

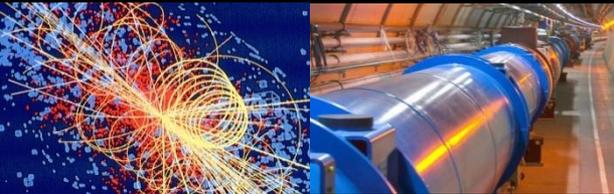
# HOW19

## HSF-OSG-WLCG workshop summary

Ian Collier

ian.collier@stfc.ac.uk

STFC Rutherford Appleton Laboratory  
GDB, April 3<sup>rd</sup> 2019



Joint HSF, OSG & WLCG Meeting

# HOW 2019

MARCH 18-22, 2019

Jefferson Lab • Newport News, Virginia, USA

### SCIENTIFIC ORGANIZING COMMITTEE

Ian Bird, CERN  
Simone Campana, CERN  
Tim Cartwright, University of Wisconsin-Madison  
Ian Collier, STFC

Michel Jouvin, LAL-CNRS  
David Lange, Princeton University  
Graeme A. Stewart, CERN  
Frank Wuerthwein, UC San Diego

LOCAL ORGANIZATION PROVIDED BY JEFFERSON LAB

# Overview

- **HSF** (HEP Software Foundation) and **WLCG** had a first combined workshop in Naples a year ago
  - Very successful, will be the standard now
- In 2019, at JLab, we were also joined by the **OSG** (Open Science Grid) all-hands meeting
- 246 registered participants
  - largest meeting JLab has hosted by some margin
- Excellent hospitality & enjoyable social events
  - The Mariner's Museum was a particular highlight

# Scientific Program

<https://indico.cern.ch/event/759388/timetable/#all.detailed>

- Plenary sessions on Monday and Friday
- 2-5 parallel tracks at different times on the other days
  - WLCG
  - HSF
  - OSG
- Some overlap in interest between the tracks
  - Sadly, not possible to take it all in
  - This is a necessarily partial view

# Monday plenary sessions

- Introductions to JLab
- Input from communities/experiments on current and future computing challenges
  - LHC experiments: [ALICE](#), [ATLAS](#), [CMS](#), [LHCb](#)
  - DUNE, Belle II
  - Dark matter
  - EIC
  - Photon/neutron sources
  - LSST
  - LIGO/VIRGO
  - IceCube
- Evolution of the WLCG collaboration

# Tuesday parallel tracks

- WLCG (+ HSF)
  - Technology watch on computing, storage and networking
  - HPC centers, clouds
  - Expt. software frameworks on heterogenous resources
  - Authentication and Authorization Infrastructure evolution
  - Security operations
- HSF
  - GPU and other accelerator technologies
- OSG
  - OSG status
  - OSG communities

# Wednesday parallel tracks

- WLCG
  - Resource and cost estimates
  - Benchmarking
  - Performance evaluation
  - Storage modeling and data popularity
  - DOMA (Data Organization, Management and Access)
    - WG topics: 3<sup>rd</sup> party copies; quality of service; access
    - Rucio, DIRAC
    - Data provisioning for HPCs and clouds
- HSF
  - Simulation, analysis, reconstruction, machine learning
- OSG
  - Infrastructure & resources

# Thursday parallel tracks

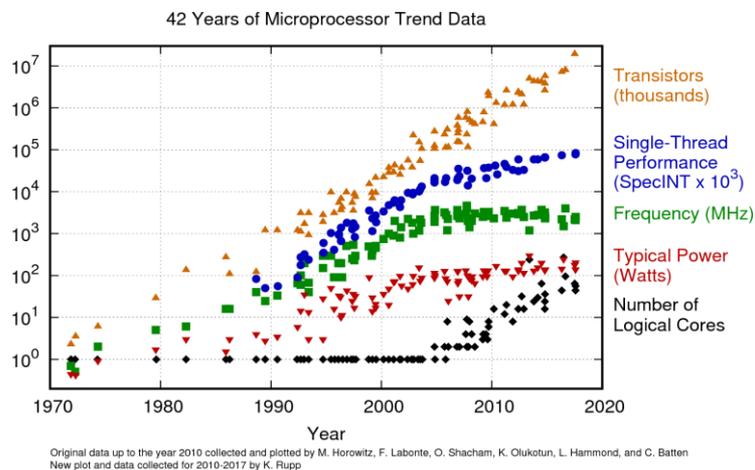
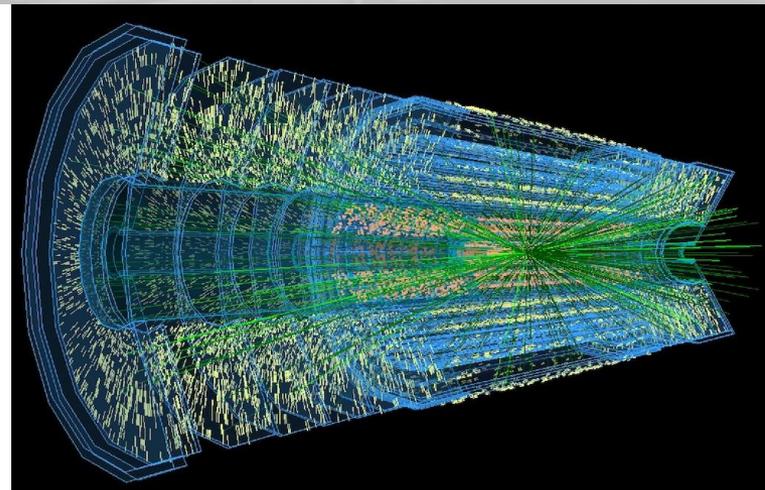
- HSF
  - Present and future technologies for data analysis
  - Notebooks, Python, ROOT, vectorization, ...
  - Training
  - Performance monitors/profilers, static analyzers
  - Packaging
- WLCG
  - Information system evolution
  - Operational intelligence
  - Long term future of the storage services at T2s
  - Lightweight sites
- OSG
  - USCMS Facilities
  - Researcher training

# Friday plenary session

- Forward look and close out
  - DE Funding Initiative
  - UK IRIS Project
  - US IRIS-HEP Project
  - The Future of Scientific Computing

# HSF Impressions

- Two drivers for software and computing
  - Ambitious experimental physics programme
    - LHC Run3 for ALICE and LHCb
    - HL-LHC for ATLAS and CMS
    - Belle II in data taking
    - DUNE has test beam data, firming up ideas
  - Technology evolution
    - CPU improvements slowing
    - Complex latency issues with data access
      - Caches, Storage, Network, ...
    - Compute accelerators becoming prevalent
      - Current and future generations of HPCs
      - Managing data rates close to the detector with software triggers



# HSF Session Highlights

- Software on Accelerators
  - Significant work now archived in the community:
    - ALICE tracking in TPC; LHCb Allan project to port the whole of HLT1 to GPU
    - Event generation on GPU looks possible; Simulation looks very hard
  - General lesson: *data layout matters a lot – make it simple and portable*
  - General frustration: *no obvious toolkit exists for maintaining heterogeneous code*
- Simulation
  - Speed is of the essence – approximate methods are needed
  - Machine Learning is helping, but details are really tricky
  - Stochastic process – not easy to adapt to modern CPU architectures
- Reconstruction
  - Real Time Analysis (close to data taking) is driving fast calibration and high quality reconstruction to throw away raw data
    - Accelerators are finding use here

# HSF Session Highlights

- Analysis and PyHEP
  - Very diverse landscape with huge dynamic range
    - Balance flexibility against costs of storage and (re)calculation
  - New ideas from data science are important, toolkit approach
  - Imperative and functional approaches look attractive – technology agnostic
- Education and Training
  - New initiatives needed to equip people with the right skills – better and wider training needed
  - LHCb StarterKit leads the way – being adopted across different experiments
    - Common training material within HEP, and even with Carpentries, looks possible
- Software Tools
  - Ripe area for collaboration in software profiling and analysis as well as packaging

# HSF Perspectives

- Software covers a high range of tasks for HEP
  - Sharing ideas is profitable
  - Sharing code is much harder, but pays off in the long term
    - E.g. ACTS, DD4hep, VecGeom/VecCore
- New working groups put together great sessions during the meeting
  - Really generating community engagement
  - This is just the start of the process
    - Next [HSF meeting](#) will discuss future perspectives (11 April)
- CWP Roadmap was [published](#) in Computing and Software for Big Science – “The end of the beginning”
- HOW2019 took us to the next phase and was a really success

# Selected observations (1)

- Can no longer assume increases in performance/capacity
  - 20% yearly increase “for free” has not held in recent years
- ATLAS and CMS Run 4 requirements are driving a lot of the activities
  - With benefits already planned for Run 3 and for other experiments and communities, e.g. through Rucio
- Other experiments and communities have requirements at least comparable scale
  - WLCG will evolve toward more explicit forms of collaboration with related communities
    - Profit better from shared efforts and investments
    - Speak to funding agencies with a common voice
- Funding agencies have finally started recognizing the importance of sustainable SW development for big science (CWP played a big role.)

# Selected observations (2)

- Use of HPC centers and clouds will increase
  - Should become easier to use – more organisation on our side will help
- Use of GPUs, other accelerators and machine learning will become more significant
- Authentication and authorization becoming easier
  - Federated identities instead of certificates
- The organization, management and access of big data will shift toward *data lakes*
  - WLCG will probably be a hybrid infrastructure for many years to come
- Sites will be able to choose between several ways to make their service deployment and operations more lightweight

We have an interesting decade ahead of us! 😊

# Questions?