

**Proceedings of the
13th Australian Space Science Conference
Sydney
September 30 - October 2, 2013**



Australian Space Science Conference Series

1st Edition

Published in Australia in 2014 by
National Space Society of Australia Ltd
GPO Box 7048
Sydney NSW 2001
Fax: 61 (02) 9988-0262
email: nssa@nssa.com.au
website: <http://www.nssa.com.au>

Copyright © 2014 National Space Society of Australia Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission from the publisher.

ISBN 13: 978-0-9775740-7-0
Editors: Wayne Short and Iver Cairns

Distributed on DVD



Preface to the Proceedings

A large number of the presenters at the conference later submitted completed written papers which form the basis of this Conference Proceedings.

All papers published in these proceedings have been subject to a peer review process whereby a scholarly judgement by at least two suitable individuals endorsed by the Program Committee determined if the paper was suitable to be published. All papers not rejected were revised until deemed suitable. Responsibility for the content of each paper lies with its author(s). The publisher retains copyright over the text. Papers appear in the Conference Proceedings with the permission of the authors.

The Editors would like to give special thanks to the Program Committee and those scholars who participated in the peer review process:

Elais Aboutanios, Alen Alempijevic, Brad Alexander, Jeremy Bailey, Daniel Bayliss, Adrian Brown, Ian Bryce, Graziella Caprarelli, Michael Chang, Manuel Cervera, Mazher Choudhury, Jong Chow, Michael Davis, Brett Dahlhaus, Andrew Dempster, Boyin Ding, Brian Fraser, James Gilmore, Eamonn Glennon, Daniel Hernandez, Jonathan Horner, Michael Ireland, Iwan Kelaiah, Barry Jones, Joseph Leach, Sarah Maddison, Tony Monger, Marc Norman, Dennis Odijk, Carol Oliver, John Olsen, John Page, Karsten Peters, Ian Petersen, Robert Pidgeon, Li Qiao, Chris Rizos, Lyle Roberts, Roy Sach, Murray Sciffer, Michael Smart, Mike Terkildsen, Matt Tetlow, Bob Vincent, Malcolm Walter, Martin Westwell, Phil Wilkinson, Rob Wittenmyer, Falin Wu, Xiaofeng Wu, Xianjie Zha and Xiyang Zhi.

Finally we would like to thank our sponsors (the Space Co-ordination Office, University of NSW and the Australian Centre for Astrobiology) for their support in funding student participation and the Organising Committee, Program Committee and colleagues Jeremy Bailey, Lucyna Kedziora-Chudczer, Andrew Dempster and Chris Rizos for giving generously of their time and efforts.

We trust that you will find the 2013 Conference Proceedings enjoyable and informative.

Wayne Short and Iver Cairns
Editors, 13ASSC Conference Proceedings, May 2014

Conference Background

The Australian Space Science Conference (ASSC) is the focus of scientific cooperation and discussion in Australia on research relating to space. It is a peer reviewed forum for space scientists, engineers, educators, and workers in Industry and Government.

The conference is of relevance to a very broad cross section of the space community, and therefore generates an enlightening and timely exchange of ideas and perspectives. The scope of the conference covers fundamental and applied research that that can be done from space and space-based platforms, and includes the following:

- **Space science**, including space and atmospheric physics, remote sensing from space, planetary sciences, astrobiology and life sciences, and space-based astronomy and astrophysics ;
- **Space engineering**, including communications, navigation, space operations, propulsion and spacecraft design, testing, and implementation ;
- **Space industry** ;
- **Space archeology** ;
- **Current and future Australian space projects** ;
- **Government, International relations and law** ;
- **Education and outreach** .

The 13th ASSC was held at the University of New South Wales (Kensington Campus) from September 30 to October 2, 2013. The Conference was opened by Pro Vice-Chancellor for Research of University of New South Wales (Research), Professor Mark Hoffman.

The conference included a comprehensive program of plenary talks and special sessions on the national context for space the foci and programs of Australian Government units with interests in space, and detailed descriptions and progress reports of research funded by the Australian Space Research Program. In addition, the program contained a special planetary science session, with a strong preponderance of projects involving the Mars Society of Australia. The program also contained multiple sessions of invited and contributed presentations, both oral and poster, on Propulsion, Planetary Science, Earth Observation and GNSS, Space Capabilities, Space Physics, Space Technology, Space Archeology, Education and Outreach and Indigenous Sky Knowledge.

Appendix A lists all abstracts accepted for presentation at the conference.

The 13th ASSC was organised by the National Space Society of Australia (NSSA), the Academy of Sciences National Committee for Space and Radio Science (NCSRS) and University of NSW. The Mars Society Australia (MSA) and Australian Space Research Institute (ASRI) also helped significantly with organising abstract submissions.

A call for papers was issued in March 2013 and researchers were invited to submit abstracts for presentation at the conference. Following the conference itself, a call for written papers was issued in October 2013: this invited presenters to submit a formal written paper for this Proceedings that covered their conference presentations.

Table of Contents

Preface to Proceedings	page iii
Conference Background	page iv
List of Proceedings Papers	page vii
Welcome to the 13th Australian Space Science Conference	page x
About the NSSA	page xii
About the NCSRS	page xiii
2013 Program Committee	page xv
2013 Organising Committee	page xvi
Conference Plenary Speakers	page xvii
Program	page xix

List of Proceedings Papers

Authors	Paper Title	
Ali Haydar Goktogan, Iwan Kelaiah, Salah Sukkarieh	Virtual Mars Yard Simulation System for Internet-Based STEM Education	pages 1– 10
Thien Nguyen	Finishing BLUEsat	pages 11– 22
Eren Gorur, Ivan Cheung, Tu Lesuma, Treves Li, Cecily Zhu, Richard McKenzie, Mathew Fenwick, Adam Ingle, Timothy Gollan and Jacob Hacker	Australian Space Policy: A Student Perspective	pages 23– 40
Jonathan Horner, Robert Wittenmyer, Tobias Hinse, Jonathan Marshall and Alex Mustill	Wobbling Ancient Binaries – Here Be Planets?	pages 41 – 54
Elodie Thilliez , Lea Jouvin , Sarah T. Maddison and Jonathan Horner	A dynamical test for the terrestrial planets in the habitable zone of HD 204313	pages 55– 68
Brett Addison	Planets in Spin–Orbit Misalignment and the Search for Stellar Companions	pages 69 – 82
Jonathan Clarke	Searching for life on early Mars: lessons from the Pilbara	pages 83 – 96
Emily Bathgate, Graziella Caprarelli, Linda Xiao, Ron Shimmon, Ross Pogson	Raman characterisation of the products of alteration of volcanic rock minerals: preliminary results and implications for Martian studies	pages 97 – 108
James Gilmore	Measurement Uncertainty in TRMM Observations of the Precipitation and Column Water Vapor Relationship	pages 109 – 116
Jonathan Horner, Dave Waltham and F. Elliot Koch	The role of Jupiter in driving periodic climate change	pages 117 – 128
David Herne, John Kennewell, Mervyn Lynch	Ionospheric Phenomena and Low–frequency Radio Astronomy	pages 129 – 140
Lewis Freeland, Fred Menk, Murray Sciffer, Michael Terkildsen	A Search for Signatures of Discrete Cosmic Radio Sources Using an Imaging Radiometer at Davis, Antarctica	pages 141 – 148
Anne Unewisse, Andrew Cool, Andrew Kirby, Dan Meehan	A New Relocatable Airglow Imager	pages 149 – 160
Naomi Altman, Sascha Schediwy, John McFerran, Michael Tobar, and Andre Luiten	Expanding Australia’s contribution to the ACES and future space missions	pages 161 – 170

List of Proceedings Papers

Authors	Paper Title	
Champlain Kenyi, Jeremy Bailey, Daniel Cotton	Retrieval of greenhouse gas concentrations from observations with a ground-based spectrometer in the near-infrared	pages 171– 182
Tony Monger	Frequency Allocation for Satellite Space Stations using Amateur Radio Bands	pages 183– 190
Mazher Choudhury, Eamonn Glennon, Andrew Dempsterh	Test Results of the Namuru Dual-GNSS Space-borne Receiver	pages 191 – 204
Eamonn Glennon, Scott O’Brien, Joon Wayn Cheong, Kevin Parkinson, Andrew Dempster	CubeSat-based GPS Reflectometry using the Namuru V3.2 GPS Receiver	pages 205 –216
Li Qiao , Min Zhang , Barnaby Osborne and Andrew G Dempster	Attitude Control System (ACS) of UNSW QB50 project “UNSW ECO”	pages 217 – 228
Liam Liyuan Li, Wei Zhang, Xiaojing Li and Linlin Ge	Satellite Vibrations and Restoration of High Resolution Pushbroom Images	pages 229 – 240
Tim Youtian Liu, Jicang Wu, Xiaojing Li and Linlin Ge	Phase Unwrapping for Interferometric Synthetic Aperture Radar Technique	pages 241 – 252
John Page, Haoyang Cheng, Monica Chi, Nathan Kinkaid	Swarm-Based Satellite Constellation Control	pages 253 – 260
Kyll Schomberg, Graham Doig and John Olsen	Computational Analysis of Pintle Variation in an Expansion- Deflection Nozzle	pages 261 – 272
William Reid, Callum McCarter, Ali Haydar Goktogan	A Highly Mobile Wheel-on-Leg Planetary Rover for use in a Mars Analogue Environment	pages 273 – 282
Anthony Tompkins, Steven Potiris, Ali Haydar Göktoğan	Development of a Low-Cost Vision-Based Localisation System for the Experimental Mars Rover	pages 283 – 294
Steven Potiris, Anthony Tompkins, Ali Haydar Göktoğan	A Vision-based Autonomous Docking System of the Experimental Mars Rover	pages 295 –306
APPENDIX A	13ASSC: List of Presentations & Posters	pages a – j



Welcome to the 13th Australian Space Science Conference

and to the University of New South Wales! This will be the seventh ASSC jointly sponsored and organised by the National Committee for Space Science (now renamed the National Committee for Space and Radio Science or NCSRS) and the National Space Society of Australia (NSSA). The ASSC is intended to be the primary annual meeting for Australian research relating to space science. It welcomes space scientists, engineers, educators, and workers in Industry and Government.

The conference was opened by Professor Mark Hoffman, Pro Vice-Chancellor for Research at the University of New South Wales.

The 13th ASSC has over 120 accepted abstracts across Australian space research, academia, education, industry, and government.

We would like thank the Space Co-ordination Office, University of New South Wales and the Australian Centre for Astrobiology for their sponsorship. Special thanks also go to the Australian Space Research Institute (ASRI), Engineers Australia and Mars Society Australia (MSA) for their support.

We look forward to an excellent meeting!

Iver Cairns
Co Chair ASSC 2013
University of Sydney

Wayne Short
Co Chair ASSC 2013
President, NSSA



The National Space Society of Australia is the coming together of like-minded space enthusiasts who share a vision for the future in which there is an ambitious and vigorous space program leading to eventual space settlement.

To this end the National Space Society (worldwide) promotes interest in space exploration, research, development and habitation through events such as science and business conferences, speaking to the press, public outreach events, speaking engagements with community groups and schools, and other pro-active events. We do this to stimulate advancement and development of space and related applications and technologies and by bringing together people from government, industry and all walks of life for the free exchange of information.

As a non-profit organisation, the National Space Society of Australia draws its strength from an enthusiastic membership who contributes their time and effort to assist the Society in pursuit of its goals.

For more information, and to become a member:

<http://www.nssa.com.au>

Ad Astra!
Wayne Short
NSSA President



The National Committee for Space and Radio Science (NCSRS), formerly the National Committee for Space Science (NCSS), is chartered by the Australian Academy of Science to foster space science, to link Australian space scientists together and to their international colleagues, and to advise the Academy's Council on policy for science in general and space science in particular. The web page is at <http://www.science.org.au/natcoms/index.htm> .

NCSRS believes that ASSC meetings provide a natural venue to link Australian space scientists and foster the associated science, two of its core goals. As well as ASSC, NCSRS is also sponsoring the VSSEC – NASA Australian Space Prize.

This is the third ASSC meeting following the launch of the first Decadal Plan for Australian Space Science. NCSRS encourages people to work together to accomplish the Plan's vision: "Build Australia a long term, productive presence in Space via world-leading innovative space science and technology, strong education and outreach, and international collaborations."

2013 ASSC Program Committee

Jeremy Bailey (University of NSW)

Iver Cairns (Chair, University of Sydney)

Jonathan Clarke (Mars Society of Australia)

Andrew Dempster (University of NSW)

Kerrie Dougherty (Powerhouse Museum)

Alice Gorman (Finders University)

Duane Hamacher (University of NSW)

Trevor Harris (Defence Science and Technology Organisation)

Trevor Ireland (Australian National University)

Carol Oliver (University of NSW)

Gordon Pike

Chris Rizos (University of NSW)

Richard Samuel (Australian Space Research Institute)

Michael Smart (University of Queensland)

Roy Sach

Murray Sciffer (University of Newcastle)

Kefei Zhang (RMIT University)

2013 ASSC Organising Committee

Jeremy Bailey

University of NSW

Iver Cairns

Co Chair ASSC 2013

University of Sydney

Chair, ASSC Program Committee

Aditya Chopra

Australian National University

Andrew Dempster

University of NSW

Scott Dorrington

University of NSW

Lucyna Kedziora-Chudczar

University of NSW

Chris Rizos

University of NSW

Wayne Short

Co Chair ASSC 2013

President, NSSA

Conference Plenary Speakers

Professor Andrew Dempster
University of NSW
“ACSER & space engineering”

John Goldsmith
Curtin University
‘The “Ilgarijiri- “Things Belonging to the Sky” project:
Collaboration between Aboriginal communities and radio astronomy in Australia’

Dr Craig O’Neil
Macquarie University
“The tectonics of exoplanets”

Professor Chris Rizos
University of NSW
“GNSS Trends: Where are we now and where are we going?”

Conference Plenary Speakers

Professor Steven Tingay
Curtin University
“The Murchison Widefield Array, Precursor for the Square Kilometre Array”

Professor Colin Waters
University of Newcastle
“Probing near-Earth Space using ULF signals”

Dr David Williams
CSIRO
“Lessons from the UK for space research & space agencies”

David Willson
NASA / Ames & MSA
“Mars sample return, Dragon-X and other missions”

13th ASSC Conference Program

September 30 - October 2

Time	Monday		Tuesday		Wednesday	
7:30	Registration		Registration		Registration	
8:00	Registration		Registration		Registration	
9:00	Opening / Nat Context <i>Physics Theatre</i>		Plenaries <i>Physics Theatre</i>		Plenaries <i>Physics Theatre</i>	
10:30	Break		Break		Break	
11:00	Plenary & Govt Units <i>Physics Theatre</i>		ASRP <i>Physics Theatre</i>	Planetary Science - MARS Room G31	Indigenous Sky Knowledge <i>Physics Theatre</i>	Planetary Science - Solar system Room G32
13:00	Lunch		Lunch		Lunch	
14:00	Space Physics I <i>Physics Theatre</i>	Space Engineering I Room G31	Earth Observation, GNSS <i>Physics Theatre</i>	Space Physics II Room G31	New Instruments, Techniques, Space Projects <i>Physics Theatre</i>	Space Physics III Room G32
16:00	Break		Break		Break	
16:30	Posters & Cocktail Reception Level 2		Space Industry <i>Physics Theatre</i>	Space Engineering II Room G31	Space Education & Humanities <i>Physics Theatre</i>	Planetary Science - Exoplanets Room G32
18:00	Close		Close		Closing Remarks	
18:15						
18:30					Close	

Virtual Mars Yard Simulation for Internet-based Science, Technology, Engineering and Mathematics Education

Iwan Kelaiah, Ali Haydar Göktoğan, Salah Sukkarieh

*Australian Centre for Field Robotics
The University of Sydney
Sydney, NSW, Australia, 2006*

Summary: The Virtual Mars Yard (VMY) is a simulation system for Internet-based Science, Technology, Engineering and Mathematics (STEM) education. It closely mimics the Internet-based teleoperation of Experimental Mars Rover (EMR)-Mawson, a six-wheeled robotic platform with a teleoperable sensor payload being operated at the Mars Yard at the Powerhouse Museum in Sydney, Australia. The VMY offers a realistic kinematic and physics-based dynamic model interacting with an accurate virtual Mars Yard terrain model. It also provides users with a similar graphical user interface to the real teleoperation system. The VMY allows students to practice driving the EMR-Mawson in the virtual environment and learn how the interface works. Potential teleoperation system users can also enhance and test their skills before using the real system at the Mars Yard, with the potential of reducing the risk of damaging the actual system, for long-term operation in STEM education. This paper presents the architecture and applications of the VMY.

Keywords: Virtual, simulation, Mars Yard, hands-on learning, STEM, technology, engineering

Introduction

Experimental Mars Rover (EMR)-Mawson is a robotic teleoperation system that engages high school students in Science, Technology, Engineering and Mathematics (STEM) education. The teleoperation interface allows users to control the pan/tilt camera, drive the rover, spot-turn and crab motion in a Mars-like environment called Mars Yard, which is located at the Powerhouse Museum in Sydney, Australia. Mars Yard resembles the surface of Mars and has features such as gypsum evaporite, a meteorite, sandstones, hematite globules, a group of basalt rocks and an eroded plain. Stromatolites are also included in the Mars Yard as an example of what past life on Mars, if it ever existed, could look like. The teleoperation of EMR-Mawson at the Mars Yard enables students to undertake simulated robotic planetary missions. For example, they can explore an area of scientific interest, investigate rocks and work out how to conduct a mission with the minimum number of commands, all of which are features of rovers actually on Mars. Minimising the number of commands required is particularly useful, as it is expensive in terms of both cost and energy to send messages to Mars.

Teleoperating EMR-Mawson is different from driving a remote-controlled car because there is a time delay between sending commands and seeing them in action. In addition, it is necessary to plan routes, as some rocks may slow the rover down or prevent it from moving. Some students may have less hands-on experience with such systems than others and thus may require more time to learn the rover's controls. Some may damage the rover by performing dangerous manoeuvres, such as driving with left wheels on a flat surface while



(a) EMR-Mawson at the Mars Yard



(b) Virtual EMR-Mawson in Virtual Mars Yard

Fig. 1: Virtual Mars Yard mimics the environmental features of Mars Yard and EMR-Mawson

the others are on a steeply inclined slope or large rocks. Such damage may prevent other students from accessing the teleoperation system, thereby affecting the overall engagement with STEM.

This paper presents the Virtual Mars Yard (VMY), which is an Internet-based virtual simulation that focuses on simulating the teleoperation of EMR-Mawson. It provides a realistic Mars Yard terrain, a similar graphical user interface (GUI) to the Mars Yard teleoperation system and a virtual EMR-Mawson with an accurate kinematic model (see Figs. 1 and 5). The benefits of the simulation are that students can take the Mars Yard experience home for extra practice, understand the limitations of the rover, and safely explore what-if missions without risking the actual system.

The following sections of this paper discuss relevant work in the field of virtual simulation for STEM education. This is followed by the architecture and application of the VMY, a conclusion and suggestions for future work.

Related Works

In the field of Modelling and Simulation (M&S), virtual simulation plays an important role in developing and testing high-risk and complex robotics systems such as space robotics [1]. Chiefly, various configurations of systems can be trialled without risking the actual system [2,3]. Virtual simulations place real users in simulated systems through which they can experience and experiment with phenomena that are not available in real life for a variety of reasons including time, ethical, physical and fiscal constraints.

Virtual simulations of planetary exploration rovers, such as Rover Analysis, Modelling and Simulation (ROAMS) [4,5], three-dimensional (3D) Virtual Rover Operation Simulator (VROS) [2] and rover simulation based on terramechanics and dynamics (ROSTDyn) [6], present accurate kinematic and dynamic simulations of rovers so that engineers can test rover designs and various configurations in them, with room for trial and error.

Virtual Astronaut (VA) [7] is an Internet-based virtual simulation of the Opportunity rover around the Santa Maria Crater in the Meridiani Planum on Mars. This simulation, for geologists, re-enacts the Opportunity rover's scientific mission by using the Planetary Data System (PDS) archives to display an accurate representation of the Santa Maria Crater.

However, the VA does not have an accurate kinematic model, as the rocker-bogie and steer actuators are static; thus, in promoting STEM education to a wider audience, the VA may give the wrong information about how the actual rover was engineered. Further, users cannot access the pan/tilt camera or drive beyond the pre-defined path.

These rover simulation systems are lacking in a number of areas in promoting STEM education. First, the realism of the behavioural aspect of the simulation is not complemented by its visual aspect or vice versa. If the behavioural aspect of the system is realistic, but the visualisation aspect is not, it may alter the perception of users. Studies show that visual realism in virtual reality increases presence, which in turn enables users to respond to the system in a realistic manner [8,9]. However, if the visualisation is realistic, but does not show a realistic behaviour of the system, the user may develop flawed knowledge about the system. For example, in the case of the rover simulation not allowing wheel steering animation during turning, users may perceive that the rover can turn without steering the wheels. By contrast, a realistic simulation would assist users in making judgements or decisions in the actual systems. The VMY is thus not geared towards a game-like, fast and instantaneous experience, but rather a simulation of the actual system.

A second area in which previous simulations are lacking is that these virtual rover simulations are not commonly available as web-based applications available to the public to support STEM education in classrooms. Conversely, major Internet browsers on Windows and Mac OSX operating systems can run the VMY, which is publicly available online. The VMY has a realistic Mars Yard terrain and virtual EMR-Mawson with a realistic kinodynamic. Users can control the virtual rover by driving, spot-turning, crabbing, operating lasers and panning/tilting the on-board camera, as can be done in the real rover.

The Architecture of Virtual Mars Yard Simulation

Fig. 2 shows the architecture of the VMY. Each block represents a model and the directional arrows represent the flow of information. In replicating the Mars Yard teleoperation of EMR-Mawson, the VMY has a number of objectives:

- **Objective 1:** A GUI should be provided that allows users to interact with virtual EMR-Mawson and virtual cameras, similar to the real teleoperation system. A similar user interface will assist users to implement the skills learned in the VMY into the real system immediately, rather than adjusting themselves to an unfamiliar system.
- **Objective 2:** The coordinates positions of the panels and title screen should be easily changed without recompiling entire project files. This objective will ensure that GUI changes in the actual system can be incorporated quickly into the VMY to support Objective 1.
- **Objective 3:** Virtual EMR-Mawson should have a realistic kinodynamic with controllable steering and driving actuators via a GUI, as in the real system. This will allow users to test mission strategies and identify problems as the virtual rover moves on the 3D Mars Yard terrain.
- **Objective 4:** Virtual Mawson should have payloads similar to the real system. For example, it should have a pan/tilt camera, laser pointers and a solar array. Teleoperation activities that require a first-person view via the pan/tilt camera and laser pointers must be replicable in the VMY. This will help users to plan their missions virtually.
- **Objective 5:** The settings of virtual cameras and on-board cameras, such as pixel resolutions and field of view, must be easily editable. This objective will enable

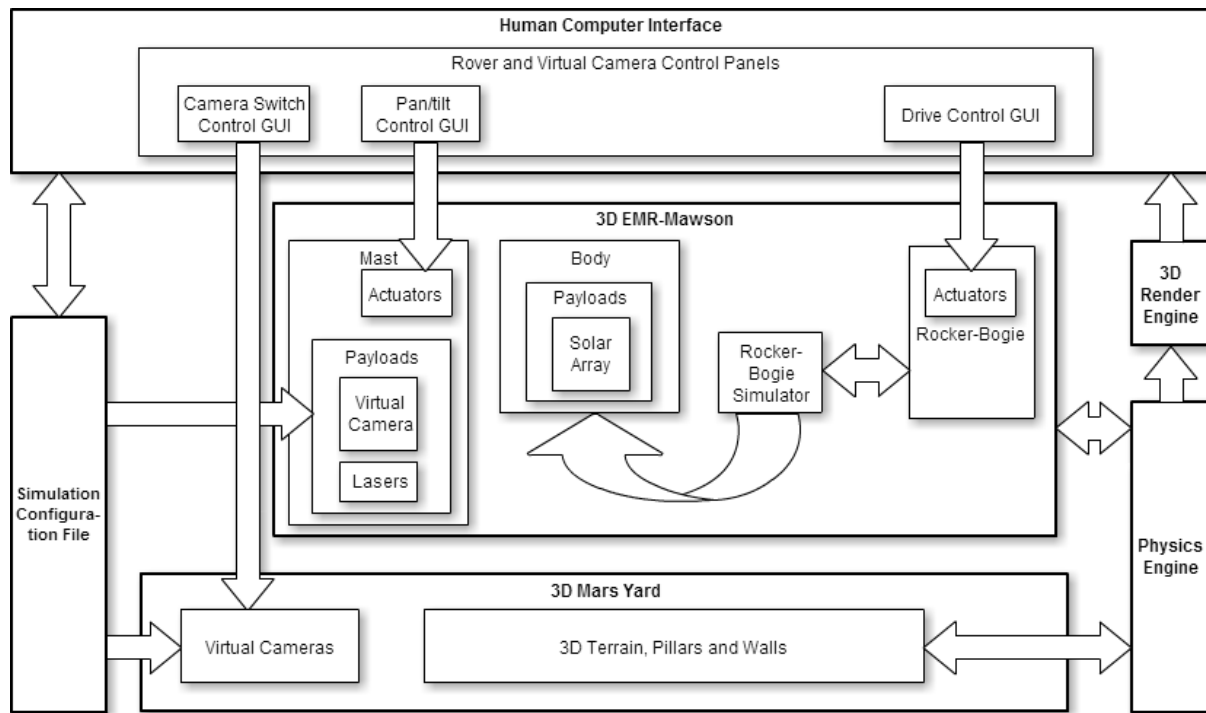


Fig. 2: Architecture of the Virtual Mars Yard application

developers to update settings quickly when incorporating changes in the camera settings at the Mars Yard.

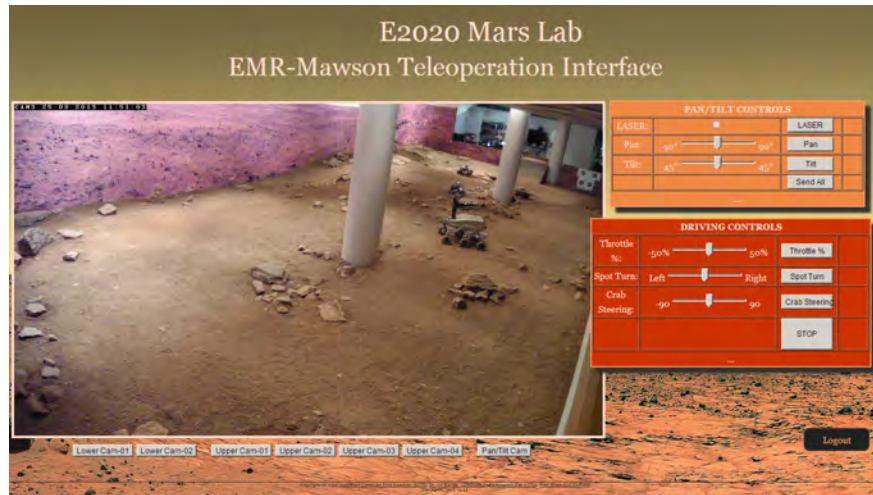
- **Objective 6:** The VMY should contain an accurate representation of the Mars Yard terrain and its features, pillars and walls, and virtual cameras. Virtual EMR-Mawson must be able to drive on the terrain. Virtual cameras must be placed at the top corners of the virtual walls to improve users' situational awareness in the VMY, similar to in the real teleoperation system. In combination with Objective 3, this will ensure the long-term operation of the real system in fostering STEM education.

Human-Computer Interface

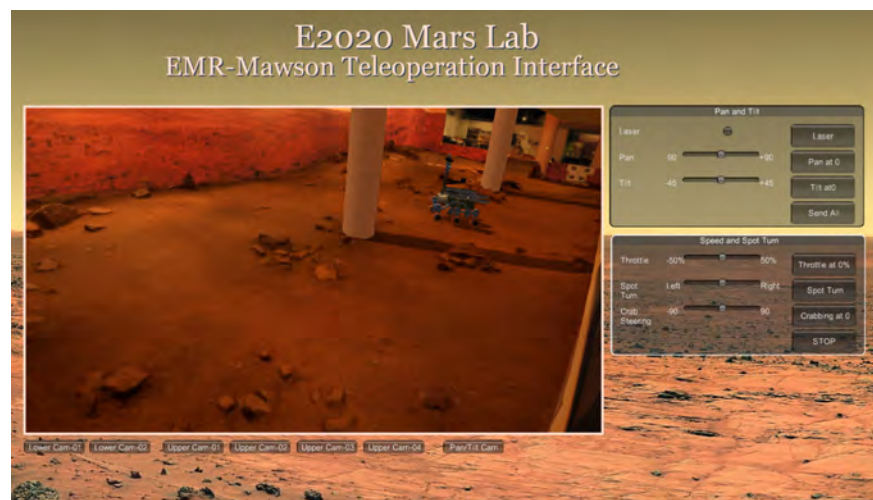
The human-computer interface refers to the medium of interaction between users and the VMY. The interface shows 3D animations of the virtual simulation from the 3D rendering engine, depending on the user's input. The rover and virtual camera control panels (see Fig. 3b) consist of floating panels and buttons for users to control the pan/tilt camera, drive the rover and switch between camera views (Objective 1). Developers can alter the simulation configuration data.

3D Rendering Engine

The 3D rendering engine generates a two-dimensional (2D) image on the screen based on the computation of lighting, textures, post-processing effects and the transformation of 3D objects as pre-computed by the physics engine.



(a) Mars Yard teleoperation GUI



(b) Virtual Mars Yard GUI

Fig. 3: The GUI of the VMY is similar to the GUI of the teleoperation system

Physics Engine

3D EMR-Mawson interacts with the terrain via a physics engine that manages collision detection and rigid body dynamics to give the rover realistic motion over the 3D terrain on the screen (supporting Objectives 4 and 6). The physics engine computes the transformation of rigid body objects in the virtual world before they are rendered on the screen.

3D Mars Yard

The 3D Mars Yard model consists of a 3D representation of the Mars Yard terrain, supporting columns and walls, and the virtual cameras (Objective 6). The VMY captures the Mars Yard features accurately (see Figs. 1b and Fig. 4). Multiple high-resolution photographs were taken with a digital camera and stitched together to re-create the surrounding walls, including the Mars diorama and the Mars Lab area.

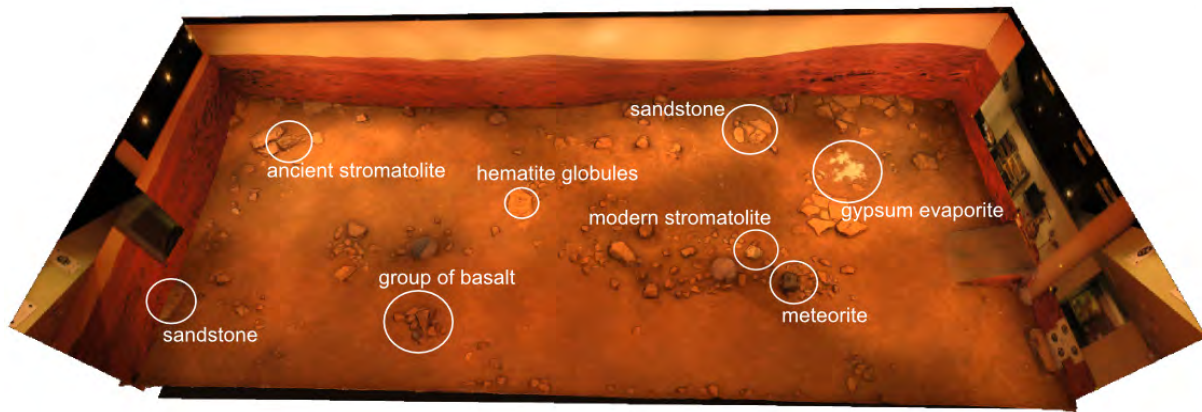


Fig. 4: The VMY shows an accurate terrain at the Mars Yard, including sites of interest as marked by white circles

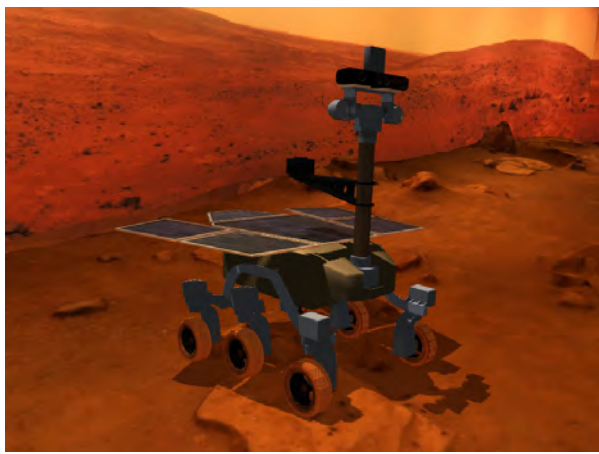
The virtual cameras model replicates the actual wall-mounted upper and lower cameras in the Mars Yard. These cameras are needed to increase the situational awareness of users in planning and controlling the EMR. There are four upper cameras and two lower cameras that give a third-person view of the Mars Yard from different angles. The upper cameras show the high-angle view of the Mars Yard from the corners of the walls, while the lower cameras are located on the wall of the Mars Lab area (see the right-hand side of Fig. 4). Users may send a command to conduct a crab motion but not see any movement. Switching to other upper/lower camera views may reveal a problem; for example, the wheels may be sitting hard against a large rock. Any camera at the Mars Yard can be substituted with another camera with different intrinsic settings, and the virtual counterpart can be copied by editing the configuration file.

3D EMR-Mawson

3D EMR-Mawson consists of 3D models of the mast, body, rocker-bogie and their components to simulate the behaviour of EMR-Mawson (Objective 3). The mast has payload components including lasers and on-board cameras. The mast also contains an actuator component to allow pan/tilt camera movements similar to the real EMR counterpart. It can be controlled via the pan/tilt control GUI.

The body has a payload component for attaching virtual instruments. The pitch angle of the body is set by the rocker-bogie simulator, which mimics the kinematic model of the EMR (see Fig. 5). The left and right rockers of the EMR are directly connected through a differential rather than a shaft; thus, if one rocker goes up because of an obstacle, the other rocker goes down. This configuration minimises the EMR's body pitch angle to keep the solar panel as parallel to the surface as possible. One end of the rocker has an actuated wheel, and the other end is a pivoted bogie with a pair of actuated wheels. This rocker-bogie simulator keeps the pitch angle of the body minimal, as in the real EMR-Mawson.

The rocker-bogie has an actuator component to simulate realistic steering behaviour. It simulates 12 actuators, six for steering and six for driving, as in the actual EMR. It has a GUI component to display a panel for driving, spot-turning, crabbing and stopping (see Fig. 3b).



(a) The rover is parallel to the ground as right rocker-bogie tilts up due to terrain elevation



(b) The left bogie tilts without affecting the overall tilting of the body

Fig. 5: Realistic kinodynamic of virtual EMR-Mawson. The rocker-bogie simulator gives realistic rover movements on virtual terrain, in line with those of the real EMR on the Mars Yard surface

Simulation Configuration File

The VMY uses an editable configuration file to modify the properties of the simulation quickly without recompiling all project files. Variables in this file are needed to set up the locations of the panels on the screen, as well as the intrinsic settings of the virtual cameras of the 3D terrain and the virtual rover (Objective 2 and 5).

The Application of the Virtual Mars Yard Simulation

Using the VMY, users can become familiar with the EMR controls and the Mars Yard terrain. Users may sometimes fail a mission in the VMY by neglecting the time delay between sending commands and seeing them in action. This is an opportunity for teachers or supervisors to ask questions such as ‘If you repeat that mission, what will you do differently?’ This can be an avenue for users to devise better strategies to control the EMR more effectively in future missions.

Realism in the VMY is important for STEM education. Firstly, the VMY contains material for STEM education from sites of interest to the rover’s kinematics and payloads. Secondly, it has the potential to minimise discrepancies between expectations and observations to support learning [10]. Just as obtaining a driver’s licence requires learners to practice under realistic traffic conditions, before letting users teleoperate the EMR, they need to be trained and tested in a similar environment. The VMY provides this training and testing environment, for users to move from learners to fully licensed drivers.

Further, the VMY offers the contextualisation of STEM-related subjects. Contextualisation via virtual simulation provides a richer experimental context and contributes to better understanding between the theoretical model and the experimental model [11]. For example, consider the Pythagorean formula $a^2 + b^2 = c^2$ in the context of Fig. 6. With a little creativity, such as by driving the virtual EMR and shooting the laser beams onto a rock and providing the height and tilting angle of the laser, teachers could ask students to find the

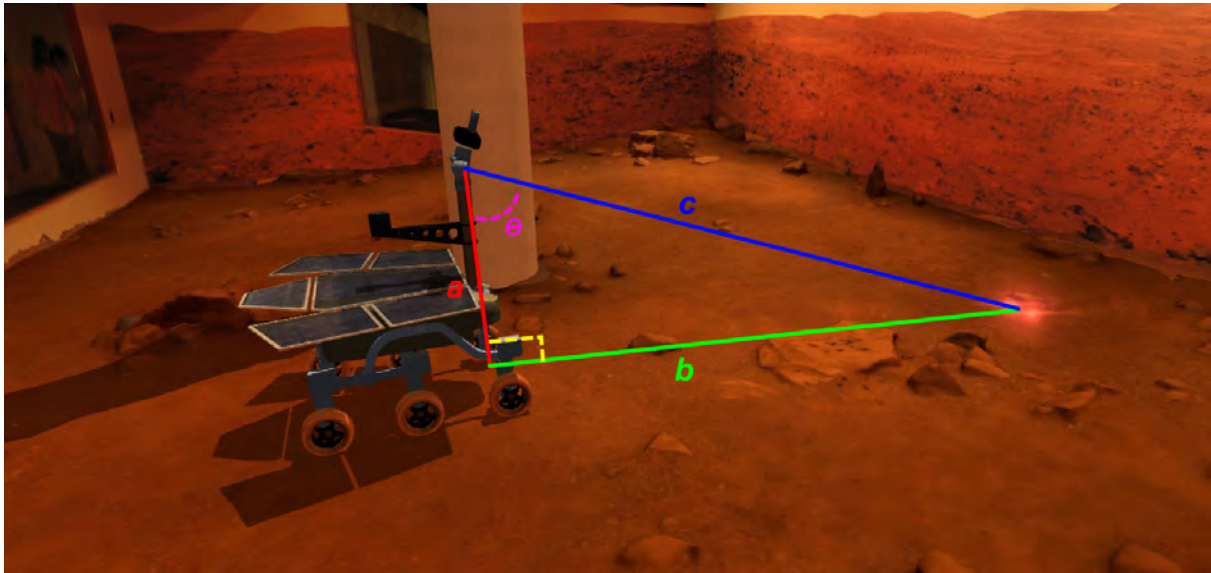


Fig. 6: The VMY can be used to learn trigonometry in the context of space robotics: approximate distance to the rock (b) can be determined if the height of the laser (a) and the angle θ (θ) is known. These lines do not appear in the VMY, instead, they show how trigonometry lessons can be achieved in a classroom setting

distance between the rover and the rock. Without placing this theorem into a genuine or interesting context, students may struggle to see its significance to their lives, potentially reducing students' interest in continuing studies in STEM-related disciplines. Another example would be for a physics lesson. A teacher could ask students to drive the virtual EMR-Mawson at the Mars Yard, the VMY is a virtual simulation that allows users to practise controlling the pan/tilt camera and lasers, as well as driving, spot-turning and crabbing actions as they would in the real teleoperation system. The VMY has the potential to engage and motivate students in STEM education through contextualisation and uncertainty whenever and wherever they are. Future work could make the VMY accessible on mobile devices and enable data exchange between widely used engineering applications.

The VMY also has elements of uncertainty that have the potential to increase users' motivation [12]. Participants may end the simulated mission in a failure, providing an opportunity for teachers to ask questions such as 'Should you run the same mission, what would you change in your rover?' This would allow students to think about the possibility of new rovers or solutions such as the addition of new sensors to avoid a similar predicament in future missions.

Conclusion and Future Work

While hands-on activities that simulate real-world phenomena can be used to engage students in solving real-world problems [13], time and physical constraints may not permit all students to have adequate hands-on experience. To solve this problem as regards teleoperating EMR-Mawson at the Mars Yard, the VMY is a virtual simulation that allows users to practise controlling the pan/tilt camera and lasers, as well as driving, spot-turning and crabbing actions as they would in the real teleoperation system. The VMY has the potential to engage and motivate students in STEM education through contextualisation and uncertainty whenever and wherever they are. Future work could make the VMY accessible on mobile devices and enable data exchange between widely used engineering applications.

Acknowledgement

This work was supported by the Broadband Enabled Education and Skills Services Programme, Department of Education, Australian Federal Government (2012–2014) called ‘Education 2020: Enabling learning in science, engineering and mathematics’. The authors thank Mr Esa Attia from the Australian Centre for Field Robotics (ACFR) for his help in generating an accurate 3D topography of the Mars Yard.

References

1. Rossmann, J. and Schluse, “Virtual Robotic Testbeds: A Foundation for e-Robotics in Space, in Industry—and in the Woods”, *Proceedings of the Developments in E-systems Engineering (DeSE) 2011 Conference*, Dubai, December 6–8, 2011, pp. 496–501.
2. Yu, C.-Q., Ju, H.-H. and Gao, Y., “3D Virtual Reality Simulator for Planetary Rover Operation and Testing”, *Proceedings of the Virtual Environments, Human-Computer Interfaces and Measurements Systems Conference*, Hong Kong, May 11–13, 2009, pp. 101–106.
3. Mao, N., Su, B., Yao, Q., Yang, S. and Xu, H., “Research on Virtual Scene Simulation for Planetary Rover”, *Proceedings of the Information and Automation Conference*, Shenzhen, June 6–8, 2011, pp. 685–689.
4. Jain, A., Guineau, J., Lim, C., Lincoln, W., Pomerantz, M., Sohl, G. and Steele, R., “ROAMS: Planetary Surface Rover Simulation Environment”, *Proceedings of the International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*, Nara, Japan, May 19–23, 2003.
5. Jain, A., Balaram, J., Cameron, J., Guineau, J., Lim, C., Pomerantz, M. and Sohl, G., “Recent Developments in the ROAMS Planetary Rover Simulation Environment”, *Proceedings of the Aerospace Conference, Vol. II*, March 6–13, 2004, pp. 861–876.
6. Li, W., Ding, L., Gao, H., Deng, Z. and Li, N., “ROSTDyn: Rover Simulation Based on Terramechanics and Dynamics”, *Journal of Terramechanics*, vol. 50, no. 3, 2013, pp. 199–210. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022489813000475>
7. Wang, J., Bennett, K.J. and Guinness, E.A., “Virtual Astronaut for Scientific Visualization—A Prototype for Santa Maria Crater on Mars”, *Future Internet*, vol. 4, no. 4, 2012, pp. 1049–1068. [Online]. Available: <http://www.mdpi.com/1999-5903/4/4/1049>
8. Slater, M., Khanna, P., Mortensen, J. and Yu, I., “Visual Realism Enhances Realistic Response in an Immersive Virtual Environment”, *Computer Graphics and Applications, IEEE*, vol. 29, no. 3, 2009, pp. 76–84.
9. Yu, I., Mortensen, J., Khanna, P., Spanlang, B. and Slater, M., “Visual Realism Enhances Realistic Response in an Immersive Virtual Environment—Part 2”, *Computer Graphics and Applications, IEEE*, vol. 32, no. 6, 2012, pp. 36–45.
10. Lindsay, E.D., Murray, S., Liu, D.K., Lowe, D.B. and Bright, C.G., “Establishment Reality vs. Maintenance Reality: How Real is Real Enough?”, *European Journal of Engineering Education*, vol. 34, no. 3, 2009, pp. 229–234. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/03043790902902906>
11. Machet, T., Lowe, D. and Gütl, C., “On the Potential for Using Immersive Virtual Environments to Support Laboratory Experiment Contextualisation”, *European Journal of Engineering Education*, vol. 37, no. 6, 2012, pp. 527–540. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/03043797.2012.721743>

12. Ozcelik, E., Cagiltay, N.E. and Ozcelik, N.S., "The Effect of Uncertainty on Learning in Game-Like Environments", *Computers & Education*, vol. 67, 2013, pp. 12–20. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360131513000481>
13. Mathers, N., Göktoğan, A., Rankin, J. and Anderson, M., "Robotic Mission to Mars: Hands-on, Minds-on, Web-Based Learning", *Acta Astronautica*, vol. 80, 2012, pp. 124–131.

Finishing BLUEsat - Lessons learnt from 15 years of managing an undergraduate space engineering project.

Thien H. Nguyen*, Ian S. Bartlett*, James Roberts-Thomson* and Chi Ieong Kam*

* *BLUEsat, The School of Electrical Engineering and Telecommunications, The University of New South Wales, NSW, Australia, 2052*

Summary: The BLUEsat project is a UNSW student led project with aims to build, launch and operate small satellites. After 15 years, the first successful iteration of the project is now complete. The project's long history provides insight into the pitfalls and successes of running a long-term student led project. BLUEsat remains an important fixture at UNSW in educating undergraduates in space engineering.

Keywords: BLUEsat, Undergraduate Space Projects, Project Management, Student Satellites

Introduction

The BLUEsat (Basic Low-Earth Orbit UNSW Experimental Satellite) Student Satellite Project has been an on-going student project at UNSW with the goal to build, launch and operate an Amateur Radio Microsatellite. In doing so, the Project had two main goals -

- 1) Increase awareness about space and space technology through the construction of space-hardware by students, and
- 2) Provide undergraduate engineers with practical experience in large engineering projects and creating space hardware.

The original project finally came to a close in 2013 after 15 years of development, with a complete functional demonstration of the fully assembled satellite. The protracted lifetime of the project revealed several issues with the management of undergraduate student resources. Key lessons learnt from the completion of BLUEsat need to be noted in order to successfully run and lead a complex engineering project using an entirely volunteer undergraduate student workforce. This paper outlines how these issues were identified and addressed in order to drive BLUEsat to its now ongoing success.

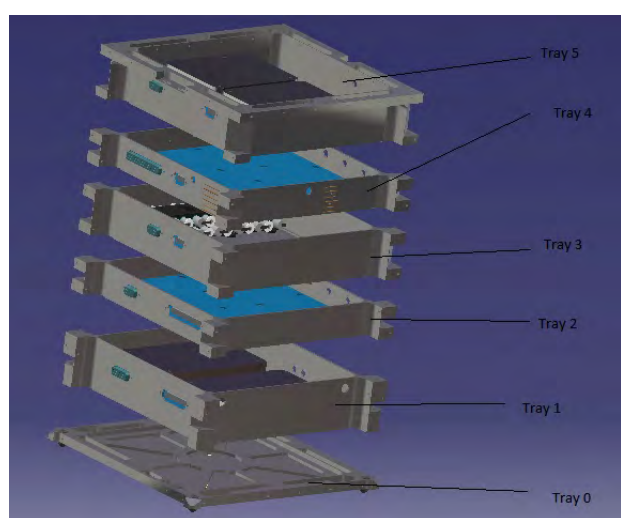
Satellite Specifications

The BLUEsat satellite was a 15kg microsatellite designed to fit in a 250mmx250mmx250mm envelope. The satellite was intended to service the amateur radio and scientific community by providing the following services

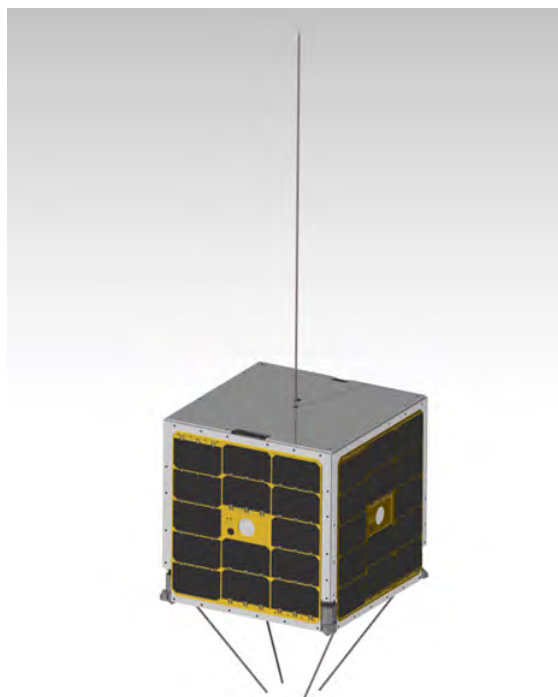
- 1) Bent-pipe RF service over Amateur Radio Bands

- 2) Extensible payload bus interface for generic scientific payloads. This includes power, control and radio-link support.

The satellite design shown in Figure 1 was based on the generic 5-tray Amateur Radio Satellite (AMSAT) design for microsatellites such as that used by Satellite "AMSAT Echo AO-51" [1]. The design consisted of a customized mechanical structure, housing a power distribution and ARM7 on-board computer designed and assembled by undergraduate BLUEsat engineering students. The two redundant radio transceiver units were commercial-off-the-shelf (COTS) components, purchased and integrated with the rest of the satellite systems.



(a) Exploded Satellite Assembly



(b) Assembled Satellite

Fig. 1: Satellite CAD Model

The Society

The BLUEsat project was made up of an entirely undergraduate workforce with some guidance from academic supervisors at UNSW. All development, construction and testing work was conducted at the BLUEsat "Ground Station" room in the Electrical Engineering building at UNSW in Kensington. At the project's peak in 2011, the project had 34 official members, with 20 students from the School of Electrical Engineering, 4 from Mechanical Engineering, 4 from Photovoltaic and Renewable Energy Engineering and 5 Computer Science students. The project was lead by a three man executive team elected at the Annual General Meeting. Historically the project has had academic supervisors from both the School of Mechanical Engineering and the School of Electrical Engineering. The last iteration of the project was supervised by Dr. Elias Aboutanios (since 2007) and Dr. Barnaby Osbourne (since 2012) from the School of Electrical Engineering.

Background

Conception

The BLUEsat project began in 1997 as a student branch of the National Space Society of Australia (NSSA) at UNSW. The group was predominantly comprised of undergraduate engineering students with a strong interest in amateur radio. Between 1997-2000, interest in creating an amateur radio satellite grew, resulting in the inception of the BLUEsat Project. During this period, the groundwork for the project was established, with funding and a ground station established for project development. The original satellite concept with external camera is shown in Figure 2. The three year period ended with the first set of engineering studies and preliminary designs completed and published as honours theses at UNSW [2]–[5]. The development of the satellite then occurred over three iterations of the project, discussed below.



Fig. 2: The original BLUEsat concept.

Iteration One, 2000-2004

The first iteration of satellite development started in 2000. The satellite was designed using the 5 tray structure from the original AMSAT-NA Microsat design [6]. The group also developed the first basic flight computer for the satellite using the ARM based StrongArm (SA) 1100 chip.

A prototype satellite structure with on-board computer (OBC) and power system was constructed. However these prototypes revealed critical flaws in the initial design. A preliminary design review revealed that the initial system budgets were not adequate for the desired satellite mission. As such the OBC needed expanded functionality and more power was required than the original silicon solar array design could provide. The extra system bus requirements then necessitated a larger and more protective mechanical structure. The first iteration ended with several key members graduating in the 2001-2002 period, causing the project to ramp down and progress to slow before 2004.

Iteration Two, 2004-2007

The second iteration of the satellite development started in 2004. The team aimed to expand the capabilities and address key issues from the first iteration. This was achieved by a major structural redesign, OBC expansion and power system overhaul. The following were implemented by the end of 2007:

- A new, larger structure that was capable of accommodating more functional sub-systems.
- Thermal and acoustic tests confirming the theoretical performance of the new structure [7]. The test apparatus is shown in Figure 3.
- Gallium Arsenide solar panels, providing the power required by the new on board systems.
- High power density NiMH battery packs.
- Expanded StrongArm 1100 computer architecture to include more interfacing capabilities.
- Temperature and power monitors for each satellite sub-system.

This period culminated in the first successful 'FlatSat' assembly of subsystems, pictured in Figure 4.

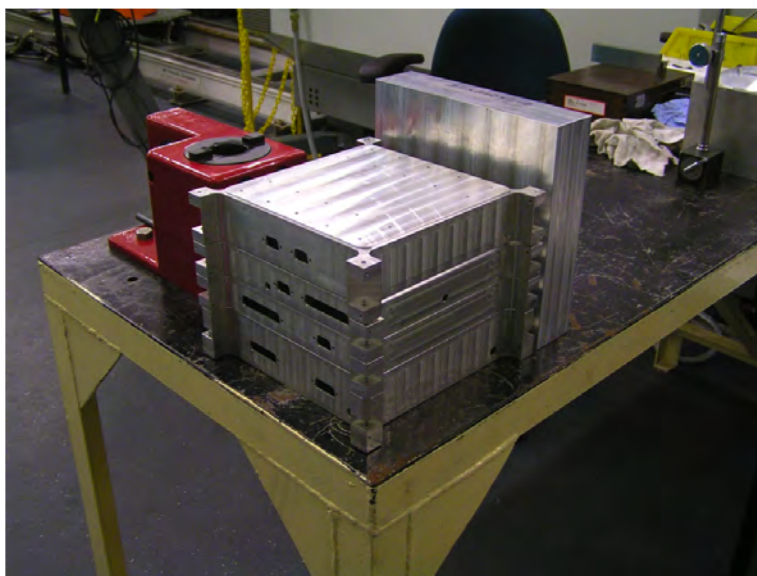


Fig. 3: Vibration testing was successfully completed.

Despite the success experienced in this iteration of the project, mis-management of timelines and human resources caused the project deadline to be pushed further and further back. This period saw many key designers enter and graduate UNSW, causing many ramp-up and ramp-down periods on the development of key subsystems. This eventually lead to the project's complete stall in 2008.

Iteration Three, 2009-2013

The majority of the BLUEsat technical leaders from second iteration had graduated and left UNSW between 2006 and 2008. The project subsequently ramped down and stalled

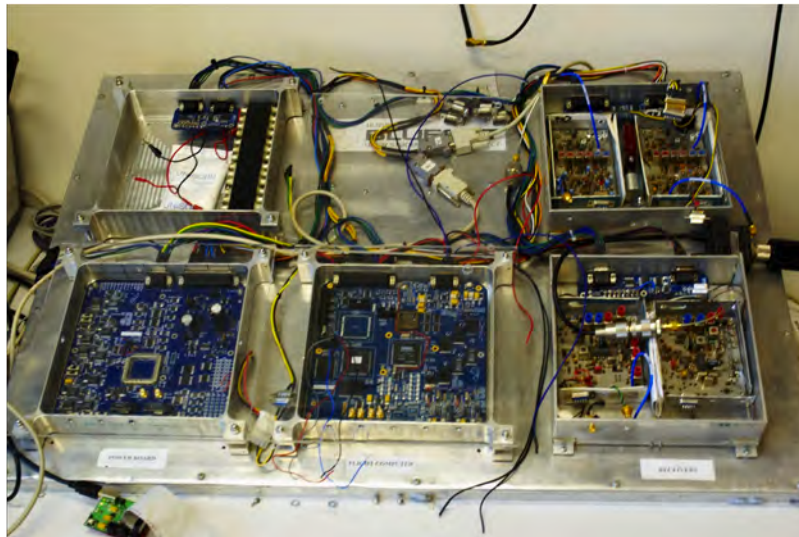


Fig. 4: The FlatSat satellite trays.

in 2008. This prompted an introspective period between 2008 and 2009 addressing key project management issues that had caused the protracted timeline and then stall. A new project management strategy was implemented (discussed in later in **Organisation and Management**) as a result of this review, ramping up the third iteration of the project over 2009.

In 2010, BLUEsat gained momentum with the inception of the Warrawal Project - a space educational program arising from the government-funded Australian Space Research Program (ASRP). The project provided the Group with funding and a deadline to complete a working model of the satellite.

Current Status

Satellite Functional Demonstration

Following successful support from the ASRP and Warrawal Consortium, BLUEsat was able to functionally complete the satellite in 2013, bringing development to a successful close. The following functions were successfully demonstrated at the July 2013 meeting of the Warrawal Consortium:

- Long-range communications
- In system health monitoring, including power consumption and temperature
- Two-way telecommand and telemetry
- System bus management

Figure 5 shows the BLUEsat team at the completion of final construction and testing before the presentation in 2013.

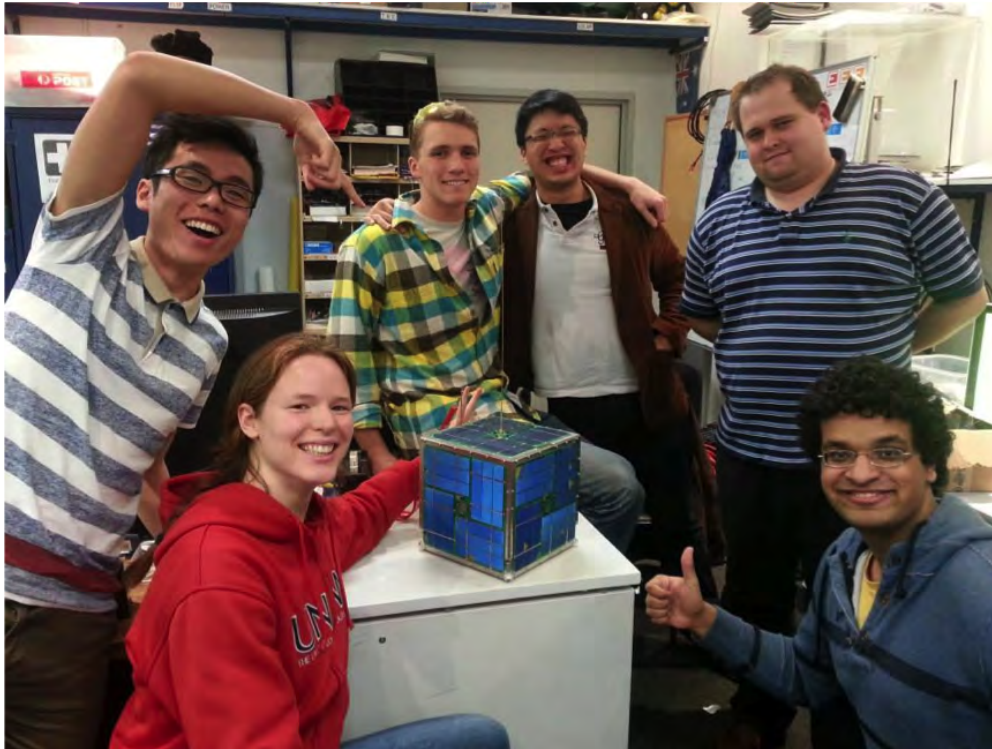


Fig. 5: The BLUEsat team at the completion of the satellite, just prior to Warrawal 2013. Pictured from left to right - Chun-kan Leung, Anne Gwynne-Robson, John Aiden Rohde, Thien Nguyen, Daniel Jedrychowski and Varun Nayyar

Test Launch and Flight Model Construction

The satellite flight model will be tested on a stratospheric balloon in late 2014. The main objective of the balloon launch is to provide a conclusive demonstration of the satellite's functionality in near-space conditions. Environmental performance of the satellite will be measured, especially thermal behaviour during the day/night cycle. An example of a similar stratospheric balloon test conducted for Envisat by CNES [8] is shown in Figure 6.

The satellite is to be suspended beneath a Raven Aerostar zero-pressure helium balloon, launched near Wagga Wagga and lifted to an altitude of roughly 35 km. It will remain at this altitude for approximately 24 hours, experiencing temperature and pressure conditions comparable to low Earth orbit. During this flight, the satellite will be in constant communication with the ground crew, providing telemetry on the satellite's function and position, as well as conducting communication experiments with amateur radio clubs throughout south-eastern Australia.

In order to ensure reliability during this test, the satellite is currently being rebuilt to the final 2013 design. The on-board electrical system will be re-assembled. The majority of the mechanical structure will stay in service, and be used for the flight test. Certain structural components will be re-manufactured for improved structural reliability, including fasteners and radio shielding.



Fig. 6: CNES stratospheric balloon launch in northern Sweden. Image from [8]

Issues and Lessons Learnt

The volunteer nature of the BLUEsats workforce introduced a number of project management and personnel organisational challenges. In particular, dealing with transiency of the membership in the project, having to engage and motivate a volunteer workforce, all whilst tracking and managing a large scale engineering project.

The project's continually delayed timeline came as a result of key project and time management failures. The scope of building a complete set of customised micro-satellite hardware was well beyond what could be achieved within one generation of BLUEsat membership. The project lapses in 2000, 2004 then 2008 coincide almost exactly with the duration of an undergraduate engineering degree at UNSW. These 'lapse' periods occurred due to a critical number of students moving on from UNSW and BLUEsat after having graduated. A lack of on-going training and recruitment created a vacuum of experience.

This occurred due to three reasons -

- Poor project scoping. The project itself should have been scoped such that it could be completed well within one student's tenure at UNSW.
- Lack of continual recruitment. The project's structure was originally run like a typical professional engineering environment. This structure could not cope with the high 'turnover' of human resources inherent in an unpaid, volunteer and largely under-skilled workforce. No mechanism was in place to deal with the constant 'ramp up' cycle caused by having to recruit volunteers to work on high-skill engineering projects in their spare time.

- Poor documentation. Lack of proper documentation of the design in the project meant that knowledge in the project was inherently tied to its senior members. When those members graduated, new members could not rely on documentation in order to carry the project forward.

These issues were identified in the 2008-2009 introspection period. A new project structure was put in place in order to push BLUEsat to its successful completion, and a new vision with more achievable goals was put in place for BLUEsat's future.

Organisation and Management

The executive structure expanded upon in 2005 and team management system introduced in 2009 were created to specifically address these issues. Furthermore, the adoption of a Scrum-like Agile project management philosophy aided in project tracking and deadline management [9].

Personnel Organisation

Members within BLUEsat were managed in teams with a structure similar to that implemented in Scrum projects under the Agile Framework [9]. Rather than focus on the immediate requirements of the satellite, development teams are formed based on availability of human resources. All teams are overseen by the executive panel in the project.

The executive panel consisted of the President, Chief Technical Officer (CTO) and Chief Operations Officer (COO). The COO was responsible for managing project funds, the efficient operation of the group, and organising out-reach events. The CTO focused on project management and provided technical design oversight within the project. The President maintained the mission of the group as a whole, overseeing decisions made by the CTO and COO to ensure that the project was always heading in the right direction.

Teams within BLUEsat are formed based upon member availability and common interest rather than purely on available projects. Projects and parts of the satellite that were not being worked on were assigned to a backlog (discussed below). The latest cycle of development (2009-2013) demonstrated that this mode of team management is very suitable for constant recruitment of new members without sacrificing project velocity. New members were integrated into existing and well developed teams, mentored by the more senior team members.

Each team had a product owner and scrum-master. Product owners were in charge of the technical scope and end requirements of the team's project. Scrum-masters maintained the project timeline and ensured smooth running of the team. All team members are encouraged to bring up any issues or ideas that they had with the project. This relationship is illustrated in Figure 7.

As new members entered and old members leave the project, teams would be re-shuffled and redistributed. With this, personnel management remained adaptive and constantly introspective allowing for peak development efficiency for an otherwise unreliable workforce.

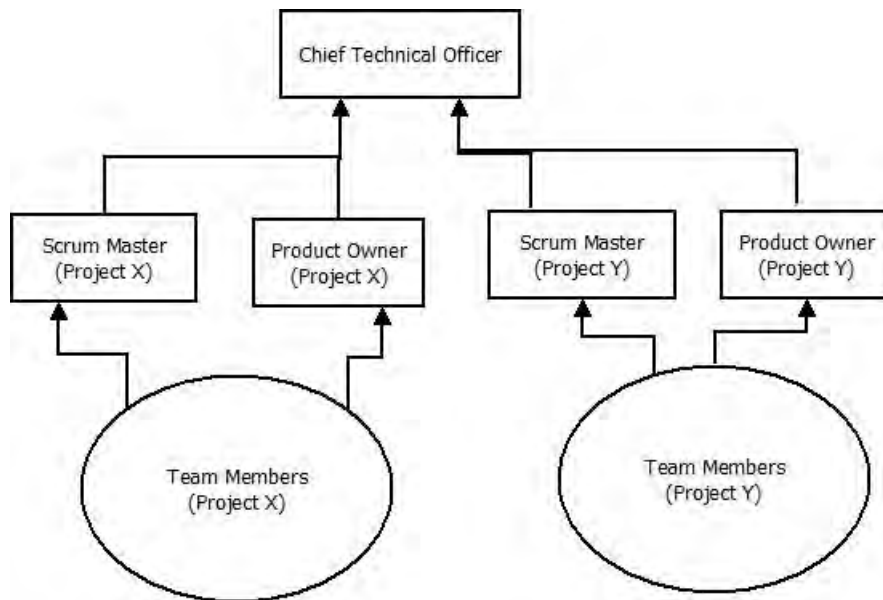


Fig. 7: BLUEsat Team Structure

Generational Handover

A key problem identified in past iterations of the project was the graduation of key senior team members without an appropriate role handover process. Senior members would leave taking away technical knowledge and project management skills that had not been passed on to members who would stay on after them. BLUEsat has implemented two policies in order to combat this problem - on-going documentation and mentoring with ongoing recruitment.

Designs and design practices in BLUEsat are documented and organised into a central repository managed using subversion (SVN)¹. When designs or policies were changed, members are required to ‘check-in’ changes to the central repository. Old versions of all files were maintained for reference and study. This way, all technical and operational knowledge in the project was logged and maintained for all members, present and future, to access.

The senior members of the project were required to pass on knowledge to younger members in a constant mentoring process. As new members joined, the CTO ensured that they were made aware of all aspects of their particular sub-project. With time, this evolved into learning more and more about the grander scope of the project. This process was facilitated by yearly re-election of new executive members. High executive turnover was encouraged as it gave junior members the opportunity to be trained as project seniors before key members had graduated.

Project Management

BLUEsat adopted a Scrum-like system of Agile project breakdown structure. The greater project was broken down into a series of sub-projects based upon sub-system design and

¹SVN is file-version control system that featured robust maintenance of database backups with each ‘check in of a new design’ [10].

logical division of skill sets. This occurred during the initial conceptual phase of the project. The overall design architecture was carried out by senior designers in conjunction with the CTO and then split into sub-projects. The BLUEsat sub-project structure is shown in Figure 8. An initial strategy was then laid out to provide a roadmap to project completion.

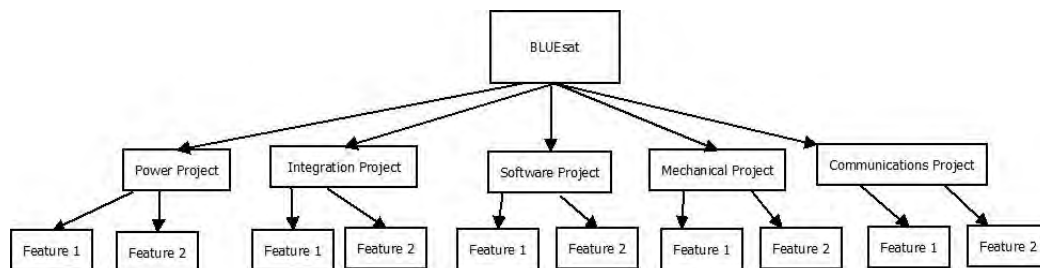


Fig. 8: The BLUEsat sub-project Structure

As teams became available, they were assigned to attack these sub-projects. Regular meetings between scrum-masters and product owners ensured that progress was made. These meetings occurred every two to three weeks with specific deliverables set for each sprint period. Scrum-masters and product owners regularly reported to the CTO, who then oversaw the entire project.

Once certain milestones are met, general meetings were held in order to consolidate progress. These meetings include all members involved with technical development and were aimed to bring everyone up to speed with the current technical progress of the project. The aim of these meetings is to maintain transparency across all levels in the project. This engaged all members at all different levels in the project and had a definitive positive effect on member motivation.

BLUEsat's Future

BLUEsat will carry on the tradition of giving UNSW undergraduate students experience in professional space engineering and project management across all levels as the 'BLUEsat Group' of undergraduate space projects. To address the scope and deadline issues which hampered the progression of the first satellite, the Group will take on multiple, smaller projects under advisement of the Australian Centre for Space Engineering Research (ACSER).

The number of projects undertaken will be dependent on the availability of student resources. Each project will have its own operations and technical lead, with central administration and central vision to be carried out by the Group President. This is illustrated in Figure 9.

Canadian Satellite Design Challenge

The CSDC is a Canadian-run CubeSat design competition. The competition calls for the design and delivery of 3U CubeSat. A 3U CubeSat belongs to the CubeSat family of satellites with a specified an envelope of 300x100x100mm, total mass of 3kg and launch interface

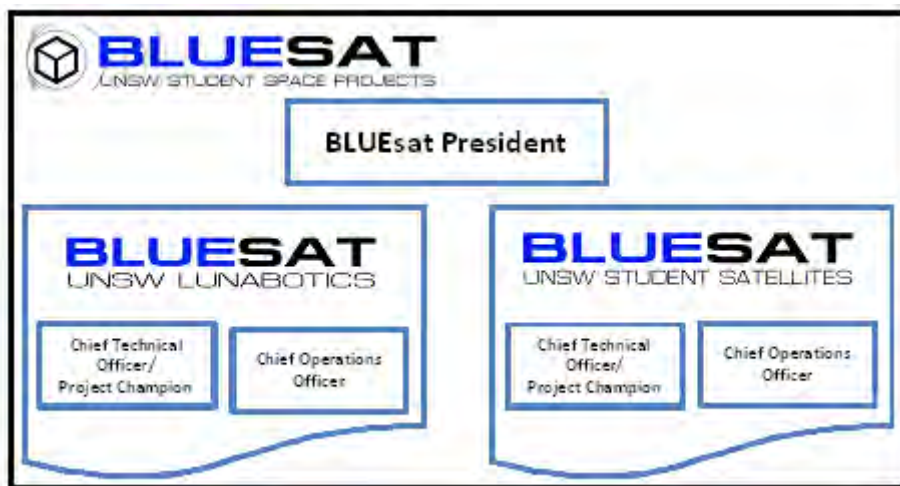


Fig. 9: The new BLUEsat Group Corporate Structure

as specified by California Polytechnic University [11]. The competition allows 21 months between registration and satellite delivery. The CSDC also requires that entrants participate in high-school outreach and science educational programs in order to increase space-technology awareness in the public community. The CSDC administrators have encouraged BLUEsat to partner with a Canadian University to field an entrant in the 2015 competition. Although details about the 2015 competition have not yet been finalised, the satellite is expected to be delivered by quarter 4 of 2015.

NASA Lunabotics Competition

The NASA Lunabotics Mining Competition is an annual competition for university student groups to design and build an un-manned rover for operation on the moon. The BLUEsat Lunabotics team is currently working on an initial conceptual design, project plan and budget. The aim is to have a semi-functional model by end of Semester 1, 2014 and a rover ready for entrance to the May 2015 competition. The different scope and requirements of the Lunabotics Competition allow students in the Group to choose the project that is most engaging to them.

Conclusions

The BLUEsat project has largely been successful in providing students with practical, hands-on space engineering experience. The first satellite allowed student members to develop their understanding of space engineering and create technology which would be suitable for a space-environment. The set-backs that the project experienced over its 15 year lifespan allowed for members to learn valuable lessons in project management and design consolidation. The subsequently developed personnel and project management structures allowed the current team to drive BLUEsat to its recent and very significant success. Positive management decisions contributing to this success included more structured project management, documentation and maintaining ongoing recruitment. BLUEsat will continue in the form of the BLUEsat Group, undertaking space engineering projects and applying lessons learnt from the first 15 years of the BLUEsat Project.

Acknowledgements

Special thanks go to Dr. Elias Aboutanios and Dr. Barnaby Osborne for their invaluable guidance in the final phase of this project. Without them or the support provided by the Warrawal Consortium, BLUEsat would not have finished as successfully as it did. Aiden Rohde, Mitch Wenke and Chun-kan Leung cannot go unmentioned as the executive team who lead BLUEsat across the finish line.

Thanks to Wayne Short, Iver Cairns and all at the NSSA for their continuing support of BLUEsat through the years.

And finally, thank you to all BLUEsat members past and present, without whom the project would not exist.

References

- [1] G. G. Smith, *AO-51 - Operation, Development and Specifications*. AMSAT, 2004.
- [2] P. Croaker, "Design and analysis of bluesat's solar panels," School of Mechanical and Manufacturing Engineering, The University of New South Wales, Anzac Parade, UNSW, 2052, Australia, Tech. Rep., October 2000.
- [3] D. Faber, "Solar arrays for a micro satellite," School of Mechanical and Manufacturing Engineering, The University of New South Wales, Anzac Parade, UNSW, 2052, Australia, Tech. Rep., October 2001.
- [4] H. Endo, "Analysis of bluesat structure," School of Mechanical and Manufacturing Engineering, The University of New South Wales, Anzac Parade, UNSW, 2052, Australia, Tech. Rep., October 2001.
- [5] P. D. Matthison, "Testing and evaluation of the bluesat prototype model," School of Mechanical and Manufacturing Engineering, The University of New South Wales, Anzac Parade, UNSW, 2052, Australia, Tech. Rep., October 2001.
- [6] H. Price. (1995, Jan.) AMSAT-NA Microsats. Accessed 2013-10-07. [Online]. Available: http://www.amsat.org/amsat-new/satellites/sat_summary/microsats/
- [7] K. Cussen, "Testing the bluesat structure for the launch environment," School of Mechanical and Manufacturing Engineering, The University of New South Wales, Anzac Parade, UNSW, 2052, Australia, Tech. Rep., October 2009.
- [8] Collabnet. (2009) subversion. Accessed 2014-01-18. [Online]. Available: <http://subversion.tigris.org/>
- [9] L. Rising and N. S. Janoff, "The scrum software development process for small teams," *Software, IEEE*, vol. 17, no. 4, pp. 26–32, 2000.
- [10] CNES. (2005, May) Envisat and CNES balloons keep track of ozone. Accessed 2013-10-12. [Online]. Available: <http://www.cnes.fr/web/CNES-en/3395-envisat-and-cnes-balloons-keep-track-of-ozone.php>
- [11] S. Lee, A. Hutputanasin, A. Toorian, W. Lan, and R. Munakata, "CubeSat design specification," *The CubeSat Program*, California Polytechnic State University, 2011.

Australian Space Policy: A Student Perspective^{*}

Eren Gorur¹, Ivan Cheung¹, Tu Lesuma¹, Treves Li¹, Cecily Zhu¹, Richard McKenzie²,
Mathew Fenwick³, Adam Ingle³, Timothy Gollan³ and Jacob Hacker³

¹ *University of New South Wales*

² *University of Technology, Sydney*

³ *University of Sydney*

Summary: This paper was developed by a team of carefully selected candidates from a range of fields and levels of study and employment. Established as a student research team within the NSW branch of the Australian Youth Aerospace Association, it consisted of undergraduate and postgraduate engineering, commerce and law students; with a shared interest in the technical, political and legal considerations associated with Australia's involvement in space. The aim of this paper is to summarise and discuss the perspectives of students and young professionals on Australian space activity, and the ensuing policy considerations, in a constructive and meaningful way. Distilling the fragmented views obtained into a response that is both relevant and productive proved challenging, however, the authors feel the paper in its final form adequately conveys the concerns and recommendations of the next generation of space industry leaders in Australia, and allows the conversation on national space involvement to extend to the space industry's inheritors, both in this country and around the world.

Keywords: Australian Youth Aerospace Association, student perspective, space industry policy, economics, national security.

Introduction

The student research team welcomes Australia's Satellite Utilisation Policy (herein referred to as ASUP), released by the Australian Government in April 2013. The team acknowledges the policy document and its significance as a crucial stepping-stone towards Australia's full involvement in space and space related activities, propelling this issue into political discourse and into the minds of Australia's general public.

In March 2008, the Senate Standing Committee on Economics was tasked with producing a report that investigated the current state of Australia's space industry, the arguments for and against the expansion of the industry, and to provide realistic policy solutions. Five years later, the Australian Government released ASUP. The student research team recognises ASUP as an achievement of the 2008 report. ASUP marks a new beginning for Australia's formal discussion on how it should conduct itself on the world stage with regards to its satellite usage. It serves to assist the general public and government in grasping the importance and extent to which space plays a part in everyday life, and how secure Australia's future is in that regard. ASUP also underlines many aspects of Australian industry, and its dependence on satellites and the information that they provide, which invokes the question why such a policy had not already been in place, considering the importance of space to Australian interests.

^{*} Student Research Team (SRT) within the Australian Youth Aerospace Association.

The policy must also be noted for its pragmatic direction of satellite utilisation, rather than an attempt to hurriedly formulate a fully-fledged space policy. This specificity circumvents the perceived financial cost that generally accompanies discussion on Australia conducting manned space flight or launching space vehicles. This provides a good foundation on which to build into the future. This student research team believes that ASUP, this paper, as well as subsequent and intermediary reports and discussions, should be conducted and viewed with the goal of bettering Australia's space industry and global involvement.

In an effort to address some of the criticism ASUP has received since its publication, the team would like to identify that ASUP provides a prime platform from which Australia can build space related policy and that it should be seen as the first iteration of a discussion that will constantly be augmented and refined.

This paper primarily addresses the issues and proposed resolutions to Australian space policy and Australia's space industry as a whole, with regards to technology, policy, legal considerations, defence and economics.

Technology

Background

Australia's historical involvement in and contribution to space utilisation has been noteworthy for a country of its size and level of space development. 1947 saw Australia's Woomera Rocket Range in South Australia established as a launch site for the United Kingdom and European launch activities. This site saw the Weapons Research Establishment Satellite (WRESAT) launch in 1967, only the seventh satellite to be launched into outer space. The Observatory at Parkes, NSW, played an integral role in the Apollo launch and subsequent Moon landings. It served as the main signal receiver and transmitted Neil Armstrong's first steps on the Moon to over 600 million people across the world. Australia has since been a founding member of many space related organisations, including the United Nations Committee on the Peaceful Uses of Outer Space in 1958. This nation has historically played a major role in space and still maintains a significant role in its existing facilities, including the newly announced Square Kilometre Array, a 96-dish survey telescope in Western Australia.

However, since this initial involvement in space activities, there has been a decline in Australia's direct involvement and contribution to the space faring community, as well as an explicit retreat from launch capacity in this country. This is not necessarily negative in the view of the research group. While it would certainly be beneficial to have access to local launch capabilities, this is a costly and delicate process that is better managed by other groups at present. There is no sense in duplicating the functionality provided by international partners, and it is widely accepted that space self-sufficiency is not something Australia can, or should, strive for in any feasible time frame. Rather, Australia should focus on developing capabilities that are readily integrated into global space supply chains, forming a consumer/provider balance. In comparison, if one was to observe the differences in space involvement between Australia with the United States of America, there exists a vast mismatch in terms of both capabilities and scale. While it has been recognised that Australia is an effective consumer of space derived services in a number of categories [1]; it is a small nation with only very specific needs from a space program. On the other hand, Canada has been a very active member in space and related activities. As a country with a very similar

gross domestic product (GDP) as Australia, it is interesting to compare the extent to which space related activities are prioritised. The Canadian Space Agency's budget for 2013-14 is CAD\$488.7 million [2]. The CAD\$3.483 billion in revenue generated by the Canadian space sector in 2011 justifies this [3]. By comparison, Australia's government previously pledged AUD\$40 million over 4 years to the Australian Space Research Program, which has now concluded with no definite continuation of funding. Australia's space sector needs the support of the government to reach a level at which the economy can directly benefit from space related industry, as exemplified by Canada. The UK Space Agency, as a member of the European Space Agency (ESA), has injected £35,000 [4] into promoting space at a grassroots level for 2013/14. Annually, their space industry turnover is £7.5 billion, with employment in the sector rising at 15% per annum [5]. Its main space operations are satellite broadcasting and telecommunications, with more niche operations in satellite manufacturing. In 2011, a £10 billion grant was made available by the government for companies involved in developing new technologies to be used in spacecraft systems.

These comparisons are made with the intention of drawing parallels between Australia and other countries with similar histories and social, economic and cultural backgrounds, and highlight the differences in commitment to space related activities. Rather than build Australia's space capabilities, effectively from scratch, these comparisons provide examples Australia can follow to streamline any progression into space or, at the very least, a space industry.

Relationships

To say that Australia's economic and scientific future relies on space technologies might be received as hyperbole by some. In reality, space and space technologies currently contribute AUD\$1-2 billion annually to the economy [6]. As a nation, Australia has cultivated relationships with a variety of countries and organisations in order to maintain these industries. However, Australia relies on the cooperation of our neighbours and allies to ensure our access to satellite systems, launches and scientific data [6].

While it is unlikely that Australia will be locked out of access in the future [6], this does not mean that the nation should rest on its laurels and assume indefinite access to the systems and data it depends on. For example, there is currently a significant reliance on US, Chinese and Japanese satellites for weather information [7] and US systems for GPS while mapping of the Great Barrier Reef in 1981 was carried out using the US satellite LANDSAT [7]. Without these negotiated concessions, the country's industries would be crippled. As an example, see BBC UK's "The day without satellites" (<http://www.bbc.com/future/story/20130609-the-day-without-satellites>).

Conversely, Australia has made a critical contribution to a number of US missions since the 1960's by providing tracking and data management [7]. Australian scientists and engineers have made valuable contributions to a number of missions such as JAXA's Hayabusa, NASA's HiFire and ESA's EnviSat, among numerous others. According to the UN, Australia has a responsibility for approximately 1/8th of the globe's viewing land and sea mass, from the Java Sea to Antarctica [7].

Thanks to its existing relationships with international agencies, Australia has been involved in a large number of space missions. However, these missions represent only a small fraction of what Australia is capable of achieving. Despite reliance on external providers, whether they

are nations or organisations, Australia's current approach to space policy integration in the civil arena lacks transparency and is thus somewhat hindered (in comparison to the coordinated and structured approaches other states have taken to policy formulation, bringing together industry and other stakeholders to maintain relevant and appropriate policy) [1]. In addition to this, Australia has been offered membership of ESA no less than four times, yet has not utilised this offer to bolster our existing relationships. Whilst membership fees (calculated as a percentage of the country's GDP) are not insignificant, justification needs to be made as to whether such cooperation and membership could be economically worthwhile. Thus, the relationships the nation fosters internationally are critical to the future of national space development. As discussed later, there are a number of ways that this can be sustained.

Skills

It is clear that Australia already possesses highly motivated and skilled individuals with the capacity to initiate a functioning and growth-gearred space industry. The calibre of Australian graduates in the STEM fields is renowned, and some of the nation's brightest minds are already equipped with relevant skillsets. As an example, the highest 'ATAR' (Australian Tertiary Admissions Rank) cut-off without supplementary requirements in any Australian university is Space Mechatronics at 99.80, placing the recipient of this rank in the top 0.15 percentile. However, the outcomes are only too clear when there is a void in the demand for their talents at home – it becomes apparent that Australia is aiding the 'brain drain' phenomenon, where these innovators find more hospitable industries overseas [6]. Contrary to this trend, in comparison with other OECD nations, there is a further decline in the take-up of STEM-related subjects in Australia, especially in secondary education [6], [8].

Essentially, the first step in extending policy will require drawing on home-grown entrepreneurs and innovators to 'space work'. In this process, Australia has the opportunity to develop a generation of young professionals who are eager to stimulate projects in a fertile space industry. However, this all rests upon the necessity of developing industry as submitted by the Space Industry Innovation Council. Subsequently, it is anticipated that tertiary institutions will respond by developing courses and curricula in line with this projected demand, and a renewed emphasis on STEM subjects in secondary schools. Directly relevant technological advances pioneered on home soil can turn around the nation's perceived disinterest in science. The trickle-down effect to the grassroots level is likely to be self-evident and it is certainly anticipated to pique the interest of young children, who are to become the engineers and scientists of tomorrow [6]. Australia's graduates in relevant fields will then be able to seek employment in space, domestically. The overall result is an environment that taps into the potential of that nation's brightest graduates to better the nation as whole in a field that will continue to directly impact and dictate the lives of the population.

In this renewed space development atmosphere, it is anticipated that with the required changes, a new culture of space start-up companies will begin addressing the narrow space-related technologies Australia currently has the capacity to develop, but it is also expected that universities will seek to establish research and development laboratories specifically addressing space technologies. The benefits of establishing R&D labs have been widely reported [See: Chapter 4 of National Innovation System]. A chain-reaction will likely follow and it is hoped that eventually an Australian space cluster will be erected: a concentration of institutions in a particular location [6], much like Silicon Valley in California. The rest of the world will become aware of Australia's burgeoning position, further attracting foreign talent and investment. The value of Australian companies specialising in niche technologies will

gradually increase as we become an important link in global supply chains. This does not necessarily equate to Australia being a sole provider of certain components or technologies, rather Australia achieving a position of technological expertise that can be called upon by other states for their own indigenous programs as well as Australia directly contributing to collaborative programs that rely on technology, applications and sophistication that Australia possesses.

Another issue with current infrastructure is that the Space Coordination Office (SCO) is currently very small, with its employees primarily having an administrative and diplomatic background. Whilst it is extremely beneficial to have administrators who speak the ‘language’ of parliament and law making, the SCO needs to increase its technical foundations in order to adequately represent the space sector in Australia on a global stage, and so there is a requirement for representatives who can also speak the ‘language’ of space and related sciences. The current structure, whilst providing a fantastic vantage point in a departmental sense, and with great potential, is restricting the access of Australian companies to international projects. Small, carefully nurtured expansion of the office will help confirm Australia’s role in the international space industry and give credence to the work that is being done in Australia. The office also needs to move from a model that is not only administrative but also includes a strong technical foundation and adequate outreach and advocacy for Australian involvement in space so that the commercial, demand and ‘future-proofing’ aspects of the space industry are also catered for. There are also perceived benefits in other organisations, such as the *Commonwealth Scientific and Industrial Research Organisation* (CSIRO) and other government bodies, taking on an advocacy role, not only the science, technology and regulatory roles they currently promote. This would be similar to the role that organisations such as the Federal Aviation Administration (FAA), the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) take. This will help work companies within in the Australia space sector into international supply chains whilst providing a justifiable cause to the Australian public. The expansion of the Space Coordination Office would be further aided by the presence of the relevant staff at international space events, such as the International Astronautical Congress as ambassadors of the industry. Delegates have recently seen this in Beijing; however a greater participation in heads of agency plenaries and bilateral meetings would be highly beneficial for Australia as other agencies will better understand Australia’s needs in space and have access to a uniform front through which Australian industry can better communicate.

Funding

The recent Australian Space Research Program was established to provide grants to bodies that developed technologies related to space activities. Over the last four years with a commitment of \$40 million, it was able to fund 14 projects, but its term has finished. Through this, a number of highly successful and hugely promising programs were started, however, with no continuation of funding, the Australian space industry finds itself divided and internally fragmented, each segment fighting for further funding from a dedicated pool that is not sufficient to fund. To sustain the fertile environment of innovation, it is the opinion of the authors that continuing funding is provided as a matter of urgency [9]. The initial AUD\$40 million served a purpose in kick-starting a number of programs into action, and the majority proved to be successful and with immense potential. Referring back to the original comparisons, AUD\$40 million is a relatively small figure for space research funding, when compared to countries like the UK and Canada. However, now the Australian Government should assess the relative success of each of the ASRP funded projects and select a “shortlist” of programs that directly relate to the needs and developing capabilities Australia wishes to

foster and then allocate new funding to these programs. In essence, there is no reason not to continue this source of funding in years to come. The National Committee for Space Science recommends an allocation of AUD\$140 million [7]. Indeed, long-term funding has proven to be integral in Canada's successful approach to space technologies [6], but the policy and justification of Australia's involvement in space needs to be finalised and confirmed so that funding can survive beyond the standard election cycle. The availability of the funding thus becomes a clear signal of Australia's dedication to innovation and its recognition of the important function it plays in the national society; this is the right message Australia should seek to broadcast to the world if it wishes to become a viable partner in space. The emphasis here is not to develop technologies which already exist and are commonplace overseas, nor to pioneer new fields of space research, requiring prohibitively high funding and resources, but to specialise and advance in niche technologies which are highly demanded on a global scale; as an example, Silanna Semiconductors, an Australian-based company that finds its products embedded in satellites, probes, and even Mars Rovers.

With a hub of institutions and businesses all working toward the development of a variety of space technologies, the potential for success is high. A key reason for promoting such a hub includes the development of professional collaboration in the enhancement and progress of space technologies, particularly when Australia is seen as an international leader in global navigation and been described as having an “underlying capacity for innovation” [1]. Developing such a hub could push the country towards global leadership in broader areas of space development. The Decadal plan 2010-2019 calls for AUD\$140 million investment in space related technologies. A similar level of investment over the coming decade in a space technology hub would likely return massive benefits to Australia [1]. Another crucial idea is that space-related technologies should not be confined to one field – instead it should be synonymous with innovation. There is enough evidence that accounts for the sundry beneficial by-products of space technologies [6], and of its contribution to other fields such as medicine. Competition for funding grants would also be a further incentive for space technologies to proliferate domestically. By increasing support and funding to companies, we would have the means to push technological boundaries from our own shores, because at the end of the day, an investment in space technology is an investment in Australia's future as a nation of innovators.

Legal

From a legal standpoint, Australia is a signatory to the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (Outer Space Treaty [OST], 1967) and must therefore meet certain international obligations under the treaty. This can impact domestic space legislation in Australia and articles within these agreements of particular concern to Australia are noted in the Appendix [A]. These international treaty clauses must be adhered to in order for Australia to build a legitimate space agency and/or be actively involved in space activity. The research group does not see these obligations as obstacles toward an Australian space effort, as these are principles that all current space capable signatory countries are accountable to.

With the rapid growth of the commercial space sector, there lies the potential for current treaties concerning space to be amended in order to clearly define parties as either a state or corporate actor and to address the ensuing liability considerations. It should be expected that any signatory to amended treaties or resolutions will be required to have a clearly defined

‘space policy’ or relevant legislation that details the relationship and sharing of liability between commercial operators and states. This would be of particular importance to Australia, as a country that appears to sign up to treaties relating to concepts that have no direct bearing on its space operations. As a nation, Australia is best served by becoming more informed on the global concerns in space industry activity and the legal obligations that follow. However, the question remains as to whether Australia should forge ahead and develop lasting and meaningful space policy, or to wait until it is required under a treaty to which it is a signatory.

Security

Defence force activity accounts for 85.1% of the total value of the Australian aerospace sector [10]. In addition, over 50% of total satellite communications systems have been developed to support the military [11]. Accordingly, any analysis of Australian space policy would be deficient if it did not consider the impacts on national security interests.

Scope and Definitions

The mission of the defence forces is to defend Australia and its national interests. Whilst it is arguable that this mandate extends beyond national security; that security is paramount to their role is axiomatic and this discussion confines itself to national security concerns. This section will look at the national security strategy; defence objectives, plans and capabilities anchored in considerations of space and specifically satellite strategies. It will make reference to Australia’s Satellite Utilisation Policy (ASUP), the Defence White Paper (‘White Paper’), the Defence Capability Plan and end with some criticism and student perspective.

Defence White Paper

The first mention of satellites within the White Paper is in reference to space situational awareness (s6.19, 8.36 White Paper) [13]. Space situational awareness refers to the monitoring and tracking of space based objects (such as debris, or satellites) [14]. It is suggested that this is an area of importance due to Australia’s growing reliance on satellite technology as well as the proliferation of counter-space activities and increasing space debris (White Paper s8.36) [13]. It is the opinion of the authors that the importance of space situational awareness should not be understated. Estimates place the economic impact from unplanned denial of service of earth observation from space capabilities (EOS) at an economic cost of AUD\$100 million in the year the denial occurs [19].

An examination of the White Paper reveals that the primary use the Defence Force intends to make is with regard to communications, although the language suggests it may perform an auxiliary function. Specifically, it states that it provides an independent source of connectivity to deployed forces (s8.35) [13] integrated into planned AUD\$1.1 billion upgrades to fixed telecommunications systems (s8.38) [13]. To assess the wisdom of this relegation it is necessary to look at the infrastructure the Defence Forces intend to use. The ADF has entered into a communication partnership with the United States granting it access to the Wideband Global Satellite Communications (WGS). This system enables flexible, high-capacity communications across both “X” and “Ka” frequencies [15]. This system is purported to offer better interoperability and operational flexibility. The chief drawback to this arrangement is the cost. As part of the deal, Australia had to fully finance the launch of one of the six satellites at a cost of AUD\$900 million [16]. The benefit being that Australia is given access

to bandwidth covering the entire constellation of 6 satellites. The drawback is that Australia receives no proprietary interest in the satellite launched and is granted access to the technology only until 2029 when its contract expires [17]. The reliance on alliances and good will of foreign nations continues to be a burden on Australian space endeavours. It is simply not feasible to construct the military communications network around technology that Australia has no proprietary interest in and cannot control, the ramifications of a denial of service are simply too great if conflict were to break out. This is unfortunate, as the WGS network is reported to be incredibly powerful, capable of transferring information around the world at more than 3 GB/s. It is the reluctant assessment of the authors that satellite communications, whilst an integral component of military communications, can never be wholly or even principally relied upon. As such, Australia should aim to achieve better, more secure, defence communications arrangements that lead to possession, control or even significant influence over space assets and their derived resources and this could be facilitated by Australia adopting a more forthright approach to space utilisation as a whole.

Defence Capability Plan

When considering defence undertakings and capabilities, according to the Defence Capability Plan 2012 [18], aerospace spending has been focussed on sustenance with acquisition expenditure falling by 23.8% in 2011-2012. While acquisition expenditures are expected to increase over 2014-2015, the bulk of these expenditures are unrelated to satellite or space investment.

Looking more closely at the plan, only one project (JP2008) specifically mentions satellites (although it is acknowledged that other projects focusing on geospatial intelligence would make use of them). JP2008 has an Acquisition Category rating of III, which indicates that it is an expenditure of moderate strategic significance to the ADF. This is an unfortunate but ultimately sensible approach, it stems from an under-utilisation of the technology. As previously stated, satellites cannot form the primary technology relied upon for our national security if there remains no proprietary interest in the technology. This is especially so when the use of the technology is for a definite period of time.

Economics

Composition of the Australian Space Industry

The Australian space industry is complex, primarily motivated by telecommunications and satellite services. Globally, the industry is heavily influenced by satellites, particularly telecommunications, accounting for 62% of space industry revenues [20]. A majority of the remaining space industry revenue comes from various government grants and projects. Australia's involvement in the space and satellite industry is largely focused on satellite service providers, including telecommunications and positioning. Other areas include ground support services, minor equipment manufacturing and research and development.

Asia Pacific Aerospace Consultants (APAC) conducted a review of Australian space activities for the former Space Policy Unit (SPU) in two consecutive years and estimated revenue to be in the order of AUD\$1 billion to AUD\$2.2 billion per annum in the year 2011 [21]. In their survey, in relation to respondent's outlook on revenues, exports, imports and employment, roughly 30% to 40% of respondents expected no change, 30% to 40% indicated up to 25%

industry growth and 10% to 15% expected small declines over a 6 year period consisting of the previous 3 and following 3 years [22]. It was estimated that the Australian space industry workforce contained approximately 8400 persons with 6000 full-time equivalent (FTE) staff. The 2011 [21] survey revealed approximately 45% of respondents reported problems with skills shortages in recruiting staff for space-related activities and approximately 62% of those encountering shortages recruited from overseas to meet these needs. Compared to the 2010 survey, approximately 42% of respondents reported problems with skills shortages and around 83% of those encountering shortages recruited from overseas [22].

APAC also identified 631 organisations involved with space-related activities in Australia in 2011 [21], up from 456 identified in 2010 [22]. The majority of these companies are involved in activities in the “Space Enabled Services & Applications” category, including telecommunications, broadcasting, positioning, navigation, earth observation, and satellite imagery. This is also consistent with the space industry composition in international trends. With an increasing trend in the growth of companies associating with the space industry in Australia, the authors of this paper believe a significant portion of the growth in these numbers are of companies belonging to the emerging small to medium enterprises (SMEs) segment, with industry growth trends in traditional segments such as telecommunication services, wireless telecommunication services and surveying and mapping services not showing signs of growth. Analysis of telecommunication services, wireless telecommunication carriers, and surveying and mapping services are contained in the Appendix [B].

The Economic Case for Policy Engagement

Although Australia’s current yearly revenue from space-related activities amounts to AUD \$1-2.2 billion, globally the economic footprint of the space industry is within the range of \$255-310 billion. Accordingly, any policy-driven expansion of Australia’s space industry must be directed at increasing Australia’s contribution to international activities whilst simultaneously developing capabilities of our own.

Further consideration must be given to:

- Assessment of economic risks of Australia’s dependence on foreign-owned and operated satellites, and
- Economic, social and environmental needs that are not being met by Australia’s existing space resources.

The global health of the space economy is sustained by demand for established space-related technologies in fields such as communications, navigation, earth observation and defence. However there is promising growth, particularly in the private sector, in areas such as natural resource management, climate change monitoring, space tourism, robotic satellite servicing and the clearing of space debris as well as the privatisation of space transportation. Furthermore, the heavy investment in space-related activities by emerging economies, particularly India and China, will continue to fuel growth in this sector.

Australia can benefit directly from participation in international space activities by utilising its engineering expertise and scientific sophistication domestically to develop high technology, nuanced components for space applications. Currently, much of the professional workforce is employed overseas due to industry demands. Australia’s stability, business environment and

research strengths make this a realistic and attractive option that has previously been identified as a focus area.

Further consideration must be given to:

- The possibility of encouraging private sector interest in the development of a potential future launch industry in Australia, and
- Economic viability of a national satellite development and deployment program (with critical consideration for cases such as FedSat).

With the aim of growing an industry that will supply high-tech components to international partners, key measures that should be undertaken include:

- Strengthening and establishing international partnerships with South East Asian nations, particularly India/China. This can be achieved through bilateral treaties, industry lobby groups and trade partnerships reducing the cost of business for high-tech industries.
- Adopting less restrictive legislation with the aim of fostering the commercialisation of launch capabilities and other space-related activities, achieved through:
 - Amending the licensing regime, launch safety requirements and registration costs enumerated in the *Space Activities Act 1998 (Commonwealth)*.
 - Adopting certain measures contained within the United States Commercial Space Act 1998, particularly those which promote a stable but minimal regulatory oversight and foster public-private partnership in the development of space-related technologies.
 - Encouraging competition and foreign investment in satellite communications by reviewing existing telecommunications legislation.
 - State mitigation of risk for those seeking to establish space-related companies.
 - Ensuring the protection of innovators and entrepreneurs by protecting intellectual property rights with regards to space-related technologies but encouraging their responsible transfer, including state subsidies for technology transfer.

Vision 2025

Considering the analysis up until this point has been focussed on the near term, the authors considered it important to present an outline on how the youth of today would like to see the Australian space industry evolve up until the year 2025 and beyond. The following concepts require the full enactment and consideration of the points detailed above in order to be remotely realisable, however, even more important is the requirement for a shift in the mindset of government and policy makers and their attitudes towards space and Australia's role in it. These suggestions, whilst challenging and not currently realistically achievable, serve to provide an ultimate goal that might exist as the ultimate milestone to consider Australia as a part of the ever growing cadre of space faring nations.

Guaranteed annual funding for space research and development

No body, government or otherwise, can operate without funding. Whilst in the past there have been funding grants and ongoing contributions to the Australian space industry, there has not

been anything supplied with a long-term guarantee. The authors feel that the general consensus is to see a guaranteed amount, gradually increasing with each subsequent budget cycle until space technologies receive an ongoing confirmed endowment commensurate with their national importance. The authors argue that Australia's involvement in space and space technologies is critical to our national development and economic performance. With this in mind, the authors expect a budget in the order of hundreds of millions per annum would be an appropriate amount as an initial constraint. This could be allocated between an Australian 'space agency' and commercial entities in the form of grants and tenders.

Australian astronauts as part of an international mission

While there have been Australians in space previously, all of them have had to become US citizens to achieve this significant feat. We would like to see Australian scientists and engineers given the opportunity to work directly in space as representatives of Australia. We envision a future Australian representative body on space affairs partnering with international agencies to place appropriate Australians in prospective astronaut programs. This fits under the umbrella of an Australian 'space agency' as well as within the current educational export programs. Potentially, Australia could offer to provide some components of the astronaut training program. In addition to this, a formalisation of exchange programs between countries and their respective space agencies could see a healthy exchange of ideas, improve collaboration and normalise operational and educational outcomes in tertiary and professional training across partnering nations.

Membership of ESA or other major international consortium

As has been repeatedly mentioned, we have turned down the possibility of membership with the European Space Agency on several occasions. Achievement of this goal would also facilitate the fulfillment of the other 2025 goals and could accelerate Australia's progress to space development and contribution. This would provide our students with more secure post-study career advancement and employment, a potential market for our indigenously developed technologies and access to a multitude of researchers and institutions to collaborate with. There is no obvious downside to closer ties with a large space related organization, save for the significant price-tag membership attracts. Membership would only truly be feasible once space becomes clearly recognised as a national priority.

Summary and Conclusion

This document aims to highlight the extent to which Australia is already involved in space on a multitude of platforms. However, what should be evident to the reader is that this involvement, at least in the civil domain, is neither coordinated nor reflective of the true capacities and needs of Australia. The overall conclusions of this paper aim to encapsulate both the progress that can be achieved in the near future as well as in the longer term. Given there have been several previous offers of membership to the ESA it would be remiss of the Australian Government not to pursue this further. Not only would this give Australian graduates access to fulfilling, and otherwise domestically unavailable, employment, it would also foster further cooperation between Australian and European industries. The potential rewards of such an initiative cannot be ignored. However, in order to achieve this there is a need for an active, representative body to advocate the needs of the Australian space sector and the Australian government in space at a domestic and international level. The closest

current possibility option is the Space Coordination Office. At present this is a relatively small, non-technical office, working within the Department of Innovation. We propose a progressive upgrade of this department starting with the introduction of technical staff and assured budget. It may also be of value to alter the name to something more publically accessible such as incorporation into the title of a federal portfolio alongside climate, weather, disaster management, environment etc., or in its current position with significant rebranding and awareness as the Australian Office of Space Activities, Australian Federal Space Directorate or another appropriate name that advises the public and interested parties as to the actual capabilities and range of authority of the office. This would also enhance the international view of Australian space capabilities by providing a known and uniform point of contact. There is general support for this endeavour from within government and industry; Lost in Space [6] highlights that an agency should evolve gradually so as to ensure that its goals are met effectively. The eventual consolidation of the various departments involved in space activities (earth observation, metrology, etc.) should be a goal of this new agency.

In conclusion, if one thing is to be gained from this article it is that there is a significantly sized, concerned body of students and young professionals frustrated with the stilted progress Australia has made in space in recent years, but equally open-minded and optimistic about the potential for the future. The reluctance for any concrete and spurring commitment to space development in Australia is endemic in the public discussion that occurs across the board and needs to change. The public and the government needs to be informed of the risks Australia could face into the future if the status quo is maintained and realistic progress needs to happen very soon. As a final remark, should nothing change, Australia risks creating a generation of disillusioned yet highly skilled graduates who will increasingly search for employment overseas. Whilst this may be a partial goal of the current Australia's Satellite Utilisation Policy, the increased contribution to a 'brain drain' may have dire consequences for the future of science and technology within Australia, leaving a gap in knowledge and thus a gap in motivation and capability, propagating a vicious cycle of over-reliance on foreign-sourced good will and cooperation. Australia is a country with a history of dedicated innovators and visionaries, and it is the youth of today who are mobilising to recapture the forward thinking and imaginative spirit that Australia once recognisably possessed. In space, from the youth perspective, the only way for Australia is up.

References

- [1] Department of Innovation, Science & Research 2012, *Assessing Australia's Use of Space Products and Services: A Comparative Benchmarking Analysis*, prepared by Futron Corporation, Bethesda, MA.
- [2] The Canadian Space Agency 2012, Report on Plans and Priorities, *2013-14 Estimates*, p.2, viewed 7th September 2013, http://publications.gc.ca/collections/collection_2013/asc-csa/ST96-7-2013-eng.pdf
- [3] The Canadian Space Agency 2012, State of the Canadian Space Sector, *Policy and External Relations*, p.9, viewed 7th September 2013, <http://www.asc-csa.gc.ca/pdf/eng/industry/state-2011.pdf>
- [4] The UK Space Agency 2013, UK Space Agency, Annual Reports and Accounts, *2012-13*, viewed 14th September 2013, <http://www.official-documents.gov.uk/document/hc1314/hc04/0460/0460.pdf>

- [5] The UK Space Agency 2010, *The Size and Health of the UK Space Industry, A Report for the UK Space Agency*, viewed 14th September 2013, <http://www.oxfordeconomics.com/publication/open/222601>
- [6] Senate Standing Committee on Economics 2008, *Lost in Space? Setting a new direction for Australia's space science and industry sector*, Senate Printing Unit, Parliament House, Canberra.
- [7] National Committee for Space Science 2010, *2010-2019 Decadal Plan for Australian Space Science: Building a National Presence in Space*, Australian Academy of Science, Canberra.
- [8] Biddington, B & Sach, R 2010, *Australia's Place in Space: Toward a National Space Policy*, The Kokoda Foundation, Kingston, ACT.
- [9] Space Industry Innovation Council 2012, *Report 2012*, Department of Industry, Innovation, Science, Research and Tertiary Education, Canberra.
- [10] MarketLine 2012, *Aerospace & Defence Industry Profile: Australia*, viewed 10th October 2013, <http://web.ebscohost.com.ezproxy2.library.usyd.edu.au/bsi/pdfviewer/pdfviewer?sid=5a2cf9a6-4f6d-4af9-a049-394fe621620c%40sessionmgr113&vid=2&hid=123>.
- [11] Abeyratne, R.I.R 2011, *Space security law*, New York: Springer.
- [12] Department of Prime Minister and Cabinet. (2013). *Strong and Secure: A Strategy for Australia's National Security*, viewed 13th October 2013, http://www.dpmc.gov.au/national_security/national-security-strategy.cfm
- [13] Australian Government. 2013, *Defence White Paper*, viewed 10th October 2013, <http://www.defence.gov.au/whitepaper2013/>
- [14] AUSMIN. 2010, *AUSTRALIA-UNITED STATES SPACE SITUATIONAL AWARENESS PARTNERSHIP FACT SHEET*, viewed 13th October, <http://www.dfat.gov.au/geo/us/ausmin/Space-Situational-Awareness-Partnership-fact-sheet.pdf>
- [15] Air Force Space Command 2012, *WIDEBAND GLOBAL SATCOM SATELLITE FACTSHEET*, viewed 13th October 2013, <http://www.afspc.af.mil/library/factsheets/factsheet.asp?id=5582>
- [16] Wroe, D. 2013, *Military ties benefit from sixth satellite*, viewed 12th October 2013, <http://www.smh.com.au/technology/sci-tech/military-ties-benefit-from-sixth-satellite-20130808-2rkp8.html>
- [17] Klotz, I. (2013). *U.S. military satellite, paid for by Australia, launched into orbit*, viewed 12th October 2013, <http://www.reuters.com/article/2013/08/08/us-space-satellite-idUSBRE97700T20130808>
- [18] Department of Defence 2012, *Defence Capability Plan*, viewed 15th October 2013 <http://www.defence.gov.au/publications/capabilityplan2012.pdf>
- [19] ACIL Tasman 2010, *The economic value of earth observation from space: a review of the value to Australia of Earth observation from space*, ACIL Tasman, Melbourne.
- [20] Satellite Industry Association June 2013, *State of the Satellite Industry Report*, viewed 8th August 2013, http://www.sia.org/wp-content/uploads/2013/06/2013_SSIR_Final.pdf
- [21] Asia Pacific Aerospace Consultants 2011, *A Review of Current Australian Space Activities May 2011*, Asia Pacific Aerospace Consultants.
- [22] Asia Pacific Aerospace Consultants 2010, *A Review of Current Australian Space Activities April 2010*, Asia Pacific Aerospace Consultants.
- [23] Department of Innovation, Science & Research 2013, *Australia's Satellite Utilisation Policy*, DISR, Canberra.
- [24] Dempster, A 2013, 'Ten reasons why Australia urgently needs a space agency', *The Conversation*, 5 August, viewed 6 September 2013, <http://theconversation.com>.

- [25] C. Shulman, *IBISWorld Industry Report J5802. Wireless Telecommunications Carriers in Australia*, IBISWorld, 2013.
- [26] C. Shulman, *IBISWorld Industry Report J5800. Telecommunications Services in Australia*, IBISWorld, 2013.
- [27] A. Kelly, *IBISWorld Industry Report M6922. Surveying and Mapping Services in Australia*, IBISWorld, 2013.
- [28] Pasfield, M. 2008, *Safeguarding Australia's Borders*, viewed 17th October 2013, http://www.ga.gov.au/webtemp/image_cache/GA11630.pdf.
- [29] Space Policy Unit, Department of Innovation, Science & Research 2011, *A Review of Current Australian Space Activities*, prepared by Asia Pacific Aerospace Consultants, APAC, St Ives, NSW.
- [30] Symbios Communications 2010, *A Review of the International Space Landscape: Executive Summary*, Vacluse, NSW.
- [31] Department of Innovation, Industry, Science and Research 2008, *Venturous Australia: Building Strength in Innovation*, prepared by Cutler & Company, Cutler & Company, North Melbourne, Vic.

--

Appendix A: UN Treaty/Resolution Articles of Particular Note to Australia

Rescue Agreement 1967

- Art. I - IV: Must notify launching authority and UN and/or render assistance to any personnel of a spacecraft that are in distress or in accident.
- Art. V: Expenses to recover and return a space object is borne by the launching authority.

Liability Convention 1971

- Art. II: Launching state is absolutely liable to pay compensation for damage caused by its space object on the surface of the earth or to aircraft in flight.
- Art. IV - V: Two or more launching states are jointly and severally liable for any damage caused.
- Art. IX and XIV: Claim for compensation for damage shall be presented to a launching state through diplomatic channels. Otherwise, the parties can establish a Claims Commission to determine amount of compensation payable.

Registration Convention 1974

- Art. II and IV: Must register any space object launched into Earth orbit in national registry that it maintains and provide such information to UN for its central register.

Use of Artificial Earth Satellites for International Direct Television Broadcasting, 1982

- Principle 13: State which intends to establish or authorise establishment of an international direct television broadcasting satellite service will notify the proposed receiving state and enter into consultation.
- Principle 14: Service must conform with the relevant ITU instruments.

Remote Sensing of Earth from Outer Space, 1986

- Principle 2: Remote sensing activities shall be carried out for the benefit and in the interests of all countries.
- Principle 12: Sensed state shall have access to primary data and processed data on non-discriminatory basis and on reasonable cost terms.

Use of Nuclear Power Sources in Outer Space, 1992

- Principle 3: Use of nuclear power sources in outer space restricted to space missions that cannot be operated by non-nuclear energy sources in a reasonable way.
- Principle 4: Launching state must conduct a thorough and comprehensive safety assessment before launching a space object with nuclear power sources.
- Principle 5: Must inform states concerned in event of re-entry of radioactive materials to the earth.

--

Appendix B: Economic Data on Space Derived Revenues and Industry

Telecommunications Services

Table 1: Telecommunications Services Key Industry Statistics [25]

Criteria	Outcome
Revenue	\$39.7 billion
Annual Growth 08-13	-1.9%
Annual Growth (projected) 13-18	-1.4%
Profit	\$5.1 billion
Wages	\$5.4 billion
Businesses	182
Employment Growth Overall	Decline

The data presented above accounts for the declining revenue growth in the telecommunications sector. In contrast, there has been increased demand and intense competition creating a fierce competitive environment. This means growth in demand has

Use of Artificial Earth Satellites for International Direct Television Broadcasting, 1982

- Principle 13: State which intends to establish or authorise establishment of an international direct television broadcasting satellite service will notify the proposed receiving state and enter into consultation.
- Principle 14: Service must conform with the relevant ITU instruments.

Remote Sensing of Earth from Outer Space, 1986

- Principle 2: Remote sensing activities shall be carried out for the benefit and in the interests of all countries.
- Principle 12: Sensed state shall have access to primary data and processed data on non-discriminatory basis and on reasonable cost terms.

Use of Nuclear Power Sources in Outer Space, 1992

- Principle 3: Use of nuclear power sources in outer space restricted to space missions that cannot be operated by non-nuclear energy sources in a reasonable way.
- Principle 4: Launching state must conduct a thorough and comprehensive safety assessment before launching a space object with nuclear power sources.
- Principle 5: Must inform states concerned in event of re-entry of radioactive materials to the earth.

--

Appendix B: Economic Data on Space Derived Revenues and Industry

Telecommunications Services

Table 1: Telecommunications Services Key Industry Statistics [25]

Criteria	Outcome
Revenue	\$39.7 billion
Annual Growth 08-13	-1.9%
Annual Growth (projected) 13-18	-1.4%
Profit	\$5.1 billion
Wages	\$5.4 billion
Businesses	182
Employment Growth Overall	Decline

The data presented above accounts for the declining revenue growth in the telecommunications sector. In contrast, there has been increased demand and intense competition creating a fierce competitive environment. This means growth in demand has

including voice and data, have seen growth along with its user base, particularly from oil and mining companies operating in remote areas. Other users of satellite services include government organisations, such as forestry and disaster recovery services, and individuals that cannot get adequate coverage from terrestrially based telecommunications services. Air and sea services are also heavy users of satellite-based telecommunications services, as they are essential for mapping, obtaining weather conditions, emergency distress signals and various voice communication services to land, sea and air points. Satellite offers ubiquitous coverage using point to multi-point communication.

Surveying and Mapping Services

Table 3: Surveying and Mapping Services Key Industry Statistics [27]

Criteria	Outcome
Revenue	\$3.8 billion
Annual Growth 08-13	2.5%
Annual Growth (projected) 13-18	2.1%
Profit	\$562.5 million
Exports	\$92.5 million
Businesses	3350
Employment Growth Overall	Stable

The surveying and mapping services industry has been experiencing uncertainty in relation to fluctuating industry growth. Key industry drivers include demand and investment in residential and commercial property development and construction. The surveying and mapping services industry is highly fragmented with little concentration in key players. This involves low barriers of entry to this industry creating high levels of competition and is expected to further increase.

Technology levels have significant influence on this industry, with computerisation improvements in streamlining office activities and introduction of satellite based technologies. Imaging and positioning satellite technology has contributed to areas of surveying, mapping and positioning providing major impacts. This contribution along with the introduction of other technologies in this area will streamline efficiency and improve the accuracy is required in different fields of the industry.

Wobbling Ancient Binaries – Here Be Planets?

Jonathan Horner ^{1,2}, Robert Wittenmyer ^{1,2}, Tobias Hinse ³, Jonathan Marshall ⁴ and Alex Mustill ⁴

¹ *School of Physics, University of New South Wales, Sydney, NSW 2052, Australia*

² *Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW 2052, Australia*

³ *Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro Yuseong-gu 305-348 Daejeon, Korea*

⁴ *Departamento de Física Teórica, Facultad de Ciencias, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

Summary: In the last few years, a number of planets have been proposed to orbit several post main-sequence binary star systems on the basis of observed variations in the timing of eclipses between the binary components. A common feature of these planet candidates is that the best-fit orbits are often highly eccentric, such that the multiple planet systems proposed regularly feature mutually crossing orbits - a scenario that almost always leads to unstable planetary systems. In this work, we present the results of dynamical studies of all multiple-planet systems proposed to orbit these highly evolved binary stars, finding that most do not stand up to dynamical scrutiny. In one of the potentially stable cases (the NN Serpentis 2-planet system), we consider the evolution of the binary star system, and show that it is highly unlikely that planets could survive from the main sequence to obtain their current orbits - again casting doubt on the proposed planets. We conclude by considering alternative explanations for the observed variation in eclipse timings for these systems.

Keywords: Planetary systems, Numerical methods: N-body simulation, Planetary systems: dynamical evolution and stability, Exoplanets, Circumbinary companions

Introduction

In the last two decades, a plethora of planets have been proposed orbiting a wide variety of stars. Depending on which of the two main online exoplanet databases you follow, the current tally of confirmed exoplanets stands at either 732 (<http://exoplanets.org>; [1]) or 992 (<http://exoplanet.eu>)¹. Of those planets, only a handful (~1%) have been detected directly, through direct imaging (e.g. [2][3]) – the rest have been found through indirect means, such as the radial velocity (e.g. [4][5][6]) and transit (e.g. [7][8][9]) techniques. Although the bulk of the discoveries are made through these indirect means, it has still been possible to learn a great deal about the variety and nature of exoplanets – with recent work revealing the composition of exoplanetary atmospheres (e.g. [10][11]), the variety of exoplanet orbits (e.g. [12][13]) and the complexity of multiple planet systems (e.g. [14][15]).

The majority of those planets orbit single stars in the prime of their lives – main sequence stars (e.g. [16][17]). Fewer have been found around evolved stars – stars who have completed their main sequence life and swollen to become giants (e.g. [18][19]). However, in recent years, a number of planets have been proposed to orbit post-common envelope binary star

¹ Exoplanet numbers taken from <http://exoplanets.org> and <http://exoplanet.eu> on 4th October 2013.

systems² (hereafter PCEBs), on the basis of observed variations in the timing of eclipses between the two components of the binary system (as discussed in detail in e.g. [20]). Such planets, if they do indeed exist, would provide fascinating new insights into the planet formation process around binary stars, as well as the likelihood of planets surviving the post-main sequence evolution of their host stars. As such, it is clearly important to examine whether those planets are truly all that they seem to be. In most cases, PCEBs have been proposed to host two planets, often moving on surprisingly eccentric orbits. As a result, we have carried out a series of dynamical studies investigating the interaction of the proposed planets. Our work serves a dual purpose – first, it allows us to determine whether the candidate planets are dynamically feasible (that is, that their orbits are not catastrophically unstable). Second, our work can provide better constraints on the orbits of the planets in a given system (e.g. [15][21]), thereby allowing better characterisation of multiple planet systems.

In this work, we summarise the results of our dynamical studies of all proposed multiple-planet systems orbiting PCEBs. In the next section, we briefly describe our methodology, before continuing to discuss the various systems studied. We follow that discussion with a brief summary of our subsequent investigations that go beyond pure dynamics, in order to better ascertain the true nature of the signals presented by the observed PCEBs, before drawing our conclusions in the final section.

Dynamically Testing Exoplanet Systems

In order to test the dynamical stability of multiple planet exoplanetary systems, we follow a well-established routine (e.g. [15][22]). We use the Hybrid integrator within the n-body dynamics package MERCURY [23] to run a large suite of simulations following the evolution of the planets within the system for a period of 100 Myr. We hold the initial orbit of the best constrained planet fixed at the best-fit solution, and vary the initial orbit of the second candidate planet across solutions spanning the full $\pm 3\sigma$ uncertainty range in orbital semi-major axis (a), eccentricity (e), argument of periastron (ω) and mean anomaly (M). The planetary systems created in this way are then integrated forward in time for a period of 100 Myr, or until the planets therein collide with one another, are flung into the central star, or are ejected from the planetary system as a result of their mutual interactions. Should the system destabilise in this manner, the time at which this occurs is recorded.

² PCEBs are tightly bound evolved binary systems, which feature a white dwarf primary and a main sequence secondary star. These binaries were once more widely separated, but when the primary star moved off the main sequence, it expanded rapidly, to become a red giant star. At some point during its time on the Red Giant Branch or, slightly later, on the Asymptotic Giant Branch of the Hertzsprung-Russell diagram, the primary swelled to such an extent that it overflowed its Roche lobe, resulting in the two stars becoming surrounded by a common envelope of material. This began a period known as the “common envelope phase”, during which the secondary star spiraled inwards until the loss of mass by the primary was completed, leaving it as a white dwarf or subdwarf B star. Where once the two stars orbited one another with a period of tens or hundreds of days, they now have a mutual orbital period of just a couple of hours. PCEBs frequently evolve to become “Cataclysmic Variable stars”, with significant quantities of material being transferred from the secondary to the primary through an accretion disk, resulting in significant variability and even infrequent nova outbursts from the system. For more information, we direct the interested reader to [46], and references therein.

In this manner, we are able to create detailed maps of the dynamical stability of the planetary system in question, as a function of the initial semi-major axis and eccentricity of the orbit of the planet with the least tightly constrained orbital elements. We can then assess whether the planetary system is dynamically feasible and, if so, determine whether additional constraints can be placed on the orbits of the planets therein on the basis of the dynamics. For more details on our methodology, we direct the interested reader to [24], and references therein.

Four Unstable Systems

Of the six PCEBs for which multiple companions have been proposed³, four feature orbits for the candidate companions that cross one another. Mutually crossing orbits are typically dynamically unstable on short timescales, unless the objects moving on those orbits are protected from close encounters by the influence of mutual mean-motion resonance (as is the case for the Jovian and Neptunian Trojans (e.g. [26][27]) and the Plutinos [28] in our Solar system). The orbital solutions proposed for two of these systems (HW Virginis [29] and QS Virginis [30]) are presented in Figure 1 to illustrate the extreme architectures put forward in the discovery works.

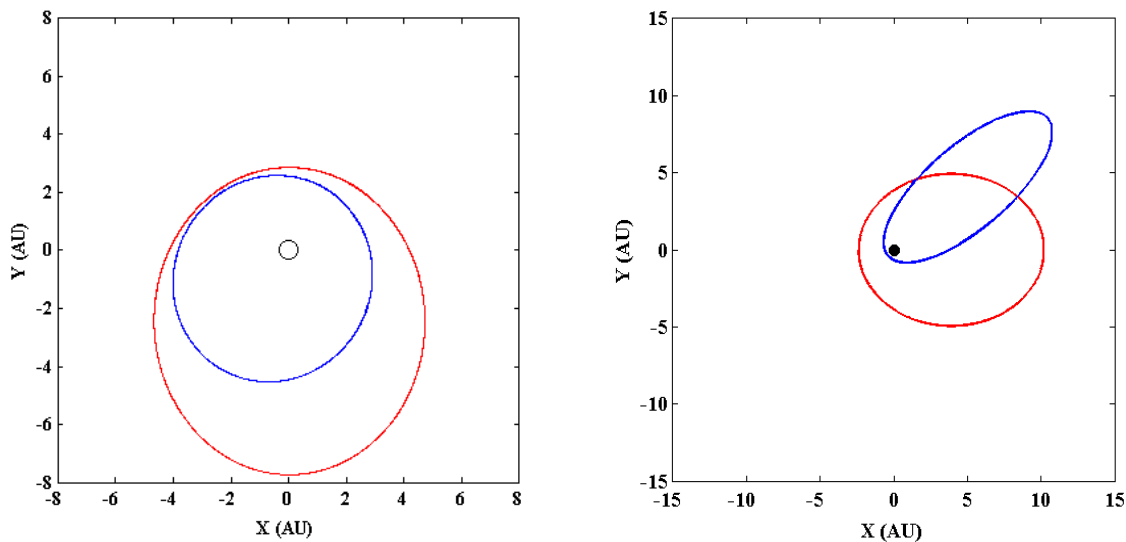


Figure 1: The best-fit orbits proposed for the candidate planets around the PCEBs HW Virginis (left, [29]) and QS Virginis (right, [30]). Both scenarios feature orbits that approach one another closely. In the case of QS Virginis, the orbits have eccentricities reminiscent of the comets in our Solar system. For HW Virginis, although the best-fit orbits do not cross one another at the current epoch, the great majority of architectures permitted within the uncertainties on the orbital solution are mutually crossing.

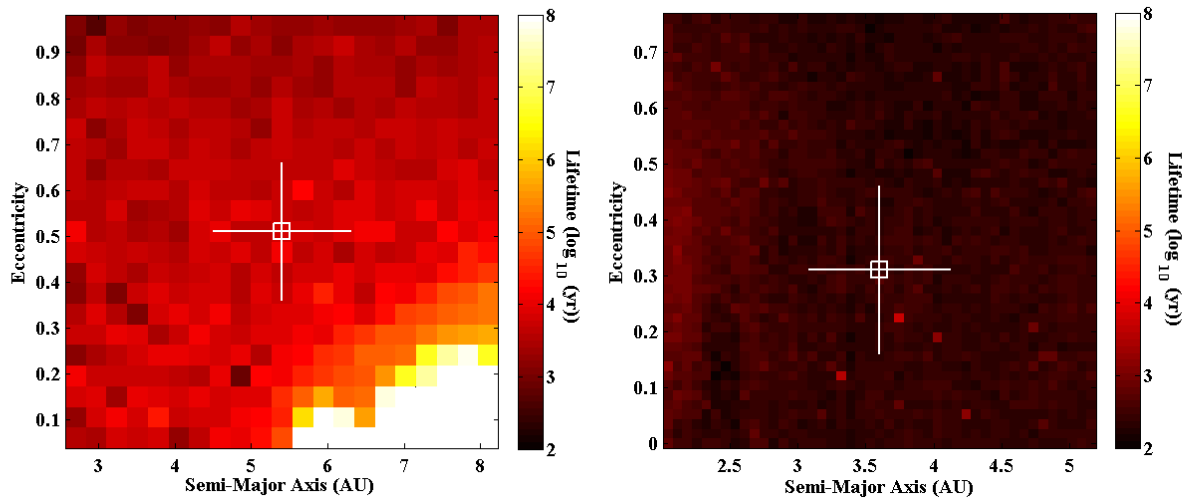
Given the extremely unusual architectures proposed for the candidate planets in these systems, it is clearly important to examine whether the systems could be dynamically feasible. We therefore studied each system in turn, to see whether any stood up to scrutiny. Table 1 presents the solutions proposed for the four systems in question in their discovery works, together with the uncertainties on the orbital parameters, and references to both the original work and our dynamical investigations. To illustrate the instability we observed for these

³ A seventh system, NY Vir, has one claimed companion, and the authors suggest there may be a second – but do not place any constraints on its orbital parameters [25]. As such, that system is not considered further in this work.

systems, we present dynamical stability maps for HU Aquarii, HW Virginis, QS Virginis and V1828 Aquilae⁴ (NSVS 14256825) in Figure 2.

Table 1: The best-fit orbital solutions for the four candidate PCEB companion systems that have proven to be dynamically unstable on astronomically short timescales.

System	Companion	Minimum Mass (M_J)	Semi-Major Axis (au)	Eccentricity	References
HU Aqr	b	5.9 ± 0.6	3.6 ± 0.8	0.0	[31][32][33]
	c	4.5 ± 0.5	5.4 ± 0.9	0.51 ± 0.15	
HW Vir	b	8.5 ± 0.42	3.62 ± 0.52	0.31 ± 0.15	[34][35]
	c	19.2 ± 0.03	5.30 ± 0.23	0.46 ± 0.05	
V1828 Aquilae	b	2.8 ± 0.3	1.9 ± 0.3	0.00 ± 0.08	[30][36]
	c	8.0 ± 0.8	2.9 ± 0.6	0.52 ± 0.06	
QS Vir	b	8	6.031 ± 0.051	0.62 ± 0.02	[29][37]
	c	57	7.043 ± 0.019	0.92 ± 0.02	



⁴ It was brought to our attention during the review process for this paper that NSVS 14256825 was recently given the official variable star designation V1828 Aquilae, as part of the 80th name list of variable stars, details of which can be found at <http://www.konkoly.hu/cgi-bin/IBVSpdf?6052>. As a result, we hereafter refer to it by its new designation.

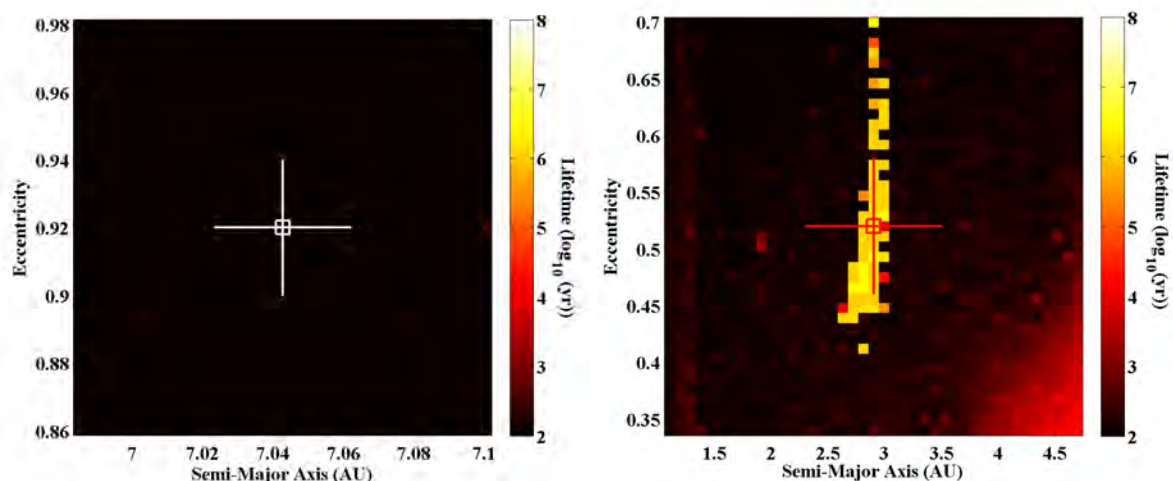


Figure 2: Dynamical stability maps for HU Aquarii (top left), HW Virginis (top right), QS Virginis (lower left) and V1828 Aquilae (lower right). In each case, the stability is mapped as a function of the semi-major axis (a , in au) and eccentricity (e) of the companion with the most poorly constrained orbital elements. The colour at each (a - e) location shows the mean lifetime at that location, taken over all tested values of ω and M . The simulations for HU Aquarii were carried out at lower resolution than those for the other three systems, since at the time we had access to significantly less computational power. As a result, the lifetimes in each small square of that plot are the mean of 21 individual trials, whilst those for the other three are the mean of 75 unique trials.

Each of the four systems detailed in Table 1 proves dynamically unstable on timescales of just a few hundred, or a few thousand years. Even in the case of V1828 Aquilae, the narrow region of enhanced stability features lifetimes of just ~ 1 Myr, far shorter than the estimated age and lifetime of the host stars. In each case, therefore, we conclude that the observed variations in the timing of eclipses between the two central stars are most likely not the result of unseen massive companions, although we acknowledge that the thin band of stability for V1828 Aquilae does suggest that, with future observations, it may be possible to find some dynamically plausible solutions for that system⁵. If such companions do exist, they must move on orbits drastically different to those proposed in the discovery works, and detailed in Table 1. It is interesting to note that three of the four systems detailed in Table 1 (HU Aqr, HW Vir and V1828 Aquilae) feature orbital semi-major axes in the ratio 1:1.5 (corresponding to orbital periods in a ratio $\sim 1:1.84$, i.e. nominal best-fit orbits just tighter than the mutual 1:2 mean-motion resonance between the proposed companions). The striking similarity between the orbital architectures proposed for these companions is interesting, and may perhaps suggest that the same physical process is responsible for the observed eclipse timing variations – if, indeed, those variations are real rather than an artefact of underestimated observational uncertainties⁶.

⁵ As further observations are obtained, the uncertainties on the fit to the orbit will decrease significantly, if the observed “wobble” is truly the result of massive companions. It is likely that such a decrease would be accompanied by some shift in the best fit values for the orbital rotation angles, which might we result in a “fine tuning” of the system towards a stable, resonant solution such as that hinted at by the narrow strip of stability shown in Fig. 2.

⁶ We note that the problem of dynamically unfeasible companions claimed on the basis of eclipse timings between close binaries is not purely limited to PCEBs. The stars SZ Dra and RZ Dra, both of which are Algol type binaries, have also recently had massive companions proposed to orbit them on the basis of the timing of their mutual eclipses. As is the case for the four unstable systems described in this work, the candidate companions are proposed on

The Stable Systems

The candidate planetary systems proposed around the PCEBs NN Serpentis ([42]) and UZ For ([43]) move on orbits that more closely resemble those of the planets in our own Solar system, and the bulk of known exoplanets. The orbits of the proposed companions have either low or moderate eccentricity and are relatively widely spaced, which in turn significantly reduces the likelihood of strong mutual perturbations on astronomically short timescales. The best-fit solutions proposed for these two candidate planetary systems are presented in Table 2.

In the case of NN Ser, the discovery work ([42]) proposed two potential architectures for the candidate planetary system, featuring companions trapped in mutual 2:1 and 5:2 mean-motion resonance. In our dynamical study of the system ([44]), we found that both these solutions would be dynamically stable for a wide range of plausible orbits for NN Ser (AB) d. However, we noted that the 5:2 resonant solution placed that planet on an orbit that was very close to a broad region of instability (see e.g. Fig. 4 in [44]). We therefore concluded that the 2:1 solution was the more likely of the two, when the stability of the orbits was considered in isolation. We present the best-fit orbits for that 2:1 solution, together with the dynamical stability map for the system, as the left-hand panels of Figure 3.

More recently, further observations of NN Ser have allowed the discoverers to significantly improve their orbital solution for the system ([45]), ruling out the 5:2 resonant solution. If the observed variations in the eclipses of NN Ser are the result of perturbations from massive companions, then it seems likely that those objects move on orbits that are trapped in mutual 2:1 mean-motion resonance, a conclusion consistent with our dynamical investigation.

The candidate planets orbiting UZ For move on orbits that are even more widely spaced than those in the NN Ser system, with orbital periods that place them just exterior to their mutual 3:1 mean-motion resonance. The closest analog to the two candidate planets in our own Solar system are the planets Saturn and Uranus, whose orbital periods differ by a factor of ~ 2.89 (compared to the ~ 3.05 difference between the periods of the UZ For companions). The orbital eccentricities of the candidate planets are almost identical to those of Saturn and Uranus – although their masses are significantly greater than the planets in our Solar system. Nevertheless, this comparison would instinctively suggest that the candidate planets should be dynamically stable on long timescales – an expectation that is borne out by the results of our simulations, as can be seen in the right hand panels of Figure 3. The great bulk of the tested solutions were dynamically stable, with the only broad region of instability lying beyond a semi-major axis of ~ 3.7 au for UZ For (AB) c. Orbits beyond this distance lie closer to that of UZ For (AB) b than the location of their mutual 2:1 mean-motion resonance, meaning that their instability is entirely to be expected (as was seen for the six systems discussed in the previous section).

Table 2: The best-fit orbital solutions for the two candidate PCEB companion systems that have been shown to be dynamically feasible.

System	Companion	Minimum Mass (M_J)	Semi-Major Axis (au)	Eccentricity	References
--------	-----------	---------------------------	-------------------------	--------------	------------

mutually crossing orbits, and prove dynamically unstable on astronomically short timescales. We direct the interested reader to [38] and [39] (SZ Dra) and [40] and [41] (RZ Dra) for more details on these systems.

NN Ser	c	6.91 ± 0.54	5.38 ± 0.20	0.0	[42][44][45][46]
	d	2.28 ± 0.38	3.39 ± 0.10	0.20 ± 0.02	
UZ For	b	6.3	5.9 ± 1.4	0.04 ± 0.05	[43]
	c	7.7	2.8 ± 0.5	0.05 ± 0.05	

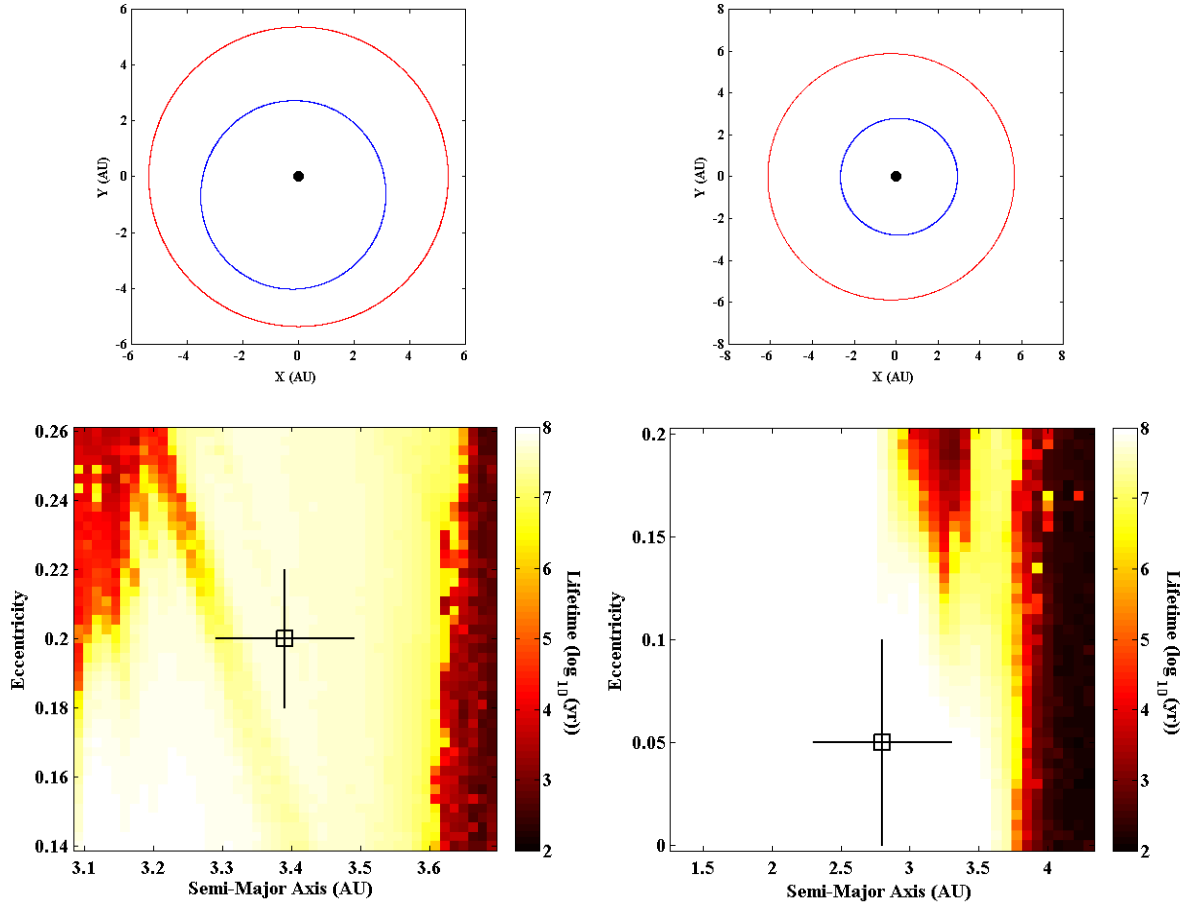


Figure 3: The best-fit orbits (top) and dynamical stability maps (bottom) for the candidate planetary systems around NN Ser (left) and UZ For (right). For NN Ser, the stability map for the 2:1 resonant solution is shown. In both cases, the orbits are significantly more widely separated, and less eccentric, than those proposed for the other six candidate PCEB planetary systems. The best-fit solutions lie in broad regions of dynamical stability, in which the majority of solutions remained stable for the full duration of the 100 Myr simulations.

Circumbinary Planets and Stellar Evolution

Given that the planetary systems proposed orbiting NN Ser and UZ For stand up to dynamical scrutiny, it is interesting to consider whether the proposed planets could have formed at the same time as the host stars, before surviving the post-main sequence evolution of those two stars before acquiring their current orbits. We are therefore now in the process of studying the formation and evolution of these two candidate systems to see whether planets formed therein could survive the evolution of the system away from the main sequence to its current compact and dynamic state. Our study of the NN Ser system was recently completed ([46]), whilst that for UZ For has just begun.

In the case of the NN Ser system, we began with the assumption that the candidate planets formed with the binary star, and attempted to reconstruct a variety of initial binaries that could have evolved to give rise to the current configuration of NN Ser. Whilst the primary was still on the main sequence, the separation of the binary itself would have been much wider than that observed today, whilst the candidate planets would have moved on significantly smaller orbits. We used the binary star evolution code BSE ([47]) to follow the evolution of a large number of potential progenitor binaries through the primary's main sequence, red giant and asymptotic giant branch phases. Our modelling took account of the binary orbital changes due to their mutual tides and the common envelope phase, in which the two stars move within an envelope of material that consists primarily of the disrupted outer atmosphere of the giant primary. These effects combine, resulting in the orbits of the binary components decaying to the compact configuration observed today. In this way we identified the range of progenitor main sequence binary systems whose evolution renders them compatible with the current binary system. For more detail on our binary reconstruction procedure, we direct the interested reader to section 2 of [46].

We then investigated the stability on the main sequence of the potential progenitor systems identified through the stellar evolution modelling described above, when the binary's orbit was much wider and the planets' much tighter. The scale of the planetary orbits whilst on the main sequence was determined under the assumption that those orbits would have expanded adiabatically during the period of mass loss from the central binary as the primary evolved off the main sequence, as codified by Equation 1 in [46]. We followed the orbits of the bodies in the progenitor binaries with the MERCURY N-body code ([23]) over the primary's main sequence lifetime. Almost all the progenitor systems compatible with the evolution of the binary were dynamically unstable on very short timescales, typically of a few hundred or thousand years⁷. We therefore conclude that the candidate planets proposed to explain the variations in eclipse timing for NN Ser are not survivors from before the Common Envelope Phase. Either the objects formed from material that was not entirely ejected during the post-main sequence evolution of the system (which seems highly unlikely, see the discussion of [42], and references therein), or some other, non-planetary, explanation must be sought to explain the observed timing variations.

Conclusion

Over the last few years, six multiple-companion systems have been proposed orbiting highly evolved binary star systems known as post-common envelope binaries (PCEBs). Those systems were all proposed based on periodic variations in the observed timing of eclipses between the two components of the evolved binary systems. We have shown that, of the six systems proposed, only two stand up to dynamical scrutiny – the rest are found to be catastrophically unstable on timescales of just a few hundred, or a few thousand years.

For the four dynamically unfeasible systems, it is clear that the proposed companions do not provide a reasonable explanation for the observed variations in the timing of the observed

⁷ The main sequence binaries corresponding to the current PCEB state of NN Ser featured binary semi-major axes between ~ 0.5 and 1.0 au, with the innermost planet residing at between ~ 1.04 and ~ 1.13 au. Such systems are inherently highly unstable, as a result of strong interactions between the planets and the secondary star. The observed instability is not unexpected – indeed, the instability can be predicted directly for all but the most compact main sequence binary configurations using the simple empirical expression given by equation 5 in [51].

eclipses. Interestingly, a recent study of the circumbinary companions proposed to orbit the Algol-type binary RZ Dra has revealed that the uncertainties on the eclipse timings used in the discovery analysis may well have been significantly under-estimated ([41]). Indeed, in that work, we show that merely increasing the uncertainties by just ± 5 seconds from their stated values is enough to make all evidence for the proposed companions disappear. Given that there are a number of factors that could deleteriously affect the precision with which the timing of mid-eclipse can be determined⁸, it is not unreasonable to suggest that the uncertainties of these measurements may sometimes be underestimated. Indeed, we note that [31] adopt an uncertainty of just ± 3.4 seconds for their eclipse timing observations, despite the fact that their observed eclipses display a large amount of variability and noise. Although we have not yet had chance to perform a similar analysis for the other unstable systems, it therefore seems reasonable to assume that the observations used in the discovery works may be victim to similar problems, and that therefore, once further observations of those systems are carried out, the evidence for unseen companions may disappear.

In the case of the remaining two systems, NN Ser and UZ For, our dynamical simulations reveal that the candidate companions would move on orbits that are stable on long timescales. As a result, the existence of companions in those systems cannot be ruled out on the basis of their current dynamical evolution alone. As a result, we have begun a program of study investigating whether planets that formed with the host binary would be able to survive the binary's post-main sequence evolution before being emplaced on the orbits proposed around the PCEB. In the case of NN Ser, our study suggests that it is highly implausible that planets could survive the transition from the main sequence to the PCEB state for architectures that would result in the proposed planetary system ([46]). However, recent observations ([48]) have added further data to the archive for NN Ser, an analysis of which strengthens the conclusion that the observed periodic variation in eclipse timings for that system is a real effect – and that therefore the unseen companion hypothesis remains strongly supported by the observational data⁹. If the proposed planets do exist, then that system will prove a fascinating test-bed for models of exoplanet formation and evolution, and also for the fine details of post-main sequence evolution of close binary star systems.

Acknowledgements

This research has made use of the Exoplanet Orbit Database and the Exoplanet Data Explorer at <http://exoplanets.org>. The work was supported by iVEC through the use of advanced computing resources located at the Murdoch University, in Western Australia. TCH gratefully acknowledges financial support from the Korea Research Council for Fundamental Science and Technology (KRCF) through the Young Research Scientist Fellowship Program, and also the support of the Korea Astronomy and Space Science Institute (KASI) grant 2013-9-400-00. JPM is supported by Spanish grant AYA 2011/26202. AJM is supported by Spanish grant

⁸ Examples of such factors include the influence of star spots on the observed flux; [49], difficulties introduced by handling old data for which the time standard used has not been explicitly stated; [50], and the asymmetric and highly variable shape of eclipse light curves in PCEBs, as is well illustrated by Fig. 1 of [31].

⁹ We note that, while this article was under review, [52] uploaded an article to the arXiv discussing a possible route by which the candidate planets orbiting NN Ser could have been formed from material ejected from the common envelope during the evolution of the system to become a PCEB – a model that will be tested by future observations of both NN Ser itself, and other PCEB systems.

AYA 2010/20630. The authors wish to thank the two anonymous referees of this work, whose comments led to changes that improved the flow and clarity of the final version.

References

1. Wright, J. T., Fakhouri, O., Marcy, G. W., Han, E., Feng, Y., Johnson, J. A., Howard, A. W., Fischer, D. A., Valenti, J. A., Anderson, J. and Piskunov, N., 2011, "The Exoplanet Orbit Database", *Publications of the Astronomical Society of the Pacific*, Volume 123, Issue 902, pp. 412-422
2. Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D. and Dyon, R., 2008, "Direct Imaging of Multiple Planets Orbiting the Star HR 8799", *Science*, Volume 322, Issue 5906, pp. 1348-1352
3. Lagrange, A.-M., Gratadour, D., Chauvin, G., Fusco, T., Ehrenreich, D., Mouillet, D., Rousset, G., Rouan, D., Allard, F., Gendron, É., Charton, J., Mugnier, L., Rabou, P., Montri, J. & Lacombe, F., 2009, "A probable giant planet imaged in the β Pictoris disk. VLT/NaCo deep L'-band imaging", *Astronomy and Astrophysics*, Volume 493, Issue 2, pp. L21-L25
4. Mayor, M. and Queloz, D., 1995, "A Jupiter-mass companion to a solar-type star", *Nature*, Volume 378, Issue 6555, pp. 355-359
5. Butler, R. P., Johnson, J. A., Marcy, G. W., Wright, J. T., Vogt, S. S. and Fischer, D. A., 2006, "A Long-Period Jupiter-Mass Planet Orbiting the Nearby M Dwarf GJ 849", *The Publications of the Astronomical Society of the Pacific*, Volume 118, Issue 850, pp. 1685-1689
6. Tinney, C. G., Wittenmyer, R. A., Butler, R. P., Jones, H. R. A., O'Toole, S. J., Bailey, J. A., Carter, B. D. and Horner, J., 2011, "The Anglo-Australian Planet Search XXI. A Gas-giant Planet in a One Year Orbit and the Habitability of Gas-giant Satellites", *The Astrophysical Journal*, Volume 732, Issue 1, article id, 31
7. Charbonneau, D., Brown, T. M., Latham, D. W. and Mayor, M., 2000, "Detection of Planetary Transits Across a Sun-like Star", *The Astrophysical Journal*, Volume 539, Issue 1, L45 – L48
8. Collier Cameron, A., Bouchy, F., Hébrard, G., Maxted, P., Pollacco, D., Pont, F., Skillen, I., Smalley, B., Street, R. A., West, R. G., Wilson, D. M., Aigrain, S., Christian, D. J., Clarkson, W. I., Enoch, B., Evans, A., Fitzsimmons, A., Fleenor, M., Gillon, M., Haswell, C. A., Hebb, L., Hellier, C., Hodgkin, S. T., Horne, K., Irwin, J., Kane, S. R., Keenan, F. P., Loeillet, B., Lister, T. A., Mayor, M., Moutou, C., Norton, A. J., Osborne, J., Parley, N., Queloz, D., Ryans, R., Triaud, A. H. M. J., Udry, S. and Wheatley, P. J., 2007, "WASP-1b and WASP-2b: two new transiting exoplanets detected with SuperWASP and SOPHIE", *Monthly Notices of the Royal Astronomical Society*, Volume 375, Issue 3, pp. 951-957
9. Penev, K., Bakos, G. Á., Bayliss, D., Jordán, A., Mohler, M., Zhou, G., Suc, V., Rabus, M., Hartman, J. D., Mancini, L., Béky, B., Csubry, Z., Buchhave, L., Henning, T., Nikolov, N., Csák, B., Brahm, R., Espinoza, N., Conroy, P., Noyes, R. W., Sasselov, D. D., Schmidt, B., Wright, D. J., Tinney, C. G., Addison, B. C., Lázár, J., Papp, I. and Sári, P.,

- 2013, “HATS-1b: The First Transiting Planet Discovered by the HATSouth Survey”, *The Astronomical Journal*, Volume 145, Issue 1, article id. 5.
10. Charbonneau, D., Brown, T. M., Noyes, R. W. and Gilliland, R. L., 2002, “Detection of an Extrasolar Planet Atmosphere”, *The Astrophysical Journal*, Volume 568, Issue 1, pp. 377-384
 11. Seager, S. and Deming, D., 2010, “Exoplanet Atmospheres”, *Annual Review of Astronomy and Astrophysics*, Vol. 48, p.631-672
 12. Fossey, S. J., Waldmann, I. P. and Kipping, D. M., 2009, “Detection of a transit by the planetary companion of HD 80606”, *Monthly Notices of the Royal Astronomical Society: Letters*, Volume 396, Issue 1, pp. L16 – L20
 13. Addison, B. C., Tinney, C. G., Wright, D. J., Bayliss, D., Zhou, G., Hartman, J. D., Bakos, G. Á. and Schmidt, B., 2013, “A Nearly Polar Orbit for the Extrasolar Hot Jupiter WASP-79b”, *The Astrophysical Journal Letters*, Volume 774, Issue 1, article id. L9, 5
 14. Vogt, S. S., Butler, R. P. and Haghighipour, N., 2012, “GJ 581 update: Additional evidence for a Super-Earth in the habitable zone”, *Astronomische Nachrichten*, Vol. 333, Issue 7, p. 561
 15. Wittenmyer, R. A., Horner, J. and Tinney, C. G., 2012, “Resonances Required: Dynamical Analysis of the 24 Sex and HD 200964 Planetary Systems”, *The Astrophysical Journal*, Volume 761, Issue 2, article id. 165, 6
 16. Wittenmyer, R. A., Tinney, C. G., Butler, R. P., O’Toole, S. J., Jones, H. R. A., Carter, B. D., Bailey, J. and Horner, J., 2011, “The Frequency of Low-mass Exoplanets. III. Toward η_{earth} at Short Periods”, *The Astrophysical Journal*, Volume 738, Issue 1, article id. 81, 6
 17. Howard, A. W. et al. (66 co-authors), 2012, “Planet Occurrence within 0.25 au of Solar-type stars from Kepler”, *The Astrophysical Journal Supplement*, Volume 201, Issue 2, article id. 15, 20
 18. Johnson, J. A., Fischer, D. A., Marcy, G. W., Wright, J. T., Driscoll, P., Butler, R. P., Hekker, S., Reffert, S. and Vogt, S. S., 2007, “Retired A Stars and Their Companions: Exoplanets Orbiting Three Intermediate-Mass Subgiants”, *The Astrophysical Journal*, Volume 665, Issue 1, pp. 785-793
 19. Wittenmyer, R. A., Endl, M., Wang, L., Johnson, J. A., Tinney, C. G. and O’Toole, S. J., 2011, “The Pan-Pacific Planet Search. I. A Giant Planet Orbiting 7 CMa”, *The Astrophysical Journal*, Volume 743, Issue 2, article id. 184, 7
 20. Zorotovic, M. and Schreiber, M. R., 2013, “Origin of apparent period variations in eclipsing post-common-envelope binaries”, *Astronomy and Astrophysics*, Volume 549, id. A95, 16
 21. Robertson, P., Horner, J., Wittenmyer, R. A., Endl, M., Cochran, W. D., MacQueen, P. J., Brugamyer, E. J., Simon, A. E., Barnes, S. I. and Caldwell, C., 2012, “A Second Giant Planet in 3:2 Mean-motion Resonance in the HD 204313 System”, *The Astrophysical Journal*, Volume 754, Issue 1, article id. 50, 9

22. Marshall, J., Horner, J. and Carter, A., 2010, “Dynamical simulations of the HR 8799 planetary system”, *International Journal of Astrobiology*, Volume 9, pp. 259 - 264
23. Chambers, J. E., 1999, “A hybrid symplectic integrator that permits close encounters between massive bodies”, *Monthly Notices of the Royal Astronomical Society*, Volume 304, pp. 793 – 799
24. Horner, J., Wittenmyer, R. A., Tinney, C. G., Robertson, P., Hinse, T. C. and Marshall, J. P., 2013, “Dynamical Constraints on Multi-Planet Exoplanetary Systems”, *Proceedings of the 12th Annual Australian Space Science Conference*, ISBN: 978-0-9775740-6-3, editors: Wayne Short and Iver Cairns, Published by the National Space Society of Australia Ltd, Sydney, Australia, p. 103 – 116
25. Qian, S.-B., Zhu, L.-Y., Dai, Z.-B., Fernández-Lajús, E., Xiang, F.-Y. and He, J.-J., 2012, “Circumbinary Planets Orbiting the Rapidly Pulsating Subdwarf B-type Binary NY Vir”, *The Astrophysical Journal Letters*, Volume 745, Issue 2, article id. L23, 5
26. Horner, J., Müller, T. G. and Lykawka, P. S., 2012, “(1173) Anchises – Thermophysical and dynamical studies of a dynamically unstable Jovian Trojan”, *Monthly Notices of the Royal Astronomical Society*, Volume 423, Issue 3, pp. 2587-2596
27. Horner, J. and Lykawka, P. S., 2010, “2001 QR322: a dynamically unstable Neptune Trojan?”, *Monthly Notices of the Royal Astronomical Society*, Volume 405, pp. 49 – 56
28. Malhotra, R., 1995, “The Origin of Pluto’s Orbit: Implications for the Solar System Beyond Neptune”, *The Astronomical Journal*, Volume 110, p. 420
29. Lee, J. W., Kim, S.-L., Kim, C.-H., Koch, R. H., Lee, C.-U., Kim, H.-I. and Park, J.-H., 2009, “The sdB+M Eclipsing System HW Virginis and its Circumbinary Planets”, *The Astronomical Journal*, Volume 137, Issue 2, pp. 3181 - 3190
30. Almeida, L. A. and Jablonski, F., 2011, “Two bodies with high eccentricity around the cataclysmic variable QS Vir”, *The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution*, Proceedings of the International Astronomical Union, IAU Symposium, Volume 276, pp. 495-496
31. Qian, S.-B., Liu, L., Liao, W.-P., Li, L.-J., Zhu, L.-Y., Dai, Z.-B., He, J.-J., Zhao, E.-G., Zhang, J. and Li, K., 2011, “Detection of a planetary system orbiting the eclipsing polar HU Aqr”, *Monthly Notices of the Royal Astronomical Society: Letters*, Volume 414, Issue 1, pp. L16-L20
32. Horner, J., Marshall, J. P., Wittenmyer, R. A. and Tinney, C. G., 2011, “A dynamical analysis of the proposed HU Aquarii planetary system”, *Monthly Notices of the Royal Astronomical Society: Letters*, Volume 416, Issue 1, pp. L11-L15
33. Wittenmyer, R. A., Horner, J., Marshall, J. P., Butters, O. W. and Tinney, C. G., 2012, “Revisiting the proposed planetary system orbiting the eclipsing polar HU Aquarii”, *Monthly Notices of the Royal Astronomical Society*, Volume 419, Issue 4, pp. 3258-3267
34. Lee, J. W., Kim, S.-L., Kim, C.-H., Koch, R. H., Lee, C.-U., Kim, H.-I. and Park, J.-H., 2009, “The sdB+M Eclipsing System HW Virginis and its Circumbinary Planets”, *The Astronomical Journal*, Volume 137, Issue 2, pp. 3181-3190

35. Horner, J., Hinse, T. C., Wittenmyer, R. A., Marshall, J. P. and Tinney, C. G., 2012, "A dynamical analysis of the proposed circumbinary HW Virginis planetary system", *Monthly Notices of the Royal Astronomical Society*, Volume 427, Issue 4, pp. 2812-2823
36. Wittenmyer, R. A., Horner, J. and Marshall, J. P., 2013, "On the dynamical stability of the proposed planetary system orbiting NSVS 14256825", *Monthly Notices of the Royal Astronomical Society*, Volume 431, Issue 3, pp. 2150-2154
37. Horner, J., Wittenmyer, R. A., Hinse, T. C., Marshall, J. P., Mustill, A. J. and Tinney, C. G., 2013, "A detailed dynamical investigation of the proposed QS Virginis planetary system", *Monthly Notices of the Royal Astronomical Society*, Advance Access, 7 pp.
38. Lee, J. W., Lee, C.-U., Kim, S.-L., Kim, H.-I. and Park, J.-H., 2012, "The Algol System SZ Herculis: Physical Nature and Orbital Behaviour", *The Astronomical Journal*, Volume 143, Issue 2, article id. 34, 8 pp.
39. Hinse, T. C., Goździewski, K., Lee, J. W., Haghighipour, N. and Lee, C.-U., 2012, "The Proposed Quadruple System SZ Herculis: Revised LITE Model and Orbital Stability Study", *The Astronomical Journal*, Volume 144, Issue 2, article id. 34, 10 pp.
40. Yang, Y.-G., Li, H.-L., Dai, H.-F. and Zhang, L.-Y., 2010, "The Near-contact Binary RZ Draconis with Two Possible Light-time Orbits", *The Astronomical Journal*, Volume 140, Issue 6, pp. 1687-1693
41. Hinse, T. C., Horner, J., Lee, J. W., Wittenmyer, R. A., Lee, C.-U. and Park, J.-H., 2013, "On the RZ Draconis Sub-stellar Circumbinary Companions", *Monthly Notices of the Royal Astronomical Society*, submitted
42. Beuermann, K., Hessman, F. V., Dreizler, S., Marsh, T. R., Parsons, S. G., Winget, D. E., Miller, G. F., Schreiber, M. R., Kley, W., Dhillon, V. S., Littlefair, S. P., Copperwheat, C. M. and Hermes, J. J., 2010, "Two planets orbiting the recently formed post-common envelope binary NN Serpentis", *Astronomy and Astrophysics*, Volume 521, id. L60, 5 pp.
43. Potter, S. B., Romero-Colmenero, E., Ramsay, G., Crawford, S., Gulbis, A., Barway, S., Zietsman, E., Kotze, M., Buckley, D. A. H., O'Donoghue, D., Siegmund, O. H. W., McPhate, J., Welsh, B. Y. and Vallerga, J., 2011, "Possible detection of two giant extrasolar planets orbiting the eclipsing polar UZ Fornacis", *Monthly Notices of the Royal Astronomical Society*, Volume 416, pp. 2202
44. Horner, J., Wittenmyer, R. A., Hinse, T. C. and Tinney, C. G., 2012, "A detailed investigation of the proposed NN Serpentis planetary system", *Monthly Notices of the Royal Astronomical Society*, Volume 425, Issue 1, pp. 749 - 756
45. Beuermann, K., Dreizler, S. and Hessman, F. V., 2013, "The quest for companions to post-common envelope binaries. IV. The 2:1 mean-motion resonance of the planets orbiting NN Serpentis", *Astronomy & Astrophysics*, Volume 555, id.A133, 8 pp.
46. Mustill, A. J., Marshall, J. P., Villaver, E., Veras, D., Davis, P. J., Horner, J. and Wittenmyer, R. A., 2013, "Main-sequence progenitor configurations of the NN Ser candidate circumbinary planetary system are dynamically unstable", *Monthly Notices of the Royal Astronomical Society*, in press: arXiv:1309.3881

47. Hurley, J. R., Tout, C. A. and Pols, O. R., 2002, "Evolution of binary stars and the effect of tides on binary populations", *Monthly Notices of the Royal Astronomical Society*, Volume 329, Issue 4, pp. 897 – 928
48. Marsh, T. R., Parsons, S. G., Bours, M. C. P., Littlefair, S. P., Copperwheat, C. M., Dhillon, V. S., Breedt, E., Caceres, C. and Schreiber, M. R., 2013, "The planets around NN Ser: still there", *Monthly Notices of the Royal Astronomical Society*, in press, arXiv: 1310.1391
49. Watson, C. A. and Dhillon, V. S., 2004, "The effect of star-spots on eclipse timings of binary stars", *Monthly Notices of the Royal Astronomical Society*, Volume 351, Issue 1, pp. 110-116
50. Eastmann, J., Siverd, R. and Gaudi, B. S., 2010, "Achieving Better Than 1 Minute Accuracy in the Heliocentric and Barycentric Julian Dates", *Publications of the Astronomical Society of the Pacific*, Volume 122, Issue 894, pp. 935-946
51. Holman, M. J. and Wiegert, P. A., 1999, "Long-term stability of planets in binary systems", *The Astronomical Journal*, Volume 117, pp. 621-628
52. Schleicher, D. R. G. and Dreizler, S., 2014, "Planet formation from the ejected of common envelopes", *Astronomy and Astrophysics*, submitted, arXiv:1312.3479

A dynamical test for terrestrial planets in the habitable zone of HD 204313

Elodie Thilliez*, Lea Jouvin[†], Sarah T. Maddison* and Jonathan Horner^{‡§}

* *Centre for Astrophysics & Supercomputing, Swinburne University of Technology,
Hawthorn, VIC, 3122, Australia*

[†] *Universite Paris-Sud 11, 91400 Orsay, France*

[‡] *School of Physics, University of New South Wales, Sydney, NSW, 2052, Australia*

[§] *Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW 2052,
Australia*

Summary: With improvements in exoplanet detection techniques, the number of multiple planet systems discovered is increasing, while the detection of potentially habitable Earth-mass planets remains complicated and thus requires new search strategies. Dynamical studies of known multiple planet systems are therefore a vital tool in the search for stable and habitable planet candidates.

Here, we present a dynamical study of the three-planet system HD 204313 to determine whether it could harbour an Earth-like planet within its habitable zone for a sufficient time to develop life. We found two semi-stable regions in the system, but neither prove stable for long enough for a terrestrial planet to develop life. Our investigations suggest that overlapping weak and high order resonances may be responsible for these semi-stable regions. This study established a framework for a larger project that will study the dynamical stability of the habitable zone of multiple planet systems, providing a list of interesting targets for future habitable low-mass planet searches.

Keywords: Planetary systems, Numerical methods: N-body simulation, Planetary systems: dynamical evolution and stability, Exoplanets, Habitability, Astrobiology

Introduction

Since the discovery of the first exoplanet, more than a thousand exoplanets have been confirmed, including 170 multiple planet systems¹. With improvements in exoplanet detection techniques, new discoveries are expected and will continue to enhance our understanding of planetary systems and the planet formation process. However the detection of habitable Earth-mass planets remains complicated, and only 12 super-Earth planets are currently known to orbit within the habitable zone of their host star [1], while 35 additional Kepler candidates² are still waiting to be confirmed.

The habitable zone (HZ) is defined as the region around a star where liquid water could be present on the surface of rocky planets [2]. The presence of liquid water on the surface depends

¹Current exoplanet tally taken from the Extrasolar Planets Encyclopaedia, <http://www.exoplanet.eu>, on the 11th November 2013

²Current estimation from the Habitable Exoplanets Catalog, <http://phl.upr.edu>, on the 11th November 2013

on a broad range of parameters, such as the stellar incident energy, spectral distribution, radiative properties of the planetary atmosphere (absorption, diffusion, emission, clouds) and the reflectivity of its surface (albedo) [2]. Earth is the only habitable planet of which we currently know, and it appears that life took ~ 1 Gyr to develop a significant atmospheric signature that would be detectable [3]. Using the Earth as a baseline, we will therefore work on the assumption that a planet would need to survive for at least 1 Gyr within a star's habitable zone to detect any life there. This therefore requires additional constraints to be placed on the orbit of the planet for it to be deemed detectably habitable - not only must it currently orbit within the habitable zone, but it must also maintain an orbit which is simultaneously dynamically stable and confined within the HZ (i.e. a small eccentricity) for this entire duration.

Dynamical studies thus play a crucial role in the hunt for habitable exoplanets, and the evaluation of the long-term stability of the HZ of known multiple exoplanet systems can help focus future observations. There are a number of examples in the literature of multiple planet systems which turn out to be dynamically unstable on short timescales. For example, the two planet system orbiting the cataclysmic variable HU Aquarii announced by [4] was shown by [5] and [6] to be unstable on timescales of order 10^4 years, strongly suggesting that the multiplanet system is not real or that the orbital parameters must be different from those reported in the literature. While numerical simulations have been commonly used and are necessary to understand the dynamics of multiple systems (e.g. [5], [7], [8]), analytical stability criteria tools can be used to provide a quick assessment of the stable regions of a multiplanet system (e.g [9], [10]).

We aim to establish a new framework to systematically investigate the stability of HZs of multiple planetary systems. Firstly, the HZ of such systems must be located using atmospheric model criteria, and then a numerical search for dynamically stable regions within the HZ can be conducted. This requires the integration of the orbits of massless test particles randomly distributed inside the HZ. Once stable regions (where test particles survive at least 1 Gyr) are identified, we can then explore their long-term stability by (i) investigating their resonance state, and (ii) producing maps of their stability as a function of orbital elements by varying the stable test particles initial conditions around the stable known initial parameters. Finally, to ensure stability of a potential Earth-mass planets in these stable and habitable zones, numerical simulations must be run by substituting the stable test particles with Earth-mass planets.

In this paper we present a study of the dynamical stability of test particles in the habitable zone of the star HD 204313. Using the N-body integrator SWIFT [11], we first isolate regions of stability within the habitable zone of HD 204313 using massless test particles. We then numerically investigate the long-term stability of test particles in the identified dynamically stable regions of the HZ and compare these locations with the stable regions derived using the analytical stability criteria of [9]. The paper is organized as follows: first, we describe the HD 204313 system, and the location of its habitable zone, in the following section. We then present the analytical and numerical methods used to investigate the stability of the system's habitable zone. We then report our findings in section 3 before discussing the potential origin of two identified semi-stable zones in section 4.

The HD 204313 System and its Habitable Zone

HD 204313 is a metal-rich Sun-like star of spectral type G5 [12], with a mass of $1.02 M_{\odot}$ and age of 7.2 Gyrs [13]. It has been the target of multiple radial velocity surveys and was observed by [14] with the 2.7 m Smith telescope over 8 years to obtain Radial Velocity measurements and a stellar spectrum. Using the MOOG local thermodynamic equilibrium line analysis and spectral analysis program [15], they determined the effective temperature, T_{eff} and by matching the abundances of the Fe I and Fe II lines, they determined the metallicity of HD 204313, $[Fe/H]$. The star's distance and absolute magnitude, m_V , were also derived by [14] and all their stellar properties are summarized in Table 1.

Table 1: Stellar properties for HD 204313 from [14].

Property	Value
Spectral type ^a	G5 V
Age ^b	7.2 Gyrs
Mass ^b	$1.02 M_{\odot}$
Distance	47.0 ± 0.3 pc
T_{eff}	5760 ± 100 K
m_V	4.63 ± 0.03
$[Fe/H]$	0.24 ± 0.06

^a From [12]

^b From [13], maximum likelihood estimate.

This system is known to harbour at least three exoplanets: [16] presented the discovery of HD 204313 b, a super-Jupiter planet ($M \sin(i) \sim 3.5 M_J$) with a period $P \sim 5$ years with the CORALIE Echelle spectrograph, while [17] revealed an interior ($P \sim 35$ days) Neptune-mass planet HD 204313 c with the HARPS spectrograph. [14] simultaneously fitted the data from the CORALIE Echelle spectrograph in conjunction with their observations from the Harlan J. Smith Telescope at McDonald Observatory. The residuals from their fully generalised Lomb-Scargle periodogram for a one-planet model (HD 204313 b) showed significant peaks, notably around $P \sim 2700$ days. By fitting an additional planet at this period, their fitting routine converged for an additional Jupiter-mass planet: HD 204313 d with mass $M \sin(i) \sim 1.68 M_J$, and a period of 2831 days. They did not include the planet HD 204313 c in their study since its radial velocity amplitude as detected by [17] was below the sensitivity limit of their data, but that planet could be responsible for the remaining scatter in their two planet fit. We will be using the orbital parameters derived by [14] for the two planets, HD 204313 b and d, in our study – see Table 2. Their best-fit orbits reveal that HD 204313 b and d are trapped in a mutual 3:2 mean-motion resonance (MMR). By performing further dynamical analysis, the authors found that outside this MMR, there is no stable configuration for these two planets, and that fine tuning of orbital elements of the two planets was required in order for this system to survive. Moreover they confirmed that adding the inner planet, HD 204313 c, to their dynamical analysis does not affect the behaviour of HD 204313 b and d.

As discussed earlier, the location of the HZ around a given star depends both on the nature of the star itself (parameters such as its luminosity and temperature) and the atmosphere of the planet in question (absorption, diffusion, albedo, emission efficiency and cloud coverage) [2].

Thus the location of the HZ is expected to migrate outward as the star evolves on the main sequence and its effective temperature and luminosity increase. Therefore the time available for a planet to develop life in its HZ depends on (i) the width of this zone and (ii) the lifetime of a star on its main sequence (which depends of the stellar mass and metallicity) [18]. There are several ways to define the inner and outer boundary of the HZ (as discussed in [19], [20], [18] and [21]). We select the inner boundary as the maximum distance from the star at which a ‘runaway greenhouse effect’ would lead to the evaporation of all surface water, while for the outer boundary we chose the ‘maximum greenhouse effect’, which defines the distance at which a cloud-free CO₂ atmosphere could maintain a surface temperature of 273 K.

To calculate the distance of these boundaries from the star, we followed the work of [22], who express the effective flux received at the HZ boundaries, S_{eff} , as a quadratic function of the difference between the effective temperature of the star and that of the Sun, $T_{eff} - T_{\odot}$. The slope of this function is given by coefficients that depend on the boundary criteria of the HZ, with all coefficients for each boundary given in [22]. We selected the coefficients for a ‘runaway greenhouse effect’ as inner boundary and the coefficients for a ‘maximum greenhouse effect’ as an outer boundary. The ratio between the luminosity of the star and our Sun, L/L_{\odot} , is then deduced from its apparent visual magnitude, m_V , and from its distance to our Sun, D , in pc. Finally, the corresponding distance, d_{HZ} , of the limits of the HZ is given by:

$$d_{HZ} = \left(\frac{1}{S_{eff}} \frac{L}{L_{\odot}} \right)^{1/2} \quad (1)$$

Using the stellar parameters from Table 1, we find the HZ boundaries for HD 204313 at:

$$\begin{aligned} HZ_{inner} \text{ (runaway greenhouse)} &= 1.1 \text{ AU} \\ HZ_{outer} \text{ (maximum greenhouse)} &= 1.9 \text{ AU} \end{aligned}$$

Numerical and Analytic methods

We expand on the work of [14], who conducted a dynamical stability study of the HD 204313 system to constrain the orbital parameters of the two outer planets using the N-body dynamics package MERCURY [23]. Because of its negligible mass compared to the other two planets and its short period, HD 204313 c was found to have little impact on the system’s dynamics: its Hill radius, which defines its zone of gravitational influence, is ~ 0.005 AU which means that it is dynamically well separated from the two outer planets and the HZ, and therefore

Table 2: *Orbital parameters of HD 204313 planets from [14]*

Elements	HD 204313 ^a c	HD 204313 b	HD 204313 d
$M \sin(i)$ (M_J)	0.054 ± 0.005	3.55 ± 0.2	1.68 ± 0.3
a (AU)	0.2103 ± 0.0035	3.04 ± 0.06	3.93 ± 0.14
e	0.17 ± 0.09	0.23 ± 0.04	0.28 ± 0.09
P (days)	34.88 ± 0.03	1920 ± 25	2830 ± 150
M ($^{\circ}$)		300 ± 0.4	137 ± 2
ω ($^{\circ}$)		298 ± 6	247 ± 16

^a From [17].

does not significantly perturb them. As a result, those authors did not include that planet in their long-term simulations. In their study, the orbital parameters of HD 204313 b were kept fixed, while those of HD 204313 d were varied: the eccentricity, e , semi-major axis, a , longitude of periastron, ω , and mean anomaly, M , of HD 204313 d were systematically varied by $\pm 3\sigma$ around their best fit values, with simulations run for a maximum 10^8 years. The orbits of the two giant planets are assumed to be coplanar. They derived two stability maps displaying the mean dynamical lifetime of the HD 204313 b and d planetary system as (i) a function of the initial semi-major axis and eccentricity of planet d, and (ii) as a function of the initial semimajor axis and longitude of periastron of planet d – see their Figure 5 [14].

We used the SWIFT integration software package to numerically integrate the orbits of HD 204313 b and HD 204313 d over a similar duration (10^8 years). We used the N-body symplectic RMVS integrator of SWIFT, a well-tested integrator first computed by [11] which conserves the Hamiltonian during the integration and integrates close encounters between planets and test particles. Their goal was to model the orbit of short-period comets under the gravitational influence of all the Solar system’s planets except Mercury. To ensure that SWIFT gives similar results to MERCURY, we repeated a subsample of the simulations of [14] with SWIFT, selecting 13 simulations along the edge of the stable zone in both a, e and a, ω space – see Table 3. We found that 11/13 of the simulations were stable for 10^8 years, with the

Table 3: Initial orbital parameters of HD 204313 d for the 13 simulations taken from [14]

	1	2	3	4	5	6	7	8	9 ^a	10	11	12	13 ^a
a (AU)	3.90	3.93	3.93	3.98	3.98	3.98	3.98	3.98	4.01	4.01	4.01	4.01	4.07
e	0.41	0.12	0.3	0.05	0.05	0.12	0.12	0.30	0.05	0.05	0.24	0.24	0.21
ω ($^\circ$)	292	256	283	238	247	247	256	292	238	247	283	292	274
M ($^\circ$)	137	137	137	137	137	137	137	137	137	137	137	137	131

^a Models 9 and 13 were found to be unstable.

two unstable models originating at the extreme edge of the stability zone of [14] (see their Figure 5). Given that MERCURY and SWIFT do not handle close encounters in exactly the same manner, it is not surprising to see some differences for critical cases. Whilst MERCURY treats a close encounter between a planet and test particle by switching from a symplectic integrator to a classic Bulirsch–Stoer integrator to ensure the precision set by the user is reached, SWIFT treats the close encounter using the same symplectic RMVS integrator but switching to the planet as the barycenter of the system instead of the star during the encounter.

Using these 11 stable models, we investigated the system’s suitability for harboring additional stable bodies by randomly distributing 2000 massless test particles in the HZ of HD 204313 and integrating for 10^8 years. The test particles had initial eccentricities e_{TP} ranging from 0.0 to 0.3, inclinations i_{TP} of 0 degrees, and random values of longitude of periastron, ω_{TP} , and mean anomaly, M_{TP} . The outcome of these simulations allowed us to identify stable regions within the HZ. We then inspected the long-term stability of these zones by determining the resonance state of the stable particles, and produced maps of the test particle lifetimes as a function of their initial orbital elements. To create these maps, we ran additional simulations with test particle orbital elements ($a_{TP}, e_{TP}, \omega_{TP}, M_{TP}$) distributed over a specific range of values: we examined 51 values of a_{TP} in a radius of 0.6 AU around the stable zone; 51 values of e_{TP} between 0.0–0.3 for each a_{TP} ; 19 values of ω_{TP} between 0.0– 360° for each e_{TP} value and 19 values of M_{TP} between 0.0– 360° for each ω_{TP} value. Thus more than 10^6 test particles with unique initial conditions were generated to derive a

high resolution stability map as a function of orbital elements inside the HZ.

By following this method, the numerical simulations can provide a complete picture of the stability state of test particles in the HZ. However, to ensure no dependence on exact initial conditions, this method requires a large number of simulations (10^6 in our experiment). Analytical criteria are another way to quickly appraise the location of stable regions. By generalizing the resonance overlap criterion for the onset of stochastic behaviour in the planar circular restricted three-body problem from [24] to multiplanetary systems, [9] provide analytical expressions that can put constraints on the location of stable zones. Their criterion is based on the fact that if the orbits of planets in a multiplanetary system cross and are not protected by some resonant mechanism, this could lead to the planets in the system colliding with one another, or being ejected from the system as a result of their mutual interactions. More precisely, around each planet j exists a region of instability, δ_j , where a close encounter with another body will perturb the dynamics enough to lead to chaotic diffusion of the eccentricity and semi-major axis, resulting in either collision or escape. The width of this region depends on the mass, m , eccentricity, e , and semi-major axis, a , of the planet, and on the mass of the encountering body. In our case, we will apply this criterion of the two planets HD 204313 b and HD 204313 d encountering an Earth-mass planet. Thus δ_j is given by [9]:

$$\delta_j = 1.57 \times a(j) \times \left(\left(\frac{m(j)}{m_{star}} \right)^{2/7} + \left(\frac{m_{earth}}{m_{star}} \right)^{2/7} \right). \quad (2)$$

In the simple case of a planet j on a circular orbit, as long as no other bodies enter the region $a_j \pm \delta_j$ around the planet, the bodies will be stable with respect to planet j . For eccentric planets such as HD 204313 b and d, this region is extended to $q_j - \delta_j$ and $Q_j + \delta_j$, where q_j is the periastron and Q_j the apastron of planet. We use the expressions derived by [9] to assess the position of the interior and exterior limit of the unstable zone around HD 204313 b and d, and compare it to the stability zone of our test particles in the HZ from our numerical simulations.

Results & Discussion

In each of our simulations, all of the test particles were removed from the HZ on timescales far shorter than the maximum possible run time (10^8 years). As a result, the integrations were terminated when the final test particle was removed - and so no simulation actually completed its full 100 million year run time. Figure 1 shows the time evolution of the semi-major axis, a , for the test particles and the two planets for the longest lasting model in which the final test particle was ejected after a period of $\sim 9 \times 10^6$ years (model 4 in Table 3). In all other models, the time at which the final test particle was removed was always significantly shorter. Figure 2 zooms into the HZ region of model 4, in which it is apparent that all test particles in the HZ are rapidly removed by the gravitational influence of the two outer giant planets, aside from two areas located around $a \sim 1.5$ AU and $a \sim 1.4$ AU. We observed the same stable regions in each of our 11 simulations, and those regions were always located at the same distance from the central star. We therefore carried out a more detailed investigation in order to determine why test particles in these two regions displayed semi-stable behaviour across our entire suite of simulations.

One potential explanation for the behaviour is resonance trapping (e.g. [25]), and so we examined the potential mean motion resonances (MMR) located between 1.4 and 1.6 AU

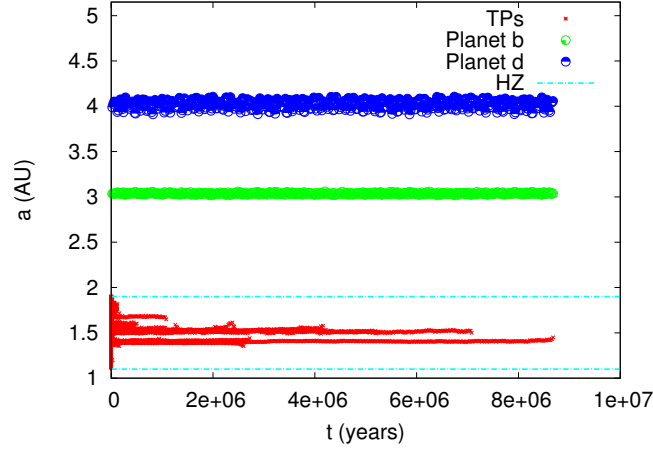


Fig. 1: Evolution of the semi-major axis, a , for the test particles and planets in model 4. The dotted lines represent the location of the habitable zone.

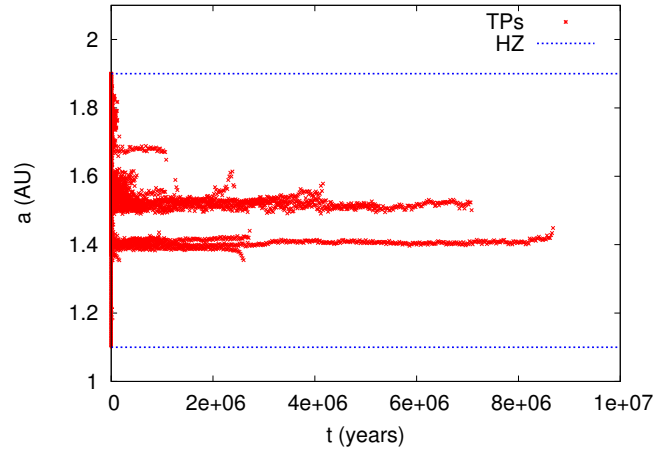


Fig. 2: Zoom of the HZ region of model 4, following the evolution of the semi-major axis, a , for the test particles. The dashed lines represent the location of the habitable zone.

resulting from the two giant planets, HD 204313 b and d. Particles orbiting around $a \sim 1.5$ AU have a period $P_{TP_{1.5}} \sim 667$ days, while particles orbiting around $a \sim 1.4$ AU have a period $P_{TP_{1.4}} \sim 601$ days. Thus the ratio between the orbital periods of the particles trapped at 1.5 AU and HD 204313 d is $P_d/P_{TP_{1.5}} = 4.2$, suggesting that test particles may possibly be trapped in a 1:4 MMR with HD 204313 d. Similarly, the ratio between the orbital periods of the particles trapped at 1.4 AU and HD 204313 b is $P_b/P_{TP_{1.4}} = 3.2$, suggesting that test particles may possibly be trapped in a 1:3 MMR with HD 204313 b³. To test this scenario, we plotted the evolution of the resonant argument, ϕ , for those MMRs:

$$\phi_d = \lambda_{TP} - 4\lambda_d + 3\omega_{TP} \quad (3)$$

$$\phi_b = \lambda_{TP} - 3\lambda_b + 2\omega_{TP} \quad (4)$$

³We note that exact commensurability in the ratio of the periods is not required for the resonant angle to librate. Resonance can occur on a small distance from exact commensurability in dimensionless units of period ratio, which is called 'resonance width' [26].

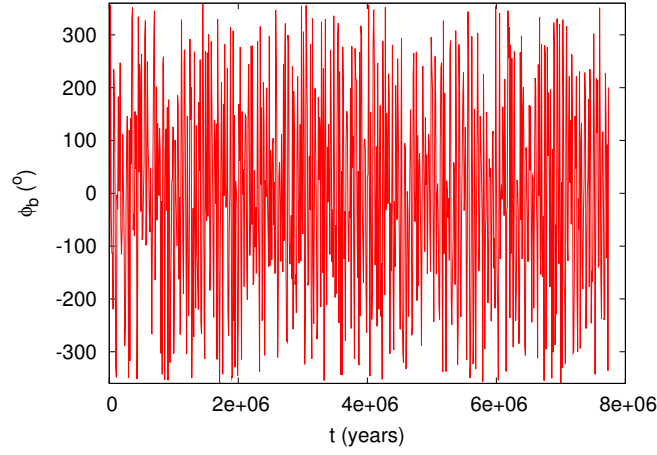


Fig. 3: Variation in the resonant argument ϕ_b given by Eq. 3 for the longest living test particle at 1.4 AU of model 4.

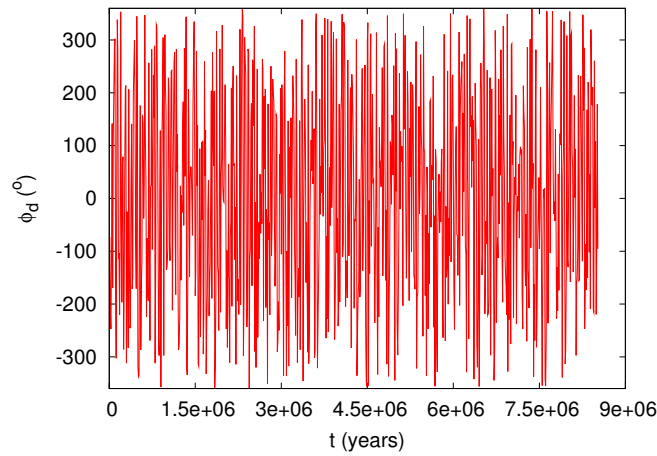


Fig. 4: Variation in the resonant argument ϕ_d given by Eq. 4 for the longest living test particle at 1.5 AU of model 4.

where λ is the mean longitude and ω the argument of the periastron. However we found that these resonant arguments did not exhibit libration or any periodic behaviour for any of the test particles tested – as can be seen in Figures 3 and 4.

Since the locations of the semi-stability regions do not match any strong MMRs with HD 204313 b or HD 204313 d, we examined whether the test particles might be trapped in regions where a number of overlapping weak and/or high order MMRs could be combining to create a region of "sticky chaos", acting to stabilise the test particles motion [25]. We calculated all possible resonances with HD 204313 b or HD 204313 d up to order 51 and plotted their individual resonant arguments with respect to both planets. However, we were unable to identify any clear resonant behaviour. We therefore conclude that no single weak/high order resonance is responsible for these semi-stable regions, although we note that test particles may be captured and stabilised as a result of the overlap between a number of these high-order resonances. In such a scenario, test particles would chaotically hop from one resonance to another, without spending any significant period of time trapped in any specific resonance. Such behaviour has been invoked in the past to explain regions of stability for

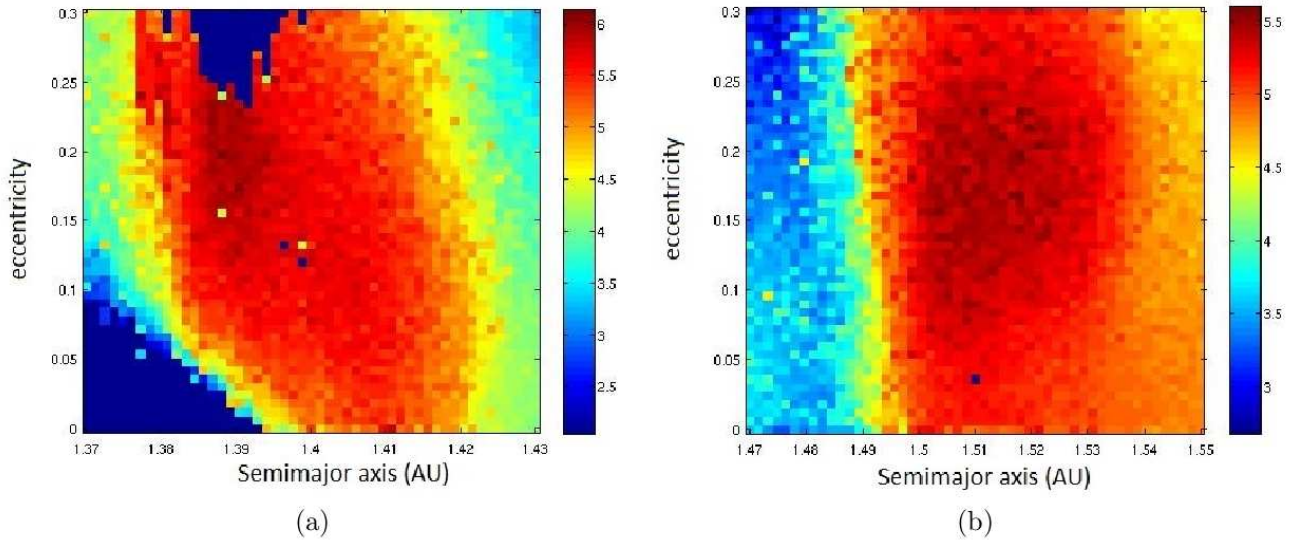


Fig. 5: The mean dynamical lifetime of the test particles as a function of (a) the initial semimajor axis, a , and eccentricity, e , around 1.4 AU, and around (b) 1.5 AU. The lifetime in years is plotted with a logarithmic color scale.

other exoplanetary systems [27].

In order to better define the two semi-stable regions, we produced high resolution maps of the mean dynamical test particle lifetime as a function of their initial orbital elements by running additional simulations with 10^6 test particles distributed around 1.4 AU and 1.5 AU. Figure 5 shows the mean dynamical lifetime of the test particles as a function of the initial semi-major axis, and eccentricity. Whilst stable zones are observed for orbits spanning the entire range of orbital eccentricities tested (between 0 and 0.3), the stability of the test particles is clearly most strongly influenced by the semi-major axis of the test particles orbit. We thus conclude that fine tuning of initial semi-major axis is required to obtain stable orbits within the HZ. However, while all simulations were run for 10^8 years, as can be seen in Figure 5, the simulations were stable for no longer than 10^6 years.

We now compare the location of the two semi-stable regions inside the HZ with the stability map analytically derived using the criterion of [9] – see Figure 6. The first remark is that planets b and d are mutually located inside their unstable regions and our work therefore supports the conclusions of [14] that the 3:2 MMR is responsible for maintaining these two planets in a stable configuration. Whilst the inner limit of the unstable zone of HD 204313 d (i.e. to the left of the innermost green curve in Figure 6) is located between 1.8 AU for highly eccentric particles and 2.8 AU for particles with $e \sim 0.3$, the inner limit of the unstable zone of HD 204313 b (i.e. to the left of the innermost red curve in Figure 6) is between 1.4 AU for highly eccentric particles and 1.8 AU for particles with $e \sim 0.25$. This map therefore explains why no stable orbits are found exterior to 1.5 AU within the HZ – beyond this semi-major axis, the proximity between additional bodies and the known planets would result in the disruption of the system.

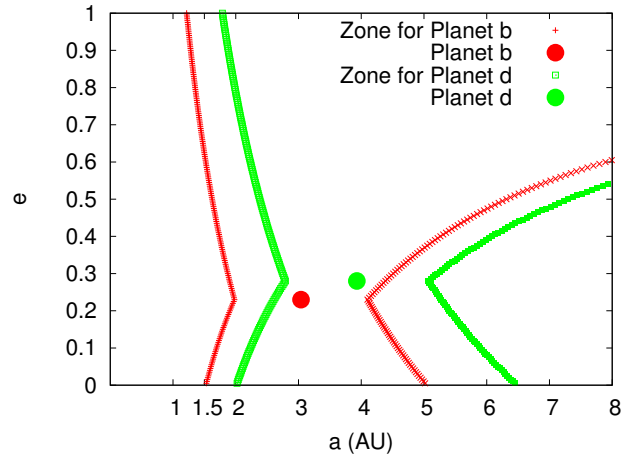


Fig. 6: Theoretical stability map for HD 204313. The dots represent the location of the planets HD 204313 b (red) and HD 204313 d (green), while the dashed lines delimit the extended orbit crossing regions around HD 204313 b (red) and HD 204313 d (green).

Conclusions

We propose a new method to investigate the stability of potential terrestrial planets in the habitable zone of multiplanetary systems. This method combines numerical integrations and an analytical criterion to assess the suitability of the HZ to host additional long-term stable objects. Using atmospheric model criterion, we first defined the location of the HZ before testing its appropriateness to host massless test particles by numerically integrating their orbits on a timescale required for the establishment of life. We examined the resulting stable regions in terms of the presence of any resonance mechanisms, and produced high resolution maps of the test particles lifetime as a function of their initial orbital elements inside the identified stable regions. As a final step, we compared the location of the numerical stable zones with the region allowed by an analytical criteria which checks for orbit crossings. This approach provides a more complete picture of the dynamics of the HZ. We applied this method to the system HD 204313, which is mainly composed of a Sun-like star and two giant planets, HD 204313 b and d, orbiting respectively at 3.04 and 3.95 AU (we ignored HD 204313 c). We report the following results and conclusions:

- Using the ‘Runaway greenhouse effect’ and ‘Maximum greenhouse effect’, we defined the HZ of HD 204313 b and d as the region between 1.1–1.9 AU.
- By distributing massless test particles throughout the HZ of the HD 204313 system and testing their stability with numerical simulations, we found two semi-stable regions near 1.4 and 1.5 AU.
- Although no single mean motion resonance up to an order 51 was identified as controlling the evolution of the test particles in those regions, our investigations do not exclude that particles could be trapped due to overlapping weak and high-order resonances with the two outer planets of the system.
- Using the analytical criterion from [9], we confirm that the 3:2 MMR between HD 204313 b and d is required to maintain this system in a stable configuration. Moreover, if not protected by any resonance mechanism, no additional planets can be located in the HZ with $a > 1.5$ AU at the risk of destabilizing the orbits of the two outer giant planets.
- However, it must be noted that these two semi-stable regions remain stable for only

a short period of $\sim 9 \times 10^6$ years, which is less than the timescale required for the emergence of life ($\sim 10^9$ years). Thus those two semi-stable regions are not suitable for a terrestrial planet to develop life.

While we did not find a zone of potential long-term stability and habitability for planets in the HZ of the HD 204313 system, this study established a framework for a larger project that will study many different systems using a similar method. One could follow this approach for all known multiple systems with well constrained orbital elements. Generally one expects low eccentricity systems would have a higher chance of hosting stable Earth-like planet in their habitable zone. If we find multiple systems which can host stable Earth-mass planets in their HZ, these systems could be targeted to search for low-mass planets in future surveys.

Acknowledgments

All simulations were run on the Swinburne supercomputer. LJ was supported by the Swinburne Centre for Astrophysics & Supercomputing. We thank Rosemary Mardling for useful discussions, and the referees for their feedback.

References

- [1] Borucki, W. J. et al., “Kepler-62: A Five-Planet System with Planets of 1.4 and 1.6 Earth Radii in the Habitable Zone”, *Science*, 2013, Vol. 340, pages 587-590
- [2] Kasting, J. F. and Whitmire, D. P. and Reynolds, R. T., “Habitable Zones around Main Sequence Stars”, *ICARUS*, 1993, Vol. 101, pages 108-128
- [3] Chyba, C. F., “The violent environment of the origin of life: Progress and uncertainties”, *Geochimica et Cosmochimica Acta*, 1993, Vol. 57, pages 3351-3358
- [4] Qian, S.-B. and Liu, L. and Liao, W.-P. and Li, L.-J. and Zhu, L.-Y. and Dai, Z.-B. and He, J.-J. and Zhao, E.-G. and Zhang, J. and Li, K., “Detection of a planetary system orbiting the eclipsing polar HU Aqr”, *Monthly Notices of the Royal Astronomy Society*, 2011, Vol. 414, pages L16-L20
- [5] Horner, J. and Marshall, J. P. and Wittenmyer, R. A. and Tinney, C. G., “A dynamical analysis of the proposed HU Aquarii planetary system”, *Monthly Notices of the Royal Astronomy Society*, 2011, Vol. 416, pages L11-L15
- [6] Wittenmyer, R. A. and Horner, J. and Marshall, J. P. and Butters, O. W. and Tinney, C. G., “Revisiting the proposed planetary system orbiting the eclipsing polar HU Aquarii”, *Monthly Notices of the Royal Astronomy Society*, 2012, Vol. 419, pages 3258-3267
- [7] Hinse, T. C. and Michelsen, R. and Jørgensen, U. G. and Goździewski, K. and Mikkola, S., “Dynamics and stability of telluric planets within the habitable zone of extrasolar planetary systems. Numerical simulations of test particles within the HD 4208 and HD 70642 systems”, *Astronomy & Astrophysics*, 2008, Vol. 488, pages 1133-1147
- [8] Sándor, Z. and Süli, Á. and Érdi, B. and Pilat-Lohinger, E. and Dvorak, R., “A stability catalogue of the habitable zones in extrasolar planetary systems”, *Monthly Notices of the Royal Astronomy Society*, 2007, Vol. 375, pages 1495-1502
- [9] Giuppone, C. A. and Morais, M. H. M., and Correia, A. C. M., “A semi-empirical stability criterion for real planetary systems with eccentric orbits”, *arXiv*: 1309.6861, 2013

- [10] Deck, K. M. and Payne, M. and Holman, M. J., “First-order Resonance Overlap and the Stability of Close Two-planet Systems”, *The Astrophysical Journal*, 2013, Vol. 774, page 129
- [11] Levison, H. F. and Duncan, M. J., “The long-term dynamical behavior of short-period comets”, *ICARUS*, Vol. 108, 1994, pages 18-36
- [12] Kharchenko, N. V. and Roeser, S., “All-sky Compiled Catalogue of 2.5 million stars”, *VizieR Online Data Catalog*, 2009, Vol. 1280
- [13] Casagrande, L. and Schönrich, R. and Asplund, M. and Cassisi, S. and Ramírez, I. and Meléndez, J. and Bensby, T. and Feltzing, S., “New constraints on the chemical evolution of the solar neighbourhood and Galactic disc(s). Improved astrophysical parameters for the Geneva-Copenhagen Survey”, *Astronomy & Astrophysics*, 2011, Vol. 530, pages A138
- [14] Robertson, P., Horner, J. Wittenmyer, R. A., Endl, M., Cochran, W. D., MacQueen, P.J., Brugamyer, E. J., Simon, A. E., Barnes, S. I. and Caldwell, C., “A Second Giant Planet in 3:2 Mean-motion Resonance in the HD 204313 system”, *The Astrophysical Journal*, Vol. 754, Issue 1, 2012, article id. 50
- [15] Sneden, C., “The nitrogen abundance of the very metal-poor star HD 122563”, *The Astrophysical Journal*, 1973, Vol. 184, pages 839-849
- [16] Ségransan, D. and Udry, S. and Mayor, M. and Naef, D. and Pepe, F. and Queloz, D. and Santos, N. C. and Demory, B.-O. and Figueira, P. and Gillon, M. and Marmier, M. and Mégevand, D. and Sosnowska, D. and Tamuz, O. and Triaud, A. H. M. J., “The CORALIE survey for southern extrasolar planets. XVI. Discovery of a planetary system around HD 147018 and of two long period and massive planets orbiting HD 171238 and HD 204313”, *Astronomy & Astrophysics*, 2010, Vol. 511, page A45
- [17] Mayor, M. and Marmier, M. and Lovis, C. and Udry, S. and Ségransan, D. and Pepe, F. and Benz, W. and Bertaux, J. - and Bouchy, F. and Dumusque, X. and Lo Curto, G. and Mordasini, C. and Queloz, D. and Santos, N. C., “The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets”, *arXiv: 1109.2497v1*, 2011
- [18] Underwood, D. R. and Jones, B. W. and Sleep, P. N., “The evolution of habitable zones during stellar lifetimes and its implications on the search for extraterrestrial life”, *International Journal of Astrobiology*, Vol. 2, 2003, pages 289-299
- [19] Jones, B. W. and Underwood, D. R. and Sleep, P. N., “Prospects for Habitable Earths in Known Exoplanetary Systems ”, *The Astrophysical Journal*, Vol. 622, 2005, pages 1091-1101
- [20] Jones, B. W. and Underwood, D. R. and Sleep, P. N., “Which exoplanetary systems could harbour habitable planets?”, *International Journal of Astrobiology*, 2006, Vol. 5, pages 251-259
- [21] Menou, K. and Tabachnik, S., “Dynamical Habitability of Known Extrasolar Planetary Systems”, *The Astrophysical Journal*, 2003, Vol. 583, pages 473-488
- [22] Kopparapu, R. K. and Ramirez, R. and Kasting, J. F. and Eymet, V. and Robinson, T. D. and Mahadevan, S. and Terrien, R. C. and Domagal-Goldman, S. and Meadows, V. and Deshpande, R., “Habitable Zones around Main-sequence Stars: New Estimates”, *The Astrophysical Journal*, Vol. 770, 2013, page 82
- [23] Chambers, J. E., “A hybrid symplectic integrator that permits close encounters between massive bodies”, *Monthly Notices of the Royal Astronomy Society*, Vol. 304, 1999, pages 793-799
- [24] Wisdom, J., “The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem”, *Astronomical Journal*, 1980, Vol. 85, pages 1122-1133
- [25] Lykawka, P. S. and Mukai, T., “Resonance sticking in the scattered disk”, *ICARUS*, 2007, Vol. 192, pages 238-247

- [26] Mardling, R. A., “New developments for modern celestial mechanics - I. General coplanar three-body systems. Application to exoplanets”, *Monthly Notices of the Royal Astronomy Society*, 2013, Vol. 435, pages 2187-2226
- [27] Gozdziewski, K. and Slonina, M. and Migaszewski, C. and Rozenkiewicz, A., “Testing a hypothesis of the ν Octantis planetary system”, *Monthly Notices of the Royal Astronomy Society*, 2013, Vol. 430, pages 533-545

Planets in Spin-Orbit Misalignment and the Search for Stellar Companions

Brett C. Addison^{*†}, C. G. Tinney^{*†}, Duncan J. Wright^{*†}, Graeme Salter^{*†}, Daniel Bayliss[‡]
and George Zhou[‡]

^{*} *School of Physics, University of New South Wales, Sydney, NSW 2052, Australia;
b.addison@unsw.edu.au*

[†] *Australian Centre for Astrobiology, University of New South Wales, NSW 2052, Australia*

[‡] *Research School of Astronomy and Astrophysics, Australian National University,
Canberra, ACT 2611, Australia*

Summary: The discovery of giant planets orbiting close to their host stars was one of the most unexpected results of early exoplanetary science. Astronomers have since found that a significant fraction of these ‘Hot Jupiters’ move on orbits substantially misaligned with the rotation axis of their host star. We recently reported the measurement of the spin-orbit misalignment for WASP-79b by using data from the 3.9 m Anglo-Australian Telescope. Contemporary models of planetary formation produce planets on nearly coplanar orbits with respect to their host star’s equator. We discuss the mechanisms which could drive planets into spin-orbit misalignment. The most commonly proposed being the Kozai mechanism, which requires the presence of a distant, massive companion to the star-planet system. We therefore describe a volume-limited direct-imaging survey of Hot Jupiter systems with measured spin-orbit angles, to search for the presence of stellar companions and test the Kozai hypothesis.

Keywords: planets and satellites: dynamical evolution and stability stars: individual (WASP-79) techniques: radial velocities and direct imaging

Introduction

Exoplanetary science is possibly the most exciting and rapidly developing branch of modern astronomy. Just over 1000 planets¹ have been discovered to date, mainly through radial velocity and transit searches. A detailed discussion of the various methods used to detect exoplanets is beyond the scope of this paper, but for more information, we direct the interested reader to [1]. In addition to finding new planets, a detailed analysis of their structure, composition, and other bulk properties is needed to understand the processes involved in their formation and migration. It is possible to probe these processes by measuring the sky-projected spin-orbit alignment (or obliquity) of exoplanetary systems. This is the angle between the planetary orbital plane and the spin vector of the host star. This is done through spectroscopic measurements of the Rossiter-McLaughlin effect (first measured for eclipsing binaries ([2],[3]) and since extended to exoplanets [4] including the recently measured planet WASP-79b [5]). The observed effect is caused by the modification of the stellar spectrum as a transiting planet occults a small region of the stellar disk of its host star, causing a radial velocity anomaly, due to asymmetric distortions in the rotationally broadened stellar line profiles [6].

¹<http://exoplanet.eu>, as of 2014 January 15. There is some debate on the fraction of planets that truly exist. The other main online exoplanet database, <http://exoplanets.org/>, list the total confirmed planets at just 763 as of 2014 January 15.

The radial velocity anomaly is sketched in Figure 1. For prograde orbits, a planet will first transit across the portion of its host star's disk that is rotating towards the observer (i.e. blue-shifted), blocking light from that hemisphere. This results in the observer receiving a greater fraction of the total flux of the star from the hemisphere that is rotating away (red-shifted) from the observer than that rotating towards the observer. The blue-shifted portion of the rotationally broadened stellar lines will appear to have less absorption, resulting in the line profile centroids being red-shifted and a positive radial velocity anomaly. A negative velocity anomaly will occur during the second half of the transit as the planet moves across the hemisphere rotating away from the observer. For retrograde orbits, a planet will transit across the red-shifted hemisphere first, resulting in the inverse velocity anomaly. By measuring the shape and magnitude of the Rossiter-McLaughlin effect, it is possible to determine the inclination of a planet's orbit relative to the spin axis of its host star.

Giant planets are thought to form within the proto-planetary disk that surrounds a protostar through the core-accretion process [7]. This model predicts that Jovian planets should form several AU² away from the protostar where the proto-planetary disk is sufficiently cool for icy volatiles to exist and to slowly accrete into planetesimals. The planetesimals continue to accrete material, growing until they reach a critical mass of $\sim 5 - 10 M_{\oplus}$, at which point they rapidly accrete gas from the surrounding disk. This leads to the formation of a large gaseous envelope of hydrogen and helium. Accretion halts and planet formation comes to an end once the gas in the local disk is exhausted or blown away by the protostar as it reaches the main sequence phase ([7],[8]).

There are two complicating factors to this simple model of planet formation. First, gas giant planets are found well inside 1 AU where they cannot have formed in situ, implying that they must have migrated in from their birthplaces (i.e. disk migration [9]). Secondly, many Hot Jupiters at small orbital radii are observed to be in spin-orbit misalignment [10]. This is unexpected as the proto-planetary disk from which planets form should be well aligned with the plane of the protostar's equator (e.g., [6]; [11]), suggesting that either the migration process, or post-migration evolution of the planet's orbit, has driven planets into spin-orbit misalignment. Several mechanisms have been proposed to produce these anomalies, such as Kozai resonances [12], secular chaos [13], and planet-planet scattering [14].

To date, of the 74 planetary systems³ with measured obliquities, 33 show substantial misalignments ($> 22.5^\circ$), 10 of which are in nearly polar orbits, and 7 are in retrograde orbits. With such a large fraction of planets in spin-orbit misalignment, there is a clear need to understand the physical mechanisms that are producing these systems.

Spin-Orbit Misalignment of WASP-79b

WASP-79b is a bloated Hot Jupiter that was recently discovered through the Wide Angle Search for Planets (WASP) [15]. We determined the spin-orbit misalignment of WASP-79b through spectroscopic measurements of the Rossiter-McLaughlin effect, using high-precision radial velocity observations taken during the transit on the night of 2012 December 23, using

²An astronomical unit, or AU, is a standard unit for measuring distances in astronomy. 1 AU is slightly less than 150 million kilometers, and is approximately the mean distance between the Earth and the Sun.

³This study has made use of René Heller's Holt-Rossiter-McLaughlin Encyclopaedia which was last updated on 2013 November; <http://www.physics.mcmaster.ca/~rheller/index.html>

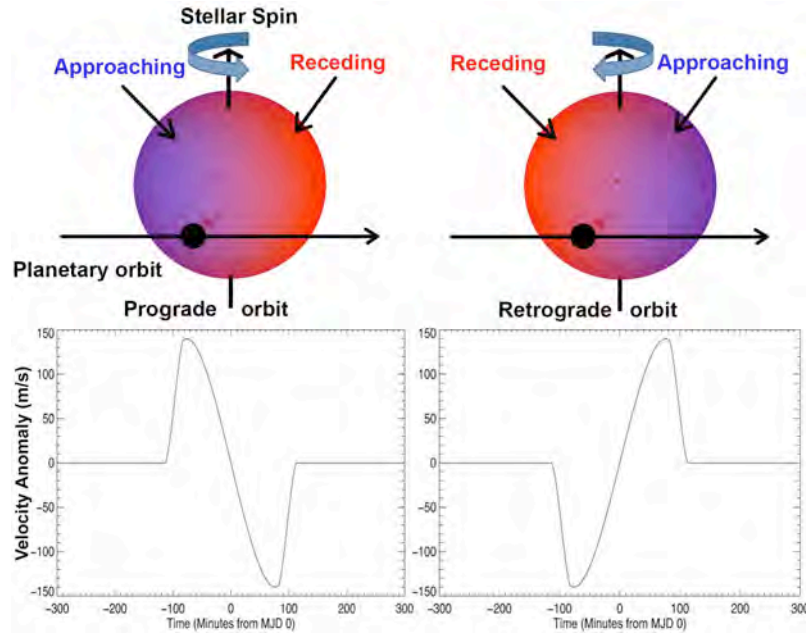


Fig. 1: Top: Artist's impression of a transiting planet on a prograde and a retrograde orbit (for clarity, the motion of the planet across its host star is the same for both prograde and retrograde orbits while the rotation direction of the star is reversed for the retrograde orbit). The stellar hemisphere shaded blue represents the side rotating towards the observer with its light blue-shifted. The stellar hemisphere shaded red represents the side rotating away from the observer with its light red-shifted. Bottom: The modeled radial velocities for the two orbits showing the differing Rossiter-McLaughlin effect as a planet transits across the disk of its host star.

the CYCLOPS2⁴ optical-fiber bundle system feeding the UCLES echelle spectrograph on the Anglo-Australian Telescope at Siding Spring Observatory. Raw spectra were reduced using custom routines developed by the authors and were wavelength calibrated from a thorium-argon observation taken at the beginning of the night and from thorium-xenon spectra taken using the simultaneous calibration fiber during each object exposure [5]. We used the IRAF⁵ task, *fxcor*, to compute radial velocities by cross-correlation with a spectrum of a bright template star (HD86264) of similar spectral type. Details on the reduction and data analysis of WASP-79 can be found in [5].

We developed the Exoplanetary Orbital Simulation and Analysis Model (ExOSAM) to fit our radial velocities of WASP-79b with a model of the Rossiter-McLaughlin effect using the Hirano et al. analytical approach [16]. The best-fitting values for the projected spin-orbit angle, λ , and the projected stellar rotational velocity, $v \sin i_*$ ⁶, along with the uncertainties in these parameters were derived using a grid search and minimizing χ^2 between the observed radial velocities and modeled radial velocities [5].

Two solutions for the stellar parameters of WASP-79 have been derived from photometric

⁴http://www.phys.unsw.edu.au/~cgt/CYCLOPS/CYCLOPS_2.html

⁵IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

⁶ v is the absolute rotational velocity of the star and i_* is the inclination of the star's rotational axis to the observers line of sight

Table 1: The relevant system parameters of WASP-79 as a main sequence and evolved star. The full listing of system parameters for WASP-79 can be found in [5]. Here, ms denotes the model that assumes WASP-79 to be a main sequence star, while non-ms denotes the model that assumes it is instead an evolved star.

Parameter	Value (ms)	Value (non-ms)
<i>Parameters as given by Smalley et al. (2012)</i>		
Mid-transit epoch (2400000-HJD), T_0	56285.03589 ± 0.00200	56285.03739 ± 0.00300
Orbital period, P	3.6623817 ± 0.0000050 d	3.6623866 ± 0.0000085 d
Semi-major axis, a	0.0539 ± 0.0009 AU	0.0535 ± 0.0008 AU
Orbital inclination, i	$85.4 \pm 0.6^\circ$	$83.3 \pm 0.5^\circ$
Impact parameter, b	0.570 ± 0.052	0.706 ± 0.031
Transit depth, $(R_P/R_\star)^2$	0.01148 ± 0.00051	0.01268 ± 0.00063
Orbital eccentricity, e	0.0 (assumed)	0.0 (assumed)
Stellar mass, M_\star	$1.56 \pm 0.09 M_\odot$	$1.52 \pm 0.07 M_\odot$
Stellar radius, R_\star	$1.64 \pm 0.08 R_\odot$	$1.91 \pm 0.09 R_\odot$
Planet mass, M_P	$0.90 \pm 0.09 M_J$	$0.90 \pm 0.08 M_J$
Planet radius, R_P	$1.70 \pm 0.11 R_J$	$2.09 \pm 0.14 R_J$
<i>Parameters determined from model fit using our velocities</i>		
Projected obliquity angle, λ	$-106^{+19}_{-13}^\circ$	$-84^{+23}_{-30}^\circ$
Projected stellar rotation velocity, $v \sin i_\star$	$17.5^{+3.1}_{-3.0} \text{ kms}^{-1}$	$16.0^{+3.7}_{-3.7} \text{ kms}^{-1}$

data [15] – one with WASP-79 on the main sequence ($R_\star = 1.64 \pm 0.08 R_\odot$) and one with it evolved just off the main sequence ($R_\star = 1.91 \pm 0.09 R_\odot$). This results in two sets of solutions for the various other system parameters including λ and $v \sin i_\star$ which are given in Table 1. Our results for the projected spin-orbit alignment and stellar rotation velocity, using the main sequence parameters, are $\lambda = -106^{+19}_{-13}^\circ$ and $v \sin i_\star = 17.5^{+3.1}_{-3.0} \text{ kms}^{-1}$. For the non-main sequence case, $\lambda = -84^{+23}_{-30}^\circ$ and $v \sin i_\star = 16.0^{+3.7}_{-3.7} \text{ kms}^{-1}$. In both cases, WASP-79b is in a significantly misaligned orbit.

The main sequence solution appears to be the most likely one [15], as the main sequence lifetime of a star is significantly longer than its post-main sequence lifetime [17]. Since photometric data neither prefers the main sequence or the evolved solution for WASP-79, and given that a star is far more likely to be observed on the main sequence, we focus on the main sequence solution. Figure 2 shows the velocity anomaly for the main sequence solution with the observed velocities over-plotted on the left. A positive hump-shaped anomaly due to the Rossiter-McLaughlin effect is clearly apparent in our velocities. This implies that the planet must be in a nearly polar orbit and transits across only the blue-shifted hemisphere (or the side rotating toward us) as depicted in the illustration on the right side of figure 2.

Observed Trends of Systems in Spin-Orbit Misalignment

One early trend observed for planetary systems with measured obliquities is that planets in spin-orbit misalignment tend to orbit hot stars with $T_{eff} \geq 6250$ K while planets in spin-orbit alignment tend to orbit cooler stars with $T_{eff} < 6250$ K [18]. The correlation between spin-orbit misalignments and stellar effective temperature may be related to the thickness of the stellar convective zone, since it has been proposed that the convective zone could act to dampen orbital obliquities over time through enhancing planet-star tidal interactions [18]. Hot stars have a thin convective layer and are expected to have very long obliquity dampening timescales, typically orders of magnitude longer than the main sequence lifetime of the star.

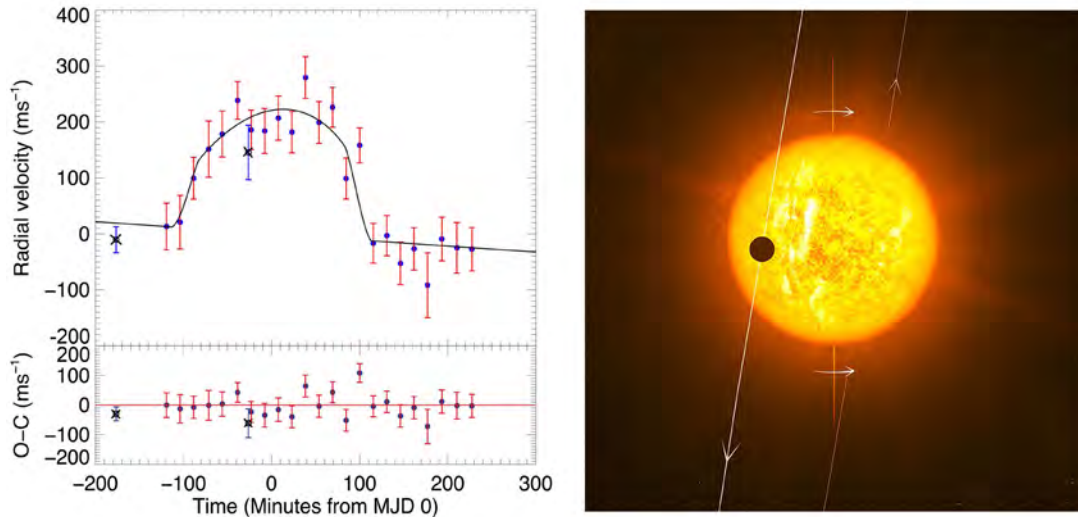


Fig. 2: Left: Spectroscopic radial velocities of the WASP-79b transit taken on 2012 December 23 using the CYCLOPS2 fiber-bundle on the Anglo-Australian Telescope [5]. Velocities from just before, during, and after the transit are plotted as a function of time along with the best fitting model (for the main sequence parameters) and corresponding residuals. The filled blue circles with red error bars are velocities we measured with our estimated uncertainty. The two black circles with an x and with blue error bars are previously published velocities using the quoted uncertainties [15]. Right: Artist impression of the WASP-79b polar orbit. Image credit: Brett Addison (modified version of the original artist impression of the WASP-8b retrograde orbit by ESO/L. Calçada).

On the other hand, stars cooler than 6250 K have much shorter obliquity dampening timescales that are fractions of the main sequence lifetime.

This interpretation of the dichotomy observed for aligned and misaligned systems has been supported by a new study [10], which measured the spin-orbit alignments for 14 new systems and computed the obliquity dampening timescales for these systems and for 39 previously published systems. A positive correlation was found between obliquity and stellar temperature as well as a positive correlation between obliquity and the timescales for dampening obliquity. The reasoning for this observed correlation is similar to that of an earlier study [18].

Recent studies of multi-planet transiting systems have revealed low stellar obliquities for five systems [19] and a high obliquity in one system [20]. The study on low stellar obliquity systems that were known at the time of [19] publication (shortly before the high obliquity system was announced [20]) suggested that the migration mechanism(s) responsible for producing Hot Jupiters is fundamentally different from the mechanism(s) producing compact close-in multi-planet systems [19]. The [19] study proposes that multi-planet systems likely migrated due to disk-planet interactions while Hot Jupiters experienced dynamical perturbations during migration from Kozai resonances or planet-planet scattering.

This view has been challenged by the discovery of a significant spin-orbit misalignment (true obliquity angle $\psi > 37^\circ$) for the Kepler-56 multi-planet system [20]. This result illustrates that spin-orbit misalignments are not restricted to Hot Jupiter systems. The presence

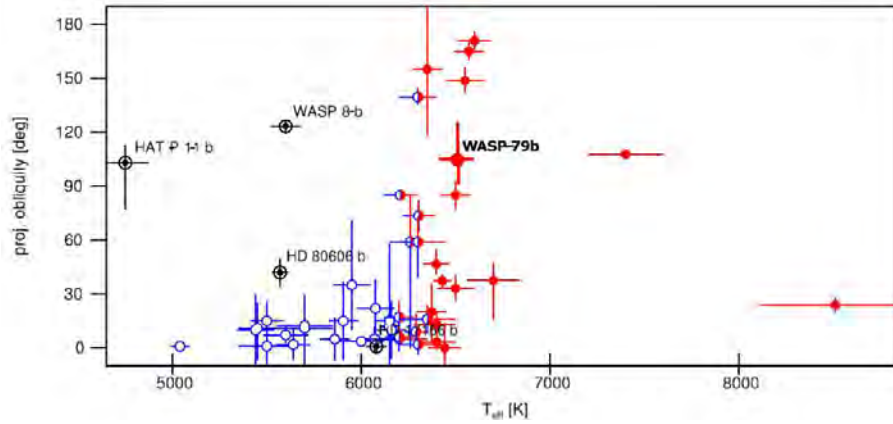


Fig. 3: Projected orbital obliquity as a function of stellar effective temperature [10]. The filled red circles with red error bars are for stars that have temperatures higher than 6250 K. The blue unfilled circles with blue error bars show stars with effective temperatures lower than 6250 K. The circles that are half red and blue show stars that have measured effective temperatures consistent with 6250 K from the 1σ interval. Systems which harbor planets with masses $< 0.2M_J$ or orbital periods longer than 7 days are shown by a black circle with a black dot in the middle and black error bars. WASP-79b has been included in this figure.

of an additional massive body (planet, brown-dwarf, or low mass star) in a wide orbit has been detected from radial velocity measurements which reveal a long term systematic trend [20]. The high stellar obliquity and the presence of a third companion is interpreted as evidence for a scenario in which torques from the outer massive companion drive the inner planets into co-planar orbits that are misaligned with the spin-axis of the host star [20]. It is unlikely that the high obliquity of Kepler-56 is due to the Kozai resonances or planet-planet scattering. Whether a similar scenario to the one proposed for Kepler-56 is responsible for spin-orbit misalignments seen in Hot Jupiters is yet to be determined. Therefore, measuring the obliquities of multi-planet systems is key to determining if such misalignments are common or rare, and whether the mechanism(s) driving misalignments are similar to or different from Hot Jupiters.

We calculated the tidal dissipation timescale for WASP-79b [5] to test its consistency with the overall trends that have been observed for other systems. Using either of the two methods presented in [10], we found that the tidal dissipation timescale for WASP-79b is between $\tau_{mcz} = 1.6 \times 10^{11}$ yr to $\tau_{RA} = 3.3 \times 10^{15}$ yr, which is longer than that calculated for 80% of the systems examined in [19]. WASP-79 has an effective temperature of $T_{eff} = 6600 \pm 100$ K, which is above the $T_{eff} > 6250$ K threshold claimed for planetary systems displaying high orbital obliquities and consistent with the very long tidal dissipation timescale we have calculated for this system. Figures 4 and 5 show the projected orbital obliquity of WASP-79b and those planets presented in [10] as a function of the effective stellar temperature and the relative tidal-dissipation timescale, respectively.

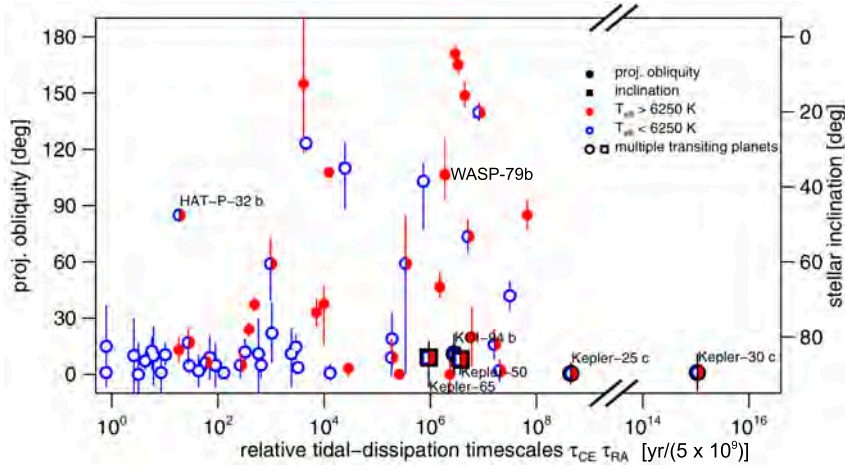


Fig. 4: Projected orbital obliquity as a function of the relative alignment timescale for stars with either convective (CE) or radiative envelopes (RA) calibrated from binary studies [19]. The same symbols are used as in Figure 3 with the addition of multiple transiting planets, indicated by the dark black borders. Systems with measured projected obliquity (λ) are shown as circles while stellar inclinations (i_*) are shown as squares. We have calculated the tidal-dissipation timescale for WASP-79b and include it in this figure.

Mechanisms Driving Spin-Orbit Misalignments

Several mechanisms have been proposed to explain the spin-orbit misalignments observed in many Hot Jupiter systems. These include disk only migration, dynamical mechanisms such as planet-planet scattering, secular chaos, Kozai resonances, or some combination of these during or post-disk migration [21]. Additional scenarios for producing misalignments include stellar internal gravity wave modulation and primordial circumstellar disk misalignments.

Disk migration was one of the earliest and most widely accepted mechanisms invoked to explain the origin of Hot Jupiters [22]. It is understood that gas giant planets form several AU from their host star. However, interactions with the circumstellar disk during and after their formation can induce type 1 and type 2 inward migration [23]. In type 1 migration, planets produce spiral density waves in the circumstellar disk, which results in non-zero tidal torques that drive them inward [23]. Planets more massive than about $10M_{\oplus}$ will quickly clear a gap in the circumstellar disk as they migrate, leading to a transition from type 1 to type 2 migration. Type 2 migration involves the exchange of angular momentum between the planet and the disk as material from the disk enters the gap [23]. If Hot Jupiters are exclusively the result of type 1 and type 2 disk migration, one would expect that their orbits to be well aligned with the stellar spin axis of their host star [9]. This, however, is simply not the case as nearly 45% of Hot Jupiters show significant spin-orbit misalignments⁷.

The standard disk migration model may also play a role in the production of misaligned Hot Jupiters, if the spin-orbit misalignments are the result of the internal gravity wave mechanism altering the orientation of the spin-axis of the host star ([24]; [25]). This model predicts that internal gravity waves are generated at the boundary between the convective core and radiative envelope in relatively hot and massive stars. These waves transport angular momentum

⁷This study has made use of René Heller's Holt-Rossiter-McLaughlin Encyclopaedia which was last updated on 2013 November; <http://www.physics.mcmaster.ca/~rheller/index.html>

outwards towards the surface of the star, causing the surface to rotate at a different speed, and even in a different direction than the interior. Furthermore, the waves themselves can vary over time causing changes in the rotation speed and direction of the stellar surface. While this model does seem to produce systems in spin-orbit misalignment, it has not been determined whether the distribution of obliquities it produces matches the observed population of Hot Jupiters. Other critical tests of this hypothesis include the measurement of time variations in spin-orbit alignments which the internal gravity wave model predicts will occur on timescales of as little as 9 to 1000 rotational periods ([24]; [25]). In addition, measuring the obliquities of multi-planet systems around hot stars with the same misalignments would provide strong evidence for the internal gravity wave model and disfavor dynamical mechanisms. This is because the observed misalignments would be created through changes in the star's spin-axis and would result in all of the planets in a multi-planetary system having nearly co-planar orbits. Misalignments produced through dynamical mechanisms would result in planets on various orbital planes and with different observed obliquities.

An additional scenario in which disk migration will play a key role in producing misaligned Hot Jupiters is through the primordial misalignments of proto-planetary disks [26]. It has been shown that short period misaligned planetary orbits can be the natural product of disk migration in binary systems if the disk is misaligned. If the orbit of a distant stellar companion is inclined by at least 45° to the plane of the disk, then gravitational torques from the companion will drive the disk into misalignment with the spin-axes of its host stars [26]. Planets forming from such a disk will have misaligned orbits. Finding stellar companions around misaligned systems and determining the distribution of obliquities produced by this model will test the validity of this mechanism.

A non-disk migration mechanism proposed for driving spin-orbit misalignments is planet-planet scattering. If there were two or more planets in a multiple planet system in initially unstable orbits, strong gravitational perturbations between them during close encounters can lead to the ejection of one or more planets and the inward migration of the surviving planets. The planets that migrate inwards will have their orbital eccentricity and obliquity increased until tidally circularized if the periastron distance reaches to within a few stellar radii [14]. This mechanism is a leading candidate for explaining the occurrence of giant planets in highly eccentric orbits and has had some success in reproducing the incidence and distribution of these planets as well as the architecture of the Solar System [27]. Solar System dynamics studies using the NICE model have revealed that Jupiter, Saturn, Uranus, and Neptune have experienced planet-planet scattering episodes in the early history of the Solar System approximately 60 My to 1.1 Gy [28]. The NICE model has been able to explain the distribution of the Kuiper belt objects and the outward migration of Uranus and Neptune through planet-planet scattering. It is widely believed that the giant planets of the early Solar System formed on circular and coplanar orbits that were packed significantly closer together (5.5 – 14 AU compared to 19 AU and 30 AU for present day locations of Uranus and Neptune respectively [28]). The initially stable orbits of the Solar System giant planets became destabilized when Jupiter and Saturn crossed their mutual 1:2 mean-motion resonance. This increased their eccentricity slightly allowing the eccentricities of Uranus and Neptune to increase to the point where they had close encounters with each other and migrated outward to their present day locations [28]. The NICE model thus highlights the importance of planet-planet interactions in shaping the Solar System into the architecture we see today and is likely involved in the diversity of observed exoplanetary orbits. Planet-planet scattering does have some shortcomings. The primary shortcoming of this mechanism is that the scattering process tends to be quite violent and sudden, which impedes the ability of slow processes such as

tides to halt the inward migration of planets into their host star [21].

Secular chaos is another non-disk migration mechanism where planetary orbits evolve over long timescales (significantly longer than the orbital periods of the planets) due to small gravitational interactions that occur in systems that have three or more well-spaced planets that (generally) are not in strong mean-motion-resonances [29] (though secular effects can occur for some objects trapped in mean-motion resonances [30]). The inner most planet in such a system will lose angular momentum (but not orbital energy) to an outer planet that will drive its pericenter towards the star as its eccentricity increases while at the same time increase its orbital obliquity [13]. Tidal circularization will then dampen the eccentricity, creating a Hot Jupiter. One success of this model is that it correctly predicts the 3-day orbital period pile-up observed for Hot Jupiters ([13]; [29]). Despite the model's success in this prediction, it hardly produces planets with obliquities $> 90^\circ$ and no planets in retrograde orbits [29], counter to the observed population of Hot Jupiter obliquities.

Kozai resonance is the most commonly adopted mechanism to explain planetary systems in spin-orbit misalignment. This mechanism involves the gravitational interaction between a planet and an outer stellar companion that has an orbit that is highly inclined relative to the orbital plane of the planet and is orbiting at large separations (up to several hundred AU though small separations are also possible [31]) from the central star. The gravitational interactions between the two objects induce Kozai oscillations which increases the inclination and decreases the eccentricity for one object while the other object's eccentricity increases and its inclination decreases. Eccentricity and inclination are therefore anti-correlated during Kozai cycles and are described by the Kozai integral [31] I_K

$$I_K = \sqrt{1 - e^2} \cos i \quad (1)$$

which remains constant through this process. If the eccentricity of the planet increases to the point where its periastron distance is only a few stellar radii from its host star, it will raise tides on its host star's surface. This will cause the semi-major axis to shrink and orbit to circularize through tidal dissipation during periastron passages [32].

It has been found that $\sim 30\%$ of the observed Hot Jupiter population and up to 100% of the misaligned systems could be produced through this mechanism from modeling the dynamical effect of Kozai resonances on Jupiter-like planets from distant stellar companions [33]. It is also one of only two migration mechanisms (the other being secular chaos) which can naturally explain the 3-day orbital period pile-up observed for Hot Jupiters [13]. If indeed the Kozai mechanism were responsible for the spin-orbit misaligned systems, one would expect to find stellar or massive substellar companions to most if not all of them. Additionally, it has been found that the formation of Hot Jupiters through the Kozai mechanism is significantly suppressed for close binaries compare to the efficiency for more distant stellar companions (> 500 AU) [33]. This conclusion is also in agreement with the observational evidence which suggests that Hot Jupiter systems in close binaries (< 100 AU) are less common [34]. Therefore, searching for distant stellar companions, as described in the next section, will provide an important test of this mechanism.

A Test for the Kozai Mechanism: Searching for Stellar Companions

Direct imaging surveys of systems with measured spin-orbit alignments can provide a critical test of the hypothesis that misalignments are driven by the Kozai mechanism. With

current technology, it is feasible to survey a substantial sample of nearby stars ($d \leq 250$ pc) for the presence of stellar and substellar companions at separations as small as ~ 100 AU. If companions are found preferentially in systems with misaligned Hot Jupiters (as opposed to systems with aligned Hot Jupiters), then the Kozai hypothesis will have substantial support as the dominant driver of these misalignments.

Companions can be sought using two obvious techniques. One way is through their radial velocity impacts on the host star, similar to radial velocity searches for exoplanets. A caveat of this method is that the timescales required to detect companions orbiting beyond 5 AU is greater than 10 yr. This is not very feasible for finding long period companions in a reasonable time frame. Alternatively, the method we are proposing is to search for companions through direct imaging surveys. With the exception of the closest stars, companions at separations of less than ~ 100 AU will be difficult to detect due to the achievable contrast being insufficient for detecting faint companions. Despite this limitation, it is thought that stellar companions are not likely to reside in orbits less than 100 – 500 AU in systems with giant planets ([33]; [34]) as stellar companions can significantly retard planetesimal accretion ([35]; [36]). However, other studies have shown that perturbations from moderately close-in massive companions are usually not strong enough to fully halt accretion ([37]; [38]). Therefore giant planets are expected to be less frequent (but not completely absent) around binaries with separations of less than 100 AU [36]. If systems with Hot Jupiters do host stellar or substellar mass companions, they should lie far enough away from the host star to be detectable. In this way, a survey such as the one we are pursuing will determine whether stellar companions are the cause of the observed spin-orbit misalignment of Hot Jupiter systems and in turn the validity of the Kozai mechanism.

The contrasts required to detect companions with masses as low as mid M-dwarf and with separations as small as $0.5''$ around solar type stars (i.e. G stars) is $> 10^2$ in the Ks-band ($\sim 2000 - 2400$ nm). Even higher contrasts ($> 10^3$) will be necessary to detect fainter companions (such as late M-dwarf or early L-dwarfs) around solar type stars and to detect companions around more massive and brighter F-type stars. These requirements are readily achievable with current technology. Therefore our group is pursuing a survey of nearby stars ($d \leq 250$ pc) with measured spin-orbit alignments in the southern hemisphere using the Magellan Adaptive Optics (MagAO) and Clio2 infrared camera instruments on the 6.5 m Magellan Telescope at the Las Campanas Observatory in Chile. The MagAO system can achieve contrasts of $\sim 10^4$ ($\Delta Ks = 10$) at $0.5''$ and $\sim 10^5$ ($\Delta Ks = 12.5$) at $1.0''$ (private communication L. Close) in the Ks-band (better than other similar surveys [35], [39], [40]), that will conclusively confirm or reject the presence of stellar mass companions at separations greater than 150 AU in our sample.

Other groups are also searching for stellar companions around nearby stars. One such group is the VLT/NACO Search for Stellar Companions to 130 Nearby Stars with and Without Planets survey [35]. They are using the Nasmyth Adaptive Optics System (NAOS) with the Near-Infrared Imager and Spectrograph (CONICA) instruments on the 8.2 m Very Large Telescope (VLT) in Cerro Paranal, Chile to search for companions around southern hemisphere stars and the PUEO adaptive optics imager on the 3.6 m Canada-France-Hawaii Telescope (CFHT) on top of Mauna Kea, Hawaii to search for companions around northern hemisphere stars. Their survey is capable of detecting stellar companions down to M5 dwarfs at $0.2''$ and early L-dwarfs (brown dwarfs) at $0.2''$. Other groups are using the Lucky Imaging technique to search for companions in the SDSS i' band with the LuckyCam camera on the 2.56 m Nordic Optical Telescope [39] and in the z' band with the AstraLux Norte imaging instrument

on the Calar Alto 2.2 m telescope and the AstraLux Sur imaging instrument on the ESO 3.5 m New Technology Telescope at La Silla, Chile [40].

Compared to other surveys, ours is unique in that it is the only one that is specifically targeting systems with measured spin-orbit alignments and is directly testing the Kozai mechanism for producing the observed misaligned Hot Jupiters. We also have a key advantage over other surveys as we will be able to achieve higher contrast that will enable us to detect fainter companions.

Conclusions

We have presented results revealing that the transiting Hot Jupiter WASP-79b is in significant spin-orbit misalignment. We find that the projected angle between the spin-axis of the host star and the orbital plane of WASP-79b is $\lambda = -106^{+19}_{-13}^\circ$, making its orbit nearly polar. WASP-79b joins the growing population of Hot Jupiters that have been found to exhibit significant spin-orbit misalignments. Several mechanisms have been proposed for producing Hot Jupiters with high obliquities. These include disk migration [9], Kozai resonances [12], secular chaos [13], planet-planet scattering [14], primordial misalignments of proto-planetary disks [26], and internal gravity wave modulation of the stellar surface of host stars [24]. The jury is still out on which mechanism(s) are responsible for producing spin-orbit misalignments. An expansion of the sample of systems with measured obliquities, in particular multiple planet systems, will allow astronomers to determine the dominant mechanisms that produce such misalignments. Systems suitable for Rossiter-McLaughlin effect follow-up observations will come from globally distributed ground-based transit searches such as HATSouth [41], the recently announced Kepler space telescope K2 mission [42], and new space-based all-sky transit surveys like the Transiting Exoplanet Survey Satellite (TESS; [43]). Finally, we propose a test for the hypothesis that the majority of misaligned Hot Jupiter systems were produced by Kozai resonant behavior resulting from perturbations by as-yet undiscovered binary companions to the planet host stars.

References

- [1] Perryman, M., “The Exoplanet Handbook”, Michael Perryman, *Cambridge University Press*, 1st edition, ISBN: 0521765595, 2011.
- [2] Rossiter, R. A., “On the detection of an effect of rotation during eclipse in the velocity of the brighter component of beta Lyrae, and on the constancy of velocity of this system”, *The Astrophysical Journal*, Vol. 60, 1924, pp. 15-21.
- [3] McLaughlin, D. B., “Some results of a spectrographic study of the Algol system”, *The Astrophysical Journal*, Vol. 60, 1924, pp. 22-31.
- [4] Queloz, D., Eggenberger, A., Mayor, M., Perrier, C., Beuzit, J. L., Naef, D., Sivan, J. P., and Udry, S., “Detection of a spectroscopic transit by the planet orbiting the star HD209458”, *Astronomy and Astrophysics*, Vol. 359, 2000, pp. L13-L17.
- [5] Addison, B. C., Tinney, C. G., Wright, D. J., Bayliss, D., Zhou, G., Hartman, J. D., and Bakos, G. Á., and Schmidt, B., “A Nearly Polar Orbit for the Extrasolar Hot Jupiter WASP-79b”, *The Astrophysical Journal Letters*, Vol. 774, 2013, pp. L9-L14.

- [6] Ohta, Y., Taruya, A., and Suto, Y., “The Rossiter-McLaughlin Effect and Analytic Radial Velocity Curves for Transiting Extrasolar Planetary Systems”, *The Astrophysical Journal*, Vol. 622, 2005, pp. 1118-1135.
- [7] Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., and Greenzweig, Y., “Formation of the Giant Planets by Concurrent Accretion of Solids and Gas”, *Icarus*, Vol. 124, 1996, pp. 62-85.
- [8] Alibert, Y., Mordasini, C., Benz, W., and Winisdoerffer, C., “Models of giant planet formation with migration and disc evolution”, *Astronomy and Astrophysics*, Vol. 434, 2005, pp. 343-353.
- [9] Bate, M. R., Lodato, G., and Pringle, J. E., “Chaotic star formation and the alignment of stellar rotation with disc and planetary orbital axes”, *Monthly Notices of the Royal Astronomical Society*, Vol. 401, 2010, pp. 1505-1513.
- [10] Albrecht, S., Winn, J. N., Johnson, J. A., Howard, A. W., Marcy, G. W., Butler, R. P., Arriagada, P., Crane, J. D., Shectman, S. A., Thompson, I. B., Hirano, T., Bakos, G., and Hartman, J. D., “Obliquities of Hot Jupiter Host Stars: Evidence for Tidal Interactions and Primordial Misalignments”, *The Astrophysical Journal*, Vol. 757, 2012b, pp. 18-43.
- [11] Winn, J. N., Noyes, R. W., Holman, M. J., Charbonneau, D., Ohta, Y., Taruya, A., Suto, Y., Narita, N., Turner, E. L., Johnson, J. A., Marcy, G. W., Butler, R. P., and Vogt, S. S., “Measurement of Spin-Orbit Alignment in an Extrasolar Planetary System”, *The Astrophysical Journal*, Vol. 631, 2005, pp. 1215-1226.
- [12] Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A., and Teyssandier, J., “Hot Jupiters from secular planet-planet interactions”, *Nature*, Vol. 473, 2011, pp. 187-189.
- [13] Wu, Y. and Lithwick, Y., “Secular Chaos and the Production of Hot Jupiters”, *The Astrophysical Journal*, Vol. 735, 2011, pp. 109-121.
- [14] Chatterjee, S., Ford, E., Matsumura, S., and Rasio, F., “Dynamical Outcomes of Planet-Planet Scattering”, *The Astrophysical Journal*, Vol. 686, 2008, pp. 580-602.
- [15] Smalley, B., Anderson, D. R., Collier-Cameron, A., Doyle, A. P., Fumel, A., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Udry, S., and West, R. G., “WASP-78b and WASP-79b: two highly-bloated hot Jupiter-mass exoplanets orbiting F-type stars in Eridanus”, *Astronomy and Astrophysics*, Vol. 547, 2012, pp. 61-68.
- [16] Hirano, T., Suto, Y., Taruya, A., Narita, N., Sato, B., Johnson, J. A., and Winn, J. N., “Analytic Description of the Rossiter-McLaughlin Effect for Transiting Exoplanets: Cross-Correlation Method and Comparison with Simulated Data”, *The Astrophysical Journal*, Vol. 709, 2010, pp. 458-469.
- [17] Schaller, G., Schaerer, D., Meynet, G., and Maeder, A., “New grids of stellar models from 0.8 to 120 solar masses at $Z = 0.020$ and $Z = 0.001$ ”, *Astronomy and Astrophysics Supplement Series*, Vol. 96, 1992, pp. 269-331.
- [18] Winn, J. N., Fabrycky, D., Albrecht, S., and Johnson, J. A., “Hot Stars with Hot Jupiters Have High Obliquities”, *The Astrophysical Journal Letters*, Vol. 718, 2010a, pp. L145-L149.
- [19] Albrecht, S., Winn, J. N., Marcy, G. W., Howard, A. W., Isaacson, H., and Johnson, J. A., “Low Stellar Obliquities in Compact Multiplanet Systems”, *The Astrophysical Journal*, Vol. 771, 2013, pp. 11-25.
- [20] Huber, D., Carter, J. A., Barbieri, M., Miglio, A., Deck, K. M., Fabrycky, D. C., Montet, B. T., Buchhave, L. A., Chaplin, W. J., Hekker, S., Montalbán, J., Sanchis-Ojeda, R., Basu, S., Bedding, T. R., Campante, T. L., Christensen-Dalsgaard, J., Elsworth, Y. P., Stello, D., Arentoft, T., Ford, E. B., Gilliland, R. L., Handberg, R., Howard, A. W., Isaacson, H., Johnson, J. A., Karoff, C., Kawaler, S. D., Kjeldsen, H., Latham, D. W., Lund, M. N.,

- Lundkvist, M., Marcy, G. W., Metcalfe, T. S., Silva Aguirre, V., and Winn, J. N., “Stellar Spin-Orbit Misalignment in a Multiplanet System”, *Science*, Vol. 342, 2013, pp. 331-334.
- [21] Nagasawa, M., Ida, S., and Bessho, T., “Formation of Hot Planets by a Combination of Planet Scattering, Tidal Circularization, and the Kozai Mechanism”, *The Astrophysical Journal*, Vol. 678, 2008, pp. 498-508.
- [22] Lin, D. N. C., Bodenheimer, P., and Richardson, D. C., “Orbital migration of the planetary companion of 51 Pegasi to its present location”, *Nature*, Vol. 380, 1996, pp. 606-607.
- [23] Ward, W. R., “Protoplanet Migration by Nebula Tides”, *Icarus*, Vol. 126, 1997, pp. 261-281.
- [24] Rogers, T. M., Lin, D. N. C., and Lau, H. H. B., “Internal Gravity Waves Modulate the Apparent Misalignment of Exoplanets around Hot Stars”, *The Astrophysical Journal Letters*, Vol. 758, 2012, pp. 6-11.
- [25] Rogers, T. M., Lin, D. N. C., McElwaine, J. N., and Lau, H. H. B., “Internal Gravity Waves in Massive Stars: Angular Momentum Transport”, *The Astrophysical Journal*, Vol. 772, 2013, pp. 21-40.
- [26] Batygin, K., “A primordial origin for misalignments between stellar spin axes and planetary orbits”, *Nature*, Vol. 491, 2012, pp. 418-420.
- [27] Raymond, S. N., Armitage, P. J., and Gorelick, N., “Planet-Planet Scattering in Planetesimal Disks”, *The Astrophysical Journal Letters*, Vol. 699, 2009, pp. L88-L92.
- [28] Levison, H. F., Morbidelli, A., Van Laerhoven, C., Gomes, R., and Tsiganis, K., “Origin of the structure of the Kuiper belt during a dynamical instability in the orbits of Uranus and Neptune”, *Icarus*, Vol. 196, 2008, pp. 258-273.
- [29] Lithwick, Y. and Wu, Y., “Secular chaos and its application to Mercury, hot Jupiters, and the organization of planetary systems”, 2013, arXiv:1311.1214.
- [30] Horner, J., Evans, N. W., and Bailey, M. E., “Simulations of the population of Centaurs - II. Individual objects”, *Monthly Notices of the Royal Astronomical Society*, Vol. 355, 2004, pp. 321-329.
- [31] Murray, C. D. and Dermott, S. F., “Solar system dynamics”, *Cambridge University Press, Cambridge*, 1999, ch. 7.
- [32] Wu, Y., Murray, N. W., and Ramsahai, J. M., “Hot Jupiters in Binary Star Systems”, *The Astrophysical Journal*, Vol. 670, 2007, pp. 820-825.
- [33] Naoz, S., Farr, W. M., and Rasio, F. A., “On the Formation of Hot Jupiters in Stellar Binaries”, *The Astrophysical Journal Letters*, Vol. 754, 2012, pp. L36-L42.
- [34] Eggenberger, A., Udry, S., Chauvin, G., Forveille, T., Beuzit, J., Lagrange, A., and Mayor, M., “Probing the impact of stellar duplicity on the frequency of giant planets: Final results of our VLT/NACO survey”, *The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution, Proceedings of the International Astronomical Union, IAU Symposium*, Vol. 276, 2011, pp. 409-410.
- [35] Eggenberger, A., and Udry, S., “Probing the Impact of Stellar Duplicity on Planet Occurrence with Spectroscopic and Imaging Observations”, *Planets in Binary Star Systems, Astrophysics and Space Science Library*, Vol. 366, 2010, pp. 19-49.
- [36] Eggenberger, A., Udry, S., Chauvin, G., Beuzit, J. L., Lagrange, A. M., and Mayor, M., “Probing the Impact of Stellar Duplicity on Planet Occurrence”, *Extreme Solar Systems, ASP Conference Series*, Vol. 398, 2008, pp. 179-186.
- [37] Marzari, F., Thebault, P., Kortenkamp, S., and Scholl, H., “Dynamics and planet formation in/around binaries”, 2007, arXiv:0705.3113.
- [38] Boss, A. P., “Gas giant protoplanets formed by disk instability in binary star systems”, *The Astrophysical Journal*, Vol. 641, 2006, pp. 1148-1161.

- [39] Faedi, F., Staley, T., Gómez Maqueo Chew, Y., Pollacco, D., Dhital, S., Barros, S. C. C., Skillen, I., Hebb, L., Mackay, C., and Watson, C. A., “Lucky imaging of transiting planet host stars with LuckyCam”, *Monthly Notices of the Royal Astronomical Society*, Vol. 433, 2013, pp. 2097-2106.
- [40] Bergfors, C., Brandner, W., Daemgen, S., Biller, B., Hippler, S., Janson, M., Kudryavtseva, N., Geißler, K., Henning, T., and Köhler, R., “Stellar companions to exoplanet host stars: Lucky Imaging of transiting planet hosts”, *Monthly Notices of the Royal Astronomical Society*, Vol. 428, 2013, pp. 182-189.
- [41] Bakos, G. Á., Csubry, Z., Penev, K., Bayliss, D., Jordán, A., Afonso, C., Hartman, J. D., Henning, T., Kovács, G., Noyes, R. W., Béky, B., Suc, V., Csák, B., Rabus, M., Lázár, J., Papp, I., Sári, P., Conroy, P., Zhou, G., Sackett, P. D., Schmidt, B., Mancini, L., Sasselov, D. D., and Ueltzhoeffer, K., “HATSouth: A Global Network of Fully Automated Identical Wide-Field Telescopes”, *Publications of the Astronomical Society of the Pacific*, Vol. 125, 2013, pp. 154-182.
- [42] Beichman, C., Ciardi, D., Akeson, R., Plavchan, P., Howell, S., Christiansen, J., Kane, S., Cody, A., Stauffer, J., Vasisht, G., and Covey, K., “New Uses for the Kepler Telescope: A Survey of the Ecliptic Plane For Transiting Planets and Star Formation”, 2013, arXiv:1309.0918.
- [43] Deming, D., Seager, S., Winn, J., Miller-Ricci, E., Clampin, M., Lindler, D., Greene, T., Charbonneau, D., Laughlin, G., Ricker, G., Latham, D., and Ennico, K., “Discovery and Characterization of Transiting Super Earths Using an All-Sky Transit Survey and Follow-up by the James Webb Space Telescope”, *Publications of the Astronomical Society of the Pacific*, Vol. 121, 2009, pp. 952-967.

Searching for Life on Early Mars via Sample Return: Lessons from the Pilbara

Jonathan D. A. Clarke^{1 2}

¹ *Mars Society Australia, 43 Michel St Monash, ACT 2904, Australia*

² *Australian Centre for Astrobiology, University of New South Wales, NSW 2052, Australia*

Summary:

The evidence for terrestrial life in ~3.5 billion year old Pilbara rocks provide insight into the likely scientific success of proposed robotic Mars sample return (MSR). It took 80 years of geological exploration of the Pilbara for evidence for early life to be discovered, and over 30 years for this to be generally accepted. A vast amount of contextual and site specific data was collected by many expeditions hundreds to thousands of km of traverses, and the collection of thousands of samples probably massing tonnes from scores to hundreds of sites was needed. By analogy the likelihood of MSR involving the return of ~38 samples, massing ~500 grams and collected over <20 km of traverse providing conclusive answers is therefore miniscule. Proposed MSR missions might be better focused on providing risk-minimising data on the martian environment preparatory to crewed missions that would out-perform robotic-only missions by several orders of magnitude.

Keywords: Astrobiology, Mars sample return, Pilbara, Archean palaeobiology

Introduction

Mars Sample Return (MSR) by unmanned missions was identified as a major goal in the exploration of Mars soon after the Viking mission [1]. More recently, the US National Research Council's (NRC) Planetary Science Decadal Survey (PSDS) for the period 2013-22 [2] identified MSR as the highest priority science goal for planetary exploration. These conclusions underlie the rationale for NASA's proposed 2020 rover mission [3] which is intended to cache sample for future collection and return to Earth.

As defined by the Mars Exploration Program Analysis Group (MEPAG) in 2011 [4] the goals of MSR are:

- Search for evidence of past life or its precursors,
- Understanding the nature of the martian surface and its materials, and of planetary and atmospheric evolution.
- Clarifying issues to do with human exploration, identifying both hazard and resource potential.

The issues of extant life on Mars or those associated with returning it to Earth are not directly addressed by either MEPAG or the PSDS, beyond stressing the need for the highest quarantine standards [5]. However, as early Mars was much more habitable than present Mars

a strategy based on searching for past life is more likely to be successful than searching for extant life.

In this paper I will compare the expectations from currently proposed MSR missions with the history of exploration of geological successions analogous in age and formational environment in the Pilbara region of Western Australia. I will argue that the history of study of life signs in 3.5 billion year old (Ga) rocks of the Pilbara provide an indication of the likelihood of finding definitive evidence of past life on Mars in the proposed suit of returned samples.

Mars sample return methodologies

Mars Sample return mission concepts vary greatly in complexity, some being comparatively simple, consisting of atmospheric dust and gas collected during a high altitude flyby (e.g. SCIM, Jones et al. 2008). However most concepts are highly complex, involving multiple assets deployed on the martian surface and in Mars orbit. For example the most recent detailed study [4] recommended a mission with one or more rovers, Mars ascent vehicles, and Mars return orbiters. The proposed 2020 rover mission [3] will, by collecting and caching some 37 short cores along a traverse of up to 20 km, with a total mass of about 500 grams, represent the first phase of MSR.

However, the MEPAG goals [4] could mostly be met by less complex missions, some not involving MSR. Almost any suite of diverse samples would improve understanding the nature of the martian surface and materials, and planetary and atmospheric evolution [1]. A single sample of regolith such as a drill core would clarify potential hazards to crews [6, 7]. The general history of Mars and its potential resources might be adequately studied using surface rovers rather than sample return, as recognised in the capabilities of both the present Curiosity mission [8] and the proposed 2020 rover [3]. The single greatest justification for MSR of the proposed complexity therefore becomes the search for evidence of past life or its precursors.

The Pilbara as Mars analogue

Significance of the Pilbara to Martian astrobiology

Experience in searching for life in a terrestrial analogue can provide insight into the likelihood of success of such missions. The Pilbara region of Western Australia [9, 10, 11] is one such region. Surface conditions on Earth and Mars were similar at 3.5 Ga during the Paleoarchean on Earth, and the Noachian/Phyllocian epochs on Mars [12, 13, 14]. Both planets appear to have possessed a global magnetic field, dense and oxygen-poor atmospheres, hydrospheres with active hydrologic cycles, active volcanism, hydrothermal activity, and diverse surface, and near-surface environments within the tolerable range of many microorganisms. , et al. (2010) have reported a similar talc-carbonate hydrothermal alteration regime at Nili Fossae, similar to type examples to be found at the North Pole Dome and throughout the Eastern Pilbara [15, 16]. The Nili Fossae carbonate bearing region is a top candidate for the Mars 2020 rover' As a result surface lithologies and surface mineralogies of the Paleoarchean of the Pilbara and the Noachian/Phyllocian of Mars are very similar. Conditions extant of the surface of both planets somewhat earlier are likewise predicted to have been similar and suitable for the emergence of life [17].

The Pilbara successions also contain evidence for the presence of microbial life, most obviously at the mesoscopic scale in the form of stromatolites (Figure 1). These were first recognised by Walter, Buick and Lowe [18] and are accompanied by other evidence, including microfossils (Figure 2), first convincingly reported by Awramik *et al.* in 1983 [19]. Given the similarity in environments it is conceivable that coeval conditions on Mars may also have hosted life. The Pilbara therefore is not only one of few terrestrial regions that similar environmental conditions to early Mars, but is of similar antiquity. It therefore provides an analogue to assess the habitability of potential environments on early Mars, a region for developing and testing life search strategies for that planet.



Fig. 1: Conical stromatolites the Strelley Pool Formation, Dawn of Life Trail site [20]), Pilbara. These examples date from about 3.4 Ga.

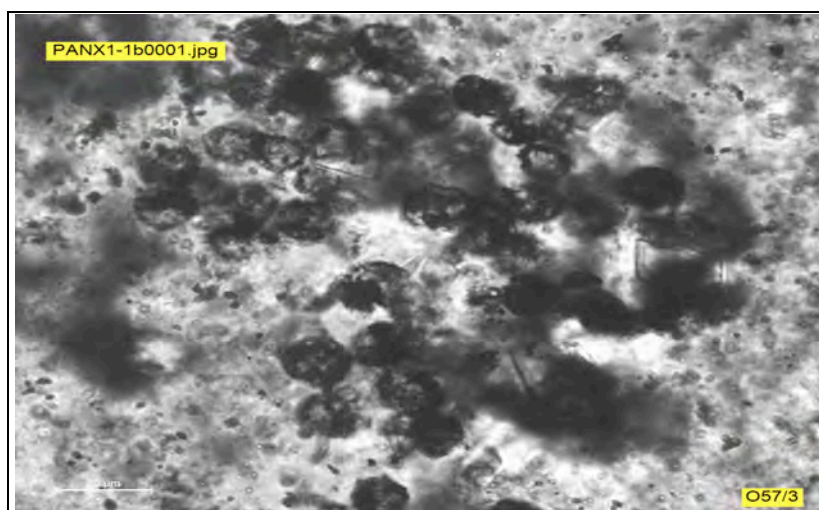


Fig. 2: Microfossils from the ~3.5 Ga Strelley Pool Formation in the Pilbara (near Trendall location, see [20]). Image from [21], used with permission of the author.

History of Palaeobiological investigations in the Pilbara

At the time of writing evidence for life had been reported from multiple horizons within the Paleoproterozoic succession of the Pilbara (Table 1). These were formed in between 3.49 and 3.315 Ga [10]. The discovery of this evidence occurred following a long history of

exploration in the region [22]. European exploration the inland Pilbara began with the expedition of Gregory in 1861. Gold was discovered in 1888 at Mallina, and lead to the first geological investigations, published by Maitland in 1908 [22].

Table 1: Distribution of biogenic indicators in the Paleoarchean of the Pilbara (modified from [10].

Rock Unit	Rock type	Age from U-Pb geochronology ¹	Biogenic indicators
Top Wyman Formation	Felsic volcanics & chert	3.325-3.315 Ga	Domical stromatolites (in chert)
Euro Basalt	Basalt	3.35 Ga	Micro-borings
Strelley Pool Formation	Chert, carbonate, & volcanic sediments		Conical and domical stromatolites
Panorama Formation	Felsic volcanic		
Apex Basalt	Basalt and chert		Microfossils (in chert)
Marble Bar Chert	Chert		
Duffer Formation	Felsic volcanic	3.465 Ga	
Mount Ada Basalt	Basalt	3.468 Ga	
Dresser Formation	Chert	~3.49 Ga (Pb-Pb age)	Domical and conical stromatolites, microfossils, carbon isotope signatures
North Star Basalt	Basalt		

The decades of geological research in support of mineral exploration led to the recognition that much of the succession in the region consists of volcanic rocks (Figure 3) with potentially fossiliferous carbonate and cherty rocks only a small part of the succession [11, 9]. Microfossils found in silicified carbonates were the first convincingly reported for Paleoarchean life in the Pilbara [19]. Subsequent studies of microfossils include those of Schopf [23], Oehler *et al.* [24], Sugitani *et al.* [21, 25]. Stromatolites were discovered in the Pilbara in 1980 [18, 26], even though elsewhere they had been named and their significance recognised by Kalkowsky in 1908 [27]. Previously named *Cryptozoon* by Walcott in 1883, they were first characterised in 1925 by Steele and recognised as marine plant remains by Hall in 1847.

Much studied since their recognition by Walter *et al.* and Lowe [18, 26], recent analysis of their characteristics and their significance can be found it are those of Allwood *et al.* [28, 29]. Microborings in the rims of well-preserved pillow lavas from the Euro Basalt are texturally identical with those formed by microbes in modern basalts and ophiolite complexes [30]. These have been dated as Archean and show that microbial colonisation of basaltic glass was established by the Paleoarchean. The composition of reduced carbon compounds associated with the microfossils is difficult, because of the low organic carbon content of these rocks probability of later contamination. A recent study is that of Marshall *et al.* [31]. These authors used a range of advanced micro-analytical techniques to demonstrate the biogenic nature of kerogens in the Strelley Pool Formation. Lastly, stable isotopes of both the organic and carbon components have been studied by many authors, with Ueno *et al.* [32] being but one example. Other studied stable isotopes include those of sulphur e.g. [33] nitrogen [34].



Fig. 3: Pillow basalts of the 4.468 Ga Mt Ada Basalt, near the Trendall locality, North Pole, Pilbara.

Given the significance of finding evidence for life in such ancient rocks, the inevitable sceptical responses soon appeared. The authenticity of the stromatolites was questioned by Buick *et al.* [35] and Lowe [36] even though previously Lowe [26] had accepted them as genuine. The putative microfossils were first dismissed by Buick [37] as pseudofossils, formed during metamorphism, a theme developed by Brasier *et al.* [38, 39], Pinti *et al.* [40], and Marshall *et al.* [41]. The validity of the organic data was again questioned by Buick [37] and Lindsay *et al.* [42]. Brasier *et al.* [39] also doubted the significance of the fractionation of light carbon isotopes, citing evidence that abiogenic processes could cause similar signatures.

A characteristic of this discussion has been the switching of positions by researchers. As already noted, Lowe [26] argued for microfossils but against them in his 1984 paper [37]. Brasier (in Brasier *et al.* [38, 39]) argued strongly against the detection of microfossils, but was a contributor to the Wacey *et al.* [43] paper than argued that microfossils were present in the Strelley Pool Formation. Buick was a contributing author in Walter *et al.* [18], in the first publication of Paleoarchean Pilbara stromatolites, but in Buick *et al.* [37] concluded that none were confirmed as stromatolites. Van Kranendonk contributed to sceptical paper of Brasier *et al.* [38] but in Van Kranendonk [9, 10] and as a contributor to Sugianti *et al.* [25] was arguing they were genuine.

Current research has tended to demonstrate that evidence for Paleoarchean life does indeed occur in the ~3.5 Ga rocks of the Pilbara. Allwood *et al.* [28, 29] with strong evidence and robust reasoning demonstrated that the stromatolites were indeed biogenic. The paper by Wacey *et al.* [43, 44] whose one of whose contributing authors was the formerly sceptical Brasier, showed that the latest NanoSIMS data and the balance of the evidence pointed to microfossils indeed being present, all although pseudofossils certainly abound. The present of genuine biogenic organic molecules and isotope signatures was argued by Marshall [45], again using relatively new techniques such as Fourier transform infrared spectroscopy, Raman spectroscopy, nuclear magnetic resonance spectroscopy, catalytic hydrolysis and gas chromatography mass spectrometry, and isotope ratio mass spectrometry of samples containing kerogen. At the present state of knowledge therefore it would appear that life was indeed present at 3.5 Ga in the Pilbara, leaving a diverse imprint on the geological record, from the mesoscale (stromatolites) as well as the microscale (microfossils, micro-borings),

and in the organic and stable isotope chemistry. This imprint reflects the presence of diverse microbial communities in the surface environment [46].

Pilbara Implications for MSR

Historical lessons from the Pilbara

However, the history of research into the evidence for early terrestrial life in the Pilbara suggests that demonstrating the presence or, equally important, the absence, of life on early Mars beyond reasonable doubt will be challenging. The relevant strata in the Pilbara are thin, of limited lateral extent, and widely scattered, as such strata are likely to be on Mars, if present at all [47]. It took ninety years of exploration before the first sites were discovered, even though the palaeobiological significance of stromatolites have been known since their first description being by Steel in 1825, although they were not named as such until Kalkowsky in 1908 [48].

Subsequently, more than thirty years of detailed investigation by scores of researchers from many institutions across the world for the presence of biogenic features and signatures to be generally accepted [43]. Together these efforts involved many years of cumulative field work, hundreds to thousands of km of traverses, and the estimated collection of thousands of samples probably massing tonnes from scores to hundreds of sites.

Challenges of sample return strategies

Assuming such features are found on Mars what would be required for the investigations to achieve a similar level of investigative confidence we have in the Pilbara? As observed by Clarke and Stoker [46], the first challenge would be reconnaissance, locating potential targets as efficiently as possible. This would minimise detailed surface reconnaissance that might require several thousand km of traverses. Hyperspectral mapping from orbit of regions of interest at the metre scale would be an ideal method, able to identify targets such as chemical sediments, mudrocks, and hydrothermal systems similar to those found in the Pilbara [49]. Such datasets are not yet available.

The second challenge would be surface scouting and documentation of the targets. This will need precision landing, probably to within 10 km of specified landing site, to minimise travel. Even so, vehicles capable of traverses of the order of hundred km in a reasonable time frame, such the 500-600 sols (martian days) between Mars landing and departure windows, or at least within a single Mars year (669 sols), will be needed to scout as many targets as possible. The targets will need to be characterised with mesoscopic and microscopic imagery, for chemistry, mineralogy, and organic using short-range remote sensing and contact data. Such capabilities are well beyond what is likely with unmanned exploration. A comparison between the planned performance of the 2020 sample collection rover and what has been actually achieved by Mars rovers to date illustrates how unrealistic performance expectations are for this mission, given its' Curiosity rover derivation (Table 2).

Thirdly, there is the sampling challenge. This might be integrated with the scout phase or could be a separate mission. A diversity of the most promising targets will need to be visited and sampled to increase likelihood of success. Each target may contain several sites of interest. Lithological diversity, including chemical sediments, mudrocks, surficial materials, and hydrothermal deposits, need to be represented in the samples. The collected material should consist of a range of sampling methods including outcrop, core, and surficial materials.

Significant total sample mass will be needed to ensure repeat analyses, alternate methods, and multiple access beyond the initial science team and investigative period. Only through such in depth investigation are conclusive results likely to be reached. This requirement indicates that sample masses in excess of 100 kg from as many widely dispersed sites as possible will be required. This is beyond the capability of any envisaged MSR mission.

Table 2: Mars rover performance, actual and predicted. Note the disparity between actual rover performance (Spirit, Opportunity, and Curiosity) and that proposed for the 2020 rover, suggesting that the 2020 Rover's objectives are unreasonably optimistic.

Mission	Sites sampled	Travel (km)	Operational (sols)	Sols/sample site	metres/sol (m)
Spirit	-----	7.73	2155	-----	3.58
Opportunity ¹	-----	38.53	3487	-----	1.11
Curiosity ¹	3	4.15	454	151	9.14
2020 Rover ²	37	20	669	18	29.11

¹As of 14/11/13

²Ref. [3]

Fourthly, there is retrieval and return of samples to Earth. This again would utilise precision landing to increase mission safety, and would also need the ability of the sample-carrying vehicles to drive significant distances (~10-20 km) in a short period of time to reach the sample return craft within the Earth return window.

Lastly, the samples should be appropriately archived and analysed by a wide range of international groups with access to suitable facilities. In addition to answering with high confidence the question as to whether or not life existed on early Mars the returned samples will provide considerable insight into volcanic, hydrothermal and sedimentary processes then operating and the nature of the early martian hydrosphere and atmosphere. However this level of investigative diversity cannot be achieved with the small sample masses envisaged by current MSR proposals.

How can these goals be best met?

These challenges are orders of magnitude beyond what could be realistically met by an unmanned sample return mission, typified by Ref [2, 3] and demonstrated by the Curiosity mission. Therefore the ability of the proposed level of MSR missions to demonstrate the presence or absence of life on early Mars beyond reasonable doubt appears unlikely. The number, spatial spread and size of samples would be too small to show to give strong negative results and the likelihood of finding samples with incontrovertible evidence for past life in them, such outcrops of stromatolitic chert rich in organic microfossils, is astronomically small.

These goals could, however be achieved by even a relatively modest crewed mission, for example the Mars Oz reference mission [50]. As shown by Crawford [51], crewed missions are more cost effective and offer far greater scientific return, by two to three orders of magnitude, than any conceivable rover or MSR mission or combination of missions. The returns from larger human missions proposed by NASA [52] are likely to be even more substantial. This is despite the encumbrance of field science while wearing a space suit [53].

Therefore crewed missions are far more likely to answer the key astrobiological questions than either unmanned rovers or unmanned sample return.

Conclusions

The conclusions from this overview are that:

- Paleoproterozoic rocks of the Pilbara provide useful analogues to possible surface environments on Noachian/Phyllochaian Mars.
- The evidence of life in the Paleoproterozoic of the Pilbara provide pointers as to how evidence for equivalent life in Noachian/Phyllochaian rocks on Mars might be expressed, should it exist.
- The Pilbara is also an ideal location to test exploration concepts and techniques applicable for the search of life on early Mars.
- The history of studies of evidence of life in the Paleoproterozoic of the Pilbara provides pointers to the likelihood of finding evidence of such life on Mars, had it been present.
- These studies show that to obtain incontrovertible evidence of life in the Paleoproterozoic of the Pilbara required decades of field investigations, the collection of thousands of samples massing tonnes, taken from scores to hundreds of sites, and involving several thousands of km of travel.
- The results of the approximately 500 grams of sample spread over about 30 samples, collected over 20 km of traverse rather optimistically predicted by MSR studies are therefore unlikely to be definitive, whether positively or negatively.
- Only a crewed mission is likely to cover the distance and sample the required diversity of material and return it in sufficient quantities to make definitive results, e.g. on whether or not life was ever present on Mars, likely.

This should not be seen as argument against either unmanned MSR or the sophisticated robotic rovers, such as the 2020 proposal, that might support them. Both represent the next stage in robotic exploration. It is, however, a caution against unrealistic expectations of what such missions are likely to achieve, expectations that are more likely to be met by crewed exploration.

Pilbara experience may also indicate that unmanned MSR efforts might better be directed at simpler objectives, for example quantifying and reducing the risk to crewed missions posed by the martian regolith, as exploration by scientist-astronauts are far more likely to yield definitive results, be they positive or negative, than even an extensive campaign of unmanned MSR. A more basic MSR mission, perhaps a single drill to sample the regolith profile, might resemble a martian version of the Luna 16, 20 and 24 sample return missions (Ref. [54], chapter 7).

While such a MSR might yield fewer results than those currently proposed, the science return will still be significant, as shown for the Luna missions [51]. The missions will yield valuable science and provide an operational bridge between unmanned surface missions and those carried out by scientist-astronauts, in addition to the important risk assessment data that will

enable the vastly more capable crewed expeditions. Such an approach may also offer lower operational risk and lower cost than the complex MSR missions currently proposed.

Acknowledgements

I would like to thank K. Sugitani for permission to use the photomicrograph in Figure 2, M. Walker and A. Brown for their helpful reviews, and the proceedings editor W. Short for his advice and support.

References

1. Nickle, N. 1980. Mars sample return: site selection and sample acquisition study. JPL Publication 80-59.
2. Committee on the Planetary Science Decadal Survey 2011. Vision and Voyages for Planetary Science in the Decade 2013-2022. National Academies Press, Washington, 382p.
3. Mustard, J.F., M. Adler, A. Allwood, D.S. Bass, D.W. Beaty, J.F. Bell III, W.B. Brinckerhoff, M. Carr, D.J. Des Marais, B. Drake, K.S. Edgett, J. Eigenbrode, L.T. Elkins-Tanton, J.A. Grant, S. M. Milkovich, D. Ming, C. Moore, S. Murchie, T.C. Onstott, S.W. Ruff, M.A. Sephton, A. Steele, A. Treiman (2013): Report of the Mars 2020 Science Definition Team, 154 pp., posted July, 2013, by the Mars Exploration Program Analysis Group (MEPAG) at http://mepag.jpl.nasa.gov/reports/MEP/Mars_2020_SDT_Report_Final.pdf.
4. McLennan, S. M., M. A. Sephton, C. Allen, A.C. Allwood, R. Barbieri, D.W. Beaty, P. Boston, M. Carr, M. Grady, J. Grant, V. S. Heber, C. D. K. Herd, B. Hofmann, P. King, N. Mangold, G.G. Ori, A. P. Rossi, F. Raulin, S. W. Ruff, B. Sherwood Lollar, S. Symes, and M. G. Wilson (2011) Planning for Mars Returned Sample Science: Final report of the MSR End-to-End International Science Analysis Group (E2E-iSAG), 101 pp., posted December, 2011, by the Mars Exploration Program Analysis Group (MEPAG) at <http://mepag.jpl.nasa.gov/reports/>.
5. Farmer, J. D., Bell, J. F., Benison, K. C., Boynton, W. V., Cady, S. L., Ferris, F. G., Macpherson, D., Race, M. S., Thiemens, M. H., and Wadhwa, M. 2009. Assessment of Planetary Protection Requirements for Mars Sample Return Missions. National Academies Press, 80p.
6. Hauck, F. H., Mcsween, H. Y., Breazeal, C., Clark, B. C., Von Eshleman, R., Haas, J., Reid, J. B., Richmond, J., Turner, R. E., and Whittaker, W. L. 2002. Safe on Mars: Precursor measurements necessary to support human operations on the martian surface. National Academies Press, 64p.
7. Beaty, D.W., Snook, K., Allen, C.C., Eppler, D., Farrell, W.M., Heldmann, J., Metzger, P., Peach, L., Wagner, S.A., and Zeitlin, C., 2005. An Analysis of the Precursor Measurements of Mars Needed to Reduce the Risk of the First Human Missions to Mars. Unpublished white paper, 77 p, posted June, 2005 by the Mars Exploration Program Analysis Group (MEPAG) at <http://mepag.jpl.nasa.gov/reports/index.html>.

8. Grotzinger, J. P., Crisp, J., Vasavada, A. R., Anderson, R. C., Baker, C. J., Barry, R., Blake, D. F., Conrad, P., Edgett, K. S., Ferdowski, B., Gellert, R., Gilbert, J. B., Golombek, M., Javier Gómez-Elvira, J., Hassler, D. M., Jandura, L., Litvak, M., Mahaffy, P., Maki, J., Michael Meyer, M., Malin, M. C., Mitrofanov, I., Simmonds, J. J., Vaniman, D., Welch, R. V., and Wiens, R. C. 2012. Mars Science Laboratory Mission and Science Investigation. *Space Science Reviews* 170, 5–56.
9. Van Kranendonk, M. F. 2006. Volcanic degassing, hydrothermal circulation and the flourishing of early life on Earth: A review of the evidence from c. 3490-3240 Ma rocks of the Pilbara Supergroup, Pilbara Craton, Western Australia. *Earth-Science Reviews* 74, 197–240.
10. Van Kranendonk, M. J. 2007. A review of the evidence for putative Paleoarchean life in the Pilbara Craton, Western Australia. In Van Kranendonk, M. J., Smithies, H., and Bennett, V. C. (eds.) *Developments in Precambrian Geology* 15, pp. 855-877.
11. Van Kranendonk, M. J., Webb, G. E., and Kamber, B. S. 2003. Geological and trace element evidence for a marine sedimentary environment of deposition and biogenicity of 3.45 Ga stromatolitic carbonates in the Pilbara Craton, and support for a reducing Archaean ocean. *Geobiology* 1(2), 91–108.
12. McKay, C. P. and Stoker, C. R. 1989. The early environment and its evolution on Mars: implications for Mars. *Reviews of Geophysics* 27, 189-214.
13. Fairén, A. G., Davila, A. F., Lim, D., Bramall, N., Bonaccorsi, R., Zavaleta, J., Uceda, E. R., Stoker, C., Wierzchos, J., Dohm, J. M., Amils, R., Andersen, D., and McKay, C. P. 2010. Astrobiology through the Ages of Mars: The Study of Terrestrial Analogues to Understand the Habitability of Mars. *Astrobiology*. 10(8), 821-843.
14. Bibring, J.-P., Langevin, Y., Mustard, J. F., Poulet, F., Arvidson, R., Gendrin, A., Gondet, B., Mangold, N., Pinet, P., Forget, F., and the OMEGA team. 2006. Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. *Science* 312, 400–404.
15. Brown, A. J., Hook, S. J., Baldridge, A. M., Crowley, J. K., Bridges, N. T., Thomson, B. J., Marion, G. M., Filho, C. R. d. S. and Bishop, J. L. 2010, Hydrothermal formation of Clay-Carbonate alteration assemblages in the Nili Fossae region of Mars, *Earth and Planetary Science Letters*, 297, 174-182.
16. Brown, A. J., Cudahy, T. J. and Walter, M. R. 2006. Hydrothermal alteration at the Panorama Formation, North Pole Dome, Pilbara Craton, Western Australia. *Precambrian Research*. 151, 211-223.
17. Brack, A. and Pillinger, C. T. 1998. Life on Mars: chemical arguments and clues from Martian meteorites. *Extremophiles* 2, 313-319.
18. Walter, M.R., Buick, R., Dunlop, J.S.R., 1980. Stromatolites, 3400–3500 Myr old from the North Pole area, Western Australia. *Nature* 284, 443–445.
19. Awramik, S. M., Schopf, J. W. and Walter, M. R., 1983. Filamentous fossil bacteria from the Archean of Western Australia. *Precambrian Research*, v. 20, 357-374.

20. Grey, K., Clarke, J. D. A., and Hickman, A. H. 2012, The proposed Dawn of Life Geotourism Trail, Marble Bar, Pilbara Craton, Western Australia — geology and evidence for early life: Geological Survey of Western Australia, Record 2012/9, 27p.
21. Sugitani, K., Mimura, K., Nagaoka, T., Lepot, K., and Takeuchi, M. 2013. Microfossil assemblage from the 3400 Ma Strelley Pool Formation in the Pilbara Craton, Western Australia: Results from a new locality. *Precambrian Research* 226, 59– 74.
22. Noldart, A. J. and Wyatt, J. D. 1962. The geology of part of the Pilbara goldfield. *Western Australia Geological Survey Bulletin* 155, 199p.
23. Schopf, J. 1993. Microfossils of the early Archaean Apex Chert: New evidence of the antiquity of life. *Science* 260, 640–646.
24. Oehler, D. Z., Robert, F., Walter, M. R., Sugitani, K., Allwood, A., Meibom, A., Mostefaoui, S., Selo, M., Thomen, A., and Gibson, E. K. 2009. NanoSIMS: Insights to biogenicity and syngeneity of Archaean carbonaceous structures. *Precambrian Research* 173, 70–78.
25. Sugitani, K., Lepot, K., Nagaoka, T., Mimura, K., Van Kranendonk, M., Oehler, D. Z., and Walter, M. R. 2010. Biogenicity of Morphologically Diverse Carbonaceous Microstructures from the ca. 3400 Ma Strelley Pool Formation, in the Pilbara Craton, Western Australia. *Astrobiology* 10(9): 899-920.
26. Lowe, D. R. 1980. Stromatolites 3,400 Myr old from the Archaean of Western Australia. *Nature* 284, 441-443.
27. Friedman, G. M. 2000. Late Cambrian cabbage-head stromatolites from Saratoga Springs, New York, USA. *Carbonates and evaporites* 15(1), 39-48
28. Allwood, A., Walter, M., Burch, I., and Kamber, B. 2007. 3.43-Billion-year-old stromatolite reef from the Pilbara Craton of Western Australia: Ecosystem-scale insights to early life on Earth. *Precambrian Research* 158:198–227.
29. Allwood, A. C., Grotzinger, J. P., Knoll, A. H., Burch, I. W., Anderson, M. S., Coleman, M. L., and Kanik, I. 2009. Controls on development and diversity of Early Archean stromatolites. *Proceedings of National Academy of Sciences* 106(24), 9548-9555.
30. Banerjee, N. R., Simonetti, A., Furnes, H., Muehlenbachs, K., Staudigel, H., Heaman, L., Van Kranendonk, M. J. 2007. Direct dating of Archean microbial ichnofossils. *Geology* 35, 487-490.
31. Marshall, C. P. 2007. Organic geochemistry of Archaean carbonaceous cherts from the Pilbara Craton, Western Australia. In Van Kranendonk, M. J., Smithies, H., and Bennett, V. C. (eds.) *Developments in Precambrian Geology* 15, pp 897-921.
32. Ueno, Y., Isozaki, Y., Yurimoto, H., Maruyama, S., 2001. Carbon isotopic signatures of individual Archean microfossils (?) from Western Australia. *International Geology Review* 43, 196–212.

33. Shen, Y., Buick, R., Canfield, D.E., 2001. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* 410, 77–81.
34. Pinti, D. L., Hashizume, K. and Matsuda, J. 2001. Nitrogen and argon signatures in 3.8 to 2.8Ga metasediments: clues on the chemical state of the Archean ocean and the deep biosphere. *Geochimica et Cosmochimica Acta* 65, 2301–2315.
35. Buick, R., Dunlop, J. S. R., and Groves, D. I. 1981. Stromatolite recognition in ancient rocks: an appraisal of irregularly laminated structures in an Early Archaean chert-barite from North Pole, Western Australia. *Alcheringa* 5, 161-181.
36. Lowe, D.R., 1994. Abiological origin of described stromatolites older than 3.2 Ga. *Geology* 22(5), 387–390.
37. Buick, R., 1984. Carbonaceous filaments from North Pole, Western Australia: are they fossil bacteria in Archean stromatolites? *Precambrian Research* 24, 157– 172.
38. Brasier, M.D., Green, O.R., Jephcoat, A.P., Kleppe, A.K., Van Kranendonk, M.J., Lindsay, J.F., Steele, A., and Grassineau, N.V. 2002. Questioning the evidence for Earth's oldest fossils. *Nature* 416, 76–81.
39. Brasier, M. D., Green, O. R., Lindsay, J. F., McLoughlin, N., Steele, A., and Stoakes, C. 2005. Critical testing of Earth's oldest putative fossil assemblage from the ~3.5 Ga Apex chert, Chinaman Creek, Western Australia. *Precambrian Research* 140 (2005) 55–102.
40. Pinti, D. L., Mineau, R., and Clement, V. 2009. Hydrothermal alteration and microfossil artefacts of the 3,465-million-year-old Apex chert. *Nature Geoscience* 2, 640-643.
41. Marshall, C. P., Emry, J. R., and Marshall, A. O. 2011. Haematite pseudomicrofossils present in the 3.5-billion-year-old Apex Chert. *Nature Geoscience* 4, 240-243.
42. Lindsay, J.F., Brasier, M.D., McLaughlin, N., Green, O.R., Fogel, M., Steele, A., Mertzmann, S.A., 2005. The problem of deep carbon; an Archean paradox. *Precambrian Research* 143(1–4), 1–22.
43. Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J., and Brasier, M. D. 2011. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience* 4, 698-702.
44. Wacey, D., Kilburn, M. R., McLoughlin, N., Parnell, J., Stoakes, C. A., Grovenor, C. R. M. and Brasier, M. D. 2008. Use of NanoSIMS in the search for early life on Earth: ambient inclusion trails in a c. 3400 Ma sandstone. *Journal of the Geological Society, London* 165, 43–53.
45. Marshall, C. P. 2007. Organic geochemistry of Archaean carbonaceous cherts from the Pilbara Craton, Western Australia. In Van Kranendonk, M. J., Smithies, H., and Bennett, V. C. (eds.) *Developments in Precambrian Geology* 15, pp 897-921.
46. Noffke, N., Christian, D., Wacey, D. and Hazen, R, M. 2013. Microbially Induced Sedimentary Structures Recording an Ancient Ecosystem in the ca. 3.48 Billion-Year-Old Dresser Formation, Pilbara, Western Australia. *Astrobiology*. 13(12): 1103-1124

47. Clarke, J. D. A. and Stoker, C. R. 2013. Search for stromatolites: The 3.4 Ga Strelley Pool Formation (Pilbara region, Western Australia) as a Mars analogue *Icarus* 224(2), 413–423.
48. Riding, R. 1999. The term stromatolite: towards an essential definition. *Lethaia* 32, 321-330.
49. Brown, A. J., 2006. Spectral Curve Fitting for Automatic Hyperspectral Data Analysis. *IEEE Transactions on Geoscience and Remote Sensing*. 44, 1601-1608.
50. Willson, D. and Clarke, J. 2007. A Practical Architecture for Exploration-Focused Manned Mars Missions Using Chemical Propulsion, Solar Power Generation and In-Situ Resource Utilisation. *Proceedings of the 6th Australian Space Science Conference*.
51. Crawford, I. 2012. Dispelling the myth of robotic efficiency. *Astronomy & Geophysics* 53, 2.22-2.26.
52. Willson, D., Rask, J. C., George, S. C., de Leon, P., Bonaccorsie, R., Blanke, J., Slocombe, J., Silburn, K., Steele, Gargarno, M. and McKay, C. P. 2014. The performance of field scientists undertaking observations of early life fossils while in simulated space suit. *Acta Astronautica* 93, 193–20.
53. Hoffman, S. J. (ed.) 2001. The Mars Surface Reference Mission: A Description of Human and Robotic Surface Activities. NASA Technical Report TP—2001–209371.
54. Harvey, B. 2007. Soviet and Russian Luna Exploration. Springer, Berlin, 317p.

Raman characterisation of the products of alteration and low-grade metamorphism of volcanic rock minerals: preliminary results and implications for Martian studies

Emily Bathgate^{1-2*}, Graziella Caprarelli¹⁻³, Linda Xiao², Ron Shimmon², Ross Pogson⁴

1. School of the Environment, Faculty of Science, University of Technology, Sydney. PO Box 123, Broadway, NSW 2007, Australia.

2. School of Chemistry and Forensic Science, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia.

3. Division of IT, Engineering and the Environment (DITEE), University of South Australia, City East Campus, BJ3-49, Frome Road, Adelaide SA 5001, Australia

4. Research and Collections, Geosciences and Archaeology, Australian Museum, 6 College St, Sydney, NSW 2010

** Corresponding Author: Emily.Bathgate@student.uts.edu.au*

Summary

Hydrated minerals outcrop in several Martian locations, and have been interpreted as formed by aqueous deposition or alteration of basalts. Hydrated minerals support interpretations of past wetter and warmer Martian climates. Many studies are aimed at generating libraries of mineral spectra using the analytical techniques that will be employed on planned future missions to Mars.

Here we present our preliminary results of a Raman spectroscopy study of primary magmatic minerals and secondary minerals typically found in altered basalts. We used a Renishaw inVia Raman spectrometer equipped with a 17mW Renishaw Helium-Neon Laser 633nm, silicon reference standard, and a Leica DMLB microscope. We acquired the spectra of the minerals along different crystallographic orientations, to account for differences in mineral structures. The results of our research highlight challenges for future exploration and sample recovery missions.

Keywords: Raman spectroscopy, alteration minerals, Mars.

Introduction

Recent missions (ESA Mars Express; NASA Mars Reconnaissance Orbiter) to Mars have revealed in layered sequences and bedrock exposures hydrated silicate minerals [1, 2], most of which are Fe- and Mg- phyllosilicates [3]. Zeolites have been detected in the Martian dust. All these hydrated minerals are common products of alteration of basalts, which cover the entire surface of Mars [3-7]. Secondary minerals on Earth are commonly associated with weathering and water/rock interaction processes. Hence, by analogy, the presence of secondary minerals on Mars has been interpreted as evidence in support of past wetter climate

conditions, highly significant for investigations of possible conditions conducive to the emergence of life on Mars [1-5, 8-21]. The principal focus of future missions to the planet is thus the specific and detailed characterisation of hydrated minerals and of mineral assemblages that can constrain post-magmatic physical and chemical conditions.

One of the ExoMars 2018 mission instruments will be a Raman field spectrometer (Raman Laser Spectrometer, RLS; [22]). Raman spectroscopy has been widely used to characterise gaseous, liquid or solid geological materials. The technique is non-destructive, it requires minimal to no preparation and it can be performed on micro-samples of no more than a few molecules in size. The essential aspect of the technique is irradiation of a sample by monochromatic light: some of the light is scattered by the material at the same frequency as the incident light (elastic scattering). However, weak photon-molecule interactions cause excitation of the molecule from its ground state to a virtual energy state. When the molecule relaxes it emits a photon and it moves to a different rotational or vibrational state. The difference between the original and new states corresponds to a shift in the emitted photon frequency away from the excitation wavelength (inelastic scattering). If detected by a spectrometer and displayed as a spectrum, weak satellite peaks corresponding to the vibrational frequencies are observed at the sides of a strong central peak corresponding to the frequency of the incident light [23]. If the final state of the molecule is more energetic than the initial state, the photon is shifted to a lower frequency. If the final state of the molecule is less energetic the photon is shifted to a higher frequency. Raman spectra are essentially continuous binary plots of peak intensity (vertical axis) against Raman shifts reported as cm^{-1} . The location and relative intensities of the peaks are diagnostic of the material.

Five regions of the Raman spectrum are diagnostic for silicates [24]

1. Lattice vibrations (shifts $< 600\text{cm}^{-1}$ for phyllosilicates, < 450 for other silicates), caused by translational motions of cations in octahedral sites and interlayer sites relative to the SiO_4 groups in tetrahedral layers, and by oxygen and the OH^- groups lattice translations parallel to X,Y,Z crystallographic directions [24, 25];
2. Vibrational Si-Ob-Si modes Si-Ob-Si (Ob = bridging oxygen) connecting SiO_4 tetrahedra ($600\text{-}800\text{ cm}^{-1}$ for phyllosilicates, $450\text{-}750\text{ cm}^{-1}$ for other silicates);
3. Si-Onb stretching (Onb = non-bridging oxygen) in SiO_4 tetrahedra ($800\text{-}1150\text{ cm}^{-1}$);
4. OH^- bending ($1500\text{-}1600\text{ cm}^{-1}$);
5. OH^- or H_2O stretching mode ($3000\text{-}3800\text{ cm}^{-1}$).

It is furthermore noteworthy that the Onb/Ob ratio vibrational mode of minerals with fewer than two Onb (chain silicates - *inosilicates*, sheet silicates - *phyllosilicates*, and frame silicates - *tektosilicates*) show a much stronger signal in the $600\text{-}800\text{ cm}^{-1}$ range than in the $800\text{-}1150\text{ cm}^{-1}$ range [25]. Therefore the relative strength of the two vibrational modes is indicative of the chemical structure of the minerals.

The application of Raman spectroscopy to the identification and classification of mineral species in the laboratory is conducted under controlled conditions: the minerals can be separated from any matrix and cleaned prior to analysis, polished, oriented along preferred crystallographic directions; the source of the excitation beam can be changed to account for optimal analysis conditions of specific material, and the beam can be pulsed or polarised, while exposure times can be varied; the spectra can be interpreted using a variety of statistical and mathematical techniques to obtain unique results. If necessary, samples can be analysed multiple times changing orientation and performing additional preparation. On the contrary,

the conditions at which Raman field spectrometers operate are much more difficult, and the spectra are much noisier than those acquired in the laboratory [26] owing to climatic conditions, preservation of rock or mineral faces, slope of bedrock, among other factors. Furthermore, for any instrument operated remotely on the surface of a distant planet there will be fewer and limited choices of adjustment of the analytical conditions. Based on mission objectives, the set-up of the instruments will probably have to satisfy a variety of target samples (e.g., silicates, hydrated silicates, oxides and sulphates, carbonates, organic matter, fluids, etc.), and it is therefore likely that field spectrometers will be built to the best possible compromise set-up, but not be optimal for any of the materials of interest.

Given the prevalence of basalt on the Martian surface, and the nature of the hydrated minerals thus far discovered on Mars, we have undertaken a Raman spectroscopy laboratory study of the most common minerals composing basalts and their alteration products, with the purpose of: (a) generating a library of minerals; (b) identifying those aspects of the technique more likely to present challenges in the interpretation of future Martian field Raman spectra.

Basalts are composed by high-Ca plagioclase $(\text{Ca,Na})[\text{Al}(\text{Al,Si})\text{Si}_2\text{O}_8]$ and high-Ca pyroxene $[(\text{Mg,Fe,Ca})(\text{Mg,Fe})\text{Si}_2\text{O}_6]$, with variable amounts of olivine $[(\text{Mg,Fe})_2\text{SiO}_4]$, various accessory minerals, and silica-rich glass. Common minerals produced by secondary processes of alteration of the primary magmatic minerals are Fe- and Mg-phyllosilicates (e.g., chlorite), and zeolites. In this paper we present the results of our analytical work on primary basaltic and secondary phases from a collection of minerals donated by the Australian Museum in Sydney, NSW, Australia.

Materials and Methods

We analysed a broad range of primary magmatic and secondary (weathering, metamorphism) minerals and prepared a library of spectra that would likely encompass minerals commonly occurring in basalts and their alteration products. This is an essential preliminary step to investigate the potential results from Raman technique, because a field Raman spectrometer will collect spectra from mineral assemblages, rather than from pure mineral species (laboratory conditions). The reference spectra library must thus contain all minerals likely to be intercepted in the field of view of a field Raman spectrometer.

A broad array of secondary minerals forming in different environments was chosen to cover a wide range of low-grade metamorphic and weathering conditions that could have occurred on Mars during its long geological history. In addition, because the positions and intensities of peaks in Raman spectra are dependent on the crystal structure of minerals, it was necessary to also analyse phyllosilicates with a high degree of crystallinity (e.g., muscovite, biotite) which, although not part of a common equilibrium paragenesis of basaltic rocks, nonetheless represent a crystallinity end-member against which to assess other phyllosilicates (e.g., chlorite, clays).

The primary and secondary minerals analysed in this study are listed in Table 1 and images of selected minerals shown in Fig 1.

Table 1: List of reference minerals obtained for the calibration database. Entries in *italics* indicate hydrated minerals. The minerals are grouped according to their most common paragenesis.

Mineral name	Museum I.D.	Sample location	Typical chemical formula
Primary (igneous)			
Analcime	D49646	Leicester, England	$\text{Na}_2(\text{Al}_2\text{Si}_4\text{O}_{12}) \cdot 2\text{H}_2\text{O}$
Anorthoclase	D45410	Mt Anakie, VIC	$(\text{Na}, \text{K})\text{AlSi}_3\text{O}_8$
Biotite	D18221	Torrington, NSW	$\text{K}(\text{Mg}, \text{Fe}^{2+})_3(\text{Al}, \text{Fe}^{3+})\text{Si}_3\text{O}_{10}(\text{OH}, \text{F})_2$
Clinopyroxene	D52577	Mt Wyangapinni, QLD	$(\text{Ca}, \text{Na})(\text{Mg}, \text{Fe}^{2+}, \text{Al}, \text{Fe}^{3+}, \text{Ti})[(\text{Si}, \text{Al})_2\text{O}_6]$
Hypersthene	D3683	Canada	$(\text{Mg}, \text{Fe})\text{SiO}_3$
Labradorite	D58053	Mallanganee, NSW	$(\text{Ca}, \text{Na})[\text{Al}(\text{Al}, \text{Si})\text{Si}_2\text{O}_8]$
Labradorite (2)	D36683	Tentfield-Casino, NSW	$(\text{Ca}, \text{Na})[\text{Al}(\text{Al}, \text{Si})\text{Si}_2\text{O}_8]$
Muscovite	D38763	Egebek, NSW	$\text{KAl}_2(\text{AlSi}_3\text{O}_{10})(\text{OH})_2$
Olivine	D59219	Wild Pig Tier, Interlaken TAS	$(\text{Mg}, \text{Fe}^{2+})_2\text{SiO}_4$
Olivine (2)	D57526	Mt Shadwell, VIC	$(\text{Mg}, \text{Fe}^{2+})_2\text{SiO}_4$
Kaersutite	D53109	Mt Anakie, VIC	$(\text{Na})(\text{Ca}_2)(\text{Mg}_3\text{AlTi})(\text{Al}_2\text{Si}_5\text{O}_{22})\text{O}_2$
Secondary: Alteration (chemical weathering)			
Chlorite	D15219	Nundle, NSW	$(\text{Mg}, \text{Fe})_3(\text{Si}, \text{Al})_4\text{O}_{10}(\text{OH})_2(\text{Mg}, \text{Fe})_3(\text{OH})_6$
Clinochlore	D29787	Pennsylvania, USA	$(\text{Mg}, \text{Fe}^{2+})_3\text{Al}(\text{AlSi}_3\text{O}_{10})(\text{OH})_8$
Kaolinite	D33263	Marulan, NSW	$\text{Al}_2(\text{Si}_2\text{O}_5)(\text{OH})_4$
Nontronite	D41296	Attunga, NSW	$\text{Na}_{0.5}\text{Fe}_2((\text{Si}, \text{Al})_4\text{O}_{10})(\text{OH})_2 \cdot n\text{H}_2\text{O}$
Secondary: Low- and medium-grade metamorphism			
Analcime	D49646	Leicester, England	$\text{Na}_2(\text{Al}_2\text{Si}_4\text{O}_{12}) \cdot 2\text{H}_2\text{O}$
Biotite	D18221	Torrington, NSW	$\text{K}(\text{Mg}, \text{Fe}^{2+})_3(\text{Al}, \text{Fe}^{3+})\text{Si}_3\text{O}_{10}(\text{OH}, \text{F})_2$
Chabazite	D45526	Mersey-Forth, TAS	$(\text{Ca}, \text{Na}_2, \text{K}_2, \text{Mg})\text{Al}_2\text{Si}_6\text{O}_{12} \cdot 6\text{H}_2\text{O}$
Chlorite	D15219	Nundle, NSW	$(\text{Mg}, \text{Fe})_3(\text{Si}, \text{Al})_4\text{O}_{10}(\text{OH})_2(\text{Mg}, \text{Fe})_3(\text{OH})_6$
Clinochlore	D29787	Pennsylvania, USA	$(\text{Mg}, \text{Fe}^{2+})_3\text{Al}(\text{AlSi}_3\text{O}_{10})(\text{OH})_8$
Muscovite	D38763	Egebek, NSW	$\text{KAl}_2(\text{AlSi}_3\text{O}_{10})(\text{OH})_2$
Natrolite	(unregistered)	Kulnura, NSW	$\text{Na}_2\text{Al}_2\text{Si}_2\text{O}_{10} \cdot 2\text{H}_2\text{O}$
Prehnite	D39021	Prospect, NSW	$\text{Ca}_2\text{Al}_2\text{Si}_2\text{O}_{10}(\text{OH})_2$
Saponite	D36778	Cornwall, U.K.	$\text{Ca}_{0.25}(\text{Mg}, \text{Fe})_3((\text{Si}, \text{Al})_4\text{O}_{10})(\text{OH})_2 \cdot n\text{H}_2\text{O}$
Serpentine	D43620	Tamworth, NSW	$\text{D}_2[\text{Si}_2\text{O}_5](\text{OH})^{4+} \text{ or } -n\text{H}_2\text{O}$

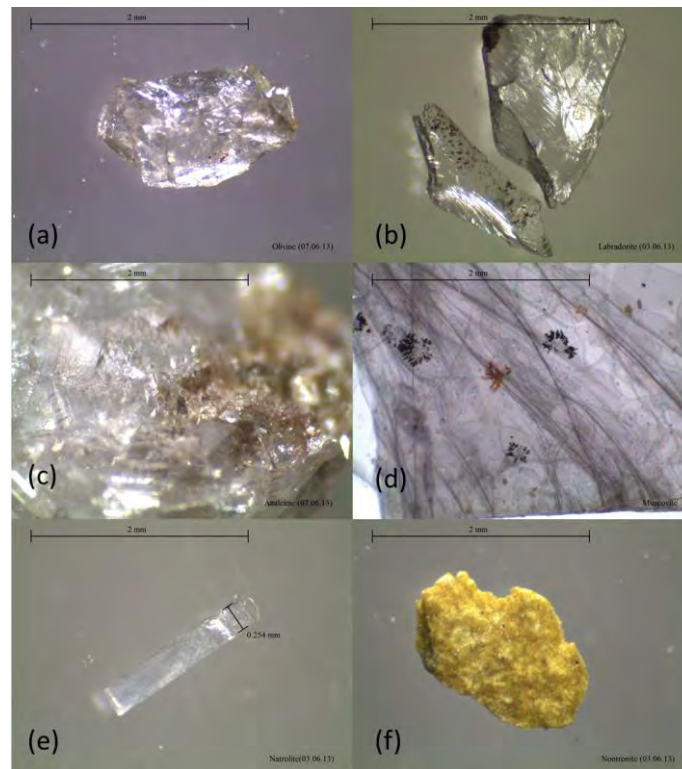


Fig. 1: Microscopic images of common primary and secondary basaltic minerals: (a) olivine; (b) labradorite; (c) analcime; (d) muscovite; (e) natrolite (crystal width: 0.254mm); (f) nontronite. All microphotographs were taken with a light microscope at 35x magnification. Scale bars: 2 mm.

The mineral samples were subdivided into fractions, and small fractions were prepared for analysis by preliminary washing and cleaning with acetone if the minerals exhibited fluorescence. Fluorescence is very common in Raman spectroscopy and in most rock forming minerals, it is most often caused by impurities such as ions of transition metals, uranium or other actinide or lanthanide (REE- Rare Earth Elements) and organic materials [27]. The element inclusions in the minerals cannot be accounted for, but organic contamination could be a major contributing factor to the high fluorescence in some of the samples. This may be mitigated by simple sample preparation such as an alcohol or acetate wash.

Raman analyses were undertaken using a Renishaw inVia Raman spectrometer (Gloucestershire, UK) equipped with a 17mW Renishaw Helium-Neon Laser 633nm and silicon reference standard, with a Leica DMLB microscope (Wetzlar, Germany) (Fig. 2). The acquisition settings varied with each mineral. Most minerals were analysed at 100% laser intensity. Others had to be analysed at lower power levels due to fluorescence effects or when the software could not plot the signal. Clinocllore was analysed at 5% standard power, kaolinite at 10% standard power and anorthoclase, serpentine and chlorite were analysed at 50%.

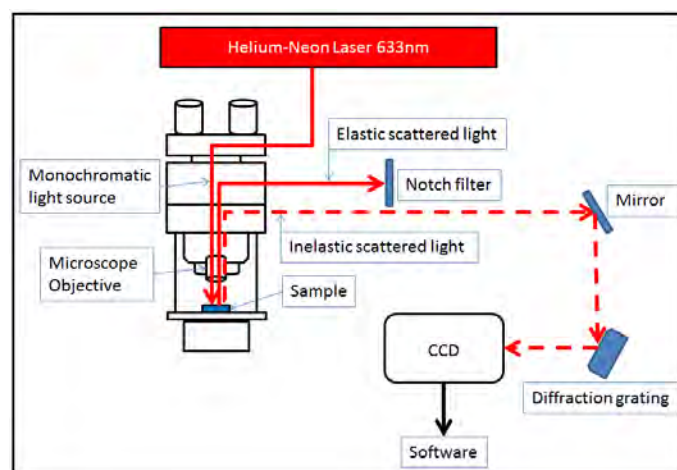


Fig. 2: Simplified schematic diagram of the Raman spectrometer used in this work. The Monochromatic light source is focused onto the sample through the microscope objective. The elastically scattered light is removed by the notch filter. The weaker Raman scattered light is focused onto a diffraction grating and then processed by the CCD. The signals are processed by the software and a binary plot of Raman shift and signal intensity is obtained as analysis output.

It has been shown that exposing a fluorescing sample to the laser beam for longer periods of time will reduce the signal to noise ratio of the sample [28]. In this study we therefore carried out multiple short readings. Fluorescing minerals, such as clays and feldspars, were acquired over an average of 4 readings of 10 s each.

Before analysing the samples, the laser was centred and calibrated against a silicone chip standard. Initial spectra for all samples were collected using a 50X objective with a spectral range of $120\text{-}3800\text{ cm}^{-1}$ subdivided into two sections: $120\text{-}1800\text{ cm}^{-1}$ and $1800\text{-}3800\text{ cm}^{-1}$. Subsequent analyses were carried out limiting the spectral range only to those shift values in which notable peaks had been initially detected, to increase the resolution of the analysis at those frequencies. Precision runs of feldspars (labradorite and anorthoclase) were further

conducted in the Raman shift range 120-800 cm^{-1} , in an attempt to overcome high background values which rendered peak identification impossible, consistently with the findings of previous researchers [29-31].

For minerals with elongated or platy crystal habits, additional analyses were performed by rotating and orientating the crystals along different crystallographic directions, identified based on the mineral cleavage and surface shapes, relative to the laser beam: for each of these minerals a measurement was undertaken, and then the sample was rotated 45° and the measurements repeated until a full 180° rotation of the sample has been completed.

Results

We obtained a set of Raman spectra that can be used as a library of spectra for the mineral groups and species expected to be present in the basalts outcropping on the surface of Mars.

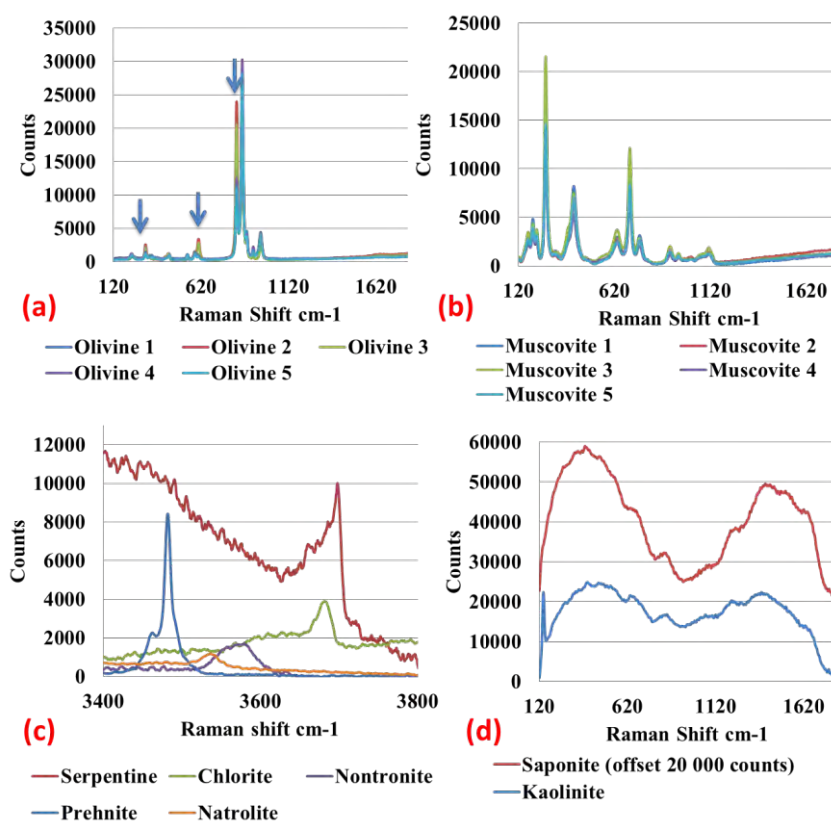


Fig. 3: Results – (a) Orientation effects for olivine in the 120-1800 cm^{-1} spectral range where spectra 1-5 are the five readings taken at different orientations of the sample, major spectral changes indicated by arrows, (b) Orientation effects for muscovite in the 120-1800 cm^{-1} spectral range where spectra 1-5 are the five readings taken at different orientations of the sample muscovite: no major spectral changes with orientation. The muscovite crystal was oriented perpendicular to the excitation source, which corresponds to a symmetric direction of the minerals in which all directions are equivalent relative to the beam. , (c) Higher range spectra of selected hydrated minerals: serpentine, chlorite, nontronite, prehnite and natrolite. All of these secondary minerals have structural water and show a distinctive peak in this range: serpentine, 3685 and 3698 cm^{-1} ; chlorite, 3681 cm^{-1} ; nontronite, 3579 cm^{-1} ; prehnite, 3482 cm^{-1} ; natrolite, 3325 and 3526 cm^{-1} , (d) Clays: Raman spectra of saponite (red) and kaolinite (blue). The sample of saponite has no distinguishing peaks. The sample of kaolinite has one distinguishing peak at 142 cm^{-1} .

Analcime: The main spectral features for analcime were identified at 390 cm^{-1} , 483 cm^{-1} and 1104 cm^{-1} with the peak at position 483 cm^{-1} showing changes in intensity with changes in orientation. This was directly comparable to spectra in the literature [32]. Analcime also shows a diagnostic peak in the higher Raman shift range at 3559 cm^{-1} (1638 counts).

Biotite: The spectrum for biotite shows major peaks at Raman shifts 186 cm^{-1} , 540 cm^{-1} and 680 cm^{-1} . The relative pattern of the peaks is consistent with spectrum published in literature [24]. The sample also exhibited diagnostic peaks in the higher Raman shift range at 3594 cm^{-1} and 3657 cm^{-1} .

Chabazite: The main spectral features for chabazite are at 329 cm^{-1} , 457 cm^{-1} and 480 cm^{-1} , with the peak at 457 cm^{-1} disappearing at some orientations. However our spectra of chabazite exhibits fluorescence, making identification of any further peaks difficult. These spectral features do not coincide with published spectra which show major spectral features at 128 cm^{-1} , 204 cm^{-1} and 465 cm^{-1} [32]. This is possibly due to a different wavelength at 1064 nm of the laser beam used. Chabazite also exhibited a diagnostic peak in the higher Raman shift range at 3465 cm^{-1} .

Chlorite: The spectrum for chlorite exhibited major peaks at 199 cm^{-1} , 355 cm^{-1} , 548 cm^{-1} and 683 cm^{-1} . Chlorite also exhibited fluorescence which made distinguishing minor peaks very difficult. A similar spectral pattern was also observed in the published literature [24]. A diagnostic peak in the higher Raman shift range at 3683 cm^{-1} was also observed (Fig. 3c).

Kaolinite and Saponite: The spectra of these minerals showed no peak (Fig. 3d), owing to the high signal noise due to fluorescence. Spectra for the clay minerals was very difficult to obtain. It is recognized that the Raman spectra of clay minerals can be difficult [33].

Clinochlore: The spectra for clinochlore exhibited high fluorescence and only peaks in the higher Raman shift range were visible. These peaks were located at 1064 cm^{-1} , 1175 cm^{-1} , 1333 cm^{-1} , and 1514 cm^{-1} .

Clinopyroxene: The spectra for clinopyroxene exhibits major peaks at 339 cm^{-1} , 398 cm^{-1} , 670 cm^{-1} and 1008 cm^{-1} , with the peak at 398 cm^{-1} disappearing at some orientations and the peak at 1008 cm^{-1} exhibiting small changes in intensity. This is comparable with published spectra [34], exhibiting peaks at 324 cm^{-1} , 391 cm^{-1} , 666 cm^{-1} , and 1011 cm^{-1} . Slight differences in Raman shift may be attributed to the use of a different laser frequency at 785 nm .

Hypersthene: The spectrum for hypersthene exhibits major peaks at 130 cm^{-1} , 231 cm^{-1} , 338 cm^{-1} , 395 cm^{-1} , 678 cm^{-1} and 1004 cm^{-1} our spectra showed major changes with orientation for the peaks at 395 cm^{-1} and 1004 cm^{-1} with some minor peaks disappearing at some orientations. Published spectra of hypersthene [35] exhibits a similar spectral pattern with major peaks at 228 cm^{-1} , 335 cm^{-1} , 392 cm^{-1} , 674 cm^{-1} and 1003 cm^{-1} , their spectral range begins at 200 cm^{-1} therefore the corresponding peak to our 130 cm^{-1} peak is not visible. Slight differences in shift may be attributed to laser frequency at 532 nm .

Labradorite and Anorthoclase - The spectra of these minerals have 4 major peaks, with the labradorite peaks occurring at 180 cm^{-1} , 282 cm^{-1} , 482 cm^{-1} and 508 cm^{-1} and the anorthoclase peaks occurring at 170 cm^{-1} , 282 cm^{-1} , 475 cm^{-1} and 511 cm^{-1} . The peak at 180 cm^{-1} in the labradorite spectrum exhibits changes in intensity with orientation as well as the minor peak at 407 cm^{-1} . This is also observable with anorthoclase as the peaks at 170 cm^{-1} and 475 cm^{-1} showing changes in intensity as well as the minor peak at 406 cm^{-1} . The feldspar minerals

were very difficult to analyse and exhibited high levels of fluorescence, especially in the higher Raman shift ranges. Earlier spectroscopic analyses of these minerals [30] show comparative peak locations with labradorite peaks at 181 cm^{-1} , 281 cm^{-1} , 484 cm^{-1} and 510 cm^{-1} and the anorthoclase peaks at 167 cm^{-1} , 199 cm^{-1} , 285 cm^{-1} , 474 cm^{-1} and 513 cm^{-1} . However they have detected extra peaks as fluorescence may have been reduced with the use of 514.5nm or 488nm laser.

Kaersutite: This mineral has 6 principal peaks at 131 cm^{-1} , 247 cm^{-1} , 378 cm^{-1} , 340 cm^{-1} , 574 cm^{-1} , 742 cm^{-1} and 1014 cm^{-1} with minor changes in intensity for the peaks at 574 cm^{-1} and 742 cm^{-1} and the peak at 430 cm^{-1} disappearing with some orientations. This spectral pattern is also observed in published spectra, with main peaks at 249 cm^{-1} , 590 cm^{-1} , 764 cm^{-1} and 1013 cm^{-1} using a laser at 532nm [36] and 243 cm^{-1} , 349 cm^{-1} , 581 cm^{-1} , 756 cm^{-1} and 1011 cm^{-1} from RRUFF database kaersutite R070128 [37].

Muscovite: In the lower Raman range the muscovite spectrum has major peaks at 196 cm^{-1} , 262 cm^{-1} , 408 cm^{-1} and 701 cm^{-1} (Fig.3b), comparable to published spectra [24, 25, 38]. Muscovite also exhibits a diagnostic peak in the higher Raman shift range at 3624 cm^{-1} .

Natrolite: In the lower Raman shift range natrolite shows 5 main diagnostic peaks at 160 cm^{-1} , 205 cm^{-1} , 441 cm^{-1} , 532 cm^{-1} and 1038 cm^{-1} , with the peaks at 160, 441 and 1038 showing changes in intensity with different orientations and the peak at 185 showing major changes with the peak disappearing in some orientations. This is comparable to published data [32]. Natrolite also exhibits diagnostic peaks in the higher Raman shift range at 3212 cm^{-1} , 3325 cm^{-1} and 3535 cm^{-1} (Fig. 3c).

Nontronite: In the lower Raman shift range the nontronite spectrum exhibits fluorescence, although distinguishing peaks are still visible at 164 cm^{-1} , 237 cm^{-1} , 285 cm^{-1} , 418 cm^{-1} and 683 cm^{-1} with the peak at 285 changing in intensity with different orientations. This spectrum is comparable to spectra published in other work [39]. Nontronite also exhibits a diagnostic peak in the higher Raman shift range 3579 cm^{-1} (Fig. 3c).

Olivine: Two different samples of olivine of similar composition were analysed. Their spectra have major peaks at 605 cm^{-1} , and 880 cm^{-1} (Fig. 3a) and the characteristic olivine doublet at 822 cm^{-1} and 854 cm^{-1} . Fig. 3a also highlights the changes in the spectra as the crystal direction is changed, 605 cm^{-1} and 880 cm^{-1} disappear in some orientations while the peak at 822 cm^{-1} changes in intensity. A study [40] on magmatic and synthetic olivines produces peaks at different shifts with the olivine doublet at 816 and 844 owing to the use of a laser beam at 532.3nm frequency.

Prehnite: The principal peaks for prehnite occur at 317 cm^{-1} , 387 cm^{-1} , 520 cm^{-1} , 934 cm^{-1} and 987 cm^{-1} with the peaks at 317 cm^{-1} and 520 cm^{-1} exhibiting changes in intensity with orientation whilst some minor peaks disappear in some orientations. This spectral pattern is consistent with published spectra [41]. In the higher Raman shift range prehnite also exhibits diagnostic peaks at 3462 cm^{-1} and 3481 cm^{-1} (Fig. 3c).

Serpentine: In the lower Raman shift range the serpentine spectrum exhibits high levels of fluorescence, drowning out the signal of most of the diagnostic peaks of the mineral. Several low intensity peaks are visible at 680 cm^{-1} , 1111 cm^{-1} , 1259 cm^{-1} and 1547 cm^{-1} . In the higher Raman shift range two peaks at 3684 cm^{-1} and 3699 cm^{-1} are very clear (Fig. 3c), and correspond to published spectra [24]

Interpretation: opportunities and challenges

The minerals analysed in this study rendered spectra consistent with those published in the literature. This is significant for two important reasons: (1) regardless of the frequency of the beam used to excite the samples, diagnostic peaks of minerals are uniquely identifiable by their relative positions, with high degree of certainty; (2) it is unlikely that the chemical compositions of the mineral samples analysed in this study are exactly the same as those of minerals from other studies. Thus, Raman spectroscopy is sensitive to mineral structures, but not to minor chemical variability in minerals.

Both these points are fundamentally important for the purpose of using field Raman spectrometers in planetary exploration. Future field spectrometers will be built to accommodate a variety of tasks, and it may not be possible to equip them with multiple light sources. Knowing that the frequency of the laser beam does not significantly affect the pattern and sequence of diagnostic peaks in the spectra, it will be possible to compare spectra across libraries and databases, including those not collected for the specific purpose of the mission. Furthermore, the higher sensitivity of Raman spectra to structure rather than to exact chemical compositions will be an asset in studies of Martian minerals, which are probably higher in Fe than their terrestrial equivalents.

This second point, though, poses a challenge in the future geological interpretation of the spectra to be acquired on the Martian surface. In order to derive information about the pressure and temperature conditions of formation of the mineral assemblages, the chemical gradients and exact compositions of the mineral phases must be known. The Mg/Fe ratios of olivine samples have been determined from Raman spectra, by precise measurements of the strongest peaks [42]. It is not clear at this stage whether the same technique could be successfully applied to other minerals, but it is undoubted that minerals with more complicated structures and chemical compositions, which is the case of most of the hydrated minerals of interest for Martian exploration, present a much bigger challenge. Some field Raman spectrometers are designed to include a laser-induced breakdown spectroscopy (LIBS) system, to obtain major element concentrations [43]. Based on the information available at the time of writing, however, the Raman Laser Spectrometer (RLS) that will participate in the ExoMars mission (scheduled to land in 2018), will analyse samples after crushing and powdering them in the landing rover, which will not include a LIBS. In our follow-up work we will investigate how lack of information on the precise chemical composition of mineral species may affect geological interpretations, using terrestrial samples as analogues of Martian rocks.

Additional challenges are tied to fluorescence and orientation. Plagioclase is one of the principal primary components in basalts. In our laboratory study the samples of feldspar (labradoritic plagioclase and anorthoclase) proved very difficult to analyse owing to the strong fluorescence of the minerals. When analysing powdered rocks, essentially composed of mixtures of minerals, the presence in the mixed spectra of the fluorescing spectra of plagioclase could mask the peaks, making interpretation of the spectra impossible. Especially significant is the effect of plagioclase fluorescence on the identification and quantification of hydrated minerals: these have OH⁻ diagnostic peaks at Raman shift values > 3000 cm⁻¹. This is the range of Raman shifts where plagioclase fluoresces more strongly.

An additional effect is due to the orientation of anisotropic minerals, as documented by the case of analcime, chabazite, clinopyroxene, hypersthene, labradorite, anorthoclase, kaersutite, natrolite, nontronite and prehnite. In our follow up study we will work on mineral mixtures to document thoroughly the orientation effect and its interpretative importance for future missions to Mars.

Conclusions

Raman spectroscopy is a very powerful non-destructive technique. It is extremely effective for the detection of micro-quantities of material without preliminary preparation, and is therefore suitable for inclusion in field instruments operated remotely. Raman spectrometers are also very versatile, and can be used to detect and identify any type of substance, solid, liquid or gaseous. However, application of field Raman spectrometry to the analysis of minerals contained in rocks requires calibration for many effects that, alone or in combination, may mask principal diagnostic peaks. In this study we showed fluorescence and mineral orientation, which will cause interpretative challenges for the data that will be acquired from the field Raman spectrometer on future Mars missions. We are performing additional investigations to constrain uniquely these effects in altered basaltic rocks in order to provide an interpretative framework for data returned from missions.

References

1. Bibring, J.-P., et al., *Global Mineralogical and Aqueous Mars History Derived from OMEGA/Mars Express Data*. Science, 2006. **312**(5772): p. 400-404.
2. Mustard, J.F., et al., *Hydrated silicate minerals on Mars observed by the Mars Reconnaissance Orbiter CRISM instrument*. Nature, 2008. **454**(7202): p. 305-309.
3. Ehlmann, B.L., et al., *Subsurface water and clay mineral formation during the early history of Mars*. Nature, 2011. **479**(7371): p. 53-60.
4. Bandfield, J.L., V.E. Hamilton, and P.R. Christensen, *A Global View of Martian Surface Compositions from MGS-TES*. Science, 2000. **287**(5458): p. 1626-1630.
5. McSween, H.Y., et al., *Basaltic Rocks Analyzed by the Spirit Rover in Gusev Crater*. Science, 2004. **305**(5685): p. 842-845.
6. McSween, H.Y., T.L. Grove, and M.B. Wyatt, *Constraints on the composition and petrogenesis of the Martian crust*. Journal of Geophysical Research: Planets, 2003. **108**(E12): p. 5135.
7. McSween, H.Y., G.J. Taylor, and M.B. Wyatt, *Elemental Composition of the Martian Crust*. Science, 2009. **324**(5928): p. 736-739.
8. Bishop, J., *Hydrated Minerals on Mars*, in *Water on Mars and Life*, T. Tokano, Editor. 2005, Adv. Astrobiol. Biogeophys. p. 65-96.
9. Bishop, J.L., H. Fröschl, and R.L. Mancinelli, *Alteration processes in volcanic soils and identification of exobiologically important weathering products on Mars using remote sensing*. Journal of Geophysical Research: Planets, 1998. **103**(E13): p. 31457-31476.
10. Carter, J., et al., *Hydrous minerals on Mars as seen by the CRISM and OMEGA imaging spectrometers: Updated global view*. Journal of Geophysical Research: Planets, 2013. **118**(4): p. 831-858.
11. Carter, J., et al., *Mineralogy of Layered Deposits in Terby Crater, N. Hellas Planitia*. 41st LPSC, 2010.

12. Carter, J., et al., *Detection of Hydrated Silicates in Crustal Outcrops in the Northern Plains of Mars*. Science, 2010. **328**(5986): p. 1682-1686.
13. Chevrier, V. and P.E. Mathé, *Mineralogy and evolution of the surface of Mars: A review*. Planetary and Space Science, 2007. **55**(3): p. 289-314.
14. Christensen, P.R., et al., *Identification of a basaltic component on the Martian surface from Thermal Emission Spectrometer data*. Journal of Geophysical Research: Planets, 2000. **105**(E4): p. 9609-9621.
15. Clark, B.C., et al., *Chemistry and mineralogy of outcrops at Meridiani Planum*. Earth and Planetary Science Letters, 2005. **240**(1): p. 73-94.
16. Ehlmann, B.L., et al., *Identification of hydrated silicate minerals on Mars using MRO-CRISM: Geologic context near Nili Fossae and implications for aqueous alteration*. Journal of Geophysical Research: Planets, 2009. **114**(E2): p. E00D08.
17. Hamilton, V.E., P.R. Christensen, and J.L. Bandfield, *Planetary science (communication arising): Volcanism or aqueous alteration on Mars?* Nature, 2003. **421**(6924): p. 711-712.
18. Haskin, L.A., et al., *Water alteration of rocks and soils on Mars at the Spirit rover site in Gusev crater*. Nature, 2005. **436**(7047): p. 66-69.
19. Poulet, F., et al., *Phyllosilicates on Mars and implications for early martian climate*. Nature, 2005. **438**(7068): p. 623-627.
20. Viviano, C.E., J.E. Moersch, and H.Y. McSween, *Implications for early hydrothermal environments on Mars through the spectral evidence for carbonation and chloritization reactions in the Nili Fossae region*. Journal of Geophysical Research: Planets, 2013: p. n/a-n/a.
21. Wyatt, M.B. and H.Y. McSween, *The Orbital Search for Altered Materials on Mars*. Elements, 2006. **2**(3): p. 145-150.
22. Vago, J.L., et al., *ExoMars - searching for life on the Red Planet*. ESA Bulletin, 2006. **126**: p. 16-23.
23. McMillan, P., *Raman spectroscopy in mineralogy and geochemistry*. Annual Review of Earth & Planetary Sciences, 1989. **17**: p. 255-283.
24. Wang, A., J. Freeman, and K. Kuebler. *Raman spectroscopic characterization of phyllosilicates*. in *Lunar and Planetary Institute Science Conference Abstracts*. 2002.
25. Wang, A., et al., *Database of Standard Raman Spectra of Minerals and Related Inorganic Crystals*. Appl. Spectrosc., 1994. **48**(8): p. 959-968.
26. Olcott Marshall, A. and C.P. Marshall, *Field-Based Raman Spectroscopic Analyses of an Ordovician Stromatolite*. Astrobiology, 2013. **13**(9): p. 814-820.
27. Haskin, L.A., et al., *Raman spectroscopy for mineral identification and quantification for in situ planetary surface analysis: A point count method*. Journal of Geophysical Research: Planets, 1997. **102**(E8): p. 19293-19306.
28. Brody, R.H., et al., *FT-Raman Spectroscopy, Applications*, in *Encyclopedia of Spectroscopy and Spectrometry (Second Edition)*, L. Editor-in-Chief: John, Editor. 1999, Academic

Press: Oxford. p. 732-740.

29. Freeman, J.J., et al., *Characterization of Natural Feldspars by Raman Spectroscopy for Future Planetary Exploration*. The Canadian Mineralogist, 2008. **46**(6): p. 1477-1500.
30. Mernagh, T.P., *Use of the laser Raman microprobe for discrimination amongst feldspar minerals*. Journal of Raman Spectroscopy, 1991. **22**(8): p. 453-457.
31. Tarcea, N., et al., *Raman Spectroscopy—A Powerful Tool for in situ Planetary Science*, in *Strategies of Life Detection*, O. Botta, et al., Editors. 2008, Springer US. p. 281-292.
32. Mozgawa, W., *The relation between structure and vibrational spectra of natural zeolites*. Journal of Molecular Structure, 2001. **596**(1–3): p. 129-137.
33. Frost, R.L., et al., *Raman spectroscopy of kaolinites using different excitation wavelengths*. Journal of Raman Spectroscopy, 2001. **32**(8): p. 657-663.
34. Muniz-Miranda, M., C. Gellini, and L. Bindi, *Surface-enhanced Raman spectroscopy for identifying rock composition*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2009. **73**(3): p. 456-459.
35. Buzatu, A. and N. Buzgar, *The Raman Study of Single-Chain Silicates*. Analele Stiintifice ale Universitatii “Al I Cuza” din Iasi Seria Geologie, 2010: p. 107-125.
36. Apopei, A.I., N. Buzgar, and A. Buzatu, *Raman and infrared spectroscopy of kaersutite and certain common amphiboles*. Analele Stiintifice ale Universitatii “Al I Cuza” din Iasi Seria Geologie, 2011. **57**(2): p. 35-58.
37. Downs, R.T., *The RRUFF Project: an integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals*. Program and Abstracts of the 19th General Meeting of the International Mineralogical Association, 2006. **003-13**.
38. Tlili, A., D. Smith, and J. Beny, *A Raman microprobe study of natural micas*. Mineralogical Magazine, 1989. **53**(Part 2 (370)): p. 165-179.
39. Wang, A., L. Haskin, and B. Jolliff, *Characterization of mineral products of oxidation and hydration by laser Raman spectroscopy: Implications for in situ petrologic investigation on the surface of Mars*. Lunar Planet. Sci. XXIV, 1998.
40. Kuebler, K.E., *A combined electron microprobe (EMP) and Raman spectroscopic study of the alteration products in Martian meteorite MIL 03346*. Journal of Geophysical Research: Planets, 2013: p. n/a-n/a.
41. Bartholomew, P.R., *Comparing the Success Rate of Raman Spectroscopy and Powder XRD for Routine Mineral Identification*. Geostandards and Geoanalytical Research, 2013: p. n/a-n/a.
42. Foster, N.F., et al., *Identification by Raman spectroscopy of Mg–Fe content of olivine samples after impact at 6kms⁻¹ onto aluminium foil and aerogel: In the laboratory and in Wild-2 cometary samples*. Geochimica et Cosmochimica Acta, 2013. **121**(0): p. 1-14.
43. Sharma, S.K., et al., *Stand-off Raman spectroscopic detection of minerals on planetary surfaces*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2003. **59**(10): p. 2391-2407.

Measurement Uncertainty in TRMM Observations of the Precipitation and Column Water Vapor Relationship

James B. Gilmore¹

¹*Climate Change Research Centre,
University of New South Wales, Sydney NSW 2052, Australia
james.gilmore@unsw.edu.au*

Summary: Precipitation and water vapor play fundamental roles in the climate system, yet the relationship between the two only gained recent attention after Peters and Neelin [1] examined satellite retrievals of column integrated water vapor and precipitation. They analyzed microwave observations taken from the Tropical Rainfall Measuring Mission (TRMM) satellite and found a signature suggesting that a continuous phase transition takes place in this system, whereby the pickup in the mean precipitation is found with a corresponding peak in the variance at this pickup. Here I study the influence of measurement uncertainty on this relationship using four different precipitation retrievals from TRMM. I show that precipitation radar retrievals do not exhibit the expected form of a continuous phase transition in the precipitation and column water relationship. I discuss this result in the context of critical theory and its implications for the proposed precipitation phase transition. I show that geophysical measurement uncertainty must be carefully considered in situations where critical phenomena may be present.

Keywords: precipitation, water vapor, measurement uncertainty, satellite observations, TRMM, phase transition, extremes

Introduction

In Earth's hydrological cycle, rainfall and atmospheric water vapour are important components [2] and understanding how these two constituents influence the hydrological cycle is an important challenge, particularly in the context of climate change and the climate system in general [3,4,5,6]. One method that has been used to understand the roles of rainfall and atmospheric water vapour is to examine instantaneous precipitation (P) as a function of column integrated water vapour (w) [7]. The resulting curve, which I call the P - w relationship, encodes important information about the nature of atmospheric convection and rainfall physics. This means that the P - w relationship can potentially shed light on the physical processes leading to rainfall in Earth's climate system.

Peters and Neelin [1] extended the work by Bretherton et al. [7] by examining the P - w relationship for instantaneous precipitation and column water vapour measurements from the tropical microwave imager (TMI) onboard TRMM [8]. They used Remote Sensing Systems (RSS) retrievals, where the column-integrated water vapour was retrieved in rainy and non-rainy scenes [9,10]. They found a strong upswing and then an apparent plateau in precipitation response for high column integrated water vapour values.

The curve Peters and Neelin [1] found seemed to exhibit the form of a continuous phase transition. The mean response in the precipitation matched the form $P \sim (w - w_c)^\alpha$, where w_c is a critical amount of column-integrated water vapour and α is an empirically determined exponent, known as a critical exponent. In addition to this, the variance showed a peak at w_c

such that a divergence in the variance (or susceptibility) was seen, i.e. $\chi^2 \sim 1/|w-w_c|^\beta$ near the critical value, where β is another critical exponent. Further work showed the curves coalesce when separated by column averaged atmospheric temperature [11].

The power law behaviour of the mean response and the peak in the variance seemed to provide evidence for the presence of a second order continuous phase transition in the atmospheric column [12]. Observing these two features in combination typically implies that the system has undergone a phase transition and there has been a change in its stable thermodynamic state. For the P-w transition suggested by [1], one would interpret the state to change from non-raining to raining as the total column water vapour was increased past the critical value w_c , with the system showing large variability in between.

Phase transitions are normally examined in carefully controlled conditions where potential noise sources are reduced, as for example in liquid Helium measurements [13]. In remote sensing measurements, this is generally not possible and observations, generally, contain uncertainty [14,15,16,17]. The satellite retrievals of oceanic precipitation and column-integrated water vapour from TMI are known to contain uncertainty [9,10] and for these microwave measurements, this happens because a number of microwave sources are combined into a single brightness temperature observation.

Given that uncertainty is present in these measurements, it is natural to ask: Just how robust is the phase transition interpretation to the presence of uncertainty in the precipitation measurements? Here I will address this question by constructing the P-w relationship using four different precipitation retrieval algorithms using TRMM observations. In the remainder of this paper, I will discuss the observations and algorithms used from TRMM to construct the P-w relationship. I will then present these results and discuss the findings and their implications for the proposed convective phase transition.

Observations

In this section, I will discuss the observations used in this paper. The TRMM satellite has two instruments that are used to retrieve instantaneous surface rainfall, the TMI and the precipitation radar (PR). Of these instruments, the TMI passively observes the microwave brightness temperature at a number of frequencies and then uses this information to reconstruct the surface rain rate from inverse modelling. The PR on the other hand actively measures the amount of scatter from rain droplets, which is then converted to a surface rainfall rate. For both instruments, two different algorithms have been developed to retrieve surface rainfall.

For TMI, RSS has developed an algorithm [10] that essentially diagnoses cloud water and then scales this to a surface rain rate. Another algorithm using TMI alone, known as GPROF has also been developed [18]. This algorithm uses a database of cloud resolving simulations with microwave brightness predictions to optimally match the observations assuming the cloud-resolving simulations provide a reasonable approximation of precipitation characteristics. This approach thereby allows the identification of surface rainfall rates through the best match to the cloud-resolving simulations.

The PR based algorithms diagnose rainfall more directly than the TMI algorithms because of the active nature of the radar measurement. The standard algorithm to achieve a surface rain rate estimate is that of Iguchi et al. [19]. There is also another algorithm that uses the PR. This

is the combined PR and TMI algorithm of Masunaga and Kummerow [20]. The latter algorithm uses PR and TMI (called PR+TMI here) measurements to constrain the radar equation through an iterative approach that reduces the mismatch between the observations and the reconstruction.

For these precipitation retrievals, differences between TMI only and PR reconstructions can be large in cyclones [21]. In addition, validation of instantaneous rain rates shows TMI retrievals can underestimate rain rates, especially at extreme rain rates [22]. Given that there are differences between the validation of the TMI and PR algorithms in terms of instantaneous rain rates, how does this influence the P-w curves? I address this in the next section.

Precipitation Retrievals and the P-w Relationship

The original phase transition interpretation by Peters and Neelin [1] used the TMI rain rate retrieval from RSS along with the RSS water vapour retrieval. Since this time, however, TMI rain rate retrievals have been shown to contain substantial uncertainties for extreme rain rates [21,22], which may potentially influence the fluctuations in the mean response for the P-w relationship. Given the importance of the variance peak to the second order continuous phase transition interpretation, I examine this quantity and the mean response across the four precipitation retrieval algorithms from TRMM.

Fig. 1 shows the P-w relationship derived from TRMM using the four different rain rate algorithms. All panels use the RSS w retrieval. The figure shows both the mean and the variance of the response for the (a) RSS retrieval, (b) GPROF algorithm, (c) PR retrieval, and (d) the PR+TMI retrieval. Each panel shows satellite swath (level 2) data from 2006 to 2011, selected for high quality retrievals, then averaged to a 0.5° by 0.5° grid. The grid is then masked so that all four data sets contain the same measurements, and is lastly combined from 20S to 20N where the average is taken and shown in Fig. 1. This procedure is equivalent to that from Peters and Neelin [1].

The mean curves of Fig. 1 (solid lines) show generally good agreement amongst the four different retrievals; however, differences are seen in the plateau at high w values. The microwave only retrievals in Fig. 1(a,b) show higher slopes in this region than the PR retrievals in Fig. 1(c,d) which are flatter. Otherwise, the four different retrievals agree about the location of the upswing in the mean response.

For the precipitation variance in Fig. 1, comparing the four different retrievals shows rather different behaviour amongst the retrievals. The PR+TMI retrieval in Fig. 1(d) does not show a reduction in the variance but remains flat as w is increased past the upswing in mean precipitation. This obviously disagrees with the RSS TMI retrieval in Fig. 1(a), which shows a peak at the location of the upswing and then a reduction for high w values. The GPROF retrieval in Fig. 1(b) shows behaviour similar to RSS, and the PR only retrieval from Fig. 1(c) also shows similar behaviour to the PR+TMI with no clear peak in the precipitation variance.

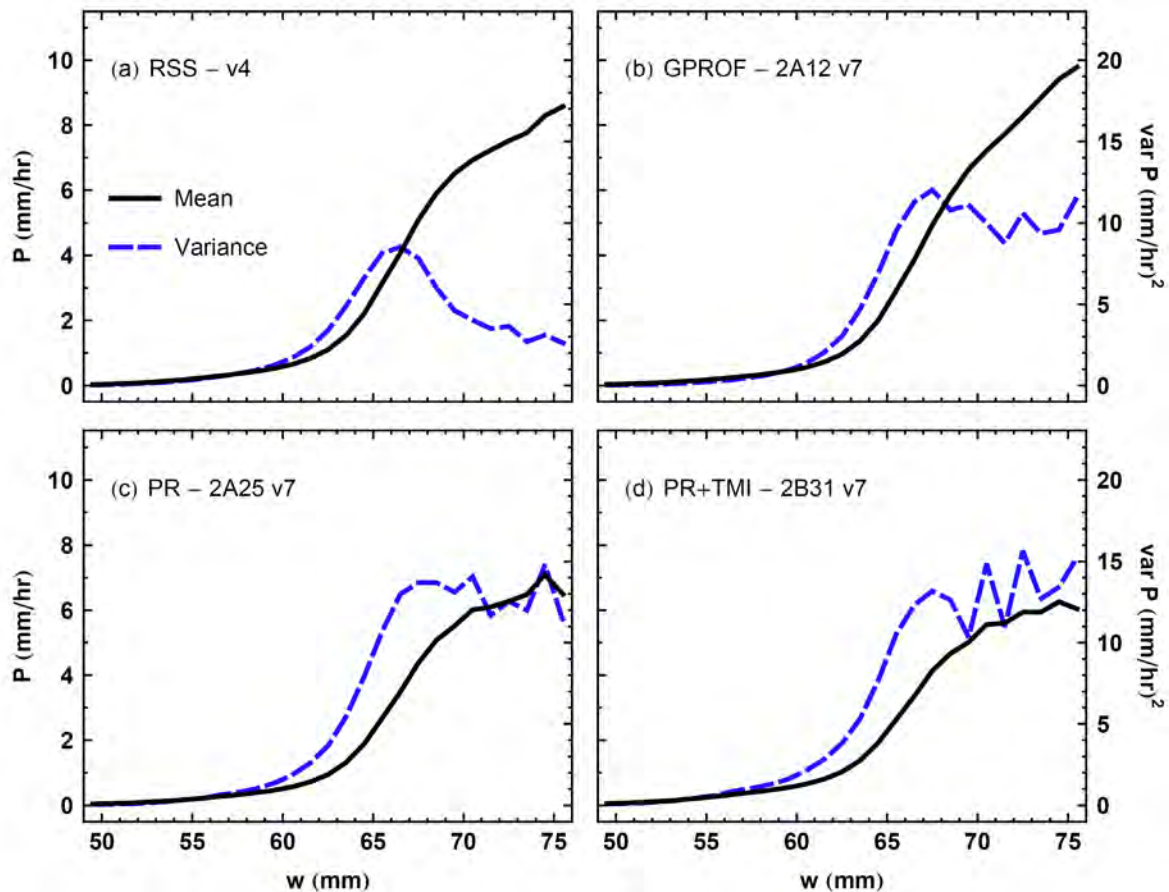


Fig. 1: The P - w relationship as observed from TRMM using four different precipitation retrieval algorithms. Shown are the mean (solid black line) and variance (dashed blue line) for instantaneous rain rate P retrievals. TMI only based retrievals are given by the (a) RSS and (b) GPROF algorithms. PR based retrievals are shown in (c) using only the PR and (d) the combined PR and TMI retrieval. Algorithm versions are also shown. All panels use the TMI RSS algorithm for column integrated water vapour w . Data is averaged to 0.5° square grids before being combined over the tropics (20°S to 20°N). The mean and variance are represented by the left and right axes respectively.

Discussion

When instantaneous rainfall rates from the TMI are compared with ground based instantaneous rain rate estimates, it is found that TMI retrievals can saturate at rain rates of 10-20 mm/hr [10,23]. This means that the TMI retrievals have a substantially reduced dynamic range in their precipitation retrieval characteristics. The dynamic range reduction can be up to a factor of 10 when compared to the PR [21]. This can be understood by noting that the PR uses a much more direct method of measuring rainfall, whereas TMI must deconstruct a number of sources in its single microwave brightness measurement. Validation against instantaneous rainfall rates does show the PR based retrievals have a realistic dynamic range and the PR+TMI retrieval is generally the most accurate at high rain rates [22].

Since the PR+TMI retrieval has a much more realistic dynamic range than the RSS TMI retrieval at extreme rain rates, how does this reduced dynamic range of RSS TMI influence the P-w curves? Fig. 1 showed the mean response is similar amongst all retrievals, which is likely due to the excellent calibration performed by RSS, thereby providing good representations of averaged quantities. However, for the variance in the presence of a reduced dynamic range, one would expect the variance curve to be reduced somehow. In the context of the RSS P-w curve, this reduced dynamic range in the RSS product occurs for high rain rates, i.e. above the threshold or critical value in the P-w curve. This will therefore naturally lead to a reduction in the variance past the so-called critical point.

Examining the RSS variance curve in Fig. 1(a) shows an increase leading up to $w_c = 66$ mm and then a reduction past this point creating a peaked variance. This reduction is a direct result of the reduced dynamic range of the RSS TMI precipitation retrieval. This reduction is not seen in the PR or PR+TMI retrievals. Since the PR-based retrievals exhibit the best ground validation, the peak in the P-w variance curve, which is vital for the second order continuous phase transition interpretation of the P-w relationship, is seemingly absent from the observations when using the PR to estimate the instantaneous rainfall rate at the surface.

For the mean response, differences are seen between the TMI and PR based algorithms at high w , with the PR based algorithms having a lower mean response and being generally flatter at high w . These differences are most likely because the PR more directly measures actual rain rates and thereby does not spread high rain rates over larger spatial regions with similar w values. Such spatial spreading, which would enhance the mean response at high w , could potentially occur for RSS and GPROF (both TMI-based algorithms) through beam filling of the microwave observations or degeneracies in the inverse modelling that are not broken by the available microwave channels. For the RSS algorithm in particular, the employed scaling of observed cloud water could potentially provide this effect because we would generally expect thick clouds over large regions during extreme precipitation events (especially in cyclones) and not just at the site of the extreme rainfall event.

Measurement uncertainty in w from the RSS algorithm has been characterized in [10] and is Gaussian with a standard deviation of 1 mm in non-rainy scenes and 3 mm in rainy scenes over the tropics. These errors will not have a substantial influence on this work because they do not change the suppressed dynamic range of the TMI retrievals relative to the PR and ground-based validation. These errors will change the number of counts at high w (with larger errors increasing the number of counts), which in turn will extend the plateau and/or flatten the mean response at high w . Only the RSS retrieval of w in rainy scenes has been validated, meaning there is no second independent w retrieval to use for comparison.

Conclusion

This work has shown that when using four different retrieval algorithms to construct the P-w relationship the mean response between the algorithms was generally consistent. However, for the variance of the P-w relationship, quite different responses were found. The RSS TMI retrieval, due to the presence of rain rate saturation at high rain rates, had a variance that reduced. On the other hand, when the PR-based algorithms are used for the rain rate retrieval in the P-w relationship, the variance peak, which is found for the TMI and is paramount for the second order continuous phase transition interpretation, is absent from the observations. Since the PR has a larger and substantially more realistic precipitation range than the RSS

TMI retrieval, this indicated that the microwave retrievals could mimic a second order continuous phase transition because of saturation at extreme rain rates.

An important question remains, and that is whether the second order continuous phase transition is present in Earth's atmosphere? Based on this work it seems unlikely that such a phase transition is present because the PR does not exhibit a variance peak and has a much better validation than TMI. This work instead suggests that the atmosphere retains high precipitation variability at high column water vapour values in the tropics.

Further observational improvements in our understanding of the P-w relationship will most likely come from experimental studies where the size of the measurement uncertainty in precipitation is reduced. The upcoming Global Precipitation Mission, which will fly with an improved microwave imager and updated radar, is well placed to improve our understanding of the P-w relationship. Given the highly variable precipitation response found here at extreme column water vapour values, more accurate measurements of instantaneous rainfall rates from the Global Precipitation Mission will provide increased knowledge of this variability inside and outside of the tropics.

Acknowledgements

I thank Remote Sensing Systems for help with their TMI data. I also thank GHRC and GES DISC from NASA for providing TRMM precipitation data. This work was supported by the Australian Research Council as part of Super Science Fellowship FS100100054.

References

1. Peters, O., and J. D. Neelin (2006), Critical phenomena in atmospheric precipitation, *Nature Phys.*, 2, 393-396.
2. Allan, R. P., and B. J. Soden (2008), Atmospheric warming and the amplification of precipitation extremes, *Science*, 321, 1481-1484.
3. Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns (2000), Climate extremes: Observations, modeling, and impacts, *Science*, 289, 2068-2074.
4. Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons (2003), The changing character of precipitation, *Bull. Amer. Meteor. Soc.*, 84, 1205-1217.
5. Held, I. M., and B. J. Soden (2006), Robust responses of the hydrological cycle to global warming, *J. Climate*, 19, 5686-5699.
6. Huntington, T. G. (2006), Evidence for intensification of the global water cycle: Review and synthesis, *J. Hydrol.*, 319, 83-95.
7. Bretherton, C. S., M. E. Peters, and L. E. Back (2004), Relationships between water vapor path and precipitation over the tropical oceans, *J. Climate*, 17, 1517-1528.

8. Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson (1998), The tropical rainfall measuring mission (TRMM) sensor package, *J. Atmos. Ocean. Tech.*, 15, 809-817.
9. Wentz, F. J. (1997), A well-calibrated ocean algorithm for special sensor microwave/imager, *J. Geophys. Res.-Oceans*, 102, 8703-8718.
10. Wentz, F. J., and R. W. Spencer (1998), SSM/I rain retrievals within a unified all-weather ocean algorithm, *J. Atmos. Sci.*, 55, 1613-1627.
11. Neelin, J. D., O. Peters, and K. Hales (2009), The transition to strong convection, *J. Atmos. Sci.*, 66, 2367-2384.
12. Yeomans, J. M. (1992), *Statistical Mechanics of Phase Transitions*, Oxford Science Publications.
13. Buckingham, M. J., and W. M. Fairbank (1961), Chapter III: The Nature of the λ -Transition in Liquid Helium, *Progress in Low Temperature Physics*, vol. 3, pp. 80-112, Elsevier.
14. Ricciardulli, L., and F. J. Wentz (2004), Uncertainties in sea surface temperature retrievals from space: Comparison of microwave and infrared observations from TRMM, *J. Geophys. Res.-Oceans*, 109, C12013.
15. Kummerow, C., W. Berg, J. Thomas-Stahle, and H. Masunaga (2006), Quantifying global uncertainties in a simple microwave rainfall algorithm, *J. Atmos. Ocean. Tech.*, 23, 23-37.
16. Hilburn, K. A., F. J. Wentz, D. K. Smith, and P. D. Ashcroft (2006), Correcting active scatterometer data for the effects of rain using passive radiometer data, *J. Appl. Meteor. Climatol.*, 45, 382-398.
17. Tian, Y., G. J. Huffman, R. F. Adler, L. Tang, M. Sapiiano, V. Maggioni, and H. Wu (2013), Modeling errors in daily precipitation measurements: Additive or multiplicative?, *Geophys. Res. Lett.*, 40, 2060-2065.
18. Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D. B. Shin, and T. T. Wilheit (2001), The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors, *J. Appl. Meteor.*, 40, 1801-1820.
19. Iguchi, T., T. Kozu, R. Meneghini, J. Awaka, and K. Okamoto (2000), Rain-profiling algorithm for the TRMM precipitation radar, *J. Appl. Meteor.*, 39, 2038-2052.
20. Masunaga, H., and C. D. Kummerow (2005), Combined radar and radiometer analysis of precipitation profiles for a parametric retrieval algorithm, *J. Atmos. Ocean. Tech.*, 22, 909-929.
21. Cecil, D. J., and M. Wingo (2009), Comparison of TRMM rain-rate retrievals in tropical cyclones, *J. Meteor. Soc. Japan*, 87, 369-380.
22. Wolff, D. B., and B. L. Fisher (2008), Comparisons of instantaneous TRMM ground validation and satellite rain-rate estimates at different spatial scales, *J. Appl. Meteor. Climatol.*, 47, 2215-2237.

23. Wolff, D. B., and B. L. Fisher (2009), Assessing the relative performance of microwave-based satellite rain-rate retrievals using TRMM ground validation data, *J. Appl. Meteor. Climatol.*, 48, 1069-1099.

The role of Jupiter in driving Earth's orbital evolution

Jonathan Horner^{1,2}, Dave Waltham³ and F. Elliot Koch^{1,2,4}

¹ School of Physics, University of New South Wales, Sydney, NSW 2052, Australia

² Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW 2052, Australia

³ Department of Earth Sciences, Royal Holloway, University of London

⁴ San Diego State University, Physics Department, San Diego, CA 92182-1233, USA, 5500 Campanile Drive

Summary: In coming years, the first truly Earth-like planets will be discovered orbiting other stars, and the search for signs of life on these worlds will begin. However, such observations will be hugely time-consuming and costly, and so it will be important to determine which of those planets represent the best prospects for life elsewhere. One of the key factors in such a decision will be the climate variability of the planet in question - too chaotic a climate might render a planet less promising as a target for our initial search for life elsewhere.

On the Earth, the climate of the last few million years has been dominated by a series of glacial and interglacial periods, driven by periodic variations in the Earth's orbital elements and axial tilt. These Milankovitch cycles are driven by the gravitational influence of the other planets, and as such are strongly dependent on the architecture of the Solar system.

Here, we present the first results of a study investigating the influence of the orbit of Jupiter on the Milankovitch cycles at Earth - a first step in developing a means to characterise the nature of periodic climate change on planets beyond our Solar system.

Keywords: Astrobiology, Exoplanets, Exo-Earths, Habitability, Climate change, Jupiter, Milankovic cycles

Introduction

Ever since we first looked at the night sky, mankind has asked the question “Are we alone?”. Until just two decades ago, the answer to that question was solely the preserve of fiction, fantasy, speculation and faith – we knew of no planets beyond our Solar system. As such, it was an open question as to whether our Solar system was a fluke, miraculously alone in the universe, or whether planets were instead common around other stars.

The discovery of the first planets orbiting other Sun-like stars (e.g. [1][2][3]) was a watershed moment. Over the years, an ever-growing catalogue of exoplanets have been discovered¹, and have revealed that the variety of planets orbiting other stars is far more diverse than we could ever have imagined – ranging from Jupiter and Earth-sized planets that orbit within a million kilometres of their host star's surface (e.g. [4][5][6]), to planets orbiting binary stars (e.g. [7][8]), proposed “diamond planets” (e.g. [9][10]), and even a few moving on highly eccentric orbits reminiscent of the comets in our own Solar system (e.g. [11]). Coupled with our ever increasing understanding of our own Solar system (e.g. [12][13][14]), these discoveries have revolutionised our understanding of the formation and evolution of planetary systems (e.g. [15][16][17]).

¹ For an up-to-date tally of the number of known exoplanets, we direct the interested reader to the two main catalogues of exoplanets on the internet – <http://exoplanets.org/> and <http://exoplanet.eu/>. As of 1st November, 2013, these websites give the current tally of “confirmed” exoplanets as 755 and 1038, respectively.

In the coming years, it is highly likely that we will discover the first truly Earth-like planets² orbiting nearby stars, and the search for life beyond our Solar system will be able to begin in earnest. However, the observations required to detect evidence of life on Earth-like planets orbiting other stars will be hugely time-consuming and costly – which will in turn mean that we will only be able to focus on the few most promising targets, at least in the early part of that search. So how will we choose which of those exo-Earths we should target in the search for life? Clearly, the proximity of the various exo-Earths to our Solar system will be an important factor in any decision – the closer an exoplanet to Earth, the more widely it will be separated from its host star, and the easier it will therefore be to observe. There are, however, a variety of additional factors that will play an important role in determining which of the exo-Earths we discover are the most promising targets for the search for life (e.g. [18][19]).

One factor which will play an important role in any decision on exo-Earth suitability is the climatic stability of that planet. It is reasonable to assume that life will have the greatest chance of becoming established and thriving on a planet with a relatively quiescent climate. In contrast, planets whose climates oscillate dramatically and rapidly will likely represent poor targets for our initial search for life. It is well established that the Earth's climate varies on timescales of tens and hundreds of thousands of years as a result of variations in our planet's orbit driven by the gravitational influence of the other planets in our Solar system. The Earth's orbit twists and flexes as it is pulled around by the other massive objects in the Solar system, which results in long-term variations in the amount of energy the Earth receives from the Sun, averaged over any given year. These oscillations, known as the Milankovitch cycles, were first recognised in the late 19th and early 20th Century (e.g. [20][21][22]), and are thought to have driven the recent series of glacial and interglacial periods that have occurred over the last few million years (e.g. [23][24]).

Beyond the most recent ice age, there is evidence that the Earth has experienced at least four other periods of glaciation through its history. The Huronian glaciation, over two billion years ago, is thought to have lasted approximately 300 Myr, and may well have been linked to the evolution of photosynthesis (which would have removed a significant amount of greenhouse gas from the atmosphere, causing the planet to cool) [25]. More recently, the Sturtian and Marinoan glaciations (~750-700 Myr and 635 Myr ago; [26][27]) are thought to have been some of the most extensive glaciations in our planet's history – the archetypal “snowball Earth” events. It is almost certain that, during these periods of global glaciation, the Milankovitch cycles will have played an important role in driving the backward and forward march of the ice-caps on timescales of tens and hundreds of thousands of years. Indeed, studies of the glaciation that occurred at the end of the Ordovician and in the early Silurian periods (c.a. 445 Myr ago; [28]) have shown strong evidence of cyclical variations that can best be explained by the Milankovitch cycles of the time (e.g. [29]).

Whilst for the Earth, the periodic orbital variations induced by the other planets are relatively small, the same cannot be said for the planet Mercury. At the current epoch, Mercury moves on an orbit with eccentricity approximately 0.21. On timescales of hundreds of thousands of years, Mercury's orbital eccentricity varies dramatically – at times falling so low as to put the planet on a circular orbit, whilst at other times being driven as high as 0.45 (e.g. [30][31]). Were the Earth's orbital excursions equally extreme, there would be times when our planet's orbit would take it significantly closer to the Sun than the planet Venus, at perihelion, and out almost to the orbit of Mars, at aphelion. Such a situation would clearly have deleterious

² i.e. planets of comparable size and mass to the Earth, orbiting at a suitable distance from their host star such that liquid water could be present and stable upon their surface.

effects on our planet's climate, and would likely render it a less hospitable place for the development of life.

In this work, we present the preliminary results of a study that aims to categorise the influence of the giant planet Jupiter's orbit on the scale and speed of the Milankovitch cycles experienced by the Earth. The key goal of this work is to build expertise that will allow future studies to determine the Milankovitch cycles that would be experienced by any newly discovered exo-Earths, in order to help determine which would be the most promising targets for the search for life. In section two, we describe our methodology, before presenting our preliminary results in section three. Finally, in section four, we conclude with a discussion of the future direction of our research project.

Testing Jupiter's Role

To examine the influence of Jupiter's orbit on the amplitude and frequency of the Milankovitch cycles of the Earth, we used the Hybrid integrator within the n -body dynamics package MERCURY [32], which has been widely used to address questions in Solar system astronomy, exoplanetary science and astrobiology (e.g. [33][34][35]). In order to properly account for the effects of general relativity on the orbit of the planet Mercury, we used a version of the MERCURY code that had been modified to take account of the relativistic correction to Mercury's orbital motion. We then set up a total of 39,601 unique simulations, each of which ran for a simulation period of one million years, from the current epoch forward in time. In each simulation, the initial orbits of Mercury, Venus, Earth, Mars, Saturn, Uranus and Neptune were held constant at their current values. The initial orbit of Jupiter, however, was systematically varied in semi-major axis and eccentricity. In total, we tested 199 unique values of the initial Jovian semi-major axis, ranging from 4.20336301 to 6.20336301 AU, in equal steps (i.e. covering a range ± 1 AU from the orbit of Jupiter in our Solar system, at $a = 5.20336301$ AU). At each of these Jovian semi-major axis values, we test 199 unique orbital eccentricities, distributed evenly between 0.0 and 0.2 (for reference, the current value of Jupiter's orbital eccentricity is 0.04839266)³. The orbital elements for the eight planets were recorded at 100 year intervals for the duration of the 1 Myr integrations.

Once the simulations were complete, we used the results to create maps of the variability of the Earth's orbital elements as a function of Jupiter's initial orbit, building on earlier work creating dynamical maps of exoplanetary systems and the orbits of Solar system objects (e.g. [36][37]). These maps provide a quick visual guide to the degree of variability in the Earth's Milankovitch cycles that can result from small scale changes to the orbit of Jupiter, and we present a number of examples of such plots in the next section.

Preliminary Results

In Figure 1, we present four exemplar plots that reveal how moderate changes to Jupiter's orbit can have significant effects on both the periodicity and the amplitude of the Milankovitch cycles experienced by the Earth. In that figure, the left hand panels show the evolution of the Earth's orbital eccentricity (top) and inclination (bottom) for a period of one million years in a Solar system where Jupiter occupies its current orbit around the Sun. The right hand panels show the variation of the same variables for a Solar system that differs from the first only in the initial eccentricity of Jupiter's orbit. In the right hand panels, rather than

³ The orbital elements for the eight planets were taken from <http://www.met.rdg.ac.uk/~ross/Astronomy/Planets.html>, and are valid for epoch JD 2451545.0 (i.e. January 1.5, 2000, UT).

Jupiter starting on the low eccentricity orbit we see in our Solar system ($e \approx 0.048$), the giant planet instead moved on an orbit with an initial eccentricity 0.2 (still far less than the eccentricities proposed for a number of candidate exoplanets; e.g. [38][39]). As can be seen, the variations in Earth's orbital eccentricity differ greatly in scale between the two scenarios (with the high eccentricity Jupiter driving oscillations approximately five times larger than those in our Solar system). On the other hand, inclination variations are comparable in scale between the two scenarios, but the more eccentric Jupiter forces the Earth's orbital inclination to vary at a higher frequency than observed in the Solar system. Whilst these are just the results of two simple simulations, they serve to illustrate the effect that variations in the orbits of the other planets can have on the Milankovitch cycles at Earth.

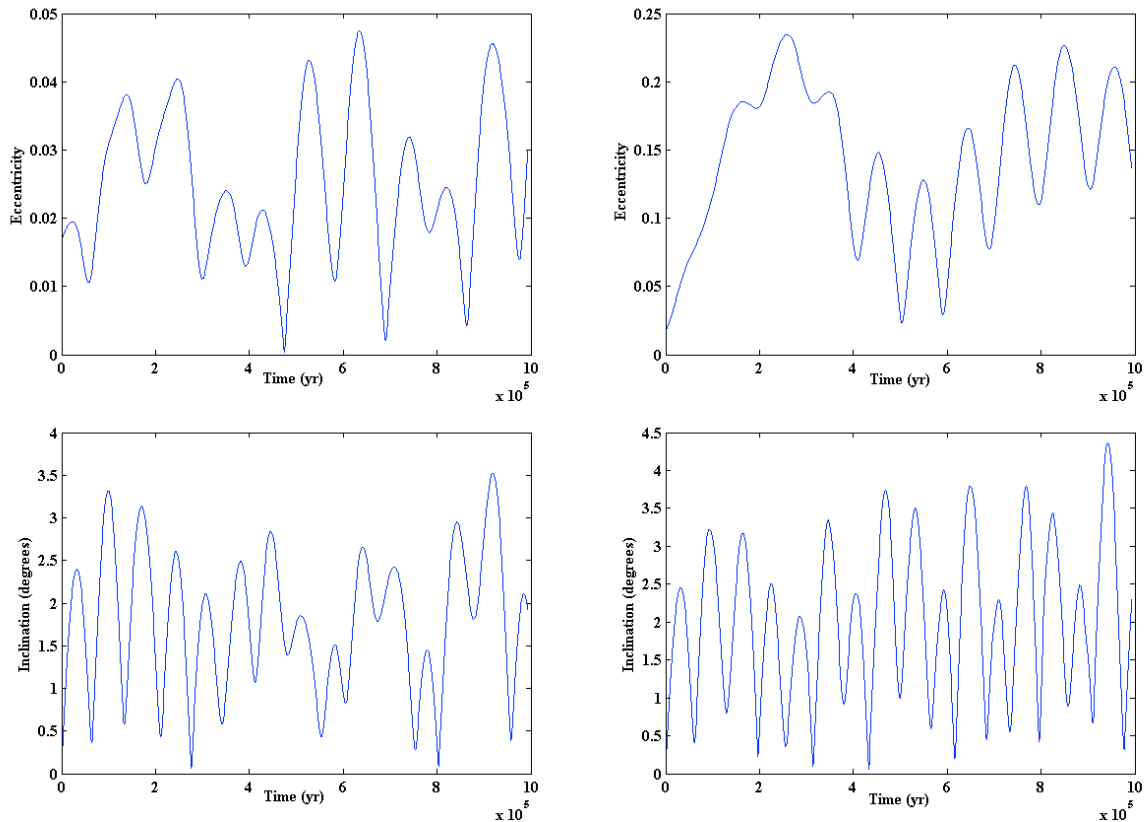


Fig. 1: The evolution of the eccentricity (top) and inclination (bottom) of the Earth's orbit, for a period of one million years. The left hand plots show the variations in eccentricity and inclination for the Solar system as we know it today, with Jupiter on its current orbit. Those to the right show the variability if Jupiter instead moved on an initial orbit with eccentricity 0.2, but with all other orbital parameters the same as their current values.

In Figure 2, we show the variation in the maximum value of orbital eccentricity achieved by the Earth over the one million years of integration time, as a function of Jupiter's orbital eccentricity and semi-major axis. Since the maximum eccentricities achieved by the Earth spanned almost two orders of magnitude, the data is plotted in a logarithmic colour scale, with the most eccentric solutions being displayed in red, and the least eccentric shown in blue. It is immediately apparent that the evolution of Earth's eccentricity varies dramatically as a function of Jupiter's initial orbit. More surprisingly, the plot reveals a great deal of fine structure in the eccentricity evolution of the Earth. Rather than the smooth variation from more stable to less stable orbits that one might expect as Jupiter is moved around, we instead see bands of relatively strong instability and stability running from left to right across the

figure. The broad pattern is simple and relatively clear cut – as Jupiter is moved outward, towards the orbit of Saturn, the typical maximum eccentricities experienced by the Earth increase. Similarly, there are more cases where the Earth’s eccentricity is driven to large values when Jupiter starts on an eccentric orbit than when it begins on a near circular orbit. However, the fine structure revealed in Fig. 2 shows that, beyond these general, broad brush stroke results, the influence of Jupiter on Earth’s orbital eccentricity is actually remarkably complex. There are regions with relatively small excursions for scenarios where Jupiter is both distant from the Sun and relatively eccentric (such as the tongue of blue that extends to the right at an initial eccentricity of ~ 0.16), and others where the Earth experiences large orbital eccentricities when Jupiter is on a near-circular orbit, close in to the Sun (e.g. the tongue of yellow extending to the left at the bottom of the figure). The cause of this fine structure remains to be tied down, but it seems most likely to be the result of long-term secular effects – and possibly the shifting of multi-body secular resonances as a function of Jupiter’s orbital eccentricity (e.g. [40]). Were the variations instead the result of mean-motion resonant⁴ interactions between Jupiter and other planets in the system, one would instead expect to see vertical structure in the figure, since mean-motion resonances are purely a function of heliocentric distance. It is also apparent that, although our Solar system (marked by the hollow circle) falls in a region of only moderate variability in eccentricity for the Earth, there are many architectures for the Solar system that would feature significantly smaller excursions in eccentricity for our planet.

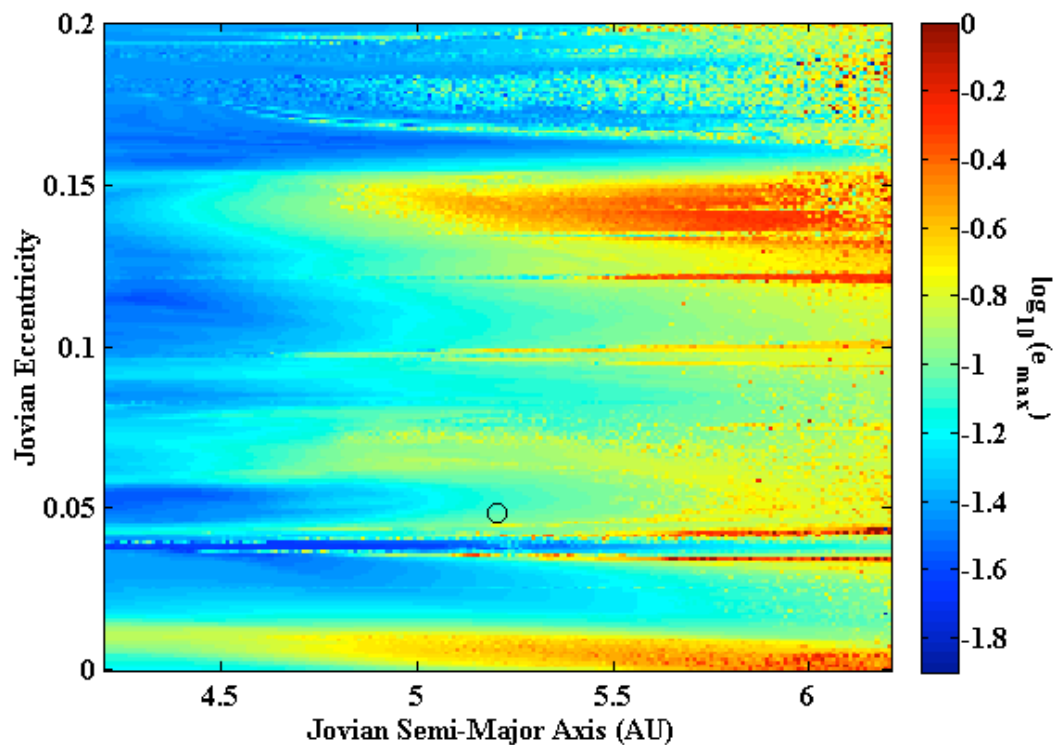


Fig. 2: The maximum eccentricity of the Earth’s orbit obtained over a period of one million years as a function of Jupiter’s initial orbital eccentricity and semi-major axis. For each of the 39,601 simulations shown here, the only variables changed in the initial conditions were

⁴ A mean motion resonance between two bodies occurs when their orbital periods are an integer ratio of one another. The most famous mean motion resonance in our Solar system is that between Neptune and the dwarf planet Pluto. Pluto completes two orbits in the time it takes Neptune to complete three. Even though the orbit of Pluto crosses that of Neptune, their commensurate orbital periods prevent their ever experiencing sufficiently close encounters to disrupt their orbits. Further discussion of resonant orbital behaviour is beyond the scope of this work, but we direct the interested reader to e.g. [41], [42] and [43] for further discussions of orbital resonance and our Solar system.

Jupiter's semi-major axis and eccentricity – the initial orbits of the other planets were held constant across the suite of integrations. The hollow circle shows the location of Jupiter's orbit within our Solar system.

In Figure 3, we present the variability of the maximum inclination of the Earth's orbit as a function of Jupiter's initial semi-major axis and eccentricity. Here, the variability is much less pronounced than was the case for the Earth's orbital eccentricity. Despite this, a number of similar features are shown, with fingers of increased inclination variability running from left to right in the plot. Once again, a concentration of solutions that feature relatively extreme excursions in the Earth's orbital inclination are concentrated at low Jovian eccentricities, for scenarios where the giant planet is located beyond around 5.5 AU from the Sun.

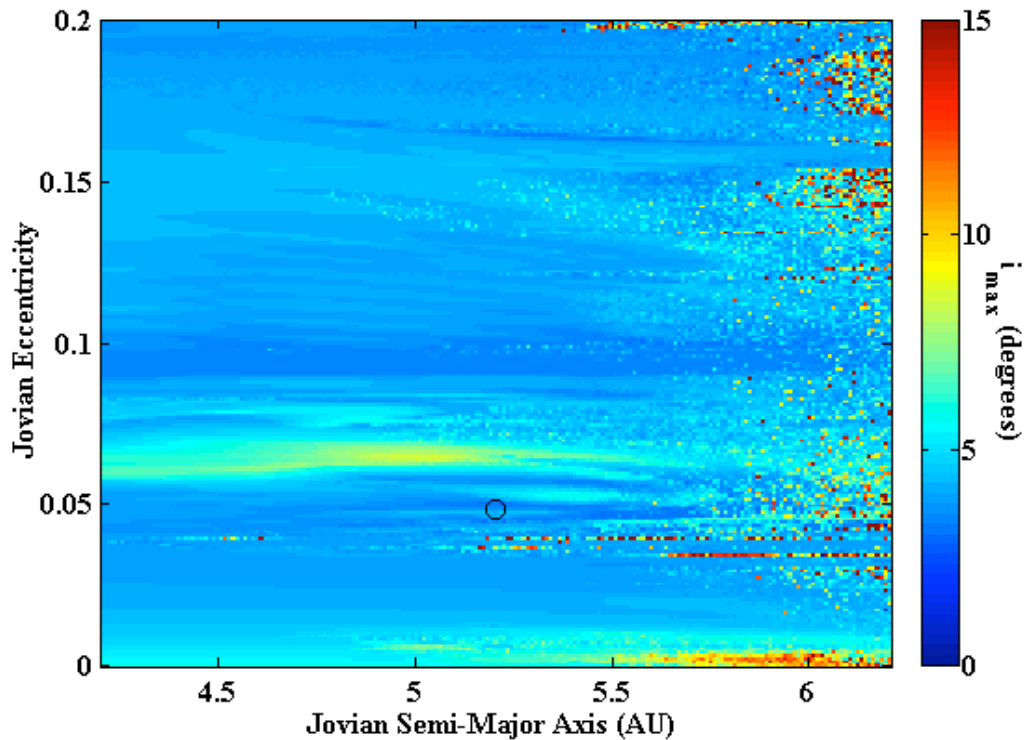


Fig. 3: The maximum inclination of Earth's orbit over a period of one million years as a function of Jupiter's initial orbital eccentricity and semi-major axis. The hollow circle again marks the location of Jupiter's orbit in our own Solar system.

Figure 4 shows the manner in which the maximum rate of change of the Earth's orbital semi-major axis varies as a function of Jupiter's eccentricity and orbital radius. In the great majority of cases, the Earth did not experience significant excursions in orbital radius – but although the results here are significantly more noisy than those presented in Figures 2 and 3, similar features are once again visible, with fingers of increased instability stretching from right to left across the figure. It is interesting to compare the results shown in Figures 2 and 4. There are a number of similarities between the two (such as the region of enhanced variability that surrounds the location of our Solar system on three sides). However, although the general structure of the two figures is very similar, it is noticeable that the strength of the different features in different locations varies significantly between the two plots. Compare, for example, the two broad regions of enhanced variability at both high Jovian eccentricity and semi-major axis, at the top right hand corner of both figures. Though the two bands of variability that are visible there display the same sculpting in both figures, the lower of the two results in far more intense variability in the eccentricity of the Earth's orbit, whilst the upper yields far faster variability in the Earth's semi-major axis.

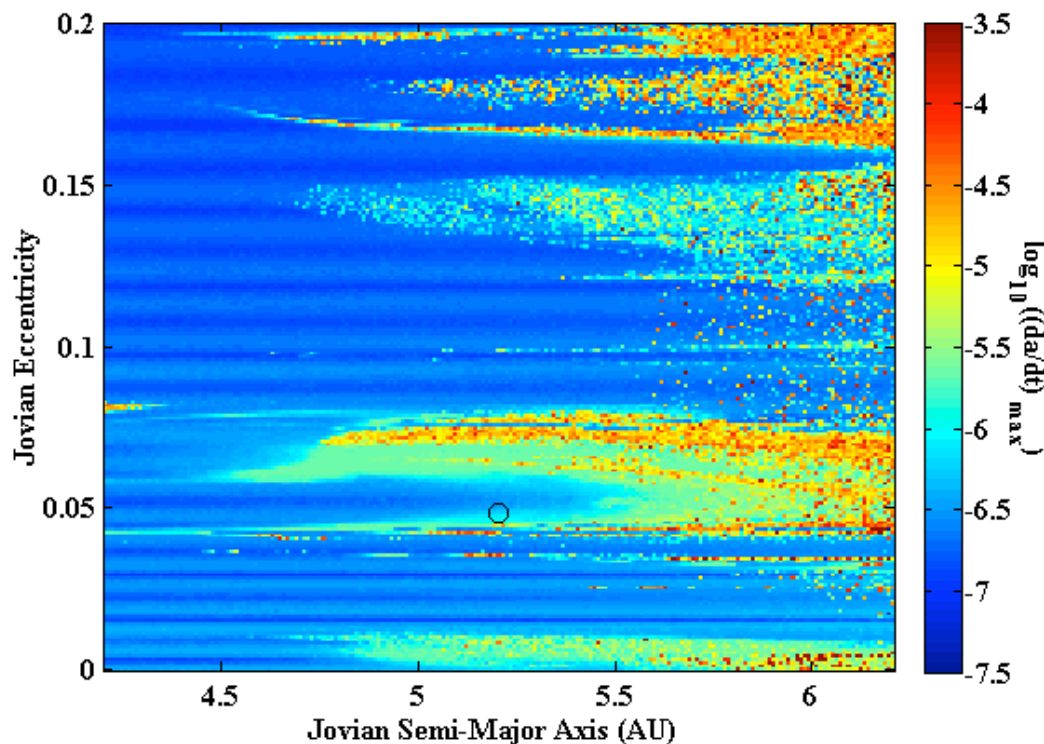


Fig. 4: The maximum rate of change of the Earth's semi-major axis over a period of one million years, as a function of Jupiter's initial orbital eccentricity and semi-major axis, in units of $\log_{10} (AU/yr)$. Variations in the semi-major axis of the Earth were relatively small, but the figure shows how minor variations in Jupiter's initial orbit can cause the rate of that change to vary by up to four orders of magnitude.

Conclusions

We have presented the preliminary results of a study that will investigate the influence of the orbit of the giant planet Jupiter on the Milankovitch cycles experienced by the Earth. Our results show that both the amplitude and frequency of the Milankovitch cycles would vary strongly as a function of Jupiter's semi-major axis and eccentricity. Whilst one might expect that those variations would be relatively smooth and easy to predict, we instead find that a great deal of fine structure is present – with Solar systems that differ by only minor changes in Jupiter's semi-major axis or eccentricity displaying significant differences in the strength and periodicity of the Milankovitch cycles.

The next step in this work is to carry out more detail, and longer, simulations of our Solar system. It is possible that our choice to limit our preliminary tests to just one million years of run time will have abridged the longer-period oscillations for some of the more stable systems tested. As a result, we intend to perform simulations that last for an order of magnitude longer over the coming months, using UNSW's *Katana* supercomputing cluster. It is also apparent from our plots (Figures 2 – 4), that there is significant noise around the more unstable solutions tested. We note that a number of our integrations resulted in critically unstable systems, in which one or other of the Solar system's planets were ejected or collided with one another within our 1 million year time frame. Clearly, such unstable systems can result in dramatic variations in the Earth's orbital elements, even when it is not the planet removed in this way – and this is the main reason for the speckle visible towards the highest eccentricities and semi-major axes. We will therefore more than quadruple our resolution in the new suite

of runs – testing 159,201 unique architectures for our planetary system, rather than the 39,601 presented in this work. In addition to calculating the variation of the Earth’s orbital elements as a function of time, we will also examine the precession of Jupiter’s orbit, which will no doubt vary significantly as a function of its initial semi-major axis and eccentricity. Given the precessional frequency of Jupiter’s orbit, it will be possible to determine the stability (or instability) of the Earth’s obliquity, following the procedure detailed in [44].

Once those simulations are complete, we will collaborate with colleagues at UNSW’s Climate Change Research Centre to bring together our dynamical models of the Solar system with simple climate models, in order to assess the degree to which the observed changes in the Earth’s orbit across the variety of Jovian orbital architectures studied would affect the climate of the Earth. This will allow us to get a true handle on the importance of planetary architecture as a driver for the climate of exo-Earths, allowing us to build a toolset that will facilitate the study of such planets as and when they are discovered in the coming years.

The long-term goal of this work is to provide a mechanism by which the climatic variability of newly discovered exo-Earths can be estimated, allowing the potential habitability of those worlds to be assessed. Such assessment will form a critical component of the target selection process for the search for life on planets beyond our Solar system. Since the observations that will be required in order to carry out that search will be extremely challenging, it will not be possible to search more than the best few candidates, and so it is imperative that every avenue be explored in order to ensure that we can best select the most promising targets to go forward with the search (e.g. [18][19]).

Acknowledgments

The work was supported by iVEC through the use of advanced computing resources located at the Murdoch University, in Western Australia.

References

1. Mayor, M. and Queloz, D., 1995, “A Jupiter-mass companion to a solar-type star”, *Nature*, 378, 355-359
2. Marcy, G. W. and Butler, R. P., 1996, “A Planetary Companion to 70 Virginis”, *Astrophysical Journal Letters*, 464, L147
3. Butler, R. P. and Marcy, G. W., 1996, “A Planet Orbiting 47 Ursae Majoris”, *Astrophysical Journal Letters*, 464, L153
4. Hellier, C., Anderson, D. R., Collier Cameron, A., Gillon, M., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smalley, B., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Udry, S. and West, R. G., 2011, “WASP-43b: the closest-orbiting hot Jupiter”, *Astronomy and Astrophysics*, 535, article id. L7
5. Howard, A. W., Sanchis-Ojeda, R., Marcy, G. W., Johnson, J. A., Winn, J. N., Isaacson, H., Fischer, D. A., Fulton, B. J., Sinukoff, E. and Fortney, J. J., 2013, “A rocky composition for an Earth-sized exoplanets”, *Nature*, early online access: <http://www.nature.com/nature/journal/vaop/ncurrent/full/nature12767.html>

6. Pepe, F., Collier Cameron, A., Latham, D. W., Molinari, E., Udry, S., Bonomo, A. S., Buchhave, L. A., Charbonneau, D., Cosentino, R., Dressing, C. D., Dumusque, X., Figueria, P., Fiorenzano, A. F. M., Gettel, S., Harutyunyan, A., Haywood, R. D., Horne, K., Lopez-Morales, M., Lovis, C., Malavolta, L., Mayor, M., Micela, G., Motalebi, F., Nascimbeni, V., Phillips, D., Piotto, G., Pollacco, D., Queloz, D., Rice, K., Sasselov, D., Segransan, D., Sozzetti, A., Szentgyorgyi, A. and Watson, C. A., 2013, "An Earth-sized planet with an Earth-like density", *Nature*, early online access: <http://www.nature.com/nature/journal/vaop/ncurrent/full/nature12768.html>
7. Doyle, L. R., Carter, J. A., Fabrycky, D. C., Slawson, R. W., Howell, S. B., Winn, J. N., Orosz, J. A., Prcaronsa, A., Welsh, W. F., Quinn, S. N., Latham, D., Torres, G., Buchhave, L. A., Marcy, G. W., Fortney, J. J., Shporer, A., Ford, E. B., Lissauer, J. J., Ragozzine, D., Rucker, M., Batalha, N., Jenkins, J. M., Borucki, W. J., Koch, D., Middour, C. K., Hall, J. R., McCauliff, S., Fanelli, M. N., Quintana, E. V., Holman, M. J., Caldwell, D. A., Still, M., Stefanik, R. P., Brown, W. R., Esquerdo, G. A., Tang, S., Furesz, G., Geary, J. C., Berlind, P., Calkins, M. L., Short, D. R., Steffen, J. H., Sasselov, D., Dunham, E. W., Cochran, W. D., Boss, A., Haas, M. R., Buzasi, D., and Fischer, D., 2011, "Kepler-16: A Transiting Circumbinary Planet", *Science*, 333, 1602
8. Welsh, W. F., Orosz, J. A., Carter, J. A., Fabrycky, D. C., Ford, E. B., Lissauer, J. J., Prvsa, A., Quinn, S. N., Ragozzine, D., Short, D. R., Torres, G., Winn, J. N., Doyle, L. R., Barclay, T., Batalha, N., Bloemen, S., Brugamyer, E., Buchhave, L. A., Caldwell, C., Caldwell, D. A., Christiansen, J. L., Ciardi, D. R., Cochran, W. D., Endl, M., Fortney, J. J., Gautier, III, T. N., Gilliland, R. L., Haas, M. R., Hall, J. R., Holman, M. J., Howard, A. W., Howell, S. B., Isaacson, H., Jenkins, J. M., Klaus, T. C., Latham, D. W., Li, J., Marcy, G. W., Mazeh, T., Quintana, E. V., Robertson, P., Shporer, A., Steffen, J. H., Windmiller, G., Koch, D. G. and Borucki, W. J., 2012, "Transiting circumbinary planets Kepler-34 b and Kepler-35 b" *Nature*, 481, 475-479
9. Bailes, M., Bates, S. D., Bhalerao, V., Bhat, N. D. R., Burgay, M., Burke-Spolaor, S., D'Amico, N., Johnston, S., Keith, M. J., Kramer, M., Kulkarni, S. R., Levin, L., Lyne, A. G., Milia, S., Possenti, A., Spitler, L., Stappers, B. and van Straten, W., 2011, "Transformation of a Star into a Planet in a Millisecond Pulsar Binary", *Science*, 333, 1717
10. Madhusudhan, N., Lee, K. K. M. and Mousis, O., 2012, "A Possible Carbon-rich Interior in Super-Earth 55 Cancri e", *The Astrophysical Journal Letters*, 759, L40
11. O'Toole, S. J., Tinney, C. G., Jones, H. R. A., Butler, R. P., Marcy, G. W., Carter, B. and Bailey, J., 2009, "Selection functions in Doppler planet searches", *Monthly Notices of the Royal Astronomical Society*, 392, 641-654
12. Malhotra, R., 1995, "The Origin of Pluto's Orbit: Implications for the Solar System Beyond Neptune", *Astronomical Journal*, 110, 420
13. Morbidelli, A., Levison, H. F., Tsiganis, K. and Gomes, R., 2005, "Chaotic capture of Jupiter's Trojan asteroids in the early Solar System", *Nature*, 435, 462-465
14. Lykawka, P. S., Horner, J., Jones, B. W. and Mukai, T., 2009, "Origin and dynamical evolution of Neptune Trojans – I. Formation and planetary migration", *Monthly Notices of the Royal Astronomical Society*, 398, 1715-1729

15. Thommes, E. W., Duncan, M. J. and Levison, H. F., 2003, "Oligarchic growth of giant planets", *Icarus*, 49, 195-236
16. Mousis, O., Alibert, Y., Hestroffer, D., Marboeuf, U., Dumas, C., Carry, B., Horner, J. and Selsis, F., 2008, "Origin of volatiles in the main belt", *Monthly Notices of the Royal Astronomical Society*, 383, 1269-1280
17. Fogg, M. J. and Nelson, R. P., 2007, "The effect of type I migration on the formation of terrestrial planets in hot-Jupiter systems", *Astronomy and Astrophysics*, 472, 1003-1015
18. Horner, J. and Jones, B. W., 2010, "Determining habitability: which exoEarths should we search for life?", *International Journal of Astrobiology*, 9, 273-291
19. Horner, J., 2013, "Beyond the Habitable Zone – Dynamics and Habitability", *Life*, *submitted*
20. Croll, J., 1875, "Climate and time in their geological relations; a theory of secular changes of the earth's climate", Published by Daldy, Tsbister & co., London
21. Köppen, W., Wegener, A., 1924, "Die Klimate der Geologischen Vorzeit", Gebrüder Borntraeger, Berlin.
22. Milankovitch, M., 1941, "Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeiten-problem", R. Serbian Acad., Belgrade
23. Roe, G., 2006, "In defense of Milankovitch", *Geophysical Review Letters*, 33, L24703
24. Hays, J. D., Imbrie, J. and Shackleton, N. J., 1976, "Variations in the Earth's Orbit: Pacemaker of the Ice Ages", *Science*, 194, 1121-1132
25. Kopp, R. E., Kirschvin, J. L., Hilburn, I. A. and Nash, C. Z., 2005, "The Paleoproterozoic snowball Earth: A Climate disaster triggered by the evolution of oxygenic photosynthesis", *Proceedings of the National Academy of Sciences of the United States of America*, 102, 11131-11136
26. Preiss, W. V., Gostin, V. A., McKirdy, D. M., Ashley, P. M., Williams, G. E. and Schmidt, P. W., 2011, "Chapter 69 The glacial succession of Sturtian age in South Australia: the Yudnamutana Subgroup", *Geological Society, London, Memoirs*, 36, 701-712
27. Shields, G. A., 2008, "Palaeoclimate: Marinoan meltdown", *Nature Geoscience*, 1, 351-353
28. Finnegan, S., Bergmann, K., Eiler, J. M., Jones, D. S., Fike, D. A., Eisenman, I., Hughes, N. C., Tripathi, A. K. and Fischer, W. W., 2011, "The magnitude and duration of Late Ordovician-Early Silurian glaciation", *Science*, 331, 903-906
29. Williams, G. E., 1991, "Milankovitch-band cyclicity in bedded halite deposits contemporaneous with Late Ordovician-Early Silurian glaciation, Canning Basin, Western Australia", *Earth and Planetary Science Letters*, 103, 143-155

30. Strom, R. G., Sprague, A. L., “Exploring Mercury: the iron planet”, eds. Strom, R. G., Sprague, A. L., Springer-Praxis Books in Astronomy and Space Sciences, London (UK): Springer, Chichester (UK): Praxis Publishing. ISBN 1-85233-731-1, 2003
31. Correia, A. C. M., Laskar, J., 2009, “Mercury's capture into the 3/2 spin-orbit resonance including the effect of core-mantle friction”, *Icarus*, 201, 1-11
32. Chambers, J. E., 1999, “A hybrid symplectic integrator that permits close encounters between massive bodies”, *Monthly Notices of the Royal Astronomical Society*, Volume 304, pp. 793 – 799
33. Horner, J., Lykawka, P. S., Bannister, M. T., Francis, P., 2012, “2008 LC18: a potentially unstable Neptune Trojan”, *Monthly notices of the Royal Astronomical society*, 422, 2145-2151
34. Robertson, P., Endl, M., Cochran, W. D., MacQueen, P. J., Wittenmyer, R. A., Horner, J., Brugamyer, E. J., Simon, A. E., Barnes, S. I., Caldwell, C., “The McDonald Observatory Planet Search: New Long-period Giant Planets and Two Interacting Jupiters in the HD 155358 System”, *Astrophysical Journal*, 739, article id. 39
35. Horner, J., Jones, B. W., 2009, “Jupiter – friend or foe? II: the Centaurs”, *International Journal of Astrobiology*, 8, 75-80
36. Horner, J., Lykawka, P. S., 2010, “2001 QR322: a dynamically unstable Neptune Trojan?”, *Monthly Notices of the Royal Astronomical Society*, 405, 49-56
37. Horner, J., Wittenmyer, R. A., Hinse, T. C., Tinney, C. G., 2012, “A detailed investigation of the proposed NN Serpentis planetary system”, *Monthly Notices of the Royal Astronomical Society*, 425, 749-756
38. Wittenmyer, R. A., Endl, M., Cochran, W. F., Levison, H. F. and Henry, G. W., 2009, “A Search for Multi-Planet Systems Using the Hobby-Eberly Telescope”, *The Astrophysical Journal Supplement*, 182, 97-119
39. Moutou, C., Hébrard, G., Bouchy, F., Eggenberger, A., Boisse, I., Bonfils, X., Gravallon, D., Ehrenreich, D., Forveille, T., Delfosse, X., Desort, M., Lagrange, A.-M., Lovis, C., Mayor, M., Pepe, F., Perrier, C., Pont, F., Queloz, D., Santos, N. C., Ségransan, D., Udry, S. and Vidal-Madjar, A., 2009, “Photometric and spectroscopic detection of the primary transit of the 111-day-period planet HD 80 606 b”, *Astronomy and Astrophysics*, 498, L5-L8
40. Minton, D. A., Malhotra, R., 2011, “Secular Resonance Sweeping of the Main Asteroid Belt During Planet Migration”, *The Astrophysical Journal*, 732, article id. 53
41. Murray, C. D., Dermott, S. F., “Solar system dynamics”, ed. Murray, C. D., Cambridge University Press, 1999
42. Horner, J., Jones, B. W., 2008, “Jupiter friend or foe? I: The asteroids”, *International Journal of Astrobiology*, 7, 251-261
43. Lykawka, P. S., 2012, “Trans-Neptunian Objects as Natural Probes to the Unknown Solar System”, *Monographs on Environment, Earth and Planets*, 1, 121-186

44. Waltham, D., 2006, "The large-moon hypothesis: can it be tested?", *International Journal of Astrobiology*, 5, 327-331

Ionospheric Phenomena and Low-Frequency Radio Astronomy.

David Herne¹, John Kennewell¹, Mervyn Lynch¹ and Charles Carrano².

¹*Curtin Institute for Radio Astronomy (CIRA), Curtin University, Perth, Western Australia,*

²*Boston College, Boston, Massachusetts, U.S.A.*

Summary: The Murchison Widefield Array radio telescope (MWA), situated on the Murchison Radio Observatory (MRO) in Western Australia, has recently commenced operations. This instrument operates over the frequency range 80-300 MHz. Further, the MRO is also the site chosen to host the low-frequency component of the Square Kilometre Array, radio telescope (SKA). Each instrument is susceptible to scintillation caused by fluctuations in ionospheric plasma density and Faraday rotation of incoming signals caused by the interaction of low-frequency radio waves with dissociated electrons in the ionosphere. Observations of these parameters over several years, across periods of both subdued and elevated solar activity have demonstrated markedly differing regimes. High-precision GPS systems, combined with purpose-written data acquisition software (SCINDA), have enabled investigation of various phenomena including the effect of solar storms on the ionosphere at highly resolved time-scales. We report on aspects of phenomena observed and their significance to low-frequency radio astronomy and note that conditions of very low scintillation encountered support the decision to site world-leading instruments on the MRO.

Keywords: ionosphere, radio-astronomy, scintillation, TEC, MWA, SKA.

Low-Frequency Radio Astronomy at the Murchison Radio Observatory

The MWA is a low-frequency, aperture-synthesis, tile-based array that comprises 2,048 crossed-dipole antennas mounted on 128 tiles arrayed in a dithered Reuleaux formation, (Lonsdale, Cappallo et al. 2009) (Tingay, Goeke et al. 2012). The MWA achieved first light with an array of three tiles located on Mileura Station in March 2005 (Bowman, Barnes et al. 2007), (Bhat, Wayth et al. 2007) and is now a commissioned observatory.



Figure 1: A tile of the Murchison Widefield Array radio telescope, consisting of 16 crossed dipole antennas.

The MWA is an exemplar for facilities that are currently operational, such as LOFAR in the Netherlands, or being planned, such as the Square Kilometre Array, low-frequency component, planned also for the MRO. The requirements of this instrument in respect of calibration of the ionosphere are typical for comparable arrays.

Radio Telescope Ionospheric Calibration

The MWA requires the ionospheric contribution to image degradation be mitigated through accurate calibration. Self-calibration employs a process called peeling and is sufficient for certain classes of observation (Mitchell, Greenhill et al. 2008). Peeling in this implementation is a real-time interferometric technique that has been observed to perform well under various ionospheric conditions. The calibration system of the MWA measures the apparent angular offsets of calibrators induced by the ionosphere. Models of the ionosphere and of instrument response are fitted to images to remove offsets. This operation proceeds with a cadence of 8 seconds. Data streaming from the array are not stored due to the volumes involved (19 Gbs⁻¹ from the correlator). Corresponding maps of ionospheric relative, total electron content (TEC) are produced. This capability also promotes use of the MWA in mapping of the ionosphere. Absolute calibration, which requires accurate knowledge of the ionospheric absolute total electron content (aTEC), is required for other important classes of observation, in particular those involving measures of Faraday rotation such as galactic magnetic field studies (Bowman, Cairns et al. 2013). The polarisation angle of a linearly polarised radio signal is rotated in transiting the ionosphere, proportional to frequency of the signal. The amount of rotation, the rotation measure (RM), is a function of ionospheric aTEC and the magnetic field strength integrated along the line of sight (Oberoi and Lonsdale 2012):

$$RM = \frac{e^3}{8\pi^2 \epsilon_0 m^2 c^3} \int_{Obs}^{Src} N_e(\mathbf{r}) \mathbf{B}(\mathbf{r}) \cdot d\mathbf{s} \quad (1)$$

where $N_e(\mathbf{r})$ is the electron density at the position vector \mathbf{r} and \mathbf{B} is the magnetic field strength along the line of sight at location \mathbf{r} . Therefore, knowledge of the free electron content of the ionosphere is required in order to determine the ionosphere's contribution to signal Faraday rotation, given as:

$$FR = RM \times \lambda^2 \quad (2)$$

Absolute TEC measured by us over several years is shown in Figures 2(a) and 2(c). However, the sky visible to the MWA exhibits varying qualities across a large field of view (610 deg²@ 150 MHz). Fluctuations in electron density induce amplitude and phase variations in signals received at the observatory (Kung Chie and Chao-Han 1982). Scintillation measured by dual-frequency GPS systems over the same interval as above, are shown in Figures 2(b) and 2(d). These are calculated as the normalised variance of the intensity of this received signal I :

$$S_4 = \sqrt{\frac{\langle I^2 \rangle - \langle I \rangle^2}{\langle I \rangle^2}} \quad (3)$$

and known as the S_4 index.

Tables 1(a) and 1(b) display effects imposed by the ionosphere on radio astronomy:

Ionospheric Effects on Radio Astronomy
First Order Effect
- Plasma opacity
Second Order Effects
- Refraction
- Dispersion
- Faraday Rotation
Third Order Effects
- Scintillations
- De-coherence
- Variable refraction
- Phase stability
Fourth Order Effects
- Emission / Radiation

Typical Ionospheric Changes in RA Measures		
Quantity	Typical value @100MHz	Frequency Dependence
Refraction	1.5 arcminutes	$1/f^2$
Polarisation	10 radians	$1/f^2$
Phase change	1000 radians	$1/f$
Path length	500 metres	$1/f^2$
Absorption	0.01 dB	$1/f^2$

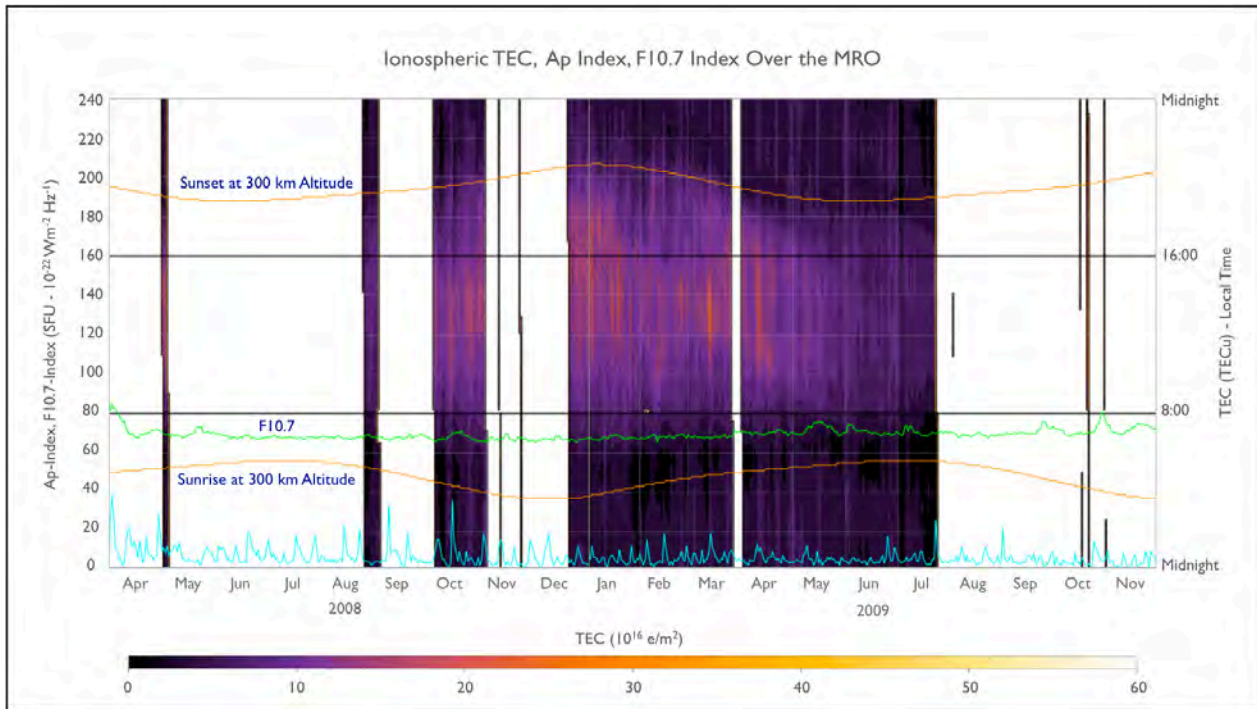
Tables 1(a), 1(b): Effects imposed by the ionosphere on radio astronomy and corresponding changes in radio astronomical measures.

The Southern, Mid-latitude Ionosphere

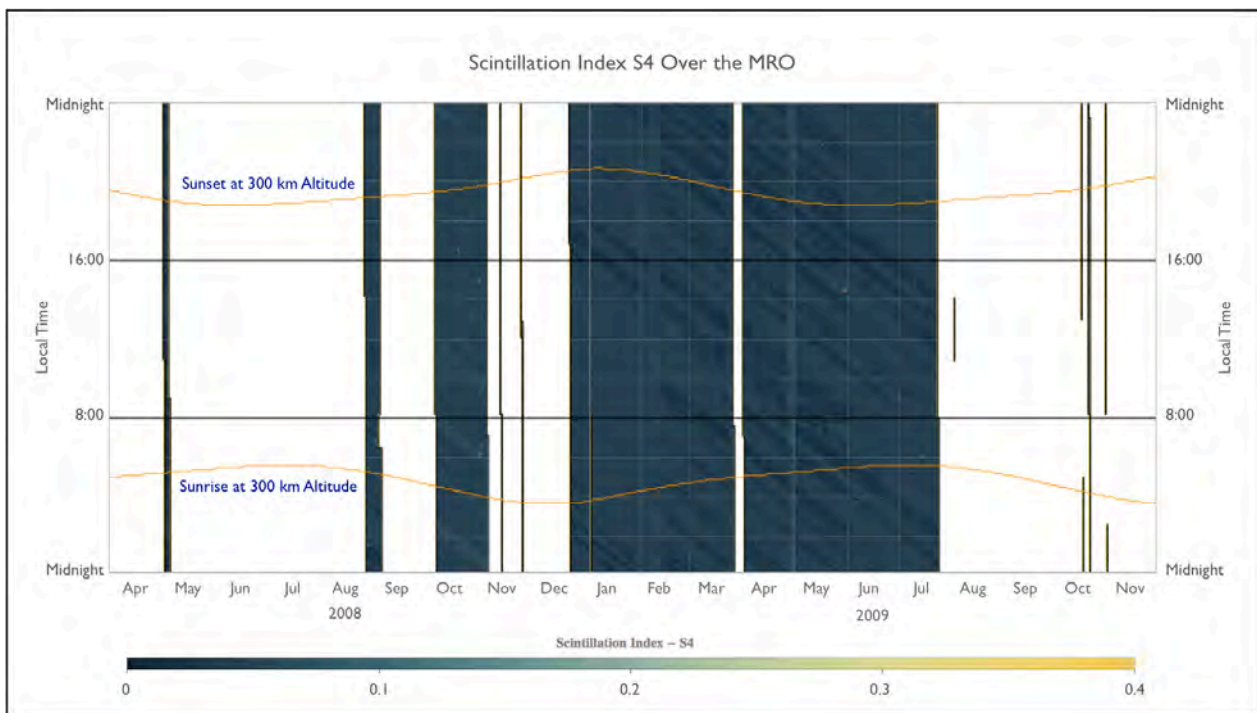
The southern, mid-latitude ionosphere has been studied over an extensive period, (Bowman 1981, Hajkowicz 1994). When therefore, recommendations were sought regarding suitable locations in Australia for siting of the Square Kilometre Array (SKA) radio telescope, the region in the mid-west of Western Australia that was eventually chosen was expected to exhibit very low levels of scintillation (Kennewell, Caruana et al. 2005) and such is our experience. As part of the initial site selection process, TEC and scintillation were modeled at the location of the Murchison Radio Observatory. Under conditions of low sunspot number, which existed at the time of this initial study (as opposed to current conditions, under which this study is continuing) TEC values were expected to be quite low, up to 20 TECU over an annual cycle. Corresponding scintillation was expected to be extremely subdued and smooth. The location for the observatory was chosen in part because, as a mid-latitude, southern hemisphere site, low ionospheric turbulence was expected. However, the total electron content of the ionosphere varies markedly not only over the course of a day but annually and over a solar (activity) cycle, as displayed below in Figures 2(a) and 2(c).

As shown in Figures 2(a) and 2(c), the index F10.7 is a measure of the sun's electromagnetic activity at wavelengths that produce photoionisation of the ionosphere, the index Ap is the three-hourly measure of geomagnetic activity.

Whereas a great deal of fine structure is observed in the plots in Figures 2(a) and 2(c), variation in parameters over much smaller scales than those discernible here occur from moment to moment over the course of a single day or a few days (Fig. 3).

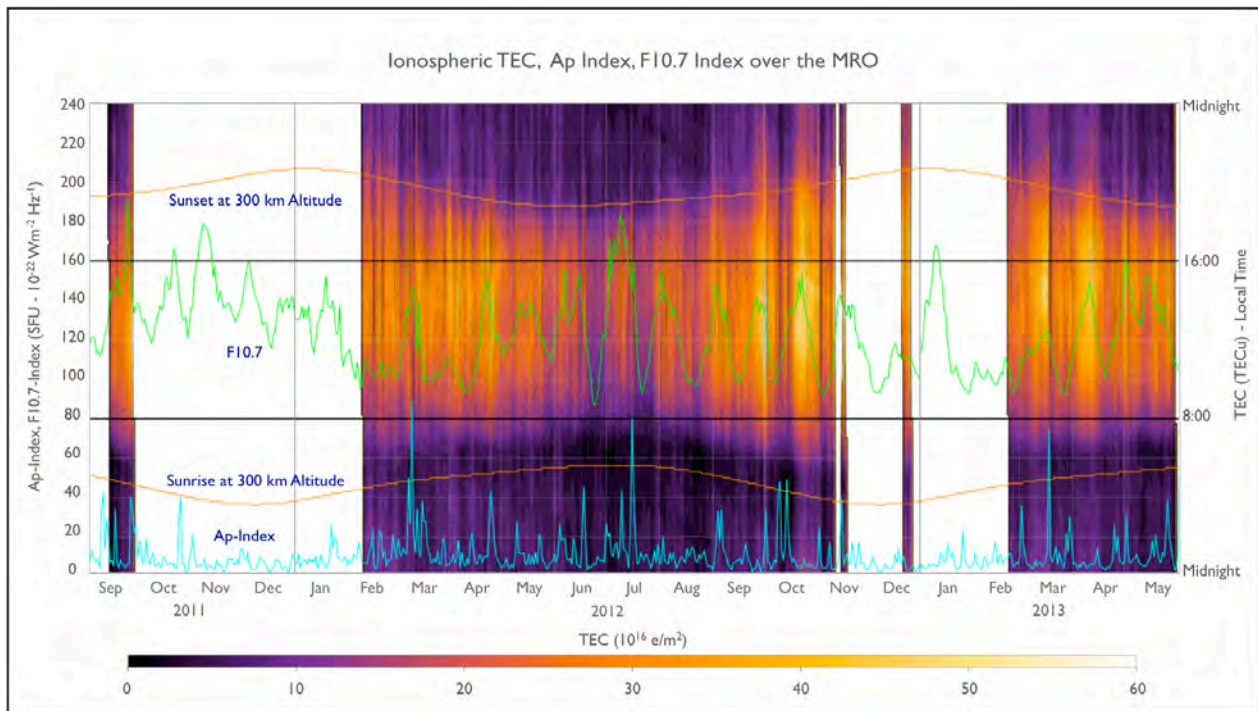


(a)

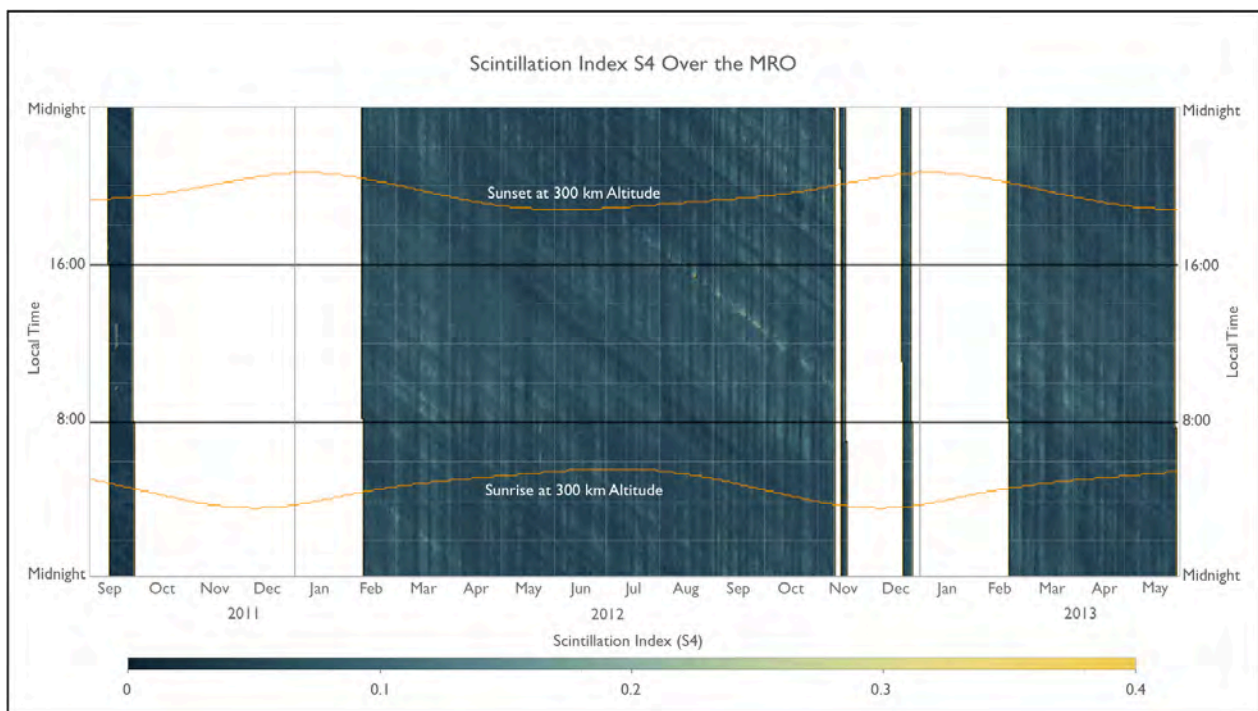


(b)

Figures 2(a), 2(b): Absolute total electron content (aTEC) and scintillation indices (S_4) (consecutively) of the ionosphere above the Murchison Radio Observatory (MRO) during an epoch of low solar activity over the period 2008-2009.



(c)



(d)

Figures 2(c), 2(d): Absolute total electron content (aTEC) and scintillation indices (S4) (consecutively) of the ionosphere above the Murchison Radio Observatory (MRO) during an epoch of high solar activity over the period 2012-2013 (continuing into the present).

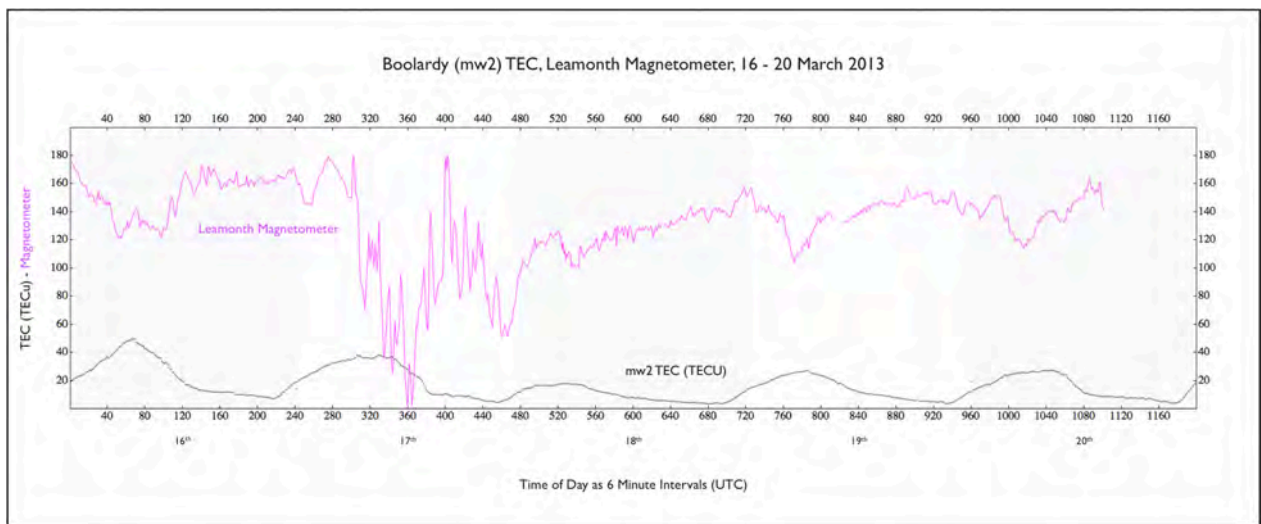


Figure 3: The effect on ionospheric TEC of a geomagnetic storm that occurred on March 17th, 2013. Geomagnetic activity was responsible for depletion of the ionospheric plasma on March 18th.

Through the use of distributed, high-fidelity GPS systems, the propagation of ionospheric disturbances are detectable. Two dual-frequency systems running purpose written software (SCINDA), (Carrano, Anghel et al. 2009), situated approximately 400 km apart on a north-south transect, recorded the effect on ionospheric TEC of the March 17th event (Fig. 4).

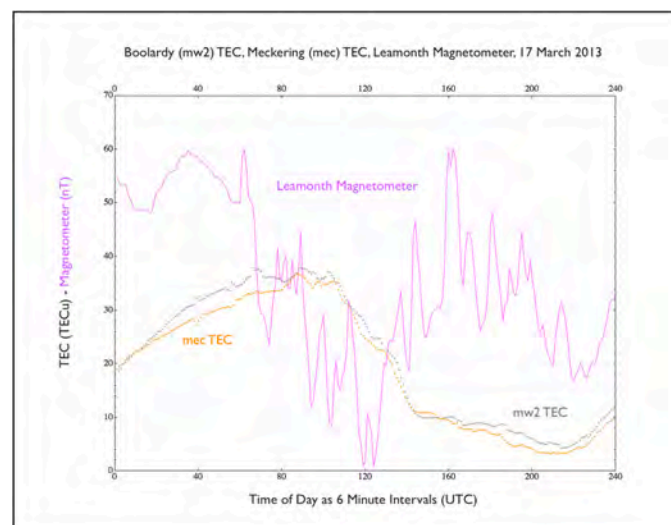


Figure 4: Absolute TEC values produced using purpose written software (SCINDA) on data recorded by high-fidelity GPS systems at two widely separated locations, superimposed with magnetometer readings captured at the Learmonth Solar Observatory, Learmonth, Western Australia, indicating a significant geomagnetic storm event. One system (mw2, lat. -26.984° , long. 116.535°) is located on the MRO, the other (mec, lat. -31.639° , long. 116.989°) on the Australian Space Academy campus at Meckering, Western Australia.

Figure 5 displays celestial object offsets derived from an actual observation conducted with the MWA precursor instrument, the MWA 32 Tile Array (32T). Vectors effectively display the magnitude of the positional offset of observed sources when compared to a known catalogue, the Molonglo Reference Catalogue.

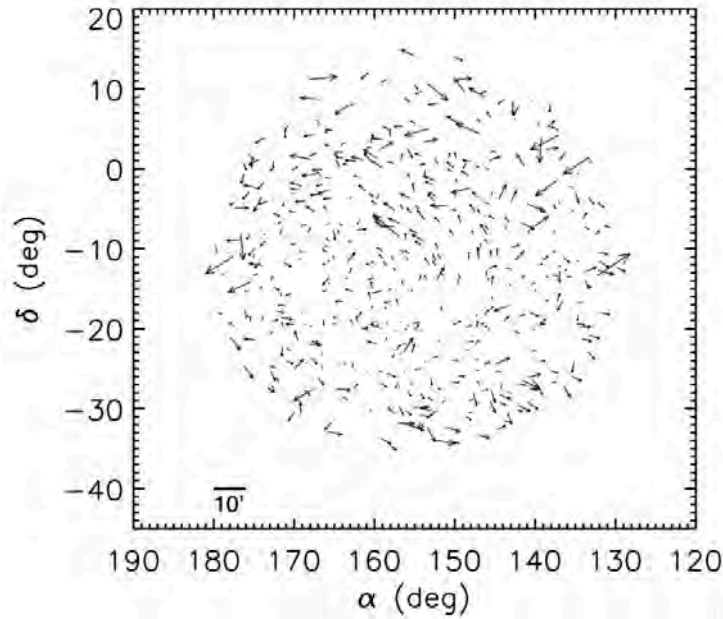
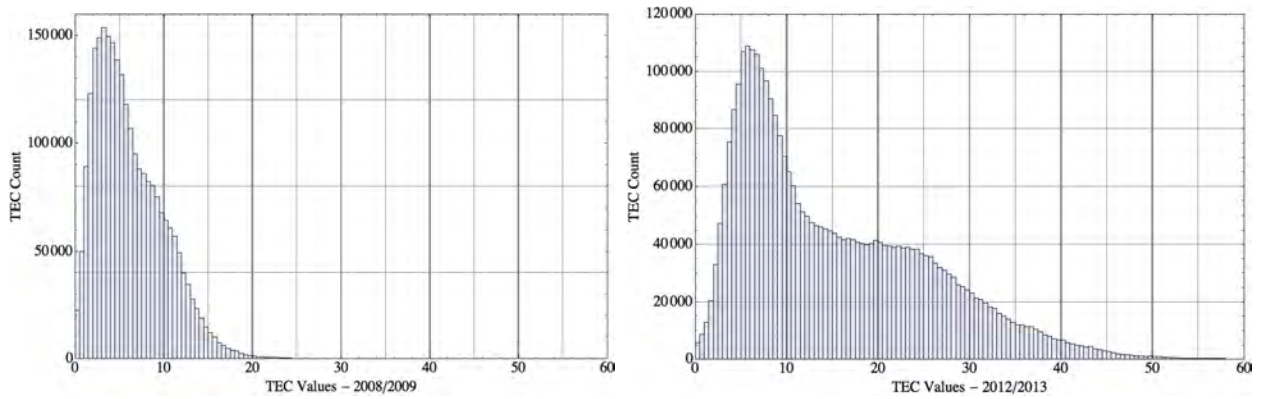


Figure 5: Spatial offsets between the position of sources imaged by the MWA (32T) and sources for which matches were found in the Molonglo Reference Catalogue (MRC). These vectors effectively demonstrate the variability in the scene attributable substantially to the effect of the ionosphere on incoming radio signals (Williams, Hewitt et al. 2012).

Ionospheric TEC varies markedly between periods of low and high solar activity. Histograms of TEC data captured over those epochs corresponding to those of Figure 2 are given in Figures 6(a) and 6(b).



Figures 6(a), 6(b): Histograms of TEC data captured over the years (a) 2008/2009 and (b) 2012/2013 respectively.

Discussion

The MWA is a low-frequency radio array that exhibits a wide field of view, high spatial and temporal resolution and simultaneous imaging capabilities, allowing high time-resolution imaging of transient events, daily observations of the sun with unprecedented capability (Oberoi, Matthews et al. 2011) and long integration time science, as required by epoch of re-

ionisation (EOR) studies. In order however, to fully achieve the observatory's potential, careful attention must be paid to the dynamics of the local ionosphere over both short (seconds) and long (days) intervals and steps taken to remove the degrading effects of the ionosphere's free electrons.

Figures 2(a), (b), (c) and (d) illustrate the response of the ionosphere under conditions of differing solar activity, proxies of which are represented by the F10.7 and A_p indices. Solar emissions capable of ionising molecular species that form Earth's ionosphere are represented by the 10.7 cm (2800 MHz) solar radio flux in units of $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ (F10.7). This emission is strongly associated with solar plage, those regions observed in the Sun's chromosphere associated with strong bunching of magnetic field lines. These are non-uniformly distributed across the solar sphere and therefore result in a periodic emission that is clearly observed in the F10.7 values exhibited during heightened solar activity (Fig. 3(c)).

In respect of scintillation (Figures 2(b) and 2(d)), the data are dominated by multi-path noise and although day-to-day variation is observed, intra-day modulation is not at the levels measured. Multipath manifests as diagonal lines in time-of-day/day-of-year waterfall plots, such as those in figures 2(b) and 2(d) in particular. The lines correspond to the rate at which GPS satellites precess in their fixed-path, daily orbits (approximately 4 minutes) and therefore result from fixed structures in the antenna environment. Inter-day variation is seen, although levels overall are very low. S_4 data derived from GPS carrier signals, ($L_1 = 1575 \text{ MHz}$), are not ideal measures of scintillation in respect of low-frequency radio astronomy. GPS signals of $\lambda \approx 0.2 \text{ m}$ correspond to a radiative, near-field (Fresnel) length of 264 m. The Fresnel length corresponding to MWA antennas is approximately 592 m at $\lambda = 1 \text{ m}$. No apparent scintillation has been observed within this considerable noise environment, an observation that supports the decision to locate the highly scintillation-sensitive MWA and in the near future, SKA Low, on the MRO. Further analysis of this data is ongoing.

Episodes of heightened geomagnetic activity, of which the A_p index presented here is a daily average, clearly overwhelms excitation of the ionosphere by solar EUV photons to induce a drop in ionospheric electron content (TEC) (Fig. 3). The greatest effect is observed over following days when peak TEC is reduced dramatically over what might have been expected. Ionospheric electron content is typically depressed for several days. Further, use of two or more GPS systems provides the opportunity to observe traveling ionospheric disturbances (TIDs). Greater insights into temporal and physical (length/magnitude) scales will be gained through continued observations and the deployment of another GPS system in closer proximity to the MRO, anticipated for coming months.

The wealth of data captured and represented in these plots demonstrates the complex electron dynamics of the ionosphere. The accuracy required of TEC measurements are challenging to meet, though important in this context. Therefore, measurements of ionospheric electron content (aTEC) and possible scintillation are continuing and should be expanded.

References

1. Lonsdale, C. J., Cappallo, R. J., Morales, M. F., Briggs, F. H., Benkevitch, L., Bowman, J. D., Bunton, J. D., Burns, S., Corey, B. E., Desouza, L., Doleman, S. S., Derome, M., Deshpande, A., Gopala, M. R., Greenhill, L. J., Herne, D. E., Hewitt, J. N., Kamini, P. A., Kasper, J. C., Kincaid, B. B., Kocz, J., Kowald, E., Kratzenberg, E., Kumar, D., Lynch, M. J., Madhavi, S., Matejek, M., Mitchell, D. A., Morgan, E., Oberoi, D., Ord, S., Pathikulangara, J., Prabu, T., Rogers, A., Roshi, A., Salah, J. E., Sault, R. J., Shankar, N. U., Srivani, K. S., Stevens, J., Tingay, S., Vaccarella, A., Waterson, M., Wayth, R. B., Webster, R. L., Whitney, A. R., Williams, A., Williams, C., (2009). "The Murchison Widefield Array: Design Overview." *IEEE Proceedings* 97: 1497-1506.
2. Bhat, N. D. Ramesh, Wayth, Randall B., Knight, Haydon S., Bowman, Judd D., Oberoi, Divya, Barnes, David G., Briggs, Frank H., Cappallo, Roger J., Herne, David E., Kocz, Jonathon, Lonsdale, Colin J., Lynch, Mervyn J., Stansby, Bruce, Stevens, Jamie, Torr, Glen, Webster, Rachel L., Wyithe, J. Stuart B. (2007). "Detection of Crab Giant Pulses Using the Mileura Widefield Array Low Frequency Demonstrator Field Prototype System." *The Astrophysical Journal* 665: 618-627.
3. Bowman, G. G. (1981). "The nature of ionospheric spread-F irregularities in mid-latitude regions." *Journal of Atmospheric and Terrestrial Physics* 43(1): 65-79.
4. Bowman, Judd D., Barnes, David G., Briggs, Frank H., Corey, Brian E., Lynch, Merv J., Bhat, N. D. Ramesh, Cappallo, Roger J., Doleman, Sheperd S., Fanous, Brian J., Herne, David E., Hewitt, Jacqueline N., Johnston, Chris, Kasper, Justin C., Kocz, Jonathon, Kratzenberg, Eric, Lonsdale, Colin J., Morales, Miguel F., Oberoi, Divya, Salah, Joseph E., Stansby, Bruce, Stevens, Jamie, Torr, Glen, Wayth, Randall, Webster, Rachel L., Wyithe, J. Stuart B. (2007). "Field Deployment of Prototype Antenna Tiles for the Mileura Widefield Array Low Frequency Demonstrator." *The Astronomical Journal* 133: 1505-1518.
5. Bowman, Judd D., Cairns, Iver, Kaplan, David L., Murphy, Tara, Oberoi, Divya, Staveley-Smith, Lister, Arcus, Wayne, Barnes, David G., Bernardi, Gianni, Briggs, Frank H., Brown, Shea, Bunton, John D., Burgasser, Adam J., Cappallo, Roger J., Chatterjee, Shami, Corey, Brian E., Coster, Anthea, Deshpande, Avinash, deSouza, Ludi, Emrich, David, Erickson, Philip, Goeke, Robert F., Gaensler, B. M., Greenhill, Lincoln J., Harvey-Smith, Lisa, Hazelton, Bryna J., Herne, David E., Hewitt, Jacqueline N., Johnston-Hollitt, Melanie, Kasper, Justin C., Kincaid, Barton B., Koenig, Ronald, Kratzenberg, Eric, Lonsdale, Colin J., Lynch, Mervyn J., Matthews, Lynn D., McWhirter, S. Russell, Mitchell, Daniel A., Morales, Miguel F., Morgan, Edward H., Ord, Stephen M., Pathikulangara, Joseph, Prabu, Thiagaraj, Remillard, Ronald A., Robishaw, Timothy, Rogers, Alan E. E., Roshi, Anish A., Salah, Joseph E., Sault, Robert J., Shankar, N. Udaya, Srivani, K. S., Stevens, Jamie B., Subrahmanyam, Ravi, Tingay, Steven J., Wayth, Randall B., Waterson, Mark, Webster, Rachel L., Whitney, Alan R., Williams, Andrew J., Williams, Christopher L., Wyithe, J. Stuart B. (2013). "Science with the Murchison

Widefield Array." PASA - Publications of the Astronomical Society of Australia 30.

6. Carrano, Charles S., Anghel, Adela, Quinn, Richard A., Groves, Keith M. (2009). "Kalman filter estimation of plasmaspheric total electron content using GPS." *Radio Sci.* 44: RS0A10.
7. Hajkowicz, L. A. (1994). "Types of ionospheric scintillations in southern mid-latitudes during the last sunspot maximum." *Journal of Atmospheric and Terrestrial Physics* 56(3): 391-399.
8. Kennewell, John, Caruana, John, Terkildsen, Michael, Wu, Jiping, Wilkinson, Phil, Cole, David. (2005). *The Australian Ionosphere - Report to CSIRO by IPS Radio and Space Services.*
9. Kung Chie, Y. and L. Chao-Han (1982). "Radio wave scintillations in the ionosphere." *Proceedings of the IEEE* 70(4): 324-360.
10. Mitchell, D. A., Greenhill, L. J., Wayth, R. B., Sault, R. J., Lonsdale, C. J., Cappallo, R. J., Morales, M. F., Ord, S. M. (2008). "Real-Time Calibration of the Murchison Widefield Array." *Selected Topics in Signal Processing, IEEE Journal of* 2(5): 707-717.
11. Oberoi, Divya, Matthews, Lynn D., Cairns, Iver H., Emrich, David, Lobzin, Vasili, Lonsdale, Colin J., Morgan, Edward H., Prabu, T., Vedantham, Harish, Wayth, Randall B., Williams, Andrew, Williams, Christopher, White, Stephen M., Allen, G., Arcus, Wayne, Barnes, David, Benkevitch, Leonid, Bernardi, Gianni, Bowman, Judd D., Briggs, Frank H., Bunton, John D., Burns, Steve, Cappallo, Roger C., Clark, M. A., Corey, Brian E., Dawson, M., DeBoer, David, De Gans, A., deSouza, Ludi, Derome, Mark, Edgar, R. G., Elton, T., Goeke, Robert, Gopalakrishna, M. R., Greenhill, Lincoln J., Hazelton, Bryna, Herne, David E., Hewitt, Jacqueline N., Kamini, P. A., Kaplan, David L., Kasper, Justin C., Kennedy, Rachel, Kincaid, Barton B., Kocz, Jonathan, Koeing, R., Kowald, Errol, Lynch, Mervyn J., Madhavi, S., McWhirter, Stephen R., Mitchell, Daniel A., Morales, Miguel F., Ng, A., Ord, Stephen M., Pathikulangara, Joseph, Rogers, Alan E. E., Roshi, Anish, Salah, Joseph E., Sault, Robert J., Schinckel, Antony, Udaya Shankar, N., Srivani, K. S., Stevens, Jamie, Subrahmanyam, Ravi, Thakkar, D., Tingay, Steven J., Tuthill, J., Vaccarella, Annino, Waterson, Mark, Webster, Rachel L., Whitney, Alan R. (2011). "First Spectroscopic Imaging Observations of the Sun at Low Radio Frequencies with the Murchison Widefield Array Prototype." *The Astrophysical Journal Letters* 728: L27.
12. Oberoi, D. and C. J. Lonsdale (2012). "Media responsible for Faraday rotation: A review. *Radio Science* 47.
13. Tingay, S. J.; Goeke, R.; Bowman, J. D.; Emrich, D.; Ord, S. M.; Mitchell, D. A.; Morales, M. F.; Boller, T.; Crosse, B.; Wayth, R. B.; Lonsdale, C. J.; Tremblay, S.; Pallot, D.; Colegate, T.; Wicenc, A.; Kudryavtseva, N.; Arcus, W.; Barnes, D.; Bernardi, G.; Briggs, F.; Burns, S.; Bunton, J. D.; Cappallo, R. J.; Corey, B. E.; Deshpande, A.; Desouza, L.; Gaensler, B. M.; Greenhill, L. J.; Hall, P. J.; Hazelton, B. J.; Herne, D.; Hewitt, J. N.; Johnston-Hollitt, M.; Kaplan, D. L.; Kasper, J. C.; Kincaid, B. B.; Koenig, R.; Kratzenberg, E.; Lynch, M. J.; McKinley, B.; McWhirter, S. R.; Morgan, E.; Oberoi, D.; Pathikulangara, J.; Prabu, T.; Remillard, R. A.; Rogers, A. E. E.; Roshi, A.; Salah, J. E.; Sault, R. J.; Udaya-Shankar, N.; Schlagenhauser, F.; Srivani, K. S.; Stevens, J.

- Subrahmanyam, R.; Waterson, M.; Webster, R. L.; Whitney, A. R.; Williams, A.; Williams, C. L.; Wyithe, J. S. B. (2012) "The Murchison Widefield Array: the Square Kilometre Array Precursor at low radio frequencies." Publications of the Astronomical Society of Australia, Volume 30, id.e007 21 pp.
14. Williams, CL and Hewitt, JN and Levine, AM and de Oliveira-Costa, A and Bowman, JD and Briggs, FH and Gaensler, BM and Hernquist, LL and Mitchell, DA and Morales, MF and Sethi, SK and Subrahmanyam, R and Sadler, EM and Arcus, W and Barnes, DG and Bernardi, G and Bunton, JD and Cappallo, RC and Crosse, BW and Corey, BE and Deshpande, A and deSouza, L and Emrich, D and Goeke, RF and Greenhill, LJ and Hazelton, BJ and Herne, D and Kaplan, DL and Kasper, JC and Kincaid, BB and Koenig, R and Kratzenberg, E and Lonsdale, CJ and Lynch, MJ and McWhirter, SR and Morgan, EH and Oberoi, D and Ord, SM and Pathikulangara, J and Prabu, T and Remillard, RA and Rogers, AEE and Roshi, DA and Salah, JE and Sault, RJ and Shankar, NU and Srivani, KS and Stevens, JB and Tingay, SJ and Wayth, RB and Waterson, M and Webster, RL and Whitney, AR and Williams, AJ and Wyithe, JSB (2012) "Low Frequency Imaging of Fields at High Galactic Latitude with the Murchison Widefield Array 32-Element Prototype." The Astrophysical Journal, 755, (1) Article 47. ISSN 0004-637X (2012).

Acknowledgements

We acknowledge and thank Dr. Ron Caton and the U.S.A.F. Research Laboratories for provision of the high-fidelity GPS systems employed in the measurement of ionospheric TEC and scintillation.

We take this opportunity to thank engineering staff of CIRA who have assisted in the deployment and maintenance of a GPS system employed in this study, in particular, Mr. David Emrich, Mr. Brian Crosse and Mr. David Pallot.

We acknowledge the contribution made to this project by Horizon Power in provision of electrical components, including solar panels, required for the construction of 3 solar power trailers.



An Investigation of Discrete Cosmic Radio Sources Using an Imaging Riometer at Davis, Antarctica

Lewis E. Freeland¹, Frederick W. Menk¹, Murray D. Sciffer¹ and Michael B. Terkildsen²

¹ School of Mathematical and Physical Sciences, University of Newcastle,
Callaghan, NSW 2308, Australia

² IPS Radio and Space Services, Bureau of Meteorology,
PO Box 1386, Haymarket, NSW 1240, Australia

Summary: This paper describes a search for discrete cosmic radio sources in data from the Southern Hemisphere Imaging Riometer Experiment (SHIRE). Procedures were developed to identify the signatures of discrete cosmic radio sources in SHIRE data, and 6 candidate sources were selected from extant astronomical survey data. The passage of each source across individual beams of the riometer was examined in selected days and statistical samples from over 4 years of SHIRE data. It was found that none of the selected sources had any significant effect on the SHIRE observations, in contrast to a previous report of detection of discrete radio sources by the imaging riometer at Poker Flat, Alaska [1]. This is likely due to two factors: the relatively large beamwidth of the SHIRE instrument compared to the Poker Flat riometer, and the relatively low flux of the strongest radio sources visible to Antarctic riometers compared to their northern counterparts.

Keywords: riometer, ionospheric absorption, cosmic noise

Introduction

Riometers are instruments which measure the level of absorption of cosmic radio noise in the ionosphere. The measured amplitude of cosmic noise varies temporally due to (1) the diurnal transit of the galaxy over the recording site, and (2) changes in absorption which are mostly due to variations in electron density low in the ionosphere. This parameter varies daily and seasonally with solar zenith angle, and is also influenced by space weather events, such as energetic electron precipitation accompanying magnetic storms. Riometers have been in widespread use as a standard geophysical observatory tool for over five decades. Imaging riometers are a more recent development, and are able to synthesise multiple narrow beams. This allows the spatial dependence of the cosmic radio absorption to be investigated. It might therefore be possible for imaging riometers to detect discrete cosmic radio sources, such as radio galaxies, as separate entities from the galactic background.

There are two main motivations for this work. First, if it were possible to routinely observe discrete radio sources using imaging riometers, the resulting data might provide a method for spatially calibrating their beam patterns. This is important because there is no direct method to achieve such calibration at present. Second, there is little information about southern astronomical radio sources at the frequencies around 40 MHz used by imaging riometers, and it might be possible to estimate the relative strengths of some radio sources in this band.

This paper aims to address these questions by testing whether discrete cosmic radio sources are visible in output from SHIRE at Davis station, Antarctica (68.58 S, 79.97 E geographic).

Instrumentation

SHIRE consists of a square grid of 64 crossed dipole antennas connected in a phased array that allows the simultaneous observation of 49 receiver beams. The spatial arrangement of these beams is illustrated in Figure 1, where the 7x7 beam pattern has been projected onto a horizontal flat plane at an altitude of 90 km. Each beam has a full width at half maximum (FWHM) of 12.8 degrees and adjacent beams in each row or column are separated by 15 degrees. The beams are labelled using an ordered pair (a,b), where a and b range from -3 to 3, so that (0,0) indicates the zenith beam, (3,0) points toward magnetic south and (0,3) to magnetic west. SHIRE digitally records the level of cosmic radio noise in each beam direction at its operating frequency of 38.2 MHz once per second. The design of SHIRE is similar to that of the IRIS (Imaging Riometer for Ionospheric Studies) instruments that operate elsewhere in the polar regions [2]. Further details on this common design and the operational principles and parameters appear in [2].

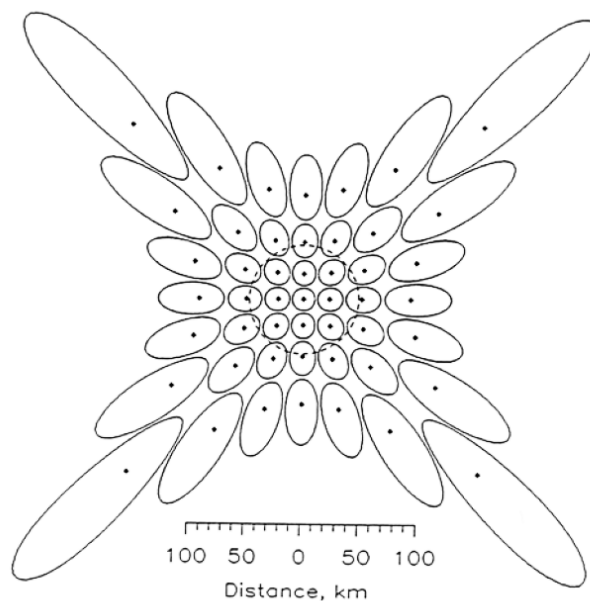


Fig. 1: The beam pattern of SHIRE, projected onto a plane at an altitude of 90km. The dashed circle at the centre shows the beam pattern of a conventional broad beam riometer for comparison. The dots indicate the centre of each beam, and the solid ellipses represent each beam's half-power contour [2].

Experimental Method

We examined SHIRE data at times when 6 discrete radio sources (listed in Table 1) were expected to pass through SHIRE's field of view. The sources were selected from the Parkes PKSCAT90 catalogue of point radio sources [3]. This catalogue contains data for flux densities at 80 MHz and not at or near 38.2 MHz, but no more appropriate catalogue exists for southern-hemisphere radio sources at this frequency. These 80 MHz measurements come from Culgoora radioheliograph surveys [4-6]. The 5 strongest 80 MHz sources that are visible from SHIRE's southern latitude were chosen. As shown in Table 1, these sources had

a flux of 100 Jy or higher at 80 MHz and a declination less than -10° . An additional sixth source, Centaurus A, was also included since it was anecdotally believed this was visible in SHIRE data.

Table 1: Sources examined in the present study, data are from [4-6].

Source Name	Object type	Declination	80 MHz Flux (Jy)
Hydra A	Galaxy	-12.1	441
Pictor A	Galaxy	-45.8	410
Milne 23	SNR	-53.2	182
J1935-4620	Galaxy	-46.3	147
Kepler's SN	SNR	-21.5	100
Centaurus A	Galaxy	-43.1	Not in catalogue

To determine the time intervals when a particular radio source was potentially observable by SHIRE a coordinate transformation algorithm was used to convert equatorial coordinates (right ascension, RA, and declination) to horizontal coordinates (azimuth and elevation) at the riometer's location. Then, using the known azimuth and elevation values for the pointing direction of each riometer beam, the distance (in degrees) of the source from each beam direction was calculated. Repeating the process every second throughout a given day and examining when this distance for a certain beam was less than the half-power radius gave the time interval when this source should be within this beam's field of view. The riometer data were then aligned in sidereal time for each day and beam of interest, and averaged using a superposed epoch process to improve the signal-to-noise ratio.

Meaningful analysis of riometer data requires removal of the diurnal variation in cosmic noise arising from the rotation of Earth and hence the transit of the galaxy and other astronomical features over the recording site. This diurnal variation generally dominates other effects in riometer data and is removed by subtracting a quiet day curve (QDC) consisting of an average over 30 days of raw data. The sidereal day is about 4 min shorter than a solar day, and this must be accounted for when performing averaging and superposed epoch analysis of riometer data.

The SHIRE instrument operates continuously, but there are occasional data gaps and other instrumental artefacts which would affect statistical studies. Ideally the data should be visually inspected to remove such potential difficulties prior to detailed analysis. This was done for 11 years of SHIRE data, after which QDC removal was undertaken and then specific intervals were examined for evidence of discrete cosmic noise sources.

Results

We illustrate the results by focusing on a specific example, Centaurus A. The same process was followed for all other sources listed in Table 1. Figure 2 illustrates how the motion of Centaurus A, depicted by the solid curve, was tracked in horizontal coordinates across the

riometer beams, represented by circles. The centres of the circles correspond to the beam azimuth and elevation, and the circle diameter is the FWHM angle. Beams of interest for a corresponding radio source are selected on the basis of the length of the segment the source traces across a beam.

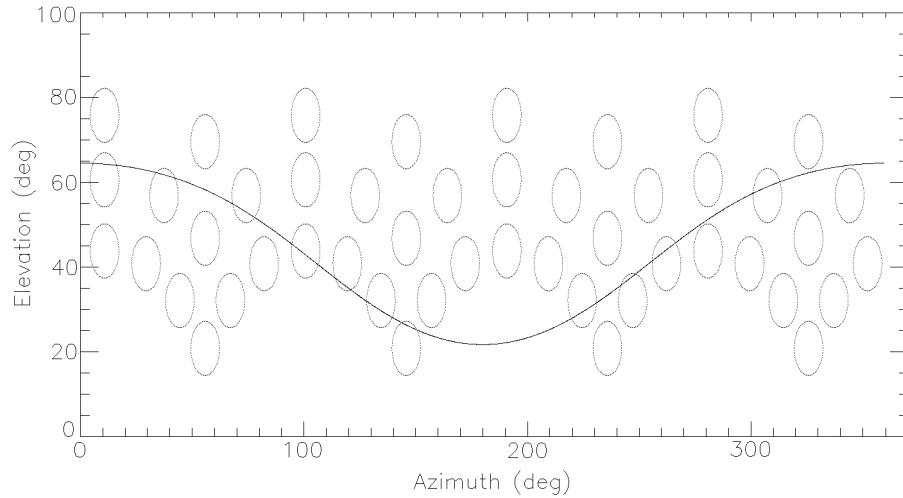


Fig. 2: The path of radio source Centaurus A over SHIRE, in horizontal coordinates.

We initially elected to examine SHIRE data from early 2007, as solar activity was low at this time, reducing instances when solar radio emissions were present in the data. Intervals of both raw and QDC-subtracted data were examined as it was unclear to what extent radio noise from the discrete sources would be subtracted along with the QDC. Here we show data from beam (3,0) for one representative and otherwise unremarkable day, chosen because Centaurus A moved almost directly across the beam centre. On 15 January 2007 this passage occurred between 1839 and 1932 UT. The raw riometer data for this time, and the hour before and after the crossing, are shown in Figure 3. The ordinate axis represents output voltage from the riometer receiver.

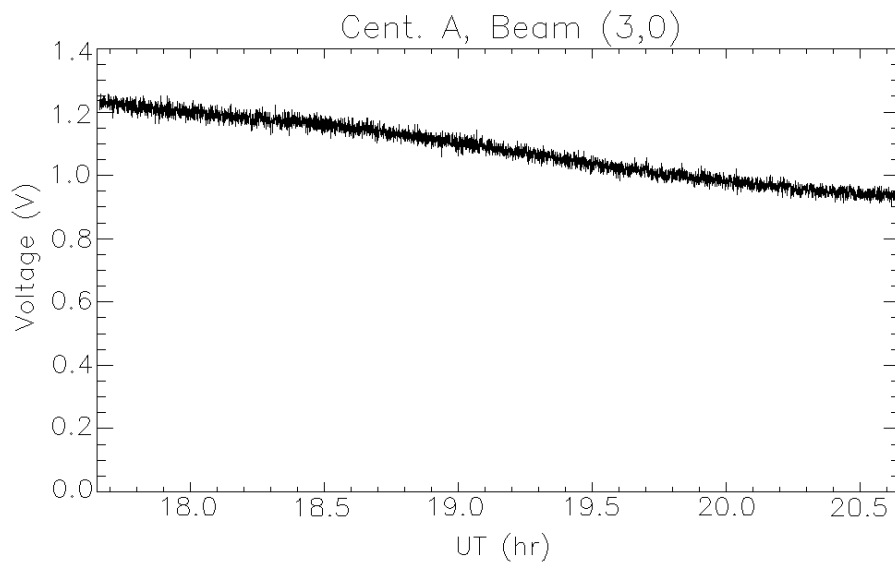


Fig. 3: Time series plot of raw SHIRE beam (3,0) data corresponding to the predicted crossing time of Centaurus A, 1839-1932 UT, on 15 January 2007.

Figure 4 shows the same beam and time period after QDC removal, accounting for the diurnal variation. Here the ordinate axis represents ionospheric noise absorption, relative to the background, in dB, and an enlarged scale has been used to better illustrate the short term variations.

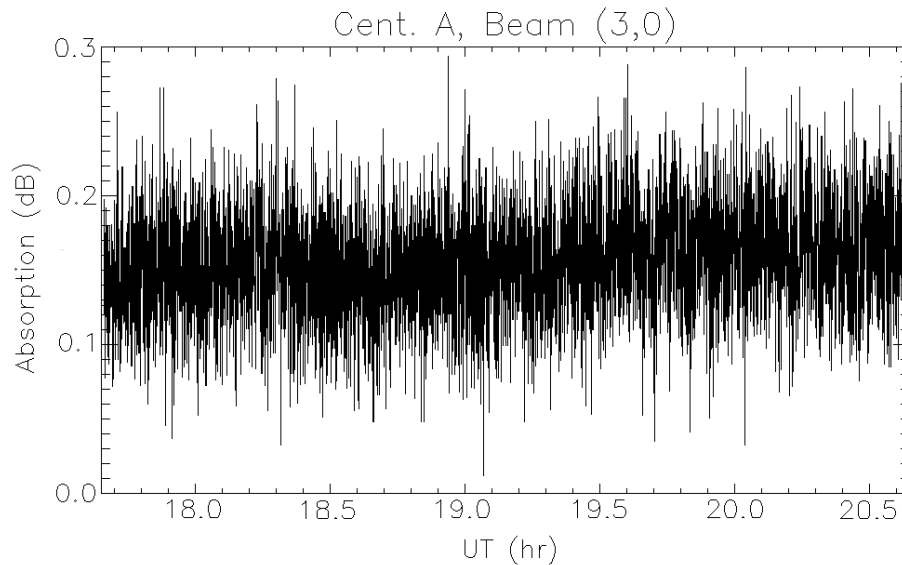


Fig. 4: Time series plot of QDC-subtracted SHIRE beam (3,0) data corresponding to the predicted crossing time of Centaurus A, 1839-1932 UT, on 15 January 2007.

Considerable noise is present in individual QDC-subtracted time series such as shown in Figure 4, masking any contribution from a discrete source. Therefore superposed epoch averaging was performed over the first 100 days of 2007, creating the averaged raw data plot shown in Figure 5. The vertical lines represent the times when the radio source entered and exited the FWHM contour, and the x- axis represents time relative to this source transit.

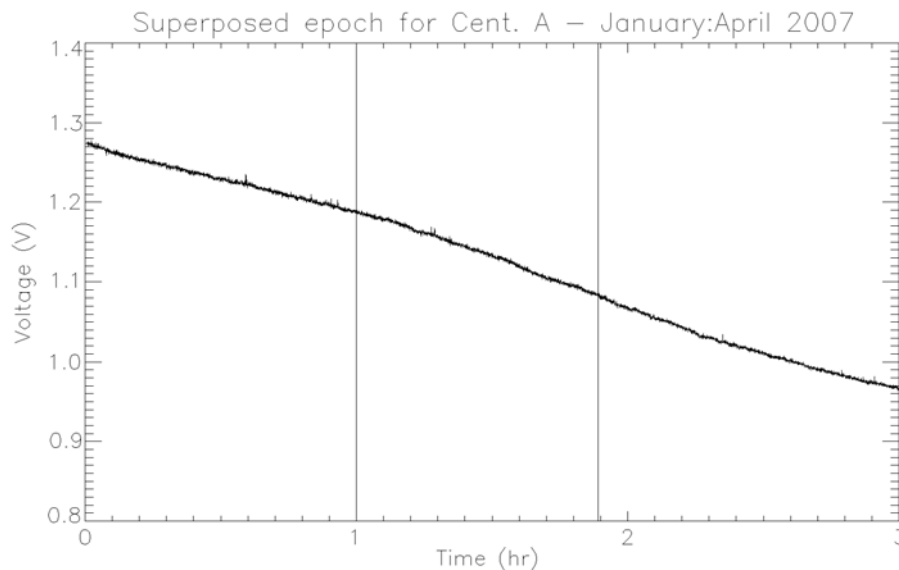


Fig. 5: A 100 day superposed epoch plot of raw data for Centaurus A crossing the (3,0) SHIRE beam.

The corresponding QDC-subtracted plot is presented in Figure 6. Here the data have been smoothed with a 60-second moving average. The error bars at the right of the plot represent 3 standard deviations calculated from the unsmoothed data noise level, which was the chosen minimum amplitude for a statistically significant detection in either raw or QDC-subtracted data.

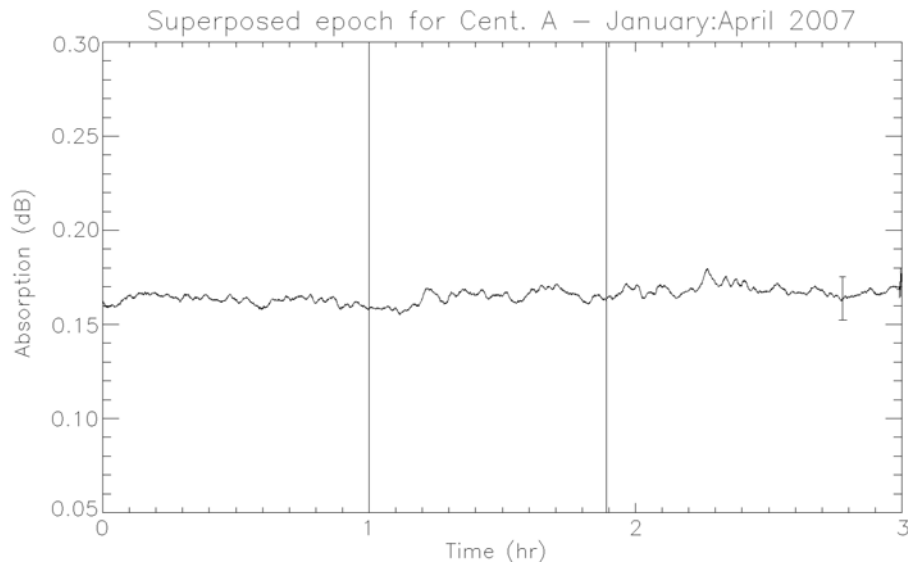


Fig. 6: A 100 day superposed epoch plot of QDC-subtracted data for Centaurus A crossing the (3,0) SHIRE beam.

Inspection of Figures 5 and 6 shows that there is no statistically significant deviation from the radio background that would indicate the presence of the radio source Centaurus A in SHIRE beam (3,0) over this 100-day interval. Thus we conclude that Centaurus A is undetectable in the output from this beam.

Similar analysis was performed for all of the radio sources listed in Table 1, and for every SHIRE beam that had a substantial intersection with each source. In all of this analysis, the 100 day epoch discussed above was used as a minimum. Data were examined from a four-year period, with different periods between 2006 and 2008 used for Centaurus, Hydra and Pictor. In all cases, the result was the same: the chosen discrete cosmic radio sources were not statistically detectable in the output from SHIRE using this method.

Discussion

The non-detection of the discrete radio sources that we expect to be brightest at SHIRE's frequency and location initially surprised us and appears to contradict the findings of Bezrodny et al. [1]. They described the detection of four strong northern hemisphere radio sources using the imaging riometer at Poker Flat, Alaska. We suggest two reasons for their different result compared to our study here.

First, the Poker Flat imaging riometer is a newer design that synthesises narrower beams than SHIRE. The FWHM beam widths for the two riometers are 6 degrees and 12.8 degrees respectively. A discrete cosmic radio source in a narrower beam will contribute

proportionally more to the received signal than it would in a wider beam, so the Poker Flat riometer has an advantage for detecting such a source.

Second, the northern radio sources detected by Bezrodny et al. have a higher flux density than the southern sources observed by SHIRE, at the low operating frequency (by radioastronomical standards) of the riometers. Table 2 is based on data from [7] and compares Centaurus A to the two strongest northern sources detected at Poker Flat. Centaurus A is of comparable intensity to the other sources at 1000 MHz but nearly an order of magnitude weaker at 100 MHz, suggesting that the sources examined by us are simply not strong enough to be detected.

Table 2: A comparison between the fluxes of the sources observed in [1], and Centaurus A. data are from [7].

Source	S100 (Jy)	S1000 (Jy)
Cas A	19 500	3 300
Cygnus A	13 800	2 340
Cent A	3 000	2 000

In order to differentiate between the effects of beamwidth and source intensity a future study could perform a similar analysis of data from a northern-hemisphere imaging riometer of similar design to SHIRE. This would remove factors related to the intensity of different radio sources visible from either hemisphere and reduce the question to one of beamwidth alone. If routine observations of the stronger northern sources are possible, then it should be possible to use these sources for calibration of the spatial beam pattern of imaging riometers.

Finally, it should be noted that many non-astronomical discrete radio sources were detected using SHIRE during this study. The most significant were intense radio bursts from the Sun accompanying large solar flares and the onset of geomagnetic storms. While such bursts were most intense in the SHIRE beams pointing most nearly to the Sun, they were visible at smaller amplitudes in all beams. Artificially produced radio signals were also frequently detected. These were most intense in SHIRE's northernmost beams, suggesting they are due to radio transmissions from ships.

Conclusion

This project examined riometer data corresponding to 6 discrete astronomical radio sources over a 4 year period and found that there are no detectable signatures of such cosmic radio sources in the data from SHIRE. This was shown by the systematic inspection of several years of data at the predicted crossing times of radio sources.

The reasons such detections using SHIRE are not possible despite reported detection of discrete cosmic radio sources with the Poker Flat imaging riometer is likely a combination of the larger beam width of the SHIRE instrument and the relatively low flux of southern radio sources compared to their northern counterparts.

This conclusion is surprising because there are references to discrete radio sources in prior literature for northern-hemisphere imaging riometers.

Acknowledgements

We thank personnel at Davis, Antarctica, and IPS Radio and Space Services, for operating the SHIRE instrument and making data available for this study.

References

1. Bezrodny, V., Charkina, O., Galushko, V., Groves, K., Kashcheyev, A., Watkins, B., Yampolski, Y. and Murayama, Y., “Application of an imaging HF riometer for the observation of scintillations of discrete cosmic sources”, *Radio Science*, Vol. 46, No. 6, 2008, RS6007.
2. Detrick, D. and Rosenberg, T., “A phased-array radiowave imager for studies of cosmic noise absorption”, *Radio Science*, Vol. 25, No. 4, 1990, pp. 325-338.
3. Otrupcek, R.E. and Wright, A.E., “PKSCAT90 - the southern radio source database”, *Proceedings of the Astronomical Society of Australia*, Vol. 9, No. 1., 1991, p. 170.
4. Slee, O., “Culgoora-3 list of radio source measurements at 80 MHz”, *Australian Journal of Physics Astrophysics Supplement*, Vol. 43, 1977, pp. 1-123.
5. Slee, O., and Higgins, “Culgoora-1 list of radio sources and measurements at 80 MHz”, Technical report, CSIRO, Epping, 1973.
6. Slee, O., and Higgins, C., “Culgoora-2 list of radio source measurements at 80 MHz”, *Australian Journal of Physics Astrophysics Supplement*, Vol. 36, 1975, 1.
7. Zombeck, M., *Handbook of Space Astronomy and Astrophysics*, Cambridge University Press, Cambridge, 2007.

TRACE: A New Relocatable Airglow Imager

Anne Unewisse ¹, Andrew Cool ¹, Andrew Kirby ² and Dan Meehan ¹

¹ *Defence Science and Technology Organisation, PO Box 1500 Edinburgh SA, 5111*

² *US Air Force, United States of America*

Summary: The ionospheric E and F layers of the Earth's atmosphere are regions of enhanced electron density important for High Frequency (HF) propagation such as HF communications and over-the-horizon radar (OTHR). Irregularities in the ionosphere are common and are observed using a variety of instruments including ionospheric sounders, satellites, GPS receivers and airglow imagers. Most Earth-based ionospheric measuring instruments are relatively easy to transport and may be used to sample the ionosphere at various locations of interest. Airglow imagers, however, are generally used to gather large volumes of data at one location. This paper describes a new joint Defence Science and Technology Organisation (DSTO) / United States Air Force relocatable airglow imager known as the Thermospheric Radar Airglow Correlation Experiment (TRACE) which is designed to image the ionosphere at wavelengths of 557.4 nm, 589.3 nm, 630.0 nm and 777.4 nm at locations around Australia. Some routinely imaged small scale ionospheric disturbances and transient ionospheric depletion events are also presented.

Keywords: Airglow imaging, ionospheric studies, HF propagation

Introduction

The Defence Science and Technology Organisation (DSTO) operates the Jindalee Over-the-horizon Radar Network (JORN); a network of three High Frequency (HF) over-the-horizon radars located around Australia. Signals from these radars propagate via regions of high electron density in the Earth's atmosphere known as the ionospheric E and F layers which have peak densities at altitudes of around 110 km and 250 km respectively. Electron density irregularities are prevalent in these layers and have implications for radar co-ordinate registration, frequency selection and signal strength.

Various atomic and molecular species also have density peaks at E and F region altitudes and some of these produce characteristic airglow emission through chemiluminescence when excited by solar or cosmic radiation. Airglow brightness is a function of species density and excitation rate so that, for a constant excitation rate, it may be used to map species density. Ionospheric disturbance phenomena, such as Atmospheric Gravity Waves (AGWs), are not species specific and will affect the density distribution of ions, molecules and electrons alike. Thus measuring the density patterns in airglow originating from E and F layer heights gives an indication of the electron density profile through which the HF radar transmission is propagating.

F region airglow is dominated by two different narrow band atomic Oxygen emissions: 630 nm (red) from 200-300 km and 777.4 nm (Near Infrared: NIR) from 300-400 km. There is also a minor contribution from 557.7 nm (green) atomic Oxygen at around 230 km. E region

airglow is dominated by 557.7 nm (green) emission due to atomic Oxygen at ~96 km [1] and ~589.3 nm (yellow) emission from a Sodium doublet line at ~92 km [2]. Other commonly imaged E-layer airglow emission includes Meinel Hydroxyl (OH) bands (720 - 910 nm) [3] and molecular Oxygen bands (~866 nm) [4]. Figure 1 compares the altitudes of dominant airglow emission regions to the night-time electron density profile showing the approximate position of the E and F layers.

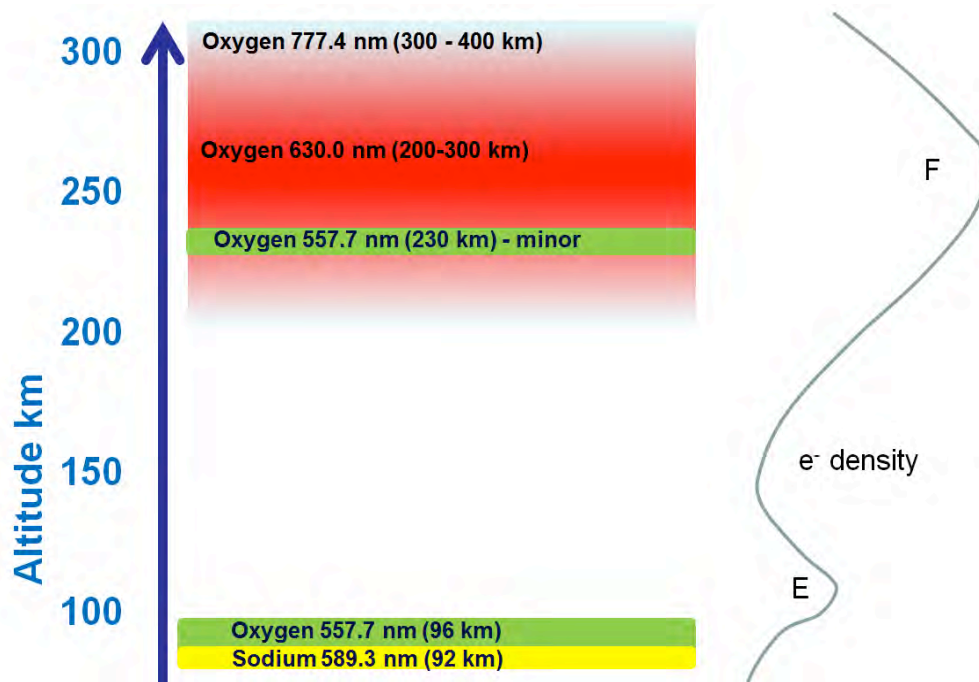


Figure 1: Diagram showing the night-time relationship between some of the brightest airglow emission regions and the ionospheric E and F layers.

The effects of AGWs and low density “bubbles” rising through the layers of the atmosphere are commonly observed in airglow images [5, 6]. The intensity of 630 nm emission has been shown to be directly proportional to F-layer electron density [1, 5]. Depletions in 630 nm emissions are also associated with the onset of Spread F [6]. The intensity of 777.4 nm emission is dependent on F-layer peak electron density although this emission is intrinsically weak and suffers from contamination from OH emission at 778.2 nm [1, 7]. There is a strong correlation between sodium 589.3 nm airglow emission and sporadic E ionization layers [8] and the detection of spread F in radar measurements [6].

This paper describes the transportable Thermospheric Radar Airglow Correlation Experiment (TRACE), a joint Defence Science and Technology Organisation (DSTO)-United States Air Force (USAF) collaboration designed to investigate the structure of the Earth’s atmosphere at E and F ionospheric heights using airglow imaging at various locations throughout Australia. It also presents examples of TRACE images collected at Alice Springs in late November 2011 and Adelaide from February 2012 onwards.

The Airglow Imager

The TRACE airglow imaging equipment is shown in Figure 2 with the camera lowered. The equipment is housed in a transportable 20 ft shipping container and consists of a KeoSentry [9] camera with 180 degree field of view (FOV), 24mm / F4.0, achromatic Mamiya fisheye lens, a temperature controlled filter wheel and a 16 Bit Princeton Instruments Acton backlit Pixis 1024B CCD camera with a pixel size of $13.3\mu\text{m} \times 13.3\mu\text{m}$ all mounted on a Pier-Tech telescopic pier. The pier is automated so that the camera rises to fit inside an Aquatica 9.25 inch BK7 glass dome at astronomical dusk and lowers at astronomical dawn. Lowering the camera in this way reduces UV damage to the lens during daylight hours.

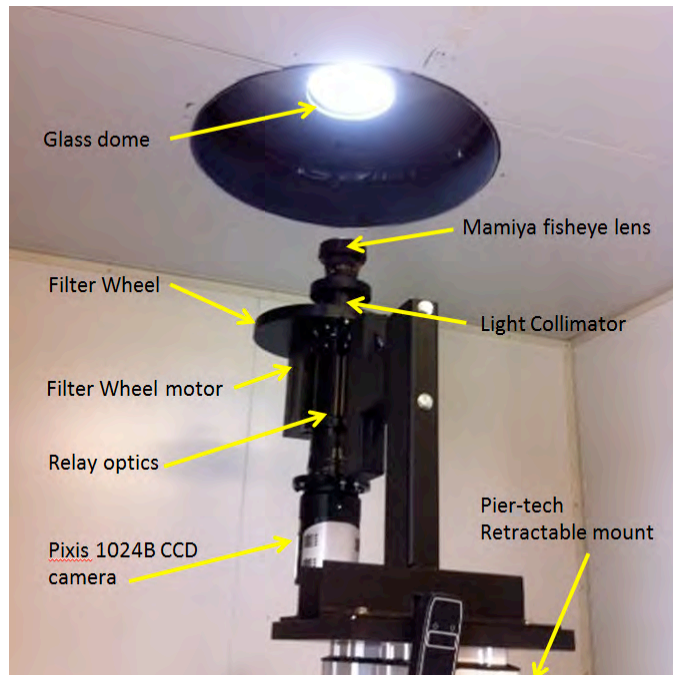


Figure 2: The airglow imager and filter wheel setup in the TRACE container

The filter wheel is kept at a constant temperature of 25 C and has six 3 inch slots, one of which is blank to allow baseline light measurements. The Keo scientific filters have a bandwidth of ~ 2 nm and are centred on: 557.7 nm (OI), 589.3 (Na), 630.0 nm (OI), 777.4 nm (OI) and 572.5 nm (background). The 777.4 NIR measurements tend to suffer from a fringing interference pattern known as echeloning which may be mitigated during image processing [10]. The CCD quantum efficiency is around 95% at 557.7 nm, 598.3 nm, 630.0 nm and 572.5 nm but falls to approximately 75% at 777.4 nm.

The Pixis camera's CCD dark charge measured at -70°C is 0.0003 e-/pixel/sec although it regularly operates at -80°C . We use the maximum readout speed of 2 MHz which equates to a gain of 1.11 e-/ADU and a readout noise of 12.85 e- rms [11].

TRACE's FOV corresponds to image diameters of ~ 400 km at an altitude of 96 km [3, 7] and ~ 1000 km at an altitude of 250 km [12]. The resolution of the inner region of the image is approximately 0.4 km/pixel at an altitude of 96 km and 1 km/pixel at an altitude of 250 km [13].

Comparison to other Australian airglow imagers

There are currently three other all-sky airglow imagers in Australia: one located in Darwin, NT which forms part of the Japanese Optical Mesosphere Thermosphere Imagers (OMTI) network [7, 14] and two older imagers operated by the University of Adelaide (ADLU) located in Alice Springs, NT and Buckland Park, North of Adelaide, SA [4]. The bit depth, CCD chip size, field of view (FOV) and available filters of these other imagers are compared to the TRACE imager in Table 1. The OMTI all-sky imager is supported by a 1024 x 1024 CCD Fabry-Perot interferometer [15, 16]. Note that although the Adelaide University's original image size is 512 x 512, they implement on-chip binning of 4 x 4 pixels to form 128 x 128 super pixels. The dark current is 0.05 counts/pixel/sec for the full 512 x 512 chip and 0.8 counts/pixel/sec for the 128 x 128 super pixels [4].

Table 1 Comparison of Australian airglow imagers

Imager	Bit depth	Chip size	FOV (deg)	E Filters (nm)				F Filters (nm)	Backgr (nm)
				OI	O ₂	Na	OH	OI	
TRACE	16	1024	180	557.7		589.3		630.0,777.4,(557.7)	572.5
OMTI	14	512	180	557.7			720-910	630.0,777.4,(557.7)	572.5
ADLU	14	512	60x40		866.0,868.0		840.0,843.0		857.0

TRACE represents state of the art airglow equipment offering greater colour depth, chip size, field of view and resolution than older imagers. Although comparable to the OMTI imager, TRACE's portability increases operational flexibility.

Environmental sensors

The TRACE is supported by two roof mounted environmental sensors: the Unihedron Sky Quality Meter (SQM) [17] which records calibrated light readings every minute and the Boltwood Cloud Sensor Mk II (BCS) [18, 19] which monitors environmental conditions such as temperature, cloud coverage, rain, humidity, wind speed and light every 2.06 seconds. The outputs from these sensors are fed back into the TRACE container via a patch panel to a sensor computer.

As well as monitoring current weather conditions, the BCS output triggers the automated pier to lower the camera during daylight hours to reduce previously experienced UV-induced lens coating degradation. Similar degradation has been reported on OMTI [7] and THEMIS Ground Based Observatory [20] imagers. The BCS output data is collected in a database which may be interrogated later to determine the observing conditions and filter wavelengths used on a particular night.

Image Processing

The airglow image obtained using TRACE (Image_M) has been modified by the CCD chip performance (as measured by the Dark) and non-uniformities in the optical chain from the lens to the CCD (as measured by the Flat). The true image (Image₀) may be recovered from the TRACE image using Eqn 1:

$$\text{Image}_0 = (\text{Image}_M - \text{Dark}) / \text{Flat} \quad (1)$$

By convention, the image is also rotated so that North is at the top and East is to the left. Darks and Flats are usually calculated from a series of images to reduce error. These are known as the master dark and master flat respectively.

A master dark is constructed by finding the pixel by pixel median of a large number of closed shutter exposures (darks) taken over many nights with the same exposure time and temperature. Figure 3 shows the number of counts across a 1024 x 1024 pixel master dark image. Each line represent the counts across a row (x pixels) or column (y pixels). The master dark image contains high and low rows and columns as well as a general slope in counts across the chip (the bias). These features are typical of most darks [21, 22].

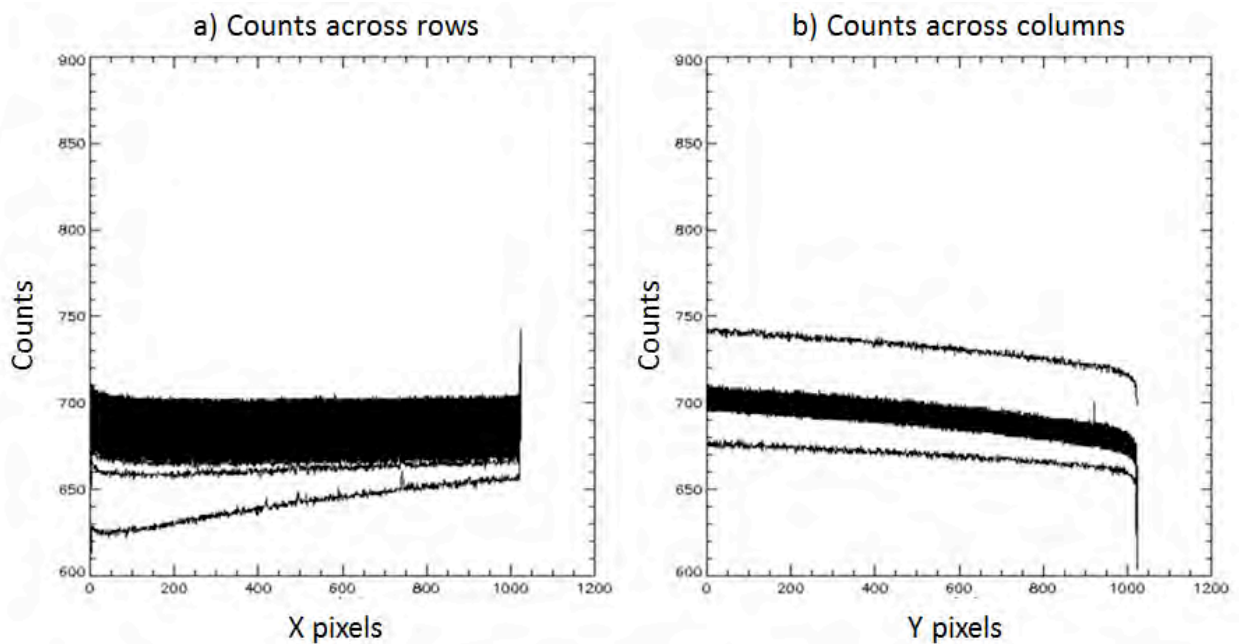


Figure 3: Number of counts across a 1024 x 1024 pixel master dark image. Each line represents a row (x pixels) or column (y pixels) on the image.

There are many methods used to create the master flat. In some cases, a sky flat is created using short exposure images taken at twilight (either dusk or dawn) [23]. Another common method involves combining a series of uniformly illuminated images (sometimes artificially illuminated with lamps etc.) but this method is suboptimal for wide-field lenses, such as that used in TRACE, because it does not model the Van Rhijn effect: elevation dependent brightness levels experienced when observing through the Earth's atmosphere [13, 24]. In this paper, master flats are constructed from the normalized pixel by pixel median of a large number of dark-subtracted images taken on the same night at the same wavelength, bandwidth, filter temperature etc. When performed with enough images (10-20), this method minimises star trails but additional star subtraction techniques are sometimes still applied.

The wavelength of the images used to create the master flat by this method can be either the airglow-free wavelength [13] (in our case 572.4 nm) or the observed wavelength: that is, the same wavelength as the image to be corrected [4, 22, 25]. Flats formed from airglow-free wavelengths remove lens to CCD non-uniformities well but do not affect aberrations that are a function of the observing wavelength such as etaloning. Flats formed from the observed

wavelength remove lens to CCD non-uniformities, including wavelength dependent features, but may inadvertently remove persistent stationary airglow during the flattening process. In this paper master flats are constructed from the observed wavelength because wavelength dependent features were not apparent and airglow-free wavelength data was not always available.

An example of the normalised counts distribution across the centre of a 557.7 nm master flat in the x (rows) and y (columns) direction is shown in Figure 4. The shape is comparable to the theoretical curves produced by Kubota et al. [26]. Values at the edges of each graph have been affected by masking and trees.

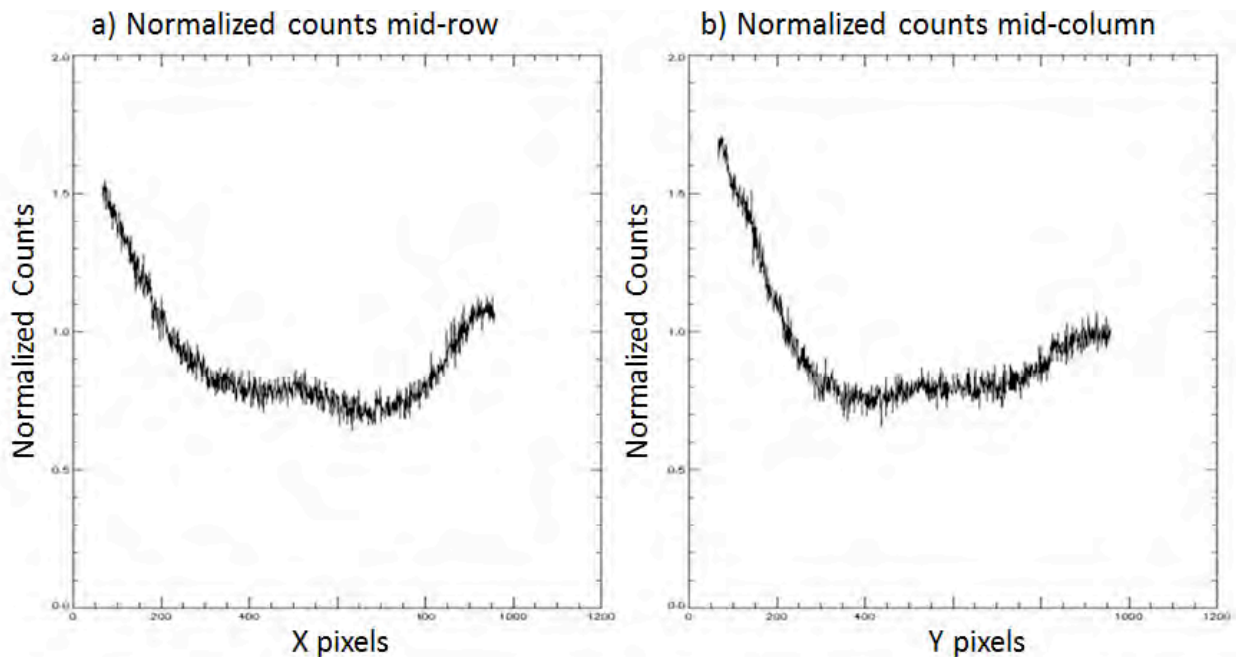


Figure 4 Normalized count value for pixels in the middle row and column of a 557.7 nm master flat image.

Further image processing such as re-gridding to geographic co-ordinates and star subtraction are planned but not yet implemented.

Results

TRACE has collected airglow images in two locations: Alice Springs in the Northern Territory in November 2011 and DSTO Edinburgh, Adelaide in South Australia from February 2012 onwards.

Alice Springs data

While stationed at Alice Springs, TRACE imaged E layer (557.7 nm) airglow using 2 minute exposures and F layer (630.0 nm) airglow using exposures of 2, 4 and 8 minutes.

An example of 557.7 nm airglow is shown in the time sequential images of Figure 5. North is at the top and East is to the left. The images are approximately 6 minutes apart from 12:27:18

UT to 12:43:29 UT. Wavefronts in the 557.7 nm emission span East to West and move from North to South (as indicated by the green arrow), as is commonly seen at this time of year in the northern parts of Australia [3].



Figure 5 North to South moving green Oxygen 557.7 nm wavefronts from 12:27:18 UT to 12:43:29 UT. The green arrow shows the direction of wave motion.

Similar moving wavefronts were observed in 630.0 nm emissions but the wavefronts were much fainter and broader than those seen in the 557.7 nm emission.

Figure 6 shows 630.0 nm airglow emission observed on November 25 at Alice Springs. The images have exposure times of 8 minutes and are approximately 12 minutes apart from 13:29:26 UT to 14:17:41. A North-East/South-West aligned depletion is seen to pass through from the SE to the NW (as indicated by the red arrow). F region depletions were more commonly detected on longer integration times (8 minutes) than shorter (2 minutes).

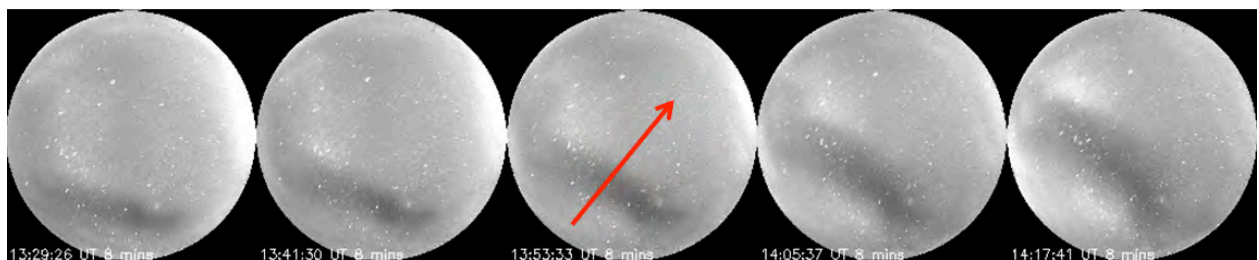


Figure 6 Oxygen 630.0 nm emission images from 13:29:26 UT to 14:17:41 UT showing a moving depletion. The red arrow shows the direction of movement.

The Darwin based OMTI imager did not record data while TRACE was stationed in Alice Springs so no comparison of F region airglow features could be made between OMTI and TRACE but the University of Adelaide's Alice Springs based imager, located 21 km East and 33 km South of the TRACE site, was operational and confirmed E region features seen in TRACE images.

Figure 7 compares The University of Adelaide's 866 nm (altitude 94 km) image and the TRACE 557.7 nm (altitude 96 km) image at approximately the same time. Similar E region features are observed by the two imagers. The University of Adelaide image has been rotated by 100 degrees to align the image to the same azimuthal bearing as the TRACE image. The TRACE FOV covers approximately 400 km whereas the Adelaide University FOV covers ~110 km x 70 km. No quantified comparison between the two images is made at this time as image processing techniques for TRACE are still in development.

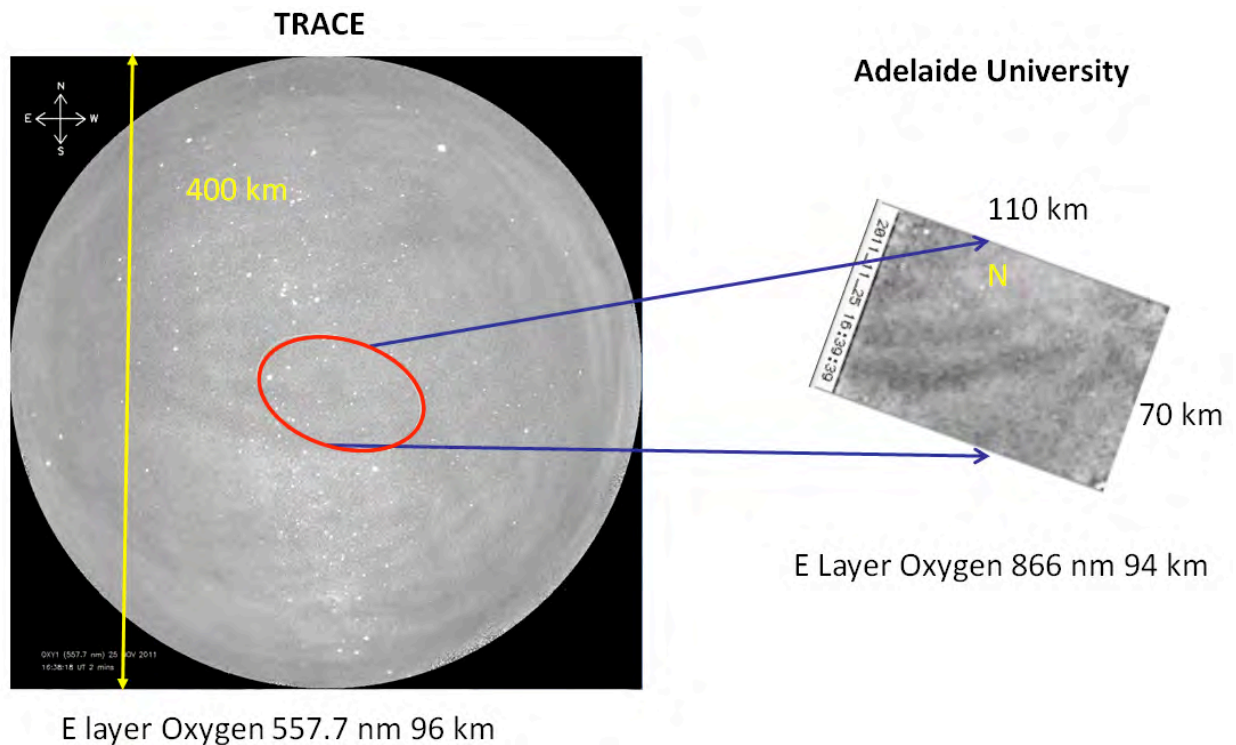


Figure 7 Comparison of TRACE (557.7 nm) and Adelaide University (866 nm) E region images. Oxygen 866 nm image reprinted with permission of The Aerospace Corporation.

DSTO Adelaide data

At the beginning of 2012, TRACE was re-located to DSTO Edinburgh where the FOV is impinged upon by trees. At this site, it monitors airglow at wavelengths of 557.7 nm, 630.0 nm and 777.4 nm as well as imaging the airglow-free wavelength 572.4 nm. In general, each filter is observed for two minutes and the open sky is imaged for 30 seconds. This section briefly covers an unusual E-layer feature observed at this location which will be covered in more detail in a subsequent paper.

Figure 8 shows a series of five 557.7 nm flat-corrected images taken approximately 7 minutes apart from 17:36:03 UT to 18:03:28 UT at DSTO Edinburgh. The figure shows a distinct North-West to South-East oriented wave packet feature (blue line) travelling from North-East to South-West (green arrow) with horizontal wavelength ~ 23 km, horizontal phase velocity ~ 20 m/s and group velocity of ~ 40 m/s. Horizontal group and phase speeds are perpendicular. These values are consistent with those produced by small-scale mesospheric waves [3, 27, 28].

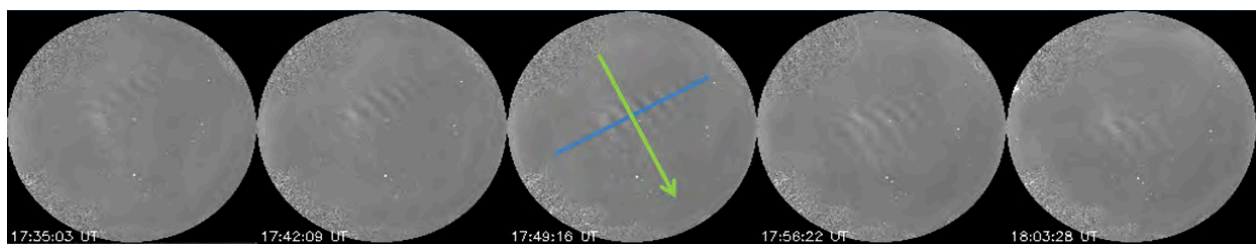


Figure 8 Oxygen 577.7 nm banding seen at DSTO, Edinburgh from 17:36:03 UT to 18:03:28 UT. The direction of phase velocity (blue) and group velocity (green) are perpendicular.

Figure 9 compares two of the TRACE 557.7 nm images from Figure 8 to an 866 nm image from the University of Adelaide's Buckland Park airglow imager situated 10 km North and 18 km West of TRACE. The approximate region of sky encompassed by the University of Adelaide's image is indicated by the ellipse on the TRACE image. Clearly the same feature is seen in both images. The next night, similar features were seen travelling in the opposite direction in 557.7 nm airglow images.

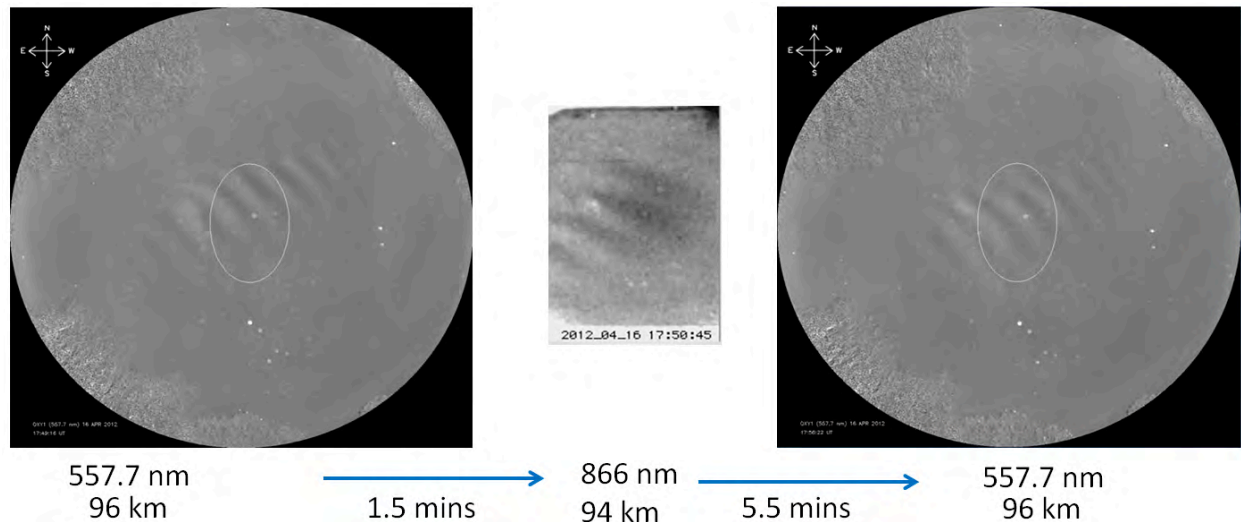


Figure 9 Comparison of E layer banding seen at DSTO Edinburgh with TRACE (557.7 nm) and Adelaide University's imager at Buckland Park (866 nm). Oxygen 866 nm image reprinted with permission of The Aerospace Corporation

Future

Having demonstrated the data quality and the re-locatability of the TRACE container, collaborations have been initiated with several interested groups including Curtin University, LaTrobe University and The University of Adelaide. Two additional low-cost airglow cameras have been acquired and a new container is being designed to accommodate the expanded experiment. An alternate long-term site at the Murray Bridge Training Area (MUTA), South Australia is under evaluation to replace the light-affected DSTO, Edinburgh site.

In 2014, it is planned that TRACE participate in a joint experiment with Curtin University at the newly commissioned Murchison Widefield Array (MWA) to compare airglow images with MWA TEC differential ionospheric calibration maps [29].

Conclusion

The Thermospheric Radar Airglow Correlation Experiment presents a unique opportunity to image both E and F region airglow emission around Australia with a large field of view and high sensitivity. It may be used alone or in conjunction with existing airglow imagers, GPS receivers, high frequency ionospheric sounders or other instruments.

Acknowledgments

We wish to thank The University of Adelaide for the use of their 866 nm airglow images which were reprinted with permission of The Aerospace Corporation.

References

- [1] G. K. Mukherjee, "Airglow and other F-layer variations in the Indian sector during the geomagnetic storm of February 5-7, 2000," *Earth Planets Space*, vol. 58, pp. 623-632, 2006.
- [2] A. E. Potter and B. S. Del Duca, "Origin of the sodium airglow," *J. Geophys. Res.*, vol. 65, pp. 3915-3923, 1960, doi:10.1029/JZ065i012p03915.
- [3] S. Suzuki, *et al.*, "Statistical characteristics of gravity waves observed by an all-sky imager at Darwin, Australia," *J. Geophys. Res.-Atmos*, vol. 109, p. D20S07, 2004, doi:10.1029/2003jd004336.
- [4] J. H. Hecht, R. L. Walterscheid, and M. N. Ross, "First measurements of the two-dimensional horizontal wave number spectrum from CCD images of the nightglow," *J. Geophys. Res.-Space*, vol. 99, pp. 11449-11460, 1994, doi:10.1029/94ja00584.
- [5] D. Fukushima, *et al.*, "Observation of equatorial nighttime medium-scale traveling ionospheric disturbances in 630-nm airglow images over 7 years," *J. Geophys. Res.*, vol. 117, p. A10324, 2012, doi:10.1029/2012ja017758.
- [6] K. J. W. Lynn, Y. Otsuka, and K. Shiokawa, "Simultaneous observations at Darwin of equatorial bubbles by ionosonde-based range/time displays and airglow imaging," *Geophys. Res. Lett.*, vol. 38, p. L23101, 2011, doi:10.1029/2011gl049856.
- [7] K. Shiokawa, Y. Otsuka, and T. Ogawa, "Propagation characteristics of nighttime mesospheric and thermospheric waves observed by optical mesosphere thermosphere imagers at middle and low latitudes," *Earth Planets Space*, vol. 61, pp. 479-491, 2009.
- [8] B. P. Williams, *et al.*, "Coincident extremely large sporadic sodium and sporadic E layers observed in the lower thermosphere over Colorado and Utah," *Ann. Geophys.*, vol. 25, pp. 3-8, 2007, doi:10.5194/angeo-25-3-2007.
- [9] T. S. Trondsen. (2012, March 29). *Low Light Level Imaging Instrumentation for Airglow - Keo Scientific*. Available: <http://www.keoscientific.com/aeronomy-imagers.php>
- [10] M. Schirmer. (2010). *About superflattening and defringing*. Available: <http://www.astro.uni-bonn.de/theli/gui/aboutsuperflattening.html>
- [11] P. Instruments, "Princeton Instruments Certificate of Performance," ed, 2011.
- [12] T. Ogawa, *et al.*, "Simultaneous ground- and satellite-based airglow observations of geomagnetic conjugate plasma bubbles in the equatorial anomaly," *Earth Planets Space*, vol. 57, pp. 385-392, 2005.
- [13] F. J. Garcia, M. J. Taylor, and M. C. Kelley, "Two-dimensional spectral analysis of mesospheric airglow imagedata," *Appl. Optics*, vol. 36, pp. 7374-7385, 1997.
- [14] K. Shiokawa, *et al.*, "Development of Optical Mesosphere Thermosphere Imagers (OMTI)," *Earth Planets Space*, vol. 51, pp. 887-896, 1999.

- [15] N. U. Solar-Terrestrial Environment Laboratory. (2013). *Station Locations, Observed Airglow Lines, and Time Resolutions* Available: <http://stdb2.stelab.nagoya-u.ac.jp/omti/stations.html>
- [16] K. Shiokawa, *et al.*, "Development of low-cost sky-scanning Fabry-Perot interferometers for airglow and auroral studies," *Earth Planets Space*, vol. 64, pp. 1033-1046, 2012, doi:10.5047/eps.2012.05.004.
- [17] Unihedron, "Sky Quality Meter -LU Users manual," Optcorp, Ed., 1.1 ed, 2006.
- [18] B. S. Corporation, "Cloud Sensor II User's Manual," ed, 2009, p. 58.
- [19] J. Marchant, R. J. Smith, and I. A. Steele, "Calibration of the Boltwood Cloud Sensor," presented at the Proc. SPIE 7012, Ground-based and Airborne Telescopes II, 2008, doi:10.1117/12.789072.
- [20] S. E. Harris, *et al.*, "THEMIS Ground Based Observatory System Design," *Space Sci. Rev.*, vol. 141, pp. 213-233, 2008/12/01 2008, doi:10.1007/s11214-007-9294-z.
- [21] M. P. Levesque and M. Lelievre, "Evaluation of the Accuracy of the Dark Frame Subtraction Method in CCD Image Processing," Defence Research and Development Canada - Valcartier TN 2007-343, December 2007.
- [22] P. LeFevre. (2003, March 29). *Chapter 2: Calibration -- Darks and Flats*. Available: <http://www.lefevre.darkhorizons.org/articles/proctutorialchap2.htm>
- [23] A. Layden. (2000). *Taking Twilight Sky Flats*. Available: http://physics.bgsu.edu/~layden/A309/ccd_obs_extras.htm
- [24] V. L. Narayanan, S. Gurubaran, and K. Emperumal, "Imaging observations of upper mesospheric nightglow emissions from Tirunelveli (8.7oN)," *Indian J. Radio Space*, vol. 38, pp. 150-158, 2009.
- [25] M. Schirmer. (2013, March 29). *CCD data reduction: Superflat*. Available: <http://www.astro.uni-bonn.de/~mischa/datareduction/superflat.html>
- [26] M. Kubota, H. Fukunishi, and S. Okano, "Characteristics of medium and large-scale TIDs over Japan derived from OI 630-nm nightglow observation," *Earth Planets Space*, vol. 53, pp. 741-751, 2001.
- [27] R. L. Walterscheid, *et al.*, "Analysis and interpretation of airglow and radar observations of quasi-monochromatic gravity waves in the upper mesosphere and lower thermosphere over Adelaide, Australia (35°S, 138°E)," *J. Atmos. Sol-Terr. Phy.*, vol. 61, pp. 461-478, 1999, doi:10.1016/S1364-6826(99)00002-4.
- [28] D. B. Simkhada, *et al.*, "Analysis and Modeling of Ducted and Evanescent Gravity Waves Observed in the Hawaiian Airglow," *Ann. Geophys.*, vol. 27, pp. 3213-3224, 2009, doi:10.5194/angeo-27-3213-2009.
- [29] A. Coster, *et al.*, "Using the Murchison Widefield Array to observe midlatitude space weather," *Radio Sci.*, vol. 47, p. RS0K07, 2012, doi:10.1029/2012rs004993.

An ultra-stable microwave transfer technique for the ACES and future space missions

Naomi Altman¹, Sascha Schediwy¹, John McFerran¹, Michael Tobar¹, and Andre Luiten²

¹School of Physics, University of Western Australia, Perth, WA 6009, Australia

²Institute for Photonics and Advanced Sensing (IPAS) and School of Chemistry & Physics,
The University of Adelaide, Adelaide, SA 5005, Australia
naomi.altman@uwa.edu.au

Summary: We present an actuation technique for suppressing uncorrelated noise between two optical frequencies to obtain ultra-stable microwave frequency dissemination over optical fibre. This technique will enable dual optical and microwave frequency transfer between the University of Western Australia and Yarragadee Satellite Laser Ranging station, which has application for the Atomic Clock Ensemble in Space (ACES) mission and other precision navigation and timing technologies. We build on the technique presented in [1], which demonstrates microwave transfer stability of 10^{-14} at 1s integration time by stabilising the group delay of an two correlated optical signals. A second actuator has been developed to suppress the remaining uncorrelated phase fluctuations between two optical frequencies by implementing optical-wavelength-scale action to vary the microwave phase. The actuator has been designed with the aid of a simulation that predicts actuation bandwidth and loss for a range of actuator parameters.

Keywords: Time and frequency transfer, ACES mission, optical simulation, fibre optics.

Introduction

The comparison of atomic clock signals is the foundation of many scientific and technological endeavours, including quantum physics, cosmology, metrology, geodesy, gravimetry, precision timing, and navigation. The latest generation of optical atomic clocks have produced timing signals with unprecedented accuracy and stability, offering an order of magnitude improvement over similar clocks based on microwave electronic transitions. Recently, an optical atomic clock demonstrated 10^{-18} fractional frequency instability for 1 day averaging time [2]. Existing techniques for the distribution of timing signals, such as two-way satellite time transfer, or through global positioning system, are limited to 10^{-15} instability at 1 day [3], thus are no longer sufficiently stable to convey the full stability of atomic clock signals.

One high profile application facing this precision bottleneck is the European Space Agency's Atomic Clock Ensemble in Space (ACES) mission, which will install atomic clocks on board the International Space Station to create a space-based time and frequency transfer system with fractional frequency stability better than 10^{-16} at 1 day [4]. The cold atom caesium clock and space hydrogen maser in the ACES payload have a fractional frequency stability of 2×10^{-13} at 10s integration time, compared to the on-board microwave time-transfer link (MWL) of 4.3×10^{-13} at 10s [5]. There is potential to improve the performance of the ACES timescale by using the intrinsically

higher stability of optical transfer techniques, for example by using the on-board European Laser Timing (ELT) optical link [6].

Researchers at the University of Western Australia (UWA) are in the process of constructing a state-of-the-art optical atomic clock, which is expected to be completed prior to the ACES launch. This high-quality frequency reference helped secure UWA's selection as the only ACES ground station in the Southern Hemisphere. In addition to this optical clock, UWA also has a unique opportunity to access a world-leading satellite laser ranging station that would enable optical time and frequency transfer. This paper focuses on the recent work to expand UWA's capabilities from the existing microwave link to include optical ground-station facilities.

Time and Frequency Transfer in the Optical Domain

Time and frequency transfer by free-space laser link has many advantages over existing microwave satellite techniques including well-defined propagation paths, low jitter noise, and most importantly high temporal stability [6]. Results from the European Laser Timing Experiment indicate that an optical ground-to-space link will be 1.5 times more stable than the current ACES microwave link over short time scales [5, 7]. However, optical ground-to-space lasers are more complex than microwave transceivers, and are generally located far away from laboratories with certain high-accuracy optical atomic clocks for improved weather conditions. Fibre-based solutions are emerging as the leading technique for transfer over regional distances, offering a reliable, low-attenuation method for connecting remote laboratories and optical ground-stations.

The infrastructure required for optical space communication is available at the Yarragadee Satellite Laser Ranging facility 330km north of the UWA campus. With the construction of a 60km fibre link to an existing network, it will be possible to transfer timing signals over the resulting 500km link between UWA and Yarragadee. However, fluctuation in optical path length due to temperature and acoustic noise cause phase fluctuations in the received signal. We expect the fractional frequency stability of the unstabilised link to be greater than 10^{-13} at 1 second, based on published results for similar link lengths [8]. Time and frequency transfer between UWA and Yarragadee requires active stabilisation of the optical fibre to compensate for the environmentally induced fluctuations in optical path length.

Recently, optical-only transfer at 4×10^{-19} stability at 100s integration time was achieved on a record 1840km fibre link across Germany [9]. We extend optical-only techniques to enable dual optical and microwave transfer, which offers improved the simplicity and versatility of microwave signals combined with the intrinsic stability of an optical carrier. Single sideband modulation of a laser produces an optical carrier f_c and sideband f_s separated in frequency by the microwave signal f_m , which for our experiment was 9GHz to match existing band-pass filters [1].

We use an optical interferometer to sense length changes in the optical fibre, as outlined in Fig. 1. The link forms the long arm of the interferometer, and the beat signal produced between the outgoing and reflected optical signals encodes the phase fluctuations accumulated in a round-trip passage of the fibre. A fibre stretcher can be

used as a single actuator to directly compensate for changes in the optical path length [1]. The stretcher stabilises a virtual optical signal that is the average of the carrier and sideband, which in turn stabilises the microwave modulation product. If the phase delay (optical) and group delay (microwave) are perfectly correlated, then in principle, noise in the microwave signal will be suppressed the same fractional amount as in the optical signal.

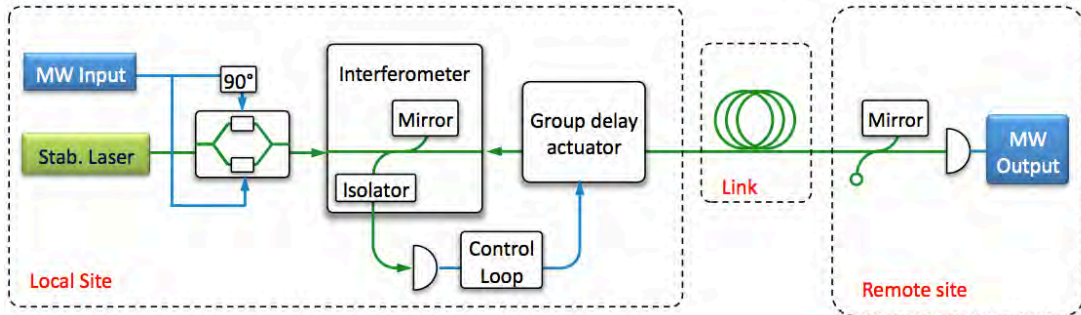


Fig. 1: A simplified view of the dual optical/microwave transfer system with using a single group delay actuator (fibre stretcher).

The preliminary results of Fig. 2 indicate suppression of the optical phase fluctuations to 1×10^{-18} at 10s on a 30km spool. However, the same performance is not reproduced in the microwave signal. While we cannot measure stability better than the noise-floor of the microwave measurement system (1×10^{-15} at 1s), we expect the actual microwave stability to follow that of the optical signal. Instead, the noise in the microwave system increases above the measurement noise-floor as a function of fibre length, indicating uncorrelated phase fluctuations between f_c and f_s . The increase of microwave noise with fibre length implies that the fluctuations are due to the material properties of the fibre, rendering this system unscalable to the lengths required to reach Yarragadee.

Similar frequency dissemination techniques have used polarization scramblers and dispersion compensating fibre to correct for microwave-only noise [10, 11]. However, chromatic dispersion (CD) and polarisation mode dispersion (PMD) should generate a constant contribution to the received phase in a continuous-wave system, suggesting that the noise is due to fluctuations in dispersion properties, and not the properties themselves.

To illustrate this point, consider SMF-28 fibre with 17ps/nm/km dispersion, which produces a 37ps difference in propagation time for a 30km link when f_c and f_s are separated by 9GHz. This additional 120° phase shift is a constant contribution on top of the optical phase shift. The observed microwave fluctuations of 10^{-14} at 1s correspond to phase fluctuations of 0.032°, or 9×10^{-5} of a cycle in that time. A change in chromatic dispersion coefficient of just 0.009% due to stress or temperature would be sufficient to cause the observed microwave noise. It is possible that the microwave noise is caused by the laser diode frequency noise, which is converted into phase noise through chromatic dispersion [11]. However, this effect is minimal in our system because we use an actively stabilised laser [1]. We confirm the above hypothesis by adding 1km of dispersion compensating fibre, resulting in no observed improvement in microwave stability.

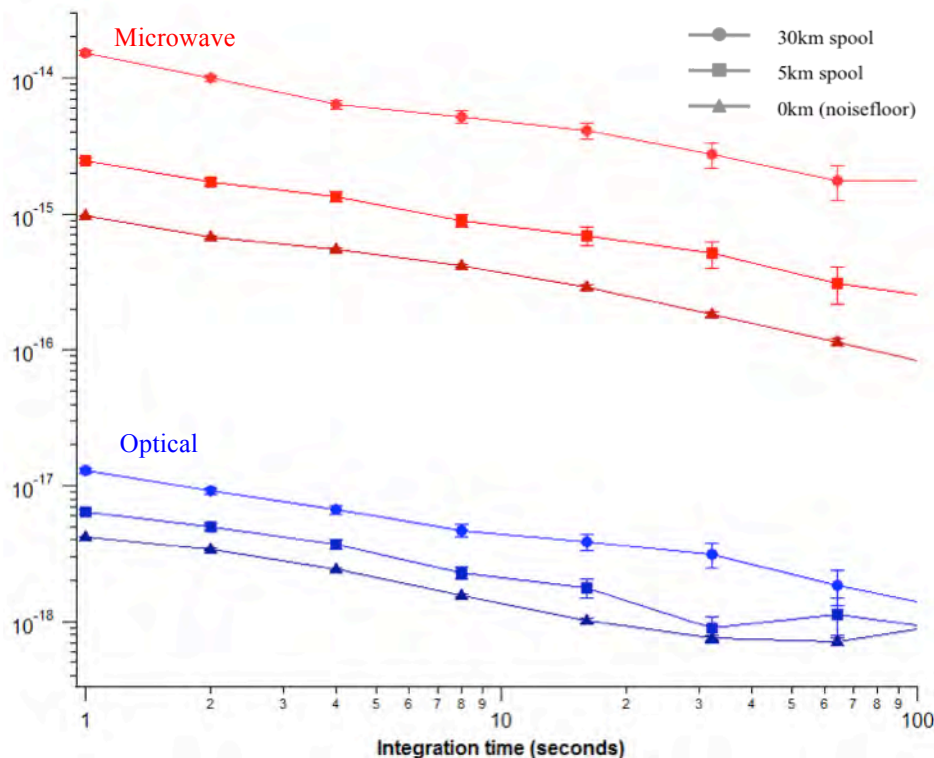


Fig. 2: Allan deviation plot of measured optical and microwave stability. The microwave instability increases about the measurement noise-floor indicating the presence of uncorrelated noise between the phases of the two optical frequencies.

An actuator for differential phase control

To correct for the uncorrelated noise between the optical carrier and sideband we require an additional actuator that implements optical-wavelength-scale action to produce a change in the microwave phase. The group delay is controlled by stabilising a virtual frequency that is the average of the two optical signals. In previous work [1], this was achieved using a fibre stretcher; relying on the high-degree of correlation between both optical frequencies. However, with the addition of a second actuator, an AOM can be used to correct the group delay, eliminating the bandwidth limitations of the fibre stretcher.

The proposed actuator splits the signal into two paths. In the configuration of Fig. 3 (a) lower path uses a Mach-Zehnder interferometer to suppress a single optical frequency through destructive interference and then a fibre stretcher to control the phase of the unsuppressed signal. The upper path transmits the signal unchanged. When the two paths recombine, the phasor addition of all optical signals produces a differential phase shift between f_c and f_s . In the alternative configuration of Fig. 3 (b), the stretcher and selective attenuator are in opposite paths. This is the preferred configuration as it minimises the differential length between the two arms of the main Mach-Zehnder interferometer, which minimises the impact of residual temperature fluctuations within the actuator.

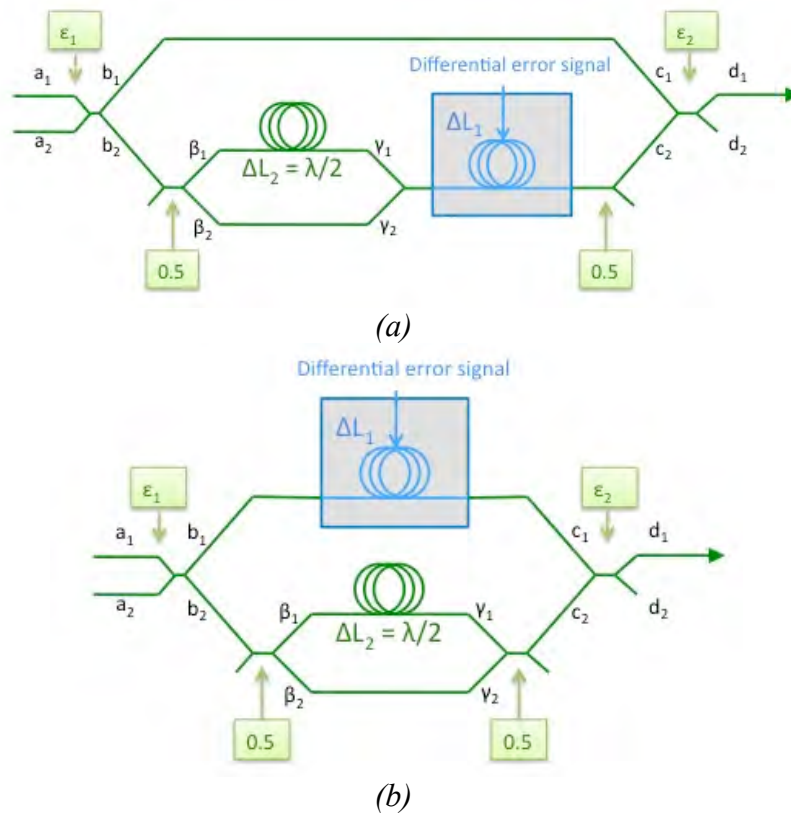


Fig. 3 (a) and (b): Two configurations for a secondary actuator to control the differential phase between a carrier and sideband. The configuration of (b) minimises the total length of fibre in the actuator.

A simulation tool for actuator design

This paper outlines a simulation tool developed within the *Mathematica* environment to guide the design of an additional actuator and verify its behaviour. The simulation displays the transfer function of the actuator system, visualises the envelope of the input and output signals, and depicts the phase relation of all signals in phasor representation. Key parameters in the transfer function, most importantly the additional length of the fibre stretcher, can be actively manipulated for complete characterisation of actuator behaviour.

At the core of the simulation is a matrix representation of the transfer function of the actuator. This is obtained by representing each optical component as an $m \times n$ matrix, where m and n are the number of output and input ports of the component respectively. The input/output relationship of a fibre section of length ΔL is:

$$b_1 = e^{j\frac{\omega n \Delta L}{c}} a_1 \quad (1)$$

where ΔL is the length of the fibre with refractive index n , and ω is the angular frequency of the signal. The input/output relationship for a 2×2 optical coupler is:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\epsilon} & j\sqrt{1-\epsilon} \\ j\sqrt{1-\epsilon} & \sqrt{\epsilon} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (2)$$

where a_1 , a_2 and b_1 , b_2 are the electrical fields at the input and output ports respectively, and ϵ is the power coupled between input port 1 and output port 1 [12]. Matrix multiplication is used to obtain a transfer function for each input-output pair of the actuator. The transfer function between input a_1 and output d_1 from Fig. 3 (b) is:

$$H(j\omega) = e^{-j\frac{\omega n \Delta L}{c}} \sqrt{1-\epsilon_1} \sqrt{1-\epsilon_2} + \frac{\sqrt{\epsilon_1} \sqrt{\epsilon_2}}{2} - \frac{1}{2} e^{-jk\frac{\omega}{\omega_{block}}} \sqrt{\epsilon_1} \sqrt{\epsilon_2} \quad (3)$$

where ϵ_1 and ϵ_2 are the splitting ratios as defined in Fig. 3 (b), ω_{block} is the angular frequency of the attenuated signal, and k an integer corresponding to the length imbalance in the attenuating interferometer.

The simulation, with example outputs shown in Fig. 4, enables a user to switch between phasor, time-domain, and transfer-function representation. Furthermore, the transfer function can be displayed both as a function of actuation length or input frequency. Within the simulation, the user can control the coarse length tuning ΔL and fine length control δL , in addition to the other actuator parameters. The transfer function representation provides the most information about actuator behaviour, showing the behaviour for both f_c and f_s as a function of actuation length, and can be used to tune the actuator to a suitable operating point ΔL . Maximum range occurs when ΔL (grey dashed line in Fig. 4) is tuned to a transmission peak of the unattenuated frequency.

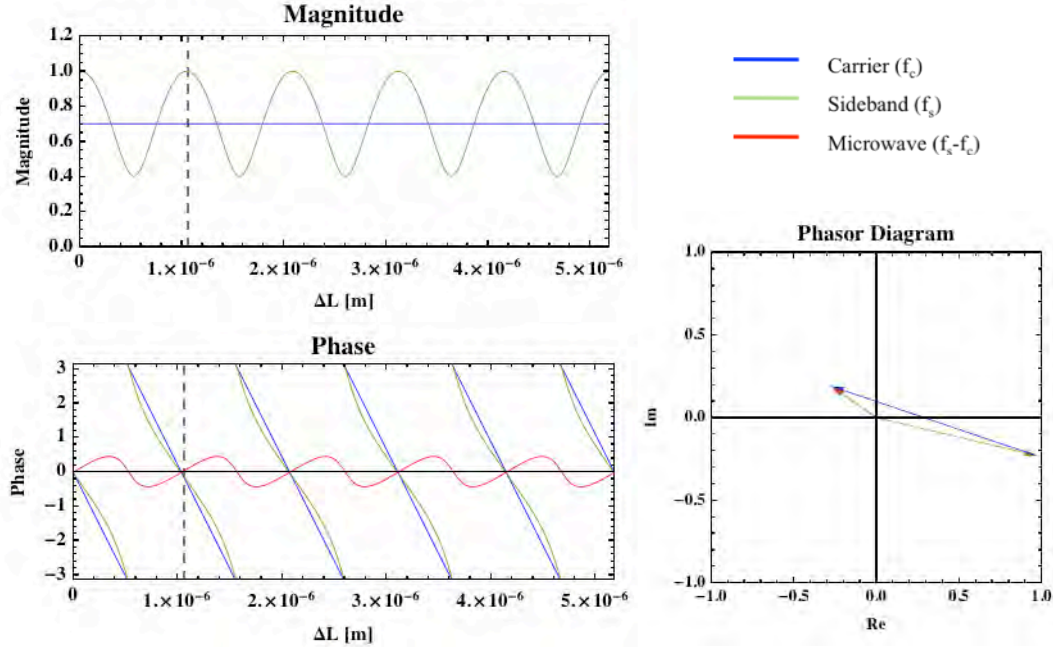


Fig. 4: Outputs of the actuator simulation, showing the transfer function of the actuator and signal phasors as a function of stretching length ΔL . The red trace shows the resulting phase of the microwave signal.

The most important utility of the simulation is the calculation of loss and actuation range, which is determined by the power coupling ratios ϵ_1 and ϵ_2 . To minimise the number of free parameters in the system, we assume $\epsilon_1 = \epsilon_2 = \epsilon$.

We illustrate the design trade-off between loss and actuation range by first considering the $\epsilon = 1$ extreme, where 100% of the power is routed into the upper path of the actuator in Fig 3. (b). The actuator is now reduced to a simple stretcher, which has negligible loss, but requires microwave-wavelength-scale action to effect a microwave phase change. The other extreme of $\epsilon = 0$ bypasses the stretcher completely, allowing for optical-wavelength-scale action, but with complete attenuation of the microwave modulation state. An optimal configuration exists between these two extremes that maximises the actuation range for a tolerable amount of loss through the actuator. The compromise between range and loss is illustrated in Fig. 5 for standard coupling ratios of $\epsilon = 0.5, 0.7$ and 0.9 .

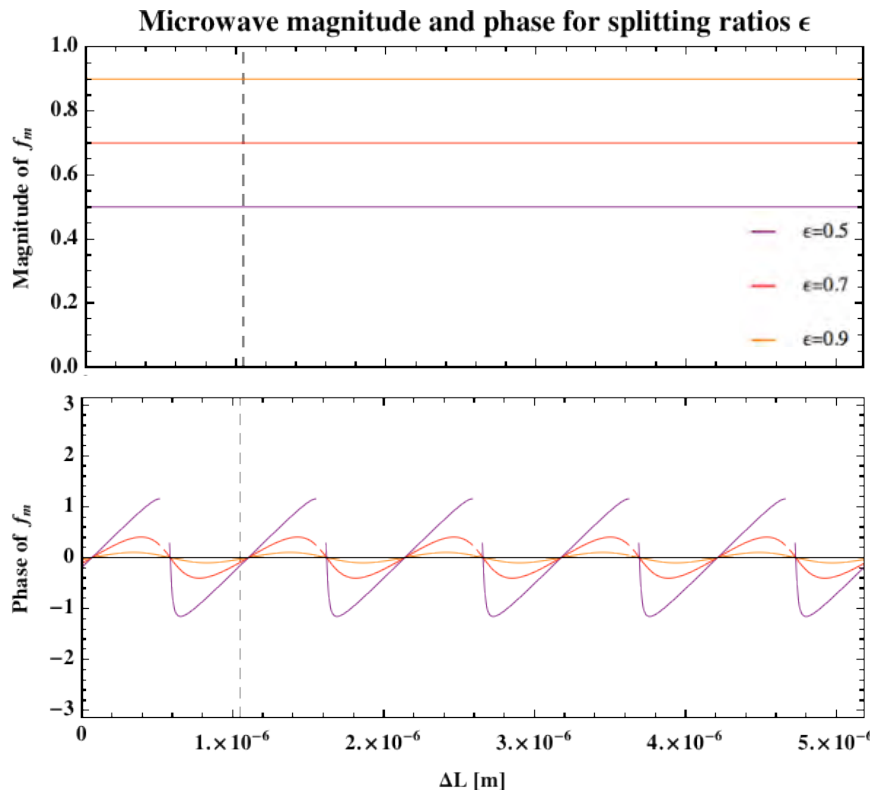


Fig. 5: Simulated microwave magnitude and phase as a function of actuation length ΔL and splitting ratio ϵ . More power to the filtering path produces greater actuation bandwidth but larger overall losses.

Simulated values for actuation range and loss are shown in Table 1 for a range of standard splitting ratios. To obtain the actuation range necessary for the Yarragadee link, the actuator is designed to correct residual microwave phase fluctuations of up to 10^{-13} at 1s, based on the measured instability of free running fibre links of similar length [8]. For a 9GHz microwave signal, this corresponds to 0.36° phase changes, or 10^{-4} of a cycle. Thus for the Yarragadee link, the 99:1 coupling ratio is the optimal choice, resulting in only -0.09dB loss for each passage through the actuator.

Table 1: Simulated range vs. loss parameters for the actuator.

Coupling ratio [path1:path 2]	Range	Loss [dB]
50:50	169.72°	-4.26
70:30	50.53°	-2.72
90:10	12.70°	-0.89
99:1	1.14°	-0.09

Simulated and real actuator performance (Results)

To embed the additional actuator in the complete system requires the extraction of a second error signal that encodes the uncorrelated phase fluctuations between f_c and f_s . This is achieved by beating the reference carrier f_c against the reflected sideband f_s , which is then passed through control circuitry and becomes the input signal to the additional actuator's fibre stretcher. The greatest challenge for implementing the actuator is thermal and mechanical isolation. Any phase fluctuations arising from the 19m of free fibre in the actuator are indistinguishable from phase fluctuations in the link. Even in an enclosed aluminium container, the actuator exhibits significant drift due to temperature instability. Further improvements are possible by using spliced components to minimise the length of fibre inside the actuator.

Preliminary experimental results show strong correlation between the simulated and experimental behaviour of the attenuating path of the secondary actuator, as shown below in Fig. 6. Here, the attenuating path (path 2 of Fig. 3 (b)) is tuned with an independent fibre stretcher to selectively filter f_c with maximum transmission of f_s . The microwave signal at the monitor port is observed directly, and degradation of fringe visibility in the experimental data is likely due to imperfect microwave-wavelength-scale matching of the interferometer arms. Work to characterise the additional actuator as part of the complete system is ongoing at the time of this publication.

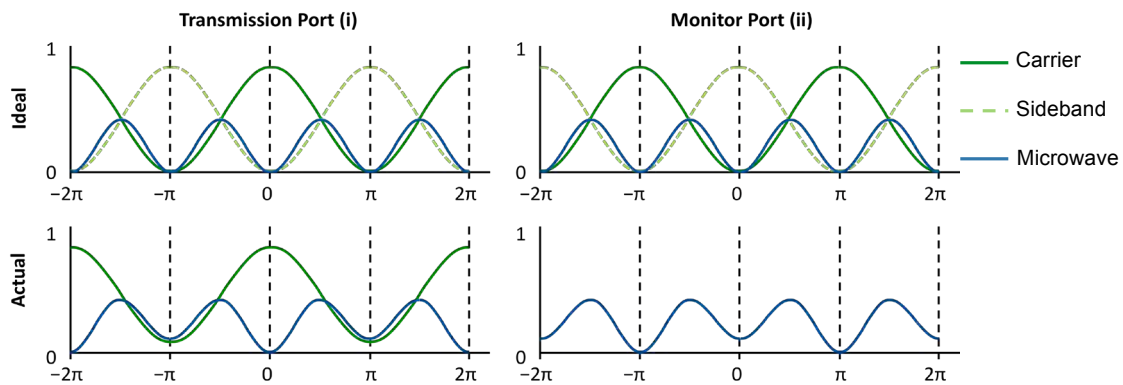


Fig. 6: Simulated and experimental results for the attenuating section of the actuator. The phase of the microwave signal at the monitor port (ii) was controlled by varying the state of the relevant fibre stretcher.

Conclusion

The simulated results of an optical fibre actuator demonstrate the suppression of residual microwave noise between two optical signals using optical-wavelength-scale

action. We outline the design of an actuator that can, in principle, correct residual microwave phase fluctuations of up to 1.14° degrees with only -0.09dB loss. This corresponds to the suppression of fractional frequency deviations of up to 3.5×10^{-13} at 1s for a 9GHz signal. The ability to transfer ultra-stable microwave signals over long distance fibre networks is central to the development of an optical ground-station for the ACES and future space missions.

Acknowledgements

This research was supported under Australian Research Council's Linkage Project (project number LP110100270) and Discovery Project (DP130100205) funding schemes. Authors Altman and Schediwy conducted the experimental work with assistance by Luiten. McFerran is developing the Ytterbium atomic clock, and Tobar is coordinating UWA's contribution to the ACES mission including the involvement with the Yarragadee Satellite Ranging Station.

References

1. Schediwy, S., et al. "Microwave frequency transfer with optical stabilisation", *European Frequency and Time Forum (EFTF)*, 2012. 2012.
2. Hinkley, N., et al., "An Atomic Clock with 10^{-18} Instability", *Science*, 2013, 341(6151): p. 1215-1218.
3. Bauch, A., et al., "Comparison between frequency standards in Europe and the USA at the 10^{-15} uncertainty level", *Metrologia*, 2006. **43**(1): p. 109.
4. Cacciapuoti, L. and C. Salomon, "Space clocks and fundamental tests: The ACES experimen" *The European Physical Journal Special Topics*, 2009, 172(1): p. 57-68.
5. Hess, M.P., et al. "ACES MWL status and test results in *Frequency Control and the European Frequency and Time Forum (FCS)*", *2011 Joint Conference of the IEEE International*, 2011.
6. Schreiber, U., et al. "The european laser timing (ELT) experiment on-board ACES. in *Frequency Control Symposium*", *2009 Joint with the 22nd European Frequency and Time forum. IEEE International*. 2009.
7. Schreiber, K.U., et al., "Ground-based demonstration of the European Laser Timing (ELT) experiment", *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 2010, 57(3): p. 728-737.
8. Lopez, O., et al., "Ultra-stable long distance optical frequency distribution using the Internet fiber network", *Optics Express*, 2012, 20(21): p. 23518-23526.
9. Droste, S., et al., "Optical-Frequency Transfer over a Single-Span 1840km Fiber Link" *Physical Review Letters*, 2013, 111(11): p. 110801.

10. Marra, G., H.S. Margolis, and D.J. Richardson, "Dissemination of an optical frequency comb over fiber with 3×10^{-18} fractional accuracy", *Optics Express*, 2012, 20(2): p. 1775-1782.
11. Lopez, O., et al., "High-resolution microwave frequency dissemination on an 86-km urban optical link", *Applied Physics B*, 2010, 98(4): p. 723-727.
12. Hui, R. and M. O'Sullivan, *Basic Instrumentation for Optical Measurement*, in *Fiber Optic Measurement Techniques*, R. Hui and M. O'Sullivan, 2009, Academic Press: Boston. p. 129-258.

Retrieval of Greenhouse Gas Concentrations from Observations with a Ground-based Spectrometer in the Near-infrared

Champlain Kenyi, Jeremy Bailey and Daniel V. Cotton

School of Physics, The University of New South Wales, NSW 2052, Australia

Summary: The accurate measurement of the atmospheric concentration of greenhouse gases such as carbon dioxide and methane is the goal of a number of space missions such as the Japanese GOSAT satellite and the forthcoming NASA OCO-2 mission. However, obtaining accurate measurements from space is challenging due to the effects of aerosols and clouds. Ground-based total column measurements largely avoid such difficulties and can thus complement, and help to validate space instruments. Here we report progress on measuring atmospheric trace gases, particularly carbon dioxide (at $1.58\ \mu\text{m}$), using a ground-based fibre Fabry-Perot spectrometer; the data is fitted with an iterative non-linear least squares fit algorithm in order to estimate the atmospheric states of interest. We describe current progress with the development of the retrieval algorithm and some of the issues that need to be overcome to reliably measure greenhouse gases with such an instrument.

Keywords: Fibre optic, near-infrared, spectroscopy, retrieval, ground-based, greenhouse gases

Introduction

The observation of atmospheric carbon dioxide (CO_2) from space platforms is important for improving our understanding of its spatial and temporal distribution in the atmosphere [1], but the existing measuring instruments have limitations. Whilst the satellites can provide global coverage, especially over the oceans and tropical regions that are poorly sampled by the ground-based networks, their ability to estimate the column-averaged gas amounts is compromised by the scattering effects of the atmosphere (e.g. aerosols and clouds that modify the light path), making it hard to meet the stringent requirements on retrieval accuracy [1]. Detecting trace gas concentrations over the oceans is also difficult – as noted with the Japanese Greenhouse-gases Observing SATellite, (GOSAT) – due to the low reflectivity of the sea. The GOSAT retrieval algorithms for determining the dry-air column-averaged mole fractions of CO_2 and methane (CH_4) are under continuing development and require reliable data for evaluation [2]. One source of reliable data for such a task is the Total Carbon Column Observing Network (TCCON) exploiting high-resolution Fourier Transform Spectrometers [3].

Recent studies of atmospheric carbon sources and sinks, and their concentrations and distributions, have largely relied on *in situ* surface measurements [2, 4] as well as from tall towers augmented by aircraft campaigns [5, 6]. The *in situ* measurements alone are insufficient because the existing network is far too sparse to resolve uncertainties associated with atmospheric mixing, and the aliasing between atmospheric transport and surface emission; local carbon sources further complicate their interpretation [5, 7]. In contrast, column measurements are much less affected by variability due to vertical transport of the tracer within the atmospheric column [2, 3] since the column vertically integrates the concentration of the gas above the surface. Column measurements also have a large sampling

“footprint”; they are less sensitive to surface sources and sinks; as a result a combination of the measurements and simulations are expected to be particularly useful for improving our understanding of the carbon cycle [2, 3, 8].

The technology of monitoring total carbon column densities of greenhouse gases from surface sites is indeed highly regarded in the carbon cycle and provides a validation check on the space-based instruments. While the TCCON data are highly reliable and used for the validation tasks [9, 10], its coverage is poor in many developing countries (e.g. in Asia, Africa and South America) and over the oceans. This probably is contributed by the fact that the FTS instruments are expensive, bulky and heavy to transport; they require stable platform to set up, air-conditioned laboratory space and support infrastructure [11, 12].

As a way to address some of these bottlenecks, there has been investigation of the possibility of making similar measurements with smaller, more portable and low cost instruments. Kobayashi et al. [13] have demonstrated the capability to operate both grating-based desktop optical spectrum analyser (OSA) and fibre Fabry-Perot interferometer (FFPI) to monitor CO₂ and CH₄ column densities. In doing so direct collimated sun light is accepted through an optical fibre via a small telescope installed on a portable sun tracker and transmitted to either the OSA or FFPI for optical analysis. A similar approach based on FFPI for measuring CO₂ column has been described by Wilson et al. [14]. Petri et al. [15] and Gisi et al. [16] have also described the use of smaller and hence lower resolution Fourier-Transform spectrometers, while Kawasaki et al. [12] describes the use of OSA based on a grating spectrometer. The instrument used for our work is fully described by Bailey [11]; it rather uses a very high finesse (~5000) Fabry-Perot etalon operating in low order (~13).

In this paper we present physical retrieval estimates of trace gas total column densities from measured radiance spectra in the near infrared, using a ground-based Fabry-Perot spectrometer built from commercially available fibre optic devices. Brief outlines and descriptions on measurement procedures, and the flux inversions from the observations to profile amounts are included in the manuscript.

Our retrieval (inversion) algorithm uses modified Levenberg-Marquardt routines [17], in which the input parameters of the forward model are optimised to generate a synthetic spectrum that best fits the observed data, subject to prior constraints (i.e. it varies the mixing ratios of the absorbing gas in the model spectrum in order to better match the absorption line profile to the observed spectrum); this is the key defining feature of our version of the retrieval scheme.

Model Description

The forward model used to predict the spectrum is an application of VSTAR (Versatile Software for Transfer of Atmospheric Radiation) software [18, 19]; it uses the line-by-line method to predict the transmission spectrum of a multilayer atmosphere. Whilst the computational costs involved are high, the retrieval scheme is fast enough for the whole spectral signature of the target trace gas molecule to be used. Here, by retrieval scheme (algorithm), we are referring to the procedure for seeking an atmospheric state (e.g. CO₂ or H₂O column density) whose simulated transmission best fits the observations.

The line-by-line modelling used in this work is preferred for its high accuracy; however its low computational efficiency (contributed to by the large number of spectral lines) poses practical difficulties when vast amounts of data have to be processed, since our forward model must be run at each iteration point. This computational burden could feasibly be

minimised by reducing the number of atmospheric layers/levels considered, but by doing so trades-off accuracy. Other practical mitigating strategies include applying a high spectral resolution to reduce the degree of spectral line overlap, cut down the number of data points and perhaps the number of fit parameters, as well as keeping a library of forward models corresponding to the expected values of the fit parameters.

Observations and Fitting Procedure

Measurement method

The measurements are made using a compact and portable fibre-optic Fabry-Perot (FP) spectrometer as described by Bailey [11]. The spectrometer is battery powered, connects to a laptop computer via a USB cable, and accepts light via a single-mode fibre-optic cable. Two of these spectrometers have now been built. Each spectrometer is based on a fibre-optic Fabry-Perot tuneable filter (Micron Optics FFP-TF2). These devices are Fabry-Perot etalons constructed entirely in single-mode fibre and can be piezoelectrically scanned by applying a computer controlled drive voltage in the range 0 to 60 V. The devices have very high finesse (5000-6000) and large free spectral range, allowing a large number of independent spectral points to be scanned. We have three FP tuneable filters available, two covering the range 1.54 – 1.67 μm (including CO_2 and CH_4 bands) and one covering the range 1.26 – 1.36 μm covering the O_2 a-X band.

We have carried out measurements with the spectrometers at two locations. The first location is the UNSW observatory site, which is on the roof of a building on UNSW's Kensington campus. At this location we use a low cost sun-tracking system based on an amateur astronomy telescope mount as described by Bailey [11]. The other location used was the UoW atmospheric chemistry laboratory, where we set up our two spectrometers alongside the TCCON FTIR spectrometer [3]. In this case we used the same sun-tracking heliostat to simultaneously feed both the FTIR spectrometer and the fibre feeds to the two UNSW spectrometers, allowing simultaneous measurements.

Our standard measurement consists of a scan of 3000 – 4000 spectral points made with an exposure time of 200 ms per point. A 4000 point spectrum takes 13.3 minutes to scan and covers about 150 cm^{-1} . The signal to noise ratio (S/N) achieved on such a solar spectrum, after recent improvements to the spectrometer, is greater than 2000. This suggests that in the future, shorter integration times and faster scans could be used while still achieving good S/N. The spectral resolution ($R = \lambda/\Delta\lambda$) of our instrument is about 70000.

For use in the retrieval software the observed spectra are processed as described by Bailey [11] to remove zero point error (based on dark measurements made at the start and end of every scan), to calibrate in wavenumber, and to calibrate the “ripple” in the transmission of the order-sorting interference filter.

Model Fitting Procedure

The fitting of models to observed spectra is carried out using a software package called ATMOF (ATMOspheric Fitting) [20] [13]. ATMOF performs iterative fitting of a model to an observed spectrum using the Levenberg-Marquardt non-linear least-squares algorithm [17].

Details of the forward modelling procedures and validation are presented in *Kenyi et al.* [21] and Bailey [11]. The forward model makes use of the VSTAR software to calculate the

atmospheric transmission using the line-by-line method with spectral line data from (High resolution TRANsmission) HITRAN 2008 [22]. The calculated atmospheric transmission is multiplied by the input solar spectrum, at a resolution that fully resolves the lines, to obtain the observed solar flux at the surface. Our input solar spectrum used for the model is discussed in detail elsewhere [11, 20, 21]; it is derived by combining the high resolution Kitt Peak Solar Atlas with the model-based Kurucz flux distribution [23].

A feature of Fabry-Perot spectrometers is that the instrumental line shape has extended wings, and there can thus be significant effects on the overall transmission from spectral lines at large distances from the line centre. To properly model these contributions the forward model must use the correct line shape function. We have found that the line shape of our spectrometer is stable and is very well modelled by the theoretical Airy profile for a Fabry-Perot etalon.

This forward model also includes the transmission function for the interference filter used for order sorting. It is important that this is included in the forward model, rather than applied as a correction to the data, because it is the filter shape that places a limit on the far-wing contributions of the instrumental line shape.

The model requires an *a priori* atmospheric profile including the pressure-temperature structure and the abundances of important species. For the model fits described here we used a 50 layer version of the widely used mid-latitude summer atmosphere taken from the ICRCCM (InterComparison of Radiation Codes in Climate Models) project [24]. This has 1 km level spacings in the lower atmosphere up to 25 km (the region in which most of the absorption occurs) and then 2.5 km and 5 km spacings for the highest levels.

The atmospheric parameters that can be fitted are the abundances of H₂O and CO₂. These are adjusted by multiplying the *a priori* profile by a scaling factor at all altitudes. Other parameters of the model are the SZA, the solar Doppler shift, a wavelength shift applied to the whole spectrum, the spectral resolving power of the Fabry-Perot, and a scaling factor and slope to match the model spectrum to the arbitrary flux units of the measured spectrum. Each of these parameters can either be included as a fit parameter, or set to a fixed value.

Results and Discussion

This paper investigates fitting simulations to spectra observed on the 24th October 2012 and on the 9th July 2013 from our UNSW observatory and from the TCCON-FTS site at the University of Wollongong (UoW), for an assumed clear-sky scenario; each spectrum was independently fitted in the retrieval algorithm and the resulting estimates evaluated for accuracy by comparing them with the corresponding prescribed input value. Although the results detailed appear promising there are caveats: the retrieval is affected by unnoticed thin cirrus clouds and variability in surface pressures that bias the radiances, and at larger SZAs the effect of refracting bending as well as aerosols tends to influence the *a posteriori* solutions. Additional errors are also introduced through the forward model by the assumption that the layered atmosphere is plane parallel. Furthermore, inadequacies in application of the instrument (e.g. uncertainties in the instrumental line shape, spectral resolution, detector nonlinearity, zero offset errors due to uncorrected excess dark current) in addition to those due to the imperfect forward model do introduce systematic errors in the CO₂ observations and bias the retrievals.

The issue of surface pressure changes and uncertainties in the atmospheric optical path length modification introduced by aerosol and cloud scattering, and pointing errors could be present

in our model. These effects are expected to eventually be minimized by using the simultaneous O₂ column measurements (i.e., ratioing the CO₂ and O₂ column densities) since the dry air mole fraction of O₂ is well known and essentially constant. Our instrumental scan time for these spectra is currently 13.3 minutes, however in future work it is planned to cut this down by at least a factor of two to minimise artefacts relating to air mass changes. We hope to address most of these model and instrumental biases prior to carrying out proper error budget analysis (which is well beyond the scope of the paper).

Example Fitted spectrum

In this section we show a typical fit of an observed spectrum obtained with the ATMOF system. The spectrum was measured from our UNSW observing site; and is one of the data sets collected on the 4th of October 2012. This is shown in the top panel of Fig. 1 in black; and the modelled spectrum is in red. The bottom panel is the residual difference between the observed and the calculated spectrum.

It can be seen from Fig. 1 that the model fits the observations quit well with residuals well below 2 % and there is a computed root mean square error of 0.0044642. This is however, substantially higher than the measured noise in the spectra which is a factor of 3 lower in this observation. The remaining structure in the residuals is thought to result from incomplete cancellation of the filter ripple, and imperfections in the wavenumber calibration. Some differences may also arise from imperfect matching of the solar spectrum.

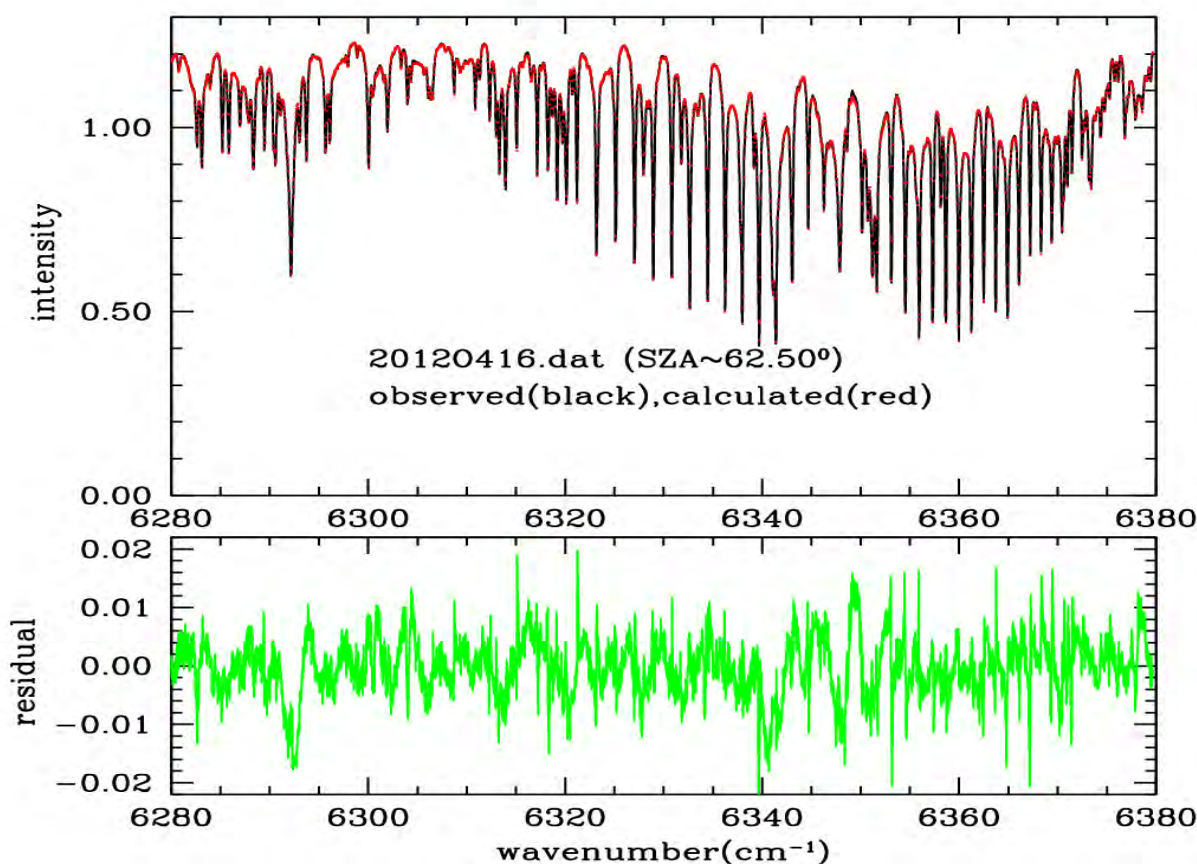


Fig. 1: an example of observed CO₂ data (at 1.58 μ m) (in black) fitted with model calculated spectrum (red); the bottom panel is the residuals difference between the observed and the calculated spectrum (green); the solar zenith angle is about 62.5°.

Retrieved CO₂ and Sensitivity to H₂O

In this section we present some results on retrieved CO₂ and on the sensitivity of CO₂ retrievals to changes in the *a priori* atmospheric profile used. The CO₂ concentrations (ppm) here do refer to total column and are determined in a similar way to TCCON.

Fig. 2 shows the concentrations of atmospheric CO₂ (in ppm) obtained with ATMOF for a set of observations at different times over October 4th 2012. It can be seen that the retrieved CO₂ varies significantly between individual observations over the same day ranging from 382 – 391 ppm. This is a larger variation than would be expected and indicates that we are not yet achieving the accuracy that is needed to be competitive with other instruments, in particular the TCCON network obtains precisions of ~0.2 ppm.

A number of factors could be responsible for the differences. Among these are the instrumental factors noted in the previous section that are leading to structure in the fit residuals, including wavelength calibration errors and filter correction artefacts. Our current detectors show significant zero point drift with temperature, which could alter apparent line strengths. Solar tracking errors could lead to significant errors in the solar zenith angle. Thin cloud at levels too low to be immediately apparent could also be responsible. Most of these factors will be addressed with improvements to the instrument currently underway.

In the results shown in Fig. 2 the fitting routine was adjusted to put all the atmospheric water vapour in one specified layer at a time rather than use the standard *a priori* water vapour profile. The retrieved abundances are plotted against the pressure of the layer containing the water vapour. The results show that the retrieved CO₂ abundance is rather insensitive to the location of water vapour in the atmosphere, because even these dramatic changes in the location of the water vapour lead to only small changes in the retrieved CO₂.

The retrieved data in Fig. 2 (b) correspond to a fitted spectrum we observed with our instrument at the TCCON site, located within the Atmospheric Chemistry Department of the UoW; the SZA is ~57.52°. The retrieved profile trend is similar to the other data sets seen in Fig. 2 (a), but reproduced with the error bars that were omitted there for clarity. These error bars are a characterisation of the spread of the observed concentrations from the model simulated data; specifically the difference between the retrieved gas concentration and the true gas concentration measured; the profile is not as smooth as for the synthetic retrievals, probably caused by passage of unnoticed thin cirrus cloud or instrumental drift during the scan.

Sensitivity to Temperature

To test the sensitivity of the retrievals to the atmospheric temperature profile, we generated synthetic spectra using our modelling code with the standard atmospheric profile and CO₂ and H₂O values of 400 ppm and 60 pp10000 respectively. We then fitted to these synthetic spectra using a modified version of the *a priori* model, in this case with the temperature changed by a fixed amount at each level.

The results are shown in Fig. 3: retrieved CO₂ increases with increasing temperature, whereas retrieved H₂O falls slightly with increasing temperature. Note that arbitrarily increasing the atmospheric temperature while pressures are left unchanged has the effect of decreasing the density of each atmospheric layer (in a way that is inconsistent with maintaining hydrostatic balance). More realistic variations of the temperature-pressure profile will need to be considered in the future. The CO₂ abundance therefore has to increase to maintain the same

CO₂ absorbing column. This column effect probably accounts for most of the change seen. The retrieved H₂O shows a different behaviour and this is probably because in the case of H₂O we are not observing complete bands, but outlying lines that may have significant intensity dependence with temperature. For CO₂ we observe a complete band that will show redistribution of absorption between its different rotational lines as temperature changes, but should show smaller changes in total absorption.

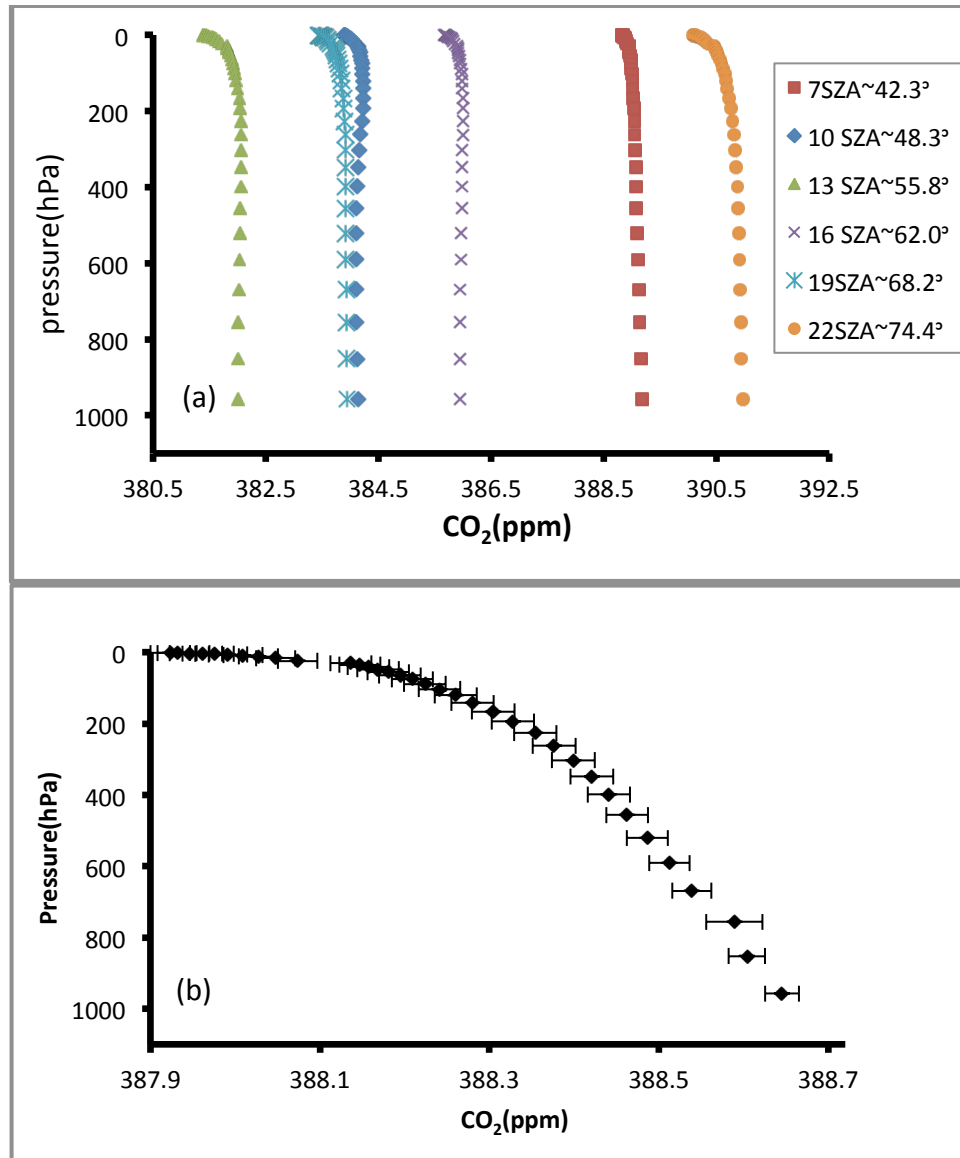


Fig. 2: variability of retrieved total column CO₂ concentrations as a function of the atmospheric layer pressure at which water vapour is placed for a number of measured spectra from our UNSW observatory (a); in the legend $n\text{SZA} \sim \theta$, n is a local data ID number, SZA denote solar zenith angle θ in degrees. The bottom panel (b) is analogous, but for a spectrum we observed with our own instrument at the TCCON_FTS location of the University of Wollongong, UoW; the SZA is 57.52°.

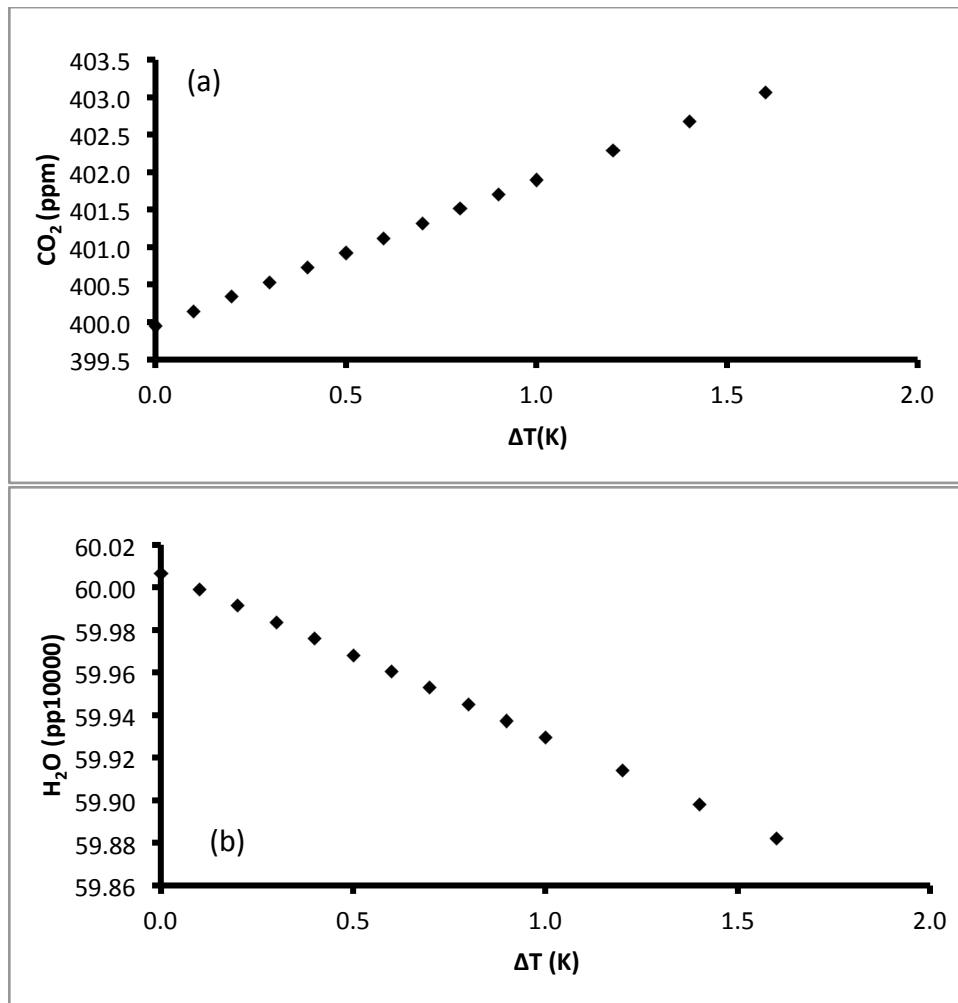


Fig. 3: sensitivity of retrieved (a) CO_2 and (b) H_2O to changes in the temperature of the a priori model.

Changes in the a priori temperature introduce errors in the retrievals. We have considered two numerical cases for temperature change ΔT before commencing the fitting procedure: firstly when ΔT is positive and therefore an enhancement to the a priori temperature –the results are apparently in Fig. 3; later ΔT was changed in sign and the fitting repeated accordingly (the estimates are not shown in the figure for convenience). Finally, the magnitudes of the partial derivatives of the retrievals versus ΔT were plotted and the Jacobians determined by integration the fitting functions in the interval 0-1.6K. The CO_2 temperature Jacobian is about 5.0ppm/K and for the H_2O -temperature it is about 0.2pp10000/K. These are large variations for the CO_2 retrievals and low for H_2O for reasons outlined above, that is, we have not used a realistic case whereby the layer pressure should change too as the temperature is perturbed.

Conclusions

We have described the current state of development of retrieval software for measuring CO_2 abundance using spectra from a compact fibre-optic Fabry-Perot spectrometer. We have developed a forward model and iterative least squares fitting software package, that can be used to fit atmospheric parameters to an observed transmission spectrum. We are able to obtain reasonably good fits to Fabry-Perot spectra, but calibration issues with the

spectrometer currently limit the quality of the fits. Investigations into the sensitivity of the retrievals to the assumed *a priori* atmospheric model have begun. Preliminary results show that CO₂ retrievals are relatively insensitive to assumptions about the atmospheric water vapour profile, but are quite sensitive to the adopted atmospheric temperature profile.

The results show that improvements are still needed to the instrument and calibration procedures to obtain competitive performance for greenhouse gas retrievals. The sensitivity tests show that it will be important to use realistic *a priori* atmospheric profiles rather than the standard atmospheres used in these preliminary studies.

Acknowledgements

We thank Nick Jones and David Griffith (University of Wollongong) for allowing us to observe with our instrument at their TCCON-FTIR observatory site.

NSO/Kitt Peak FTS data used here were produced by NSF/NOAO.

References

1. Oshchepkov, S., Bril, A., Yokota, A., Yoshida, Y., Blumenstock, T., et al., "Simultaneous retrieval of atmospheric CO₂ and light path modification from space-based spectroscopic observations of greenhouse gases: methodology and application to GOSAT measurements over TCCON sites," *Appl. Opt.*, vol. 52, Iss. 6, pp. 1339-1350, 2013.
2. Belikov, D.A., Maksyutov, S., Sherlock, V., Aoki, S., Deutscher, N.M., et al., "Simulations of column-averaged CO₂ and CH₄ using the NIES TM with a 1 hybrid sigma–isentropic (σ – θ) vertical coordinate," *Atmos. Chem. Phys. Discuss.*, vol. 12, pp. 8053-8106, 2012.
3. Wunch, D., Toon, G.C., Blavier, J.-F., Washenfelter, A., Notholt, J., Conner, B.J., Griffith, D.W.T., Sherlock, V. and Wennberg, P.O., "The Total Carbon Column Observing Network (TCCON)," *Phil. Trans. R. Soc. London, Ser. A*, vol. 369, pp. 2087-2112, 2011.
4. GLOBALVIEW-CO₂, "Cooperative atmospheric data integration project – carbon dioxide", NOAA ESRL, Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW, retrieved 21/6/2013.
5. Connor, B.J., Boesch, H., Toon, G.C., Sen, B., Miller, C. and Crisp, D., "Orbiting carbon observatory: inverse method and prospective analysis," *J. Geophys. Res.*, vol. 113, art. no. D05305, 2008.
6. GLOBALVIEW-CO₂, "Cooperative atmospheric data integration project – carbon dioxide", NOAA ESRL, Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW, retrieved 21/5/2013.
7. Gurney, K.R., Law, M.R., Scott-Denning, A., Rayner, P.J., Baker, D., et al., "TransCom 3 CO₂ inversion intercomparison: 1. Annual mean control results and sensitivity to transport prior flux information," *Tellus, Ser. B*, vol. 55, pp. 555-570, 2003.
8. Yang, Z., Washenfelter, R.A., Keppel-Aleks, G., Krakauer, N.Y., Randerson, J.T., Tans, P.P., Sweeney, C. and Wennberg, P.O., "New constraints on Northern Hemisphere growing season net fl P," *Geophys. Res. Lett.*, vol. 34, pp. 1-6, 2007.
9. Washenfelter, R.A., Toon, G.C., Blavier, J.-F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D.M. and Daube, B. C., "Carbon dioxide column abundances

- at the Wisconsin Tall Tower site”, *J. Geophys. Res.*, 111, D22305, doi: 10.1029/2006JD007154, 2006.
10. Wunch, D., Toon, G.C., Wennberg, P.O., Wofsy, S.C., Stephens, B.B., Fischer, M.L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K.P., Browell, E.V., Campos, T., Connor, B.J., Daube, B.C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jimenez, R., Keppel-Aleks, G., Kort, E., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A., “ Calibration of the total carbon column observing network using aircraft profile data”, *Atmos. Meas. Tech.*, vol. 3, pp. 1351–1362, 2010, doi:10.5194/amt-3-1351-2010.
 11. Bailey, J., "A low-cost portable fibre-optic spectrometer for atmospheric absorption studies," *Atmos. Meas. Techniq. Discuss.*, vol. 6, pp. 1067-1092, 2013
 12. Kawasaki, M, Yoshioka, H, Jones, N. B., Macatangay, R., Griffith, D.W.T., Kawakami, S., Ohyama, H., Tanaka, T., Morino, I., Uchino, O. and Ibuki, T., “Usability of optical spectrum analyzer in measuring atmospheric CO₂ and CH₄ column densities: inspection with FTS and aircraft profiles in situ”, *Atmos. Meas. Tech.*, vol. 5, pp. 2593–2600, 2012 doi: 10.5194/amt-5-2593-2012.
 13. Kobayashi, N, Inoue, G., Kawasaki, M., Yoshioka H., , Minomura, M, Murata, I, Nagahama, T, Matsumi, Y., Tanaka, T, Morino I, and Ibuki, T., “Remotely operable compact instruments for measuring atmospheric CO₂ and CH₄ column densities at surface monitoring sites” *Atmos. Meas. Tech.*, vol. 3, pp. 1103–1112, 2010, doi:10.5194/amt-3-1103-2010.
 14. Wilson, E.L., Georgieva E.M. and Heaps, W.S. “Development of a Fabry–Perot Interferometer for ultra-precise measurements of column CO₂”, IOP PUBLISHING, *Meas. Sci. Technol.* 18 (2007) 1495–1502, doi:10.1088/0957-0233/18/5/040.
 15. Petri, C., Warneke, T, Jones, N, Ridder, T., Messerschmidt, J., Weinzierl, T., Geibel, M. and Notholt, J., “Remote sensing of CO₂ and CH₄ using solar absorption spectrometry with a low resolution spectrometer”, *Atmos. Meas. Tech.*, vol. 5, pp. 1627–1635, 2012.
 16. Gisi, M, Hase, F., Dohe, S., Blumenstock, T., Simon, A., and Keens, A., “XCO₂-measurements with a tabletop FTS using solar absorption spectroscopy”, *Atmos. Meas. Tech.*, vol. 5, pp. 2969–2980, 2012 doi: 10.5194/amt-5-2969-2012.
 17. Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., "Numerical Recipes in FORTRAN: The Art of Scientific Computing," 2nd ed. Place Published: Cambridge University Press, 1992.
 18. Bailey, J., "VSTAR - A new high-spectral-resolution atmospheric radiative transfer code for Mars and other planets," presented at Second workshop on Mars atmosphere modelling and observations, *Granada, Spain*, pp. 148-151, 2006.
 19. Bailey, J. and Kedziora-Chudczer, L., "Modelling the spectra of planets, brown dwarfs and stars using VSTAR," *Mon. Not. R. Astron. Soc.*, vol. 419, pp. 1913-1929, 2012.
 20. Cotton, D.V., Bailey, J. and Kedziora-Chudczer, L., "Atmospheric modelling for the removal of telluric features from infrared planetary spectra," accepted to *Mon. Not. R. Astron. Soc.*, 2013.
 21. Kenyi, C., Cotton, D.V. and Bailey, J., "Retrieval software for total column greenhouse gas measurements from ground and space," presented at *12th Australian Space Science Conference*, Melbourne, 13: 978-0-9775740-6-3, pp. 117-126, 2013.
 22. Rothman, L.S., Gordan, L.E., Barbe, A., Benner, D.C., Bernath, P.F., et al., "The HITRAN 2008 molecular spectroscopic database," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 110, pp. 533-572, 2009.

23. Kurucz, B., "The solar irradiance by computation", <http://kurucz.harvard.edu/papers/irradiance/solarirr.tab>, retrieved 5/11/2011.
24. RFM atmospheric profiles", <http://www.atm.ox.ac.uk/RFM/atm/>, retrieved 21/6/2013.

Frequency Allocation for Satellite Space Stations using Amateur Radio Bands

Anthony Monger

School of Physics
University of Sydney
NSW, Australia

Summary: The cost of using commercial Radio Frequency (RF) bands is prohibitive for University and most private spacecraft telemetry and command link applications. Amateur bands are available for this purpose at no cost, but require frequency reservation through the International Amateur Radio Union (IARU). The process leading to this reservation is complex and needs to be carried out well in advance of the mission. It also requires the applicant to possess at least a standard Amateur Radio Call Sign. Information required for the application includes details such as output transmitter power, International Telecommunications Union (ITU) emission designation, antenna gain, radiation pattern, noise temperature, and signal power attenuation link budget. In addition mission parameter details such as mission duration, tele-command sequences, details of the positive transmitter command procedures, power supply budget, Keplerian elements and launch parameters are required. The assembly of this information is time consuming and requires intimate knowledge of the mission, and substantial liaison with the design team during the design process. The preparation of this document should be viewed as an integral part of the design phase of the mission, up to and including the Critical Design Review. I present some of these details regarding the University of Sydney i-INSPIRE mission and the QB50 spacecraft.

Keywords: frequency allocation, spacecraft, picosat, amateur radio, commercial radio

I. Introduction and Summary

Amateur radio operators are very familiar with the concept of self-regulation. Under normal circumstances this works very well for locally operated terrestrial repeaters and personal radio stations. Satellite applications are obviously different because they have a global effect due to the motion of the satellite around Earth. This global effect must be taken into account when assigning bands to the satellite. The International Amateur Radio Union (IARU) co-ordinates this process when the satellite operators choose to use amateur bands for this purpose. Amateur bands are particularly popular because they are, as the name suggests, assigned free of charge. The author has investigated the use of commercial bands for the i-INSPIRE project. Informal discussions with the Australian Communications and Media Authority (ACMA) revealed that the likely costs of commercial frequency assignment for a pico-sat mission of this type, substantially exceeded the cost of the launch. Figures of the order of \$32,000.00 were suggested in the course of these discussions. The form referred to here can be found at <http://www.iaru.org/uploads/1/3/0/7/13073366/iarucoordinationrequestversion28.pdf>. The frequency co-ordination application must be led by

<i>World Region</i>	<i>Member Continents</i>
Region 1	Europe, Africa, Middle East and Northern Asia
Region 2	The Americas
Region 3	Southern Asia and the Pacific

Table 1: IARU world regions

a licensed amateur radio operator who holds least an Australian Standard call sign. It is highly recommended that the person who carries out this process is experienced and knowledgeable in matters related to radio communication, and ideally holds an Australian Advanced call sign. The advanced level call sign permits the operator to use transmitter powers up to and including 400 Watts peak envelope power (PEP), while a standard operator is limited to a maximum of 100 Watts PEP. These figures apply for both the space station and the ground station. Standard and Advanced operators can use digital and other non-vocal transmission modes. The latter modes are very useful in communication with a satellite, as they permit the exchange of telemetry and experimental data as well as images. This paper proceeds while summarising the IARU forms and information in Section II, the more important and difficult sections are described in Section II-A, and illustrated using examples from i-INSPIRE. Section III contains some concluding remarks. The application process is very complex when it is being made on behalf of a higher degree granting institution. Approval by numerous entities within Universities are required. These begin with the departmental heads, then to faculty deans, University lawyers and finally the Vice Chancellor or proxy. The whole process can take in excess of 3 months to complete.

II. Radio Link Information Required by the IARU

Information required to complete the frequency co-ordination request is supplied in 11 sections. The document control is covered in section 0 and contains minor information about the documents used in the submission and the expected launch date. Section 1 covers information about the spacecraft, which will be published openly, including the mission names before and after launch, if different, and the country of origin of the license. This latter piece of information is particularly important since it allows the IARU to determine which of 3 World Groups the licensee belongs to. Separate frequency bands are applied to the different world groups II The IARU world regions are broadly the same as the International Telecommunications Union (ITU) world regions. Australia is in region 3.

For each of these regions the 70cm amateur band between 435MHz to 438MHz is allocated world wide for satellite use on the basis that its use causes no interference to other commercial and non-amateur operations. There are other bands, some as high as 400GHz which are available under limited circumstances. The amateur bands contain a mixture of primary and secondary sub bands. Bands referred to as Primary, are to all intents and purposes owned by the amateur radio system. The 70cm band is not primary, although it forms part of the amateur band plan. Details of the person applying for the allocation are contained in section 2. This consists of details about the identity of the licensee and that person's amateur radio qualifications and call sign. Section 3 of the application is concerned with the details of the organization that owns the satellite or is responsible for it's design, construction and operation.

In addition, details of the local Amateur Radio national authority are also dealt with, in this case the Australian authority is AMSAT-VK. This part of the document can also be thought of as a form of identification. Section 4 contains the more significant details of the proposed space station (or satellite). This is the portion for which close liaison with the design team is most important. In effect it requires the applicant to lay out a detailed design of the communications link with the spacecraft. In the case of the i-INSPIRE mission the author simply used the web site <http://sydney.edu.au/i-inspire-satellite/> to cover most of the questions asked in this part of the document. This includes most of the details of the mission known at the time of writing, and all of the papers published relating to the project [2], [3], [4], [5]. Within the description of the satellite, information required includes the overall mission plan, duration, proposed transmit and receive frequencies and ITU emission designators (see the relevant section below), a description of the spacecraft physical structure, a functional description of each section of the spacecraft and a power budget for each system in the spacecraft. Section 5 of the application requires the provision of operationally sensitive information which is not published. This includes the frequency plan and emission designators plus a description of any cyphers used in the communications link and direct space station transmitter controls. This latter point requires details of the space station control sequences and any turnoff or turn on commands. Also required in this section are details of any ground stations to be used for control of the spacecraft. Section 6 is concerned with telemetry details. These include frequencies, emission designators, link budget, telemetry equations and transmission format. Each of these will be touched upon later in this paper. Launch details and orbit parameters are covered in section 7. This includes the launch agency, the location of the launch, the Keplerian elements of the orbit and a manifest of other payloads on the proposed launch. Sections 8 and 9 cover the details of the Earth or Ground stations likely to be supporting the mission as for the space station but with the tele-command sequences covered earlier in section 5. Section 10 is a catchall for any other details.

A. RF Details

The radio link information for this or any mission consists of Radio Frequency (RF) link parameters including antenna gain and radiation pattern as well as signal power link budget. For the i-INSPIRE mission these details are quite straight forward. The mission is primarily a technology demonstrator and is not commercial. The link is simplex. The antenna is a simple centre fed dipole, which will have a gain of 1.5, (the typical gain for antennas of this type). The radiation pattern is also typical of centre fed dipole antennas: It is expected to be a typical “donut” pattern of the type seen in figure 1, modified only by the presence of the body of the spacecraft. The link budget is as follows, these figures are current at the time of publication, but may change when the orbital parameters are better known. The total attenuation leaves a link margin of $8.55dB$ above the detection threshold with a signal to noise power ratio of $20.55dB$ at the ground station. The uplink figure of merit is $-24.99dB$. The figures quoted here are averaged over one acquisition period.

B. ITU Emission Designators

Amateur radio was originally referred to as experimental radio. This was, of course, in the early days when amateurs actually were experimenting with the process. In the course of this experimentation they tried many different types of modulation and transmission and,

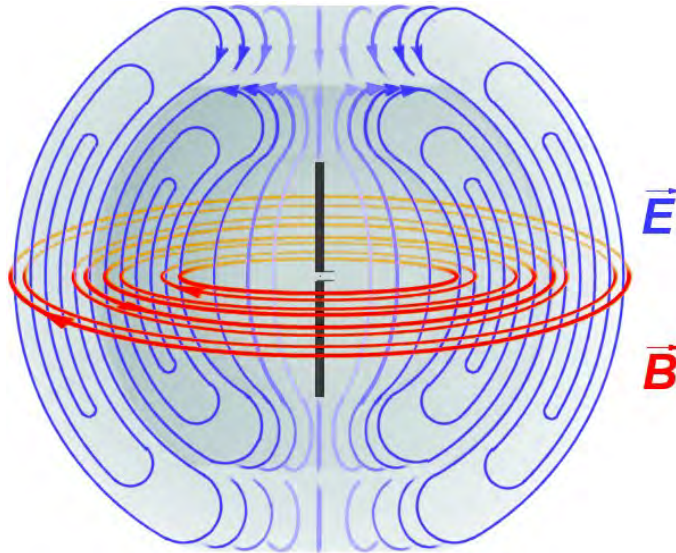


Fig. 1: Electric and magnetic field radiation patterns for a simple dipole antenna

accordingly, required a means of conveying what type of transmission they were using. A fairly complex system of symbols consisting of letters and numbers were developed. The details of this system as it applies to our application is below. The emission designator for i-INSPIRE is 8K00F2DBN indicating that we expect to need a bandwidth of 8 KHz. Bandwidth is expressed as H (hertz) K (kilohertz) M (megahertz) G (gigahertz) with the letter placed where the decimal point would be in the respective bandwidths; for example an 8 KHz bandwidth is expressed as 8K00. The next symbol indicates the type of modulation of the main carrier, in our case F for frequency modulation. The second character indicates the nature of the modulating signal; for i-INSPIRE the number 2 is used, indicating a single digital channel with carrier modulation. The third symbol is the letter D indicating data or telemetry information is to be transmitted. The fourth symbol indicates that we are using a two condition code of constant duration (simple binary), expressed using the letter B. The last symbol indicates the nature of any multiplexing, in this case the letter N is used to indicate no multiplexing is applied. At present we expect that the emission designator for the subsequent QB50 mission will be the same as this, but this may change depending on the final design. The interested reader is referred to the web site [1] for a more detailed explanation of this system of symbols.

C. Tele-command Equations

The IARU requires that all of the data likely to be transmitted by the space station to the ground station be expressed as a series of (usually simple) equations. These equations are not necessarily analytical in any sense, nor are they required to be. They are simply a set of polynomial equations empirically fitted to the likely functional form of the data to be exchanged. The limits within which these data fits are valid are also required. A typical example of this type of these is equation 1 where N is the number transmitted, V is a battery voltage, and the range of validity is between $6.4V$ and $16.4V$.

$$V = 0.1N + 6.4(\text{volts}) \quad (1)$$

<i>Subsystem</i>	<i>Power (mW)</i>	<i>Duty cycle (hrs/day)</i>	<i>mWh/day</i>
Power	558	19.4	10825
Onboard data handling	150	24	3600
Transmitting	110 to 600	12	1320
Receiving	27	12	324
Payloads	55	24	1320
Total Energy Available (mWh): 10825			
Total Used (mWh): 6564			
Energy Remaining (mWh): 4261			

Table 2: i-INSPIRE power budget

D. Power Budget

The likely power budget for this mission depends on the solar panel output, the period of the orbit, the time and proportion of the orbit in eclipse and the load required by the payloads and telemetry systems. The size of the battery is also a consideration. The power budget for i-INSPIRE is contained in table 2 corrected from reference [4].

The largest single drain on the power used by the satellite is due to the onboard computer for data handling. The instantaneous load applied to the transmitter is greater but has a much lower duty cycle. In reality we expect that the spacecraft will be within range of the likely ground stations for periods not exceeding seven minutes, and not more than four times per day.

E. Noise Temperature

The noise temperature of the transmitter is also one of the required parameters. The Noise Temperature is defined as the temperature which would produce the expected Johnson-Nyquist noise level for the transmitter circuit. It is in effect, the emission temperature which could produce the emitted power assuming a Rayleigh Jeans power law approximation. The transmitted power P in watts and the bandwidth in hertz B are related to the temperature T in kelvins in equation 2:

$$P/B = k_B T \quad (2)$$

where k_B is Boltzman's constant ($= 1.38 \times 10^{-23} J/K$). Making T the subject of the equation reveals

$$T = (1/k_B)(P/B). \quad (3)$$

In the case of i-INSPIRE this results in a total noise temperature of $5.4 \times 10^{18} K$.

F. Tele-Command Sequences

These details are not published for security reasons, but are required by the IARU in case there is a need for the IARU, or other national or international authorities to command the spacecraft transmitter off. While this is not yet known to have happened, a spacecraft transmitter would need to be turned off if it could be shown to be the likely cause of substantial interference in the same or in different frequency bands.

G. Orbit Details

The launch service for i-INSPIRE is to be provided by Inter Orbital Systems (IOS) from a site about 100Km to 200Km off the coast of California [5]. This mission is expected to be flown in a low altitude (310Km), high inclination (98°), essentially polar orbit. At this altitude the expected mission lifetime is heavily dependant of solar activity, which at the time of writing is fairly high. High levels of solar activity cause high levels of ionisation in the ionosphere and also the exosphere. This tends to expand the atmosphere to higher altitudes which results in higher levels of drag for Low Earth Orbit (LEO) spacecraft. The expected lifetime of this mission is 24 days, but is subject to considerable uncertainty. Small differences in altitude have a large impact on lifetime due to varying drag in addition to the solar/terrestrial atmospheric influences mentioned above. The full set of Keplerian elements are not known at this stage for this mission. The final parameters will be supplied, either by the launch service provider or by the US NORAD, after orbital insertion.

III. Concluding Remarks for i-INSPIRE and QB50

The frequency allocation process for the University of Sydney i-INSPIRE satellite is currently underway, with the formal application under consideration by the IARU. The system will consist of a simplex 70cm amateur link in the frequency range 435MHz to 438MHz. Completion of this allocation is expected by the end of 2013 or early 2014. The University of Sydney QB50 satellite, probably to be called INSPIRE-2, will be launched with the rest of the QB50 constellation in 2015 (<http://www.qb50.eu>). This will be a double unit cube sat with a 2m uplink system and a 70cm downlink. The design is not complete yet so these systems are not fixed. The frequency allocation will be handled as part of the QB50 system by the Von Karman Institute (VKI) in Belgium.

Acknowledgments

The author would like to thank Iver Cairns, Jiro Funamoto, Chris Betters, Xiaofeng Wu and Adrian Xiao for assistance in the preparation of this paper.

References

- [1] <http://life.itu.ch/radioclub/rr/ap01.htm>

- [2] Fogarty, L., Cairns, I.H., Bland-Hawthorn, J., Wu, X., Betters, C., Funamoto, J., Leon-Saval, S.G., Monger, A.G. and Xiao, S., "The initial - INtegrated SPectrograph, Imager and Radiation Explorer (i-INSPIRE) - a university satellite project., *Proceeding of the 11th Australian Space Science Conference*, 2011
- [3] Betters, C., Cairns, I.H., Bland-Hawthorn, J., Wu, X., Fogarty, L., Funamoto, J., Leon-Saval, S.G., Monger, A.G. and Xiao, S., "Instrumentation of the i-INSPIRE satellite., *Proceeding of the 11th Australian Space Science Conference*, 2011
- [4] Xiao, S., Wu, X., Cairns, I.H., Bland-Hawthorn, J., Betters, C., Funamoto, J., Leon-Saval, S.G., Fogarty, L., Monger, A.G. and Bai, X., "i-INSPIRE Tube-Satellite Bus Design, *Proceeding of the 11th Australian Space Science Conference*, 2011
- [5] Funamoto, J., Wu, X., Cairns, I.H., Bland-Hawthorn, J., Betters, C., Fogarty, L., Leon-Saval, S.G., Monger, A., Xiao, S., "Engineering i-INSPIRE - a Pico-Satellite from Australia", *Proceeding of the 11th Australian Space Science Conference*, 2011

Test Results of the Namuru Dual-GNSS Space-borne Receiver

Mazher Choudhury, Joon Wayn Cheong, Nagaraj C Shivaramaiah, Chris Rizos and Andrew G Dempster

Australian Centre for Space Engineering Research (ACSER), School of Surveying and Geospatial Engineering, The University of New South Wales, Sydney, Australia.

Summary: The Garada Project is funded under the Australian Space Research Project (ASRP). Garada project – a Synthetic Aperture Radar (SAR) Satellite Formation Flying constellation – requires precise relative position, timing and attitude knowledge for its satellites to allow the SAR operation to be executed cooperatively and synchronously. This paper describes the navigation performance of Namuru V3.3, a dual GNSS (GPS and Galileo) receiver designed specifically for in-orbit operations. Testing for low earth orbit operation was performed using a Spirent multi-GNSS simulator (GSS8000) that was used to generate scenarios using both GPS and Galileo constellations. Data from the Namuru V3.3 receiver was processed using Matlab and compared against Spirent-derived truth data. In this paper, the test results and a detailed analysis of the quality of pseudorange- and carrier phase-derived navigation and position solution in orbit scenarios is presented. The NMEA output stream from the receiver was processed using Matlab to investigate the navigation solution's performance. The inter-satellite (baseline) solution was processed using RTKLibPro, an improved version of RTKLIB developed by UNSW.

Keywords: Namuru V3.3, dual-constellation GNSS receiver, navigation performance.

Introduction

The Namuru family of FPGA-based GPS receivers is one of the key research outputs of the Australian Centre for Space Engineering Research (ACSER) at the University of New South Wales. Different versions of Namuru receivers provide different capabilities in the GNSS environment, with the Namuru V3.2 GPS L1 receiver having been developed specifically for spaceborne applications such as Project Biarri [1]. On the other hand, the Namuru V3.3 used for the Garada project is a dual-constellation GNSS (i.e. GPS L1 and Galileo E1) receiver that inherits all the characteristics of Namuru V3.2. The key difference is not only the firmware that runs into the receiver but also the hardware.

Although the navigation performance of the Namuru V3.2 GPS receiver, as well as earlier version of the Namuru (i.e. V2.4), was analysed by [2] and [3], Namuru V3.3 GNSS receiver's performance has not had not to this date been investigated. The Namuru V2.4 dual-constellation GNSS receiver was analysed by [4]. In this paper the most recent dual-constellation GNSS (i.e. GPS and Galileo) Namuru receiver's navigation performance was studied. As this V3.3 receiver is a new development from the ACSER team, the navigation performance needed to be analysed as it was specifically designed for in-orbit operations.

Test setup

To test the performance of Namuru V3.3, GPS (L1) and Galileo (E1) signals were generated using a GSS8800 Spirent GPS simulator as shown in Figure 1(a). The scenario in Table 1 generates signals that match closely with the signals that would be received by a spacecraft in low earth orbit. The RF output from the Spirent was fed to the Namuru V3.3 receiver's antenna port, while the receiver RS232 port was connected to a desktop computer. A Windows-based terminal program (called Termite) was then used to log serial data to a hard disk for later analysis. Figure 1(b) is a block diagram of this test setup.

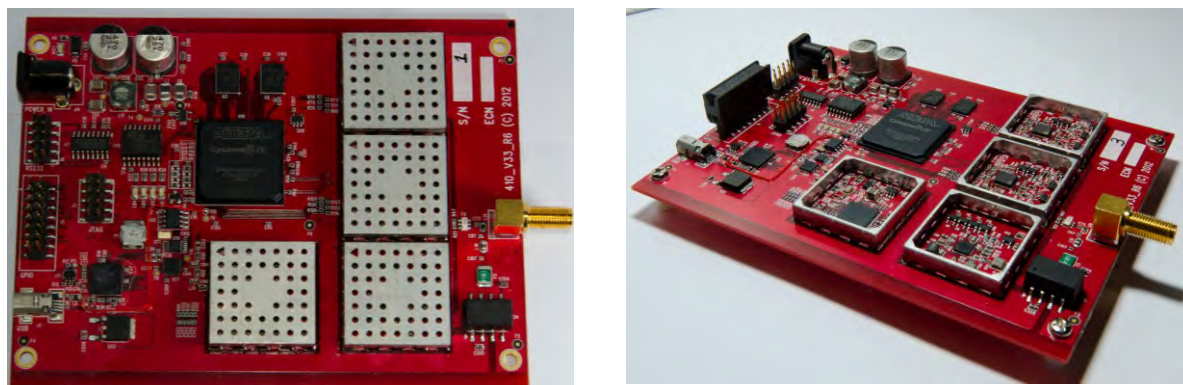


Fig. 1(a): Namuru V3.3

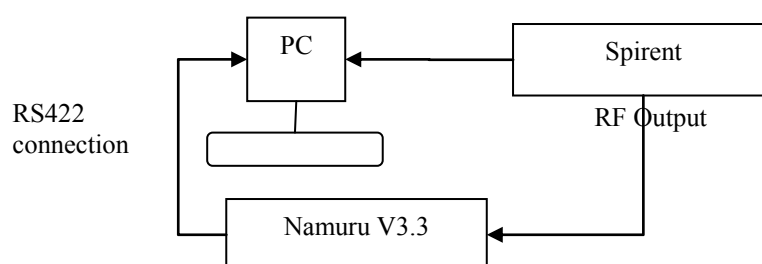


Fig. 1(b): Block diagram of the test setup

Table 1: Scenario parameters

Epoch start: 17:00:00, 25 th August, 2011	
Orbit parameter:	Value
Semi-major axis (a)	7058140m
Eccentricity(e)	0
Inclination (I)	98.0443°
Right ascension (Ω)	-90.046°
Argument of perigee (ω)	0
Mean anomaly (M0)	0

Three different application scenarios were developed for the navigation performance analysis. A very short baseline (i.e. 1m) application scenario was used to check the applicability of

Namuru V3.3 in multi-GNSS attitude determination, which requires a precise relative positioning solution. This test only focuses on the navigation and relative position solutions. However, altitude determination using multi-GNSS is out of this paper's scope. Interested readers are referred to [5] for further details of the multi-GNSS altitude determination algorithm and the performance of Namuru V3.3 for such an application.

Next, a short baseline (500m) application scenario was used for testing receiver applicability in a bi-static SAR mission. This mission consists of a constellation of formation flying satellites, with 3 or 4 satellites in each formation [5]. A short baseline (i.e. 500 to 1000m) between two satellites was investigated.

Finally, a zero-baseline test was proposed to study the precision of the receiver measurements. Availability of several Namuru V3.3 receivers for testing was difficult at the time of this study due to limited availability of the hardware prototype. As a result a single Namuru V3.3 was used to define a "virtual" zero-baseline. This virtual zero-baseline was formed using two runs of the same test scenario with the simulator. After collecting two sets of raw measurements, navigation and baseline position solutions were generated for analysis. This virtual zero-baseline analysis methodology is adopted from [3]. Table 2 summarises the three application scenarios.

Table 2: Application scenarios

Case	Description	Note
1	Very short baseline (1m)	By changing antenna position to 1m in 'X' direction in WGS84 reference frame.
2	500m baseline	By changing antenna position to 500m in 'X' direction in WGS84 reference frame.
3	Zero-baseline	Using virtual zero-baseline

Navigation Solution

The Namuru V3.3 produces National Marine Electronics Association (NMEA) message output along with other proprietary messages. Proprietary messages are used to obtain a full-fledged RINEX message format which was later used in post-processing. Limitation of the current Namuru V3.3 is that it cannot decode the Galileo ephemeris or almanac data. The navigation solution was obtained using RTKLIBPro which is a modified and upgraded version of an open source software package for GNSS carrier phase and pseudorange positioning [6]. Details of the algorithms implemented for GNSS positioning can be found in [6,7].

To validate the quality of these solutions, the navigation solution output from the three application scenarios (mentioned in the previous section) was analysed. All tests were conducted using an in-orbit scenario assuming no ionospheric or ephemeris errors. Each test provided two sets of data. One from the base antenna (hereafter referred to as the 'base') and the other from the rover antenna (hereafter referred to as the 'rover'). This rover antenna is in a different position from the base antenna, except for the zero-baseline application scenario. All the accuracy results presented here are in $x \pm y$ format where x is mean and y is standard deviation (at 1σ) of the data set. All the navigation results are presented in the WGS84 XYZ coordinate system, which were compared with the truth data used for the Spirent simulator.

First the navigation solution from “Case1: very short baseline” is presented. This test was designed to analyse Namuru V3.3’s applicability for GNSS attitude determination [5]. From this test it was observed that the 3D accuracy of the pseudorange-based navigation solution from the base and the rover was $6\pm5\text{m}$, and $5\pm5.7\text{m}$, respectively. One point to note is both cases did not finish at exactly same point. During the experiment of “rover case” simulator stopped right before where it suppose to finish. However, this did not hamper the comparison analysis as only the common epochs are considered for comparison. Figure 2(a) and 2(b) show the navigation solutions obtained from the receiver for both base and rover. Figure 2(c) and 2(d) show the difference from the true orbit for both cases. Although the result from the navigation solution has lower accuracy and higher standard deviation than the previous Namuru version (i.e. Namuru V3.2), which was $0.5\pm4\text{m}$ [3], it satisfies the mission design requirement. As the noise level (i.e. standard deviation) was more than the baseline length, data from both the base and the rover needs to be processed using carrier phase observations. From Figure 3 it can be seen that the 1m baseline cannot be identified using only the pseudorange-based navigation solution. The accuracy of the velocity was $0.1\pm3\text{m/s}$, $0.1\pm3\text{m/s}$ and $0.01\pm3\text{m/s}$ for the X, Y and Z components, respectively. Figure 2(e-f) shows the velocity solution with respect to the true values.

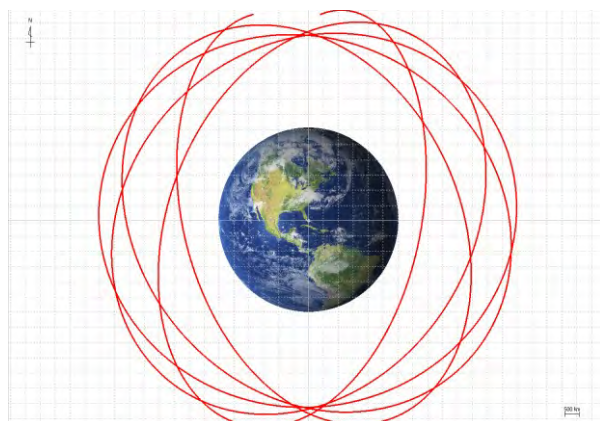


Fig. 2(a):Navigation solution (Base)

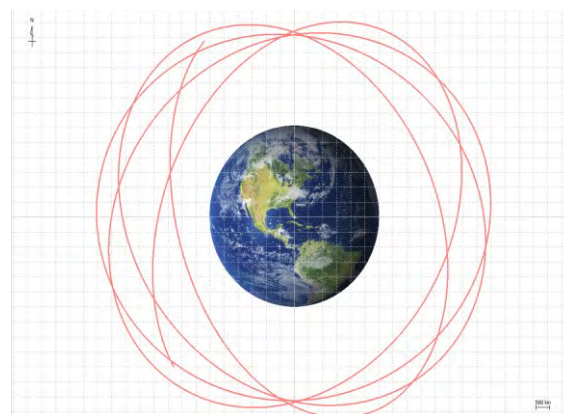


Fig. 2(b):Navigation solution (Rover)

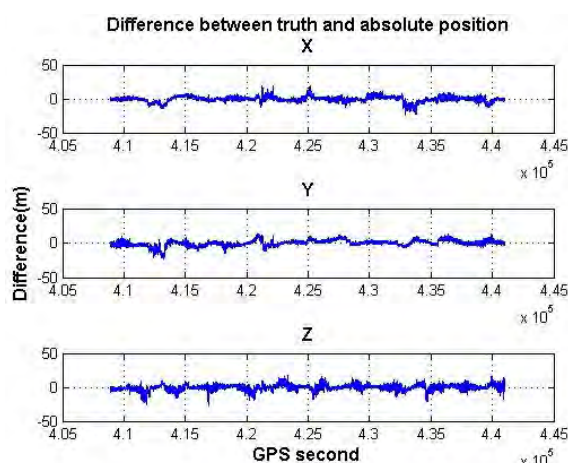


Fig. 2(c):Coordinate difference(Base)

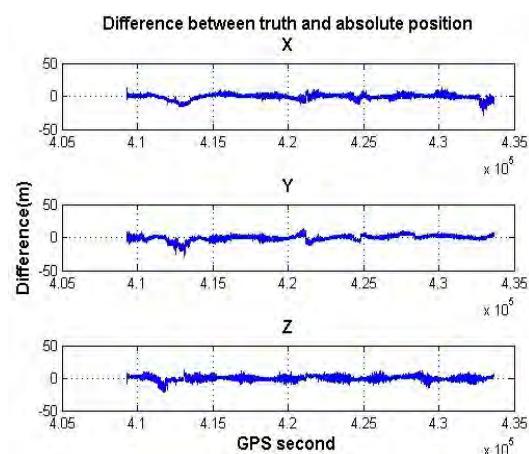


Fig. 2(d):Coordinate difference(Rover)

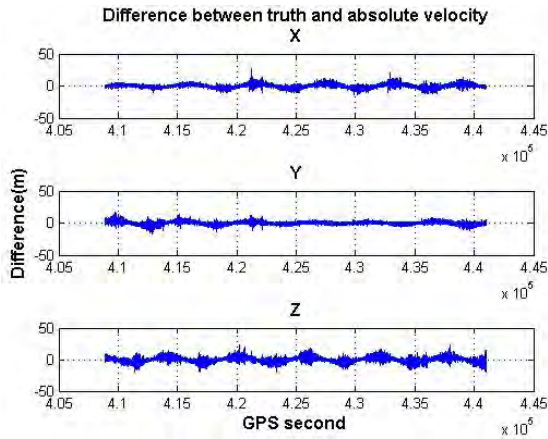


Fig. 2(e): Velocity difference(Base)

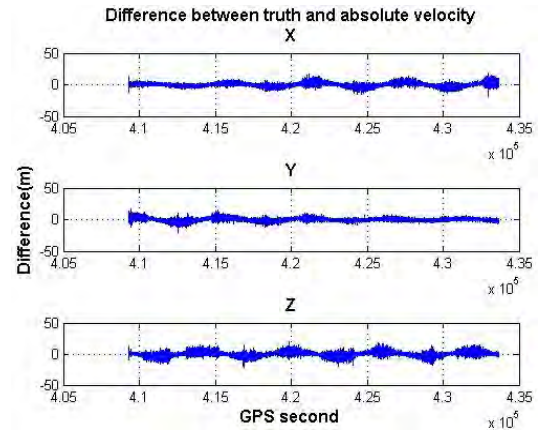


Fig. 2(c): Velocity difference(Rover)

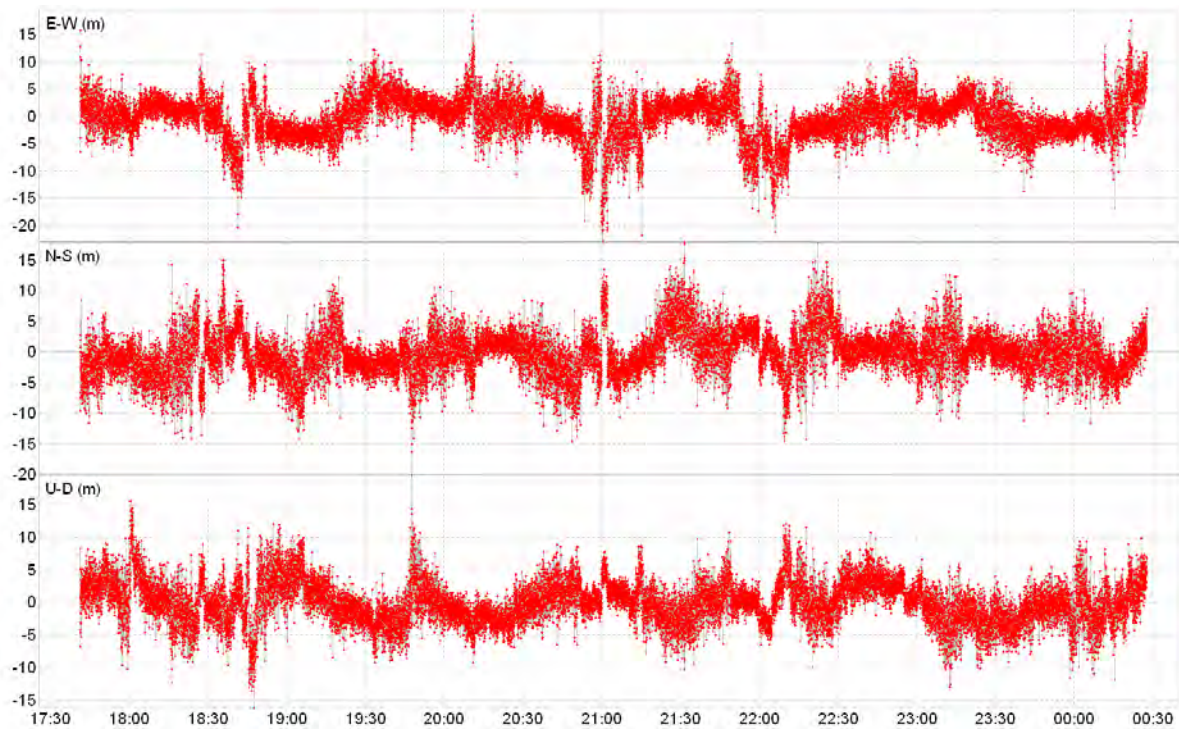


Fig.3: Baseline-pseudorange solution

Next, the navigation solution from “Case2: short baseline” is presented. The Namuru V3.3 is designed for SAR formation flying [5], As a result, operation across a range of baseline lengths is necessary. (In this paper, a short baseline (i.e. 500m) was analysed, whereas long baseline (i.e. >1km) tests will be conducted after Namuru V3.3’s hardware and firmware are upgraded.) It was observed that the 3D accuracy of the pseudorange-based navigation solution from the base and the rover was $6\pm 5\text{m}$ and $6\pm 6\text{m}$, respectively. Figure 4(a-d) shows the navigation solutions obtained from the receiver for base and rover, as well as the difference from the truth for both cases. On the other hand, from Figure 5 it can be seen that the 500m baseline can be identified using navigation solution. However, the 3D accuracy of baseline was $-1\pm 5\text{m}$, hence carrier phase observations should be used in the solution. Velocity accuracy was similar to that of Case1. It was observed that both base and rover cases were $0.1\pm 4\text{m/s}$, $0.2\pm 4\text{m/s}$ and $0.1\pm 4\text{m/s}$ for the X, Y and Z components respectively. Figure 4(e-f) shows the velocity solution with respect to the true values.

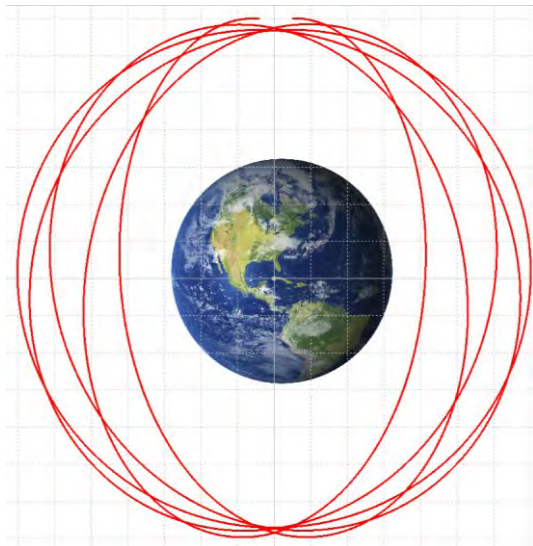


Fig. 4(a): Navigation solution (Base)

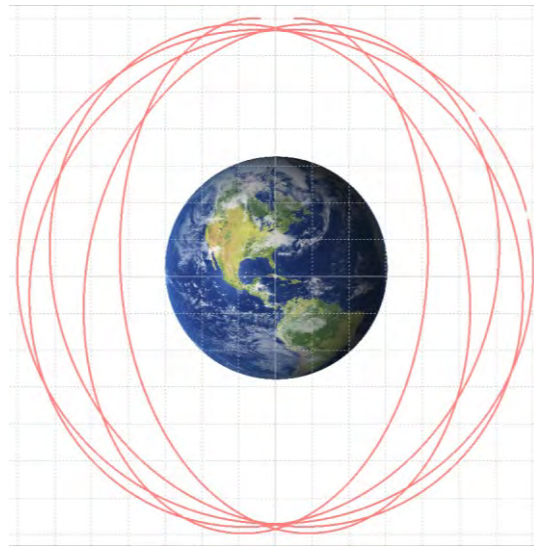


Fig. 4(b): Navigation solution (Rover)

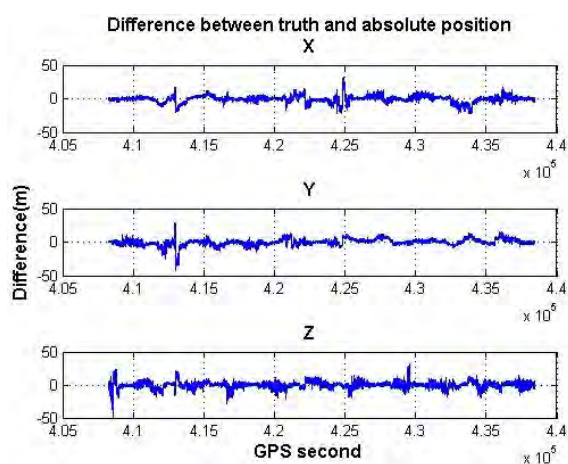


Fig. 4(c): Coordinate difference (Base)

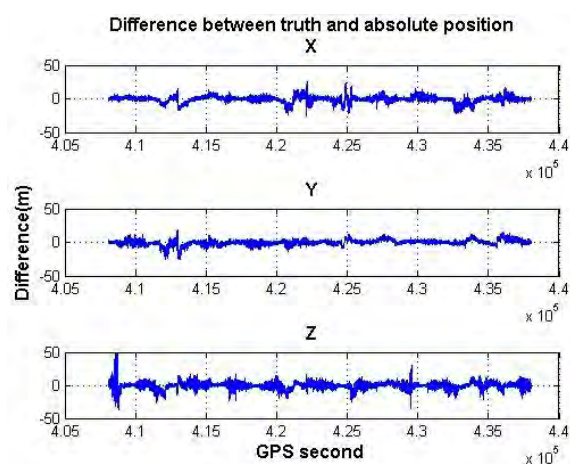


Fig. 4(d): Coordinate difference (Rover)

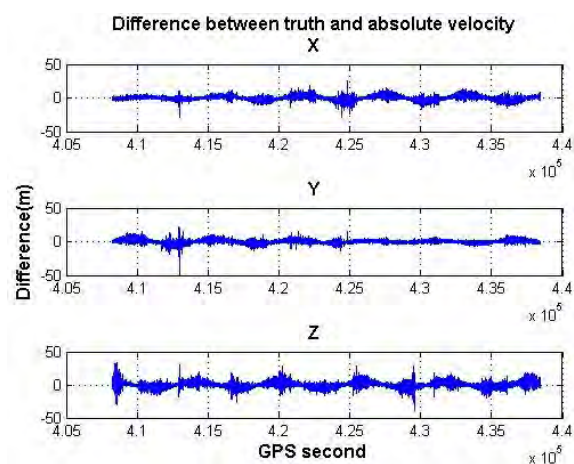


Fig. 4(e): Velocity difference (Base)

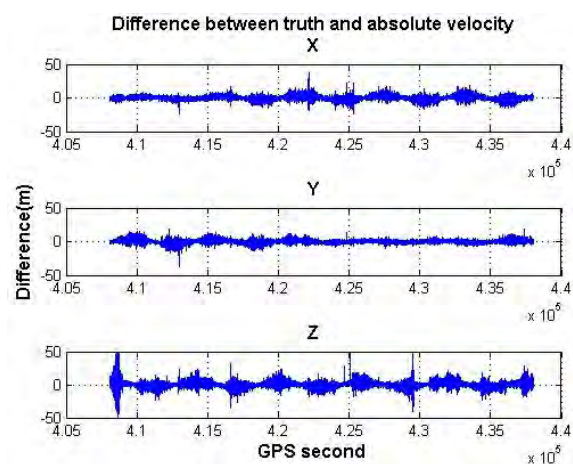


Fig. 4(c): Velocity difference (Base)

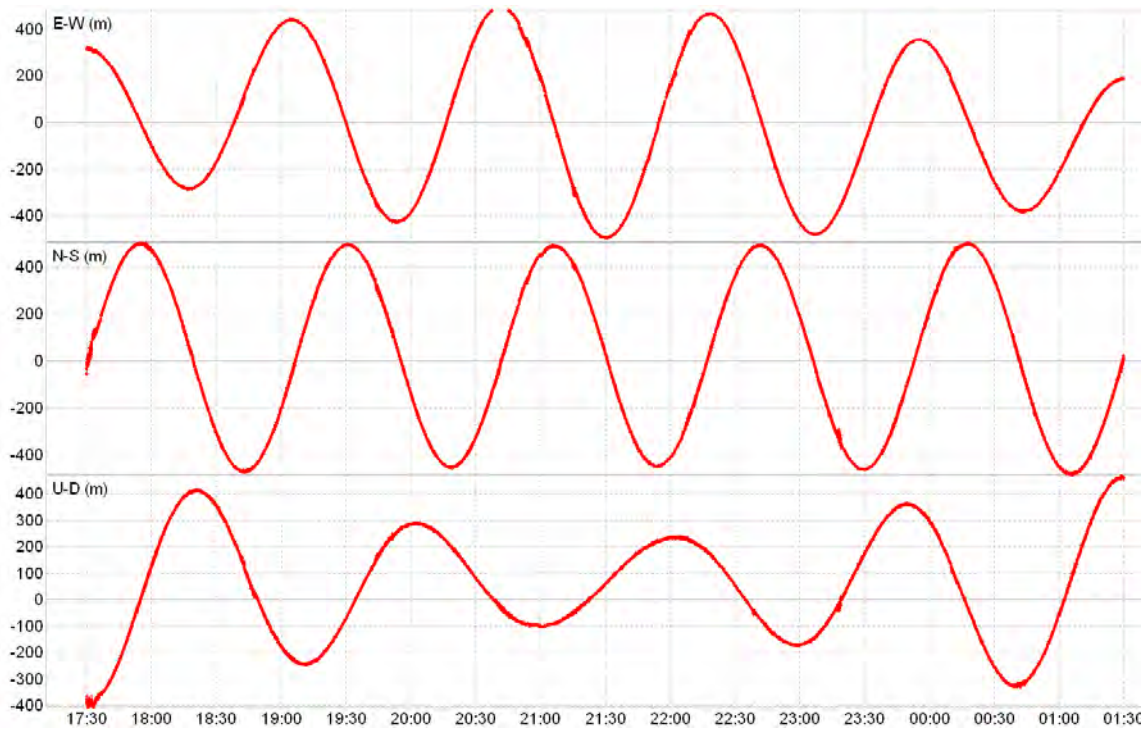


Fig.5: Baseline pseudorange-based navigation solution

Finally, the results of navigation solution from “Case3: zero-baseline” are presented. This test was designed to study the system noise level in the baseline position solution. It was observed that the 3D accuracy of the navigation solution from both the base and the rover was $7\pm 5\text{m}$. As this is a zero length baseline, both base and rover antenna are positioned in the same physical place. The pseudorange-based navigation solution’s accuracy and precision was low hence carrier phase observations should be used. The zero-baseline navigation solution had an accuracy of $-0.06\pm 2\text{m}$, $-0.06\pm 2\text{m}$ and $-0.05\pm 2\text{m}$ for the East, North and Up components (with respect to base), respectively. Figure 6 shows the zero-baseline solution.



Fig.6: Zero-baseline navigation solution

Baseline solution

Baseline solutions were also generated using RTKLIBPro with assumed standard deviations of the carrier phase and pseudorange observables are 0.003m and 0.030m, respectively. “Fix and hold” ambiguity resolution method was used for ambiguity resolution. Both carrier phase and pseudorange observables were used to generate accurate relative position solutions. The baseline solution from “Case1: very short baseline” now has high accuracy, being $0.01 \pm 0.06\text{m}$ for ambiguity-float solutions and $0.007 \pm 0.008\text{m}$ when the ambiguities were fixed to their integer values. For this test the integer ambiguities were resolved for 99.2% of the time. One drawback was the required time to obtain an ambiguity-fixed solution – on average 10 minutes. After analysing the residuals of the solutions it was noticed that there was a high level of noise present at the beginning, which did not satisfy the statistical validation threshold (i.e. minimum “ratio” for acceptance of ambiguity-fixed solution was set to 0.33 [8]). After improvements identified for the hardware are available this result is expected to be of even higher precision and accuracy. Figure 7(a) shows the ambiguity-fixed baseline solution, while Figure 7(b) shows the difference between the truth and the baseline solution. Figure 7(c) shows the scatter plot of the error, where it can be observed that results are indeed of high precision.

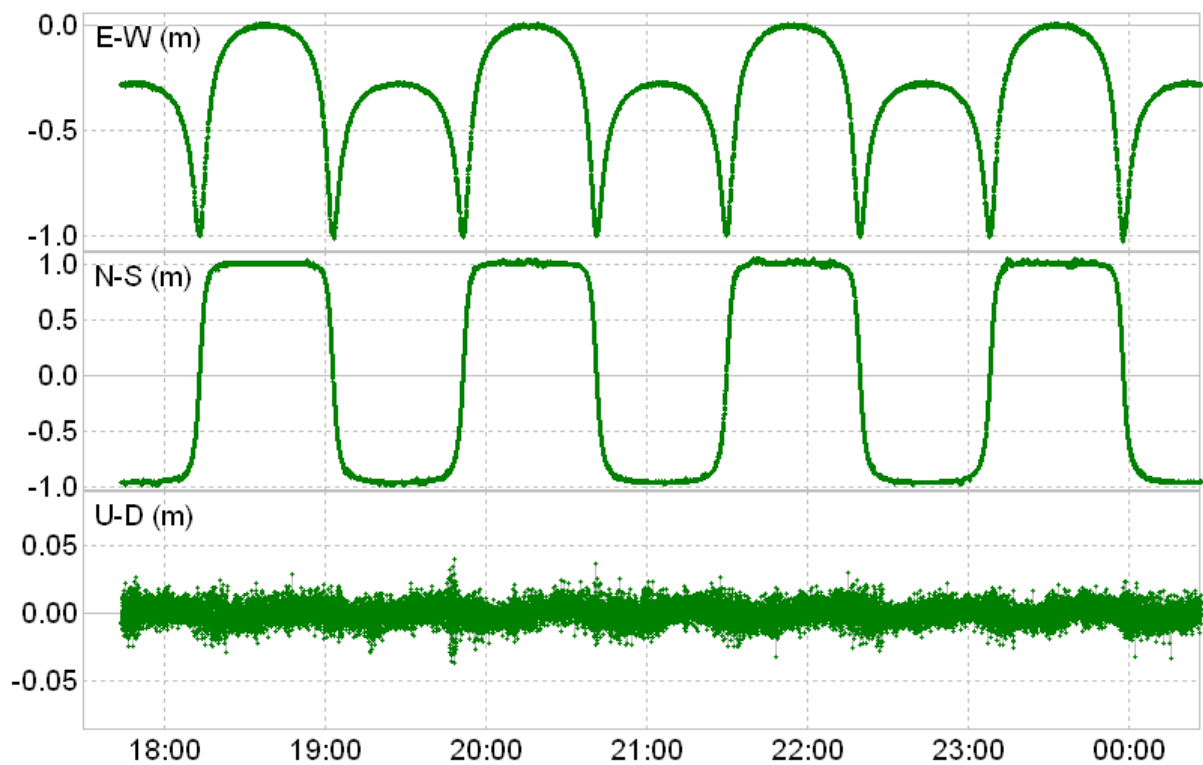


Fig.7(a): Very short baseline solution with ambiguity fixed

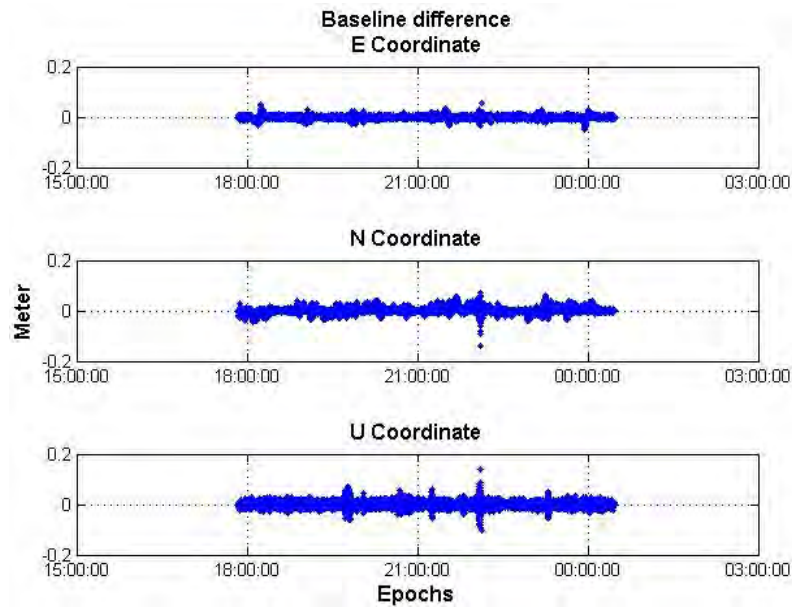


Fig.7(b): Very short Baseline difference from truth

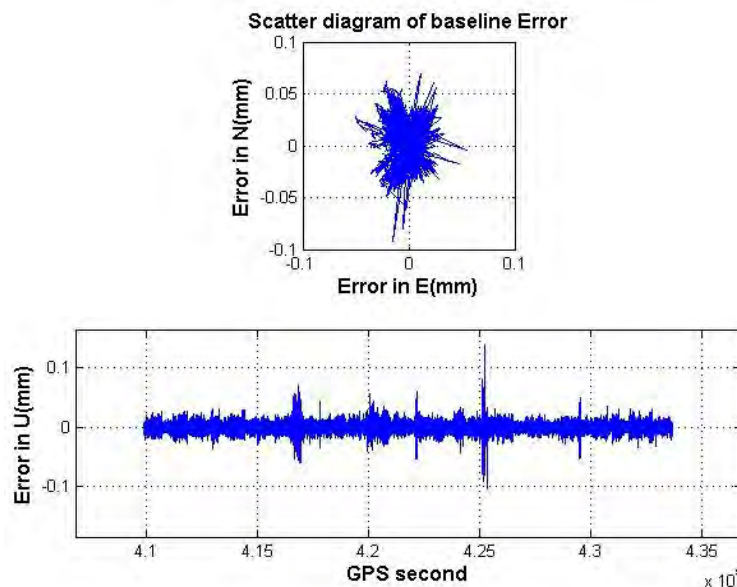


Fig.7(c): Very short baseline error scatter plot

Next, the baseline solution from “Case2: short baseline” was analysed. It was observed that 3D accuracy of $0.03 \pm 0.16\text{m}$ was obtained for ambiguity-float solutions and $0.006 \pm 0.004\text{m}$ was obtained when for ambiguity-fixed solutions. However from time period 21:58:17 to 22:10:51 only an ambiguity-float solution was possible (*yellow dots in Figure 8(a)*). This sudden “ambiguity loss” issue is currently under investigation by the authors. The Namuru V3.3 receiver design requirement was $0.0 \pm 0.10\text{m}$, which was not achieved for ambiguity-float solutions. On the other hand, the Namuru V3.2 did meet the mission design accuracy requirement of $0.01 \pm 0.10\text{m}$ [3]. Figure 8(a-c) shows the baseline solution, the difference between the truth and the baseline solution, and the scatter plot of the test, respectively.

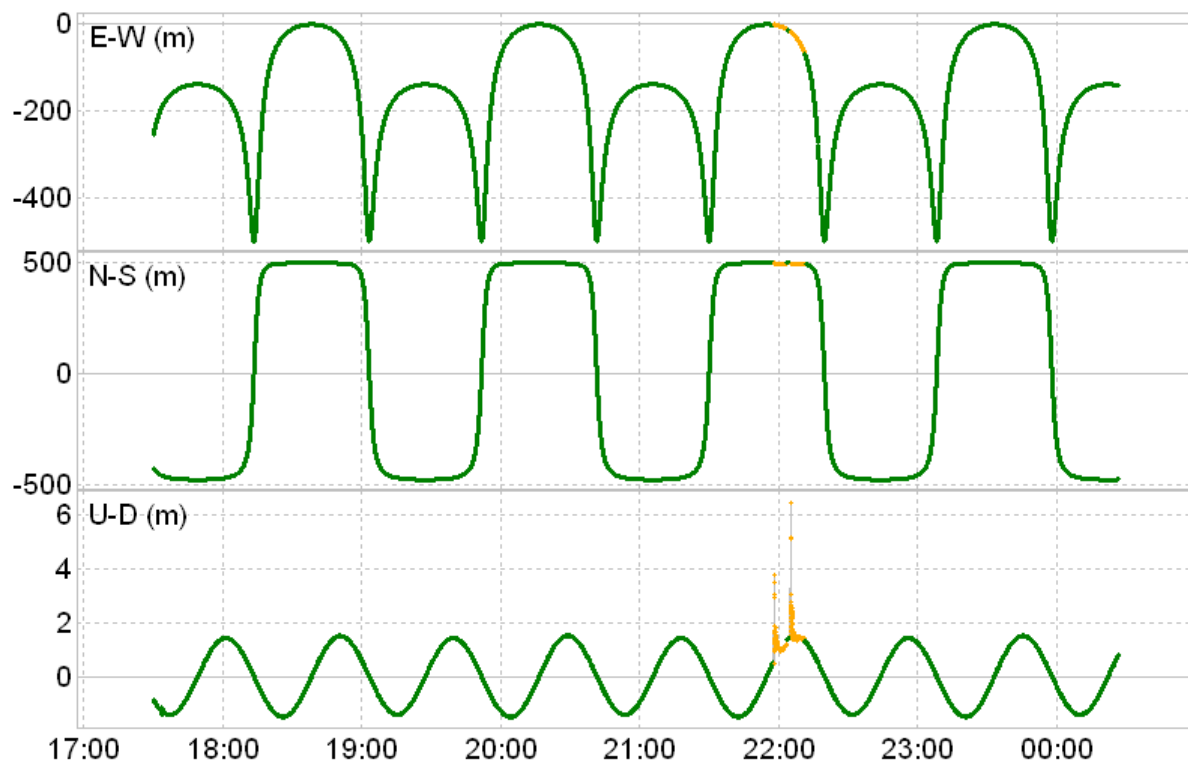


Fig.8(a): Short baseline solution
Note: yellow dots represents float ambiguity

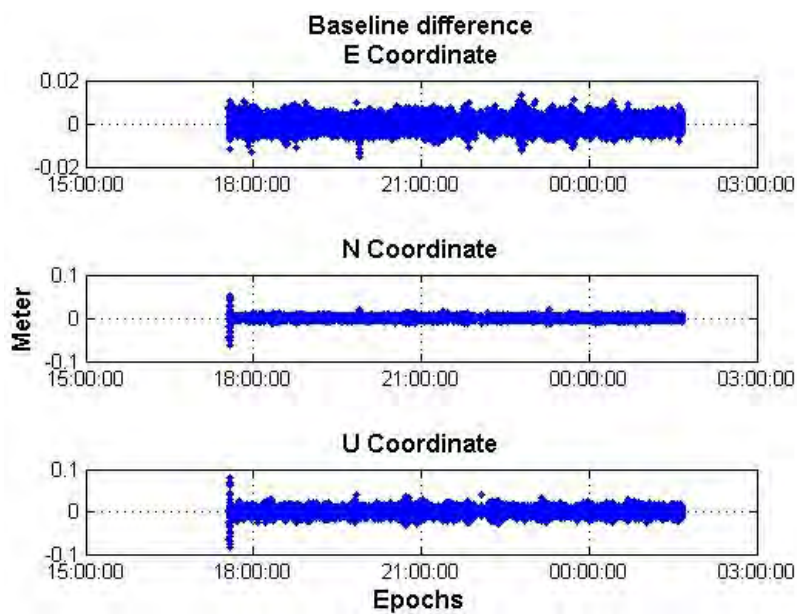


Fig.8(b): Baseline difference

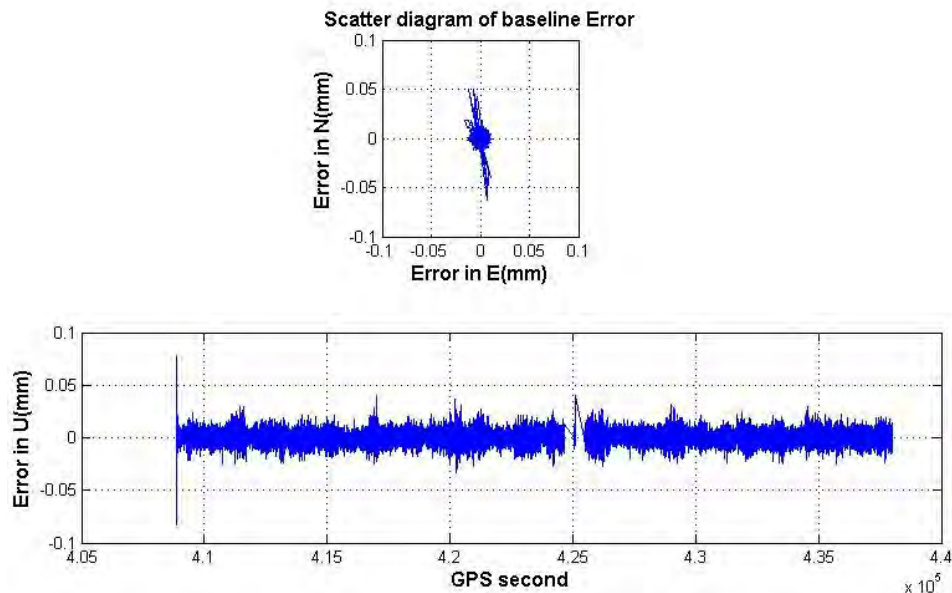


Fig.8(c): Short baseline error scatter plot

Finally, baseline solution from “Case3: zero baseline” was analysed. In this test the zero-baseline solution had a 3D accuracy of $0.002 \pm 0.002\text{m}$, with ambiguity-fixed solutions obtained 100% of the time. Residuals of the GPS L1 and Galileo E1 pseudoranges were $0.60 \pm 1.2\text{m}$ and $0.031 \pm 1.2\text{m}$, respectively, while the carrier phase residuals were $0.0 \pm 0.002\text{m}$ and $0.0 \pm 0.002\text{m}$, respectively. These results confirm that random observation error and the propagation of receiver biases (if any present) cancelled during zero-baseline carrier phase measurement processing. On the other hand, pseudorange measurements from both GNSSs indicate the presence of a certain level of bias. It is suspected that these biases might be generated by the receiver clock because this test involved a single receiver and a virtual zero-baseline setup. Figure 9(a) shows the zero-baseline solution. Figure 9(b) shows the scatter diagram of the error, and Figure 9(c-d) show the GPS L1 and Galileo E1 residuals, respectively.

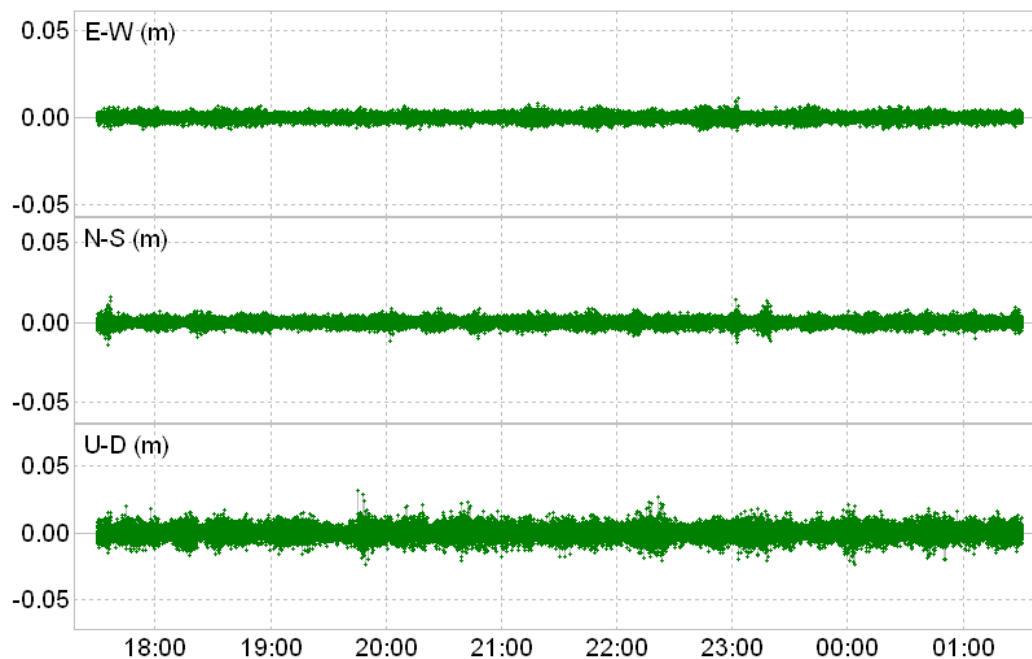


Fig.9(a):Zero-baseline solution with ambiguity fixed.

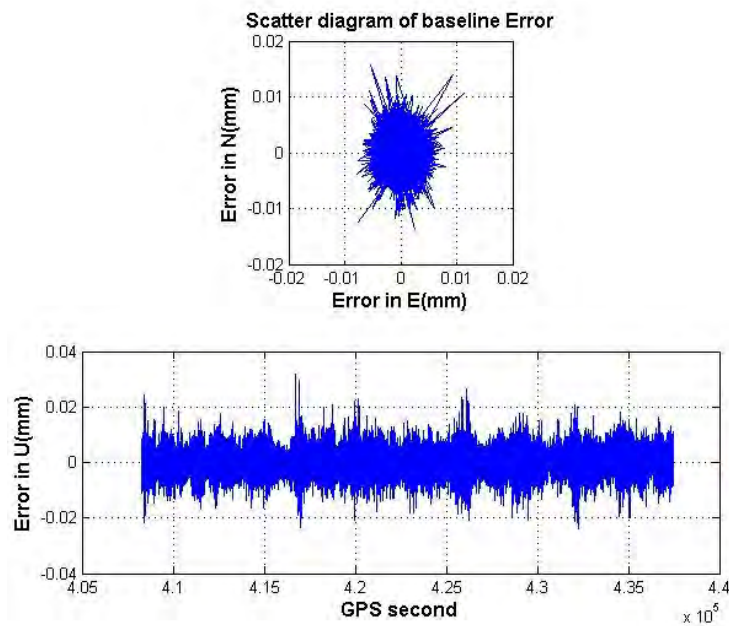


Fig.9(b): Baseline error scatter plot

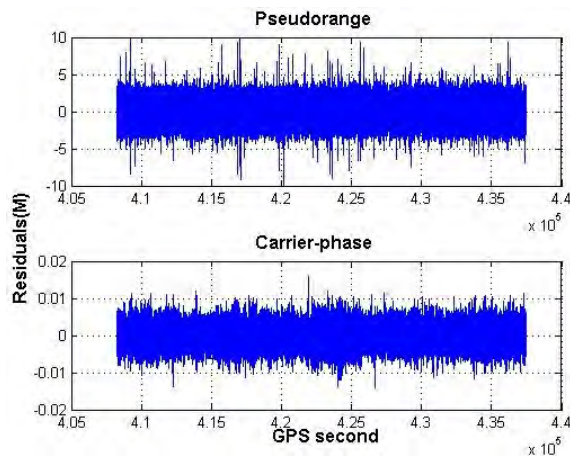


Fig.9(c): GPS L1 residuals

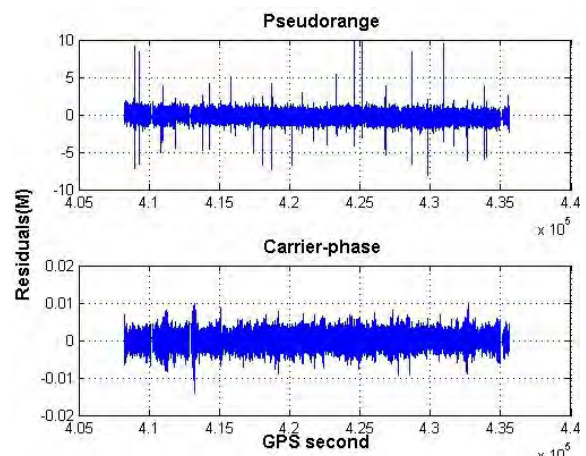


Fig.9(d): Galileo E1 residuals

Concluding remarks

In this paper the Namuru V3.3 dual GNSS spaceborne receiver was tested in orbital scenarios. Navigation and velocity solutions were assessed using two different orbital scenarios. Furthermore, the baseline solution was investigated, in particular for the validity and applicability of carrier phase measurements for in-orbit applications where the receiver experiences high dynamics. Going forward, after the improved hardware is available; the zero-baseline solutions need to be re-analysed. Longer baselines (i.e. 1 to 100 kilometres) also need to be analysed to determine whether the Namuru receiver can satisfy the SAR mission requirements.

Acknowledgment

The work was supported by the Australian Research Council and the Australian Space Research Program.

References

1. Newsam, G., and Gordon, N., "An Update on SSA in Australia", *12th Annual Advanced Maui Optical and Space Technologies (AMOS) Conference*, Maui, Hawaii, USA, September 13-15, 2011, <http://www.amostech.com/TechnicalPapers/2011/Poster/GORDON.pdf>.
2. Grillenberger, A., Rivas, R., Markgraf, M., Mumford, P., Parkinson, K., and Rizos, C., "The Namuru receiver as development platform for spaceborne GNSS applications", *4th ESA Workshop on Satellite Navigation User Equipment Technologies*, Noordwijk, The Netherlands, December 10-12, 2008, Session 11, Paper 2.
3. Choudhury, M., Glennon, E.P., Dempster, A.G. & Mumford, P., "Characterization of the Namuru V3.2 Spaceborne GPS Receiver", *Proceedings of the 12th Australian Space Science Conference*, Melbourne, Australia, September 24-27, 2012, 285-294.
4. Choudhury, M., Wu, J., Cheong, J.W., and Dempster, A.G., "Initial test results of Namuru multi GNSS spaceborn receiver", *International GNSS Society Symp.*, July 16-18, 2013, Gold Coast, Australia.
5. Australian Center for Space Science (ACSER), "Final Report: A National Soil Moisture Monitoring Capability - Volume I: Technical Results", June, 2013, Sydney, Australia.
6. Takasu, T., and Yasuda, A., "Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB", *International Symposium on GPS/GNSS*, Jeju, south Korea, November 4-6, 2009.
7. Li, Y., Glennon, E.P., Li, R., Jiao, Y., and Dempster, A.G., "Development of a spaceborne GPS receiver for precise relative navigation of formation flying small satellites", *Lecture Notes in Electrical Engineering*, Vol 161, 2012, 467-476.
8. Teunissen, P.J.G., "GNSS integer ambiguity validation: Overview of theory and methods", *Proceedings of the ION 2013 Pacific PNT Meeting*, Honolulu, Hawaii, USA, April 22-25, 2013, 673-684.

CubeSat-based GPS Reflectometry Using the Namuru V3.2 GPS Receiver

Éamonn P. Glennon¹, Kevin Parkinson⁴, Joon-Wayn Cheong³, Andrew G. Dempster^{2,3},
Chris Rizos¹, Scott O'Brien², and Kegen Yu¹

¹ *School of Civil and Environmental Engineering (CVEN)*

² *School of Electrical Engineering and Telecommunications (EET)*

³ *Australian Centre for Space Engineering Research (ACSER)*
University of New South Wales, Sydney, NSW 2052, Australia

⁴ *General Dynamics Pty Ltd, New Zealand*

Summary: The Namuru V3.2 GPS receiver is a single frequency, FPGA based receiver that has been designed specifically for navigation, timing and high precision relative navigation in low earth orbit. With its intended use on the UNSW 'EC0' CubeSat to be deployed as part of the QB50 constellation, there exists an opportunity to use the receiver to perform radio-occultation and GPS reflectometry experiments.

In this paper, we discuss the modifications necessary to allow the receiver to be used to perform these experiments. Particular emphasis is placed on the GPS reflectometry experiment, which has software and hardware requirements that differ substantially from the standard requirements of navigating GPS receivers. Design considerations for the FPGA correlation logic will be discussed in detail, as well as our solution to meet those design constraints.

Keywords: QB50, space-borne GPS receiver, CubeSat, Namuru V3.2, GPS reflectometry, GNSS-R

Introduction

The 2010 launch of the Australian Centre for Space Engineering Research (ACSER) at the University of New South Wales (UNSW) has heralded a significant change in the perception of space-related projects within the university. Since then, ACSER's Australian Space Research Program (ASRP) funded Garada project, 'Synthetic Aperture Radar (SAR) Formation Flying', has been successfully completed and the GPS receivers for the Defence Science and Technology Organisation (DSTO) Biarri project [1] are nearing final delivery.

One of ACSER's newest projects is participation in the *von Karman Institute of Fluid Dynamics* (VKI) QB50 program. This program, funded through the European Union Framework Programme 7 (FP7) research grants, has VKI funding the launch of up to 50 2U and 3U cubesats into low earth orbit in order to gain an unprecedented temporal and spatial picture of the lower thermosphere. Each participant is required to fund its own QB50 compliant CubeSat design [2], with each CubeSat required to contain one of three sensor packages, as selected by the program participant. Spare payload capacity is available for the

participant's own experiments, thereby allowing the participant to fly their own satellite payloads without having to absorb any launch costs.

UNSW has four different experiments planned for its Educational Cubesat 0 (EC0) QB50 entry. The first of these is the RApid Manufacture of Space Exposed Structures, abbreviated to RAMSES, in which the satellite superstructure will be constructed using rapid prototyping 3D printing techniques using PAEK HP3 thermoplastic material. The material will be thoroughly characterized for its use in space before launch, and will be subject to additional monitoring throughout the flight. The second 'Rapid recovery from SEUs in Reconfigurable Hardware' experiment, abbreviated to RUSH, will investigate techniques by which Field Programmable Gate Arrays (FPGAs) may automatically recover from single event upsets (SEUs), as occurs in the harsh environment of space. The third experiment, seL4 Reliable Optimized Critical Systems or seL4 ROCS, is the software equivalent of RUSH, except that in this case the proven microkernel operating system developed by National ICT Australia (NICTA) will be flown and operated on board the spacecraft. The fourth experiment, and the topic of this paper, is to fly the UNSW developed Namuru V3.2 GPS receiver, which will be used primarily for providing navigation solutions for the spacecraft. However, unlike standard space-based GPS receivers, the Namuru V3.2 receiver will also incorporate modifications that will allow it to be used to perform GPS-based remote sensing, with experiments relating to GPS radio-occultation (GPS-RO) of the ionosphere [3] and GPS reflectometry of the earth's surface being planned.

GPS reflectometry (GPS-R) is a form of remote sensing that detects and processes forward scattered GPS reflections from the earth's surface [4], with the directly received signal sometimes also used in addition to the reflected signals. Characteristics relating to the reflecting surface may then be inferred, including sea surface wind speed and direction, the detection of sea ice, altimetry, land categorization and soil-moisture.

Reflected GPS signals differ from directly received GPS signals in several respects. The first difference is signal polarization, which is mostly left-hand circularly polarized (LHCP) rather than right-hand circularly polarized (RHCP), this being due to the significant polarization reversal associated with each reflection event. A second difference is the power levels of the signals, with reflected signals experiencing some level of attenuation during the reflection process. The third difference is distortion of the detected signal, which in this case is due to the received signal actually being the superposition of a multitude of reflections that occur close to the specular reflection point. This causes the measured GPS correlation triangle to have a trailing edge that extends well beyond the one chip correlation code delay when reflected by a rough surface. One form of GPS-R measurement involves analysis of the trailing edge of the correlation curve in the time-domain, thereby allowing the sea surface roughness to be estimated. A second GPS-R measurement is similar to the first, except that the measurements are performed at a wide range of Doppler frequencies, leading to the construction of a delay-Doppler map (DDM).

In this paper, we first introduce the Namuru V3.2 GPS receiver, followed by an examination of some of the requirements imposed by the GPS-R experiment on the system. We then discuss our solution to the problem, taking into account the resources available within our hardware, with the focus of this paper being the FPGA modifications required to support this application.

Namuru V3.2 GPS Receiver

The Namuru V3.2 GPS receiver is a single frequency, FPGA-based GPS receiver that has been designed specifically for operation onboard a Boeing Colony 2 CubeSat platform [1, 5, 6]. The receiver is manufactured by General Dynamics Pty Ltd, a privately owned NZ based company with which the Surveying and Geospatial Engineering (SAGE) group at UNSW has had a long-standing relationship. UNSW owns the FPGA correlator design that allows the GPS signals to be down-converted, de-spread and converted to a state suitable for being controlled by the GPS firmware. UNSW also holds licenses and full source code for the Aquarius GPS firmware [7] that allows, but is not limited to, use of the firmware for its own research purposes. This powerful combination of intellectual property provides an unparalleled capability for highly skilled engineers to modify the receiver for a variety of unusual applications that are not available using commercial off the shelf (COTS) equipment. It also allows development of receiver designs without being hampered by US International Traffic in Arms Regulations (ITAR), which are particularly problematic for space-based GPS receiver development due to the high speed and altitude requirements for such receivers.

The Namuru V3.2 GPS receiver is one of a family of Namuru-based designs, with this particular variant employing Flash based MicroSemi FPGA's that were selected due to the improved resistance afforded by flash memory-based designs against single event upsets (SEUs) and single event latch-ups (SELs). The main GPS correlator is located within a MicroSemi ProASIC3E1500 device, while a MicroSemi SmartFusion A2F500 incorporates an ARM Cortex M3 processor that runs the Aquarius GPS firmware and Biarri boot-loader at a speed of 80 MHz. The SmartFusion also includes 512 kB of onboard Flash that is used to store the firmware, 64 kB of high speed SRAM and FPGA fabric approximately 1/3 the size of the ProASIC device. It was originally hoped that the GPS correlator could be incorporated into the SmartFusion FPGA fabric, but unfortunately the fabric was too small and an additional FPGA ended up being required. In a similar vein, the internal 64 kB SRAM was also too small to support the software requirements and an additional external SRAM was also included in the design. As it turns out, the unused SRAM and the unused SmartFusion FPGA fabric have been used in our GPS-R solution.

GPS-R Requirements and System Constraints

A receiver designed for GPS-R differs substantially from a standard GPS receiver designed for position, velocity and timing (PVT). First, the signals to be processed are LHCP signals that typically originate from the nadir, compared to the RHCP signals that typically originate from directly satellites located anywhere from the horizon to the zenith. Those LHCP signals will also be significantly attenuated. This means that not only are non-standard, non-COTS antennas required, but any antenna system is required to incorporate additional gain in order to counteract some of the attenuation inherent in the raw signals. As a consequence, if GPS-R is to be carried out in addition to standard navigation, two different antenna systems are required; a standard-zenith pointing RHCP antenna and a nadir pointing high-gain LHCP antenna. Additionally, each of these antennae is required to be connected either to its own individual RF front end, or alternatively, may be switched to a single RF front-end provided time multiplexing of the signals is acceptable.

Following reception, down-conversion and quantization of the received signals, the next step involves the processing of those signals. Again, this differs completely from standard PVT processing, which requires that each signal be acquired then tracked (usually with a phase locked loop), ephemeris, clock correction and timing data extracted from those signals and pseudorange, pseudorange-rate and carrier phase measurements generated at a rate of at least 1 Hz, after which the navigation solution is calculated. In contrast, the GPS-R processing need

not be performed in real time and ideally are processed with a flexible software receiver, perhaps using hardware acceleration, in order to maximize the versatility of the system and allow for future reprocessing of the data when new techniques become available. This is especially the case for systems that are in the research phase. Furthermore, the type of processing also differs substantially from standard receivers, with the main observables being either the delay waveforms or delay-Doppler maps (DDMs) for each specular reflection.

Transmitting any captured IF data to ground for post-processing is the ideal solution, although the data-rates and bandwidths of the GPS signals may exceed the storage and communications capacity of the spacecraft. For example, the GP2015 RF front-end used by the Namuru V3.2 samples the GPS L1 signal at a rate of 5.714 MHz at 2-bit (sign & magnitude) resolution, leading to a data-rate of 11.428 Mbps. With this IF sample rate, even 1 second of data consumes 1.428 MB, and although high capacity flash storage may be externally available, there may not be sufficient bandwidth to that system to permit saving large periods of data in real-time.

A space-borne GPS-R observation from a CubeSat has additional complications, including limitations on the downlink communications capacity from the spacecraft. This is definitely the case for the EC0 spacecraft, which had a 19,200 bps UHF downlink capacity when this project was first proposed, but may also be equipped with a 100 000 bps (limited to 25,000 bps) S-band downlink capacity. At 19,200 bps, the minimum transmission time of 595 seconds for 11.428 Mb was barely unacceptable, but even with the upgraded capability, transmitting 1 second of captured IF signals from a GP2015 would take at least 457 seconds or longer, depending on the transmission messaging format and overhead. In both cases, on-board processing of the signals may be beneficial even if the flexibility of the data is reduced. This means that the period between successive GPS-R observations is over 450 seconds even in the best scenario.

In addition to communications constraints, a 2U cubesat such as EC0 is subject to significant size and weight restrictions. These become especially apparent when considering the LHCP antenna array, which is a typical solution to the problem of increasing the gain of the nadir-looking antenna footprint. Not only is the area available for the array limited to 20 cm x 10 cm, but some of that area is reserved for the CubeSense earth horizon sensor camera and an S-band antenna for communications downlink. Furthermore, the dimensions available for the antenna array are near the limit for receipt of the GPS L1 signal wavelength of 19 cm. The space available for the payload electronics is also limited, with a consequence that the single GPS receiver board is required to fulfill all of the associated requirements (PVT, GPS-R and GPS-RO), despite the fact that they differ in many respects.

EC0 Namuru GPS Reflectometry Solution

The primary mission for the EC0 Namuru GPS receiver is to provide PVT functionality for the spacecraft, thereby allowing us to gain flight heritage for our receiver design. Both GPS remote sensing experiments are considered secondary experiments that add scientific and engineering value to the mission, but should not increase the risks associated with the mission.

To ensure this outcome, we consider the minimum requirement for including the EC0 Namuru receiver to be one in which the receiver can be interfaced to the on-board-computer (OBC) for power and communications, and a zenith pointing L1 RHCP active patch antenna for navigation. This will be achieved through the construction of a daughter board that will form the electrical and mechanical interface between the Boeing Colony 2 CubeSat interface that the Namuru V3.2 is currently designed for, and the CalPoly CubeSat design that the EC0

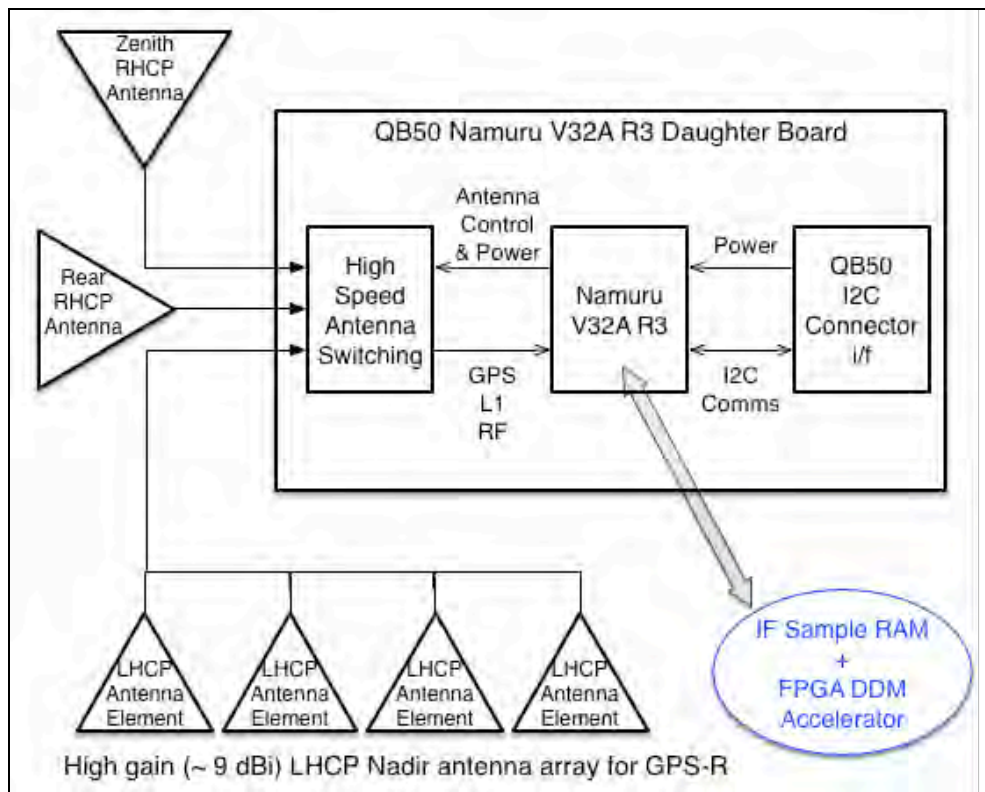


Fig. 1: EC0 Namuru V3.2 Payload Configuration

is required to meet. The advantage of this approach is that it allows the stock Biarri-designed receiver to be used without major modification, save for minor firmware modifications to provide the I2C functionality that is the cornerstone of the CalPoly interfacing system.

However, use of a daughterboard also allows the inclusion of additional functionality required to support the GPS remote sensing. This will be done by including an RF switching arrangement whereby solid-state high-speed RF switching circuitry will be incorporated into the design. This will allow one of up to four active antennas to be connected to the Namuru GPS receiver RF front-end. The default configuration will be for the RF switch to connect to a zenith-facing antenna, thereby providing maximum visibility to the GPS signals during normal PVT operation. A second RF connection will connect to a nadir pointing LHCP antenna array allowing observation of the reflected GPS signals. A third RF connection will allow a rear facing RHCP antenna to be included for the purposes of performing GPS-RO, while the fourth will be reserved for possible use on a nadir looking RHCP antenna. Control of the RF switches will be via the main GPS correlator embedded within the ProASIC3E1500 FPGA, thereby allowing rapid switching between antennas should the need arise. This might turn out to be useful in allowing the receiver to operate if the spacecraft is not roll stabilized by virtue of being able to switch to an upward angled antenna at any time. It will also allow the GPS-R observations to be made with only a momentary interruption of normal navigation. Because this momentary interruption occurs within 1 second of an actual PV solution, any prediction errors associated with a predicted PVT solution at the time instant of the GPS-R observation will also be negligible. The payload configuration is illustrated in Fig. 1. Both the antenna array and the daughterboard are currently in the design phase and will not be subject to further discussion.

Two different approaches are available for the processing of the data. The first is provision of a capability allowing real time streaming and capture of raw IF samples as output from the GP2015 RF front end to the Aquarius firmware. This has been done by including within the

main GPS correlator a 32-bit First-In-First-Out (FIFO) buffer together with a mechanism for converting a set of 16 two-bit (sign/magnitude) samples into 16 sample batches. When the IF capture process is initiated, buffering of these batches commences at the next measurement instant, which may be selected by the firmware to be the measurement instant coincident with a 1 second GPS boundary. The buffering process then continues to load samples into the buffer up until the buffer overflows, which should not occur if the firmware is emptying the buffer as required, or the required number of batches has been downloaded, or the internal memory available for the batches has been filled. Features within the firmware allow the captured IF samples to be output to the user via proprietary NMEA sentences, after which they may be post-processed. Verification & validation of this process was carried out by writing a MATLAB conversion routine to convert the data into a form suitable for post-processing using the Kai-Borre MATLAB software suite [8]. This ability to extract IF samples from the receiver will also be beneficial during assembly, integration and test (AIT) because it will allow any potential GPS interference to be more easily identified. During AIT, other payloads can be operated while capturing GPS IF data. The captured IF data can then be post-processed to determine whether any in-band GPS interference is being generated by any other payload.

One limitation of the IF capture process is the amount of memory available for storage of the IF data within the receiver. As previously discussed, the Namuru V3.2 receiver includes a Cypress Semiconductor CY7C1051DV33 8 Mb static RAM (configured as $512k \times 16$ -bit) that supplements the 64 kB of internal SRAM inside the SmartFusion. However, as much of this external memory is not required, 4 Mb has been allocated for the storage of IF samples. This equates to approximately 0.36 seconds of IF data. This could be increased by populating the board with a higher capacity SRAM with the same package footprint.

The second approach for processing the data is inclusion of an on-board processing capability within the receiver, although this feature is only beneficial in a severely constrained downlink scenario. Several methods are available for performing the on-board processing, namely a full software receiver, a full hardware receiver or some combination of the two. For our application, a full software receiver is not feasible because the main processor does not have the required processing capability. However, the Namuru V3.2 receiver is also severely constrained in the availability of FPGA resources, with the capacity of the ProASIC3E1500 running at close to 80%. This is down from near 99% utilization following a recent optimization exercise required in order to add the IF data capture feature. Furthermore, making additional changes to the GPS correlator FPGA is also undesirable due to the risk of introducing bugs into what is a stable design. Our approach to solving this dilemma was to make use of the FPGA fabric that was previously unused within the SmartFusion device itself, thereby allowing inclusion of a new feature without impacting the previously proven designs required for standard navigation. This may also permit GPS-R processing to be performed in parallel with standard navigation, except for the short time-interval when the IF data is captured.

The design and features of our delay-Doppler map accelerator (DDMA) are described in the following section.

SmartFusion Fabric Delay-Doppler Map Accelerator

DDMA Requirements

The SmartFusion A2F500 FPGA fabric is equipped with 11,520 VersaTiles (D-flipflops) and 24 memory blocks, with each memory block having a 4,608-bit capacity. Although this was not sufficient to allow the instantiation of a 12-channel Namuru GPS correlator, it is

Table 1: DDMA Configuration Options

Chip Spacing Chip Resolution (chips)	Dump Interval (ms)	Taps	Dumps Per ms	Chip Range (chips)	Bandwidth (kHz)	"1ms Look" Compute Time (ms)
1/8	1/8	80	8	10	8	10
1/8	1/8	256	8	32	8	32
1/4	1/8	256	8	64	8	32
1/2	1/8	256	8	128	8	32
1/2	1/4	512	4	256	4	64
1/2	1/2	1024	2	512	2	128
1/2	1	2048	1	1024	1	256

sufficiently large for a single channel with 8 code-phase taps, as well as additional logic to control the use of this correlator. The availability of the memory blocks also provide a means by which IF samples can be loaded into the device for processing, and a place to store the outputs from the accelerator.

To determine the requirements for the DDMA, the published DDM outputs from the GPS-R experiments performed by the UK-DMC mission were examined [4, 9]. From these results, it was decided that the DDMA should be capable of supporting 1/8 chip resolution in code space, as well as a frequency bandwidth of 8 kHz. This frequency bandwidth is achieved by allowing the correlator to support 1, 2, 4 or 8 in-phase and quadrature dump samples every millisecond, giving 1, 2, 4 or 8 kHz bandwidths, respectively. A maximum coherent integration period of 1 ms was selected because this is the maximum size that may be used for space-borne GPS-R observations (see Section 4.4, Figure 4.6 of [4]).

Several other operational requirements were deemed appropriate for this hardware. First, any data loaded into the device is required to be fully utilized in order to ensure the processor is not required to repeatedly reload the same data into the device. Second, the DDMA should be capable of processing the IF samples in 1 ms batches; this being the maximum coherent integration interval needed for a space-borne platform. This requirement defines the minimum size of each IF ping pong buffer, which in this case is required to be at least 11,428 bits in length or 357 16-sample batches. We provide two 384-length buffers in our implementation, so that one buffer can be loaded while the other buffer is being used. Creating a $384 \times 2 \times 32$ -bit two-port memory consumes 8 of the 24 memory blocks available within the device.

All of the remaining memory blocks within the FPGA have been allocated to the memory used to store the correlations calculated by the correlator, which in this case is a 4096×16 bit two-port memory. Table 1 describes some of the possible configurations, assuming 16-bit integrate-and-dump samples. The "1 ms look" compute time does not include the processor time required to perform an FFT on each set of dump samples occurring at the same code phase tap, which has the effect of creating a set of 1 ms dump samples separated by 1 kHz. This would allow construction of a 10 chip, 1/8 chip resolution, 8 kHz wide DDM with 200 non-coherent rounds (or 200 one-millisecond looks, to use radar terminology) to be calculated using 2000 ms of DDMA time. This is similar to the DDM shown in Section 4.6, Figure 4-10 of [4].

The flexibility of this hardware also permits its use in accelerating the cold start detection of satellites. Such a scenario applies when the receiver is first powered up and the receiver does not have GPS almanac, an estimate of time from its real-time clock or a set of two-line-element (TLE) orbit parameters that describe the orbit of the receiver. In such cases, the time

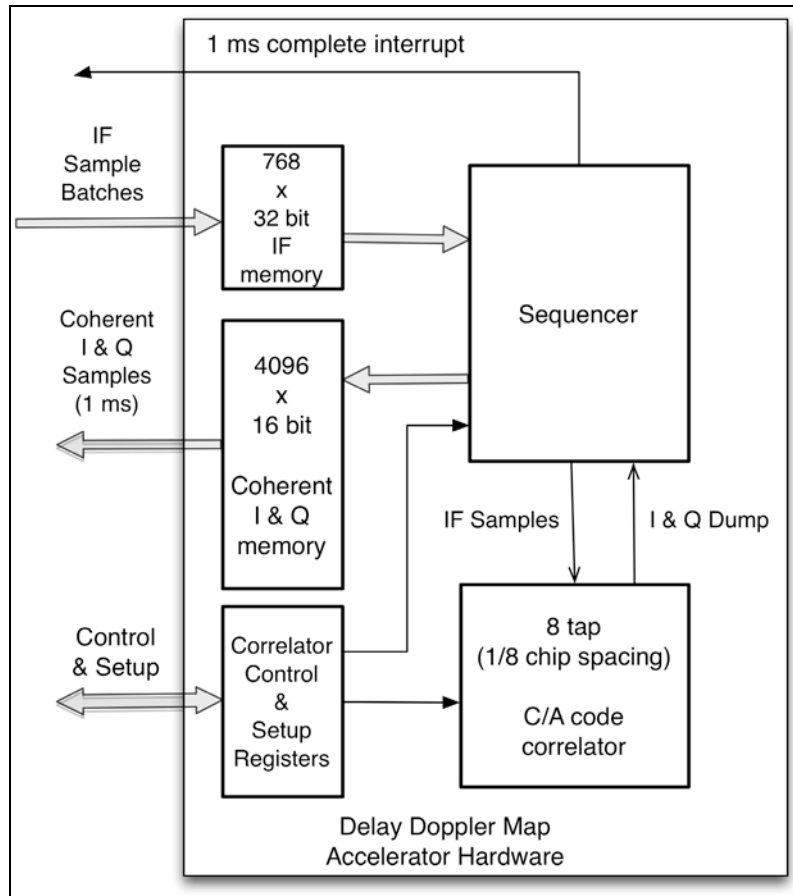


Fig. 2: DDMA Hardware Architecture

to acquire the GPS satellites received by the zenith facing antenna can take up to 900 seconds if using the standard 12-channel correlator present in the receiver. However, this can be substantially reduced if the DDMA hardware is also used for search acceleration. For this application, a wide chip spacing of 1/2 a chip and a wide bandwidth of 8 kHz is applicable, allowing 128 chips and 8 kHz of bandwidth to be examined every 32 ms, with at least 2 non-coherent rounds of integration required to improve the output signal to noise ratio. Searching for a single satellite across ± 40 kHz of bandwidth would then take at least $5 \times 10 \times 0.032 \times 2$ or 5.12 s (excluding processor FFT and integration time), which equates to 163.84 s for the entire constellation. The ± 40 kHz of bandwidth is required because the Doppler frequencies measured by a GPS receiver in low earth orbit and travelling at approximately 7.5 km/s are much larger than the typical ± 4 kHz Doppler ranges that are applicable for a terrestrial receiver [5].

DDMA Architecture and Implementation

The DDMA design uses the Namuru V3.2 FPGA GPS correlator as the starting point, although this has been evolved in order to make it suitable for use in this application. These changes include making the correlator run on an input 'sample by sample' basis, rather than being driven by the 40 MHz clock that would otherwise be employed, the ability to make GPS measurements has been removed and the state of the correlator can be initialised, saved and restored, in addition to being allowed to run when performing correlations. Because the correlator runs at the input sample rate of 40/7 MHz (period of 175 ns) rather than 40 MHz, changes were also necessary in order to provide the 1/8 code phase resolution (equivalent to a code phase tap separation of 122 ns) required of the DDM application. Getting narrow code-phase resolution working correctly ended up being particularly challenging.

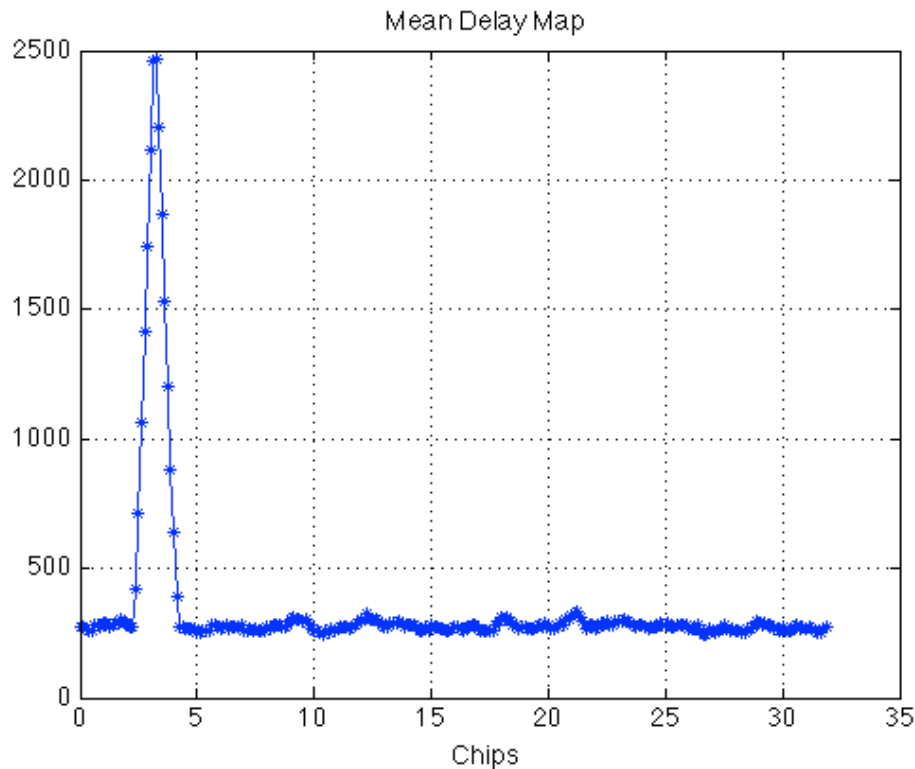


Fig. 3: Matlab generated "Test Data" correlation calculated using the DDMA (1/8 chip spacing, 1 ms samples, 140 non-coherent rounds, 50 dBHz SNR)

For this application, most of the complexity in the design is associated with the sequencing of the single channel correlator. Because the correlator is only capable of providing an 8-tap by 8 kHz DDM segment each iteration, the results from multiple iterations have to be stitched together in a patchwork manner in order to generate the final picture. Following each millisecond of sample processing, the results need to be saved in the coherent I&Q memory, the correlator reset and the input samples shifted by an amount equivalent to the required code phase shift. In the event that the very first code phase tap/set has been completed, the sample index of the last processed sample also needs to be recorded in order to permit processing to continue from that point when the next non-coherent-round/look commences. The sequencer also interrupts the processor at the end of each round and pauses the process. This allows time for the IF buffers to be replenished with new samples and I&Q coherent samples to be copied from the DDMA memory into external SRAM.

This design consumes 80% of the available FPGA fabric, but uses 100% of the available memory blocks.

Testing

Two different forms of testing were undertaken using MATLAB-generated datasets validated using the Kai-Borre software suite. The advantage of using MATLAB-generated test data is that the signal-to-noise ratios of the test data can be made arbitrarily large and the results of the tests are known to be free of any real-world artifacts.

The first form of testing was to run the test data through the ModelSim simulator supplied with the MicroSemi Libero FPGA design tools. This technique allows many internal signals to be examined in order to validate the design logic. Once acceptable results were observed in ModelSim, the same data was loaded into the real hardware and a check for consistency with ModelSim was made. Firmware supporting the upload of either previously captured data, or

MATLAB generated data into the receiver was added to the Aquarius code base, as well as commands allowing the DDMA to be configured and intermediate outputs from the DDMA to be output from the receiver for further processing and display using MATLAB.

Using these features, consistency between ModelSim and the real FPGA designs was demonstrated. Fig. 3 shows a result obtained using 50 dBHz test data generated using MATLAB and correlated using the DDMA. The 1 ms correlator outputs were then output via serial port and non-coherently integrated using MATLAB in order to generate the plot. The plot shows that the required 1/8 chip separation on the correlation curve has been achieved, with equispaced taps being apparent in the plot. The reader should recognise that real GPS-R observations would probably have a lower signal to noise ratio, although this would be compensated by performing more non-coherent rounds. That said, the maximum non-coherent integration is ultimately limited by the amount of data that is captured during each IF snapshot.

Conclusions

In this paper, we have described how we have adapted the Namuru V3.2 space-borne GPS receiver to be capable of performing GPS reflectometry. Our solution allows for the capture and download of a small segment of captured IF data, and allows for onboard processing of that data on-board the GPS receiver for applications where the downlink communications budget is severely constrained.

Some aspects of our solution are incomplete at the time of writing. This includes an ability to predict the location of the specular reflection points, which is needed if on-board processing is to be carried out. That said, this calculation could also be performed prior to each experiment and the receiver simply instructed to capture data at a specific time or location, which would have the additional benefit of reducing the complexity of the Aquarius firmware. The firmware modifications to convert a sequence of 1/8, 1/4 or 1/2 ms dump samples into 1 ms dump samples with 1 kHz frequency separation using an FFT, and the non-coherent accumulation of those samples to build up the DDM also need to be done, although this should be straightforward.

On the hardware front, the daughterboard design is yet to be completed, as is the LHCP antenna-array, with the antenna-array being the more challenging of the remaining tasks.

Acknowledgements

Several grants contributed to the development of this work, including ARC Discovery Grant DP1093982 (Preparing for the Next Generation Global Navigation Satellite System Era: Developing and Testing New User and Reference Station Receiver Designs), the ASRP Garada Project (SAR Formation Flying) and a UNSW Faculty Research Grant (GPS Reflectometry Experiment for QB50 Cubesat Mission).

The porting of the Namuru V2 GPS receiver designs to the Namuru V3.2 variant using ARM/Microsemi components, as required for the Biarri project was partially funded by the Defence Science and Technology Organisation, Department of Defence.

References

1. E. Glennon, J. P. Gauthier, M. Choudhury, K. Parkinson, and A. G. Dempster, "Project Biarri and the Namuru V3.2 Spaceborne GPS Receiver," in *IGNSS 2013 Symposium*, Outrigger Gold Coast, Australia, 2013.
2. "QB50 System Requirements and Recommendations: Interface Control Documents: Issue 4," 2013.

3. E. Kahr, O. Montenbruck, K. O'Keefe, S. Skone, J. Urbanek, L. Bradbury, and P. Fenton, "GPS Tracking on a Nanosatellite - The CanX-2 Flight Experience," in *8th International ESA Conference on Guidance, Navigation and Control Systems*, Karlovy Vary, Czech Republic, 2011.
4. S. Gleason, "Remote Sensing of Ocean, Ice and Land Surfaces Using Bistatically Scattered GNSS Signals From Low Earth Orbit," PhD University of Surrey, December 2006
5. E. P. Glennon, K. Parkinson, P. Mumford, N. Shivaramaiah, Y. Li, R. Li, and Y. Jiao, "A GPS Receiver Designed for Cubesat Operations," in *Australian Space Science Conference*, 2011, pp. 213-222.
6. E. P. Glennon, J. P. Gauthier, M. Choudhury, K. J. Parkinson, and A. G. Dempster, "Synchronization and Syntonization of Formation Flying Cubesats Using the Namuru V3.2 Spaceborne GPS Receiver," in *Pacific PNT 2013*, Honolulu, HI, 2013.
7. E. P. Glennon, P. J. Mumford, and K. J. Parkinson, "Aquarius Firmware for UNSW Namuru GPS Receivers," in *IGNSS Symposium 2011*, Sydney, Australia, 2011.
8. K. Borre, D. M. Akos, N. Bertelsen, P. Rinder, and S. H. Jensen, *A Software-Defined GPS and Galileo Receiver: A Single-Frequency Approach*: Birkhauser, 2007.
9. S. Gleason, M. Adjrad, and M. Unwin, "Sensing Ocean, Ice and Land Reflected Signals from Space: Results from the UK-DMC GPS Reflectometry Experiment," in *ION GNSS 18th International Technical Meeting of the Satellite Division*, Long Beach, CA.

Attitude Control System (ACS) of UNSW QB50 project “UNSW EC0”

Li Qiao ¹, Min Zhang ², Barnaby Osborne ¹ and Andrew G Dempster ¹

¹ *Australian Centre for Space Engineering Research, University of New South, Sydney, NSW, 2052, Australia*

² *College of Information Science and Control, Nanjing University of Information Science and technology, Nanjing, Jiangsu 210044, China*

Summary: The primary design of the Attitude Control System (ACS) for an Australian 2U CubeSat is presented. The satellite UNSW EC0, is being designed and constructed at the Australian Centre for Space Engineering Research (ACSER). The Control part is solved by using an ISIS magnetorquer board, which is a 3-axis magnetorquer system. The control system operates in stabilization modes during mission. The ACS will provide attitude control with a pointing accuracy of $\pm 10^\circ$ and pointing knowledge of $\pm 2^\circ$. The control law is performed using a time varying linear quadratic (LQ) controller. Simulation results of the stabilization process are presented and indicate the designed ACS should meet the time and power consumption requirements.

Keywords: Attitude control system, CubeSat, QB50, Magnetic control

Introduction

The University of New South Wales (UNSW) established itself as a leading centre for satellite system engineering research and education in Australia following receipt of a major grant under the Australian Space Research Program (ASRP). One of the major projects under ACSER is the Warrawal project, a Comprehensive Tertiary Education Program in Satellite Systems Engineering. As part of the space engineering related research and education program, UNSW proposed and was awarded the design and construction of one of the QB50 constellation satellites. UNSW's QB50 2U CubeSat is called UNSW QB50 Educational CubeSat 0 (UNSW EC0).

UNSW EC0 will be part of a larger constellation of satellites launched through the QB50 project run by the Von Karman Institute (VKI). The primary goal of the QB50 mission is to carry out a series of atmospheric sounding experiments in the thermosphere. To this end, all participating CubeSats will carry a VKI payload in addition to the host institute payloads. The network of 50 CubeSats will be launched into a 98 degree inclined circular orbit at an altitude of 350-380 km in a string-of-peals configuration [1]. The constellation of CubeSats are made up of a mix of 2U and 3U CubeSats, with the majority carrying a set of standardised sensors for multi-point, in-situ, long-duration measurements of key parameters and constituents in the largely unexplored lower thermosphere and ionosphere.

Satellite Configuration and Attitude Requirements

The UNSW EC0 is one of the 2U CubeSats, which has the measurements 101mm × 101mm × 227mm and a maximum allowed weight of 2 kg. The concept design of UNSW EC0 is shown in *Fig. 1*. Besides QB50 atmospheric mission, one of the host experiments is to implement an

attitude control system (ACS) onboard a single board computer running the NICTA seL4 operating system (see Fig. 2). The ACS will run in a microkernel operating system designed for security and safety-critical systems on the satellite as a test of its most critical systems functionality.

The design requirements imposed by the payloads necessitate precise 3-axis control of the spacecraft when it is in orbit. In addition to the pointing requirements, meeting the size and mass restrictions of a 2U CubeSat represents the main technical challenge of the work. The complete mission includes the payload, ADCS with actuators and sensors, deployable solar panels, on board computer and power system. Miniaturization is a key approach in order to meet the mass and volume budget.

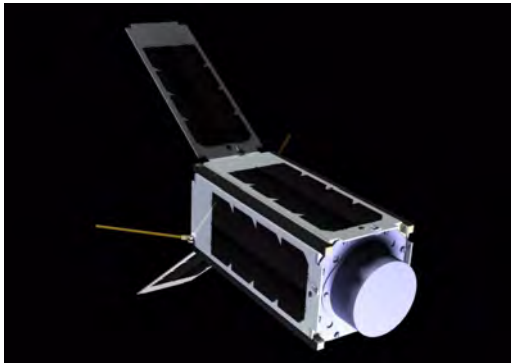


Fig. 1: Conceptual design of UNSW EC0 in the deployed configuration

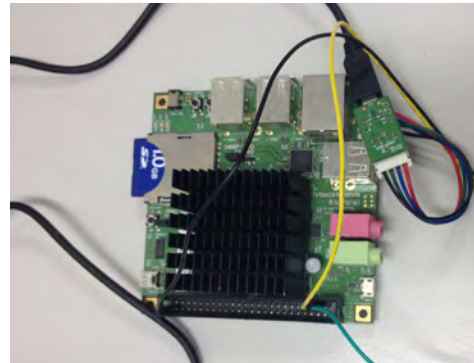


Fig. 2: seL4 computer

The ACS of the satellite requires two subsystem components: 1). the determination of the attitude of the satellite through on-board measurement and; 2). the control of its attitude through passive or active actuators. The first component determines the satellite's orientation relatively to the Earth, Sun or other object and controls the attitude actuators to point the satellite as desired. In our case, it is important to know the orientation of EC0 relatively to the Earth so it can receive the GPS signals. There are many ways to build an ACS component, which will be discussed later.

The attitude dynamics study is performed considering the yaw, pitch and roll angles of the satellite body reference frame with respect to a right-handed orbiting reference frame. Table. 1 shows the CubeSat main rotational parameters.

Table. 1: EC0 satellite mass and orbit parameters

Satellite mass	1.92 kg
Body size	10.1cm×10.1cm×22.7cm
Moments of inertia (solar panels folded)	$I_x = I_y = 0.0115 \text{ kgm}^2$ $I_z = 0.00369 \text{ kgm}^2$
Typical Orbit altitude	350.00 km
Typical Orbit period	91.39 minutes

From the experimental and communications requirements, the desired nominal attitude is nadir pointing mode when implement mission task. The attitude control system should limit the satellite attitude rotation about the body axes according to the attitude requirements. When the satellite is released from the launcher, it will have an initial angular velocity. Before the satellite switches on its payload, the body frame must be aligned with the orbit frame.

Consequently, there are two ACS operation models that need to be considered, namely detumbling and stabilization. It is noted that this paper only shows the modelling and simulation for the stabilization mode.

Mode 1: Detumbling mode

The purpose of detumbling mode (DM) is to suppress the satellite angular velocity after separation from launcher and align the body co-ordinate system with the orbital co-ordinate system. The DM consists of rate detumbling and angle detumbling. The maximum initial angular velocity when the satellite is released is $10^\circ/\text{s}$ per axis [2]. The rate after DM should be reduced to a value about several $10^{-3}^\circ/\text{s}$ per axis. After the rate detumbling, the satellite may have arbitrary attitude. The proposed control strategy is to use the B-dot algorithm [3] with magnetic torque to minimize the derivative of the magnetic field vector measured by the magnetometer.

Mode 2: Stabilization Mode (SM)

This mode is intended to keep the alignment between the body frame (b-frame) with the non-inertial orbit reference frame (*o* frame) within an acceptable pass-band. During the stabilization, the ACS will provide attitude control with a pointing accuracy of $\pm 10^\circ$ and pointing knowledge of $\pm 2^\circ$.

ACS Hardware

In the UNSW EC0 mission, the attitude sensors include Coarse sun sensing via the solar panels, fine sun sensing via the CubeSense board, rotation rates from a 3-axis Gyroscope and earth magnetic field vector sensing via an externally mounted 3-axis magnetometer. The attitude control torques are provided by three axis magnetorquers (MT) and MT based 3-axes active control techniques are used for its simple design and low power consumption.

Attitude Sensors

An integrated sun and nadir sensor, a CubeSense module (see Fig. 3) from Innovus [4], will be used as primary attitude sensor. Two 640×480 pixel CMOS cameras are dedicated to sun sensing and Earth horizon detection, respectively. The size of the module is $96\text{mm} \times 96\text{mm} \times 10\text{mm}$ (excluding cameras) and the weight is 100g , including cameras. Both cameras have wide field-of-view optics (180°) based on a fisheye lens for increased operating range. The primary outputs of the sensor are the measured sun vector and nadir vector in the sensor's coordinate frame. The measured vectors are output as azimuth/elevation angles relative to the camera bore-sight. Mapping the azimuth and elevation from spherical to three-dimensional Cartesian coordinates, the measured vector can be expressed with the two angles. For instance, the sun vector \hat{S} in *b* frame can be found as

$$\hat{S} = \begin{bmatrix} \cos(\text{Elevation}) \cdot \cos(\text{Azimuth}) \\ \cos(\text{Elevation}) \cdot \sin(\text{Azimuth}) \\ \sin(\text{Elevation}) \end{bmatrix}$$

The main problem is the sun sensor needs direct sunlight. When the satellite is in the Earth shadow, the attitude determination could use the measurements of earth horizon.



Fig. 3: CubeSense module

Attitude control actuator

For actuation of the satellite, magnetic torque coils will be used. The actuators are an iMTQ board that provides three magnetic actuators for attitude control operations. The board incorporates two torque rods with a metallic core (see Fig. 4 a)) and one air core torque (Fig. 4 b)). The benefit of using magnetic coils is that they both inexpensive and lightweight compared with reaction wheels and thrusters, and do not suffer from saturation or require a stored propellant. Three-axis stabilization of the 2U CubeSat can be achieved by using feedback from magnetic field measurements.

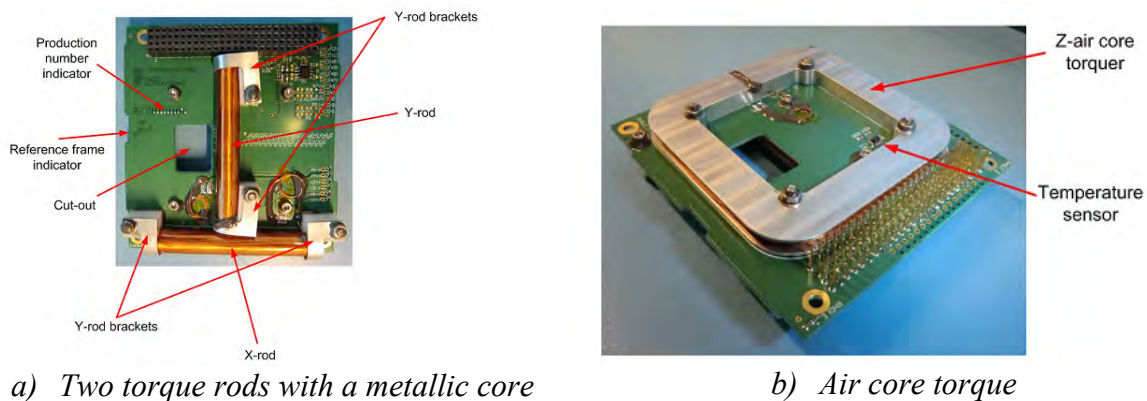


Fig. 4: ISIS magnatorquer board

ACS Modelling

Basics in Satellite Attitude Control

Before inducing the ACS, this section describes the definitions of the different reference frames used throughout the paper.

Non-inertial Orbit Reference Frame: Orbit frame is a kind of Vehicle Velocity Local Horizontal (VVLH) coordinate frame as shown in Fig. 5. The origin of the orbit frame also lies with the satellite centre of mass and it has the x-axis pointing along the positive velocity vector direction tangentially to the orbit, the z-axis pointing toward the centre of the Earth, and the y-axis completes the right-handed Cartesian coordinate system. The origin rotates at an angular velocity ω_o relative to the earth centred inertial (ECI) frame. The satellite attitude

is described by roll, pitch and yaw, which is the rotation around the x -, y -, z - axis respectively. This frame is represented by the letter o .

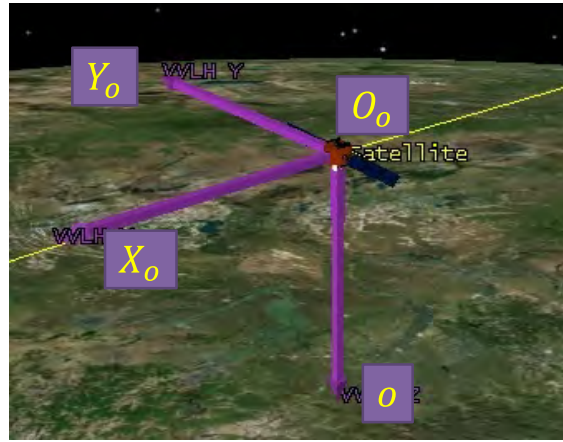
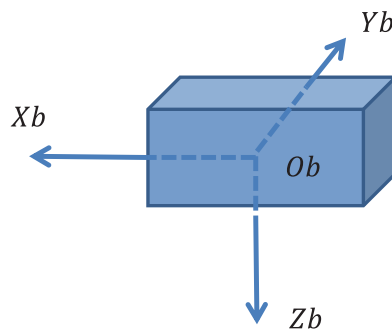
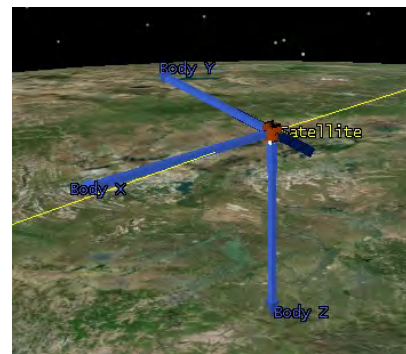


Fig. 5: Orbit frame (o)



a) Body frame



b) Body frame during stabilization

Fig. 6. Body frame (b)

Body Reference Frame: The body frame (see Fig. 6 a)) is fixed to the geometry of the satellite and moves and rotates with the satellite. Its origin lies with the centre of satellite mass, the x -axis points in the forward direction, the z -axis points down side, and the y -axis completes the right-handed orthogonal system. The rotation between the orbit frame and the body frame is used to represent the attitude of the satellite. This frame is represented by b . As aforementioned, the desired nominal attitude is nadir alignment with ECF velocity constraints when implement mission task. Fig. 6 b) shows the body frame when the satellite is in the stabilization mode.

Euler Angles: Roll, pitch and yaw. Euler angles are commonly used to represent attitude. There are twelve possible sets of Euler angles. The first, second and third rotation is about the zyx – axis of the body frame, respectively. Roll is the angle about the x – axis. Pitch is the angle about y – axis. Yaw is the angle about the z – axis.

Attitude quaternion: As an alternative to the Euler angle representation, a quaternion is a four row-vector commonly used to represent attitude [5]. Noted that both quaternions and Euler angles are representatives of attitude rotations and conversions between the two can be performed.

An assumption when we study the attitude control in this paper is that the satellite can be treated as a rigid body with a constant inertia matrix and a non-moving centre-of-mass.

Attitude kinematical and dynamical model

Use the attitude quaternion $q = [\eta \quad \varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3]^T$ to present the satellite kinematics.

$$\begin{cases} \dot{q} = \begin{bmatrix} \dot{\eta} \\ \dot{\varepsilon} \end{bmatrix} \\ \dot{\eta} = \frac{1}{2} \varepsilon^T \omega_{ob}^b \\ \dot{\varepsilon} = \frac{1}{2} [\eta I_{3 \times 3} + S(\varepsilon)] \omega_{ob}^b \end{cases} \quad (1)$$

η and $\varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3]$ are the real part and the three imaginary part, respectively. $S(\varepsilon)$ is the skew-symmetric matrix, indicating the cross product of ε . ω_{ob}^b is the angular velocity of frame b , relative to o frame, represented in frame b . The dynamical equations for the satellite are the Euler's equations [5], written as

$$I \dot{\omega}_{ib}^b + \omega_{ib}^b \times (I \omega_{ib}^b) = T_m^b + T_g^b \quad (2)$$

$I = \text{diag}[I_x \quad I_y \quad I_z]^T$ is the inertia matrix of the satellite, ω_{ib}^b is the body rate measured about the body axes with respect to the inertial frame. $T = T_m^b + T_g^b$ is the sum of all external torques on the satellite. T_m^b is the magnetic torque, and T_g^b is the gravity torque. $\omega \times$ is given by

$$\omega \times = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (3)$$

The ω_{ob}^b in Equation (1) and ω_{ib}^b in Equation (2) have equation as

$$\omega_{ob}^b = \omega_{ib}^b - R_o^b \omega_{io}^o = \omega_{ib}^b + \omega_o c_i^b \quad (4)$$

$R_o^b = [c_1 \quad c_2 \quad c_3]$ is rotation matrix from o frame to b frame. c_i ($i = 1, 2, 3$) is the column vector of c_i^b . ω_o is the orbit rotation velocity which is a constant to a certain orbit.

$$T_m^b = m^b \times B^b \quad (5)$$

m^b is the total dipole moment, which is the product of the number of windings N , the cross section area A and the current i through the respective inductors

$$m^b = [N_x A_x i_x \quad N_y A_y i_y \quad N_z A_z i_z]^T \quad (6)$$

The momentary power consumption of the torques are computed as

$$P = V(i_x + i_y + i_z) \quad (7)$$

and where V is the voltage of the coils.

\mathbf{B}^b is the local Earth magnetic field vector. It can be seen from Equation (5) that the earth magnetic field must be known or estimated. The International Geomagnetic Reference Field (IGRF) is an approximation near and above the Earth's surface, working with an orbit estimator [6]. Since the orbit determination part is not added to the ACS simulation, a simplified model is used to estimate the magnetic field [7]. In this model, the geomagnetic field is near periodic and only related to the latitude. The advantage of using this model is that the mean value of magnetic field can be used to make the attitude-system equation time-invariant.

$$\mathbf{T}_g^b = 3\omega_0^2 \mathbf{u}_e \times (\mathbf{I} \mathbf{u}_e) = \frac{3\mu}{R_c^3} \mathbf{u}_e \times (\mathbf{I} \mathbf{u}_e) \quad (8)$$

where $\mu = 3.987 \times 10^{14}$ is the Earth's gravitational coefficient. R_c is the distance to the Earth's center.

Control law

According to attitude kinematic Equation (1) and dynamic Equation (2), the attitude system equation can be written as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad (9)$$

Where the state $\mathbf{x} = [\varepsilon_1 \ \dot{\varepsilon}_1 \ \varepsilon_2 \ \dot{\varepsilon}_2 \ \varepsilon_3 \ \dot{\varepsilon}_3]^T$ and the control variable $\mathbf{u} = [m_x \ m_y \ m_z]^T$ which are the torque vectors. \mathbf{A} and \mathbf{B} are given as

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -4k_x\omega_0^2 & 0 & 0 & 0 & 0 & (1-k_x)\omega_0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -3k_y\omega_0^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -(1-k_z)\omega_0 & 0 & 0 & -k_z\omega_0^2 & 0 \end{bmatrix}_{6 \times 6} \quad (10)$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2I_x} B_z^0 & \frac{1}{2I_x} B_y^0 \\ 0 & 0 & 0 \\ \frac{1}{2I_y} B_z^0 & 0 & \frac{1}{2I_y} B_x^0 \\ 0 & 0 & 0 \\ \frac{1}{2I_z} B_y^0 & \frac{1}{2I_z} B_x^0 & 0 \end{bmatrix}_{6 \times 3} \quad (11)$$

where $k_x = \frac{I_y - I_z}{I_x}$, $k_y = \frac{I_x - I_z}{I_y}$ and $k_z = \frac{I_y - I_x}{I_z}$ in Equation (10)

The optimal controllers, which is the same control law as in [8], is used here. It is a combination of an angular velocity feedback and a time varying linear quadratic (LQ)

controller chosen for the stabilization phase of the ACS system. LQ controllers are known to be reliable and robust about the reference, so it can be used for satellites with magnetorquer. LQ problem in control theory minimizes a quadratic cost function, which penalize the deviation from the reference attitude as well as actuator usage subject to weighting matrices:

$$J = \frac{1}{2} \int_{t_0}^T [\tilde{\mathbf{x}}^T \mathbf{Q} \tilde{\mathbf{x}} + \mathbf{u}^T \mathbf{P} \mathbf{u}] dt \quad (12)$$

where $\tilde{\mathbf{x}}$ is the deviation from the reference attitude, which in this case is zero, \mathbf{u} is actuator usage and \mathbf{Q} , \mathbf{P} are positive semidefinite weighting matrices. The solution to the LQ problem is found by solving the continuous Riccati equation:

$$\mathbf{R} = -\mathbf{R}\mathbf{A} - \mathbf{A}^T \mathbf{R} + \mathbf{R}\mathbf{B}\mathbf{P}^{-1}\mathbf{B}^T \mathbf{R} - \mathbf{Q} \quad (13)$$

which gives the control law:

$$\mathbf{u}(\mathbf{t}) = -\mathbf{P}^{-1}\mathbf{B}^T \mathbf{R}\mathbf{x}(\mathbf{t}) \quad (14)$$

The weighting matrices \mathbf{P} and \mathbf{Q} in Equation (12) are given as

$$q_{ii} = \frac{1}{(\Delta x_i)^2} \quad (15)$$

$$p_{ii} = \frac{1}{(\Delta u_i)^2} \quad (16)$$

where Δx_i and Δu_i are nominally acceptable deviations in the state and actuator use respectively [8]. It should be noted that the optimal controllers are used on the system linearized around the equilibrium, when the deviations from reference are small. The matrix $\mathbf{R}(\mathbf{t})$ has to be computed for each time step, therefore LQ control demands computation. This will be tested in the seL4 board in the following work.

Simulation and Conclusion

In this section, the controllers for stabilization will be simulated. The parameters of the model used in the simulation are listed in Table. 2.

Table. 2 Simulation parameters

Coil voltage	5 V
Initial angles	$[\phi \quad \theta \quad \psi] = [-3.2^\circ \quad 2^\circ \quad -3.3^\circ]$
Initial angular velocity	$\omega = [0.0001^\circ/s \quad 0.0006^\circ/s \quad -0.0003^\circ/s]$
Weighting matrices in Equation (15) and Equation (16)	$\mathbf{Q} = \text{diag}[1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0] \frac{1}{\left(\frac{10\pi}{180}\right)^2}$ $\mathbf{P} = \text{diag}[1 \quad 1 \quad 1] \frac{1}{(0.01)^2}$

In Fig. 7 and Fig. 8, the Euler angles and angular rates in three axis of the satellite using the stabilizing controller (14) are shown. It can be seen that the satellite attitude is stabilized

within 1 orbit and meets the control requirements where the desired pointing accuracy is $\pm 10^\circ$. The momentary and total power consumption is shown in Fig. 9 and can be seen to be within the specification. Consequently, the LQ controller presented in the paper should meet the time and power consumption requirements.

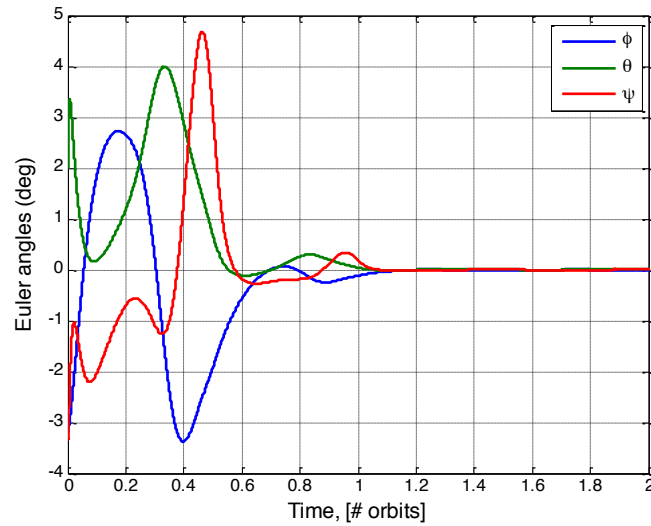


Fig. 7: The stabilized attitude of EC0 in terms of Euler angles

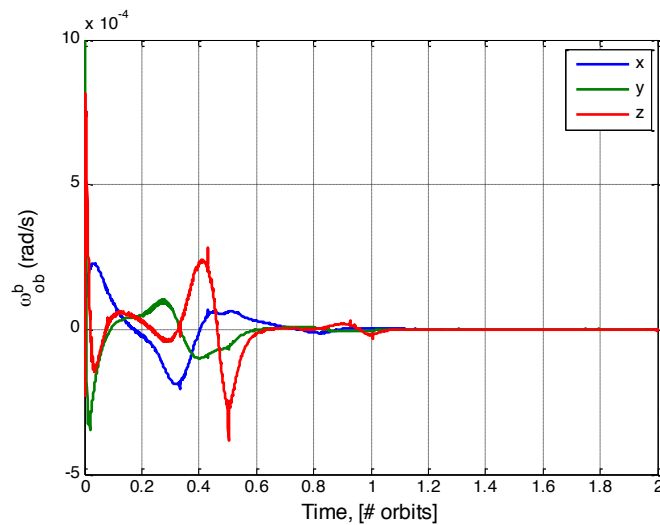


Fig. 8: The stabilized attitude of EC0 in terms of Euler angular rates

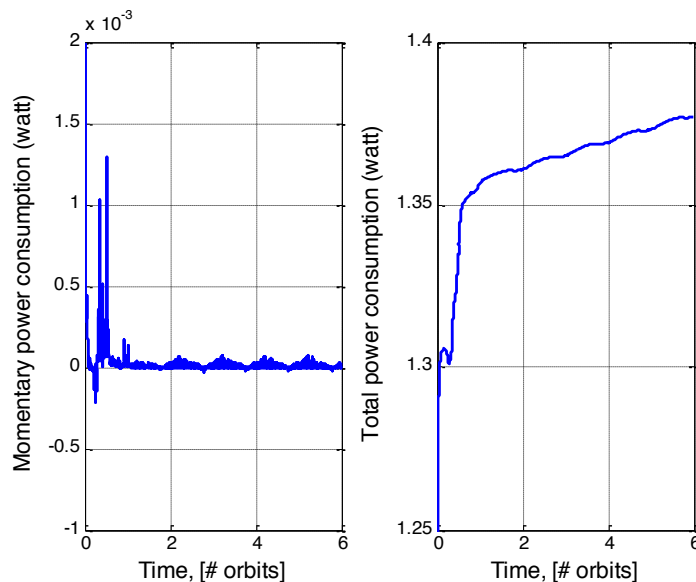


Fig. 9: The momentary and total power consumption

Further work

The paper shows a concept design of the ACS work in UNSW EC0 project. There are some simplifications in the control algorithm, such as use of periodic model to estimate the earth magnetic field, rather than using the IGRF model. Further work includes integration of the orbit determination part and inclusion of IGRF, as well as adding attitude determination part to ACS. The work also includes the testing ACS on seL4bit microkernel board. The final satellite will undergo assembly and testing during later 2014 and launch is planned for 2015.

Acknowledgements

The authors gratefully acknowledge the support of the following partners. The UNSW EC0 is part of the education programme Space Engineering supported by UNSW and Australian Space Research Programme (ASRP).

References

- 1 . QB50 Project Description, <<https://www.qb50.eu/index.php/project-description>> (Retrieved November 2013)
2. QB50 System Requirements and Recommendations. Interface Control Documents. Issue 3, 5 February 2013. <<https://www.qb50.eu/download/requirements/QB50SystemRequirementsDocument-20130206.pdf>> (Retrieved November 2013)
3. Shou, H., Sheu, J., and Wang, J., "Micro-satellite detumbling mode attitude determination and control : UKF approach", *Procedings of Control and Automation (ICCA), 2010 8th IEEE International Conference*, Xiamen , China, June , 9-11 2010pp. 673,678, 9-11
- 4 . Innovus. CubeSense <http://www.innovus.co.za/media/component-sales-pdfs/CubeSense_web.pdf> (Retrieved November 2013)

5. Ruiter, A.D., Damaren, C.J., Forbes, J.R., *Spacecraft Dynamics and Control: An Introduction*, Wiley, 2013.
6. Gravdahl, J.T., Eide and E., Skavhaug, A., et.al, "Three Axis Determination and Control System for a Pico-Satellite: Design and Implementation", Proceedings of the 54th International Astronautical congress of the international astronautical federation, the international academy of astronautics, and the international institute of space law, Bremen, Germany, September 29 – October 2, 2003.
7. Kane, P.C., and Carlisle, P., *Spacecraft attitude dynamics*. J. Wiley, 1986.
8. Holberg, F.S., Optimal Attitude Control of a Double Cubesat using Magnetorquers, Final project report, Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), May 19, 2011.

Satellite Vibrations and Restoration of High Resolution Pushbroom Imagery

Liyuan Li ¹, Wei Zhang ², Linlin Ge ^{1*} and Xiaojing Li ¹

¹ *Geoscience and Earth Observing Systems Group (GEOS), School of Civil and Environmental Engineering, UNSW Australia, Sydney NSW 2052, Australia*

² *Research Center of Space Optical Engineering, Harbin Institute of Technology, Harbin 150001, China*

Summary: Pushbroom scanners are widely applied in passive optical remote sensing owing to their high spatial resolution and radiation responsivity. Satellite vibrations, however, result in geometric distortion and blur in the images. As the spatial resolutions go higher and the number of TDI (Time Delay and Integration) lines increased, these effects become more noticeable. In view of the precision limits of satellite platform control, image restoration is necessary to mitigate these effects. Firstly, satellite vibrations and their effects are studied. These facts, in combination with the other degradation factors in remote sensing, are then used to build an imaging model. On the basis of the model, a new restoration algorithm that takes vibration effects into consideration is proposed, so that high quality imagery can be delivered while the design of critical control is relieved. The simulation results indicate that the proposed algorithm can correct vibration effects and restore images effectively.

Keywords: Image restoration, high resolution imagery, satellite vibrations, passive optical imaging, pushbroom scanner

I. Introduction

Passive optical remote sensing is widely applied in the missions of earth observation, owing to its capability of acquiring various information of land objects. The information can be pre-processed, transmitted and analyzed so that vivid pictures of lands are revealed. In order to observe the scenes of interest in finer details, remote sensing satellites of higher spatial resolution are developed in the past few decades. This, however, brought about a new problem in passive optical imaging – satellite vibrations cause irregular angular shift in the principle axis of optical system, which result in geometric distortion and blur in images. These effects brought errors to subsequent image processing such as registration. They become noticeable as spatial resolution goes higher and, what is more important, the effects will be further aggravated by the use of pushbroom scanners. The idea of using pushbroom scanners such as TDI-CCD (Time Delay Integration Charge-Coupled Device) in passive optical remote sensing is that they can acquire more radiation from land scenes so as to maintain high SNR (Signal-to-Noise Rate) even in finer resolution. Hence, for the sake of mitigating vibration effects and acquiring high resolution imagery of good quality, two approaches which focus on different areas have been considered – to suppress satellite vibration by improving the control precision of satellite platform at very beginning, or to restore the acquired images using certain algorithms afterwards. Though solving the problem fundamentally, it may be less cost-effective by using the first approach, and a comparatively

* Corresponding author. Email: l.ge@unsw.edu.au

long period could be also required to make breakthrough in the control precision. Image restoration, accordingly, can work as an alternative approach.

Vibrations in different satellite platforms have been studied in the last few decades mainly through experiments [1,2,3,4] and FEM (Finite Element Method) [5,6,7]. It can be concluded that the vibration of 10Hz or lower frequency occupied the majority part of vibration energy, and nearly all the energy lies within 100Hz. Therefore, though many efforts have been paid on the study of vibrations effects [8,9,10], especially on high frequency vibrations and vibration blur, it could deliver better results for the very case of satellite remote sensing if more attention could be given on lower frequency vibrations and their major effect (vibration distortion). Besides, some restoration algorithms that focus mainly on vibration distortion have been proposed [11,12]. Although there do have some comprehensive restoration algorithms that considered both kinds of vibration effects [13,14,15], further adaptation is possible to be made in the area of pushbroom imaging.

In view of the current status of related research, in order to solve the problem facing by high resolution imagery and relieve the requirement of precise control, this paper analyzes satellite vibrations and reveals their effects on imagery, constructs the imaging model for high resolution imagery, and aims to propose a new restoration algorithm based on the imaging model so that vibration effects can be took into consideration. The paper is structured in the following order. In Section II, satellite vibrations and their influence on imaging is discussed. These are followed by the analysis of vibration degradations and, in combination of the other degradation factors of imaging chain, an overall imaging model is constructed in Section III. Based on the model, the inverse problem of imaging is studied in Section IV and a new restoration algorithm for high resolution pushbroom imagery is proposed. Finally, Section V concludes the significants of the paper.

II. Satellite Vibrations

It has been known that satellite vibrations will cause degradation in high resolution pushbroom imagery. To mitigate their effects and restore imagery comprehensively, this section traces back to the specific sources of image degradation, and analyzes their patterns.

Sources of Vibration

Vibrations in satellite platform are not uniform; instead, they come from various kinds of sources. Regardless the fact that most of them carry merely a small amount of energy, they do cause distortion and blur on high resolution imagery. According to the different of vibration frequency, these sources can be categorized into to several groups. While sources such as solar panel, control loop rate error, uncompensated IMU (Inertial Measurement Unit) bias, flexible body dynamics and propellant slosh generate low frequency vibrations, vibrations of higher frequency can be brought by reaction wheel imbalance, cryocooler, and high-gain antenna gimbal drives [6] etc. The majority of vibration energy lies within 10Hz, and most of it is lower than 100Hz [3]. In addition to these, perturbation forces (e.g. aerodynamic forces, solar radiation pressure and non-spherical mass effects [16]) will also cause some long-period changes in satellite altitude and direction.

Vibration Model

To conclude, vibration patterns vary from one to another in satellite platform. All the vibrations are generated irrelevantly with random initial phase and most of them can be expressed as sinusoidal wave after platform controlling. In spite of the difference of their

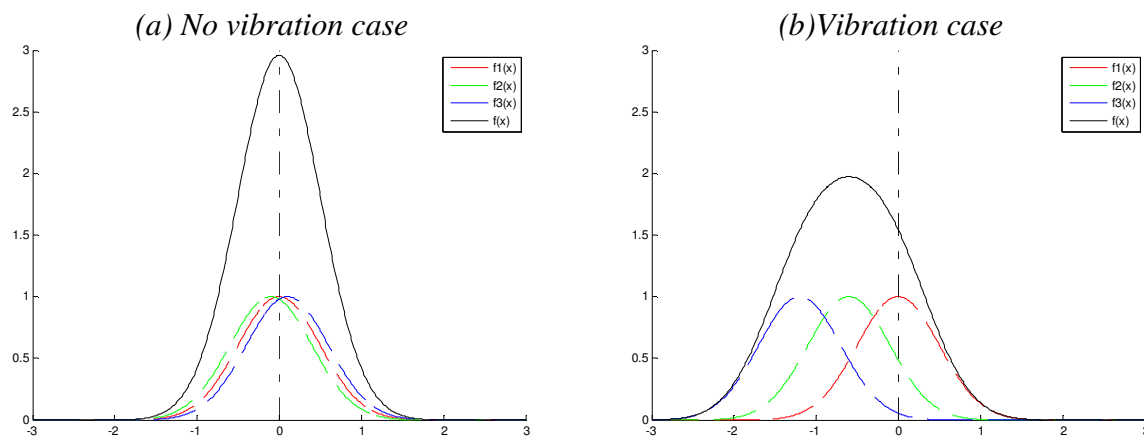
patterns, vibration in satellite platform can be described by applying sinusoidal decomposition:

$$A(t) = \sum_{i=1}^{\infty} A_i \cos(2\pi f_i t + \varphi_i), \quad (1)$$

where $A(t)$ is the vibration of satellite platform, t is vibration time, and A_i , f_i and φ_i are the amplitude, frequency and initial phase of the i -th vibration component respectively. In practice, vibration information can be measured by comparing with gyroscope (vibrations lower than 1Hz) and angular vibration sensors (1~1000Hz vibrations) [17].

The Influence of Satellite Vibration

It can be noticed that Eq. (1) contains a variety of vibration components with random initial phase, which indicates that vibration in platform tends to be random. Consequently, it causes irregular shift in the image plane of optical system. The irregular shift is captured by pushbroom CCD as the device is scanning through image plan, and hence line-wise irregular sampled data (geometric distorted image) is acquired. In addition, because image plane keep shifting within the integration time of CCD, a motion blur is also brought by the vibration. On top of these, what aggravates the effects is the application of TDI mode – integrates a single line of scene by multiple lines of CCD, which cause the superstition of vibration effects. *Fig. 1* demonstrates the corresponding process in image plane. Note that the energy of a single point spreads out as it goes through imaging chain. When there is no vibration, the centre of spread point falls accurately in the expected location; when the imaging system is affected by vibration, however, the centre of spread point shifts while CCD integrates, and hence a the overall distribution is shifted and further spread, which corresponding to distortion and additional blur respectively.



*Fig. 1: The spread of radiation energy. The radiation energy of a single point goes through the imaging chain and is acquired by CCD array. Assume that for one integration period T , the energy distributed on CCD array is $f(x)$, and the energy distributions in $[0, T/3]$, $[T/3, 2T/3]$ and $[2T/3, T]$ are $f_1(x)$, $f_2(x)$ and $f_3(x)$ respectively, *Fig. 1(b)* shows that due to the image shift cause by vibration, the distribution is displaced and further spread, which result in distortion and additional blur in images.*

Imaging Model of High Resolution Pushbroom Imagery

Apart from the degradations caused by satellite vibrations, there are numbers of factors which will result in distortion, blur and noise in the acquired images. This section studies vibration effects. Then, by combing the other degradation factors in imaging chain, an overall imaging model for high resolution pushbroom imagery is built.

Degradations Caused by Vibration

Satellite vibrations result in irregular geometric distortion and motion blur in images. Vibration distortion can be modeled by the following equation when the CCD sampling rate and its pixel size are known:

$$\varepsilon(x) = \sum_{i=1}^M \frac{A_i}{a} \cos(2\pi \frac{f_i}{1/T_{\text{int}}} x + \varphi_i), \quad x \in \mathbb{R}, \quad (2)$$

where $\varepsilon(x)$ is the shift of CCD sampling in x direction, a is the size of single CCD pixel, T_{int} is integration time, and A_i , f_i and φ_i are the amplitude, frequency and initial phase of the i -th vibration component of optical system respectively. f_1, f_2, \dots, f_M follow such an order that $f_1 < f_2 < \dots < f_M$, and f_M is smaller than the Nyquist frequency of sampling. Similarly, $\varepsilon(y)$ can be modeled. Specifically, for the case of pushbroom CCD, pixels of the same column should have the same amount of displacement, which is illustrated by Fig. 2.

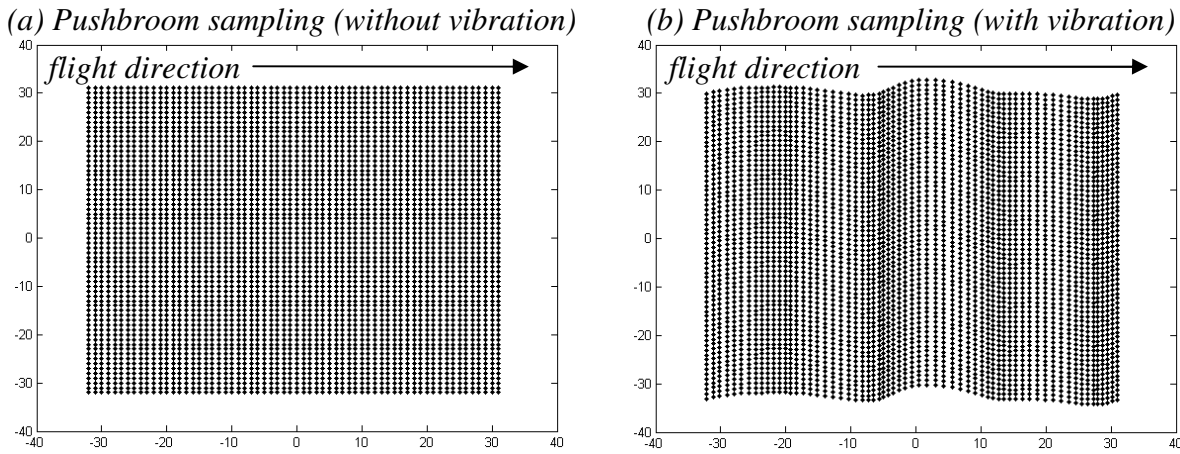


Fig. 2: Sampling displacement caused by satellite vibrations, in pushbroom system.

Despite the motion blur cause by scanning, LSF (Line Spread Function) of vibration blur is

$$\text{LSF}(x) = \frac{1}{T_{\text{int}}} \int_{t_0}^{t_0+T_{\text{int}}} \delta(x - A(t)) dt, \quad (3)$$

where t_0 is the initial time of CCD integration, and $A(t)$ is the vibration of optical system. Therefore, by applying Fourier transform, the OTF (Optical Transfer Function) of vibration blur is obtained:

$$\begin{aligned} \text{OTF}(u) &= \int_{-\infty}^{\infty} \text{LSF}(x) \exp(-j2\pi ux) dx \\ &= \frac{1}{T_{\text{int}}} \int_{-\infty}^{\infty} \int_{t_0}^{t_0+T_{\text{int}}} \delta(x - A(t)) \exp(-j2\pi ux) dt dx \\ &= \frac{1}{T_{\text{int}}} \int_{t_0}^{t_0+T_{\text{int}}} \exp(-j2\pi u A(t)) dt, \end{aligned} \quad (4)$$

and the corresponding MTF (Modulation Transfer Function) can be approximated by

$$\begin{aligned} \text{MTF}(u) &= |\text{OTF}(u)| \\ &= \frac{1}{T_{\text{int}}} \left| \int_{t_0}^{t_0+T_{\text{int}}} \exp(-j2\pi u A(t)) dt \right| \\ &\approx \frac{1}{T_{\text{int}}} \sum_{n=1}^N \left[\exp(-j2\pi u A(t_n)) \cdot \frac{T_{\text{int}}}{N} \right], \end{aligned} \quad (5)$$

where N is the number of sections divided from $[t_0, t_0+T_{\text{int}}]$, and t_n is the initial time of the n -th

section. Two dimensional MTF can be derived in similar way.

Eq. (2) and Eq. (5) indicate that both vibration distortion and blur is determined by $f_i/(1/T_{\text{int}})$ (i.e. vibration frequency to sampling frequency ratio) and A_i/a (i.e. vibration amplitude to CCD pixel size ratio). In most of cases, the higher the ratios, the more obvious the vibration effects becomes (except when the vibration frequency is higher than sampling frequency [18]). These can be understood by considering the case that $f_i/(1/T_{\text{int}})$ and A_i/a increased, in which the shift in the image plane of optical system will become faster and larger.

Degradation Factors in Imaging Chain and Imaging Model

The original information of land scenes go through imaging chain to image. Degradations will be brought to the acquired images when this process is not ideal – all atmosphere, satellite attitude (in which satellite vibrations belongs to), optical system, subsequent electronic devices etc. could be the degradation factors. Generally, they cause distortion (e.g. vibration distortion, optical aberration and distortion of turbulence), blur (nearly all of the factors will cause flatten the MTF of imaging chain, i.e. cause image blur) and noise (e.g. stray light, shot noise and the noise in electronic devices), and therefore, the imaging process of high resolution pushbroom imagery can be expressed as follow:

$$g = SHu_c + n, \quad (6)$$

where $u_c: \mathbb{R}^2 \rightarrow \mathbb{R}$ (i.e. u_c is a function from \mathbb{R}^2 to \mathbb{R}) is the original information of land scene, $g: \Lambda \rightarrow \mathbb{R}$ is the acquired image, Λ is the set of sampling coordinates, S is sampling operator and will takes the form of un-equal space sampling when geometric distortion (e.g. vibration distortion) exist, H is the blur kernel of imaging chain, and noise n is assumed to be additive.

III. Restoration Algorithm

Once the imaging model is built, method to restore high resolution pushbroom imagery can be derived by studying the inverse problem of imaging. This section starts by explaining the discrete expression of imaging and comes to the study of image restoration afterwards.

When there is no vibration and the other distortion factors in imaging chain, the sampling coordinate of CCD should be equally-spaced and takes the form $\lambda_{mn} = (m, n)$, where $m, n \in \mathbb{N}$; when distortion occurs, however, it brings displacement $\varepsilon_{mn} = (\varepsilon_x(m, n), \varepsilon_y(m, n))$ to each sampling point, which result in the change of sampling coordinate:

$$\Lambda = \Omega + \varepsilon(\Omega) = \{(m + \varepsilon_x(m, n), n + \varepsilon_y(m, n))\}_{m=1, n=1}^{M, N}, \quad (7)$$

where $\Lambda = \{\lambda_{mn}\}_{m=1, n=1}^{M, N}$ is the set of actual sampling coordinates, and $\Omega = ([1, M] \times [1, N]) \cap \mathbb{Z}^2$ is the set of ideal sampling coordinates. Thus, assuming that the size of perfect image u (i.e. discrete form of the original information of land scene) is $N \times N$, the process of distortion can be approximated by:

$$g_{mn} = g(\lambda_{mn}) = \sum_{(p, q) \in \Omega} s_{pq}(\lambda_{mn}) u(p, q), \quad (8)$$

where $\{g_{mn}\}_{m=1, n=1}^{N, N}$ is the distorted image, $\{s_{pq}\}_{p=1, q=1}^{N, N}$ is resampling matrix, and $\Lambda = \{\lambda_{mn}\}_{m=1, n=1}^{N, N}$ is irregular when distortion exist. To simplify, let $\lambda_k = \lambda_{mn}$ iff $k = (n-1)N + m$ (g_k, s_i and u_i are set similarly):

$$g_k = g(\lambda_k) = \sum_{i \in \Omega} s_i(\lambda_k) u_i, \quad (9)$$

and the equation system can be written in matrix form as

$$g = Su, \quad (10)$$

or

$$\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{N^2} \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1N^2} \\ s_{21} & s_{22} & \cdots & s_{2N^2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N^2 1} & s_{N^2 2} & \cdots & s_{N^2 N^2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N^2} \end{bmatrix}, \quad (11)$$

where $S = \{s_{ki}\}_{k=1, i=1}^{N^2, N^2}$ is the resampling matrix in which $s_{ki}=s_i(k)$ is satisfied, and the matrix can be built by applying Lagrange interpolation.

Thus, the discrete expression of Eq. (6) is written as

$$g = SHu + n, \quad (12)$$

and the inverse problem of imaging turns to the following minimization problem:

$$\arg \min_u \{ \|u\|^2 + \lambda \|SHu - g\|^2 \}, \quad (13)$$

where λ is Lagrange multiplier. This can be solved by finding the optimal solution of u which satisfies $\|SHu - g\| \leq N^2 \sigma_n^2$ (where σ_n is the standard deviation of noise), and therefore it approximates to the solution of $\arg \min_u \|SHu - g\|^2$, or

$$H^T S^T WSHu = H^T S^T Wg, \quad (14)$$

in which the superscript “ T ” means transpose. In Eq. (14), both sides of equation is multiplied by $H^T S^T W$, where $W = \text{diag}\{w_i\}_{i=1}^{N^2}$ is a weight matrix and used to compensate the variation of sampling density [13]. Suppose that for the coordinate $(k, l) \in \mathbb{Z}^2$ there are n_{kl} sampling points fall in $[k, k+1] \times [l, l+1]$, the weight of these points will be:

$$w_i = 1/n_{kl}, \quad \lambda_i \in [k, k+1] \times [l, l+1]. \quad (15)$$

Now, by replacing $H^T S^T WSH$ with A and $H^T S^T Wg$ with B , Eq. (14) becomes

$$Au = b. \quad (16)$$

Since A is symmetric positive-definite matrix, Eq. (16) can be solved by applying conjugate gradient (CG) method and the restoration algorithm is as follows:

- 1) Build the resampling matrix S and calculate the vector $b = H^T S^T Wg$;
- 2) Set the initial value u_0 and stop criteria η for conjugate gradient method;
- 3) Approximate the solution of $Au = b$ by conjugate gradient method. Desired results could be obtained when the approximation u' get close enough to the solution and the stop criteria η is chosen appropriately.

IV. Simulation and Evaluation

To validate the proposed algorithm, simulation is run in this section. Beforehand, a set of high quality images is assumed to be perfect images of land scenes and degraded by applying the imaging model of section III. The algorithm is then used to restore the degraded images. Finally, images from different scenes and of different extents of degradation are tested to evaluate the overall performance of the algorithm.

Simulation

As the degraded images are designed to be simulated by using Eq. (12), resampling matrix S , blur kernel H and noise n must be specified initially.

For geometric distortion, satellite vibrations are assumed to be the major source here. It can be derived from section II and Eq. (2) that the sampling displacement caused by vibration is highly random and declined as the frequency become higher. Accordingly, a simplified model of Eq. (2) can be built by regarding displacement as colored noise [14]:

$$\begin{cases} \hat{\varepsilon}_x(u, v) \sim N(0, \tilde{\sigma}^2), & \|(u, v)\|_2 \leq N/T', \\ \hat{\varepsilon}_x(u, v) = 0, & \|(u, v)\|_2 > N/T', \end{cases} \quad (17)$$

where ε_x is the displacement in x direction and written as $\hat{\varepsilon}_x$ in frequency domain, T' is the minimum period of distortion and should satisfy the condition $T' \geq 2$ pixels (due to the limit of Nyquist frequency), and the standard deviation of ε_x is A' . ε_y is define similarly. The larger $1/T'$ and A' are, the wider the frequency distribution and the large the overall amplitude are. Once the displacement of each pixel is simulated, bilinear interpolation is used to build the resampling matrix (for quality, bicubic [19], visual oriented interpolation [20], etc. can be applied).

For image blur, since the overall MTF of imaging chain is the product of each section and the MTF distribution of most of the sections are similar to Gaussian distribution or sinc function, the blur kernel here is approximated by the following equation:

$$\text{MTF}(u, v) = -\alpha \exp(u^2 + v^2), \quad (18)$$

where α is the variable relates to standard deviation. Standard deviation (the width of distribution) of MTF decrease as α increase, and consequently, image is further blurred.

Finally, additive white Gaussian noise is assumed to be the form of noise in imaging chain. A set of simulation, including images of farmland, urban area and river (size: 256×256, panchromatic, 8-bit)*, is run and demonstrated below. The land scenes corresponding to perfect image 1, 2 and 3 are denoted as scene 1, 2 and 3.

Overall Performance Evaluation

Since the information of perfect images is known, a full reference image quality assessment method, Structural Similarity (SSIM), is applied here to evaluate the performance of the proposed algorithm. Except luminance and contrast information of image, SSIM takes the structure information of images into consideration. This is importance as the simulated images contain geometric distortion and is the reason why it is used to replace Signal-to-Noise Ratio (SNR) to evaluate image quality here. According to [21], SSIM is defined as:

$$\text{SSIM}(f, g) = \frac{(2\mu_f\mu_g + c_1)(2\sigma_{fg}^2 + c_2)}{(\mu_f^2 + \mu_g^2 + c_1)(\sigma_f^2 + \sigma_g^2 + c_2)}, \quad (19)$$

where μ_f and μ_g are the means of image f and g , σ_f and σ_g are the standard deviations of image f and g , σ_{fg} is the co-variance between f and g , $c_1 = k_1 L^2$, $c_2 = k_2 L^2$ in which L is the dynamic range of image and c_1 and c_2 are the constants that are usually set to 0.01 and 0.03 respectively.

To evaluate the performance of the proposed algorithm, three group of images of different scenes (i.e. scene 1, 2 and 3) and different extents of degradation are restored using the algorithm, and the corresponding evaluation results are demonstrated in *Table 2* and *Fig. 4*†.

* Image source: SPOT 6 1.5m Ortho Bundle, Barcelona, Spain, 12/07/2012.

Available from: <http://www.astrium-geo.com/en/23-sample-imagery>

† The computer configuration is as follows: processor - Intel(R) Core(TM) i5-3210M CPU @ 2.50GHz, RAM - 6 GB, software platform – MATLAB R2011b.

Each group contains three images with increasing extent of degradation. And according to the extent of degradation, the degraded images are denoted by No.1, No.2 and No.3 respectively. Parameters that used to simulate the images are list below in *Table 1*.

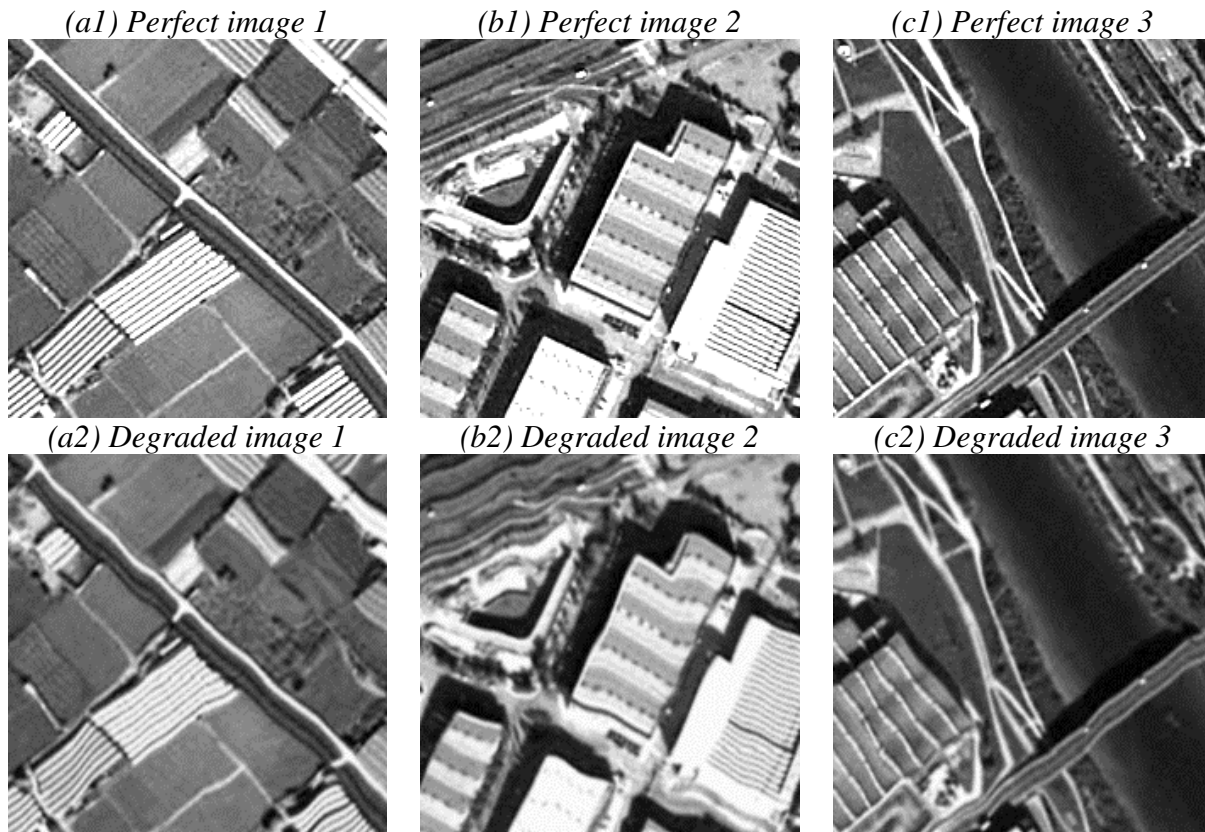


Fig. 3: Simulation results of a set of images (256×256). The information of scenes are distorted, blurred and contaminated by noise (less noticeable) after imaging.

Table 1: Degradation parameters of degraded image No.1, No.2 and No.3

No.	A'	$1/T'$	α	σ_n
1	0.5	1/32	1×10^{-4}	0.25
2	1	1/16	2×10^{-4}	0.5
3	2	1/8	4×10^{-4}	1

Table 2: Evaluation results of images of different scenes, with different extents of degradation. The SSIMs between restoration results u' and perfect images u are generally above 0.98 and much higher than the one between degraded image g and perfect images u , which indicates a good performance of the proposed algorithm.

Scene	Degradation extent	SSIM(u, g)	SSIM(u, u')	Iteration times	Time (s)
Scene 1	No. 1	0.930	0.994	10	5.326
	No. 2	0.824	0.988	12	5.876
	No. 3	0.736	0.984	15	6.965
Scene 2	No. 1	0.939	0.997	11	5.617
	No. 2	0.910	0.995	13	6.140
	No. 3	0.809	0.991	13	6.563
Scene 3	No. 1	0.935	0.997	10	4.890
	No. 2	0.866	0.996	10	4.908
	No. 3	0.745	0.992	15	6.271

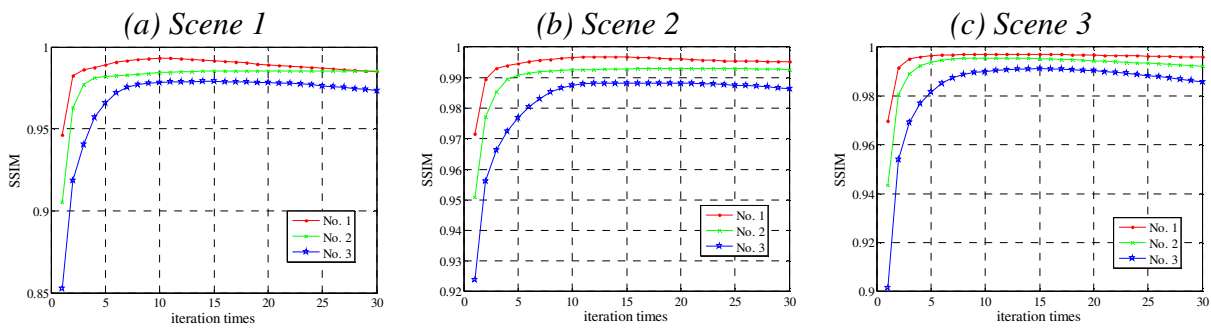


Fig. 4: The relationship between $SSIM(u, u')$ and iteration times. For images (256×256) of different scenes and different extents of degradation, the proposed algorithm takes 10~20 times of iteration (within 10 seconds) to approach the optimal results.

For visual clarity, the following figure compares a group of restoration results with the corresponding degraded images (with the degradation extent of No. 2). Note that the edges of images are truncated here due to the imbalance of sampling positions between vibration and non-vibration cases (which results in the absence of information when interpolating edge points).

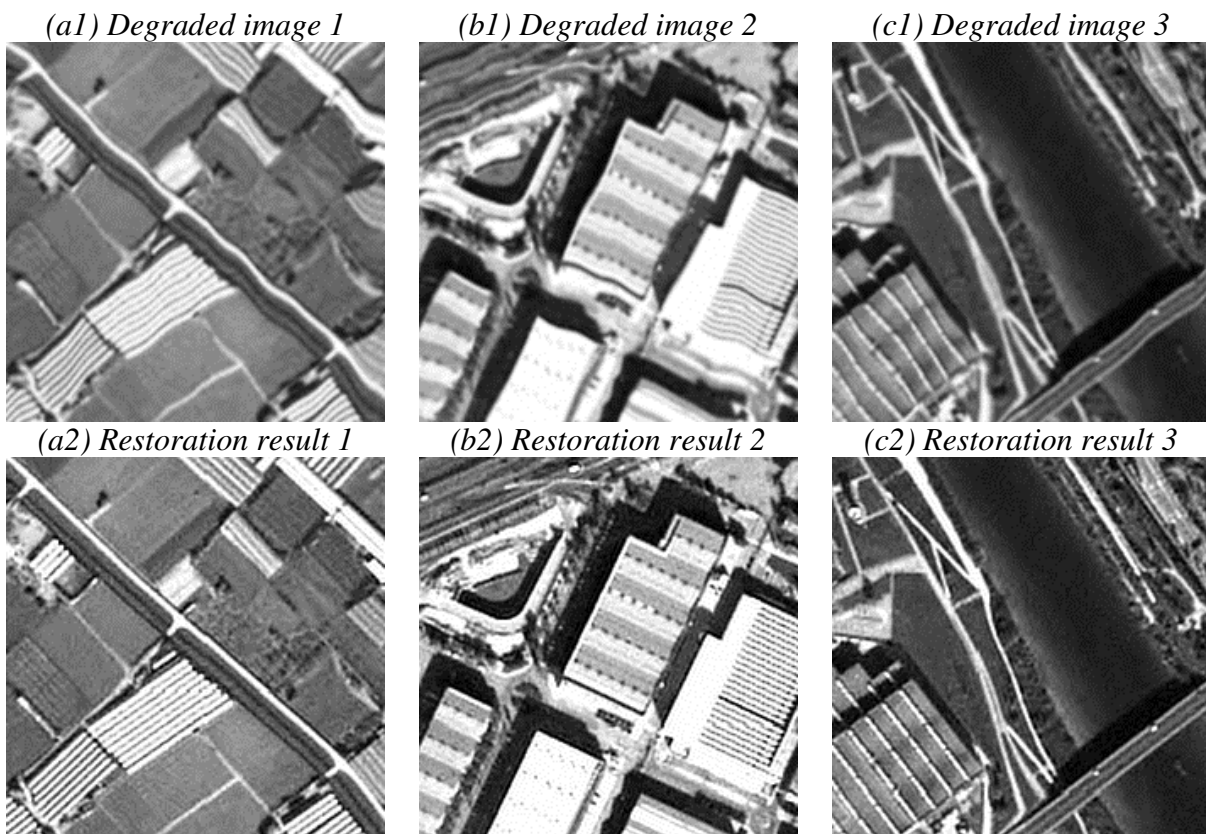


Fig. 5: Degraded mages (with the degradation extent of No.2) and the corresponding restoration results. Geometric distortion and blur in the degraded images is significantly reduced after restoration.

It can be concluded from the above cases that, for different categories of images (i.e. farmland, urban area and river) of different extent of degradation, the proposed algorithm takes 10~20 times of iteration to approximate the optimal solution and can restore the degraded images (256×256) within 10 seconds. Geometric distortion and blur in the degraded images are significantly reduced, and $SSIM(u, u')$, the SSIM between perfect and restored images, is generally above 0.98 – much higher than that of the degraded images. These indicate that the

algorithm can restore high resolution pushbroom imagery effectively with moderate time consumption, and is broadly applicable to different categories of land scenes and to different extents of degradation.

V. Conclusion

Satellite vibrations can cause displacement of pushbroom sampling and bring geometric distortion and blur to the acquired images. These effects not only affect subsequent image processing (e.g. registration), but also become more noticeable as image spatial resolution and TDI level increasing. In order to reduce image degradation and achieve high quality imagery without raising the requirement of precise control, the authors studied satellite vibrations and their effects. An imaging model for high resolution pushbroom imagery which considered vibration effects has been built, and then a new restoration algorithm has been proposed based on the model. Simulation results in Section IV illustrate that this algorithm can restore 256×256 images within 10 seconds, and SSIM of the restoration results are higher than 0.98.

Acknowledgments

The authors would like to acknowledge Dr. Ming Zhao, the School of Information Science and Technology, Dalian Ocean University, and Dr. Xiyang Zhi, the Research Centre of Space Optical Engineering (RCSOE), Harbin Institute of Technology (HIT), for their support at the beginning of this research. The authors would also like to thank the people in the RCSOE, HIT, who offered their advice and comments for this research.

References

1. Sudey Jr, J. and Schulman, J. R., "In-orbit Measurements of Landsat-4 Thematic Mapper Dynamic Disturbances", *Acta Astronautica*, Vol. 12, No. 7, 1985, pp. 485-503.
2. Wittig, M. E., Van Holtz, L., Tunbridge, D. E. L., et al., "In-orbit Measurements of Microaccelerations of ESA's Communication Satellite OLYMPUS", *OE/LASE'90*, Los Angeles, CA., US, Jan. 14-19, 1990, pp. 205-214.
3. Toyoshima, M, and Kenichi A., "In-orbit Measurements of Short Term Attitude and Vibrational Environment on the Engineering Test Satellite VI using Laser Communication Equipment", *Optical Engineering*, Vol. 40, No.5, 2001, pp. 827-832.
4. Toyoshima, M., Takayama, Y., Kunimori H., et al., "In-orbit Measurements of Spacecraft Microvibrations for Satellite Laser Communication Links", *Optical Engineering*, Vol. 49, No. 8, 2010, pp. 083604-083604-10.
5. Neeck, S. P., Venator, T. J. and Bolek, J. T., "Jitter and Stability Calculation for the ASTER Instrument", *Proc. SPIE 2317, Platforms and Systems*, 1994, pp. 70-80.
6. Lee, S. W., Skulsky, E. D., et al., "Mars Reconnaissance Orbiter Design Approach for High-resolution Surface Imaging", *26th Annual AAS Guidance and Control*, Breckenridge, CO, 2003, AAS 03-067.
7. Lieber, M., "Space-based Optical System Performance Evaluation with Integrated Modeling Tools", *Proc. SPIE 2317*, 2004, pp. 85-96.
8. Som. S. C., "Analysis of the Effect of Linear Smear on Photographic Images", *Journal of the Optical Society of America*, Vol. 61, No. 7, 1971, pp. 859-864.
9. Hadar, O., Fisher, M., Kopeika, N. S., "Image Resolution Limits Resulting from Mechanical Vibrations. Part III: Numerical Calculation of Modulation Transfer Function", *Optical Engineering*, Vol.31, No. 3, 1992, 581-589.
10. Hadar, O., Dror, I., Kopeika, N. S., "Numerical Calculation of Image Motion and Vibration Modulation Transfer Functions - A New Method", *Proc. SPIE 1533*,

Optomechanics and Dimensional Stability, San Diego, CA., US, 1991, pp. 61-74.

11. Tchernykh, V., Dyblenko, S., Janschek, K., et al., "Airborne Test Results for Smart Pushbroom Imaging System with Optoelectronic Image Correction", *Proc. SPIE 5234, Sensors, Systems, and Next-Generation Satellites VII*, 2004, pp. 550-559.
12. Janschek, K., Tchernykh, V. and Dyblenko, S., "SMARTSCAN-Smart Pushbroom Imaging System for Shaky Space Platforms", *20th Annual AIAA/USU Conference on Small Satellites*, 2006, SSC06-VI-3.
13. Gröchenig, K. and Strohmer, T., "Numerical and Theoretical Aspects of Nonuniform Sampling of Band-limited Images", *Nonuniform Sampling*, Spring, US, 2001, 283-324.
14. Almansa, A., Caselles, V., Haro, G., et al., "Restoration and Zoom of Irregularly Sampled, Blurred, and Noisy Images by Accurate Total Variation Minimization with Local Constraints", *Multiscale Modeling & Simulation*, Vol. 5, No. 1, 2006, pp. 235-272.
15. Facciolo, G., Almansa, A., Aujol, J. F., et al., "Irregular to Regular Sampling, Denoising, and Deconvolution", *Multiscale Modeling & Simulation*, Vol. 7, No. 4, 2009, 1574-1608.
16. Tan, W., *Spacecraft Systems Engineering*, China Science and Technology Press, Beijing, China, 2009, pp. 63-65.
17. Huo, H., Ma, M., Li, Y., et al., "High Precision Measurement Technology of Satellite's Angle Microvibration", *Transducer and Microsystem Technologies*, Vol.30, No. 3, 2011, pp. 4-9.
18. Zhang, X., *The Effect Analysis and Simulation of Platform Motion on Image Quality of Spaceborne TDICCD*, Master Thesis, Harbin Institute of Technology, China, 2011.
19. Keys, R., "Cubic Convolution Interpolation for Digital Image Processing", *Acoustics, Speech and Signal Processing*, Vol. 29, No. 6, 1981, pp. 1153-1160.
20. Wang, Q., Ward, R. K., "A New Orientation-Adaptive Interpolation Method", *IEEE Transactions on Image Processing*, Vol. 16, No. 4, 2007, pp. 889-900.
21. Wang, Z., Bovik, A. C., Sheikh, H. R., "Image Quality Assessment: From Error Visibility to Structural Similarity", *Image Processing*, Vol. 13, No. 4, 2004, pp. 600-612.

Phase Unwrapping for Interferometric Synthetic Aperture Radar Technique

Youtian Liu ¹, Linlin Ge ¹, Xiaojing Li ¹ and Jicang Wu ²

¹ School of Civil and Environmental Engineering, UNSW, Sydney, NSW, 2052, Australia

² College of Surveying and Geo-Informatics, Tongji University, Shanghai, 20092, China

Summary: Radar Interferometry provides height information on earth topography at meters level, monitors earthquake deformation, volcano activity and so on at centimetric level from phase information. As the continuous phase information is sampled in a discrete phase from $-\pi$ to $+\pi$, which is called wrapped phase, phase unwrapping turns to be a crucial step for rebuilding the DEM and analysing the deformation data. This paper gives an overview on major algorithms for phase unwrapping problem. There are plentiful phase unwrapping algorithms from many aspects: (1) Branch cut, (2) Minimum-norm methods, (3) Statistical cost flow networks and so on. This paper groups these algorithms into local and global algorithms. Besides the simple review, a new supplementary algorithm based on branch cut has been researched, which can put a more detailed judgment for unwrapping phase around residues and try to find better route for unwrapping.

Keywords: Radar Interferometry, phase unwrapping

1. Introduction

Radar provides two-dimensional image independent of sunlight, and is able to penetrate the cloud cover. Synthetic Aperture Radar (SAR) has higher resolution than Real Aperture Radar, and SAR image data contain important information about amplitude and phase of the backscattered signal from the imaged objects while the height of the object is unknown. By using the differences in the phase of two SAR images through the geodetic method called interferometric SAR (InSAR) a digital elevation model (DEM) can be generated [1]. With external DEM (Digital Elevation Model) or more SAR images, we can utilise differential InSAR (DInSAR) to obtain surface variation or deformation. We call these methods as Radar Interferometry and the image is called interferogram.

The measurement of digital elevation or deformation is thus obtained from the interferogram which is measured in radians of phase difference and due to the cyclic nature of phase and it is recorded as repeating fringes from $-\pi$ to $+\pi$. Once the basic interferogram has been produced, the consecutive fringes present in it will have to be unwrapped for most quantitative applications. We call this step as phase unwrapping which involves interpolating over 0 to 2π phase jumps to produce a continuous topography or deformation field. After phase unwrapping, we can utilise the unwrapped phase data and compute it into height or deformation.

Nowadays there are many significant developments in understanding interferogram phase wrapping problem theoretically and practically but not completely [2]. There are many papers with numerous algorithms for the solutions of phase unwrapping aiming to obtain a satisfying result, but it seems none of them gives a perfect solution.

In this paper, we do not intend to cover all the existing algorithms but present an overview to these major algorithms that are popular in use, such as Branch cut which developed by

Goldstein [1], the two-dimensional phase unwrapping by Ghiglia and Pritt [3], and SNAPHU (Statistical-cost Network-flow Approaches to Two-dimensional Phase Unwrapping) developed by Chen and Zebker [4].

2. Phase unwrapping theory and development

Topography is directly related with the resultant phase of the interferogram and the phase of the radar signal is a number of cycles of oscillation that the wave travels between the radar and the surface and back again. However, the phase detected by SAR sensors is not absolute phase, but relative phase, and the phase data modulo 2π , called wrapped phase from $[-\pi, \pi)$. Phase unwrapping aims to recover the underlying continuous phase information from the discrete wrapped phase that has been sampled, as the figure 1 shown below:

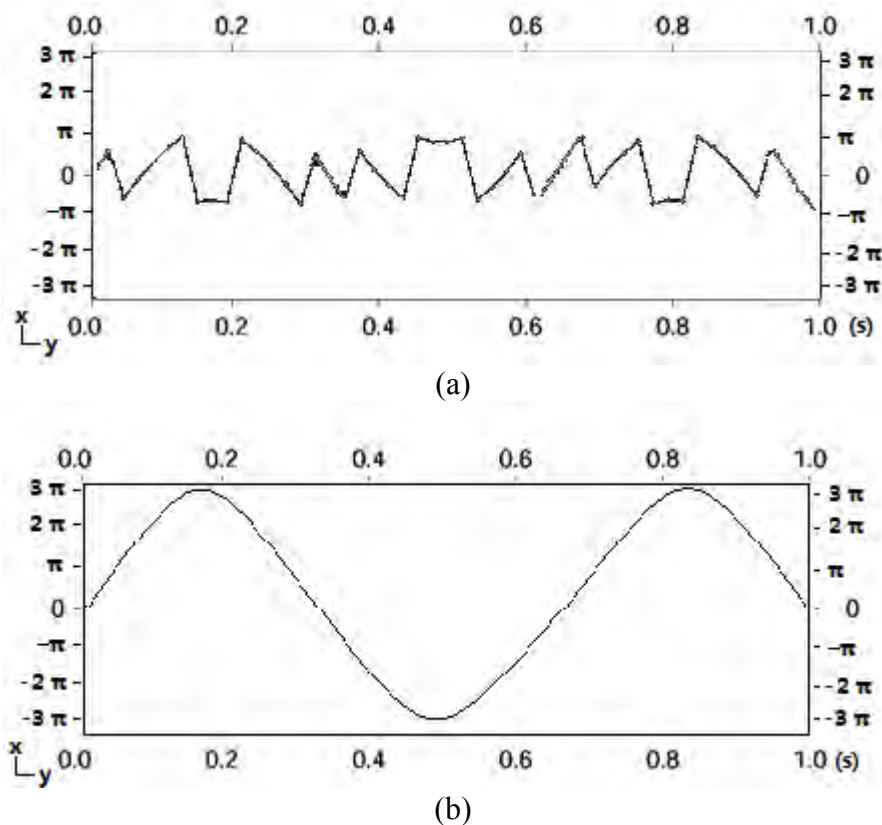


Fig. 1: Wrapped Phase (a) and Original Phase (b) examples, where x is the amplitude of phase and y is time

The actual phase images contain noise and discontinuities which can cause fringes to merge together, separate into segments, or generate isolated segments, for example, the geometric displacement such as layover and shadow can cause the fringes merge or isolated, and those places are called residues by Goldstein et al. and discontinuity sources by Huntley [5].

All the popular phase unwrapping algorithms are based on the assumption that the complex interferogram spatial sampling rate is sufficient to avoid aliasing in most area, which is defined as the Nyquist frequency. Theoretically in most places, the unwrapped adjacent pixel phase gradients are assumed to equal their wrapped phase counterparts, where the latter are less than one-half cycle (π rad) in magnitude [4]. If this assumption were true everywhere, then phase unwrapping could be integrated from arbitrary path by the wrapped phase to recover the original phase. However, the inherent radar speckle and geometric displacement

can cause discontinuities, which are simply displayed as unwrapped gradients do exceed π . These discontinuities make the unwrapping path dependent on the choice of integration paths of wrapped gradients.

To overcome this impact, in 1988 Goldstein et al. [1] applied branch cut algorithm related with residue to prohibit any integration across the branch, ensuring the conservation of integration. After that, many researchers develop new algorithms. Ghiglia and Romero [6] in 1994 used preconditioned conjugate gradient (PCG) for weighted least squares phase unwrapping. Full multi-grid algorithm for partial differential equation was invented in 1996 by Pritt [7], which was claimed faster than PCG method. Quality map was introduced in 1996 for guiding the unwrapping path involving residues [8]. Later on Flynn [5] in 1997 applied minimum discontinuity algorithm focusing on how to solve the discontinuities and wrapped phase, this algorithm is regarded as an early minimum cost flow algorithm [9]. A good basic understanding of phase unwrapping theory and principles book is directed to Ghiglia and Pritt [3], the *Two-dimensional phase unwrapping for radar interferometry: Developments and new challenges* in 1998, as the book contains excellent introductions of concepts for two dimensional phase unwrapping methods. Besides, in this book Ghiglia and Pritt [3] introduced mask-cut algorithm in the book for comparing with quality map. Masks provide a good type of weighting factor. After that, in 2000 the SNAPHU [4] algorithm based on minimum spanning tree and residues gives a relatively satisfying result and this algorithm is widely used in many softwares such as Doris. Nowadays there are various algorithms and models on phase unwrapping, the following is a glimpse at those algorithms:

Table 1: Some recent phase unwrapping algorithms

	Time	Algorithm idea
1	2008, M. Lachaise, T. Fritz et al.[10]	A new dual baseline phase unwrapping algorithm (MCF related)
2	2008, Xiaowei Li and Xiang-Gen Xia. [11]	Robust Chinese Remainder Theorem (CRT)
3	2009, Juan J. Martinez-Espla, T. Martinez-Martin et al. [12]	A Particle Filter Approach (Extended Kalman filter)
4	2010, Heping Zhong, Jinsong Tang et al. [13]	Based on Minimum Discontinuity by Blocking
5	2011, A. Pepe, L. D. Euillades et al. [14]	Extended Minimum Cost Flow Algorithm
6	2011, Hanwen Yu, Zhenfang Li et al. [15]	A Cluster-Analysis-Based Efficient Multibaseline Algorithm
7	2011, Kui Zhang, Linlin Ge et al. [16]	Phase Unwrapping for Very Large Interferometric Data Sets
8	2012, Davide Chirico and Gilda Schirinzi [17]	Nonlinear Kalman Smoother
9	2013, Heping Zhong, Jinsong Tang et al. [18]	Quality-Guided and Local Minimum Discontinuity

3. Basic Phase Unwrapping Methods

3.1. Local phase unwrapping algorithms

1) *Branch Cut*

Theoretically, if the radar image was sufficiently sampled, arbitrary integrating path should get the same result. However, the inherent speckles in radar image may lead residues. As mentioned above, residues make the integration different by choosing different unwrapping paths. Branch cut then is introduced by Goldstein et al. [1] in 1988, which is a classical path-following method, and this algorithm raised the concept of residue. Any clockwise 2×2 adjacent gradients summation should be zero, while in fact they are not, some turn out to be -1 or +1 which are in different charge, and this is the definition of residues. As shown below in figure 2:

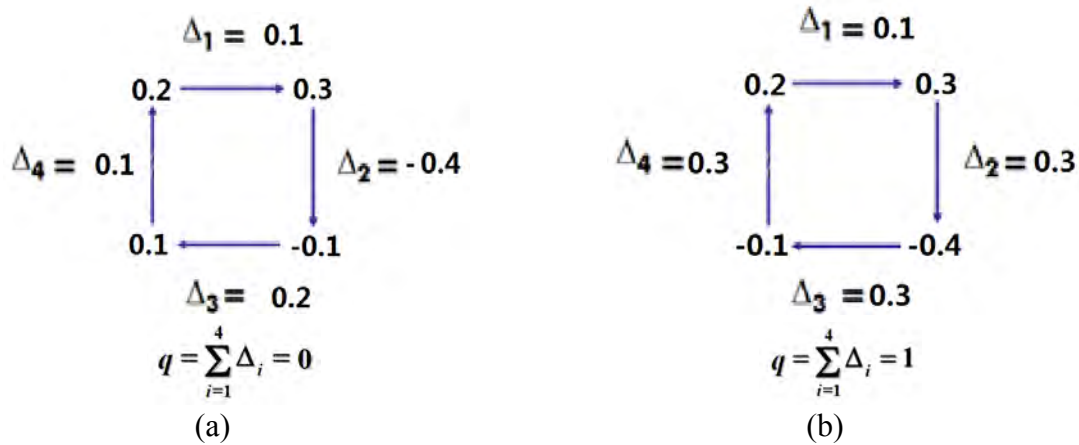


Fig. 2: A sufficient sample (a) and a possible residue (b) (the value is divided by 2π)

After calculating all the residues and record their charges, this algorithm connects the residues of different charges which are called branch cuts, and makes the total charge in the branch is zero, then prohibits any integration paths through the branch to make sure the result conservative. This procedure is depicted in figure 3:

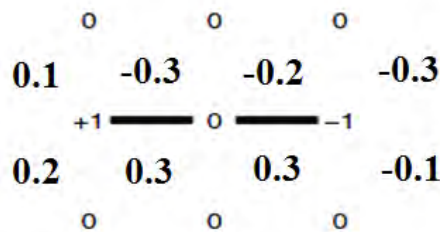


Fig. 3: Connection of branch cuts from different charges

Basic branch cut procedure follows the steps: the left loop has a charge of +1 while the right loop is -1, then the aim is to connect them to build a branch and cut this area into different parts. After that, the integration paths now cannot pass the branch, for example, it can go from 0.1 to -0.3 then -0.2, but not to 0.3. The advantages of this algorithm are that it is mathematically easy and computational fast as it costs low memory. The disadvantage of this algorithm is the lack of weighting factor for guiding the placement and connection of branch cuts. This would result in block or the called island when unwrapping at areas of low coherence and where densely branch cuts are put. However, the island caused by cuts appears to be caused by the way they are represented internally.

Many new researches are applied to enhance branch cut, such as using quality map [8], minimum the spanning tree [19] and so on.

2) Quality map guided phase unwrapping

The phase of the correlation forms the data to be unwrapped, while its magnitude indicates the quality of the phase data, ranging from 0 (poor) to 1 (perfect), this is a quality map [8], which is also called coherence in radar interferometry. As interferometry with a quality map is available, it indicates the reliability of the measurements. Flynn used the quality map for designing the mask at low quality area, which is to grow the residues into areas where the quality is below a threshold.

The constraint is to form the residues into clusters with zero charge summation and requires the path enclose either all or none of the residues in each cluster. This algorithm searches global for the best set of clusters. The steps for quality map guided phase unwrapping are firstly the regions with very high quality pixels are seeded, and then low quality areas grow mask, after that the algorithm tracks the residues for unwrapping.

However, the threshold for high quality is not standard and this algorithm ignores the detection of discontinuities. A coarse quality map can definitely block the unwrapping procedure, as low coherence areas will be put more discontinuities and phase unwrapping cannot pass that area.

3) *Flynn's minimum discontinuity algorithm*

Different from utilising residues for putting branch cuts and trying to minimize the length of branch cut, Flynn uses the concept of discontinuity which was noticed by Lin et al. [20] along the fringe lines, and regards phase unwrapping as a process of adding a multiple of 2π to the phase difference to minimize the discontinuities across the lines. The expense of this algorithm comparing with branch cuts is higher memory and computational requirement. However, nowadays the computer technology is sufficient enough for this algorithm.

The concept of discontinuity, or jump, is defined as where the difference of two adjacent unwrapped phases is over than one-half cycle, and Flynn finds the relationship among the wrapped phase, discontinuities and unwrapped phase. The discontinuity can be removed by adding a multiple of 2π to the phase gradients and this multiple is defined as jump count for the pair of phase gradient. Jump count includes vertical and horizontal jump counts. That is, if we got the vertical jump count as the following formula:

$$v_{mn} = c_{mn} - c_{m-1,n} + \left\lceil \frac{\phi_{mn} - \phi_{m-1,n} + \pi}{2\pi} \right\rceil \quad \text{For } (m,n) \in V \quad (1)$$

Where v is vertical jump counts, c is integer array, called the wrap-count array, is calculated from the phase image ϕ . $V = \{1, 2, \dots, m-1\} \times \{0, 1, \dots, n-1\}$ and $[X]$ is the largest integer less than or equal to X . And horizontal jump counts:

$$z_{mn} = c_{mn} - c_{m,n-1} + \left\lceil \frac{\phi_{mn} - \phi_{m,n-1} + \pi}{2\pi} \right\rceil \quad \text{For } (m,n) \in V \quad (2)$$

Where z is the horizontal jump counts, c is described as above.

Then the wrap-count E equal to the total jump counts:

$$E(c; \phi) = \sum_{(m,n) \in V} w_{mn}^v |v_{mn}| + \sum_{(m,n) \in V} w_{mn}^z |z_{mn}| \quad (3)$$

Where w_{mn}^v and w_{mn}^z are arrays of nonnegative integer weights derived from the quality information.

So the wrapped phase gets unwrapped:

$$\hat{\phi}_{mn} = \phi_{mn} + 2\pi c_{mn} \quad (4)$$

The pros for this algorithm are that it removes fringes quickly, and if possible, the weight chosen may not divide the image into islands by zero-weight. However, the threshold is not always easy to choose, and at low coherence area there may lead large trees of jumps. The imagery algorithm is depicted in figure 4:

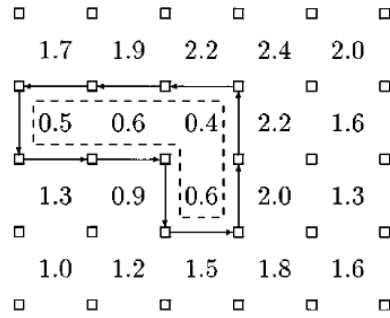


Fig. 4: Jumps and discontinuity connection (from [5])

This minimum weighted discontinuity (MWD) algorithm is regarded as an early research on minimum cost flow algorithm, but this algorithm does not consider residues, and as it does not require quality map, then weighting factors is optional.

4) Minimum Cost Flow algorithms

Minimum cost network flow algorithm framework is invented by Costantini [9], though Flynn attributed earlier, and the algorithm is combined with residues and similar MWD algorithm, connecting positive and negative residues then focusing on minimizing the object. Under Costantini's framework, the phase unwrapping problem itself turns out to be a general network flow problem, allowing the interchangeable use of fast existing network optimization paths as well as many other ideas in the rich and well-developed area of network theory [19].

This algorithm is entirely different to put branch cuts between residues, as it uses the fact that phase differences of neighbouring pixels can be estimated with possibly an error that an integer multiple of 2π . Then it defines the discrete derivative residuals looking for the solution of the following minimization problem:

$$k1 = \frac{|\Delta\phi_{(i,j)}^1 - \Delta\psi_{(i,j)}^1|}{2\pi} \quad (5)$$

$$k2 = \frac{|\Delta\phi_{(i,j)}^2 - \Delta\psi_{(i,j)}^2|}{2\pi} \quad (6)$$

$$J = \min_{\{k1, k2\}} \sum_{i=0}^{N-2} \sum_{j=0}^{M-1} c_{(i,j)}^1 k1 + \sum_{i=0}^{N-1} \sum_{j=0}^{M-2} c_{(i,j)}^2 k2 \quad (7)$$

Where $k1$ and $k2$ are the residuals, while the nonzero discrete derivative residuals identify the branch cuts. MCF algorithms can be more efficient for dealing with large data images, which can be referred to Zhang et al. [16]. And Marie et al. [10] use MCF for TanDEM-X mission and get good results.

5) Minimum Spanning Tree (MST) algorithm

Minimum spanning tree approximates a minimum Steiner tree, aiming to connect the residues at least expense for connections. This algorithm proposed by Curtis W. Chen and Zebker [19], which adapts to branch cuts and approximates minimizing the length of

spanning tree. By containing all the charges of residues to the nearest to draw a neutral tree, it uses Dijkstra's [21] shortest path algorithm for searching. With the user-defined weights on phase gradients, it is possible guiding the placement of tree branches.

Based on MST, SNAPHU algorithm was developed and put in large SAR data unwrapping, and as mentioned above, it is used in Doris and so on for phase unwrapping.

6) *Path-following algorithm conclusion*

There are various path-following phase unwrapping algorithms, and in general, these algorithms are local phase unwrapping methods. Ghiglia and Romero[22] suggested the minimum L^p – norm frame work, where L is the difference between a pair of corresponding wrapped and unwrapped phase gradients, and p is a power to which L is raised in order to establish an error metric. Goldstein's branch cut algorithm can be regarded as L^0 algorithm while MCF is regarded as L^1 solutions.

Other path-following methods such as mask cut [3] also make contributions to unwrapping problem. Local algorithms focus on typical areas but where the noise is too much local unwrapping may be blocked. Then Global phase unwrapping gives better performance.

3.2. Global phase unwrapping algorithms

1) *Least squares method*

As described above, the minimum L^p norm algorithms have the power of p , then least squares can be regarded as when $p=2$, that is L^2 norm situation. This is different with L^0 and L^1 methods, as this involves all the values to the partial differential equations (PDEs) using mathematical methods for minimization. L^2 norm turns the PDEs to a linear equation for solution, and minimizing the gradient differences.

Least squares method is requested as weighted and unweighted algorithms, while unweighted phase unwrapping can be described as a discretised Poisson equation [23]. To solve this problem efficiently, fast Fourier transforms (FFTs) is used and also discrete cosine transform (DCTs) [6]. Based on Gauss-Seidel relaxation schemes, the unweighted full multi-grid (FMG) [7] is working iteratively and efficiently.

In the case of weighted least squares problem, the preconditioned conjugate gradient (PCG) method [6] is applied, while weighted FMG is claimed that 25 times faster than PCG. However, it does not guarantee to cluster and needs a mass of iterations.

The advantages of this algorithm are that as all the value can be added, it ensures the smoothness and continuity of the result; and at the low SNR or densely residue area, L^2 can be quite effective for anti-noise, permitting use on noisy data that would have been difficult or impossible to unwrap by path-following algorithms.

2) *Minimum L^p norm methods*

Minimum L^p norm algorithm is a generalization of the weighted L^2 phase unwrapping approach, while the optimization equation is:

$$\text{minimize}\{\sum_{i,j} w_{i,j}^x |\Delta\phi_{(i,j)}^x - \Delta\psi_{(i,j)}^x|^p + \sum_{i,j} w_{i,j}^y |\Delta\phi_{(i,j)}^y - \Delta\psi_{(i,j)}^y|^p\} \quad (8)$$

Where w is weight, $\Delta\phi_{(i,j)}^x - \Delta\psi_{(i,j)}^x$ is the gradient between corresponding unwrapped and wrapped phase, and p is power. This method generates data-dependent weights if $p \neq 2$, and some discussion about weight calculation is described by Pritt [7]. However, this algorithm is highly intensive for computation due to its doubly iterative structure [1].

3) Other global algorithms

FFTs and DCTs techniques are widely used to make the global phase unwrapping methods computationally efficient. It has been proved that the Green's formulation method is mathematically equivalent to the least squares algorithm [24]. PCG and FMG for weighted least squares are efficient, while Zebker and Lu [25] combine local and global approaches in a synthesis algorithm.

Global unwrapping algorithms have the problem that they may lead a shifting for all the data comparing with local phase unwrapping algorithms. But at noisy area global algorithms have better performance.

4. Our research for unwrapping

Considering residues and discontinuities, this algorithm aims to combine both branch cuts and discontinuities when doing phase unwrapping. For high dense noise area, there could be blocked by closed branch, the algorithm firstly connects the residues and then when unwraps the phase, it also considers the unwrapping path under the instruction of discontinuity lines.

We define the block lines by 2×2 pixels and mark a number to the left top pixel, these numbers are chosen at the basic five types, namely no block, left block, down block, right block, and up block. Then add their value together if it is combining situation, for example, down and right block it sets 8 as 3 plus 5. To avoid the same number stands for multi type, so up block is set to 10. The figure connecting the line is in the following:

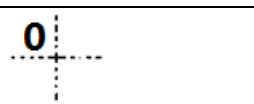


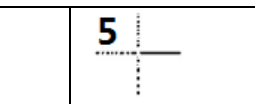
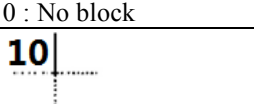
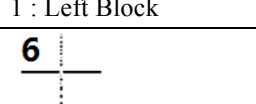
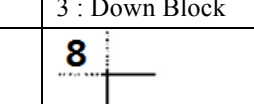
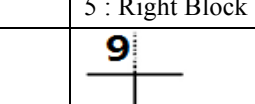


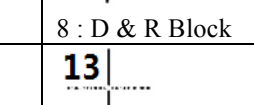
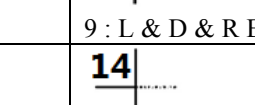
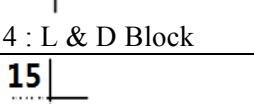

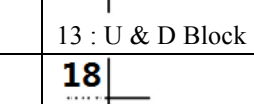
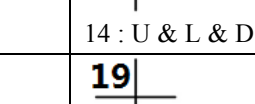
 0 : No block	 1 : Left Block	 3 : Down Block	 5 : Right Block
 10 : Up Block	 6 : L & R Block	 8 : D & R Block	 9 : L & D & R Block
 4 : L & D Block	 11 : U & L Block	 13 : U & D Block	 14 : U & L & D Block
 15 : R & U Block	 16 : R & U & D Block	 18 : D & L & U Block	 19 : All Block

Figure 5: All situations of discontinuities

L, R, U and D are short for left, right, up and downward block lines. The dotted line means that it allows the integration pass while the solid line prohibits the integration. This somewhat like mask and minimum weighted discontinuity, but it does not consider coherence and jump counts. This algorithm tries to find better way to break through the branch cut when the cuts come to close as the way connecting cuts are different. However, this method do solve the island at the expense of time consuming and in real data if the image is separated by heavy

layover or river then there may lead error propagation. Figure 6 shows the simulated topography results:

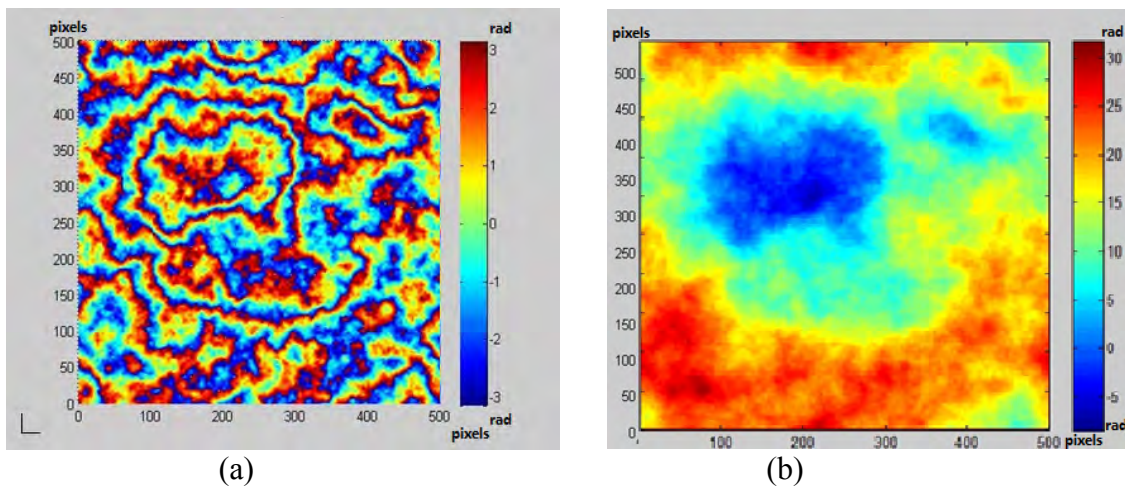


Figure 6: Simulated wrapped interferogram (a) and unwrapped by the algorithm (b)

5. Conclusion

There is variety of phase unwrapping methods and those algorithms mentioned above are concluded as two-dimensional phase unwrapping problem. It is complex and difficult to obtain completely correct answer. Two-dimensional phase unwrapping method then remains to accuracy and efficiency problem. As this turns to be a NP hard problem, it is impossible for efficient algorithms to solve exactly [4]. Some researchers believe that L^0 or L^1 norm such as MCF is better for phase unwrapping [4] while other believe that minimum L^2 norm can give more exact answer at high noise area [26, 27].

Another challenge for any algorithm is its ability to deal with large data. Although the computer technology is developing rapidly and more memory is becoming available; the data from new satellites are also becoming larger and larger. As mentioned above, MCF and L^p norm can handle this problem, while L^2 may cause shifting at some area. Chen and Zebker [4] have applied SNAPHU for large data, and Zhang et al. [16] processed large data using MCF.

Further, InSAR and DInSAR images need to incorporate the temporal and geometrical decorrelation to form interferometric fringe. The algorithms discussed in this paper are conventional spatial unwrapping algorithms without long time series. For high decorrelation images, such as in processing persistent scattering InSAR data, the algorithms should consider the temporal impacts [28, 29]. One example algorithm is STUN algorithm (Spatial-Temporal Unwrapping Network) [30].

Combining these phase unwrapping methods and balancing their disadvantages make new algorithms produce better and satisfying results. For example, Zebker and Lu [25] applied synthesis algorithm using both residues and least squares. Dai and Zha [31] combined reliability sorting and residue mask to get accurate phase unwrapping. Many ongoing researches on optimal weighting factors are aiming to find the most appropriate threshold. Even though it is always approximation to the whole problem, there are algorithms offering acceptable result and they should be further enhanced in both efficiency and precision.

6. Acknowledgement

The authors would like to thank the two anonymous reviewers for their constructive comments and suggestions. Their valuable comments make this paper more completed.

7. References

1. Goldstein R. M., H. A. Zebker, and W. C. L., "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Science*, vol. 23, 1988, pp. 713-720.
2. R. Gens, "Two-dimensional phase unwrapping for radar interferometry: Developments and new challenges," *International Journal of Remote Sensing*, vol. 24, 2003, pp.703-710.
3. D. C. Ghiglia and M. D. Pritt, *Two-dimensional phase unwrapping: theory, algorithms and software*. New York: John Wiley, 1998.
4. C. W. Chen and H. A. Zebker, "Phase unwrapping for large SAR interferograms: statistical segmentation and generalized network models," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, 2002, pp. 1709-1719.
5. T. J. Flynn, "Two-dimensional phase unwrapping with minimum weighted discontinuity," *Journal of the Optical Society of America A*, vol. 14, 1997, pp. 2692-2701.
6. D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *Journal of the Optical Society of America A*, vol. 11, 1994, pp. 107-117.
7. M. D. Pritt, "Phase unwrapping by means of multigrid techniques for interferometric SAR," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 34, 1996, pp. 728-738.
8. T. Flynn, "Consistent 2-D phase unwrapping guided by a quality map," in *Geoscience and Remote Sensing Symposium, 1996. IGARSS '96. 'Remote Sensing for a Sustainable Future.'*, *International*, vol.4, 1996, pp. 2057-2059.
9. M. Costantini, "A novel phase unwrapping method based on network programming," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 36, 1998, pp. 813-821.
10. M. Lachaise, T. Fritz, and M. Eineder, "A new dual baseline phase unwrapping algorithm for the TanDEM-X Mission," in *Synthetic Aperture Radar (EUSAR), 2008 7th European Conference on*, 2008, pp. 1-4.
11. L. Xiaowei and X.-G. Xia, "A Fast Robust Chinese Remainder Theorem Based Phase Unwrapping Algorithm," *Signal Processing Letters, IEEE*, vol. 15, 2008, pp.665-668.
12. J. J. Martinez-Espla, T. Martinez-Marin, and J. M. Lopez-Sanchez, "A Particle Filter Approach for InSAR Phase Filtering and Unwrapping," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, 2009, pp. 1197-1211.
13. Z. Heping, T. Jinsong, and L. Dandan, "A fast phase unwrapping algorithm based on minimum discontinuity by blocking," in *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, 2010, pp. V1-717-V1-721.
14. A. Pepe, L. D. Euillades, M. Manunta, and R. Lanari, "New Advances of the Extended Minimum Cost Flow Phase Unwrapping Algorithm for SBAS-DInSAR Analysis at Full Spatial Resolution," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, 2011, pp. 4062-4079.
15. Y. Hanwen, L. Zhenfang, and B. Zheng, "A Cluster-Analysis-Based Efficient Multibaseline Phase-Unwrapping Algorithm," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, 2011, pp. 478-487.

16. Z. Kui, G. Linlin, H. Zhe, A. H. Ng, L. Xiaojing, and C. Rizos, "Phase Unwrapping for Very Large Interferometric Data Sets," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, 2011, pp. 4048-4061.
17. D. Chirico and G. Schirinzi, "InSAR phase unwrapping using nonlinear Kalman smoother," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 2012, pp. 5610-5613.
18. H. Zhong, J. Tang, S. Zhang, and X. Zhang, "A Quality-Guided and Local Minimum Discontinuity Based Phase Unwrapping Algorithm for InSAR/InSAS Interferograms," *Geoscience and Remote Sensing Letters, IEEE*, vol. PP, 2013, pp. 1-1.
19. C. W. Chen and H. A. Zebker, "Network approaches to two-dimensional phase unwrapping: intractability and two new algorithms," *Journal of the Optical Society of America A*, vol. 17, 2000, pp. 401-414.
20. Q. Lin, J. F. Vesecky, and H. A. Zebker, "New approaches in interferometric SAR data processing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 30, 1992, pp. 560-567.
21. R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, "Network flows: theory, algorithms, and applications," 1993.
22. D. C. Ghiglia and L. A. Romero, "Minimum Lp-norm two-dimensional phase unwrapping," *Journal of the Optical Society of America A*, vol. 13, 1996, pp. 1999-2013.
23. B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *Journal of the Optical Society of America*, vol. 69, 1979, pp. 393-399.
24. G. Fornaro, A. Pauciullo, and E. Sansosti, "Phase difference based multiple acquisition phase unwrapping," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol.2, 2003, pp. 948-950.
25. H. A. Zebker and Y. Lu, "Phase unwrapping algorithms for radar interferometry: residue-cut, least-squares, and synthesis algorithms," *Journal of the Optical Society of America A*, vol. 15, 1998, pp. 586-598.
26. M. A. Richards, "A Beginner's Guide to Interferometric SAR Concepts and Signal Processing [AESS Tutorial IV]," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 22, 2007, pp. 5-29.
27. L. C. Graham, "Synthetic interferometer radar for topographic mapping," *Proceedings of the IEEE*, vol. 62, 1974, pp. 763-768.
28. M. Costantini, F. Malvarosa, F. Minati, L. Pietranera, and G. Milillo, "A three-dimensional phase unwrapping algorithm for processing of multitemporal SAR interferometric measurements," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, vol.3, 2002, pp. 1741-1743.
29. H. Zebker, P. Shankar, and A. Hooper, "InSAR Remote Sensing Over Decorrelating Terrains: Persistent Scattering Methods," in *Radar Conference, 2007 IEEE*, 2007, pp. 717-722.
30. B. M. Kampes, *Radar interferometry: persistent scatterer technique* vol. 12: Springer, 2006.
31. D. Zhiyang and Z. Xianjie, "An Accurate Phase Unwrapping Algorithm Based on Reliability Sorting and Residue Mask," *Geoscience and Remote Sensing Letters, IEEE*, vol. 9, 2012, pp. 219-223.

Swarm-Based Satellite Constellation Control

John Page, Haoyang Cheng, Monica Chi, and Nathan Kinkaid

*School of Mechanical and Manufacturing Engineering, The University of New South Wales,
Sydney, NSW, 2052, Australia*

Summary: Over the last few years there has been a lot of research interest in space satellite swarms. An area of particular interest, to us, is highly autonomous swarms, referred to as self-organising swarms. This approach does not require any centralised command and control. The coordination is achieved through agent-level interactions. They are able to take action to ensure the most efficient use of fuel, avoid impacts and adapt to individual degradation and failure without earth based control. We are now investigating applying these capabilities for constellations of spacecraft. In this paper, two applications of satellite swarm are presented. The potentials of applying swarm technologies to the satellite constellations are discussed.

Keywords: Satellite Constellation, Bio-Inspired Swarm, Behaviour Rules.

Introduction

Many space applications such as global navigation and cell based space communication require a number of platforms to collaborate. Currently this collaboration requires sophisticated ground control stations usually situated around the earth and copious amounts of data transfer. This presents two limiting factors the expense of the ground stations, which if manned as they usually are is very costly and the system reliability. There are two factors inherent in this approach that limit the ruggedness of the system, the ground station itself is vulnerable to technical or malicious damage and the data is subject to interference and corruption. There is another aspect of data transfer that is becoming a global issue and that is the security of the data. Both nations and individuals have increasing concerns about repurposing of data by third parties. Due to the complexity of managing the system it is also often not possible to exploit the assets to their full capability.

Self-Organising Swarms

Swarm technology has its origin in biomimicry of social insects. As often observed, biological system such as the swarms of ants, bees, fish and birds reveal some amazing cooperative behaviour. These insects are able to carry out complex tasks which emerge from a very limited cognitive capability. Every single insect that lives in colonies seems to appear in the right place and do the right task so that the whole colony appears to have some global organisation which it does not. The continuous integration of all individual activities does not

seem to require any supervision [1]. Such systems are often referred as a Self-Organised (SO) system or swarm intelligent system. The terms 'SO system' and 'intelligent system' are used interchangeable in this thesis. After decades of research, the mechanism that drives SO systems has been identified. Recent research suggest that the individuals within the swarm system are neither able to assess a global situation nor aware of the tasks to be carried out by the other agents and thus no global control, i.e. there is no supervisor in a swarm system. Their actions are determined by sets of behaviour rules on the basis of local information. For example, each time an agent performs an action, the local environment is modified by this action. The modified environmental configuration will then influence the actions of the other agents.

Ants and termites which were first studied by the pioneers of this area of research carry out complex behaviour including major construction work their activity based on a simple rule set that can be easily replicated in a simulation. Honey bees have a far more complex rule set mainly due to their need to deal with a much richer sensory capability. A bee is able to fly out searching for a food source and then return to its nest and describe the location to its fellows. The flying out and return alone requires the bee to replay the visual information in the reverse order and compare it to data from another sensor, the other eye, this behaviour is complex and again carried out with limited cognitive and data transfer capability. Reynolds in particular was able to demonstrate that the flocking of birds or the shoaling of fish could be reproduced with just three simple rules. While later studies have shown these original rules required some refinement to match real swarms the principal holds true. The concept of SO system has been applied to the formation flying spacecraft [2, 3], in which the spacecraft, modelled as a swarm of agents, follow three biological rules, namely 'avoidance' of both each other and the threat, 'gather' to maintain the formation and 'attraction' towards target locations according to pre-defined artificial potential functions. In this study, the scenario is more complicated than converging to target locations. Therefore the agents will be required to switch between multiple modes and their behaviours are governed by different rule sets under each mode.

The adaption of these techniques to solving engineering problems is still in its infancy. We have investigated the search part of marine search-and-rescue operations [4], military ground attack [5], collision avoidance for orbiting spacecraft [6] and fuel utilisation for life maximisation of swarms of spacecraft [7]. The conclusions from this work is that SO swarms offer real advantages in terms of the response time and system ruggedness, but this is to some extent the cost of not optimising the assets available.

Control of Swarms in Orbit

A number of papers have been produced examining the viability of the use of swarms of spacecraft [8, 9]. These have mainly concentrated on the proposed advantages of distributing the assets and the cost savings that may be achieved from producing a number of small cheap and possibly replaceable spacecraft. There is much less exploration of how swarm technology

might be applied to these vehicles. For this reason we investigated two scenarios where a self organised swarm might yield benefit.

Debris Avoidance

In this scenario we looked at a swarm of spacecraft in the same orbit and following a line astern pattern. When debris, such as space junk, threatened one of the craft it automatically adjusted its location to avoid the impact. This led to the other spacecraft adjusting their positions to accommodate the translation. In Figure 1, the spacecraft avoids the debris region by changing their along-track position through three impulsive thrusts in the orbital plane, as indicated by the red dots, to their final configuration. The selected avoidance region in this simulation is arbitrarily taken to be one kilometre for scalability such that any changes could be easily accommodated within the methodology.

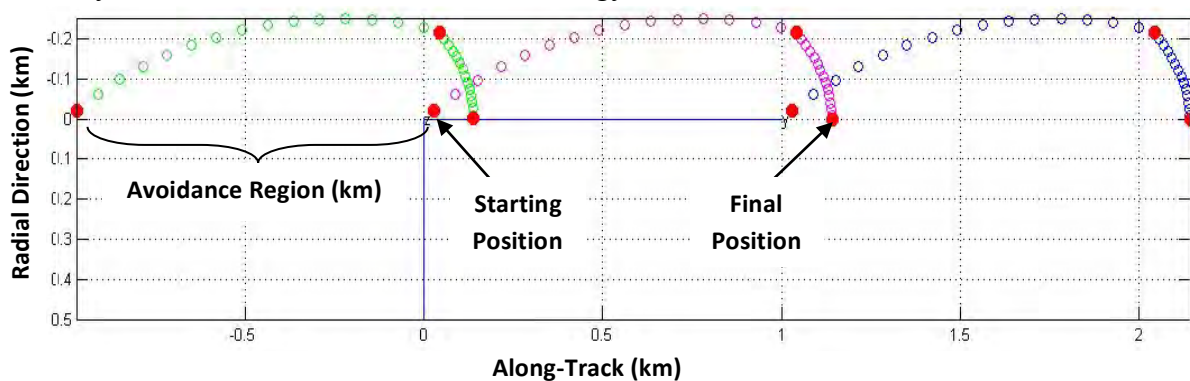


Figure 1 In-plane debris avoidance manoeuvre

This could be the basis for a classic swarm readjustment based on a changed environment. When the threatened vehicle moved it effectively changed the other vehicles in the swarm's environment which led to them modifying their location to accommodate the changed environment.

Mission Life Maximisation

As the useful life of a space system is dependent on the fuel available for manoeuvring, one of the weaknesses of using a swarm of satellites to carry out a mission is that the mission ceases or is at least degraded when the first vehicle runs out of fuel. Under the influence of the J_2 perturbation, the agents with larger orbital element differences with respect to the reference orbit consume more energy in maintaining their position within the configuration. This suggests that the fuel consumption of individual spacecraft will not be uniform across the swarm as it depends on initial conditions and the location of the individual spacecraft within the formation. To overcome this we produced an algorithm that led to the vehicle moving from a high fuel-consumption zone to a lower fuel consumption area by utilising swarm technology.

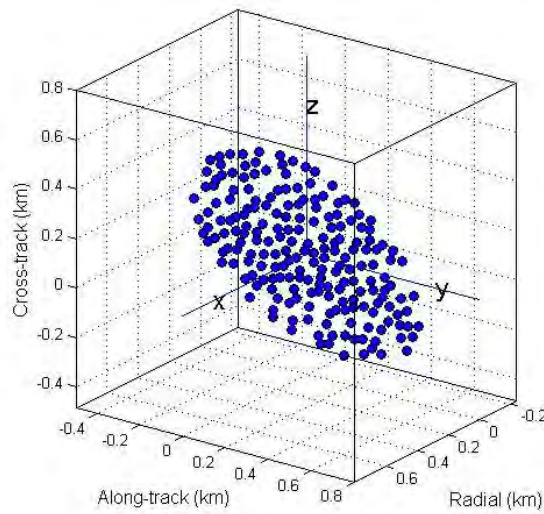


Figure 2(a): The initial position of the spacecraft swarm.

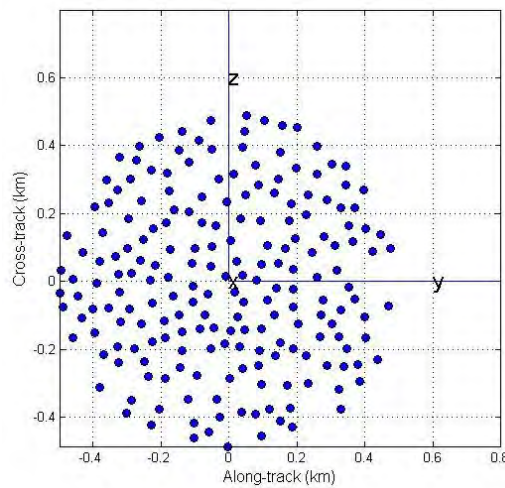


Figure 2 (b): The initial position of the spacecraft swarm on $y - z$ plane.

Figure 2 shows a swarm of satellites viewed in three dimensions and the distribution when viewed along the track. As can clearly be seen from figure 3 this orbital structure leads to a considerable distribution in fuel-usage rates. It is related to its orbital elements differences with respect to the reference orbit. If the coordination strategy is not applied to balance the fuel consumption, the mission lifetime will be affected as the spacecraft on the outer perimeter of the configuration will run out of fuel before those nearer to the centre.

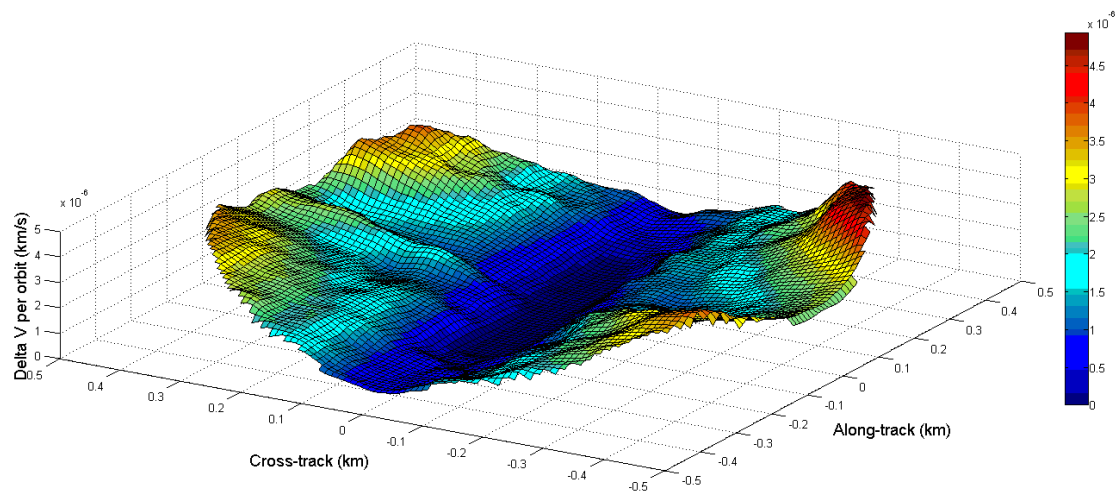


Figure 3: Fuel cost map.

The swarm algorithm allows the individual vehicles to modify their behaviour as shown in Figure 4. The common objective is to extend the mission lifetime by allowing the spacecraft in the high-fuel-consumption positions to switch with those in the low-fuel-consumption positions.

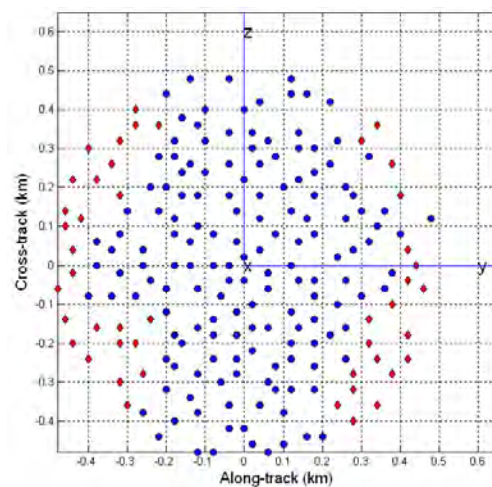


Figure 4(a): Initial swarm.

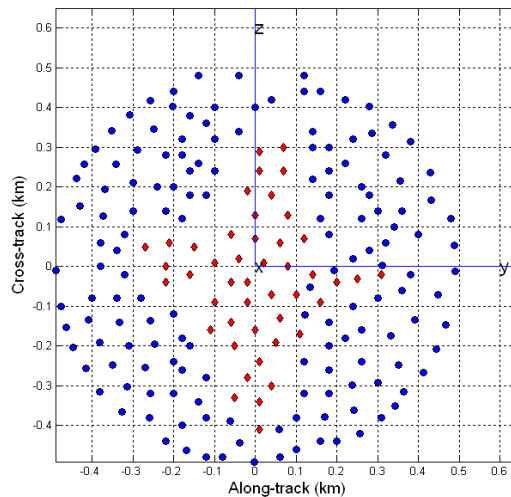


Figure 4 (b): Swarm after 1000 orbits.

Thus after 1000 orbits, the vehicles that have the highest initial usage, those marked in red, have moved to the less fuel demanding locations.

Applying Swarm Technology to Constellations of Spacecraft

Spacecraft that operate as constellations are often distributed over a wide area of space. Some of the approaches that we have used with closely clustered swarms can be directly applied in this new situation. For example, if a vehicle is required to take avoiding action in response to a potential collision other agents can automatically fill any gap that might result. Also if it is necessary to geographically redistribute the constellation the vehicle can respond so as to follow a fuel usage plan. However there are other ways in which swarm technology can be applied to Constellations which we are currently investigating.

Transponder configuration

In order to maximise their revenue, communication satellites reconfigure their transponders. For example, a crude plan can be implemented such that communication types are prioritised so real time communication such as voice and personal communication would be given the highest priority and charged accordingly. The next priority might be given to large data sets such as video, television, transfer and the lowest to downtime data transfer. In practice this is much more complex involving decisions not only based on transponder availability but also on commercial and operational considerations. It should be possible to produce a swarm rule set to maximise revenue which may or may not maximise usage.

Relocation

There may well be situations when the constellations performance could be improved by changing the orbital parameters of the vehicles. An example of this might be a space

capability for detecting and monitoring bushfires. Such a system would be required to provide regular overfly to keep the data current, leading to a requirement for a constellation of spacecraft. While such a capability could be extremely valuable in reducing losses caused by wildfires, the bushfire season is relatively brief making high utilisation difficult. In a general sense one could shift the sensor platforms to look at the East of Australia in the Southern Summer and West Coast of USA in the Northern summer. This is, however, a crude approach. A far better one would be to deploy the assets based on fire risk which can be determined from local surface temperature, humidity and wind. A swarm-type approach would match sensor deployment to real time conditions.

Conclusion

The application to Bio-inspired swarms to engineering control and management problems is still in its infancy. There appears to be significant potential gains over more traditional methods, but a great deal of work is still required before these methods can be widely deployed. This fact is evident as there are few if any currently deployed systems. One area where potential advantages can be postulated is in the control and management of space assets, and it is towards this end our research is directed.

References

- [1] E. Bonabeau, *et al.*, *Swarm Intelligence- From Natural to Artificial Systems*: Oxford University Press, 1999.
- [2] D. Izzo and L. Pettazzi, "Autonomous and Distributed Motion Planning for Satellite Swarm," *Journal of Guidance, Control, and Dynamics*, vol. 30, April 2007.
- [3] S. Nag and L. Summerer, "Behaviour based, autonomous and distributed scatter manoeuvres for satellite swarms," *Acta Astronautica*, 2012.
- [4] T.-C. Chi, *et al.*, "Self Organised UAV Swarms," in *15th Australian International Aerospace Congress* Melbourne, Australia 2013.
- [5] H. Cheng, *et al.*, "Cooperative control of UAV swarm via information measures," *International Journal of Intelligent Unmanned Systems*, vol. 1, p. 19, 2013.
- [6] T.-C. Chi, *et al.*, "Debris Avoidance through Autonomous Spacecraft Scattering," in *12th Australian Space Science Conference*, Melbourne, Australia, 2012.
- [7] H. Cheng, *et al.*, "Behaviour-Based Control Law for Spacecraft Swarm Operation," in *5th International Conference on Spacecraft Formation Flying Missions and Technologies*, Munich, Germany, 2013.
- [8] C. J. M. Verhoeven, *et al.*, "On the origin of satellite swarms," *Acta Astronautica*, vol. 68, pp. 1392-1395, 2011.
- [9] R. Burch, "The case for disaggregation of U.S. MILSATCOM," in *MILITARY COMMUNICATIONS CONFERENCE, 2011 - MILCOM 2011*, 2011, pp. 2286-2291.

Computational Analysis of Pintle Variation in an Expansion-Deflection Nozzle

Kyll Schomberg¹, Graham Doig¹ and John Olsen¹

¹ *School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW 2052*

Summary: The effect of varying the geometry of the flow deflector or pintle in an expansion deflection (ED) nozzle has been explored through a computational analysis. Flow behaviour was evaluated using axisymmetric, RANS-based numerical models and validated using experimental static pressure values and schlieren photography. The length and angle of the pintle were varied in order to assess the potential for altitude-adaptive behaviour within the nozzle design. It was found that the pintle length had a far greater effect on flow behaviour compared to variation of the pintle angle. Evidence of altitude-adaptive behaviour was observed in the increased length ED nozzle through the existence of a wake-dominated flowfield and avoidance of flow separation. The results demonstrate that an altitude-adaptive flowfield incorporating an open wake can be achieved with careful design. The salient effect of turbulence within the ED nozzle confirmed the unsuitability of traditional inviscid numerical techniques when describing complex nozzle flowfields.

Keywords: Propulsion, nozzle, expansion-deflection, pintle, CFD

Introduction

Conventional rocket propulsion systems are currently subject to significant nozzle inefficiencies when operating over a range of altitudes. The source of this inefficiency stems from the use of fixed-area convergent-divergent nozzles that yield a constant exit pressure for the exhaust. Optimal nozzle efficiency is achieved when the exit pressure of the exhaust is equal to that of the receiver, or local atmospheric pressure. As atmospheric pressure continually varies with altitude, this condition is only satisfied at one discrete point and losses are present at all other altitudes. The loss of efficiency due to changes in receiver pressure can reach as high as 15% within nozzles designed to operate between sea level and the vacuum of space [1].

Advanced nozzle concepts capable of compensating for the variation in receiver pressure due to altitude have existed for over half a century [2]. These concepts can be categorised with respect to the adaptive mechanism employed and generally fall into two general classes: variable area nozzles [3, 4] or controlled flow separation devices [5]. Predicting and controlling the flow separation phenomenon is difficult [6], and separation of exhaust flow from the nozzle wall can result in damaging side loads [7]. Additionally, compensation through flow separation can only produce a finite number of available area ratios. Therefore, altitude compensation using these devices is not a continuous process and still subject to pressure-induced efficiency loss [8].

Nozzle concepts utilising a passive approach to exit area variation continually manipulate the effective nozzle area ratio to inherently compensate for any changes in receiver pressure, maintaining high efficiency irrespective of altitude. Variation of the exit pressure is achieved

through manipulation of the effective area ratio of the nozzle through utilising the atmospheric air itself. This interaction occurs externally to the nozzle shroud in the plug nozzle [3] or internally in the expansion-deflection (ED) [4] nozzle. Early work on both nozzle concepts suggested the altitude adaptive potential of the plug nozzle was far greater than the ED nozzle [9]. However, the configuration used in the study was different to that proposed in the original design [4] and followed a methodology inherently unable to compensate for altitude [10]. Variations in these design philosophies can be seen in Fig. 1. Recent work has shown comparable altitude adaptive behaviour between both concepts [11]. The lack of understanding of the ED nozzle flowfield coupled with the increased heat flux and transonic region instability of the plug nozzle [12] justify further investigation into the ED nozzle concept

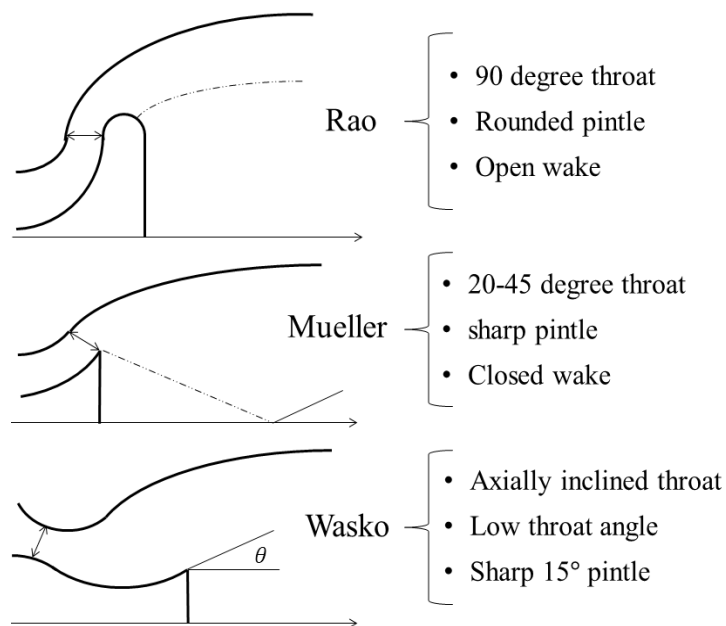


Figure 1: Variations in ED nozzle design philosophies

Altitude compensation is achieved in the ED nozzle through use of a central flow deflector, commonly referred to as a pintle. The function of the pintle is to deflect the exhaust radially outwards towards the nozzle wall, resulting in the creation of a viscous wake region at the base of the pintle comprised of local ambient air. The interaction between the supersonic exhaust and stagnant wake creates a shear layer that varies the effective area ratio of the nozzle with respect to the wake region pressure. While the wake remains ‘open’ to the atmosphere, the effective area ratio will be equal to that required for complete expansion of the exhaust. However, as altitude increases, the required effective area ratio will also increase until the physical maximum is reached. At this point, the wake becomes ‘closed’ to the atmosphere and nozzle behaviour will be similar to that of a conventional nozzle. Operation in open and closed mode is shown in Fig. 2.

In this work, variations in pintle geometry through length and angle manipulations were made to a Wasko-type ED nozzle to determine if altitude-compensating, or open mode operation could be observed. It is important to note that the traditional numerical tools used to analyse the ED nozzle are restricted to analytical methods that require an inviscid, irrotational flowfield in order to produce a solution [11]. While these conditions do not limit accuracy for conventional nozzle flows, their use to describe the ED nozzle flowfield is limited and requires the entire wake region to be assumed. An approach using computational fluid

dynamics (CFD) is inherently able to describe all flow regions within the ED nozzle and was therefore selected for this work.

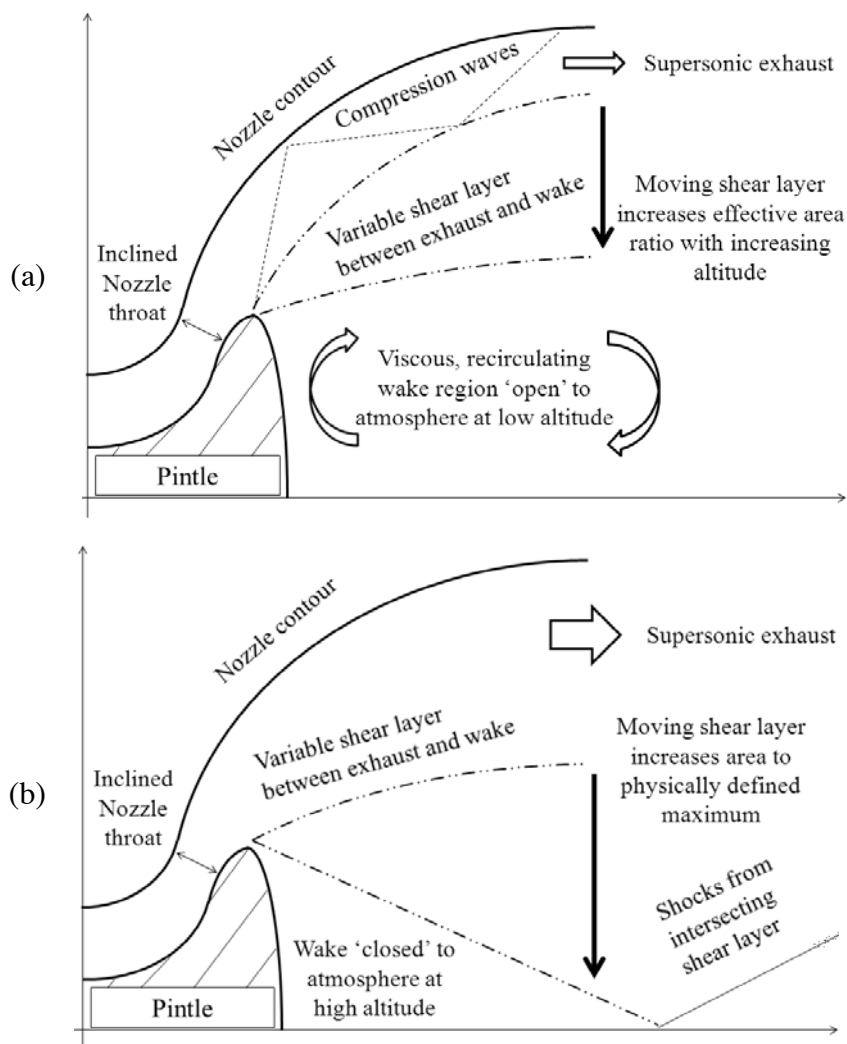


Figure 2: Operation of the ED nozzle in (a) open wake and (b) closed wake mode

Numerical model

All numerical results were generated using the commercially available software, ANSYS Fluent 14.5. Due to the quasi-steady nature of the full-flowing nozzle flowfield, a time or Reynolds averaged (RANS) approach to turbulence modeling was used. A two-dimensional axisymmetric pressure-based coupled solver was used in conjunction with second order spatial discretisation schemes for all models. Turbulent parameters were derived from the Reynolds number and boundary layer thickness resulting in a turbulent intensity of 3.6% and length scale of 1.68mm at the nozzle inlet and a three co-efficient Sutherland model was used for viscosity [13]. To determine solution convergence, the mass flow rate was recorded at the nozzle inlet, exit and outlet and satisfied after a variation of mass flow less than 1% between locations coupled with a change in magnitude of less than 0.1% over 500 iterations had been achieved. The outflow domain was sized with respect to the nozzle throat radius r_t and was kept consistent between configurations. Details of the computational domain can be seen in Fig. 3.

The geometry utilised in all simulations was consistent with that used in a prior experiment [14]. A baseline pintle length of 20mm and angle of 10° were used in order to follow the

nozzle contour and achieve a pintle base to nozzle exit area of 0.1. The pintle length was varied by $0.5L_p$, L_p and $1.5L_p$ and the pintle angle varied by $0.5\theta_p$, θ_p and $2\theta_p$ where L_p and θ_p were the baseline pintle length and angle respectively. Discretisation of the domain was completed using a fully-structured scheme consisting of quadrilateral mesh cells. A multiblock technique was employed to ensure cells were aligned with the direction of fluid flow. In order to accurately model boundary layer effects, prism inflation layers were implemented in all models and a non-dimensional wall distance (y^+) value of 1 was maintained. The presence of cells within the viscous sub-layer was necessary to describe the strong pressure gradients within the flowfield as well as model separation of the flow from the nozzle wall.

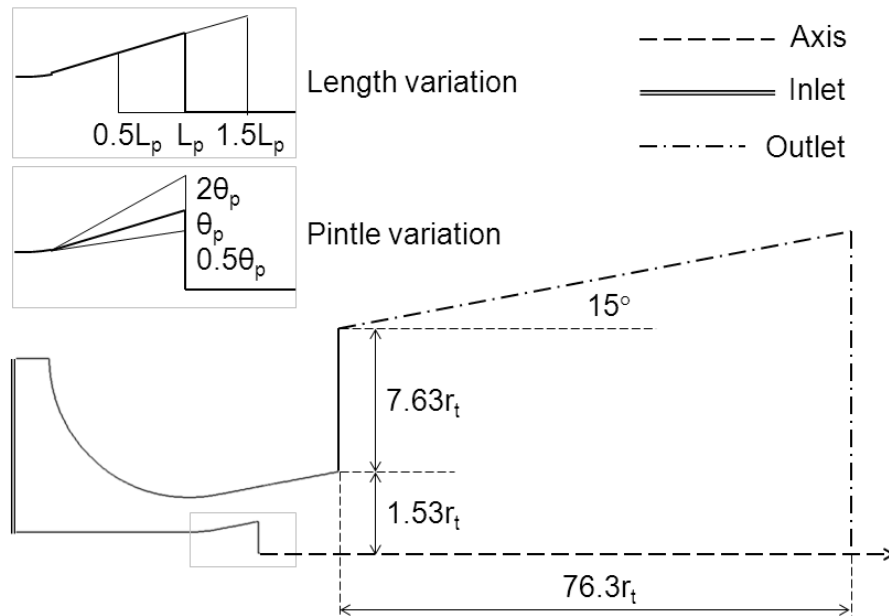


Figure 3: (a) Computational domain with pintle variations

A grid convergence index (GCI) analysis was used in order to determine mesh independence and quantify numerical uncertainty [15]. The GCI analysis was carried out across four discretisation schemes: coarse, refined, standard and fine. The standard and fine levels of refinement were achieved using a quad region adaption technique and resulted in a cell count of 1.01 , 4.04 and 16.2×10^5 elements for the coarse, standard and fine mesh levels respectively.

Flow through all nozzle configurations was treated as compressible and turbulent. The baseline turbulence model selected for the GCI analysis was the Spalart-Allmaras (SA) one-equation turbulence model developed specifically for aerodynamic flowfields involving wall bounded flows [16]. Variation of averaged exit velocity and quantification of numerical error using the GCI analysis is shown in Fig. 4. A safety factor of 3 was applied to all numerical error, as per the recommendations in [15]. This error was converted to a percentage and was found to be less than 1% between all mesh levels. This combined with the greatly reduced computational time resulted in the coarse mesh being selected for use in all models. To improve resolution of the shock waves, the coarse mesh was periodically adaptive with respect to a normalised density gradient. This resulted in an overall cell count of 2.18×10^5 elements.

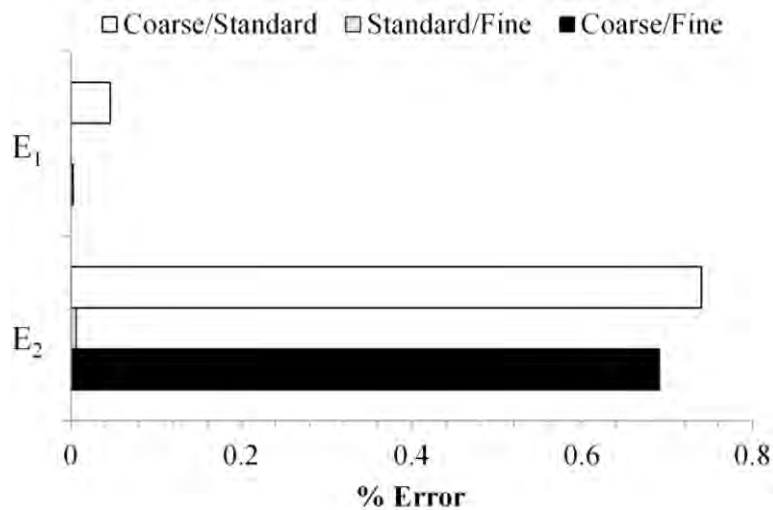


Figure 4: Quantification of exit velocity fine mesh error E_1 and coarse mesh error E_2

To evaluate the influence of turbulence on the solution, a comparison was made between the SA, k-omega shear stress transport (k- ω SST) [17], k-epsilon realizable (k- ϵ realizable) [18] and inviscid models. The evaluation was conducted using static pressure readings taken using tappings at the nozzle wall throughout the divergence section and a comparison of experimental and numerical schlieren images immediately downstream of the nozzle exit. The pressure distribution and schlieren comparison are shown in Figs. 5 and 6 respectively. Quantification of error for the experimental pressure readings were taken from the values reported in [14] and determined to be 2.5% throughout the divergence section and 5% at the throat due to the high rate of pressure change in the region

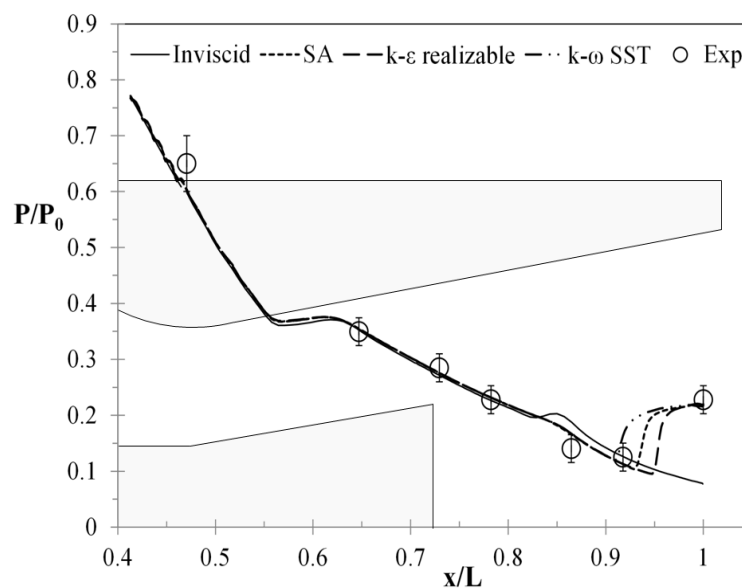


Figure 5: Static pressure over stagnation pressure against non-dimensionalised nozzle length

A clear variation in pressure distribution was evident between each of the turbulence models used. The inviscid model predicted a full-flowing nozzle whereas separation of the flow from the nozzle wall was modeled in all turbulent simulations. The location of separation varied between models, although all were within the recorded experimental pressure values. The variation in numerical schlieren was far more pronounced; both two-equation turbulence models generated a wake-dominated flowfield. Comparatively, the experiment and SA turbulence models produced a shock-dominated flowfield. Due to the high agreement between both performance measures, the SA model was used in all models.

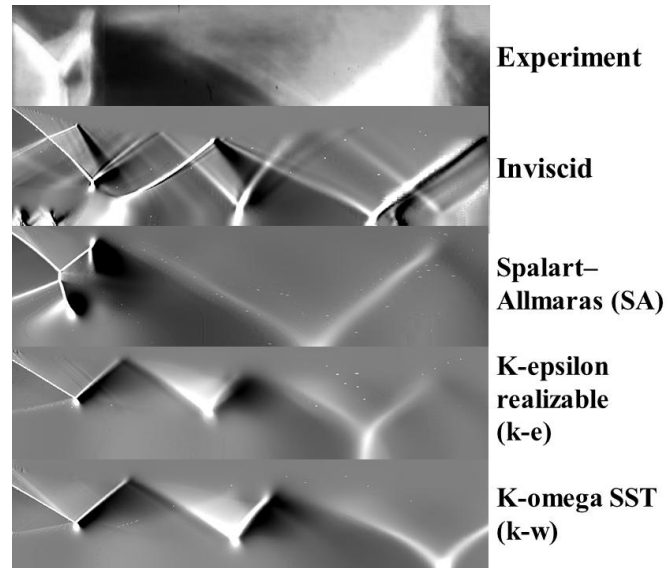


Figure 6: Comparison of experimental and numerical schlieren images immediately downstream of the nozzle exit

Results

To observe flow behaviour for all pintle variations, operation of each nozzle configuration was conducted across two operating conditions: overexpanded (OX) flow and grossly overexpanded (GOX) flow. These flow conditions were categorised by the theoretical threshold at which separation of the flow from the nozzle wall would be expected to occur [19]. As mentioned previously, all experimental results were taken from previous work [14].

Length Variation

In the length-varied configurations, the OX and GOX operating conditions were evaluated at a stagnation pressure of 4.56 and 6.29 atm respectively. Static pressure distributions are shown in Fig. 7 and 8 and a comparison of numerical (upper) and experimental (lower) and numerical schlieren images are given in Fig. 9.

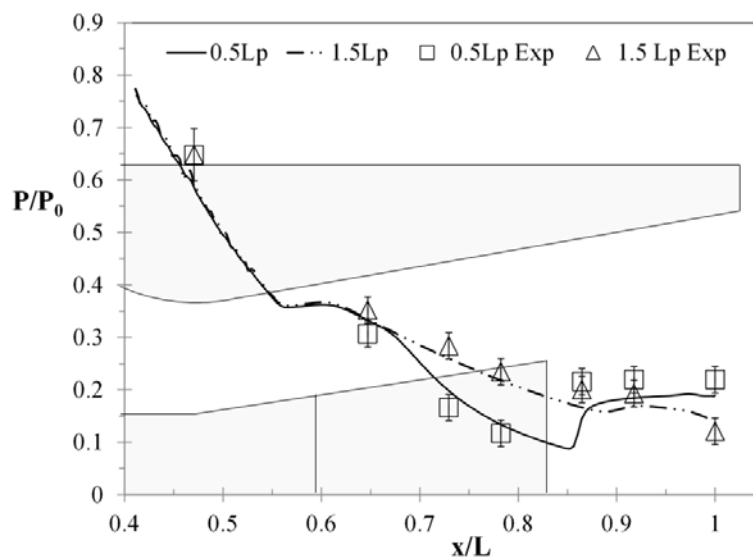


Figure 7: Length-varied normalised static pressure distribution under GOX flow conditions

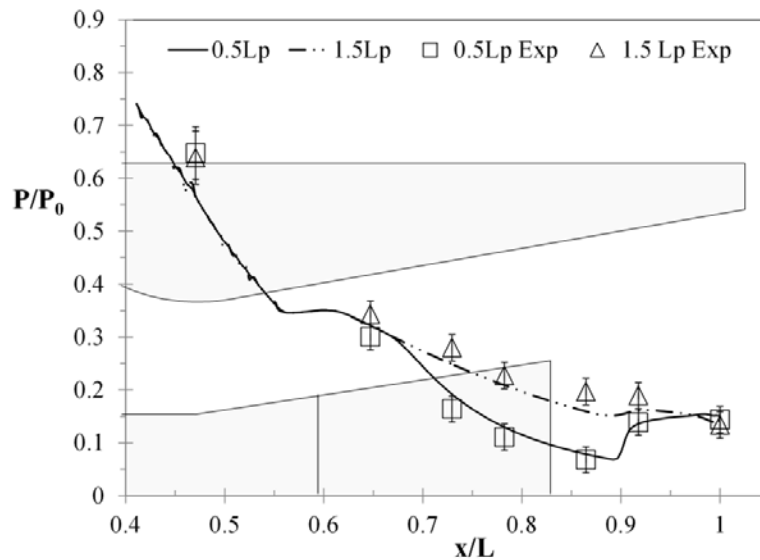


Figure 8: Length-varied normalised static pressure distribution under OX flow conditions

The correlation between experimental and numerical pressure values was within the predicted error range for both configurations under both operating conditions. Flow separation was avoided in the increased length ED nozzle which suggested open wake or altitude compensating behaviour. In comparison, separation occurred under both OX and GOX conditions in the reduced length ED nozzle and the location of separation varied by less than 5% between these operating conditions. The relatively consistent distributions supported the inherent quality of the ED nozzle to operate independently of operating pressure.

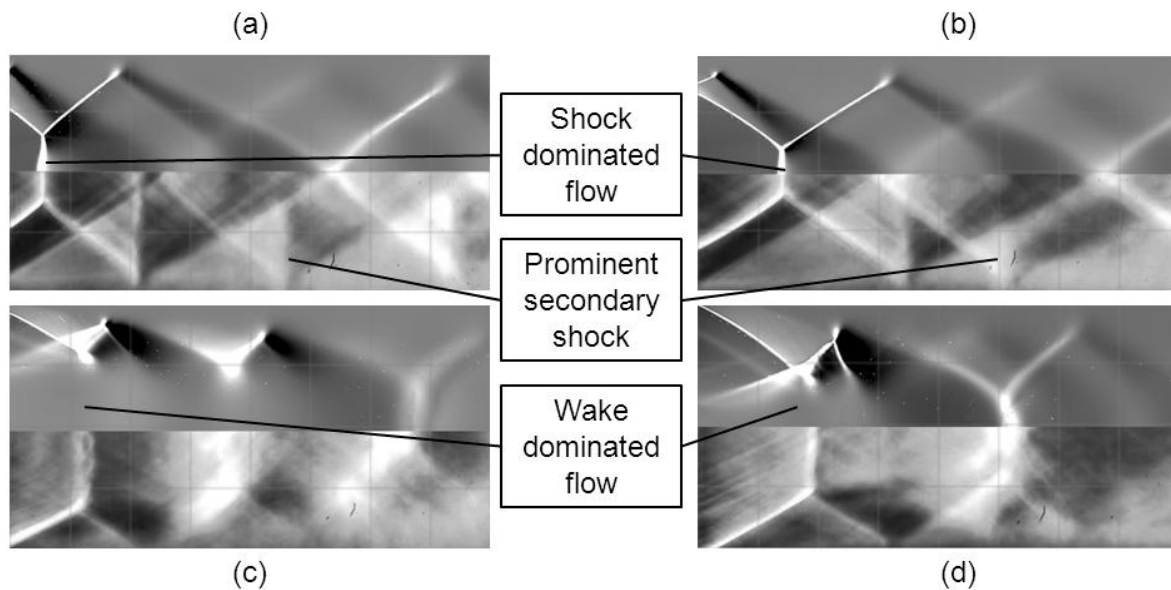


Figure 9: Length-varied schlieren comparison for (a) GOX $0.5L_p$, (b) OX $0.5L_p$, (c) GOX $1.5L_p$ and (d) OX $1.5L_p$.

Similar to the pressure distributions, a generally high correlation between density gradient was evident in the experimental and numerical schlieren images. A prominent secondary ‘trailing’ shock induced by the pintle was evident in the reduced length ED nozzle under both flow conditions. There was a slight offset of the secondary shock between the numerical and experimental images, although the primary or separation-induced shock was well captured. The shock-dominated flowfield for both conditions suggested that wake closure occurred in the reduced length ED nozzle. However, a wake dominated flowfield was apparent in the increased length ED nozzle which was consistent with the avoidance of flow separation.

Correlation between schlieren images in the increased length OX flow condition was reduced, with a significant variation in primary shock angle and Mach diamond structure.

Angle Variation

In the angle-varied configurations, the OX and GOX operating conditions were evaluated at a stagnation pressure of 4.92 and 6.21 atm respectively. Static pressure distributions are shown in Fig. 10 and 11 and a comparison of numerical and experimental schlieren images is given in Fig. 12.

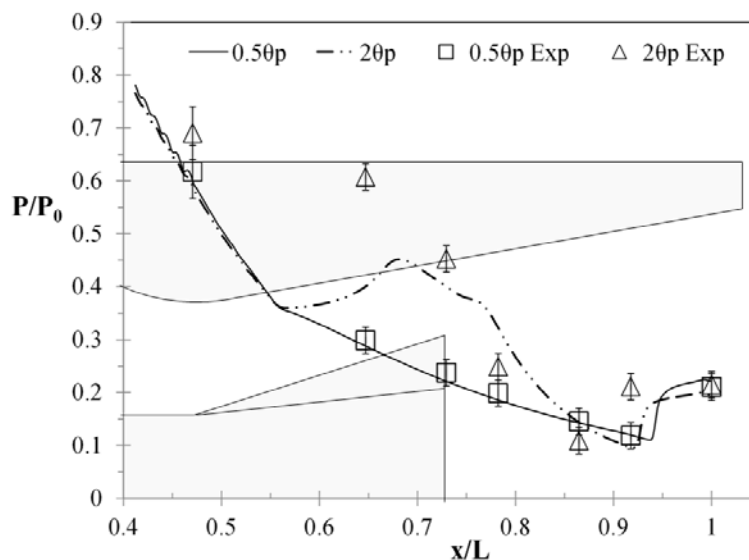


Figure 10: Angle-varied normalised static pressure distribution under GOX flow conditions

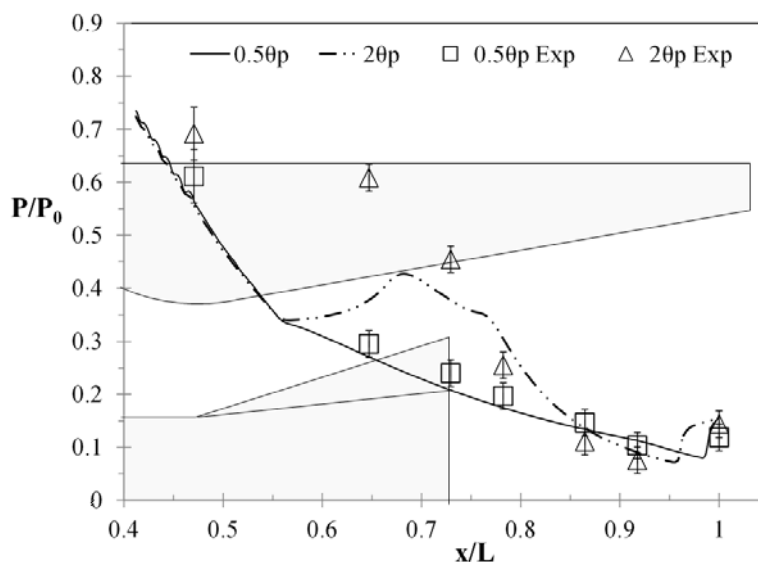


Figure 11: Angle-varied normalised static pressure distribution under OX flow conditions

The agreement between experimentally measured and numerical pressure values were high in the reduced angle ED nozzle with values within error ranges for both operation conditions. Correlation was extremely poor for the increased angle ED nozzle which can be attributed to the inconsistent rate of area increase throughout the divergence section. In the experimental nozzle, the pressure distribution showed that the throat was displaced axially towards the end of the pintle. Comparatively, the numerical model predicted throat conditions at the geometric throat resulting in a compressive increase in static pressure at the pintle end before further

reduction. The occurrence of flow separation was again consistent between operating conditions as shown through a difference of location of less than 5%.

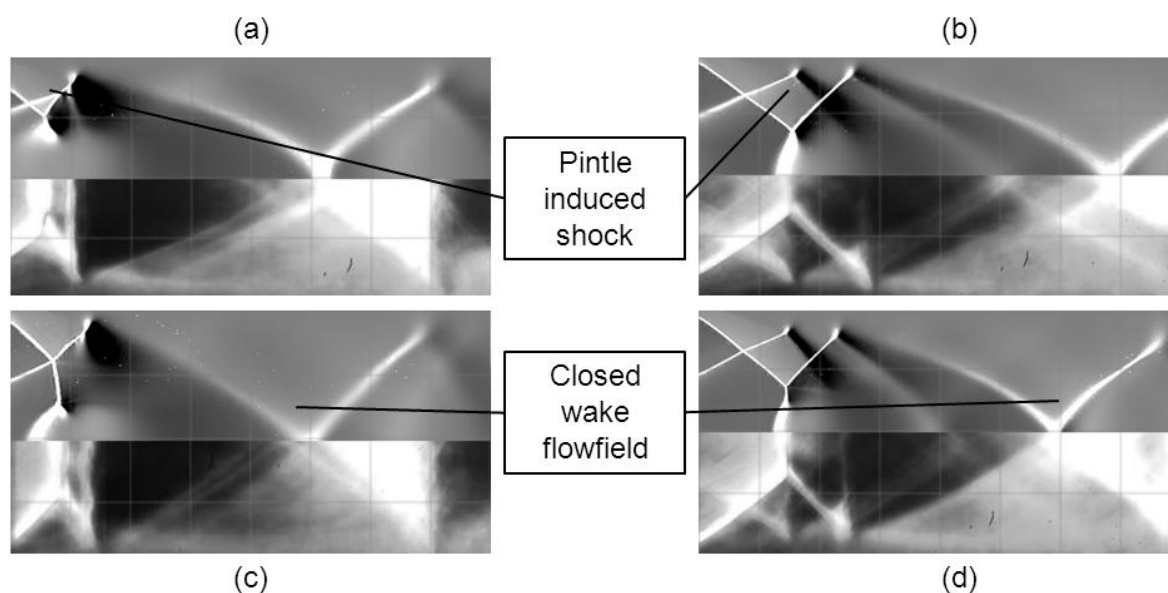


Figure 10: Angle-varied schlieren comparison for (a) GOX $0.5\theta_p$, (b) OX $0.5\theta_p$, (c) GOX $2\theta_p$ and (d) OX $2\theta_p$.

The flowfield was well described by the numerical model under all conditions. A secondary pintle-induced shock was observed for all configurations but was less prominent in the GOX flow condition due to the increased turbulence levels amplifying the influence of the upstream wake. Operation under all conditions was restricted to closed wake or non-altitude compensating flow regardless of pintle angle. Elongation of the Mach diamonds was greater for the reduced angle attachment which was expected due to the location of flow separation being closer to the nozzle exit.

Conclusion

A high level of agreement was evident between the computational and experimental results for all configurations and flow conditions tested. The primary discrepancy was the correlation between static pressure values in the increased angle ED nozzle. This occurred due to an inconsistent rate of area ratio increase within the nozzle and resulted in axial displacement of the throat in the experimental testing. Additionally, the agreement between numerical and experimental schlieren images was compromised in the increased length ED nozzle due to the turbulent nature of the wake-dominated flowfield. The observed inaccuracies in the computational results may represent a limitation in RANS modelling to accurately describe the flow behaviour in an open wake ED nozzle flowfield. This result further reinforces the unsuitability of traditional inviscid numerical techniques and highlights the salient effect of turbulence within the ED model flowfield and that an open wake and therefore a nozzle capable for compensating for altitude can be achieved with careful design.

Acknowledgements

The authors would like to acknowledge Mr. Charles Queriaud for his assistance with the experimental schlieren imagery and Mr. Terry Flynn for his assistance with setting up the experimental rig and obtaining experimental pressures. The contributions of Mr. Ian Cassapi, Mr. Andrew Higley and Mr. Seetha Mahadeven in the manufacture of all experimental components are also acknowledged.

References

1. Sutton, G.P., Biblarz, O., *Rocket Propulsion Elements*. 7th ed., John Wiley & Sons, New York, 2001 Ch. 2, 3.
2. Rao, G.V.R., "Recent developments in rocket nozzle configurations," *ARS journal*, 1961, pp. 1488–1494.
3. Rao, G.V.R., "Spike nozzle contour for optimum thrust", *Ballistic missile and space technology*, Vol. 2, 1961, pp. 92-101.
4. Rao, G.V.R., "Analysis of a new concept rocket nozzle," *Journal of Liquid Rockets and Propellants*, 1960, pp. 669–682.
5. Nasuti, F., Onofri, M., "Flow analysis and methods of design for dual bell nozzles", in *37th joint propulsion conference and exhibit*, AIAA: Salt lake city, UT, 2001.
6. Ostlund, J., "Supersonic flow separation with application to rocket engine nozzles," Ph.D. Dissertation, Royal Institute of Technology, Stockholm, Sweden, 2004.
7. Onofri, M., Nasuti, F., "The Physical Origins of Side Loads in Rocket Nozzles," *35th Joint propulsion conference and exhibit*, 99-2587, AIAA, Los Angeles, CA, 1999.
8. Hagemann, G., Frey, M., Manski, D., "Critical assessment of dual bell nozzles," *33rd Joint propulsion conference and exhibit*, 1997-3299, AIAA, Seattle, WA, 1997.
9. Wasko, R.A., "Performance of annular plug and expansion-deflection nozzles including external flow effects at transonic mach numbers," NASA TN D-4462, 1968.
10. Mueller, T.J., Sule, W.P., Hall, C.R., "Characteristics of separated flow regions within altitude compensating nozzles variation," UNDAS TN-029-FR-9, University of Notre Dame, 1971.
11. Taylor, N.V., Sato, T., "Experimental and Computational Analysis of an ED nozzle in open wake mode," *26th Applied Aerodynamics Conference*, 2008-6924, AIAA, Honolulu, HI, 2008.
12. Hagemann, G., Immicht, H., Terhardt, M., Dumnov, G., "Critical assessment of the linear plug nozzle concept," *37th Joint propulsion conference and exhibit*, 2001-3683, AIAA, Salt lake city, UT, 2001.
13. Sutherland, W., "The viscosity of gases and molecular force," *Philosophical magazine*, Vol. 36, 1893, pp. 507–531.
14. Schomberg, K., Olsen, J., "Investigating pintle geometry in expansion-deflection nozzles," *15th Australasian Fluid Mechanics Conference*, Launceston, Australia, 2012.
15. Roache, P.J., "Quantification of uncertainty in computational fluid dynamics," *Annual review of fluid mechanics*, Vol. 29, 1997, pp. 123–160.

16. Spalart, P., Allmaras, S., "A one-equation turbulence model for aerodynamic flows," *La Recherche Aerospatiale*, Vol. 1, No. 5, 1992, pp. 5-21.
17. Menter, F.R., "Two equation eddy-viscosity turbulence models for engineering applications," *AIAA journal*, Vol. 32, No. 8, 1994, pp. 1598-1605.
18. Shih, T.H., Liou, W.W., Shabbir, A., Zhu, J., "A New k-epsilon eddy-viscosity model for high Reynolds number turbulent flows - model development and validation," *Computers Fluids*, Vol. 24, No. 3, 1995, pp. 227-238.
19. Summerfield, M., Foster, C., Swan, N., "Flow separation in overexpanded supersonic exhaust nozzles." *Journal of Propulsion*, 1954: p. 319-321.

A Highly Mobile Wheel-on-Leg Planetary Rover for use in a Martian Analogue Environment

William Reid, Ali Haydar Göktoğan and Salah Sukkarieh
*Australian Centre for Field Robotics, The Rose Street Building J04
The University of Sydney, NSW, Australia, 2006*

Summary: Rovers used for Martian exploration have traditionally employed passive suspension systems that limit the step obstacle height that they can traverse to the diameter of their wheels. They are also susceptible to getting their wheels stuck in soft soil. To overcome these problems, hybrid wheel-on-leg planetary rovers capable of operating in multiple mobility modes have been designed. In this paper we present such a concept. The rover's mobility system is composed of four legs with a wheel at the end of each leg, resulting in a platform with 16 degrees of freedom. The rover is capable of energy efficient driving as well as high traversability walking. In addition, the rover can orient its body into a stable configuration by raising and lowering its legs. Presented is the selection process of the wheel-on-leg mobility system, followed by a summary of the design of the rover and the proposed rover operations in a Mars analogue environment.

Introduction

In 2013 the University of Sydney, the The University of New South Wales and the Powerhouse Museum were awarded a \$2.9 million research grant over two years from the NBN-Enabled Education and Skills Services program, with the aim of addressing the decline in science and engineering uptake, using the capabilities of the National Broadband Network to link users from anywhere in Australia to educators, researchers, engineers, and scientists. As part of this project the University of Sydney's role is to design and build two new rovers that both meet the needs of the project education requirement as well as its research needs.



Fig. 1: The MAMMOTH (Mars Analogue Multi-MODE Traverse Hybrid) Rover. The rover is shown next to a 1.8 m tall human figure for scale.

The MAMMOTH (Mars Analogue Multi-MODE Traverse Hybrid) Rover, shown in Figure 1, is a wheel-on-leg system comprising four legs with a wheel at each end. Each leg has four degrees of freedom: hip rotation, thigh raise/lower, ankle rotation and wheel rotation. This provides the rover with a large set of possible joint positions, or configuration space, compared to traditional Mars rovers. The MAMMOTH rover is designed to test highly mobile operations in a Mars analogue environment.

Four mobile robotic rovers have explored or are still exploring Martian sites. These include the Sojourner, Mars Exploration Rover and Mars Science Laboratory platforms [1]. Each of these rovers utilizes the rocker-bogie mobility system composed of four or six independently steered and driven wheels. On either side of the rovers there are two wheels at the middle and back that are attached to the two ends of a bogie, which is in turn connected to the front wheel via a rocker. These rocker-bogie assemblies are connected via a differential at the centre of the rover body. The two main advantages of this mobility system are that all six wheels are kept in contact with the surface while driving on uneven terrain, and that the wheels' ground contact pressures are equalised [2]. Disadvantages of a rocker-bogie rover include the limitation of only being able to safely traverse obstacles that have a step height less than or equal to the rover's wheel diameter [3]. These rovers are also susceptible to getting stuck in soft soil, or "embedding events". The most significant example of this is the Mars Exploration Rover "Spirit" getting embedded in soft soil, causing an end of its mobile operations [4].

Rovers that are capable of operating in multiple mobility modes are not as susceptible to these problems. An example of this type of platform uses the hybrid wheel-on-leg mobility system, which combines the high mobility of legged vehicles for rough terrain traversal with more energy efficient wheeled locomotion over benign terrain. Examples of wheel-on-leg rovers include the SHERPA and ATHLETE rovers. As discussed in [5], the SHERPA rover is a 24 degree of freedom quad wheel-on-leg system that has been designed as a highly mobile robot used to explore lunar craters. The SHERPA's operational tasks are to traverse harsh terrain and manipulate communication infrastructure and other robots as part of a robotic team. It is capable of conducting all of these tasks given its multiple locomotion modes and re-configurability. The ATHLETE rover described in [6] is a hex wheel-on-leg system that was also designed for lunar operations. The ATHLETE has 42 degrees of freedom, with which it is able to perform many different exploration and construction tasks. Both rovers are capable of traversing obstacles that have a step height equal to or larger than their chassis height. They are capable of lifting their wheels off the surface and clambering or walking out of an embedding event. The disadvantages of these wheel-on-leg platforms come with their mechanical and control complexities. Having large numbers of actuators and a complex leg mechanism increases possible failure modes, which introduces a reliability risk. A large number of actuators presents complexities in the form of a larger possible configuration space than conventional rovers, resulting in complex planning and control schemes.

To select an appropriate mobility system for the MAMMOTH rover, this article evaluates various mobility system designs based on specific research and educational objectives, operational constraints and engineering requirements. The technical design of the MAMMOTH rover is then presented, including a higher level explanation of the mechanical and electrical sub-systems. This is followed by a description of the conceptual operations of the MAMMOTH rover, highlighting the various locomotion modes that the rover is capable of operating in. Lastly, conclusions and a discussion of further work to be performed are provided.

Design Selection

The conceptual design of the MAMMOTH rover has been selected based on an evaluation of various existing rover concept designs against certain design criteria that are based on the rover's design objectives, operational constraints and engineering requirements. This section discusses the various aims, constraints and engineering requirements imposed on the MAMMOTH rover followed by a discussion of existing rover concept designs and their applicability to the MAMMOTH rover design.

Aims, Constraints and Engineering Requirements

The aims for the MAMMOTH rover are to:

- develop a platform that will allow high-school students the ability to control a highly mobile platform and explore a Mars analogue environment;
- research highly mobile rover operations on variable terrain;
- research semi-autonomous and autonomous Mars analogue mobile rover operations with the aid of tele-operation.

The majority of the rover's operations will be conducted within Sydney's Powerhouse Museum Mars Yard, a Mars analogue environment. The yard is a 140 m² space that is composed of highly variable terrain that includes boulder fields, loose soil, and a crater with highly inclined walls. This environment is a relatively confined space and limits the allowable size of the rover. During operation, it is foreseen that the rover will need to be handled by human operators, which introduces a limit on what the rover can weigh.

With these objectives and constraints in mind, a list of critical engineering requirements was formulated. The rover shall:

- traverse over soft sandy soils;
- be able to traverse over obstacles with step sizes equal to the rover's nominal body height;
- have a mass less than 80 kg;
- have an envelope volume less than 1.5 m³;
- have a maximum driving speed less than 0.5 m/s;
- operate on inclines up to 20°;
- be able to drive continuously for 0.5 hours and conduct static operations for 5.5 hours on a single battery charge;
- operate semi-autonomously with operator supervision;
- operate autonomously given a goal position and orientation.

Concept Design Evaluation

Ultimately, the criterion that is most consistent with the design aims, constraints and requirements is the rover's mobility. The main purpose of the rover is to conduct complex traverse operations given the operational environment. The educational and research potential

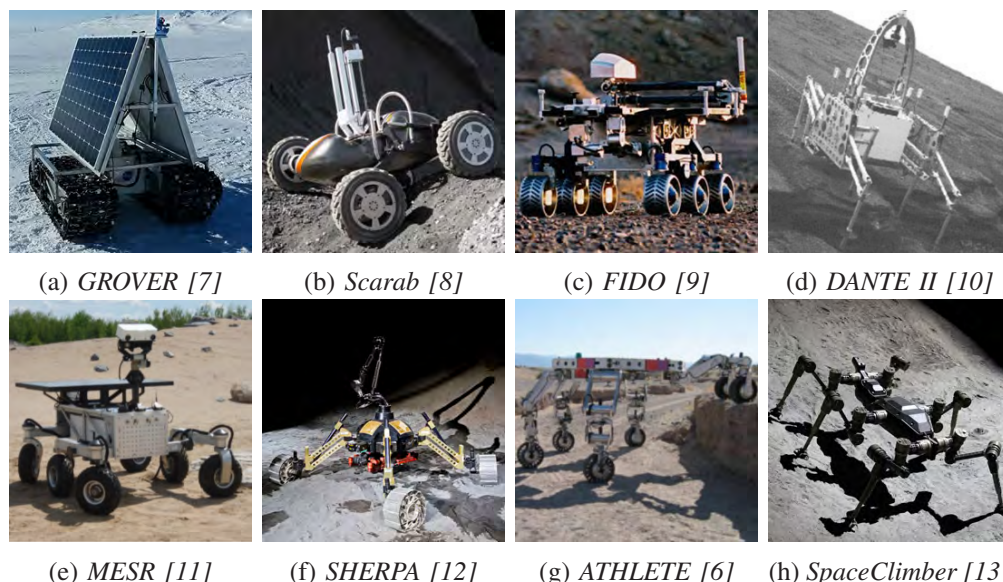


Fig. 2: The various rover platforms studied to choose a mobility system for the MAMMOTH rover.

of the rover is tightly coupled with the rover's ability to traverse complex terrain. Mass and volume are the next most important criteria considering the strict environmental constraints of the Mars Yard and the desired mission duration times. Cost follows given a limited development budget. Rover power consumption is constrained by the 6 hour operation time requirement. Control simplicity and mechanical simplicity are desired, however not strictly necessary. It is ultimately a research goal to determine what degree of control and mechanical complexity must be sacrificed to achieve high mobility.

A planetary rover's operational abilities are most significantly determined by its mobility system. A large taxonomy of rover mobility systems exists, and it is arranged into continuous, discrete, and hybrid classes [8]. Continuous mobility systems include tracked and wheeled systems such as the GROVER [7], SCARAB [8], FIDO [9] and MESR [11] rovers shown in Figure 2. There are significant limits on the type of terrain that wheeled rovers are capable of traversing. Their technological maturity and relative simplicity allows for reduced development time, reduced cost, control simplicity and mechanical simplicity. Continuous mobility systems are also considered the most energy efficient [6].

Discrete mobility systems include walking rovers such as the DANTE-II frame walker [10] and the SpaceClimber hexapod [13] shown in Figure 2. The advantage of these rovers is that they have increased mobility by being able to walk over obstacles. DANTE-II is an example of a robot that has been used for extremely harsh terrain such as high inclination craters with loose soil and large rocks. Discrete mobility systems are of significant research interest given their uncommon use in planetary or lunar exploration studies and their potential to explore harsh planetary terrain. Discrete mobility system masses are generally larger compared with continuous systems. Large mass, energy inefficiencies and mechanical and control complexities are the main disadvantages of these systems.

Hybrid wheel-on-leg platforms, such as the ATHLETE [6] and SHERPA [5] rovers, shown in Figure 2, take advantage of the energy efficiency of wheeled systems and the increased manoeuvrability of legged platforms by incorporating both into a single system. Their research

potential is also promising given the high mobility of the platforms demonstrated in [5], [6], [14] and [15]. Disadvantages come with increased mass, mechanical complexity and control complexity. Based on its high manoeuvrability and research potential, a hybrid quad wheel-on-leg mobility system most similar to the SHERPA rover was selected for the MAMMOTH rover. To reduce cost and increase mechanical simplicity a platform with four wheel-on-legs as opposed to a five or six wheel-on-legs was chosen.

Detailed Design

This section describes the higher level design of the MAMMOTH Rover. An overview of the electrical sub-system is given in addition to a summary of key elements of the mobility system and the rover's structure. The rover's mass is 80 kg, while its envelope volume is variable due to the re-configurability of the wheel-on-leg mobility system. In its compact configuration, the rover has a minimum footprint of 600×600 mm, while when fully extended the footprint area is 1300×1300 mm. The height of the rover, from the ground surface to the height of the mast can vary between 700 and 1200 mm. During typical operations the rover is capable of driving at a maximum speed of 0.4 m/s.

Electrical Sub-System

The MAMMOTH rover's electrical sub-systems are composed of the power, mobility, communications, control and sensor sub-systems, shown in the architectural diagram in Figure 3. The rover is powered by eight 95 Wh, 6.6 Ah rechargeable Li-Ion cells. The requirement that the rover can operate for a 6 hour period is highly dependent on the duty cycle of each of the actuators. Assuming 10%, 5%, 5%, and 10% duty factors for the hip, thigh, ankle and wheel actuators respectively, it is estimated that the eight batteries will be sufficient. Given that typical operations have yet to be determined, extra volume for two more batteries has been included in the battery housing. Each of the batteries provides 14.4 V, which is converted to 24 V, 12 V, 5 V and 3.3 V busses. Each of the mobility system actuators run off the 24 V bus, while the majority of the sensors and control electronics run off the 12 V, 5 V and 3.3 V busses.

The control of the rover is handled by four on-board processors. A 1 GHz Linux computer acts as an embedded controller, which handles all lower level mobility control, telemetry processing and sensor processing. An 8-bit microcontroller is used as the secondary embedded controller. Its responsibility is to control the movement of each of the linear actuators, which do not have built in speed or position controllers. One 2.7 GHz and another 1.6 GHz computer are used for all high level guidance, navigation and control processing. All communications with ground control stations will take place over Wi-Fi and ZigBee wireless interfaces.

Rover monitoring and environment monitoring are performed by an array of sensors all over the rover. Each actuator joint houses either an encoder or potentiometer for monitoring angular or linear position. Force sensors are present at the end of each linear actuator, reporting the loading at each leg. Load measurements will allow the rover to know whether a leg is making contact with the surface or not and also provide data for linear actuator velocity and position control. Voltage, current and thermal sensors are included in each of the rotary actuators, while voltage and current readings are also available from the linear actuators. An

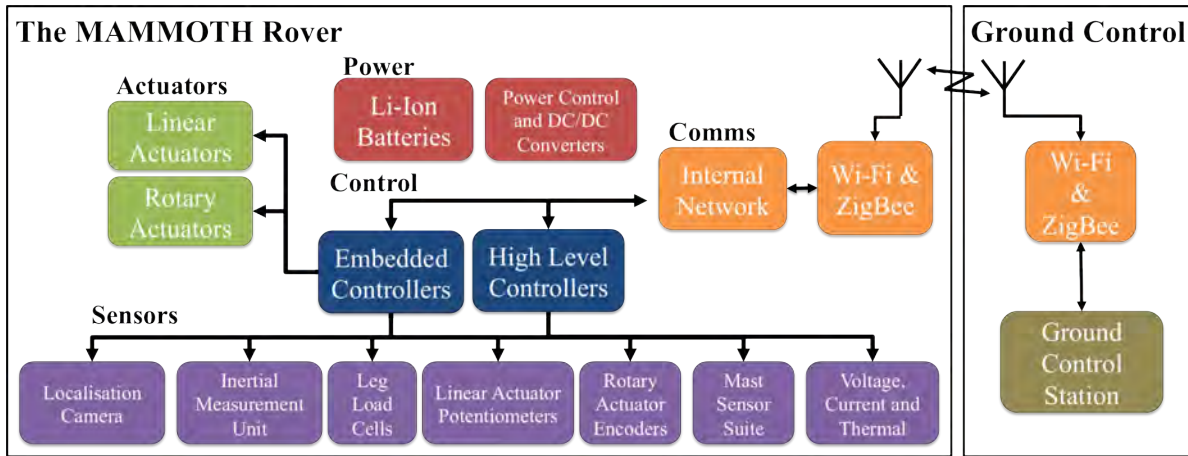


Fig. 3: Architectural diagram of the MAMMOTH rover's power, mobility, control, communications and sensor systems.

inertial measurement unit (IMU) is located inside the rover chassis, and provides orientation data. Data from a camera on the rover enclosure lid is fused with IMU data to form an accurate estimate of the rover position and orientation within the Mars Yard. The camera data is processed to identify accurately mapped distinct fiducial markers on the ceiling of the Mars Yard. Based on the markers that the rover identifies and the orientation data from the IMU, a centimetre accurate position solution can be triangulated. Additional sensors will be placed in the head of a mast, 0.5 m above the rover torso. These sensors will include laser range finders and a stereo camera pair. This sensor suite will provide valuable data for classifying the surrounding environment.

Mobility Sub-System

The MAMMOTH rover mobility system is composed of four legs, each with a wheel attached at its end. A single leg has four actuated degrees of freedom resulting in 16 degrees of freedom for the whole mobility system. A leg is capable of being moved at the hip joint, which is located between the rover torso and the top of the leg. The middle of the leg, or the thigh, is made up of a parallel structure that has a linear actuator on its diagonal. The leg is raised and lowered by changing the length of the linear actuator. Given the thigh's parallel structure, the hip joint and wheel always remain parallel to one another. At the base of the leg is an ankle joint that steers the wheel. Connecting the ankle joint to the wheel is a C-bracket that allows the steering axis to align with the centre of the wheel. This arrangement allows for the control of the steering motion to be simplified given that there is no lever arm between the ankle steering axis and the wheel's vertical axis. The last leg joint is located along the driving axis of the wheel. A labelled view of the torso and leg is shown in Figure 4.

Each of the four legs is attached to the rover torso at the hip joints. The hip joints are located at the four corners of the 600 mm by 600 mm torso. The torso is the main structural element of the rover as well as the payload housing. The structure is a rectangular box with a flange around its perimeter. The interior of the box is used to house the power, control and telemetry electronics. The flange is used as an attachment point for payload items such as the rover mast and sensor suites. The saddle bags on the sides of the rover can be used to house temporary experiment payloads that may be designed by students. Each saddle bag

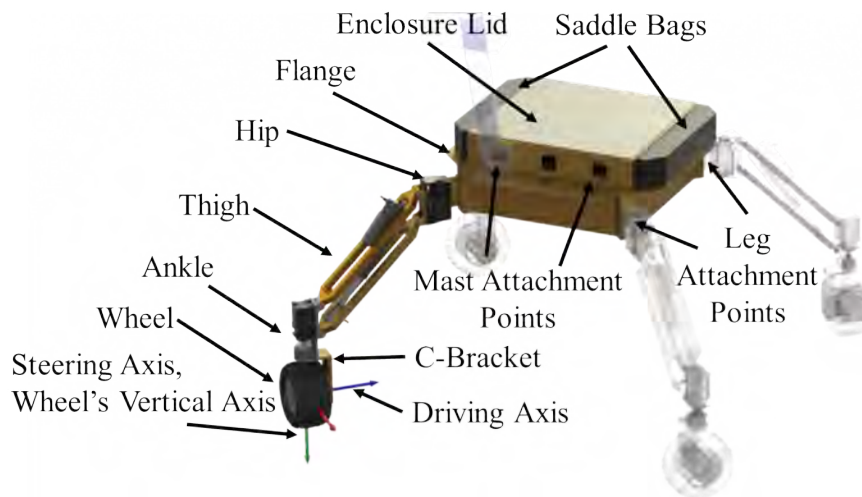


Fig. 4: The MAMMOTH leg and torso. The leg is made up of the wheel, ankle, thigh and hip joints. The torso is the central part of the rover body, where all of the limbs attach and the majority of the electronics are housed.

has mechanical attachment points along with data and power interfaces. The enclosure lid is removed to access the rover electronics. Given that there are data and power interfaces to the enclosure lid, additional experiment payloads may be mounted on the top of the lid.

A critical part of the mechanical design was the selection of appropriate actuators used to drive the rover's joints. The actuators have been selected based on loading, volume, power, mechanical simplicity, control simplicity and cost requirements. The maximum loading case for the wheel actuator is 20 Nm, assuming an 80 kg rover driving up a 20° incline. Determination of maximum loading experienced by the wheel steering actuator at the ankle joint is challenging given the varied types of soil that the rover will operate in. Experiments in the Mars Yard with a four-wheeled 35 kg rover show that in soft soil on a 20° incline the steering actuators begin to stall at approximately 8 Nm. It is estimated that an 80 kg four wheeled rover would require actuators capable of 20 Nm to steer in similar conditions. A maximum loading on the linear thigh joints would occur when one wheel is lifted off the ground and the three other legs are taking the full load of the rover. Each of the load bearing thigh joints may experience a maximum load of 800 N in this case. The maximum loading on the hip joint may be significant, approximately 50 Nm, when the rover is climbing 20° inclines and a maximum lever arm is created by the hip's angular position. These high torque situations can be avoided by limiting the lever arm length during high incline operations.

A commercial off the shelf (COTS) rotary servo-unit has been selected that not only meets the majority of the rotational load cases, but also meets the strict volume ($\leq 10 \text{ cm}^3$), mass ($\leq 1 \text{ kg}$), and power ($\leq 150 \text{ W}$ continuous operation) requirements. The selected solution also met strict budgetary requirements. The selected rotary actuator is used at each of the rotary joints and houses its motor, cycloid harmonic gearbox and control electronics within a $54 \times 54 \times 108 \text{ mm}$ volume with a 732 g mass. The maximum continuous torque of each unit is 21.1 Nm, while the maximum peak torque is 36 Nm. This actuator selection meets all requirements except for the most extreme hip joint loading cases. This will limit the operations of the rover so that the legs cannot be placed in high torque situations where the required hip actuator torque would be above 36 Nm. For the thigh linear actuator, a COTS linear actuator that matches the dimensions of the leg's parallel structure that is capable of lifting 1500 N

at a maximum speed of 11 mm/sec has been selected.

The angular limits for each of the rotary actuators have been chosen based on the geometric constraints of the rover. The wheel actuator can rotate continuously to drive the wheel. The ankle actuator is also able to rotate continuously given that a slip ring is used at the ankle joint to provide electrical connections to the drive actuator. The hip actuator is limited to a 270° movement, allowing the legs to be packed under the rover to form a compact configuration, as shown in Figure 5a. The linear actuators have a stroke length of 110 mm. This allows for a maximum leg angle between the thigh and a level ground of 70° and a minimum angle of -20° . Figures 5b and 5c show the rover in configurations with a maximum leg angle and minimum leg angle respectively. Figures 5b and 5c show the rover in configurations with a maximum leg angle and minimum leg angle respectively.



(a) Compact configuration. (b) Tallest configuration. (c) Shortest configuration. (d) Kineo-static walking.

Fig. 5: Various operational configurations of the MAMMOTH rover.

Conceptual Operations

Given the many degrees of freedom available to the MAMMOTH rover, it is able to move around its environment in many different ways. These movements can be characterised by a movement hierarchy, with increasing levels on the hierarchy representing increasingly complex movements. The movements presented form a basis for the control of the rover in its different locomotion modes. Any movement, except for the most primitive, is composed of movements belonging to lower levels in the hierarchy. The hierarchy is composed of the following:

- 1) **Actuator movement:** driving of a single rotational joint at the hip, ankle or wheel drive joint, or driving a linear thigh joint.
- 2) **Leg movement:** movement of multiple actuators in a single leg. These include raising and lowering a leg while the wheel is in contact with the surface and swinging the hip while the wheel is in contact with the surface.
- 3) **Primitive full rover movement:** the basic set of motions that rely on the simultaneous movement of joints in all four legs. Examples of these are rover steering by arranging all of the wheels so that the rover can turn about a specific point, arranging the legs to meet a specified yaw, pitch or roll angle of the body and driving all wheels at coordinated angular rates.
- 4) **Full rover movement:** sequences of movements to create full rover rotations and translations. Examples are driving while keeping the rover body in a stable orientation, compacting the rover by folding each leg underneath the body and lifting a single leg while re-positioning the other three legs so that the rover remains stable.

These motion primitives allow the MAMMOTH rover to take advantage of multiple locomotion modes. A nominal driving mode is shown in Figures 5b and 5c. This configuration allows the legs to raise and lower when rough terrain is encountered while keeping a stable torso orientation. The range of leg motion to accommodate for variable terrain is highlighted in Figures 5b and 5c, which show the maximum and minimum angular positions of the legs respectively. The rover has independently steered wheels, hence the steering angles are constrained by an instantaneous centre of rotation (ICR). The ICR forms the centre of a circle that intersects with the curve that the rover moves along. Steering angles are selected so that each wheel is perpendicular to the ICR. If the rover were to drive in a straight direction, the ICR would be located at infinity in a perpendicular direction to the rover's movement direction. If the rover were to turn on the spot, the ICR would be located directly below the centre of the rover. A discussion of ICR steering and constraints imposed on the placement of the ICR given a specific rover configuration is provided in [16]. When the rover needs to be stored in compact spaces, such as when it is brought back to a ground station for charging, the rover may fold its legs in towards the chassis and adopt a short, low footprint area configuration as shown in Figure 5a.

When driving operations are not sufficient for traversing rough terrain, the rover may engage in kineo-static walking or clambering as shown in Figure 5d. The rover can lift one leg over obstacles, while the other three legs remain in contact with the surface to retain static stability. These three contact legs must be positioned so that the projection of the rover's centre of mass onto the surface is bounded by the three leg contact points. This procedure may be repeated to form a walking or clambering gait.

Conclusions and Future Work

A design and operational analysis of a highly mobile rover that is capable of exploring terrain that is inaccessible to conventional planetary rovers has been performed. The work presented summarises the selection of the mobility system design for the MAMMOTH rover, comparing the merits of alternative concept designs against criteria that the rover must fulfill. It also outlines the design of the MAMMOTH wheel-on-leg rover. Lastly, the motion hierarchy of the rover and various locomotion modes are discussed. Compared to existing wheel-on-leg systems, the MAMMOTH rover has a reduced number of degrees of freedom in an effort to reduce platform complexity, while still maintaining the wheel-on-leg platform's high mobility and relative energy efficiency. Additionally, the MAMMOTH wheel-on-leg platform is capable of stable body orientation while traveling over rough terrain. Future work includes development of guidance, navigation and control schemes that take full advantage of the high mobility of the platform.

Acknowledgments:

This work was supported by the Broadband Enabled Education and Skills Services Department of Education program, Australian Federal Government (2012-2014) entitled Education 2020: enabling learning in science, engineering and mathematics, and the Australian Centre for Field Robotics (ACFR) at the University of Sydney. The authors would like to thank Javier Martinez and Muhammad Esa Attia from ACFR for their help in designing and building the MAMMOTH platform.

References

- [1] B. K. Muirhead, "Mars Rovers, Past and Future," in *IEEE Aerospace Conference Proceedings*, pp. 128–134, 2004.
- [2] R. A. Lindemann, D. B. Bickler, B. D. Harrington, G. M. Ortiz, and C. J. Voorhees, "Mars Exploration Rover Mobility Development," *IEEE Robotics & Automation Magazine*, 2006.
- [3] B. D. Harrington and C. Voorhees, "The Challenges of Designing the Rocker-Bogie Suspension for the Mars Exploration Rover," in *37th Aerospace Mechanisms Symposium*, (Houston), 2004.
- [4] R. E. Arvidson, J. F. Bell, P. Bellutta, N. A. Cabrol, J. G. Catalano, J. Cohen, L. S. Crumpler, D. J. Des Marais, T. a. Estlin, W. H. Farrand, R. Gellert, J. A. Grant, R. N. Greenberger, E. a. Guinness, K. E. Herkenhoff, J. A. Herman, K. D. Iagnemma, J. R. Johnson, G. Klingelhöfer, R. Li, K. A. Lichtenberg, S. A. Maxwell, D. W. Ming, R. V. Morris, M. S. Rice, S. W. Ruff, A. Shaw, K. L. Siebach, P. A. de Souza, A. W. Stroupe, S. W. Squyres, R. J. Sullivan, K. P. Talley, J. A. Townsend, A. Wang, J. R. Wright, and A. S. Yen, "Spirit Mars Rover Mission: Overview and selected results from the northern Home Plate Winter Haven to the side of Scamander crater," *Journal of Geophysical Research*, vol. 115, p. E00F03, Sept. 2010.
- [5] F. Cordes, A. Dettmann, and F. Kirchner, "Locomotion modes for a hybrid wheeled-leg planetary rover," *2011 IEEE International Conference on Robotics and Biomimetics*, pp. 2586–2592, Dec. 2011.
- [6] B. Wilcox, T. Litwin, J. Besiadecki, J. Matthews, M. Heverly, J. Morrison, J. Townsend, N. Ahmad, a. Sirota, and B. Cooper, "ATHLETE: A Cargo Handling and Manipulation Robot for the Moon," *Journal of Field Robotics*, vol. 24, no. 5, pp. 421–434, 2007.
- [7] S. Cole and M.-J. Vinas, "NASA's GROVER Debuts On Greenland Ice Sheet," 2013.
- [8] P. W. Bartlett, D. Wettergreen, and W. Whittaker, "Design of the Scarab Rover for Mobility & Drilling in the Lunar Cold Traps," in *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, pp. 3–6, 2008.
- [9] P. S. Schenker, T. L. Huntsberger, P. Pirjanian, and E. T. Baumgartner, "Planetary Rover Developments Supporting Mars Exploration, Sample Return and Future Human-Robotic Colonization," *Autonomous Robots*, vol. 14, pp. 103–126, 2003.
- [10] J. E. Bares and D. S. Wettergreen, "Dante II: Technical Description, Results, and Lessons Learned," *The International Journal of Robotics Research*, vol. 18, pp. 621–649, July 1999.
- [11] C. Langley, L. Chappell, J. Ratti, N. Ghafoor, C. Ower, C. Gagnon, T. D. Barfoot, and N. G. Orr, "The Canadian Mars Exploration Science Rover Prototype," in *Proc. 2012 International Symposium on Artificial Intelligence Robotics and Automation in Space.*, 2012.
- [12] J. Hidalgo and F. Cordes, "Kinematics Modeling of a Hybrid Wheeled-Leg Planetary Rover," in *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, (Turin), 2012.
- [13] S. Bartsch, T. Birnschein, R. Malte, J. Hilljegerdes, K. Daniel, and F. Kirchner, "Development of the Six-Legged Walking and Climbing Robot SpaceClimber," *Journal of Field Robotics*, vol. 29, no. October 2008, pp. 506–532, 2012.
- [14] A. Suzumura and Y. Fujimoto, "Control of dynamic locomotion for the hybrid wheel-legged mobile robot by using unstable-zeros cancellation," *2012 IEEE International Conference on Robotics and Automation*, pp. 2337–2342, May 2012.
- [15] C. Grand, F. Benamar, and F. Plumet, "Motion kinematics analysis of wheeled-legged rover over 3D surface with posture adaptation," *Mechanism and Machine Theory*, vol. 45, pp. 477–495, Mar. 2010.
- [16] S. Chamberland, E. Beaudry, L. Clavien, F. Kabanza, F. Michaud, and M. Lauria, "Motion Planning for an Omnidirectional Robot with Steering Constraints," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Taipei), pp. 4305–4310, 2010.

Development of a Low-Cost Vision-Based Localisation System for the Experimental Mars Rover

Anthony Tompkins*, Steven Potiris* and Ali Haydar Göktoğan†

* *School of Aerospace, Mechanical and Mechatronic Engineering*

† *Australian Centre for Field Robotics, J04, The University of Sydney, NSW, Australia, 2006*

Summary: Accurate localisation is a valuable tool for the validation of many autonomous and teleoperated vehicles in both indoor and outdoor environments. For planetary robotic research platforms such as the Experimental Mars Rover (EMR) in the Powerhouse Museum in Sydney, gathering consistent and accurate pose information is essential for validation of experiments in control, planning and teleoperation. This paper describes the development of a vision based, low cost, real-time localisation system, suitable for indoor robotics applications. The system utilises the existing constellation of fiducial markers in the Mars Yard and it is robust to adverse lighting conditions. Pose estimation is obtained by fusing vision data from a single camera and the inertial data from an Inertial Measurement Unit (IMU). The system is successfully integrated with the EMR platform and used for a number of autonomous and teleoperation missions in the Mars Yard.

Keywords: Experimental Mars Rover, Fiducials, IMU, Vision-based pose

Introduction

Simulated environments of Martian and Lunar surfaces provide valuable foundations for research and educational activities. Increasing relevance to both on and off-Earth exploration allows the use of such environments to be exploited without having to directly interact with a foreign location which may be difficult to reach both in terms of cost and time. The Powerhouse Museum's Mars Yard is one such analog environment simulating a Martian terrain and provides both students and researchers alike to perform simulated missions and research with the Experimental Mars Rover (EMR) platform.

Knowledge of a vehicles global position and orientation is essential for many robotics projects operating in simulated environments. Although the EMR has an existing commercial localisation system (CLS), it is often unable to operate in multiple areas of the Mars Yard where adverse lighting is caused by exhibit constraints. Furthermore the system is proprietary and operating system limited, relatively costly to replicate, and therefore difficult to use on additional robotics platforms that may be operated within the Mars Yard. There is thus significant motivation to develop an integrated localisation system using low cost commercial off-the-shelf (COTS) components capable of operating successfully within the environmental constraints of the Mars Yard.

This paper explores the development and evaluation of the resulting system. The following section compares existing vision and inertial sensor based pose estimation techniques. It is followed by a section outlining the experimental components. The steps taken to obtain the

visual pose are then presented. After that, methods used to fuse the vision and IMU data are explained. Finally an evaluation of the system validity and performance are presented. The presented system will be referred to as the Low Cost Localisation System (LCLS)

Related Work

Using artificial visual markers, also known as fiducials, is a common practice for determining pose for use in various areas such as virtual reality, robotics and human-computer interfaces [3] [10]. A primary limitation of vision only systems is inability dealing with abrupt changes in motion. Similarly, vision only systems are limited by their vision sensor's field of view (FOV) and are often hampered by occlusions as well as the vision sensor low-light performance. In order to overcome these limitations, an IMU is integrated to provide high update rates for translational accelerations and angular velocities [6]. In [9] the authors define the methods used for real time tracking and registration of the fiducials in low light. However, they do not show differing lighting conditions which produce system failure. The paper identifies a non-homomorphic image processing method consisting of a Sobel edge detector combined with an experimentally defined reflectance difference between black and white print. Although the system is fused via a statistical framework, other papers quantitatively define the advantage of fusing with IMU [6] [7] [8].

Various applications and methods have been developed in recent times that take advantage of the globally accurate but low rate of vision only systems, and the high sampling rate but long term inaccuracies of IMUs. Considering the vision and IMU fusion approach in detail, two primary methods of taking advantage of vision and inertial sensors have been defined in [1]. In the loosely coupled approach [4] [5], vision and inertial data occurs independently and their results are fused. The inertial system is thus used to speed up fiducial tracking by assisting in next-frame prediction. The tightly coupled sensor fusion method incorporates statistical filtering, such as the Extended Kalman Filter (EKF).

Experimental Setup

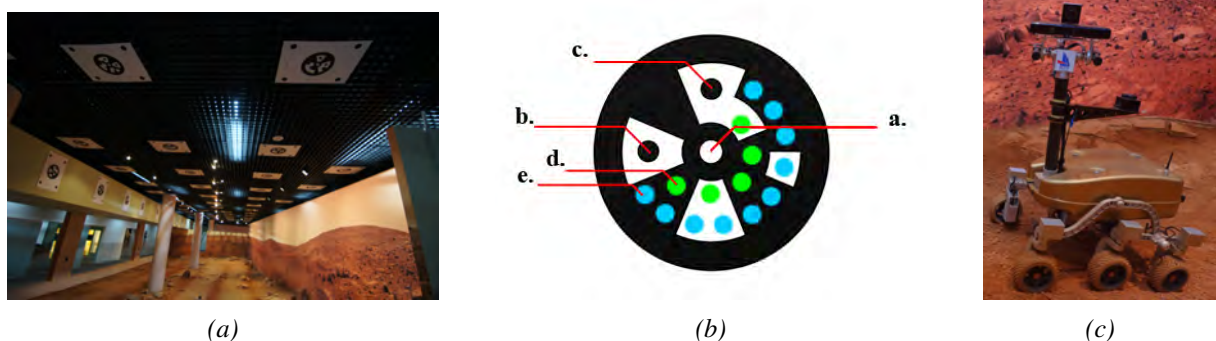


Fig. 1: (a). The Mars Yard, (b). The circular matrix fiducial design used in the Mars Yard. a. Center, b. Start encode, c. End encode, d. Inner data bit ring, e. Outer data bit ring, (c) The Experimental Mars Rover.

The Experimental Mars Rover: The EMR in Fig. 1 (c) is a rocker-bogie type robotics platform equipped with six independently-steered wheels and a mast containing a collection of various sensors. The camera and IMU are rigidly attached to the mast and the EMR body respectively.

Sensors: The navigation system consists of two sensors: A single monocular RGB camera ultra-wide 120° FOV webcam, and an IMU consisting of a 3-axis accelerometer and a 3-axis rate gyroscope. The camera is used to observe the constellation of markers on the ceiling, while the IMU provides high-frequency short-term accurate measurements for inertial navigation and attitude determination. The measurements from both devices are read on board the EMR to extract global pose information which is then sent via the WiFi network to a control station.

Fiducials: Fiducials are identifiers which allow the association of observed data with real world locations. The constellation of fiducials in Fig. 1 (a) on the ceiling are encoded markers which each have a unique 15 bit identifiable encoding. These fiducials enable the EMR to navigate through the Mars Yard as it decodes each fiducial and correlates a real world position registered to each identifier.

Coordinate Frames

The coordinates frames used by the LCLS are defined in Fig. 2 where the R^3 frames are shown with their relative orientations.

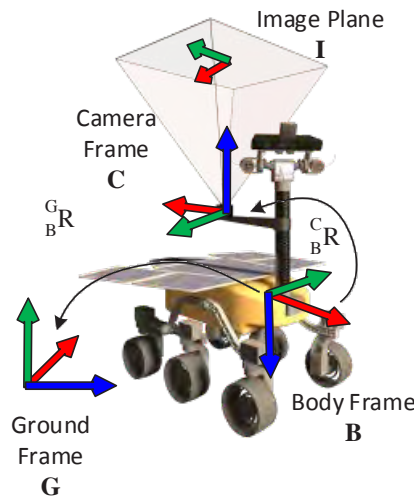


Fig. 2: R^3 coordinate frames in the LCLS

- 1) Ground Frame **G**: This coordinate is fixed to the Mars Yard floor and used as an absolute reference.
- 2) Body Frame **B**: The R^3 coordinate system attached to the body; this frame is collocated the IMU sensor frame.
- 3) Camera Frame **C**: The R^3 coordinate system of the camera is defined with its position at the optical center of the camera and its orientation such that the z-axis is aligned with the optical axis.

- 4) Image Plane I: The R^2 plane on which the camera projects the R^3 scene, which is perpendicular to the optical axis of the camera and is offset along the optical axis from the optical center by the focal length.

Image Processing

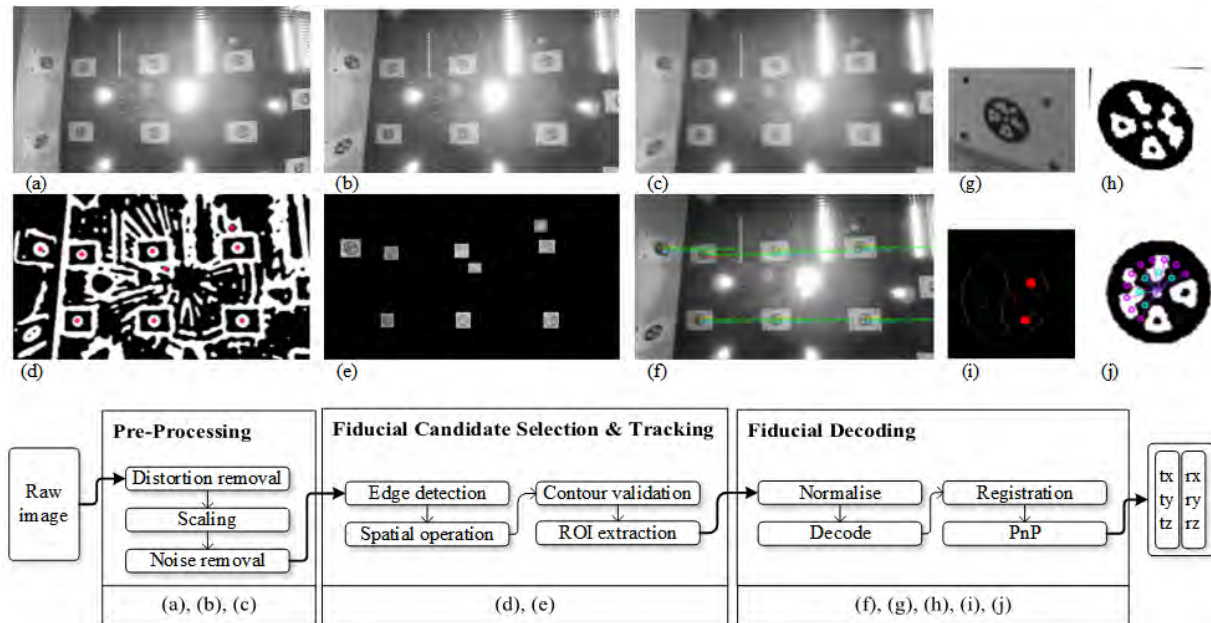


Fig. 3: Image processing algorithm overview. (a) Raw image with barrel distortion, (b) Processed image with distortion removed, (c) Fiducial image after applying Gaussian blur and resizing, (d) Fiducial Image after edge detection, (e) ROI extraction for potential fiducials, (f) Final candidate fiducials decoded, (g) Raw image of candidate fiducial, (h) Candidate ROI and Otsu, (i) Rotation and skew normalisation, (j) Fiducial decoding

Pre-Processing

Distortion Removal: A raw colour image frame is captured from the camera and converted to grayscale. The grayscale image is undistorted using the camera intrinsic parameters found by calibration to remove the barrel distortion created by the ultra-wide angle lens. This removes any adverse effects of distortion in the proceeding tracking and decoding processes and is shown in Fig. 3(b).

Scaling: The image is resized by a scale factor s_f to create the primary search image. The resolution of the input image is $1280 \times 720px$. Scaling the original image reduces the computation for the next stages in the process.

Noise Removal: In low light conditions the camera is prone to high-frequency noise. As this noise affects the detection rate, a Gaussian blur is applied globally to the image to remove any high-frequency noise. A median blur is applied to the raw image with a convolution window of $3 \times 3px$ to achieve a similar effect of noise mitigation. Median blurring is used in this instance to retain overall feature detail while filtering high-frequency noise. The result can be seen in Fig. 3(c).

Fiducial Candidate Selection and Tracking

Edge Detection: A local neighbourhood pixel-based adaptive threshold process is applied to the image. This returns a binary image exaggerating the shapes of the fiducial contours.

Spatial Operations: A further median blur of window $5 \times 5px$ is performed followed by a morphological erosion of kernel size $3 \times 3px$. This process removes a large portion of the outliers in the edge detection stage, reduces the candidate set size and speeds up image processing. The result can be seen in Fig. 3(d).

Contour validation: A border following contour [12] is applied to the result of the morphological processes. After contours have been extracted, the image is evaluated against various metrics to further classify contours as fiducials:

- *Contour area:* Knowledge of the camera distance from the fiducials causes all observed fiducials to fall within an experimentally defined pixel size $[60, 400]px^2$.
- *Area ratio:* A least-squares fitting ellipse obtained on the edge image pixels in a local window around the candidate fiducial. This ellipse is characterised by its major and minor axes lengths which are denoted e_M , and e_m . The equation

$$M_e(a_f, e_M, e_m) = \left| \frac{a_f}{\frac{\pi}{4}l_1l_2} - 1.00 \right| \leq \varepsilon \quad (1)$$

is a condition [9] indicating whether or not a candidate fiducial has an area representative of an ellipse with equal area, where $\varepsilon = 0.30$ corresponding to a 30% threshold.

- *Roundness:* The function

$$M_r = \frac{4\pi a_f}{p_f^2} \quad (2)$$

is a metric for the roundness of a contour given the fiducial pixel area and fiducial contour perimeter p_f and aids in rejecting false positives given by the previous step. A result of $M_r = 1.00$ corresponds to a perfect ellipse while $M_r = 0.70$ corresponds to a relatively smooth ellipse and $M_r < 0.5$ corresponds to an extremely rough ellipse.

High resolution ROI extraction: Given the previously found least-fit ellipse the Region of Interest (ROI) is found for each candidate fiducial. The low resolution ROI is padded with a constant ratio of pixels and scaled back to the original size to obtain the high resolution ROI. The result of this operation is observed in Fig. 3(e)

Fiducial normalisation: After the high resolution ROI is acquired, this ROI has another contour search performed for the fiducial as well as the start and stop bits. The start and stop bits are defined as the two smallest contours that are not at the centre. The fiducial outline is the largest contour within the ROI. The normalisation process is split into two parts:

1) Skew normalisation:

- Obtain the least fit ellipse giving the major and minor axis
- Find the ratio between the major and minor axis
- Scale along the major axis by the ratio found in 2

2) Rotation normalisation:

- Obtain the signed angle θ_{ss} between the two start or stop bit contours (c_{1x}, c_{1y}) and (c_{2x}, c_{2y}) , and the ROI centre (c_x, c_y) .

- Select the start and stop bits as follows:

$$(b_1, b_{15}) = \begin{cases} (c_{1x}, c_{1y}), (c_{2x}, c_{2y}) & \text{if } 0 < \theta_{ss} \leq 180 \\ (c_{2x}, c_{2y}), (c_{1x}, c_{1y}) & \text{if } -360 < \theta_{ss} < -180 \end{cases}$$

where b_1 is the 'start bit' and b_{15} is the 'stop bit'.

- Once the start and stop bits have been determined the rotation is normalised such that the start bit is always at (b_{1x}, c_y) , in line with the centre of the fiducial.

Least fit ellipse: The detected contours are significantly pixelated which results noisy centroids which are not optimum for tracking the candidate fiducial. To combat the noisy contour, a least squares ellipse is fitted to the major contour. The centre of this ellipse is henceforth considered as the high resolution tracking centroid.

Fiducial Decoding

Thresholding: Once a candidate fiducial is normalized, it is ready to be processed for its unique value. One of the major considerations for the system robustness is the ability to deal with adverse lighting scenarios created by flaring light sources and lighting transitions. Otsu/s thresholding method [11] allows us to create an optimum automatic global threshold on a per-fiducial basis. This method is superior to whole-image global thresholding that cannot deal with changing lighting conditions.

Decoding: There are 15 standard locations. The first five bits are the inner ring each offset by 45° , and the remaining 10 bits are made up of the outer ring each offset by 22.5° . The radius of the bit information is calculated dynamically based on the ratio of the center of the encoded bits to the radius of the fitted ellipse centroid of the start bit. The binary identifier is extracted by iterating through each of the 15 possible positions and adding to a list the boolean value corresponding to the pixel color at each position; 0 if the pixel colour is black or 1 if the pixel color is white. Once all the 15 possible positions have been sampled and the corresponding boolean bit added to the bit list it is treated as a binary number and then converted to a base-10 unsigned integer. Fig. 1(b) shows the standard positions of the bits on the fiducial.

Fiducial Registration: The 15 bit fiducials are identified and registered to 3D positions with respect to the ground frame. Fig. 3(g,h,i,j) show the result of the fiducial decoding operations.

Perspective-n-Point Localisation The solution of the Perspective-n-Point (PnP) problem [13] yields the ground frame position and orientation with respect to the a camera. A set of at least four 3D-to-2D correspondences between known features in the reference frame and pixel positions occupy the 2D projection of the features on to the image frame. To solve the algorithm requires n 3D feature positions gF_j in the ground frame. $j = 1 \dots n$, n represents the number of fiducials in frame, and ${}^i f_j$ represents the associated 2D pixel positions of the features. It is then possible to localise the ground frame with respect to the camera frame and obtain position ${}^c p_g$ and orientation ${}^c R$ with respect to the ground frame.

One solution to the PnP problem is the ePnP (Efficient Perspective-n-Point) algorithm; in this project the OpenCV implementation of the ePnP algorithm is used. Finally a Median Absolute Deviation (MAD) filter [14] is applied to the PnP output to remove outliers.

$\alpha\beta$ Filter

The $\alpha\beta$ F [2] is a filter that can estimate two states where one state is the integral of the other. In this case two filters are used; one to estimate the translation state and one to estimate the rotational state. The respective translational and rotational state vectors are

$$x_T = [p^\top v^\top]^\top \quad (3) \quad x_R = [\theta^\top \omega^\top]^\top \quad (4)$$

where p , v , θ and ω are the position, velocity, orientation Euler-angles and angular velocity respectively. In a traditional $\alpha\beta$ F observer for a position-velocity state vector the infrequent position measurements correct the position and velocity estimates, and the frequent prediction stage estimates the position while keeping the velocity estimate constant between measurement intervals. The prediction stage is modeled by equations (5) and (6), and the update stage involves first determining the residual in equation (7) then the state correction is modeled by the equations (8) and (9).

$$\hat{p}_k = \hat{p}_{k-1} + \hat{v}_{k-1} dt_P \quad (5) \quad \hat{v}_k = \hat{v}_{k-1} \quad (6)$$

$$\hat{r}_T = p - \hat{p} \quad (7) \quad \hat{p}_k = \hat{p}_{k-1} + \alpha_T \hat{r}_T \quad (8) \quad \hat{v}_k = \frac{\beta_T \hat{r}_T}{dt_U} \quad (9)$$

where α and β are the filter correction gains, $\hat{\cdot}$ refers to the estimate, subscripts $k-1$ and k refer to the time before and after the prediction or update and dt_P and dt_U are the time between predictions and updates respectively, and ${}^G_B R$ is the Direction Cosine Matrix (DCM) of orientation $\hat{\theta}_{k-1}$ that transforms the body coordinates into global coordinates. For the orientation estimation $\alpha\beta$ F, a similar model is used where the prediction stage is modeled by:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \hat{\omega}_{k-1} dt_P \quad (10) \quad \hat{\omega}_k = \hat{\omega}_{k-1} \quad (11)$$

and the update stage involves first determining the residual in equation (12) then the state correction is modeled by the equations (13) and (14).

$$\hat{r}_R = \theta - \hat{\theta} \quad (12) \quad \hat{\theta}_k = \hat{\theta}_{k-1} + \alpha_R \hat{r}_R \quad (13) \quad \hat{\omega}_k = \hat{\omega}_{k-1} + \beta_R \hat{r}_R dt_U \quad (14)$$

$$\hat{v}_k = {}^G_B R^B a_m dt_U \quad (15) \quad \hat{\omega}_k = \hat{\omega}_{k-1} + {}^G_B R^B \omega_m dt_P \quad (16)$$

where a_m is the measured acceleration in the body frame and ${}^B \omega_m$ is the rate gyro measurement in the body frame. This model uses only the camera PnP pose estimate and does not take advantage of the IMU and its high-frequency updates. The model is subject to significant error during camera outages. A change is made to the prediction models to include the IMU accelerometer and rate gyroscope data to achieve smoother position and orientation estimates during camera outages: states \hat{v}_k and $\hat{\omega}_k$ will be provided from the integral of the accelerometers over the last IMU update for each translational and rotational states. The velocity prediction then becomes equation (15) and the angular velocity is propagated by (16).

Results and Discussion

This section will both qualitatively and quantitatively show the results of experiments performed with the unfiltered PnP, outlier rejected PnP, and $\alpha\beta$ filtered PnP. Unless specified otherwise, the following results were recorded on the EMR in the Mars Yard with simultaneous acquisition from both the CLS and the LCLS. Where relevant the X-Y plots of the position and orientation of the EMR have been plotted on a close-to-accurate map of the Mars Yard.

Adverse Lighting



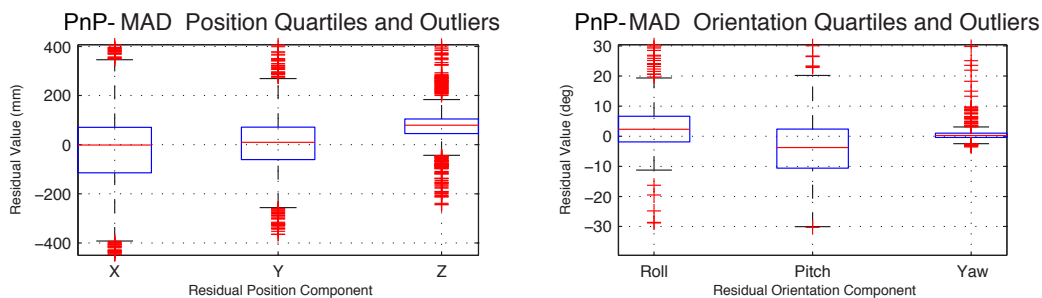
Fig. 4: Comparison between Otsu and global thresholding to identify fiducials

It was observed that in the docking area as well as besides some tall obstacles the CLS did not function. Low lighting conditions were not handled well by the CLS whereas the LCLS was more robust in all areas when dealing with adverse lighting conditions. When observing the CLS image feed, it is almost completely black, unlike the LCLS camera which achieves a noisy but visible image. To verify the validity of Otsu thresholding on fiducial region of interests, it was compared against a typical global threshold. Fig. 4 shows that Otsu is able to successfully threshold all cases across the image plane.

Unfiltered PnP Localisation

Table 1: Statistics for PnP-MAD position and orientation residuals

PnP-MAD	$p_x(m)$	$p_y(m)$	$p_z(m)$	Roll ($^{\circ}$)	Pitch ($^{\circ}$)	Yaw ($^{\circ}$)
Mean	-0.03	-0.01	0.07	1.87	-4.14	4.26
Standard Deviation	0.22	0.20	0.07	9.78	8.52	1.08



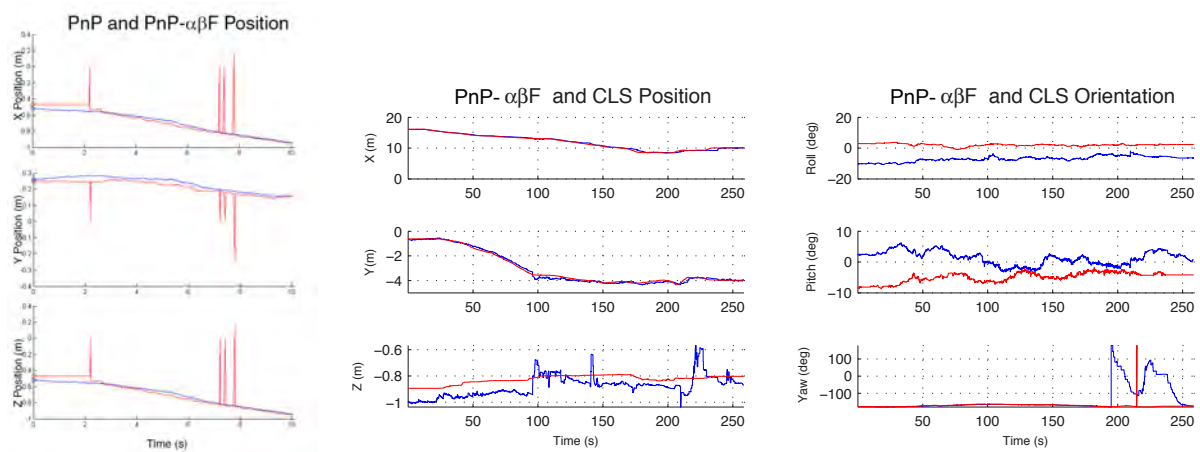
(a) Statistics for PnP-MAD position error residual (b) Statistics for PnP-MAD orientation error residual

Fig. 5: Statistics for PnP-MAD error residuals

The unfiltered PnP localisation method works well and PnP position generally follows the CLS position however it is subject to significant noise of often several metres. The following experiments are conducted with the same PnP localisation system, however outliers

are rejected using the MAD algorithm. By examining Fig. 6 (b) it is evident that the X and Y positions very closely match the CLS positions, however the Z position contains less magnitude in noise than the other components. Fig. 6 (c) shows the rotational components are offset by a constant angle in the roll and pitch angles, however the yaw angle seems to very closely match the CLS orientations. This constant offset in rotation may be attributed to the physical misalignment of the CLS and LCLS. Table 1 and 5 show the residual error between the CLS and the PnP-MAD path. The outliers are skewed differently in each component of position and rotation; this may be attributed to the geometry of the path and the heading of the EMR. The standard deviations reflect the accuracy of the PnP-MAD system compared with the CLS.

PnP- $\alpha\beta$ Filter



(a) PnP (red) and PnP- $\alpha\beta F$ position (blue) vs. time (b) Position of PnP- $\alpha\beta F$ (blue) and CLS (red) vs. time (c) Orientation of PnP- $\alpha\beta F$ (blue) and CLS (red) vs. time

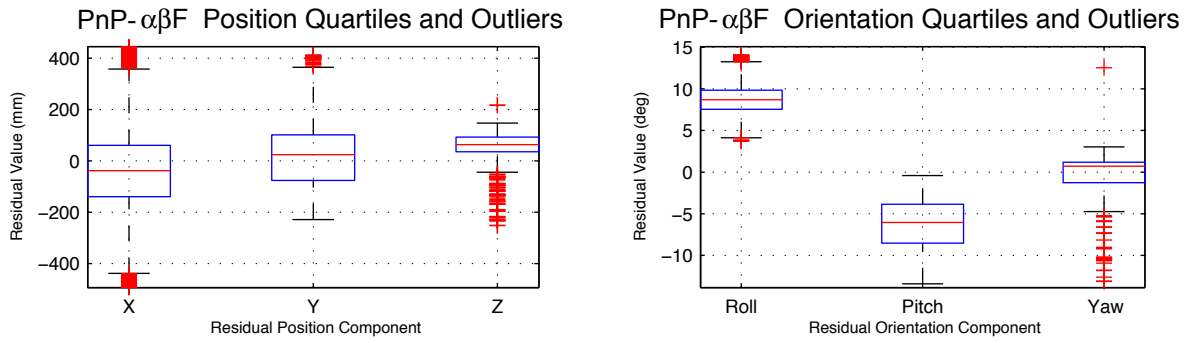
Fig. 6: (a) short and (b), (c) extended runs of position and orientation propagation with an $\alpha\beta$ filter

Table 2: Statistics for PnP- $\alpha\beta F$ position and orientation residuals

PnP- $\alpha\beta F$	$p_x(m)$	$p_y(m)$	$p_z(m)$	Roll (°)	Pitch (°)	Yaw (°)
Mean	-0.03	0.02	0.05	8.99	-6.25	-3.02
Standard Deviation	0.22	0.12	0.06	2.13	3.15	1.56

The results of a small-scale test are presented for proof-of-concept, then the results from the implementation of the filter on the EMR are evaluated. Fig. 6 (a) shows the result of a small scale test done with the same IMU and camera as used for the rest of these experiments. The figure shows the $\alpha\beta$ filtered position is noticeably smoother than the PnP equivalent which contains four large spikes that the $\alpha\beta$ filter is unaffected by. At about $t = 8s$ the $\alpha\beta F$ position follows the PnP and ‘jumps’ from one position to another. The spikes that occur at $t = 2, 10$ seconds do not influence the $\alpha\beta$ filter at all; the state continues to be propagated by the accelerometer and then the next available vision update that is not rejected is applied and any error accumulated from the accelerometer drift is corrected.

Comparing the results of the the PnP-MAD system in Fig. 5 and Table 1, with the the PnP- $\alpha\beta F$ system in Fig. 7 and Table 2 it is evident from the standard deviation that the PnP- $\alpha\beta F$ system more closely matches the CLS. One very important consideration is the discontinuous



(a) Statistics for PnP- $\alpha\beta F$ residual position error (b) Statistics for PnP- $\alpha\beta F$ residual orientation error

Fig. 7: Statistics for PnP- $\alpha\beta F$ error residuals

representation of the orientation as Euler angles; this becomes evident in the standard deviation and mean of the yaw component where the data is orders larger than the other orientation components due to the two angle being on either side of the domain $[-180, 180]^\circ$. Also important is to notice the mean value; this includes in it the offset of the cameras between the LCLS and the CLS, however it is not a good estimate of the offset because of the constant EMR heading rotation during the path following. Evaluating the standard deviation results of the PnP-MAD against the PnP- $\alpha\beta F$ system we deduce the following: in 1 and 2 there is a 0%, 40%, 14% improvement in the respective x, y, z translations and a 78%, 63%, -44% improvement in the respective roll, pitch, yaw orientations. The reason there is a degradation in the yaw is due to the absence of accounting for angle wrap when calculating the standard deviation. This would be a consideration for future testing. Similarly there was the expectation of an improvement in all translations, however there was none in the x. This may be attributed to the above comparison being only a single run. In future work it would be valuable to include extended comparisons across multiple runs to more confidently confirm performance. Furthermore, the disparity in the results may be attributed to different magnitude biases between the accelerometers and the rate gyroscopes. In general the magnitude of bias in an accelerometer will be orders higher than a rate gyroscope, and as such the position is effected by bias drift much more than the rate gyroscopes due to the double integration of the accelerometer outputs to achieve position.

$\alpha\beta F$ Vision-Denied Results

Experiments were done to see the effect of vision loss on the $\alpha\beta F$; this involved covering the lens of the camera of the LCLS for a short period. During this time the position is being propagated only by the accelerometer dead-reckoning.

Fig. 8 and Table 3 show the result of vision denials on position for various short periods of time; for each path the LCLS deviation from the CLS path is measured. In Fig. 8 there are two vision denials to the LCLS (circled in magenta); note that the CLS also deviates when denied vision and similarly requires its IMU to propagate the position. The results in Table 3 were achieved by comparing the uncovered CLS with the covered LCLS.

Experiments of increasingly longer vision denials were run within 2 minutes of each other. The time between experiments is important as larger differences would cause changes in

Table 3: Table of vision denial results

Denial Time (s)	Maximum Deviation (cm)
2	10.24
3	21.64
7	47.01

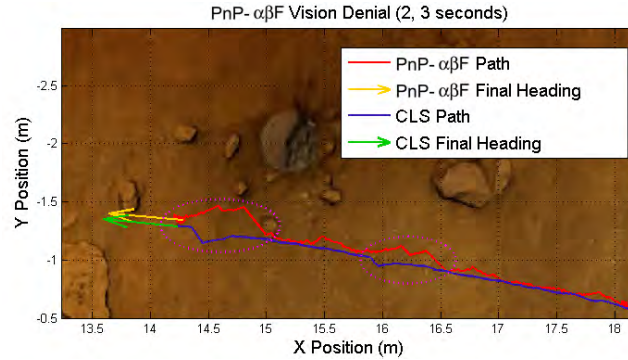


Fig. 8: $\alpha\beta F$ (red) with vision loss of 2 and 3 seconds

accelerometer and rate gyroscope bias. Finally, the linear relationship between denial time and maximum deviation is both intuitive and expected as it is driven primarily by the last known velocity plus the result of the slow random walk process inherent in the accelerometer bias.

Conclusion and Future Work

Considering the experimental results of the LCLS, it can be seen that the use of a vision and IMU based provides a valid method for pose determination within the constrained environment of the Powerhouse Museum Mars Yard. The presented system successfully fuses the globally accurate visual pose with the short-term accurate IMU information. It has been shown that the visual pose method allows for superior detection to the CLS in the same lighting conditions allowing the LCLS to operate in more lighting adverse areas of the Mars Yard.

Although ultimately successful, there exist various shortcomings of the navigation system. The primary shortcoming is the systems relative residual error to the assumed ground truth which causes positional errors occasionally in excess of 0.4m. Although these offshoots are intermittent, an improved statistical filtering framework based on the Extended Kalman Filter as opposed to $\alpha\beta F$ would provide superior results. Therefore an improved filtering system would be a priority as an extension of the existing system.

Acknowledgments

The authors would like to thank Mr Daniel Wilson from the Australian Centre for Field Robotics (ACFR) for his support throughout the project. This work was partially supported by the Broadband Enabled Education and Skills Services Department of Education program,

References

- [1] Corke, P. and Lobo, J. and Dias, J., “An introduction to inertial and visual sensing”, *International Journal of Robotics*, Vol. 26, pp. 519–535, 2007
- [2] Hussain, Akbar, Hicks, David, Ortiz-Arroyo, A Haq, Shaiq, Ahmed, Zaki, “A Case Study: Kalman & Alpha-Beta Computation under High Correlation”, *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. I, pp. 19–21, 2008
- [3] Hoff, W. and Nguyen, K. and Lyon, T., “Computer Vision-Based Registration Techniques for Augmented Reality”, *Proceedings of IRCV SPIE*, Vol. 2904, pp. 538–548, 1996
- [4] Hwangbo, M. and Kim, J.-S. and Kanade, T., “Gyro-aided feature tracking for a moving camera: Fusion, auto-calibration and gpu implementation”, *International Journal of Robotics Research*, Vol. 30, pp. 1755–1774, 2011
- [5] Klein, G.S.W. and Drummond, T.W., “Tightly integrated sensor fusion for robust visual tracking”, *Journal of Image and Vision Computing*, Vol. 22, pp. 769–776, 2004
- [6] Ligorio, G. and Sabatini, A. M., “Extended Kalman Filter-Based Methods for Pose Estimation Using Visual, Inertial and Magnetic Sensors: Comparative Analysis and Performance Evaluation”, *Sensors Journal, IEEE*, 2013
- [7] Mirzaei, F. M. and Roumeliotis, S. I., “A Kalman Filter-based Algorithm for IMU-Camera Calibration”, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007
- [8] Mirzaei, F. M. and Roumeliotis, S. I., “A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation”, *IEEE Transactions on Robotics*, 2008
- [9] Naimark, L. and Foxlin, E., “Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker”, *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, pp. 27–36, 2002
- [10] Neumann, U. and Cho, Y., “A Self-Tracking Augmented Reality System”, *ACM VRST*, 1996
- [11] Otsu, Nobuyuki, “A Threshold Selection Method from Gray-Level Histograms”, *Systems, Man and Cybernetics, IEEE Transactions on*, Vol. 9, No. 1, pp. 62–66, 1979
- [12] Suzuki, S and Abe, K., “Topological Structural Analysis of Digitized Binary Images by Border Following”, *CVGIP*, Vol. 30, No. 1, pp. 32–46, 1985
- [13] Lepetit, Vincent and Moreno-Noguer, Francesc and Fua, Pascal, “Epnnp: An accurate o (n) solution to the pnp problem”, *International Journal of Computer Vision*, Vol. 81, No. 2, pp. 155–166, 2009
- [14] Ruppert, David “Modeling Univariate Distributions”, *Statistics and Data Analysis for Financial Engineering*, pp. 118–110, 2010

A Vision-Based Autonomous Docking System of the Experimental Mars Rover

Steven Potiris*, Anthony Tompkins* and Ali Haydar Göktoğan†

* *School of Aerospace, Mechanical and Mechatronic Engineering*

† *Australian Centre for Field Robotics, The University of Sydney, Rose Street Building J04, Sydney, NSW, 2006, Australia*

Summary: The Experimental Mars Rover (EMR) at the Mars Yard in the Powerhouse Museum, Sydney, is a battery powered platform being used extensively by experienced engineers and researchers from universities as well as teleoperated by high school students from different states of Australia. In order to maximise the EMRs operational time, occasionally, its batteries need to be recharged. Connecting the EMR to the charging station requires a set of delicate manoeuvres which are difficult to perform for inexperienced users. In this paper we present a vision-based, autonomous docking system for the EMR for use in the Mars Yard. The system is closely integrated with the EMR and its teleoperation system. The autonomous charging process can be initiated by either the EMR when its battery charge is below a predetermined level or by a human operator via the teleoperation system. Once the process is activated, the system calculates an obstacle free, feasible route from the rovers current position and orientation to the charging station. The control laws for the docking manoeuvres utilise the highly flexible mobility subsystem and brings the rover to the docking bay in a particular heading suitable for charging. This paper also shows the results of docking experiments successfully performed at the Mars Yard.

Keywords: Vision-based Docking, Autonomous Docking, Mars Rover, Terrain Characterisation

Introduction

This paper develops the guidance and control system of the Experimental Mars Rover (EMR) to plan an obstacle free, feasible path in a controlled research environment from the rovers current position and orientation to its docking bay and track the planned trajectory. In order to maximise the EMRs operational time occasionally its batteries need to be recharged. Returning the EMR to its docking bay in the correct position and heading requires a series of delicate manoeuvres which may be difficult for inexperienced users to perform. By implementing an autonomous docking system for the EMR the requirement of an experienced external operator to be present is removed. The value of the EMR is sustained by ensuring it is charged and returned to its docking station after each experiment, ready for the next researcher to use. Many planetary rover guidance systems have used elevation maps [1] and roughness maps [2] in order to characterise terrain as part of minimising risk to rover failure. Also common in planetary rover control systems is the implementation of Proportional-Integral (PI) controllers [3] to follow time invariant paths. The use of monocular cameras in vision-based docking has been researched in applications where the vision system tracks either a fiducial marker or a salient feature representing a docking station [4] [5]. More extensive literature develops the

docking control for mechanically and electrically connecting the charging mechanism [12]. This paper details the development of a system for use in the Mars Yard that can characterise the terrain from an *a priori* 3D model, calculate a reduced-risk path, generate and track the trajectory to the docking bay, and finally perform adjustments to achieve a suitable docking pose.

This article first introduces the EMR, the Mars Yard, the on-board navigation system and the docking procedure in *System Setup and Environment*. The terrain characterisation, path planning and trajectory generation processes is detailed in *Guidance System*. *Control System* introduces the mobility subsystem of the EMR and details the PI controllers for both the trajectory follower and the spot-turn manoeuvre. The guidance and control systems results are presented and discussed in *Results and Discussion*. In *Conclusion and Future Work* the paper is evaluated with respect to the aims of the research and future work is set forth.

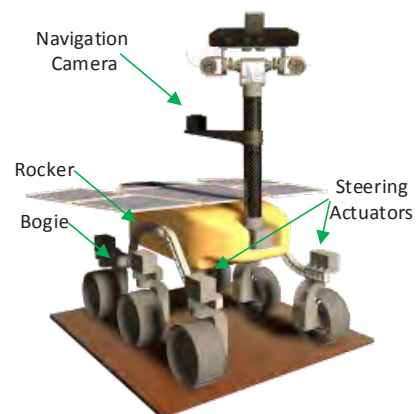
System Setup and Environment

The Mars Yard is a simulated Martian environment in the Powerhouse Museum, Sydney and is a controlled research environment in which the EMR operates. It contains obstacles and hazards that would be found on the Martian surface such as traversable and untraversable rocks, a crater, hills and loose rocks, and in addition stromatolites and hematites for education and research purposes. In this environment the EMR is used by both high school students as an educational platform, and by researchers for experiments in Guidance, Navigation, Control (GNC), autonomy and developing sensing systems, amongst other research areas.

The EMR is a robotic platform equipped with six independently steered wheels on a rocker-bogie mechanism. It has a mast carrying a set of mission sensors, including an upwards-facing monocular camera used for navigation.



(a) The EMR in the Mars Yard traversing a small rocky obstacle



(b) A 3D rendering of the EMR with labelled components

Fig. 1: The Mars Yard with fiducial constellation on the ceiling

The rocker-bogie mechanism makes the EMR highly mobile and able to traverse small obstacles while keeping the roll and pitch angles of the body relatively low. In Figure 1(a)

the EMR is pictured traversing small rocky obstacles without significant roll or pitch. This is important for the navigation system camera as it is mounted rigidly to the mast of the EMR and must remain facing upright towards the fiducials on the ceiling.

Docking Procedure

Autonomously docking the EMR in order to have its charger connected involves three processes:

- 1) Calculating the path to the docking bay from the current EMR position
- 2) Generating and tracking the trajectory to an area in close proximity of the docking bay
- 3) Accurately controlling the EMR heading to a suitable heading for connecting the charger

Once the EMR has reached an acceptable position and heading in the docking bay it is ready to have its charger connected.

Vision-Based Navigation System

The navigation system of the EMR uses fiducial markers on the ceiling of the Mars Yard to localise the EMR with respect to a ground reference frame. It is aided by an Inertial Measurement Unit (IMU) to increase the frequency and robustness of the pose determination. While in motion, the navigation system is accurate to within $\pm 10\text{ cm}$ of the true position of the EMR. The navigation system provides the control system with updates at 50Hz of the EMR states: the EMR heading ψ_m and the EMR body position expressed in the ground frame, ${}^g\mathbf{p}_b$. In field applications where fiducials are not available, pose measurements may be acquired from a sun sensor [7], star trackers and fused with wheel odometry [8] and inertial measurements. Other methods include matching scans from a 3D laser range finder with a pre-existing Digital Surface Model (DSM) [9][10].

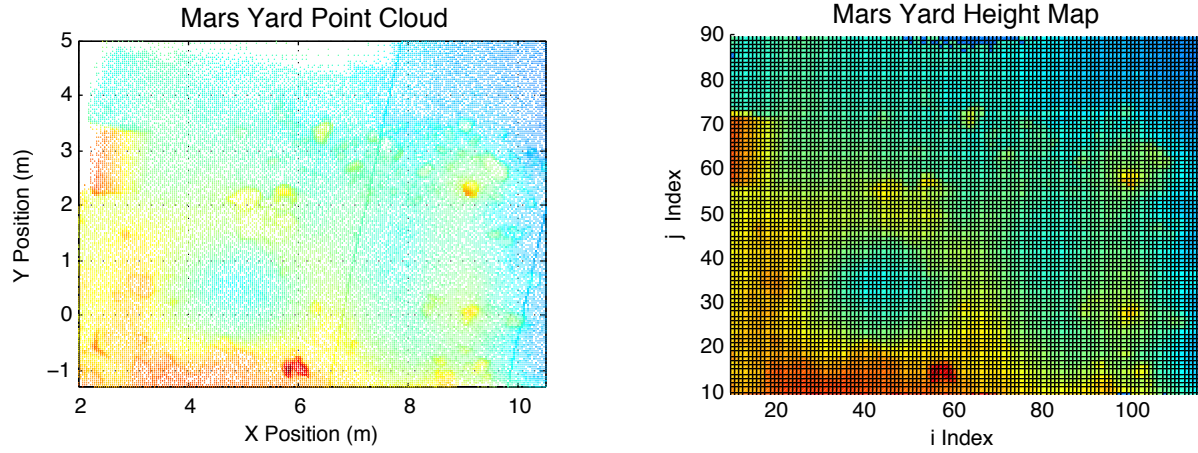
Guidance System

In order to autonomously move the EMR from its current position to the docking bay a safe, obstacle free path must be determined. The guidance system evaluates the heuristic and cost functions over the discretised terrain, produces a cost map and identifies the minimum-cost path. The cost function incorporates the risk of the EMR colliding with or becoming stuck on obstacles; the most common way this can occur is by having the wheel brackets become stuck while trying to traverse large obstacles. As the Mars Yard is indoors and limited in size there are no areas with slope that can pose any significant threat of tip-over.

Terrain Discretisation and Characterisation

The Mars Yard terrain is given *a priori* as a high-polygon 3D model. The 3D vertices are extracted and used as a point cloud \mathbf{T} which is pictured in Fig. 2(a). In field applications

such a point cloud could be assembled using an on-board 3D ranging laser scanner, a stereo vision camera pair [11] or at a larger scale using DSM extraction from multi-modal satellite imagery [6]. The terrain is discretised into a grid \mathbf{G} with node sizes equal to the smallest expected obstacle; in this case a node size of $75 \times 75\text{mm}$ is used. Fig. 2 shows the point cloud and resultant grid.



(a) A portion of the point cloud extracted from 3D model of the Mars Yard

(b) A grid representation of a portion of the Mars Yard

Fig. 2: Transformation of terrain from a point cloud to a grid based representation. Hotter colours represent high terrain height and cool colours represent lower terrain height

The 3D point cloud \mathbf{T} is evaluated over the grid \mathbf{G} and each grid node $\mathbf{G}_{i,j}$ is assigned a cost vector $\mathbf{c}(i,j)$, where each element in the vector represents a different normalised terrain characteristic. The subset of points in \mathbf{T} that lie in grid node $\mathbf{G}_{i,j}$ is denoted $\mathbf{P}_{i,j}$.

A number of the terrain characteristics can be found by fitting a plane to the local set of points $\mathbf{P}_{i,j}$. This can be accomplished using orthogonal residual regression, in which the set of points $\mathbf{P}_{i,j}$ are fit by least squares minimisation [13] to a plane $\mathbf{S}_{i,j}$ represented by the equation (1) where $\mathbf{n}_{i,j}$ is the normal vector of the plane fit to $\mathbf{P}_{i,j}$ and $\mathbf{r}_{i,j}$ is the geometric centroid of $\mathbf{P}_{i,j}$.

$$(\mathbf{P}_{i,j} - \mathbf{r}_{i,j}) \cdot \mathbf{n}_{i,j} = 0 \quad (1)$$

The mean of the individual position components of the points in $\mathbf{P}_{i,j}$ is the geometric centroid $\mathbf{r}_{i,j}$. The height characteristic of the grid node at index (i,j) is given by the Z component of $\mathbf{r}_{i,j}$. The residuals $\bar{\mathbf{p}}_{i,j}$ are given by equation (2) and then forms the vector equation (3) which represents a plane formed by the residuals.

$$\bar{\mathbf{p}}_{i,j} = \mathbf{P}_{i,j} - \mathbf{r}_{i,j} \quad (2) \quad \bar{\mathbf{p}}_{i,j} \cdot \mathbf{n}_{i,j} = 0 \quad (3)$$

The sum of squares of the residuals is given by the equation (4) which holds true **iff** the \mathbb{R}^3 set of points $(\bar{\mathbf{p}}_{i,j}^\top \bar{\mathbf{p}}_{i,j})$ are rank deficient; in which case the set $\bar{\mathbf{p}}_{i,j}$ must be coplanar, assuming that the set is not strictly collinear or a single unique point.

$$(\bar{\mathbf{p}}_{i,j}^\top \bar{\mathbf{p}}_{i,j}) \cdot \mathbf{n}_{i,j} = 0 \quad (4) \quad (\bar{\mathbf{p}}_{i,j}^\top \bar{\mathbf{p}}_{i,j}) \cdot \mathbf{n}_{i,j} = \lambda \mathbf{n}_{i,j} \quad (5)$$

By solving equation (4) the normal $\mathbf{n}_{i,j}$ to the least-squares best-fit plane $\mathbf{S}_{i,j}$ is found. In practice equation (4) is commonly solved by singular value decomposition [14] to find $\mathbf{n}_{i,j}$ computationally, where $\mathbf{n}_{i,j}$ is the smallest singular value of the eigenvector equation (5). The

first and second components of the plane normal $\mathbf{n}_{i,j}$ are the slope of the plane at grid node (i, j) in the x and y directions, $S_X(i, j)$ and $S_Y(i, j)$ respectively. The slopes S_X and S_Y are evaluated at every node in \mathbf{G} and are used to form the directional slope maps $C_{SX}(i, j)$ and $C_{SY}(i, j)$. A linear combination of $C_{SX}(i, j)$ and $C_{SY}(i, j)$ is used depending on the proposed direction of travel during the path cost determination. The height map $C_H(i, j)$ indicates the normalised vertical displacement of the terrain. The surface roughness $C_R(i, j)$ quantifies how rough the terrain is at each node. The residuals given by equation (6) are computed and projected on to the vector normal of the plane $\mathbf{S}_{i,j}$ given by equation (7). The value of the surface roughness at (i, j) is then given by equation (8) [13].

$$\hat{\mathbf{q}}_{i,j} = \mathbf{P}_{i,j} - \mathbf{r}_{i,j} \quad (6) \quad \hat{\mathbf{n}}_{i,j} = \frac{\mathbf{n}_{i,j}}{|\mathbf{n}_{i,j}|} \quad (7) \quad C_R(i, j) = std(\hat{\mathbf{n}}_{i,j} \cdot \hat{\mathbf{q}}_{i,j}) \quad (8)$$

The remaining characteristic map $C_B(i, j)$ is read from an *a priori* image where the binary pixel intensity $C_B(i, j)$ corresponds to the grid node $\mathbf{G}_{i,j}$. This map provides input to the guidance system which enforces a policy that forbids the system from planning a path through the high-intensity areas. The cost map $C_B(i, j)$ is a convenient way to change the behaviour of the path planning algorithm and is much less costly than to reconstruct the *a priori* 3D model in the case of significant changes to the terrain. The characteristic maps are shown in Fig. 3 for reference.

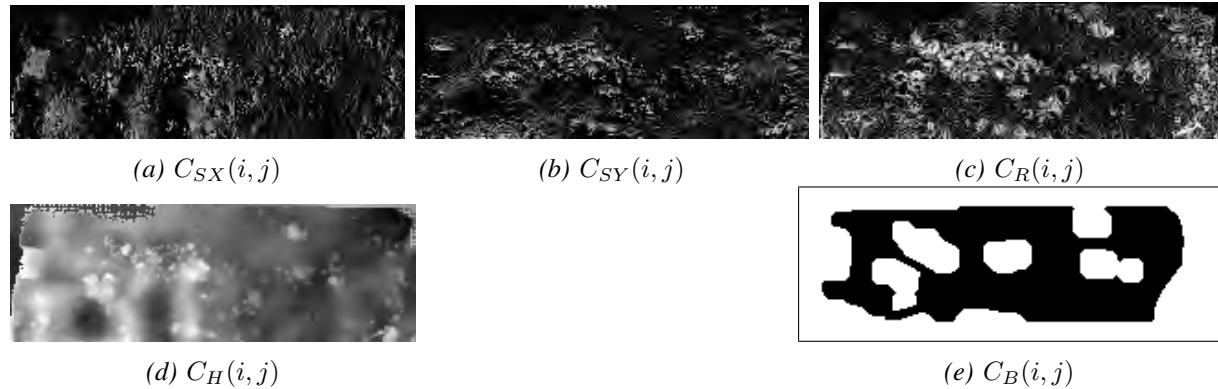


Fig. 3: The evaluated cost maps of the Mars Yard

Heuristics, Costs Functions and Path Planning

Using the cost maps $C_H(i, j)$, $C_R(i, j)$, $C_{SX}(i, j)$ and $C_{SY}(i, j)$ the A* algorithm is used to calculate a path from the current position of the EMR to the docking bay. The path planning algorithm evaluates the terrain grid using cost function $g(x_1, x_2)$, and a heuristic function $h(x_1, x_g)$ to find a reduced-risk path based on the terrain characterisation information. The heuristic function is used to guide the A* algorithm towards the docking bay and specify a general shape for the path in order to reduce computation. The heuristics evaluated in this instance are the Euclidean, Manhattan [15] and Chebyshev [16] distances from a node x_1 to the docking bay x_g where p_1 and p_g are the positions of nodes x_1 and x_g in the grid.

The cost function defines the cost associated with moving from a node x_1 to an adjacent node x_2 and uses the values of the cost maps at both nodes in order to determine the total cost of the movement. It is important that the cost function appropriately weights the values of each map in order to achieve the desired path in terms of risk to the EMR. The cost function used is given by equation (9) where \mathbf{W} is given by equation (10), \mathbf{u} is the vector of travel from

the position of \mathbf{x}_1 to \mathbf{x}_2 given by equation (12) and $\mathbf{c}(x) = [C_H(x), C_{SX}(x), C_{SY}(x), C_R(x)]$ is the terrain characterisation vector of node x .

$$g(\mathbf{x}_1, \mathbf{x}_2) = |u| \mathbf{W} \cdot (\mathbf{c}(\mathbf{x}_2) - \mathbf{c}(\mathbf{x}_1)) \quad (9)$$

$$\mathbf{W} = [w_H, w_{SX}, w_{SY}, w_R] \quad (10)$$

w_H, w_R, w_{SX}, w_{SY} are the weights on height, roughness and slopes respectively and are bound by the condition given by equation (11) which forces normalisation.

$$w_H + w_R + \frac{w_{SX} + w_{SY}}{2} = 1 \quad (11)$$

The boundary map $C_B(i, j)$ is not used in the cost function but rather is used by the A* solver to determine whether or not to consider a given node. The vector \mathbf{u} is important in determining the cost contributed by each slope map, also in correctly scaling the cost for diagonal movements, and is given by equation (12), where \mathbf{p}_{x2} and \mathbf{p}_{x1} are the positions of the node in consideration and the current node respectively.

$$\mathbf{u} = \mathbf{p}_{x1} - \mathbf{p}_{x2} \quad (12)$$

The heuristic chosen was the Manhattan distance heuristic because the paths generated using the heuristic were favourable to the geometry of the Mars Yard for the purpose of docking. Figure 4 shows various paths generated with the Manhattan distance heuristic. Assuming that a linear combination of the three weights will yield somewhat of a linear combination of the three full-weighted paths; the cost function was chosen to have a weighting of $\mathbf{W} = [w_H, w_{SX}, w_{SY}, w_R] = [0.1, 0.4, 0.4, 0.5]$.

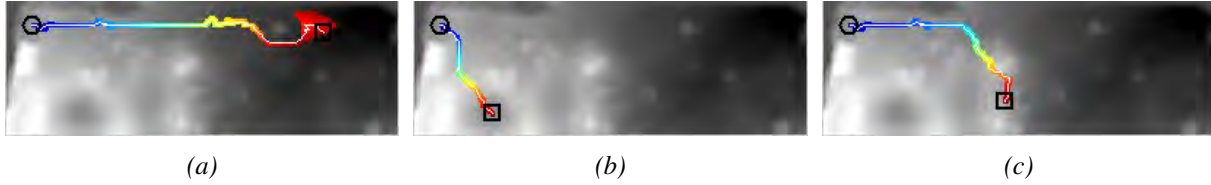


Fig. 4: Paths calculated using the Manhattan distance heuristic from various starting positions. The paths tend to consist of straight segments rather than curved segments

Trajectory Follower

The trajectory is generated using a constant nominal speed by linearly interpolating over the path calculated by the guidance system. A nominal speed s_n of 0.10 m/s is used, which is double the nominal speed of the NASA Mars Exploration Rovers [17].

The time taken to traverse a path segment between adjacent nodes x_i and x_{i+1} at the nominal speed is given by equation (13). This is extended to a path between any two indirectly connected nodes k and K ; time taken is then given by equation (14). In practice this is achieved via iterative Newton integration by summation of smaller time-steps.

$$T(i, i+1) = \frac{\sqrt{(\mathbf{p}_{i+1} - \mathbf{p}_i)^\top (\mathbf{p}_{i+1} - \mathbf{p}_i)}}{s_n} \quad (13) \quad T_d(k, K) = \sum_{i=k}^K T(i, i+1) \quad (14)$$

The trajectory $\mathbf{p}(t)$ is generated by linearly interpolating between the points in sequence:

$$\mathbf{p}(t) = (t_{k+1} - t)(\mathbf{p}_{k+1} - \mathbf{p}_k) \quad (15)$$

where k is the current path node number, \mathbf{p}_k and \mathbf{p}_{k+1} are the positions of the current path node and next path node, respectively. The time that the controller should arrive at path node k is denoted t_k and can be found by using equation (14) evaluated between limits $i = 0 \dots k$.

Control System

The control system consists of three major components: the trajectory follower that propagates the desired position with respect to time using a constant speed, the PI controller that tracks the trajectory and computes the control input to the actuators, and the PI controller that adjusts the heading of the EMR during a spot turn.

Mobility Control

The mobility subsystem of the EMR is actuated by 12 digital servo motors; 6 of which control the steering angle of each wheel and 6 of which control the throttle to each wheel. Figure 5(a) identifies the IDs of the actuators of the EMR mobility subsystem. The steering servo motors are set to joint mode and can have a commanded angle set in the range of -150° to 150° at a commanded angular speed. The driving motors can have a commanded percentage throttle set at a desired angular speed in the range of $0^\circ/s$ to $78^\circ/s$ which the internal controller tries to seek. The actuator shaft is not guaranteed to rotate that angular speed; throttle roughly refers to maximum torque to be used in achieving the desired angular speed.

The steering angle of each wheel η is constrained to the range -90° to 90° due to the mechanical design of the EMR rocker-bogie connections. This however does not restrict the motion of the EMR; any steering angle outside the valid range can be achieved by negating the throttle and using the supplementary steering angle given by equation. These conditional functions are given in equations (18) and (22). Using this control strategy the EMR can travel in all directions using crab motion as pictured in Figure 5(a). A disadvantage of this control strategy is that the kinematic model becomes discontinuous when η approaches $\pm\pi$ rad and as such η should avoid being driven near that angle.

Proportional-Integral Trajectory Follower

The PI controller determines the input $\mathbf{u}(\mathbf{x}, \mathbf{x}_0)$ to the actuators for the crab motion model to follow the trajectory $\mathbf{p}_t = \mathbf{p}(t)$. Due to the slow speed of the EMR a derivative term is omitted from the controller.

$$\mathbf{x} = [{}^g\mathbf{p}_b^\top, \psi, I]^\top \quad (16)$$

$$\theta = \text{atan2}(e_{py}, e_{px}) \quad (17)$$

$$\eta = \begin{cases} \psi - \theta & \text{if } |\psi - \theta| \leq \frac{\pi}{2} \\ \psi - \theta - \pi & \text{if } |\psi - \theta| > \frac{\pi}{2} \end{cases} \quad (18)$$

$$I_t = I_{t-1} + k_i e_v dt \quad (19)$$

$$\mathbf{e}_p = \mathbf{p}_t - {}^g\mathbf{p}_b \quad (20)$$

$$\mathbf{e}_v = \sqrt{\mathbf{e}_p^\top \mathbf{e}_p} - d \quad (21)$$

$$v = \begin{cases} k_p e_v + k_i I_{t-1} & \text{if } |\psi - \theta| \leq \frac{\pi}{2} \\ -(k_p e_v + k_i I_{t-1}) & \text{if } |\psi - \theta| > \frac{\pi}{2} \end{cases} \quad (22)$$

$$\mathbf{u} = [v, \eta]^\top \quad (23)$$

The control law is given by equation (23). ${}^g\mathbf{p}_b$, ψ and I_t are the state vector components position, EMR heading and the driving throttle integral variable at time t . dt is the time step between actuator inputs. k_p , k_i , k_h are the speed proportional and integral constants and the steering angle proportional constant. e_p and e_v are the errors in position and speed and d is the 'look-ahead' distance. d is used to let the heading control 'look ahead' of the trajectory which in effect makes the controller produce smoother paths, however introduces positional error.

Proportional-Integral Spot-Turn Controller

The spot-turn controller is a PI controller that can achieve accurate desired headings while remaining relatively stationary in position. The procedure involves three parts: positioning steering actuators in spot turn configuration; executing the spot-turn controller until the heading error residual of equation (24) is within the acceptable error bounds ϵ_ψ ; and finally turning the steering actuators to the forward configuration. The spot turn steering actuator

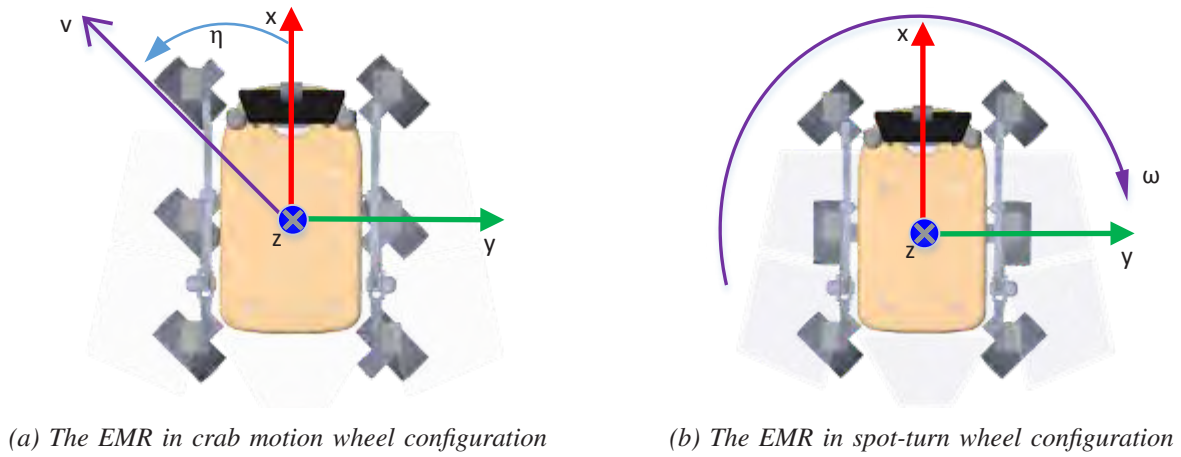


Fig. 5: The wheel configuration of the EMR in crab motion and spot turn manoeuvres

configuration is pictured in Figure 5(b). The spot-turn controller uses only heading measurement information ψ_m to control the driving speed ω of the wheels using the control law in equation (25) and integral state update equation (26).

$$\psi_e = \psi_d - \psi_m \quad (24) \quad \omega = k_p \psi_e + k_i I_{\psi|t} \quad (25) \quad I_{\psi|t+1} = I_{\psi|t} + \psi_e dt \quad (26)$$

where $I_{\psi|t}$ and $I_{\psi|t+1}$ are the integral heading state immediately before and after the measurement ψ_m is taken, and k_p and k_i are the proportional and integral error constants respectively.

Results and Discussion

Figure 6 shows a path calculated with the different heuristic functions using the same cost function. The images show a grayscale height map of the Mars Yard with a colour overlay on the nodes that were evaluated by the A* algorithm. The colours represent the normalised value of the heuristic where hotter colours correspond to further away from the docking bay.

The starting position is marked with a black square and the docking bay is marked with a black circle. The path taken is shown in white or magenta for reversed paths. The path

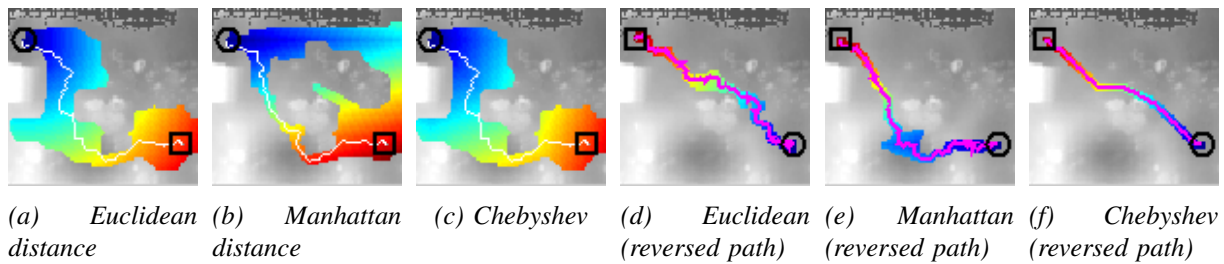


Fig. 6: Path evaluated with various heuristics

calculated with the Euclidean distance heuristic follows a more direct path to the docking bay while the path calculated with the Manhattan distance heuristic first minimises the cost in one direction then the other, unless impeded by high costs.

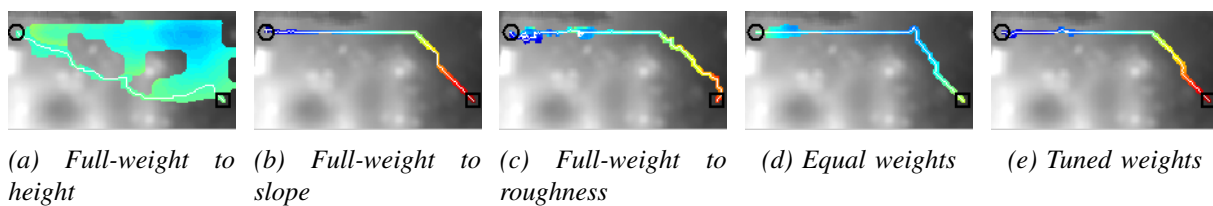


Fig. 7: Example paths with variations in weightings of terrain costs

In Figure 7 each path was generated using the Manhattan distance heuristic and the same initial and docking bay positions were specified. The result of full-weighting a single characteristic is shown in Figure 7(a), (b), (c). In Figure 7(a) using full weight on C_H yields smooth paths and avoids the crater but subjects the rover to unnecessary risk by ignoring the small and medium sized rocks and slopes. Full weight on C_S yields flat paths and avoids the crater and medium size rocks. Full weight on C_R yields paths that minimise the terrain roughness of the path, which favours avoiding rocks of all size and favours paths without obstacles. The tuned cost function is not subject to the faults of the full-weighted cost functions and is a good balance between minimising traversal over rough terrain, obstacles and large changes in height.

In Figure 7(d) equal weights are used and in 7(e) a path is calculated with a experimentally tuned weights. Figure 7(e) shows a good compromise between minimising terrain roughness and slope. The weighting of terrain height did not have significant effect on the path.

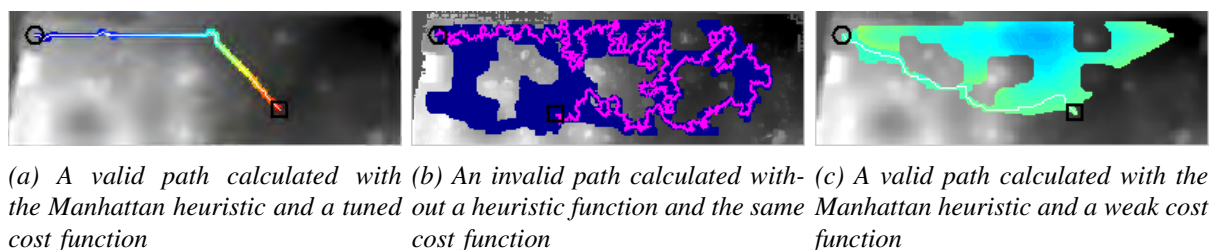


Fig. 8: Calculated invalid paths and a valid path

The control of the EMR in both the position and heading are measured using the onboard

localisation system and the measurements are used to compute the actuator inputs. The result of this is discussed against the nominal trajectory for the trajectory follower, and the desired heading for the spot-turn controller.

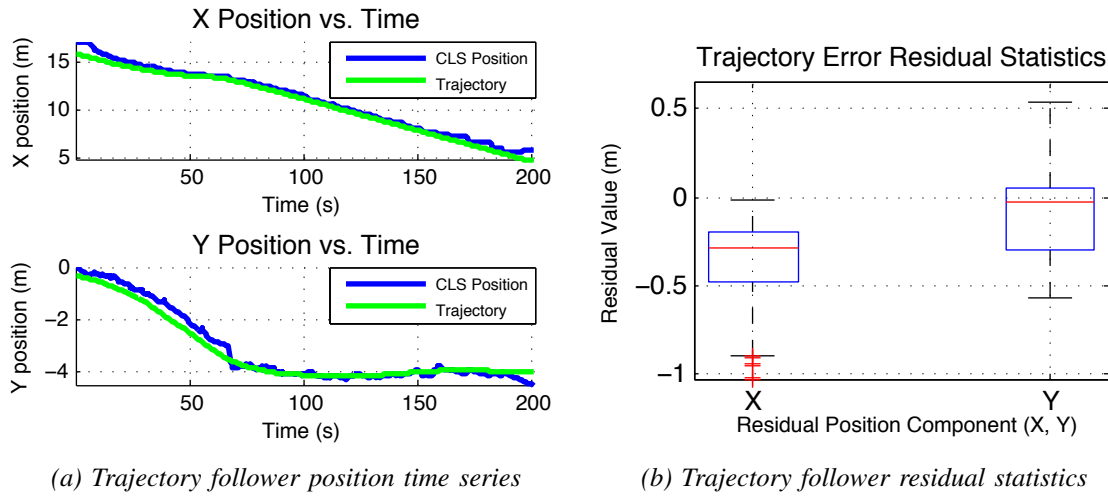


Fig. 9: Position results of a trajectory tracking experiment

Figure 9 shows the position and position residual statistics of an experiment in which the EMR followed a predetermined path. The residual error statistics are summarised in Table 1 and shown in Figure 9. The mean errors correspond to the expected offset behind the trajectory tracking point and standard deviation loosely corresponds to the controller oscillation. The

Table 1: Summary of statistics for trajectory follower residuals

Mean X Position	Mean Y Position	Std. X Position	Std. Y Position
-38 cm	-9 cm	27 cm	22 cm

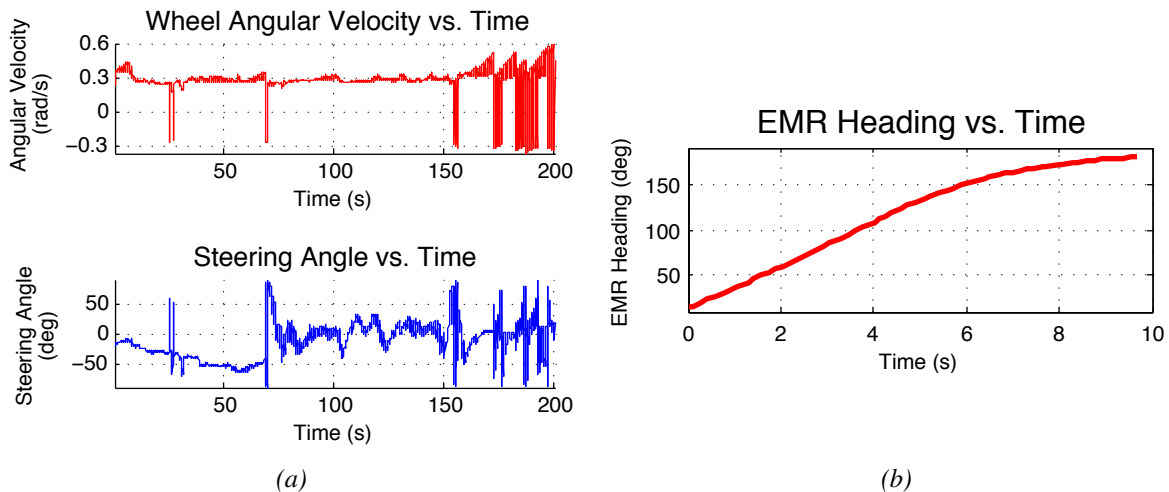


Fig. 10: (a) Control input time series during trajectory tracking; (b) Spot turn time series

trajectory follower control inputs are shown in Fig. 10(a), where the effects of the discontinuity of the steering actuators is clearly visible on the red and blue lines, and corresponds with oscillations in 9(a). Figure 10(b) shows a spot turn from 10° to -180° performed over 10 seconds. The proportional action of the controller is most evident between 0 and 6 seconds,

after which the integral action begins to dominate the controller and the curve becomes non-linear. This enables the controller to achieve accurate desired headings less than 2° .

Conclusion and Future Work

This paper has described an autonomous docking utility system that has been successfully tested at the Mars Yard and is capable of safely returning the EMR to its docking bay with an acceptable position and heading error. The guidance system was successfully used to produce valid, obstacle-free trajectories in the Mars Yard. The PI control systems performed well without any knowledge of the EMR dynamics. This is an advantage in that the GNC system may be used with other robots given that a crab movement actuation routines exist. The final position error is attributed to both how well the controller tracks the trajectory and to the accuracy of the localisation system. The final heading error of a spot turn manoeuvre executed over 10 seconds was within an acceptable bounds of 2° . Standard deviation from the trajectory was within 35cm of the nominal path.

Future work on the guidance and control systems include integrating the dynamic model of the EMR into the path planning algorithm and PI controllers. This would allow the implementation of LQR (Linear Quadratic Regulator) or MPC (Model Predictive Controller) in the control system to more accurately follow the trajectory and reduce docking error in both position and heading. Further development of the autonomous docking procedure involves the design, development and integration of a docking station to automatically connect the charger to the EMR.

Acknowledgements

The authors would like to thank Mr. Daniel Wilson from the Australian Centre for Field Robotics (ACFR) for his support throughout the project. This work was partially supported by the Broadband Enabled Education and Skills Services Department of Education program, Australian Federal Government (2012-2014) entitled Education 2020: enabling learning in science, engineering and mathematics.

References

- [1] Kubota, T., Kuroda, Y., Kunii, Y. and Yoshimitsu, T., "Path planning for newly developed microrover", *Proc. IEEE Int. Conf. Robotics and Automation*, Seoul, Korea 2001, Vol. IV, pp. 3710–3715
- [2] Maimone, M. W., Leger, P. C. and Biesiadecki J. J., "Overview of the mars exploration rovers autonomous mobility and vision capabilities", *IEEE Int. Conf. Robotics and Automation Space Robotics Workshop*, Rome, Italy, 2007
- [3] Kulkarni, N., Ippolito, C., Krishna-Kumar, K. and Al-Ali, K. M., "Adaptive inner-loop rover control", *2nd IEEE Int. Conf. Space Mission Challenges for Information Technology*, Pasadena, USA, 2006 pp. 8–504,
- [4] Low, E. M. P., Manchester, I. R. and Savkin, A. V., "A method for vision-based docking of wheeled mobile robots", *IEEE Int. Symp. Intelligent Control*, Munich, Germany, 2006 pp. 614–619,

- [5] Luo, R. C., Liao, C. T. and Lin, K. C., "Vision-based docking for automatic security robot power recharging", *IEEE Workshop Advanced Robotics and its Social Impacts*, 2005, pp. 214–219,
- [6] Lee, H., Kim T., Park, W. and Lee, H. K. "Extraction of digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry", *Image and Vision Computing*, Vol. XXI, 2003, pp. 789–796
- [7] Eisenman, A.R.; Liebe, C.C.; Perez, R., "Sun sensing on the Mars exploration rovers," *IEEE Aerospace Conference Proceedings*, 2002, Vol. 5, pp. 5–2249
- [8] Gammell, J. D. et. al., "Rover odometry aided by a star tracker", *Proc. IEEE Aerospace Conference*, 2013, pp. 1–10
- [9] Soon-Yong, P., Sung-In, C., Jaekyoung, M., Joon, K. and Yong-Woon, P., "Localization of an unmanned ground vehicle using 3D registration of laser range data and DSM", *Workshop on Applications of Computer Vision (WACV)*, 2009, pp. 1–6,
- [10] Yong-Hoon, J., Sung-Ho, H., Jae-Bok, S. and Ji-Hoon, C., "DSM update for robust outdoor localization using ICP-based scan matching with COAG features of laser range data", *IEEE/SICE Int. Symp. System Integration*, 2011, pp.1245–1250,
- [11] Goldberg, S.B.; Maimone, M.W.; Matthies, L., "Stereo vision and rover navigation software for planetary exploration", *Aerospace Conference Proceedings, 2002. IEEE*, 2002, Vol. V, pp.5-2025,
- [12] Silverman, M. C., Nies, D., Boyoon, J. and Sukhatme, G., "Staying alive: a docking station for autonomous robot recharging", *IEEE Int. Conf. Robotics and Automation Proceedings*, Washington D.C, 2002, Vol. I, pp. 1050–1055
- [13] Hoffman, R. and Krotkov, E., "Terrain Roughness Measurement from Elevation Maps", *Proc. SPIE Mobile Robots*, Vol IV, Philadelphia, USA, 1989, pp/ 104–114
- [14] Madsen, R. E., Hansen, L. K. and Winther, O., "Singular value decomposition and principal component analysis", *Neural Networks*, 2004, pp. 1–5
- [15] Krause, E. F., *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, New York, USA: Dover, 1987
- [16] Cantrell, C. D., *Modern Mathematical Methods for Physicists and Engineers*, NY, USA: Cambridge University Press, 2000
- [17] Biesiadecki, J. J, Leger, P C. and Maimone, M. W., "Tradeoffs between directed and autonomous driving on the mars exploration rovers", *Int. Journal of Robotics Research*, Vol. XXVI, 2007, pp. 91–104



APPENDIX A

13ASSC List of Presentations & Posters

List of Presentations

Abstracts are listed in alphabetical order of first author

Name of Presenter	Names of all authors	Title
Barnaby Osborne	Elias Aboutonios, Barnaby Osborne	A Comprehensive Tertiary Education Program in Satellite Systems Engineering
Brett Addison	Brett Addison	The recently discovered exoplanet WASP-79b in a Nearly polar orbit
Naomi Altman	Naomi Altman	Expanding Australia's contribution to the ACES and future space missions
Jeremy Bailey	Jeremy Bailey	30 Years of Studying Venus with the Anglo-Australian Telescope
Lewis Ball	Lewis Ball	CSIRO Astronomy and Space Science
Emily Bathgate	Emily Bathgate, Graziella Caprarelli, Linda Xiao, Ron Shimon, Ross Pogson	Raman characterisation of the products of alteration of volcanic rock minerals: preliminary results and implications for Martian studies.
Jayanta K. Behera	Jayanta K. Behera, Anand K. Singh, A. K. Sinha, R. Rawat, C. Selvaraj, K.U. Nair	Substorm associated dayside CNA and simultaneous Pc5 pulsations in geomagnetic field variations
Ragbir Bhathal	Ragbir Bhathal	Perspectives on Aboriginal Sky Knowledge
Brett Biddington	Brett Biddington	Australia's bid to host the International Astronautical Congress (IAC) in 2017
Brett Biddington	Brett Biddington	Role of Australian SMEs in the global space economy
Kimberly Bott	Kimberly Bott, Lucyna Kedziora-Chudczer, Jeremy Bailey	Modelling a Hot Jupiter Atmosphere with VSTAR
Zahra Bouya	Zahra Bouya, Michael Terkildsen, Matthew Francis	Regional climatological model of TEC over Australia
Russell Boyce	Russell Boyce	Outcomes of the SCRAMSPACE Scramjet-based Access-to-Space Systems project
Robert Brand	Robert Brand	Triple Play in the Space Sector
Tim Broadbent	Tim Broadbent, Jason Held, Daniel Bunker	The DragEN tether deployer experiments for the NASA Flight Opportunities Program
Hugh Cairns	Hugh Cairns	An Ancient Aboriginal Astronomy from the Northern Territory
Iver Cairns	Iver Cairns, Vasily Lobzin, Alina Donea, Steven Tingay, Divya Oberoi	MWA and Spacecraft Observations of Type III Solar Radio Bursts, X-ray Bursts, and Plasma Flows
Iver Cairns	Iver Cairns, Xiaofeng Wu, Joss Bland-Hawthorn, Chris Betters, Tony Monger, Jeremy Soh, Jiro Funamoto, Adrian Xiao, Lisa Fogarty, Sergio Leon-Saval	Status of the University of Sydney's i-INSPIRE spacecraft project
Manuel Cervera	Manuel Cervera, Trevor Harris	Generation of TIDs by Atmospheric Gravity Waves and their effect on VI ionograms
Sarah Chamberlain	Sarah Chamberlain, Jeremy Bailey, David Crisp, Vikki Meadows	Water Vapour near the Venus Surface

Name of Presenter	Names of all authors	Title
Anthony Cheetham	Anthony Cheetham, Nick Cvetojevic, Peter Tuthill, Anand Sivaramakrishnan	Optical testbed demonstration of Fizeau Interferometric Cophasing of Segmented Mirrors
Mazher Choudhury	Mazher Choudhury, Joon Cheong, Jinghui Wu, Nagaraj Shivaramaiah, Andrew Dempster	Test Results of the Namuru Dual-GNSS Space-borne Receiver
Jonathan Clarke	Jonathan Clarke	Searching for life on early Mars: lessons from the Pilbara
Daniel Cotton	Daniel Cotton, Jeremy Bailey, Lucyna Kedziora-Chudczer	Spectroscopic Fitting of the Terrestrial Atmosphere for Improved Planetary Atmosphere Observations
Julie Currie	Julie Currie, Colin Waters	ULF Waves and Earthquake Precursors
Michael Davis	Michael Davis	Southern Hemisphere Summer Space Program
Andrew Dempster	Andrew Dempster	Past and present research activities at ACSER
Andrew Dempster	Andrew Dempster	The Garada "SAR Formation Flying" project
Kerrie Dougherty	Kerrie Dougherty	Before Woomera: Australia's Early Rocketeers
Kerrie Dougherty	Kerrie Dougherty, Carol Oliver, Jennifer Fergusson	Pathways to Space: ASRP Project Review
Rod Drury	Rod Drury	Lockheed Martin and space industry in Australia
Brian Fraser	Brian Fraser, Harlan Spence, Geoffrey Reeves, Craig Kletzing	New Understanding of Inner Magnetosphere Physics: Initial Van Allen Probe Results
Roland Fleddermann	Roland Fleddermann	Laser Ranging Instrument for the GRACE Follow-on mission
Lewis Freeland	Lewis Freeland, Fred Menk, Murray Sciffer, Michael Terkildsen	A Search for Signatures of Discrete Cosmic Radio Sources Using an Imaging Radiometer at Davis, Antarctica
Robert Fuller	Robert Fuller	The Sky Knowledge of the Kamilaroi People and Their Neighbours
Stuart Gilchrist	Stuart Gilchrist	Nonlinear Force-free Modeling of the Coronal Magnetic Field in Spherical Geometry
James Gilmore	James Gilmore	Exo-Earth Habitability and Radiative Properties in Energy Balance Models
James Gilmore	James Gilmore	Measurement Uncertainty in Satellite Observations of Precipitation and Column Water Vapor
Eamonn Glennon	Eamonn Glennon, Scott O'Brien, Joon Wayn Cheong, Kevin Parkinson, Andrew Dempster	CubeSat-based GPS Reflectometry using the Namuru V3.2 GPS Receiver
Ali Haydar Goktogan	Ali Haydar Goktogan, Salah Sukkarieh	An Internet-based Collaborative Teleoperation System for the Mars Lab and Distance STEM Education
Ali Haydar Goktogan	Ali Haydar Goktogan, Iwan Kelaiah, Salah Sukkarieh	Virtual Mars Yard Simulation System for Internet-Based STEM Education
John Goldsmith	John Goldsmith	The "Ilgarjiri- Things Belonging to the Sky" project: Collaboration between Aboriginal communities and radio astronomy in Australia.
Alice Gorman	Alice Gorman	Beyond the Morning Star: the Voyager spacecraft and Australian Aboriginal culture
Eren Gorur	Eren Gorur, Ivan Cheung	Australian Space Policy: A Student Perspective
Duane Hamacher	Duane Hamacher	Are Supernovae Recorded in the Astronomical Traditions of Aboriginal Australians?

Name of Presenter	Names of all authors	Title
James Harding	James Harding, Iver Cairns, Vasili Lobzin	New Density Model for the Solar Corona and Constraints on the Corona's Origin and Outflow
Trevor Harris	Trevor Harris, Manuel Cervera, Adrian Snell	EXOTIC: A small experiment on polarisation effects of oblique HF radiowaves (separating O/X modes on oblique soundings)
Yi He	Yi He, Chao Zhang	Titan Stream Mapping – Morphometric analysis of channel on Titan
Steven Hobbs	Steven Hobbs, David Paull, Jon Clarke	A Comparison Of Arid, Temperate And Sub-Humid Terrestrial Gullies With Gullies On Mars – Implications For Martian Gully Erosion
Jonathan Horner	Jonathan Horner, Dave Waltham	The role of Jupiter in driving periodic climate change.
Jonathan Horner	Jonathan Horner, Robert Wittenmyer, Tobias Hinse, Jonathan Marshall, Alex Mustill, Chris Tinney	Wobbling Ancient Binaries – Here Be Planets?
Intan Idrus	Intan Idrus, Mardina Abdullah, Alina Hasbi	Observation of Large-scale Traveling Ionospheric Disturbance in Peninsular Malaysia using GPS receivers
Lucyna Kedziora-Chudczer	Lucyna Kedziora-Chudczer, Jeremy Bailey, Daniel Cotton, Kim Bott, Graeme Salter	Inflated 'hot' Jupiters and their atmospheres
John Kennewell	John Kennewell	Satellite Overpass Statistics
John Kennewell	John Kennewell, Phil Bland, Martin Towner	The desert fireball network and orbital space debris reentries
Champlain Kenyi	Champlain Kenyi, Jeremy Bailey, Daniel Cotton	Retrieval of greenhouse gas concentrations from observations with a ground-based spectrometer in the near-infrared
Jeff Kingwell	Jeff Kingwell	Geoscience Australia activities in space science
Alexey Kondyurin	Alexey Kondyurin, Irina Kondyurina, Marcela Bilek, Alexander Svistkov	Composite materials cured in stratosphere during stratospheric balloon flight
Trevor Kwan	Trevor Kwan	Position Determination of Satellites in Formation Flying
Andrew Layden	Andrew Layden, Iver Cairns, Bo Li, Peter Robinson	Langmuir electrostatic decay in magnetized plasmas
Brett Layden	Brett Layden, Iver Cairns, Peter Robinson	Exact Evaluation of the Rates of Electrostatic Decay and Scattering off Thermal Ions for an Unmagnetized Maxwellian Plasma
Trevor Leaman	Trevor Leaman, Duane Hamacher	Ooldea Nights: Daisy Bates and the Aboriginal Sky Knowledge of the Great Victoria Desert, South Australia
John Le Marshall	John Le Marshall, Jin Lee, Paul Gregory, Jim Jung, Robert Norman, Rolf Seecamp	An Analysis Of The Impact Of Earth Observations From Space On Australian And Global Numerical Weather Prediction
Liam Liyuan Li	Liam Liyuan Li, Wei Zhang, Xiaojing Li and Linlin Ge	Satellite Vibrations and Restoration of High Resolution Pushbroom Images
Tim Youtian Liu	Tim Youtian Liu, Jicang Wu, Xiaojing Li and Linlin Ge	Phase Unwrapping for Interferometric Synthetic Aperture Radar Technique
Shyeh Tjing Loi	Shyeh Tjing Loi, Iver Cairns, Bo Li	Production of fine structures in type III solar radio bursts due to turbulent density profiles

Name of Presenter	Names of all authors	Title
Helen Maynard-Casely	Helen Maynard-Casely, Max Avdeev, Helen Brand, Kia Wallwork	Icy moons in the laboratory, re-creating the outer solar system at ANSTO
Fred Menk	Frederick Menk, Brian Fraser, Yonghua Liu	Pc2 Electromagnetic Ion Cyclotron Wave Source Regions in the Outer Magnetosphere
Haritina Mogosanu	Haritina Mogosanu, Jonathan Clarke, Elf Eldridge	"Using analogue Mars expeditions to create school resources for New Zealand's "Planet Earth and Beyond" Curriculum."
Anthony Monger	Anthony Monger	Frequency Allocation for Satellite Space Stations using Amateur Radio Bands
Robert Moore	Robert Moore	Economical Liquid Rocket Engine Production Using Thermal Spray Techniques
Shaun Moss	Shaun Moss	Blue Dragon: A Humans-to-Mars Mission Architecture
Sai Dheeraj Nadella	Sai Dheeraj Nadella, Ananya Pramanik	Computational Requirements in Interfacing Attitude Determination And Control System, onto the On Board Data Handling System in a Nano-satellite
Dave Neudegg	Dave Neudegg	IPS Radio and Space Services and space weather in cycle 24
Thien Nguyen	Thien Nguyen	Finishing BLUEsat
Craig O'Neill	Craig O'Neill	The tectonics of exoplanets
Jon Ozer	Jon Ozer	ArduSat: Bringing Hands-On Space Technology Into The Classroom
John Page	John Page, Haoyang Cheng, Monica Chi, Nathan Kinkaid	Swarm-Based Satellite Constellation Control
Tim Payne	Tim Payne	DSTO Space Program
Lenard Pederick	Lenard Pederick, Manuel Cervera	Semi-empirical Model for Ionospheric Absorption based on the NRLMSISE-00 atmospheric model (SiMIAN)
Steven Potiris	Steven Potiris, Anthony Tompkins, Ali Haydar Göktoğan	A Vision-based Autonomous Docking System of the Experimental Mars Rover
David Pross	David Pross, Robert Pankhurst, Duane Hamacher	Sky Knowledge and Rock Art in the Sydney Basin
Li Qiao	Li Qiao, Min Zhang, Barnaby Osborne, Andrew Dempster	Attitude Determination and Control System (ADCS) of UNSW QB50 project "UNSW ECO"
Jon Rask	Jon Rask, Patti Zeidler-Erdely, Terence Meighan, Mark Barger, William Wallace, Bonnie Cooper, Dale Porter, Erin Tranfield, Lawrence Taylor, David Mckay, Vincent Castranova, Yang Liu, David Loftus	Lung Macrophage Responsiveness to Chemically Reactive Lunar Dust
William Reid	William Reid, Callum McCarter, Ali Haydar Goktogan	A Highly Mobile Wheel-on-Leg Planetary Rover for use in a Mars Analogue Environment
Chris Rizos	Chris Rizos	GNSS Trends: Where are we now and where are we going?
Elyse Schinella	Elyse Schinella, Craig O'Neill	Venusian Admittance and Correlation: Clues to Topographic Compensation Mechanisms and Mantle Dynamics

Name of Presenter	Names of all authors	Title
Fiona Schleyer	Fiona Schleyer, Iver Cairns, Eun-Hwa Kim	Linear mode conversion of upper hybrid waves to radiation: Averaged energy conversion efficiencies, polarization, and applications to Earth's continuum radiation.
Joachim Schmidt	Joachim Schmidt, Iver Cairns	Type II solar radio bursts predicted by 3D MHD CME and kinetic radio emission simulations
Kyll Schomberg	Kyll Schomberg, Graham Doig, John Olsen	Computational analysis of pintle variation in an expansion-deflection nozzle
Gareth Sciffer	Gareth Sciffer, Colin Waters, Murray Sciffer, Yuki Obana	Modelling ULF Quarter-wave modes
Murray Sciffer	Murray Sciffer, Colin Waters, Robert Lysak	Phase distortions of space to ground electromagnetic signals due to ULF waves
Leila Norouzi Sedeh	Leila Norouzi Sedeh, Colin Waters	Properties of ULF waves observed in TIGER radar
Harrison Steel	Harrison Steel, Christopher P. McKay, Duncan Steel	NASA's Spaceward Bound Program: Expeditions in Australia and New Zealand
Jospeph Stefano	Jospeph Stefano, Xiaofeng Wu	QB50 Power Management Optimization
David Lingard	Leon Stepan, Iain Cartwright, David Lingard	COWPATT: The Value of an Integrated Modelling Approach to Designing Cubesats and their Missions
Paul Stewart	Paul Stewart, Peter Tuthill, Phil Nicholson, Matt Hedmann	High Resolution Imaging of Evolved Stars with Cassini Stellar Occultations
Steven Tingay	Steven Tingay	The Murchison Widefield Array, Precursor for the Square Kilometre Array
Dion Tiu	Dion Tiu	New model for the ion reflection process as quasiperpendicular collisionless shocks
Anthony Tompkins	Anthony Tompkins, Steven Potiris, Ali Haydar Göktoğan	Development of a Low-Cost Vision-Based Localisation System for the Experimental Mars Rover
Anne Unewisse	Anne Unewisse, Andrew Cool, Andrew Kirby, Dan Meehan	A New Relocatable Airglow Imager
Colin Waters	Colin Waters	Probing near-Earth Space using ULF signals
Bob Buxton	Martin Westwell, Bob Buxton	Place and Space: Perspectives on Earth Observation
Mike Wheatland	Mike Wheatland	Revisiting the event statistics method for solar flare prediction
David Williams	David Williams	Space Research capability in Australia. Are there lessons from the UK and the wider Europe?
David Willson	David Willson, Carol Stoker, Lawrence Lemke, Steven Hu, Jeffrey Bowles, Andrew Gonzales, Loc Huynh, Nicolas Faber, Cyrus Foster and Bernardus Helvensteijn	Mars Exploration and Sample Return using Dragon: A New low cost paradigm for Mars science missions
David Willson	David Willson	Off-World field science performance: Results from the Pilbara space suit trials
Geoff Wyatt	Geoff Wyatt	Star wheel and signals: Sydney Observatory's Shared Sky education program
Yang Yang	Yang Yang, Andrew Dempster, Xiaokui Yue	Inter-satellite Link Augmented GPS Navigation System for Autonomous Satellites Formation Flying

Name of Presenter	Names of all authors	Title
John Le Marshall	Kefei Zhang, Robert Norman, John Le Marshall, Brett Carter, Simon Alexander, Suqin Wu, Shaocheng Zhang, Ying Li, and Congliang Liu	The Platform Technology ASRP Project – Completion and Future Directions

Posters

Posters are listed in alphabetical order of first author

Name of Presenter	Names of all Authors	Title
Shaila Akhter	Shaila Akhter, Maria Cunningham, Paul Jones	Clustering of dense molecular gas in the Milky Way Galaxy
Jeremy Bailey	Jeremy Bailey	Greenhouse Gas Measurement using Fibre-Optic Spectrometers
Mazher Choudhury	Mazher Choudhury, Eamonn Glennon, Andrew Dempster	Analysing Navigation Solutions of the Namuru V3.2 GPS Receiver for Spaceborne Applications
Suelynn Choy	Suelynn Choy	Precipitable Water Vapour Estimation using Observations from the Australian Regional GNSS Network for Climate Monitoring
Samuel Francis	Samuel P. Francis, Kirk McKenzie, Timothy T-Y. Lam, Andrew J. Sutton and Daniel A. Shaddock	Weak-Light Phase Locking for Space-Based Interferometry
Ali Haydar Goktogan	Ali Haydar Goktogan, Salah Sukkarieh	Sydney Lunabotics: A Project Based Teaching and Learning in Space Robotics
David Herne	David Herne, John Kennewell, Mervyn Lynch	Ionospheric Phenomena and Low-frequency Radio Astronomy
Lea Jouvin	Lea Jouvin, Sarah Maddison, Jonathan Horner	Dynamical stability of terrestrial planets in the habitable zone of HD204313
Lucyna Kedziora-Chudczer	Lucyna Kedziora-Chudczer, Jeremy Bailey, Jonti Horner, Daniel Cotton	The D/H ratio of giant planets as a diagnostic of planet forming conditions
Alexey Kondyurin	Alexey Kondyurin, Liudmila Komar, Alexander Svistkov	Modelling of the composite materials cured under free space conditions
Bo Li	Bo Li, Iver Cairns	Type III Bursts Produced by Power-law Distributed Electrons Injected into Thermal Corona
Vasily Lobzin	Vasily Lobzin	Automatic recognition of type III solar radio bursts in STEREO/WAVES data
Frank Mills	Frank Mills	Modelling and observations of mesospheric sulfur chemistry on Venus
Dave Neudegg	Dave Neudegg	Antarctic polar cap ionosphere and effects of solar EUV, magnetosphere-ionosphere coupling and thermospheric transport
Jeremy Soh	Jeremy Soh, Xiaofeng Wu	A modular FPGA-based Kalman filter for nanosatellite attitude determination
Samira Tasnim	Samira Tasnim, Iver Cairns	Generalized theory for the solar wind
Marek Zbik	Marek Zbik, David Williams	Discovery of nano in size particles within glassy bubbles in lunar soil, likely source of the ultra-fine lunar regolith fraction

