Simulation of Pile-up in the ATLAS Experiment

Zachary Marshall for the ATLAS Collaboration
Lawrence Berkeley National Laboratory and University of California, Physics Division, MS50B-6227, 1 Cyclotron Road, Berkeley, CA 94720, United States of America
E-mail: zach.marshall@cern.ch

Abstract. The high luminosity of the LHC results in a significant background to interesting physics events known as pile-up. ATLAS has adopted two independent methods for modeling pile-up and its effect on analyses. The first is a bottom-up approach, using a detailed simulation of the detector to recreate each component of the pile-up background. The second uses specially recorded data events to emulate it. This article reports on the experience using both of these methods, including performance considerations, for simulating pile-up in ATLAS.

1. Introduction
Proton-proton collisions in addition to the collision of interest, collectively referred to as “pile-up,” present a serious challenge to physics analyses at the Large Hadron Collider (LHC). In the ATLAS detector [1], many of the subsystems have sensitivity windows longer than 25 ns, which is the interval between proton-proton bunch crossings. As a result, every physics object is affected by pile-up in some way, from additional energy contributions in jets to the mis-reconstruction of background as high-momentum muons. During 2011, the number of proton-proton collisions per bunch crossing increased from 5 to 15, and during 2012 the number increased from 10 to almost 35. During the next high-energy run of the LHC in 2015, it is expected that the number of collisions per bunch crossing will peak at values over 50 and may be as high as 80. Thus, understanding and modeling this background is critical for performing analyses in ATLAS.

There are five components of the pile-up background:

- In-time pile-up: additional proton-proton collisions occurring in the same bunch-crossing as the collision of interest;
- Out-of-time pile-up: additional proton-proton collisions occurring in bunch-crossings just before and after the collision of interest. When detectors are sensitive to several bunch-crossings or their electronics integrate over more than 25 ns, these collisions can affect the signal in the collision of interest;
- Cavern background: the gas of neutrons and photons inundating the cavern during a typical run of the LHC. These mostly contribute random hits in the muon system.
- Beam halo events: the effect of protons from a bunch scraping against an upstream collimator. The scraping results in sprays of muons running approximately parallel to the beam-line.
- Beam gas events: collisions between the proton bunch and residual gas inside the beam-pipe. These generally occur off-center in the detector.
There are two basic approaches to modeling these backgrounds. The first, described in Section 2, is an attempt to simulate each component of the background individually. These components are then added on top of signal events prior to the conversion of energy deposits to detector signals in a step known as “digitization.” The second, described in Section 3, is an attempt to use specially-collected data to model the background. Each approach has advantages and disadvantages; alone, neither is adequate for all applications. Performance issues with each method are discussed in Section 4. Future plans for pile-up simulation improvements are outlined in Section 5.

2. Component-wise simulation
Each component of the pile-up described in Section 1 can be individually simulated. This allows the simulation of future detector layouts that have not been used for any data-taking, as well as the direct simulation of high-luminosity scenarios that have not yet been realized at the LHC. In addition, it is possible to trace particles through the simulation, providing an easy route to understanding background origins. However, this method is deeply dependent upon an accurate detector description and Monte Carlo (MC) modeling of particle spectra. Significant work is also required to incorporate the changing detector conditions over the course of a data-taking run. Transient effects, for example detector components that have intermittent failures, can be difficult to include except in an average way.

Each of the above components can be overlaid at configurable rates on the signal events simulated in ATLAS. Typical events simulated for 2012 data analyses included at least in-time and out-of-time pile-up. These events are overlaid as additional energy deposits in each detector before the conversion from energy to detector signal (e.g. voltages and times) is made. Thus, all saturation effects, masking effects, etc. are taken into account.

2.1. In-time pile-up
Additional proton-proton collisions occurring during the same proton-proton bunch-crossing present a significant background for all physics objects. These additional collisions are generated with Pythia8 [2], including single-, double-, and non-diffractive components, and are passed through the standard ATLAS detector simulation [3]. They are then overlaid at a rate to match that of the data. Variations are allowed to model both the bunch structure of the LHC beam and the bunch-to-bunch variations.

In order to save both memory and disk space, much of the history of the MC record, known as “truth,” is removed from the background files. This allows considerable disk space savings, as well as a reduction of CPU time when reading the events from disk.

The inclusion of unique minimum-bias events for every signal event would require a prohibitive amount of disk space. Therefore, the sample is divided into events with and events without an anti-$k_t$ jet with $p_T > 35$ GeV built from the stable particles provided by the MC event generator. The events without such a jet are re-used regularly, and the events with such a jet are only used once for a given dataset. This prevents the accidental creation of excessive fluctuations in the reconstructed jet transverse momentum spectrum that are caused solely by the repeated use of a background event with a high-$p_T$ jet.

2.2. Out-of-time pile-up
Additional proton-proton collisions occurring in proton-proton collisions just before or after the collision of interest can affect the detector response to the in-time collision. The effect of these additional interactions depends strongly on the detector technology, and therefore different time windows are used to include the appropriate level of background in each detector.

The liquid argon calorimeter, for example, has a bi-polar pulse-shape that makes it sensitive to $\sim 250$ ns after the collision of interest and many bunch-crossings before. Signals before the
collision of interest can result in a reduced total pulse height. Similarly, in tracking detectors like the transition radiation tracker, particles passing through a tube can mask, or “shadow,” particles from a collision in the next bunch-crossing passing through the same tube. The inclusion of these effects is critical to an accurate simulation of the detector response.

These collisions are treated similarly to in-time pile-up, overlaying a number of collisions that varies to model the bunch structure and bunch-to-bunch variations in the LHC. The sensitive time window is set individually for each detector, with the longest (±32 bunch crossings in the monitored drift tubes in the muon system) determining the total number of events that must be read from disk. Some detectors, like the pixel detector, are sensitive to no more than two crossings in each direction.

2.3. Cavern background

The cavern background comprises a gas of thermal neutrons and photons filling the ATLAS cavern during LHC operation. This background has a long lifetime (particles can propagate for seconds without interacting) and consists mostly of low-energy particles with kinetic energies at or below 1 MeV. Thus, the simulation of this background requires the inclusion of many effects that are not normally considered important for event simulation in the LHC. For example, this simulation is sensitive to the inclusion of boronated polyethylene cladding on the forward shielding, which must, therefore, be included in the simulation and to which the “standard” simulation is completely agnostic. The cavern itself is simulated, along with the detailed shielding in the forward region.

Two separate approaches to simulating this background exist in ATLAS. The first uses the FLUKA simulation engine [4, 5, 6] with a simplified detector geometry. Scoring volumes around the muon chambers record particle fluxes at their faces. The second uses the Geant4 simulation toolkit [7, 8] with the standard detector geometry, as used in ATLAS simulation. Thermal neutrons are recorded during simulation of the event. In both cases, the simulations must include high-precision models of low energy neutron physics that are too slow to be included in standard event simulation. Following this, the fluxes from FLUKA or the particles from Geant4 are fed back into standard detector simulation, where energy deposits in the muon system are recorded. These energy depositions are then “time wrapped” to emulate many turns of the LHC beam in a single pass. Without this time wrapping, it would be necessary to overlay many millions of these simulated cavern-background events on every signal event.

As the high-luminosity of some upgrade scenarios creates a large background, particularly for forward muons, this background is quite important. However, for typical events in 2012 it was not included as a part of the pile-up simulation. The detailed simulations that have been implemented have led to the understanding of the origin of some of the background flux in the muon system. In order to mitigate the background, special pieces of shielding were introduced between 2011 and 2012, and more shielding will be introduced before Run 2.

2.4. Beam halo events

The beam halo events originate from proton bunches scraping against collimators far upstream of the ATLAS detector. As such, the simulation of the interactions and their propagation to the ATLAS cavern is done by the LHC machine group [9]. The particles reaching ATLAS are mostly muons and can traverse the entire detector parallel to the beam-pipe. In doing so, the particles can leave background signals in the muon system and, if they pass horizontally through a piece of the calorimeter, can ionize sufficiently to produce significant signals there.

The rate of such events is typically tens of Hz and is asymmetric between the two sides of the detector because of the configuration of upstream collimators. However, as these events generally leave signals that can be specifically recognized (e.g. out-of-time signals), they can
be removed from analyses. Thus, they were not typically simulated for ATLAS signal events in 2012.

2.5. Beam gas events

Beam-gas events consist of collisions between protons in the LHC beam with residual gas in the beam-pipe inside of ATLAS. These highly asymmetric proton-carbon, proton-oxygen, proton-nitrogen, or proton-proton (hydrogen) collisions create collision vertices well outside of the standard interaction region, which is only 12 cm long. They also typically produce out-of-time signals that can be cut out from analyses. Therefore, they are also not typically simulated for ATLAS signal events.

3. Data overlay

Rather than simulating each component of the pile-up background, data can be collected during running as special “zero-bias” events. This method has the advantage that the background is determined in a truly data-driven way, and the dependence on MC event generator modeling of particle spectra is eliminated. If such data is collected throughout a run period, the detector conditions during the run can be trivially re-created as well. All components are included at the appropriate rates. However, this method cannot be used to simulate new detector configurations or additional subsystems that have not yet been used in data-taking.

In order to collect this data, a random trigger was used in 2012. Several Hz of such data were recorded in order to be able to fully reproduce the run conditions throughout 2012 with reasonable statistics. The data must not include zero-suppression in the calorimeters, as it is important to include any noise or trace of signal that might be present.

These recorded background events are then overlaid one-by-one on top of simulated signal events. The signal must be simulated with the same conditions as the background event in order to ensure detector consistency. This requirement is generally straightforward in the calorimetry but can produce difficulties in the inner detector, where module alignment may not be perfectly known. The simulation must avoid introducing overlaps between volumes that can be introduced when applying detector alignment conditions to the simulated detector. If a different alignment is used, then small shifts in hit positions in the inner detector can lead to dramatic changes in measured track momenta.

This method is extensively used in ATLAS for heavy ion event simulation. The enormous particle background in a heavy ion collision can be taken directly from the data, and an additional collision can be overlaid on top of it. This allows the embedding of hard di-jets, Z or W boson, and other interesting events within heavy ion collisions. Of course, such embedding requires that the position of the simulated collision be the same as that of the background event. Therefore, the positions of the collisions in the recorded zero-bias heavy ion collisions are passed to the detector simulation, and each background event has a signal event simulated specially for it.

4. Computing performance

The performance of pile-up digitization is best viewed as a function of the average number of additional proton-proton collisions per bunch crossing ($\langle \mu \rangle$). With increasing $\langle \mu \rangle$, the number of background events required by a single signal event increases. This can be dealt with in one of two ways. Either the events can be read into memory, resulting in a strong $\langle \mu \rangle$-dependence of the memory consumption of pile-up digitization, or the events can be flushed from memory and read back from disk, resulting in a strong $\langle \mu \rangle$-dependence of the CPU consumption of the pile-up digitization for I/O. For very high luminosity conditions, the memory requirements become prohibitive. Thus, the choice has been made to read events from disk more frequently and introduce a stronger $\langle \mu \rangle$-dependence in the CPU use.
The memory use for several configurations as a function of $\langle \mu \rangle$ is shown in Figure 1. The configurations labeled “MC12” (“MC14”) refer to the standard configuration used in 2012 (foreseen for 2014). The change between the “Algorithms” approach and the “PileUpTools” approach was a move from keeping a cache of background events in memory to loading only those that are actually needed. This change allows the jobs to run with $\langle \mu \rangle$ higher by 50 for the same memory consumption. The importance of such a change can be seen in the 32-bit jobs, where at high $\langle \mu \rangle$ the Algorithms jobs failed after reaching the memory limit of the machines. The CPU consumption per event for the same configurations is also shown in Figure 1. Changes to SLC6 and 64-bit running allowed faster running, though at the cost of somewhat increased memory consumption. The CPU consumption exhibits more erratic behavior than the memory consumption, partially due to threshold effects like the use of swap space rather than active memory.

5. Future work and plans

ATLAS plans to move in the coming run to an integrated simulation framework that will allow additional and advanced flexibility in detector simulation [10]. This includes several approaches to fast detector simulation. Although the detector simulation is currently the slowest part of the MC simulation chain, once it is made significantly faster it becomes important to improve the speed of the simulation of pile-up background and the reconstruction. In order to hasten the simulation of pile-up, the possibility of including only in-time pile-up, with an adjusted level to mimic the effect from out-of-time pile-up, is currently under investigation. For very fast simulations, it might be possible and practical to simulate additional proton-proton collision events on-the-fly, rather than saving them to disk. Alternatively, a parameterization of the effect of pile-up could be developed. However, such parameterizations are generally difficult to develop in a generic context.

The LHC operating conditions are continuing to evolve. The next run of the LHC is expected to include well over 50 proton-proton collisions per bunch-crossing, and some upgrade scenarios include as many as 200 additional collisions. Such an enormous amount of pile-up requires major performance improvements in both the pile-up digitization and reconstruction. Parallel processing approaches might help reduce the memory required for a common event cache, but the potential problems with event-re-use and I/O limitations are still not well understood.

References

[1] The ATLAS Collaboration 2008 JINST 3 S08003
[2] Sjostrand T, Mrenna S and Skands P Z 2008 Comput. Phys. Commun. 178 852–867 (Preprint 0710.3820)
[3] 2010 Eur. Phys. J. C 70 823–874 (Preprint 1005.4568)
[4] Fassò A et al. 2003 Computing in High Energy and Nuclear Physics 2003 (CHEP2003), La Jolla, CA, USA, March 24-28, 2003, (paper MOMT005), eConf C0303241 (Preprint hep-ph/0306267)
[5] Battistoni G et al. 2007 AIP Conference Proceeding 896 31–49 proceedings of the Hadronic Shower Simulation Workshop 2006, Fermilab 6–8 September 2006
[6] Fassò A, Ferrari A, Ranft J and Sala P 2005 Fluka: a multi-particle transport code CERN 2005-10 also INFN/TC_05/11, SLAC-R-773
[7] Agostinelli S et al. 2003 Nucl. Instr. and Meth. A 506 250
[8] Allison J et al. 2006 IEEE Transactions on Nuclear Science 53 270–278
[9] Bruening O et al. (eds) 2004 LHC Design Report vol v.1: The LHC Main Ring (Geneva: CERN) ISBN 9789290832249 also CERN-2004-003-V-1
[10] Debenedetti C 2014 This volume
Figure 1. The average memory use (top), in GB, for several configurations of the digitization, as a function of the average number of proton-proton collisions per bunch crossing (⟨µ⟩). Bottom, the CPU use per event, in ms, for the same configurations. Two jobs in the MC14 SLC6 configuration appear to have suffered from competition with other jobs in the batch node. For the detailed differences between the configurations, see the text.