HPC WALES • CYMRU

Powering Success
Pweru Llwyddiant

HPC Wales User Group Meeting

Multi-Site Video Conference

1st April 2015: 14.00 – 16.00

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop
Europe & Wales:
Investing in your future
European Regional Development Fund
ERDF
Llywodraeth Cymru
Welsh Government

Agenda

Time	Item	Speaker
14.00	Introduction	Chris Headleand
14.15	HPC Wales Update – Past, Present & Future	Martyn Guest
User Presentations		
14.30	User Presentation I	Denis Murphy
14.50	User Presentation II	Carwyn Frost
15.10	User Presentation III	Muhanad Hayder
Breakout Session		
15.30	Local User Group Discussion Session	
15.45	Feedback and Conclusion	Martyn Guest

HPC WALES • CYMRU

Powering Success
Pweru Llwyddiant

Welcome

Mr Christopher Headleand

HPC Wales User Group Meeting

Wednesday 1st April 2015

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

Overview

- **Introduction**
- **User Satisfaction Survey Analysis**
  - Initial Engagement
  - On Boarding and Training
  - Ease of Access
  - Technical Requirements, Challenges & Support
  - Usage
- **Conclusions**
- **Special Interest Groups**

User Survey Introduction

- Launched on 29th October 2014.
- 26 Primary Questions with additional follow-ups.
- 73 users out of 1722 individuals responded.
- Full analysis available, some key areas highlighted here.

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

Initial Engagement

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

How Did You Hear About HPC Wales

- Many through recommendations, direct contact and events.

- 45% through (“Other”) stakeholder Universities, via courses, bursary workshops and studentship projects. Some joined projects or businesses who were already engaged with HPC Wales.

Main Benefits of Working with HPC Wales

- Responses included: Speed, ease of use, a bursary, reliability, training, power, size, accessibility, support, cutting edge technology, ability to parallelize, efficient, further research, free, focus on improving the local economy.

How Much Initial Support Did You Require?

- Responses to this question followed the shape of a bell curve with a normal distribution. The largest proportion of respondents (42.62%) required an average amount of support. 37.7% required either a fair amount of support or not very much support and the remaining 19.68% either required a lot or no support.

- Responses to this question indicate the necessity for HPC Wales' support services and highlight the mix in HPC ability of the user community.

HPC WALES • CYMRU

Powering Success
Pweru Llwyddiant

On-Boarding & Training

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

Support Running up to a Project

- Responses to this question were positively skewed, with very few requiring no support, many requiring a little support or an average amount of support, less requiring a lot of support and very few requiring lots of support.

- Responses indicate there is demand for technical support to varying degrees across almost all of the user community once a project is up and running, regardless of existing expertise.

- More support is required to outline a project and to get a user onto a system, with a reduced amount required once a project has begun.

Time to be On-Boarded

- 45.28% of responders were on-boarded within 1 month of entering into discussions with HPC Wales. 75.47% had been on-boarded within 3 months of initially engaging with HPC Wales.
- 5.66% of respondents took over 9 months to get on to the system.
- 50.94% found it fairly or very fast. 28.4% of responders felt the time frame was fairly or very slow. No further detail exists on why delays were encountered.
- Local groups should discuss why many felt the time frame was slow and identify whether speeds can be improved / expectations managed differently to ensure a higher level of customer satisfaction.

Preferred Method of Learning

- 14.29% preferred remote learning, 26.53% preferred face to face learning and 59.18% preferred a mix of the two.

- Responses highlight the requirement for at least some element of face to face training in order to satisfy the majority of the user base.

HPC Understanding

- Results showed that responders increased understanding post engagement with HPC Wales. This is a testament to the work done by the Training and Outreach and support teams.

- This is highlighted in the results with 69% responding that they are fairly or very satisfied with the expertise of the support staff. Also 74.42% of respondents felt the training courses on offer met their needs.

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

Ease of Access

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

Rating Ease of Access

- Over 60.38% of respondents found HPC Wales' services fairly or very easy to access.
- 16.98% found them to be fairly or very difficult to access.
- Further investigation is underway to identify how these figures could be improved. This would be worth discussing with the local groups.

Use of Gateways

- 83.02% of users have not used the Gateways.16.98% found them to be fairly or very difficult to access.
- Most Popular are the Life Sciences Gateway and the AM&M Gateway.
- Do the gateways need to be promoted better? Why do you think that they are not being used? Would you use them if you were aware of them?

Synfiniway

- Only 18.7% of users have tried using it.
- 69.8% of users are not aware of it.
- Further discussion needed at local user groups.

How useful is the HPC Wales Website and Portal

- The website and portal appears to be of limited importance to responders (24.49% found it very or fairly important, 36.73% found it of average importance, 38.78 found it very or fairly unimportant.

- When asked if users could access to the information they required through the portal 32.65% answered yes and 16.33% answered no. 51.02% haven't accessed the portal.

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

Technical Requirements, Challenges & Support

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

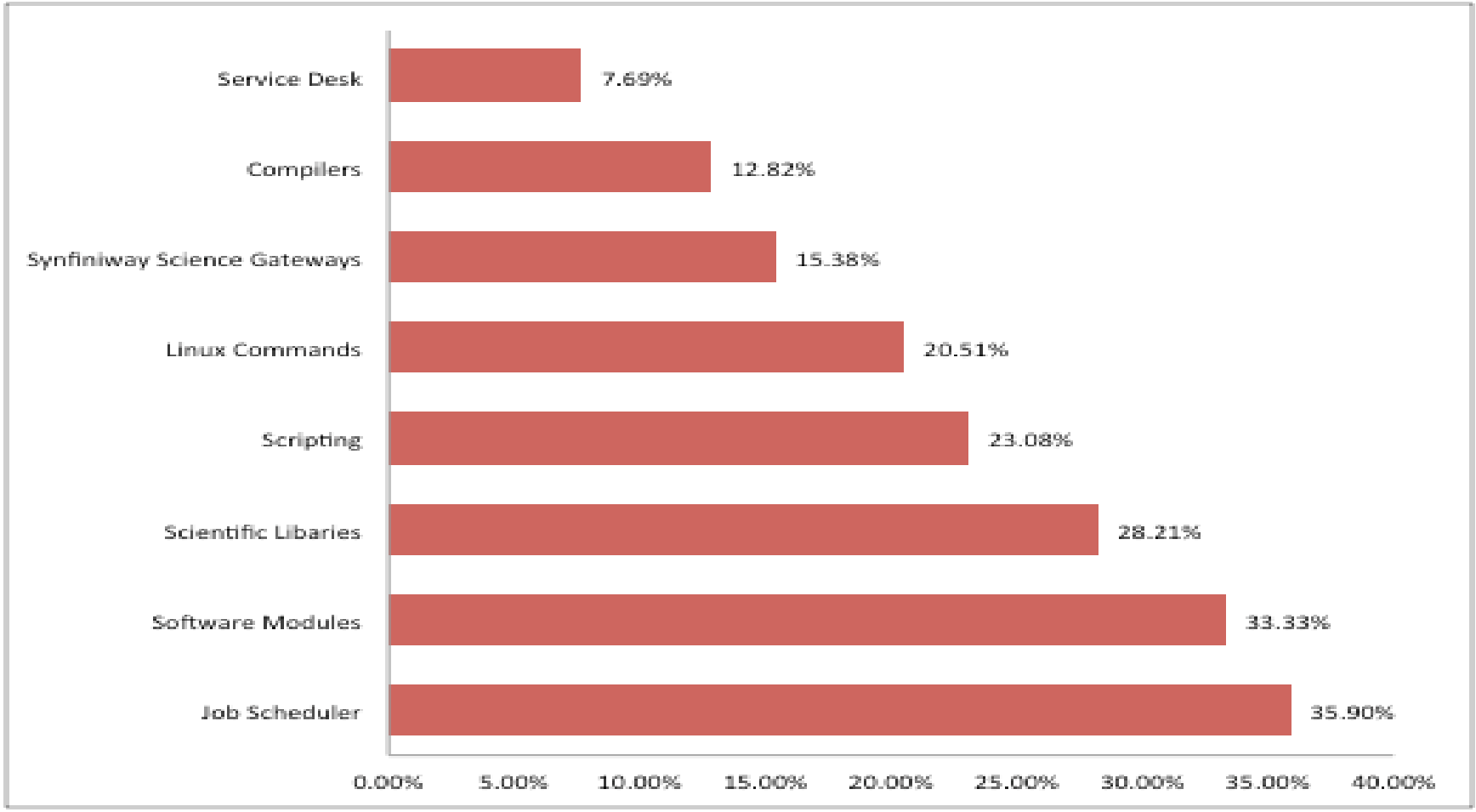
ERDF

Llywodraeth Cymru
Welsh Government

Does HPC Wales Meet Your Technical Needs

- HPC Wales is meeting the needs of many of the respondent's requirements in terms of data storage (69.81%), number of processors/power (71.70) and memory (80.77%).

Challenges


Service Desk
7.69%
Compilers
12.82%
Synfiniway Science Gateways
15.38%
Linux Commands
20.51%
Scripting
23.08%
Scientific Libaries
28.21%
Software Modules
33.33%
Job Scheduler
35.90%
0.00%
5.00%
10.00%
15.00%
20.00%
25.00%
30.00%
35.00%
40.00%

How Satisfied are you with the Support Desk & The Specialist Support Provision? (NAG and Fujitsu)

- Question did not apply to 32.6% of responders
- The vast majority of the remaining respondents were very or fairly satisfied (55.1%)
- 24.49% of respondents had received support from NAG. The overwhelming majority were very or fairly satisfied (92.67%) with the support. No respondent was unsatisfied.
- 6.12% of the user community have had FLE support. 66.66% were very or fairly satisfied. No respondent was unsatisfied with this support.

Have You Run a Job?

- Current Maximum runtime for serial jobs (72 – 240 hours) was appropriate for 41.46% and not appropriate for 29.72% of respondents. The questions was not relevant to 29.27% of responders.

- Users were asked to rank on a scale of 1 to 3, the importance of large jobs (>256 cores), development style jobs and fast turnaround for mid-range (32-128 cores) jobs. They scored 2.08, 1.48 and 2.45 respectively.

- 16% have yet to run a job

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

Conclusion

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

- Generally the overall sentiment of the survey was positive. However, worryingly few people responded to the survey. Those who responded were typically highly active members of the community which could bias the results.
- In some instances, where negative comments have been made, HPC Wales has not been in a position to improve the situation. For example, there is not a large commercial software budget to satisfy the requirements of all the user community.
- The User community has grown substantially over the last 6 months and feedback from these may not have all been captured.

- It was very encouraging to see the improvement in respondents understanding of HPC, post engagement with HPC Wales, highlighting the impact HPC Wales is having on the Welsh knowledge economy.
- The user group had been attended by almost half of respondents, and of those that hadn't attended, almost half again were interested in attending future sessions. It was pleasing to see high levels of satisfaction with regards to NAG and FLE support.
- There was a lack of awareness about Synfiniway within the user community, and it appears, a general lack of interest or motivation to try it amongst those that were aware. The website and portal need to be worked on and promoted better to be of any significant value to the user community.

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

HPC Wales Update – Past , Present and Future

Professor Martyn Guest

HPC Wales User Group Meeting

Wednesday 1st April 2015

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

Overview

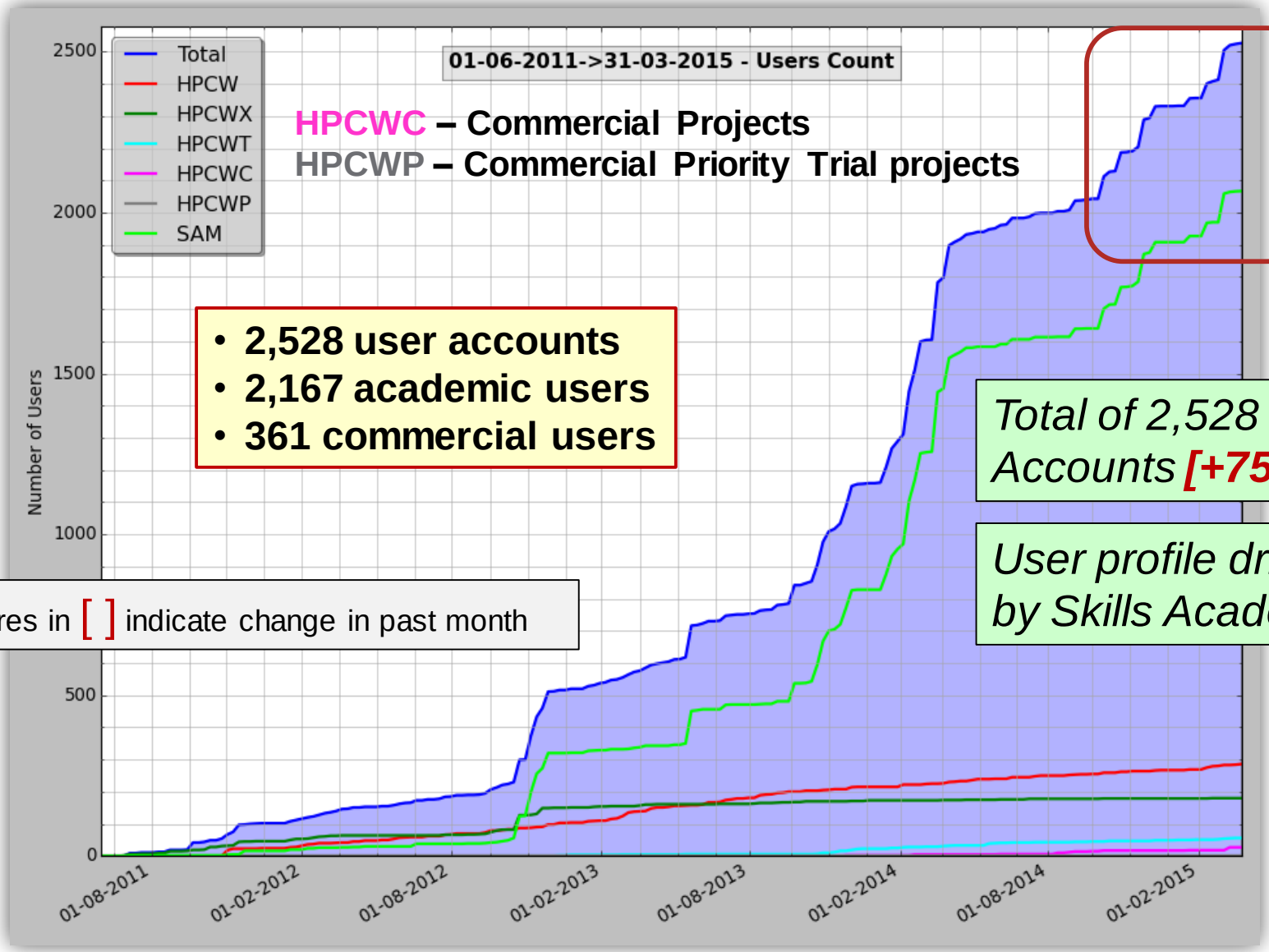
- **Update on Usage of HPC Wales Systems**
  - User Accounts, Usage Profiles, Service Desk & User Documentation
- **Measuring the Research Impact**
  - Journal Publications, Conference Proceedings, Presentations, Posters, PhDs, Prestige Factors etc.
- **HPC Wales – Next Steps**
- **Four Components of Stage 2**
  - Research and Innovation
  - SME Access
  - Training and Knowledge Transfer
  - Business Engagement
- **Technology Roadmap and Challenges**
- **Summary**

January 2012 – March 2015

OVERALL USAGE OF THE HPC WALES SYSTEMS

User Account Types

31th March 2015



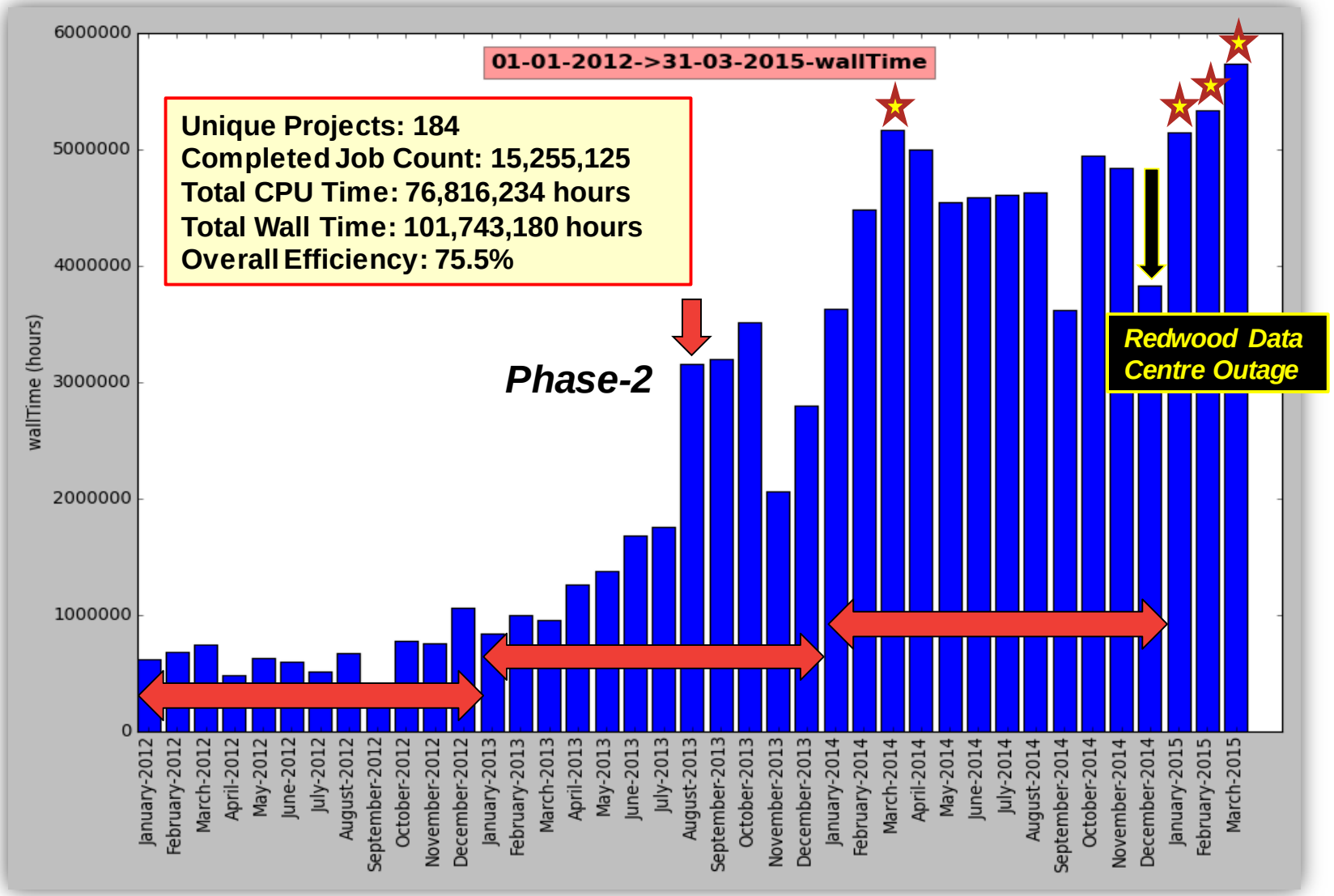
- **2,528 user accounts**
- **2,167 academic users**
- **361 commercial users**

*Total of 2,528 User Accounts **[+75]***

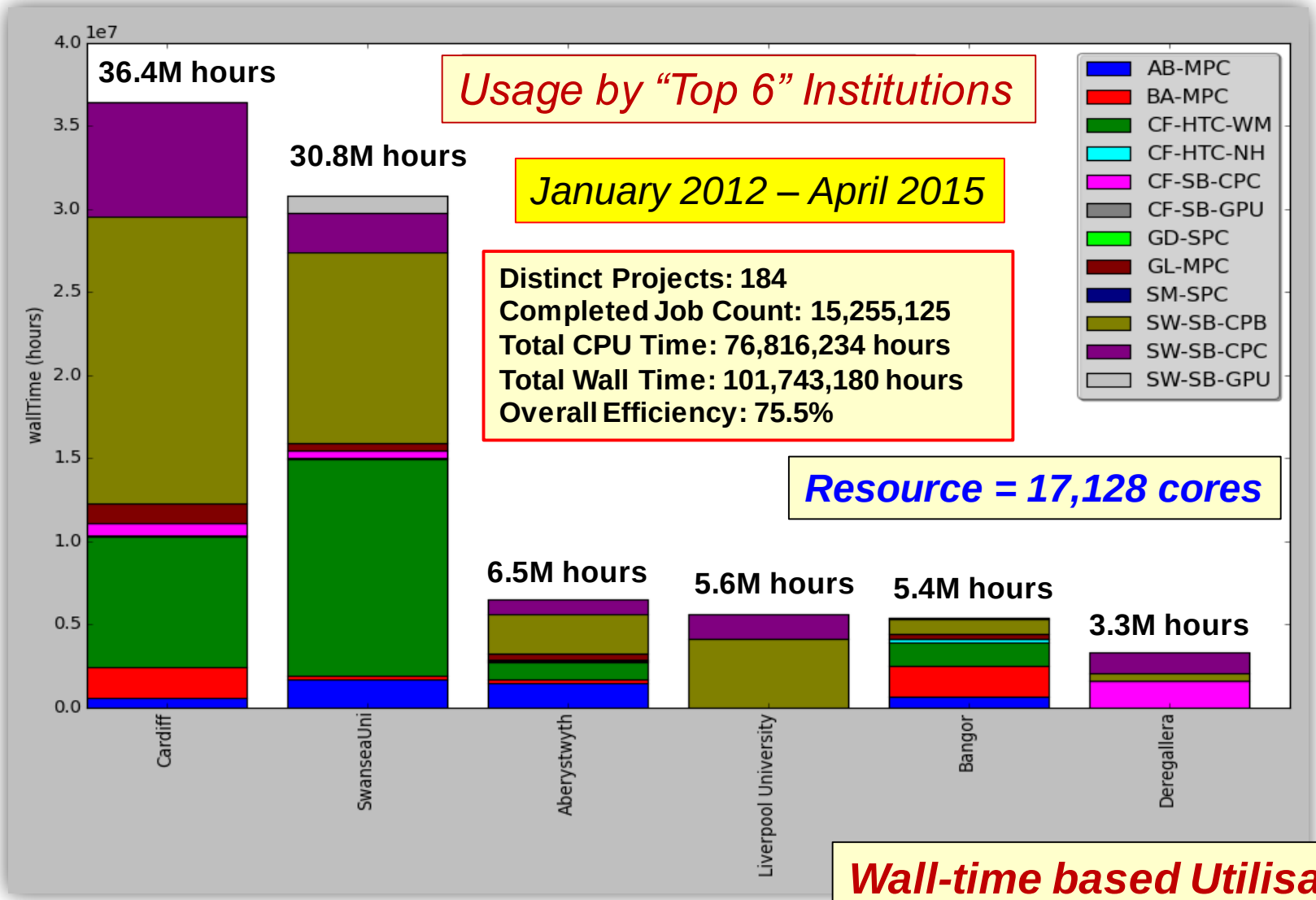
User profile driven by Skills Academy

Figures in [ ] indicate change in past month

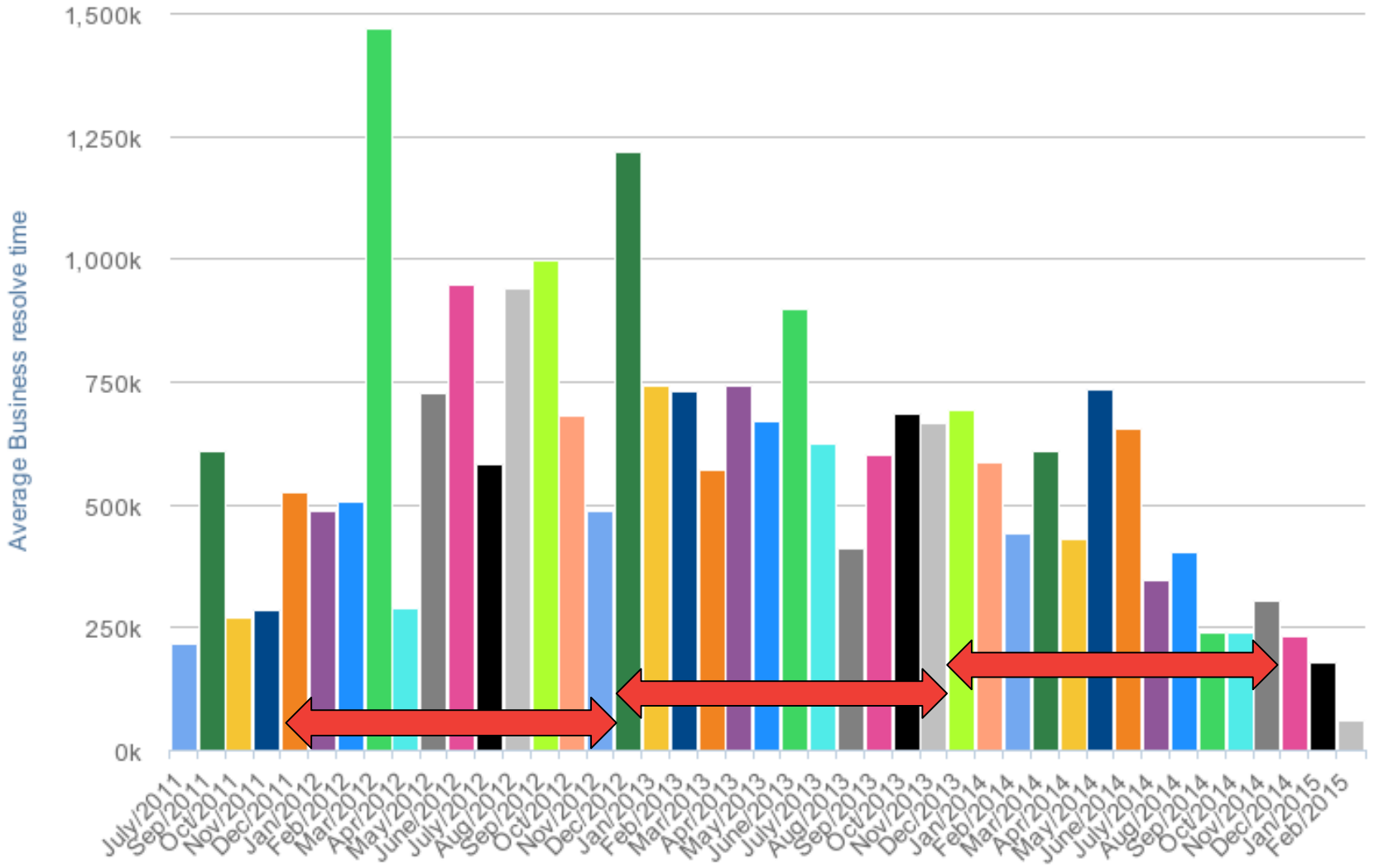
Overall Usage of the HPC Wales Systems



Overall Usage of the HPC Wales Systems



Average Resolution Time by Month



Enhanced User Documentation

Application Guides	
Tools and Languages	DDT, IPM, OSS, Scalasca, Tau, Valgrind
	Introduction to R
	MATLAB
Molecular SImulation	LAMMPS, Gromacs, DL_POLY classic, DLPOLY_4
Electronic Structure	GAMESS US, GAMESS UK, NWChem, GAUSSIAN
Materials Modelling	QuantumESPRESSO, VASP, CP2K
Energy and Environment	ROMS, SWAN, TELEMAC, WRF, Fluidity, Gerris
Engineering	OpenFOAM, ANSYS
Life Sciences & Genomics	JAGS, ABySS, BLAST, ClustalW
Creative Industries	Maya, MentalRay, Povray
Technical Notes	
GROMACS GPU Performance at HPC Wales	
Running gene sequences in parallel using I-TASSER	
Parallel Profiling of the FLITE CFD Software	
Rheological Modelling on HPC Wales	
Optimising the Performance of the TERRA software	
Papers and Presentations	
Application Performance – Westmere vs. Sandy Bridge	
A Guide to Benchmarking	

NAG, FLE and HPC Wales Technical Team

Traditional Output Metrics

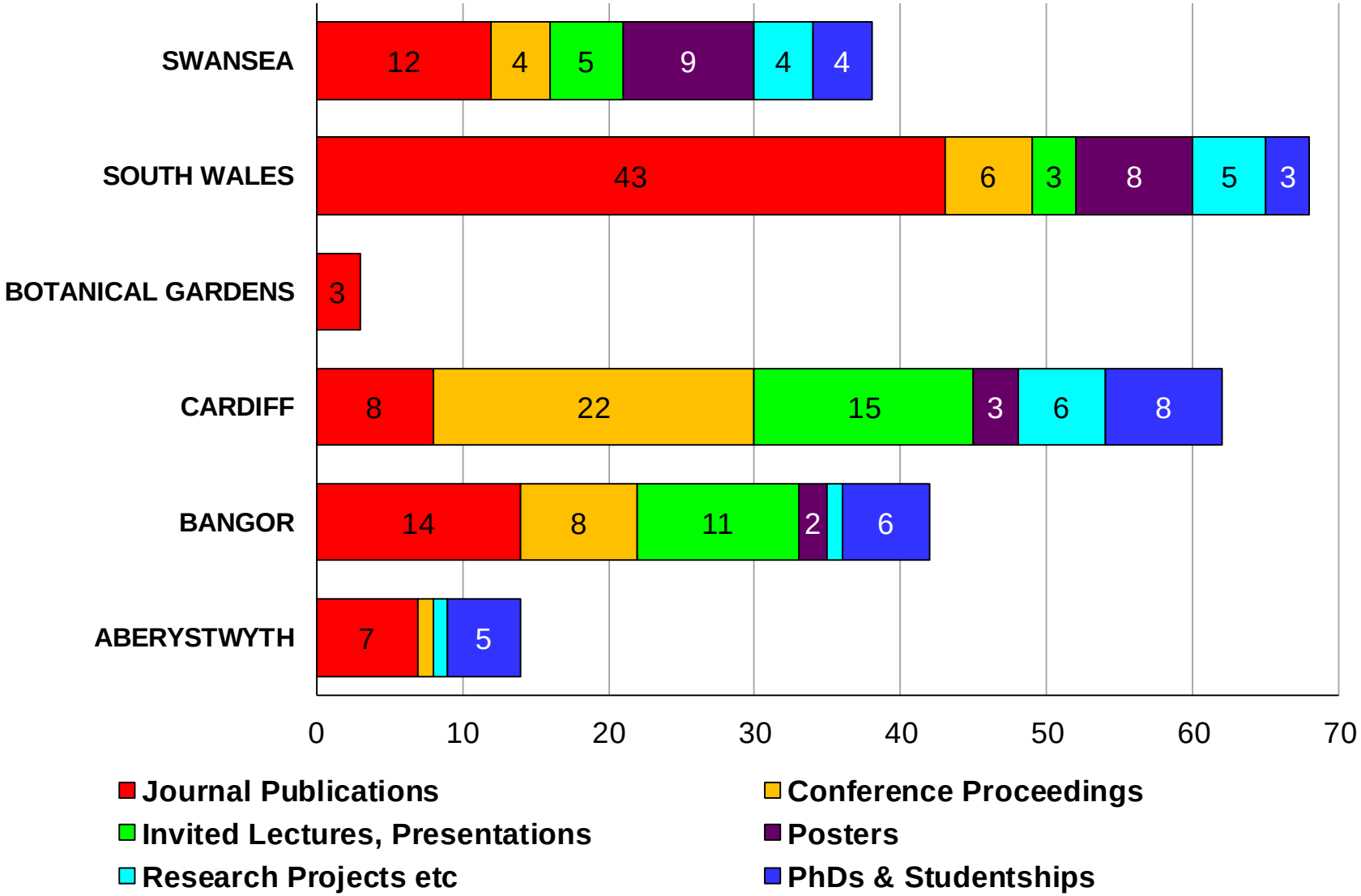
MEASURING THE RESEARCH IMPACT

Measuring the Research Impact

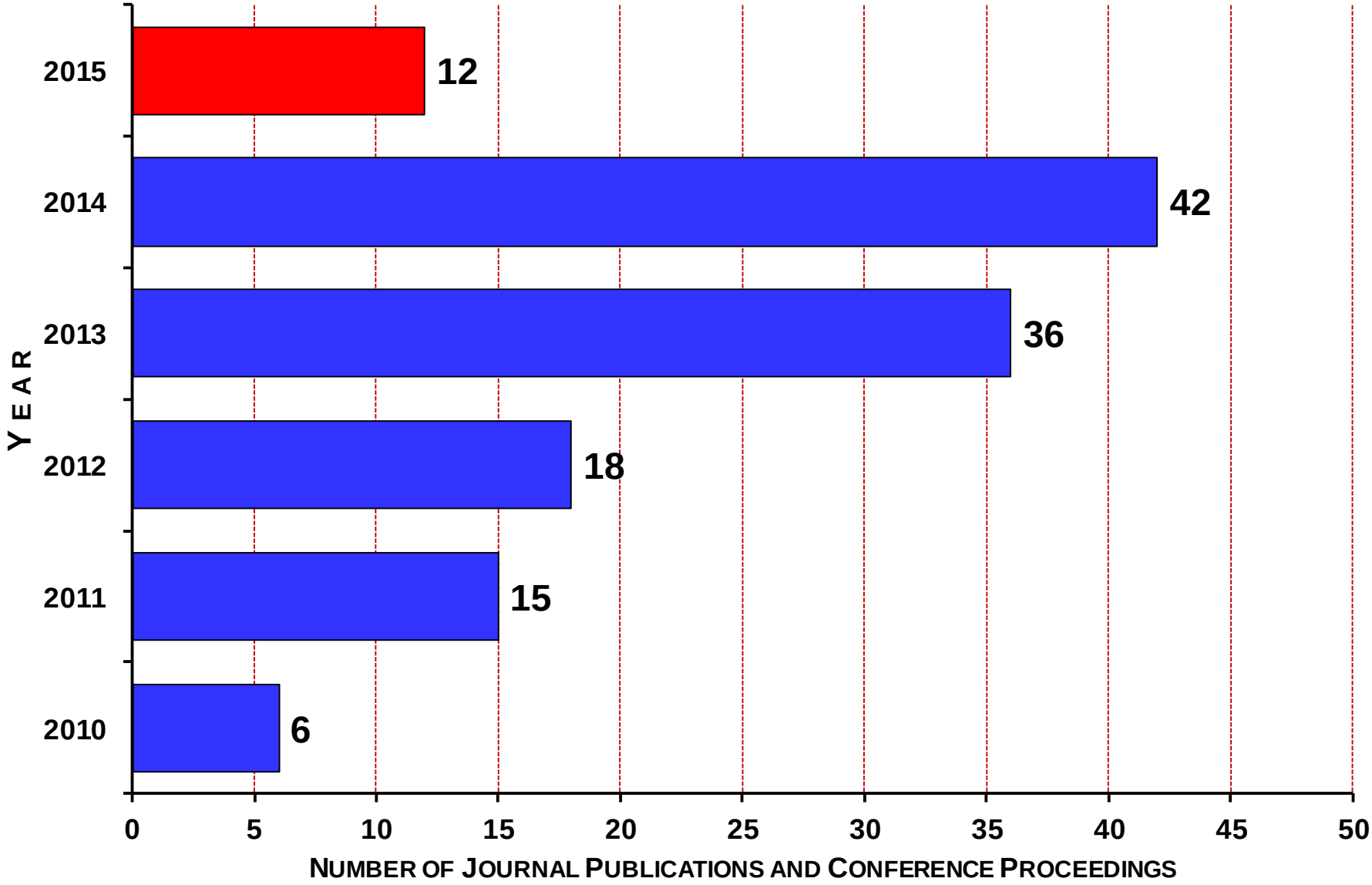
- A list of all Journal Publications, including those accepted / in press;
- A record of Conference Proceedings;
- A summary of presentations and posters given at national and international meetings/conferences, including invited lectures given at other Universities;
- Other activities and outcomes e.g., Research Projects, Application Support, Prestige Factors, Events, Reports etc. ;
- Specific research projects that have been enabled by access to the facilities; PhD studentships and Undergraduate projects.
- Posters given at national and international meetings / conferences.

…. where the associated research has at least in part been facilitated using HPC Wales resources

Number of Research Outcomes per Institution by Category



Number of Journal Publications & Conference Proceedings



The Transition to Stage 2

HPC WALES – PAST, PRESENT AND FUTURE

Background

- **HPC Wales communication of 25 March** – an update on how the venture is progressing and how our services will develop over the next five years.
- HPC Wales is now a key part of Wales' Research and Innovation (R & I) infrastructure and a major component in the delivery of a 'Digital Wales' through enhanced connectivity. The venture is already acknowledged as:
  - A major regional HPC Centre, in the UK (EPSRC[1]), EU and globally, and
  - A 'first wave' delivery partner in the UK Jisc's HPC Brokerage Service[2].

[1] *http://www.epsrc.ac.uk/research/ourportfolio/themes/researchinfrastructure/subthemes/einfrastructure/highperformance/computingcentres/*
[2] *http://www.slideshare.net/comth/introducing-the-jisc-national-hpc-framework-agreement*

An Evolving Strategy

- Our strategy for taking HPC Wales into the next phase is to **continue to provide access to supercomputing technology, training and support to existing and new users**, with the core objective of enabling and enhancing research to deliver both income and innovation to grow Wales' economy.
- Based on recent feedback from the user community, plans are underway to
  - **Increase the capacity of the HPC Wales systems** to reduce queue times & enable more jobs to be run across a larger number of cores.
  - Purchase **further software licenses and updates** are being installed to support research and innovation across a variety of disciplines.
  - Upgrade the **systems' storage capability** to meet current and emerging demands.
  - It is critical we ensure our systems have suitable **capacity and capability** to support scientific discovery and deliver the innovative solutions of the future.

We anticipate four distinct areas of business and funding areas moving into the second phase of the venture:

1. **Research and Innovation**
2. **SME Access**
3. **Up-skilling and Knowledge Transfer**
4. **Business Engagement**

1. Research and Innovation

- Extending access to **improved HPC Infrastructure services and expertise** in Wales will drive research and innovation within universities, public services and business alike.
- HPC Wales will continue to be underpinned by a collaborative agreement jointly sharing the assets, benefits and risks across the stakeholder institutions in Wales, transitioning to a position as a **Research and Innovation Institute**.
- The key objective in creating this Institute is to build upon the substantial infrastructure investment (e.g. skill base and large capital equipment) for HPC in Wales, to address future technology challenges, and to raise the overall international standing in terms of research quality and scientific impact in areas of HPC in Wales.
- It will have an impact on research in all associated departments and institutions, with the collective outcome being that Wales will ultimately become a global leader in **HPC-enabled research**.

2. SME Access

- HPC Wales will **continue to work with SMEs** to deliver economic impact and build Research, Development and Innovation collaborations very much in line with the objectives above.
- HPC Wales has been selected by the European Commission to complete a grant agreement for the EU's Horizon 2020 R&I programme to lead Europe's **Network of HPC Competence Centres for SMEs**.
- This Network will :
  - Promote access to computational facilities, pool expertise and resources across Europe and share best practice in HPC industrial use,
  - Raise awareness of the benefits of HPC and contributing to the implementation of the European HPC Strategy.

3. Up-skilling and Knowledge Transfer

- **Training and up-skilling of the work force**, along with raising HPC awareness, is essential in enabling businesses to capitalise on the benefits of the technology and contribute to economic growth.
- Maximising the benefits of HPC requires skills development and knowledge transfer to increase the capabilities of the user community.
- The success of the **HPC Wales Skills Academy** in this field now widely recognised, both within Wales and further afield.
- Objective is to build on these accomplishments and step up the delivery of high quality, high level skills training to the current user base, future academics and researchers, students, and industrial partners.
- In addition, we will continue to deliver and support programmes for young people e.g. **HPC4Schools**.

4. Business Engagement

- The fourth area of the business is focussed on developing activities more widely to **businesses and other organisations**.
- This area will draw on other technology focused programmes, such as Innovate UK, to develop new products and services that form the next generation of HPC related services, as well as directly funded activities by partner organisations.
- Critical component to deliver **sustainability of HPC Wales by the end of Stage 2**.

Funding

- The first five years of the venture have been funded through European Structural funds, UK BIS, Welsh Government (HEFCW) and University match funding.
- HPC Wales is now seeking a **second continuation round of funding from the 2014-20 European Structural Funds Programme**.
- In addition to this, and in line with the target principles, we anticipate financing the operations from a range of sources, including
  - **RCUK bids, university grants, partnership bids e.g. the Catapult Centres, TSB (now Innovate UK), and Horizon 2020.**

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

HPC Wales Stage 2

Maintaining Service & Future Technology Adoption

HPC Wales User Group Meeting

Wednesday 1st April 2015

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

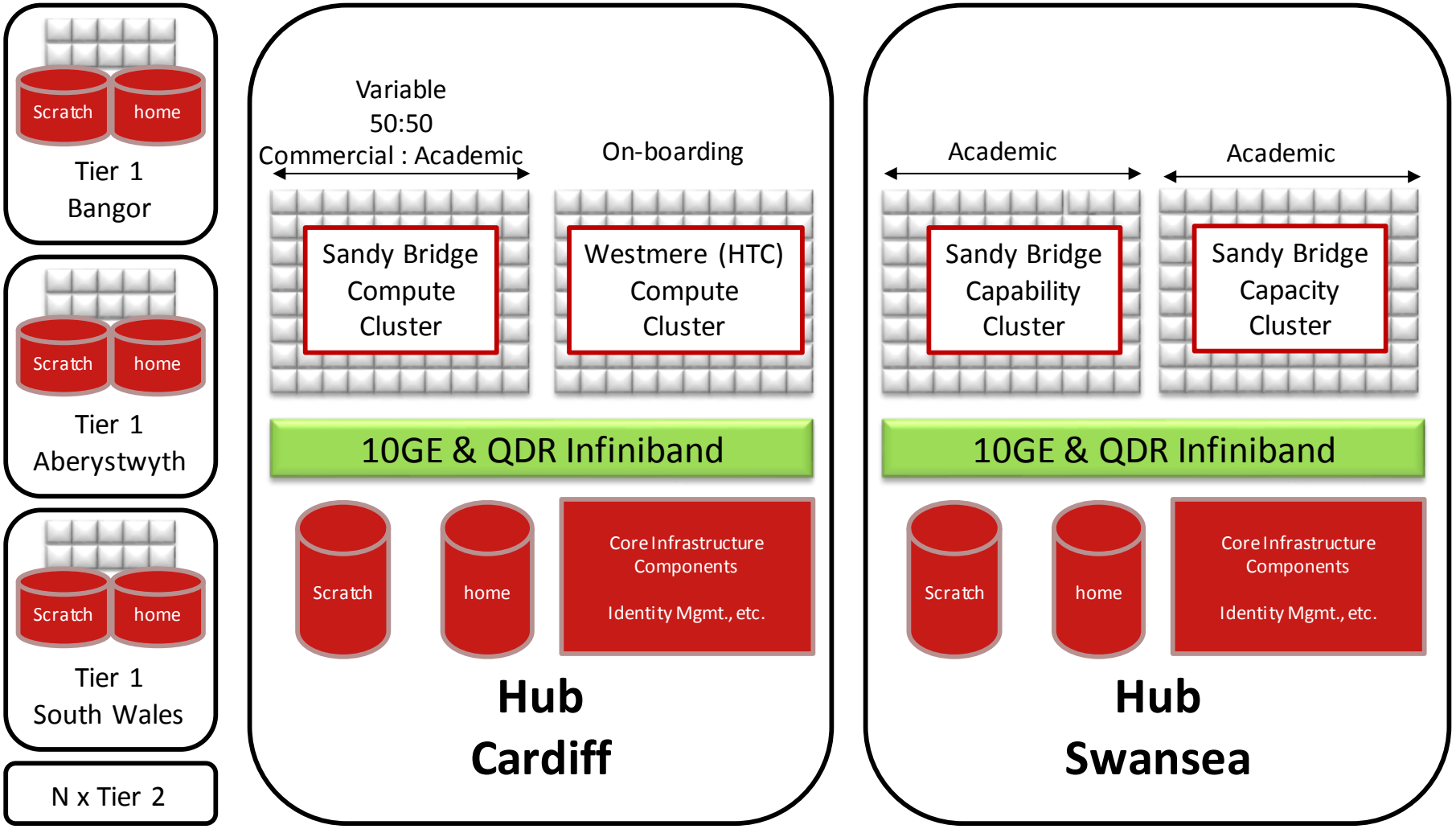
Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

Current Technology Platforms

… May 2015 – January 2016



Future Technology Challenges

- **Business as usual?**
  - We've all got used to new machines being relatively simple evolutions of our previous machines…
  - … This will **no longer be true from 2016 onwards**
- **What's changing?**
- Mainstream multi-core CPUs will continue to evolve, but more slowly…
  - Intel Ivy Bridge → Haswell → Broadwell CPUs
  - 12 core → 18 core → >20 core …
- To retain the levels of performance increase we have historically enjoyed, we will have **no choice** but to **adopt radically different architectures**

What are the options?

- **Many-core CPUs:**
- Intel Xeon Phi
  - Knights Landing (KNL) launching late 2015
  - Large KNL machines going into US national labs
  - O(70) cores, 512-bit wide vectors, HBM
- Other many-core CPUs expected to emerge
  - Based on ARM architecture
  - Multiple vendors – AMD, Broadcom, Cavium, AMCC, …

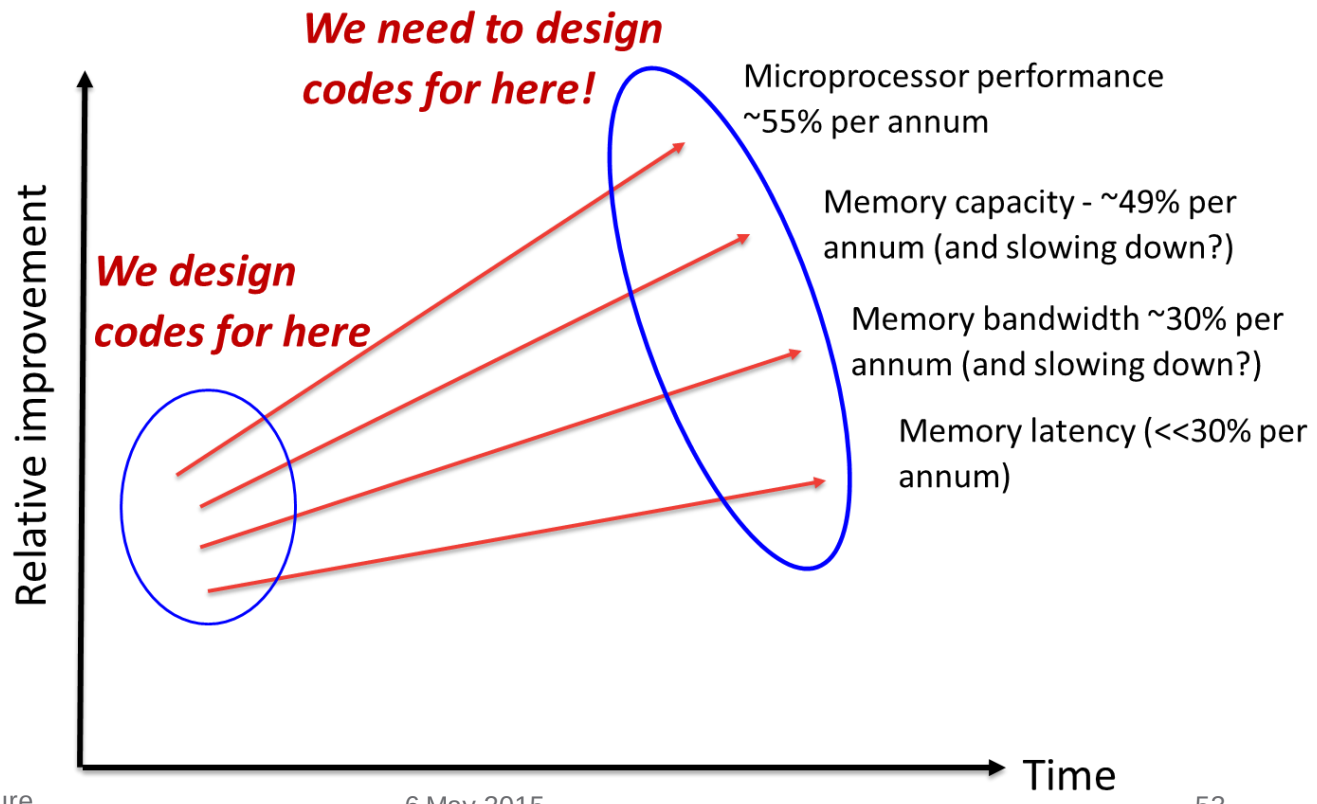
What are the options?

- **GP-GPUs:**
- Nvidia Pascal
  - Tightly couples with IBM Power9
- AMD
  - Actually have more memory bandwidth & FLOPs than Nvidia
  - Interesting focus on tight CPU/GPU integration ("fat APUs")

What other big changes are coming?

- Deeper memory hierarchies
  - Stacking (HBM), non volatile memories etc.
- Integrated interconnects
  - E.g. Intel Omni-Path fabric
- FPGAs?
  - Now supporting OpenCL…

Long-term fundamental trends



Summary

- A "***business as usual***" approach to scientific software development will result in being **left in the slow lane**
- Developers are faced with the challenging issue of developing performant code on increasingly complex architectures
- HPC Wales will need a major injection of CSE/Applications expertise

Assumptions on Usage Profile

- Starting position for Stage 2 at 30% Commercial / 70% Academic
- Commercial growth will need 50% - 50%, with larger scale system for both
- 20 projects (@ 2M core hours p.a), growing to 60 projects by Y3.
- Ease of use + Higher security for commercial usage remains an issue
- Dynamic allocation of jobs from Academic to Commercial resource
- “Application catalogue” for both commercial and non-expert users

Scratch home
Tier 1
Bangor

Scratch home
Tier 1
Aberystwyth

Scratch home
Tier 1
South Wales

N x Tier 2

Phase 1 Upgrade - Late 2016

Workload distribution to be determined
Based on evolving demand profile
Common standards to be adopted to allow dynamically flexed approach

Sandy Bridge Compute Cluster

Future Compute Cluster KNL

10GE & FDR Infiniband

Scratch

home

Updated Core Infrastructure Components

Identity Mgmt., etc.

Hub Cardiff

Workload distribution to be determined
Based on evolving demand profile
Common standards to be adopted to allow dynamically flexed approach

Sandy Bridge Compute Cluster

Acceleration Cluster GPU and/or Phi

10GE & FDR Infiniband

Scratch

home

Updated Core Infrastructure Components

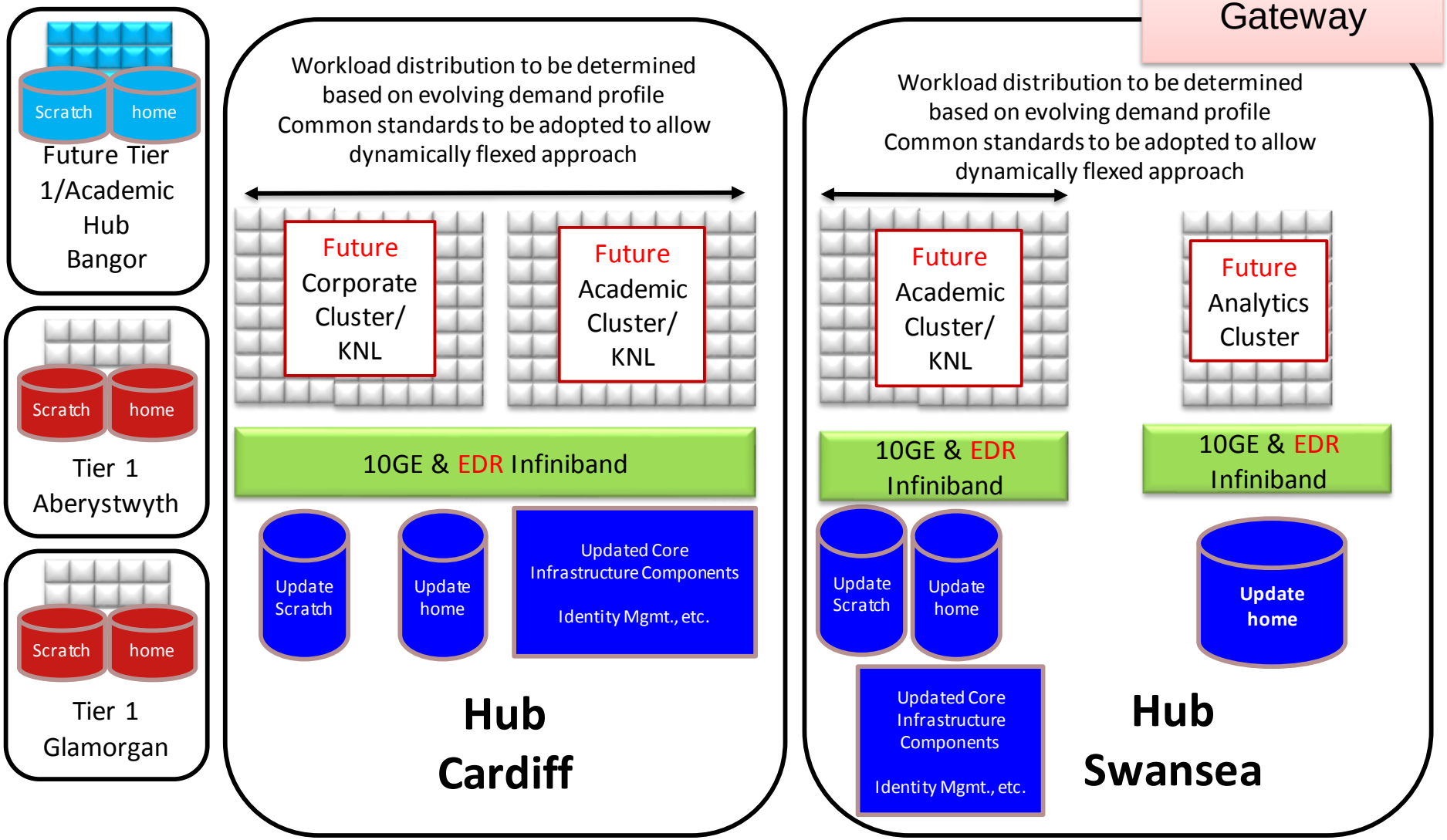
Identity Mgmt., etc.

Hub Swansea

Phase 2 Upgrade - 2018

Application Catalogue and Portal

Health research Gateway



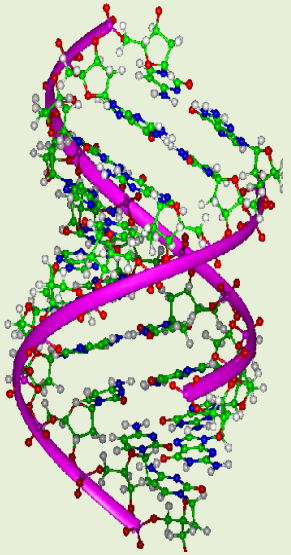
Summary

- **Update on Usage of HPC Wales Systems**
  - User Accounts, Usage Profiles, Service Desk & User Documentation
- **Measuring the Research Impact**
  - Journal Publications, Conference Proceedings, Presentations, Posters, PhDs, Prestige Factors etc.
- **HPC Wales – HPC Wales – Next Steps**
- **Four Components of Stage 2**
  - Research and Innovation
  - SME Access
  - Training and Knowledge Transfer
  - Business Engagement
- Technology Roadmap and Challenges
- **Summary**

University of South Wales
Prifysgol De Cymru

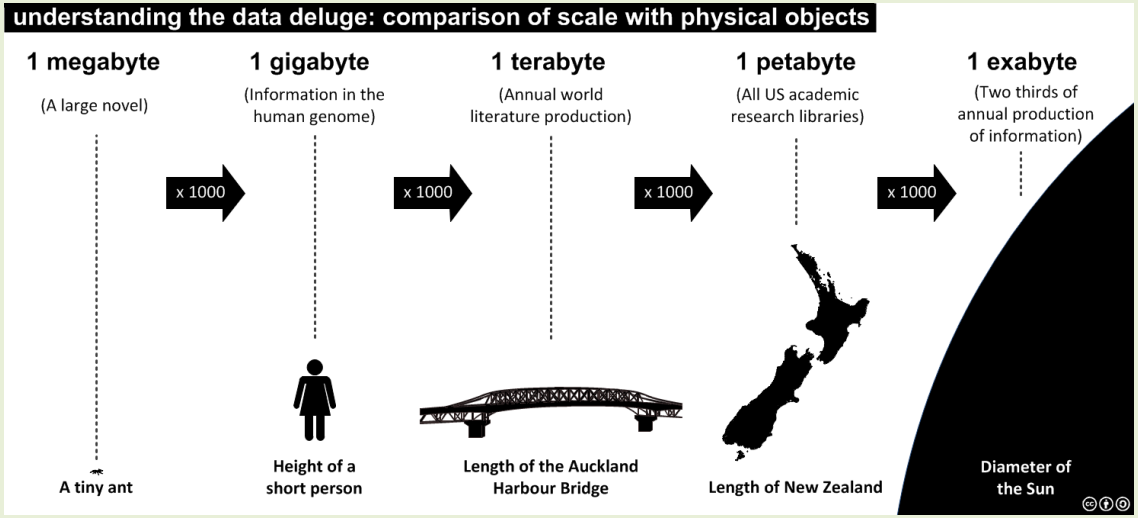
Developing Improved Bioinformatic Pipelines for Genome Analysis and Practical Applications in Crop Breeding – a big data problem

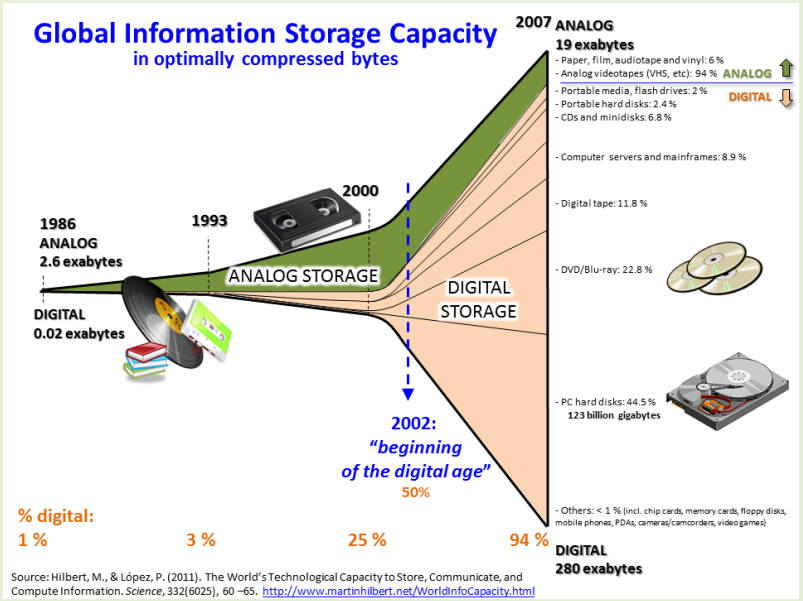
Denis J Murphy*
Head of Genomics & Computational Biology
University of South Wales, UK



* Biotechnology Advisor, United Nations Food & Agriculture Organization
* Chair, Biology Advisory Committee, Malaysian Palm Oil Board
* Consultant on crop breeding to European Union Joint Research Council

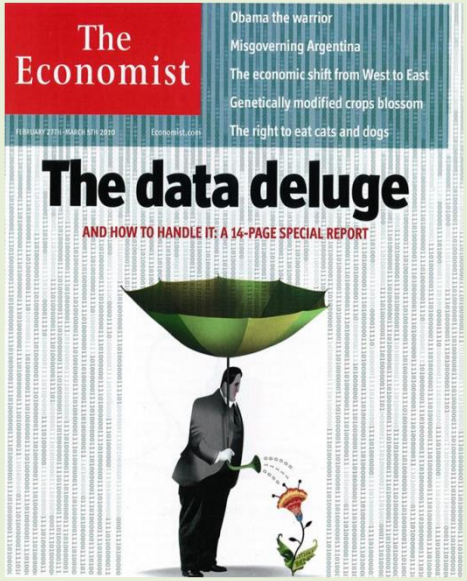
Data generation has increased exponentially over the past decade

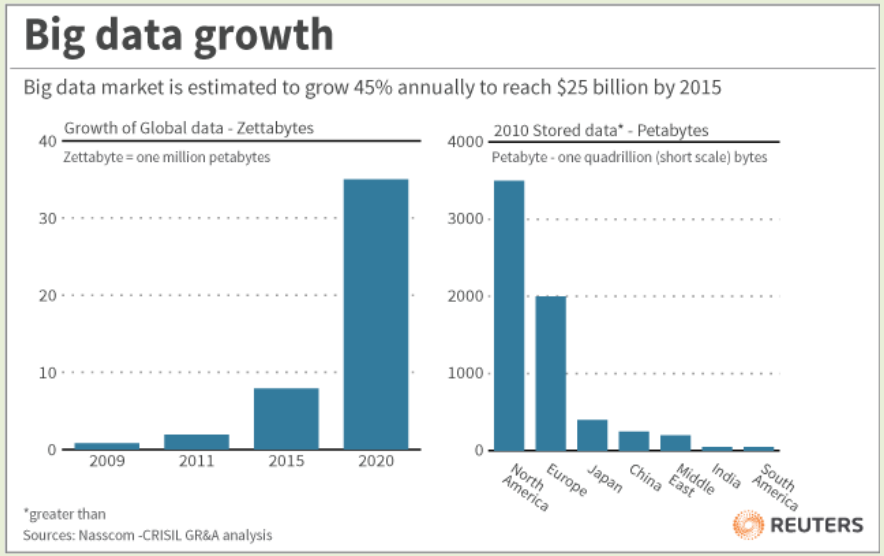




The 'big data' problem

- In recent years, advances in computing and information technology have led to the so-called 'big data' problem.
- This involves the creation of massive quantities of data in many fields including medicine, agriculture, commerce, communications, state security and basic science.
- The analysis of big data and its reduction to meaningful knowledge is a major challenge in today's information-rich society.



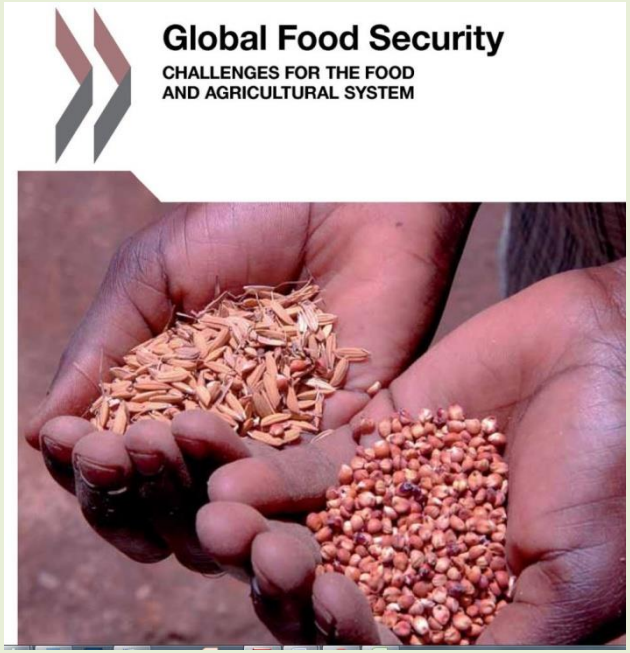


Reuters graphic/Catherine Trevethan 05/10/12

The 'big data' problem

- **Bioinformatics involves the use of high performance computational systems and novel mathematical tools to make sense of big data in fields related to bioscience.**

- **Our work is particularly related to the use of genomic data analysis in two areas, namely the emerging field of personalised medicine to address health concerns and in the improvement of key crops to address global food security**



Bioinformatics and the DIKW hierarchy

- By developing tools such as algorithms and computational pipelines bioinformatics allows us to transform raw data into genuine information
- We then need to use our biological expertise to transform this information into useful knowledge
- The final process, the transformation of knowledge into true wisdom is something that scientists need to do in collaboration with society as a whole

For example, we are now on the verge of being able to create synthetic life forms – but is this really a wise thing to do?

Scope & Application of our research

- Genomics
- Biotechnology
- Computational biology
- Bioinformatics
- Database design and mining
- User interfaces
- Network modeling
- Software analysis/development including algorithm & pipeline design

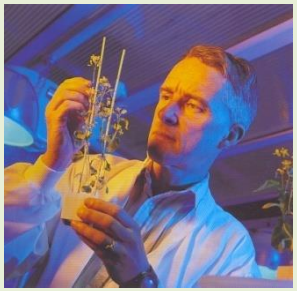
- Medicine – eg cancers & bacterial infections
- **Agriculture – eg tropical crop improvement**
- Business – eg commercial ancestry prediction

Our work is based in the Welsh Valleys but is very much part of an international network of collaborators

- Prof Alexandar Bolshoy
  (University of Haifa)
- Prof Tatiana Tatarinova
  (
- Prof. Eran Elhaik
  (University of Sheffield & JHU)
- Prof. John Harwood
  (University of Cardiff)
- Dr. Arunachalam Vadivel
  (Indian Council of Agricultural Research)
- Dr. Natasha de Vere
  (National Botanic Garden of Wales)
- Dr. Nickolai Alexandrov
  (International Rice Research Institute)
- HPC Wales consortium
- Fujitsu Laboratories Europe Limited
- Malaysian Palm Oil Board
- National Botanic Gardens Wales

Our Team at Univ South Wales

- the Computational Biology Group



Dmitriy Sonkin

Negusse Kitaba

Mehedi Hassan

Rozana Rosli

Hannah Garbett

Farzana Rahman

Martin Triska

University of South Wales
Prifysgol De Cymru

Assessment of the Database of Clusters of Orthologous Genes (COGs)

A step toward bacterial toxicity prediction

Farzana Rahman[1], M. Hassan[1], T .Tatarinova[2], A. Bolshoy[3] and D. Murphy[1]

Our work can be seen online at **http://mendel.fat.glam.ac.uk/projects/bpiCOG**

Background

Rapid emergence of drug-resistant bacteria became a global issue over last four decades. This is generating speculations like 'end of the antibiotic era', 'crisis of modern medicine' and giving way for application of the chaos theory to medicine.

Bacteria are fast growing, rapidly evolving organism which exist everywhere. There are approx 10 times more bacterial cells than human cells in our body.. Mostly harmless, bacterial strains can become toxic and lethal within a short time.

Therefore, using genomics we are on a mission to describe rules of bacterial evolution and thus aid medical science in bringing about better medicine and medical techniques to understand and control bacteria.

Project Aim

We aim to **re-classify** bacterial strains based on their toxicity, find **pattern** and **association** between **genomic** characteristic and **toxicity**, devise a method to quickly sample and identify pathogens in wounds, food, bio-materials by use of computational methods.

Objectives

- Study genomic properties of sequenced bacteria.
- Recognise pattern in association between genetic makeup and toxicity.
- Use the clusters to build a family tree for bacteria based on toxicity.
- Explore the families to fine tune the logic.

Entero Haemorrhagic *Escherichia coli* (EHEC)

Artist's impression on *Salmonella typhimurium*

What is a COG?

What it has to do with my work?

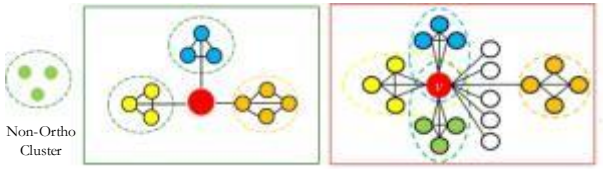
When a new gene and/or associated proteins is discovered, a comprehensive database of gene family is used to extract maximum amount of information from existing and rapidly growing new genome collections.

This database is called "**C**lusters of **O**rthologous **G**enes".

This database was developed to classify all conserved genes as per their homologous relationship. As described by Tatusov *et al.,* COGs database is a growing and useful resource to identify genes & groups of orthologs in different species related by evolution (Tatusov, Koonin *et al.*, 1997).

Despite the principles, in recent years non-Orthologous groups were discovered within the COGs database (Dessimoz, Boeckmann *et al.*, 2006).

At the outset of my project, we reviewed literature covering recent findings of non-orthologous groups. This led us to assess the database and find a way of curation.



Ideal COGs Database

COGs Database with Erroneously merged Clusters of non-orthologous protein (green)

Structure of COGs Database

Current COG database has 4,699 clusters, containing 56,666 genes with 4,959,417 proteins in 2784 sequences from over 1,400 genomes providing 250,824 products. Proteins within same COG are supposed to present similar characteristics. Since 2011, we have been developing algorithms and procedures to handle these data to check whether there is any discrepancy in the dataset.

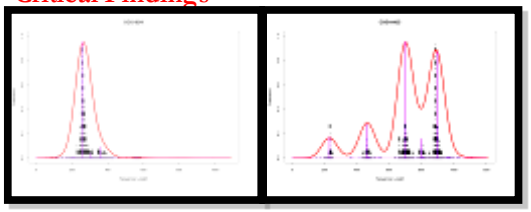
Assessment Framework

We examined the database for genes within same COG for their lengths. We have recently optimised
ranking of bacterial genomes using Kendall Tau rank-correlation coefficient.

Using Nonparametric Bayesian Approach & Mixture models, we detected **erroneously merged clusters** in the database. We developed a model and applied bayesian inference using gibbs sampling to find parameters of this model for each cluster.

Programmes on HPC Wales supercomputer, made possible, efficient MCMC computation in feasible timescale. For each COG we collected lengths of all proteins contained in the COG to find model parameters.

Critical Findings



Unimodal distribution of lengths in a good COG, as expected.

Multimodal distribution by length, in a bad COG, COG1042C, which contains a number of protein families from *Bordetella* sp.

Our algorithm, implemented in R, showed distribution of gene lengths for **64%** COG as **unimodal distribution** (as expected). However, there were, **36%** COGs contained **multimodal distribution,** which is unexpected. These COGs where erroneously merged clusters existed (like the illustration on left). We spotted, in one such COG1042C, high volume of proteins were assigned belonging to bacteria genuses known for toxicity (i.e. *Bordetella bronchiseptica*).

Also within some healthy COGs, we have found outliers, belonging to the genuses known forstrains with volatile toxic nature. We found six genomes contributing to the outliers, these are *Bacillus coagulans* 36D1, *Bacillus clausii* KSM-K16,*Bacillus cellulosilyticus* DSM 2522, *Bacillus selenitireducens* MLS10, *Bacillus halodurans* C-125, *Bacillus pseudofirmus*. Two member of this lethal to us, *B. anthracis* for causing anthrax and *B. cereus* causing food poisoning.

What's next???

With the observations and fault detection completed, I am now working on the reasoning of the error and looking forward to work closely with the developers of the database for its curation.

The Role of Supercomputer

As much as it is dependent on efficient algorithms, computational biology research, heavily relies upon computational power. Over a period of 18 months, we used well above 40 thousand hours of computational power on High Performance Computing (HPC) Wales's distributed supercomputing clusters.

Acknowledgements

This work has been supported by first studentship award, sponsored by *HPC Wales & Fujitsu Laboratories Europe Limited.* I would like to thank them for the funding.

HPC WALES - CYMRU

FUJITSU

Author Affiliations: (1) University of South Wales, Pontypridd. (2) University of Southern California, USA (3) University of Haifa, Israel

This poster was presented at the Annual Post Graduate Research Day 2013, at the University of South Wales, Pontypridd

Development of an integrated bioinformatics platform with methods for Translation Start Site (TSS) prediction, DNA motif detection and improved gene annotation.

The work is partially funded by HPC Wales & Fujitsu and started in autumn 2012

The major targets are:

- To develop novel tools for precise TSS identification
- Using the TSS identification tool, build an exhaustive database of the gene promoters

- To continue development of the DNA motif prediction method – MotifExpresser. Implement new ideas and make use of new findings. One of the challenges will be detection of the aggregates of motifs, which work together.
- Using the improved motif detection method, create exhaustive database of the DNA motifs and their aggregates.
- To improve annotation of GC_3-rich genes using functional and compositional properties.

- **Create a complex genomics platform, that would allow scientists worldwide to:**

1) access and enrich the database of promoters, DNA motifs & expression data

2) analyse their data using our tools

3) provide their data and results to the scientific community

Mehedi/Rozana Projects

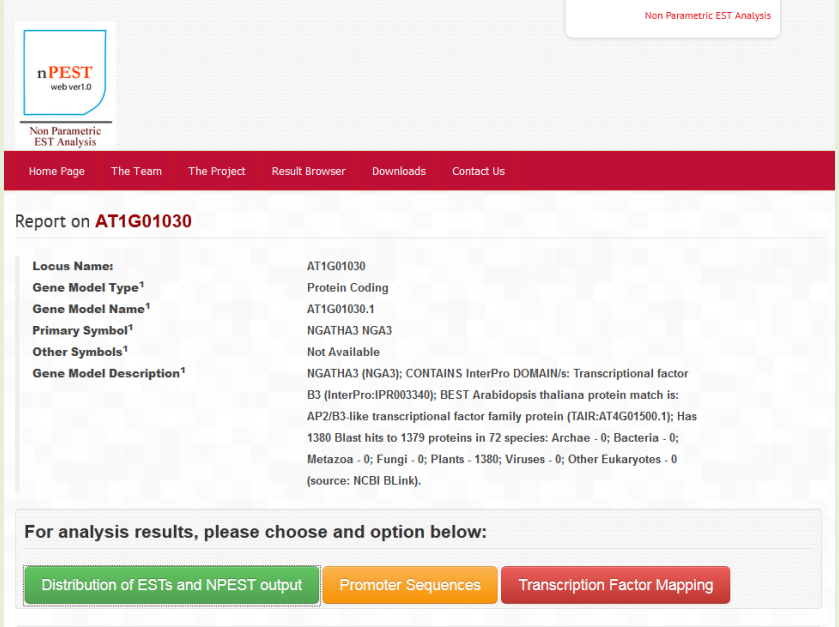
- NPEST – Non Parametric EST Mapping
- Published in ***J Quantitative Biology,*** Jan 2014
  - A collaborative project to develop a novel tool for the analysis of Expressed Sequence Tags (EST) distributions and Transcription Start Site (TSS) prediction.
  - Studied model plant genome of *Arabidopsis thaliana*
  - Using our statistical tool we analyzed 16,520 loci and developed a database of TSS, which is now publicly available at www.glacombio.net/NPEST.
  - Modeled, designed and constructed the database.

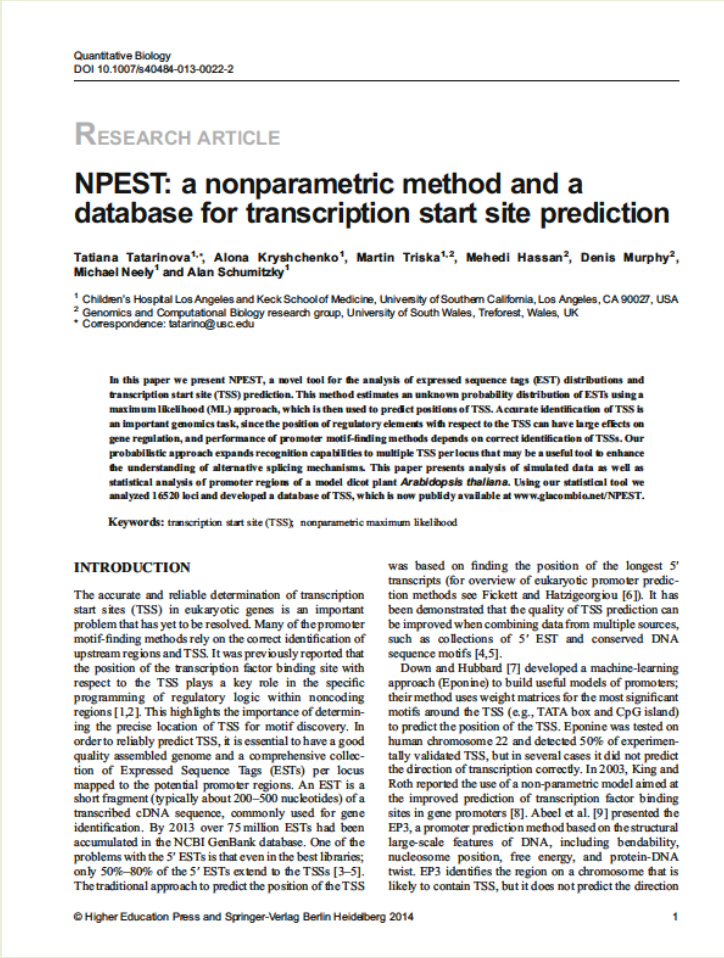
Currently expanding research into other species (oil palm)

- Rozana Rosli started new PhD project funded by Malaysian Govt in 2015 that includes use of HPCW facilities
- Mehedi funded by MPOB to work on oil palm genomics and database/portal design

NPEST – web portal

- Web implementation using PHP, R, Perl, Oracle and ImageMagick
- Received >4,000 page views since launch in 2013





Quantitative Biology
DOI 10.1007/s40484-013-0022-2

RESEARCH ARTICLE

NPEST: a nonparametric method and a database for transcription start site prediction

Tatiana Tatarinova[1,*], Alona Kryshchenko[1], Martin Triska[1,2], Mehedi Hassan[2], Denis Murphy[2], Michael Neely[1] and Alan Schumitzky[1]

[1] Children's Hospital Los Angeles and Keck School of Medicine, University of Southern California, Los Angeles, CA 90027, USA
[2] Genomics and Computational Biology research group, University of South Wales, Treforest, Wales, UK
* Correspondence: tatarino@usc.edu

In this paper we present NPEST, a novel tool for the analysis of expressed sequence tags (EST) distributions and transcription start site (TSS) prediction. This method estimates an unknown probability distribution of ESTs using a maximum likelihood (ML) approach, which is then used to predict positions of TSS. Accurate identification of TSS is an important genomics task, since the position of regulatory elements with respect to the TSS can have large effects on gene regulation, and performance of promoter motif-finding methods depends on correct identification of TSSs. Our probabilistic approach expands recognition capabilities to multiple TSS per locus that may be a useful tool to enhance the understanding of alternative splicing mechanisms. This paper presents analysis of simulated data as well as statistical analysis of promoter regions of a model dicot plant *Arabidopsis thaliana*. Using our statistical tool we analyzed 16520 loci and developed a database of TSS, which is now publicly available at www.glacombio.net/NPEST.

Keywords: transcription start site (TSS); nonparametric maximum likelihood

INTRODUCTION

The accurate and reliable determination of transcription start sites (TSS) in eukaryotic genes is an important problem that has yet to be resolved. Many of the promoter motif-finding methods rely on the correct identification of upstream regions and TSS. It was previously reported that the position of the transcription factor binding site with respect to the TSS plays a key role in the specific programming of regulatory logic within noncoding regions [1,2]. This highlights the importance of determining the precise location of TSS for motif discovery. In order to reliably predict TSS, it is essential to have a good quality assembled genome and a comprehensive collection of Expressed Sequence Tags (ESTs) per locus mapped to the potential promoter regions. An EST is a short fragment (typically about 200–500 nucleotides) of a transcribed cDNA sequence, commonly used for gene identification. By 2013 over 75 million ESTs had been accumulated in the NCBI GenBank database. One of the problems with the 5′ ESTs is that even in the best libraries; only 50%–80% of the 5′ ESTs extend to the TSSs [3–5]. The traditional approach to predict the position of the TSS was based on finding the position of the longest 5′ transcripts (for overview of eukaryotic promoter prediction methods see Fickett and Hatzigeorgiou [6]). It has been demonstrated that the quality of TSS prediction can be improved when combining data from multiple sources, such as collections of 5′ EST and conserved DNA sequence motifs [4,5].

Down and Hubbard [7] developed a machine-learning approach (Eponine) to build useful models of promoters; their method uses weight matrices for the most significant motifs around the TSS (e.g., TATA box and CpG island) to predict the position of the TSS. Eponine was tested on human chromosome 22 and detected 50% of experimentally validated TSS, but in several cases it did not predict the direction of transcription correctly. In 2003, King and Roth reported the use of a non-parametric model aimed at the improved prediction of transcription factor binding sites in gene promoters [8]. Abeel et al. [9] presented the EP3, a promoter prediction method based on the structural large-scale features of DNA, including bendability, nucleosome position, free energy, and protein-DNA twist. EP3 identifies the region on a chromosome that is likely to contain TSS, but it does not predict the direction

Presented at 22nd ISMB and 10th ISCB Student Council Symposium, 11 -15 July 2014, Boston, USA

BarcAlign: An improved computational pipeline for the alignment of DNA barcode sequences

University of South Wales
Prifysgol De Cymru

Author: **Hannah Garbett**; T V Tatarinova; N de Vere; D J Murphy

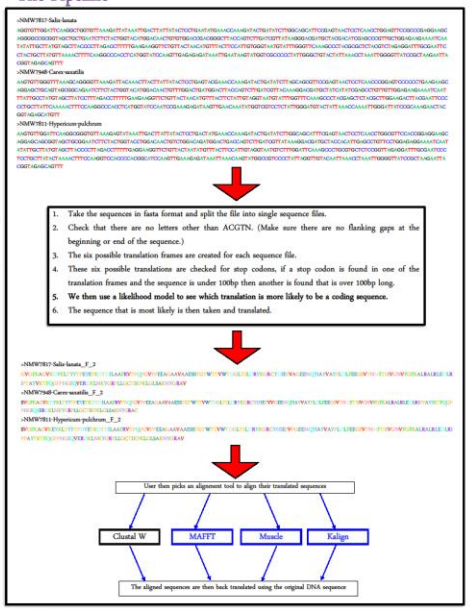
What is a DNA Barcode?

A DNA Barcode is a short DNA sequence; usually under 1000bp, that can be extracted from an organism cheaply and quickly. It is used for species identification (Herbert, et al., 2003). The animal DNA barcode region is the gene encoding **C**ytochrome **O**xidase subunit **1**, (**CO1**), that is found in the mitochondrial genome. The DNA barcode regions for plants are the *rbcL* and *matK* gene regions, which are located within the chloroplast genome (CBOL Plant Working Group, 2009). It is thought that, thanks to DNA barcoding, one day a hand-held device could be used to identify any species of plant from a DNA-containing tissue sample such as seed, root, leaf or pollen.

Aligning DNA Barcodes.

The first step of analyzing DNA barcode data is to accurately align the sequences that have been extracted. This is a crucial step in the analysis as if there is an error here then incorrect conclusions could be drawn from results that follow this step. When doing extensive research on alignment tools so that the *rbcL* and *matK* sequences could be accurately aligned, there was no problem when it came to aligning the *rbcL* sequences. However, a significant problem occurred when trying to align the *matK* sequences, because the *matK gene* is very diverse between different species (Lahaye, et al., 2008). We found that there was one tool that gave a better alignment that any that had been tried to date, this was called *transAlign* (Bininda-Emonds, 2005), but the results from this were not optimal for this type of data. Therefore we took the idea of *transAlign* and made some improvements to the method. One of the improvements made is the way in which the **O**pen **R**eading **F**rames (**ORF**s) are found, here we used likelihood models to calculate the likelihood of each possible **ORF**. The likelihood models use training statistics that were created from a collection of *matK*, *rbcL*, *trnK* and non-coding sequences, downloaded from the NCBI database http://www.ncbi.nlm.nih.gov/nuccore/.

The Pipeline



References

Bininda-Emonds, O. R. O., 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics*, 6(156).

CBOL Plant Working Group, 2009. A DNA Barcode for Land Plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), pp. 12794 - 12797.

Herbert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R., 2003. Biological identifications through DNA barcodes.. *Proceedings of the Royal Society B: Biological Sciences*, 7 February, 270(1512), pp. 313-321.

Lahaye, R. et al., 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences of the United States of America, 105(8), pp. 2923 - 2928.

Conclusions

From the information content in table 1, it can be seen that the pipeline proposed here is better than transAlign without manual editing.

Alignment Tool	Number of Sequences	Average Information Content
transAlign	3121	1.017997588
BarcAlign (using MAFFT alignment tool)	3121	**1.039812834**
Clustal W	3121	1.036667481
MAFFT	3121	1.035498804
KAlign	3121	0.990383914

Table 1: Average information content when aligning the matK barcode

Presenting Author's Contact

E-mail: hannah.garbett@southwales.ac.uk

Linked in: https://www.linkedin.com/pub/hannah-garbett/82/747/727

Acknowledgements

This work is part-funded by the European Social Fund (ESF) through the European Union's Convergence programme administered by the Welsh Government.

kess Sgoloriaethau Sgiliau: Economi Gwybodaeth / Knowledge Economy Skills Scholarships

ESF Europ & Chyfoeth Buddsoddi yn eich dyfodol / Europe & Wales: Investing in your future

Llywodraeth Cymru Welsh Government

National Botanic Garden of Wales Gardd Fotaneg Genedlaethol Cymru

Conference presentation Boston, July 2014

HPCW and bio-computational skills in Wales

- Presentation to National Assembly for Wales, Cross-Party Group on Science and Technology, July 2014
- All-Wales Life Sciences Skills Group, established Sept 2014 – WG support with potential future funding for bioinformatics (KESS?)
- Student symposium on computational biology – HPCW/USW initiative, Oct 2014. Very successful, next meeting in Norwich in Oct 2015
- Computational skills training at Welsh Universities – HPCW/USW pilot, Feb 2015. Very successful, next events planned in late 2015
- BioWales, March 2015 – identified skills requirements for informatics, especially in life sciences SMEs

8 October 2014

9:00am - 5:00pm

THE 1ST STUDENT SYMPOSIUM ON COMPUTATIONAL BIOLOGY AND LIFE SCIENCES

Treforest, Pontypridd

Website: rsg-uk.iscbsc.org/sym2014/

VENUE

Glamorgan Business Centre,
Treforest Campus,
University of South Wales,
CF37 1DL, Wales, UK

CHAIR OF THE SYMPOSIUM:

Farzana Rahman,
University of South Wales

CO-CHAIR:

Sayoni Das,
University College London

ORGANISED BY:

The UK Regional Student Group of International Society for Computational Biology and Genomics & Computational Biology Research Group at University of South Wales

KEY SPEAKERS

Manuel Corpus
The Genome Analysis Centre

Alex Bateman
EMBL-EBI, Hinxton, Cambridge

Christopher Creevey
Aberystwyth University

Natasha De Vere
National Botanic Garden of Wales

Tatiana Tatarova
University of Southern California

FUJITSU

HPC WALES · CYMRU

University of South Wales Prifysgol De Cymru

REGISTER NOW ON
http://rsguk.eventbrite.co.uk

ISCB REGIONAL Student GROUP United Kingdom

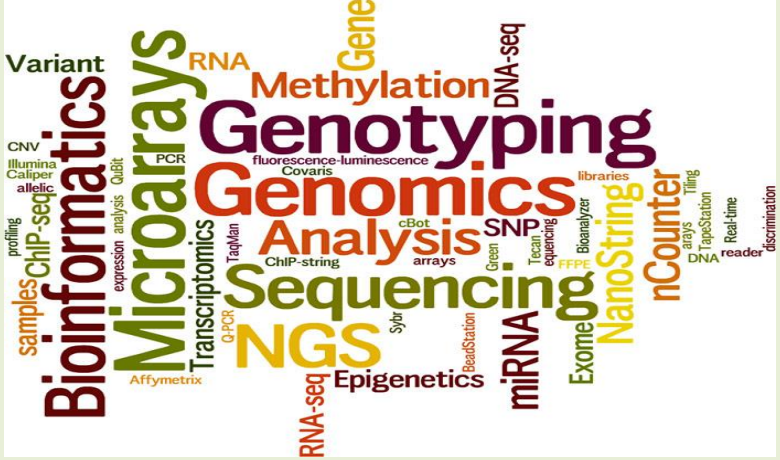
ISCB Student COUNCIL

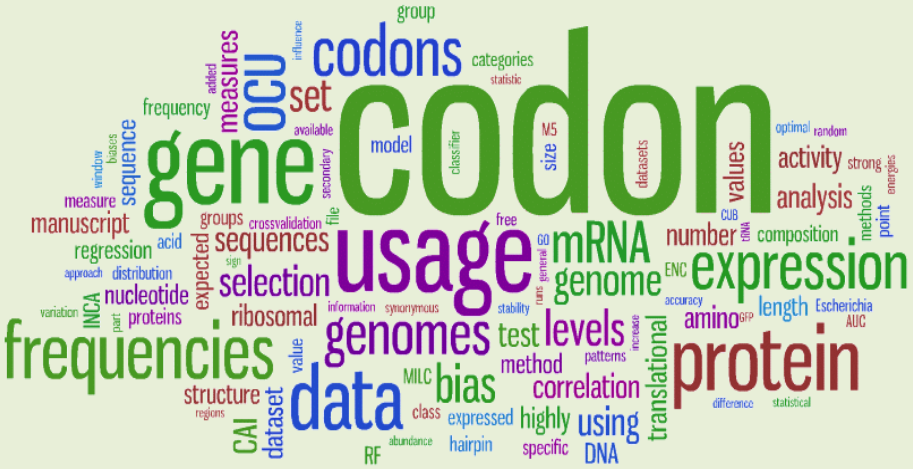
ISCB

BioWales, March 2015

A few conclusions …..

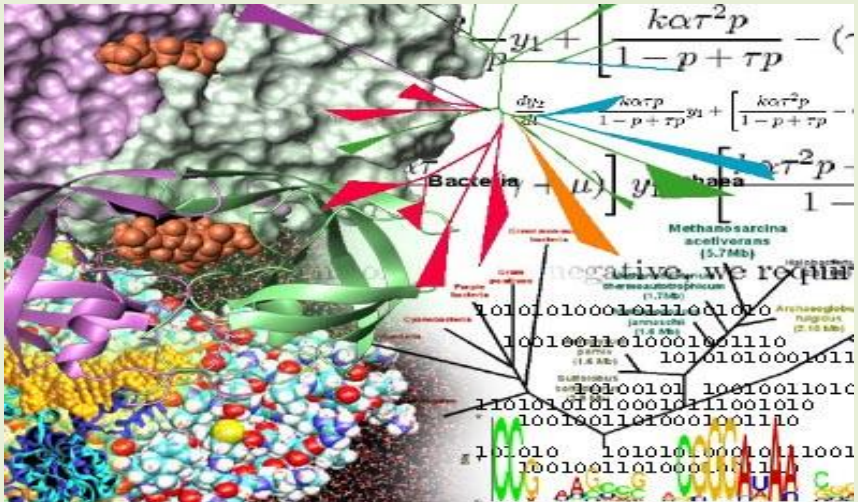
- We hope that HPCW can continue to provide what has been a very useful service for our computational biology researchers at USW
- Future developments should address capacity and connectivity issues as discussed at previous meetings
- HPCW is already carrying out a very useful role in skills training and this could be extended to address skills gaps, eg in life science students and SMEs
- We have previously discussed having a life sciences user group as part of the HPCW researcher network – this has received strong support from AWLSSG and merits further discussion

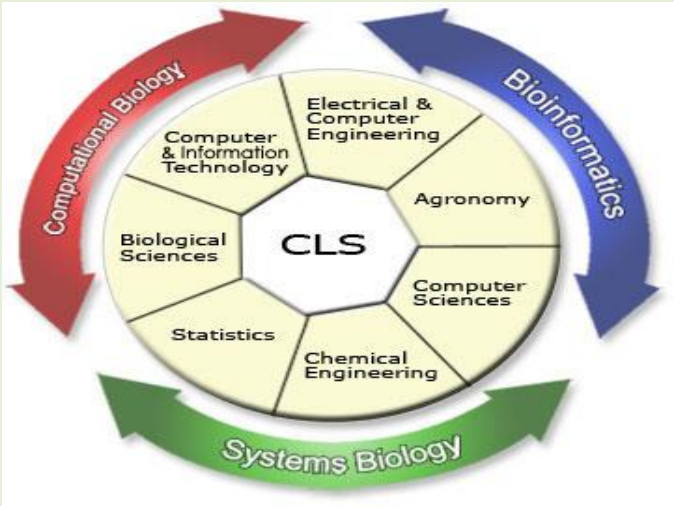




Thank you for your attention

Any questions?





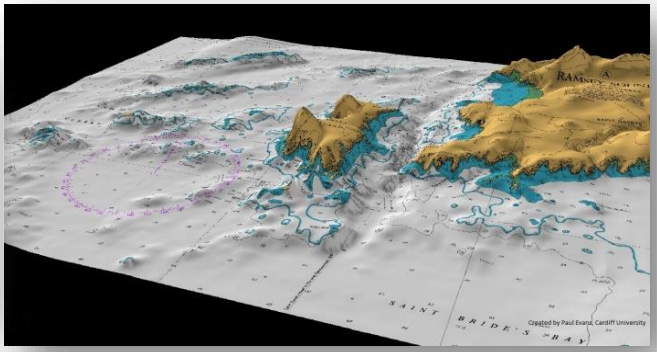
The Effect of Axial Flow Misalignment on Tidal Turbine Performance

Carwyn Frost

P.S. Evans, S. Tatum, C.E. Morris, A. Mason-Jones, T. O'Doherty & D.M. O'Doherty

Cardiff Marine Energy Research Group

School of Engineering / School of Earth and Ocean Sciences, Cardiff University



LCRI LOW CARBON RESEARCH INSTITUTE MARINE

Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales

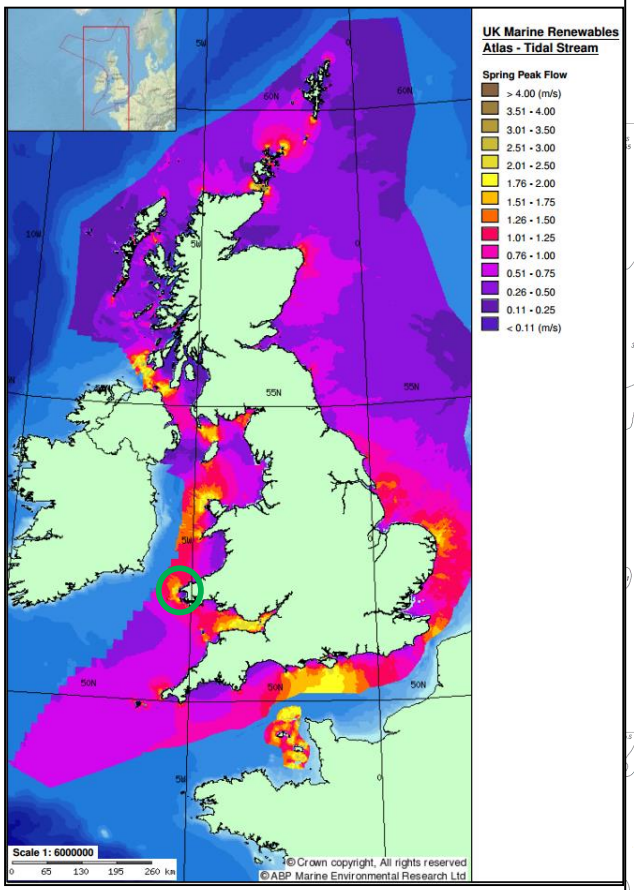
hefcw

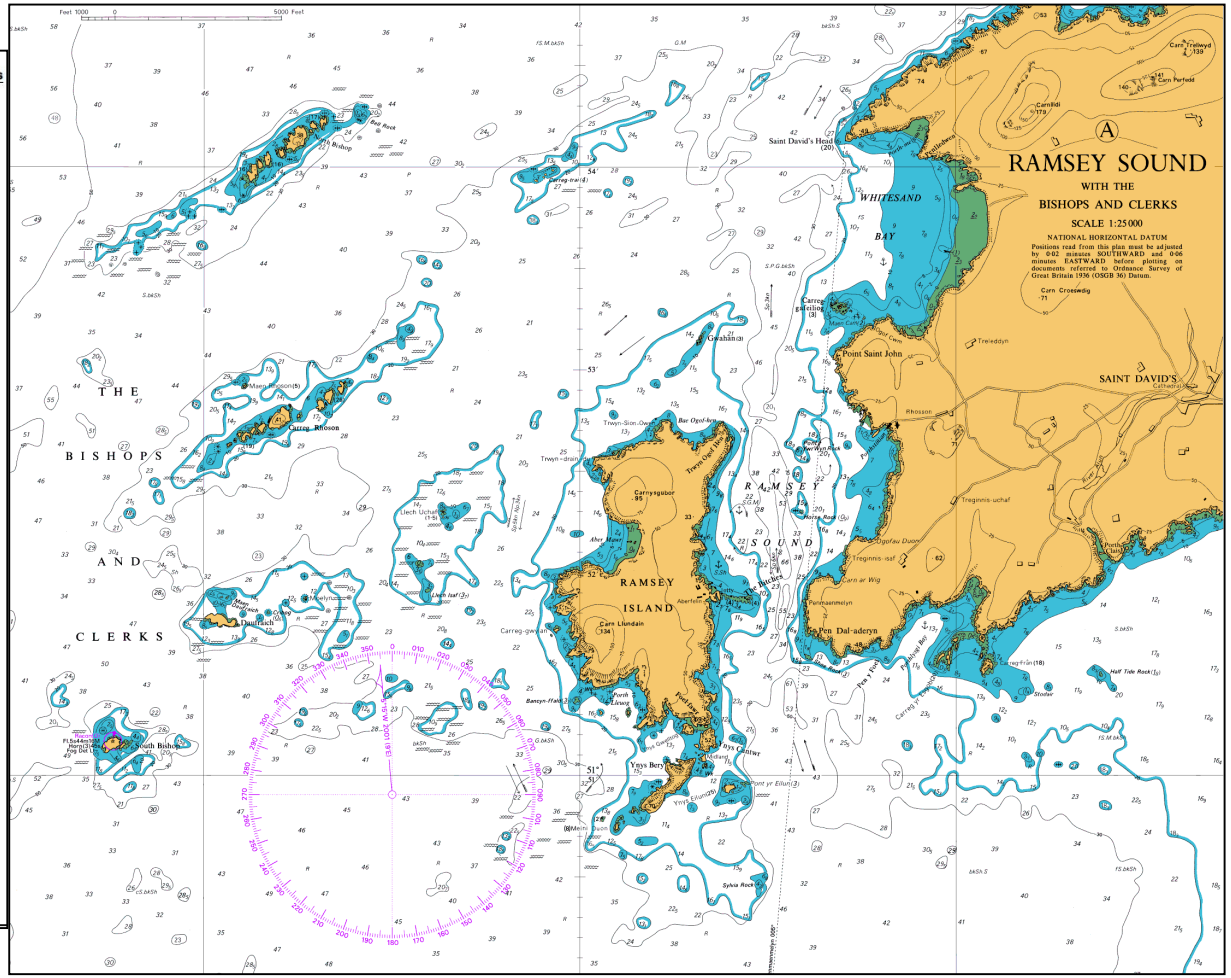
ERDF
Ewrop & Chymru Buddsoddi yn eich Dyfodol
Europe & Wales: Investing in your Future

Outline

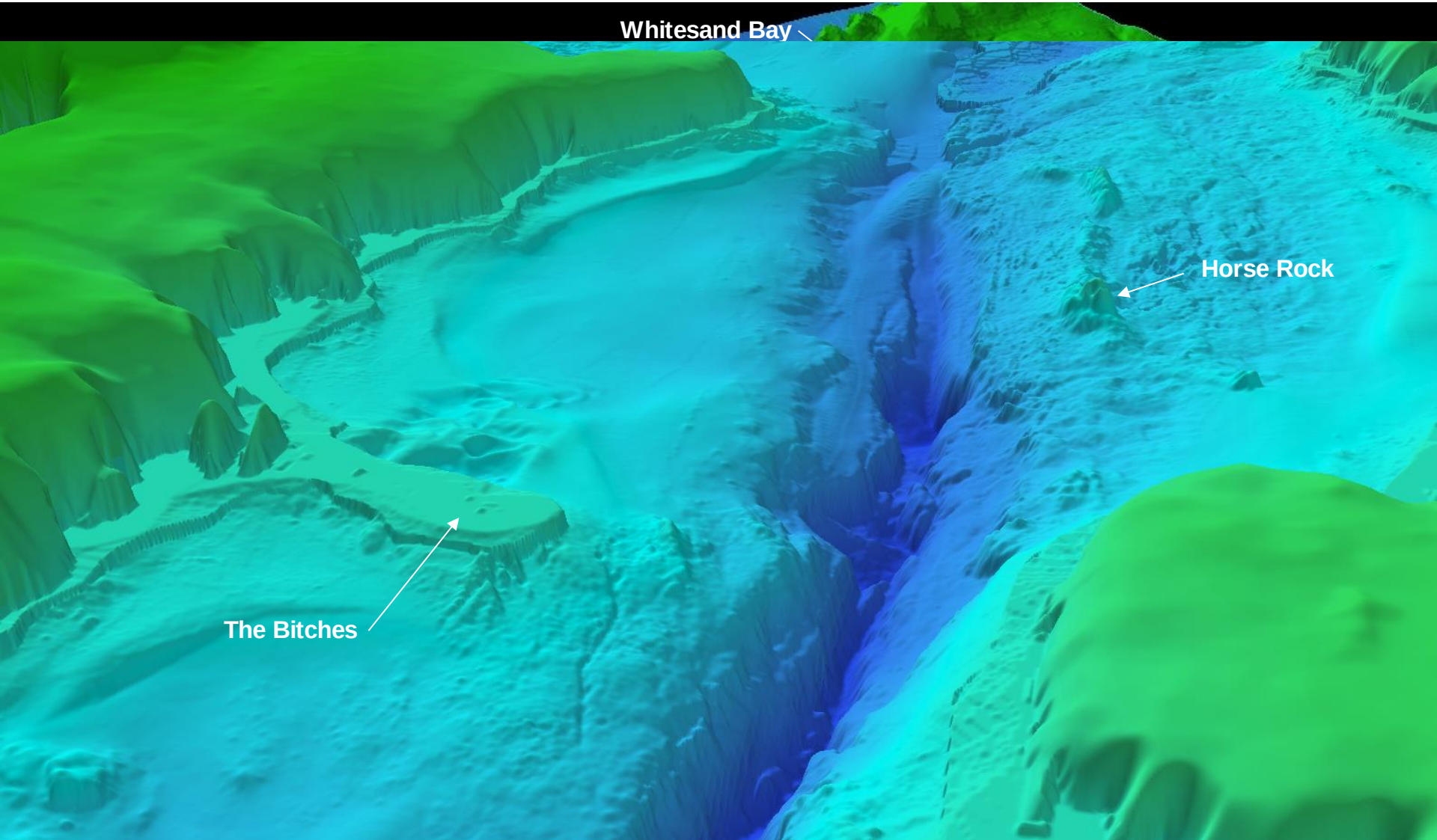
1. Case Site
2. Field Data
3. Numerical Model
4. Results
5. Conclusions

Case Site


UK Marine Renewables Atlas - Tidal Stream
Spring Peak Flow
> 4.00 (m/s)
3.51 - 4.00
3.01 - 3.50
2.51 - 3.00
2.01 - 2.50
1.76 - 2.00
1.51 - 1.75
1.26 - 1.50
1.01 - 1.25
0.76 - 1.00
0.51 - 0.75
0.26 - 0.50
0.11 - 0.25
< 0.11 (m/s)
Scale 1: 6000000
0 65 130 195 260 km
© Crown copyright. All rights reserved
© ABP Marine Environmental Research Ltd


A
RAMSEY SOUND
WITH THE
BISHOPS AND CLERKS
SCALE 1:25000
NATIONAL HORIZONTAL DATUM
THE BISHOPS AND CLERKS
RAMSEY SOUND
RAMSEY ISLAND
WHITESAND BAY
SAINT DAVID'S
Point Saint John

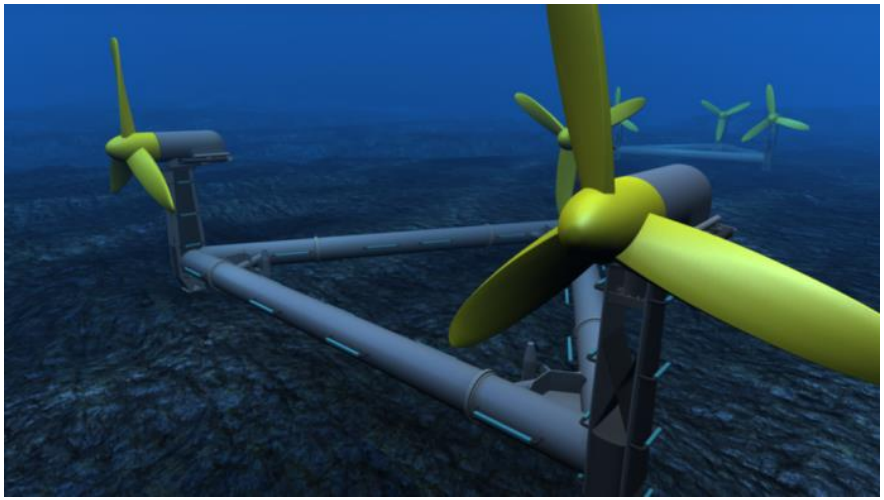
Case Site



Case Site

Ramsey Sound:

- Macrotidal strait
- 3 km Long
- 500-1600 m Wide
- Typical Water Depths: 20 - 40 m
- Tidal Range: 1.5 – 5 m



Tidal Energy Limited:

- DeltaStream deployment Site
- 400 kW Demonstration Project
- 12 m Diameter Turbine
- 12 Month Deployment

SUPERGEN GRAND CHALLENGES
CMERG CARDIFF MARINE ENERGY RESEARCH GROUP
CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD
LCRI MARINE LOW CARBON RESEARCH INSTITUTE
Cyngor Cyllido Addysg Uwch Cymru Higher Education Funding Council for Wales hefcw
ERDF Ewrop & Chymru Buddsoddi yn eich Dyfodol Europe & Wales: Investing in your Future

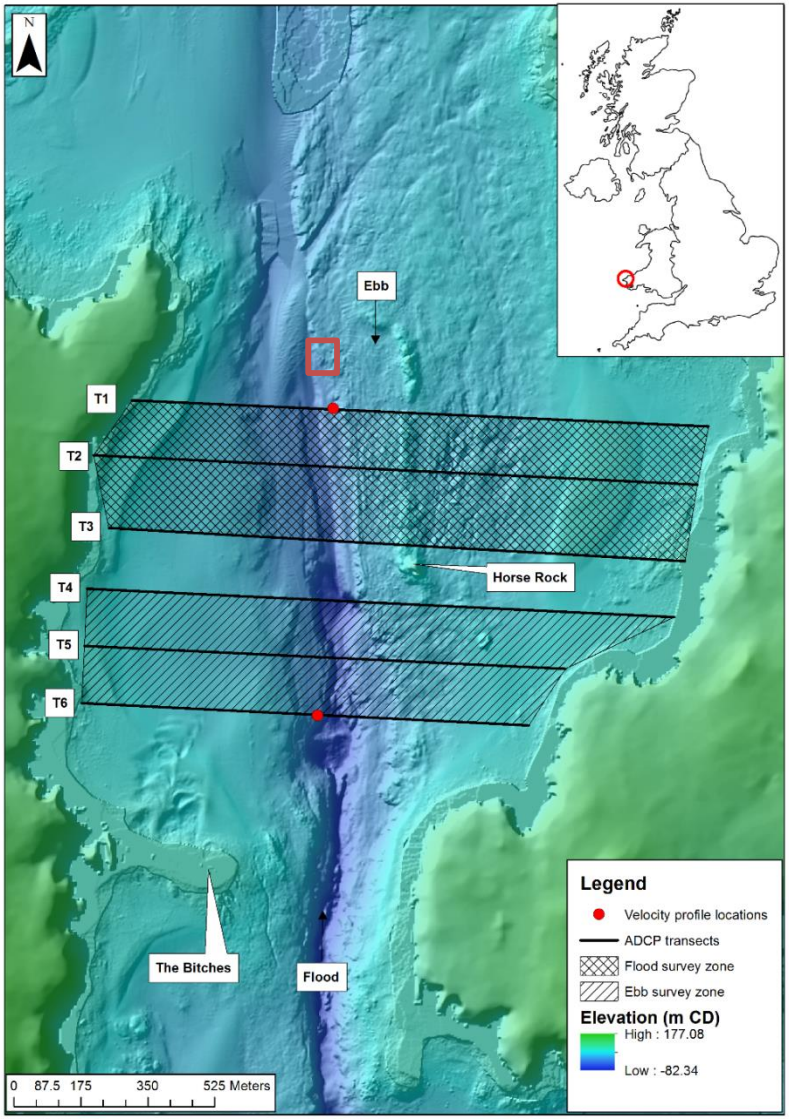
Field Data

Methodology:

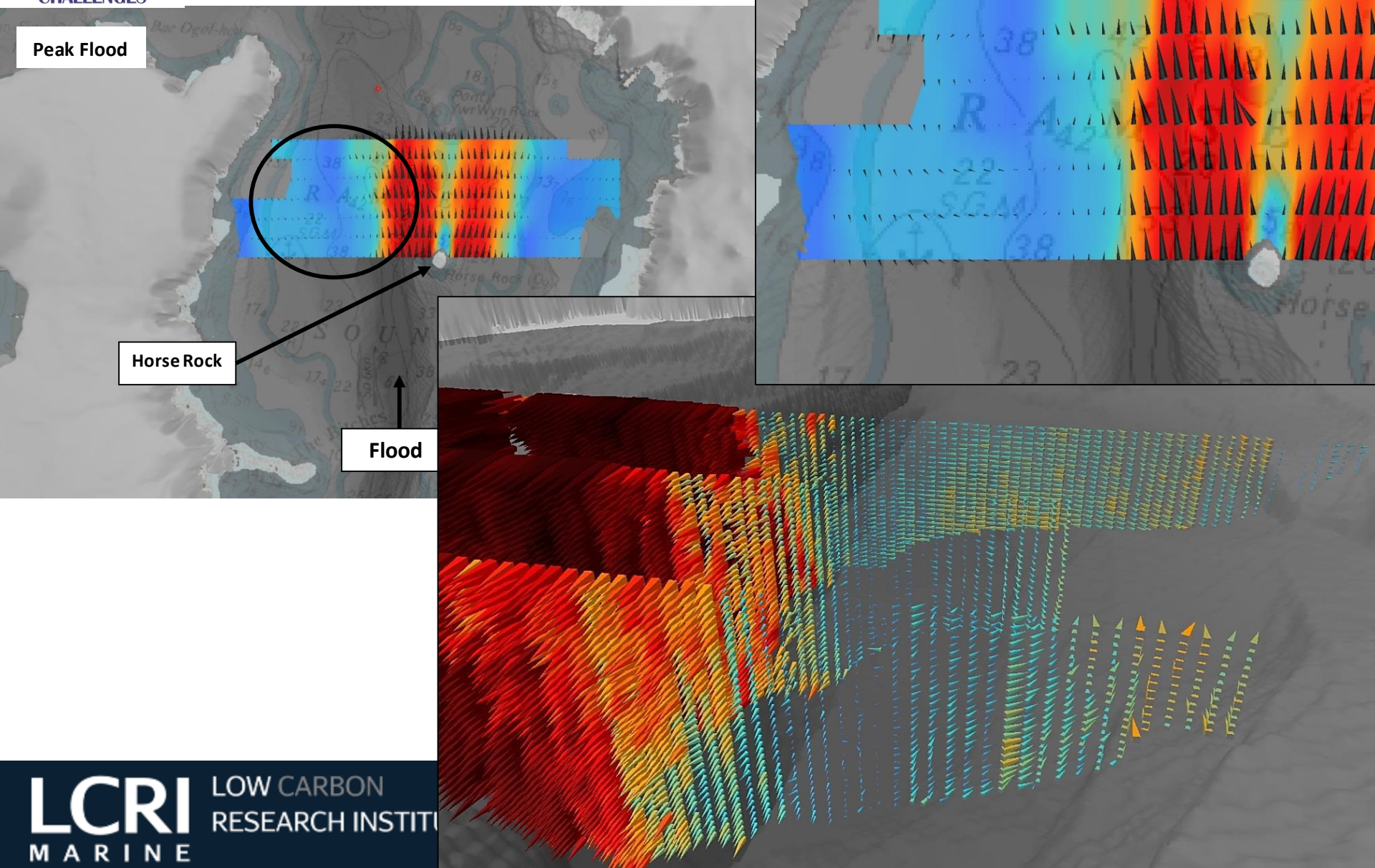
- Vessel-mounted 600-kHz ADCP
- 3 x survey tracks
- 2 x zones (flood & ebb – half tidal cycle)

Data Collected:

- Current velocity and direction
- Depth



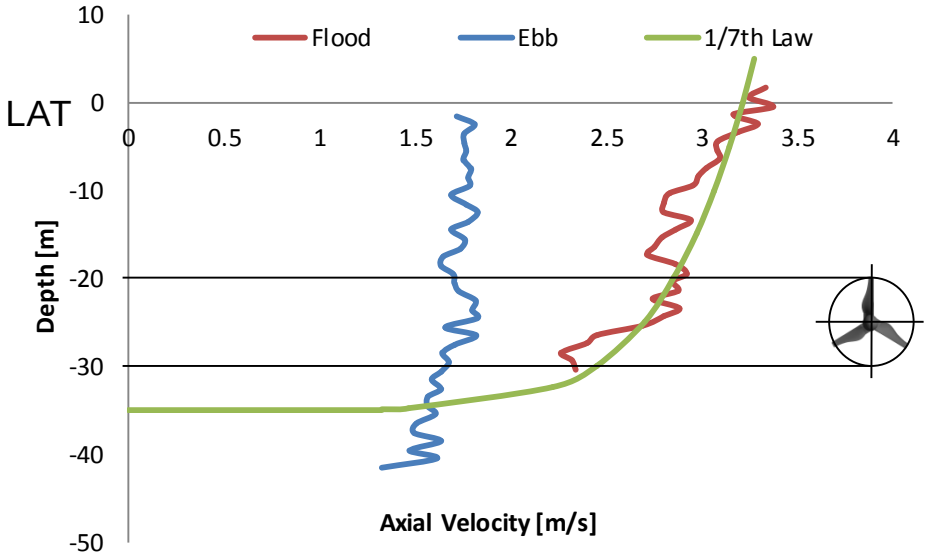
Field Data

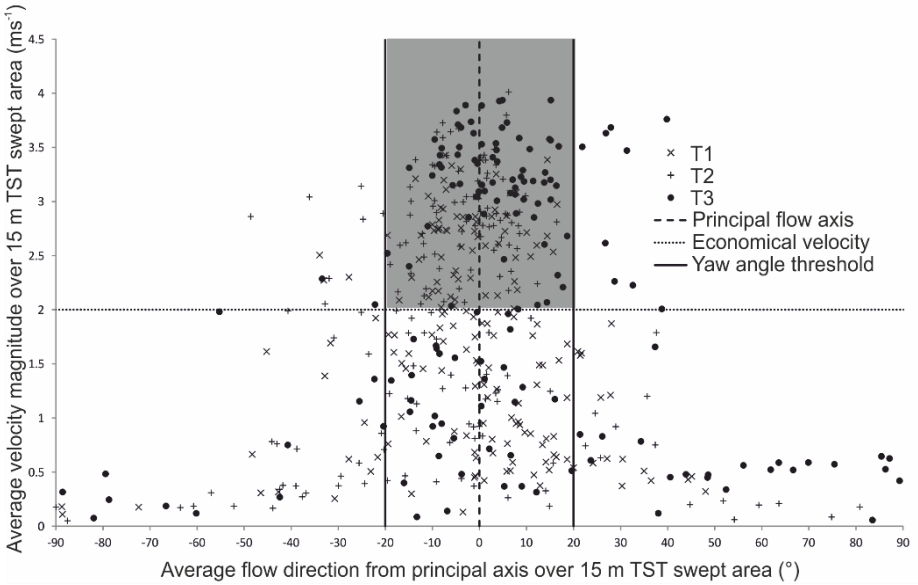


Field Data

Results:

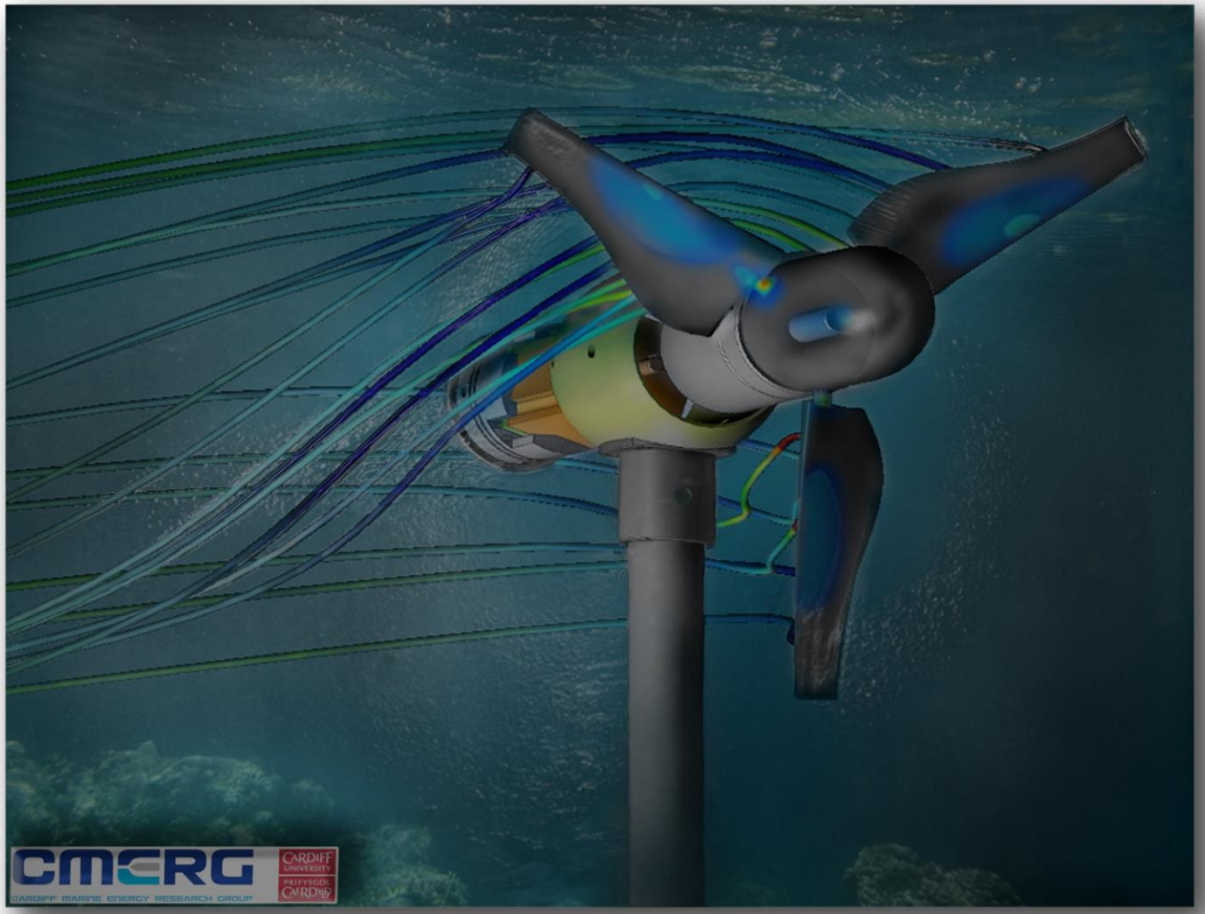
- Economical viable threshold of 2 m s^{-1}
- High proportion of velocities fall within ±20°





- Velocity Profiles (1/7th Power Law)
- Results are highly site specific

Numerical Modelling

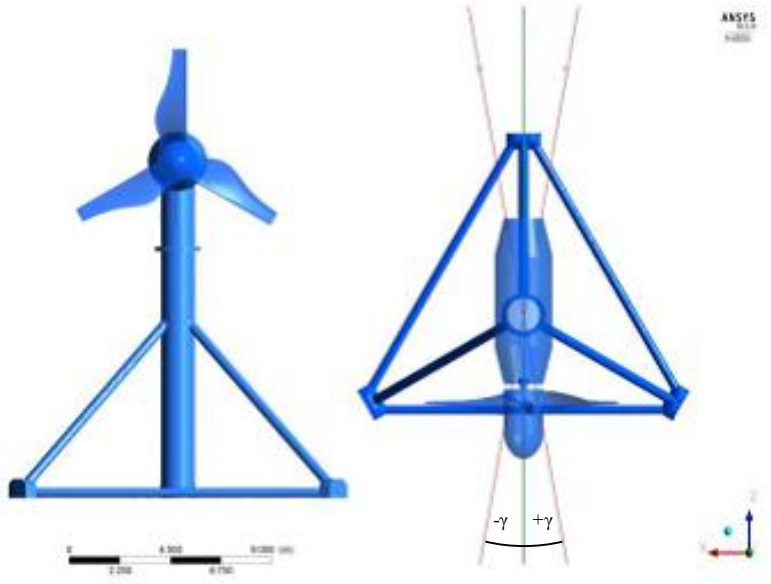


Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales

Turbine Design

Table 1 – Geometry Parameters*

Turbine Specification	Value
Turbine Diameter (D)	10 m
Hub Diameter	1.8 m
Blade Profile	Wortman-FX-63-137
Blade Twist	33 °
Optimum Blade Pitch	6 °
Stanchion Diameter	1.8 m
Nacelle Diameter	2.6 m
Tower height	15 m



Aligned Turbine Design

Table 2 – Mesh Parameters

Mesh Specification	Value
Elements in Sea Domain	2 million
Elements in Rotating Domain	7 million

*Mason Jones, A. 2010. Performance Assessment of a Horizontal Axis Turbine in a High Velocity Shear Environment. *PhD Thesis*

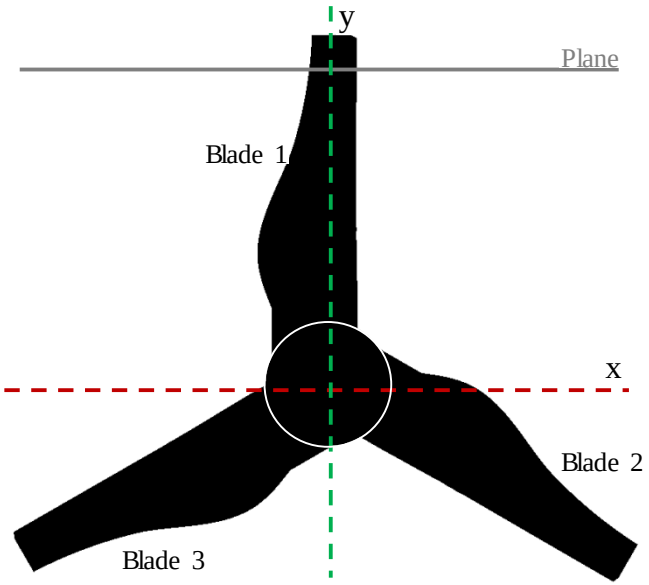
Performance Characteristics

Coefficient of Power $C_P = \frac{T\omega}{1/2 \rho A V^3}$

Coefficient of Torque $C_\theta = \frac{T}{1/2 \rho A R V^2}$

Coefficient of Thrust $C_T = \frac{F}{1/2 \rho A V^2}$

Tip Speed Ratio $\mathrm{TSR} = \frac{\omega R}{V}$

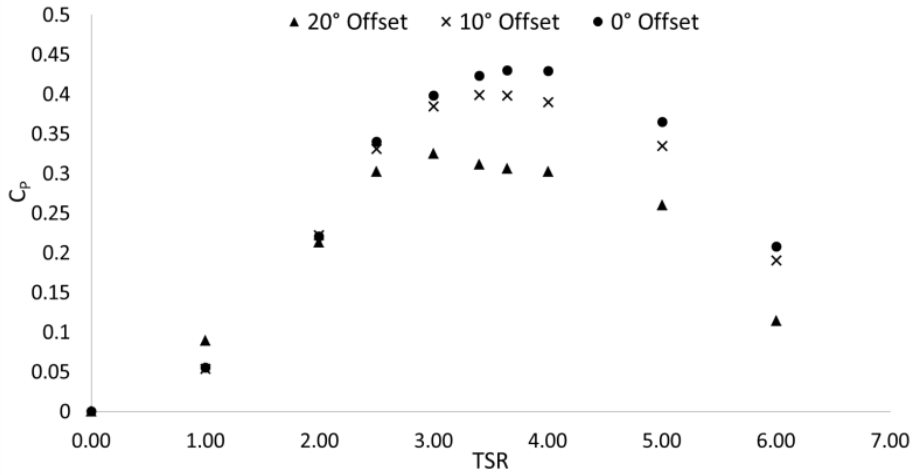


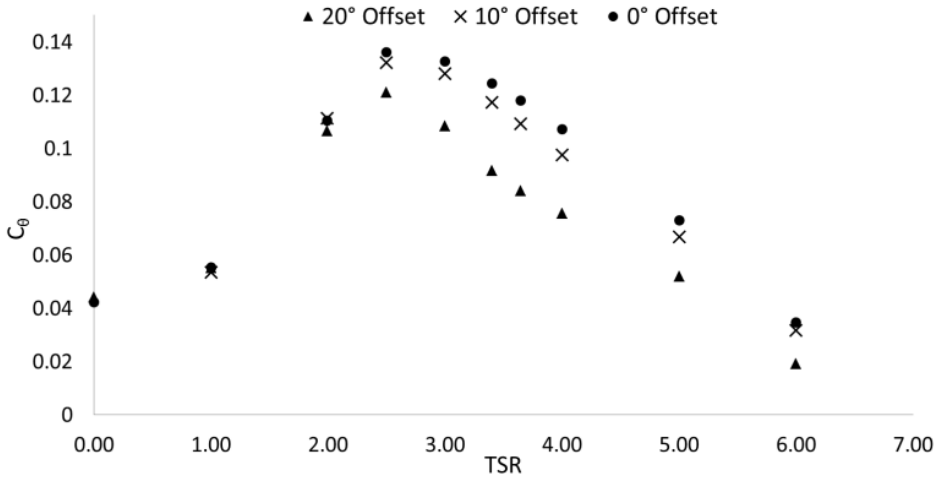
Axis about which the bending moments are taken.

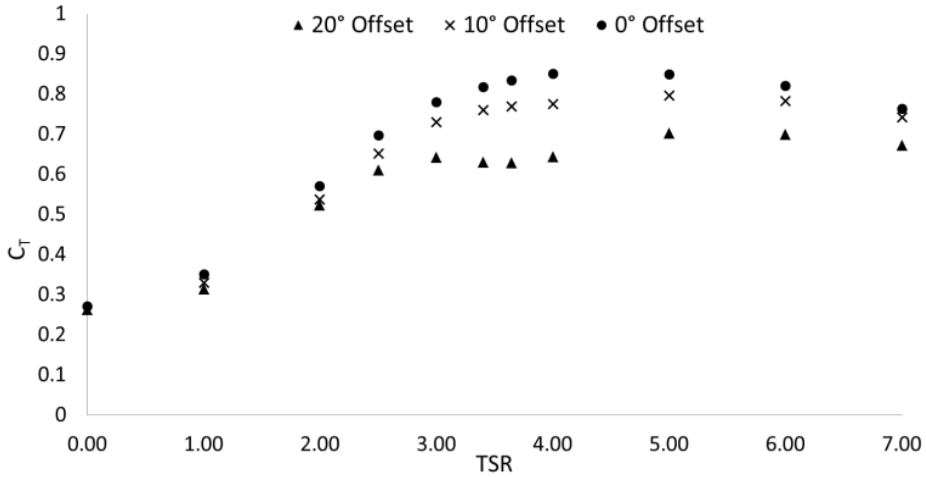
The bending moments are calculated by taking the torques about the x and y axis of the turbine, as seen in figure (opposite). The plane at which the pressure contours around blade 1 are taken is shown in figure (opposite).

T – Torque (Nm)
ω – Angular Velocity (rad/s)
ρ – Density (kg/m^3)
A – Area (m^2)
R – Radius (m)
V – Free Stream Velocity (m^3)

Results







Percentage drop at:	Peak C_P	Peak C_θ	Peak C_T
10° Offset	7.5%	3%	9%
20° Offset	29%	11%	24.5%

SUPERGEN GRAND CHALLENGES

CMERG CARDIFF MARINE ENERGY RESEARCH GROUP

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

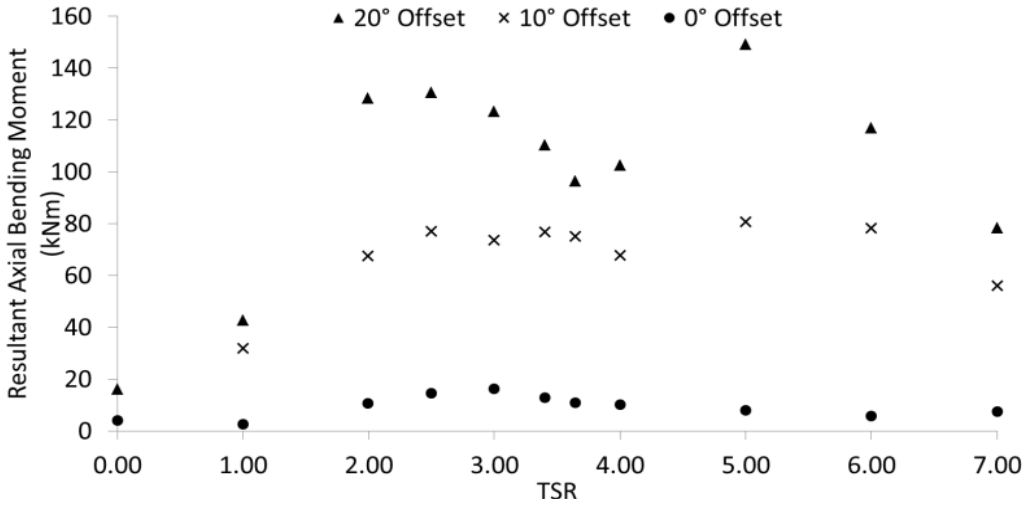
LCRI LOW CARBON RESEARCH INSTITUTE MARINE

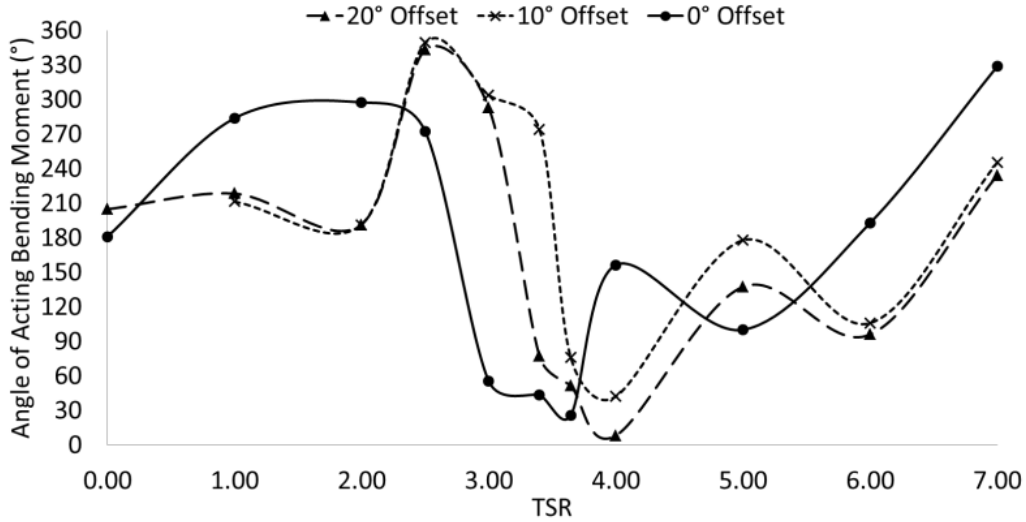
Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales

hefcw

ERDF Ewrop & Chymru Buddsoddi yn eich Dyfodol Europe & Wales: Investing in your Future

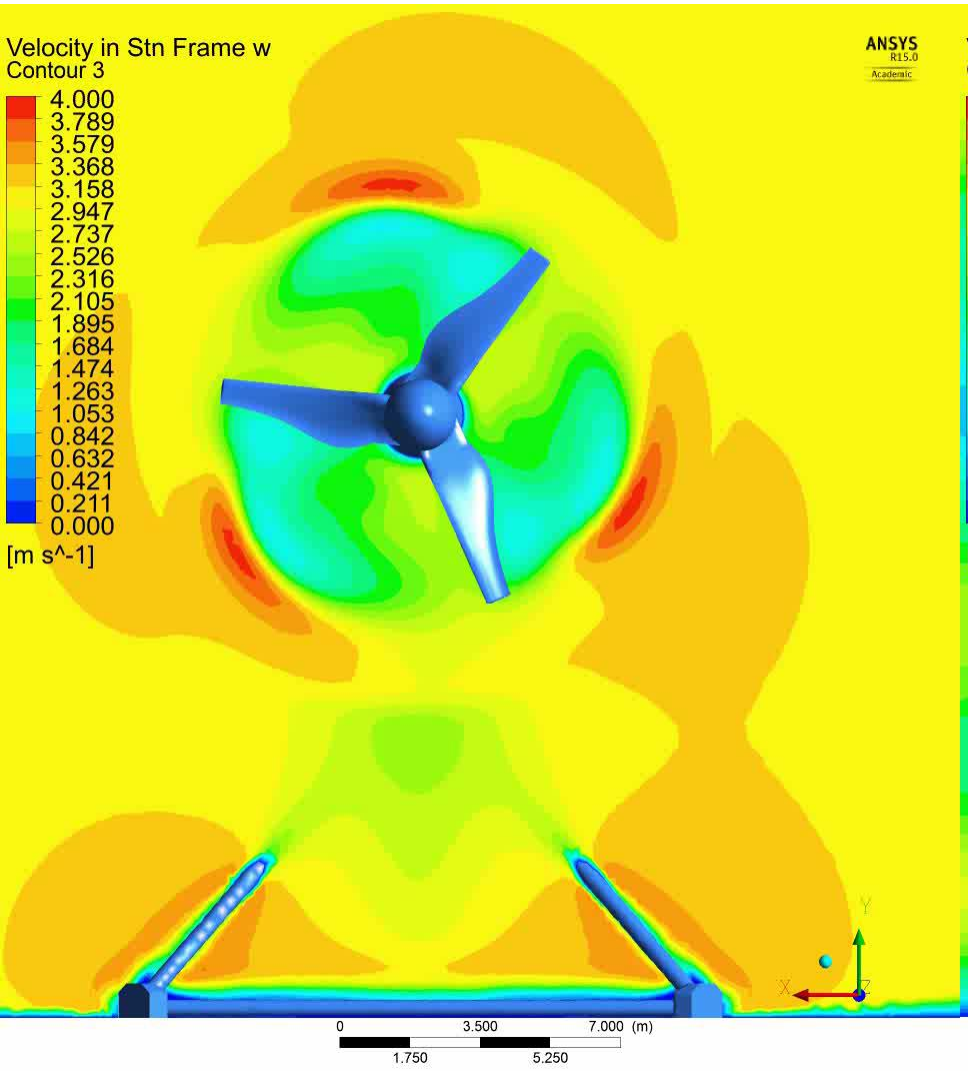
Results

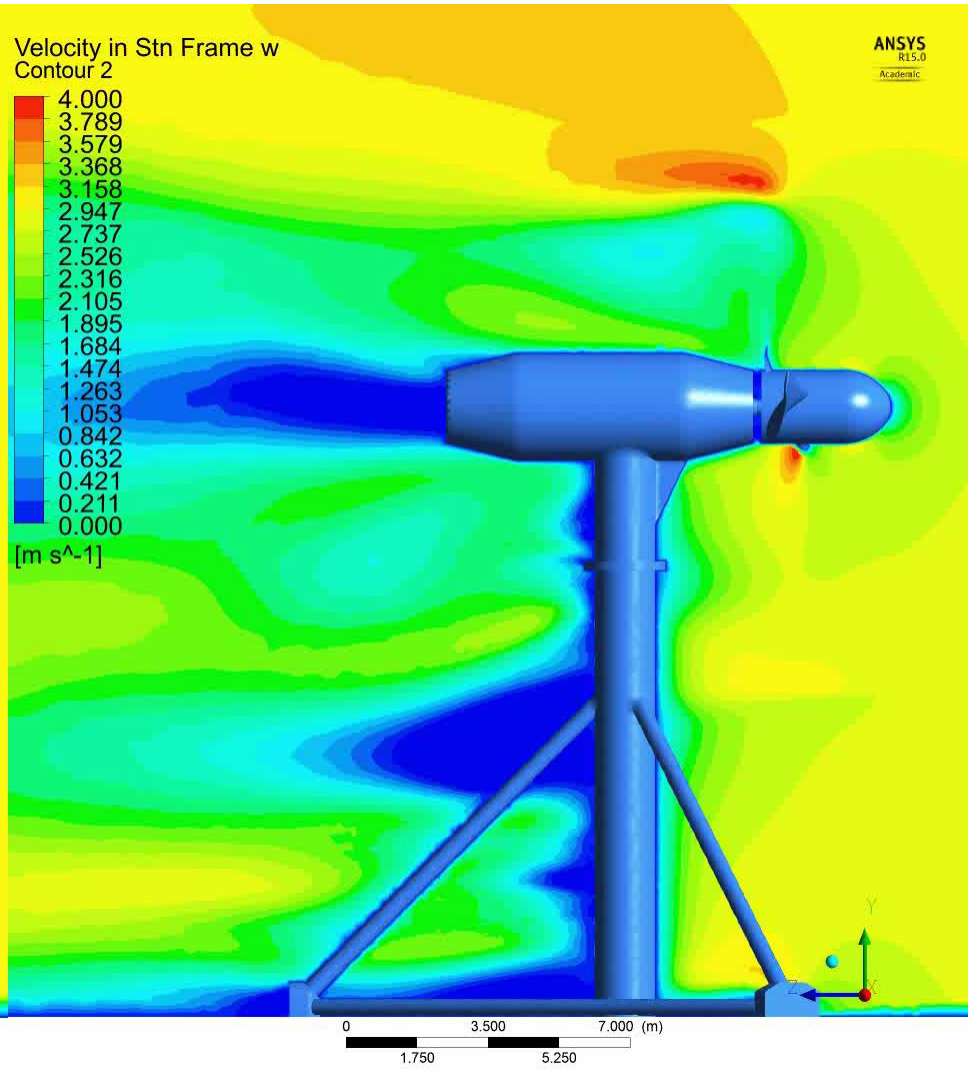




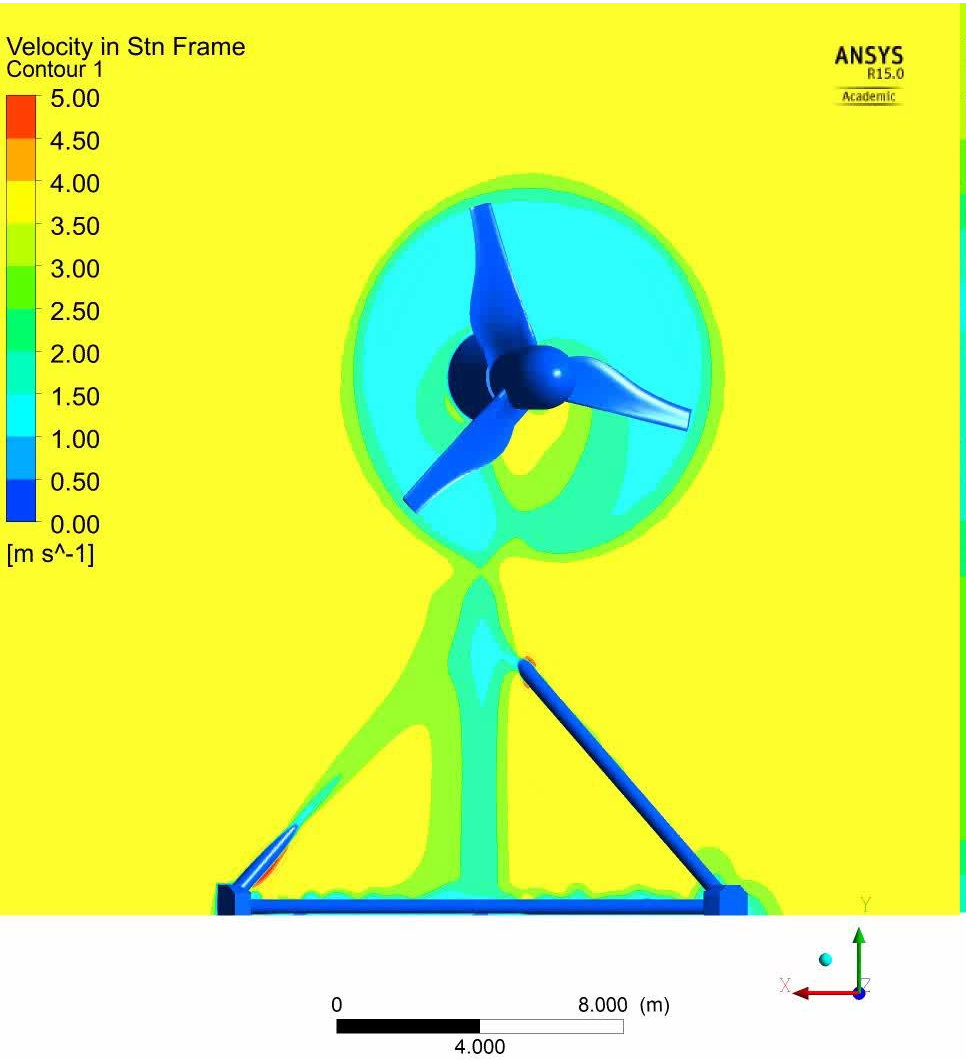
The resultant magnitude of the bending moments about the head of the driveshaft for the 10° and 20° misaligned turbines were found to be five times and nine times greater, respectively, than the aligned turbine.

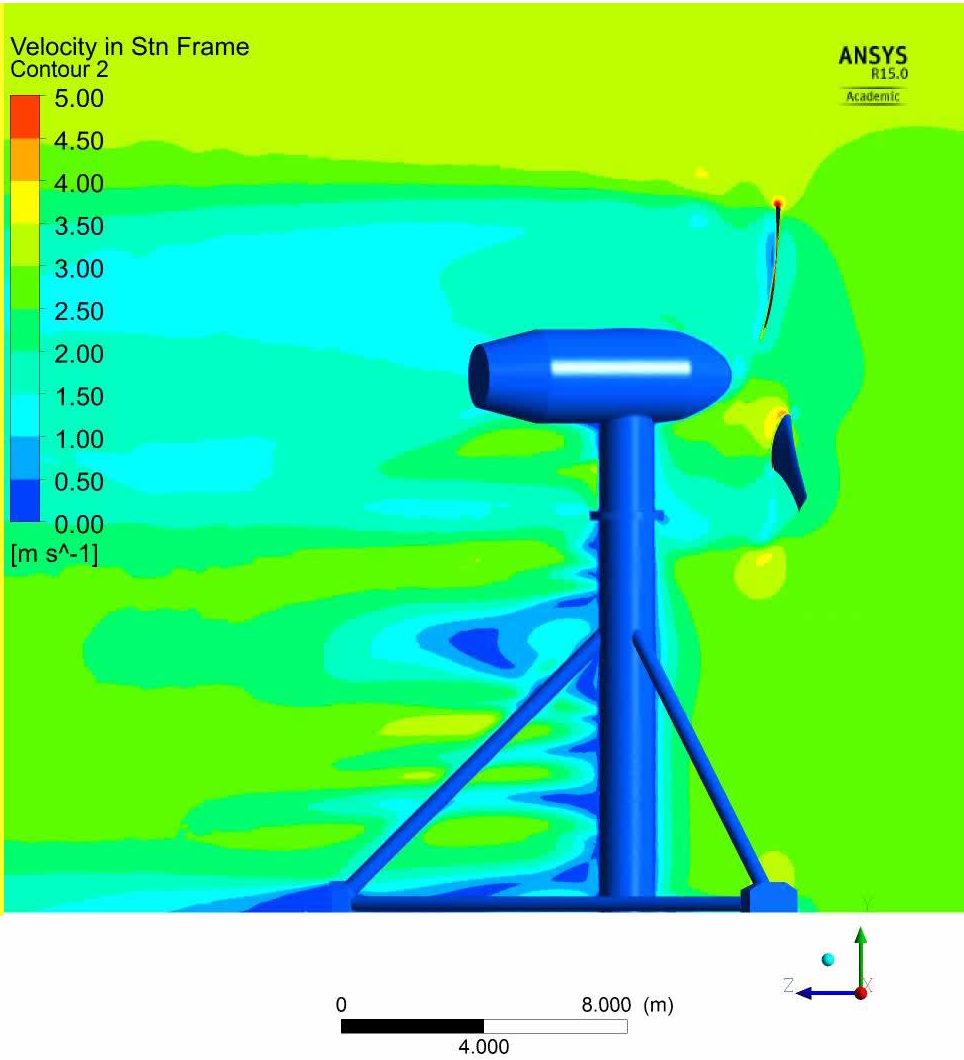
Results


Velocity in Stn Frame w
Contour 3
4.000
3.789
3.579
3.368
3.158
2.947
2.737
2.526
2.316
2.105
1.895
1.684
1.474
1.263
1.053
0.842
0.632
0.421
0.211
0.000
[m s^-1]
ANSYS
R15.0
Academic
0
3.500
7.000 (m)
1.750
5.250


Velocity in Stn Frame w
Contour 2
4.000
3.789
3.579
3.368
3.158
2.947
2.737
2.526
2.316
2.105
1.895
1.684
1.474
1.263
1.053
0.842
0.632
0.421
0.211
0.000
[m s^-1]
ANSYS
R15.0
Academic
0
3.500
7.000 (m)
1.750
5.250

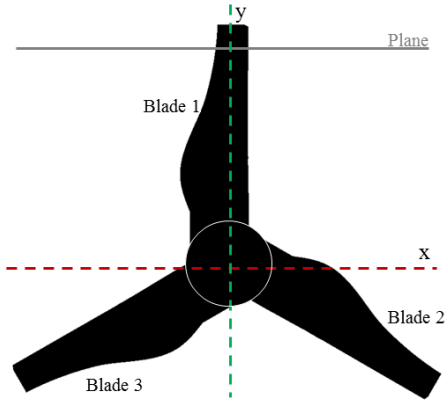
Results


Velocity in Stn Frame
Contour 1
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
[m s^-1]
ANSYS
R15.0
Academic
0
8.000 (m)
4.000

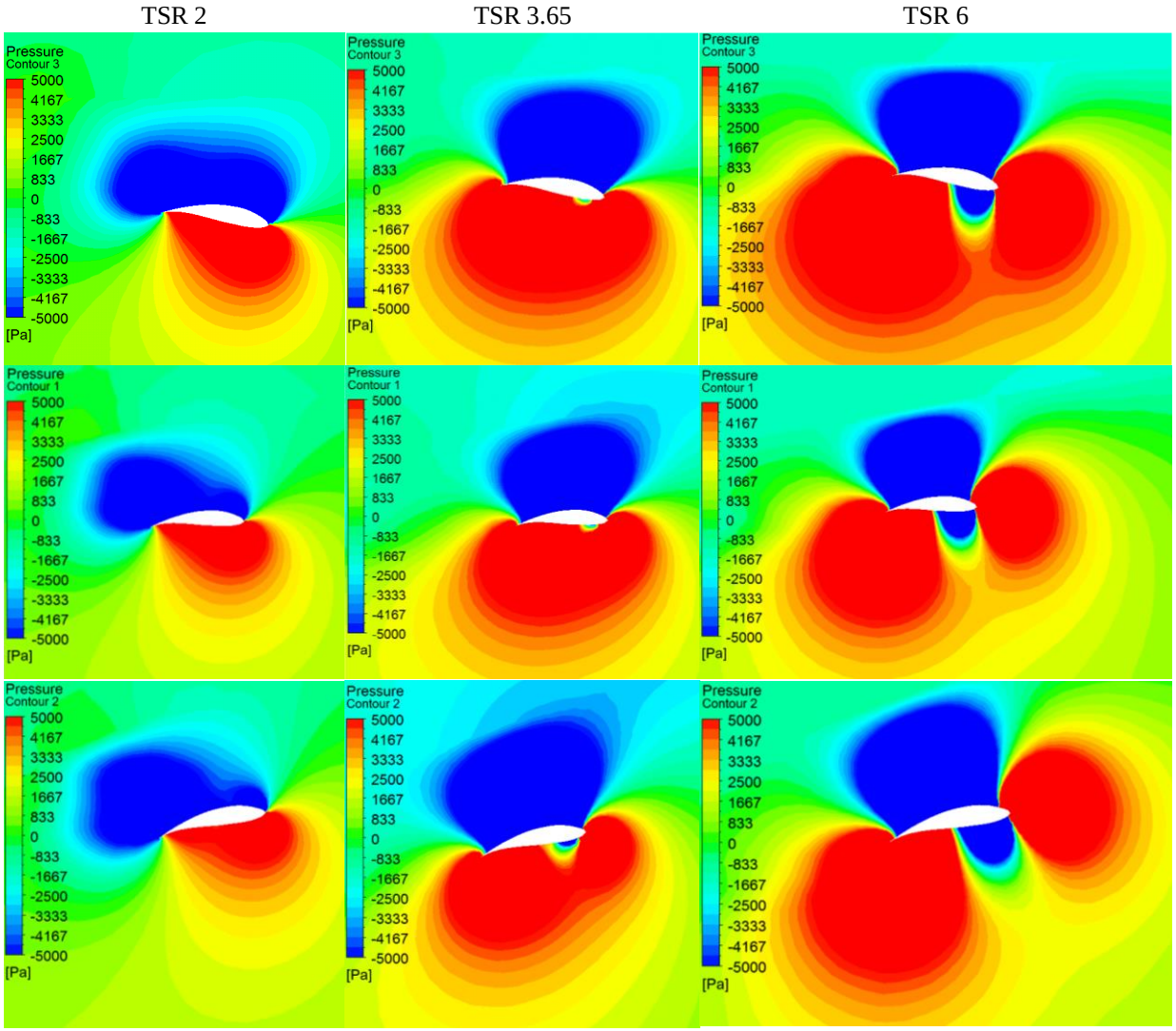

Velocity in Stn Frame
Contour 2
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
[m s^-1]
ANSYS
R15.0
Academic
0
8.000 (m)
4.000

Results

Pressure contours about Blade1 for three TSR values and various angles of turbine misalignment. Stall occurs as TSR increases, this happens sooner for the 20° misaligned turbine.



Axis about which the bending moments are taken.



Conclusions

Field Data:

- Misalignment threshold of γ= ±20°
- Economic Velocity threshold of 2 m s^{-1}

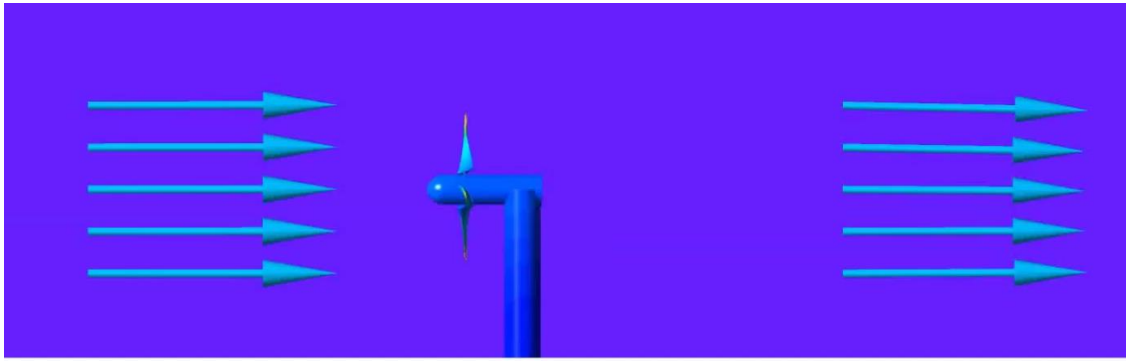
Numerical Modelling:

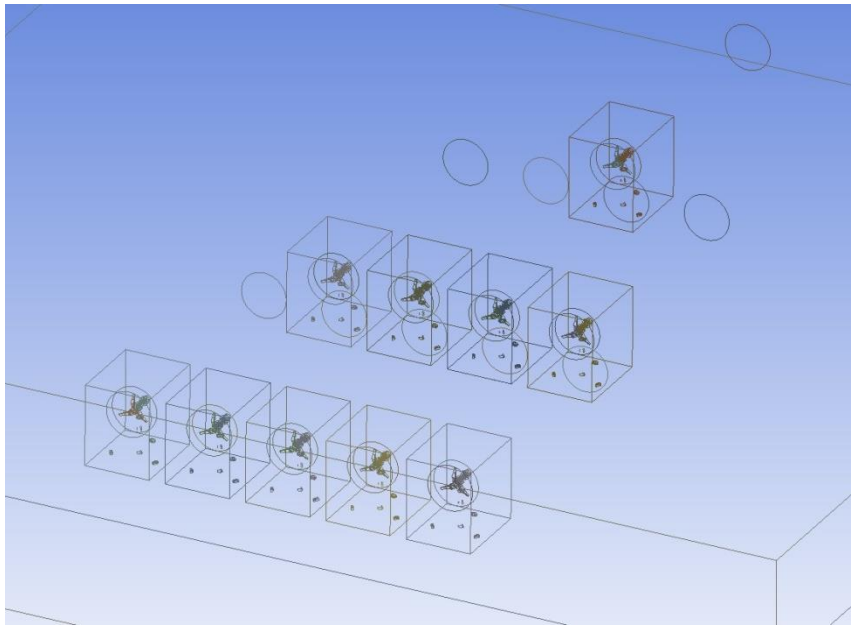
- Misalignment between the axis of rotation and free stream flow will reduce the overall performance of a turbine and will reduce the life expectancy or maintenance period of a turbine. The increased loading on the shaft and bearings will increase the likelihood of failure in the shaft support bearings and potentially the gearbox.
- The tolerance to axial flow misalignment between the free stream velocity and axis of rotation of a HATT requires defining, in order to avoid the clearly outlined detrimental effects it has on performance and loading.

SUPERGEN GRAND CHALLENGES
CMERG CARDIFF MARINE ENERGY RESEARCH GROUP
CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD
LCRI LOW CARBON RESEARCH INSTITUTE MARINE
Cyngor Cyllido Addysg Uwch Cymru Higher Education Funding Council for Wales hefcw
ERDF Ewrop & Chymru Buddsoddi yn eich Dyfodol Europe & Wales: Investing in your Future

Further Work

- Time Dependent
- Velocity Profiles
- Drive Shaft Loading
- Array Spacing





SUPERGEN GRAND CHALLENGES
CMERG CARDIFF MARINE ENERGY RESEARCH GROUP
CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD
LCRI MARINE LOW CARBON RESEARCH INSTITUTE
Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales
hefcw
ERDF Ewrop & Chymru Buddsoddi yn eich Dyfodol
Europe & Wales: Investing in your Future

My Experience

LCRI MARINE LOW CARBON RESEARCH INSTITUTE
Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales
hefcw
ERDF
Ewrop & Chymru: Buddsoddi yn eich Dyfodol
Europe & Wales: Investing in your Future

SUPERGEN GRAND CHALLENGES

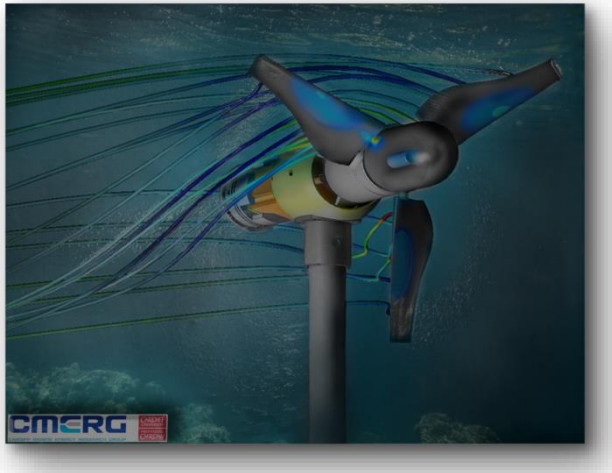
ANSYS

HPC WALES · CYMRU

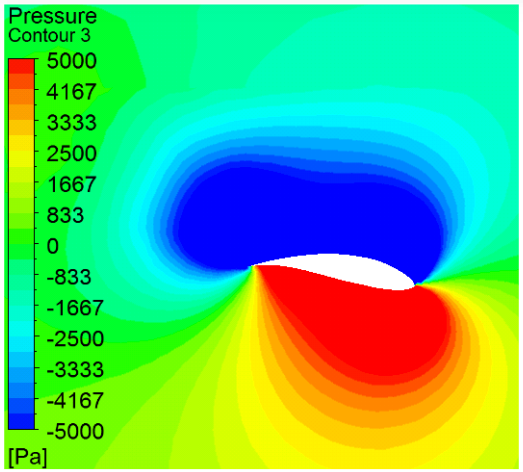
CMERG CARDIFF MARINE ENERGY RESEARCH GROUP

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Thank You



Questions?



LCRI LOW CARBON RESEARCH INSTITUTE MARINE

Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales

hefcw

ERDF
Ewrop & Chymru Buddsoddi yn eich Dyfodol
Europe & Wales: Investing in your Future

Cooperative Object Transport Using Swarm Evolutionary Robotics Methods

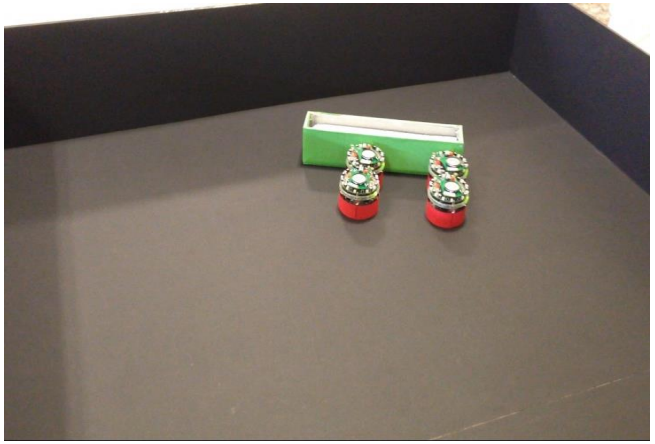
Muhanad H. Mohammed
mhm1@aber.ac.uk

Supervised by: Elio Tuci

Intelligent Robotics Group
Department of Computer Science

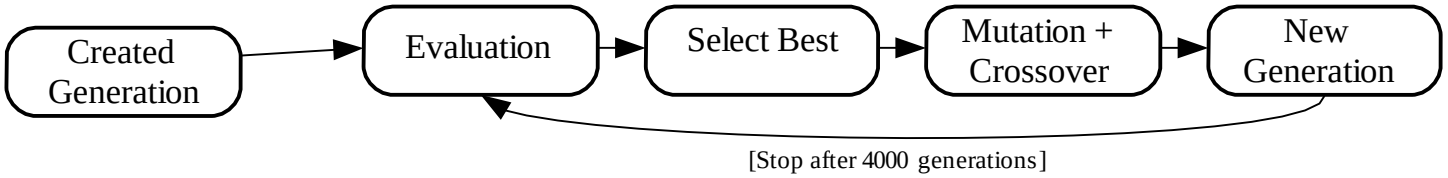
Inspiration From Biological Swarm

- Main challenges in Swarm Robotics design is to define the individual rules that produce the desired global behaviour.
- Drawing inspiration from social insects.
- Try to understand and exploit the power displayed by these species.



Evolutionary Algorithm

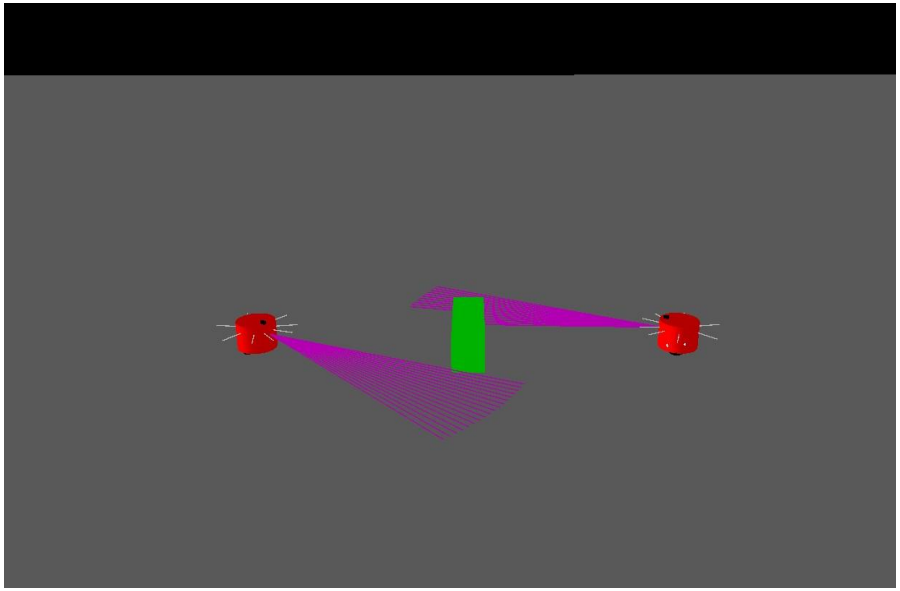
Inspired by Darwinian theory of evolution in nature.



- A **generation** of individuals are **created** randomly.
  - Individuals are **evaluated** with a performance fitness score.
  - The best scored individuals are **selected**.
  - The selected individuals undergo **mutation** and **crossover** operations to **breed** a **new generation**.
- Continue for hundreds or thousands of generations.

The Task and The Simulation Model

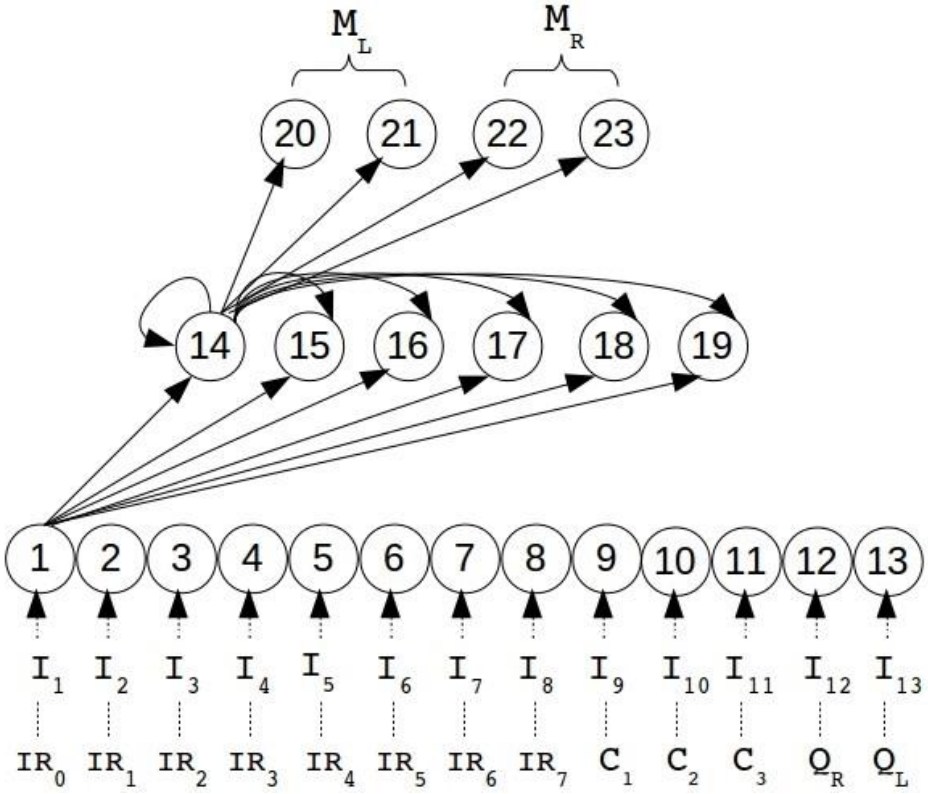
- • A group of two robots is required to push anelongate cuboid object which is too heavy to be Transported by part of the group.
- • Sensors:
  - camera, 50cm range.
  - 8 proximity sensors, 4 cm range.
  - 2 torque sensors.



Snapshot taken from the simulator showing the robots and the object

The Robot Controller

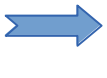
- Use an **artificial neural network** , as inspired by human brain.
- Continuous Time Recurrent Neural network (CTRNN) used to control the robot.



Evolutionary Algorithm and Fitness Score

- Evolutionary Algorithm Parameters:
  - population = 100, elite =8, mutation = 0.05 , crossover =0.4
  - Each stops after 4000 generations.
- Individual Fitness Score: $F_e = \sum_{t=0}^{T} (f_t^1 + f_t^2) + f^3$ *with* $T = 600 \simeq 60$ *seconds,* $t =$ *current timestep*

Rewards groups in which the robots approach the object and attach to it**:**



$f_t^1 = \sum_{r=1}^{R} IR^r{}_1 + IR^r{}_2 + (1 - d_r)$

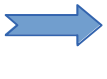
with R=2, d=distance from robot to object

Rewards groups that transport the object at maximum speed regardless of the object's trajectory:

$f_t^2 = O_{vel}$

where O_{vel} *= object linear velocity*

Rewards groups that transport the object as far as possible:



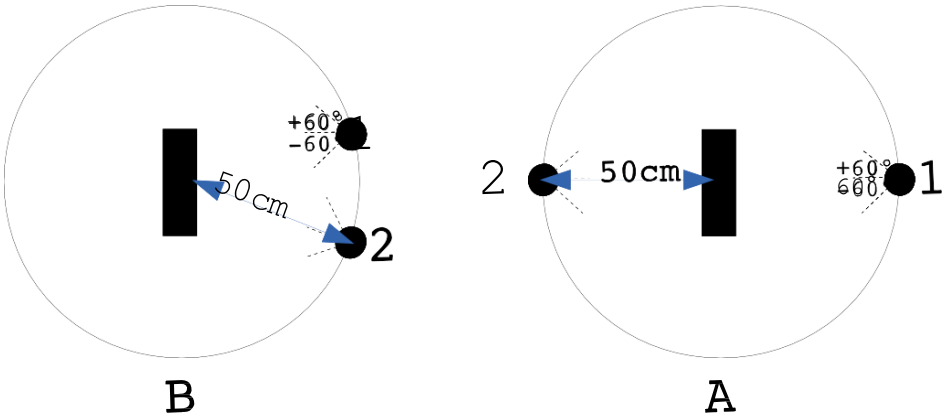
$f^3 = O_{start} - O_{final}$

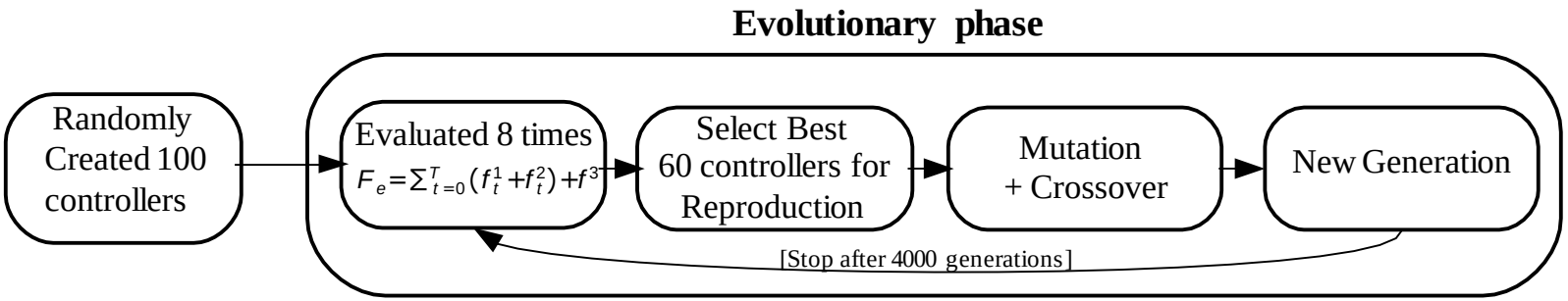
Experiment Phases

- Experiment undergoes two phases:
  - **Evolutionary phase** (i.e training phase)
    - Generate individuals
    - Evaluate individuals.
  - **Post-Evaluation phase** (i.e test phase)
    - Robustnes
    - s Flexibility
    - Scalability

Evolutionary Phase (Training)

- **1 controller** is a **group** of **2 robots**.
- **One object type** were used with **length=30cm** and **mass=280g.**
- Each group evaluated **8 trials** with **different** starting **positions** and **orientations.**

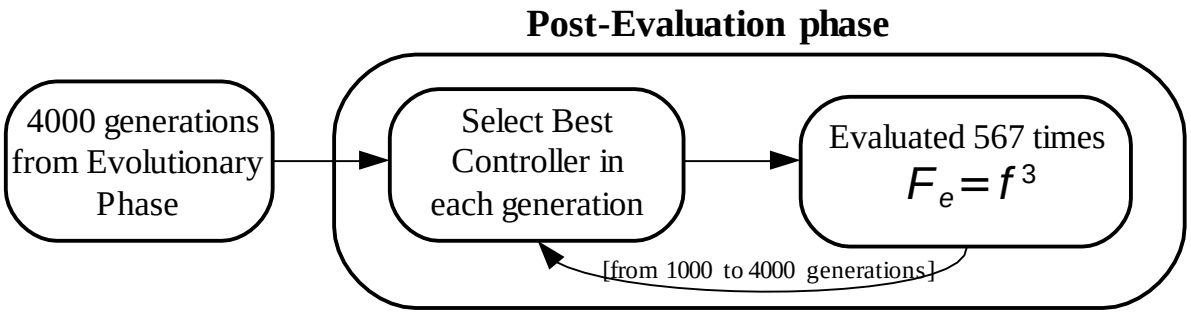




Post-Evaluation Phase (Testing)

- The **group cardinality** of 2, 3 and 4 robots were used.
- Objects with 3 different **length** 20cm, 30cm,40cm and three different **masses** per group

  (i.e 9 object types per group):
  - 230g, 280g, 330g for 2 robots cardinality.
  - 380g, 430g, 480g for 3 robots cardinality.
  - 530g, 580g, 630g for 4 robots cardinality.
- 21 different starting **position** and **orientation** per group.
- Thus, each group **evaluated** 567 times (i.e. 3 group cardinalities * 9 object types * 21 starting conditions).



Computation Expense

- Computation extensive process for both evolutionary phase and re-evaluation phase.
- We make heavy use of **HPC-Wales** services.
- Example:

HPC-Cluster

- One generation in Evolutionary phase last for 30 second.
- 4000 generation ≈ 33.3 hours using 100 cores.

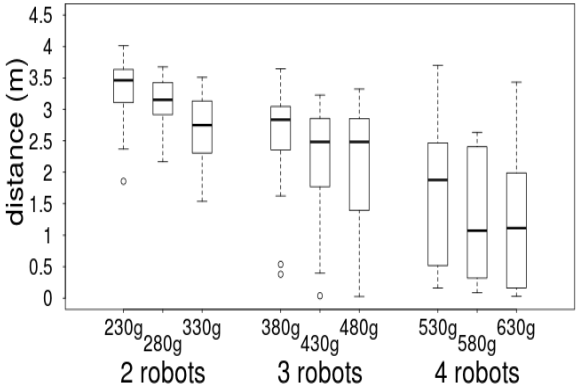
Local Machine

- One generation in Evolutionary phase last for 20 minutes
- 4000 generation ≈ 55.5 days

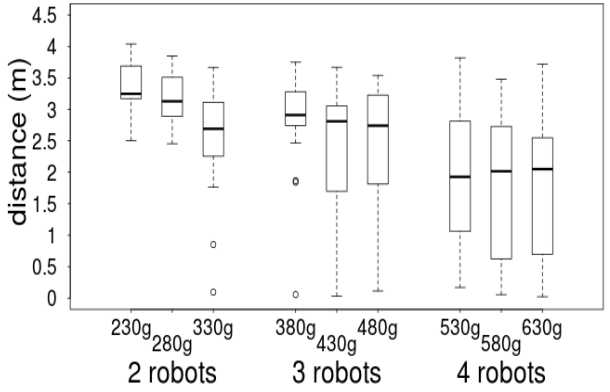
Simulation Results

& Publication Submission, March 2015

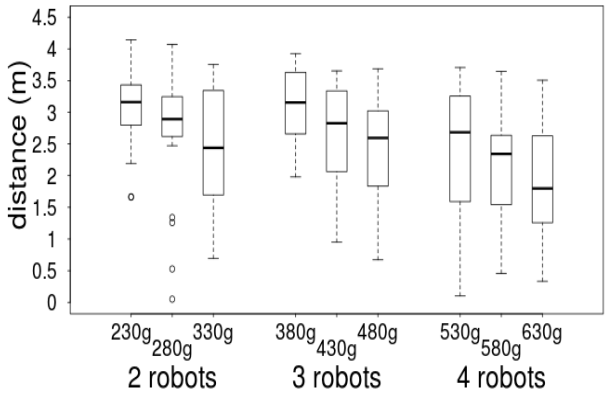
Results-RobustnessofBestGroupinT-conditionw.r.ttoObject&GroupCardinality



(a) 20cm object length

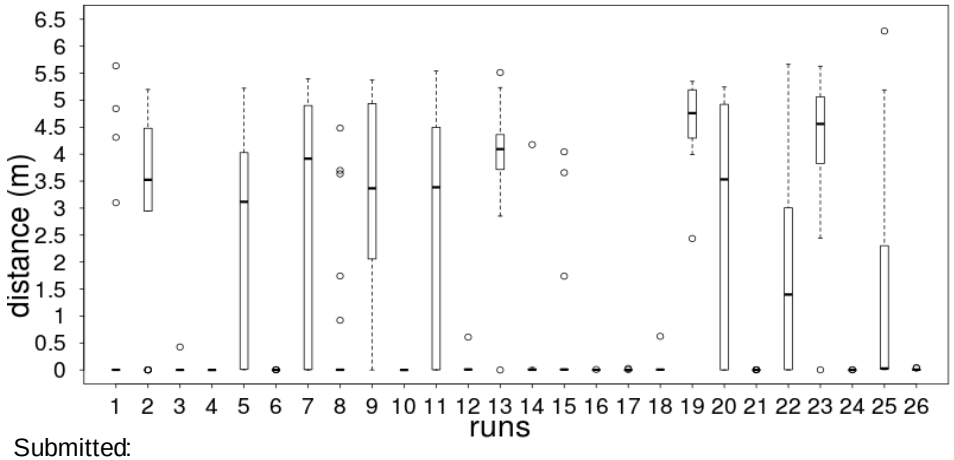


(b) 30cm object length

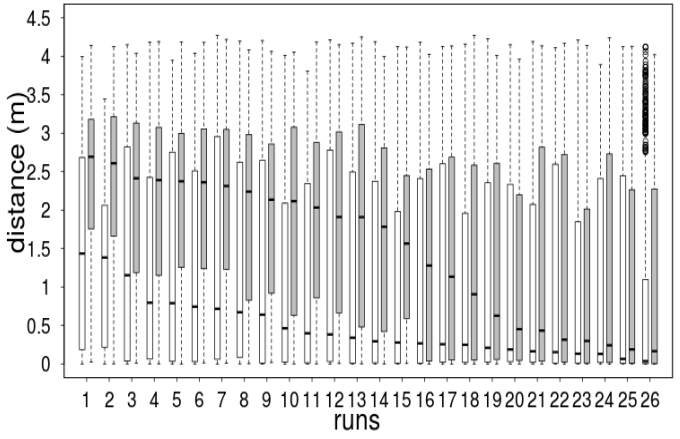


(c) 40cm object length

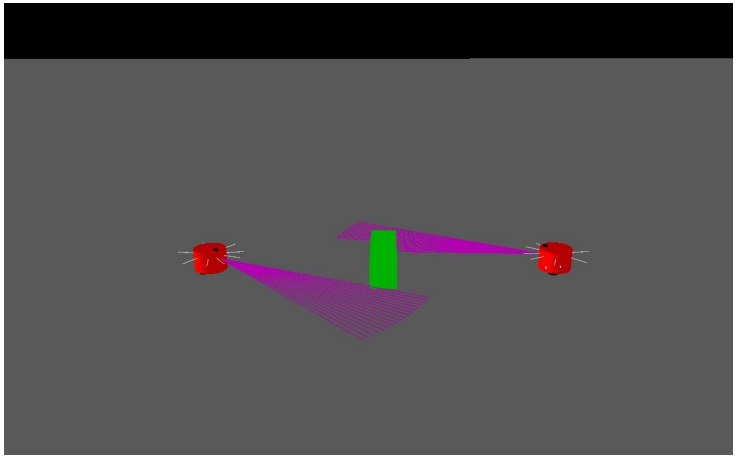
Results-Recruitment-test

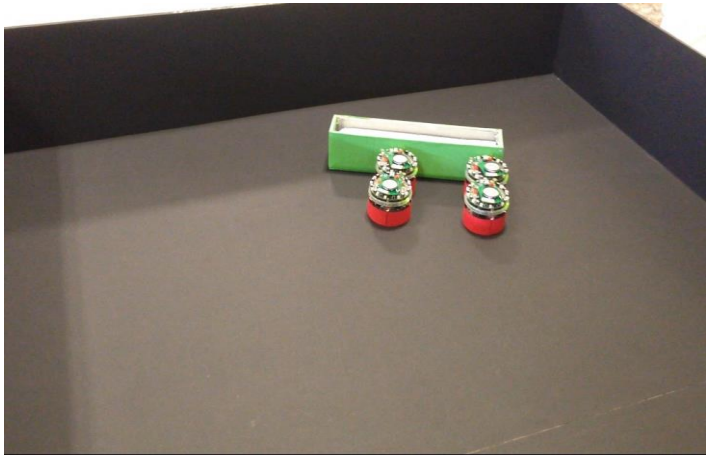


Results–Torque(T)vs.Non-Torque(NT)



Submitted:
Mohammed, M., Lu, C., & Tuci, E. (2015). Cooperative Object Transport Using Swarm Evolutionary Robotics Methods. In Advances in Artificial Life, ECAL (Vol. xx, pp. xx).





Muhanad H. Mohammed
mhm1@aber.ac.uk

Supervised by: Elio Tuci

Intelligent Robotics Group
Department of Computer Science

PRIFYSGOL ABERYSTWYTH UNIVERSITY

HPC WALES · CYMRU

Powering Success
Pweru Llwyddiant

Forum Discussion

HPC Wales User Group Meeting

Wednesday 1st April 2015

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government

User Group Forum - Questions

- Where would you like to see technical documents and information stored?
- What additional technical information would you like to see?
- How valuable do you find FAQs?
- Do you have any strong feelings regarding the future stage 2 of HPC Wales?
- Was gaining access to HPC Wales overly difficult? Why?
- Why are gateways not used more?

Q&A & Close