Fast Monte Carlo at ATLAS

2nd Fast Monte Carlo Workshop in HEP

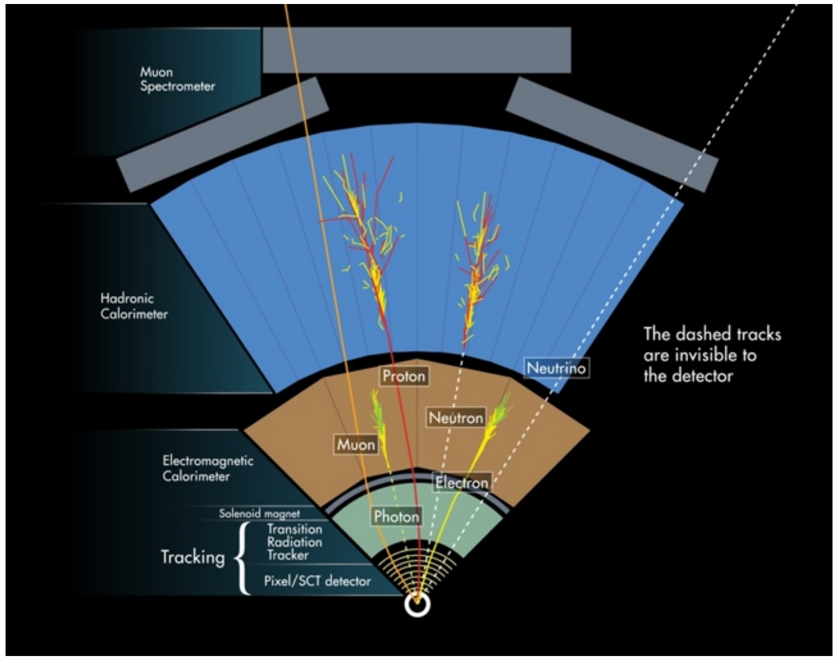
Flavia de Almeida Dias

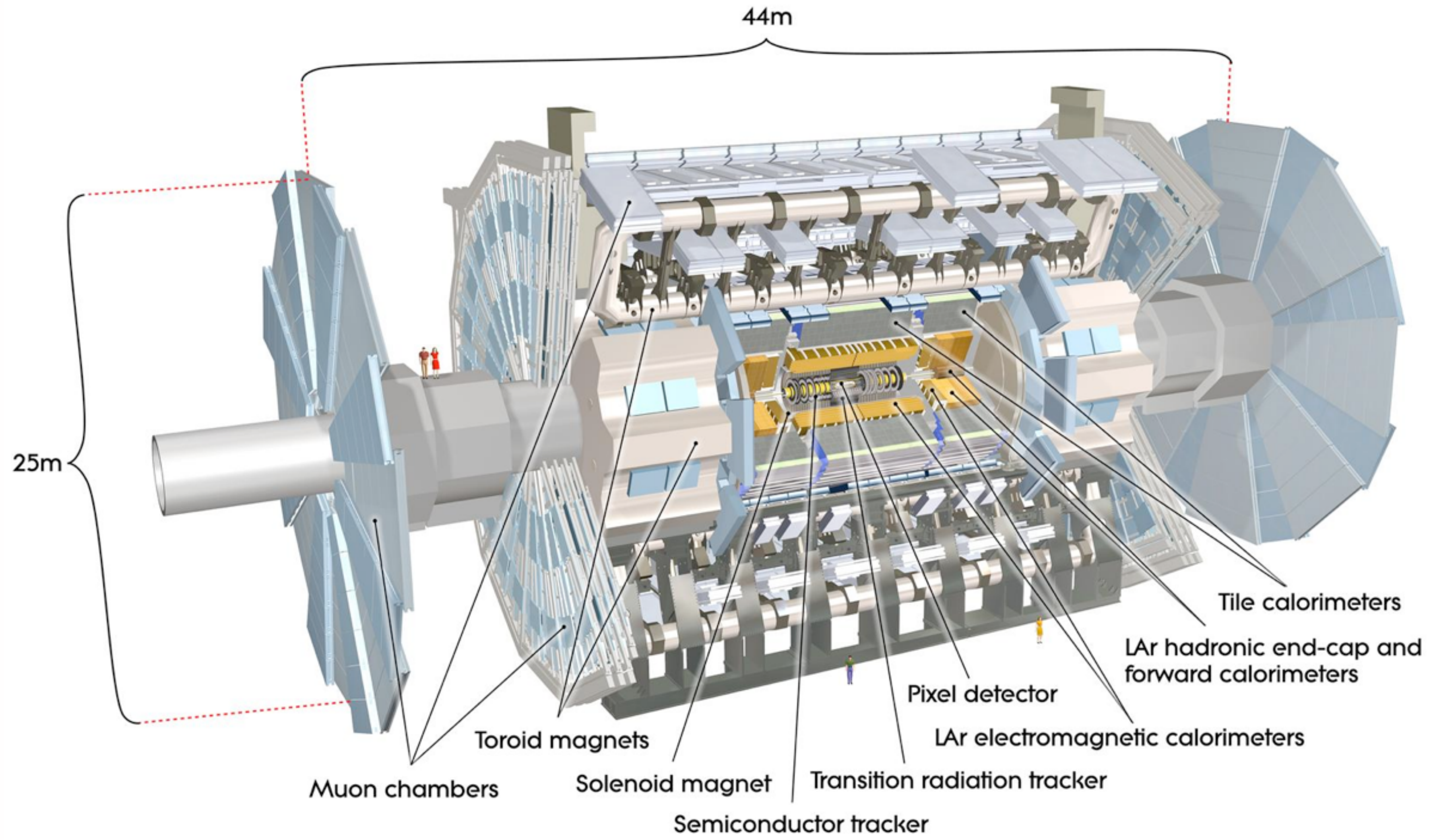
14 - January - 2014

Outline

- ATLAS detector overview
- MC production chain
- Fast simulators on ATLAS
  - ➡ Frozen Showers
  - ➡ AF2 / FastCaloSim
  - ➡ AF2F / Fatras
  - ➡ Parameterization
- Integrated Simulation Framework
- Fast digitization and fast reconstruction
- Final Product
- Summary and outlook

ATLAS Detector Overview

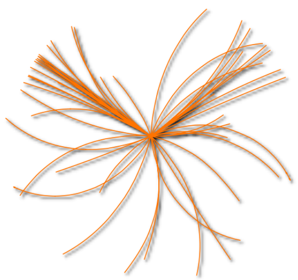


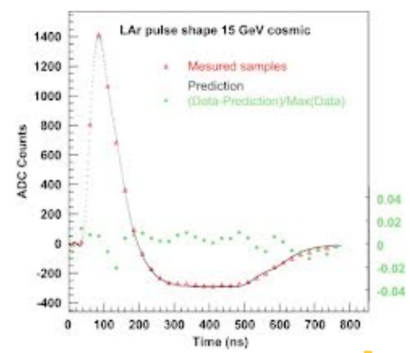


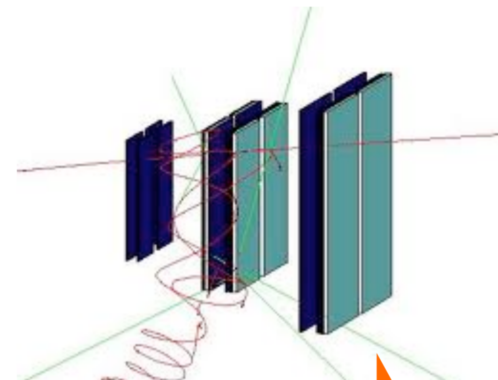
- Main subdetectors
  - ➡ **Inner Detector** => Silicon and transition radiation technologies, in solenoidal magnetic field
  - ➡ **Calorimeters** => LAr EM calorimeter (in central and forward regions) and hadronic calorimeter (tile in central and LAr in forward region)
  - ➡ **Muon system** => spectrometer in toroidal magnetic field

Monte Carlo Production Chain

From 4-vectors to ROOT

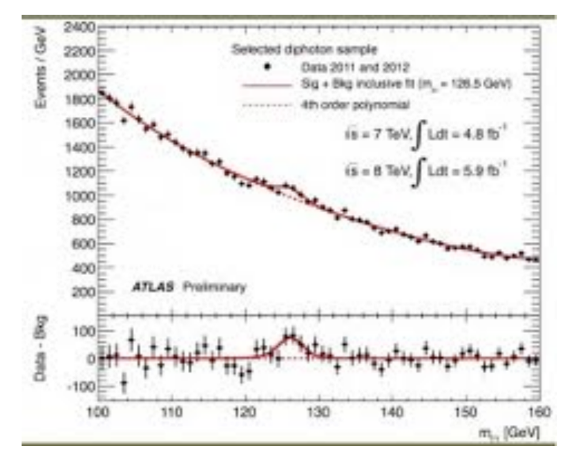


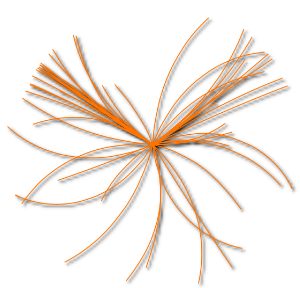





Event Generation → **Detector Simulation** → **Digitization** → **Reconstruction** → **Rootification**

Format: EVGEN | HITS | RDO | ESD/AOD | NTUPLE/HIST



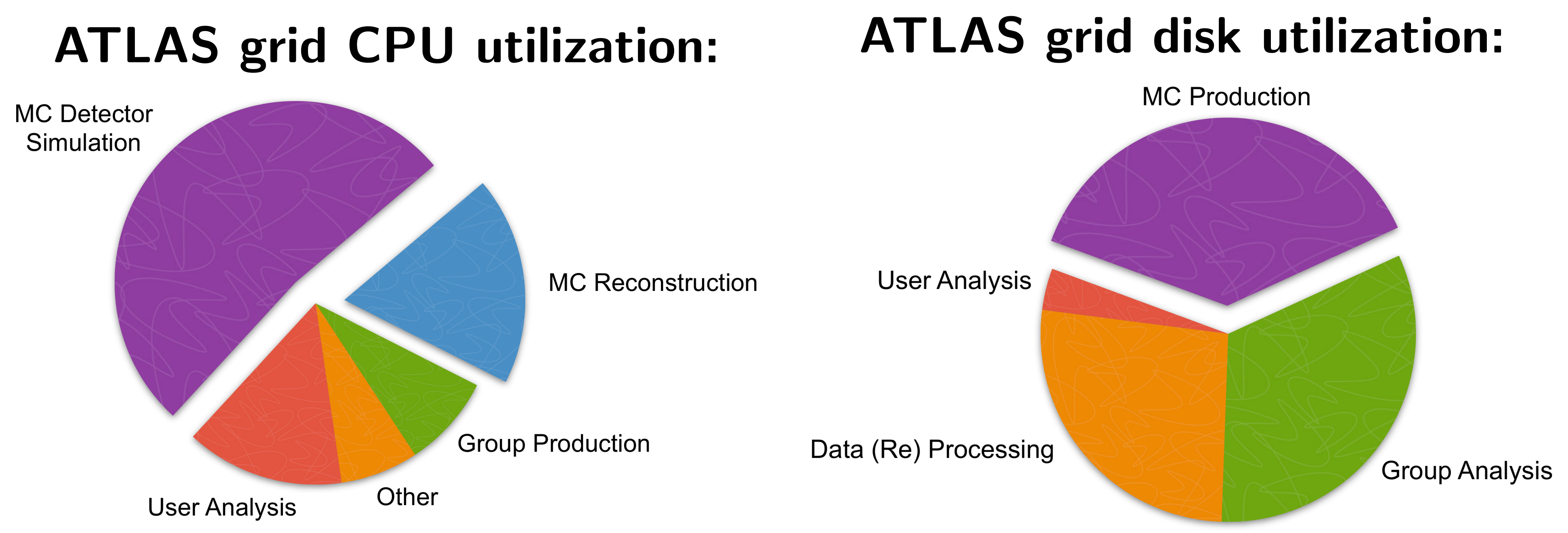



overlay up to 40 individual detector simulation events to emulate pileup

MC Production on the Grid

- Grid CPU usage dominated by MC production
- MC production takes up large fraction of Grid disk usage => limitation
- Precise detector simulation => highly CPU intensive
- Obstacle for physics analyses in need of large MC statistics => sensitivity limitation
- Higher luminosity and pileup => larger MC production needed

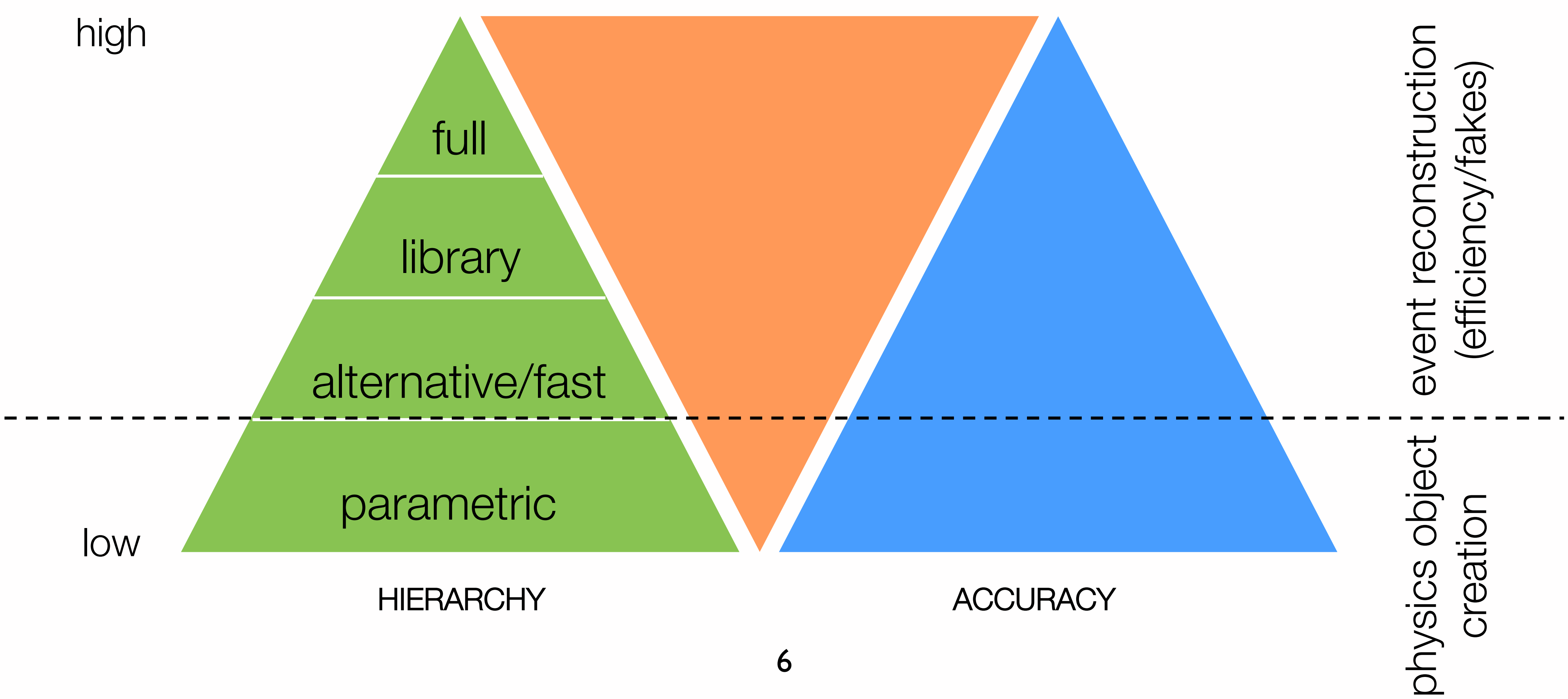
ATLAS Grid usage in 2012



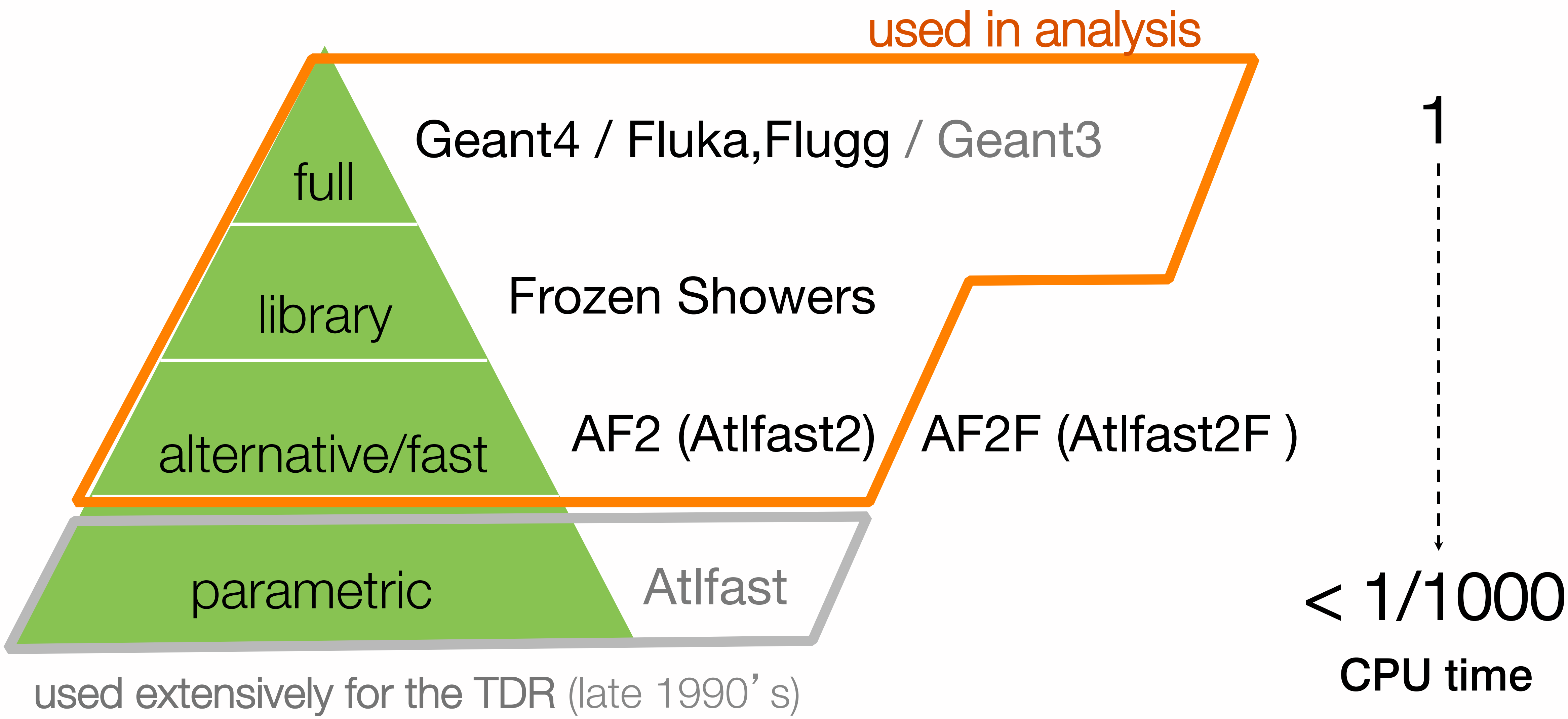
Faster chain is necessary for RunII

Simulation Hierarchy Pyramid

- Simulate interactions of particles with sensitive and non sensitive detector material
- Produce sensitive detector hits with position and deposited energy information => input to digitization
- More accurate simulation means slower simulation
  - ➡ tradeoff between accuracy and speed



Simulation History and Potential Speed-Ups

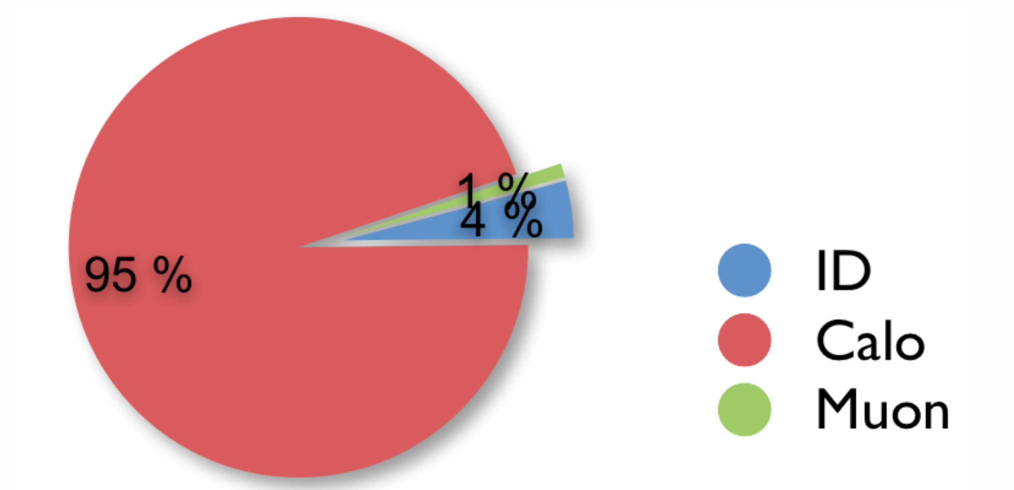


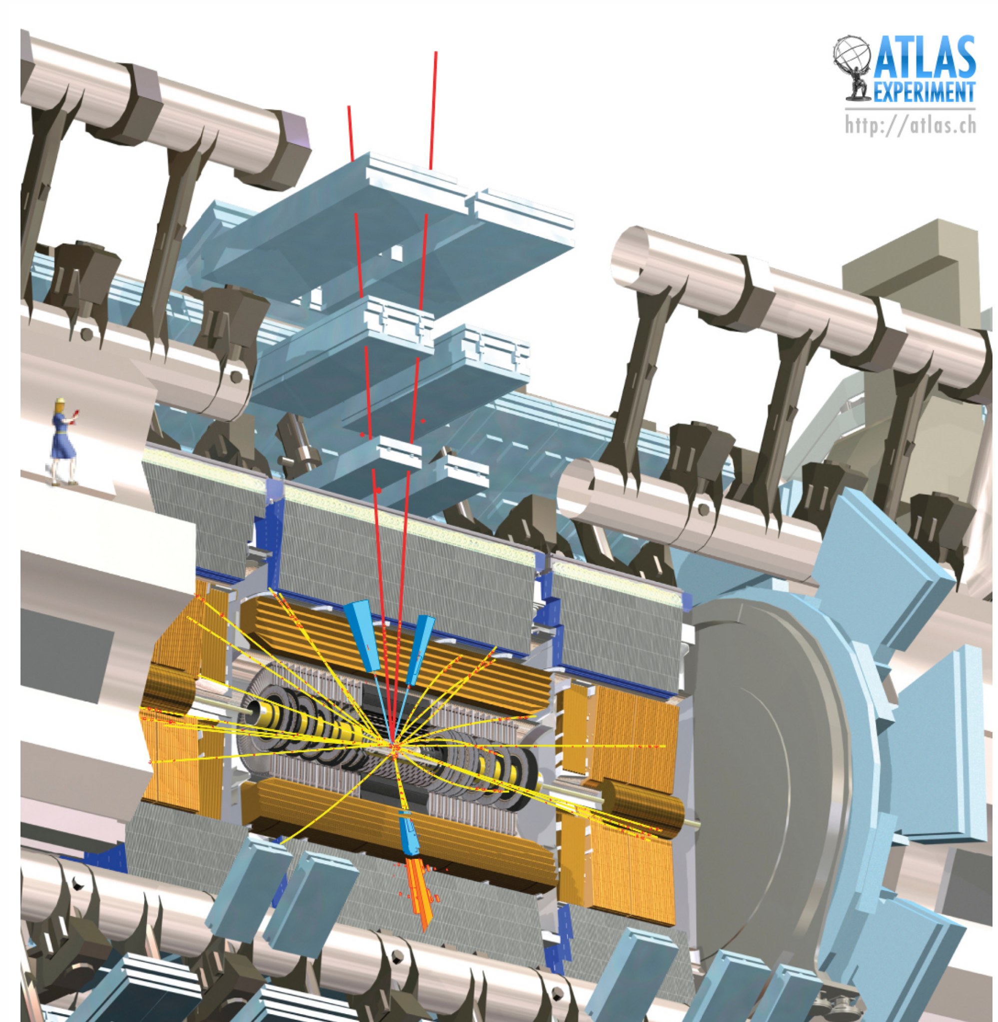
- Unfortunately these all have “grown” independently
  - ➡ different configuration, steering
  - ➡ different output format
- Fast simulation sets the simulation into the ~ Hz level regime
- Has many more consequences (see later)

FullSim - Geant4

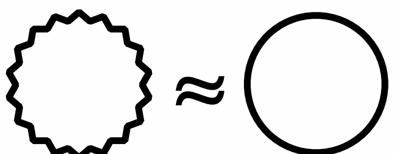
- Stable, fully validated and precise simulation
- High CPU consumption
  - ➡ mostly in EM calorimeters
  - ➡ simulation of ~30M volumes
- Also Geant4 can be/should be speed-optimised
  - ➡ Runge-Kutta-Nystroem propagator into Geant4
    - significantly faster
    - higher accuracy in long extrapolation tests
- Complete rework of Magnetic field access in ATLAS

G4 simulation time per subdetector

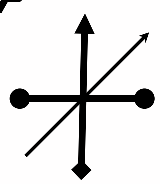




Ideas to Speed Up Simulation



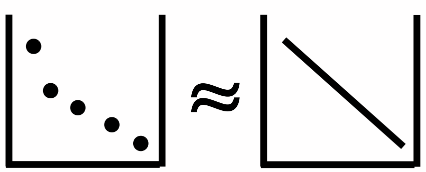
approximate geometry



optimise transport and navigation

$\pi \approx 3$

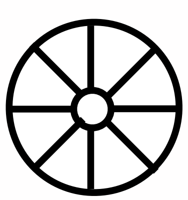
approximate models



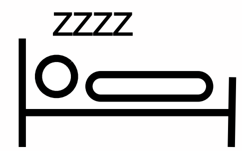
parameterisations

ctrl C/V

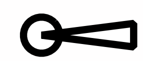
take shortcuts



use new technologies



don't do anything

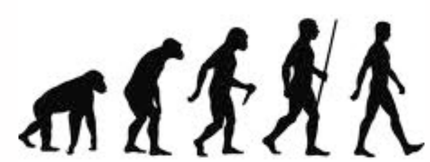
on  off

work only on demand

1 €	2 DM
2 €	4 DM

use look-up tables

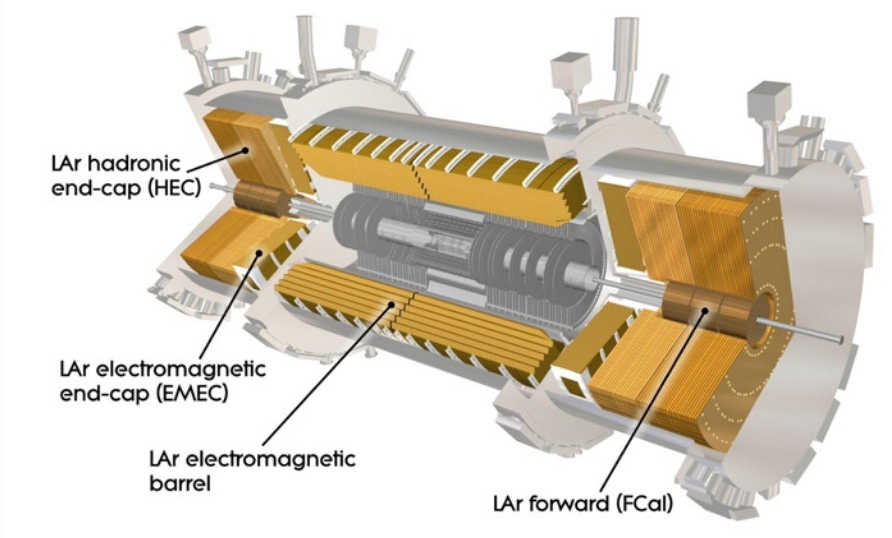
throw away things

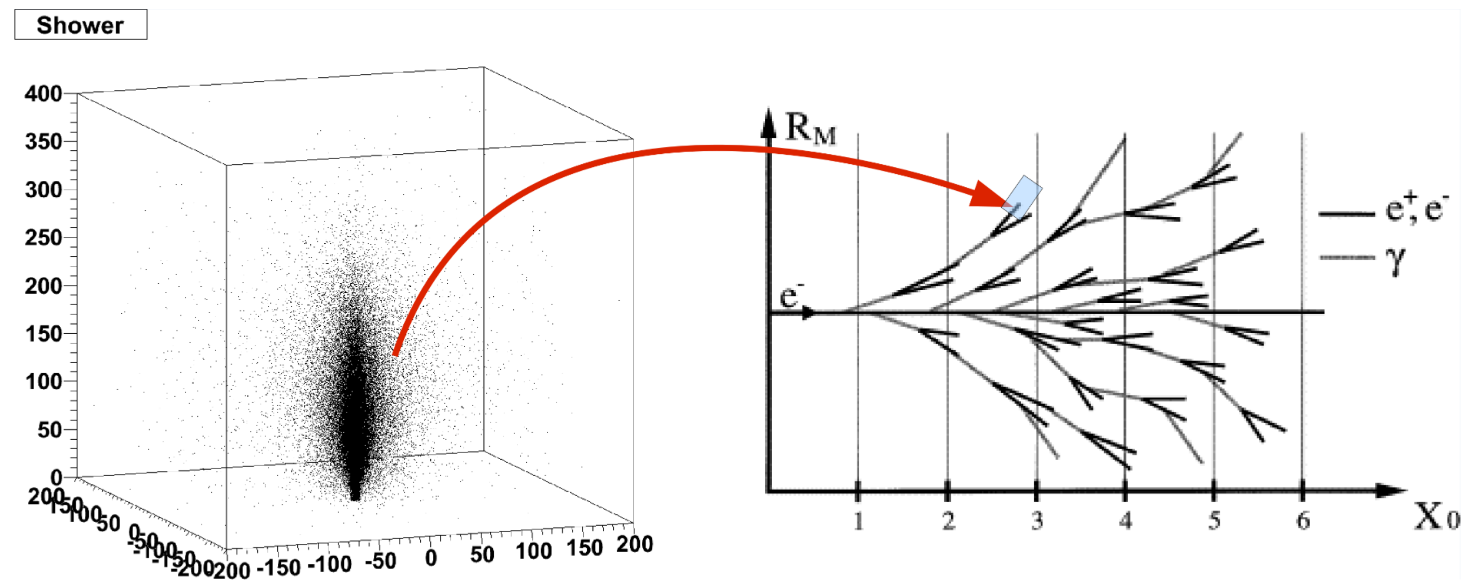


ignore the truth

Frozen Showers

- Many high energetic particles in forward direction => high CPU demand
- Specific to forward EM calorimeters
- Idea: replace low-energetic particles in developing particle showers with pre-simulated EM showers
  - ➡ libraries of frozen Geant4 showers assigned based on particle characteristics

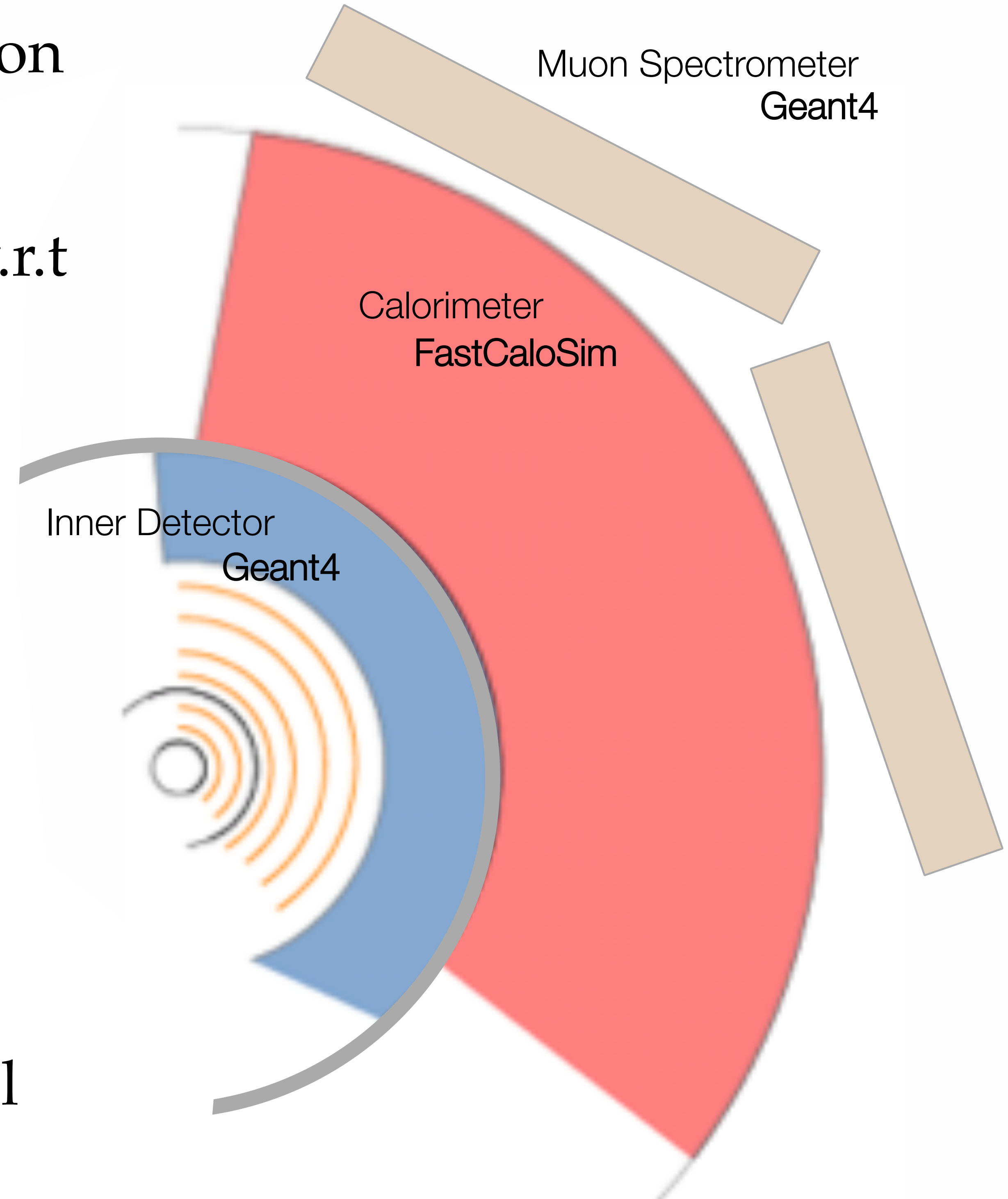




Default in ATLAS “Full simulation”
=> Geant4 + frozen showers for forward calorimetry

AF2 - FastCaloSim

- Replacement of calorimeter simulation with parameterised **FastCaloSim**
- Relative CPU speed improvement w.r.t full Geant4 simulation:
  ~ 20
- Drawbacks:
  - ➡ simplifications in shower shapes (less fluctuations)
  - ➡ per se no hadronic leakage into Muon Spectrometer (can be and is parameterised in ATLAS)
- Used in MC12 production for several physics groups, such as SUSY



AF2 Usage in ATLAS Physics Groups - SUSY

- AF2 used in all SUSY signal samples, except:
  - ➡ Long lived scenarios
  - ➡ Substructure analyses
- Hadronic taus in AF2 shows good agreement with full simulation
- Dedicated jet calibration
- Also used in many Standard Model backgrounds

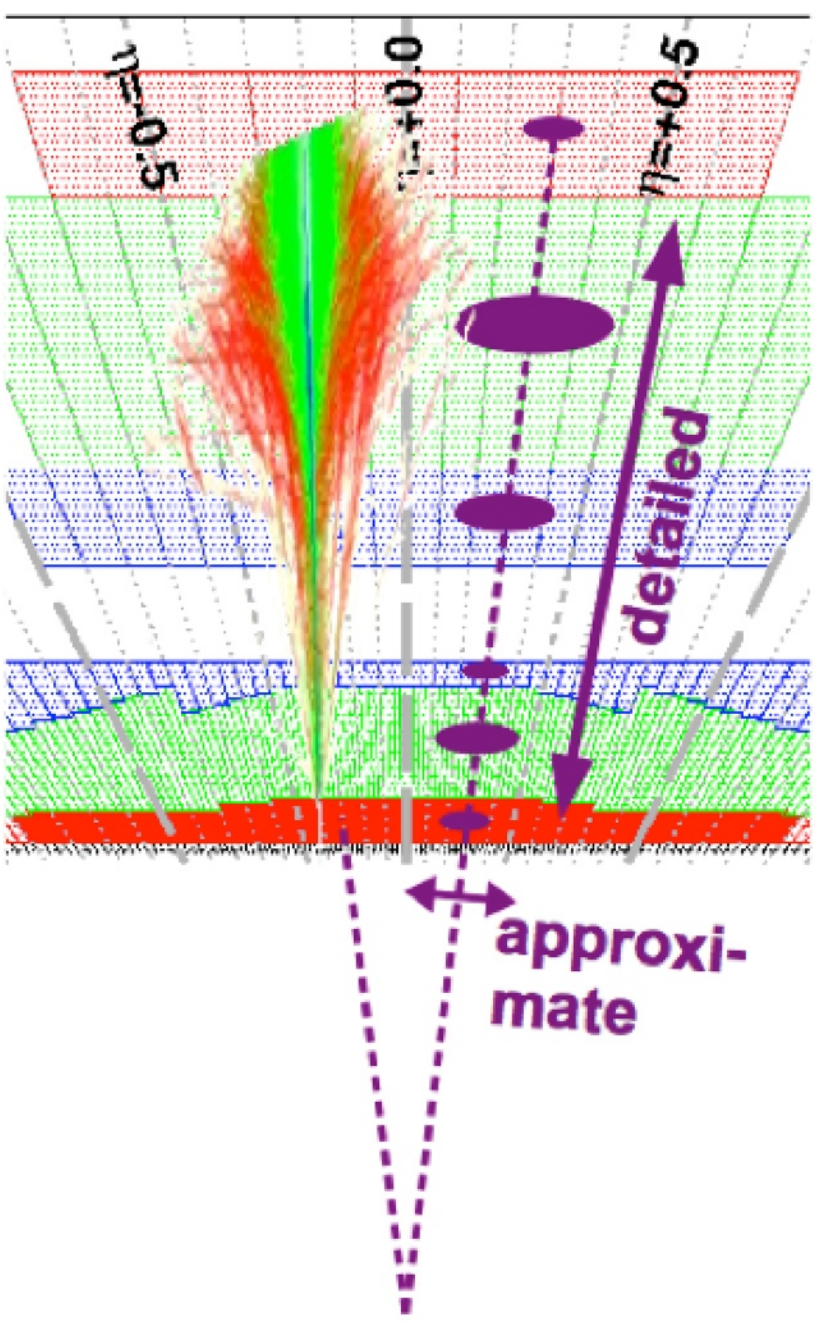
SUSY Sample	# AF2 Events
Model Specific	~ 25 Million
Simplified model (no 3rd generation)	~ 57 Million
Gluino mediated sbottom/stop	~ 25 Million
Stop pair production	~ 50 Million
EWK Production	~ 50 Million

SM Sample	# AF2 Events
ttbar	~ 145 Million
Single top	~ 40 Million
W+jets	~ 500 Million
Z+jets	~ 35 Million

Total: ~930 Million events
And this is low estimate of total #

FastCaloSim

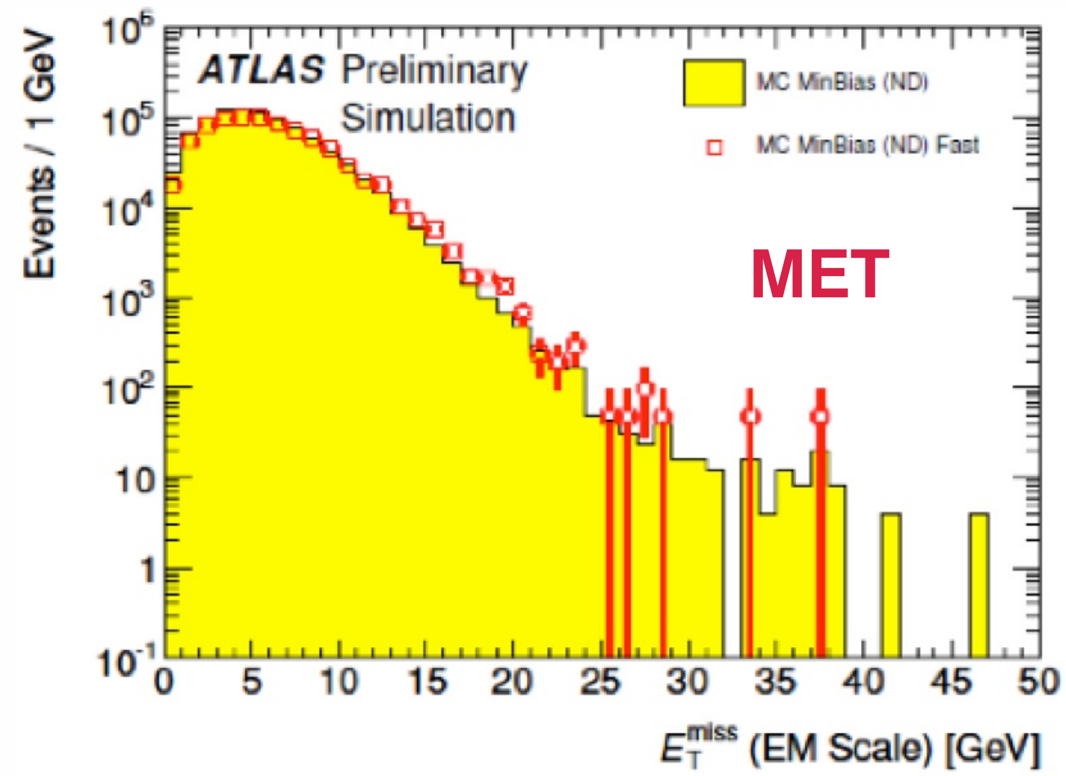
- **Full Simulation**
  - ➡ Full detector geometry
  - ➡ All physics processes for all primary and secondary particles
- Tracking of shower development through the calorimeter in fine steps

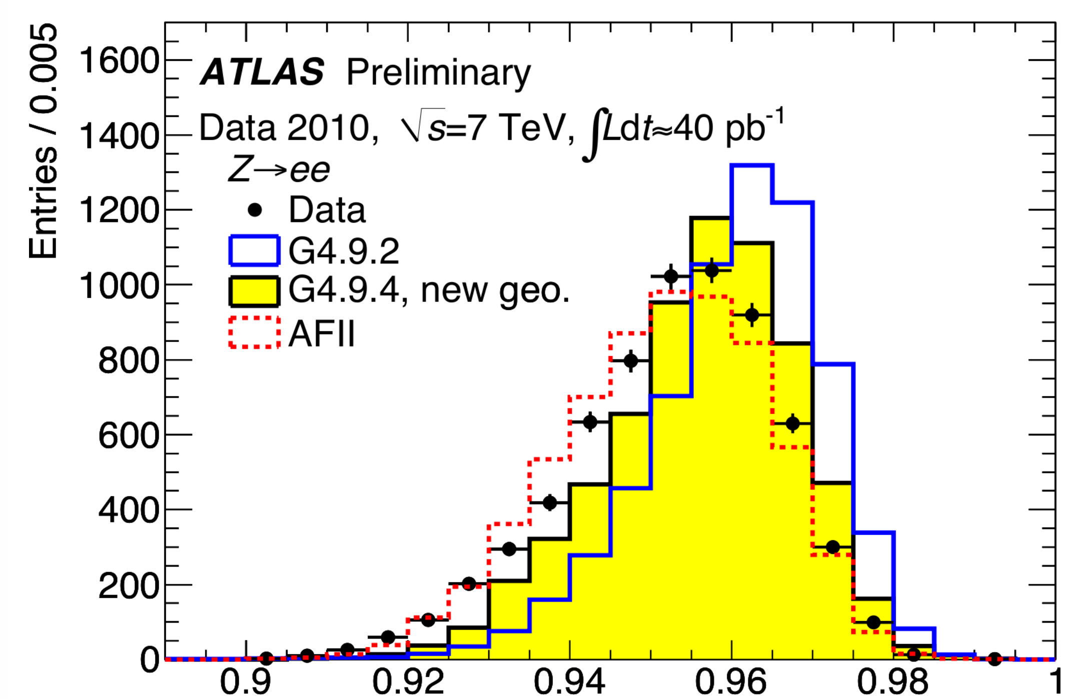


- **FastCaloSim**
  - ➡ Simple geometry with only ~185000 cells
  - ➡ Energy and shape parameterization only for photons, electrons and charged pions
  - ➡ Derived from FullSim
- Deposit of the particle energy in each calorimeter layer in one step

FastCaloSim (II)

- Approximations/shortcuts cause loss of accuracy
  - ➡ usually lead to worse data/MC compatibility
- Parameterizations tuned to data for the EM shower shapes for latest production campaign (MC12)
- Dedicated jet calibrations to use in physics analysis
- Use of pile-up corrections which improves agreement of the jets and MET

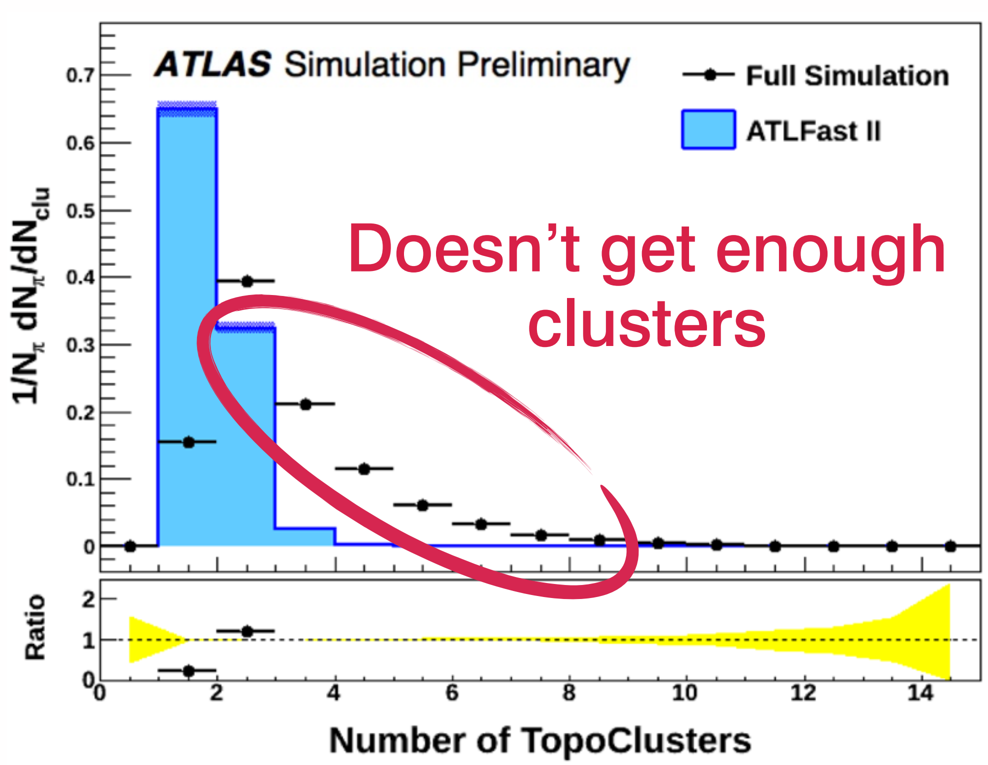


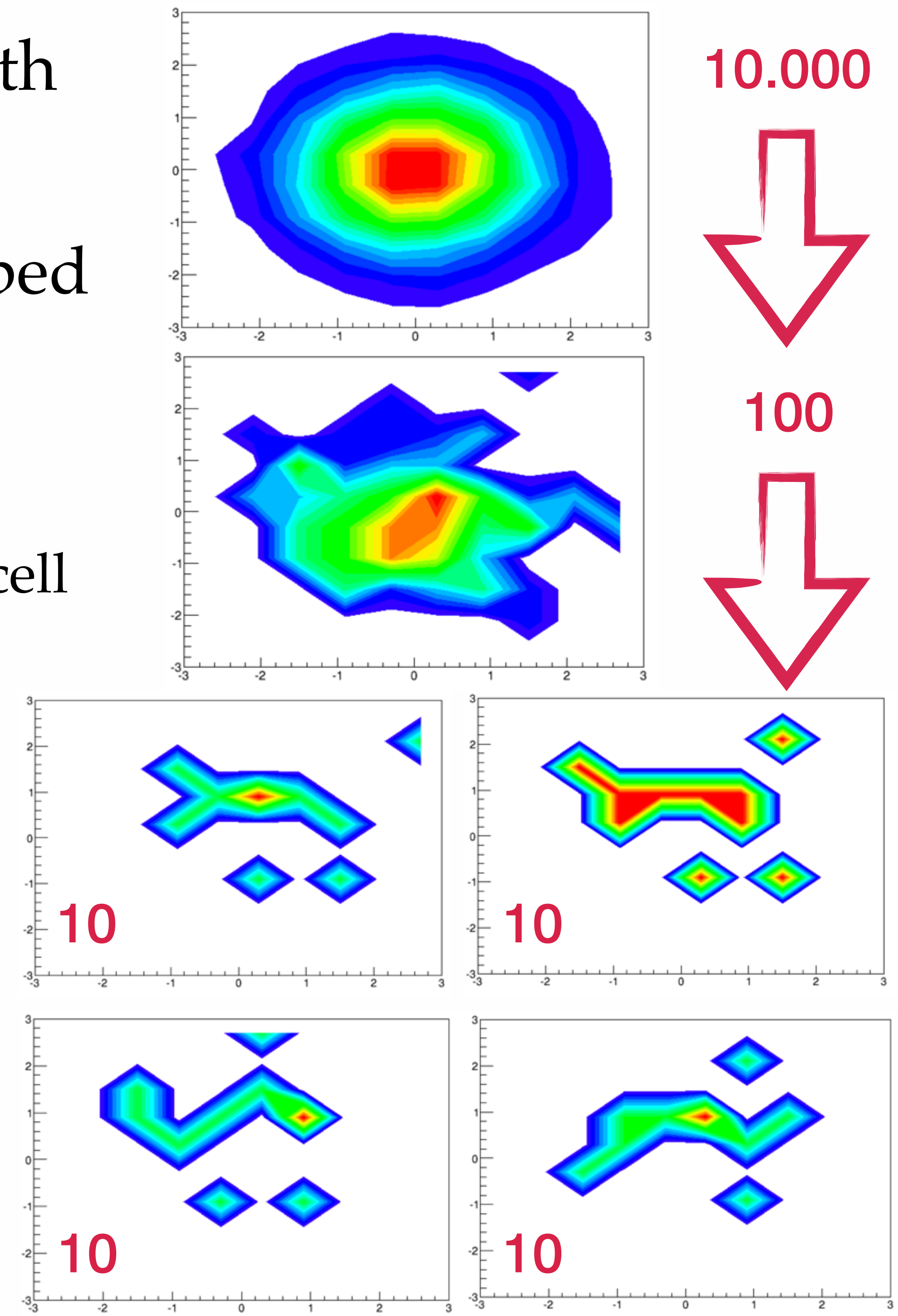


Energy ratio in cell cluster

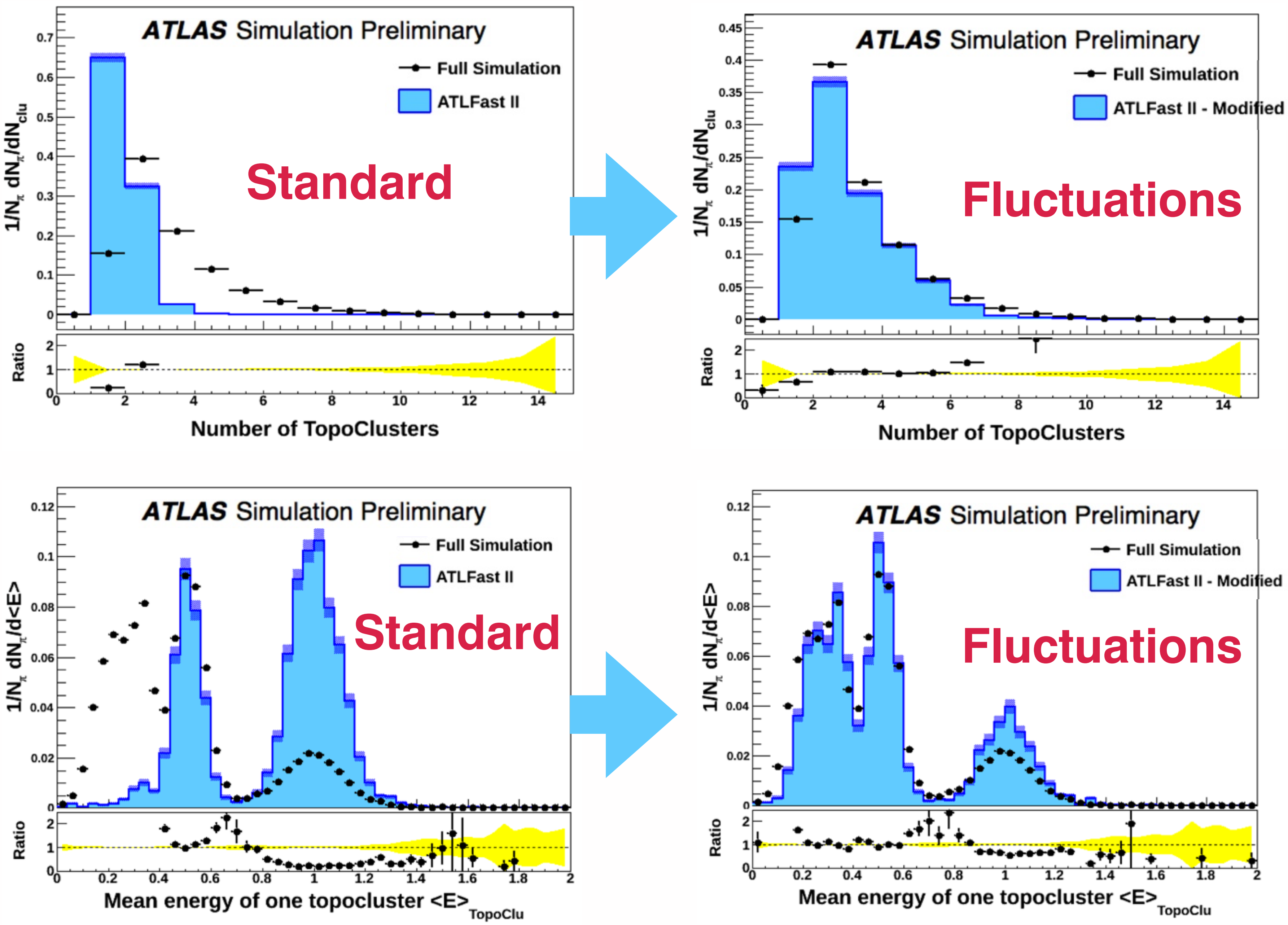
Shower Fluctuations in FastCaloSim

- Overall shower shape is compatible with full simulation
- Sub-cluster structure is not well described by FastCaloSim
  - ➡ Not enough secondary clusters
  - ➡ Introduction of random fluctuations in the cell energy deposits





Shower Fluctuations - Proof of Concept



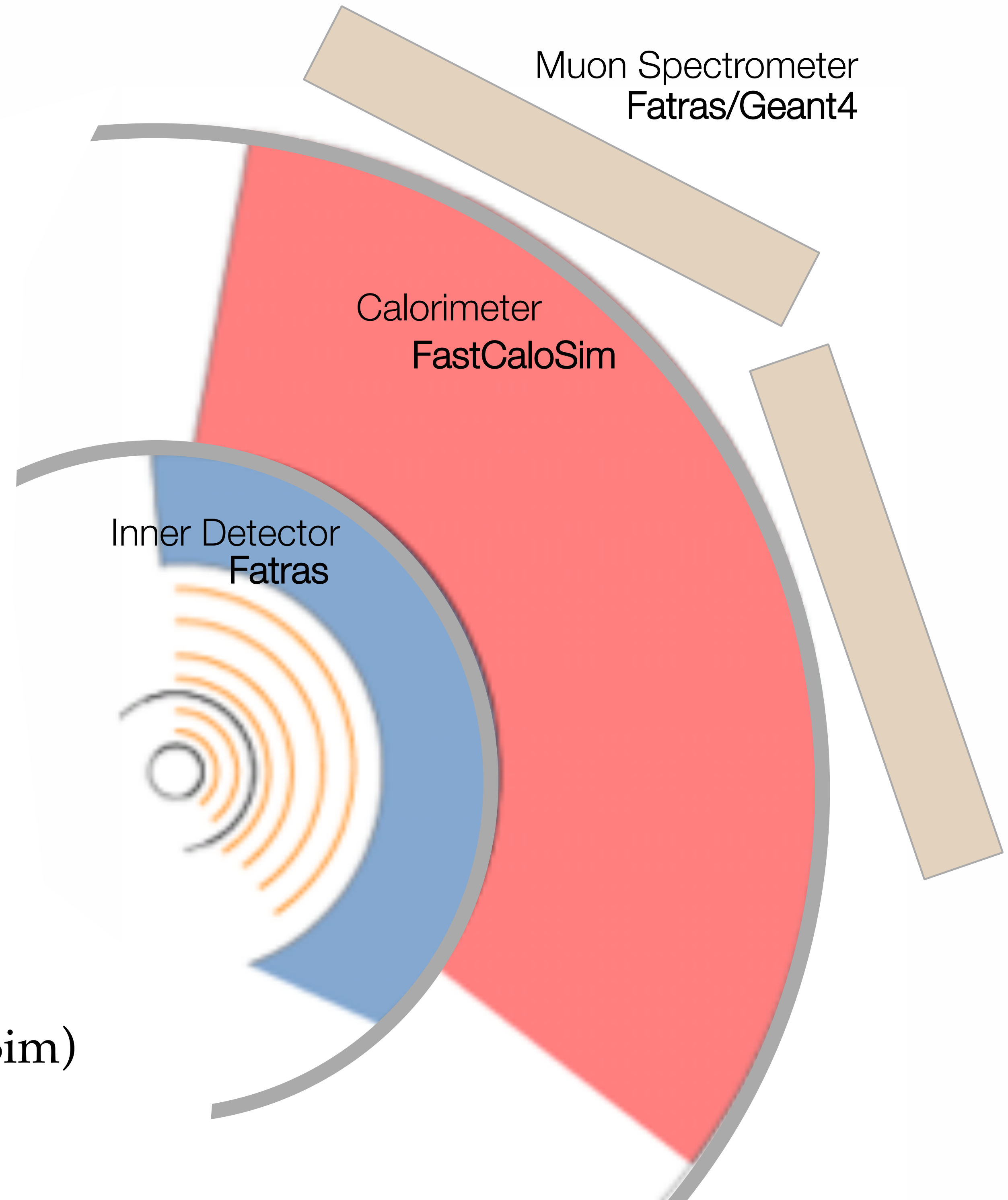
- First attempt shows great improvement in several variables
- Ongoing:
  - ➡ Tuning of fluctuation parameters layer by layer
- To be tested with new FastCaloSim parameterization

AF2F

- Replacement of calorimeter simulation with parameterised **FastCaloSim**
- Replacement of Track simulation with **Fast Track Simulation (Fatras)**

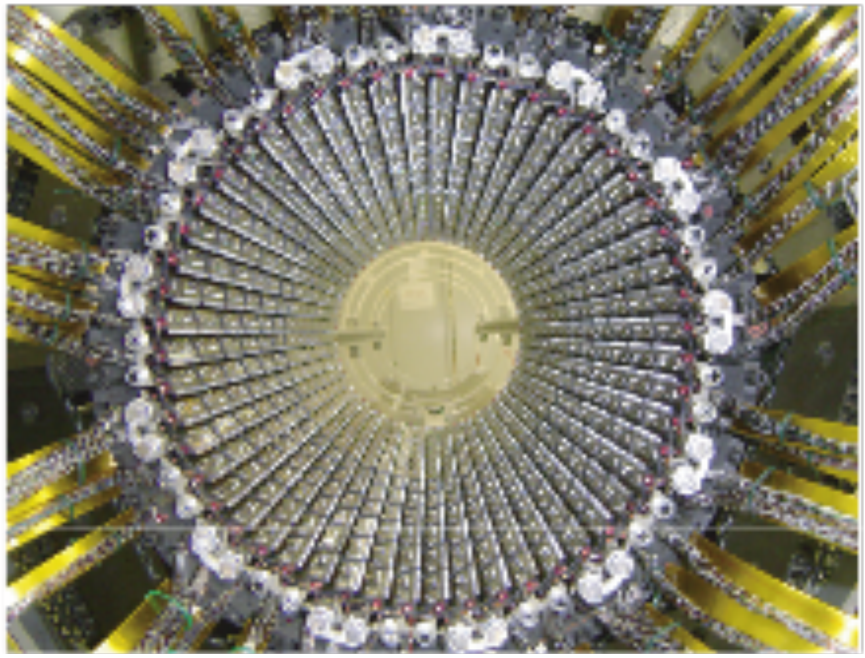
- Relative CPU speed improvement w.r.t full Geant4 simulation:
  > 100

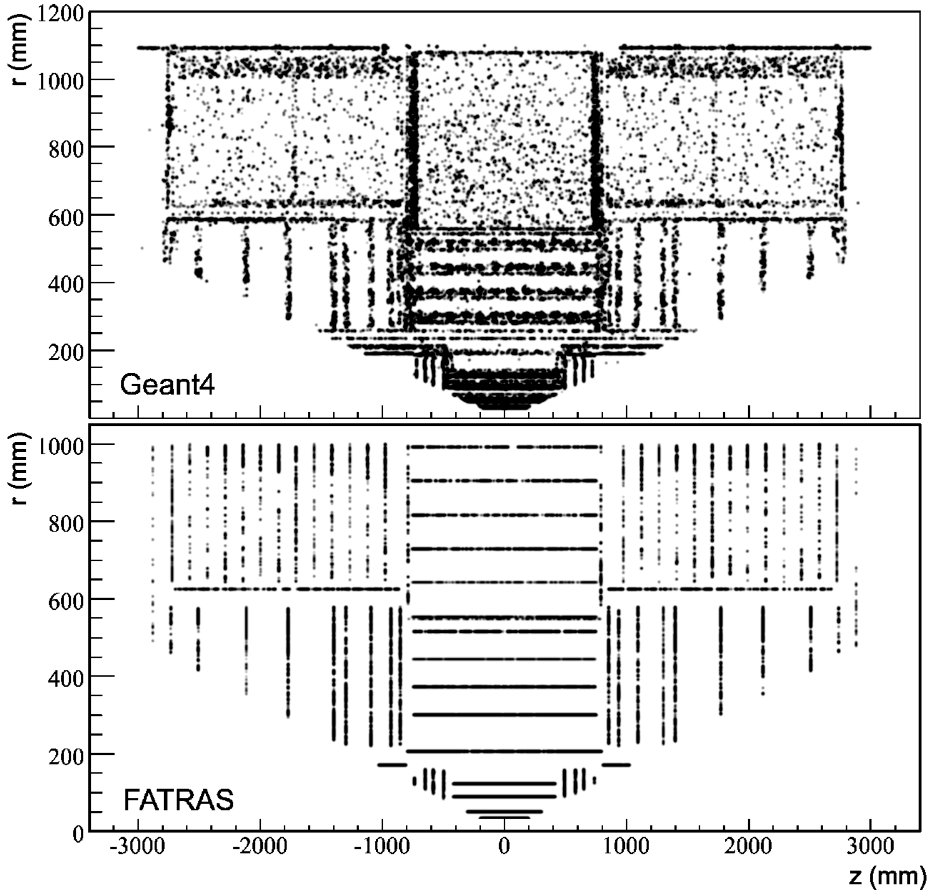
- Drawbacks:
  - ➡ simplifications of material integration (less tail effects in resolutions)
  - ➡ usually slightly higher simulation thresholds (affects hand-over to FastCaloSim)



Fatras

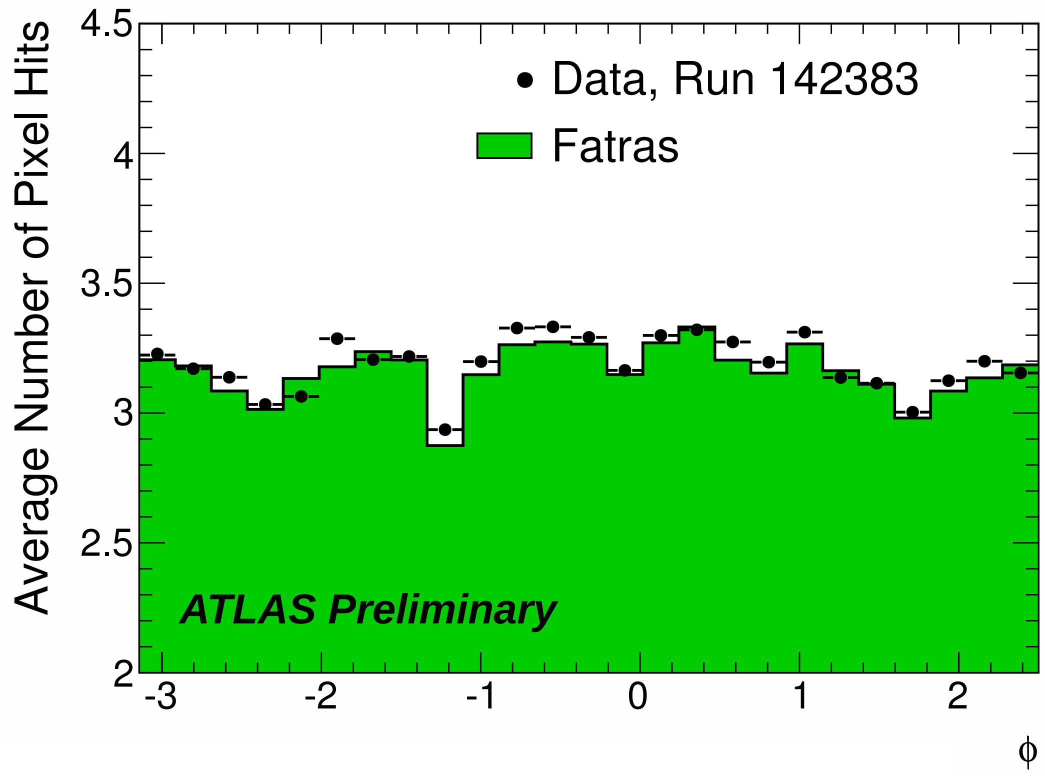
- Treats inner detector, muon spectrometer, and the muons interaction in the calorimeters
- Simplified detector geometry and interaction processes
    - ➡ Keeps the exact description of sensitive elements
    - ➡ Navigation using layers and volume boundaries, modules found by intersection with layer
    - ➡ Material is mapped onto layers using Geant4 description and geantinos

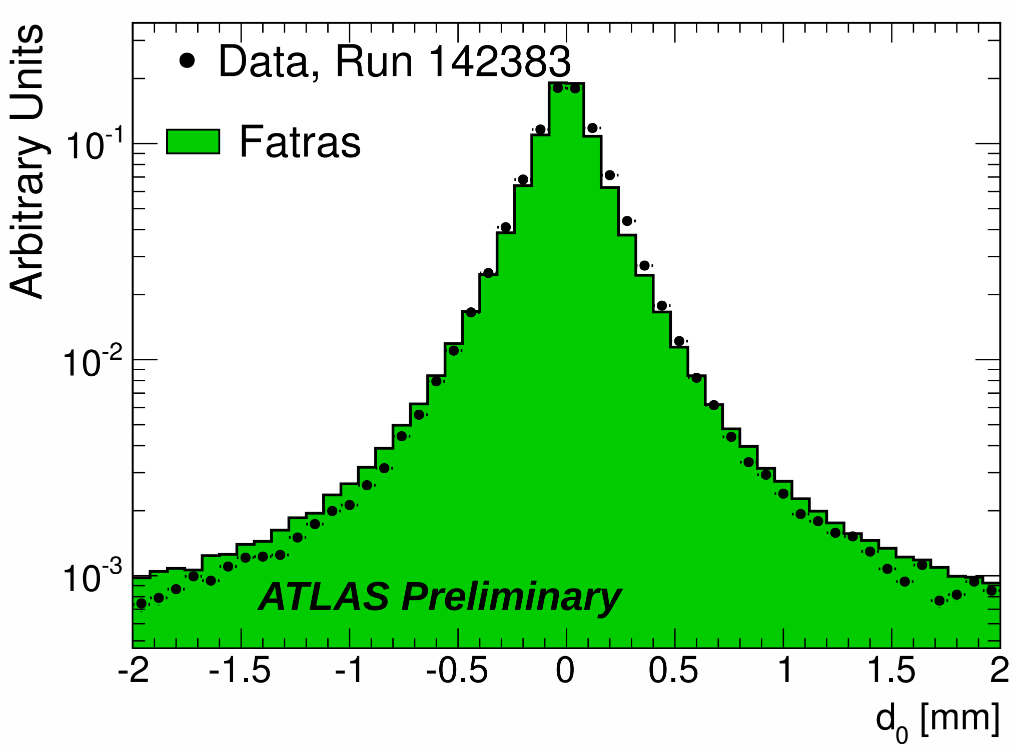


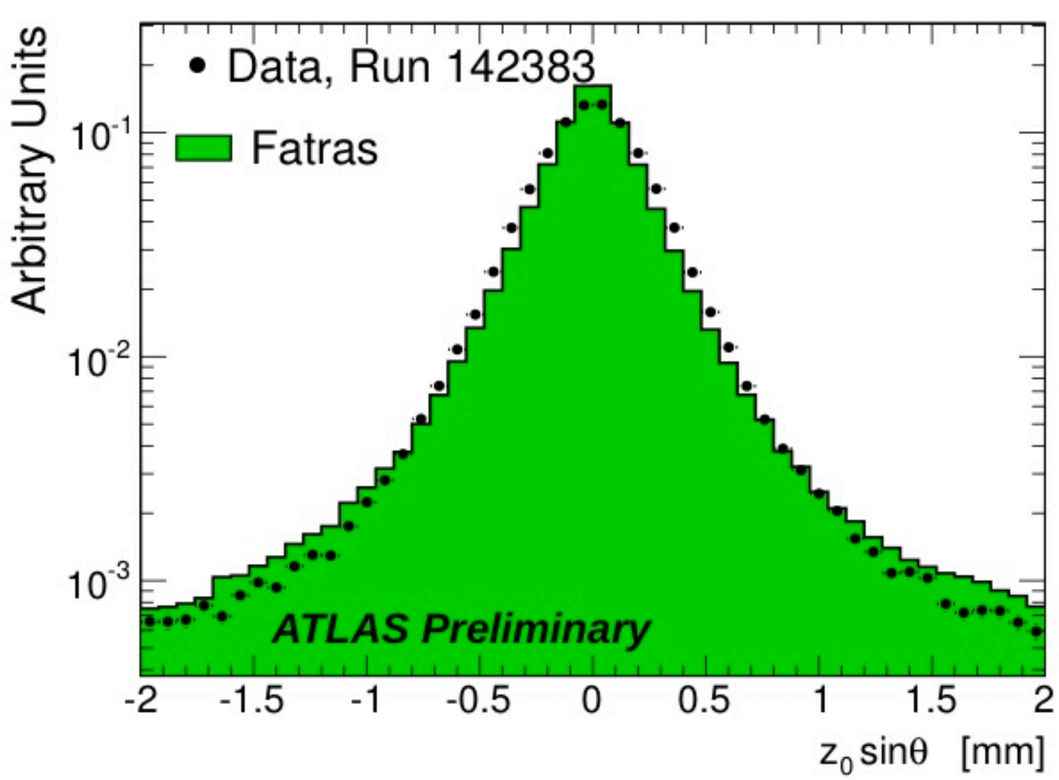


Fatras Performance

- FATRAS in comparison to data
  - ➡ ID reconstruction, tracks with $p_T >$ 500 MeV
  - ➡ using exact same sensitive detector elements:
    - conditions data being fully integrated

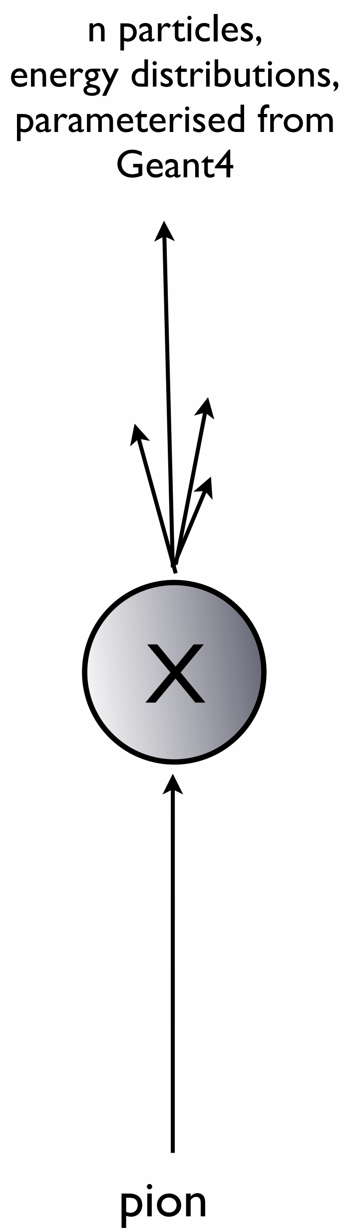


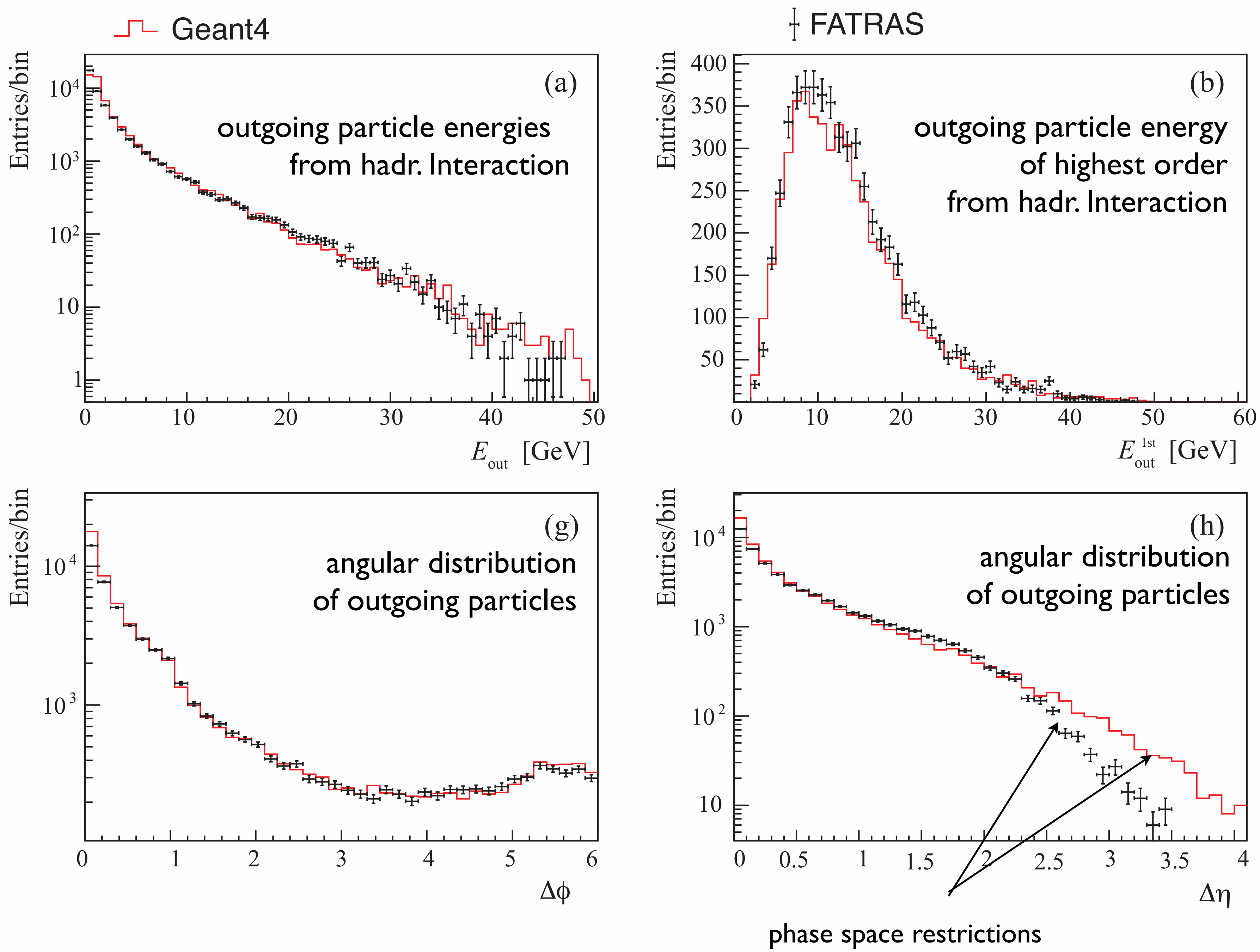




Fatras Simplified Material Effects

nuclear interactions (parametric model implemented)



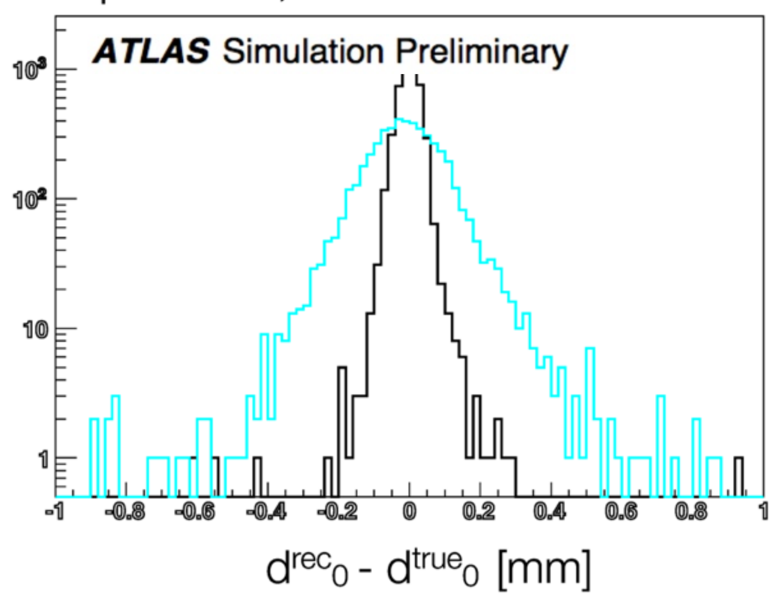


Currently testing a Geant4 based hadronic interaction processor

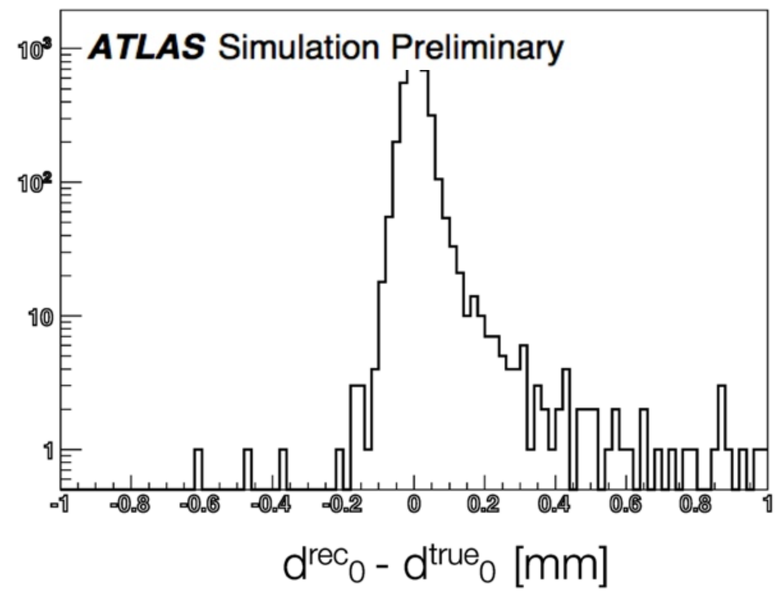
G4 Hadronic Interactions into Fatras

- Hadronic interactions reasonably quick in full simulation
  - ➡ Better accuracy with G4
  - ➡ ~ 50% slower than Fatras parametric but ~ 20 times faster than G4 for 10 GeV pions
- Fatras simplified material description introduces complications in the implementation
  - ➡ Tests been made now to overcome all issues

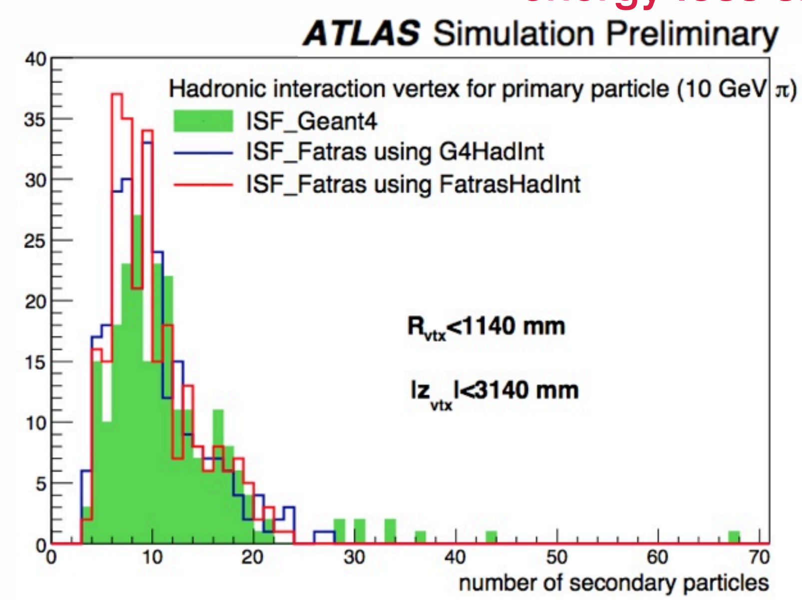
μ 10 GeV, 1 GeV



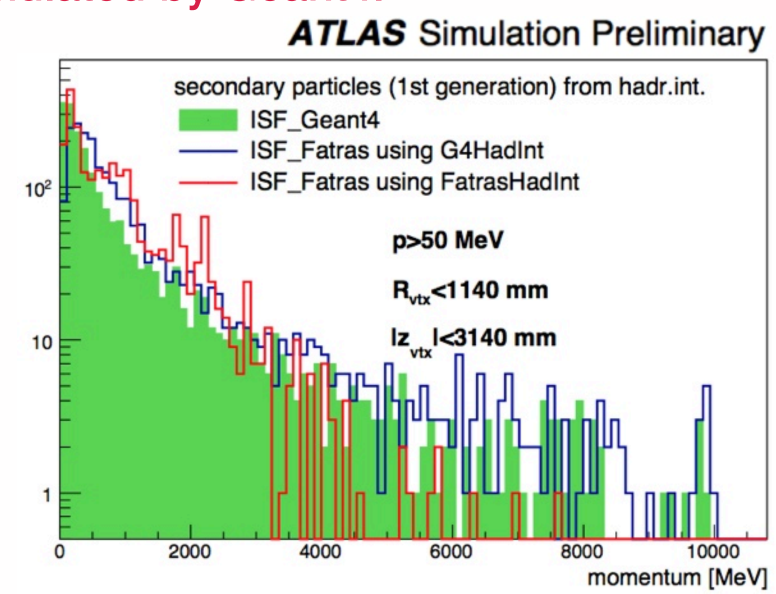
e 10 GeV



Single μ and e processed with Fatras, but multiple scattering and energy loss simulated by Geant4.



Secondary track multiplicity

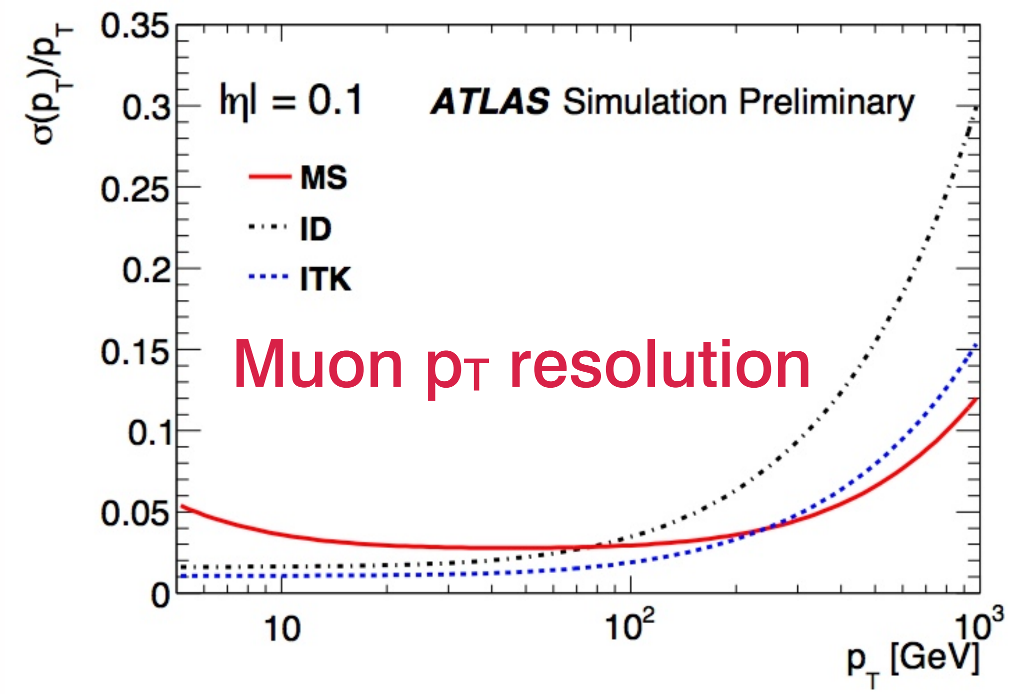


Secondary particle momenta

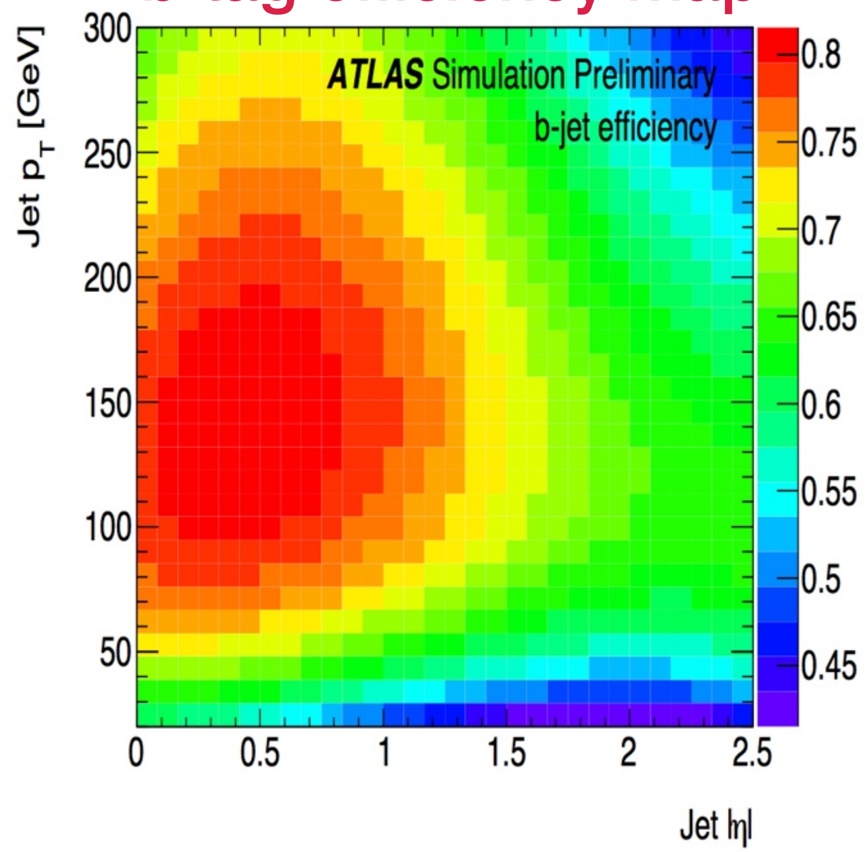
Parameterized Simulation

- Parametric smearing functions
- Detector resolution, reconstruction and trigger efficiencies
  - ➡ extrapolations from existing data sample, and full Monte Carlo simulations with high pile-up scenarios
- Currently used in High Luminosity studies
  - ➡ ES and ECFA functions
  - ➡ See upgrade physics studies public page

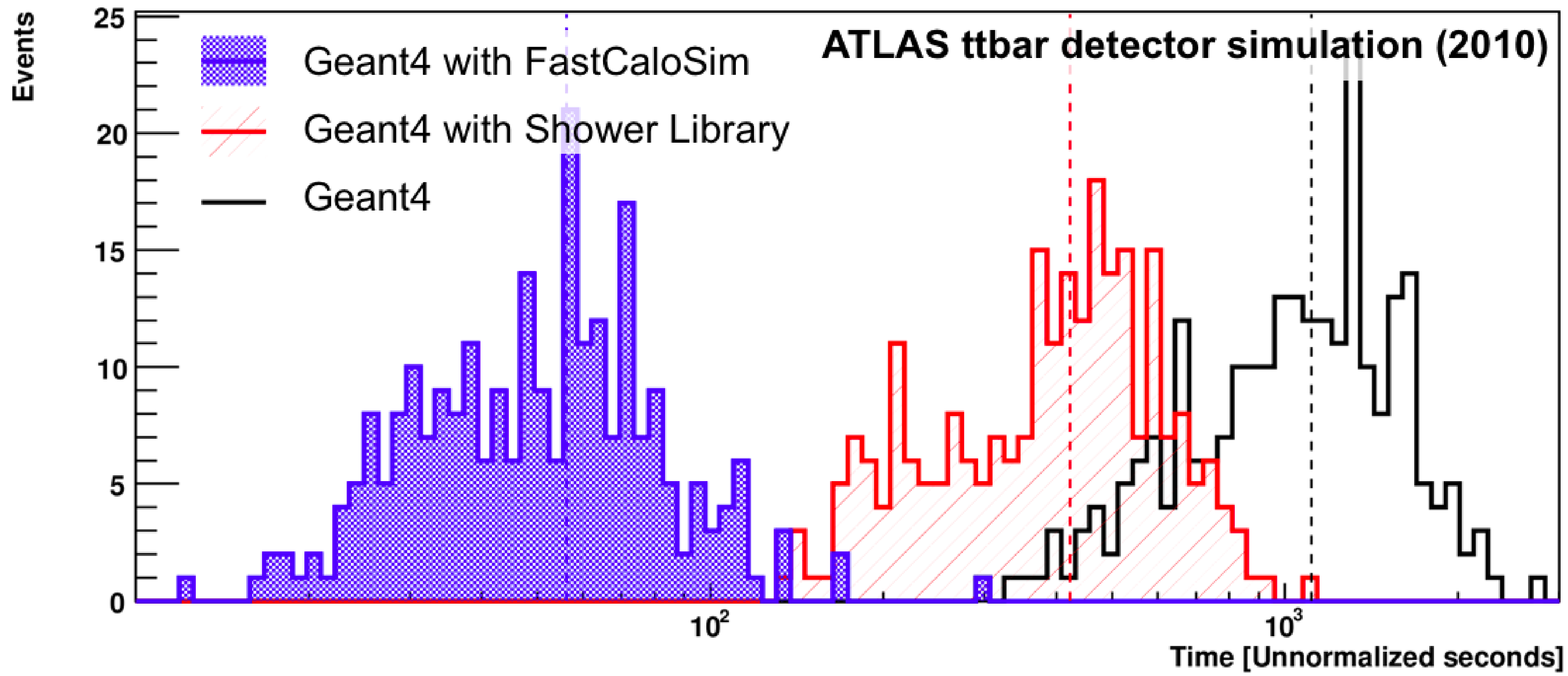
    https://twiki.cern.ch/twiki/bin/view/AtlasPublic/UpgradePhysicsStudies



b-tag efficiency map



Current Simulation Performance

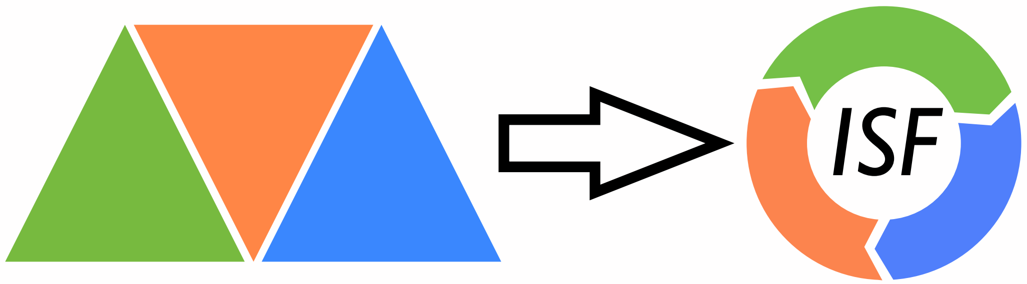


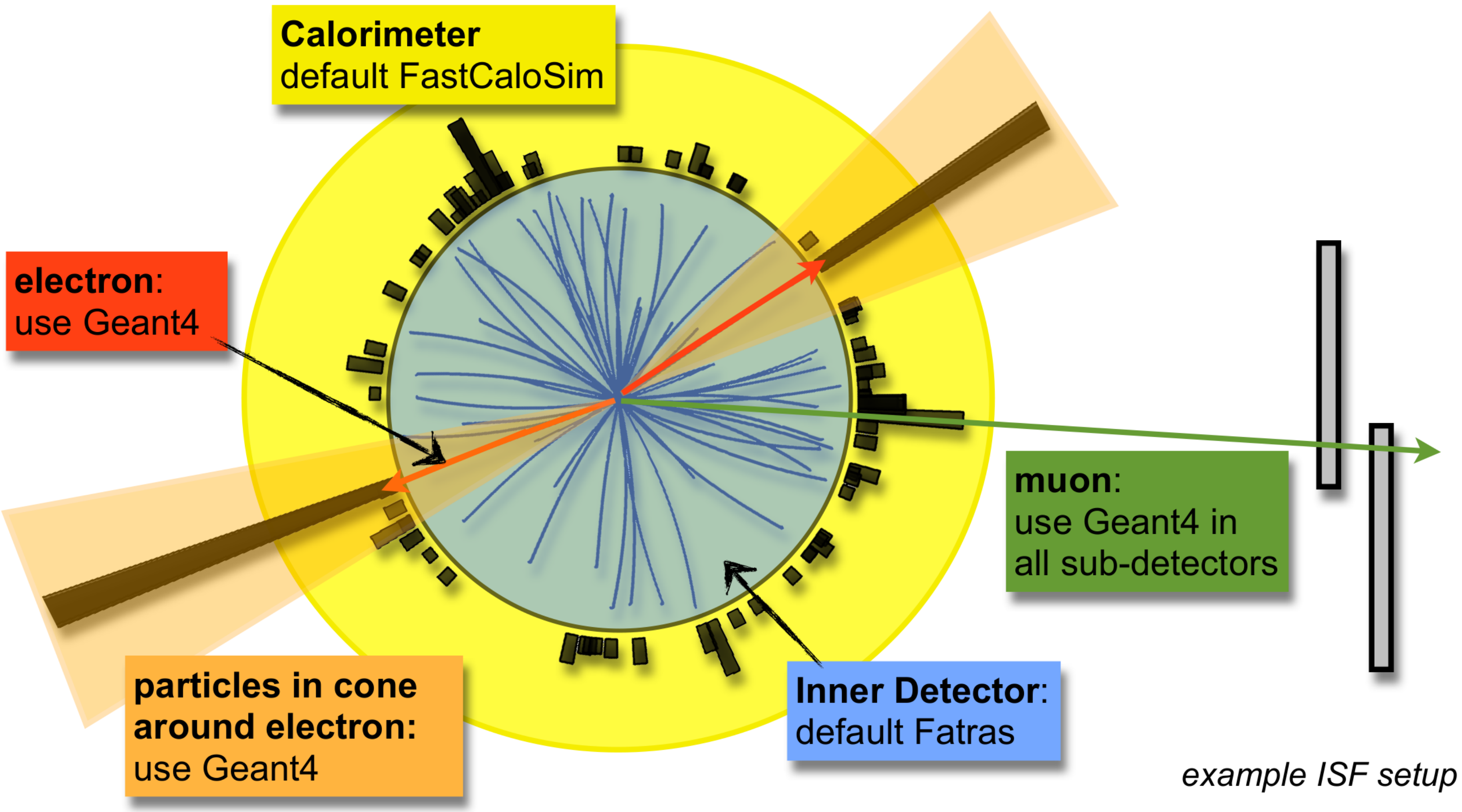
Number of events simulated in ATLAS for the MC12 campaign:

- ➡ 3.9 G => full simulation (Geant4 + frozen showers)
- ➡ 3.0 G => fast simulation (Geant4 + FastCaloSim)

ISF - Integrated Simulation Framework

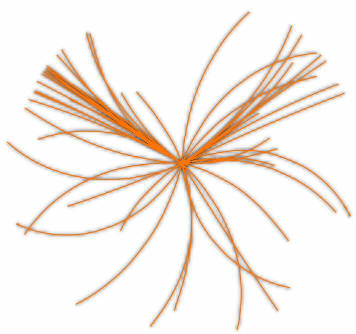
- Combines different simulation approaches in ATLAS into one framework
  - ➡ Output format is always the same independent of simulation chosen
  - ➡ Configuration is done at one central place and standardized
  - ➡ Fast and full simulation setup can be mixed and used alongside
- Compatible with multithreading and multiprocessing

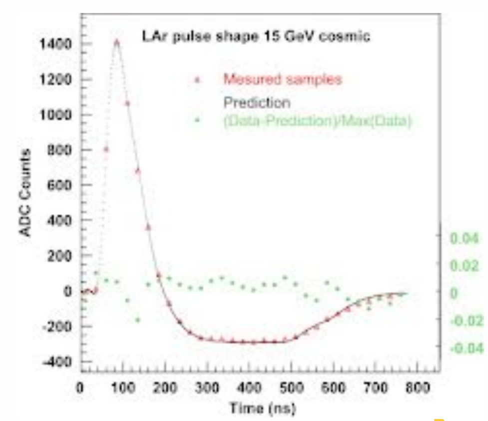


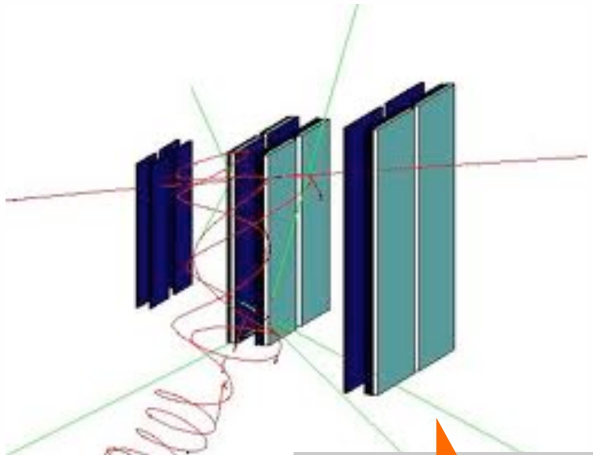


example ISF setup

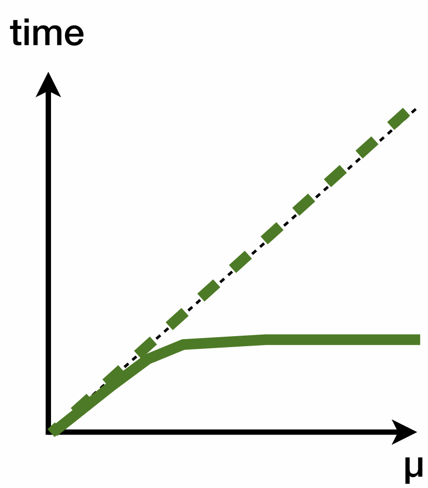
Fast MC Chain vs Pile-Up

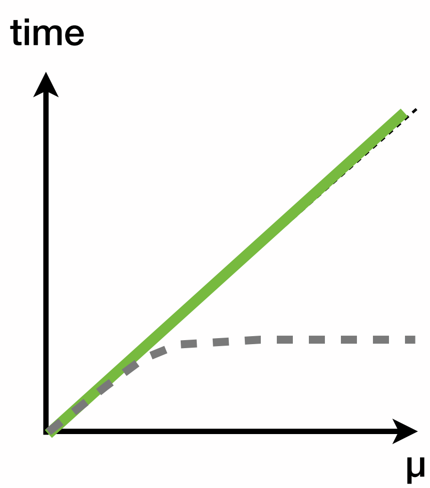


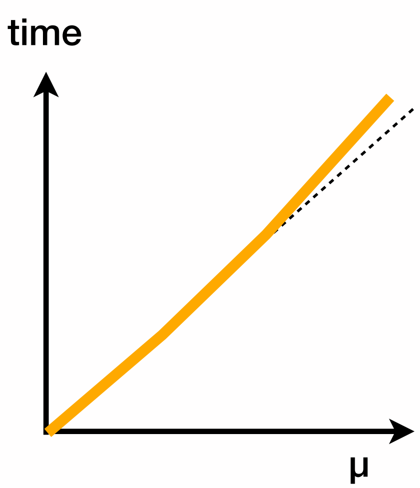


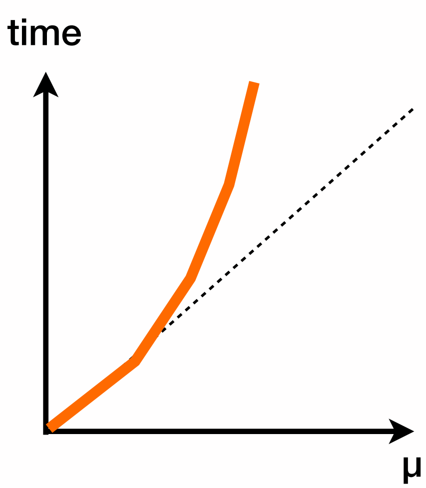


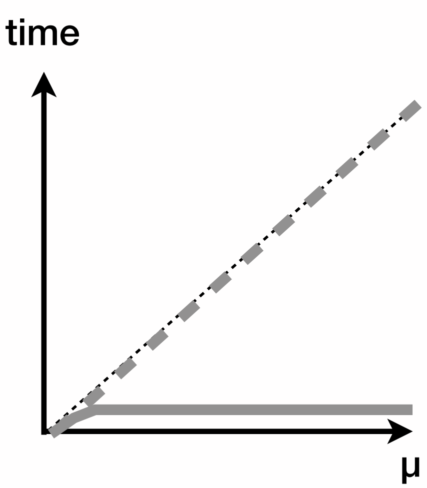
Event Generation → **Detector Simulation** → **Digitization** → **Reconstruction** → **Analysis**



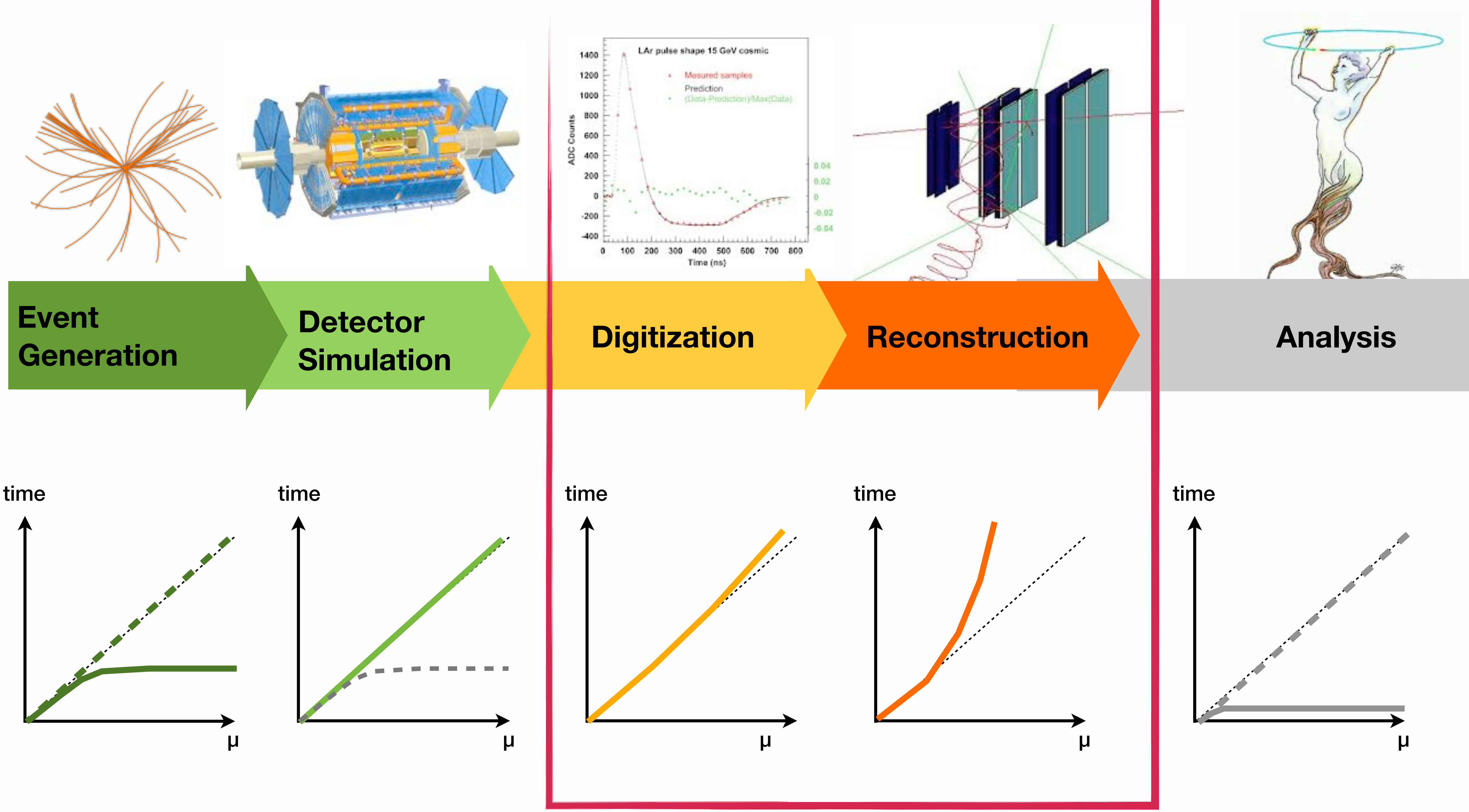








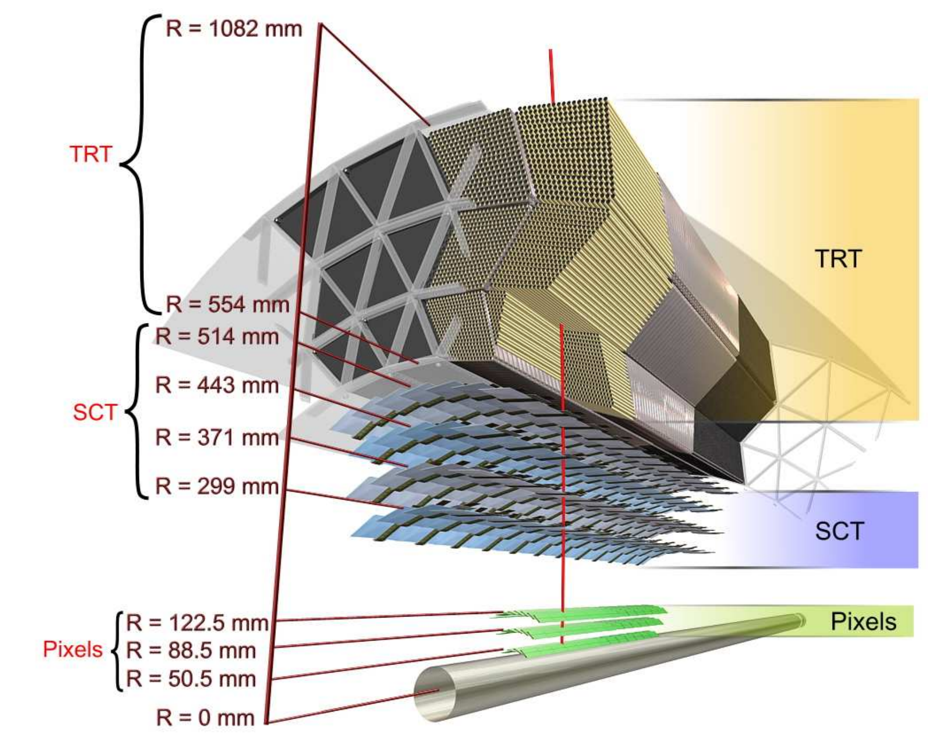
Fast MC Chain vs Pile-Up



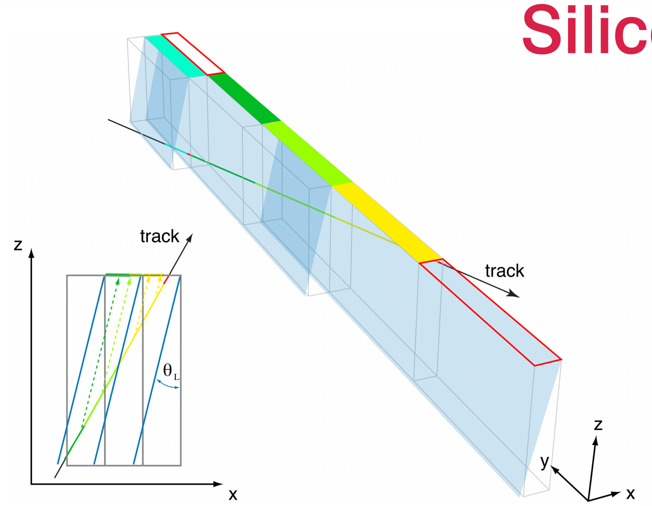
ATLAS is also working on fast digitization and fast reconstruction

Fast Digitization

- Detector Simulation main bottleneck for current MC production
- Speedup of simulation up to ~3k times
  - ➡ next bottleneck are Digitization and Reconstruction
- Digitization time dominated by **Inner Detector** (~50%)
- **Fast Pileup** - Use in time pileup to model out of time pileup contributions using detector specific weights

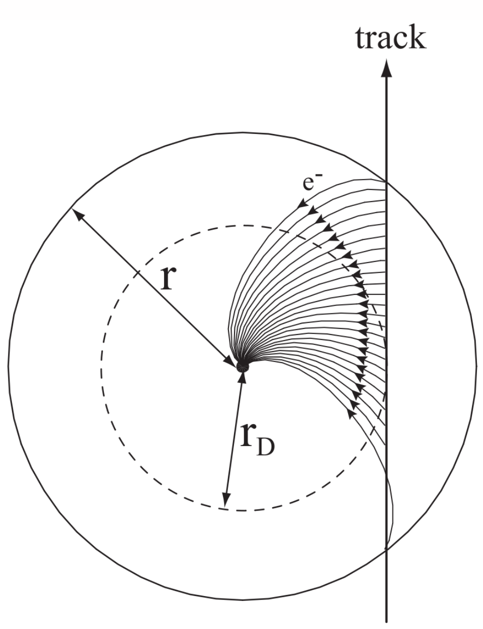


Silicon



Fast digitization method for silicon detectors: the particle path length is projected on the readout surfaces, corrected for Lorentz angle drifts and smeared.

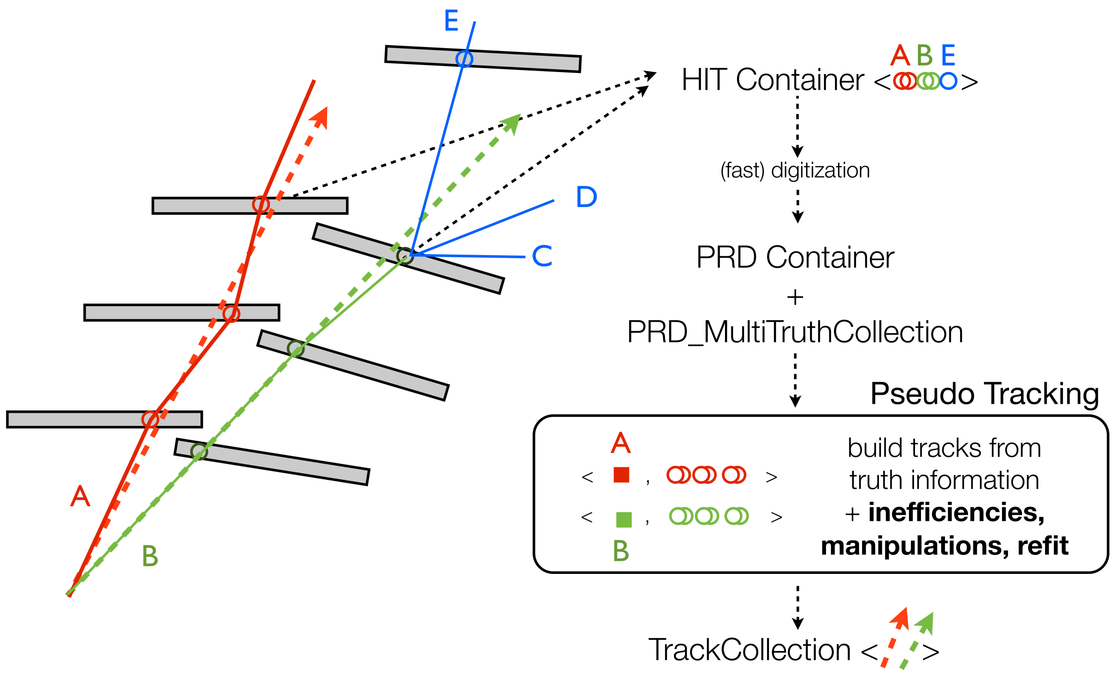
TRT



Fast digitization method used for the TRT: the closest approach radius is computed together with the uncertainty on its measurement from the simulated HITS, giving an estimate for the drift radius r_D used at reconstruction level.

Fast Reconstruction

- **ID**: most time consuming because of combinatorics in pattern recognition
- Fast tracking:
  - ➡ Seed track from MC truth
  - ➡ Skip most time consuming steps:
    - pattern recognition
    - track seeding
    - ambiguity treatment
  - ➡ reconstructed track fit to hits from truth
  - ➡ high efficiency at high pileup



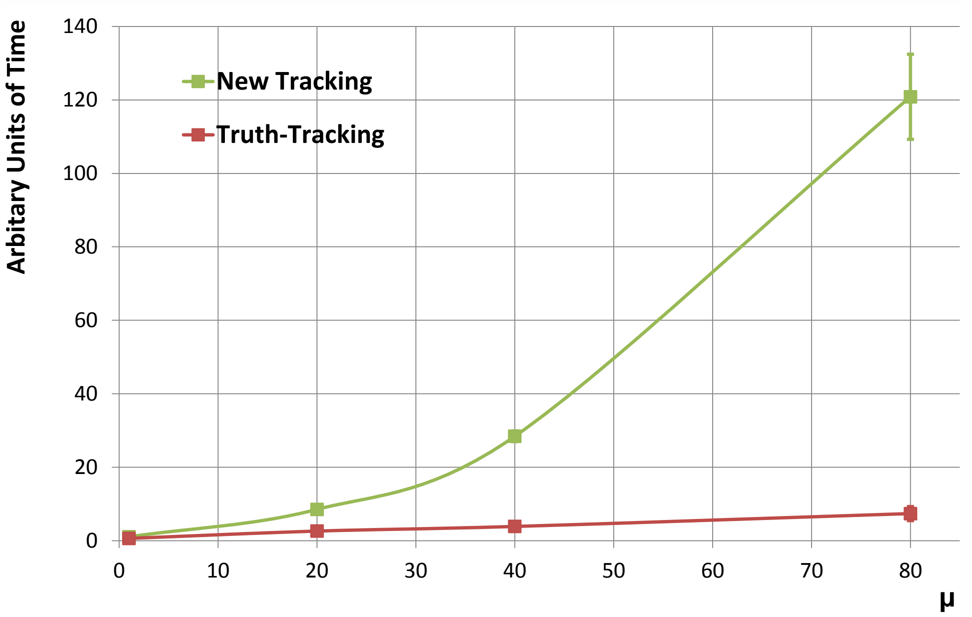
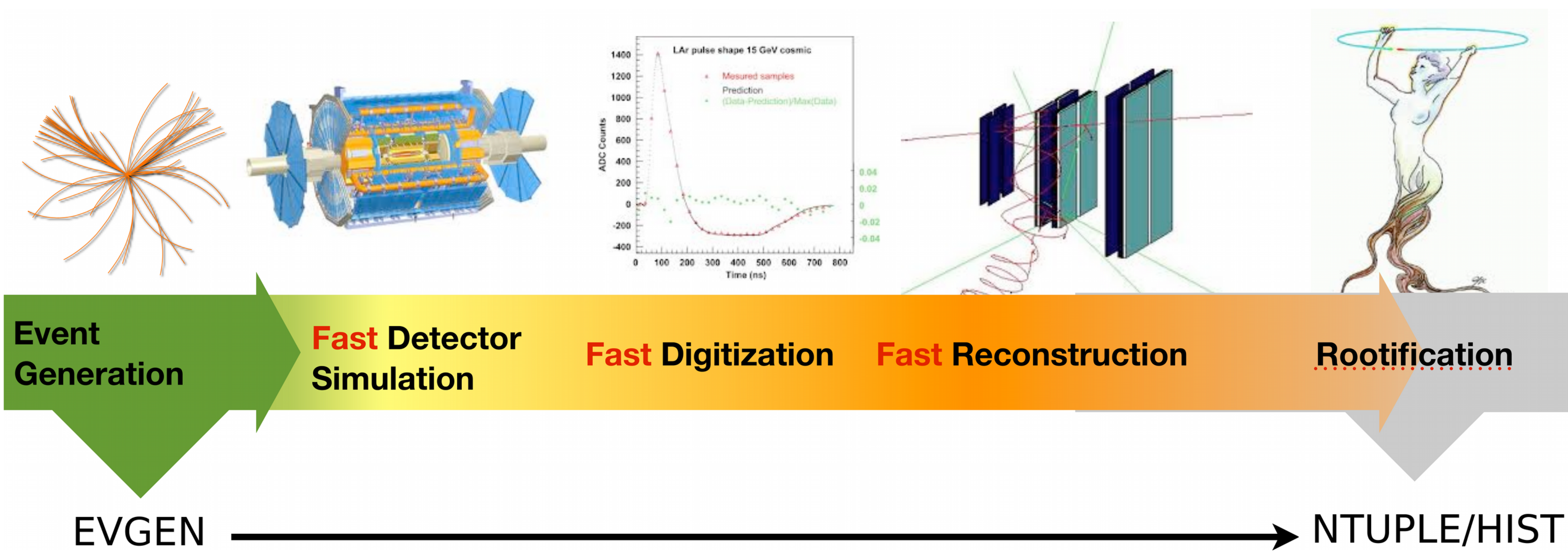


Figure 30: Overview of μ dependance of the reconstruction time for TT and NT.

Final Product



- Evgen to ROOT in one go
  - ➡ I/O writing next bottleneck after Fast Sim/Digi/Reco
  - ➡ No intermediate output (minimisation of I/O overhead and storage disk space)
  - ➡ Fast Simulation + Fast Digitization + Fast Reconstruction
- Estimated time per event: a few seconds
  - ➡ Possibility for large scale MC production with substantially lower resources

Summary and Outlook

- ATLAS is developing and maintaining both full and fast simulation
- A lot of work ongoing on the Integrated Simulation Framework
  - ➡ dynamic use of different Simulation technologies based on event characteristics
- Fast digitization for silicon and transition radiation tracking technologies
- Fast reconstruction => tracking based on seeding from MC truth
- Fast MC production chain:
  - ➡ combination of Fast Simulation, Digitization and Reconstruction
  - ➡ 4-vectors → ROOT in one step
  - ➡ only a few seconds necessary to process an event

Backup Slides