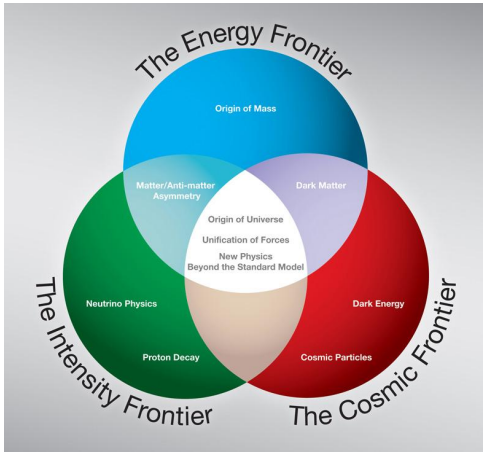
HSF Community White Paper: Lessons and Future Work

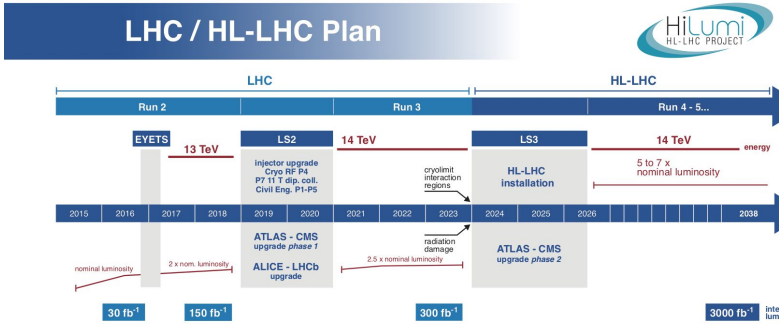
Michel Jouvin - CNRS/LAL

CHEP 2018, Sofia

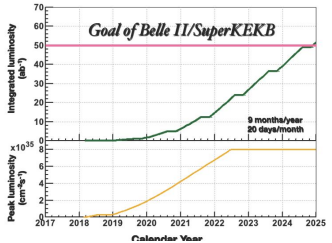
2018-07-11

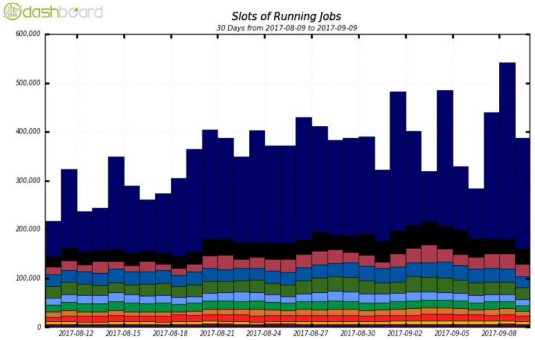
HEP: Landscape and Frontiers





SuperKEKB luminosity projection





>50M LOC

Fermilab Program Planning
20-Feb-17

LONG-RANGE PLAN: DRAFT Version 7a

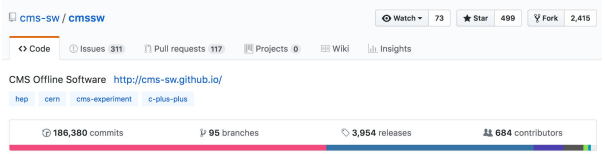
		FY16	FY17	FY18	FY19	FY20	FY21	FY22	FY23	FY24	FY25	FY26
LBNF/PIP II	LBNF/PIP II							LBNF	LBNF	LBNF / PIP II		
	SANFORD						DUNE	DUNE	DUNE	DUNE	DUNE	DUNE

Summer shutdown Construction / commissioning Run

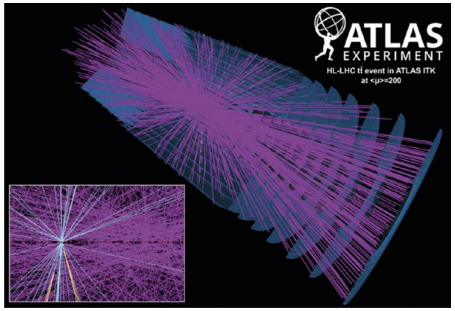
FNAL Intensity Frontier

athena

ATLAS Experiment main repository for Athena code



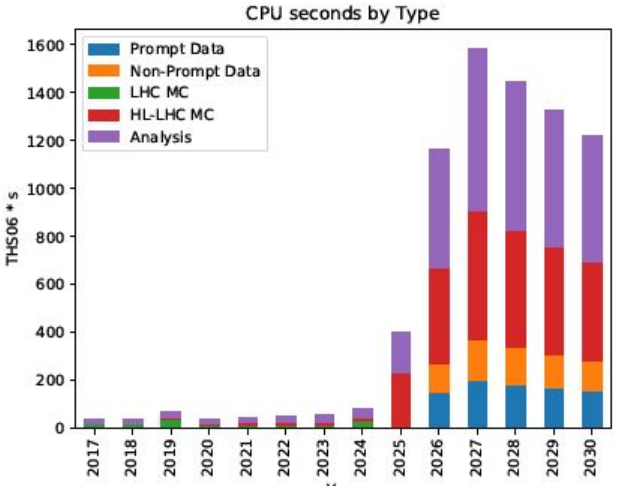
Software Challenges for HL-LHC



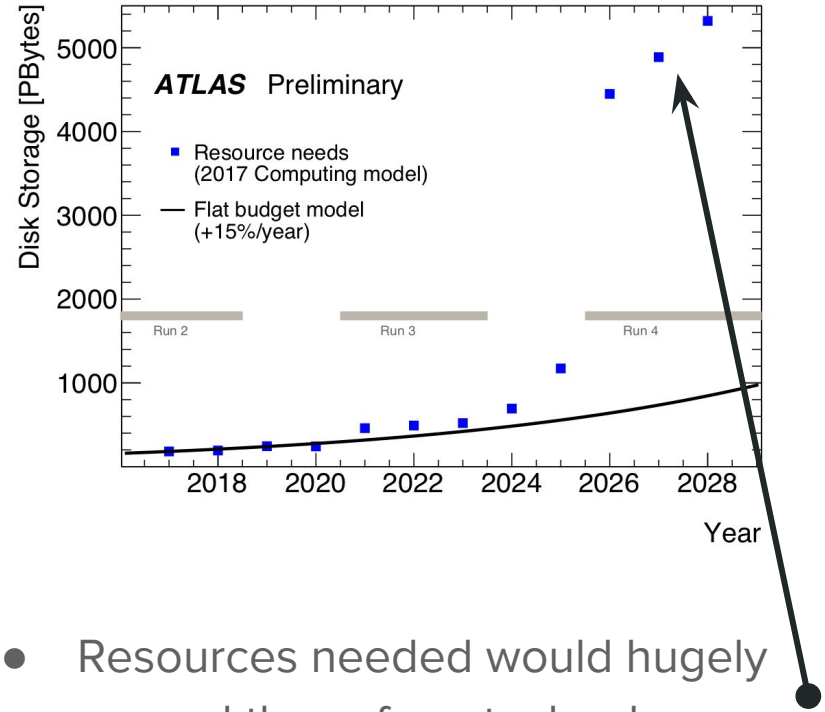
- Pile-up of ~200 ⇒ particularly a challenge for charged particle reconstruction (40x compute resources)
- With a flat budget, Moore's lawish improvements (x10) are the real maximum we can expect on the HW side
- HEP software typically executes one instruction at a time (per thread)
  - Since ~2013 CPU (core) performance increase is due to more internal parallelism
  - x10 probably achievable with the same HW if using the full potential of processors
    - major SW re-engineering required (but rewriting everything is not an option)
  - Accelerators 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

High Luminosity LHC

- Large rise in rate (~10kHz) and complexity (mu~200) : Run 2 SW & computing will not scale

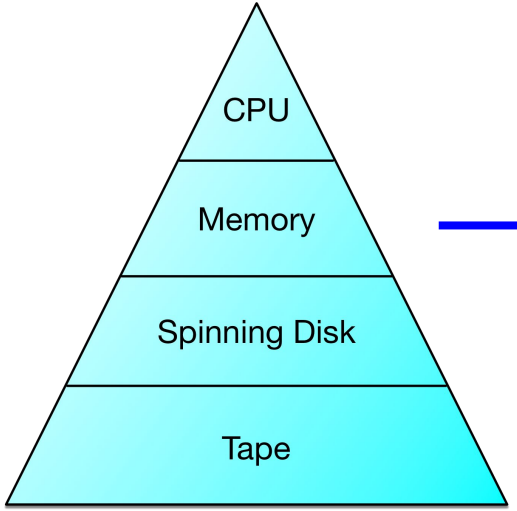


CMS CPU requirements

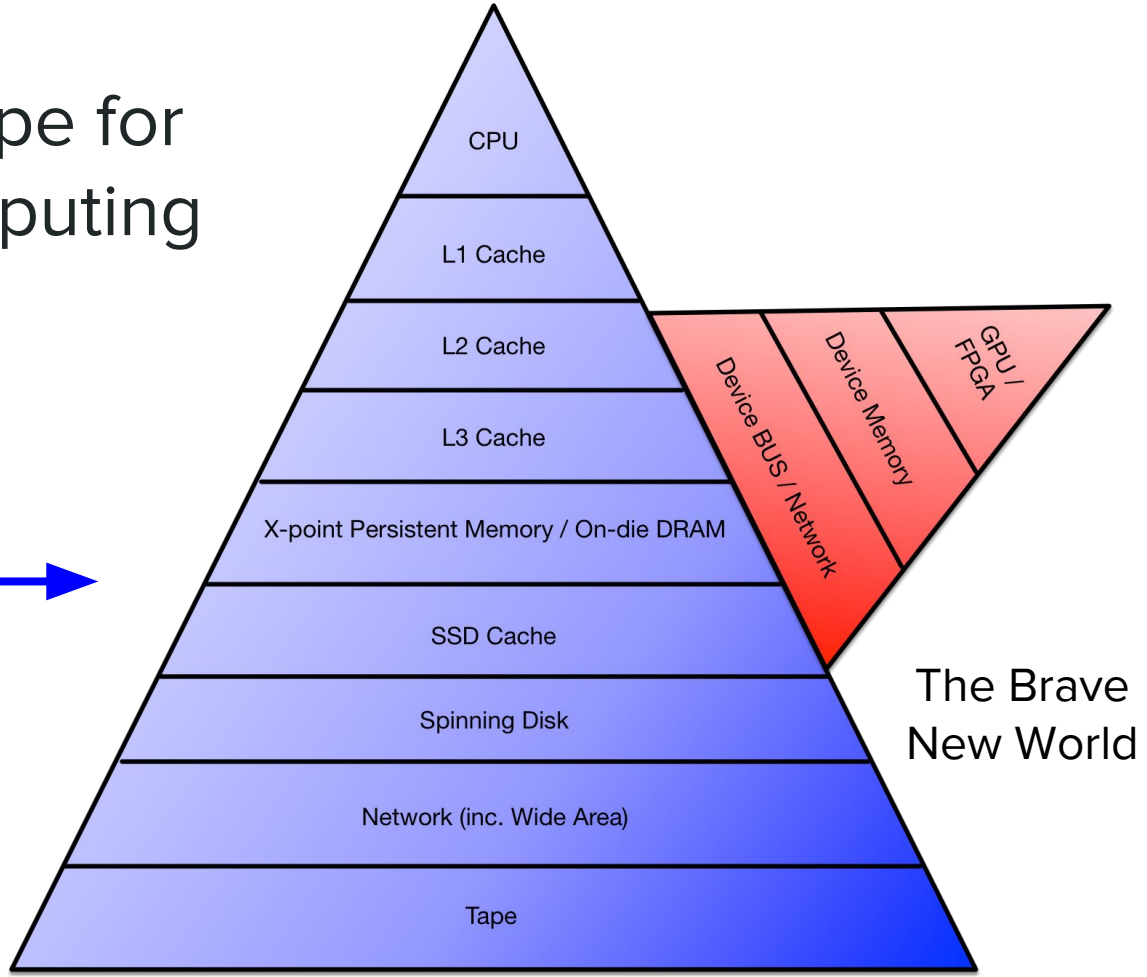


- Resources needed would hugely exceed those from technology evolution alone alone with a flat budget (close to Run 2+3 evolution)

Shifting landscape for end-to-end computing



The Good Old Days



The Brave New World

Courtesy Graeme Stewart, CERN

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 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*
- HSF already started with a number of workshops and working groups on common topics (packaging, licensing, analysis)

CWP: Making a Roadmap for the Future

- Community White Paper objective: describe a global vision for SW and computing in HEP in the 2020s (aka HL-LHC era)
  - Focus: achieve improvements in SW efficiency, enable new approaches for an extended physics reach, long term sustainability of our SW
  - Shared community view: bottom-up process built upon several general and topical workshops
    - Kick-off in San Diego, Jan. 2017; closing workshop in Annecy, June 2017
  - Official charge from WLCG in July 2016: CWP as an input to the HL-LHC “software upgrade”
  - Editorial work done during Fall 2017
- Significant community involvement in the CWP process and writing
  - ~100 participants to workshops, ~250 in the writing of the topical papers
  - 2 public drafts of the global roadmap: 100s of commenters
  - Final release on Dec. 20, 2017 (arXiv: 1712.06982)
    - Publication in progress in Computing and SW for Big Science journal

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

HSF-CWP-2017-01
December 15, 2017

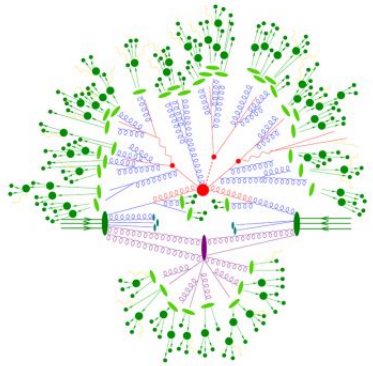
- 70 page document
- **13 topical sections** summarising R&D in a variety of technical areas for HEP Software and Computing
  - Almost all major domains of HEP Software and Computing are covered
  - For each section, a topical paper with more details also (being) published in arXiv (e.g. 50-page detailed review about Detector Simulation)
- **1 section on Training and Careers**
- **310 authors** (signers) from 124 HEP-related institutions
- More details on the HSF web site

Contents

Guiding Strategy for the Roadmap

- 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 is a model/driver for large-scale computing adopted by some other disciplines
- 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
- 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 and long-term sustainability
  - As illustrated in CWP chapters, not one single approach for every topic: several paths for moving in this direction are part of the roadmap

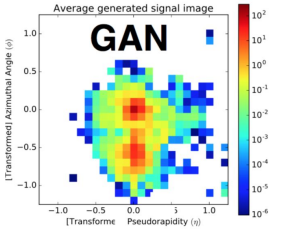
Physics Event Generators

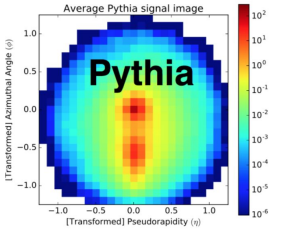


- Physics event generation starts our simulation chain to enable comparisons with detector events
  - Depending on the precision requested, CPU for event generation ranges from modest to huge
  - At Next-to-Leading Order (NLO) precision used today, CPU consumption can become important
  - Study of rare 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**

R&D Outlook: dedicated re-engineering workshop planned Fall 2018

Detector Simulation



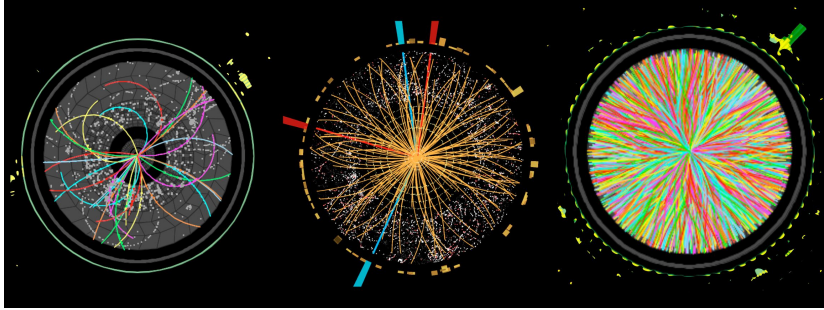


Machine learning simulated calorimiter

- **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 computing architectures
    - Can a vectorised transport engine actually work in a realistic prototype (GeantV early releases)? How painful would evolution be (re-integration into Geant4)?
  - Faster simulation - develop a common toolkit for tuning and validation of fast simulation
    - How can we best use Machine Learning profitably here? Multi-level approach, from processes to entire events
  - Geometry modelling
    - Easier modelling of complex detectors, targeting new computing architectures
  - **CWP brought a more consistent view and workplan** between the different projects

R&D Outlook: Community is well organised and actively pursuing many lines

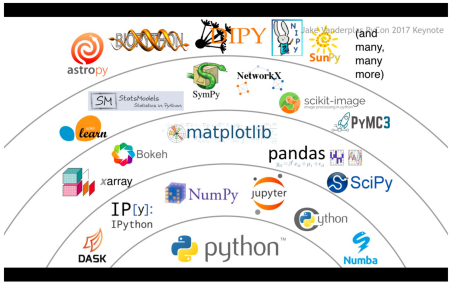
Software Trigger and Event Reconstruction



- **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 more efficient (storage and CPU)
- **Main R&D topics**
  - Controlling charged particle tracking resource consumption and maintaining performance
    - Do current algorithms’ physics output hold up at pile-up of 200 (or 1000)
    - Can tracking maintain low p_T sensitivity within budget?
  - Detector design itself has a big impact (e.g., timing detectors, track triggers)
  - Improved use of new computing architectures: multi-threaded and vectorised CPU code, GPGPUs, FPGAs
  - Robust validation techniques when information will be discarded
    - Using modern continuous integration, multiple architectures with reasonable turnaround times
  - Reconstruction toolkits can help adapt to experiment specificities: ACTS, TrickTrack, Matriplex

R&D Outlook: A lot of projects in healthy states - keep up level of cooperation and sharing (Connecting the Dots; TrackML Challenge)

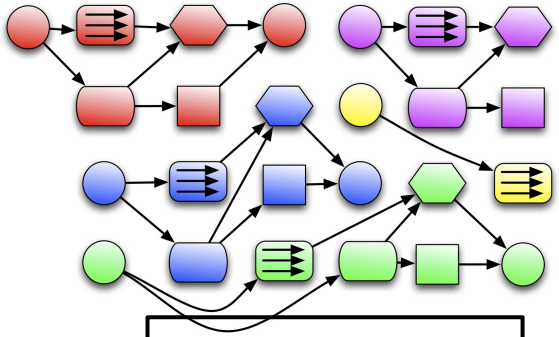
Data Analysis and Interpretation



- **Today we are dominated by many cycles of data reduction**
  - Aim is to 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’
- **Main R&D topics**
  - How to use the latest techniques in data analysis that come from outside HEP?
    - Particularly from the Machine Learning and Data Science domains
    - Need ways to seamlessly interoperate between their data formats and ROOT
      - Python is the *lingua franca* here, thus guaranteeing our python/C++ bindings is critical
  - New Analysis Facilities
    - Skimming/slimming cycles consume large resources and can be inefficient
    - Can interactive data analysis clusters be set up? SWAN, Spark, Dask interesting
      - Characterised by rapid column-wise access reads, with writes of new columns

R&D Outlook: many potential directions, no clear overall structure yet, needs good exchange of information and collaboration with the non-HEP world

Data Processing Frameworks

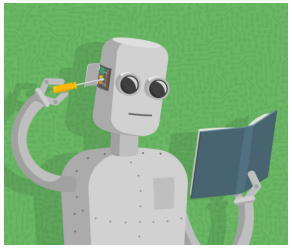


Event processing framework schematic (colours are events, boxes algorithms)

- 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
  - Reasons for so many frameworks are not really related to experiment specificities…
- 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
  - Evolution strategy: promote commonalities with common libraries and services
    - Discussions still going on about possible framework consolidation

R&D Outlook: general agreement that it is an area for consolidation in the future, even if no clear path has been identified yet

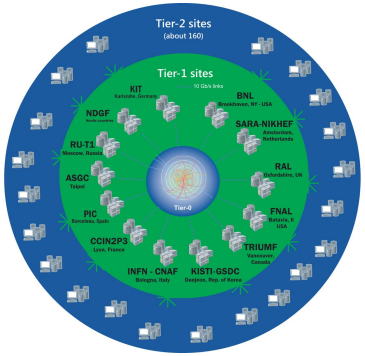
Machine Learning



- 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

R&D Outlook: good links with the broader Machine Learning and Data Science communities required

Distributed Computing and DOMA



- 2 different sections covering several topics from facilities to data organisation, management and access (DOMA)
  - From technology to more organisational issues
  - Data storage costs are a major driver for LHC today: next decade experiments (HL-LHC, DUNE) will bring a step change in data being acquired
  - WLCG operational model as a too high cost for too specific solutions
  - Sites have to support multiple experiments and cannot afford specific technologies
- Main R&D topics
  - Data-lake approach: (federated) large data centers accessible from any kind of computing resources. At the heart of the WLCG DOMA project launched in June.
  - WLCG performance and cost model: a WG formed and active since 6 months to build a (not too complex) model to assess the global impact of computing model changes.

R&D Outlook: strengthen links to other big data sciences (e.g. SKA) and computer science

Other Technical Areas

- Conditions Data and Visualisation: many different products/projects
  - 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, Software and Analysis Preservation
  - Challenge is both to preserve physically bits and to preserve knowledge: DPHEP has looked into both
  - Preserving knowledge is the most challenging: CERN Analysis Preservation Portal forms a good basis for further work
- Security: new requirements, new threats, new technologies
  - Must protect our work and our reputation
  - HEP is a structured community and often acts as a driver in common efforts with others

Training and Careers

- To address the technical challenges, we need to raise the SW&Computing expertise in our community
  - Investment in SW is 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

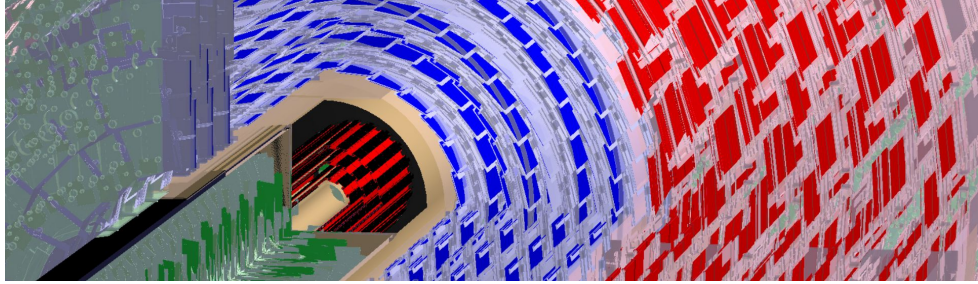
The CWP: an Important Milestone for HEP SW&C

- The Community White Paper process is concluded and has been a success
  - 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**

New HSF Working Groups

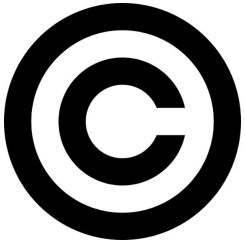
- HSF is forming working groups in this three key areas of HEP software:
  - Simulation
  - Reconstruction
  - Analysis
- Building on R&D topics proposed in the CWP
- Raise awareness of work being done in these areas
  - Not all projects are as known as they should be
  - New projects can begin with a broad scope and common goals
- These will be areas reviewed by the LHCC next year
  - These groups will able to answer the charge of whether we really have learned to work together or not
  - These WGs will not be HL-LHC specific, but the review can help us drive forwards

Software Forum



- HSF has been established to foster sharing of expertise and increase collaboration around common projects
  - Needed a place to do it on a regular basis, independently of specific projects
- HSF has (re)launched the Software Forum
  - Showcase common software projects
  - Introduce tools that help us face challenges like concurrency or vectorisation
  - Open dialogue with other like-minded communities
- Bi-weekly meeting on Wednesday 5pm CET, odd weeks
  - Not restricted to HL-LHC topics: embrace the whole community requirements, experience and tools (FAIR, Intensity Frontier, ILC, FCC, CLIC...)
  - Managed to have only 2 meetings before the summer
    - DD4hep geometry modeling package adopted by CLIC, FCC, CMS with LHCb interested
    - VecCore and SOAContainer: in fact next week

Copyright and Licensing

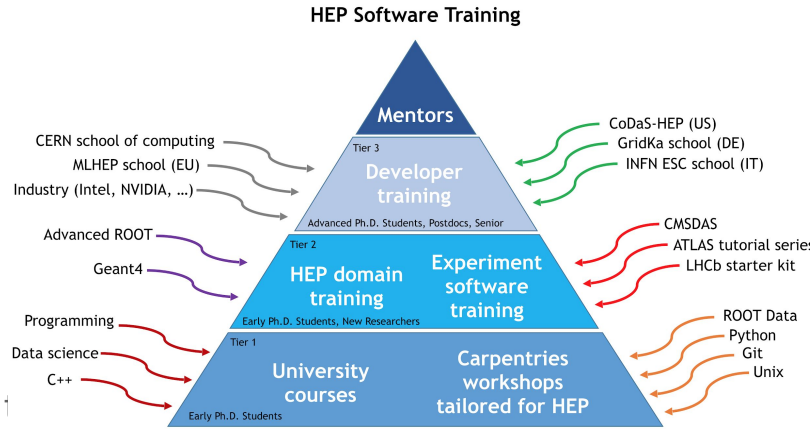


- HSF has took in charge this much neglected area in HEP software
  - Much code exists with no clear copyright or licence
  - The issues of large and deep stacks of experiments' software and license combinations were often neglected up to now
    - *Does impact on our ability to collaborate*
- Experiments started to worry about licensing issues (LHC, Belle II)
  - Goal is to maximise our useful user base and foster collaboration with others, including outside HEP
- GPL licenses have become disfavoured as they place obligations on any users that can inhibit collaboration (e.g., with commercial companies)
  - ATLAS and CMS want non-GPL licenses
    - Matches shift at CERN, e.g., Indico moving from GPL to MIT
  - We made significant progress in moving packages like HepMC and DD4hep to *LGPL*
  - Widespread use of GPL by theory community still affects us greatly (Fastjet, Pythia8 among others)
  - HSF continuing the discussion with authors of "problematic packages"

Other HSF Initiatives...

- Citation: 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)
- Software Development: the more commonality in the tools and techniques, the more training and expertise we can share
  - Build upon LHCb StarterKit: ALICE and LHCb common training recently
  - This helps with preservation and propagation of knowledge
- Software Tools WG: performance analysis (profiling) tools and data
  - Many available tools: try to converge on a few ones to help sharing expertise and data
  - Common work on warehousing and visualisation possible

… Other HSF Initiatives



- Packaging is one of the de facto areas of common interest between experiments
  - Building and deploying our software is a significant
  - Packaging WG decided to formalize the problem around actual use cases we have
  - Several R&D projects looking at possible directions for the future (Nix, Portage, Spack)
- Training: address the training “pyramid” from core skills to experts
  - Organising a federation of training schools
  - Concrete work started on a curated set of training materials
- Topical workshops and events, like those organised around CHEP this year
  - PyHEP, frameworks
- Umbrella organisation for the Google Summer of Code (GSoC)
  - 2018: 29 projects funded (51 proposed), 25 organisations, 64 mentors

Advancing from Here

- Main areas for our **Software Upgrade** identified and concrete actions started
  - HL-LHC is a driver: LHC 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
  - CERN led the way with an HSF position in SFT group and CERN/EP R&D plans include SW projects
  - US project of an Institute for Research and Innovation in SW (IRIS-HEP) may play an important role
  - More similar initiatives are needed

Conclusions

- CWP brought us a long way forwards in understanding the problems ahead of us
  - And the areas where we can work together profitably
  - Each project/experiment must build its own strategy out of it, e.g. WLCG Strategy towards HL-LHC
- HSF continues to act as a focal point for common software efforts
  - Organisation of common work around the main technical areas (simulation, reconstruction, analysis)
  - Continued work on important technical matters: licensing, packaging, software tools
  - Inventory of software projects and tools; advice on publication and dissemination; training
  - Communication channels (hsf-forum, hsf-tech-forum lists) are vital
- New working groups will form nuclei of solving the grand challenges
- The program of work built from the CWP must be refined as concrete actions
  - New projects should be agile and cooperative from the outset
  - Support from organisations and funding agencies will be needed

There are many opportunities to be involved and shape our common work in the field

Useful Links...

- HSF Community White Paper web site : links to topical papers, status of their publication to arXiv, updates on related activities, presentations about the CWP
  - https://hepsoftwarefoundation.org/organization/cwp.html
- CWP and related work presentations with additional details
  - CHEP2106: CWP Status and Plan
  - 4th CERN Scientific Computing Forum (Jan. 2018): CWP Lessons and Future Work
  - WLCG Workshop, Naples, March 2018: CWP Roadmap
  - ATLAS SW&C Week, DESY, June 2018: R&D and Activities after Naples
- WLCG Strategy towards HL-LHC