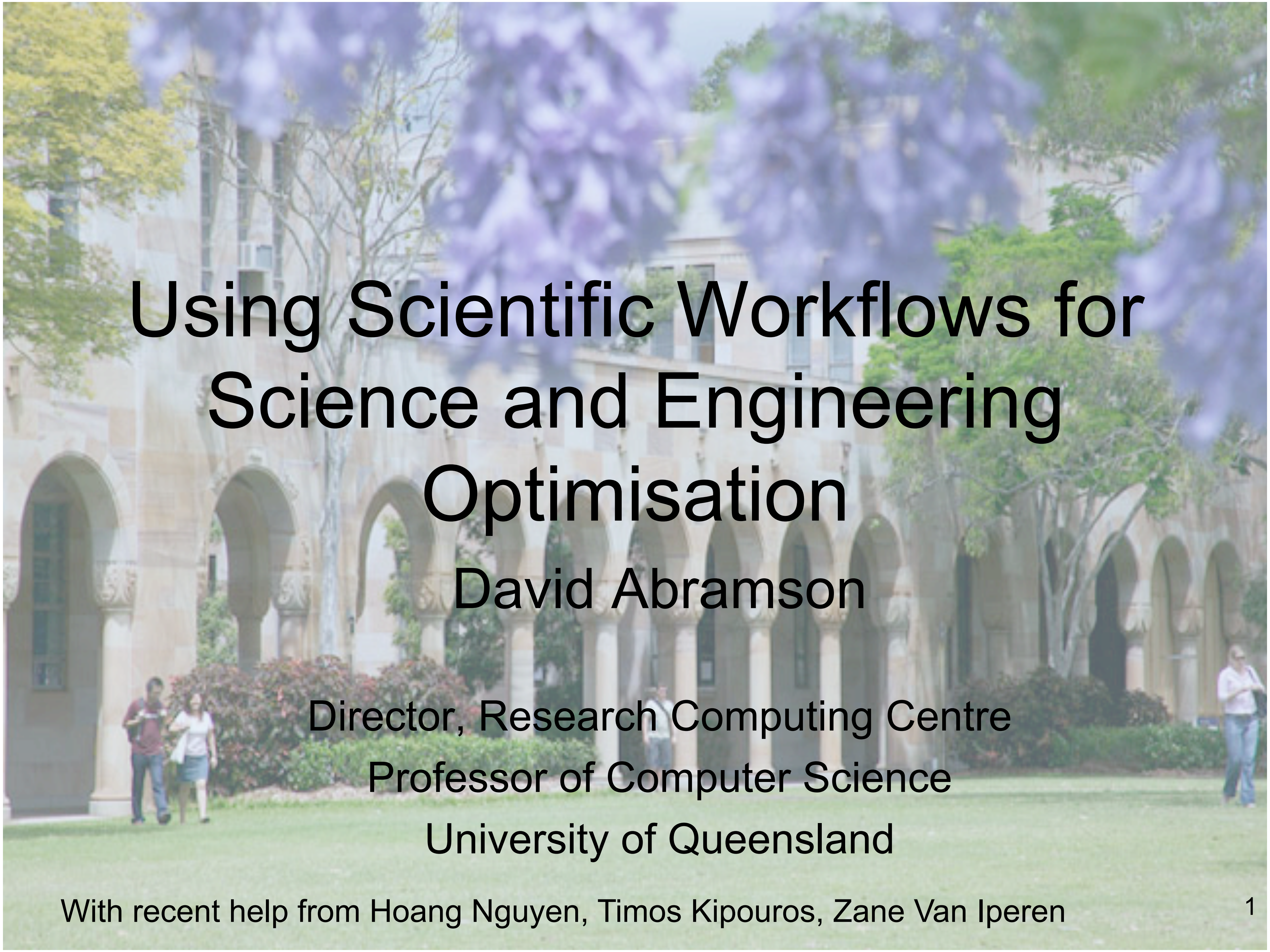
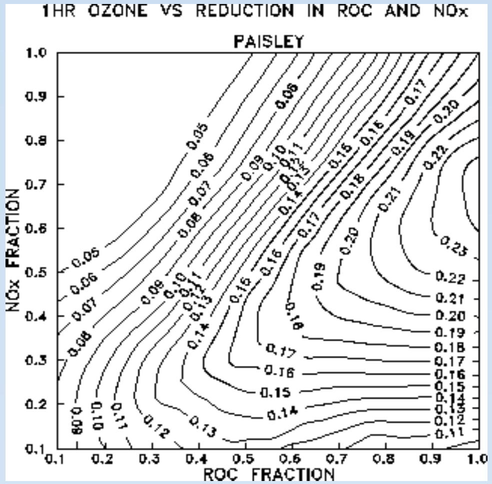

Using Scientific Workflows for Science and Engineering Optimisation
David Abramson
Director, Research Computing Centre
Professor of Computer Science
University of Queensland
With recent help from Hoang Nguyen, Timos Kipouros, Zane Van Iperen
1

Introduction

- Some Engineering Optimization Problems
- The Nimrod tool family
- Scientific Workflows
- Execution Engines
  - Clusters & Grids & Clouds
- Interacting with designs
- Conclusions

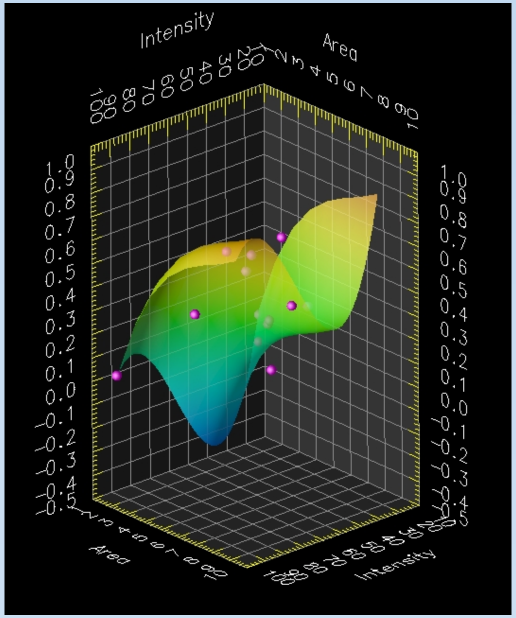
Some Engineering Optimization Problems

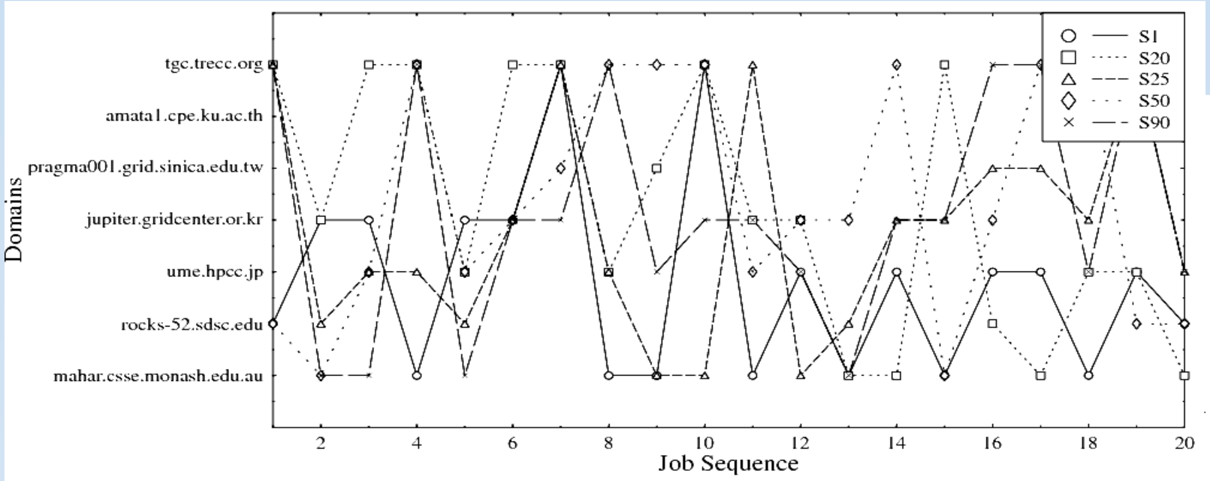
Air pollution
Cope, Victorian EPA (in 1990)

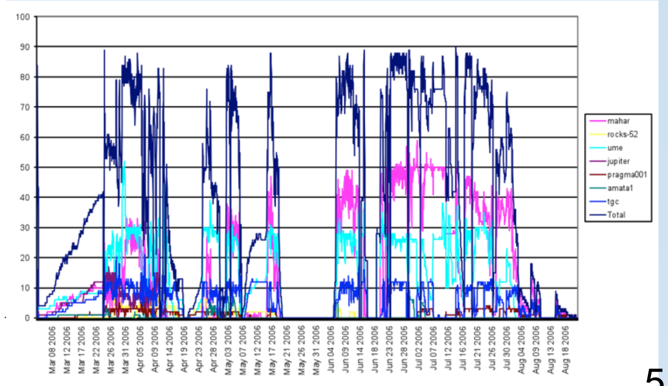


Wildfires

Lynch, Beringer, Uotila Monash U, AU



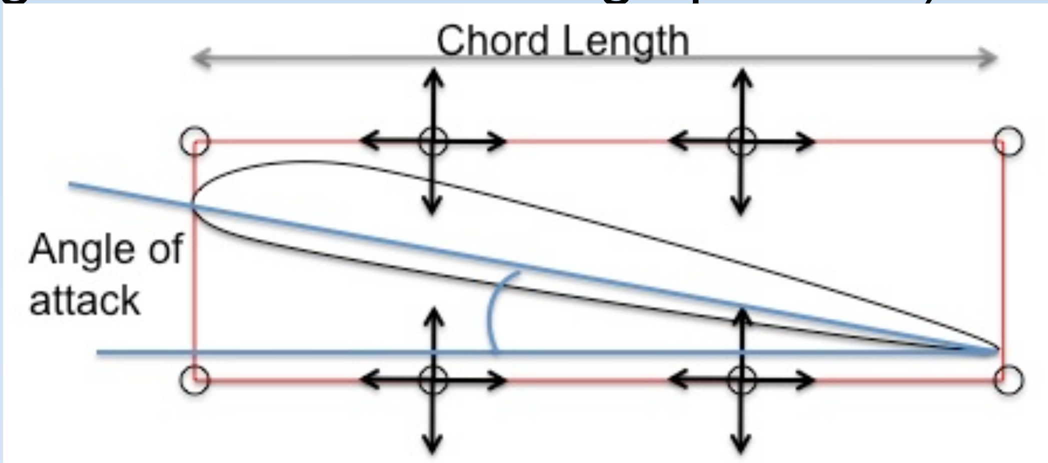


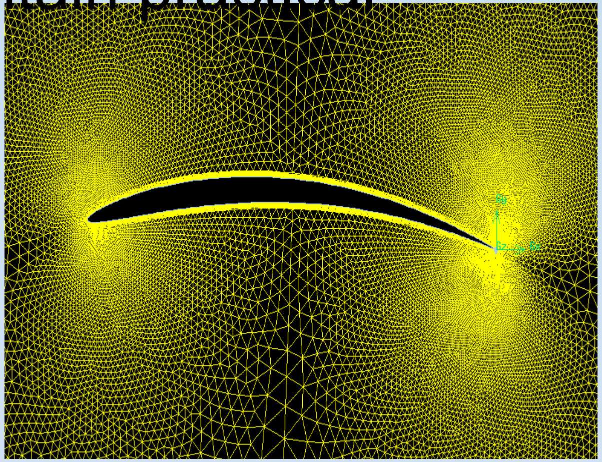


Aerofoil Design

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)



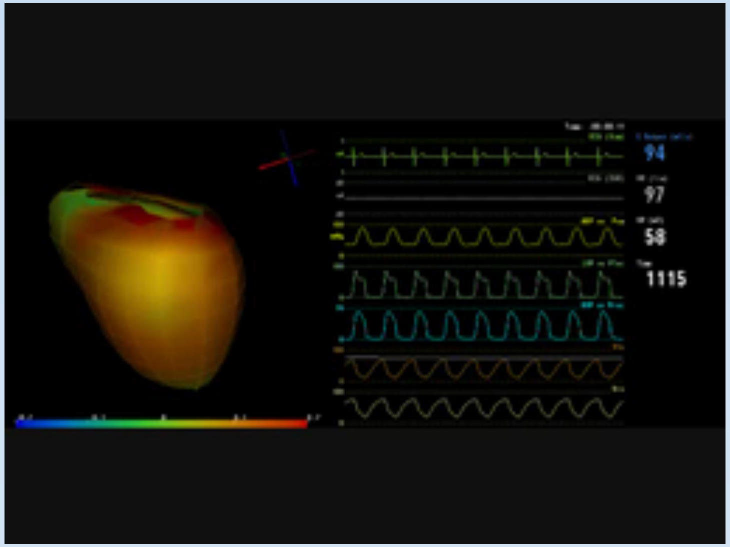


Cardiac Science

Sher, Gavaghan, Rodriguez, Oxford

Mcculloch, Mihaylova, Kerckhoffs, UCSD

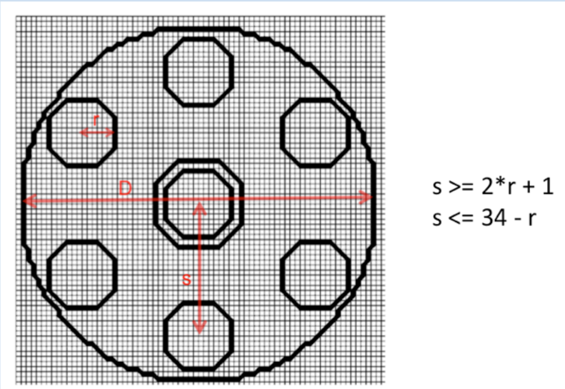
- Heart disease still leading cause of death
- Understanding the underlying physiological mechanisms is cheaper and faster when experimental studies are performed together with mathematical models & computer simulations
- Studying pathologies
- Developing & Testing drugs

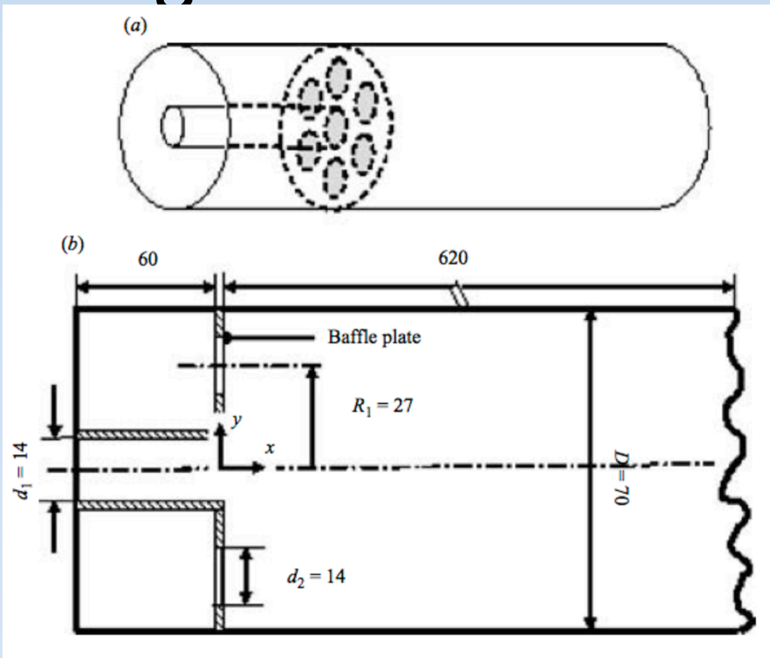


Micromixer optimization

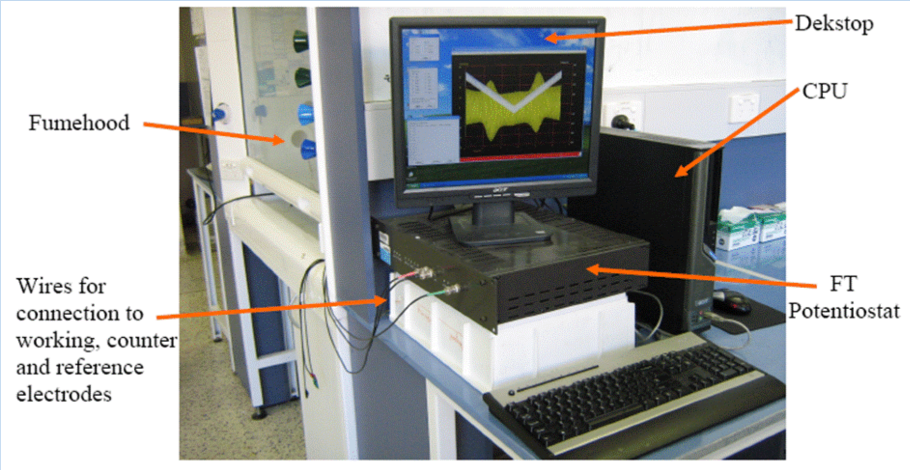
Kipouros, Cambridge, UK

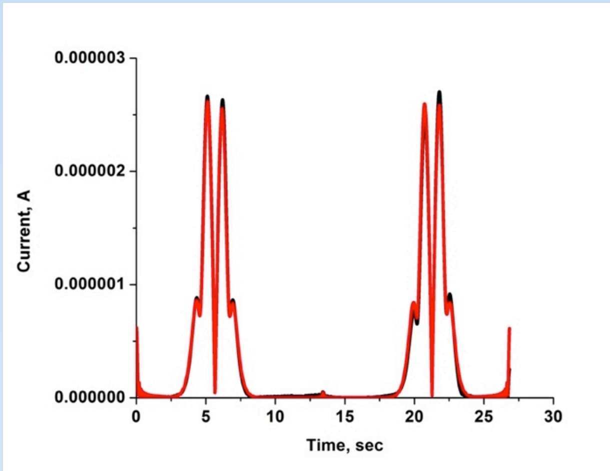
- Microfluidics
  - 10^{-9} to 10^{-18} litres amounts of fluids
  - Gaining importance in various fields
- Micromixer deals with mixing fluids in the smallest scale
- Active vs. passive

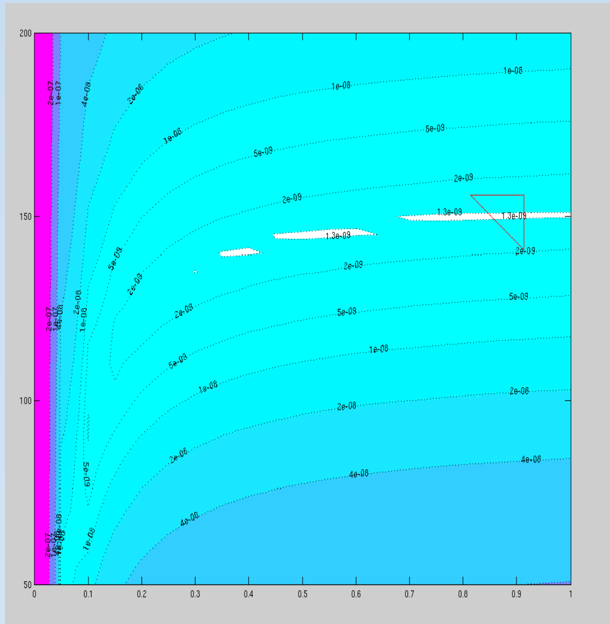




Electro-chemistry
Bond, Gavaghan: Monash, Oxford



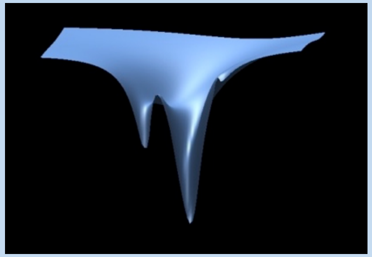


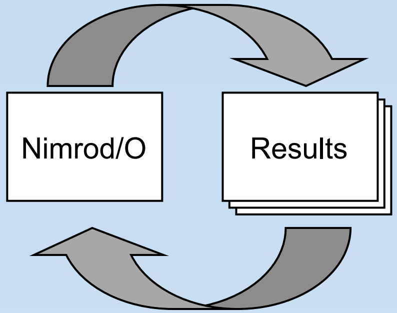


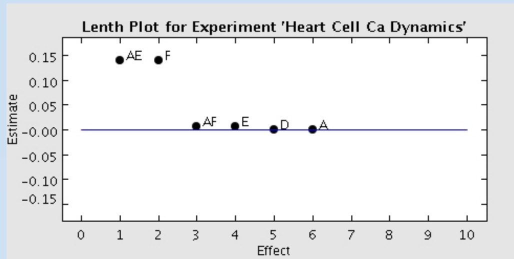
The Nimrod Tools Family

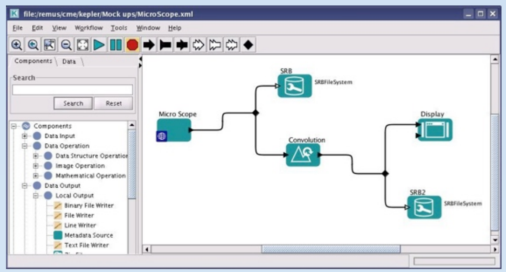
Nimrod supporting “real” science

- 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)

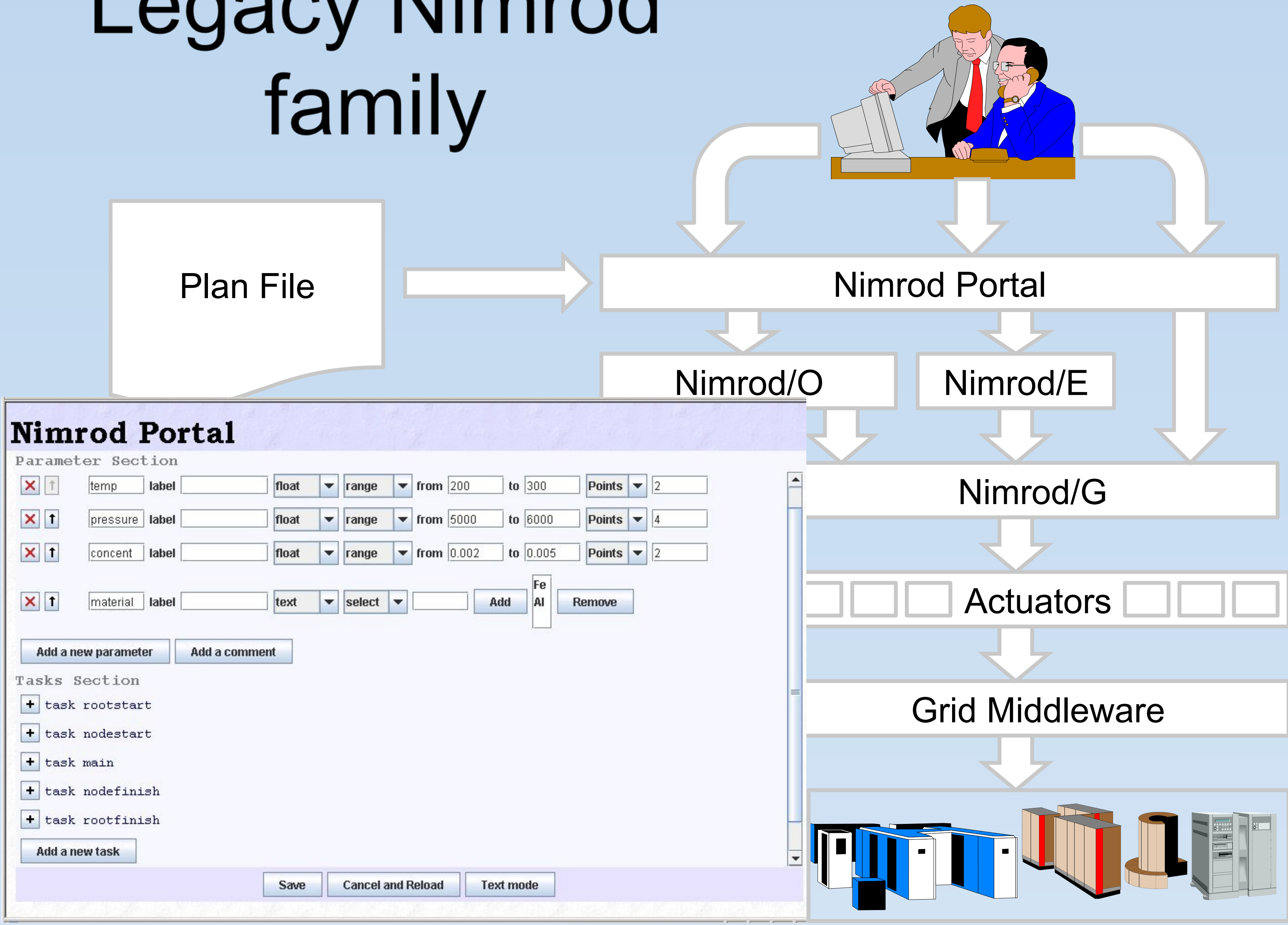




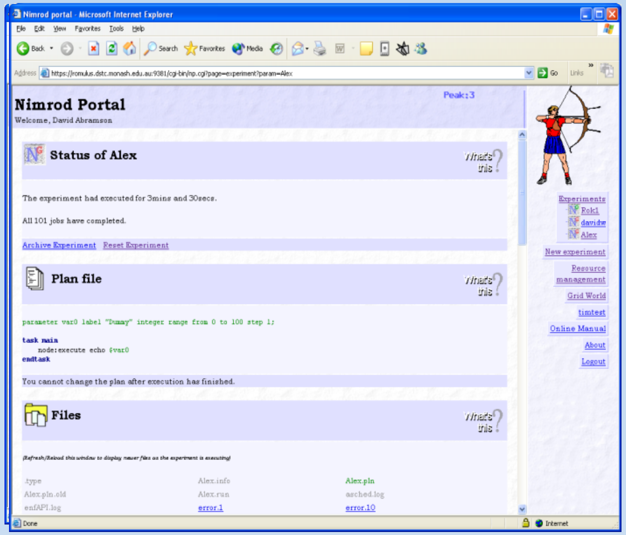




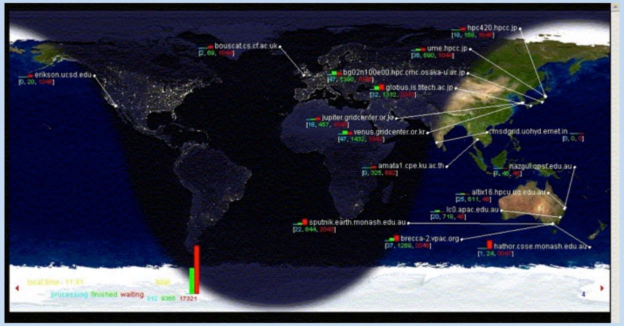
Legacy Nimrod family



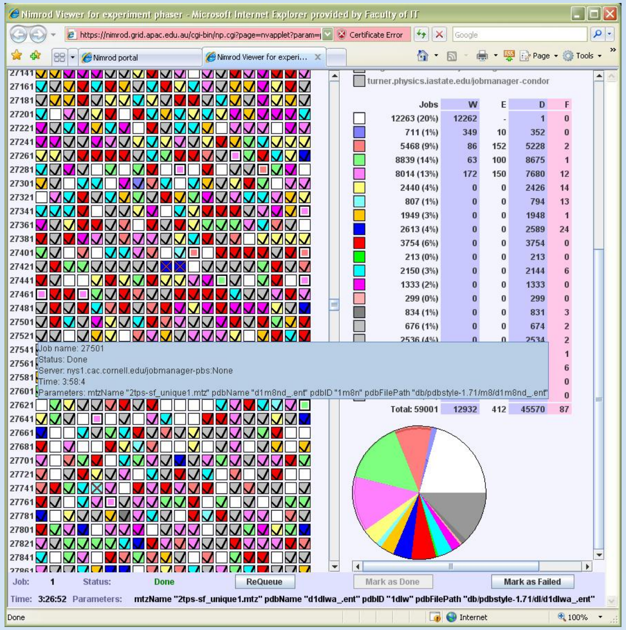
Nimrod Development Cycle



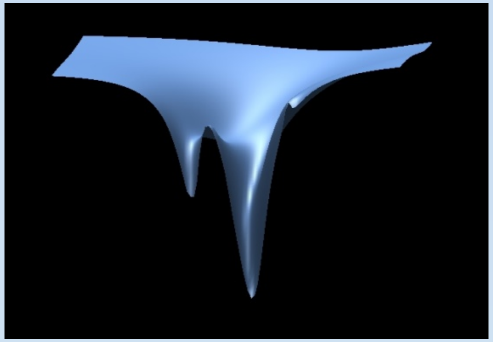
Prepare Jobs using Portal



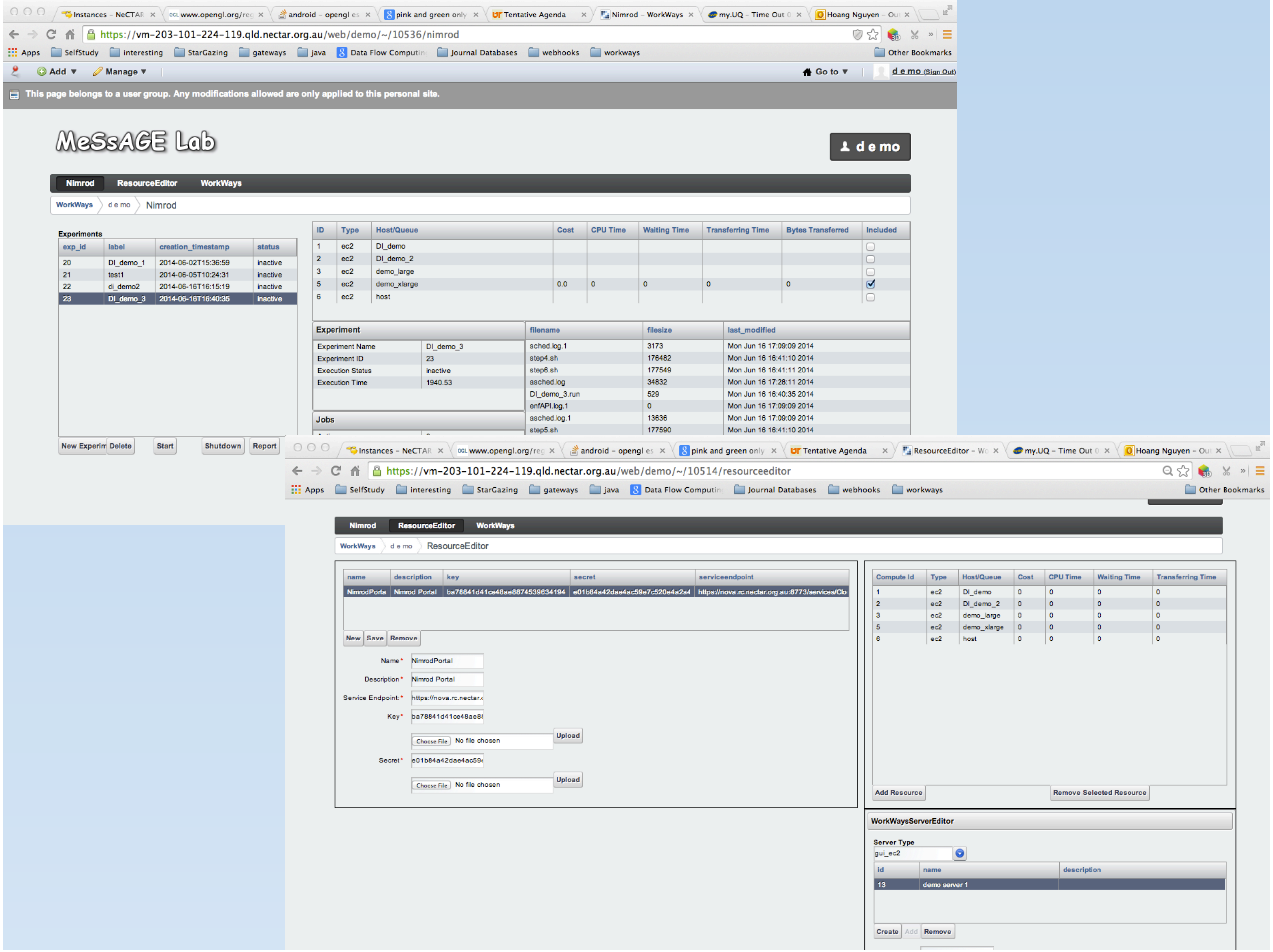
Sent to available machines

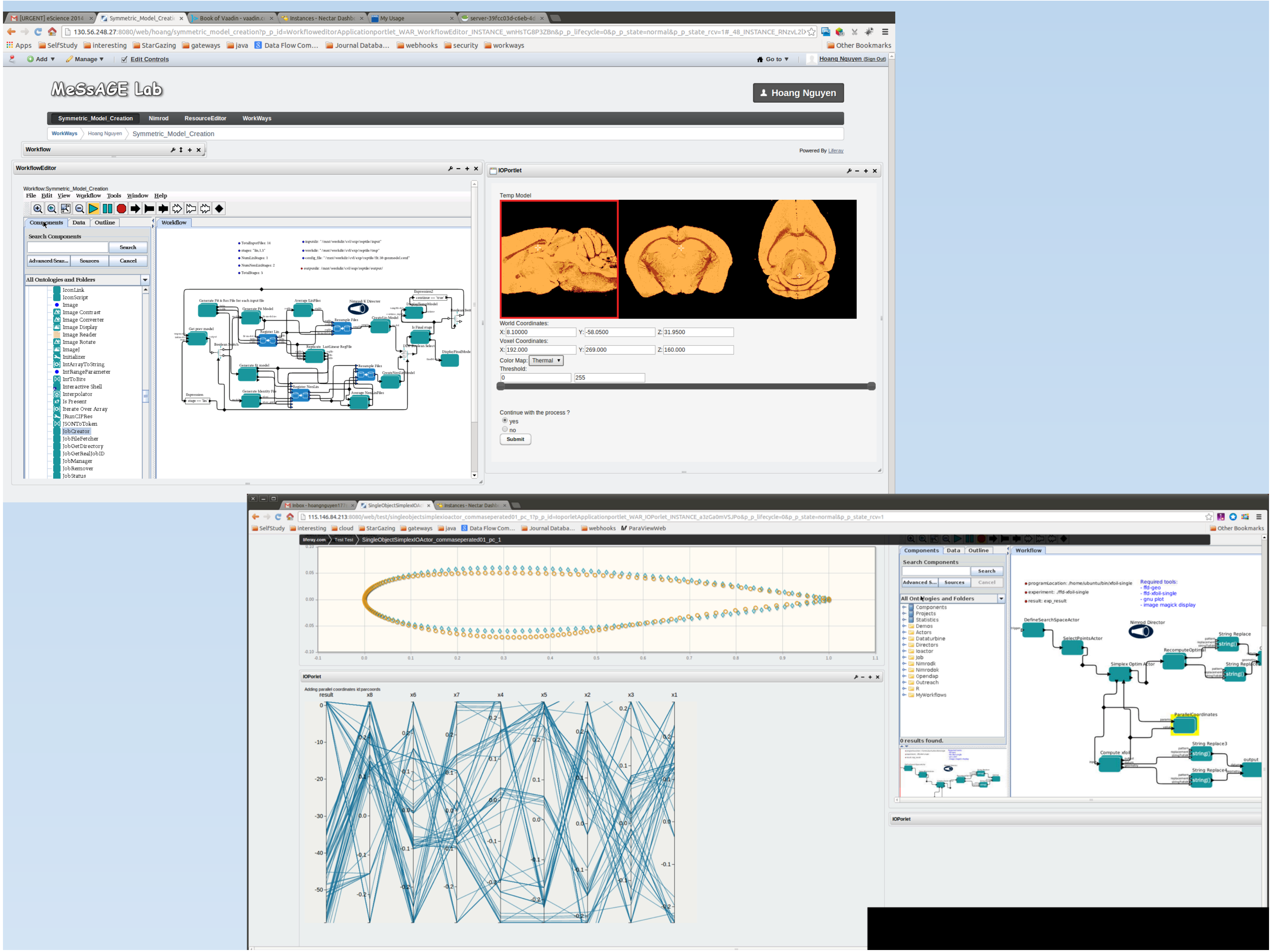


Jobs Scheduled Executed Dynamically



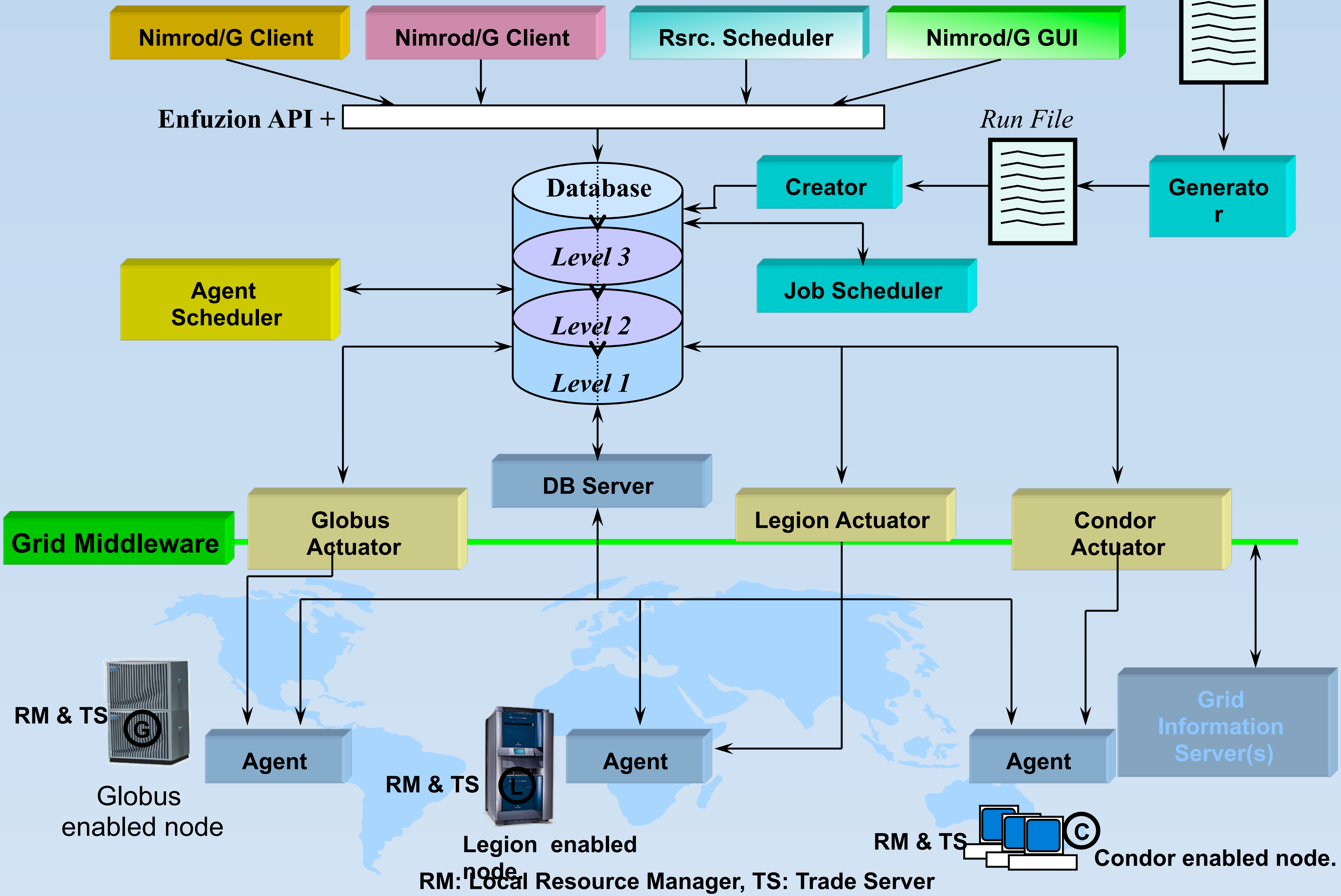
Results displayed & interpreted


https://vm-203-101-224-119.qld.nectar.org.au/web/demo/~/10536/nimrod
This page belongs to a user group. Any modifications allowed are only applied to this personal site.
MeSsAGE Lab
d e mo
Nimrod
ResourceEditor
WorkWays
Experiments
exp_id label creation_timestamp status
20 DI_demo_1 2014-06-02T15:36:59 inactive
21 test1 2014-06-05T10:24:31 inactive
22 di_demo2 2014-06-16T16:15:19 inactive
23 DI_demo_3 2014-06-16T16:40:35 inactive
ID Type Host/Queue Cost CPU Time Waiting Time Transferring Time Bytes Transferred Included
1 ec2 DI_demo
2 ec2 DI_demo_2
3 ec2 demo_large
5 ec2 demo_xlarge 0.0 0 0 0 0
6 ec2 host
Experiment
Experiment Name DI_demo_3
Experiment ID 23
Execution Status inactive
Execution Time 1940.53
Jobs
filename filesize last_modified
sched.log.1 3173 Mon Jun 16 17:09:09 2014
step4.sh 176482 Mon Jun 16 16:41:10 2014
step6.sh 177549 Mon Jun 16 16:41:11 2014
asched.log 34832 Mon Jun 16 17:28:11 2014
DI_demo_3.run 529 Mon Jun 16 16:40:35 2014
enfAPI.log.1 0 Mon Jun 16 17:09:09 2014
asched.log.1 13636 Mon Jun 16 17:09:09 2014
step5.sh 177590 Mon Jun 16 16:41:10 2014
New Experim Delete Start Shutdown Report
https://vm-203-101-224-119.qld.nectar.org.au/web/demo/~/10514/resourceeditor
ResourceEditor
name description key secret serviceendpoint
NimrodPorta Nimrod Portal ba78841d41ce48ae8874539634194 e01b84a42dae4ac59e7c520e4a2a4 https://nova.rc.nectar.org.au:8773/services/Clo
New Save Remove
Name* NimrodPortal
Description* Nimrod Portal
Service Endpoint:* https://nova.rc.nectar.
Key* ba78841d41ce48ae8
Choose File No file chosen Upload
Secret* e01b84a42dae4ac59
Choose File No file chosen Upload
Compute Id Type Host/Queue Cost CPU Time Waiting Time Transferring Time
1 ec2 DI_demo 0 0 0 0
2 ec2 DI_demo_2 0 0 0 0
3 ec2 demo_large 0 0 0 0
5 ec2 demo_xlarge 0 0 0 0
6 ec2 host 0 0 0 0
Add Resource
Remove Selected Resource
WorkWaysServerEditor
Server Type
gui_ec2
id name description
13 demo server 1
Create Add Remove



MeSsAGE Lab
Hoang Nguyen
Symmetric_Model_Creation
Nimrod
ResourceEditor
WorkWays
Workflow
WorkflowEditor
IOPortlet
Temp Model
World Coordinates:
Voxel Coordinates:
Color Map:
Thermal
Threshold:
Continue with the process ?
yes
no
Submit
Powered By Liferay
Components
Data
Outline
Search Components
Search
Sources
Cancel
All Ontologies and Folders
Adding parallel coordinates id:parcoords
ParallelCoordinates

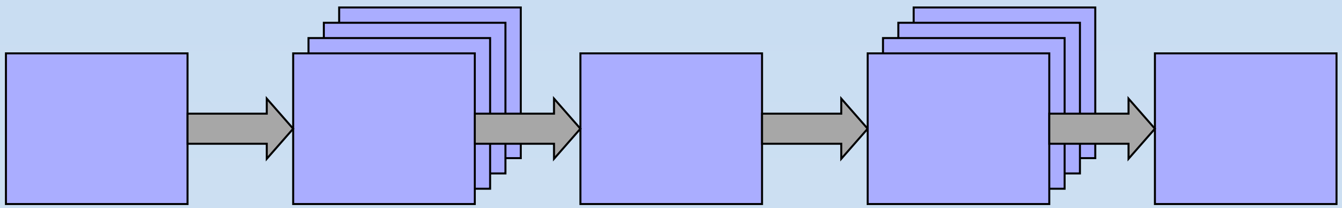
Nimrod/G Architecture

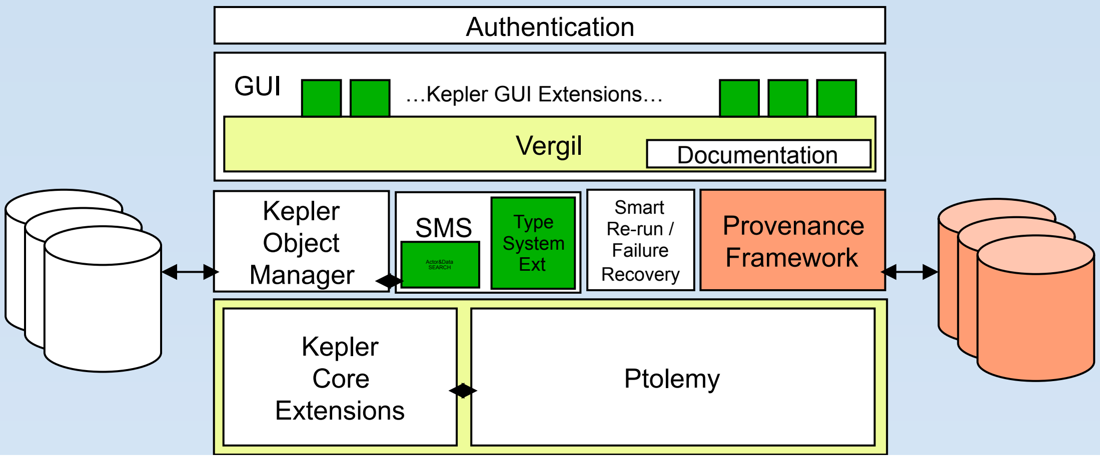


Scientific Workflows and Nimrod

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

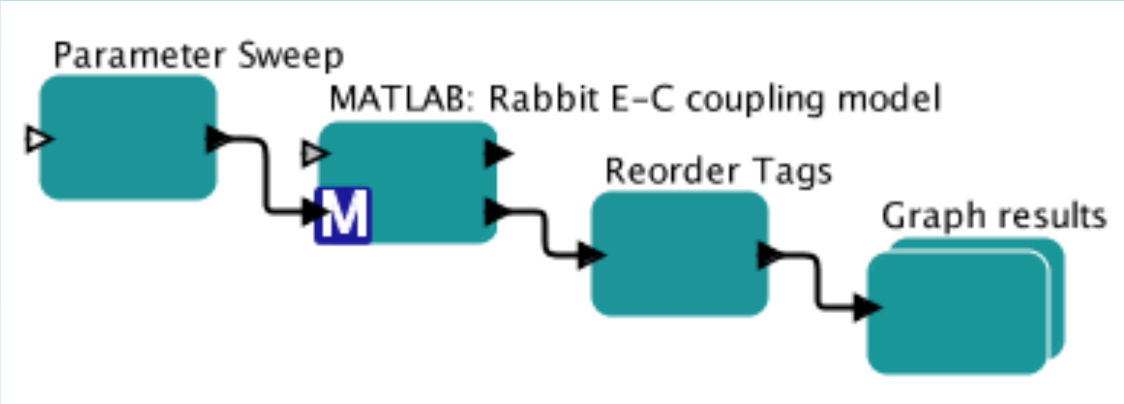


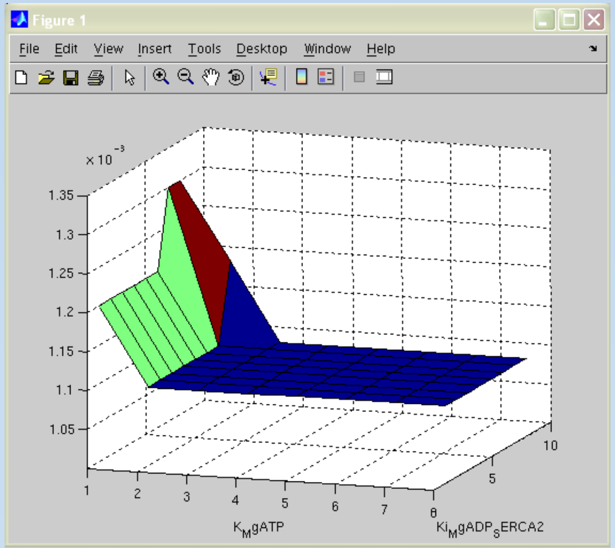


Workflow Threading

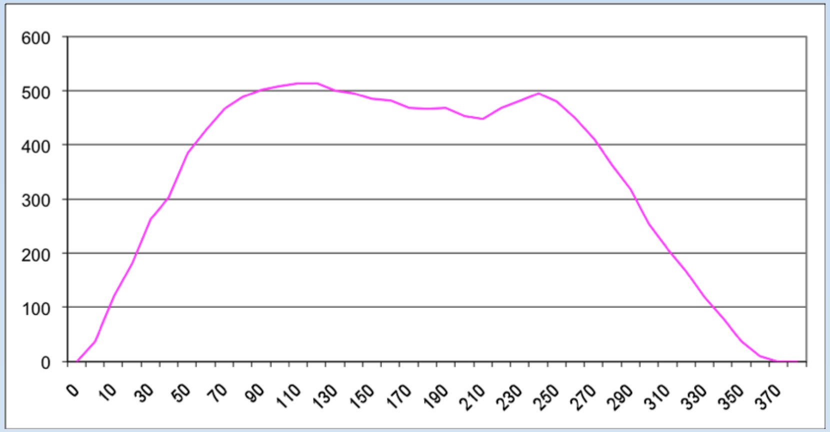
- Nimrod parameter combinations can be viewed as threads
- Multi-threaded workflows allow independent sequences in a workflow to run concurrently
  - This might be the whole workflow, or part of the workflow
- Tokens in different threads do not interact with each other in the workflow

Complete Parameter Sweep

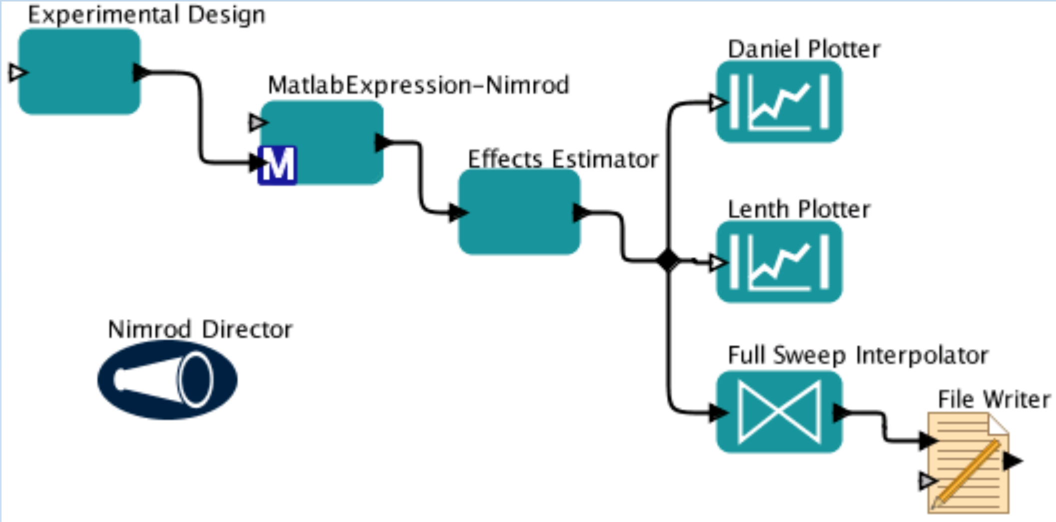


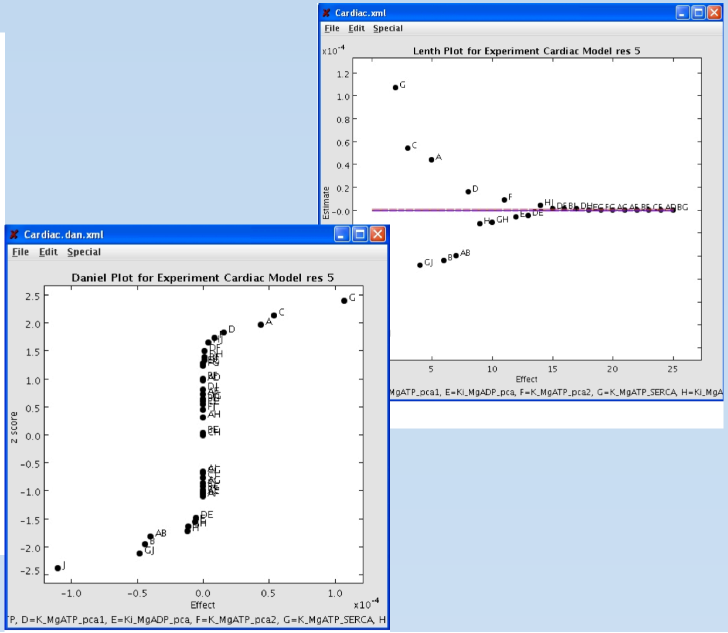


- Using a MATLAB actor provided by Kepler
- Local spawn
  - Multiple thread ran concurrently on a computer with 8 cores (2 x quads)
  - Workflow execution was just under 8 times faster
- Remote Spawn
  - 100's – 1000's of remote processes

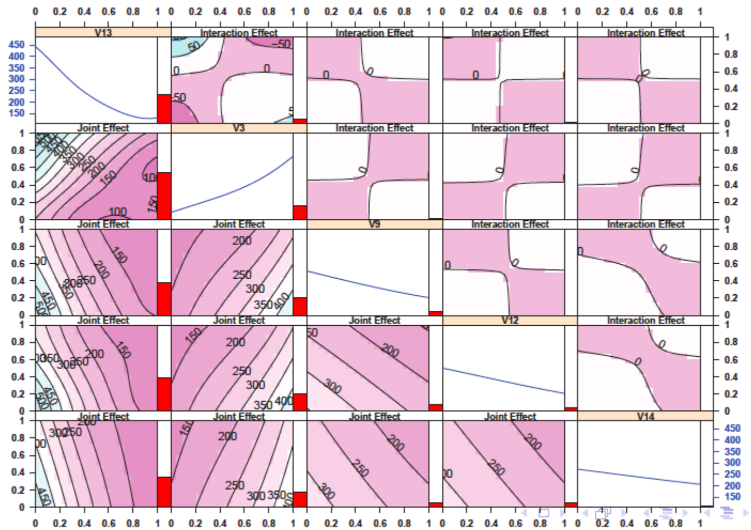


Nimrod/EK Actors





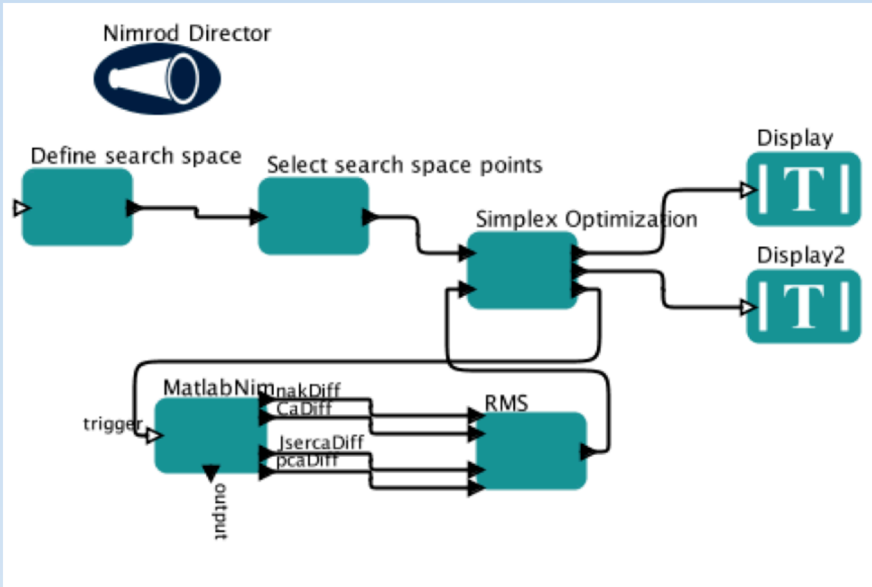
- Actors for generating and analyzing designs
- Leverage concurrent infrastructure



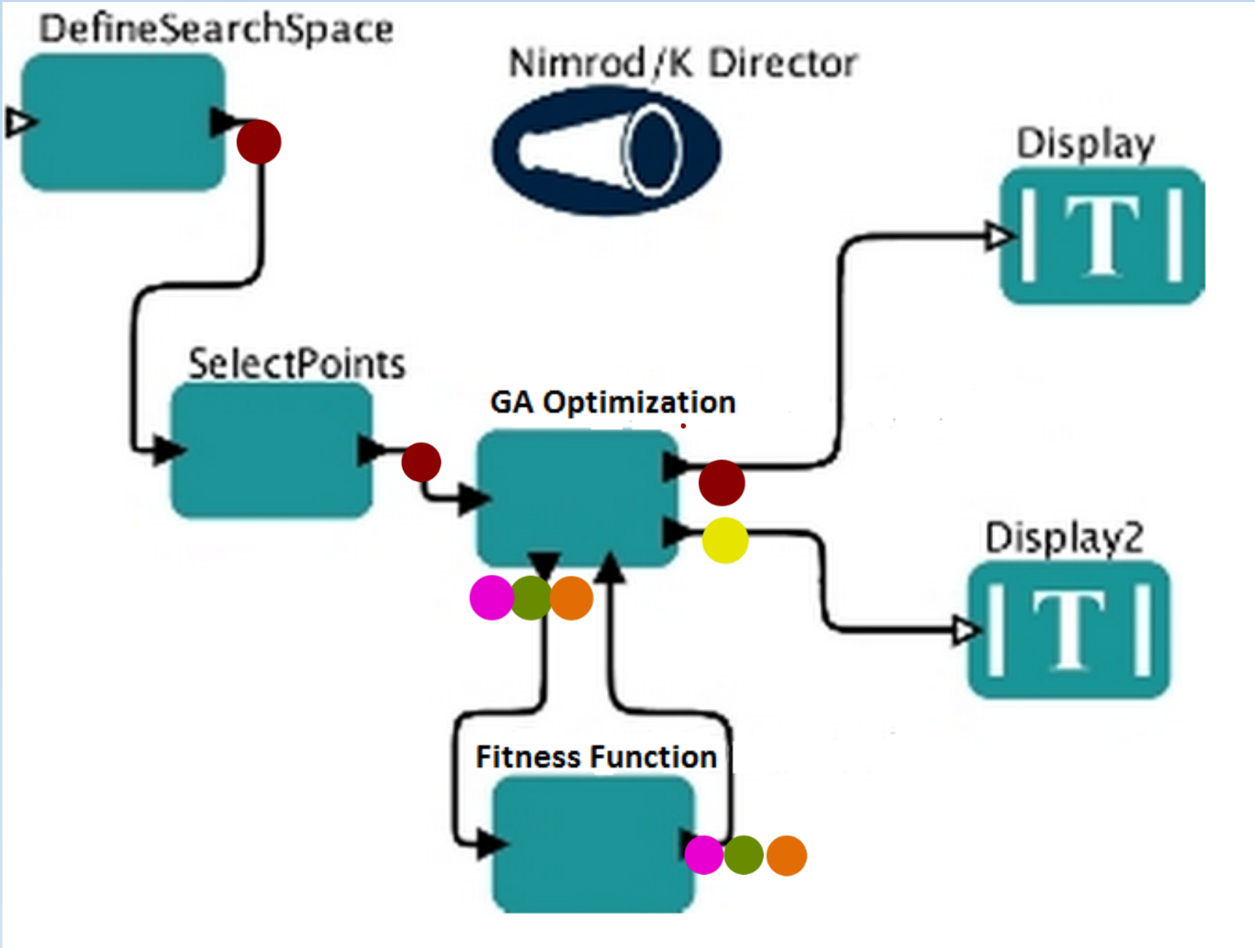
Optimization with Nimrod/OK

Nimrod/OK Workflows

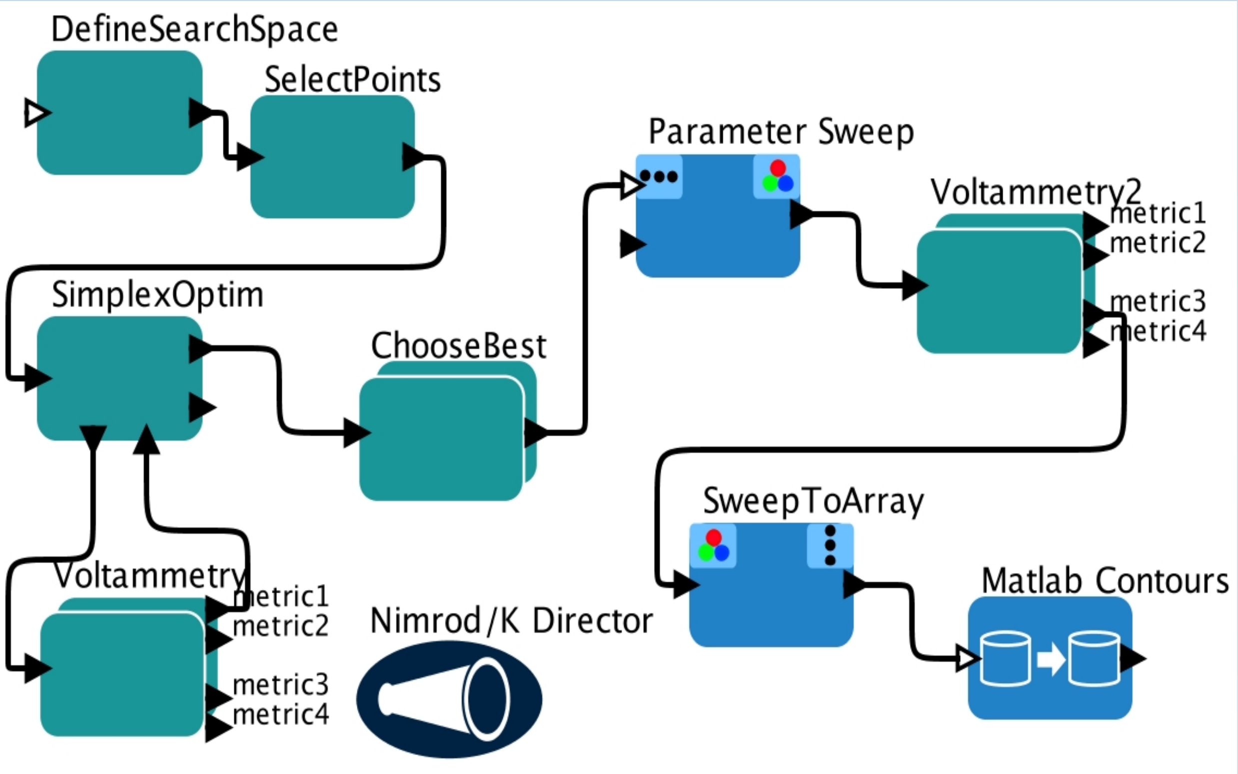
- Nimrod/K supports parallel execution
- General template for search
  - Built from key components
- Can mix and match optimization algorithms



Using different optimisers



Hybrid Optimization



Execution Engines

Nimrod over Clusters

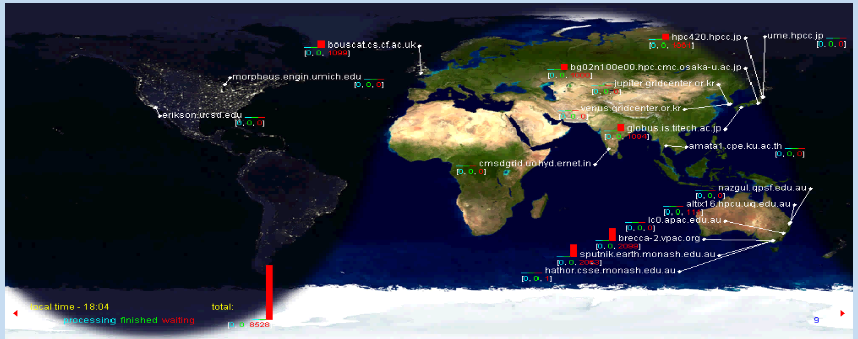
Jobs / Nimrod experiment

Nimrod

Actuator, e.g., SGE, PBS, LSF, Condor

Local Batch System

Nimrod over Grids



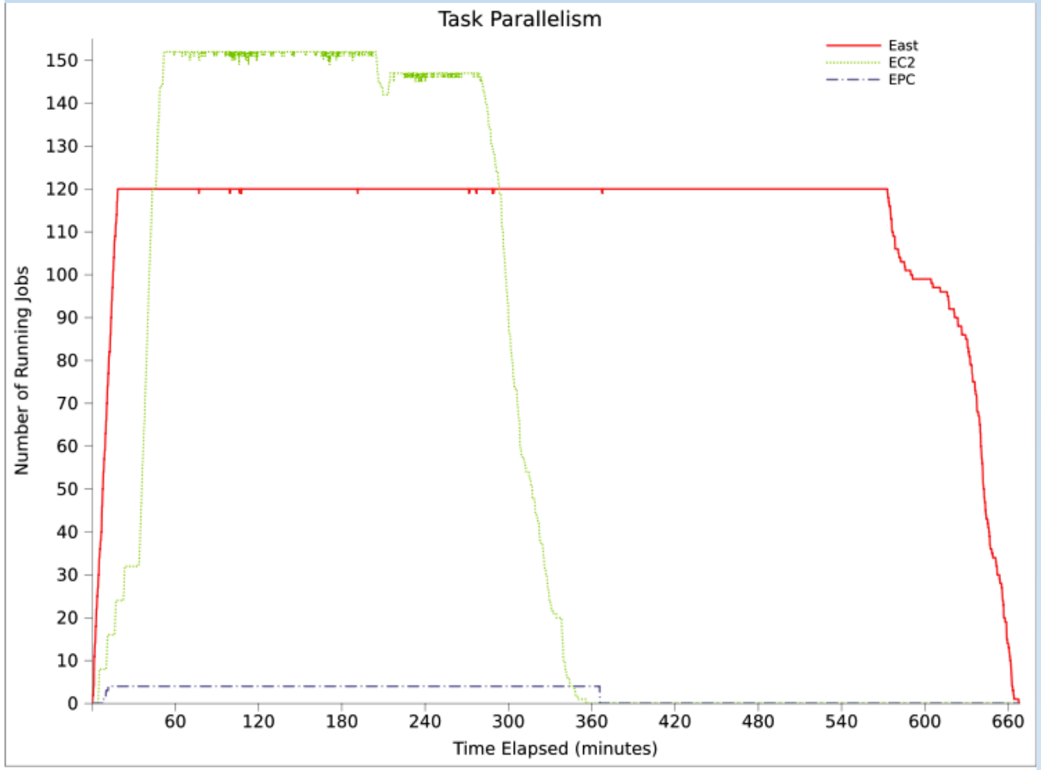
- Advantages
  - Wide area elastic computing
  - Portal based point-of-presence independent of location of computational resources
  - Grid level security
  - Computational economy proposed
    - New scheduling and data challenges
  - Virtualization proposed (Based on .NET!)
- Leveraged Grid middleware
  - Globus, Legion, ad-hoc standards

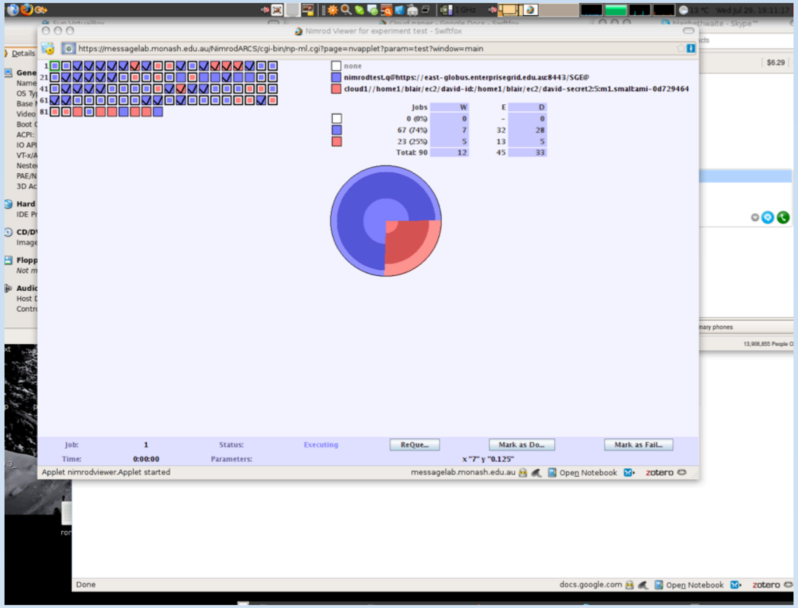
Leveraging Cloud Infrastructure

- Centralisation is easier
  - (Clusters vs Grid)
- Virtualisation improves interoperability and scalability
  - Build once, run everywhere
- Computational economy, for real
  - Deadline driven
    - "I need this finished by Monday morning!"
  - Budget driven
    - "Here's my credit card, do this as quickly and cheaply as possible."
- Cloud bursting
  - Scale-out to supplement locally and nationally available resources

Nimrod's Grid Economy is Cloud bursting

Resource	#jobs completed	Total job time (h:m:s)	μ / σ Job runtime (mins)
East	818	1245:37:23	91/5.7
EC2	613	683:34:05	67/14.2

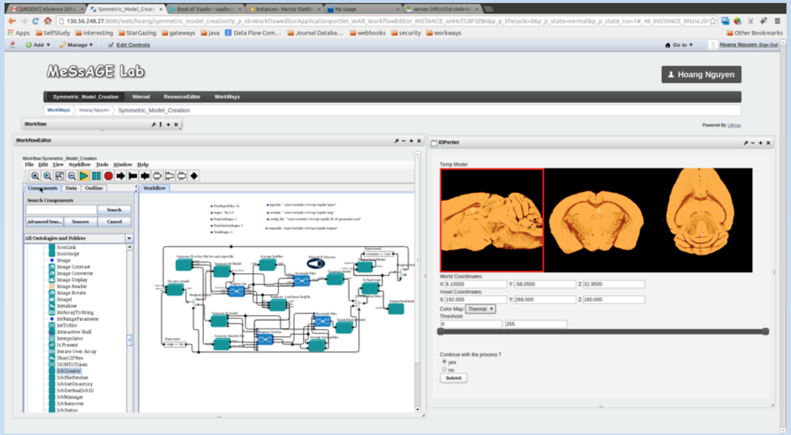




Workways

Interacting with Workflows

WorkWays



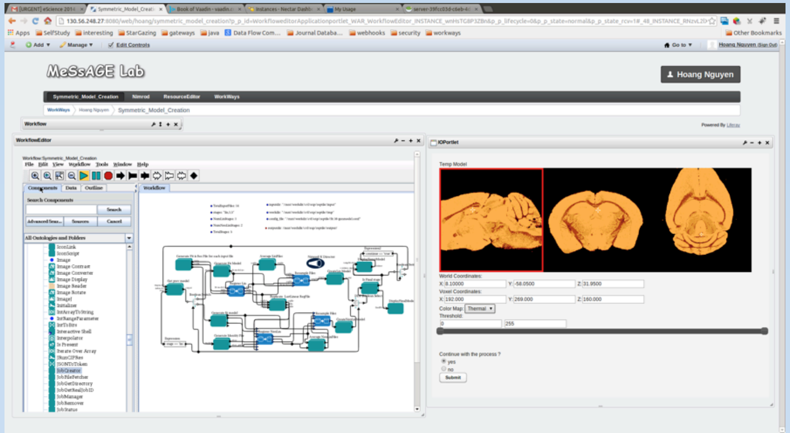
- Ease of use of Science Gateway
- Workflows as service
- IO through portlets
- Extensibility
  - Different IO mechanisms, protocols and topolgy
  - Different UI clients
- Currently Kepler as the workflow engine

Motivation

- Support human-in-the-loop workflows
  - Ability to perform IO operations with a continuously running workflow
- Benefits
  - Insights into workflow execution
  - Steer the execution

WorkWays

- Leverage various existing technologies
  - Kepler workflow
  - Nimrod family toolkit
  - Liferay portal
- Virtual desktops
- AAF
- Various Web-based visualization tools
  - Parallel coordinates
  - Para-view Web



IOActor

- Generic actor
  - Simplify the creation of (new) IO actors
  - Instantiate an IOActor & provide the actor definition
- IOActor definition
  - Actor name
  - Number of (supported) clients
  - Operation: input/output/inout
  - Additional information

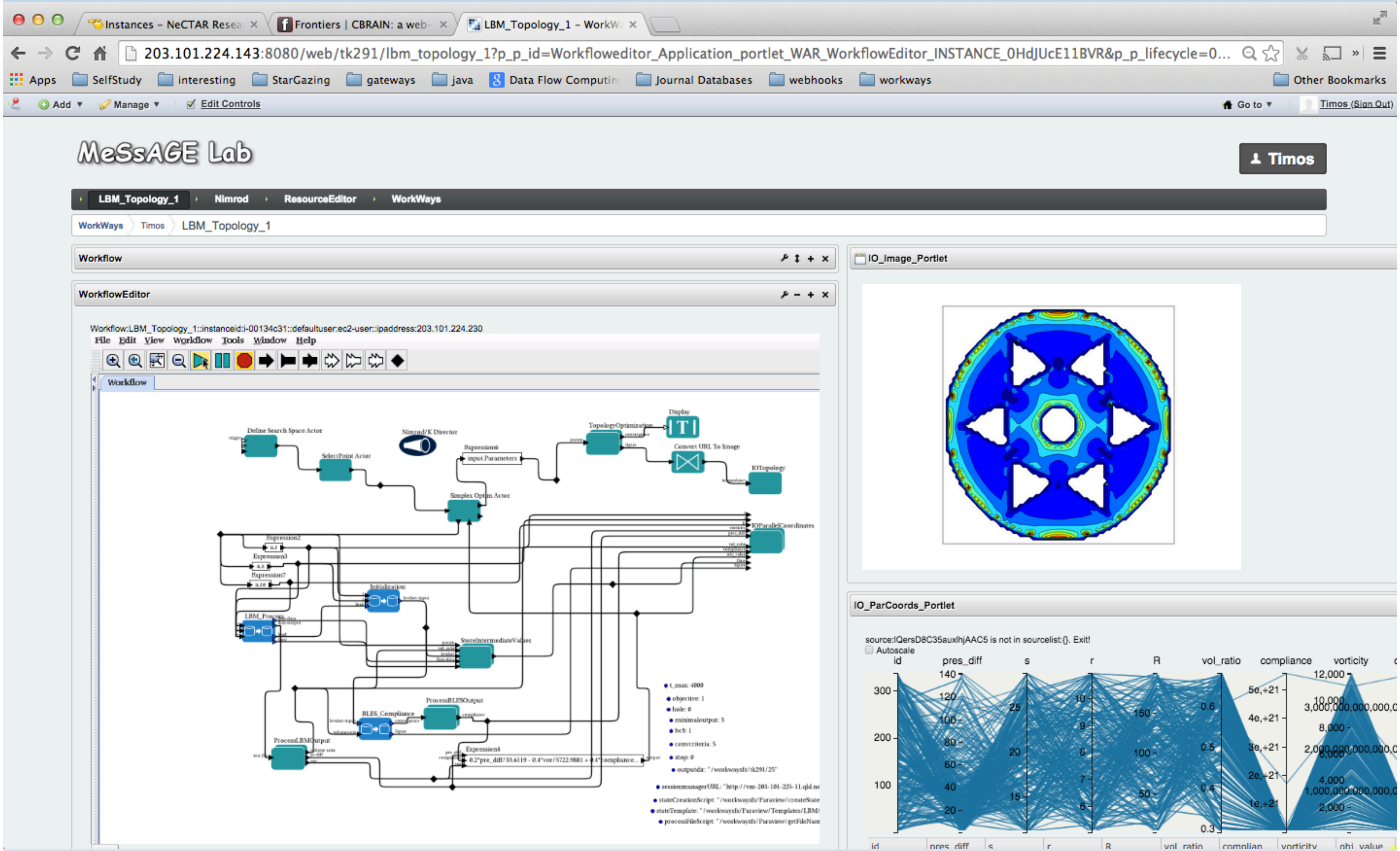
```
{
    actor: "ImageInOutActor"
    {
        operation: "inout"
        input_type: "text"
        prompt: "Choose an area in the image"
        output_type: "binary"
        display_type: "image"
        action: "subarea"
        wait_client: true
        wait_for_input: true
    }
}
```

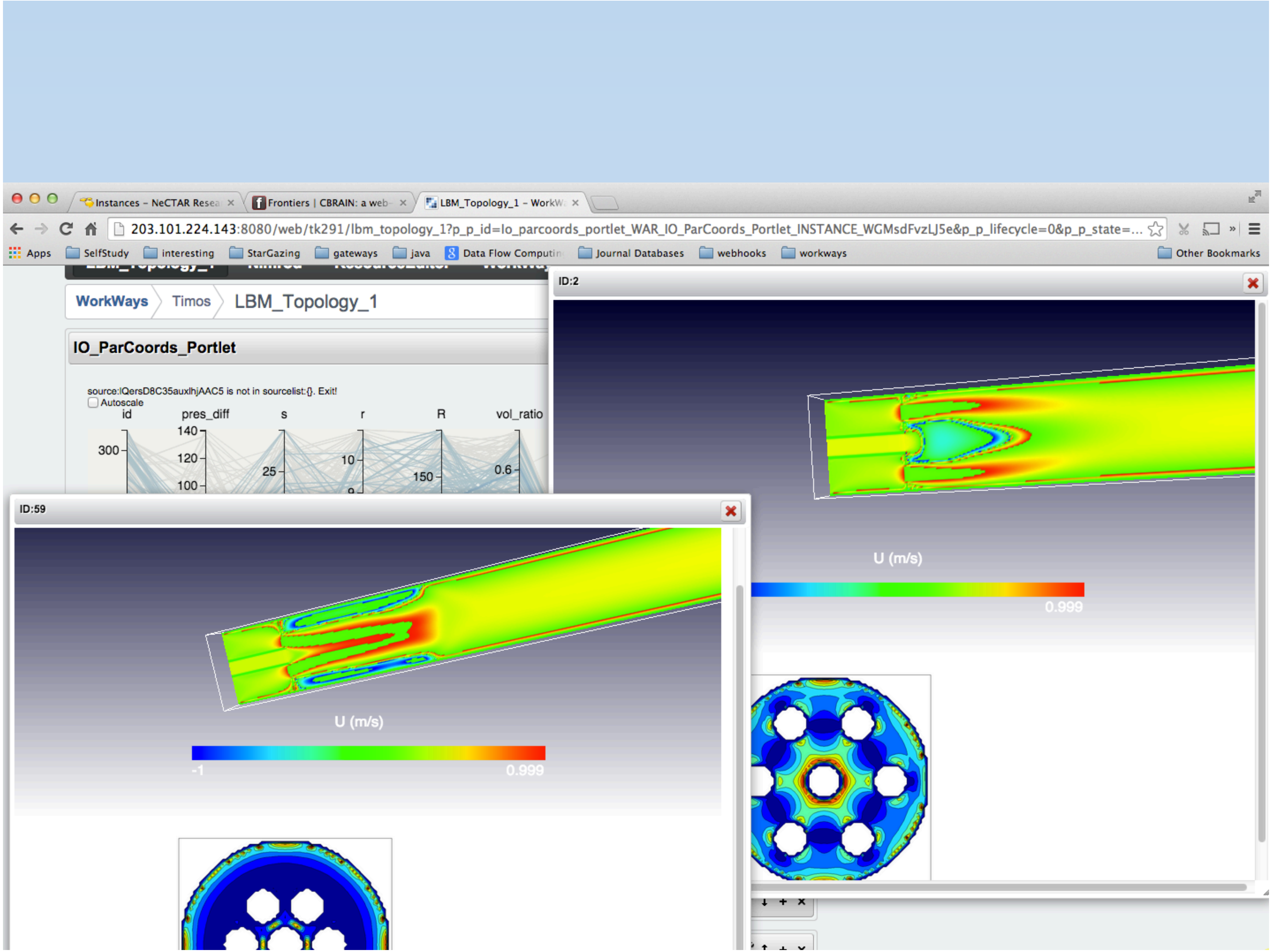
IOPortlet

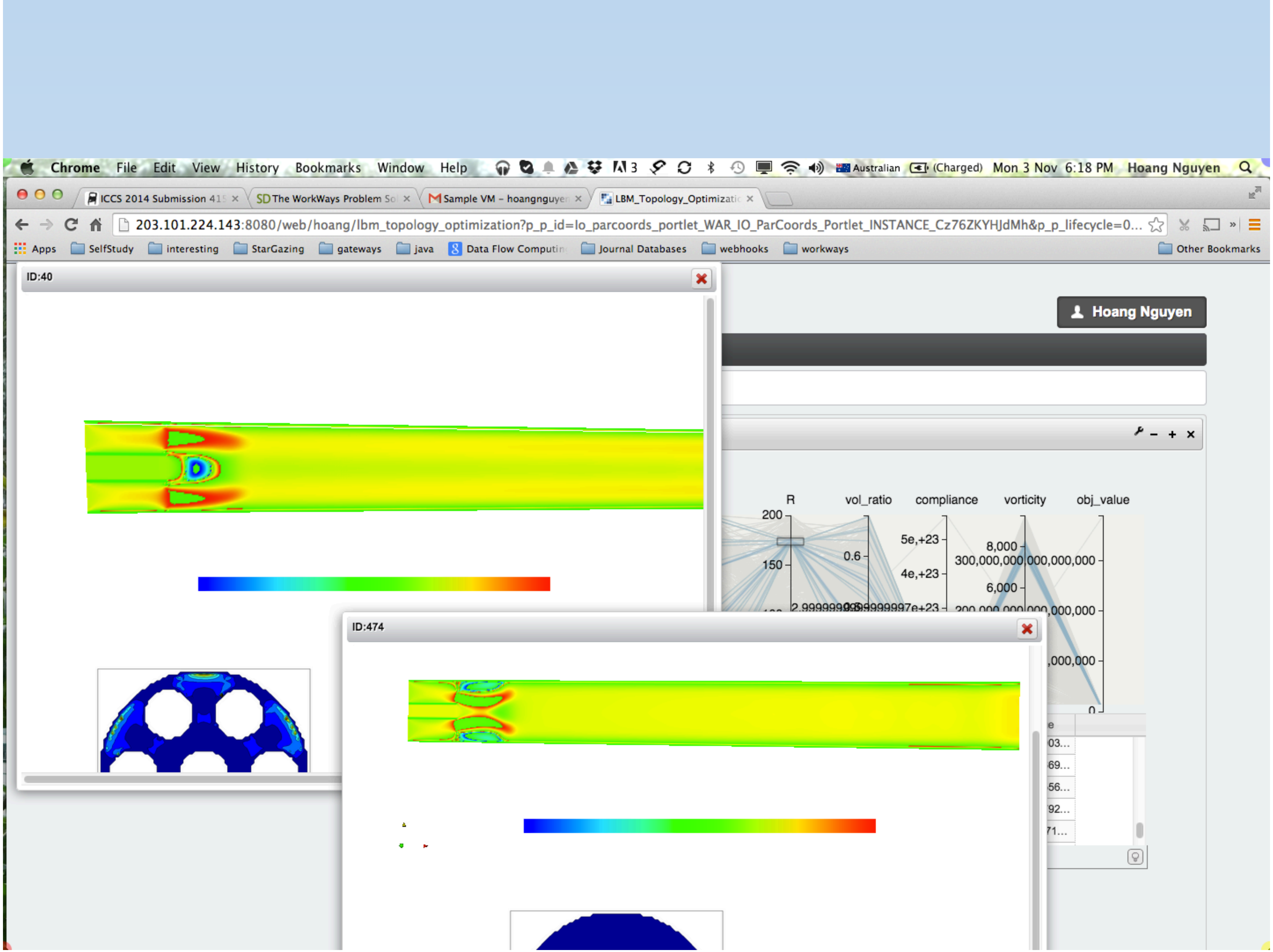
- Web UI client
- Vaadin framework
  - framework for building rich Web applications
- JSR-286 portlet
  - UI elements generated based on requests from connected IOActor
  - Limited UI elements

Examples

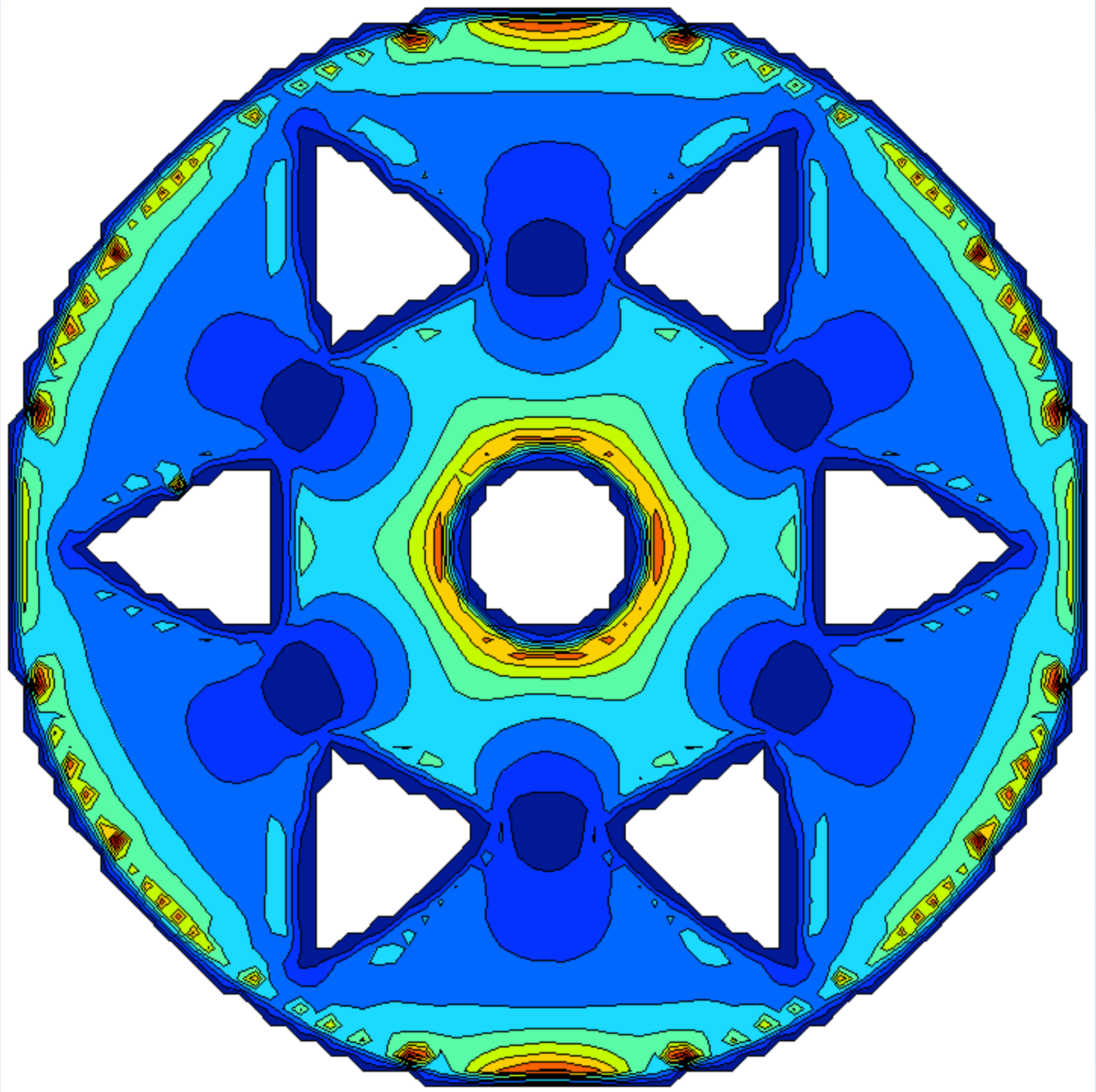
Micro-mixer design



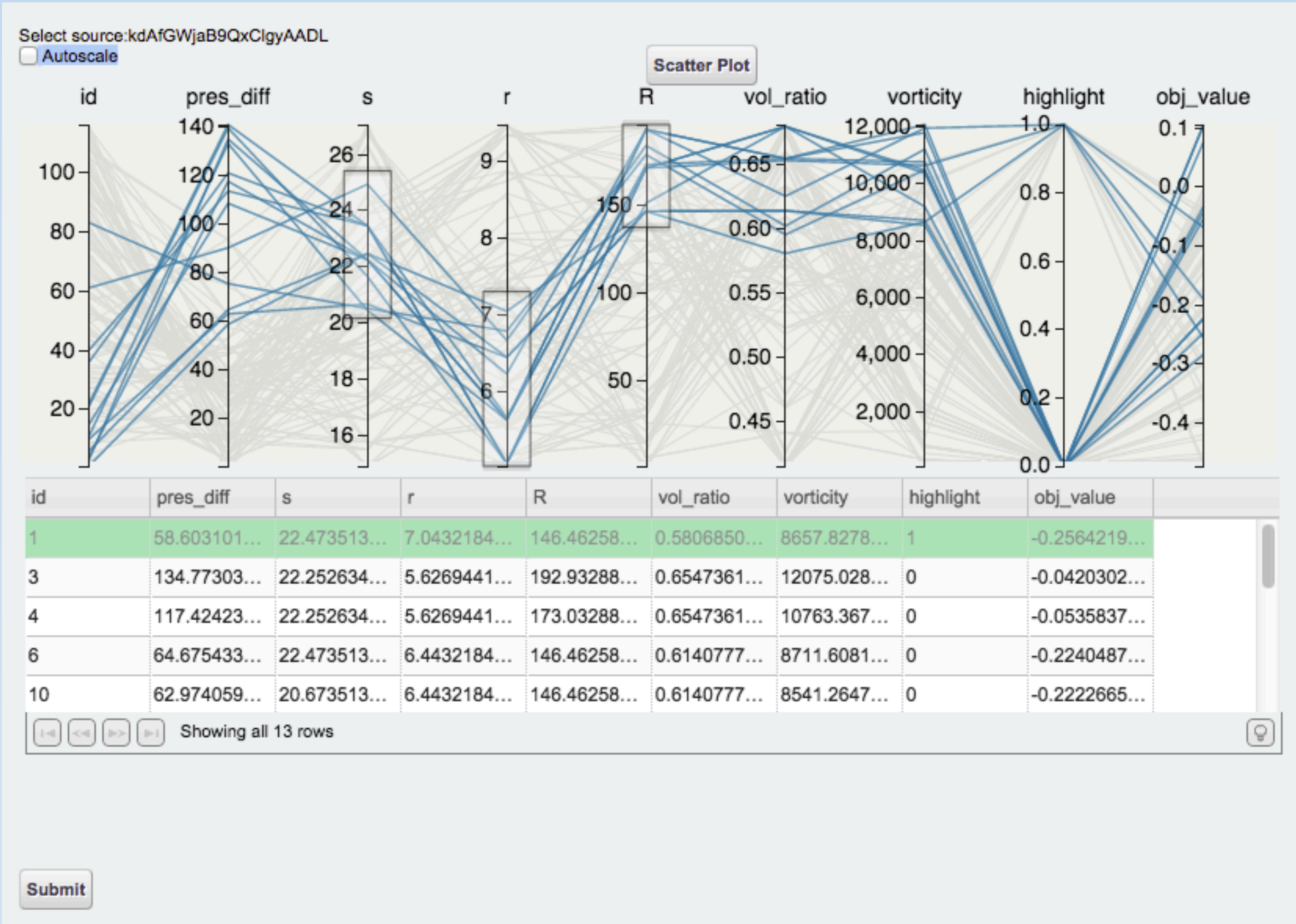

Instances – NeCTAR Resea
Frontiers | CBRAIN: a web
LBM_Topology_1 – WorkW
203.101.224.143:8080/web/tk291/lbm_topology_1?p_p_id=lo_parcoords_portlet_WAR_IO_ParCoords_Portlet_INSTANCE_WGMsdFvzLJ5e&p_p_lifecycle=0&p_p_state=...
Apps
SelfStudy
interesting
StarGazing
gateways
java
Data Flow Computin
Journal Databases
webhooks
workways
Other Bookmarks
WorkWays
Timos
LBM_Topology_1
IO_ParCoords_Portlet
source:lQersD8C35auxlhjAAC5 is not in sourcelist:{}. Exit!
Autoscale
id
pres_diff
s
r
R
vol_ratio
ID:2
U (m/s)
0.999
ID:59
U (m/s)
-1
0.999


Chrome File Edit View History Bookmarks Window Help
Mon 3 Nov 6:18 PM Hoang Nguyen
ICCS 2014 Submission 41
The WorkWays Problem So
Sample VM – hoangnguyen
LBM_Topology_Optimizatio
203.101.224.143:8080/web/hoang/lbm_topology_optimization?p_p_id=lo_parcoords_portlet_WAR_IO_ParCoords_Portlet_INSTANCE_Cz76ZKYHJdMh&p_p_lifecycle=0...
Apps SelfStudy interesting StarGazing gateways java Data Flow Computin Journal Databases webhooks workways Other Bookmarks
ID:40
ID:474
Hoang Nguyen
R
vol_ratio
compliance
vorticity
obj_value
200
150
0.6
5e,+23
4e,+23
8,000
6,000
300,000,000,000,000,000
200,000,000,000,000,000
0

Optimal Topology

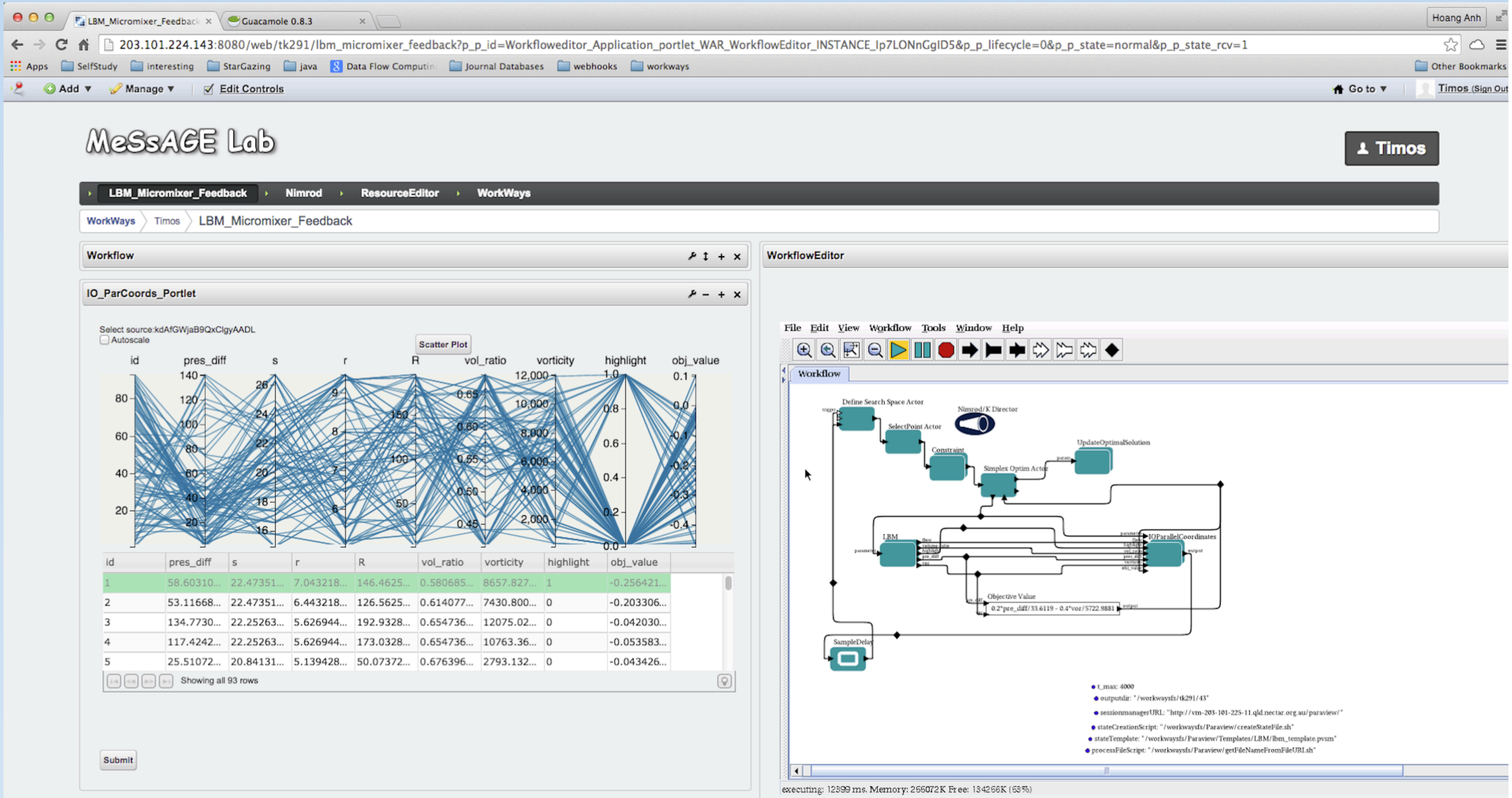


Parallel Coordinates

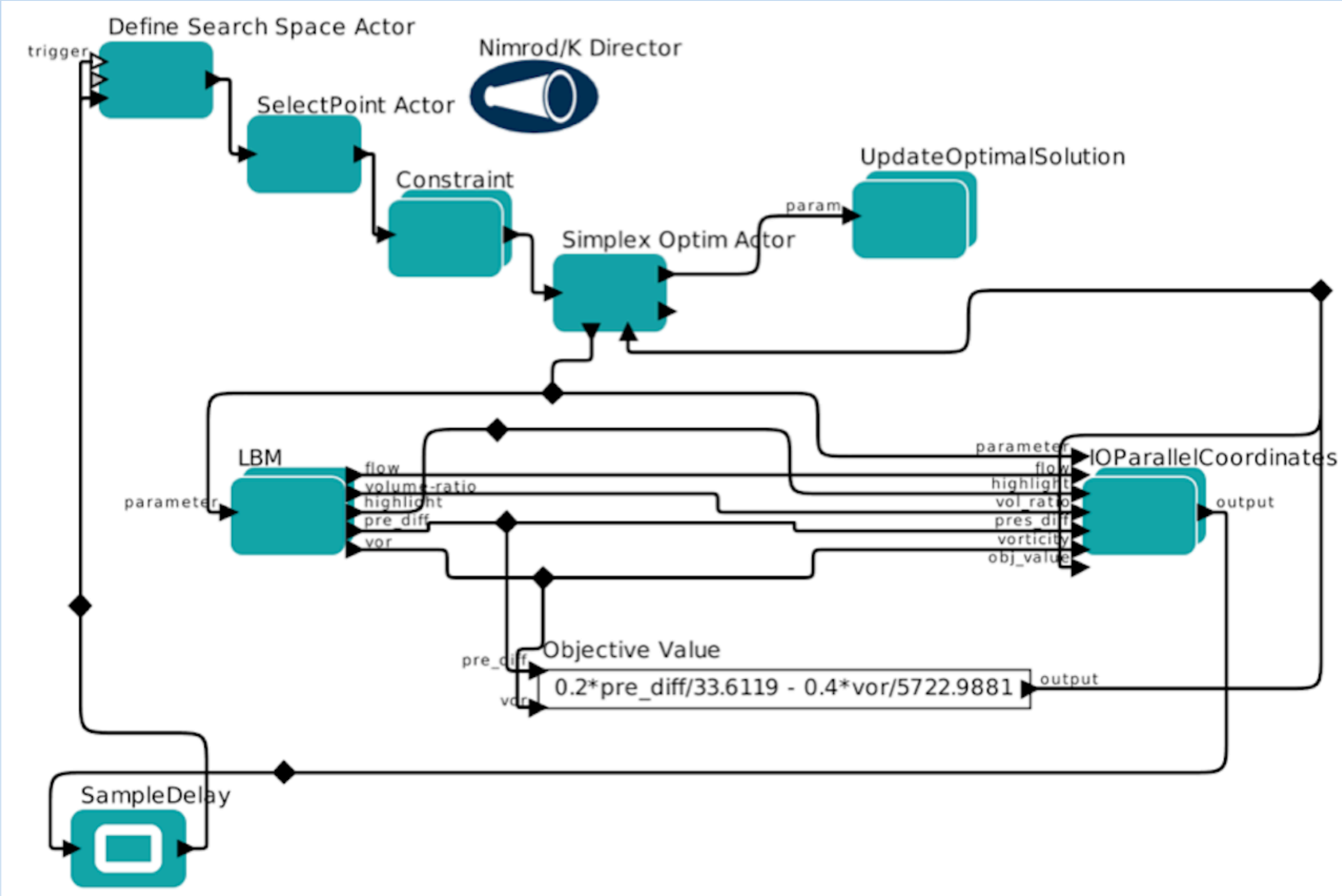


id	pres_diff	s	r	R	vol_ratio	vorticity	highlight	obj_value
1	58.603101...	22.473513...	7.0432184...	146.46258...	0.5806850...	8657.8278...	1	-0.2564219...
3	134.77303...	22.252634...	5.6269441...	192.93288...	0.6547361...	12075.028...	0	-0.0420302...
4	117.42423...	22.252634...	5.6269441...	173.03288...	0.6547361...	10763.367...	0	-0.0535837...
6	64.675433...	22.473513...	6.4432184...	146.46258...	0.6140777...	8711.6081...	0	-0.2240487...
10	62.974059...	20.673513...	6.4432184...	146.46258...	0.6140777...	8541.2647...	0	-0.2222665...

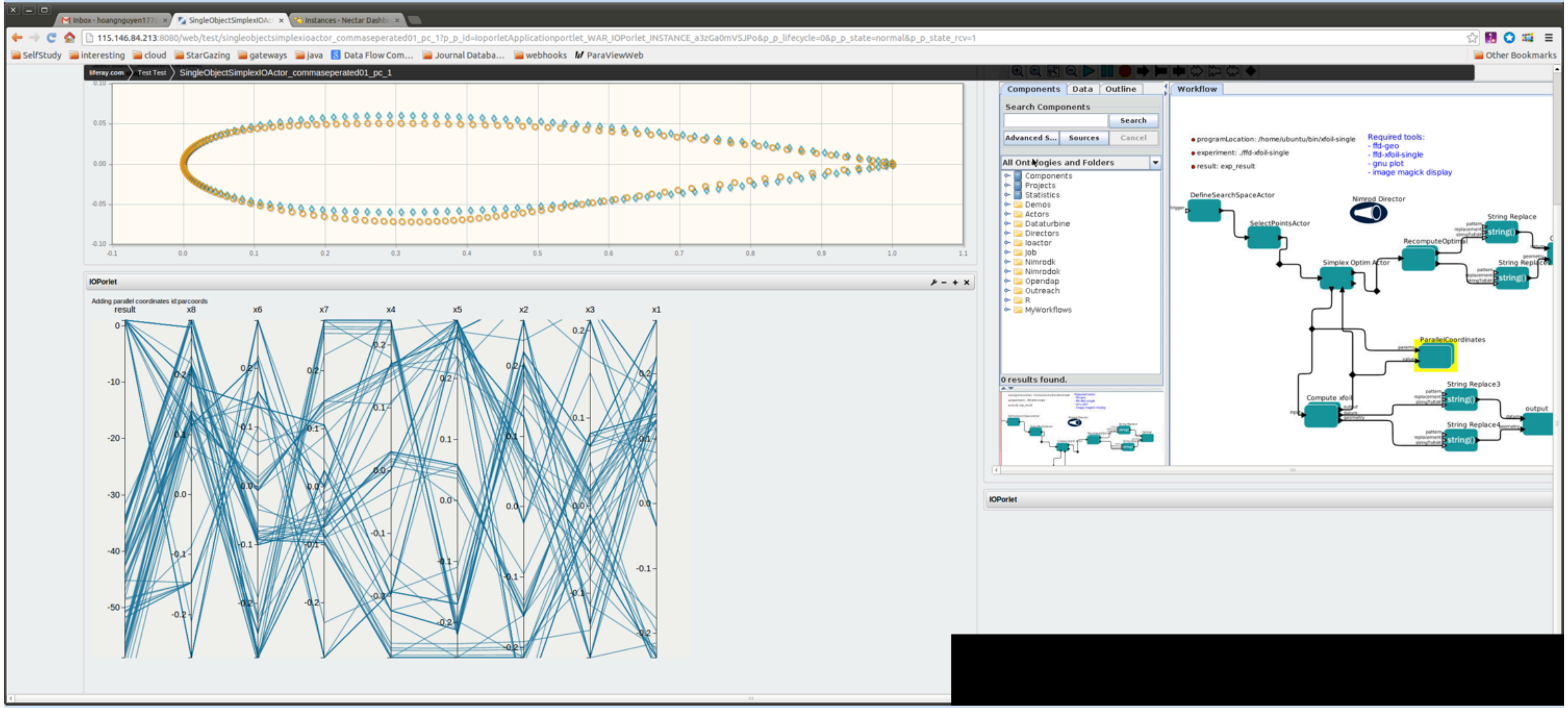
Micromixer Optimization



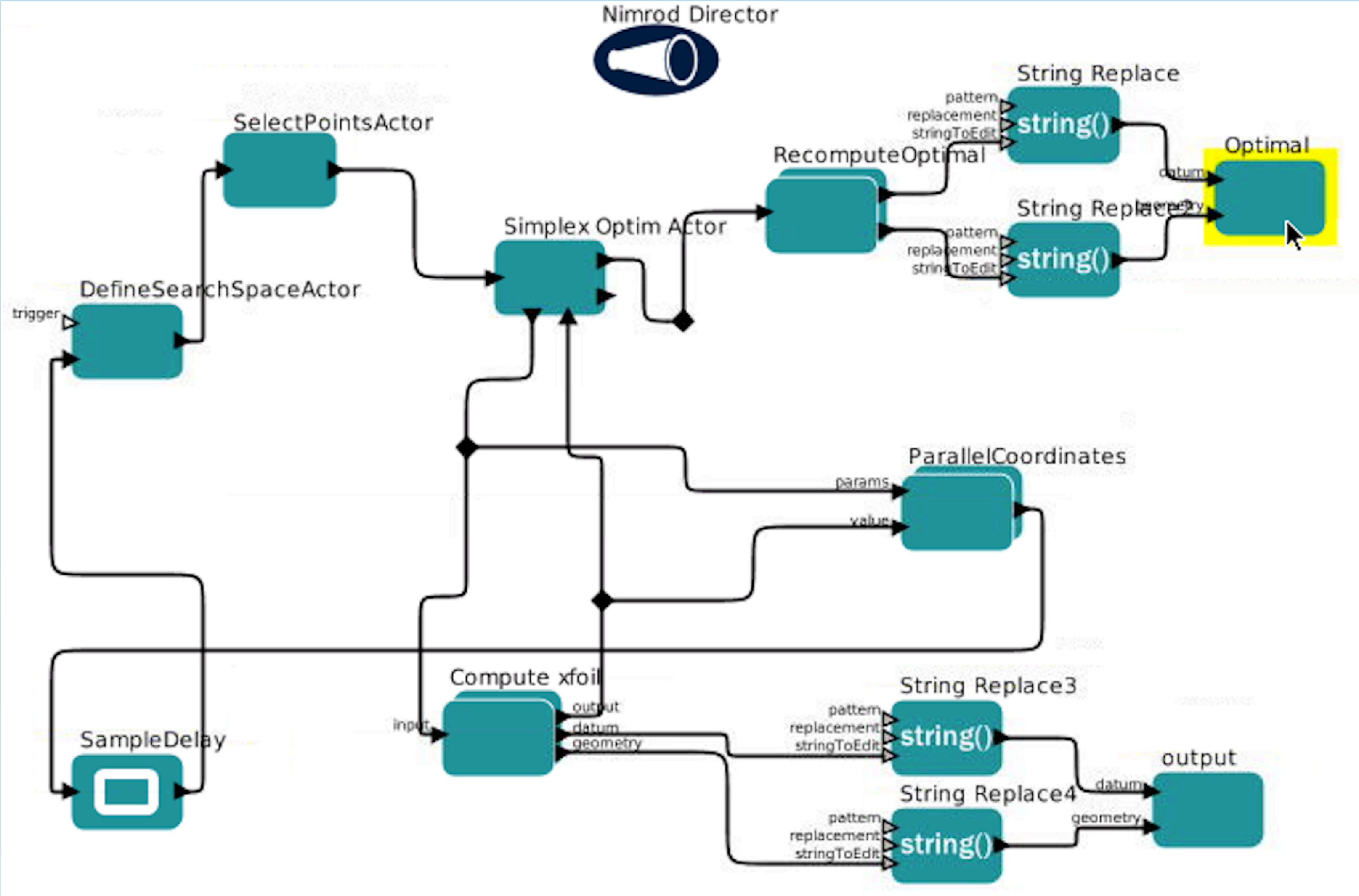
Micromixer optimization Workflow



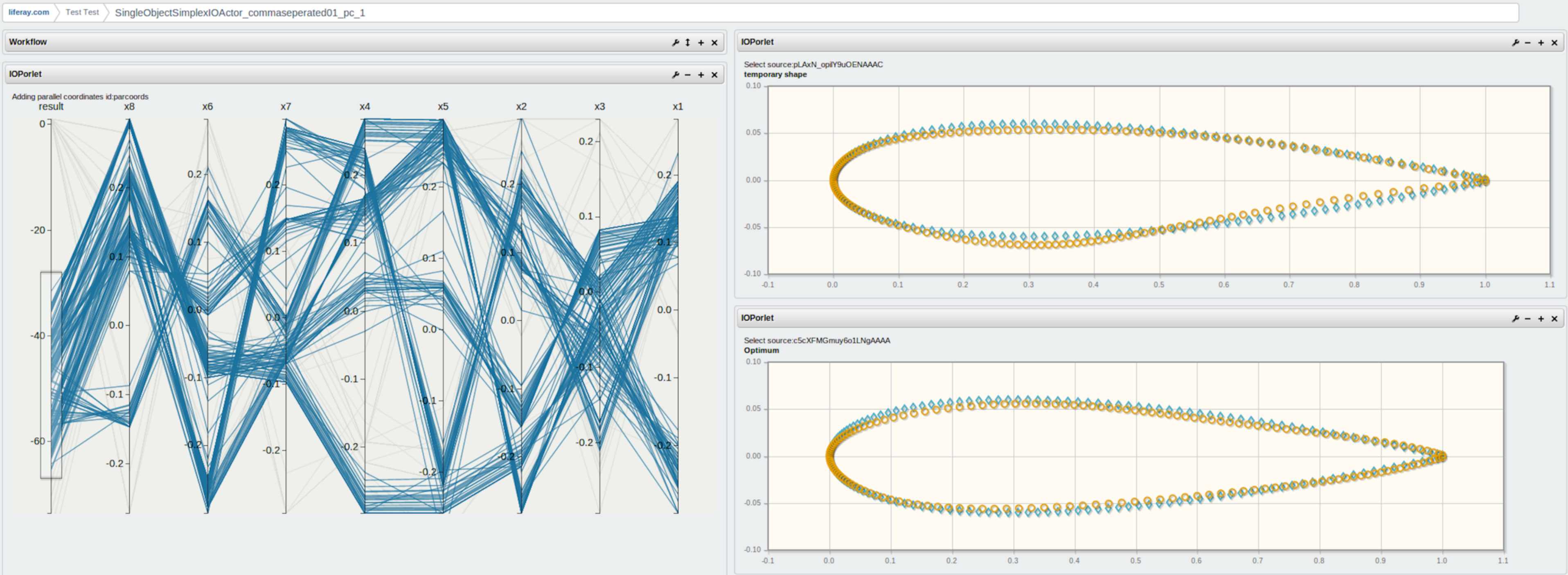
Airfoil Design



2D airfoil optimization



2D airfoil optimization



Conclusions

- Workflows are useful for scripting complex computational science and engineering problems
- Conceptually easy to add optimization
- User interaction requires new workflow actors
- Integration to a Science Gateway allows very powerful workflows to be exposed to wider communities.