Expected $b$-tagging Performance with the upgraded ATLAS Inner Tracker Detector at the High-Luminosity LHC

The ATLAS Collaboration

In view of the operation of the ATLAS detector under High-Luminosity LHC conditions, the central tracking system will be upgraded to maintain high levels of performance with a peak instantaneous luminosity of $7.5 \times 10^{34}$ cm$^{-2}$s$^{-1}$, corresponding to an average of about 200 inelastic proton-proton collisions per beam-crossing. The new Inner Tracker (ITk), with an extended pseudo-rapidity coverage up to $|\eta|=4$, is planned to operate for more than ten years, during which time the LHC aims to deliver $4 \text{ ab}^{-1}$ of integrated luminosity. The algorithms used for the identification of the hadronic jets stemming from $b$-quarks rely heavily on track reconstruction, and their performance is therefore strongly dependent on the design of this detector. This document presents the $b$-tagging performance expected with the latest ITk design, updated since the ITk Pixel Detector Technical Design Report.
1 Introduction

The scientific programme of the LHC spans the next 20 years and includes an ambitious series of upgrades, culminating in the High-Luminosity LHC (HL-LHC) phase [1], that will ultimately result in an accumulated integrated luminosity for proton-proton collisions of around 4 ab$^{-1}$. This represents an order of magnitude more data than what will be collected prior to the HL-LHC run. The foreseen increased luminosity for this phase, up to $7.5 \times 10^{34}$ cm$^{-2}$s$^{-1}$, with its associated data rate and accumulated radiation damage, would lead to unacceptable levels of track reconstruction performance loss with the current inner detector (ID). Therefore the ATLAS collaboration plans to replace the ID with a new all-silicon tracker to maintain tracking performance in these conditions. The upgraded detector installation (Phase-II) will take place during the so-called Long Shutdown 3, starting at the end of 2024.

The Inner Tracker (ITk) combines precision central tracking in the presence of 200 pile-up events, with the ability to extend the tracking coverage to a pseudo-rapidity $|\eta|$ of 4 (from the current 2.5) with excellent tracking performance [2]. The ITk comprises two subsystems; a Strip Detector [3] surrounding a Pixel Detector [4]. The Strip Detector has four barrel layers and six end-cap disks, covering $|\eta| < 2.7$. It is complemented by the Pixel Detector with five barrel layers extending the coverage to $|\eta| < 4$, and multiple ring-shaped end-cap disks, the number of which depends on the pseudo-rapidity. Given the harsh radiation environment expected for HL-LHC, the two innermost layers of the Pixel Detector are replaceable.

The identification of jets arising from $b$-hadrons, referred to as $b$-tagging, is key to many physics analyses: top physics, the measurement of Higgs boson couplings to heavy quarks, probing the Higgs boson self-coupling and the searches for new physics where new heavy particles tend to decay to heavy quarks. Therefore it is essential to improve or at least maintain the current Run 2 $b$-tagging performance for the ITk, while extending it to larger rapidities. As such, the $b$-tagging performance is a key benchmark for the detector design. Results on this topic have been presented in the ATLAS Inner Tracker Pixel Detector Technical Design Report (TDR) [4]. Since then, the simulation of the ITk has been refined to take into account, in particular, a reduction of the detector active volume required for the introduction of the proposed High-Granularity Timing Detector [5] and design choices regarding the local support structures for the inner Pixel system. This document focuses on the $b$-tagging performance obtained with the so-called “ITk layout” introduced in [2].

This note is organized as follows. Section 2 gives details on the Monte Carlo (MC) samples and the object and event selections used to estimate the $b$-tagging performance. Section 3 briefly presents the $b$-tagging algorithms investigated, while their performance is presented in Section 4. The impact of the $b$-tagging performance on some key HL-LHC analyses is outlined in Section 5.

2 Monte Carlo Samples and Object Selections

The studies presented in this note are performed using a Monte Carlo sample of simulated $t\bar{t}$ events, with inclusive top quark decays, produced in proton-proton collisions generated at a centre-of-mass energy $\sqrt{s} = 14$ TeV. This sample is generated with POWHEG [6] interfaced with PYTHIA 6 [7] for parton shower and hadronization. An average of 200 pile-up interactions per event are simulated, in line with the HL-LHC operating environment. All simulated event samples were processed with the GEANT4 ATLAS full detector simulation [8, 9], including the ITk layout as described in [2], displayed in Fig. 1. The
corresponding number of radiation lengths encountered by particles going through the ITk is presented in
Fig. 2.

Figure 1: A schematic depiction of the ITk layout used in simulated samples considered in this document [2]. The
active elements of the Strip Detector are shown in blue, and those of the Pixel Detector are shown in red. The
horizontal axis is along the beam line with zero being the interaction point. The vertical axis is the radius measured
from the interaction region.

Some of the key inputs for $b$-tagging are: the jet direction usually obtained from calorimeter jets, the
tracks reconstructed in the ITk and the primary vertex of the hard-scattering collision of interest. Tracks
are reconstructed from clusters of signals in the silicon pixel and strip sensors, collectively referred to as
hits. For the studies shown in this document, the pixel hits are clustered using the analogue clustering
algorithm [4], for which the center of the cluster is computed using a weighted sum based on the charge
information of the pixels belonging to the clusters. Minimal quality requirements are applied to the
reconstructed tracks based on pseudorapidity, hit content and displacement with respect to the mean position
of the reconstructed beam spot in the transverse plane and along the beam axis. Those requirements are
detailed in Table 1. An average track reconstruction efficiency between 85% and 95% is associated with
those selections in $t\bar{t}$ events [2], similar to the current Run-2 detector. The number of nuclear interaction
lengths encountered by a track until enough detector layers have been crossed such that the reconstruction
hit requirements are met is displayed in Fig. 3 and compared to the current ATLAS detector. A similar
amount of material is traversed by successfully reconstructed particles in the barrel region, while it is
significantly lower in the end-cap and forward regions.

The $b$-tagging performance is expected to be strongly correlated with the impact parameter resolutions
obtained with the ITk. The transverse impact parameter $d_0$ used for $b$-tagging algorithms is the distance of
closest approach of the track to the primary vertex point, in the $r$-$\phi$ projection, while in the longitudinal
direction, the parameter used is the difference between the $z$ coordinates of the primary vertex position and
of the track at this point of closest approach (i.e. the longitudinal impact parameter $z_0$) multiplied by the
Figure 2: Number of radiation lengths $X_0$ versus the pseudorapidity $|\eta|$ for the ITk Layout [2]. The Patch Panel 1 is located behind all of the ITk active layers and does not impact the tracking performance.

| Requirement | $|\eta| < 2.0$ | $2.0 < |\eta| < 2.6$ | $2.6 < |\eta| < 4.0$ |
|-------------|---------------|-----------------|-----------------|
| pixel+strip hits | $\geq 9$ | $\geq 8$ | $\geq 7$ |
| pixel hits | $\geq 1$ | $\geq 1$ | $\geq 1$ |
| holes | $\leq 2$ | $\leq 2$ | $\leq 2$ |
| double holes | $\leq 1$ | $\leq 1$ | $\leq 1$ |
| $p_T$ [MeV] | $> 900$ | $> 400$ | $> 400$ |
| $|d_0|$ | $\leq 2$ mm | $\leq 2$ mm | $\leq 10$ mm |
| $|z_0|$ | $\leq 20$ cm | $\leq 20$ cm | $\leq 20$ cm |

Table 1: Set of selection requirements applied during the track reconstruction in different pseudorapidity intervals. Holes are counted if track candidates cross active sensors on which no hit was found, double holes are two consecutive active sensors crossed without a hit found. Layers before the first or after the last hit are excluded from the hole counting. The longitudinal and transverse impact parameters, $z_0$ and $d_0$, are defined with respect to the mean position of the beam spot.

sine of the track polar angle $\theta$. The performance results presented in this document have been obtained for the two pixel pitches considered for the ITk pixel detector ($50 \times 50 \mu m^2$ and $25 \times 100 \mu m^2$). The corresponding impact parameter resolutions, as well as the performance obtained with the current ATLAS detector with a pixel pitch of $50 \times 250 \mu m^2$ for the innermost pixel layer (IBL [10]), are illustrated in Fig. 4. For low transverse momenta, this version of the ITk geometry is not expected to match the impact parameter resolutions of the Run-2 ATLAS detector, due to the larger radius of the innermost pixel layer (39 mm for ITk compared to 33 mm for the IBL). An option of reducing the radius of the ITk innermost
Figure 3: Comparison of the number of nuclear interaction lengths seen by a particle as a function of pseudorapidity up to the position where enough detector layers have been crossed such that the reconstruction hit requirements of Table 1 are met [2]. A comparison with the Run-2 detector is presented. For the Run-2 detector the hit requirement is 7 hits.

The primary and pile-up interaction vertices are reconstructed using the Adaptive Multi-Vertex Finder (AMVF) algorithm [11]. The use of this algorithm, instead of the Iterative Vertex Finder used during Run-2 [12], was motivated by the increased pile-up. The inputs to the AMVF algorithm (beam spot properties and reconstructed tracks), along with the procedure employed to reconstruct vertices, are described in detail in the Pixel TDR [4]. The vertex reconstruction efficiency, defined as the fraction of events in which the vertex corresponding to the hard-scatter interaction is successfully reconstructed, is close to 100% in $t\bar{t}$ events. Around 100 vertices are reconstructed per event and the reconstructed vertex with the largest sum of the squared transverse momenta of the associated tracks ($\Sigma p_T^2$) is chosen as the signal primary vertex. Simulation studies indicate that the probability of choosing the correct signal primary vertex, applying this criterion, is around 95% in $t\bar{t}$ events.

As the aim of the studies presented here is to focus on the $b$-tagging performance and factor out primary
vertex finding and selection performance, some events are discarded based on the true position of the hard scatter vertex. Events where the chosen reconstructed signal primary vertex is not within 0.1 mm of the true position of the hard scatter vertex along the beam axis make up around 7% of the μ sample, and are rejected.

Jets are reconstructed by clustering energy deposits in the calorimeter with the anti-\( k_t \) algorithm [13] with a radius parameter \( R = 0.4 \), as implemented in the FastJet package [14]. Jet transverse momenta are further corrected to the corresponding particle-level jet \( p_T \), based on the simulation [15]. Only jets with \( p_T > 20 \) GeV and \(|\eta| < 4.0\) are considered. In order to decouple the \( b \)-tagging performance from the ongoing optimisation of the pile-up jet tagging performance with the ITk detector, generator-level selections are applied to discard pile-up jets. Selected jets are required to be matched within \( \Delta R < 0.3 \) to a truth jet. Truth jets are built from the stable particles produced in the hard interaction, using the same FastJet implementation of the anti-\( k_t \) algorithm with \( R = 0.4 \). Jets failing this criteria are labelled as pile-up jets and are not considered, unless stated explicitly. Therefore, for the studies shown in this note, light-flavour jets originate either from light quarks produced by the \( W \) from top decays or from QCD initial

Figure 4: Track parameter resolution in \( d_\theta \) (a-b) and \( z_0 \) (c-d) as a function of \( \eta \) for muons with \( p_T = 1 \) GeV and \( p_T = 100 \) GeV without pile-up [2]. For comparison the results for the current Run-2 detector are shown as a black line. The ratio in the lower part of the plots is defined as the results using \( 50 \times 50 \mu m^2 \) pixels over those obtained using \( 25 \times 100 \mu m^2 \) pixels.
or final state radiation but not from pile-up induced jets. However, inside the considered jets, the overlay of energy deposits and tracks coming from pile-up is present. In particular, 25% of the jets are associated with at least one pile-up track and are expected to display degraded $b$-tagging performance with respect to jets associated with only hard-scatter tracks.

In order to label jets as being $b$, $c$- or light-flavour jets in the simulation, the jets are matched to the $b$ and $c$-hadrons within a cone of radius $\Delta R = 0.3$ around the jet axis. The transverse momenta of the $b$- and $c$-hadrons are required to be larger than 5 GeV. The labelling procedure is performed by first searching for $b$-hadrons within the cone centered on the axis of the jet. If no $b$-hadron is found, the algorithm is repeated for $c$-hadrons, and for $\tau$ leptons. Ultimately the remaining jets which do not fall in the previous categories are considered as light-flavour jets. Results presented here differ from those presented in [2] or [4] due to the implementation of an additional generator-level jet matching to quarks for light-jets, in order to remove the contamination from electrons produced in top quark decays. Reconstructed jets are thus required to be matched within $\Delta R < 0.3$ with a generator-level light quark produced in the hard scattering interaction or the top quark hadronic decays.

3 $b$-tagging Algorithms Studied

Algorithms typically used in ATLAS for $b$-tagging fall into two broad categories. The so-called low-level taggers focus on specific features of basic jet constituents, such as large impact parameter tracks or secondary vertices. The IP3D algorithm [16], presented in Section 3.1, exploits specifically the large impact parameters of tracks originating from $b$-hadron decays, while the SV1 algorithm [17], presented in Section 3.2, attempts to reconstruct secondary vertices from the tracks and exploits kinematic variables associated to those objects. The performance of those low-level taggers are to a large extent uncorrelated and they can therefore be combined to improve the $b$-tagging performance into so-called high-level taggers, such as IP3D+SV1 or MV2 [12], detailed in Section 3.3 and 3.4, respectively. The algorithms have been widely used during Run 2 and have been reoptimised to take into account the new ITk geometry and the HL-LHC data-taking conditions.

3.1 IP3D Algorithm

Impact parameter (IP) based algorithms aim to identify $b$-jets using the impact parameters of the charged-particle tracks from the $b$-hadron decay products. The uncertainties associated to the transverse impact parameter $d_0$ and the longitudinal impact parameter $z_0$ multiplied by the sine of the track polar angle $\theta$, $z_0 \sin \theta$ are denoted as $\sigma_{d_0}$ and $\sigma_{z_0 \sin \theta}$. The tracks from the $b$-hadron decay products tend to have large impact parameter significance. The IP3D tagger [16] is a low-level discriminant taking advantage of those properties and has been widely used for both the current $b$-tagging algorithms and ATLAS upgrade studies.

IP3D is based on a log-likelihood ratio (LLR) method. For each track, the measurement of the impact parameter significances $(d_0/\sigma_{d_0}, z_0 \sin \theta/\sigma_{z_0 \sin \theta})$ is compared to pre-determined simulation-based two-dimensional probability density functions (PDFs) obtained from simulation for both the $b$- and light-flavour-jet hypotheses. As such, the IP3D discriminant is not specifically optimised to reject the background from $c$-jet but a $c$-jet probability is still computed to be used as input of the MV2 discriminant described in Section 3.4. The ratio of probabilities defines the track weight. The jet weight is then defined as the sum.
of the logarithms of the individual track weights. The LLR formalism allows exclusive track categories to be used by defining different dedicated PDFs for each category depending on the quality of the tracks. 17 track categories are defined based on the hit content and expected impact parameter resolution. They have been reoptimised for the Pixel TDR studies [4] to exploit the geometry of the new ITk detector. Only tracks with $p_T > 1$ GeV are considered. The transverse and longitudinal impact parameters with respect to the primary vertex must fulfil $|d_0| < 1$ mm and $|z_0| \sin \theta < 1.5$ mm. Those selections are still inherited from the Run 2 configuration and will benefit from a dedicated re-optimisation for the ITk in the future. The tracks are associated to jets following the same $p_T$-dependent $\Delta R$ criteria as those used in Run 2 [12]. The projections of the PDF, associated with the impact parameter significances, in the category of central tracks with $|\eta| < 1$ corresponding to the best expected impact parameter resolutions, are presented in Fig. 5. The sign of the impact parameter significances is defined as positive if the track intersects the jet axis in front of the primary vertex along the flight direction, and as negative if the intersection lies behind the primary vertex [16]. In the case of $b$-hadrons, a positive sign is expected while for tracks originating from the primary vertex, the experimental resolution results in a symmetric distribution around zero.

![Figure 5: One-dimensional projections of the probability density function used for the IP3D algorithm in the category of central tracks with $|\eta| < 1$ corresponding to the best expected impact parameter resolutions.](image)

### 3.2 SV1 Algorithm

The SV1 algorithm [17] is based on the reconstruction of the secondary vertex formed by the decay products of the $b$-hadron. A single inclusive secondary vertex is reconstructed by the Single Secondary Vertex Finder algorithm [17]. The transverse impact parameter of the tracks considered must fulfil $|d_0| < 3.5$ mm. Vertices whose associated invariant mass is compatible with the kaon mass are discarded, as they would otherwise be associated with a significant mistagging rate for light flavour jets. Due to the complex ITk geometry, no attempt to discard vertices close to the layers of the pixel detector has been implemented to reduce the impact of hadronic interactions with the detector material but it will be considered for future improvements. Subsequently, properties of that vertex are used via a likelihood ratio formalism similar to
the one used for the IP3D algorithm to define a discriminant. The maximum efficiency reachable with SV1 is limited by the secondary vertex reconstruction efficiency in real $b$-jets ($\sim 75\%$-$85\%$ for the considered $\bar{t}t$ sample, against $\sim 10\%$ for light jets) but the discrimination power of SV1 provides additional handles with respect to IP3D when combined into high-level taggers. The SV1 discriminant exploits three vertex properties: the vertex mass (the invariant mass of all charged-particle tracks used to reconstruct the vertex, assuming that all tracks are pions), the energy fraction (the ratio of the sum of the energies of tracks used in the vertex to the sum of the energies of all tracks in the jet), and the number of two-track vertices. In addition, the $\Delta R$ separation between the jet direction and the direction of the line joining the primary vertex and the secondary vertex is used in the LLR. SV1 relies on a two-dimensional PDF of the first two variables and on two one-dimensional PDFs of the latter variables. The projections of those PDFs are presented in Fig. 6.

### 3.3 IP3D+SV1 Algorithm

The two previous algorithms provide low-level discriminants focusing on different properties of the jets and can be combined in a straightforward manner by summing their respective weights, thus defining the so-called IP3D+SV1 algorithm. This discriminant benefits from an improved mistag rate reduction with respect to the individual IP3D and SV1 outputs, thanks to the limited correlation between those. In particular, for a 70% $b$-tagging efficiency working point, the rejection obtained with the IP3D discriminant alone can be improved by $\sim 60\%$ using the combined IP3D+SV1 discriminant.

### 3.4 MV2 Algorithm

MV2 [12] is a multivariate algorithm based on a Boosted Decision Tree (BDT) that combines inputs from the low-level impact parameter and secondary vertex algorithms described above, as well as from the decay chain multi-vertex algorithm JetFitter [18], which attempts to reconstruct the topological structure of cascade $b$- and $c$-hadron decays inside the jet and is used to define additional topological input variables for MV2. A dedicated reoptimisation of the JetFitter reconstruction parameters has not been performed for those studies and may further improve the MV2 performance in the future. The version of the algorithm used for the upgrade studies includes a $10\%$ $c$-jet fraction in the background training sample. The $\bar{t}t$ sample used has been split in two independent subsamples, one being used for the training and the other for testing the performance of the algorithm. The BDT training parameters have been optimized to reflect the difference between the detectors and limited size of the training sample (500k as opposed to 5M for Run 2). In particular, the depth of the trees has been reduced from 30 to 12. While these parameters have been shown to ensure absence of over-training (by checking the difference in performance between the training and testing samples), the algorithm performance is expected to improve when larger training samples become available.

### 4 Expected Performance

In the following section, the expected performance of the $b$-tagging algorithms introduced above are presented, using the recent ITk layout simulation introduced in [2]. In all cases, the PDFs used in the likelihood-based algorithms have been re-derived to optimize their discrimination power.
Figure 6: One-dimensional projections of the probability density functions used for the SV1 algorithm, corresponding to the vertex mass (a), the energy fraction (b), the number of two-track vertices (c) and the ΔR separation between the jet direction and the direction of the line joining the primary vertex and the secondary vertex (d). Distributions are normalised to the same number of jets with a reconstructed secondary vertex.

The light-jet and c-jet rejection, defined as the inverse of the corresponding background mis-identification probabilities, are shown as a function of the b-jet efficiency for various η and $p_T$ ranges in Fig. 7-10 for the different b-taggers investigated.

The b-tagging performance is in general better in the central region of the detector than in the forward region, as it benefits from better track impact parameter resolutions due to the lower amount of material. Low-$p_T$ jets also suffer in general from the impact parameter resolutions of low-$p_T$ tracks. The worse impact parameter resolutions for forward and low-$p_T$ tracks not only reduce the discrimination between
tracks associated with light jets and $b$-jets but also induce a larger contamination of such pile-up tracks in jets, which affects as well the $b$-tagging performance. The performance for very high-$p_T$ jets is instead limited by the larger fraction of $b$ hadrons flying beyond the innermost pixel layer, leading to a degradation of the impact parameter resolutions for the tracks associated to their decay products, as well as by the dense environment in the core of those jets, making the track reconstruction more challenging [19].

As shown, for all the investigated algorithms, even in the very forward region, a significant discrimination power between $b$-jets and light-jets can be achieved with the ITk. Moreover, the use of a pixel pitch of $25 \times 100 \mu m^2$ improves the light-jet rejection for all the algorithms, with an increase between 10 and 35% compared to $50 \times 50 \mu m^2$, except in the region $2 < |\eta| < 3$. This improvement is expected, considering the improved transverse impact parameter resolution achieved with this configuration, illustrated in Fig. 4. The region $2 < |\eta| < 3$ is associated to a larger amount of material as illustrated in Fig. 1, which counteracts the improvement due to the intrinsic pixel resolution for the $25 \times 100 \mu m^2$ pitch. The improvement observed with the $25 \times 100 \mu m^2$ pixel pitch in terms of light-jets allows flexibility, since operating points which improve $c$-jet rejection at a small cost to the light-jet performance (obtained via dedicated tuning of high-level taggers) can be considered while still matching or improving on the current light-jet rejection. Regarding the choice of the pixel pitch to be used in the ITk pixel detector, the benefits of the $25 \times 100 \mu m^2$ pixel pitch for $b$-tagging will also be balanced against the better expected performance of the pile-up jet rejection algorithms with the $50 \times 50 \mu m^2$ pixel pitch, in particular in the forward region which suffers from a larger pile-up jet multiplicity. The $50 \times 50 \mu m^2$ pixel pitch is the current ITk baseline and used in the rest of this document. The use of a $25 \times 100 \mu m^2$ pixel pitch is considered for a future evolution of the ITk design.

The performance of the IP3D and MV2 taggers, retrained for the ITk detector in Phase-II conditions, with an average of 200 pile-up events, is also compared with the one obtained with the Run 2 algorithms [20], with an average of 30 pile-up events, in Fig. 11. As shown, the Phase-II performance within the acceptance of the current detector is expected to be improved with respect to the Run 2 performance. However it should be noted that the jet selections used to assess the Run 2 performance slightly differ from the ones used for those Phase-II studies. In particular, light jets also include contributions from gluon jets in the $t\bar{t}$ events considered, while this is not the case for the Phase 2 performance. Moreover the Run 2 $b$-tagging performance is assessed after applying the Jet Vertex Tagger (JVT) algorithm, which is used in Run 2 ATLAS analyses [21]. The performance of the JVT algorithm has been measured in collision data to have an efficiency of 92% for jets originating from the hard-scatter vertex and a mis-identification rate of pileup jets, which is approximately 5%. It is assumed here that jets originating from pileup interactions will be rejected at a very high rate with the ITk in an HL-LHC scenario; they are therefore removed from the sample of jets used to assess the b-tagging performance in these studies. Using pile-up jet rejection algorithms optimised for the ITk, the $b$-tagging performance is therefore not expected to be significantly worse than in Run 2, in spite of the significantly larger pile-up expected for HL-LHC. MV2 performance is also expected to be improved in the future thanks to the dedicated optimisation of the JetFitter reconstruction parameters.

For the purposes of Phase-II physics analyses, the MV2 tagging rates for $b$-, $c$-, light-, and pile-up jets are parameterized as functions of jet $p_T$ and $|\eta|$, which are then applied to the particle-level quantities [22]. Figs. 12 and 13 show the efficiency parameterizations for the fixed $\epsilon_b = 70\%$ operating point obtained with the $50 \times 50 \mu m^2$ and $25 \times 100 \mu m^2$ pixel pitch, respectively. Some representative numbers in terms of light-jet rejection are also highlighted in Table 2.
Figure 7: Light-jet and charm-jet rejections vs $b$-tagging efficiency for the IP3D tagging algorithms, evaluated in $t\bar{t}$ events with 200 pile-up events, compared for a pixel pitch of $50 \times 50 \mu m^2$ (full line) and $25 \times 100 \mu m^2$ (dashed line). The results are compared in different $\eta$ regions (a-b) and $p_T$ regions (c-d).
Figure 8: Light-jet and charm-jet rejections vs $b$-tagging efficiency for the SV1 tagging algorithms, evaluated in $t\bar{t}$ events with 200 pile-up events, compared for a pixel pitch of $50 \times 50 \mu m^2$ (full line) and $25 \times 100 \mu m^2$ (dashed line). The results are compared in different $\eta$ regions (a-b) and $p_T$ regions (c-d).
Figure 9: Light-jet and charm-jet rejections vs $b$-tagging efficiency for the IP3D+SV1 tagging algorithms, evaluated in $tt$ events with 200 pile-up events, compared for a pixel pitch of $50 \times 50 \, \mu m^2$ (full line) and $25 \times 100 \, \mu m^2$ (dashed line). The results are compared in different $\eta$ regions (a-b) and $p_T$ regions (c-d).
Figure 10: Light-jet (a) and charm-jet rejections (b) vs $b$-tagging efficiency for the MV2 tagging algorithm, evaluated in $t\bar{t}$ events with 200 pile-up events, compared for a pixel pitch of 50 $\times$ 50 $\mu$m$^2$ (full line) and 25 $\times$ 100 $\mu$m$^2$ (dashed line). The results are compared in different $\eta$ regions (a-b) and $p_T$ regions (c-d).
Figure 11: Light-jet and charm-jet rejections vs $b$-tagging efficiency for the IP3D (a-b) and MV2 (c-d) tagging algorithms, evaluated in $t\bar{t}$ events with 200 pile-up events, obtained with the analogue clustering and a pixel pitch of $50 \times 50 \, \mu m^2$. The results are compared in different $\eta$ regions. For comparison, the performance obtained with the current Run 2 Inner Detector with an average pile-up of 30 is also shown [20].
Figure 12: Light-jet (a), charm-jet (b), b-jet (c), and pile-up jet (d) tagging rates for the MV2 tagging algorithm, parameterized as functions of jet $p_T$ and $|\eta|$. The efficiencies are evaluated in $t\bar{t}$ events with 200 pile-up events, obtained with the analogue clustering and a pixel pitch of $50 \times 50 \text{ m}^2$, for the fixed $\epsilon_b = 70\%$ operating point.

Table 2: Light-jet rejection achieved with the MV2 tagging algorithm for jets in different $\eta$ regions and $p_T$ regimes for operating points corresponding to an average $b$-jet efficiency of $\epsilon_b = 70\%$ and 85% in the relevant $\eta$ region, obtained with the analogue clustering and a pixel pitch of $50 \times 50 \text{ m}^2$. 

| $|\eta| < 2.5$ | $20 < p_T < 100 \text{ GeV}$ | $p_T > 100 \text{ GeV}$ |
|----------------|------------------------------|--------------------------|
| $\epsilon_b = 70\%$ | light-jet rej. = 381 | light-jet rej. = 329 |
| $\epsilon_b = 85\%$ | light-jet rej. = 41 | light-jet rej. = 51 |
| $|\eta| < 4.0$ | $2.5 < |\eta| < 4.0$ | $|\eta| > 4.0$ |
| $\epsilon_b = 70\%$ | light-jet rej. = 51 | light-jet rej. = 59 |
| $\epsilon_b = 85\%$ | light-jet rej. = 6.8 | light-jet rej. = 10.9 |
Figure 13: Light-jet (a), charm-jet (b), $b$-jet (c), and pile-up jet (d) tagging rates for the MV2 tagging algorithm, parameterized as functions of jet $p_T$ and $|\eta|$. The efficiencies are evaluated in $t\bar{t}$ events with 200 pile-up events, obtained with the analogue clustering and a pixel pitch of $25 \times 100 \ \mu m^2$, for the fixed $\epsilon_B = 70\%$ operating point.
5 Impact on di-Higgs Analyses

As highlighted in the previous section, the $b$-tagging performance is expected to be improved over the Run 2 performance with the upgraded ATLAS detector, in spite of the more challenging running conditions at HL-LHC. This will benefit physics analyses involving final states with $b$-quarks, such as the ones targeting the production of a pair of Higgs bosons [23]. At a given light-jet rejection rate of about 300 in the central region of the detector $|\eta| < 2.5$, the $b$-jet efficiency with the MV2 tagger can be increased by around 2% with respect to Run 2, resulting in an increase in acceptance by 4% for final states with two $b$-quarks, such as the $HH \rightarrow b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states and by up to 8% for final states with four $b$-quarks, like in the $HH \rightarrow b\bar{b}bb$ final state. Those numbers are still expected to be improved in the future but represent an evolved understanding of the upgraded detector with respect to Ref. [4]. Taking into account the performance results presented in this note, the expected sensitivity of those analyses with $3 \text{ ab}^{-1}$ has been re-assessed and are presented in Table 3. This in particular includes a more realistic description of the detector material, representing slightly less optimistic conditions than those considered in Ref. [23].

The analysis strategy is unchanged with respect to the one presented in Ref. [23]. In particular, an MV2 working point with a $b$-jet efficiency around 70% is used to identify $b$-jets in the region $|\eta| < 2.5$. The sensitivity of the $HH \rightarrow b\bar{b}bb$ and $b\bar{b}\tau^+\tau^-$ analyses is extrapolated from the existing Run 2 analyses [24, 25], while the sensitivity of the $HH \rightarrow b\bar{b}\gamma\gamma$ is estimated using dedicated Phase-II simulations.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Statistical-only</th>
<th>Statistical + Systematic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$HH \rightarrow b\bar{b}bb$</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>$HH \rightarrow b\bar{b}\tau^+\tau^-$</td>
<td>2.3</td>
<td>2.0</td>
</tr>
<tr>
<td>$HH \rightarrow b\bar{b}\gamma\gamma$</td>
<td>2.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Combined</td>
<td>3.3</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Table 3: Standard Model significance of the individual $HH \rightarrow b\bar{b}bb$, $HH \rightarrow b\bar{b}\tau^+\tau^-$ and $HH \rightarrow b\bar{b}\gamma\gamma$ channels as well as their combination, taking into account the $b$-tagging performance results presented in this note, corresponding to the $50 \times 50 \mu m^2$ pixel pitch. The analysis strategy is unchanged with respect to the one presented in [23].
6 Conclusion

This note presents the $b$-tagging performance expected with the ATLAS detector at HL-LHC, taking into account the latest developments in the ITk simulation. All the $b$-tagging algorithms investigated have benefited from a dedicated reoptimisation to exploit the new layout of the ITk detector. Thanks to this, the $b$-tagging performance is not expected to be significantly worse than in Run 2, in spite of the significantly larger pile-up expected for HL-LHC. This optimisation will be pursued in the future, opening the possibility to further improve the $b$-tagging performance. The study of the performance obtained with a pixel pitch of $25 \times 100$ $\mu$m$^2$ has also been shown to improve the light jet rejection with respect to the $50 \times 50$ $\mu$m$^2$ configuration by 10 to 35% in most of the phase space. The expected $b$-tagging performance will strongly benefit the search for the di-Higgs process, for which the combined sensitivity is expected to be close to $3\sigma$ by the end of the HL-LHC data-taking. Further improvements related to $b$-tagging with the Phase-II ATLAS detector are also expected by then. First of all, the option to reduce the radius of the ITk innermost pixel layer is currently being studied and is expected to improve the $b$-tagging performance significantly. In addition, the use of discriminants based on Deep Neural Networks, such as the DL1 tagger introduced in Run 2 [20], will become possible as larger simulated samples become available for their training and the performance in the forward region is expected to improve thanks to the use of the timing information brought by the High-Granularity Timing Detector [5], which was not yet included in these studies.

References


