A palette of fast simulations in LHCb

Mark Whitehead[*][†]

RWTH Aachen University

E-mail: mwhitehe@cern.ch

Generating simulated data samples currently dominates the LHCb CPU usage and this fraction will grow further as the data samples grow. Fast simulation options will become crucial in order to provide the samples analysts need in Run III and beyond. The available and under development fast simulation options at LHCb are discussed, including switching off parts of the detector simulation, reusing parts of the event, and fully parametrising the detector response and reconstruction.

---

[*]Speaker.

[†]On behalf of the LHCb collaboration.

1. Introduction

The LHCb simulation application, Gauss [1], is based on the Gaudi [2] framework and incorporates particle generation and transport in the detector. It depends on a number of external libraries, including Geant4 [3] for particle transport. A full description of the LHCb detector is given in Ref. [4] Producing simulated data samples currently uses most of the LHCb CPU resources, and in the LHCb upgrade era (LHCb Run III and beyond) it is expected to be more dominant. LHCb upgrade data samples will be a factor of five larger than those from Run II, so larger simulated samples will also be required. This necessitates a move towards producing a significant amount of simulated events with fast simulation techniques to produce the necessary statistics within the CPU resources available to the experiment. Disk space considerations are less critical, with several solutions deployed in over the past few years.

2. Fast simulation options

In the last few years several fast simulation options have been made available or are under development at LHCb as follows.

2.1 Simulation without RICH physics

The simulation of the optical photons in the two RICH detectors accounts for a significant part of the CPU time taken by Geant4: one option is to switch off the generation of the optical photons. The detector material is still included, therefore there is no impact on the downstream detectors. This approach speeds up the simulation by around 30% and leads to a small reduction in the disk usage as the PID information is not present. Figure 1 compares the nominal and RICHless samples to show that the remaining kinematic variables are unaffected. This option is already available in production and is suitable for studies that do not require hadronic particle identification variables from the simulation.
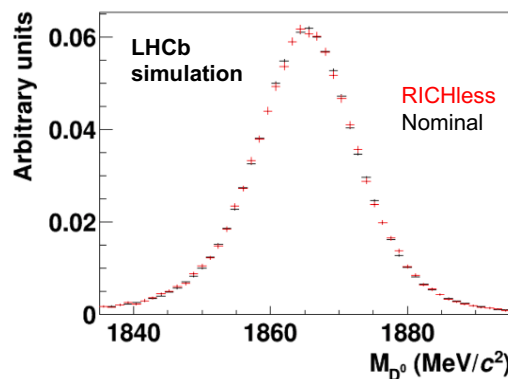
Figure 1: Comparison of the reconstructed mass of D^0 candidates generated using nominal and RICHless simulation samples.

2.2 Tracker only simulation

Tracker only simulation is a more aggressive option: the RICH detectors are switched off but remain in the material description, while the calorimeter and muon systems are entirely removed, including their material. This gives a speed up of around a factor of 10 and reduces the disk space required by roughly a factor of four. The tracker only option has already been used in production, and validated to perform well in the kinematic variables. These samples are suited to high statistic productions where no particle identification variables are needed in the simulated samples.

2.3 Signal particle guns

Signal particle guns are an alternative approach where just the signal particles are generated and simulated, including smearing of the primary vertex during the reconstruction phase. Particle gun simulations are between 50–100 times faster and use 100 times less disk space than nominal simulation. They are, however, only useful for specific studies and can't fully replace nominal samples due to detector occupancy effects. Particle gun samples are available in production.

2.4 Signal ReDecay

ReDecay [5] combines the full simulation and particle gun approaches. A complete event is generated and reused N times, while the signal decay is replaced each time. Since the rest of the event remains the same, it is only processed once by the Geant4 detector simulation which saves more time as N increases. Typically a value of $N \approx 100$ is used. The origin vertex and momentum of the signal particle remains the same to keep the correlation with the rest of the event, but its flight distance and all of the final state particle kinematics change. The ReDecay procedure is independent of the generator used and is compatible with other fast simulation options. It is typically 10–50 times faster than nominal simulation, depending on the decay modes requested, and retains the complexity of fully simulated events. ReDecay samples are available in production and a comparison of nominal and ReDecay simulation showing excellent agreement is given in Fig. 2. ReDecay samples should be suitable for most physics analyses alongside much smaller nominal samples.
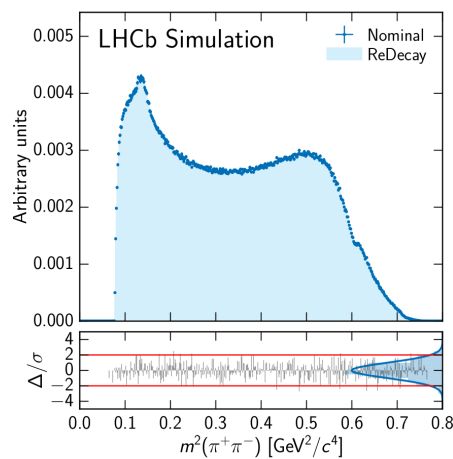
Figure 2: Comparison of the reconstructed mass squared $m^2(\pi^+\pi^-)$ from D^0 decays generated using nominal and ReDecay methods.

2.5 Fast simulation of the Calorimeter system

The calorimeter dominates the Geant4 CPU timing budget, so faster alternatives are under study. A generic infrastructure has been developed in the Geant4 interface of the current Gauss framework to allow the Geant4 transport for a given particle to be interrupted to inject a fast simulation component. Two approaches are under study, implementing frozen shower libraries and hit generation using generative adversarial networks, with a view to speed up the simulation by a factor of 3–10.

2.6 Fully parametric fast simulation

Work is ongoing to implement an ultra-fast, fully parametric simulation based on the DELPHES package [6], that parametrises both the detector response and the reconstruction. This is expected to be a crucial development to cope with the huge samples required by the end of Run III and beyond. The goal is to achieve a simulation $\mathcal{O}100$ times faster than the nominal detailed simulation. A functional prototype is integrated into the current Gauss framework and includes the tracking efficiency and resolution, primary vertex reconstruction and photon calorimetric objects. A first beta release is expected soon as well as including additional features such as covariance matrices for charged particle tracking, calorimetric objects for all electromagnetic particles and particle identification information. The parametrisation is fully experiment dependent and specific to a given data taking period. Proper calibration requires extremely reliable nominal simulation and precise data-driven measurements of the detector performance.

3. Summary

The role of fast simulation in LHCb is becoming more and more important to face the limitation of computing resources and the need for larger samples to keep up with the huge datasets available. Significant progress has been made in the last two years, with several options now available in production. The goal is to produce a flexible framework which allows users to select fast and detailed simulation options for different particle types and subdetectors, depending on their needs.

References

[1] M. Clemencic et al., The LHCb Simulation Application, Gauss: Design, Evolution and Experience, J. Phys. Conf. Ser. 331 032023 (2011).

[2] G. Barrand et al., GAUDI - A software architecture and framework for building HEP data processing applications, Comput. Phys. Commun. 140 (2001) 45. M. Clemencic et al. Recent developments in the LHCb software framework Gaudi, J. Phys. Conf. See. 219 (2010) 042006.

[3] S. Agostinelli et al., GEANT4 - A Simulation Toolkit, Nucl. Instrum. Methods Phys. Res. Sect. A, 506, 250 (2003). J. Allison et al. Geant4 developments and applications, IEEE Trans. Nucl. Sri. 53 (2006) 270.

[4] LHCb Collaboration, The LHCb detector at the LHC, JINST 3 (2008) S08005.

[5] D. Müller et al., ReDecay: A novel approach to speed up the simulation at LHCb, in preparation.

[6] J. de Faverau et al., DELPHES 3: a modular framework for fast simulation of a generic collider experiment, JHEP, 2(57), 2014.