Advancements in the ATLAS Fast Chain for HL-LHC: Towards Efficient MC Production

Fang-Ying Tsai
(Stony Brook University)
On behalf of the ATLAS collaboration

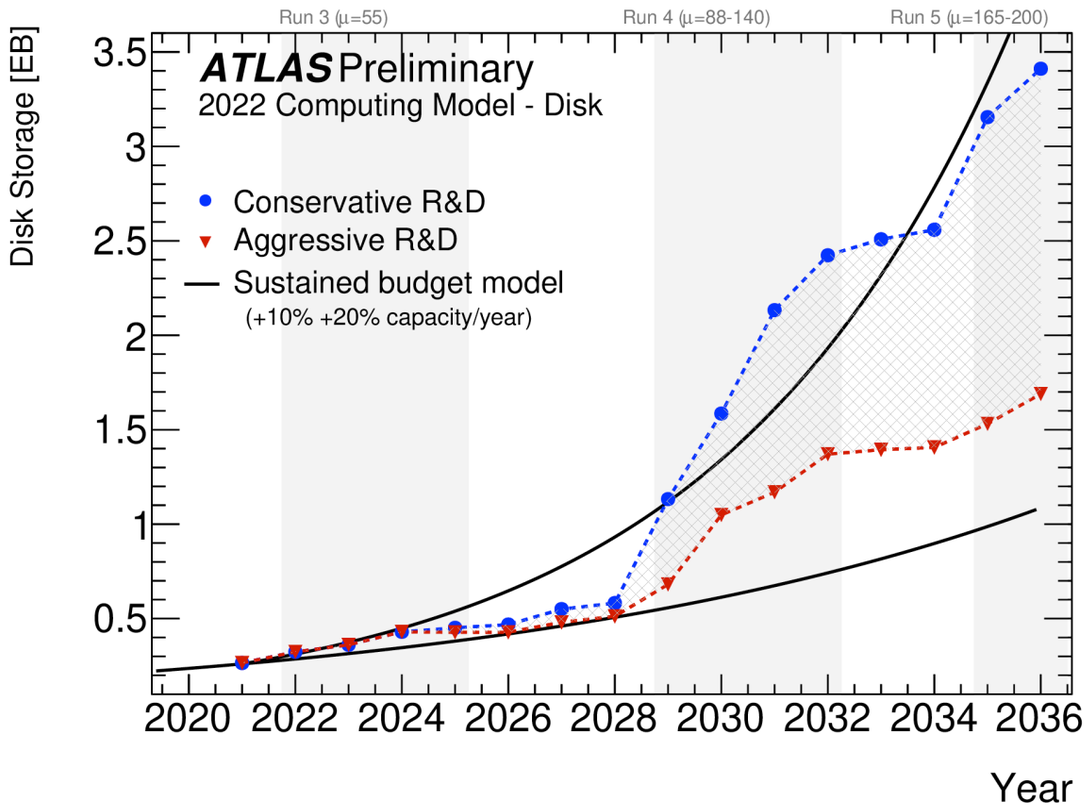
CHEP 2024

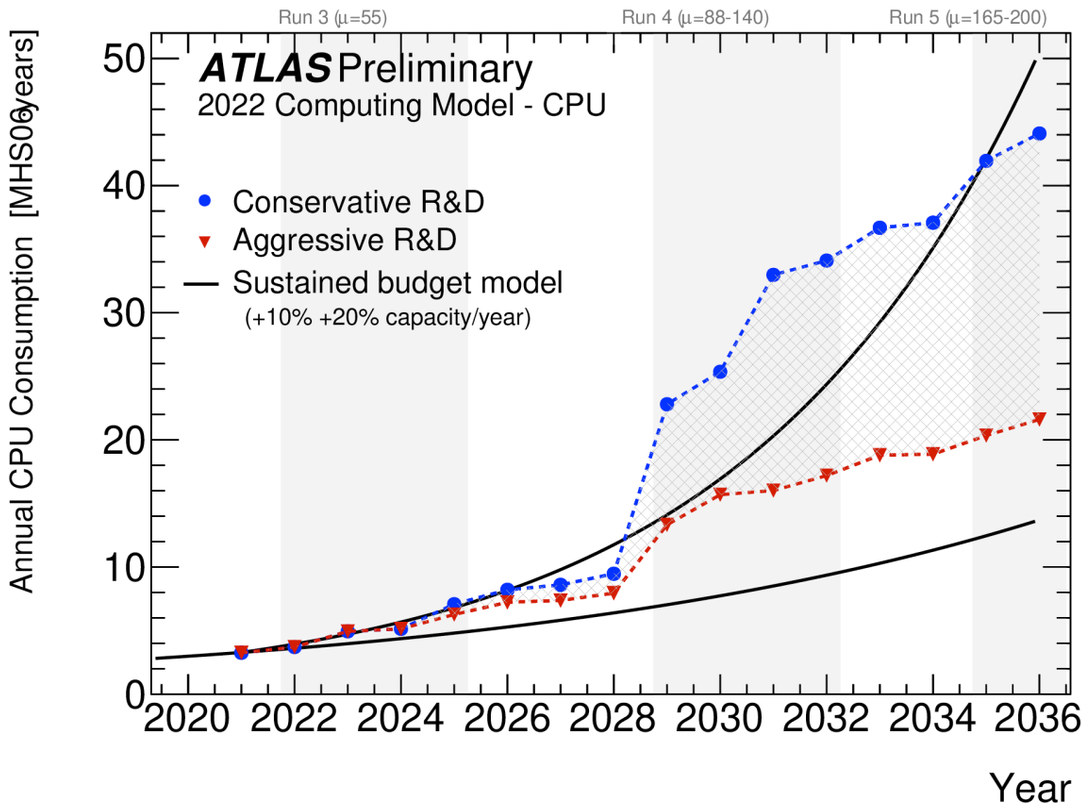
Outline

- **Intro: Accelerating simulation for the HL-LHC era**
  - **The need for simulation speed-up**
  - **ATLAS detector overview**
  - **Fast Chain workflow for MC production**
  - **Strategies to speed up in ATLAS**
- **FATRAS: Fast ATLAS Tracking simulation**
- **Track Overlay**
- **Conclusions**

The need for simulation speed-up

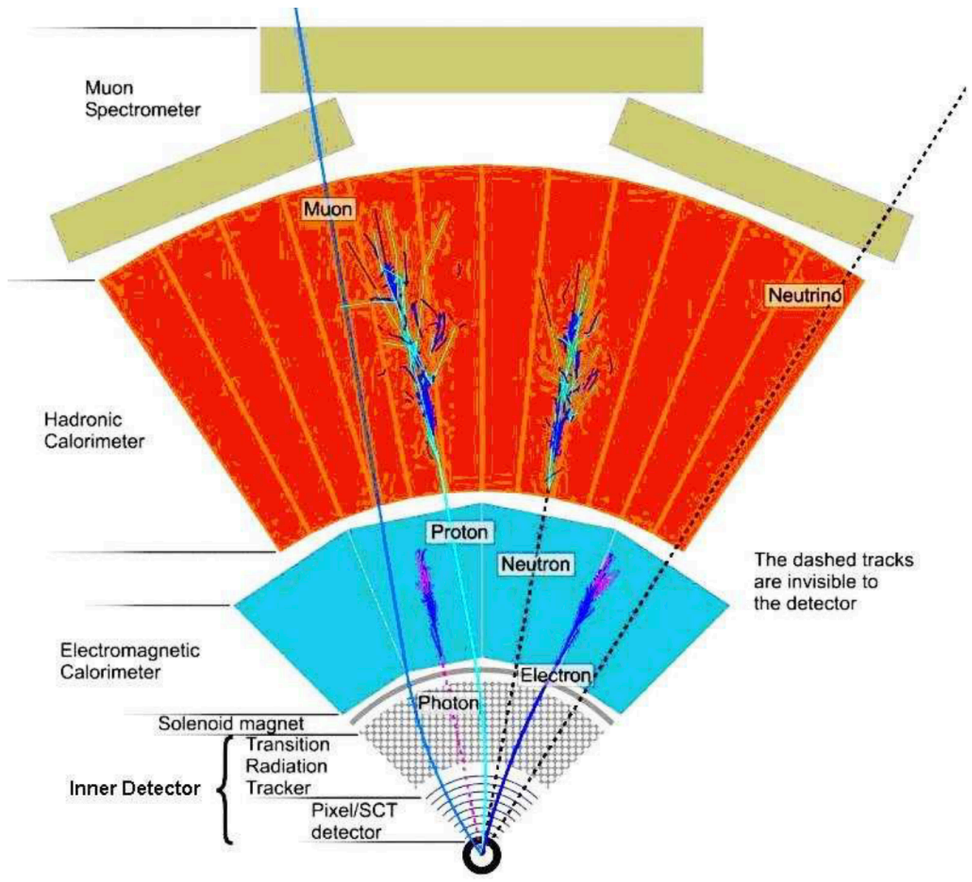
- Massive data and storage constraints.
  - simulation speed is critical for the HL-LHC, especially with increased pile-up and event rates.
- FastChain helps to meet precision demands in Physics modeling while operating within resource limitations.





Source: CERN-LHCC-2022-005

ATLAS detector overview



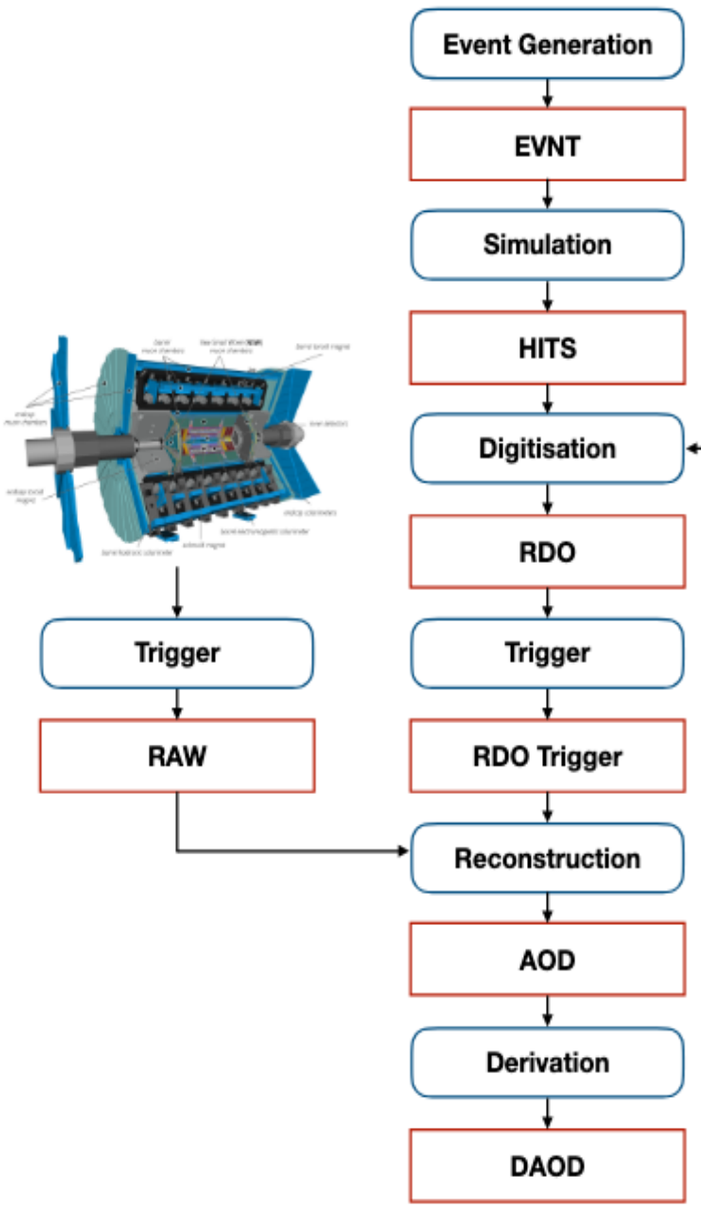
- Inner Detector:
  Tracks charged particles, measures momentum, and reconstructs vertices.
  - Challenges: high particle density and complex geometry require sophisticated simulation techniques.

- Calorimeters (ECal and HCal):
  ECal measures energy of electrons and photons.
  HCal measures energy of hadrons.
  - Challenges: Electromagnetic and hadronic showers produce complex (and very different) cascades of particles.

- Muon Spectrometer:
  Fewer particles interact with muon spectrometer than in the ID or calorimeters, full Geant4 is often still feasible here.

Fast Chain workflow for MC production

Standard software workflow



Source: 2404.06335

Fast Chain components in development

Simulation → Fast Simulation:

FastCaloSim & FastCaloGAN (AF3) + Fatras (first half of talk)

Simplified Geometry: Reduce the number of boundary crossings, leading to larger **step length** and faster navigation.

The material properties influence the **step length** and interactions (energy loss, scattering, photon conversion), which can affect the **direction.**

Trigger → Fast trigger tracking: under investigation.

Reconstruction: factorize tracking from extra pp interactions using track overlay technique. (2nd half of talk)

- From Event Generation → directly to user outputs: through eliminating intermediate files (e.g. HITS and AODs) and optimizing CPU utilization.

Strategies to speed up simulation

- Wall time speed up
  - Multi-threading or Multi-processing.
  - Resource usage optimization

CPU usage

* Optimized Detector Geometry description for FullSim.
* Fast Simulation.

GPU usage (e.g. use Geant4 API to offload part of Geant4 workload to the GPUs)

→ e.g. Celeritas & AdePT on HPCs.

(Work in progress)

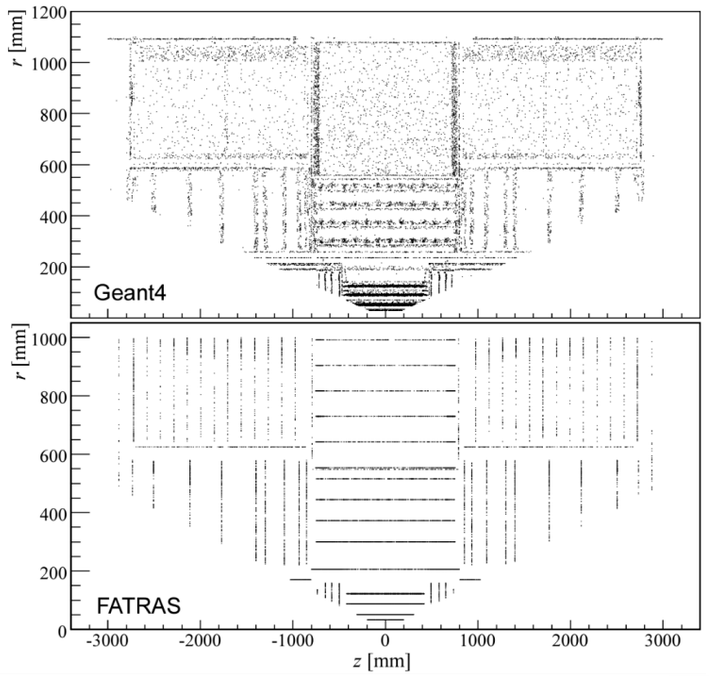
- Calorimeter (AF3): uses parametrized models or ML GAN predictions to simulate the energy deposits in the calorimeters.
- Inner Detector: Combines parametrization and simplified geometry + Track overlay.

+ Scalable performance across diverse computing infrastructures, including traditional GRID and HPC clusters, ensuring efficient resource utilization for fast simulation.

FATRAS

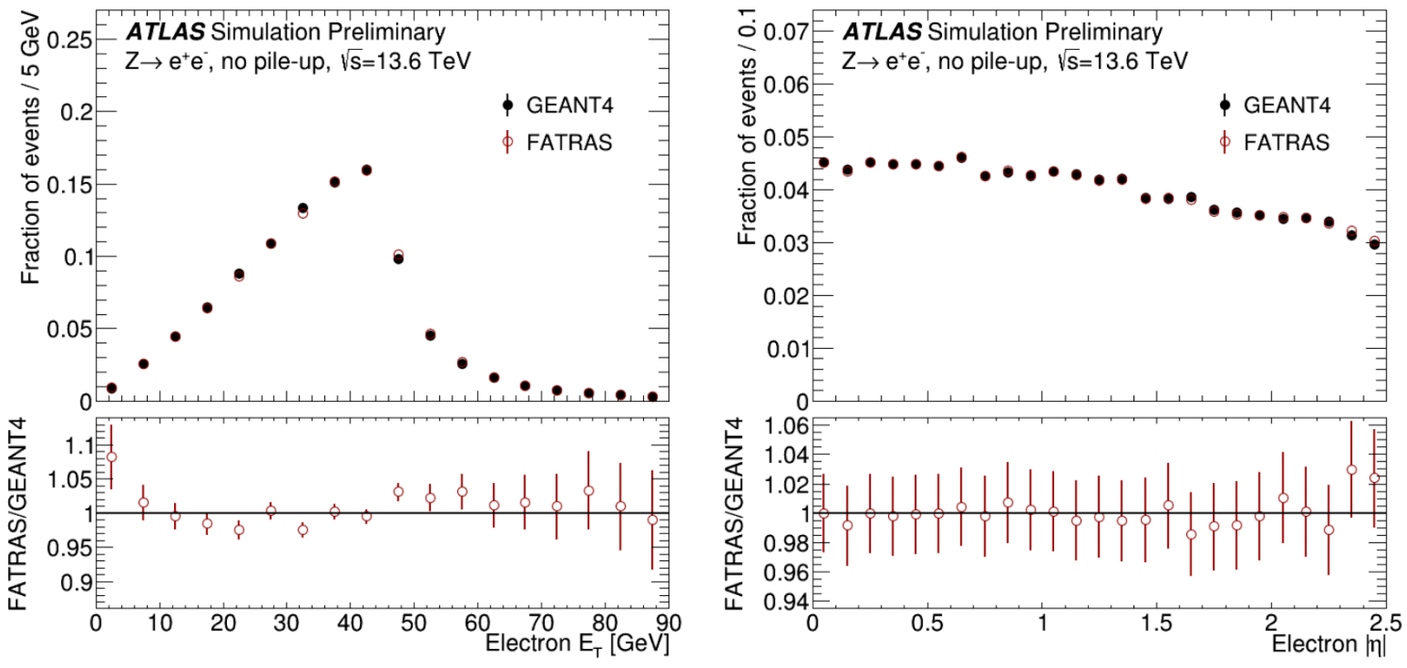
FATRAS

- Fatras, a fast simulation tool, designed to model charged particle interactions in the ID using simplified approaches:
  - Simplified Geometry.
  - Approximation of interactions.
- Fatras currently achieves accuracy within ~10% compared to Geant4 results for EM interactions. We aim to improve the accuracy to around ~1%.



Reference: ATL-COM-SOFT-2008-002

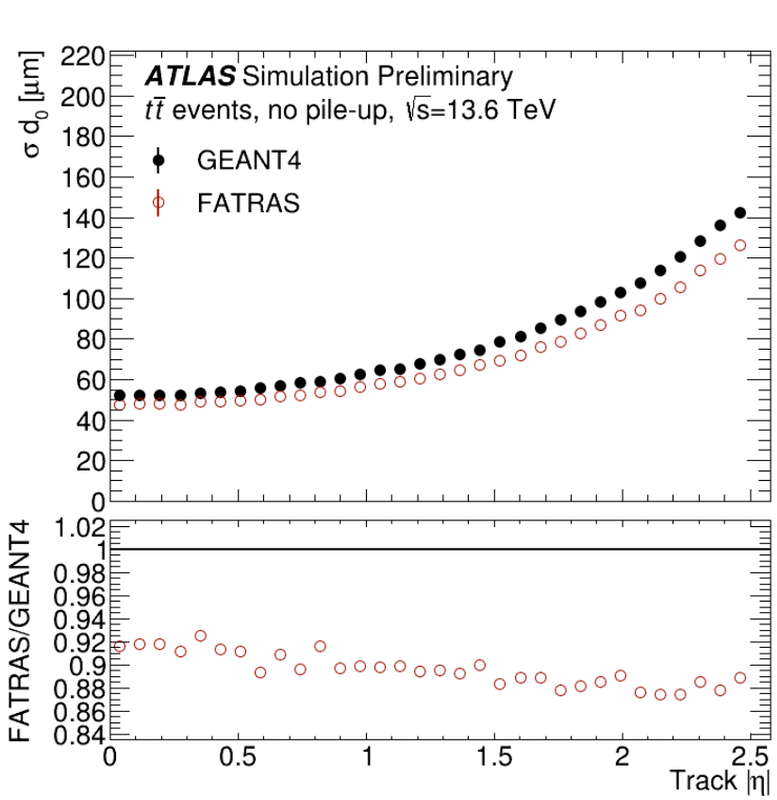
We want to further improve the photon conversion model by fine-tuning the parameters (ongoing).

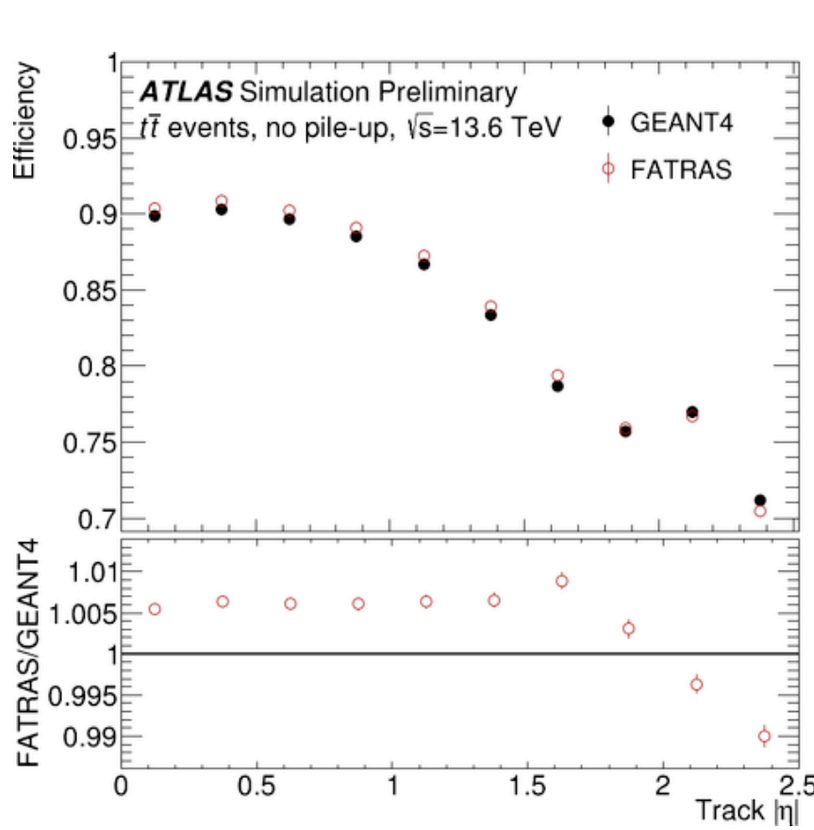


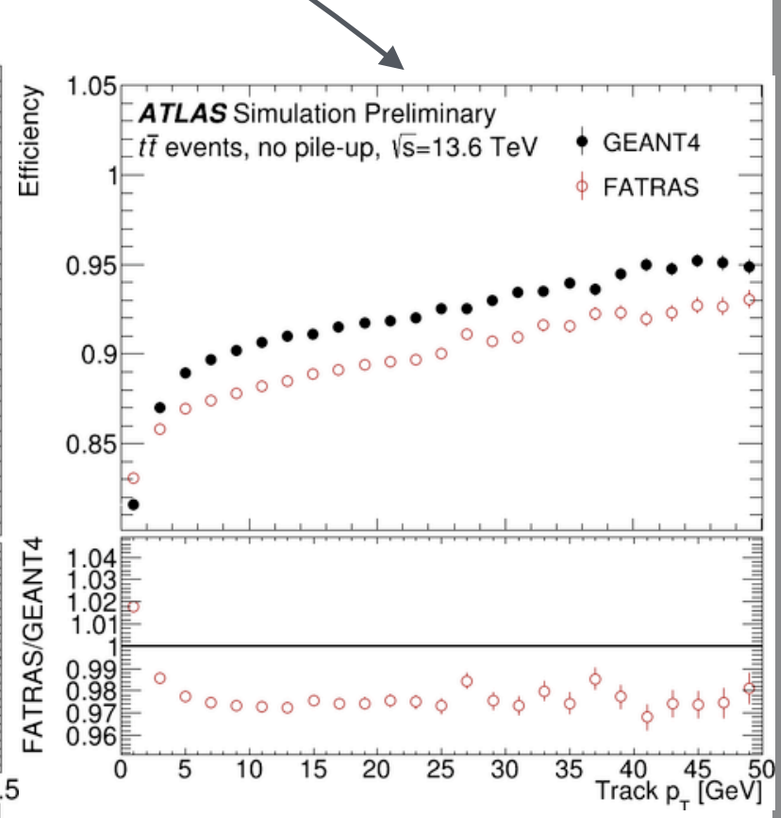
Source: ATLAS PLOTS SIM-2024-002

FATRAS: track reconstruction

- Due to limitations in Fatras' ability to accurately simulate hadronic interactions, we plan to use Geant4 for these processes (ongoing).
- Fatras shows about 10% better than expected resolution in transverse impact parameter, d0. Over-optimism due to mismodeling of rare hadronic interactions.
- While good agreement in track reconstruction across pseudorapidity (agrees with Geant4 within 1%.), there are notable efficiency discrepancies.







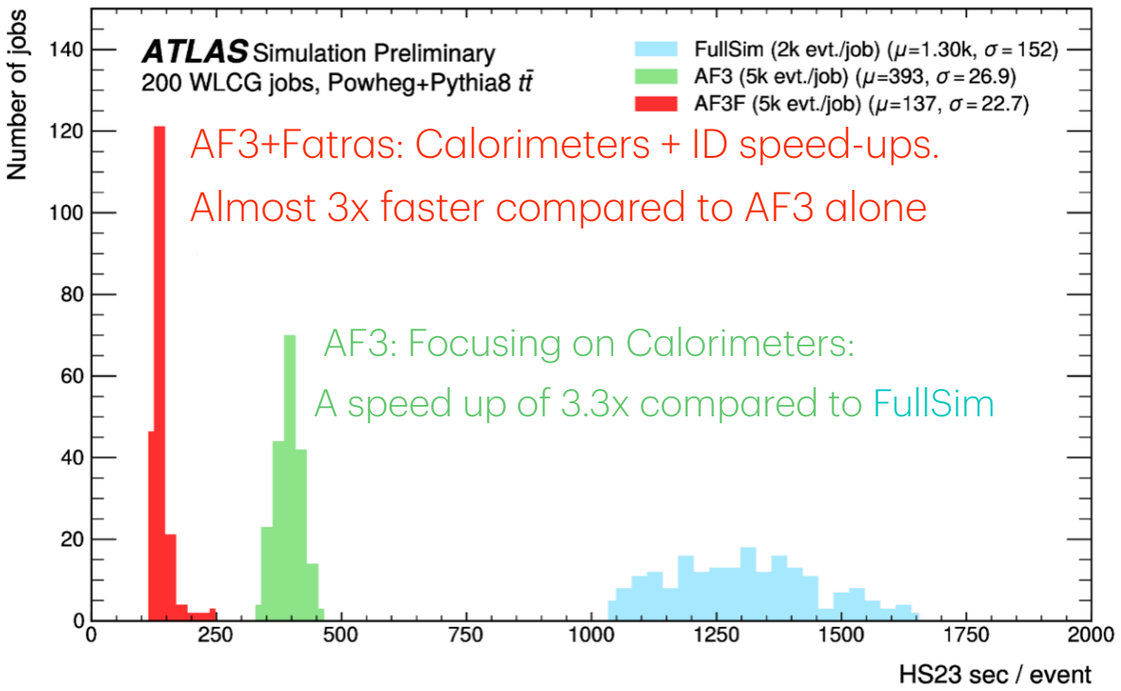
Source: ATLAS PLOTS SIM-2024-002

FastChain on GRID and HPCs

- The performance of Fatras on large-scale computing infrastructures. Our 5 ongoing workflows:

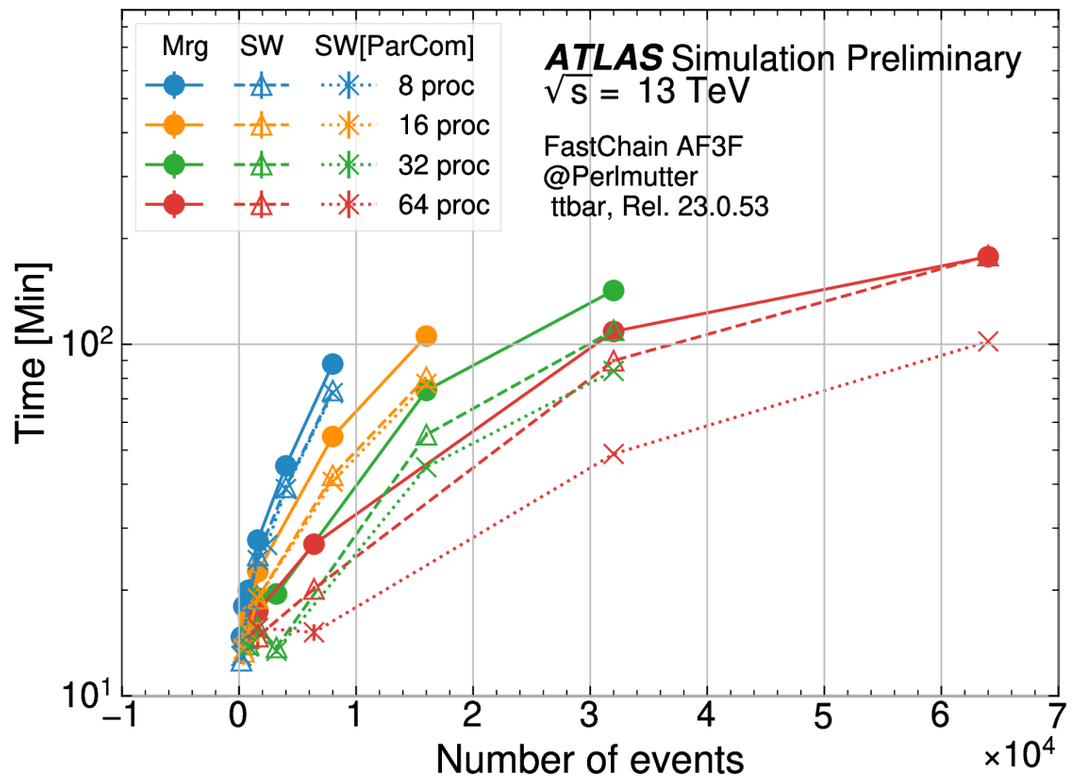
Scenario 1: AF3F + MC Overlay (Sim + Reco + Derivation); Scenario 2: AF3F + MC Overlay (FastChain_tf);
Scenario 3: AF3F + Track Overlay (Sim + Reco + Derivation); Scenario 4: AF3F + Track Overlay (FastChain_tf); Scenario 5: AF3 + Track Overlay (FastChain_tf)

- Process scaling on HPCs: increasing the number of processes (e.g., from 8 to 64) improves the processing speed for large event counts because more events are handled in parallel. However, the writing method (I/O strategy) plays a critical role in ensuring that scaling provides real performance benefits.



Source: ATLAS PLOTS SIMU-2024-006

AthenaMP write modes: Merge(Mrg), Shared Writer (SW), Shared Writer + parallelCompression (SW[ParCom])



More performance plots: ATLAS PLOTS SIMU-2024-07

Fatras components and ongoing enhancements

- Expect comprehensive results on the accuracy of physics models in 2025! Stay tuned for future performance results.

Testing for accuracy in physics models through observables in H→4l analysis and photon studies.

Geometry conversion into ACTS format required

Enhancing accuracy of physics modules, e.g. photon conversions

Fatras

Porting classes from Fatras.

Thread safe

ACTS Track reconstuction ToolKit

ACTsFatras

Simplified geometry & interaction: e.g. TRT & ITk geometry description

Technology

Geant4

A simplified tracking geometry (R&D)

Simulates Hadronic interaction

Our plan is to focus on Run 4 geometry. The most significant difference in Run 4 is the new Inner Tracker (ITk).

Track Overlay

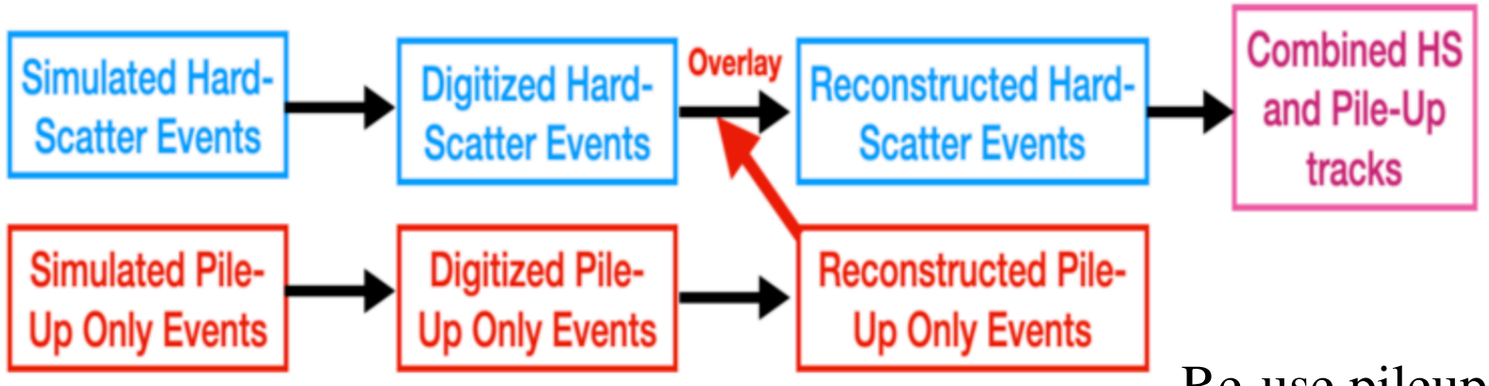
Overview of track overlay: simulation with pile-up

- Goal: Optimizing computational efficiency in reconstruction while maintaining the physics performance.
- Reusing reconstructed pile-up tracks

  - This skips several time-consuming steps (e.g. clustering and track finding for pile-up events).

Schematic of the MC Overlay process (ATLAS default)

Schematic of the Track Overlay process

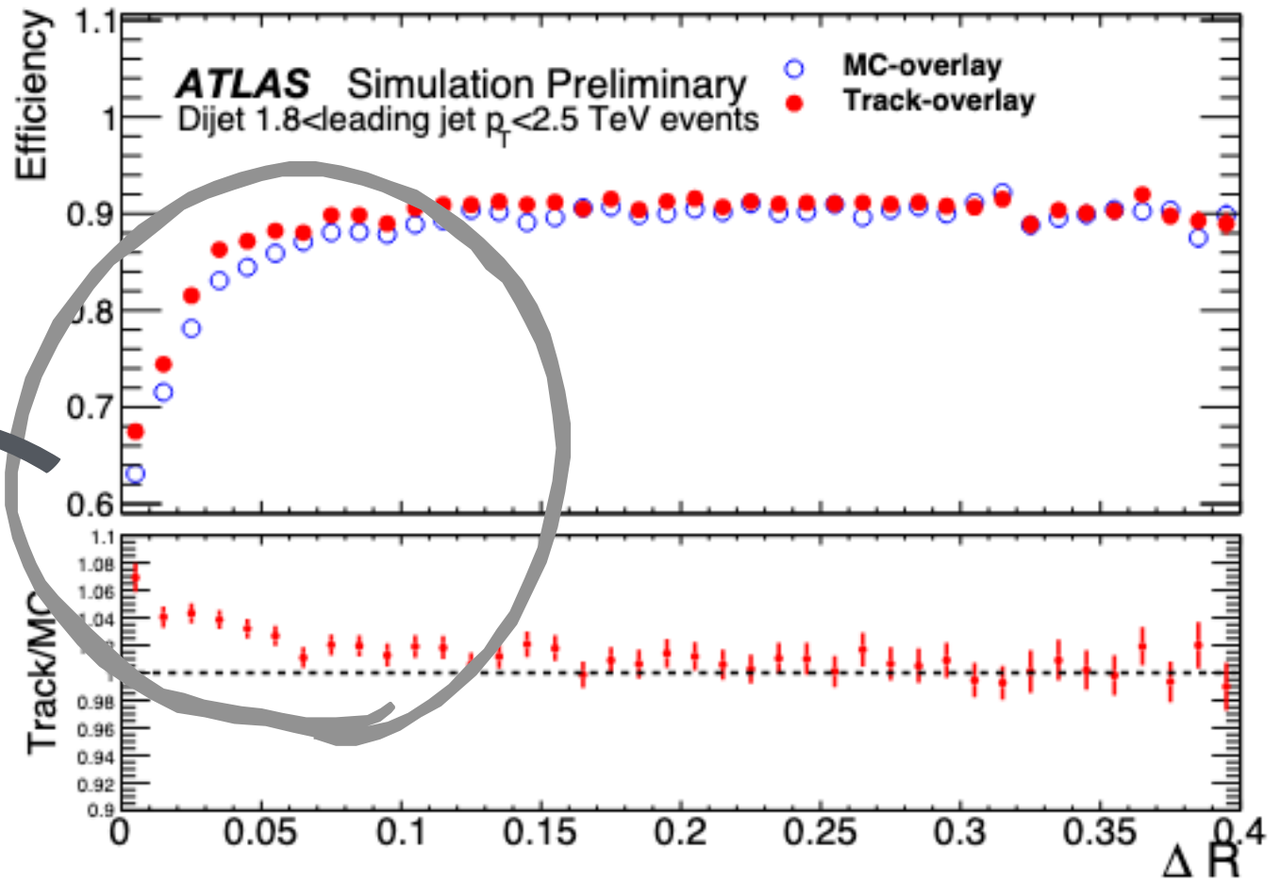


Re-use pileup tracks!

Overview of track overlay: simulation with pile-up

- The limitation: Track overlay is not suitable for all event types:
  - Dense track environment: track misinterpretation or hit sharing is common when multiple tracks overlap in high density area, complicating pattern recognition.
  - Pile-up events introduce additional noise, making it difficult to separate true hard-scatter from pile-up tracks, especially inside jets.

These high-pT jets are composed of multiple particles, resulting in a dense environment with many overlapping tracks.

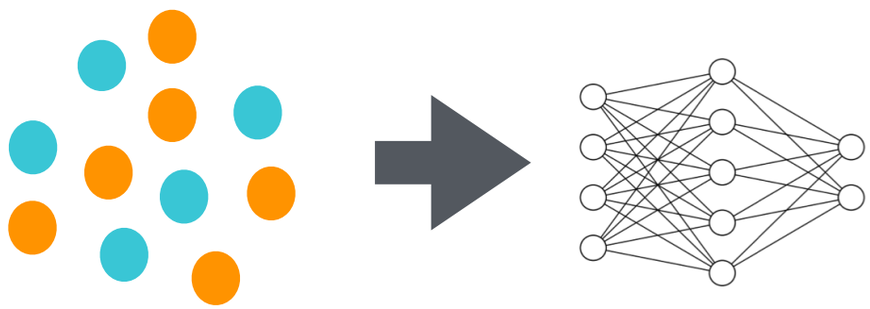


Source: EPJ Web of Conferences 295, 03014 (2024)

DNN ML model training

- We developed a NN to identify unsuitable events.
  The NN is used to determine, on an event-by-event basis, whether track overlay or MC overlay should be applied.
- Step 1: Network classification
  - Features: truth information including kinematics of generator-level particles, event topology (i.e. local track density), Pile-up information.

Diff label
No Diff label



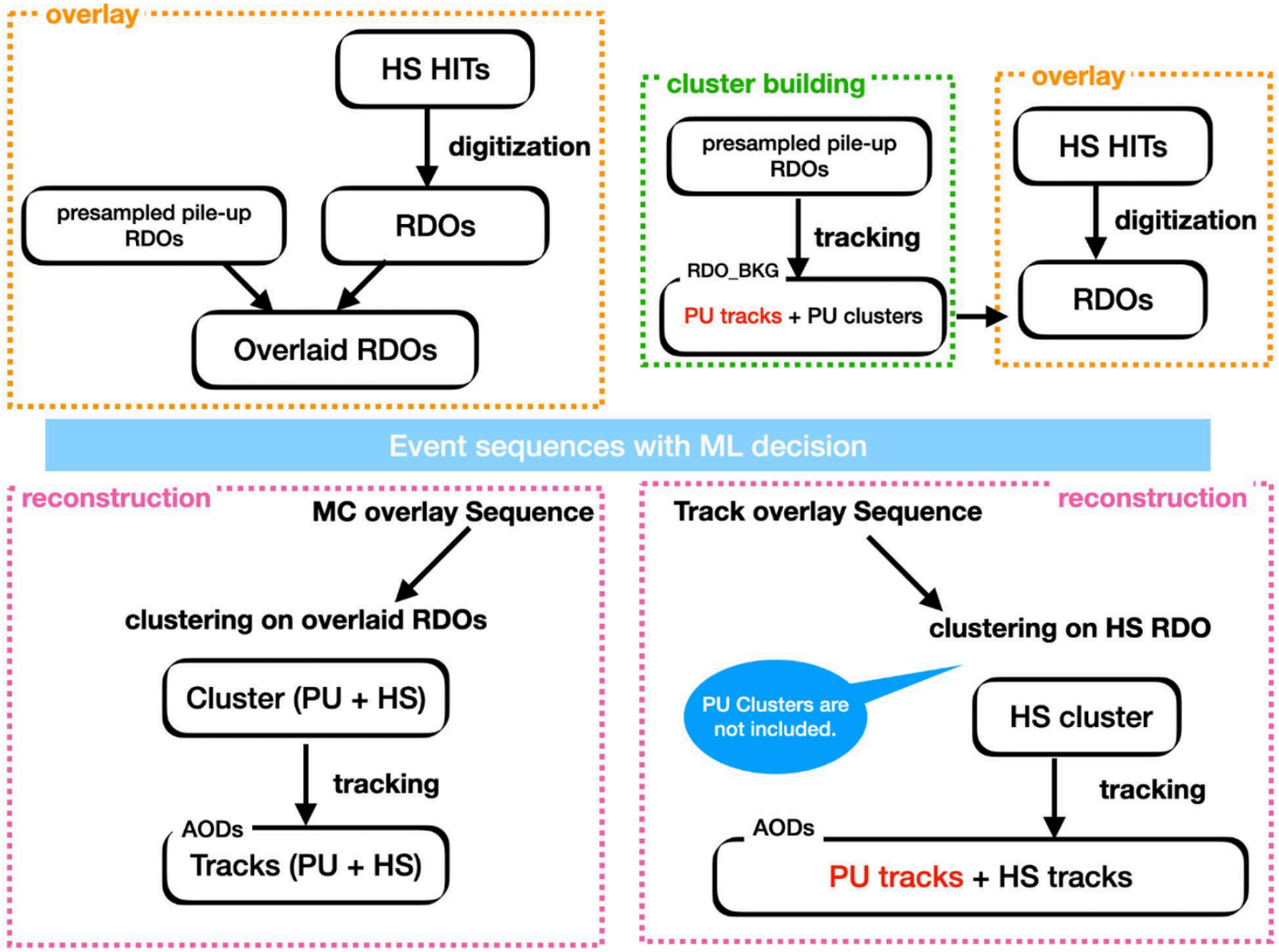
A track is labeled as "Diff" if the track is truth-matched in Track Overlay but fails to match in MC Overlay.
Other tracks are assigned a "no Diff" label

The output of the NN: the probability of a track receiving a Diff label.

- Step 2: Identifying "Bad" tracks
  - A threshold is applied to determine which tracks are considered as "Bad" tracks. The bad tracks are associated with events that have differences between the Track overlay and MC overlay.
  - If an event contains a single "bad" track, it is redirected to MC overlay.

DNN ML model training

- We integrate the trained NN model into Athena. The model predicts whether each event should be reconstructed using track overlay or redirected to MC overlay.



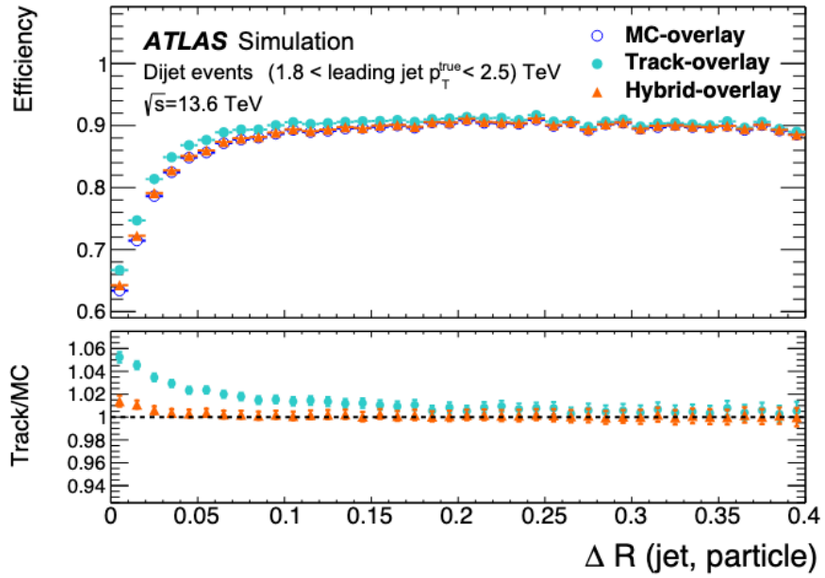
MC- and Track overlay work the same in the muon spectrometer and calorimeter.

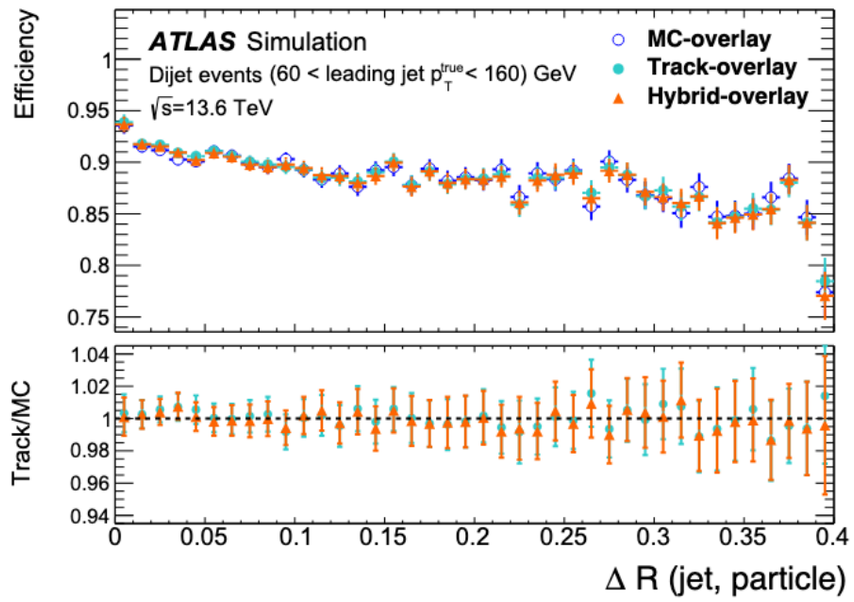
On the fly decision per event!

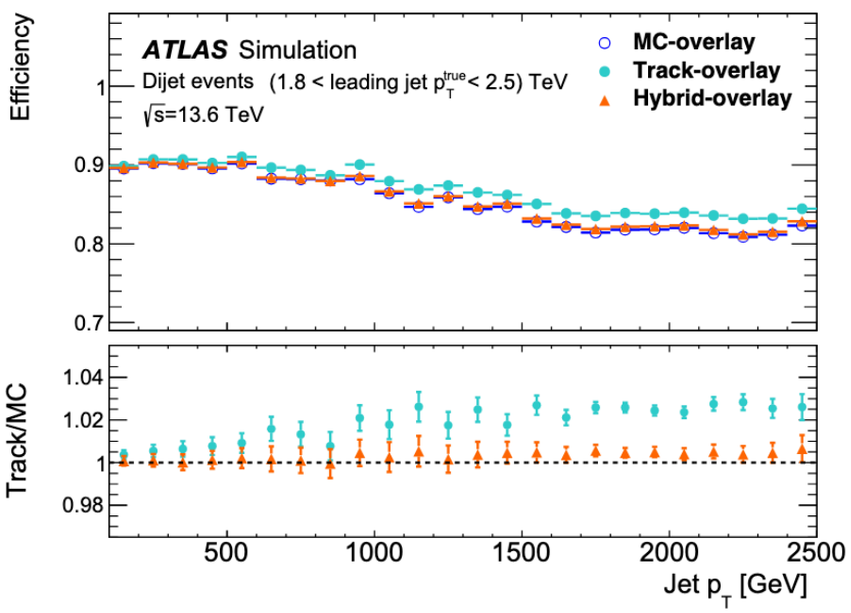
The result of this integration is our Hybrid overlay.

Accuracy: Hybrid v.s. standard workflows

- Evaluating the accuracy of hybrid compared to MC overlay in numerous physics process: QCD multi-jet events, top quark pairs, W'→ Wh → lvbb (for $m_{W'}$=3 TeV).







Event fractions for track overlay
Dijets (high pT): 35.3% (3530/10000)
Dijets (low pT): 93.5% (9350/10000)
ttbar:90.15% (9015/10000)

Source: 2404.06335

More performance plots: ATLAS PLOTS SIM-2024-001

Speed analysis

- Assessing the speed improvements with track overlay in terms of CPU usage with <mu> ~ 60 pileup file. The evaluation has been tested on the ttbar process.

Configuration	Overlay	Reconstruction
MC overlay	2.21s (file size: 4 GB)	5.01s
Track overlay	3.00s (file size: 8.8 GB)	2.84s
Track overlay Less compression algorithm	2.18s (file size: 13 GB)	2.66s

(Plan to run multiple instances to get a reasonable average.)

- The track overlay reduces the reconstruction CPU usage by approximately 45%.
- Eased the compression setting to speed up processing.
  - Trade-off: file size increased from 8.8 GB to 13 GB.
  - This is acceptable because the files are intermediate and the performance gain outweighs the cost of larger file sizes.

Conclusions

- The Fast Chain has demonstrated a notable reduction in wall-time and CPU usage (for Run 3):

  AF3: speed-ups ranging from 3 to 15 times compared to fullSim.

  Fatras: a factor of ~3 compared to AF3.

  Track overlay: ~45% faster in reconstruction relative to standard MC overlay approach.

- Ongoing work:

- Improving the accuracy of the Fatras model, especially for challenging physics observables like photon conversions.
- Integrating ACTS-based geometry and material maps.
- Exploring a simplified tracking geometry used in Geant4.
- Developing the Track overlay workflow for Run 4+.
- Setting up scalable workflows on both GRID and HPC clusters.

- Once the Fast Chain is optimized to be efficient and accurate enough, we aim to avoid generating and storing intermediate files. Re-simulating everything in a single streamlined process would significantly decrease storage requirements.

Backups