New techniques for pile-up simulation in ATLAS

Tadej Novak^{1,*}, on behalf of the ATLAS Collaboration

¹Jožef Stefan Institute, Ljubljana, Slovenia

Abstract. The high-luminosity data produced by the LHC leads to many proton-proton interactions per beam crossing in ATLAS, known as pile-up. In order to understand the ATLAS data and extract physics results it is important to model these effects accurately in the simulation. As the pile-up rate continues to grow towards an eventual rate of 200 for the HL-LHC, this puts increasing demands on the computing resources required for the simulation and the current approach of simulating the pile-up interactions along with the hard-scatter for each Monte Carlo production is no longer feasible. The new ATLAS “overlay” approach to pile-up simulation is presented. Here a pre-combined set of minimum bias interactions, either from simulation or from real data, is created once and a single event drawn from this set is overlaid with the hard-scatter event being simulated. This leads to significant improvements in CPU time. This contribution will discuss the technical aspects of the implementation in the ATLAS simulation and production infrastructure and compare the performance, both in terms of computing and physics, to the previous approach.

1 Introduction

In addition to the hard-scatter pp interaction which causes the event to be triggered, the ATLAS detector [1] is also sensitive to proton-proton collisions in the same or surrounding bunch crossings. This is collectively known as “pile-up”. The average number of these interactions per bunch crossing has been increasing each year of Run 2 data taking, from 13.4 in 2015 to 38.3 in 2018. This is foreseen to rise further to more than 200 during high-luminosity LHC running, scheduled to start in 2026 [2].

As ATLAS subdetectors are also sensitive to not only the triggered bunch crossing, but the surrounding bunch crossings in time, pile-up is difficult to model accurately and represents a significant computational time in the detector Monte Carlo (MC) simulation. As digitisation CPU requirements are directly proportional to the number of soft collisions, this aspect needs to be improved.

The overlay technique is an alternative method of pile-up simulation, where pile-up can be estimated either with real data (data overlay) or simulated events (MC+MC overlay). This contribution will focus on the MC+MC overlay only. Compared to the current digitisation, simulated pile-up events are pre-mixed in an independent step and hard-scatter events are then overlaid on the merged background. Benefits and drawbacks of the method will be presented including the physics performance and an overview of the modifications needed to achieve it.

*e-mail: tadej.novak@cern.ch

2 MC+MC overlay method

The ATLAS simulation infrastructure [3] is used to produce Monte Carlo samples to be used in physics and performance studies. The simulation software chain is divided into three steps: generation of the event and immediate decays, simulation of the detector and physics interactions, and digitisation of the energy deposited in the sensitive regions of the detector into voltages and currents for comparison to the readout of the ATLAS detector. The simulation program is integrated into the ATLAS software framework, Athena [4]. Each stage can run individually, but the digitisation step is usually run together with the reconstruction step, where real physics objects such as jets, muons and electrons are reconstructed.

The ATLAS digitisation software converts the hits produced by the core simulation into detector response objects — “digits.” A digit is produced when the voltage or current on a particular readout channel rises above a pre-configured threshold within a particular time window. While most of the subdetectors record only that the threshold has been exceeded, some also contain information on the signal shape. The digits of each subdetector are written out as Raw Data Objects (RDOs).

Figure 1. Current digitisation stage workflow diagram. [5]

For any given hard scattering interaction the additional pile-up interactions must be included in a realistic model of detector response. They are simulated separately at the event generation and simulation stages. At the digitisation step, hits from the hard scattering are combined with those from pile-up before the detector response is calculated. Most subdetector responses are affected also by interactions from neighbouring bunch crossings, up to 32 before and 6 after the triggering one. Taking the average value of pile-up $\langle \mu \rangle = 34$ for Run 2 data taking, more than 1000 so-called “minimum bias” events need to be selected at random and processed. The workflow overview is presented in Fig. 1. The current method suffers from two main shortcomings: random event selection causes large random I/O, which can additionally also damage hard drives, and the whole digitisation stage takes a lot of time for large μ values, since the digitisation of pile-up events is repeated for each hard-scatter process.

A new MC+MC overlay method is proposed to replace the current digitisation, outlined in Fig. 2. An additional step is introduced where pile-up events are pre-mixed. A standard pile-

Figure 2. MC+MC overlay workflow diagram. [5]

up digitisation is run using zero-hard-scatter events (e.g. single neutrinos) where additional digits are stored to avoid losing information. In the future, only minimum bias events will be required for this step. Each simulated hard-scatter event is then digitised and overlaid on pre-mixed pile-up digits at the overlay stage.

The MC+MC overlay implementation is based on the data overlay technique that was already used for heavy-ion collisions simulation [6]. The basic infrastructure was already in place but it had to be adapted for MC background instead of data. Proper event information and metadata propagation is now ensured independent of the input type. Particle and detector truth information otherwise not present in data needed to be copied over and merged if needed. To ensure better precision, pre-mixed digits are now stored as floats for some of the subdetectors (e.g. the Tile Calorimeter). The new method is also transparently integrated in the digitisation and reconstruction chain to allow a smooth migration.

Figure 3. Average CPU time per event comparison between the standard digitisation (black circles) and the MC+MC overlay (red crosses) as a function of the number of proton-proton collisions per bunch crossing (μ). Horizontal error bars represent the width of the bin where a flat distribution of μ has been used. Vertical error bars represent the standard deviation of multiple measurements of the CPU time. [5]

The new method has two main benefits: the background dataset only needs to be digitised once per production campaign; and CPU and I/O requirements are also much lower and have almost no dependence on μ as seen in Fig. 3. There are also a few drawbacks. Two sub-threshold signals from different events causing a signal above threshold are now lost, which mostly affects the Inner Detector. The inner tracking detector covers the pseudorapidity range $|\eta| < 2.5$ and consists of silicon pixel, silicon micro-strip, and transition radiation tracking detectors. Additionally, different datasets can have an identical set of merged pile-up events, the effects of which still need to be investigated.

Pre-digitising of the pile-up RDOs together with the MC+MC overlay method is only beneficial if the events are reused. Bookkeeping is needed to ensure unique pile-up events for related samples, which are grouped in the Production System. For example, the same background processes produced using different MC generators or their systematic variations can use the same pool of pre-mixed events. Also signal samples from different analyses could reuse the background events.

Still, approximately 1 billion pre-mixed RDOs will be needed, when the new method will be used in production, estimated to require 4 PB of disk space per copy. As only one input file is needed per overlay job it can be requested and copied on demand. Thus only 2 or 3 copies will be needed globally. This will not present a large increase in needed storage since the current pileup simulation requires 80 copies of the minimum bias dataset, which at 40 TB each give rise to 3.2 PB in total.

3 Physics performance

To ease the migration to the new method and to reduce the amount of work needed for physics analyses, the requirement for MC+MC overlay is to give statistically compatible results to the standard digitisation. The current implementation is very close to nominal physics performance and the last remaining issues are being fixed.

Figure 4. Track reconstruction efficiency (left) and muon momentum resolution (right) comparisons between the standard digitisation (black circles) and the MC+MC overlay (red crosses) in simulated $t\bar{t}$ events. All primary charged particles with $p_T > 500$ MeV, $|\eta| < 2.5$ and a production radius < 110 mm are selected. The momentum resolution is computed using the difference between the reconstructed track p_T and the actual particle p_T . [5]

Tracking is especially affected by pile-up but good track reconstruction efficiency is preserved as shown in Fig. 4. The Semi-Conductor Tracker (SCT) is particularly sensitive to events from earlier bunch crossings. No hit in the channel in the previous bunch-crossing is explicitly required at readout. To avoid losing signals below threshold, all digits are stored regardless of previous hits at the pre-mixing stage. This yields performance comparable with standard digitisation (Fig. 5).

The Transition Radiation Tracker (TRT) energy deposits in drift tubes from pile-up and the hard-scatter event cannot be directly added together, because the information is stored with coarser granularity in the pre-mixed RDOs. TRT high-threshold drift circles are consequently incorrect. Besides tracking, they are also used for particle identification. The solution would be to store the energy and timing information in the RDOs, significantly increasing the file size. Instead, a detector occupancy based correction to the high-threshold hits in the TRT was implemented, tuned separately for electrons and non-electrons (Fig. 6), because electrons have higher probability to exceed the high threshold than other particles [7]. Although a perfect match cannot be achieved, TRT tracks have a lower weight in the tracks fits at high pile-up. Remaining differences then only affect the fit quality but not the fit results.

The Liquid Argon and the Tile Calorimeter also show good performance. Calorimeter overlay is performed at the digit level. During pre-mixing, Tile Calorimeter digits are now stored as floats to reduce rounding errors when the digits are merged. Good overall agreement for jets and electrons is observed and energy measurements are consistent between the methods as displayed in Fig. 7.

Figure 5. Distributions of the average number of SCT clusters (left) and holes (right) on tracks produced by muons from simulated $t\bar{t}$ decays as a function of the number of proton-proton collisions per bunch crossing (μ). The standard digitisation (black circles) is compared to the MC+MC overlay (red crosses). A hole is defined as a missing cluster in a sensitive detector area. [5]

Figure 6. Distributions of the average number of TRT high-threshold drift circles for tracks produced by electrons (left) and muons (right) from simulated $t\bar{t}$ decays as a function of the number of proton-proton collisions per bunch crossing (μ). The standard digitisation (black circles) is compared to the MC+MC overlay (red crosses). [5]

Figure 7. The total deposited energy distributions in the electromagnetic barrel of the Liquid Argon Calorimeter from simulated dijet events. The standard digitisation (black circles) is compared to the MC+MC overlay (red crosses). [5]

Other detector components also show good performance giving properly reconstructed physics objects – jets, photons, electrons, muons, and tau jets. A special calorimeter-based trigger system overlay had to be implemented for MC. Both the hardware-based Level-1 trigger and the software-based High Level Trigger match well the standard simulation results.

4 Conclusions and outlook

MC+MC overlay is a promising method of pile-up simulation which will help ATLAS computing cope with the ever increasing number of simulated proton-proton collisions. The CPU requirements and the stress on storage is much lower compared to the current method used in production. This will be especially useful in the future when the number of collisions in the same bunch-crossing will increase.

Overlaying hard-scatter events on background events at the digit level does not present significant differences to the standard digitisation method. Although precision is lost due to sub-threshold effects, CPU and storage requirements are much lower than using simulated hits directly. A similar approach has also been successfully used by the CMS collaboration [8].

Overlay is close to being validated for ATLAS physics. After promising production-grade tests it could be used in the simulation chain in 2019, after the remaining open issues are fixed. Data overlay could also be considered for pp collisions which would eliminate the need for CPU-expensive pile-up pre-mixing. Long-term plans for the future of overlay include making the code thread-safe and better integration between the data and MC+MC overlay. The integration with ATLAS fast simulation is also a must for the future of the ATLAS simulation infrastructure.

References

- [1] ATLAS Collaboration, JINST **3**, S08003 (2008)
- [2] ATLAS Collaboration, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, CERN-LHCC-2012-022 (2012), <https://cds.cern.ch/record/1502664>
- [3] ATLAS Collaboration, Eur. Phys. J. **C70**, 823 (2010), 1005.4568
- [4] ATLAS Collaboration, *ATLAS computing: Technical design report* (2005)
- [5] ATLAS Collaboration, *MC+MC Pile-up Overlay Performance Plots* (2018), <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2018-002/>

- [6] A. Haas, ATLAS Collaboration, Journal of Physics: Conference Series **898**, 042004 (2017)
- [7] ATLAS Collaboration, Tech. Rep. ATL-INDET-PUB-2014-001, CERN, Geneva (2014), <https://cds.cern.ch/record/1669603>
- [8] D.J. Lange, M. Hildreth, V.N. Ivantchenko, I.O. and, Journal of Physics: Conference Series **608**, 012056 (2015)