Reconstruction of primary vertices at the ATLAS experiment
in Run 1 proton–proton collisions at the LHC

ATLAS Collaboration
CERN, 1211 Geneva 23, Switzerland

Received: 1 December 2016 / Accepted: 5 May 2017 / Published online: 19 May 2017
© CERN for the benefit of the CMS collaboration 2017. This article is an open access publication

Abstract This paper presents the method and performance of primary vertex reconstruction in proton–proton collision data recorded by the ATLAS experiment during Run 1 of the LHC. The studies presented focus on data taken during 2012 at a centre-of-mass energy of \( \sqrt{s} = 8 \) TeV. The performance has been measured as a function of the number of interactions per bunch crossing over a wide range, from one to seventy. The measurement of the position and size of the luminous region and its use as a constraint to improve the primary vertex resolution are discussed. A longitudinal vertex position resolution of about 30 \( \mu m \) is achieved for events with high multiplicity of reconstructed tracks. The transverse position resolution is better than 20 \( \mu m \) and is dominated by the precision on the size of the luminous region. An analytical model is proposed to describe the primary vertex reconstruction efficiency as a function of the number of interactions per bunch crossing and of the longitudinal size of the luminous region. Agreement between the data and the predictions of this model is better than 3% up to seventy interactions per bunch crossing.

Contents

1 Introduction ............................................. 1
2 The ATLAS detector and LHC beam parameters ... 2
  2.1 The ATLAS inner detector .......................... 2
  2.2 The minimum-bias trigger .......................... 2
  2.3 Determination of pile-up interactions .......... 3
  2.4 Parameters affecting the luminous region
    at the LHC ..................................... 3
3 Data and Monte Carlo samples ...................... 4
4 Primary vertex reconstruction ..................... 5
  4.1 Track reconstruction ............................ 5
  4.2 Primary vertex finding and fitting .......... 6
  4.3 Beam-spot reconstruction ...................... 7
  4.4 Beam-spot stability ............................ 8
5 Hard-scatter interaction vertices ................ 10

1 Introduction
Efficient and precise reconstruction of primary vertices, defined as the points in space where proton–proton (pp) interactions have occurred, is an important element of data analysis at the LHC. It is of direct relevance to the reconstruction of hard-scatter interactions, in which the correct assignment of charged-particle trajectories to the hard-scatter primary vertex is essential in reconstructing the full kinematic properties of the event. An aspect of primary vertex reconstruction requiring special attention is the superposition of multiple inelastic pp interactions reconstructed as a single physics event with many primary vertices. These additional primary vertices, which are usually soft-QCD interactions related to the dominant components of the total cross section, are referred to as pile-up. The average number of inelastic pp interactions per bunch crossing under constant beam conditions is denoted as \( \mu \) and is directly related to the instantaneous luminosity [1]. The primary vertex reconstruction is also important for the determination of the luminous region, or beam spot, where collisions take place within the ATLAS detector.
This paper describes the performance of primary vertex reconstruction with the ATLAS detector, during Run 1 of the LHC from 2010 to 2012. The studies presented here are based on the data collected in 2012 at a proton–proton centre-of-mass energy $\sqrt{s} = 8$ TeV. Averaged over the 2012 dataset, $\mu$ was approximately 20. The 2012 data are representative of the full set of data taken from 2010 to 2012 in terms of the primary vertex performance. Studies in this paper make use of dedicated datasets recorded at very low values of $\mu$ ($\mu = 0.01$), thereby providing a measurement of the performance in the absence of pile-up. Data recorded with the highest number of interactions per bunch crossing, leading to values of $\mu$ up to 72, are used to study the various mechanisms that lead to a degradation of the primary vertex reconstruction as pile-up increases.

The paper is organised as follows: Sect. 2 provides a brief description of the ATLAS detector, a description of pile-up determination and a discussion of the parameters of the LHC accelerator that determine the size of the luminous region. Section 3 describes the data and Monte Carlo (MC) simulation samples used. Section 4 presents the algorithms for primary vertex reconstruction in ATLAS. The measurement and stability of the beam-spot parameters and their use as a constraint in primary vertex reconstruction are discussed. The predicted impact of pile-up contamination on the reconstruction and selection of primary vertices from hard-scatter processes is discussed in Sect. 5. Studies of single vertex reconstruction in minimum-bias data and the related comparisons to MC simulation are presented in Sect. 6. Section 7 describes the performance of vertex reconstruction in high pile-up conditions. In Sect. 8, the results of studies presented in Sects. 5 through 7 are used to model the efficiency of primary vertex reconstruction in simulation, to predict its behaviour at high pile-up, and to compare the predictions to data. Summary and conclusions are presented in Sect. 9.

2 The ATLAS detector and LHC beam parameters

The ATLAS detector [2] is a multi-purpose detector with a cylindrical geometry. It is comprised of an inner detector (ID) surrounded by a thin superconducting solenoid, a calorimeter system and a muon spectrometer embedded in a toroidal magnetic field. The ID is the primary detector used for vertex reconstruction and it is described in further detail below in Sect. 2.1. Outside of the ID and the solenoid are electromagnetic sampling calorimeters made of liquid argon as the active material and lead as an absorber. Surrounding the electromagnetic calorimeter is the iron and scintillator tile calorimeter for hadronic energy measurements. In the forward regions it is complemented by two end-cap calorimeters made of liquid argon and copper or tungsten. The muon spectrometer surrounds the calorimeters and consists of three large superconducting eight-coil toroids, a system of tracking chambers, and detectors for triggering.

2.1 The ATLAS inner detector

The inner detector covers the pseudorapidity\(^1\) range $|\eta| < 2.5$. Schematic views of the Run 1 inner detector are presented in Fig. 1. Particle trajectories are identified using the combined information from the sub-detectors of the ID: the innermost silicon pixel detector, the surrounding silicon microstrip semiconductor tracker (SCT), and the transition radiation tracker (TRT), made of straw tubes filled with a Xe–CO\(_2\) gas mixture [3]. All three sub-systems are divided into a barrel section and two end-caps. The barrel sections consist of several cylindrical layers, while the end-caps are composed of radial disks and wheels. The sensitive regions of the three sub-detectors cover radial distances in the barrel section from 50.5 to 122.5, 299 to 514, and 554 to 1082 mm. Typical position resolutions are 10, 17, and 130 $\mu$m for the transverse coordinate in the pixel detector, the SCT, and the TRT respectively. In the case of the pixel and SCT, the resolutions in the z-coordinate are 115 and 580 $\mu$m. The superconducting solenoid coil around the tracking system produces a 2 T axial magnetic field. A track from a charged particle traversing the barrel detector would typically have 11 measurements in the silicon detector\(^2\) (3 pixel clusters and 8 strip clusters) and more than 30 measurements in the TRT [4].

2.2 The minimum-bias trigger

A minimum-bias trigger was used to select the data presented in this paper. This trigger is designed to record a random selection of bunch crossings, unbiased by any hard physics produced in the bunch crossing, by using a signal from the minimum-bias trigger scintillators (MBTS). The MBTS are mounted at each end of the detector in front of the liquid-argon end-cap calorimeter cryostats at $z = \pm 3.56$ m, covering the range $2.09 < |\eta| < 3.84$. The MBTS trigger used for this paper requires one hit above threshold from either side of the detector, referred to as a single-arm trigger [4].

\(^1\) The ATLAS experiment 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 direction. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates ($r, \phi$) are used in the transverse ($x, y$) plane, $\phi$ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$.

\(^2\) Measurements of charged particle trajectories in the pixel, SCT and TRT are called ID hits.
2.3 Determination of pile-up interactions

Depending on the length of the read-out window of a sub-detector, signals from neighbouring bunch crossings can be present simultaneously when the detector is read out. The impact of interactions from the neighbouring bunch crossings is referred to as out-of-time pile-up, while in-time pile-up results from the presence of multiple \( pp \) interactions in the same bunch crossing.

During most of Run 1 of the LHC, the separation of proton bunches was 50 ns. The timing resolution of the inner detector components is about 25 ns. This is sufficient for the out-of-time pile-up to have a much smaller impact on ID measurements than the in-time pile-up. As a consequence the number of reconstructed vertices is a direct measure of the amount of in-time pile-up on an event-by-event basis.

The instantaneous luminosity, \( L \), can be expressed in terms of the visible interaction rate, \( R_{\text{vis}}^{\text{inel}} \), and the visible inelastic cross section, \( \sigma_{\text{vis}}^{\text{inel}} \), as:

\[
L = \frac{R_{\text{vis}}^{\text{inel}}}{\sigma_{\text{vis}}^{\text{inel}}}. \tag{1}
\]

The inelastic cross section, \( \sigma_{\text{inel}} \), and the visible inelastic cross section are related through: \( \sigma_{\text{vis}}^{\text{inel}} = \epsilon \sigma_{\text{inel}} \). Here \( \epsilon \) is the efficiency of the detector to record an inelastic collision. The inelastic cross section is defined as the total cross section minus the elastic cross section.

In practice, the full rate of inelastic collisions is never directly measured. Only a fraction of it is observable in the detector due to the \( \eta \) acceptance. The luminosity is measured using a set of dedicated detectors which allow bunch-by-bunch measurements. The luminosity detectors are calibrated using dedicated Van der Meer scans [5]. The uncertainty in the luminosity measurement is 1.9% [1].

The number of \( pp \) inelastic interactions per bunch crossing follows a Poisson distribution with mean value \( \mu \). Assuming that the \( pp \) collider operates at a revolution frequency \( f_r \) with \( n_b \) interacting bunches per beam, the luminosity can also be expressed as:

\[
L = \frac{\mu n_b f_r}{\sigma_{\text{inel}}}. \tag{2}
\]

The value of \( \mu \) changes during data-taking as a function of time: it decreases with decreasing beam intensity and increasing emittance. The highest value is at the start of the stable beam period of the fill. For the studies presented in this paper, \( \mu \) is calculated using Eq. (2). The value of the inelastic cross section at 8 TeV centre-of-mass energy is 71.5 mb, taken from the PYTHIA8 MC generator [6]. Experimental measurements [7,8] are found to be compatible with the cross section predicted by PYTHIA8. The overall uncertainty in \( \mu \) is 4%, which is derived from the quadratic sum of the uncertainties in the luminosity and in the inelastic cross section.

2.4 Parameters affecting the luminous region at the LHC

The size, position and shape of the luminous region, or beam spot, are determined by the operating parameters of the beams and magnets of the LHC [9]. The transverse size is determined by the focusing of the LHC beams near the interaction region and by the spread in position–momentum phase space of the protons within the colliding bunches. The latter is quantified by the geometric emittance \( \epsilon \) of the beams, or equivalently by the normalised emittance defined as \( \epsilon_N = \beta_v \gamma \epsilon \), where \( \beta_v \) and \( \gamma \) are the relativistic functions \( \beta_v = v/c \simeq 1 \) and \( \gamma = E_{\text{beam}}/m_p \). \( E_{\text{beam}} \) is the beam energy and \( m_p \) is the mass of the proton. The focusing of the beams is characterised by the \( \beta \)-function, and especially its minimum value.

---

Fig. 1  Schematic views of the ATLAS Run 1 inner detector: a barrel and end-cap sections; b cross section of the barrel section showing the TRT, SCT, and pixel sub-detectors.
Table 1  Summary of LHC parameters for typical pp collision fills and corresponding expected sizes of the luminous region. Emittance and bunch length values (and the corresponding beam-spot sizes) refer to values expected at the start of a fill. The two values given for expected transverse and longitudinal beam-spot size in 2011 correspond to the two \( \beta^* \) settings of 1.5 and 1.0 m. Measured average beam-spot parameters are presented in Table 3 (Sect. 4.4).

<table>
<thead>
<tr>
<th>Year</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam energy (TeV)</td>
<td>3.5</td>
<td>4.0</td>
</tr>
<tr>
<td>( \beta^* ) (m)</td>
<td>1.5, 1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Normalised emittance ( \varepsilon_N ) (( \mu \text{m rad} ))</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Full crossing angle ( \phi ) (( \mu \text{rad} ))</td>
<td>240</td>
<td>290</td>
</tr>
<tr>
<td>( 4\sigma ) bunch length ( T_z ) (ns)</td>
<td>1.20</td>
<td>1.25</td>
</tr>
<tr>
<td>Bunch length ( \sigma_z ) (mm)</td>
<td>90</td>
<td>94</td>
</tr>
<tr>
<td>Expected transverse beam-spot size ( \sigma_x, \sigma_L ) (( \mu \text{m} ))</td>
<td>22, 18</td>
<td>22, 18</td>
</tr>
<tr>
<td>Expected longitudinal beam-spot size ( \sigma_z ) (mm)</td>
<td>60, 59</td>
<td>54</td>
</tr>
</tbody>
</table>

\( \beta^* \). The longitudinal size of the luminous region is determined by the bunch length and by the angle \( \phi \) (full crossing angle) at which the two beams are brought into collision. In the following discussion it is assumed that the emittances and \( \beta \)-functions in the horizontal and vertical direction are the same for each of the two beams. These assumptions lead to a circular transverse beam profile, as has been observed to be approximately the case at the LHC.

The particle densities in proton bunches can be described by three-dimensional Gaussian distributions with transverse and longitudinal sizes given by \( \sigma_x = \sigma_y = \sqrt{\varepsilon} \beta \) and \( \sigma_z = c T_z/4 \) respectively, where \( T_z \) is the “four \( \sigma \) bunch length” (in ns) customarily quoted for the LHC. Because the ratio \( \sigma_z/\beta^* \) was small during Run 1, the quadratic form of the \( \beta \)-function around the interaction region had a negligible effect over the length of the luminous region and the transverse beam size along the beam axis remained constant. As a result the luminous region is described well by a three-dimensional Gaussian distribution. With the assumption of pair-wise equal bunch sizes mentioned above, the transverse size \( \sigma_x \) (and equivalently \( \sigma_y \)) of the luminous region is given by \( \sigma_x = \sigma_x/\sqrt{2} \). For a crossing angle in the vertical plane as is the case for ATLAS, and assuming equal longitudinal bunch sizes \( \sigma_z \) in both beams, the longitudinal size of the luminous region is given by:

\[
\sigma_L = \frac{c T_z/4}{\sqrt{2}} \frac{1}{\sqrt{1 + \left( \frac{\sigma_z \phi}{\sigma_y} \right)^2}}. \tag{3}
\]

A summary of typical LHC parameters for pp collisions at \( \sqrt{s} = 7 \text{ TeV} \) in 2011 and at \( \sqrt{s} = 8 \text{ TeV} \) in 2012 is shown in Table 1 together with the resulting expected sizes of the luminous region. The measured sizes of the luminous region are discussed in Sect. 4.4 and Table 3.

3 Data and Monte Carlo samples

This paper uses pp collision data with \( \sqrt{s} = 8 \text{ TeV} \) recorded during the LHC Run 1 period. Data were collected using the minimum-bias triggers described in Sect. 2. The data-taking conditions of the corresponding data samples are summarised in Table 2. The studies presented here aim to cover the full range of Run 1 \( \mu \) values and use both a special high-\( \mu \) data sample as well as a range of lower-\( \mu \) data. The distribution of the average number of interactions per pp bunch crossing in Run 1 is shown in Fig. 2. This does not include the special high and low \( \mu \) runs listed in Table 2. Most data taken in Run 1 had pile-up near \( \mu = 20 \). The low pile-up dataset was taken at average \( \mu \) around 0.01, while the special high pile-up run featured peak collision multiplicities up to \( \mu = 72 \). The results presented in this paper use MC simulation of hard-scatter interactions and soft inelastic pp collisions. The collection of soft inelastic interactions is referred to here as the minimum-bias sample. These are events that would have been collected with the minimum-bias trigger, described in Sect. 2.2, and they represent an average beam crossing, without selection of a specific hard-scatter interaction.

Table 2  The data-taking conditions of the pp collision data samples used in this paper

<table>
<thead>
<tr>
<th>Pile-up conditions</th>
<th>( \mu ) range</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low ( \mu )</td>
<td>0–1</td>
<td>April 2012</td>
</tr>
<tr>
<td>High ( \mu )</td>
<td>55–72</td>
<td>July 2012</td>
</tr>
<tr>
<td>Run 1 data range</td>
<td>7–40</td>
<td>2012</td>
</tr>
</tbody>
</table>

Fig. 2  The average number of interactions per proton bunch crossing, \( \mu \), during 8 TeV data-taking in Run 1, weighted by the luminosity
Minimum-bias samples were simulated with the PYTHIA8 MC generator, with the A2 set of tuned parameters [10] and the MSTW2008LO parton density function set [11]. The PYTHIA8 model for soft QCD uses a phenomenological adaptation of $2 \rightarrow 2$ parton scattering to describe low transverse momentum processes. Samples were generated for non-diffractive, single-diffractive, and double-diffractive interactions. These contributions were combined according to the PYTHIA8 generator cross sections.

To study the collective effects of multiple primary vertices reconstructed in one beam crossing, MC simulation with no hard-scattering process but only pile-up was created for $\mu$ up to 72. These samples mimic randomly triggered events, and were also generated with PYTHIA8 using the A2 tune. A special configuration was used to match 2012 data-taking conditions, including the beam spot with $z$-direction size equal to the average observed in data.

Hard-scatter interactions were simulated with POWHEG [12] interfaced to PYTHIA8 for the $Z \rightarrow \mu\mu$ and $H \rightarrow \gamma\gamma$ processes, and MC@NLO [13], HERWIG [14] and Jimmy [15] for top-quark pair production ($t\bar{t}$). The CT10 parameterisation [16] of the parton density functions was used. The top-quark pairs were generated with a lepton filter, requiring a lepton in the final state. The hard-scatter interaction samples were generated for a range of pile-up between $\mu = 0$ and 38. The overlaid pile-up collisions were simulated with the soft QCD processes of PYTHIA8 in the manner of the minimum-bias simulation described above.

All generated events are processed with the ATLAS detector simulation framework [17], using the GEANT 4 [18] toolkit. After full detector simulation, the MC events are reconstructed and analysed in the same manner as data.

When comparing data with simulation in the presence of pile-up interactions, the average number of collisions per bunch crossing in simulation is re-weighted to match that measured in data. In order to obtain the same visible cross section for $pp$ interactions for the simulation and data, a $\mu$-rescaling is also applied before the re-weighting. The rescaling factor is calculated by comparing the ratio of the visible cross section to the total inelastic cross section, $\epsilon_{vis} = \sigma_{inel}^{vis}/\sigma_{inel}$, for data with that for simulation. The value of $\epsilon_{vis}^{data}$ is computed from independent measurements of these cross sections in data [19,20]. The value of $\epsilon_{vis}^{MC}$ is computed from events simulated with the PYTHIA8 MC generator with the A2 tune. The final scale factor is corrected to match the visible cross section within the ATLAS inner detector acceptance, resulting in $\epsilon_{vis}^{MC}/\epsilon_{vis}^{data} = 1.11$. The uncertainty in this scale factor is 5%. It is calculated from the quadrature sum of the uncertainties in the cross-section measurements, 3.5 and 2.6% from Refs. [19,20] respectively, and a 2% uncertainty in the extrapolation from 7 to 8 TeV and to the inner detector acceptance.

4 Primary vertex reconstruction

This section describes the method for reconstructing primary vertices. The input to the vertex reconstruction is a collection of reconstructed tracks. A brief summary of the main steps of track reconstruction is presented in Sect. 4.1. The vertex reconstruction is presented in Sect. 4.2. This is followed by a description of how primary vertices are used to reconstruct the shape of the luminous region, or beam spot, in Sect. 4.3, and a description of the stability of the beam spot in Sect. 4.4.

4.1 Track reconstruction

The reconstruction of charged-particle trajectories in the inner detector is based on fitting a trajectory model to a set of measurements. The reconstructed charged-particle trajectories are hereafter referred to as tracks. The general structure and performance of ATLAS track reconstruction is described in detail in Refs. [21,22] and a brief overview is given below.

Track seeds consist of three measurements in different layers of the pixel detector and SCT. Tracks are propagated out from the seed towards the TRT (“inside-out”) using a combinatorial Kalman filter [22], and additional silicon hits are added to the seed. An ambiguity solving procedure is applied to remove track candidates with incorrectly assigned hits. The candidate tracks are scored in a reward–penalty schema with respect to one another. To favour fully reconstructed tracks over short track segments, each additional measurement associated with a track leads to a better score value. The measurements from different sub-detectors are weighted differently, preferring the precision measurements (e.g. pixel clusters) and downgrading measurements from less precise detector parts. To provide a realistic description of detector acceptance and efficiency, the concept of a hole on a track is introduced. A hole represents a measurement on a detector surface that is expected, given the trajectory predictions, but not observed (holes are not considered on the first and last surfaces in the measurement). The presence of holes reduces the overall track score. The $\chi^2$ of the track fit is also used to penalise poor-quality candidates. Finally, the logarithm of the track transverse momentum $\ln(p_T)$ is considered as a criterion to promote energetic tracks and to suppress the larger number of tracks formed from incorrect combinations of clusters, which tend to have low measured $p_T$. After the reconstruction of tracks in the pixel and the SCT detectors, the successful candidates are extrapolated into the TRT volume and combined with measurements there.
During data-taking at $\sqrt{s} = 8$ TeV, the input to the vertex reconstruction algorithms consisted of charged-particle tracks selected according to the following criteria:

- $p_T > 400$ MeV; $|d_0| < 4$ mm; $\sigma(d_0) < 5$ mm; $\sigma(z_0) < 10$ mm;
- At least four hits in the SCT detector;
- At least nine silicon (SCT or pixel) hits;
- No pixel holes.

Here the symbols $d_0$ and $z_0$ denote the transverse and longitudinal impact parameters of tracks with respect to the centre of the luminous region, and $\sigma(d_0)$ and $\sigma(z_0)$ denote the corresponding uncertainties [21]. The impact parameter requirements are applied to reduce contamination from tracks originating from secondary interactions. The above requirements are tighter than the standard ATLAS track selection criteria in order to maintain a low rate of fake tracks (tracks mistakenly reconstructed from a random combination of hits) at Run 1 pile-up levels (up to $\mu = 40$). The track reconstruction efficiency under this selection is between 75 and 85% for central rapidities ($|\eta| < 1.5$) and track $p_T$ above 500 MeV; the efficiency falls to about 60% at higher rapidities or about 65% for tracks with $p_T$ between 400 and 500 MeV.

4.2 Primary vertex finding and fitting

The procedure of primary vertex reconstruction is divided into two stages: vertex finding and vertex fitting [23]. The former stage generally denotes the pattern recognition process: the association of reconstructed tracks to vertex candidates. The vertex fitting stage deals with reconstruction of the actual vertex position and its covariance matrix. The strategy is explained in detail in this section, and can be briefly outlined in these steps:

- A set of tracks satisfying the track selection criteria is defined.
- A seed position for the first vertex is selected.
- The tracks and the seed are used to estimate the best vertex position with a fit. The fit is an iterative procedure, and in each iteration less compatible tracks are down-weighted and the vertex position is recomputed.
- After the vertex position is determined, tracks that are incompatible with the vertex are removed from it and allowed to be used in the determination of another vertex.
- The procedure is repeated with the remaining tracks in the event.

Each of these steps (except the track selection described in the previous section) is expanded on below.

1. The seed position of the vertex fit is based on the beam spot in the transverse plane. The $x$- and $y$-coordinates of the starting point are taken from the centre of the beam spot, reconstructed as discussed in Sect. 4.3. The $z$-coordinate of the starting point is calculated as the mode of the $z$-coordinates of tracks at their respective points of closest approach to the reconstructed centre of the beam spot. The mode is calculated using the Half-Sample Mode algorithm [24].

2. After the seed has been determined, the iterative primary vertex finding procedure begins. The vertex position is determined using an adaptive vertex fitting algorithm with an annealing procedure [25]. Using the seed position as the starting point and parameters of reconstructed tracks as input measurements, the algorithm performs an iterative $\chi^2$ minimisation, finding the optimal vertex position. Each input track is assigned a weight, reflecting its compatibility with the vertex estimate. The vertex position is recalculated using the weighted tracks, and then the procedure is repeated, recalculating track weights with respect to the new vertex position. The individual track weights are calculated according to the following equation:

$$\omega(\chi^2) = \frac{1}{1 + \exp \left( \frac{\chi^2 - \chi^2_{\text{cutoff}}}{2\sigma} \right)}.$$  \hspace{1cm} (4)

Here $\hat{\chi}^2$ is the $\chi^2$ value calculated in three dimensions between the last estimated vertex position and the respective point of the closest approach of the track. Tracks with lower weights are less compatible with the vertex and will have less influence on the position calculation. The constant $\chi^2_{\text{cutoff}}$ defines the threshold where the weight of an individual track becomes equal to 0.5. Tracks with low weights are not removed, but will have less impact on the calculated vertex position. The value of $\chi^2_{\text{cutoff}}$ is set to nine, which corresponds to about three standard deviations. The temperature $T$ controls the smoothness of the weighting procedure. For low values of $T$, $\omega(\chi^2)$ approaches a step function, and for large values of $T$ the function flattens, progressively losing its $\chi^2$ dependence. To avoid convergence in local minima, the weighting procedure is applied progressively by decreasing the temperature $T$ during the fit iterations. The temperature is lowered from some high starting value in a pre-defined sequence of steps that converges at $T = 1$. A typical distribution of track weights is shown in Fig. 3. It widens as $T$ decreases, reaching an optimal separation of track outliers for $T = 1$.

3. After the last iteration, the final weight of each track used in the vertex fit is evaluated. Tracks found incompatible with the vertex by more than seven standard deviations are removed from the vertex candidate and returned to the
Track weight within vertex fit

![Histogram showing the weights applied to tracks in the vertex reconstruction fit.](image)

Fig. 3 Histogram showing the weights applied to tracks in the vertex reconstruction fit. The fitting algorithm iterates through progressively smaller values of the temperature $T$, effectively down-weighting outlying tracks in the vertex fit. The vertical axis is on a logarithmic scale.

Pool of unused tracks. This loose requirement is intended to reduce the number of single $pp$ interactions which are reconstructed as two distinct primary vertices due to the presence of track outliers, while maintaining a high efficiency.

4. After the vertex candidate is created, the rejected tracks are considered as input for a new vertex finding iteration. The procedure described above is then repeated starting from step 1, calculating the new starting position from remaining tracks, until no unassociated tracks are left in the event or no additional vertex can be found in the remaining set of tracks.

All vertices with at least two associated tracks are retained as valid primary vertex candidates. The output of the vertex reconstruction algorithm is a set of three dimensional vertex positions and their covariance matrices. Figure 4 shows a typical distribution for the number of reconstructed vertices per event in Run 1 for minimum-bias data collected in the pile-up range $21 < \mu < 23$.

The reconstructed position and width of the beam spot can be used as an additional measurement during the primary vertex fit. It is taken as a three-dimensional Gaussian measurement centred around the beam-spot centre and with the beam-spot size as the width. Tracks outside the beam spot have low compatibility with the vertex fit and are thus removed in the iterative fitting procedure. This procedure is hereafter referred to as the beam-spot constraint.

4.3 Beam-spot reconstruction

The beam-spot reconstruction is based on an unbinned maximum-likelihood fit to the spatial distribution of primary vertices collected from many events. These primary vertices are reconstructed without beam-spot constraint from a representative subset of the data called the express stream during the detector calibration performed approximately every ten minutes. In each event only the primary vertex with the highest sum of squares of transverse momenta of contributing tracks, denoted hereafter as $\sum p_T^2$, is considered. In order to be used in the beam-spot fit, this vertex must include at least five tracks and must have a probability of the $\chi^2$ of the vertex fit greater than 0.1%. The requirement of at least five tracks ensures that most vertices have a transverse vertex resolution better than 50 $\mu$m with a most probable value of about 15 $\mu$m that is comparable to the transverse beam-spot size. At least 100 selected vertices are required to perform a beam-spot fit, and in a typical fit several thousand vertices collected over a time period of about ten minutes are available. The fit extracts the centroid position ($x_L$, $y_L$, $z_L$), the tilt angles $x'_L$ and $y'_L$, in the $x$–$z$ and $y$–$z$ planes respectively, and the luminous sizes ($\sigma_x L$, $\sigma_y L$, $\sigma_z L$), which are the measured sizes of the luminous region with the vertex resolution deconvoluted from the measurements.
In the transverse plane the width of the distribution of primary vertices is the convolution of the vertex resolution with the width of the luminous region. This is modelled by the transverse covariance matrix

\[ V_i = V_B^2 + k^2 V_i^V, \]  

(5)

where \( V_B \) describes the transverse beam-spot size and allows for a rotation of the luminous-region ellipsoid in the transverse plane in case of non-circular beams. The transverse vertex resolution \( V_i^V \) estimated by the vertex fit for each primary vertex \( i \) is scaled by a parameter \( k \) determined by the beam-spot fit in order to account for any differences between fitted and expected vertex resolutions. The parameter \( k \) is expected to be close to unity as long as the vertex fitter provides good estimates of the vertex position uncertainty, the contamination from secondary vertices among the primary vertex candidates used in the beam-spot fit is small, and the Gaussian fit model provides an adequate description of the beam-spot shape. During 2012, the average value of \( k \) was 1.16. No vertex resolution correction and no error scaling is applied in the longitudinal direction because the longitudinal beam-spot size of about 50 mm is much larger than the typical \( z \) resolution of 35 \( \mu \)m for the vertices selected for the beam-spot fit.

The beam-spot fit assumes a Gaussian shape in \( x, y \) and \( z \) and the corresponding probability density function (PDF) is maximised using the MINUIT \[26\] minimisation package after an iterative procedure removes a small number of outliers incompatible with the fit. The effect of this outlier removal on the fitted beam-spot parameters is negligible but brings the error scaling factor \( k \) closer to 1.

As an example of the beam-spot fit, Fig. 5 shows the distribution of primary vertices selected as input to the beam-spot fit (before outlier removal), together with the projection of the fit result. The fitted beam spot, i.e. the distribution of primary vertices after unfolding of the vertex position resolution, is also shown. The impact of the vertex position resolution is clearly seen in the transverse direction, whereas in the longitudinal \((z)\) direction the vertex resolution is negligible compared to the beam spot and therefore fitted beam spot and fit projection are identical.

### 4.4 Beam-spot stability

The evolution of the beam-spot position and size as a function of time during a typical LHC fill is shown in Fig. 6. The coordinates of the beam-spot position are given with respect to the ATLAS coordinate system. The precise origin location and the orientation of the ATLAS coordinate system is defined through the detector alignment procedure. The origin was chosen to be at the nominal interaction point with a \( z \)-axis along the beam direction, ensuring that the coordinates of the beam-spot centroid position are close to zero. In the early Run 1 data, a tilt angle of \( x_L \approx 500 \mu \)rad was observed.
Fig. 6 Position (a $x_L$, c $y_L$, e $z_L$) and size (b $\sigma_x L$, d $\sigma_y L$, f $\sigma_z L$) of the luminous region in ATLAS during a typical fill at $\sqrt{s} = 8$ TeV. The transverse sizes are corrected for the transverse vertex resolution.

In 2011 the ATLAS coordinate system was rotated in order to align the coordinate system more precisely with the beam line.

The downward movement of the beam-spot position during the first 40 min of the run followed by a gradual rise as seen in Fig. 6c is typical and is attributed to movement of the pixel detector after powering up from standby. The increase in transverse size during the fill (Fig. 6b, d) is expected from the transverse-emittance growth of the beams. The magnitude of the changes in longitudinal beam-spot position (Fig. 6e) is typical and is understood to be due to relative RF phase drift. The increase in longitudinal size (Fig. 6f) reflects bunch lengthening in the beams during the fill. The tilt angles $x_L'$ and $y_L'$ (not shown in Fig. 6) were stable at the level of about 10 $\mu$rad.

The long-term evolution of the beam-spot position during 2012 is shown in Fig. 7. The large vertical movement at the beginning of May visible in Fig. 7b was associated with movement of the ID. Apart from variations in each fill due to transverse-emittance growth and bunch lengthening, both the transverse and longitudinal beam-spot sizes remained unchanged during 2012.

Table 3 summarises the beam-spot position and size in 2010, 2011 and 2012 for $pp$ collision data.

Data from special runs is excluded. As expected, the average transverse beam-spot size scales approximately with $\sqrt{\beta^*/E_{beam}}$, but is also influenced by changes in the normalised emittance and by the amount of emittance growth during the fills. In 2010 and 2011 the centre-of-mass energy was 7 TeV. In 2012 it increased to 8 TeV. During this time the crossing angle $\phi$ was increased from zero at the start of 2010 to 290 $\mu$rad in 2012.

The measured transverse size of the beam spot at the start of a run is in good agreement with the values expected from the LHC machine parameters at the start of a fill (Table 1). This can be seen in Fig. 6. The average transverse size in 2012 shown in Table 3 (15 $\mu$m) is larger than the expected size of 13 $\mu$m from Table 1 due to emittance growth dur-
Table 3 Average beam-spot position and size for \( pp \) collision data in 2010, 2011 and 2012 for different \( \beta^* \) settings. The errors given in the table are the RMS spread of the parameters during the corresponding time period.

<table>
<thead>
<tr>
<th>Year</th>
<th>( \beta^* ) (m)</th>
<th>( x_L ) (mm)</th>
<th>( y_L ) (mm)</th>
<th>( z_L ) (mm)</th>
<th>( \sigma_x ) (µm)</th>
<th>( \sigma_y ) (µm)</th>
<th>( \sigma_z ) (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>11</td>
<td>(-0.347 \pm 0.015)</td>
<td>(0.611 \pm 0.018)</td>
<td>(0.9 \pm 3.5)</td>
<td>(49 \pm 8)</td>
<td>(60 \pm 12)</td>
<td>(29 \pm 3)</td>
</tr>
<tr>
<td>2010</td>
<td>2</td>
<td>(-0.364 \pm 0.031)</td>
<td>(0.647 \pm 0.009)</td>
<td>(-1.2 \pm 2.2)</td>
<td>(30 \pm 5)</td>
<td>(39 \pm 12)</td>
<td>(36 \pm 3)</td>
</tr>
<tr>
<td>2010</td>
<td>3.5</td>
<td>(0.081 \pm 0.033)</td>
<td>(1.099 \pm 0.029)</td>
<td>(-3.0 \pm 4.6)</td>
<td>(41 \pm 4)</td>
<td>(44 \pm 6)</td>
<td>(63 \pm 3)</td>
</tr>
<tr>
<td>2011</td>
<td>1.5</td>
<td>(-0.050 \pm 0.018)</td>
<td>(1.059 \pm 0.051)</td>
<td>(-6.2 \pm 3.8)</td>
<td>(26 \pm 2)</td>
<td>(24 \pm 2)</td>
<td>(57 \pm 3)</td>
</tr>
<tr>
<td>2011</td>
<td>1.0</td>
<td>(-0.052 \pm 0.009)</td>
<td>(1.067 \pm 0.013)</td>
<td>(-6.7 \pm 1.5)</td>
<td>(21 \pm 2)</td>
<td>(20 \pm 1)</td>
<td>(56 \pm 3)</td>
</tr>
<tr>
<td>2012</td>
<td>0.6</td>
<td>(-0.291 \pm 0.016)</td>
<td>(0.705 \pm 0.046)</td>
<td>(-7.3 \pm 4.7)</td>
<td>(15 \pm 2)</td>
<td>(15 \pm 1)</td>
<td>(48 \pm 2)</td>
</tr>
</tbody>
</table>

5 Hard-scatter interaction vertices

This section describes how both the reconstruction and identification efficiencies of hard-scatter primary vertices are evaluated using simulation. The impact of pile-up tracks and vertices on the performance is also estimated. A classification scheme based on MC generator-level information, denoted hereafter as truth-level information, is used to describe the level of pile-up contamination in reconstructed vertices from hard-scatter processes.

5.1 Monte Carlo truth matching and classification of vertices

To study the performance of primary vertex reconstruction using MC simulation, a truth-matching algorithm has been developed, based on the generator-level particles associated to tracks contributing to reconstructed vertices. The procedure first classifies each reconstructed track used in a vertex fit. The compatibility criteria for track truth-matching are based on the fraction of hits used to reconstruct the track in each sub-detector that were produced by the generated primary particle as discussed in Ref. [21]. Each reconstructed track is classified as one of the following:

- A track matched to a hard-scatter interaction.
- A track matched to a pile-up interaction.
- An unmatched track. Such a tracks are considered random combinations of detector hits falsely identified as charged particle trajectories. These are referred to as fake tracks.

Tracks are matched to their primary generating interaction, i.e. tracks from secondary interactions are traced back to a hard-scatter or pile-up interaction. Based on the above classification, reconstructed vertices can be categorised. For each vertex, the sum of the weights assigned to all contributing tracks is normalised to unity. The fractional weights of

![Fig. 7 Position of the luminous region in ATLAS over the course of \( pp \) running in 2012 at \( \sqrt{s} = 8 \) TeV. The data points are the result of a maximum likelihood fit to the spatial distribution of primary vertices collected over ten minutes. Errors are statistical only.](image)
individual tracks in each vertex are calculated. Vertices can then be put into one of the following exclusive categories:

- **Matched vertex** Tracks identified as coming from the same generated interaction contribute at least 70% of the total weight of tracks fitted to the reconstructed vertex.
- **Merged vertex** No single generated interaction contributes more than 70% of track weight to the reconstructed vertex. Two or more generated interactions contribute to the reconstructed vertex.
- **Split vertex** The generated interaction with the largest contribution to the reconstructed vertex is also the largest contributor to one or more other reconstructed vertices. In this case, the reconstructed vertex with the highest fraction of track $\Sigma p_T^2$ is categorised as matched or merged and the vertex or vertices with lower $\Sigma p_T^2$ are categorised as split.
- **Fake vertex** Fake tracks contribute more weight to the reconstructed vertex than any generated interaction.

This classification schema allows detailed studies of vertex reconstruction in a pile-up environment. The effects of splitting and merging of primary vertices as well as the influence of these effects on the vertex reconstruction efficiency and primary vertex resolution can be studied. This schema also allows the reconstructed vertices to be associated either with the primary hard-scatter $pp$ collision or with pile-up interactions.

When studying the hard-scatter $pp$ collisions, the reconstructed events are classified based on the following mutually exclusive definitions:

- **Clean** The event contains one matched vertex corresponding to the hard-scatter interaction. The hard-scatter interaction does not contribute more than 50% of the accumulated track weight to any other vertex.
- **Low pile-up contamination** The event contains one and only one merged vertex where the hard-scatter interaction contributes more than 50% of the accumulated track weight.
- **High pile-up contamination** The event does not contain any vertex where the hard-scatter interaction contributes more than 50% of the accumulated track weight. It does however contain at least one merged vertex in which the hard-scatter interaction contributes between 1 and 50% of the accumulated track weight.
- **Split** The event contains at least two merged vertices in which the hard-scatter interaction contributes more than 50% of the accumulated track weight.
- **Inefficient** The event does not contain any vertex where the hard-scatter interaction contributes more than 1% of the accumulated track weight.

In the current analysis, all categories except “Inefficient” are considered as successful in reconstructing the hard-scatter primary vertex. All of these categories thus contribute to the calculation of total vertex reconstruction efficiency.

5.2 Vertex reconstruction and selection efficiency for hard-scatter interactions

The efficiency to reconstruct and also to correctly identify the hard-scatter primary vertex is used to quantify the impact of pile-up contamination. Assuming that the hard-scatter primary vertex produces reconstructed tracks, the efficiency of hard-scatter primary vertex reconstruction is predicted to be larger than 99%. This includes interactions with low or high pile-up contamination, and split event categories as defined in Sect. 5.1. The corresponding contributions to the reconstruction efficiencies as a function of simulated $\mu$ are shown in Fig. 8 for the processes $Z \rightarrow \mu\mu$, $H \rightarrow \gamma\gamma$ and $t\bar{t} \rightarrow l+X$ ($t\bar{t}$ decays that include a lepton).

The fraction of events with low and high pile-up contamination increases with growing $\mu$, while the fraction of clean events decreases with $\mu$. The fraction of events containing split vertices remains negligible for all $\mu$. For $\mu = 38$ the fraction of high pile-up contamination vertices is 8% for $Z \rightarrow \mu\mu$ events, 5% for $H \rightarrow \gamma\gamma$ events, and 2% for $t\bar{t}$ events.

The effect of pile-up contamination on the reconstruction efficiency for the hard-scatter primary vertex clearly depends on the nature of the physics process under study. The hard-scatter interactions corresponding to $Z$-boson production leave on average fewer charged particles within the detector acceptance than those corresponding to $t\bar{t}$ production. Hard-scatter vertices from $Z$-boson production can therefore be expected to be more affected by pile-up contamination than those from $t\bar{t}$ events. Indeed, Fig. 8 shows that the low and high pile-up contamination fractions are always higher for $Z \rightarrow \mu\mu$ than for $t\bar{t}$ events.

Pile-up tracks contaminating reconstructed hard-scatter vertices lead to a degradation of position resolution. Figure 9 shows the distribution of residuals of the primary vertex position in a $Z \rightarrow \mu\mu$ sample for different classes.

The residuals are calculated as the distance between the position of the hard-scatter primary vertex at generator level and its reconstructed position obtained from the primary vertex reconstruction as described in Sect. 4.2. Only the vertices matched according to the definition presented in Sect. 5.1 are taken into account. The results are obtained using the MC simulation including detector acceptance without further selection criteria. The categories of clean reconstruction, low and high pile-up contamination show progressively degrading resolution. This effect is visibly largest for the $z$-coordinate, because the transverse coordinates are constrained by the beam-spot width. The events categorised as...
containing split vertices do not suffer from a degraded resolution compared to the clean event category.

In addition to the degradation of the spatial resolution, the presence of significant pile-up makes it more difficult to correctly identify the hard-scatter primary vertex among the many pile-up vertices reconstructed in most bunch crossings. For most hard-scatter physics processes, it is effective to identify the hard-scatter primary vertex as the primary vertex with the highest sum of the squared transverse momenta of contributing tracks: \( \sum p_T^2 \). This criterion is based on the assumption that the charged particles produced in hard-scatter interactions have on average a harder transverse momentum spectrum than those produced in pile-up collisions. The efficiency of the hard-scatter identification using this criterion depends on the kinematics of the hard-scatter process. Distributions of \( \sqrt{\sum p_T^2} \) of the tracks in various hard-scatter processes are shown in Fig. 10, including \( H \rightarrow \gamma\gamma \), \( Z \rightarrow \mu\mu \), and \( t\bar{t} \) decays in which a filter has been applied to select decays with leptons. These are compared to a minimum-bias sample, which can be taken to have the same \( \sqrt{\sum p_T^2} \) distribution as pile-up.

---

**Fig. 8** Contributions to the predicted primary vertex reconstruction efficiency as a function of the average number of interactions per bunch crossing, \( \mu \). The mutually exclusive categories of events are defined in Sect. 5.1. The black circles show the contribution to the efficiency from events categorised as clean, and the blue and red circles show the contributions from events with low and high pile-up contamination respectively. The open crosses show the sum of the contributions from events that are clean and those with low pile-up contamination; the filled crosses show the sum of the contributions from all categories and represent the overall efficiency. The hard-scatter processes considered are Higgs-boson decay into \( \gamma\gamma \), \( t\bar{t} \) production with a lepton in the decay, and \( Z \)-boson decay into \( \mu\mu \).
Fig. 9 The residual distributions in a $x$ and b $z$ coordinates for reconstructed primary vertices in a sample of simulated $Z \rightarrow \mu\mu$ events for the four classes of events defined in Sect. 5.1. The distributions are normalised to the same area. The RMS values of these residuals are provided for each class.

Fig. 10 The distributions of the sum of the squared transverse momentum for tracks from primary vertices, shown for simulated hard-scatter processes and a minimum-bias sample. In the case of the $Z \rightarrow \mu\mu$ process, only events with at least two muons with $p_T > 15$ GeV reconstructed within the ATLAS inner detector acceptance are shown. The $t\bar{t}$ process is filtered to select decays with leptons. The distributions are normalised to the same area.

In the case of $Z \rightarrow \mu\mu$ and $t\bar{t}$, there is significant transverse momentum carried by charged particles even in the case of inclusive samples. In contrast, in the case of $H \rightarrow \gamma\gamma$ events, most of the transverse momentum is carried by the photons from the Higgs boson decay. The remaining charged particles in the acceptance of the detector are produced in the underlying event and have a much softer $p_T$ spectrum. The efficiency to correctly select the hard-scatter vertex among many pile-up vertices by choosing the vertex with the highest $\sum p_T^2$ is thus inferior for $H \rightarrow \gamma\gamma$ decays compared to most other hard-scatter processes. A more efficient method for choosing the primary vertex in the case of $H \rightarrow \gamma\gamma$ decay is described in Ref. [27].

For hard-scatter processes, the primary vertex selection efficiency is defined as the fraction of events in which the highest $\sum p_T^2$ vertex is the vertex associated with the MC simulation hard scatter. The MC hard scatter is taken as the vertex with the highest weight of hard-scatter tracks, as described in Sect. 5.1. The efficiency to reconstruct and then select the hard-scatter primary vertex is shown as a function of $\mu$ in Fig. 11a for different physics processes. The highest efficiency is achieved for $t\bar{t}$ events for all values of $\mu$. This observation is attributed to the high multiplicity of high transverse momentum tracks produced in top-quark decays. The selection efficiency for $Z \rightarrow \mu\mu$ events is greatly improved when additional criteria reflecting the kinematics of the physics process are imposed. Figure 11b shows the selection efficiencies after requiring at least two muons with $p_T > 15$ GeV to be reconstructed within the ATLAS inner detector acceptance. The $t\bar{t}$ sample shows a selection efficiency above 99% with or without the muon acceptance requirement (the points are overlapping in the figure). A clear selection efficiency improvement for the $Z \rightarrow \mu\mu$ process is visible when muons are reconstructed in the acceptance, resulting in at most 2% of events with a wrongly selected hard-scattering primary vertex for $\mu$ of 38. These losses are primarily due to the small but non-zero probability that the $\sum p_T^2$ of tracks from one of the inelastic interactions in the minimum-bias sample is larger than in the $Z \rightarrow \mu\mu$ interaction, as illustrated in Fig. 10. A more quantitative prediction of this loss is given in Sect. 8.

6 Primary vertices in minimum-bias data

This section presents a study of single primary vertex reconstruction in soft interactions which are characteristic of the
pile-up events superimposed on the hard-scatter event of interest. This study is based on a minimum-bias data sample with a single primary vertex reconstructed in each event and corresponding to an average number of interactions per bunch crossing $\mu = 0.01$. These data are compared to a simulation of inelastic interactions using the PYTHIA8 event generator.

The reconstruction efficiency for primary vertices produced in soft $pp$ interactions varies depending on the nature of the soft interaction process. If the majority of final-state charged particles are produced outside the detector acceptance, the reconstruction of the corresponding primary vertex may be unsuccessful. The vertex reconstruction efficiency may be further reduced by the inefficient reconstruction of very low $p_T$ trajectories, characteristic of these soft interactions. Table 4 shows the efficiencies for reconstructing the primary vertex in events from a minimum-bias sample with only single interactions. These efficiencies are obtained from PYTHIA8 MC simulation separately for the three processes which produce minimum-bias triggers in the experiment, namely non-diffractive, single-diffractive, and double-diffractive interactions. Without selection cuts the reconstruction efficiency depends strongly on the process: increasing from 46% for single-diffractive to 93% for non-diffractive interactions. Taking into account the relative contributions of each process to inelastic interactions, the average efficiency is estimated to be about 80%. The difference in the efficiencies estimated for the different processes is primarily due to the different distributions of transverse momenta and pseudorapidities of charged particles produced in each process. In diffractive processes, the charged particles are mostly produced at large pseudorapidities, often outside the acceptance of the ATLAS tracking system. The very soft transverse momentum spectrum of these charged particles is an additional complication in their reconstruction. As shown in the second row of Table 4, basic geometrical and kinematic requirements on the generated particles remove most of the differences in efficiency among the non-diffractive, single- and double-diffractive processes. The overall vertex reconstruction efficiency increases to 95% in this case. The

![Graph](image1)

**Fig. 11** Efficiency to reconstruct and then select the hard-scatter primary vertex as a function of the average number of $pp$ interactions per bunch crossing, $\mu$, for different physics processes: a all reconstructed events; b events with at least two muons with $p_T > 15$ GeV reconstructed within the ATLAS inner detector acceptance. The points showing the $t\bar{t}$ efficiency with and without acceptance criteria overlap

<table>
<thead>
<tr>
<th></th>
<th>Non-diffractive (%)</th>
<th>Single-diffractive (%)</th>
<th>Double-diffractive (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency without any selection cuts</td>
<td>92.9</td>
<td>45.7</td>
<td>49.0</td>
</tr>
<tr>
<td>Efficiency requiring at least two charged particles with $p_T &gt; 400$ MeV and $</td>
<td>\eta</td>
<td>&lt; 2.5$</td>
<td>96.1</td>
</tr>
<tr>
<td>Efficiency requiring at least two charged particles reconstructed in the inner detector</td>
<td>99.6</td>
<td>99.5</td>
<td>99.3</td>
</tr>
</tbody>
</table>

Table 4 Vertex reconstruction efficiencies, at various selection levels, for non-diffractive, single-diffractive, and double-diffractive interactions in PYTHIA8 minimum-bias simulation
Fig. 12 Distributions of a number of tracks per vertex, b track transverse momentum $p_T$, c track pseudorapidity $\eta$ and d $\sqrt{\Sigma p_T^2}$ of the tracks associated with each vertex. Distributions are shown for tracks associated with primary vertices in low $\mu$ minimum-bias data and in simulation samples.

remaining differences in efficiencies are mostly due to the dependence of the track reconstruction efficiency on $\eta$ and $p_T$. The third row of Table 4 shows that the primary vertex reconstruction efficiency further increases to about 99% for all processes after requiring that at least two tracks are reconstructed within the inner detector, in addition to the requirements listed in the second row. The intrinsic efficiency of the ATLAS vertex reconstruction algorithm is thus expected to be very high if at least two charged particles are produced within the inner detector acceptance.

Figure 12 compares the simulation to data for the distributions of the number of fitted tracks, the track $p_T$, track $\eta$, and $\sqrt{\Sigma p_T^2}$ of tracks in primary vertices. The figure illustrates how soft the pile-up interactions are: only 0.4% of the tracks belonging to a reconstructed primary vertex have $p_T > 4$ GeV and only 1.2% of the reconstructed vertices have a total $\sqrt{\Sigma p_T^2}$ above 10 GeV. There are small discrepancies between simulation and data at very high values in the track $p_T$ spectrum and at high $\eta$. As described in Refs. [4,10], these are due to deficiencies in the physics modelling of these distributions and not related to the primary vertex reconstruction algorithm. The dominant sources of systematic uncertainties relevant to the comparisons in Fig. 12 are the knowledge of the beam-spot size, the modelling of fake tracks, and the dependence of the track reconstruction efficiency on $p_T$, $\eta$ and $\mu$. These sources are not included in the error bars of the corresponding plots, but contribute to the observed discrepancies between data and simulation.

The position resolution of single vertices is estimated either from MC simulation or from data using the split-vertex
Fig. 13 Resolution of the primary vertex position in a $x$ and $b$ $z$ as function of the number of fitted tracks, estimated using the split-vertex method (SVM) for minimum-bias data (black circles) and MC simulation (blue squares). Also shown is the resolution obtained from the difference between the generator-level information and reconstructed primary vertex position in MC simulation (labeled “truth”), with and without the beam-spot constraint (pink and red triangles respectively). The bottom panel in each plot shows the ratio of the resolution found using the split-vertex method in data to that obtained using the MC generator-level information without the beam-spot constraint.

The split-vertex method is also used to calculate the resolution for the minimum-bias simulation sample. There is good agreement between the data and simulation distributions, showing that the reconstructed track parameters used in the vertex reconstruction are well modelled in the simulation. Figure 13 also shows the primary vertex resolution calculated as the difference between the true and reconstructed vertex position in the MC simulation. The good agreement between the split-vertex method and the resolution calculated with the MC generator-level information gives confidence that the split-vertex method provides a reliable measurement of the primary vertex resolution. At very low track multiplicity the result of the split-vertex method deviates slightly from the resolution obtained using the generator-level information. Here the resolution obtained from the generator-level information benefits from the perfect knowledge of vertex position decreasing the resolution spread, compared to the resolution obtained from the two reconstructed vertices in the split-vertex method. When the beam-spot constraint is included the resolution improves considerably in the transverse direction, staying below 20 $\mu$m for the full range of $\mu$ studied. The longitudinal resolution reaches 30 $\mu$m at high track multiplicity. Figure 13 also shows the resolution calculated using MC generator-level information with and without beam-spot constraint.

7 Performance in the high pile-up regime

In this section, the study of the primary vertex reconstruction performance at low $\mu$ is extended to the high pile-up regime. A dedicated data sample of minimum-bias events collected with values of $\mu$ between 55 and 72 was used to study the performance of the primary vertex reconstruction in the presence of multiple vertices. The simulation samples spanned values of $\mu$ from 0 to 22, typical of the standard 2012 data-taking conditions, and from 38 to 72 to emulate the high $\mu$ data sample.

The efficiency of primary vertex reconstruction decreases with increasing pile-up. In addition to the inefficiencies affecting single vertex reconstruction described in Sect. 6, effects related to the merging of adjacent primary vertices start to play a significant role as pile-up increases. Figure 14a shows the average number of vertices lost due to merging and to other effects, such as track reconstruction and detector acceptance.
Merging has a small effect on overall vertex reconstruction efficiency for $\mu$ values below 20, but it is a dominant effect for $\mu$ values above 40. Figure 14b shows the average number of expected reconstructed primary vertices as a function of $\mu$, for the two main classes of vertices defined in Sect. 5, matched vertices, consisting of tracks mostly coming from a single interaction, and merged vertices. For the highest values of $\mu$ around 70, where one expects about 60 primary vertices with at least two charged particles with $p_T > 400$ MeV within the detector acceptance, a total of 30 primary vertices are expected to be reconstructed on average, out of which about 10 are merged vertices. About 20 additional primary vertices are lost due to merging and about 10 due to other inefficiencies as shown in Fig. 14a. Vertices classified as “Fake” or “Split”, according to the definitions presented in Sect. 5.1, are not shown in Fig. 14b, since they represent a very small contribution of at most 2% of the total number of reconstructed vertices at $\mu = 70$.

The main observables relevant to the primary vertex reconstruction performance are in reasonable agreement between data and simulation with only small discrepancies attributed to the physics modelling of soft interactions (see Fig. 12). To quantify the agreement between data and simulation at high values of $\mu$, the same observables are studied and the ratios of data to simulation are compared between low and high values of $\mu$. This is shown in Fig. 15 for the track $p_T$, the number of tracks per primary vertex, and the $\sqrt{\Sigma p_T^2}$ per primary vertex. The data to simulation ratios are overlaid for low and high $\mu$ samples in the upper panels. The lower panels show the double ratios of data to simulation between high and low values of $\mu$.

The double ratios agree with unity, showing that there is similar agreement between data and simulation at low and high $\mu$. In the case of track multiplicity, the agreement between data and simulation for high track multiplicities is somewhat better at high $\mu$ than at low $\mu$. This arises possibly because discrepancies in physics modelling are diluted by the contributions from merged vertices as $\mu$ increases.

8 Efficiency of vertex reconstruction as a function of pile-up

An analytical model to predict the number of reconstructed vertices as a function of event multiplicity has been developed. This model is based on the measured primary vertex reconstruction efficiency and on the the probability of vertex merging.

8.1 Modelling the number of reconstructed vertices

In the ideal case of perfect reconstruction efficiency, the number of reconstructed vertices would scale linearly with $\mu$. In reality there are a number of effects that cause the relation to be non-linear. As discussed in Sect. 7, one of the most important effects is vertex merging, when two or more vertices are merged and reconstructed as one vertex. Other effects include reconstruction inefficiencies, detector acceptance, and, at a small level for low track multiplicities, non-collision background. As already mentioned, the impact of fake and split vertices is negligible.
Fig. 15 Ratios of data to MC simulation for observables relevant to the primary vertex reconstruction performance: a) track transverse momentum \( p_T \), b) number of tracks per vertex, \( \sqrt{\sum p_T^2} \) of the tracks in each vertex. Error bars represent only statistical uncertainties. The ratios are shown for low (0–1) and high (55–72) values of \( \mu \). The bottom panel in each figure shows the double ratio of high to low \( \mu \).

The average number of reconstructed vertices, \( \langle n_{\text{vertices}} \rangle \), can be parameterised as a function of \( \mu \) as follows:

\[
\langle n_{\text{vertices}} \rangle = p_0 + \epsilon \mu - F(\epsilon \mu, p_{\text{merge}}),
\]

(6)

where \( \epsilon \) is the efficiency of the vertex reconstruction algorithm before including vertex merging effects, and \( p_0 \) accounts for any small offset arising from non-collision background. Based on the results shown in Sects. 5, 6, and 7, the value of \( \epsilon \) is considered to be independent of \( \mu \). The quantity \( \epsilon \mu \) represents the average number of vertices that would be reconstructed in the absence of any pile-up induced vertex merging effects. This quantity is referred to, hereafter, as the number of reconstructible vertices. In this study the parameter \( \epsilon \) is obtained from a fit to the MC simulation. The function \( F(\epsilon \mu, p_{\text{merge}}) \) represents the average number of vertices lost due to merging effects, taking into account the number of reconstructible vertices and the vertex merging probability, \( p_{\text{merge}} \). These effects are primarily responsible for the non-linear dependence of the number of reconstructed vertices as a function of \( \mu \). The evaluation of this function is described in the next section.

The proposed model only describes the primary vertex reconstruction and does not account for pile-up effects in the reconstruction of tracks. The model assumes that the track reconstruction efficiency and the corresponding fake rate are constant for the studied range of pile-up values.

8.2 Determination of correction for merging of primary vertices

The effects of vertex merging are studied using the longitudinal separation, \( \Delta z \), between pairs of adjacent reconstructed primary vertices. The distribution of \( \Delta z \) in a typical Run 1 minimum-bias data sample is shown in Fig. 16 together with the prediction from simulation.

At low values of \( \Delta z \) close-by vertices can no longer be separated and are reconstructed as a single vertex. In Fig. 16, this effect is visible as a steep decrease of the number of reconstructed vertices at values of \( \Delta z \) below a few mm. The small peak around \( \Delta z = 0 \) is due to the effect of splitting of primary vertices: in this case, close-by vertices are reconstructed with longitudinal separations well below the typical primary vertex resolution. The distribution of \( \Delta z \) measured
in a low pile-up data sample ($\mu$ below 10) is used to derive a two-vertex merging probability density function $p_{\text{merge}}(\Delta z)$. This function can then be combined with a given beam-spot shape to derive an analytical relationship between the number of reconstructible vertices per event, $\epsilon \mu$, and the average number of reconstructed vertices, $\langle n_{\text{vertices}} \rangle$. Using this approach, the effect of different beam-spot sizes on the merging probability can then also be evaluated.

The analytical function is derived as follows:

1. The $\Delta z$ distribution for pairs of adjacent vertices reconstructed in low pile-up data is fitted with a Gaussian function in a range where the merging of vertices is negligible: $|\Delta z| > 30$ mm. The Gaussian has an expected width of $\sqrt{2} \sigma_{\Delta z}$, where $\sigma_{\Delta z}$ is the longitudinal beam-spot RMS, assuming the beam spot has a Gaussian shape distribution along the $z$-axis.

2. A merging probability density function, $p_{\text{merge}}(\Delta z)$, is constructed by taking the difference between the distribution of $\Delta z$ observed in data in the range $|\Delta z| < 30$ mm and the prediction obtained from the Gaussian fit, $f_{\text{exp}}(\Delta z)$. This difference is then normalised to the prediction probability density function:

$$p_{\text{merge}}(\Delta z) = \frac{f_{\text{exp}}(\Delta z) - f_{\text{obs}}(\Delta z)}{f_{\text{exp}}(\Delta z)}. \quad (7)$$

Here, $f_{\text{obs}}(\Delta z)$ represents the observed probability density function of $\Delta z$ in the range $|\Delta z| < 30$ mm. An example of the observed distribution $f_{\text{obs}}(\Delta z)$ is shown in Fig. 16.

The $p_{\text{merge}}(\Delta z)$ PDF is parameterised using a step function convolved with a Gaussian function with parameters fit to the observed distribution. The $p_{\text{merge}}(\Delta z)$ PDF is derived in the low pile-up regime, where only the merging of adjacent pairs of vertices is assumed to be significant. The possible effects of merging more than two $pp$ collisions into a single reconstructed primary vertex are assumed to be negligible in this low pile-up regime.

3. The total merging probability $p_{\text{merge}}$ for two independent reconstructible vertices is computed from the product of the merging PDF and the expected $f_{\text{exp}}(\Delta z)$ distribution:

$$p_{\text{merge}} = \int f_{\text{exp}}(\Delta z)p_{\text{merge}}(\Delta z)d(\Delta z). \quad (8)$$

It is assumed that the merging PDF for a pair of adjacent vertices $p_{\text{merge}}(\Delta z)$ is independent of the beam conditions. The overall probability of merging two random reconstructible vertices depends on the particular beam-spot distribution, and therefore on $f_{\text{exp}}(\Delta z)$.

4. The total number of vertices lost due to merging effects is given by:

$$F(\epsilon \mu, p_{\text{merge}}) = \epsilon \mu - \sum_{N_{\text{vertices}}} P(N_{\text{vertices}}, \mu)\phi_{\text{merge}}(N_{\text{vertices}}, p_{\text{merge}}), \quad (9)$$

where $P(N_{\text{vertices}}, \epsilon \mu)$ is a PDF, representing the probability of reconstructing $N_{\text{vertices}}$ vertices given $\epsilon \mu$ potentially reconstructible vertices. Since the number of visible $pp$ collisions varies according to Poisson with the mean of $\mu$, this function $P(N_{\text{vertices}}, \epsilon \mu)$ is a Poisson with a mean $\epsilon \mu$.

The function $\phi_{\text{merge}}(N_{\text{vertices}}, p_{\text{merge}})$ represents the number of reconstructed vertices after taking into account merging effects, for a number, $N_{\text{vertices}}$, of vertices which would be reconstructed in the absence of any merging. This number is defined as follows:

$$\phi_{\text{merge}}(N_{\text{vertices}}, p_{\text{merge}}) = \sum_{i=1}^{N_{\text{vertices}}} p_i, \quad (10)$$

where $p_i = p_{i-1}(1 - p_{i-1}p_{\text{merge}})$, $i \geq 2$ and $p_1 = 1$. The $p_i$ represents the probability to reconstruct $i$ vertices in the presence of merging effects.

8.3 Comparison of data to simulation

To quantitatively compare data with simulation, additional effects and systematic uncertainties need to be taken into account. To account for the difference in visible cross section between data and simulation discussed in Sect. 3, the parameter $\epsilon$, extracted from the simulation fit, is scaled by a factor 1/1.11, which is equivalent to a scaling of $\mu$. A 6\% uncertainty is assigned to this procedure, where the dominant contribution comes from the uncertainty in the measured value of $\mu$.

The impact of possible discrepancies in longitudinal beam-spot size between data and MC simulation was also assessed since the observed data values represent an average over a range of different and non-uniform experimental values. The MC simulation samples used in this study were generated with a beam-spot size equal to the average observed in data. The effect of a change in beam-spot size on the merging probability can be evaluated with Eq. (8). A small additional uncertainty is assigned to account for the variations of up to $\pm 2$ mm in beam-spot size in data.

A fit using Eq. (6) was performed on MC simulation, allowing parameters $p_0, \epsilon$, and $p_{\text{merge}}$ to vary. The efficiency, $\epsilon$, and merging probability, $p_{\text{merge}}$, are extracted from the fit to simulation and found to be, $0.618 \pm 0.004(\text{stat.}) \pm 0.037(\text{syst.})$ and $0.0323 \pm 0.0002(\text{stat.}) \pm 0.0013(\text{syst.})$ respectively, after correcting $\epsilon$ with the $\mu$-rescaling factor and taking into account the systematic uncertainties, as described above. The fit to MC simulation is shown in Fig. 17a.
The overall agreement between the data and the prediction is within 3%, with the largest observed discrepancies well within the systematic uncertainty bands.

This comparison shows that the simulation describes the primary vertex reconstruction efficiency dependence on $\mu$ accurately. Vertex merging is the effect that has the largest impact on primary vertex reconstruction efficiency as $\mu$ increases. The analytical description proposed to describe this effect is validated by the measurements based on minimum-bias data. This confirms that the main factors related to the vertex reconstruction in pile-up conditions are correctly taken into account and that the remaining effects related to the presence of fake and split vertices are negligible, as expected.

The predicted average number of reconstructed vertices, as obtained from data for a given value of $\mu$ in Fig. 17b, can be used to estimate the primary vertex selection efficiency for a specific hard-scatter process. This is done by combining the prediction with the simulated distributions of track $\sqrt{\sum p_T^2}$ for this process and for minimum-bias events, as shown in Fig. 10. For the highest $\mu$ value ($\mu = 40$) studied in terms of hard-scatter primary vertex reconstruction and selection efficiencies in Sect. 5, Fig. 17b predicts an average number of reconstructed vertices from pile-up interactions of $17 \pm 1$. Of all the reconstructed vertices, the one with highest $\sum p_T^2$ is selected as the hard-scatter vertex with a very high efficiency for most processes. To estimate the small probability that a pile-up vertex is selected by this procedure instead, the simulated distribution of $\sqrt{\sum p_T^2}$ is compared to the much harder one expected for the hard-scatter process of interest. For $Z \rightarrow \mu\mu$ events, a randomly selected point on the $\sum p_T^2$ distribution is found to be lower than the largest of the values found for 17 random samplings of the distribution for minimum-bias events in approximately 4% of the cases. This estimate, which is partially based on data but does not account for all experimental effects such as the distortion of the track $\sum p_T^2$ distribution of minimum-bias events due to merging of primary vertices, is in reasonable agreement with the estimate of 2% obtained based on simulation in Fig. 11.

### 9 Conclusion

This paper presents primary vertex reconstruction and selection methods and their performance for proton–proton collision data recorded by the ATLAS experiment at the LHC during Run 1. The primary vertex position resolution measured in data is consistent with the predictions from simulation. A longitudinal vertex position resolution of about 30 $\mu$m has been achieved for events with high track-multiplicity. A significant improvement of the vertex transverse-position reso-
ution is obtained using the beam-spot constraint in the vertex fit, giving a resolution below 20 μm for all multiplicities.

The primary vertex reconstruction efficiency has been measured using MC simulation. For minimum-bias events, the single vertex reconstruction efficiency is above 99% for all processes, provided at least two charged particles are reconstructed within the ATLAS inner detector. For hard-scatter interactions, the reconstruction and selection efficiency has been studied for a number of benchmark processes as a function of pile-up. In all cases, the overall signal vertex reconstruction efficiency exceeds 99%. A significant contamination from pile-up minimum-bias vertices is however observed for high values of μ in the case of hard-scatter processes with a small number of charged-particle tracks, such as $H \rightarrow γγ$ and $Z \rightarrow μμ$. The efficiency to reconstruct and then correctly select the primary vertex at $μ = 40$ in the case of $Z \rightarrow μμ$ is predicted to remain very high, namely 98%, when both muons are reconstructed within the inner detector acceptance.

The impact of multiple $pp$ interactions in the same bunch crossing on the reconstruction of primary vertices has been studied in detail. Comparisons of the modelling of vertex input quantities were made for low and high values of μ and good agreement between data and the MC simulation is observed for values of μ up to 70. The largest impact of pile-up is the merging of nearby vertices, which has been quantified precisely by studying the relationship between μ and the number of reconstructed vertices. The corresponding non-linear effects due to merging are well modelled within the uncertainties in the MC simulation for values of μ as high as 70, confirming the validity of the proposed model.

Acknowledgements We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CPNPq and FAPESP, Brazil; NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DSNRC, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; SRNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNISW and NCN, Poland; FCT, Portugal; MINEJA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZS, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [28].

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP3.

References

7. TOTEM Collaboration, G. Anelli et al., The TOTEM Experiment at the CERN Large Hadron Collider. JINST 3, S08007 (2008)
az Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia
ba Also at LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
* Deceased