Lamarr: LHCb ultra-fast simulation based on machine learning models

in *21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research* (ACAT 2022)

L. Anderlini[a], **M. Barbetti**[a,b], G. Corti[c], A. Davis[d], D. Derkach[e], N. Kazeev[e], A. Maevskiy[e], S. Mokhnenko[e], B. Siddi[f], Z. Xu[g]

[a]INFN Firenze, [b]University of Florence, [c]CERN, [d]University of Manchester, [e]HSE University, [f]University of Ferrara, [g]Université Clermont Auvergne

1. Motivation

During the LHC Run 2, the LHCb experiment has spent **more than 80% of the pledged CPU time** to produce simulated samples. The foreseen needs for Run 3 will far exceed the CPU resources available to the LHCb Collaboration, that is spending huge efforts in developing **faster options for simulation**, like the new Lamarr framework.

2. What is Lamarr?

The new *ultra-fast simulation* framework for LHCb is named **Lamarr** and is embedded within the LHCb simulation framework Gauss. Lamarr consists of a pipeline of (ML-based) **modular parameterizations** designed to replace both the physics simulation and the reconstruction steps.

- Compatibility with LHCb-tuned **generators** (e.g. Pythia8, Particle Guns);
- Promotion of generator-level particles to successfully **reconstructed** candidates;
- Possibility of submitting *Lamarr jobs* through the LHCb **distributed computing** middleware Dirac;
- Capability of producing datasets with the same **persistency** format as the LHCb physics analysis framework DaVinci.

3. ML-based parameterizations

Efficiencies: *Gradient Boosted Decision Trees* (GBDT) trained on simulated data to predict the fraction of accepted / reconstructed / selected candidates.

High-level quantities: Conditional *Generative Adversarial Networks* (GAN) trained on either simulated or calibration data to synthetize the high-level response of LHCb sub-detectors.

4. How to deploy models in Gauss?

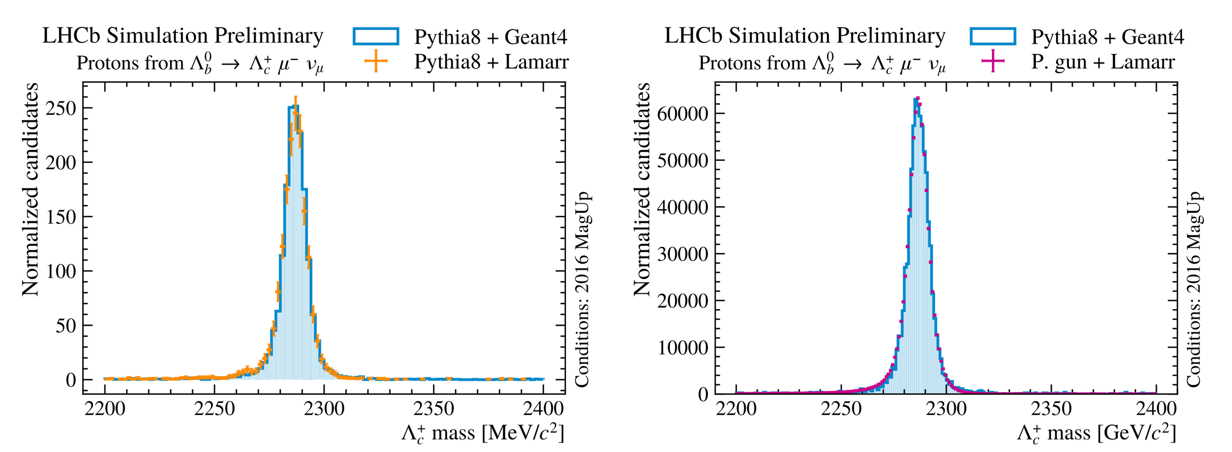
Best-performing parameterizations can easily replace specific modules without recompiling the whole pipeline using the deployment tool `scikinC`.

`scikinC` translates ML-based models to be dynamically linked to the main application (Gauss). In this way, parameterizations can be developed and released **independently**.

5. Validation campaign

Lamarr is currently under validation, comparing the distributions of the **analysis-level reconstructed** quantities parameterized with what obtained from *detailed simulation* for $\Lambda_b^0 \to \Lambda_c^+ \mu^- X$ decays with $\Lambda_c^+ \to pK^-\pi^+$.

- Abundant decay widely studied by LHCb, and that is part of *PID calibration samples*;
- It is described by a complex decay model including many feed-down modes;
- It provides examples for **muons**, **pions**, **kaons** and **protons** in a single decay mode.



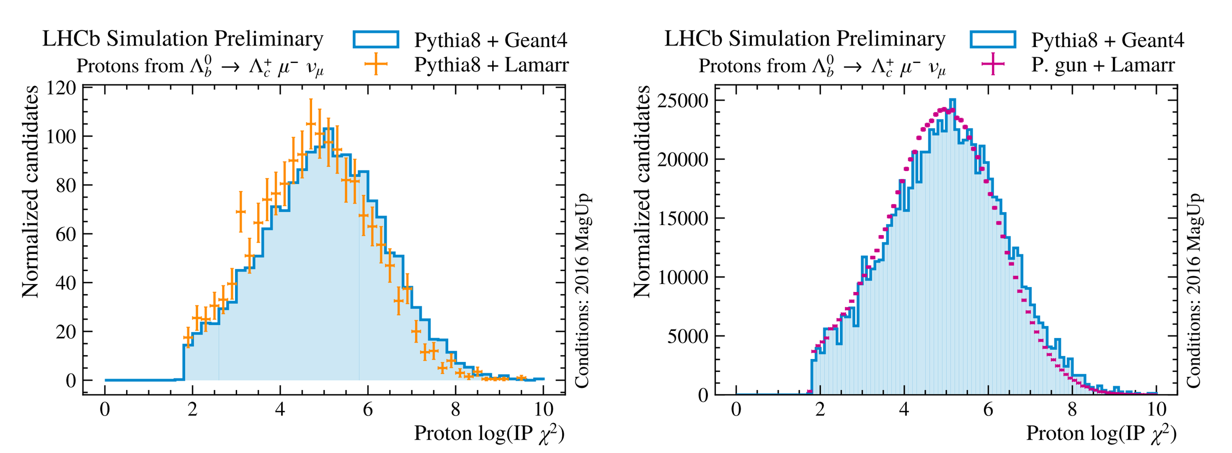
Λ_c^+ mass obtained from Pythia8 (left) and Particle Gun (right) generators by Lamarr against *detailed simulation*. Reproduced from LHCB-FIGURE-2022-014.

6. Results: Tracking system

The momentum and the point of closest approach to the beams of the generated particles **get smeared**: a GAN-based model predicts effects as *multiple scattering*, imperfections of alignment, calibration, and so on.

Another GAN-based model is used to predict the **uncertainties** associated to the track reconstruction. Track uncertainties, like the *impact parameter* (IP) χ^2, are crucial in LHCb to define the **(in)consistency** of trajectories with vertices.

Smeared quantities and track uncertainties can be used within the LHCb offline reconstruction programme Brunel to compute **higher-level quantities**, like the reconstructed mass.



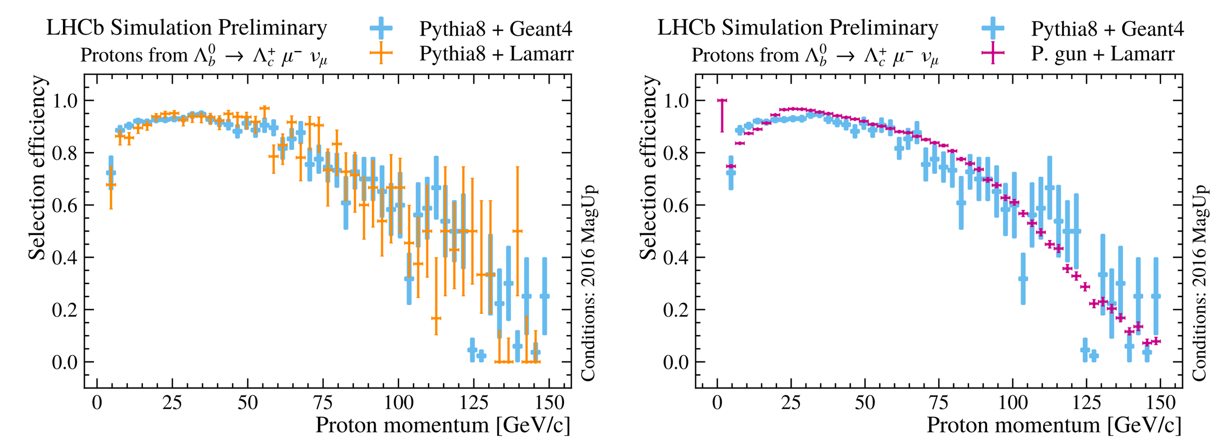
Proton IP χ^2 obtained from Pythia8 (left) and Particle Gun (right) generators by Lamarr against *detailed simulation*. Reproduced from LHCB-FIGURE-2022-014.

7. Results: PID system

Smeared *track kinematics* and *detector occupancy* are used by two sets of GAN-based models to parameterize the **high-level response** of the RICH and MUON systems.

Further GAN-based models are trained to reproduce the **higher-level PID classifiers** typically used in physics analyses, relying only on the input and the output of RICH and MUON parameterizations.

The adopted **stacked GAN structure** is designed to simulate both single-system detector response (RICH and MUON) and higher-level PID classifiers, enabling analysts to define new higher level classifiers based on the underlying basic quantities.



Proton ID efficiency obtained from Pythia8 (left) and Particle Gun (right) generators by Lamarr against *detailed simulation*. Reproduced from LHCB-FIGURE-2022-014.

8. Timing performance

Given the decay channel chosen for validation, the overall time needed for producing simulated samples passes from being dominated by **physics simulation** (Geant4) in *detailed simulation* to being dominated by **particles generation** (Pythia8) in *ultra-fast simulation*.

Preliminary studies show that Lamarr ensure a **CPU reduction of at least 98%** for the physics simulation phase. Further improvements in terms of timing can be achieved relying on Particle Guns instead of Pythia8.

Detailed simulation: Pythia8 + Geant4
1M events @ 2.5 kHS06.s/event ≃ 80 HS06.y

Ultra-fast simulation: Pythia8 + Lamarr
1M events @ 0.5 kHS06.s/event ≃ 15 HS06.y

Ultra-fast simulation: Particle Gun + Lamarr
100M events @ 1 HS06.s/event ≃ 4 HS06.y

9. Conclusions

Great progress has been made on developing a **fully parametric simulation** of the LHCb experiment, aiming to reduce the pressure on the CPU computing resources.

Model development, tuning and specialization will continue taking great advantage of **opportunistic GPU resources** made available to the LHCb Collaboration.

References

1. M. Borisyak and N. Kazeev, *Machine Learning on data with sPlot background subtraction*, JINST **14** (2019) P08020, arXiv:1905.11719
2. A Maevskiy *et al.*, *Fast Data-Driven Simulation of Cherenkov Detectors Using Generative Adversarial Networks*, J. Phys. Conf. Ser. **1525** (2020) 012097, arXiv:1905.11825
3. L. Anderlini, *Machine Learning for the LHCb Simulation*, arXiv:2110.07925
4. L. Anderlini *et al.*, *Towards Reliable Neural Generative Modeling of Detectors*, arXiv:2204.09947
5. L. Anderlini and M. Barbetti, *scikinC: a tool for deploying machine learning as binaries*, PoS **CompTools2021** (2022) 034
6. L. Anderlini *et al.*, *Lamarr: the ultra-fast simulation option for the LHCb experiment*, PoS **ICHEP2022** 233 (in preparation)