

### David Abramson

Director, Research Computing Centre

Professor of Computer Science University of Queensland

#### **About David**

#### PURELY ACADEMIC

Cast

Prof John Holywell Southern University academic in

his early 50's

Prof Martin Godson Middleton University academic

in his mid 50's

Prof Mary Long Southern University academic in

her early 40's

Charles Mittleman Initially a 30 year old PhD

student at Wooton College and Southern University, but then moves to Middleton University

as a young academic.

Joanne Southern University software

developer in her mid 30's. Works in Prof Holywell's lab

and is pregnant.

Prof Max Williams St George College academic in

his late 50's. Serves as the chair of the Shaw Trust, a not

for profit society that

supports research projects with

grants.

Anna Middleton University

administrator in her 30's. She also serves as an administrator on the Shaw Trust, taking notes and helping with the grant

and neiping with the gr assessment exercises.

Mark Early career academic

Robin, Cheryle, Newsreader (voices only. Can be played by other actors)

- - -

This play starts in mid to late 1990's.

david.abramson@uq.edu.au



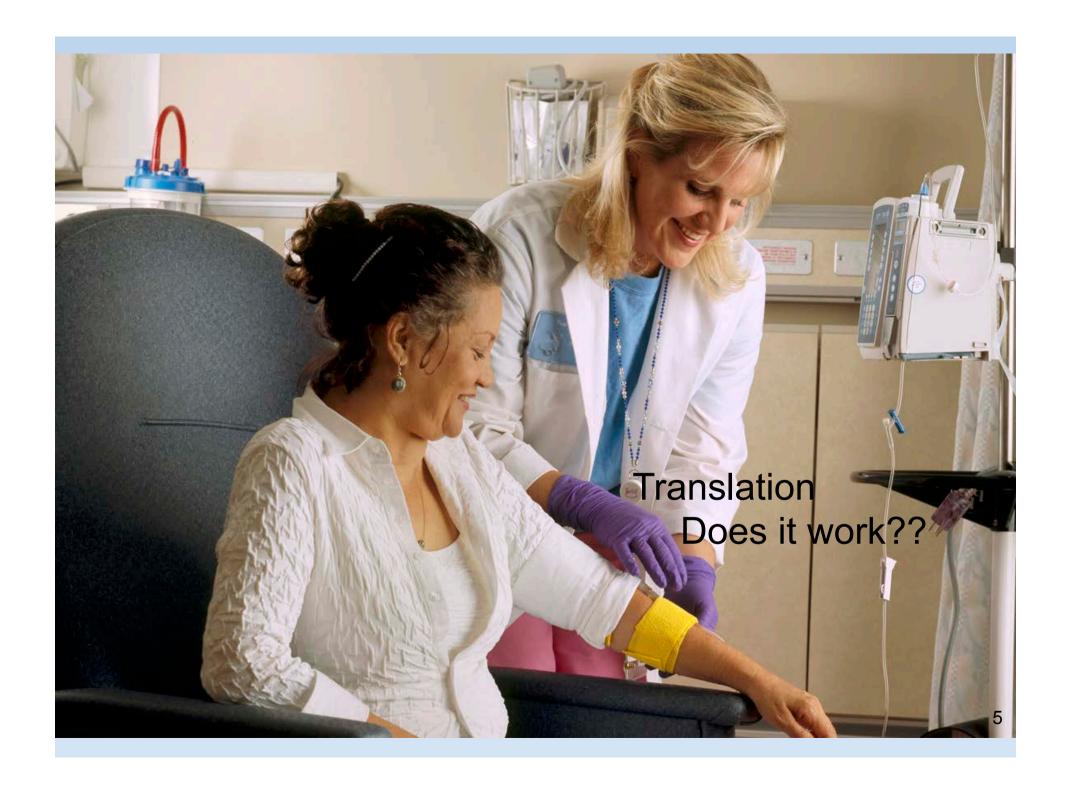




#### Introduction

- Research Methods
  - Translational Research in Medicine
  - Traditional Computer Science Research
  - Translational Computer Science
- Framework
  - Applications that matter
  - Structures that scale
  - New Computer Science
  - Novel Platforms

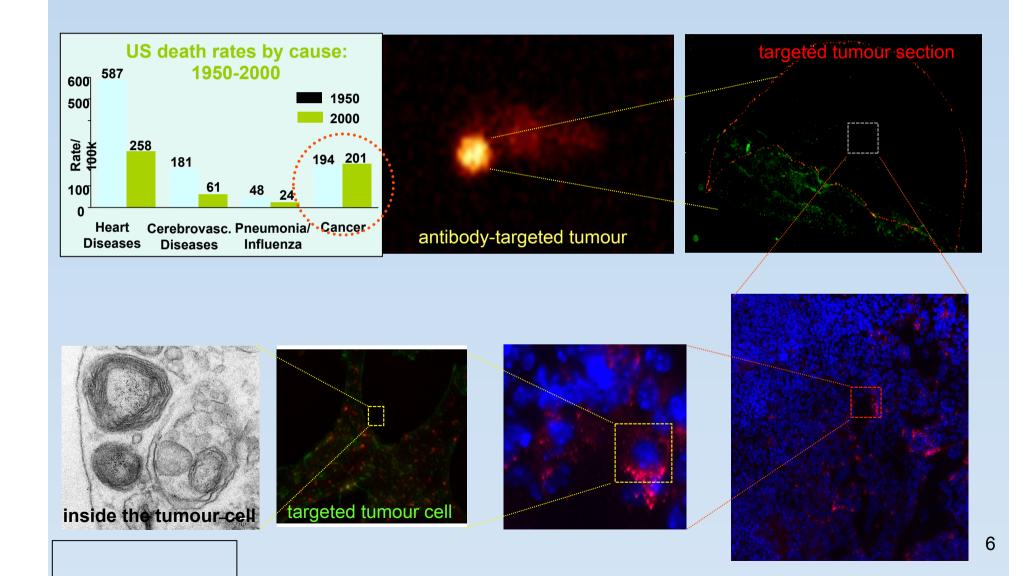
## **Translational Research in Medicine**



#### Antibody-based anti-cancer therapeutics development

Imaging strategies 'Zooming into the molecular world of tumour cells?'

Lackmann et al



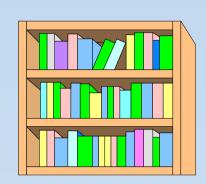
## Traditional Computer Science research

#### Traditional IT Research

- Can range from
  - theory to practice
  - "soft" to "hard"
- Theory
  - Might be some fundamental advance in theory of computation
  - New algorithms (ways of solving problems)
  - Leads to better "science"
- Practice
  - New solutions (i.e. practical algorithms)
  - Applying new technologies
  - Leads to better "engineering"
- IT tends to be multi-disciplinary
  - E.g. e-science, bio-informatics
- Translational Research
  - New Computer Science
  - Applied to real world

#### Research Outcomes

- Theories and experimental results
  - Papers published in reviewed journals, conferences
- Patents
  - Published by the patents office
- Prototypes
  - Computer programs
  - Hardware
- Commercial products
- Translation?
  - May not be commercialised

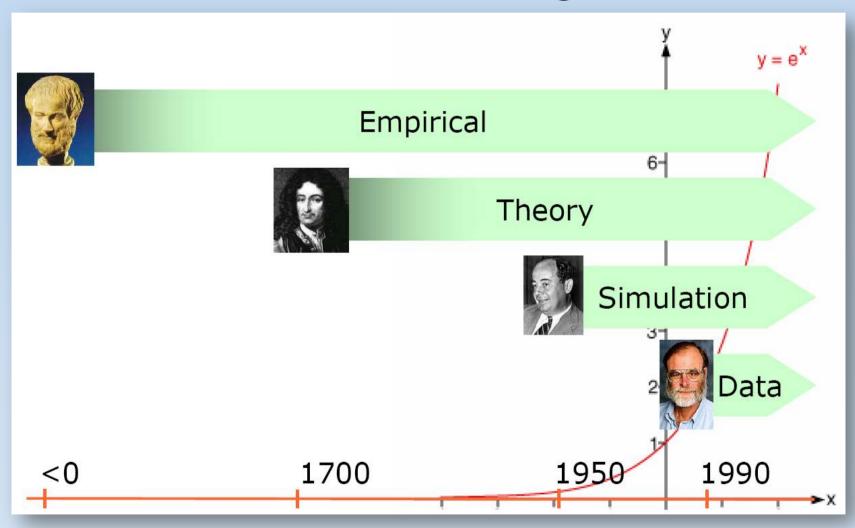






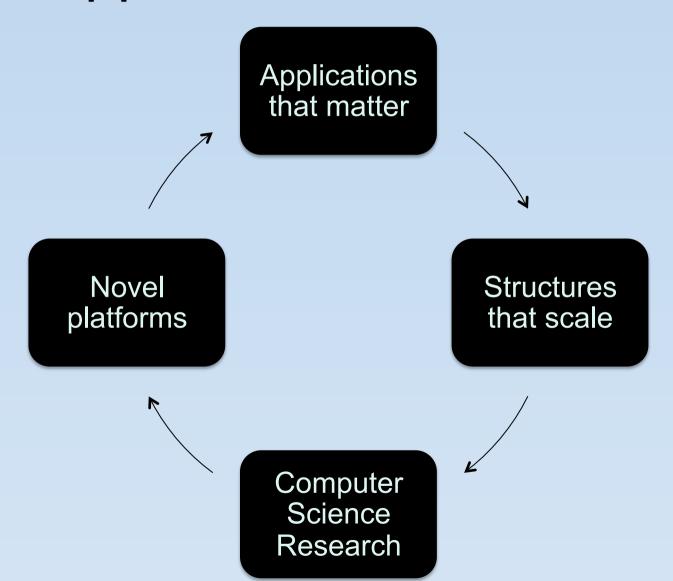


# Potential for translation has never been higher



10

### **Applications Driven CS**



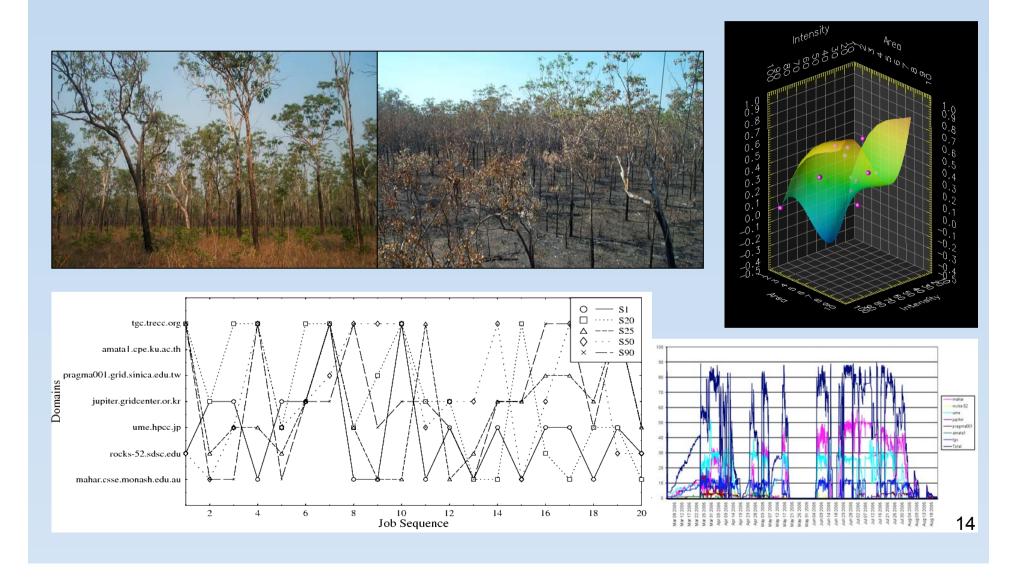
### **Applications that matter**

# Computer Science is increasingly important

- Systems Biology and Health
  - Human Genome Project, Protein function, Virtual Physiological Human, Blue Brain, ...
- Engineering
  - Aerospace, civilian, automotive, domestic, ...
- Environment
  - Climate, weather, pollution, ....
- Chemistry
  - Drug design, novel pathways, ...
- Physics
  - Particle Physics, Xray treatment, Astrophysics, ....
- Business
  - Manpower planning, Logistics, Resource allocation

#### **Environmental Science**

Lynch, Beringer, Uotila Monash U, AU



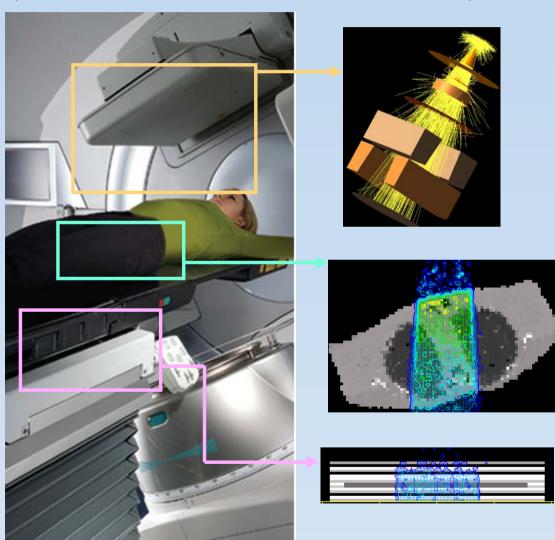
## Radiotherapy planning

Giddy, Chin, Lewis, Welsh e-Science Centre, UK

**BEAMnrc** 

**EGS** 

**DOSXYZnrc** 



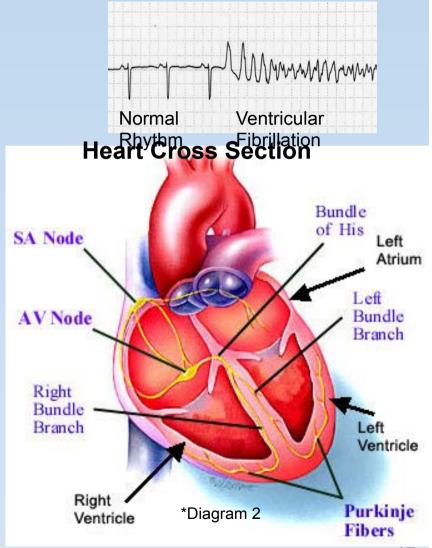
#### Cardiac Science

Sher, Gavaghan, Rodriguez, Oxford Mcculloch, Mihaylova, Kerckhoffs, UCSD

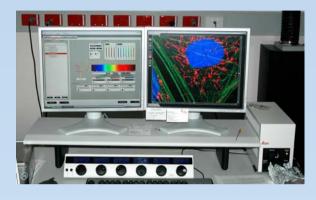
- Heart disease still leading cause of death
- Understanding the underlying physiological mechanisms is cheaper and faster when experimental studies are performed together with mathematical models & computer simulations
- Studying pathologies
- Developing & Testing drugs

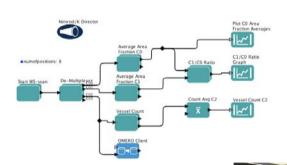
## Why study heart failure?

- Leading Causes of HF Include:
  - High blood pressure
  - Myocardial Infarctions
  - Mitral valve disorders
  - Conduction disorders
    - Revelli studied
      - Left Bundle Branch Block
      - Effects 25% of all HF patients
      - Increases risk of sudden death considerably
    - Dan Stokeley studied
      - Long QT



## Cancer Imaging and Therapy

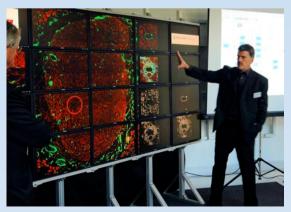




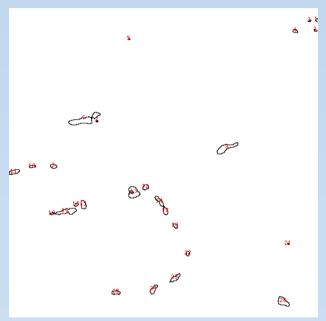


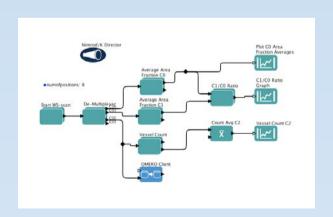






## Cancer Imaging and Therapy





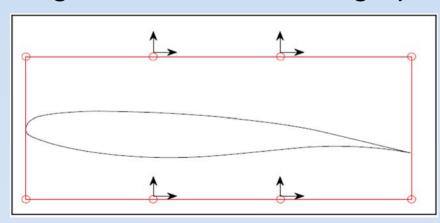


## Aerodynamic Design

#### Kipouros, Cambridge, UK

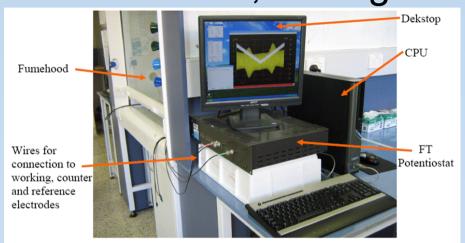
- Geometry management using Free Form Deformation 8 design variables
- Evaluation of the aerodynamic characteristics, Cl, Cd, and Cm coefficients using Xfoil
- Investigation of the lift to drag trade-off subject to hard geometrical constraints to the thickness of the airfoil at 25% and 50% of the chord (in order to maintain practical

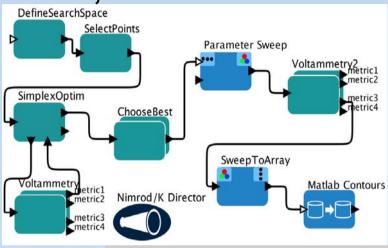
significance to the design problem)



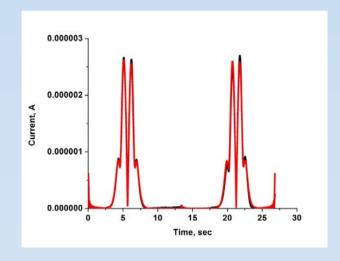
## Electro-chemistry

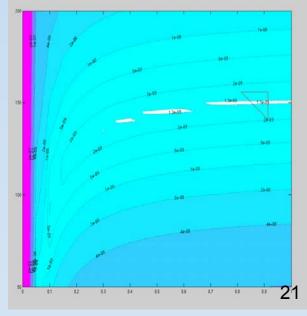
Bond, Gavaghan: Monash, Oxford





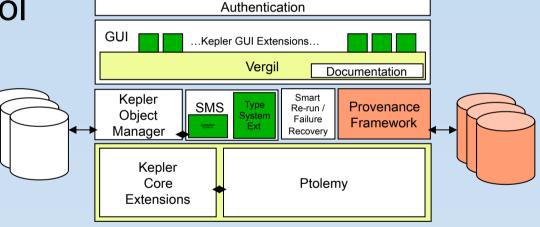


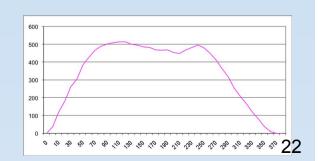




## But are there computer science outcomes?

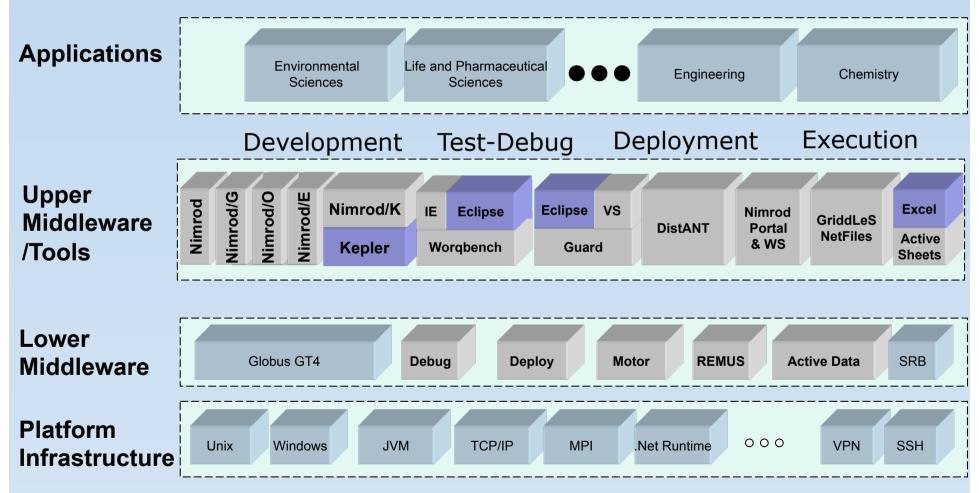
- Workflows
  - Complex experiments
  - Instrument control
  - Optimization
- Parallelism
- Grid and Cloud infrastructure
- Science gateways
- New optimization algorithms
- Scalable Parallel Debugging





#### **Structures that scale**

### MeSsAGE Lab: A living Lab



#### **Applications**

#### Core Technology

**UCSD** PRAGMA/PRIME/MURPA

Oxford

Pharmacy

Nimrod/G

Guard

Arts

Nimrod/K

Nimrod/E

Nimrod/O

Eclipse/PTP

DPI

Engineering

**Univ Melb** 

Univ

Amsterdam

Information Technology

Business and **Economics** 

Medicine

**Univ Zurich** 

Science

MeSsAGE Lab

Axceleon

Guardsoft

Cray

Leica

HP

**IBM** 

Microsoft

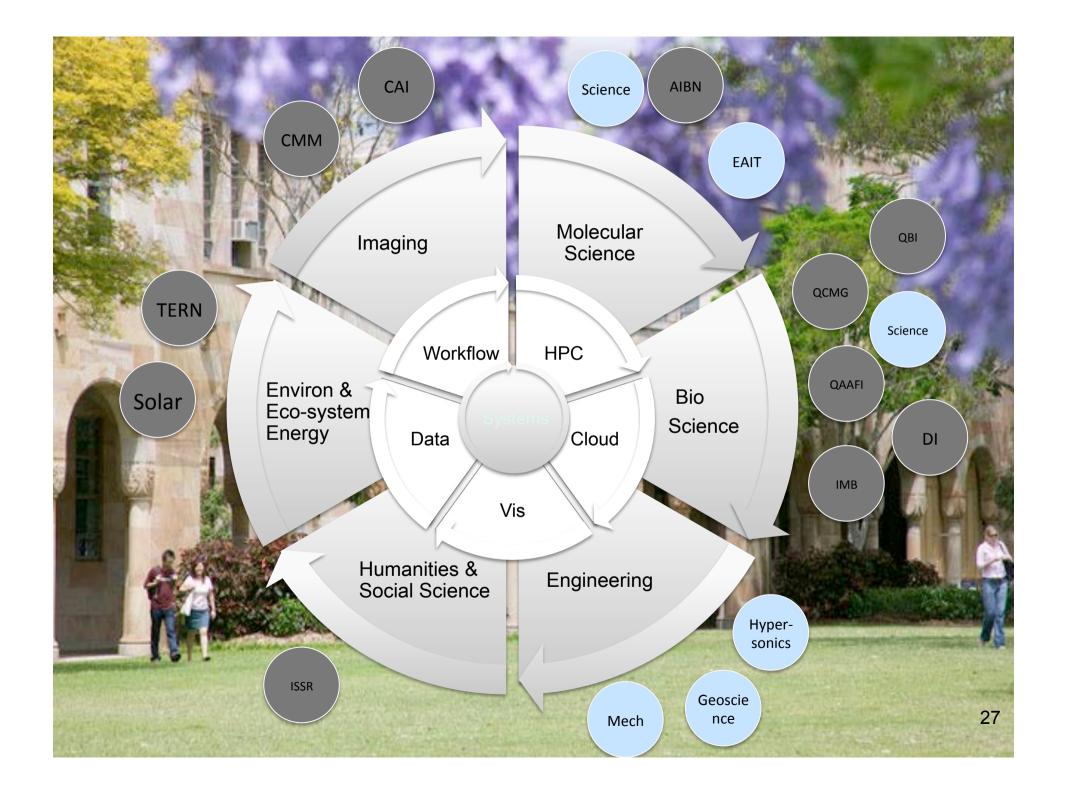
External Industry Collaborators 25

Research Collaborators

External

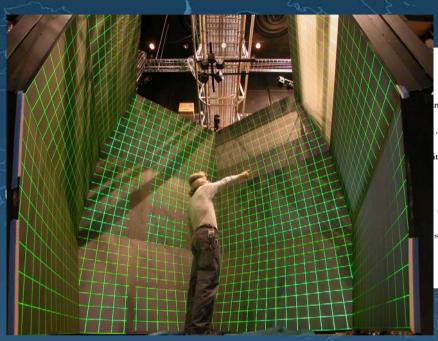
Monash Research Collaborators



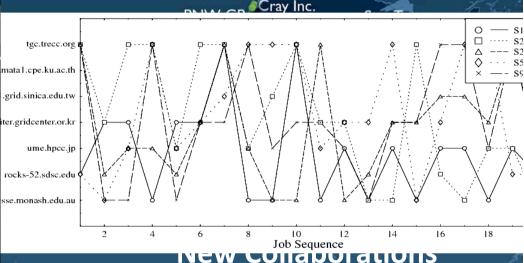


#### **PRAGMA**

**A Practical Collaborative Framework** 



http:/



Privation

O Animals

Shed

Shed

Shed

She Firmed

She Firmed

This Step

O Charling

O Charling

Animals

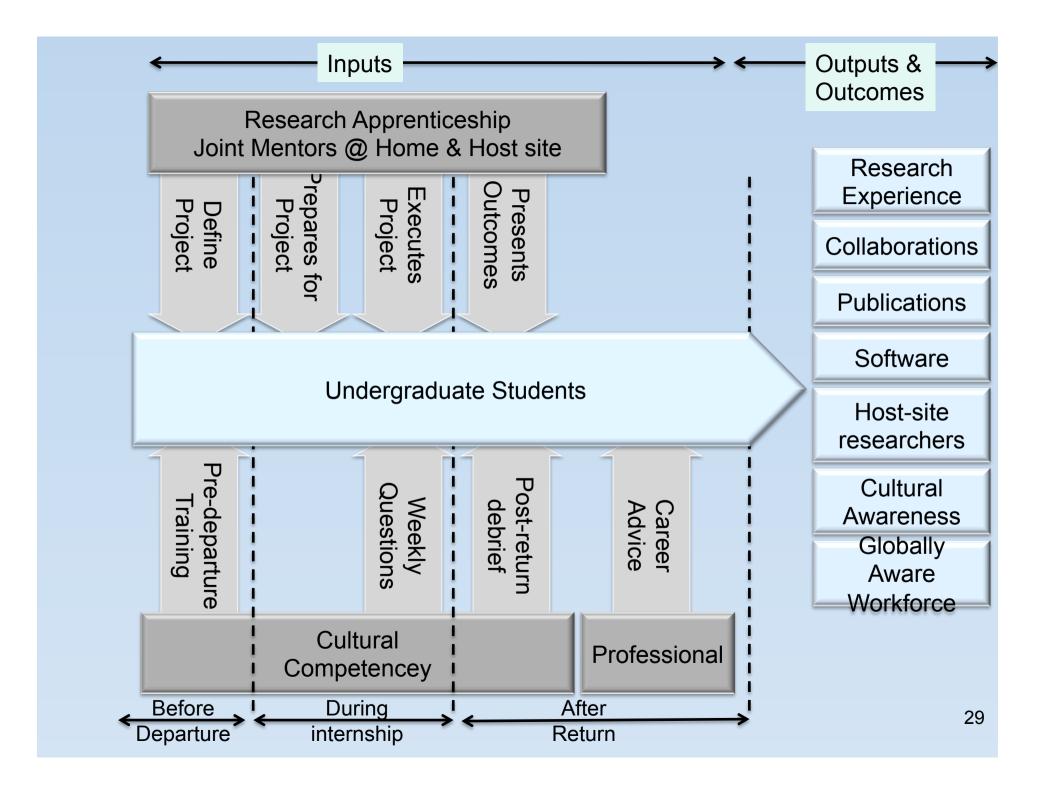
O Symphoritie

O S

Work with Science Teams to Advance Grid Technologies and Improve the Underlying Infrastructure

BeSTGRID

In the Pacific Rim and Globally



# MURPA & QURPA

- Engaged in PRIME since 2004
- Projects range from bio-engineering, theoretical chemistry to computer science
- Has underpinned long lasting academic collaborations
  - Publications
  - Presentations at conferences
- Undergraduate students without research experience!



#### Leveraging Novel Platforms



I've participated in numerous video conferences to date but nothing like this. The quality was so high that the experience was almost as if we were all in the same room.

The massively increased bandwidth was transformational. Quantity begat quality.

Alan Finkel, Chancellor, Monash Univ

### MURPA & QURPA Lectures

Nancy Wilkins-Diehr	Science Gateways and their tremendous potential for Science and Engineering
Jurgen Schulze	Latest Developments in Virtual Reality at Calit2
Ania Sher	Mathematical models of cardiac muscle cells: Predicting drug-induced arrhythmias.
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 the Old Problem
Philip Bourne	New Modes of Scholarly Communication
Mike Norman	Gordon: A New Kind of Supercomputer for Data-Intensive Applications
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



This is NOT about development

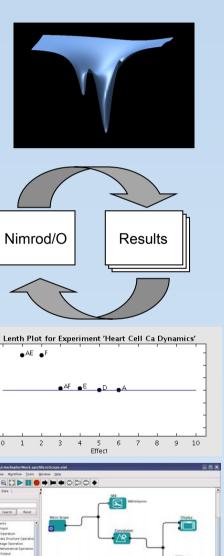
### Computer Science Research

Distributed computing on the Grid

#### **Nimrod**

#### Nimrod supporting "real" science

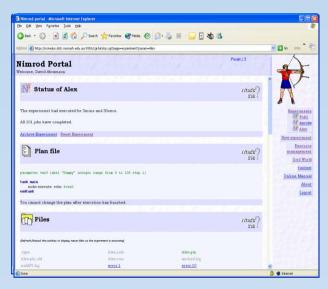
- A full parameter sweep is the cross product of all the parameters (Nimrod/G)
- An optimization run minimizes some output metric and returns parameter combinations that do this (Nimrod/O)
- Design of experiments limits number of combinations (Nimrod/E)
- Workflows (Nimrod/K)



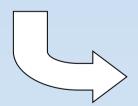


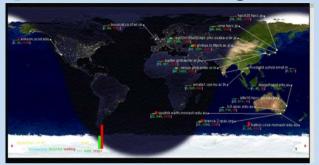
Nimrod Portal Plan File Nimrod/O Nimrod/E Nimrod Portal Parameter Section ▼ from 200 to 300 Points v 2 Nimrod/G range pressure label ▼ from 5000 to 6000 Points - 4 concent label ▼ from 0.002 to 0.005 Points **v** 2 **Actuators** material label ▼ select ▼ Remove Add a comment Add a new parameter Tasks Section **Grid Middleware** + task rootstart + task nodestart + task main + task nodefinish + task rootfinish Add a new task Cancel and Reload Text mode

# Nimrod Development Cycle

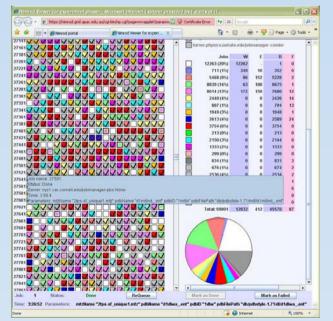


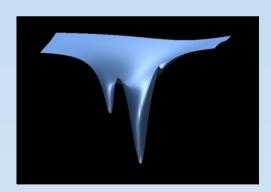
Prepare Jobs using Portal





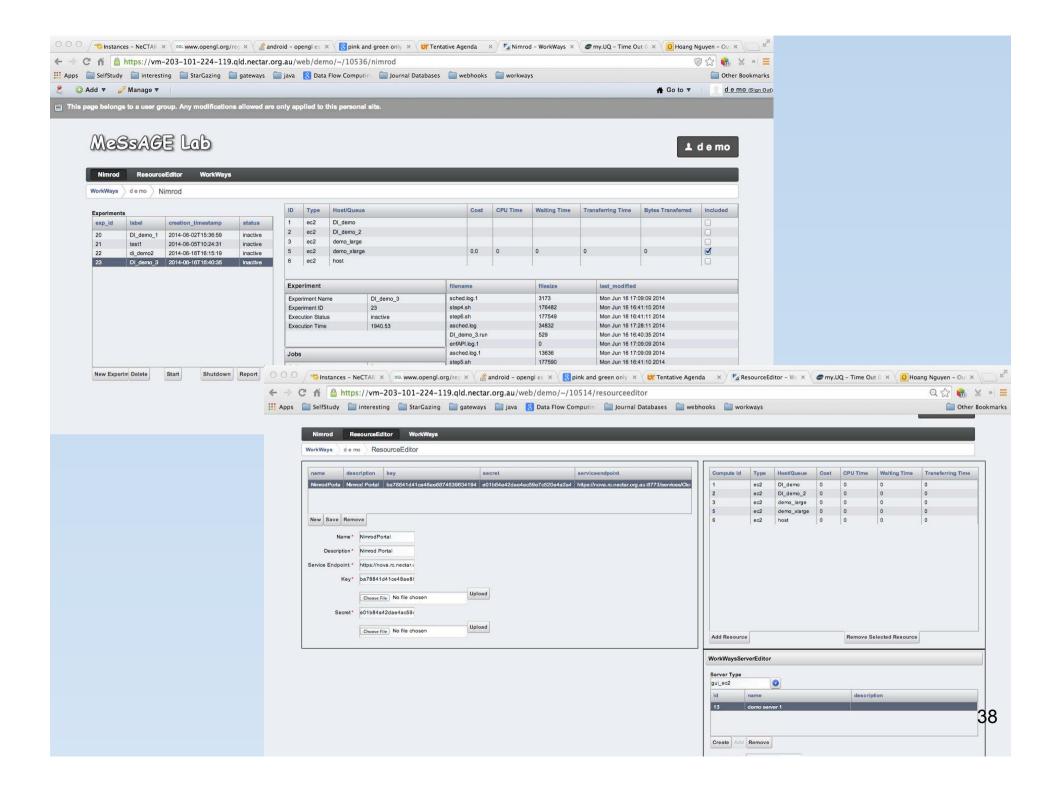
Sent to available machines

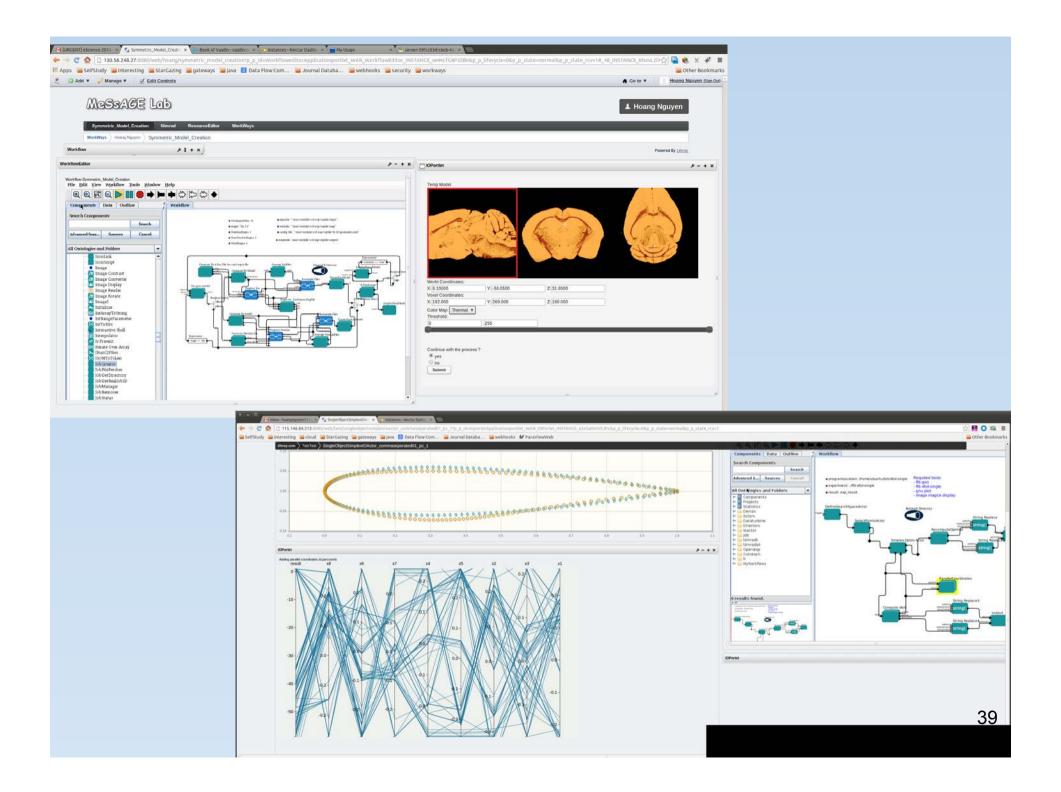


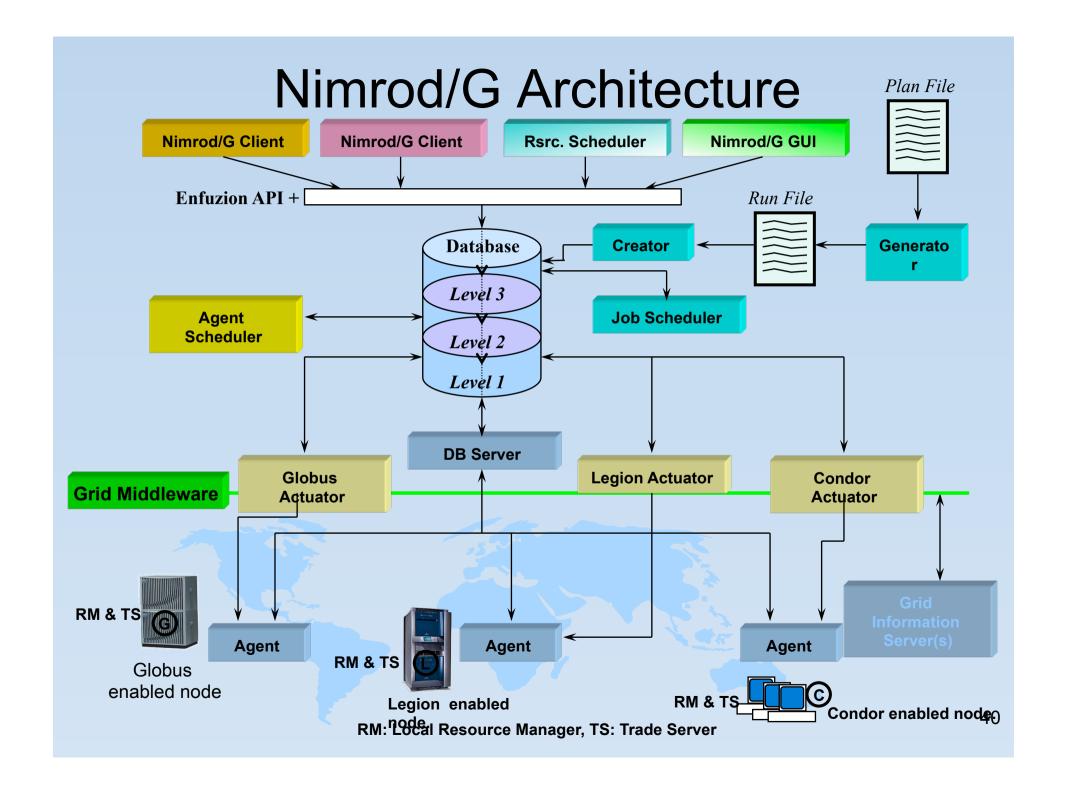


Results displayed & interpreted

37

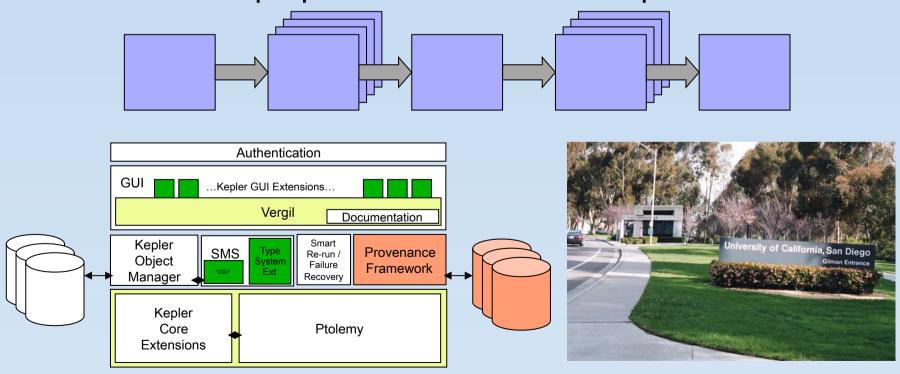




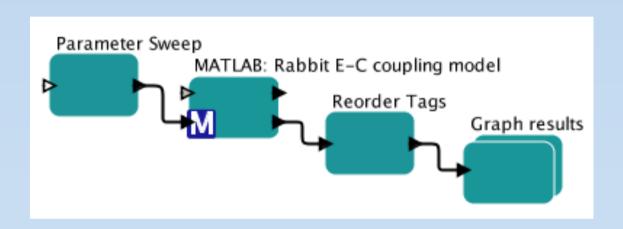


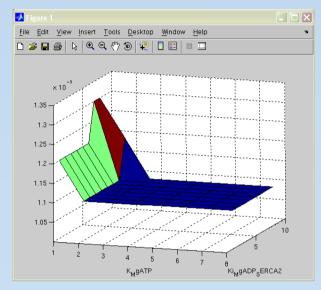
#### Nimrod/K Workflows

- Nimrod/K integrates Kepler with
  - Massively parallel execution mechanism
  - Special purpose function of Nimrod/G/O/E
  - General purpose workflows from Kepler

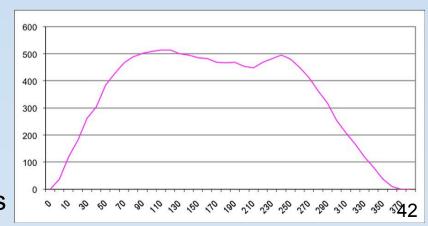


## Nimrod/K Workflows

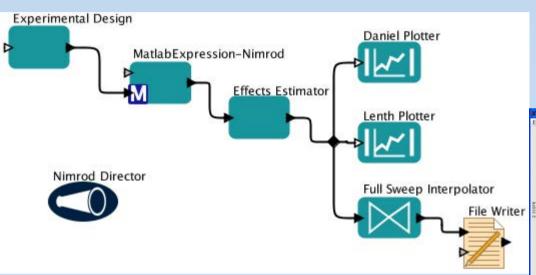


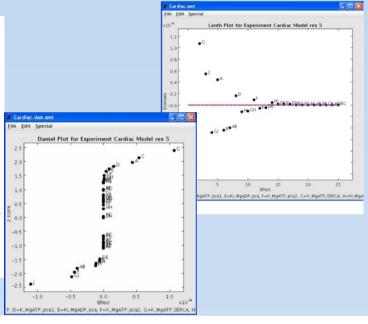


- Using a MATLAB actor provided by Kepler
- Local spawn
  - Multiple thread ran concurrently on a computer with 8 cores (2 x quads)
  - Workflow execution was just under 8 times faster
- Remote Spawn
  - 100's 1000's of remote processes

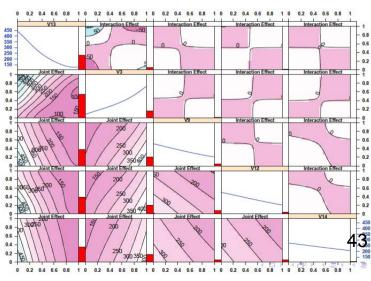


## Nimrod/EK Actors

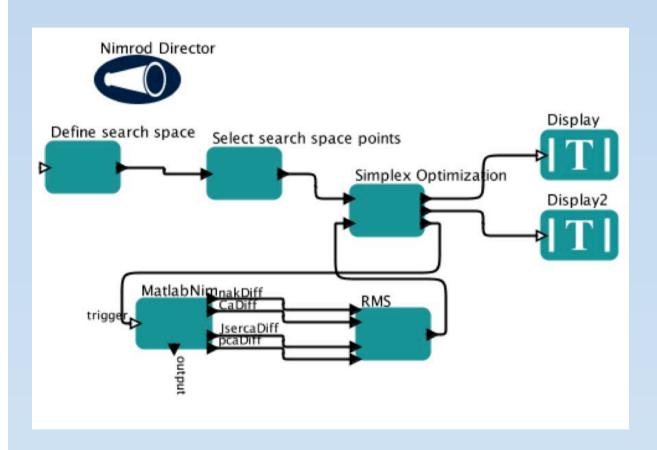




- Actors for generating and analyzing designs
- Leverage concurrent infrastructure



### Nimrod/OK Workflows



- Nimrod/K supports parallel execution
- General template for search
  - Built from key components
- Can mix and match optimization algorithms

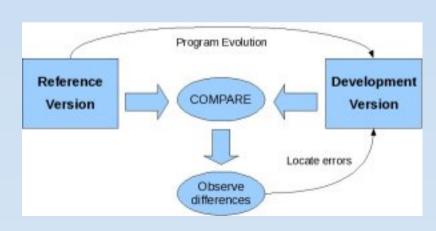


New ways to debug code

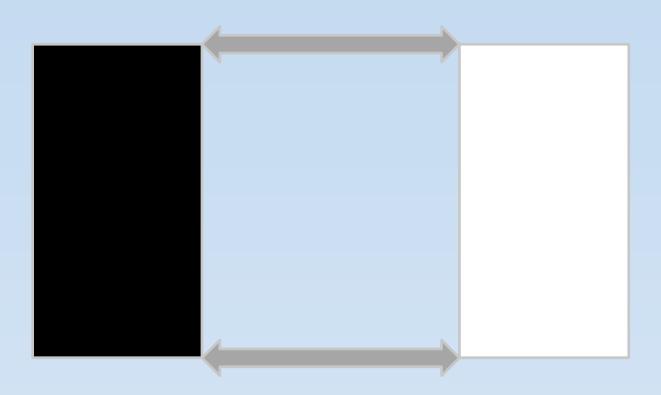
## Comparative debugging

# Comparative Debugging

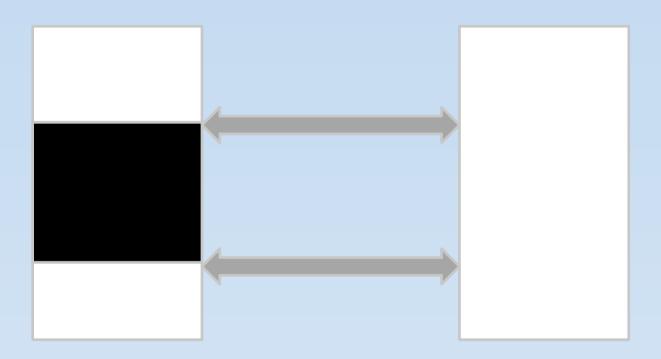
- What is comparative debugging?
  - Data centric approach
  - Two applications, same data
  - Key idea: The data should match
  - Doesn't actually locate bugs
  - Quickly isolate deviating variables
  - Focus is on where deviations occur
- How does this help me?
  - Algorithm re-writes
  - Language ports
  - Different libraries/compilers
  - New architectures



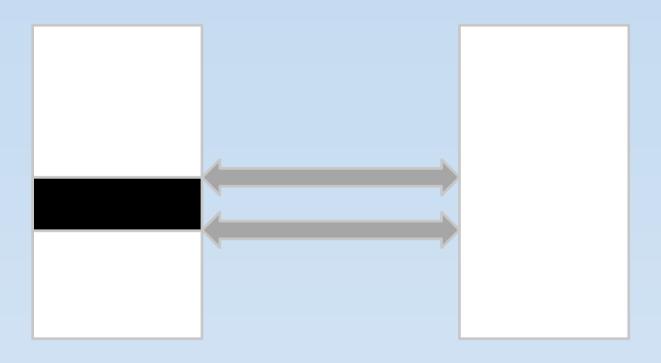
Iterative refinement of problem area



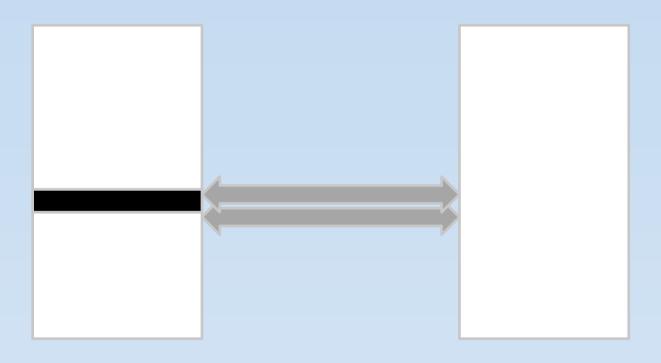
Iterative refinement of problem area



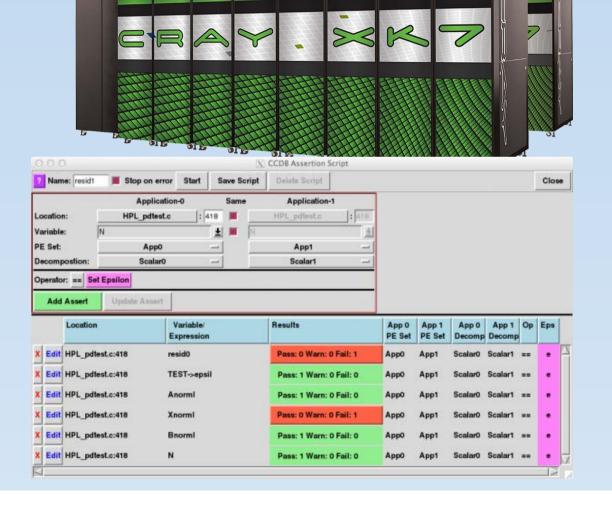
• Iterative refinement of problem area

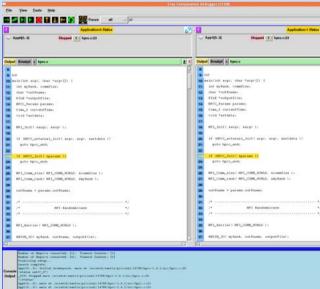


Iterative refinement of problem area



### **CCDB**

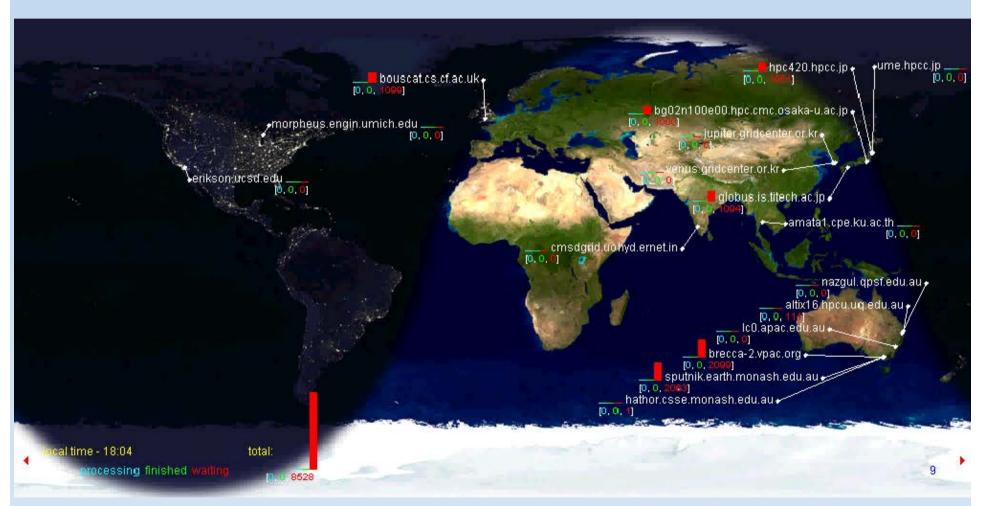




Stretch our thinking and enable innovation

#### **Novel Platforms**

#### PRAGMA testbed: 2004



# Supercomputing requires extreme debugging

#### Titan

- 299,008 AMD Opteron cores
- 18,688 Nvidia Tesla K20 GPU Accelerators
- 710 TB system memory, 32 GB + 6 GB per node (w/accelerator)
- 18,688 compute nodes

#### BlueWaters

- XE partition
  - 362,240 AMD Bulldozer cores (physical, 724480 w/hyperthreads that share fpu)
  - 1.382 PB system memory, 64 GB per node
  - 22,640 compute nodes

#### XK partition

- 33,792 AMD Bulldozer cores (physical, 67584 w/hyperthreads)
- 135 TB system memory, 32 GB per node
- 25 TB accelerator memory, 6 GB per node
- 4,224 compute nodes



# FlashLite A Data Intensive Platform

Parallel

High performance

network

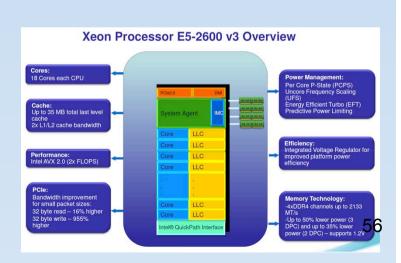
Good numeric performance

Massive memory

- Ability to hold whole data sets or data bases in memory
- High IO throughput

#### What is FlashLite?

- FlashLite
  - ~ 70 compute nodes (~1600 cores)
    - Dual socket Intel E5-2680v3 2.5GHz (Haswell)
    - 512 GB DDR-2
    - 4.8 TB NVMe SSD
  - ScaleMP vSMP virtual shared memory
    - 4TB RAM aggregate(s)



#### Conclusion

- Translation of computer science results to practice is important
  - Not necessarily commercialisation
- Good outcomes in
  - Computer science
  - Host discipline
- Structures need to be designed with translation in mind
- Novel platforms stretch our thinking and enable innovation





