Proceedings of CTD 2020
PROC-CTD2020-32
August 3, 2020

Low-p_{T} tracking for ATLAS in nominal LHC pileup conditions

Patrick McCormack
On behalf of the ATLAS Collaboration[1]

Physics Division
Lawrence Berkeley National Laboratory, USA

ABSTRACT

In the most recent years of data-taking with the ATLAS detector at the Large Hadron Collider (LHC), the minimum transverse momentum (p_{T}) of default track reconstruction was 500 MeV. This bound was set to reduce the event reconstruction time and to save disk space, which are challenges in high pileup environments. However, most proton-proton collisions at the LHC result in a large number of soft particles, and the reconstruction of these soft particles in high pileup conditions can provide important information for many analyses. This note explains a method of tracking in high pileup conditions where low-p_{T} tracks are reconstructed in a second tracking pass after the default tracking is performed. In order to prevent a large increase in the per-event reconstruction time, tracks with p_{T} below 500 MeV are only reconstructed within a "region of interest", which is defined event-by-event. Examples of applications for this technique are searches for photon-induced physics at the LHC, charm tagging, and SUSY searches.

PRESENTED AT

Connecting the Dots Workshop (CTD 2020)
April 20-30, 2020

1 Introduction

In an average proton-proton (pp) interaction with center-of-mass energy of $\sqrt{s} = 13$ TeV at the Large Hadron Collider (LHC) [1], more than half of the charged particles produced have a momentum transverse to the beamline (p_T) of less than 500 MeV [2]. The production of such low-p_T particles involves momentum transfers (q^2) in the non-perturbative regime of Quantum Chromodynamics (QCD), making the prediction of distributions associated with such particles difficult, including their multiplicity and p_T. However, by the same principle of asymptotic freedom in QCD, a high-q^2 interaction between partons within the colliding protons can be treated as decoupled from the low-q^2 physics associated with the interactions between the remnants of the protons, which is called the underlying event (UE). Because of this, experiments which focus on the study of high energy phenomena at the LHC, such as ATLAS [3], optimize their instrumentation and algorithms for the reconstruction of particles with p_T typically of order $\geq$1 GeV rather than 0.1 GeV. In most analyses, particles with $p_\mathrm{T} < 500$ MeV are not relevant, so for the most recent years of data-taking with ATLAS, the minimum reconstructed track p_T allowed was 500 MeV. Having this threshold reduces the time taken by the reconstruction algorithm and reduces the storage space required for analyzed events. However, there are many analyses in ATLAS that can benefit from the reconstruction of low-p_T particles; for example, searches for photon-induced physics can use better UE reconstruction to isolate events of interest, and charm-tagging and some SUSY searches use low-p_T particles in reconstructing particle decays that proceed through small mass differences.

Crucially, most analyses that can benefit from the use of low-p_T tracking need to use the bulk of ATLAS data, which in Run 2 averaged more than 30 pp interactions per bunch crossing ("nominal pileup"). There have been analyses in ATLAS looking at low-p_T physics in the past, such as [2]; however, these studies are performed with specialized beam conditions that make crossings with more than 1 pp collision very rare, rendering the integrated luminosity of these runs very small (151 μb^{-1} in [2], compared to 140 fb^{-1} collected in Run 2). Because of these special conditions, the normal restrictions due to event reconstruction time and storage space are avoided.

In this note, an algorithm for the reconstruction of low-p_T tracks designed to be used in nominal LHC pileup conditions is presented. The performance of the algorithm is explored in detail, and potential means of improvement are investigated.

2 Tracking in ATLAS

Tracking in ATLAS is performed as follows [5]:

1. Create "seeds" from space points in 3 different layers (either 3 pixel hits [6, 7] or 3 semiconductor tracker (SCT) hits [8], potentially with extra confirmation hits).

2. Extend seeds through additional layers using a Kalman filter to create "candidates".

3. Perform ambiguity solving on candidates to create final tracks, with the goal of reconstructing exactly 1 track per particle.

Certain restrictions are placed on the number of hits that must be used to make a track, on the number of hits missing where expected along a track, and on the amount of hit sharing between tracks. Nominally, the seeds must have a predicted p_T greater than 500 MeV, where the predicted p_T is determined by interpolating a curve between the hits that comprise the seed, corresponding to the momentum-dependent curvature of a charged particle in a magnetic field. In the study performed in [2], this algorithm was adjusted to allow tracks down to 100 MeV, but otherwise the steps involved were the same.

When considering high-p_T particles alone, seed finding is a question of finding three hits that form a line in space (or small deviations from a line), as high-p_T particles curve relatively little. As tracks with lower and lower p_T are allowed, an increasing number of hit combinations must be considered, as more significantly non-linear sets of hits may have been caused by a curving charged particle. This is problematic in nominal LHC pileup conditions, because a large number of hits are created on the active components of the ATLAS inner detector (ID) [3, 7] in each bunch crossing.

In order to alleviate this combinatorics problem, the new algorithm for low-p_{T} tracking involves two stages. First, the default tracking algorithm is run as described above with a minimum track p_{T} threshold of 500 MeV. The detector hits used by these tracks are then removed from consideration, and the track reconstruction algorithm is run again on the remaining hits, where seeds with a predicted p_{T} down to 100 MeV are accepted and the hit requirements for final tracks are slightly relaxed.

In addition to the two-pass tracking setup, a "region of interest" (RoI) for tracking has been implemented to reduce the reconstruction combinatorics problem. The RoI is a region of set distance longitudinally along the beamline centered around a location of interest to the analysis. This center can be set either based on reconstructed objects, such as leptons or jets, or based on the location of some truth-level collision of interest in a simulated sample. In the second low-p_{T} tracking step, seeds are found iteratively and rejected if their longitudinal beamline projection falls outside of the RoI. Seeds that are accepted are then potentially extended into candidates and eventually finalized tracks. The size of the RoI is adjustable, and determining what value to use will be discussed below.

To guide the determination of the RoI size and to evaluate the algorithm's performance overall, three metrics are used:

- Efficiency of reconstructing truth-level charged particles (with matching as defined in [5]):

$$\text{Efficiency} = \frac{\text{Number of charged particles with at least one matched track}}{\text{Number of charged particles}}. \tag{1}$$

- Fake rate, as shown in Eq. (2), where a fake track is one constructed from detector hits created by multiple different charged particles, which is to say that it does not correspond to any truth level object's trajectory. A more exhaustive explanation of the metric used to determine a fake track is given in Sec. 6 in the context of seeds; the arguments there generalize to tracks.

$$\text{Fake rate} = \frac{\text{Number of fake tracks}}{\text{Number of tracks}}. \tag{2}$$

- Impact on the reconstruction time.

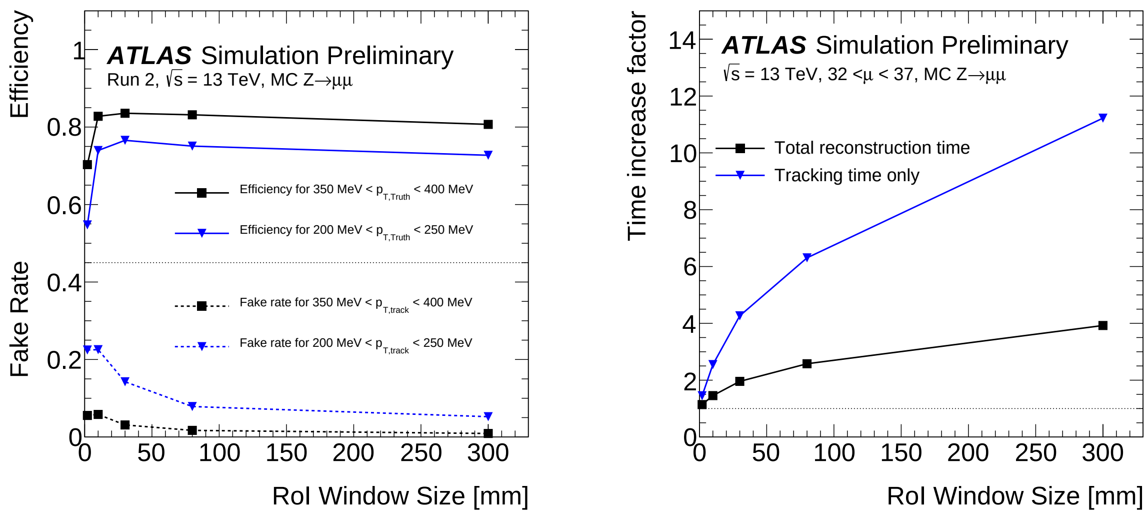
The following study was performed on a Powheg+Pythia8 [9, 10] Z boson sample, where the Z boson was forced to decay to two muons with invariant mass between 60 and 110 GeV. The sample had a charged particle filter applied, such that there were fewer than 12 truth particles with $p_{\mathrm{T}} > 500$ MeV and $\eta < 2.5$ originating from the Z boson-producing interaction. The pileup conditions of the sample approximate those of data taking at the LHC from 2015 to 2018. In this study, the RoI center for each event is set based on the truth-level muon information.

3 Setting the RoI size

Figure 1 illustrates the above performance metrics in the context of determining an RoI size. The top half of Figure 1a shows the efficiency as a function of the RoI size for hard-scatter tracks in a simulated sample, and the bottom half of that plot shows the fake rate for tracks near the hard-scatter interaction location. Similarly, Figure 1b shows the impact on the reconstruction time. By considering Figure 1, the following conclusions are reached:

- A small RoI (< 10 mm) is fast but has lower efficiency and high fake rate.
- A large RoI (> 100 mm) has low fake rate, has generally lower efficiency, and is time intensive.
- A medium RoI size (around 30 mm) has close to maximal efficiency, at the cost of medium fake rate and a moderate impact on the reconstruction time.

Because of this, 30 mm is chosen for the RoI size parameter of the algorithm, though it can be adjusted at run time. To some extent, the size of the RoI is driven by the fact that seeds have much poorer pointing resolution than tracks. Though tracks typically have a resolution better than $\mathcal{O}(1)$ mm, using a too small RoI results in lost efficiency, as seeds rejected due to pointing outside the RoI may end up being extended into tracks that would point into the RoI.



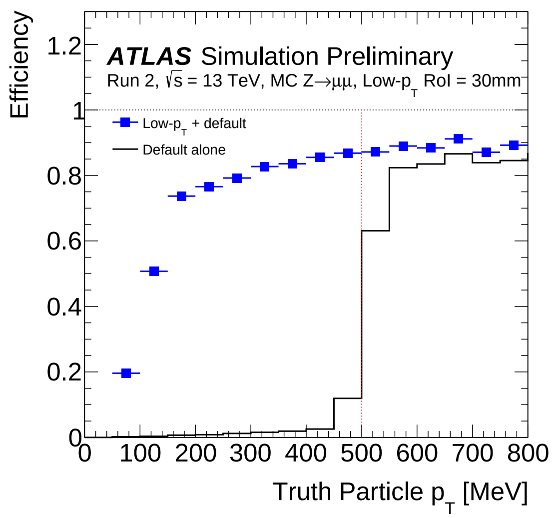
(a) (b)

Figure 1: (a) Tracking efficiency and fake rate as a function of the Region of Interest (RoI) size for two different truth particle or track p_{T} ranges. Only truth particles originating from the hard-scatter interaction are considered for the tracking efficiency, and only tracks pointing to the hard-scatter interaction location within 1 mm longitudinally along the beamline and with a transverse displacement from the beamline of less than 1 mm are considered for the fake rate. (b) Impact of RoI size choices on the reconstruction time. The black squares show the fractional increase in the total reconstruction time over the default setup. The blue triangles show the fractional increase in the tracking time alone, where the tracking time is defined as the time spent in both the default tracking pass and in the low-p_{T} track pass [4].

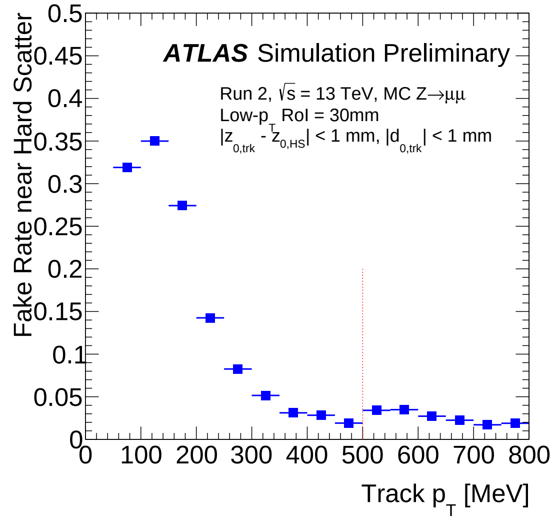
4 Efficiency and fake rate

Figure 2 shows the efficiency and fake rate now as a function of $p_{\mathrm{T,truth}}$ and $p_{\mathrm{T,track}}$, respectively. In Figure 2a, the efficiency is also shown for default-tracking alone. It can be seen that the low-p_{T} tracking algorithm gives a smooth and gently decreasing efficiency as $p_{\mathrm{T,truth}}$ decreases down to about 200 MeV. In the very low-$p_{\mathrm{T,truth}}$ regime below 200 MeV, the efficiency falls off quickly. The fake rate stays below 10% down to $p_{\mathrm{T,track}}$ of 250 MeV. The fake rate increases quickly in the very low-$p_{\mathrm{T,track}}$, where approximately one out of three tracks are expected to be fake.

The natural question that is raised then is how to improve the efficiency and/or decrease the fake rate. Given the structure of the ATLAS track reconstruction algorithm, to increase the efficiency one would naturally want to create more seeds or else to accept more candidates. Alternatively, to decrease the fake rate, one would prefer fewer seeds or perhaps to accept fewer candidates. To investigate what avenues of improvement are most promising, let's examine where in the reconstruction process efficiency is being lost and the general quality of the seeds being created.



(a)



(b)

Figure 2: (a) Tracking efficiency as a function of the truth particle p_{T}. The black line indicates the efficiency when default tracking alone is run, and the blue points indicate the efficiency when both default and low-p_{T} trackings are run. The truth particles considered here originate from the hard-scatter interaction. (b) Fake rate as a function of the reconstructed track p_{T}, where the tracks are restricted to point near the hard-scatter interaction location. The reconstructed tracks considered here have to point to within 1 mm of the hard-scatter interaction longitudinally along the beamline and be displaced from the beamline by less than 1 mm in the transverse direction. Both default and low-p_{T} tracks are included when the considering fake rate [4].

5 Where is efficiency lost?

Figure 3 shows the seed efficiency, candidate efficiency, and the previously shown track efficiency as a function of $p_{\mathrm{T,truth}}$. The definitions for seed efficiency and candidate efficiency are the same as that in Eq. (1), but with “matched seed” and “matched candidate” respectively instead of “matched track”. Matching here is a question of which truth particle contributed the most energy to the hits comprising the seed/candidate/track [5]. It can be seen that consistently across almost all $p_{\mathrm{T,truth}}$ values, over 90% of truth particles have at least one matched seed, and that above about 200 MeV, typically less than half of the seed-to-track efficiency loss occurs in the seed-to-candidate stage. In the very low-$p_{\mathrm{T,truth}}$ regime, the drop in seed-to-candidate efficiency is more dramatic. Given that such a high fraction of truth particles have at least one matched seed, does that mean that the efficiency can be improved simply by accepting a higher fraction of seeds? And why is the very low-$p_{\mathrm{T,truth}}$ efficiency drop between seed and candidate efficiency so dramatic? Answering these questions requires an investigation of the seed quality.

6 Seed quality

The primary metric for understanding the seed quality is the “truth match probability”. This is the weighted fraction of hits on a track created by the truth particle to which the seed is matched, where pixel hits receive a higher weight than SCT hits. Seeds are formed from 3 pixel hits or 3 SCT hits, though occasionally extra confirmation hits are used, typically resulting in a seed with 4 or 5 SCT hits. If all 3 hits on a seed are from the same truth particle, the match probability will be 3/3 or 1. If all three hits on the seed are caused by different truth particles, the match probability will be 1/3 or 0.333. This latter case is considered a fake seed as it is comprised of a random collection of hits. A seed is considered “fake” if the match probability

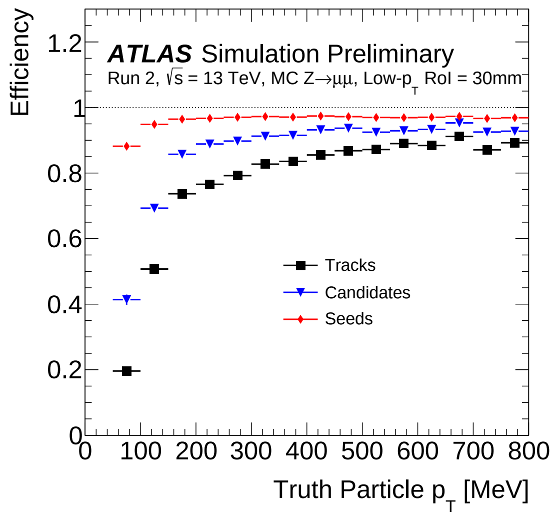


Figure 3: Efficiency for the three steps of ATLAS tracking as a function of the truth particle p_{T}. Seed efficiency (red diamonds) is defined here as the likelihood for a stable, charged truth particle within tracking acceptance to have at least one matched seed. Candidate efficiency (blue triangles) and track efficiency (black squares) are defined analogously for candidates and tracks respectively. The truth particles considered here originate from the hard-scatter interaction [4].

is less than 0.5.

Figure 4 shows the proportion of good and fake seeds as a function of the predicted $p_{\mathrm{T,seed}}$. In the very low-$p_{\mathrm{T,seed}}$ regime, almost half of the seeds are fake. This calls into question the idea of simply "accepting more seeds" to improve the efficiency. If the seed is fake, one cannot expect it to be extended through additional hits to make a candidate, which could explain the large seed-to-candidate efficiency drop. Given the high fake rate, loosening selection criteria would likely lead to a large increase in fake rate rather than a gain in efficiency.

7 Reducing the fake rate

Whereas increasing the efficiency requires creating additional seeds or loosening criteria, decreasing the fake rate is a matter of tightening seed-to-candidate-to-track selections. In order not to suffer a large loss in efficiency by doing this, one must start out with an adequate number of seeds per track. Figure 5a shows the difference in z_0 between low-p_{T} seeds and tracks, where both the seed and track are matched to the same truth particle. While the fake seeds can point almost anywhere, low-p_{T} tracks typically have about two good seeds pointing within 5 mm along the beamline. Figure 5b shows the average number of seeds per track as a function the p_{T} of the truth particle to which both are matched, where the seeds must have z_0 within 5 mm of their track. For truth particles that have a track, there are typically about two good pointing seeds per track regardless of $p_{\mathrm{T,truth}}$. This suggests that there might be some redundancy allowing for tightened selection criteria in the tracking algorithm.

Increasing the efficiency for very low-p_{T} tracks will be difficult; there is a large combinatorial background such that almost half of the seeds are comprised of hits of random origin. An increase in efficiency will come at a high cost in fake rate. However, reducing the fake rate without a drastic loss of efficiency might be possible by tightening seed or track selection criteria. One option is to do this at reconstruction level, which would likely have the added benefit of reducing the reconstruction time. In order not to throw out any information though, this could also be done offline at track level. For example, there could be a set of track selections based on p_{T}, η, number of pixel hits, number of SCT hits, or number of holes. Any reduction

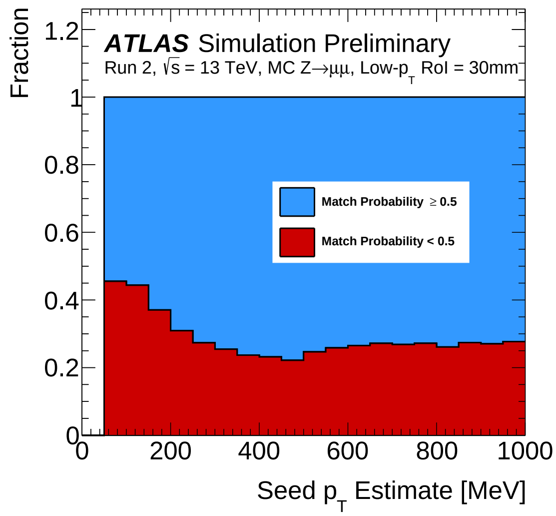


Figure 4: Seed composition as a function of the seed p_{T}, where the seed p_{T} is an estimate from the 3 (or more) hits that comprise the seed. The seeds considered for this plot are required to point to within 5 mm of the hard-scatter interaction longitudinally along the beamline and to have a transverse displacement of less than 2 mm. The blue region shows the fraction of seeds where the truth matching probability is greater than 0.5 ("good"), and the red area shows the fraction of seeds where the truth matching probability is less than 0.5 ("fake") [4].

in fake rate will come with a loss of efficiency, but the appropriate balance of efficiency and fake rate is a somewhat analysis-specific question.

8 Conclusions

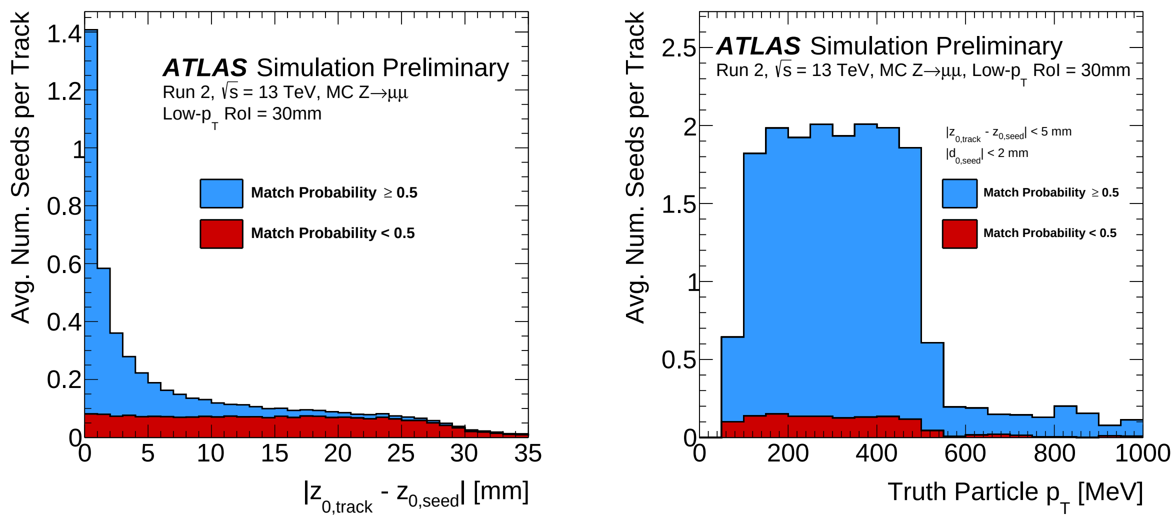
Low-p_{T} tracking can be a useful tool for many analyses, and the algorithm presented in this note can extend tracking down to about 200 MeV in nominal LHC pileup conditions with relatively high efficiency and low fake rate. In the very low-p_{T} regime below 200 MeV, the efficiency tends to be low and the fake rate high, but some tightening of selection criteria for seeds or tracks can likely alleviate the fake rate at least. The working point presented increases the track reconstruction time by 300% and the total event reconstruction time by 100%. Therefore, this would not be viable to run on all events but is acceptable for pre-selected events passing initial analysis cuts, which is a typical use case. The algorithm is currently neither intended to be used for a full reprocessing of data nor to be run by default on all ATLAS simulation analyses.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy, Office of Science under contract DE-AC02-05CH11231. Patrick McCormack was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1752814. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

[1] L. Evans and P. Bryant, "LHC Machine", JINST **3**, S08001 (2008).



(a) (b)

Figure 5: (a) Distribution of the absolute difference in longitudinal distance along the beamline (z_0) of seeds and tracks created in the low-p_{T} tracking pass, where both the seed and track are matched to the same truth particle. The blue region shows “good” seeds, and the red area shows “fake” seeds. This is a stacked plot; the y-axis represents the average number of matched seeds per track for a given track-seed position difference. (b) Average number of seeds created in the low-p_{T} tracking pass per track, where both the seed and track are matched to the same truth particle, as a function of the p_{T} of the truth particle to which both are matched. The seeds are required to be within 5 mm of their track longitudinally along the beamline and to have a transverse displacement from the beamline of less than 2 mm. The blue region shows “good” seeds, and the red area shows “fake” seeds [4].

[2] ATLAS Collaboration, “Charged-particle distributions at low transverse momentum in $\sqrt{s}$=13 TeV pp interactions measured with the ATLAS detector at the LHC”, Eur. Phys. J. C **76** (2016) 502 [arXiv:1606.01133 [hep-ex]].

[3] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, JINST **3**, S08003 (2008).

[4] ATLAS Collaboration, “Low p_{T} tracking in nominal pileup”, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2020-001/.

[5] ATLAS Collaboration, “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2”, Eur. Phys. J. C **77** (2017) 673 [arXiv:1704.07983 [hep-ex]].

[6] ATLAS Collaboration, “ATLAS pixel detector electronics and sensors”, JINST **3**, P07007 (2008).

[7] ATLAS Collaboration, “ATLAS Insertable B-Layer Technical Design Report”, ATLAS-TDR-19, https://cds.cern.ch/record/1291633 (2010).

[8] A. Ahmad, et al., “The Silicon microstrip sensors of the ATLAS semiconductor tracker”, Nucl. Instrum. Meth. A **578**, 98-118 (2007).

[9] S. Alioli, et al., “NLO vector-boson production matched with shower in POWHEG”, JHEP **07**, 060 (2008) [arXiv:0805.4802 [hep-ph]].

[10] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, Comput. Phys. Commun. **191** 159 (2015) [arXiv:1410.3012 [hep-ph]].