GridPP
UK Computing for Particle Physics

UNIVERSITY of GLASGOW

Future Directions for Computing in Particle Physics

RAL@50
13th Nov 2014

Prof. David Britton
GridPP Project leader
University of Glasgow

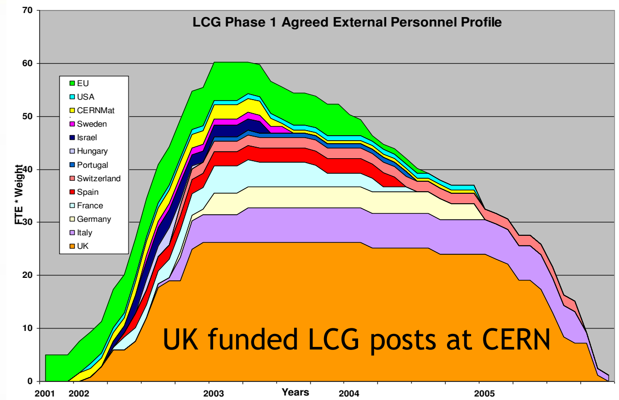
GridPP
UK Computing for Particle Physics

Foreword

- In terms of funding and physicists, Particle Physics is currently dominated by the LHC which has produced the largest data-sets in the field.
- Computing for Particle Physics is, therefore, also dominated by the LHC computing, so this is the focus of this talk.
- LHC computing in the UK is provided by the GridPP project; we also support many (~30) other groups but the LHC accounts for 90% of what we do.

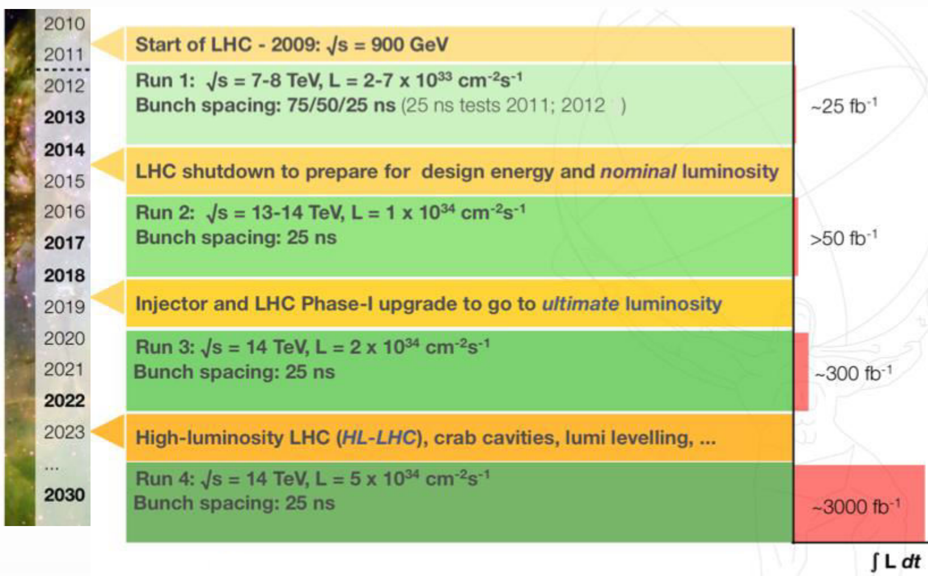
GridPP Mission

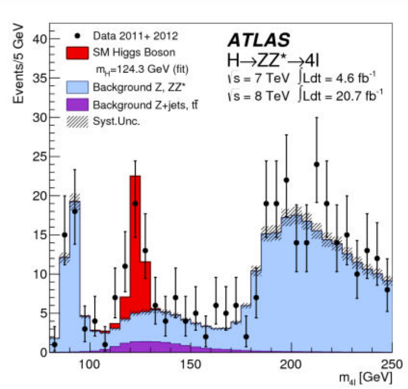
Mission: To deliver resources to the UK and worldwide particle physics community in accordance with the WLCG MOU, by means of a large-scale computing Grid in the UK.

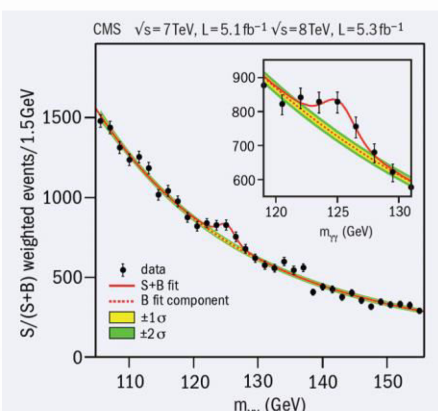


The UK kick-started WLCG in 2002 with a £5.6m investment.

2001 GridPP1 – *From Web to Grid*
2004 GridPP2 – *From Prototype to Production*
2007 GridPP2+ *(6-month extension)*
2008 GridPP3 – *From Production to Exploitation*
2011 GridPP4 - *Computing in the LHC era*
2015 GridPP4+ *(One year extension)*
2016 GridPP5 – *Computing beyond the Higgs*



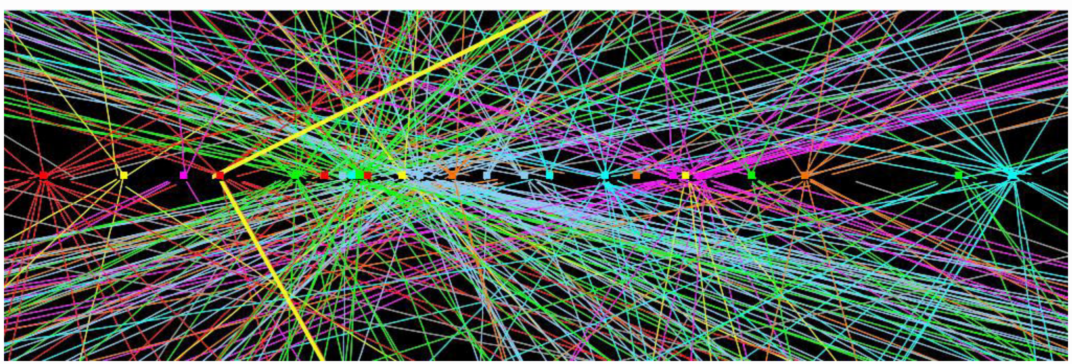


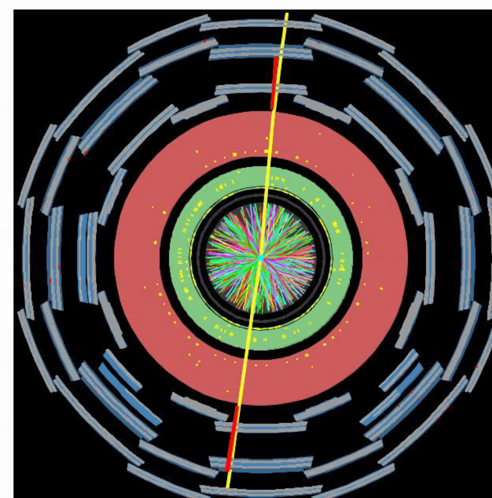


Future

- Current detectors designed for 10 years with <μ>=23.
- Performing well beyond this specification (<μ>=40 expected in Run-2), but have limited life and will not handle HL-LHC <μ>=140.

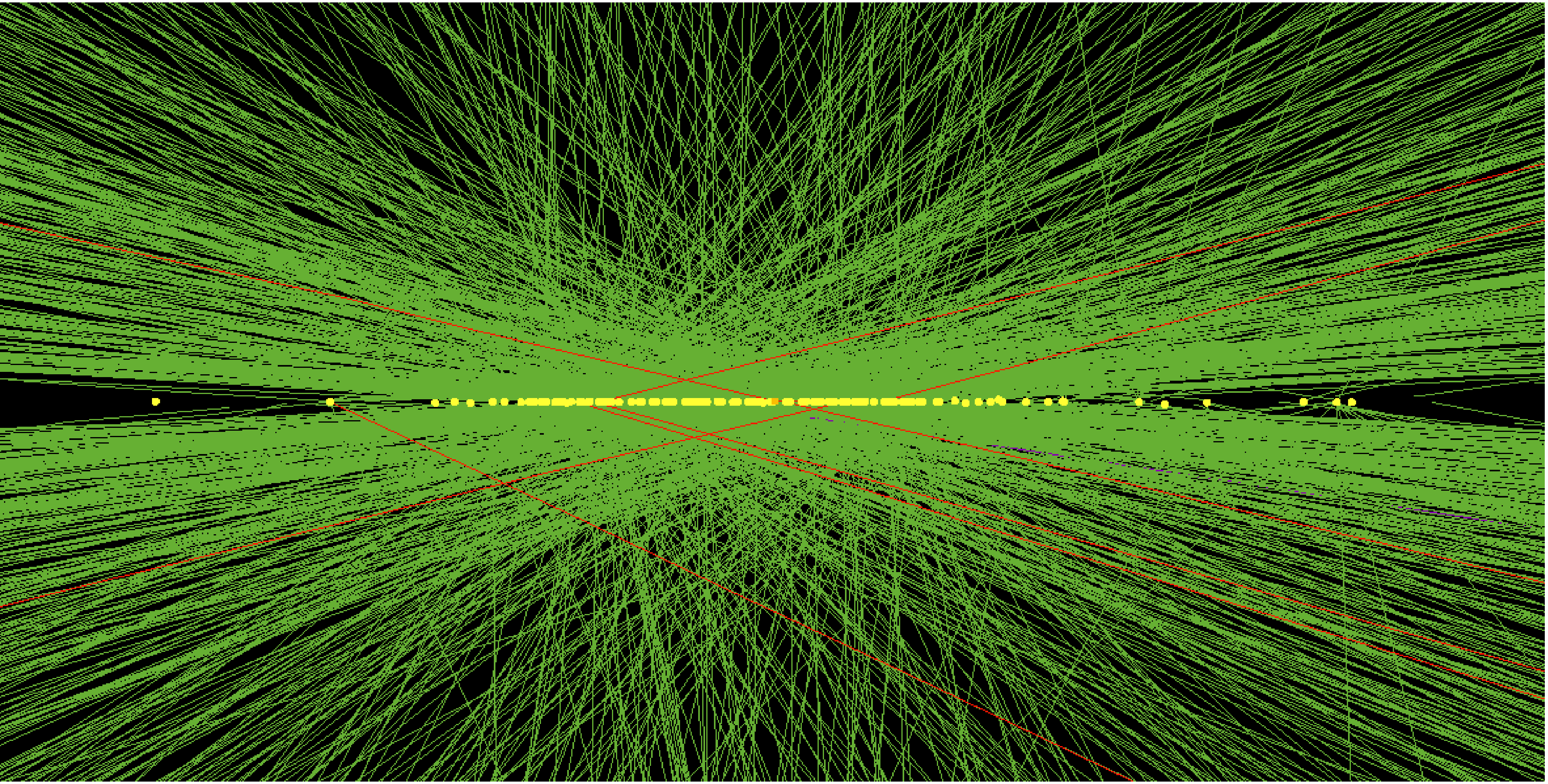
Z → μμ decay with 25 vertices (April 15th 2012)





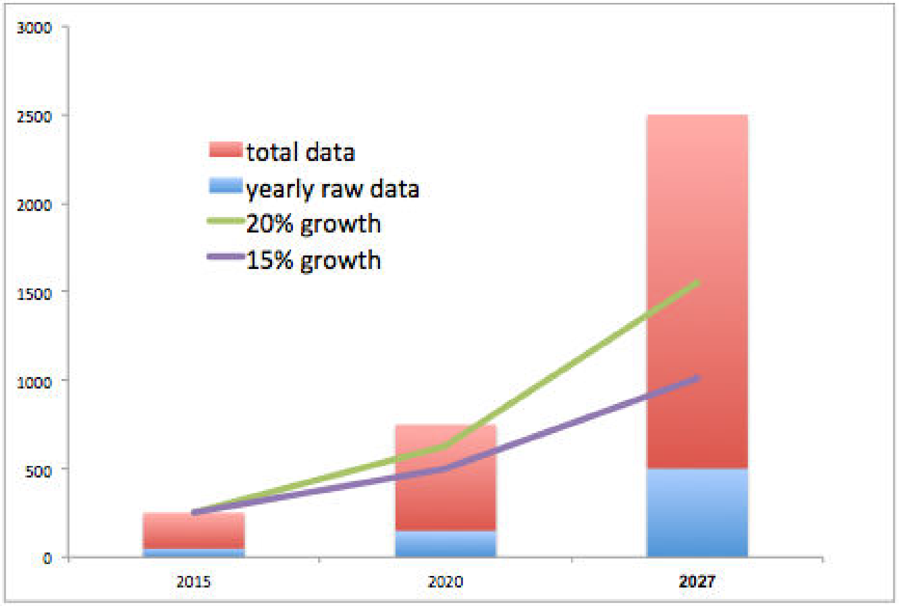
Future

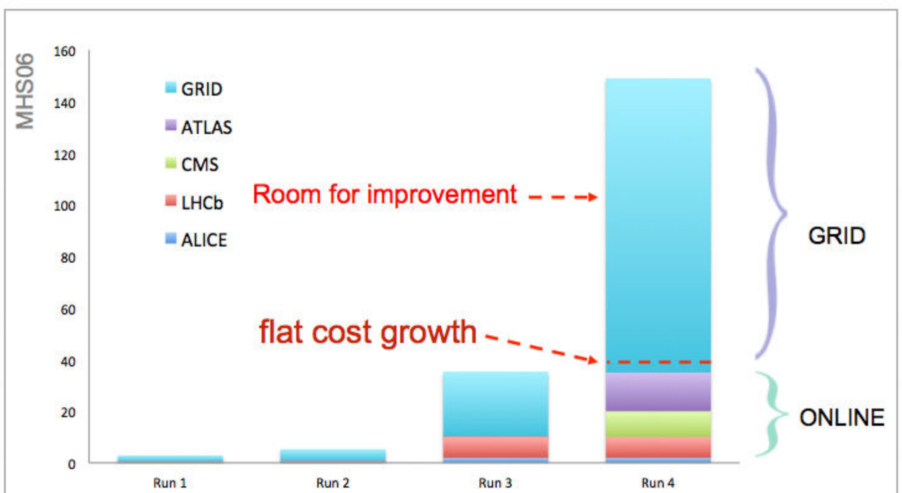
- Simulated Event Display at 140 PU (102 Vertices)



The Challenge

GridPP
UK Computing for Particle Physics



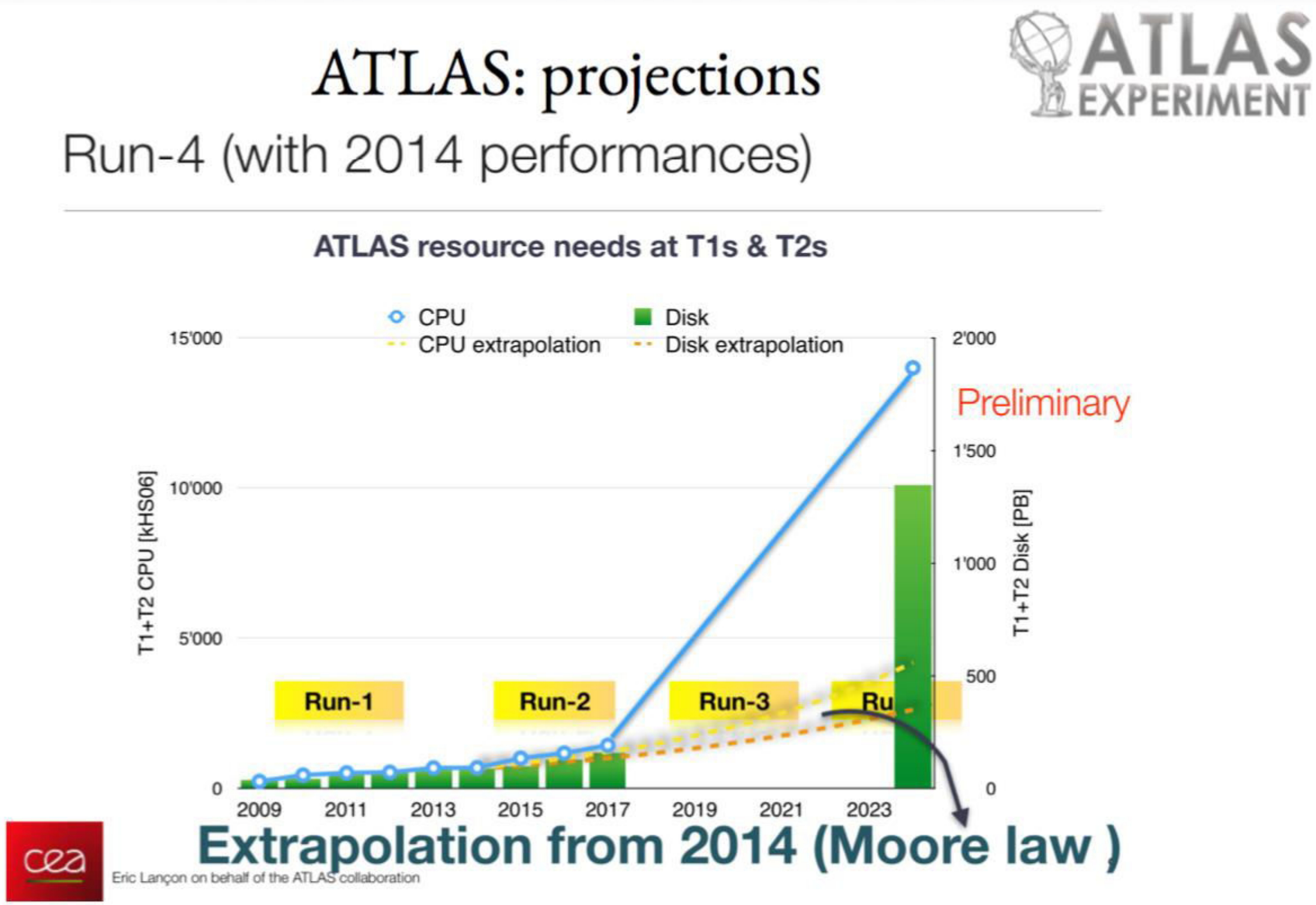


Volume: >200 PB at present (on disk) will grow by factor of >10x by Run-4.

Complexity: Pile up increasing from 23 to 140 by Run-4 increases computational problem super-linearly.

GridPP
UK Computing for Particle Physics

The Challenge



- Need to worry about disk and CPU usage for HL-LHC as well as access to disk (IO and capacity!).

How to meet the future requirements?

- W
  -
- W
  -
- Ge fr
  -

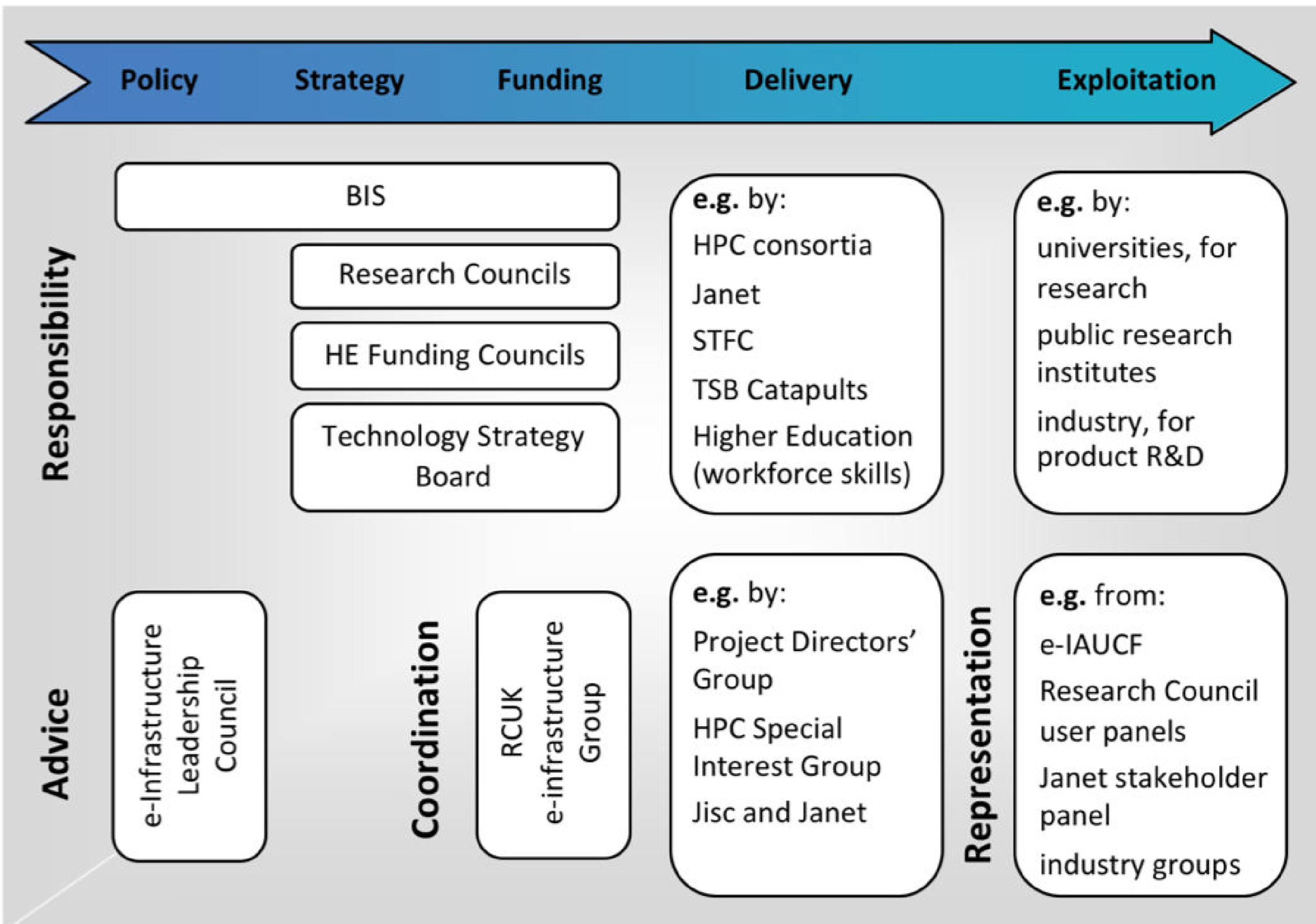
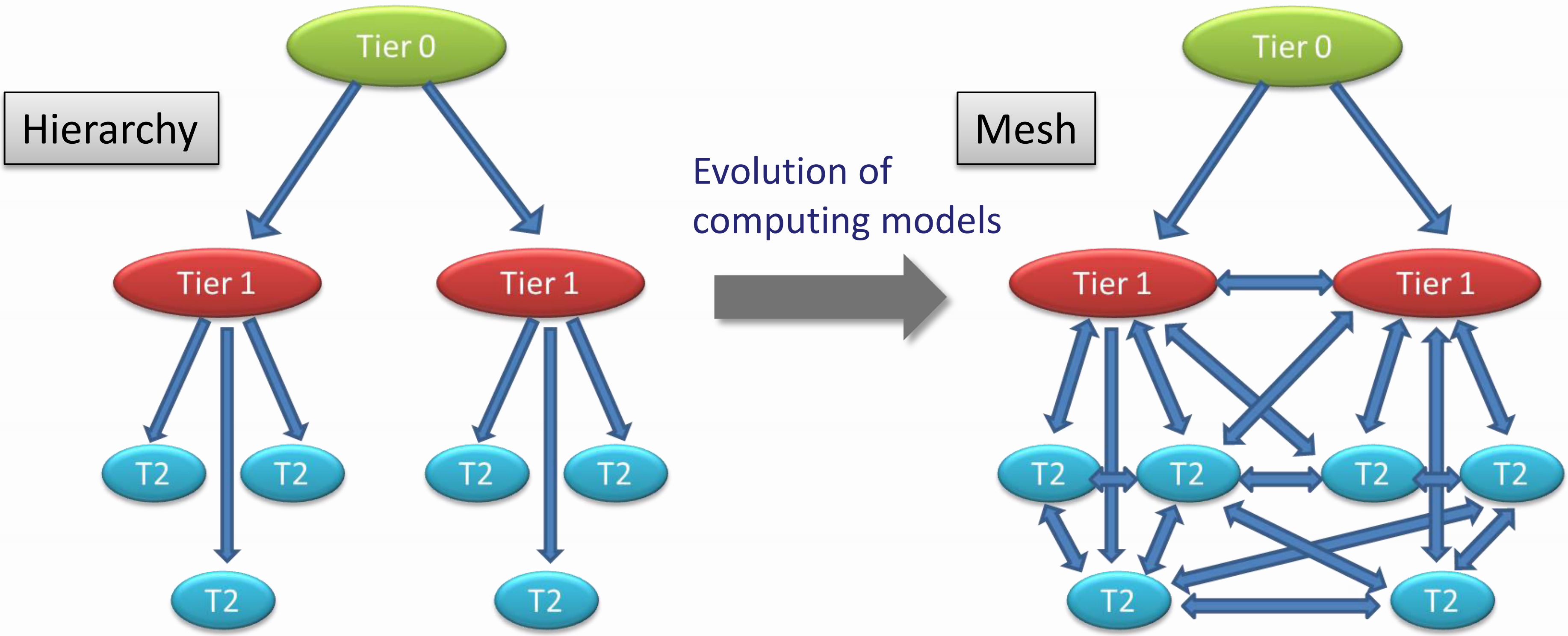


Figure 8: The e-infrastructure pipeline

Working Smarter



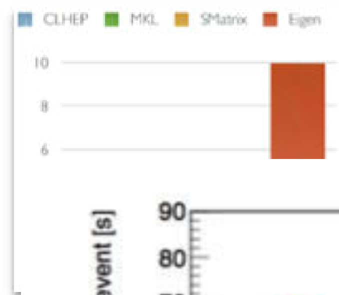
- Network capabilities and data access technologies have significantly improved our ability to use resources independent of location.

Improvements: ATLAS Example

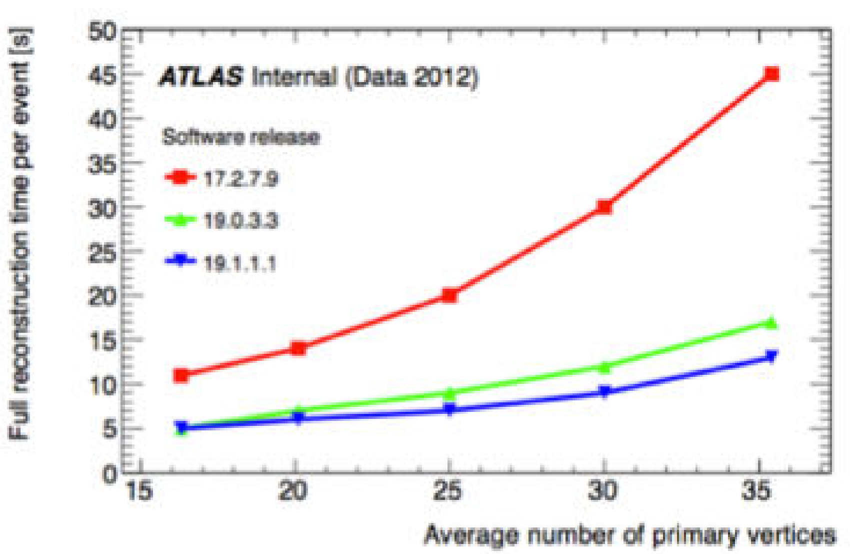
(Graeme Stewart)

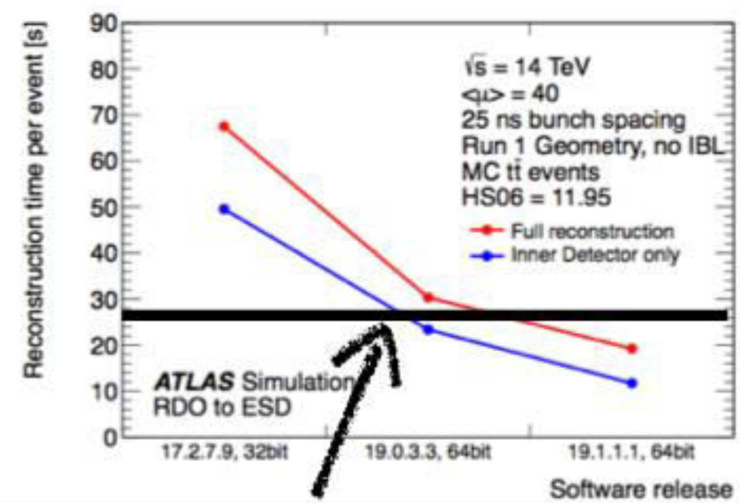
Run 2 Challenges: Reconstruction

- Run 2 sees the HLT output rate rise to 1kHz
  - Need to maintain single lepton triggers
- Tier-0 budget is around 11k cores



Linear Algebra

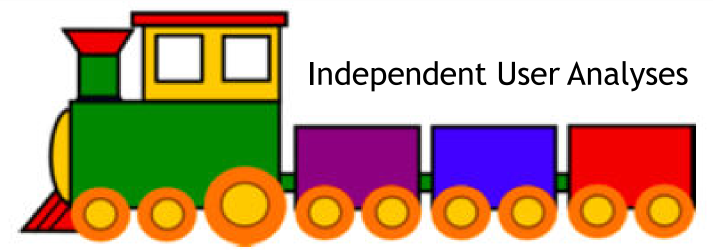


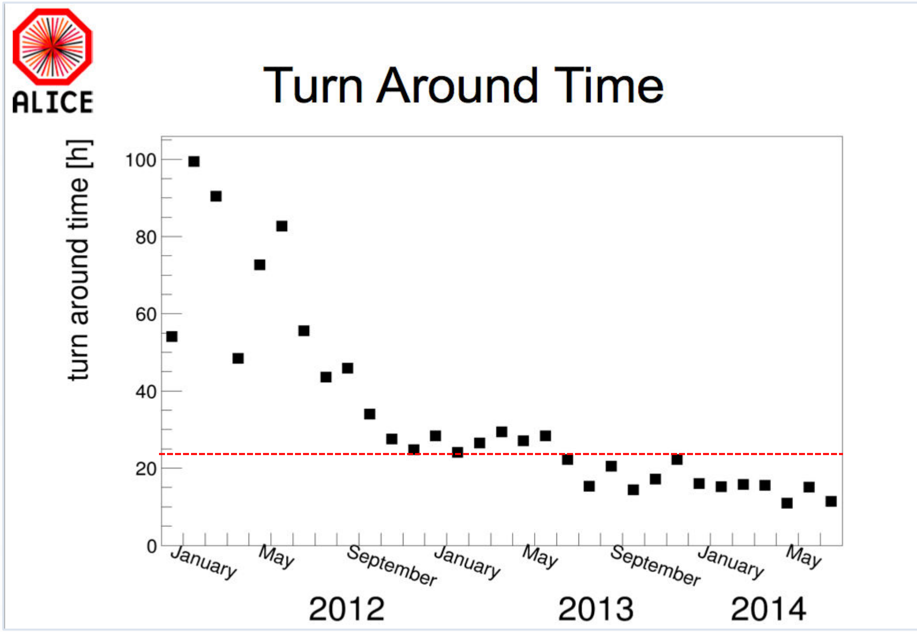


Tier-0 budget (μ=40, 0.4)

Analysis Trains

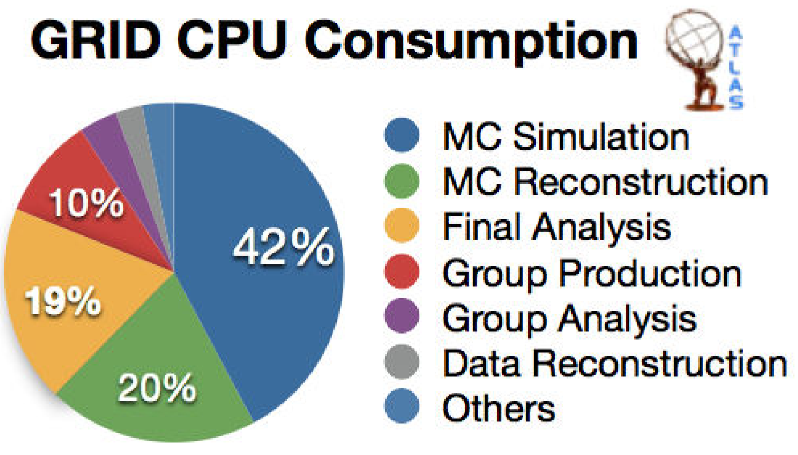
- In Run-1, we read the data 50x to produce 50 derived data-sets!
- Combine separate user analyses into a Centrally managed ‘Train’.
- Each ‘carriage’ is tested before departure and removed if it fails.
- Central book-keeping makes derived data more accessible to all in a standard way.
- Can be multiple trains (eg Fast and Slow).
- Reduces “chaotic” user analysis.

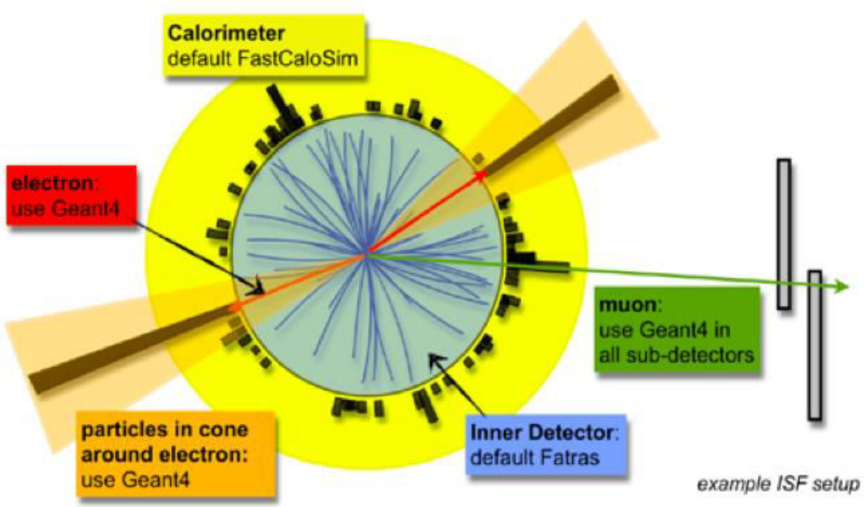




Monte Carlo

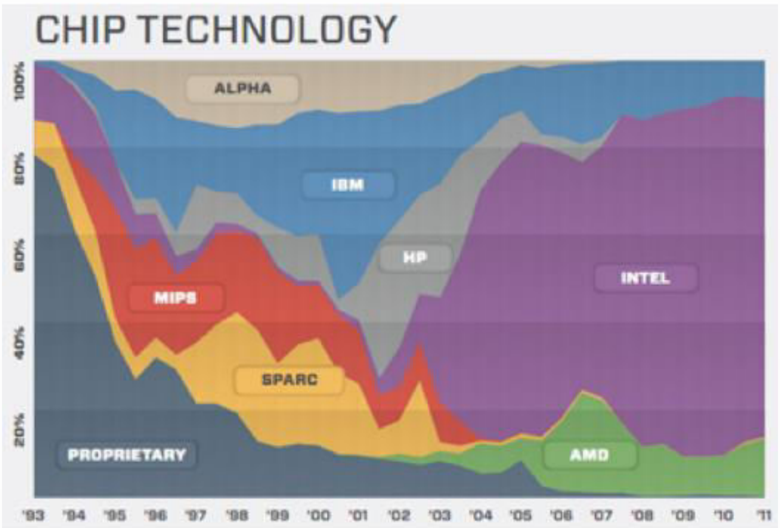
- GEANT-4 full simulation is very expensive: e.g. up to 1000s/evt for ATLAS.
- Fast simulations use techniques such as Smearing, Frozen-Showers, Parametric Responses but are not good enough for all situations.
- Development of an Integrated Simulation Framework allows an appropriate mixture of Fast and Full in the same event and x100 speed-ups are possible.

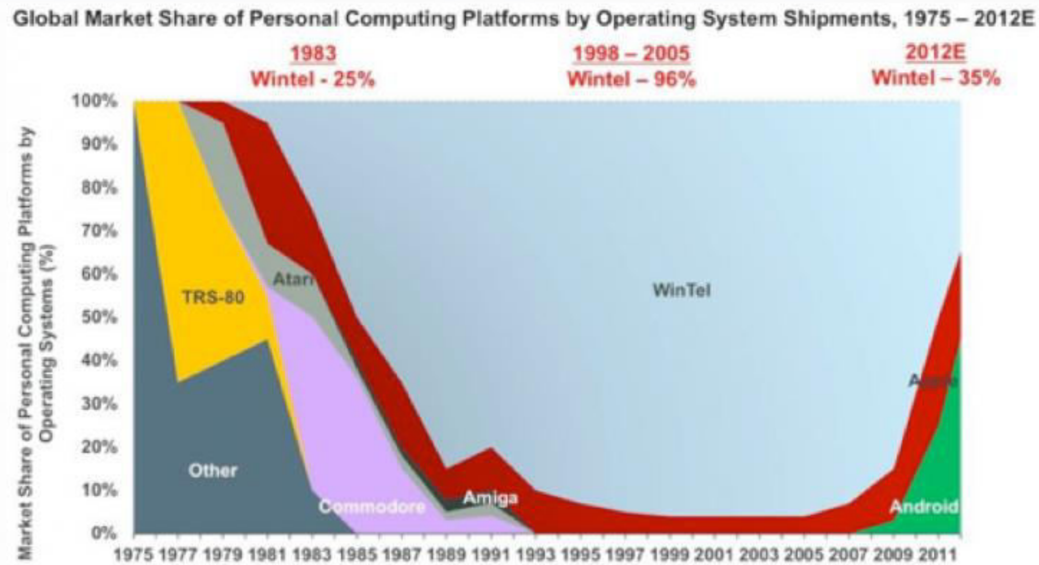




GridPP
UK Computing for Particle Physics

Hardware Trends

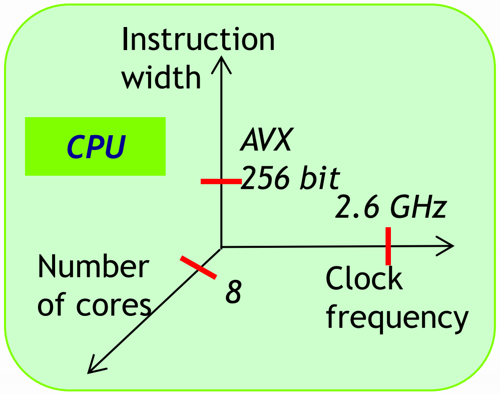


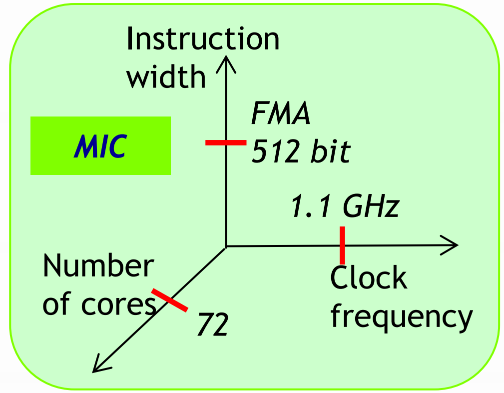


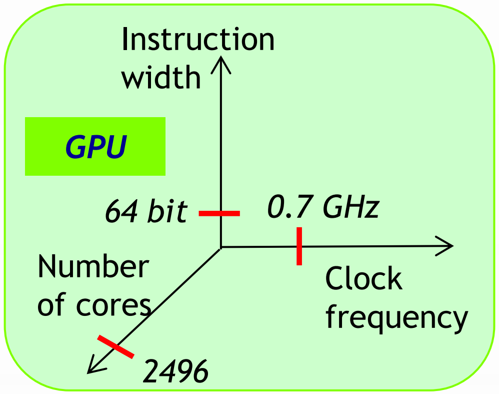
- HEP utilises commodity hardware which is driven by external forces.
- Key driver is now power-consumption for both:
  - Portable devices, which are then supported by
  - Backend data-centre machines providing compute and data services.
- X86 mono-culture is breaking down.
  - ARM64, PowerPC, GPUs, etc.

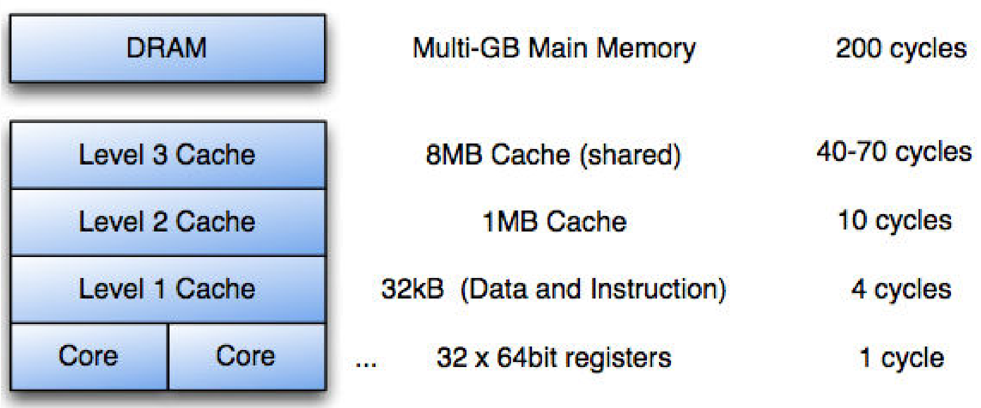
Reduced clock speed

Hardware Directions









Costly to move data from main memory – can far outweigh benefits of SIMD. The software design needs to understand what data is needed in the caches. No abstract solution works for all data on all hardware.

Using More Cores

- Xeon Grid Servers: 2-4 GB/core
  - Hyper-threaded (2): 1-2 GB/core
- Xeon-phi (MIC): 256 MB/core
  - Hyper-threaded (4): 64 MB/core
- Tesla K40 (GPU): 4MB/core

“Memory Wall”
Typical HEP job requires ~2GB memory. Can’t simply run independent jobs on each core.

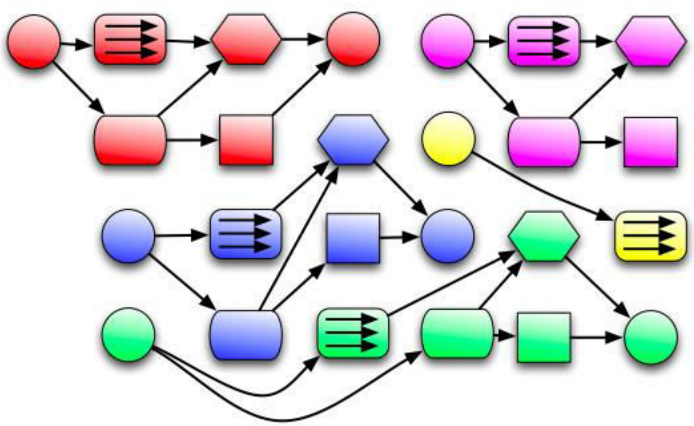
→ Reduce memory use via “threading”

→ Memory savings can be huge (“heap” is shared)

→ But programming more difficult (races, deadlocks etc)

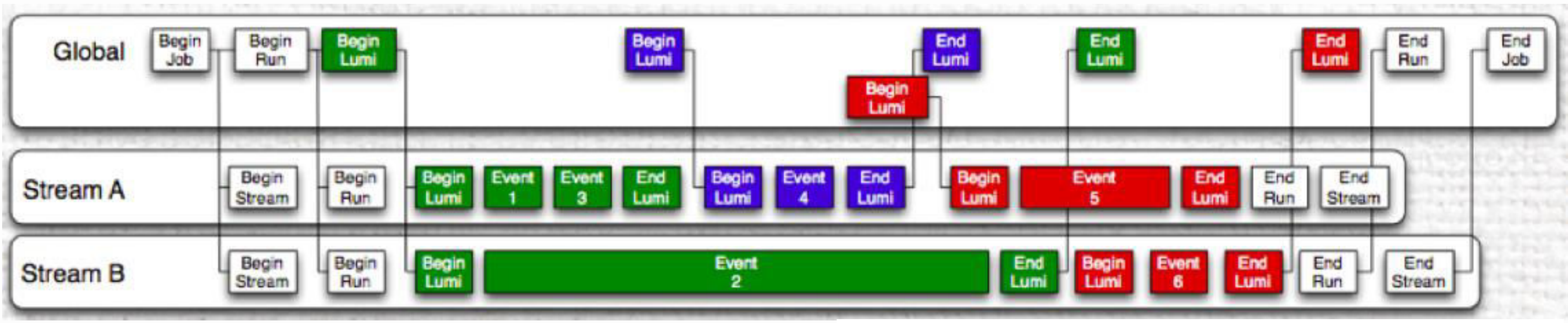
→ Back-porting to millions of lines of legacy code hard.

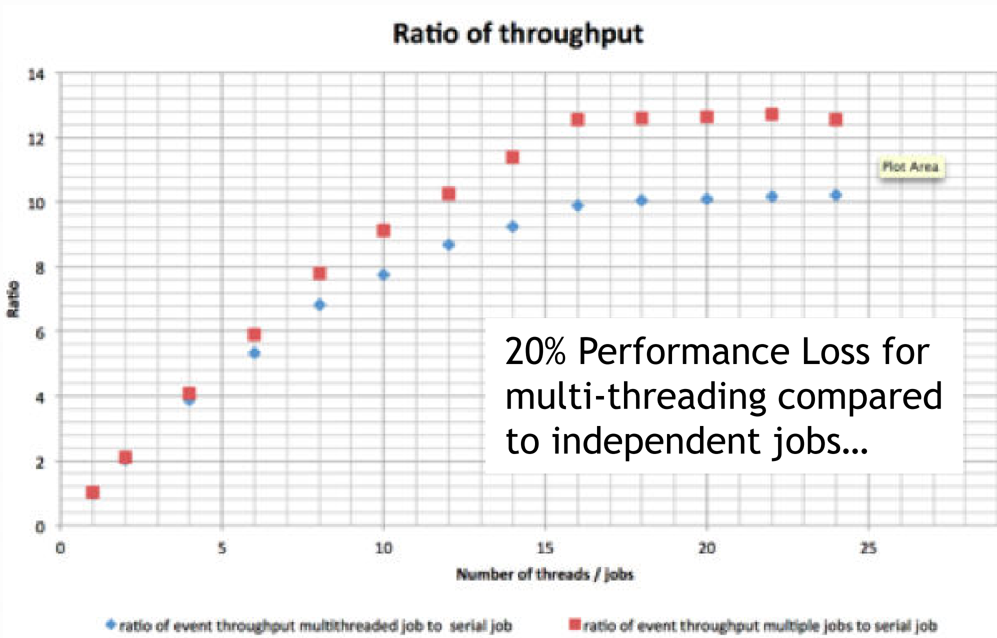
GaudiHive



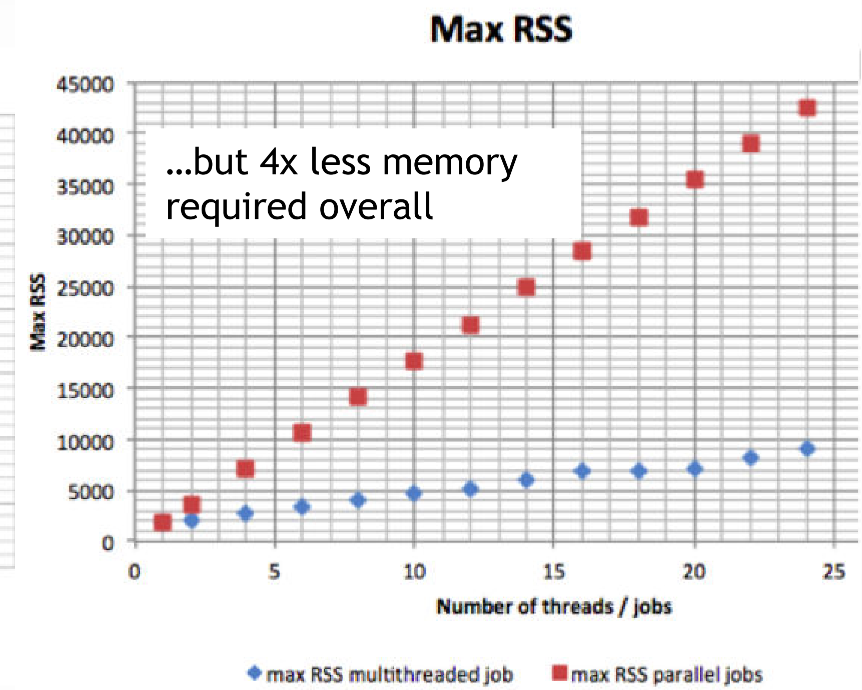
Colours represent different events, shapes different algorothms

CMS Multi-Threading



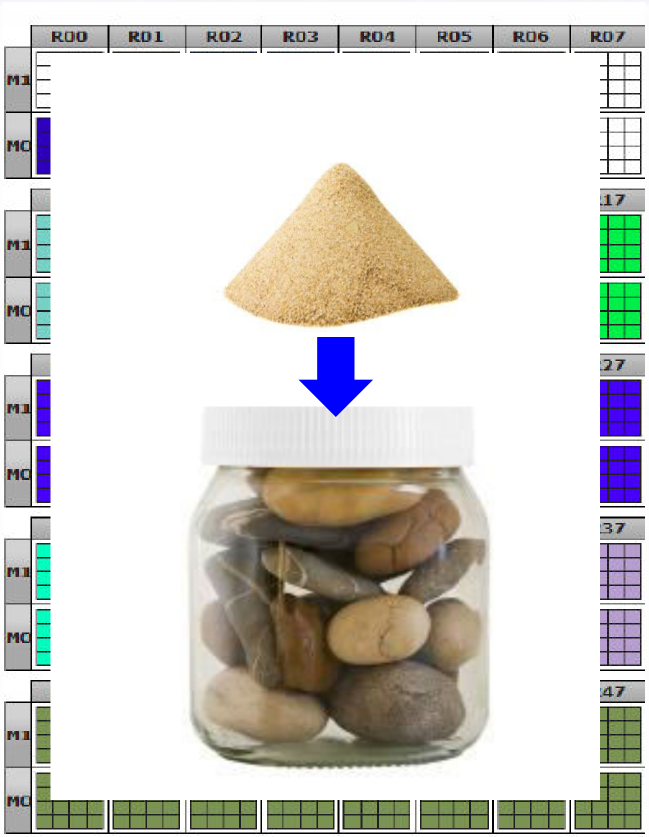


20% Performance Loss for multi-threading compared to independent jobs…

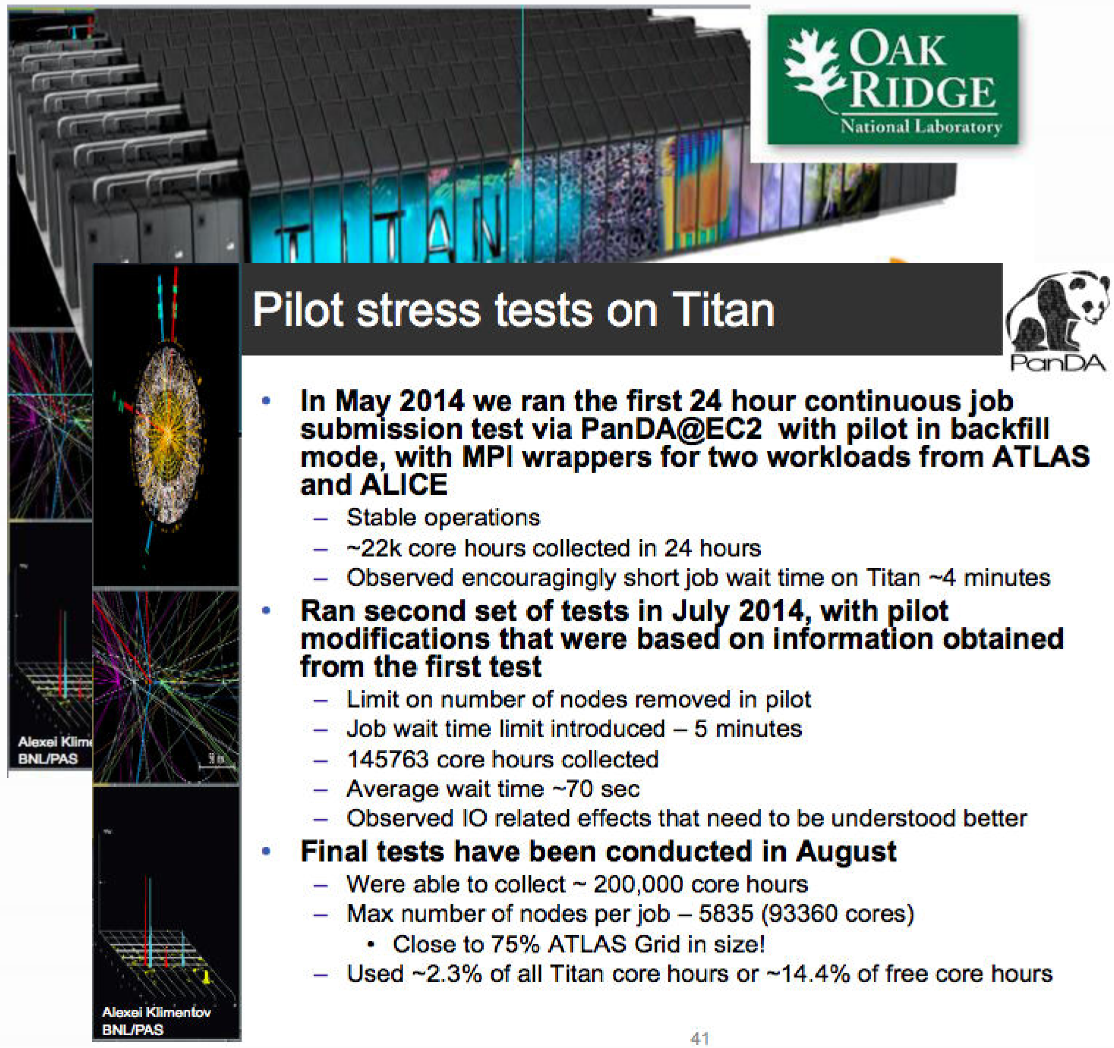


…but 4x less memory required overall

Getting Help



"Full" HPC machine: Only 85% utilised because shortest job in queue requires larger partition than is available.



UK/EU Eco-System

Our community are involved in many of the new UK/European/Global initiatives:

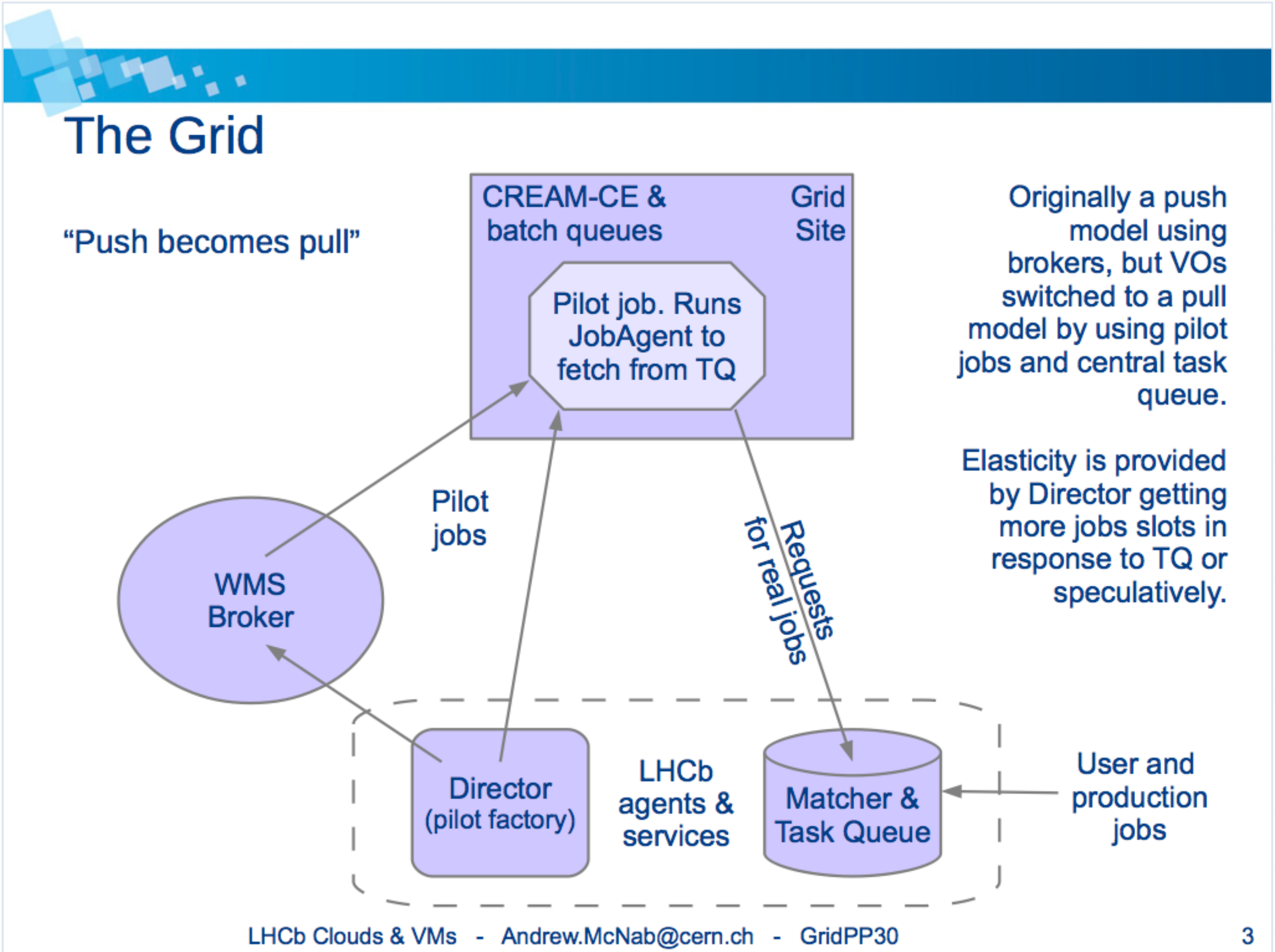
- EGI Engage (H2020)
- VLData (H2020)
- Zephyr (H2020)
- HEP Software Foundation
- EU-T0 Initiative
- …and more.

egi

DIRAC
THE INTERWARE

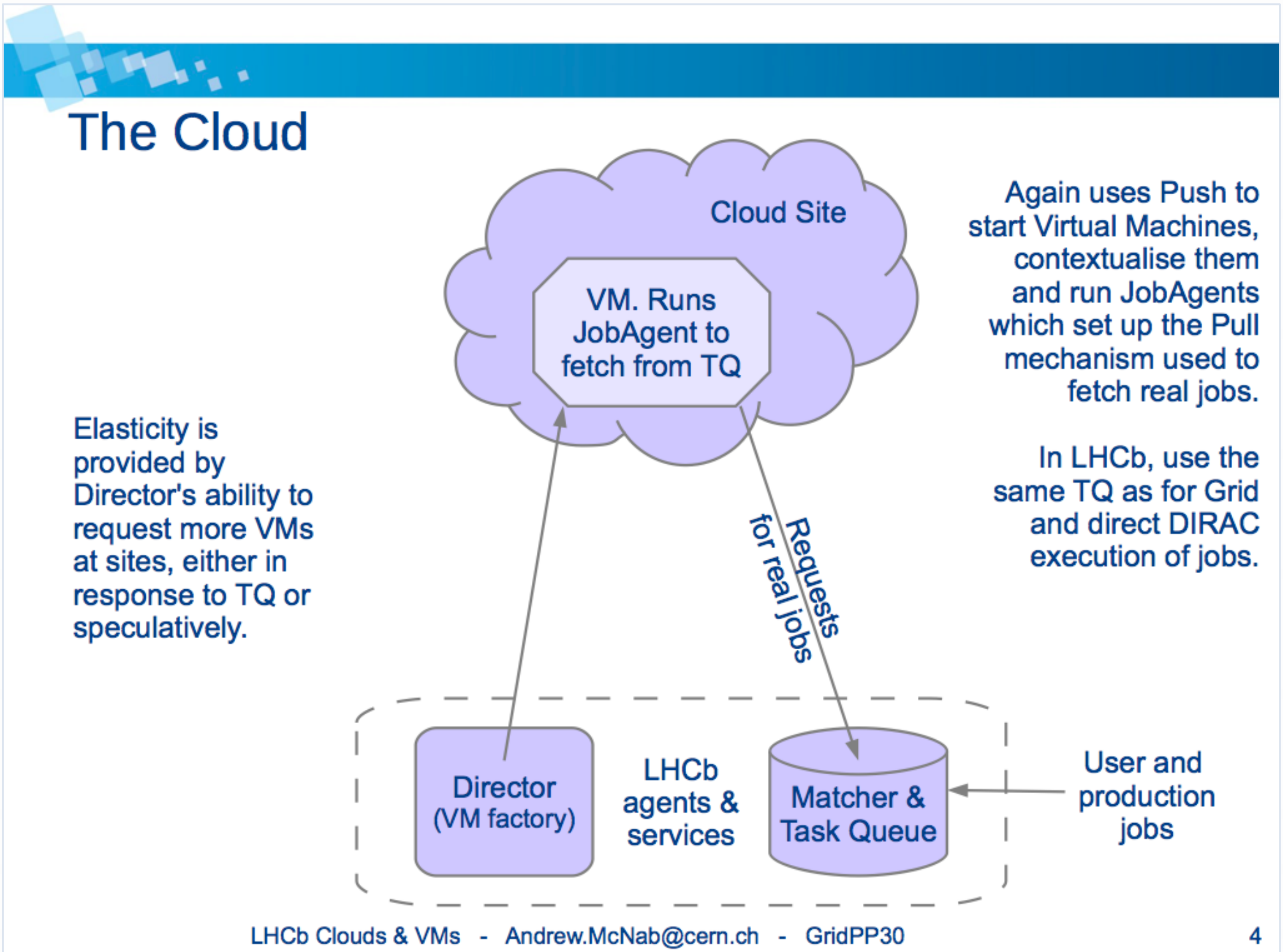
THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION
HORIZON 2020

Accessing Resources



GridPP
UK Computing for Particle Physics

Accessing Resources



Grid sites increasingly using Cloud Technology internally to manage resources (e.g. CERN T0 split between Geneva and Budapest). But cloud interface is not typically exposed externally.

Grid of Clouds?

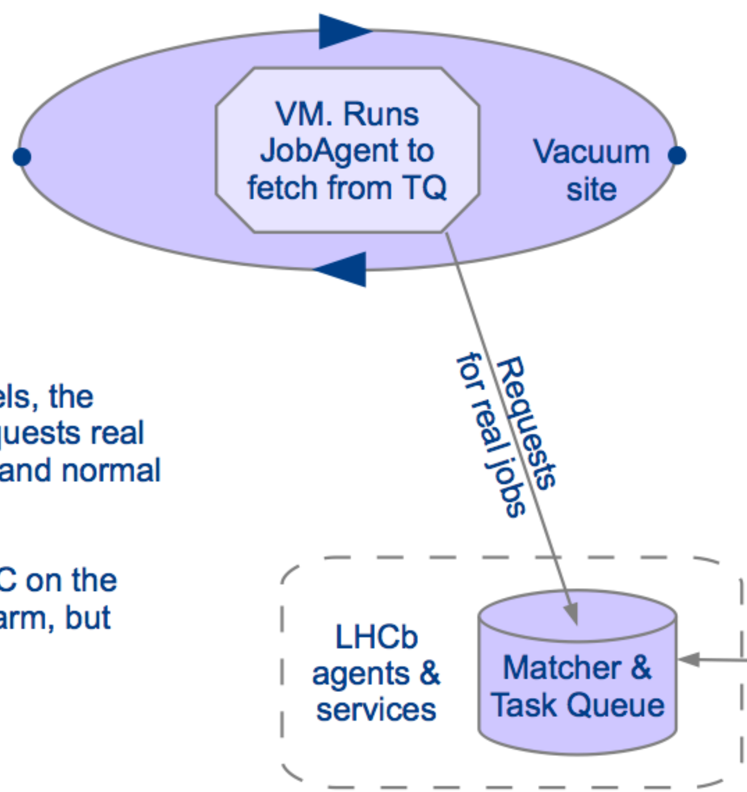
Accessing Resources

"The Vacuum"

Instead of being created by VOs, the Virtual Machines appear spontaneously "out of the vacuum" at sites.

As with the other models, the JobAgent runs and requests real jobs from the Matcher and normal Task Queue.

This is similar to DIRAC on the existing LHCb online farm, but with VMs added.



User and production jobs

Hypervisors/hosts can run VMs for particular VOs depending on work available and target shares for each VO.

Elasticity comes directly from matching resources to demand from TQ.

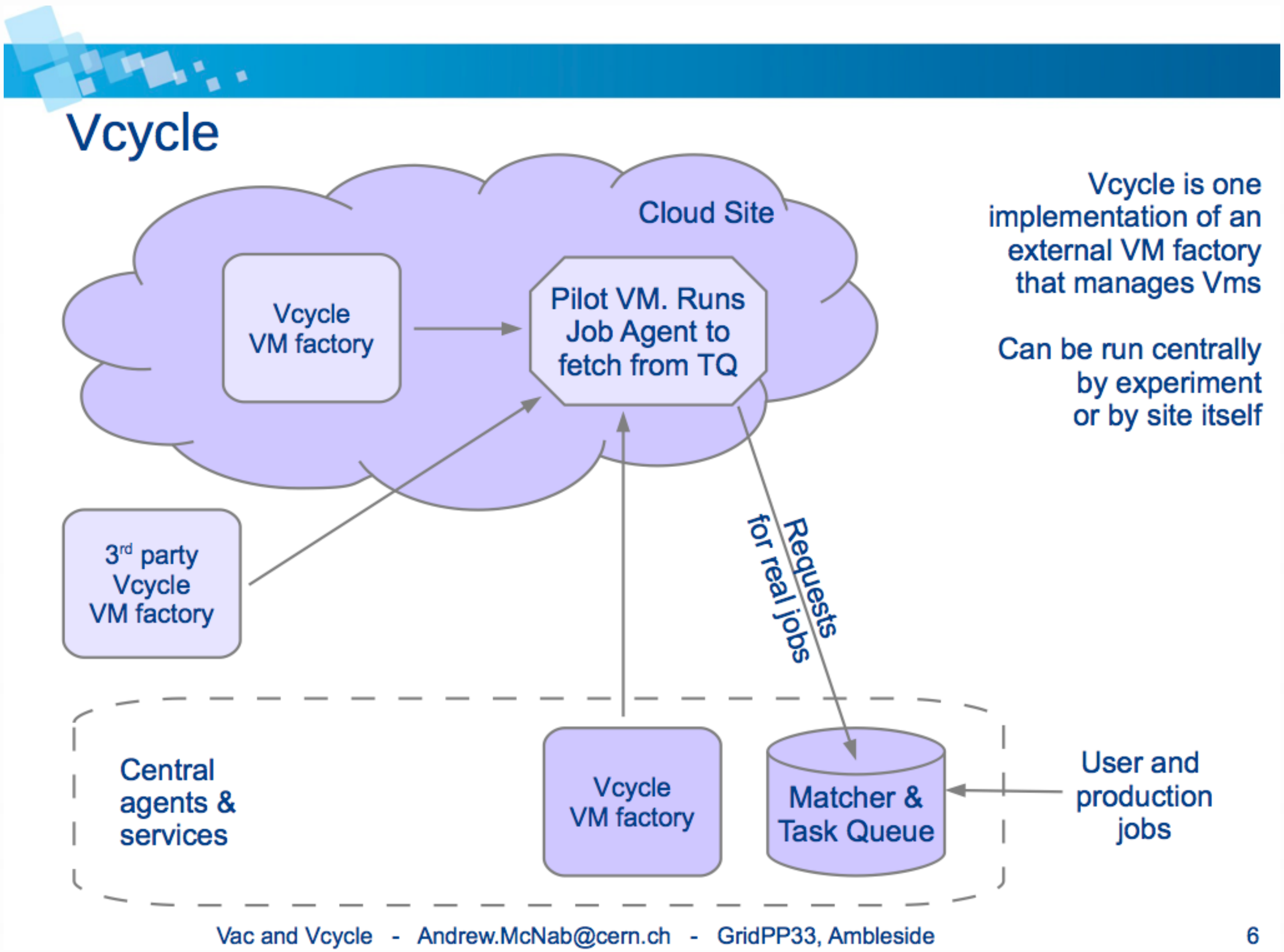
LHCb Clouds & VMs - Andrew.McNab@cern.ch - GridPP30 5

VMs created and contextualised by the sites themselves.

Suitable for sites dedicated to a few VOs.

Containers instead of Virtual Machines?

Accessing Resources



Applies VAC idea to Cloud Resources (OpenStack).

Can be run by site, experiment, or 3rd party.

Could we use VAC or VCycle to flexibly integrate Tier-3 resources?

Summary

- HEP computing resource needs are large and will continue to grow to meet the volume and complexity of LHC data.
- Evolution of Computer Hardware is making it harder to realise the ‘Moore’s Law’ type growth.
  - We can cope with Run-2 until 2020
  - But we need to prepare for runs beyond that
  - A lot of work ongoing/needed to adapt frameworks
  - More work is needed; e.g. within Algorithms.
  - Work → Manpower which is increasingly expensive c.f. hardware
- Joining up the e-infrastructures can help.
- Evolving the Grid to use Cloud technologies where appropriate.