Update on Software

Report on HSF Workshop

Benedikt Hegner
(EP-SFT, CERN)

LHCC Referee Meeting
24.5.2016

Using input from HSF and GeantV

3rd HSF Workshop

- **This workshop was held in LAL Orsay, May 2-4 2016**
  - Kick-off Workshop was held at CERN, April 2014
  - 1st follow up workshop was held at SLAC, January 2015

- **Goals of Workshop (see agenda)**
  - Topical sessions to review progress made in supporting collaboration on common software projects
  - Hands-on sessions to work concretely on specific problems
  - Session to gather input from other communities

- **Good Participation : ~70 people in room, more on Vidyo**
  - New involvement from all parts of the community, including intensity frontier and Belle II, covering ~30 institutes
  - Good balance Europe vs. North America

- **Co-located event: Geant4 Technical Forum as example of how to organise interaction between large projects and their user community**

Achievements from work done in 2015

- **Project Support**
  - This is the incubator for new projects wanting to join the HSF
    i. Interoperability of software is an essential requirement!
  - Four concrete results in the last year
    i. **Software Best Practices Document** (TN draft)
    ii. **Project Starter Kit** (https://github.com/HEP-SF/tools)
    iii. **Knowledge Base** gives visibility to projects and efforts (link)
    iv. **Licensing guidelines** (TN)

- **Infrastructure - Packaging**
  - HSF finished a review of packaging tool landscape (HSF Technical Note)
    i. What tools are around - inside and outside the community?
    ii. What are the requirements of the various stakeholders?
  - Identified the HPC tool spack as a very promising candidate for widespread adoption
    i. Dedicated hands-on session to demonstrate capabilities of this solution
    ii. Allows to provide entire stacks of HSF software packages with guaranteed interoperability
    iii. HEP people actively contribute to spack developer community already
  - Now active follow up by new experiments and individuals

List of projects reporting at workshop

- **DIANA-HEP (funded by NSF - USA)**
  - Elaborate and enhance existing analysis tools with ROOT as its core
- **AIDA2020-WP3 (funded by EU)**
  - Development of generic software tools for use in future experimental programmes (LC, FCC, LHC, Dune,...)
- **Future Conditions Database**
  - Started a common project between ATLAS and CMS for next-generation conditions data handling
  - Belle-II and LHCb are following this as well
- **HEP Software and Computing Knowledge Base**
  - All information about known software packages used at the LHC and elsewhere
  - **Now in production**
- **WikiToLearn**
  - New platform for organising training material: tutorials, “Collaborative Textbooks”,
  - Not HEP-specific but committed to support HSF
  - Growing community (160 contributors)
- **All these projects working in the spirit of the HSF on shared solutions**

Topical Session: SW Performance

- **Contributions by ALICE, ATLAS, CMS, GeantV, ROOT, Art/LArSoft and the astroparticle community**

- **Followed by a discussion session:**
  - What hardware to focus on? Commodity vs. GPU vs. HPC
  - Still no common understanding and idea on what to expect
  - Can we decouple low-level optimization (experts) from high-level code (physicists) via libraries?
    - Interesting input from Astrophysics how far one can get

- Lead to ***more questions than answers***
  - All the details in the meeting notes and the upcoming workshop summary

- **Outcome**
  - Needs a more visible and continuous activity in the HSF
    ⇒ Planning a focused follow-up workshop
  - Important input to the Community Whitepaper (see next)

Community White Paper

- **Discussed the need for a longer-term strategy for HEP software**
  - Idea is to produce a community white paper to define goals, priorities and possible collaborations on tackling the problems (similar to P5 for HEP experiments in the US)
  - Proposing a series of HSF-branded workshops working towards the whitepaper, with a conclusion mid-2017
  - Essential for providing a coherent picture and matching deliverables towards funding agencies when applying for funding

- **Discussion outcome**
  - Should go in line with discussions on HL-LHC computing TDR
  - Overall consensus that this is high priority
  - Kick-off workshop in fall

Communication and Recognition

- **Depsy.org**
  - Platform / infrastructure to promote credit for software in science
    - http://www.nature.com/news/the-unsung-heroes-of-scientific-software-1.19100
  - Automatic analysis of papers and code repositories for usage of SW
  - Gives a much better picture on software re-usage than information given in papers
  - Not yet applied/tried for C++ code bases

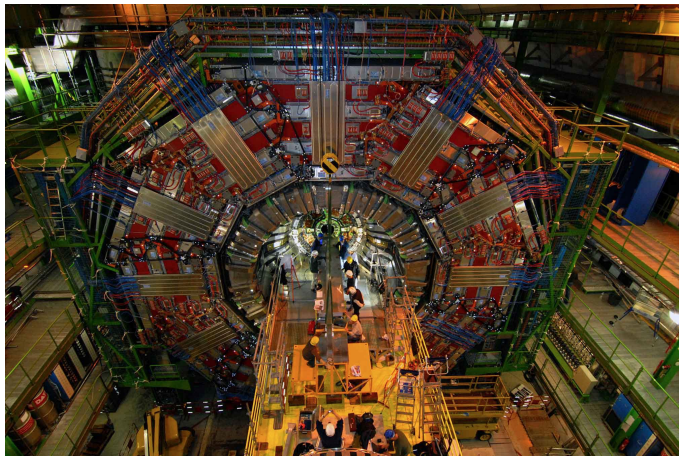
- **Experience of the Bioconductor project**
  - Project portal for bioinformatics software
    - Actively supported by a core team of people
  - Openness of software important
    - Development on GitHub, open peer-review
    - Research papers get corresponding paper about SW
    - After initial doubts/fears the openness helps people in careers

- **Software and Computing Journal by Springer**
  - Proposed by German community and e.g. presented to HEPiX mid-April
  - Authoritative and central reference archive for articles, reviews, advanced tutorials
  - Several open questions (e.g. on business model and level of OA)
  - **General agreement that this is a good idea!**

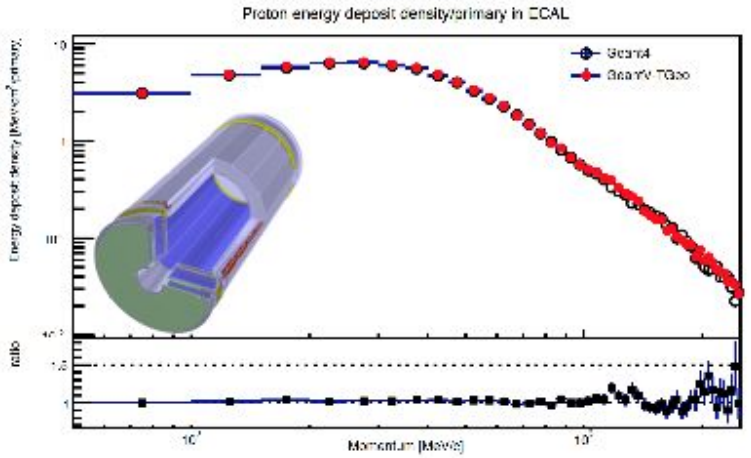
- **Idea to create a StackExchange forum for HEP SW related questions**
  - Well established approach outside HEP; allows promoting good questions and answers

GeantV Review

- **The GeantV R&D project is now ~3 years old**
- **As of now, the prototype is able to run at the scale of the entire CMS detector**
  - Simplified (tabulated) physics, but full geometry
  - Still validating, but even in single-threaded mode, significant speed improvements
    - > *3 speedup with next VecGeom version
    - Some potential for backporting to Geant4



- **The GeantV team approached the HSF for a peer-review of the project to**
  - Make sure the project is ‘on track’
  - Assess whether the goal of a ‘proof of principle’ of the design chosen was achieved
  - Look for possible show-stoppers until completion
  - Give suggestions for improvements and next steps
  - Make the project known to a wider community and potential new contributors
  - Convince clients and funding agencies



- **Review around 25. - 27. October at CERN**
  - Collecting proposals for reviewers
    - Clients of Geant4
    - Performance Experts

HSF Resources: Need for Effort

- **HSF needs dedicated resources**
  - Effort so far is the spare time of a tiny number of overcommitted people
  - This is clearly not sufficient if the HSF is to have a serious role

- **Much work ahead** in (especially) **common software infrastructure, performance, and new approaches** to effectively using new architectures
  - *"Improving software efficiency is essential to meet growth in requests. CRSG strongly supports funding to continue this work."* (LHC computing resources scrutiny group report)

- **Can the HSF be made a legal entity?**
  - Still far from a consensus but agreed to explore the possibility with funding agencies and lawyers
  - Main initial goal is IPR management similar to the approach of the Apache SW foundation

- **If common software efforts are an important activity they should be funded**

Workshop Conclusions

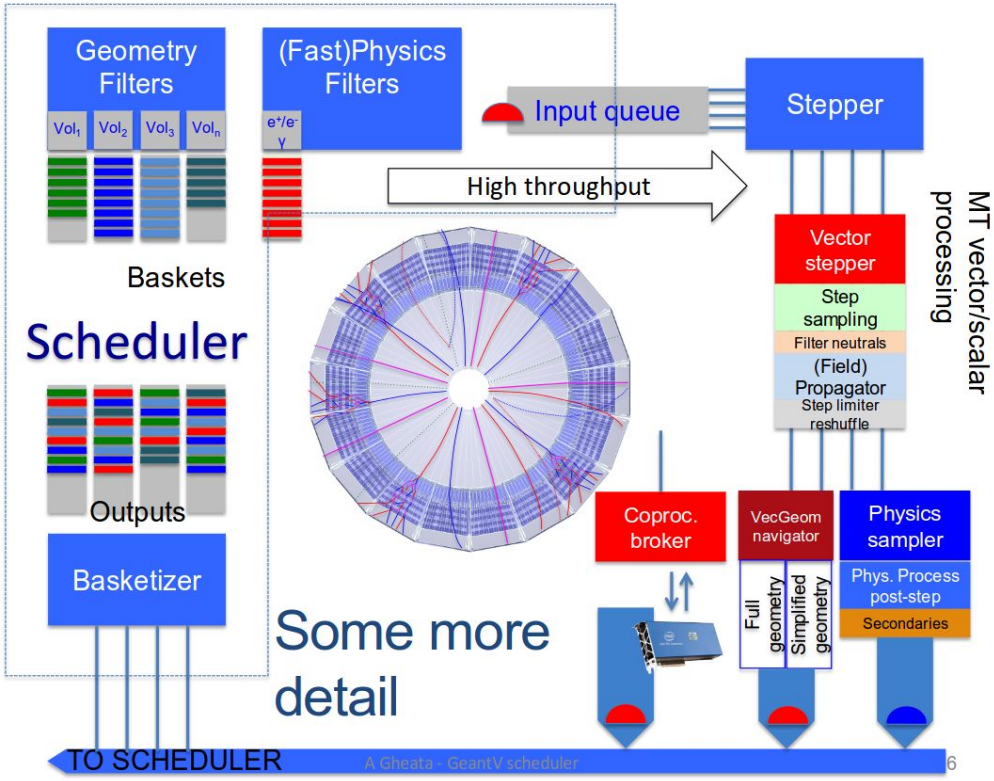
- **HSF seems well alive**
  - Significant progress and increasing motivation compared to one year ago
- **Need for resources is much more visible than last year**
- **Community White Paper a useful incentive to progress towards more common views and projects**
  - Series of topical workshops to organize
- **Develop/increase project support of HSF**
  - GeantV review a good example for what the HSF can offer as additional value
- **Look for an “official blessing” of the HSF by bodies like ECFA/ICFA**
- **Should the HSF be a legal entity?**

Backup

GeantV - Basic Motivation

- **Fetching data and instructions from RAM has a cost**
  - Often much higher than expected
  - Caches misses are not something to take lightly
- **What NOT to abuse**
  - Sparse access over large data structures
  - Frequent incoherent low granularity allocations
  - Piping small data through highly branching processing logic
  - Virtual interfaces on top of fine grain data units
- **Keeping up with technology**
  - Instruction-level parallelism
  - SIMD
  - Prefetching
  - Out of order execution

- **The GeantV project tries to address all these points with a complete re-design of simulation software by grouping multiple items and operations into *baskets***



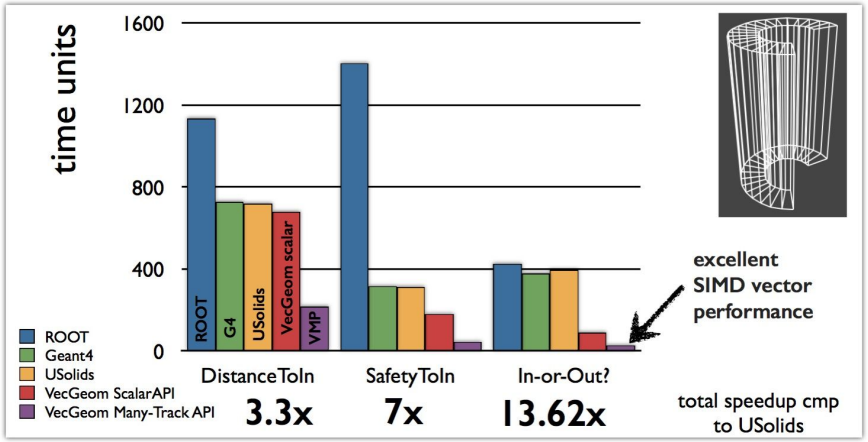
Preparing Simulation for the Future

- **Detector simulation with Geant4 is one of the essential ingredients**
  - It is the largest (50%) single consumer of the WLCG cycles.

- **Geant4 has been (naively) parallelized with Geant4 Version 10**

- **The potential of the current Geant4 design for long-term adjustment is limited**
  - Software design more than 20 years old (but serves the community well!)
  - Triggered new project - *GeantV*(ectorized)
    - Currently in R&D phase
    - 2016 review milestone is a CMS-like detector

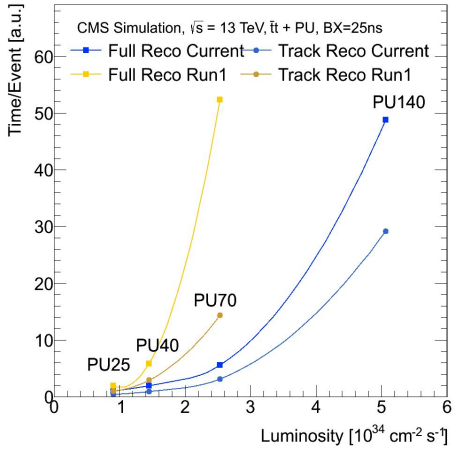
- **Necessary revolution of software can even be backported**
  - Geometry navigation takes 30-40% of CPU time in simulation
  - Addressed with new “VecGeom” library in GeantV
  - Substantial performance gains also in scalar mode & Geant4



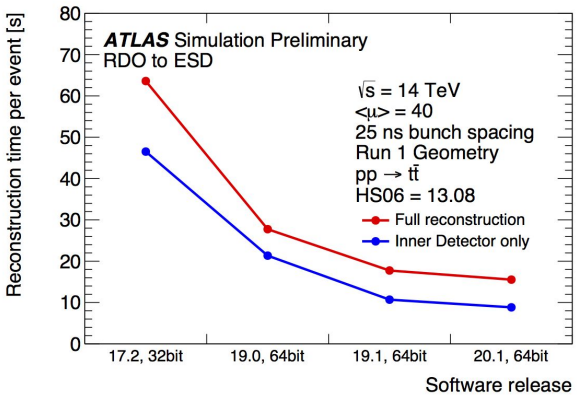
(Physics) Code Optimization

- **During LS1 all experiments invested heavily in improving their software**
  - Required detailed expert knowledge rare
    - Important to invest on the right topic
  - Significant gains achieved

- **One example - ATLAS and CMS Tracking**
  - Better/vectorized implementations of calculations
  - Addressing hot-spots, e.g. magnetic field in ATLAS
  - Tuned reconstruction strategy for higher pileup, e.g. new seeding algorithms in CMS
  - Up to x4 improvements in speed



- **Gained expertise fed back into community**
  - Ongoing efforts to make improved tracking code available to the wider community



- **Tracking performance will stay an important topic for HL-LHC**

HSF Topical Fora

- **Software Technology Forum**
  - Technical issues to embrace new technology in our software
  - Ongoing activity
- **Reconstruction Algorithms Forum**
  - All matters of event reconstruction and pattern recognition software; 3 in-person meetings
- **Machine Learning Forum**
  - ML discussions and code development in the context of HEP
  - Development of relevant tools, methodology and applications

HSF Topical Fora

- **Software Technology Forum**
  - Technical issues to embrace new technology in our software
  - Ongoing activity
- **Reconstruction Algorithms Forum**
  - All matters of event reconstruction and pattern recognition software; 3 in-person meetings
- **Machine Learning Forum**
  - ML discussions and code development in the context of HEP
  - Development of relevant tools, methodology and applications

Cross-experiment Collaborations

- **There are quite a few (more or less) new cross-experiment collaborations, with involvement or moderation of the HSF - going beyond WLCG**

- **Experiment frameworks**
  - Gaudi (ATLAS, LHCb, FCC)
  - FAIRRoot (FAIR, ALICE)
- **Common Conditions Data Project**
  - Discussion/cooperation between ATLAS, Belle II, CMS and LHCb
- **Common Software Build and Packaging Tool efforts**
  - Working group of HSF comparing HEP and non-HEP solutions
  - Starting point was LCG's Librarians and Integrators Meeting
- **Cooperation on Reconstruction Software**
  - "Connecting the Dots" tracking workshop extended by HSF session about common tracking implementations

- **AIDA2020** (EU funded)
    - DD4hep for detector description (LCD, FCC, potentially LHCb)
    - PODIO data model library (FCC, LCD, potentially LHCb)
- **DIANA (Data Intensive ANAlysis)** (NSF funded)
  - 4-year project on on analysis software, including ROOT and its ecosystem

Current Status and Activities II

- **Software performance**
  - Simulation: parallelisation of Geant4; GeantV
  - Reconstruction: HSF common tracking SW forum + Machine Learning Forum
  - I/O: parallel ROOT I/O, key-value-store evaluations
  - Mathematics: MetaLibm
  - Ad-hoc improvements and parallelization in various SW projects
  - Performance tools (e.g. igprof, FOM tools)

- **Supporting developers and participating projects**
  - Providing best practices to facilitate integration into HEP eco-system
  - Project templates for bootstrapping new projects
  - Development services
  - Help in selecting the proper SW license

- **Quite some activity in HSF, even though participation in the startup-team is on volunteer/best-effort level**

Support for multiple Architectures

- **For many years we have been developing for/running on x86 (Intel) architectures**
  - Other platforms mainly for code-checking

- **The landscape is much more complex now**
  - ARM coming from the low-power consumption market
  - Power8 a common platform in super-computing/HPC environment
  - Dedicated vector processors (e.g. GPUs)

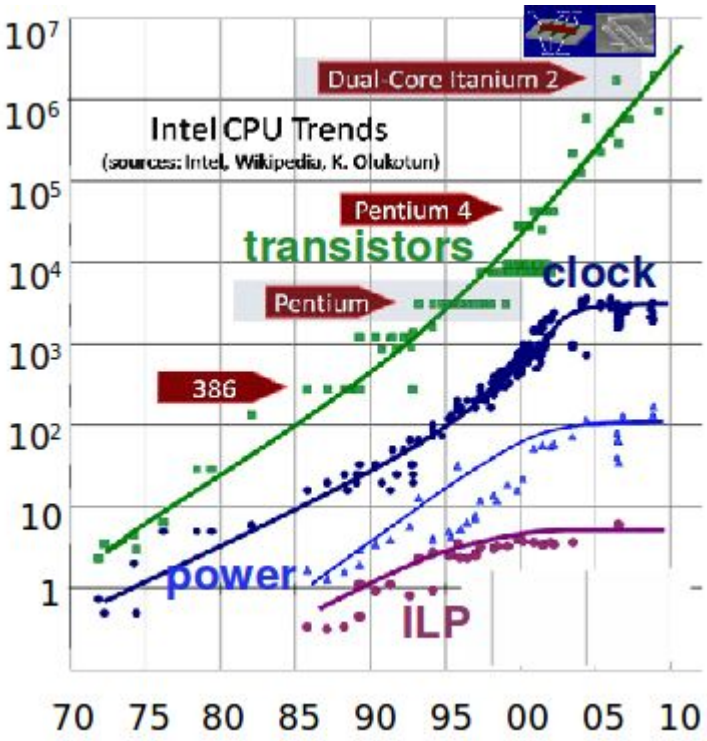
- **Ongoing activities**
  - Projects like Geant4 / ROOT already being tested on a variety of platforms
  - CERN IT providing multiple platforms on best-effort basis (TechLab)
  - Porting efforts within the experiments
  - Concurrency forum

HSF Timeline

- Jan 2014: HEP software collaboration proposed
- Apr 2014: HEP software collaboration meeting
- Spring/Summer 2014: gathering White Papers from the community.
- Oct 1 2014: Startup plan approved and startup team established. Agreement communities and software domains to focus on initially.
- Nov 11 2014: White Paper Analysis and Proposed Startup Plan released, followed by discussions with many parts of the community prior to the SLAC workshop.
- Jan 20-22 2015: SLAC HSF workshop established concrete activities and next steps
- Apr 17, 2015: HSF meeting at CHEP 2015, Okinawa to present progress, assess opportunities emerging from CHEP, and discuss next steps.
- June/July 2015: Intensive discussions in Packaging Working Group
- Sep 2015: Technical Notes policies published - more in the queue
- Sep 2015: HSF on WikiToLearn
- Nov 2015: HEP Knowledge Base finished
- May 2016: HSF Workshop in Paris

Stagnation in Speedup

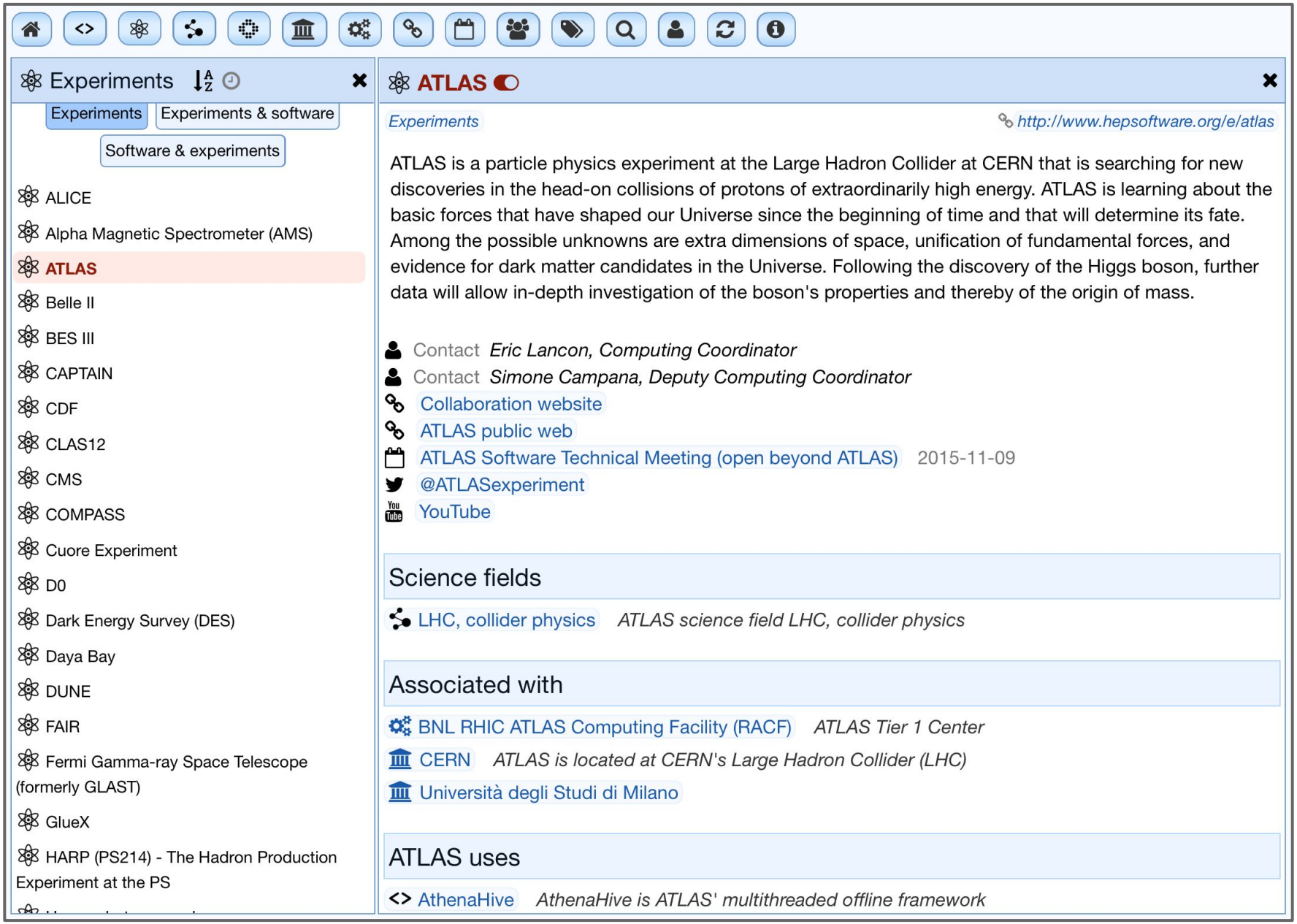
- Performance of our code scales with clock cycle (hence is stagnant!)
- Needs will increase more than tenfold and the budget will be constant at best
- HEP code needs to exploit new architectures and to team with other disciplines to share the optimization effort
  - Data & instruction locality and vectorisation
- Portability, better physics and optimization will be the targets
- Simulation can lead the way to show how to exploit today's CPU's resources more effectively in complex applications



Knowledge Base

- **Software catalog, software categories, science fields, community, and events**
  - implementation is a browser-based app (javascript client, node.js server, json in between, MySQL)
  - authentication is via github, google, facebook etc.
  - **emphasizes easy adding/editing of content, and extensibility. Adding content should be fun.**
- Available at http://hepsoftware.org
  - Comments/feedback are welcome!
  - Just start filling it!

- *Implementation based on ATLAS' DKB (data knowledge base)*


Experiments
Experiments
Experiments & software
Software & experiments
ALICE
Alpha Magnetic Spectrometer (AMS)
ATLAS
Belle II
BES III
CAPTAIN
CDF
CLAS12
CMS
COMPASS
Cuore Experiment
D0
Dark Energy Survey (DES)
Daya Bay
DUNE
FAIR
Fermi Gamma-ray Space Telescope (formerly GLAST)
GlueX
HARP (PS214) - The Hadron Production Experiment at the PS
ATLAS
Experiments
http://www.hepsoftware.org/e/atlas
ATLAS is a particle physics experiment at the Large Hadron Collider at CERN that is searching for new discoveries in the head-on collisions of protons of extraordinarily high energy. ATLAS is learning about the basic forces that have shaped our Universe since the beginning of time and that will determine its fate. Among the possible unknowns are extra dimensions of space, unification of fundamental forces, and evidence for dark matter candidates in the Universe. Following the discovery of the Higgs boson, further data will allow in-depth investigation of the boson's properties and thereby of the origin of mass.
Contact Eric Lancon, Computing Coordinator
Contact Simone Campana, Deputy Computing Coordinator
Collaboration website
ATLAS public web
ATLAS Software Technical Meeting (open beyond ATLAS) 2015-11-09
@ATLASexperiment
YouTube
Science fields
LHC, collider physics ATLAS science field LHC, collider physics
Associated with
BNL RHIC ATLAS Computing Facility (RACF) ATLAS Tier 1 Center
CERN ATLAS is located at CERN's Large Hadron Collider (LHC)
Università degli Studi di Milano
ATLAS uses
AthenaHive AthenaHive is ATLAS' multithreaded offline framework

HSF Activities and Working Groups

Working Group	Objectives	Forum - Mailing list
Communication and information exchange	Address communication issues and building the knowledge base Technical notes	hep-sf-tech-forum
Training	Organization of training and education, learning from similar initiatives	hep-sf-training-wg
Software Packaging	Package building and deployment, runtime and virtual environments	hep-sf-packaging-wg
Software Licensing	Recommendation for HSF licence(s)	hep-sf-tech-forum
Software Projects	Define incubator and other project membership or association levels. Easy-start project templates	hep-sf-tech-forum
Development tools and services	Access to build, test, integration services and development tools	hep-sf-tech-forum

Software Packaging

- Topics
  - package building, deployment, runtime environment, new technologies like Dockers, cmake best practices

- Organized a series of discussions/presentations on packaging and build tools (8 meetings)
  - Current practices inside and outside HEP
  - Document to summarize findings being prepared

- Trying a hands-on approach to increase share of actual code even if existing experiments and projects locked-in to a certain packaging solution
  - Common “build recipes” protocol

join the hep-sf-packaging-wg

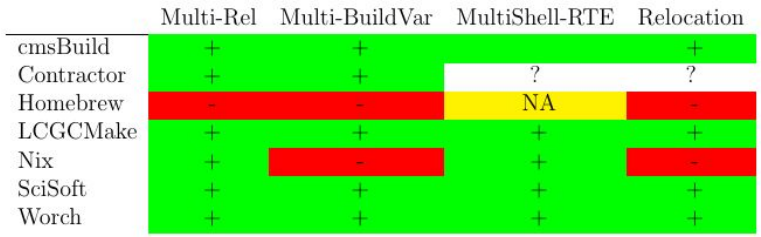
Build and Packaging Software Review

Looked at many tools, in particular

- worch, cmsBuild, aliBuild, LCGCMake, SciSoft, contractor (HEP), homebrew, Nix, conda

Main problems in HEP software

- reinvention of the square wheel
- non share even within the communit



	Multi-Rel	Multi-BuildVar	MultiShell-RTE	Relocation
cmsBuild	+	+		+
Contractor	+	+	?	?
Homebrew	-	-	NA	-
LCGCMake	+	+	+	+
Nix	+	-	+	-
SciSoft	+	+	+	+
Worch	+	+	+	+

Main problems in non-HEP software:

- non HEP-tools prefer rolling releases / care less about reproducibility
- little support for multi-environment setups

Evolving document available at https://github.com/HEP-SF/documents/tree/master/HSF-TN/draft-2015-PKG

Fostering Collaboration

HSF may have the role of foresting and publicising common software development initiatives. Some examples:

- Next-generation conditions data
  - Belle II / ATLAS / CMS / LHCb
  - 2 meetings: Dec 10, Jan 21
- Track reconstruction
  - Huge challenges ahead
  - Should try to not only share ideas, but concrete code
- Frameworks: Gaudi
  - ATLAS / FCC / LHCb

Satellite projects in the HSF constellation. Examples:

- DIANA (Data Intensive ANAlysis), 4-year NSF funded
  Focus on analysis software, including ROOT and its ecosystem

Technical Notes

- Technical Notes can be proposals, ideas, whatever people want to add
- First TN with the TN policy has been published
- Some more in preparation:
  - Licence Guidelines, Naming conventions, packaging tools landscape, ...
- Repository and version control in GitHub

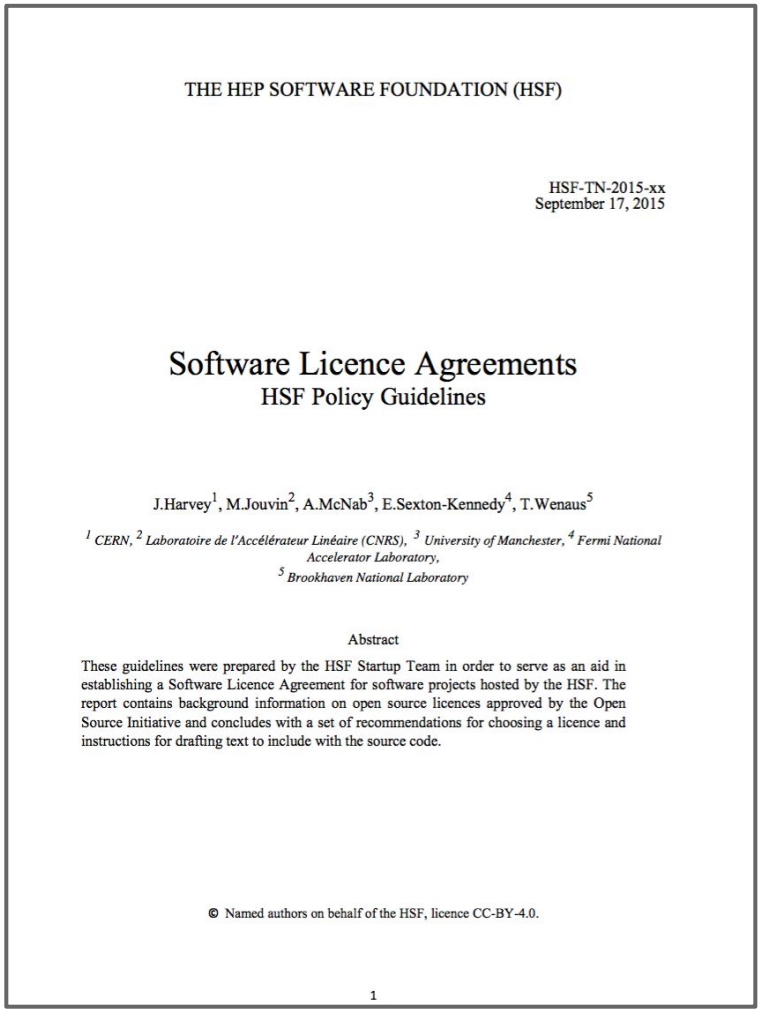
TN Number	Title	Authors	Download
HSF-TN-2015-01	HSF Technical Notes policy	A. McNab	PDF GitHub

Drafts in the acceptance process:

Draft TN Reference	Title	Authors	Download
HSF-TN-2015-LIC	(Draft) Software Licence Agreements HSF Policy Guidelines	J. Harvey et al.	GitHub
HSF-TN-2015-NAM	(Draft) HSF Platform Naming Conventions - A Proposal	B. Hegner	GitHub
HSF-TN-2015-PKG	(Draft) HSF Packaging Working Group Report	B. Hegner, L. Sexton-Kennedy	GitHub

Software Licensing

- TN for the HSF Licensing Guidelines is being finalized in the technical forum
  - Plan to publish it this week
- Background information on open source licences approved by the Open Source Initiative
- Set of recommendations for choosing a license and instructions for drafting text to include with the source code



THE HEP SOFTWARE FOUNDATION (HSF)

HSF-TN-2015-xx
September 17, 2015

Software Licence Agreements
HSF Policy Guidelines

J.Harvey[1], M.Jouvin[2], A.McNab[3], E.Sexton-Kennedy[4], T.Wenaus[5]

[1] CERN, [2] Laboratoire de l'Accélérateur Linéaire (CNRS), [3] University of Manchester, [4] Fermi National Accelerator Laboratory,
[5] Brookhaven National Laboratory

Abstract

These guidelines were prepared by the HSF Startup Team in order to serve as an aid in establishing a Software Licence Agreement for software projects hosted by the HSF. The report contains background information on open source licences approved by the Open Source Initiative and concludes with a set of recommendations for choosing a licence and instructions for drafting text to include with the source code.

© Named authors on behalf of the HSF, licence CC-BY-4.0.

WikiToLearn

- **WikiToLearn** is a wiki-based platform tailored at training and teaching
- Initiated in the context of italian universities
  - Basic idea was that students can improve and extend the material of their professors, while still being quality-controlled
- HSF jumped onto that to see whether we can take advantage of it
  - Started adding material to this site
- Now investing in providing interactive tutorials
  - think of the combination of jupyter style notebooks and a privately owned sandbox - start tutorial now, resume later (this even triggered a new collaboration w/ the ROOT team)
- This is only the shell, content has to come by the community

Some Challenges

- **Status of Technology Tracking**
  - Future HW architectures
  - New technologies/trends worth looking at
    (cloud based analysis, e.g. Data Mining-as-a-Service?)
  - Instrumentation and tools for measuring and improving SW performance
  - What else?

- **Evolution vs. Revolution**
  - Parallelism / vectorization implies revolutions in our SW
  - Challenge to backwards compatibility
  - Results of Revolution can still be included as evolution
    (GeantV's VecGeom as "preview" in Geant4 10.2)

- **Managing available manpower efficiently**
  - build up more commonality in software projects and procedures
  - prepare a common curriculum of development essentials
  - provide an easy entry point for people to apply best practices
    (HSF project template + infrastructure to set up + tools)
  - improve on quality and ease-of-use of the software we develop
    (less incentive on reinventing the wheel!)