Alignment of the ATLAS Inner Detector in Run 2

ATLAS Collaboration
CERN, 1211 Geneva 23, Switzerland

Received: 15 July 2020 / Accepted: 28 October 2020 / Published online: 24 December 2020
© CERN for the benefit of the ATLAS collaboration 2020, corrected publication 2020

Abstract The performance of the ATLAS Inner Detector alignment has been studied using pp collision data at $\sqrt{s} = 13$ TeV collected by the ATLAS experiment during Run 2 (2015–2018) of the Large Hadron Collider (LHC). The goal of the detector alignment is to determine the detector geometry as accurately as possible and correct for time-dependent movements. The Inner Detector alignment is based on the minimization of track-hit residuals in a sequence of hierarchical levels, from global mechanical assembly structures to local sensors. Subsequent levels have increasing numbers of degrees of freedom; in total there are almost 750,000. The alignment determines detector geometry on both short and long timescales, where short timescales describe movements within an LHC fill. The performance and possible track parameter biases originating from systematic detector deformations are evaluated. Momentum biases are studied using resonances decaying to muons or to electrons. The residual sagitta bias and momentum scale bias after alignment are reduced to less than $0.1 \text{ TeV}^{-1}$ and $0.9 \times 10^{-3}$, respectively. Impact parameter biases are also evaluated using tracks within jets.

Contents

1 Introduction .................................. 1
2 The ATLAS detector .......................... 2
  2.1 Inner Detector structure ................. 2
  2.2 Local coordinate system ................. 3
3 Alignment principles and formalism .......... 4
  3.1 Track fitting with the Newton–Raphson method 4
  3.2 The Global $\chi^2$ method for alignment ...... 5
  3.3 Solving the linear system of alignment equations 7
4 Inner Detector alignment .................... 8
  4.1 Time-dependent alignment ............... 9
  4.2 Baseline alignment constants .......... 9
  4.3 Residuals ................................ 10
5 Detector stability and time-dependent alignment 10

1 Introduction

The precise reconstruction of the trajectories of charged particles created in proton–proton (pp) and heavy-ion collisions at CERN’s Large Hadron Collider (LHC) is a key ingredient in many of the physics processes studied by the ATLAS Collaboration. Almost every measurement performed using the ATLAS detector [1], from Standard Model processes to searches for new physics phenomena, relies on the accurate reconstruction of charged particles.

In order to reconstruct the trajectories of charged particles, ATLAS uses the Inner Detector (ID) tracking system to provide efficient, robust and precise position measurements of charged particles as they traverse the detector. The energy deposits from charged particles (hits) recorded in individual detector elements of the ID are used to reconstruct their trajectories (tracks) and estimate the associated track parameters. The precision achieved for the track parameters is determined by several factors: the intrinsic resolution of sensitive devices; the knowledge of the magnetic field; the distribution of material in and before the ID and the knowledge of it; and the knowledge of the geometry, i.e. the location and orientation, of the detector elements. The purpose of the detector alignment is to determine, as precisely as possible, the actual geometry of the active detector elements of the tracking system, and to follow changes in the geometry with time.
Poor knowledge of the actual geometry of the active detector elements results in a deterioration of the resolution of reconstructed track parameters. The criteria for the minimum precision required were defined in order to limit the degradation of the resolution of the track parameters for high-momentum tracks to less than 20% in comparison to a perfectly aligned detector [2]. In addition, correlated geometrical distortions can lead to systematic biases in the reconstructed track parameters. Correlated systematic biases can be introduced either by real detector deformations to which the alignment procedure has little sensitivity or by the procedure used to determine the alignment parameters. These correlated biases are referred to as ‘weak modes’ of the alignment.

In this document, the ATLAS ID alignment procedure and its performance during Run 2 of the LHC is presented. A new layer of pixel sensors was included in the detector for Run 2, which posed additional challenges for the alignment of the detector compared to those faced during Run 1 [3, 4]. The greatest new challenge was the short-timescale movement of parts of the detector during data taking.

This paper is organised as follows: a brief description of the ATLAS detector is given in Sect. 2. Section 3 presents the formalism of the ATLAS track-based ID alignment. Section 4 introduces the different alignment levels and Sect. 5 discusses the detector stability and describes the time-dependent alignment. The performance of the ATLAS Run 2 alignment is presented in terms of track parameter biases in Sects. 6 and 7. Concluding remarks are made in Sect. 8.

2 The ATLAS detector

The ATLAS detector [1] at the LHC is a multipurpose particle detector with a forward–backward symmetric cylindrical geometry that covers nearly the entire solid angle around the collision point. The global ATLAS reference frame is a right-handed Cartesian coordinate system, where the origin is at the nominal pp interaction point, corresponding to the centre of the detector. The positive x-axis points to the centre of the LHC ring, the positive y-axis points upwards and the z-axis points along the beam direction. Polar coordinates \( r, \phi \) are used in the transverse plane, \( \phi \) being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle \( \theta \) as \( \eta = -\ln \tan(\theta/2) \). Angular distance is measured in units of \( \Delta R \equiv \sqrt{\Delta \eta^2 + (\Delta \phi)^2} \).

ATLAS consists of the ID (described in Sect. 2.1), electromagnetic and hadronic calorimeters, a muon spectrometer and a magnet system. Lead/liquid-argon sampling calorimeters provide electromagnetic energy measurements with high granularity and a steel/scintillator-tile hadronic calorimeter covers the central pseudorapidity range of \( |\eta| < 1.7 \). The endcap and forward regions are instrumented with liquid-argon calorimeters for measurements of both electromagnetic and hadronic showers up to \( |\eta| = 4.9 \). The outer part of the detector consists of a muon spectrometer with high-precision tracking chambers for coverage up to \( |\eta| = 2.7 \), fast detectors for triggering over \( |\eta| < 2.4 \), and three large superconducting toroid magnets with eight coils each. The ATLAS detector has a two-level trigger system to select events for offline analysis [5].

2.1 Inner Detector structure

The ATLAS ID [2, 6] consists of three subdetectors utilising three technologies: silicon pixel detectors, silicon strip detectors and straw drift tubes, all surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field [7]. The ID is designed to reconstruct charged particles within a pseudorapidity range of \( |\eta| < 2.5 \) (see Fig. 1 for a schematic view of the ID barrel region and Table 1 for a list of the main detector characteristics). The material distribution inside the ID has been studied in data through use of hadronic interactions and photon conversion vertices [8, 9]. During the second LHC data-taking run (2015–2018) with \( pp \) collisions at a centre-of-mass energy \( \sqrt{s} = 13 \) TeV, the ID collected data with an efficiency greater than 99% [10].

The innermost part of the Inner Detector consists of a high-granularity silicon pixel detector and includes the insertable B-layer (IBL) [11, 12], a new tracking layer added for Run 2 which is closest to the beam line and designed to improve the precision and robustness of track reconstruction. The IBL consists of 280 silicon pixel modules arranged on 14 azimuthal carbon fibre staves surrounding the beam pipe at
Table 1 Summary of the main characteristics of the ID subdetectors. The intrinsic resolution of the IBL and the Pixel sensors are reported along \( r-\phi \) and \( z \), while for SCT and TRT only the resolution along \( r-\phi \) is given [1,11]. For SCT and TRT the element size refers to the spacing of the read-out strips and the diameter of the straw tube, respectively.

<table>
<thead>
<tr>
<th>Subdetector</th>
<th>Element size</th>
<th>Intrinsic resolution (( \mu \text{m} ))</th>
<th>Barrel layer radii [mm]</th>
<th>Disk layer ( l_z ) [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBL</td>
<td>50 ( \mu \text{m} \times 250 \mu \text{m} )</td>
<td>10 \times 60</td>
<td>33.25</td>
<td></td>
</tr>
<tr>
<td>Pixel</td>
<td>50 ( \mu \text{m} \times 400 \mu \text{m} )</td>
<td>10 \times 115</td>
<td>50.5, 88.5, 122.5</td>
<td>495, 580, 650</td>
</tr>
<tr>
<td>SCT</td>
<td>80 ( \mu \text{m} )</td>
<td>17</td>
<td>299, 371, 443, 514</td>
<td>From 839 to 2735</td>
</tr>
<tr>
<td>TRT</td>
<td>4 mm</td>
<td>130</td>
<td>From 554 to 1082</td>
<td>From 848 to 2710</td>
</tr>
</tbody>
</table>

Fig. 2 Schematic representation of the ATLAS global reference frame \((x, y, z)\) and the local reference frame of each component of the ID. The Pixel, IBL, and SCT modules are grouped in the ‘Silicon’ category. For each component, the local-\( x \) axis points along the most sensitive direction; the local-\( z \) axis points away of the ATLAS centre; and the local-\( y \) direction is chosen according to the right-handed frame. For TRT tubes, the local reference frame is determined by the orientation of the module they are mounted on. For visualisation purposes only, the local reference frame is referred to as \((x', y', z')\) in the drawing.

The local coordinate system of an individual sensor of the detector is a right-handed system frame with the origin placed in the geometrical centre of the sensor. The local coordinate system for each subsystem component is illustrated in Fig. 2. The convention used is the following: the local-\( x \) axis points along the most sensitive direction of the sensor. This corresponds to the shorter pitch side for Pixel and IBL modules, and perpendicular to the strip-orientation for the SCT. In the silicon detectors, the local-\( y \) axis is oriented along the long side of the sensor (i.e. longer pitch direction for the Pixels on the disks. The barrel SCT sensors have a uniform pitch strip of 80 \( \mu \text{m} \), while the endcap sensor strips run radially with a 161.5 \( \mu \text{rad} \) angular pitch.

The Transition Radiation Tracker (TRT) [19–21] is the outermost subdetector and extends track reconstruction radially outwards to a radius of 1082 mm. It is made of 350,848 gas-filled straw tubes of 4 mm diameter. The tubes are arranged in 96 barrel modules in 3 layers (32 modules per layer) and 40 disks in each endcap. The expected hit resolutions for each subdetector are summarised in Table 1.

2.2 Local coordinate system

The local coordinate system of an individual sensor of the detector is a right-handed system frame with the origin placed in the geometrical centre of the sensor. The local coordinate system for each subsystem component is illustrated in Fig. 2. The convention used is the following: the local-\( x \) axis points along the most sensitive direction of the sensor. This corresponds to the shorter pitch side for Pixel and IBL modules, and perpendicular to the strip-orientation for the SCT. In the silicon detectors, the local-\( y \) axis is oriented along the long side of the sensor (i.e. longer pitch direction for the Pixels on the disks. The barrel SCT sensors have a uniform pitch strip of 80 \( \mu \text{m} \), while the endcap sensor strips run radially with a 161.5 \( \mu \text{rad} \) angular pitch.
and IBL and the strip direction in the SCT), while the local-$z$ direction is orthogonal to the local $x$-$y$ plane. In the case of the TRT, the local-$y$ axis points along the wire: either in the same direction as the global $z$-axis (barrel) or radially outwards (endcaps). In the barrel, the local-$z$ axis points radially outwards (from the origin of the global frame to the straw centre). In the endcaps, the local-$z$ axis points outwards (parallel to the beam line). The local-$x$ axis is perpendicular to both the TRT wire and the radial direction.

Hits are reconstructed in the local reference frame. The TRT measures the radial distance of the primary ionisation from the wire as $\sqrt{x^2 + z^2}$, taking both $x$ and $z$ in the local frame.

3 Alignment principles and formalism

This section reviews the formalism for in situ alignment of the ATLAS ID using reconstructed tracks. The concept of Global $\chi^2$ alignment is introduced, followed by a discussion of ‘weak modes’ (Sects. 3.2.4, 6) and how they can be avoided by adding constraints on track parameters. The section closes with a detailed description of the alignment procedure and its implementation within the ATLAS software.

The approach used is based on the Newton–Raphson method and determines both the trajectory parameters and a set of alignment parameters, $\alpha$. In this context, $\alpha$ are chosen as the six degrees of freedom (DoF) of each alignable structure that uniquely define its position and orientation in space. These correspond to three translations ($T_x$, $T_y$, $T_z$) and three rotations ($R_x$, $R_y$, $R_z$). Translations are relative to the origin of the reference frame of each alignable structure and rotations are around the Cartesian axes.

3.1 Track fitting with the Newton–Raphson method

The Newton–Raphson method uses an iterative approach to find the best fit to a set of measurements of a track left in the detector by a charged particle traversing active detector elements. The quality of the fit is characterised by a track $\chi^2$, determined from the distances between the hits in the detector, which constitutes the track measurements, and the fitted track (residuals). The trajectory of a track in a magnetic field is parameterised by a set of five parameters. The chosen parameterisation in ATLAS is: $\tau = (d_0, z_0, \phi_0, \theta_0, q/p)$, where $d_0$ and $z_0$ are the transverse and longitudinal impact parameters and $\phi_0$ and $\theta_0$ the azimuthal and polar angles of the track, all defined at the point of closest approach to the $z$-axis of the reference frame $[22]$. The ratio $q/p$ is the inverse of the particle momentum $(p)$ multiplied by its charge $(q)$ (see Ref. [4] for more details).

The track $\chi^2$ is calculated from all measured track-hit residuals, $r_i = e_i(\tau) - m_i$, where $m_i$ is the position of the $i$th measurement, and $e_i$ is the position of the intersection of the fitted track with the surface on which the $i$th measurement is made. The determination of the intersection position ($e_i$) includes the measurement in question, which causes $r_i$ to be a biased residual. The track $\chi^2$ is defined, using vector notation, as

$$\chi^2 = r^T \Omega^{-1} r,$$

where $r$ is the vector of track residuals and $\Omega$ is the covariance matrix of the corresponding measurements. The parameters of a track’s trajectory, $\tau$, are those that minimise this $\chi^2$. The minimisation is done using the first and second derivatives of the $\chi^2$ with respect to $\tau$. Defining the derivative $G = dr/d\tau$, the condition for the minimisation of the $\chi^2$ is

$$\left( \frac{d\chi^2}{d\tau} \right)^T = 2 G^T \Omega^{-1} r = 0.$$  

In practical terms, the values of $\tau$ satisfying Eq. (2) are found using an iterative procedure by evaluating the first and second derivatives of $\chi^2$ with respect to the track parameters of the current iteration, $\tau_0$. If the derivative $G$ were constant, then

1 The local position and uncertainty of each measurement are provided by the corresponding subsystem after applying its own clustering and hit reconstruction techniques. The Pixel detector uses an artificial neural network, trained on simulation, to determine the position of a cluster and its uncertainty $[23]$. The SCT parameterises the position of the cluster and its uncertainty, using simulation, as a function of the number of strips in the cluster and the incident angle of the particle. For the TRT the drift radius and its uncertainty is calibrated using an iterative procedure in data and simulation $[21]$. 

 Springer
In general, the derivative $G$ depends on the track parameters themselves. Therefore, the procedure is repeated until a convergence criterion is met.

The track fit is further improved by taking into account the impact of material interactions on the trajectory of the particle. Energy loss is treated as point-like at the center of material layers and for hadrons and muons, which are used during the alignment, deterministic as the variance of the energy loss processes is small. Additional parameters, $\theta$, are added to account for the effects of multiple Coulomb scattering (MCS) of the particle with the detector components, as detailed in Appendix A. Consequently, the residuals now also depend on $\theta$ and the variance of the scattering angles, $\Theta$:

$$\chi^2_{\text{Track}} = r^T \Omega^{-1} r + \theta^T \Theta^{-1} \theta.$$  \hspace{1cm} (3)

Thus, $\chi^2_{\text{Track}}$ has to be minimised for $r$ and $\theta$ simultaneously. The derivatives of residuals with respect to track and scattering parameters are defined as $G \equiv \partial r / \partial r$ and $S \equiv \partial r / \partial \theta$, respectively. In the following, the Global $\chi^2$ method for alignment including MCS is described and the following simplified notation is adopted:

$$\rho \equiv \begin{pmatrix} r \\ \theta \end{pmatrix}, \quad V \equiv \begin{pmatrix} \Omega & 0 \\ 0 & \Theta \end{pmatrix}, \quad \pi \equiv \begin{pmatrix} \tau \\ \theta \end{pmatrix}, \quad \text{and} \quad H \equiv \begin{pmatrix} G & S \\ \frac{\partial \theta}{\partial \tau} = 0 \quad & \frac{\partial \theta}{\partial \theta} = 1 \end{pmatrix}.$$  \hspace{1cm} (4)

It should be noted that refinements are made to the track fit by performing multiple iterations of the fit during which: material effects are recalculated; measurements are recalibrated based on the particles incident angle; and outliers are removed.

3.2 The Global $\chi^2$ method for alignment

3.2.1 General definition of the Global $\chi^2$ method for alignment

The Global $\chi^2$ is a track-based alignment method which uses a $\chi^2$ built from a large sample of reconstructed tracks and their associated hits in the detector elements being aligned. The alignment parameters are determined by minimising the Global $\chi^2$ with respect to the alignment parameters:

$$\chi^2_{\text{Global}} = \sum_i \chi^2_{\text{Track}i},$$  \hspace{1cm} (5)

where $\chi^2_{\text{Track}i}$ is the $\chi^2$ of the $i$th track as given by Eq. (3). The residuals used in Eq. (5) depend on the alignment parameters ($\alpha$) as both, the measurements and the track extrapolations depend on $\alpha$, the former directly and the latter through the fitted track parameters. Therefore, the minimisation of $\chi^2_{\text{Global}}$ with respect to $\alpha$ uses the total derivative operator with respect to $\alpha$, which can be expressed as:

$$\frac{d}{d\alpha_i} = \frac{\partial}{\partial \alpha_i} + \sum_j \frac{d\pi_j}{d\alpha_i} \frac{\partial}{\partial \pi_j}. \hspace{1cm} (6)$$

The $d\pi / d\alpha$ term is determined from the condition that, once $\chi^2_{\text{Global}}$ is at a minimum, $\chi^2_{\text{Track}}$ is also at a minimum with respect to the track parameters:

$$\frac{d}{d\alpha} \frac{\partial \chi^2_{\text{Track}}}{\partial \pi} = 0.$$  \hspace{1cm} (7)

Using Eq. (7) in Eq. (6), this results in:

$$\frac{d\pi}{d\alpha} = - \left( \frac{\partial^2 \chi^2_{\text{Track}}}{\partial \pi^2} \right)^{-1} \frac{\partial^2 \chi^2_{\text{Track}}}{\partial \alpha \partial \pi},$$

which allows the nested dependence of the $\pi$ on $\alpha$ to be resolved, thereby removing the need to determine both (the track parameters and alignment parameters) simultaneously.

Ignoring second-order derivatives in the residuals, using the covariance matrix of the track parameters, $C$, expressed as

$$C = 2 \left( \frac{\partial^2 \chi^2_{\text{Track}}}{\partial \pi^2} \right)^{-1} = (H^T V^{-1} H)^{-1},$$  \hspace{1cm} (8)

and defining $A$ as the derivative of the residuals with respect to the alignment parameters:

$$A \equiv \frac{\partial \rho}{\partial \alpha},$$  \hspace{1cm} (9)

the total derivative operator with respect to $\alpha$ can be written as:

$$\left( \frac{d}{d\alpha} \right)^T = \left( \frac{\partial}{\partial \alpha} \right)^T - A^T V^{-1} H C \left( \frac{\partial}{\partial \pi} \right)^T.$$  \hspace{1cm} (10)

The first- and second-order derivatives of $\chi^2_{\text{Global}}$ with respect to $\alpha$ are thus:

$$Y \equiv \left( \frac{d\chi^2_{\text{Global}}}{d\alpha} \right)^T = 2 \sum_{\text{tracks}} A^T V^{-1} (V - HCH^T) V^{-1} \rho,$$  \hspace{1cm} (10)

$$M \equiv \left( \frac{d^2\chi^2_{\text{Global}}}{d\alpha^2} \right) = 2 \sum_{\text{tracks}} A^T V^{-1} (V - HCH^T) V^{-1} A.$$  \hspace{1cm} (11)
Here, the term $H C H^\top$ represents the covariance of the track parameters in the measurement space, whereas the covariance of the residuals of the track fit is given by

$$R = V - H C H^\top.$$  

### 3.2.2 Newton–Raphson method for Global $\chi^2$ alignment

In analogy to the general method for track fitting (Sect. 3.1) an iterative approach is used to solve for the alignment parameters. The first- and second-order derivatives are obtained using Eqs. (10) and (11) and evaluated for an initial set of alignment parameters, $\alpha_0$. Such an initial geometry description is available from design drawings, survey measurements, or previous alignment results. The alignment corrections, to the initial geometry, are given by

$$X \equiv \Delta \alpha = - \left( \frac{d^2 \chi^2_{\text{Global}}}{d \alpha^2} \right)^{-1} \left( \frac{d \chi^2_{\text{Global}}}{d \alpha} \right)^\top \bigg|_{\alpha_0}$$  

$$\equiv - M^{-1} Y.$$  

(12)

The above step is repeated for successive iterations until a convergence criterion is met and $\Delta \alpha$ is negligible. This requires re-fitting the tracks using the updated geometry (initial alignment constants $\alpha_0$ plus their corrections $\Delta \alpha$), to obtain new residuals and new derivatives, and solving again to compute the next set of corrections to the alignment constants.

### 3.2.3 Locality ansatz

If the initial track parameters, $\pi_0$, minimise $\chi^2_{\text{Global}}$ for a given $\alpha_0$, Eq. (10) simplifies to

$$\left( \frac{d \chi^2_{\text{Global}}}{d \alpha} \right)^\top \bigg|_{\pi_0, \alpha_0} = 2 \sum_{\text{tracks}} A^\top V^{-1} \rho,$$  

(13)

as the term $H^\top V^{-1} \rho$ is zero. Consequently, if the measurements are independent and $V$ is diagonal, the derivative with respect to a particular parameter $\alpha_i$ only receives contributions from residuals for which the related entries in the derivative matrix $A$ are non-zero. In other words, if $\alpha_i$ is an alignment parameter of a given detector module, only the measurements in this module contribute to the first-order derivative of $\chi^2_{\text{Global}}$ with respect to $\alpha_i$. Therefore, contributions to the $\chi^2_{\text{Global}}$ from measurements in other subdetectors and MCS effects can be ignored. This useful property is labelled as the so-called locality ansatz [24] and provides an important simplification for the software implementation.

### 3.2.4 Adding constraints on track parameters

It is of particular importance to assure that the determination of the track parameters is free from systematic biases that can occur due to poorly determined ‘weak modes’ of the alignment. These modes are geometry distortions that leave the $\chi^2$ of the fitted tracks nearly unchanged and typically lead to an incorrect solution of the alignment. They can be controlled by imposing constraints on track parameters [25]. Examples of such constraints, discussed in detail in Sects. 6 and 7, are the beam-spot constraint, track parameter constraints from external detector systems (e.g. calorimeters), and contraints determined using reconstructed physics events (e.g. mass constraints from narrow resonances). These constraints are included in the Global $\chi^2$ method by adding extra terms to the expression for the $\chi^2$ in Eq. (5). For one track the modified contribution to $\chi^2$ is

$$\chi^2_{\text{cons}} = \rho^\top V^{-1} \rho + (\pi - q)^\top T^{-1} (\pi - q),$$  

(14)

where $q$ is a vector defining the constraint on $\pi$ and $T$ is its covariance matrix.

In the ATLAS implementation, this constraint is implemented by adding a pseudo-measurement on a track [22]. The solution for the alignment parameters is given by Eq. (12), where for each constrained track the covariance matrix is now defined as

$$C = 2 \left( \frac{d^2 \chi^2}{d \pi^2_{\text{cons}}} \right)^{-1} = (H^\top V^{-1} H + T^{-1})^{-1}.$$  

In this context, the first-order derivative of the Global $\chi^2$ is given by

$$\left( \frac{d \chi^2}{d \alpha} \right)^\top \bigg|_{\alpha_0} = 2 \sum_{\text{tracks}} A^\top V^{-1} (V - H C H^\top) V^{-1} \rho(\alpha_0)$$

$$- A^\top V^{-1} H C T^{-1} (\pi(\alpha_0) - q).$$  

(15)

If the tracks have been re-fitted with the imposed constraint, the locality ansatz drastically simplifies Eq. (15), reducing it to Eq. (13). This property is used in the ATLAS implementation.

### 3.2.5 Constraints on alignment parameters

Often one has some prior knowledge of the geometry from either survey measurements or mechanical constraints. These constraints can be included by adding terms to the $\chi^2$ in

---

2 This typically refers to a correction threshold of $< 0.1 \mu m$. However, it should be noted that the convergence criteria can vary between specific alignment level and degrees of freedom chosen.
In the general case, one can write

\[ \chi^2_{\text{cons}} = \sum_{\text{tracks}} V^T V^{-1} \rho + (\alpha - \alpha_{\text{cons}})^T W^{-1} (\alpha - \alpha_{\text{cons}}), \] (16)

where \( \alpha_{\text{cons}} \) is a vector defining the constraint on \( \alpha \) and \( W \) is its covariance matrix. The added term leads to extended expressions for the first and second derivatives of \( \chi^2 \) with respect to \( \alpha \) (Eqs. (10), (11)):

\[ Y \rightarrow Y + 2 W^{-1} (\alpha - \alpha_{\text{cons}}), \]
\[ M \rightarrow M + 2 W^{-1}, \] (17)

while the solution is computed using \( Y \) and \( M \) in Eq. (12).

The special case when \( a = \alpha_0 \) and \( W \) is diagonal, i.e. when the alignment parameters are constrained to their initial values, is further discussed in Sect. 3.3.3.

3.2.6 The Local \( \chi^2 \) method

The main advantage of the Global \( \chi^2 \) method arises from its rigorous treatment of correlations between alignable objects through the tracks connecting them. However, this approach becomes technically challenging when the number of alignment parameters is very large, such as in the case of the alignment of individual TRT straws (~700,000 parameters). In order to overcome this challenge, a simplified version of the \( \chi^2 \) approach (the Local \( \chi^2 \) method) is used. It is based on the minimisation of the same \( \chi^2 \), Eq. (5), but the implicit dependence on the fitted track parameters is dropped, reducing Eq. (6) to a simpler form:

\[ \frac{d}{d\alpha} = \frac{\partial}{\partial\alpha}. \]

Consequently, Eqs. (10) and (11) are reduced to:

\[ \left( \frac{d\chi^2_{\text{Local}}}{d\alpha} \right)^T = 2 \sum_{\text{tracks}} A^T \Omega^{-1} \mathbf{r} \]
\[ \frac{d^2\chi^2_{\text{Local}}}{d\alpha^2} = 2 \sum_{\text{tracks}} A^T \Omega^{-1} A. \]

In addition, the problem is reduced to separate systems of equations describing individual alignable modules. The Local \( \chi^2 \) method eliminates the numerical challenges of the Global \( \chi^2 \) since only systems of equations with up to six parameters (albeit many of them) need to be solved. However, due to the loss of the correlations between alignable objects, the Local \( \chi^2 \) method needs a much larger number of iterations to converge.

3.3 Solving the linear system of alignment equations

In general, the properties of the matrix representing a system of linear equations determine the most suitable solution technique. The matrix \( M \) in Eq. (11) as defined in the Global \( \chi^2 \) ansatz is found to be symmetric and singular and to have a poor matrix condition number if no constraints are applied. The addition of appropriate constraints generally renders the matrix positive definite. The singular nature of the matrix is the result of detector movements that leave a track’s \( \chi^2 \) unchanged. The simplest examples are global transformations of the detector (either translations or rotations), which are generally singular modes. A trivial way to remove these global degrees of freedom is to fix a detector element, making it the reference point for all other detector elements. This method has the unwanted drawback of arbitrarily selecting one detector module as the reference frame. In the following section, two methods used to obtain a solution to the alignment system of linear equations are discussed along with how ‘weak modes’ are removed or mitigated.

3.3.1 Diagonalisation of the matrix

The symmetric matrix \( M \) is decomposed into its diagonal basis: \( PDP^T \) where \( D \) is a diagonal matrix containing the eigenvalues of \( M \), and \( P \) is a matrix containing the eigenvectors of \( M \). Of course, in the diagonal basis all parameters (directions) are linearly independent, and the solutions plus their associated uncertainties are given by the eigenvalues (\( \lambda_i \)) with:

\[ X_D^j = -\frac{1}{\lambda_j} Y_D^j \quad \text{and} \quad \sigma(X_D^j) = \frac{1}{\sqrt{\lambda_j}}, \] (18)

where \( X_D^j \) and \( Y_D^j \) are the \( i \)th component of vectors \( X_D \) and \( Y_D \) in the diagonal basis, with \( Y_D = P^T Y \).

Singular and weak modes must be excluded as their eigenvalues are zero or have an arbitrarily large associated uncertainty, respectively. Although this can be achieved in many ways, the primary method employed is to set \( D_{i,i} = 0 \) for the modes that need to be removed, thereby creating a new diagonal matrix \( D^{-1} \) which provides the solution:

\[ X = -PD^{-1}P^T Y. \]

The DSPEV function in the LAPACK [26] software package is used as a baseline in the ATLAS implementation to diagonalise large matrices. Alternative implementations

\footnotesize\(^3\) Rotations within a magnetic field or translations in an inhomogeneous magnetic field may not be singular modes but for practical purposes may essentially be so. They are typically extremely poorly constrained because track trajectories are not significantly modified by small changes in the magnetic field.
using ROOT [27], EIGEN [28] and CLHEP [29] linear algebra classes are also available. In general, the computation time for matrix diagonalisation scales as $O(\text{DoF}^3)$ and solutions for very large systems become untenable on a single machine. If the initial matrix is poorly conditioned, the accuracy of the numerical solution can be limited by the precision of 64-bit floating-point computations for problems exceeding $O(10,000)$ DoFs.

### 3.3.2 Direct solving

Even for very large problems, direct solvers offer an accurate and CPU-efficient method for solving sparse linear equations. In addition, less memory is required as no matrix is inverted or diagonalised in the process. The LDLT Cholesky factorisation method provided within EIGEN [28] is used within the ATLAS ID alignment and takes less than 10 min to solve an alignment problem with 35,000 parameters (the approximate number of parameters needed to align all modules in the ID simultaneously) on a modern CPU. Direct solving is used when aligning thousands of degrees of freedom (usually when aligning at individual module level). Obtaining a direct solution does not offer the possibility of eliminating specific eigenmodes. Thus, other preconditioning techniques are used in order to extract a meaningful solution (e.g. Sect. 3.3.3). It is noteworthy that, although not extensively utilised within ATLAS, it is possible to iteratively find the eigenvalues and associated eigenvectors of large systems by solving $M\mathbf{x} = \lambda\mathbf{x}$ for $\mathbf{x}$ and $\lambda$ [30], which can be useful in understanding the weak modes of very large systems and identifying the underconstrained degrees of freedom.

### 3.3.3 Constraining alignment parameters in the solution (the Soft Mode Cut)

As introduced in Sect. 3.2.5, setting $a \equiv a_0$ and having a diagonal $W$ constrains the alignment parameters to their initial values. Here, $W$ denotes a diagonal matrix with diagonal elements: $\sigma(a_i)^2$, providing the tolerances to the corrections of the alignment parameters.

For this special case, the top row of Eq. (17) simplifies to Eq. (10) and the diagonal of the matrix $M$ in Eq. (11) is incremented by the reciprocal of assumed variances of alignment corrections:

\[
(M + 2W^{-1})X = -Y.
\]

The above equation can be rearranged as

\[
\sum_j \left(\frac{\sigma(a_i)\sigma(a_j)M_{ij} + 2I_{ij}}{\sigma(a_j)}\right) \frac{X_j}{\sigma(a_j)} = -\sigma(a_i)Y_j \quad (19)
\]

yielding an equation in which the corrections to the alignment parameters are normalised to their assumed uncertainties $\Delta a_i \rightarrow \Delta a_i/\sigma(a_i)$. Apart from the extra identity matrix $I$, Eq. (19) is exactly equivalent to Eq. (12).

To illustrate the effect of such a constraint, consider the case that all $\sigma(a_i)$ are equal ($\sigma(a_i) = \sigma_c$). The extra identity matrix does not affect the eigenmodes of $M$, but adds an offset to its spectrum of eigenvalues:

\[
M' = M + 2I/\sigma_c^2, \quad D' = D + I/\sigma_c^2, \quad \lambda_i' = \lambda_i + 1/\sigma_c^2.
\]

The solution in the diagonal basis, Eq. (18), takes the form:

\[
X_{D}' = \frac{1}{\lambda_i + 1/\sigma_c^2}Y_{D}' \quad \text{and} \quad \sigma(X_{D}') = \frac{1}{\sqrt{\lambda_i + 1/\sigma_c^2}}. \quad (20)
\]

Hence, one obtains a solution explicitly free from ill-defined (weak) modes. This operation does not require an explicit diagonalisation and can be used as preconditioning prior to fast solving, providing powerful control over solutions for an arbitrarily large number of DoFs. Due to the typically exponential nature of the eigenspectrum, Eq. (20) represents a solution with a clear cut-off in the diagonal basis for $\lambda_i \ll 1/\sigma_c^2$. This technique is extensively used in the ATLAS implementation.

### 4 Inner Detector alignment

The ID is composed of a large number of active detector components (see Sect. 2.1 for details). Each component or grouped collection of modules (e.g. a subdetector) can be treated as an alignable structure. The alignment is performed at different hierarchical levels following the assembly structure of the ID. Starting with the largest physical structures at level 1, the detector subsystems are aligned separated into endcaps and barrel regions in order to correct for collective movements. Level 2 treats individual barrel layers and endcap disks as physical structures (barrel modules and endcap wheels in the case of the TRT). Level 3 corresponds to a silicon module or TRT wire alignment. In this context, the SCT modules are considered as a single element in the alignment procedure due to their high construction precision [16, 17]. The levels are addressed sequentially during the alignment procedure, see Table 2.

In Run 2, the alignment levels were updated to accommodate the IBL. These changes are straightforward for levels 2 and 3, as the IBL represents merely an additional Pixel layer or additional silicon modules, respectively. The support structure of the IBL is mechanically independent from the previously installed Pixel subdetector, so IBL movement is not expected to be correlated with collective Pixel move-
Table 2  Typical alignment configurations used throughout Run 2 data taking to derive baseline alignment constants. Translational degrees of freedom (DoF) are denoted by a $T$, rotational ones by an $R$. As shown in Fig. 2, TRT barrel straws run parallel to the beam line. That corresponds to $T_z$ at level 1 and $T_y$ at level 2.

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Structures</th>
<th>DoF</th>
<th>Additional constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IBL</td>
<td>1</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pixel detector</td>
<td>1</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCT endcaps (SCT barrel fixed)</td>
<td>2</td>
<td>All except $T_z$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT split into barrel and 2 endcaps</td>
<td>3</td>
<td>All except $T_z$</td>
<td></td>
</tr>
<tr>
<td>Si 2</td>
<td>Pixel and IBL barrel split into layers</td>
<td>4</td>
<td>All</td>
<td>Beam spot, momentum bias, and impact parameter bias</td>
</tr>
<tr>
<td></td>
<td>Pixel endcaps split into disks</td>
<td>6</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCT barrel split into layers</td>
<td>4</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCT endcaps split into disks</td>
<td>18</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td>Si 3</td>
<td>Pixel and IBL barrel modules</td>
<td>1736</td>
<td>All</td>
<td>Beam spot, momentum bias, impact parameter bias, and module placement accuracy</td>
</tr>
<tr>
<td></td>
<td>Pixel endcaps modules</td>
<td>288</td>
<td>$T_z$, $T_y$, $R_z$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCT barrel modules</td>
<td>2112</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCT endcaps modules</td>
<td>1976</td>
<td>$T_z$, $T_y$, $R_z$</td>
<td></td>
</tr>
<tr>
<td>TRT 2</td>
<td>TRT barrel split into barrel modules</td>
<td>96</td>
<td>All except $T_y$</td>
<td>Momentum bias and impact parameter bias</td>
</tr>
<tr>
<td></td>
<td>TRT endcaps split into wheels</td>
<td>80</td>
<td>$T_z$, $T_y$, $R_z$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pixel and SCT detectors fixed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRT 3</td>
<td>TRT straws</td>
<td>351k</td>
<td>$T_z$, $R_z$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pixel and SCT detectors fixed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ments. Consequently, the IBL is treated as a separate physical structure at level 1.

4.1 Time-dependent alignment

Time-dependent alignment is performed for each LHC fill prior to data reconstruction to determine if the detector, or individual subsystems, have moved significantly compared to a reference alignment. Such detector movements occur on different timescales, which are classified as short, medium, or long.

Short timescales describe movements during a single LHC fill that are a result of variations of the thermal load of the ID. These movements are caused by fluctuations in the power consumption of the front-end electronics, due to variations in the trigger rate, that additionally affect the temperature of the cooling system. On medium timescales, in the range of days to a month, changes to the environmental conditions of the detector, such as ramping the magnetic field or cycling the power or cooling systems, often lead to significant movements of the detector. Slow gradual movements of the subsystems over several months (long timescales) were also observed and attributed to mechanical relaxations after sudden changes.

An automated time-dependent level 1 alignment is performed within the ATLAS prompt calibration loop [10] to address all known time-dependent movements, as detailed in Sect. 5. These results are monitored and new alignment corrections are automatically obtained during the calibration period. They serve as input for the bulk reconstruction of the corresponding dataset.

4.2 Baseline alignment constants

The baseline alignment constants are a set of reference constants that serve as initial estimates for the time-dependent refinements of the alignment. In order to achieve an accurate detector alignment and a minimisation of track parameter biases over a data-taking period, a large quantity of data are used (typically $\sim 2 \text{ fb}^{-1}$). The levels of alignment performed are summarised in Table 2. The alignment using the global $\chi^2$ method typically converges within two to four iterations for levels 1 and 2, while at least four iterations are required at level 3 (silicon). The TRT level 3 (straw level) uses the local $\chi^2$ method and requires up to 30 iterations to converge, owing to the large number of DoFs.

Depending on the alignment level, some DoFs may be fixed during the alignment procedure if poor sensitivity is
expected. Alignment levels targeting the silicon subdetectors use all tracks, whereas alignment levels including the TRT require tracks based on silicon and TRT hits. In order to remove weak modes from the alignment solution, appropriate constraints are added to the global $\chi^2$ method (see Sect. 3.2.4). Different constraints are considered depending on the expected misalignment and DoF for each alignable structure, listed in Table 2. Additionally, each subsystem can be aligned at any required level independently from the others. Further subdivision of alignment levels into smaller physical detector components, e.g. the division of individual barrel layers into staves, is also supported and used. At level 1, the SCT barrel is kept fixed due to its good stability and to serve as reference for the rest of the structures.

4.3 Residuals

As described in Sect. 3.2, the solution of the Global $\chi^2$ is the one that minimises the unbiased track-hit residuals. Figures 4, 5 and 6 show track-hit residual distributions for data and simulation in different ID subdetectors.

Data and simulation correspond to a set of muons selected in $Z \rightarrow \mu^+\mu^-$ candidate events triggered by the lowest-threshold unprescaled single and dimuon triggers. The simulation sample was generated with the POWHEG-Box v1 Monte Carlo event generator [31–33] at next-to-leading order (NLO) in $\alpha_S$ interfaced to PYTHIA 8.186 [34] for the modelling of the parton shower, hadronisation, and underlying event, with parameter values set according to the AZNLO tune [35]. The CT10 (NLO) set of parton distribution functions (PDF) [36] was used for the hard-scattering processes, whereas the CTEQ6L1 PDF set [37] was used for the parton shower. Events are required to contain two muons (satisfying ‘medium’ quality criteria as defined in Ref. [38]) with opposite charge and $p_T > 20$ GeV. In addition, requirements on the opening angle between the two muons, $\gamma(\mu^+ , \mu^-) > 45^\circ$, and their invariant mass, $70$ GeV $< m_{\mu^+\mu^-} < 110$ GeV, are imposed. In Figs. 4, 5 and 6, both data and simulation correspond to $2 \, \text{fb}^{-1}$ of data collected during 2018. Statistical uncertainties in data and simulation are included in all the figures, although barely visible as they are negligible.

Adequate agreement is seen between data and simulation in the residual distributions, where differences are quantified in terms of the ‘full width at half maximum’ (FWHM) figure of merit. A similar level of agreement is observed for the data collected during the other years of Run 2. In the case of the IBL, Pixel and SCT barrel, larger residual widths are observed in data. As shown in Sect. 5.3, the Run 2 alignment accuracy and stability in the central pseudorapidity range for both the Pixel and SCT barrel modules is controlled to a precision better than $0.5 \mu$m and $2 \mu$m in local-$x$ and local-$y$, respectively. Consequently, several other possible causes of the observed discrepancy between data and simulation are considered, such as imperfect modelling of the interactions of muons with detector material in the simulation, the material description, delta ray production modelling, mis-modelling of the detector response (and resolution) in simulation, and residual biases not uniform across individual modules in data. The latter particularly impacts the local-$y$ track-hit residuals in Fig. 4. The poorest agreement is seen for the IBL residuals, which have not yet been corrected for sensor distortions, in contrast to the Pixel layers. The sensor distortion can result in track-hit residual biases of up to $10 \mu$m within a given module, thus causing a broadening of the overall distribution [39]. The shape of the IBL modules was recently parameterised with Bernstein–Bézier functions and will be corrected in the track fitting procedure for Run 3 data taking. The cause of the small bias of $4 \mu$m in the IBL local-$y$ track-hit residuals in simulation in Fig. 4 is currently unidentified. Simulated samples use a perfectly aligned detector with no track-based alignment correction, hence this bias originates from the track or cluster reconstruction. On data, this small reconstruction bias is removed by the alignment without a significant effect on alignment precision.

5 Detector stability and time-dependent alignment

This section discusses the main sources of time variation in ID geometry and the methods implemented to mitigate these effects within the ATLAS prompt calibration loop [10]. In addition, the stability of the ID in Run 2 is summarised, final time-dependent corrections for all subsystems are presented, and the precision of the alignment is determined. All results use $pp$ collision data at $\sqrt{s} = 13$ TeV. The alignment precision for heavy-ion data in Run 2 is at least as good as the final precision of $pp$ collision data, as the instantaneous luminosity, and therefore the thermal load variations in the ID, is typically lower.

5.1 Short-timescale movements

Detector movements on short timescales are particularly challenging, since the ID track-based alignment calculates an average position correction for the time interval under study.\footnote{This time interval varies from a few minutes to several hours depending on the configuration of the alignment task.}

The procedure used to correct for rapid movements must balance two competing effects: the alignment corrections must be determined in time intervals that are short enough
Fig. 4 The IBL local-x (left) and local-y (right) residual distributions for the $Z \rightarrow \mu^+\mu^-$ data sample compared with simulated data. The distributions are integrated over all hits on tracks in barrel modules.

Fig. 5 The Pixel local-x (left) and local-y (right) residual distributions for the $Z \rightarrow \mu^+\mu^-$ data sample compared with simulated data. The distributions are integrated over all hits on tracks in barrel modules.

Fig. 6 The local-x residual distributions in the SCT (left) and TRT (right) for the $Z \rightarrow \mu^+\mu^-$ data sample compared with simulated data. The distributions are integrated over all hits on tracks in barrel modules.
to capture the motion of the particular deformation, but long enough to include sufficient data to obtain precise corrections.

5.1.1 Temperature-dependent IBL distortions

During the Run 2 commissioning of the IBL, it was already noticed that the IBL staves can be distorted by hundreds of micrometers relative to the design geometry. It was soon observed that these distortions depend on the operating temperature and correspond to module displacement in the azimuthal direction of the staves, equivalent to their local-x direction. The distortion was understood to be caused by tight, asymmetric mechanical coupling of materials with different coefficients of thermal expansion (CTE). The correlation between temperature and the size of the IBL distortion was studied using cosmic-ray data in March 2015 with a controlled variation of the IBL temperature, $T_{\text{set}}$, in the range $-20^\circ\text{C}$ to $+15^\circ\text{C}$ [40]. The size of the distortion was measured in situ using the track-based alignment and a fit to a model determined from a three-dimensional finite-element analysis. This model parameterises the IBL distortion in local-x, $\delta x(z)$, using a parabolic function,

$$\delta x(z) = B - \frac{M}{z_0} \left( z^2 - z_0^2 \right), \quad (21)$$

where $z$ is the global-z coordinate of the module, $z_0 \equiv 366.5 \text{ mm}$ is the coordinate of the stave mount at both ends, $B$ is the baseline describing the overall translation of the stave in local-x, and $M$ is the magnitude of the distortion at the stave centre. The temperature gradient of $M$ with respect to $T_{\text{set}}$ is found to be $\mathrm{d}M/\mathrm{d}T_{\text{set}} = (-10.6 \pm 0.7) \mu\text{m/K}$. The local-y position shows no temperature-dependent effect within 20 $\mu\text{m}$ uncertainty, whereas the local-z (bending out of the plane of the stave) was not included in this study. The IBL distortion is shown in Fig. 7 for different $T_{\text{set}}$ values using 2015 and 2016 pp collision data.

From the initial data taking in Run 2 through September 2015, the IBL power consumption per module was found to be stable, and fluctuations in $T_{\text{set}}$ were within $\sim 0.2$ K resulting in a stable detector ($\delta x(z) < 3 \mu\text{m}$) [40]. This situation changed with the rapid increase in integrated luminosity per LHC fill after September, which induced an increase in the low-voltage (LV) currents in the IBL module front-end electronics. This increase was traced back to radiation-induced leakage current in transistors [41]. The change in LV currents depends on the total ionisation dose. Studies show that the increase reaches a peak value for radiation doses between 10 and 30 kGy and decreases for higher doses to a value close to the pre-irradiation case.

These variations in the LV currents caused an increase in IBL module temperatures that resulted in changes in IBL distortions on short timescales. In this context, values of $\delta x(z)$ of up to $30 \mu\text{m}$ were observed between LHC fills and up to $10 \mu\text{m}$ within a single fill, corresponding to a variation of $0.5 \mu\text{m/h}^{-1}$.

5.1.2 Vertical movements of the Pixel detector

Another systematic deformation on short timescales is a change in the vertical position (global-y direction) of the Pixel detector by up to $8 \mu\text{m}$ at the start of an LHC fill. Figure 8 shows the Pixel detector vertical movement from the
start of an LHC fill. The position is computed every 20 min, which is the shortest time interval used in the ATLAS prompt calibration loop. As is evident from Fig. 8, the average position across an LHC fill does not accurately describe the position of the Pixel detector.

The cause of this movement is understood to be the following. When the Pixel detector is switched on at the start of a fill, modules reach their new temperature almost immediately as a result of the strong thermal coupling between the modules and the evaporative cooling system [1,42]. The LV current in the read-out electronics also increases immediately, while the temperature in the Pixel detector volume rises gradually during the first 60 min. The smaller mass load due to the change in density of the bi-phase cooling liquid causes the Pixel detector to rise. After this initial rise, as the instantaneous luminosity and thus the occupancy decreases over the course of the fill, LV digital currents, module temperature and Pixel volume temperature gradually decrease as well. This in turn causes an additional slow drift in the direction opposite to the initial movement. The speed of this slow drift depends on the peak luminosity per LHC fill. This speed increased during 2016 to reach values of 0.2 μm h⁻¹, as shown in Fig. 9. The vertical speed is determined as the average speed of the Pixel detector excluding the first hour after the start of data taking. This vertical drift was monitored and corrected for throughout Run 2.

5.2 Dynamic alignment on short timescales

In an effort to mitigate the effects of systematic short-timescale distortions and ensure adequate data quality for all analyses relying on tracking, conceptual improvements within the alignment framework and strategy were made. A key improvement was the introduction of a new alignment DoF, $B_x$, to parameterise the IBL distortion deformation magnitude $M$. The $B_x$ DoF correlates the local-x coordinate of each module along the IBL stave using the parabolic function defined in Eq. (21). Minimising the global $\chi^2$ with respect to $B_x$ provides corrections for varying degrees of IBL stave distortion using a single DoF, which can be done with small amounts of data. In contrast, a full level 3 alignment, which relies on a large amount of data, had been required previously, which did not allow short-timescale movements to be determined.

The automated alignment scheme that is performed within the ATLAS prompt calibration loop in Run 2 data taking determines level 1 and IBL $B_x$ (per stave) dynamic alignment constants every 20 min at the start of a fill and every 100 min for the rest of the fill. This level of granularity in time is adequate to mitigate the effects of short-timescale vertical movements on track parameter resolution. The alignment is performed in two iterations of the level 1 calibration loop (level 1 CL) followed by two dedicated iterations to correct for IBL distortions. The $B_x$ correction in the level 1 CL corresponds to a collective, uniform correction for all IBL staves. The dedicated IBL bowing iterations determine $B_x$ individually for each stave, as summarised in Table 3. The SCT barrel is used as the reference in the dynamic alignment.

5.3 Inner Detector stability during Run 2 data taking

5.3.1 Time-dependent corrections for all subsystems

The performance of the dynamic alignment scheme using 2016 pp collision data is shown in Figs. 10 and 11. The average bowing magnitude of the 14 IBL staves relative to the baseline alignment is compared with the results of the dynamic alignment in Fig. 10. Figure 11 shows the average IBL distortion computed after different alignment corrections versus time in the form of luminosity blocks (LB), which correspond to stable data-taking conditions in periods of approximately 1 min. It also compares the unbiased local-x residuals computed using the a fill-averaged correction (for illustration only) with those obtained after computing the full dynamic alignment correction, which is derived in short time-intervals. A clear improvement in the residual distributions is seen after applying dynamic alignment corrections. Figures 10 and 11 illustrate that, averaged over an LHC fill, even very large values of $M$ (up to 30 μm) are accurately corrected for using $B_x$ as an alignment DoF. These features were present for all Run 2 data, although there was some saturation of the effect in the later years of Run 2, as observed in the radiation damage studies of the IBL [41].
Table 3  Typical alignment configurations used throughout Run 2 data taking to derive dynamic alignment corrections

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Structures</th>
<th>DoF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CL IBL</td>
<td>IBL</td>
<td>1</td>
<td>All DoF incl. $B_x$, except $R_z$</td>
</tr>
<tr>
<td></td>
<td>Pixel detector</td>
<td>1</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>SCT endcaps (SCT barrel fixed)</td>
<td>2</td>
<td>All except $T_z$</td>
</tr>
<tr>
<td></td>
<td>TRT split into barrel and 2 endcaps</td>
<td>3</td>
<td>All except $T_z$</td>
</tr>
<tr>
<td>IBL bowing</td>
<td>IBL staves</td>
<td>14</td>
<td>$B_z$</td>
</tr>
<tr>
<td></td>
<td>Pixel, SCT, and TRT detectors fixed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The long-term trend of the Pixel and IBL detector movements relative to the baseline alignment correction is shown in Fig. 12 for the average $B_x$ correction, the global $T_x$ correction, and the global $T_y$ correction. For the sake of clarity, the plots in Fig. 12 show only a fraction of the Run 2 data; the remaining data follow the same trend.

5.3.2 Final alignment precision of each subsystem

The final alignment precision of each ID subsystem is determined from the track-hit residuals of individual silicon modules for each LHC fill in 2015 and 2016 data taking after the dynamic alignment corrections are applied. These dynamic
alignment corrections are computed either for large structures (e.g. the Pixel detector) as a collective movement of all modules or using a simplified parameterisation (like $B_x$). In this context, less significant module-to-module movements remain uncorrected by the dynamic alignment. This effect is seen as a residual time-dependent misalignment or ‘instability’ of the modules. This instability is estimated for each silicon layer and module $z$-position by integrating modules over $\phi$ into one group. Results are presented for the ‘in-plane’ translation DoFs only (local-$x$ and local-$y$).

For each module, the average track-hit residual, $\langle r_{x,y} \rangle$, is computed for each LHC fill, for both local-$x$ and local-$y$, on a set of calibration data, whose size is approximately independent from the fill conditions. Its statistical uncertainty, $\sigma_{r_{x,y}}/\sqrt{N}$, where $N$ is given by the number of tracks per
module and $\sigma_{x,y}$ is the standard deviation of the residuals, is computed assuming that the residual distribution is approximately Gaussian. The dispersion $\sigma_{(x,y)}$ of the distribution in $(x_y)$ obtained from all LHC fills is an estimate of the total instability of the module position after all alignment corrections are applied. This total uncertainty can be divided into a statistical component ($\sigma_{(x,y)}/\sqrt{N}$) and a component describing residual instability due to uncorrected time-dependent movements and stochastic fluctuations, $\sigma_{x,y}^{\text{time}}$:

$$\sigma_{(x,y)} \sim \sigma_{x,y}^{\text{time}} + \frac{\sigma_{x,y}}{\sqrt{N}}.$$  

As the size of the statistical contribution per module per LHC fill is generally small, $\sigma_{x,y}^{\text{time}}$ is estimated by

$$\sigma_{x,y}^{\text{time}} \equiv \sqrt{\sigma_{x,y}^{2} - \left(\frac{\sigma_{x,y}}{\sqrt{N}}\right)^{2}}.$$  

Figures 13, 14 and 15 show the estimated instability in local-$x$ and local-$y$ of the Pixel and SCT barrel layers as a function of module $z$-position. Figure 16 shows the local-$x$ and local-$y$ instabilities of the modules in the Pixel endcap layers. The alignment in the central pseudorapidity range for both the Pixel and SCT barrel modules is controlled to a precision better than $0.5\, \mu m$ and $2\, \mu m$ in local-$x$ and local-$y$, respectively. This level of control is considered to be very good given the time-dependent corrections of $O(10\, \mu m)$ due to the IBL distortion and the vertical movement of the Pixel detector. The same level of precision is not achieved for the outermost IBL modules (3D sensors) corresponding to the range $|\eta| > 2.5$. There the alignment precision in local-$x$ (local-$y$) is measured with an uncertainty better than $3\, \mu m$ ($15\, \mu m$). This region is particularly challenging due to the low number of tracks. Moreover, the tracks in this region have only small overlap with other ID tracking layers. Furthermore, the large IBL local-$y$ uncertainty may be related to a deformation of IBL staves in local-$z$ over time, which is not corrected for in the prompt calibration loop.

The instability for Pixel endcap modules is larger than for barrel modules; the local-$x$ and local-$y$ instabilities are 2–4 $\mu m$ and 4–7 $\mu m$, respectively. This instability corresponds to the size of the movements of the Pixel endcap modules relative to the baseline alignment over time. The precision achieved is nearly one order of magnitude better than the required precision [2]. This required precision was defined in order to limit the degradation of the resolution of the track parameters for high-momentum tracks to less than 20% in comparison with a perfectly aligned detector. While succeeding in its primary goal, these results, specifically the residual Pixel endcap movements, also imply that the current dynamic alignment scheme, which allows time-dependent alignment of the entire Pixel detector as one unit, is not optimal. A higher level of precision might be achieved if the Pixel endcap disks were aligned individually. This improvement is under study for LHC Run 3, including detailed cross-checks for new weak modes that may be introduced due to the additional DoFs within the calibration loop.

6 Momentum biases

The alignment corrections described in Sects. 4 and 5 target misalignments that change the $\chi^{2}$ of the track fit in Eq. (1). In contrast, correlated geometrical distortions referred to
as weak modes leave the $\chi^2$ of the fitted tracks virtually unchanged and can systematically bias the reconstructed track parameters. Momentum biases induced by correlated detector misalignments can generally be classified into two categories:

- **Sagitta** deformations consist of detector geometry distortions in the bending plane that affect the reconstructed track curvature differently for positively and negatively charged particles (Fig. 17 left).
- **Length scale** biases are characterised by detector geometry distortions along the track trajectory and affect the reconstructed curvature identically for positively and negatively charged particles (Fig. 17 right).

These biases can be mitigated through the use of constraints either on track parameters (Sect. 3.2.4) or on alignment parameters (Sect. 3.2.5), or on both simultaneously. Residual momentum biases, after corrections to the detector alignment have been made, are sufficiently small that they can be accounted for by directly correcting track parameters.

### 6.1 Sagitta bias

Displacements of the reconstructed hits in the bending plane orthogonal to the track path result in a charge-antisymmetric alteration of the track curvature, which is parameterised as

$$p' = p \left(1 + q \ p_T \ \delta_{\text{sagitta}}\right)^{-1},$$  \hspace{1cm} (22)

where the un-primed quantities correspond to the true values, the primed quantities correspond to the reconstructed values, $q$ refers to the sign of the electric charge of the particle and $\delta_{\text{sagitta}}$ is a bias parameter common to all measured momenta and uniquely defines the detector geometry deformation.

Two iterative methods are used to determine the sagitta biases. The first method uses $Z \rightarrow \mu^+\mu^-$ decays. The second uses the electromagnetic calorimeter as a reference, and utilises the ratio of the measured energy deposited in the calorimeter ($E$) to the momentum ($p$) measured by the ID for electrons and positrons. Both methods allow the detector to be segmentated arbitrarily in $\eta$ and $\phi$, allowing the study of localised sagitta biases. Sagitta biases have, to a great extent, been corrected for during the determination of the alignment constants by adding constraints to the parameters of the tracks used to perform the detector alignment, as given by Eqs. (14) and (16) and explained in Sects. 3.2.4 and 3.2.5, and also in Refs. [3,43]. The methods used to calculate the constraints are described below, and the residual sagitta biases after alignment corrections are shown.

#### 6.1.1 Measuring sagitta biases using $Z \rightarrow \mu^+\mu^-$ decays

The invariant mass, $m$, of two highly relativistic opposite-charge particles is given approximately by

$$m^2 = 2p^+p^- (1 - \cos \gamma),$$

where $p^+$ and $p^-$ are the magnitudes of the momenta of the positively and negatively charged particles, and $\gamma$ is defined as their opening angle. In the following, $+$ and $-$ superscripts refer to the properties of the positively and negatively charged muons respectively. Sagitta biases can be measured using any particle (of reasonably narrow width) that decays into pairs of stable particles. In LHC conditions, resonances that decay into pairs of muons (such as $J/\psi$, $\Upsilon$ and $Z$) present the advantage that the dimuon signature can be clearly distinguished from the large hadronic background. For $\delta_{\text{sagitta}}$ studies, $Z \rightarrow \mu^+\mu^-$ decays are preferred due to the high momentum of the $Z$ decays products. Data quality selection criteria, summarised in Table 4, are applied to both the selected muon candidates and the dimuon system. In total, more than 70 million $Z \rightarrow \mu^+\mu^-$ candidate events were selected.
In general, geometrical distortions that bias sagitta measurements can be localised in specific regions of the detector. As a result, the sagitta bias parameter explicitly depends on the path of the track, which can be approximated by the direction of the track at the \( pp \) interaction point, given by \( \eta \) and \( \phi \): \( \delta_{\text{sagitta}} \rightarrow \delta_{\text{sagitta}}(\eta, \phi) \). The difference at leading order in \( \delta_{\text{sagitta}}(\eta, \phi) \) between the reconstructed dimuon invariant mass using the uncorrected geometry \( (m_{\mu\mu}) \) and the expected mass \( (m_Z) \) for each event is given by:

\[
\begin{align*}
    m^2_{\mu\mu} - m^2_Z & \approx m^2_Z \left( p^\tau_\mu \delta_{\text{sagitta}}(\eta^+, \phi^+) \right. \\
    & \left. - p^\tau_\mu \delta_{\text{sagitta}}(\eta^-, \phi^-) \right).
\end{align*}
\]

(23)

The corrections are updated by adding the average of the detector layers depending linearly on the radius. The real (dashed black line) and fitted (solid black line) particle trajectories are shown. Red stars indicate real measurement positions and grey stars show the reconstructed hit positions (biased measurements).

---

**Fig. 17** A simplified representation of two common weak modes that bias the track momentum. A *sagitta* bias (left) is caused by a deformation in the bending plane of the tracks, e.g. a rotation of the detector layers depending linearly on the radius. A *length scale* bias (right) caused by a deformation along the track trajectory, e.g. a radial expansion of the detector layers depending linearly on the radius. The real (dashed black line) and fitted (solid black line) particle trajectories are shown. Red stars indicate real measurement positions and grey stars show the reconstructed hit positions (biased measurements).

---

**Table 4** Event selection criteria for \( Z \rightarrow \mu^+\mu^- \) (Sect. 6.1.1) and \( Z \rightarrow e^+e^- \) (Sect. 6.1.2) candidate events for the analyses of the sagitta biases in data. Events triggered by the lowest-threshold unprescaled single and double electron and muon triggers are used to select \( Z \rightarrow \mu^+\mu^- \) and \( Z \rightarrow e^+e^- \) event candidates. \( \gamma(\mu^+, \mu^-) \) is the opening angle between the muons. \( \Delta d_0(\mu^+, \mu^-) \) and \( \Delta z_0(\mu^+, \mu^-) \) are defined as the difference in \( d_0 \) and \( z_0 \) between the two muons.

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th>( Z \rightarrow \mu^+\mu^- )</th>
<th>( Z \rightarrow e^+e^- )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepton selection</td>
<td>Two muons(^a) associated with the primary vertex(^b)</td>
<td>Two electrons(^e) associated with the primary vertex(^b)</td>
</tr>
<tr>
<td>( p_T &gt; 12 \text{ GeV} )</td>
<td>( E_T &gt; 25 \text{ GeV} )</td>
<td></td>
</tr>
<tr>
<td>(</td>
<td>\eta</td>
<td>&lt; 2.5)</td>
</tr>
<tr>
<td>(d_0) significance(^d) &lt; 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dilepton selection</td>
<td>(70 \text{ GeV} &lt; m_{\mu^+\mu^-} &lt; 110 \text{ GeV} )</td>
<td>(</td>
</tr>
<tr>
<td>(</td>
<td>\Delta d_0(\mu^+, \mu^-)</td>
<td>&lt; 0.1 \text{ mm} )</td>
</tr>
<tr>
<td>(</td>
<td>\Delta z_0(\mu^+, \mu^-)</td>
<td>&lt; 0.6 \text{ mm} )</td>
</tr>
<tr>
<td>(\gamma(\mu^+, \mu^-) &gt; 45^\circ)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Both muon candidates are required to satisfy the ‘medium’ quality criteria as defined in Ref. [38]  
\(^b\)The reconstructed vertex with the largest \( \sum p^2_T \) of its tracks  
\(^c\)Excluding the transition region between the barrel and forward calorimeters  
\(^d\)The significance is defined as \( |d_0|/\sigma(d_0) \), where \( \sigma(d_0) \) is the uncertainty on the \( d_0 \) from the track fit  
\(^e\)Both electron candidates are required to satisfy the ‘loose’ quality criteria as defined in Ref. [46]
current iteration to the result of the previous iteration:

\[
\delta_{\text{sagitta},i}(\eta, \phi) = \delta_{\text{sagitta},i-1}(\eta, \phi) + \langle \Delta \delta_{\text{sagitta},i}(\eta, \phi) \rangle.
\]

where \( \langle \Delta \delta_{\text{sagitta},i}(\eta, \phi) \rangle \) is the average bias in a \( (\eta, \phi) \) region. The value of \( m_{\rho \rho}^2 \) is computed using Eq. (23) with the values of \( \delta_{\text{sagitta}} \) from the previous iteration. The iterations are repeated until convergence is reached.

The method, as described by Eq. (23), is only sensitive to relative sagitta biases in different sectors of the detector. An alternative method, comparing the \( p_T \) spectrum of the \( \mu^+ \) and \( \mu^- \) [45] was also tested. This method is sensitive to global sagitta biases, although it is also subject to detector acceptance effects and requires more data to achieve the same statistical precision as the mass-based method.

Figure 18 shows the measured sagitta distortions depending on the track direction using this technique. The central barrel region of the detector is largely free of sagitta bias, while the endcap regions exhibit some areas of small residual sagitta bias. The distribution of \( \delta_{\text{sagitta}} \) for the full Run 2 data is shown in Fig. 18. Figure 19 shows the average \( \delta_{\text{sagitta}} \) versus \( \eta \) and \( \phi \), as well as its RMS. The distributions, split by data-taking year, have compatible shapes indicating a consistent and stable detector geometry during Run 2.

6.1.2 Measuring sagitta biases using the \( E/p \) ratio of electrons and positrons

Assuming that the calorimeter response is independent of the charge of the incoming particle and that a perfectly aligned detector reconstructs the momentum of charged particles correctly, charge-dependent momentum biases are expected to result in differences in the \( E/p \) ratio of positive and negative particles. This ratio is defined as the ratio of the calorimeter energy measurement \( (E) \) to the track momentum measurement \( (p) \). This technique is mainly suitable for electrons and positrons. In the presence of a sagitta bias, the \( \langle E/p \rangle \) ratio would be modified as \( \langle E/p' \rangle = \langle E/p \rangle + q \langle E_T \rangle \delta_{\text{sagitta}} \), where \( E_T \equiv E/cosh \eta \) is referred to as the transverse energy. Assuming that the average transverse momentum of positrons and electrons is equal, the sagitta bias can be estimated [3] as

\[
\delta_{\text{sagitta}} = \frac{\langle E/p' \rangle^+ - \langle E/p' \rangle^-}{2 \langle E_T \rangle}.
\]

To take into account any biases introduced by the aforementioned assumptions the value of \( \delta_{\text{sagitta}} \) is determined iteratively, correcting the momentum using Eq. (22) at each iteration. It should be noted that biases in the calorimeter energy scale cancel out to first order and any residual dependence would be reduced by this iterative procedure. In addition, this method is, by construction, sensitive to global sagitta biases. Data quality selection criteria are applied to both the selected electron candidates and the electron–positron system and are summarised in Table 4.

Figure 20 shows the \( \delta_{\text{sagitta}} \) as obtained from the \( E/p \) method. This results support the observations from Sect. 6.1.1: the central barrel region of the detector is largely free of sagitta bias, while the endcap regions exhibit regions of small residual sagitta bias. Compared to Fig. 18 (right) a global offset of \( \sim 0.05 \text{TeV}^{-1} \) can be seen in Fig. 20 (right) indicating the presence of a small global sagitta bias. Figure 21 shows the average \( \delta_{\text{sagitta}} \) versus \( \eta \) and \( \phi \), as well as its RMS. The \( \delta_{\text{sagitta}} \) distributions from the \( E/p \) method split by data-taking year have comparable shape to those obtained from the \( Z \to \mu^+ \mu^- \) mass method, further supporting the observation of a consistent and stable detector geometry during Run 2. The change in position due to the residual sagitta bias \( \sim 0.1 \text{TeV}^{-1} \) when extrapolating a track from the detector origin to the outermost SCT endcap disk (radius of 500 mm and a \( z \)-axis distance of 2720 mm from the detector origin) is less than 10 \( \mu \)m.

6.2 Length scale biases

Displacements of the reconstructed hits parallel to the track direction result in a charge-symmetric alteration of the measured track curvature. In a tracker with a solenoidal magnetic field these can be induced by changes in the radial or longitudinal length scale of the detector with little impact on the track fit quality.

If the actual radius of a detector module, \( R \), is assumed to be \( R \left( 1 + \varepsilon_r \right) \), then for small distortions \( |\varepsilon_r| \ll 1 \), the reconstructed momentum will be:

\[
\begin{align*}
p'_r & = p_T \left( 1 + \varepsilon_r \right) \\
p'_z & = p_z.
\end{align*}
\]

Equation (26) assumes that the length scale in the bending plane also expands by a factor \( (1 + \varepsilon_r) \), which implies that dimensions of sensitive detector modules would also expand by the same factor. If it is assumed that detector modules do not expand in the bending plane then the reconstructed transverse momentum will be biased by a factor of \( (1 + 2 \varepsilon_r) \).

Similarly, if the actual longitudinal dimension of a detector module, \( z \), is assumed to be \( z \left( 1 + \varepsilon_z \right) \), the reconstructed momentum will be:

\[
\begin{align*}
p'_r & = p_T \\
p'_z & = p_z \left( 1 + \varepsilon_z \right).
\end{align*}
\]

Linear combinations of both the radial and longitudinal biases are also considered. It is worth noting that there is a
Fig. 18 Sagitta biases versus $\eta$ and $\phi$ (left) for 2018 data and the overall sagitta biases in the Run 2 data (right) for the $Z \rightarrow \mu^+\mu^-$ method. The error bars represent the statistical uncertainty.

Fig. 19 Measured sagitta biases versus $\eta$ (top) and $\phi$ (bottom) using the $Z \rightarrow \mu^+\mu^-$ method. The average (left) and the RMS (right) of the sagitta bias is shown. The markers of the data points of the different years are slightly shifted in $\eta$ and $\phi$ for better visibility. The error bars represent the statistical uncertainty.

degeneracy between the effects of a bias in the magnetic field and a global scaling of the detector (radial and longitudinal: $\varepsilon_r$, $\varepsilon_z$), as both lead to a momentum bias of the form $p (1 + \varepsilon_r)$. Consequently, if $B$ is assumed to be $B (1 + \varepsilon_r)$ the particle momentum scales as $p (1 + \varepsilon_r)$.

The relationship between the reconstructed invariant mass of a particle decaying into two muons ($m'_{\mu\mu}$), and the true mass ($m_{\mu\mu}$), assuming that the radial and longitudinal biases in Eqs. (26) and (27) are both small, is given by:

$$m^{'2}_{\mu\mu} \approx m^2_{\mu\mu} + 2E^+E^- \left[ (\beta_T^+)^2 - \beta_T^+ \cdot \beta_T^- \right] \varepsilon_r(\eta^+, \phi^+)$$
$$+ 2E^+E^- \left[ (\beta_T^+)^2 - \beta_T^+ \cdot \beta_T^- \right] \varepsilon_r(\eta^-, \phi^-)$$
$$+ 2E^+E^- \left[ (\beta_T^-)^2 - \beta_T^- \cdot \beta_T^+ \right] \varepsilon_z(\eta^+, \phi^+)$$
$$+ 2E^+E^- \left[ (\beta_T^-)^2 - \beta_T^- \cdot \beta_T^+ \right] \varepsilon_z(\eta^-, \phi^-),$$
Fig. 20 Sagitta biases versus $\eta$ and $\phi$ (left) for 2018 data and the overall sagitta biases of the Run 2 data (right) for the $E/p$ method. The error bars represent the statistical uncertainty.

Fig. 21 Sagitta biases versus $\eta$ (top) and $\phi$ (bottom) for the $E/p$ method. The average (left) and the RMS (right) of the sagitta bias is shown. The markers of the data points of the different years are slightly shifted in $\eta$ and $\phi$ for better visibility. The error bars represent the statistical uncertainty.

where the $\beta = p/E$ is the velocity of the particle. This approximation is valid to first order in $\varepsilon$.

In a simpler case, where only a global radial and longitudinal bias are present, the reconstructed mass is:

\[
m^{2}_{\mu\mu} \approx m^{2}_{\mu\mu} + 2E+E^{-}[\beta_{T}^{+} - \beta_{T}^{-}]^{2}\varepsilon_{r} + 2E^{+}E^{-}[\beta_{z}^{+} - \beta_{z}^{-}]^{2}\varepsilon_{z},
\]

which, in the limit where the muon mass is ignored leads to

\[
m^{2}_{\mu\mu} \approx m^{2}_{\mu\mu} + 2m^{2}_{\mu\mu}\varepsilon_{r}\sin^{2}\alpha + 2m^{2}_{\mu\mu}\varepsilon_{z}\cos^{2}\alpha
\approx m^{2}_{\mu\mu} + 2m^{2}_{\mu\mu}(\varepsilon_{s} + \varepsilon_{r}^{'\sin^{2}\alpha}) ,
\]

where

\[
\sin^{2}\alpha = E^{+}E^{-}[\beta_{T}^{+} - \beta_{T}^{-}]^{2}/m^{2}_{\mu\mu} ,
\]
Fig. 22 Ratio of the measured mass to the reference as a function of \( \sin^2 \alpha \). Due to event kinematics, \( J/\psi \rightarrow \mu^+\mu^- \) events (left) cover the entire \( \sin^2 \alpha \) range while \( Z \rightarrow \mu^+\mu^- \) events (right) cover a smaller range. Error bars represent the statistical uncertainty. The red lines show the fit to Eq. (28) from which the values of \( \epsilon_s \) and \( \epsilon_r' \) are extracted.

Fig. 23 The measured momentum scale bias \( \epsilon_s \) as a function of track \( p_T \). Error bars represent the statistical uncertainty. Left: \( J/\psi \rightarrow \mu^+\mu^- \) decays; right: \( Z \rightarrow \mu^+\mu^- \) decays.

\( \epsilon_s = \epsilon_z \), and \( \epsilon_r' = \epsilon_r - \epsilon_z \) is the difference between the radial and longitudinal components of the momentum scale.\(^6\)

Thus, by measuring the mass as function of \( \sin^2 \alpha \) it is possible to differentiate between radial and scale biases. Figure 22 shows the measured scale using \( J/\psi \) and \( Z \) decays into \( \mu^+\mu^- \) in the barrel of the ID. The results show a clear momentum scale bias but no significant radial scale \( (\epsilon_r') \) as the reconstructed mass is constant as a function of \( \sin^2 \alpha \).

An analysis using an iterative procedure similar to the \( \delta_{\text{sagitta}} \) method, Eq. (24), is also performed. Here, the momentum scale factor \( (\epsilon_s) \) is computed and consequently used to update the momentum of the tracks at the next iteration. This method allows biases to be measured as a function of any kinematic or geometric parameter. The results as a function of the track \( p_T \) are presented in Fig. 23. The magnitude of the momentum scale bias is observed to be constant as a function of track \( p_T \) as expected from a length scale or magnetic field strength bias.

\( \epsilon_s \approx -9 \times 10^{-3} \). This result is in agreement with the momentum scale at the ID for muons [38]. As previously highlighted, the origin of such a global momentum scale bias cannot be unambiguously resolved by these studies. It should be noted that: the measurement of the absolute scale of the magnetic field has an uncertainty, which is about four times smaller than the observed scale bias [7].

7 Impact parameter biases

The weak modes of the alignment can also lead to a bias in the transverse \( (d_0) \) and longitudinal \( (z_0) \) impact parameters. For example, a rotation of the IBL or radial distortions of the Pixel layers can lead to transverse impact parameter biases. The quality of the detector alignment can be assessed by analysing impact parameter biases as a function of track \( p_T \) and \( \eta \). For this study, events are selected using a combination of single-jet triggers with several jet \( p_T \) thresholds.
starting at 100 GeV. The standard ATLAS event cleaning selection is applied, ensuring that all detectors were fully operational. In order to disentangle the biases due to residual misalignment from those originating from the track reconstruction algorithms, recorded data are compared with a dijet Monte Carlo simulation sample generated with Pythia [47]. The primary vertex of each selected event must have at least three tracks associated with it. Tracks are selected by requiring them to be assigned to jets using ghost association [48], a procedure that treats them as four-vectors of infinitesimal momentum magnitude during the jet reconstruction and assigns them to the jet with which they are clustered. Jets are reconstructed using the anti-$k_t$ algorithm [49] with radius parameter $R = 0.4$. In addition, tracks are required to have at least 9 silicon (Pixels + SCT) hits for $|\eta| \leq 1.65$, at least 11 silicon hits for $|\eta| > 1.65$, a maximum of 2 SCT holes, no Pixel hole, $p_T > 3$ GeV, $|\eta| < 2.5$, and an opening angle $\Delta R(\text{track}, \text{jet}) < 0.4$ relative to the reconstructed jet axis. A track $p_T$ of 3 GeV corresponds to the lowest momentum threshold typically used within the alignment to reduce MCS effects (see Sect. 3.1). The impact parameters are obtained relative to the primary vertex by extrapolating the particle trajectory to its position. This is particularly relevant for the longitudinal impact parameter, as the width of the luminous region in the $z$ direction is very broad. The impact parameter biases are extracted by iteratively fitting the distribution of impact parameters relative to the primary vertex with a Gaussian function within a $\pm 2\sigma$ range until the fitted $\mu$ and $\sigma$ are stable within 1%. The resulting value of the Gaussian mean ($\mu$) represents the estimate of the impact parameter bias.

Figure 24 shows the transverse and longitudinal impact parameter biases as a function of the delivered luminosity in Run 2. Data collected in 2016 have a period-dependent $d_0$ bias of $-4 \mu$m (early 2016) and $+3 \mu$m (late 2016). This bias was introduced by a change in the underlying geometry description of the ATLAS ID and a misconfiguration of the beam-spot constraint. Data collected in 2017 and 2018 show overall $d_0$ biases of less than 1 $\mu$m. The longitudinal impact parameter bias is negligible and constant across the years (below 0.5 $\mu$m). In Figs. 25 and 26 the transverse and longitudinal impact parameter biases are shown as function of the track transverse momentum and track $\eta$, respectively. The small bias in the longitudinal impact parameter as a function of track $\eta$ is present in simulation and data and is consequently not introduced by the track-based alignment because it is not applied to simulation (where perfect alignment is assumed). The resulting bias has no significant effect on the ATLAS tracking performance as the longitudinal impact parameter resolution is on the order of 100 $\mu$m for tracks with $p_T > 10$ GeV.

8 Conclusion

This paper describes the precision alignment of the ATLAS Inner Detector (ID) for Run 2 and quantifies the impact of alignment uncertainties on track parameter biases. The alignment procedure consists of a track-based algorithm that minimises track-hit residuals. It calculates the track parameters at each measurement surface and encodes the relationship between track-hit residuals and the alignment parameters of each alignable structure. To resolve ambiguities, it imposes
externally determined constraints on track parameters, e.g., using tracks from resonance decays. The alignment procedure is performed at different hierarchical levels, starting from the largest physical structures and proceeding to individual detector modules or sensor elements. The number of degrees of freedom increases for each subsequent alignment level. In total, more than 36,000 degrees of freedom are considered when aligning all silicon modules (IBL, Pixel and SCT) and more than 700,000 degrees of freedom are added for the TRT.

It has been observed that operational conditions affect the positions of ID elements. The Pixel detector moves rapidly upwards every time the data acquisition is activated. The staves of the IBL bow depending on the temperature; the degree of variation depends on the thermal load and is a function of the accumulated radiation dose and of the luminosity. The remaining detector structures are quite stable during an LHC fill; the movements of individual modules in the barrel have an RMS at the micrometer level while those in the endcap regions range from 2 µm to $O(10\,\mu\text{m})$.

An automated alignment procedure that corrects for relatively rapid movements of the Pixel detector and IBL and the relative positions of all of other subdetectors is executed for every LHC fill for which the ID collects data. The detailed alignment of all the other structures (subdetectors, barrel, endcaps, layers, disks, modules or wires) is determined in dedicated alignment campaigns. The impact of alignment weak modes, namely distortions that leave the track fit quality largely unchanged and can bias the measured track parameters, is minimised during these campaigns by employing external constraints on track parameters. Independent measurements are performed to quantify potential biases, enabling them to be largely removed. The residual sagitta bias and momentum scale bias after the full Run 2 alignment are reduced to less than $\sim 0.1\text{ TeV}^{-1}$ and $0.9 \times 10^{-3}$, respectively. Remaining track parameter biases do not significantly impact ATLAS physics analyses.
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; CNPq and FAPESP, Brazil; NSERC, 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 DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRT, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNISW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russia Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, 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, USA. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; ERC, ERDF, Horizon 2020, Marie Skłodowska-Curie Actions and COST, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes cofinanced by EU-ESF and the Greek NSRF, Greece; BSE-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya and PROMETEO Programme Generalitat Valenciana, Spain; Gábor Gustafsson’s Stiftelse, Sweden; 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. [50].

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: All ATLAS scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by CERN. Data associated with journal publications are also made available: tables and data from plots (e.g. cross section values, likelihood profiles, selection efficiencies, cross section limits, ...) are stored in appropriate repositories such as HEPDATA (http://hepdata.cedar.ac.uk/)]. ATLAS also strives to make additional material related to the paper available that allows a reinterpretation of the data in the context of new theoretical models. For example, an extended encapsulation of the analysis is often provided for measurements in the framework of RIVET (http://rivet.hepforge.org/).” This information is taken from the ATLAS Data Access Policy, which is a public document that can be downloaded from http://opendata.cern.ch/record/413 [opendata.cern.ch].]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as 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. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-right holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Funded by SCOAP3.

Appendix A: track fitting with multiple Coulomb scattering effects

The track fit can be improved by considering the charged particle scattering as it passes through material in the detector. Knowing the detector material description, one can estimate and fit the scattering angles in the sensor planes [50]. To include the effects of multiple scattering, terms are added directly to the track $\chi^2$ as it is done in Eq. (3), which can be written as:

$$\chi^2_{\text{Track}} = \sum_i \left( \frac{r_i(\tau, \theta)}{\sigma_i} \right)^2 + \sum_j \left( \frac{\hat{\theta} - \theta_j}{\Theta_{jj}} \right)^2.$$  

It should be noted that the residuals now also depend on the scattering angles, $\theta$. The scattering expectation value, $\hat{\theta}$, is zero and its variance, $\Theta_{jj}$, depends on the particle momentum and amount of material traversed. The uncertainty of the $i$-th measurement is denoted by $\sigma_i$.

The $\chi^2$ has to be minimised for $\tau$ and $\theta$ simultaneously. Defining the derivative of residuals with respect to track and scattering parameters to be:

$$G \equiv \frac{\partial r}{\partial \tau} \quad S \equiv \frac{\partial r}{\partial \theta}$$

the derivatives of $\chi^2$ with respect to the track and scattering parameters are:

$$\frac{1}{2} \frac{d\chi^2}{d\tau} = G^T \Omega^{-1} r,$$
$$\frac{1}{2} \frac{d\chi^2}{d\theta} = S^T \Omega^{-1} r + \Theta^{-1} \theta.$$  

Neglecting second-order derivatives of residuals, the second derivatives of $\chi^2$ with respect to perigee and scattering parameters are:

$$\frac{1}{2} \frac{d^2\chi^2}{d\tau^2} = G^T \Omega^{-1} G,$$
$$\frac{1}{2} \frac{d^2\chi^2}{d\theta^2} = S^T \Omega^{-1} S + \Theta^{-1},$$
$$\frac{1}{2} \frac{d^2\chi^2}{d\theta d\tau} = G^T \Omega^{-1} S.$$
The above can be written down in a compact form using Eq. (4):

\[
\frac{1}{2} \frac{d\chi^2}{d\varpi} = H^\top V^{-1} \rho, \\
\frac{1}{2} \frac{d\chi^2}{d\varpi^2} = H^\top V^{-1} H.
\]

References


**ATLAS Collaboration**

Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
Also at National Research Nuclear University MEPhI, Moscow, Russia
Also at Physics Department, An-Najah National University, Nablus, Palestine
Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
Also at The City College of New York, New York NY, USA
Also at TRIUMF, Vancouver BC, Canada
Also at Universita di Napoli Parthenope, Naples, Italy
Also at University of Chinese Academy of Sciences (UCAS), Beijing, China
* Deceased