# Advancing in HEP SW and Computing: CWP Lessons and Future Work

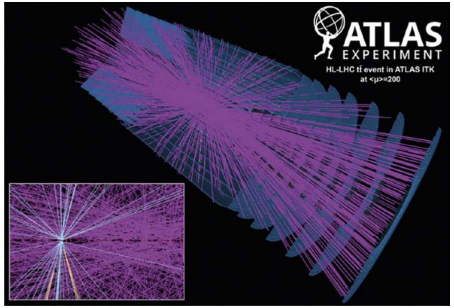
Michel Jouvin - CNRS/LAL

4th CERN Scientific Computing Forum - 2018-02-23

# Disclaimer

- Apologies for not being able to be at CERN (family problems)
- This presentation is a follow-up for the presentation/discussion introduced by G. Stewart at the previous Scientific Computing Forum, October 27
  - Happened a few days after the release of the first CWP draft, with an important focus on the CWP community process itself
- This presentation is more focused on what has been delivered finally and on the future
  - For details on the HEP Computing Challenges and the CWP process, refer to the previous presentation

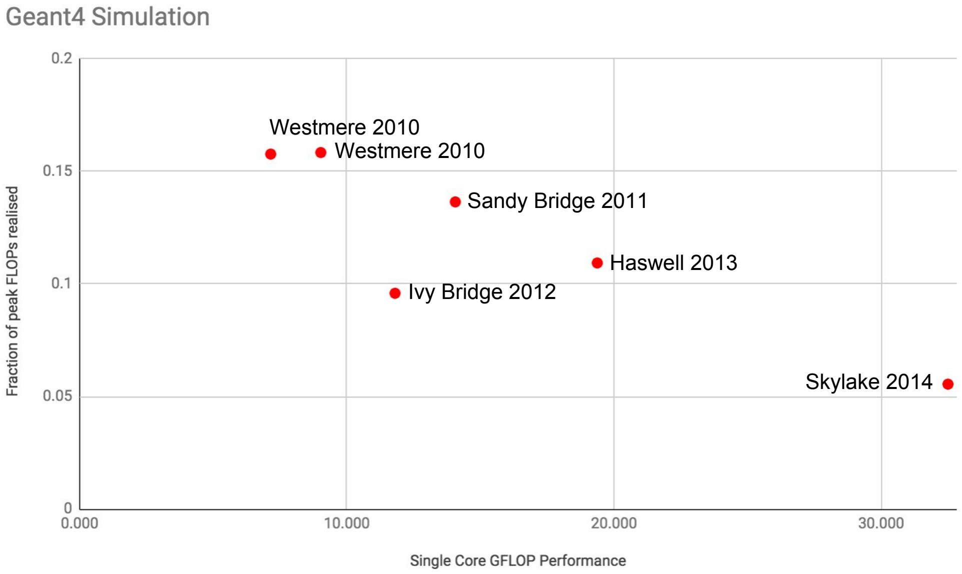
# Quick Recap: Software Challenges for HL-LHC




- Pile-up of ~200 ⇒ particularly a challenge for charged particle reconstruction
- With a flat budget, Moore's lawish improvements are the real maximum we can expect on the HW side
- HEP software typically executes one instruction at a time (per thread)
  - Since ~2013 CPU (core) performance increase is due to more internal parallelism
  - x10 with the same HW only achievable if using the full potential of processors
    - major SW re-engineering required (but rewriting everything is not an option)
  - Co-processors like GPUs are of little use until the problem has been solved
- Increased amount of data requires to revise/evolve our computing and data management approaches
  - We must be able to feed our applications with data efficiently
- HL-LHC salvation will come from software improvements, not from hardware

# How is our Code Doing? Simulation on 5 years of Intel CPUs

- Fraction of the potential floating point performance we use has been dropping over time
- CPU manufacturers add wider vectors that we do not take advantage of, or deep pipelines where cache misses are very costly
- Confirms what we have long suspected about the growing performance gap on modern architectures




Private ttbar event test in the ATLAS detector with Geant4 10.1

# HEP Software Foundation (HSF)

- The LHC experiments, Belle II and DUNE face the same challenges
  - HEP software must evolve to meet these challenges
  - Need to exploit all the expertise available, inside and outside our community, for parallelisation
  - New approaches needed to overcome limitations in today’s code
- **Cannot afford any more duplicated efforts**
  - Each experiment has its own solution for almost everything (framework, reconstruction algorithms, …)
- HSF already started with a number of workshops and working groups on common topics (packaging, licensing, analysis)
- HSF established in 2015 to facilitate coordination and common efforts in software and computing across HEP in general
  - Our philosophy is bottom up, a.k.a. *do-ocracy*

# CWP Process Reminder

- Formal charge from the WLCG in July 2016
- Kick-off workshop 23-26 January 2017, San Diego
- Groups held workshops and meetings in the subsequent months
    - Broadening the range of participation, often with non-HEP experts participating
- Workshop in Annecy 26-30 June started to draw the process to a close
- Both workshops involved ~100 people, mainly US and EU
    - Total number of people involved in the writing process of the various topical papers was about 250

# CWP - Making a Roadmap for the Future

- Editorial Board set up in September, with the aim of encompassing the breadth of our community
    - Wide regional/experimental representation
- First draft released Oct. 20, second draft Nov. 17
- These drafts elicited a **substantial response from the community**, leading to many improvements
    - **100s of commenters**, some sections deeply rewritten
- Final version of the document published arXiv: 1712.06982 on 20 December (as announced in Oct. !)
    - Likely publication into the Computing and Software for Big Science journal created last year
- Several topical WG chapters giving significant additional details are starting to be pushed to arXiv

- Predrag Buncic (CERN) - ALICE contact
- Simone Campana (CERN) - ATLAS contact
- *Peter Elmer (Princeton)*
- *John Harvey (CERN)*
- *Benedikt Hegner (CERN)*
- Frank Gaede (DESY) - Linear Collider contact
- Maria Girone (CERN Openlab)
- Roger Jones (Lancaster University) - UK contact
- *Michel Jouvin (CNRS/LAL)*
- Rob Kutschke (FNAL) - FNAL experiments contact
- *David Lange (Princeton)*
- Dario Menasce (INFN-Milano) - INFN contact
- Mark Neubauer (U.Illinois Urbana-Champaign)
- *Eduardo Rodrigues (University of Cincinnati)*
- Stefan Roiser (CERN) - LHCb contact
- *Liz Sexton-Kennedy (FNAL) - CMS contact*
- Mike Sokoloff (University of Cincinnati)
- *Graeme Stewart (CERN, HSF)*
- Jean-Roch Vlimant (Caltech)

*(italic: core team aka "ghost writers")*

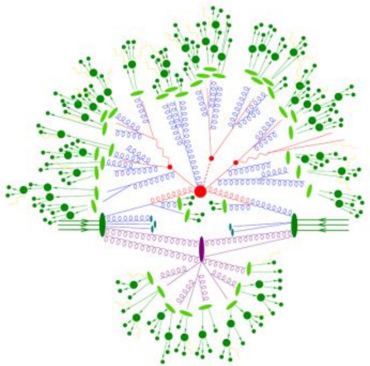
# A Roadmap for HEP Software and Computing R&D for the 2020s

- 70 page document
- **13 topical sections** summarising R&D in a variety of technical areas for HEP Software and Computing
  - Almost all major domains of HEP Software and Computing are covered
  - Including Generators and Security, absent in the initial draft, recognizing the importance of the CWP
- **1 section on Training and Careers**
- **296 authors** (signers) from 120 HEP-related institutions
- Signing policy: sign the document if you agree with the main observations and conclusions
  - Remains open up to the WLCG Naples workshop (March 26-29, 2018)

HSF-CWP-2017-01
December 15, 2017

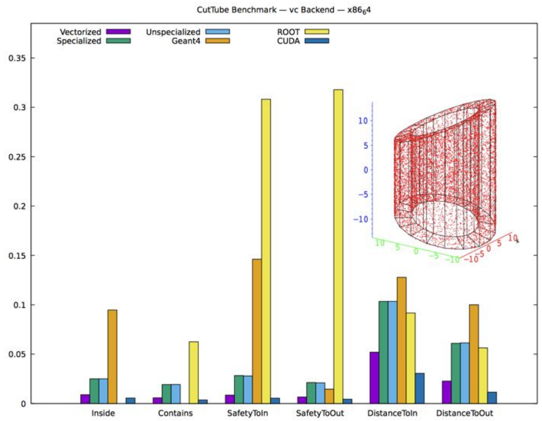
**Contents**

# Physics Event Generators



- Physics event generation starts our simulation chain to enable comparisons with detector events
  - Depending on the precision requested, CPU for event generation ranges from modest to huge
  - At Next-to-Leading Order (NLO) precision used today,  CPU consumption can become important
  - Study of rare processes at the HL-LHC will require the more demanding NNLO for more analyses
- **Generators are written by the theory community**
  - Need expert help and long term associations to achieve code optimisation
  - Even basic multi-thread safety is problematic for many older, but still heavily used, generators
  - Ongoing maintenance of tools like HepMC, LHAPDF, Rivet is required and needs rewarded
- Writing this section was the result of **intense contacts between HEP experts and the main people in the generator community**
  - Several actions identified to foster the collaboration

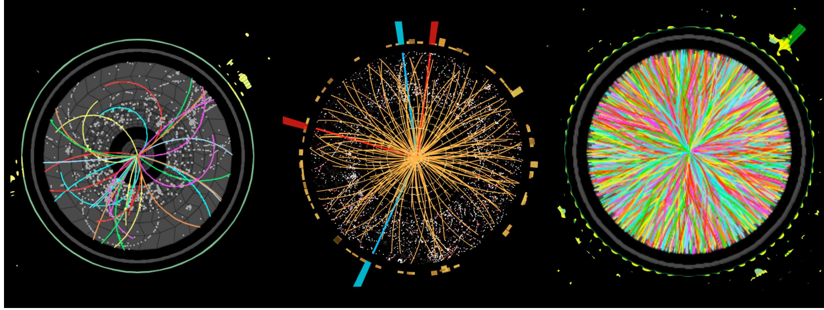
# Detector Simulation




Vectorised geometry speed-ups, Mihaela Gheata

- Simulating our detectors consumes huge resources today
  - Remains a vital area for **HL-LHC and intensity frontier experiments** in particular
- Main R&D topics
  - Improved physics models for higher precision at higher energies (HL-LHC and then FCC)
  - Adapting to new, in particular vectorised, computing architectures
  - Fast simulation - develop a common toolkit for tuning and validation
  - Geometry modelling: easier and efficient modelling of complex detectors
- **All main experts and frameworks contributed** to a 50 page detailed review of the detector simulation challenges and required R&D actions
  - **CWP brought a more consistent view and workplan** between the different projects, in particular Geant4 and GeantV

# Software Trigger and Event Reconstruction

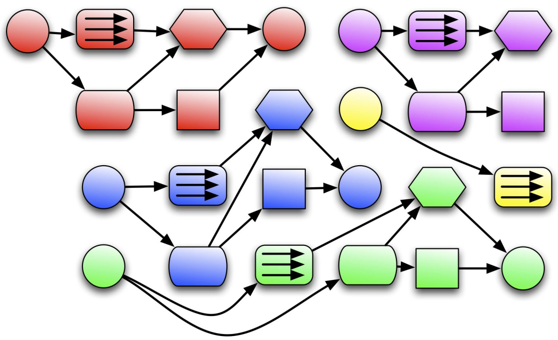


- Move to software triggers is already a key part of the program for LHCb and ALICE already in Run 3
  - ‘Real time analysis’ increases signal rates and can make computing much more efficient (storage and CPU)
- Main R&D topics
  - Controlling charged particle tracking resource consumption and maintaining performance at pile-up of 200 (or 1000)
  - Detector design itself has a big impact (e.g., timing detectors, track triggers)
  - Improved use of new computing architectures, extending the use of GPGPUs and possibly FPGAs
  - Robust validation techniques when information will be discarded
- Section rather HL-LHC specific
  - Trigger is highly dependent on the kind of experiments
  - One area where experience of Run 1 + Run 2 gives a better idea of challenges

# Data Analysis and Interpretation

- HEP analysis currently dominated by many cycles of data reduction
    - Goal: reduce the input to an analysis down to a manageable quantity that can be cycled over quickly on ~laptop scale resources
    - Key metric is ‘time to insight’
    - ROOT playing a central role
- Main R&D topics
    - Can we benefit, and how, from the latest techniques/tools in data analysis developed/used outside HEP: skimming/slimming cycles consume large resources and can be inefficient
    - For this, need ways to seamlessly interoperate between their data formats and ROOT
    - New analysis facilities: can interactive data analysis clusters be set up in our budget constraint
- Area with a strong potential for collaborating with the non-HEP world
    - In particular, links to be strengthened with the data science community

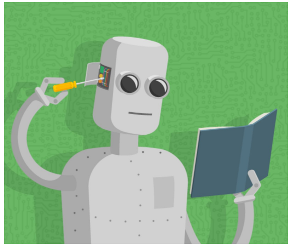
# Data Processing Frameworks



Event processing framework schematic (colours are events, boxes algorithms)

- Experiment software frameworks provide the scaffolding for algorithmic code
  - Currently there are many implementations of frameworks, with some (limited) sharing between experiments (e.g. ATLAS/LHCb Gaudi)
  - Ongoing efforts in all these frameworks to support concurrency
- Main R&D topics
  - Adaptation to new hardware, optimising efficiency and throughput
  - Incorporation of external (co)processing resources, such as GPGPUs
  - Interface with workload management to deal with the inhomogeneity of processing resources
- General agreement that it is an area for consolidation in the future
  - Reasons for so many frameworks are not really related to experiment specificities...
  - But also the hardest component to change in the experiment SW stack: need to identify the best approaches to promote commonalities, e.g., in underlying components

# Machine Learning



- Neural networks and Boosted Decision Trees have been used in HEP for a long time
  - e.g., particle identification algorithms
- The field has been significantly enhanced by new techniques (Deep Neural Networks), enhanced training methods, and community-supported (Python) packages
  - Very good at dealing with noisy data and huge parameter spaces
  - A lot of interest from our community in these new techniques, in multiple fields
- Main R&D topics
  - Speeding up computationally intensive pieces of our workflows (fast simulation, tracking)
  - Enhancing physics reach by classifying better than our current techniques
  - Improving data compression by learning and retaining only salient features
  - Anomaly detection for detector and computing operations
- Good links with the broader Machine Learning and Data Science communities required
  - Required match the efforts and expertises needed to make effective use of these techniques

# Other Technical Areas

## Conditions Data

- Growth of alignment and calibration data is usually linear in time
  - Per se, this does not represent a major problem for the HL-LHC
- Opportunities to use modern distributed techniques to solve this problem efficiently and scalably
  - Cacheable blobs accessed via REST
  - CVMFS + Files
  - Git

## Visualisation

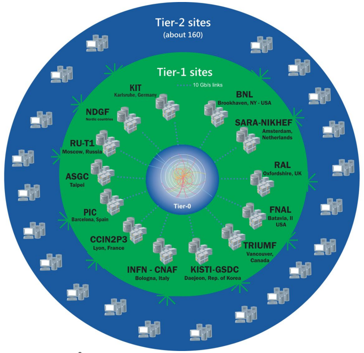
- Many software products developed for event visualisation
  - Part of the framework, with full access to event and geometry data
  - Standalone as a lightweight solution
- New technologies for rendering displays exist, e.g., WebGL from within a browser
- A new joint R&D project starting between CMS and ATLAS exploring advanced imaging techniques from biomed community

- These areas are examples of where we can refocus current effort towards common software solutions and some actions started because of the CWP
- This should improve quality, economise overall effort and help us to adapt to new circumstances

# Data Management and Organisation

- Data storage costs are a major driver for LHC physics today
  - HL-LHC will bring a step change in the quantity of data being acquired by ATLAS and CMS
- Main R&D topics
  - Adapt to new needs driven by changing algorithms and data processing needs, e.g., fast access to training datasets for Machine Learning, high granularity access to event data, rapid high throughput access for a future analysis facilities...
  - Consolidate storage access interfaces and protocols
  - Efficient hierarchical access to data, from high latency tape and medium latency network
- HEP should be able to benefit from advances in industry standards and implementations, such as Apache Spark-like clusters
  - Not a drop-in replacement for our solutions: what they do is not exactly like what we do (structured access to complex data)

# Facilities and Distributed Computing




- Storage and computing today are provided overwhelmingly from WLCG resources
  - Expected to continue for HL-LHC, but to be strongly influenced by developments in commodity infrastructure as a service, e.g. (commercial) Cloud Computing
- Main R&D topic: understand the effective costs involved in delivering computing
  - Activity started during autumn 2017: a WLCG WG formed and meeting fortnightly
  - This needs to be sensitive to regional variations in funding and direct and indirect costs like resources “beyond the pledge”, power and human resources
  - Full model unfeasible: only need a model good enough to answer questions about future investments, e.g. network or CPU vs. storage investments, impact of more concentrated storage resources (aka “data lake”)...
- Strengthen links to other big data sciences (SKA) and computing science
  - Common solutions where possible: cannot expect sites to deploy per-experiment technologies
  - Understand how to benefit from Software-Defined Networks or Storage

# Security Matters

- We have a large infrastructure that is an important resource for us
  - Protecting is is necessary for both our work and for our reputation
- HEP cannot live alone: cooperation with others is a requirement
  - Already exists: HEP as a structured community often a driver in common efforts
  - Need to evolve to integrate new requirements from new big science communities and from new legislation, e.g; EU Data protection
  - Facing new threats: requires better intelligence sharing for threat monitoring and response
- Main R&D topics
  - Trust and policy, in particular integration of commercial resources and hybrid clouds that we don’t control
  - Operational security: R&E Forum for incident response, broaden regional participation
  - Authentication and Authorisation, e.g. generic authentication services (e.g., eduGAIN) help users and are easier than X.509

# Data, Software and Analysis Preservation

- We seem to be doing well compared to other fields
  - Already involved in cross-community efforts, like in RDA forum
- Challenge is both to physically preserve bits and to preserve knowledge
  - DPHEP has looked into both
- Knowledge preservation is very challenging
  - Experiment production workflows vary in significant details
  - Variety of different steps are undertaken at the analysis stage, even within experiments
- Need a workflow that can capture this complexity
  - Technology developments that can help are, e.g., containers
- CERN Analysis Preservation Portal forms a good basis for further work
  - Needs to have a low barrier for entry for analysts
  - Can provide an immediate benefit in knowledge transmission within an experiment

# Software Development

- Experiments have modernised their SW development models a lot recently
  - Tools adopted from the open source community: Git and CMake, social coding (GitLab/GitHub) coupled with Continuous Integration
  - More tools and expertise sharing will benefit the community: need to (re-)establish the development forum
- **The more commonality in the tools and techniques, the more training we can share**
  - Using new tools requires investing in training for the community
  - Recently put in practice by ALICE and LHCb, using LHCb StarterKit material
  - This helps with preservation and propagation of knowledge
- **Our environment is becoming more complex: input required from physicists *whose concerns are not primarily in software***
  - Sustainability of these contributions is extremely important
  - We should become better at publication and citation of work to help this (and use new tools like Zenodo)

# Master Direction: Avoid HEP-specific Solutions

- HEP faced many computing challenges before other communities and has developed over the decades a lot of community-specific solutions…
  - Mainly for good reasons!
  - Several HEP-tools adopted by some other communities, e.g. GEANT4 and ROOT, and WLCG itself was a model/driver for large-scale computing adopted by some other disciplines (e.g., EGI)
- But the world changed: other scientific communities and industry facing some similar challenges and HEP must be able to benefit from them
  - Machine learning, distributed analysis, distributed infrastructure (e.g., clouds…)
- Does not mean that we have drop-in replacements for our solutions
  - Challenge: find the proper integration between our community tools and the available technologies outside, maintain the necessary backward compatibility/continuity
  - As illustrated in CWP chapters, not one single approach for every topic: several paths for moving in this direction are part of the roadmap

# Training and Careers

- To address the technical challenges, we need to raise the SW&Computing expertise in our community
  - Investment in SW critical to match HL-LHC requirements with a “flat-budget” scenario
  - Sharing between experiments is still an exception: training must become a first class activity
- Historically, many different profiles involved in HEP computing from physicists, PhDs to real SW&Computing experts
  - Required by the cutting-edge challenges we face that require all the expertises to collaborate
  - No way to “outsource” the challenging problems to a few experts…
  - Recognition of the contribution of our specialists in their careers is extremely important
- A critical role played by people with a strong physics background + a strong computing expertise
  - Difficult career paths for this profile: neither outstanding physicists nor outstanding SW experts
  - The community does not really have control over this: we depend on national/organisation policies

# The CWP: an Important Milestone for HEP SW&C

- The Community White Paper process is concluded and has been a success
    - Engaged more 250+ people and produced 300+ pages of detailed description in many areas
    - A real step forward compared to the situation before the CWP, thanks to the fruitful discussions: not a shopping list of all the possible ideas
- **But the CWP is a milestone, not a final step**
    - Links fostered between the people involved in the SW&C of the major HEP experiments
    - R&D program proposed in each area should serve as the basis for future work
    - Concrete paths identified to move towards more common solutions in HEP and to benefit from solutions developed outside the community
- Each experiment must build its own prioritized R&D program out of the CWP
    - Priorities are different between all experiments: not facing the same challenges at the same time or scale: not possible to have a prioritized R&D program at the HEP level
    - On each topic relevant to several experiments, **must work together**

# Advancing from Here

HEP Software Foundation

- Main areas we need to invest in for the future for our **Software Upgrade** identified: we need to turn them into concrete actions
  - Explicit goal of the WLCG/HSF workshop in Naples (March 26-29)
  - HL-LHC is a driver: the Run 1 and Run 2 experience helps to better identify the challenges
  - Must be inclusive of the whole HEP community: better links with Intensity Frontier and Belle II
- **HSF, with its bottom-up approach, has proved its worth in delivering this CWP**
  - Managed to build a community consensus: not an easy and usual process in HEP
  - The challenges are formidable, working together will be the most efficacious way to succeed
  - Now a recognized organization to spread knowledge of new initiatives, to encourage collaboration and to monitor progress
- **Organisations and funding agencies support is required** for marshalling and refocusing the R&D efforts, and helping to **attract new investment** in critical areas
  - Career path of the needed experts is of critical importance for the medium/long term

# Useful Links...

- CWP process and first results, CERN 3d Scientific Computing Forum, G. Stewart, October 2017
    - https://indico.cern.ch/event/663273/contributions/2708178/attachments/1545100/2431717/HSF-CWP-Roadmap.pdf
- HSF Community White Paper web site : links to topical papers, status of their publication to arXiv, updates on related activities, presentations about the CWP
    - http://hepsoftwarefoundation.org/activities/cwp.html