L. Brown, J. Sebby-Strabley, W. D. Phillips, J. V. Porto, Nature (London) **448**, 452 (2007).

[12] R. Dumke, M. Volk, T. Müther, F. B. J. Buchkremer, G. Birkl, and W. Ertmer, Phys. Rev. Lett. **89**, 097903 (2002).

[13] J. Kruse, C. Gierl, M. Schlosser, G. Birkl, Phys. Rev. A **81**, 060308(R), (2010).

[14] S. Bergamini, B. Darquié, M. Jones, L. Jacubowiez, A. Browaeys, and P. Grangier, J. Opt. Soc. Am. B **21**, 1889 (2004).

[15] D. McGloin, G. C. Spalding, H. Melville, W. Sibbett, and K. Dholakia, Opt. Express **11**, 158 (2002).

[16] C. García-Segundo, H. Yan, and M. S. Zhan, Phys. Rev. A **75**, 030902 (R) (2007).

[17] J. Goldwin and E. A. Hinds, Opt. Express **16**, 17808 (2008).

[18] O. Alloschery, R. Mathevet, and J. Weiner, Opt. Express **14**, 12568 (2006).

[19] V. Yannopapas, N. V. Vitanov, J. Phys. Condens. Matter **21**, 245901 (2009).

[20] T. N. Bandi, V. G. Minogin, and S. N. Chormaic, Phys. Rev. A **78**, 013410 (2008).

[21] L. Chen and J. Yin, Phys. Rev. A **80**, 065401 (2009).

[22] K. Christandl, G. P. Lafyatis, S.-C. Lee, and J.-F. Lee, Phys. Rev. A **70**, 032302 (2004).

[23] X. Ji and J. Yin, J. Opt. Soc. Am. B **22**, 1737 (2005).

[24] R. Mu, J. Lu, S. Xu, X. Ji, and J. Yin, J. Opt. Soc. Am. B **26**, 80 (2009).

[25] V. I. Balykin, V. V. Klimov, and V. S. Letokhov, J. Phys. II (France) **4**, 1981 (1994).

[26] V. V. Klimov and V. S. Letokhov, J. Mod. Opt. **42**, 1485 (1995)

[27] V. V. Klimov and V. S. Letokhov, Opt. Commun. **121**, 130 (1995)

[28] V. I. Balykin, V. V. Klimov, and V. S. Letokhov, JETP Lett. **78**, 8 (2003)

[29] V. I. Balykin, V. V. Klimov, and V. S. Letokhov, Opt. Photonics News **16**, 44 (2005).

[30] V. V. Klimov, S. K. Sekatskii, and G. Dietler, Opt. Commun. **259**, 883 (2006)

[31] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, Nature (London) **465**, 74 (2009).

[32] G. D. Gillen and S. Guha, Am. J. Phys. **72**, 1195 (2004).

[33] G. D. Gillen, S. Guha, and K. Christandl, Phys. Rev. A **73**, 013409 (2006).

[34] I. H. Deutsch and P. S. Jessen, Phys. Rev. A **57**, 1972 (1998).

[35] R. Grimm, M. Weidemüller, and Y. B. Ovchinnikov, Adv. At. Mol. Opt. Phys. **42**, 95 (2000).

[36] I. H. Deutsch and P. S. Jessen, Opt. Commun. **283**, 681 (2010).

[37] J. M. Geremia, J. K. Stockton, and H. Mabuchi, Phys. Rev. A **73**, 042112 (2006).

[38] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, New York, 1957), p. 88.

[39] P. S. Jessen and I. H. Deutsch, Adv. At. Mol. Opt. Phys. **37**, 95 (1996).

[40] G. Bekefi, J. Appl. Phys. **24**, 1123 (1953).

[41] S. Guha and G. D. Gillen, Opt. Express **13**, 1424 (2005).

[42] For information regarding the MATHCAD code used to calculate the electric field distribution behind a circular aperture using Hertz vector diffraction, contact Glen D. Gillen, ggillen@calpoly.edu.

[43] H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping* (Springer-Verlag, New York, 1999).

[44] For information regarding the MATHEMATICA code used to calculate the potential energy from the electric field using Eq. (11), contact Katharina Gillen-Christandl.

[45] K. Christandl, Ph.D. thesis, The Ohio State University, 2005.

[46] O. Mandel, M. Greiner, A. Widera, T. Rom, T. W. Hänsch, and I. Bloch, Phys. Rev. Lett. **91**, 010407 (2003).

[47] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, New York, 1935), p. 63.

[48] G. Racah, Phys. Rev. **62**, 438 (1942).

[49] A. Messiah, *Quantum Mechanics* (Dover, Mineola, NY, 1999), p. 573.

[50] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1996), p. 75.

[51] M. S. Safronova, C. J. Williams, and C. W. Clark, Phys. Rev. A **69**, 022509 (2004).

USING MONOCULAR VISION AND IMAGE CORRELATION TO ACCOMPLISH AUTONOMOUS LOCALIZATION

A Thesis
Presented to the
Faculty of California Polytechnic State University
San Luis Obispo

In Partial Fulfillment
of the Requirements for the Degree
Masters of Science in Computer Science

By
Matthew P. Schlachtman
June 2010

COMMITTEEE MEMBERSHIP

TITLE: Monocular Vision and Image Correlation to Accomplish Autonomous Localization

AUTHOR: Matthew P. Schlachtman

DATE SUBMITTED: June 2010

COMMITTEE CHAIR: Christopher M. Clark, Ph.D.

COMMITTEE MEMBER: Franz Kurfess, Ph.D.

COMMITTEE MEMBER: Saeed Niku, Ph.D.

Abstract

Monocular Vision and Image Correlation to Accomplish Autonomous Localization

Matthew P. Schlachtman

For autonomous navigation, robots and vehicles must have accurate estimates of their current state (i.e. location and orientation) within an inertial coordinate frame. If a map is given a priori, the process of determining this state is known as localization. When operating in the outdoors, localization is often assumed to be a solved problem when GPS measurements are available. However, in urban canyons and other areas where GPS accuracy is decreased, additional techniques with other sensors and filtering are required.

This thesis aims to provide one such technique based on monocular vision. First, the system requires a map be generated, which consists of a set of geo-referenced video images. This map is generated offline before autonomous navigation is required. When an autonomous vehicle is later deployed, it will be equipped with an on-board camera. As the vehicle moves and obtains images, it will be able to compare its current images with images from the pre-generated map. To conduct this comparison, a method known as image correlation, developed at Johns Hopkins University by Rob Thompson, Daniel Gianola and Christopher Eberl, is used. The output from this comparison is used within a particle filter to provide an estimate of vehicle location. Experimentation demonstrates the particle filter's ability to successfully localize the vehicle within a small map that consists of a short section of road. Notably, no initial assumption of vehicle location within this map is required.

Contents

List of Figures vii

1 INTRODUCTION 1

2 BACKGROUND 2

3 PROBLEM DEFINITION & SOLUTION 9

3.1 Map Creation 9

3.1.1 Video Capture 11

3.1.2 Grey Scale 11

3.1.3 Edge Detection 12

3.1.5 Map Image Selection 13

3.2 Localization 14

3.2.1 Video Capture 15

3.2.2 Grey Scale 16

3.2.3 Edge Detection 16

3.2.4 Correlation 16

3.2.6 Map Image Match Determination 16

3.2.8 Particle Filter 17

4 IMPLEMENTATION 19

4.1 Implementation of Map Creation 19

4.1.1 Video Capture 19

4.1.2 Video Editing 21

4.1.3 Grey Scale & Edge Detection 22

4.1.4 Converting Video Format 24

4.1.5 Image Extraction 25

4.1.6 Map Image Selection 25

4.2 Localization 26

4.2.1 Video Capture 27

4.2.2 Video Editing 27

4.2.3 Grey Scale & Edge Detection 28

4.2.4 Converting Video Format 28

4.2.5 Image Extraction 28

4.2.6 Image Correlation 29

4.2.7 Map Image Match Determination 34

4.2.8 Particle Filter 35

5 RESULTS **36**

6 CONCLUSION **41**

7 FUTURE WORK & POTENTIAL APPLICATIONS **42**

Bibliography **43**

Appendix A: 'smooth_try3.m' **45**

Appendix B: 'IP_PF_0.m' **46**

List of Figures

Figure 1: City Vs. Street Map 10
Figure 2: Map Creation block diagram 10
Figure 3: Top down view of camera mounting configuration 11
Figure 4: Localization block diagram 15
Figure 5: Camera Mounted 20
Figure 6: Route Map 21
Figure 7: Trim Tool 22
Figure 8: Grey scale Effect 23
Figure 9: Canny Edge Detection 23
Figure 10: WinAVI Convert WMV to MPEG-2 24
Figure 11: IrfanView used to extract all frames 25
Figure 12: Geo-Located Image Locations 26
Figure 13: Geo-Located Image Four 26
Figure 14: Google Street View Map Image 4 26
Figure 15: Start of Driving Run 27
Figure 16: Visual Compare, 3rd Geo-located frame on left, Frame 587 from Localization Video on right 28
Figure 17: Pointing to the first image in a series 29
Figure 18: Saving file list 30
Figure 19: Select grid type 31
Figure 20: Grid Selection 31
Figure 21: Grid Resolution 32
Figure 22: Automate Image Running 33
Figure 23: Visualize Data Menu 34
Figure 24: Image Correlation Strain 36
Figure 25: *K(t)* results for each geo-located image 37
Figure 26: Initial State of Particle Filter 38
Figure 27: First Grouping at $t = 95$ 38
Figure 28: A bit of random dispersion at $t = 955$ 39
Figure 29: Particle Filter Right after a good set of matches at $t = 1055$ 39
Figure 30: Error Graph 40

Chapter 1

INTRODUCTION

One of the design objectives of computer vision localization is to provide a low cost method for outdoor localization using a single camera. [3] This relaxes the need for global positioning system (GPS), which may experience degraded reliability in urban settings like in downtown settings with more tall buildings that could block the signals from the GPS satellites from getting to the GPS receiver. It is possible to imagine TomTom or Garmin, both commercial GPS manufacturers, implementing a small camera on the GPS system that could be looking out in front of the car and perform the kind of visual localization that has been discussed. Moreover, this type of localization could be useful in the operation of future autonomous vehicles.

The goal of this thesis is to demonstrate that by using a combination of existing technology, autonomous localization can be accomplished with a single camera obtaining images of the view outside of the car. This thesis intends on demonstrating that through the use of image processing, edge detection, and image correlation and comparison, one image can be compared with another, image from a database of geo-located images to determine the location of a vehicle. While only a proof of concept, the idea could be expanded on and can be adapted to serve real world purposes and potentiality a profitable idea or product.

Chapter 2

BACKGROUND

Localization is the aet of determining the loeation of an (autonomous) agent within a map given a priori. In almost every type of mobile robot system loealization is neeessary. Regardless of the size of the robot, the system will most likely need to know where it or something else is. Various aspeets to roboties eount on this ability, ineluding but not limited to, robotic navigation. Sueeess in navigation requires sueeess at the four building blocks of navigation:

- Pereeption - the robot must interpret its sensors to extraet meaningful data from its environment
- Loealization - the robot must determine its position in the environment, or the position of some element of interest in the environment.
- Cognition - the robot must deeide how to aet to aehieve its goals.
- Motion Control - to robot must modulate its motor outputs or other aetuators to aehieve the desired trajeetory [2].

Autonomous vehiele localization can be aeeomplished with various different hardware and software components. There are several different families of hardware that ean be used to aeeomplish localization. There are also algorithms to faeilitate the loealization using the various localization hardware types [1]. Some of the most popular are Markov Localization, Partiele Filter Localization, and Kalman Filter Localization [1].

Markov loealization uses an explieitly speeified probability distribution across all possible robot positions [2]. Instead of maintaining a single hypothesis as to where in the world a robot might be, Markov loealization maintains a probability distribution over the spaee of all sueh hypotheses. The probabilistie representation allows it to weigh these different hypotheses in a mathematieally sound way [10]. For the sake of simplieity, let's assume that the spaee of robot positions is one-dimensional, that is, the robot ean only move horizontally (it may not rotate). Now suppose the robot is plaeed somewhere in this environment, but it is not told its loeation.

Markov localization represents this state of uncertainty by a uniform distribution over all positions. Now let's assume the robot queries its sensors and finds out that it is next to a door. Markov localization can modify the belief by raising the probability for places next to doors, and lowering it anywhere else. Now let's assume the robot moves a meter forward. Markov localization incorporates this information by shifting the belief distribution accordingly. To account for the inherent noise in robot motion, which inevitably leads to a loss of information, the new belief more evenly distributes likelihoods across all robot states (and less certain) than the previous one. Finally, let's assume the robot senses a second time, and again it detects another door. At this point, the belief is updated to again increase likelihoods of all states next to the doors [10] [11].

The particle filter is an implementation of the Baye's filter using a finite number of particles in the continuous state space to describe the belief state probability density distribution. The algorithm propagates particles through time using the *survival of the fittest* concept [3]. The main objective of particle filtering is to track a variable or interest as it evolves over time, typically with a non-Gaussian and potentially multi-modal probability density function. The basics of the method are to construct a sample-based representation of the entire probability density function. A series of actions are taken, each one modifying the state of the variable of interest according to some model. Moreover, certain observations can constrain the state of the variable of interest at that time [9]. Multiple particles of the variable of interest are used, each one with an assigned weight that signifies the likelihood of that specific particle. An estimate of the variable of interest can be obtained by the weighted sum of all of the particles. Like most localizations algorithms, the particle filter algorithm is recursive in nature and operates in two phases; prediction and correction. After each action, each particle is modified according to the existing model (prediction stage), including the addition of random noise in order to simulate the uncertainty in the model. Then, each of the particle's weight is evaluated based on the latest sensory information available (update/correction stage). Particles with high weights have a higher likelihood of remaining when the resampling occurs [9].

Kalman Filter Localization, another commonly used approaeh, uses a Gaussian probability density representation of robot position [2]. The Kalman filter itcrates over the consecutive cycles of prediction and correction. The dynamics of these cycles is dcrived and interpreted in the framework of Gaussian probability density functions. Under additional conditions on the system dynamics, the Kalman filter dynamics converge to a steady-state filter and the steady-state gain is derived. The innovation process associated with the filter, that represents the novel information conveyed to the state estimate by the last system measurement, is introduced. The states estimates are interpreted in terms of the covariance associated with thc Gaussian probability density function involved in the filter dynamics [12].

All of the localization algorithms, including those mentioned above, are able to be set up in such a way that they can be adapted with various motion and sensor modcls to correctly implement any particular mode of locomotion and sensing. Pieking the algorithm is usually based on thc actual hardware, system, and environment. Once one knows the requircments for a project the optimal algorithm ean be ehosen [1].

Sonar (originally an aeronym for sound navigation and ranging) is a tcchniquc that uscs sound propagation (usually underwater) to navigate, communicatc with or detccts other vessels. There are two kinds of sonar: aetive and passive. Sonar may be used as a means of acoustic loeation and of measurement of the echo charaeteristies of "targets" in the water. Aeoustie location in air was used before the introduetion of radar. Sonar may also be used in air for robot navigation, and SODAR (an upward looking in-air sonar) is used for atmospheric investigations. The term sonar is also used for the equipment used to gencrate and receive thc sound. The frequencies used in sonar systems vary from infrasonic to ultrasonic. The study of underwater sound is known as underwater acoustics or sometimes hydro acoustics [13]. Sonar is a good source for localization but usually in smaller environments and can be interfered with by many things that emit sound in nature. Sonar works best for localization when applied to underwater robots. In fact, the VideoRay ROV robot uses sonar and this robot has bcen used quite a bit to ereate maps and perform SLAM algorithms (Simultaneous Localization and Mapping) [5]. As well this technology is thc primary sensor for the Malta Cistern Mapping project founded by Dr. Christopher Clark [14].

Infrared Range sensors are commonly found on smaller robots that are used for exploring places like buildings, caves, and other such relatively small environments. These sensors work on the premise that invisible (to humans) light is emitted by a LED, bounces off a surface, and is perceived by Infrared photocells. The time that it takes for the light to travel, bounce off of something, and travel back to the photocell can be used to calculate the distance that the light hit something at. This type of sensing technology is likely to be the least expensive but is also very unreliable. The technology is subject to errors caused by simple room lights, the sun, and various electronic devices. The problem is that the photocell may not be able to differentiate between the light the LED emits and various other sources in the environment [2].

A laser rangefinder is a device that uses a laser beam in order to determine the distance to a reflective object. The most common form of laser rangefinder operates on the time of flight principle by sending a laser pulse in a narrow beam towards the object and measuring the time taken by the pulse to be reflected off the target and returned to the sender. Due to the high speed of light, this technique is not appropriate for high precision sub-millimeter measurements, where triangulation and other techniques are often used [13]. Although Laser Scanners are very accurate; their cost is very high. Even the simplest models go for thousands of dollars [15]. As well, the technology is not really the best sensor for mapping, and its localization ability is based of past data or some type of three-dimensional map [1].

GPS (Global Positioning System) works by having a network of twenty-four satellites in orbit around earth. Global Positioning System satellites transmit signals to equipment on the ground. GPS receivers passively receive satellite signals; they do not transmit. GPS receivers require an unobstructed view of the sky, so they are used only outdoors and they often do not perform well within forested areas or near tall buildings. GPS operations depend on a very accurate time reference, which is provided by atomic clocks at the U.S. Naval Observatory. Each GPS satellite has atomic clocks on board [16]. Each GPS satellite transmits data that indicates its location and the current time. All GPS satellites synchronize operations so that these repeating signals are transmitted at the same instant. The signals, moving at the speed of light, arrive at a GPS receiver at slightly different times because some satellites are farther away than others. The

distance to the GPS satellites can be determined by estimating the amount of time it takes for their signals to reach the receiver. When the receiver estimates the distance to at least four GPS satellites, it can calculate its position in three dimensions [16].

In current technology, GPS is seen as being one of the best ways to localize one's position [2]. With accuracy down to a few meters, and costs in the 100 to 200 dollar range it is no wonder it's used so heavily. The accuracy of a position determined with GPS depends on the type of receiver. Most hand-held GPS units have about ten to twenty meter accuracy. Other types of receivers use a method called Differential GPS (DGPS) to obtain much higher accuracy. DGPS requires an additional receiver fixed at a known location nearby. Observations made by the stationary receiver are used to correct positions recorded by the roving units, producing accuracy greater than 1 meter [16]. When the system was created, timing errors were inserted into GPS transmissions to limit the accuracy of non-military GPS receivers to about 100 meters. This part of GPS operations, called Selective Availability, was eliminated in May 2000. If one could attach an accurate GPS sensor to a mobile robot, much of the localization problem would be obviated [16].

The GPS would inform the robot of its exact position, so the answer to the question "Where Am I?" would always be available immediately. Unfortunately, such a sensor is not always practical. The existing GPS network provides accuracy to within several meters, which may be unacceptable for localizing on much smaller areas. Furthermore, GPS technology cannot function indoors or in obstructed areas and are thus limited in their workspace. [2] In large city areas there can sometimes be issues when localizing using GPS due to the fact that the GPS signals can be impeded or blocked completely by buildings that get in the way of the line of sight which is needed for GPS to function properly. Hypothetically speaking, an alternate sensor payload could be used in conjunction with GPS to help with the localization when GPS signals become unreliable or nonexistent.

There are other solutions besides GPS that are very useful and have various benefits that GPS can't match. Various small solutions include sonar and infrared distance sensors, cameras, and even laser range finders and scanners [4]. The intended are of focus with the aforementioned

area studies, is the use of monocular computer vision. This type of vision would be a great alternative for localization when convention GPS has failed due to any number of circumstances. Other researchers have also used monocular vision in localization because of the simplicity of hardware involvement.

It is possible that computer vision could be a good alternative or supplement to GPS. In recent publications, researchers have tried using a database of images to serve as the map in localization. Zhang et al. [4] captured images at various points in an operating workspace and tagged them with GPS position readings to create an image database [3]. Similar work has also been done by looking at the details of building facades. In the most related and recent work presented in [7], a robot is first guided through a course as it records a video of the surrounding. This information is used offline where distinct image features are selected to generate a three dimensional map [3].

Using computer vision and edge detection a reference can be made to detect location based on images or edges detected compared to another image and set of edges in a stored database. The first step in obtaining a feature map is the processing of raw image data using edge detector to create an edge map which highlights the pixels that are likely to be part of building boundaries [3]. A comparison of the edges detected to pre processed images yields a probability distribution that would be associated with all possible locations the robot could be in within the environment. [2].

The orientations of observed building walls need to be interpreted from camera images. This information can be recovered from the effect of perspective using vanishing point analysis [7]. A benefit of using vanishing point analysis and selecting building orientation as a measure for comparison is that it makes the system tolerant to significant camera rolling and tilting. Edge detection is used to extract useful features from an image scene [3].

There exists a free library for computer vision known as OpenCV, which has the ability to implement the computer vision procedures mentioned above. OpenCV is a computer vision library originally developed by Intel. It is free for commercial and research use under a BSD

license. The library is cross-platform, and runs on Windows, Mac OS X, Linux, PSP, VCRT (Real-Time OS on Smart camera) and other embedded devices. It focuses mainly on real-time image processing, as such, if it finds Intel's Integrated Performance Primitives on the system, it will use these commercial optimized routines to accelerate itself [13][17].

Many computer vision techniques have also been programmed into the math simulator known as Matlab. MATLAB stands for "MATrix LABoratory" and is a numerical computing environment and fourth-generation programming language. Developed by The MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, and FORTRAN [18]. Unlike the OpenCV libray, the matlab image processing and computer vision libraries enable a user to focus more on the theory of the project and less on the computer programming and syntax of the language, as using MATLAB is a very simple tool for performing simulations, and requires less programming skill. While MATLAB is not primarily used for computer vision applications, in more recent builds, MATLAB has been including various computer vision and image processing libraries allowing the necessary mathematical operation to function over videos and images [19].

There is a MATLAB library called *Digital Image Correlation and Tracking* posted by Christopher Eberl, which was of particular interest for the purpose of this paper as it is the primary method of which image correlation is accomplished in the experiments talked about in the latter sections. The Library's given description is that is calculates displacement and strain from a series of images. Strain can be calculated in horizontal as well as vertical direction. The functions were developed at the Johns Hopkins University by Rob Thompson, Daniel Gianola and Christopher Eberl. The code is a continued work in progress and was originally released in September of 2006, and was most recently updated in September of 2008. The code was specifically designed to either produce the calculations necessary for image correlation, or to assist in a form of feature tracking. For the cases of this thesis, correlation section of the library will be used.

Chapter 3

PROBLEM DEFINITION & SOLUTION

The problem addressed in this thesis is defined as:

Given a previously recorded video set parameterized within a geolocalized coordinate system, determine the geo-referenced location of a vehicle using only an on-board monocular vision system.

The proposed solution to this problem can be broken into two basic components: map creation and localization. This section sets the scene for building the low level detail upon how each one of the tasks should be accomplished. The implementation chapter on the other hand, goes into the technical details of the experiment.

3.1 Map Creation

A map must be created with geo-located points that correspond to specific frames in a pre-recorded video such that:

$$V_m \Rightarrow M \tag{1}$$

$$M = \{[I_{m1}, x_{m1}, y_{m1}], [I_{m2}, x_{m2}, y_{m2}], \ldots, [I_{mn}, x_{mn}, y_{mn}]\} \tag{2}$$

The above equations symbolize the goal of the map creation part of the Solution. In equation (1) and (2), M refers to the map itself, V_m refers to the prerecorded video. I_{m1} refers to a geo-located image frame from the video such that x_{m1} and y_{m1} represent the corrdinates on the map. The following flow chart explains the necessary steps that need to be taken to construct the map.

Figure 1: City Vs. Street Map

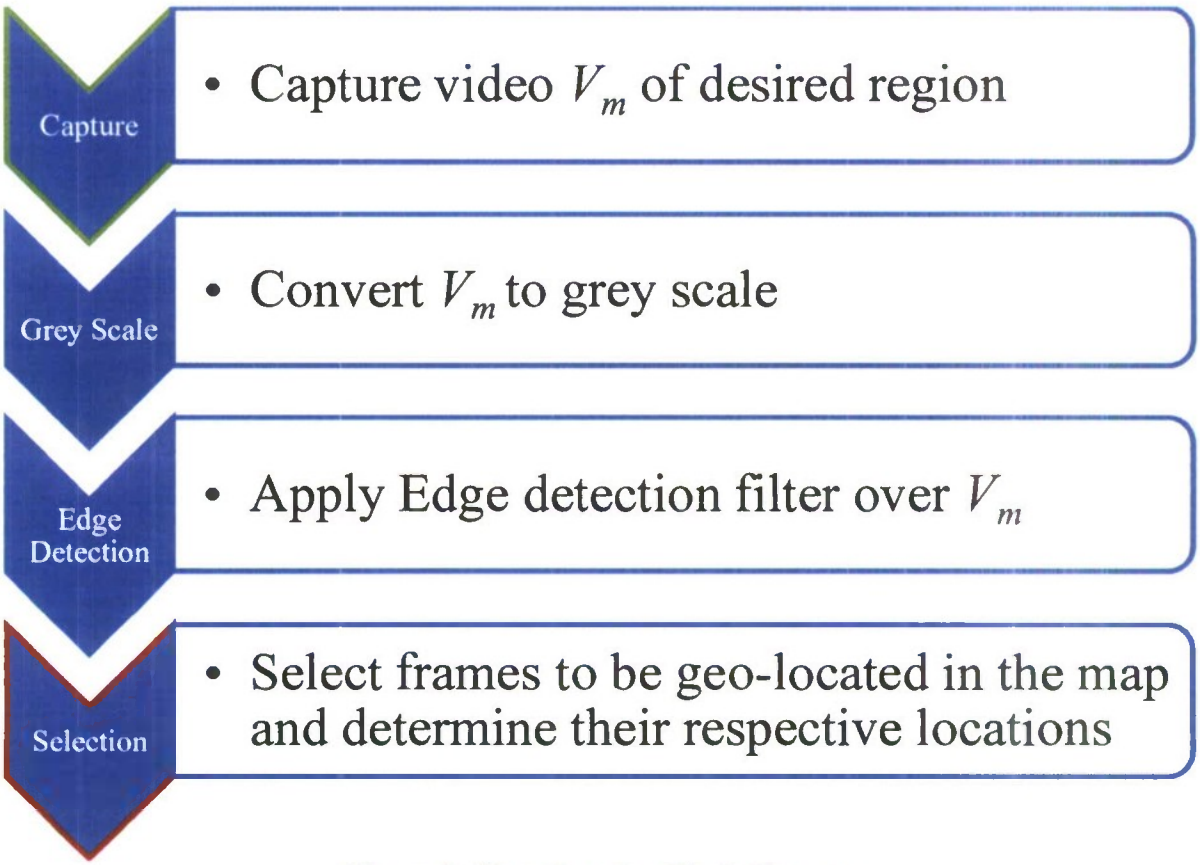


Figure 2: Map Creation block diagram

The above block diagram (figure 2) demonstrates how the map creation process is split into four phases: Capture, Grey Scale, Edge Detection, and Selection.

3.1.1 Video Capture

The first step in constructing a map is obtaining a video V_m. Video is recorded while the vehicle is driven down a length of road. Camera configuration should capture the view outside of the passenger side window as shown in Figure 3 below.

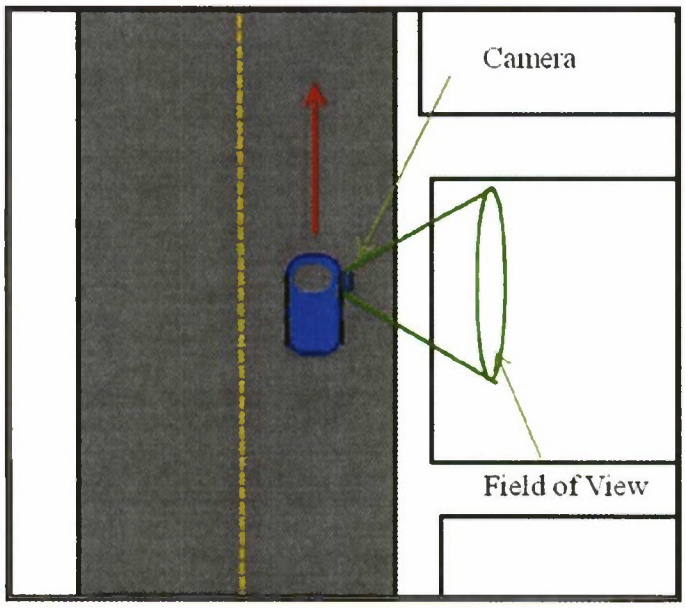


Figure 3: Top down view of camera mounting configuration

3.1.2 Grey Scale

Computer vision and/or video processing tools are used to convert the videos to grey scale. Conversion of a color image to grey scale is not unique; different weighting of the color channels effectively represents the effect of shooting black-and-white film with different-colored photographic filters on the cameras. A common strategy is to match the luminance of the grey scale image to the luminance of the color image. To convert any color to a grey scale representation of its luminance, first one must obtain the values of its red, green, and blue (RGB) primaries in linear intensity encoding, by gamma expansion. Then, add together 30% of the red

value, 59% of the green value, and 11% of the blue value (these weights depend on the exact choice of the RGB primaries, but are typical), as seen in (3). The resultant number is the desired linear luminance value which is applied to a white to black scale [22].

$$g_{ij} = 0.30R_{ij} + 0.59G_{ij} + 0.11B_{ij} \text{ for } i = 0 \ldots \text{Image Width}, \quad j = 1 \ldots \text{Image Height} \tag{3}$$

3.1.3 Edge Detection

Using computer vision and/or video processing tools a Canny edge detection filter is applied over all of the frames of the video. The grey scale is applied first as it is typically easier to apply an edge detection filter when dealing with a single intensity value (i.e. Black). The Canny edge detector uses a filter based on the first derivative of a Gaussian, because it is susceptible to noise present on raw unprocessed image data. First, the raw image is convolved with a Gaussian filter. The result is a slightly blurred version of the original which is ideally not affected by a single noisy pixel to any significant degree. Secondly, an edge in an image may point in a variety of directions, so the Canny algorithm uses four filters to detect horizontal, vertical and diagonal edges in the blurred image. The edge detection operator (Roberts, Prewitt, Sobel for example) returns a value for the first derivative in the horizontal direction (G_y) and the vertical direction (G_x), as shown in (4) and (5).

$$G = sqrt(G_x^2 + G_y^2) \tag{4}$$

$$\Theta = \arctan(G_x / G_y) \tag{5}$$

The edge direction angle is rounded to one of four angles representing vertical, horizontal and the two diagonals (0, 45, 90 and 135 degrees for example). Given estimates of the image gradients, a search is then carried out to determine if the gradient magnitude assumes a local maximum in the gradient direction. So, for example,

- if the rounded angle is zero degrees the point will be considered to be on the edge if its intensity is greater than the intensities in the **north and south** directions,
- if the rounded angle is 90 degrees the point will be considered to be on the edge if its intensity is greater than the intensities in the **west and east** directions,
- if the rounded angle is 135 degrees the point will be considered to be on the edge if its intensity is greater than the intensities in the **north east and south west** directions,
- if the rounded angle is 45 degrees the point will be considered to be on the edge if its intensity is greater than the intensities in the **north west and south east** directions.

This stage referred to as non-maximum suppression, produces a set of edge points in the form of a binary image.

Thresholding with hysteresis requires two thresholds – high and low. Making the assumption that important edges should be along continuous curves in the image allows us to follow a faint section of a given line and to discard a few noisy pixels that do not constitute a line but have produced large gradients. Therefore we begin by applying a high threshold. This marks out the edges we can be fairly sure are genuine. Starting from these, using the directional information derived earlier, edges can be traced through the image. While tracing an edge, we apply the lower threshold, allowing us to trace faint sections of edges as long as we find a starting point. Once this process is complete we have a binary image where each pixel is marked as either an edge pixel or a non-edge pixel. The edge detection algorithm is used as it is time efficient to implement and though personal experience has proven to revel distinctive features from the geo-located frames that will be selected later in section 3.1.4.

3.1.4 Map Image Selection

Using specific criteria, frames are selected from the video that will be used as the geo-located frames in the localization part of the solution. Such criteria include:

- Maximizing detail of a structure (e.g. houses)
- Minimizing trees
- Minimizing dynamic elements (e.g. cars, people, etc.)

The coordinates are then found on the map that correlate to the aforementioned selected frames, then they were label on the map to complete its construction.

3.2 Localization

After the map has been created, one can localize the position of the vehicle by processing the video data that is recorded as the vehicle moves along the path. For every image frame taken at time step t from an on board camera, the correlation function is calculated:

$$k_i(t) = IC(I_t, I_{mi}) \text{ for } i = 1 \ldots n \tag{6}$$

Where $k_i(t)$ the correlation between image I_t, at the current frame at time step t provided by the vehicle camera, and image I_{mi} the ith geo-located selected image frame from the map, n represents the number of geo-located points in the map. Once all of these values have been calculated the values can be stored such that:

$$K(t) = [k_1(t), k_2(t), k_3(t), \ldots, k_n(t)] \tag{7}$$

$K(t)$ is then used in the map image match determination to create $H(t)$ as defined such that:

$$H(t) = [h_1(t), h_2(t), h_3(t), \ldots, h_n(t)] \quad , \qquad h_i(t) \in [0,1] \tag{8}$$

Where $h_x(t)$, represents if a match has been determined with the x^{th} geo-located, at time t. $H(t)$ is needed for the particle filter to localize the position of the vehicle. The following flow chart explains the necessary steps that need to be taken to accomplishing the localization portion of the solution.

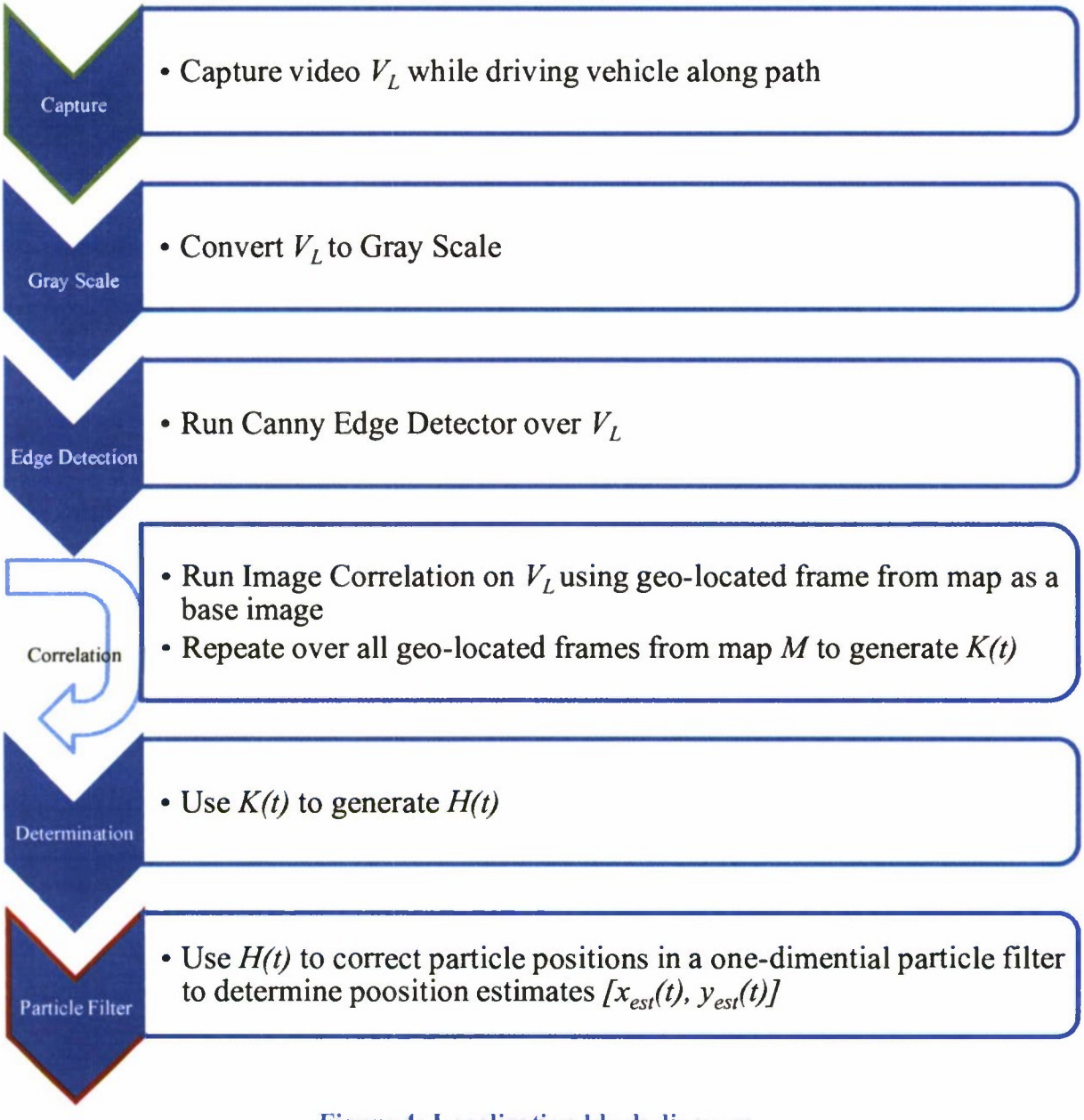


Figure 4: Localization block diagram

The above block diagram demonstrates how the localization process is split into six phases; Capture, Grey Scale, Edge Detection, Correlation, Determination, and Particle Filter.

3.2.1 Video Capture

Images must be captured as the vehicle is driven down a specific length of road. Similar to configuration as in section 3.1.1

3.2.2 Grey Scale

Using computer vision and/or video processing tools, the video gets converted to grey scale. See details from section 3.1.2.

3.2.3 Edge Detection

Using eomputer vision and/or video processing tools, a Canny edge detection filter is applied over all of the frames of the video. See details from section 3.1.3.

3.2.4 Correlation

A combination of the image correlation functions provided in the Image Correlation and Tracking Library [20], were used to calculate the strain of features within the frames over time. The correlation process uses discretized grid sections to measure how similar or different two images are. Onee an image has been discretized each grid cell is traeked for properties that would eonstitute a mateh. The properties are the number of line sections and relative position of features. The differenees to an assumed fit(s) of the aforementioned properties are measured as strain along a single direction (x-axis). The movement of the grid sections is averaged and a line is fitted to the different positions. When the slope of this fitted line is elose to zero it reflects an image frame in which there is minimal displacement (or differences) between the base image and the image being compared. The result is the average strain slope over each one of the frame pairs being correlated. This data, $K(t) = K(t) = [k_1(t), \ldots , k_n(t)]$, is used by the next phase (section 3.2.5).

3.2.5 Map Image Match Determination

After obtaining strain $K(t)$ from the image eorrelation proeess, it must be proeessed to determine matehes $H(t)$ between current frames and map frames. This data is then used as input data for the particle filter explained in section 3.2.6. For each map image strain calculated, there

is random fluctuation due to the fact that the geo-located images, and the frames from the localization data set, are not identical. The data is smoothed and the derivative taken. There exists plateaus in the strain time plots and the goal is to try and find these plateaus as part of the data image match determination. Each one of these plateaus occurs at (or around) where there is a match in the image correlation process. This occurs because the change in strain is minimal when the base image is correlating well with specific frames in the localization data set. The derivative of the data will reveal sections of the data where the slope is close to zero, which represents a match in the image correlation. These steps are summarized below:

- Smooth out variance
- Derivation
- Smoot Derivative
- Find regions within threshold of zero

3.2.6 Particle Filter

As a visual demonstration of the localization process, a one dimensional particle filter is constructed to take the matching results, *H(t)*, in as input and show a visual representation of the vehicle moving though time (frames) and depict its actual and estimated position on a graph. A particle filter estimates an agent's position using the following pseudo code:

Particle Filter Pseudo Code

```
FOR t = 1 to NumberOfFrames
  Propagate Particles
  IF MATCH
      FOR  i = 1 to NumberOfParticles
            Update weights according to Equation 4
      END FOR
      FOR i = 1 to NumberOfParticles
            Normalize weights using Equation 5
            Add particles x times to temp array, where x is higher depending on weight of
            particles
      END FOR
      FOR i = 1 to NumberOfParticles
            Resample by randomly selecting particles from temp array
      END FOR
```

END IF
END FOR

When a match in the data is found, the weights of the particles are calculated and the particles are re-sampled. Over each frame, the particles are propagated using a motion model which adds randomness in its movement. The weights for the particles in the particle filter are calculated using (10) and the standard deviation, σ_i of matches for a geo-located image, where the match occurs, and the distance from where a match is discovered and where its actual frame number is.

$$\sigma_i = stddev(h_{1i}, h_{2i}, \ldots, h_{ji}) \qquad \begin{array}{l} for\ i = 1 \ldots n \\ j = 1 \ldots numOfMatches\ in\ V_m \end{array} \tag{8}$$

$$D_i(t) = X(t) - GoalX(i) \tag{9}$$

$$W(p) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(\frac{-D(t)^2}{\sigma^2}\right)} \tag{10}$$

In which h_i represents a match from the i^{th} geo-located point and n is number of matches for that geo-located point. The t is the frame number where particle p is at, and W represents the weight of particle p. After all of the weights have been calculated all of the weights get normalized using the following function:

$$W(p) = \frac{W(p) - minWeight}{maxWeight - minWeight} \tag{11}$$

The *minWeight* and *maxWeight* are found in the initial loop of calculating the weights.

Chapter 4

IMPLEMENTATION

This section of the thesis explains how map creation and localization were implemented in an experiment. Titles of software, images, and other low level information will be used to explicitly explain how each of the phases of the experiment was accomplished. Several videos were recorded and two were selected to be used for the map creation and localization.

4.1 Implementation of Map Creation

The block diagram shown in Figure 2 demonstrates how the map creation process is split into five phases; Capture, Edit, Pre-Process, Extract, and Select. The following sub-sections describe how each one of the phases from the corresponding section in the solution was implemented in this experiment.

4.1.1 Video Capture

A Logitech Quick Cam Pro 5000 was used for obtaining all video footage. At the time of purchase, it had very good reviews and was one of the best web cameras for capturing fast motion without blurring images [21]. The Camera was attached using double sided tape to the passenger side window of a 1997 Honda CR-V (Figure 5).

Figure 5: Camera Mounted

The selection of road selected for this experiment began just outside of 2600 El Cerrito Street, San Luis Obispo, California. The video was set to record and the vehicle was driven as slowly as possible (approximately 5 miles per hour) down El Cerrito Street, until the intersection of El Cerrito Street and El Paseo Street (as shown in figure 6) was reached.

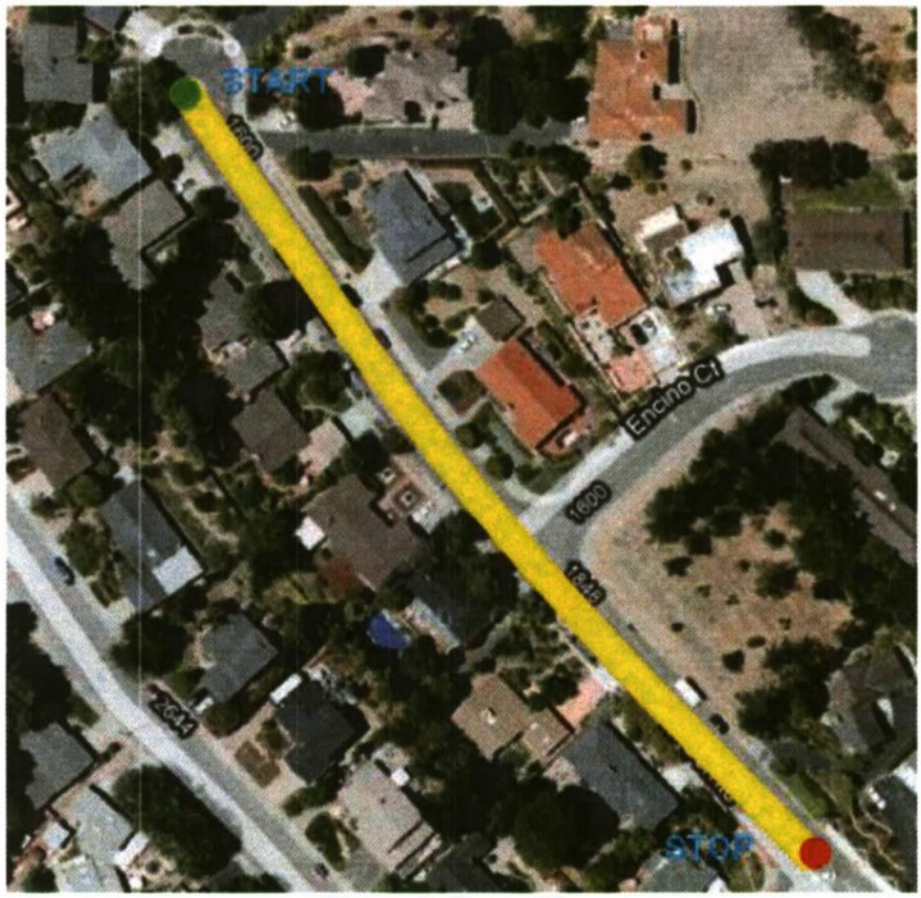


Figure 6: Route Map

4.1.2 Video Editing

Initially the video was opened and edited using Microsoft Windows Movie Maker from Windows Vista (not XP or Windows 7). This specific version was used because of its unusual stability while processing visual effects and for its ability to use both included and custom filters. These filters can seamlessly be applied to videos, and proved helpful in section 4.1.3. Using the trim function in Movie Maker (shown in figure 7), the video was trimmed to only include the visual data from the desired region of the experiment.

Figure 7: Trim Tool

4.1.3 Grey Scale & Edge Detection

Using Microsoft Windows Movie Maker, the video was converted into grey scale (Figure 8). A Microsoft provided implementation of Canny Edge Detection filter was easily added to the Movie Maker program as a free add-on. This filter was then available as a selection choice from the visual effect menu (Figure 9). Applying the edge detection filter took several seconds to apply over an entire video, and had similar if not greater resolution and quality than the OpenCV alternative.

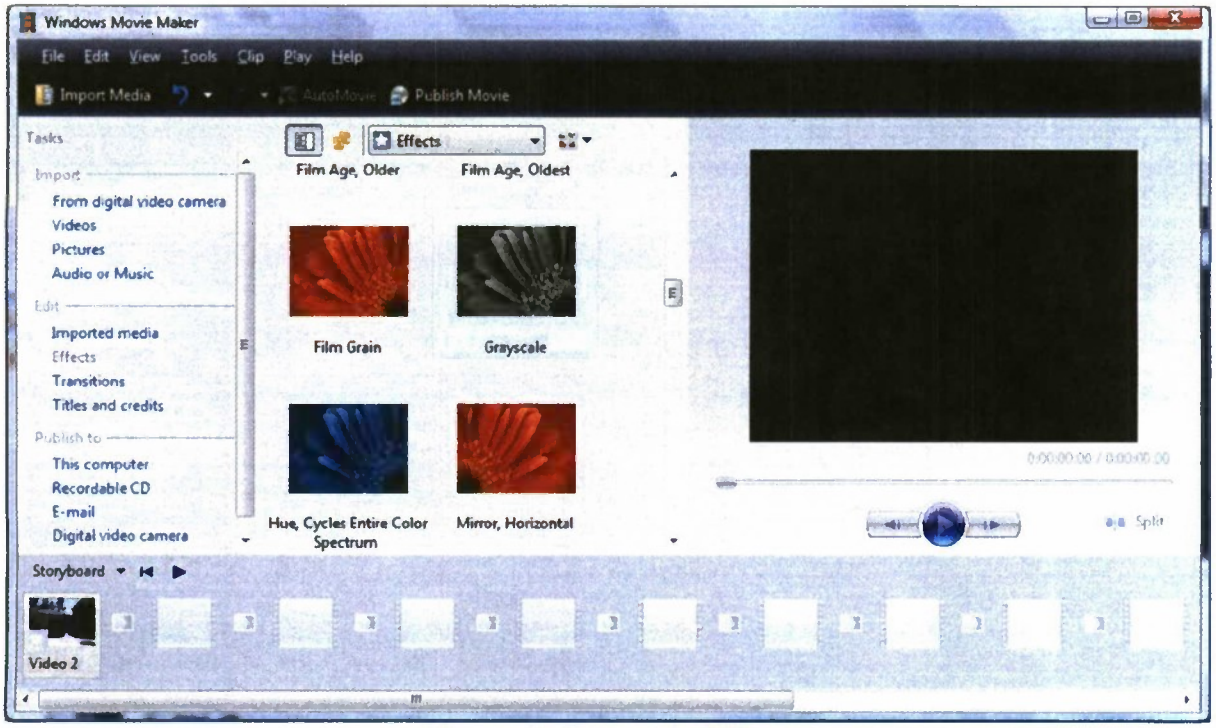


Figure 8: Grey scale Effect

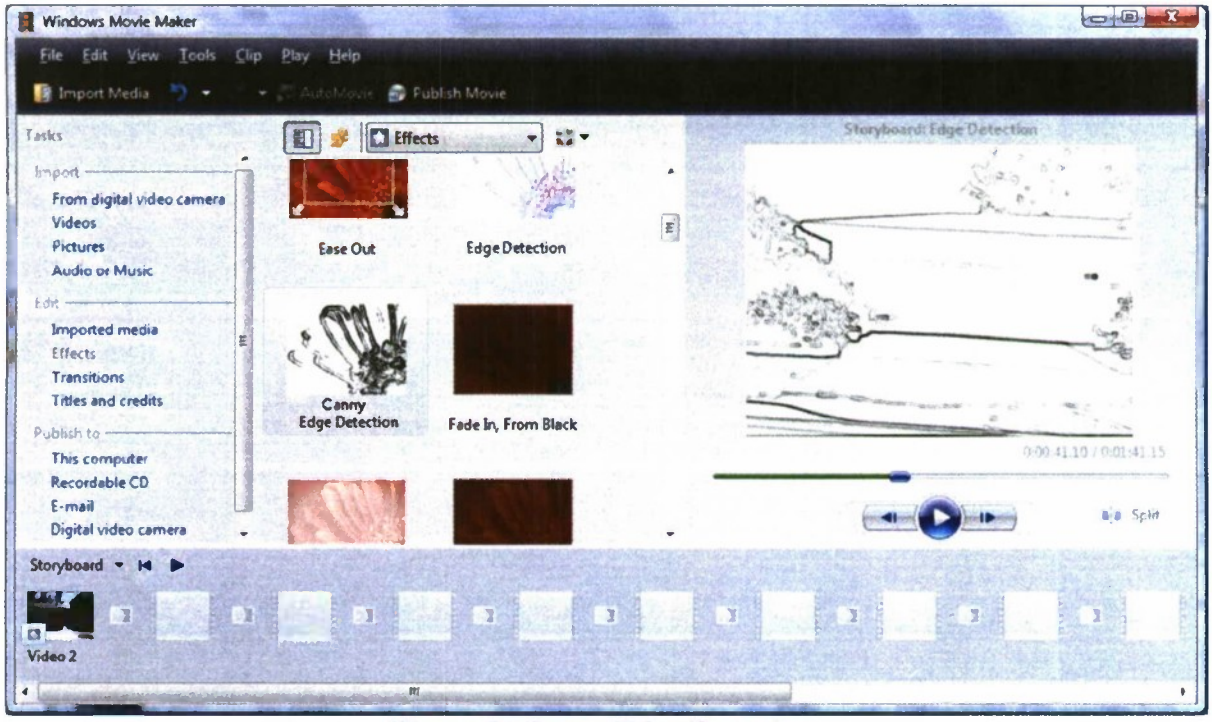


Figure 9: Canny Edge Detection

4.1.4 Convert Video Format

In order to continue with the rest of the experiment, the video had to be converted from its original video format used in Windows Movie Maker (WMV) into a format in which no compression is used. This was done so that all of the frames could be extracted in section 4.1.5. This phase was not originally part of the solution but was a minor step in the implementation of the experiment. To do this, a program called WinAVI was used. WinAVI is designed to convert various video formats into DVD format, and vice versa. In this case the software was used to convert the WMV file into MPEG-2 video format (Figure 10). MPEG-2 format uses no video compression and thus allows the extraction of all the frames from entirety of the video.

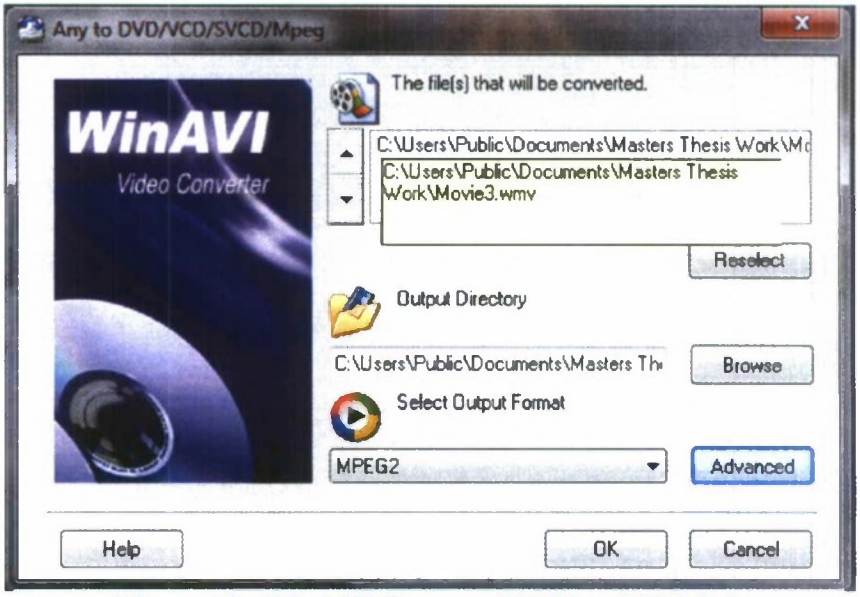


Figure 10: WinAVI Convert WMV to MPEG-2

4.1.5 Image Extraction

IrfanView, a multi faceted image application with batch conversion abilities, was used to extract all of the frames from the MPEG video file produced in section 4.1.4 (figure 11). This process apparently is not always without error. The program had failed multiple times in completing the extraction process. Such failures included skipped frame numbers, corrupted frames, or frames left blank. The solution to this problem was to repeat the process on multiple computers until the process was completed. The cause of this malfunction is yet unknown, but irrelevant to the true purpose of this experiment.

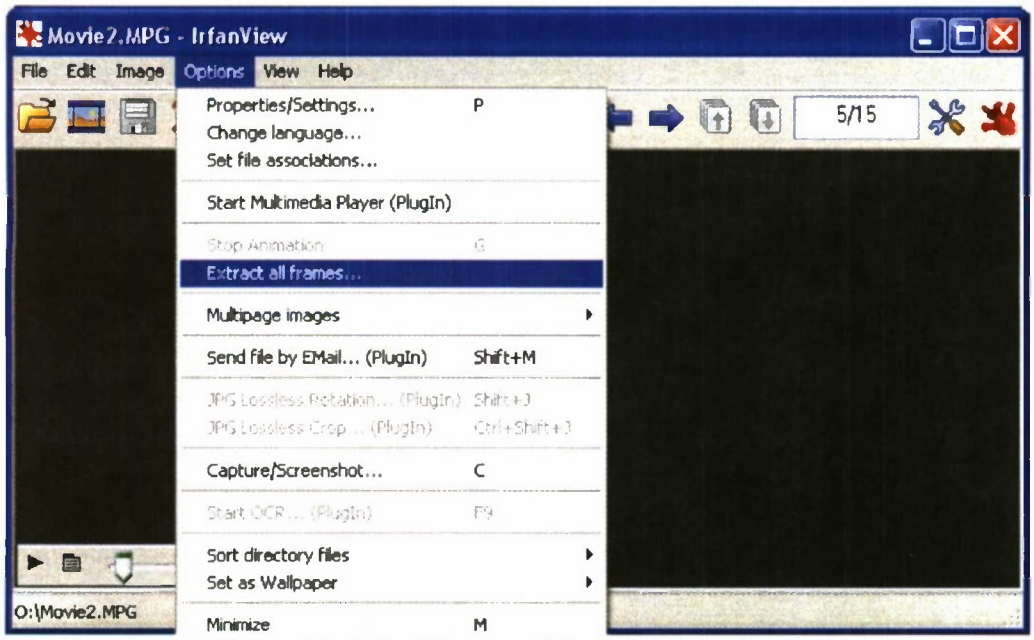


Figure 11: IrfanView used to extract all frames

4.1.6 Map Image Selection

For this experiment, five specific frames were selected from the video data set and their correlated geo-located points on the map were determined. These images were selected using the criteria discussed in section 3.1.4. These frames were labeled image 1 though 5 and are noted below by the image number as well as the corresponding frame number (figure 12). The geo-location information for each point was determined by matching the image with the view on Google Street View and finding the coordinates of that particular location, demonstrated in figure 13 and 14.

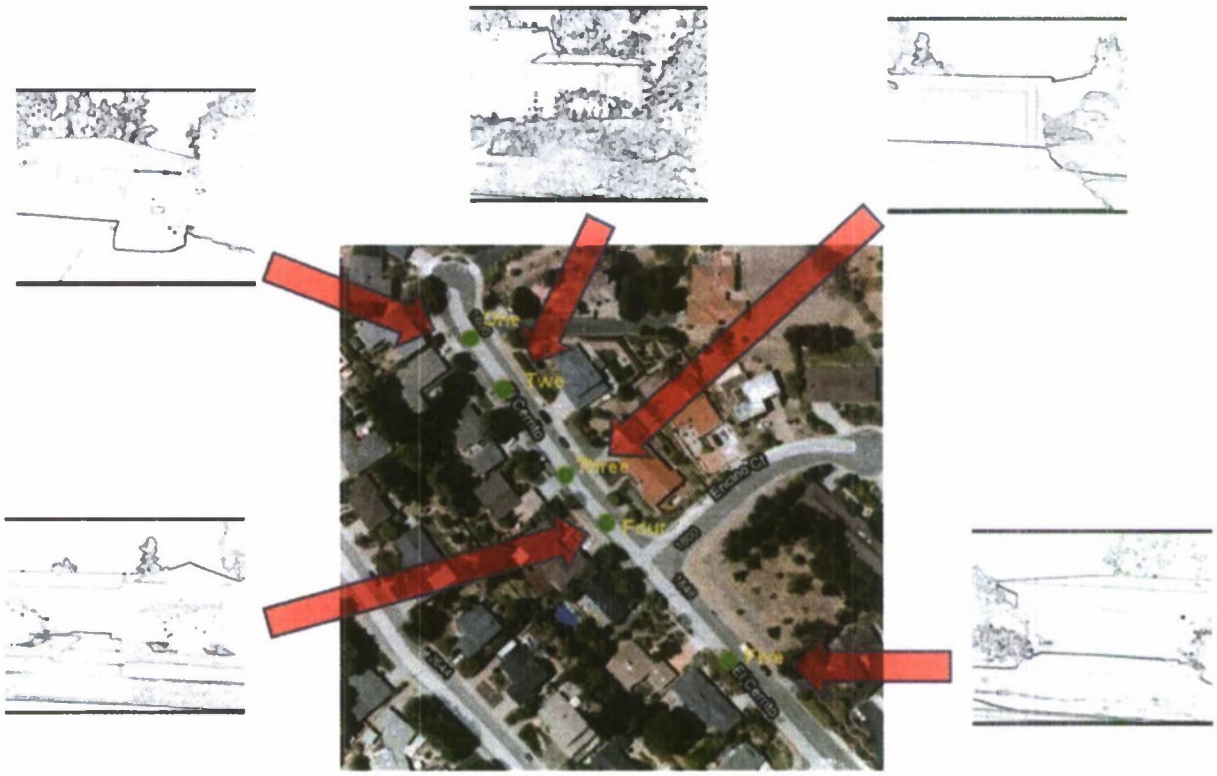


Figure 12: Geo-Located Image Locations

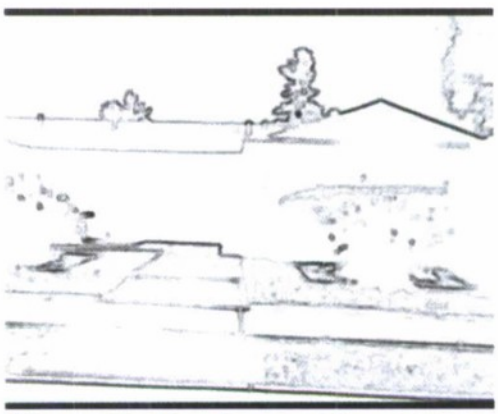


Figure 13: Geo-Located Image Four

Figure 14: Google Street View Map Image 4

4.2 Implementation of Localization

The block diagram shown in Figure 4 demonstrates how the localization process is split into eight phases: Capture, Edit, Pre-Process, Extract, Correlate, Data, Graph, and Particle Filter. The following sub-sections describe how each corresponding phase of the solution was implemented in the experiment..

4.2.1 Video Capture

The same camera used in the map creation part of the implementation was once again attached using double sided tape to the passenger side window of a 1997 Honda CR-V (see Figure 3). Once again, starting just outside of 2600 El Cerrito Street, the video was set to record and the vehicle was driven down El Cerrito Street (Figure 16) until the intersection of El Cerrito Street and El Paseo Street (Figure 4) had been reached.

Figure 15: Start of Driving Run

4.2.2 Video Editing

The video was opened and edited using Microsoft Windows Movie Maker. Using the trim function in Movie Maker (shown in figure 5), the video was trimmed to include the visual data from the desired region of the experiment which corresponds to the desired region used in the map creation from section 4.1.2.

4.2.3 Grey Scale & Edge Detection

While still using Microsoft Windows Movie Maker, the video was converted into grey scale and had the Canny Edge Detection filter mentioned in section 4.1.3 applied. Applying the edge detection filter took a matter of seconds to apply over the entire video. Refer to section 4.1.3 for further detail on this process.

4.2.4 Converting Video Format

The WMV video file needed to be converted to MPEG-2. IrfanView was used to convert the WMV videos into MPEG-2 video format (Figure 10). For further information on this process refer to section 4.1.4.

4.2.5 Image Extraction

Similar to the step taken in section 4.2.5, IrfanView was used to extract all of the frames from the video data set. For more information on this process refer to section 4.1.5. This section differs from section 4.1.5 because right after extracting the images from the video, the frame numbers that correspond to the geo-located frames from the map video had to be found. This is due to the fact that the exact speed throughout both of the videos cannot be regulated nor synced with one another. A visual inspection id used, the frames in the localization video are compared against the geo-located images from the map (Figure 16). When a match was found the frame number was noted elsewhere for use in the Particle Filter described in section 4.2.8.

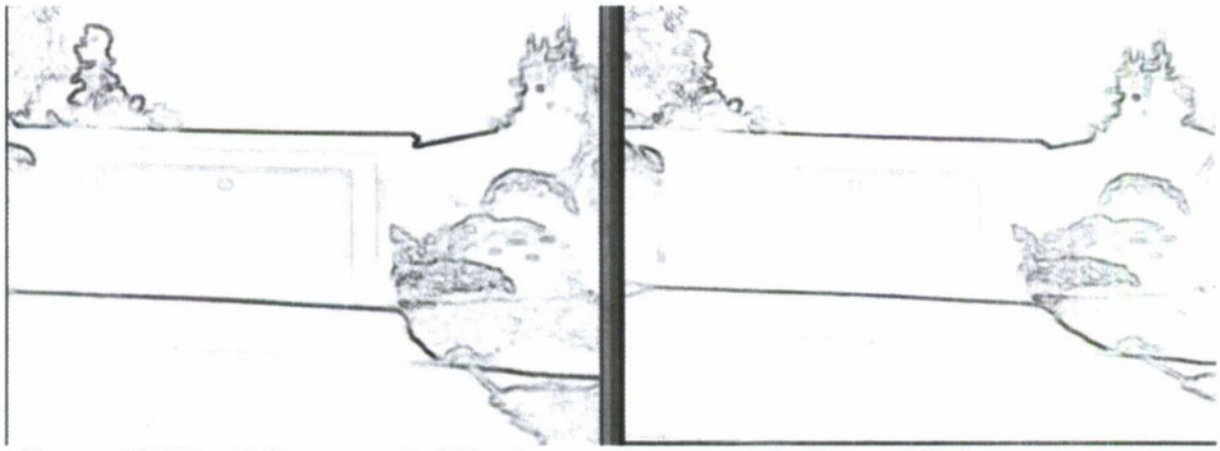


Figure 16: Visual Compare, 3rd Geo-located frame on left, Frame 587 from Localization Video on right

4.2.6 Image Correlation

This particular phase implements the core image processing function of this solution. The process of running the localization video against a geo-located image from the map is actually a process that needs to be done five times; One time for each base image selected to be geo-located in the map creation. To accomplish all of the following tasks mentioned in this section, the Image Correlation and Tracking Library and functions must be installed into Matlab [20]. Initially, a "file list" has to be generated; This is a file that contains the names of all of the frame image files from the localization video. This is accomplished while in MATLAB by typing 'filelist_generator;' and pressing 'ENTER' at the command line of MATLAB. There are two ways to proceed now: Either choose to 'manually' type in the image numbers desired for processing or 'automatically' generate a list of images by pointing out the first image of the set, and the function will find all images with increasing number before the point in the name (e.g. 'PIC00001.Tif') (figure 17). The latter choice was selected and MATLAB was pointed to the first image of the set (dataset2_0001.tif). The function was then able to generate a list of the file names from "dataset2_0001.tif" all the way to "dataset2_1164.tif.

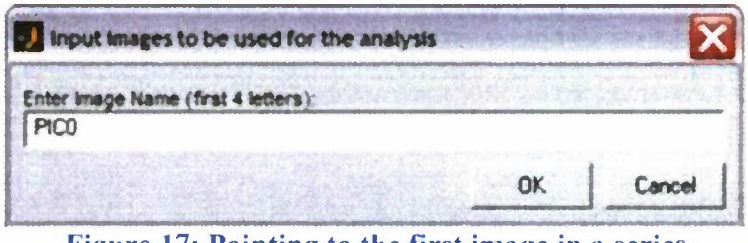


Figure 17: Pointing to the first image in a series

After the file list is generated, MATLAB asks where to save the list. In this case the files were saved to the same directory to which the library and images are found. (Figure 18)

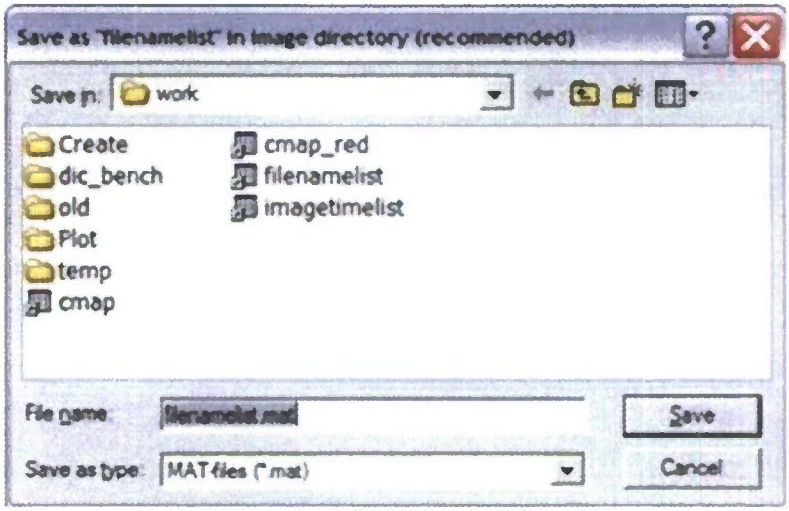


Figure 18: Saving file list

The next step in the image correlation process is to generate a grid (or section) of the frames that the correlation algorithms will be looking. It was decided early on in the experiment that the region of interest would be the top half of the frame as it included roof detail of the houses and avoided the region of the video that had shown cars either parked on the street or in the drive way. In the creation of the grid, one would be asked to point to the base image to which the images in the file list will be compared against. This is where the repeating steps would begin. Initially, the first geo-located image from the map would be selected which should correlate with the image from the localization data set at frame 88 and end with the last geo-located image from the map that correlates with frame 1052 from the localization data set. Once the image was selected a dialog box was presented that asks the user what type of grid they would like to make (Figure 19).

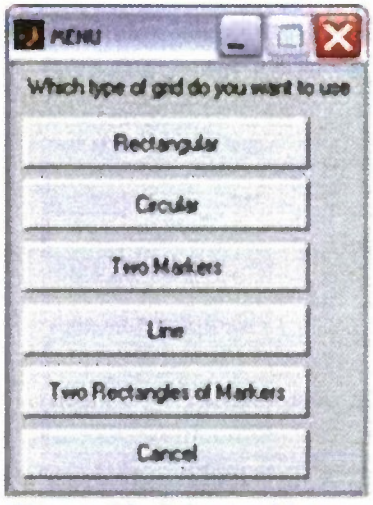


Figure 19: Select grid type

A Rectangular grid was selected for every image. After that selection was made an additional window appears that shows the base image that was selected. Using the mouse, the top half section of the frame was selected.

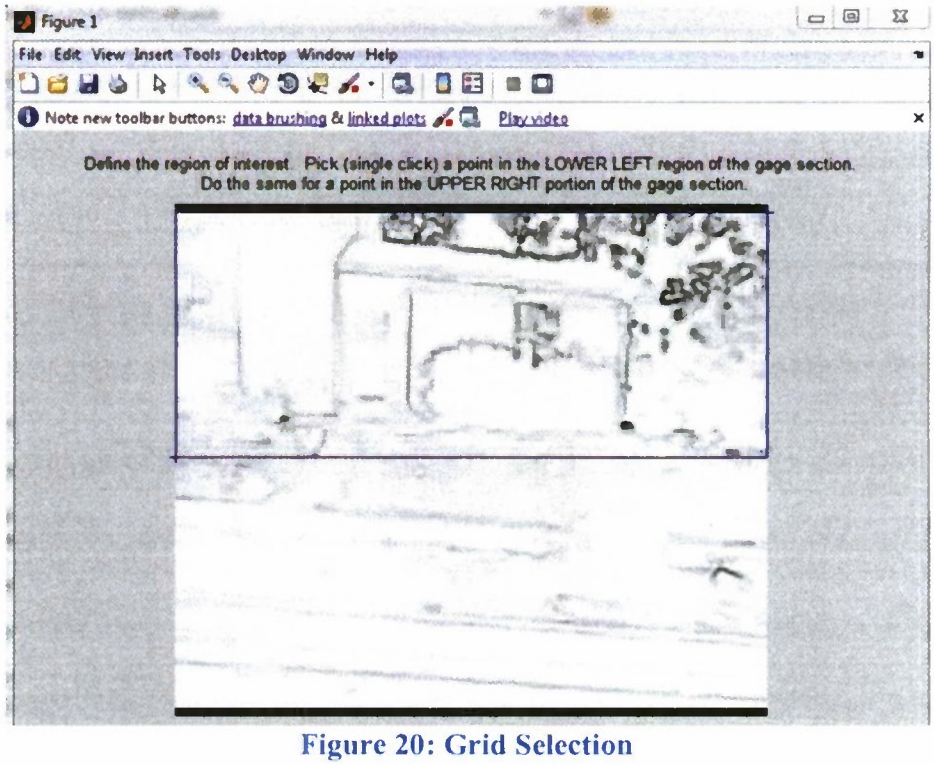


Figure 20: Grid Selection

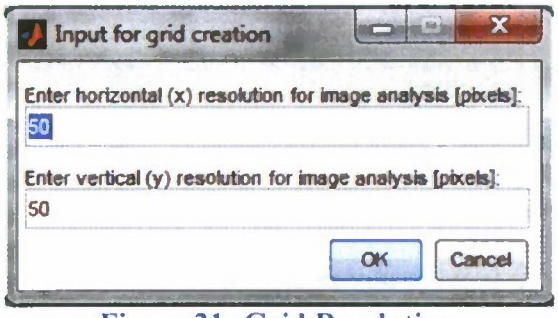


Figure 21: Grid Resolution

After selecting the region of interest, MATLAB then prompted the user for a desired x and y resolution (Figure 21). The default resolution is 50 by 50. Double the default resolution was used as the effective region of interest from the whole image had been cut down to half. Essentially, the section of the frame is being broken up into discretized parts that will be tracked and correlated later. After this was done, the grid files are created, and the image correlation function was ready to be executed.

The automation function is the central function and processes all markers and images. Therefore the 'Current directory' in MATLAB has to be the folder where automate_image.m finds the filenamelist.mat, grid_x.dat and grid_y.dat as well as the images specified in 'filenamelist.mat'. 'automate image;' is typed and the 'ENTER' key is pressed at the command line of MATLAB. At first, automate image opens the first image in the filenamelist.mat then 'automate_image.m' will process all images in the 'filenamelist.mat' and plot the grid with green and red crosses on top. The function is attempting to find similarities in the discretized sections made by the grid generation, from frame to frame, tracking their movement (or strain) over time (Figure 22). Depending on the number of images and markers you are tracking, this process can take between seconds and days. In this case it took about five minutes per base image to run the process over the entire localization data set. When all images are processed, automate image will write the files validx.mat, validy.mat, validx.txt and validy.txt. The text files are meant to store the result in a format which can be accessed by other programs and functions that can be used later and potentially in future applications.

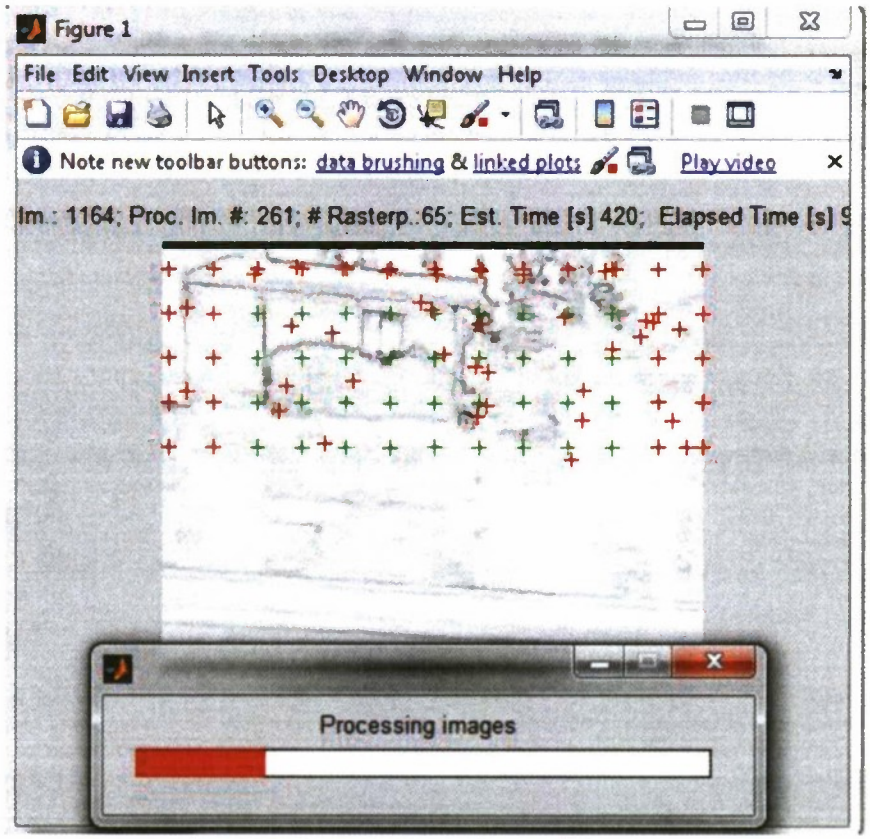


Figure 22: Automate Image Running

The final step involved with the image correlation and tracking library is to analyze the displacement and strain of the features calculated in the previous function. The post processing is the most interesting and rewarding step since one can analyze the collected displacement data. The displacement.m function is a small collection of functions which allows the user to review the displacement field, calculate the strain or delete markers which were not correlated or tracked very well. To start, one types "displacement" into the command window. A window will pop up asking for the validx.dat file which contains the x-displacement off all markers in all images, followed by a dialog for the validy.dat containing the y data. After "displacement.m" has loaded both files, a new window pops up that allows one to choose between different options of visualizing the data (Figure 23).

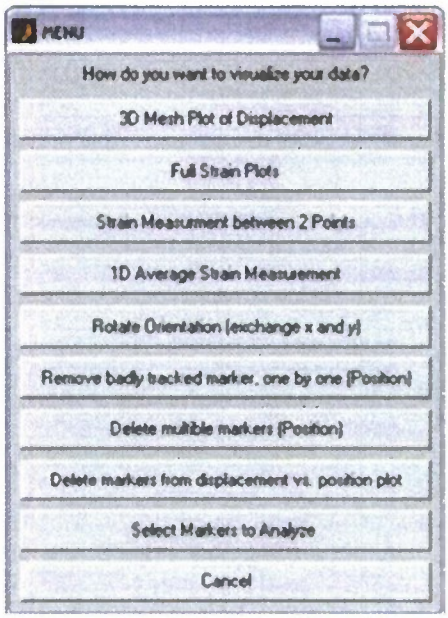


Figure 23: Visualize Data Menu

For the purposes of this experiment, the 1D average Strain Measurements are needed. After clicking on the "1D Average Strain Measurement" button, the x-displacement versus x-direction will be plotted for each image and then fitted by a linear function. The slope is the true strain which will be plotted versus the image number after all images are processed. If you choose to save the strain versus image number, you the data will be saved as an ASCII file which can be opened with MATLAB, Excel, or notepad. In this case the y-strain is negligible as the vehicle is moving predominantly horizontally through space.

The output of the final graph can then be saved as raw data. It is this data that gets smoothed and post-processed later is section 4.2.7 in order to demonstrate how this experiment can be a viable means of autonomous localization.

4.2.7 Map Image Match Determination

The data that is obtained from the previous section depicts a plateau that occurs in the plotting of strain for each of the base images. The goal is to find all of these plateaus as part of the map image match determaination method. One of these plateaus should occur on or around where there is a match in the image correlation process. This happens because the change in strain is minimal when the base image actually is correlating well with specific frames in the

localization data set. There is an entire set of data for each one of the results from the previous section, containing small amounts of error, and expected random fluctuation due to the fact that the geo-located images, and the frames from the localization data set, are not identical. Steps from section 3.2.5 are used to smooth the data in question and to pick out the regions where these plateaus appear. [Appendix A] The code is designed to remove small amounts of fluctuation to make the graph smoother and to extract regions in which the slope of the graph is close to zero. The smoothing section is done by creating a small window and averaging the points within each window across the entire set of points. Once the graph has been sufficiently smoothed, MATLAB takes the derivative of the graph and locates points where the slope is close to 0 within another window. The desired final result is a set of data that has a domain of 1 to 1164, representing the frame numbers, and a range of 0 or 1. The 1's get placed in areas in which are considered a match for that image, and 0's elsewhere. For example, if the data from the first results data processing reads that image 1 has been found at frame 110 then a 1 would be the value at index 110 out of 1164. In reality, the frame is taken at frame 118, so 110 would be quite close. This match determination method is run on each one of the resulting data sets retrieved from section 5.2.6 (a total of 5 times). The resulting data sets can be graphed one on top of another in different colors to show when the whole system thinks the vehicle is at a certain point at a certain time. In addition, this data is pipelined into a one-dimensional particle filter to visually demonstrate the viability of this system.

4.2.8 Particle Filter

Using the Post-Processed data from the previous section and the value of σ_i calculated by taking the standard deviation of the matches in *H(i)* using Microsoft Excel, a one-dimensional particle filter (as described in section 3.2.6) was constructed in MATLAB [Appendix B]. Once this algorithm was implemented into MATLAB, it became a simple deed of pulling the data from section 4.2.7 into the Particle Filter. The particle filter visually shows over time (frames) the estimated position of the vehicle in 1D space as well as the error in position estimation over time.

Chapter 5

RESULTS

The following are the results of the image correlation between map images one though five with each frame of a second video set, V_L (Section 4.2.6).

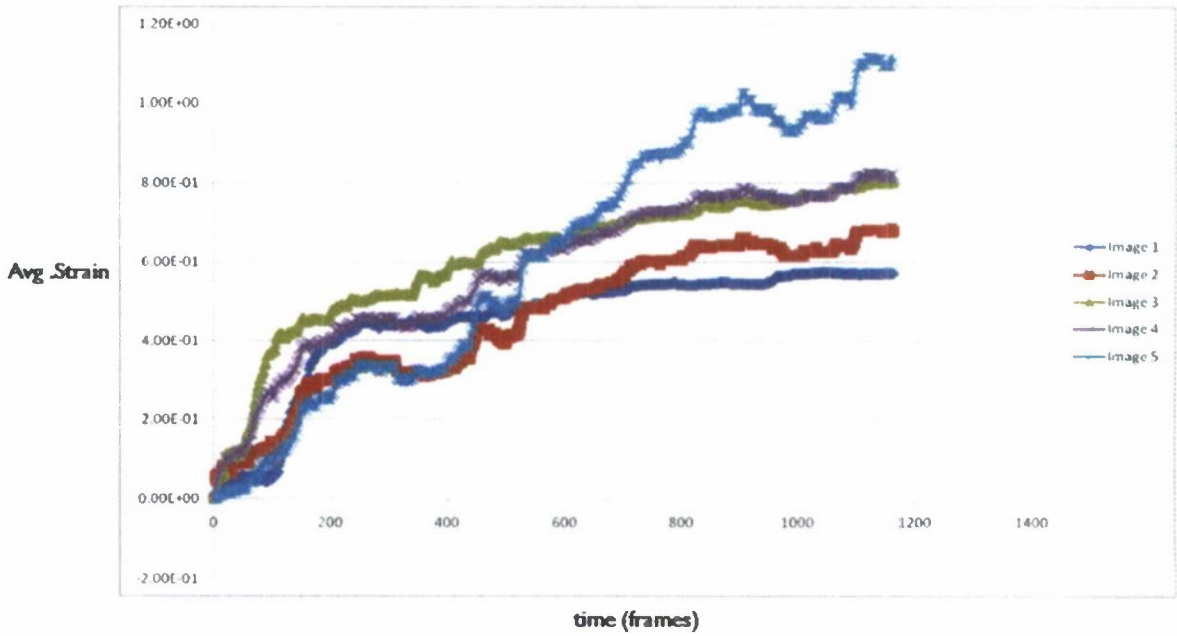


Figure 24: Image Correlation Strain

Figure 24 above represents strain (y-axis) over frames (x-axis). While some of the plateaus are easy to distinguish in the above figure (Figure 24), there is enough variance and random error to warrant the use of the data processing described in section 4.2.7. Image matching described in section 4.2.7 produced matches shown in figure 25.

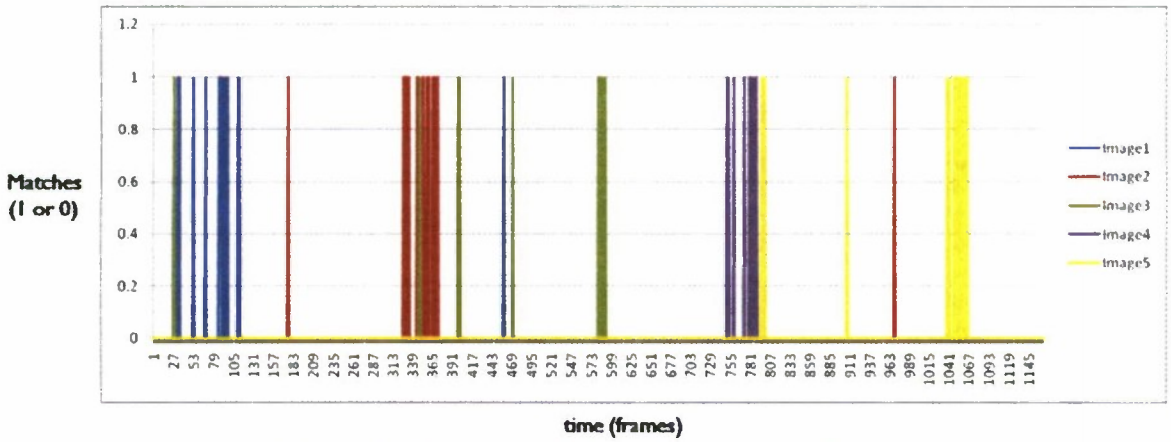


Figure 25: *K(t)* results for each geo-located image

Figure 25 above shows matches as a function of time (frames) in the x-axis. Figure 26 shows when a match is detected for each of the geo-located images over time (frames). It can be seen that there are a few false positives but the matches generally cluster are around the map image location:

- Image 1 => Frame 88
- Image 2 => Frame 353
- Image 3 => Frame 587
- Image 4 => Frame 777
- Image 5 => Frame 1052

These frame numbers were found using visual comparison with the map's geo-located images explained in more detail in section 4.2.5.

This data shows in a simple yes or no (1 and 0 respectively) form, if the system thinks that the vehicle is at the point of the indicated image at a specific time value (frame number). This data is further utilized as it is fed into the simple one-dimensional particle filter. The results of that filter are demonstrated by showing some select frames of the Particle Filter animation, displayed below (Figures 26 through 29), where the particles are represented by the red circles and the actual position is the blue line, the y-axis is irrelevant, and the x-axis is time (frames):

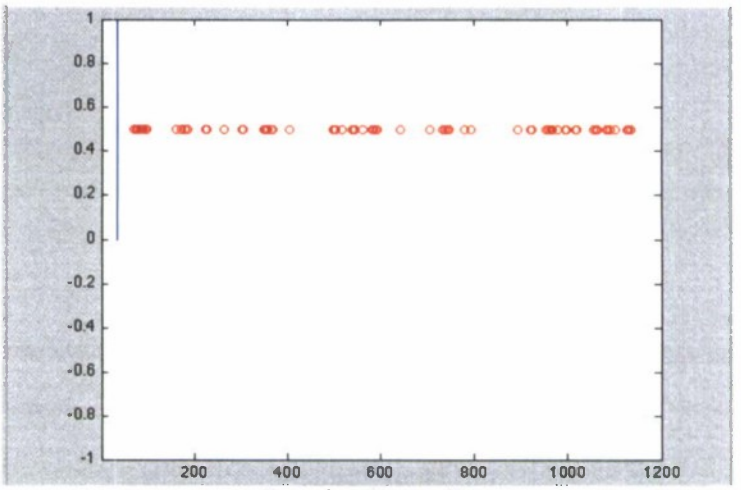


Figure 26: Initial State of Particle Filter

It can be seen in figure 26, near the very beginning of the localization the particles can be found randomly dispersed over the domain. This is because the particles the initial position is unknown. The actual location is indicated by the blue line and the particle locations are indicated by the red circles. The Results improve as the vehicle progress though time (frames).

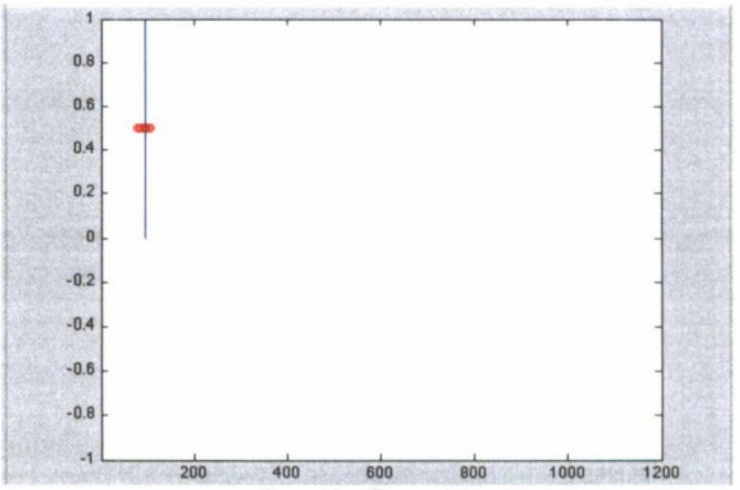


Figure 27: First Grouping at $t = 95$

Figure 27 shows how the particles group around the actual position of the vehicle early on.

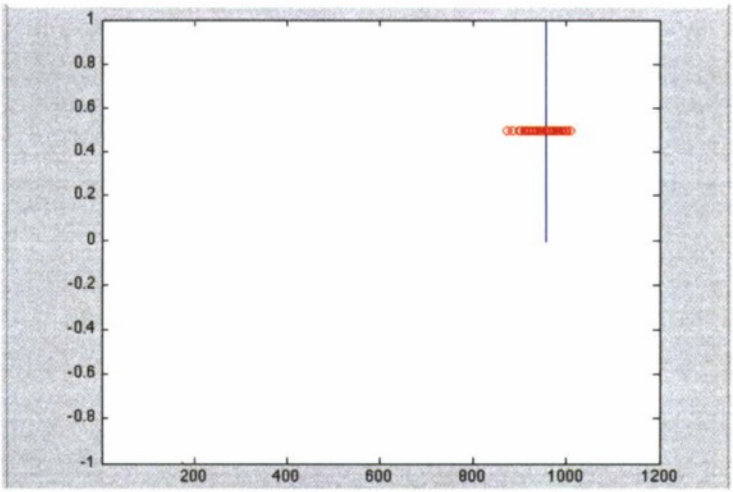


Figure 28: A bit of random dispersion at $t = 955$

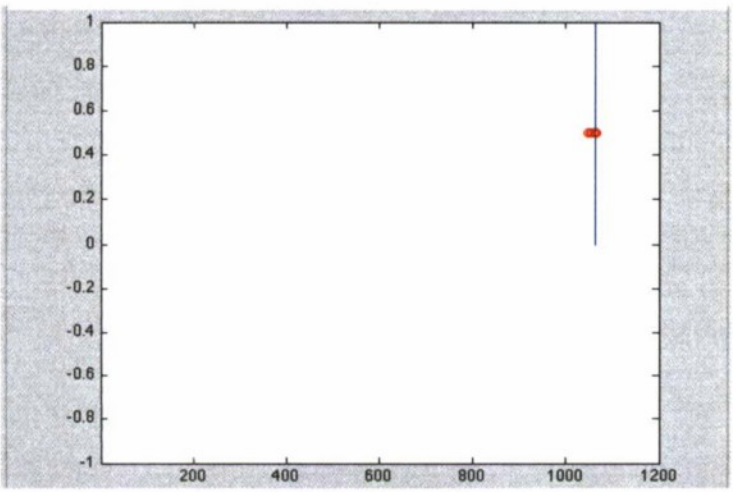


Figure 29: Particle Filter Right after a good set of matches at $t = 1055$

Figure 28 shows a wide dispersion of particles that still average close to the blue line. This dispersion occurs cumulatively in regions where there are no matches in the data, but can be seen once again forming a tight grouping in figure 29, as the particles get re-sampled due to the matches of the 5th image. As the particle filter is running, the position of all the particles is averaged, and the error from that average (estimated position) compared to the actual position is stored.

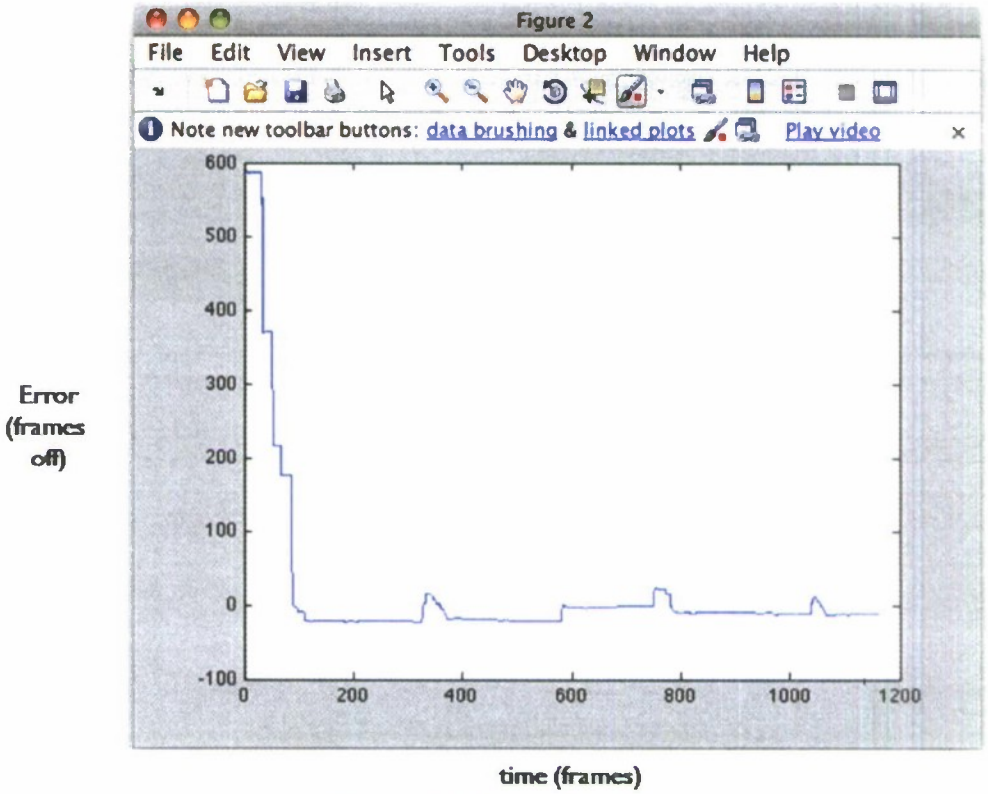


Figure 30: Error Graph

Figure 30, above, shows the error (y-axis) in number of frames off, over time (x-axis) in units of frames in the video. It can be seen that as soon as the localization data revels a good match, the error very quickly drops down to essentially zero, and continues to stay in that area over time.

Chapter 6

CONCLUSION

It can be seen in the data shown in chapter 5 that the particular configuration of technology described in this thesis is a viable means of accomplishing autonomous localization. The image correlations technology successfully finds matches between any two images. The particle filter not only shows how one can localize a vehicle's position from a known starting point, but demonstrates how the particle filter can be implemented to satisfy the *Kidnapped Robot* scenario, where the initial position is unknown.

Experiments were a successful in that the position error decreases towards zero over time. The results demonstrate the ability to determine within X_m where the vehicle was as a function of time while the car was in motion for this experiment. This experiment has shown the potential for this system to be scaled up and to potentially aid in addressing common issues with autonomous outdoor localization.

While not currently implemented in real time, it is expected that real time localization is possible if one or more changes are made. It is believed that current hardware couldn't process the data fast enough to perform the necessary calculations in real-time. With fast enough hardware, and perhaps changes to the correlation and other processing functions, there exists the possibility that certain parts of the implementation can be changed or perhaps upgraded in multiple ways to accomplish this future goal. Additional ideas for further improvements are contained within the next chapter of this thesis.

Chapter 7

FUTURE WORK & POTENTIAL APPLICATIONS

A relatively new feature of Google Maps is an interesting application called Google Street View. Google Maps and Google Earth have proved incredibly popular with Internet users and are used by public and private businesses and individuals as essential and informative tools. Street View provides an added experience by enabling users to see a location as they would if they were standing on the street. [8] There are huge benefits from the Street View feature, ranging from the promotion of tourism, to helping people find local businesses or meeting points, assisting property buyers and prospective tenants. [8] My assumption is that such a database like the one mentioned above which provides one with the street view, could be loaded dynamically as one is driving, and in the event that the GPS signals start degrading or fail, the software within the GPS unit, combined with a camera, could continue to localize one's position by comparing its images that it is recording with those of the Google Street View database.

There could potentially be an issue with constantly scanning large amounts of images so either smaller amounts of data would have to be sent at a time, or a faster storage medium could be used. The most likely method for satisfying this caveat is to use only the images located in the vicinity of GPS measurements. This is related to another issue; should the images and the respective location database be streamed via the Internet using a form of wireless broadband, or should a large storage medium be employed. Both ideas have their merits. Broadband seems more convenient and has the possibility of being updated as the database is updated. On the other hand, the storage medium queries would be faster, and would work in any area (ie. without the need for a wireless infrastructure).

Bibliography

[1] S. Thrun, W. Burgard and D. Fox, "Probabilistic Robotics", MIT Press, 2005.

[2] R. Siegwart and I.R. Nourbakhsh, "Autonomous Mobile Robotics", MIT Press, 2005.

[3] K. Y. K. Leung, C. M. Clark, and J. P. Huissoon, "Monocular Vision based Particle Filter Localization in Urban Environments" , Proc. of the 2008 IEEE International Conference on Robotics and Automation, May 19-23, 2008

[4] W. Zhang and J. Kosecka, "Image based localization in urban environments," in Proceedings of the International Symposium on 3D Data Processing, Visualization, and Transmission, 2006.

[5] W. Wang, "Autonomous Control of a Differential Thrust Micro ROV", University of Waterloo M.Sc. Thesis, 2006.

[6] S. Haykin. Kalman Filters, chapter 1, pages 1–21. John Wiley & Sons, Inc., 2001.

[7] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," International Journal of Computer Vision, vol. 74, no. 3, pp. 237–260, 2007.

[8] "Street View on Google Maps FAQ" http://maps.google.com/help/maps/streetview/faq.html, Google 2009

[9] Ioannis M. Rekleitis, "A Particle Filter Tutorial for Mobile Robot Localization", Center for Intelligent Machines, McGill University. Thesis, 2002

[10] Dieter Fox, Wolfram Burgard, Sebastian Thrun, "Markov Localization for Robots in Dynamic Environments", Carnegie Mellon University, University of Freiburg. Thesis 1999

[11] W. Burgard, A. Derr, D. Fox, A. Cremers, "Integrating Global Position Estimation and Position Tracking for Mobile Robots: The Dynamic Markov Localization Approach", IEEE conference on Intelligent Robots and Systems, 1998

[12] Maria Isabel Ribeiro, "Kalman and Extended Kalman Filters: Concept, Derivation and Properties", Institute for Systems and Robotics in Portugal. Thesis, 2004

[13] Sonar, OpenCV, Accuracy and Laser Rangefinder on Wikipedia, March 16th, 2008

[14] "Cistern Exploration Project", http://users.csc.calpoly.edu/~cmclark/MaltaMapping/, Cal Poly San Luis Obispo, 2008

[15] SICK – New Products, http://www.sickusa.com/gus/products/new/new_products/en, SICK, March 16th, 2008

[16] How Does GPS Work?, http://www.nasm.si.edu/gps/work.html, Smithsonian - National Air and Space Museum, 2005

[17] OpenCV on SourceForge, http://sourceforge.net/projects/opencvlibrary/, Source Forge, 2009

[18] "MATLAB", Wikipedia Coverage, http://en.wikipedia.org/wiki/MATLAB

[19] Image Processing Toolbox - Documentation, The MathWorks Help Desk, http://www.mathworks.com/access/helpdesk/help/toolbox/images/

[20] Image Correlation and Tracking, MATLAB Central – File Exchange, http://www.mathworks.com/matlabcentral/fileexchange/12413

[21] Review of Logitech QuickCam Pro 5000, cNet Reviews, http://reviews.cnet.com/webcams/logitech-quickcam-pro-5000/4505-6502_7-31529173.html?tag=mncol;lst;1

[22] Conversion to Grey scale, Grokking the Gimp, http://gimp-savvy.com/BOOK/index.html?node54.html

Appendix A: 'smooth_try3.m'

```
clear;

% Load data file, change file name accordingly
data = load('data1.txt');

acc = data(:,1);
siz = size(acc,1);
winSize=5

for i=winSize+1:siz

    % Calculate the mean over last winSize steps
    sum = 0;
    for j=i-winSize+1:i
            sum = sum + acc(j);
    end
    mean = sum /winSize;

    % Calculate the mean over last winSize steps
    sum = 0;
    for j=i-winSize+1:i
            sum = sum + (acc(j)-mean)^2;
    end
    var(i) = sum/winSize;

    % Calculate the derivative
    sum = 0;
    for j=i-winSize+1:i
            sum = sum + acc(j)-acc(j-1);
    end
    der(i) = sum/winSize;

    % Threshold der
    if (der(i) < 0.015)
        t_der(i) = 1;
    else
        t_der(i) = 0;
    end

    % Threshold var
    if (var(i) > 0.00004)
        t_var(i) = 1;
    else
        t_var(i) = 0;
    end

    % Combined
    com(i) = t_var(i)*t_der(i);
end
```

Appendix B: 'IP_PF_0.m'

```
clear;

[num text raw] = xlsread('SAD.xls', 'Sheet2');
numMapImages = size(num,2);
videoSize = size(num,1);
speed = 1;

mapImageX(1) = 88;
mapImageX(2) = 353;
mapImageX(3) = 587;
mapImageX(4) = 777;
mapImageX(5) = 1052;

% Initialize filter
numParticles = 200;
for i=1:numParticles
   particleX(i) = rand*videoSize;
end
figure(1);
axis([1 videoSize 0 1.5]);
hold on;


%Loop over video
for t=1:videoSize;


            %If any hits
    if (sum(num(t,:))>0)
        hits = true;
        goalId = 0;
        for j=1:5
           if num(t,j) == 1
              goalId = j;
           end
        end
        [ t goalId]
    else
        hits = false;
    end

    sumWeights = 0;
    maxWeight = 0;
    minWeight = 1;
    for i=1:numParticles

        % propagate particles
        r = rand;
        if r>0.9
            r_delta = speed + 4;
        elseif r<0.1
```

```
            r_delta = speed -4;
        else
            r_delta = speed;
        end
        particleX(i) = particleX(i) + r_delta;


        %If any hits
        if (hits == true)


            sigma=10.7; %600.7 makes things work good
            dist(i) = particleX(i) - mapImageX(goalId);
            Weight(i) = 1/sqrt(2*pi)/sigma * exp(-(dist(i)/sigma)^2)+0.01;
            sumWeights = sumWeights + Weight(i);
            maxWeight = max(maxWeight, Weight(i));
            minWeight = min(minWeight, Weight(1));
        end
    end


     % Resample
    if(hits == true)
        for i=1:numParticles
            Weight(i) = (Weight(i)-minWeight)/(maxWeight-minWeight);

        end

        PTEMP = [];
        PTEMPINDEX = 1;
        for i=1:numParticles
            if Weight(i) < 0.2
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
            elseif Weight(i) < 0.4
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
            elseif Weight(i) < 0.6
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
            elseif Weight(i) < 0.8
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
```

```
            else
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
                PTEMP(PTEMPINDEX) = i;
                PTEMPINDEX = PTEMPINDEX + 1;
            end
        end

        tempParticle = [];
        for i=1:numParticles
            r = ceil(rand*size(PTEMP,2));
            tempParticle(i) = particleX(PTEMP(r));
        end

        particleX = tempParticle;

    end

    x_est(t) = sum(particleX)/numParticles;
    error(t) = x_est(t) - t;
    % plot actual position
    clf;
    plot([t t],[0 1]);
    axis([1 1200 -1 1]);
    hold on;

    % plot particles;
    plot(particleX,0.5*ones(numParticles),'ro');
    pause(0.0001);

end

figure(2)
plot(error);
```

Wind Turbine Nacelle Senior Project

For:

Patrick Lemieux

Prepared by:

[Hurakan Tech]

Alvaro Martinez (amartincz1218@gmail.com) _____

Francisco Martinez (ciscojncz@gmail.com) _____

David Nevarez (davidnevarez@gmail.com) _____

Zachary Taylor (zacharywtaylor@gmail.com) _____

Statement of Disclaimer

Since this project is a result of a class assignment, it has been graded and accepted as fulfillment of the course requirements. Acceptance does not imply technical accuracy or reliability. Any use of information in this report is done at the risk of the user. These risks may include catastrophic failure of the device or infringement of patent or copyright laws. California Polytechnic State University at San Luis Obispo and its staff cannot be held liable for any use or misuse of the project.

Contents

List of Tables 4

List of Figures 5

1. Abstract 6

2. Introduction 7

3. Background 8

4. Specification Development 11

5. Management Plan 14

6. Design Development and Fabrication 15

6.1. Base Structure 16

6.2. Yaw System 21

6.2.1. Yaw Bearing 21

6.2.2. Tail Vane 24

6.2.3. Tail Boom 26

6.3. Fairing 27

6.4. Drivetrain 30

6.4.1. Fatigue Analysis 32

6.4.2. Bearing Analysis 36

6.4.3. Coupling 39

6.5. Overspeed Protection 39

6.5.1. Calipers 42

6.5.2. Hydraulic Power Unit 44

7. Testing Plan 45

8. Conclusion 46

9. References 47

10. Appendix 48

List of Tables

Table 4.1.....12
Table 5.1.....14
Table 6.2.2.1.....25
Table 6.4.1.....31
Table 6.4.1.1.....33
Table 6.4.1.2.....33
Table 6.4.1.3.....34
Table 6.5.1.....40
Table 6.5.1.1.....43

List of Figures

Figure 2.1.....7
Figure 3.1.....8
Figure 3.2.....9
Figure 6.1.....16
Figure 6.1.1.....17
Figure 6.1.2.....18
Figure 6.1.3.....19
Figure 6.1.4.....20
Figure 6.1.5.....20
Figure 6.1.6.....21
Figure 6.2.1.1.....22
Figure 6.2.1.2.....23
Figure 6.2.1.3.....23
Figure 6.2.2.1.....25
Figure 6.2.2.2.....25
Figure 6.2.3.1.....26
Figure 6.2.3.2.....26
Figure 6.3.1.....27
Figure 6.3.2.....27
Figure 6.3.3.....28
Figure 6.3.4.....28
Figure 6.3.5.....29
Figure 6.4.1.....30
Figure 6.4.2.....31
Figure 6.4.1.1.....32
Figure 6.4.1.2.....35
Figure 6.4.2.1.....36
Figure 6.4.2.2.....36
Figure 6.4.2.3.....37
Figure 6.4.2.4.....38
Figure 6.4.2.5.....39
Figure 6.5.1.1.....43
Figure 6.5.2.1.....44

1. Abstract

The following report outlines the analysis, construction and test plan of a wind turbine nacelle. The wind turbine nacelle was given as a senior project by Dr. Patrick Lemieux to four mechanical engineering students; Alvaro Martinez, Francisco Martinez, David Nevarez and Zachary Taylor during the 2008-2009 academic school year at Cal Poly San Luis Obispo. The wind turbine nacelle is to serve as an electricity power producing wind turbine for the college of agriculture to power a pump and a teaching and research tool for the Cal Poly Wind Power Research Center.

2. Introduction

San Luis Obispo is currently the location of the Cal Poly Wind Power Research Center (CPWPRC). The CPWPRC is directed by Dr. Patrick Lemieux and through his efforts has acquired land in Escuela Ranch to construct a wind turbine, see figure 2.1.

Figure 2.1: Google Image of the Escuela Ranch site in relation to the Cal Poly campus.

Dr. Lemieux has requested that a mechanical engineering senior project design team construct a wind turbine nacelle (WTN) to act as a platform for the CPWPRC and serve as a fully functional wind turbine to produce electricity to power a pump to irrigate land for the College of Agriculture. The project requirements are that the WTN contain all the vital parts of a WTN such as a; drivetrain, overspeed protection, frame, yawing mechanism and fairing. The WTN must also leave room for future upgrades and modifications as seen fit by Dr. Lemieux.

3. Background

A traditional electricity producing wind turbine has three blades and a drivetrain parallel to the wind direction. The being constructed for the CPWPRC has four main components; rotor, nacelle, tower and electrical load bank. The rotor is where the energy from the wind is captured and transferred into a torque. The nacelle is where all the vital subsystems are housed such as the drivetrain, generator, overspeed protection, yawing mechanism and sensor or monitoring devises. The load bank is where the electricity produced by the generated is transmitted and used for various uses such as; charging a battery bank, resistive heating, tying into the existing grid, etc. The tower is the structure that supports the rotor, nacelle and allows for a safe conduit to transmit electrical wires from the generator or to any electrical device in the nacelle. For the wind turbine being built for the CPWPRC the rotor is being designed and fabricated by two mechanical engineering master students; Bryan Edwards and Devin Gosal. The tower is designed by Dr. Lemieux. The load bank is designed and constructed by the mechanical engineering senior project team Electric Breeze, comprised of; Michael Kulgevich, Travis Robins and Wayland Li. The WTN is designed and constructed by the mechanical engineering senior project team Hurakan Tech comprised of Alvaro Martinez, Francisco Martinez, David Nevarez and Zachary Taylor, figure 3.1.

Figure 3.1: Photograph of Dr. Lemieux, left, with Hurakan Tech (from left to right; Francisco Martinez, David Nevarez, Alvaro Martinez, and Zachary Taylor) at the site of the wind turbine.

Because the wind turbine will serve as a research and teaching tool for Cal Poly the WTN will be designed such that it is representative of the technologies used in industry. For the design of the WTN the S-250 5kW wind turbine made by Endurance Wind Power [13] was used as the main benchmark, figure 3.2.

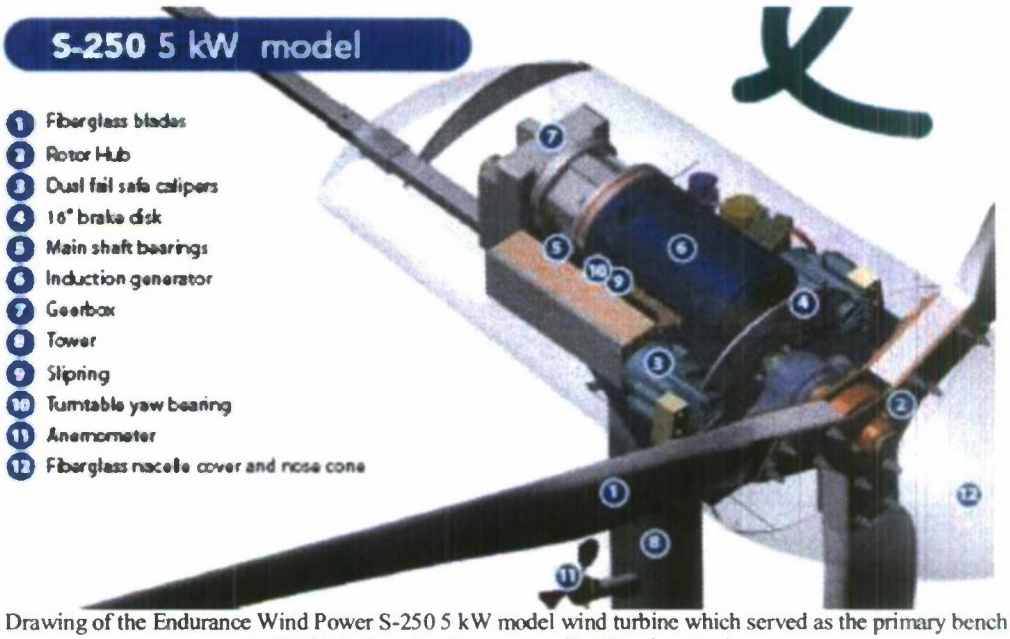


Figure 3.2: Drawing of the Endurance Wind Power S-250 5 kW model wind turbine which served as the primary bench mark for the WTN designed and constructed by Hurakan Tech.

The S-250 was designed and constructed through the efforts of Dean Davis, a Cal Poly alumni and current vice president of engineering for Endurance Wind Power. The S-250 features a mechanical dual fail safe overspeed protection and a passive yaw control, both of which are present of the WTN designed by Hurakan Tech. The size of the S-250 model is comparable to the size of the WTN designed by Hurakan Tech.

Other manufactures and associations that focus on the development and implementation of small scale horizontal axis wind turbines were also used as a reference, such as; Bergey Windpower Company, Iskra Wind Energy, Kestrel Wind and American Renewable Energy. All of the following manufactures produce small three blade horizontal axis wind turbines with various components similar to those used by the WTN designed by Hurakan Tech.

Bergey Windpower Company has been involved with wind energy technologies for nearly 40 years. Bergey began work with their small horizontal axis 1kW XL.1 battery charging wind turbine and now have a 7.5 kW Excel-R and 10 kW Excel-S wind turbines [6]. The XL.1 and Excel-R are on the same scale as WTN and were good size and weight benchmarks.

Iskra is a wind energy company from the United Kingdom and currently has an AT5-1 5kW wind turbine, which features electro-dynamic overspeed protection, direct-drive drivetrain and a lightweight simple design [9]. The Iskra AT5-1 is similar to Hurakan's WTN in that it contains an electro-dynamic braking system and contains no gearbox.

Kestrel Wind Turbines is a manufacturer based out of South Africa. Kestrel features an e400*i* 3kW wind turbine with a direct-drive drivetrain, lightweight design and passive yaw control. The wind turbine can be either grid integrated or used to power pumps [10]. The Kestrel e400*i* was a useful comparison for relative tail vane and yaw bearing sizing.

Abundant Renewable Energy (ARE) is a wind turbine manufacturer based out of Oregon with branches all over the United States. ARE features the ARE110 2.5 kW wind turbine. These turbines contain direct-drive drivetrains, passive yaw, robust design, and quiet operation [11]. ARE was a helpful resource because they have a representative in Atascadero who was able to answer some of Hurakan Tech's initial questions during the brainstorming phase of the design.

Aside from manufacturer's, research centers such as the National Renewable Energy Laboratory, and the German Wind Energy Institute were helpful in preliminary research in current technologies that exist for small scale wind turbines. As a design criterion the International Electrotechnical Commission (IEC) served as a reference for factors of safety and load predictions.

The National Renewable Energy Laboratory (NREL) has performed extensive research in wind energy technology, and served as a great resource for the design of the wind turbine nacelle. NREL's National Wind Turbine Technology Center has published many research papers on wind turbines from understanding the basics of wind energy to wind development system modeling [5]. NREL was helpful during preliminary research and brainstorming.

The German Wind Energy Institute (DWEI) is one of the world's leading consultants in wind energy and "is accredited to EN ISO/IEC 17025 and MEASNET and is recognized as an independent institution in various measurement and expertise fields..." [7]. The DWEI provides the public with scientific publications regarding many aspects of wind turbine technology and served as an excellent source for preliminary research and brainstorming.

The International Electrotechnical Commission (IEC) is the leading global organization in preparing and publishing international standards for all electrical and electronic related technologies. The publications serve as a basis for national standardization and as references when drafting international contracts [4]. The IEC provided Hurakan Tech with the baseline standards that were used for the analysis of the WTN. International Standard IEC 61400 parts: 1, 2, 12, 12-1, 13 and 24 acted as the foundation for the WTN design.

4. Specification Development

The main objective throughout the 2008-2009 academic school year was to produce a fully functional WTN with the specifications given by Dr. Lemieux and fulfill the requirements of the IEC. These requirements are summarized in a Quality Function Deployment (QFD) map presented as table 4.1. The QFD lists the requirements not only of Dr. Lemieux but also of; the College of Agriculture, students, and maintenance and manufacturing personnel. The first row lists the engineering requirements that Hurakan Tech selected to accomplish. Each engineering requirement is compared to a user requirement and rated on how well they correlate. A solid black circle represents a strong correlation, a hollow circle represents a medium correlation, a triangle represents a low correlation and a blank represents no correlation.

Structural dynamics and specified situations that produce worst case scenario loads drove the design. The conceptualization of solutions for undesired situations was another important part of the design. The adequacy of the design was verified by performing calculations and a testing plan [4].

Wind turbines are subjected to environmental conditions that may affect their loading, control system functionality, durability, corrosion, etc [4]. It was anticipated that wind conditions would be the primary external conditions affecting structural integrity. The design wind loads were supplied by Dr. Lemieux (Appendix B). To ensure the appropriate level of safety and reliability the environmental conditions were taken into account in the design and are explicitly stated in the design documentation.

Once subsystems and components were selected the design consisted primarily of stress analysis. To ensure the components were able to withstand the predicted loads, references such as; "Shigley's Mechanical Engineering Design" [3] and manufacture engineering design documentation such as; Timken design documentation for bearing load determination and selection, WC Branham engineering documentation for brake system design and Lovejoy coupling sizing and selection documentation. During stress analysis, safety factors and industry standards were taken into account, these factors and standards are referenced from the IEC standards on wind turbine design [4].

An overspeed protection system is integral to the design of the WTN. An overspeed protection system will be activated in the event of a control system failure, from effects of an internal/external failure, or dangerous events [4]. A manual stop is also designed in case the wind turbine needs to be stopped for any reason. Based on IEC 61400-1, protection systems will activate in cases such as; overspeed, generator overload or fault, and excessive vibration. All protection systems are designed for fail-safe operation.

Table 4.1: Quality Function Deployment Map for the WTN.

Correlation ●=Strong ○=Medium Δ=Low Blank=None	Weighting	Max blade tip speed	Max WTN weight	Rotor braking time (5-10 seconds) from max rpm	Maintenance (hand tools only) 5 tools max	Yaw accuracy	Cooling System	rotor blade attachment clearance	Max hub stress	max drag coefficient of 0.4	Redundant braking system
Patrick Lemieux											
Direct-drive drivetrain	3		○	Δ	●						Δ
Fail safe overspeed protection	10	●	○	●	Δ			Δ	○		●
Remote braking system	8	Δ	○	●	Δ				○		●
Cooling system	7		○	●	○		●			Δ	Δ
Adjustable yaw bearing mount	7		○		Δ	●					
Wire ducting path	8		Δ		○		○			○	
Passive yaw control	9		○		Δ	●				○	
Modular mounting for components inside WTN	10		●		○		○				○
Ag dept											
Power pumps	8	Δ				○					
Students											
Simple layout	6		○		●	Δ	○				○
Maintenance											
Easy to maintain	7		Δ		●	Δ	○	Δ			●
Simple disassembly	3		●		●	○	○	Δ			○
Simple installation	6				●	○	Δ	Δ			
Manufacturing											
Low number of tools needed	6				●		Δ				○
Small time frame to build	7			○	Δ						○
Common materials	4		●		○		○		●	○	Δ
Lightweight materials	7		●	○	○	○			○		
Low cost	6		●	Δ	○	○	●	Δ	○	○	○
Units		ft/sec	lb	sec	Tools	Degrees	Btu/hr	in	psi	-	>1 Braking system

Referring to the IEC 61400-1, the WTN design follows two major points of safety:

- A normal safety class that applies when a failure results in risk of personal injury or other social or economic consequence.
- A special safety class that applies when the safety requirements are determined by local regulations and/or the safety requirements are agreed between the manufacturer and Dr. Lemieux.

When looking at different types of failures, the IEC 61400 served as the primary reference. The IEC 61400-1 divides all components into three classes when determining safety factors:

- Component Class 1: "fail-safe" structural components whose failure does not result in the failure of a major part of a wind turbine; for example, replaceable bearings with monitoring.
- Component Class 2: "non fail-safe" structural components whose failures may lead to the failure of a major part of a wind turbine.
- Component Class 3: "non fail-safe" mechanical components that link actuators and brakes to main structural components for the purpose of implementing non-redundant wind turbine protection functions.

These safety factors served as a reference and guided the design toward a safe and long lasting WTN.

In general the safety factors and partial safety factor for loads used throughout the design were consider deviations and uncertainties of loads from predetermined max values for loads. These factors gave way to safe design values for uncertainties and variability in loads.

Referring to IEC 64100-1, the partial safety factors for materials used in the design of the WTN take into account:

- Possible unfavorable deviations/uncertainties of the strength of material from the characteristic value.
- Possible inaccurate assessment of the resistance of sections or load-carrying capacity of parts of the structure.

When conducting any kind of analysis of the WTN four types of analysis drove the design; analysis of the ultimate strength of parts and materials, analysis of the fatigue failure due to the wind turbines rotating systems, stability analysis, and critical deflection analysis.

Throughout preliminary analysis the WTN followed a steady state model in order to simplify calculations and to quickly gather useful loads for preliminary design. Steady state analysis assumed extreme wind conditions and a vibration provided by Dr. Lemieux (Appendix B). Based on the steady state model of the WTN fatigue damage was estimated using Miner's rule [4] and the fatigue analysis calculations given in "Shigley's Mechanical Engineering Design" [3].

5. Management Plan

Planning was crucial for this project in properly delegating out tasks to each member to establish a minimum progress. Following tasks developed on the Gantt chart Appendix A, the team attempted to maintain a proper time table.

The following were the responsibilities of each team member throughout the 2008-2009 academic school year:

- Alvaro Martinez: Primary contact for Dr. Lemieux and other members collaborating on the WTN design and responsible for organizing testing plans.
- Francisco Martinez: Responsible for maintaining the documentation current and up to date and drivetrain analysis
- David Nevarez: Responsible for documentation of project progress, and prototype fabrication and overspeed protection analysis
- Zachary Taylor: Responsible for majority of prototype fabrication and overseeing manufacturing considerations

Table 5.1 identifies the members of Hurakan Tech and their contribution to the design and fabrication of the WTN.

Table 5.1: Hurakan Tech members and contribution to the design and fabrication of the WTN

Team Member	Contribution
David Nevarez	Frame vibration analysis and FEA. Mechanical overspeed control system, sizing and FMEA.
Francisco Martinez	Drivetrain vibration analysis, bearing and coupling analysis and sizing.
Alvaro Martinez	Yaw system analysis and sizing, fairing design and fabrication, and tail vane design.
Zachary Taylor	Frame, tail boom, brake caliper mounts and shaft fabrication.

6. Design Development and Fabrication

The main subsystems that are identified:

- Base structure
- Yaw System
- Fairing
- Drivetrain
- Overspeed protection

The subsystems are listed by which received the most attention during our design and analysis. The hierarchy was decided based on the likelihood of a subsystem being completed. The base structure, yaw system, and fairing are likely to be unchanged. Based on input from Dr. Lemieux, some subsystems may be changed or modified in the future for other projects. For example the rotor may be modified to include blade pitching. The overspeed protection may be replaced with a larger or completely different system.

The following is a brief explanation of the final design of the WTN. Section 7 includes a more detail of the analysis and fabrication of the respective subsystems. The design includes all major systems and subsystems. As seen in Figure 6.1 the design of the WTN consists of a 43 inches long by 27.5 inches wide by 20.5 inches tall fairing. Including the tail vane and boom the total length of the nacelle is 85 inches or about 7 feet. Inside the fairing is found a direct-drive drivetrain that is coupled to a Ginlong permanent magnet generator, GL-PMG-3500, via a Lovejoy LC110 Jaw Type coupling. In order to protect the generator from thrust loads a direct mounted bearing design was chosen. The driveshaft is 1.5 inches in diameter beginning at the generator through the bearing closest to the generator and stepping up to 1.75 inches in diameter to accommodate the braking system and providing a contact surface for the bearing to isolate the generator from the thrust force. The driveshaft steps back down to 1.5 inches in diameter through the bearing farthest from the generator continuing out the fairing into the spinner. The second step provides a safety mechanism that protects the generator and the breaking system from thrust force in either direction in case the tail boom or tail vane breaks off causing the nacelle to orient itself so that the rotor is downwind.

The brake disk and calipers are from WC Branham. The brake disk is 10 inches in diameter. The brake calipers are FS200 hydraulic spring applied brake calipers. The spring applied features causes the calipers to close on the disc stopping the rotor in their lowest energy state. In order to allow the rotor to rotate the calipers are pressurized with at least 750 psi of pressure. If pressure is lost or the hydraulic pump losses power the calipers will close and bring the rotor to rest. This is where the overspeed protection gains its failsafe feature.

The bearings chosen for this design are Timken 308W Medium Series. The frame of the nacelle is of a step design welded with 2 inch square tubing 1/8 inch thick. This design provides rigidity and strength while keeping the nacelle considerably light weight. The generator is bolted down on a ½ inch thick steel plate. The steel plate has an 8 bolt-hole pattern to accommodate the yaw bearing. The yaw bearing is a Gear Tek swing bearing. This is a ball bearing assembly that is capable of withstanding large moment, radial, and thrust loads. The bearing will be mounted to the tower on a 10 bolt-hole pattern. The yawing of the entire nacelle is controlled by the Cal Poly

mustang design tail vane. A boom connects the tail vane to the nacelle frame. The tail boom is bolted on the nacelle frame, and is designed to take high moment loads due to wind forces.

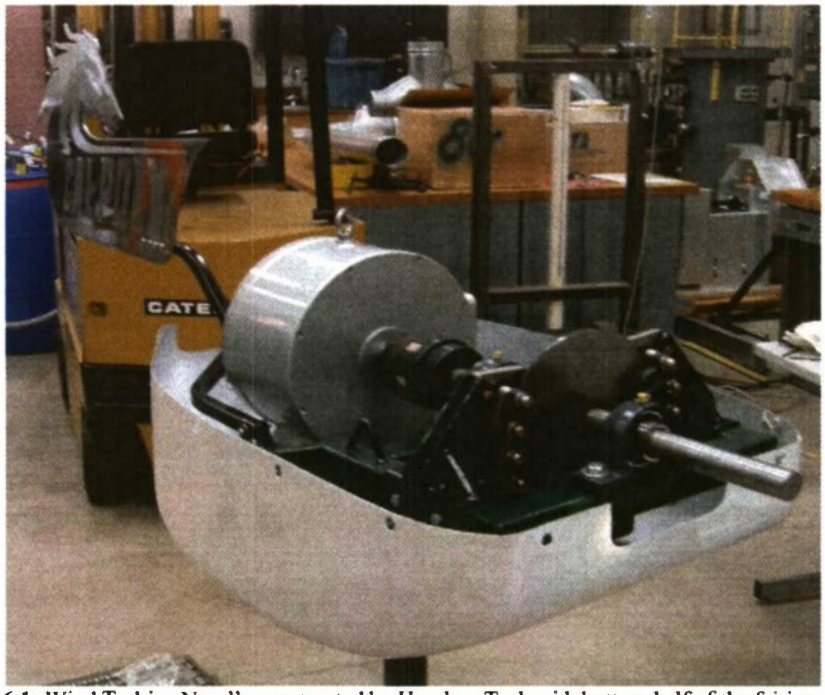
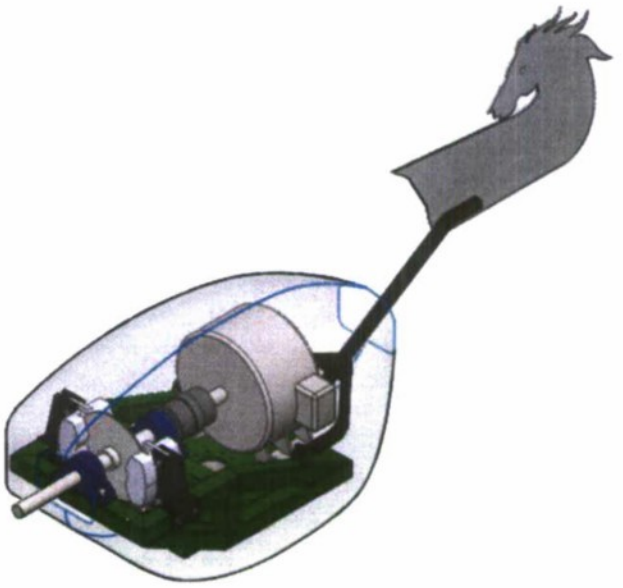


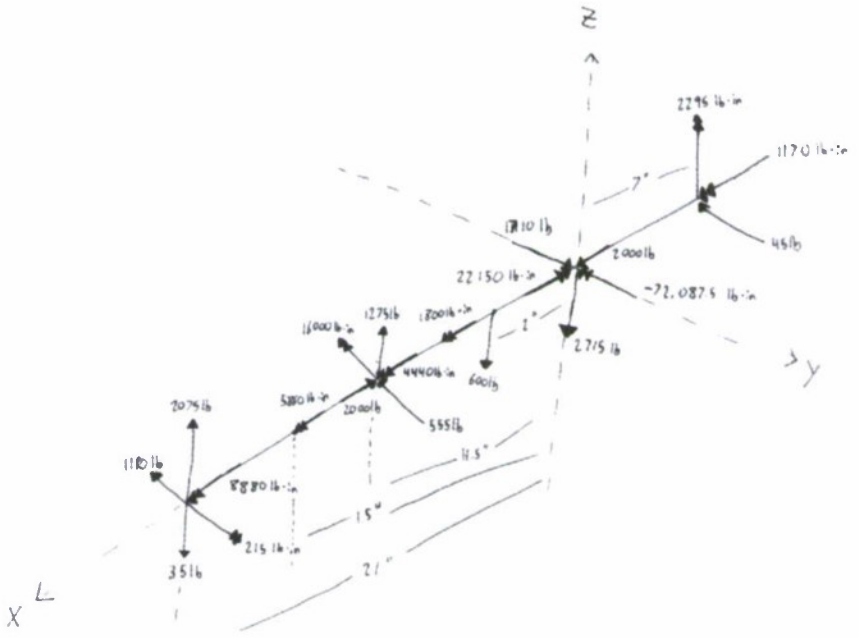
Figure 6.1: Wind Turbine Nacelle constructed by Hurakan Tech with bottom half of the fairing present.

6.1. Base Structure

The base structure depends primarily on the forces the WTN will experience during operation. Figure 6.1.1 (a) and (b) show the Solid works model and free body diagram of the WTN, respectively, in steady state operation with an unbalanced load while braking and yawing. The unbalanced load is modeled using the 5-5 rule explained in more detail in the drivetrain section. The case depicted in figure 6.1.1 is a worst-case scenario that acted as the design condition. For simplicity the mass of all the components are lumped together and applied at the WTN center of mass.



(a)



(b)

Figure 6.1.1: (a) Isometric view of the entire WTN assembly (b) FBD of the WTN. The system is taken to be the shaft, generator and yaw bearing. Center of the yaw bearing is set at the origin. Point 1 is the location of the first bearing located closest to the rotor and Point 2 is the location of the second bearing located closest to the generator.

In figure 6.1.1 (b) it should be noted that the forces at 1 and 2 in the Z and Y direction are due to the reactions in the bearings and alternate with respect to each other. The values given in figure 6.1.1 (b) are the maximum values that occur during the WTN operation. For the purpose of designing the frame, the loads were directed in the direction that produced the maximum stress in the orientation expected during steady state operation. The thrust load does not change direction relative to the WTN orientation during steady state operation. The thrust load will always be acting in the negative X direction causing a moment about the negative Y-axis according to the right hand rule. By having the reaction forces at 1 and 2 acting in the positive Z direction they add to the moment caused by the thrust load and produce a worst case scenario for the analysis. The fatigue is due to the unbalanced load rotating at 300 RPM causing the alternating reactions at 1 and 2 previously mentioned.

For the design of the frame the two main design criteria were deflection and fatigue. The base structure is designed to be stiff enough so that the bearings, brake calipers, and generator are not moving excessively compromising the performance of the WTN. The maximum allowable deflection was determined from the caliper opening minus the brake disc thickness divided by two, 0.039 inches. This clearance ensured that the disc would not slip against the calipers during operation.

Figure 6.1.2: Final manufactured base.

Figure 6.1.2 is an image of the final base structure. The base structure is fabricated from 1/8 inch thick 2 inch square A36 steel tubes welded together and a ½ inch thick A36 steel plate. Low carbon steel was selected due to the ease of manufacturability, stiffness and cost. The base structure also contains six tabs on the bottom to attach to the fairing, three ear tabs on the rear to mount the tail boom, two ½ inch steel triangles welded to give a location to hoist the WTN and two 1 inch thick aluminum spacers bolted between the frame and generator to align the shaft of the generator with the drivetrain shaft and allow for clearance under the generator to run wires.

For the analysis of the frame, hand calculations were first performed using beam deflection theory to determine the maximum deflection of the frame under the given design loads, Appendix H. In the analysis the plate was ignored and only the steel tubes were analyzed to simplify the analysis and add a design factor of safety. The hand calculations showed that maximum deflection of the frame is equal to 0.02041 inches. Next a finite element (FE) model

was created in Abaqus, figure 6.1.3. The model uses 3-D extrusions for the members and connects them using a pin joint. The reason for using pin joints over weld joints is to add an additional factor of safety in the analysis. With pin joints the load will be carried primarily by the members rather than the welds. This meant that a weld analysis was not necessary and the welds did not have to be done by a certified welder to ensure consistency with the analysis.

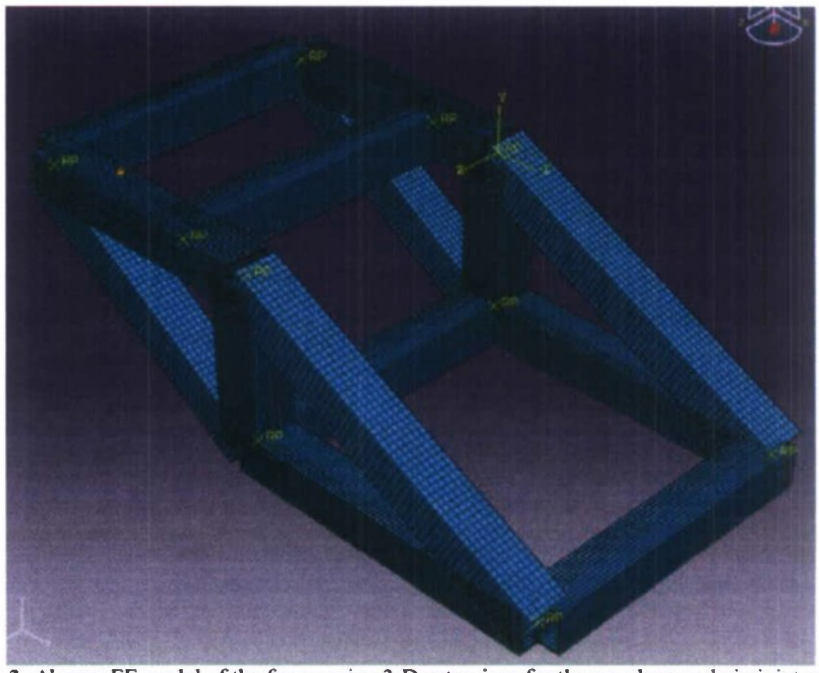


Figure 6.1.3: Abaqus FE model of the frame using 3-D extrusions for the members and pin joint connections.

From the FEA the two critical locations were found to be the mounting location of the rear bearing and the location where the base structure is fixed to the tower. Figure 6.1.4 illustrates this more clearly with a FEA performed in Solidworks. The Solidworks FE model applied a 2000 pound design thrust load to the member supporting the bearing closest to the generator. At this location the bearing pillow block is bolted directly to a steel tube. The location of the yaw bearing is where the bending moment is largest due to the increased stiffness. The yaw bearing is mounted to a 1/2 inch thick steel plate as specified by the manufacturer at Gear Tek. From the FEA performed in Abaqus the von misses stress has a maximum value of 23 ksi due to a stress concentration on the bolt-pattern for the yaw bearing. The von misses stress at the location where the rear bearing is mounted is equal to 15 ksi.

For the dynamic analysis performed in Abaqus the reaction forces at the bearings were alternating with respect to one another at a rate of 300 RPM. The fatigue strength of the frame was found to be 18 ksi and the first harmonic frequency is equal to 2435 Hz.

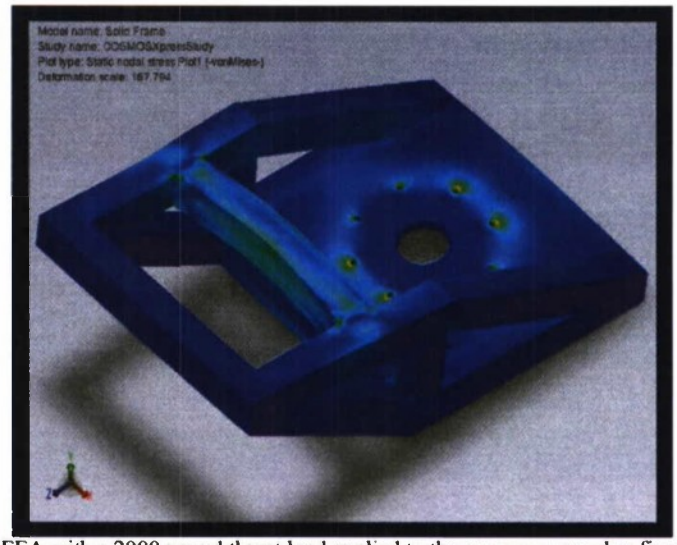


Figure 6.1.4: Solidworks FEA with a 2000 pound thrust load applied to the rear cross member fixed at the 8 hole bolt-pattern.

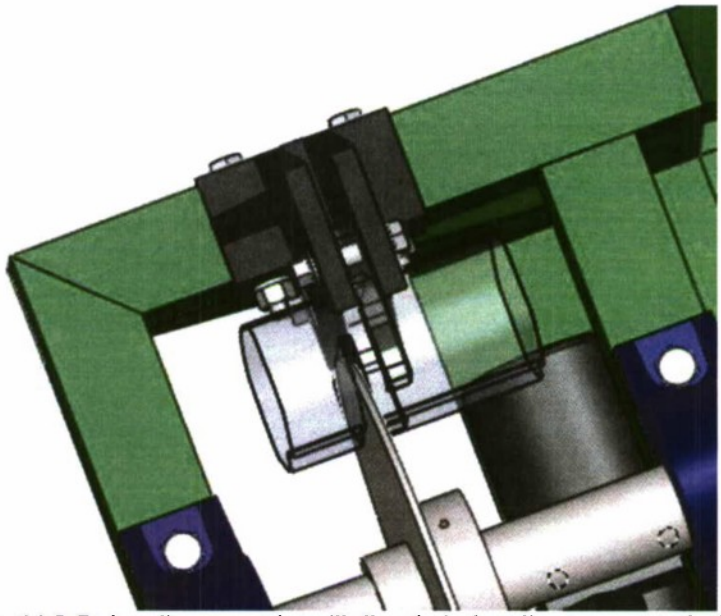


Figure 6.1.5: Brake caliper mount that will allow the brake caliper to mount to the frame.

Figure 6.1.5 is an image of the brake caliper mount and the brake caliper. The mount is shown in black and the brake caliper is transparent for a better view of the mounting structure. The brake caliper mount is fabricated out of 1/8 inch steel plate welded into a C-channel to bolt onto the frame. Two 1/2 inch thick steel plates are used for the main structural component to bolt to the caliper and a 1/2 steel gusset for added strength when the brakes are applied. The mount is bolted to the frame with two 7/16 inch bolts. Bolts were selected over welding to allow the mounts to be replaced easily in the future if a different caliper is installed.

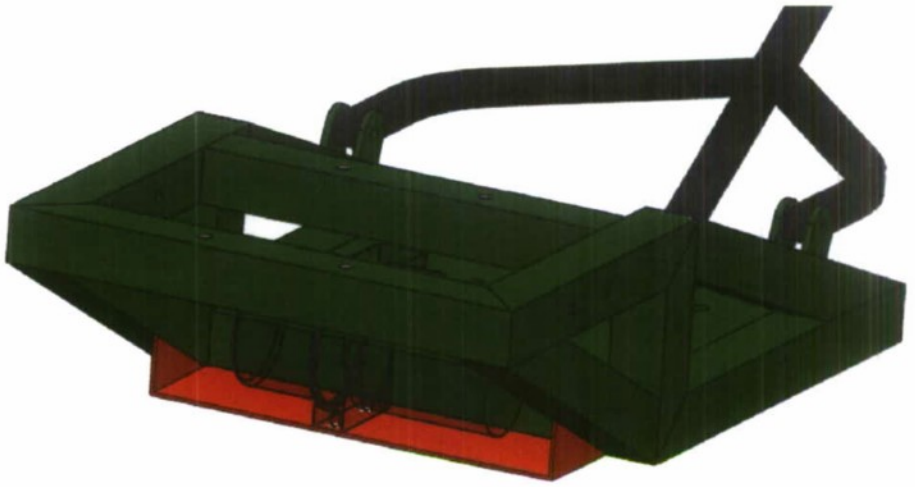


Figure 6.1.6: Hydraulic power unit mount that attaches the hydraulic power unit to the frame.

Figure 6.1.6 shows an image of the hydraulic power unit mount. The mount is red and the power unit is transparent for a better view of the mounting structure. More information on the hydraulic power unit can be found in the overspeed protection section. The mount for the power unit is welded to the frame out of 11 gauge steel plate. The mount has enough space to even allow for extra space for a controller for future applications or upgrades.

The entire structure; steel square tubes, steel plate, hydraulic power unit mount, brake caliper mounts, tabs and hooks are 29.75 inches by 19 inches by 8.15 inches and weigh a total of 87 pounds.

6.2. Yaw System

The yaw system of the WTN is a passive system and consists of a yaw bearing, tail vane, and a tail boom that connects the tail vane to the frame.

6.2.1. Yaw Bearing

The yaw bearing of the WTN will experience loads due to the thrust load, wind loads, startup and shutdown of the WTN, and the weight of the WTN. The free body diagram (FBD) shown in Figure 6.2.1.1 was used to determine the loads likely to be seen by the yaw bearing.

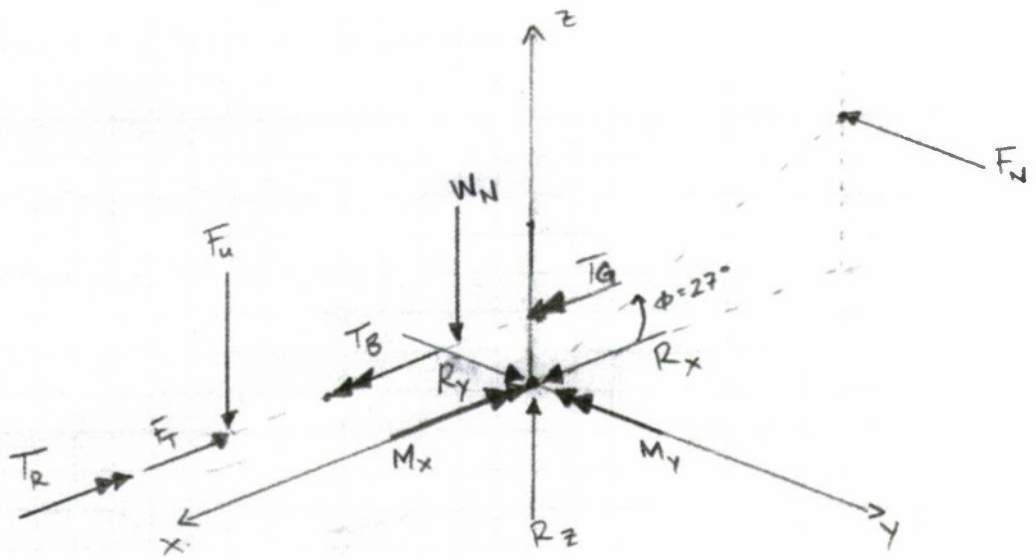


Figure 6.2.1.1: FBD of the WTN. The system is taken to be the shaft, generator and yaw bearing. Center of the yaw bearing is set at the origin.

Resolving the moments and forces to the axis of the yaw bearing yielded that the yaw bearing must have the capacity to operate under the following conditions:

Axial Loading: The yaw bearing is designed to withstand an axial load of 640 pound force. The main forces that will be supported in the axial direction by the yaw bearing are the weight of the WTN (W_N) and the force that is caused by the imbalance of the blades (F_u). F_u is an alternating load and its direction will change as the rotor rotates. For the purpose of determining the maximum axial load, the force F_u was taken to be acting in a downward direction (negative Z-direction). Also, the weight of the WTN was taken to be 600 pound force for design purposes. In actuality the weight is 460 pound force.

Radial Loading: The yaw bearing is designed to withstand a resultant radial force of 2000 pound force. The main radial force that the WTN will have to withstand is the 2000 pound force due to the wind thrust load (F_T). The load due to the wind acting on the tail vane (F_N) is negligible compared to the thrust load and can be neglected.

Moment: The yaw bearing will experience a resultant moment load of 778 ft-lbf. This was found by resolving all forces at the yaw bearing and accounting for the moments that each force will impart on the yaw bearing. To calculate this load, the unbalance load at the blades (F_u) was taken to act in the negative Y-direction. The torque generated by the rotor was neglected and it was assumed that the generator was producing a 150 ft-lbf torque (T_G) that was felt by the yaw bearing. These assumptions generated the largest moment at the yaw bearing for design purpose.

The bearing that has been selected for the WTN is the swing bearing, part number SY7B0101N-10-b, made by Gear Tek [15]. Gear Tek's swing bearing, figure 6.2.1.2 is a ball bearing assembly that is capable of withstanding large moment, radial, and thrust loads. Figure 6.2.1.3 is a static limiting load graph given by Gear Tek for the swing bearing.

Figure 6.2.1.2: Image of the Gear Tech swing bearing used in the WTN

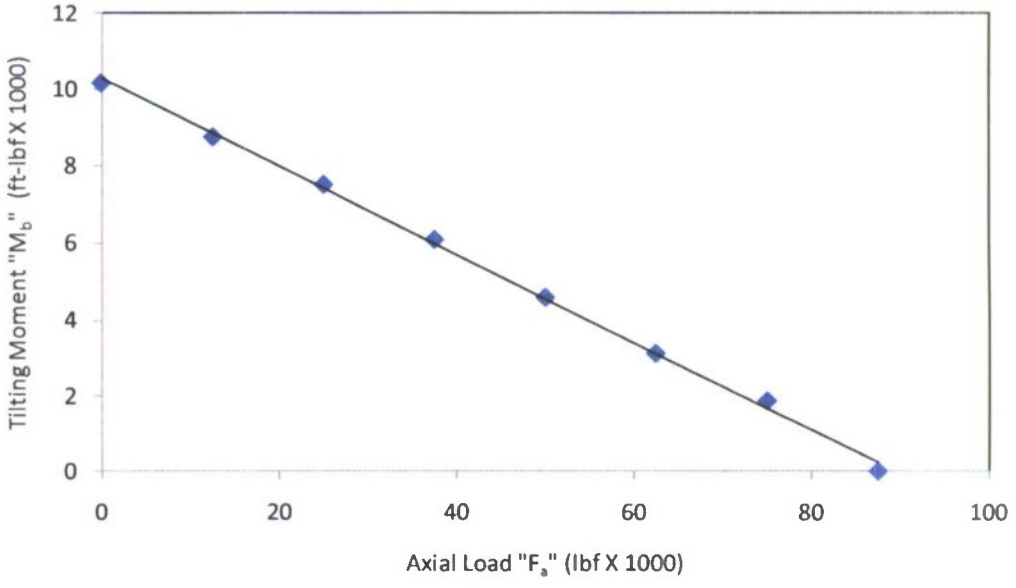


Figure 6.2.1.3: Static limiting load graph for the SYB0101N-10-b Gear Tek swing bearing.

From Figure 6.2.1.3 the static limiting load line equation was found to be:

$$M_b = -0.1144F_a + 10.264 \qquad (6.2.1.1)$$

The axial loading that the yaw bearing will experience, F_a is equal to 640 pound force. From equation 6.2.1.1, this corresponds to a maximum moment, M_b, of 10,191 ft-lbf that the bearing is

capable of supporting. This value is well above the maximum moment determined of 778 ft-lbf that the yaw bearing will experience under the worst case scenario presented. Based on this, the Gear Tek swing bearing was determined sufficient for use as the WTN yaw bearing.

The Gear Tek bearing also has an external gear ring that is not being used in the current design. This gear ring can be used to accommodate a pinion and yaw motor in the future to give the WTN an active yaw system.

6.2.2. Tail Vane

The tail vane incorporated into the WTN design is for the purpose of aligning the WTN into the headwind. The tail vane is light to reduce the weight of the WTN. Being that the tail vane is the sole source of yaw control for the WTN, it is strong enough to withstand the elements so that it does not brake causing the WTN to operate in an undesirable condition. A prominent characteristic considered throughout the analysis of the yaw system is its tendency to misalign. The maximum misalignment allowable is ±8° [4].

Table 6.2.2.1 is a decision matrix that compares various materials considered for use in the construction of the tail vane. The way the matrix is setup is as follows; a level of importance from 1-5 was given to each of the categories, 1 being low importance and 5 being high importance. Each material was then given a rating for how well it satisfied the given characteristics. A value of 1 corresponds to the material having a low correlation to the specific characteristic while a value of 5 corresponds to the material having a high correlation with specific characteristics. To get the total, the characteristic values were multiplied by the level of importance for the characteristics and the values were added across.

The top choice and the material used for the construction of the tail vane is aluminum. A sheet of 3/16 inch of aluminum was cut using a water jet cutter in Paso Robles from Advance Adapters to the shape of the mustang tail vane design, figures 6.2.2.1 and 6.2.2.2.

Once the tail vane was cut it was polished and etched with 60 grit sand paper. The lettering is only on one side and the mustang is on both.

Table 6.2.2.1: Decision Matrix for the WTN Tail Vane Comparing Different Material Options for Fabrication.

	Strength	Lightweight	Stiffness	Ease of Construction	Price	Withstand Elements	Total
Importance	4	3	4	4	2	5	
Fiber Glass With High Density Foam Core	2	5	4	2	4	4	75
Carbon Fiber With High Density Foam Core	4	5	4	2	1	4	77
Stainless Steel Sheet Metal	2	3	3	3	2	2	52
Aluminum	4	2	5	5	3	3	78
Steel	5	1	5	5	2	2	72
Rubber	1	5	1	1	5	1	41
Polyethylene Terephthalate Plastic	2	4	1	4	5	4	65
Clay	2	2	5	2	5	3	75
Wood	4	3	4	5	4	2	74
Canvas With Steel Frame	1	5	1	5	4	2	56

Figure 6.2.2.1: Proposed Cal Poly Mustang logo to be used for the tail vane.

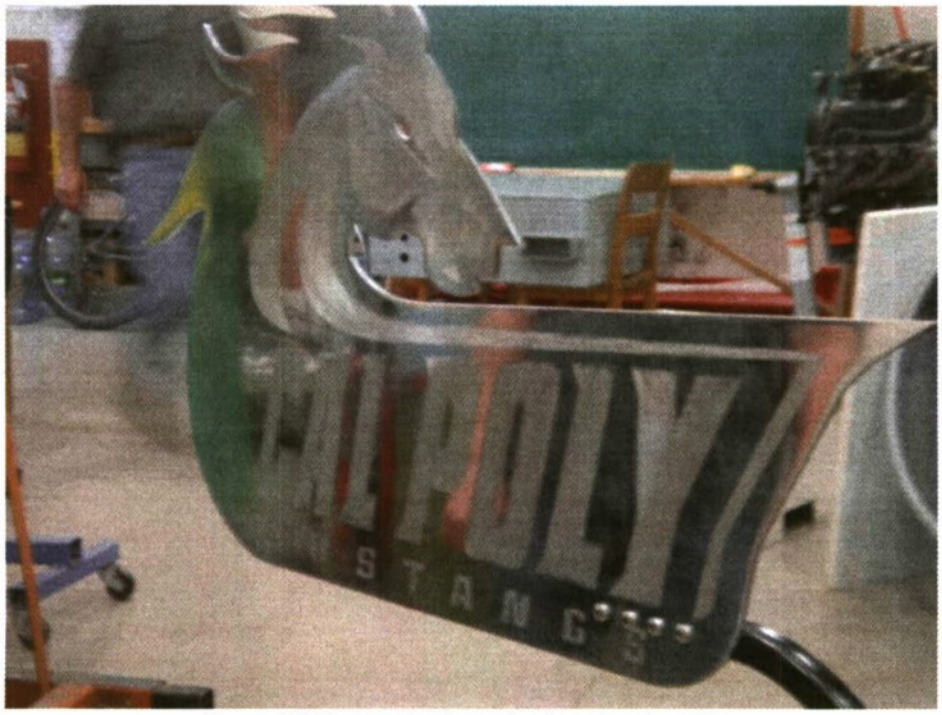


Figure 6.2.2.2: Actual tail vane that is used on the WTN.

6.2.3. Tail Boom

The tail boom is the connection for the tail vane to the base structure. The largest load that the tail boom will experience is due to the wind force acting on the tail vane. This load has been calculated to be 45 pound force, Appendix AA. Static analysis of the tail boom has given a factor of safety of 1.86 for the structure. The tail boom is a three member structure that was bent and welded together using 1 inch 11 GA A513 square steel tubing. Three 3/8 inch bolts support the structure to the base and four ¼ bolts support the tail vane to the tail boom.

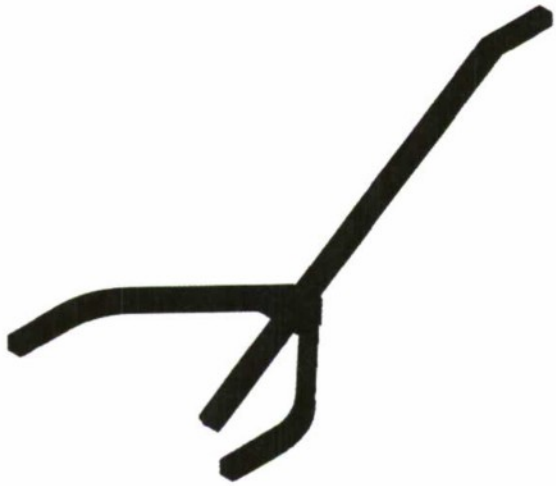


Figure 6.2.3.1: Tail boom mounted to the frame of the WTN.

Figure 6.2.3.2 shows the FE model made in Solidworks to determine maximum stresses seen by the tail vane with a 45 pound load representing the wind and fixed at the three bolt holes used to mount to the frame.

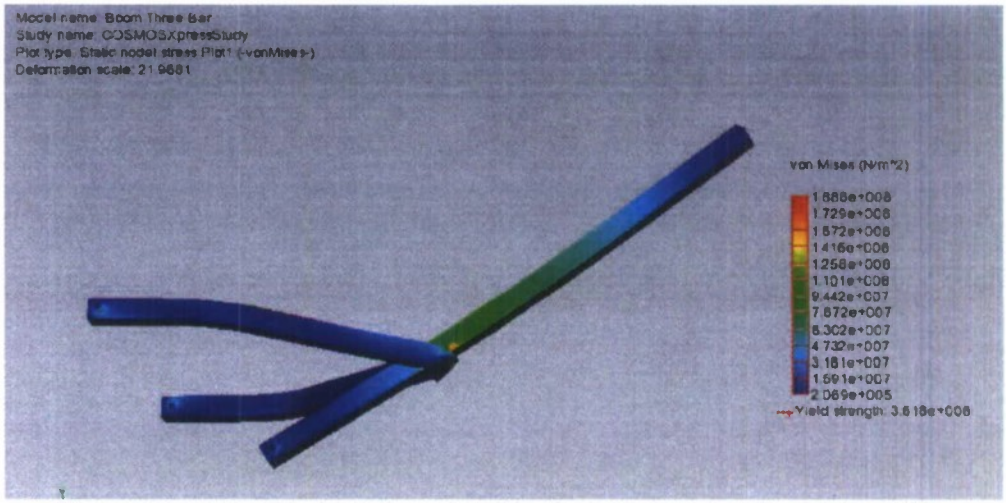


Figure 6.2.3.2: FE model of the tail vane in Solidworks. Von misses stress equal to 188.6 MPa.

6.3. Fairing

The function of the fairing is to cover and protect the components of the WTN. The fairing is durable enough to withstand impact forces, but also lightweight to minimize the stress on the yaw bearing. The enclosure also functions as the main form of heat removal from within the WTN. The aerodynamics of the design are such to minimize boundary layer separation as much as possible to reduce the drag force on the entire structure. Figure 4.2.2 shows a picture of the fairing.

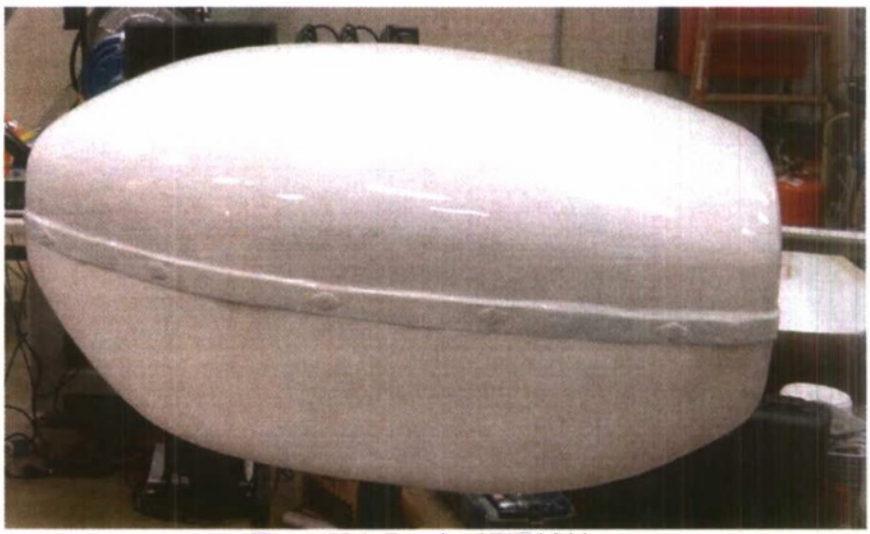


Figure 6.3.1: Completed WTN fairing.

Modified twill fiberglass was chosen as the material for construction of the fairing. This choice is based on fiberglass's high strength relative to the application, the ability to easily form fiberglass through use of molds, and it's lightweight.

Fabricating the fairing was a four step process step one was to create a foam plug. The foam plug was made out of two pound density foam and hand shaped to the dimensions of the Solidworks model, figure 6.3.2. Once the foam plug was shaped it was coated with bondo and sanded smooth, figure 6.3.3.

Figure 6.3.2: Alvaro Martinez shaping the foam plug of the fairing to be used for the WTN.

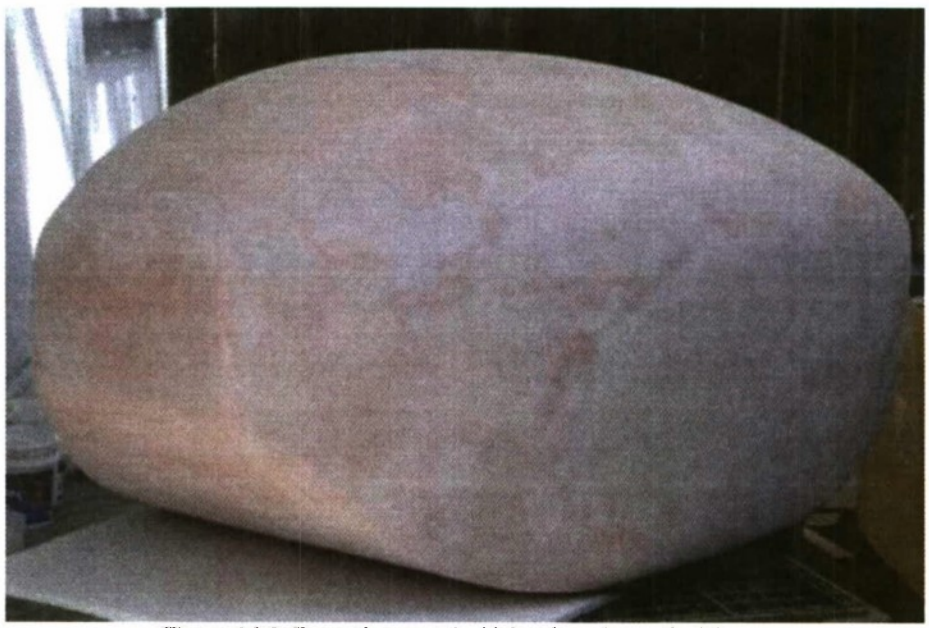


Figure 6.3.3: Foam plug coated with bondo and smoothed down.

Steps two involved making two female molds, top half and bottom half, of the fairing using the foam plug from step one. The mold was made with 1.5 oz fiberglass mat and polyester resin, figure 6.3.4. Once cured the mold was pulled off the plug and reinforced with fiber board to keep its shape. The inside of the mold was prepped before the part was made. Preparation of the mold included applying five coats of carnauba wax and then spraying polyvinyl alcohol (pva) over the wax. The wax and the pva helped in removing the part from the mold. The final stage of prep was to apply a gel coat to the inside of the mold. The gel coat is the main source of protection from the weather for the fairing. The gel coat prevents moisture from penetrating into the fiberglass fairing causing degradation. The coats of wax and pva will prevent the gel coat from bonding to the polyester resin used in the mold fabrication.

Figure 6.3.4: Image of the top half of the mold used for the WTN fairing.

Step three involved creating the actual parts using the molds. Once the gel coat dried, the actual part was made. The fiberglass selected for the fairing is a 7725 modified twill fiberglass. The twill is a woven cloth and was fairly easy to work with. Being that it is a heavier fabric, 8.8 oz / sq. yard, it provides good strength and rigidity. The resin used in constructing the fairing is a vinyl-ester resin. This specific resin has low water absorption rating and good adhesive properties which provide a strong bond to hold the fiberglass together. Four layers of fiberglass were laid into the mold using a vacuum infusion layup, figure 6.3.5. Once cured, the final parts were pulled out of the mold. Since the gel coat was blocked from bonding to the mold by the wax and the pva, it bonded to the vinyl-ester resin used in the fairing fabrication and came out of the mold with the fairing.

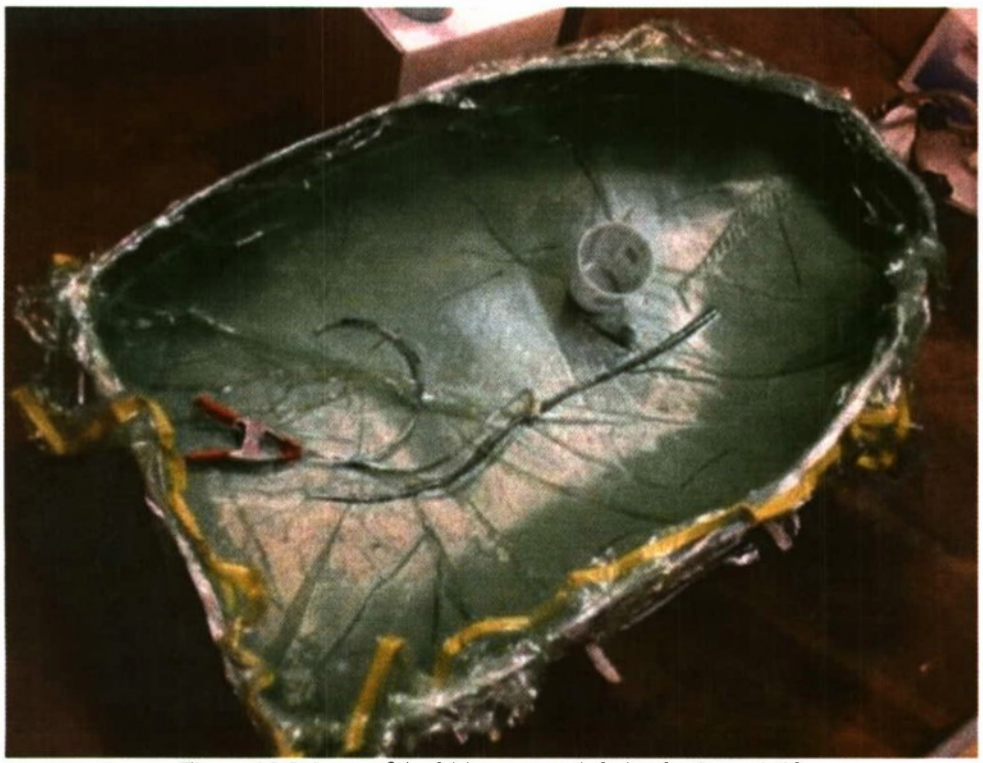


Figure 6.3.5: Image of the fairing vacuum infusion for the top half.

Step four of the fairing was to finalize the parts in order to get them ready to attach to one another and the frame. The bottom half of the fairing was cut in half lengthwise and three holes were cut into the fairing; one on the bottom for the tower, one in the front for drivetrain to extend out to attach to the spinner and one in the rear to allow the tail boom to extend out. The fairing attaches together with ½ inch stud quarter turn fasteners, the same type used for formula race car fairings. The fairing is sealed with weather stripping to prevent the fairing from getting scratched when it is assembled and prevent water from entering in through the cracks. Six 1 inch thick blocks of wood are attached to the fairing lining up with the six protruding tabs off the frame and attached to the fairing with two layers of fiberglass.

6.4. Drivetrain

In the design of the drivetrain consideration was given to how the brakes, generator, coupling, rotor, and bearings affect its overall performance, and design requirements.

The drivetrain layout consists of five primary elements that were thoroughly analyzed. The drive shaft was the main focus throughout the analysis as it influences, and in some cases, drives the design of some of the components. The rotor is connected to the drive-shaft via a taper lock. The end of the shaft was left turned down to 1.5 inches. Two bearings isolate the generator from axial, radial, and moment loads. A coupling joins the drive-shaft and generator as it transfers torque produced from the rotor. Analysis began with assuming a shaft size and determining a safety factor. Because the generator a 40 millimeter (approximately 1.5 inch) diameter drive-shaft, analysis of the shaft began assuming a 40 millimeter diameter shaft.

In the shaft design consideration was given to both static and dynamic loading. For both cases a clear analysis was performed to ensure failure would not occur. Static loading analysis looked at how the shaft is loaded once all sub-systems and sub-components are integrated to form the entire drivetrain. Bearing reaction forces and loads from sub-system and sub-component weights were the primary focus. Basic shaft design equations and methodology are referenced from "Shigley's Mechanical Engineering Design" [3].

A basic static analysis of the shaft accounting for the axial load on the shaft produced from the thrust and the torque transferred from the rotor to the generator produced factors of safety and provided a 40 millimeter diameter shaft performance. As seen on table 6.4.1, torque was the driver in the design. A factor of safety of 11 seems high, but as stated before, consideration to fatigue, notch concentrations, and forces such as the braking torque are not taken into account. Figure 6.4.1 provides more details on the calculations of these safety factors.

Table 6.4.1: Factors of safety from obtained from simple analysis of 4140 anncaled steel

Factors of Safety	
Basic Analysis of Shaft Performance	
Axial	54
Bending	52
Torsion	11

```
%Safety Factor due to Torsion
        J = (pi*(dia)^4)/32;
        T = 150*12;
        r = 1.5/2
        taumax = T*r/J;

n = Sy/(2*taumax)        %Factor of Safety

%Safety Factor due to Bending
        I = (pi*(dia)^4)/64;
        moment = 76*8;
        sigma_a = moment*r/I;

nbend = Sut/sigma_a;     %Factor of Safety

%Safety factor due to Axial
        Fthrust = 2000;
        A = pi*(dia^2)/4;
        sigma_axial = Fthrust/A;

naxial = Sy/sigma_axial %Factor of Safety
```

Figure 6.4.1: Matlab code used to determine factors of safety

The shaft is sized for max loads from wind load data (Appendix B); also bearing mounts were specified based on this information. The drivetrain connects to the generator through a coupling. This connection is specified under Appendix C in the technical drawings for the generator using a 12 mm key. The drivetrain was designed to prevent axial thrust loads from being felt by the generator while maintaining the integrity of the entire drivetrain. Referring to IEC 61400-1 all

design loads contain a minimum factor of safety of at least 1.35 [4], and "non fail-safe" mechanical/structural components with non-ductile behavior have a general factor of safety no less than 1.3 for ruptures from exceeding tensile or compression strength.

In the case of dynamic loading a more extensive analysis was performed. A fatigue analysis accounting for fatigue induced from rotation at steady state operation. Also vibrations analysis was required to account for mass imbalances caused by imperfection in the rotor hub, blades, and any other sub-system or sub-component directly attached to the shaft. The system is designed to either completely eliminate or significantly minimizes thrust interactions between the generator and the shaft.

Determination of an appropriate shaft size and design of the shaft accounts for all forces and factors. The analysis was conducted using a worst case scenario accounting for loads produced by a rotating imbalance at the rotor, and breaking torque via fatigue, and startup and shutdown operation.

Dr. Patrick Lemieux provided a 5-5 rule, Appendix C and L, as a worst case scenario and was used to determine the forces acting at the rotor and the reactions at the bearings. Figure 6.4.2 provides the Matlab code used to determine the resulting force from the mass unbalance.

```
%Vibrations Analysis of Unbalance Force


m = Mb*.05;        %Mass of imbalance (slugs)
R = f;             %Distance from shaft center line to imbalance (ft)
Ws = Wx;           %Operating speed of shaft (rad/sec)

famp = m*R*Ws^2;   %Amplitude of force caused by imbalance in z-direction(lbf)
```

Figure 6.4.2: Determination of the unbalance force produced by the 5%-5% worst case scenario.

6.4.1. Fatigue Analysis

Based the wind loads provided by Dr. Lemieux, alternating moments and torque, and mean torque were found. These loads were used to calculate fluctuating stresses due to bending and torsion on the shaft.

$$\sigma_a = \frac{K_f M_a c}{I} \tag{6.4.1.1}$$

$$\tau_a = \frac{K_{fs} T_a c}{J} \tag{6.4.1.2}$$

$$\sigma_m = \frac{K_f M_m c}{I} \tag{6.4.1.3}$$

$$\tau_m = \frac{K_{fs} M_m c}{I} \tag{6.4.1.4}$$

These stresses were combined with the distortion energy failure theory, to obtain von misses stresses for rotating solid shafts. These stresses in conjunction with the modified Goodman line generated an acceptable shaft diameter. Modifying factors, for different cases such as reliability, loading factor, etc., were applied. Please refer to Appendix O-Q for a more detailed analysis of the shaft loads. The following table presents the fatigue stresses calculated.

Table 6.4.1.1: Fatigue stresses due to alternating and mean bending moments and applied torques.

Shaft Stresses					
Ta (in-lb)	2940	Tm (in-lb)	2940	Ma (in-lb)	649.6
Kf	1.6096	D (in)	1.5	Mm (in-lb)	0
Kfs	1	I (in4)	0.2485		

Stresses due to bending			Stresses due to torsion		
σ_a	3155.66075		τ_a	8873.064916	
σ_m	0		τ_m	8873.064916	

Theses bending and torsion stresses showed that torsion effects induced by torque loads were the driver of the shaft design. Comparing these bending and torsion stresses with the ones calculated using the suggested formulas provided in IEC 61400-2, Table 6.4.1.2, to the von misses stresses, Table 6.4.1.1, it was evident that the von misses stresses due to torsion being used to design the shaft were larger than the design stress due to torsion obtained using IEC 61400-2 formulas, therefore the design von misses stresses provide a relatively larger safety factor to prevent failure.

Table 6.4.1.2: IEC suggested design stresses

Moment due to bending	M_{shaft} (in-lb)	649.6
Torque applied to shaft	$M_{x\text{-}shaft}$ (in-lb)	1800
radius/moment of inertia	W_{shaft} (in^3)	1.27389
Bending		
	$\sigma_{m\text{-}shaft}$ (lb/in^2)	477.28
Torsion		
	$\tau_{m\text{-}shaft}$ (lb/in^2)	1413

Table 6.4.1.3: Calculated diameter based on 4140 annealed steel for two load cases

Case 1 - Alternating breaking torque applied over 10^6 cycles			
Sut (lb/in^2)	95000	Ta (lb-in)	2940
n	2	Tm (lb-in)	2940
ka	0.702	Ma (lb-in)	649.6
kb	0.842		
kc	1		
kd	1		
ke	0.868		
Kf	1.6096		
Kfs	1		
se'	47500		
se	24370.39332		
d (in)	1.330340432		

Case 2 - Alternating torque due to wind speed on rotor applied over 10^6 cycles			
Sut (lb/in^2)	95000	Ta (lb-in)	900
n	2	Tm (lb-in)	900
ka	0.702	Ma (lb-in)	649.6
kb	0.842		
kc	1		
kd	1		
ke	0.868		
kf	1.6096		
kfs	1		
se'	47500		
se	24370.3933		
d (in)	1.03433109		

Table 6.4.1.3 presents calculated shaft diameters based on two different load cases. Case 1 took into account the torque produced when a brake force is applied to the disk brake. Assuming this action will take place an infinite amount of cycles, the torque load will fluctuate between 5880 in-lb of torque when braking is applied to 0 in-lb of torque when the shaft comes to a complete stop. Case 2 focused on a situation where no braking torque was present, but rather the turbine experienced start-up and shut-down loads due to wind speeds reaching 0 mph at complete shut down and 40 mph at steady state operation. These wind speeds produced a fluctuating torque between 0 in-lb at complete shutdown to 1800 in-lb at steady state operation. Analyzing cases 1 and 2 while applying conservative fatigue failure theory generated a 1.33 inch nominal shaft diameter.

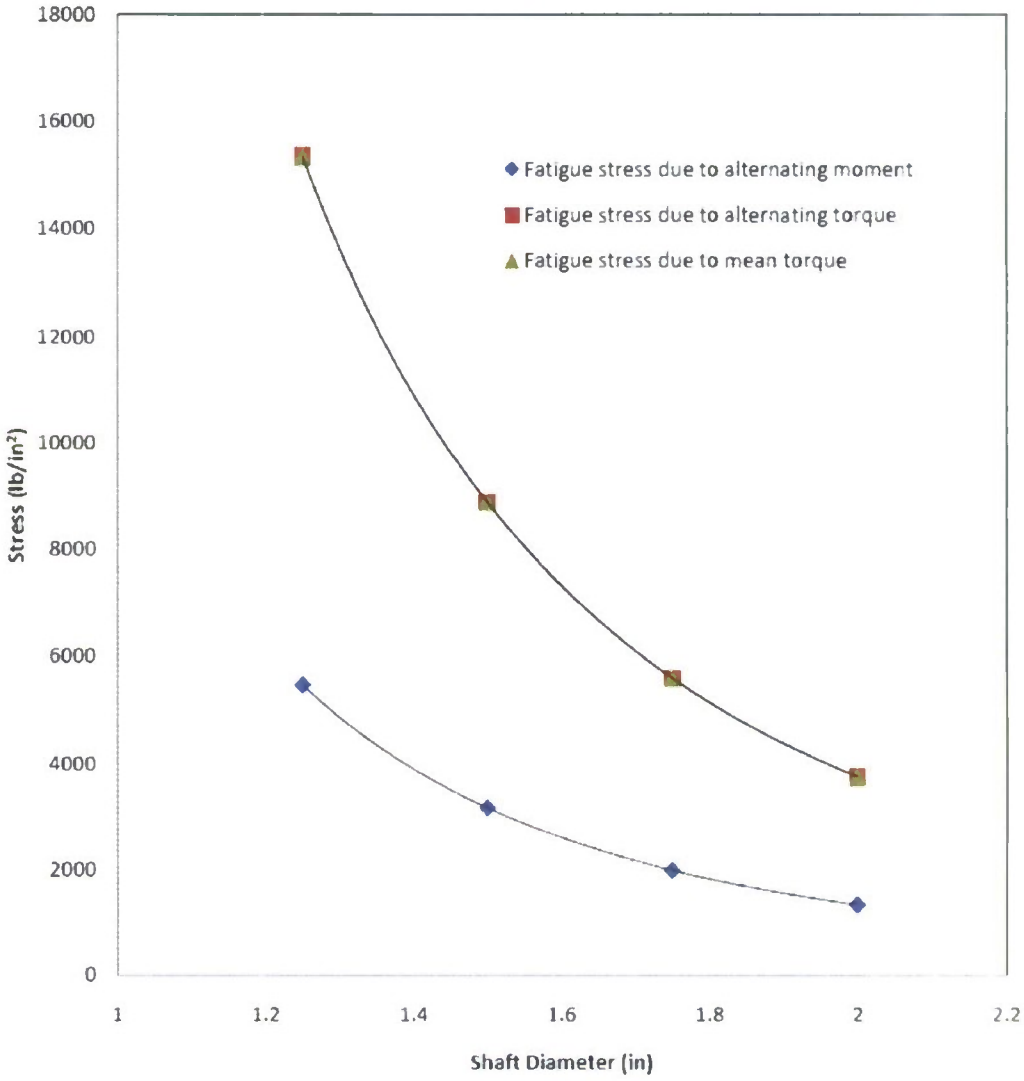


Figure 6.4.1.1: Graph of fatigue stress alternating and mean loads as a function of the shaft diameter.

Figure 6.4.1.1 provided information regarding the shaft diameter effects on the fatigue stresses experienced at the critical point on the shaft. It is visible that as the size of the shaft increase the stresses experienced decrease. Because the stress decreases as the size of the shaft increases, it was safe to assume that a larger shaft size compared to the calculated shaft size could be used in the design of the WTN. Recalling that the generator has a 40 millimeter diameter shaft it was safe to assume the drivetrain required a shaft diameter of 40 millimeters or greater. Taking into consideration the analysis done to determine the shaft size a shaft of 1.5 inches in diameter with an internal step of 1.75 inches was selected.

Finally the shafts critical speed was found to determine if the drive-shafts first mode of vibration would be reached during operation. The natural frequency was found to be 24262 RPM; therefore any failure due to the resonance was ignored since the operating speed of the shaft is

designed to never exceed 300 rpm. Figure 6.4.1.2 provides the Matlab code and analysis used to determine the critical speed of the shaft.

```
%Determination of shaft critical speed

Wtot = 3*Wb + Wr + Wshaft;   %Weight of shaft, rotor, and blades (lbf)
D =1.5;                      %Ideal shaft diameter(in)
F = Wr+3*Wb;                 %Force due to weight of rotor and blades (lbf)
l = 3;                       %Distance from bearing A to Force F (in)
E = 29*10^6;                 %Modulous of elasticity for low carbon steel
I = (pi*D^4)/64;             %Sahft area moment of inertia(in^4)
ymax = (-F*l^3)/(3*E*I);     %Maximum deflection (in)
K = (3*E*I)/(l^3);           %Sahft equivalent spring constant (lbf/in)
Wn = ((((K*386.4)/Wtot)^.5)/(2*pi))*60; %Shaft critical speed (rpm)
```

Figure 6.4.1.2: Matlab code used to determine the shaft critical speed.

6.4.2. Bearing Analysis

The bearings for the drivetrain are required to withstand loads in the axial and radial directions. A comparison between single-row, tapered roller bearings and single-row ball bearings was performed. The first stage of bearing analysis determined the reaction forces at both bearing locations. The reaction forces were found using 3-D angular momentum equations. The free body diagram in Figure 6.4.2.1 shows the reaction at both bearings.

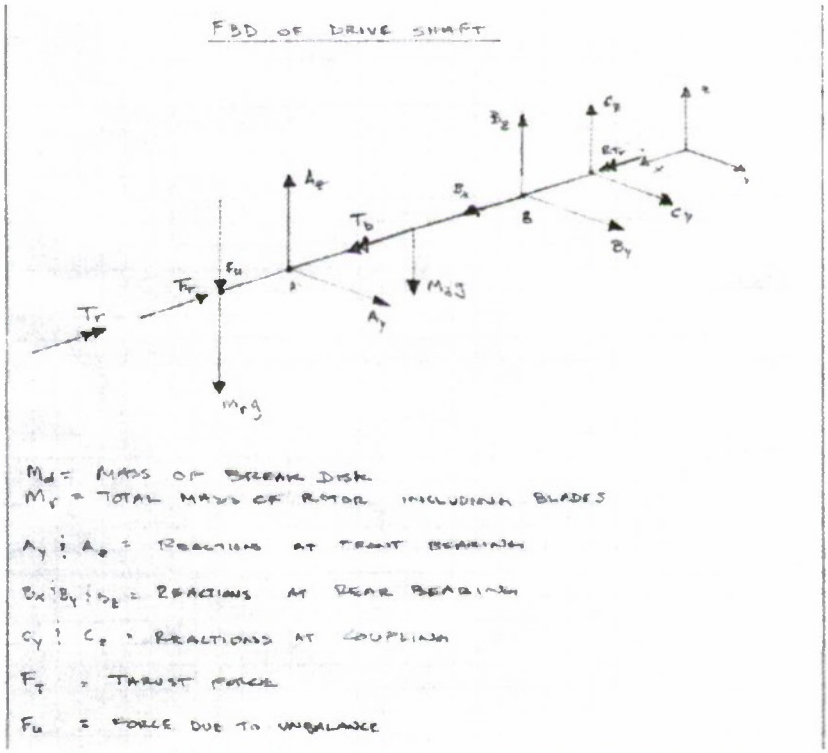


Figure 6.4.2.1: FBD of WTN providing reaction forces at bearings 1 and 2.

These reactions were found by applying the moment relations of equation 7/23 from Engineering Mechanics Dynamics by Meriam and Kraige for rigid body motion with the axes attached to the

body, and assuming parallel plane motion with the products of inertia of the components on the shaft equal to zero. Figure 6.4.2.2 provides the Matlab code used to determine the reactionary forces.

```
Wr = 18;  %Weight of Rotor (lbs)
Wd = 13;  %wiehgt of disk brake (lbs)
Wb = 8;   %Weight of blades (lbs)
Wbi = .05*8+8; %weight of inbalanced blade
Wshaft = 5.93;

Mr = Wr/32.2;  %mass of rotor (slugs)
Md = Wd/32.2;  %mass of disk (slugs)
Mb = Wb/32.2;  %mass of blades (slugs)
Mbi = Mb*0.05 +Mb;

a = 7.16/12;     % distance from rear bearign to front bearing
b = 5.60/12;     % distance from front bearing to rotor
c = 10.75/12;    % distance from rear bearing to rotor
d = 2.08/12;     % distance from rear bearing to disk
e = 36/12;       %distance from centerline of shaft to point mass
f = .05*e +e;

Wx = 31.4;       %shaft speed (rad/sec)
g= 32.2;         %Acceleration due to gravity

theta1(1) = 0;  %Blade 1 angle
theta2(1) = 30; %Blade 2 angle
theta3(1) = 30; %Blade 3 angle (

for q = 1:360

   theta1(q+1) = theta1(q)+1;
   theta2(q+1) = theta2(q) + 1;
   theta3(q+1) = theta3(q) - 1;

%Mass moments of inertia and Reactions at bearings


    %Mass Moments of Inertia with respect to 'B'
        IAxy = Mb*(-c)*(e*cosd(theta1)) +Mb*(-e*sind(theta3))*(-c) + Mbi*(-f*sind(theta2))*(-c);

        IAzx = Mb*(-c)*(e*sind(theta1)) + Mb*(-c)*(-e*cosd(theta3)) + Mbi*(-c)*(f*cosd(theta2));

    %Reactions at Bearing 'A'
        Az = -(IAzx*Wx^2)/a;

        Ay = (IAxy*Wx + b*Mb*g +b*Mb*g + b*Mbi*g + b*Mr*g + d*Md*g)/a;


    %Mass Moments of Inertia with respect to 'A'
        IBxy = Mb*(-b)*(e*cosd(theta1)) +Mb*(-e*sind(theta3))*(-b) + Mbi*(-f*sind(theta2))*(-b);

        IBzx = Mb*(-b)*(e*sind(theta1)) + Mb*(-b)*(-e*cosd(theta3)) + Mbi*(-b)*(f*cosd(theta2));
    %Reactions at Bearing 'A'
        Bz = (-IBzx*Wx^2)/a;

        By = (IBxy*Wx + b*Mr*g +b*Mb*g + b*Mb*g + b*Mbi*g - d*Md*g)/a;
       end
```

Figure 6.4.2.2: Matlab code used to determine the reactionary forces at bearings 1 and 2 defined in the code as bearings A and B respectively.

Timken's engineering section catalog was used to determine the appropriate size bearings. As stated earlier single row tapered roller bearings were compared to single-row ball bearings. The comparison showed that single-row ball bearings are suitable for the drivetrain design. In the comparison an equivalent radial load due to axial and radial forces acting on the bearing was found and used to determine the bearings catalog load rating and life rating. Timken roller bearing life rating is based on a $90x10^6$ cycles. Timken ball bearing life is based on $1x10^6$ cycles. The following figures provide the reaction forces at the bearings, catalog load rating information,

life rating, and equivalent radial loads as a function position of the rotor blades for ball bearing analysis.

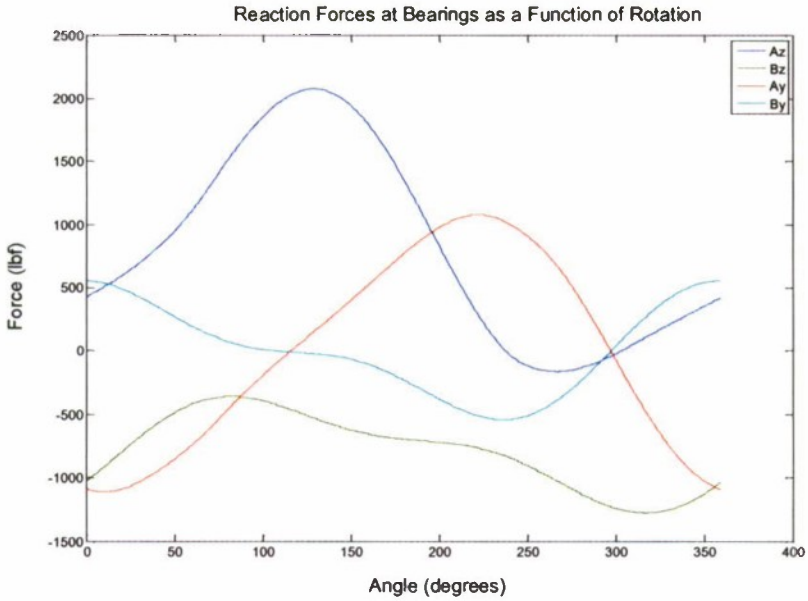


Figure 6.4.2.3: Graph of reactions at bearing as a function of position.

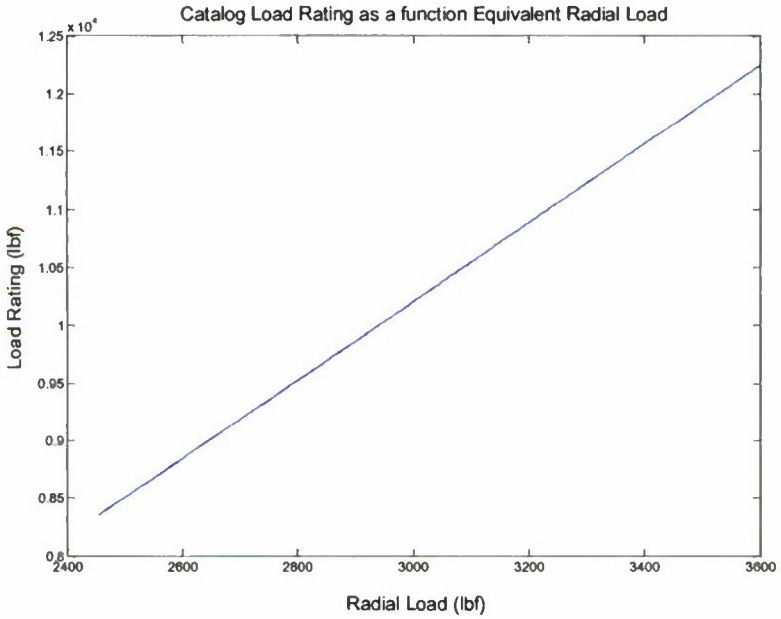


Figure 6.4.2.4: Graph of catalog load rating vs. radial load shows that at a maximum radial load of about 3600 pound force a bearing catalog load rating need to be under12500 pound force.

Based on the calculated maximum catalog rating of 12500 pound force Timken's 308W Medium Series bearing was chosen. The 308W Medium Series bearings catalog rating is 15000 pound force.

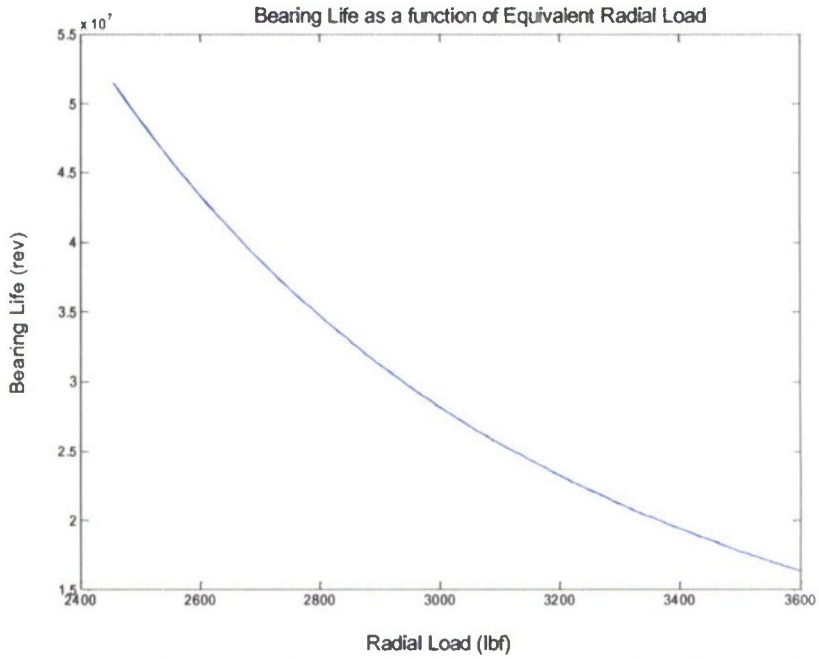


Figure 6.4.2.5: Graph of bearing life vs. radial load shows that at a maximum equivalent radial load of about 2500 pound force the life of the bearing is approximately 4.85×10^7 cycles.

6.4.3. Coupling

From analysis the torque loads on the coupling are no more than 4059 in-lbf and maximum operating speed of the shaft is 300 rpm, the coupling selected is the Lovejoy LC 190 Hytrel Jaw Type coupling. This coupling is rated for torque loads above 4600 in-lbf and is rated for a maximum operating speed of 5000 rpm. Please refer to Appendix U for more details.

6.5. Overspeed Protection

The purpose of the overspeed protection is to bring the rotor to a complete stop from any operating condition without damaging components of the WTN. The braking system will activate in the case that the following parameters operate out of their safety limits; power, rotor speed, connection of the electrical load or wind speed. The overspeed protection will also have a manual override command to allow a user to halt the wind turbine remotely for any instant including improper start-up and shutdown procedures, yaw misalignment, and any other reason that the operator feels the WTN should halt. Table 6.5.1 shows the decision matrix for overspeed designs.

The factors considered for the decision matrix are rated from 0-5 with a 5 meaning that it can perform best for the given parameter. The first column lists some of the initial ideas for overspeed protection. The overspeed can be divided into two categories, passive and active. Passive systems have the benefit that they do not need a separate monitoring device for actuation. If the WTN begins to operate outside of safe parameters then the passive overspeed device will activate and bring the rotor to a complete stop. One device that is used in industry [8] and IEEE [14] is an electro-dynamic braking system. An electro-dynamic braking system is a

circuit connected to the generator that increases the resistance of the generator once the generator begins producing a voltage that exceeds the design voltage output. This system has its benefits that it can protect the generator from overheating and is passive since the device is actuated as the voltage and amperage of the generator increase. However in the case that the rotor becomes disconnected from the generator the electro-dynamic braking system will not be able to bring the rotor to a stop since the generator will not be generating any power. Therefore electro-dynamic braking will not be included for the design of the WTN.

Under passive braking systems, electrical connection and friction are listed under centripetal acceleration. What centripetal acceleration refers to is a braking system that activates once the angular velocity of the rotor reaches a cut out angular speed. This cut our speed comes from technical specifications for the PL-PMG 3500. Once the cut out speed is reached, the system will force a rotating mass to move outward in the radial direction causing the rotor to stop. There are two ideas for how centripetal acceleration can stop the rotor. The first is for the objects to be battery packs with metal brushes that make contact with a metal drum. When the metal drum and battery packs come in contact, a circuit will be closed activating a separate braking system. This system is not necessarily passive since the actual braking force of the rotor remains a function of a separate braking system and not that of the metal brushes.

Table 6.5.1: Decision Matrix for overspeed protection devices for WTN.

	Means of isolating or discharging accumulated energy	Resistance to Rusting	Cracking	Scarring	Warping	No material properties degradation at high temperatures	Frictional performance	Heat Dissipation	Recover quickly From Immersion	Resistance to brake fade	Lubrication requirements	Life expectancy	simplicity	Weighted total
Importance	5	3	3	3	3	4	5	5	2	2	3	5	5	-
Passive														
Electro-Dynamic	5	5	5	5	5	4	0	3	0	5	5	3	0	156
Centripetal Acceleration														
Electrical Connection	2	4	5	5	5	5	0	5	0	5	4	2	0	144
Friction	2	3	3	3	1	3	4	3	2	2	2	2	3	126
Active														
Internal Drum	3	3	4	2	3	2	3	2	3	4	3	2	4	137
External Drum	2	2	4	2	3	2	3	2	3	4	3	2	4	129
Disc	4	3	3	4	2	3	3	3	5	4	4	4	5	173
Direct Shaft	2	3	3	4	1	2	2	1	3	3	3	1	5	117
Actuation														
Hydraulic	4	4	3	3	3	3	3	3	3	3	4	4	5	170
Pneumatic	4	2	3	3	3	3	3	2	3	3	2	3	5	148
Leverage	0	2	2	3	3	3	4	3	2	2	3	2	3	119
Electromagnetic	5	5	4	5	5	4	4	3	0	3	5	4	2	184

Another approach is similar to internal drum brake systems. Once the cut out angular velocity is reached a shoe will move outwards and make contact with a rotating drum. This system is better

because the actual braking system is included in the mechanism. The system can also be self energizing once the shoe comes in contact with the drum as long as the shoe is placed in the correct position relative to the direction of the rotors spin axis. If the rotor does spin in the opposite direction, e.g. tail vane fails and the blades move in a downwind position, then the system will become self de-energizing and will be unable to stop the rotor.

Another approach is to use an active braking system that acts in a passive manner. The key to having an active braking system act passive is to set the system so that the active mode allows the rotor to spin and the in-active mode stops the rotor. This would be similar to using spring applied brake calipers. The calipers are forced open and once energy is lost in the system the calipers will brake the rotor.

Having the braking torque directly applied to the shaft was another consideration. Direct torque application can assure that the shaft will stop, for disc and drums there is a possibility for the surface connected to the shaft to fail and the braking torque will have nothing to be applied to. Torque being applied directly to the shaft may damage the shaft, causing complications during various operating conditions of the WTN.

For the WTN a disc brake will be used as the main overspeed protection. Because the WTN needs to be as similar to large wind turbines that are used in industry a disc brake system is the most commonly used [13], [15], and [16].

For an active mechanical braking system an actuation is needed to energize the system. The three modes considered are hydraulic, pneumatic, and electric. Each of the three actuations will now be considered in more detail.

Pneumatic systems are rather loud while they are compressing the air. Because a pneumatic is an open system air is taken from the environment and compressed energize the brake. When the air is compressed moisture is condensed out of the air. The condensed water must be purged from the system periodically. An outlet will need to be added to the fairing design to ensure that excess moisture does not accumulate within the nacelle. Pneumatic devices cannot deliver the same pressure as hydraulic since air can compress more than oil. Finally a compressor needs to be lubricated periodically so either a large oil reservoir must be added to the compressor or the WTN must be lowered to be lubricated periodically.

Electric systems have more complicated actuators then pneumatic or hydraulic. An electric system also draws more power than a pump of compressor. For the WTN two calipers will be used for redundancy adding complexity in the actuation. Either two electric motors will be needed or a leverage system will be needed to split the actuation from a single motor to the two calipers. Electric actuation is cleaner then pneumatic and hydraulic because it requires less heat removal, does not rely on a fluid and are typically are more efficient. However for larger wind turbines a pressurized fluid is typically used over an electromagnetic actuation [13], [15], and [16].

Hydraulic systems are quieter than pneumatic systems and because the moving parts are submerged in oil they do not require periodic lubrication. Hydraulic devices have a larger range of temperature that they can operate in and will not freeze up like pneumatic systems. Hydraulic systems can provide more power than a pneumatic system and are typically more robust then

pneumatic systems. A hydraulic system is a closed system their fore if a leak occurs in the hydraulic line the system will de-energize whereas a pneumatic device will continue to draw power to run the compressor and attempt to re-energize the system. Also a hydraulic system does not need to purge condensed liquid so the system can be completely enclosed so long as sufficient heat is dissipated from the system.

The primary actuation for the brakes will be hydraulic due to its familiarity among the team and the current brake calipers considered require hydraulic actuation.

6.5.1. Calipers

For disc brakes three requirements were met; first the brakes must be able to supply sufficient torque to stop the WTN, second the disc is large enough to dissipate the heat generated during operation and finally the operating temperature of the disc must be acceptable to avoid brake fade and reduced performance. The design overspeed condition was when the rotor is rotating at 1200 RPM, decoupled from the generator and only one caliper is used with only 90% of its effective braking force available. The design condition is to ensure that towards the end of the brakes life the wind turbine will still be able to be brought to rest even if one of the calipers fails.

The braking torque required can be determined using equation 6.5.1.1 from WC Branham brake catalog.

$$T = \frac{WK^2N}{308t} \tag{6.5.1.1}$$

Where W is the weight of the rotating members in pounds force, K^2 is the radius of gyration in ft, N is the angular velocity in RPM's and t is the stopping time required by the brakes. The weight and radius of gyration were pulled off the mass properties from the Solidworks model. The angular velocity was set equal to 1200 RPM, four times the design rotation velocity and the braking time was set equal to 20 seconds. Equation 6.5.1.1 gives a braking torque of 146.8 ft-lbf. Equations 6.5.1.2, 6.5.1.3 and 6.5.1.4 calculate the Btu per stop, Btu per hour and needed disk area to ensure the heat is properly dissipated respectively.

$$\frac{Btu}{stop} = \frac{\pi TNt}{46{,}680} \tag{6.5.1.2}$$

$$\frac{Btu}{hour} = \left(\frac{Btu}{stop}\right)\left(\frac{stops}{hr}\right) \tag{6.5.1.3}$$

$$sq.ft.of\ disc = \frac{Btu/hr}{660} \tag{6.5.1.4}$$

For equation 6.5.1.2 all the values from equation 6.5.1.1 are used to determine the Btu/stop which equals 237 Btu/stop. For equation 6.5.1.3 the stops per hour were assumed to be 2 per day giving an average of 19.7 Btu/hr. Equation 6.5.1.4 gives a minimum disc area of .36 ft^2. Table 6.5.1.1 is taken from the WC Branham brake catalog and lists the available disc sizes and the amount of exposed area.

Table 6.5.1.1: Diameters, exposed areas, weights and Btu/hr of commonly used disc from WC Branham.

Disc Dia. (in)	Exposed Area (ft^2)	Weight (lbf)	MAX Btu/hr @ 300°F
6 5/16	0.43	1.37	283.8
8	0.7	3.52	462
10	0.109	5.46	719.4
12	1.57	7.91	1036
16	2.79	14.07	1840.4

From table 6.5.1.1 any of the available disc will work, for the WTN a 10" disc was selected.

Once the disc was selected the FS200 hydraulic spring applied brake calipers were chosen. Spring applied brakes refer to the fact that a spring is located inside the caliper that forces the brake closed and will only open when the caliper is pressurized to compress the spring. Figure 6.5.1.1 is the specifications for the brake caliper. A single FS200 brake caliper will be sufficient to stop the WTN but a second caliper will be used for redundancy and to extend the life of the brakes

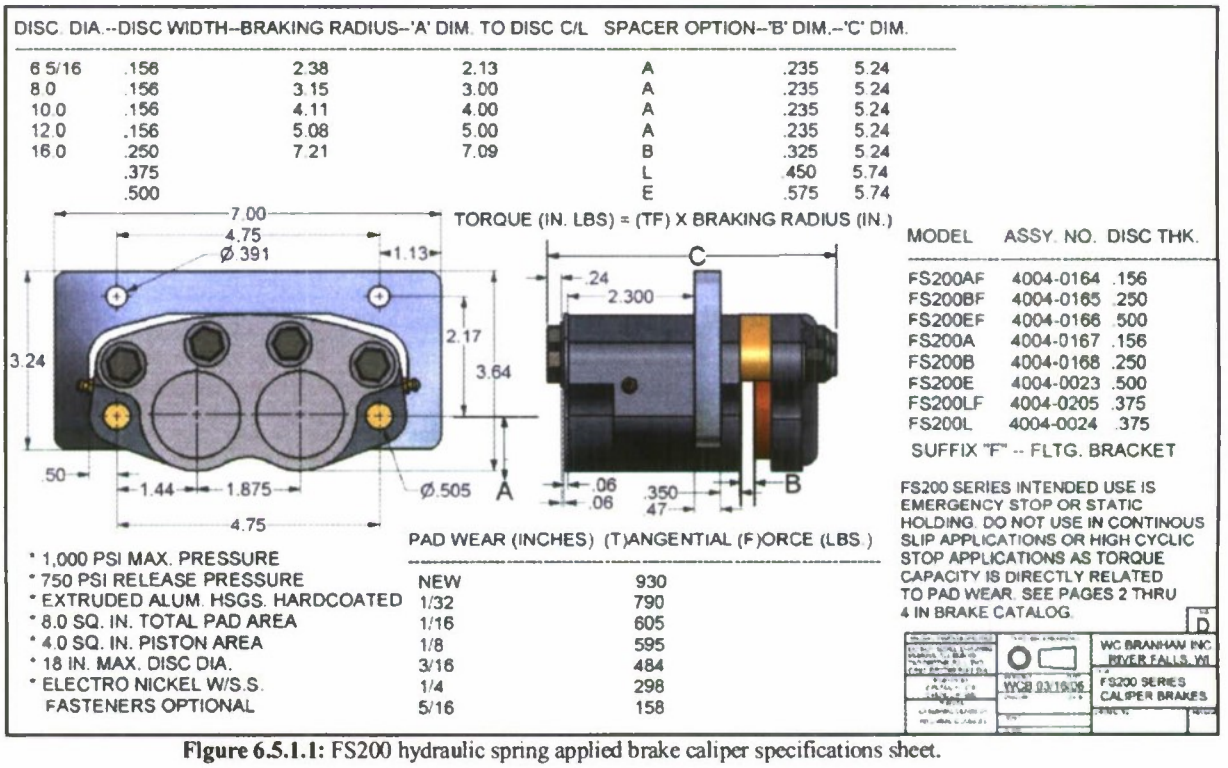


Figure 6.5.1.1: FS200 hydraulic spring applied brake caliper specifications sheet.

The brake calipers are floating so they will be mounted directly to the frame. For hydraulic spring applied brakes they are only as effective as the wear on the pad. WC Branham also provides a series of equations and tables to determine the life expectancy of the brake calipers. For spring applied brakes the wearable friction material is significantly less than other calipers.

For FS200 spring applied brakes and an operating temperature of 300°F the expected life of the two calipers is 6 years. See Appendix E for a more in depth look into the brake caliper analysis.

6.5.2. Hydraulic Power Unit

The calipers require a minimum 750 psi release pressure to open and will close below that amount. A hydraulic power unit used is a 108 series oildyne power unit from Parker. The power unit includes a 24 volt DC motor, pump, 23 cubic inch reservoir tank, 12 volt DC normally open solenoid valve, accumulator to sustain the pressure above 750 psi for an entire day, throttling valve to regulate the flow into the calipers and a pressure switch to shut off the pump once the pressure reaches 1000 psi within the calipers. Figure 6.5.2.1 shows and image of the 108 oildyne unit plumbing diagram provided by Parker.

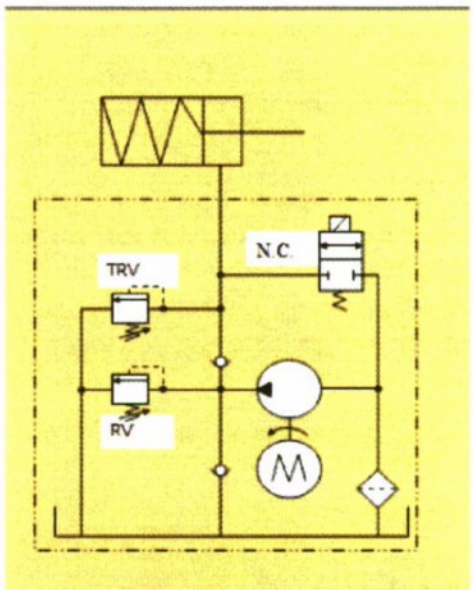


Figure 6.5.2.1: Parker 108 oildyne power unit plumbing diagram. Note the unit used will have a normally open solenoid valve where the diagram depicts a normally closed solenoid valve.

The power unit mounts with two 3/8-16 UNC-2B bolts in a horizontal configuration shown in figure 6.5.2.1 The supply and return ports are threaded 1/8 inch. The unit is wired with 11 gauge wire transmitting a maximum of 180 watts of power to the pump. The throttling valves are 1/8 threaded to connect directly into the caliper ports. The oildyne unit is built with pressure switches so that once the pressure falls below 750 psi the pump will turn on and turn off on the pressure reaches 1000 psi.

7. Testing Plan

The wind turbine was completed just in time for senior design expo on June 4th 2009. The hydraulic power unit was not included since it had not arrived. The first plan is to connect the throttling valves to the calipers and the oildyne power unit to pressurize the brakes. The throttling valves will need to be adjusted to ensure that pressure is not instantaneously lost in the calipers bringing the rotor to a sudden stop and damaging the blades. In order to accurately determine the proper valve opening the rotor should be attached to a dynamometer and ran at 300 RPM transmitting 150 ft-lb of torque. The calipers should bring the rotor to rest no sooner than 10 seconds to avoid damaging the blades and no longer than 30 seconds to avoid excessive heat generation.

Once the wind turbine is mounted on top the tower the accuracy of the yaw should be verified, The IEC suggests an accuracy of plus or minus 8 degrees. This is to ensure maximum power production. The maximum yaw rate is also given in the IEC as a function of the rotor radius. According to the IEC for a rotor 12 feet in diameter the yaw speed should not exceed 18 RPM.

In order to ensure that the fairing can protect the components of the WTN from the environment prior to installation the fairing should be placed in an oven at least 110°F for an entire day. Once removed from the oven there should be no damage to the finish of the fairing occurs. The fairing should also be exposed to a full day of down pour to ensure that water will not leak into the fairing from the seams getting the internal components wet.

Verification of the stress levels in the frame and brake caliper mounts should also be verified with the application of strain gauges. Strain gauges attached to the frame would also be helpful in optimizing the frame and caliper mounts if more wind turbines are raised in Escuela Ranch.

8. Conclusion

The basic design requirements specified by Dr. Lemieux have been satisfied. The plan is to combine the nacelle with the rotor, load bank and tower to have a fully functional wind turbine by fall of 2009. To help Dr. Lemieux at least one member from Hurakan Tech will likely assist Lemieux over the summer to complete any work that needs to be completed before the wind turbine can run such as; connection the oildyne power unit, setting the foundation for the tower, adding weather stripping to the fairing and the addition of a wire mesh in the openings of the fairing to prevent birds from entering the nacelle and making nest. The completed wind turbine will be used as both a teaching and research platform for the CPWPRC. The frame, overspeed, and drivetrain systems are all overdesigned to withstand loads larger than those seen by a 3kW wind turbine. This will guarantee future upgradeability. The yaw system, although currently passive, includes a yaw bearing ring gear, which can eventually accommodate an active control system. The nacelle has extra room for additional monitoring and control devices.

9. References

[1] Bianchi, Fernando D., Battista, Hernan De, Mantz, Ricardo J., "Wind Turbine Control Systems", Springer, 2007.

[2] Mathew, Sathyajith, "Wind Energy Fundamentals, Resource Analysis and Economics", Springer, 2006.

[3] Budynas, Richard G., Nisbett, Keith J., "Shigley's Mechanical Engineering Design", 8th Edition, McGraw Hill, 2008.

[4] International Electrotechnical Commission 61400-1. "Wind Turbines-Part 1: Design Requirements" IEC 2007

[5] www.nrel.gov. Monday, October 30, 2008National Renewable Energy Laboratory. October 30, 2008. http://www.nrel.gov/about/

[6] www. Bergey.com. Friday, October 10, 2008. Bergey Windpower Company October 30, 2008. http://www.bergey.com/

[7] www.dewi.de. Monday, November 03. German Wind Energy Institute. November 3, 2008. http://www.dewi.de/dewi/index.php?id=131&L=0

[8] www.iec.ch. Monday, July 28, 2008. International Electrotechnical Commission. October 24,2008. http://www.iec.ch/ourwork/who-e.htm

[9] www.iskrawind.com. Wednesday, October 17, 2007. Iskra. November 1, 2008. http://www.iskrawind.com/about.htm

[10] Kestralwind.com. Tuesday, April 08, 2008. Kestral Wind Turbines. November 1, 2008. http://www.kestrelwind.co.za/assets/brochures/kestrel%20e400i%20info%20leaflet_final%20for%20print.pdf

[11] www.abundantre.com Monday, September 22, 2008. Abundant Renewable Energy. November 1, 2008. http://www.abundantre.com/ARE_Wind_Turbines.htm#ARE110

[12] The A to Z of Materials. November 1, 2008. *Resin Properties for Composite Materials.* www.azom.com/details.asp?ArticleID=997

[13] www.endurancewindpower.com. Monday, November 03, 2008. Endurance Wind Power. November 03, 2008. http://www.endurancewindpower.com/better_approach.html.

[14] L M Craig, Z Saacl-Saoud, N Jenkins, "Field Testing of a Passive Electrodynamic Brake for Wind Turbines", EMD97 1-3 September 1997 Conference Publication No. 444.

[15] www.icpltd.co.uk Friday, October 17, 2008. Industrial Clutch Parts. October 29, 2008. http://www.icpltd.co.uk/index1.html?lang=en-uk&lmd=39378.443924&gclid=CLzBnNy23JYCFQ8QagodvCwu4Q

[16] www.wcbranham.com Tuesday, November 04, 2008. WC Branham. November 01, 2008. http://www.wcbranham.com/brakes-caliper-disc/.

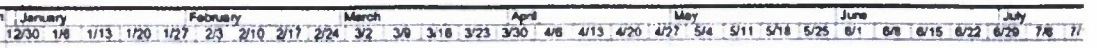
10. Appendix

A. Gantt Chart
B. Wind Loads Given From Dr. Lemieux
C. Ginlong Permanent Magnet Alternator (PL-PMG-3500) Technical Specs
D. Bill Of Materials
E. Overspeed Protection Analysis and Sizing
F. FS200 Hydraulic Spring Applied Brake Spec Sheet
G. Frame FBD
H. Frame Deflection Hand Calculations
I. Wind Turbine Nacelle Strength and Deflection FEA Report
J. Yaw Bearing FBD
K. Gear Tech SY7B0101N-10-b Swing Bearing Specification Sheet
L. 5-5 Vibration Diagram
M. FBD of Drive-Shaft
N. Reaction Forces at Bearings Hand Calculations
O. Shaft Stress Hand Calculations
P. Shaft Stress Concentration Hand Calculation
Q. Shaft Shear and Bending Moment Diagram
R. Timken Bearing Selection Analysis
S. Timken Bearing Medium 300W Series Specs
T. FBD of WTN with unbalanced load, braking while yawing
U. Lovejoy Coupling L 190 Specs
V. Shaft Layout
W. Maximum Shaft Forces
X. Shaft Diameter Sizing Hand Calculations
Y. Static Factor of Safety Calculation for Shaft
Z. Shaft Endurance Limit
AA. Wind Force Acting on a Flat Plate
BB. Drag Force on Nacelle

Appendix A

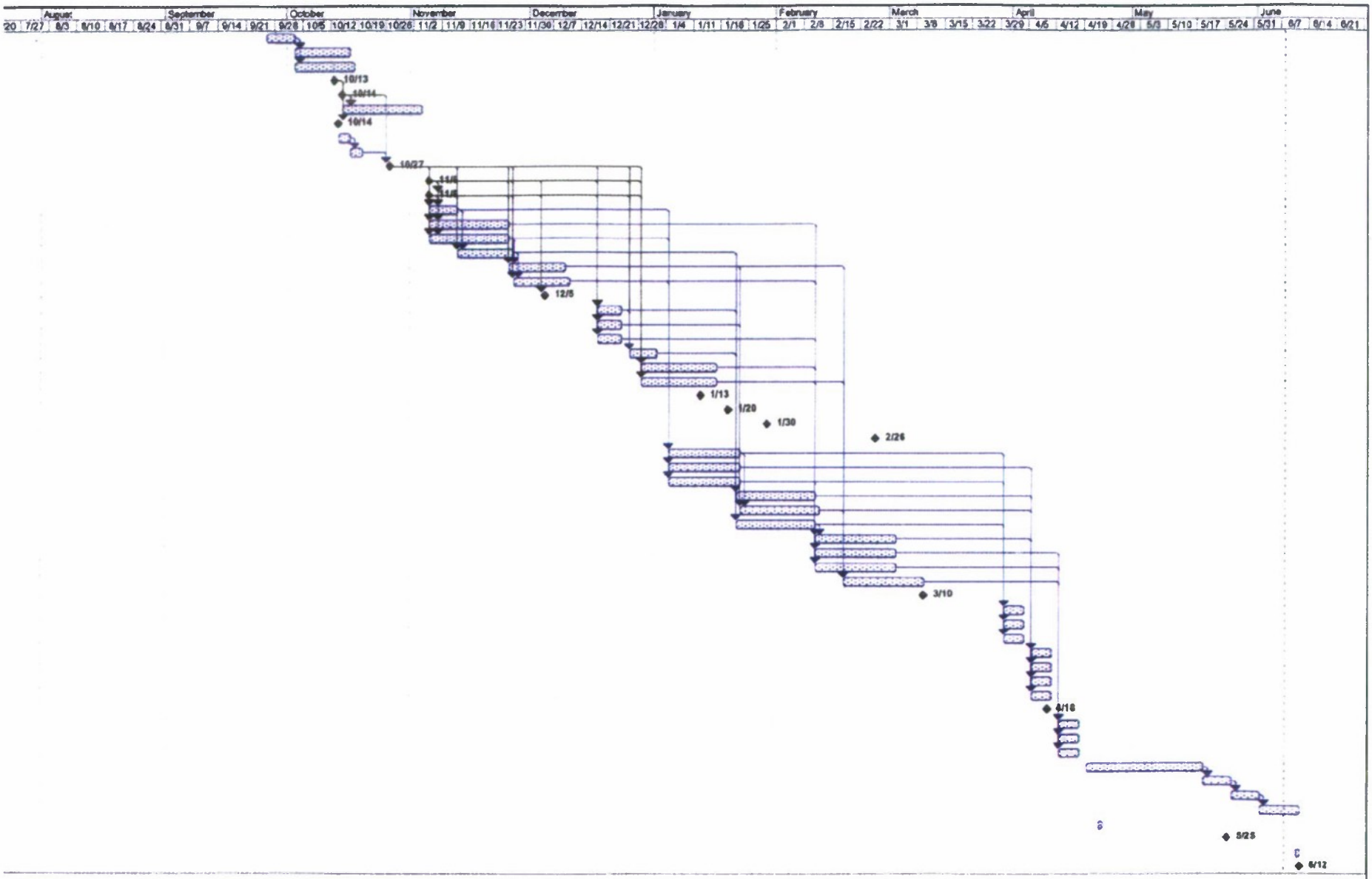
Gantt Chart

ID	Task Name	Duration	Start	Finish	Predecessors
1	Consider Projects	1 wk	Fri 9/26/08	Thu 10/2/08	
2	Background Research	2 wks	Fri 10/3/08	Thu 10/16/08	1
3	Develop Specification	11 days	Fri 10/3/08	Fri 10/17/08	1
4	Project Schedule	0 days	Mon 10/13/08	Mon 10/13/08	
5	Project Proposal Due	0 days	Tue 10/14/08	Tue 10/14/08	
6	Conceptualization	3 wks	Wed 10/15/08	Mon 11/3/08	5
7	Finalize Project Schedule	0 days	Tue 10/14/08	Tue 10/14/08	4
8	QFD	3 days	Tue 10/14/08	Thu 10/16/08	
9	Main Brainstorming	2 days	Fri 10/17/08	Sun 10/19/08	8
10	Decision Matrix	0 days	Mon 10/27/08	Mon 10/27/08	5,9
11	Interim Design Report	0 days	Thu 11/6/08	Thu 11/6/08	
12	Evaluate Sponsor Feedback	0 days	Thu 11/6/08	Thu 11/6/08	11
13	Wind Vane Design	5 days	Thu 11/6/08	Wed 11/12/08	10,11,12
14	Braking System Design	14 days?	Thu 11/6/08	Tue 11/25/08	10,11,12
15	Hub Design	14 days	Thu 11/6/08	Tue 11/25/08	10,11,12
16	Yaw Bearing Design	10 days	Thu 11/13/08	Wed 11/26/08	10,11,12,13
17	Drivetrain Desing	10 days	Wed 11/26/08	Tue 12/9/08	10,11,12,15
18	Nacelle Base Design	10 days?	Thu 11/27/08	Wed 12/10/08	10,11,12,16
19	Draft Report	0 days	Fri 12/5/08	Fri 12/5/08	11,12
20	Nacelle Façade Design	4 days	Thu 12/18/08	Tue 12/23/08	10,11,12
21	Drivetrain bearings Analysis	4 days	Thu 12/18/08	Tue 12/23/08	10,11,12
22	Yaw Bearing Mount Desing	4 days	Thu 12/18/08	Tue 12/23/08	10,11,12
23	Motor Cooling System Design	5 days	Fri 12/26/08	Thu 1/1/09	10,11,12,20
24	Braking System Controls Design	15 days?	Mon 12/29/08	Fri 1/16/09	10,11,12
25	Remote Braking System Desing	15 days	Mon 12/29/08	Fri 1/16/09	10,11,12
26	Design Presentation	1 day	Tue 1/13/09	Tue 1/13/09	
27	Final Design Report	1 day	Tue 1/20/09	Tue 1/20/09	
28	Critical Design Review	0 days	Fri 1/30/09	Fri 1/30/09	
29	Ethics Memo	0 days	Thu 2/26/09	Thu 2/26/09	
30	Hub Fabrication	14 days	Mon 1/5/09	Thu 1/22/09	15
31	Wind Vane Fabrication	14 days	Mon 1/5/09	Thu 1/22/09	13
32	Nacelle Base Fabrication	14 days	Mon 1/5/09	Thu 1/22/09	18
33	Yaw Bearing Fabrication/Purchase	14 days	Thu 1/22/09	Tue 2/10/09	16
34	Drivetrain Fabrication	14 days	Fri 1/23/09	Wed 2/11/09	17,21,30
35	Nacelle Façade Fabrication	14 days	Thu 1/22/09	Tue 2/10/09	20,23
36	Yaw Bearing Mount Fabrication/Purchas	14 days	Wed 2/11/09	Mon 3/2/09	22,18,32,35
37	Braking System Fabrication/Purchase	14 days	Wed 2/11/09	Mon 3/2/09	24,14
38	Braking System Controls Fabrication/Purchase	14 days	Wed 2/11/09	Mon 3/2/09	25,24
39	Remote Braking System Fabrication/ Purchase	14 days	Wed 2/18/09	Mon 3/9/09	26,17,34
40	Project Update Report	0 days	Tue 3/10/09	Tue 3/10/09	
41	Nacelle Façade Testing	5 days	Mon 3/30/09	Fri 4/3/09	35
42	Nacelle Base Testing	5 days	Mon 3/30/09	Fri 4/3/09	32
43	Hub Testing	5 days	Mon 3/30/09	Fri 4/3/09	30
44	Wind Vane Testing	5 days	Mon 4/6/09	Fri 4/10/09	31
45	Drive Train Testing	5 days	Mon 4/6/09	Fri 4/10/09	34
46	Yaw Bearing Mount Testing	5 days	Mon 4/6/09	Fri 4/10/09	36
47	Yaw Bearing Testing	5 days	Mon 4/6/09	Fri 4/10/09	33
48	Manufacturing and Test Review	0 days	Fri 4/10/09	Fri 4/10/09	
49	Remote Braking System Testing	5 days	Mon 4/13/09	Fri 4/17/09	39
50	Braking System Controls Testing	5 days	Mon 4/13/09	Fri 4/17/09	38
51	Braking System Testing	5 days	Mon 4/13/09	Fri 4/17/09	37
52	Necessary Redesign and Fabrication	21 days	Mon 4/20/09	Mon 5/18/09	
53	System Assembly	5 days	Tue 5/19/09	Mon 5/25/09	52
54	System Testing	5 days	Tue 5/26/09	Mon 6/1/09	53
55	Necessary System Redesign/Fabrication	6 days	Tue 6/2/09	Thu 6/11/09	54
56	Senior Exit Exam	1 day	Thu 4/23/09	Thu 4/23/09	
57	Senior Survey	0 days	Mon 5/25/09	Mon 5/25/09	
58	Senior Design EXPO IV	1 day	Thu 6/11/09	Thu 6/11/09	
59	Final Project Report	0 days	Fri 6/12/09	Fri 6/12/09	



Project: simpleexample
Date: Mon 6/8/09

Task | Progress | Summary | External Tasks | Split
Split | Milestone | Project Summary | External Milestone



oject: simplesample
te: Mon 6/8/09

Task
Split
Progress
Milestone
Summary
Project Summary
Split

Appendix B

Wind Loads Given From Dr. Lemieux

Hello Dr Lemieux,

We have not received the wind turbine load data for the wind turbine project. Could you please send this information as we need it for our analysis.

--
Thank you,
Alvaro Martinez

Hi,
Sorry for the delay guys.
- The thrust load on the shaft is 2000lbs max.
- The rotor max speed is 300rpm (your control system should brake it if there are any excursions past that speed)
- Max torque on the driveshaft: 150ft-lb; it carries 3kW at 210rpm.

For shake force, it's hard to tell. I suggest we use a 5-5 rule as worst case scenario: one blade is 5% heavier than the others, and its center of mass is 5% further outbound. Feel free to come up with a better method.

Patrick Lemieux, Ph.D., P.E.
Associate Professor of Mechanical Engineering
California Polytechnic State University,
San Luis Obispo
tel.: (805)756-5583