Preparing ATLAS reconstruction software for LHC’s Run 2

Jovan Mitrevski on behalf of the ATLAS Collaboration
Excellence Cluster Universe, Ludwig-Maximilians-Universität München, Boltzmannstraße 2, 85748 Garching, Germany
E-mail: Jovan.Mitrevski@cern.ch

Abstract. In order to maximize the physics potential of the ATLAS experiment during LHC’s Run 2, the reconstruction software has been updated. Flat computing budgets required a factor of three improved execution time, while the new xAOD data format forced changes in the reconstruction algorithms. Physics performance was also made better. This paper presents an overview of the improvements made to the reconstruction software during the long shutdown of the LHC.

1. Introduction
Run 2 of the LHC will be a challenging software environment for the ATLAS experiment [1]. The High Level Trigger (HLT) accept rate will be 1 kHz, instead of 400 Hz prompt and about 150 Hz delayed as at the end of Run 1. The software should also be prepared for higher pile-up, and Fig. 1 shows that the per-event execution time increases as a function of the average number of interactions per bunch crossing [2]. Furthermore, the physics performance must not be compromised, but improved where possible. At the same time, we have a flat computing budget, so we cannot just buy more computing resources to be able to deal with the incoming data. Meeting these goals requires a factor of three reduction in processing time per event for reconstruction compared to what we had in Run 1. This paper presents the reconstruction software improvements for Run 2, both in computing and in physics performance.

2. ATLAS Detector
The ATLAS detector is a multi-purpose apparatus with a forward-backward symmetric cylindrical geometry and nearly 4π solid angle coverage. Closest to the beamline is the inner detector, consisting of pixel and microstrip trackers covering |η| < 2.51 and a transition radiation tracker (TRT) covering |η| < 2.0. For Run 2, an additional pixel layer, called the Insertable B-Layer (IBL), was added to increase tracking robustness in the face of module failures, tracking efficiency in high pile-up, and tracking precision [3]. The TRT uses straw tubes for discrimination between electrons and charged hadrons based on transition radiation, as well as for tracking.

1 ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, φ) are used in the transverse plane, φ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle θ as η = −ln tan(θ/2).
Figure 1. The time per event to reconstruct data events triggered by the presence of jets, missing transverse momentum, or tau leptons at the end of 2012 as a function of the average number of interactions per bunch crossing and software release. Release 17.2.7.9 was used in Run 1, while releases 19.0.3.3 and 19.1.1.1 incorporate some improvements. Plot from Ref [2].

The inner detector is located inside a thin superconducting solenoid that provides a 2 T magnetic field.

Outside the solenoid, a fine-granularity lead/liquid-argon (LAr) electromagnetic calorimeter measures the energy and position of electrons and photons in the region $|\eta| < 3.2$. A presampler, covering $|\eta| < 1.8$, is used to correct for energy lost by particles before entering the calorimeter. An iron/scintillating-tile hadronic calorimeter covers the region $|\eta| < 1.7$, while a LAr hadronic end-cap calorimeter covers $1.5 < |\eta| < 3.2$. In the forward region, $3.1 < |\eta| < 4.9$, LAr calorimeters with copper and tungsten absorbers measure both the electromagnetic and hadronic energy.

A muon spectrometer surrounds the calorimeter system, consisting of trackers, and three superconducting toroidal magnet systems each comprising eight toroidal coils. There are four types of trackers in the muon spectrometer, the monitored drift tubes (MDT) and cathode strip chambers (CSC) for precision measurement (optimized for different pseudorapidity ranges), and the resistive plate chambers (RPC) and thin gap chambers (TGC) for triggering and providing a measurement in a second coordinate orthogonal to the precision measurement. A number of additional chambers were added for Run 2 in order to improve the performance.

3. General Improvements
A number of general improvements were made to the software to speed up reconstruction [2, 4, 5]. The method to access the magnetic field strength was rewritten, reducing the call depth, reducing the number of unit conversions, and caching recent results in order to exploit the fact that magnetic field requests are often for locations close to those already requested. The CLHEP linear algebra library was replaced by Eigen [6], which by using expression templates significantly improves the performance. We also replaced GNU libm with the Intel math library, since benchmarks showed an improvement. Finally, newer, more optimizing versions of GCC [7] improved the reconstruction software execution time, as did compiling the software in 64-bit.
mode.

To improve the usability of the software for performing physics analyses, a new event data
model (EDM) was created for analysis-level objects, with the resulting output file format, called
xAOD [8], replacing the old AOD and D3PD formats. This required many changes to the
reconstruction software. Simplicity and harmonization were major themes in the new EDM,
which we also strived to follow in the reconstruction software. For example, we now have one
code to do isolation for electrons and muons instead of two.

4. Improvements in reconstruction domains

4.1. Tracking and vertexing

Tracking and vertexing changes were quite extensive. They are covered in more detail in separate
contributions [9, 10] and briefly summarized here.

Given how time-consuming tracking is in high pile-up situations, much effort was put towards
optimizing the tracking software. The Eigen migration and the magnetic field provider rewrite
were largely motivated by the need to speed up tracking, and in turn required the most changes
in the tracking software. Additionally, the tracking EDM was greatly simplified. A new triple-
template structure reduced the code size by more than 90% while keeping the same functionality.
The new EDM reduced the call depth and the number of dynamic casts, speeding up the
performance.

More algorithmically, the track seeding was optimized for high pileup, making use of the new
IBL. This leads to both higher purity seeds and faster execution. Since photon conversion is
the only user of TRT-seeded tracking, and TRT-seeded tracking is expensive in high pile-up
scenarios due to high TRT occupancy, in Run 2 TRT-seeded tracking will only be performed
in regions of interest (ROIs) created around electromagnetic calorimeter clusters. Improved
ambiguity resolution for Tracking In Dense Environments (TIDE) improves flavor tagging and
tau reconstruction and at the same time improves the execution time. Finally, tracking inside
of the calorimeter was improved, using a dense volume description of the calorimeter. The
extrapolation of tracks to the calorimeter is now cached so that it is not repeated by various
object reconstruction algorithms.

For vertexing, a new seeding algorithm based on imaging techniques was implemented. The
algorithm attempts to simultaneously identify all potential vertices in one bunch crossing using
the tracks as input, to be subsequently used as seeds to the vertex finding and fitting. It is more
robust to pile-up. The algorithm will continue to be developed, though it will not be used by
default at the start of Run 2.

4.2. Electron and photon reconstruction

The need to modify the reconstruction of electrons and photons to use the new analysis EDM
gave an opportunity to revisit and simplify the software, including it’s configuration, which was
refactored and simplified.

One of the major improvements in electron and photon reconstruction for the 2012 run was
that electromagnetic calorimeter clusters were made to seed electron-optimized pattern matching
and bremsstrahlung-aware refitting. As mentioned in the previous section, electromagnetic
calorimeter clusters now also seed the TRT-seeded tracking. The isolation calculation is now
harmonized with the muon isolation, using energy density pile-up corrections for isolation in the
calorimeter. Some effort was also devoted to making the offline trigger reconstruction match as
closely as possible the offline reconstruction.

During Run 1, it was found that an MVA-based energy calibration [11] performed better than
the default, calibration-hits method [12], so for Run 2, an MVA-based calibration will become
the default.
The TRT is filled with xenon gas in order to use transition radiation to differentiate electron from other tracks. However, parts of the TRT have started leaking to the point where it is not feasible to keep them filled with xenon gas. Instead, they are filled with argon gas, at the expense of not being able to detect transition radiation. The electron and photon reconstruction and identification was made to be tolerant to the loss of transition radiation for parts of the TRT.

4.3. Muon reconstruction
For Run 2, the muon software will run a unified reconstruction chain, with simplified steering and configuration. The Hough transform has a new tuning, and there is in addition new TGC-seeded segment finding, as a complement to the MDT-seeded segment finding. A large effort was devoted to improve the energy loss tuning in the calorimeter, with a goal of having 10 MeV precision for the energy loss. A new calorimeter geometry description was created, and detailed energy loss tuning was performed.

Additionally, the muon software now provides more support for “exotic” reconstruction. Vertexing within the muon system is now run by default to make analyses depending on displaced vertices easier to perform. In the MDT, dE/dx information is now stored in the xAODs, as is more information for low-β particles. Finally, a new identification method is defined for high-pT muons, which is useful for exotic searches.

4.4. Jet reconstruction
The jet software is now “dual-use,” meaning that it can run both within and outside of Athena [13], ATLAS’s reconstruction and analysis framework. Fig. 2 gives a diagram of how this is accomplished. Dual-use tools need to inherit from a new AsgTool base type and IAsgTool interface. When compiled with Athena, this tool acts as a native Athena tool, while when it is compiled outside of Athena, there is special support added to provide the missing functionality. The full jet reconstruction can now be performed outside of Athena. More information about dual-use tools can be found in Ref. [14].

Algorithmically, there are the following changes. A new version of FastJet [15] (currently 3.1.1) is now used, resulting in an improved execution time. The interface to FastJet is also improved, making it easier to use external tools with it. Given that boosted scenarios are becoming more and more important as our mass reach goes further, substructure calculations, including subjects, are now well integrated in the software for easier analysis.

Ghost association is a method to associate particles to jets by giving them negligible momentum and clustering them within the jets. It is now integrated in the reconstruction flow and EDM to easily ghost-associate any particle of interest. This in particular is used by flavor tagging.

4.5. Missing Transverse Momentum reconstruction
Missing transverse momentum (and its magnitude $E_T^{\text{miss}}$) is calculated as:

$$E_T^{\text{miss}} = E_T^{\text{miss}}(e) + E_T^{\text{miss}}(\gamma) + E_T^{\text{miss}}(\tau) + E_T^{\text{miss}}(\text{jet}) + E_T^{\text{miss}}(\mu) + E_T^{\text{miss}}(\text{soft})$$

where each component is the negative sum of the momenta of the objects of that type in the transverse plane. $E_T^{\text{miss}}(e)$ is the negative sum of the transverse momenta of electrons, and similarly the other components are for photons, taus, jets, muons, and a soft-term for energy not associated with any of the objects. For Run 2, both a calorimeter-based and a track-based $E_T^{\text{miss}}(\text{soft})$ will be provided. The track-based term is more resistant to pile-up.

Double-counting by including the same object in multiple terms must be avoided. The order of the terms in Eq. 1 indicates the default precedence, but the object selection criteria is often
Figure 2. The “dual-use” class diagrams. In orange are the “dual-use” AsgTool base class and interface, from which the concrete tool (in blue) inherits. The base class is written so that its inheritance structure depends on whether the code is compiled in Athena or not. Athena classes are shown in green, while in pink are the replacement classes used when the code is compiled outside of Athena.

Figure 3. The $E_T^{\text{miss}}$ resolution as a function of the truth vertex multiplicity for simulated $Z'$ events.

modified. In Run 1 recomputing the $E_T^{\text{miss}}$ with a different selection was a computationally expensive and error-prone task because the overlap removal between the objects had to be redone. For Run 2 the overlap is now encoded in one structure per jet type, thus allowing for an easy customization of the $E_T^{\text{miss}}$ calculation. Analysts can easily and efficiently compute the $E_T^{\text{miss}}$ terms with an arbitrary object selection and avoid double-counting or missing objects.

At the end of Run 1, there was significant effort devoted to pile-up suppression. The methods continue to be refined. Fig. 3 shows the good pile-up performance.
obtained by varying the cut on the jet weight obtained from $p_T$ impact parameter measurements with information about secondary vertex properties in the jet. This tagger combines the rather independent information from the tagging variables.

Figure 27. Improved light jet rejection for the benchmark.

The instantaneous luminosity during Phase I is expected to exceed $10^{34}$ cm$^{-2}$ s$^{-1}$, which demands that hadron lifetimes and decay properties be very significant improvements in light jet rejection for the benchmark.

Table 6. Rejection of light jets in $\bar{t}t$ events without pileup for the benchmark.

- Run 1: MV1/
- Run 2: MV2c20

Figure 4. Light jet rejection versus $b$-tagging efficiency with and without the IBL. Plot from Ref [3].

Figure 5. Diagrams showing the structure of the multivariate $b$-taggers in Run 1 (MV1) and Run 2 (MV2). In blue are algorithms, in green likelihoods, and in red multivariate algorithms.

4.6. Flavor tagging

Flavor tagging benefits from the new IBL, as can be seen by Fig. 4. A new multivariate tagger, MV2, was created for Run 2, providing 30-50% better light jet rejection at the same $b$-tagging efficiency compared to the multivariate tagger, MV1, used in Run 1. At the same time, the MV2 is simpler in structure than MV1, as shown in Fig. 5.

Flavor tagging in boosted and high-$p_T$ conditions is becoming more important as higher mass ranges are probed. Therefore, there are two new taggers for such conditions. One is based on $b$-tagging track-jets and matching those track-jets to calorimeter jets [16]. In this way, the track-jets can be small in radius, which is desirable for pile-up tolerance, while there is flexibility to use large, groomed jets, which are preferable in certain boosted scenarios. Another tagger is the new multivariate MVb tagger for boosted scenarios [17]. The training in this tagger is modified to better match boosted scenarios.
4.7. Tau reconstruction
In addition to migrating the tau reconstruction to the new analysis EDM, there has been work towards reconstructing the individual decay products as a way to improve the energy resolution, position resolution, and identification efficiency of hadronic taus. Known as substructure reconstruction, the method uses both tracking and calorimeter information to identify and measure the charged and neutral pions in a particle-flow-inspired method. Care must be taken to assign the energy associated with each constituent and to not double-count when determining the full tau properties. The improvement in the tau energy resolution is promising.

5. Conclusion
As can be seen in Fig. 6, the reconstruction execution time improved by a factor of four, better than the factor of three goal that was set at the beginning of the long shutdown. Physics performance has also improved, with more pile-up tolerant tracking and vertexing, better calibrated electrons and photons, more precisely measured muons, more user-friendly jets and $E_T^{\text{miss}}$, and better performing flavor tagging and tau reconstruction.

References
[1] ATLAS Collaboration 2008 JINST 3 S08003
[2] ATLAS Collaboration 2014 Timing improvements for the ATLAS offline reconstruction for run-2 Tech. Rep. ATL-SOFT-PUB-2014-004 CERN Geneva URL http://cds.cern.ch/record/1955923
[3] ATLAS Collaboration 2010 ATLAS Insertable B-Layer Technical Design Report Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19 CERN Geneva URL http://cds.cern.ch/record/1291633
[4] Langenberg R et al. (ATLAS) 2014 Journal of Physics: Conference Series 513 022018
[5] Chauhan N et al. (ATLAS) 2014 Journal of Physics: Conference Series 513 052022
[6] Eigen URL http://eigen.tuxfamily.org
[7] GCC URL http://gcc.gnu.org/
[8] Buckley A et al. (ATLAS) 2015 Journal of Physics: Conference Series these proceedings
[9] Salzburger A (ATLAS) 2015 *Journal of Physics: Conference Series* these proceedings
[10] Borissov G *et al.* (ATLAS) 2015 *Journal of Physics: Conference Series* these proceedings
[11] ATLAS Collaboration 2014 *Eur.Phys.J.* C74 3071 (*Preprint* 1407.5063)
[12] ATLAS Collaboration 2012 *Eur.Phys.J.* C72 1909 (*Preprint* 1110.3174)
[13] ATLAS Collaboration 2005 ATLAS computing: Technical design report Tech. Rep. CERN-LHCC-2005-022, ATLAS-TDR-017 URL http://cds.cern.ch/record/837738
[14] Adams D *et al.* (ATLAS) 2015 *Journal of Physics: Conference Series* these proceedings
[15] Cacciari M, Salam G P and Soyez G 2012 *Eur.Phys.J.* C72 1896 (*Preprint* 1111.6097)
[16] ATLAS Collaboration 2014 Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector Tech. Rep. ATL-PHYS-PUB-2014-013 CERN Geneva URL http://cds.cern.ch/record/1750681
[17] ATLAS Collaboration 2014 b-tagging in dense environments Tech. Rep. ATL-PHYS-PUB-2014-014 CERN Geneva URL http://cds.cern.ch/record/1750682