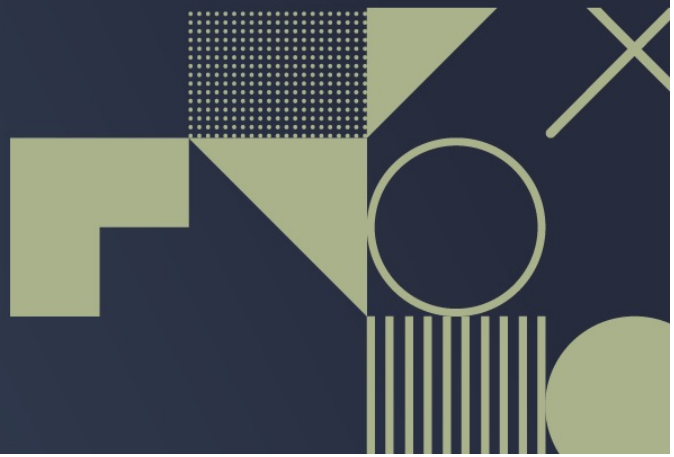
# Translational Computer Science and its Application to Supercomputing

David Abramson

Research Computing Centre

University of Queensland, Brisbane, Australia



# Ken Kennedy Award

As advertised

> For innovation in parallel and distributed computing tools with broad applications, as well as leadership contributions to professional service, creating international technical communities, and mentoring.

Hollingsworth's first amendment

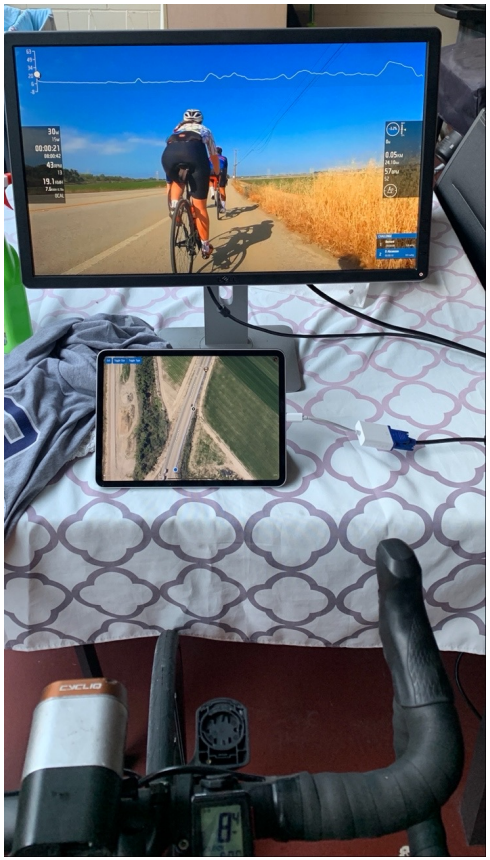
> For diligence and tensity in navigating international travel regulations to foster community building

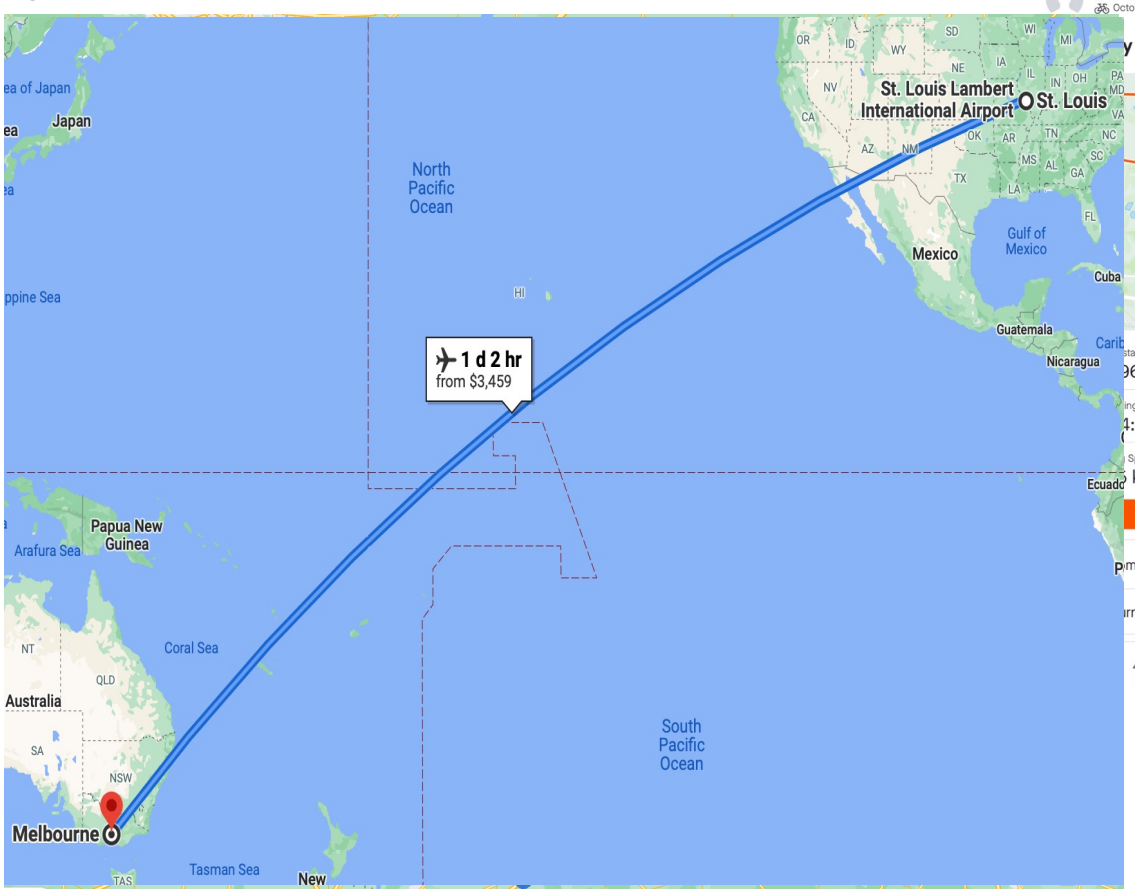
AWARDS & RECOGNITION

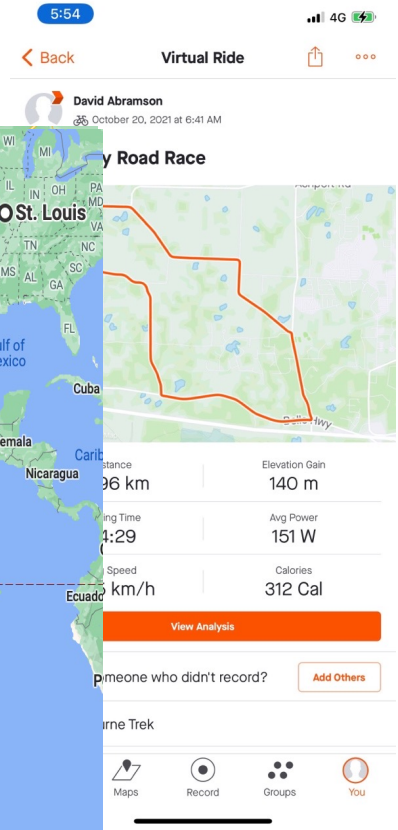
## David Abramson Recognized with Ken Kennedy Award

David Abramson of the University of Queensland has been named the recipient of the 2021 ACM-IEEE CS Ken Kennedy Award. Abramson is recognized for contributions to parallel and distributed computing tools, with application from quantum chemistry to engineering design. He also cited for his mentorship and service to eld. The award will be presented at The International Conference for High ce Computing, Networking, Storage virtual awards plenary session

# The SC community knows about virtualisation






FULGAZ

# The role of a French chateaux and red wine … with Manish Parashar

Workshop on CCDSC, Lyon France

Jack Dongarra, Bernard Tourancheau

# Introduction

- Translational Research
  - Medicine
  - Computer Science
- Exemplars
  - Nimrod
  - Guard
  - Others …
- Important issues for TCS
- Laboratory Scale Matters
- The role of students
- Translation to the arts …

# Background: Translational Medicine

- An “interdisciplinary branch of the biomedical field supported by three main pillars:
  - Benchside, Bedside and Community.
  - Combine disciplines, resources, expertise, and techniques within these pillars to promote enhancements in prevention, diagnosis, and therapies.
- Differs subtly from applied biomedical research, in which a research problem has a potential real-world application (driver).
  - Findings are applied as a specific phase of the research plan.
  - This not only demonstrates applicability and practicality, but also generates tangible outcomes.
- Now well understood and has become a de-facto standard for much of biomedical research.
- Intrinsically helps generate outcomes because the research is applied as part of the original plan, as opposed to being an afterthought once the project has completed

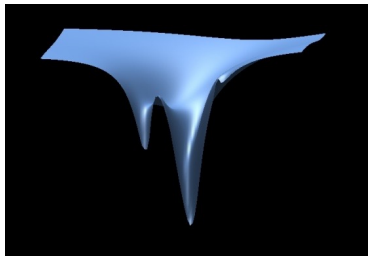
# Translational Computer Science

- In TM, translation relies on
  - Taking research from the laboratory Bench to the Bedside
  - More recent refinements involve Community
    - healthy populations, patients and medical practitioners.
- In TCS, translation relies on
  - Taking research from the laboratory Laboratory to the Locale
    - might be physical or virtual
  - Community
    - users and early adopters who work with the technology, and can include public bodies that would help in the evaluation

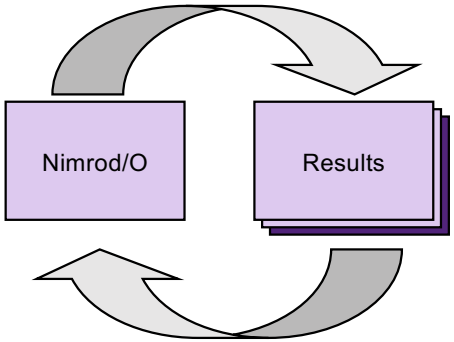
# Exemplars

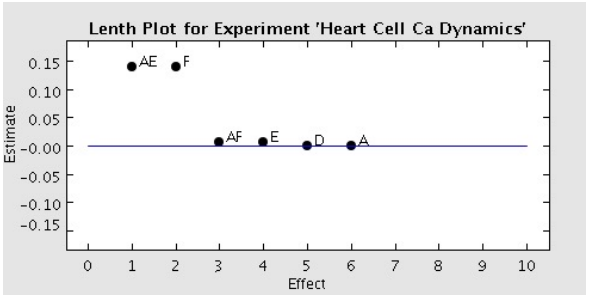
## A personal perspective

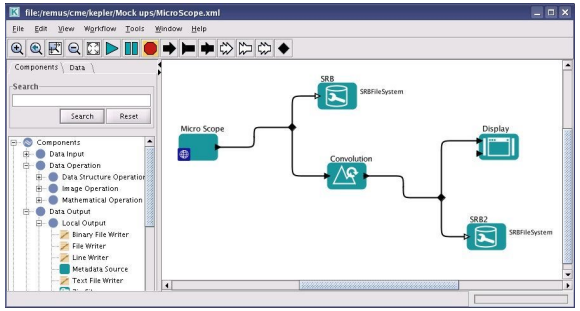
# Nimrod supporting “real” science



- Niche distributed programming environment
  - A full parameter sweep is the cross product of all the parameters (Nimrod/G)
  - An optimization run minimizes some output metric and returns parameter combinations that do this (Nimrod/O)
  - Design of experiments limits number of combinations (Nimrod/E)
  - Workflows (Nimrod/K)
- Has survived many distributed computing platforms
  - Workstations, Clusters, Grids, Clouds
- Has contributed to the understanding of HPC and distributed computing



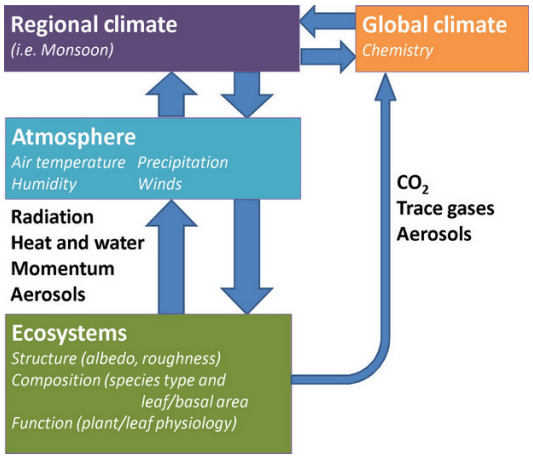



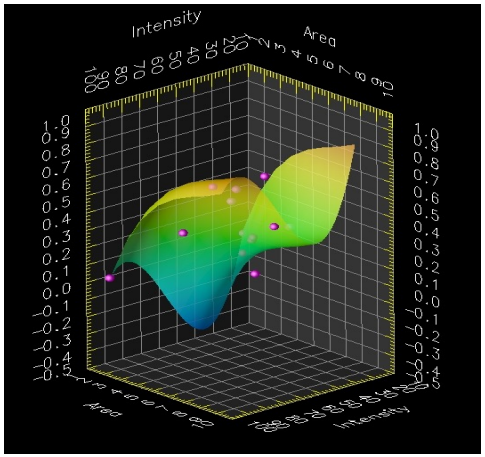



# Fire in Australian savannas: from leaf to landscape

Lynch, Beringer, Uotila Monash U, AU

# Science outcomes



Global Change Biology

Global Change Biology (2015) **21**, 62–81, doi: 10.1111/gcb.12686

RESEARCH REVIEW

## Fire in Australian savannas: from leaf to landscape

JASON BERINGER[1,2], LINDSAY B. HUTLEY[3], DAVID ABRAMSON[1], STEFAN K. ARNDT[4], PETER BRIGGS[5], MILA BRISTOW[3], JOSEP G. CANADELL[5], LUCAS A. CERNUSAK[6], DEREK EAMUS[7], ANDREW C. EDWARDS[3], BRADLEY J. EVANS[8], BENEDIKT FEST[4], KLAUS GOERGEN[9,10,11], SAMANTHA P. GROVER[1,3], JORG HACKER[12], VANESSA HAVERD[5], KASTURI KANNIAH[1,13], STEPHEN J. LIVESLEY[14], AMANDA LYNCH[1,15], STEFAN MAIER[3], CAITLIN MOORE[1], MICHAEL RAUPACH[5], JEREMY RUSSELL-SMITH[3], SIMON SCHEITER[16], NIGEL J. TAPPER[1] and PETTERI UOTILA[17]

*[1]School of Earth and Environment, The University of Western Australia, Crawley, WA 6009, Australia, [2]School of Geography and Environmental Science, Monash University, Melbourne, Vic. 3800, Australia, [3]School of Environment, Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT 0909, Australia, [4]Department of Forest and Ecosystem Science, The University of Melbourne, Melbourne, Vic. 3121, Australia, [5]CSIRO Marine and Atmospheric Research, GPO Box 3023, Canberra, ACT 2601, Australia, [6]School of Marine and Tropical Biology, James Cook University, Cairns, Qld 4878, Australia, [7]School of the Environment, University of Technology, Sydney, NSW 2007, Australia, [8]Department of Biological Sciences, Macquarie University, North Ryde, NSW 2113, Australia, [9]Meteorological Institute, University of Bonn, Bonn, D-53121, Germany, [10]Juelich Supercomputing Centre, Research Centre Juelich, Juelich 52425, Germany, [11]Centre for High Performance Scientific Computing in Terrestrial Systems, Research Centre Juelich, Juelich 52425, Germany, [12]Airborne Research Australia|Flinders University, Salisbury South, SA 5106, Australia, [13]Faculty of Geoinformation & Real Estate, Department of Geoinformation, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia, [14]Department of Resource Management and Geography, The University of Melbourne, Melbourne, Vic. 3121, Australia, [15]Department of Geological Sciences, Brown University, Providence, RI 02912, USA, [16]Biodiversity and Climate Research Centre (LOEWE BiK-F), Senckenberg Gesellschaft fur Naturforschung, Senckenberganlage 25, 60325, Frankfurt am Main, Germany, [17]Finnish Meteorological Institute, Helsinki FI-00101, Finland*

**Abstract**

Savanna ecosystems comprise 22% of the global terrestrial surface and 25% of Australia (almost 1.9 million km²) and provide significant ecosystem services through carbon and water cycles and the maintenance of biodiversity. The current structure, composition and distribution of Australian savannas have coevolved with fire, yet remain driven by the dynamic constraints of their bioclimatic niche. Fire in Australian savannas influences both the biophysical and biogeochemical processes at multiple scales from leaf to landscape. Here, we present the latest emission estimates from Australian savanna biomass burning and their contribution to global greenhouse gas budgets. We then review our understanding of the impacts of fire on ecosystem function and local surface water and heat balances, which in turn influence regional climate. We show how savanna fires are coupled to the global climate through the carbon cycle and fire regimes. We present new research that climate change is likely to alter the structure and function of savannas through shifts in moisture availability and increases in atmospheric carbon dioxide, in turn altering fire regimes with further feedbacks to climate. We explore opportunities to reduce net greenhouse gas emissions from savanna ecosystems through changes in savanna fire management.

*Keywords:* biomass burning, climate feedbacks, greenhouse gas exchange, net ecosystem carbon balance, savanna

*Received 19 December 2013; revised version received 16 April 2014 and accepted 8 June 2014*

**Introduction**

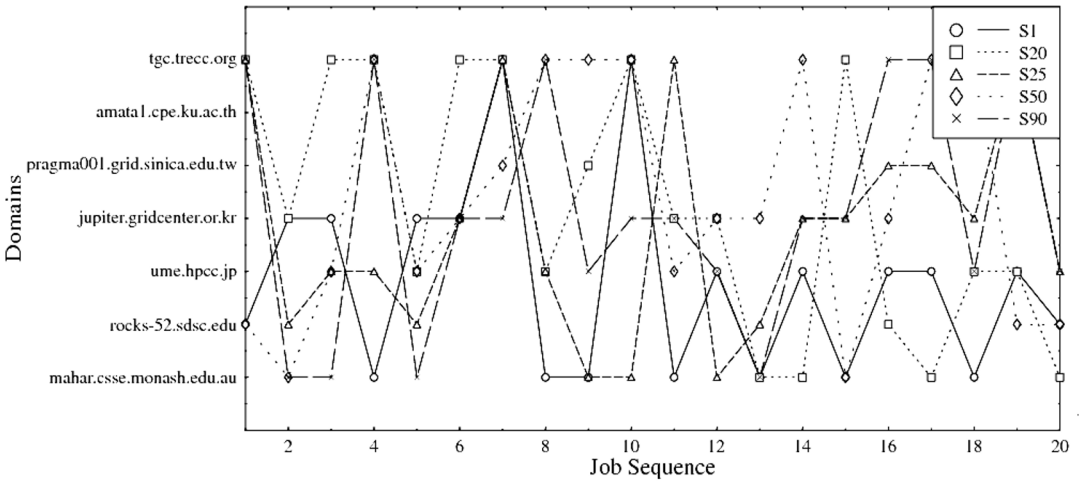
Tropical savanna ecosystems account for around 22% of the global land surface (Ramankutty & Foley, 1999). Annually, up to 75% of global tropical savanna landscapes are burned either by natural or anthropogenic fires (Hao *et al.*, 1990) and accordingly, 50% of the total annual amount of biomass burned globally takes place in the savanna region (Hao & Liu, 1994). The wet-dry tropics of northern Australia include extensive areas of savanna vegetation, which occupy approximately 1.9 million km². This area accounts for 12% of the world's tropical savanna ecosystems, making this

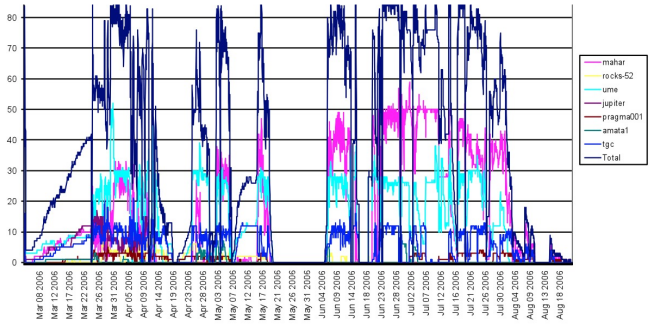
Correspondence: Jason Beringer, tel. +61 409355496, fax + 61 86488103, e-mail: jason.beringer@uwa.edu.au



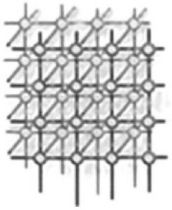
© 2014 The Authors. *Global Change Biology* Published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

# CS outcomes




CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
*Concurrency Computat.: Pract. Exper.* (2008)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cpe.1353



## Fault-tolerant execution of large parameter sweep applications across multiple VOs with storage constraints

Shahaan Ayyub*,†, David Abramson, Colin Enticott, Slavisa Garic and Jefferson Tan

*Faculty of Information Technology, Monash University, Melbourne, Vic. 3145, Australia*

**SUMMARY**

**Applications that span multiple virtual organizations (VOs) are of great interest to the e-science community. However, our recent attempts to execute large-scale parameter sweep applications (PSAs) for real-world climate studies with the Nimrod/G tool have exposed problems in the areas of fault tolerance, data storage and trust management. In response, we have implemented a task-splitting approach that facilitates breaking up large PSAs into a sequence of dependent subtasks, improving fault tolerance; provides a garbage collection technique that deletes unnecessary data; and employs a trust delegation technique that facilitates flexible third party data transfers across different VOs. Copyright © 2008 John Wiley & Sons, Ltd.**

*Received 11 December 2007; Revised 30 March 2008; Accepted 7 May 2008*

KEY WORDS: e-science; parameter sweep applications; Grid

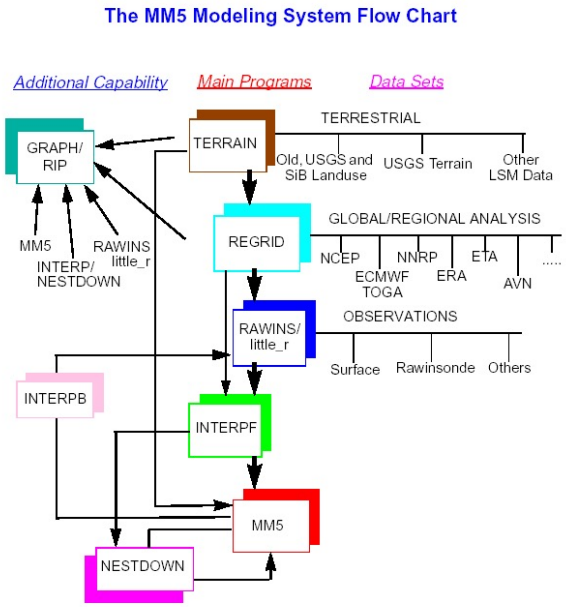
### 1. INTRODUCTION

The computational Grid aggregates computational power and storage capacity by coupling together distributed CPU, network and storage resources [1]. The scale and nature of Grid testbeds make it possible to solve particular challenging problems in science and engineering using *parameter sweep*

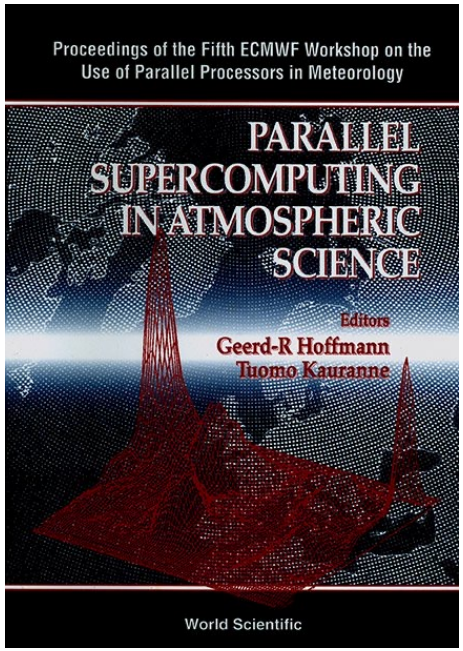
*Correspondence to: Shahaan Ayyub, Faculty of Information Technology, Monash University, Melbourne, Vic. 3145, Australia.
†E-mail: ayyub.shahaan@infotech.monash.edu.au

Contract/grant sponsor: Australian Research Council
Contract/grant sponsor: CSIRO Division of Atmospheric Research

# Debugging a climate model




MM5




MPMM

MPMM: A MASSIVELY PARALLEL MESOSCALE MODEL

Ian Foster
John Michalakes

Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, Illinois 60439

**1. Introduction**

Static domain decomposition is a technique that provides a quick path to porting atmospheric models on distributed-memory parallel computers. However, parallel inefficiencies in the form of load imbalances and ill-tuned communication are difficult to correct without complicated and explicit recoding. Reconfiguring the code to run on larger or smaller numbers of processors may require recompiling. Modularity and machine independence may also suffer. If full advantage is to be taken of massively parallel processing (MPP) technology, tools and techniques that allow for dynamic performance tuning and reconfiguration are required.

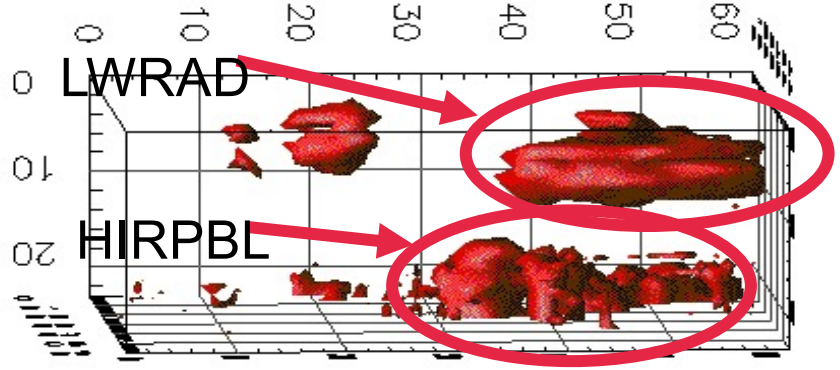
Program Composition Notation (PCN) is a language and run-time system developed at Argonne and at the California Institute of Technology for expressing parallel programs [2, 3]. It provides an intermediate layer between the application program and the physical processors of a computer. It allows the model to be statically decomposed over a *virtual* machine, but this virtual machine can be mapped and remapped dynamically over the physical computer. Programs are portable to as many machines as PCN itself, modularity is easily preserved, and communication tuning for a particular computer is encapsulated within the PCN run-time system.

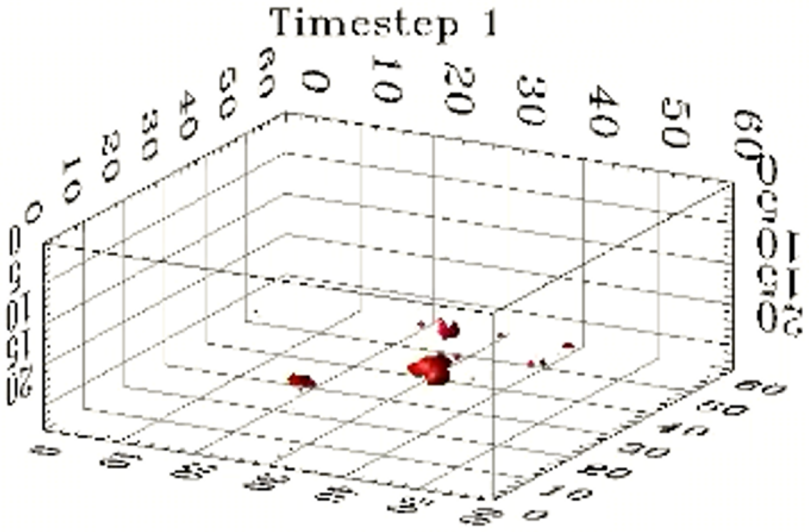
In this paper we report on a project at Argonne National Laboratory to parallelize the Penn State/NCAR Mesoscale Model version 5 using a fine grain decomposition dynamically mapped and managed under PCN.

*This work was supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38, and was performed in part using the Intel Touchstone Delta System operated by Caltech on behalf of the Concurrent Supercomputing Consortium. Access to this facility was provided by Argonne National Laboratory.

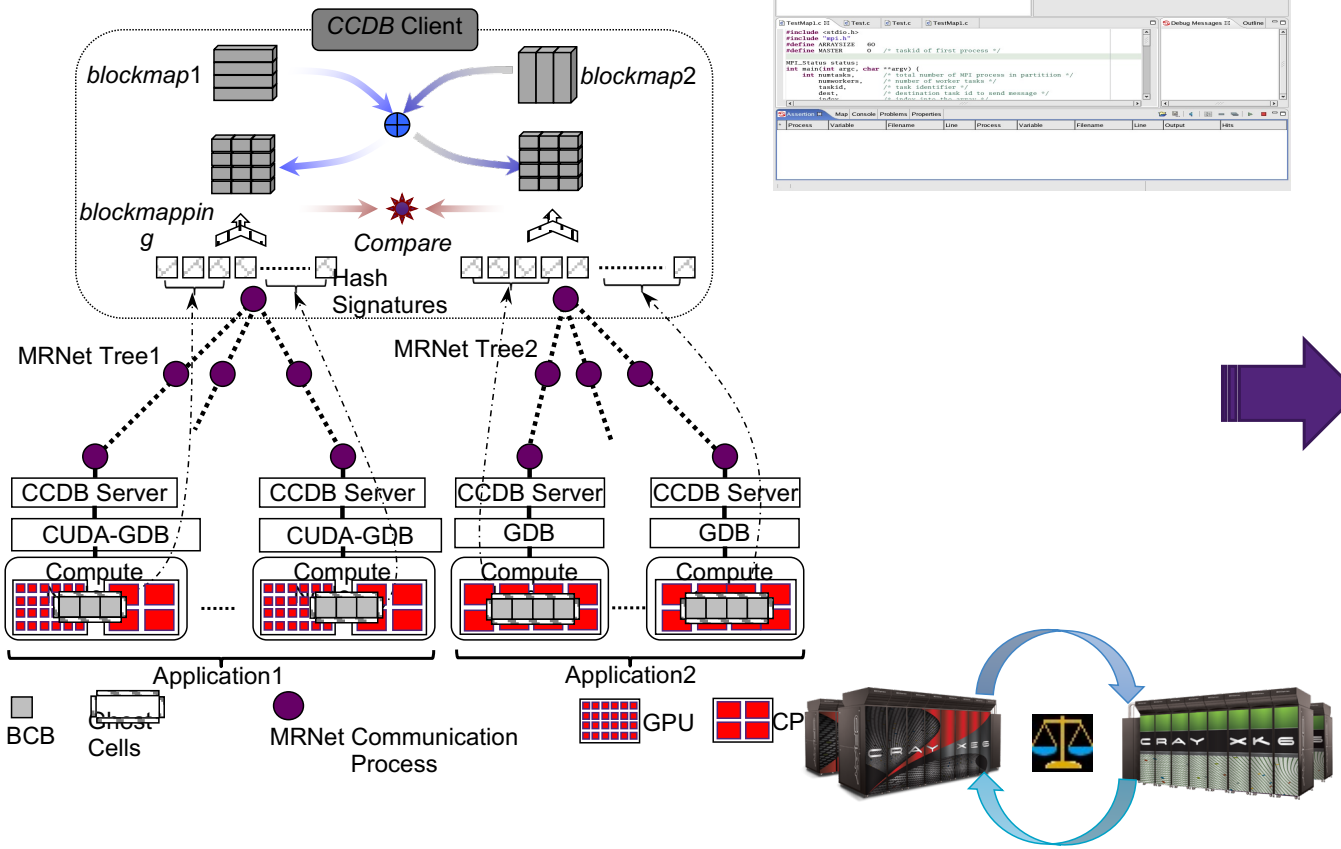
# Science outcomes

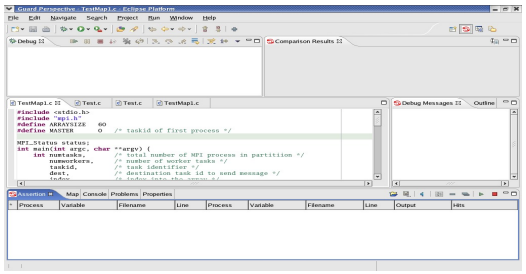
- Difference in physics of planetary boundary layer
  - Computation of #steps suited to parallel execution
  - Evident in 3 dimensional visualisation
- Error in radiation time step computation
- More complete physics in long wave radiation



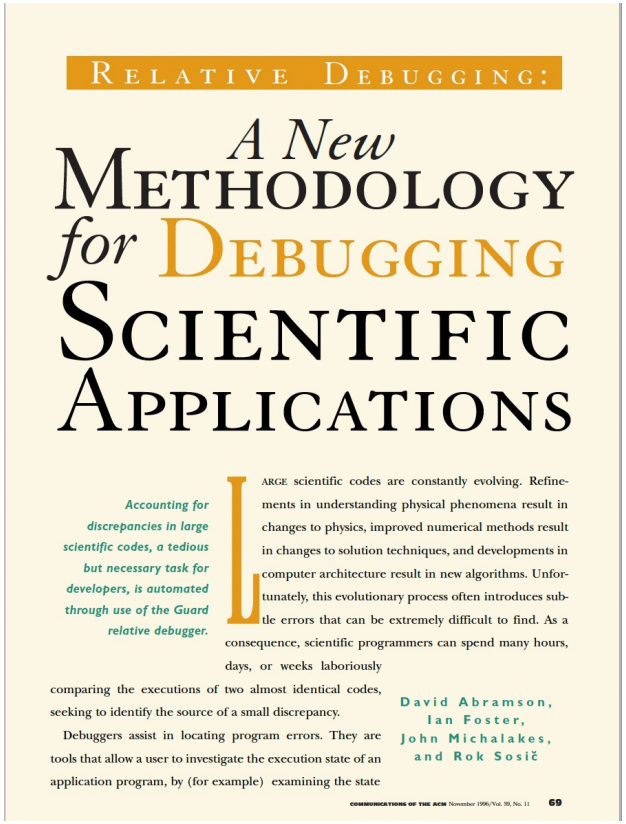



# CS outcomes





RELATIVE DEBUGGING:

A New METHODOLOGY for DEBUGGING SCIENTIFIC APPLICATIONS

*Accounting for discrepancies in large scientific codes, a tedious but necessary task for developers, is automated through use of the Guard relative debugger.*

LARGE scientific codes are constantly evolving. Refinements in understanding physical phenomena result in changes to physics, improved numerical methods result in changes to solution techniques, and developments in computer architecture result in new algorithms. Unfortunately, this evolutionary process often introduces subtle errors that can be extremely difficult to find. As a consequence, scientific programmers can spend many hours, days, or weeks laboriously comparing the executions of two almost identical codes, seeking to identify the source of a small discrepancy.

Debuggers assist in locating program errors. They are tools that allow a user to investigate the execution state of an application program, by (for example) examining the state

David Abramson, Ian Foster, John Michalakes, and Rok Sosič

COMMUNICATIONS OF THE ACM November 1996/Vol. 39, No. 11 69

# Exemplars

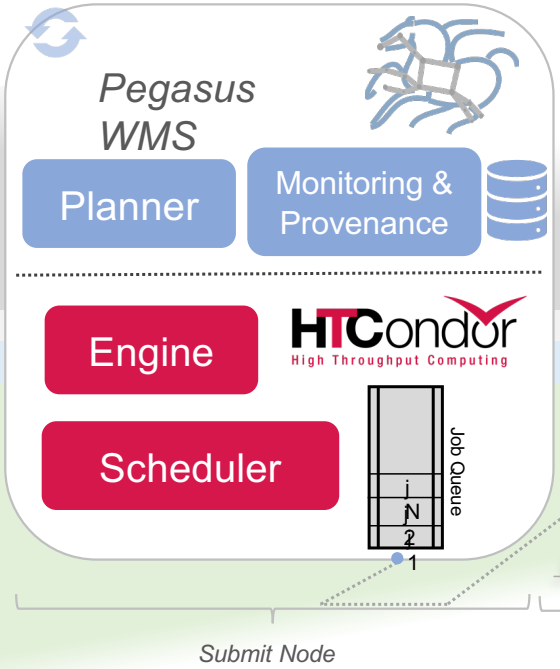
## Others

# Pegasus Workflow Management System

Ewa Deelman - University of Southern California - PI - deelman@isi.edu

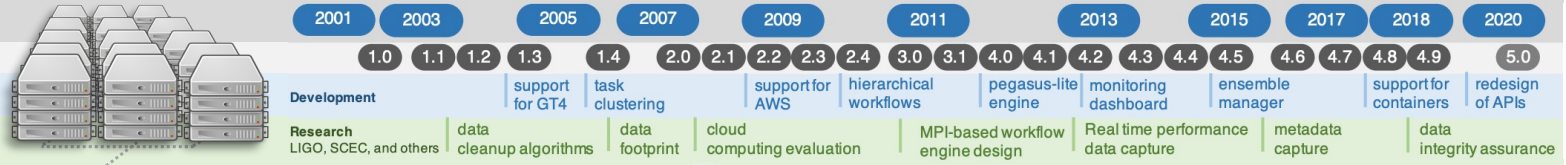
Funded Under National Science Foundation Grant #1664162

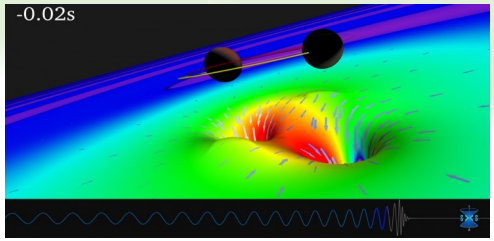
pegasus.isi.edu



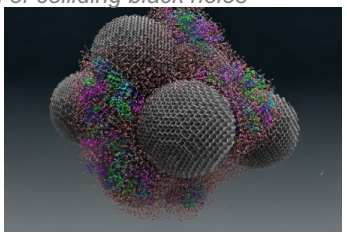

## End To End Workflow Management and Execution

- Develop ***portable scientific workflows*** in Python, Java, and R
- Compile workflows to be ***run on heterogenous resources***
- Monitor and debug workflow execution via ***CLI and web-based tools***
- Recover from failures with built-in ***fault tolerance mechanisms***
- ***Regular release schedule*** incorporating latest research and development

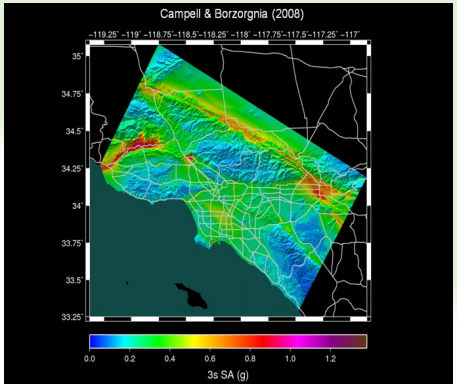



*LIGO observation of colliding black holes*



*Visualization of water on nanodiamond spheres from*



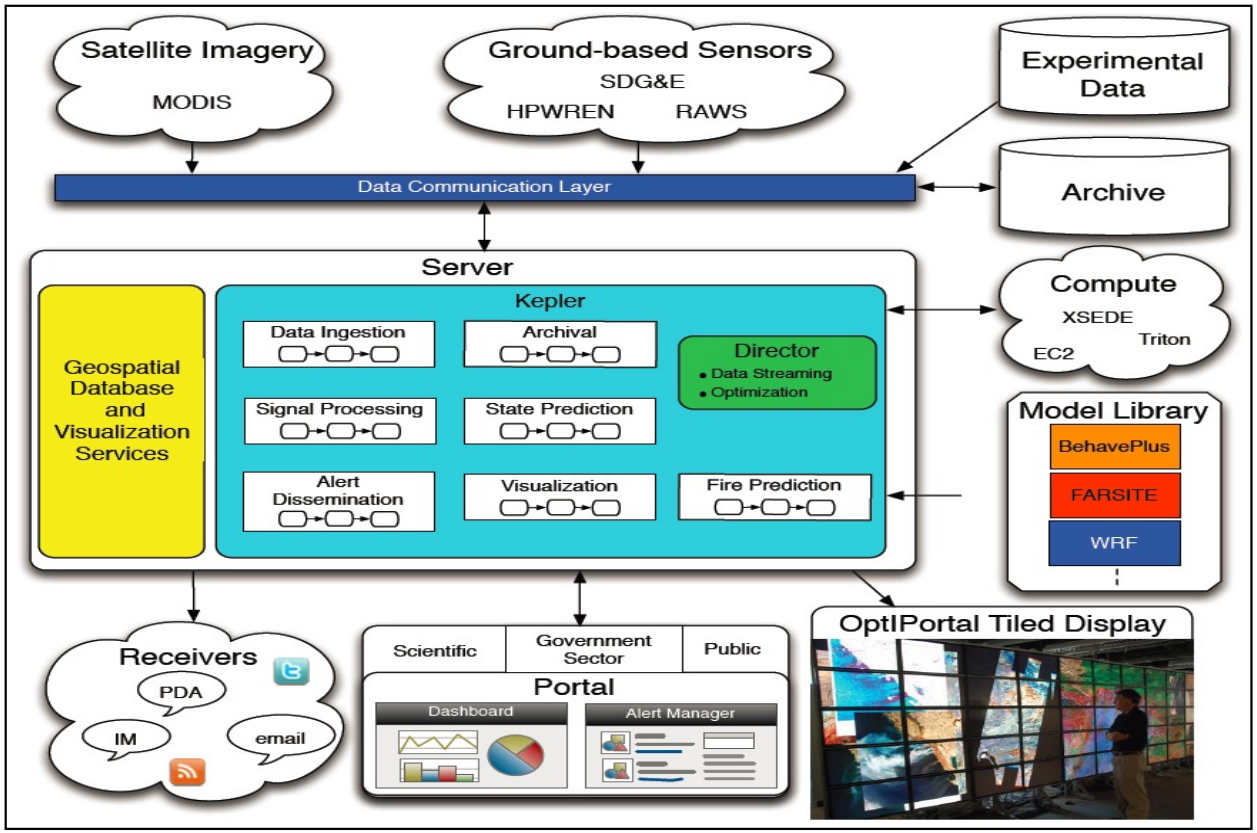

*Hazard map indicating maximum amount of shaking at a particular geographic location generated from SCEC's CyberShake Pegasus workflow*

## Pegasus In Practice

- Laser Interferometer Gravitational Wave Observatory (LIGO) develops large scale analysis pipelines used for gravitational wave detection.
- Southern California Earthquake Center (SCEC) CyberShake project generates hazard maps using hierarchical workflows .
- Oak Ridge National Lab (ORNL) conducted studies on tRNA and nanodiamonds to improve drug delivery design principles.

# NSF WIFIRE Project: 2013-2018

WIFIRE



NSF-1331615

NSF

# Grid Computing as Translational Research

Globus community testbed, 1998

**Goal**: Identify simple mechanisms to enable collaborative science in a hyper-connected world

**Translational approach**: Work closely with domain scientists, application developers, and resource providers on deployment, application, and evaluation of proposed mechanisms

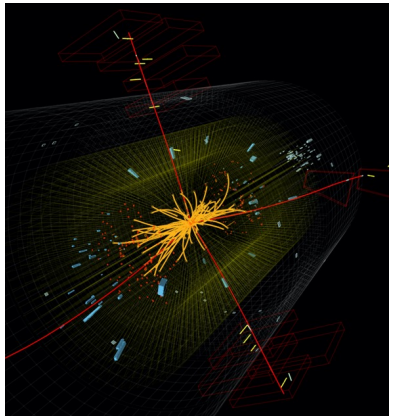
**Artifacts**: Globus Toolkit (1998-2015); Globus service (2010-)

**Science impact**: 10,000s of scientists worldwide via dozens of Grid deployments and application projects; 100s of science projects, including Nobel-winning physics, astronomy, climate

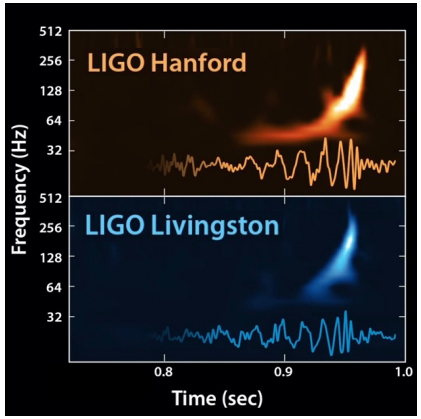
**Computer science impact**: Virtual organizations, secure and reliable remote computation, performant and reliable data sharing, workflow specification and execution

**Thanks**: Globus team, grid deployments, science collaborators, computer science collaborators, DOE, NSF, NASA, NIH

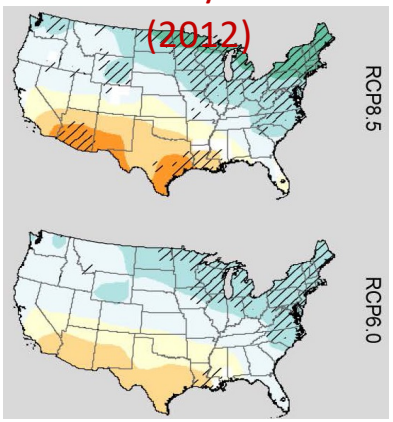
**More info**: Ian Foster (foster@anl.gov)
Carl Kesselman (carl@isi.edu)



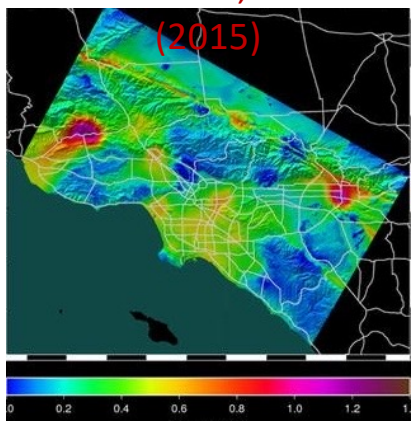
Higgs boson discovery at LHC (2012)




Gravitational wave detection, LIGO (2015)




IPCC climate science: Earth System Grid




CyberShake earthquake hazard map

https://doi.org/10.1016/j.jocs.2020.101214 – https://globus.org

# So how does this differ from traditional research pipeline?

# Typical CS research workflow

Real world

Problem Definition

Propose Solution

Prototype

Evaluate & Refine

Publish

Code

Papers

Commercialize

# Translational CS research workflow

Real world

Problem Definition

Propose Solution

Prototype

Code

Publish

Evaluate & Refine

Papers

Engage Community

Commercialize

# TCS Important issues

# Roadblocks

1. In computer science, translation is often confused with commercialization
2. Open source techniques are often confused for translation
3. Funding agencies typically don't provide support for translation
4. PhD programs don't allocate time and resources to translation
5. Traditional publication venues don't value translation
6. There are a lack of exemplars

# Translation is not commercialization

- Commercialisation almost always occurs after the research has been completed,
  - almost never funded as part of the original research proposal.
- Commercialisation implies a financial angle that has little to do with the research per-se.

# Use of Open Source

- Helps with distribution of a software system, but doesn't intrinsically drive impact
- No direct link between the way the software is used, and the research program. Thus, there is no explicit feedback from lessons learned in the adoption into the research itself.
- More focussed on producing software that is maintained in a sustainable way, by building a distributed workforce.

# Funding bodies don’t typically support translation

- Evaluation criteria typically focus on the quality of the investigator team, the project quality and innovation, the feasibility and the benefit.
- Translation is not usually highlighted as a desirable property, thus a proposal might be marked down for including translational activities.
- A budget that allocates resources to items such as a community trial, software distribution, software maintenance, may be pruned back to the basic research program.

# PhD timelines don't support translation

- Typical PhD projects in computer science follow a very standard and often rigid template.
  - Students engage in a project of interest to them
  - Execute a plan much like any other research project.
  - Milestones and deliverables include software prototypes, experiments and tests, producing publication outputs along with possibly software and data artefacts.
  - At the very least, a PhD student needs to produce a thesis.
- TR adds complexity by requiring a translation phase,
  - might extend the timeline beyond that of current PhD programs.

# Traditional publication venues don’t value translation

- Many editorial boards would argue translation is secondary to their scope,
- More focussed on primary research outcomes in computer science
- Many translational research projects are interdisciplinary,
  - Outcomes might not align well with the journal’s primary focus.
- Most journals do not publish failures.

# Lack of exemplars

- Numerous examples of computer science research being commercialised and adopted
- Few examples of successful translational research projects
- Changing the culture in an organisation is difficult because people don't know what a good TCS project looks like.

# Funding

- Currently ad hoc funding.
- Sustained funding programs and mechanisms focussed on fostering and nurturing TCS
- More money alone not solution
- Need to *build* translation into the research plan
- Funding must be used to experiment with solutions and prototypes.
- TCS typically involves substantial interaction with end users.
- Additional travel, user engagement, and provisioning of computing resources
- Translation process feeds back into the research,
  - may be a loop of research and translation rather than a linear waterfall style of workflow
- Should be free to report on both research successes, but also translation **success** or **failures**.

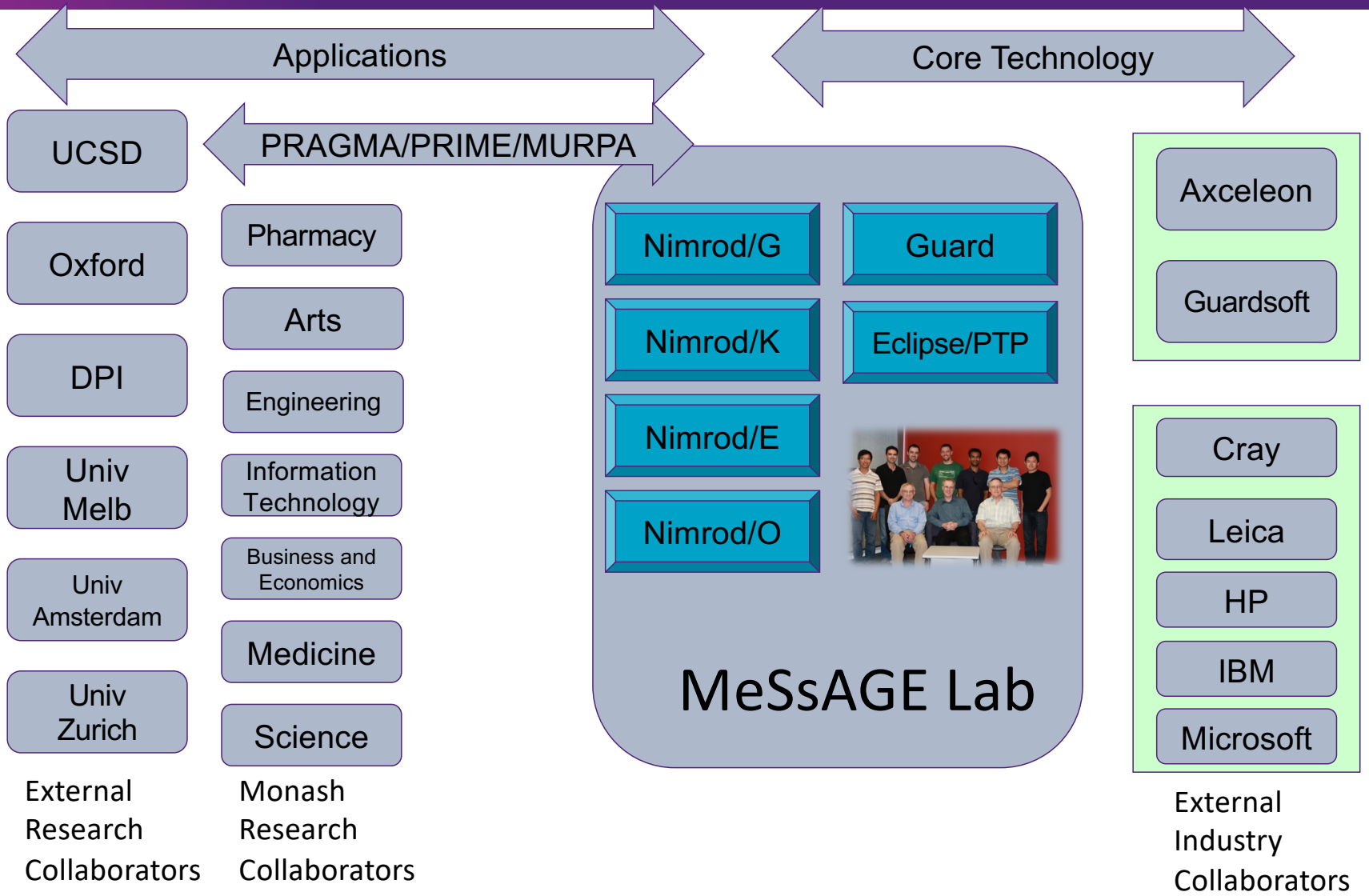
# Venues, metrics and reward structures

- Traditional publications are not well suited to TCS.
  - Drawing on TM, new journals have been created that explicitly target translational medicine.
  - New set of similarly scoped journals and conferences.
  - Metrics, recognitions and rewards structures, especially in the academic community.
    - software and data and to track their use, citations and impact are a step in the right direction
  - Metrics that report uptake of their work, and measure how many of these have resulted in successful translation
  - Integrate metrics into promotion processes

# Education and Training

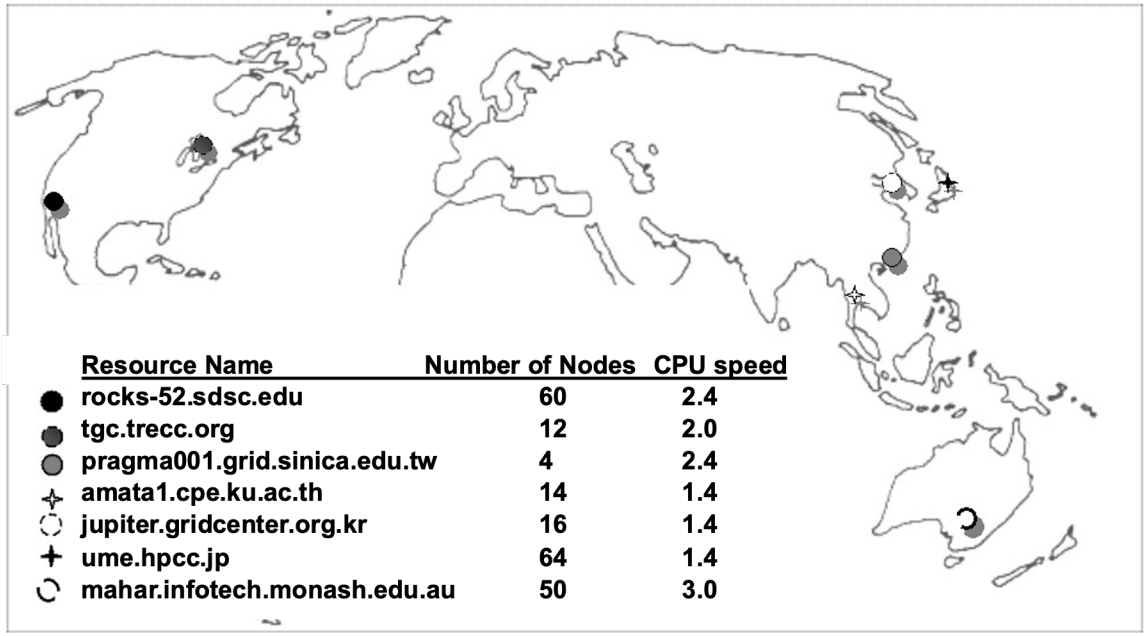
- Integration of translational approaches and methodologies into more formal computer science curricula
- New materials and mechanisms for providing translational skills to practitioners, in both computer and other disciplines.
  - doctoral training centers have been established that encourage and enable trans-disciplinary research
- Extreme example, a PhD could be entirely devoted to the translation of work performed by another researcher, with no original research on the background IP per se

# Laboratory Scale Matters

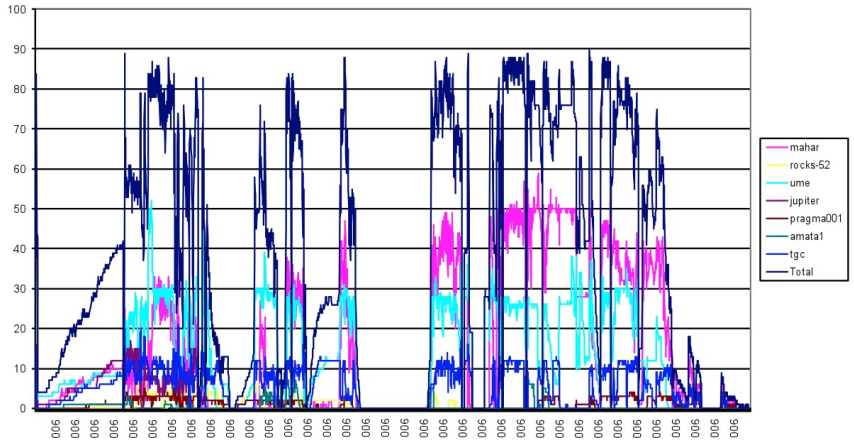

Applications
Core Technology
PRAGMA/PRIME/MURPA
UCSD
Oxford
DPI
Univ Melb
Univ Amsterdam
Univ Zurich
External Research Collaborators
Pharmacy
Arts
Engineering
Information Technology
Business and Economics
Medicine
Science
Monash Research Collaborators
Nimrod/G
Nimrod/K
Nimrod/E
Nimrod/O
Guard
Eclipse/PTP
MeSsAGE Lab
Axceleon
Guardsoft
Cray
Leica
HP
IBM
Microsoft
External Industry Collaborators

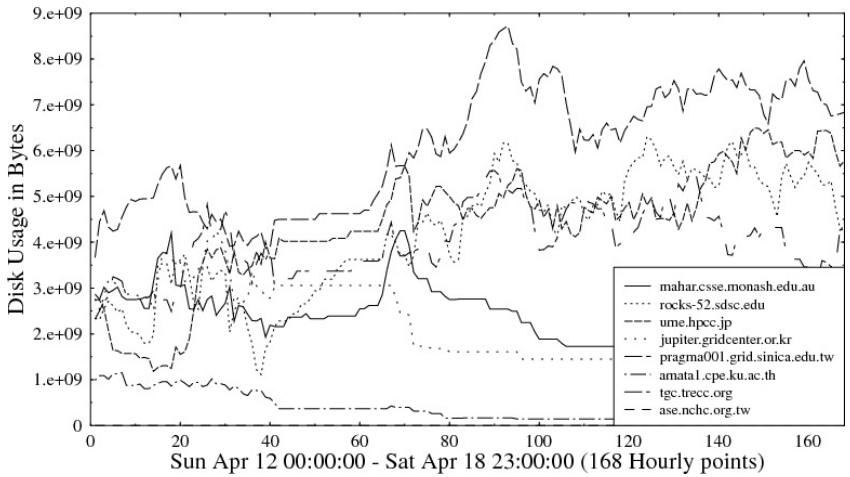
# PRAGMA

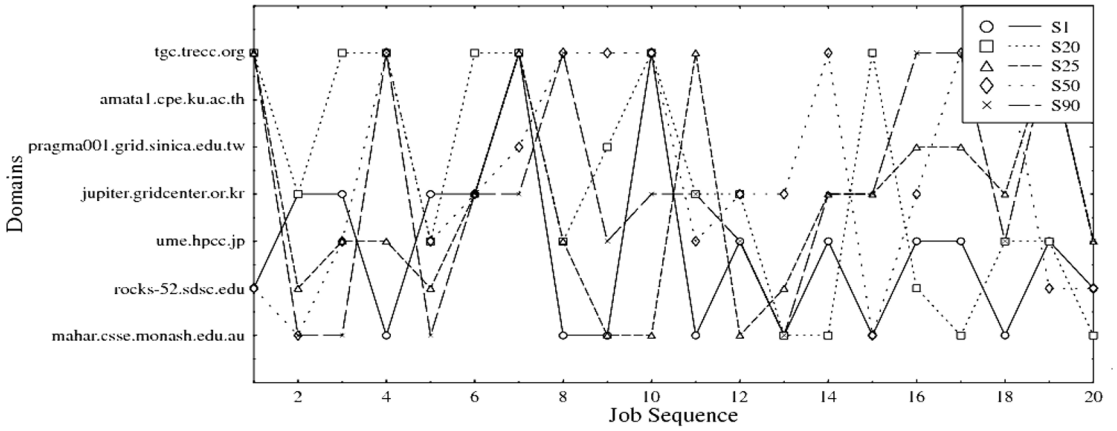
- Strengthen Existing and Establish New Collaborations
- Work with Science Teams to Advance Grid Technologies and Improve the Underlying Infrastructure
- In the Pacific Rim and Globally



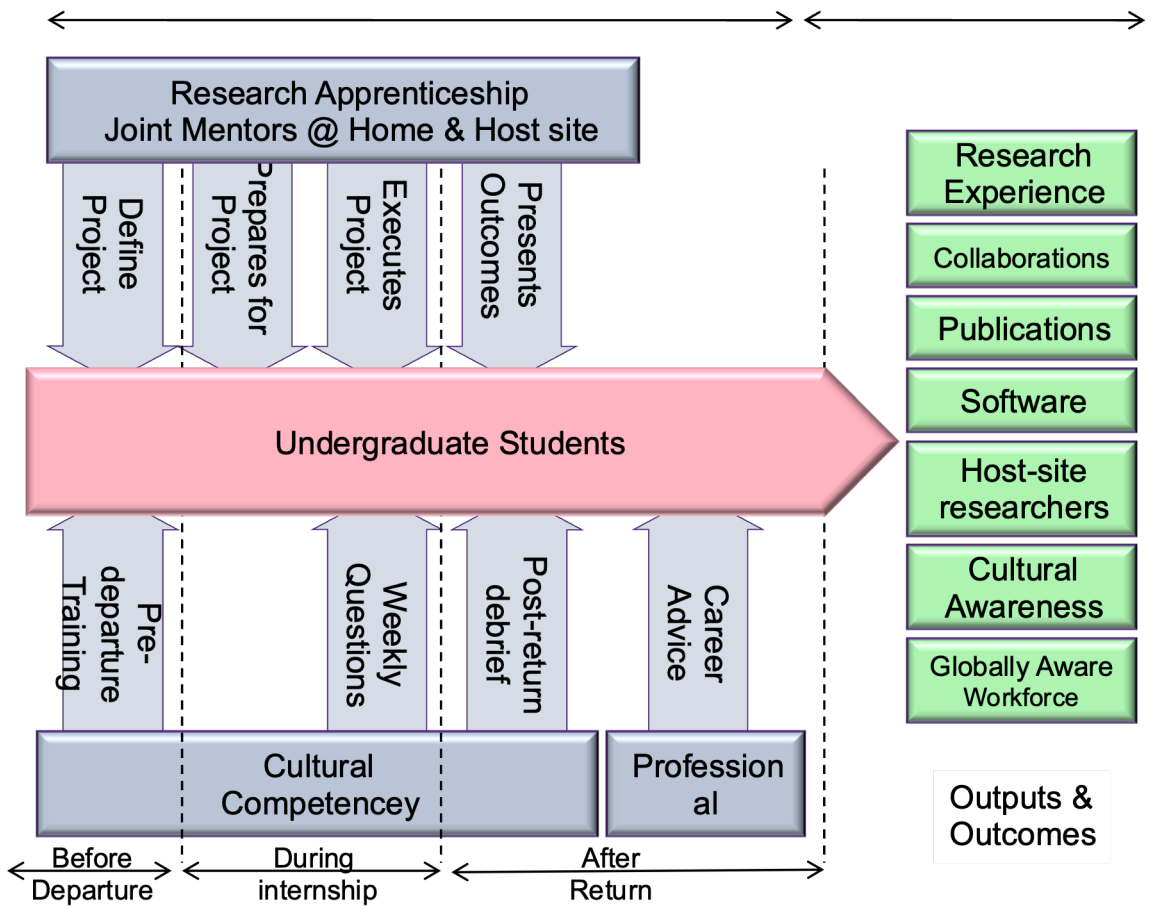
# Leveraging the PRAGMA testbed: Technical

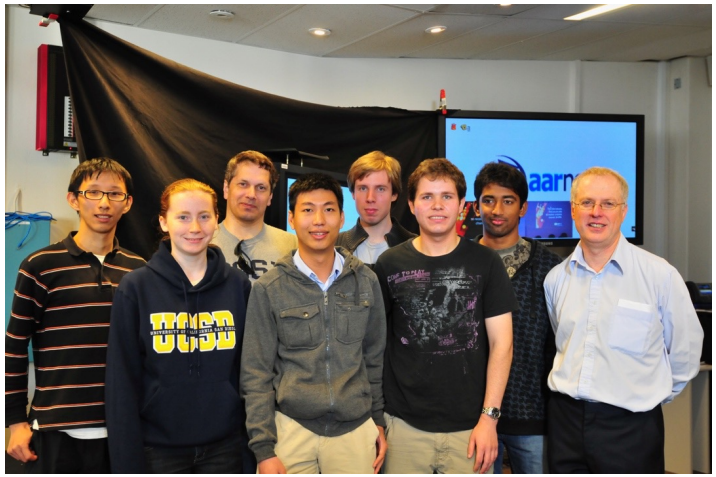


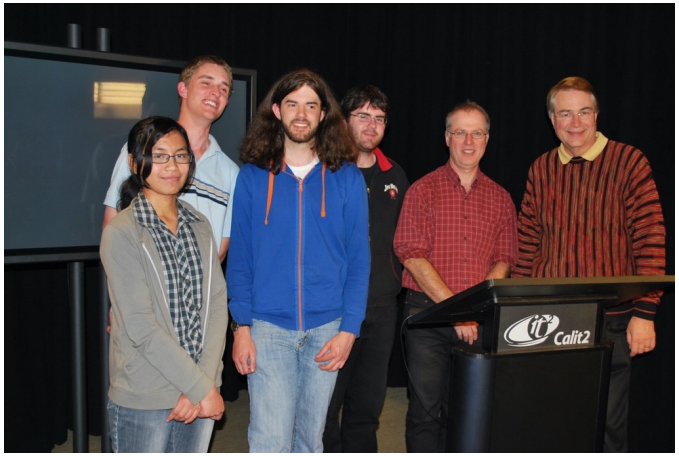



# Leveraging the PRAGMA testbed: Social

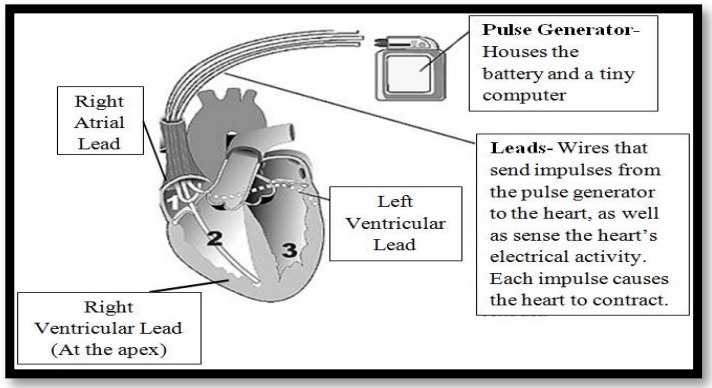


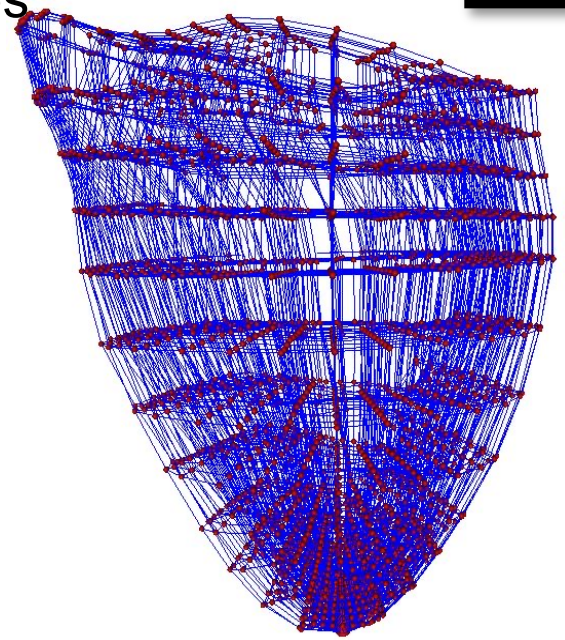





# A typical PRIME project
# Lead placement for CRT
## Revelli, Kirchoffs, McCulloch UCSD

- Study left bundle branch block
- Determine the optimal pacing sites for CRT
- 2 or 3 leads are inserted into the heart
- Challenges:
  - Placing the leads
  - Scar tissue effects pacing
- Continuity and Nimrod to explore different pacing sites



# MURPA/QURPA annual seminar stream

| | |
|---|---|
| Nancy Wilkins-Diehr | Science Gateways and their tremendous potential for Science a |
| Jurgen Schulze | Latest Developments in Virtual Reality at Calit2 |
| Ania Sher | Mathematical models of cardiac muscle cells: Predicting drug-ir |
| Robert Konecny | Multiscale Modeling of Proteins |
| Phil Papadopolous | Extending Rocks Clusters into Amazon EC2 Using Condor |
| Wilfred Li | Workflows for Computer Aided Drug Discovery: New Twitter for |
| Philip Bourne | New Modes of Scholarly Communication |
| Mike Norman | Gordon: A New Kind of Supercomputer for Data-Intensive Appli |
| Larry Smarr | Global Climatic Disruption and its Impact on Victoria and California |
| Sameer Tilak | Cyberinfrastructure for Large-Scale Environmental Observing Systems |
| Bill Gropp | Enabling the Next Generation of Scalable Systems |
| Dr. Edee Wiziecki | Using Virtual Machines to accommodate Computational Chemistry in courses and classrooms |
| Alan Craig | Augmented Reality |
| Dr. Robert A. Fiedler | Applications on the Blue Waters Sustained Petascale System |
| Dr. Radha Nandkumar | Global connections with Child Health Informatics |
| Dr. Brett Bode | Software Development in Petascale Computing |
| Dr. Steven Gottlieb | Lattice QCD: Challenges of Scaling to Peta and Exaflop Speeds |

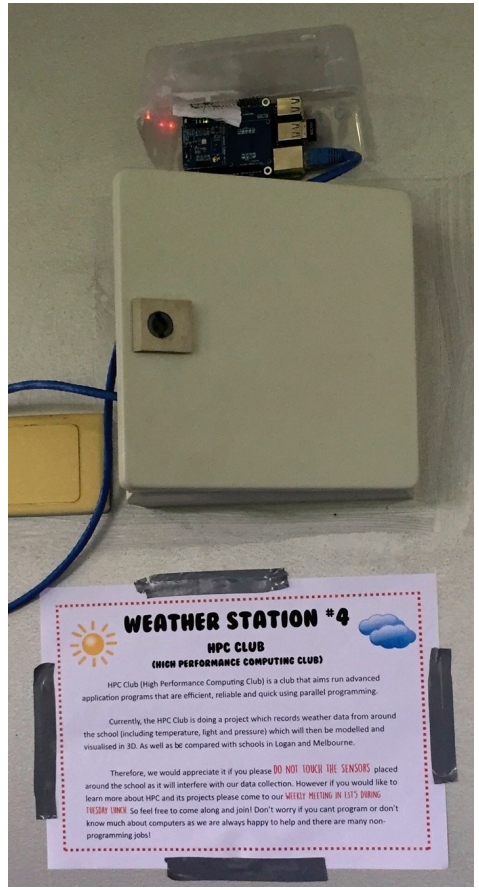
# Kids@SC

U.S. DEPARTMENT OF
ENERGY

SAN DIEGO SUPERCOMPUTER CENTER at the UNIVERSITY OF CALIFORNIA SAN DIEGO
GATEWAY TO DATA-DRIVEN DISCOVERY
SDSC
SAN DIEGO SUPERCOMPUTER CENTER

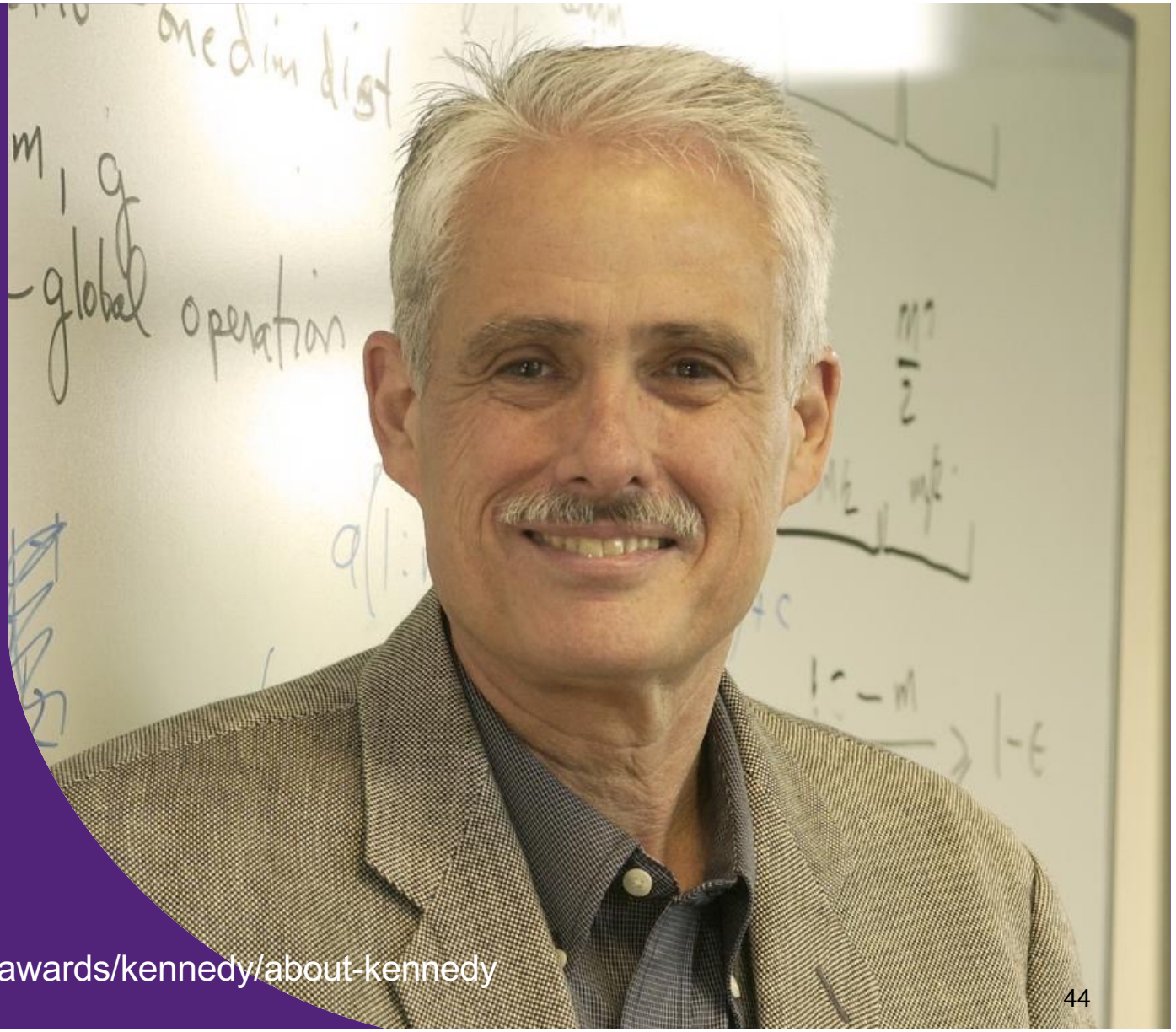
# Kids@SC

SC2011
2012
SC2013


WEATHER STATION #4
HPC CLUB
(HIGH PERFORMANCE COMPUTING CLUB)
HPC Club (High Performance Computing Club) is a club that aims run advanced application programs that are efficient, reliable and quick using parallel programming.
Currently, the HPC Club is doing a project which records weather data from around the school (including temperature, light and pressure) which will then be modelled and visualised in 3D. As well as be compared with schools in Logan and Melbourne.
Therefore, we would appreciate it if you please DO NOT TOUCH THE SENSORS placed around the school as it will interfere with our data collection. However if you would like to learn more about HPC and its projects please come to our WEEKLY MEETING IN E5T5 DURING THEORY LUNCH. So feel free to come along and join! Don't worry if you cant program or don't know much about computers as we are always happy to help and there are many non-programming jobs!

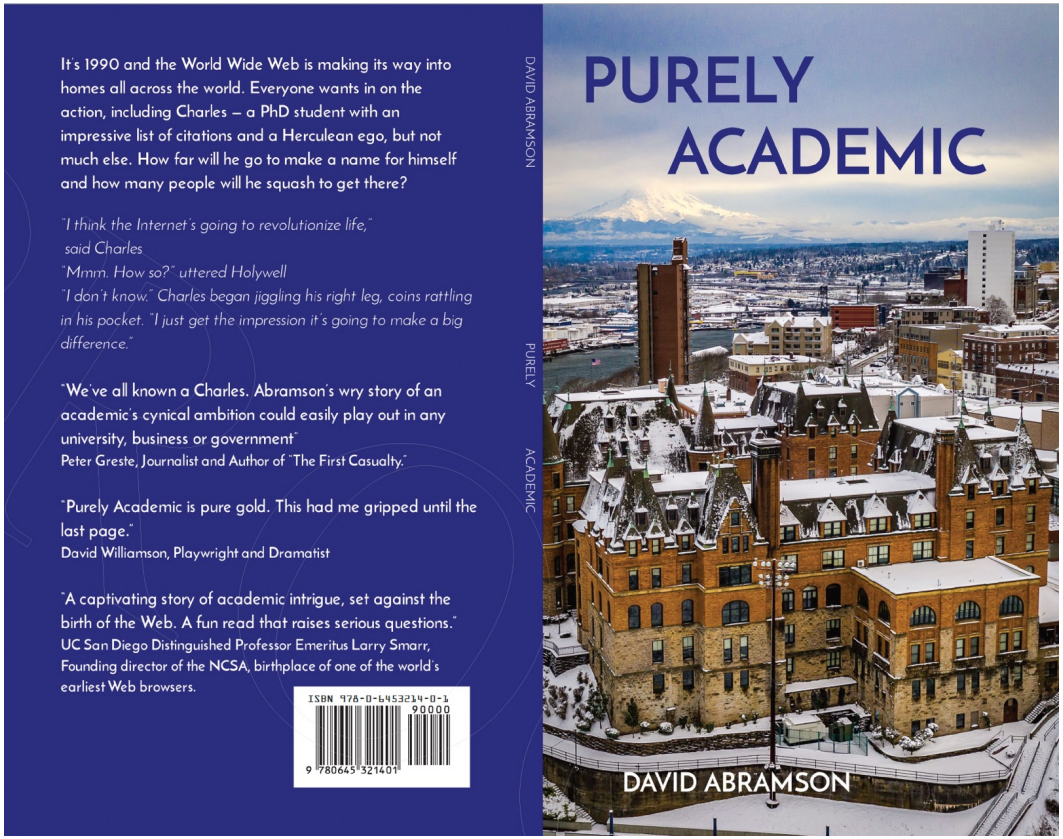
# In memory of Ken Kennedy

… the theoretical impact of this work was matched by a tremendous practical impact, as Kennedy's work has had a profound influence in the commercialization of automatic vectorizing and parallelizing compilers.

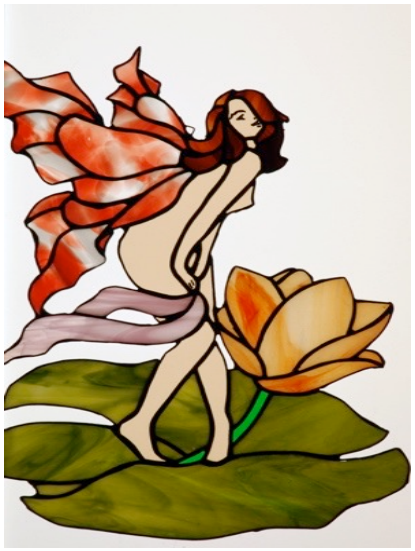



https://www.computer.org/volunteering/awards/kennedy/about-kennedy

# Translation to the arts




Coming soon!

purely-academic.com

# Thank you and Questions