Summary of ATLAS and LHCb SW Workshops

Benedikt Hegner
CERN

SFT Group Meeting
23.11.2015

The Events

ATLAS Software Technical Meeting, Nov 9-13 (Berkeley)

48 participants (including external invitees)

https://indico.cern.ch/event/395887/

6th LHCb Computing Workshop, Nov 16-20 (Paris)

~70 participants (including external invitees)

https://indico.cern.ch/event/337568/

The summary on the following slides is subjective and highly selective !

ATLAS Software Technical Meeting (a.k.a. TIM)

3 days of “plenary” technical discussions + 2 days of “hackathons” aimed at Run-3

- Core-Software
- Analysis
- Simulation
- Reconstruction
- Testing and Infrastructure

One presentation by a group member:

HEP Software Foundation (HSF) Initiatives (B. Hegner)

Core Software Session I

Project between ATLAS and CMS for a future conditions data project

- Addressing scaling issues and database complexity
- Interest as well by LHCb
- **After-workshop follow up with Belle-II (initiated by HSF)**

Various additions to Gaudi/Athena for GPU and many-core support

- Integrating the lessons learned from GaudiHive
- How to approach Regions-Of-Interest in a common scheduler with offline
- **ATLAS and LHCb on the same track for data handles in Gaudi again**

Core Software Session II

Discussing I/O challenges and problems

- Readiness for multi-threaded environment
- Callbacks causing coupling of Athena and Data Models
- No proper strategy for meta-data yet

Event Service

- Dynamically distributing ~15min of work items to worker nodes
- HPC as one use-case
- While solving some problems it may increase the overall complexity

Reconstruction Session I

Invited presentation: Writing Xeon Phi Optimized Code

- Not so straight-forward as our vectorizable loops are tiny

Two presentations about tracking and muon reconstruction SW

- Details of implementation
- Work on code optimization
  ⇒ already lots of work done
- Chances for parallelization
- **Lots** of technical details

Challenges imposed by Run-3 pile-up

Algorithm	Concurrent Cell	Number of cells	Tests	Rel. CPU of ID	Comments
Cluster creation	per module	O(1000)	no tests exist	O(5%)	output merging into identifiable containers
SpacePoint creation	per space point	O(10000)	no tests exist	O(<1%)	identifiable container, SIMD ?
SpacePoint seeded track finding	detector region	O(100)	GPU based test version from 2011	O(50%)	overlaps are dangerous
Ambiguity Solving	fitting: per track ambiguity: per tracks through shared	O(1000)	no tests exist	O(20%)	book keeping of hits is shared
TRT extension	per track	O(100)	no tests exist	O(10%)	

by A. Salzburger

Reconstruction Session II

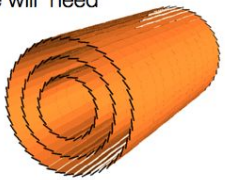
What are the chances of Machine Learning Methods for our tracking

⇒ New Machine Learning Challenge

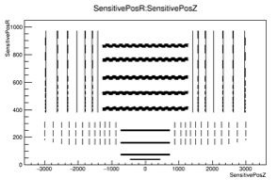
A HEP Tracking Machine Learning Challenge (3)

- Challenge will have to take several stages

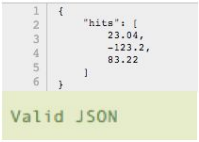
- we will need



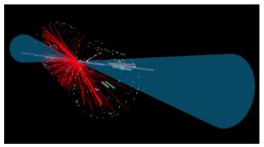
detector geometry
planar barrel/EC type detector
pixel/strip system



simulation
with the possibility to
simplify where possible



event data
easily readable,
platform independent



visualisation
of geometry,
hits & found tracks

well defined goal
what is success
and how we measure them

different categories
for different
solutions

by A. Salzburger

Simulation Session

Current simulation framework and the fast-/full simulation integration

coffee-chat: in retrospect using the Geant4 fastsim interface would have been better

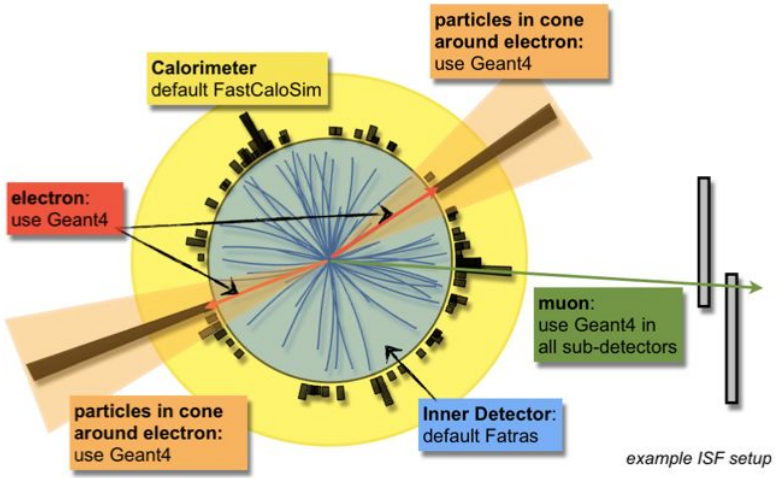
External status report about Geant4

Integration efforts of parallel Geant4 and GaudiHive

- Work induced by impedance mismatch of multi-threaded vs. task-based designs

Discussion on special needs for Run-3 simulation

- Pile-Up handling
- Resource needs



by E. Ritsch

Analysis Session

IMO more a summary of status-quo than a discussion session

Current status of standalone analysis vs. Athena based one

- Well-known problem of users going for (perceived) easiness
- Requirements for better decoupling of components in Athena

ROOT 7 ideas (by Philippe Canal)

- Good that he was presenting it! People were surprised

Analysis-focused software releases

- Yet another solution presented
- Virtualization / CernVM not considered a proper solution
- Requires follow up in LIM and HSF

Testing and Infrastructure

Presentations about code checking tools and principles

- Static code checking based on GCC
- Dynamic code checking using various sanitizers
- CMS' and Danilo's work on the static code checker SAS unknown

Touched general strategies

- Need for more unit testing
- More active quality control
- Discussed strategies for proper teaching
- Jenkins as replacement for ATLAS' nightly infrastructure (NICOS)
- Proposal to move to a GitHub/GitLab centered integration procedure
    - would replace tag collector, foster code reviews

IMO there was consensus on importance of code checking and the need to replace current procedures

ATLAS “Hackathon”

I did not participate so I only know that they tackled

- Conditions Data Access
- The ML Tracking Challenge

6th LHCb Computing Workshop

5 days mixed between plenaries and parallel working groups

Plenaries

- Lessons from Run1
- Beyond the Looking Glass
- Detector Descriptions
- Offline/Grid/Simulation

Parallel working groups

- Collaborative Working
- Data processing/Analysis Models
- Event Model
- Scheduling and Framework
- Hardware and Dataflow

Contributions by our group

- Geant4 in 2020 - J. Apostolakis
- Detector Description in Geant4/V - J. Apostolakis
- Simulation Software for FCC - B. Hegner
- Defining Control Flows - B. Hegner
- Using PODs for Data Models - B. Hegner
- Towards ROOT 7 (part 1; part2) - A. Naumann

Summary and Highlights of the Plenaries

- Many state-of-the-art summaries of Geant4, ROOT, ...
- Geometry seems one of the hot topics for Run-3
    - not clear whether DD4Hep could be something for LHCb
- Thoughts on how to tackle simulation just starting
    - ATLAS presented their approach
    - Plan to go for a common FCC/LHCb effort on simulation
- Presentation about HPX library by Th. Heller
    - Queuing for becoming C++ standard
    - Huge overlap w/ the TBB features we use
    - **Worth having the speaker coming to CERN again**
- Handling large-code bases in Chrome by Th. Nagel
    - Still on the train then, but told it was good

Event Model WG Summary

Very exploratory approach with presentations about

- Clear that current data model has to change
- POD usage proposal and prototype
- AoS and SoA prototypes
- Vectorization options
- Packaging needs (i.e. mapping precision needs to #bits stored)
- Role and options of meta-data
- Data preservation

Hacking sessions gave concrete results

- Prototype PODification of MC part of LHCb data model
- Simplification of KeyedContainer / improving movability
- LHCb::Math moved to using intrinsics

Scheduling WG Summary

As an 80% overlap of participants partially merged w/ Event Model

Prepared many potential work items

Group’s actual work centered around

- Proper work scheduling approaches
- Re-inforcement that multi-threading is the way to go
- Truly (event-)stateless GaudiAlgorithms - “TransformAlgorithm”
- The importance of testing, integration, and code reviews

Conveners seemed not too happy about the hands-on part

- Hands-on **really** needs preparation and more concrete goals
- Pick’n’choose only works for real experts, not for new volunteers

Collaborative Working WG Summary

In “poster sessions” people were supposed to present their projects and ideas. following the

https://en.wikipedia.org/wiki/Open_Space_Technology

In the end the following two main themes emerged

- Supporting analysers:
  - Automated analysis workflows
  - Reproducible/reusable code - Everware project
  - Integration of machine learning challenges
- Development workflows and procedures
  - Using Git and GitLab
  - Coding rules and coding review guidelines

Consensus on goals and candidates for technical solutions. Sociological “implementation” not yet clear

Data Access WG Summary

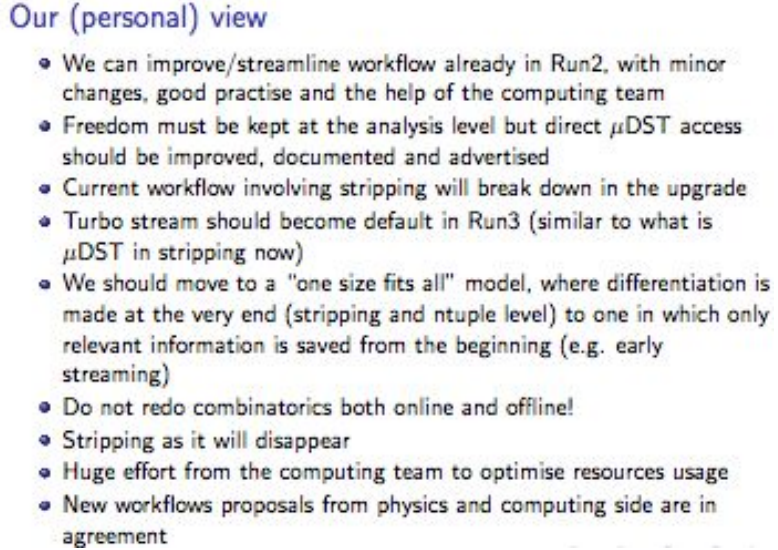
Input from various physics working groups on current patterns and problems

- Baseline for LHCb is full-reco at trigger-stage
- How to continue from there?

Surprisingly quick convergence (?)
on ideas and proposals

Nice summary of LHCb computing history and future by Ph. Charpentier

- Showing project planning plots created by John Harvey more than 15 years ago



WG summary by A. Contu

Conclusions of both workshops

A few general directions are clear

- Static code checking seems to gain momentum
- Both collaborations plan to move to a GitHub/GitLab based approach
  - CMS serves as very successful example
- Questions of procedure efficiencies popping up all over the place
- Task-based model remains the way to go for parallelization

IMHO (almost) all the external experts gave valuable input

Both collaborations still need to define their direction based on the input

Differences between the two workshops

Quite different dynamics

- “What I did so far” (ATLAS) vs. “let’s try out a few new things!” (LHCb)
- Only little discussion in LHCb plenary, but plenty of in breaks
- Level of team spirit

Funding agency forces more visible in ATLAS

- Strong “boundaries” for who works on what
- HPC boundary conditions mentioned many, many times

LHCb surprisingly bigger than ATLAS !

Take away messages for our group

The projects our group is involved in were well represented
and it was important that they were!

It seems CVMFS is blindly assumed to be available everywhere

Some of our efforts weren't known well enough

- ROOT 7 ideas (despite so many presentations!)
- SAS static code checker

Both LHCb and ATLAS are very worried about PyROOT future

Having a closer look at the HPX library is worthwhile