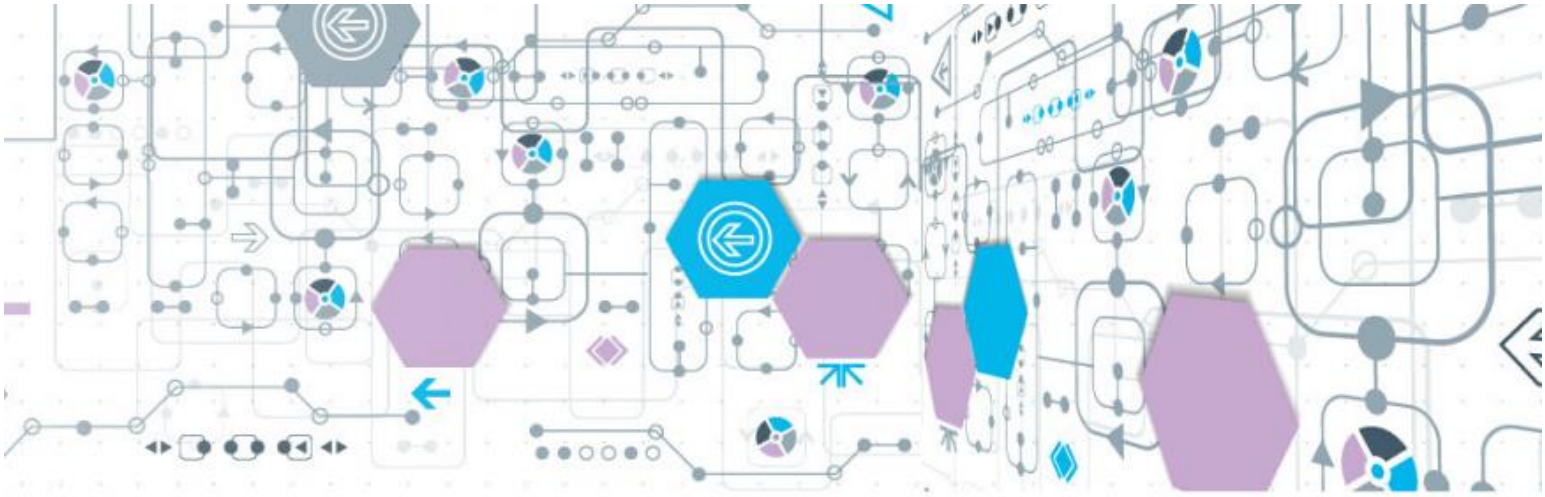
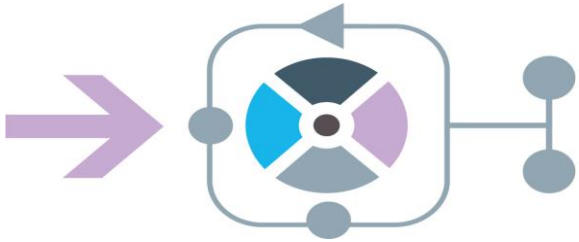


New Zealand Research Software Engineering Conference | 2020

Abstract Booklet

Click on the session title/presenter to read the abstract

Programme 2 - 3
Abstracts 4 - 28



New Zealand Research Software Engineering Conference | 2020

Tuesday 8 September (Pre-conference event)

16:00 - 17:00		**Social event/Welcome to the NZRSE Virtual Conference Platform**

Wednesday 9 September (Day 1)

10:00 - 10:30		**Conference Welcome Address**
10:30 - 10:35		**Mini Break**
10:35 - 11:25		**Presentation Session 1**
	10:40	Dr David Nickerson - Session intro
	10:50	Hugh Sorby - Reproducible and FAIR modelling with CellML and libCellML.
	11:00	Dewan Sarwar - Epithelial Modelling Platform: A Tool for Model Discovery and Construction of Computational Models of Epithelial Transport
	11:10	Dr Karin Lundengård - Physiome curation system for reproducible models
11:25 - 11:45		**Morning Tea/Networking**
11:45 - 12:45		**Keynote**
	11:45	Dr Hilary Oliver, Senior Research Software Engineer, NIWA - Scaling: it's all (well, quite a lot!) about the workflow.
12:45 - 13:45		**Lunch/Networking**
13:45 - 14:15		**Networking Activity**
14:15 - 15:15		**Presentation Session 2**
	14:15	Dr Pablo Higuera - Parallel CFD Modelling to Explore Scale Effects in Coastal Engineering Wave-Structure Interaction (WSI) simulations
	14:35	Dr Wolfgang Hayek - Parallel Computing with Dask
	14:55	Maxime Rio
15:15 - 15:35		**Afternoon Tea/Networking**
15:35 - 17:05		**Presentation Session 3**
	15:35	Dr Chris Scott & Dr Damien Mather - OpenACC pgfortran: substantial speedups and beyond for the O(3) Condensation algorithm
	15:55	Rere-No-A-Rangi Pope & Rhys Owen- NeSIssity is the mother of invention
16:15 - 16:25		**Mini Break**
	16:25	Dr John Rugis - Processing Calcium Signalling Fluorescence Microscopy Image Stacks
	16:45	Mingrui Yang - Native distributed and MPI parallelism in the high-level language Julia for quantum Monte Carlo
17:05 - 17:35		**Networking happy hour**
17:35		**End of Day One**

Thursday 10 September (Day 2)

10:00 - 10:10		**Housekeeping welcome**
10:20 - 10:30		**Mini Break**
10:20 - 11:20		**Presentation Session 4**
	10:20	TBC
	10:40	Dr Alexander Pletzer - Harnessing more compute power on NeSI's platforms with OpenMP
	11:00	Dr Chris Seal - Customising an open source project - the tension between convergence and divergence
11:20 - 11:40		**Morning Tea/Networking**
11:40 - 13:10		**Presentation Session 5**
	11:40	Dr Kyle Chard - funcX: A Federated Function Serving Fabric for Science
	12:00	Dr Louise Ord - Developing an Interactive Genome Analytics Tool – an Interface to Scientific Insight
12:20 - 12:30		**Mini Break**
	12:30	Eric Burgueno - Practical recipe standards for Singularity containers to make life easier for Data
	12:50	Scientists Marko Laban - Running Web Applications on HPC using containers
13:10 - 14:00		**Lunch/Networking**
14:00 - 14:35		**Lightening Talks**
	14:00	Dr Michelle Barker - FAIR for Research Software working group
	14:05	Bruno P. Kinoshita - Cylc user interface scalability
	14:10	Nilani Algiriyage - Traffic Flow Estimation based on Deep Learning using CCTV Images
	14:15	Dr Chris Scott - Jupyter on NeSI
14:35 - 14:55		**Afternoon Tea/Networking**
14:55 - 15:55		**Keynote**
	14:55	Dr Denis Bauer, Principal Research Scientist, CSIRO - How computational science and data-driven health decisions help in the fight against COVID-19
15:55 - 16:05		**Mini Break**
16:05 - 16:30		**Closing remarks & hackathon winner(s) announced**
16:30 - 17:00		**Networking Activity**
17:00		**End of Conference**

Reproducible and FAIR modelling with CellML and libCellML.

Hugh Sorby, Alan Garny, Keri Moyle, David Nickerson
Auckland Bioengineering Institute, University of Auckland
h.sorby@auckland.ac.nz and d.nickerson@auckand.ac.nz

ABSTRACT / INTRODUCTION

Modelling complex processes is a difficult, costly, and error-prone undertaking. We know from experience that over 90% of computational models of biological systems published in the scientific literature are neither reproducible nor reusable. Creating reproducible and reusable models can be made easier and less error prone through the use of modular, reproducible, and interoperable model description formats. CellML (https://cellml.org/) is one such format that can be used to encode models in a FAIR way (https://go-fair.org/fair-principles/). libCellML (https://libcellml.org/) is the software library that implements the principles of CellML in a cross-platform, multi-language manner. With CellML models and their annotations stored in the Physiome Model Repository (https://models.physiomeproject.org/), we make our models findable and accessible. The harmonised annotation of our models and the availability of libCellML on all operating systems, and through Web technologies, make our models interoperable. CellML is inherently modular, enabling reuse, and by making libCellML easy for tool developers to use, we encourage the reuse of existing models. Through the evolution of CellML and the robustness of libCellML, our goal is to improve the credibility of computational modelling in the domain of life sciences.

ABOUT THE AUTHOR(S)

- Alan Garny is a Senior Software Engineer at the Auckland Bioengineering Institute, University of Auckland. His background is in both software engineering and computational biology. He is the project manager and lead developer of OpenCOR (https://opencor.ws/), a modelling environment for organising, editing, simulating and analysing CellML documents on Windows, Linux and macOS.

- Keri Moyle is a Software Developer at the Auckland Bioengineering Institute, University of Auckland and has worked as a programmer for libCellML since March 2019. Before then, her background involved research in computational fluid dynamics, dynamic mesh generation, and teaching & learning.

- David Nickerson is a Senior Research Fellow at the Auckland Bioengineering Institute, University of Auckland, where he is involved in several computational physiology projects. David is a co-director of the NIH-funded Center for Reproducible Biomedical Modeling, where he leads a curation service to provide scientific journals an evaluation of reproducibility for modelling manuscripts submitted for publication.

Return to Programme

- Hugh Sorby is a Senior Software Engineer at the Auckland Bioengineering Institute, University of Auckland, where he is involved in a multitude of software projects from front-end Web development to Python scripts to continuous integration systems to high performance computing.

Return to Programme

Epithelial Modelling Platform: A Tool for Model Discovery and Construction of Computational Models of Epithelial Transport

Dewan Sarwar and David Nickerson
Auckland Bioengineering Institute, University of Auckland
dewan.sarwar@auckland.ac.nz and d.nickerson@auckand.ac.nz

ABSTRACT / INTRODUCTION

Scientists often leverage computational models of biological systems to investigate hypotheses which are difficult or prohibitively expensive to achieve experimentally. Such investigations are best achieved by utilizing suitable computational models, reusing existing validated models where possible and creating novel models consistently as needed. This requires tools which enable the discovery and exploration of existing models matched with assistance in constructing and testing new models.

We have developed a web-based tool, the Epithelial Modelling Platform, for scientists to discover relevant models and then assemble these into a novel model customized for investigating their hypotheses. While our tool specifically focuses on epithelial transport, by utilizing relevant community standards and publicly accessible knowledge repositories, it is extensible to other areas of application. The platform abstracts underlying mathematics of the computational models and provides a biologically inspired visual environment for creating custom epithelial cell models.

To help automate the verification of novel mathematical models created using our tool, we have implemented a feature to discover existing simulation experiments which match the features of the model. By executing these simulation experiments with the novel models and comparing to previous model predictions we are able to provide the user with some measure of automated testing.

ABOUT THE AUTHOR(S)

- Dewan Sarwar is a Research Assistant and a PhD Candidate at the Auckland Bioengineering Institute at the University of Auckland. Prior to this, Dewan did his Master of Science from Chalmers University of Technology, Sweden and Bachelor of Science from Khulna University of Engineering and Technology, Bangladesh.

- David Nickerson is a Senior Research Fellow at the Auckland Bioengineering Institute, University of Auckland, where he is involved in several computational physiology projects. David is a co-director of the NIH-funded Center for Reproducible Biomedical Modeling, where he leads a curation service to provide scientific journals an evaluation of reproducibility for modelling manuscripts submitted for publication.

Return to Programme

Physiome curation system for reproducible models

Karin Lundengård, Auckland Bioengineering Institute, karin.lundengard@auckland.ac.nz
David Nickerson, Auckland Bioengineering Institute, d.nickerson@auckland.ac.nz
Anand Rampadarath, Auckland Bioengineering Institute, a.rampadarath@auckland.ac.nz
Tommy Yu, Auckland Bioengineering Institute, tommy.yu@auckland.ac.nz

ABSTRACT / INTRODUCTION

Reproducibility and confirmation of results is crucial for useful science and should be one of the supporting pillars of good research. However, less than 10% of the computational physiological models published in scientific journals reproduce the published results when implemented by another group. *Physiome* is a journal committed to reproducibility and reusability of mathematical models of physiological processes. Every model published in *Physiome* is connected to a curated and permanent version of the model code with a persistent identifier. The code necessary to run the model is easily accessible to be reused as it is or as a module in a novel model. Model validation and scientific value is ensured by being connected to a primary paper published in a domain-specific journal. The Physiome Project have collaborated with Digital Science to build an open-source curation system (based on the collaborative knowledge foundation software), with journal articles published open access in figshare. The effort in demonstrating reproducibility by building reproducibility into scientific workflows from the beginning can also be reduced through the use of technologies such as gigantum. Using these systems, *Physiome* publications are a complement to your primary article that ensures reproducibility, reusability and discoverability of you model.

ABOUT THE AUTHOR(S)

- Karin Lundengård is a Post-doc from Linköping University, Sweden. She is an Editor and Curator for the new scientific journal *Physiome*. She also models the human cervical spine to help whiplash patients.
- David Nickerson is a Senior Research Fellow at the Auckland Bioengineering Institute, University of Auckland, where he is involved in several computational physiology projects. David is a co-director of the NIH-funded Center for Reproducible Biomedical Modeling, where he leads a curation service to provide scientific journals an evaluation of reproducibility for modelling manuscripts submitted for publication.
- Tommy Yu is a Software Developer at the Auckland Bioengineering Institute, University of Auckland, where he took the role as the lead developer of the Physiome Model Repository and is still the maintainer of the associated infrastructures for its day-to-day operation.
- Anand Rampadarath is a Model Curator/Research Assistant at the Auckland Bioengineering Institute, University of Auckland. He is the main curator for *Physiome*

Return to Programme

as well as the NIH-funded Center for Reproducible Biomedical Modeling, and currently serves as a Technical Advisor for a Reproducibility pilot project with PLoS Computational Biology.

Return to Programme

Scaling: it's all (well, quite a lot!) about the workflow

Hilary Oliver

Senior Research Software Engineer, NIWA

Hilary.Oliver@niwa.co.nz

In scientific computing "scaling" normally refers to large number-crunching models running efficiently in parallel on many compute nodes. After touching on this topic, Hilary will shamelessly take a much looser interpretation of the term so that he can talk about his own experience in using workflow orchestration to get better use of the hardware in a more holistic sense, and in scaling a small research software project up to a much bigger one with critical large-scale production use.

Return to Programme

Parallel CFD Modelling to Explore Scale Effects in Coastal Engineering Wave-Structure Interaction (WSI) simulations

Pablo Higuera
The University of Auckland
pablo.higuera@auckland.ac.nz

ABSTRACT / INTRODUCTION

Small-scale testing in coastal engineering is a valuable tool to aid in the design of structures at a reasonable cost. Laboratory tests are usually performed maintaining the balance between gravity and inertial forces between scales (Froude scaling). As a result, the balance between inertial and viscous forces (Reynolds scaling) is not maintained, leading to scale effects when extrapolating results. This makes, for example, turbulence to become decreasingly important at smaller scales and can introduce substantial uncertainties in the design process. In principle, CFD models allow performing simulations at any scale, hence eliminating scale effects. However, prototype-scale CFD simulations present challenges on their own. For example, large-scale CFD simulations require a much finer mesh resolution relative to small-scale modelling, to account for turbulence properly. This leads to massive meshes that require efficient parallel models to be solved.

In this presentation we will show how the CFD code olaFlow (developed within OpenFOAM) can help to quantify scale effects for non-reflecting breakwaters subjected to a wide range of wave conditions. We will present laboratory-scale simulations to validate the model and prototype-scale simulations (50 times larger) with which to identify changes in the fluid behaviour caused by scale effects.

ABOUT THE AUTHOR(S)

- Pablo Higuera
- Dr Pablo Higuera is a Lecturer in Coastal Engineering at the Civil and Environmental Engineering Department of the University of Auckland. His research field is Computational Fluid Dynamics (CFD) applied to coastal hydrodynamics, wave-structure interaction, hydraulics and environmental flows. Pablo is currently developing the open-source CFD model olaFlow within OpenFOAM to simulate coastal and offshore wave processes. His research interests include adaptation of coastal structures to climate change, numerical simulation of renewable energies (e.g., Oscillating Water Column -OWC- devices), artificial surf reefs and machine learning.

Return to Programme

Parallel Computing with Dask

Wolfgang Hayek, Maxime Rio, Chris Scott
NeSI
wolfgang.hayek@nesi.org.nz, maxime.rio@nesi.org.nz, chris.scott@nesi.org.nz

ABSTRACT / INTRODUCTION **(Up to 200 words)**

Parallel computing has become a necessity for a wide range of modern scientific computing problems, including data-oriented computing at large scale to achieve reasonable processing times. Implementing parallel computing can be challenging and time-consuming - APIs such as the Message Passing Interface (MPI) are powerful but can be hard to learn and implement.

Dask is a popular toolkit for the Python programming language that addresses this issue. While requiring very little programming effort, it offers a variety of parallelisation paradigms, including work sharing via parallel function evaluation, task graphs, and direct integration with packages such as NumPy, Pandas, and Scikit-Learn. Dask can be used interactively and as a batch processing tool. The Dask-MPI package adds MPI as a parallelisation backend, enabling scalability and high throughput on high-performance computing (HPC) systems.

In this presentation, I will introduce Dask, discuss some of its parallelisation mechanisms, and demonstrate how to use the MPI backend for batch processing.

[Note: This presentation should precede Maxime Rio's demo of using Dask with SciKit-learn in Jupyter notebooks as it will cover off the basics of Dask.]

ABOUT THE AUTHOR(S)

Wolfgang Hayek is a research software engineer at NeSI and NIWA, and group manager of NIWA's scientific programming group, with many years of experience in scientific computing and HPC.

Maxime is a data scientist at NeSI and NIWA. He enjoys helping researchers to analyse their data, from visualisation to probabilistic modelling.

Chris Scott is a Research Software Engineer at NeSI with a background in scientific computing and HPC.

Return to Programme

OpenACC pgfortran: substantial speedups and beyond for the O(3) Condensation algorithm for determinants and estimation

Authors names: Drs. Damien Mather and Chris Scott
Organisations: University of Otago and NeSI
Authors Emails: damien.mather@otago.ac.nz; chris.scott@nesi.org.nz

ABSTRACT / INTRODUCTION

This milestone achievement report for NeSI project uoo02741 "Extracting D-efficient training samples.." and is an informative case study in applying Open acceleration (OpenACC) directives to MPI fortran using PGI's smart pgfortran compiler on NeSI's Mahuika platform. We utilise intel MPI libraries and up to 4 of Mahuika's P-100 GPUs per batch job and show (a) how substantial speedups can be had with four additional OpenACC compiler directives, and (b) how evolving the algorithm to optimise data locality and reduce process blocking can achieve further substantial speedup. We also demonstrate how these can be achieved consistently in practice across a wide variety of computing platforms from legacy CPUs and Nvidia accelerator cards through to NeSi's HPC platforms. The Condensation algorithm used in this demonstration has superior scaling performance and immunity to ill-conditioning for both the calculation of determinants, and, by extension, to the estimation of large predictive analytic systems of linear and linearised equations, whilst retaining O(3) computational complexity similar to the widely used Gaussian Elimination based methods.

ABOUT THE AUTHOR(S)

Damien Mather Bio: Damien is a keen runner and Senior Lecturer in the Department of Marketing in the University of Otago Business School. He teaches a postgraduate course in predictive analytics and has had a persistent interest in statistics and computer science, especially FORTRAN, C/C++, SAS, and has App Dev Expertise in electromagnetic and electroacoustic dynamics, embedded systems design, modelling and visualisation and text analytics for the social sciences.

Chris Scott Bio: Chris is a keen swimrunner and Research Software Engineer working in NeSI's Consultancy Service. He has a background in materials modelling and has many years' experience in scientific computing and HPC.

Return to Programme

NeSIssity is the mother of invention

Rhys Owen, Rere-No-A-Rangi Pope, Sydney Shep
Te Herenga Waka, Wellington University of Wellington
rhys.owen@vuw.ac.nz rere.pope@vuw.ac.nz sydney.shep@vuw.ac.nz

The impact of nineteenth-century Māori land confiscations is a lived experience in Aotearoa New Zealand today. Identifying, contacting, engaging missing shareholders constitutes an enormous challenge for Māori corporations, iwi and hapū across Aotearoa New Zealand. Without accurate data or tools to harmonise existing fragmented or conflicting data sources, issues around land succession, opportunities for economic development, and maintenance of whānau relationships are all negatively impacted. Researchers at Victoria University of Wellington have developed a culturally-tuned semantic web/linked open data (CIDOC-CRM) information architecture as a framework to knit together and explore relevant and accessible data in order to find missing shareholders for Taranaki-based Parininihi ki Waitotara Corporation, Inc (PKW). In order to wrangle messy historic data, we built our pipelines using environments such as GitLab, Docker and Singularity on our individual workstations, then moved to VUW's new HPC Rāpoi. Complexities in agile scheduling, data intensive compute requirements, and Bayesian inference development meant we, by necessity, shifted to NeSI. Our project has been an exercise in scaling up, scaling out and portability that has made us adopt advanced new toolings. Throughout the process, we maintain our core kaupapa of reconnecting whānau to whenua via whakapapa.

Rhys Owen https://orcid.org/0000-0001-9078-2381

Rhys Owen (Te Rarawa ki Hokianga) is a Software Engineer with an interest in computational logic and formal languages. Current research areas are knowledge representation and machine learning, and Māori information and knowledge systems.

Rere-No-A-Rangi Pope https://orcid.org/0000-0003-2304-7111

Rere (Ngāruahinerangi) is a Research Software Engineer with a special interest in the intersection between Māori knowledge systems and Western Science. At the core of his thinking is how digital technologies can sustain indigenous knowledge.

Sydney Shep https://orcid.org/0000-0003-0699-3739

is Co-Team Leader and Co-Principal Investigator of the National Science Challenge, Science for Technological Innovation project "Kimihia te Matangaro | Finding the Missing." She specialises in digital humanities and cultural history.

Return to Programme

Processing Calcium Signalling Fluorescence Microscopy Image Stacks

John Rugis, James Sneyd, David Yule
University of Auckland, University of Rochester
j.rugis@auckland.ac.nz

ABSTRACT / INTRODUCTION

Calcium signalling plays an important role in the functioning of the cells in salivary glands. Recent advances enable real-time *in vivo* microscopy imaging of fully intact cells. With fluorescence microscopy each resultant image is a 2D cut-plane through the sample and these are generally accumulated over time. Images can also be acquired at levels of increasing depth to provide 3D structural detail. The image stacks from a single experiment can easily consume over 1GB of storage and, in the past, analysis of the data from a single experiment took several weeks.

In this presentation we show details of a collection of python based jupyter notebooks that we created for pre-processing and analysing our image stacks as well as for post-processing the results to extract additional information. Initial analysis of an experiment can now be done in under twenty minutes. The notebooks are used by lab staff in a “cookbook” like fashion with very little training required.

In the interest of reproducibility across different labs as well as open science, which our funding agency encourages, the code for our notebooks is hosted on GitHub.

ABOUT THE AUTHOR(S)

- Dr John Rugis
  John is currently a Scientific Programmer in the Department of Mathematics at the University of Auckland. His primary research interest is in computer based 3D modelling and visualisation.
- Professor James Sneyd
  James is a Professor in the Department of Mathematics at the University of Auckland. He specialises in the numerical modelling of calcium signalling in biological cells and has authored many papers and several books on the topic.
- Professor David Yule
  David holds a number of roles at the University of Rochester Medical Center in New York and heads his own lab there. He specialises in experimental cell biology with an emphasis in a range of aspects of calcium signalling and has published numerous papers on the topic. His research work is funded by the National Institutes of Health.

Return to Programme

Native distributed and MPI parallelism in the high-level language Julia for quantum Monte Carlo

Mingrui Yang[1,2,3], Elke Pahl[4,3,5], Joachim Brand[1,2]

[1]New Zealand Institute for Advanced Study and Centre for Theoretical Chemistry and Physics, Massey University, Auckland, New Zealand
[2]Dodd-Walls Centre for Photonic and Quantum Technologies, New Zealand
[3]MacDiarmid Institute for Advanced Materials and Nanotechnology, New Zealand
[4]Department of Physics, University of Auckland, Auckland, New Zealand
[5]School of Natural and Computational Sciences, Massey University and Centre for Theoretical Chemistry and Physics, Auckland, New Zealand

M.Yang4@massey.ac.nz; Elke.Pahl@auckland.ac.nz; J.Brand@massey.ac.nz

ABSTRACT / INTRODUCTION (Up to 200 words)

Julia is a very young programming language but has already proven powerful. Its parallel computing feature provides essential tools for efficient implementation of the quantum Monte Carlo method in computational physics research. The quantum Monte Carlo method is a parallelisable approach for finding the eigenvalue and vectors of a gigantic matrix which appears very often in quantum many-body problems. For example, finding the ground state energy (the eigenvalue) and wave function (the eigenvector) of a 200-boson chain involves a 10^{118}-element vector and a 10^{118}x10^{118} matrix. The Monte Carlo method utilises random numbers to work the magic of diagonalising a matrix without building the full matrix. In this talk, we showcase an implementation of the Monte Carlo algorithm for solving quantum many-body problems in Julia, and how the parallelisation is done with the built-in native distributed computing and the external MPI.jl package. We also present experiences with profiling and running the parallelised Julia code on the NeSI cluster as the result of a NeSI consultancy project, as well as the challenges we have faced within the consultancy project.

ABOUT THE AUTHOR(S)

- Name: Mingrui Yang
- Bio: Mingrui is a PhD student from Massey University, with the background of computational chemistry and physics.
- Name: Elke Pahl
- Bio: Dr. Elke Pahl is a senior lecturer from University of Auckland, Her research spans topics in computational physics and chemistry and solid state physics with an emphasis on computational modelling.
- Name: Joachim Brand
- Bio: Prof. Joachim Brand is a theoretical physicist from New Zealand Institute for Advanced Study, Massey University. His research expertise is in ultracold atomic gases.

Return to Programme

Harnessing more compute power on NeSI's platforms with Open Multi-Processing (OpenMP)

Alexander Pletzer, Wolfgang Hayek and Chris Scott
NeSI
Alexander.Pletzer@nesi.org.nz

ABSTRACT / INTRODUCTION

You know the feeling -- your application used to hum happily but after increasing the problem size it seems to have hit a wall. No matter how much you tweak the algorithm, play with the compiler options or tune the input parameters, your code now takes too long to execute. This is the point where more computational power is required. Here we explore a few strategies that will enable a serial problem to scale up, either by leveraging more cores or by offloading computations to a Graphical Processing Unit (GPU). With the most recent OpenMP standard, you can now do both. The pros and cons of each approach will be discussed and comparisons with Open Accelerators (OpenACC), an alternative technology, will be provided.

ABOUT THE AUTHOR(S)

- Alex Pletzer is a HPC research software engineer for NeSI and based at NIWA in Wellington. Alex enjoys playing ping pong with his colleagues, sailing and windsurfing while running scientific code on NeSI platforms.

Return to Programme

Practical recipe standards for Singularity containers to make life easier for Data Scientists

Eric Burgueño
The New Zealand Institute for Plant & Food Research Ltd.
eric.burgueno@plantandfood.co.nz

ABSTRACT / INTRODUCTION **(Up to 200 words)**

Reproducible research is a fundamental goal of scientific computing, and containers are tremendously helpful at this. Ensuring the portability of an analysis environment is fundamental for the peer-review process, even when data gravity continues to be challenging.

It is no secret that software packagers face the daunting task of dealing with dependency and integration hell. Yet, creators of scientific software continue to miss the memo and leave the problem of software distribution unsolved, or at the whims of a sysadmin who must often "hack the code" to deploy it in their HPC environment. Container technologies such as Docker and Singularity can definitely help, but they come with their own set of limitations which are often not known or poorly understood.

In this talk I will share my learnings from publishing the containers we created for powerPlant, our in-house HPC cluster. I will cover some of the strategies that make container recipes as small as possible, how to deal with raw data, how we integrate with Environment Modules, and in general how to make life easier for Data Scientists so that technology does not get in the way of Science.

ABOUT THE AUTHOR(S)

- **Name:** Eric Burgueño
- **Bio:** Hello there! I am an IT professional specialising in GNU/Linux and Open Source. I also have a Law degree but computers are my true passion, so I wrangle a bunch of them at the HPC Services Team at Plant & Food Research.
  I have approximate knowledge of many things. I am a science enthusiast and an aspiring polyglot (in both human and computer lingos). I use Oxford commas and indent my code with spaces. When I take a break from being a geek, I enjoy discovering the world and other cultures, particularly if there's food, wine, or beer involved.

Return to Programme

funcX: A Federated Function Serving Fabric for Science

Kyle Chard*^, Yadu Babuji*^, Zhuozhao Li*^, Tyler Skluzacek*, Anna Woodard*, Ben Blaiszik*^, Daniel S. Katz+, Ian Foster*^

* University of Chicago, ^ Argonne National Laboratory, +University of Illinois at Urbana-Champaign

chard@uchicago.edu, yadu@uchicago.edu, zhuozhao@uchicago.edu, skluzacek@uchicago.edu, annawoodard@uchicago.edu, blaiszik@uchicago.edu, dsk@illinois.edu, foster@anl.gov

ABSTRACT / INTRODUCTION

Exploding data volumes and velocities, new computational methods and platforms, and ubiquitous connectivity demand new approaches to computation in the sciences. These new approaches must enable computation to be mobile, so that, for example, it can occur near data, be triggered by events (e.g., arrival of new data), be offloaded to specialized accelerators, or run remotely where resources are available. They also require new design approaches in which monolithic applications can be decomposed into smaller components, that may in turn be executed separately and on the most suitable resources. To address these needs we present funcX—a distributed function as a service (FaaS) platform that enables flexible, scalable, and high-performance remote function execution. funcX's endpoint software can transform existing clouds, clusters, and supercomputers into function serving systems, while funcX's cloud-hosted service provides transparent, secure, and reliable function execution across a federated ecosystem of endpoints. We motivate the need for funcX with several scientific case studies, present our prototype design and implementation, show optimizations that deliver throughput in excess of 1 million functions per second, and demonstrate, via experiments on two supercomputers, that funcX can scale to more than more than 130 000 concurrent workers.

ABOUT THE AUTHOR(S)

Kyle Chard is a Research Assistant Professor in the Department of Computer Science at the University of Chicago and a joint appointee at Argonne National Laboratory. He received his Ph.D. in Computer Science from Victoria University of Wellington, New Zealand. Kyle received the IEEE TCHPC Award for Excellence for Early Career Researchers in HPC, was part of the Globus team that won an R&D100 award, and was awarded a New Zealand Top Achiever Doctoral Scholarship. He co-leads the Globus Labs research group which focuses on a broad range of research problems in data-intensive computing and research data management. He leads NSF- and DOE-funded projects related to distributed and parallel computing, scientific reproducibility, research automation, and cost-aware use of cloud infrastructure.

Return to Programme

Developing an Interactive Genome Analytics Tool – an Interface to Scientific Insight

Louise Ord[1], Brint Gardner[2], Shannon Dillon[3]
[1]CSIRO IM&T Scientific Computing, South Eveleigh, NSW, Australia, [2]CSIRO IM&T Scientific Computing, Clayton, VIC, Australia, [3]CSIRO A&F, Black Mountain, ACT, Australia
[1]louise.ord@csiro.au

ABSTRACT / INTRODUCTION (Up to 200 words)

Scientists in Agriculture and Food at CSIRO are applying machine learning techniques to resolve biological factors and interactions underpinning variation in crop traits. We established a toolset capable of visualising the multi-layered output of these models. However, existent tools require unique data processing pipelines and manual manipulation of the data in order to produce interpretable results. We worked to develop a self-contained tool in Shiny, a visualisation framework in R, that incorporates JavaScript visualisation libraries as HTML widgets. This Genome Analytics tool enables CSIRO researchers to interactively explore their data in real-time and extract useful meaning in an accessible and automated way. We will describe our development approach, from the initial phase of exploration through to coding the visualisation framework. We will also discuss the considerations that were made to ensure we developed an application capable of informing scientific research.

ABOUT THE AUTHOR(S)

Dr Louise Ord

Data Analytics and Visualisation Specialist

Louise enjoys bringing data to life through analysis, exploration and visualisation. Gaining her doctorate at the University of Oxford, she spent six years as a cosmologist, analysing cosmic microwave background anisotropies. She then moved into the field of data analytics and predictive modelling where she developed a passion for data visualisation. Now at CSIRO, Louise designs and creates analytics and visualisation tools to bring data insights to scientists.

Return to Programme

Customising an open source project - the tension between convergence and divergence

C. Seal[1,2], M. Laverick[1] and Y. Wharton[1]
[1] Centre for eResearch, The University of Auckland
[2] corresponding author: c.seal@auckland.ac.nz

ABSTRACT / INTRODUCTION **(Up to 200 words)**

A recent project at the University of Auckland has focussed on implementing a scientific instrument data repository. An open source solution, MyTardis (http://www.mytardis.org/), has been selected as a basis for this repository, yet there are core features identified by end users, that are not implemented in the base system.

We have chosen to develop these features, which touch many aspects of the existing code base, and are contributing these back to the maintainers of the project. This decision was made on the basis that they are modifications that will be broadly useful to the community using MyTardis,

While we are confident that most, if not all, of our changes will be merged into the community code base, there is an implicit tension between choosing to use software, out of the box, thus limiting its potential usefulness but minimising the challenges of ongoing support and maintenance, and customising the code to meet end user requirements.

In this presentation we are seeking to share our experiences and engender discussion around this tension, including the techniques that we are using to mitigate the potential for ongoing maintenance headaches and the considerations made prior to choosing to modify the code base.

ABOUT THE AUTHOR(S)

- Dr Chris Seal
- Chris gained his PhD in Materials Engineering, at the University of Auckland, studying the effect of earthquake loading on steel framed buildings. He subsequently took a postdoctoral position in the field of computer simulation of fracture process, at the University of Manchester. Returning to New Zealand with his family, Chris works as a Senior eResearch Solutions Specialist, a role that enables the application of computational skills and knowledge of research workflows to the development of the university's scientific instrument data repository.
- Dr Mike Laverick

Return to Programme

- Mike is a former atomic astrophysicist turned eResearch Solutions Specialist at the University of Auckland. Hailing from the UK, he completed his PhD entitled “Fundamental atomic data deduced using stellar spectroscopy” at KU Leuven, Belgium, before moving to New Zealand and joining the Centre for eResearch. Mike is currently working on the development of the university’s new instrumentation data platform, and the Space Payload Operation Centre. Mike is also keen advocate of all things Python-related, contributing to digital research skills training and community events.

- Yvette Wharton

- Yvette Wharton is the eResearch Solutions Lead in the Centre for eResearch at the University of Auckland, working on the research data solutions and researcher enablement projects. She has extensive experience in University teaching, research and IT environments and is passionate about using her broad knowledge to facilitate people to achieve their aspirations.

    -

Return to Programme

Running Web Applications on HPC using containers

Marko Laban, Chris Scott
NeSI
marko.laban@nesi.org.nz, chris.scott@nesi.org.nz

ABSTRACT / INTRODUCTION **(Up to 200 words)**

NeSI supports the use of Singularity containers, which can be used to package up entire scientific workflows and software, on its High Performance Computers (HPCs). For researchers wanting to run web applications on the HPC in a scalable manner, we have come up with a process for bundling the web app into a Singularity container, running it via Slurm and exposing it to the end user's web browser via SSH tunnel.

We will share our experiences and talk about some challenges we faced as well as about future opportunities in this space.

ABOUT THE AUTHOR(S)

- Chris Scott is a Research Software Engineer working for NeSI with a background in scientific computing and HPC.
- Marko Laban is a Software Engineer working for NeSI

Return to Programme

FAIR for Research Software Working Group

Michelle Barker
Research Software Alliance (ReSA)
michelle@researchsoft.org

ABSTRACT / INTRODUCTION (Up to 200 words)

A new FAIR for Research Software Working Group (FAIR4RS WG) has been jointly convened by Research Data Alliance (RDA), FORCE11 and Research Software Alliance (ReSA) Taskforce, in recognition of the importance of this work for the advancement of the research sector. FAIR4RS WG will enable coordination of a range of existing community-led discussions on how to define and effectively apply FAIR principles to research software, to achieve adoption of these principles.

The working group will deliver: 1) A document developed with community support defining FAIR principles for research software; 2) Guidelines on how to apply the FAIR principles for research software; and 3) A document summarising the definition of the FAIR principles for research software, implementation guidelines and adoption examples.

Four initial subgroups are now 1) defining research software, 2) taking a fresh look at the FAIR principles in the context of research software, 3) examining recent work in this area, and 4) looking at how FAIR is being applied to other types of objects. This talk will provide an update on the results of these four subgroups, in the context of the entire working group's activities and plans, and share ongoing opportunities for engagement with this work.

ABOUT THE AUTHOR(S)

Dr Michelle Barker is the Director of the Research Software Alliance (ReSA). She has extensive expertise in open science, research software, digital workforce capability and digital research infrastructure. As a sociologist, Michelle is passionate about building collaborative partnerships to achieve system change.

She recently chaired the OECD Global Science Forum expert group on digital skills for the research sector, is an Advisory Committee Member of the US Software Sustainability Institute (URSSI), and has co-convened conferences including the IEEE International Conference on e-Science, Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE), International Workshop on Science Gateways and eResearch Australasia.

Michelle is a former Director of the Australian Research Data Commons, where she led the strategic planning for the Australian government's $180 million, five-year investment in ARDC, the national research software infrastructure investment program, and developed a national strategy to enhance digital workforce capacity in the research sector.

Return to Programme

Cylc user interface scalability

Bruno P. Kinoshita
NIWA
bruno.kinoshita@niwa.co.nz

ABSTRACT / INTRODUCTION **(Up to 200 words - 84)**

In this lightning talk we will go through what is happening with the new Cylc 8 user interface. The general structure of the user interface will be explained, as well as the challenges we faced when trying to display varied and large workflows. We will share lessons learned during the development, tools and techniques used, and more about our future work. The material shared in this talk might interest other research software engineers and researchers facing similar issues in web development and data visualization.

ABOUT THE AUTHOR(S)

- Bruno P. Kinoshita
- Research Software Engineer at NIWA, working with Java, PHP, Python, and JavaScript.

Return to Programme

Traffic Flow Estimation based on Deep Learning using CCTV Images

Nilani Algiriyage[1], Raj Prasanna[1], Kristin Stock[2], Emma Hudson-Doyle[1], David Johnston[1]
Joint Centre for Disaster Research, Massey University[1] ,Institute of Natural and Mathematical Sciences, Massey University[2]
R.Nilani@massey.ac.nz, R.Prasanna@massey.ac.nz, K.Stock@massey.ac.nz, e.e.hudson-doyle@massey.ac.nz, D.M.Johnston@massey.ac.nz

ABSTRACT

Traffic flow estimation is the first step in the management of road traffic infrastructure and is essential for the successful deployment of intelligent transportation systems. Closed-circuit television (CCTV) systems are now popular and are mounted in many public places to support real-time surveillance. The data generated by CCTV cameras can be used as the foundation for accurate traffic flow estimation.

The lightning talk is based on research carried out seeking to answer the questions; 1) What object detection algorithm is best suited to the CCTV image data set for vehicle detection? 2) Can traffic flow be estimated by counting the number of vehicles in CCTV images using an object detection algorithm?.

We collect real-time CCTV imagery from traffic cameras through the New Zealand Transport Agency's (NZTA) traffic cameras Application Programming Interface (API). In the first experiment, we compare the performance and accuracy of faster R-CNN, Mask R-CNN and YOLOv3 algorithms in vehicle detection task. Then, we select a case study at one of the busiest roads in Christchurch Central Business District (CBD) to estimate the traffic flow. The results can be used by city authorities to understand traffic flow patterns, make traffic predictions, understand anomalies, and make management decisions.

ABOUT THE AUTHOR(S)

- Nilani Algiriyage
- Nilani is a PhD candidate at the Joint Centre for Disaster Research at Massey University, Wellington

- Raj Prasanna
- Raj is a senior lecturer at the Joint Centre for Disaster Research at Massey University, Wellington

Return to Programme

- Kristin Stock
- Kristin is a Senior Lecturer in Information Technology and Director of the Massey Geoinformatics Collaboratory at Massey University, Auckland

- Emma Hudson-Doyle
- Emma is a senior lecturer at the Joint Centre for Disaster Research at Massey University, Wellington

- David Johnston
- David Johnston is the Professor of Disaster Management and Director of the Joint Centre for Disaster Research at Massey University, Wellington

Return to Programme

Jupyter on NeSI

Chris Scott, Yuriy Halytskyy, Marko Laban
NeSI
chris.scott@nesi.org.nz, yuriy.halytskyy@nesi.org.nz, marko.laban@nesi.org.nz

ABSTRACT / INTRODUCTION **(Up to 200 words)**

NeSI supports the use of Jupyter for interactive supercomputing via our new JupyterHub service. Through this service users can connect to JupyterLab, the next generation web based user interface for Project Jupyter, running on NeSI's HPC infrastructure using just a web browser.

We are currently in the early stages of making our JupyterHub service available to researchers. Here we would like to discuss details of, and share learnings from, our implementation of JupyterHub on NeSI infrastructure. We will also explore the different use cases researchers have for the service and the future direction and additional features we are planning to implement.

ABOUT THE AUTHOR(S)

- Chris Scott is a Research Software Engineer working for NeSI with a background in scientific computing and HPC
- Marko Laban is a Software Engineer working for NeSI
- Yuriy Halytskyy is a Systems Engineer working for NeSI

Return to Programme

How computational science and data-driven health decisions help in the fight against COVID-19

Dr. Denis Bauer
Head cloud computing bioinformatics, CSIRO

COVID-19 has highlighted the need to substantially improve disease preparedness to resolve this pandemic and avoid future ones. Digital health and the use of technology at large will be a key contributor to this preparedness. This session talks about how CSIRO has leveraged cloud-native technologies to advance three areas of the COVID-19 response: firstly we worked with GISAID, the largest data resource for the virus causing COVID-19, and use standard health terminologies (FHIR) to help collect clinical patient data. This feeds into a docker-based workflow that creates identifying "fingerprints" of the virus for guiding *vaccine developments* and investigating whether there are more pathogenic versions of the virus. Secondly, we developed a fully serverless web-service for *tailoring diagnostics* efforts, capable of differentiating between strains. Thirdly, we are creating a serverless COVID-19 analysis platform that allows distributed genomics and patient data to be shared and analysed in a privacy- and ownership-preserving manner and functioning as an *surveillance system* for detecting more virulent strains early.

Bio:
Dr Denis Bauer is CSIRO's Principal Research Scientist leading the Transformational Bioinformatics group. She is an internationally recognised expert in machine learning and cloud-based genomics, having keynoted AWS Summits ('18), Open Data Science Conference (India, '18) and International conference on Bioinformatics (Indonesia '19). Her achievements include developing open-source machine-learning cloud services that are used by 10,000 researchers and contributing to Australia's COVID-19 response.

Return to Programme