Summary of Track-2

Offline Software

Andrew Norman, **E**lizabeth Sexton-Kennedy, **H**isaya Kurashige, **I**van Kisel and **J**ames Kowalkowski

Overview

Track-2 Sessions:

#1 Mon: Reconstruction
#2 Mon: Exp. Computing Models
#3 Tue : Frameworks
#4 Tue : Simulation
#5 Thu : Analysis
#6 Thu : Tools

Mon-Thu: Enjoyed Okinawa and OIST

#1
- ptimisation of the ATL...
- Simulation and Recon...
- CMS reconstruction...
- Optimization of the LHC...
- Event Reconstruction...
- Kalman Filter Trac...
- Cellular Automaton...
- 4-Dimensional...
- Poster session A: unnel Gallery

#2
- tatus and future evol...
- Implementation.
- ATLAS Metadata I...
- The Data Quality Mo...
- MiniAOD: A new analys...
- MAUS: The MICE Anal...
- The NOvA Simulation...
- The Heavy Photon Se...

#3
- nning ATLAS wor...
- A New Petabyte-s...
- Requirements...
- Development of a Next G...
- Using the CMS Threa...
- New developments...
- ALFA: The new ALICE...
- Physics Analysis S...
- Coffee break ST

#4
- eant4 Version 10 ...
- Geant4 VMC 3.0
- slic : A full-featured G...
- The GeantV project: pr...
- CMS Full Simulation...
- CMS Detector D...
- Continuous Readout Si...
- Energy Reconstruction...

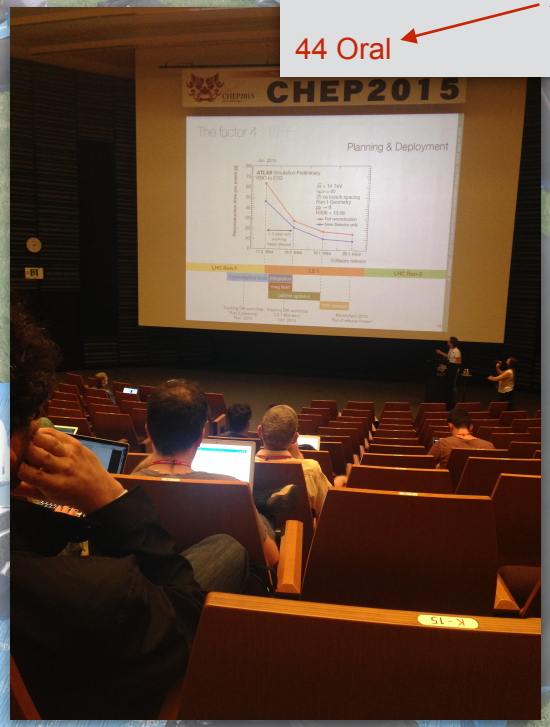
#5
- e ATLAS Higgs Mac...
- How the Monte Carl...
- The Library Event Matc...
- Interpolation ...
- CosmoSIS: a system for ...
- The Bayesian a...

#6
- Poster session B: nel Gallery
- sing R in ROOT with ...
- HistFitter: a flexible fra...
- Accelerating Scientific A...
- DDG4 A Simulation...
- BESIII Physics Da...
- Breaking the Silos: The ...

95 Track-2 related Abstracts

44 Oral

51 Poster



Optimisation of the ATLAS Track Reconstruction Software for Run-2

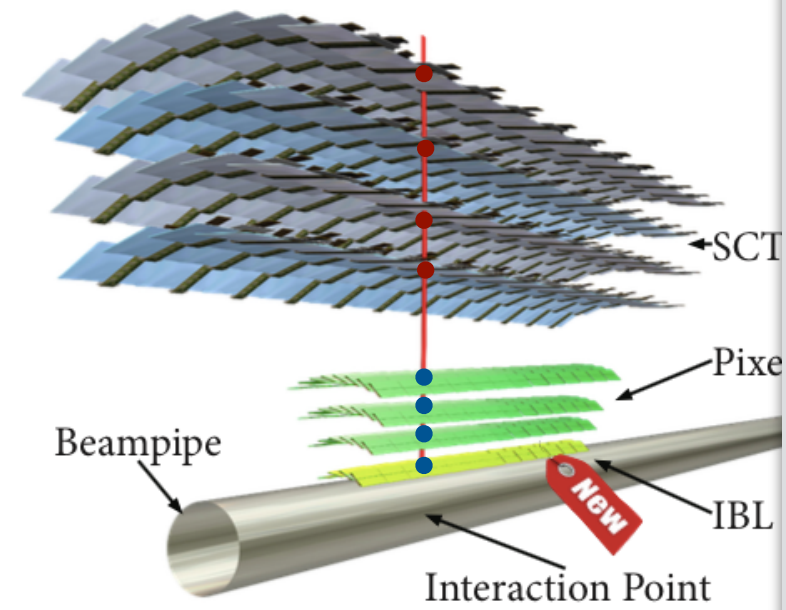
#1 Mon: Reconstruction

A. Salzburger

Kalman Filter based track following

- Track reconstruction software for Run-1 was designed with a lot of redundancy and safety margin
- Run-1 performance convinced us that our system was understood
  - and extremely well modelled by MC
- Re-investigation of track seeding
  - taking high purity seeds from strip
  - make optimal use of new innermost Pixel layer (IBL)
- Greatly improved the seed purity

*managed to be **more efficient** in less time: ~25 % saving*



triple seeds can we built as:
- pixel space points only (PPP | ●●●)
- strip space points only (SSS | ●●●)
- a combination of both, e. g. (PSS | ●●●)

Summary:
- We achieved a factor 4 speedup of the overall event reconstruction time
  - at the same time improving physics performance on all ends
  - mainly achieved by the Inner Detector track reconstruction
  - ready for Run-2 data taking

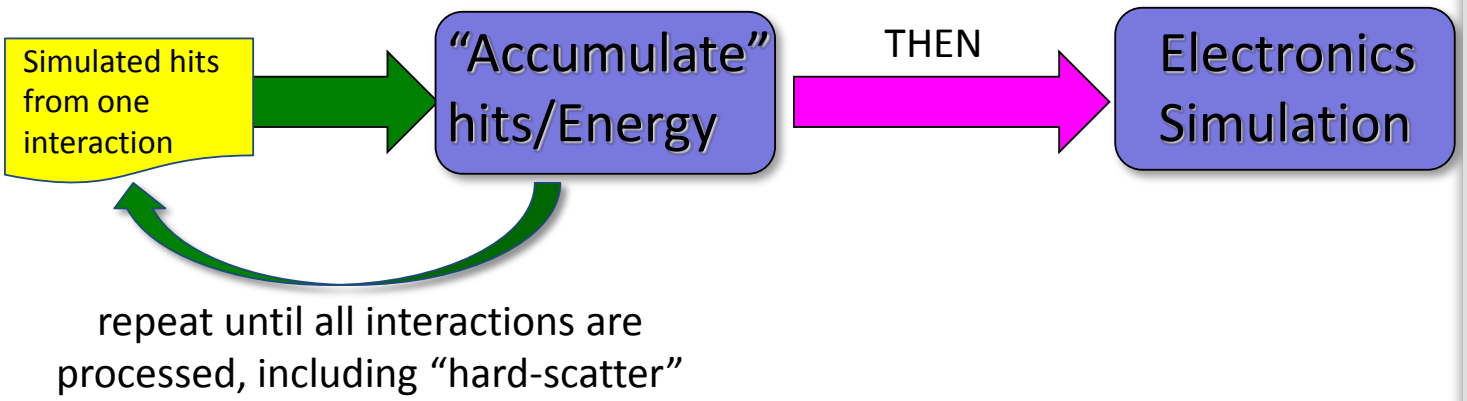
Upgrading the CMS Simulation and Reconstruction

#1 Mon: Reconstruction

D. Lange

We have achieved significant progress on the resource needs of our simulation during LHC shutdown

- We re-factored the pileup simulation to process each interaction sequentially
  - Required substantial rewrite of digitization code, and the re-organization of internal event processing



- The content of each event is dropped from memory once processed:
  - Only 1 event in memory at any given time, so arbitrarily many pileup events can be included in the digitization
- Next challenge in pileup simulation for CMS: Reduce the I/O burden from the pileup events to open up more resources for processing

Summary:

- CMS detector upgrades push us to use today's software/computing for tomorrow's event complexity.
  - The detector upgrade developments have proven to be an excellent platform for the quick deployment of new simulation features

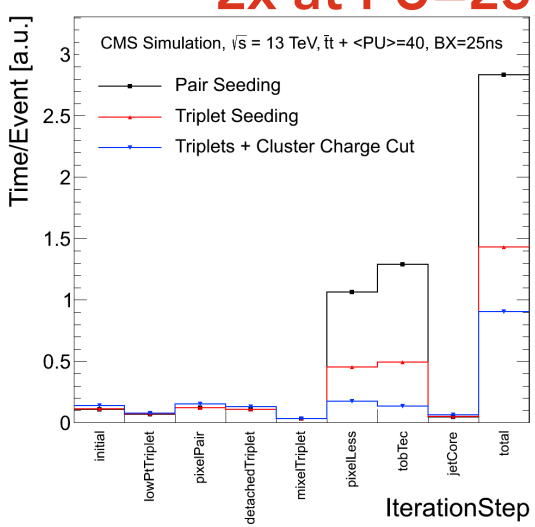
CMS Reconstruction Improvements for Tracking in Large Pile-up Events

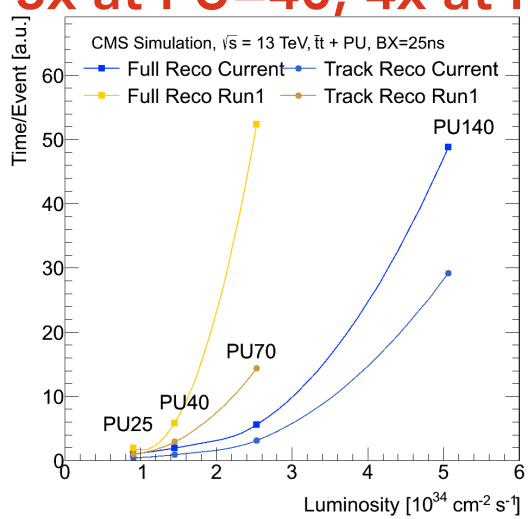
#1 Mon: Reconstruction

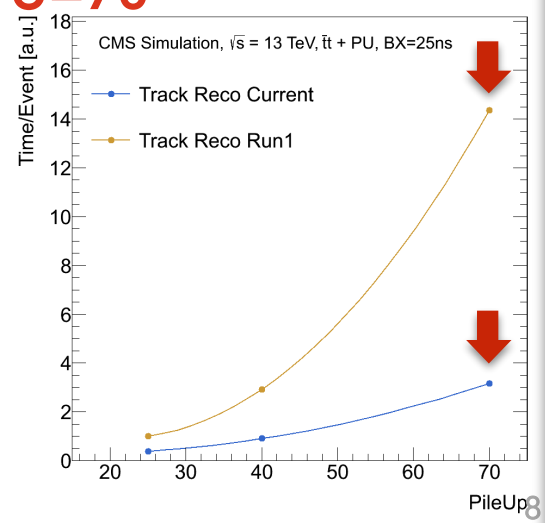
M. Rovere

Kalman Filter based track following

- The new seeding and the cluster charge cut reduce timing of PixelLess and TobTec by 2x
- Physics performances and timing in different conditions:
  - TTbar samples with
  - BX=25 ns, <PU>=25, 40, 70, 140
  - BX=50 ns, <PU>=25
- Iterative tracking time reduction @25 ns:
  - **2x at PU=25, 3x at PU=40, 4x at PU=70**







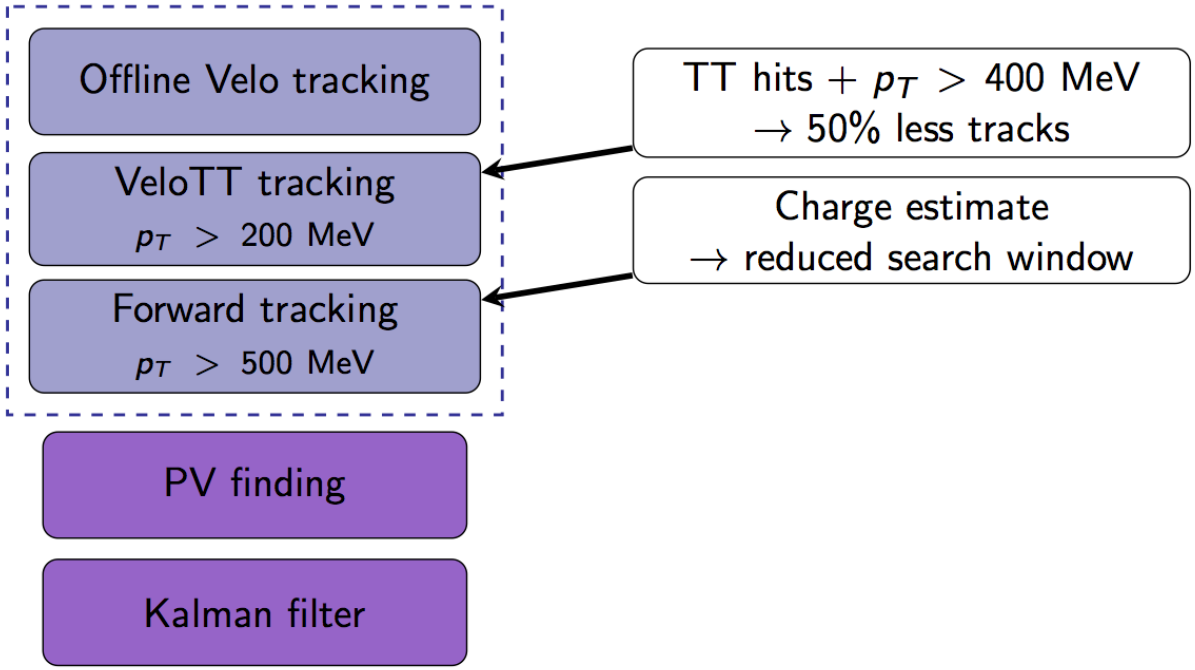
Summary:

- Many developments have been included in CMS' tracking code for Run2
  - Timing is now under control
  - Should expect the same or better physics performances as in Run1
  - Work is not over and many other developments are on their way

Optimization of the LHCb Track Reconstruction

B. Storaci

Kalman Filter based track following



- 3 times faster reconstruction chain $\rightarrow$ HLT1 runnable in $\sim$32ms with looser requirements than in run I
  - 500MeV p_T tracks available already at HLT1 level!
- 4 times less ghost rate, i.e. 6%
- 87% single track efficiency for $p > 3$GeV, $p_T > 500$MeV

Summary:
- Several studies done to improve the performance of the reconstruction chain to face the new challenging condition of Run II

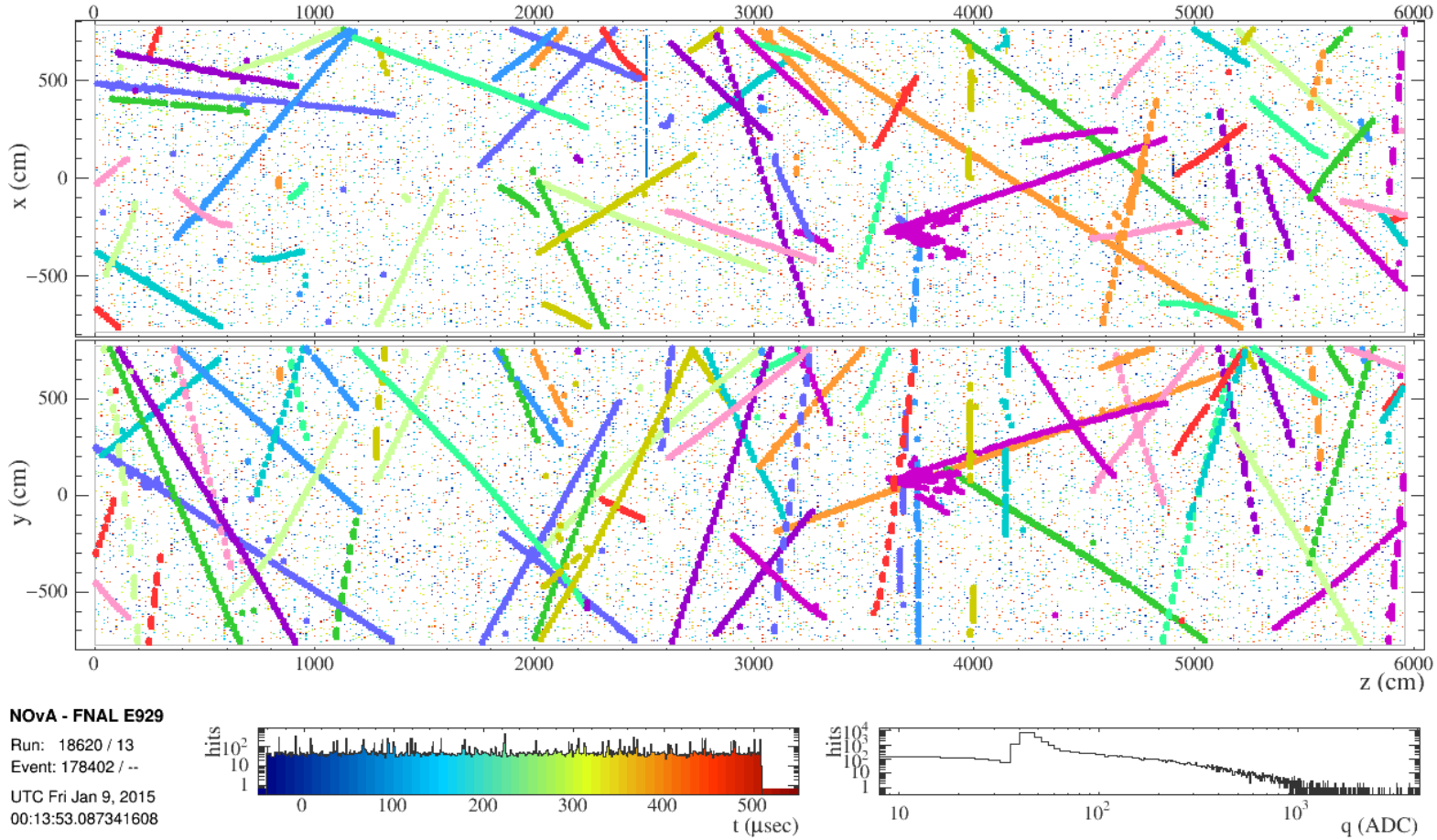
Event Reconstruction Techniques in NOvA

#1 Mon: Reconstruction

D. Rocco

- Slice borders are defined by regions where the neighborhood density drops below some critical value.
- The algorithm expands from neighbor to neighbor to find all borders.

A far detector event after slicing has been applied.



Summary:

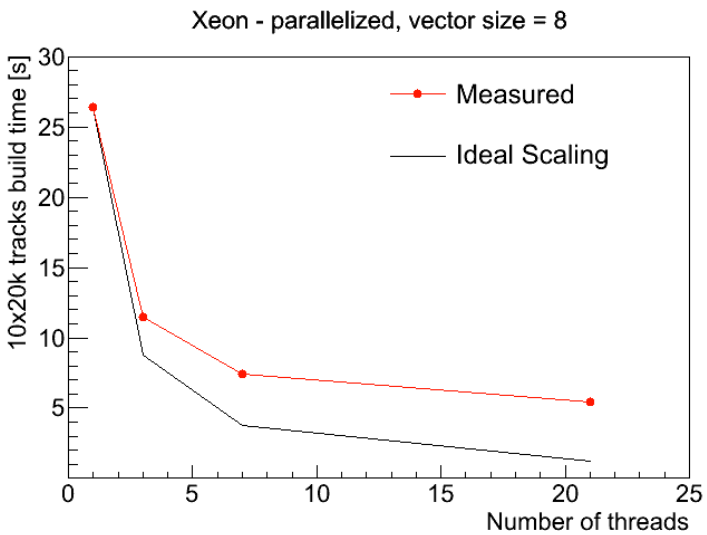
- A series of pattern-recognition algorithms have been adapted and chained into a vertex-first approach that works well for short and long tracks or showers
- Reconstruction forms basis of neural net for electron neutrino event classification

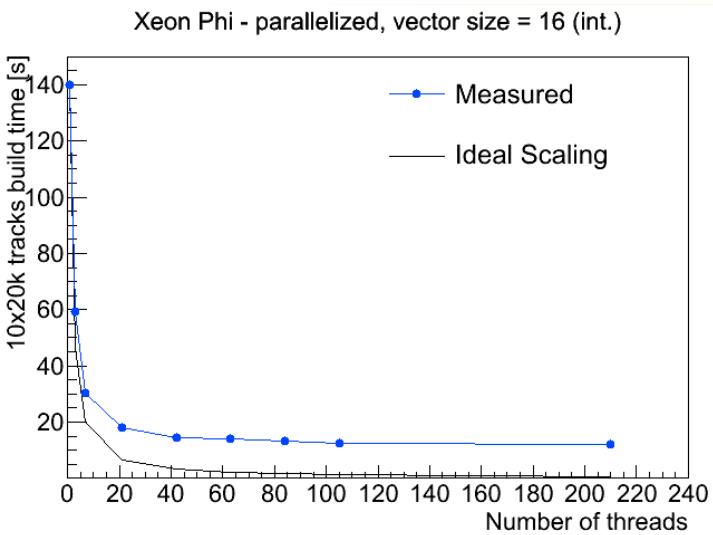
Kalman Filter Tracking on Parallel Architectures

#1 Mon: Reconstruction

G. Cerati

Kalman Filter based track following





- Run **full track building** with combinatorial expansion of candidates
  - ultimate physics performance, slower
  - 85% (95%) of tracks found with ≥90% (60%) of the hits
- Parallelization is implemented by **distributing threads across 21 eta bins**
  - for nEtaBin multiple of nThreads, split eta bins in threads
  - for nThreads multiple of nEtaBin, split seeds in bin across nThreads/nEtaBin threads
- Large **speedup** achieved, both on Xeon and Xeon Phi
  - up to **~5x on Xeon and >10x Xeon Phi**
  - speedup saturates above nThreads=42

Summary:
- Significant speedup achieved both on Xeon and Xeon Phi.
- Ideal scaling indicates a large margin for further improvements.

Cellular Automaton based Track Finding for the Central Drift Chamber of Belle II

O. Frost

Approximate the Hopfield-Network more closely

- Allowing weights on vertices θ_i and edges w_{ij}
- Weighted evaluation rule picking the neighbor with highest weight gain

$$E_i = \theta_i + \max_{\text{neighbor } j} (w_{ij} + E_j) = \sum_{\text{best path to } i} w_{ij} + \sum_{\text{best path to } i} \theta_i$$

Inherited properties

- High evaluation speed by single pass over the graph (forward pass, $O(n)$ - algorithm)
- Generation of valuable tracks following cells with highest state (backward pass)
- Iteration for additional disconnected tracks
- Emphasis on local connections - robustness against separated background and energy loss
- Agnostic to start position and direction of flight - seeding phase unnecessary

Upgraded properties

- Refined track model by encoding probability of positions and propagations into the weights

Summary:

- Best weighting model under investigation
- First application to the data from cosmic data taking starting this spring

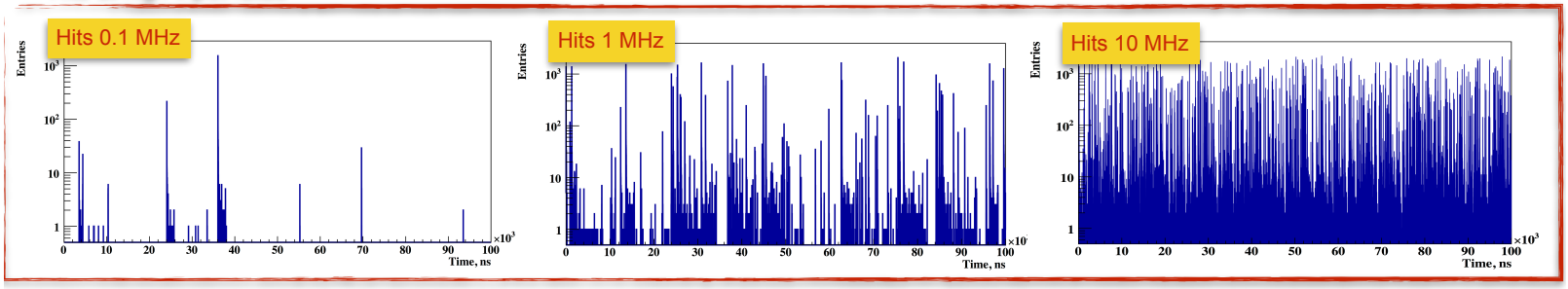
4-Dimensional Event Building in the First-Level Event Selection of the CBM Experiment

#1 Mon: Reconstruction

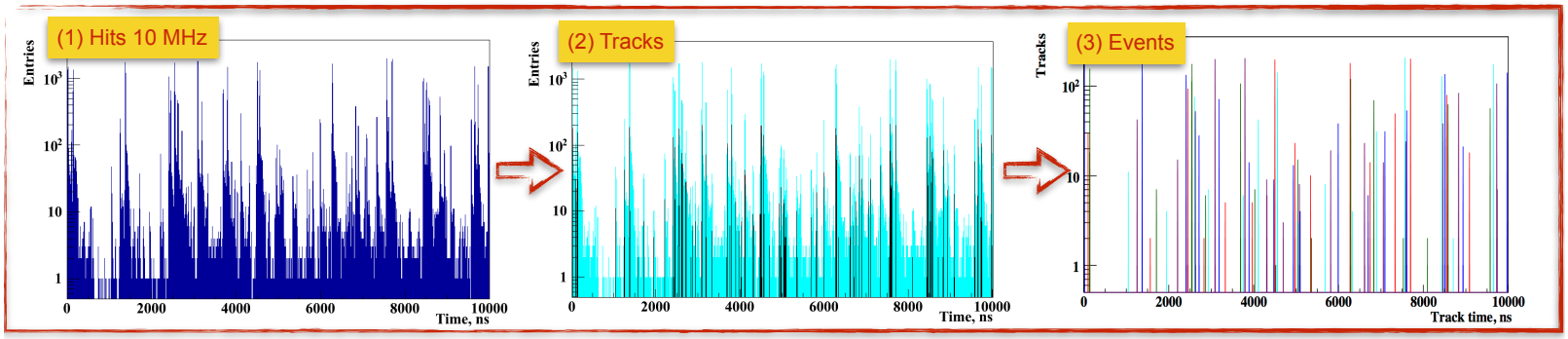
I. Kisel

Cellular Automaton based track finder

Hits at high input rates



From hits to tracks to events



Reconstructed tracks clearly represent groups, which correspond to the original events
83% of single events, no splitted events, further analysis with TOF information at the vertexing stage

Summary:

- The Kalman Filter track fit library is vectorized, parallelized and portable to CPU/Phi/GPU architectures.
- The Cellular Automaton track finder is vectorized, parallelized and updated for time-based (4D) track finding in time-slices.

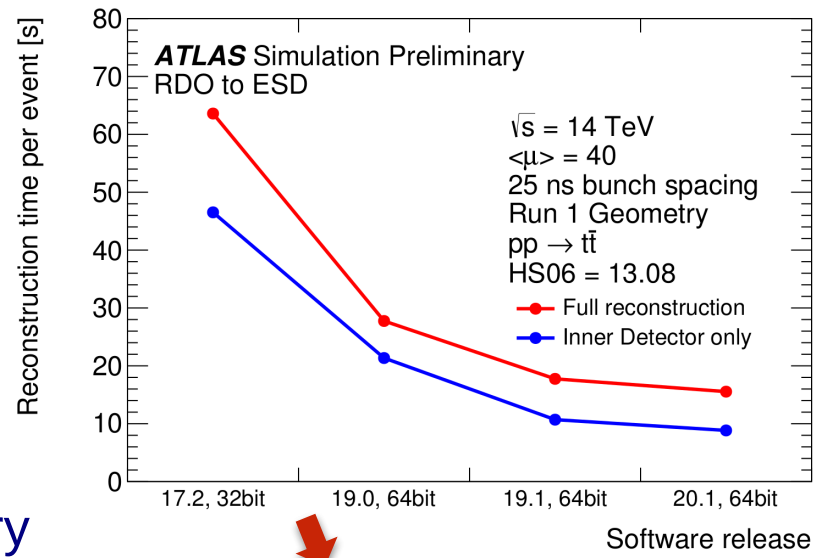
Status and Future Evolution of the ATLAS Offline Software

#2 Mon: Experiment's Computing Models

R. Seuster

Since the end of LHC Run 1, ATLAS improved the full analysis chain from RAW data to final ntuples, addressing a few issues seen as potential problems for Run 2

- Significant speedup of reconstruction to cope with 1kHz EF rate: ~4x !!
- In Run 1 Inner Detector dominated processing time, most work done:
  - more modern Matrix library
    Eigen replaced CLHEP: faster due to SIMD intrinsics, ...
  - general code cleanups, Magnetic field rewritten in c++
  - 'free' lunch: newer gcc, Intel mathlib, SLC6, 64bits



Summary:

- Since end of Run 1, ATLAS updated offline Software significantly to match new requirements
  - factor of ~4 speedup of data reconstruction
  - overhauled analysis model including EDM
  - multi-processing software in Production

Implementation of the ATLAS Run 2 Event Data Model

S. Snyder

The new event data model is one of the contributors to the improvements in reconstruction time

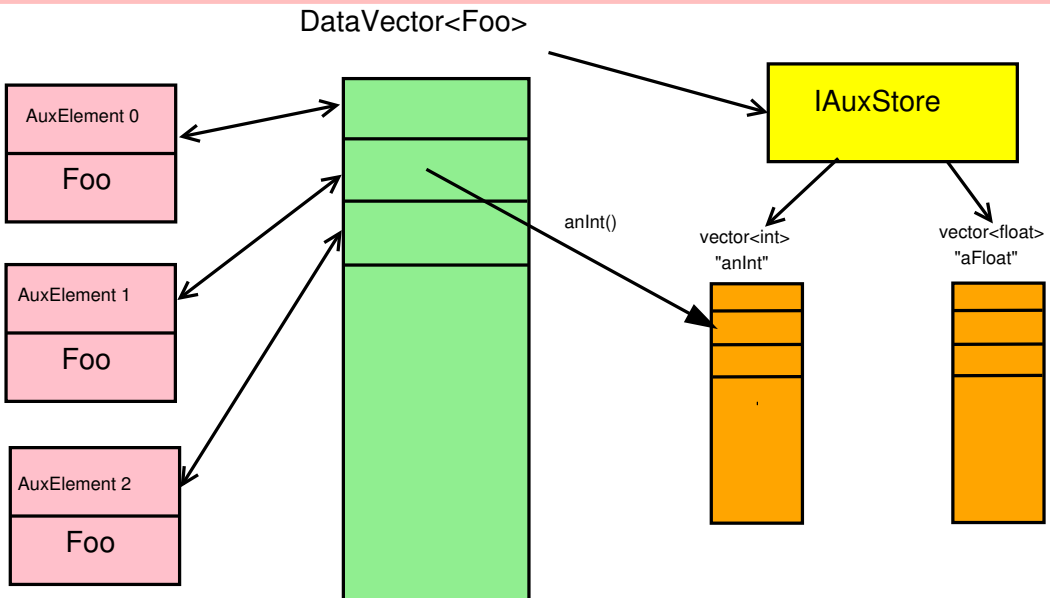
Analysis-level event data model redesigned for Run 2. ("xAOD")

- Simplify, and make more directly usable with ROOT.

Based on new "auxiliary data" feature of `DataVector`.

- Attach data of arbitrary type to elements of `DataVector`.
- Data are stored as vectors, managed via separate "auxiliary store" object via abstract interface.
- Object data stored as auxiliary data rather than in the object itself.

Almost all object data stored as auxiliary data.



Summary:

- Based on a unified "auxiliary store" design, providing:
  - Data stored internally as a set of vectors.
  - Data are directly readable from ROOT.
  - User analysis code can add arbitrary new data items to objects.

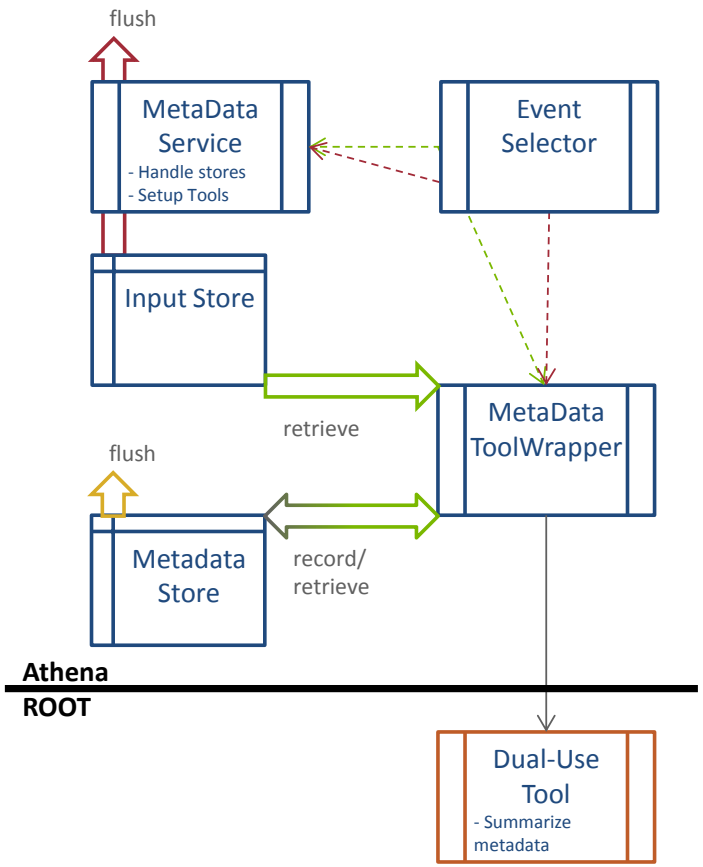
ATLAS Metadata Infrastructure Evolution for Run 2 and Beyond

#2 Mon: Experiment's Computing Models

P. van Gemmeren

A common Event Data Model for Athena and ROOT analyses requires to share metadata and tools handling metadata.

- ATLAS has changed its event data model to unify Athena and ROOT analyses (xAOD).
- To allow the reuse of metadata components in downstream analyses that are not utilizing the ATLAS control framework, Dual-Use Tools are being developed to summarize metadata records.
  - Transfer some functionality from the framework MetaDataTools to Dual-Use Tools
  - Provide generic MetaDataTool-Wrapper to allow framework integration
    - Listen to incidents
    - Interact with stores



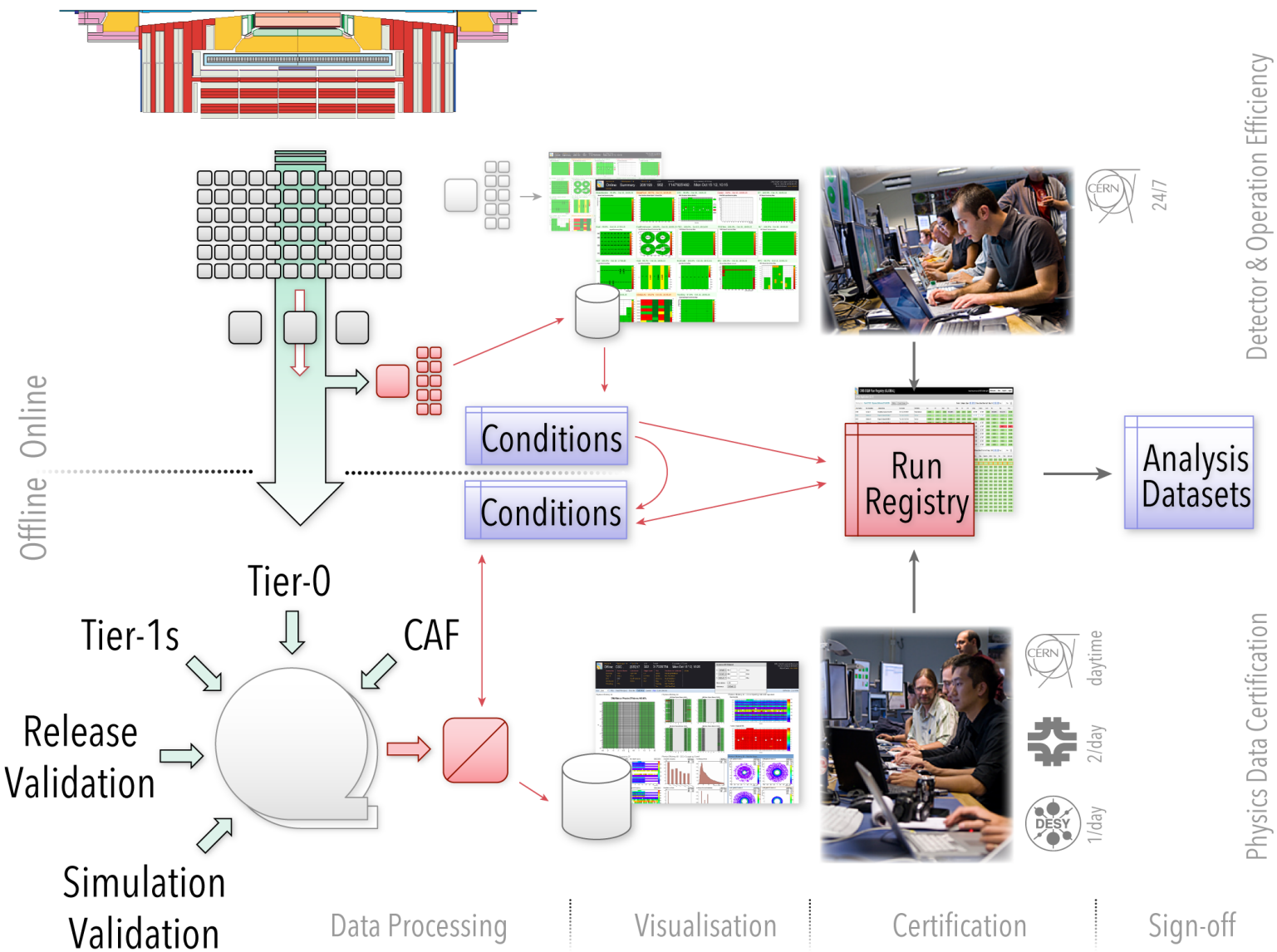
Summary:

- Run 2 conditions further emphasize the importance of metadata for distributed data processing and analysis.
- At the same time, the move to new computing architectures requires extensions to the metadata infrastructure.

The Data Quality Monitoring Software for the CMS experiment at the LHC

M. Rovere

DQM framework proved to be extremely flexible and stable: DQM is used everywhere in CMS



Summary:

- DQM Framework adapted and improved in the face of fundamental changes in the Online (DAQ2) and Offline (Multithreading) environments

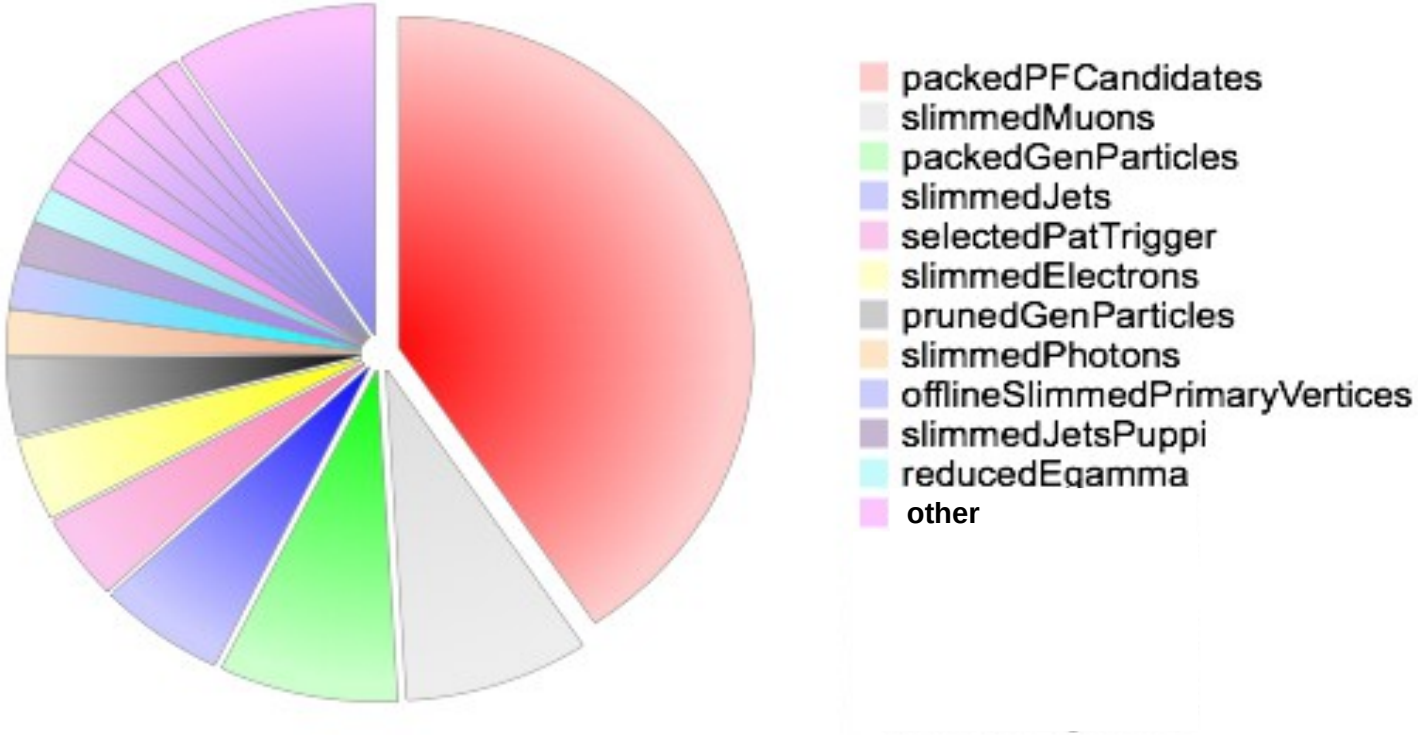
Mini-AOD:
A New Analysis Data Format for CMS

#2 Mon: Experiment's Computing Models

C. Vuosalo

Previous method of storing AOD and many versions of intermediate ntuples would overflow storage capacity many times over

Overall size for ttbar MC: ~40 kB/event



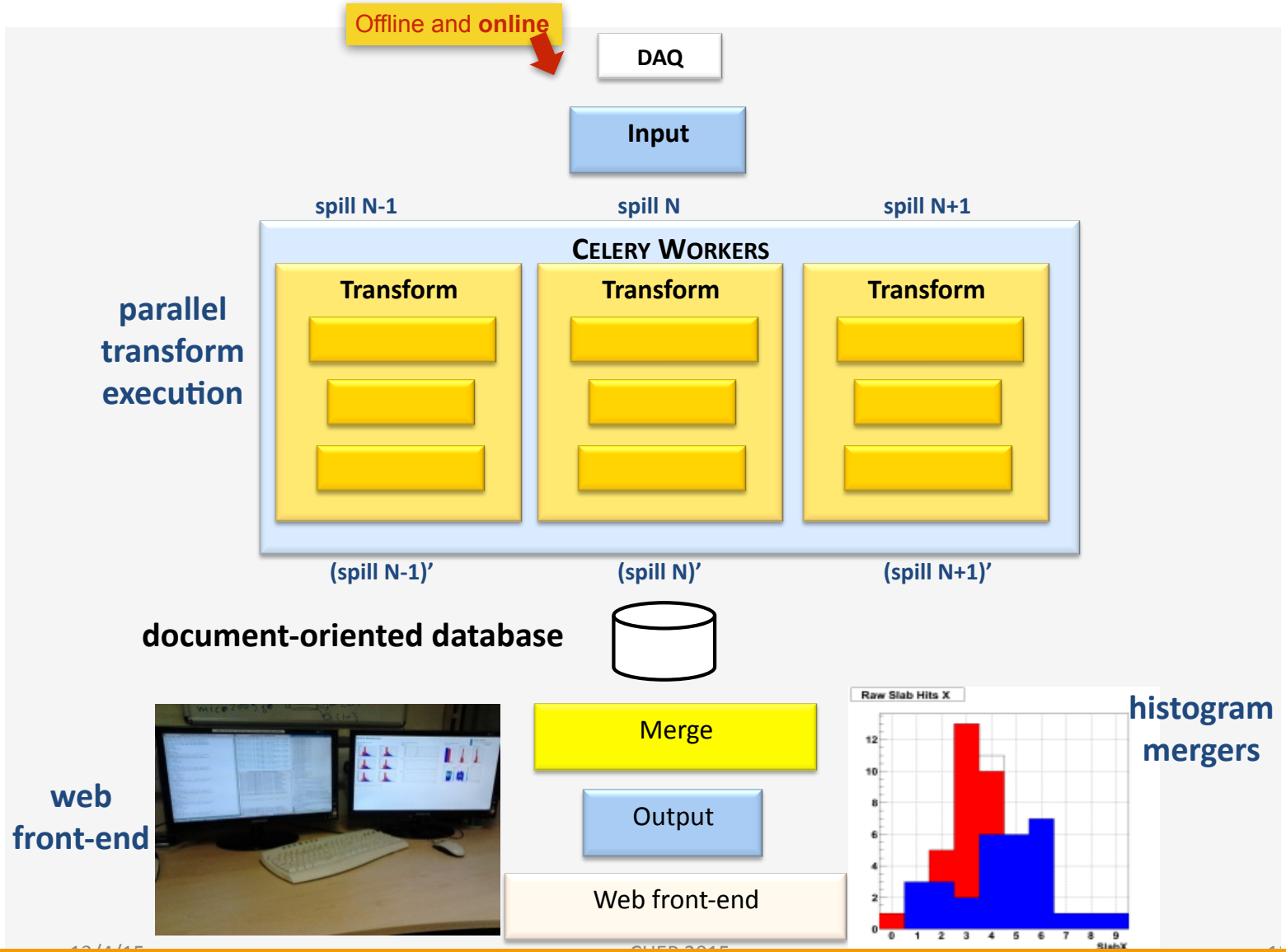
Summary:

- Solution is compressed Mini-AOD
  - T10% of size of AOD
  - Replaces intermediate ntuples with standard format for most analyses

MAUS:
The MICE Analysis and User Software

D. Rajaram

MICE: Muon Ionization Cooling Experiment



Summary:

- MAUS provides a simulation, reconstruction, and accelerator physics analysis framework for MICE

The NOνA Simulation Chain

A. Aurisano

NOvA is a long-baseline neutrino oscillation experiment

Simulation Chain

Beam simulation → FLUKA/FLUGG

Simulates the incoming neutrino flavor and energy spectrum from π, K, and μ decays

Neutrino interactions → GENIE

Cosmic rays → CRY

Produces particle lists and kinematics to be propagated through the detector

Detector simulation → GEANT4

Propagates particles through the detector and produces energy deposits in active material

Parametrized front-end simulation

Converts energy deposits into scintillation light, transports scintillation light to the APD, and simulates the readout response.
→ output looks like raw data

Summary:

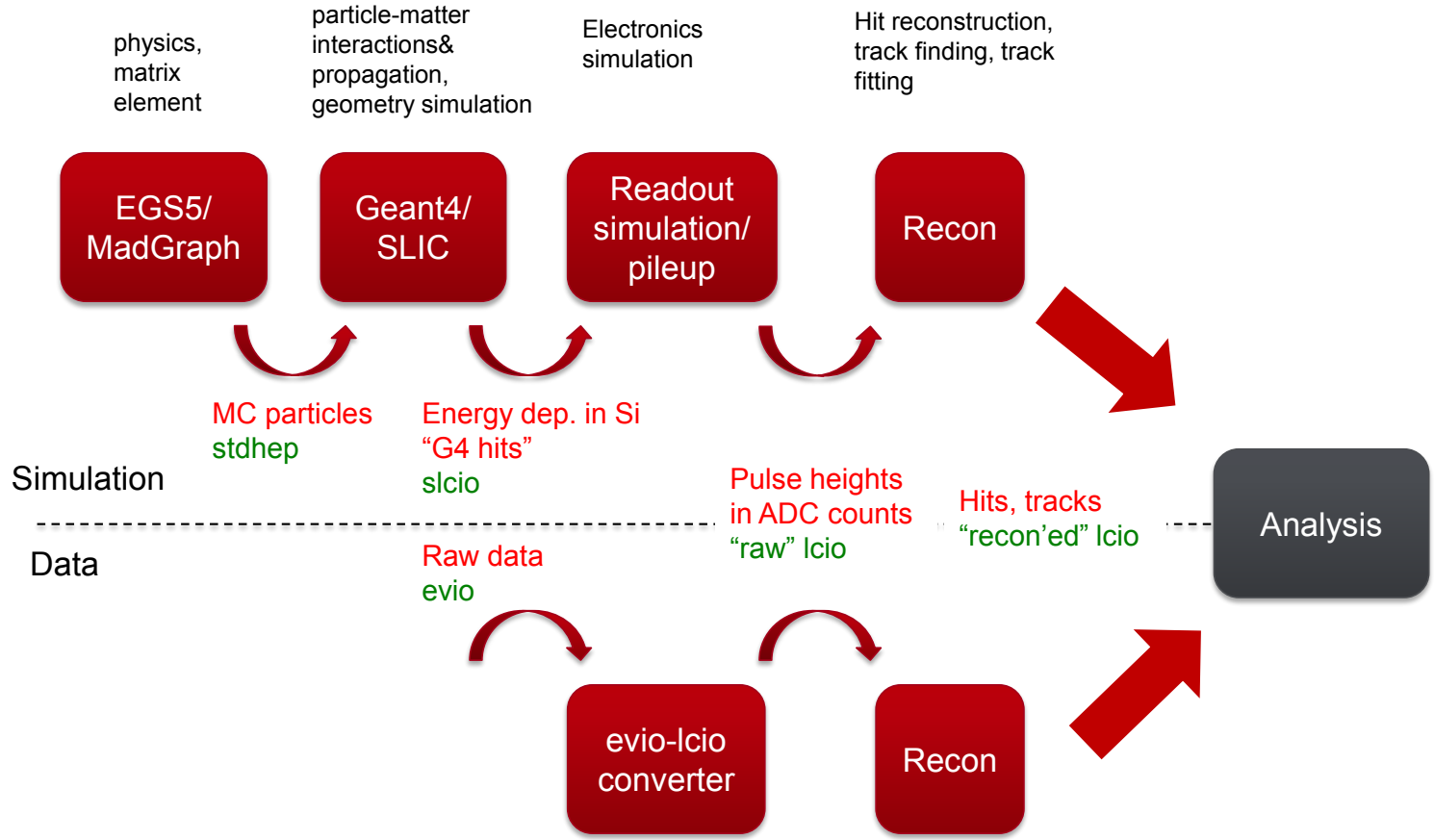
- Simulation at NOvA is a multi-stage process.
- Preliminary results show that low level quantities in cosmic ray data and Monte Carlo match well.

The Heavy Photon Search Experiment Simulation & Reconstruction

#2 Mon: Experiment's Computing Models

N. Graf

- Heavy photons have recently become popular since they could explain experimental anomalies in particle physics (g-2) and astrophysics (e+ excess).
- The Heavy Photon Search (HPS) Experiment is a fixed-target experiment that uses the JLab electron beam to search for heavy photons with masses 10-200 MeV.



Summary:

- Use of existing simulation and reconstruction software (developed for the ILC collider detectors) minimized the time needed to design and optimize detector.
- Java-based reconstruction software working well.

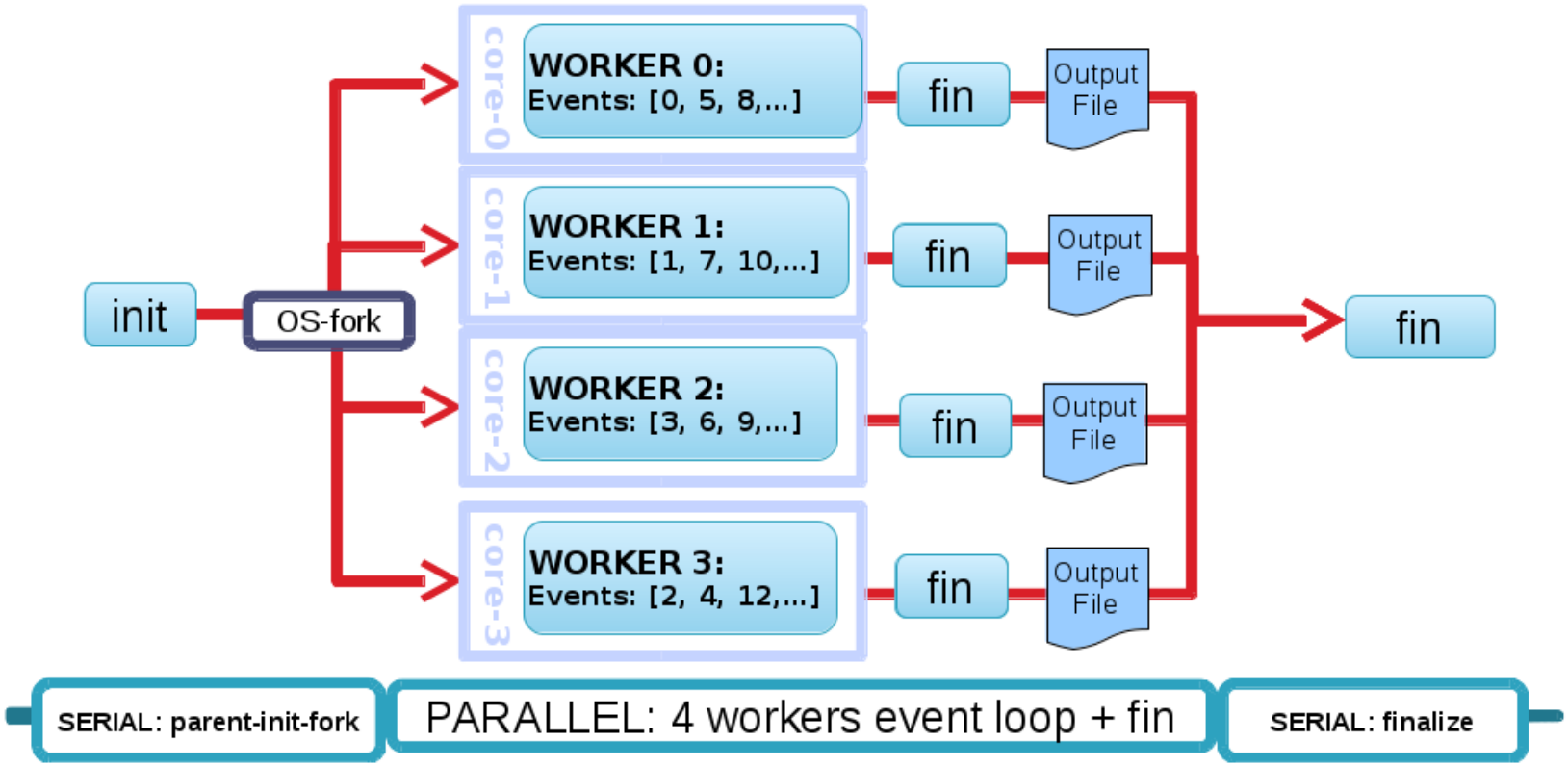
Running ATLAS Workloads within Massively Parallel Distributed Applications using Athena Multi-Process Framework (AthenaMP)

#3 Tue: Frameworks

V. Tsulai

Currently is actively used for running ATLAS production jobs on multi-core resources on the Grid

Schematic view of ATLAS AthenaMP



Summary:

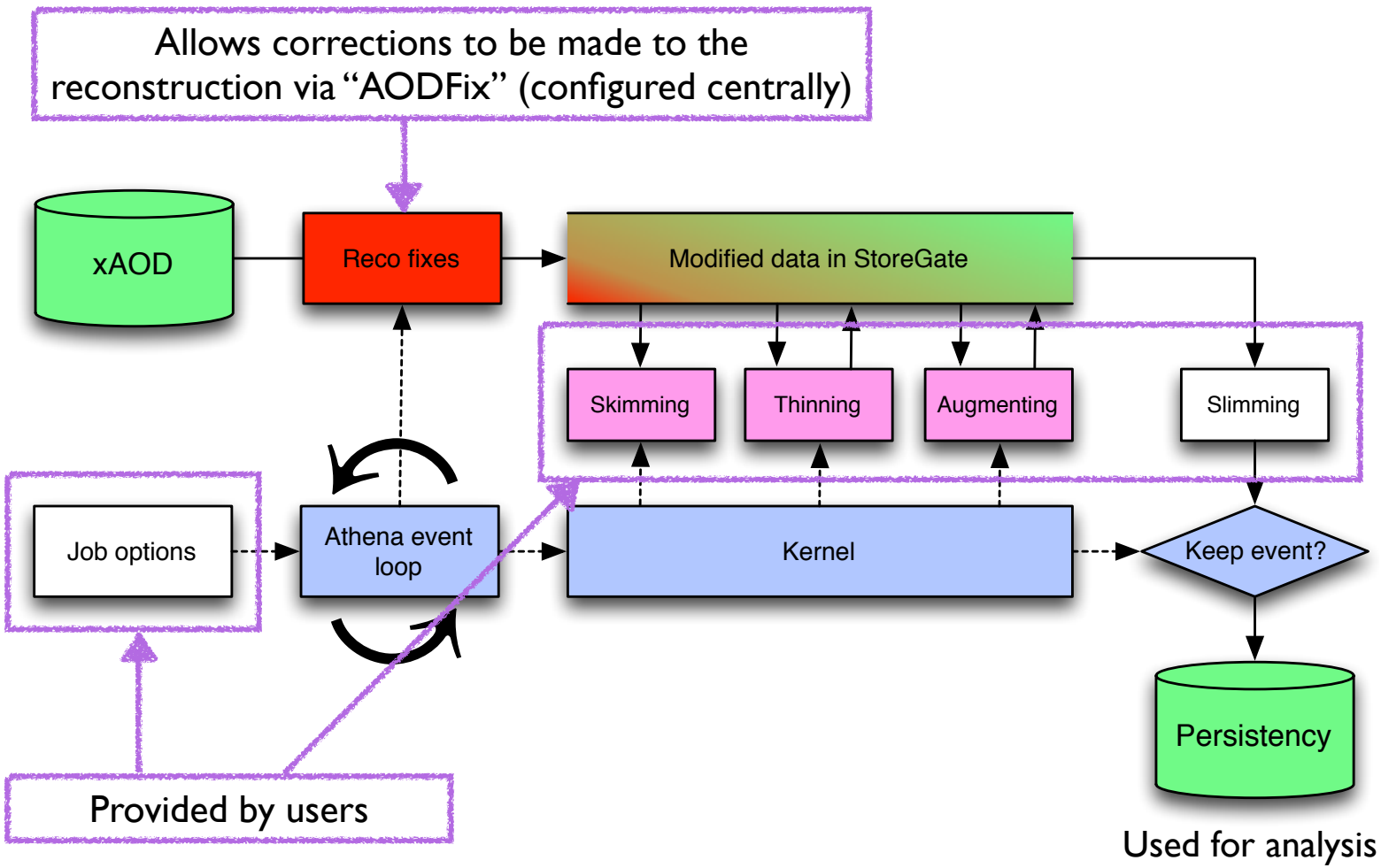
- Recent tests on supercomputers demonstrated, that AthenaMP can run efficiently within distributed applications at large scale.

A New Petabyte-scale Data Derivation Framework for ATLAS

#3 Tue: Frameworks

J. Catmore

The framework is built on the main ATLAS data processing framework (Athena)



Summary:

- The framework is in use already and the output data products are within the resource limits.
- It will be deployed from the start of data taking next month.

Requirements for a Next Generation Framework: ATLAS Experience

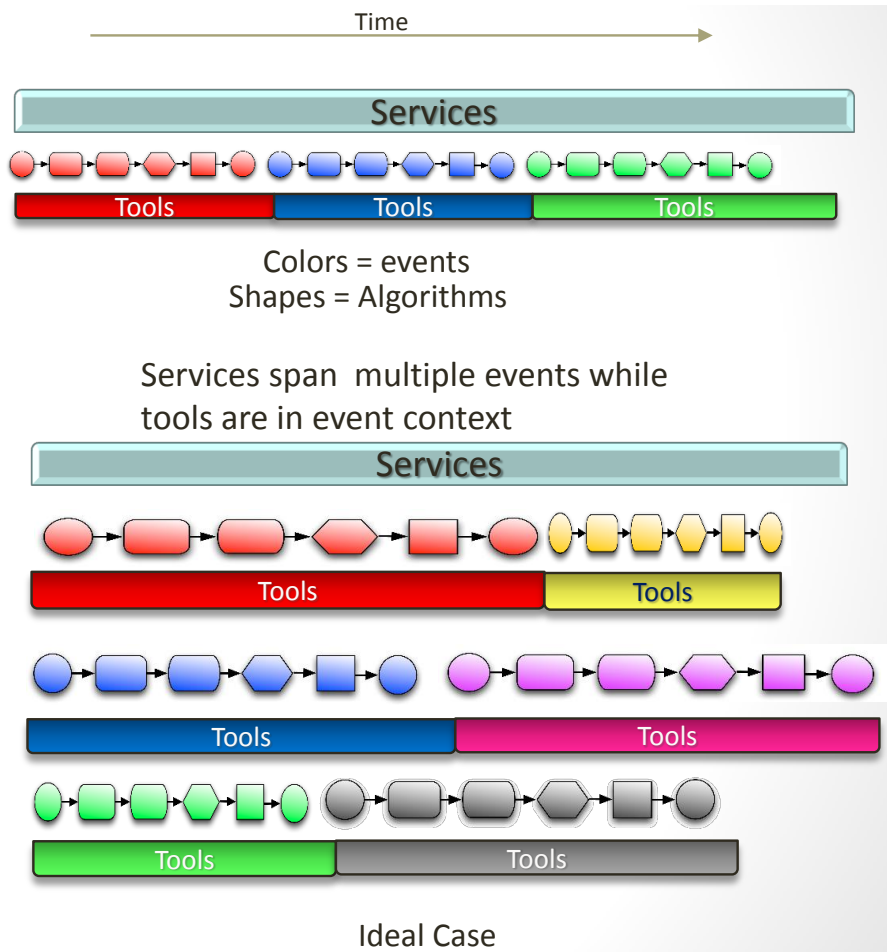
#3 Tue: Frameworks

S. Kama

Future Frameworks Requirements Group started working spring 2014.

- Single thread event processing is well established
- Running multiple instances of single threaded process is wasting memory
- Running event processing in multiple threads require separation of *global* components like multi-event services and *local* components such as per-event or sub-event algorithms and tools

Multiple Threads



Summary:

- Multi-threaded multi-event processing is required.
- Same framework should handle both offline and trigger use cases.
- Scheduler needs to be smarter and more complex than serial case.
- Whiteboard should handle multiple events on flight transparently to algorithms.

Development of a Next Generation Concurrent Framework for the ATLAS Experiment

#3 Tue: Frameworks

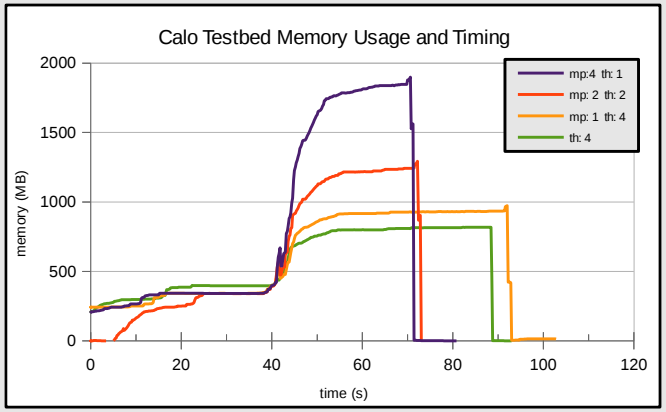
C. Leggett

Expect evolutionary rather that revolutionary code changes.

Hybrid MP/MT

- Hybrid multiprocess / multi-threading framework implementation
  - events are distributed to worker processes *via* AthenaMP shared event queue
  - multi-threading within each worker
    - only one concurrent event, but multiple parallel algorithms in different threads
      - reduce thread safety issues
      - initial modeling on toy simulation shows ~8 parallel data paths for full reconstruction
    - full multi-threading, with multiple concurrent events and algorithms
- Maximizes processor utilization while reducing memory footprint compared with pure event level concurrency
  - *eg* 4 concurrent events:

	time /s	memory /MB
mp: 4 proc, 1 conc. event	32.8	1847
mp: 2 proc, 2 conc. events	34.4	1241
mp: 1 proc, 4 conc. events	53.6	935
hive: 4 concurrent events	47.8	817



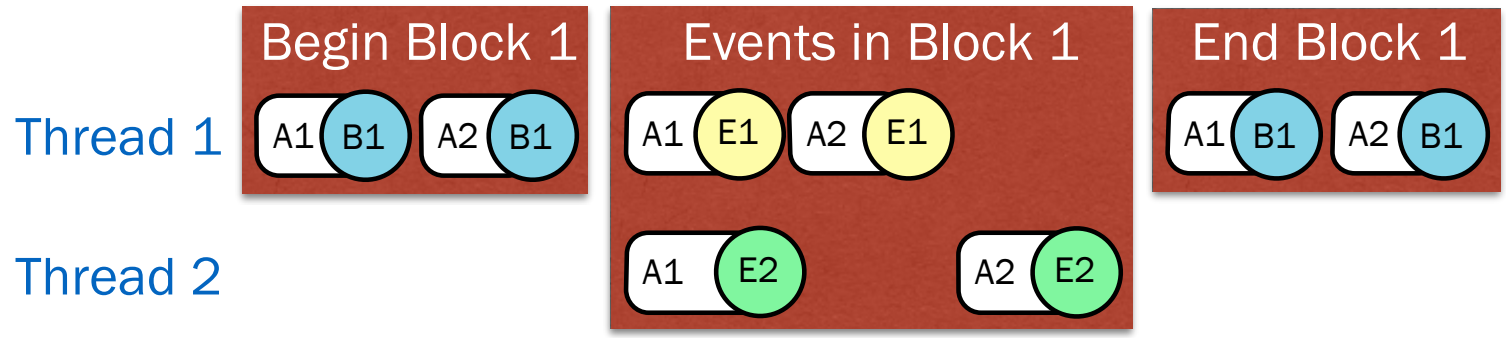
Summary:

- Vast majority of the changes are backward compatible, and will likely make current serial code more efficient.

Using the CMS Threaded Application in a Production Environment

#3 Tue: Frameworks

C. Jones



Evolved from original single-threaded application

- Updated algorithms can run concurrently (A1)
- Only one legacy algorithm is allowed to run at a time (A2)

Processes events in *block* concurrently

- Synchronizes on *block* boundaries (B1)

Goal: reduce memory needed per CPU core

Summary:

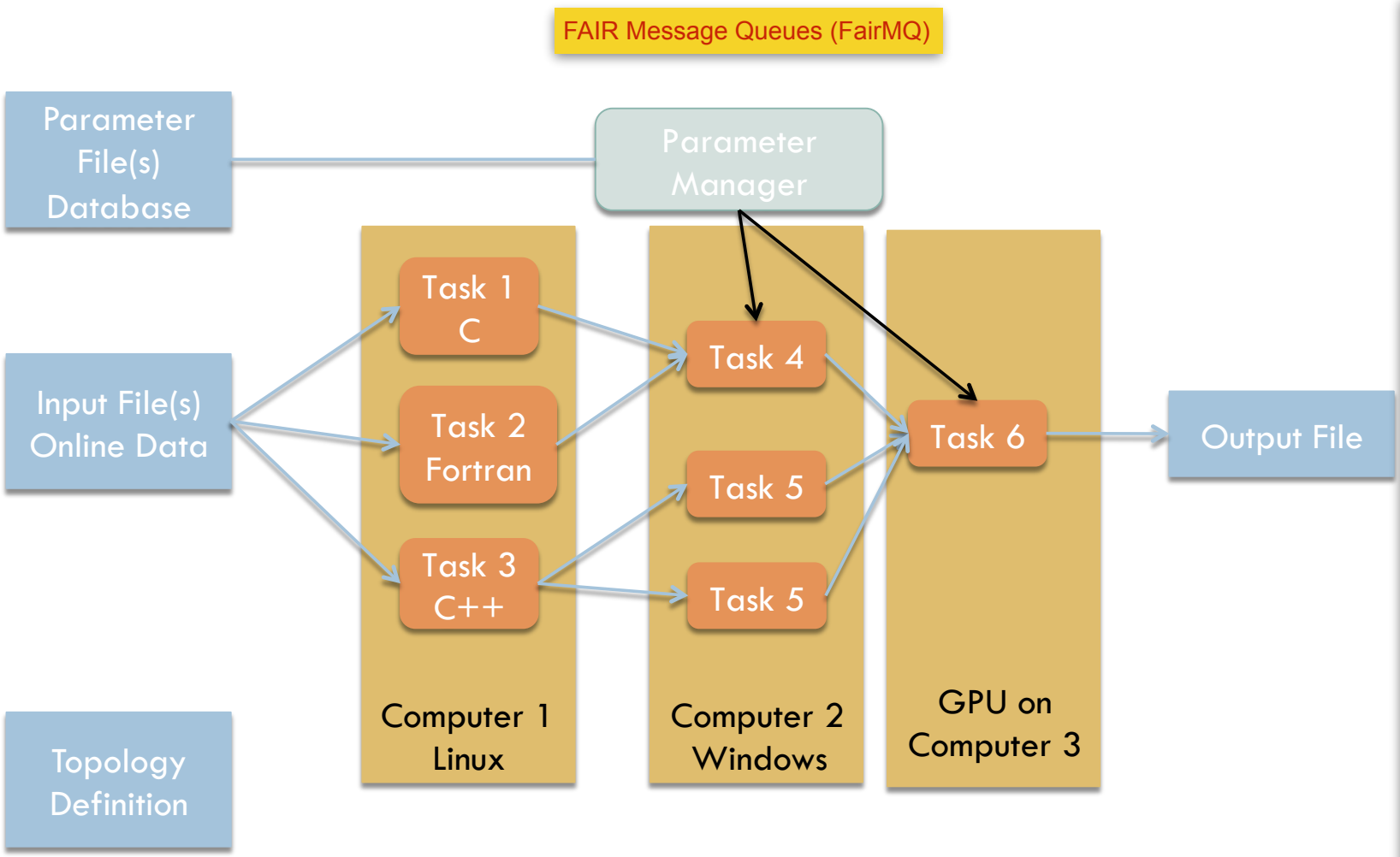
- CMS has successfully transitioned to using multiple threads.
- Job success rate is sufficient for Run 2 data processing.

New Developments in FAIRROOT

#3 Tue: Frameworks

F. Uhlig

FAIRROOT used by ~10 experiments, mostly at GSI, some at CERN.



Summary:

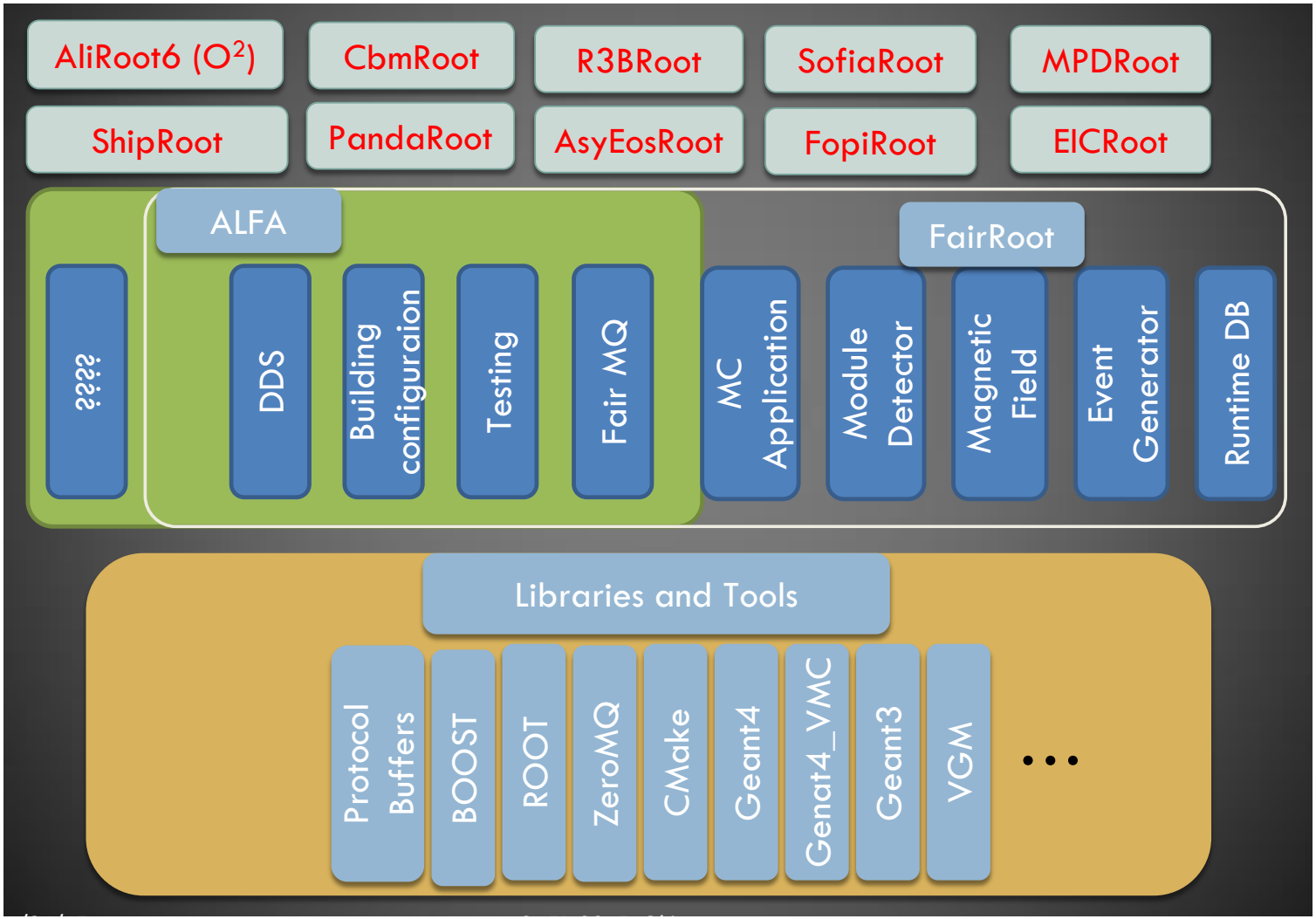
- Each task is a single independent process.
- Each Task sends its results to the next task in the row.
- Allow to setup exactly the needed chain

ALFA: The New ALICE-FAIR Software Framework

#3 Tue: Frameworks

M. Al-Turany

Strategy: massive data volume reduction; coupling between online and offline.



Summary:

- ALFA is a tools set library (transport layer, configuration tools, management and monitoring tools);
- a data-flow based model (Message Queues based multi-processing);
- provides unified access to configuration parameters and databases.

The framework utilizes the Belle II software framework Basf2.

Steering example: reconstruction of $D^{*+} \to D^0\pi^+, D^0 \to K^-\pi^+$

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-

from basf2 import *
from modularAnalysis import *
from stdLooseFSParticles import *

inputMdst('DstarSignalMC.mdst.root') # define mdst input file

stdVeryLoosePi() # make lists of very loosely selected pions (pi+:all, pi-:all)
stdLoosePi()     # make lists of loosely selected pions (pi+:loose, pi-:loose)
stdLooseK()      # make lists of loosely selected kaons (K+:loose, K-:loose)

# reconstruct D0 -> K- pi+ + cc
reconstructDecay('D0 -> K-:loose pi+:loose', '1.7 < M < 2.0 and p_CMS > 2.2')
vertexKFit('D0', 0.001)
applyCuts('D0', '1.81 < M < 1.91')

# reconstruct D*+ -> D0 pi+ + cc
reconstructDecay('D*+ -> D0 pi+:all', 'Q < 0.05')
vertexKFit('D*+', 0.001)
applyCuts('D*+', 'Q < 0.02 and p_CMS > 2.5')

outputUdst('recDstar.udst.root', ['D*+']) # write selected events to micro dst

process(analysis_main) # process events
```

Summary:

- Although the framework is still under development, a user can already perform most of the physics analysis steps like decay reconstructions, vertex fits, tag the flavor of a B meson and perform TMVA-based continuum suppression.

Geant4
Version 10 Series

#4 Tue: Simulation

M. Asai

Geant4 is a general purpose Monte Carlo simulation tool for elementary particles passing through and interacting with matter.

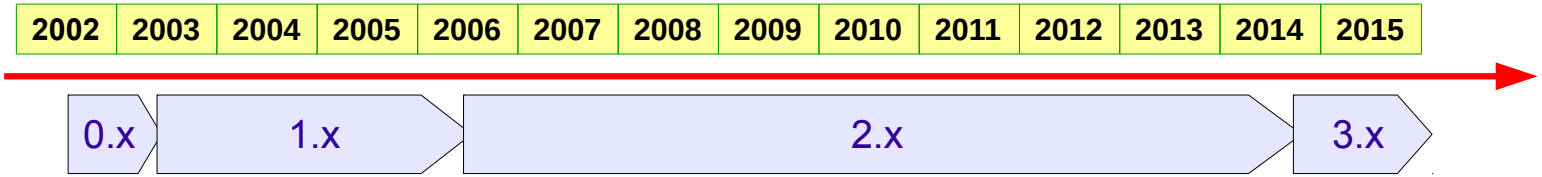
- Major release Geant4 version 10.0 was released on December 6th, 2013.
  - The first major release since June 2007.
- There are several highlighted features including
  - Multithreading capability with event parallelism
  - Isomer production
  - Enhancements in biasing options
  - Introduction of phonon transport with a new concept of crystal
- Version 10.1 released on December 5th, 2014 has lots of improvements in both physics and computing aspects.
  - Current version : 10.1-patch01 released on April 1st, 2015
- Release of version 10.2 is scheduled on December 4th, 2015

Summary:
- Given Geant4 is nowadays mission-critical for many users including all HEP experiments, space missions, medical applications, etc., Geant4 is to be kept maintained and still evolving for at least next decade.

Geant4 VMC 3.0

I. Hřivnáčová

The Virtual Monte Carlo interface (VMC) allows to run simulation with different Monte Carlo codes from the same user application.



Oct 2002	Jul 2003	Dec 2006	Nov 2014
v0.1	**v1.0**	**v2.0**	**v3.0**
First version	Geometry convertors (roottog4, …)	G4Root Navigation	Multithreading CMake Root 6
.. v0.5	.. v1.9	.. 2.15	.. 3.1 ..
Root 3.03/09 Geant4 4.1	Root 3.05/06 Geant4 5.2	Root 5.14/00 Geant4 8.2	Root 5.34/23 and 6.02/01 Geant4 10.00.p03

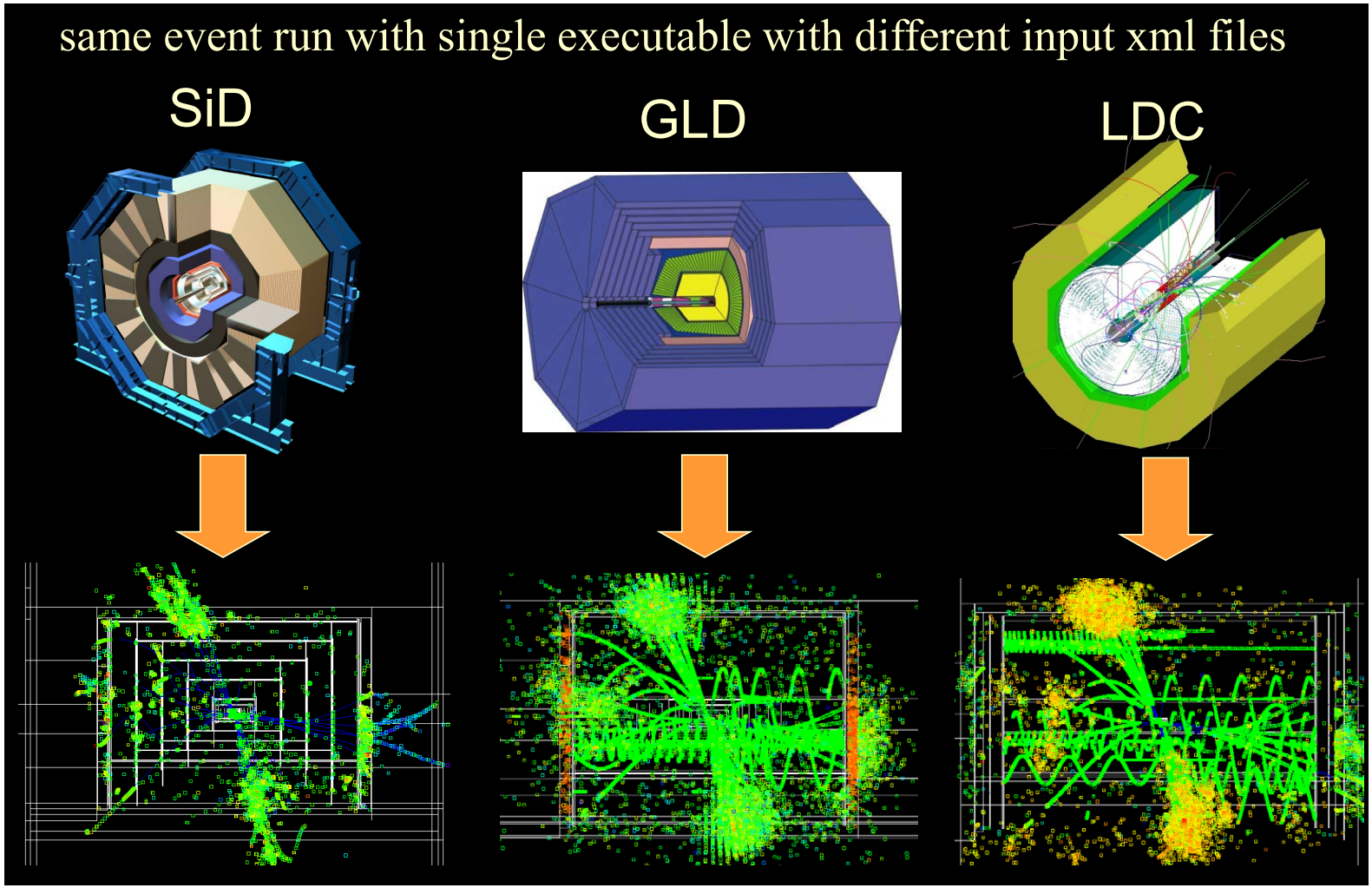
Summary:
- Geant4 VMC 3.0, multithreading capable, is available since November 2014.
- The interest in Geant4 multithreading in both ALICE and FAIR.
- Several improvements in build system and testing.

slic: A Geant4-based Detector Response Simulation Program

#4 Tue: Simulation

N. Graf

The goal of *slic* is to develop and distribute an executable program, which provides the full functionality and flexibility of the Geant4 toolkit, but insulates the end user from having to write any C++ code.



Summary:

- *slic* provides a complete and flexible detector simulation package capable of simulating arbitrary complex detectors with runtime detector description.
- Being used by LC (ILC & CLIC) detector community and by HPS experiment.

The GeantV Project: Preparing the Future of Simulation

F. Carminati

- We developed three main components
  - 1. A multithread scheduler to handle the particle baskets
  - 2. A vectorised geometry library and navigator
  - 3. A vectorised Compton scattering
- Our results indicate that
  - Basket handling introduces a minimal overhead
  - SIMD gains half an order of magnitude in performance
- We believe we are on track with our objective
- An optimistic success-oriented interpretation of our results could be 1.4 (scheduler) x 4 (physics & geometry) ~ 6
- Of course the proof of the cake will be in the eating

Summary:
- Work towards a credible demonstrator for a high-performance particle transport code is progressing.
- The techniques developed have the potential to benefit to many other codes within and outside of HEP.

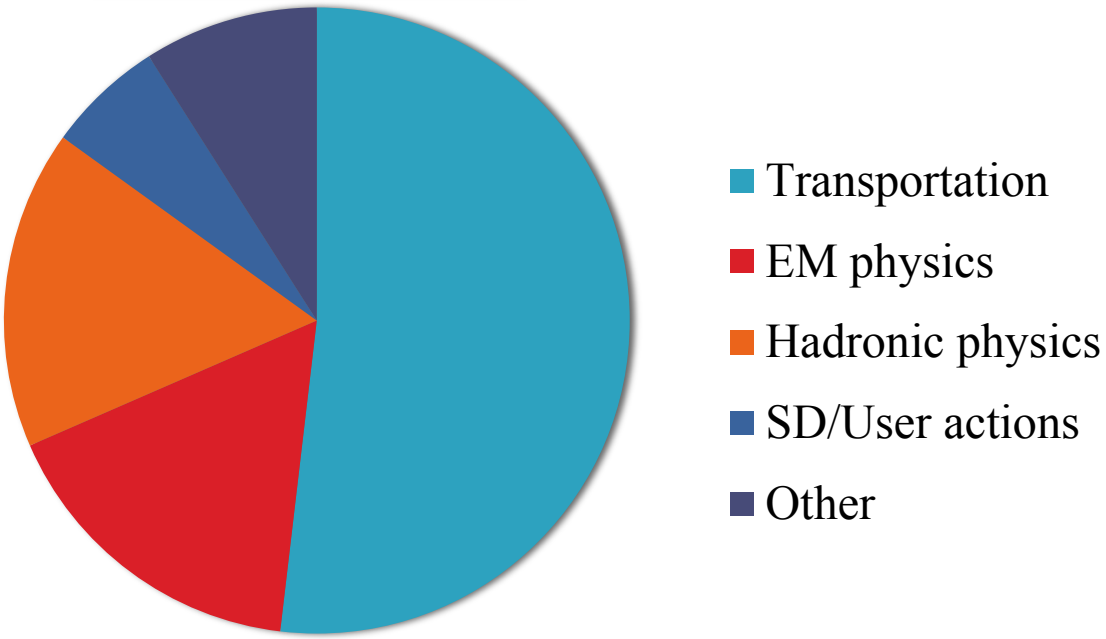
CMS Full Simulation for Run-2

#4 Tue: Simulation

D. Lange

Run 2: increasing of energy from 8 to 13 TeV provides more simulation steps and more hits in detectors.

Where our simulation CPU goes



- Technical performance improvements for Run 2 simulation:
  1. Upgrade to Geant4 10.0p03 (~5%)
  2. Implementation of Russian Roulette technique (~30%)
  3. CMSSW code optimization (~15%)
  4. Library repackaging (~10%)

Summary:

- Large gains achieved since Run 1. Now investigating other areas including interesting new components of Geant4.
- Deployment of multithreaded simulation planned fro 2016.

CMS Detector Description for Run II and Beyond

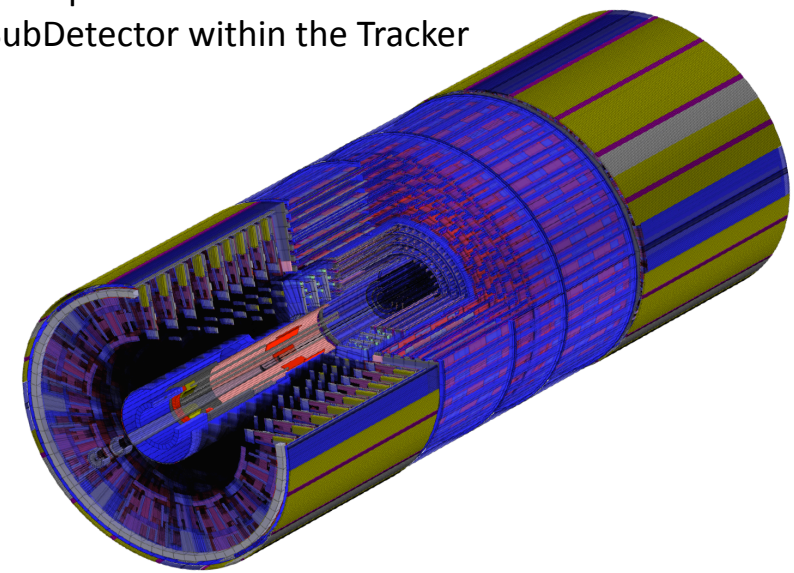
#4 Tue: Simulation

G. Boudoul

The DD is a common source of information for Simulation, Reconstruction, Analysis, and Visualisation.

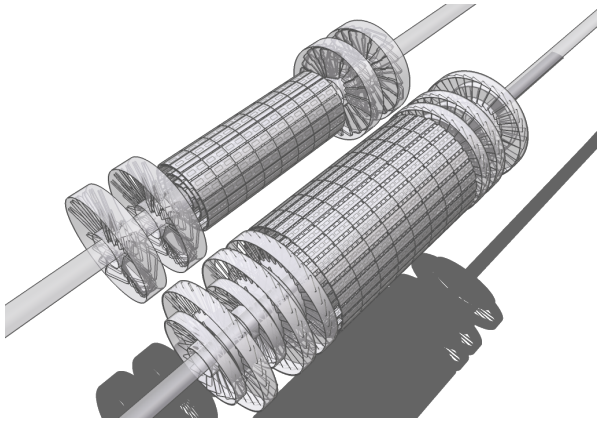
Modular CMS Detector Description

Example of the Pixel SubDetector within the Tracker



Present Pixel or the future Pixel description can be exchanged within the Tracker

Present Pixel Geometry (3 layers- 2disks)



Future Pixel Geometry (4 layers- 3disks)

12

Summary:

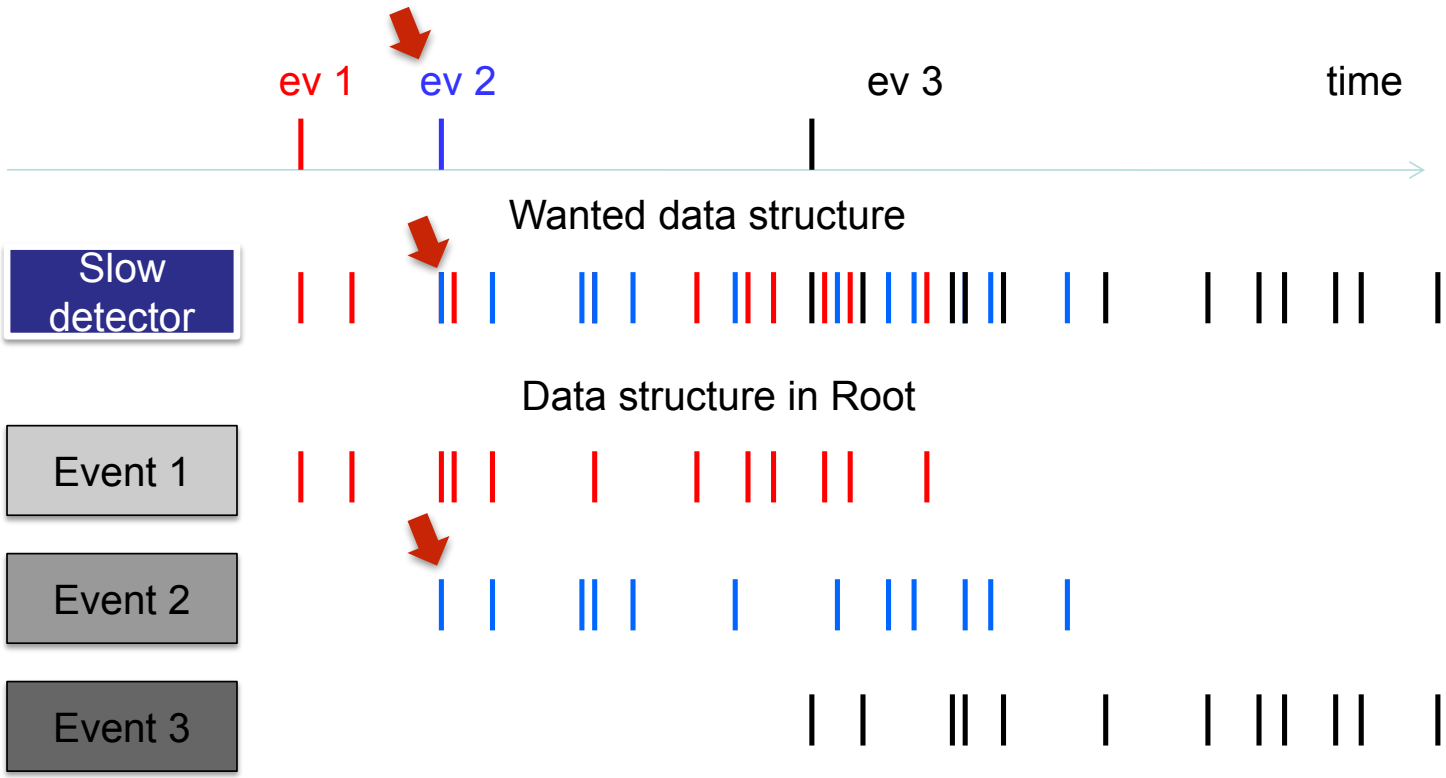
- Many developments occurred during the long Shutdown.
- Both Run 1 and Run 2 Detector Descriptions coexist in the same software release.

Continuous Readout Simulation with FairRoot for the PANDA Experiment

#4 Tue: Simulation

T. Stockmanns

Signal and background-events very similar → no hardware trigger possible.



Summary:

- Time-based simulation part of FairRoot-Framework.
- Reconstruction based on time slices.

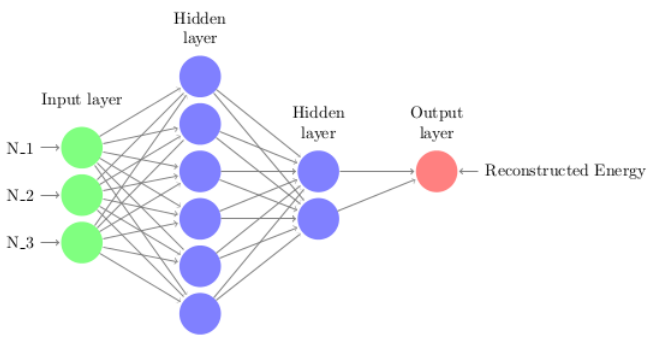
Energy Reconstruction Study in a Semi-Digital Hadronic Calorimeter for ILC

S. Mannai

Motivation: ILC should be equipped with high precision detectors.

Energy reconstruction: **A**rtificial **N**eural **N**etworks

- TMultiLayerPerceptron of root package.
- 2 hidden layers with 6 and 2 neurons.
- The input variables: N_1,N_2,N_3 .
- The output variable is the reconstructed energy: E_{rec} .
- Monte Carlo Simulation
  - Training Samples: Odd energies, 1-99 GeV (50 training samples)
  - Test Samples: Even energies, 10-90 GeV (40 test samples)



Summary:
- Analytic energy reconstruction method: energy resolution reaches < 6% at 80 *GeV* with satisfactory linearity.
- ANN technique also giving promised results.

The ATLAS Higgs Machine Learning Challenge

#5 Thu: Analysis

G. Cowan

Optimal analysis uses information from all (or in any case many) of the measured quantities → Multivariate Analysis (MVA).
But much recent work in Machine Learning only slowly percolating into HEP (deep neural networks, random forests, ...)

Typical Machine Learning goal is event classification; try to minimize e.g. classification error rate.

Goal in HEP search is to establish whether event sample contains only background; rejecting this hypothesis ≈ discovery of signal.

Often approach in HEP is to use distribution of MVA classifier.

- Simplest case, use classifier to define "search region" and count:
  - s = expected number of signal events (assuming it exists)
  - b = expected number of background events
- Goal: Minimize Approximate Median Significance of discovery:

$$\mathrm{AMS} = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}$$

(Modified in the Challenge to prevent small search region where estimate of b may fluctuate very low: $b \to b + b_{\mathrm{reg}}$.)

Summary:
- Re-importing into HEP all the ML developments; e.g., discussions on-going with TMVA experts.

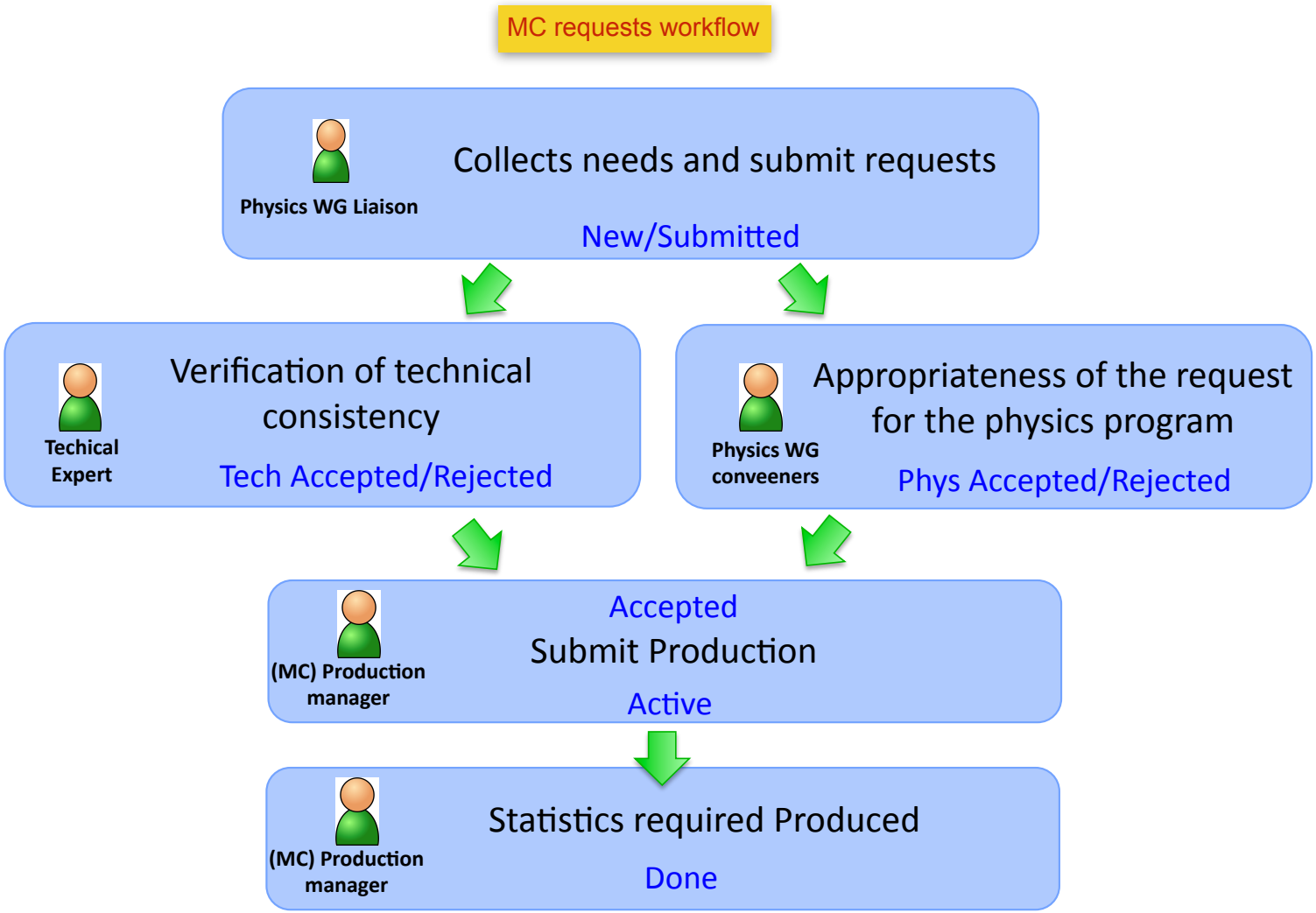
How the Monte Carlo Production of a Wide Variety of Different Samples is Centrally Handled in the LHCb Experiment

#5 Thu: Analysis

G. Corti

In LHCb a wide variety of Monte Carlo samples need to be produced for the experiment's physics program.

Procedures based on common infrastructures have been setup to handle Monte Carlo productions centrally.



Summary:

- Allows for massive transparent and efficient production on a world-wide distributed system with very little manpower.

The Library Event Matching Classifier for ν_e Events in NOvA

#5 Thu: Analysis

D. Rocco

Concept: compare an unknown trial event to a library of known events.

- Library is 77M simulated neutrino interactions
  - 18M ν_e charged-current signal events
  - 29M ν_μ charged-current events and neutral current events
  - 30M π^0-enriched neutral-current events ($\pi^0 \rightarrow \gamma\gamma$ is primary background)
- Library events are from NOνA Monte Carlo
  - Flux from FLUKA/FLUGG simulation of beamline
  - ν interactions simulated by GENIE
  - Particle propagation in detector from GEANT4
  - NOνA simulation code converts energy depositions to electronic signals
- Performance scales roughly logarithmically with library size
- Flipping events transversely effectively quadruples library size
  - Events can be flipped in either view (x,y) or both
  - Symmetry is not perfect, but close enough
  - Beam direction is slightly upward (~ 3°)
  - Attenuation in fiber varies with distance from readout

Summary:

- Library Event Matching performs well for ν_e event selection in NOvA
- Classifier uses full hit information rather than reconstructed quantities

Moment Morphing: Interpolation between Multi-dimensional Histograms

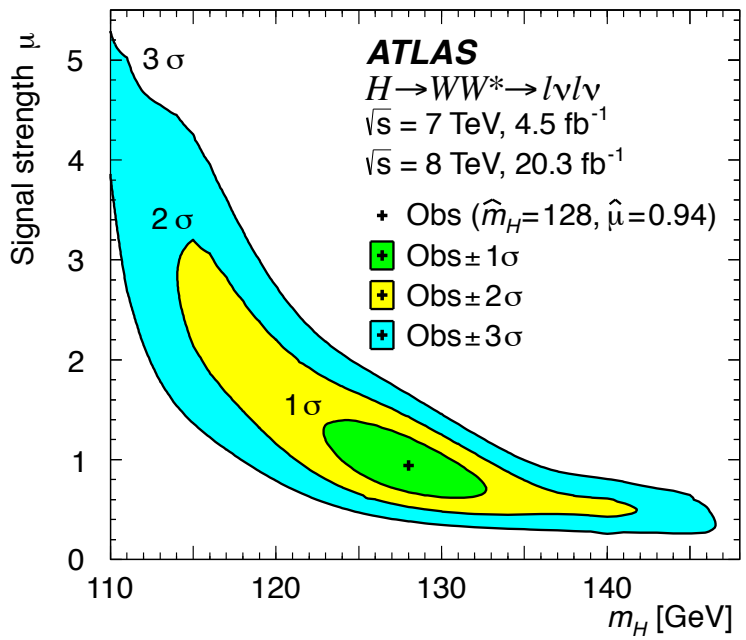
S. Gadatsch

Allows to interpolate between available samples

Real world examples

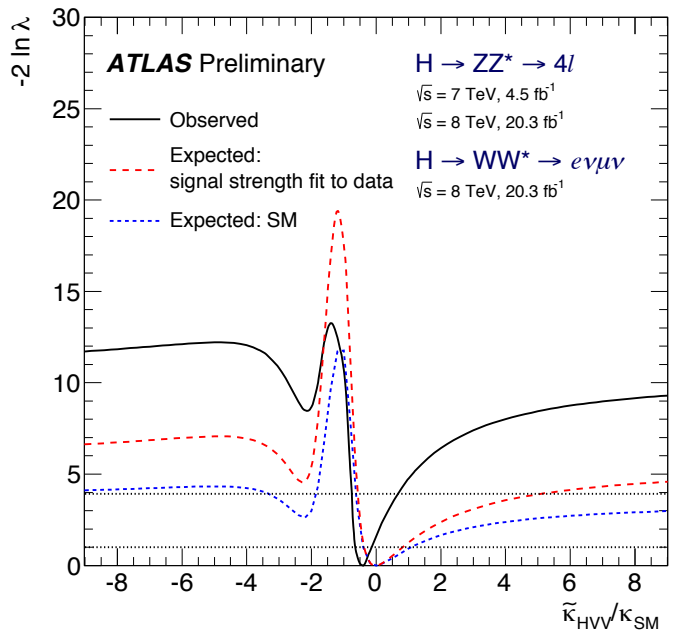
Higgs boson rate in $H \to WW^*$

- Simulated samples available in 5 GeV in m_H



Higgs boson *CP* measurement

- Combination of 1d ($H \to WW^*$) and 2d ($H \to ZZ^*$) moment morphing



Summary:

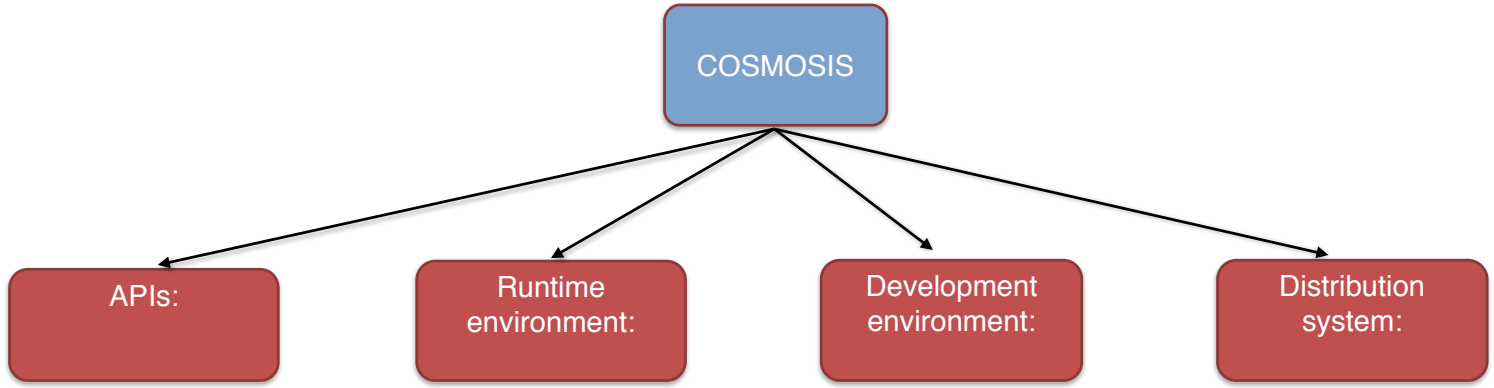
- New algorithm to interpolate between Monte Carlo sample distributions: generic, simple, fast, stable and accurate.

CosmoSIS: Modular Cosmological Parameter Estimation

A. Manzotti

CosmoSIS is a modular framework for parameter estimation

Modular solution



Define module configuration, how to obtain their input data, how they interact with the framework, how outputs are organized, and now new modular components can be added.

To reliably configure and run programs that use these modules.

To write, build, and test analysis software, and which makes it easy to share what they have developed.

Registration of attribution information (required citations, etc.) for use of any contributed code.

To make it easy to install the code and ensure compatibility.

Summary:

- HEP might consider adopting a similar to CosmoSIS attribution concept to help encourage sharing useful software.

Given data from LHC, what are likely values of masses, cross sections. . . ?

BAT components

$$P(\theta|D,M) \propto P(D|\theta,M)P_0(\theta|M)$$

USER DEFINED

- create model
- read data

COMMON TOOLS

- `Normalize()`
- `FindMode()`
- `MarginalizeAll()`
- `PrintAllMarginalized()`
- `PrintKnowledgeUpdatePlots()`

DEFINE MyModel : BCModel

- `AddParameter("mu", 0, 1)`
- `LogLikelihood()`
- `LogAPrioriProbability()`

Summary:

- BAT well established
- More powerful sampling algorithms in BAT 2.0

Using R in ROOT with the ROOT-R Package

#6 Thu: Tools

L. Moneta

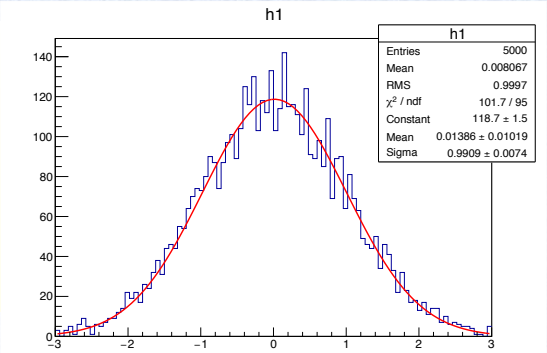
- R is an open source language and environment for statistical computing and graphics.
- ROOT-R package gives access to ROOT users to the R capabilities and its rich functionality.

Example of RMinimizer

- ROOT plugin for Minimisation implemented using R
  - *developed by Kirby Hermann (GSOC student 2014)*
  - give access to R optimisation tools when fitting or multi-dimensional function minimisation
  - based on R optim and optimx packages

```
ROOT::Math::MinimizerOptions::SetDefaultMinimizer("RMinimizer","L-BFGS-B");
hist->Fit("gaus");
```

```
root [4] h1.Fit("gaus")
Value at minimum =101.673
****************************************
Minimizer is RMinimizer / L-BFGS-B
Chi2                      =      101.673
NDf                       =           95
NCalls                    =          265
Constant                  =      118.694   +/-   1.47659
Mean                      =    0.0138555   +/-   0.0101907
Sigma                     =     0.990906   +/-   0.00741443
```



Summary:
- ROOTR provides easy access to R tools in ROOT and C++
- Easy to use directly R from ROOT prompt
- Package is ready do be released in the next ROOT production version (6.0.4)

HistFitter
A Flexible Framework for Statistical Data Analysis

G.-J. Besjes

- Software framework for statistical data analysis built on top of the HistFactory/RooFit (construction of models) and RooStats (statistical tests) software packages.
- Used in (almost) all searches for supersymmetry in the ATLAS collaboration.

Interpreting results: testing models

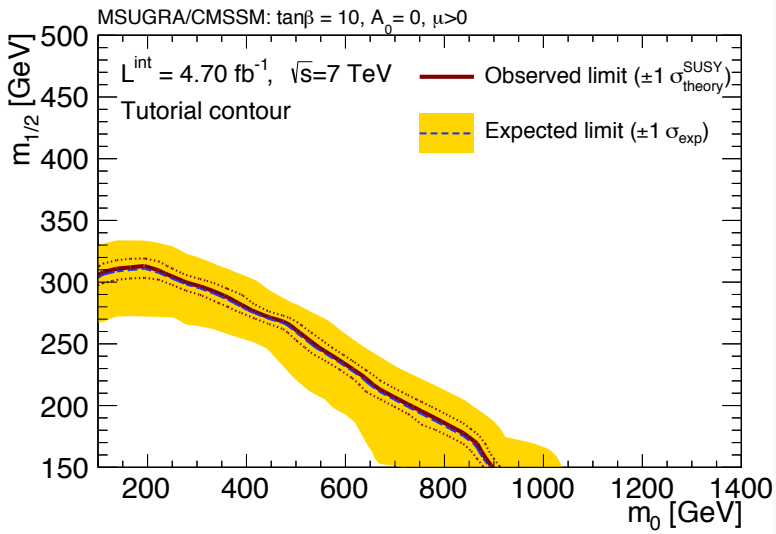
Several interpretations/hypothesis tests available (RooStats-based) and macros to interpret and present the results.

- **Hypothesis tests**: test a specific alternative (signal) model

  HistFitter provides tools to run over sets of models and graphically present the results.

- **Upper limit calculations**: performs repeated hypothesis tests

  Determines 95% CL cross-section upper limit for specific signal models



Note: dummy figure!

Summary:

- ***Easy configuration***: single user-defined configuration file for an entire analysis
- ***Essential concepts of regions built in***: control, validation and signal regions deeply woven into the design and extrapolation of results is automatically taken care of
- ***Bookkeeping***: automatic management of multiple configurations, statistical tests, underlying data, etc.
- ***Presentation and interpretation***: large set of easy-to-use tools to present data and interpret results available

Accelerating Analysis with SciDB

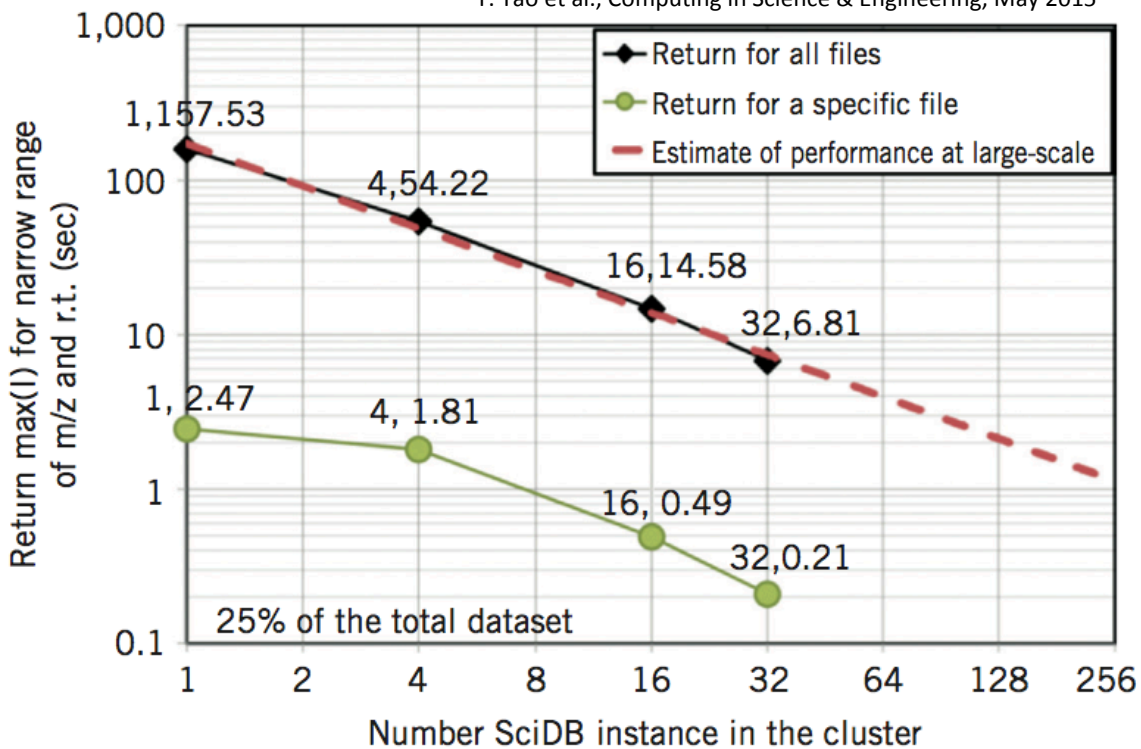
#6 Thu: Tools

L. Gerhardt

SciDB is an Array Data Base management system and analytic platform, scalable out to 1,000s of processors and TBs - PBs of data

SciDB scales well

Y. Yao et al., Computing in Science & Engineering, May 2015



Summary:

- **Lots of optimized functions**: sampling, matrix operations (ScaLAPACK), statistics, user designed functions
- **Parallel architecture** that is transparent to the user

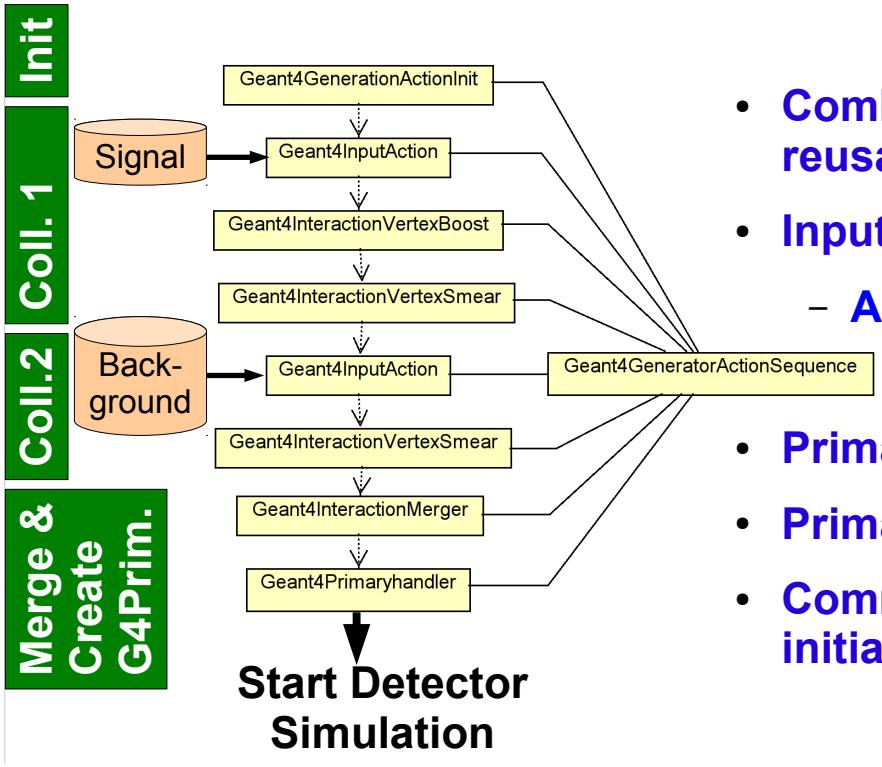
DDG4 - a Simulation Framework based on DD4hep and Geant4

#6 Thu: Tools

M. Frank

- **DD4hep**: a general purpose Detector Description toolkit
- **DDG4**: simulation support for DD4hep, requires minimal user effort

Example of an Action Sequence: Event Overlay with Features



- **Combine simple and reusable modules**
- **Input module**
  - **Any data format**
- **Primary vertex smearing**
- **Primary vertex boost**
- **Common: initialization, final merge**

Summary:

- Toolkit to simulate particle collisions in HEP detectors with minimal effort: simple, easy, flexible
- Can host user plugins: extensible

BESIII Physics Data Storing and Processing on HBase and MapReduce

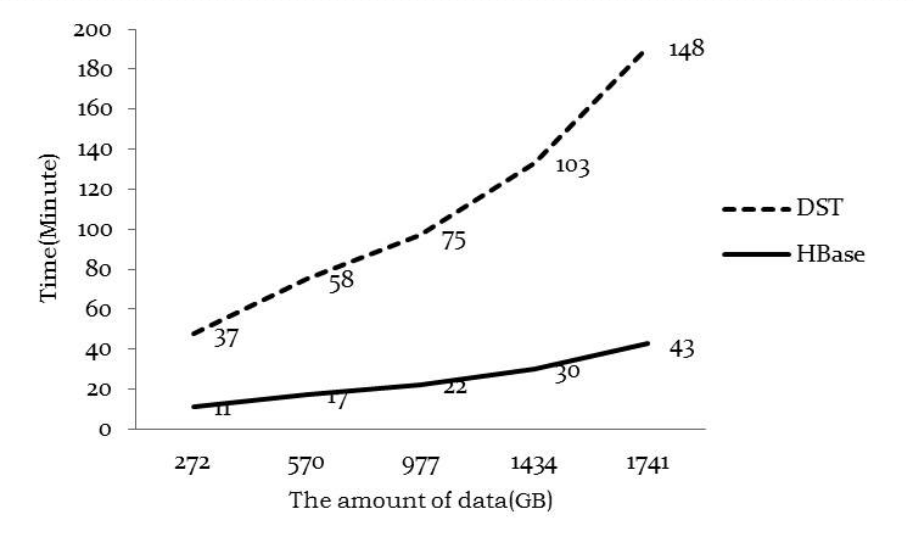
#6 Thu: Tools

X. Lei

- Data need to be analyzed is stored in HBase
- A MapReduce model and a region split algorithm are designed to satisfy HBase's locality principle

Performance

- Data stored in DST files vs. data stored in HBase
  - 3.5 times faster with pre-selection



  - 20% faster without pre-selection

Summary:

- The new data storing and processing platform for BESIII is faster, stable and scalable

Breaking the Silos: The *art* Documentation Suite

#6 Thu: Tools

R. Kutschke

- *art*: event processing framework used as an external product
- In the same sense as ROOT, Geant4, CLHEP ...

Status of the documentation suite

90% Complete

25% Complete

5% Complete

Introduction
1. What is a framework
2. Define Prerequisites
3. Overview of documentation
4. ...

Workbook
- Exercise 1
  - Activity 1
  - Activity 2
- Exercise 2
  - Activity 1
  - Activity 2
- ...
- Exercise N

Reference Manual

Technical Reference

External Refs
Root, C++, STL, G4
...

Users Guide
Table of Contents
1. Users view of *art*
2. Modules
3. Services
4. FHiCL
5. Interface to G4
6. Interface to SAM
7. .....

Appendices
I. Best Practices
II. Trouble Shooting
III. CLHEP gap Filler Docs
IV. Glossary
Index

Fermilab

Summary:
- There is a plan for an integrated *art* documentation suite, which still a mostly volunteer effort, new volunteers welcome.

Summary

We can more:

- more optimisation
- more compactness
- more vectorization
- more cores
- more events/s
- more expect at CHEP2016