Fast Calorimeter Simulation in ATLAS with Modern Generative Models

Florian Ernst
on behalf of the ATLAS collaboration

CERN / University of Heidelberg

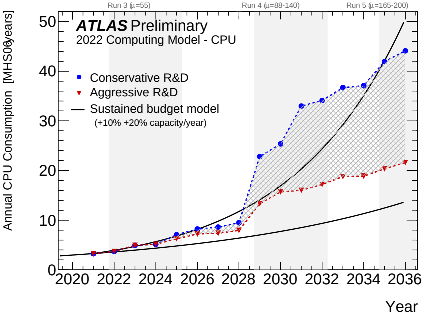
florian.ernst@cern.ch

September 16, 2026

This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA)

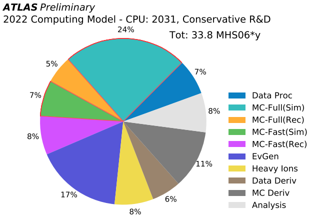
Introduction

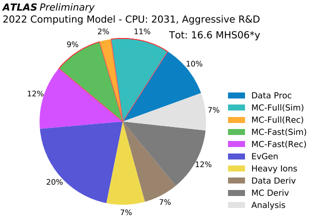
Why better fast calorimeter simulations?



ATLAS has to aggressively speed up its simulations to stay within its computing budget.
Therefore, the fast simulation must be usable for a majority of events which requires accuracy improvements wrt. to the current fast calorimeter simulation.

CERN-LHCC-2022-005

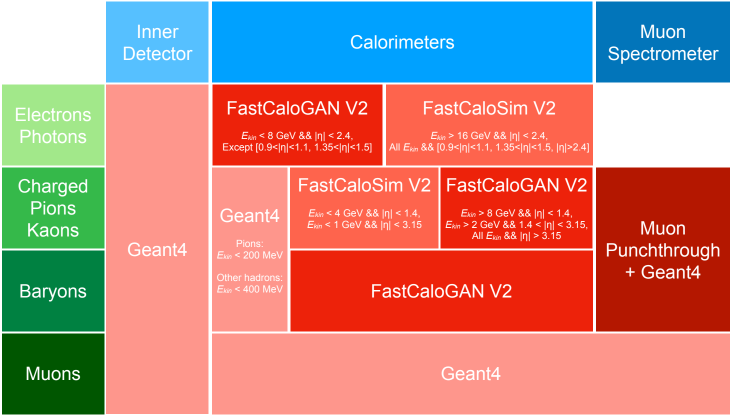




Conservative: 67% FastSim

Aggressive: 90% FastSim

Current fast simulation: AtlFast3



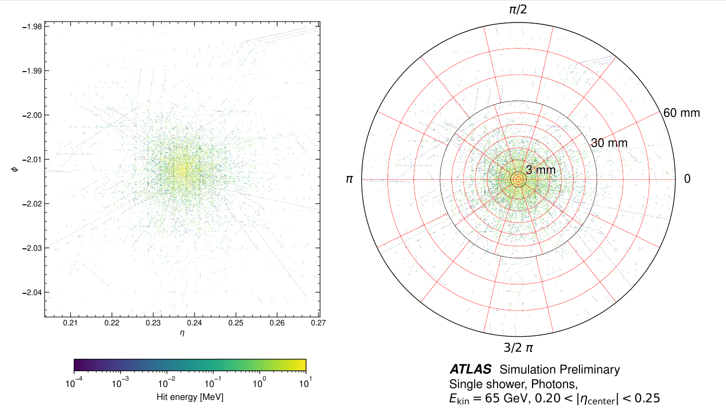
- AtlFast3 employs a mixture of classical parametric simulations (FCSV2) and generative adversarial networks (FastCaloGAN)
- FastCaloGAN mostly inferior for electromagnetic particles
- Current effort to improve the AtlFast generative models

Dataset and Preprocessing

CaloChallenge and dataset considerations

- The CaloChallenge identified more expressive alternatives for the FastCaloGAN:
  - (Continuous) normalizing flows
  - Diffusion models
  - Vision transformers
- Need a suitable ATLAS-specific training dataset
- Original FastCaloGAN voxelization introduced problematic artifacts
- For now focus only on photon showers
- First version of the dataset publicly available at CERN opendata

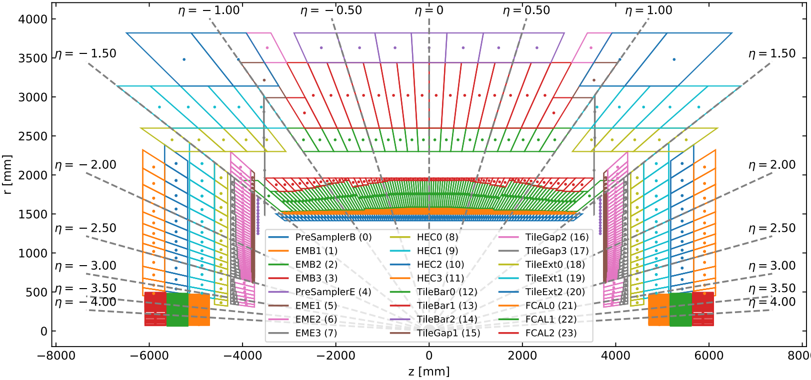
Used voxelization for an example shower



SIMU-2026-14

- Red lines indicate the non-uniform voxelization for a single photon shower in EMB1
- A regular voxelization with a fixed number of angular bins and the same voxelization for each layer used for the transition regions

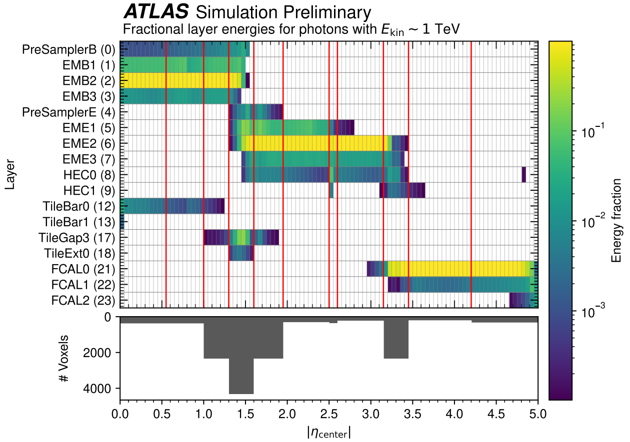
Calorimeter geometry overview



The Acts project: track reconstruction software for HL-LHC and beyond

- Three main regions: EMB (barrel), EME (endcap) and FCAL (forward calorimeter) that are relevant depending on the shower center position η_{center}.
- Overlapping at the intersections ⇒ Transition regions

Define the calorimeter layers to be simulated

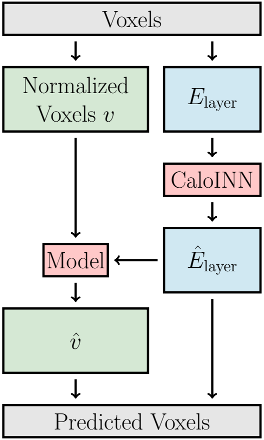


SIMU-2026-02

- Set of contributing layers has a strong dependence on η_{center}
- Identified 11 regions (defined by the red vertical lines) with constant number of required layers and voxels $\rightarrow$ Fixed input dimensionality
- Within one region train only single models conditioned on η_{center}

Machine learning models

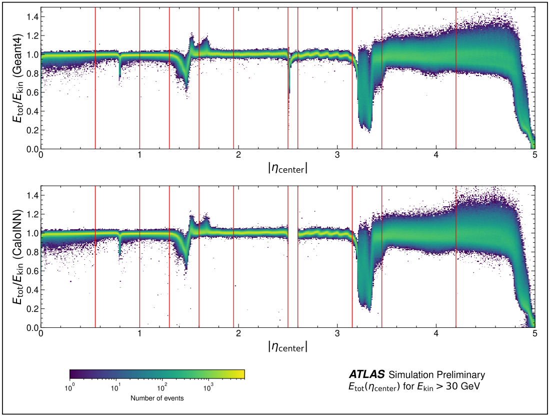
ML setup



- Use a single 2-stage model (like CaloFlow) per detector region.
- Successfull $\eta-$ & $E_{\text{kin}}-$conditioning
- (Layer) energy model: CaloINN
- 2 different shape models:
  - CaloCFM with MLP or Conv backbone.
  - Diffusion Transformer (CaloDiT-2) using patches (only for the regions with ≥ 6 layers, $1.0 \leq |\eta_{\text{center}}| \leq 1.95$ and $3.15 \leq |\eta_{\text{center}}| \leq 3.45$).

Performance evaluation

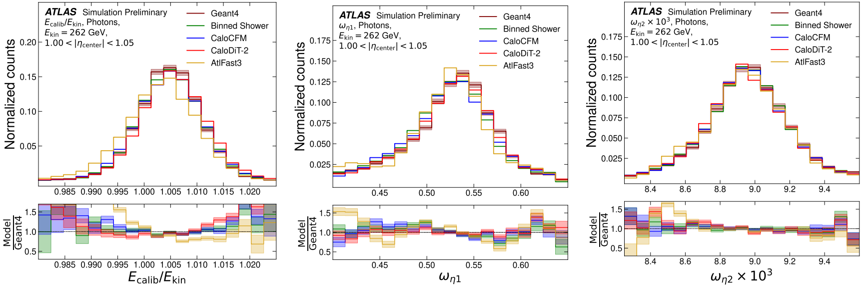
Measured energy vs incident energy - CaloINN



SIMU-2026-02

- Ratio between the total deposited energy E_{tot} and the photon's incident energy E_{kin}
- The CaloINN (bottom) is reproducing Geant4 (top) highly accurately and is used for all following ML-based simulations ($\frac{\chi^2}{N_{dof}} \sim 1 - 2$)
- Training dataset smaller for $2.5 \leq \eta_{center} \leq 2.6 \Rightarrow$ no ML model

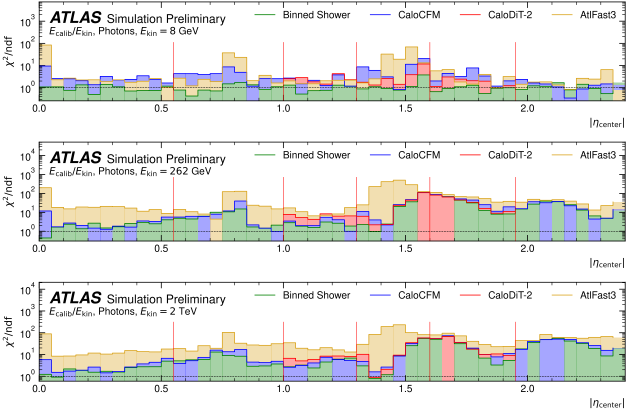
Reconstructed shower shapes for $1.0 \leq |\eta_{\text{center}}| \leq 1.05$



SIMU-2026-02

- Depicted: Deposited energy E_{calib}, η shower widths wrt. EMB1 or EME1 ($\omega_{\eta 1}$) and η shower widths wrt. EMB2 or EME2 ($\omega_{\eta 2}$)
- Binned shower: Upper bound on ML accuracy
- Slight reconstruction version inconsistencies $\Rightarrow$ AtlFast3 $E_{\text{calib}}/E_{\text{kin}}$ shifted
- E_{calib} width difference is a real artifact
- For $\omega_{\eta 1}$ and $\omega_{\eta 2}$ both models improve over AtlFast3
- Only a small η-region

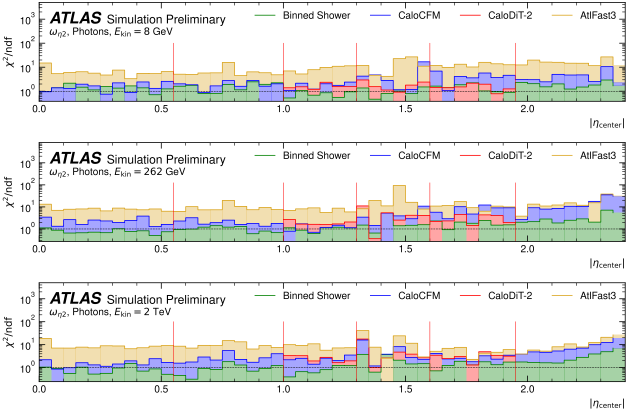
E_{calib} improved in most $\eta-$ and $E_{\text{kin}}-$regions



SIMU-2026-14

- χ^2-study using multiple histograms like before - comparing to Geant4
- E_{calib}-mean of AtlFast3 corrected to Geant4
- Low energies: ML predicted means slightly off
- Larger energies: Visible improvement over AtlFast3

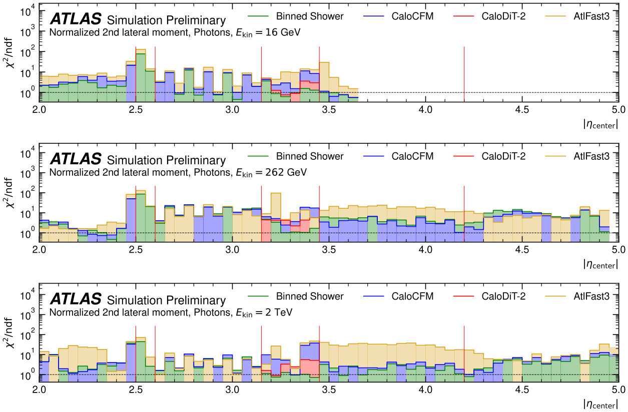
$\omega_{\eta 2}$: Both models improve over AtlFast3 everywhere



SIMU-2026-14

- Models always improving over currently used AtlFast3
- In the regions with ≥ 6 layers both models seem comparable

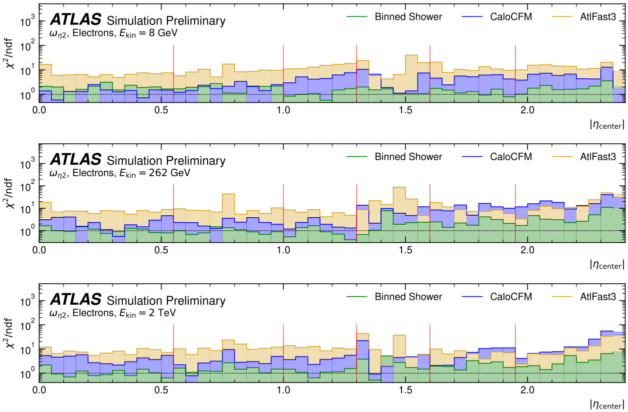
Similar performance for $|\eta_{\text{center}}| > 2.5$



SIMU-2026-02

- For $|\eta_{\text{center}}| > 2.5$, the definition of the $e - \gamma$ shower shapes ends
  $\Rightarrow$ Use topo-cluster observables, e.g. second lateral cluster moment
- Topo-clusters are the basic building block reconstructed in the calorimeter
- New models generally improving over AtlFast3
- Small differences for $3.45 \leq |\eta_{\text{center}}| \leq 4.2$ for few other observables

$\omega_{\eta 2}$: Nice out of the box results for electrons



SIMU-2026-14

- Models mostly improving over currently used AtlFast3
- Currently using exactly the photon architecture and hyperparameters
- Probably better results with tuned models in the future

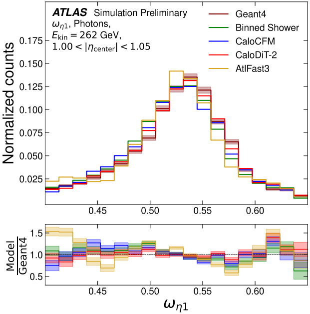
Preliminary computing benchmarks

	MLP-CaloCFM	Conv-CaloCFM	CaloDiT-2[1]	FastCaloGAN
$t\,[s]$	~ 0.013	~ 0.045	~ 0.018	~ 0.00004
mem $[\mathrm{MB}]$	~ 175	~ 55	~ 60	~ 14

- Measurements done for pure ONNX model evaluation
  $\Rightarrow$ General AtlFast3 overhead missing
- New models combine multiple slices while hundreds of GANs are needed
  $\Rightarrow$ Different number of used models
  $\Rightarrow$ Expected total memory usage similar to AtlFast3 (including FastCaloSim)
- FastCaloGAN was no CPU bottleneck
- Currently investigating impact on fast simulation of realistic event

[1] The CaloDiT-2 benchmarks were estimated based on the measurements for ATL-SOFT-PUB-2025-003 and the runtime and memory might be underestimated

Conclusion



SIMU-2026-02

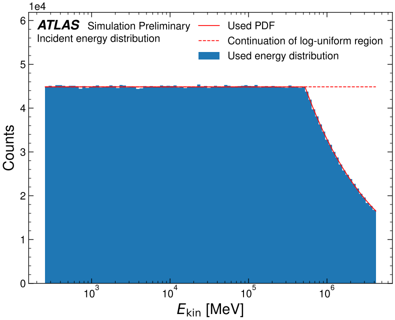
- Introduced a new photon and electron shower dataset for full ATLAS calorimeter
- New state of the art generative ML models are able to reproduce these datasets very accurately
- Improving over AtlFast3 almost everywhere in the ability to successfully reproduce Geant4.
- Memory consumption expected similar to AtlFast3
- Inference time probably not critical
- Next step: Hadronic showers

Questions?

Backup

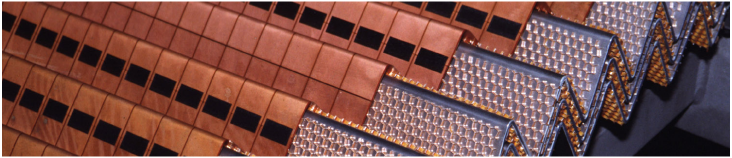
New dataset

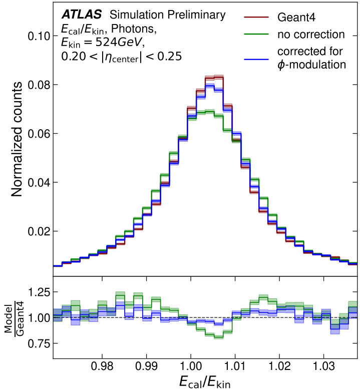
- Created 15 M new single photon and electron Geant4 events.
- No calorimeter noise and pileup.
- Incident kinetic energy $256~\text{MeV} < E_{\text{kin}} < 4.2~\text{TeV}$.
- Shower center $|\eta_{\text{center}}| < 5.0$.



SOFT-PUB-2025-03

ϕ-Modulation

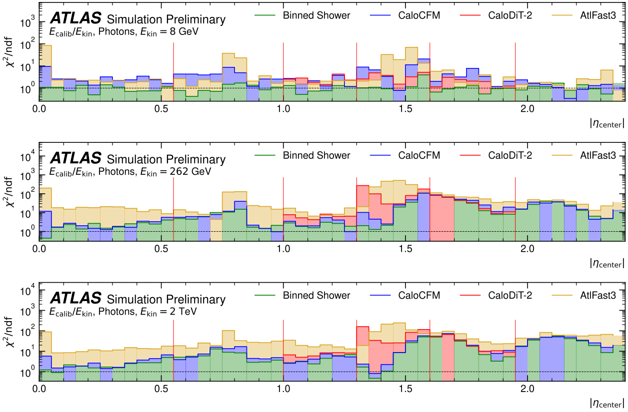




SIMU-2025-02

- The accordion structure of the ATLAS calorimeter introduces absorber-periodic fluctuations in the calibrated energy response E_{cal}/E_{kin}.
- By randomly shifting the shower wrt. its original ϕ-position this random modulation is increasing the width of the observed calibrated energy (green curve) compared to Geant4 (brown curve).
- In the new dataset, this modulation was corrected for as a pre-processing step (blue curve).

E_{cal} improved in most $\eta-$ and $E_{\mathrm{kin}}-$regions



SIMU-2026-02

- χ^2-study using multiple histograms like before - comparing to Geant4
- E_{cal}-mean of AtlFast3 corrected to Geant4
- Low energies: ML predicted means slightly off
- Larger energies: Visible improvement over AtlFast3