The Fast Simulation Program of ATLAS at the LHC

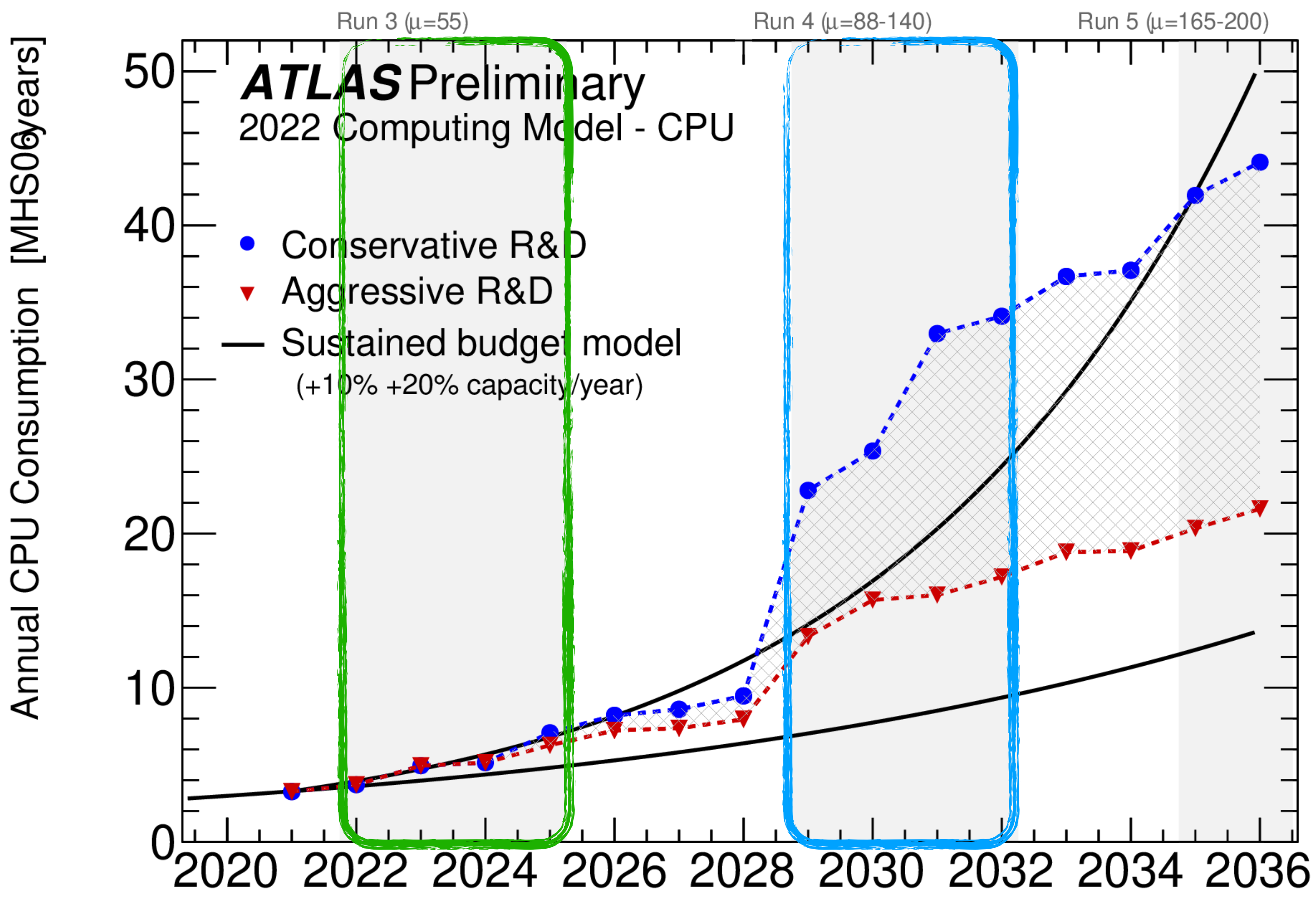
Martina Javurkova

(University of Massachusetts-Amherst)

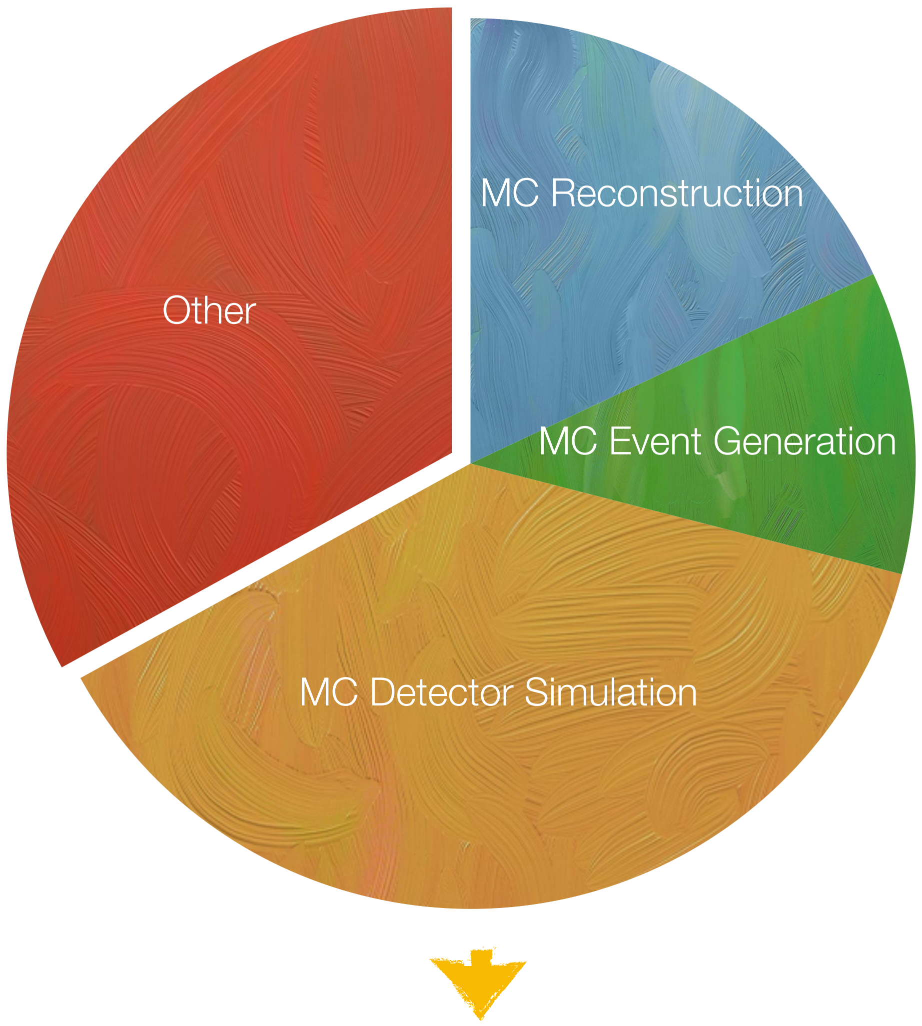
On behalf of the ATLAS collaboration

Why Fast Simulation?

- Not enough computing resources to produce sufficient samples with detailed simulation
- **MC production** takes ~**70%** of the GRID CPU time in ATLAS
  - Dominated by **MC full detector simulation** done with `Geant4`



Projected evolution of **compute** usage



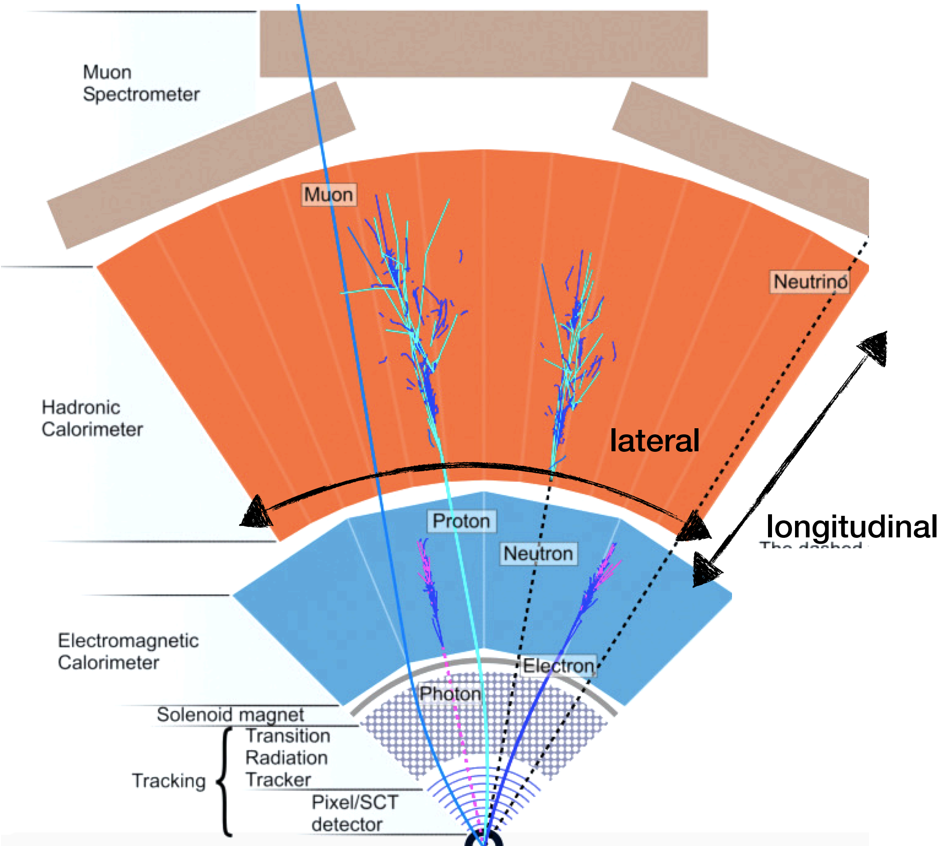
Energy deposits for each sub-detector (*HITS*)

Sources: ATLAS S&C HL-LHC Roadmap, CERN-LHCC-2020-015

Fast Simulation in **Run 3**

Fast Calorimeter Simulation

- The most CPU-intensive component of the `Geant4` simulation is the calorimeter shower simulation
  - ~80% of the total simulation time
- ATLAS developed a fast simulation tool called **AtlFast3**
  - To replace the slow propagation and interactions of incident particles inside the calorimeter volume with the direct generation of *energy deposits* based on an underlying parametrisation
  - Introduced for Run 2 [1] and further improved for Run 3 [2]
  - Uses a **simplified geometry** of the calorimeter cells to simplify its complex and non-homogeneous structure
  - Employs **two techniques** (complementary in different parts of detector)
    - `FastCaloSim`: parametric approach for the shower development
    - `FastCaloGAN`: generative adversarial networks (`GANs`)
  - The goal is to reproduce the `Geant4` full simulation in the calorimeters
    - Speed-up + high-accuracy

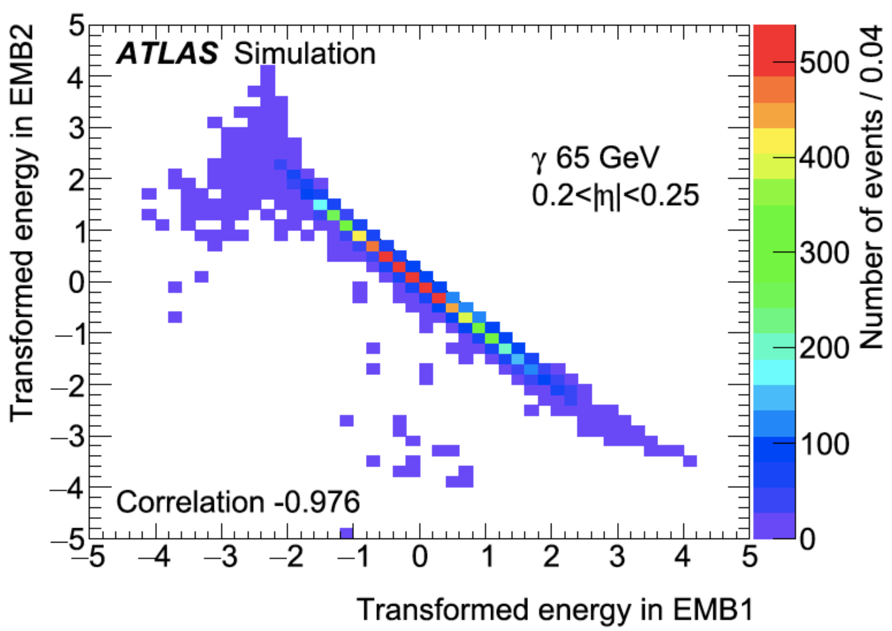


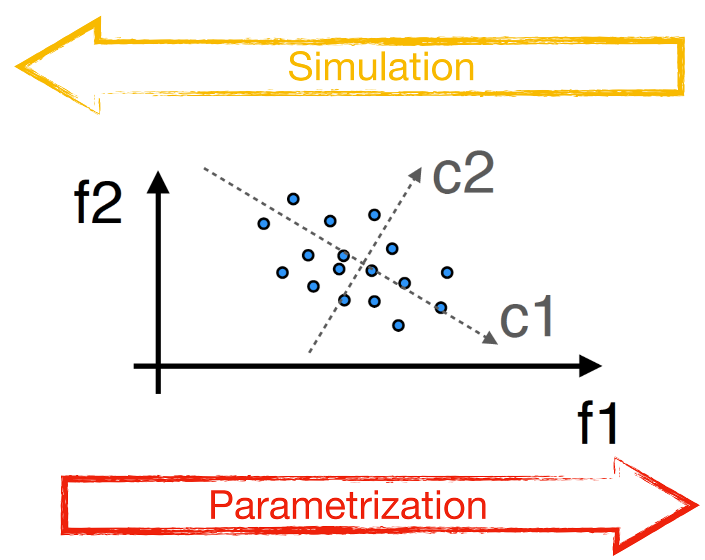
Sources: [1] COMPUT SOFTW BIG SCI 6, 7 (2022), [2] arxiv:2404.06335, ATLAS detector

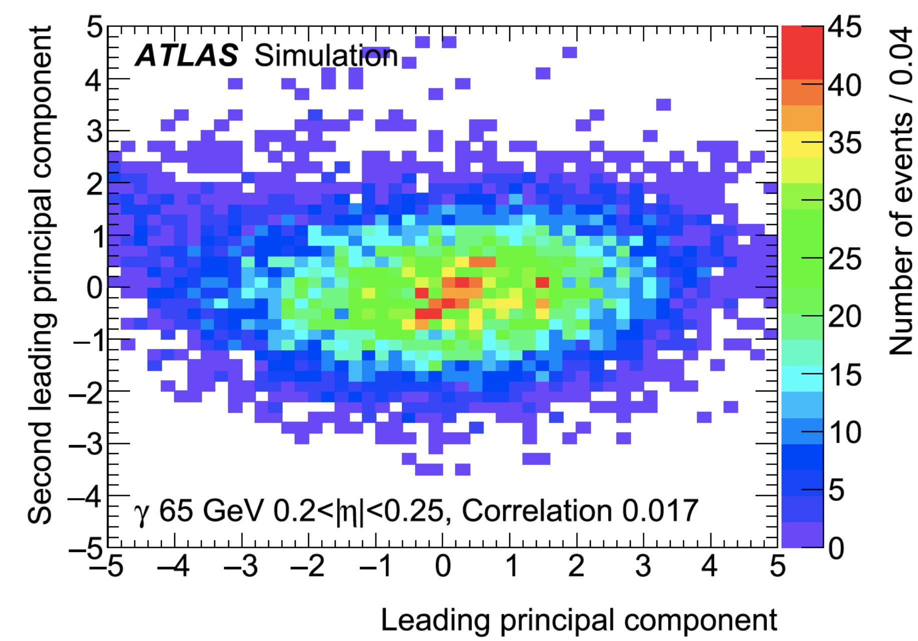
AtlFast3 strategy

Parameterisation used for `FastCaloSim` and **training** of the `GANs` are based on **single particles simulated with `Geant4`**

- Various particle types — *photons* and *electrons* for electromagnetic showers
  - *pions* for hadronic showers (in the case of `FastCaloGAN` also *protons*)
- Fine bins of η — 100 linearly spaced bins to provide coverage up to $|\eta| = 5$
- Different energy ranges — 17 logarithmically spaced energy bins
- `FastCaloSim`
  - Separate parametrisation of *longitudinal* and *lateral* shower development
  - Energy deposited in layers are decorrelated using **Principal Component Analysis** (PCA)
  - Average lateral shower profile is parametrised as 2D probability density functions







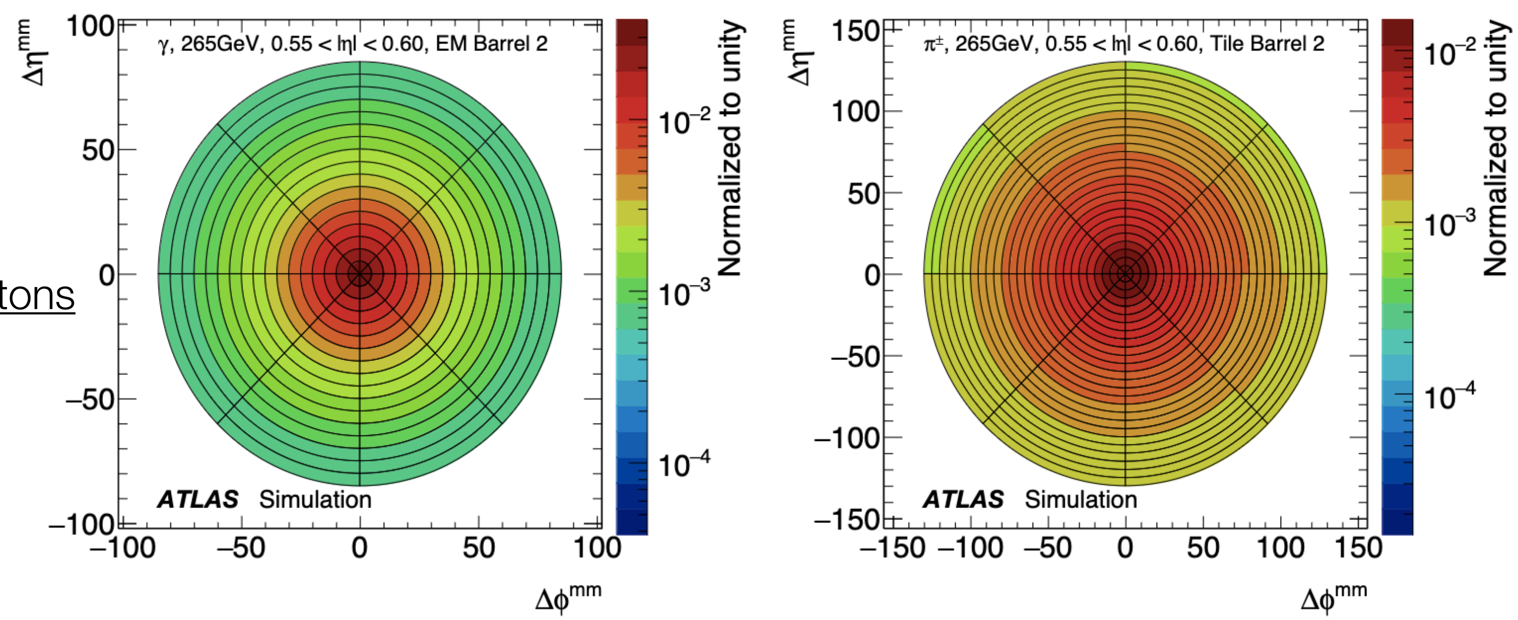
Sources: COMPUT SOFTW BIG SCI 6, 7 (2022), Rui Zhang

AtlFast3 strategy

Parameterisation used for `FastCaloSim` and **training** of the `GANs` are based on **single particles simulated with `Geant4`**

- Various particle types - *photons* and *electrons* for electromagnetic showers
  - *pions* for hadronic showers (in the case of `FastCaloGAN` also *protons*)
- Fine bins of η - 100 linearly spaced bins to provide coverage up to $|\eta| = 5$
- Different energy ranges - 17 logarithmically spaced energy bins
- `FastCaloSim`
  - Separate parametrisation of *longitudinal* and *lateral* shower development
  - Energy deposited in layers are decorrelated using Principal Component Analysis (PCA)
  - Average lateral shower profile is parametrised as **2D probability density functions**

photons

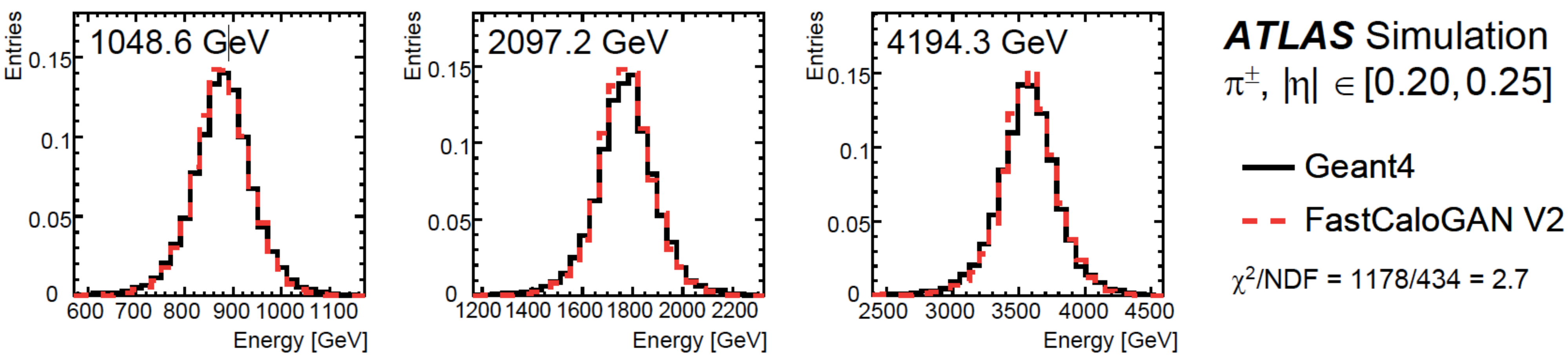


pions

Source: COMPUT SOFTW BIG SCI 6, 7 (2022)

AtlFast3 strategy

- FastCaloGAN
  - One GAN trained for each particle type and $|\eta|$ bins, and conditioned on the true momentum: **600** GANs
  - Trained to reproduce energy deposited in *voxels*, *layers* as well as *total energy* in the calorimeter in a single step
    - Calorimeter *HITS* are grouped into **voxels**
      - Their granularity is optimised and finer than that of the calorimeter cells, which improves the modelling
  - Architecture of the networks and the hyperparameters were optimised



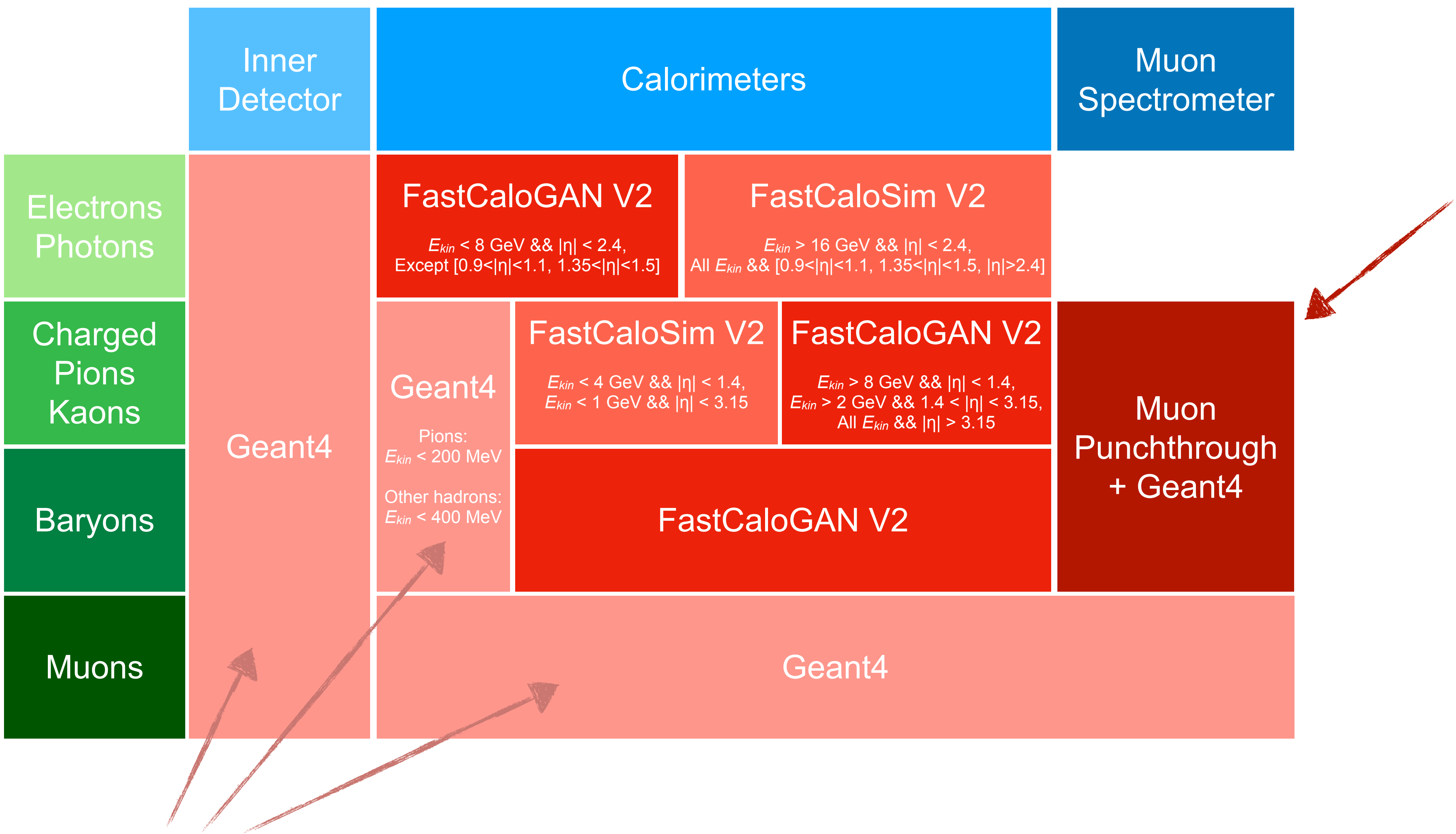
During the simulation, *HITS* are produced in the calorimeter based on the chosen technique given particle type and energy

- Additional corrections are then applied to match the accuracy of Geant4

Source: arxiv:2404.06335

AtlFast3 configuration in Run 3

- **AtlFast3** combines the strengths of the `FastCaloSim` and `FastCaloGAN` approaches by selecting the most appropriate algorithm depending on the properties of the shower-initiating particles



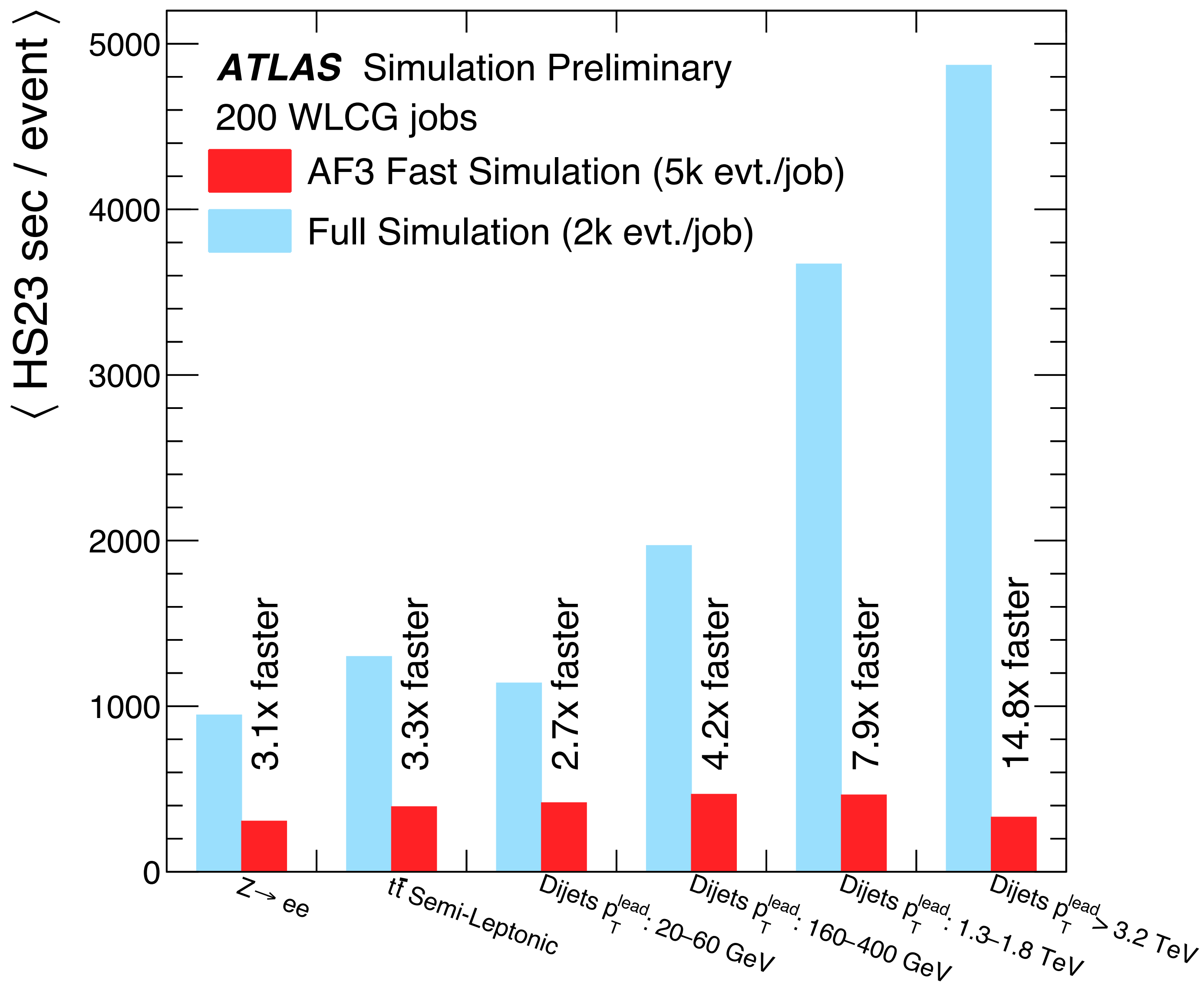
`Muon punch-through` (the spray of particles into the muon spectrometer resulting from late interacting high-energy hadrons) is modelled with a tool based on DNNs

`Geant4` is used to simulate all particles in the Inner Detector, low energy hadrons in the calorimeters and muons

Source: arxiv:2404.06335

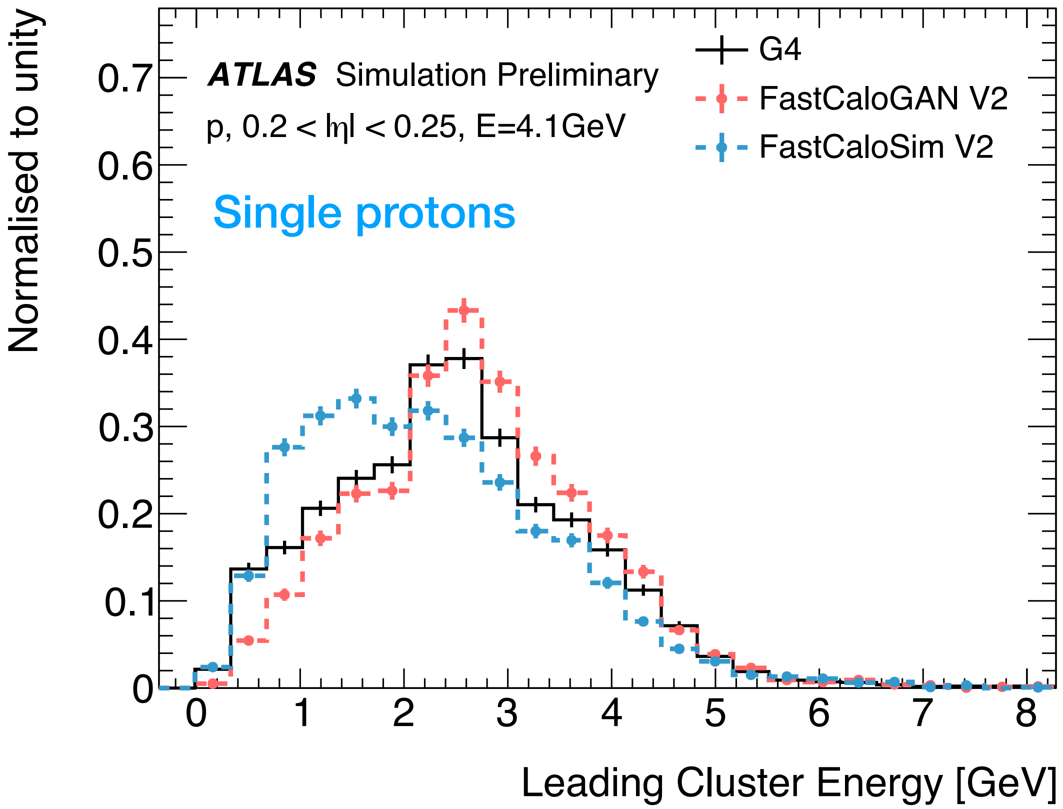
AtlFast3 performance: speed-up

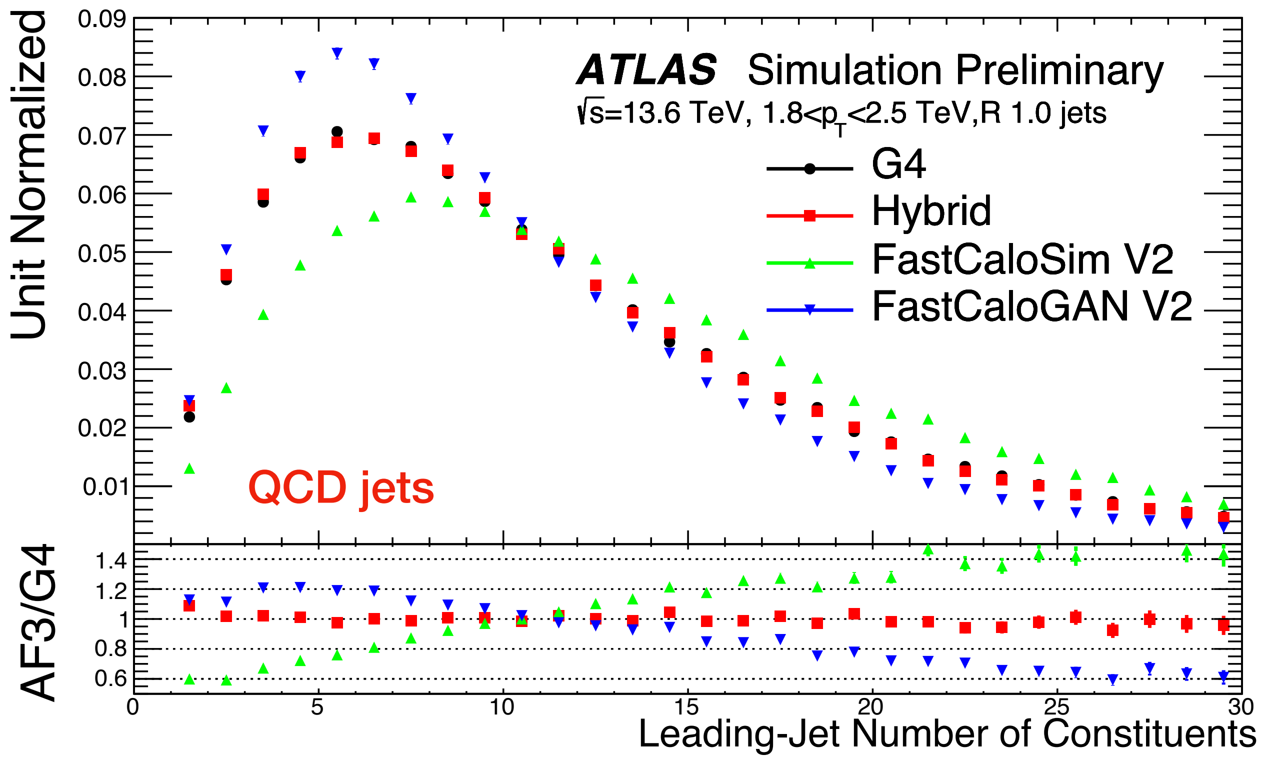
- **AtlFast3** is 3 (for $Z \to ee$ events) to 15 (for high-p_{T} di-jet events) times faster than `Geant4` simulation of the ATLAS Run 3 detector covering 2023 data taking (corresponding to **mc23c** production campaign)

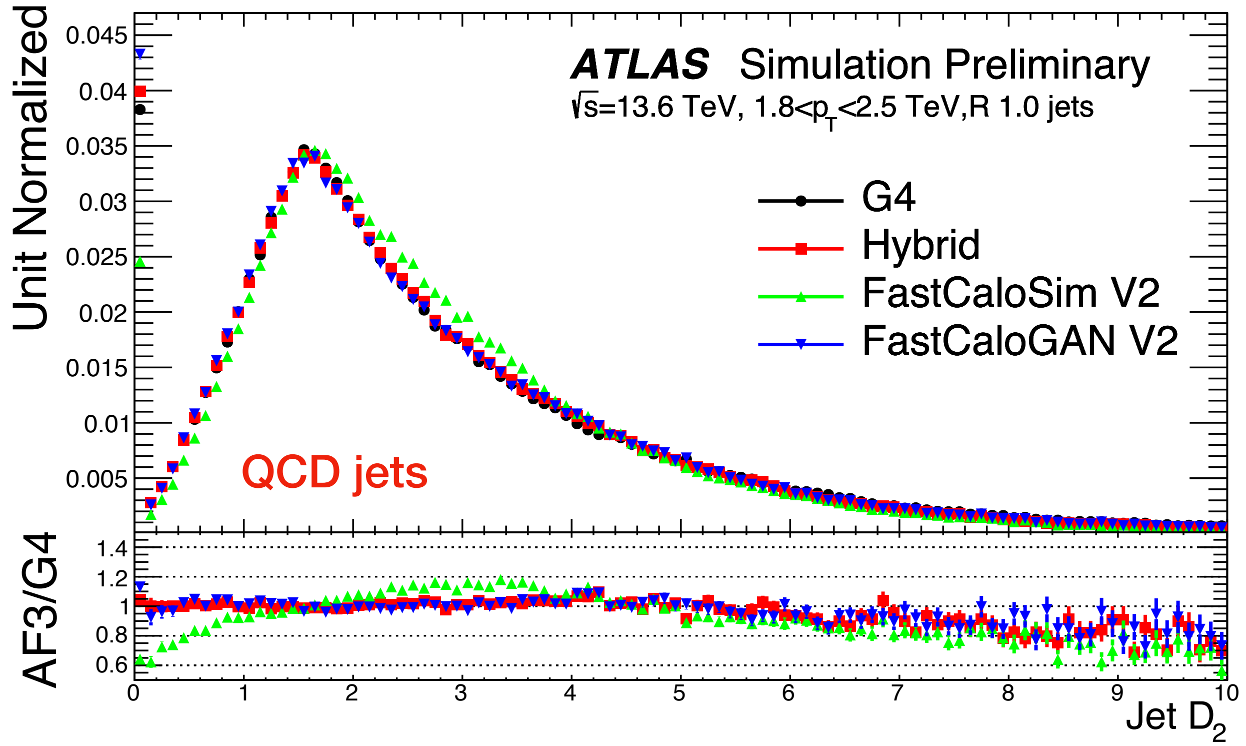


Sources: SIM-2023-005

AtlFast3 performance: accuracy







- Very accurate modelling of the leading cluster energy for **protons** with `FastCaloGAN` where dedicated `GAN` was trained on protons as well as of the number of constituents for the leading jet and D2 variable in **di-jet** events with the hybrid approach

- For most observables used in physics analyses, **AtlFast3** and `Geant4` agree within <u>a few percent</u>

  - **AtlFast3** can be used for almost every analysis (not only signals but also backgrounds)

Sources: SIM-2023-004

Fast Simulation towards **HL-LHC**

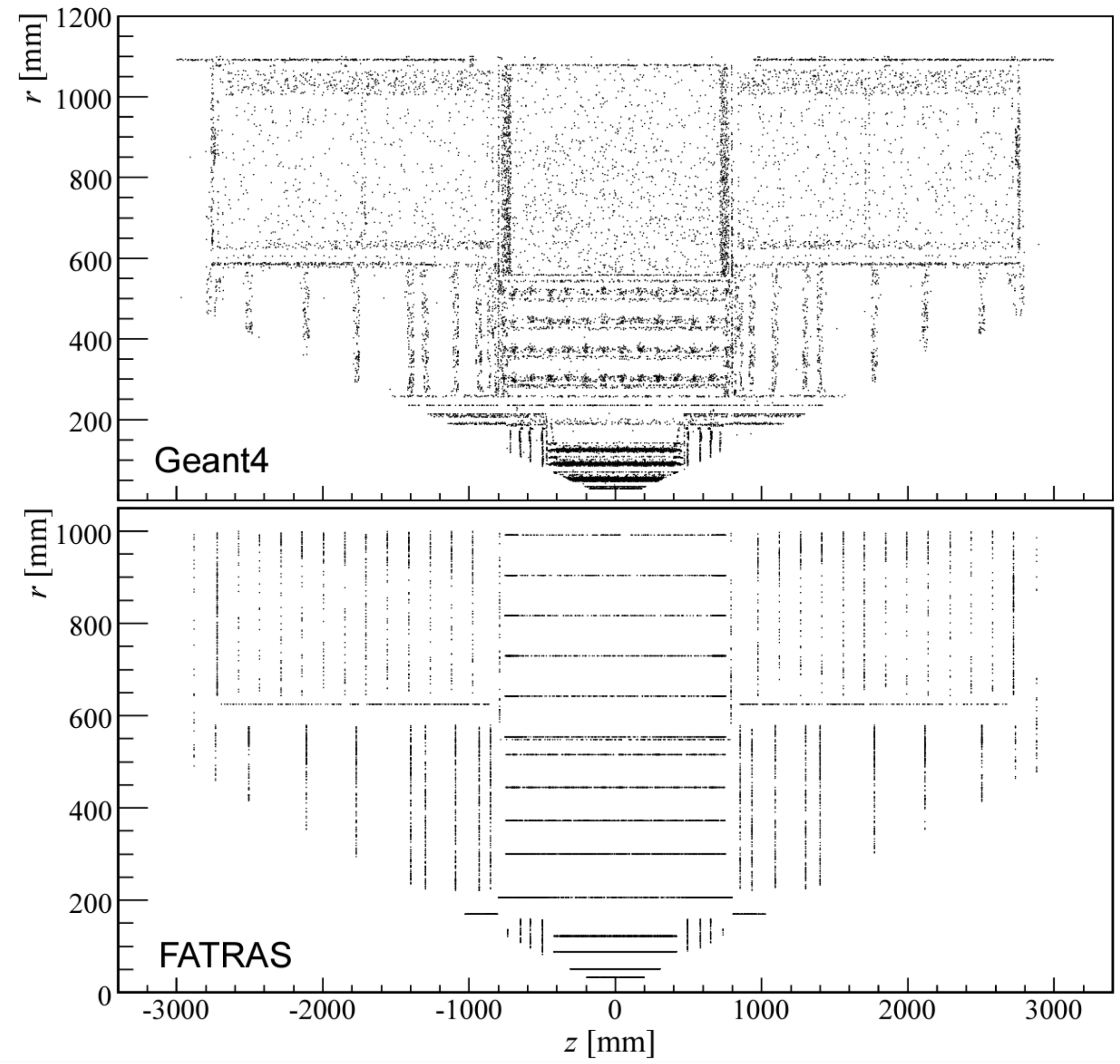
Fast Track Simulation

- When **AtlFast3** is used, most time is spent on the simulation of the inner detector with `Geant4`

- Fast ATLAS Track Simulation (**FATRAS**) utilizes

- **Simplified detector geometry**
    - Volume material properties projected to layer surfaces

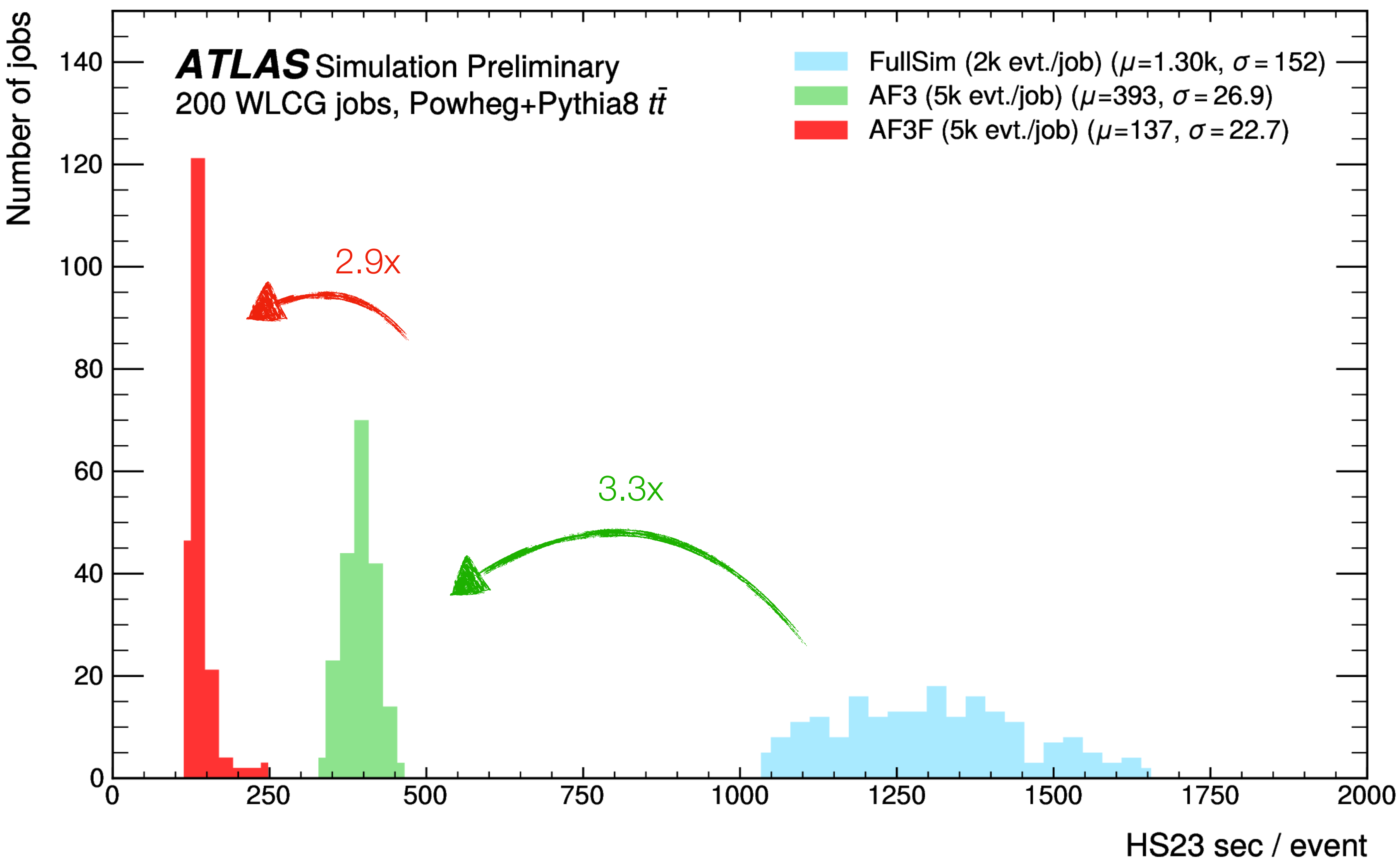
- **Fast algorithms** to parameterise material effects
    - Ionization loss: Bethe-Bloch formula
    - Radiation loss: Bethe-Heitler formula
    - Multiple coulomb scattering: Gaussian mixture model
    - Hadronic interactions: param. from simulated `Geant4` events



Source: ATL-SOFT-PUB-2008-001

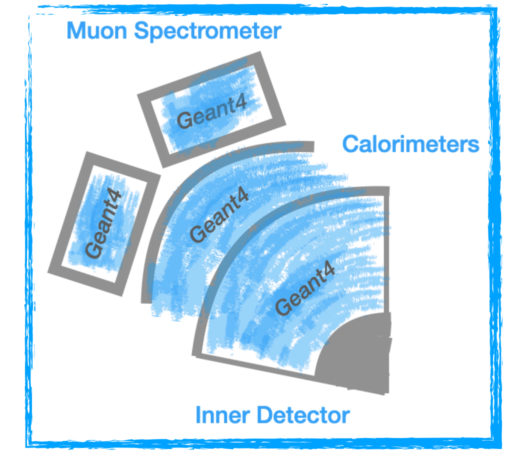
FATRAS integration and speed-up

- Integration into **ACTS** (A Common Tracking Software) ongoing
  - ACTS is an experiment-independent software package [1], [2]
  - **Thread-safe** fast simulation of the inner detector

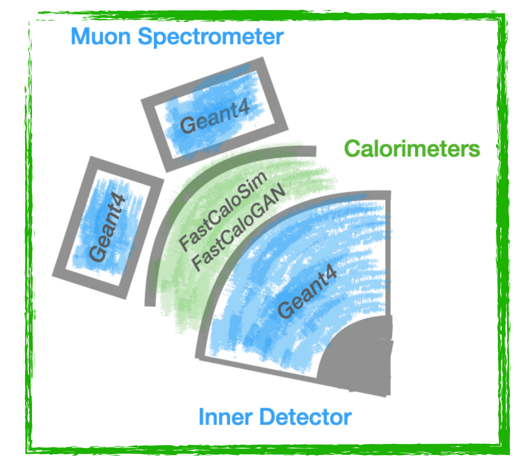


- **AF3F** is almost 3 times faster (for $t\bar{t}$ semi-leptonic events) than **AF3**

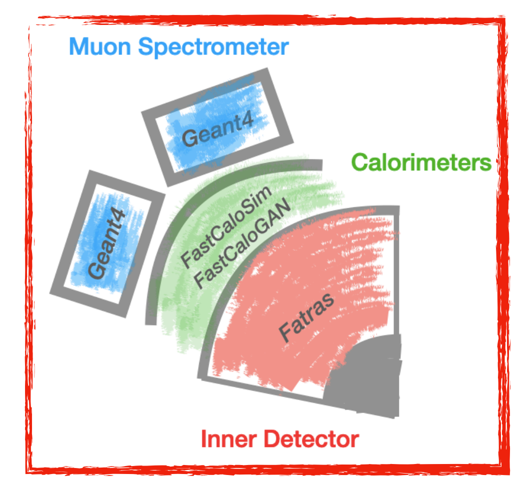
FullSim



AF3

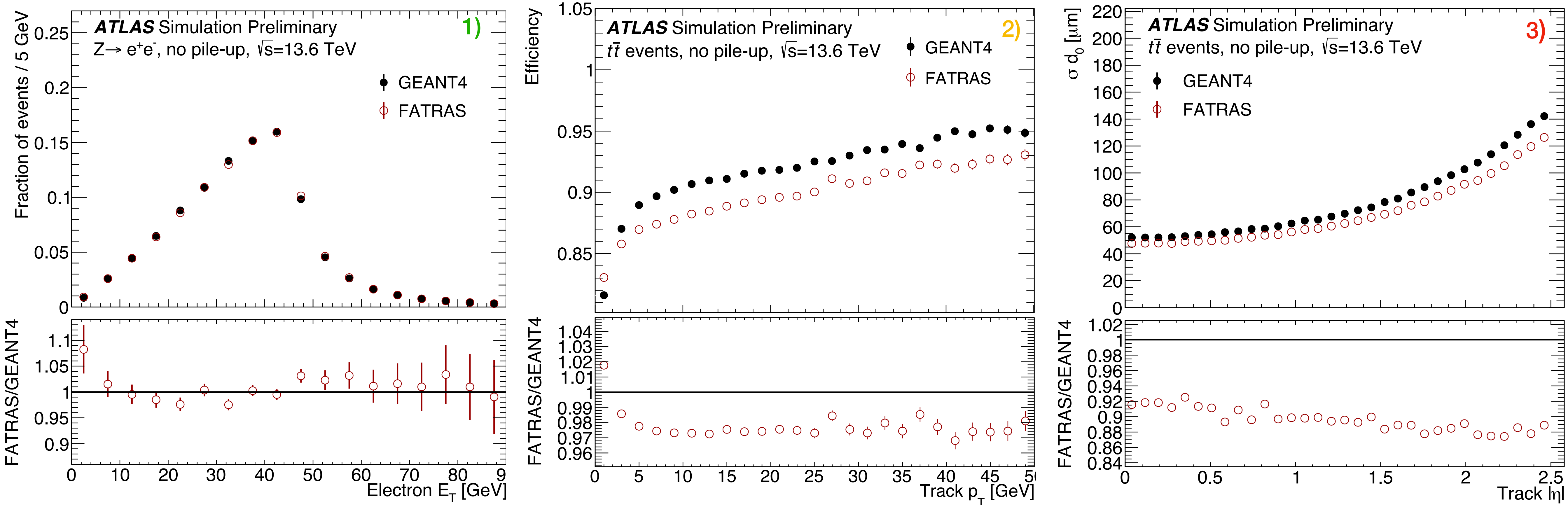


AF3F = AF3 + FATRAS



Sources: SIMU-2024-06, [1] ATL-SOFT-PROC-2017-030, [2] ACTS documentation

FATRAS performance



1) **EM physics**: photon conversion model needs to be improved

2) **Track reconstruction efficiency**: within 3% vs pT (too high @ low pT / too low @ pT>2GeV)

3) **Transverse impact parameter (d0) resolution**: 10% @ large $|\eta|$ (due to FATRAS's inability to simulate rare hadronic interactions)

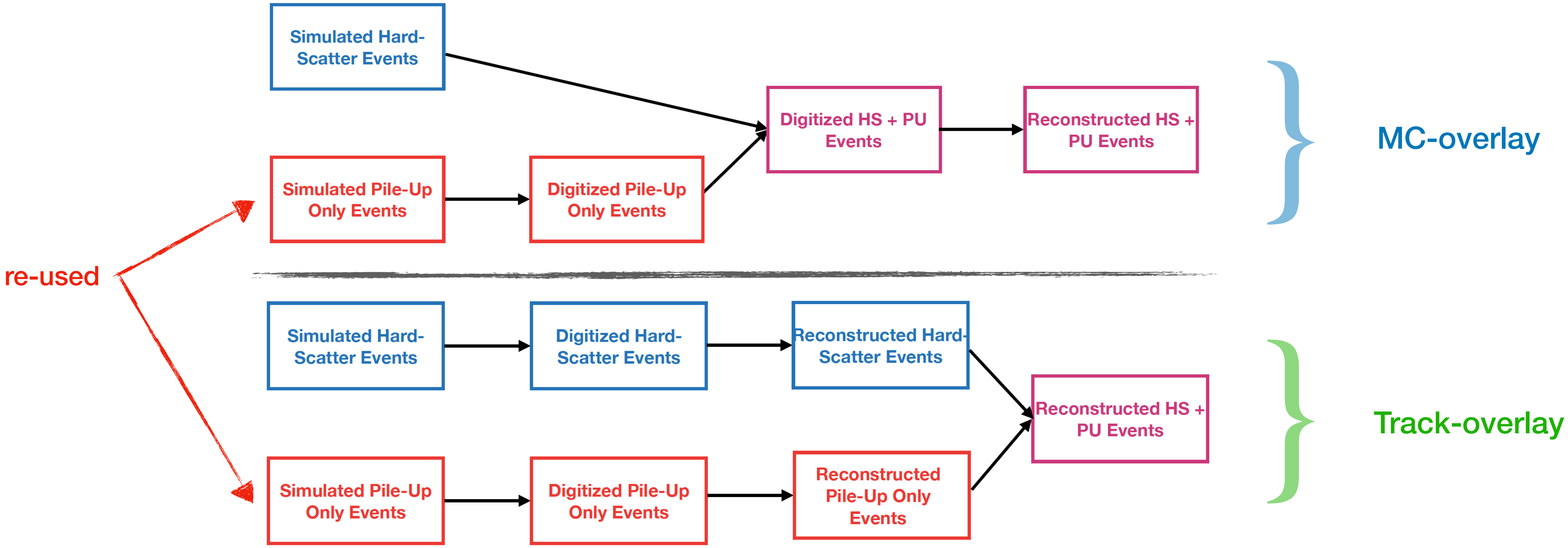
- **FATRAS** reproduces `Geant4` with ~10% accuracy
  - Ongoing work to improve the physics modelling performance to be within ~1% of the `Geant4`

Sources: SIM-2024-002

Fast Reconstruction

- **Most CPU-consuming reconstruction algorithm** is the inner detector track reconstruction, which identifies track candidates from tracking detector *HITS*, as it is greatly slowed down by the presence of pile-up (PU) collisions

- **ATLAS PU model**
  - **MC-overlay**: PU collisions simulated, digitised and overlayed onto hard-scatter (HS) events during digitisation (standard)
  - **Track-overlay**: PU collisions simulated, digitised, reconstructed and overlayed onto HS events during reconstruction



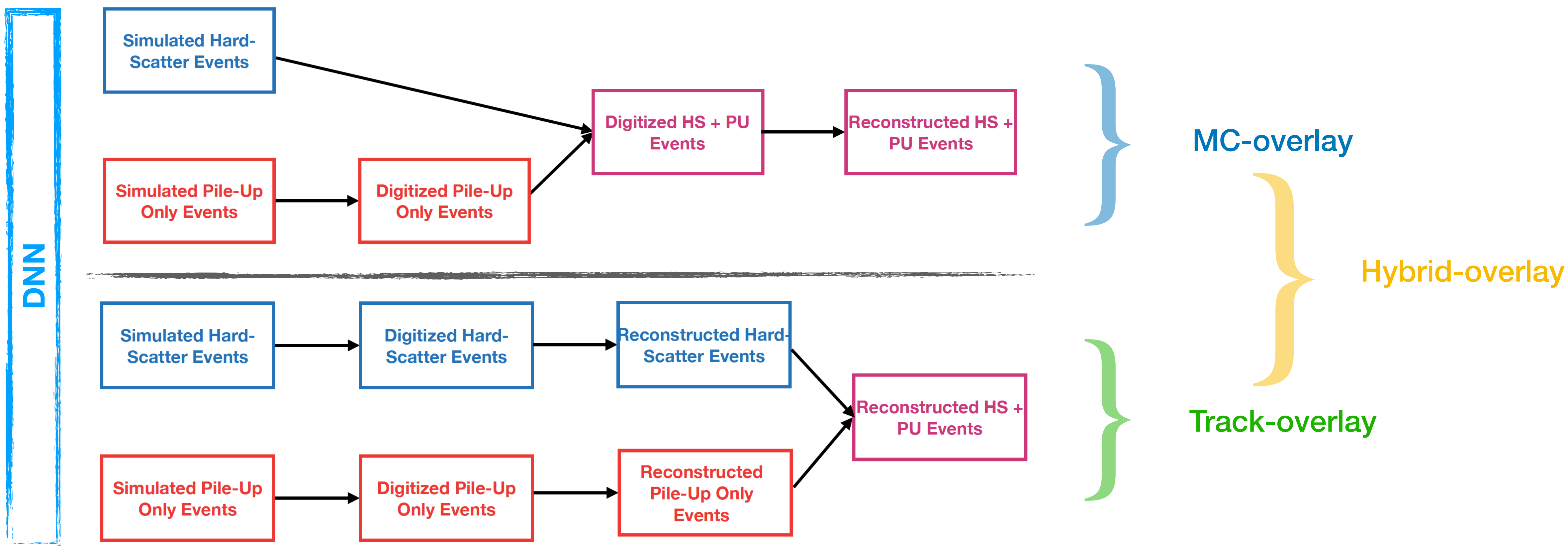
Glossary

- Digitisation: simulation of detector electronics and read-out
- Reconstruction: used to find tracks and identify physics objects (particles, jets)

Source: CHEP 2023 proceedings

Hybrid-overlay

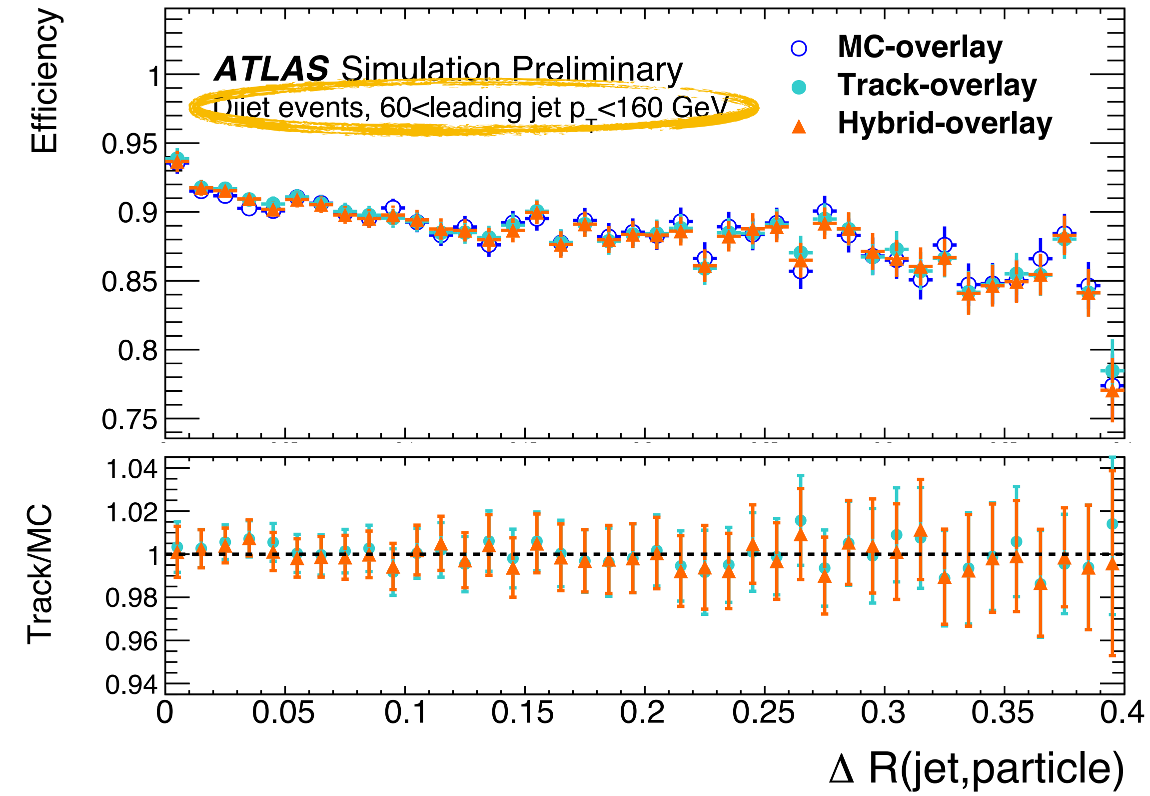
- Track-overlay is a good approximation as long as HS tracks do not pick up PU *HITS* (**no dense environment**)
- Use a **Deep Neural Network** to decide on an event-by-event basis whether Track-overlay can be used
  - Input features: kinematics of generator-level particles, event topology (e.g. local track density), PU information

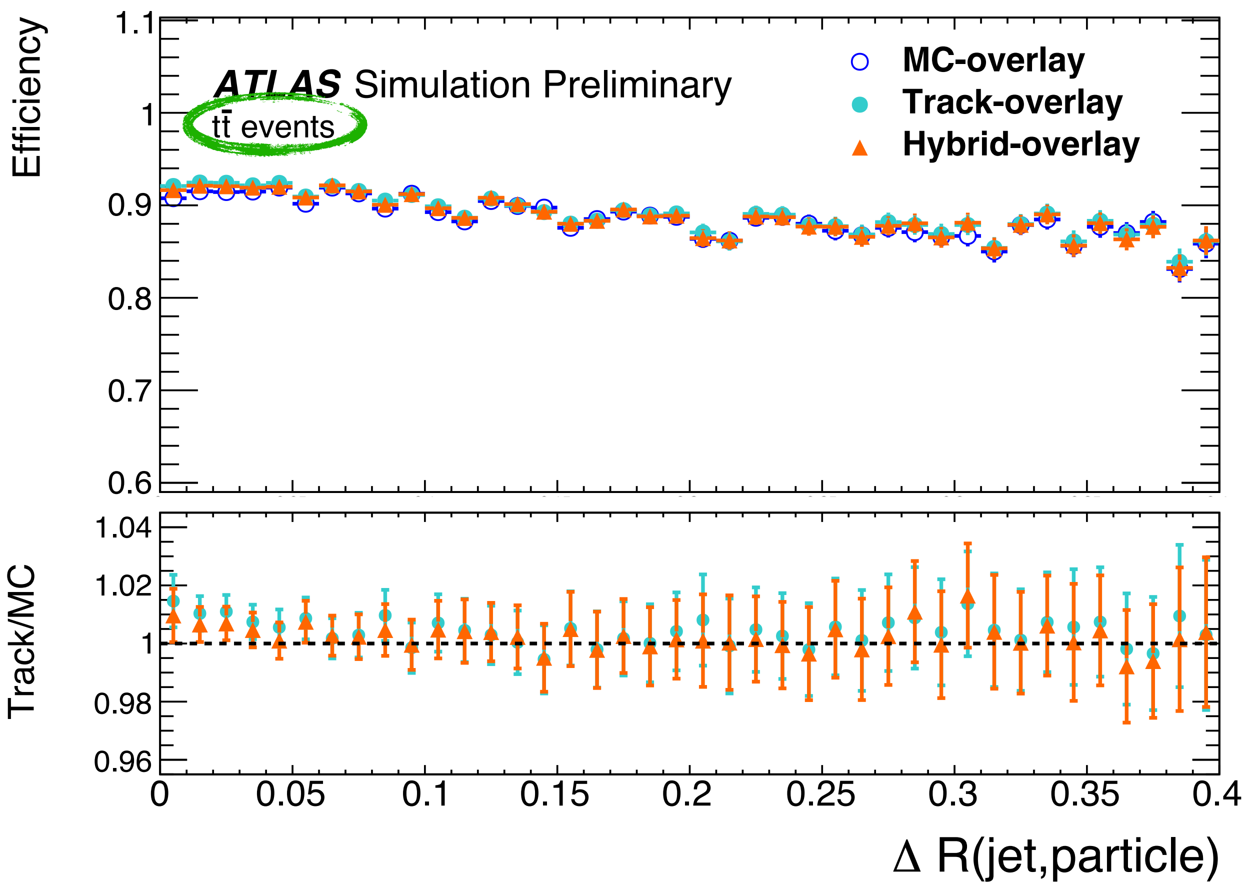


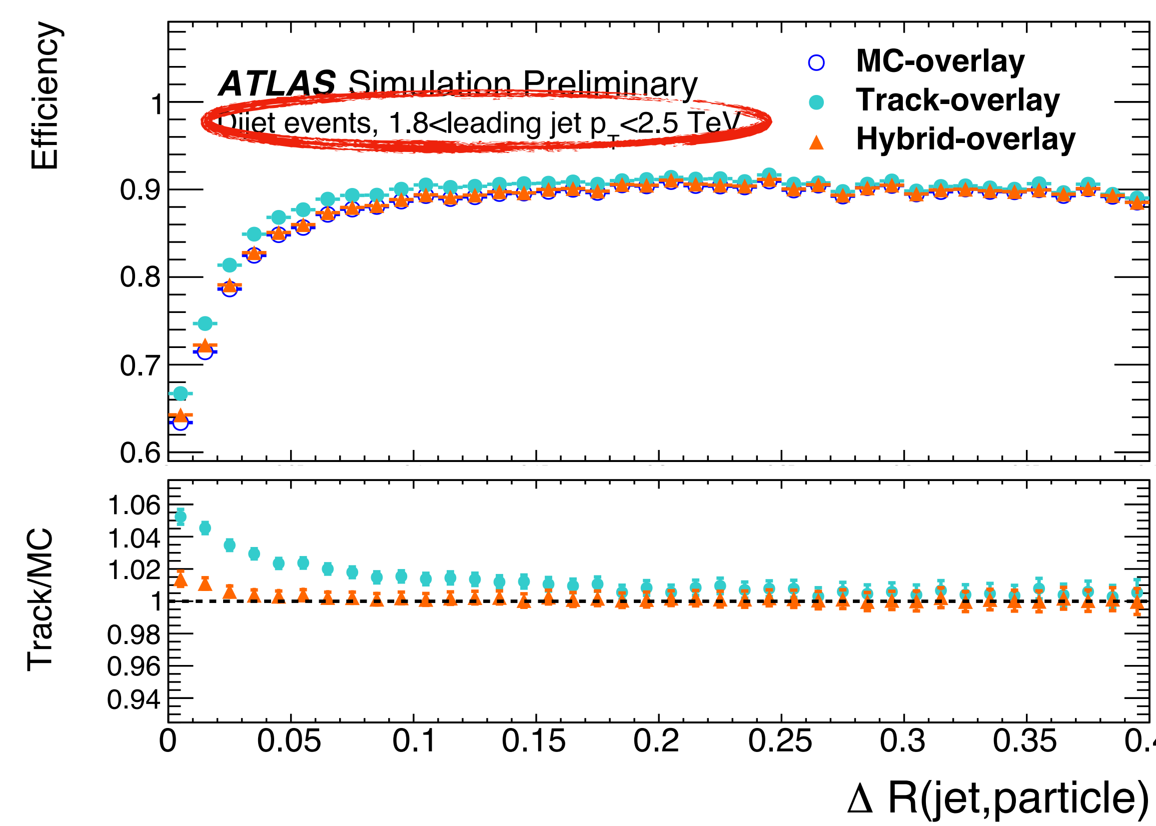
Source: CHEP 2023 proceedings

Hybrid-overlay performance

- **Top quark pairs**: 86% of events sent to Track-overlay
- **Low-p_{T} QCD multijets**: 94% of events sent to Track-overlay
- **High-p_{T} QCD multijets**: 35% of events sent to Track-overlay
- Negligible degradation in physics performance wrt MC-overlay
  - Ready to be validated for official MC production



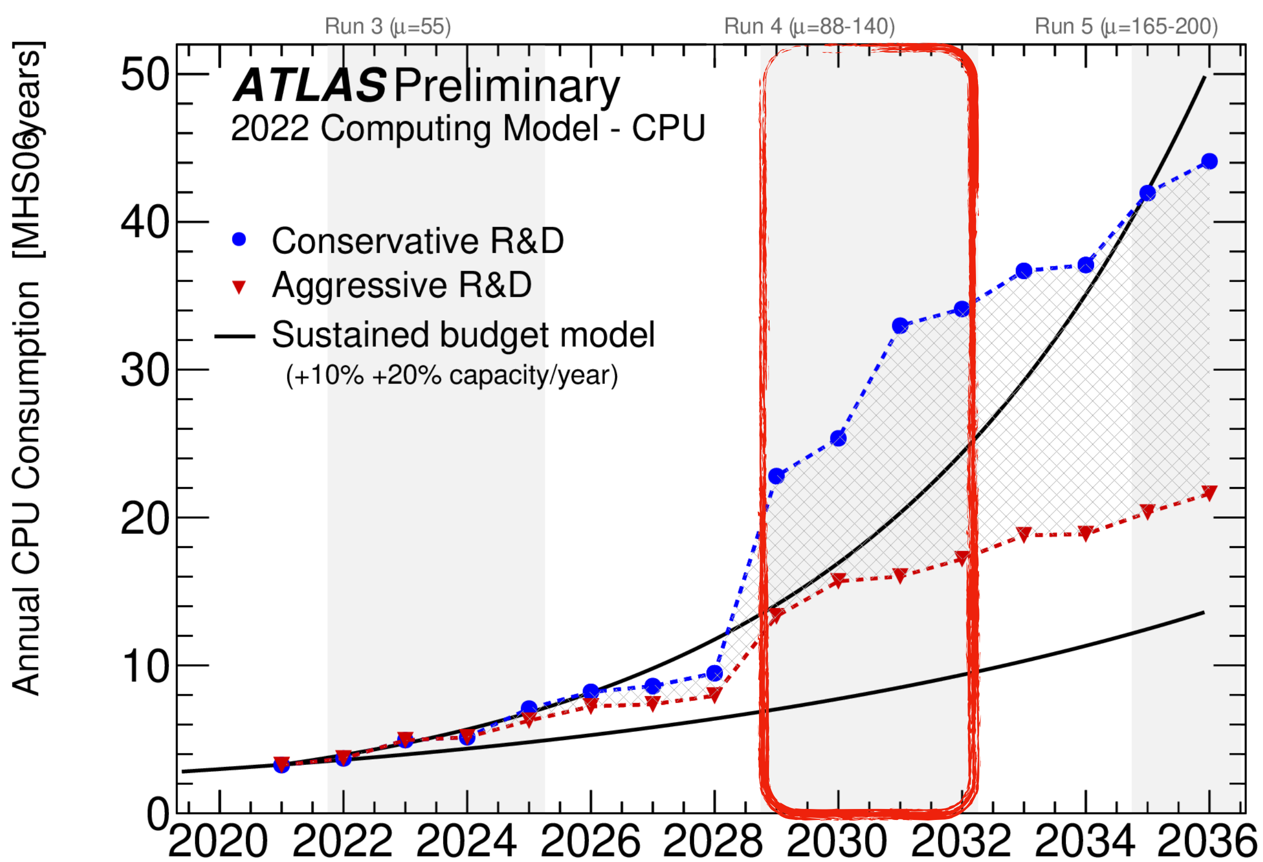


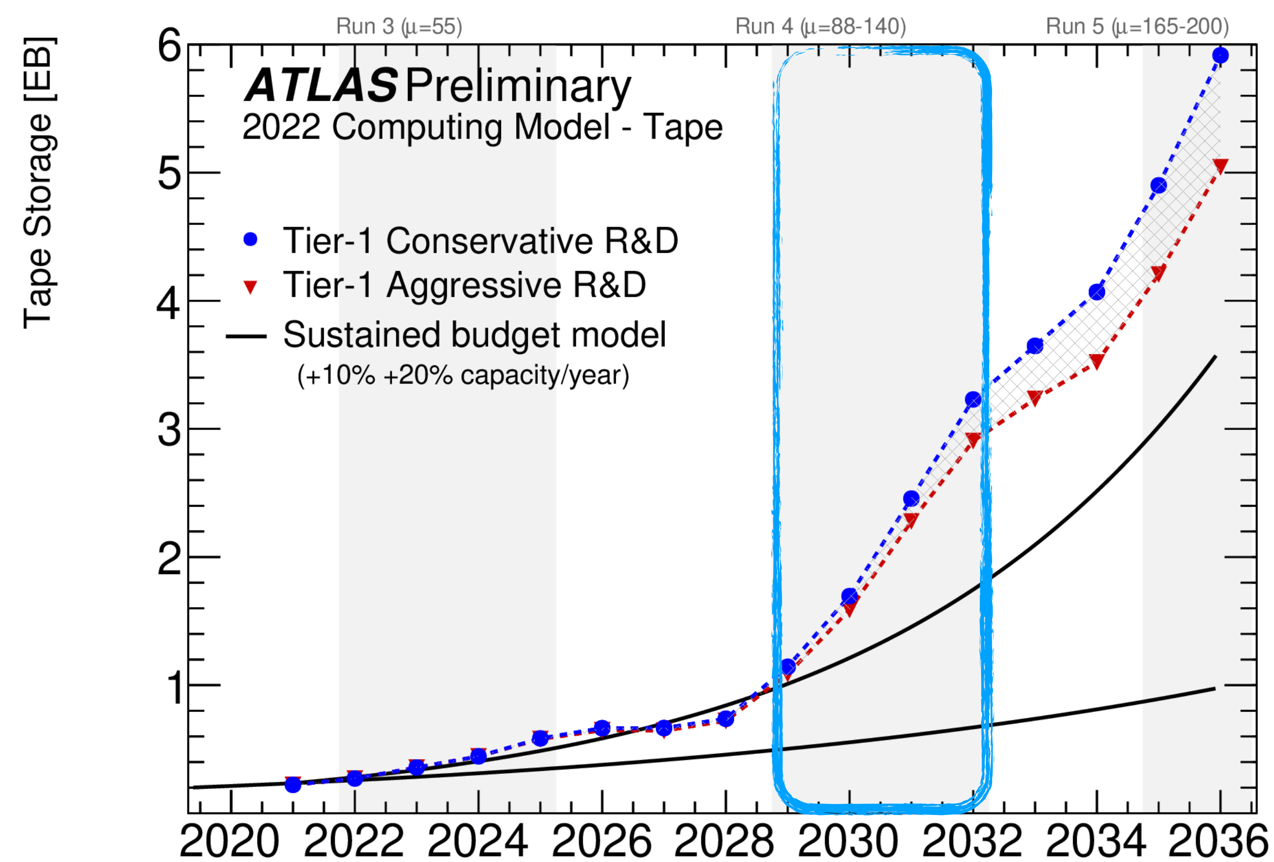


Sources: SIM-2024-001/

How to Effectively Use Fast Simulation Tools?

- **Fast Chain workflow** replaces CPU-intensive parts of the MC production chain with fast simulation tools and allows to
  - Prepare various production deployment scenarios running different simulation tools: flexible
  - Generate outputs used as input to physics analysis directly without storing intermediate files: tape/disk savings





Projected CPU savings

- 1.2 MHS06*year from **FATRAS:** tradable for **>200 PB of tape space** when skipping *HITS* and re-simulating everything each year
- 0.8 MHS06*year from **Track-overlay**

Sources: ATLAS S&C HL-LHC Roadmap, CDR

Conclusions

- **Fast Simulation Program of ATLAS** aims to provide a faster alternative to the standard MC production chain with more efficient handling of I/O and CPU resources
  - Essential to meet the **computational requirements** of the future runs of the LHC, as well as **physics modelling accuracy** needs
- **AtlFast3** for Run 3 reaches high precision for many objects and **3-15 times** CPU speed-up
  - Can be used for many physics analyses (not only signals but also backgrounds)
  - Default for HL-LHC (plan to study the feasibility for long-lived particles)
- **FATRAS** aims to further speed-up the simulation time by a **factor of ~3**
  - Work ongoing to improve physics modelling performance
- **Hybrid-overlay** is designed to increase CPU efficiency in the inner detector reconstruction by a **factor of ~1.8**
  - Negligible degradation in physics performance, deployment planned for Run 3
- By combining these fast simulation tools effectively into different **Fast Chain workflow** scenarios, it is possible to meet current ATLAS resource requirements (CPU/storage)

Conclusions

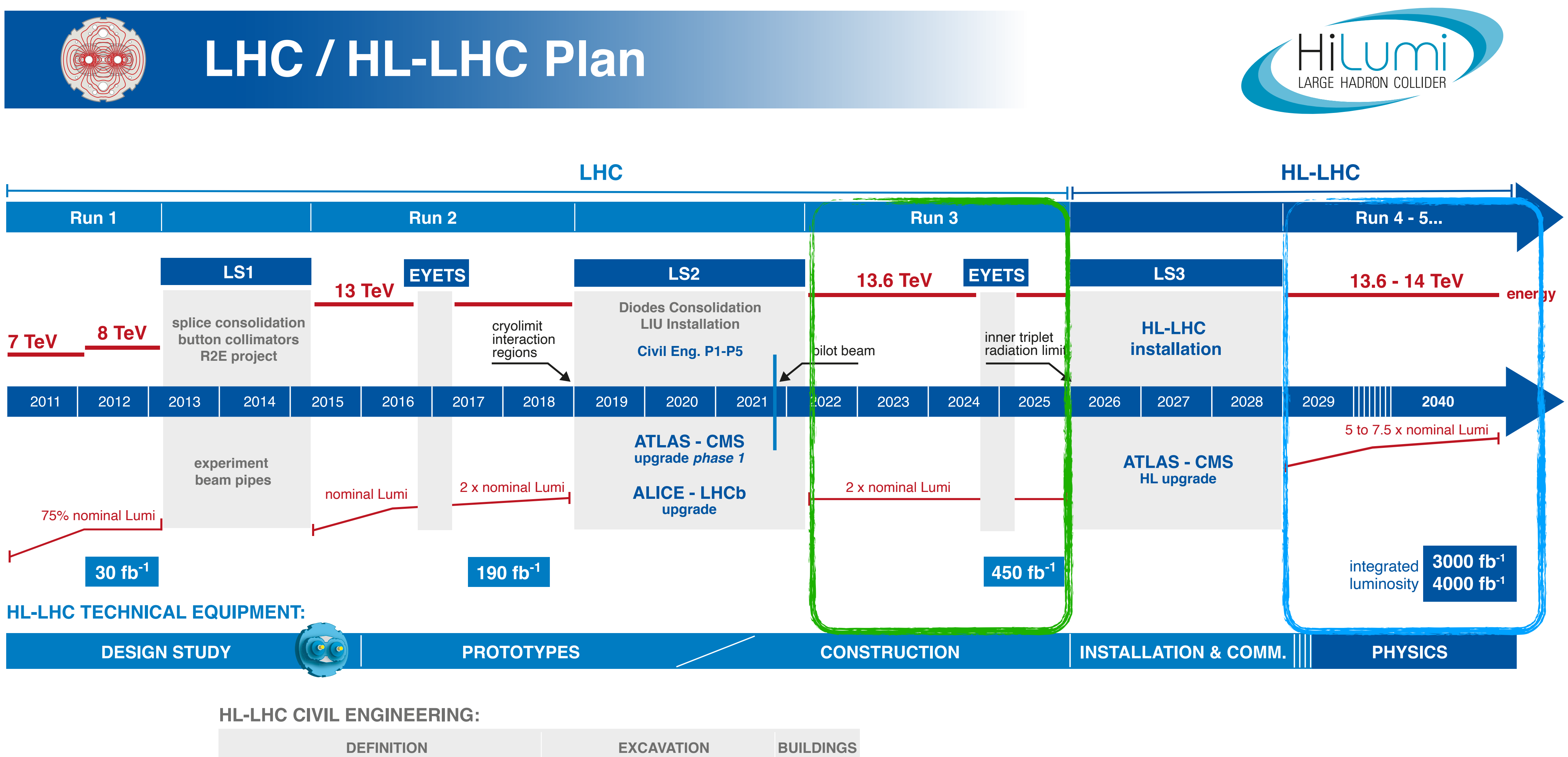
- **Fast Simulation Program of ATLAS** aims to provide a faster alternative to the standard MC production chain with more efficient handling of I/O and CPU resources
  - Essential to meet the computational requirements of the future runs of the LHC, as well as physics modelling accuracy needs
- AtlFast3 for Run 3 reaches high precision for many objects and **3-15 times** CPU speed-up
  - Ca
  - De

Thank you for your attention and stay tuned!

- FATRAS aims to further speed-up the simulation time by a **factor of ~3**
  - Work ongoing to improve physics modelling performance
- Hybrid-overlay is designed to increase CPU efficiency in the inner detector reconstruction by a **factor of ~1.8**
  - Negligible degradation in physics performance, ready to be validated for official MC production
- By combining these fast simulation tools effectively into different Fast Chain workflow scenarios, it is possible to meet current ATLAS resource requirements (CPU/storage)

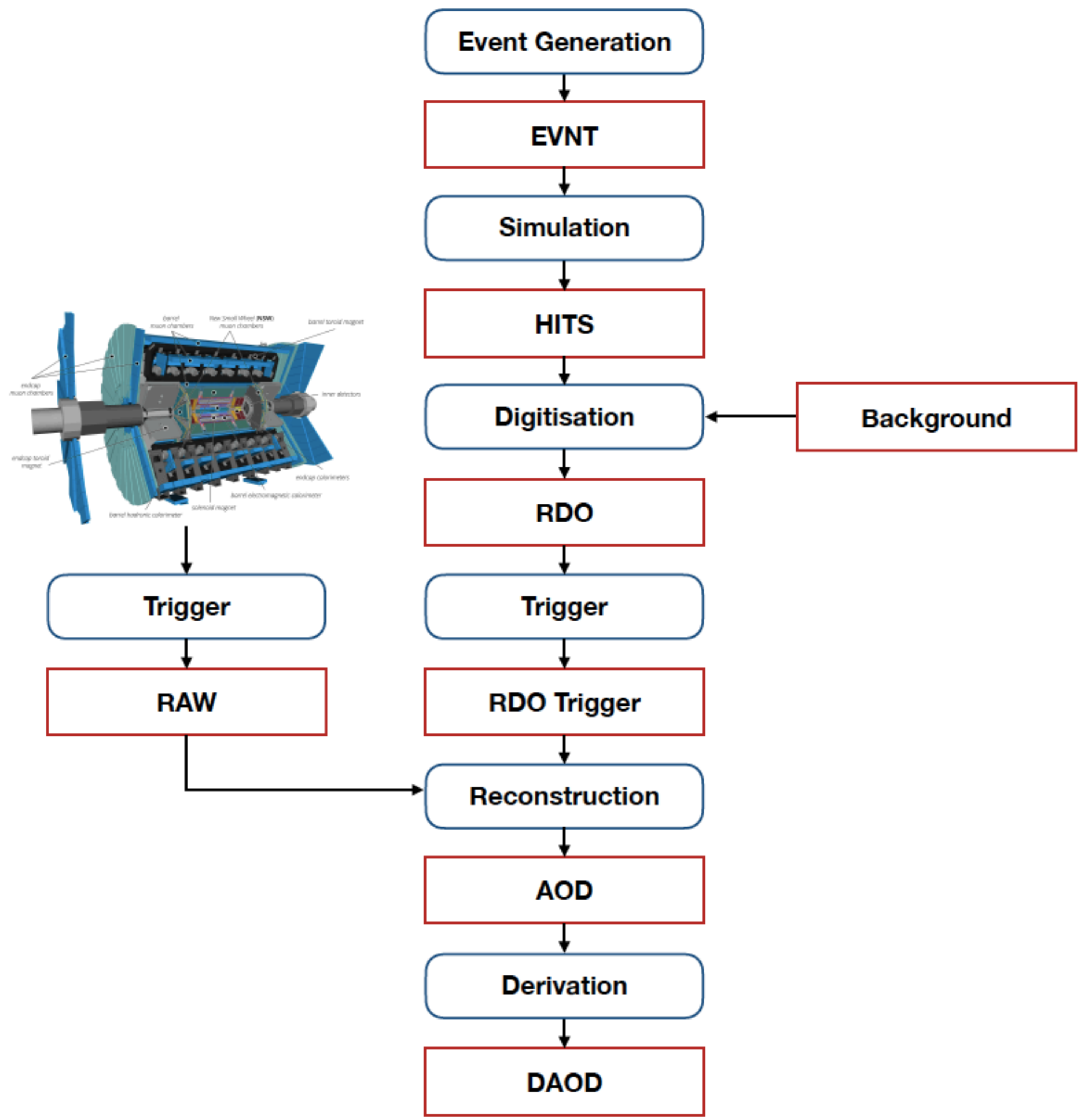
Backup

LHC program



Source: HL-LHC project

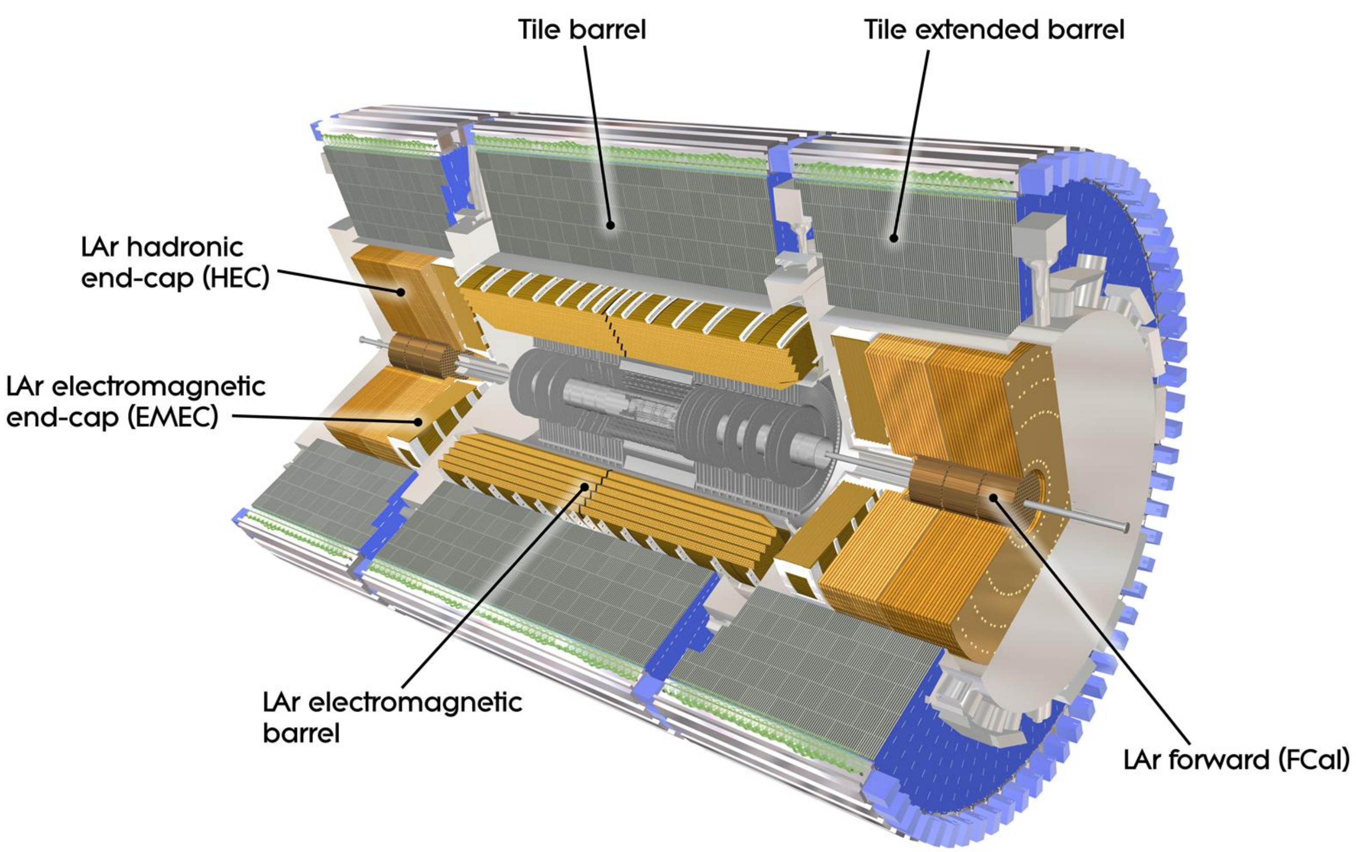
ATLAS Software Workflow

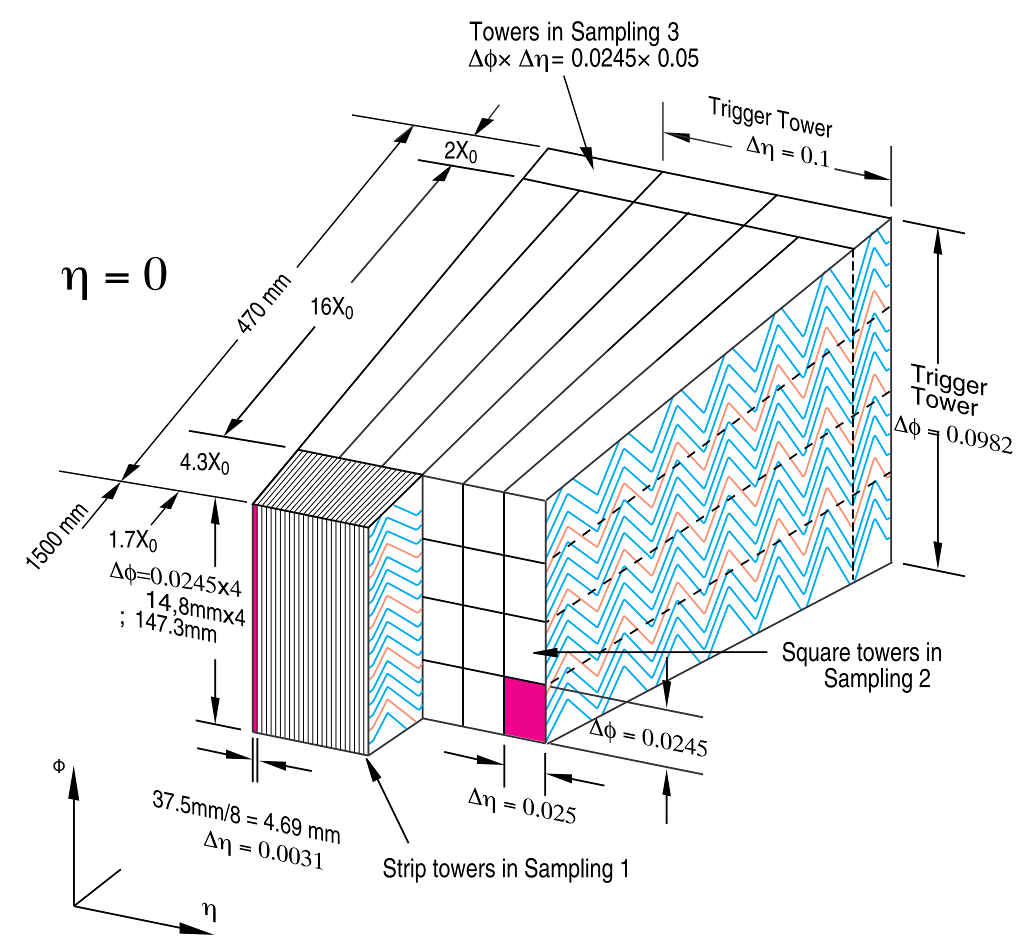


Sources: arxiv:2404.06335

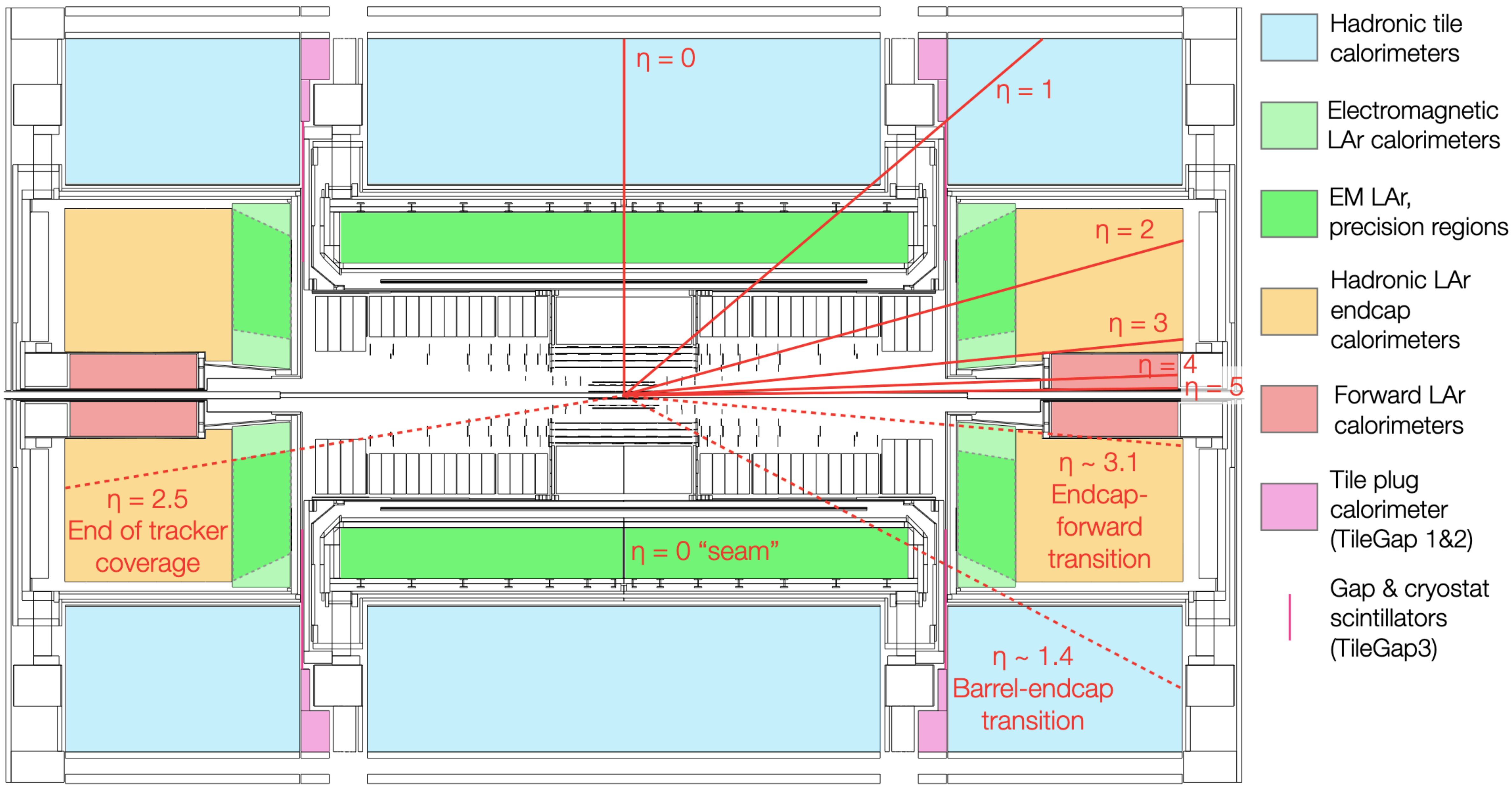
ATLAS Calorimeter System

Calorimeter	Layers	Module Name	η-coverage	Sampling Layer
Electromagnetic calorimeters	4	Electromagnetic Barrel (EMB)	$\lvert\eta\rvert < 1.5$	PreSamplerB, EMB1, EMB2, EMB3
	4	Electromagnetic Endcap (EMEC)	$1.5 < \lvert\eta\rvert < 1.8$	PreSamplerE
			$1.5 < \lvert\eta\rvert < 3.2$	EME1, EME2
			$1.5 < \lvert\eta\rvert < 2.5$	EME3
Hadronic calorimeters	4	Hadronic Endcap (HEC)	$1.5 < \lvert\eta\rvert < 3.2$	HEC0, HEC1, HEC2, HEC3
	3	Tile Barrel (TileBar)	$\lvert\eta\rvert < 1.0$	TileBar0, TileBar1, TileBar2
	3	Tile Extended Barrel (TileExt)	$0.8 < \lvert\eta\rvert < 1.7$	TileExt0, TileExt1, TileExt2
	3	Tile Gap (TileGap)	$1.0 < \lvert\eta\rvert < 1.6$	TileGap1, TileGap2, TileGap3
Forward calorimeter	3	FCal	$3.1 < \lvert\eta\rvert < 4.9$	FCal0, FCal1, FCal2
Transition regions	-	between barrel and endcap	$\lvert\eta\rvert \approx 1.45$	-
	-	between outer and inner wheel of endcap	$\lvert\eta\rvert = 2.5$	-
	-	between endcap and FCal	$\lvert\eta\rvert \approx 3.2$	-





Layout of ATLAS with pseudorapitidy



Source: arxiv:2007.02645

Fast Calorimeter Simulation History

AltFast → 5 years → **AF2** → 10 years → **AF3 Run2 version** → 2.5 years → **AF3 Run3 version**

AltFast	AF2	AF3 Run2 version	AF3 Run3 version
(Before data taking)	(Run 1, Run 2)	(Run 2)	(Run 3)
	Much improved wrt AF	Much improved wrt AF2	Improved wrt AF3 Run2

AltFast:

Simplistic, e.g. Gaussian smearing

AF2:

FastCaloSim
- Based on parametrisation
- 2D splines for lateral shape

- Limited shower fluctuations
  - Not usable for boosted topologies
- No FCAL parametrisation
  - Not usable for VBF
- No Muon Punch-Through

AF3 Run2 version:

FastCaloSimV2
- PCA for longitudinal energy deposits (better correlation)
- Hit-based lateral shape with more fluctuations to resemble energy resolution

FastCaloGAN
- First use of GANs in a large-scale HEP experiment)
- Used for mid-energy pions

Muon Punch-Through
- Simple (based on a look-up)

- $0 < |\eta| < 5$ fully covered
- Derived from G4 10.1
- Released in Spring 2021

AF3 Run3 version:

FastCaloSimV2
- Same as Run2 version

FastCaloGANV2
- Improved GANs
- Introduce proton GANs

Muon Punch-Through
- Improved based on DNN

- $0 < |\eta| < 5$ fully covered
- Derived from G4 10.6
- Released in Fall 2023

Source: Thanks to Rui Zhang and Jana Schaarschmidt for inputs.

AtlFast3 corrections

Energy Resolution Correction: only `FastCaloSim`

- Simulation of the resolution of the total energy is improved by reweighting to the distribution from `Geant4`

Energy ϕ-Modulation Correction: `FastCaloSim` and `FastCaloGAN`

- Due to the accordion structure of the EM calorimeter, the total deposited energy is modulated in the ϕ-direction: this modulation is not reproduced because it does not have a functional dependence on ϕ
- Resolution of showers is corrected by deriving the energy parameterisation after removing the modulation of the energy as a function of ϕ in the reference samples

Hadron Total Energy Correction: only `FastCaloSim`

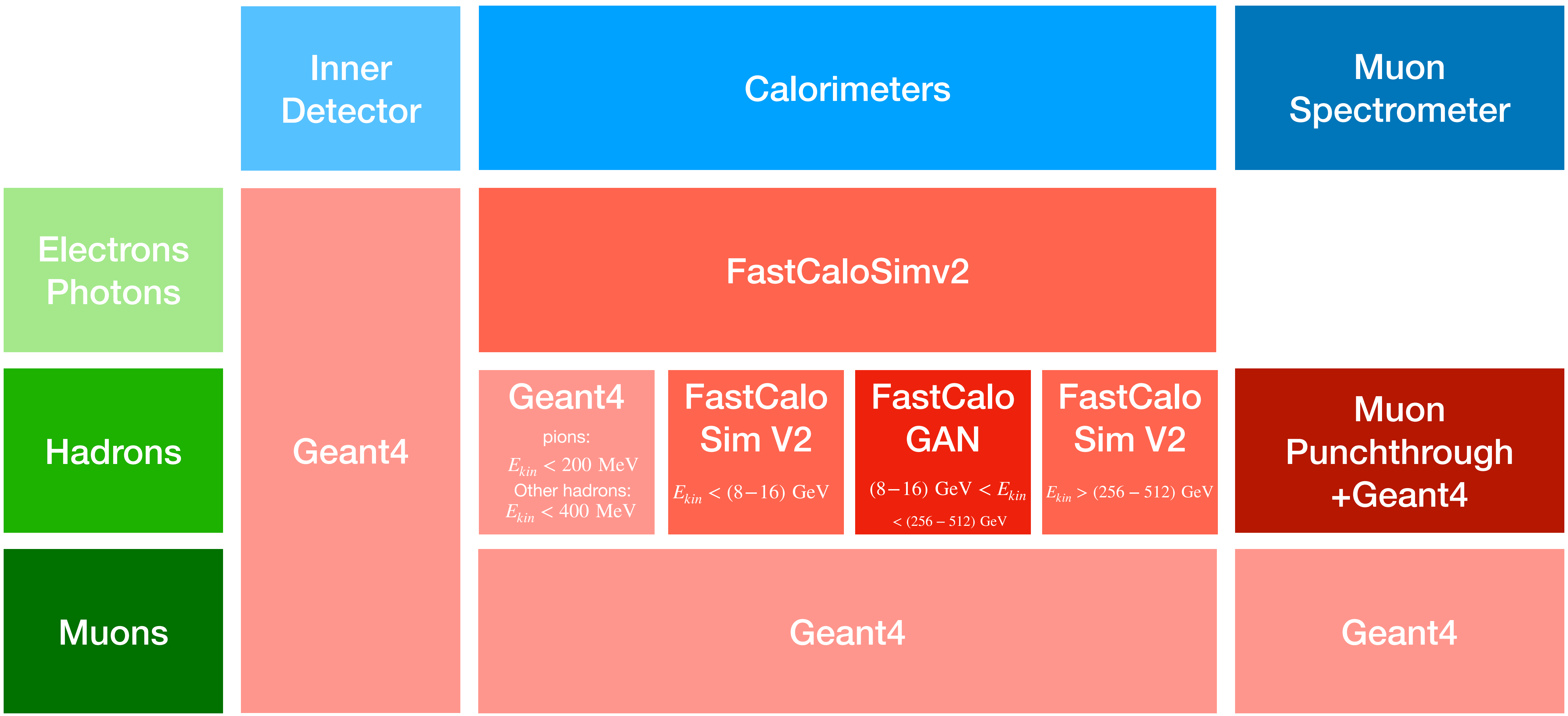
- Accounts for the difference between the charged-pion response, which is used to derive the calorimeter parameterisations, and the response to other hadron species

Simplified Geometry Shower Shape Correction: `FastCaloSim` and `FastCaloGAN`

- *HITS* generated by `FastCaloSim` and `FastCaloGAN` are assigned to calorimeter cells using a simplified cuboid geometry
- Small displacement in ϕ is assigned to each *HIT* before geometrically matching it to a cell with the simplified geometry

Fast Calorimeter Simulation

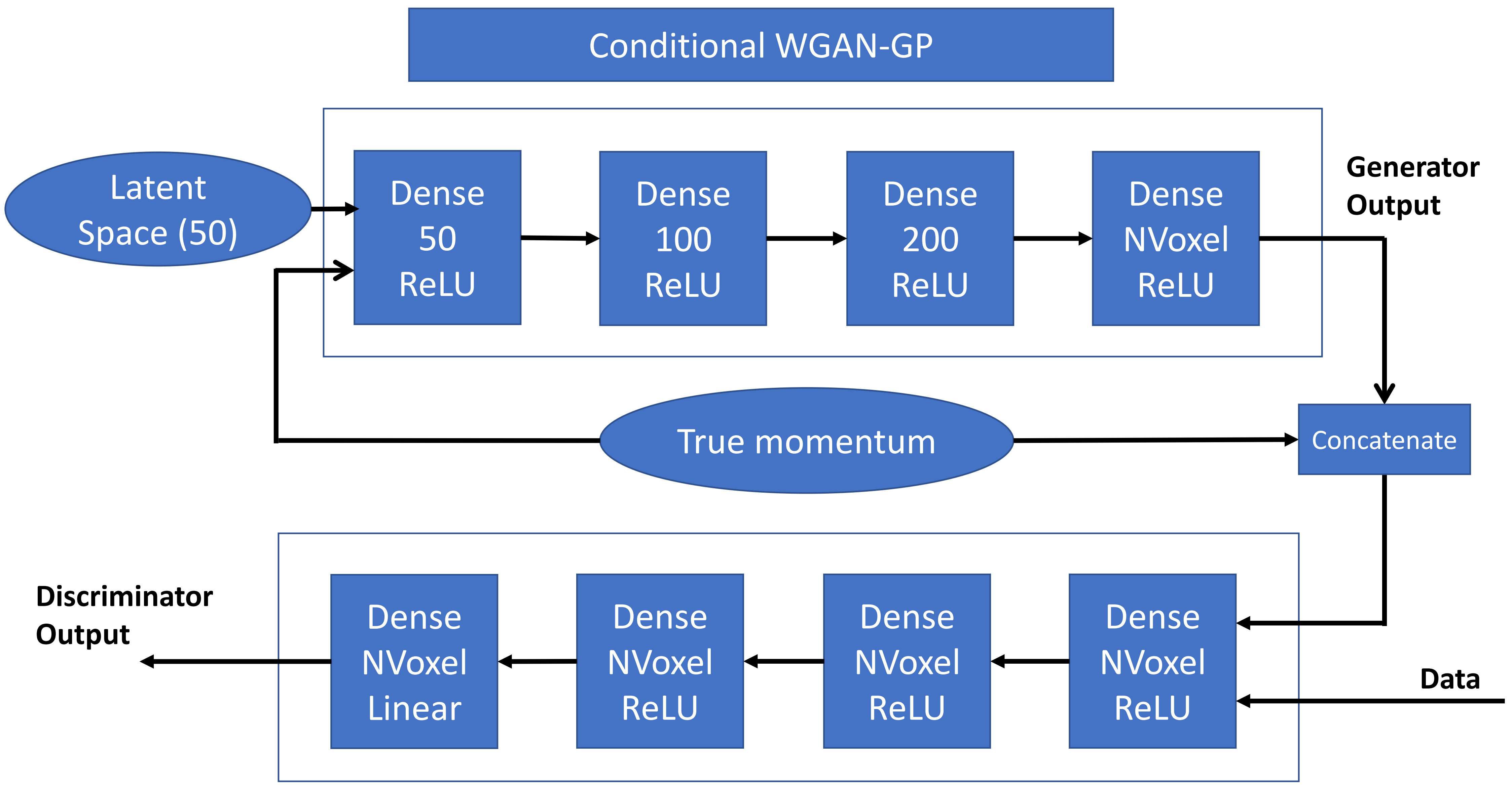
- Run 2 schema



Source: COMPUT SOFTW BIG SCI 6, 7 (2022)

FastCaloGAN model

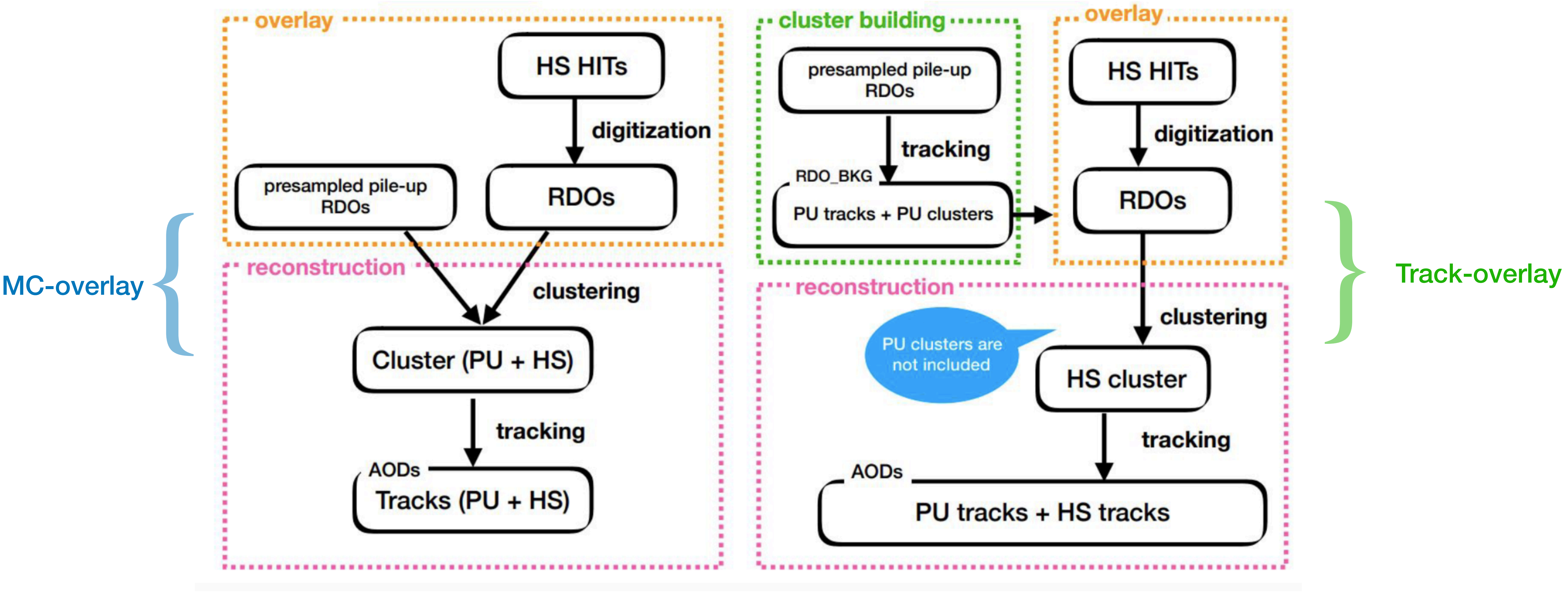
- FastCaloGAN uses the Wasserstein GAN with a gradient penalty (WGAN-GP) term in the loss function of the discriminator = good performance and training stability



- Rectified Linear Unit (ReLU) activation function l used in all layers of the discriminator with the exception of the last

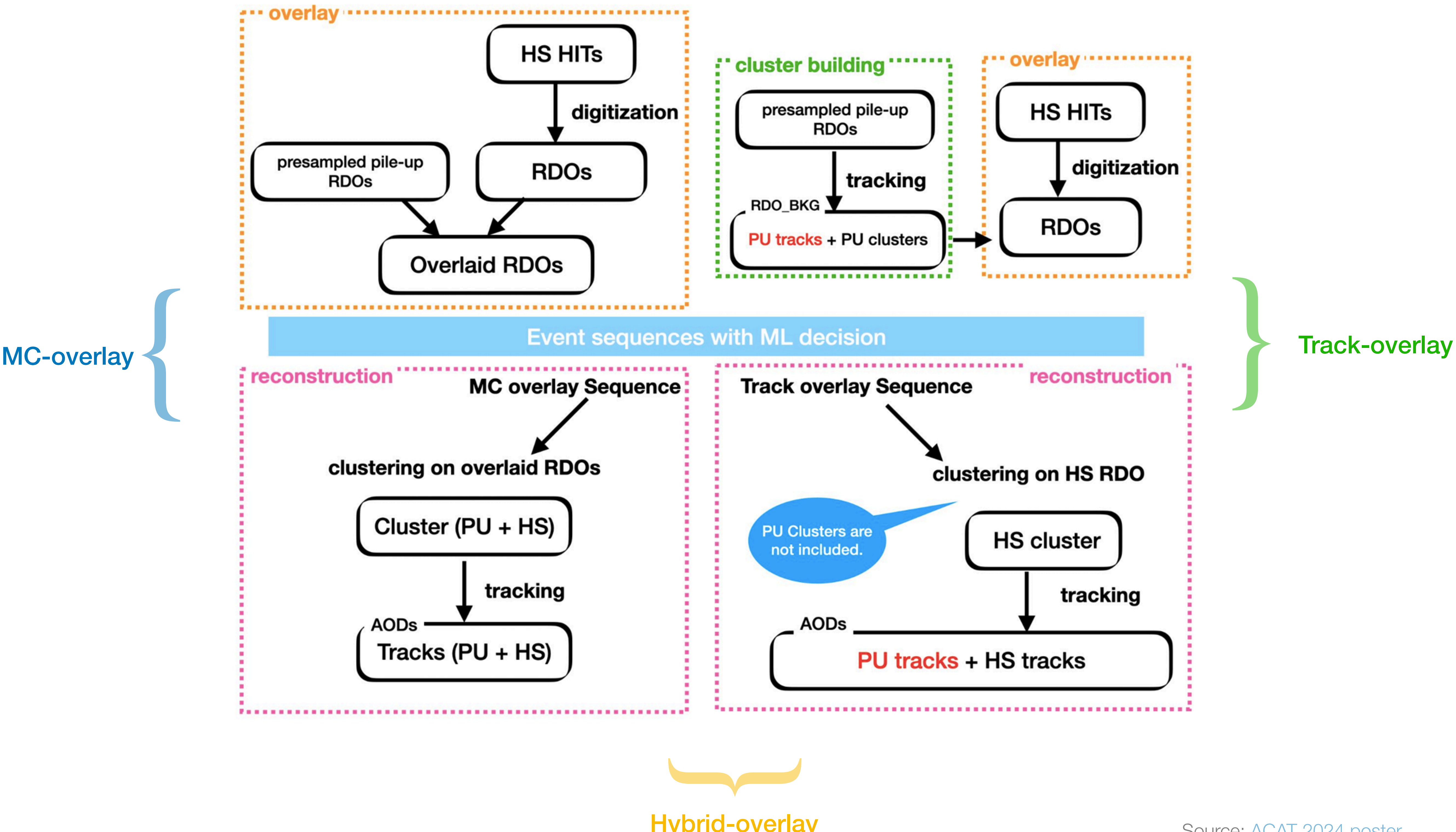
Source: COMPUT SOFTW BIG SCI 6, 7 (2022)

Fast Reconstruction



Source: [ACAT 2024 poster]

Hybrid-overlay



Source: ACAT 2024 poster