

Advanced computational techniques for fast simulation of the ATLAS experiment calorimeter^(*)

F. A. G. CORCHIA⁽¹⁾⁽²⁾ and L. RINALDI⁽¹⁾⁽²⁾ on behalf of the ATLAS COLLABORATION

⁽¹⁾ *Dipartimento di Fisica e Astronomia, Università di Bologna - Bologna, Italy*

⁽²⁾ *INFN, Sezione di Bologna - Bologna, Italy*

received 2 December 2024

Summary. — Detector simulation uses the largest amount of computational resources of the ATLAS experiment at the LHC. About 80% of allocated resources are used for the simulation of calorimeters. Action to ease this burden is required, also in view of High Luminosity-LHC. This work presents two solutions in continuous development within the Collaboration. The first one foresees usage and development of fast simulation tools, which result in a faster simulation of the calorimeter response and with a smaller resource footprint with respect to traditional tools; this solution includes using techniques based on Machine Learning. The second strategy envisages the deployment of fast simulation training to resources external to CERN, including cutting-edge supercomputers like Leonardo at CINECA in Bologna.

1. – Introduction

Detector simulations uses the largest amount of computational resources of the ATLAS experiment [1] at the LHC [2], as shown in fig. 1. About 80% of this amount is required by simulation of calorimeters [3] and this load is expected to increase in absolute value with the forthcoming activation of the High Luminosity LHC (HL-LHC). Therefore, there is great need for solutions to ease this burden.

The first way is through the development of *fast simulation* systems, able to simulate calorimeter response faster with respect to the *full simulation* software Geant4 [4-6], guaranteeing at the same time good accuracy. These systems are mainly based on Machine Learning (ML) techniques, harnessing their potential for this task.

These ML systems require large resources for their training, making this the heaviest load among fast simulation tasks, therefore an additional improvement can be achieved by distributing their training on other resources than the ones commonly used at CERN (the CERN batch system LXBATCH and the Worldwide LHC Computing Grid [7]), in order to obtain further room for training and, if cutting-edge resources like supercomputers are used, a remarkable performance boost.

^(*) IFAE 2024 - “Poster” session

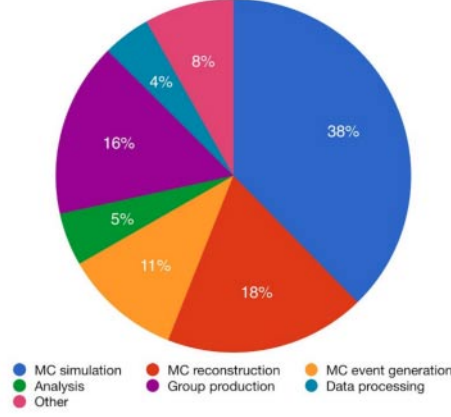


Fig. 1. – CPU-hours usage by ATLAS activities [2].

This work presents the ongoing work on the proposed solutions. For the first approach, the ATLAS fast simulation system *AtlFast3* is presented, while for the second one the *BoloGANtainer* system is introduced.

2. – AtlFast3

AtlFast3 is the ATLAS fast simulation system, which is already in production for LHC Run 3 [3]. It combines two fast simulation tools:

- *FastCaloSim*, based on parametrisations of the longitudinal and the lateral development of showers in the calorimeter;
- *FastCaloGAN*, based on Generative Adversarial Networks (GANs) [8].

The reference for evaluation of performance is the full simulation system Geant4. Depending on particle type, energy and pseudorapidity interval, *AtlFast3* employs the spe-

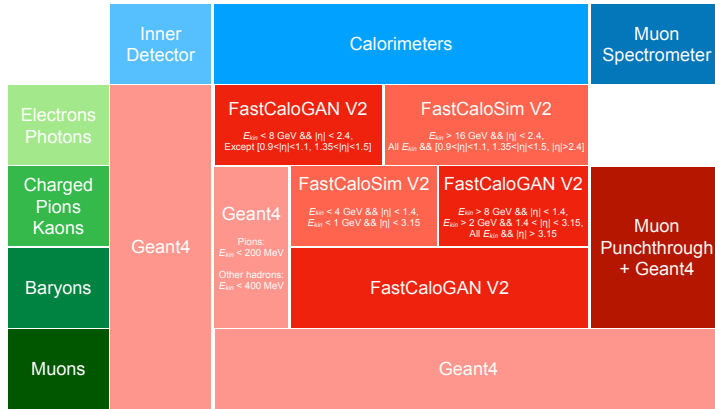


Fig. 2. – Fast simulation tools invoked by *AtlFast3* depending on particle type, energy and pseudorapidity interval [9]. The chosen tool is the one returning the most similar simulation to the one of Geant4. Geant4 keeps being used in specific cases.

cific simulation that gives the best results in comparison to Geant4 original distributions, as shown in fig. 2 [9]. Geant4 keeps being used for specific cases, as shown in the same figure.

A great improvement in performance is observed, as AtlFast3 runs simulation between 3 and 15 times faster than Geant4, with the greatest improvements being observed for processes with the highest energy particles [10].

2'1. FastCaloGAN. – The tool uses the Wasserstein GANs with a gradient penalty (WGAN-GP) term in the loss function of the discriminator, providing good performance and training stability [11]. It is based on the simultaneous training of two neural networks:

- a *generator*, that aims at generating samples as similar as possible to Geant4 datasets;
- a *discriminator*, that aims at distinguishing Geant4 data from the ones produced by the generator.

The two neural networks try improving during each iteration. Once an equilibrium is reached between the two, the FastCaloGAN generator simulates calorimeter response much faster than Geant4, preserving good accuracy at the same time, as shown in fig. 3.

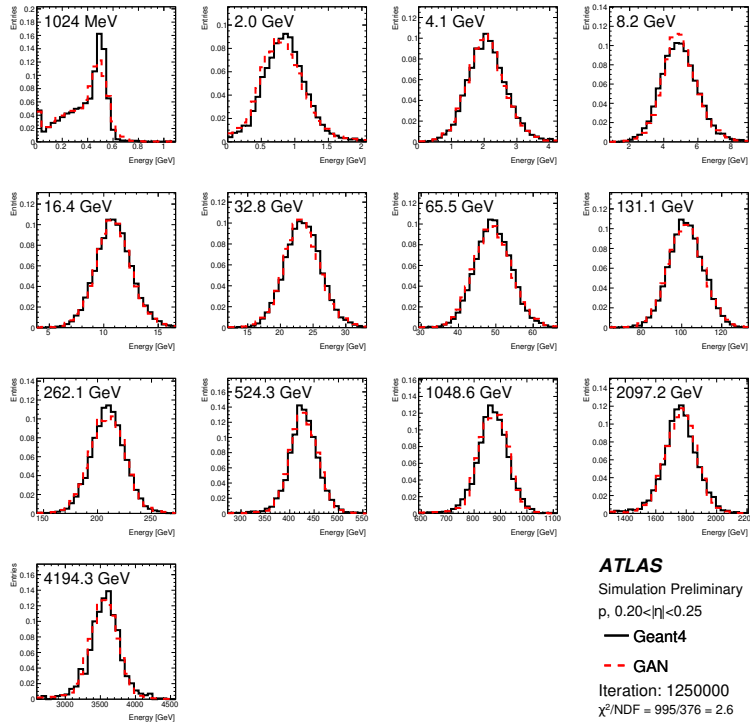


Fig. 3. – Comparison between full simulation (*Geant4*) and fast simulation (*GAN*) for the sum of the energy in all voxels for single protons generated at the calorimeter surface, with absolute value of pseudorapidity η between 0.20 and 0.25 [12]. Remarkable agreement is observed.

TABLE I. – *BoloGANtainer performance on tested clusters. For pions one single GAN was trained for all incident energy values, while for photons two GANs were, one for energies below or equal to 4 GeV, one for energies above. NDF is the number of degrees of freedom.*

Resource	Type and Owner	Hardware and Software	Pion Results	Photons Results
LXBATCH	CERN batch system, reference cluster	CentOS 7 (for used nodes), CVMFS, HT-Condor, NVIDIA V100 GPUs	Runtime: 12 h $\chi^2/\text{NDF} \sim 2$	Runtime: 30-31 h $\chi^2/\text{NDF} \sim 5$
Leonardo [13]	The 6 th most powerful cluster in the TOP500 ranking [14] at CINECA in Bologna, Italy	RHEL 8.7, no CVMFS, SLURM, NVIDIA A100 GPUs, isolated nodes	Runtime: 6-7 h $\chi^2/\text{NDF} \sim 2$	Runtime: 10-11 h $\chi^2/\text{NDF} \sim 5$

3. – BoloGANtainer

Usage of FastCaloGAN requires its GANs to be trained, but such training requires a large amount of resources. A possible solution involves distributing this task on other resources than the ones commonly used at CERN (the CERN batch system LXBATCH and the Worldwide LHC Computing Grid).

BoloGANtainer makes it possible for FastCaloGAN training to run on other resources as just described, providing resource saving and also an additional performance improvement if this distribution makes use of cutting-edge devices like supercomputers. BoloGANtainer is based on a container developed with Apptainer and on the official ATLAS CentOS 7 image, which replicates the operating system and software of the CERN batch system; it requires CUDA and CuDNN libraries for GPU usage. The rest of required software is installed directly into the container, in order for that to be independent from the system on which it is deployed.

BoloGANtainer has been deployed on the clusters shown in table I. Training on Leonardo (fig. 4) runs in about half the time needed on LXBATCH; remarkable acceleration is observed thanks to NVIDIA A100 GPUs, showing how the usage of supercomputers brings great advantage.

This work shall be further developed with distribution on other resources than the ones on which BoloGANtainer has already been tested (including cloud resources), architectures (ARM) and for more particle types. Code optimisation shall also be investigated, both as a general improvement and to employ multi-core and multi-GPU nodes in a more efficient way.

4. – Conclusion

Detector simulation takes the largest load of the computational resources of the ATLAS experiment at the LHC, with a major fraction taken by calorimeter simulation. As this burden is relevant and set to increase in view of HL-LHC, two solutions have been presented: ML-based fast simulation systems, with AtlFast3 being the one used in the ATLAS Collaboration, and distribution of the training of fast simulation ML systems

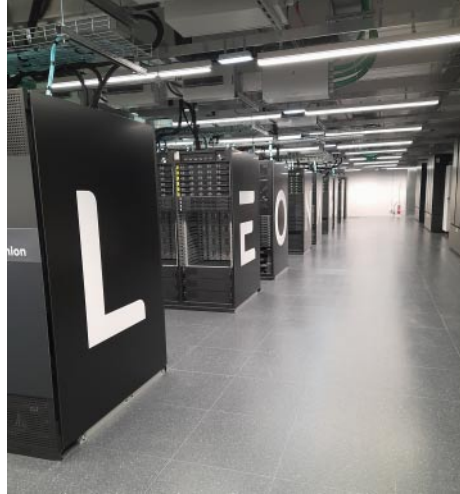


Fig. 4. – The Leonardo supercomputer at CINECA in Bologna, Italy. [15]

on other resources than CERN commonly used ones, for which BoloGANTainer has been introduced. Their good results in performance and accuracy significantly reduce the load on computing resources and allow for faster analyses.

REFERENCES

- [1] THE ATLAS COLLABORATION *et al.*, *JINST*, **3** (2008) S08003.
- [2] THE ATLAS COLLABORATION, *ATLAS HL-LHC Computing Conceptual Design Report*, CERN-LHCC-2020-015 (2020).
- [3] THE ATLAS COLLABORATION, *Comput. Softw. Big Sci.*, **6** (2022) 7.
- [4] ALLISON J. *et al.*, *Nucl. Instrum. Methods Phys. Res. A*, **835** (2016) 186.
- [5] ALLISON J. *et al.*, *IEEE Trans. Nucl. Sci.*, **53** (2006) 270.
- [6] AGOSTINELLI S. *et al.*, *Nucl. Instrum. Methods Phys. Res. A*, **506** (2003) 250.
- [7] The Worldwide LHC Computing Grid website, wlcg.web.cern.ch.
- [8] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A. and BENGIO Y., *Adv. Neural Inf. Process. Syst.*, **3** (2014) 11.
- [9] Public plots, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2024-004>.
- [10] Public plots, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2023-005>.
- [11] THE ATLAS COLLABORATION, *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks*, ATL-SOFT-PUB-2020-006.
- [12] Public plots, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2023-004/>.
- [13] Leonardo supercomputer website, leonardo-supercomputer.cineca.eu.
- [14] TOP500 Ranking - November 2023, top500.org/lists/top500/2023/11/.
- [15] LEONARDO - Il supercomputer al servizio delle geoscienze, ingv.it/it/stampa-e-urp/stampa/comunicati-stampa/5269-leonardo-il-supercomputer-al-servizio-delle-geoscienze (photograph by INGV - Istituto Nazionale di Geofisica e Vulcanologia).