Alignment of the ATLAS Inner Detector Tracking System

John Alison[*][†]

Department of Physics and Astronomy, University of Pennsylvania

E-mail: `johnda@hep.upenn.edu`

These proceedings present the track-based method of aligning tracking systems such as the AT-LAS Inner Detector at the CERN Large Hadron Collider. This alignment is crucial for commissioning and operating the spectrometer. We discuss one of the most difficult challenges of the method, alignment weak modes. Initial results from cosmic-ray data collected with the ATLAS Inner Detector are presented.

---

[*]Speaker.

[†]On behalf of the ATLAS Collaboration.

1. Introduction

ATLAS [1] is one of two general purpose experiments at CERN's Large Hadron Collider. It will record proton-proton collisions at a center-of-mass energy of up to 14 TeV. The ATLAS detector is comprised of a central tracking system (the Inner Detector or ID) surrounding the interaction point, electromagnetic and hadronic calorimeters, muon detectors, and a magnet system made of a central solenoid and a series of toroid magnets. The ID and solenoid form a spectrometer in which the 2T magnetic field of the solenoid curves the trajectory of charged particles into a helix. The solenoid defines an axis along which the magnetic field points and is the basis for the cylindrical coordinate system used in ATLAS. In this coordinate system the interaction point is defined as the origin. The z-axis points along the magnetic field; r is defined as the distance from the z-axis, and ϕ is the angle around the z-axis. The ID is built around this z-axis with a cylindrical geometry. It is composed of central barrel layers and endcap wheels or disks. The barrel layers are centered on the origin, are symmetric in ϕ and are located at constant r. The endcap wheels are also symmetric in ϕ, however, they are extended in r and located at fixed z beyond the extent of the barrel. The purpose of the ID spectrometer is to measure the trajectory and the momentum of charged particles. These reconstructed trajectories are referred to as "tracks".

The ATLAS ID consists of three sub-detectors built using two technologies: silicon sensors and straw drift tubes. Collisions at the interaction point produce charged particles that travel outward through the ID, first crossing the Pixel detector, then the Semiconductor Tracker (SCT), and finally the Transition Radiation Tracker (TRT). The Pixel detector is the closest sub-detector to the interaction point and provides the finest granularity. Comprised of over 80 million channels, the Pixel detector provides on average three measurements per track and has an intrinsic resolution of 10 μm (115 μm) in the Rϕ (Z) plane. The SCT is made of silicon strips with ∼6 million channels that provide between four and nine measurements per track with an intrinsic resolution of 17 μm (580 μm) in the Rϕ (Z) plane. The TRT, the largest of the sub-detectors in the ID, is made of straw drift tubes which have a single hit resolution of 130μm in the Rϕ plane, and provides an average of 36 measurements per track. In total the TRT is comprised of nearly 350,000 read-out channels. As is typical in the design of tracking detectors, the ATLAS ID has been constructed and assembled in a modular fashion. The individual sub-detectors are first assembled in such a way that larger structures are comprised of smaller collections of detector elements. For example, individual silicon wafers are mounted in rows forming ladders, which are then combined to form modules, which are part of a barrel layer or endcap wheel. After their separate construction, the barrel and endcaps are then assembled together to complete the ID. At each stage in this assembly procedure optical and mechanical survey measurements are made to control the quality of the built detector. Depending on the stage of the assembly, the installation accuracy of the ATLAS ID varies from hundreds of microns to a few millimeters.

These proceedings introduce the problem of detector alignment, a commissioning challenge crucial for operating and understanding a tracking device such as the ATLAS ID. They go on to describe a method implemented in the ATLAS experiment, track-based alignment, that provides a means of meeting this challenge and present initial results from cosmic-ray data.

2. The Alignment Problem

The trajectories of charged particles are reconstructed from the signals recorded by the tracking detectors. The trajectory is a helix, and the parameters of this helix are determined by fitting a track to the particle positions determined from the signals in the tracking detectors. The track fit determines the track parameters, from which the position, direction, and momentum of the particle can be determined. This track reconstruction involves not just the locally measured hits associated to the track, but requires the combination of the local measurements *and* the assumptions about the relative positions of the detector elements making the measurements. The alignment problem is simply that the initial assumptions are wrong. That is, detector positions used in the reconstruction algorithms do not correspond to the actual relative positions of the installed detector. This is a problem because spatial misalignments can affect the determination of track parameters which, in turn, compromise physics results.

The alignment problem in the ATLAS ID is exacerbated by both the overall size and the high granularity of the tracking detectors. Each of the 1744 modules in the Pixel detector, 4088 modules in the SCT, and 176 modules in the TRT, need to be aligned in six degrees of freedom: three rigid translations locating the module and three rotations orienting it. Thus, there are over 35,000 total degrees of freedom (DoF), alignment parameters, in the ID alignment problem.[1] The size and modular design of the ID implies that different scales of possible misalignments need to be addressed. Large, order millimeter, relative sub-detector misalignments and misalignment of the barrel layers with respect to endcap wheels, are expected from survey measurements. Misalignments on this scale will have the largest impact on physics and will significantly reduce the track reconstruction efficiency. Internal sub-detector misalignments are considerably smaller than relative sub-detector misalignments, but need to be corrected to achieve the ultimate detector performance. Internal misalignments in general require a much larger number of reconstructed tracks, to be handled properly, and are much more sensitive to possible biases discussed in the next section. The alignment objective in the ATLAS ID is to measure the relative position of the in-situ detectors with the precision that mis-alignment effects contribute less than 20% to the overall track parameter resolution[2]. This requirement, coupled with the small intrinsic resolution of the tracking devices, requires that the detector positions are determined to the accuracy of up to tens of microns, up to an order of magnitude more precise than survey measurements.

Various complementary methods of addressing the alignment problem have been implemented in the ATLAS ID; the remainder of these proceedings, however, will focus on one particular solution: track-based alignment. Track-based alignment has the advantages that it can easily be applied to all sub-detectors and scales of misalignment in the ATLAS ID and provides a means of gaining the required precision to meet the alignment objective. Track-based alignment also provides a flexible, extensible framework in which solutions, complementary to the basic track-based approach, can easily be included.

3. Track-Based Alignment

During track fitting, the degrees of freedom of a parametrized curve, a helix in the ATLAS

---

[1]Currently the TRT barrel modules (96 out of the 176) are only aligned in two rigid translations, not three.

ID, are fit to the spatial positions of a collection of measurements. These measurements are made by the individual tracking detector elements, either pixels in the Pixel detector, silicon strips in the SCT, drift or tubes in the TRT. Detector misalignments affect these spatial positions and will, in turn, affect the determination of the track parameters from the track fit. The track-based alignment exploits this interdependence in order to correct these very misalignments. The crucial step in allowing for the recovery of the misalignments is the definition of a statistic sensitive to the alignment problem. That is, the definition of a statistic that is sensitive to *both* the measurements made by the individual tracking detector elements *and* the assumptions about the relative position of these elements. This statistic is the χ^2 defined in 3.1,

$$\chi^2 = \sum_{hits} \left(\frac{m_i - h_i(\vec{x})}{\sigma_i} \right)^2, \tag{3.1}$$

m and σ are the position and uncertainty of the measurement made by the detector elements; h is the value of the measurement predicted by the track fit and is a function of the track parameters $\vec{x}$; the sum is over the hits associated to the reconstructed tracks. χ^2 is a function of the detector element measurements explicitly and of the alignment parameters through the $\vec{x}$ dependence in the fitted track prediction, h.

The key property of the χ^2 statistic is that, thought of as a function of the alignment parameters, χ^2 has a minimum at the true value of the detector alignment. Any detector misalignment will cause the track fit to be pulled away from its correct value (correct meaning the fit if there was no misalignment or if the misalignment was properly accounted for). This compromised track fit will result in predictions of measurements, $h(\vec{x})$, that are systematically more discrepant from those actually measured, m, and thus will result in an increase in χ^2.

The solution to the true detector misalignment is given by 3.2, where $\vec{\alpha}$ represents the alignment parameters[2]

$$\frac{d\chi^2}{d\vec{\alpha}} = 0. \tag{3.2}$$

In general, $\vec{x}$, and thus $h(\vec{x})$, and thus χ^2, is a highly nonlinear function of the alignment parameters, α. We use a linear expansion and an iterative process:

$$\frac{d\chi^2(\alpha)}{d\alpha} \approx \frac{d\chi^2(\alpha_0)}{d\alpha} + \left. \frac{d^2\chi^2}{d\alpha^2} \right|_{\alpha_0} (\alpha - \alpha_0), \tag{3.3}$$

where α_0 are the values of the alignment parameters of the current iteration. This approximation is, in general, more accurate as α_0 approaches the true values, and thus allows for an iterative solution for the true misalignment, α, given by 3.4.

$$\alpha = \alpha_0 - \left(\left. \frac{d^2\chi^2}{d\alpha^2} \right|_{\alpha_0} \right)^{-1} \frac{d\chi^2(\alpha_0)}{d\alpha} \tag{3.4}$$

The α's really are vectors of the alignment parameters and $\left(\left. \frac{d^2\chi^2}{d\alpha^2} \right|_{\alpha_0} \right)^{-1}$ is the inverse of a *NxN* matrix, with *N* being the number of DoF. In ATLAS, with $N \sim 35,000$, this inversion is non-trivial both in terms of CPU and memory requirements.

[2]In the following the $\vec{\,}$'s will be dropped from $\vec{\alpha}$.

The determination of the detector misalignments is done by iteratively solving equation 3.4, which amounts to inverting a *NxN* matrix. In general, there are several methods of handling this matrix inversion. In ATLAS the two that are used primarily are diagonalization and a class of techniques know as fast solvers. In diagonalization the *NxN* matrix is decomposed into a singular value decomposition: a product of three matrices. The middle matrix is a diagonal matrix and the diagonal elements are the eigenvalues of the original *NxN* matrix. The outer two matrices are composed of the eigenvectors of this original matrix. After this diagonalization the solution is a trivial multiplication of matrices. Diagonalization also has the advantages that the uncertainties on the alignment parameters are directly available and that it provides a means of handling weak modes, which are discussed later. Diagonalization is CPU intensive and can only be used when the number of alignment parameters is small (< 1000) as when aligning large structures or sub-detectors individually. When aligning the full detector, with its 35K DoF, fast-solver techniques must be used. Instead of calculating the inverse to the *NxN* matrix, and thus the alignment solution, directly, the fast-solver techniques minimize the distance to the solution, d, defined in 3.5.

$$d \equiv \left| \frac{d^2 \chi^2}{d\alpha^2} (\alpha - \alpha_0) + \frac{d\chi^2}{d\alpha} \right| \tag{3.5}$$

This solution is iterative, only approximate after a finite number of iterations, and, because the matrix inversion is not done, does not provide the uncertainties on the alignment parameters. These techniques also typically exploit unique properties of the large matrix to be inverted. The $\left(\left. \frac{d^2 \chi^2}{d\alpha^2} \right|_{\alpha_0} \right)^{-1}$ matrix is symmetric because it is a second derivative and it tends to be sparse because most alignment DoF are uncorrelated. These facts are exploited to gain both CPU and memory performance when dealing with the large number of DoF in the ATLAS ID.[3]

A major obstacle one has to confront when using the track-based alignment solution are weak modes. Weak modes correspond to detector deformations, either physical deformations in the in-situ detector, or in the alignment constants used for the track reconstruction, that result in little, if any, impact on the χ^2. These misalignments, by definition, are inherently problematic for the track-based method, which is based solely on χ^2. The $\frac{d\chi^2}{d\alpha} = 0$ solution is blind to classes of misalignments that result in solutions that are local minima in χ^2. Thus, track-based alignment will not be able to correct detector misalignments that result in weak modes. Moreover, these detector distortions are physically important in that they leave the χ^2 unchanged by biasing the reconstructed track parameters; the determination of these track parameters is the ultimate goal of the track reconstruction. There is a whole class of such distortions that plague detectors with cylindrical symmetry, such as the ID. An example, the clocking effect, is shown in Figure1. This coherent misalignment results in real tracks, indicated by the solid arrows, which are systematically reconstructed with incorrect curvature, represented as the the dashed arrows. The consequence is a biasing of the reconstructed component of the momentum transverse to the z-axis (p_T), which is dependent on the charge of the track and is larger for tracks with higher p_T. The effects of weak modes, such as that seen in the clocking effect, are imperceptible to the track-based alignment and can have a profound impact on the physics results.

---

[3]Practically, for a large number of DoF ($> 1,000$), the fast-solver techniques can only be used if the matrix is significantly sparse.
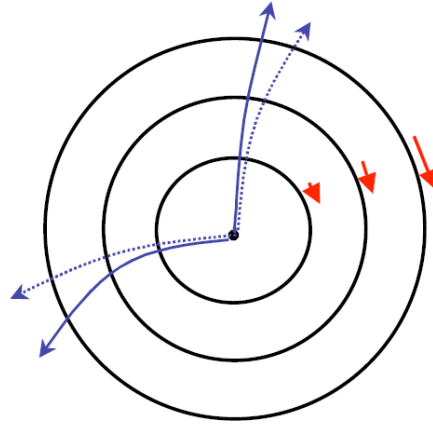
Figure 1: The Clocking Effect, an example of a weak mode giving rise to a p_T dependent p_T biasing of reconstructed tracks. Here, the deformation is such that detector elements are misaligned in ϕ by an amount which depends on r, as indicated by the red arrows.

The track-based alignment solution rests on inverting the *NxN* matrix, due to the weak modes however, this solution need not correspond to the actual detector alignment. In many ways, the handling of the weak modes is the *real* alignment problem, in that, once the χ^2 is defined and a means of dealing with the matrix is decided upon, the solution is straightforward. On the other hand, ensuring that the resulting solution is weak mode free is, in general, a much harder problem. As mentioned above diagonalization provides of means of coping with weak modes. Weak modes are particular detector deformations and can be thought of as a list of values for the alignment parameters. This list of alignment parameters defines a vector in the *N*-dimensional solution space. If this vector corresponds to a weak mode, it will be an eigenvector of the matrix being diagonalized.[4] Furthermore, the corresponding eigenvector will have an eigenvalue near zero.[5] When the matrix is diagonalized one has direct access to the eigenvectors and the corresponding eigenvalues. Eigenvalues near zero,[6] indicate that the corresponding eigenvector is a weak mode. If $\left(\left. \frac{d^2\chi^2}{d\alpha^2} \right|_{\alpha_0} \right)^{-1}$ has an eigenmode with eigenvalue zero, the alignment solution is degenerate. Weak modes correspond to eigenmodes that are nearly degenerate. While the linear, iterative solution has a hard time distinguishing a weak mode from the real alignment solution, the physics that results from the two detector alignments can be dramatically different. Although, through diagonalizing, the presence of weak modes is fairly straightforward, removing them is a much harder and less direct problem.

In general there are three methods of eliminating weak modes, each of which are implemented in the ATLAS ID alignment. The first is to simply ignore eigenmodes with eigenvalues below a certain threshold. The six DoF corresponding to rigid movements of the entire ID are unconstrained by the track-based alignment. Although this method is often used for eliminating these

---

[4]The basis that diagonalizes the matrix is the one in which the uncertainties on the alignment parameters are uncorrelated. The weak modes are those basis vectors in which the uncertainties are infinite. See [3]

[5]The actual values of the eigenvalues are arbitrary in that the can all be scaled by an arbitrary factor. However, the ratios of the eigenvalues have significance.

[6]Again, the definition of "near zero" is somewhat arbitrary, typically one looks for eigenvalues that are orders of magnitude smaller than the majority.

trivial, global DoF, it can be dangerous: the cutoff that defines the eigenmodes that will be ignored is arbitrary, and ignoring detector movements in these directions limits the ultimate control one has in the alignment.[7] This method is also not generally applicable, diagonalization is needed to get the eigenvalues and cannot be done at the highest granularity, which also happens to be the most susceptible to weak modes. Another potential means of eliminating weak modes is through enhancing the definition of χ^2. χ^2 as defined in 3.1 is only dependent on position discrepancies, measured and reconstructed. For a mode to be weak track parameters in the fit are required to be biased. By adding terms to the χ^2 that are dependent on these track parameters, one gains sensitivity to weak modes. Through minimizing the enhanced χ^2 function, as presented above, the true detector alignment can be recovered. This method is complicated by the fact that it requires multiple measurements of the track parameters used in the χ^2. There are several ways in which this can be done, fitting track segments in different sub-detectors, using calorimeter information, or exploiting known decay kinematic properties. However, they all require that the detectors making the other track parameter measurements are well understood, for example: the knowledge that these detectors do not contain a similar potential weak mode. The final, and most promising method of eliminating potential weak modes is in combining events with various track topologies. The χ^2 landscape is highly dependent on event properties: track origin, track direction, number of detectors crossed. These different types of events will lead to different types weak modes. Combining the information, the χ^2, from these different events allows the inherent ambiguity of the individual weak modes to be resolved. Beam collisions, single beam events, and cosmic rays, naturally provide events with a wide range of track topologies. Although there is as of yet no indication of the presence of weak modes in the ATLAS ID, there is a large effort focused on being able to detect, diagnose, and eliminate their presence in multiple, independent ways.
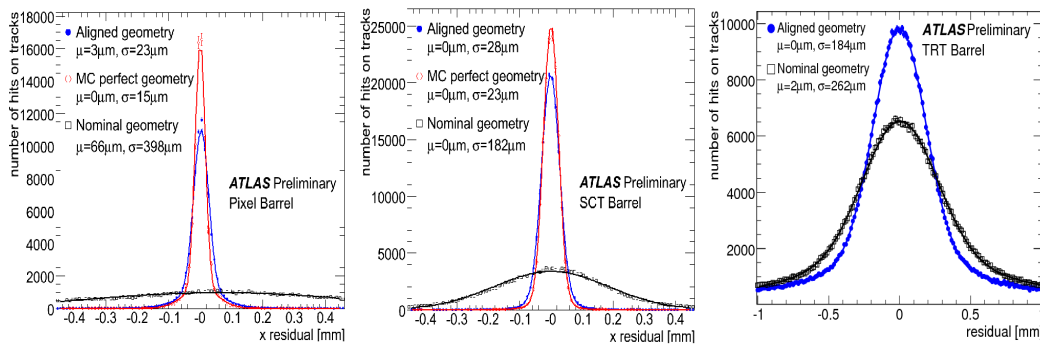
4. Results

In the fall of 2008, ATLAS held a dedicated cosmic-ray data taking period, during which over two million reconstructed tracks in the ID were recorded. This period proved extremely useful for the ID alignment effort. These cosmic events provided the first data for which the alignment algorithms were commissioned, allowed for the removal of large subsystem misalignments, and will be crucial for the ultimate understanding of weak modes. The track-based alignment algorithm presented in these proceedings was performed at various levels and in several steps. As a first step the Pixel barrel was aligned with the SCT barrel. This resulting alignment was used to align modules in both the Pixel and SCT barrel detectors internally, however, due to statistical limitations, not to their highest granularity. Due to the endcap geometry and the cosmic-ray spectrum, relatively few cosmic-rays are reconstructed in the detector endcaps, and thus an endcap alignment was not performed. Finally, the resulting alignment was used to align the TRT barrel and endcaps with respect to the silicon detectors. The full diagonalization of the χ^2 matrix was possible because the alignment did not include the highest granularity and not all of the ID was aligned.

---

[7]One can condition the matrix before diagonalizing it in such a way that the magnitude of the eigenvalues have a statistical interpretation. A cutoff can then be defined such that DoF in which there is not enough data to determine the alignment parameter to a specified accuracy are eliminated. See [3]

The alignment parameters converged with iteration, resulting in values consistent with survey measurements, while the number of reconstructed tracks and measured hits on track increased as a result of the alignment. The improvements in the residual distributions defined as $m_i - h_i(\vec{x})$, the difference between the measured and reconstructed track position, for the respective barrel detectors can be seen in Figure 2. The results before the alignment, in black, are shown with the results after the alignment, in blue. Here it is seen that the alignment enhances the peak of the residual distributions for each sub-detector, with improvements in the width of up to 350 μm. For Pixel and SCT residual distributions of the MC perfect geometry results are also shown in red. These are the results of simulated cosmic-ray events that are reconstructed using a simulated response of the ID. There are no effects of mis-alignment in the red distributions that are shown, which represent the design resolution of the Pixel and SCT. For the TRT the design resolution is 130 μm and the quality of the residual before alignment reflects the accuracy with which the TRT was built and installed. Nevertheless, there remains a considerable amount to be done before the ultimate alignment accuracy and the designed detector resolutions are reached. The alignment of the ATLAS ID has only begun, however promising early results provide strong encouragement both in terms of validating the overall track-based alignment procedure and in terms of understanding potential problems likely to arise when running the alignment with first collision data.

Figure 2: Residual distributions for the Pixel detector (left) (residual in R-ϕ plane), SCT (center), and TRT (right) barrel detectors. The black distributions are the residuals before the alignment, whereas blue is the resulting distributions after the alignment. For the Pixel and SCT, the residual distributions of cosmic simulation with an ideally aligned detector are shown in red. Tracks are selected to have $p_T > 2$ GeV, and go through the first Pixel layer.

5. Conclusions

The overall scale and ultimate precision of the ATLAS Inner Detector poses a challenging problem in terms of understanding the detector. Alignment is crucial in order to reach the full physics potential and detector performance. Track-based alignment provides a means of addressing the alignment problem, has been implemented in ATLAS, and has been successfully tested on early data. Cosmic-ray data has, and will continue to, provide a strong starting point from which the Inner Detector alignment can be performed, as well as guiding the way through the weak modes to the ultimate systematic free alignment. As the commissioning of the LHC and its associated

experiments rapidly progresses, ATLAS is poised for meeting the ID alignment challenge with the first collisions.

References

[1] The ATLAS Collaboration, *The Atlas Experiment at the CERN Large Hadron Collider*, 2008 JINST 3 S08003.

[2] The ATLAS Collaboration, *Inner Detector: Technical Design Report*, CERN/LHCC/97-016/017 (1997).

[3] A. Bocci and W. Hulsbergen. *TRT Alignment for the SR1 Cosmics and Beyond*, ATL-INDET-PUB-2007-009, 2007.
http://cdsweb.cern.ch/record/1039585/files/indet-pub-2007-009.pdf