Particle Tracking Machine Learning Challenge: Detector and Dataset

David Rousseau[1], Sabrina Amrouche[2], Paolo Calafiura[3], Steven Farrell[3], Cécile Germain[4], Vladimir Vava Gligorov[5], Tobias Golling[2], Heather Gray[3], Isabelle Guyon[4,11], Mikhail Hushchyn[6], Julia Hrdinka[7,12], Vincenzo Innocente[7], Moritz Kiehn[2], Edward Moyse[8], Andreas Salzburger*[7], Andrey Ustyuzhanin[9], Jean-Roch Vlimant[10], and Yetkin Yilmaz[1]

[1]LAL, Orsay
[2]University of Geneva
[3]Lawrence Berkeley National Laboratory
[4]UPSud and INRIA, University Paris-Saclay
[5]LPNHE-Paris
[6]Yandex, MIPT
[7]CERN
[8]University of Massachusettes
[9]Yandex, HSE
[10]California Institute of Technology
[11]Chalearn
[12]TU Vienna

Abstract

This document describes the layout of the detector and the structure of the dataset of the Particle Tracking Machine Learning dataset. It outlines the characteristics of the detector layout, the features of the measurements and describes the magnetic field. Furthermore, the composition of the generated physics events is described and the simulated interactions of the particles with the detector material explained. A detailed description of the dataset structure and format is provided.

Keywords

Particle Tracking Machine Learning Challenge, Detector, Dataset.

1 The Particle Tracking Machine Learning Challenge

The finding of trajectories of charged particles in the tracking detectors used in modern high energy physics experiments is one of the most complex tasks in event reconstruction. The widely used combinatorial approach of seed finding and track following (for the LHC experiments ATLAS and CMS described in [1] and [2], respectively) requires high CPU consumption due to the increasing event complexity foreseen in the future high luminosity program and beyond. Machine learning (ML) is a rapidly growing sector in academics and industry, which has large potential overlap and opportunities for application in track reconstruction. One such attempt to foster external expertise for high luminosity track reconstruction is the upcoming Particle Tracking Machine Learning Challenge (TrackML) [3].

This document describes the detector setup, the simulation and the structure of the TrackML dataset. The simulation has been done using the fast simulation extension of the ACTS tracking toolkit [4].

*corresponding author: Andreas.Salzburger@cern.ch

2 Detector Description

Particle detectors at hadron colliders such as the LHC are complicated multilayer structures that aim to identify the types and measure the kinematic properties of particles emerging from beam-beam collisions that take place in the center of the detector, see Sec. 3. The innermost detection devices are usually tracking detectors that are designed to measure the trajectories (and hence production origins) of charged particles. Neutral particles do not leave traces in the innermost detectors. In order to do so, the charged particles are localised with special detectors at discrete positions (layers) while they follow their path through the detector volume. A magnetic field bends the particle trajectories which allows to determine the particle momenta from the curvature, see Sec. 2.8. Tracking detectors are typically designed as follows: the closer to the beam-beam interaction region, the better is the resolution of the detection devices, as the estimation of the position where the particle emerges from (in the following referred to *vertex*) is one of the primary tasks of track reconstruction.

Modern particle detectors are often realised as planar solid state detectors, and use silicon wafers as the main detection devices. The principle of such planar silicon detectors is relatively simple: a charged particle, when traversing a Silicon module, induces ionisation charge in the Silicon structure. This charge is then read out by a segmented readout chip. By finding single channels or *clusters* of channels that have collected charge on such a Silicon wafer, a measurement of the particle position is done. Higher readout granularity, i.e. smaller structures of channels on the Silicon wafer hereby translates into a better measurement resolution. Figure 1 shows a simplified view of the detector, indicating the layer positions.

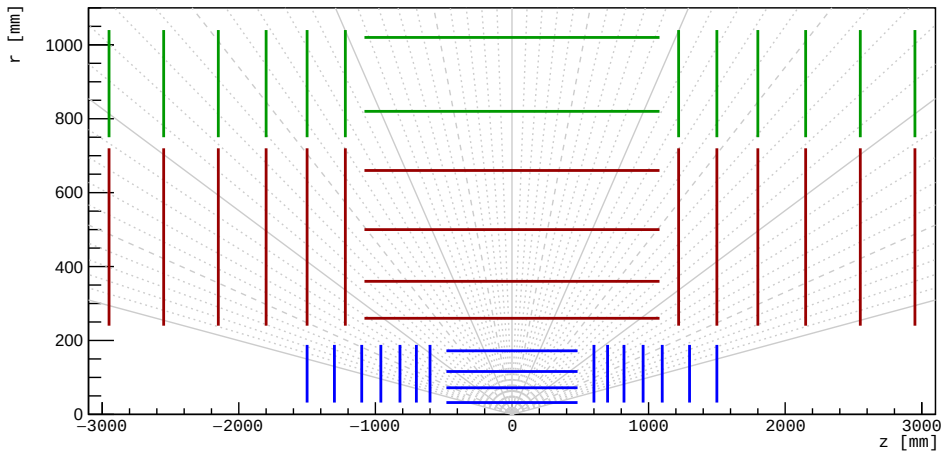

Fig. 1: Illustration of the TrackML detector: the different detector technologies are shows with different colors. Innermost pixel detector (blue), followed by a short (red) and long strip (green) detector.

Central cylindrical barrel structures and so-called *end-cap* disc structures are built from planar modules (mostly rectangular, or of trapezoidal shape in the end-caps). A certain overlap of the modules is needed to ensure hermetic coverage of the full phase space. Figure 2 shows a transverse view of the innermost barrel structures showing the module overlap in the transverse direction, which is introduced to guarantee a hermetic coverage of the detection region. In addition, overlap in the longitudinal direction is given for the barrel region, and radially overlapping modules on the end-cap discs are present. This means that a single particle can indeed produce several measurements on one single layer, while several measurements on one single module is impossible.

Figure 3 shows an overview of on the average number of measurements in the sub detectors (described in Secs. 2.3 to 2.4.2) and the average total number of measurements for particles that traverse the

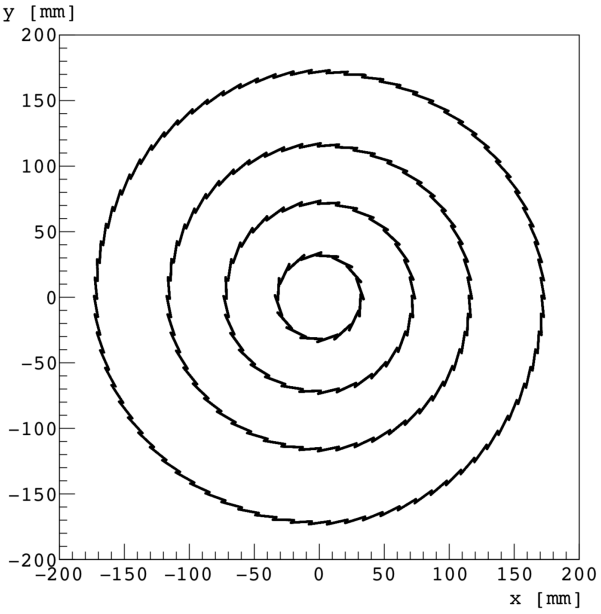

Fig. 2: Transverse view onto the pixel detector showing the four barrel layers with their overlapping modules in the transverse plane.

entire detector. Particle production vertices at larger radii, and nuclear interaction process (described in Sec. 2.7.3) generally reduce the number of measurements produced per particle.

2.1 Identification schema

The barrel and end-cap volumes are uniquely identified with volume identifiers (`volume_id`). Within a single volume, the individual layers are uniquely label with a layer identifier (`layer_id`), which for technical reasons only takes even values. Modules within a layer are uniquely labeled with a module identifier (`module_id`), which runs from 1 to N, when N denotes the number of modules on that layer.

2.2 Axes and angle definitions

The coordinate system is a right-handed cartesian coordinate system (x, y, z) with the global z axis defined along the beam direction, i.e. the ML detector builds a cylindrically shaped object around the global z axis. The $(x-y)$ plane is referred to as the transverse plane, and the azimuthal angle $\phi \in [0, 2\pi)$ is defined in the transverse plane with $\phi = 0$ denoting the x-axis. The polar angle θ is measured from the z-axis and is defined to be within $[0, \pi]$. In particle physics θ is often replaced by the so-called *pseudo-rapidity* η, where η is defined as

$$\eta = -ln(\tan(\frac{\theta}{2})). \tag{1}$$

2.3 Pixel detector

The innermost detector is a Silicon pixel detector. It is built from four cylinders that resemble a barrel system and 7 discs on each side complete the hermetic coverage of the phase space. A three dimensional illustration of the pixel detector is shown in Fig. 4.

The two local axes of the silicon sensor are denoted with l_0 and l_1, respectively. The pixel sizes are also referred to *pitch* sizes, and are denoted as p_0 and p_1. For the TrackML pixel detector, roughly

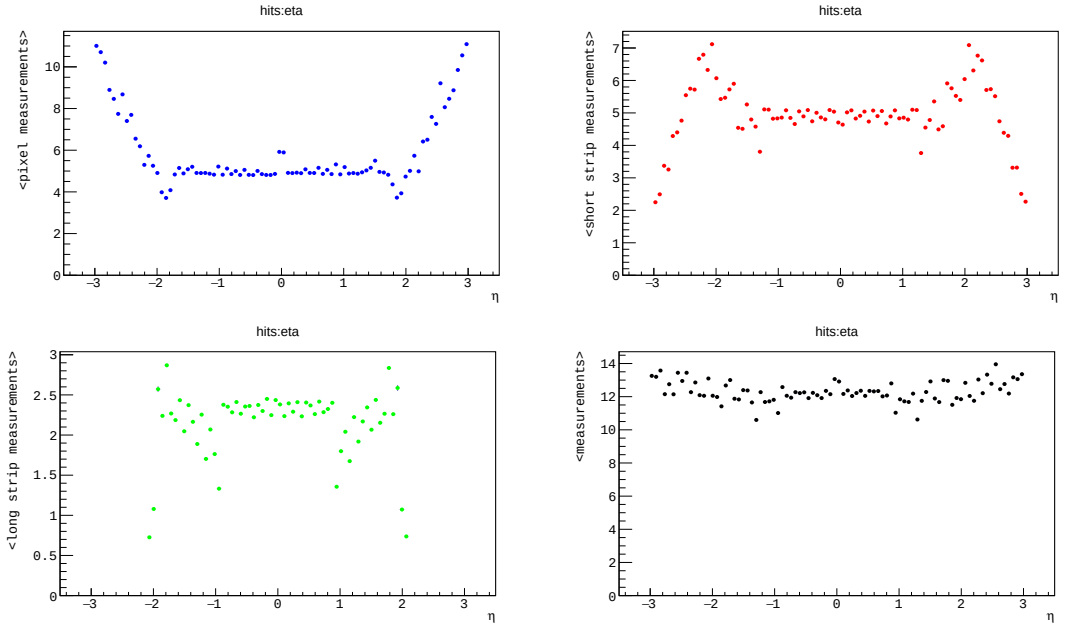

Fig. 3: Average number of measurements in the pixel (top left), short strip (top right) and long strip detector (bottom left) versus pseudo-rapidity. Bottom right shows the average total number of measurements for particles from the interaction region and traversing the entire TrackML detector.

squared pixels of $p_0 \times p_1 = 50\mu\text{m} \times 56.25\mu\text{m}$ have been chosen. The individual pixels are indexed with a pair of channel identifiers running from 0 to n along l_0, and 0 to m along l_1, see Fig. 7.

Cluster shapes and measurement accuracy

If only one pixel is traversed by the particle, say the pixel with cell index pair (i, j), the center position of the pixel, called $(c_0, c_1)_{ij}$ is taken as the measurement position, denoted as $(m_0, m_1) = (l_0^{rec}, l_1^{rec})$, where *rec* indicates that the position was *reconstructed* from the cell position. The difference between the true intersection of the particle with the silicon sensor's mid plane, $(t_0, t_1) = (l_0^{true}, l_1^{true})$, and the measurement position is defined as the two dimensional residual

$$r_0 = m_0 - t_0 = l_0^{rec} - l_0^{true}, \tag{2}$$

and

$$r_1 = m_1 - t_1 = l_1^{rec} - l_1^{true}, \tag{3}$$

respectively. In the case of a single pixel measurement, the residual distribution is basically a box distribution between $[\frac{-p_i}{2}, \frac{p_i}{2}]$, because all possible intersections between the pixel borders would end up being assigned to the pixel center[1]. Figure 5 shows the residual distribution in the outermost pixel disk where due to geometric constraints only clusters/measurements of size 1 are present.

Due to the incident angle of the particle on the silicon sensor, typically more than one pixel is traversed, see Fig. 6 which shows the cluster size distributions for clusters in the innermost pixel barrel layer. The shallow incident angle in the forward direction hereby creates very large cluster sizes in the l_1 direction.

[1] In reality, geometrical constraints shape the box distribution slightly: given a certain detector geometry particles from the interaction point can only have restricted incidence angles into the silicon sensor, which in turn restricts the possible mid plane intersections and henceforth residual.

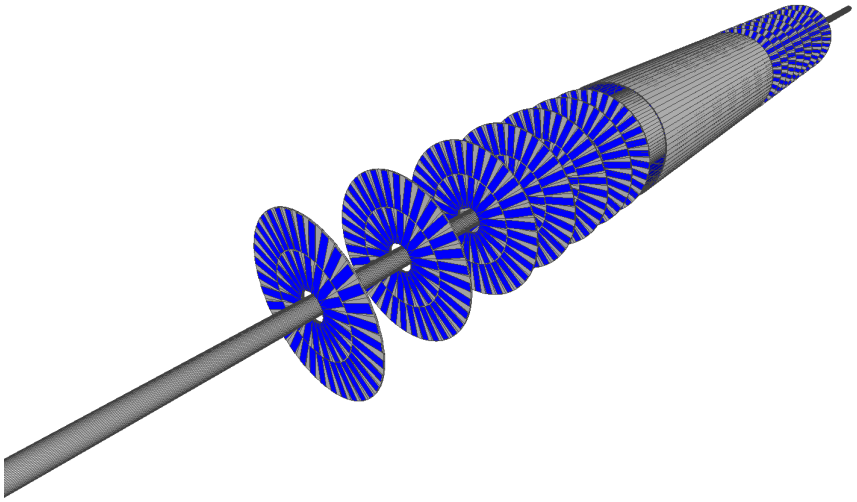

Fig. 4: 3D model of the TrackML detector showing the beam pipe and the pixel detector. The pixel detector is entirely built from rectangular Silicon pixel modules.

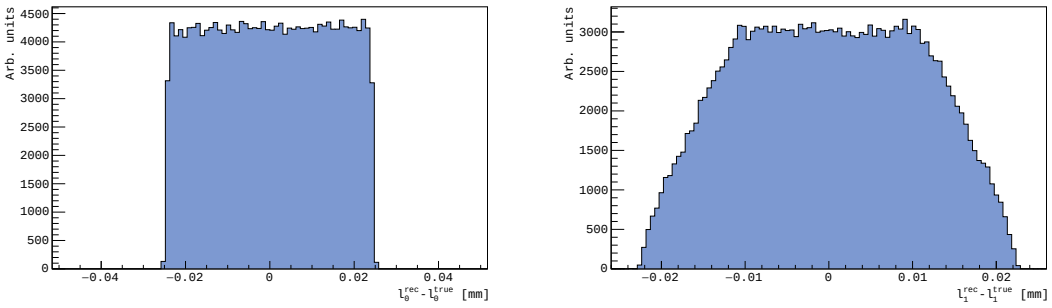

Fig. 5: Examples for residual distribution in the outermost disc of the Pixel detector endcap. Due to geometric constraints only clusters with a width of 1 are possible in the l_0 direction, leading to a perfect box distribution. In the l_1 coordinate cluster sizes of 1 and 2 are present, leading to a slightly improved resolution.

When several pixels in a module are traversed by the particle, the measurement position is reconstructed from all involved pixels that are grouped together. The readout electronics does not only respond if a pixel was traversed, but also how much charge was induced by that particle. The deposited charge is roughly linear to the path length in the pixel, and hence this information can be used to refine the reconstructed cluster position by building a weighted mean position:

$$(m_0, m_1) = \frac{1}{Q_i} \sum_i q_i (c_0, c_1)_i \tag{4}$$

where

$$Q_i = \sum_i q_i, \tag{5}$$

and q_i is the individual read out charge value of pixel i. This method is in the following referred to

Table 1: Volume and layer information and identification for the Pixel Detector

Negative EC, `volume_id` **= 7**	**average z position [mm]**	`layer_id`	**modules (ring 1, ring 2)**
	-1500	2	40, 68
	-1300	4	40, 68
	-1100	6	40, 68
	-960	8	40, 68
	-820	10	40, 68
	-700	12	40, 68
	-600	14	40, 68
Barrel, `volume_id` **= 8**	**average radius r [mm]**	`layer_id`	**modules (ϕ x z)**
	32	2	16 x 14
	72	4	32 x 14
	116	6	52 x 14
	172	8	78 x 14
Positive EC, `volume_id` **= 9**	**average z position [mm]**	`layer_id`	**modules (ring 1, ring 2)**
	600	2	40, 68
	700	4	40, 68
	820	6	40, 68
	960	8	40, 68
	1100	10	40, 68
	1300	12	40, 68
	1500	14	40, 68

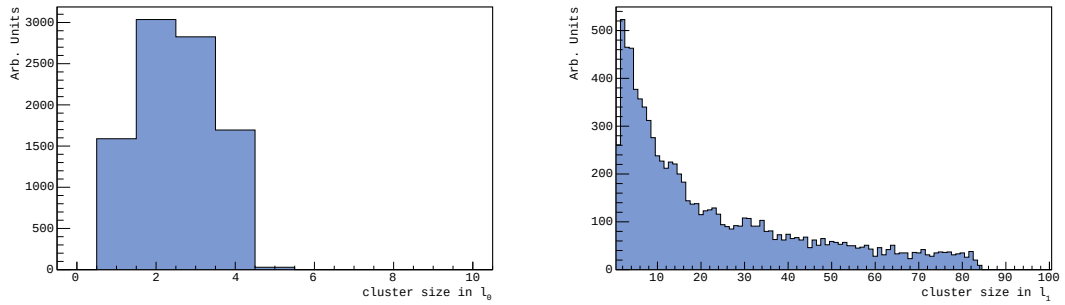

Fig. 6: Cluster size distribution in the innermost pixel barrel layer in l_0 and l_1, respectively. The shallow incident angle into the pixel modules in the outer z regions of this barrel detector results in large cluster sizes in l_1.

analog clustering.

Figure 7 illustrates the reconstruction of a pixel cluster center from the individual pixel centers. Figure 8 shows residual distributions for several cluster sizes. As a rule of thumb, bigger clusters estimate the true position better, however, such clusters are also often prone to be at the edge of modules or interrupted by inefficient pixels that introduce significant mis-measurements. The charge can not be read out with unlimited accuracy, to emulate a readout uncertainty, the path length in the pixel is smeared by 10 % following a normal distribution.

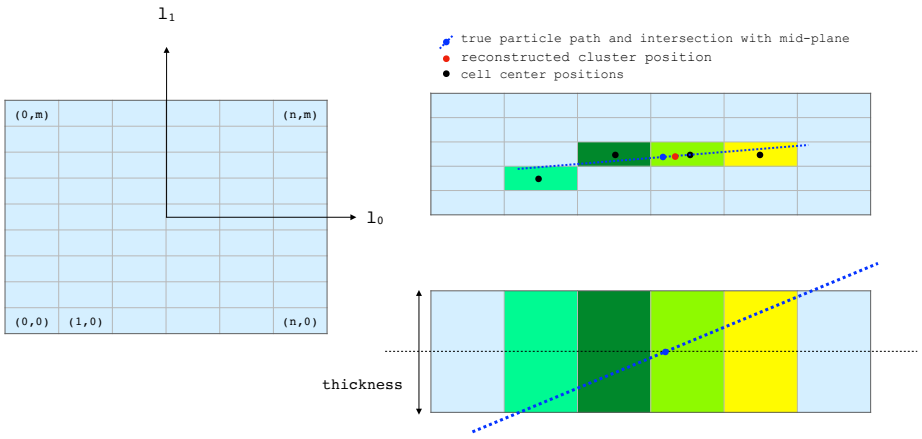

Fig. 7: Illustration of the clusterisation in the Pixel detector, the silicon module is shown from a bird eyes view. The particle traverses the silicon and crosses several pixels. The individual path lengths in the pixels corresponds to the read out charge. The true intersection can then be reconstructed by as the weighted position of the read out cells. If a the path length in a cells falls below a certain threshold, the cell is not read out.

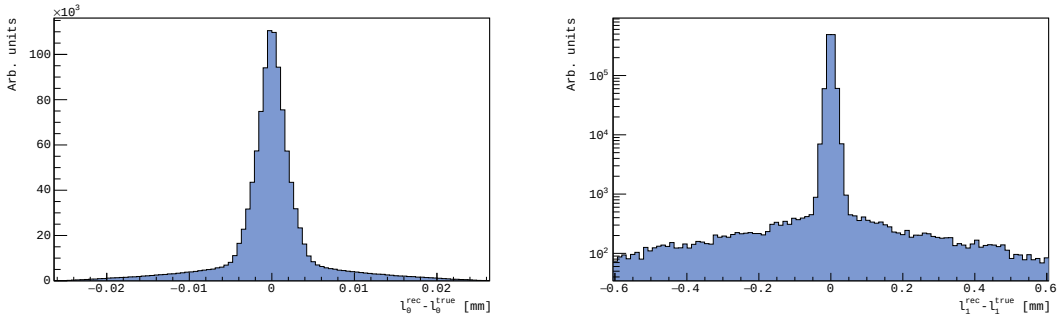

Fig. 8: Residual distributions for the innermost barrel layer of the pixel detector, where several different cluster sizes are present.

Directional information

The cluster shape in the pixel detector carries some limited information about the particle direction at intersection, see Fig. 7. The local incident angle of the particle into the silicon module can be calculated with a limited accuracy from its cluster shape and size in the l_0 and l_1 direction. This can be translated into a global angle using the modules relative rotation matrix with respect to the global frame.

2.4 Strip detector

The particle density per detection area decreases rapidly with the distance to the interaction point. The particle localisation at outer radii is mainly needed for the curvature measurement (contrary to the innermost measurements that are needed for the particle origin measurement). This allows outer tracking detectors to have a reduced granularity, which directly translates in bigger measurement structures. Silicon pixel detectors are thus often followed by Silicon strip detectors that tend to measure only the coordinate related to the particle bending with a high precision, while reducing the precision in the non-bending plane, see Sec. 2.8. The lower granularity reduces the complexity of the detector by vastly

reducing the number of necessary readout channels, which in turn translates often to less material such as cables, readout chips and cooling support needed. Additionally, since the highest possible precision is not required in the outer strip detector, the readout is limited to a digital readout system, i.e. a strip either receives a signal from a traversing particle or not and no charge information is present. This simplifies equation Eq. 4 to

$$(m_0, m_1) = \frac{1}{N} \sum_{i=1}^{N} (c_0, c_1)_i, \tag{6}$$

to the arithemtic mean of the individual cell centers, and is consequently called *digitial clustering*. Again, however, a minimum path length of the particle in the silicon is required for the strip cell to be above threshold in order to be read out. The actual path length is again smeared by an average of 10 $\%$ with a gauss distribution to emulate a readout uncertainty.

2.4.1 Short Strip detector

The pixel detector is surrounded by a short strip detector with four barrel layers and six endcap discs on each side. The strip pitch p_0 has been chosen to be $80\mu m$, while the strip length is given with 1.2mm, yielding in strongly reduced resolution in the longitudinal direction. Due to the larger segmentation, cluster sizes are basically limited to 1 or 2, and digitial clustering is used, i.e. no charge information is used to reconstruct the cluster position. Figure 9 shows a three-dimensional model of the short strip detector, and Fig. 12 shows typical residual distributions for the strip detector measurements.

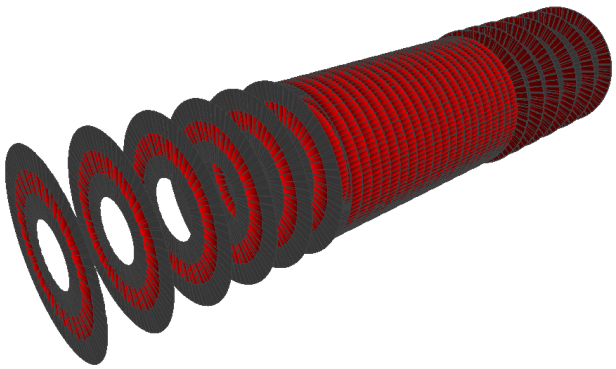

Fig. 9: 3D model of the TrackML detector showing only the short strip section of the detector.

2.4.2 Long Strip detector

The outermost detector device has the coarsest granularity of the Tracking ML detector, it is a long strip detector that builds the outermost 2 barrel layers and the outermost 2 rings of the end-cap disks. The

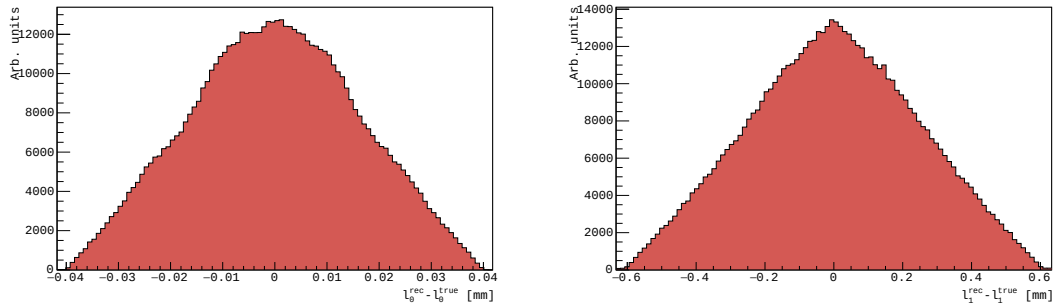

Fig. 10: Residual distributions for the strip detector, innermost barrel layer.

Table 2: Volume and layer information and identification for the short strip detector

Negative EC, `volume_id` **= 12**	**average z position [mm]**	`layer_id`	**modules (ring 1, 2, 3)**
	-2950.	2	52, 56, 60
	-2550.	4	52, 56, 60
	-2150.	6	52, 56, 60
	-1800.	8	52, 56, 60
	-1500.	10	52, 56, 60
	-1220.	12	52, 56, 60
Barrel, `volume_id` **= 13**	**average radius r [mm]**	`layer_id`	**modules (ϕ x z)**
	260.	2	40 x 21
	360.	4	56 x 21
	500.	6	78 x 21
	660.	8	102 x 21
Positive EC, `volume_id` **= 14**	**average z position [mm]**	`layer_id`	**modules (ring 1, 2, 3)**
	1220.	2	52, 56, 60
	1500.	4	52, 56, 60
	1800.	6	52, 56, 60
	2150.	8	52, 56, 60
	2550.	10	52, 56, 60
	2950.	12	52, 56, 60

strip pitch is chosen to be $p_0 = 120\ \mu\text{m}$ and the strip length $p_1 = 10.8\ \text{mm}$.

2.5 Detector inefficiency, spurious measurements and merged measurements

Particle detectors can not be 100 % efficient in detecting particles. There are many reasons for a particle to not leave a measurement in a Silicon wafer, a description of such is outside the scope of this measurement. To emulate inefficiencies, however, the dataset contains a random inefficiency of 1%. Also, some spurious measurements are injected in the dataset. In reality, these can either be from electronic noise, or - more prominently - from low energetic particles that emerge from detector interactions for the particles coming from the initial collision, but are either below detection threshold of the detector or simply left-over products of prior collisions and thus not interesting for the actual trajectory finding. In the presented dataset, those measurement represent about 10 % of all measurements per event and are indicated by a missing link to a truth particle, see Sec. 4. In reality, when particles are very close-by on

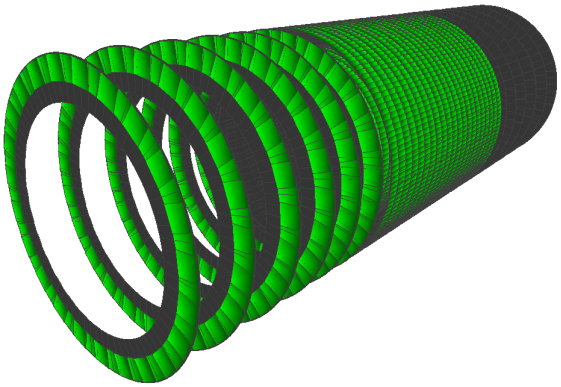

Fig. 11: 3D model of the TrackML detector showing only the long strip section of the detector.

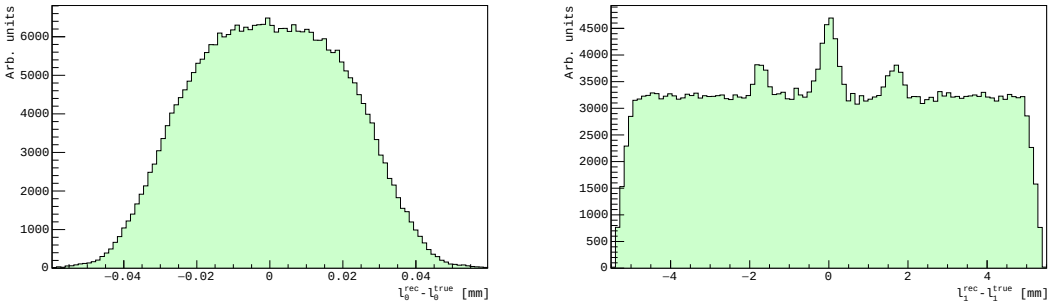

Fig. 12: Residual distributions for the strip detector, innermost barrel layer.

detection wafers, there is a real chance that the individual clusters ares merged and a unique association to a particle is not possible anymore. However, in aspect of simplicity of the dataset, this cluster merging is not applied and in such a case two or more clusters with (partly) overlapping features are present in the dataset.

2.6 Material budget and consequences

A real particle detector is a very complicated object involving many components: active detector elements for particle detection, mounting and support structures, cooling systems to maintain the optimal operation temperature, readout electronics, and power and data link cables to operate the detector while data taken. All those components have significant material which disturbs the actual measurement and particles that traverse the detector undergo interaction with the detector material. In the construction, tracking detectors are carefully designed to minimize the material budget in order to minimize the influence on the particle. The material is commonly quantified by the so called electromagnetic X_0 and nuclear interaction length Λ_0, as they can be related to the strength and probability of certain interactions

Table 3: Volume and layer information and identification for the long strip detector

Negative EC, `volume_id` **= 16**	**average z position [mm]**	`layer_id`	**modules (ring 1, 2)**
	-2950.	2	48,50
	-2550.	4	48, 50
	-2150.	6	48, 50
	-1800.	8	48, 50
	-1500.	12	48, 50
	-1220.	12	48, 50
Barrel, `volume_id` **= 17**	**average radius r [mm]**	**layer_id**	**modules (ϕ x z)**
	820.	2	120 x 21
	1020.	4	152 x 21
Positive EC, `volume_id` **= 18**	**average z position [mm]**	**layer_id**	**modules (ring 1, 2)**
	1220.	2	48, 50
	1500.	4	48, 50
	1800.	6	48, 50
	2150.	8	48, 50
	2550.	10	48, 50
	2950.	12	48, 50

of the particle to occur with the detector material. The amount of material is then given by the path length (or thickness) of the particle in units of either the electromagnetic or nuclear interaction length, respectively. Roughly, each layer in the TrackML detector has a 2-3 % of thickness in terms of electromagnetic interaction, and less than 1 % in terms of nuclear interaction length. The incident angle of the particle into the layer acts as a geometrically multiplying factor to the material *seen* by the particle. Figure 13 shows the distribution of the material along the pseudorapidity η.

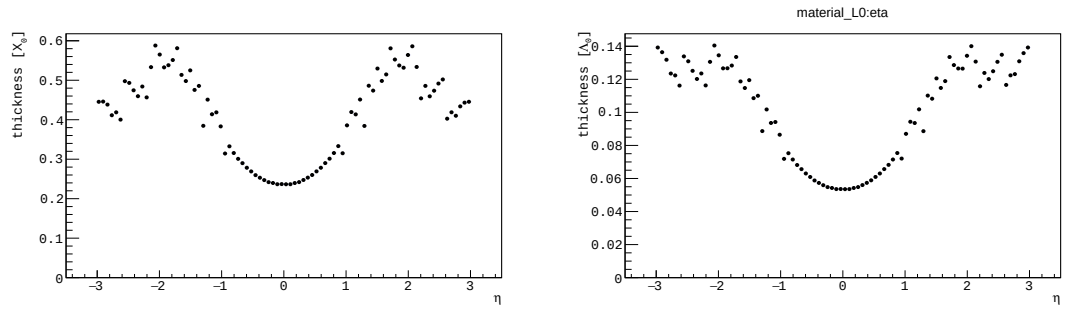

Fig. 13: Material distribution in the detector in terms of radiation length X_0 (left) and Λ_0 (right). The maximum material is in the transition region between barrel and end-cap sectors, where the incident angle of the particle into the detector material gives contributes most.

For charged particles, three different processes describing their interaction with matter are taken into account for this dataset, and are described in the following.

2.7 Particle interaction with the detector

A variety of interactions can occur when a charged particle traverses the detector, mainly caused by the detector material. A detailed description can be found in [5], and only an overview of the applied

interactions in the presented dataset is given here.

2.7.1 Multiple Coulomb Scattering

When a charged particle traverse dense matter, multiple Coulomb Scattering interactions appear, resulting in a net deflection of the original particle direction. Each individual scattering process is random and usually leads to only a small deflection from the initial particle direction. Following the central limit theorem, the sum of many small deflections results in a stochastic deflection following a normal (gauss) distribution with mean 0 and a certain width. The width is given by the amount of the traversed material and the momentum of the particle. The dependency is rather intuitive: the more material is traversed the broader is the resulting spread. On the other hand, the *faster* the particle moves, the smaller the spread.

2.7.2 Energy loss

Particles that traverse material lose part of the energy, mainly due to two processes: *ionisation loss* and *radiation loss*. The latter is only relevant for particles with small particle masses, which build only a small part of the particles in this dataset, see Sec. 3.1.1. Ionisation loss is a relatively small effect, and the net result on the particle energy after traversing the inner tracker stays often well beyond 1 %. Caused by their low mass compared to all other particles, electrons actually tend to lose a lot of energy due to *bremsstrahlung*, however, for simplicity of the dataset this effect is not simulated and electrons only undergo ionisation loss.

2.7.3 Hadronic interactions

A type of charged particles, so-called *charged hadrons*, interact with the nuclei of the detector material such, that the initial particle is destroyed, leading to an emerging spray of outgoing particles, again predominantly hadrons. This process is very common, because the overwhelming majority of charged particles in the dataset are hadrons, see Sec. 3.1.1. The probability with which a hadronic interaction occurs is given by the amount of material expressed by its hadronic interaction length Λ_0. Products of hadronic interactions are evidently less energetic than the incoming or causing particle.

2.8 Magnetic Field Setup

Tracking detectors are embedded in a strong magnetic fields to measure the particle momentum. A charged particle, when moving with constant velocity through a constant magnetic field follows a helical trajectory, which is illustrated in Fig. 14. For inner tracking detectors, the field is usually (as much as possible) aligned with the beam direction, such that the particle is bent in the transverse plane.

The dataset is simulated with a non-ideal solenoidal magnetic field (i.e. the main component of the magnetic field vector points into the global z direction), with a central field strength of 2 Tesla.

A perfect constant magnetic field, however, is not possible to produce. For example the limited size of the solenoid coil will result in an inhomogeneous magnetic field, and hence in a trajectory that differs from a helical track.

3 Event Topology

3.1 Generated Events

Present and future experiments and proton colliders such as the LHC are operated such that multiple proton-proton collisions occur during a single beam-beam crossing, a feature that is often referred to as *pile-up*. This mode is chosen to enhance the probability of creating an *interesting* event to take place in the vast amount of *ordinary* events that had been studied to a large extend in the past. These interesting events are usually characterised by a higher energy transfer involved, which results in general to more and higher energetic particles to traverse the detector. Without loss of generality, the dataset is chosen to

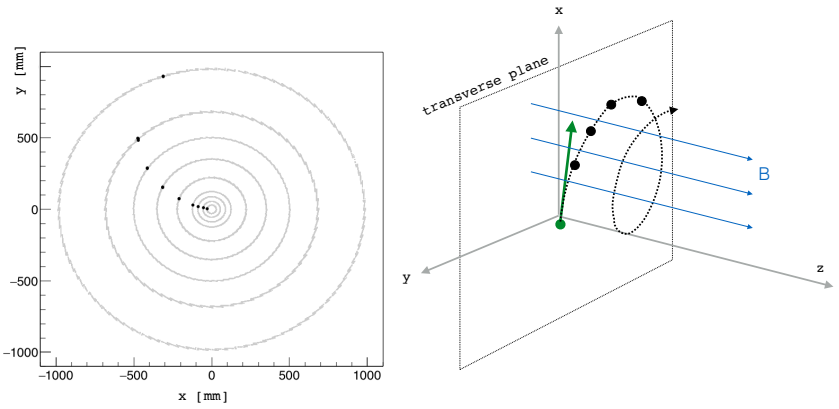

Fig. 14: Illustration of a particle moving through a constant magnetic field. In the transverse projection, this corresponds to the particle moving along a circle.

consist of one *interesting* event, overlaid with m lower energetic pile-up events, where m is drawn from a Poisson distribution with a mean of 200[2]. The individual collisions are randomly placed in small region, the so called *luminous region* around the nominal center position of the detector (0,0,0). The luminous region follows a three dimensional gauss distribution with transverse width $(\sigma_x, \sigma_y) = (15\mu\mathrm{m}, 15\mu\mathrm{m})$ and a longitudinal width of $\sigma_z = 55\mathrm{mm}$. In the following, the *interesting* collisions will be referred to as *hard scatter* (HS) and *pile up* (PU) collisions, following the usual nomenclature used in high energy physics. The events used in this dataset are generated with the PYTHIA8 [6] event generator. Gluon-gluon fusion to a top-antitop pair is chosen as the HS topology, the PU events are drawn as soft QCD events[3]. The center of mass energy of the proton beams at collision is chosen to be 14 TeV in order to resemble future HL-LHC conditions.

3.1.1 Kinematics and particle composition

Only stable charged particles from both the HS and PU collisions are further considered for this dataset, as only those leave traces in the tracking detector. Potential further decay of particles is not considered, only hadronic interaction with the detector material can alter the particle composition of the dataset, see Sec. 2.7.3.

4 Dataset structure in detail

4.1 Event hits

The hits file contains the following values for each hit/entry:

– `hit_id`: numerical identifier of the hit inside the event.

[2]This configuration resembles to a good approximation the future running conditions of the high luminosity extension of the LHC, the so-called *HL-LHC*.

[3]The corresponding PYTHIA8 configuration stings are `Top:gg2ttbar=On` and `SoftQCD=On`, respectively.

- `x, y, z`: measured x, y, z position (in millimeter) of the hit in global coordinates.
- `volume_id`: numerical identifier of the detector group.
- `layer_id`: numerical identifier of the detector layer inside the group.
- `module_id`: numerical identifier of the detector module inside the layer.

The volume/layer/module id could be in principle be deduced from x, y, z. They are given here to simplify detector-specific data handling.

4.2 Event hit cells

The cells file contains the constituent active detector cells that comprise each hit. The cells can be used to refine the hit to track association. A cell is the smallest granularity inside each detector module, much like a pixel on a screen, except that depending on the `volume_id` a cell can be a square or a long rectangle. It is identified by two channel identifiers that are unique within each detector module and encode the position, much like column/row numbers of a matrix. A cell can provide signal information that the detector module has recorded in addition to the position. Depending on the detector type only one of the channel identifiers is valid, e.g. for the strip detectors, and the value might have different resolution.

- `hit_id`: numerical identifier of the hit as defined in the hits file.
- `ch0, ch1`: channel identifier/coordinates unique within one module.
- `value`: signal value information, e.g. how much charge a particle has deposited.

4.3 Event particles

The particles files contains the following values for each particle/entry:

- `particle_id`: numerical identifier of the particle inside the event.
- `vx, vy, vz`: initial position or vertex (in millimeters) in global coordinates.
- `px, py, pz`: initial momentum (in GeV/c) along each global axis.
- `q`: particle charge (as multiple of the absolute electron charge).
- `nhits`: number of hits generated by this particle.

All entries contain the generated information or ground truth.

4.4 Event truth

The truth file contains the mapping between hits and generating particles and the true particle state at each measured hit. Each entry maps one hit to one particle/track.

- `hit_id`: numerical identifier of the hit as defined in the hits file.
- `particle_id`: numerical identifier of the generating particle as defined in the particles file. A value of 0 means that the hit did not originate from a reconstructible particle, but e.g. from detector noise.
- `tx, ty, tz` true intersection point in global coordinates (in millimeters) between the particle trajectory and the sensitive surface.
- `tpx, tpy, tpz` true particle momentum (in GeV/c) in the global coordinate system at the intersection point. The corresponding vector is tangent to the particle trajectory at the intersection point.

– textttweight per-hit weight used for the scoring metric; total sum of weights within one event equals to one. Dataset submission information

The dataset is structured in four different file classes, written out as comma separated value format. A single file describes the detector geometry: `detector.csv`.

The training dataset consists from four files per event, e.g.

– `event000000000-hits.csv`
– `event000000000-cells.csv`
– `event000000000-particles.csv`
– `event000000000-truth.csv`

4.5 Additional: Detector descriptions

The detector is built from silicon slabs (or modules of rectangular or trapezoidal shapes), arranged in cylinders and disks, which measure the position (or hits) of the particles that cross them. The detector modules are organized into detector groups or volumes identified by a volume id. Inside a volume they are further grouped into layers identified by a layer id. Each layer can contain an arbitrary number of detector modules, the smallest geometrically distinct detector object, each identified by a `module_id`. Within each group, detector modules are of the same type have e.g. the same granularity. All simulated detector modules are so-called semiconductor sensors that are build from thin silicon sensor chips. Each module can be represented by a two-dimensional, planar, bounded sensitive surface. These sensitive surfaces are subdivided into regular grids that define the detectors cells, the smallest granularity within the detector.

Each module has a different position and orientation described in the detectors file. A local, right-handed coordinate system is defined on each sensitive surface such that the first two coordinates u and v are on the sensitive surface and the third coordinate w is normal to the surface. The orientation and position are defined by the following transformation:
`pos_xyz = rotation_matrix * pos_uvw + offset`
that transform a position described in local coordinates u, v, w into the equivalent position x, y, z in global coordinates using a rotation matrix and an offset. The local coordinates u and v correspond to the readout coordinates l_0 and l_1 from Sec. 2, respectively, w points along the thickness of the module.

– `volume_id`: numerical identifier of the detector group.
– `layer_id`: numerical identifier of the detector layer inside the group.
– `module_id`: numerical identifier of the detector module inside the layer.
– `cx, cy, cz`: position of the local origin in the described in the global coordinate system (in millimeter).
– `rot_xu, rot_xv, rot_xw, rot_yu, ...`: components of the rotation matrix to rotate from local u, v, w to global x, y, z coordinates.
– `module_t`: thickness of the detector module (in millimeter).
– `module_minhu, module_maxhu`: the minimum/maximum half-length of the module boundary along the local u direction (in millimeter).
– `module_hv`: the half-length of the module boundary along the local v direction (in millimeter).
– `pitch_`u`, pitch_v`: the size of detector cells along the local u and v direction (in millimeter).

References

[1] Cornelissen, T. et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, ATL-SOFT-PUB-2007-007, https://cds.cern.ch/record/1020106

[2] Adam, W. et al., *Track reconstruction in the CMS tracker*, CERN-CMS-NOTE-2006-041, CMS-NOTE-2006-041

[3] https://www.kaggle.com/c/trackml-particle-identification

[4] https://cern.ch/acts

[5] http://pdg.lbl.gov/2014/reviews/rpp2014-rev-passage-particles-matter.pdf

[6] PYTHIA8, http://home.thep.lu.se/ torbjorn/pythia81html/Welcome.html