Fast simulation in ATLAS for LHC Run 3 and beyond

F.A.G. Corchia on behalf of the ATLAS Collaboration

Dipartimento di Fisica e Astronomia, Università di Bologna - Bologna, Italy
INFN, Sezione di Bologna - Bologna, Italy

Summary. — Simulation of detector response takes the largest amount of the computational resources of the ATLAS experiment at the LHC, with a large part used for simulation of calorimeters. As demand of CPU resources is expected to increase during Run 3 and in view of HL-LHC Run 4, action to ease this burden is required. One of the solutions developed by the Collaboration is the *fast simulation* system AtlFast3, which uses classical parametric and machine learning-based approaches to run simulation faster and with a smaller resource footprint with respect to Geant4. This work presents the version of AtlFast3 currently in production for simulation of Run 3 samples, discussing technical details, computing and physics performance, and the ongoing work in view of Run 4.

1. – Introduction

Simulation of Monte Carlo (MC) events in the detector is a major computing challenge for LHC experiments, taking about 40% of the total load on the computing resources [1] of the ATLAS experiment [2]. A large part of simulation resources is taken by the detailed simulation of electromagnetic and hadronic showers in the ATLAS calorimeter system (about 80% of CPU consumption for $t\bar{t}$ processes) [3]. In addition, ATLAS resource demand is expected to increase significantly in the coming years (see fig. 1), already during the current data taking run of the LHC (Run 3) and in view of Run 4, which will be the first one of the High Luminosity LHC (HL-LHC) [4].

This situation calls for significant effort in research and development and one of the solutions is the introduction of *fast simulation* tools. These are programs able to simulate detector response faster than the standard full process simulation tool Geant4 [5, 6, 7], but keeping the loss of accuracy to a minimum.

2. – ATLFAST3 and Its Components

AtlFast3 is the fast simulation system developed by the ATLAS Collaboration [3, 8]. Already introduced for Run 2, it has been further improved in preparation for the current Run 3 and is now in production. This tool replaces the slow propagation and interactions

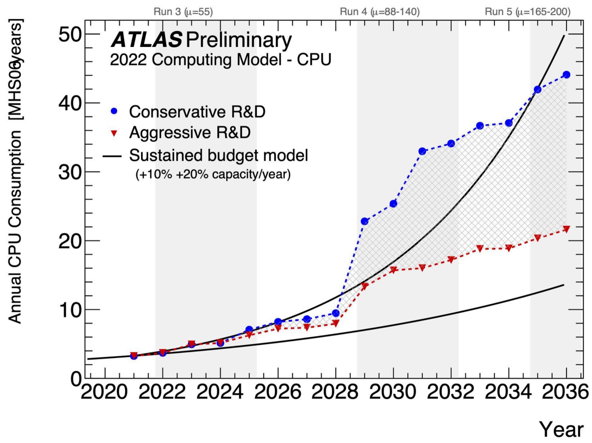

Fig. 1. – Projected evolution of compute usage, under a "conservative" and "aggressive" R&D scenarios. The hatched shading shows the range of resource consumption if the aggressive scenario is only partially achieved; the solid lines indicate the impact of sustained year-on-year budget increases and improvement in new hardware amounting together to a capacity increase of 10% (lower line) and 20% (upper line). The vertical shaded bands indicate the periods of data taking for ATLAS. MHS06 are 10^6 HS06 (as defined in [9]) [4].

of particles inside the calorimeter volume with the direct generation of energy deposits, by means of an underlying parametrisation.

AtlFast3 is composed of two subsystems, corresponding to two different approaches to fast simulation, which are presented later in this section: FastCaloSimV2 (doing longitudinal and lateral parametrisation of showers) and FastCaloGANV2 (machine learning-based, employing Generative Adversarial Networks - GANs). AtlFast3 runs fast simulation through one of the two approaches, according to the configuration in fig.

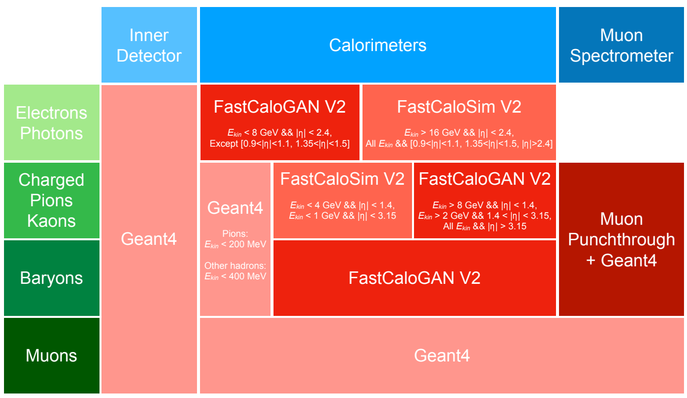

Fig. 2. – Configuration of the various subsystems of AtlFast3, as used for Run 3, depending on detector region, particle type and particle energy. Geant4 is still used in specific cases. Muon Punchthrough (the spray of particles into the Muon Spectrometer resulting from late interacting high-energy hadrons) is modelled with a tool based on Deep Neural Networks (DNNs) [8].

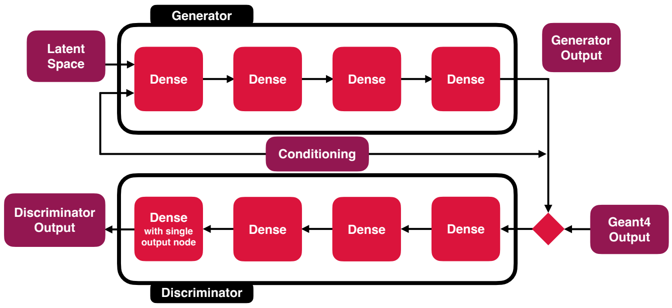

Fig. 3. – FastCaloGANV2 architecture.

2: the choice of the simulator was carefully tuned through extensive validation and is optimised to give the best physics performance.

2˙1. FastCaloSimV2. – FastCaloSimV2 is a fast simulation tool parametrising separately the longitudinal and lateral shower development [3]. During simulation, energy is directly deposited into the calorimeter cells using the parametrised responses. Parametrisation is done using Geant4 single photon, electron and charged pion samples, in 17 logarithmically spaced energy bins from 64 MeV to 4 TeV and 100 linearly spaced bins in pseudorapidity from 0 to 5 in absolute value. For the longitudinal shower development, as the amount of energy deposited in each layer by the shower particles (which depends on how deep in the calorimeter the shower was initiated) is highly correlated between layers (making an independent parametrisation of the response of each layer difficult), *Principal Component Analysis* (PCA) is used to classify showers for each slice of energy, pseudorapidity bin and particle type. Two PCA transformations are performed, the first one classifying showers into bins, the second one acting on each bin of the first PCA to generate uncorrelated and approximately Gaussian distributions. During simulation, PCA bins are randomly selected, followed by the generation of uncorrelated random numbers, which are then mapped back to the total energy and the energy fractions deposited in each layer with the inverse transformation. The lateral shower shape is instead parametrised as bidimensional probability density functions.

2˙2. FastCaloGANV2. – FastCaloGANV2 is a machine learning-based fast simulation system based on GANs [10]. This architecture, first introduced in [11], involves the simultaneous training of two neural networks, one called *generator* and the other one *discriminator*. The generator aims to generate samples as similar as possible to Geant4-generated ones, while the discriminator is fed samples from both Geant4 and the generator and aims to distinguish actual Geant4 samples from the ones produced by the generator. Both try improving their abilities and, once a Nash equilibrium is reached between the two, the FastCaloGANV2 generator is ready to simulate calorimeter showers.

FastCaloGANV2 employs specifically Wasserstein GANs with gradient penalty (WGAN-GP) [12], an evolution of the original GAN architecture that improves train-

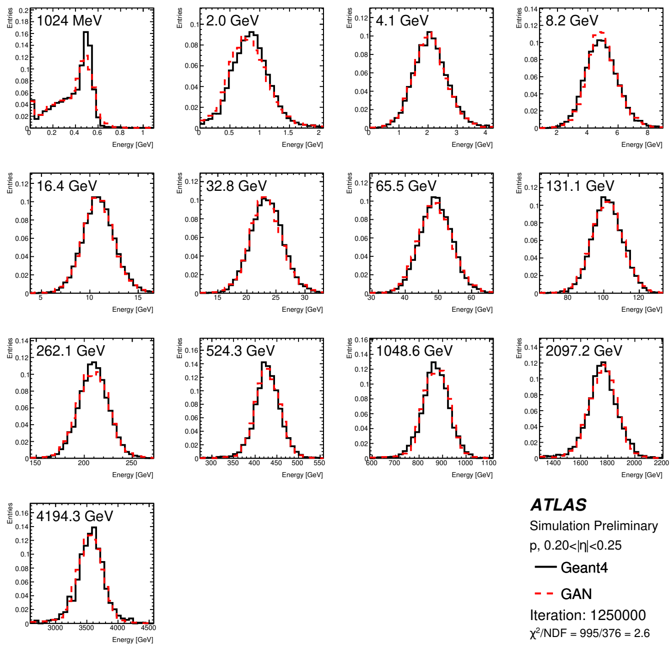

Fig. 4. – Sum of the energy in all voxels for single protons generated at the calorimeter surface in the pseudorapidity range between 0.2 and 0.25 in absolute value. Geant4 is compared to the GAN trained within FastCaloGANV2 [13].

ing by making it more stable and performant. The architecture is shown in fig. 3. FastCaloGANV2 was trained for electrons, photons, protons and charged pions as shower-initiating particles, in each of the 100 bins in pseudorapidity, with conditioning (the GAN being supplied additional information acting as class labels) on truth momentum; calorimeter hits are grouped into three-dimensional bins (*voxels*). Fig. 4 compares simulations as done by Geant4 and FastCaloGANV2; reduced χ^2 equal to 2.6 is observed.

3. – Performance

3˙1. *Computing Side*. – AtlFast3 is from 3 to 15 times faster than the Geant4 simulation of the ATLAS Run 3 detector, the lowest and highest speedup being observed respectively for $Z \to ee$ events and for high transverse momentum dijet events [14]. Simulation time in AtlFast3 is dominated by simulation of the Inner Detector, which, as shown in fig. 2, is fully handled by Geant4. In addition, for the average ATLAS MC simulation event, AtlFast3 only needs 20% of the CPU of the full simulation [8].

3˙2. *Physics Side*. – AtlFast3 provides very accurate modelling of the leading cluster energy and, in dijet events, of the number of constituents for jet and substructure variables (see fig. 5). For most observables used in physics analyses, AtlFast3 and Geant4 agree within a few percent and AtlFast3 can be used for nearly every process, be it signal or background. This result is particular useful considering that, for Run 4,

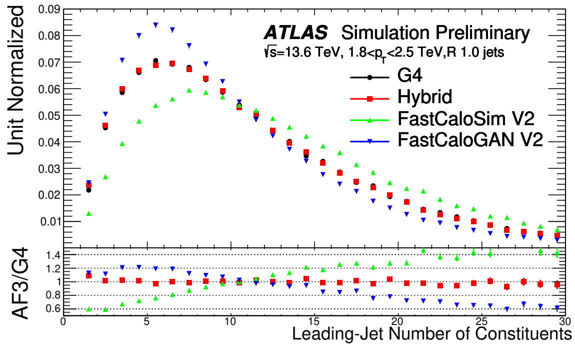

Fig. 5. – Number of constituents for the leading reconstructed jet in dijet events with transverse momentum between 1.8 and 2.5 TeV [13]. Results are compared for samples simulated with GEANT4, FASTCALOSIMV2 only, FASTCALOGANV2 only and the combination of the latter two ("Hybrid") according to the scheme in fig. 2.

more than 90-95% of analyses are expected to require the use of fast simulation, as there will not be the CPU capacity to allow full simulation for them [1].

4. – Towards Run 4

Now that ATLFAST3 is in production for Run 3, further development is in progress in view of the final part of Run 3 and the first part of HL-LHC Run 4. Improved voxelisation techniques aimed at reducing bias are currently being tested and additional machine learning models are under investigation, notably transformer-based diffusion models (*CaloDiT* [15, 16]) and Invertible Neural Networks (INNs [17]). These models had been identified as top performers in the *CaloChallenge* study [15] and are under consideration for integration into the fast simulation framework, in order to, if future developments maintain or improve their simulation quality, potentially supplement or replace FASTCALOSIMV2 and FASTCALOGANV2.

Voxelisation has been re-optimised in order to better emulate GEANT4; it features finer granularity at the shower centre to obtain higher spatial precision while accounting for voxel-to-voxel correlations. For the new models, current results show potential for better simulation with GANs, and INNs also yield an excellent simulation performance, accurately reproducing GEANT4 distributions with the optimised voxelisation [18].

5. – Conclusion

This work presented ATLFAST3, the state-of-the-art fast simulation tool of the ATLAS Collaboration at the LHC. ATLFAST3 for Run 3 can simulate a broad range of physics processes with high precision, with a great improvement in computing performance, as the tool runs with a CPU speedup of a factor 3-15 and only one fifth of the CPU with respect to GEANT4. ATLFAST3 is therefore essential to meet the computing requirements of the future runs of the LHC, as well as physics modelling accuracy needs.

REFERENCES

[1] ATLAS Collaboration, *ATLAS HL-LHC Computing Conceptual Design Report*, CERN-LHCC-2020-015 (2020) cds.cern.ch/record/2729668

[2] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST*, **3** (2008) S08003 doi.org/10.1088/1748-0221/3/08/S08003

[3] ATLAS Collaboration, *AtlFast3: The Next Generation of Fast Simulation in ATLAS*, *Comput Softw Big Sci*, **6** (2022) 7 doi.org/10.1007/s41781-021-00079-7

[4] ATLAS Collaboration, *ATLAS Software and Computing HL-LHC Roadmap*, CERN-LHCC-2022-005 (2022) cds.cern.ch/record/2802918

[5] Allison J. et al., *Recent Developments in Geant4*, *Nucl. Instrum. Meth. A*, **835** (2016) , pp. 186-225 doi.org/10.1016/j.nima.2016.06.125

[6] Allison J. et al., *Geant4 Developments and Applications*, *IEEE Trans. Nucl. Sci.*, **53** (2006) , pp. 270-278 dx.doi.org/10.1109/TNS.2006.869826

[7] Agostinelli S. et al., *Geant4 - A Simulation Toolkit*, *Nucl. Instrum. Meth. A*, **506** (2003) , pp. 250-303 dx.doi.org/10.1016/S0168-9002(03)01368-8

[8] ATLAS Collaboration, *Software and computing for Run 3 of the ATLAS experiment at the LHC*, *Eur. Phys. J. C*, **85** (2025) 234 doi.org/10.1140/epjc/s10052-024-13701-w

[9] w3.hepix.org/benchmarking/HS06.html

[10] ATLAS Collaboration, *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks*, ATL-SOFT-PUB-2020-006 (2020) cds.cern.ch/record/2746032

[11] Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A. and Bengio Y., *Generative Adversarial Nets*, *Advances in Neural Information Processing Systems*, **3** (2014) 11 dx.doi.org/10.1145/3422622

[12] Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V. and Courville A., *Improved Training of Wasserstein GANs*, arXiv:1704.00028 (2017) doi.org/10.48550/arXiv.1704.00028

[13] ATLAS Collaboration, ATLAS-SIM-2023-004 atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2023-004

[14] ATLAS Collaboration, ATLAS-SIM-2023-005 atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2023-005

[15] Krause C. et al., *CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation*, arXiv:2410.21611 (2024) doi.org/10.48550/arXiv.2410.21611 to be published in *Rep. Prog. Phys.*

[16] Peebles W. and Xie S., *Scalable Diffusion Models with Transformers*, arXiv:2212.09748 (2022) doi.org/10.48550/arXiv.2212.09748

[17] Ernst F., Favaro L., Krause C., Plehn T. and Shih D., *Normalizing Flows for High-Dimensional Detector Simulations*, *SciPost Phys.*, **18** (2025) 081 doi.org/10.21468/SciPostPhys.18.3.081

[18] ATLAS Collaboration, ATLAS-SIMU-2024-10 atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIMU-2024-10