EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

The *flash-simulation* paradigm and its implementation based on Deep Generative Models for the LHCb experiment at CERN

Author: Matteo Barbetti *Supervisor:* Lucio Anderlini

INFN and Università degli Studi di Firenze

Abstract. The LHCb experiment is dedicated to precision measurements of hadrons containing b and c quarks at the Large Hadron Collider (LHC) at CERN. During the first two Runs of the LHC, spanning from 2010 to 2018, the LHCb Collaboration invested more than 90% of the computing budget to simulate the detector response to the traversing particles produced in heavy hadron decays. Since 2022, the LHCb experiment has relied on a renewed detector and a novel data-acquisition strategy designed to acquire data at a rate enhanced by a factor of ten. Enabling an equivalent increase in simulation production is a major challenge, requiring a technological shift and diversifying the simulation strategies for specific purposes. Data processing and data analysis technologies have been evolving quickly during the last ten years. New industrial standards and huge communities behind open-source software projects arose, transforming the landscape of computer science and data processing. The fast development of *Machine Learning* and *Cloud* technologies provides modern solutions to address challenges well known to the High Energy Physics community, operating distributed data processing software on the nodes of the *Worldwide LHC Computing Grid* for the last three decades. In this Thesis, I present a study to adopt these new technologies to evolve the LHCb simulation software using machine learning models trained on multi-cloud resources to parameterize the detector response and the effects induced by the reconstruction algorithms. The resulting detector simulation approach is known as *flash-simulation* and represents the most challenging and radical option in the landscape of opportunities to accelerate the detector simulation. To encode in a machine learning model the intrinsic randomness of the quantum interactions occurring within the detector, the experimental uncertainties, and the effect of missing variables, parameterizations are designed as *Generative Models*, and in particular as *Generative Adversarial Networks*. The LAMARR project, arising as the official flash-simulation option of the LHCb experiment, enables connecting the trained models in long data-processing pipelines to simulate various effects in the detection and reconstruction procedure. Pipelines can be deployed in the LHCb Simulation software stack by relying on the same physics generators as the other simulation approaches and serializing the results with the format of the official reconstruction software. In this Thesis, I address the most compelling challenges in the design of a flash-simulation solution, including the identification of a strategy to train and validate reliable parameterizations, the definition and distribution of heavy hyperparameter optimization campaigns through opportunistic computing resources, the combination of multiple parameterizations in a data processing pipeline, and its deployment in the software stack of one of the major experiments at the LHC. Future work will extend the validation of flash-simulation techniques for additional heavy hadrons, decay modes, and data-taking conditions, paving the way to the widespread adoption of flash-simulations and contributing to a significant decrease in the average computational cost of detector simulation.