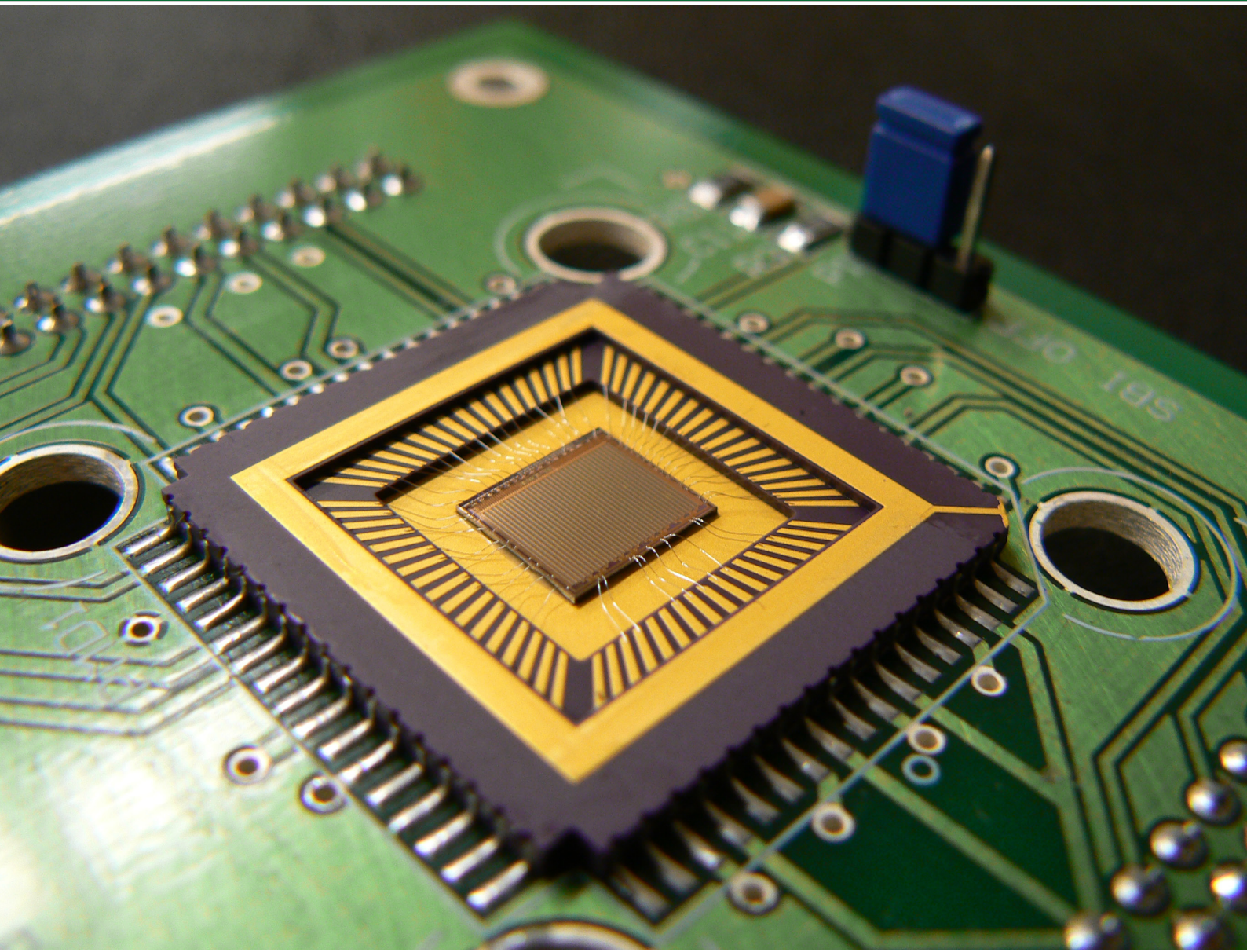


ENZCon 2010

CONFERENCE PROCEEDINGS
22ND - 23RD NOVEMBER



UNIVERSITY OF WAIKATO
HAMILTON, NEW ZEALAND

Published by:

School of Engineering
The University of Waikato
<http://eng.waikato.ac.nz/enzcon10/>

Edited by:

Adrian Dorrington and Andrew Payne
School of Engineering
The University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand

ISSN 1175-2351 (Print)**ISSN 1177-6951 (CD-ROM)****ISSN 2230-2239 (Online)**

Copyright remains with the author(s) of each paper contained in these Proceedings. Please contact the authors if you wish to reproduce their work. Responsibility for the contents of these articles rests upon the authors and not the editors nor ENZCon Inc.

Sponsors:

RF Test Solutions
Nichecom Limited
ElectroTest Ltd
Hoare Research Software Ltb
Wireless Design
WaikatoLink Ltd.

Credits:

Cover photo: PMD3k Time-of-Flight Sensor, taken by Richard Conroy.
LaTeX editor: Refael Whyte
LateX scripting: Michael Cree

Printed by Waikato Print - November 2010

Copyright © 2010 Electronics New Zealand Conference

Welcome from Adrian Dorrington, Conference Chair

Welcome to the seventeenth Electronics New Zealand Conference, The University of Waikato, and to Hamilton. We are delighted to host academics, students, and industry; and to provide the opportunity for sharing knowledge, developing and maintaining contacts and collaborations, and, of course, engaging in social activities.

We are pleased that the conference has returned to its home town for only the second time since the inaugural event here in 1994. I'm sure that those who attended that first ENZCon can attest to the progress that has been made. The University, Hamilton, and the Conference have changed and matured greatly in the intervening sixteen years.

This year we are privileged to hear keynote presentations from two accomplished individuals with experience both in academia and industry. Professors Jonathan Scott from the School of Engineering at The University of Waikato and Ian Graham from Endace Limited. We also have fifteen oral paper presentations and nine poster presentations submitted from eight institutions country wide.

I wish to sincerely thank the conference organising committee, the reviewing committee, and our trade exhibitors and sponsors. In particular, the conference would not have been possible without the day-to-day organisational support from Andrew Payne, Richard Conroy, Robert Larkins, Refael Whyte, and Stefanie Edler; or the financial support of our trade exhibitors RF Test Solutions, Nichecom Limited, ElectroTest Limited, Wireless Design, and WaikatoLink Limited, as well as our sponsor Hoare Research Software Limited; and administration support from various groups at The University of Waikato.

Adrian Dorrington
Convenor, ENZCon'10

Organising Committee

Adrian Dorrington (Chairman)
Richard Conroy
Michael Cree
Stefanie Edler
John Godbaz
Howell Round
Rainer Künnemeyer
Andrew Payne
Howell Round
Sadhana Talele
Refael Whyte

Programme Committee

Fakhrul Alam (Massey)
Donald Bailey (Massey)
Morteza Biglari-Abhari (Auckland)
Richard Blaikie (Canterbury)
Dale Carnegie (Victoria)
George Coghill (Auckland)
Dejan Djukic (Massey)
Pawel Dmochowski (Victoria)
Robin Dykstra (Victoria)
Colin Fox (Otago)
Gourab Sen Gupta (Massey)
Richard Harris (Massey)
Michael Hayes (Canterbury)
Patrick Hu (Auckland)
Nihal Kularatna (Waikato)
Rainer Knnemeyer (Waikato)
John Morris (Auckland)
Wyatt Page (Massey)
Howell Round (Waikato)
Jonathan Scott (Waikato)
Liqiong Tang (Massey)
Steve Weddell (Canterbury)

Contents

Keynote 1 - Prof. Jonathan Scott

<i>Electronics: A Potted History and a Glimpse of the Future</i>	1
J. Scott	

Oral Session 1 - Applications

<i>Analysis of Lighting System in an Educational Facility for Efficient Lighting</i>	3
S. Kancharla, N. Yadla, D. Kacprzak	
<i>Communication System for Satellite Ground Station Using Software Defined Radio</i>	9
M. Abrar, X. Gui	
<i>Bus Stop</i>	13
H. Gulati	

Oral Session 2 - Mechatronics

<i>Quadrotor Helicopters for Visual Flight Control</i>	21
J. Stowers, M. Hayes, A. Bainbridge-Smith	
<i>Embedded Linux Controlled Quadrotor Helicopter</i>	27
A. Millane, M. Hayes, J. Stowers	
<i>Autonomous Anthropomorphic Robotic Arm using Vision and Colour to Monitor Plant Growth in a Laboratory</i>	33
M. Seelye, G. Sen Gupta, J. Seelye	
<i>Simulation and Optimisation of a 2-DOF Parallel Planar Manipulator</i>	39
B. P. Haughey, D. A. Carnegie	

Oral Session 3 - Electronics

<i>Design of a True 3D Digital Display</i>	45
B. Vercoelen, C. Hollitt, R. Dykstra	
<i>Design of XOR gates in VLSI implementation</i>	51
N. Ahmad, R. Hasan	
<i>Scalable inductively coupled power transfer platform</i>	57
D. Kacprzak, A. P. Hu, P. Raval	
<i>Design and Simulation of Biosensor for DNA Detection By CMOS Technology</i>	63
M. Alipour, R. Hasan	

Oral Session 4 - Sensors & Instrumentation

<i>Optimising antenna locations in large radio telescopes using iterative projection algorithms</i>	69
J. Chen, V. Elser, R. P. Millane	
<i>Cytometer: A Low Cost Approach Using an HD DVD Optical Pickup Head</i>	75
R. P. Dayal, R. Künnemeyer	

<i>Crack detection of eggshells using self-mixing phase locked interferometer</i>	79
P. Shrestha, R. Künnemeyer	
<i>Tree Disk Machine</i>	85
S. Knopp, M. Hayes	
Poster Session	
<i>A Flexible, Low-Power, Programmable Unsupervised Neural Network Based on Microcontrollers for Medical Applications</i>	91
R. Długosz, T. Talaśka, P. Przedwojski, P. Dmochowski	
<i>Humanoid Service Robot</i>	97
P. Barlow, G. Sen Gupta, S. David	
<i>Parameter Tuned Digital Filter</i>	103
I. R. Scott, T. C. A. Molteno	
<i>Volume Measurement Using 3D Range Imaging</i>	109
V. Shrivastava, M. J. Cree, A. A. Dorrington	
<i>Proof of Concept of Diffuse Optical Tomography Using Time-of-Flight Range Imaging Cameras</i> ...	115
A. Hassan, R. Künnemeyer, A. A. Dorrington, A. Payne	
<i>Performance Analysis of Microstrip Line Matching Network for 866 MHz UHF LNA</i>	121
J. Li, R. Hasan	
<i>Intelligent Wireless Greenhouse Management System</i>	129
V. M. Quan, G. Sen Gupta, S. Mukhopadhyay	
<i>Mobile Traceable Robot for Large Area Operation</i>	135
M. B. Hitchings, G. Sen Gupta	
<i>Radio Astronomy and eVLBI using KAREN</i>	141
S. D. Weston, T. Natusch, S. Gulyaev	

Electronics: A Potted History and a Glimpse of the Future

Professor Jonathan B. Scott*

*School of Engineering, The University of Waikato
Hamilton, New Zealand.
Email: scottj@waikato.ac.nz

ABSTRACT

Even electronics people are surprised to learn that the vacuum device industry is still a thriving US\$1B business. The tube presided over the electronics revolution, including the fastest-ever penetration of a technology into society—that is the arrival of broadcast radio in the 1920s, not the cellphone in the 1990s—but the vacuum tube also remains far ahead of solid-state alternatives in power-bandwidth. Electronics people are often surprised to learn that Silicon, as a technology for transistors, is superior only in price (and volume). It was not the first solid-state vehicle, Germanium was. It is not the fastest, the so-called III-V family of Gallium-Arsenide and Indium-Phosphide is. Cray supercomputers had a heart of GaAs. In fact, compound semiconductors (CS) in the form of GaAs enabled satellite communications, satellite TV, LEDs, radar, and fibre-optics. It is only recently that pure-silicon cellphones and TV dishes appeared. Hold a complex cellphone in your hand and you hold over half the elements of the periodic table. Most high-performance electronics relies on CS ICs, GaAs, InP, GaN, SiGe.

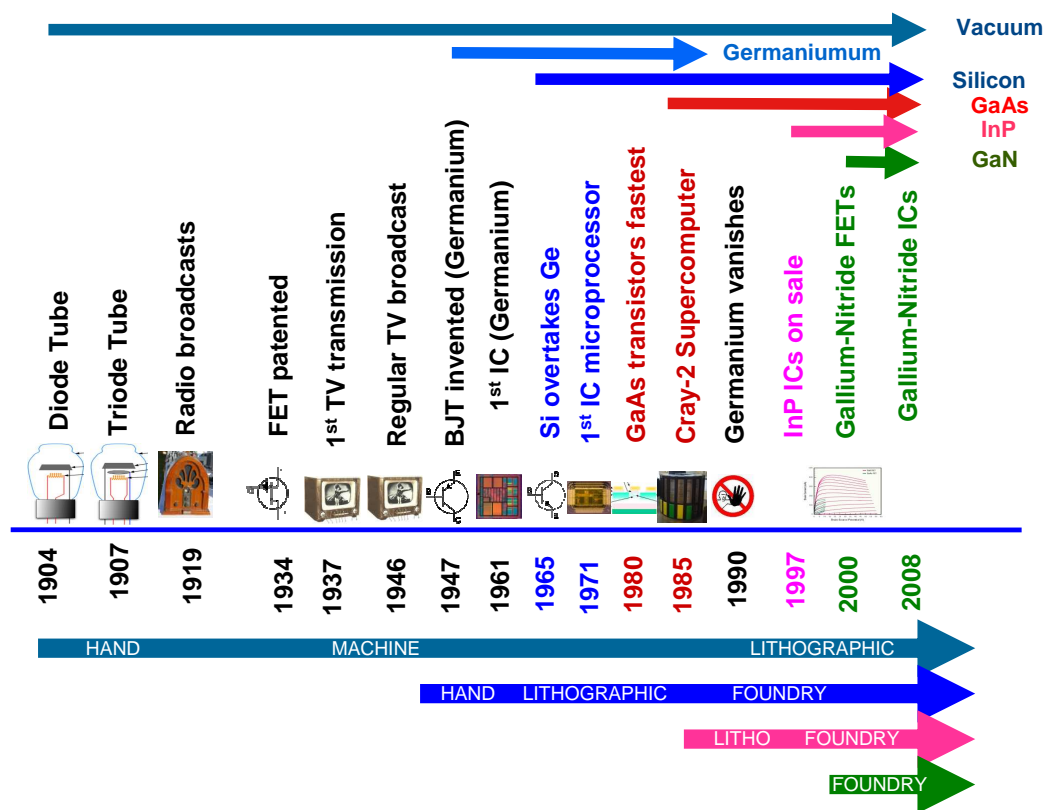
Looking at the history another way, transistors appeal because they can be manufactured lithographically, in Silicon for less than 1 millionth of \$1, RRP. That requires a huge capital investment. In the 1960s & 1970s famous companies built fabs and sold ICs and some became very rich. Fabs that did not sell enough volume folded; in Australia that was AWA through to Peregrine. Will Sapphicon/Silanna change that? In the 1980s & 1990s the industry went largely “fabless”, and foundries/fabs made ICs for many customers, protecting their IP, sharing costs. Except for some huge customers, most CS fabs are “pure play”. It is the same in silicon, with more “captive”, large-volume fabs such as Intel, NXP (Philips), Freescale (Motorola), etc. This fabless trend means that small players get access. In Sydney, Mimix is an example in the CS world. Modest-dimension silicon ICs can be fabricated for a few US\$k, fine-dimension silicon or GaAs ICs for a few US\$10k's. This really happened in the 1990s, so that the “noughties” are thought of as the fabless decade. Education in electronics involves knowing about all devices and how you make things out of them. It means knowing about ICs, not simply the latest silicon VLSI, but vacuum tubes, CS technologies, fab access, the whole “lay of the land” in electronics.

The latest CS technology to appear is Gallium-Nitride, or

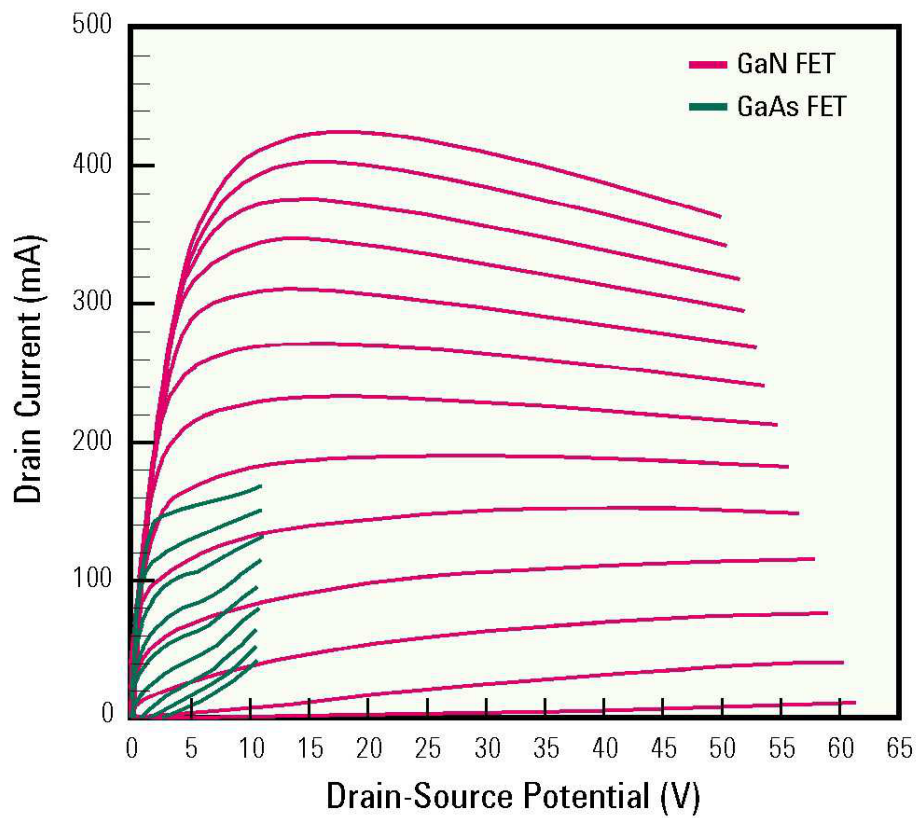
GaN. What does it offer? The speed of GaAs and an order of magnitude more power, and a few other properties that make it very tough and versatile. The second figure here compares two transistors of equal size and equal speed. The small green characteristic of the GaAs FET is that of a $100\mu\text{m}$ 25GHz- F_T GaAs MESFET, the best power III-V process available (circa 2003). The huge, red GaN FET characteristic dwarfs the GaAs one, yet was also that of a $100\mu\text{m}$, 25GHz- F_T device, but in GaN not GaAs. GaN may displace some vacuum tubes, it could revolutionise medical in-vivo ablation techniques, it already has displaced some GaAs power amplifiers, and I believe it will eventually do to GaAs what Silicon did to Germanium.

I have a PhD student working on the automation of coffee roasting. Apart from being a subject that seems to interest just about everyone, it is a subject “whose time has come”. Where once upon a time beans were roasted centrally and shipped across the country, there is a shift towards every coffee shop, and perhaps soon every house with a coffee maker, roasting its own beans, in the interest of ultimate freshness. A few hundred dollars buys you a roaster on Trademe, but it does not give you the expertise to roast well on varying-quality beans. That is a matter of expertise, human or, in our hope, artificial and automatic. There is a “sea change” in the coffee-roasting landscape.

New Zealand has a thriving electronics industry with a long history. The renowned Kiwi ingenuity fits in well with the creative, understanding-based aspects of electronics, a discipline that is both an art and a science. However, the face of electronics is changing with the passing decades. There comes a “sea change” in the electronics landscape. Increasingly the ingenuity and added value in electronics is shifting from circuits composed of discrete components to ICs. These ICs are routinely assembled on circuit boards and require much less ingenuity, and return less value. It is in the design of the ICs that the ingenuity lies, and also the potential for profit. The fabless design model, in silicon and CS, is the affordable entry for the ingenious designer to the electronic world of the future. “Affordable” is a relative term, but we talk of \$100s discrete, \$10k's fabless, but \$1e8 for the 1970s captive model. If New Zealand does not manage to buy into this, the long history of a Kiwi electronics industry will go the way of Germanium. Tell your local funding agency, before the last train leaves the station.



Scaled FET Comparison



Analysis of Lighting System in an Educational Facility for Efficient Lighting

Spandana Kancherla, Neelima Yadla, Dariusz Kacprazk

Department of Electrical and Computer Engineering
The University of Auckland
Auckland, New Zealand
skan078@aucklanduni.ac.nz, nyad003@aucklanduni.ac.nz

Abstract: Lighting systems play a significant role in energy consumption in commercial buildings. Therefore, using efficient lighting systems will offer energy savings. This report outlines the three main components of a lighting system and the design strategies focus mainly on the use of lamps and luminaires to obtain efficient lighting. The lighting in all the different places in the ECE Department at The University of Auckland was re-designed in accordance to the lighting standards and modelled in DIALux. The total energy consumption was then evaluated and the results compared with the current energy usage in department to calculate the energy savings obtained. It can be concluded that approximately 47 % of the energy consumption can be reduced by proposed lighting system which involved precision selection of lamps and luminaires, without sacrificing required illuminance levels in the interiors. It can also be concluded that the proposed system is economically viable and eco-friendly. Overall a sustainable solution was proposed.

Keywords: Lumen; illuminance; uniformity

1 INTRODUCTION

In today's world there is a major demand to save energy. With increasing electricity demands, focus should not be concentrated on how to generate more energy to meet the needs but attention should be directed at the ways in which the available energy can be utilized efficiently. One such sector where with the use of modern technology and design techniques energy can be utilized efficiently is the lighting system. Electricity is a major component of a lighting system. For the past decade, lighting design has been researched into extensively as lighting contributes to a large part of energy consumption in the commercial, which includes office based sectors. In fact, about 30% to 50% of the electricity consumption is due to lighting and it also indirectly increases cooling requirements causing an increase in energy consumption to maintain the temperature levels in buildings [1] [2]. Because lighting systems are known to consume lots of energy, if they can be designed to be as efficient as possible then energy can be used efficiently.

This report outlines the project scope, followed by a short review of the various elements of a lighting system. An overview of the lighting standards is given and the design software used "DIALux" is introduced. Following that, design strategies used to achieve energy savings are outlined, followed by in depth information of the design process and the results obtained. Future work for the project is identified and the conclusions are stated.

2 BACKGROUND INFORMATION

The Electrical and Computer Engineering (ECE) Department at The University of Auckland is an educational facility which can be classified as a commercial building. The reason for the choice of location was due to the easy accessibility to the building for the authors. The building has three floors and consists of many com-

puter laboratories, offices, corridors, staircases and rest-rooms. After examining the current lighting system at the department, it was found that there is a lot of old technology still in place. The notion of using modern day technology and re-designing the lighting system to be energy efficient is seen as the main incentive for commissioning the project.

2.1 Project Objective

The main aim of the project was to investigate and propose energy efficient lighting system solution that results in lower energy usage, reducing expenses and also as a direct result contributes to the environment, by reducing the carbon emissions [1]. Divided into three stages, the first step in the project was to learn basic photometric details. Secondly, with the use of a commercial lighting simulation package "DIALux", re-design the lighting system. Finally, quantify the energy savings for the entire ECE department and evaluate the cost effectiveness of the proposed system.

3 LIGHTING SYSTEM COMPONENTS

The basic elements of a lighting system are the lamp, luminaire and ballast.

3.1 Lamp

Lamps or most commonly called 'light bulbs' are the sources of light. Electrical energy gets converted to light to illuminate the surroundings. With changing technology seen over the years there exist many varieties of lamps such as the incandescent, fluorescent, high intensity discharge and many more. Fluorescent lamps can be used for efficient lighting due to their high efficacy, long life, moderate initial costs and low operating costs [3].

In the category of fluorescent lamps, T-5, T-8 and T-12 lamps are among the most commonly used for commer-

cial purposes. They come in various sizes as shown in Fig. 1 [4].

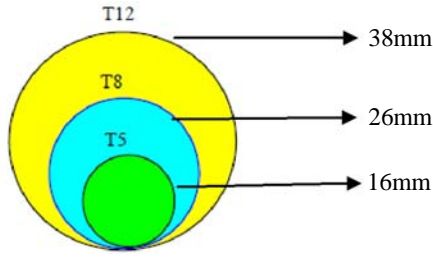


Figure 1: Diameters of fluorescents lamps ^[2]

T-5 lamps have higher efficacy and also higher maintenance factor than T-8 lamps. This means that T-8's need to be more regularly replaced than T-5's and therefore increases maintenance costs.

3.2 Ballast

Ballast is a control device that is used to start a lamp and control the electrical current during its operation. Fluorescent lamps and HID lamps require ballast to operate. They are available in two types, magnetic and electronic. Magnetic ballasts are heavy and noisy. Electronic ballasts on the other hand are quieter, weigh less and some operate on high frequency power to improve the efficiency of the lamp-ballast system [5].

3.3 Luminaire

A luminaire or a luminary is a lighting fitting that comprises of a light source (lamp) and connects the lamp to the power supply. The primary function of a luminaire is to protect the lamps and ballast from dirt build-up and mechanical damage and to direct light to appropriate locations [6]. Depending on the type of interior they need to be used in and the lighting required for a task, a luminaire with the appropriate distribution should be selected. Since the ECE department is an education facility, most of the tasks are screen based and usually requires task lighting, i.e. direct form of light that enables its users to see the critical details of an activity.

An important parameter to be considered when choosing luminaires is the LOR of a luminaire. The Light Output Ratio (LOR) takes the loss of light energy by the luminaire into account. It is expressed as:

$$\text{LOR} = \frac{\text{Light output of luminaire}}{\text{Light output of lamp}} \times 100 \quad (1)$$

From the above equation, it can be deduced that the higher the LOR rating for a luminaire, the more efficient it is. LOR ratings are provided by the luminaire manufacturers as part of the photometric data of a luminaire. This data is necessary because it is used to select luminaires to use, specifies candlepower distribution curves and other important parameters used to perform calculations and provide maintenance recommendations to clients [3]. Fig. 2 shows the candlepower distribution

curve for a luminaire that fits a 28W, T-5 fluorescent lamp.

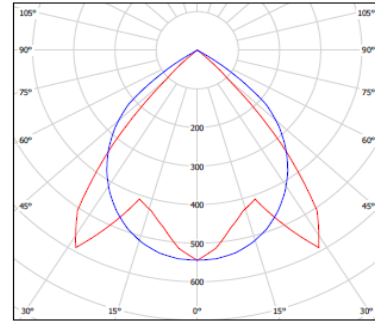


Figure 2: Candlepower distribution curve ^[7]

The candlepower distribution curve is a graph that shows how the luminous intensity of a source from a luminaire varies with direction. Luminous intensity is the concentration of luminous flux emitted in a specific direction [8]. The centre of the chart is the location of the light source, the concentric circles indicate the light intensity in candelas and the radiating lines are the angles at which light is distributed. Important design decisions are made using the photometric data provided keeping the end users of the place in mind.

4 LIGHTING STANDARDS

Regulations followed to design lighting systems are called lighting standards. The lighting standards are official documents that set out the minimum electric lighting requirements for various places and various tasks. This particular project followed Standards Australia, Standards New Zealand and a specific European Standard (EN12464). It was important to comply with the standards in the project to ensure that a safe and quality environment was being designed. The standards are designed to acquire practical results as the work environment and user's needs are taken into consideration. The lighting system designed was based on the minimum lighting requirement values for offices and educational facilities, hence usage of surplus amount of illumination was prevented and energy savings were obtained.

The sections of the standards most relevant to the project were interior and workplace lighting for specific applications, maintenance of lighting systems for indoor applications and safety standards. Table 1 adapted from different sections in the standards shows the minimum maintained illuminance values for different types of interiors and activities.

Table 1: Recommended Lighting Requirements

Types of interior/ activity	Maintained illuminance (lx)
Computer rooms	320
Offices	320
Corridors	40
Foyers	160
Stairs (internal)	80
Restrooms	80

5 DIALux

DIALux is a lighting simulating package which was selected to redesign the lighting system in the ECE department. DIALux is Windows-based radiosity lighting calculation software. Radiosity is a calculation method that divides each surface into small pieces, called patches. Each patch is calculated individually for the amount of light that enters or leaves that surface. The program then solves the system of equations in the model by determining the quantity of light on each patch as a result of the total sum of all the patches. Radiosity is based on Lambertian reflectance calculations. Lambertian reflectance refers to surfaces that have reflected light scattered in such a way that the apparent brightness of the surface is the same regardless of the observer's angle of view. Because of the surface dependency of the calculation, the radiosity method can calculate a model once and produce any desired view [9]. DIALux has interior lighting design feature which is widely used by lighting designer in Europe and North America. Apart from interior lighting design, DIALux also features emergency and exterior street lighting assistants. It also supports certain CAD files which can be useful to duplicate the interiors accurately. DIALux also supports luminaire data sheets from many manufacturers globally which gives designers a wide variety of luminaires to choose from. DIALux is economical package as there is no license fee for this software; it can be easily downloaded by lighting designers all around the world. All the above features made DIALux a professional, suitable and reliable option for this project [9].

5.1 Modelling in DIALux

Few of the parameters that need to be known before modelling in DIALux are the maintenance factor and the room surface reflectance. Maintenance factor was included in the light output calculations according to the luminaire type, for reflector 0.85 and for soft light 0.77, to account for light depreciation. The reduction in light output can be due to ageing of the lamps or accumulation of dirt on the luminaries or both. Reflection factors for room surfaces were set as per the standards as following; 0.7 for the ceiling, 0.5 for the walls and 0.2 for the floor [10], which accounted for the naturally occurring reflections in a room structure. To duplicate the room dimensions accurately, ".dwg" AutoCAD files can be used as DIALux supports these file extensions. This feature made DIALux the ideal simulation package for this pro-

ject as the ECE Department floor plans were available from the ECE-IT Department in DWG file formats. Lastly calculation surface was positioned 0.7 meters from the floor to ensure sufficient illuminance was acquired at a height at which tasks are performed.

6 DESIGN STRATEGIES

Design strategies were aimed to achieve energy efficient, economic and quality lighting system. Proper combination of lamp, luminaire and ballast will allow energy efficient lighting. Choosing luminaries that efficiently provide appropriate luminance patterns for the application is an important part of energy efficient lighting design [11].

Grouping fluorescent T-5 lamps with direct light luminaries that are efficient and have high LOR output allows achieving the minimum illumination levels set in the Standards using the minimum number of lamps. Efficiency of a single-fluorescent lamp is higher than that with two or more lamps in one luminaire [6]. Therefore, it will be efficient to use single lamp luminaries as often as possible and use few multiple lamp luminaries to achieve the required illumination levels.

For quality lighting in the interior, importance should be given to aesthetics. Soft lights can be used as their light distribution allows for bright walls and makes the interior a comfortable environment to work in. Spacing between the luminaries should be such that the light intensity from two consecutive luminaries does not cause any glare (visual discomfort). Uniformity in illuminance is obtained by proper spacing of the luminaries in the ceiling. Good levels of uniformity are 0.6 or greater [10].

According to standards [10], distance of 0.5m from the walls of the room can be omitted from the results. Typical workplane height for tasks is 0.7m. When modelling a room in DIALux, calculation surfaces can be added. Calculation surfaces allow the user to define the geometry of the calculation plane. Simulations were carried out in DIALux taking into account all of the above factors.

7 DESIGN and ANALYSIS

The interiors in the ECE department were categorized into three main sections, offices, labs and others. The total power consumption of the present lighting system in the department is about 46kW and was calculated using a bottom-up approach based on lamp wattage. DIALux simulations were performed for all the rooms in these categories to achieve good uniformity levels and the required maintained illumination levels as set by the standards according to the type of interior. The present and the proposed lighting systems were then compared to evaluate the energy savings obtained.

7.1 Design - Offices

All of the lecturer's offices and other areas such as the department reception office were included in the "Offices" section. The tasks involved are mainly screen based tasks. Each room in this section was designed for the

required maintained illuminance of 320lx as stated by the lighting standards. Fig. 3 shows the proposed geometric distribution of the luminaires in one of the offices.

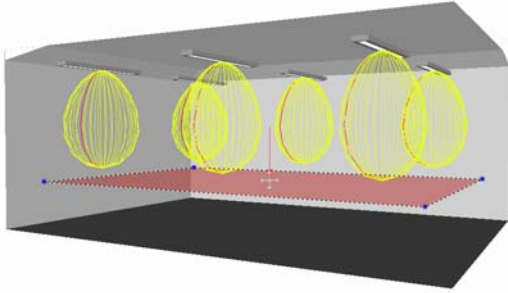


Figure 3: Proposed layout of luminaires in an office

After modelling the room with various types of luminaires, the best solution was obtained using Etap soft lights which gave very high uniformity in illumination of 0.8 as well as meeting the required maintained illuminance of 320lx. From Fig. 3 shown above, it can be seen that the luminaires have been spaced out evenly which avoids the overlap of the candlepower distribution curve giving high uniformity.

Linear surface mount luminaires were used with LOR of 97%. Because high LOR luminaires are efficient, only one 35W lamp was needed in each of the luminaire to achieve the required illumination levels. This was compared to the present lighting system where each of the luminaires housed two 28W lamps. Each of the other office rooms were designed similarly using other varieties of luminaires where necessary and energy savings in this category was thus largely possible due to the reduction in the number of lamps.

Table 2: Comparing the present and proposed lighting system in “Offices” category

Offices	Present	Proposed
Power Consumption (kW)	11.2	5.9
Total number of luminaires	173	171
Total number of lamps	379	179

Table 2 compares the present and the proposed lighting system of all the rooms in the Offices section combined. From the table, it can be seen that the number of lamps in the proposed system have decreased, and this has resulted in total power consumption of 5.9kW, which is lower than the present system.

7.2 Design - Labs

This category mainly included the undergraduate labs and the various specialist labs. The tasks in these labs are either screen based or involve various types of electronic assembly. The required maintained illuminance for electronic assembly is 600lx as stated by the standards. Because of the high precision work involved in electronic assemblies, it requires much higher illumina-

tion than screen based tasks. Fig. 4 shows the proposed arrangement of the luminaires in the SMT facility lab which is an electronic assembly area.

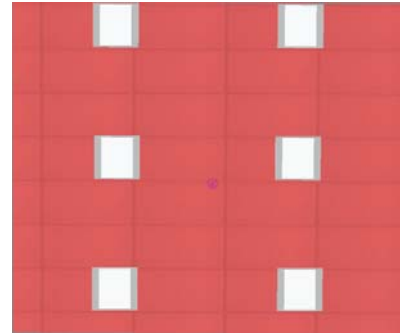


Figure 4: Proposed luminaire arrangement in SMT facility

Recessed square troffers were used, each housing three lamps of 14W each in this particular laboratory. Comparing this to the present system, nine recessed luminaires are currently in place. But in the proposed system, only six Etap luminaires were required to achieve the required illumination of 600lx along with a uniformity level of 0.68.

The SMT facility has a modular ceiling, where the modules are 1.2m×0.6m in size. The room had to be modelled with the luminaires placed in the right co-ordinates where the actual modules are located in the ceiling so that the results can be accurate during simulations. Using the modular ceiling feature in DIALux, it was ensured that the proposed placement of the recessed luminaires were at practical locations in the ceiling.

Table 3: Comparing the present and proposed lighting system in “Labs” category

Labs	Present	Proposed
Power Consumption (kW)	25.4	13.5
Total number of Luminaires	424	338
Total number of Lamps	875	412

Table 3 compares the present and the proposed lighting system of all the rooms in the Labs section combined. From the table, it can be seen that both, number of lamps and luminaires in the proposed system have decreased. Thus energy savings of about 47% were obtained in all the labs.

7.3 Design - Others

The rest of the areas in the department that do not fall into the categories of Offices and Labs were included in the “Others” category. This section comprised of areas such as the corridors, stairs, lobbies, lunchrooms and restrooms. Apart from the lunchrooms, the rest of the areas do not require a calculation surface. In areas such as the corridors and stairs calculations for the amount of illumination was done on a ground level at height zero

rather than at desk height. This was to ensure that there was enough illumination on the floor for safe movement around these areas.



Figure 5: 3-D view of set of stairs

Fig. 5 shows a 3-D view of a set of stairs in the department modelled in DIALux. Etap recessed downlights were used to achieve the required maintained illuminance of 80lx. Each luminaire included only one 13W lamp where as in the present system there is more than one lamp with higher wattage in each luminaire.

Table 4: Comparing the present and proposed lighting system in “Others” category

Others	Present	Proposed
Power Consumption (kW)	9.2	4.6
Total number of Luminaires	228	248
Total number of Lamps	326	248

Table 4 compares the present and the proposed lighting system of all the different areas in the “Others” section combined. From the table, it can be seen that the proposed system consists of more number of luminaires than the present system. But the power consumption of the proposed system is lower than the present system. This was possible because of the use of lesser wattage lamps and the use of only one lamp in each luminaire. As seen in the table above, there are equal number of lamps and luminaires in this category for the proposed system, and the acquired energy savings accounted to 4.6kW, almost 50% lower than the present system.

8 RESULTS

To provide evidence that the re-designed lighting system for the ECE department is superior to the present lighting system, the energy savings obtained were determined and a cost analysis was carried out.

8.1 Total Energy Savings

In order to determine the energy savings obtained from the proposed system in the entire department, power consumption of all the categories for both the present and the proposed lighting system was combined. A bar graph was used to easily compare the two systems. Fig. 6 shows the graph of the percentage of energy savings.

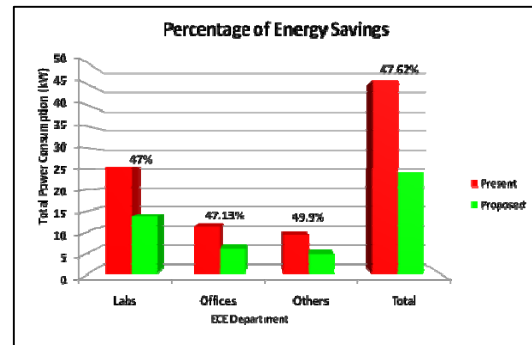


Figure 6: Percentage of Energy Savings

The energy savings in each category is represented as a percentage above the respective category bars. The total power consumption of the proposed system is about 24kW, which is much lower than the total consumption of the present system. This gives a total of about 47% of energy savings on an average. This indicates that the proposed lighting system is energy efficient and the energy savings will lead to a significant reduction in the energy costs.

8.2 Life Cycle Costs

This analysis was necessary in order to allow a decision to be made about whether the proposed system is actually economically feasible. Fig. 7 shows the graph of the life cycle costs of the present and the proposed lighting system.

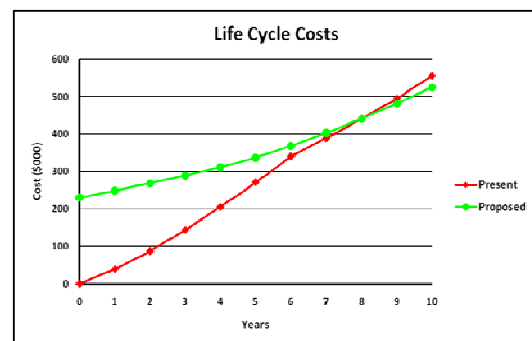


Figure 7: Life Cycle Costs

The costs included installation, energy and maintenance costs. Installation cost for the proposed system was calculated using actual market prices and represented on the graph at year zero. Minimum varieties of luminaires were chosen to keep the initial installation costs to a bare minimum. The installation cost for the present system

was taken to be zero because the cost has already been incurred.

The energy costs for both, the present and the proposed system was calculated based on the total power consumption of each system. The energy calculations were based on bottom up approach. The wattage value used for each lamp includes the losses due to ballast. Total of 4050 hours were calculated assuming that the lights are used 15 hours a day, 6 days a week and 45 weeks a year. Cost of one kilo Watt hour (kWh) of energy used was \$0.185. The cost of power was calculated for each year and then accumulated over a period of ten years, taking into account a 6% increase in energy prices each year [12].

Maintenance cost includes lamp replacement costs and the labour costs incurred upon replacing the lamps. The labour costs are assumed to be \$50 per hour, with a capability of replacing an average of 5 lamps per hour. As done with energy costs, maintenance costs are calculated yearly and added to the energy cost to calculate the total yearly expense. This expense was accumulated over ten years for both the systems. From the graph in Fig. 7, it can be seen that the return on investment from the proposed system is in about eight years. Savings can be expected from then on, making the proposed system a cost effective solution.

9 ENVIRONMENTAL BENEFITS

The proposed system is environment friendly when compared with the present lighting system. Managing the quantity of lighting not only reduces the consumption of electricity but also reduces the use of natural resources such as the raw materials required to manufacture lamps and luminaires and for generating electricity.

Since the total power consumption of the proposed lighting system is lesser than the present system, there will also be a decrease in the amount of CO₂ emissions. In New Zealand, every kWh of electricity consumed generates 0.23kg of CO₂ [13].

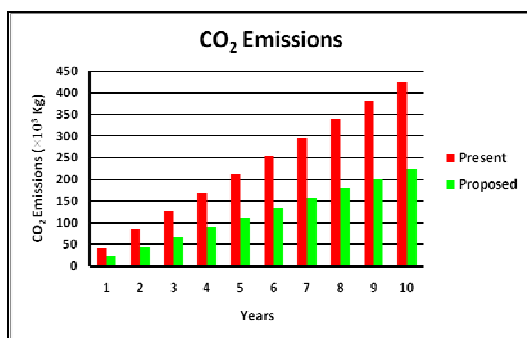


Figure 8: CO₂ emissions

Fig. 11 shows the accumulated CO₂ emissions in thousands of kilograms for the present and the proposed lighting systems. From the figure, it can be seen that there is a significant reduction in the amount of CO₂ emissions from the proposed system, about 45% to 50% lower when compared with the present system.

10. FUTURE WORK

Further reduction in energy consumption can be obtained using dimming controllers. The proposed lighting system was designed to accommodate for the inclusion of dimming in the future. Day light sensors can also be incorporated into the proposed lighting system to take into account natural light. Newer technologies such as LED's can be investigated into, to analyze the advantages of their use in a lighting system.

11. CONCLUSIONS

From research, proper understanding of the different types of lamps, luminaries and ballasts was achieved. Lighting standards were strictly followed to ensure that the proposed lighting system would provide a safe and quality lighting environment. The use of DIALux in simulating lighting for a variety of interiors in the ECE department was successfully incorporated. A new lighting system was proposed for the ECE department which utilized fluorescent T-5 lamps in conjunction with high LOR luminaires to obtain energy efficient lighting. The total energy savings obtained from the proposed system accounted to about 47%, proved to be economically viable to implement with a breakeven point in about eight years and also very environment friendly. Overall, the proposed lighting system proved to be a sustainable solution.

12. References

- [1] W. N. W. Muhamad, *et al.*, *Energy Efficient Lighting System Design for Building*, ISMS International Conference on, 2010
- [2] EECA. (2009). Retrieved from Right-Light.govt.nz/business: <http://www.rightlight.govt.nz/business/efficient-workplace-lighting>
- [3] S.M. Winchip, *Fundamentals of lighting*, New York: Fairchild Publications, Inc 2008.
- [4] NLPPI. (July 2002, *T5 Fluorescent Systems*. Available: <http://www.lrc.rpi.edu/programs/nlpip/lightingAnswers/lat5/pc9.asp> Technical report, "Lighting Fundamentals Handbook," Electric Power Research Institute, 1992.
- [5] I. M. W. Kwot-tin, "T5 Lamps and Luminaires -The 3rd Generation in Office Lighting," p. 13.
- [6] P.C. Sorcar, *Energy Saving Lighting Systems*, Van Nostrand Reinhold Company, 1982.
- [7] "Lighting: ETAP" Retrieved 10 September, 2010, from <http://www.etaplighting.com/products/productdetail.aspx?code=R1600/128HFW&vid=41&cid=3&l=en>.
- [8] Standard, A. N. Z. (2009). Interior Lighting. Safe Movement. Australia, Standards Australia.
- [9] Bickford, Jen. Lighting Software Tools. <http://www.archlighting.com>. [Online] 1 November 2008. [Cited: 8 September 2010.] <http://www.archlighting.com/industry-news.asp?sectionID=1340&articleID=816067>.
- [10] Standard, A. N. Z. (2009). Interior Lighting. Safe Movement. Australia, Standards Australia.
- [11] A Pacific Energy Center Factsheet, "Luminaries," Pacific Energy Center, 1997.
- [12] F. Wolfe, "Power prices to increase by 6pc," Waikato Times, Retrieved 11 September, 2010, from <http://www.stuff.co.nz/waikato-times/business/3362760/Power-prices-to-increase-by-6pc>.
- [13] CO₂ emissions, "Carbon Dioxide Generated in kg/MWh," UNEP: Senoko, 2007.

Communication System for Satellite Ground Station Using Software Defined Radio

Muhammad Abrar, Xiang Gui

School of Engineering and Advanced Technology,
Massey University, New Zealand
M.Abrar@massey.ac.nz

Abstract: The satellite communication industry is also in the process of rapid growing like all other communication and computer industries around the world. The European Student Earth Orbiter (ESEO) is a micro-satellite mission of European Space Agency (ESA) to the Low Earth Orbit. Massey University New Zealand is also taking part in this ESEO project and will establish a second Ground Station. An effort by authors is made to develop the communication system of the uplink transmitter of the ESEO Ground Station. For this development an application of Software Defined Radio (SDR) is being considered. The hardware platform known as Universal Software Radio Peripheral (USRP) is used for up-conversion to radio frequency. As the USRP motherboard requires an appropriate daughterboard for the specific needs of this project with regard to the operating frequency and bandwidth of the ESEO ground station, the RFX-1800 transceiver daughterboard is used for this purpose. GNU radio, a free software toolkit, is used as an interface between the USRP and the PC while Python is employed as the programming language. The objective of this paper is to highlight the use of different free and low cost tools for development and research at universities with limited budget facilities.

Keywords: *SDR, Ground Station, ESEO, GNU Radio, Satellite Communication.*

1. INTRODUCTION

Satellite communication is one of the most progressive and advanced field in the era of modern Science and Technology. Different Universities and Research centers are providing a lot of opportunities for research and development in this area. Many of these organisations have funding from major space agencies like National Aeronautics and Space Agency (NASA) and European Space agency (ESA). Engineering students are encouraged to participate in space engineering projects at these research institutes and universities. Satellites are being used in different applications such as weather prediction, military and civil surveillance and telecommunication etc.

2. ESEO PROJECT

The ESEO project is the third mission within ESA's Education Satellite Program after successful experience of SSETI Express (2005) and the YES2 (2007). The development, integration, and testing of this project is planned to be carried out by European university students as part of the European Space Agency (ESA) Education Office projects. According to project management the satellite will orbit the Earth in Low Earth Orbit (LEO) for taking pictures, measuring radiation levels and testing technologies for future education satellite missions.

The main objective of this project is to provide valuable and challenging hands-on space project experience to the students. In this way well qualified workforce can be prepared in the field of space engineering for the future. There are about 100 students from 13 universities across all Europe are actively involved. It is proposed that the ESEO spacecraft will be launched into LEO in 2012. Carlo Gavazzi Space is an industrial system prime contractor and is managing the ESEO project with the ESA Education Office. They are providing basic system-level and specialist technical support to the student teams during this project. Most of students obtained training and will get benefit from contractor and ESA expertise. They are also allowed to use Industry and European Space Research and Technology Centre (ESTEC) who provide facilities for spacecraft assembly, integration and testing. It is expected that students will provide most of the spacecraft subsystems, payload and ground support systems in coordination with their universities and European space industry in order to deliver their elements of the mission as part of their academic studies [1].

The Ground Station 2 of this ESEO project will be established by Massey University New Zealand, while Ground Station 1 is going to be established at TUM University in Germany.

The Massey University New Zealand Ground Station shall meet the following requirements [2]:

- The Ground Station shall be able to track the ESEO satellite.
- The Ground Station shall be able to communicate with the ESEO satellite with its onboard transceiver.
- The Ground Station shall be able to communicate at any of the bit rates available at the satellite side.
- The Ground Station shall be able to exchange telemetry and telecommand data with the Mission Control Centre via secure connections over the Internet.

To meet the above requirements, the proposed Ground Station will include the following major subsystems as indicated in the Figure 1 [2]:

- Satellite Tracking & Signal Receiving and Transmission
- Amplification
- Frequency Conversion
- Modulation and Coding
- Data Processing and Networking

In this paper our main focus is on uplink Frequency conversion and Modulation & Coding subsystems.

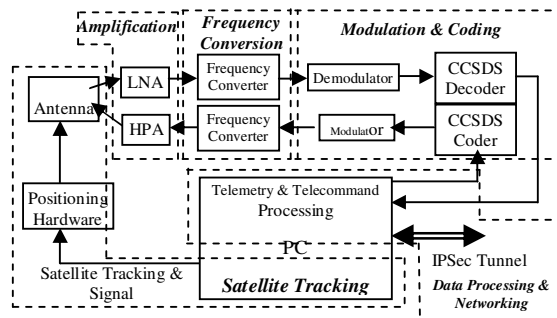


Figure 1: Block Diagram of Ground Station

2.1 Frequency Conversion

This subsystem is a major part of both the transmitter and receiver. However, the operation of this subsystem is opposite in both cases: in the transmitter tract the frequency is up-converted while it is down-converted in the receiver one. For ESEO mission the uplink frequency is supposed to be 2050 MHz, and the downlink frequency is 2250 MHz. The frequency converter needs to convert between these radio frequencies and the intermediate frequencies, much in the same way as a superheterodyne transceiver before the demodulation in the downlink, and after the modulation in the uplink, [2].

2.2 Modulation and Coding

Modulation and coding is a second important subsystem in both the uplink and the downlink for

telecommand data transmission and telemetry data reception respectively. The data rate of 4 kbps is proposed in the uplink. The Non-return-to-Zero Mark (NRZ-M) line coding along with Phase Shift Keying (PSK) on 16 KHz sub-carrier is specified by the ESEO project team. Phase Modulation (PM) is specified as Pass-band modulation. No channel coding is applied to the telecommand data before modulation according to the first design of the system. The target bit error rate (BER) is $1E-5$ with a corresponding signal to noise ratio defined as E_b/N_0 of 9.6 dB [2].

3. SOFTWARE DEFINED RADIO

No standard definition of the term software radio exists and it is associated with a large number of different technologies. The term is usually used to refer to a radio transceiver in which the key parameters are defined in software so that fundamental aspects of the radio's operation can be reconfigured by the upgrading of that software [3]. A software-defined radio (SDR) is a radio communication system that performs radio signal modulation and demodulation in software. The philosophy behind the SDR concept is that the software should operate as close to the antenna as possible (in terms of intervening signal processing stages), and this software should run on a general purpose computer [3]. The objective of SDR research is a desire to shift the radio engineering problem from the hardware domain to the software domain.

4. IMPLEMENTATION

This project is implanted by using SDR. The Universal Software Radio Peripheral (USRP) is used as hardware for up-conversion of frequency. The RFX-1800 transceiver daughterboard is used in the USRP motherboard to meet specific needs of this project with regard to the operating frequency and bandwidth of the ESEO ground station. This daughterboard can receive and transmit the signal in the range of 1.5 GHz up to 2.1 GHz with 100mW transmitting power [4].

The philosophy behind the USRP design is that all of the waveform-specific processing, like modulation and demodulation, should be performed on the host CPU, and all of the high-speed general purpose operations like digital up- and down-conversion, decimation, and interpolation should be performed on the FPGA [5]. The USRP motherboard and daughter boards are shown in Figure 2.

GNU Radio is an open source software framework for writing SDR applications under Linux, UNIX, and Windows and is used for this project along with python as programming language. In addition to a set of base classes from which custom applications can be derived, GNU Radio provides a host of signal processing, hardware interface, graphical user interface, and utility libraries. The GNU Radio framework is built around the graph design pattern. In each GNU Radio application, a flow graph is constructed from nodes

called blocks, which are connected by edges called ports. Each block performs some aspect of the signal processing, and data transport between blocks is facilitated by connecting the input port of one block to the output port of another. Under this framework, the application programmer can create custom blocks and connect them together to form a graph. Once all the blocks have been appropriately connected, the application is sent into a loop, and the GNU Radio framework takes care of streaming data, from a data source through the blocks and out to a data sinks [4].

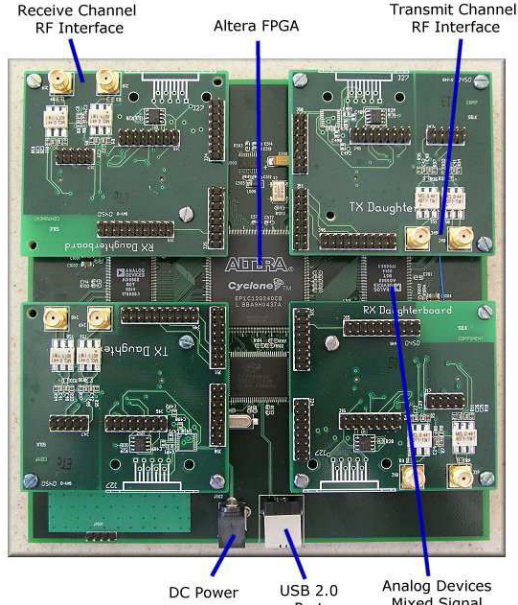


Figure 2: USRP Motherboard with Daughterboards [3]

GNU Radio is a hybrid system; the performance critical portions such as signal processing blocks are written in C++ while non-critical portions such as graph construction and policy management are written in the Python programming language. This allows developers to write highly optimised signal processing code in C++, but use the much more friendly language Python to construct applications. GNU Radio applications written in Python access the C++ signal processing blocks through interfaces automatically generated by SWIG (the Simple Wrapper Interface Generator) for Python. More details on SWIG can be found at www.swig.org.

In this project required software such as Linux, GNU Radio and Python are upgraded to the latest versions to obtain the new features known as GNU Radio Companion (GRC). GNU Radio Companion (GRC) allows users to interact with GNU Radio signal blocks in a manner similar to Labview or Simulink [6]. The project is implemented in Python command line code as well as in GRC.

In the transmitter binary data (0, 1) are encoded into digital signals by using line codes. NRZ-M/PSK/PM modulation on 16 KHz Sub-carrier is specified by ESEO project team. This type of modulation is widely used to avoid phase ambiguities of half cycles or

multiples and is often referred to as differential encoding [7]. The NRZ-M/PSK modulation is achieved in GNU Radio by using the available Differential Encoding block. The differential encoder outputs the difference between current and the last input. With all PSK modulations, using differential encoder makes reception possible even if we do not know the start of phase of the signal. The data signal is interpolated by using an FIR filter. The interpolation of signal is done in order to increase the Nyquist frequency above the sub-carrier frequency and to get to the minimum sampling rate supported by the USRP.

A 16 KHz subcarrier is generated by the signal source available in GNURADIO package and a mixer is used to move the signal from baseband to this sub-carrier. This is just a complex multiply. The analog modulation technique Phase Modulation (PM) is specified as Pass-band modulation. GNU Radio is one of a class of radios known as “Zero-IF” or direct conversion. The NRZ-M/PSK signal is phase modulated by using the available block of GNU Radio. The phase sensitivity is kept at 0.8 as specified. The USRP accepts digital samples that are baseband, centred at 16 KHz in this project. The USRP hardware contains the necessary down and up converters to receive or generate the correct carrier with the help of a daughter card.

5. RESULTS

The preliminary results are shown in Figures 3-8. Figure 3 and 4 show the NRZ-M/PSK modulated signal in the time domain and frequency domain, respectively.

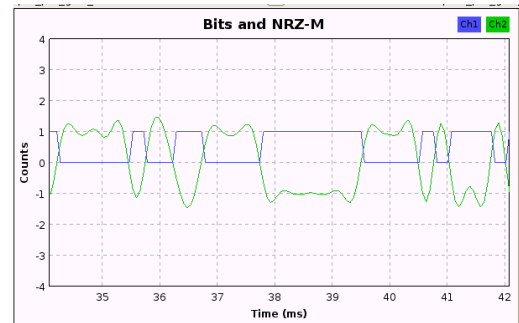


Figure 3: NRZ-M Modulation

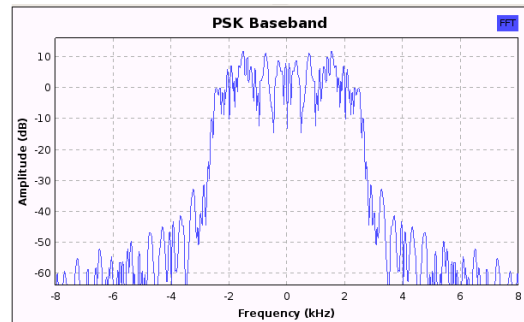


Figure 4: NRZ-M/PSK Baseband Signal

Figure 5 presents constellation plot of PSK modulated signal while an interpolated signal using FIR filter is shown in the Figure 6.

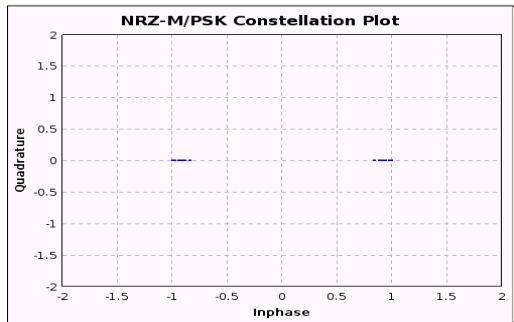


Figure 5: NRZ-M/PSK Constellation Plot

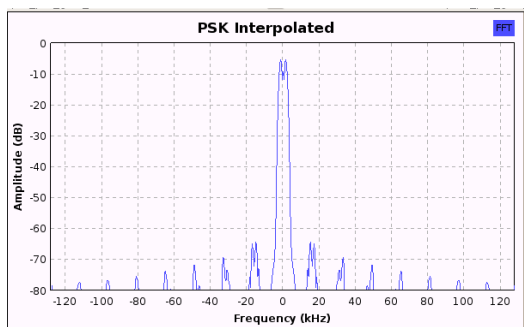


Figure 6: NRZ-M/PSK Interpolated Signal

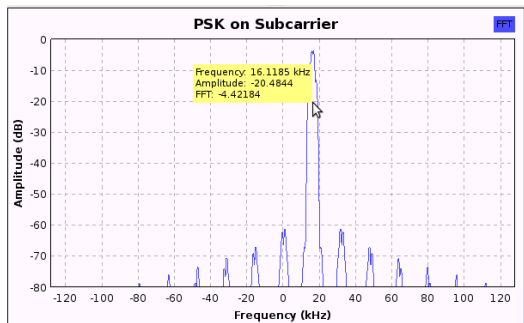


Figure7: PSK signal on Sub-Carrier

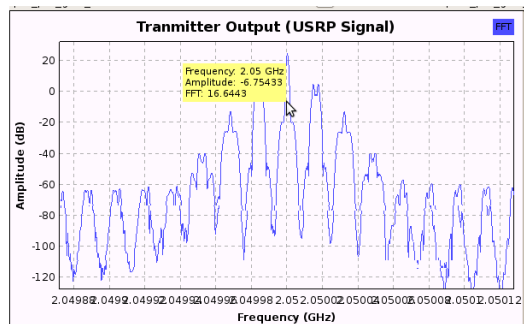


Figure 8: PM output Signal from USRP

Figure 7 and 8 show the frequency domain representation of PSK signal on 16 KHz Sub-Carrier and phase modulated output signal from USRP, respectively.

6. CONCLUSIONS

In this project, the communication system of the uplink transmitter of the ESEO Ground Station is developed by using software defined radio. We explore GNU Radio with the objective of using it for a wireless network. In the process, we discovered that GNU Radio is a sophisticated & rich programming environment, but lacks good references for beginners. Anyone who wants to start must have extensive knowledge of signal processing and basic programming to get to the stage of creating new projects in the field of SDR.

A big advantage for implementing an algorithm in software is the open availability of existing code for a similar problem. Often, only minor adjustments of existing code are needed to reuse it for a new problem. While making such adjustments still requires a complete understanding of the used algorithms, the availability of open code for a similar problem facilitates both the processes of learning and implementing the algorithm.

From all the experimentation and exploring, we can conclude that GNU Radio along with USRP is a low cost platform with immense implementation flexibility and support for modular development and easier sharing of new blocks in the development community. Also, the GNU Radio Companion (GRC) is an excellent GUI tool to test drives the capabilities of GNU Radio without having to write any code.

The implementation to be tested with real data applications, Doppler compensation, Development of ground station receiver and Interfacing with other subsystems are our future work.

REFERENCES

- [1] "ESA Education ESEO Mission" Accessed by <http://www.esa.int> on August 4, 2009.
- [2] "ESEO System Design Summary," Technical Note by Carlo Gavazzi Space SpA, July 7, 2009.
- [3] Peter B. Kenington, "RF and Baseband Techniques for Software Defined Radio," Artech House, London, 2005.
- [4] Ettus Research LLC. "Universal Software Radio Peripheral - The Foundation for Complete Software Radio Systems," Accessed by <http://www.ettus.com> on August 12, 2009.
- [5] Eric Blossom, "Gnu Radio", Accessed by, <http://www.comsec.com> on August 9, 2009.
- [6] GNU Radio Companion," Accessed by <http://gnuradio.org> on August 20, 2009.
- [7] L.Deutsch, F.Pollara and L. Swanson, "Effects of NRZ-M Modulation on Convolutional codes Performance," The Telecommunications and Data Acquisition Progress Report, TDA PR 42-77, January-March 1984, pp. 33-40.

Bus Stop

Hersh Gulati

Manukau Institute of Technology
Private Bag 94006, Manukau.
gulati.hersh@gmail.com

Abstract:

The objective of this paper is to design an application for Android OS (Operating System) based smart phones, which can show bus timetables for Auckland's bus stops on the go. It retrieves the user's present location using in-built GPS (Global Positioning System) or Wi-Fi (Wireless Fidelity) capabilities, or the cell address of their network provider, and displays a selectable list of bus stops within a defined radius, which is configured by the user in the application's settings. From that the user can opt to either see the desired bus stop's real or scheduled timetable using MAXX's VPID (Virtual Path Identifier) service, or look up the route from their location to the selected bus stop on a map. In addition to that, it provides an interface for the user to look up the timetable or directions to any particular bus stop in Auckland. It also consists of an embedded database of all the street names in Auckland, and their GPS coordinates, so that the location can be set manually. It comes with a step-by-step user's guide to help consumers navigate their way around.

Keywords:

Bus Stop, Android, Java, Embedded Database, SQLite, Eclipse, Programming.

1 INTRODUCTION

Technology has been constantly evolving over the last few decades and the world is becoming more and more digital. In Auckland, with a growing number of individuals owning a smart phone, access to information has become a lot more mobile and reachable. People can place orders for their meals, play games, listen to music, chat, browse the net, and perform a lot of other handy tasks with their smart phones these days.

Since a lot of people use buses as a mode of transport to get around within Auckland, this app (application) has been developed for Auckland's users to be able to locate a bus stop and the due buses, accordingly. The information is available to users in both real-time and scheduled-time. Auckland district's bus stop addresses, numbers, and co-ordinates have been stored in an embedded database. All this information has been retrieved from the MAXX website, which is a brand name for ARTA (Auckland Regional Transport Authority).

Using the location API (Application Programming Interface) of the Android library, the application retrieves the user's active location from the network provider's cell address, or the in-built GPS or Wi-Fi capabilities of the phone. It then locates the bus stops within a desired radius from the user's location and displays them in a selectable list, which is sorted in the order of nearest to farthest. From there users have the option of viewing a stop's timetable or step-by-step directions to it.

Sometimes people are unaware of their bus stop's exact address or number. Keeping this in mind, the app has been designed in such a manner that it provides the feature of being able to search through a list of all the bus stops in Auckland. It also allows users to choose whether they would like to specify their location manually, or whether they would like the application to determine it for them. If former is chosen, they just have to key in the street name and the app does the rest. Another key feature of the app is that it allows users to adjust its font size to match their eyesight levels, which is quite handy for the elderly or people with short sightedness.

This app has been limited to Auckland because New Zealand consists of 59 districts and every district has a different transport authority [8]. To work with the entire country's bus stops, databases would have to be retrieved individually from every district. And most of the other districts do not offer a VPID service for their buses. This would also mean a much larger database, and due to memory and performance limitations, it wouldn't be feasible for mobile phones to handle such a large amount of data, as it would hinder the app's performance extensively. If the database is placed on a server instead of embedding into the application, it would increase data costs and thus, make it less appealing to New Zealand users. The application will be tested for reliability and performance in the near future by launching it on the Android Market, which is an online software store designed by Google for its Android devices. Changes shall be made depending on the feedback received from users.

1.1 Google Android

Mobile phones are turning into smart phones, and Google's Android is one of the steadily and successfully growing OS's for smart phones around the globe. It was initially designed by Android Inc., which was then bought over by Google, who then released the OS on 5th November, 2007. It is based on the Linux kernel and GNU software, and uses Java as its core language which runs on the Dalvik Virtual machine [1]. It comes with drivers and support for Eclipse Galileo IDE and emulators for all the different versions of Android. It also has in-built support for SQLite, and allows app debugging on both the emulator and a physical device. This application has been designed using the Android 2.2 platform.

1.2 Eclipse IDE

The application has been designed in the Eclipse Galileo IDE, which is a product of one of Eclipse's projects. Eclipse is an open source non-profit community whose projects aim at creating an open development platform. Eclipse was a project which was initially commenced by IBM back in November 2001 [2].

1.3 SQLite

SQLite is an ACID (atomicity, consistency, isolation, durability) – compliant embedded relational database management system, contained in an approximately 225kB C programming library [11]. All Auckland's bus stops' codes, names and co-ordinates have been taken from the MAXX website. They are available as a Google Transit Feed in a file named stops.txt, which is in a CSV format (Comma Separated Value) [4], [10]. The file contains the addresses, codes and coordinates of the 6109 bus stops in Auckland and its sub-districts.

2 SOFTWARE DESCRIPTION

The project has been programmed in Java using Eclipse Galileo IDE, and an embedded database called SQLite. The sub sections below describe these aspects of the project in detail.

2.1 Eclipse Galileo IDE

Eclipse Galileo is a highly capable IDE which comes with the usual functionalities. But, a few things need to be put in order before it can be used as a development environment for Android OS. It needs version 5 or 6 JDK (Java Development Kit), ADT (Android Development Tools) Plugin, and Android SDK (Software Development Kit) to be able to design and debug apps for Android. The following points give a brief descriptive guide of the steps involved in setting up the IDE.

- Download and install the JDK and Eclipse [6].
- Download and unzip the Android SDK in a location, which is available for Windows, Mac and Linux [6].
- Download and install the ADT Plugin and set the "Android" preferences in Eclipse to point to where the SDK has been unzipped. The user al-

so needs to set up a remote update site, in Eclipse, at:

<https://dl-ssl.google.com/android/eclipse/>

to be able to receive periodic updates for the SDK [6].

- Add Android platforms and other components to the SDK using Android SDK and AVD (Android Virtual Device) manager [6].

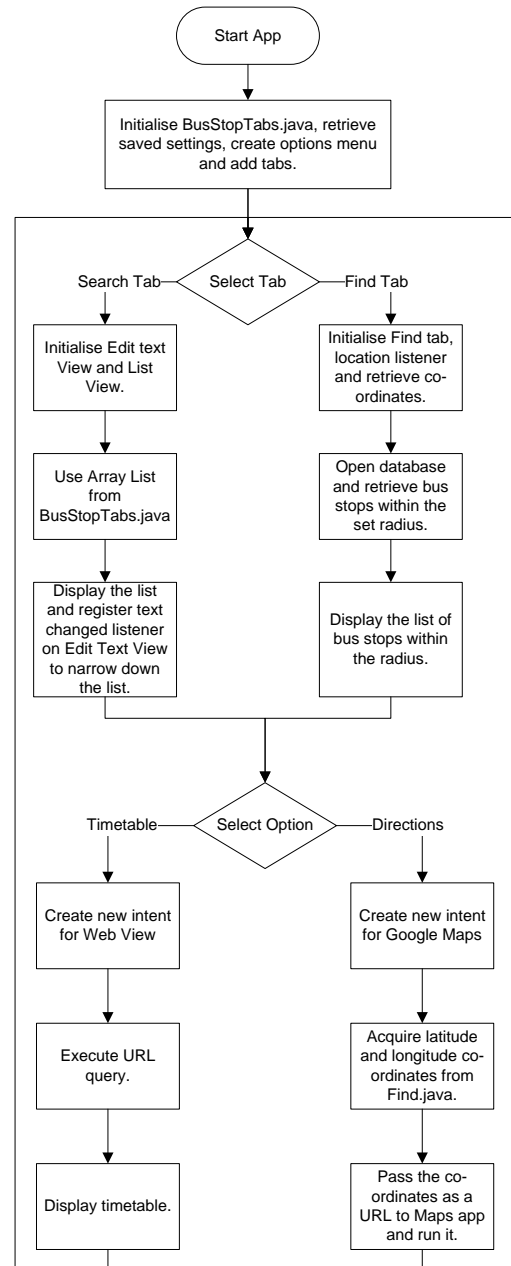


Figure 1: Application Flowchart

The project consists of eight different class files, six of which extend an activity. An activity is a class that creates a window frame in which the UI (User Interface) components are placed [5]. All activities have been placed in a package named *com.project.busstop*. It also consists of seven different layout files in XML (Extensible Markup Language) format. Figure 1 illustrates an overview of the application. All the class files, XML files and their functionalities have been discussed below.

2.1.1 BusStopTabs.java

The application's UI has been based on two tabs – Find and Search. This class extends the Tab Activity - an activity designed to display tabs. This activity is the one that bundles *Find* and *Search* in a tabbed layout and, therefore, forms the basic UI of the app. Both tabs' titles and image sources have been defined in this class. This activity also creates the application's menu using *app_menu.xml* layout file, and initiates different activities based on selections.

Since it is the first activity that is initialised on the application's start up, it extracts information about all the bus stops and street names from the database and stores them in static *ArrayLists*. Bus stops' *ArrayList* is then used for display in *Search* tab, and streets' *ArrayList* is shown when the user opts to set their location manually. Along with these *ArrayLists*, it also extracts all the saved settings like - font size, search radius, number of results to display, sorting order of bus stops in *Search* tab, and whether to use user's actual or specified location. All these settings are stored using *static* variables, so that they retain the same value globally. It consists of a nested class named *InsertDataTask* which extends *AsyncTask*. This class mainly aids in showing a *ProgressDialog*, while information is being retrieved from the database at the same time.

2.1.2 Find.java



Figure 2: Find Tab

When the application is started, it opens up with the Find tab selected, as shown in Figure 2, and displays a *progress dialog* while it's acquiring the coordinates. Once acquired, it shows a list of bus stops within the specified radius. This class extends *ListActivity* for list view, and implements a location listener which retrieves the latitude and longitude co-ordinates of the user's location using the phone's GPS, Wi-Fi or Network Provider. To be able to use the location services, it uses Android's *Location* API. It extracts the bus stops from the database and stores them in an *ArrayList*, which is bound to the list using an *ArrayAdapter*. It consists of a nested class - *ListItemsAdapter*, which extends the *ArrayAdapter* class. It defines the layout of each row, using the *row_layout.xml* layout file, the text on each row and the font size.

Furthermore, it implements a listener for when a bus stop in the list is selected. Upon selection, it pops up a dialog box with the options of viewing either the stop's *Timetable*, or *Directions* to the bus stop from the user's location. If user selects the former option, it creates a new intent for *StopInfo* activity and displays the timetable. But if user chooses the latter option, it passes the user's coordinates and the selected bus stop's coordinates to Android's Google Maps application, and the application comes up with step-by-step directions to the bus stop. It also consists of a method named *Calc*, which is used to work out the bird's-eye view distance between the user and each stop.

2.1.3 Search.java

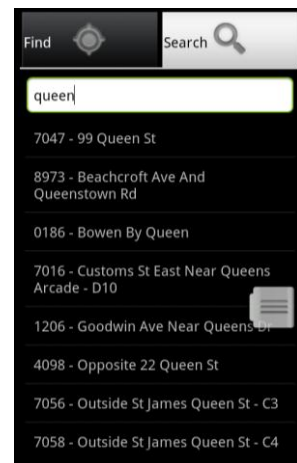


Figure 3: Search Tab

This class also extends Activity, and its main components are the EditText view and ListView, which are bound to an *ArrayList* containing information of all the bus stops in the database. The edit text view is essentially a text field, which narrows down the list as characters are typed into it, as shown in Figure 3. The Find tab and Search tab are similar, in the sense that they both contain a nested *ArrayAdapter* class and an options method for Timetable and Directions. The main difference is that this class consists of a *TextChangedListener*, implemented on the EditText view, so that the list can be narrowed down as the characters are being typed into the text field.

2.1.4 DataHelper.java

This class extends the *SQLiteOpenHelper* class, which controls the connectivity and communication between the app and the database. As mentioned in the earlier sections, information about the bus stops has been stored in an embedded database called SQLite. Android's existing compatibility with SQLite made it a lot easier to connect and query the database. The data was imported from a CSV file to the database using a SQLite Database Browser, which was then stored in a project's subfolder named *assets* as *stops.sqlite*.

This class creates an empty database into the default system path of the app and then overwrites it with the existing *stops.sqlite*. It contains a method to check

whether the database already exists, in order to avoid re-copying the file each time the app is opened. It also contains a method to copy *stops.sqlite* from the local assets folder to the just created empty database via a byte stream, and a method to open the database. Moreover, it contains different methods to extract an individual column or all the columns from different tables in the database using cursors, and stores the information in *ArrayLists*. From the *stopsTable* table in the database, it retrieves the bus stop's address, number, latitude and longitude values, calculates the straight line distance between the user's location and the bus stop's location and then saves all of it as a *stops* object. Then each *stops* object is saved in an *ArrayList* of *stops* objects, which are then extracted using the class's *toString* method.

From *settingsTable*, it retrieves the saved settings at the start of the application, and saves them when the user clicks on the *Save* button in the settings screen. It also retrieves Auckland's street names and their coordinates from *streetsTable*, when the user wants to set their location manually in settings.

2.1.5 Stops.java

This class file doesn't extend any other classes. An instance of this class is created to represent each stop. All the information about every bus stop that is retrieved from the database in the *DataHelper.java* class is stored as an instance of this class, and then all the instances are stored in *ArrayLists*. It contains two different constructors – the first one saves objects excluding distance for the *Search* tab, and the second one saves objects including distances for *Find* tab. The rest of the class basically consists of *getter* and *setter* methods for bus stops' names, addresses, latitudes, longitudes and distances.

The *toString* method of this class has been modified to control the formatting of how the bus stop's information is displayed in each row. It shows *m* or *km* next to the distance for each object in *Find* tab, based on the *toString* method. One key feature of this class is that it implements Java's *comparable* interface, which compares the distance value for each object of *Find* tab, and sorts the list in the order of their distance value.

2.1.6 StopInfo.java

ROUTE	DESTINATION	DUE
472	REDHILL	9min
455	WEYMOUTH	11min
471	PAHUREHURE	40min
472	REDHILL	50min
471	PAHUREHURE	59min

Figure 4: Bus Stop Timetable

This class extends the *Activity* class and uses a *WebView* to show timetables for upcoming buses. *WebView* is a view that is used on an *Activity* to encode and display web pages that are in HTML (Hyper Text Markup Language) format, and it does so by using the WebKit rendering engine [9], as shown in Figure 4.

The timetables for the desired bus stops are taken from MAXX's website by executing a query on the link: <http://www.maxx.co.nz/vpid.cfm?stopNumber=7050>. It uses either the phone's network or Wi-Fi to connect to the internet and retrieves the information. The last 4 digits (7050 in this example) are the bus stop's number, and are replaced every time a new bus stop is chosen. It doesn't communicate with the webpage unless Wi-Fi or mobile network has been connected. If they are connected, it shows a loading progress bar at the top of the screen, while loading the webpage. Keyboards have been disabled on this activity so that the user cannot edit the stop's number straight from the webpage itself.

MAXX uses the on-board GPS on each bus and the cellular data network to generate real-time stop information. Real-time information is currently available for the majority of the bus stops in Auckland. Ultimately it will be available for all stops on all routes (including train stations).

2.1.7 Settings.java

Figure 5: Settings Screen

This class file also extends an *Activity*, and it is one of the most important classes in the project. When the user selects *Settings* from the application's menu, it returns this activity. The class selects components individually from *settings_layout.xml* file, as shown in Figure 5, and modifies them by attaching *adapters* to *spinners*, creating custom *dialogs*, displaying the pre-set settings, and eventually saving the settings. This class also contains a nested class, *ListItemsAdapter*, like *Find* and *Search* classes in order to control the fonts and row layouts of all the *spinners*.

2.1.8 Streets.java

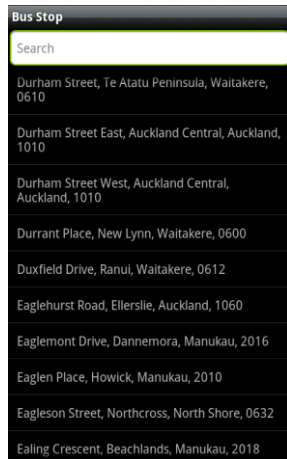


Figure 6: Streets Screen

This class is essentially the same as *Search.java*. It consists of an *EditText* view and *ListView*, as shown in Figure 6, where the list keeps narrowing down as the user keys in characters. This *Activity* is used so that only a street within Auckland is selected for the search criteria.

2.1.9 XML Files

As mentioned earlier, the application consists of seven XML files that define layouts for different components. Android uses XML to design the application's GUI (Graphical User Interface), and automatically allocates a unique ID to all the app's resources and components. These unique IDs can be found in an auto-generated java file named *R.java*, where *R* stands for Resources.

- The main XML file has been named *main.xml*, and it consists of a *TabHost* that hosts multiple *TabWidgets*, as in *Activities* or classes.
- The *list_layout.xml* file defines the *ListView* in *Find* tab, and shows a *TextView* with "No results found." if there's nothing to display.
- *Row_layout.xml* defines the layout of each individual row in both the tabs. For example – background colour, text colour, row's height, etc.
- XML file *row.xml* defines the layout of *Search* tab. It places the *EditText* field and the *ListView* in a *LinearLayout*, and specifies their different properties.

The layout for *Settings.java* has been defined in *settings_layout.xml*. It consists of:

- Five *TextViews* to label each setting.
- Three *spinners* to control the radius, number of results and font size attributes.
- Two *radio groups* with two *radio buttons* each, to control the sorting order of bus stops in *Search* tab, and to control whether the location is to be determined automatically or whether the user will define it manually.

- One *EditText* view to put in the street name. When user clicks on the text field, it opens a new activity which looks and works just like *Search* tab. However, the list is made up of all the street names in Auckland district from *streetsTable* and an *EditText* view to search through the street names. When user selects a particular street, it retrieves the respective coordinates and uses them for the rest of the activities. This *EditText* view is enabled only if the user selects to put in their location manually.
- Lastly, it consists of a *button* to save the settings. It has got an *OnClickListener* attached to it that shows different warning and confirmation dialogs, depending on how the user interacts with it.

The application consists of four additional XML files:

- *strings.xml* – defines the application's name, and consists of three string arrays for the three *spinners* in settings screen. It is placed in a folder called *values* in the resources directory of the application.
- *images.xml* – defines the file names and location of images that have been used in the tabs. It is placed in a folder called *drawable* in the resources directory of the application. The images for tabs, application's icon, and menu's icons had to be placed in the same folder for the file to be able to read them.
- *app_menu.xml* – defines the layout of the menu by specifying the icons and titles for each menu option. A new folder named *menu* had to be created in the resources directory, so that the application could recognise that the file specifies the design of the menu.
- *AndroidManifest.xml* – This is the most important XML file in the entire application. It defines the application's name, version, icon, screen orientation, etc. It also specifies permissions for the application to allow it to access the location API, access the network state, and access the internet. Any activity defined in the application will not work, until it has been defined in this XML file.

2.2 SQLite Embedded Database

Database Structure			
Browse Data			
Execute SQL			
Name	Object	Type	Schema
stopsTable	table		CREATE TABLE stopsTable (_id INT...
android_metadata	table		CREATE TABLE "android_metadata"...
streetsTable	table		CREATE TABLE streetsTable (_id IN...
settingsTable	table		CREATE TABLE settingsTable (stree...

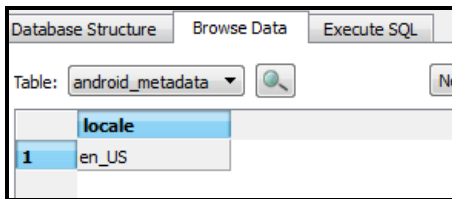
Figure 7: Database

The database is a fundamental part of this app. Without the database there would be no bus stops to locate, or display timetables or routes for. Normally databases reside on servers, and a connection has to be established between the client and the server to be able to communi-

cate with the database. If the same practice was implemented on this app, it could cost the users high data charges to be able to connect to the server, as it would have to be placed on the web. Therefore, this app has been designed using SQLite which makes the database a part of the app, and its small footprint does not slow down the application either [3].

As mentioned earlier in sections 1.3 and 2.1.4, the database was obtained from MAXX's website in CSV format files and imported into *stops.sqlite* using the SQLite Database Browser. The database structure consists of four tables – *stopsTable*, *android_metadata*, *settingsTable* and *streetsTable*, as shown in Figure 7.

2.2.1 android_metada

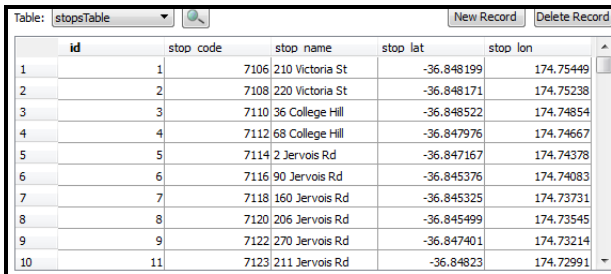


	locale
1	en_US

Figure 8: Android Metadata Table

This table has been added so that Android is able to recognise and use the database with its SQLite libraries. It consists of one column called *locale*, which is of text type. The column contains one value – *en_US*, as shown in Figure 8, which helps Android in recognising the encoding format.

2.2.2 stopsTable



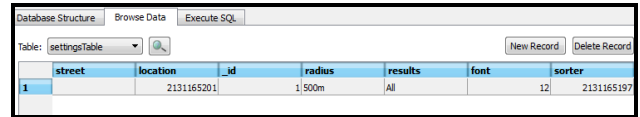
	id	stop code	stop name	stop lat	stop lon
1	1	7106	210 Victoria St	-36.848199	174.75449
2	2	7108	220 Victoria St	-36.848171	174.75238
3	3	7110	36 College Hill	-36.848522	174.74854
4	4	7112	68 College Hill	-36.847976	174.74667
5	5	7114	2 Jervois Rd	-36.847167	174.74378
6	6	7116	90 Jervois Rd	-36.845376	174.74083
7	7	7118	160 Jervois Rd	-36.845325	174.73731
8	8	7120	206 Jervois Rd	-36.845499	174.73545
9	9	7122	270 Jervois Rd	-36.847401	174.73214
10	11	7123	211 Jervois Rd	-36.84823	174.72991

Figure 9: Stops Table

This table consists of 6109 rows of data and five columns, a preview of which has been shown in Figure 9. The columns are:

- *_id* – An auto incrementing integer type primary key column.
- *stop_code* – A text type column of stops' numbers.
- *stop_name* – A text type column of stops' names.
- *stop_lat* – A floating point value type column of stops' latitudes.
- *stop_lon* – A floating point value type column of stops' longitudes.

2.2.3 settingsTable



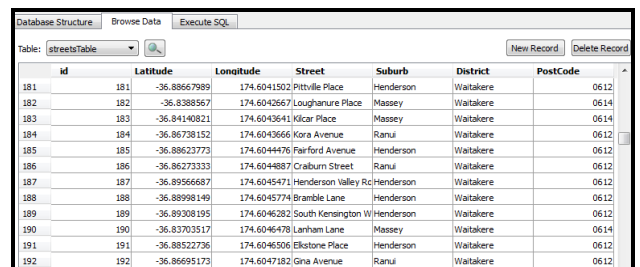
	street	location	_id	radius	results	font	sorter
1		2131165201		1,500m	All		12

Figure 10: Settings Table

As mentioned earlier, this table consists of one row, which is updated every time the user saves the settings, as it can be seen in Figure 10. It consists of eight columns:

- *_id* – An auto incrementing integer type primary key column.
- *radius* – A text type column to hold the radius value set by user. It has been reserved as a text as the radius could be in metres or kilometres.
- *results* – A text type column to hold the value for the number of results to show. It has been reserved as a text because the initial value of this field is set as *All*.
- *font* – A floating point value type column for what font size to use in the application.
- *sorter* – A floating point value type column to store the ID of the *Search* tab sorting radio buttons. Depending on the value in here, the list in *Search* tab is sorted by bus stop address, or bus stop number.
- *location* – A floating point value type column to store the ID of the radio buttons, which determine whether user wants the application to find its location manually or whether the user wants the application to use the manually typed in location.
- *street* – A text type column to store the street name if user selects a manual location for the application.

2.2.4 streetsTable



	id	Latitude	Longitude	Street	Suburb	District	PostCode
181	181	-36.88667989	174.6041502	Pittville Place	Henderson	Waitakere	0612
182	182	-36.8388567	174.6042667	Loughanure Place	Massey	Waitakere	0614
183	183	-36.84140821	174.6043641	Kicar Place	Massey	Waitakere	0614
184	184	-36.86738152	174.6043666	Kora Avenue	Ranui	Waitakere	0612
185	185	-36.88623773	174.6044476	Fairford Avenue	Henderson	Waitakere	0612
186	186	-36.86273333	174.6044887	Craiburn Street	Ranui	Waitakere	0612
187	187	-36.89566687	174.6045471	Henderson Valley Rd	Henderson	Waitakere	0612
188	188	-36.88998149	174.6045774	Bramble Lane	Henderson	Waitakere	0612
189	189	-36.89308195	174.6046282	South Kensington W	Henderson	Waitakere	0612
190	190	-36.83703517	174.6046478	Lanham Lane	Massey	Waitakere	0614

Figure 11: Streets Table

This table consists of data for 9750 streets in Auckland and its outskirts. A small preview of this table can be seen in Figure 11. It consists of seven columns:

- `_id` – An auto incrementing integer type primary key column.
- `Street` – A text type column of streets' names.
- `Suburb` – A text type column of the streets' corresponding suburbs.
- `District` – A text type column of the streets' corresponding districts.
- `PostCode` – A floating point value type column of the streets' corresponding post codes.
- `Latitude` – A floating point value type column of streets' latitudes.
- `Longitude` – A floating point value type column of streets' longitudes.

3 FURTHER WORK

Though the app currently incorporates a lot of functionalities and features, it holds strong potential for future developments.

The first major expansion could be integrating the database for all districts in New Zealand. This would also mean moving the database to a web based server, as a phone will not be able to handle such a large amount of data. But by doing this the user will be limited to scheduled timetables mostly, as most of the districts apart from Auckland do not offer a VPID service.

Second option could be including timetables for trains, ferries and planes. This information could again be based for either just Auckland, or entire New Zealand, and it would definitely expand the application's horizon.

Finally, the last major advancement could be extending this application to other platforms like iOS for iPhones, iPod Touches, iPads, and Maemo for Nokia's tablet smartphones. And with such versatility in platforms, developers around the globe could develop the same application for other countries, using it as the basis.

4 CONCLUSIONS

With the increasing fuel prices coupled with recession, more and more people are turning towards sharing vehicles and public transport. This application is a useful tool for anyone who catches the bus, as it is very user friendly, and provides information on the go.

People can look up when their bus will be due at their usual bus stop, while sitting in the convenience of their home or work or university, by just tapping a couple of times. With ever increasing Wi-Fi networks and reducing data charges, it offers a promising future.

5 REFERENCES

- [1] J. Eitel. (2008, November 3). *All About Android, Google's New Mobile Operating System* [Online]. Available: <http://www.seochat.com/c/a/Search-Engine-News/All-About-Android-Google's-New-Mobile-Operating-System/1/>
- [2] Eclipse.org. (n.d.). *About the Eclipse Foundation* [Online]. Available: <http://www.eclipse.org/org/>
- [3] SQLite.org. (n.d.). *About SQLite* [Online]. Available: <http://www.sqlite.org/about.html>
- [4] Google.com. (2010, January 11). *General Transit Feed Specification* [Online]. Available: http://code.google.com/transit/spec/transit_feed_specification.html#stops_txt_Field_Definitions
- [5] Android.com. (2010, September 23). *Activity* [Online]. Available: <http://developer.android.com/reference/android/app/Activity.html>
- [6] Android.com. (n.d.). *Quick Start* [Online]. Available: <http://developer.android.com/sdk/index.html>
- [7] Android.com. (2010, September 23). *Intent* [Online]. Available: <http://developer.android.com/reference/android/content/Intent.html>
- [8] Statoids.com. (2005, March 23). *Districts of New Zealand* [Online]. Available: <http://www.statoids.com/ynz.html>
- [9] Android.com. (2010, September 23). *WebView* [Online]. Available: <http://developer.android.com/reference/android/webkit/WebView.html>
- [10] MAXX. (n.d.). *Google Transit Feed (GTFS)* [Online]. Available: <http://www.maxx.co.nz/information/about-maxx/google-transit-feed.html>
- [11] Wikipedia. (2010, October 21). *SQLite* [Online]. Available: <http://en.wikipedia.org/wiki/SQLite>

Quadrotor Helicopters for Visual Flight Control

John Stowers*, Michael Hayes, Andrew Bainbridge-Smith

*Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand.

Email: john.stowers@ieee.org

Abstract: The University of Canterbury has developed a suite of software, the *wasp* system, to facilitate visual flight control research on flying robots. This paper introduces the system through the use of quadrotor helicopters, and provides a comparison with the state of the art of research occurring in this area worldwide.

Keywords: Quadrotor, UAV, Visual Flight Control, Flying Robotics

1 INTRODUCTION

Visual flight control testbeds were historically expensive and complex platforms [1, 4], often requiring large open spaces to operate. Fixed-wing aircraft also have constraints on maneuverability whereas conventional helicopters are dynamically and structurally complex, expensive, and hard to control. In order to circumvent these issues the *wasp*¹ system was developed; a suite of software, hardware and infrastructure for visual flight control research with quadrotor helicopters used as the test vehicle.

This paper will proceed as follows; Section 2 details the basics of quadrotor flight. Section 3 provides a discussion of important considerations in visual flight control and the state of the art today. Section 4 explains the *wasp* system and its use at the University of Canterbury. Section 5 concludes with a discussion of research and results currently being undertaken by the author.

Figure 1 shows the *wasp* system in use on a hovering quadrotor helicopter. In this scenario, the *wasp* software autopilot provides the high-speed real-time attitude stabilization of the craft, and time correlated synchronized camera images and inertial measurement unit (IMU) readings to an onboard single board computer (SBC). The SBC then executes the visual control algorithms, taking into consideration the vehicle state computed using the IMU, before sending control commands to the autopilot at a much slower rate. The decoupling of the two systems, both for reliability and for developmental efficiency is facilitated by the *wasp* system. Motivations for such a separation, and the specific implementation used here are described in sections 3 and 4 respectively.

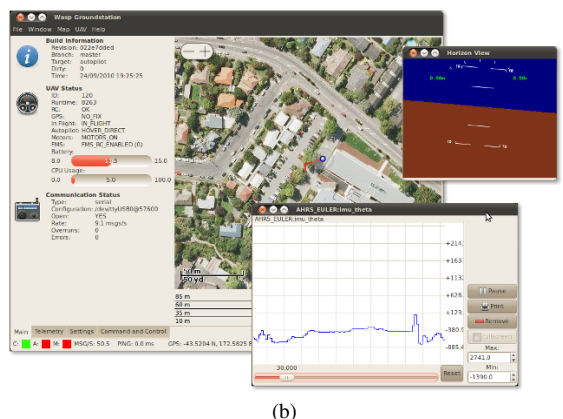
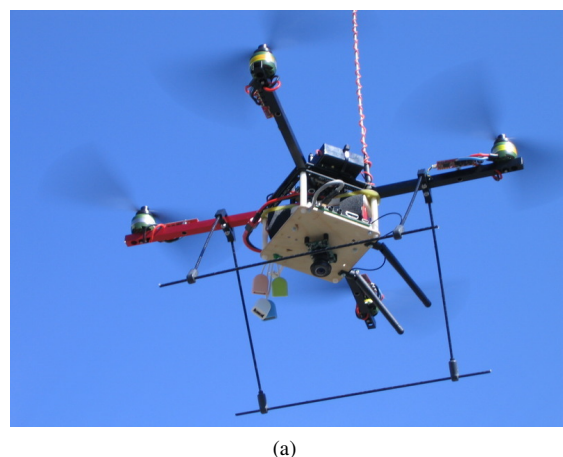


Figure 1: (a) Quadrotor hardware and camera for onboard vision processing. (b) Groundstation software for monitoring real time controller performance.

¹<http://www.waspuav.org>

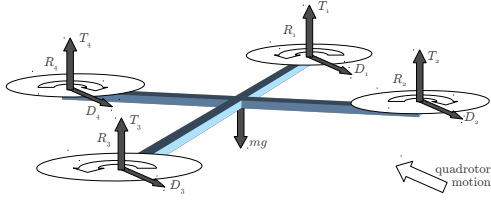


Figure 2: The coordinate system and free body diagram of a quadrotor helicopter.

2 QUADROTOR HELICOPTERS

The first quadrotors for UAV research were developed by Pounds *et al.* [1], Bouabdallah *et al.* [2], and are now a popular rotorcraft concept for unmanned aerial vehicle (UAV) research platforms. The vehicle consists of four rotors in total, with two pairs of counter-rotating, fixed-pitch blades located at the four corners of the aircraft (Figure 1a).

Due to its specific capabilities, quadrotors also provide a good basis for visual flight control research. First, quadrotors do not require complex mechanical control linkages for rotor actuation, relying instead on fixed pitch rotors and using variation in motor speed for vehicle control. This simplifies both the design and maintenance of the vehicle. Second, the use of four rotors ensures that individual rotors are smaller in diameter than the equivalent main rotor on a helicopter, relative to the airframe size. The individual rotors, therefore, store less kinetic energy during flight, mitigating the risk posed by the rotors should they collide with people or objects. These combined, greatly accelerate the design and test flight process by allowing testing to take place indoors, by inexperienced pilots, with a short turnaround time for recovery from incidents. Finally, the improvement of Lithium-polymer battery technology has enabled longer flight times with heavier payloads, increasing the computational power that can be carried onboard, thus increasing the complexity of visual algorithms that can be experimented in realtime.

2.1 Dynamic Modelling

Consider the free body diagram of Figure 2. Each rotor produces a thrust T_i , a drag moment Q_i about its axis of rotation, a rolling moment R_i perpendicular to its axis of rotation, and a drag force D_i in the direction e_{wi} , which is opposite to direction of its relative velocity with the oncoming wind. The drag moments Q_i are proportional to the rotor thrust, and are given by $Q_i = K_r T_i$. Let the direction of flight be e_W . Let x_B , y_B and z_B denote the body axes. Denote the position vector from the center of gravity of the vehicle to the rotor disc center by r_i . Let the free stream dynamic pressure be q_0 , the drag coefficient of the aircraft in forward flight C_D , the reference area S , the

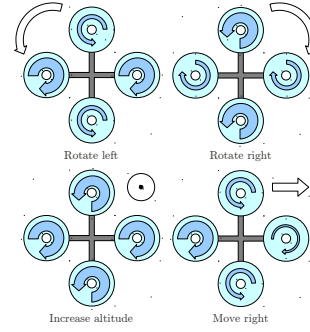


Figure 3: A description of quadrotor motion observed in response to different combinations of motor speed (arrow width is proportional to motor speed).

mass of the aircraft m , and gravity g . The total force \mathbf{F} and moment \mathbf{M} acting on the vehicle in the body frame are as follows;

$$\mathbf{F} = -q_0 S C_D \mathbf{e}_W + m\mathbf{g} + \sum_{i=1}^4 [-T_i \mathbf{z}_B - D_i \mathbf{e}_W] \quad (1)$$

$$\mathbf{M} = \sum_{i=1}^4 [Q_i \mathbf{z}_B + R_i \mathbf{e}_{W_i} + D_i (\mathbf{r}_i \times \mathbf{e}_{W_i}) + T_i (\mathbf{r}_i \times \mathbf{z}_B)]. \quad (2)$$

The reason for the four-rotor² design now becomes apparent. This configuration takes advantage of the symmetry in the system and cancels out several moments. The rotors can be grouped into the front-back pair, and the left-right pair. Consider Figure 3, suppose the pairs rotate in opposite directions. There are then two rotors spinning clockwise and two rotors spinning anti-clockwise. The yawing moments generated by the rotors cancel out to zero. A rolling moment can be generated by speeding up one of the motors in the left-right pair and slowing the other by the same amount. Similarly, a pitching moment can be generated by speeding up one of the motors in the front-back pair and slowing down the other by the same amount. A yawing moment can be generated by speeding up one pair and slowing down the other by the same amount. Note that in each of the above cases, the only effect is the generation of a single moment. The sum of rotor thrusts is held constant.

2.2 Control

When considering attitude control of quadrotors, modelling work undertaken [3, 4, 5] found that secondary aerodynamic effects such as blade flapping³, thrust dependence on free stream velocity, and interference caused by the ve-

² Actually, by extension, any even-numbered symmetric rotor design.

³ The flexion of rotor blades due to difference in relative air velocity in translational flight.

hicle body in the slip-stream of the rotor could be disregarded for hover and slow translational flight. Furthermore, Pounds *et al.* [6] found that the regulation performance of simple proportional, integral, derivative (PID) control schemes was within the $\pm 2^\circ$ of level tracking, and that furthermore, PID may offer an advantage due to their simplicity and potential robustness to parameter variation [7].

In the positional flight control case, Hoffmann *et al.* [5] experimented with positional control implemented using a PID controller with the vehicle's pitch and roll as control inputs. Tilting the vehicle in any direction causes a component of the thrust vector to point in that direction, so commanding pitch and roll is directly analogous to commanding accelerations in the X-Y plane. However, a key weakness of this and similar position controllers was the assumption that the velocity of the free stream and attitude control are decoupled. This is in fact only true for very small velocities.

Insofar as this concerns visual flight control, one can conclude with confidence that positional control and navigation can be implemented on top of, at minimum, a PID attitude controller. By commanding the attitude controller with pitch and roll deviations from hover, one can regulate position with good accuracy providing the flight dynamics are slow (close to hover), and wind is small (such as in indoor flight).

3 VISUAL FLIGHT CONTROL RESEARCH

According to flying principle and the propulsion mode, one can classify aircraft in multiple categories as shown in Figure 4. This classification is important, as different flight vehicles pose different challenges for visual control, for example, Srinivasan *et al.* [8] utilise a dominant motion assumption, and outdoor flight conditions to construct a biologically inspired control system for a fixed wing plane. Similarly Bachrach *et al.* [9] implement exploratory navigation and mapping possible only on a hovering platform with freedom to move in all directions.

3.1 Existing Research Platforms

Multiple universities and commercial companies are currently researching visual flight control, and are using quadrotor helicopters as part of that research. The following list includes notable contributors to this field. It is instructive to consider each groups technical solutions to visual flight control challenges. Technical terms not explained below, and where appropriate, similarities and differences with the *wasp* system will be further expanded in a comparison (section 4.3).

Pixhawk The Pixhawk⁴ *cheetah* quadrotor is a custom

⁴http://pixhawk.ethz.ch/micro_air_vehicle/

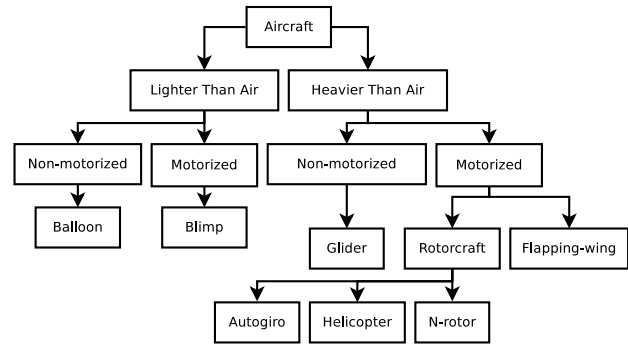


Figure 4: Aircraft classification depending on propulsion and flying principle.

quadrotor platform, featuring a bare-metal attitude controller and IMU, Core2Duo single board computer, and supporting software. Visual flight control algorithms run onboard.

CityFlyer The CityFlyer⁵ project uses the Asctec *pelican* quadrotor for real time attitude control, and custom higher level control processes integrate with the system using the ROS framework.

GRASP The GRASP⁶ at University of Pennsylvania uses the Asctec *hummingbird* quadrotor frame and a custom controller for attitude stabilization. Visual control is performed off-board, indirectly through the use of a Vicon⁷ system.

Universitat Tübingen The Universitat Tübingen⁸ autonomous flying robots research project uses the Asctec *hummingbird* quadrotor, and visual control is performed onboard using a microcontroller and low resolution infrared tracking via a Nintendo wii camera.

sFly The sFly⁹ project at ETH Zürich focuses on visual flight control and co-ordination of multiple robots. The quadrotor platform is provided by Asctec. Visual flight control occurs onboard using the same SBC as in Pixhawk.

Ascending Technologies (Asctec) Asctec¹⁰ is a commercial company providing a number of large quadrotors for the research arena. The *pelican* and *hummingbird* models are frequently seen in literature as the Asctec provided attitude controller is robust, and the

quadrotor/cheetah

⁵<http://robotics.ccny.cuny.edu/blog/node/20>

⁶<http://www.grasp.upenn.edu/>

⁷Vicon is a motion capture system that is capable of providing real time localisation in 6DOF of multiple bodies at 200 Hz.

⁸http://www.ra.cs.uni-tuebingen.de/forschung/flyingRobots/welcome_e.html

⁹<http://projects.asl.ethz.ch/sfly/doku.php>

¹⁰<http://www.asctec.de/>

real time telemetry available is well suited to integration with visual flight control systems, for either onboard or offboard processing.

MIT The robust robotics group¹¹ at MIT uses an Asctec *pelican* quadrotor, and a distributed off-board vision system [9].

Microcopter MikroKopter¹² is the largest open-source quadrotor system, and provides another excellent basis for research. The typical suite of MikroKopter software includes an autopilot with attitude and GPS position modes, and groundstation software.

Paparazzi The paparazzi project is an established system hardware and software, featuring autopilot implementations and groundstation software for fixed wing, and more recently through the *booz* project, quadrotor helicopters too. It does not offer visual flight control facilities. *Wasp* was originally derived from paparazzi code in 2008.

3.2 Important Considerations

Historically visual flight control has been simulated or performed off-board. As computing power increases, and improved visual techniques are discovered, it is becoming more possible to move this computing onto the flying robot. Bachrach *et al.* and the MIT implementation [9] represent the state of the art for offboard visual processing; the quadrotor is able to autonomously explore its environment, while mapping and searching along the way. The Pixhawk project represents state of the art for onboard visual processing; it is able to autonomously hover and navigate a predefined map.

While both onboard and offboard approaches seem distinct, they both share similar considerations;

Position control updates There is a lower bound on the acceptable rate of position control¹³ if one wants to achieve stable hovering. In the authors experience¹⁴ any vision based position controller should, assuming a reasonable degree of accuracy, provide attitude commands at ≥ 10 Hz

State estimation A combined (visual and inertial) state estimator requires synchronized IMU/camera images for best performance.

Insensitivity to computation time The difference in computation times between simple visual processes,

¹¹<http://groups.csail.mit.edu/rrg/research.html>

¹²<http://www.mikrokoetter.de/>

¹³Assuming that the position controller is implemented atop a PID attitude controller, as described in section 2, and the craft is operating in low wind and near hover conditions.

¹⁴Gathered while working on the Pixhawk project early January 2010, through his own experiments, and through discussions with other researchers at micro air vehicle conferences.

such as obstacle avoidance, and complex ones, such as simultaneous localisation and mapping (SLAM) on a large map can differ by a factor of 10 *x* or more. For this reason, it is often a practical necessity to run these processes only as needed. Furthermore, a complex robotic system often has additional overhead in the form of processes for debugging, visualisation and testing.

The *wasp* system was designed to make these three considerations easy to implement.

4 THE WASP SYSTEM

Wasp is a collection of software and hardware (collectively referred to as the *wasp* system), available under an open source, free software licence¹⁵, that facilitates rapid implementation of control systems for robotics research. *Wasp* has been successfully used to control quadrotor helicopters, fixed wing aircraft and rockets. The system includes a number of components;

1. quadrotor autopilot featuring attitude, hover and altitude hold modes,
2. fixed-wing autopilot with GPS position control capability,
3. groundstation software for visualisation, command and control of UAV flights,
4. software in the loop simulation for fixed wing and quadrotor flight vehicles,
5. calibration software for IMU and magnetometer,
6. hardware designs for quadrotor autopilot, IMU, GPS, and Gumstix based single board computer,
7. utility libraries of software for visual flight control.

Section 4.1 describes the overall system architecture, Section 4.2 will describe aspects of the *wasp* software that are most useful for visual flight control research, particularly in reference to the considerations expressed in section 3.2 and section 4.3 will compare the *wasp* system with other approaches.

4.1 Architecture

Consider Figure 5, the common element which connects the *wasp* system together is communications, a standardised message description language. This intermediate language makes it easy to add new messages, for debugging, telemetry or control, containing information one might

¹⁵*Wasp* was originally derived from Paparazzi in 2008, and has subsequently diverged in capability, design and focus, although code still flows between both projects occasionally.

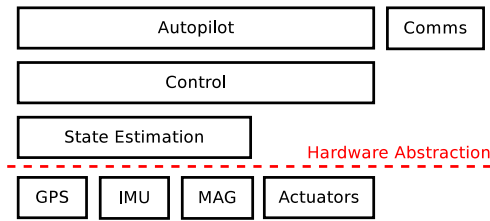


Figure 6: The *wasp* onboard software architecture. Each block has defined interfaces for code re-use on other aircraft, and other hardware.

want to send between robots or to the groundstation. C-code for parsing, sending and receiving these messages is generated programmatically. In dynamic languages, it is easy to parse this message description directly (such as is done in the Python groundstation).

The transport of these messages is sufficiently abstracted, and their content opaque, that they are utilised in current research (Figure 1a) to allow visual control code to run either on a desktop PC or a Gumstix computer on the quadrotor. More specifically and for example, *wasp* defines a command messages instructing the autopilot which attitude it should hold; as explained earlier, this is a sufficient platform upon which to implement visual position control.

4.2 Software

The *wasp* software is a mixture of C/C++ and Python code. The attitude controller includes an extended Kalman filter and a complementary filter state estimator. Onboard code is abstracted in terms of common elements of robotic control (Figure 6), making re-use on different aircraft relatively easy. Hardware implementations for ARM7 (section 4.4) and Gumstix exist, as do controllers for fixed-wing airplanes and rockets.

Additionally, *wasp* also includes software utilities and libraries for preparing synchronized IMU measurements and images acquired from various cameras. This allows fluid mixing of visual control code, it can either run on the same SBC as the attitude controller, or separately.

4.3 Comparison With Other Systems

The Pixhawk and MIT quadrotor systems also utilise an intermediate message description language, but extend the concept further by defaulting to peer-to-peer or broadcast-subscription transport models. This is more able to facilitate distributed computation on multiple nodes (such as different computers running different vision algorithms), but it is debatable whether mandating such a transport makes things simpler in the case of a single onboard computer.

The GRASP and CityFlyer projects utilise ROS¹⁶, which is a standardised version of the approach implemented by Pixhawk¹⁷. ROS provides numerous building blocks implemented atop of a standardised message description and transport layer. This allows the use of tested highly functional blocks such as SLAM, laser scanners, groundstation software, and much more. ROS is a robust and healthy open source project, but is of an order of magnitude more complex¹⁸ and capable than the *wasp* system. Similar criticisms to the Pixhawk and MIT approach apply with respect to the necessity of such complexity for a single onboard computer.

4.4 Quadrotor Hardware and Construction

The attitude controller is equipped with a NXP LPC2148 32-Bit ARM 7 Microprocessor running at 60 MHz. For height control the controller board provides a MPX4115 absolute air pressure sensor also connected to the 10 Bit ADC. For autonomous flight an additional PCB is used, providing a uBlox GPS-Module and antenna. An inertial measurement unit (IMU) is also present, and contains a 3-axis accelerometer, 3x single-axis gyroscopes, and a 3-axis magnetometer (Figure 5).

The quadrotor hardware consists of a cross-frame made of square aluminum tubes joined by plates in the center. This design has proved to be robust, usually only requiring a propeller replacement after a crash. On this frame are mounted four brushless motor / propeller combinations, four brushless motor controllers, the avionics and the battery. The two pairs of opposite rotating propellers are mounted at a distance of 0.37 m.

4.5 Quadrotor Drive Section

Pounds *et al.* [6] noted that actuator performance is a critical element in attitude controller performance. The Roxxy2824-34 (48 g, 1000 rpm/V, $i_{max}=8$ A) brushless motors are driven by sophisticated brushless controllers¹⁹. The timing of the commutations, as well as the shape of generated signals greatly affects the performances of the rotor system, both in efficiency and dynamics. The performance of the propulsion system is also affected by the refresh rate and synchronous/asynchronous nature of the commands sent to the controller. Commercial hobby brushless controllers typically use a servo PPM signal with a refresh rate of 40 Hz, introducing a 25 ms delay. The controllers use digital communications for a refresh rate up to 1 kHz, reducing the delay to 1 ms. Simulation [5]

¹⁶<http://www.ros.org/wiki/>

¹⁷Note, Pixhawk is anticipating moving to ROS in the near future.

¹⁸In terms of lines of code, levels of abstraction and time to learn.

¹⁹http://www.torcmann.de/peterslrk/SPEEDY-BL_eng.html

²⁰<http://www.mikrokopter.de/ucwiki/en/BrushlessCtrl>

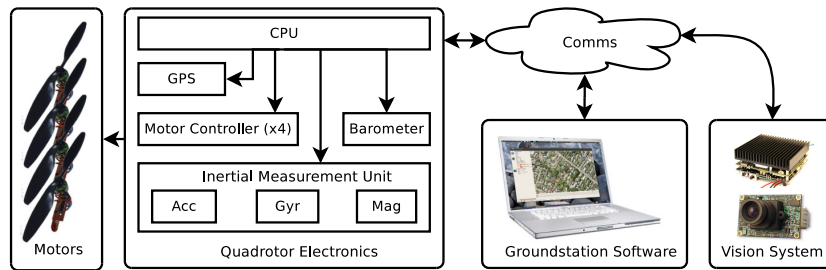


Figure 5: The *wasp* system and its application to a quadrotor helicopter.

and experimental results show that this parameter plays a significant role in the performance of the craft.

The brushless controllers are connected via I2C. The I2C interface has a data rate of 400 kHz and control loop refresh rate of 200 Hz. The motor and controller generate, in combination with 10"x4.5" propellers, a maximum thrust of 5.6 N for each motor. The whole drive train is powered by a LIPO 3 cells battery (11.4 V, 2100 mAh, 180 g). The total, ready to fly, mass of the quadrotor is 750 g yielding 25 minutes flight time. The maximum payload is approximately 250 g. The measured maximum flight speed is approximately 60 km/h.

5 DISCUSSION OF CURRENT WORK AND CONCLUSIONS

The author has implemented, using the *wasp* system, visual attitude estimation [10], biologically inspired obstacle avoidance²¹, and is currently working on improved visual-pose estimation for hovering and exploration. Colleagues and students have also used the *wasp* system for research on SLAM, embedded Linux, and have applied (non visual control only thus far) to different aircraft.

The *wasp* system is smaller in scope and features compared to other frameworks like ROS but is easier to understand, and in the authors opinion, more suited to onboard processing. Additionally *wasp* also provides lower level building blocks for controlling other aircraft, facilitating their use too as flight control vehicles.

REFERENCES

- [1] P. Pounds, R. Mahony, P. Hynes, and J. Roberts, "Design of a Four-Rotor Aerial Robot," (Auckland, New Zealand), November 2002.
- [2] S. Bouabdallah, P. Murrieri, and R. Siegwart, "Design and control of an indoor micro quadrotor," vol. 5, pp. 4393–4398, apr. 2004.
- [3] P. Pounds, R. Mahony, and P. Corke, "Modelling and control of a quad-rotor robot," in *Proceedings Australasian Conference on Robotics and Automation 2006*, Australian Robotics and Automation Association Inc., 2006.
- [4] G. Hoffmann, D. G. Rajnarayan, S. L. Wasl, P. C. C. J. Tomlin, and A. Professor, "The Stanford Testbed of Autonomous Rotorcraft for Multi Agent Control (STARMAC)," in *In Proceedings of the 23rd Digital Avionics Systems Conference*, 2004.
- [5] G. M. Hoffmann, H. Huang, S. L. Waslander, and C. J. Tomlin, "Quadrotor Helicopter Flight Dynamics and Control Theory and Experiments," in *AIAA Guidance, Navigation and Control Conference*, (Hilton Head, South Carolina), AIAA, aug 2007.
- [6] P. Pounds, R. Mahony, and P. Corke, "Modelling and control of a large quadrotor robot," *Control Engineering Practice*, vol. 18, no. 7, pp. 691–699, 2010. Special Issue on Aerial Robotics.
- [7] S. Bouabdallah, A. Noth, and R. Siegwart, "PID vs LQ control techniques applied to an indoor micro quadrotor," vol. 3, pp. 2451–2456, sep. 2004.
- [8] M. V. Srinivasan, S. Thurogood, and D. Soccol, "An Optical System for Guidance of Terrain Following in UAVs," *Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on*, pp. 51–51, 2006.
- [9] A. Bachrach, R. He, and N. Roy, "Autonomous Flight in Unknown Indoor Environments," *International Journal of Micro Air Vehicles*, vol. 1, no. 4, p. 217, 2010.
- [10] J. Stowers, A. Bainbridge-Smith, M. Hayes, and S. Mills, "Optical Flow for Heading Estimation of a Quadrotor Helicopter," *International Journal of Micro Air Vehicles*, vol. 1, no. 4, p. 229, 2010.

²¹Paper currently under review. It introduces an avoidance strategy which works by detecting objects from a relative depth map generated via optical flow.

Embedded Linux Controlled Quadrotor Helicopter

Alex Millane*, Michael Hayes, John Stowers

*Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand

Email: ami61@uclive.ac.nz

Abstract: Quadrotor helicopters make useful flight platforms for computer vision research. To ease the implementation of vision controlled flight it is desirable to have vision software and flight control software running on the same platform. This paper details the design of a quadrotor helicopter controlled by a Gumstix single board computer (SBC) running embedded Linux.

Keywords: Embedded Linux, Quadrotor, Helicopter, UAV

1 INTRODUCTION

A quadrotor helicopter is a helicopter propelled by four rotors[1]. Control is achieved by varying the speeds of each of the rotors. For use as unmanned aerial vehicles (UAVs), quadrotors offer a number of advantages over single-rotor helicopter designs. Firstly, quadrotor helicopters do not require mechanical linkages to vary the blade pitch whilst spinning. This reduces the mechanical complexity of the design [1]. Secondly, because the propellers are controlled such that pairs of blades counter-rotate there is no need for an anti-torque tail rotor, which again reduces the machine's complexity. Thirdly, as they have four blades, the size of each of the blades is able to be reduced, which makes the craft more suitable for indoor flight.

This paper details the design of quadrotor helicopter designed for use as a platform for computer vision research. To increase the intimacy between computer vision and flight control it is desirable that the two software subsystems run on the same computing platform. As such this quadrotor is controlled by a Gumstix¹ Single Board Computer (SBC) running embedded Linux. This required porting the *wasp*² quadrotor control software so it ran on a Gumstix SBC. Production of a mechanical design was required and is discussed in Section 2. Following that Section 3 discusses the production of the electrical hardware that gives the Gumstix SBC the ability to control the quadrotor in flight. Elements of the software architecture are discussed in Section 4.

2 MECHANICAL DESIGN

Composite sandwich panels were used to simplify the design and construction of the quadrotor's frame. Such panels have a high strength to weight ratio [2] and the panel geometry offers the advantage that all planar body elements can be manufactured from a single piece. Figure 1 shows the solid model prototype for the final design using an aluminum honeycomb panel. Aluminum honeycomb-panel sandwiches a honeycomb shaped aluminum alloy extrusion with facing plates made of another aluminum alloy.

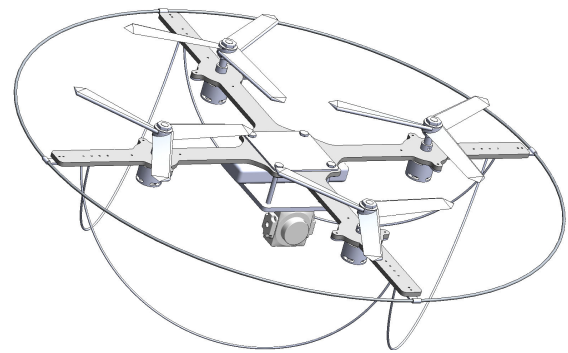


Figure 1: An solid model of an aluminium honeycomb prototype.

The size and shape of the design shown in Figure 1 were chosen to minimize stress in the aluminum honeycomb material. Stress analysis was carried out using a twofold process. Finite element analysis was used to show the effect of the body geometry on its strength assuming an isotropic material. Following this analysis, the areas of in-

¹<http://www.gumstix.com/>

²<http://www.waspuav.org>

terest, identified in the finite element tests, were analyzed for stress levels using techniques for analyzing composite honeycomb paneling described by Piak et al. [2]. New iterations of the design were then produced to reduce stress levels in the body.

Since the body structure and associated components are lightweight the static stresses in the frame are small and it was deemed important to analyze the impact loading of the structure. Castiglianos theorem [3], [4] was used to calculate the bending of the protection beam to ensure that its motion under impact loading was acceptable.

Following its design, the frame shape was water-cut from an aluminum honeycomb panel. Subsequently, the frame was assembled and the propulsion system and electrical hardware were attached. Figure 2 shows the completed quadrotor with all components and electronic hardware attached.

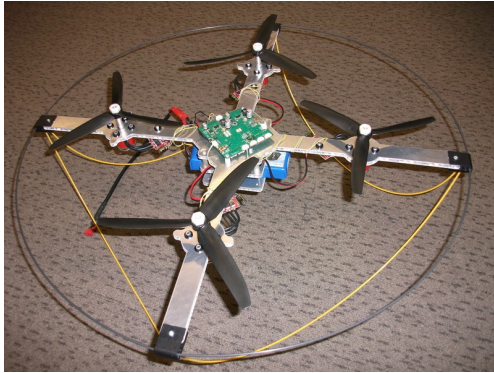


Figure 2: The final quadrotor helicopter.

The second part of the mechanical design was the selection of a propulsion/control system. The main components of that system are listed below:

- **Brushless DC motors** were used to provide a sufficiently high power to weight ratio to lift the craft. Additionally, brushless DC motors were selected to allow closed loop control of the motor speed [5]. They also offer efficiency advantages over other motor types and are powerful enough to lift the projected mass of the craft. The completed quadrotor mass is around 800 g while the predicted maximum thrust output of the proposed motor/propeller set up is around 20 N.
- Two pairs of contra-rotating three-bladed props were selected for use in the quadrotor. Three-bladed props have been used in preference to two bladed props to reduce the blade diameter to power ratio so that the overall footprint of the craft is smaller, allowing a lighter construction. Contra-rotating props are a basic requirement of quadrotor helicopters to allow yaw control of the craft [1].

- To provide sufficient current output and high power density, lithium polymer batteries were selected. The selected batteries give the craft a projected flight-time of around 15 minutes. The battery makes up around 40% of the quadrotor's mass.

3 ELECTRICAL HARDWARE DESIGN

The electrical hardware provides the necessary interfaces to the Gumstix SBC so that it is able to control the quadrotor's flight. Additionally, so that the quadrotor can fulfill its role as a research platform, the electrical hardware provides generic interfaces to additional devices, for example computer vision equipment. An overview of the components of the quadrotor system necessary for flight and their interconnects is shown in Fig. 3. A description of these devices follows.

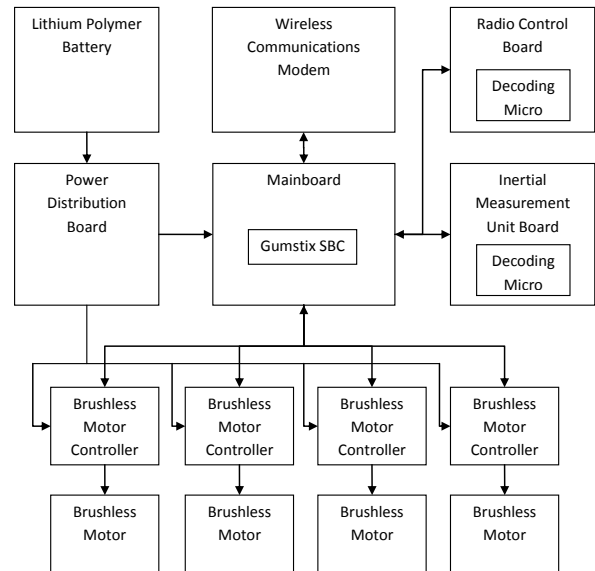


Figure 3: An overview of the electrical hardware of the quadrotor system.

- **Gumstix:** The Gumstix SBC integrates the core computer hardware onto a single PCB. The Gumstix Overo range features TI OMAP35xx processors with ARM[®] Cortex[™]-A8 architecture to give laptop-like performance.
- **Main Board:** This provides the Gumstix SBC with the interface circuitry required for communication with the other system components as well as power at the appropriate voltage levels for the Gumstix and the interfaced devices. The specifics of the interfaces required of this board are discussed in Section 3.3.

- **Power Distribution Board:** The lithium polymer battery is connected to a power distribution PCB which distributes power to the mainboard and each of the four brushless motor controller boards.
- **Brushless Motor Controller:** The four brushless motor controller PCBs produce the tri-phase waveforms to power the brushless motors. These controllers also handle the closed loop control of the motor speed, maintaining a commanded speed in the presence of a dynamic load.
- **Radio Control Board:** The radio control board provides the wireless link to a radio control unit.
- **Wireless Communications Modem:** This modem enables communication between the craft and a ground-station.
- **Inertial Measurement Unit Board:** The IMU provides acceleration data to allow attitude control and navigation of the quadrotor helicopter.

3.1 Inertial Measurement Unit Board Design

The inertial measurement unit (IMU) includes several inertial measurement devices which produce data asynchronously. These devices include a three-axis accelerometer, a three-axis gyroscope, a three-axis magnetometer and a barometer.

3.2 Radio Control Board Design

This board houses a radio control (RC) receiver. The pulse-position modulation (PPM) signal received by this receiver is passed to the timer capture input of the OMAP processor for decoding of the pulse width.

3.3 Main Board Design

To provide interfaces to flight control devices as well as vision devices, a custom carrier board for the Gumstix single board computer was designed to provide all of the core interface circuitry. In accordance with the project goals, the hardware provides generic interfaces for use with devices used in research.

3.3.1 Interfaces

Figure 4 shows the interfaces provided by the quadrotor main board. A description of each interface follows.

- **Power Supply Devices:** The Gumstix SBC and the devices to which it connects require power at several voltage levels. The chosen power supply system includes two switch-mode regulators as well as a linear regulator to produce power for devices on and off the board.

- **Console Debugging:** The Gumstix provides console output via a UART port for communication and development with the computer. This data has been passed through an FTDI serial to USB chip so that development can be carried out over USB.
- **USB Host/OTG:** USB ports have been provided so that computer vision equipment can be connected. Furthermore, these USB ports ease development by providing a means of transferring data as well as by allowing connection to a keyboard etc.
- **I²C Port:** The Gumstix quadrotor uses I²C to interface to the brushless motor controllers and the IMU. All of these devices are provided with connections to the OMAP3 I²C port that the Gumstix exposes through its connections to the main board. Additional connectors to the I²C bus have been included for addition of further I²C devices.
- **UART Port:** The Gumstix's UART ports are to be used for communication with a GPS receiver unit as well as a GSM modem, should it be required. The UART ports are passed through a logic level translator for selectable logic levels.
- **High Current Outputs:** High current output drivers have been included so that actuators can be interfaced to the SBC. These drivers are a pair of totem-pole MOSFET drivers driven by a MOSFET gate driver IC.
- **Servo Outputs:** To allow articulation of the camera in-flight, several dedicated servo control outputs have been included. 5V power is provided on the mainboard for the servo connections and offers up to 3A current draw.
- **GPIO/SPI Ports:** As the quadrotor is to serve as a research platform it is impossible to know what exactly it will be used for. As such provision for additional devices has been provided. SPI and GPIO ports with selectable voltage levels have been provided to allow connection of additional devices to the quadrotor.

3.3.2 Board Layout

When designing the layout of the Gumstix interface board there were two main concerns. Firstly, the size of the quadrotor's frame imposed limits on the maximum size of the mainboard PCB. Secondly, because the OMAP3 housed on the Gumstix SBC runs at 600 MHz, reducing the noise produced by high speed digital signals emanating from this chip was an important objective. For a higher component density, to reduce the board's size, and increase the board's immunity to noise, a four layer board was produced. Furthermore, the switch mode power devices, which are a source of switching noise, were placed

performance. Additionally, doing so reduces the difficulty of synchronizing IMU measurements with the associated image frame captured from the camera.

Running the flight software on top of the Linux kernel, like any operating system, offers the advantage of abstraction from the computer platform's hardware. As a result, the APIs required by the Gumstix to run the quadrotor flight software will be generic to any Linux system exposing the same interfaces no matter the platform it is running on. Conversely, by relinquishing direct control of the computer hardware, the flight control software has less control over the timing of events. This impacts the software's performance in a real-time system. This topic is discussed in Section 6.

The architecture of the flight software running on the quadrotor is shown in Figure 7. The dashed box indicates the part of the existing flight control software that needs to be written to port the current control code for use with the Gumstix Overo. Note that in the case of the previous quadrotor, the software APIs interact directly with the microprocessor hardware, whereas in the case of this hardware, the APIs must interact with the Linux kernel.

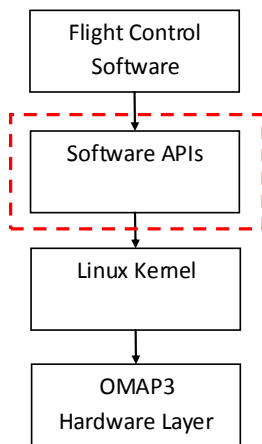


Figure 7: The architecture of the new quadrotor software.

4.1 Linux Kernel Production

The Gumstix Overo SBC is intended to run embedded versions of the Linux kernel. The kernel distribution selected for use with the quadrotor project is the Angstrom Distribution³ because it offers much support for the Gumstix SBC. To allow for the control of the quadrotor and to ease development for the Gumstix, custom versions of the Angstrom Linux kernel were produced.

Development of packages to run on the Gumstix SBC and the production of images, consisting of the root file

system and Linux kernel, are achieved using the OpenEmbedded⁴ and BitBake⁵ tools. The BitBake build-tool is a tool for cross compiling packages for distribution on embedded linux systems. A BitBake file which specifies how to build a package is called a BitBake recipe. A BitBake recipe specifies package sources as well as instructions on compilation and installation. Executing the BitBake tool on such a file will collect the sources and their dependencies, cross compile the package and package it for distribution. The OpenEmbedded project is a build framework for embedded Linux. It maintains a collection of BitBake recipes which produce embedded Linux packages for different platforms, one of those being the Gumstix SBC.

By using these two packages a custom version of the Angstrom Linux kernel was developed. A number of packages/modules were included in this kernel and were useful for development. USB Ethernet Gadget was used for connecting the Gumstix SBC to the Internet through a host computer's Ethernet bridge. Dropbear SSH server was used on the Gumstix to allow the transfer of files between the host file system and the Gumstix's file system. Lastly the I²C tools package was used for debugging the I²C bus used by the Gumstix for serial communication. Note that these tools were useful for development but will consume processor time if left to run alongside the flight software.

4.2 Flight Control Consolidation

To port flight control software for application on the Gumstix SBC a number of APIs need to be written. For a minimal implementation of the flight control software APIs connecting the software to the radio control unit, the wireless communications unit, the inertial measurement unit, and the brushless motor controllers need to be implemented. Some of these APIs require the Gumstix to communicate with external devices. Therefore, a number of drivers to allow communication with these devices are also needed. The quadrotor requires drivers for I²C, SPI, UART, USB, GPIO, and ADC peripherals which also need reimplement on the Gumstix architecture (see Figure 3).

For example, consider the actuator API. This allowed communication with the brushless motor controllers. As the quadrotor communicates with the motor controllers over an I²C bus connection this required communicating with the Linux I²C driver in Linux user space.

5 TESTING

To test the quadrotor hardware, a user interface to the actuator API was written to allow control of the quadrotor's

³<http://www.angstrom-distribution.org/>

⁴http://wiki.openembedded.net/index.php/Main_Page

⁵<http://developer.berlios.de/projects/bitbake>

brushless motors from the terminal. Using this user interface motor speed commands were sent to the motors via the quadrotor's USB console connection. This testing showed that the propulsion system produced enough thrust to lift the raise the flyer's frame at around 50% motor speed.

6 DISCUSSION

In this application the Gumstix is being tested for its ability to meet the real-time requirements of flight. By having the Linux kernel running underneath the flight control software, the way the program runs is less deterministic than if the software was running directly on microcontroller hardware, as it has in the past. This is because as well as running the flight control software the Linux kernel also has to service a host of other processes. Therefore, there will be times when the flight control software is not responding the changes in the state of the quadrotor. It will be important to ensure that a Linux kernel of minimum complexity is running on the Gumstix so that no unnecessary processes are running in parallel with the flight control software and sharing processor time.

A hard real-time system is one where the completion of an operation after its deadline is considered useless, while a soft real-time system tolerates such lateness but may respond with reduced service quality. Of the tasks that the flight control software has to perform, the only task with a hard real-time requirement is the updating of the motor speeds based on the attitude of the quadrotor to maintain stability. Failing to update the motor speeds based on changes in attitude, as communicated by the IMU, in a short time frame will result in the quadrotor losing stability.

The Linux kernel provides tools to mark a process as one with real-time requirements. This indicates to the kernel that the process "should have a short response time and, most important, such response time should have a minimum variance" [8]. Marking a process like this will mean that it will never be blocked by lower priority processes from access to shared resources. This does not mean, however, that the Linux kernel is able to support the most demanding real-time processes. This is because the Linux kernel is non-preemptive which means that processes which enter Kernel Mode, while handling an interrupt or an exception, are unable to be interrupted by higher priority processes. This reduces the predictability of response times. It remains to be seen whether these delays may prevent the flight software from meeting the real-time requirements of the attitude control system.

7 CONCLUSIONS AND FURTHER WORK

Significant progress has been made towards a Linux computer controlled UAV for computer vision research. A full mechanical design as well as electrical hardware which gives the Gumstix SBC the requisite interfaces to control the quadrotor in flight have been developed. Porting of flight control software has been started with a version of the Angstrom Linux kernel produced using the BitBake and OpenEmbedded build tools. The first of the APIs required to control the quadrotor has been written and tested. From here, the remaining software APIs and device drivers need to be written to finish the port of the flight control software. A testing/analysis will then need to be conducted to determine if the flight control software is able to meet the real-time requirements of flight.

I am grateful for the help given to me by Michael Cusdin and David Healy for the effort they have put into helping with the electrical design and fabrication and the mechanical design and fabrication respectively.

REFERENCES

- [1] P. Castillo, R. Lozano, and A. Dzul, *Modelling and Control of Mini-Flying Machines*. Springer, 2005.
- [2] J. Piak, A. Thayamballi, and G. Kim, *Strength Characteristics of Aluminum Honeycomb Sandwich Panels*. Elsevier Science Ltd, 1999.
- [3] A. Boresi and R. Schmidt, *Advanced Mechanics of Materials*. John Wiley & Sons, 2003.
- [4] J. Shigley, C. Mischke, and R. Budynas, *Mechanic Materials Design*. McGraw-Hall, 2004.
- [5] L. O. Connor, "Brushless motors find their niche," *Mechanical Engineering*, 1994. Retrieved September 26, 2010, from ABI/INFORM Global. (Document ID: 24501).
- [6] J. Stowers, A. Bambridge-Smith, M. Hayes, and S. Mills, *Optical Flow for Heading Estimation of a Quadrotor Helicopter*. EMAV, 2009.
- [7] A. Drouinm, T. Miquel, and F. Mora-Camino, "Non-linear control structures for rotorcraft positioning," in *Proc. of the AIAA Guidance, Navigation and Control Conference and Exhibition*, August 2008.
- [8] D. Bovet and M. Cesati, *Understanding the Linux Kernel*. O'Reilly Media, 2002.

Autonomous Anthropomorphic Robotic Arm using Vision and Colour to Monitor Plant Growth in a Laboratory

Mark Seelye¹, Gourab Sen Gupta¹, John Seelye²

¹School of Engineering and Advanced Technology (SEAT),
Massey University, Palmerston North, New Zealand

²The New Zealand Institute for Plant & Food Research Limited
Private Bag 11 600, Palmerston North, New Zealand
Email: marksnz@xtra.co.nz, G.SenGupta@massey.ac.nz

Abstract:

An autonomous anthropomorphic robotic arm has been designed and fabricated to automatically monitor plant tissue growing in a modified clonal propagation system being developed by Plant & Food Research. The custom-fabricated aluminium arm uses a vertical linear ball shaft and high speed stepper motors to provide the various arm joint movements that enable the arm to swivel 180 degrees horizontally. Sensors located at the end of the arm monitor plant growth and the immediate growing environment. These include a compact colour zoom camera on a pan and tilt mount for image capturing, RGB (red, green and blue) colour sensors to monitor leaf colour as well as temperature, relative humidity and carbon dioxide sensors. The robotic arm can reach anywhere over multiple trays (600mm x 600mm) of plantlets. Captured plant tissue images are processed using innovative algorithms to determine tissue, or whole plant, growth rates over specified time periods. Leaf colour sensors provide information on the health status of tissue compared to predetermined optimum values. Custom software fully automates the operation of the arm and the captured data, allowing the arm to return to specified sites (i.e. individual plantlets) at set time intervals to identify subtle changes in growth rates and leaf colour. This will allow plant nutrient levels and the immediate environment to be routinely adjusted in response to this continuous sensing resulting in optimised rapid growth with minimal human input.

Keywords:

Micropropagation, non-destructive sensing, plant growth monitoring, remote sensing, robotic arm

1. INTRODUCTION

Robotics and automated systems are becoming increasingly common in all economic sectors. The International Organization for Standardization defines a robot in ISO 8373 as: "an automatically controlled, reprogrammable, multipurpose, manipulator programmable in three or more axes, which may be either fixed in place or mobile for use in industrial automation applications".

In the past decade improvements in the sophistication and reliability of robotic applications have led to their use in the horticultural sector for routine, often repetitive tasks, which are expensive to do using a highly paid labour force. As this technology moves forward more and more applications are becoming realities

Recent published research includes: the current status of the core technologies required to successfully develop a robotic system for weed control [1], a system for automatic harvesting of numerous agri-science products such as cutting greenhouse-grown flowers [2], and automating cucumber harvesting in greenhouses [3]. Advancements in electronics has empowered engineers to build robots that can operate in unstructured environments [4]. Camera-in-hand robotic systems are becoming popular wherein a camera is mounted on the robot, usually at the hand, to provide images of objects in the robot's work-

space [5]. In the horticulture industry robots are increasingly being used to sort, grade, package and even pick fruits. Fruit can be identified and classified on a continuously moving conveyer belt [6]. Kiwifruit can be picked from orchard vines using an autonomous wheeled robot [7]. Robotic techniques for producing seedlings have been developed. These researchers identified a need to add a machine vision system to detect irregularities in seed trays using a 5-arm robot [8].

Automation of micropropagation for the rapid multiplication of plants and micro-dosing for the precise application of chemicals, such as herbicides, has also gained the attention of researchers. For example, an autonomous robotic system has been described for the micropropagation of a grass species replacing the costly and tedious manual process [9], and a system has been detailed that combines plant recognition, micro-dosing and autonomous robotics [10].

Investigations into non-destructive methods to measure the health status of plants using colour sensors are gaining popularity. Colour sensors can identify subtle changes in leaf colour to help measure plant health. Although limited work has been carried out for real-time systems, images of plants are captured by a digital camera and scanned to identify the colour of selected pixels [11]. In the agri-science industry these colour sensors are

being used to grade specific fruits and vegetables [14,15,16]. Colour sensors are also gaining popularity in other areas of science and engineering, including evaluating bakery products using a colour-based machine vision [12] and in the health sector to determine blood glucose levels [13].

Apart from collecting numerical data, such as pH, temperatures, moisture content and light levels, images of plants are increasingly being used to identify possible changes (or irregularities) over time. This allows plant growers to optimize growth and modify plant liquid media. The work detailed in this paper focuses on the design and development of an anthropomorphic robotic arm for monitoring the growth of plantlets and the immediate growing environment.

The specifications for the arm included the ability to move over multiple 600mm x 600mm trays of tissue on a shelving unit with a high resolution zoom camera and colour sensors mounted at the end of the arm. Images are captured and plant leaf colour monitored during various growth stages.

Section 2 of this paper provides an overview of the system while hardware is detailed in Section 3 and software in Section 4. Robot testing and implementation are detailed in Section 5. The discussion and enumeration of future work are presented in Section 6.

2. SYSTEM OVERVIEW

The basic concept of this project required an autonomous anthropomorphic robotic arm to be designed and fabricated, with a number of sensors including RGB colour, temperature and proximity sensors as well as a compact colour camera. Custom software created in Microsoft Visual Studio (VB.net) allowed for a completely automated operation with very little human input required. A system engineering approach was employed taking the robotic system from a concept to a reality, making use of engineering techniques and integrating standard components to make the final product.

In the following sections the implementation of the robotic system is presented with descriptions of the different hardware components used and the software.

3. SYSTEM HARDWARE

The complete system consists of several modules, each of which is described here.

3.1 Computer

A PC with three COM ports (or USB ports) running Microsoft Windows operating system (XP, Vista or 7) is required and was used. The programming software was Microsoft Visual Studio 2005 (VB.net 2005).

3.2 Robotic Arm

The anthropomorphic robotic arm was designed in the 3D CAD package SolidWorks (2010), with design input from Massey University machining technicians. All components were fabricated in-house using lathes and mills, with more complex and precise components being machined on an Acerman 500 3-Axis Computer Numerical Control (CNC) machine.

Each fabricated part machined on the CNC machine required G-code. This was generated using SolidCAM software (SolidCAM Inc., PA, USA) with the code based on the SolidWorks drawing and dimensions. The CNC machine was then able to cut very precise (to approx 0.001mm) cuts and profiles using a range of cutting tools.

To allow the robotic arm to move vertically, a ball screw and shaft assembly was incorporated, converting rotational motion into vertical movement. The ball shaft has a 5mm lead that, when combined with a microstep driver, can move the arm vertically by fractions of 1 mm. This is essential to ensure that all sensor readings are taken at 'fixed heights'. A number of design changes have been made to the arm to make it lighter, more robust and to give more movement, with designs incorporating Finite Element Analysis (FEA) on the selected parts to ensure safety factors are sufficient for the intended operation.



Fig. 1. SolidWorks rendered photo of completed robotic Arm.

3.3 Motor Controller



Fig. 2. The Ocean Controls motor controller (KTA - 190) [17]

KTA-190 motor controllers (Ocean Controls, VC, Aus) provide an interface between the computer and up to 4 stepper motors as well as the ability to control each motor independently or collectively. Utilizing a RS-232 9600, 8N1 ASCII serial communication protocol up to 4 controller boards can be linked, giving control of up to 16 stepper motors. A motor is controlled by a simple address, followed by the appropriate ASCII commands.

The KTA-190 controller has as interface to allow limit switches to be used to prevent the motors from travelling out of range. With a total of 17 commands it is possible

to tailor the operation and move the motors. Commands include: setting the position of the motors, returning the current positions of the motors, moving the motors by a relative or absolute amount and acceleration settings. A status command returns a 12-bit binary representation on the status of the controller board at any given time, providing information on the movement, direction and status of the limit switch respectively.

3.4 Stepper Motors

The robotic arm required a total of five stepper motors sourced from Ocean Controls. Three FL57STH56-2804B (MOT-122) high torque stepper motors have been used, each with a 1.8° step angle and a holding torque of 12.6kg/cm. This provides the necessary movements of the first 2 joints and vertical movement. The motors moving the arm to the required positions are used in conjunction with a 2:1 gear ratio, while the motor being used to move the arm vertically uses a 3:1 gear ratio, giving the user very precise and fine movement. The remaining two joints utilize FL42STH38-1684B (MOT-120) stepper motors, again with a 1.8° step angle, and holding torque of 3.6 kg/cm.



Fig. 3. High torque stepper motors FL57STH56-2804B (MOT-122 [17]

3.5 Camera

A Sony colour camera (model: FCB-IX11AP), (Figure 4), was used. It features a $\frac{1}{4}$ " CCD (charge coupled device) image sensor using PAL (Phase Alternating Line) encoding system. The camera has a 40x zoom ratio (10x optical, 4x digital) that is controllable from a PC via Sony's VISCA command protocol. The camera's macro capability allows it to capture images as close as 10mm from the subject.

This compact camera (65mm x 44.8mm x 39.3mm), weighing 95g produces high quality images. It has numerous controllable functions (up to 38,000 different command combinations) including digital and optical zoom, aperture setting, shutter speed settings and iris settings. The camera operates on a 6V to 12V DC input.



Fig. 4. Sony FCB-IX11AP colour camera

The camera can operate in light conditions as low as 1.5 Lux and temperatures ranging from 0 to 50°C . The electronic shutter speed is controllable from 1/10 to 1/10000 of a second allowing for clarity in photographs. In order for the camera's functions to be controlled, hexadecimal commands (as serial data) are sent to the camera. These serial commands require 8 data bits, 1 start bit, 1 (or 2) stop bit and no parity. They have a communication speed of 9.6, 19.2 or 38.4 kbps. These specifications are contained in the data packet sent to the camera to control the zoom.

3.6 Sensors

To monitor the colour of the plant material, software can be designed to determine the colour of each pixel on an image. However, a faster way to obtain the RGB colour content of the plant material is to utilize a basic RGB colour sensor. Parallax (Parallax Inc. CA, USA) has two such sensors (ColorPAL and TSC3200) that interface directly with their BASIC Stamp microcontroller, giving quick readings of the RGB values. Both the sensors incorporate a programmable TAOS (Texas Advanced Optoelectronic Solutions, TX, USA) chip, with the ColorPAL using a light-to-voltage sensor while the TSC3200 utilizes a light-to-frequency sensor.

The TCS3200 colour sensor operates by illuminating the object with 2 white LEDs, while an array of photo detectors (each with a red, green, blue and clear filter) interpret the colour being reflected by means of a square wave output whose frequency is proportional to the light reflected. The TCS3200 Colour sensor has a 5.6mm lens, which is positioned to allow an area approx 3.5mm^2 to be viewed. Since the TCS3200 is mounted 20 mm above a sample surface, and therefore not in direct contact with the sample, it was more suited for our application than the full contact required by the ColorPAL sensor.



Fig. 5. Parallax ColorPAL (left) and Parallax TSC3200 (right) colour sensors [18]

Other sensors in use include an ultrasound proximity sensor (Parallax Ping #28015) to ensure the robotic arm is always at a fixed height above the plant material in order to ensure colour readings are taken from a fixed point, a temperature and humidity sensor (Parallax Sensirion SHT11 Sensor Module #28018) and CO_2 sensor (Parallax CO_2 Gas Sensor Module #27929) are also incorporated for constant monitoring of the immediate growing environment. The Parallax sensor modules readily integrate with the Parallax Basic Stamp microcontroller.

4. SOFTWARE IMPLEMENTATION

Microsoft Visual Basic 2005 was selected for this application due to its ease of use and the research team's familiarity with this software. Our program makes use of multiple timer application, which ensures that it runs continuously, repeatedly sending commands and allowing for smooth movements of the arm, camera control and sensors.

The program contains a number of GUI pages giving the user a vast array of options, including the ability to manually move the arm to any position or to specify start and end positions of a tray. The software calculates positions using an inverse kinematic algorithm to determine the required joint angles and appropriate movements of each motor to allow the appropriate sensor to reach the required (selected) position. The user has full control over the camera settings, so that optimal values can be set to get the "best" image. All information pertaining to the position, motor coordinates and sensor readings are stored in an Access database, giving the user the ability to review the collated data on screen or export it to an Excel spreadsheet. There is also the option to watch time-lapsed videos of the growth of individual plants.

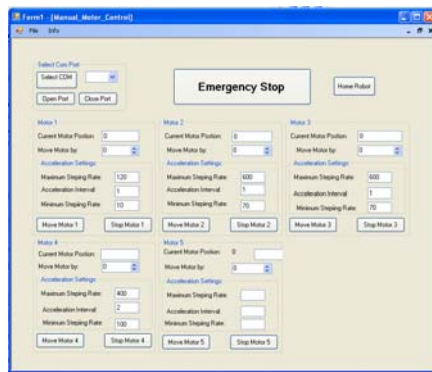


Fig. 6. GUI giving the user manual control of the arm.

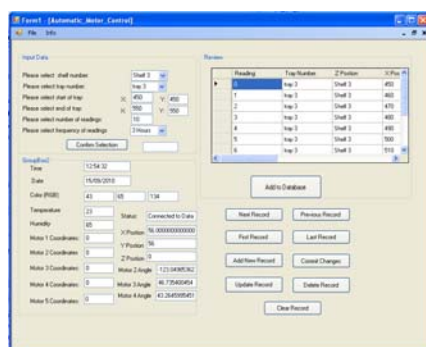


Fig. 7. GUI giving the user automatic control of the arm and reviewing collated data.

4.1 Motor Control

The Ocean Controls serial stepper motor controllers are programmed by sending ASCII commands from a serial port. The command contains the motor being used (@AA), a command required to be executed (CMND), followed by the parameters of the command (XXXX) and finally a carriage return byte (CR).

@AA CMND XXXXCR

It is possible to simultaneously communicate and control all 4 motors on a controller board by sending one command, @AA CMND WWWX XXXX YYYY ZZZZCR where @AA refers to the controller board being used, CMND the command to send to the controller, WWWX is parameter of motor 1, XXXX parameter of motor 2, YYYY is the parameter of motor 3, ZZZZ is the parameter of motor 4, followed by the carriage return byte (CR).

The information is sent from the serial port by setting up a byte array using the required protocol. Below is an example to move motor 1 a relative movement of 1000 steps (using a micro-step of 3600, and 2:1 gear ratio; this would move the motor 50 degree).

```
Motor_1(0) = Asc("@")
Motor_1(1) = Asc("0")
Motor_1(2) = Asc("1")
Motor_1(3) = Asc(" ")
Motor_1(4) = Asc("R")
Motor_1(5) = Asc("M")
Motor_1(6) = Asc("O")
Motor_1(7) = Asc("V")
Motor_1(8) = Asc(" ")
Motor_1(9) = Asc("1")
Motor_1(10) = Asc("0")
Motor_1(11) = Asc("0")
Motor_1(12) = Asc("0")
Motor_1(13) = &HD
```

All commands are sent to the controller in the same format, with the exception of the 'position' and 'status' command, which do not require parameters. When the motor controller receives the status command it replies with an integer number that, when converted is a 12-bit binary representation of the controller. The bits that make up the binary number show which of the motors are moving, in which direction and the status of the limit switch.

4.2 Camera Control

The camera is controlled using Sony's VISCA command protocol. The basic unit of VISCA communication is a packet of 3 to 16 bytes. The first byte of the packet is the header consisting of the sender and header addresses; the remainder of the packet holds the command (or message) to be sent to the camera (Figure 8). The final byte comprises a terminator signifying the end of the packet. It usually has the hexadecimal value FF (or 11111111 binary).

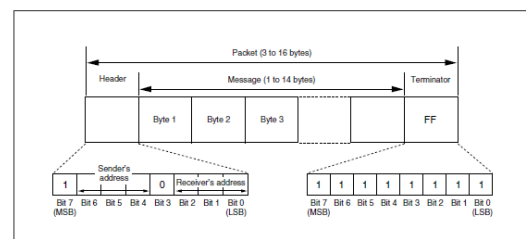


Fig. 8. Serial command data packet layout required to communicate with camera

Each of the commands required for the camera controls is sent via the serial port to the camera, in a byte format of the appropriate hexadecimal values. Each command has its own unique command, which is required to be sent to the camera each time. A timer function allows each command to be executed almost instantaneously by the camera.

4.3 Sensors

Parallax design and manufacture sensors allow a large range of applications. These sensors easily integrate into the Parallax Basic Stamp microcontroller, with most sensors coming with some source code. The Basic Stamp 2e microcontroller uses PBASIC language with a processor speed of 20 MHz. The microcontroller integrates to the PC using a serial input and serial output pin. However a Board of Education (#28850) was used that connects the microcontroller to the PC through a USB connection.

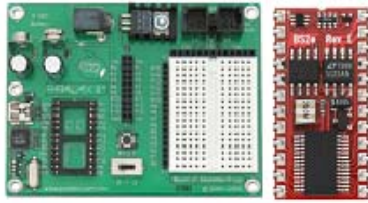


Fig. 9. Parallax Board of Education (left), BASIC Stamp 2e microcontroller (right) [18]

After connecting the sensors to the microcontroller it was obvious that the microcontroller was unable to perform some of the calculations quickly so more complex calculations are solved within the software on the PC.

5. TESTING AND INTEGRATION

Testing was done in parts using a modular approach. The sub-systems that are required to be programmed, the construction of the anthropomorphic robotic arm, control of the motor through the motor controller, implementation of all sensors and control of the camera were all tested separately. The testing setup and procedure are described below.

5.1 Anthropomorphic Robotic Arm

The robotic arm, designed in SolidWorks, was tested as parts were created but it was soon discovered that initial designs were too heavy. A quick solution was implemented utilizing finite element analysis (FEA) to ensure minimal material was used on specific components while still retaining sufficient strength.



Fig. 10. Original motor mount (left), redesigned motor mount with excess material removed (right)

5.2 Motor Controller

In the first instance a simple program was written in Visual Studio, allowing one stepper motor to be controlled by clicking a button. This action sent the corresponding command to the motor controller board. The next step was to control the motor by selecting the required steps from a numerical up/down box, which increased or decreased the relative movement position, allowing for simple movement based on the position byte. Once one stepper motor was successfully implemented, multiple motors could then be controlled in the same manner. Finally several algorithms were set up to allow the motors to move to a specific position calculated via inverse kinematics algorithms.

To ensure that the robotic arm was functioning correctly, it was set up in the working environment with a number of different-shaped objects replacing the plants. The robotic arm was then programmed to move to the same coordinates autonomously and capture images which, once reviewed, confirmed that the arm had returned to the same positions each time, with all the images showing the correct object located in the same position.

5.3 Control of Camera

The camera can be programmed and controlled using a TTL or RS232C level serial interface. To test the camera features it was directly wired to the PC using the RS232C interface via a USB-to-RS232 converter as shown in figure 11. The video output signal from the camera was fed to a frame grabber/digitizer that is interfaced with the PC via USB.

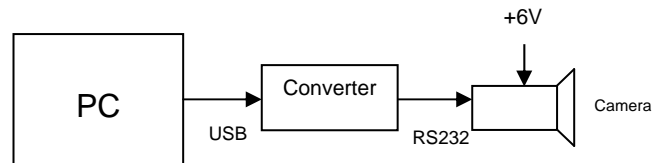


Fig. 11. Block diagram showing wired connection of PC to camera's RS232 inputs

Using a standard communication program (Terminal v1.9b), commands were sent to the camera to become familiarized with the VISCA commands and to test the various camera functions, especially the programming commands for controlling the zoom. Once successfully mastered, a more complex program was created.

5.4 Sensors

Parallax sensors come with source code which allows easy integration and testing. Once the sensors were implemented, custom software in Visual Studio was created to communicate with the microcontroller and sensors. Custom software allowed the proximity sensor to be integrated with the temperature sensor, which is necessary since the speed of sound in air is affected by the ambient temperature $C_{air} = 331.5 + (0.6 \times T_c)$ m/s. Results from Parallax show that over a temperature range of 0 to 70°, the error percentage is in the magnitude of 11-12%.

Testing the TCS3200 sensor involved comparing the RGB values obtained to those from a CM-700d Spectrophotometer (Konica Minolta Sensing Americas, Inc, NJ, USA), which gives colour in the XYZ colour space. RHS colour charts (Royal Horticulture Society, London, UK) were used for a range of green colours. To make the comparisons, a linear transformation matrix was required to convert the XYZ colour to an RGB colour space.

The XYZ to RGB transformation matrix equation (alternatively the inverse can be taken to solve for RGB) is:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1.436603368435593 & -0.118534007680723 & -0.226698746676605 \\ -0.681279626157456 & 1.618476401223029 & 0.026671792203898 \\ 0.042919641610377 & -0.102206221605993 & 0.974614036323184 \end{pmatrix} \times \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

In order to validate the results a calibration factor (gamma) was required to minimise any differences between ideal results (from the spectrophotometer) and actual results (from TCS3200).

$$R_N' = R_N^\gamma$$

Where: R_N' = Spectrophotometer (desired RGB, known)
 R_N = TCS3200 RGB (not calibrated, known)
 γ = Gamma (required calibration factor)

This required the data to be white calibrated, which involves offsetting the data to ensure the white reading has an RGB value of 255,255,255. The following equation is then solved:

$$\gamma_R = \frac{\log(R_N' / 255)}{\log(R_N / 255)}, \gamma_G = \frac{\log(G_N' / 255)}{\log(G_N / 255)}, \gamma_B = \frac{\log(B_N' / 255)}{\log(B_N / 255)}$$

Data is required to be normalised (i.e. put in a range from 0 to 1). From this the gamma values can be solved for the corresponding red, green and blue values.

Results show that over 16 random samples of green, from the RHS colour chart, the gamma was calculated to be ($\gamma_R = 0.50$, $\gamma_G = 0.38$, $\gamma_B = 0.59$). This led to an absolute error of red 5.67%, green 3.01% and blue 3.98%, showing the TCS3200 sensor to be accurate to within an average error of 4.22% across the three colours.

5.6 Image Analysis

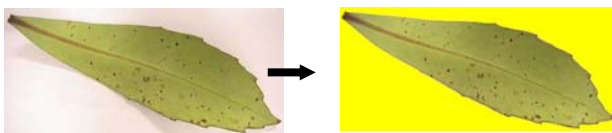


Fig. 12. Initial testing of the image analysis algorithm to determine quantity of pixels that make up the plant material content

Images captured by the camera were imported into VB.net, where algorithms were applied to determine the quantity of plant material present. The non-plant pixels in the image were counted and replaced with a contrasting colour pixel. Initial leaf testing, by making the background yellow, showed that the algorithm worked correctly (Figure 12).

6. DISCUSSIONS AND FUTURE DEVELOPMENT

The system is currently being tested in the working environment. A number of issues that arose during implementation and testing are being addressed. Future work involves completing the entire robotic system and fully integrating the sensors. On completion, investigations based on the continuous information obtained from the sensors and processed through algorithms will be used to optimise the growth of plants by manipulating liquid nutrient solutions and the immediate growing environment. Opportunities exist to set up the TCP/IP connections so that monitoring data can be accessed from anywhere in the world.

7. REFERENCES

- [1] Slaughter, D. C., Giles, D. K., & Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture*, 61(1), 63-78.
- [2] Kawollek, M., & Rath, T. (2008). Robotic harvest of cut flowers based on image processing by using *Gerbera jamesonii* as model plant. *Proceedings of the International Symposium on High Technology for Greenhouse System Management*, Vols 1 and 2(801), 557-563.
- [3] van Henten, E. J., Hemming, J., van Tuijl, B. A. J., Kornet, J. G., Meuleman, J., Bontsema, J., et al. (2002). An autonomous robot for harvesting cucumbers in greenhouses. [Article]. *Autonomous Robots*, 13(3), 241-258.
- [4] Garcia, G. J., Pomares, J., & Torres, F. (2009). Automatic robotic tasks in unstructured environments using an image path tracker. *Control Engineering Practice*, 17(5), 597-608.
- [5] Kelly, R., Carelli, R., Nasisi, O., Kuchen, B., & Reyes, F. (2000). Stable visual servoing of camera-in-hand robotic systems. *Transactions On Mechatronics*, 5(1), 39-48.
- [6] Reyes, JF; Chiang, LE., (2009). Location and classification of moving fruits in real time with a single colour camera, *Chilean Journal of Agricultural Research*, Vol. 69, 179-187.
- [7] Scarfe, A. J., Flemmer, R. C., Bakker, H. H., & Flemmer, C. L. (2009). Development of an autonomous kiwifruit picking robot. *Proceedings of the Fourth International Conference on Autonomous Robots and Agents*, 639-643.
- [8] Wang HonYong, Cao QiXin; Masateru, N., Bao JianYue, (1999) Image processing and robotic techniques in plug seedling production, *Transactions of the Chinese Society of Agricultural Machinery*, Vol. 30, 1999, 157-162.
- [9] Otte, C., Schwanke, J., & Jensch, P. (1996). Automatic micro-propagation of plants. *Optics in Agriculture, Forestry, and Biological Processing*, 2907, 80-87.
- [10] Sogaard, H. T., & Lund, I. (2007). Application accuracy of a machine vision-controlled robotic micro-dosing system. *Biosystems Engineering*, 96(3), 315-322.
- [11] Yadav, S. P., Ibaraki, Y., & Gupta, S. D. (2010). Estimation of the chlorophyll content of micropropagated potato plants using RGB based image analysis. *Plant Cell Tissue and Organ Culture*, 100(2), 183-188.
- [12] Abdullah, M. Z., Aziz', S. A., & Mohamed, A. M. D. (2000). Quality inspection of bakery products using a color-based machine vision system. *Journal of Food Quality*, 23.
- [13] Raja, A. S., & Sankaranarayanan, K. (2006). Use of RGB Color Sensor in Colorimeter for better clinical measurements of blood Glucose. *BIME Journal* 6(1), 23-28.
- [14] Miranda, C., Girard, T., & Lauri, P. E. (2007). Random sample estimates of tree mean for fruit size and colour in apple. *Scientia Horticulturae*, 112, 33-41.
- [15] Kang, S. P., & Sabarez, H. T. (2009). Simple colour image segmentation of bicolour food products for quality measurement. *Journal of Food Engineering*, 94, 21-25.
- [16] Yam, K. L., & Papadakis, S. E. (2004). A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering*, 61, 137-142.
- [17] Ocean Controls. Retrieved 12/04/2010, from www.oceancontrols.com.au
- [18] Parallax Home. Retrieved 05/07/2010, from www.Parallax.com

Simulation and Optimisation of a 2-DOF Parallel Planar Manipulator

B.P. Haughey and D.A. Carnegie

Victoria University of Wellington
School of Engineering and Computer Science
P.O. Box 600, Wellington, New Zealand
benhaughey@gmail.com

Abstract:

A machine building company wishes to develop bespoke robotic manipulators for pick-and-place applications. To do this a simulation model of a 2-DOF parallel planar manipulator has been developed in Matlab's SimMechanics[®]. This has allowed the detailed analysis of the manipulator's dynamics. A trajectory planner is detailed that utilises cubic splines to form a smooth path in joint space. By using a hill climbing optimisation algorithm, a system is developed that allows a near optimal mechanical configuration to be found for any given pick-and-place task with a pre-defined trajectory.

Keywords: Simulation, Optimisation, Trajectory Planning, Manipulator, Hill-climber, SimMechanics[®]

1 INTRODUCTION

The drive to improve factory efficiencies by increasing throughput has led to a large amount of development in robotic product manipulators. A New Zealand based company, RML Engineering Ltd., is currently investigating the development of a parallel robot with two degrees of freedom (DOF). Specifically they would like to have the tools and knowledge to create be-spoke manipulators that are optimised for a given a task. While there exist commercially available 2-DOF manipulators, these are highly priced and are not optimised for any particular application, hence they do not offer the high-end performance RML Engineering Ltd. seeks.

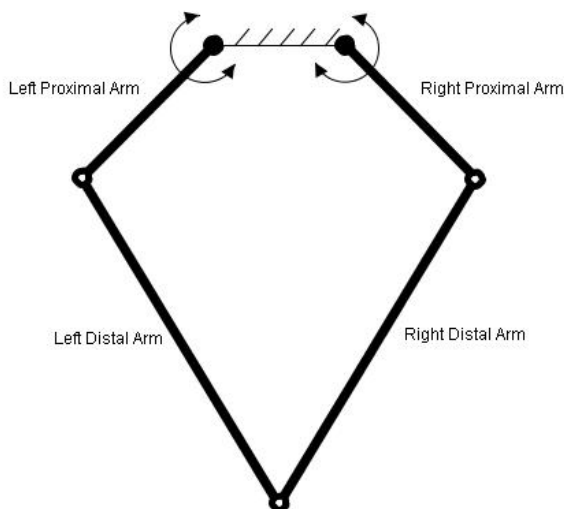


Figure 1. Simple Configuration of a 2-DOF Planar Parallel Manipulator

The manipulators are often referred to as pick-and-place robots due to the task that they perform. A popular mechanical configuration for these pick-and-place robots is of a parallel architecture. Parallel manipulators are defined as being of a closed loop parallel structure where the end-effector is supported by at least two chains controlled by separate actuators [1]. A simplified diagram of the specific configuration being examined is presented in Figure 1. These parallel manipulators offer several advantages over the alternative serial configurations, namely greater rigidity [2], higher payload/weight ratio and reduced inertia [3]. They also feature higher precision, due to positioning errors being averaged rather than compounded as they are in serial manipulators [4]. However, the closed loop structure requires more complex mathematical analysis and additionally possesses the disadvantages of having more singularities and a smaller workspace [5].

1.1 Objective

The objective is to develop software tools that will aid in the design of application specific 2-DOF parallel planar manipulators. This firstly involves simulating the mechanical structure to gain a greater understanding of the torques, forces and dynamics of the system. The second stage involves optimising the mechanical dimensions to obtain a manipulator that will perform a given task in a near-minimum period of time.

1.2 Previous Work

Several researchers have analysed the 2-DOF parallel planar manipulator. Cervantes-Sánchez et al. [6] presents a simplified method for obtaining the workspace of the manipulator without finding the kinematic equations. Deng et al. [7] uses SimMechanics® to create a simplified simulation model. Huang et al. [8] introduces the idea of a conditioning index to find a well-shaped workspace. While these methods are useful in their own right, this project seeks to not only find the workspace but also the optimal dimensions for the manipulator. Huang et al. [9] also presents a method for finding a time minimum trajectory for the manipulator utilising the system's Jacobian matrix. This method however, required several parameters that need to be estimated based on the system's dimensions. Stan et al. [10] uses a genetic algorithm and a set of conditioning indexes to optimise the manipulator.

2 SIMULATION

Matlab's® Simulink® package was used to build a model of the system and simulate a pick-and-place manoeuvre. The model uses the building blocks of *bodies* and *joints*,

with referential constraints to form a 2-D layout of the manipulator as can be seen in Figure 2. *Revolute1* and *Revolute5* represent the motor actuated joints and are driven by *Joint Actuator* blocks with inputs of angular position, velocity and acceleration. These *joints* are then connected to the upper arms labelled *Upper Arm A* and *Upper Arm B*. The upper arms are then connected to the lower arms (*Lower Arm A* and *Lower Arm B*) by passive revolute joints (*Revolute2* and *Revolute4*). The lower arms are joined together by the *Revolute3* joint, thus completing the closed loop system. Additional components are added to form a gripper head and a stabilising arm to ensure the gripper head remains parallel to the worksurface. Several sensors are included in the model, namely *Torque Sensors* on the motors and a *Body Sensor* on the gripper to measure the position, velocity and acceleration in Cartesian space.

The model is acted on by external inputs consisting of time varying angular position, velocity and acceleration commands to the two actuated joints (motors) and a constant gravity vector. SimMechanics'® physics engine

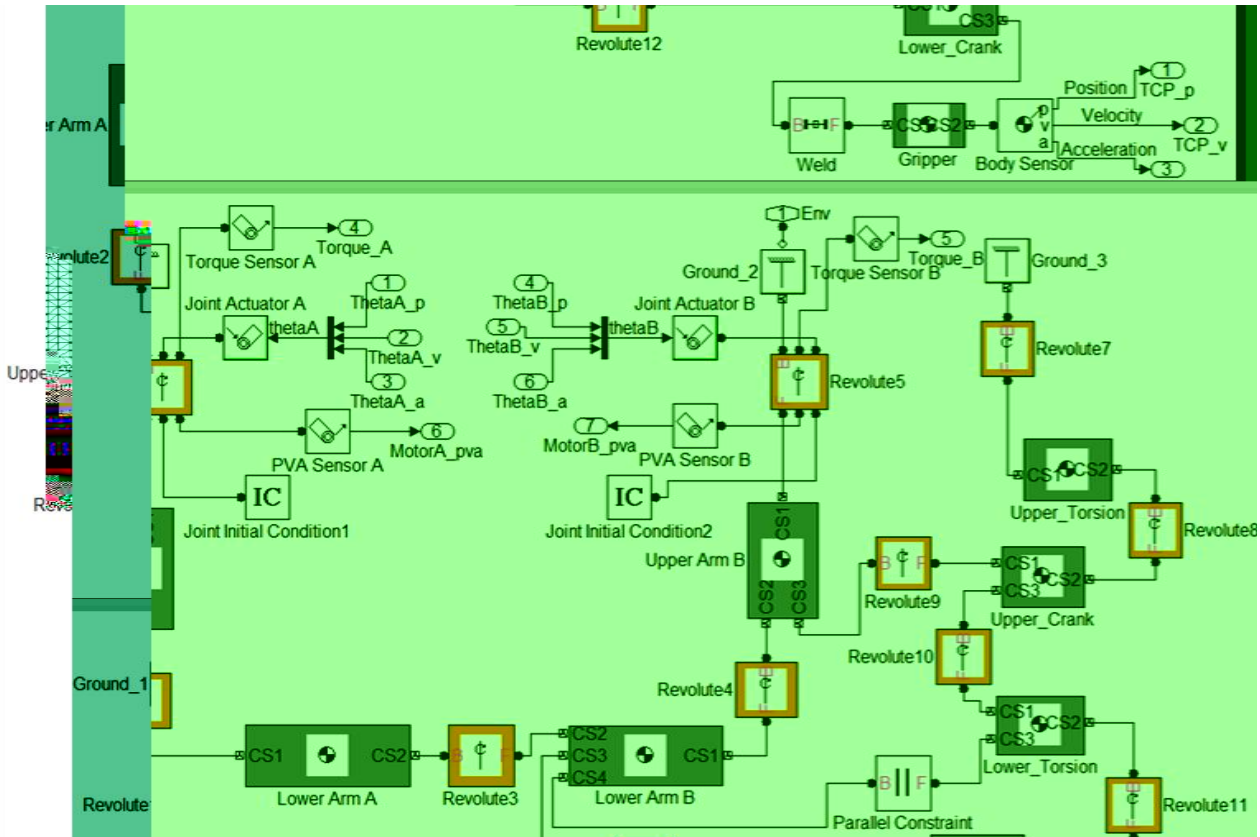


Figure 2. Construction of the manipulator model in SimMechanics®

takes these input commands and, while maintaining the body and joint constraints of the model, determines the resulting mechanical changes to the system. A number of velocities, accelerations, forces and torques are measured throughout the simulation and are provided as feedback to indicate how near the design limitations the model came.

Figure 3 shows a screen capture of the simulation running. Note the addition of a stabilising arm down the right hand chain. This non-actuated chain ensures the end-effector/gripper is always parallel to the base of the robot.

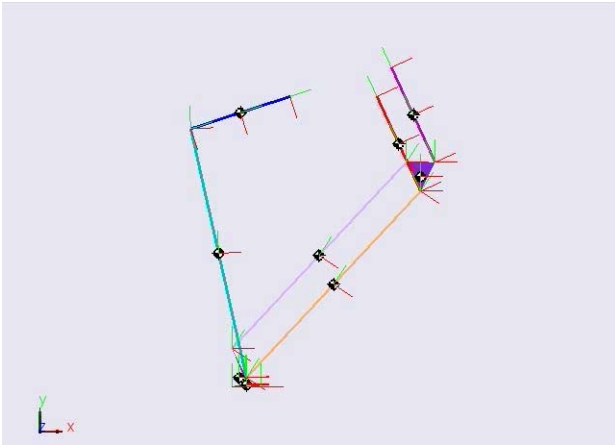


Figure 3. Visulation of the manipulator in SimMechanics®

3 Trajectory Planning

In order to provide meaningful input commands to the motors in the simulation a trajectory planning process must be done. In this project, trajectory planning is done ‘off-line’ prior to running the simulation. Several objectives must be met; namely the trajectory generated must follow the desired path to within a certain degree of accuracy and it must do this smoothly so as not to cause large jerking motions for either the motors or the end-effector/gripper. As a time-minimum cycle is of importance, the trajectory planner must maximise the performance of the motors to traverse the path as fast as possible.

The system allows a path to be defined by specifying a set of *move* commands consisting of:

- Cartesian coordinates for the end-effector to move through (*target*).
- The method of traversing to these *targets*. This can either be a *linear* move (moving in a direct line) or a *joint* move (no constraint on how the manipulator reaches the *targets*).

- A *zone* around the *target* indicating how close to the actual coordinate the end-effector must be before it can move towards the next *target*.
- The maximum velocity permitted of the end-effector during the move.

Figure 4 gives a visualisation of the above definitions. The *targets* are represented by the four crosses, with *zones* around them indicating a region through which the manipulator must pass through before moving to the next *target*. The vertical moves have been specified as *linear* moves indicating a direct linear tracking of the path between two *targets*. The horizontal *join* move is a weaker constraint allowing the trajectory to be determined in a way most efficient to the joints.

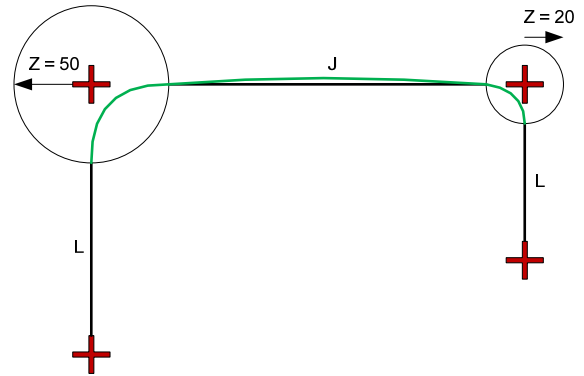


Figure 4. A path defined using the convention

Additional constraints are also specified such as available motor torques, speeds and maximum/minimum permissible angles between arms etc.

The method to form the trajectory is outlined in the steps below:

1. Inverse kinematics are used to take the Cartesian coordinates of the *targets* and convert them into joint-space (motor angle positions).
2. *Knots* are then calculated based on the *targets*, *zones* and whether the *moves* are *linear* or *joint*. *Knots* are reference points defined in joint-space which are used as the interpolation points in the spline fitting process. Note that additional *knots* are added to *linear* moves to ensure a straight path is followed.
3. Time dependent cubic splines are then formulated between the *knots*. The cubic polynomial coefficients for all path segments are solved simultaneously. Additional constraints are added to the polynomials to ensure that each adjoining path segment maintains continuous angular position and velocity of the motors.

4. Iteratively alter time between *knots* until motor velocity, acceleration, jerk, estimated motor torque and end-effector velocity are within the permissible bounds. Also check at every iteration that each cubic path segment is near the limit of at least one of the constraints to ensure the system is being maximised to its full potential.

Once the above process has completed, motor position, velocity and acceleration values are then taken at frequent time intervals and parsed into the SimMechanics® simulation. Figure 5 shows the components used in the path planning as well as the actual path traversed by the simulated manipulator. The '+' represent the *targets* with the circles surrounding them being the *zones*. The small circles are the interpolated *knots*. Note that the vertical movements have additional *knots* as they are *linear moves*. The 'x' represent the position of the end-effector at which the simulation was evaluated.

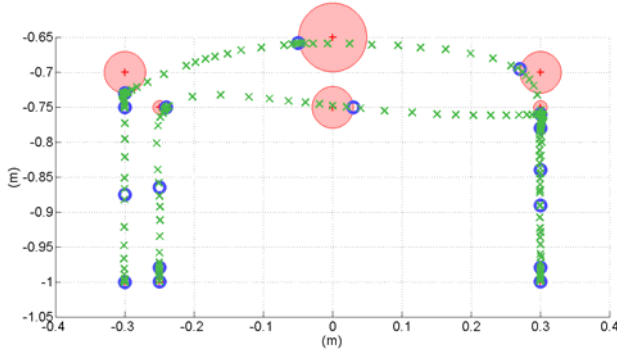


Figure 5. Plot of the targets, zones, knots and actual trajectory followed

3.1 Estimating the Motor Torque

As part of the trajectory planning process it is a requirement that the motor torque is able to be calculated. Fitting a cubic spline over a long time period results in a slow cycle time far below the potential of the motors, however fitting the spline over too shorter time period will require the motors to output a torque in excess of their capabilities. The highly coupled and non-linear nature of parallel manipulators results in complex systems of equations to calculate the torque exactly [11]. To avoid these complexities a simplified formulation is used which takes into account the mass and lengths of the arms as well as the load's mass, to estimate the inertia of each chain (left or right). The inertia is then multiplied by the angular acceleration of the motors to find the torque. The torque due to gravity is also added to give a total torque for each chain.

$$I = M_{prox} * \left(\frac{L_{prox}}{2} \right)^2 + \left(M_{dist} + \frac{M_{gripper}}{2} \right) * L_{prox}^2 \quad (1)$$

$$T_{gravity} = M_{prox} * \left(\frac{L_{prox}}{2} \right) * \sin(\theta) * g + \left(M_{dist} + \frac{M_{gripper}}{2} \right) * L_{prox} * \sin(\theta) * g \quad (2)$$

$$T_{total} = I * \alpha + T_{gravity} \quad (3)$$

where M_{prox} , L_{prox} are the mass and length of the proximal arm. M_{dist} is the mass of the distal arm. $M_{gripper}$ is the mass of the gripper/end-effector. θ is the angle of rotation of the proximal arm from the vertical. α is the angular acceleration of the proximal arm. g is the acceleration due to gravity defined as 9.8 ms^{-2} .

By using a simplified formulation, an approximate torque estimate is obtained for the path. This provides us with results as seen in Figure 6. The solid line represents the torque as calculated by the SimMechanics® physics engine. This is closely mapped by the estimated torque calculated using equation (3) as plotted with the dotted line. It can be clearly seen that the torque estimation does give an indication of the torque expected from the motors but could still be further improved since the extreme peaks are not reached.

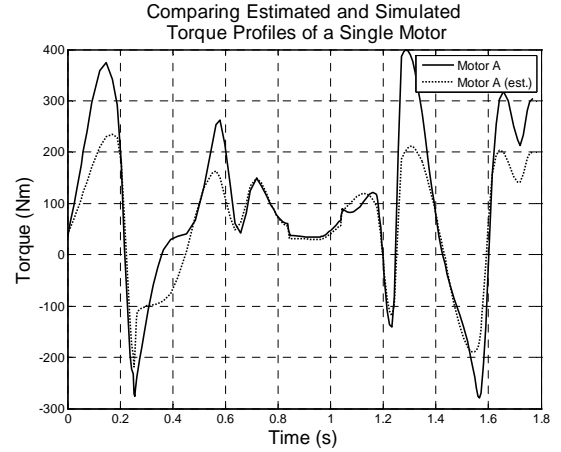


Figure 6. Comparison of estimated and simulated torque profiles for a single motor over a given trajectory

4 OPTIMISATION

While the simulation combined with the trajectory planner provides insight into the expected performance of a given manipulator, of greater value is the ability to maximise that performance for a given task. As the lengths of arms and the mounting positions of the motors are changed, the cycle-time of the system also changes. For industrial production lines even the smallest improvement in a robotic manipulators cycle-time can result in large improvements in efficiency. It is therefore of great value to design a manipulator which is optimised for a given task with a given trajectory.

By altering the dimensions of the four main components in an optimising routine, an optimal or near-optimal solution can be found. The dimensions to be altered are:

- Proximal (upper) arm length.
- Distal (lower) arm length.
- Separation distance of the motors.
- Height of the manipulator above the workspace.

In addition to these dimensions, the optimiser will be able to select motors with various performance characteristics from a database of available motors.

3.1 Search Space

Before determining what optimising algorithm to use it is advantageous to know what the solution space looks like. This then gives an indication as to the type of optimising method best suited to the application.

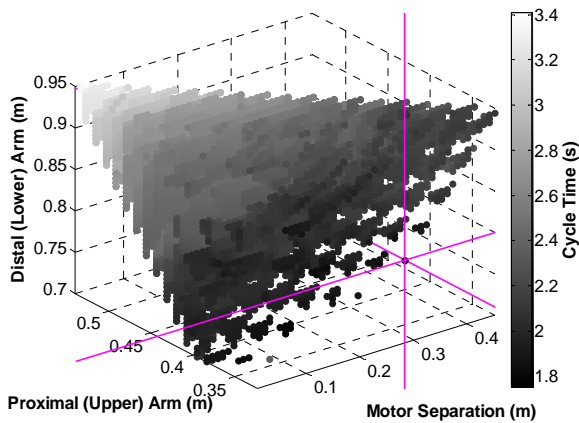


Figure 7 A four dimensional plot of three variable dimensions (arm lengths and motor separation) and the associated cycle-time plotted in gray-scale.

Figure 7 presents the solution space for a given trajectory and set of constraints. This was formed by coarsely iterating through all possible configurations and simulating their cycle time results. For ease of viewing only three of the four dimensions are plotted, with the workspace height being omitted. The darker points indicate a faster cycle-time whereas the lighter points indicate a slow cycle-time. It can be seen that the solution space has a regular shape with a generally smooth transition from fast to slow cycle-times, i.e. there is no obvious or significant local optima.

Cross hairs have been plotted to indicate the global optima. It can be seen that the global optima lies on the boundary of the solution space (near infeasible solutions). It would therefore be of great importance to consider the implications of using this optimal solution, as a

slight change in trajectory requirements may result in a manipulator configuration that can no longer fulfil its task. It is then imperative that the optimal solution not be used blindly but rather considered in view of the optimisations solution space and its reachable workspace.

3.2 Hill Climbing Algorithm

An exhaustive search of the solution space is time consuming and inefficient. Therefore an optimising algorithm must be used to resolve the optimal solution(s) in a more efficient manner. While there exists many advanced optimising algorithms, the regularity of the solution space shown in Figure 7 lends itself well to the simple Hill Climbing Algorithm with a random restart functionality added to it. This process is outlined in the steps below:

1. Randomly select a suitable set of dimensions and a motor.
2. Evaluate the cycle-time performance of these dimensions and motor within the simulation. Store results in a database.
3. Select 16 new sets of dimensions that are varied only slightly from the previous dimensions. This equates to a higher and lower value for each dimension and the combination of varying all four dimensions.
4. Evaluate each of the new sets of dimensions within the simulation.
5. Select the best (lowest cycle-time) set of dimensions from the previous 16 and repeat steps 2 & 4 with it.
6. Continue until no further improvement is made to the minimum cycle-time. At which point repeat steps 1-6.
7. Repeat this for a set number of iterations. The dimensions used to achieve the overall minimum cycle-time is considered the near-optimal solution for the given task.

A summary of the system can be seen in Figure 8. This shows how the data flows between Matlab®, SimMechanics® and the MySQL database.

Figure 9 presents the results of the hill climbing algorithm. It can be seen that an optimal solution has been found where the crosshairs intersect.

Whilst this Hill Climbing optimisation algorithm is providing useful results, on-going work involves the consideration of alternative algorithms, as potentially the Hill-Climber is susceptible to the local maxima problem which may exist in other solution spaces.

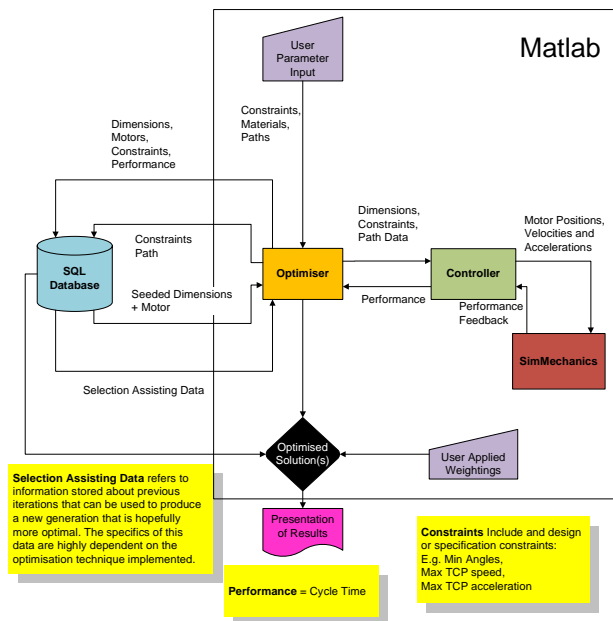


Figure 8. High level overview of the system

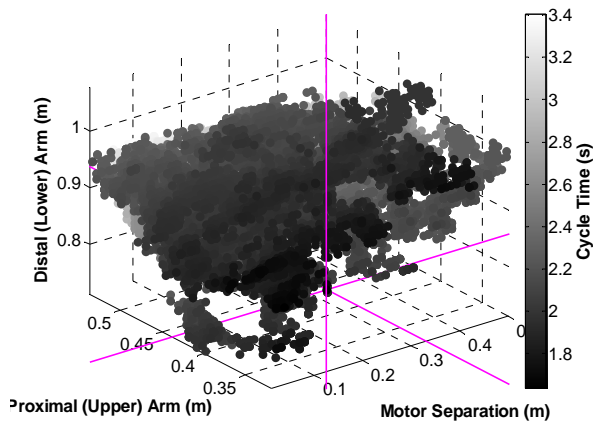


Figure 9. Results of Hill-Climbing Algorithm

4 CONCLUSIONS

The system developed allows the simulation of a 2-DOF planar parallel manipulator in SimMechanics®. This provides detailed analysis of joint torques and forces which assist in the designing of such a manipulator. A near-time-minimum trajectory planning methodology has been developed that utilises cubic splines fitted to knots defined in joint space. This allows for smooth trajectories with continuous angular velocities for the motors throughout the cycle. A hill climbing algorithm has been applied to permit finding a near optimal mechanical configuration for a given task and trajectory.

4.1 On-going Work

Further improvements to this system include developing a more accurate formula for estimating the motor torque as this would improve the validity of the simulation as it applies in reality. Also, as previously mentioned, alternative optimising algorithms are being explored and their performances compared.

5 REFERENCES

- [1] J.P. Merlot. *Parallel Robots*. Dordrecht, The Netherlands : Kluwer Academic Publishers, 2000. ISBN 0-7923-6308-6
- [2] B. Dasguptaa, T.S. Mruthyunjaya. *The Stewart platform manipulator: a review*. Pergamon Press, 1998.
- [3] C.M. Gosselin, S. Lemieux, J.P. Merlet. "A new architecture of planar three-degree-of-freedom parallel". *IEEE*, 1996.
- [4] E. Pernette, S. Henein, I. Magnani, R. Clavel. "Design of parallel robots in microrobotics". Lausanne, Switzerland : *Cambridge University Press*, 1997, Vol. 15.
- [5] P. Vischer, R. Clavel. "A Novel 3-DoF Parallel Wrist Mechanism". *The International Journal of Robotics Research*, 2000, Vol. 19.
- [6] J.J. Cervantes-Sánchez, J.G. Rendón-Sánchez. "A simplified approach for obtaining the workspace of a class of 2-dof planar parallel manipulators". *Pergamon Press* 1999.
- [7] W. Deng, J. Lee, H. Lee. "Kinematic Simulation and Control of a New 2 DOF Parallel Mechanism Based on Matlab/SimMechanics". *ISECS International Colloquium on Computing, Communication, Control, and Management*, 2009.
- [8] T. Huang, M. Li, D.G. Chetwynd, D.J. Whitehouse. "Optimal Kinematic Design of 2-DOF Parallel Manipulators With Well-Shaped Workspace Bounded by a Specified Conditioning Index". *IEEE Transactions on Robotics and Automation*, 2004, Vol. 20, No. 3.
- [9] T. Huang, P.F. Wang, J.P. Mei, X.M. Zhao, D.G. Chetwynd. "Time Minimum Trajectory Planning of a 2-DOF Translational Parallel Robot for Pick-and-place Operations". *Annals of the CIRP*, 2007, Vol. 56, No. 1.
- [10] S. Stan, V. Mătieș, R. Bălan "Optimization of a 2 DOF Micro Parallel Robot Using Genetic Algorithms". *Frontiers in Evolutionary Robotics*, H. Iba, 2008, pp 456-490.
- [11] Z. Yang, J. Wu, J. Mei, J. Gao and T. Huang, "Mechatronic model based computed torque control of a parallel manipulator" in *International Journal of Advanced Robotic Systems*, Vol. 5, No. 1 (2008) pp. 123-128

Design of a True 3D Digital Display

B. Vercoelen, C. Hollitt, R. Dykstra

*School of Engineering and Computer Science
PO Box 600, Wellington 6140, NEW ZEALAND.
Email: christopher.hollitt@vuw.ac.nz
Email: robin.dykstra@vuw.ac.nz

Abstract:

A prototype digital display has been designed and built allowing the user to view data in three dimensions. The display uses a 16 by 16 by 8 matrix of tri-colour (Red/Orange/Green) LEDs. Using a simple multiplexing technique, each LED in the display can be controlled individually. Shift registers that can be latched on allow for accurate images and animations to be played on the display. The display circuitry is driven by a Field-Programmable Gate Array allowing the display to be programmed directly using gate logic, or used to interface with a computer allowing for more complex, higher level programming. The display has many possible uses including 3D animations, 3D games and data visualisation.

Keywords:

LED, 3D, FPGA, Display

1 INTRODUCTION

Present three dimensional (3D) display technologies are excellent at adding depth to a video, giving the perception of three dimensions. We can see the results in recent blockbuster films such as *Avatar*. As the video data is still only presented on a two-dimensional (2D) plane, it is not true 3D. Depth is added by 3D glasses and not the display itself. This means all viewers still see the same 3D image no matter what angle they view on, which is undesirable as three dimensional data should be viewable in three dimensions as if it were a real object.

The goal of this project was to design and build a display system to present three dimensional data in *true* three dimensional form without the aid of devices such as 3D glasses. To avoid the false 3D effect, the display must be physically three dimensional. A purely digital solution was needed, using only off-the-shelf components. This meant there could be no use of holograms, mirrors or any moving parts. As an electronic method for displaying the 3D data is required, the obvious choice was to use Light Emitting Diodes (LED). There were many challenges with designing such a complex circuit due to the large number of LEDs required.

A true 3D LED display, although not having the high resolution of consumer 3D televisions, still has many uses. 3D object representation would allow designers, architects or product developers to visualise their 3D designs without needing to build a physical prototype or rely on 3D design

software. Not only can still 3D objects be shown, but they can also be animated. There is a vast amount of uses for 3D animations, from simple 3D video to interactive simulations allowing the display to respond to user input or other input, such as sound. This would allow for 3D audio visualisations or 3D games, such as *Snake*.

The 3D display has been built and is controllable via a Field-Programmable Gate Array (FPGA) that controls the display using low level gate logic. In order to program in a higher level language, the FPGA can be connected to another device such as a microcontroller or a computer using the serial port on the FPGA development board. The active display is shown in Figure 1.

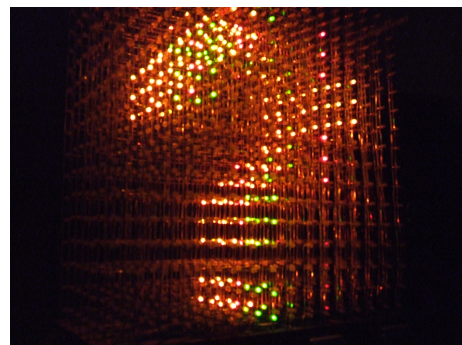


Figure 1: The constructed LED display

2 BACKGROUND

LED cubes are three dimensional matrices of LEDs allowing 3D image data to be displayed in the form of still images or moving animations. LED cubes range in both size and colour capability where each LED represents a 3D pixel (also known as a *voxel*, a volumetric pixel). Some produce only on-off effects, while others produce a full colour range using Red/Green/Blue (RGB) LEDs. Most LED cubes are built as hobbyist projects reaching sizes of 4x4x4. They can be easily controlled by a simple microcontroller, making them popular with electronic hobbyists. *Seekway*, a company in China, produces large LED cubes for sale and provides software to control them from a computer [1]. It is difficult to find detailed information on these existing LED cubes, as plans for these large designs are unavailable due to being proprietary and commercially sensitive. Smaller cube designs are unsuitable as they do not have the ability to scale to the required size. This is because the designs are based on individual column control which is impractical for larger displays as a large number of output pins would be needed.

3 DESIGN

Designing the LED display required considerable design preparation before construction could begin. Both electronic and structural design decisions were important to the design of the display. As the LEDs form a 3D matrix, the structure and circuit of the cube needed to allow for this type of construction. This affects circuit design as limitations are placed on where the circuit can be held. No circuitry can be held in the lattice of LEDs as it would obscure the visibility of the layers. Wiring and circuitry must therefore be hidden outside the matrix.

3.1 Specifications

The goal of this project was to design and build a suitable display in three dimensions, allowing for 3D images and video to be represented digitally. Measurable requirements are then as follows:

- **Size:** The display needs to be large enough to display a recognisable 3D image. The minimum size was determined by the smallest size still big enough to view a simple 2D image. From this test, it was decided 8x8x8 would be the smallest reasonable display size.
- **Colour:** The aim was to show 3D animations and to do that optimally, changes in colour would be needed. Therefore the requirement was to have a number of possible colours as opposed to just one.
- **Refresh Rate:** The display requires a refresh rate greater than 30Hz to be able to show smooth animations without a noticeable update.

3.2 Display Parameters

Before considering how the cube was to be designed and constructed, the parameters of the display needed to be decided. This would influence how the display would be designed. The first of these choices was size. The size would require enough pixels to display a recognisable image, without being so big that it exceeds both the budget and construction feasibility. The chosen size had a width, depth and height of 16 LEDs, giving a total of 4096 LEDs in the cube where one LED represents one pixel. This was determined primarily by the cost of the LEDs to keep within budget. This was also considered the largest that could be feasibly assembled in the time given as a lot of manual soldering was required.

There is a wide range of LEDs available at varying costs. The three main types of LEDs that were considered:

- **single-colour LEDs** are available in a full range of colours. This would result in a limitation to a single colour and would require multiple coloured LEDs to meet the specifications;
- **bi-colour/tri-colour LEDs**¹ provide typically Red and Green and in the case of tri-colour, orange (a combination of red and green); and
- **full-colour RGB LEDs** allow independent control of the three primary light colours: red, green and blue.

The choice of LED type was primarily based on cost. As the aim was to purchase upwards of 4000, an LED was needed that would be cost effective and yet still provide a decent colour range (i.e. not a single colour). Ideally RGB LEDs would be used, however their cost was too high due to the high cost of blue LEDs. Instead tri-colour LEDs were chosen, providing two of the three primary colours: red and green.

All of the acquired quotes from New Zealand suppliers were more than the budgeted \$200. This meant that further investigation was required to find a cheaper supplier. *Satistronics*, a Chinese manufacturer, was able to provide tri-colour LEDs for 3.2 US cents each. This was much cheaper than any quote found from a New Zealand supplier and this quote was accepted for 4500 tri-colour LEDs.

3.3 Display Construction

Physical construction of the display was a vital concern as the lattice of LEDs must be able to hold its own weight while remaining transparent to the LEDs behind. The first decision made was to make each layer modular, so that

¹Bi-colour LEDs have two diodes in 'inverse parallel' allowing only one to be lit at a time. Tri-colour LEDs have separate pins for the anodes allowing each one to be lit separately or both together to give a third colour.[2]

layers could be added or removed at any time without disruption to the system, either structurally or electronically. This would be useful for both repair of the layers and repair of the circuit as well as for portability.

As the structure was designed on a layer-by-layer basis, the structure of the LEDs had to be formed in a way that maximises transparency between the layers. The chosen method used the legs of the LEDs to form the structure of the layer, removing the need for any additional supporting materials. This design also functioned well electrically, as the anode pairs could be connected in a column and the common cathodes connected in a row. This allows the LEDs to be multiplexed, simplifying the LED control. This concept is shown in the diagram in Figure 2 and is described in more detail in section 3.4.

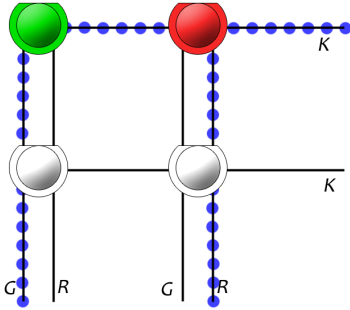


Figure 2: LED Structure, blue dots indicate current flow

3.4 Circuit Design

Controlling a large number of LEDs requires complex circuitry. This is because there are not enough output pins on any embedded controller to control 4096 LEDs individually, therefore data was needed to be modulated to control the display using minimal output pins. To control the large number of LEDs, it was decided that each layer would be controlled independently using a single output pin. This allows the system to be modular and layers to be added at a later time, as it requires only one additional output pin from the embedded controller. At layer level, a method for modulating 256 LEDs in a way that can be easily controlled was an important requirement. To do this, a method of reading and controlling serial data was required. This was achieved using a combination of shift registers and multiplexing.

As each layer was 16 LEDs high and wide, multiplexing over 16 rows would allow each column to be controlled simply. Each row will be on for one 16th of a refresh period, allowing rows to be controlled separately. Using shift registers to control both the rows and column multiplexing meant that a single digital data line could be used to control all 256 LEDs on one layer.

As the LEDs require a high current due to them being on for only one 16th of a time period, the shift registers would not have enough current to directly drive the LEDs. Therefore transistors were needed to drive them. These transistors provided a constant current source for the LEDs. As each LED can be on for a maximum of 1/16 of a refresh cycle, the current source required is 16 multiplied by the maximum LED current, as LED intensity is proportional to average current [3]. For this project, a source of 500mA is required for each column as this allows the LED to be at full brightness when on for only 1/16 of a refresh cycle. With such a high current, it is important power loss is minimised. The LEDs themselves will be dissipating a maximum of 32 watts per layer. A naive design using a 5V supply would dissipate a further 48 watts per layer as resistive heat loss. Power supply and circuit design were therefore vital to reduce the resistive power loss. A simplified circuit diagram is shown in Figure 3.

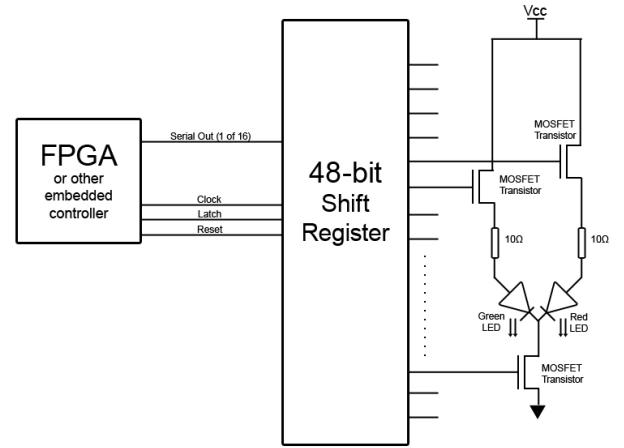


Figure 3: Simplified circuit diagram

To ensure the data rate is fast enough to control the display in this way, the minimum clock speed needed to be found to ensure the embedded controller could provide the data in sufficient time. The calculation for the minimum required transfer rate for a 50Hz display refresh, as shown in (1), needs 48 clock cycles to load the shift registers 16 times for multiplexing each row.

$$50Hz \times 48 \times 16 = 38.4kHz \quad (1)$$

The embedded controller must therefore allow for an output clock speed of at least 38.4 kHz. In practice, the clock frequency used was much faster than this, allowing for Pulse Width Modulation (PWM) of the LEDs to control the intensity. PWM works by reducing the width of the LED on-time, that in turn reduces the apparent brightness of the LED. This is required as the red diode is much brighter than the green diode and when both are on to-

gether, the red overpowers the green. Using PWM to dim the red will produce an accurate orange.

3.5 Choosing an Embedded Controller

The embedded controller is not only required to process all the 3D data but also to control a very large number of LEDs. The controller is required to send the data serially to the display with a clock speed of at least 40 kHz. Three obvious choices for the embedded controller were available. The first option considered was a micro-controller. This would allow for programming in the C programming language as well as the ability to interface with a variety of peripherals. Another alternative was using a Field-Programmable Gate Array (FPGA). An FPGA has many advantages over a micro-controller. It is at least an order of magnitude faster and has the ability to process more complex data due to its use of binary gate logic [4], making it better for 3D video. However its drawback is it must be programmed in a low level language such as Very-High-Speed Hardware Description Language (VHDL) rather than C. The third alternative was a PC, as a serial port could potentially control the cube directly, but would require additional RS232 connection circuitry.

The choice was made to use an FPGA, not only for its speed advantages but also for the ability to create functional logic blocks. This allowed many areas of the controller to function independently, without requiring complex threading or interrupts on the micro-controller. FPGAs can support large internal shift registers, which is important as thousands of bits need to be shifted out at a rapid rate. The chosen FPGA was the Spartan-3 Starter Kit Board from Xilinx as these are abundantly available at Victoria University's Department of Engineering and Computer Science.

3.6 Design Testing

LEDs are conventionally driven with 20-30mA, but would be very dim if driven for one 16th of a refresh period at this current. The LEDs needed to be driven with pulses of much higher current to achieve adequate display intensity. The first test was to ensure the LEDs could produce adequate brightness when only on for 1/16 of a period. The aim was to push the LEDs to 500mA forward current without damaging the LEDs. The LEDs from Satistronics claim a maximum continuous current of 30mA and a maximum peak current of 100mA, therefore 500mA far exceeds these ratings. As the LED would only be on for a fraction of a period, a test was required to ensure the LEDs were able to withstand the high current.

The test was a success, proving that the maximum ratings could be exceeded for one 16th of a 20ms period (50Hz), which will allow the LED display to be multiplexed over 16 rows with an adequate brightness.

3.7 PCB Design

Designing the Printing Circuit Board (PCB) was a complex challenge. Excluding the LEDs, there were still 768 transistors, 512 resistors, 96 shift registers and 48 D Connectors all to be placed and routed on the board. Using the PCB milling machine at Victoria University imposed a limitation of an A4 (210mm x 297mm) board size. This size was too small for the entirety of the required circuitry. The decision was made to use two circuit boards, each controlling eight layers of the cube. The two halves would then be placed side-by-side to produce the desired cube. Having two layers of copper on the PCB allowed components to be placed on either side of the board and gave extra room to fit all the components.

As the circuit for each layer was kept separate, for reasons explained in section 3.4, the entire circuit had to fit within a narrow 20mm strip (due to the layers being 20mm apart). This proved very complex, as most of the components were through-hole, requiring two sides and many signals needing to travel the narrow circuit. As each layer of the cube was identical, the problem was simplified into a single layer circuit on a 20mm strip and the circuit was then repeated eight times. This was far easier said than done, and a lot of time was spent juggling components and tracks to get them to fit into the narrow strip.

4 IMPLEMENTATION

4.1 Layer Assembly

Soldering 256 LEDs into a tidy matrix was predicted to be a long and painstaking process. The best solution considered was to use a custom made LED spacer. LEDs were able to be placed at even intervals into the custom grid and be soldered into a tidy lattice. The legs could then be bent evenly using a custom made tool. The spacer grid was designed in *SolidWorks* and was sent to a professional laser cutter to be made. Using the custom spacer ensured that the LED spacing was accurate. This was vital for the LEDs to appear in an accurate lattice and for the 3D image to be visualised accurately.

After the LED lattices were made, a frame was made to hold the lattices together. As each lattice was to be individually removable, a narrow frame was required to hold a lattice in place and connect with the D Connectors on the PCB. This frame was designed and laser cut from plywood. Once the LED lattice and D Connectors were held in place, wires were needed to join the LED branches to the buckets on the D Connectors, completing the circuit. As each column and row connected to a specific transistor, it was important these were kept in the correct order.

4.2 Putting it all together

Assembling the display was simple, by design. The D Connectors on the layers plugged directly into the board and the circuit could be connected directly to a DC power supply. The board was elevated by gluing legs onto the bottom of the board. A problem was found with layers clinging together so small clear perspex rods with evenly spaced slits were used to hold the layers in place. These could easily be removed when it came time to disassemble the display. These features provided a simple method for assembling and disassembling the display. The assembled display is shown in Figure 4.

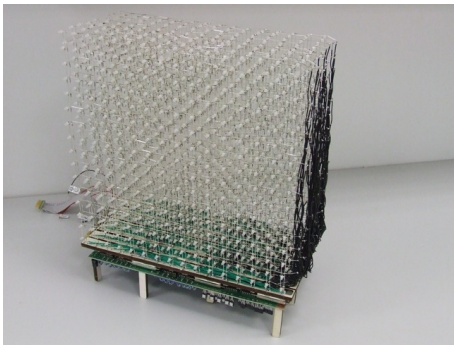


Figure 4: Assembled Half-Cube

4.3 Programming the FPGA

The hardware provided the ability to control every LED of the display individually, and the embedded controller was required to interface to the display as well as managing and manipulating the data for display. Some implementations are listed below:

- **Falling Rain** The first program that was written to test the cube was called Falling Rain. It is essentially a set of random lights that are enabled on the top rows, then at the following clock cycle, the data is shifted to the layer below and a new random set is generated for the top rows. Doing this continuously gives the impression of rain. A Linear Feedback Shift Register (LFSR) was used to generate the pseudo-random data as it could easily be implemented on the FPGA [5].
- **3D Object** To fully show the possibilities of the display, a 3D object was displayed on the cube. Using the Layer Configuration Tool (section 4.4), data was able to be generated that represented an object in three dimensions by drawing a shape, layer by layer.
- **3D Snake** To show some of the other possibilities of the display, a simple snake program was written on

the FPGA, reimplementing the classic arcade game *Snake* in three dimensions.

4.4 Layer Configuration Tool

To generate the data for 3D and 2D objects to place in the cube, a simple program was written in JavaScript to provide a GUI interface allowing the user to effectively draw their desired shape. The GUI then exports the binary code that represents the image to be imported directly into the VHDL. The program also supported loading in previous image data to modify if required. To design 3D objects, the user just needed to draw each layer in 2D and load each layer separately into the VHDL.

5 EVALUATION

5.1 Evaluation of Specifications

After completion of the cube, the display met all requirements set out in the original specifications (Section 3.1).

- **Size:** The display prototype exceeded the minimum 8x8x8 size, being a size of 16x16x8 with the full design planned being 16x16x16.
- **Colour:** The display allows for a range of colours between Green and Red. It presently only supports three specific colours, Green, Orange and Red, however little modification would be required to allow for a smoother colour range between Green and Red.
- **Refresh Rate:** The display updates at a frequency determined by the embedded controller but supports refresh speeds greater than 50Hz as determined by (3.1), which is greater than the minimum 30Hz.

5.2 Design

Finding a way to control such a large number of LEDs with minimal circuitry was a difficult challenge. It required finding a method of providing individual control of the LEDs, while keeping the circuit simple and compact. The design concept was based on the researched hobbyist designs, however as they did not scale well, they needed a lot of adaptation for a large scale design, making the design more complex than hoped for.

5.3 Implementation

Each of the layers took a full day to assemble and the circuit board itself took many days to assemble due to the large number of components. Many days were spent debugging the board. This was required to find where tracks were accidentally joined, or more commonly in this case, finding where connections were broken. Most of the broken connections were a result of the solder paste (required

to join the top layer of the PCB to the bottom layer) not being sufficiently applied and failing to make some required connections. In these cases, wires had to be run to make these connections.

5.4 Testing

The goal of the project was to implement a 3D object inside the display. Before testing how well it could represent a three dimensional object, 3D object data needed to be generated. To do this, the Layer Configuration Tool discussed in section 4.5 was used. However as this creates only 2D shape data, a method was needed to generate the 3D data. Using multiple layered 2D shapes, a 3D shape could be formed. This required some perception to how the data would look in 3D as the tool could not provide such a view until the data was transferred to the display. Assuming correct 3D data, using only 3 colours and pixel spacing of 20mm, it was important the LED pixels accurately represented the desired 3D geometry.

6 CONCLUSIONS AND FUTURE WORK

The project was successful in creating a 3D prototype display. The display forms half of a 16 by 16 by 16 LED cube made up of 4096 tri-colour LEDs. Each of the LEDs can be individually controlled for precise image generation and refreshed rapidly allowing for fluid animations.

Only half the cube was built, producing a 16 by 16 by 8 pixel display. However the design prepared for this and allowed for two halves to be built and assembled side-by-side to form the full cube as each layer was controllable separately. The reason for only half the cube being built was because of time constraints as it required a lot of manual labour to assemble the layers and PCB. The design is also scalable, with the potential of making a smaller or larger display from the same design concepts.

The display was able to render 3D shapes using individual LED control. However, without the full cube, these shapes were limited to half the space and therefore not as impressive as would be seen on the full cube. Brightness was also an issue as the cube required darkness for shapes to be seen well. Unfortunately this was a result of toning down the red in order to mix with the green to make orange. If this was not done, the colour red would appear fine in daylight but green and orange would barely be noticeable. This is a limitation of the LEDs that were purchased from China, due to budget reasons, and higher quality LEDs should provide a brighter colour green.

6.1 Future Work

Some uses of the LED cube have been explored, however there are many more possibilities to be tested. A number

of potential projects and how they may be implemented are listed below.

- **3D Game of Life** Using the cube to show some artificial life algorithms such as the *Game of Life* would be a great demonstration of the potential of the display.
- **Audio Visualisation** This would require implementing an animation based on the frequencies in an audio sample.
- **3D Mirror** Creating a virtual 3D mirror requires gathering three dimensional data and redisplaying this data in real-time. This would allow the viewer to see themselves or an object as if it were a 3D mirror.
- **Range Imaging Integration** An alternative method for gathering the 3D data would be to use the Heterodyne Range Imaging system [6, 7] currently being developed by Victoria University. This could be integrated to show captured data on the display in real-time.

All these projects will work fine on the designed hardware, requiring only further embedded programming.

REFERENCES

- [1] Jiangmen Seekway Technology Ltd., "Seekway 3D LED Cube," May 2010. <http://www.seekway.com.cn/e/ledsystem1.htm>.
- [2] J. Hewes, "Light Emitting Diodes (LEDs)," September 2010. <http://www.kpsec.freeuk.com/components/led.htm>.
- [3] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*. Oxford University Press, international sixth ed., 2011.
- [4] R. J. Tocci, N. S. Widmer, and G. L. Moss, *Digital Systems*. Prentice Hall, tenth ed., 2007.
- [5] N. Weste and D. Harris, *CMOS VLSI Design*. Addison-Wesley Publishing Company, 3rd ed., 2005.
- [6] A. Jongenelen, D. Carnegie, A. Dorrington, and A. Payne, "Heterodyne range imaging in real-time," in *Sensing Technology, 2008. ICST 2008. 3rd International Conference on*, pp. 57–62, nov. 2008.
- [7] A. Jongenelen, D. Carnegie, A. Payne, and A. Dorrington, "Development and characterisation of an easily configurable range imaging system," in *Image and Vision Computing New Zealand, 2009. IVCNZ '09. 24th International Conference*, pp. 79–84, 23-25 2009.

Design of XOR gates in VLSI implementation

Nabihah Ahmad¹ and Rezaul Hasan²

School of Engineering and Advanced Technology
Massey University, Auckland

¹N.Ahmad@massey.ac.nz, ²hasanmic@massey.ac.nz

Abstract:

Exclusive OR (XOR) gate is a fundamental building block in various digital applications such as full adder, comparator, parity generator and encryption processor, which leads to increased interest in enhanced performance of XOR gate. This paper proposes a new design of XOR gate using six transistors for low power application. The new XOR gate has been compared with previous designs in term of power, delay and power-delay product (PDP). The XOR gate is simulated using Cadence Virtuoso Analog Environment in 65nm Complementary Metal Oxide Semiconductor (CMOS) technology at different supply voltages with a range of 0.6V to 1.2V.

Keywords:

XOR gate, power, delay, PDP

1 INTRODUCTION

As the essential unit in digital logic design, XOR gate contribute to the overall performance and power of the system. Therefore it is required to design the XOR gate which satisfies the low power dissipation and delay with small size.

In this paper, we propose a novel design of 2 input XOR gate using six transistors. The paper is organized as follows: in Section II, previous work is reviewed. Subsequently, in section III, the proposed design of XOR gate is presented. In section IV, the simulation results are given and discussed. The comparison and evaluation for proposed and existing designs are carried out. Finally a conclusion will be made in the last section.

1.1 Previous Work

There are varieties of XOR gate design have been reported in literature [1-7]. The conventional design of XOR gate is based on eight transistors in static CMOS [8]. It can operate with full output voltage swing but it requires more numbers of transistors. Emphasis has been done on the design of four transistor XOR gate [1, 3, 6, 7, 9]. Radhakrisnan [3] proposed a combination of

XOR and Exclusive NOR (XNOR) circuit using 6 transistors.

Wang et al. [6] proposed four transistor XOR gate architecture shown in Figure 1(a) and Figure 1(b). These architecture gives a poor signal output for a certain input. The average delay for Figure 1(a) and 1(b) was 3.84ns and 1.42ns respectively with 400uW and 310uW. They improved the level output by cascading a standard transistor inverter as a driving output to achieve a perfect output. This XOR gate is consist of six transistors in total as shown in Figure 2.

In [7], the set of four transistors P-XOR circuit called powerless is proposed which consumes less power than other design because it has no power supply connection. The drawback is it causes a large delay but better than conventional CMOS. The delay was 350ps with maximum input frequency 200MHz.

XOR gate based on Gate-Diffusion-Input(GDI) cell was reported in [9] which requires 4 transistors..

Figure 3 shows the XOR gate circuit in [10] by using three transistors which modify a CMOS inverter with a PMOS pass transistor. It have a voltage degradation when the input A=1 and B=0, but can be minimized by

increasing the W/L ratio of PMOS transistor. It offer less power-delay product compared to four transistors design in [6] using 0.35um technology.

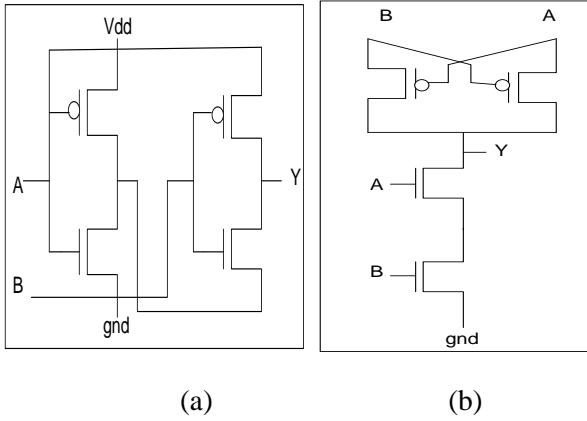


Fig. 1(a) and (b) Design of 4 transistors XOR gate by [6]

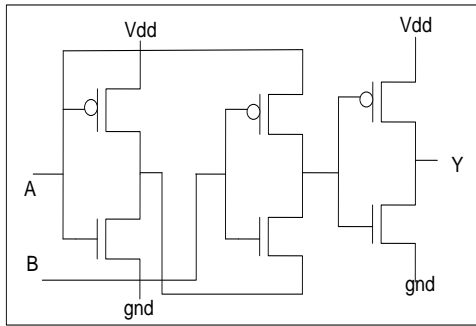


Fig. 2 Improved 6 transistors XOR gate by [6]

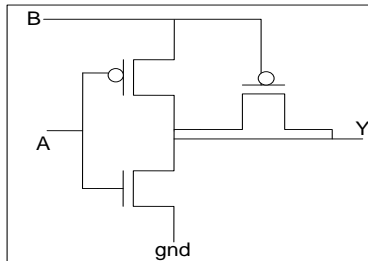


Fig. 3 Design of 3 transistors XOR gate by [10]

2 DISCUSSION

The XOR gate functions is shown in Table 1 and denoted by \oplus . The logic expression for XOR is

$$A \oplus B = A'B + AB' \quad (1)$$

TABLE I
XOR GATE FUNCTION

A	B	XOR
0	0	0
0	1	1
1	0	1
1	1	0

The proposed design of XOR gate using six transistors is shown in Figure 4. It uses a concept of pass transistor and CMOS inverter. The inverter is used as a driving output to achieve a perfect output swing. Vdd connection to transistor M3 and M4 are used to drive a good output of '1'. Transistor M4 is used to drive the output signal when Y=0 when input signal A=B=1. In this condition, when transistor M1 or M2 are ON, it will pass a poor signal '1' with respect to the input to the inverter. The output Y will also be degraded and to achieve a good output signal, transistor M4 is ON when Y=0, then pass the perfect signal '1' from Vdd. The W/L ratio of transistor M2 is bigger than the W/L ratio of transistor M3 to get the output Y=1 as both of the transistors will be ON when A=0 and B=1.

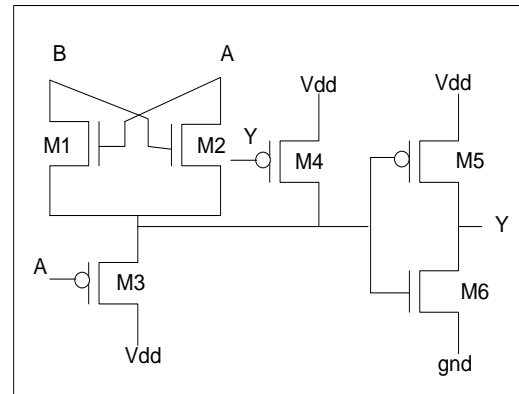


Fig. 4 Proposed 6 transistor XOR gate

The input and output waveform of XOR gate are shown in Figure 5. From the figure it is found that the output waveform for each input combination is full output voltage swing compared to previous design of XOR gate. It eliminates the voltage degradation in certain input.

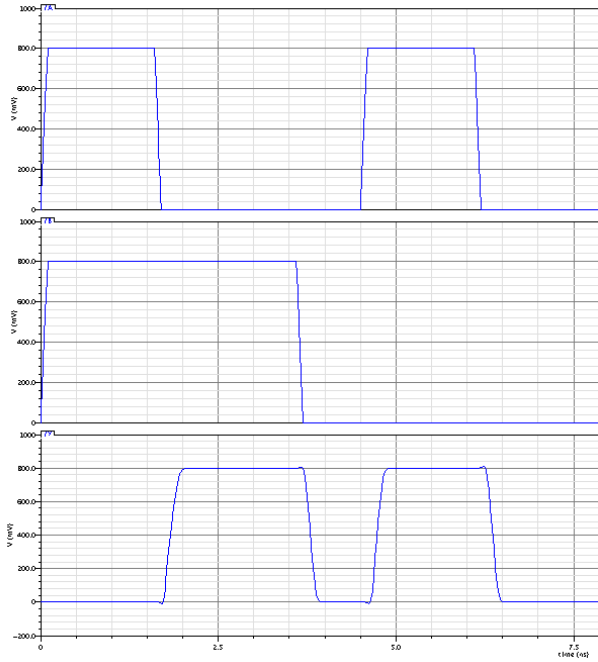


Figure 5 Waveform of the proposed 6 transistor XOR gate

Comparative analysis has been carried out on the different types of XOR gates based on previous XOR design. By using Spectre Cadence, each of the design is simulated using the same conditions to measure propagation delay and power dissipation. The DC and transient analysis of the circuits were performed at a supply voltage at range of 0.6 to 1.2V using 65nm technology with load capacitance of 50fF at 500MHz waveform. The results of simulation which included a delay, power dissipation and power delay product are represented in Table 1 and Figures 6, 7 and 8. The delay has been computed between the time the changing input reaches 50% of voltage level to the time it output reaches 50% of voltage level for both rising and falling transition. The power-delay product (PDP) is measured as the product of the average delay and the average power.

The results indicate that the delay of the proposed XOR gate is between the four transistors XOR gates in [6] and three transistors in [10], and less than six transistors in [6]. The proposed circuit give a lower power dissipation in a low voltage compared to the other design in 4T [6] and 3T [10] which is slightly more than [6] in high voltage but less than [10]. But when compared to 6T in [6], it consumes less power than it. Compared to the design in [6] and [10], the power-delay product of proposed XOR gate is less at the supply voltage between 0.6V and 0.8V. The PDP of six transistors in [6] is the highest from other design including the proposed design.

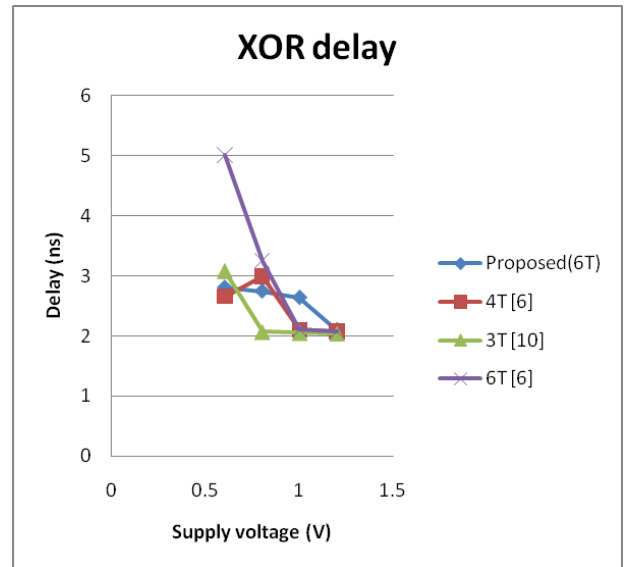


Figure 6 Delay of different XOR gates

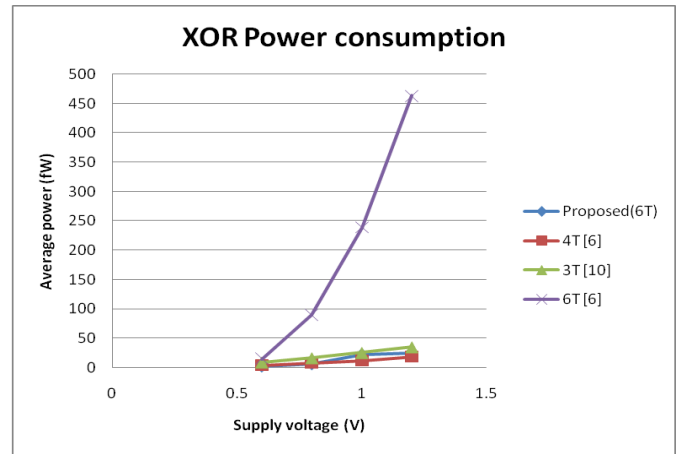


Figure 7 Power dissipation for different XOR gates

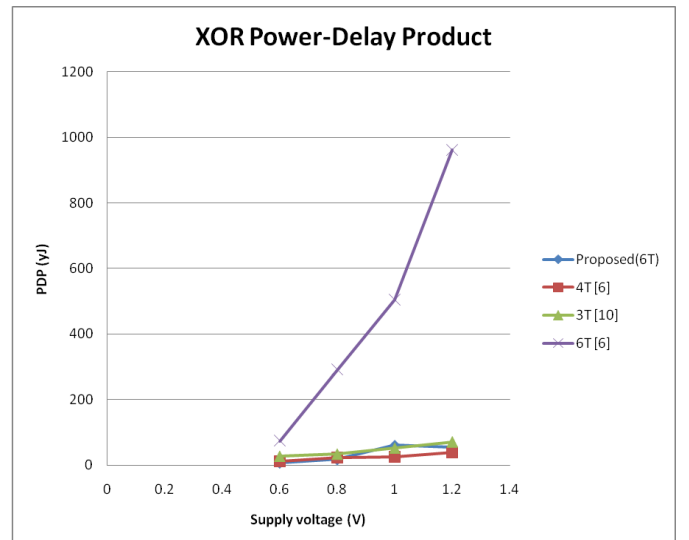


Figure 8 Power-delay product (PDP) for different XOR gates

Table 1 Simulation Results of XOR gate

	V(v)	Proposed(6T)	4T [6]	3T [10]	6T [6]
Delay (ns)	0.6	2.81	2.656	3.08	5.006
	0.8	2.75	2.994	2.061	3.251
	1	2.642	2.094	2.045	2.111
	1.2	2.099	2.068	2.035	2.075
Average power (fW)	0.6	2.312	4.189	8.703	14.66
	0.8	6.069	7.613	16.25	89.57
	1	22.77	12.12	25.36	238.9
	1.2	25	18.81	34.96	463.3
PDP (yJ)	0.6	6.497	11.13	26.81	73.36
	0.8	16.679	22.79	33.49	291.2
	1	60.158	25.38	51.86	504.2
	1.2	52.475	38.91	71.15	961.5

Table 2 Noise Margin of different XOR gate

Type of XOR	V(v)	Voh(V)	Vih(V)	Vil(V)	Vol(V)	Nmh(V)	Nml(V)
Proposed (6T)	0.600	0.600	0.300	0.295	0.000	0.300	0.295
	0.800	0.800	0.400	0.390	0.000	0.400	0.390
	1.000	1.000	0.523	0.477	0.000	0.477	0.477
4T [6]	0.600	0.600	0.340	0.300	0.000	0.260	0.300
	0.800	0.800	0.400	0.400	0.000	0.400	0.400
	1.000	0.908	0.500	0.500	0.000	0.408	0.500
3T [10]	0.600	0.600	0.270	0.330	0.260	0.330	0.070
	0.800	0.800	0.370	0.430	0.300	0.430	0.130
	1.000	1.000	0.470	0.530	0.000	0.530	0.530

Based on DC analyses, the noise margins are measured. The noise margin is the ability to tolerate noise without affecting the correct operation of the circuit [11]. The low noise margin, Nml and high noise margin, Nmh are in the following equation (1) and (2) respectively.

$$Nml = Vil - Vol \quad (1)$$

$$Nmh = Voh - Vih \quad (2)$$

The noise margin of XOR gate has been studied at the different supply voltage as shown in Table 2. The proposed XOR gate indicates acceptable values of noise margin compared with other two XOR gates.

3 FURTHER WORK

The proposed XOR gate will be use in the design of Substitution Box (S-BOX) in Advanced Encryption System (AES) to achieve a good performance because it offer low power and reliable output.

4 CONCLUSIONS

In this paper, a new design of XOR gate has been proposed using six transistors. The performances of this circuit have been compared to previous reported XOR design based on delay, power dissipation and PDP. The proposed circuit give a perfect output signal in all input combinations and better performance especially in low supply voltage compared to the previous designs.

5 REFERENCES

- [1] B. Hung Tien, W. Yuke, and J. Yingtao, "Design and analysis of low-power 10-transistor full adders using novel XOR-XNOR gates," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 49, pp. 25-30, 2002.
- [2] A. M. Shams, T. K. Darwish, and M. A. Bayoumi, "Performance analysis of low-power 1-bit CMOS full adder cells," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 10, pp. 20-29, 2002.
- [3] D. Radhakrishnan, "Low-voltage low-power CMOS full adder," *Circuits, Devices and Systems, IEE Proceedings*, vol. 148, pp. 19-24, 2001.

- [4] L. Hanho and G. E. Sobelman, "New XOR/XNOR and full adder circuits for low voltage, low power applications," *Microelectronics Journal*, vol. 29, pp. 509-517, 1998.
- [5] L. Hanho and G. E. Sobelman, "New low-voltage circuits for XOR and XNOR," in *Southeastcon '97. 'Engineering new New Century'*, *Proceedings. IEEE*, 1997, pp. 225-229.
- [6] W. Jyh-Ming, F. Sung-Chuan, and F. Wu-Shiung, "New efficient designs for XOR and XNOR functions on the transistor level," *Solid-State Circuits, IEEE Journal of*, vol. 29, pp. 780-786, 1994.
- [7] B. Hung Tien, A. K. Al-Sheraidah, and W. Yuke, "New 4-transistor XOR and XNOR designs," in *ASICs, 2000. AP-ASIC 2000. Proceedings of the Second IEEE Asia Pacific Conference on*, 2000, pp. 25-28.
- [8] N. Weste and K. Eshraghian, "Principles of CMOS VLSI Design: A Systems Perspective," Addison-Wesley Longman Publishing Co., Inc. 1993.
- [9] W. Dan, Y. Maofeng, C. Wu, G. Xuguang, Z. Zhangming, and Y. Yintang, "Novel low power full adder cells in 180nm CMOS technology," in *Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on*, 2009, pp. 430-433.
- [10] A. B. Shubhajit Roy Chowdhury, Aniruddha Roy, Hiranmay Saha, "A high Speed 8 Transistor Full Adder Design using Novel 3 Transistor XOR Gates," *International Journal of Electronics, Circuits and Systems* 2;4 2008.
- [11] S. Brown and Z. Vranesic, "Fundamentals of Digital Logic with VHDL Design," McGraw-Hill Higher Education 2005.

Scalable inductively coupled power transfer platform

Dariusz Kacprzak^{*}, Aiguo P. Hu[†], Pratik Raval[‡]

Department of Electrical and Computer Engineering, The University of Auckland
Private Bag 92019, Auckland 1142, New Zealand.

Email: d.kacprzak@auckland.ac.nz^{*} a.hu@auckland.ac.nz[†] prav010@aucklanduni.ac.nz[‡]

Abstract:

This paper presents the design, simulation and implementation of the concept of a scalable inductively coupled power transfer (ICPT) platform. The design and simulations are done by deriving an analytical magnetoquasistatic solution and using finite element method analysis software. Finally, the ICPT platform is implemented on a single layer of printed-circuit-board of rectangular structure and a custom-designed pick-up core is constructed using ferrite. The resulting system has been experimentally shown to induce a sufficiently constant average total open-circuit voltage across the platform, as required by low-power charging applications, to justify the concept of a scalable ICPT platform.

Keywords:

inductive coupling, scalable platform, wireless power, battery charging

1 INTRODUCTION

The popularity of inductively coupled power transfer (ICPT) is ever increasing due to the associated advantages of wireless power transfer such as; freedom of mechanical movement, low maintenance and longevity. More specifically, ICPT in consumer applications such as low-power inductive charging has witnessed the development of various commercial ICPT charging platforms [1]. Such developments are aided by the recent formation of the wireless power consortium, which aims to standardize ICPT charging for short distances between transmitter and receiver.

To date, a vast proportion of current electronics incorporate rechargeable power cells. The motivation for an ICPT platform in particular, arises to remove the inconveniences caused from physical wires and the need to purchase and replace new cells while translating additional non-contact power transfer benefits to the particular application. The idea behind the ICPT platform is to charge the cells via inductive coupling rather than direct electrical connection. This idea has already been used in the most relevant recent publications [2] [3] to propose an ICPT platform.

In [2], Prof. Ron Hui from City College of Hong Kong publishes a new concept of multi-layer printed-circuit-board (PCB) winding matrices of hexagonal structures. Although such multi-layer windings result in a more even magnetic flux distribution, it is difficult to extend the PCB platform into space. This is because the windings in each layer are required to maintain a certain phase-

displacement relationship. This would make scalability difficult. Nonetheless, useful concepts are introduced towards the development of an ICPT platform. Furthermore; in [3], Casanova, Low, Lin and Tseng, propose a rectangular shaped transmitter coil. Their concept suggests a relatively uniform magnetic field distribution may be attained with a single layer. However, the feature of scalability is absent. Various other related papers solely focus on the development of an ICPT platform [4][5]. Furthermore, the majority of published papers related to the ICPT platform do not utilise the modern finite element method (FEM) analysis softwares of today. Therefore, while there has been much research in the development of ICPT Platforms, there is a need for more extensive study in scalability while also making use of modern software packages. This paper extends the reviewed concepts to design, simulate and finally implement the full ICPT platform system to demonstrate the concept of scalability.

2 ICPT PLATFORM SYSTEM

Figure 1 shows a block diagram of the ICPT platform system. Evidently, the power converter inverts a DC input to distribute an AC current through the PCB track coils. The particular inverter configuration used in this study is the autonomous current-fed push-pull resonant inverter developed in [6]. The operating frequency of the power converter was 65kHz. Following the inverter switching network is the primary compensation, the purpose of which

is to maximize power transfer and efficiency. In this case, a parallel compensating capacitor, otherwise termed the resonating capacitor, was used. The PCB platform consisting of arrays of identical rectangular traces follows, noting that each piece of platform is composed of 2x4 rectangular traces. The custom-designed pick-up is placed on the platform, wrapped in 4 coils of Litz wire to reduce losses from skin effects. In the complete system, the induced voltage in the pick-up coils is resonant with a parallel compensating capacitor. The resulting voltage is then regulated through a power conditioner to satisfactorily operate any load(s). The bridge-full-wave rectifier is suitable for this purpose. The ICPT platform system is shown in Figure 2.

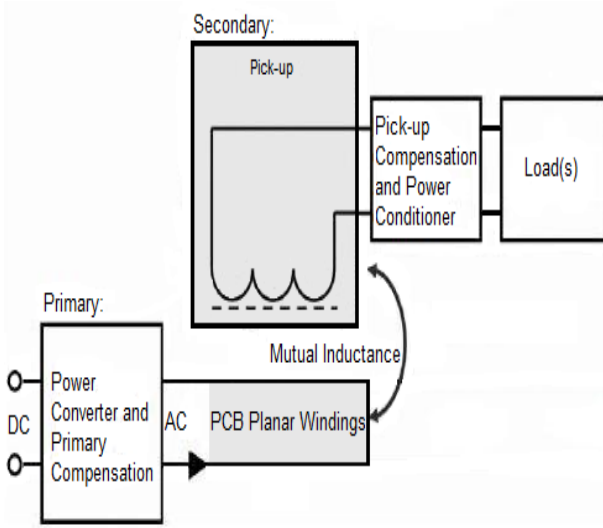


Figure 1: Structure of an ICPT system.

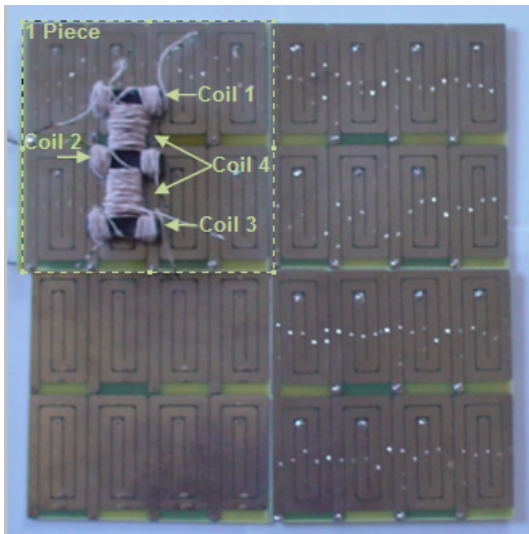


Figure 2: ICPT platform system.

3 DESIGN

This study uses a rectangular design of primary track loop. As part of the design analysis, the analytical magnetoquasistatic solution [7] was derived in cartesian co-ordinates, as in [8], for the 8 infinitesimally-thin-wire current configuration, all with a current of 1A, shown in Figure 3. Firstly, only wires 1-4 were excited from the input current face as indicated in diagram and the average magnetic field (B_1) in each component (x,y,z) computed, with the help of MATLAB, across area 5x5cm in the xy plane at $z=0.1$ cm.

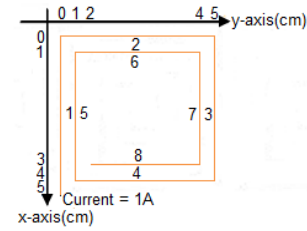


Figure 3: Design configuration.

Secondly, wires 1-8 were excited and the average magnetic field in each component computed (B_2) across the same area. Ideally, without interference effects, one may expect that $B_2=2B_1$. However, this is clearly not the case. As explained in [9], there exists an Inter-Conductor Cancellation Factor (ICCF). Such is the case in this configuration that, adjacent wires with different polarity of current results in interference effects in both the normal and tangential components. This means that in practice, $B_2=2\beta B_1$, where β is defined as the constant of proportionality resulting from interference and is typically less than 1. In other words, β is the loss resulting from interference, compared to an ideal case without interference. This constant is expressed as a percentage loss in each component (B_x, B_y, B_z), in Figure 4, for various z-planes. Evidently, the z or normal component encounters the least loss upon additional turns in the rectangular spiral. Furthermore, it was observed the z-component to have the highest average magnitude across the majority of z-planes considered on the plot. Therefore, it was established that the normal component is of most interest to this study.

Thirdly, in order to consider scalability there are two cases of different current polarities (1A: Clockwise; South pole (S) or -1A: Anticlockwise; North pole (N)) for exciting each corresponding rectangular coil. If a single rectangular coil is extended to all 4 edges and corners, the resulting configuration is shown in Figure 5. Next, only rectangular coil 5 is excited with current in the clockwise direction and the average z-magnetic field (B_5) is computed across area 17x17cm for various z-planes as reference.

As a first case, all coils are excited with a clockwise di-

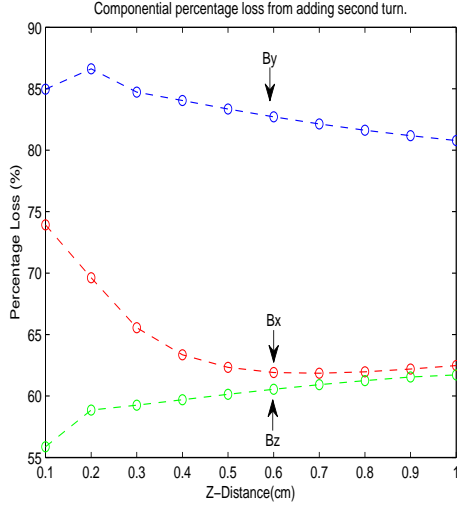


Figure 4: Percentage loss from 2nd turn.

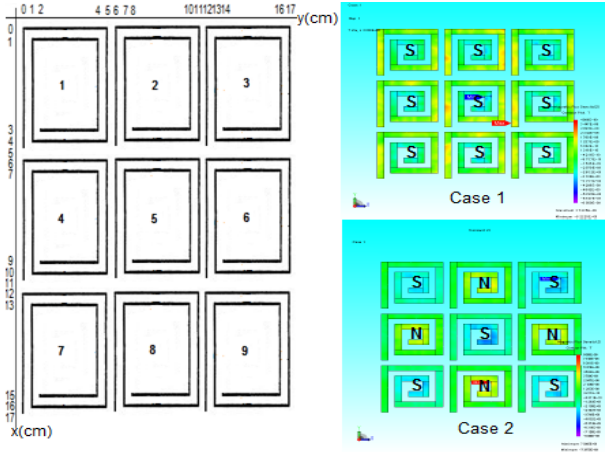
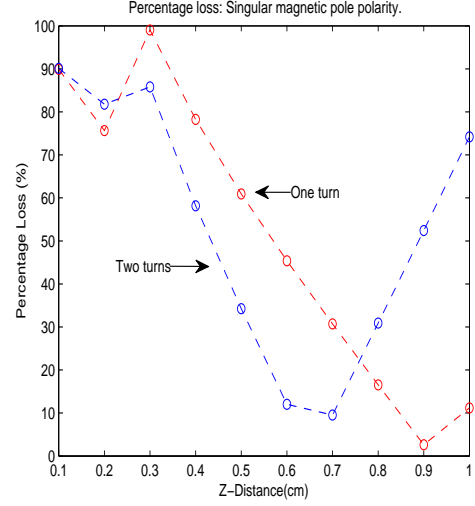


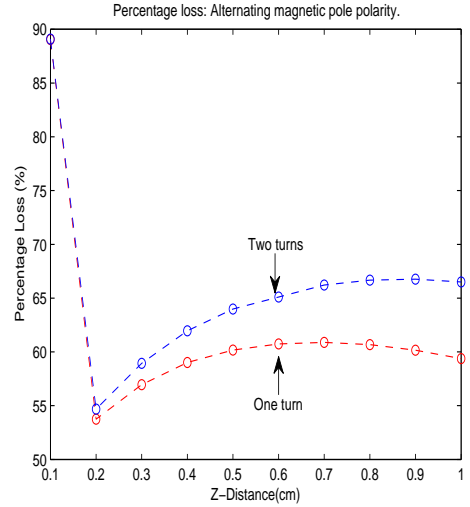
Figure 5: Extended case and magnetic pole polarities.

rected current and the average z-magnetic field similarly computed across the same area. The percentage loss, with respect to B_z , is shown in Figure 6(a). As a second case, coils 1, 3, 5, 7 and 9 are excited with clockwise current while coils 2, 4, 6 and 8 are excited with anti-clockwise current. This results in an alternating north-south magnetic flux pole orientation. The percentage loss was then calculated, in a similar way to the first case, as shown in Figure 6(b). Notably, this was done for the rectangular coils having 1 and 2 turns as informative trends arising from increasing turns. Evidently, the first case encounters a much higher loss, before encountering a minimum, in the near-field region ($z(\text{cm})=[0.2,0.5]$) than the second case. This is an especially critical region as the pick-up is located near-field. In contrast, the percentage loss in the second case is gradually higher and more constant in the far-field region. As the near-field is of greatest interest, and the magnitudes of the fields directly upon each coil are the same but of op-

posite polarity, it was chosen to implement scalability with each rectangular coil having opposite polarity of current (i.e. resulting in north-south pole alternating orientation).



(a) Current in same direction in each coil.



(b) Current in opposite direction in each coil.

Figure 6: Extended case: B_z -component % loss.

4 SIMULATION

The FEM software package from JMag was used during simulations in this study to simulate the chosen primary track loop configuration and to design and simulate the pick-up.

4.1 Primary track loop

The fully extended primary track loop configuration was modelled and simulated for a current of 5A@25kHz. The 3D model dimensions require the rectangular coils to have finite width, chosen to be 10mm, so the overall dimensions are slightly different to the design analysis for infinitesimally thin wires. The z-component of magnetic flux density graph and contour are shown in Figure 7.

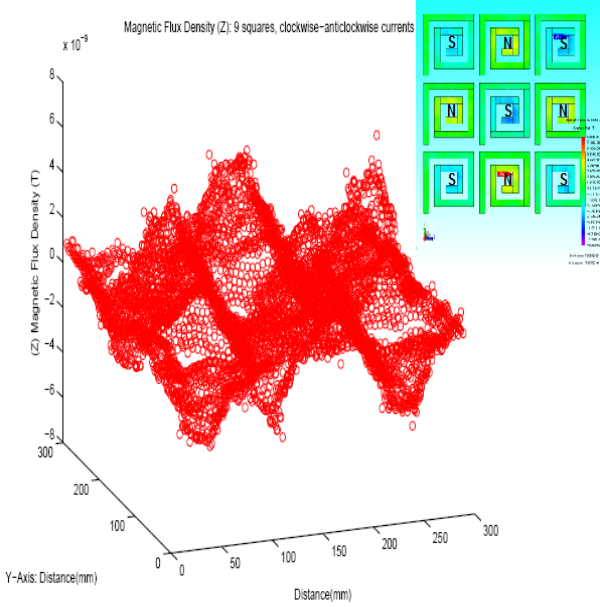


Figure 7: Normal magnetic flux density.

4.2 Pick-up

The pick-up shape was 3D-modelled in JMAG with the objective of ensuring relatively uniform power induction, irrespective of pick-up location on the primary track loop. The resulting proposed design is shown in Figure 8, evidently, 8 power coils wrap around the ferrite pick-up core. The dimensions of the pick-up are optimized to increase the induction of normal component of magnetic flux in the ferrite arms, as shown by the magnetic flux density vector plot in Figures 9(a) and 9(b). In other words, the core legs extend 5mm outside the primary track loop. Evidently, these core legs successfully intercept the normal portion of magnetic flux generated on the outer side by the primary track. Furthermore, the centrally located ferrite arm successfully intercepts the central magnetic flux peak. This demonstrates a potential robustness of the pick-up, in terms of uniform magnetic flux induction.

4.2.1 Performance

In order to test the robustness of the developed pick-up core, it must be evaluated at various locations on an ex-

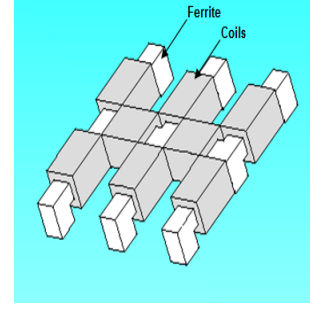
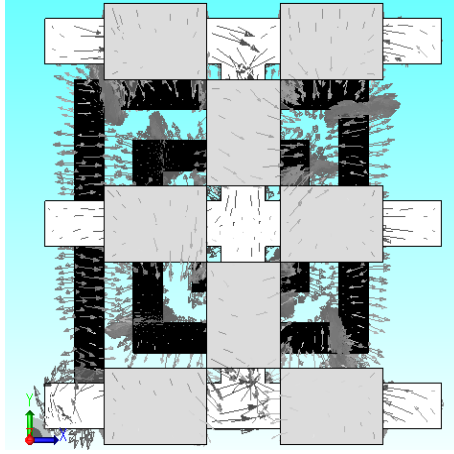


Figure 8: Proposed pick-up.

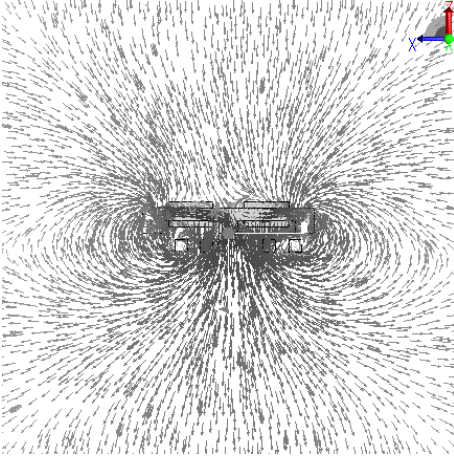
tended platform. This was done by moving the pick-up from positions 1-8 as shown in Figure 10. These positions were specifically chosen to evaluate performance including a combination of various magnetic flux peak and valley zones. As such, positions 1-5 consist of moving the pick-up along a straight segment of 3 rectangular coils in 5 steps. This includes a combination of magnetic flux peak-valley zones on each arm. In positions 6-7, the central arm of the pick-up coincides with valley zones. However, it is evident that at least 2 pick-up legs still coincide with peak zones. In position 8, the pick-up is angularly rotated. Evidently, 4 arms coincide with peak zones. Furthermore, the north-south alternating magnetic flux orientation enable magnetic flux to flow through the ferrite. In this way, magnetic flux induction is ensured virtually irrespective of pick-up location on the platform. The uncompensated power [9], S_u , was then calculated. The main result is an average S_u , as summed in all 8 power coils, across the 8 positions of magnitude 16.77W, with a spread of 52%, or a co-efficient of variation ($COV = \frac{stddev}{average}$) of 18.74%.

5 IMPLEMENTATION

The primary track loop was readily implemented on the top layer of PCB with a typical one ounce of copper thickness. The width of the traces are 4.5mm, resulting in a current density of $6.3 \frac{A}{mm^2}$. Traces on the bottom layer of PCB were used to connect adjacent coils in the chosen north-south pole orientation. Upon implementation, a single piece of platform was measured to have a series inductance and resistance of $1.268 \mu H$ and $283.8 m\Omega$. The implementation of the ferrite core was not so straight forward. In order to construct the custom-designed shape; 3 E-core and 2 bar-core ferrites were used to construct the pick-up via super-glue, with the result shown in Figure 11. Notably, the dimensions of the pick-up are such that all ferrite legs extend 5mm beyond the platform traces. This is done to be consistent with the simulation findings of increasing interception of normal magnetic flux. Overall, the induced magnetic flux undergoes 4 ferrite-glue transitions. Notably, the glueing process is not exact and results in non-uniform air-gaps which decreases the overall



(a) View 1: Without air.



(b) View 2: xz-plane, with air.

Figure 9: Magnetic flux density vector plots.

relative permeability of the core. However, such compromise is inevitable as the pick-up design is specific. 4 Litz wire power coils were then wrapped around the pick-up as shown in Figure 2. Coils 1, 2, 3 have 18 turns while coil 4 has 52 turns. For completeness, coils 1, 2, 3 and 4 were measured to have a series inductance of $14.54 \mu H$, $10.03 \mu H$, $18.10 \mu H$ and $153.1 \mu H$.

As experimental verification, the resonating voltage output from the power converter displaying switching frequency is shown in Figure 12.

5.1 Implemented System Results

This study focuses on the induced open circuit voltage, V_{oc} , in the power coils. This is because the magnetic flux density is directly proportional to the induced open circuit

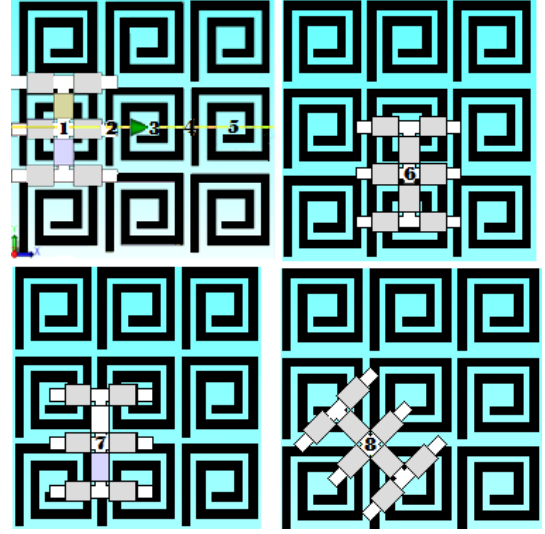


Figure 10: Various pick-up positions.

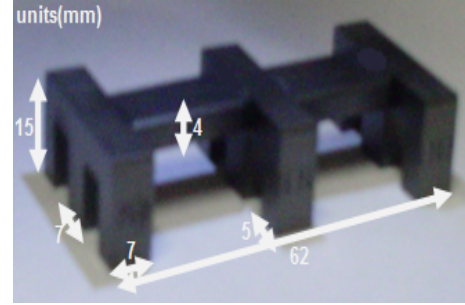


Figure 11: Constructed pick-up.

voltage [9]. As such, the open circuit voltage is measured across a single piece of platform at 0.5cm steps until the pick-up reaches the end of the platform. This means that measurements in the horizontal direction range from 0.5 to 8cm, while vertical measurements range from 3 to 8cm. As a result, a total of 616 measurements were recorded. The result is shown in Figure 13.

Evidently, the average value of the total V_{oc} is 5.747V. This is a result of a total V_{oc} spread of 1.849V. This essentially arises due to the fluctuation in the proportion of voltage induced in each coil. For completeness, the average value of V_{oc} in each coil (1, 2, 3 and 4) is 1.417V, 1.419V, 1.355V and 1.556V.

6 CONCLUSIONS AND FURTHER WORK

This paper has demonstrated the feasibility of the concept of a scalable ICPT platform system, composed of a custom-designed ferrite pick-up core shape to suit a custom-designed planar primary track loop on PCB of rectangular structure, through design, simulation and implementation. The design analysis section has two conclu-

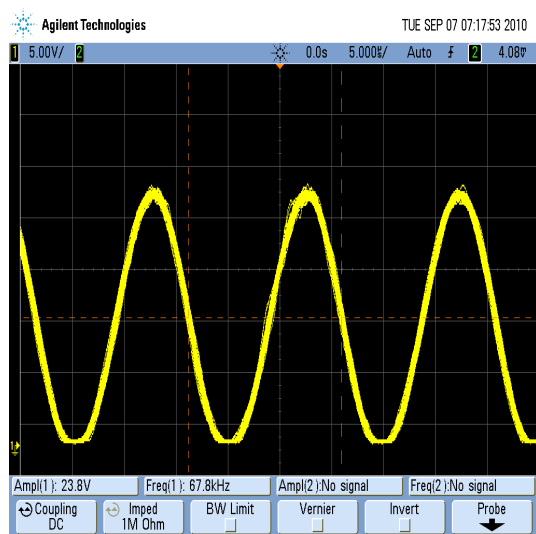


Figure 12: Power converter: Resonant voltage.

sions. First, the normal component of magnetic flux was determined most suitable for multiple turns of rectangular primary coil. Second, the extended platform considered would benefit from current in each adjacent coil to be of opposite polarity. These conclusions were derived from considering the loss from ICCF. The simulation section successfully 3D-modelled the proposed platform and developed a customized pick-up core shape. The proposed pick-up was simulated across 8 different positions on the platform inducing an average total power of 16.77W, with a spread of 52% or COV of 18.74%. Finally; the implemented ICPT system, operating at 65kHz, was successfully tested. It was shown to induce a total average voltage, as summed in 4 Litz-wire power coils, of 5.747V with a spread of 1.849V across a single piece of the platform. This can be considered relatively uniform for low-power charging applications such as mobile phones and music players. As additional pieces of the platform can readily be connected to adjacent sides, an application requiring arrays of primary coils is the simultaneous charging of multiple low-power electronic devices on a single platform.

Future work can include determining and improving the induced power and efficiency of the developed system. In order to do this, a method to yield a strong quality factor is required for a high compensated power. Furthermore, a receiver detection feature could be developed so that only the used portion of the primary track remains excited, reducing electromagnetic interference effects.

7 REFERENCES

- [1] Jurrien I., Wireless chargers from Powermat. CES Report, CES Tradeshow. Thursday 14th, January 2010.
- [2] Hui S.Y.R. Fellow IEEE and Ho W.C., Member

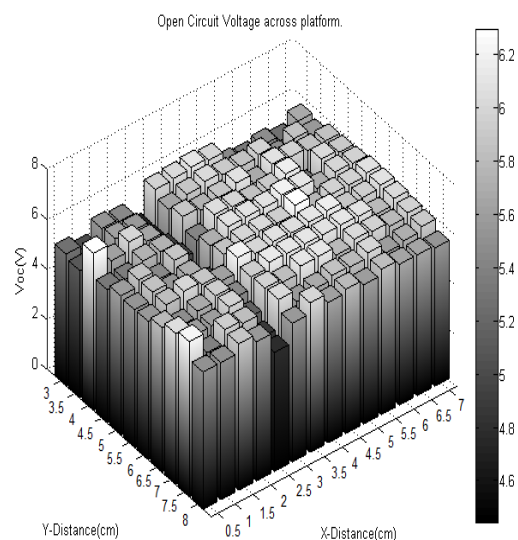


Figure 13: Measured open circuit voltage.

IEEE, A New Generation of Universal Contactless Battery Charging Platform for Portable Consumer Electronic Equipment.

- [3] Casanova J.J., Low Z.N., Lin J., and Tseng R., Transmitting Coil Achieving Uniform Magnetic Field Distribution for Planar Wireless Power Transfer System, Proceedings of the 4th international conference on Radio and wireless symposium, 2009, San Diego, CA, USA.
- [4] Choi B., Cha H., Noh J., Park S., A new contactless battery charger for portable telecommunication computing electronics, ICCE. International Conference on Consumer Electronics, 2000, Page(s): 58-59
- [5] Tang S C., Member, IEEE, Hui S.Y.R., Senior Member, IEEE, and Shu-Hung Chung H., Member, IEEE, Coreless Planar Printed-Circuit-Board (PCB) Transformers A Fundamental Concept for Signal and Energy Transfer in IEEE Transactions on power electronics, vol. 15, no. 5, September 2000
- [6] Hu, A.P., Selected resonant converters for IPT power supplies, PhD thesis, Department of Electrical and Computer Engineering, University of Auckland, Oct 2001.
- [7] Zahn M., Electromagnetic field theory: a problem solving approach. New York: Wiley, 1979
- [8] Raval P., Scalable IPT Platform with receiver detection, ME thesis, Department of Electrical and Computer Engineering, University of Auckland, Dec 2010.
- [9] Kacprzak D., Sykulski J.K., Magnetic design considerations to improve nonlinear characteristics of inductively coupled power transfer systems. COMPEL International Journal. ISSN: 0332-1649.

Design and Simulation of Biosensor for DNA Detection By CMOS Technology

Massoud Alipour¹ and S. M. Rezaul Hasan²

Massey University. School of Engineering and Advanced Technology (SEAT)

¹m.alipour@massey.ac.nz, ²s.m.hasan@massey.ac.nz

Abstract:

This paper focuses the real time detection of the DNA via hybridization and capacitance sensor technique. The detection sensor is designed using two electrodes on a monolithic CMOS wafer in close proximity to each-other. The electrodes act as the plates of a parallel-plate capacitor. ssDNA is bound to the plates in a covalent bond through oxides on a chemical polymer. The function of the designed chip is to detect hybridization between the two electrodes of the sensor. Simulation results show that the circuit has the ability to measure capacitance changes as small as a tenth femto Farad. The proposed design delivers about a hundred times more accuracy as compared to the previously published work.

Keywords:

DNA detection, Biosensor, CMOS, Capacitance sensor, Binding ssDNA.

1. INTRODUCTION

DNA (deoxyribonucleic acid) contains unique genetic information for every human being and living organism present on earth. When obtaining DNA sequencing data, small errors in this information could result in large mistakes in identification. For the medical profession, failure to provide correct DNA data in such areas as identifying pathogenic organisms could result in fatality. Fast and accurate DNA sequencing tools will provide assurance for both professionals and the private individual.

2. The Sensor

The Sensor has been designed to work as a parallel-plate capacitor as shown in Figure 1. The gap between the electrodes is denoted by 'r' where,

$$0.5\mu\text{m} < r < 0.6\mu\text{m} \quad (1)$$

The capacitance of the sensor changes when DNA is attached to the electrodes. In this sensor, first ssDNA is bonded to the electrodes. This first ssDNA acts as our first dielectric and exhibits the dielectric constant denoted by ϵ_1 . After hybridization, the dielectric properties of the ssDNA is changed and a new dielectric value is obtained which is denoted as ϵ_2 .

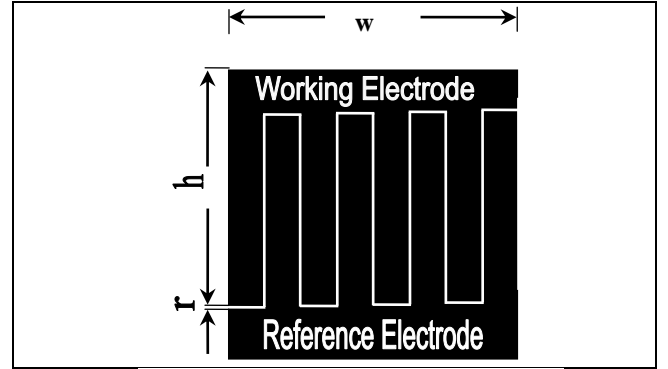


Figure 1: The basic design of the sensor

In the sensor "h" is the height of the electrodes and "w" is the width of the electrodes. The dimensions of the interdigitated sensor electrodes are mostly dependent on the minimum channel length of the integrated circuit process technology.

For a $0.5\mu\text{m}$ CMOS technology the width of the sensor ("w") usually in range of 100 to $200\mu\text{m}$.

$$100\mu\text{m} < w < 200\mu\text{m} [1]. \quad (2)$$

"When C_{ef} is the capacitance per electrode finger, P is the pitch of the electrode finger, d is width of each electrode, t is thickness of each electrode, w is diameter of electrode array and $\epsilon_{analyte}$ is dielectric permittivity of the analyte assay"[1].

$$C_{ef} = \frac{\epsilon_{analyte} \times t \times \pi \times w}{4 \times (p - d)} \quad (3)$$

$$\frac{\pi \times w}{4} \rightarrow \text{average electrode finger height}$$

$$(p - w) \rightarrow \text{is the interelectrode separation}$$

3. Bonding ssDNA to the electrode

This study offers two methods to bond DNA to the electrodes.

3.1. Option 1: Carbodiimide

Ethyl-3-(3-dimethylaminopropyl)-carbodiimide can be used to bond ssDNA to the sensor by way of coupling 5' prime phosphate group end [2]. As shown in figure 2, the reaction in binding activates an EDC (Ethyl-dimethylaminopropyl -carbodiimide) the 5'-phosphate group on the solid condition of primer (It could be a group of phosphoryl on any single or double-stranded of DNA).

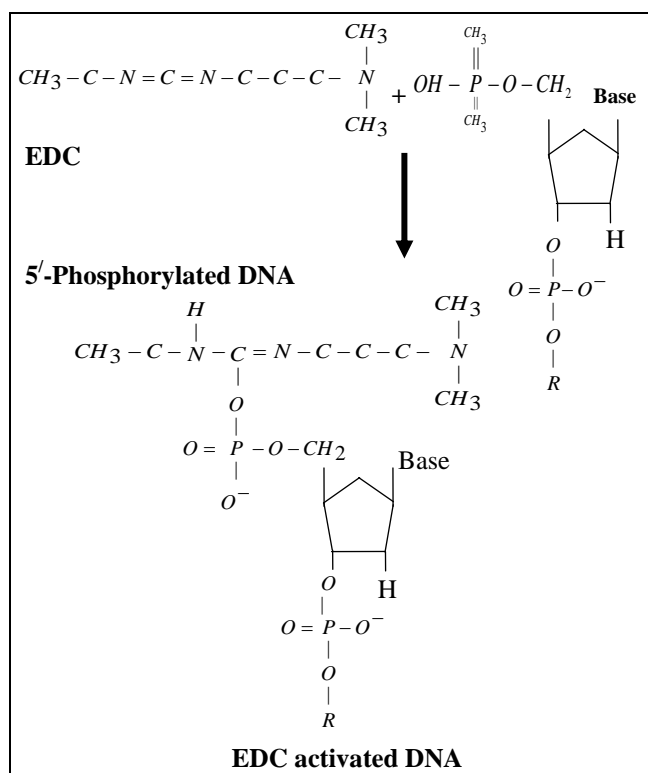


Figure 2: Ethyl-dimethylaminopropyl-carbodiimide

Because the half-life of EDC activated DNA is too short in H_2O (water), further conversion with 1-methylimidazole is necessary. It is more stable in aqueous solution. The reaction of 1-Methylimidazole with the EDC activated DNA and forms another active group.

After attachment to the phosphate group, 1-methylimidazole is capable of leaving the group on the 5'-end of the DNA and investigation for increase of the joint reaction have produce good results with this reagent [3]. On the NucleoLink surface, a reactive group with the 1-methylimidazole activated DNA to form a covalent bond between the electrode fingers and the ssDNA. This bond is thermostable and is not affected by high temperatures in the thermal cycling during the DNA amplification of DIAPOPS¹.

¹ Detection of Immobilized Amplified Products in a One Phase System

3.2. Option 2: SAM and Organic surfaces

Bare metal surfaces, with a coating of metal or metal oxide, will allow bonding of organic materials very quickly because the adsorption reduces free energy between the metal oxides and metal [4]. Self-Assembled Monolayer (SAM) is a simple system, flexible and appropriate for changing the underside surface properties of metals, metal oxides and semiconductors. The group of SAMs that have been studied most are SAMs that resulted from adsorption of alkanethiols and DNA on the surfaces of gold [5], silver [6], copper [7], palladium [8, 9], platinum [10] and mercury (Hg) [11].

4. Design of the capacitance detector circuit

The circuits in this study use the performance characteristics of relaxation oscillators, using very accurate voltage and current sources. An analogy integrated circuit has been designed to convert the capacitance changes of the sensor to a period. Then a digital circuit design is used to convert the changes in period to a digital number.

In relaxation oscillators, proportional current is inversely related to the amount of voltage and capacitance of the DNA sensor. So the frequency of these oscillators can change linearly by with a wide range of external components. Because the frequency or the period of this oscillator is controlled by current, voltage and capacitance values, this type of oscillator can be used in a capacitance meter. Changes in capacitance of the sensor occur by adding ssDNA or hybrid DNA, which in turn changes the oscillator period.

In figure 3, if the DNA sensor is charged and discharged with a constant current, a triangular wave is generated. To measure the period of the triangular wave, it is converted to a square wave by a Schmitt trigger. Measurement of the period of the square wave is achieved by comparison of a counter with a known clock speed. The frequency of clock counter must greater than the maximum frequency created by the sensor - oscillator. Maximum frequency of the sensor - oscillator occurs when the input load is zero.

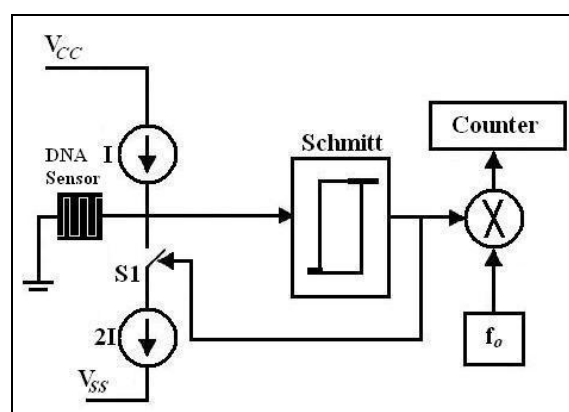


Figure 3: The block diagram of DNA detection.

5. Triangular and square wave period

The diagram designed in the method act as relaxation oscillators with constant current, so the period of their fluctuation (the output of the Schmitt trigger) is proportional to the amount of sensor capacitance and inversely related to the amount of current flow.

If the maximum voltage in charge mode is V_H and minimum voltage for discharge mode is V_L (these values are given by the Schmitt trigger) and charge t_{ch} and discharge t_{ds} of the capacitor can be expressed in following equations:

$$\text{Equation charge: } I \times t_{ch} = C(V_H - V_L) \quad (4)$$

$$\text{Equation discharge: } I \times t_{ds} = C(V_H - V_L) \quad (5)$$

Total times of t_{ch} and t_{ds} equal the triangle or square wave period so, the sum result of the above two equations obtains the following time period.

$$T = t_{ch} + t_{ds} = 2 \frac{V_H - V_L}{I} C \Rightarrow T \propto C \quad (6)$$

If the current charge and discharge are different to each other then:

$$T = (V_H - V_L) \left(\frac{1}{I_1} + \frac{1}{I_2} \right) C \Rightarrow T \propto C \quad (7)$$

The period of square or triangular wave is dependent on the amount of capacitance provided by the sensor; the period has a proportional linear relationship with it. So by measuring the period, the presence of DNA on the sensor can be confirmed.

6. Disclosure period changes with changes in capacitance

As mentioned before, the time period is proportional to the amount of capacitance provided by the sensor. The sensor in normal mode has a capacitance of C_0 . The input load changes with the application of DNA resulting in $C = C_0 + \Delta C$. The goal in this study is to determine the sensor capacitor changes, namely ΔC .

There are many methods to extract ΔC from T. Among of these methods, the method which has been used in this project is a mirror circuit. The purpose of designing a mirror circuit is to use two parallel circuits acting in tandem (figure4). Usually when constructing capacitor sensors to eliminate the environmental impact (e.g. temperature, humidity), a reference capacitor C_0 and a capacitance sensor C_s is used. Thus, using two parallel circuits, one of the counter, measures the period of the reference sensor C_0 and other counter to measures the

period provided by the DNA sensor of $C_s = C_0 + \Delta C$

If $K = 2 \frac{V_H - V_L}{I}$ is assumed, then:

$$T_1 = KC_0 \text{ And } T_2 = K(C_0 + \Delta C) \quad (8)$$

T_1 and T_2 are periods that measured by the counters. When the outputs from the two counters is taken, then the time difference ΔT is calculated as a linear function proportional to the sensor capacitance changes (ΔC) obtained. Equation (9) shows this theory. The advantage of using the two different time period is to minimise real-time losses due to environmental conditions. Environmental factors such as temperature, humidity, etc. on both C_0 and C_s will be equal in both circuits. In this design the output result is displayed as digital number.

$$\Delta T = T_2 - T_1 = K\Delta C \quad (9)$$

Thus, by measuring the two different time period, the changes in capacitance are evident ΔC . Also instead of using the mirror circuit and the reference capacitor C_0 , can use only a counter which this counter is able to count the periods that created by the DNA sensor and by a simple processor, a constant value is subtracted from the output (equal to the period established in the oscillator with the capacitor C_0). Disadvantage of this method is to ignore the environmental impact. To resolve this problem, it is possible to modify the processor so that the measured the period of the sensor before DNA is added and held it as a value in a register. After applying DNA, the time period in the register is subtracted to give the result as ΔT in the output.

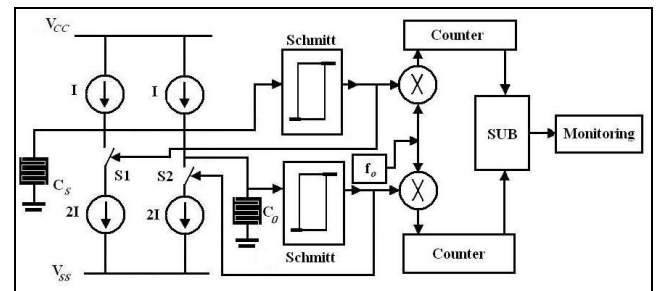


Figure 4: Mirror circuit block diagram of DNA detector

7. Design of current sources

The main determinant of current sources is the minimum capacitance changes of the sensor, which is also related to the speed of counter, the speed of comparator and off - on timing of the switches.

Other factors determining current are noise and leakage currents. Noise current and leakage currents are very small (less than hundredth nano amps).

Maximum capacitance change in the sensor is 0.6 of initial value (The DNA sensor with no load) for example

if an accuracy of 50fF is required to detect the sensor changes, the counter must add a number one for 50fF.

If an 8-bit counter is used a primary capacitor (256×50fF÷1.6=8pF) is selected. The main factors in determining the amount of current source are counter frequency and speed of comparator. For example if the counter works with frequency of 2.5MHz, in this case:

$$T = \frac{1}{2.5\text{MHz}} = 0.4\mu\text{s} \quad (10)$$

So the minimum time detectable by the counter is larger or equal to 0.2μs and is less than 0.4μs. If capacitance changes ΔC=5fF and time ΔT=0.4μs the capacitance changes of the sensor require an accuracy of ΔC=5fF.

$$\Delta T = \frac{2 \times \Delta V}{I} \times \Delta C \quad (11)$$

If we assume ΔV = 2, we have:

$$0.4\mu = \frac{2 \times 2}{I} \times 5\text{fF} \Rightarrow I = 50\text{nA} \quad (12)$$

Now if ΔT = 0.4μs, in this case should be S.R¹ in the Schmitt trigger assuming output changes size by ±2V equal or greater than 100, because:

$$t_r \text{ or } t_f \ll 0.4\mu\text{s} \quad (13)$$

By assuming t_r=t_f=0.05×0.4μs this can be written:

$$t_r = t_f = 0.02\mu\text{s} \Rightarrow \text{S.R} = 2/0.02 = 100\text{V}/\mu\text{s} \quad (14)$$

The result shows that the numbers for S.R and “T” are reasonable numbers. Whenever S.R and the work frequency of counter are larger, smaller changes can be measured in capacitors.

8. Op-Amp circuits with internal feedback

All Schmitt trigger circuits use some kind of positive feedback. By adding two transistors to the input stage of Op-Amp as in figure (5) positive feedback is created.

The first transistor is current feedback which is connected in series in the source loop of M₁ and M₂ transistors by M₅. This route provides negative feedback. The second route is voltage feedback which is parallel connecting to gate-drain of M₁₀ and M₁₁ transistors.

This route provides positive feedback. If the ratio of positive feedback is less than ratio of negative feedback the median feedback will be negative and producing no hysteresis. If positive feedback ratio is larger than the negative feedback coefficient, hysteresis is created. Creation of hysteresis can be expressed as fol-

lowed. Until comparing $\frac{\beta_{10}}{\beta_3}$ and $\frac{\beta_{11}}{\beta_4}$ is less than one,

hysteresis will not be created in the conversion function,

but when the proportions are larger than one then hysteresis is created.

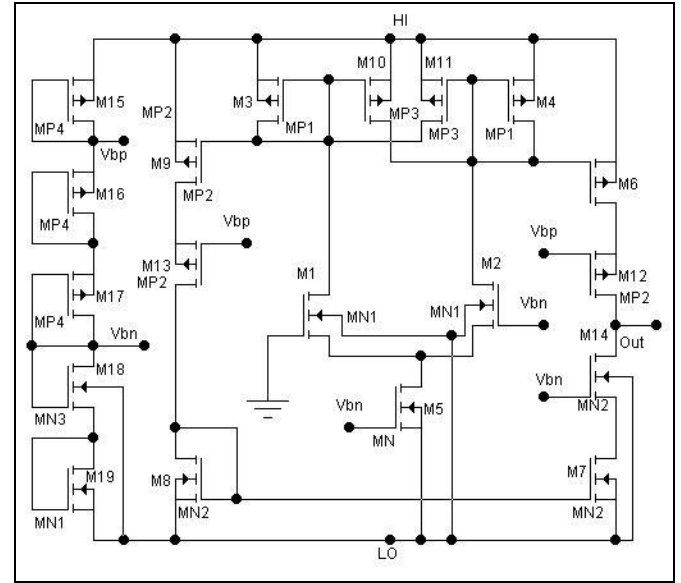


Figure 5: Hysteresis comparator circuit

In the circuit shown in figure 5, if positive and negative of the power supply is connected and gate of M₁ connected to the ground, much less than zero is applied to the input of the transistor, M₁ will be connected and M₂ will be cut, thus M₃, M₁₀ connected and M₄, M₁₁ cut. In this case, all current of I₅ passes from M₁ and M₃ (M₁₀ is connected but it has zero current) so V_{D2} is high. In that moment M₁₀ is trying to pass the following current:

$$i_{10} = \frac{(W/L)_{10}}{(W/L)_3} i_5 \quad (15)$$

By increasing V_{in} to the threshold point (which is not yet known), the remaining current I₅ starts to pass through the M₂. This practice is repeated up to the point where current through M₂ equals to the current of M₁₀. Just after this point the comparator's mode will change. To calculate the voltage at one of the circuit points we must analyse the circuit at the point where the current is equal for i₂ and i₁₀. It is calculated as follows:

$$i_{10} = \frac{(W/L)_{10}}{(W/L)_3} i_5 = i_3 \quad (16)$$

$$\text{and } i_2 = i_{10} \text{ and } i_5 = i_2 + i_1 \text{ and } i_1 = i_3 \quad (17)$$

$$i_3 = \frac{i_5}{1 + \left[\left(\frac{W}{L} \right)_{10} / \left(\frac{W}{L} \right)_3 \right]} = i_1 \text{ and } i_2 = i_5 + i_1 \quad (18)$$

By having currents for M₁, M₂ we can easily calculate the voltages for V_{GS}. Since the gate M₁ is connected to ground, the voltage gate - source is a positive voltage and its value is as follows:

¹ Stochastic Resonance

$$V_{GS1} = \left[\frac{2i_1}{\beta_1} \right]^{V_2} + V_{TH1} \text{ and } V_{GS2} = \left[\frac{2i_2}{\beta_2} \right]^{V_2} + V_{TH2} \quad (19)$$

$$V_{THR} = V_{GS2} - V_{GS1} \quad (20)$$

After reaching to the threshold voltage, the comparator mode will change which most remaining current passes through M_2 and M_4 . M_{11} will connect (on) and M_1 , M_3 , M_{10} are cut. When the input is negative, the current for M_1 will increase until be equalled to the current of M_{11} . The input voltage at this point is V_{THR} . This voltage is obtained in the following equation same as the previous mode:

$$i_4 = \frac{i_5}{1 + \left[\left(\frac{W}{L} \right)_{11} / \left(\frac{W}{L} \right)_4 \right]} = i_2 \text{ and } i_1 = i_5 \cdot i_5 \quad (21)$$

Using the equation (19) to calculate V_{GS} , the negative voltage at this point is equal to:

$$V_{THR} = V_{GS2} - V_{GS1} \quad (22)$$

9. Effect of voltage supply on the output period of the Schmitt Trigger

In the Schmitt trigger figure (5) threshold voltages are dependent on current in transistor M_5 . In this circuit, bias voltage of transistor M_5 by voltage division of power supplied by transistors M_{15} to M_{19} is produced. In this case by changes in voltage of power supply, bias voltage of transistor M_5 will be changed and it causes changes in the current of the transistor M_5 or in other words it will change Schmitt trigger threshold voltage. Changes in the threshold voltage also cause change in the output period in the Schmitt trigger circuit. To prevent changes in threshold voltages it is necessary to stabilise the bias voltage or in other words, the current in the transistor M_5 . Stabilizing current bias M_5 using the schematic of figure (4) with a mirror current.

Because of errors by way of voltage drain-source in transistors M_5 and M_{20} , it is better to mirror current to stabilise bias voltage in transistor of M_5 . To stabilise the bias voltage of M_5 , figure (6) has been used.

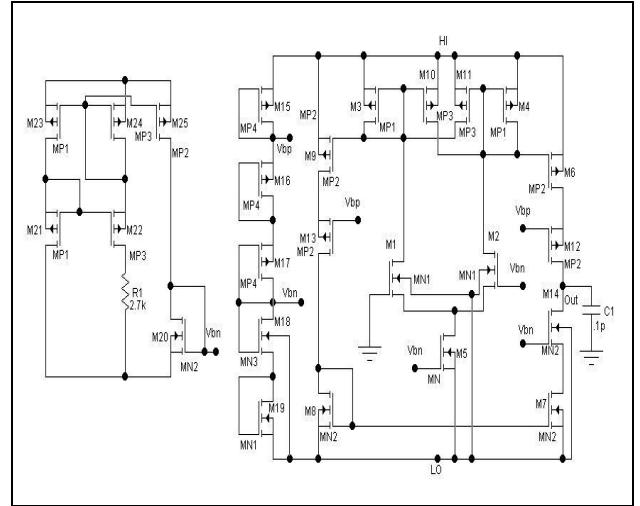


Figure 6: Stabilizing bias voltage in figure (5) in order to stabilise Schmitt trigger threshold voltage.

10. Effect change of temperature on capacitance in the sensor

Ambient temperature changes, in addition to changing the overall capacitance in the integrated circuit also causes changes in the internal capacitance of the sensor. Because the mirror circuit design calculates differential performance in the DNA detector circuit, errors through the effect of changes in temperature or other environmental factors is eliminated because both sensors are subject to the same conditions.

11. Effect of noise on wave period of the Schmitt Trigger Output

Noise voltage creates jitter in the output of the Schmitt trigger circuit or oscillator. Jitter causes random changes in the threshold voltage of the Schmitt trigger. Because of the use of MOSFET in current sources and switches in this circuit, effective noise is thermal noise. The amount of jitter in the DNA sensor is in the range of 2pf to 20pf and in the output of Schmitt trigger is in range of picoseconds. Because the minimum periods established are in the microsecond level, jitter does not create any appreciable error in the period.

12. The Digital section of the DNA detector

To measure the time period of the upper pulse for the Schmitt trigger output, a counter with a certain frequency is required. The frequency counter value is dependent on the minimum and maximum upper pulse time in the output of Schmitt trigger, the number of bits in the counter and accurate measurement.

The frequency that selected must not overflow the counter for $T_0 + \frac{1}{2}T_0$. For example for a counter $n=12$ bits, initial capacitance sensor $C_0=2pF_0$, the sensor charge and discharge current $I=1nA$, Threshold voltages

$V_{IH}=1V$ and $V_{IL}=-1V$ so the counter frequency can be obtained by the following:

Maximum period created by the sensor capacitor is equal to:

$$T = \frac{2(V_{IH} - V_{IL})}{I} C_0 \left(1 + \frac{1}{2}\right) = 3C_0 \frac{(V_{IH} - V_{IL})}{I} \quad (23)$$

Value upper pulse time of Schmitt trigger output is:

$$t = \frac{T}{2} = 1.5C_0 \frac{(V_{IH} - V_{IL})}{I} \quad (24)$$

With values above $t = 6ms$. In this time counter must have a maximum count up to 2^n-1 , so:

$$f = \frac{2^n}{t} = 682666.67 Hz \quad (25)$$

In this case, measurement accuracy is constant and obtained as follows:

$$\frac{1}{f} = 2 \frac{V_{IH} - V_{IL}}{I} \Delta C \Rightarrow \Delta C = \frac{1}{2} \frac{I}{f(V_{IH} - V_{IL})} \quad (26)$$

ΔC indicates the accuracy of measurement. With the above values, the accuracy of measurement level is 732.4aF (Atto= 10^{-18}). By using 16-bit counter $\Delta C = 45.8aF$ and $f=10922666.67Hz$. But if the value of ΔC is known, in this case, ΔC determines the frequency counter and the number of the counter bits. The following equations should be accurate between precision measurements, the counter frequency, the number of counter bits, and the time signal is high.

$$\Delta t = \frac{1}{f} \text{ and } t = 2^n \Delta t \text{ and } \Delta C = \frac{1}{2} \frac{I \Delta t}{(V_{IH} - V_{IL})} \quad (27)$$

Control pulses of the counter should be designed so that it only counts when the pulse of the Schmitt trigger is in upper cycle. For this reason a counter which has a start and reset pin can be used. Thus, when the pulse has a rising edge, the counter is reset and starts counting on the pull-down edge. The output of the counter is recorded in a latch.

The buffer used in the input of the latch is to increase the accuracy. It is better that the delay of the buffer is the same or slightly larger than delay time of the counter. This part of circuit is designed with a clock frequency 680 kHz and $I=50nA$.

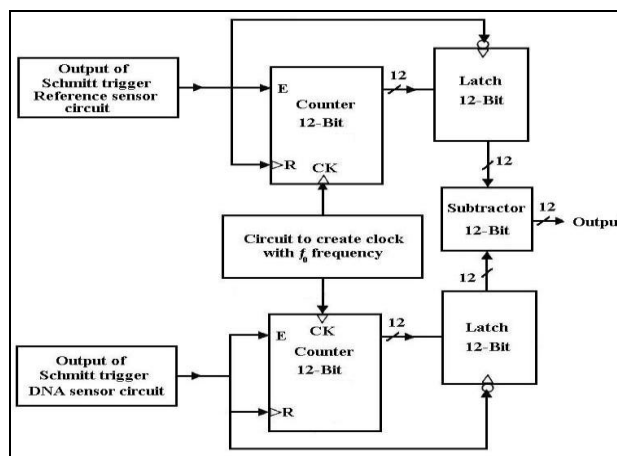


Figure 7: Digital circuit of changing capacitance detection

13. References

- [1] Hasan, S. M. R., "Stability analysis and novel compensation of a CMOS current-feedback potentiostat circuit for electrochemical sensors," *IEEE Sensors J.*, vol. 7, no. 5, pp. 814-824, May 2007.
- [2] Tedder JM, Nechvatal A, Murray AW, et al. (1972) Amino-acids and proteins. In: Basic organic chemistry. London: John Wiley & Sons, Chapter 6:305-342.
- [3] Rasmussen SR, Larsen MR, Rasmussen SE. (1991) Covalent immobilization of DNA onto polystyrene microwells: The molecules are only bound at the 5' end. *Anal Biochem.* 198:138-142.
- [4] Adamson, A. W.; Gast, A. P. *Physical Chemistry of Surfaces*, 6th ed.; Wiley-Interscience: New York, 1997.
- [5] Poirier, G. E.; Pylant, E. D. *Science* (Washington, D.C.) 1996, 272, 1145.
- [6] Walczak, M. M.; Chung, C.; Stole, S. M.; Widrig, C. A.; Porter, M. D. *J. Am. Chem. Soc.* 1991, 113, 2370.
- [7] Laibinis, P. E.; Whitesides, G. M.; Allara, D. L.; Tao, Y. T.; Parikh, A. N.; Nuzzo, R. G. *J. Am. Chem. Soc.* 1991, 113, 7152.
- [8] Love, J. C.; Wolfe, D. B.; Haasch, R.; Chabinyc, M. L.; Paul, K. E.; Whitesides, G. M.; Nuzzo, R. G. *J. Am. Chem. Soc.* 2003, 125, 2597.
- [9] Carvalho, A.; Geissler, M.; Schmid, H.; Micel, B.; Delamarche, E. *Langmuir* 2002, 18, 2406.
- [10] Li, Z.; Chang, S.-C.; Williams, R. S. *Langmuir* 2003, 19, 6744.
- [11] Muskal, N.; Turyan, I.; Mandler, D. *J. Electroanal. Chem.* 1996, 409, 131.

Optimising antenna locations in large radio telescopes using iterative projection algorithms

Joe Chen^{*}, Veit Elser[†] and Rick Millane^{*}

^{*}Computational Imaging Group,
Department of Electrical and Computer Engineering,
University of Canterbury,
Private Bag 4800,
Christchurch, New Zealand.

[†]Physics Department,
Cornell University,
Ithaca, NY 14853,
USA.

Email: rick.millane@canterbury.ac.nz

Abstract:

A novel approach to optimising antenna locations in radio telescopes using iterative projection algorithms (IPAs) is presented. The problem is formulated in simple terms with terrain and visibility constraints. A simple linear approach using IPAs is first described although this is too unwieldy for a large number of antennas. Another approach is described in which the antenna locations are described probabilistically and the problem is mapped to a phase retrieval problem which is also solved using IPAs. Simulations for snapshot imaging and a simple terrain constraint indicates the potential of this approach.

Keywords:

Radio telescope, radio astronomy, square kilometer array, antenna, phase retrieval, iterative projection algorithm

1 INTRODUCTION

Radio telescopes are a synthetic aperture device, i.e. they achieve high resolution imaging by coherently combining signals from a number of widely spaced small antennas to synthesise a large antenna [1]. Of course the synthesis is imperfect due to the gaps between the antennas where no signal is received. A key component of radio telescope design is selection of the antenna locations in order to optimize the imaging performance. In practice, the optimization has multiple objectives for a variety of observation situations that include source positions, observation durations, spectral coverage, scientific objectives (single field, survey, detection, etc.), cost, ground coverage, and operational requirements. Here we address only one of these, the problem of selecting antenna locations to achieve a desired distribution of Fourier, or visibility, samples, subject to constraints on the antenna locations (sometimes called terrain constraints). Terrain constraints may be necessary to exclude particular regions (rivers, mountains, inaccessible regions, etc.) but also for engineering requirements (e.g. power and telecommunications access). We consider this problem in the context of “large” radio telescopes, i.e. telescopes with a large number of antennas. Current synthesis radio telescopes have less than about 50 anten-

nas. However, planned radio telescopes such as the Square Kilometer Array (SKA) [2, 3] will have of the order of thousands of antennas. The nature of the design problem changes with a large number of antennas and has been little explored for such a large number.

In a synthesis radio telescope, the signals received at pairs of antennas are cross-correlated. The correlation for a particular antenna pair gives the value of the Fourier transform (the so-called visibility function in radio astronomy) of the image (the radio brightness on the celestial sphere) at a position in Fourier space that is a scaled version of the vector between the two antennas (the so-called “baseline”) (Fig. 1). With measurements from multiple baselines (antenna pairs), the data can then be gridded and inverse Fourier transformed to obtain the image. For high resolution imaging, it is desirable to have a smooth distribution of samples in Fourier space. The antenna beam pattern is the inverse Fourier transform of the Fourier space sampling pattern and it is also desirable to reduce the sidelobe level in the beam pattern. A compromise has to be reached between uniform sampling and low sidelobes which can be defined by selecting a target Fourier sampling pattern distribution.

To avoid confusion with a different Fourier transform used below, the term visibility will now be used exclu-

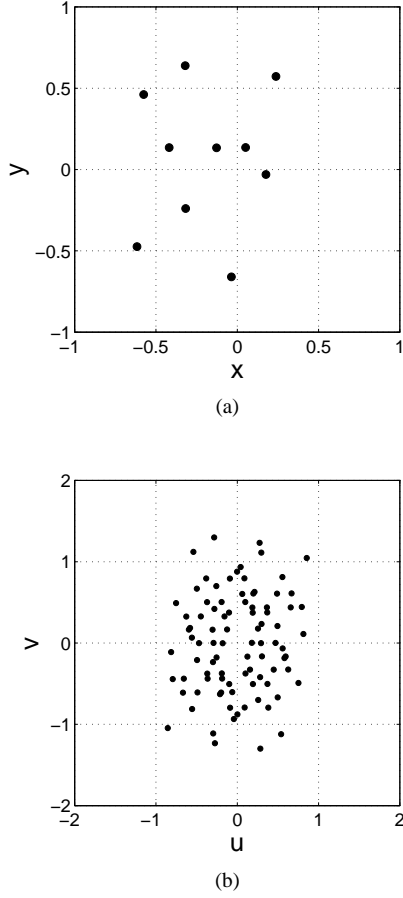


Figure 1: (a) Antenna locations and (b) positions where the visibility function is measured.

sively for the baseline space. The position of a visibility sample is actually proportional to the projection of the baseline onto a plane normal to the direction of observation. The relationship between baseline and position in visibility space is therefore constant for a fixed observation direction, i.e. for short-exposure, or snapshot, imaging. For long exposure imaging, with the telescope tracking the source, the projection of the baseline changes with time and so the position of the visibility space sample traces out a path in visibility space. The shape of the path depends on the declination of the source. A long exposure therefore has significant advantages as these paths give a much denser sampling in visibility space than do the single samples for a short exposure. However, for the sake of simplicity, we consider here only snapshot imaging.

Currently, various methods are used for planning antenna placements. Most of these address, at least, a target visibility sample distribution. The simplest case is snapshot mode, but in general long-exposure modes must be considered for a range of source declinations. Most methods employ some kind of numerical optimization scheme to search for solutions that satisfy constraints on the dis-

tribution of baselines. They generally move individual antennas in order to gradually reduce a cost function that is related to the difference between the current and desired properties of the baseline distribution. An inherent problem is that moving one antenna to change one baseline affects $M - 2$ other baselines. Various optimization methods are used including simulated annealing [4], neural networks [5], gradient methods [6], tomographic projection and a line search [7], and particle swarm methods [8]. The problem in general is highly complex and most methods operate on a restricted problem domain of some kind.

Here we discuss the problem of optimizing antenna locations subject to terrain constraints and a desired visibility distribution. The problem is first formulated in terms of antenna positions and visibility samples. Iterative projection algorithms (IPA), a global optimisation method for constraint satisfaction problems, are then briefly reviewed. The IPAmethod is then applied to the problem at hand. An alternative approach suitable for a large number of antennas, also based on IPAs, is then described.

2 FORMULATION

Consider a telescope with M antennas at positions (x_j, y_j) , described by the vector \mathbf{x} of length $2M$, where

$$\mathbf{x} = (x_1, x_2, \dots, x_M, y_1, \dots, y_M)^T. \quad (1)$$

The baselines (u_j, v_j) are given by the separations between pairs of antennas, i.e.

$$(u_j, v_j) = (x_k, y_k) - (x_l, y_l) \quad \text{for } l > k, \quad (2)$$

with some convenient mapping from (k, l) to j . This gives $P = M(M - 1)/2$ baselines which can be described by the vector \mathbf{u} of length $2P$, where

$$\mathbf{u} = (u_1, u_2, \dots, u_P, v_1, \dots, v_P)^T. \quad (3)$$

The relationship between the baselines and the antennas can be written as the system of linear equations

$$\mathbf{u} = \mathbf{T}\mathbf{x}, \quad (4)$$

where \mathbf{T} is an easily determined sparse matrix with non-zero elements 1 and -1. Equation (4) is an overdetermined system which has solution

$$\mathbf{x} = \mathbf{T}^{-1}\mathbf{u}, \quad (5)$$

in the absence of noise. Note however that the baselines only define the relative positions of the antennas so that there are two degrees of freedom (corresponding to, say, the position of one antenna or the centre of gravity of the array) in the solution Eq. (5) and \mathbf{T}^{-1} has rank $2M - 2$. In practice, in the presence of errors, Eq. (5) is replaced by

$$\mathbf{x} = \tilde{\mathbf{T}}^{-1}\mathbf{u}, \quad (6)$$

where $\tilde{\mathbf{T}}^{-1}$ is a pseudoinverse of \mathbf{T} that minimises an appropriate norm and filters out the two zero singular values. In this formalism, the objective is to find a set of antenna positions \mathbf{x} that satisfies a given set of terrain constraints and also a desired, or target, distribution of visibility samples \mathbf{u} . We use an iterative projection algorithm (IPA), described in the following section, to solve this problem.

3 ITERATIVE PROJECTION ALGORITHMS

An IPA is an algorithm for finding an image, or any unknown function, that satisfies two sets of constraints [9, 10]. In the case at hand, the two constraints are the terrain and visibility distribution constraints.

It is convenient to formulate iterative projection algorithms as operations on points $\mathbf{x} = (x_1, x_2, \dots, x_N)$ in an N -dimensional Euclidean vector space R^N . A point \mathbf{x} in this vector space represents a particular image on a discretised domain of N grid points; components x_j representing the value at the j -th point. The projection of a point \mathbf{x} onto a constraint set A (the set of all points in R^N that satisfy the constraint A), denoted $P_A\mathbf{x}$, is the point in A that is closest (in the Euclidean sense) to \mathbf{x} .

An iterative projection algorithm is an algorithm for finding a point in the intersection of the two constraint sets. At each iteration, an “iterate” \mathbf{x} is updated to form a new iterate \mathbf{x}_{n+1} . The update rule is generally a combination of the projections $P_A\mathbf{x}$ and $P_B\mathbf{x}$ and \mathbf{x} itself. Different IPAs are distinguished by different update rules. The simplest projection algorithm (referred to as the error reduction (ER) algorithm [11]) is that in which P_A and P_B are sequentially applied to \mathbf{x}_n , i.e.

$$\mathbf{x}_{n+1} = P_B P_A \mathbf{x}_n \quad (7)$$

The update Eq. (7) is repeated many times in the hope that the iterate will converge to a solution that satisfies both constraints, i.e. is in the intersection of A and B . The ER algorithm is known to converge (although possibly slowly) to a point (solution) satisfying both constraints if both the constraint sets A and B are convex. However, if one of the constraint sets is non-convex, then the algorithm will usually stagnate at a non-solution unless started close to the solution. The problem with the ER algorithm is that it becomes trapped at fixed points (where $\mathbf{x}_{n+1} = \mathbf{x}_n$) that do not satisfy both constraints. In other areas of image reconstruction, more sophisticated IPAs have been developed that move away from fixed points that are not solutions and explore the vector space until a true solution is found. These include the hybrid input-output (HIO) algorithm [11], the generalized HIO algorithm [12] and the difference map (DM) algorithm [9]. The HIO algorithm, for a particular parameter values, takes the form

$$\mathbf{x}_{n+1} = \mathbf{x}_n + P_A(2P_B\mathbf{x} - \mathbf{x}) - P_B\mathbf{x}. \quad (8)$$

The DM algorithm takes the form

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \beta[P_A F_B \mathbf{x}_n - P_B F_A \mathbf{x}_n], \quad (9)$$

where F_A and F_B are the “relaxed projections”

$$F_A\mathbf{x} = (1 - 1/\beta)P_A\mathbf{x} + (1/\beta)\mathbf{x} \quad (10)$$

$$F_B\mathbf{x} = (1 + 1/\beta)P_B\mathbf{x} - (1/\beta)\mathbf{x}, \quad (11)$$

and β is a constant generally chosen such that $0.7 < |\beta| < 1$. For the DM algorithm the iterate is not an estimate of the solution, and an estimate of the solution \mathbf{x}_{soln} is obtained as

$$\mathbf{x}_{\text{soln}} = P_A F_B \mathbf{x}_n. \quad (12)$$

The DM algorithm has good global convergence properties.

4 A DIRECT APPROACH

With the IPA as described above, the solution to our problem can now be formulated as follows. We equate the vector \mathbf{x} to the antenna positions \mathbf{x} . The antennas are required to be inside some region D (the terrain constraints), and we denote this constraint by the set A . The constraint on the distribution of visibility samples is incorporated by defining a required histogram for the sample positions \mathbf{u} and this is denoted constraint B . The bin widths of the histogram bins are chosen based on the number of antennas and the maximum dimension of the region D .

The projection onto constraint A is given by

$$\begin{aligned} P_A\mathbf{x} &= P_A(x_j, y_j) \\ &= (x_j, y_j) && \text{if } (x_j, y_j) \text{ is inside } D \\ &= \text{closest point on } \partial D && \text{if } (x_j, y_j) \text{ is outside } D. \end{aligned} \quad (13)$$

The projection onto constraint B involves making the minimum adjustment to the antenna positions so that the visibility samples conform to the target histogram. It is convenient to perform this projection onto the two marginal densities in u and v , denoted $h_u(u)$ and $h_v(v)$, derived from the 2D histogram. The projection is written

$$P_B\mathbf{x} = \tilde{\mathbf{T}}^{-1} P'_B \mathbf{T}\mathbf{x}, \quad (14)$$

where the projection P'_B involves sorting the current values of $\{u_j\}$ and assigning them to $h_u(u)$, and the same for $\{v_j\}$.

Simulations show the potential of this approach. Both the ER and HIO algorithms were applied to a simple problem with 20 antennas restricted to a circular region and a circular Gaussian distribution of visibility samples, giving Gaussian marginal densities. The HIO algorithm is asymmetric in the constraints A and B and it was found that the best performance was obtained with the projections P_A

and P_B in Eq. (8) reversed. The HIO algorithm gave better results than the ER algorithm. Example results are shown in Fig. 2 which shows the final antenna distribution and the histogram in v . Note that some of the antenna overlap in Fig. 2a but these are easily separated in practice, and in this case is an artifact of the small number of antennas and bins used.

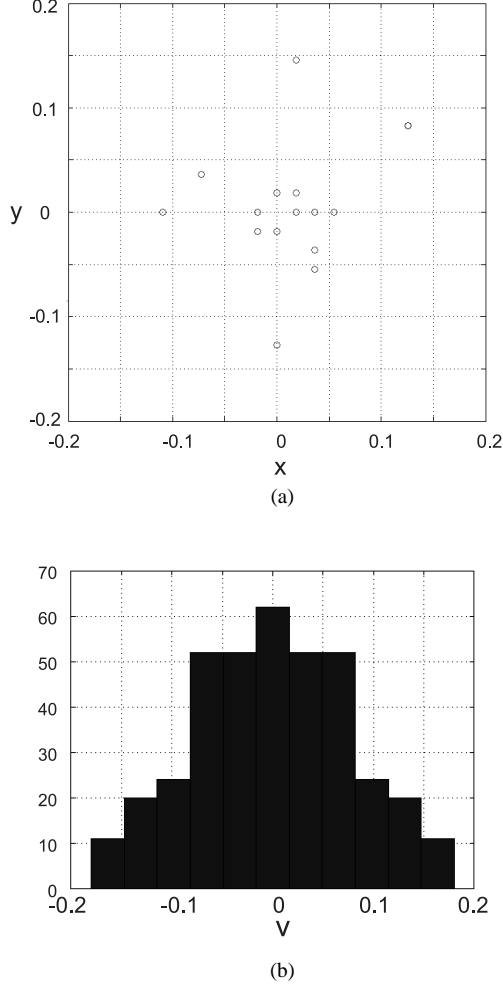


Figure 2: (a) Antenna positions and (b) histogram of visibility samples in the v direction, after application of the HIO IPA.

5 A PROBABILISTIC APPROACH

The approach described above becomes computationally too intensive for a large number of antennas. To circumvent this problem, we propose an alternative approach in which the antenna positions are described by a probability density function (pdf) denoted $f(x, y)$. The antenna pdf is easily mapped to a pdf for the visibility samples denoted $w(u, v)$. The algorithm proceeds in two steps. In the first step, the visibility and terrain requirements are expressed

as constraints on $f(x, y)$ and $w(u, v)$ and a solution for $f(x, y)$ is found by mapping the problem to a phase retrieval problem and using an iterative projection algorithm. In the next step, a set of antenna locations is found by sampling $f(x, y)$.

Since the visibility samples are at the vector separations of pairs of antennas, the pdf $w(u, v)$ is the autocorrelation of the pdf $f(x, y)$, i.e.

$$w(x, y) = f(x, y) \odot f(x, y), \quad (15)$$

where \odot denotes correlation. Note that since $f(x, y)$ and $w(u, v)$ are pdfs, they are both positive real.

The mapping from $f(x, y)$ to $w(u, v)$ is well-defined, but that from $w(u, v)$ to $f(x, y)$ is one-to-many. A convenient way to separate out the degrees of freedom in the inverse mapping is to map the problem to a phase retrieval problem [13]. Denote the two-dimensional Fourier transforms of $f(x, y)$ and $w(u, v)$ by $F(\xi, \eta)$ and $W(\mu, \nu)$, respectively. We then have that

$$W(\mu, \nu) = |F(\mu, \nu)|^2. \quad (16)$$

We note that $W(\mu, \nu)$, in an appropriate transformed coordinate system, is effectively the antenna beam pattern, but this is not relevant here as we are simply using the former as a convenient parameterisation of our problem. Specifying a target distribution $w(u, v)$ then corresponds to a constraint on $|F(\mu, \nu)|$ and, together with constraints on $f(x, y)$, this corresponds to a Fourier phase retrieval problem; a problem encountered in diffraction imaging and often solved using iterative projection algorithms.

The first step in the algorithm is to determine $f(x, y)$ given a desired, or target, $w(u, v)$, which is converted to $W(\mu, \nu)$, by solving the resulting phase problem with appropriate constraints on $f(x, y)$ and $|F(\mu, \nu)|$. The constraints on $f(x, y)$ are the terrain constraints, i.e. a support constraint where the pdf is zero, and positivity (since a pdf is positive). The constraint on $|F(\mu, \nu)|$ is the square root of the target $W(\mu, \nu)$. The corresponding projections are easily defined. The projection for $f(x, y)$ consists of setting the iterate to zero outside the support region and inside the support region where it is negative. The projection onto the Fourier magnitude constraint consists of Fourier transforming the current $f(x, y)$, to obtain $F(\mu, \nu)$, replacing $|F(\mu, \nu)|$ by $\sqrt{W(\mu, \nu)}$ and leaving the phase unchanged, and inverse Fourier transforming to obtain the updated $f(x, y)$. Starting with a random initial $f(x, y)$, one of the IPAs described above is then used to determine $f(x, y)$. Note that the Fourier magnitude constraint is non-convex so that simple algorithms such as the ER algorithm are not expected to be effective in finding the solution.

The second step in the algorithm is to generate a set of antenna positions by sampling $f(x, y)$. This is achieved using a simplified form of the Metropolis-Hastings (MH) algorithm for generating a sequence of samples from a

probability distribution. The antenna pdf is first discretised onto a grid (this can be the same or different to the grid size used above). The M antennas are first distributed randomly on the grid. As will be seen below, it is necessary to first calculate the total probability in each terrain region and then divide up the antennas as per these total probabilities and distribute this number in each region. The algorithm then systematically goes through each antenna and decides whether to move the antenna. Antennas for which all nearest neighbour sites are occupied by antennas are skipped. Let the antenna position be (x_0, y_0) . First select at random one of the nearest neighbours to (x_0, y_0) . If there is an antenna on this grid point then choose a new adjacent grid point at random. Repeat until an empty site is chosen. Denote the chosen site as (x_1, y_1) . The antenna is moved to (x_1, y_1) based on the following rule.

If $f(x_1, y_1) > f(x_0, y_0)$ then make the move.

If not, generate a random number α uniformly distributed on the unit interval, and if $f(x_1, y_1) / f(x_0, y_0) > \alpha$ then make the move, otherwise leave the antenna at its original position.

As the algorithm proceeds the antennas drift such that they eventually sample the distribution. The grid spacing enforces a minimum spacing between the antennas as is desired. The reason for the initial distribution of antennas in each terrain region is now apparent: the update rule will not allow antennas to migrate across regions of zero probability. The MH algorithm can be applied to each region independently or to all regions in parallel. The algorithm can be terminated based on monitoring a suitable metric on the distribution.

Results from a simulation using the above algorithm are shown in Fig. 3. A terrain support in the form of annulus is shown in Fig. 3a, which was discretised onto a 1024×1024 grid and zero-padded out to 2048×2048 . The target visibility pdf $w(u, v)$ is a circularly symmetric Gaussian as shown in Fig. 3b. The algorithm was started with a random pdf $f(x, y)$ as shown in Fig. 3c. The corresponding visibility is shown in Fig. 3d and although it has the same general shape as the target, it has too small values at small baselines. The DM algorithm was found to be effective in this case and the resulting pdf is shown in Fig. 3e and the resulting visibility in Fig. 3f. Although the resulting visibility is a little narrower than desired, its overall structure is a better match to the target than is Fig. 3d. A total of 200 antennas were randomly distributed in the support region in proportion to the total probabilities as described above and the MH algorithm applied. A sample set of antenna locations is shown in Fig. 3g, and the corresponding baselines in Fig. 3h. Overall, the algorithm has produced an antenna pdf that compensates for the “hole” in the support region.

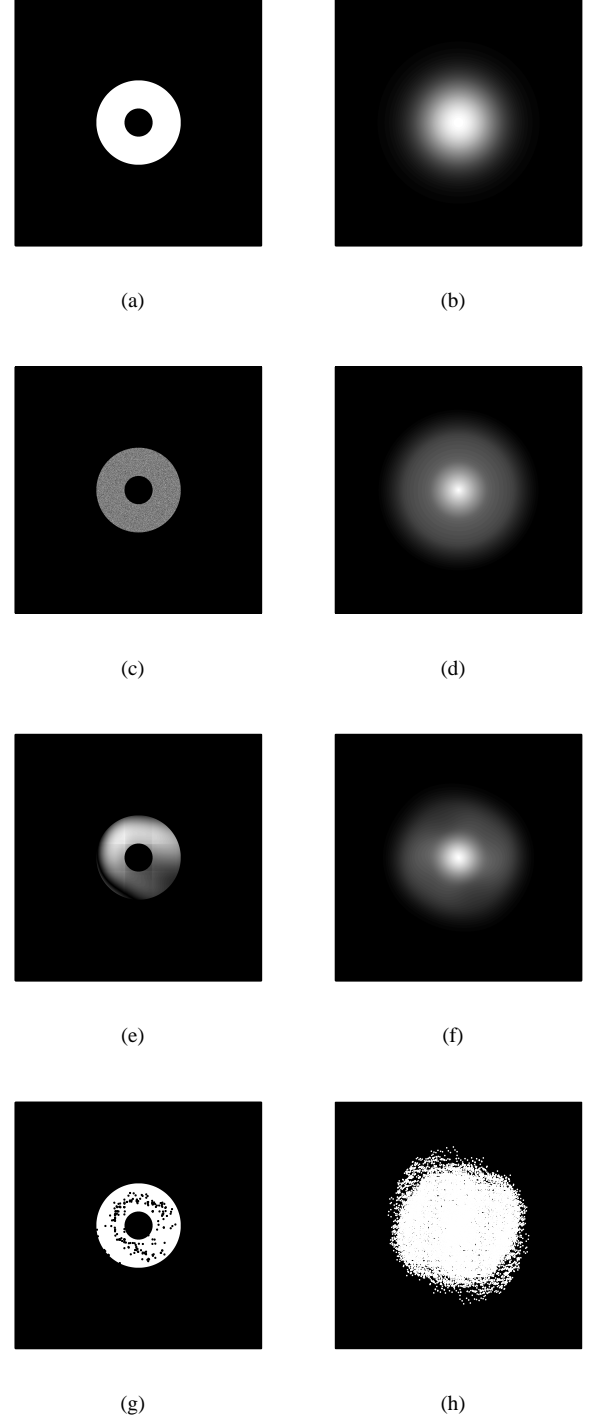


Figure 3: Simulation of the probabilistic method. (a) The terrain constraints and (b) the target visibility pdf. (c) The starting antenna pdf and (d) the corresponding visibility pdf. (e) The final antenna pdf and (f) the corresponding visibility pdf. (g) The final antenna locations (black dots) and (h) the final visibility samples (white dots).

6 CONCLUSIONS

Determination of optimal antenna locations for radio telescopes is an outstanding problem as there are numerous different constraints depending on particular observation scenarios. Iterative projection algorithms appear to offer a potentially useful global optimisation method for solving this kind of problem. For a large number of antennas, an algorithm based on an IPA and using a pdf to describe the antenna locations appears to have potential, as evidenced by application to a simple stylised problem. Further assessment of the utility of this approach will require application to more realistic scenarios.

A number of aspects of this problem require further investigation and possible extension. The sampling scheme used here treats the antenna positions as independent. There may be an advantage to introducing some dependency between the positions, to maintain a minimum spacing for example. An open problem is existence of a solution, i.e. given a particular terrain support and target visibility distribution, how well can the distribution of visibility samples match the target distribution? Optimal use of different IPAs and their convergence behaviour for these kinds of problems also deserves more investigation.

REFERENCES

- [1] A. Thompson, J. Moran, and G. Swenson, *Interferometry and Synthesis in Radio Astronomy*. Wiley, NY, 2nd ed., 2001.
- [2] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, "The Square Kilometre Array," in *Proceedings of the IEEE*, vol. 97, pp. 1482–1496, 2009.
- [3] M. Johnston-Hollitt, V. Kitaev, C. P. Hollitt, N. Jones, and T. C. Molteno, "Current New Zealand activities in radio astronomy: building capacity in engineering & science for the Square Kilometre Array," in *Proc. ENZ Con (16th Electronics NZ conference)*, pp. 107–112, 2009.
- [4] T. J. Cornwell, "A novel principle for optimization of the instantaneous Fourier plane coverage of correlation arrays," *IEEE Trans. Antenn. Propag.*, vol. 36, no. 8, pp. 1165–1167, 1988.
- [5] E. Keto, "The shapes of cross-correlation interferometers," *Astrophys. J.*, vol. 475, no. 2, pp. 843–852, 1997.
- [6] F. Boone, "Interferometric array design: Optimizing the locations of the antenna pads," *Astron. Astrophys.*, vol. 377, no. 1, pp. 368–376, 2001.
- [7] M. de Villiers, "Interferometric array layout design by tomographic projection," *Astron. Astrophys.*, vol. 469, no. 2, pp. 793–797, 2007.
- [8] N. Jin and Y. Rahmat-Samii, "Analysis and particle swarm optimization of correlator antenna arrays for radio astronomy applications," *IEEE Trans. Antenn. Propag.*, vol. 56, no. 5, pp. 1269–1279, 2008.
- [9] V. Elser, "Phase retrieval by iterated projections," *J. Opt. Soc. Am. A*, vol. 20, no. 1, pp. 40–55, 2003.
- [10] R. P. Millane, "Iterative projection algorithms for solving inverse problems," *Proc. OCEANS 2003, CD-ROM*, pp. 2714–2719, 2003.
- [11] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.
- [12] R. P. Millane and W. J. Stroud, "Reconstructing symmetric images from their undersampled Fourier intensities," *J. Opt. Soc. Am. A*, vol. 14, no. 3, pp. 568–579, 1997.
- [13] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A*, vol. 7, pp. 394–411, March 1990.

Cytometer

A Low Cost Approach Using an HD DVD Optical Pickup Head

Rajneel P. Dayal and Rainer Künnemeyer

School of Engineering, University of Waikato,
Private Bag 3105, Hamilton, New Zealand
rpd5@waikato.ac.nz

Abstract:

A flow cytometer is an instrument that measures microscopic particles as they flow in single file through a sensing region. In this paper we present a particle detector using a HD DVD optical pickup head. The system uses a micro flow cell that allows particles to be trapped inside. The reflected optical signal intensity is influenced by moving the trapped particles under the beam and is measured directly in the pickup. The advantage of using a HD DVD optical pickup head is not only the low cost optical system but also the ability to use the integrated magnetic actuators to adjust for alignment errors between the micro flow cell and the pickup. Measurements confirm that the edges of the micro flow cell are detected as well as $5\mu\text{m}$ beads trapped inside the channel.

Keywords:

Cytometer, HD, Blu-ray, optical pick-up.

1 INTRODUCTION

In optical flow cytometry characteristics of single biological cells are determined as light from a laser beam is scattered by the cells [1, 2]. The biological and physical parameters of the cell can be found from the properties of the scattered light. Optical flow cytometer detection systems comprise of five main parts: a light source (laser beam), a micro flow cell, a detector, an amplification system and a computer for analysis.

The principle of flow cytometry is that single particles suspended within a stream of fluid are interrogated individually in a very short time as they pass through a light source, usually laser light that is focused into a very small region. The measured optical signals are laser light scattered by the particles or fluorescence of the particles in response to the laser excitation. The fluorescence properties of cells are often used for labelling specific cells in a population.

In this paper we present a system that allows the optical detection of microscopic particles. This system is based on a low cost approach using a HD DVD optical pickup head. This optical pickup head contains an infrared laser of wavelength 780 nm which is used for CDs, a red laser of wavelength 650 nm which is used for DVDs and a violet laser of wavelength 405 nm which is used for Blu-ray or HD DVDs.

In a feasible study the optical beam was pointed to a plastic micro flow cell which contained trapped particles and the laser beam on the pickup was moved by an integrated magnetic actuator to scan the flow cell at a particular point. We present the results when the flow cell was moved along while being scanned to simulate particle movement.

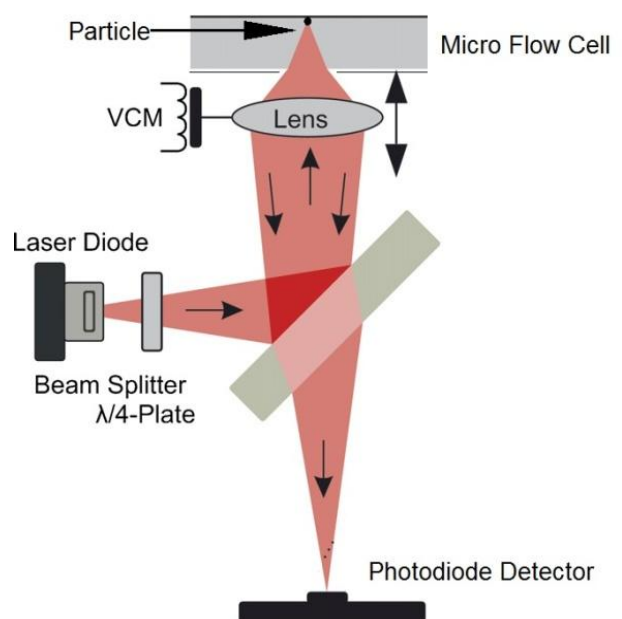


Figure 1: Principle of a DVD cytometer [3].

2 MATERIALS AND METHODS

The experimental setup is based around a HD DVD optical pickup head (PHR-803T) removed from an Xbox 360 gaming console. The main components of an optical pick-up are illustrated in Figure 1. A laser diode is used to generate an optical beam (405 nm) that passes a beam splitter and is focused to a small spot of $0.26\ \mu\text{m}$ size by the lens outside the micro flow cell. The lens position can be controlled by voice coil motors (VCMs) to scan across a flow channel and to focus into the channel. The micro flow cell was a Topas chip with 4 parallel $100\ \mu\text{m}$ by $100\ \mu\text{m}$ channels (Microfluidic ChipShop).

After back scatter the beam passes the beam splitter again and propagates to the photodiode detector fitted onto the pickup via a fibre optic cable. The original photo detector array has been replaced by this arrangement as no information was available on the photodiode array. Instead, a super blue enhanced (PIN 5DP/SB) photodiode detector was used as it was ideal for the violet wavelength (405 nm) laser beam.

The experimental setup is shown in Figure 2 where the micro flow cell is placed under the pickup head. $5\ \mu\text{m}$ Latex particles (Proscitech), suspended in water, were trapped inside the flow channel.

In this setup focusing is manually adjusted to get the biggest scatter peak and also the micro flow cell could be moved length wise under the laser beam. A program was written in LabVIEW to scan the tracking VCM at a set velocity across a flow channel. The micro flow cell was moved while being scanned and the back scattered light was collected and analyzed.

In this experiment the reflected intensity signal changes when a particle influences the optical path. Due to the high numerical aperture of the beam (0.85) it is important to have control over the position at which the particles hit the laser beam. In commercial cytometers a separate sheath flow positions the particles in the detection region. In our case sheath flow was not provided. Instead, particles were trapped, viewed under the microscope to mark the particle positions, and the region with the particles was scanned to get the particle scatter.

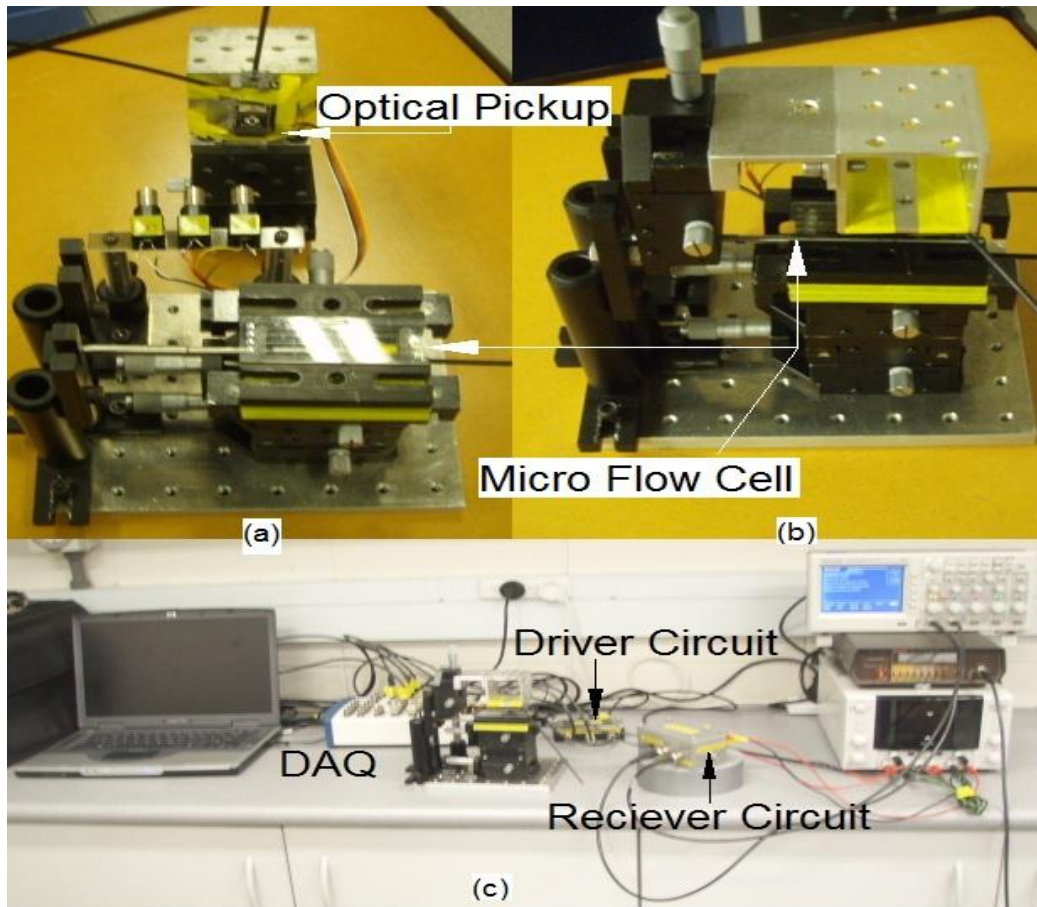


Figure 2: Experimental Setup. (a) side view of the disassembled cytometric setup. (b) top view of the assembled cytometric setup. (c) complete experimental setup.

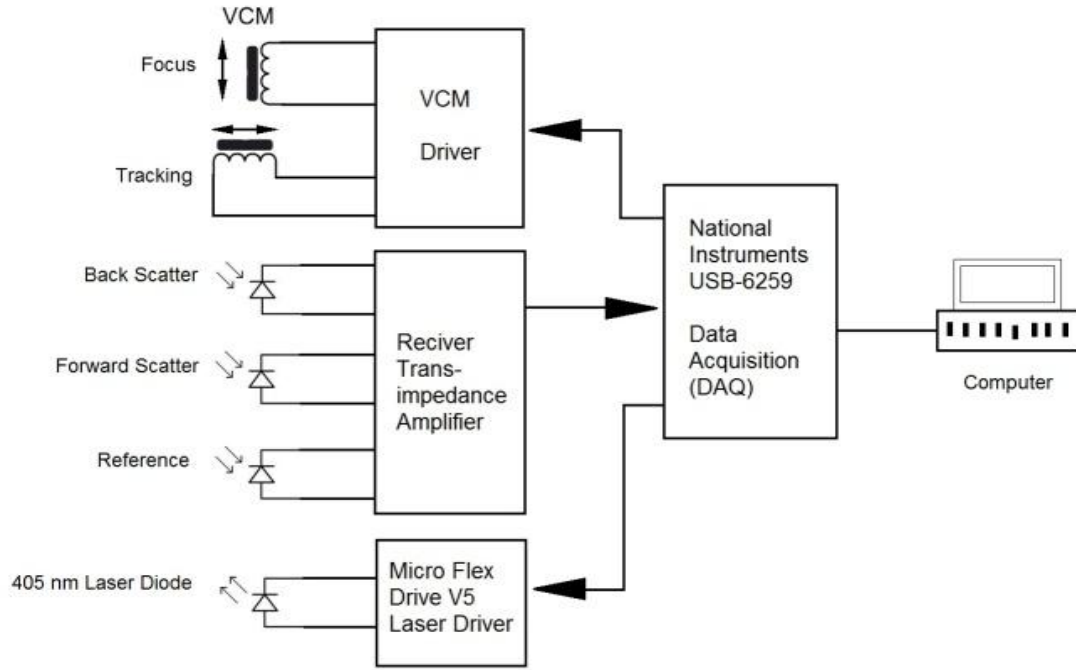


Figure 3: Electronic Setup. Custom electronics were developed to drive the optical pickup head as well as detect the scattered signal.

Figure 3 shows the electronic setup. It consists of driver circuits to operate the VCM and the laser diode of the optical pickup. It also has a receiver circuit that allows detection of back scatter, forward scatter, and reference signal which can be monitored individually as well as used to remove laser noise from the scattered signals. The National Instruments USB-6259 data acquisition module was the most appropriate one to use for this purpose. Data had to be sent by the DAQ to operate the optical pickup. The scattered signals had to be sent to the computer by the DAQ card for analysis.

The VCM driver circuit takes input from the DAQ of 0 V – 1.6 V which is created in LabVIEW, where both the focus and tracking coil experiences a range of -800 mV to +800 mV across them. To drive the 405 nm laser diode the Micro Flex drive V5 was used as this laser driver is a high-efficient switching current regulator that can be powered by battery combinations which result in a voltage of 2 – 5.5 V DC. The driver operates as a current regulated source, varying the output voltage above and below the battery voltage as necessary to meet the current output setting as there are five current output settings available in this driver.

The receiver circuit is a laser noise reduction circuit where the inputs are either back scatter and reference signal or forward scatter and reference signal depending on which is being looked at. The purpose of the circuit is to remove the laser variation from the scattered signal. This circuit consists of a differential pair, a transim-

pedance amplifier and an integrating servo amplifier that adjusts the ΔV_{BE} to force the DC output to be zero [4].

3 RESULTS AND DISCUSSION

The principle of operation of the DVD optical pickup has been shown with measurements on polystyrene beads in flow and on cells that were adhered on a mirror surface [2] or in a micro fluidic channel [5]. Now we show that measurements from HD DVD optical pickup can also be acquired.

Two areas were marked on the micro flow cell using a microscope where one had trapped 5 μm particles and the other was a particle free area. The two areas were scanned and the signals were acquired as shown in Figure 4. In both scans the micro flow cell's edges can be seen as the two big peaks. There are also two smaller peaks in the scan of the particle area. These peaks indicate particles which was verified by removing the particle free signal from the particle signal. The scatter signals from the walls correspond to a distance of about 100 μm . The scatter signal from the particles is then approximately 5 μm wide which agrees well with the nominal size of the particles.

The advantage of using a HD DVD optical pickup head is that a very precise system can be bought from the shelf for a very low cost. On the other hand, modifica-

tions to that system are hardly possible due to the casted package, which makes the system less flexible. Also, datasheet information is not readily available.

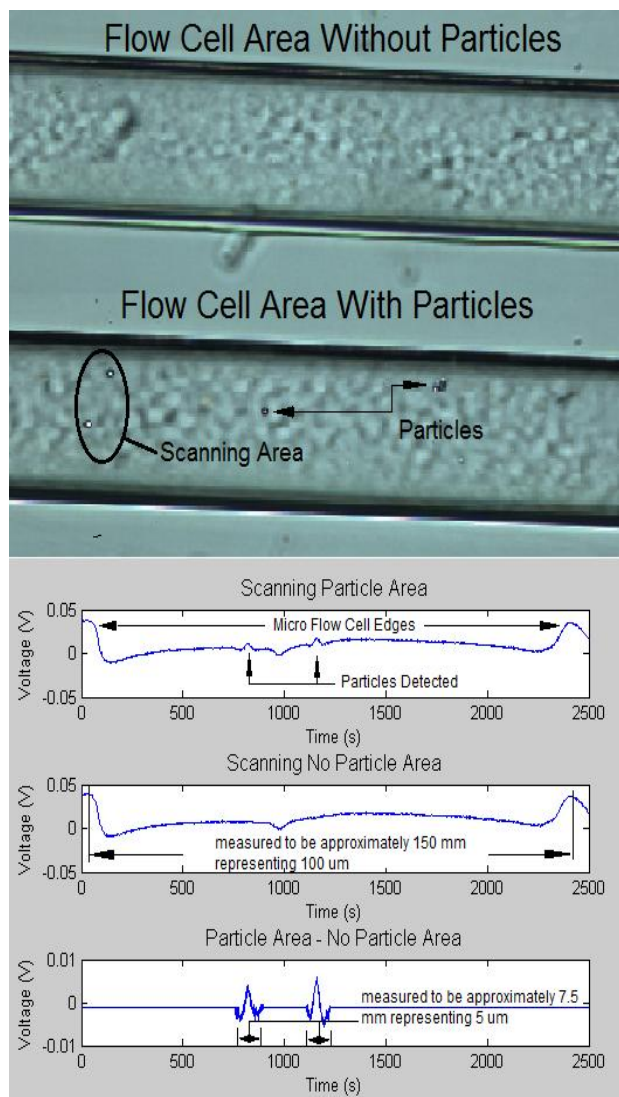


Figure 4: 5 µm Particle Detection. Two areas were marked; one with trapped particles and one without. Scanning was done over two areas and the signals analyzed. The bottom trace shows the difference signal.

4 CONCLUSIONS

We have successfully shown that particles can be detected by our cytometer setup utilizing a HD DVD optical pickup head. The instrument has sufficient sensitivity to detect 5 µm sized latex particles.

Our current research was focused on detecting particles with the aim to progress towards a cell counter. More work is required to realise a reliable system. As this optical pickup utilizes a violet laser, the work can be extended to fluorescence detection in the future.

5 ACKNOWLEDGEMENTS

The authors would like to thank the technicians and the 4th year students from the School of Engineering, The University of Waikato.

6 REFERENCES

- [1] Sharpe, J. C. "A new optical configuration for flow cytometric sorting of bovine spermatozoa by sex." PhD thesis, University of Waikato, Hamilton, New Zealand: 4-13, 1997.
- [2] Kostner, S. and Vellekoop, M. J. "Low cost cytometer based on a DVD pickup head." Eleventh International Conference on Miniaturized Systems for Chemistry and Life Sciences. Paris, FRANCE: 739-741, 2007.
- [3] Kostner, S. and Vellekoop, M. J. "Cell analysis in a microfluidic cytometer applying a DVD pickup head." Sensors and Actuators B (Chemical) 132(2): 512-17, 2008.
- [4] Hobbs, P. C. D. "Building Electro-Optical Systems. Making It All Work." 2nd edition. Wiley, Hoboken, New Jersey: 273-276, 2008.
- [5] Clow, A., Künnemeyer, R., Gaynor, P., and Sharpe, J. "Low cost optical particle detection for Lab on Chip systems based on DVD technology." Proc. SPIE 6799: 67990S1-9, 2008.

Crack detection of eggshells using self-mixing phase locked interferometer

Pawan Kumar Shrestha and Rainer Künnemeyer

The University of Waikato
Private Bag 3105, Hamilton 3240, New Zealand.
pawan@waikato.ac.nz

Abstract:

We present a method for detecting a crack on chicken eggs, using a phase locked self-mixing diode laser interferometer. The egg is excited using a single tone frequency, and its vibration is measured using a phase locked self-mixing diode laser interferometer. Vibration measured on or near the crack of the eggshell has a significantly higher number of harmonic components of the excitation frequency compared to the vibration measured on the intact part of the eggshell. Using white noise for excitation also results in different vibration spectra for the cracked or intact part of the eggshell.

Keywords:

Phase locked, self-mixing, modal testing, crack detection, egg shell.

1. INTRODUCTION

New Zealand's estimated production of eggs in 2009 was 972 million [1]. In the study of commercial hatcheries, Bell et al. [2] estimated about 4% of eggs are cracked prior to processing. Cracked eggs are prone to harmful bacteria like *Salmonella* and can pose a risk to the health of the consumer. Also, leaky eggs contaminate other eggs and mechanical parts of the egg packing and sorting plant. Egg quality in the egg sorting and packing industry is controlled by candling where eggs are examined against a bright light by a human. Candling is a slow process and the inspection is subjective. Therefore, the operation is a bottleneck in commercial egg-grading machines [3].

An automated sensor capable of differentiating between cracked and intact eggs is desirable to improve quality control and increase profitability of the egg producer. Automation of the process is approached mainly by two different techniques: Machine vision and mechanical testing.

Elster and Goodrum [4] proposed a machine vision technique to detect open cracks and hairline cracks in stationary eggs in 1991. A success rate of 96% was achieved. However, the average time required to classify each egg using this method was 25.3 s. Later Goodrum and Elster [5] extended their work to inspect continuously rotating eggs. An incandescent lamp is placed be-

low the egg and three images at an interval of 120° are taken. Then the images are processed to classify the egg as cracked or intact. With this method 94% success rate is achieved in determining the cracks. However, the software requires calibration. Using colour computer vision combined with artificial neural networks, Patel et al. [4] claim an average crack detection of 87.8% (i.e. detection of 84.4% of cracked eggs and 91.1% of intact eggs).

One of the drawbacks of the machine vision technique is the processing time of the computers. With increasing processing speed of modern computers, this technique is looking promising for detection of dirt on the eggshells, broken shells and odd shapes. However undamaged eggs with scratches or calcium deposits are often rejected [3].

In the other commonly employed technique, the mechanical behaviour of the eggshell is checked by exciting the eggshell using a small mechanical impactor. The integrity of the eggshell is examined by measuring the amplitude or the number of rebounds of an impactor on the eggshell. Several elastic rebounds indicate an intact eggshell. The rebounds of the impactor are damped near the crack of the eggshell as the elasticity of the eggshell is highly impaired near this location. Measuring rebounds from several places of the egg, a crack in the eggshell can be detected [3].

Instead of measuring rebounds of the impactor, Coucke [6] proposed to measure the frequency response of the

egg itself using a microphone. The authors report the detection of cracked eggs to be 90% and false positives below 1% [3].

We propose to extend the mechanical technique to a completely non-contact method, employing acoustic waves for the excitation of eggs and a phase locked self-mixing optical interferometer for the measurement of the response of the egg.

2 BACKGROUND THEORY

2.1 Self-mixing interferometry

The self-mixing effect in a laser occurs when some part of the light emitted by a laser re-enters the laser cavity. The phase shifted light back-scattered from a remote target mixes with the unperturbed source light generating a modulation of both the amplitude and the frequency of the lasing field [8].

The modulation period corresponds to a half-wavelength displacement ($\frac{\lambda_0}{2}$) of the target along the laser beam axis, where λ_0 is the wavelength of the laser without feedback [7]. The shape of the self-mixing signal depends on the feedback parameter, C , which is determined by the amount of feedback and the distance between the laser and the target. The shape of the self-mixing signal varies from the cosine shape for weak feedback ($C \ll 1$) to sawtooth like for strong feedback ($1 < C < 4.6$). For the very strong feedback regime ($C > 4.6$), the laser diode enters the mode-hopping regime and interferometric measurements are no longer possible [8].

The change in the spectral properties of the laser due to backscattered light coupling back into the laser cavity can be calculated from [10]:

$$(\omega_0 - \omega_c)\tau_{ext} = C \sin(\omega_c \tau_{ext} + \arctan \alpha), \quad (1)$$

where $\omega_0 = 2\pi\nu_0$ is the angular frequency of the laser without feedback, ν_0 is the optical frequency without feedback, ω_c is the angular frequency of the laser with feedback, τ_{ext} is the round trip delay time of the external cavity, and α is the linewidth enhancement factor. The optical output power with feedback, P_c , can be expressed as a function of the optical power of the unperturbed laser, P , as [7],

$$P_c = P[1 + m \cos(2\pi\nu_c \tau_{ext})], \quad (2)$$

where m is the modulation index.

2.2 Phase-locked self-mixing interferometer

By locking the interferometric phase to half a fringe a phase locked self-mixing interferometer can be achieved [9]. In the phase locked interferometer (or vibrometer), no signal processing is necessary to retrieve the target movement. This type of interferometer is capable of measuring displacements far less than half a laser wavelength.

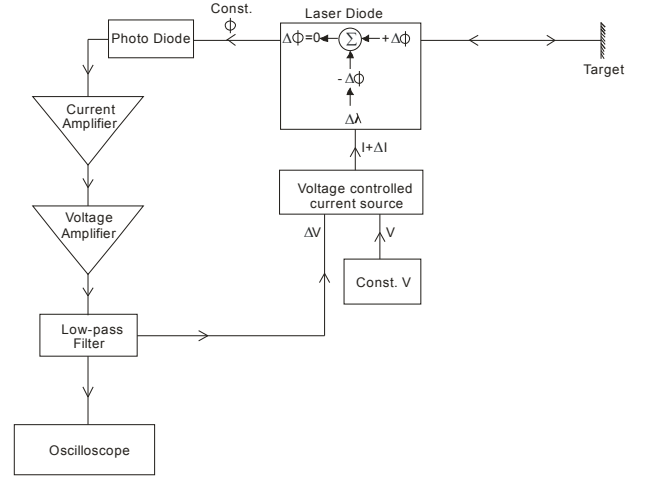


Figure 1: Phase locked self-mixing interferometer.

Feedback is used to lock the system to half an interference fringe. The feedback is adjusted so that the wavelength of the diode laser light is modulated to null out any change in phase [9], i.e.,

$$\Delta\lambda_0 = \frac{\lambda_c}{L_{ext0}} \Delta L_{ext}, \quad (3)$$

where λ_c is the wavelength of the laser with feedback, L_{ext0} is the distance from the laser diode to target at rest, and ΔL_{ext} is the target displacement.

Figure 1 shows the basic components of the phase locked self-mixing interferometer. Since wavelength and frequency are inversely proportional to each other, equation (3) also illustrates the dependence of the frequency of laser to the target displacement. Possible emission frequency of an unperturbed Fabry-Perot cavity laser can be written as [11],

$$\nu_0 = \frac{nc}{2L\mu_e}, \quad (4)$$

where $n = 0, 1, 2, \dots$, c is the speed of light, μ_e is the effective refractive index, and L is the distance between laser diode's facets. Equation (4) shows that the frequency of the laser is inversely proportional to the length of the cavity and the effective refractive index. If the temperature of the laser diode is maintained constant, the length of the laser diode cavity will remain constant. The refractive index then only depends on the carrier density. The change in laser frequency with carrier density can be written as [11],

$$(\nu_0 - \nu_{th}) = -\frac{\nu_{th}}{\bar{\mu}_e} \frac{\partial \mu_e}{\partial n} (n - n_{th}), \quad (5)$$

where, ν_{th} is the resonance frequency for $n = n_{th}$, n is the carrier density, n_{th} is the carrier density at the lasing threshold, and $\bar{\mu}_e$ is the effective group refractive index.

The diode laser frequency can be modulated by the injection current. The frequency change results in a change of phase in the self-mixing interferometer which can be used to compensate the change in phase introduced by the external target displacement. The output signal is directly proportional to the displacement of the target.

The main advantage of the self-mixing configuration is, it is very simple, compact (part count saving) and robust. Since this configuration has only one optical axis, no alignment is necessary.

3. EXPERIMENTAL SETUP

The phase locked self-mixing interferometer was constructed using a 780 nm, 10 mW laser diode (L780P010, Thorlabs, USA). The laser diode was operated at constant temperature of 25 °C ($\pm 0.2\%$) using a thermoelectric temperature controller (TEC2000, Thorlabs, USA) and a temperature-controlled laser diode mount (TCLDM9, Thorlabs, USA). The optical power (containing the vibration or velocity information) was detected by the monitor photodiode housed inside the laser diode package.

The target used to the test of the vibrometer was yellow insulation tape attached to a piezoelectric transducer driven by the sine wave at 40 Hz. Figure 2 and Figure 3 show the signal driving the target (upper trace) and the vibrometer signal (lower trace). The vibrometer signals are out of phase to the driving signal. The phase difference is introduced by the piezo transducer. The vibrometer setup was then calibrated by using the setup without feedback and counting the number of fringes (and multi-

plying them by $\frac{\lambda_0}{2}$) for the set of driving signal voltages. Assuming the piezoelectric oscillator's amplitude was changing linearly with the driving voltage, the target displacement was obtained to be 1 μm for 0.747 V.

As illustrated in Figure 2 and Figure 3 no signal processing is necessary to retrieve the target movement. The vibrometer constructed for this experiment has the dynamic range of 1 nm to 2.5 μm compared to 0.1 nm to 90 μm claimed by [9].

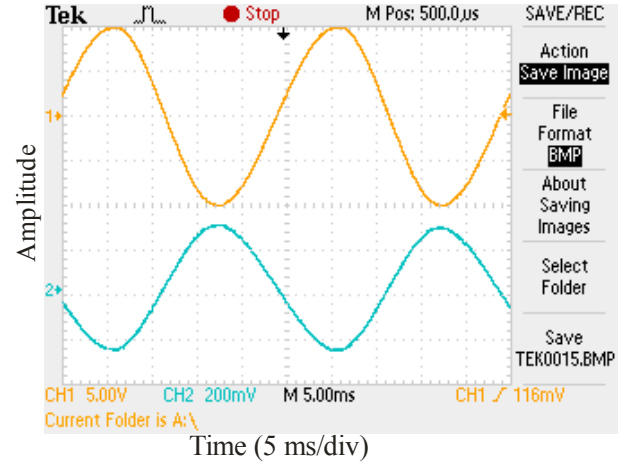


Figure 2: Upper trace: signal driving the target; Lower trace: output from the vibrometer.

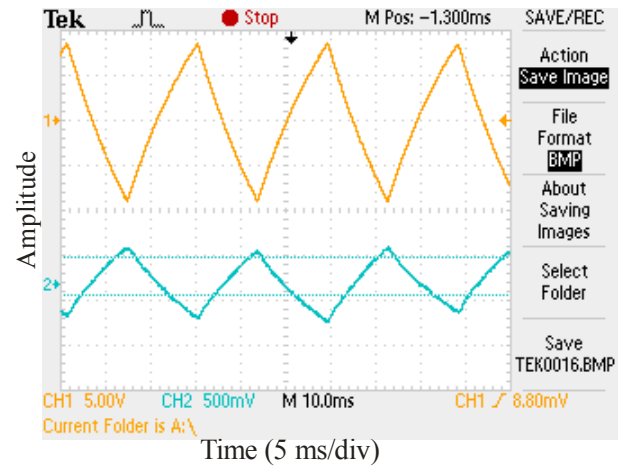


Figure 3: Upper trace: signal driving the target; Lower trace: output from the vibrometer.

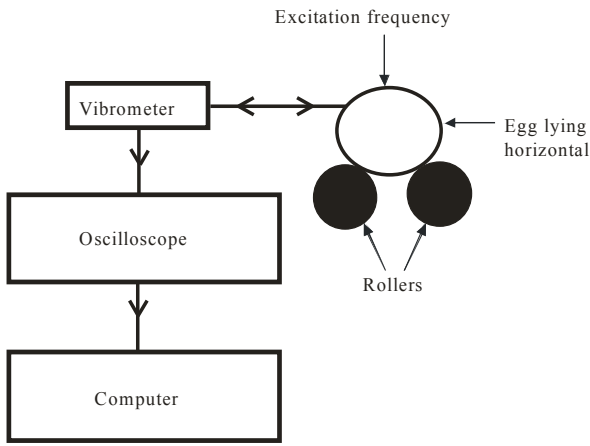


Figure 4: Block diagram of experimental setup.

Figure 4 depicts the basic block diagram of the experimental setup for crack detection of chicken eggshell. 10 fresh chicken eggs of size 6 (minimum weight of 53 g [1]) from the local supermarket were used in the experiment. 5 samples each of cracked and intact eggs were used. Hairline cracks (from a few μm to few mm in width and 25 mm to 40 mm in length) were artificially inflicted on the eggshell. The cracks were on or near the equator. Crack width was measured using a microscope (M38, Wild Heerbrugg, Switzerland), digital camera coupled to the microscope (Digital Sight DS-SM, Nikon, Japan) and the software (Image Pro Plus 6.3, Media Cybernetics, Inc., USA). The length of the crack was measured using vernier callipers. The response was determined near the crack of an egg.

The egg was placed horizontally on top of the roller. The surface of the roller was covered with 6 mm thick PVC foam. The stiffness of the egg platform was much lower than the stiffness of the intact eggshell. A loudspeaker connected to the function/arbitrary waveform generator (33220A, Agilent Technologies Inc., USA) via custom made amplifier was used to excite the egg. The distance between egg and vibrometer was set to be 15 cm and that between egg and the loudspeaker to about 5 cm.

Output of the vibrometer was visualised and recorded using an oscilloscope (TDS2004B, Tektronix, USA). A screen shot of the signal can be saved directly on USB flash drive, and a LabVIEW® VI was used to interface the TDS2004B with the computer. Data was analysed using a program written in LabVIEW®, and Matlab® was used for signal processing.

4. RESULTS AND DISCUSSION

4.1 Excitation by white noise

Egg samples were excited using white noise with the

vibrometer pointing on or close to the crack in the eggshell. Figure 5 shows the spectrum recorded by the vibrometer from an intact egg. A peak can be observed at 1.4 kHz. De Ketelaere et. al [3]. reported one dominant resonance peak on exciting an egg with an impactor and recording the response signal with a microphone. The authors report this peak to be highly repetitive when measuring on different locations around the equator of an egg and to vary from 3 to 6 kHz depending on mass of egg and shell stiffness [3].

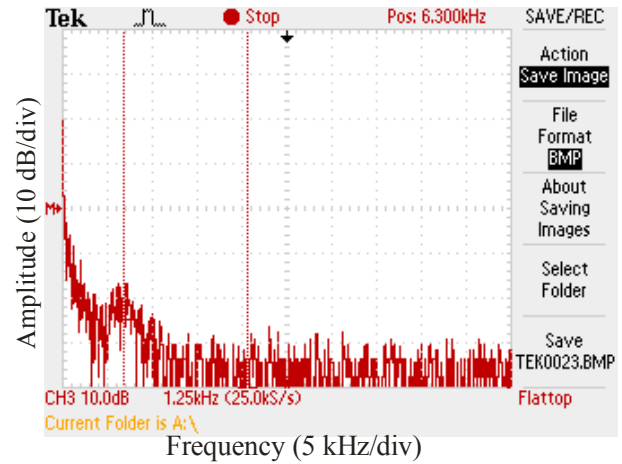


Figure 5: Intact egg excited by white noise.

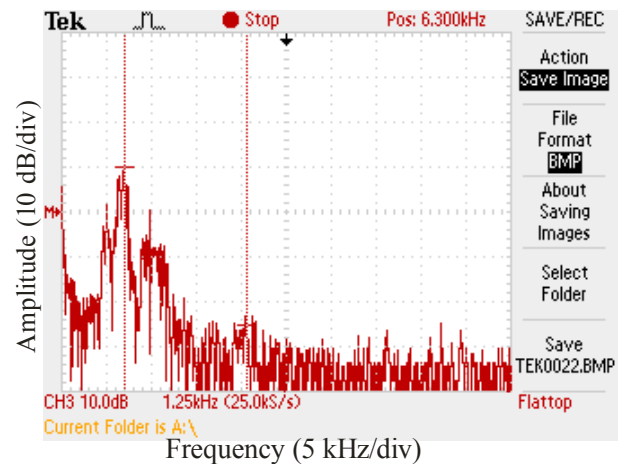


Figure 6: Cracked egg excited by white noise.

Figure 6 shows the response of a cracked egg to white noise excitation. The vibrometer was pointing directly at the crack. The power spectrum is different compared to the intact egg with dominant peaks at 1.4 kHz, 2.5 kHz and 5.2 kHz. De Ketelaere et. al. [3] also report heterogeneous spectra resulting from the cracked eggs. In similar experiments carried out by Wang et.al. [12], authors

report the frequency domain signal of the cracked egg near the crack to be different from that of an intact egg.

4.2 Excitation by single tone frequency

Since the sound energy is spread out in white noise, we applied a single tone to excite the eggs more efficiently. A couple of intact/cracked eggs were first excited using a single tone from 400 Hz to 13 kHz at 100 Hz interval step size to choose the best excitation frequency. As a result 1.35 kHz was selected as the excitation frequency.

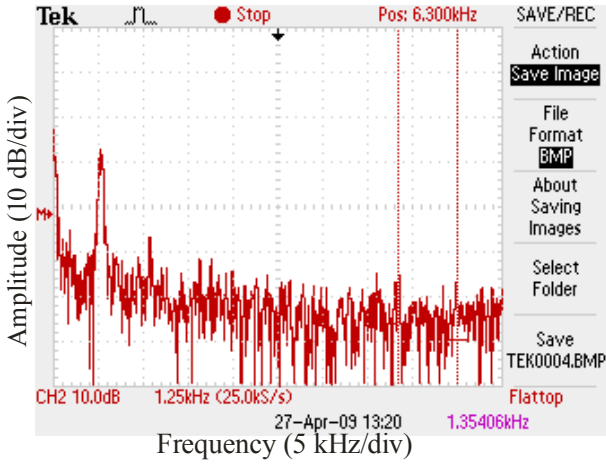


Figure 7: Frequency response of the intact eggshell.

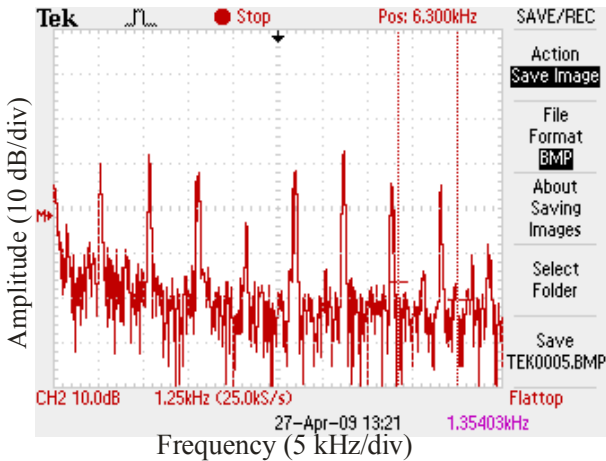


Figure 8: Frequency response of cracked eggshell

Figure 7 and Figure 8 show the frequency response when the intact egg and cracked egg respectively are excited using a single tone frequency of 1.35 kHz. Intact eggs were investigated with the vibrometer pointing at the eggshell; in the case of cracked eggs, data was collected by pointing the vibrometer directly at the crack. When the egg is excited using the single tone frequency, the

cracked eggshell produces far more harmonic components compared to intact eggshell. The number of harmonic component produced depends on the amplitude and frequency of excitation wave, size of the crack and distance between the laser spot of the vibrometer and the crack. The number of harmonic components decreases as the laser spot of the vibrometer moves away from the crack.

Data were recorded in the computer using oscilloscope and LabVIEW Virtual Instrument (VI). Another VI was written to read the stored data from the files and to compute an FFT of the signals and the cross correlation between spectra from various parts of the egg.

Let x be the first data set, y be the second data set and n be the number of points in the data set. The Pearson's correlation coefficient, r , is given by,

$$r = \frac{\sum z_x z_y}{n} \quad (6)$$

where $z_x = \frac{x - \bar{x}}{S_x}$ and $z_y = \frac{y - \bar{y}}{S_y}$ and

$$S_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \text{ and } S_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}.$$

Correlation coefficient, r , always is in the interval $[-1, 1]$. If r is $+1$, x and y have a complete positive correlation; if r is -1 , x and y have a complete negative correlation and if r is 0 , x and y have no correlation.

Five sets of data (D1, D2, D3, D4, D5) were collected for each egg by rotating the egg each time by 72° and taking measurement. r for ten different combinations from five data sets (i.e. (D1,D2), (D1,D3), (D1,D4), (D1,D5), (D2,D3), (D2,D4), (D2,D5), (D3,D4), (D3,D5), (D4,D5)) were computed, and the minimum value of r was compared with a threshold level.

For an intact egg, r is close to 1 whereas for a cracked egg, it is close to 0. If the lowest r obtained is less than the chosen threshold value, then the egg is classified as cracked. The threshold value was set to be 0.3 for this experiment. The results are shown in Table 1. Dataset A corresponds to intact egg samples and dataset B to cracked egg samples. The method was able to distinguish cracked eggs from the intact eggs.

Data set A			Data set B		
Egg	Min. r	Result	Egg	Min. r	Result
1	0.54	Intact	1	0.13	Cracked
2	0.78	Intact	2	0.14	Cracked
3	0.79	Intact	3	0.16	Cracked
4	0.55	Intact	4	0.11	Cracked
5	0.61	Intact	5	0.12	Cracked

Table 1: Classification of eggs by correlation coefficient and a threshold value of 0.3.

5. CONCLUSION

Modal testing is a common technique to establish differences in the resonant behaviour of mechanical structures that have and do not have flaws or cracks. We have presented a method to determine cracks in chicken egg shells. In our approach vibrations in the shell are excited by acoustic waves; the response of the shells is detected using a phase locked self mixing diode laser interferometer. To our knowledge this is the first time that a fully non-contact modal testing method has been applied to the assessment of chicken egg shells.

We were able to distinguish between cracked and intact shells when the egg was excited by acoustic white noise or a single tone, and the vibrometer was aimed at the crack. The signal deteriorated with distance from the hairline crack.

The cross correlation between vibration spectra recorded on various part of an egg indicated that cracks could be identified for the samples used in this proof of concept experiment.

There is more room to develop a faster algorithm. The threshold value to classify an egg as cracked can be optimised, and a statistical study should be carried out to prove the effectiveness of our system.

It is important to mention that even though the egg was rotated on the roller, it was kept stationary during the time of excitation and measurement. So the system needs to be improved to achieve real time crack detection of egg in motion.

6. REFERENCES

- [1] Egg Producers Federation of New Zealand. "Egg Quality," 22 August 2010; <http://www.eggfarmers.org.nz/egg-quality.asp>.
- [2] D. D. Bell, W. D. Weaver, and M. O. North, *Commercial chicken meat and egg production*, Norwell, Mass.: Kluwer Academic Publishers, 2002.
- [3] B. De Ketelaere, P. Coucke, and J. De Baerdemaeker, "Eggshell Crack Detection based on Acoustic Resonance Frequency Analysis," *Journal of Agricultural Engineering Research*, vol. 76, no. 2, pp. 157-163, 2000.
- [4] V. C. Patel, R. W. McClendon, and J. W. Goodrum, "Color computer vision and artificial neural networks for the detection of defects in poultry eggs," *Artificial Intelligence Review*, vol. 12, no. 1-3, pp. 163-76, 1998.
- [5] J. W. Goodrum, and R. T. Elster, "Machine vision for crack detection in rotating eggs," *Transactions of the American Society of Agricultural Engineers*, vol. 35, no. 4, pp. 1323-1328, 1992.
- [6] P. Coucke, "Assessment of some physical quality parameters of eggs based on vibration analysis," Thesis, Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen, Katholieke Univ. Leuven (Belgium), Leuven (Belgium), 1998.
- [7] T. Bosch, N. Servagent, and S. Donati, "Optical feedback interferometry for sensing application," *Optical Engineering*, vol. 40, no. 1, pp. 20-7, 2001.
- [8] G. Giuliani, M. Norgia, S. Donati *et al.*, "Laser diode self-mixing technique for sensing applications," *Journal of Optics A: Pure and Applied Optics*, vol. 4, no. 6, pp. 283-294, 2002.
- [9] G. Giuliani, S. Bozzi-Pietra, and S. Donati, "Self-mixing laser diode vibrometer," *Measurement Science & Technology*, vol. 14, no. 1, pp. 24-32, 2003.
- [10] S. Donati, G. Giuliani, and S. Merlo, "Laser diode feedback interferometer for measurement of displacements without ambiguity," *IEEE Journal of Quantum Electronics*, vol. 31, no. 1, pp. 113-119, 1995.
- [11] K. Petermann, *Laser diode modulation and noise*, Dordrecht, Netherlands: Kluwer, 1991.
- [12] J. Wang, and R. Jiang, "Eggshell crack detection by dynamic frequency analysis," *Zeitschrift für Lebensmittel Untersuchung und Forschung A*, vol. 221, no. 1-2, pp. 214-220, 2005.

Tree Disk Machine

Simon Knopp^{*}

^{*}Mechatronics Engineering,
University of Canterbury.
Email: sjk114@uclive.ac.nz

Dr Michael Hayes[†]

[†]Electrical and Computer Engineering,
University of Canterbury.
Email: michael.hayes@canterbury.ac.nz

Abstract: A second-generation machine was developed to scan the stiffness of disks of wood. The machine uses a computer vision algorithm to automatically identify and measure the disk's features. The disk is mounted on a stepper motor-driven X-Y table to provide accurate position control in any direction. The machine is faster, easier to operate, and more accurate than the first generation.

Keywords: Tree disk, computer vision, automation, forestry.

1 INTRODUCTION

This paper describes the University of Canterbury School of Forestry's Tree Disk machine. The machine scans cross-sectional disks cut from trees to measure the stiffness profile of the wood. The project has two main goals:

- To reduce operator workload and increase speed,
- To improve the accuracy of the acquired data.

1.1 Background

To measure the stiffness of the wood, the machine clamps two ultrasonic transducers on opposite faces of the disk. A *TreeTap* device connected to the transducers sends an ultrasonic pulse through the wood and records the time-of-flight. The transducers are mounted on pneumatic cylinders with linear encoders, so the thickness of the disk can be measured. From the time-of-flight and thickness measurements, and by assuming the density of the wood, the stiffness can be calculated. A scan consists of multiple measurements taken along eight radii from the centre to the bark. The foresters are interested in these data because stiffness and its variation throughout the log is a measure of wood quality.

1.2 Operator Workload

Operating the existing machine is a labour-intensive process:

1. The operator must accurately align the disk to laser cross-hairs and clamp it in place,

2. The machine scans from the centre of the disk to the edge of the bark,
3. The operator rotates the disk by 45° , realigns it to the cross-hairs and re-clamps it,
4. The process is repeated until eight radii have been scanned.

The operator is required to come back to the machine every 30–90 seconds, limiting their capacity to complete other work concurrently. The new control system automates the operation of the machine.

1.3 Accuracy of Data

The transducers are mounted on a 'C'-shaped arm, moved along rails by a pneumatic stepping cylinder. The cylinder moves in nominal 2 mm increments, but the distance is not known precisely. Since the machine scans the variation in stiffness across the radius of the disk, the accuracy of the data is limited by the accuracy to which the position of the transducers is known. Currently, position is calculated by dead-reckoning from the number of steps commanded, which introduces error due to the uncertainty in the step size. The arm also has a quadratic height error as it extends; this is compensated using a lookup table.

1.4 Project Scope

The purpose of this project was to redesign the machine's control system to provide the above improvements and to implement any hardware and software needed to realise the design. Section 2 describes the chosen design and its theoretical basis, followed by a discussion of the results in Section 3.

2 SYSTEM DESIGN

To automatically align the disk and scan it in multiple directions, several degrees of freedom are required. On the existing machine, the only automated degree of freedom is the linear motion of the scanning arm, with all other motion provided manually by the operator. Given the way that the disks are manually rotated between scans, initial concepts for moving the disk focussed on being able to rotate it. A mechanism of this type would require at least two additional degrees of freedom: one to move the disk side-to-side for alignment, and at least one to clamp and rotate it for scanning.

A simpler alternative would be to add a linear transverse axis to allow Cartesian motion. By coordinating the motion of the disk and the arm, the machine can scan in any direction. A linear encoder could be mounted on the scanning arm to allow closed-loop control and thus increase the accuracy of the data as required. Unfortunately, the existing machine has limitations inherent to its design that make such modifications impractical. The linear bearings that support the scanning arm are of low quality and sag as the arm extends. This misalignment means that a linear encoder or some such measuring device cannot be used on the arm due to their tight mounting tolerances. Additionally, the fixed step size of the arm's pneumatic cylinder means that the distance between points scanned at an angle will be different to those scanned along the axis of the arm. That asymmetry would make it difficult for the researchers to extract useful information from the raw data.

The extent of the modifications required to improve this machine was such that it was easier to design an entirely new machine.

2.1 Machine Design

The concept design for the new machine, Fig. 1, was done in conjunction with one of the School of Forestry's technicians. In this design, the transducers are fixed in place while the disk moves on an X-Y table. Such an arrangement allows the whole disk to rest on a table (with a hole for the transducers), ensuring that the disk is level at all times. The linear axes are driven by stepper motors geared down through toothed belts, providing fine-grained position control with high torque.

The pneumatic cylinders that carry the transducers were retained for the new machine. For the measurements to be consistent, it is important that the contact force between the transducers and the disk is the same for each measurement. The cylinders ensure that this is the case regardless of the thickness of the disk due to the nature of pneumatics.

The technician was responsible for the detailed design and construction of the machine, while the remainder of this project focussed on the control system.

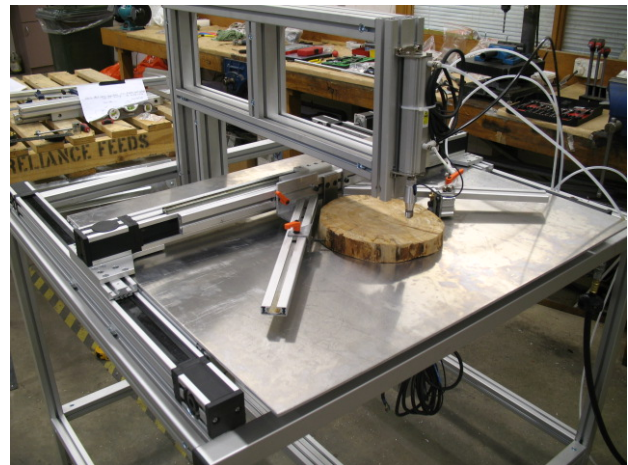


Figure 1: The new tree disk machine.

2.2 Hardware

Two key pieces of hardware make up the core of the machine's control system: a Linux PC and an embedded controller. All of the sensors and actuators interface to one of these devices.

The main I/O interface is an *Arcus Performax-2ED* stepper motor controller and driver. It features two microstepping motor drivers, quadrature encoder inputs, opto-isolated inputs for home and limit switches, and 16 general purpose digital inputs and outputs. The device is capable of functioning in both interactive and standalone modes—receiving commands from a host PC over USB or running programs from its internal flash. The Performax was chosen because it provides all of the required I/O capacity as well as the stepper motor drivers and encoder inputs. Using such a device allowed development effort to focus on system design rather than low-level hardware design.

The control system software runs on a Linux PC. The machine contains a USB hub so that the Performax controller, camera, and TreeTap can interface to the PC over a single cable. The machine is situated in a workshop, so the number of cables to the machine was kept to a minimum for safety and neatness.

2.3 Computer Vision

The computer vision software has the most influence on the autonomy of the tree disk machine. By using a camera, several features of the wood can be identified from a single input. A vision-based control system eliminates the need for the operator to accurately align the disk and can detect the disk boundary regardless of bark thickness. It also allows an image of the disk to be stored so that researchers can investigate anomalous data by visually inspecting for features such as knots and compression wood.

2.3.1 Algorithm Description

There are two major requirements for the machine’s vision software. Firstly, it must locate the pith—the centre of growth—because the scanning process starts at this point. Secondly, it must locate the boundary between the wood and the bark so that the machine can stop each radial scan at the correct point. In addition, it should be able to calibrate itself to counteract any small knocks that may cause misalignment.

Coordinate transformation In general, the bark boundary is assumed to be a closed contour of unknown shape surrounding the centre of the image. Since it is easier to search for approximately straight lines than unknown contours, the captured image is transformed to log-polar coordinates; that is,

$$(x, y)_{\text{dst}} = \left(\tan^{-1} \frac{y_{\text{src}}}{x_{\text{src}}}, M \log \sqrt{x_{\text{src}}^2 + y_{\text{src}}^2} \right) \quad (1)$$

where M is some constant scale factor. This has the effect of “unwrapping” the image about its centre as in Fig. 2b, mapping circles to straight lines.

Edge detection The transformed image is median filtered to smooth out fine details while retaining sharp edges. After converting to greyscale, this image is passed to a Canny edge detector [1]. The Canny thresholds were manually tuned over a sample of approximately one hundred images taken in different lighting conditions to produce complete edges with minimal noise. Since the identified edges often have small gaps that change from frame to frame, the process is repeated for several frames and the edges added to an accumulator. This process results in a binary image containing the bark boundary as well as some unwanted edges, Fig. 2c.

Bark boundary detection The third step is to extract the one important line, the bark boundary, from the edge image. To start the process, the left-most foreground pixel is located in the top row of the image. This point is equivalent to the innermost edge point in the usual rectangular coordinates. In the next row of the image, a window of interest is created on either side of the edge pixel from the previous row. The window is scanned for a foreground pixel and the process is repeated for the remaining rows of the image, Fig. 2d. If no foreground pixel is found in a row then the search window is widened for subsequent rows. Finally, the boundary line is transformed back to rectangular coordinates by the inverse of Eq. (1) to obtain Fig. 2e.

A check is included at this point to ensure that the returned contour is likely to be the required bark boundary.

In general, the convex hull of the bark boundary (the outline without any concavities) should be approximately elliptical. To check this condition, the area enclosed by the contour is compared to the area of its bounding box; if the contour area is less than $0.7 \times$ the bounding box area then the contour is rejected. The deciding factor of 0.7 is slightly less than the ratio of the area of an ellipse to the area of its bounding rectangle.

Pith location The pith is usually near the geometric centre of the disk. Although the disk’s outline may be an arbitrary shape, the innermost growth rings are always a good approximation to a circle. By considering a region of interest in the centre of the disk’s bounding box that is large enough to include at least one complete ring, the pith can be defined as the centre of a circular feature in this region. A Hough transform for circle detection [2], [3] is used to locate a circular feature, and the centre of the circle is then a close approximation to the pith, Fig. 2f.

The relationship between image coordinates and physical coordinates was established through offline calibration, so at this point all of the necessary information has been retrieved. The location of the pith and therefore the starting point of each radial scan is known, and the distance from the pith to the bark boundary in each of the scan directions can be calculated.

2.4 Software

The control system software is written mostly in Python, a high-level language that allows fast development. Rather than modifying the code for the existing machine, the software was written from scratch; given the completely redesigned control system there was not much scope for code reuse.

2.4.1 Other Languages

The Camera class was written in C++ and makes extensive use of image processing functions from the free OpenCV library¹. While OpenCV does have a Python interface, it does not expose the Hough circles function necessary for locating the pith, so the C++ interface had to be used instead. SWIG, the Simplified Wrapper and Interface Generator, was used to compile the C++ class as a shared object and create the necessary bindings to use it in Python.

SWIG was also used to wrap the Performax device driver. A higher level bridge was written to hide the device’s proprietary command syntax, providing intuitive methods such as `performax.move(100,200)` instead of `driver.SendRecv(handle, 'X100:Y200')`.

Finally, a simple program was written in the Performax’s own language to run on the device. It acts as a

¹<http://opencv.willowgarage.com/wiki/>

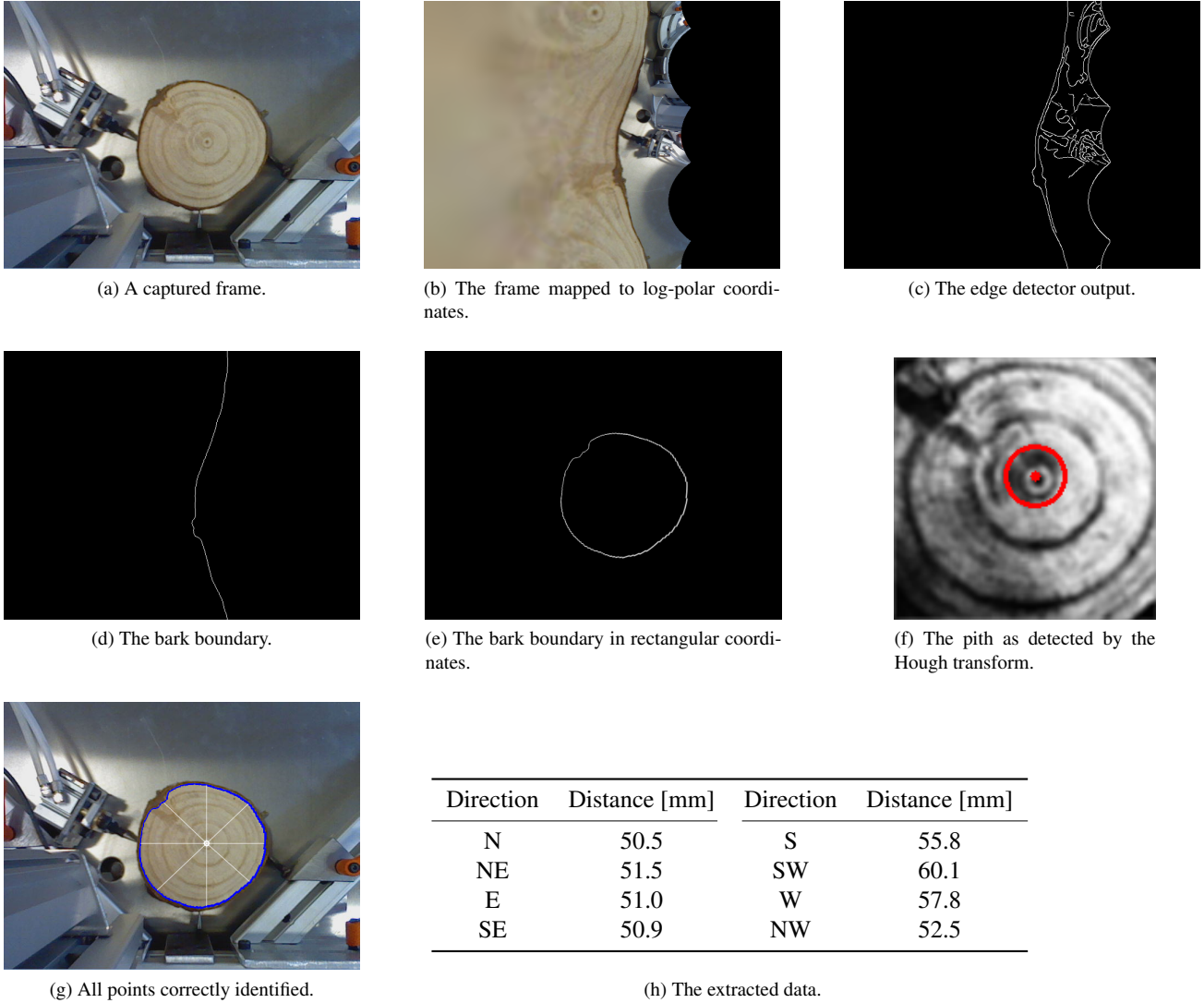


Figure 2: A step-by-step view of the computer vision algorithm.

buffer for the physical user interface, monitoring for button presses and setting corresponding flags. This design means that the buttons can be continuously polled without using any of the CPU or USB resources on the host computer.

2.4.2 Software Architecture

The main control loop calls `TreeDiskController.run()` repeatedly until it returns false. The `TreeDiskController` implements a state pattern [4] to change the functionality of its `run()` method at runtime. Each state has a `run()` method and optional setup, teardown, and transition methods. State transitions can be triggered by button presses or by the completion of the current state's `run()` method. This model ensures that the underlying hardware is in the same

physical state whenever the software is in a particular logical state.

Similarly, the `TreeView` class implements a strategy pattern [4]. The class contains a list of available strategies which it can use to identify points in an image. Initially, it uses a `Camera` object (implementing the `TreeViewStrategy` interface) to attempt to locate the points automatically, as described in Section 2.3.1. If that strategy fails, the `TreeView` class moves on to the `InteractiveTreeView` strategy. This class pops up a window displaying the captured frame and allows the operator to mark points by clicking on the image. Other strategies could be implemented too, but these two are sufficient to ensure that the key points will always be found, usually without operator intervention.

The main control loop is single-threaded for simplicity. Because of the standalone program running on the Performax device, though, the machine as a whole constitutes a

simple distributed system, bringing with it the dangers of concurrent programming. It is possible that the PC could attempt to read or clear a “button pressed” flag on the device at the same time as the device tries to write to it. In reality a collision such as this would be rare, and the worst-case outcome would be that a button press is not picked up by the PC. The buttons serve no safety-critical role, so this is an issue that can safely be ignored.

3 RESULTS

3.1 Computer Vision Algorithm

Several variations of the computer vision algorithm were implemented with varying degrees of success. To evaluate the algorithm’s performance, a set of 65 sample images was created. The samples included images of 22 different disks taken under a variety of lighting conditions: fluorescent, incandescent, LED, and DC incandescent. The disks were of varying size and surface finish and were placed in positions throughout the camera’s field of view. For each sample, the true locations of the pith and bark edge points were manually located and compared to the output of the algorithm.

3.1.1 Pith Identification

Figure 3 shows the distribution of errors between the calculated location and the true location of the pith. The pith was located with a mean error of 4.4 pixels, equivalent to about 1.2 mm. This accuracy is comparable to a human operator manually aligning the disk to laser crosshairs, as was done with the existing machine. 84% of the pith locations were found to within 2 mm (6 pixels) and 96% were found to within 4 mm (13 pixels). The errors in the upper

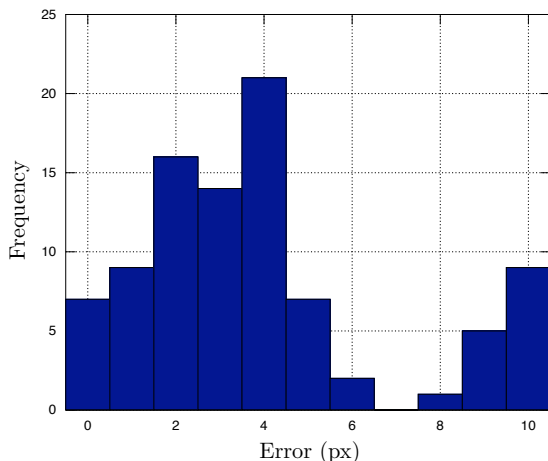


Figure 3: The distribution of pith location errors.

tail of the distribution occurred in images of small disks,

which have fewer circular features, and disks with very rough surfaces. Interestingly, neither knots or writing on the disk face posed a problem.

3.1.2 Bark Edge Detection

Initially, bark edge detection was done by passing the image through a Canny edge detector and searching for edge pixels near the disk’s bounding box. Several frames were accumulated to try and close any small gaps, but this was not always successful. When a gap in the edge image occurred along a scan direction, the bark edge was detected incorrectly. Such occurrences were rare, with 97% of the 520 edges in the sample set identified to within one pixel of the correct location. The errors occurred most commonly when the sample was strongly lit; using a more diffuse light source softened the shadows and gave more reliable edge detection. Although this method found most edge points exactly, the errored points covered a wide range because if the correct edge is not found the next closest one could be anywhere.

The bark edge detection algorithm described in Section 2.3.1 is an improvement on this approach. By converting the image to polar coordinate and searching for a line, it is easy to use linear interpolation to ensure that a continuous edge is returned. The interpolation provides a good-enough approximation to correct most of the 3% of points that the simpler approach missed. This change to the algorithm was made at a late stage, and so has not been as rigorously tested as the simpler variant. Since it is merely an improvement to the algorithm, though, it is assumed that it performs at least as well as the better-tested one. This certainly appears to be the case from informal testing.

3.1.3 Other Performance Parameters

The computer vision system as a whole was implemented for only the cost of a generic consumer webcam; the PC from the existing machine was reused for the new one. The camera is capable of capturing up to 30 frames per second at a resolution of 640×480 pixels, which is plenty for the application. Its automatic exposure and white balance adjustments allow for successful operation in changing lighting conditions. A variety of light sources were tested at a range of angles, intensities, and colours. Over all of these combinations, the only significant variable was whether the light produced shadows or glare. The workshop that the machine is installed in uses standard overhead fluorescent tube lighting which is sufficient to provide a well-lit scene with no sharp shadows. The machine is placed between rows of overhead lights rather than directly underneath them to avoid glare on the smooth aluminium table.

The algorithm takes approximately 0.5 s to capture eight frames and identify the pith and the bark edge points. If any of the checks built into the algorithm fail, it starts again

up to a maximum of ten times. This means that the operator will almost never have to wait longer than a few seconds after pressing the 'Run' button before receiving confirmation that the disk has been successfully identified. In the worst case when all attempts fail, the operator will have to wait for approximately 5 s for the ten scans to complete, and then manually mark the points on the computer. Using the strategy pattern minimises operator interaction time by usually working automatically, while maximising robustness by providing a fall-back to manual point identification.

The camera's field of view is approximately $50^\circ \times 40^\circ$, giving a viewable area of 330×250 mm on the machine's table. The machine is specified to scan disks up to 600 mm diameter, so in those cases the whole disk will not be visible in the frame. There is an option in the software's configuration file that disables the computer vision system and instead allows the operator to mark points by physically moving the disk. The operator uses the jog toggle switches to move the disk and presses a button on the control panel to mark the current location as a key point. This is a much slower process than the automatic method, but is only required occasionally since the majority of disks are less than 250 mm diameter.

3.2 Usability

The machine has a physical user interface that non-technical users can easily understand. Most operator interaction is with a control panel. Two buttons allow the operator to begin and stop a scan, and the third allows them to mark key points when using the manual mode. Two LEDs indicate whether the machine is idle or in the process of scanning. A foot switch engages and releases the clamp that holds the disk in place. The technicians asked for this physical user interface because they prefer interacting with machines rather than computers.

Ideally, the operator will only have to interact with the computer once, to start the software, and the rest of the time it just sits in the background. If the computer vision algorithm fails to find the key points then the operator does have to use the computer to mark them manually, but the advantages of this method outweigh the technicians' aversion to computers.

The control system can be configured through a well documented text file. This file is used to set such options as which radii are scanned and the step size between measurements. A graphical interface to these options was offered to the technicians, but they were happy to edit the text file as necessary.

3.3 Speed

The usual sequence of events for scanning a disk is as follows: place the disk on the table between the points of the clamp, tap the foot switch to close the clamp, press

run, wait to check that all of the key points are found, and then wait or go away until the disk has finished scanning. The process of loading a disk and beginning a scan usually takes about 20 s compared to 30 s for the existing machine. If a disk is a substantially different size to the previous one, the points of the clamp must be moved to accommodate it, adding another 15–20 s. In this case the disk may also have to be moved using the jog switches to locate it under the camera. If the computer vision algorithm fails to identify the key points then marking the points manually adds another 30 s or thereabouts. The greatest length of time required of the operator is therefore one period of approximately 70 s as opposed to eight periods of approximately 30 s for the existing machine.

As well as reducing the setup time, the scanning process runs faster on the new machine. The previous machine moved one step and made a measurement every 1.8 s, while the new machine does it in 1.3 s.

4 CONCLUSION

This project has produced a machine that addresses the shortcomings of the previous version. The mechanical design enables motion with high torque and high position accuracy. Scans can be made in any direction with any step size, providing the forestry researchers with the capability to gather as much data as they wish.

The machine is very easy to operate. The operator only needs to place a disk in the jaws of the clamp and press a button to begin the scanning process; the rest of the machine's operation is autonomous. The total time required of the operator is 20–70 s per disk down from eight periods of 30 s per disk for the previous machine. The computer vision system is capable of locating the necessary features of a disk with comparable accuracy to a human operator.

REFERENCES

- [1] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, pp. 679–698, Nov 1986.
- [2] J. Illingworth and J. Kittler, "A survey of the Hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87–116, 1988.
- [3] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*, ch. 6: Image Transforms, pp. 144–192. O'Reilly Media, Inc., 2008.
- [4] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns*. Boston: Addison-Wesley, 1995.

A Flexible, Low-Power, Programmable Unsupervised Neural Network Based on Microcontrollers for Medical Applications

Rafał Długosz^{1,2}, Tomasz Talaśka², Paweł Przedwojski², Paweł Dmochowski³

¹Institute of Microtechnology,
Swiss Federal Institute of Technology, Lausanne,
Neuchâtel, Switzerland
rafal.dlugosz@epfl.ch

²Faculty of Telecommunication and Electrical Engineering
University of Technology and Life Sciences (UTP)
Bydgoszcz, Poland
talaska@utp.edu.pl, pawel.przedwojski@wp.pl

³Victoria University of Wellington
School of Engineering and Computer Science
Wellington, New Zealand
pawel.dmochowski@vuw.ac.nz

Abstract:

We present an implementation and laboratory tests of a winner takes all (WTA) artificial neural network (NN) on two microcontrollers (μC) with the ARM Cortex M3 and the AVR cores. The prospective application of this device is in wireless body sensor network (WBSN) in an on-line analysis of electrocardiograph (ECG) and electromyograph (EMG) biomedical signals. The proposed device will be used as a base station in the WBSN, acquiring and analysing the signals from the sensors placed on the human body. The proposed system is equipped with an analog-to-digital converter (ADC), and allows for multi-channel acquisition of analog signals, preprocessing (filtering) and further analysis.

Keywords:

WTA network, microcontrollers, wireless sensor networks, medical applications

1. INTRODUCTION

Artificial Neural Networks (NN) are commonly used in tasks requiring processing, classification and recognize of "difficult" signals, such as, for example, heuristic data and non-stationary signals. They find applications in medical health care, telecommunication, and various other electrical engineering areas. In literature one can find various implementation techniques of various NNs, both the software- and the hardware-based [1, 2, 3].

Taking into account such criteria as energy consumption and calculation capacity, *fully-custom* designed networks are the most efficient solutions. Such networks enable parallel data processing, and thus are faster than their software-based counterparts. The circuits, which are designed on transistor level, enable a good matching of the internal structure to a given task, and therefore usually consume much less energy [4, 5]. On the other hand, they require relatively complex and time consuming design process, which is expensive, especially in case of realization of short series.

In this paper we present an implementation of the Winner Takes All (WTA) NN using an alternative approach that is based on microcontrollers (μCs). Such

networks are significantly less efficient than those realized in the *full-custom* style, but are up to ten times more efficient compared to similar NNs realized on PC [5]. Using the Eagle 5.6 environment, we developed a prototype testing board with two μCs , namely the 8-bits AVR and the 32-bits ARM CortexM3. The device is described in details in Section 3.

The motivation behind the realization of this device was manifold. The main purpose was to realize a flexible, programmable and convenient to use system for the application in various medical healthcare areas. The device is to be used as a base station in the Wireless Body Sensor Network (WBSN) for an on-line monitoring of patients. For this purpose, future versions will be equipped with the wireless I/O module. The WBSNs become more and more popular in recent years and there promises to be more interest in the future [6].

An interesting aspect is using the NNs implemented in WBSNs in the analysis of various biomedical signals, such as electrocardiograph (ECG) and electromyograph (EMG) signals. In [7] it was demonstrated that NNs are a very efficient tool in the analysis of such signals. Three different learning rules were investigated, specifically the self organizing map (SOM), the back-propagation

(BP) NN and the learning vector quantization (LVQ) NN. The unsupervised SOM algorithm was shown to be more efficient than, for example, the BP NN, although the learning process takes more time in this case.

The other reason behind realization of this device was to build a system that could be used as a hardware model of the analog and mixed (analog-digital) WTA NNs realized by the authors as Application Specific Integrated Circuits (ASIC) in the CMOS technology [5, 8]. As mentioned above, the overall design process of such networks is expensive and time consuming. The authors' experience shows that modeling of such chips using only Matlab or C++ is not always sufficient, especially in case when analog signals are to be processed and analysed. For this reason the proposed device is equipped with the ADC and the DACs that enable multi-channel processing of analog signals in the real time. Transistor-level designed NNs can be tested using the Spice environment, but simulations of even relatively short times require hours on a typical PC, which is the bottleneck in optimization of such chips. The proposed platform enables fast reprogramming and thus optimization of even large WTA NN in a short time.

It is worth mentioning that *fully-custom* designed NNs enable realization of the WBSN of different types. In the first approach, in which the NN is implemented in the base station, the sensors can be very simple. Their role is reduced in this case to data collection, and some preprocessing, such as data compression. In an alternative approach the ultra-low power unsupervised NNs can be used directly in particular sensors. In this case data exchange with the base station is reduced to the necessary minimum. This approach enables significant increasing the battery life time, as the communication usually consumes more than 90% of total energy consumed in such systems.

In the literature one can find the attempts to realize various NNs using μ Cs and microprocessors (μ Ps) [9, 10]. An example realization by use of the Single Instruction Multiple Data (SIMD) processor has been described in [9], where different methods of detection of the winning neuron were studied. In general, the SIMD μ Ps are suitable for large and fast networks with even hundreds neurons. The main reason for this is relatively large power dissipation, as well as large cost of a single device. Another realization of a multi-layer network using the PIC18F45J10 μ C has been described in [10]. In this case the maximum of 256 weights (connections) can be realized, which in practice means about 50 neurons.

NNs realized on μ Cs find the application in different areas, mostly in control and in diagnostics. For example, a device described in [11] has been used as an intelligent wireless electronic nose node (WENN) used in classification and quantification of binary gas mixtures NH₃ and H₂S. A NN described in [12] has been used to control temperature of a furnace.

A very important aspect in hardware realized NNs is the complexity of the learning algorithm. It has the influence on the power dissipation and achievable data rate.

Microcontrollers are rather suitable for simple arithmetic operations. For example, the unsupervised trained WTA NN requires only simple operations, such as multiplications, summations and subtractions. For the comparison, the NN described in [10] requires *tanh* activation functions, relatively more difficult in hardware realization.

2. BASICS OF THE WTA NETWORKS

In this section we outline the fundamentals of the WTA learning algorithm. WTA NN [13] belongs to the group of networks trained without the supervision, making them relatively fast, which is important in applications such as telecommunications [14, 15]. The training process relies on presenting the NN with learning patterns, X , in order to make the neurons' weight vectors, W , resemble presented data. For each new pattern the network calculates the distance between the X and the W vectors in all neurons. Different measures of the similarity between both vectors can be found in the literature. One of them is the Euclidean distance (L2) defined as:

$$d(X, W_i) = \sqrt{\sum_{l=1}^n (x_l - w_{il})^2} \quad (1)$$

Another popular measure is the Manhattan (L1) distance, defined as:

$$d(X, W_i) = \sum_{l=1}^n |x_l - w_{il}| \quad (2)$$

In the L1 measure the squaring and rooting operations have been omitted, which allows for simplification of the learning algorithm. Both these measures have been implemented in the proposed device for the comparison. The adaptation of the winning neuron in the t^{th} iteration is performed in accordance with the formula:

$$W_i(t+1) = W_i(t) + \eta \cdot (X(t) - W_i(t)) \quad (3)$$

where η is the learning rate. Other neurons in the network that lose the competition remain unchanged.

A significant problem encountered in the WTA networks are the, so-called, dead neurons i.e. the neurons that take part in the competition but never win and their weights remain unchanged. One of the reasons for this problem are badly selected initial values of the weights [16]. Such neurons reduce the number of classes that can be discriminated, thus increasing the mapping (quantization) error of the network. For this reason reducing the number of dead neurons is an important objective. One of the very efficient methods in this task is using the conscience mechanism [5, 17]. Its role is to increase the likelihood of winning the competition for all neurons in the NN. For this reason, the conscience mechanism has been implemented in the proposed device. This mechanism can be described by:

$$d_{\text{cons}}(X, W) = d_{\text{L1/L2 norm}}(X, W) + L_{\text{count}} \cdot K \quad (4)$$

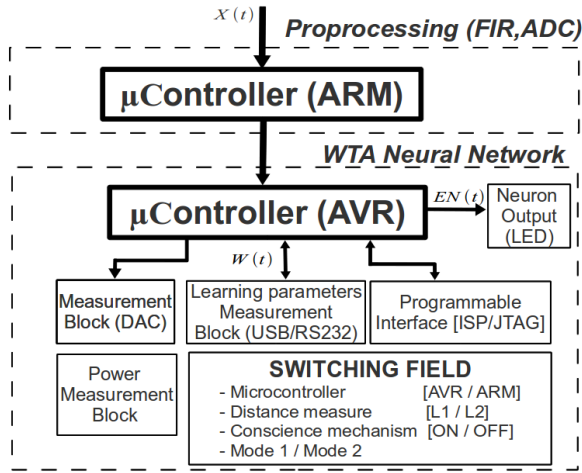


Figure 1. The proposed testing board with the programmable WTA NN based on two microcontrollers

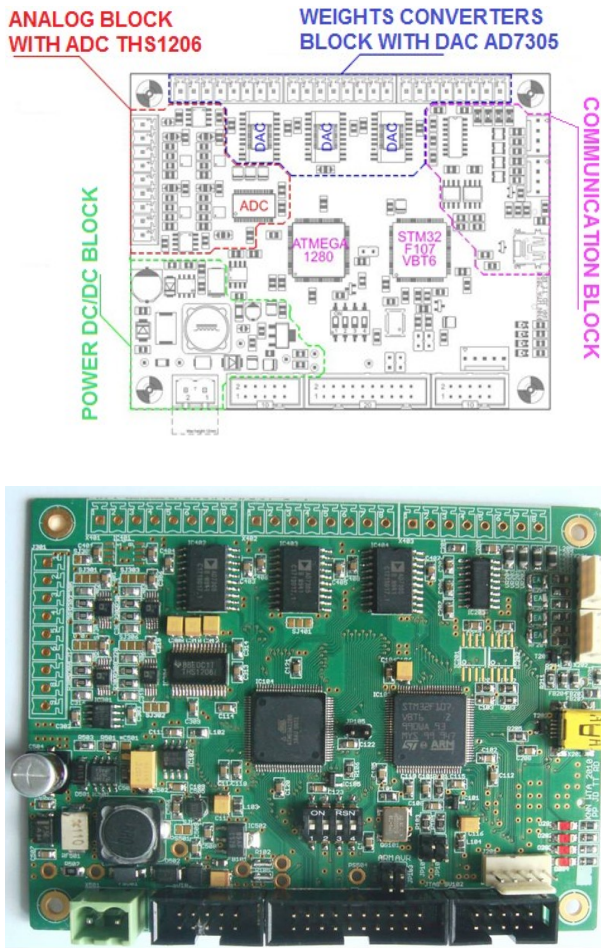


Figure 2. The proposed testing board of the programmable WTA neural network based on microcontrollers

The real distance $d_{L1/L2}(X, W)$ between the W and the X vectors is increased by adding a signal that is proportional to the number of the wins for a given neuron. Finally, in detection of the winning neurons a modified distance $d_{\text{cons}}(X, W)$ is being used. The L_{count} parameter is the number of the wins of a given neuron. The K coefficient is the gain factor that allows for controlling and optimizing the learning process by adjusting the strength of the conscience mechanism.

3. THE PROPOSED DEVICE

An overview of the realized system with two μCs is shown in Fig. 1, while the layout and the photograph of the testing board shown in Fig. 2. It is composed of several important blocks. One of them is the interface block, mentioned in Section 1, composed of the ADC and the DAC circuits that enable multi-channel data processing. A single 4-channel THS1206 ADC and three 4-channel AD7305 DACs have been used. This 12-bits ADC converts the analog input learning signals X , while the DACs enable a direct observation of selected neuron weights, w , on the oscilloscope. A direct observation is possible, for example, for 3 inputs and 4 outputs corresponding to 12 neuron weights. The higher number of the weights can be implemented, but in this case only selected weights can be observed directly, while the others can be viewed on PC. The interface block enables sending all weights, as digital signals to PC for a more detailed analysis. Digital signals are acquired throughout the USB and RS232 serial ports. The serial ports allow also for acquiring the learning signals X (in digital form), the calculated distances, d , between the X and the W vectors, the numbers of the wins of particular neurons and the quantization error for detailed analysis of the network performance. The numbers of the wins enable creation of statistics, which is very useful in many applications.

The overall device has been realized in a manner to enable on-line measuring of the power dissipation, separately for the ADC, the DAC blocks and for both μCs .

The core blocks are the μCs that are programmed by the use of the ISP/JTAG interfaces. The μCs can operate in different modes. In the first mode particular μCs work separately, performing the learning algorithm of the NN. In this case only one μC is being used, while the second one is turned off to reduce the power dissipation. This mode enables a direct comparison of both devices in terms of the attainable data rate and power dissipation. In this mode the active μC receives data from the ADC or directly the external digital signals. In the second case the ADC is turned-off to save energy.

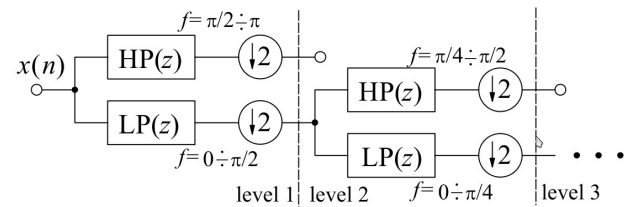


Figure 3. Cascaded discrete wavelet transform (DWT) realized using the QMF filter bank with the FIR filters

In the second mode, which is currently tested and optimized, both μCs are used at the same time, connected in series. The first μC (ARM) is in this case used as a signal preprocessing / conditioning block, while the second one (AVR) performs the learning algorithm. Using the ARM μC as a first block is necessary, as data preprocessing is usually more

complex than the subsequent WTA learning algorithm, requiring more computing resources. The used Cortex μ C is a 32-bits device, which is more convenient in data preprocessing (mostly filtering) than the 8-bits AVR μ C.

Table 1. An example realization of the wavelet Daubechies (Db10) transfer functions with reduced precision

LP theor.	LP round	LP binary	HP theor.	HP round	HP binary	HPx round
-0.0076	-1	1000001	-0.0189	-2	1000010	-2
0.0010	0	0000000	0.1331	15	0001111	17
0.0026	3	0000011	-0.3728	-41	1101001	-48
-0.0208	-2	1000010	0.4868	53	0110111	62
-0.0505	-6	1000110	-0.1988	-22	1010110	-25
0.0658	7	0000111	-0.1767	-19	1010011	-23
0.0901	10	0001010	0.1386	15	0001111	18
-0.1386	-15	1001111	0.0901	10	0001010	12
-0.1767	-19	1010011	-0.0658	-7	1000111	-8
0.1988	22	0010110	-0.0505	-6	1000110	-6
0.4868	53	0110111	0.0208	2	0000010	3
0.3728	41	0101001	0.0026	3	0000011	3
0.1331	15	0001111	0.001	0	0000000	0
0.0189	2	0000010	-0.0076	-1	1000001	-1

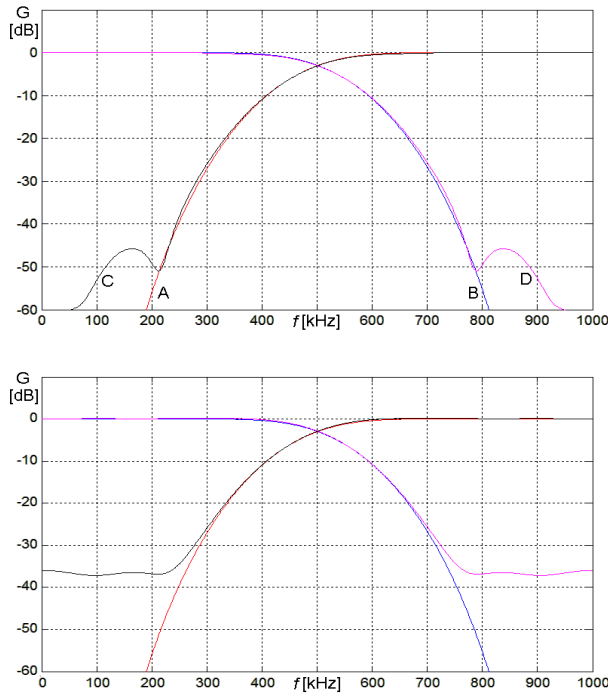


Figure 4. Frequency responses of the LP (B, D) and the HP (A, C) Daubechies (Db10) filters used in the DWT. The upper plot is for the optimal rounding, while the lower for not optimal rounding (HPx) of the filter coefficients.

Both of the above modes make the realized device a powerful autonomous system, suitable for WBSN. The second mode is important in case of the analysis of the ECG/EMG biomedical signals. Such signals must be first de-noised, and then some characteristic points must be extracted from the complexes [18].

Data preprocessing is based on the Finite Impulse Response (FIR) filtering. The Infinite Impulse Response (IIR) filters were also considered, but the FIR filters offer the attractive property of linear phase response. Data extraction is performed by means of a multistage discrete wavelet transform (DWT), shown schematically

in Figure 3. DWT is a series of the filtering operations performed by use of the quadrature mirror filter (QMF) bank composed of the lowpass, $LP(z)$, and the highpass, $HP(z)$, halfband FIR filters. Each stage is followed by decimation by a factor of 2.

The problem which must be considered at this stage is data resolution. The analog input signals X are converted into 8 – 12 bit digital form. As mentioned above, the proposed device is also used as the hardware model of the system that will be realized as an ASIC. In such implementations each bit significantly increases the number of transistors (and consequently, the chip area), and the power dissipation. For this reason data resolution has been limited to 16 bits only, and thus the FIR filter coefficients are rounded to 6 bits + 1 bit representing the sign, as shown in Table 1 for an example of Daubechies wavelet. In this situation the rounding operation must be performed very carefully to avoid the loss of the dynamic range of the filter, as shown in Figure 4. The non-optimal HPx case in Table 1 is shown at the bottom diagram. In this case the rounding factor is even less restrictive, but the loss of the attenuation exceeds 10 dB.

Data preprocessing is required as it simplifies the input signals provided to the NN, thus making the analysis performed by this network feasible. In particular, it minimizes the number of required network inputs, as particular features are provided to separate inputs. In [19] the ECG signals were decomposed into only four features, specifically, the span of the QRS wave, the interval of the R-R segment, the voltage and the slope of the S-T segment. The results obtained in [19] are in the good agreement with the diagnosis made by the medical staff. The analysis in [19] was performed using the WTA network. The results for the WTA NN were compared with the results obtained in case of using of the BP network. This conclusion is important, as the WTA NN is much simpler in the hardware realization than the BP algorithm.

Summarizing, the realized device enables operation in different modes, with either the L1 or the L2 distance measures, with or without the conscience mechanism, with different numbers of neurons, with the analog or digital input data. The system is still being developed and optimized. One of the modules that will be added in the next version is the wireless I/O module, to enable the application of the device in the WBSN.

4. LABORATORY TESTS

One of the important tests was to determine the maximum achievable data rate. This parameter depends on the number of neurons in the network, since all calculations inside the μ C are performed serially. This is a disadvantage in comparison with the NNs realized as ASIC, in which a fully parallel data processing can be easily implemented. The input data rate can be estimated by use of the following equation:

$$f_{\text{data}}(n) = f_{\text{max}} / (N_O \cdot n) \quad (6)$$

In (6) f_{\max} is the maximum clock frequency of a given μC , equal to 16 MHz and 72 MHz for the AVR and the ARM μCs , respectively. The N_0 parameter is the number of the clock cycles per a single input pattern X , in the NN with n neurons. The N_0 parameter differs for the AVR and the ARM μC . The ARM μC is more efficient, so although f_{\max} is in this case only 5 times larger, the NN operates more than 7 times faster. The data rate depends on the distance measure, and is approximately two times larger for the L1 case, as no squaring / rooting operations are required in this case. This is shown in Figure 5, as a function of the number of neurons.

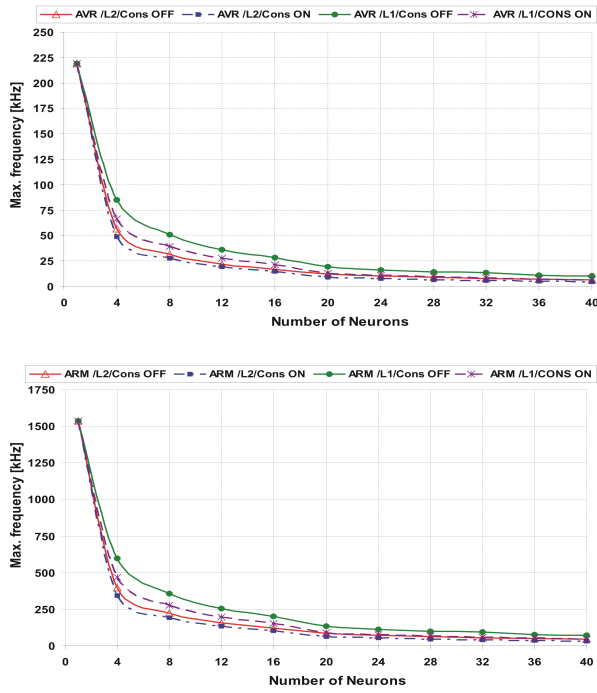


Figure 5. Achievable data rate of the WTA NN as a function of the number of neurons for both μCs for (top) the AVR, and (bottom) the ARM microcontroller.

Fig. 6 presents selected measurement results of the NN with 3 inputs and 4 outputs realized on the ARM μC sampled at 135 kHz. An example adaptation process has been presented for two selected neurons. The EN is the output signal of the WTA block that determines the winning neuron, i.e. the neuron the most resembling the input pattern X . Only this neuron can adapt its weights. As can be observed, when the EN signal becomes a logical '1', the weights of the corresponding neuron are modified. The following input signals were provided to the NN:

- x_1 – the sine signal with the frequency of 5 kHz;
- x_2 – the triangular signal with the frequency of 10 kHz
- x_3 – the rectangular signal with the frequency of 2 kHz.

Figure 7 illustrates the influence of the conscience mechanism on the learning quality of the network. In this case data acquired throughout the serial port for the NN with 10 neurons are shown. The conscience mechanism is able to activate the neurons that otherwise would remain inactive. Figure 7 illustrates the influence of dead neurons on the mapping properties. In case (a) all neurons are active, becoming representatives of particular data classes. Case

(b) is shown for the “too weak” conscience mechanism. In this case the number of dead neurons is smaller than in (a), but is not zero. In the worst case (c) only several neurons took part in the competition.

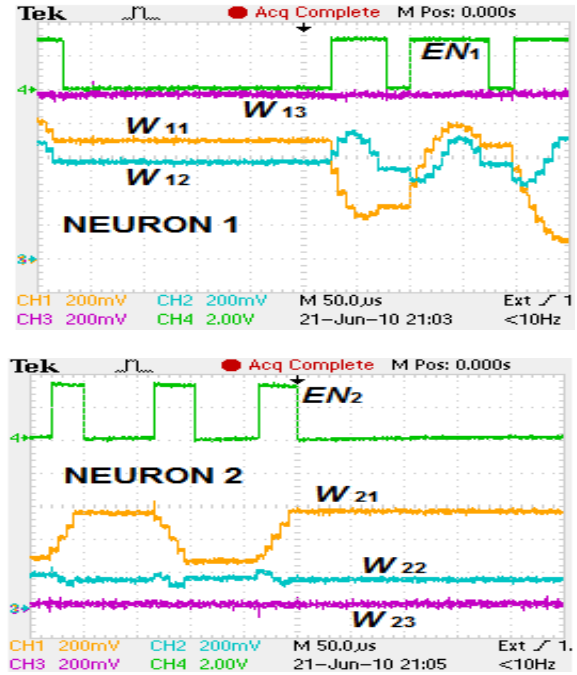


Figure 6. Example measurement results of the WTA NN with 3 inputs and 4 outputs (4 neurons). The diagrams present selected neuron weights, w , as analog signals.

An interesting aspect is the comparison of the NN realized on μC with the analog network realized earlier by the authors in the CMOS 180nm technology. The realized device consumes an average power of 300 mW and 500 mW, for the AVR and the ARM μC , respectively. For the comparison, the analog WTA network with 12 weights, sampled at 1 MHz, dissipated the power of 700 μW , i.e. approximately 500 times less than in case of the realization on μCs . Taking into account that the sampling frequency is now two times smaller, the analog NN are 1000 times more efficient.

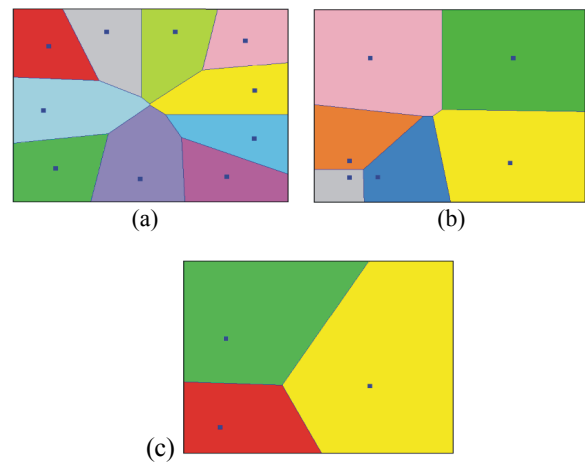


Figure 7. Voronoi's diagrams illustrating the influence of the conscience mechanism on the final placement of neurons. Depending on the strenght of this mechanism the number of dead neurons varies in-between 0 and 70%. Te results are shown for the NN with 10 neurons and digital weights transferred to PC.

5. CONCLUSION AND FURTHER WORK

A new implementation of the unsupervised trained Winner Takes All neural network (WTA NN) on micro-controllers (μ C) with the AVR and the ARM cores has been presented. We realized a prototype testing board with both μ C operating either separately or cooperatively. In the second case the ARM μ C is used in data preprocessing / conditioning, which relies on the finite impulse response (FIR) filtering and extraction of useful information from the input signals. The output signals of the ARM μ C become training signals provided to the AVR μ C that in this case performs the WTA classification algorithm. The prospective application of this device is in the Wireless Body Sensor Network (WBSN) in an on-line analysis of the biomedical ECG and the EMG signals.

The measurement results show that in the comparison with the same NN realized on PC, the network realized on μ C is even 10 times more efficient taking into account such parameters as the achievable data rate and the power dissipation. On the other hand, such network is even 1000 times less efficient than the same network realized as analog chip. The advantage of the proposed realization is relatively low cost of a single device and small sizes.

In future work, the board will be equipped with the wireless I/O ports, enabling applications in the wireless medical diagnostics systems, serving as a base station acquiring biomedical signals from the sensors.

6. REFERENCES

- [1] Macq, D. *et al.*, "Analog Implementation of a Kohonen Map with On-Chip Learning", *IEEE Transactions on Neural Networks*, Vol. 4, No. 3, May 1993, pp. 456-461
- [2] M. Holler *et al.*, "An electrically trainable artificial neural network (ETANN) with 10240 'floating gate' synapses", *International Joint Conference on Neural Networks*, 1989, pp. 191-196
- [3] J. Choi *et al.*, "A programmable analog VLSI neural network processor for communication receivers", *IEEE Transactions on Neural Networks*, Vol. 4, No. 3, 1993 r.
- [4] Dlugosz R., Talaska T., "Low power current-mode binary-tree asynchronous Min/Max circuit", *Microelectronics Journal*, Vol. 41, no 1, Jan. 2010, pp. 64-73
- [5] R. Dlugosz *et al.*, "Realization of the Conscience Mechanism in CMOS Implementation of Winner-Takes-All Self-Organizing Neural Networks", *IEEE Transactions on Neural Networks*, Vol. 21, No. 6, 2010, pp. 961-971
- [6] C.C. Enz *et al.*, "WiseNET: An Ultra low-Power Wireless Sensor Network Solution", *Computer*, August 2004, Vol. 37, No. 8
- [7] Lin He *et al.*, "Recognition of ECG Patterns Using Artificial Neural Network", *International Conference on Intelligent Systems Design and Applications (ISDA)*, Vol. 2, 2006, pp. 477-481.
- [8] Dlugosz R., *et al.*, "Programmable triangular neighborhood functions of Kohonen Self-Organizing Maps realized in CMOS technology", *European Symposium On Artificial Neural Networks (ESANN)*, Bruges, Belgium 2010, pp.529-534
- [9] B. Mailachalama and T. Srikanthan, "Area-time issues in the VLSI implementation of self organizing map neural networks", *Microprocessors and Microsystems*, Vol. 26, no. 9-10, 2002, pp. 399-406.
- [10] N.J. Cotton *et al.*, "A Neural Network Implementation on an Inexpensive Eight Bit Microcontroller", *International Conference on Intelligent Engineering Systems*, USA, 2008, pp. 109-114.
- [11] Young Wung Kim *et al.*, "An Intelligent Wireless Electronic Nose Node for Monitoring Gas Mixtures Using Neuro-Fuzzy Networks Implemented on a Microcontroller", *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, Italy, 2007, pp.100-104.
- [12] H. M. Mousa *et al.*, "On-Line Neurocontroller Based on Microcontrollers", *IEEE International Conference on Industrial Technology*, Hong-Kong, 2005, pp. 1252-1256.
- [13] T. Kohonen, *Self-Organizing Maps*, Springer Verlag, Berlin, 2001.
- [14] C. Amerijckx *et al.*, "Image Compression by Self-Organized Kohonen Map," *IEEE Transactions on Neural Networks*, Vol. 9, No. 3, 1998, pp. 503-507.
- [15] C. Chang *et al.*, "New Adaptive Color Quantization Method Based on Self-Organizing Maps," *IEEE Transactions on Neural Networks*, Vol. 16, No. 1, 2005, pp. 237-249.
- [16] T. Talaska and R. Długosz, "Initialization mechanism in Kohonen neural network implemented in CMOS technology", *European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, 2008, pp. 337-342.
- [17] D. DeSieno, "Adding a conscience to competitive learning", *IEEE Conference on Neural Network*, Vol. 1, 1988, pp. 117-124.
- [18] Paul S Addison P. S., "Wavelet transforms and the ECG: a review", *Physiological Measurement*, Vol. 26. No. 5, 2005
- [19] Chen Tian-hua *et al.*, "The Sorting Method of ECG Signals Based on Neural Network", *International Conference on Bioinformatics and Biomedical Engineering*, 2008, pp. 543 - 546.

Humanoid Service Robot

Peter Barlow

Dr Gourab Sen Gupta

Sanjay David

Massey University
School of Engineering and Advanced Technology
pwebarlow@gmail.com, g.sengupta@massey.ac.nz, sanjaydav@gmail.com

Abstract: This article focuses mainly on the mechanical design of a low cost humanoid robot, with a brief background of sensors used for maintaining balance. In order to produce a low-cost humanoid with similar flexibility to a high end humanoid a dual push rod actuated joint articulation was implemented. To further increase human like similarities a toe joint is implemented. This gives the humanoid the ability to stride more elegantly, increase speed control, and reduce energy used for each step. Torque tests show that the serial controlled WCK servo should be able to produce sufficient torque when equipped with a full metal gearbox. An 'Attitude Heading Reference System' is used to monitor the pelvis roll, pitch and yaw, along with an inclinometer in the body. Four sensors in each foot measure pressure which is converted through on-board processing to hexadecimal serial data. This monitors pressure on each corner of each foot. All the sensor information is sent to a RB110 micro computer which is used for controlling the humanoid gait through Visual Basic.NET programming language.

Keywords: Humanoid, biped, gait process, gyroscope, inclinometer, micro computer, roboard RB110, WCK 1111 servo

1 INTRODUCTION

This project focuses on designing and building a biped humanoid robot to be used as a service robot, concentrating on the lower body, the legs, and the hips, making it possible for the robot to stand and walk. In the long run this robot will be used to help disabled people perform daily tasks that they would otherwise have trouble with.

At present there are two main areas of production, very high end humanoids, and very low end humanoids. There are very few middle range humanoids. The reason for this is at present, using a classic humanoid joint design, joint actuators are hugely expensive. To achieve a middle price range product with the flexibility and performance of higher price range products, new and revolutionary joint actuation must be implemented.

2 CURRENT DESIGN ISSUES

In order for an electric actuator to produce similar torque to a human muscle the motors must be very large. A large motor is obviously heavy which means the corresponding motors in the biped not only have to be strong enough to achieve necessary joint torques, they must also be able to lift their own weight. This produces a vicious cycle. For example, a motor used in the ankle increases rotational inertia in the hip, so the bigger the motor in the ankle, the bigger the motor in the hip.

Since human muscles are tremendously powerful for their size it makes mechanical representation very difficult as actuation modules do not have the same power to weight ratio.[1]

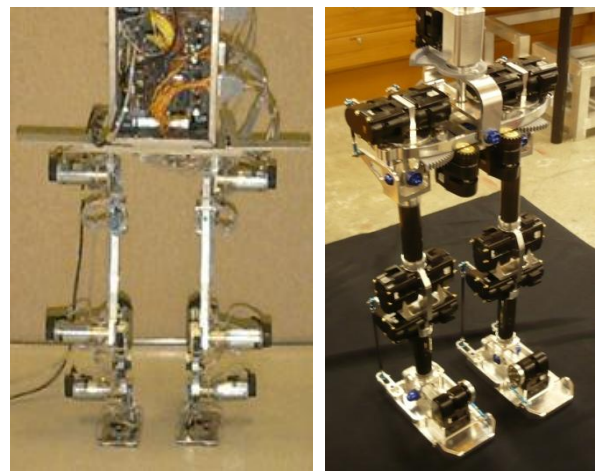


Figure 1: The UNH Biped (left) illustrates a classical humanoid design. It can be seen how large and bulky the design is compared to the Massey University Biped (right).

Since actuator modules are much larger, it makes it very difficult to make a humanoid as compact as a real human. Instead, modules are spaced very far apart in order to eliminate interference which alters the center of grav-

ity (COG) making control harder. [6] Also when actuators are spaced far apart the joints pivot axis is moved making joints swing, as opposed to pivot at a single point.

2.1 SOLUTIONS TO DESIGN ISSUES

It is recommended to bring motor weight up as high as possible. i.e. if possible bring the ankle motors up into the knee, bring the knee into the hip, and the hip into the body. However this layout cannot be completely implemented as it is not entirely necessary and increases complexity. Therefore only two characteristics were used in this project, raising the ankle into the knee, and raising the hip into the body. Below figure 1 are basic calculations to show the inertia has decreased by 50%. To control the joints some form of linear actuation would have to be used in order to translate movement from the knee to the ankle. Linear actuator/screws/push rods would be the most cost effective solution. Since a linear pushrod solution is being used it is possible to use two actuators to control both movements. i.e. Instead of one motor for roll, and one for pitch, both motors will manipulate both roll and pitch. This technique doubles the power of each joint meaning that less expensive modules can be implemented.

In order to make the pelvis and hip area as compact as possible, gears are implemented in the hip vertical axis, and pelvis roll and yaw axis. Figure 2 shows classic actuator positioning and proposed positioning.

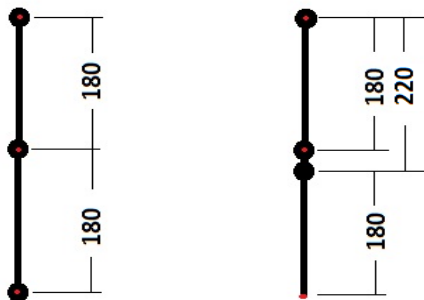


Figure 2: Left shows classic actuator positioning with motors at the bottom of the leg where as the right shows the proposed layout.

Calculating leg inertia: $I = mr^2$

Servo weight, 55g each

$$0.11 \times 0.18^2 = 0.00356 \text{ kg/m}$$

$$0.11 \times 0.36^2 = 0.01426 \text{ kg/m}$$

$$\text{Total inertia: } 0.00356 + 0.01426 = 0.01782 \text{ kg/m}$$

$$0.11 \times 0.18^2 = 0.00356 \text{ kg/m}$$

$$0.11 \times 0.22^2 = 0.00532 \text{ kg/m}$$

$$\text{Total inertia: } 0.00356 + 0.00532 = 0.00889 \text{ kg/m}$$

$$0.01782 \times \% = 0.00889$$

50% less inertia.

This is only including the weight of the servos assuming the rest of the leg is weightless.

3 ACTUATOR MODULE SELECTION

There are two types of readily available servos for hobby robotics. The classic PWM controlled servo and the new serial controlled servos. The serial controlled servo means they can be linked in one long daisy chain so there is no need for individual PWM or driver circuits.

The serial controlled servos are much more desirable due to their easier control and built in features. There are two servos to choose from that meet torque specifications and are affordable. The WCK 1111 and the AX-12 are pretty similar with good torque and priced at about \$55 each.

They both have a nice network protocol. The WCK1111 is similar to AX-12 but with extended functionalities. It has full PID control so they are very quiet and don't have the noise issue that sometimes plagues worn AX-12s. PID also lets you tune for an elastic feel, which is desirable for a humanoid limb as exact positioning is not ideal.

They have a nice dynamic brake feature which is a third power state. 'Holding Position', 'Torque' and 'Free Motion', just like AX-12 and the third Dynamic Brake mode where they are in between the two. In this mode there is no power consumption but they still offer substantial amount of strength, although with some force you can move them. Torque feedback works better on Robobuilder WCK1111 servos than on AX-12. However it is still not very accurate.

The main issue with Robobuilder servos are gears breaking. The WCK 1111 already come with 2 metal gears and some of the new servos come with the 4th gear, the main one, in metal too. The only gear that needs replacing is gear number 1 which then makes for a full metal gear box eliminating gear breakage and upping the maximum torque from 11kg/cm to 14.5kg/cm.

Something as basic as an Arduino or 8051 micro can be used to operate Robobuilder WCK servos, whereas control for AX-12 is a little more complicated as you need to split the single control pin into standard TX RX control lines. Robotis AX-12 is still the leader in the servo area. They are seen as the best quality and the best product if you're into the "real deal". Robobuilder is portrayed as a "cheaper" alternative to the AX-12.

4 JOINT DESIGN

The joints, especially hips and ankles, need to have as little slop as possible. The amount of slop in a joint is greatly magnified over the length of a limb. This cannot be completely eliminated but it can be greatly reduced by implementing ball bearing races in each joint. The manufacturing time, and cost, increases considerably but the precision and life expectancy of the joint increases. Usually ball bearings have a large amount of slop axially. This means dual bearings must be used on a shaft and spaced sufficiently to stop the slop. This in some cases can be impractical, for example in a hip joint where space is very scarce, in this case a large diameter low profile bearing can be used. A larger diameter bearing does not necessarily have less slop but due to the larger diameter will reduce the angle the slop produces. Figure 3 and 4 show the implementation of roller bearings in the hip and ankle joints.

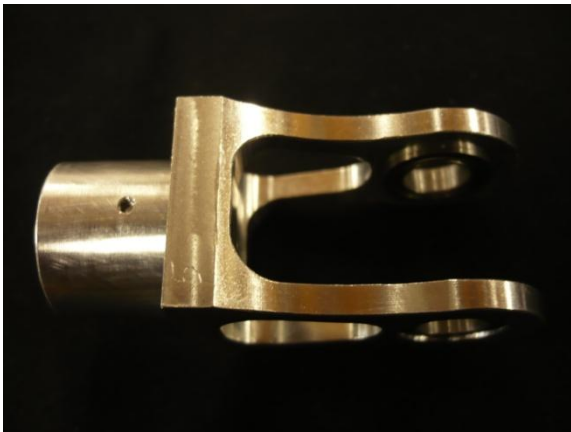


Figure 3: Ankle bearing joint.



Figure 4: Hip bearing joint.

4.1 FOOT AND TOE

A neurological disease known as Tabes dorsalis, which is due to a vitamin B12 deficiency, results in a severe loss of reflexes in the foot affecting the sufferer's sense of position resulting in difficulty in walking. Since sensory neurons are damaged, the syndrome underlines the importance of sensors in the feet in maintaining balance. [8]

All bipeds to date lack the complexity of a real human foot. The human foot is very advanced consisting of 28 bones, 20 muscles and approximately 7800 nerves. The fact that there are 7800 nerves for force sensing, which is in turn used for balance, implies that the foot is a very important part of a humanoid robot.[7]

Corrective movements are made at both the hip and ankle joints. Some postural corrections are made through articulation of individual toes. With a flat foot the degrees of freedom (DOF), are limited to only the ankle which reduces stride length. With a toe joint a stance with both feet on the ground can be maintained for a longer time allowing the COG to be transferred onto the leading foot. This toe arrangement allows the humanoid to walk more elegantly, like a human, as opposed to the more common shuffle type of walking.[2]

The human foot is so complex that it would be almost impossible to develop a mechanically operated foot with the same flexibility, strength, and size as the real one. A singular splitting foot, one toe, is sufficient for the development of a low cost humanoid with a highly intelligent gait process. [3] Figures 5 and 6 show the articulation of the single splitting foot.

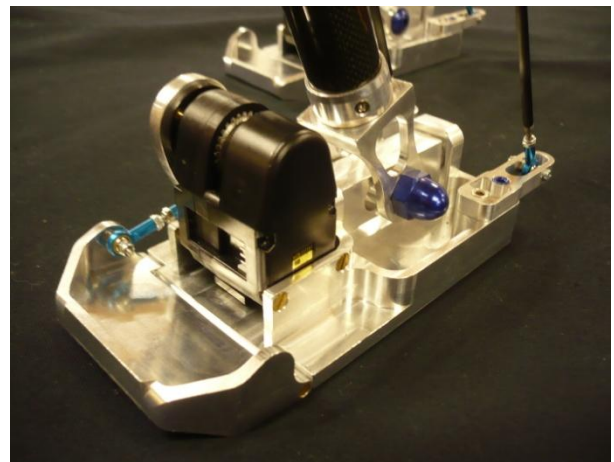


Figure 5: Foot showing toe flat.

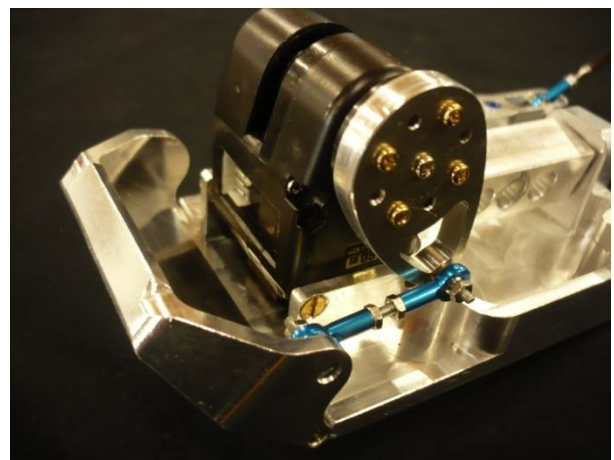


Figure 6: Foot showing toe up.

4.2 KNEE CAP

To simplify height control, a knee cap was implemented. This would mean that the drive modules would drive to a particular position then apply a low continuous torque. This would ensure that the knee is completely straight and not partially inverted. Bipedal robots that do not use knee caps require larger power consumption to make sure the knee is positioned correctly. [3] Figure 7 shows the implementation of the knee cap.

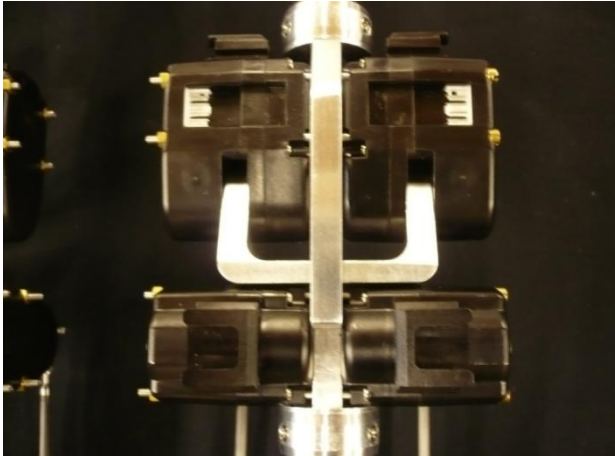


Figure 7: Knee configuration showing dual servos and knee cap.

4.3 HIP

A human hip is a ball joint. This means that there are 3 DOF. To achieve this in a mechanical application an extra vertical axis is added to a standard pitch and roll type configuration essentially producing a yaw axis.

4.4 PELVIS PIVOT AND TWIST

When walking, the pelvis does not remain static. It is manipulated with the same number of DOF as a hip joint. However, to control the pitch of the pelvis the hip pitch axis is used. This means there are only two DOF in the pelvis yet there are three to control in the system. It is essential that the roll of the pelvis is as close to the centre of the hip axis. This is because the further the pelvis axis moves away from the hip the greater the swing effect.[11] i.e. instead of the pelvis rolling side to side, lifting and dropping the hips, the pelvis will swing the entire lower body side to side. This makes for harder gait control.[4]

The second axis of the pelvis is not as crucial. This is the vertical axis which is positioned between the lower body and upper body. It is preferred to have the pivot point close to the lower body as this gives a higher weighted upper body which helps with reducing momentum in the lower body.[10]

5 GEAR DRIVE

Large drive gears are used to increase torque but most importantly to concentrate the COG and make the body more compact. These gears are used for pivoting the hips through the vertical axis, tilting the pelvis horizontally, and twisting it on the yaw axis. With a low gear ratio, 20:75, the actuator module can only rotate the axes about 65 degrees. This is more than enough for the rotation of the hip and pelvis roll and yaw as they only need very small movements.

6 ELECTRONIC SYSTEM

6.1 3 AXES AHRS

To monitor the position of both body and pelvis some form of sensor must be implemented. An Attitude Heading Reference System, (AHRS), consists of 3 gyroscopes, 3 accelerometers and 3 magnetometers. These systems are currently used in intelligent gait humanoids. [9] An on board processor calculates the rate of change with respect to time and outputs an absolute position in hexadecimal. The magnetometer is used as a horizontal reference to compute the rotational angle whereas gravity is used for the roll and pitch.

6.2 2 AXES INCLINOMETER

An AHRS cannot be used near any source of magnetic interference due to its internal magnetometers. A sensor is needed on the pelvis which is surrounded by aluminium and electric motors. To calculate pelvic angle a 2 axes inclinometer is used which works similarly to the AHRS except it does not rely on magnetic sensors.

6.3 FSR FOOT PRESSURE SENSOR

The Bioloid foot pressure sensor is a PCB that has inputs for four resistive sensors and has built in onboard processing. With onboard processing, i.e. analog to digital to serial conversion, this reduces the processing needed to be done by the onboard computer. Also, since these sensors are essentially voltage dividers, the change in voltage is very small which would result in an amplified, or even completely lost, signal when transmitted to the controller in the body.

However, complication arises due to these boards being half duplex, i.e. having a single pin for both transmit and receive. Unlike standard double pin serial communication where one pin is dedicated to transmit, and one for receive, this board must have an additional logic controller to switch between sending and receiving.

6.4 MICRO COMPUTER

The amount of processing that will be done is so large and complex that a standard micro controller will be insufficient. The RB 110 comprises of a 1GHz processor, 256Mb DDR2 ram, 10Bit ADC, 4 regular Com ports, I2C, 2 High Speed Com ports, Windows XP operating system, and open source C++ library with unique I/O functions. This is basically a miniature computer dedicated for robotics use and has been tailored to suit humanoid robots such as Bioloid and Robobuilder which use serial controlled servos and the Kondo which uses Hitech PWM servos.

Microsoft Robotics Developer Studio is a commonly used programming language for simple robotics control. Serial communication in this language is very simple, however more complex scripts are very time consuming and difficult to develop. Therefore for this project it has been decided that Visual Basic.NET will be used.

7 TESTING

To ensure that the WCK servos were strong enough, two sample servos were purchased. A mock up knee was manufactured to the calculated dimensions. The leg was then mounted to a test rig which could support various weights. Figure 8 on the left is a picture of the leg at full crouch. This would determine the maximum weight the servos could lift. On the right is a picture of the knee just below fully straight, this position is where the leg will be operating the most. At full crouch the servos can lift 800g, and within the operating parameters they can lift about 1.5-1.9 kg depending on the supply voltage.



Figure 8. Test rig for determining WCK module strength.

Below in figure 9 shows a very simplified functional block diagram of the system.

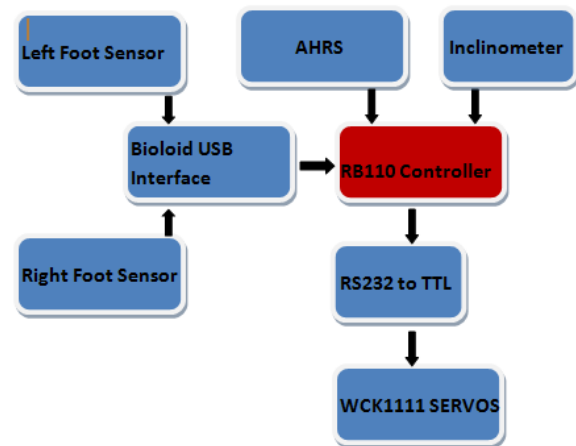


Figure 9. Block diagram of system

8 CONCLUSION

With classical joint articulation, humanoid robots are limited in their functionality. In order to produce torque even close to what a human muscle can produce generally motors and actuators have to be very large. To keep them compact the cost is significantly increased which increases the cost of research projects that can be undertaken in this field.

In order to be able to compete with larger and more technologically advanced humanoids a new type of articulation must be developed. A dual module/servo joint control will double joint torque and with the ability to position the modules away from the actual joint will further reduce joint inertia.

To reduce joint slop, ball bearing races are used at every pivot point, minus the toe and to reduce weight, carbon fibre is used where possible. Due to this humanoid being the first prototype there are many possible modifications that can be done in order to reduce weight and increase flexibility.

Sensors such as inclinometers, 'Attitude Heading Reference Systems' and pressure sensors are used to monitor characteristics of the robot in order to maintain balance.

This project will be developed over many years. At present the first prototype of the lower body has been constructed and the control process has begun. The development of a suitable gait control system is very complex and time consuming and therefore a good foundation for control of servos and monitoring of sensors must first be established.

9 REFERENCES

- [1] T. I. J. Tsay, C. H. Lai, “*Design and Control of a Humanoid Robot*” Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems October 9 - 15, 2006, Beijing, China
- [2] L. Wang, Z. Yu, Q. Meng and Z. Zhang “*Influence Analysis of Toe-joint on Biped Gaits*” Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation June 25 - 28, 2006, Luoyang, China
- [3] J. E. Pratt “*Exploiting Inherent Robustness and Natural Dynamics in the Control of Bipedal Walking Robots*” Massachusetts Institute of Technology 2000
- [4] J. H. Kim, S. W. Park, I. W. Park, and J. H. Oh, “*Development of a Humanoid Biped Walking Robot Platform KHR-1 –Initial Design and Its Performance Evaluation–*”, in Proc. of 3rd IARP Int. Work. on Humanoid and Human Friendly Robotics, pp.14–21, 2002.
- [5] W. Shi, S. D. Hall, “*A novel six-axis force sensor for measuring the loading of the racing tyre on track*” International Conference on Sensing Technology 2005
- [6] K. Masayuki, M. Takahiro, M. Noriaki, I. Hiroshi and H. Norihiro “*How does a Balancing Motion of a Humanoid Robots Affect a Human Motion and Impression?*”
- [7] F. R. Sias, Y. F. Zheng “*How many degrees of freedom does a humanoid need?*” IEEE International Workshop on Intelligent Robots and Systems IROS 1990
- [8] P. D. Donofrio, F. O. Walker “*Tabes dorsalis: electrodiagnostic features*” Department of Neurology, Wake Forest University, Bowman Gray School of Medicine, Winston Salem, North Carolina, USA 1988
- [9] S. Kim, S. Hong, D. Kim “*A Walking Motion Imitation Framework of a Humanoid Robot by Human Walking Recognition from IMU Motion Data*” 9th IEEE-RAS International Conference on Humanoid Robots December 7-10, 2009 Paris, France
- [10] “*Workshop on Mechanical Design of Humanoids*” Workshop of the 9th IEEE RAS International Conference on Humanoid Robots (Humanoids09), Paris, December 7 10, 2009.
- [11] I. W. Park, Y. Y. Kim, S. W. Park, and J. H. Oh, “*Development of Humanoid Robot Platform KHR-2(KAIST Humanoid Robot - 2)*”, Int. Conf. on Humanoid 2004.

Parameter Tuned Digital Filter

Ian R. Scott, T. C. A Molteno

University Of Otago Physics Department, University Of Otago

Dunedin, P.O. Box 56

New Zealand

scoia295@student.otago.ac.nz, tim@physics.otago.ac.nz

Abstract:

Digital filters are widely used in many signal-processing systems. They are divided into FIR and IIR categories. Each category then has multiple implementations. Although referred to as “digitally accurate” these filters remain inflexible. They typically require a large number of coefficients that must be predetermined prior to application. This limits their use in software-defined radios (SDR) that must process many diverse modulation formats simultaneously. Each new modulation format therefore demands human intervention. Even so, determining a common assortment of integer related clock frequencies and filter architectures may not be feasible. We propose an alternate Parameter Tuned Digital Filter or PTDF using integer arithmetic. We show that this filter can provide a flexible alternative to traditional FIR and IIR designs. Few parameters are required to define PTDF's and determining these parameters is straightforward and intuitive. Once determined, the PTDF's frequency and time domain behaviours allow simple parameter scaling. These favourable characteristics suggest its inclusion in highly flexible, software defined radios.

Keywords: SDR, “Digital Filter”, “Integer Arithmetic”, “Nested Loop”, Biquad, Butterworth, Chebyshev, Decimation, Interpolation,

1 INTRODUCTION

Previous analogue filters suffered from performance variation caused by component tolerances. In contrast, the digital filter is as stable as its clock sampling frequency F_s . Further they achieve high ratios of stop to pass band rejection that would be impractical for analogue filters. However digital filters have two weaknesses; large numbers of coefficients have to be predetermined by various mathematical methods to implement a target frequency or time response. New coefficients have to be recalculated when requirements change. These new coefficients do not resemble the previous set. Simple scaling approaches are ineffective and human intervention is needed. Further, the fractional bandwidth Φ for a cut frequency F_c defined as $\Phi \equiv 2 \pi^{F_c} / F_s$ has a limited range of feasible values e.g. $\pi/5 < \Phi < \pi$. Low Φ forces excessive filter length. Therefore many coefficients and digital multiplies are needed. This results in excessive use of FPGA resources.

This inflexibility seems counter-intuitive. Many people would expect that a digital filter should be software configurable. Whilst true, human intervention is needed for modifications. Even so, multiple requirements often conflict in applications. For example, a filter designed for wide band service becomes excessively large in narrow band applications. Simply recalculating coefficients is inadequate. Proposing a bank of application specific filters is also impractical – the excessive size of a narrow bandwidth filters may exceed the FPGA resources designed for wide bandwidth applications. Increasing FPGA size might suggest a solution but then additional cost would be incurred. Even if these costs are accommodated, specific clock frequencies are required for each data rate. These may not be integer related or have an unattractively low common factor.

The SDR requires flexibility as a fundamental requirement. Further the SDR is required to process conflicting modulation standards in parallel. Each standard demands specific frequency and time response characteristics. There is no guarantee that a common clock frequency can be found. Even if phase locked loops are used these still require a common integer related reference fre-

quency. Again, even if a common frequency is found, specialist design and potentially expensive hardware is needed for suitable low phase noise clocks. The generic SDR remains seeking suitable enabling technologies.

The motivation of this document is to propose an alternate Parameter Tuned Digital Filter, or PTDF architecture. The PTDF avoids many FIR and IIR weaknesses. The PTDF operates equally well at low and high values of Φ . Successful operation has been obtained with values of Φ less than 0.001. Further, Φ is linearly scaleable by a single global parameter that we will just call k . In fact,

$$\Phi' \equiv k \cdot \Phi \quad (1)$$

This inherent flexibility simplifies the design of a SDR that could, in principle, process multiple simultaneous modulation formats. In contrast to FIR and IIR filters the PTDF complexity is independent of Φ .

The PTDF is inherently tuneable by k . Its frequency domain transfer function is also continuously tuneable with extra parameters k_1, k_2, \dots, k_N . The number of parameters equals the filter order N . The PTDF emulates conventional analogue filters of the same order N . It has an IIR response and closely emulates Butterworth or Chebyshev filter types. Additionally, similar design methods can be employed to design the PTDF based on analogue filter design methods

The PTDF requires little FPGA resource compared to its FIR and IIR counterparts. For example, a $N=4$ Chebyshev PTDF only uses four multipliers and parameters. A corresponding IIR requires far more FPGA resource especially at low values of Φ . Even if these extra resources are accommodated, the IIR remains inflexible whilst the PTDF would not. Further, the PTDF can use low precision fixed-point arithmetic. Not all FIR or IIR are well suited to fixed-point arithmetic. We show stable operation on a Spartan 3 FPGA platform programmed with 8-bit resolution.

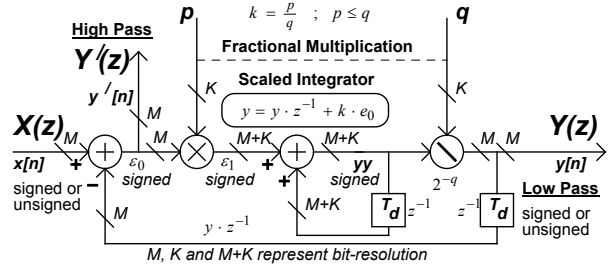
1.1 First Order PTDF Topology

The first order PTDF is simplest. It requires an integrator, integer input scaling, power of two-output division and feedback. Division by two corresponds to an array shift operation and uses minimal FPGA resource.

The PTDF is tuned by parameter p with fixed $q = 2^Q$ where Q is an integer 1, 2, \dots, N . The parameter k is defined as $k \equiv p/q$. The lowest value for Φ occurs when $p = 1$ but q can be made arbitrarily large. Therefore Φ can be made arbitrarily small without increasing filter length.

The first order PTDF can be implemented using the following architecture

1st Order Parameter Tuned Digital Filter (PTDF)



This PTDF produces simultaneous high and low pass outputs. It implements fractional multiplication despite using fixed-point arithmetic. We call this fractional multiplying integer a FMI.

1.1.1 First Order PTDF Analysis In The z-Domain

Z-domain analysis reveals

$$LPF(z) \equiv \frac{y}{x} = \frac{k}{1 + (k-1) \cdot z^{-1}} = \frac{k \cdot z}{z + (k-1)} \quad (2)$$

The corresponding high pass response is

$$HPF(z) \equiv \frac{\varepsilon_0}{x} = \frac{z-1}{z + (k-1)} \quad (3)$$

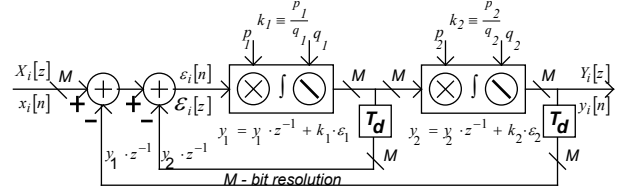
The conventional definition for z is.

$$z \equiv e^{j \cdot 2 \cdot \pi \cdot \frac{f}{F_s}} \quad (4)$$

1.2 Second Order PTDF Topology

Adding an additional integrator and extra feedback path creates a second order PTDF. It therefore has a nested loop feedback structure. The following diagram shows two FIM's. Bit resolution is shown as M .

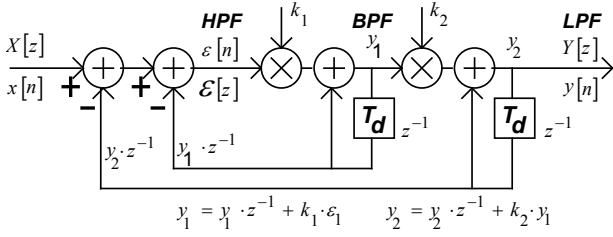
Second Order PTDF Using Two FIMs With Dual Feedback



1.2.1 Z-Domain Analysis For The 2nd Order PTDF

The second order PTDF provides an additional band-pass output where the two integrators are joined.

Simplified Second Order PTDF For Analysis



The z-domain equations used for analysis are

$$\begin{aligned}\varepsilon &= x - y_2 \cdot z^{-1} - y_1 \cdot z^{-1} \\ y_1 &= y_1 \cdot z^{-1} + k_1 \cdot \varepsilon \\ y_2 &= y_2 \cdot z^{-1} + k_2 \cdot y_1\end{aligned}\quad (5)$$

We can derive the low pass z-domain transfer function from (4) by algebraically eliminating variables.

$$LPF_2(z) \equiv \frac{y_2}{x} = \frac{k_1 \cdot k_2}{1 + (k_1 \cdot k_2 + k_1 - 2) \cdot z^{-1} + (1 - k_1) \cdot z^{-2}} \quad \dots(6)$$

$$BPF_2(z) \equiv \frac{y_1}{x} = \frac{k_1 \cdot (1 - z^{-1})}{1 + (k_1 \cdot k_2 + k_1 - 2) \cdot z^{-1} + (1 - k_1) \cdot z^{-2}} \quad (7)$$

$$HPF_2(z) \equiv \frac{\varepsilon}{x} = \frac{1 - 2 \cdot z^{-1} + z^{-2}}{1 + (k_1 \cdot k_2 + k_1 - 2) \cdot z^{-1} + (1 - k_1) \cdot z^{-2}} \quad (8)$$

1.3 Low Frequency Analysis

It is convenient to transform equations in z to frequency f especially when Φ is small. Consider the following definition for normalised frequency Ω .

$$\Omega \equiv 2 \cdot \pi \cdot \frac{f}{F_s} \quad (9)$$

The complex exponential in z can be expanded using a Taylor series and approximated for small Ω .

$$z \equiv e^{-j \cdot \Omega} \cong 1 - j \cdot \Omega \quad (10)$$

Substituting (10) into (6) results in

$$LPF_2(\Omega) \cong \frac{1}{1 + j \cdot \Omega \cdot \left(\frac{1 - k_2}{k_2} \right) - \Omega^2 \cdot \left(\frac{1 - k_1}{k_1 \cdot k_2} \right)} \quad (11)$$

Equation (11) represents a second order frequency response approximation based on Ω . The same approach can equally be applied to band and high pass transfer equations (7) and (8).

1.5 Higher Order PTDF.

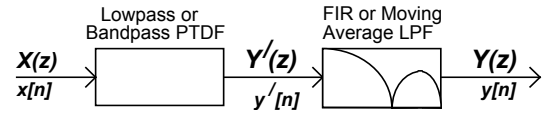
High orders PTDF are formed by cascaded lower order PTDF's. This approach is identical to conventional high order analogue filter design. Although a single high order PTDF could be formed using multiple feedback paths and FMIs it is simpler to multiply lower order PTDF's. The composite z transform becomes

$$H_N(z) = \prod_{n=0}^N H_n(z) \quad (12)$$

1.6 Periodic Frequency Response Removal

The PTDF has a repeating frequency response. For example, characteristics at Ω and $1 - \Omega / (2 \pi)$ are equivalent. In some cases a simple moving average filter can be combined with the PTDF to remove this response.

Removing The PTDF Periodic Frequency Response



This "roofing filter" is required to remove energy above $1 - \Omega / (2 \pi)$ and will not require adjustment.

1.7 Second Order Frequency Response Simulation

1.7.1 Simultaneous LPF, BPF and HPF Outputs

We used Mathcad to predict the frequency response behaviour for a second order PTDF corresponding to equations (6), (7) and (8). The simulation was implemented in time. The Mathcad algorithm, shown below, is simple.

```
PTDF_2(k1, k2, x) :=
Samples ← rows(x)
yy ← 0
yout ← 0
for n ∈ 0.. Samples - 1
    ε ← xn - yout
    ε ← xn - yy - yout
    yy ← yy + k1 · ε
    yout ← yout + k2 · yy
    yn,0 ← yout
    yn,1 ← yy
    yn,2 ← ε
y
```

A time domain impulse “ x ” was used as a stimulus. Three outputs corresponding to low pass, band pass and high pass transfer functions are shown. The final variable y is an $N * 3$ matrix where N represents the number of time domain samples. Parameters $k1$ and $k2$ control the PTDF characteristics.

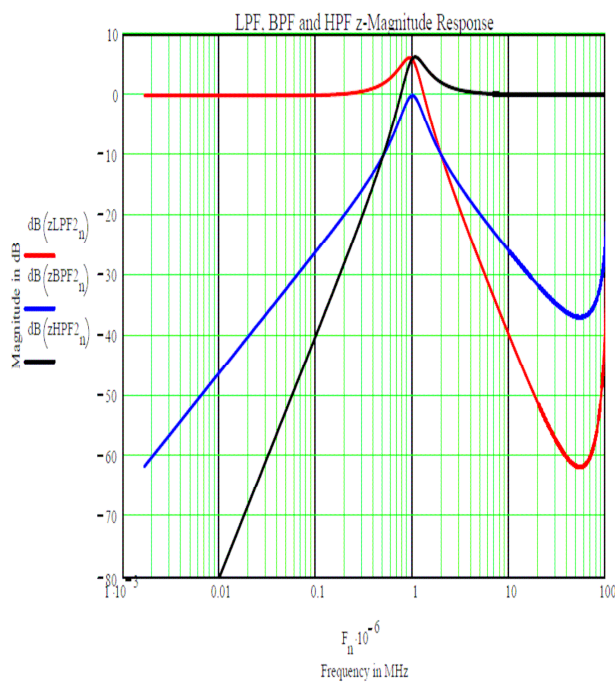
Matrix y contains individual transient responses generated by this second order PTDF. These are applied to a FFT function, creating three frequency domain responses. The results obtained were identical to direct z -domain simulation, shown below in Mathcad syntax.

```

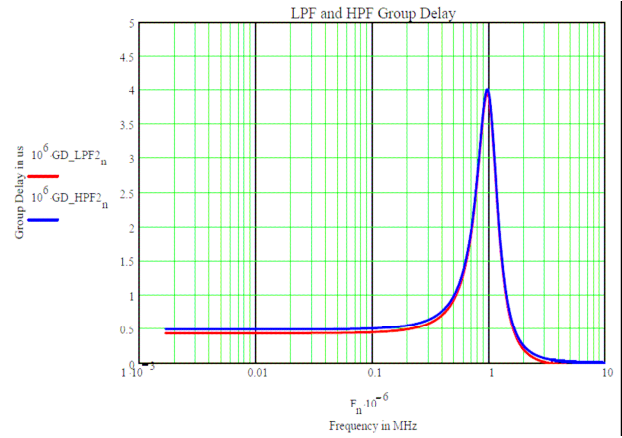
z_PTDF_2(Fs, k1, k2, F) :=
  N ← rows(F) - 1
  for n ∈ 0..N
     $z \leftarrow e^{j2\pi \frac{F_n}{F_s}}$ 
     $D \leftarrow 1 + (k1 \cdot k2 + k1 - 2) \cdot z^{-1} + (1 - k1) \cdot z^{-2}$ 
     $PTDF_{n,0} \leftarrow \frac{k1 \cdot k2}{D}$ 
     $PTDF_{n,1} \leftarrow \frac{k1 \cdot (1 - z^{-1})}{D}$ 
     $PTDF_{n,2} \leftarrow \frac{1 - 2 \cdot z^{-1} + z^{-2}}{D}$ 
  PTDF

```

The following graph shows three frequency domain amplitude responses in dB. We found that both approaches gave identical responses so only one version is shown.



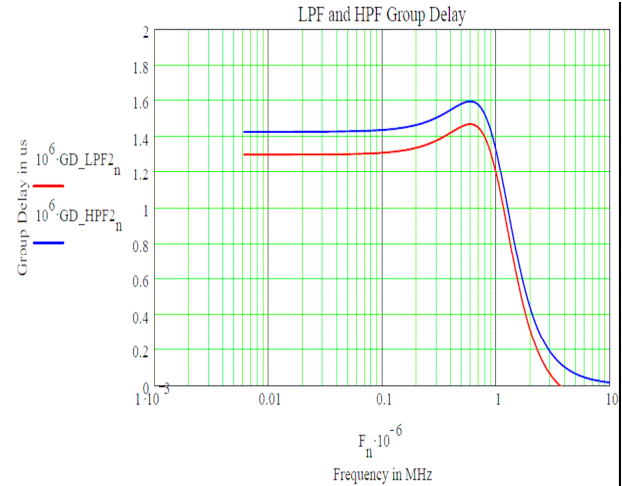
The next graph shows differential group delay for high filter “ Q ” values. This represents a Chebychev response with high pass band ripple.



The previous graphs used the following parameters

$$Q := 2 \quad k := 2 \cdot \pi \cdot \frac{F_c}{F_s} \quad k1 := \frac{k}{Q} \quad k2 := k \cdot Q$$

Reducing the PTDF Q to 0.7 improves group delay significantly and converts a Chebychev response to a Butterworth version.

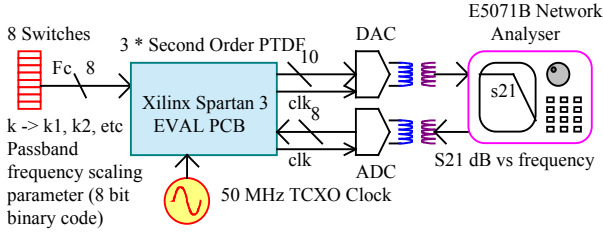


The familiar concern over IIR group delay is perhaps ill conceived based on this second graph using $Q = 0.7$. Low and High pass outputs are shown to be very similar.

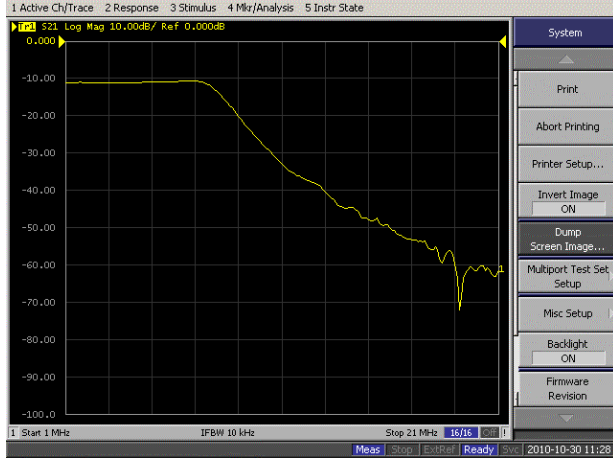
2 DISCUSSION

First and Second order PTDF's have been analysed and simulated in Mathcad. We used Verilog to implement both in a Xilinx Spartan 3 Starter Kit. Preliminary tests have been conducted at 50 MHz sample rate with 8 bit DC input test vectors. A dual hexadecimal alphanumeric display was used to indicate 8 bit output values. We plan to update the test configuration with an 8-bit ADC, 10-bit DAC and ancillary components. This is planned for the next few weeks. However the actual PTDF appears to remain well behaved with this test configuration.

RF Test Configuration For The Sixth Order PTDF



S21 Plot Using Agilent E5071B Network Analyser



Preliminary results are encouraging. The plot above shows a 6th order PTDF frequency response when implemented on the Spartan 3 FPGA. This PTDF used three-cascaded 2nd order PTDF's with a common frequency scaling factor k . An 8-bit 50 MHz ADC was used for signal acquisition and a 10-bit DAC was used to output the filtered result.

The PTDF is motivated by the need for complete parameter defined characteristics needed in a SDR. However it is also well suited for audio applications such as Automatic Level Control (ALC). Such systems require “fast attack-slow decay” gain behaviour. The parameter-controlled bandwidth of the PTDF can be switched between increasing and decreasing audio levels without introducing transients. When the audio level increases, a high k is used. When the audio level falls k takes on a smaller value. This ALC application is less well suited to FIR and IIR filter types with pre-determined coefficients. Simply switching between two standard filters is likely to introduce ALC transients and show up as audio clicks

Digital systems typically require exact integer ratios between internal sampling frequencies. To illustrate consider a SDR operating with two system bandwidths. The use of fixed cascaded decimator and interpolation filters places high demands on the design engineers. They need to find suitable common multiples of clock frequencies that allow simultaneous signal processing across both channel bandwidths. If a third channel bandwidth is added then finding a 3-way solution becomes exponen-

tially difficult both for the engineer and potentially the digital processing hardware. Given that a generic SDR should not be limited to three simultaneous formats, the conventional approach may be unattractive.

This inflexibility represents the Achilles heel of current digital radio systems intended for SDR. One outcome from this research project is to find digital processing topologies that avoid inflexibility without adding excessive complexity. The PTDF has uncomplicated implementation compared to FIR or IIR approaches. It uses far less FPGA resource. It provides potentially high decimation or interpolation ratios in a single filter. This is because Φ can be arbitrarily small without increasing PTDF size and corresponding FPGA resource. In contrast conventional decimation filters are usually limited to ratios of 2 to 5. To illustrate, consider an ADC sampling at 50 MHz. Also assume that a final demodulation bandwidth need only require a sample rate of 100 kHz. The decimation ratio is therefore 500:1. In principle a single PTDF could be used. The alternative approach might require decimation ratios of 5, 5, 5 and 4 to achieve this overall decimation ratio. Obviously four FIR filters would be needed.

We observe further difficulties that this approach imposes. When we consider a generic SDR, each one of these decimation filters requires a common clock reference frequency. If finding a common clock reference for a single filter is difficult, extending the requirement to 4 filters may make this task almost impossible.

We do not propose the PTDF as a single solution to a generic SDR. We do propose it as one potential solution of many that require invention and development. Nor do we claim that the PTDF is free of unwelcome characteristics not yet experienced. Further research is needed to assess its complete characteristics. However, it must be said, that any filter has its pro's and con's. The most we can hope for is that the pros outweigh the cons or that a suitable compromise is available.

3 FURTHER WORK

The PTDF is intended for Software Defined Radio (SDR) applications. The author's project is intended to show “proof in principle” for a smaller scale SDR. This project requires both technology innovation and application in a FPGA, combined with suitable RF processing circuits and systems. It is not simply a software programming exercise executed in a FPGA. Some imagination and creativity is required for a successful SDR outcome. Further, divergent disciplines are needed to finish a working prototype SDR for *demonstration*.

A prototype test system is planned for operation at 13.45 MHz but with parameter controlled frequency, bandwidth and data rate. This architecture will be scalable for much higher frequency operation. In this view the current investigation is intended to provide “proof in principle” for the PTDF filter architecture and other software structures with high levels of flexibility that elude conventional approaches.

Other software structures also need refinement for this example SDR. For example, the standard Numerically Controlled Oscillator (NCO) requires a phase accumulator, relatively expensive sine-cosine Look Up Table (LUT). These NCOs often need ancillary enhancement based on dither noise injection and polynomial correction. They can achieve exceptional performance but use significant FPGA resource. Recent investigations reveal that a potentially simpler architecture is feasible by changing the way the NCO is mathematically implemented. Outcomes may offer further benefit for SDRs.

The PTDF however requires a great deal of further analysis, measurement and refinement. We will investigate optimum bit resolutions based on SNR requirements. Although limit cycles have not yet been observed, critical assessment is required. The integration stages employed by the PTDF may offer great benefit however by introducing fundamental smoothing functions.

Further, the use of cascaded first and second order PTDF requires comparison and evaluation with a single high order PTDF. We will investigate the feasibility of a PTDF with orders much higher than 2. Our interest is in both the comparative performance as well as potential reduction in FPGA resource demand.

The tuneable nature of the PTDF may also offer significant benefit in systems that require ad-hoc adaptability. For example, consider a receiver experiencing a strong interferer close to its reception channel. The PTDF could simply reduce its channel bandwidth to exclude this interferer. Although this would usually increase BER on a “clear” channel, the actual BER on an interfered channel could significantly improve given this strategy. Equally the PTDF could shift its centre frequency to the opposite side of the interferer.

4 CONCLUSIONS

The results to date have been encouraging. The PTDF may not be an essential component for a SDR but it does suggest potential benefits in flexibility. It appears to overcome many of the weaknesses associated with conventional approaches. This inherent flexibility and low FPGA resource overhead is attractive. Operation at low Φ may also simplify decimation and interpolation processes. Its introduction in SDR systems requires strategies for optimal placement and design. Further, these *generic* SDRs need to progress beyond “proof in principle” demonstrators to real world products used commercially.

If further analysis and testing remains favourable, the PTDF may well provide a useful addition to traditional FIR and IIR digital filter technologies.

5 REFERENCES

Modelling, Synthesis and Rapid Prototyping with the Verilog HDL – Michael D. Ciletti, University of Colorado, Colorado Springs,

Digitally Removing A DC Offset: DSP Without Mathematics, Ken Chapman – Xilinx, WP279 (v1.0) July 18, 2008

ELEC 358: Electronic Design Techniques, Tim Molteno, Physics Department, University Of Otago, Dunedin, September 3, 2008

http://en.wikipedia.org/wiki/Software-defined_radio

http://www.wirelessinnovation.org/page/Defining_CR_and_DSA

Acronyms:

FPGA – Field Programmable Gate Array

SDR – Software Defined Radio

FMI – Fractional Multiplication Integrator

DGD – Differential Group Delay

FIR – Finite impulse Response (filter)

IIR – Infinite Impulse response (filter)

PTDF – Parameter Tuned Digital Filter

Ω - Filter Fractional Bandwidth (f_c/f_s)

Volume Measurement Using 3D Range Imaging

Vipul Shrivastava, Michael J. Cree, Adrian A. Dorrington*

*Department of Engineering, University of Waikato, Hamilton, New Zealand
Email: cree@waikato.ac.nz

Abstract:

The use of 3D Range Imaging has widespread applications. One of its applications provides us the information about the volumes of different objects. In this paper, 3D range imaging has been utilised to find out the volumes of different objects using two algorithms that are based on a straightforward means to calculate volume. The algorithms implemented successfully calculate volume on objects provided that the objects have uniform colour. Objects that have multicoloured and glossy surfaces provided particular difficulties in determining volume.

Keywords:

range imaging, volume measurement

1 INTRODUCTION

The Range Imaging Cameras are a class of ranging device that acquire a full-field image of range data simultaneously of the scene from one vantage point. Currently available commercial cameras, such as the Mesa Imaging SwissRanger, enable efficient and easily obtainable low-resolution (up to 200×200 pixels) range data with one or two centimetre precision in ranging. In this paper we explore the use of the Mesa Imaging SwissRanger SR4000 [1] to measure the volume of objects in the field of view (FOV) of the camera.

The use of range data acquired via mechanically scanned laser range finder for volume measurement was reported by Zhang et al. [2]. They apply a meshed surface to the point cloud data and from that estimate the volume. We, in contrast, are interested in exploring straightforward methods to volume measurement. For example, if one captures a single view of the scene, then it is relatively straightforward to estimate the volume enclosed by the camera sensor origin to the viewed surface expressed in the camera coordinate system. By subtracting off the volume of the scene with the object of interest removed gives the volume of the object provided there are no concavities in the object that are out of view.

In the following we present two algorithms for calculating the volume of an object in the FOV. All algorithms require a background scene with the object removed. Testing is performed on a number of rectangular solid objects of known volume.

2 ALGORITHMS IMPLEMENTED

The basic principle behind the algorithms investigated is that by subtracting the volumes occupied by the surfaces measured from the range images with the object and without the object the volume of the object is determined. This has the advantage that no segmentation of the object is needed. The description of the two algorithms follow.

2.1 Algorithm 1

The SwissRanger SR4000 provides range data and pixel coordinates calibrated in metres with the camera sensor at the origin of the coordinate system. It is therefore relatively straightforward to calculate the volume enclosed by the observed scene (treated as a three-dimensional surface) back to the camera sensor. If two views of the scene are captured, one with the object and one without, then the difference between the two volumes calculated is the volume of the object under view (under certain assumptions such as it has no concavities out of view).

To calculate the volume encompassed by the scene and the camera origin adjacent pyramids were formed by choosing three neighbouring pixels in the scene and the camera's origin. By choosing four points the pyramids are all tetrahedra. Three pixels in the viewed scene that form the base of the tetrahedra were always chosen so that the sides of the tetrahedra abut each other so that the full volume can be exactly filled with tetrahedra. In Figure 1 the way three pixels were chosen to make sure that the tetrahedra fill the volume is shown. The total volume is the sum of the volumes of the tetrahedra.

A systematic approach was used to position the tetra-

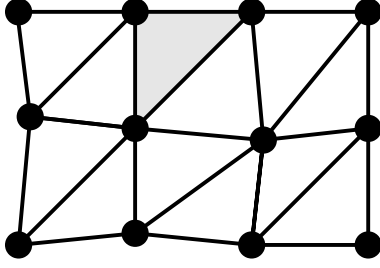


Figure 1: The grid pattern used to form the bases of the tetrahedra for volume measurement. The black dots are the pixels (which are not exactly on a regular grid but are ordered) and the lines represent the base of each tetrahedron.

hedra bases on the scene surface. First two points were selected from the first row in the horizontal direction and a third point was taken as the point in the row below the first point. In this way the selection was done moving in both directions along the x -axis and along the y -axis. This way of taking the points ensured that the tetrahedra completely fill the volume with no overlap and no gaps.

The volume V of each tetrahedron is given by

$$V = \frac{A_b h}{3}. \quad (1)$$

where A_b is the area of base and h is the height of the tetrahedron. The area of base here corresponds to the triangle formed by the three points taken on the range image and can be calculated using Heron's (or Hero's) formula.

Consider a triangle whose sides are of length a , b and c , then the area of the triangle formed by the three points using Heron's formula is given by

$$A = \sqrt{s(s-a)(s-b)(s-c)}, \quad (2)$$

where s , the semi-perimeter of the triangle, is

$$s = \frac{a+b+c}{2}. \quad (3)$$

The height h of the tetrahedron is given by the perpendicular distance of the origin from the plane formed by the three scene points forming the base of the tetrahedron. The plane formed by the three base points is

$$Ax + By + Cz + D = 0 \quad (4)$$

with A , B , C and D calculated as follows. Let us take the three points from the range image to be (x_1, y_1, z_1) , (x_2, y_2, z_2) and (x_3, y_3, z_3) , then A , B , C and D are

$$A = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2), \quad (5)$$

$$B = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2), \quad (6)$$

$$C = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2), \quad (7)$$

$$-D = x_1(y_2 z_3 - y_3 z_2) + x_2(y_3 z_1 - y_1 z_3) + x_3(y_1 z_2 - y_2 z_1). \quad (8)$$

The height of the tetrahedron, namely the perpendicular distance h of the point $P(a, b, c)$ from the base plane is,

$$h = \left| \frac{Aa + Bb + Cc - D}{\sqrt{A^2 + B^2 + C^2}} \right| \quad (9)$$

Summing over all tetrahedra gives the total volume between the camera sensor and the scene. The volume of the object of interest is the difference of the volume calculation with the object in the scene and the volume calculation of the scene without the object.

2.2 Algorithm 2

The objective of the second algorithm is to find the dimensions of the object and to calculate the volume from the object's dimensions.

In order to find the dimensions of the object two range images were taken, one with and one without the object. The idea is that only the portions of the range image which changes appreciably from the first to the second range contribute to the volume. This idea works best for the objects with no bulging protrusions.

With the two range images, the z coordinates of the range image with the object was subtracted from the corresponding z coordinates of the range image without the object. This results in a new range image, with the same values for the x coordinates and the y coordinates but subtracted z coordinates.

A threshold is applied to the data so that if the z coordinate's difference is less than the threshold, then it is set to zero, otherwise it is left unchanged. If the z coordinate of a particular pixel is non-zero after thresholding that means that pixel of the object is sufficiently high and together with the neighbouring pixels contributes to the overall volume of the object.

It is important to choose a threshold not too high else pixels of the object contributing to the overall volume may be missed, likewise it cannot be too small, else it will take background points into the object. A threshold of 0.15 m was used in this project.

The difference image was processed pixel by pixel. Any pixel whose z coordinate difference is non-zero and whose neighbours have non-zero z coordinate contribute to the volume calculation. On finding such a pixel a cuboidal block formed by the four pixels (i.e. the pixel and its neighbours) contributes to the volume of the object. The volume of all such cuboids are summed together to give the total volume of the object. Note that this method copes with holes in the object.

The volume of each cuboid is calculated in the following manner. With four neighbouring pixels that have non-zero z coordinates differences, their average was taken as the height h of the cuboid. The length l and breadth b were taken in the same way by taking the average of the dimensions along the x -axis and the y -axis. The volume V of

the rectangular cuboid is then simply given by

$$V = lbh. \quad (10)$$

3 MATERIALS AND METHODS

First the algorithm's correctness was checked with simulated data for which the volume could be exactly calculated. Range images of surfaces and objects were simulated on Matlab R2010 and then the algorithms were tested using those data as inputs. These confirmed that the algorithms were correct for precise and uncontaminated range data (see Figure 2).

The MESA SwissRanger SR4000 range imaging camera used produces calibrated range data in metres over a 176×144 pixel FOV. The camera was mounted on a stand about 800 mm above the table to point directly downwards at the table on which objects were placed. A total of 12 rectangular cuboid objects varying in size from $4 \times 10^{-4} \text{ m}^3$ to 0.019 m^3 were imaged and their volumes calculated. The object's volume were determined first by measurement with a ruler and are listed in Table 1. Some of the objects (being hardcover books) had glossy multi-coloured covers which, as discussed below, caused problems for volume measurement. A second set of images were made of the objects when they were covered in a uniform coloured matt paper.

Table 1: List of the objects imaged.

Object	Volume (m^3)
1	3.93×10^{-4}
2	7.17×10^{-4}
3	7.60×10^{-4}
4	9.55×10^{-4}
5	1.1×10^{-3}
6	1.1×10^{-3}
7	1.5×10^{-3}
8	2.2×10^{-3}
9	2.5×10^{-3}
10	8.02×10^{-3}
11	1.62×10^{-2}
12	1.89×10^{-2}

The camera used is an early version of the SR4000 and unfortunately has a calibration error resulting in pixels being misplaced by up to 4cm at the edges of the FOV due to radial lens distortion [3] hence the objects were always placed in the centre of the FOV of the camera to reduce error in volume measurement.

4 RESULTS

The plot obtained when the real covered data (i.e. the objects with uniform matt colouring) was given as the input

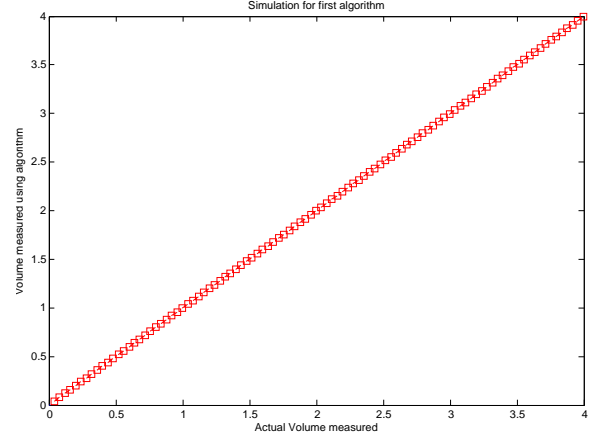


Figure 2: Volume calculation on simulated data using algorithm 1.

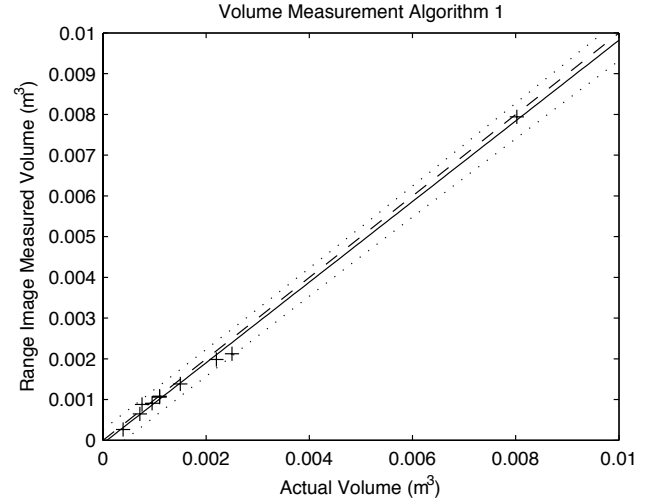


Figure 3: Volume calculation on real covered data using algorithm. 1.

to the first algorithm is shown in Figure 3. The best fit line of the points obtained using the first algorithm is given by

$$y = 0.990x + -7.80137E-05 \quad (11)$$

The second algorithm was tested using covered data and the results are presented in Figure 4. There were some of the points that didn't lie on the line specially as the size of the objects increased. The behaviour of the algorithm showed similarity with the previous algorithm's behaviour. The line of best fit to the data is

$$y = 0.944x + 4.17017E-05 \quad (12)$$

From the above obtained results, on comparison of the two algorithms for the covered data, it was seen that the second algorithm gave closer results to the actual volume as compared to the first algorithm. The slope of the best fit

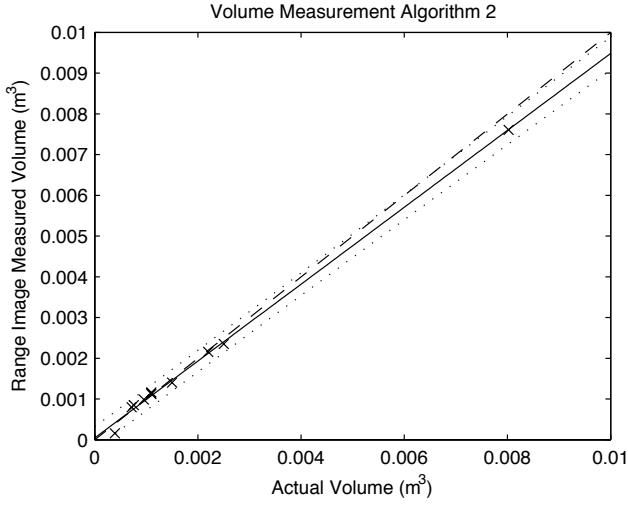


Figure 4: Volume calculation on real covered data using algorithm. 2

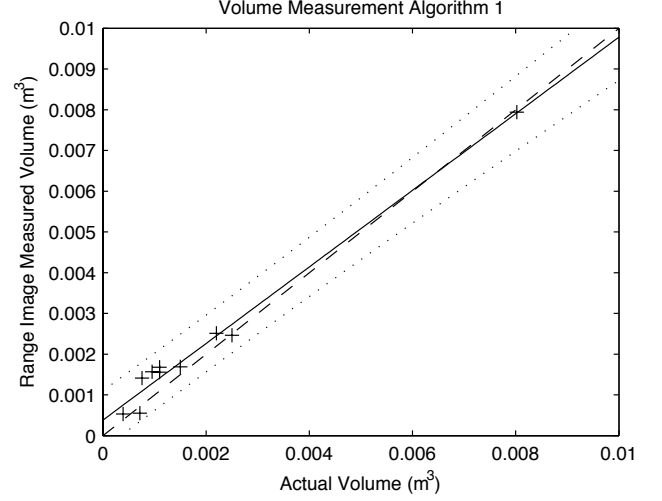


Figure 6: Volume calculation on real covered data when taken from larger height using algorithm. 1.

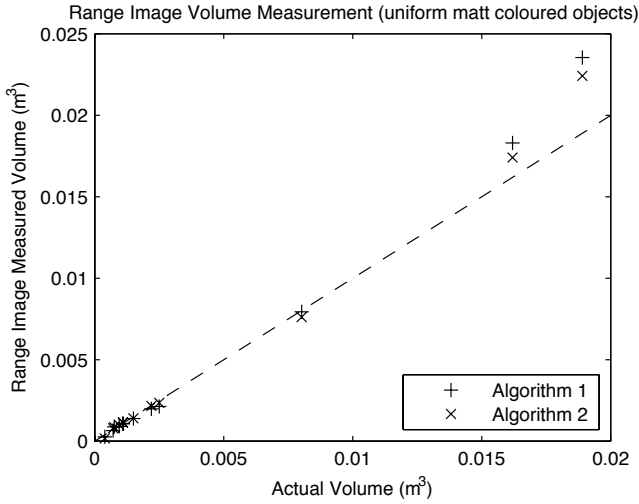


Figure 5: Comparison of two algorithms for uniform matt coloured objects for lesser height of the Camera.

line and offset in second case was found to be more closer to the ideal result than the first algorithm's results. Plot of the comparison for the two algorithms is presented in Figure 5.

To test the robustness of volume measurement, the height of the camera above the table was increased and the captures of the objects repeated. Results were obtained for the uniformly coloured objects and are presented in Figure 6 for the first algorithm. The best fit line of all the points obtained was

$$y = 0.940x + 0.000379772 \quad (13)$$

which shows a little deterioration from the original measurements.

For the second algorithm with camera at higher height,

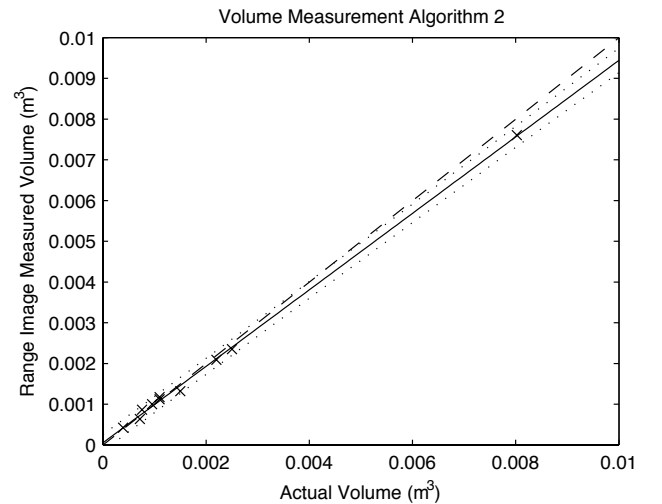


Figure 7: Volume calculation on real covered data for larger height using algorithm. 2.

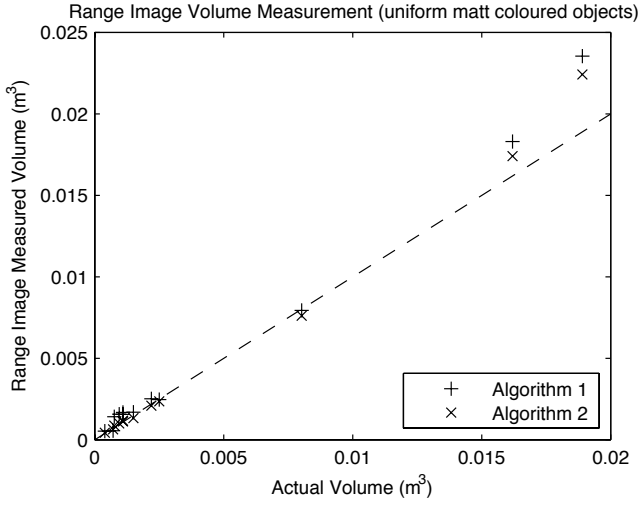


Figure 8: Comparison of algorithms for uniform matt coloured objects for higher height of the Camera.

using the covered data, the behaviour obtained is shown in Figure 7. The equation of best fit line for all the data points is close to that obtained for the lower height of the camera and is given by Eqn. 14.

$$y = 0.939x + 5.18871e - 05 \quad (14)$$

On comparison for the algorithms for higher height of the camera, it was again found that algorithm. 2 gave much accurate results than algorithm. 1. Plot of the comparison is presented in Figure 8.

The volume measurement was repeated on the image data captured of the objects when their somewhat glossy and colour surfaces were exposed. The volumes obtained using both the algorithms are considerably noisy as shown in Figure 9. The results depended on the colour of the surface and so the behaviour gave quite a lot of error in the calculation.

As seen in the Figure.9, the error is high and the results show quite a lot of deviation. It is seen that the algorithms seem to be following the theoretical predictions as expected earlier for the covered data quite nicely. But when it came to uncovered data, the results varied drastically and contained high errors.

5 DISCUSSION

From the above practical experiments and the results observed, quite a number of conclusions can be made. It is seen that the algorithms developed worked quite nicely and gave results quite close to the actual volumes when the range images were taken of objects that are of uniform colour and of matt surfaces. Moreover it is seen that the colour of the sheets used to cover the objects had to be white in order to reduce the errors. In case of multi-coloured and

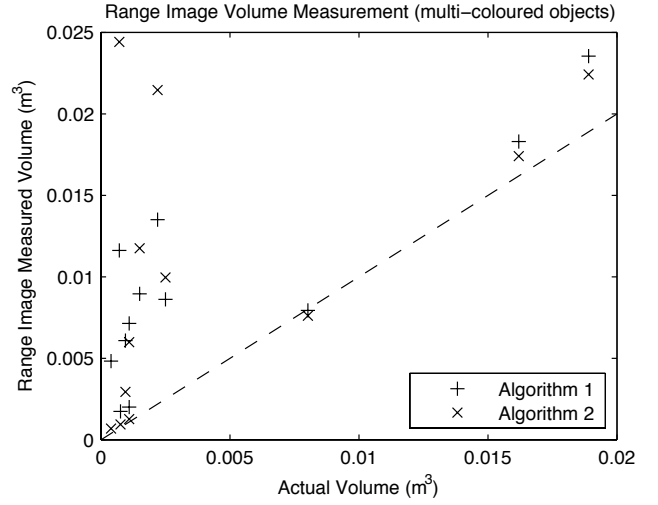


Figure 9: Volume calculation comparison on real uncovered data taken from larger height of the Camera.



Figure 10: Book without shiny surface.

somewhat glossy surfaces, the results deviated far from the ideal and volume measurement was unreliable.

Range imaging data captured with full-field range imaging cameras is subject to multipath effects. This is where light from some other part(s) of the scene makes its way into the wrong pixels, whether by multiple scattering of light within the scene, or scattering of light within the camera optics. This contamination causes phase errors in the active light modulation received back at the camera. The phase error leads to a misdetermination of the range of the object in the view of the pixel. If the view of the object is bright in the pixel then the error introduced into the range is negligible, but if the view of the object is dark in the pixel then the error in ranging can be appreciable, even centimetres in extreme cases. It is likely that this is occurring in the scenes analysed, however the authors do admit they are surprised by the magnitude of the effect on volume measurement and wonder whether other factors are in play.

The errors in volume measurement are reduced immensely when the surface of the objects are covered in uniform white paper. It was also observed that the light coloured books or the less shiny ones gave results closer to the actual volume than the darker and the shiny ones. For example, the experiments performed on the book shown

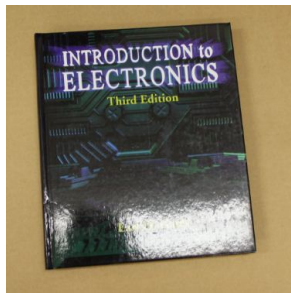


Figure 11: Book with specular reflection due to shiny surface.

in Figure 10 gave the volume that was correct up to three decimal places when using the second algorithm.

In contrast the book shown in Figure 11 with darker shade and noticeable specular reflection of light due to a glossy surface had a large error as measured using the second algorithm. Erroneous volume measurements occur because of the specular reflections occurring when the light from the camera strikes the surface of the object. So the reduction of specular reflections using white sheets as the surface of the objects helps in the reduction of errors.

One more interesting thing can be observed in the above experiments. It can be seen that the algorithms may not provide the correct results for the objects with certain parts coming out of it and not touching the surface like the protrusions bulging out. In such situations the volume measured using the second algorithm will also include the volume of the space between that part and the surface which may introduce a lot of error in the results.

When the object has a hole in it, then again there will be inclusion of some more errors arising due to the boundaries of the hole. But the volume calculation yields closer results. The error in second algorithm again is less and provides closer results.

Therefore, it can be concluded that the above developed algorithms works best with very less error when the surface of the objects are covered with sheets having white colour without any shine. Presence of any one of them will deviate results and will include quite a lot of error in them.

REFERENCES

- [1] M. Imaging, "Swiss ranger sr4000 overview." <http://www.mesa-imaging.ch/prodview4k.php>, accessed on 9th July 2010.
- [2] X. Zhang, J. Morris, and R.Klette, "Volume measurement using a laser scanner," in *Proceedings of Image and Vision Computing New Zealand (IVCNZ'05)*, (Dunedin, New Zealand), pp. 177–182, November 2005.

- [3] A. A. Dorrington, A. D. Payne, and M. J. Cree, "An evaluation of time-of-flight range cameras for close range metrology applications," in *ISPRS Commission V Mid-Term Symposium: Close Range Image Measurement Techniques*, vol. XXXVIII of *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, (Newcastle upon Tyne, UK), pp. 201–206, June 2010.

Proof of Concept of Diffuse Optical Tomography Using Time-of-Flight Range Imaging Cameras

Ahmad Hassan, Rainer Künnemeyer, Adrian Dorrington,
and Andrew Payne

School of Engineering, University of Waikato,
Private Bag 3105, Hamilton, New Zealand.
ash12@waikato.ac.nz.

Abstract:

Diffuse optical tomography is an optical technique to create 3-dimensional images of the inside of highly scattering material. Research groups around the world have been developing imaging systems using various source-detector arrangements to determine optical properties of biological tissue with a focus on medical applications. In this paper we investigate whether a range imaging camera can be used as a detector array. We used time-of-flight range imaging cameras instead of the conventional source-detector array used by others. The results provided in this paper show reconstructed images of absorption and reduced scattering of an object submerged in a tissue simulating phantom. Using the ranging camera XZ422 Demonstrator and the NIRFAST software package, we reconstructed 2D images of a 6 mm metal rod submerged in the centre of a 5 cm deep tank filled with 1% IntralipidTM. We have shown for the first time that range imaging cameras can replace the traditional detectors in diffuse optical tomography.

Keywords:

Diffuse optical tomography, range imaging, NIRFAST

1 INTRODUCTION

Diffuse Optical Tomography (DOT) is a medical imaging techniques that has been under research and development since the early 1990s [1]. It is an alternative to current imaging techniques such as Magnetic Resonant Imaging (MRI) and X-ray computed tomography. DOT is non-invasive, non-ionising, low cost and also has an order of magnitude faster temporal resolution than that of (MRI) [2]. For example, DOT has the ability of producing information about the oxy-haemoglobin and deoxy-haemoglobin concentrations by measuring the optical properties (absorption and reduced scattering) of tissues, and hence it is being used in functional brain imaging and optical mammography [1-3].

There are three types of DOT measurement techniques: time domain, continuous wave domain and frequency domain [3, 4]. The time domain technique passes pulsed light through the area under test and measures the time delay between the incident and received pulse. The continuous wave technique measures only the attenuated light intensity passed through the tissue, whereas the frequency domain technique passes intensity modulated light through the area under test and measures both phase and intensity of the received modulated light relative to the incident [5].

Many research groups around the world have been developing DOT systems using different measurement techniques, source-detector arrangement and reconstruction algorithm. The frequency domain technique has become very popular over the recent years. The basic components needed are modulated light sources placed at various locations and discrete detectors which measure phase and intensity relative to the source modulation. In this research we present a proof of concept DOT system using time-of-flight 3D range imaging cameras and an open source reconstruction software (NIRFAST) [6] to detect and reconstruct images of objects submerged in tissue-simulating phantom.

2 BACKGROUND

2.1 Time-of-Flight Range Imaging

Time-of-flight range imaging is a technique used by several off-the-shelf camera modules such as those developed by Canesta Inc. (Sunnyvale, California, USA). Each of these camera modules work by illuminating a scene with a modulated light source, then measures the time it takes for the modulated light to reach an object and reflects back to the camera. The time delay is indi-

rectly determined by the phase shift created between the reflected and the incident modulated light. Knowing the speed of light c and the modulation frequency f_m (hence the modulation wavelength λ_m), and the phase shift ϕ , the distance d on each pixel can be calculated as [7]:

$$d = \frac{\phi c}{4\pi f_m} = \frac{\phi \lambda_m}{4\pi} \quad (1)$$

The intensity or active brightness can also be calculated for each pixel using Equation 2, where A_0 , A_1 , A_2 , A_3 are pixel brightness values recorded at phase offsets of 0° , 90° , 180° , and 270° respectively.

$$Intensity = \frac{\sqrt{(A_0 - A_2)^2 + (A_1 - A_3)^2}}{2} \quad (2)$$

2.2 Optical Properties of Biological Tissues

2.2.1 Absorption and Scattering

When light travels in biological tissues it exhibits multiple scattering and absorption. Therefore, biological tissues are considered turbid media. These phenomena (absorption and scattering) occur due to the physical properties of the turbid medium. As light enters a turbid medium, it is scattered by particles, biological cells, or other discontinuities in refractive index. This diffusion causes the photons to take multiple paths. In addition, some photons are absorbed by molecules, causing the light to be attenuated as it propagates through the medium (Figure 1).

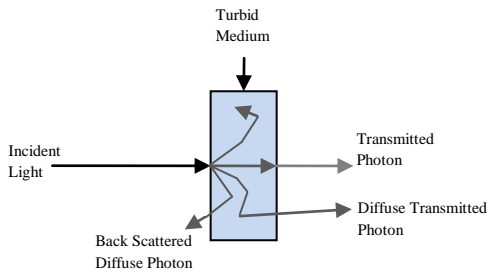


Figure 1: Light scattering and absorption in turbid medium.

Penetrating deep tissues is the main limitation in optical tomography due to the strong attenuation caused by absorption. However, absorption in the near-infrared region (between 610 to 910 nm) is orders of magnitude lower than in the visible spectrum [9, 10].

When a scattering medium is illuminated by a modulated light source, the light intensity and phase change with distance from the source. The slope of the amplitude of the AC component, I_{AC} , and the slope of the phase, ϕ , is given in Equation 3 and 4 [11].

$$I_{AC} = -\sqrt{\frac{3}{2} \mu_a (\mu_a + \mu'_s)} \left[\sqrt{1 + \left(\frac{\omega}{v \mu_a} \right)^2} + 1 \right]^{\frac{1}{2}} \quad (3)$$

$$\phi = \sqrt{\frac{3}{2} \mu_a (\mu_a + \mu'_s)} \left[\sqrt{1 + \left(\frac{\omega}{v \mu_a} \right)^2} - 1 \right]^{\frac{1}{2}} \quad (4)$$

where ω is the angular frequency of the modulation, v the speed of light in the medium, μ_a the absorption, and μ'_s the reduced scattering coefficient.

When the slopes of AC intensity and phase are known, the equations can be solved for the absorption and scattering properties of the turbid medium.

2.2.2 Tissue-Simulating Phantom

Tissue-simulating phantoms are materials used in optics research labs. They exist in solid and liquid forms. IntralipidTM is a commercially available milky solution with concentration 10%, 20% and 30%. It contains Soybean oil, Lecithin, Glycerine and water, and it is normally used in hospitals to feed weak patients. It is also used to simulate body tissue when conducting experiments. The absorption coefficient of IntralipidTM at 1% concentration is very similar to water and ranges from 0.02 cm^{-1} to 0.3 cm^{-1} . However, the scattering coefficient is approximately constant (10 cm^{-1}) over the near-infrared wavelength range [11].

3 METHOD

In this study a 6 mm metal rod was submerged in a transparent plastic tank filled with 1% IntralipidTM. A range imaging camera was placed facing one end of the tank, and a laser beam was shone on the other side.

3.1 Hardware Setup

The hardware setup shown in figure 2 consists of:

- Range imaging camera (XZ422 Demonstrator) made by Canesta Inc. (Sunnyvale, California, USA) with image resolution of 160×120 .
- Near-infrared laser, 850 nm, with an average output power of 50 mW. The laser beam was focused to a 3mm spot on the target.
- Clear plastic tank made out of acrylic sheet filled with tissue-simulating phantom (1% IntralipidTM).

- 6 mm diameter metal rod submerged in the tank.

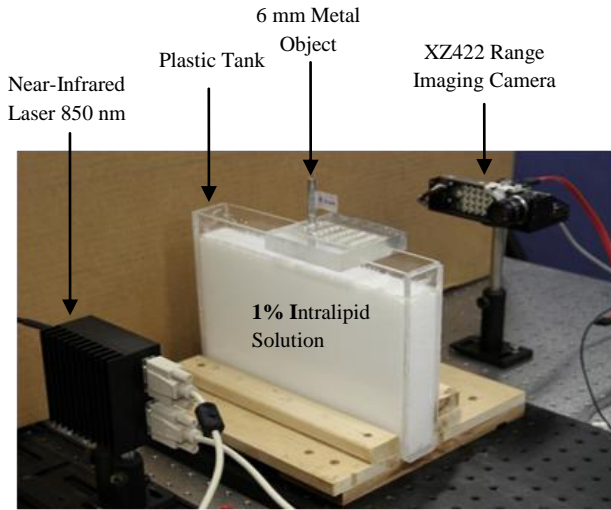


Figure 2: Hardware setup.

The laser beam is collimated and shone onto the tank from one end. Light will then diffuse as it enters the tank. Some of the light propagates through to the other side of the tank. These transmitted photons are captured by the camera and ultimately carry phase and intensity information about the submerged object.

3.2 Processing and Reconstruction

MATLABTM was used for processing the data produced by the camera (active brightness and range). However, image reconstruction is done using NIRFAST (Near-Infrared Florescence and Spectral Tomography) [6]. NIRFAST is a software package which simulates light propagation in biological tissues based on the finite element method. It can process single wavelength absorption and reduced scattering, multi-wavelength and also fluorescence models [6]. In this study we only used the 2D standard single wavelength absorption and reduced scattering image reconstruction tool. Hence, we used a 2D mesh to simulate a horizontal slice of our tank as shown in figure 3.

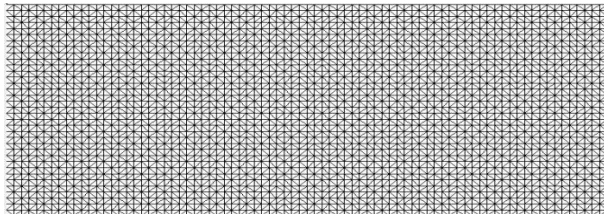


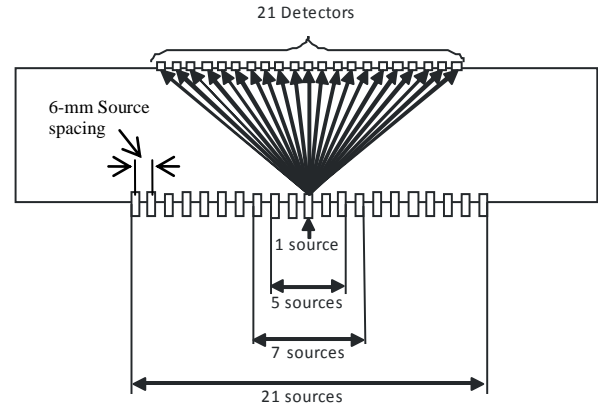
Figure 3: 2D mesh created by NIRFAST

Four experiments were repeated using one, five, seven and twenty-one sources with 6 mm spacing, keeping the number of detectors fixed (twenty-one detectors). The detectors were configured on the range image by choosing pixels across the captured images.

The sources are turned on one after the other and data is recorded for a set of detectors. For the three cases (one-source, five-sources and seven-sources) the data is obtained from all the 21-detectors. In contrast, the data collected from the 21-sources were obtained from only 14-detectors (Figure 4).

(a)

Case 1, 5 and 7 sources, all the detectors are used by all the sources



(b)

Case 21-sources, only 14-detectors are used by all the 21-sources

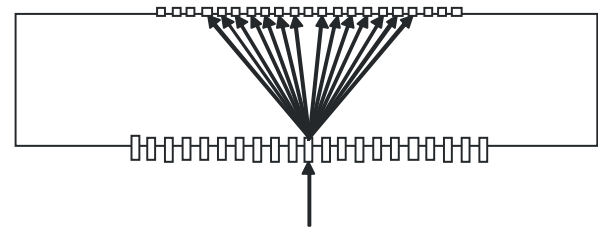


Figure 4: Source-detector configuration, (a) in case 1, 5 and 7 sources, (b) in case 21 sources.

3.3 Calibration

The software was used to generate simulated results of phase and intensity. These results were then compared with the results produced by our range imaging camera. Due to phase delay in the camera electronics, the software and the camera have different scales, therefore, the phase and intensity data produced by the camera had to be fitted to match the software data. This was really important because one of the software requirements in reconstructing images of absorption and reduced scattering is the phase and intensity data obtained from the camera. This data fitting process was achieved by the polynomial fitting tool box available in MATLABTM.

4 RESULTS AND DISCUSSION

4.1 Calibration

Figure 5 shows a plot of phase and intensity obtained from the camera and NIRFAST [6]. Clearly, there is an offset between the camera and NIRFAST scales. This issue was resolved by using a second degree polynomial function to fit the data as shown in figures 6 and 7. The result is shown in figure 8, where the camera and software functions are fitted nicely.

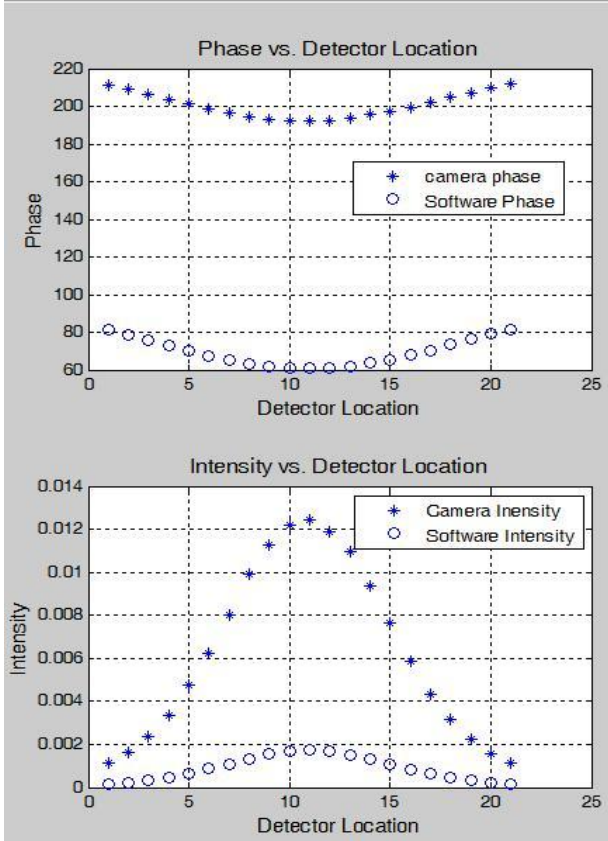


Figure 5: Un-calibrated phase and Intensity functions.

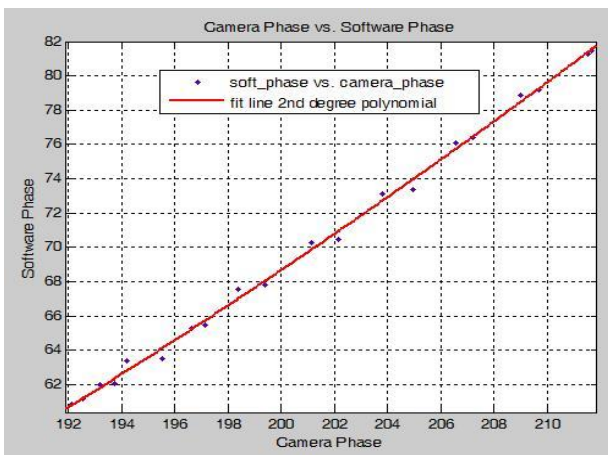


Figure 6: 2nd degree poly-fit of phase function.

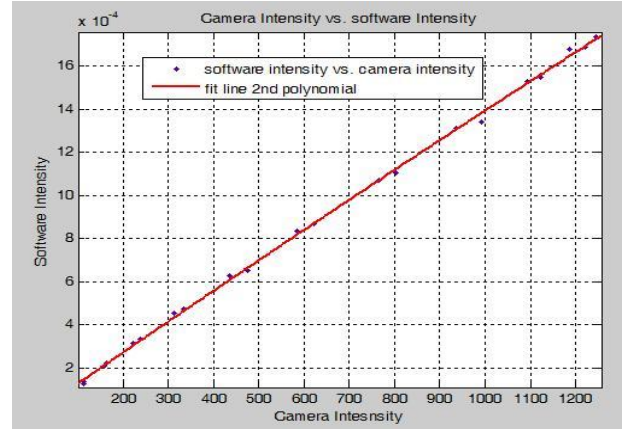


Figure 7: 2nd degree poly-fit of intensity function.

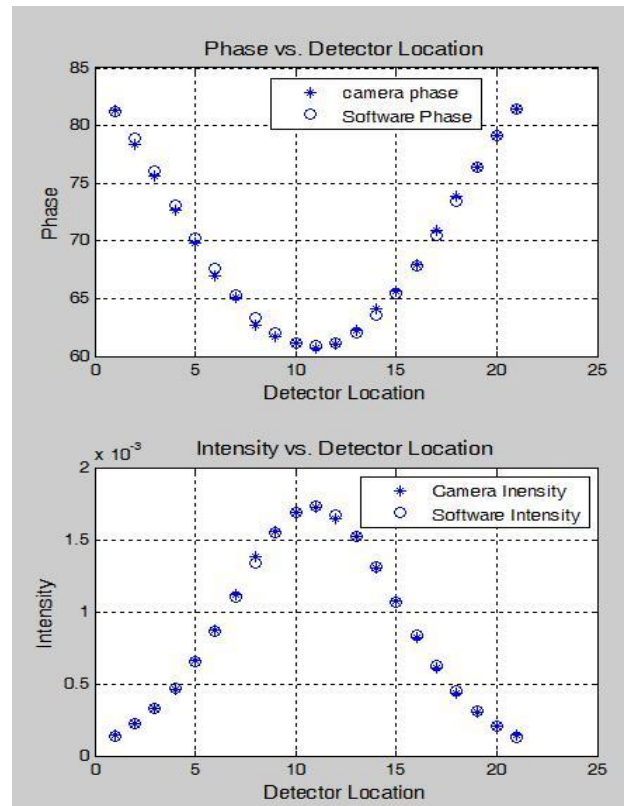


Figure 8: Calibrated phase and intensity functions.

4.2 Reconstruction Results

Figure 9 shows a photographic image of the location of the object. Figure 10, in contrast, shows a sequence of reconstructed images with different number of sources. The reduced scattering images μ_s show almost no evidence of the object as it is considered a black object when submerged in the tank, and therefore no scattering has occurred. However, the absorption images μ_a show an outline of the object in all the cases, but the reconstructed image (figure 10d) is the closest estimation of object's size and location when compared with the real image (figure 9).

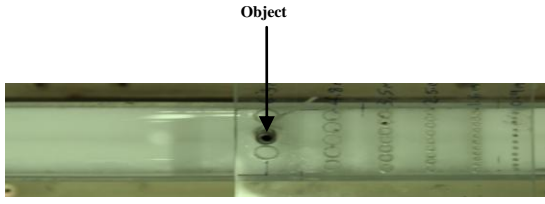


Figure 9: Photographic image (top view) of the tank with the object submerged.

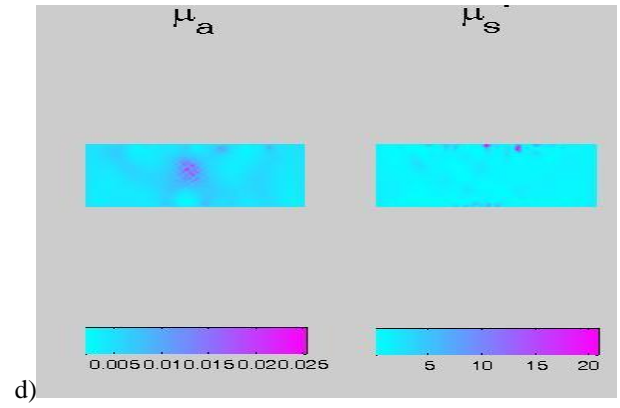
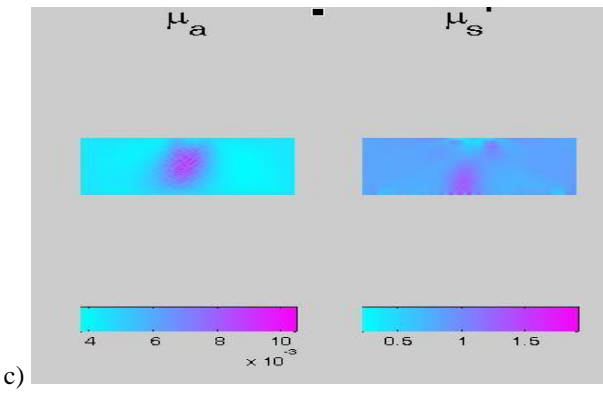
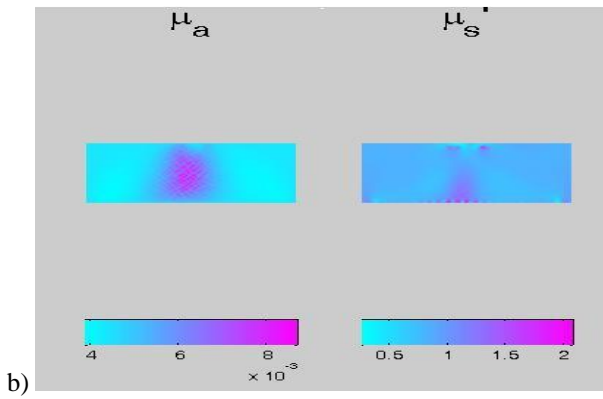
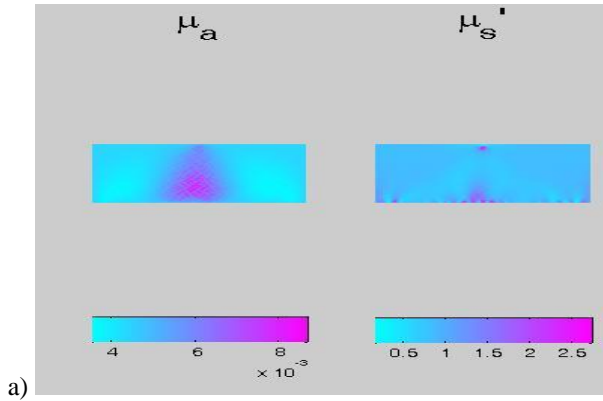


Figure 10: Reconstructed images of absorption, μ_a , and reduced scattering, μ_s (a) with a single source (b) with 5 sources (c) with 7 sources (d) with 21 sources.

These results indicate that increasing the number of sources will improve the reconstructed image, i.e., more information is obtained about the object from multiple angles as shown in figure 11. However, the source location has not been optimised in this work, and will be considered in future work.

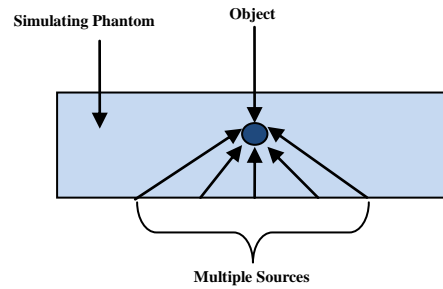


Figure 11: Top view model showing how light is obtaining spatial information about the object.

5 FURTHER WORK

In this study it was found that diffuse optical tomography is possible using 3D range imaging technology. However, more research is required to characterize this imaging system. The results highlighted that multiple sources are required to perform diffuse optical tomography. The optimum number of sources is yet to be determined with further experiments. Furthermore, NIRFAST incorporates 3D meshes which potentially can be used to reconstruct 3D images of tissue optical properties.

6 CONCLUSIONS

We have demonstrated experimentally as proof of concept the practicality of using time-of-flight range imaging cameras for non-contact signal detection in DOT medical imaging. Reconstructed images of absorption and reduced scattering are presented in this research. These results show that multiple sources are required to obtain spatial information of objects submerged in a simulating phantom. Although further investigation is needed, this could be the starting point to further develop the current set up to a fully integrated diffuse optical tomography system.

7 REFERENCES

- [1] D. Piao, H. Dehghani, S. Jiang, S. Srinivasan, and B. W. Pogue, "Instrumentation for video-rate near-infrared diffuse optical tomography," *Review of Scientific Instruments*, vol. 76, pp. 1-13, 2005.
- [2] N. Cao and A. Nehorai, "Tumor localization using diffuse optical tomography and linearly constrained minimum variance beamforming," *Opt. Express*, vol. 15, pp. 896-909, 2007.
- [3] A. Corlu, R. Choe, T. Durduran, K. Lee, M. Schweiger, S. R. Arridge, E. M. C. Hillman, and A. G. Yodh, "Diffuse optical tomography with spectral constraints and wavelength optimization," *Appl. Opt.*, vol. 44, pp. 2082-2093, 2005.
- [4] C. Li, S. R. Grobmyer, L. Chen, Q. Zhang, L. L. Fajardo, and H. Jiang, "Multispectral diffuse optical tomography with absorption and scattering spectral constraints," *Appl. Opt.*, vol. 46, pp. 8229-8236, 2007.
- [5] S. J. Erickson and A. Godavarty, "Hand-held based near-infrared optical imaging devices: A review," *Medical Engineering and Physics*, vol. 31, pp. 495-509, 2009.
- [6] H. Dehghani, M. E. Eames, P. K. Yalavarthy, S. C. Davis, S. Srinivasan, C. M. Carpenter, B. W. Pogue, and K. D. Paulsen, "Near infrared optical tomography using NIRFAST: algorithm for numerical model and image reconstruction," *Communications in Numerical Methods in Engineering*, vol. 25, pp. 711-32, 2009.
- [7] A. A. Dorrington, M. J. Cree, A. D. Payne, R. M. Conroy, and D. A. Carnegie, "Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera," *Measurement Science and Technology*, vol. 18, pp. 2809-2816, 2007.
- [8] A. A. Dorrington, C. D. B. Kelly, S. H. McClure, A. D. Payne, and M. J. Cree, "Advantages of 3D time-of-flight range imaging cameras in machine vision applications," in *Proceeding of the 16th Electronics New Zealand Conference (ENZCon), Dunedin, New Zealand*, Dunedin, New Zealand, 18-20 November, 2009, pp. 95-99.
- [9] C. Balas, "Review of biomedical optical imaging-a powerful, non-invasive, non-ionizing technology for improving in vivo diagnosis," *Measurement Science & Technology*, vol. 20, p. 104020 (12 pp.), 2009.
- [10] B. W. Pogue, S. C. Davis, S. Xiaomei, B. A. Brooksby, H. Dehghani, and K. D. Paulsen, "Image analysis methods for diffuse optical tomography," *Journal of Biomedical Optics*, vol. 11, pp. 33001-1, 2006.
- [11] B. Cletus, 2010, "Wavelength Tuneable Frequency Domain Photon Migration Spectrometer for Tissue-like Media", PhD thesis, University of Waikato, Hamilton, New Zealand.

Performance Analysis of Microstrip Line Matching Network for 866 MHz UHF LNA

Jie Li¹ and Rezaul Hasan²

Center for Research in Analog and VLSI Microsystems Design (CRAVE),
School of Engineering and Advanced Technology (SEAT)
Massey University, Albany, Auckland

Email: ¹J.Li.2@massey.ac.nz ²S.M.Hasan@massey.ac.nz

Abstract: This paper presents a method to design the matching networks by using single stub microstrip line for an UHF LNA that operates on 866 MHz. The superior performance characteristics of the Microstrip line make it one of the most important medium of transmission in microwave transistor amplifiers in both microwave integrated-circuit and PCB-circuit technology. The parallel-plate microstrip line configuration is primarily used for low impedance, high-power applications. MATLAB programming is used to obtain the Smith chart results for input and output Microstrip line matching networks. The LNA simulation result exhibits a voltage gain of 17 dB, drawing 1.8 mW power consumption from a 1V power supply. .

Keywords: Microstrip Line, Impedance Matching, UHF, LNA, MATLAB

I. INTRODUCTION

Matching circuits analyse can be troublesome at microwave frequencies. Ideally, an optimum matching network is lossless, which in another word, the impedance seen into the matching network Z_0 is equivalent to the load Z_L to eliminate reflections, especially at tuning frequency. This allows maximum power delivery and minimize the signal noise of the system. Conventional matching networks contain a number of lump elements. However, the influence of parasitic in those discrete elements becomes more noticeable when the frequency increases. Especially, inductors usually are avoided in such design since they tend to have higher resistive losses than capacitors. Alternatively, distributed components are widely been used when wavelength becomes significantly small comparing with the characteristic circuit component length. Among all transmission line technologies, microstrip line is widely applied. Fig. 1 demonstrates a traditional parallel-plate microstrip line. The conducting strips are etched on one side of the PCB, while the other side of the PCB is usually treated as ground plane to help preventing excessive field leakage.

This structure reduces radiation loss [1]. The current and voltage flow is confined to two plates separated by a dielectric medium.

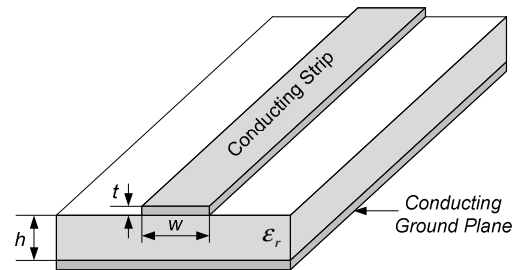


Fig. 1: Microstrip transmission line geometry

The shortcomings of this kind of single layered PCB are high radiation loss and interference between neighbouring conductors traces. Hence substrates with high dielectric constants ϵ_r is prefer for achieving high board density of the components layout since they minimize field leakage and cross coupling. Multilayer techniques are also widely used to achieve balanced circuit board design, with the increasing on cost and the higher complex of the fabrication process instead.

The Smith Chart, even was developed in early 20th century, has retained its popularity due to its easy and intuitive

tive display of the reflection coefficient as well as the line impedance. The reading accuracy from the Smith Chart is sufficient for most practical microwave transistor amplifier design issue. Matching circuit design for amplifier can be easily and quickly approached by using the normalized impedance and admittance Smith Chart from computer-aided software [2, 3]. Furthermore, the Smith Chart is also used to present the frequency dependence of scattering parameters and other characteristics [4]. In this paper, a complete procedure of constructing an impedance matching network for an UHF LNA by using transmission microstrip line associate with the Smith Chart is approached.

II. SOME USEFUL EQUATIONS

[4-6] defines all the formulas and equations used in this paper. For the purposes of obtaining fast and generally reliable estimations of the line parameters, the thickness t of the conductor forming the line is assumed negligible compared to the substrate height h . In this case, the empirical formulas that depend only on the line dimensions (w and h) and the dielectric constant ϵ_r can be adopted. Depending on the ratio w and h , two separate regions applies.

For narrow strip lines, $w/h < 1$, the line impedance can be represented in [4-6]

$$Z_0 = \frac{Z_f}{2\pi\sqrt{\epsilon_{eff}}} \ln\left(8\frac{h}{w} + \frac{w}{4h}\right)$$

where $Z_f = \sqrt{\mu_0/\epsilon_0} = 376.8\Omega$ is the wave impedance

in free space, and ϵ_{eff} is the effective dielectric constant

that was given by

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[\left(1 + 12\frac{h}{w}\right)^{-1/2} + 0.04\left(1 - \frac{w}{h}\right)^2 \right]$$

while for a wide line when $w/h > 1$, Z_0 has:

$$Z_0 = \frac{Z_f}{\sqrt{\epsilon_{eff}} \left[1.393 + \frac{w}{h} + \frac{2}{3} \ln\left(\frac{w}{h} + 1.444\right) \right]} \quad (1)$$

$$\text{with } \epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + 12\frac{h}{w}\right)^{-1/2} \quad (2)$$

There is only a small discontinuity when $w/h = 1$. For practical design purposes, w/h ratios that based on a given characteristic impedance Z_0 and dielectric constant ϵ_r of the substrate is approached as well. Assuming in an infinitely thin line conductor, this ratio can be written in a form of

$$\frac{w}{h} = \frac{8e^A}{e^{2A} - 2} \quad \text{for } \frac{w}{h} \leq 2 \quad (3)$$

$$\frac{w}{h} = \frac{2}{\pi} \left\{ B - 1 - \ln(2B - 1) + \frac{\epsilon_r - 1}{2\epsilon_r} \left[\ln(B - 1) + 0.39 - \frac{0.61}{\epsilon_r} \right] \right\} \quad \text{for } \frac{w}{h} \geq 2$$

$$\text{where } A = 2\pi \frac{Z_0}{Z_f} \sqrt{\frac{\epsilon_r + 1}{2}} + \frac{\epsilon_r - 1}{\epsilon_r + 1} \left(0.23 + \frac{0.11}{\epsilon_r} \right) \quad (4)$$

$$B = \frac{Z_f \pi}{2Z_0 \sqrt{\epsilon_r}}$$

The achieved effective dielectric can further evaluate the phase velocity, wavelength and the related wave number to the phase velocity of the microstrip line in following expressions:

$$v_p = \frac{1}{\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{\mu_0\epsilon_0\epsilon_{eff}}} = \frac{c}{\sqrt{\epsilon_{eff}}} \quad (5)$$

$$\lambda = \frac{v_p}{f} = \frac{c}{f\sqrt{\epsilon_{eff}}} = \frac{\lambda_0}{\sqrt{\epsilon_{eff}}} \quad (6)$$

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi f}{v_p} \quad (7)$$

where c is the speed of light.

III. MATCHING NETWORK IMPLEMENTATION

A generic LNA matching network configuration is shown in Fig. 2. Input and output matching networks are needed to reduce undesired reflections to improve the power capabilities. In this work, matching networks are designed for a prefabricated CMOS cascode LNA [7] with known VSWR, measured input power reflection S_{11} and output power reflection S_{22} as shown in Table I. A matching network that consists of a series microstrip transmission line connected to a parallel open-circuit or short circuit stub is used for standard port impedance matching (50Ω) [5]. Four adjustable parameters are counts to consider or calculate in this case: length l_s and characteristic impedance Z_{0S} of the stub (either open- or short-), and length l_T and characteristic impedance Z_{0T} of the series transmission line segment.

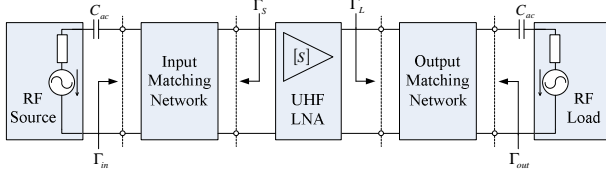


Fig. 2: Generic LNA with matching network.

TABLE I: Typical parameters of a pre-fabricated LNA and its expected result at 866 MHz, $V_{dc}=1V$, $Z_0 = 50\Omega$

Operating Frequency (MHz)	866
Noise Figure (dB)	2
VSWR	1.5
Measured input power reflection S_{11}	$0.6886 - j0.4774$
Measured output power reflection S_{22}	$0.82526 + j0.289$
Assumed reverse isolation S_{12}	≈ 0
Expected Power gain S_{12}	$>15\text{dB}$

In order to achieve maximum unilateral transducer gain, where signal flows only in one direction, we make the assumption of reverse isolation S_{12} approaching 0. From the defining equation for S-parameters [8], the source and load reflection coefficient towards the LNA needs to

become the complex conjugate of S_{11} and S_{22} . The matching procedure can be quite confusing without a very clear train of thought. The general method used in matching circuit is to choose either “toward the load” or “toward the generator” on Smith chart depending on the output or input matching network respectively. The source reflection coefficient needs to take the conjugate of S_{11} while the load reflection coefficient doesn’t [9]. Transformation under conditions brings up the error. The method used in this paper follows a unified procedure that can be adopted in both input and output matching network. Take both source and load reflection coefficient looking from the amplifier back toward the input or output line, which is the complex conjugate of S_{11} and S_{22} . Only one direction “toward to the load” from the reflection coefficient points on the Smith chart need to focus.

Therefore, the source reflection coefficient can be represented as $\Gamma_s = S_{11}^* = 0.6886 + j0.4774$ for an input matching, or $S_{11}^* \approx 0.8379 \angle +34.73^\circ$ in angular format. Similarly, the load reflection coefficient is $\Gamma_L = S_{22}^* = 0.82526 - j0.289$ for output matching, or represented in angular form as $0.87445 \angle -19.308^\circ$. Converting into impedance format can be easily approached. Hence source impedance format now can be calculated as $Z_s \approx 45.9 + j147$, or $z_s = 0.918 + j2.94$ in normalized format, and load impedance $Z_L \approx 103 - j253$, or $z_L = 2.06 - j5.06$ in normalized format. In order to better understand and utilize the use of Smith Chart, one more transformed representation step is approached to find the normalized admittances, that can be expressed as $y_s = 0.0968 - j0.3099$ and $y_L = 0.0690 + j0.1695$.

Input Matching Network

The basic concept for input impedance matching is to select the length l_s of the stub such that it produces a susceptance B_s sufficient to move the normalized input admittance $y_{in} = 1$ (or $z_{in} = 1$) to SWR circle that passes through the normalized source impedance point $z_s = 0.918 + j2.94$, as shown in Fig. 3.

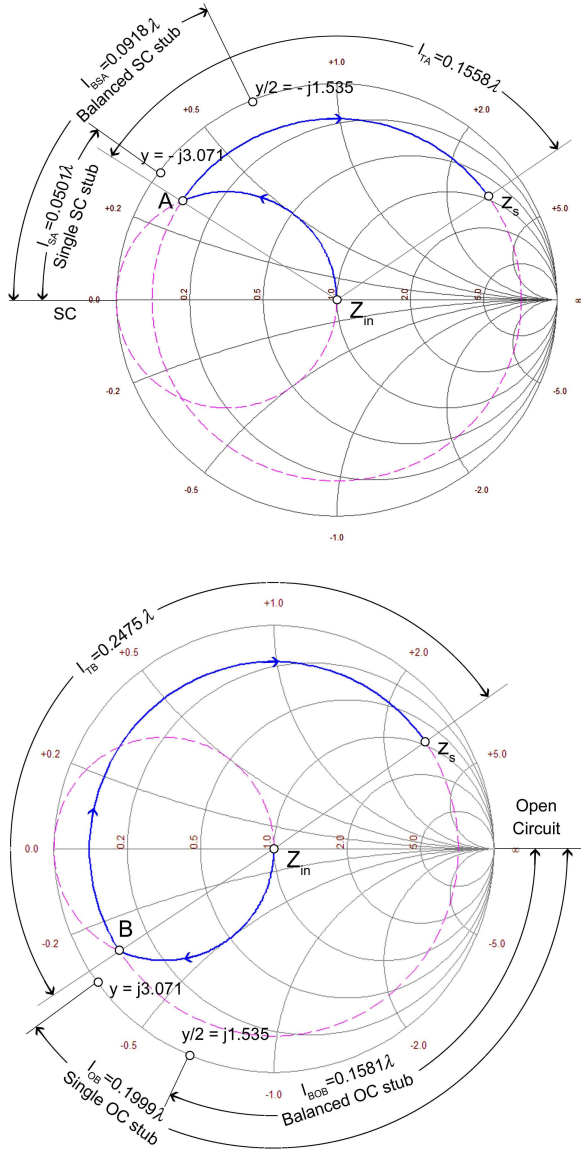


Fig. 3: Analysis of two possible (a) short-circuit single-stub and (b) open-circuit single-stub matching networks for input matching system.

To start with input impedance matching in Fig. 2, the source impedance point $z_s = 0.918 + j2.94$ is located in Smith Chart first. It can be identified in the Smith Chart as the intersection of the circle has radius 0.8379 with angular -34.73° shown as point z_s in Fig. 3. It is noticeable that the source SWR circle associated with $z_s = 0.918 - j2.94$ intersects the constant conductance circle $g=1$ in two points (at $y_A = 1 - j3.071$ and at $y_B = 1 + j3.071$) suggesting two possible solutions. The corresponding susceptance values for the stub are $jb_A = y_A - y_{in} = -j3.071$ and $jb_B = y_B - y_{in} = j3.071$, respectively. In the first case, the length of an short-

circuit stub can be found in the Smith Chart by measuring the length l_{SA} , starting from the $\Gamma_0 = -1$ point on the real axis (short circuit), and moving along the outer perimeter of the Smith Chart $g=0$ toward the generator (clockwise) to the point where $y = -j3.071$. The length in this case is $l_{SA} = 0.0501\lambda$. The short-circuit stub can be replaced by an open-circuit stub if its length is increased by a quarter wavelength $l_{OA} = 0.3001\lambda$. This can be also proved from the formula $\frac{\lambda}{2\pi} [\tan^{-1}(-3.071) + \pi] = 0.3001\lambda$.

This substitution may become necessary in printed circuit design, open-circuit stubs are sometimes preferred because they eliminate the deployment of a via, which is otherwise necessary to obtain the ground connection for a short-circuit stub. This through-hole connection in printed circuit boards will result additional parasitic inductances. However, at high frequencies, it is very difficult to maintain perfect open-circuit (as well as short-circuit) conditions because of changing temperatures, humidity, and other parameters of the medium surrounding the open transmission line. Moreover, in order to minimize the circuit layout area, the first priority is to select the shortest line segments.

Studying from the open-circuit as shown in Fig. 3(b), very similar to the first solution, $jb_B = j3.071$ yields the length $l_{OB} = 0.1999\lambda$ for the open-circuit stub, and $l_{SB} = 0.4499\lambda$ for the short-circuit stub. In this case, creating an open-circuit stub requires a shorter length than a short-circuit stub.

The length of the series transmission line segment can be approached by the phase rotation along the SWR circle from either point A or B on the Smith Chart “toward generator” to the Z_s (or sometime it’s preferable to say from Z_s “toward load” to either point A or B on the Smith Chart). It is easy to get both transmission line segment length $l_{TA} = (0.2018 - 0.0460) = 0.1558\lambda$ and $l_{TB} = (0.2018 + 0.0460) = 0.2478\lambda$ respectively.

Output Matching Network

To approach the output impedance matching network shown in Fig. 2, similarly locating $z_L = 2.06 - j5.06$ in the Smith Chart as shown in Fig. 4. Two intersect point C and D are also detect from the corresponded load SWR circle and the constant conductance circle $g=1$. The corresponding susceptance values for the stub are $y_C \approx 1 - j3.58$ and $y_D \approx 1 + j3.58$. The corresponding length of the short-circuit stub and open-circuit stub are found in the same way as $l_{SC} = 0.0432\lambda$, $l_{SD} = 0.4569\lambda$ and $l_{OC} = 0.2931\lambda$, $l_{SC} = 0.2069\lambda$ respectively. The shortest length of the series microstrip transmission line segment plus stub is obtained by using a short-circuit shunt stub. The length of two series transmission line segments are found by moving from point C and C on the Smith Chart “toward generator” to the z_L as $l_{TA} = 0.2365\lambda$ and $l_{TC} = 0.31718\lambda$.

Single stub to Balanced stub Network

To minimize the transmission interaction between shunt stub and transmission line segment, the shunt stub must be balanced along the transmission line. For this, admittance of each side of balanced shunt stub must be equal to half of the total admittance.

For input matching network, admittance of each side of balance stub will be equal to $\pm j3.071/2 = \pm j1.5355$. Using the Smith Chart in Fig. 3, the length of each side of stub $l_{BSA} = 0.0918\lambda$ for balanced short-circuit stub from point A and $l_{BOB} = 0.1581\lambda$ for balanced open-circuit stub from point B. It is noticeable that the length for balanced open-circuit stub is reduced from the single open-circuit stub instead of increasing in short-circuit stub situation. Similar approaching is processed for output matching network in Fig. 4., admittance value $\pm j3.58/2 = \pm j1.79$ of each side of balance stub lead to the length of each side of stub $l_{BSC} = 0.0806\lambda$ for balanced short-circuit stub from point C and $l_{BOD} = 0.1693\lambda$ for balanced open-circuit stub from point D.

The complete LNA matching schematic diagram is shown in Fig. 5 and Fig. 6 for single and balanced stub.

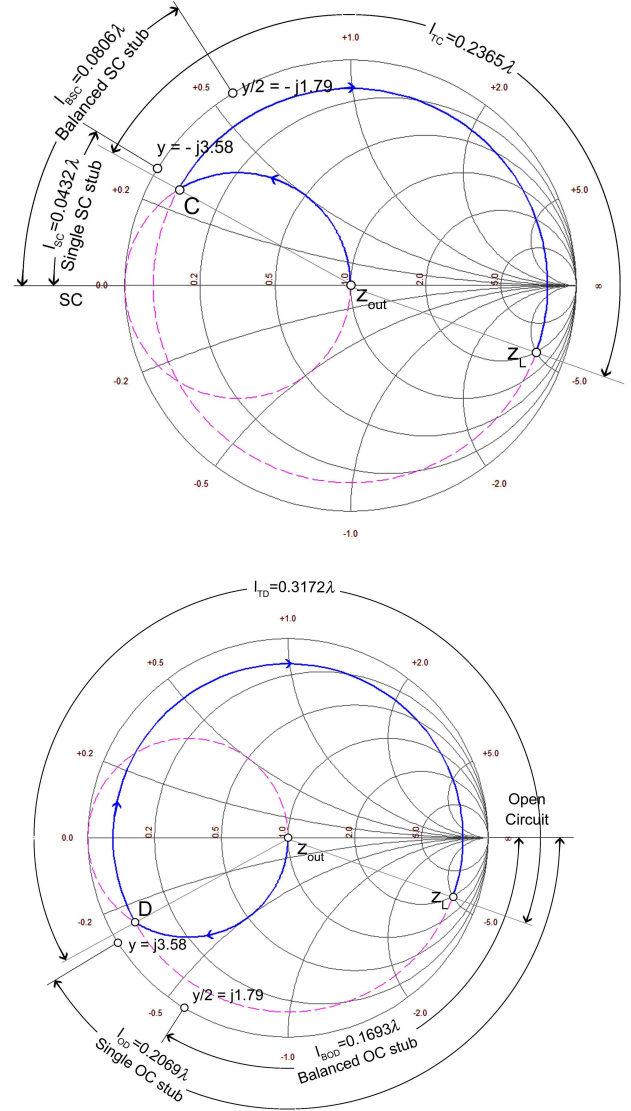


Fig. 4: Analysis of two possible (a) short-circuit single-stub and (b) open-circuit single-stub matching networks for output matching system.

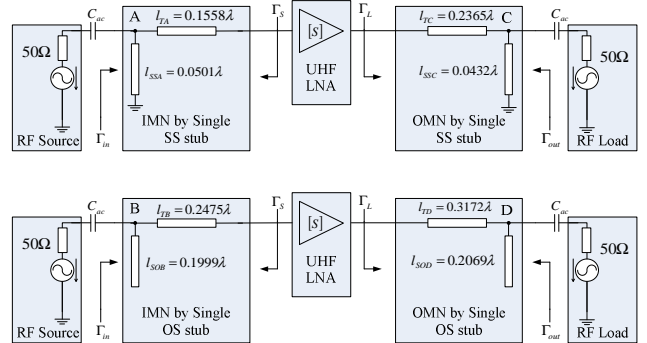


Fig. 5: Complete LNA matching schematic using single short-circuit and open-circuit stub

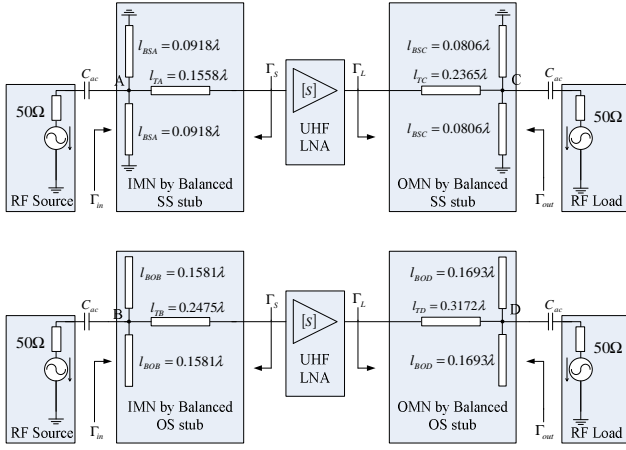


Fig. 6: Complete LNA matching schematic using balanced short-circuit and open-circuit stub

IV. RESULT

MATLAB is used to confirm the matching network. The code in section VI demonstrates one example for both input and output matching by using single open-circuit microstrip transmission line stubs. The MATLAB simulated Z_{in} and Z_{out} values after the matching networks are very close to the desired Z_S and Z_L values as shown at the bottom of the code. Fig. 3 and Fig. 4 are the output figures after running the MATLAB codes.

Here a 50 Ω characteristic Microstrip line is desired to construct for this research. The relative dielectric constant ϵ_r and the board height w are provided by FR4 PCB foundry in an approximated value of 4.6 and 1.45mm, respectively [6]. For a wide line, $w/h > 1$, a value of 3.4575 is obtained for the effective dielectric constant from (2). The more precise computed value of the characteristic impedance is 50.2243 Ω from (1). Therefore, by choosing the case where $w/h \leq 2$ from (3) and (4), a value of 1.5583 and 1.8477 is calculated for A and w/h ratio. Using the obtained ratio of w/h , the trace width equals to 2.698mm. From equations (5-7), the corresponding phase velocity, wave length and related wave number can be found as 1.61×10^8 m/s, 0.1859m and 33.7965 m^{-1} , respectively. The value of

$\lambda/10$ is less than 20 cm, which is comparable with typical circuit component dimensions. Fig. 7 shows one example of input / output impedance matching network by using single open-circuit stub from figure 8 (b).

Fig. 8 shows the s-parameter results of this impedance matching network from Fig. 7 by using TDR calibration of S-parameters. With a power supply of 1V and 50 Ω port resistor, at turning frequency, the input reflection coefficient S_{11} gives a value around -30 dB; the forward power gain S_{21} is around 17 dB. The power dissipation is approximate 1.8 mW with 1.8 mA drain current through the LNA transistor. Measured S-parameter values in line with our expected results, proves that the impedance matching method used here is appropriate.

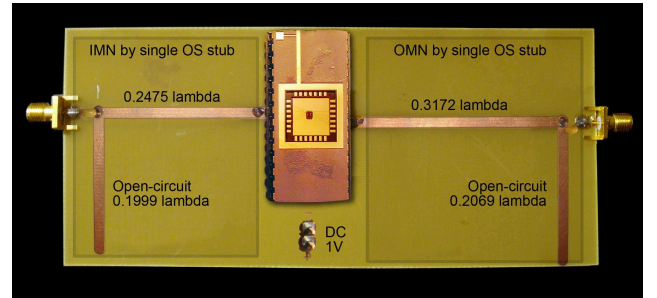


Fig. 7: Impedance matching network by both using single open-circuit microstrip transmission line stubs

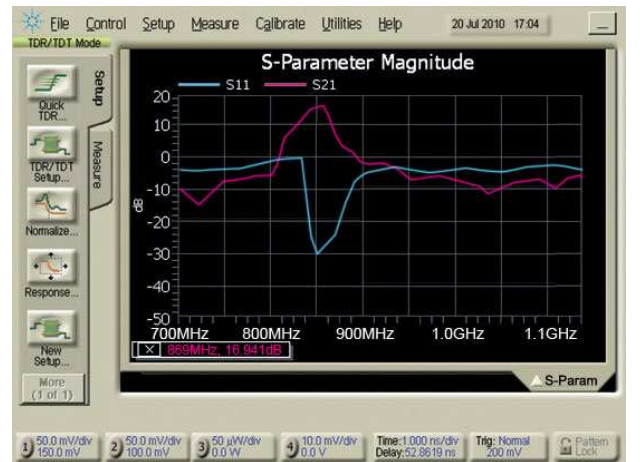


Fig. 8: S-Parameters for microstrip transmission line impedance matching network in Fig. 7

V. CONCLUSION

Input and output matching networks are designed for an UHF RFID LNA that operates on 866 MHz using the Smith Chart through MATLAB programming. In this design, the unilateral transducer gain reaches around 17 dB with -30 dB input reflection coefficient at the turning frequency. Presentation of the results show the Smith Chart impedance matching theory and practice implementation is correct and feasible for UHF LNA.

VI. SAMPLE CODE

```
global Z0;
set_Z0(50);
f=866e6;
%define desired input and output impedances
Zin=Zout=50;
%define source and load impedances
ZL=103+j*253
ZS=45.9-j*147
%create smith chart, SWR circle and G=1 circle
sm=smith_chart;
const_SWR_circle(ZL,'m--');
const_SWR_circle(ZS,'m--');
const_G_circle(real(1/Zout),'m--');
const_G_circle(real(1/Zin)
global rf_Network;
init_network;
%achieved length values for transmission line l_B & l_D
l_TB=0.2478;
l_TD=0.31718;
%achieved length value for open-circuit stub l_B & l_D
l_OB=0.1999;
l_OD=0.20690;
% locates desired input and output impedances values
Add_shunt_impedance(Zout);
Add_shunt_impedance(Zin);
%add open-circuit stub for input and output networks
Add_shunt_oc_stub(360*l_OD,f,Z0)
```

```
Add_shunt_oc_stub(360*l_OD,f,Z0);
%add transmission line stub for input and output net-
works
Add_trans_line(360*l_TB,f,Z0);
Add_trans_line(360*l_TD,f,Z0);
%input and output impedances value after matching
network
Zin_A=rf_imp_transform(f,sm)
Zout_A=rf_imp_transform(f,sm)
```

Calculated input and output impedances value after running above matching networks in MATLAB:

```
ZS = 4.5900E+001 -1.4700E+002I
ZIN_A = 4.6011E+001 +1.4725E+002I
ZL = 1.0300E+002 +2.5300E+002I
ZOUT_A = 1.0303E+002 -2.5303E+002I
```

VII. REFERENCES

- [1] L. F. Tiemeijer, *et al.*, "Low-Loss Patterned Ground Shield Interconnect Transmission Lines in Advanced IC Processes," *Microwave Theory and Techniques, IEEE Transactions on Microwave Theory and Techniques*, vol. 55, pp. 561-570, 2007.
- [2] M. Vai and S. Prasad, "Computer-aided microwave circuit analysis by a computerized numerical Smith chart," *Microwave and Guided Wave Letters, IEEE*, vol. 2, pp. 294-296, 1992.
- [3] H. J. Delgada and M. H. Thursby, "Derivation of the Smith chart equations for use with MathCAD," *Antennas and Propagation Magazine, IEEE*, vol. 40, pp. 99-101, 1998.
- [4] R. Ludwig and G. Bogdanov, "Transmission Line Analysis" in *RF Circuit Design: Theory and Applications*, 2 ed. Upper Saddle River, New Jersey: Pearson Education, Inc., 2009.

- [5] G. Gonzalez, "Matching Networks and Signal Flow Graphs" and "Microwave Transistor Amplifier Design" in *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1996.
- [6] T. G. Bryant and J. A. Weiss, "Parameters of Microstrip Transmission Lines and of Coupled Pairs of Microstrip Lines," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 16, pp. 1021-1027, 1968.
- [7] S. M. R. H. Jie Li, "A 0.7 V 850 μ W CMOS LNA for UHF RFID reader," *Microwave and Optical Technology Letters*, vol. 52, pp. 2780-2782, December 2010 2010.
- [8] J. Choma and W. Chen, *Feedback networks: theory and circuit applications*. Singapore: World Scientific, 2007.
- [9] D. H. Schrader, "Microwave Circuit Analysis Methods" in *Microstrip circuit analysis*. Upper Saddle River, New Jersey: Prentice Hall PTR, 1995.

Intelligent Wireless Greenhouse Management System

Vu Minh Quan, Gourab Sen Gupta, Subhas Mukhopadhyay

Massey University
School of Engineering and Advanced Technology,
Riddet Building, Turitea Campus, Palmerston North
QuanVu.nz@gmail.com
G.SenGupta@massey.ac.nz
S.C.Mukhopadhyay@massey.ac.nz

Abstract:

A greenhouse provides shelter and protects plants from harsh environment and external interferences. It allows plants to grow under an optimum condition which maximizes the growth potential of the plants. The existing systems only allow for the monitoring of climate variables such as temperature or humidity and often overlook many other important factors such as CO₂, light, soil moisture, soil temperature etc. Neglecting these climate factors leads to inaccurate observation of the overall greenhouse climate condition. Therefore to make up for this weakness, the prototype designed for this particular research will allow for a better monitoring of the climate condition in the greenhouse by integrating several sensor elements such as CO₂, Temperature, humidity, light, soil moisture and soil temperature into the system. Often when environmental parameters such as temperature or humidity become higher or lower than usual, some corrections are required. Fuzzy logic computing is integrated into the system to serve this particular purpose. In general this prototype provides a complete automated greenhouse management system that requires almost no human intervention. The purpose of this paper is to give an overview of the design and construction of the prototype and emphasises the integration aspects of the SHT75 humidity and temperature sensor.

Keywords:

Humidity sensor, temperature sensor, wireless sensor network, environmental monitoring.

1. INTRODUCTION

A greenhouse allows the growers to produce plants in places where the climate would otherwise be unfeasible for the growing of plants. The producing of plant is independent of the geographic location or the time of the year. It also provides shelter for the plants, protects them from harsh weather conditions, insects and diseases. It allows plants to grow under an optimum condition, which maximizes the growth potential of the plants.

Various environmental parameters are essential for the quality and productivity of plant growth. Continuous monitoring of these parameters gives valuable information to the grower for better understanding of how each factor affects the quality and the rate of plant growth and how to manage maximum crop yield.

Traditional wired monitoring system have been used in the past generation greenhouses. The system itself is usually simple and has limited control over local activities such as: heating, lighting, irrigation, ventilation, etc. A wired monitoring system transmits data by means of cables. With cables there are many potential problems such as; cable aging, data loss, cable barrier and loose terminals. This results in lower reliability and complex construction of the cable network. Moreover, measurement points are difficult to relocate once they are installed. It is a disadvantage to install this system in modern greenhouse. A modern greenhouse is typically big, resulting in a larger number of measurement points

spread across the greenhouse. A dramatic increase in installation and automation cost is almost certain. Therefore, a more convenient and cost effective system is required.

The wireless greenhouse management system reported in this paper was developed to meet this particular need. The wireless system does not need cables to run and has lower power consumption. It comprises of embedded wireless sensor nodes that can be used to collect real-time environmental data and it allows communication between the central control station and actuators that are located in different parts of the greenhouse. Moreover the system is easy to relocate once installed and maintenance is relatively cheap and easy. The only additional cost occurs when the batteries run out. But the lifetime of the battery can be extended if an efficient energy saving control algorithm is implemented.

In this work, our objective is to design and develop a wireless control network prototype for environmental management in a commercial greenhouse and test its feasibility and reliability with a simple experimental set up. The system designed in this work is able to monitor up to six different environmental parameters. Below are the environmental parameters that this project will be focused on:

- Temperature
- Humidity
- Carbon Dioxide (CO₂)

- Light Intensity
- Soil Temperature
- Soil Moisture Level

The paper is arranged as follows: section 2, presents some of the related research on the application of wireless sensor network in environmental monitoring. Section 3, gives an overview of the wireless control network prototype. Section 4 will be focused on the SHT75 temperature and humidity digital sensor used in this project. Section 5, presents the experimental setup and results and finally section 6 will present the conclusions and future challenges.

2. RELATED WORKS

Teemu Ahonen et al [1] developed a wireless sensor network for greenhouse monitoring by integrating a sensor platform provided by Sensinode Ltd with four different commercial sensors. The hardware design of the system consists of: Sensinode's Micro.2420 U100 operates as a basic measuring node and four commercial sensors (humidity, temperature, light and CO₂). The idea behind this development is to test the reliability and feasibility of a prototype wireless environment monitoring system in a commercial greenhouse.

A study on greenhouse auto control system based on wireless sensor network was carried out at Sunchon National University [2]. The system consists of sink nodes, database server computer and actuator nodes. The sink nodes are responsible for collecting environmental data and sending the collected data to a local database server. The actuator nodes act according to the collected environmental data. This research also focused on examining the correlation between variation of condition of leaf and the incidence of harmful insects [2].

Othman Sidek et al [3] presented a preliminary infrastructure development of a greenhouse wireless sensor network. This system allows for continuous data monitoring of carbon dioxide and oxygen level in environments at remote areas. The system also adopts Wireless Radio Frequency method which allows the user to receive report messages from the greenhouse via Global Sort Message Technology (GSM).

3. SYSTEM OVERVIEW

The proposed system consists of three parts: Greenhouse Station, Coordinator Station, and Central Control Station. The greenhouse Station is responsible for the collection of climate measurements in the greenhouse (Temperature, Humidity, CO₂ etc). Coordinator Station acts as a router that controls the flow of data and instructions between the Greenhouse Station and the Central Control Station in a pre-programmed manner. It also manages the local activities such as turning on/off: sprinkler, humidifier etc. Central Control Station is the main controller of the system. Its jobs are to -

- Issue instruction to lower computers (Coordinator Station, Greenhouse Station)

- Process incoming data and provide an easy and convenient way that allows the users to easily access the data
- Provide control commands to adjust the greenhouse climate condition in accordance with the requirements.

The overall system is shown in figure 1.

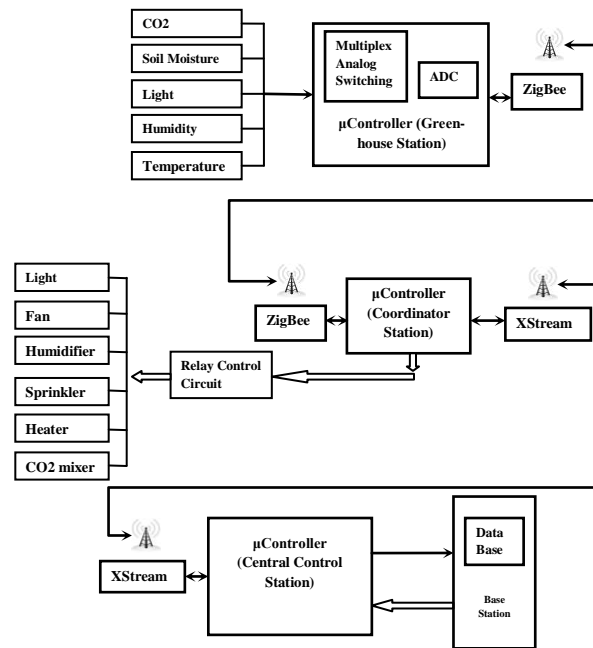


Figure 1: Block diagram of the system

3.1. Hardware Design of the System

The data acquisition unit consists of five commercial sensors:

- SHT75 humidity and temperature sensors are Sensirion's family of relative humidity and temperature sensors. These sensors integrate sensor elements plus signal-processing in a compact format and provide a fully calibrated digital output [5]. They have humidity operating range from 0 – 100% and temperature operating range from -40C – 125C. SHT75 sensors have low power consumption and fast response time.
- TGS4161 CO₂ sensors are solid electrolyte CO₂ sensors which exhibit a linear relationship between the changes in voltage output and the CO₂ gas concentration on a logarithmic scale. They have low power consumption, good long term stability and excellent durability against humidity, making them ideal for indoor air control applications such as this project.
- VG400 soil moisture sensors are low frequency soil moisture sensors with low power consumption. Their output voltage is proportional to the moisture content in the soil. The VG400 sensors are insensitive to salinity, small in size and are water proof, making them ideal for both indoor and outdoor applications.

- THERM200 soil temperature sensors are similar to VG400 in terms of low power consumption and robust to external interferences. These sensors have temperature span from -40°C to 85°C. The output voltage is linearly proportional to the changes in the temperature in the soil.
- NORP12 light dependent resistance sensors have spectral response similar to the human eye. Their internal resistance increases or decreases dependent on the level of light intensity impinging on the surface of the sensor.

The CPU used in this system is a C8051F020 single chip computer, which can process data at a high speed. C8051F020 is a 20 MIPS mixed-signal, 64K in-system programmable FLASH 8051 with 256 bytes RAM; 4K-bytes XRAM; SPI; SMBus/I2C; 2 UARTS; 12-bit 8-channel A/D; on-chip temperature sensor; 8-bit 8-channel A/D; 12-bit 2-channel D/A; 5 timers; PCA; 8 ports with 64 digital port I/Os [4].

The wireless transceiver modules used for the communication employed the XBee2.5 and XStream PKG 2.4GHz purchased from Digi Company. These transceivers operate within the ZigBee protocol and they require minimal power consumption.

3.2. Software Design of The Greenhouse Station

Figure 2 shows the flow diagram of the software design of the Greenhouse Station.

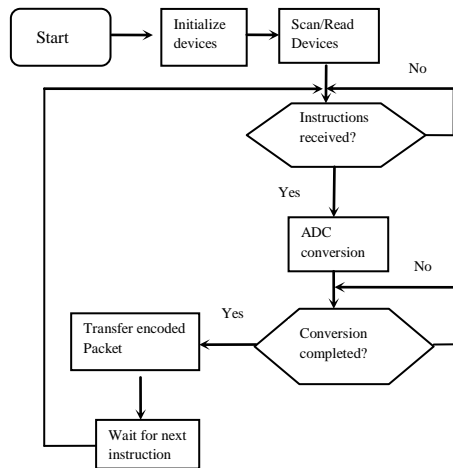


Figure 2: Software design of Greenhouse Station

The Greenhouse Station is used for collecting and processing raw data from the sensors and transform them into a more understandable format. Collected data is encoded into a data packet and is sent to the central control station via the coordinator.

3.3. Software Design of The Coordinator Station

Figure 3 shows the flow diagram of the software design of the coordinator station

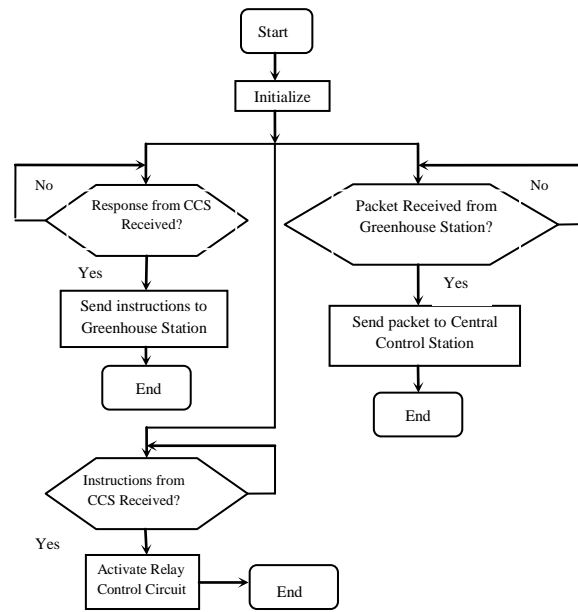


Figure 3: Software design of Coordinator Station

The Coordinator station controls the flow of incoming and outgoing data and instructions from the Greenhouse Stations and the Central Control Station.

3.4. Software Design of The Central Control Station

Figure 4 shows the flow diagram of the software design of the Central Control Station

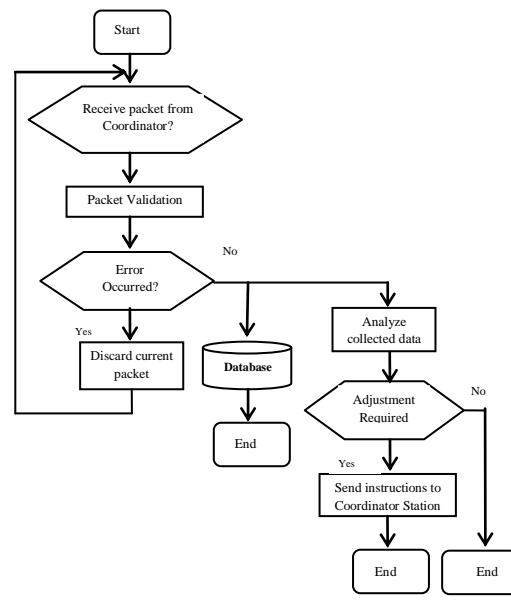


Figure 4: Software design of Central Control Station

The Central Control Station is the main controller of this system. It can be used for data control and storage as well as providing instructions to lower computers (Greenhouse Station, Coordinator Station). When environmental parameters such as temperature or humidity become higher or lower than usual, some artificial controls are needed. The controller station has been specifically designed to tackle this issue. It is able to automatically control the greenhouse climate by sending control

commands, in accordance with the requirements, through to the Coordinator Station. Upon receiving the commands, the Coordinator Station immediately controls the electric-relay switches in the greenhouse to activate local activities such as: heating, lighting, ventilation etc.

3.5. Communication Protocol

A communication protocol is necessary to enable the recipient of the data to determine where the packet starts and ends since there is no hardware synchronization to indicate the start/end of the data. A graphical representation of this protocol is depicted in Figure 5

Header bytes (x2)	Type ID (x1)	Battery Level (x2)	Data Bytes	Checksum bytes (x2)

Figure 5: Communication Packet

Two bytes are used to indicate the start of the packet. One byte is used to indicate the type of the packet (respond type or command type). Two bytes are used to represent the battery level of the sensor station. Four bytes are assigned for each sensor reading. The packet ends with two check sum bytes. These check sum bytes are for error checking. It allows the recipient to detect any errors that may have occurred during the transmission.

4. SHT75 HUMIDITY/TEMPERATURE SENSOR

SHT75 sensors are Sensirion's family of relative humidity and temperature sensor. The chip on these sensors integrates sensor elements plus digital processing in a compact format and provides a fully calibrated digital output [5]. A unique capacitive element is used for measuring relative humidity while temperature is measured by a band gap sensor.

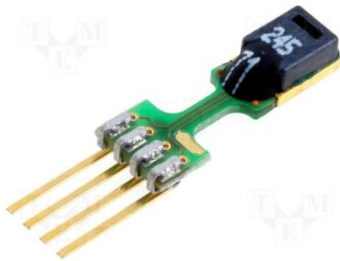


Figure 6: SHT75 Humidity and Temperature Sensor [5]

Besides having humidity and temperature sensors, the chip also contains an amplifier, a converter, One Time Programmable memory (OTP) and a digital interface. SHT75 sensors are pre-calibrated in a precision humidity chamber and the calibration coefficients obtained are directly programmed into an OTP memory on the chip [5]. These coefficients are used to internally calibrate the signals from the sensors. Figure 7 shows the sensor and its pin connections

Pin	Name	Comment
1	SCK	Serial Clock, input only
2	VDD	Source Voltage
3	GND	Ground
4	DATA	Serial Data, bidirectional

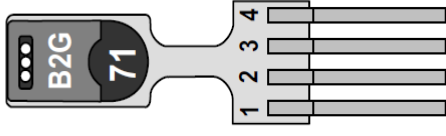


Figure 7: SHT75 pins assignment [5]

SHT75 comes with four pins: VCC, GROUND (GND), DATA, and SERIAL CLOCK (SCK). The serial clock pin on the sensor is used for input signal only. Serial data pin on the other hand is bidirectional. For sending command to the sensor, the DATA line needs to be valid on the rising edge of the SCK and must remain stable while the SCK line is high. For reading data from the sensor, DATA line needs to be valid after the SCK line has gone low and remains valid until the next falling edge of the SCK line [5].

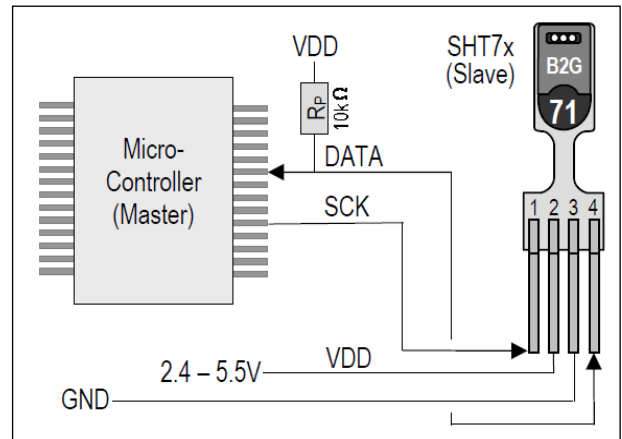


Figure 8: Application Circuit [5]

A voltage source between 3.3 - 5.5V is required to power the sensor. Both serial data and serial clock pins are connected to separate pins on the microcontroller. In order to avoid signal contention, the microcontroller must drive the DATA signal low. Therefore, a pull up resistor of 10k ohm is needed to be connected between the serial data line and the microcontroller to pull the signal high.

5. EXPERIMENTAL SETUP

An experiment was carried out to test the sensitivity and the accuracy of SHT75 sensor. A set of 1.5V batteries, SHT75 sensor, SPDT switch, 3.3v regulator and a C8051F020 microcontroller development board were used in this experiment. The SCK and DATA pins of SHT75 are connected to pin 0 and 1 of Port 1 on the development board. The development board is connected to a laptop via USB Debug Adaptor. Figure 9 shows the experimental setup.

Figure 9: SHT75 Experimental Setup

To initiate a transmission, a special transmission start sequence has to be issued to the sensor. This transmission start sequence wakes up the sensor and allows the sensor to be ready for any incoming instructions or commands.

Void Transmission_Start (void)

```
// _____
// DATA |_____|
//
// SCK: ____| |____| |____|
{
    DATA = 1;          SCK = 0;
    small_delay (50);   SCK = 1;
    small_delay (50);   DATA = 0;
    small_delay (50);   SCK = 0;
    small_delay (100)   SCK = 1;
    small_delay (50);   DATA = 1;
    small_delay (50);   SCK = 0;
    small_delay (50);
}
```

Figure 10: Start-up Transmission

The output of the start-up transmission is shown in figure 10. The program lowers the DATA line while the SCK line is high, followed by a low pulse on SCK line and raising DATA line again. The start-up transmission is followed by a series of command bits. The subsequent command consists of three address bits and five command bits ('00000101' for relative humidity, '00000011' for Temperature). Figure 11 and 12 show examples of sending command to the sensor.

Figure 11: Relative Humidity command (00000101)

Figure 12: Temperature command (00000011)

The experiment was carried out for one hour. The data sampling interval is 10 minutes and the results were compared with the results obtained from other temperature and humidity measurement devices such as SHT10 and Fluke 975 Air Meter. The test results are shown in Table 1.

Table 1: Experimental results

The test results show that the temperature obtained using SHT75 sensor is relatively similar (approximately $\pm 0.7^{\circ}\text{C}$ and $\pm 0.9\%\text{H}$ compared to SHT10 and $\pm 1.3^{\circ}\text{C}$ and $\pm 2.3\%\text{H}$ compared to Fluke975) to the results obtained from other sensor devices. The humidity measurements are slightly lower than humidity measurements obtained from Fluke 975 Air Meter. However these differences are very small. Therefore it is possible to make an assumption that the SHT75 is relatively accurate.

6. CONCLUSIONS AND FUTUREWORKS

This paper proposes a prototype WSN for greenhouse environment control. The prototype allows the growers to monitor various greenhouse environment parameters as well as controlling the local activities in the greenhouse. An experiment was carried out to determine the accuracy of the SHT75 humidity and temperature sensor. The SHT75's experimental results were compared to the results obtained from other humidity and temperature measurement instruments (SHT10, Fluke 975 Air Meter). The SHT75 sensor has proven to be relatively accurate compared to other measurement instruments. However, in our future work, we will conduct experiments using SHT75 sensors in various locations with different climate characteristics in order to achieve more detailed results.

7. REFERENCES

- [1] Ahonen, T., Virrankoski, R., & Elmusrati, M. (2008). *Greenhouse Monitoring with Wireless Sensor Network*. Proceeding of Mechtronic and Embedded Systems and Applications, 2008. MESA 2008. IEEE/ASME International Conference, pp. 403 – 408.
- [2] BeomJin, K., DaeHeon, P., KyungRyung, C., ChangSun, S., SungEon, C., & JangWoo, P. (2008). *A Study on the Greenhouse Auto Control System Based on Wireless Sensor Network*. Proceeding of Security Technology, 2008. SECTECH '08, pp. 41 – 44.
- [3] S.Othman. "Preliminary Infrastructure Development for Greenhouse Accounting of Malaysian Rainforest using Wireless Sensor Network". European Journal of Scientific Research, Vol.33 No.2, pp.249-260, 2009.
- [4] Silicon Laboratories Inc. (2010). C8051F020 Development Kit. [online]. Available: [http://www.silabs.com/Support Documents/TechnicalDocs/C8051F020_Short.pdf](http://www.silabs.com/Support%20Documents/TechnicalDocs/C8051F020_Short.pdf).
- [5] Sensirion Inc. (2010). SHT75 – Humidity & Temperature Sensor. [online]. Available: http://www.sensirion.com/en/pdf/product_information/Datasheet-humidity-sensor-SHT7x.pdf
- [6] K.Anuj et al., "Prototype Greenhouse Environment Monitoring System," Proceeding of the international MultiConference of Engineering and Computer Scientist, Vol 2, pp.17-19, March.2010.
- [7] Abdul Aziz, I., Ismail, M. J., Samiha Haron, N., & Mehat, M. (2008). *Remote monitoring using sensor in greenhouse agriculture*. Proceeding of Information Technology, 2008. ITSIm 2008. International Symposium.
- [8] Zhou, Y., Yang, X., Guo, X., Zhou, M., & Wang, L. (2007). *A Design of Greenhouse Monitoring & Control System Based on ZigBee Wireless Sensor Network*. Proceeding of Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007, pp. 2563 – 2567.
- [9] Yang, S., & Zhang, Y. (2010). *Wireless Measurement and Control System for Environmental Parameters in Greenhouse*. Paper presented at the Measuring Technology and Mechatronics Automation (ICMTMA), 2010, Vol 2, pp. 1099 – 1102.

Mobile Traceable Robot for Large Area Operation

M. B. Hitchings, G. Sen Gupta

Massey University, School of Engineering & Advanced Technology, Palmerston North, New Zealand.
Email: matthew.hitchings@gmail.com, g.sengupta@massey.ac.nz

Abstract: This paper details the design of a robot that can be manoeuvred through dangerous outdoor environments, wirelessly sending video and GPS coordinates. The robot system contains a 433 MHz RF transceiver, C8051F020 microcontroller, specifically designed circuitry, and a 2.4 GHz wireless camera unit. The base-station contains a PC, a matching RF transceiver, a 2.4 GHz wireless video receiver unit, a virtual serial port, and a video capture card. Components in the robot are run off a single 11.1 V battery, and base-station components are run off a D size battery and the computer's power through USB ports. The robot is intended to be used in land surveying operations.

Keywords: GPS, microcontroller, robot, RS232, surveying, wireless communication.

1 INTRODUCTION

A fully functional mobile robot has been designed and built. It allows the user to control the robot from a computer while being able to see a live video feed from the camera, as well as a live-updated Google Maps screen showing the robot's position. Since the aim of this project was to design and build a prototype with a limited budget, the performance specifications would improve if more expensive components were used. The completed system is shown in Figure 1.

The major advantage of this system is the ability to identify hazards in a dangerous environment that is not

sheltered from the sky. It allows the gathering of images without the risk of harm for humans that would otherwise have to perform the inspection.

The applications of this robot range from the surveillance of damaged buildings to bomb sites, to live volcanoes. This robot is also useful in environments that are not exposed to the sky for gathering images, however the Google Maps screen will not be functional.

Because the position and images from the camera are frequently sent to the computer from the robot, in the future, software can be developed that uses this information to automatically guide the robot to a location on the map decided by the user.



Figure 1 – Two views of the fabricated robot.

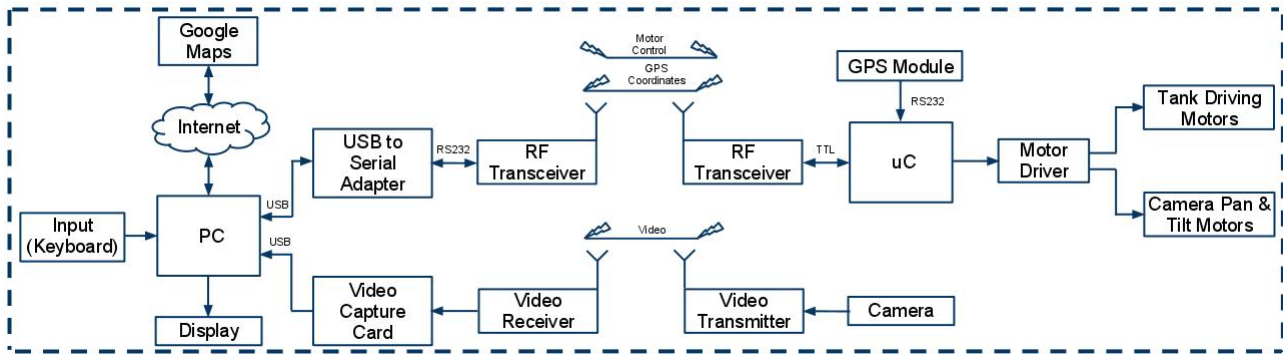


Figure 2 – Functional Block Diagram of the hardware in the system.

2 SYSTEM OVERVIEW

As illustrated in Figure 2, each of the system's inputs coming from the user, the camera and the GPS module pass through a number of function blocks, being of different protocols depending on where in the process they are. For example the signal from the keyboard that controls the tank's motors passes through six function blocks before it reaches its destination, and is transmitted as USB, RS232 and TTL along the way.

2.1 Microcontroller and Development Board

The microcontroller used in the system is a Silicon Laboratories C8051F020[1], which is clocked using a 22 MHz external oscillator. This microcontroller development board was used because of its RS232 port, its Pulse Width Modulation generation capabilities, and familiarity working with it.

In the earlier stages of this project a large development board that contained the microcontroller was used. This board contained an LCD screen, RS232 port, eight 8-bit digital I/O ports and more. The LCD screen on this larger board proved useful for testing and debugging of the wireless communication to the robot. As the project went on the larger development board was replaced with a smaller one that did not contain the LCD screen, but still had the RS232 port and the eight ports, shown in Figure 3. Using a smaller development board meant it could fit in the tank body without needing to create more room.

Two UART lines on the development board were used for communication, one through the RS232 port to the GPS module, and the other through one of the ports to the RF module. Four of the internal PWM generators the 8051 microcontroller can provide were used to control the speed of each motor. Through another port on the development board eight enable lines were sent to the motor driver board which controlled which motors were enabled or disabled.

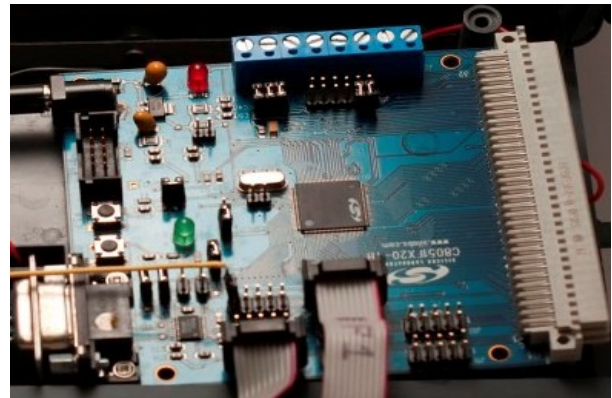


Figure 3 – C8051F020 Microcontroller on smaller Development Board.

2.2 Wireless Camera System and Video Capture

The wireless camera and receiver unit used in this project was sold for security purposes, but functioned adequately for this project. It included a small monochrome camera with a 2.4 GHz RF transmitter enclosed in its case, and a separate 2.4 GHz receiver unit with RCA outputs. It was chosen because of its low price, and 100 m rated line of sight transmission distance. The RCA outputs from the receiver unit connect to the inputs on a USB video capture unit, which plugs into the computer.

2.3 GPS Module

The QuadroUFO GPS100 module used in this project has -165 dBm tracking and an internal antenna. It was chosen because on its breakout board there are pins for RS232, TTL and USB communication. This gave two options for microcontroller communication (RS232 or TTL).

The module is interfaced with the microcontroller through the RS232 slot, and at a configurable frequency sends a variety of information acquired from the satellite to the computer in five separate messages. The microcontroller reads this information but only sends part of one of the messages across to the computer so the wireless communication throughput is maximised.

2.4 Motor Driver and Voltage Supply

The robot is powered by a 3-cell 11.1 V Lithium Polymer battery which is connected to the motor driver board designed specifically for the system. On the robot there is an RF module as well as logic gates which required 5 V, the wireless camera which required 9 V, and the microcontroller board and motors that were running on 11.1 V. This required the circuitry to include a 5 V regulator and a 9 V regulator.

Four motor drivers were required to drive the motors on the robot, and interfacing with the microcontroller to receive the control commands and pulse width modulation signals. The board also needed to interface the output signals to the motors. The schematic for the board is shown in Figure 4.

The L298 Dual Full-Bridge Driver chips used in the circuit were chosen because they had an attached heat-sink and could provide up to 2 A for each motor in DC operation. The maximum expected current for each of the two motors driving the tank in normal operating conditions was 1 A, and no more than 100 mA for the other two motors used for the turret. This gave a safety factor of 2 for one motor driver, and a safety factor of 20 for the other.

The motor driver board needed to be large enough to house all of the components required to deliver all of its functions, but small enough to be able to fit inside the tank body. The first board design to be printed performed all functions correctly, but was too large. The revised board delivers on all requirements including size, shown in Figure 5.

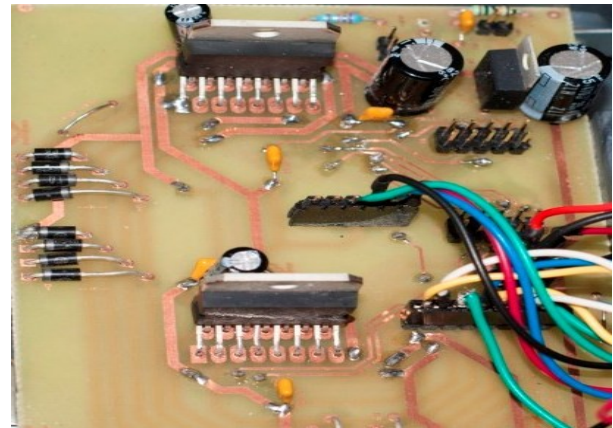


Figure 5 – Motor Driver Printed Circuit Board

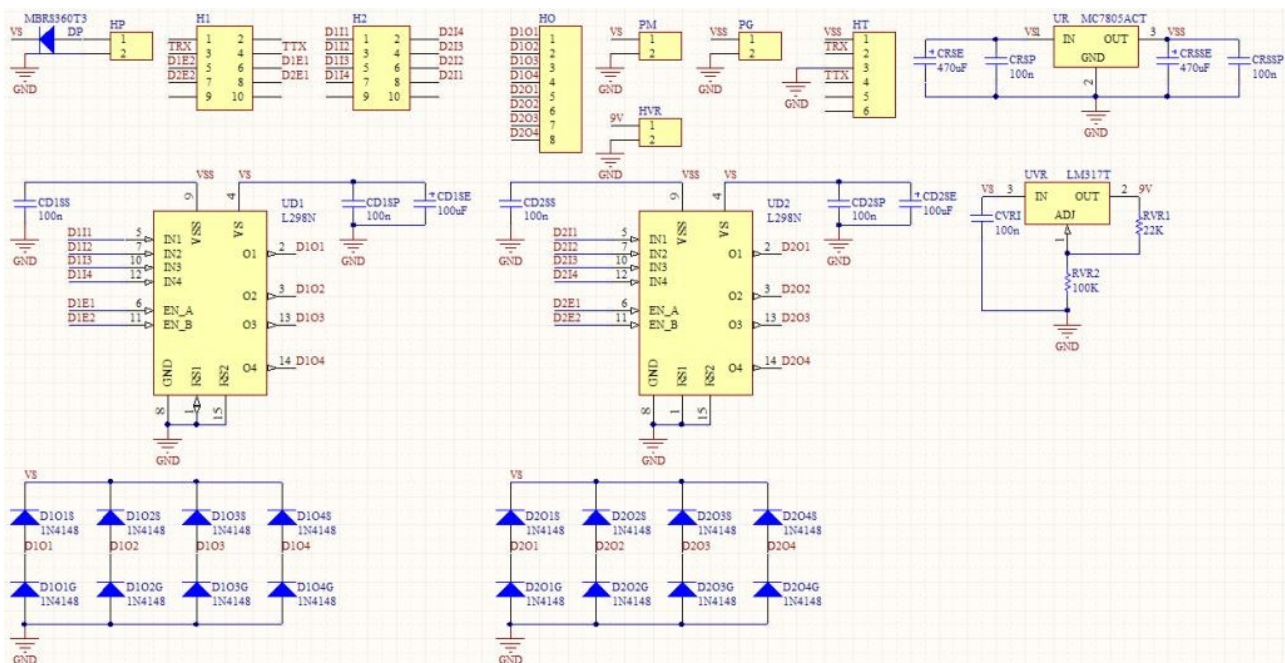


Figure 4 – Motor Driver Schematic.

2.5 RF Transceiver Modules

To send across control commands to the robot and to send GPS information back to the computer, RF transceiver modules were used that use the 433 MHz band. It was important that the carrier frequency of the control signals and GPS information is different to that

of the wireless video signal so that interference is minimised[2].

The RF module connected to the robot's microcontroller uses TTL signals to communicate and is powered off the motor driver board. The module connected to the computer uses RS232 signals to communicate, and since

RS232 serial ports are becoming less common on modern computers a USB to serial adapter is used. To power the base-station RF module a second USB plug is used.

Because the transceiver modules both transmit and receive information simultaneously, a protocol had to be implemented so that the data received by a transceiver would not be corrupted by data that it sent, and vice versa. The block diagram illustrating this protocol is shown in Figure 6. It can be seen that the control information in the system takes priority over the GPS information. This is because it is not so important if the map location takes a little longer to update, but if there is a long delay on the control then the response to the user will be undesirable.

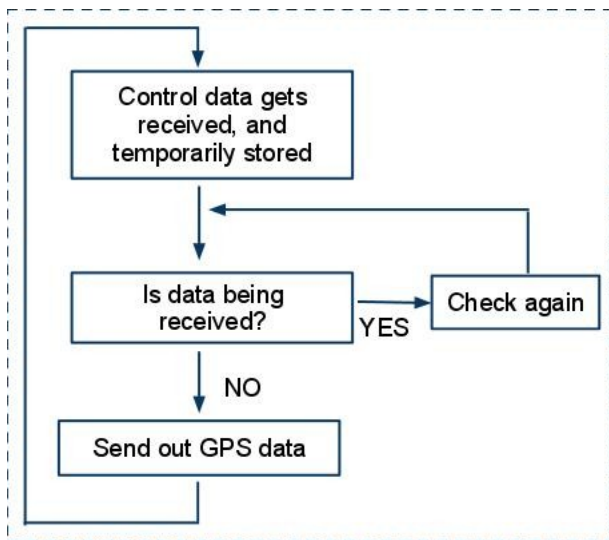


Figure 6 – Block diagram of the protocol used by the transceiver on the base-station.

2.6 Motors and Gearing

The robot required the tank to have full-directional movement, as well as the turret-mounted camera to be able to pan and tilt. This meant that a total of four motors were required.

The chassis, motors and gearing for the tank were used from a remote controlled tank unit, so no mechanical design was required to drive the motor or camera. However, it was important to find an appropriate duty ratio for the pulse width modulation fed to the motors. A balance needed to be found so the motors would be fed enough power to be able to drive the tracks, but not too much so that the current drawn would be too high, draining the battery and possibly damaging components[4]. The two motors and their gearing which are used for driving the tank are shown in Figure 7.

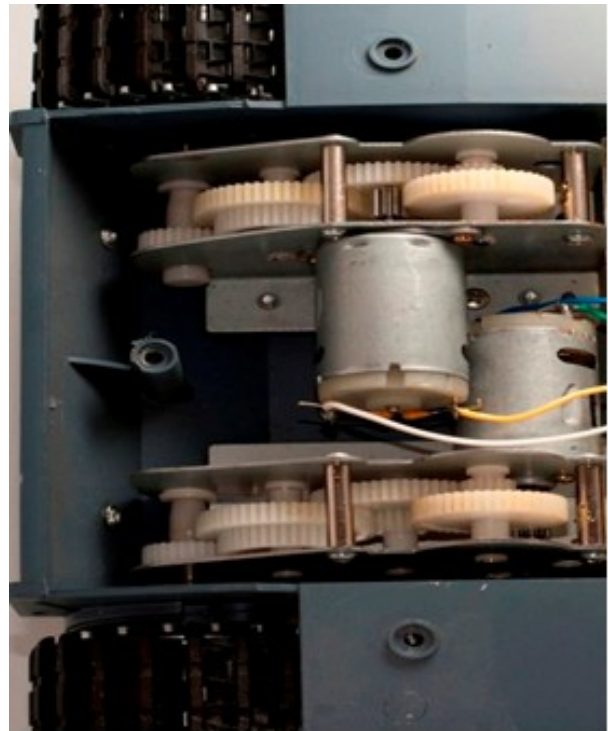


Figure 7 – DC Brushed Motors and gearing that drives the tank tracks.

3 HUMAN INTERFACE

The system had to have a tidy and attractive interface that is simple to use, as it is an interactive system that relies on the user for control of the motors. To achieve this focus had to be put into the functionality of the features of the program, the layout and how the user would control the motors of the tank.

3.1 Functionality

The computer program which the user interacts with was written in Microsoft Visual C#. All of the code was written especially for this project except the displaying and recording of the video from the video card. This code was downloaded from an open source website[3].

The communication through the virtual serial port could be performed using an inbuilt library with C#[5]. It allowed the sending of commands and the receiving of GPS coordinates required from this project.

Using the GPS coordinates to display the current position on the Google Maps screen was performed by manipulating the URL. The GPS coordinates of the tank are only sent across the wireless communication every 5 seconds to maximise the throughput required.

3.2 Layout

The Graphical User Interface for the program was designed so that the position and live video feed were clear to the user.

The layout also had to have more advanced features available to the user such as video recording, without

seeming cluttered. This was achieved by having a button which toggled the visibility of the more advanced features that are not always required, shown in Figure 8.

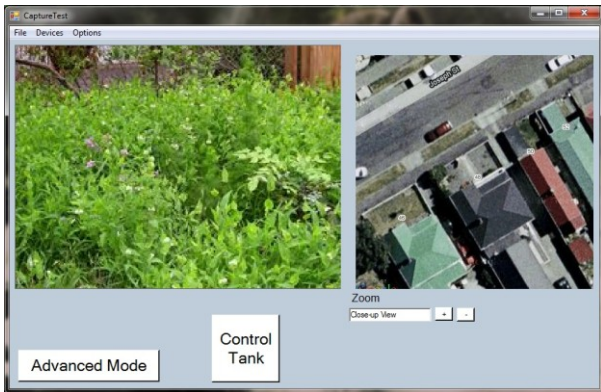


Figure 8 – Graphical User Interface of the base-station controller.

3.3 User Commands

The commands to the robot are sent through the keyboard, and require the left hand to send the tank-driving commands and the right hand to send camera pan and tilt commands.

4 SYSTEMS INTEGRATION

Because this project combines the functions of many different components, a major focus of this project was to successfully integrate them all together to get the system functional. This integration often caused problems, so a systematic process of testing had to be implemented. As described in Section 2, the signal sent from the keyboard has to go through six functional blocks with varying complexity before it arrives at the motors. This systematic process was practiced between the functional blocks, and within some of the blocks to get the system functioning.

Another important aim of this project was to have the final product built in a way that is up to industry standards. This meant that all components had to be safely secured, and wiring done in a neat and aesthetically pleasing way[6], shown in Figure 9. It was also important for the components on the outside to be mounted in an aesthetic manner, as shown in Figure 10.



Figure 9 – Internal view of the tank showing inter-connections of the microcontroller board, motor driver and motors.

5 FUTURE DEVELOPMENTS

The scope of this project was to develop and build a robotic vehicle which can be remotely controlled by a user, and shows its video and position on a screen. If this project were to be further developed, there are a number of directions to take.

In the system's current state, the video information that is received is not manipulated except for the purposes of displaying and recording. The images sent through could have image processing applied to them which could perform a variety of useful functions such as obstacle avoidance and target following.

The Google Maps screen in the GUI of this project only displays the position of the tank from the last time it received the GPS coordinates. If a point on the map were to be clicked, the robot could automatically guide itself to that location within a certain range, using image processing on the map to find the best route and processing on the video feed's images to avoid obstacles.



Figure 10 – Components secured on top of the tank including wireless camera, RF transceiver and GPS module.

6 RESULTS AND CONCLUSIONS

The outcomes of this project were a remotely controlled tank which sent back video and GPS information to the computer, complete with a computer program which displayed all relevant information in a clear and simple manner.

The outputs of this project satisfied all of the aims set out at the beginning. A systems-engineering approach was taken to ensure all components of the system came together to function correctly.

To satisfy all of the aims of the project, the tank uses RF transmission, GPS reception and wireless video transmission.

The system was tested in a spacious outdoor environment to see what range the system had. With a clear line of sight from the tank to the base-station of the system the tank-control, video and electronic map functioned correctly up to 30 m. When there was an obstacle between the robot and base-station of the system the range was reduced, depending on the obstacle's size and density.

Possible ways to further develop this project include adding image processing capabilities to perform functions such as map-manoeuvring and obstacle-avoidance.

7 REFERENCES

- [1] M.T. Chew and G. Sen Gupta, *Embedded Programming with Field-Programmable Mixed-Signal μ Controllers*. New Zealand: Massey University, 2005.
- [2] D. Christiansen, *Electronic Engineers' Handbook*. United States: McGraw-Hill, 1996.
- [3] B. Low. (2004). *DirectX video stream and frame capture* [Online]. Available: http://www.codeproject.com/KB/directx/DirXVidStrm/DirectXCapture_demo_Enhanced.zip
- [4] Y. Mei and Y. Lu and Y.C. Hu and C.S.G. Lee, (2005). A case study of mobile robot's energy consumption and conservation techniques. *International Conference on Advanced robotics, 2005. ICAR '05. Proceedings.*, 12(1), 492-497. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1507454/>
- [5] Microsoft MSDN. (2008). *System.IO.Ports namespace*. Available: <http://msdn.microsoft.com/en-us/library/system.io.ports.aspx>
- [6] N. Storey. *Electronics: A Systems Approach*. 2nd Edition, Coventy: Prentice Hall, 2000.

Radio Astronomy and eVLBI using KAREN

S.D.Weston*, T. Natusch and S. Gulyaev

Institute for Radio Astronomy and Space Research

AUT Univeristy

Private Bag 92006

Auckland 1142

New Zealand

*stuart.weston@aut.ac.nz

Abstract:

Kiwi Advanced Research and Education Network (KAREN) has been used to transfer large volumes of radio astronomical data between the AUT Radio Astronomical Observatory at Warkworth, New Zealand and the international organisations with which we are collaborating and conducting observations. Here we report on the current status of connectivity and on the results of testing different data transfer protocols. We investigate new UDP protocols such as "tsunami" and UDT and demonstrate that the UDT protocol is more efficient than "tsunami" and ftp. We report on our initial steps towards real-time eVLBI and the attempt to directly stream data from the radio telescope receiving system to the correlation centre without intermediate buffering/recording.

Keywords:

Radio astronomy, VLBI, eVLBI, data transfer protocol, network, UDP, UDT, tsunami, KAREN

1 INTRODUCTION

The New Zealand 12-m radio telescope is located some 60 km north of the city of Auckland, near the township of Warkworth as shown in Figure 1. Currently it operates in three frequency bands centred around 1.4, 2.4 and 8.4 GHz. This fast slewing antenna (5 degrees per second in Azimuth) is well suited to the purposes of geodetic radio astronomy and spacecraft navigation/tracking [1]. It also can be effectively used for astrophysical VLBI (Very Long Baseline Interferometry) research in conjunction with large radio telescopes [2].

VLBI is a technique that uses several radio telescopes separated by continental distances, synchronised and working together, simultaneously tracking the same celestial object, therefore creating a virtual radio telescope of a continental/trans-continental scale. This method was successfully used in April-May 2010 when a combination of six radio telescopes in Australia and New Zealand (Warkworth) observed the core of the radio galaxy Centaurus A [3]. Following the installation of the KAREN (Kiwi Advanced Research and Education Network) [4] GigaPoP at the AUT radio telescope, the New Zealand Radio Telescope data was transferred to Western Australia, where it was correlated and calibrated, and images of a sub-parsec resolution were created.

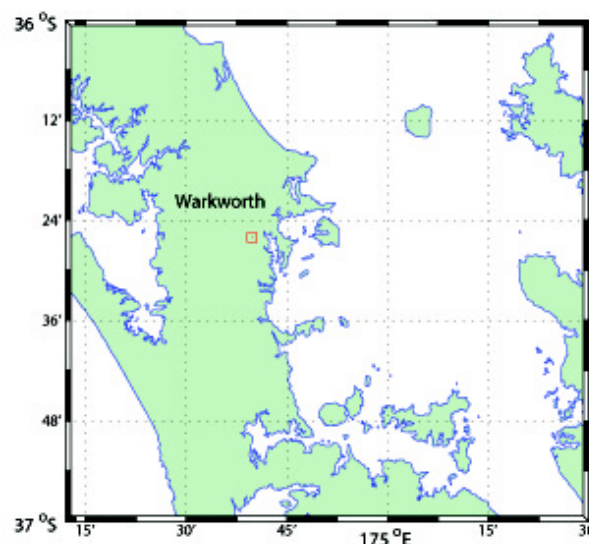


Figure 1. \square Shows geographic location of the AUT Radio Telescope south of Warkworth.

Figure 2 shows a screenshot of KAREN Weathermap for Centaurus A data transfer between Warkworth and Sydney on 9-13 May 2010. Data transfer is an issue of great importance for VLBI. Previously data was recorded to magnetic tapes, these were then sent via traditional means such as post/courier to the correlation centre. More recently with the reduced cost of hard disk

storage, the data has been recorded to removable disk arrays which can then also be sent physically to the correlation centre. Recently with the advent of eVLBI the data is sent via high speed networks to the correlation centres.

With the connection of KAREN to the AUT Radio Telescope we wish to optimize the use of KAREN for transferring large volumes of observational data to our partners in Australia, the United States and Europe and for conducting eVLBI. Here we discuss the use of FTP over TCP/IP for sending these files to Australia, as well as new protocols which are being used by our partners such as tsunami [5] and UDT [6] (UDP-based Data Transfer) via the network protocol UDP.

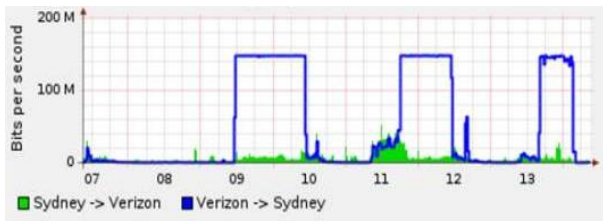


Figure 2. KAREN Weathermap data transfer between Warkworth and Sydney on 9-13 May 2010 (VLBI observations of Centaurus A radio galaxy) using ftp.

2 BACKGROUND

Point to point with no hops FTP is efficient, but as the number of hops in the route increases and the incidence of lost packets and collisions increases the TCP congestion avoidance algorithm becomes a severe limitation to the throughput that can be achieved.

2.1 Overview

The Institute for Radio Astronomy and Space Research collaborates with a number of international partners. This research collaboration can be conditionally broken into three major research topics/groups:

1. Astrophysical VLBI observations with Australian Long Baseline Array (LBA), data from the AUT Radio Telescope has to be sent to CSIRO [7] and then to Curtin University, Perth for correlation and imaging. The main goal of this collaboration is the study of the physics and origin of both extragalactic and galactic radio sources (active galactic nuclei, radio galaxies, supernova remnants, star formation regions and others).
2. Observation and navigation of inter-planetary missions/spacecraft, as well as ground tracking services for a variety of space missions. VLBI observations of Japan Aerospace Exploration Agency (JAXA) [8] IKAROS and Akatsuki spacecraft were conducted in July 2010. In September 2010 a signal from European

Space Agency (ESA) [9] Mars Express was successfully detected. Using a PCEVN recorder and a modified receiver backend a 32 MHz band of data was collected for analysis. All data was transferred by network directly to Metsähovi, Finland [10] and to the Joint Institute for VLBI in Europe (JIVE) [11] using the international connections of KAREN.

3. The forthcoming regular IVS (International VLBI Service for Geodesy and Astrometry) [12] observations of a range of quasars, which uniquely support the International Celestial Reference Frame. We will be required to send observational data to one of two correlation centres located in the United States Naval Observatory and the Max-Planck Institute for Radio Astronomy in Germany.

Figure 3 gives a simple visual overview of destinations we are trying to establish connectivity with via KAREN for current and future data transfers from the Warkworth Radio Astronomical Observatory site. The double ended arrow lines are solid to indicate that connectivity has been established between the Observatory and the remote centres. The dashed line shows a connection needed but not yet established.

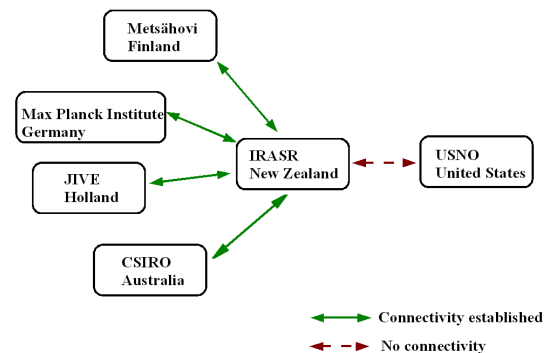


Figure 3: Established and required connectivity from IRASR to partners via KAREN

2.2 Network Protocols

The first UDP file transfer protocol we investigated was tsunami developed by Jan Wagner of the Metshovi Radio Observatory in 2007 [5]. This will be the protocol of choice for sending files to Bonn for the coming IVS observations, as stipulated by the Max-Planck Institute.

Another UDP protocol called UDT was developed in the University of Illinois in 2005 [13]. UDT was investigated by our Australian partners (private communication Chris Phillips, CSIRO) in 2008-09. This now appears to have matured and is of further interest and warrants further investigation.

2.3 Connectivity Status

Table 1 gives an overview of destinations that connectivity has been achieved with and the protocols that have been verified for data transfer and command line access to remote servers for initiating data transfers.

Destination	Protocol		Command	Date
CSIRO	ftp		ssh	01/04/10
Bonn	UDT	tsunami	ssh	01/06/10
Metsähovi	UDT	tsunami	ssh	21/07/10
JIVE	UDP		iperf	27/07/10
USNO	-	-	-	-

Table 1: Connection details between AUT and partners via KAREN

3 DATA TRANSFER TESTS and RESULTS

3.1 Max-Planck Institute, Bonn, Germany

The results in Table 2 were obtained in June 2010 transferring an actual VLBI file (16 bit) produced by an observation using the AUT Radio Telescope. The data was transferred from the AUT's IBM Blade server via the KAREN network to the Max-Planck Institute, Bonn, Germany. The second column shows the protocol used, third column gives the amount of data sent in bytes, the forth column provides the time it took to transfer the data and the fifth column is the result of the division of data volume by time, therefore providing an average throughput rate over the period.

Route	Protocol	Bytes	Time (s)	Throughput (Mbps)
AUT – Bonn	Ftp	65G	8016	65
AUT – Bonn	Tsunami	65G	3466	151
AUT – Bonn	UDT	65G	1920	273

Table 2: Data Transfer results AUT to Bonn

This exercise was conducted repeatedly over several days and at different times giving slightly different average rates.

The following is the traceroute from the AUT Blade to the IP address given for Bonn. This provides a snapshot in time for a path to the destination which can change.

```
[evlbi@ska-blade1 vt14b]$ traceroute viola2

traceroute to viola2 (194.94.199.165), 30 hops max, 40 byte packets
 1 156.62.231.65 (156.62.231.65) 0.478 ms 0.531 ms 0.609 ms
 2 210.7.32.1 (210.7.32.1) 0.368 ms 0.372 ms 0.438 ms
 3 210.7.36.67 (210.7.36.67) 0.663 ms 0.773 ms 0.823 ms
 4 210.7.47.22 (210.7.47.22) 143.304 ms 143.417 ms 143.468 ms
 5 abilene-1-lo-jmb-706.sttlwa.pacificwave.net (207.231.240.8)
 143.210 ms 143.216 ms 143.239 ms
 6 xe-2-0-0.0.rtr.salt.net.internet2.edu (64.57.28.104) 159.758
 ms 159.704 ms 159.694 ms
 7 xe-1-1-0.0.rtr.kans.net.internet2.edu (64.57.28.25) 186.049
 ms 184.231 ms 190.001 ms
 8 ge-1-2-0.0.rtr.chic.net.internet2.edu (64.57.28.37) 200.868
 ms 200.927 ms 200.921 ms
 9 xe-3-0-0.0.rtr.wash.net.internet2.edu (64.57.28.101) 217.103
 ms 217.101 ms 217.143 ms
10 abilene-wash.rtl.fra.de.geant2.net (62.40.125.17) 310.233 ms
310.269 ms 310.307 ms
11 dfn-gw.rtl.fra.de.geant2.net (62.40.124.34) 310.920 ms
310.932 ms 310.927 ms
12 xr-bir1-tel-1.x-win.dfn.de (188.1.145.45) 313.058 ms 313.165
ms 313.403 ms
13 xr-bon1-te2-1.x-win.dfn.de (188.1.144.10) 307.749 ms 308.166
ms 307.499 ms
14 viola2 (194.94.199.165) 308.847 ms 308.850 ms 308.826 ms
```

It has been found that issuing this command repeatedly over several months the route appears static and has not changed. The total number of hops is 14 on the route from AUT to Bonn via KAREN, demonstrating the complex path to be negotiated and why data transfers via protocols such as FTP will not be efficient.

3.2 Metsähovi Observatory, Finland

The results in Table 3 were obtained in July 2010 transferring an actual VLBI file (16 bit) produced by the AUT Radio Telescope during our VLBI observations. The data was transferred from the AUT's IBM Blade server via the KAREN network to the Metsähovi Observatory, Finland. The layout and description of Table 3 is the same as for Table 2 for easy comparison.

Route	Protocol	Bytes	Time (s)	Throughput (Mbps)
AUT – Metsähovi	Ftp	3.1G	7458	61
AUT – Metsähovi	Tsunami	65G	4979	105
AUT – Metsähovi	UDT	65G	1157	453

Table 3: Data Transfer results AUT to Metsähovi

At the end of August 2010 observations were conducted of the ESA's Mars Express spacecraft orbiting Mars. This data was transmitted to Metsähovi via KAREN. The data of 24 August which totaled 86 GB was sent using the UDP protocol tsunami immediately after the observational session (Figure 4). The next set of observational data (26 August) totaling 187 Gbytes was sent via UDT (Figure 5). Average rates sustained were 250 and 300 Mbps respectively.

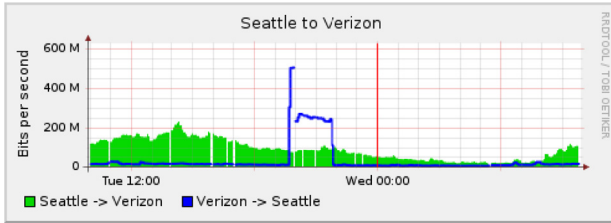


Figure 4: Snapshot of KAREN weathermap 24/8/2010 for period showing file transfers to Metsähovi using tsunami

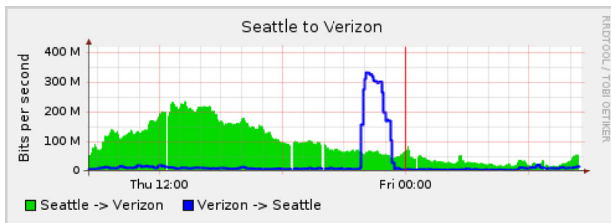


Figure 5: Snapshot of KAREN weathermap 26/8/2010 for period showing file transfers to Metsähovi using UDT

It was found that UDT performed better than tsunami by transmitting more data in a shorter period and obtaining a higher bandwidth utilization during that period.

3.3 Comparison of results for file transfer

The main purpose of our tests was to compare the performance of different data transfer protocols. Figure 6 shows the result of this comparison for two different end points, namely Bonn and Metsähovi. The histogram clearly demonstrates the advantage of the UDT protocol. It is up to 4 times faster than tsunami protocol and almost one order of magnitude faster than the standard ftp. The difference in data transfer rate between end points can be connected with the varying number of hops and performance of sections along the routes.

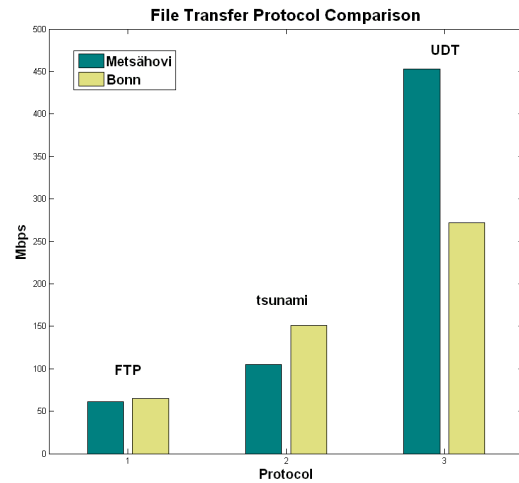


Figure 6: Comparison of the different file transfer protocols to different network end points over KAREN

3.4 Data streaming results

Another set of experiments was conducted in September 2010 aiming to test the tsunami protocol for streaming VLBI data directly from the radio telescope receiving system via KAREN to Metsähovi. This test is an important step towards real-time eVLBI.

Initial tests from the Observatory at Warkworth to the IBM Blade server at the AUT City Campus showed that the tsunami rates of 485 Mbps with no lost packets was sustainable. However, when streaming data from Warkworth to Metsähovi many thousands of lost packets occurred and a sustainable rate of 350 Mbps was achieved. This is significantly lower than the rate of 512Mbps required for the real-time eVLBI streaming of 8bit data to Metsähovi.

The next step we are planning to undertake is to test the more efficient UDT protocol for direct data streaming. It requires the VLBI software to be modified to make use of UDT instead of tsunami for streaming. This will require some programming effort as the network code has been merged within the application code rather than being maintained in a separate library.

4 CONCLUSIONS

It was clearly demonstrated that the use of the UDT protocol for radio astronomical data transfer has a number of advantages compared to the protocols currently used for VLBI and eVLBI. In particular, UDT has some advantages over tsunami:

- UDT is a better citizen on the network leaving bandwidth for TCP and other UDP protocols, which is very important on a shared network such as KAREN.
- UDT has an application programming interface (API) allowing easy integration with existing or future applications.

We have also investigated the ability to stream collected data via UDT and have modified the Curtin 16 bit to 2 bit conversion program for data streaming to a remote server in a fast and efficient manner ready for the correlator. This has been successfully demonstrated between Warkworth and the IBM Blade Server at AUT.

We have found KAREN to be a very useful tool for transmitting data to our international partners, and our Institute will be extending its use over the coming months. The next challenge will be to establish the IVS regular observational sessions and transmit the even larger volumes of data to the correlation centres, such as Bonn and USNO. Of future benefit to our work to stream data real-time to the correlation centres would be the ability to reserve bandwidth as a logical pipe within the KAREN bandwidth for the duration of an experiment.

ACKNOWLEDGEMENTS

This research was supported by the Ministry for Economic Development's SKA Grant. Warkworth connectivity via KAREN was made possible by the Remote Site Connectivity Fund provided by the Ministry of Research Science and Technology (MoRST) on behalf of the New Zealand Government. We are grateful to IBM for providing the IRASR with the IBM Blade Centre in the framework of the Shared University Research (SUR) Grant.

We thank the reviewers for many useful comments that led to an improvement of the paper. In addition we wish to thank Donald Clark, Bill Choquette, John Raine and Chris Litter from REANNZ/KAREN, and Ben Yeldon from AUT's ICT Department for their support and assistance. We thank Guifre Molera of the Metsähovi Radio Observatory, Simone Bernhart of the Max Planck Institute for Radio Astronomy at Bonn and Paul Boven from JIVE, who all assisted us in establishing connectivity to their sites. We are grateful to Hiroshi Takeuchi of JAXA, Japan and all our VLBI colleagues in Australia, in particular Steven Tingay, the International Centre for Radio Astronomy Research, Curtin University of Technology and Tasso Tzioumis, CSIRO.

REFERENCES

- [1] S. Gulyaev, T. Natusch and D. Wilson "Characterisation and Calibration of the 12-m Antenna in Warkworth, New Zealand," in *The 5th IVS General Meeting Proc.*, Hobart, Tasmania, February 2010.
- [2] A. Tzioumis, et al., "Evolution of the pc-scale structure of PKS 1934-638 revisited: first science with the ASKAP and New Zealand telescopes," *Astronomical Journal*, 2010, 140, 1506.
- [3] S. Tingay et al. (2010, May) *Trans-Tasman Antennas Successfully Linked* [Online]. Available: <http://www.ska.gov.au/news/Pages/100526-Trans-TasmanAntennasSuccessfullyLinked.aspx>
- [4] *Kiwi Advanced Research and Education Network*. <http://www.karen.net.nz/home/> (2010, Sep. 23)
- [5] J. Wagner *Tsunami UDP Protocol* [Online]. <http://tsunami-udp.sourceforge.net/> (2010, Sep. 23)
- [6] *UDT UDP-based Data Transfer* [Online]. Available: <http://udt.sourceforge.net/index.html> (2010, Sep. 23)
- [7] *The Australia Telescope National Facility* [Online]. Available: <http://www.atnf.csiro.au/> (2010, Sep23)
- [8] Japan Aerospace Exploration Agency [Online]. Available: http://www.jaxa.jp/index_e.html (2010, Sep 23)
- [9] European Space Agency [Online]. Available: <http://www.esa.int/esaCP/index.html> (2010, Sep 23)
- [10] *The Metsähovi Radio Observatory* [Online]. Available: <http://www.metsahovi.fi/en/> (2010, Sep 23)
- [11] *Jive - Joint Institute for VLBI in Europe* [Online]. Available: <http://www.jive.nl/> (2010, Sep 23)
- [12] "International VLBI Service for Geodesy & Astrometry". <http://ivscc.gsfc.nasa.gov/> (21st July 2010)
- [13] Yunhong Gu (2005). UDT A High Performance Data Transport Protocol (Unpublished doctoral dissertation). University of Illinois, Chicago, USA.

Keyword Index

- 3D, 45
- Android, 13
- antenna, 69
- automation, 85
- battery charging, 57
- Binding ssDNA, 63
- Biosensor, 63
- Biped, 97
- Biquad, 103
- Blu-ray, 75
- Bus Stop, 13
- Butterworth, 103
- Capacitance sensor, 63
- Chebyshev, 103
- CMOS, 63
- computer vision, 85
- crack detection, 79
- Cytometer, 75
- data transfer protocol, 141
- Decimation, 103
- delay, 51
- Diffuse optical tomography, 115
- Digital Filter, 103
- Display, 45
- DNA detection, 63
- Eclipse, 13
- egg shell, 79
- Embedded Database, 13
- Embedded Linux, 27
- environmental monitoring, 129
- ESEO, 9
- eVLBI, 141
- Flying Robotics, 21
- FPGA, 45
- Gait Process, 97
- GNU Radio, 9
- GPS, 135
- Ground Station, 9
- Gyroscope, 97
- HD, 75
- Helicopter, 27
- Hill-climber, 39
- Humanoid, 97
- Humidity sensor, 129
- illuminance, 3
- Impedance Matching, 121
- Inclinometer, 97
- inductive coupling, 57
- Integer Arithmetic, 103
- Interpolation, 103
- iterative projection algorithm, 69
- Java, 13
- KAREN, 141
- LED, 45
- LNA, 121
- Lumen, 3
- Manipulator, 39
- MATLAB, 121
- medical applications, 91
- Micro Computer, 97
- Microcontroller, 135
- microcontrollers, 91
- Micropropagation, 33
- Microstrip Line, 121
- modal testing, 79
- Nested Loop, 103
- network, 141
- NIRFAST, 115
- non-destructive sensing, 33
- optical pick-up, 75
- Optimisation, 39
- PDP, 51
- Phase locked, 79
- phase retrieval, 69
- Plant growth monitoring, 33
- power, 51
- Programming, 13
- Quadrotor, 21, 27
- Radio astronomy, 141
- radio astronomy, 69
- Radio telescope, 69

- range imaging, 109, 115
- Remote sensing, 33
- Roboard RB110, 97
- Robot, 135
- Robotic arm, 33
- RS232, 135

- Satellite Communication, 9
- scalable platform, 57
- SDR, 9, 103
- self-mixing, 79
- SimMechanics, 39
- Simulation, 39
- SQLite, 13
- square kilometer array, 69
- surveying, 135

- temperature sensor, 129
- Trajectory Planning, 39
- Tree disk, 85
- tsunami, 141

- UAV, 21, 27
- UDP, 141
- UDT, 141
- UHF, 121
- uniformity, 3

- Visual Flight Control, 21
- VLBI, 141
- volume measurement, 109

- WCK 1111 servo, 97
- Wireless Communication, 135
- wireless power, 57
- wireless sensor network, 129
- wireless sensor networks, 91
- WTA network, 91

- XOR gate, 51

Author Index

- Abrar, M., 9
Ahmad, N., 51
Alipour, M., 63

Bainbridge-Smith, A., 21
Barlow, P., 97

Carnegie, D. A., 39
Chen, J., 69
Cree, M. J., 109

Długosz, R., 91
David, S., 97
Dayal, R. P., 75
Dmochowski, P., 91
Dorrington, A. A., 109, 115
Dykstra, R., 45

Elser, V., 69

Gui, X., 9
Gulati, H., 13
Gulyaev, S., 141

Hasan, R., 51, 63, 121
Hassan, A., 115
Haughey, B. P., 39
Hayes, M., 21, 27, 85
Hitchings, M. B., 135
Hollitt, C., 45
Hu, A. P., 57

Künnemeyer, R., 75, 79, 115
Kacprazk, D., 3
Kacprzak, D., 57
Kancherla, S., 3
Knopp, S., 85

Li, J., 121

Millane, A., 27
Millane, R. P., 69
Molteno, T. C. A., 103
Mukhopadhyay, S., 129

Natusch, T., 141

Payne, A., 115
Przedwojski, P., 91

Quan, V. M., 129

Raval, P., 57

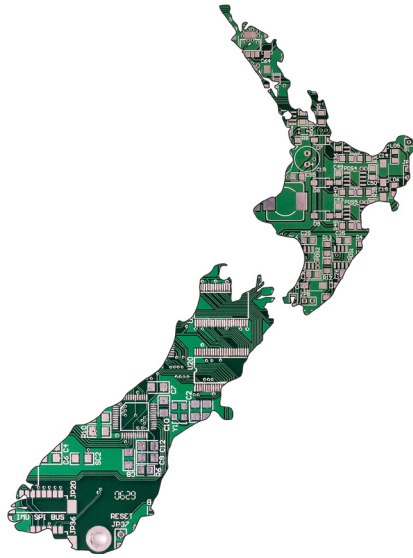
Scott, I. R., 103
Scott, J., 1
Seelye, J., 33
Seelye, M., 33
Sen Gupta, G., 33, 97, 129, 135
Shrestha, P., 79
Shrivastava, V., 109
Stowers, J., 21, 27

Talaśka, T., 91

Vercoelen, B., 45

Weston, S. D., 141

Yadla, N., 3



niche*com*

