# Faster simulated track reconstruction in the ATLAS Fast Chain

*William* Leight[1,*], *Martina* Javurkova[1], *Fang-Ying* Tsai[2], *Debajyoti* Sengupta[3], *Tobias* Golling[3], and *John* Chapman[4], on behalf of the ATLAS Computing Activity

[1]University of Massachusetts Amherst, Amherst, MA, USA
[2]Stony Brook University, Stony Brook, NY, USA
[3]Université de Genève, Geneva, Switzerland
[4]University of Cambridge, Cambridge, UK

**Abstract.** The production of simulated datasets for use by physics analyses consumes a large fraction of ATLAS computing resources, a problem that will only get worse as increases in the instantaneous luminosity provided by the LHC lead to more collisions per bunch crossing (pile-up). One of the more resource-intensive steps in the Monte Carlo production is reconstructing the tracks in the ATLAS Inner Detector, which takes up about 60% of the total detector reconstruction time. This talk discusses a novel technique called track overlay, which substantially speeds up the Inner Detector reconstruction. In track overlay the pile-up Inner Detector tracks are reconstructed ahead of time and overlaid onto the Inner Detector tracks from the simulated hard-scatter event. We present our implementation of this track overlay approach as part of the ATLAS Fast Chain simulation, as well as a method for deciding in which cases it is possible to use track overlay in the reconstruction of simulated data without performance degradation, and in which it is preferable to use the current approach.

## 1 Introduction

The reconstruction of tracks from hits in the ATLAS [1] Inner Detector (ID) from LHC collision data is a basic task underlying almost all physics analysis carried out by the ATLAS experiment. This task is greatly complicated by the presence of pile-up, additional low-energy proton-proton collisions that occur at the same bunch crossing as the high-energy collision(s). The high-energy (referred to as hard-scatter (HS)) collision or collisions contain the interesting processes, in which pile-up forms a large background. Pile-up is measured by calculating the mean number of interactions per bunch crossing, $\mu$. This quantity increases with luminosity: in Run 2, the $\mu$ distribution peaked at 25, while so far in Run 3 the peak is at 50, and in Run 4 the expectation is that $\mu$ will peak at 200. Though they are low-energy, the pile-up collisions produce large numbers of tracks which can potentially overlap with the tracks from the HS collision, making accurate track reconstruction harder.

The presence of pile-up also represents a significant issue for simulation, due to the extra resources demanded to account for it. Simulated Monte Carlo (MC) events must match data

*e-mail: wleight@cern.ch

events, so the MC naturally includes pile-up collisions, with the same $\mu$ distribution as the data. However, computing time for simulation is limited, and ideally, the vast majority of the available CPU resources would not be spent on simulating pile-up collisions which do not represent any physics process that we are interested in measuring. The solution to this dilemma is to simulate the pile-up separately from the HS, in the form of pile-up events containing only pile-up collisions, and reuse it. Since ATLAS simulates many billions of events across a large range of physics processes, there is ample scope for reusing pileup events without introducing biases.

The ATLAS MC production chain consists of several steps, carried out using the ATLAS Athena software [2]. In the first step, particles from simulated collisions are propagated through a Geant4 [3] simulation of the detector, and the appropriate energy deposits are determined. In the second step, referred to as digitization, the detector response is simulated to turn the energy deposits into outputs known as Reconstructed Data Objects (RDO), with a format matching that of real data. Finally, in the reconstruction step, the RDOs are processed using the same algorithms that are used to process real data. The ATLAS Fast Chain attempts to speed up this process at each of the stages.

One key question here is when in the simulation chain the pile-up collisions must be merged with the HS collision(s) to create a complete event. The longer this can be put off, the more computing time can be saved, as fewer operations have to be carried out on the products of the pileup collisions once the complete event is created. However, there is a potential trade-off with the accuracy of the simulation. Hits from pile-up will sometimes interfere with the reconstruction of HS collisions. The longer the creation of the complete event is delayed, the less scope there is for such interference, which will lead to a divergence between the simulated events and real data. Therefore, any method for reusing simulated events needs to be carefully validated.

## 2 Reusing simulated pile-up: MC overlay and track overlay

The simplest approach to reusing the simulated pile-up collisions is to do so after the first stage of the production chain, merging the energy deposits from a pile-up event with those from the simulated HS collision(s). Then the digitization and reconstruction steps are carried out on the complete event. A schematic of this method is shown in Figure 1. This method saves CPU by reducing the number of pile-up collisions that need to be simulated. Merging the events before digitization ensures that there is no loss of accuracy.

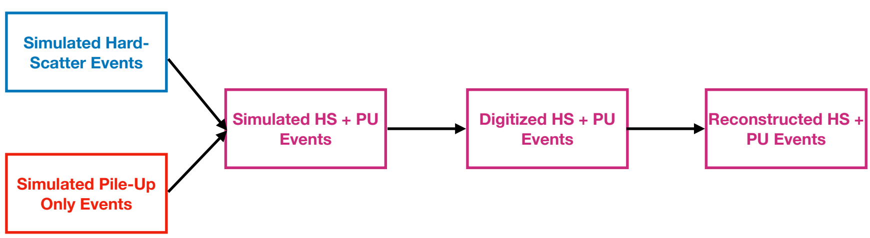



Figure 1: Schematic of the simplest process by which simulated pile-up (PU) collisions can be re-used, by merging with the HS following the first step of the production chain.

Additional CPU savings can be obtained by digitizing the pile-up events before merging. This method, known as MC-overlay, is currently the standard for ATLAS simulation [4]. A diagram of this method is shown in Figure 2. In this approach, pile-up events go through the first two steps of the production chain. HS collisions are then combined with pile-up events when the former are digitized. This allows the digitization of the HS to take into account the presence of pile-up hits, maintaining a high level of accuracy.

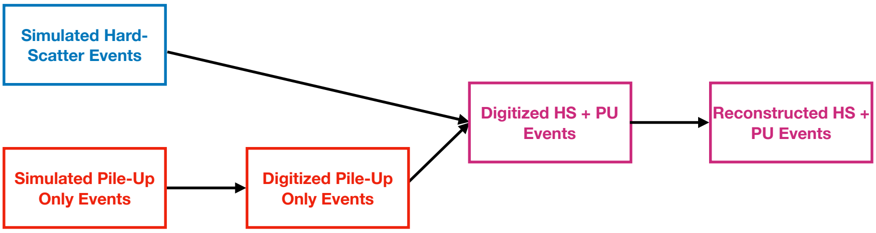



Figure 2: Schematic of the MC-overlay process, in which simulated pile-up collisions are separately digitized and then merged with the HS during the HS digitization step.

Further CPU savings can be obtained by going further and reconstructing pile-up events before merging with the HS, as outlined in Figure 3. As the ID tracking is by far the largest contributor to event reconstruction, taking up approximately 60% of event reconstruction time at $\mu = 50$ [5], it is the logical candidate for separate pile-up reconstruction. In the method referred to as track overlay, the ID is thus treated differently from the other sub-detectors. As in the MC-overlay method described above, the full pile-up event is digitized. Then, for the calorimeter and muon spectrometer, the same MC-overlay process is followed, with the HS hits digitized together with the pile-up RDO objects. However, the HS ID hits are digitized without reference to the pile-up. Instead of being used in MC-overlay, the pile-up ID RDOs have the ID track reconstruction algorithms run on them, producing sets of pile-up tracks. The resulting increase in RDO size is negligible, so the extra storage needed is minimal. Then, during the reconstruction step, the ID track reconstruction is run again on the HS ID RDOs, and the resulting tracks are combined with the pile-up tracks. The combined ID track collection is then used in further reconstruction sub-steps that include information from other sub-detectors (e.g., muon or electron reconstruction).

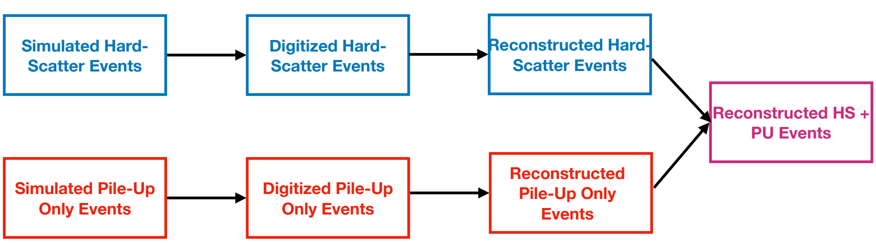
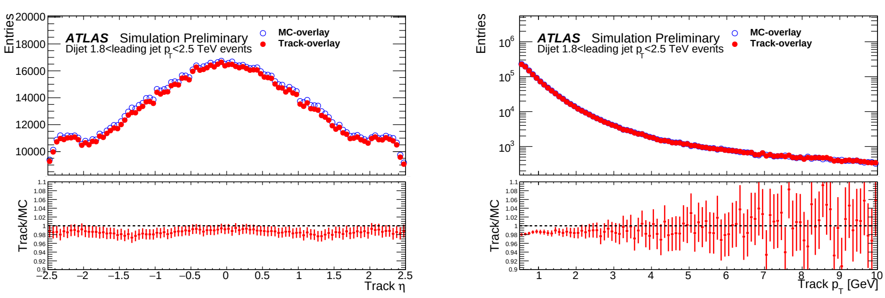



Figure 3: Schematic of the track overlay method, in which both pile-up and HS are digitized and reconstructed before being combined. For simplicity, this diagram only shows the treatment of ID hits in the track overlay method: even in track overlay, other sub-detectors are handled the same way as in MC-overlay.

Track overlay can generate substantial CPU savings because each pile-up track is only reconstructed once. The downside is that the tracks from the HS collision are reconstructed without taking the pile-up into account at all. This presents problems of accuracy for events in which there are a large number of HS tracks clustered closely together, such as events that contain one or more high-$p_T$ jets. In this case, the presence of additional hits from pile-up is quite likely to make the track reconstruction worse. Neglecting this effect will bias track overlay. Therefore, we compare the track overlay to the MC overlay, which properly accounts for the effect of pile-up in HS track reconstruction.

## 3 Track overlay validation

Validation of the track overlay method is carried out by simulating the same events using the MC and track overlay methods. Figure 4 shows a comparison of the ID track transverse momentum ($p_T$) and $\eta$[1] distributions for simulated dijet events in which the $p_T$ of the leading jet is between 1.8 and 2.5 TeV. The figure shows that the difference between the two methods is very small and concentrated at extremely low values of $p_T$. Track overlay reconstructs fewer tracks than MC overlay because there are fewer fake tracks reconstructed when HS and pile-up are not allowed to mix. However, the agreement here is overall very good. Similarly good agreement is observed in other kinematic variables, and when looking at other samples.

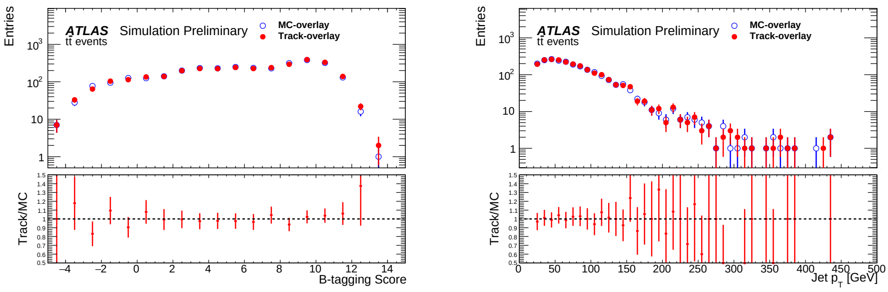



(a) (b)

Figure 4: Reconstructed Inner Detector track $\eta$ (left) and $p_T$ (right) for tracks passing the loose working point from dijet events from simulated collisions using Run 3 conditions at $\sqrt{s} = 13.6$ TeV containing a leading jet with $1.8 < p_T < 2.5$ TeV. Events reconstructed using MC-overlay are shown with empty blue markers and those reconstructed using track-overlay with filled red markers.

Figure 5 shows a comparison of the ATLAS $b$-quark tagging algorithm [6], carried out in simulated $t\bar{t}$ events. This algorithm is only dependent on ID tracks, making b-tagging a useful test of the potential impact of changes in the ID track reconstruction on physics analyses (which rarely make use of ID tracks directly). The left-hand figure shows the distribution of the b-tagging score, while the right-hand one shows the distribution of the $p_T$ values of the jets that pass one of the standard working points. The agreement between the two methods is quite good, showing that track overlay is reliable for this important application of track reconstruction.

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the $z$-axis along the beam pipe. The $x$-axis points from the IP to the center of the LHC ring, and the $y$-axis points upward. Cylindrical coordinates $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the $z$-axis. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln\tan(\theta/2)$.




(a) (b)

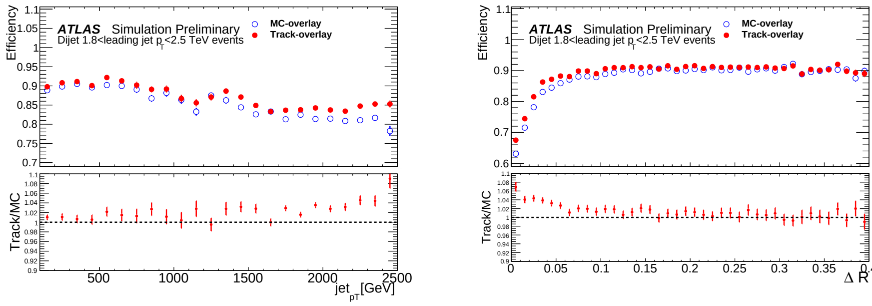
Figure 5: Final discriminant used for the DL1d b-tagging algorithm for jets (left) and $p_{\mathrm{T}}$ of jets passing the 70% b-tagging working point using the DL1D algorithm (right) for jets from $t\bar{t}$ events from simulated collisions using Run 3 conditions at $\sqrt{s} = 13.6$ TeV. Events reconstructed with MC-overlay are shown with empty blue markers and those reconstructed with track-overlay with filled red markers.

However, as expected, differences appear in high-$p_{\mathrm{T}}$ jets. Figure 6(a) shows the efficiency of reconstructing HS tracks for tracks inside jets, using the same simulated dijet events as in Figure 4. As the figure demonstrates, as the $p_{\mathrm{T}}$ of the jet increases, track overlay overestimates the track reconstruction efficiency, due to its neglect of the impact of pile-up on HS track reconstruction. Figure 6(b) plots the efficiency against the distance from the track to the center of the jet, showing that, again as expected, this impact is strongest in the core of the jet, where the density of the hits is highest. At the edge of the jet, there is effectively no impact. While these differences are not large, maximum of 10% in track reconstruction efficiency, and are concentrated in the center of high-$p_{\mathrm{T}}$ jets, they cannot be neglected. As track overlay produces biases in certain types of events, the natural approach is to avoid using it for those events.

This is still a work in progress, but studies are ongoing to train a neural network (NN) using generator-level information to identify events that track overlay is unlikely to work for. A cutoff in the NN score can be used to assign events either to track overlay or to MC overlay. The cutoff can be varied depending on the type of events being reconstructed: e.g., events including high-$p_{\mathrm{T}}$ jets could have a lower cutoff, while other types of events might receive a higher one (or even have no cutoff at all). To ensure that the transition between the track overlay and MC overlay regime is smooth, the assignment of events will be probabilistic. Further work to identify the best NN architecture, cutoff point, and assignment method is currently being carried out.

## 4 Computing resource savings with track overlay

In Run 3, the expectation is that a considerable reduction in CPU can be achieved using track overlay. However, the real goal of developing this method is to have it available for the HL-LHC, when the demand for both CPU and disk will increase exponentially [7]. Current developments in fast tracking are already expected to produce a considerable speed-up in tracking CPU usage for the HL-LHC [8]. Track overlay produces an 80% reduction in CPU usage for tracking in Run 3: if a similar reduction can be achieved for the Run 4 fast tracking,




(a)

(b)

Figure 6: ID track reconstruction efficiency for tracks from the HS collision passing the loose working point in jets as a function of jet $p_{\mathrm{T}}$ (left) and track distance from the center of the jet (right). Reconstructed HS tracks are identified by matching to truth information. Shown are dijet events from simulated collisions using Run 3 conditions at $\sqrt{s} = 13.6$ TeV containing a leading jet with $1.8 < p_T < 2.5$ TeV, reconstructed with MC-overlay (empty blue markers) and track-overlay (filled red markers).

then the tracking CPU requirement in Run 4 will be negligible. Alternatively, the savings from faster reconstruction, including track overlay, could be taken in the form of disk space. If the reconstruction step of simulation becomes sufficiently fast, then there would be no need to save the files, in an intermediate data format, that are its immediate product. This could offer a savings of 27% of the total – that is, including data – disk usage, as well as significant savings on tape storage [7]. While this scenario relies on reconstruction speed-ups in areas that track overlay cannot help with, the presence of track overlay would be an essential component.

## 5 Summary and outlook

In summary, ATLAS has developed, implemented, and validated a method of speeding up the production of MC simulation by running ID tracking on simulated pile-up collisions and then merging the resulting tracks with those from the HS collision(s). This method shows few differences from the standard approach, in which pile-up collisions are merged with HS after the former have been digitized. However, as expected, it results in worse performance, in that the simulated events disagree with data, in the case that the simulated event contains a high-$p_{\mathrm{T}}$ jet. This is due to the high density of ID hits in the core of such jets, which means that the addition of pile-up hits has a noticeable effect on the reconstruction of tracks. To compensate for this problem, a neural network is being trained to predict, on the basis of truth information, whether a given simulated HS collision is suitable for reconstruction with track overlay. In the final implementation, the output value of this NN will be used to determine whether or not track overlay is used on each event. Combined with other improvements in track reconstruction, this should mean that in the HL-LHC the contribution of ID tracking to the total CPU usage of event reconstruction should drop from the current 60% to a few percent or less. If speed-ups can be achieved in other areas of the reconstruction, the possibility exists to increase the frequency with which the reconstruction step is run, in order to avoid the necessity for saving the intermediate files that it produces. This could save 27% of the total predicted disk usage, including space for data storage. Track overlay is thus a key part of the way that ATLAS plans to achieve the CPU and disk savings needed for HL-LHC operations.

## References


[1] ATLAS Collaboration, JINST **3**, S08003 (2008)

[2] ATLAS Collaboration, *Athena* (2021), https://doi.org/10.5281/zenodo.4772550

[3] GEANT4 Collaboration, S. Agostinelli et al., Nucl. Instrum. Meth. A **506**, 250 (2003)

[4] ATLAS Collaboration, Comput. Softw. Big Sci. **6**, 3 (2022)

[5] ATLAS Collaboration, Tech. rep., CERN, Geneva (2021), https://cds.cern.ch/record/2766886

[6] ATLAS Collaboration (ATLAS), Eur. Phys. J. C **83**, 681 (2023), 2211.16345

[7] ATLAS Collaboration, Tech. rep., CERN, Geneva (2022), https://cds.cern.ch/record/2802918

[8] ATLAS Collaboration, Tech. rep., CERN, Geneva (2020), https://cds.cern.ch/record/2729668