Test beam track reconstruction and analysis of ATLAS 3D pixel detectors

Håvard Gjersdal

June 29, 2014

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1573

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Acknowledgments

I would like to thank my supervisors Ole Røhne, Are Strandlie and Steinar Stapnes.

I am grateful to the ATLAS 3D collaboration, especially the test beam group, for providing a friendly, interesting and challenging work environment.

I would also like to thank Rudolf Frühwirth and Moritz Nadler for including me in their work, and collaborating with me on material and resolution estimation.

Finally I would like to thank Hilde Skjerdal for all her support.

Summary

3D silicon pixel sensors are a new radiation detection technology with electrodes etched into the silicon wafer. The technology offers some advantages over the originally installed ATLAS planar pixel sensors in radiation tolerance, as well as the possibility of being sensitive all the way to the edge of the sensor. ATLAS 3D sensors are used in the ATLAS insertable B-layer, a layer of pixel detectors that is inserted closer to the interaction point than the original innermost pixel layer.

Several test beam experiments have been performed to characterize ATLAS 3D silicon devices in a particle beam produced by the CERN Super Proton Synchrotron. The object of the experiments was to study the response of the devices to particles with known trajectories. The particle trajectories are reconstructed from a beam telescope, an instrument made from position sensitive detector planes.

This thesis summarizes the methods that are used, and the results obtained, in the analysis of ATLAS 3D silicon sensor data from the test beam experiments. This includes the reconstruction of particle trajectories from the beam telescope, and the characterizations of the devices under test.

Some theoretical background is given in the first three chapters. Chapter 1 gives a brief introduction to the physics studied at the LHC, as well as the detector systems of the ATLAS experiment. Chapter 2 gives background on semiconductor detectors and the interactions between fast particles and matter. Chapter 3 provides background on least squares estimation and track fitting with the Kalman filter.

The experimental setups of the test beam characterizations are summarized in Chapter 4. This includes descriptions of the two different beam telescopes that have been used in the characterization.

The methods used for reconstructing particle trajectories from the beam telescope are presented in Chapter 5. This includes the steps needed to prepare the data for track reconstruction, obtaining a description of the detector system, and the track reconstruction procedure. Results showing excellent track quality in real data are presented.

Chapter 6 discusses some of the methods used for the analysis of the performance of the ATLAS 3D silicon detectors.

This thesis is based on four papers. The ATLAS public note “Straight line track reconstruction for the ATLAS IBL testbeam with the EUDET telescope” is a description of the methods that have been implemented for straight line track reconstruction in the EUDET beam telescope. This includes introducing a new method for track finding, and validation of all the implemented methods on real and simulated data.

The paper “Optimizing track reconstruction by simultaneous estimation of material and

resolutions" introduces novel methods for simultaneous estimation of detector resolution and material distribution. A correct description of the detector system is a requirement for the Kalman filter to give optimal estimates of the particle trajectories, and for the uncertainties of the estimates to be correct.

Results from the characterization of 3D devices are presented in "Tracking Efficiency and Charge Sharing of 3D Silicon Sensors at Different Angles in a 1.4 Tesla Magnetic Field" and "Test Beam Results of 3D Silicon Pixel Sensors for the ATLAS upgrade". 3D silicon pixel sensors have a detection efficiency similar to planar pixel sensors in conditions similar to the insertable B-layer. The signal distribution in planar pixel sensors is affected by the presence of the magnetic field in the insertable B-layer. For 3D silicon pixel sensors, this effect is much smaller.

My work has consisted of taking part in the mounting and operation of the test beam experiments. I have held a leading role in the group working on analysis of the test beam data, and was the initial and main developer of the analysis software framework tbmon that was used for the analysis. This framework has since been used and further developed by other test beam experiments, including ATLAS planar pixels for the insertable B-layer. I have worked on data reconstruction, providing aligned and fitted tracks for the analysis group. I have implemented and developed several methods for track reconstruction of data from the EUDET telescope, including the information formulation of the Kalman filter, the combinatorial Kalman filter, the Deterministic Annealing Filter, and a new method called the cluster track finder. I have worked on improving the alignment estimation in the framework used by the EUDET telescope by configuring it to estimate tilt angles through scale factors, and by implementing a pre-alignment step. I have performed research and development of new methods for simultaneous estimation of sensor resolution and thicknesses for optimization of tracking performance.

Contents

Chapter 1

Introduction

The performance of ATLAS 3D silicon sensors have been studied in a beam of high energy pions from the Super Proton Synchrotron. This was done as a part of studying the suitability of 3D sensor technology in the insertable B-layer, the new innermost detector layer of the ATLAS detector.

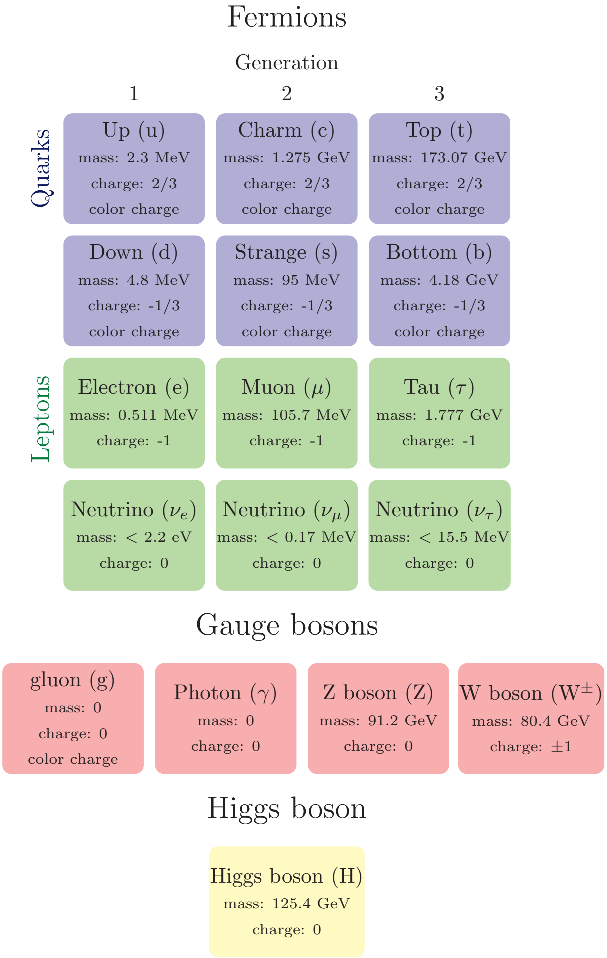
In this chapter, some context is given for the studies that have been performed. A brief summary of the Standard Model of particle physics is given, followed by an introduction to the ATLAS detector. Finally, motivation for, and a brief description of, the test beam experiment performed on ATLAS 3D silicon sensors is given.

1.1 The Standard Model of particle physics

The study of subatomic particles has been an important part of physics for more than a century. This has led to the development of the Standard Model of particle physics, a model that is able to describe all particles and particle interactions, except gravity, that have been observed in particle physics experiments. This model is a relativistic quantum theory description of nature, and is made up of particles that are assumed to be fundamental.

The stable matter that surrounds us consists of fermions that obey the Pauli exclusion principle, the principle that two particles cannot occupy the same quantum state simultaneously. The fermions in the Standard Model are electrically charged leptons, neutral leptons, and quarks that carry an electrical charge as well as a different type of charge, called a color charge. In the Standard Model there are three generations of fermions, each generation consisting of two leptons and two quarks. The particles in the different generations have the same properties, but the particles in the second and third generations are heavier, and have short lifetimes before they decay to lighter particles. Only the light particles in the first generation create stable matter. All the matter particles have antiparticles, i.e. particles with the same mass, but opposite charges.

In addition to the matter particles, several force mediating particles called gauge bosons exist. The electromagnetic force is carried by photons that couple to particles with electrical charge. The weak force, that causes particles to decay into lighter particles, is mediated by W and Z bosons. All the matter particles can interact through the weak force. The description of electromagnetic and weak interactions have been unified to a description of electroweak interactions, modeling the interactions as two aspects of the same force.



Fermions

	Generation 1	2	3
Quarks	Up (u) mass: 2.3 MeV charge: 2/3 color charge	Charm (c) mass: 1.275 GeV charge: 2/3 color charge	Top (t) mass: 173.07 GeV charge: 2/3 color charge
	Down (d) mass: 4.8 MeV charge: -1/3 color charge	Strange (s) mass: 95 MeV charge: -1/3 color charge	Bottom (b) mass: 4.18 GeV charge: -1/3 color charge
Leptons	Electron (e) mass: 0.511 MeV charge: -1	Muon (μ) mass: 105.7 MeV charge: -1	Tau (τ) mass: 1.777 GeV charge: -1
	Neutrino (ν_e) mass: $<$ 2.2 eV charge: 0	Neutrino (ν_μ) mass: $<$ 0.17 MeV charge: 0	Neutrino (ν_τ) mass: $<$ 15.5 MeV charge: 0

Gauge bosons

gluon (g) mass: 0 charge: 0 color charge	Photon (γ) mass: 0 charge: 0	Z boson (Z) mass: 91.2 GeV charge: 0	W boson ($W^{\pm}$) mass: 80.4 GeV charge: ±1

Higgs boson

Higgs boson (H) mass: 125.4 GeV charge: 0

Figure 1.1: The particles in the standard model are the three generations of fermions, five gauge bosons, and the Higgs boson.

The strong force is mediated by gluons, and couples to particles carrying color charge. It binds quarks together in composite particles called hadrons. The protons and neutrons that make up the nuclei of atoms are hadrons made out of three quarks. Another type of hadron, called mesons, are made from pairs of quarks and anti-quarks.

One of many predictions that have been made by the Standard Model is the existence of the Higgs boson. The particles in the Standard Model obtain their mass by interacting with a scalar field with a non-zero vacuum expectation value, called the Higgs field. The Higgs boson is an excitation of this field, and an indication that such a field exists. With the discovery of the Higgs particle at the ATLAS and CMS experiments at CERN [2, 15], all the particles of the Standard Model have been observed in experiments.

The particles in the Standard Model and some of their properties are summarized in Figure 1.1.

Although the Standard Model has been a great success, several questions remain unanswered. Studies of the motion of galaxies suggest the existence of so-called dark matter. This matter is assumed to consist of stable, massive particles that only interact through the weak interaction and gravity. No good candidate for dark matter have been observed in high energy physics experiments.

The world is made up almost entirely of particles, not antiparticles. This asymmetry is not completely understood from the Standard Model.

Particle masses and coupling strengths are not predicted by theory. The reason for these values being what they actually are is not understood, and there is no explanation for why exactly three generations of matter particles exist.

Several theories try to answer these questions. Some are just extensions of the Standard Model adding more particles and interactions. Others are attempts of making more fundamental models than the Standard Model. A more fundamental model can be a Grand Unified Theory, a theory that unifies the electroweak and strong interaction, or a theory of everything, that also includes a quantum description of gravity.

1.2 LHC and ATLAS

Many of the particles in the Standard Model, like the Higgs boson, are not commonly found in nature. These particles have very short lifetimes, and if created, they will almost immediately decay to lighter particles. These particles can be created by colliding lighter particles at very high energies. The Large Hadron Collider at CERN on the border between France and Switzerland was created to collide protons with energies similar to the thermal collisions occurring only moments after the Big Bang. The accelerator was built to be able to accelerate two beams of protons in opposite direction with energies up to 7 TeV. The protons in the beams are arranged in bunches, and the two beams are crossed at four points along the ring. In these points bunches of protons in opposite directions will collide at a rate of 40 MHz. Around each of the bunch crossing points, large and complex particle detectors have been constructed to record the particles produced by the collisions. One of these detectors is the ATLAS detector [1], illustrated in Figure 1.2.

Many of the processes that are studied in ATLAS involve particles with short lifetimes

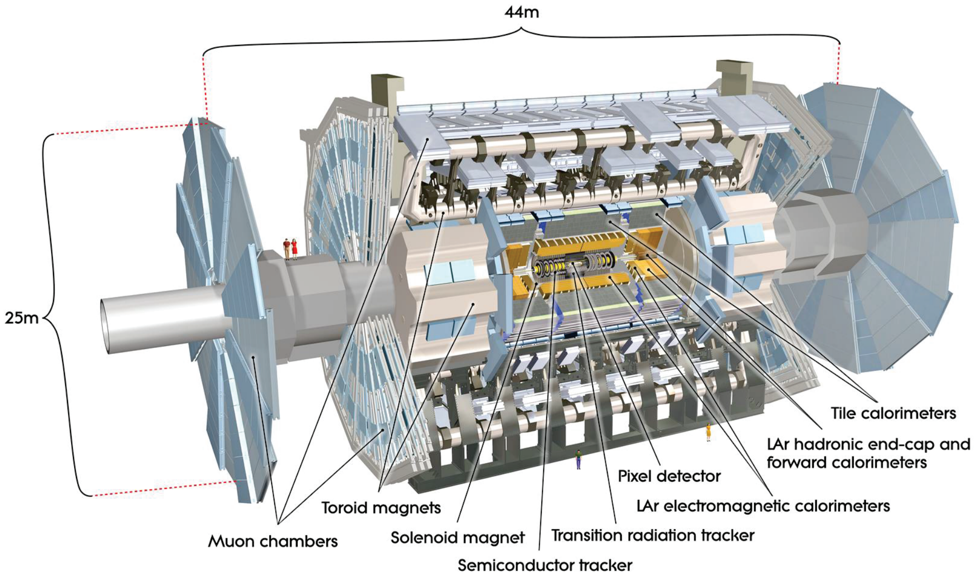


Figure 1.2: The ATLAS detector. (ATLAS Experiment ©2013. ATLAS images are under CERN copyright.)

that never reach any detector element. The processes must then be recreated by studying the decay products of the short lived particles. For this to be possible, the detector must be able to determine the energy, position of origin, momentum, and type of the decay products.

Fast charged particles emitting from the interaction point of ATLAS go through a series of interactions with the material in the detector, causing stochastic changes to the state of the particles and the matter it passes through. Fast charged particles go through inelastic interactions with the electrons in the material which cause ionization and excitation. These interactions are the cause of signal in the sensors that surround the interaction point.

The ATLAS detector is composed of subsystems, detector systems that are in place to fulfill specific needs of the ATLAS detector. These are the inner detector, the electromagnetic and hardonic calorimeters, and the muon spectrometer. The inner detector and the muon spectrometer are tracking detectors, detectors that estimate the charge, momentum and point of origin of particles from their trajectories in a magnetic field. The calorimeters measure the energies of particles by stopping them and measuring the deposited energy. Electrons, photons and hadrons are absorbed in the calorimeters. The interactions between the particles and the different subsystems are illustrated in Figure 1.3.

Neutrinos are stable particles that only interact through the weak force, making it very unlikely that they create signal in any detector layer. These particles are non-ionizing, but still carry energy and momentum. As the protons collide head-on, the momentum in the plane orthogonal to the beam of the particles emitting from the collision should add up to zero. The interaction point of ATLAS is as close to hermetically sealed as possible, so that no ionizing particles should be able to avoid detection. The presence of neutrinos can then be inferred

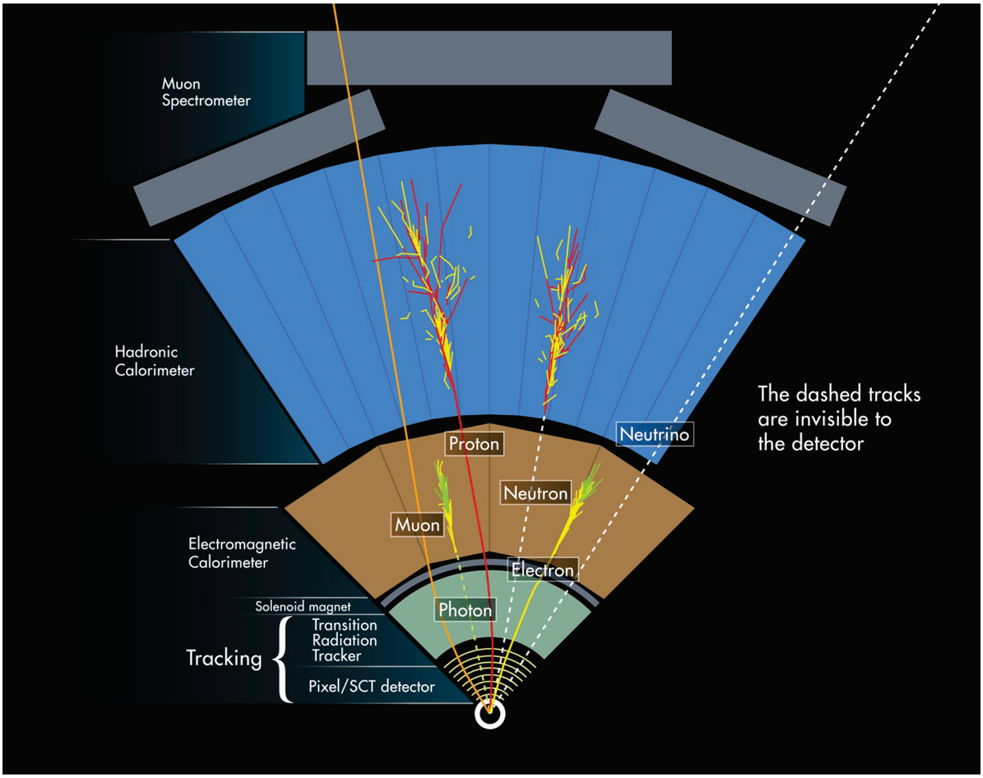


Figure 1.3: Particles passing through ATLAS. (ATLAS Experiment ©2013. ATLAS images are under CERN copyright.)

through non-symmetric momentum of the detected particles. A dark matter candidate, a particle that only interacts through gravity and weak interactions, would also avoid detection and lead to missing momentum in the reconstructed event.

1.2.1 The muon spectrometer

The outermost detector system in ATLAS is the muon spectrometer. Here, all the particles emitting from the interaction point have been absorbed in calorimeters, except muons and neutrinos. The muon spectrometer is a tracking detector consisting of gaseous detectors in three barrel layers and four end-cap wheels inside a magnetic field generated by a toroidal magnet. The barrel region consists of straw tubes, the end-caps are made from cathode strip chambers.

1.2.2 Calorimetry

Between the inner detector and the muon chamber are the electromagnetic and hadronic calorimeters. Both the calorimeters are sampling calorimeters, where layers of sensitive material are separated by absorbers with high stopping power.

Of the two calorimeters, the electromagnetic calorimeter is closest to the interaction point.

In the lead absorbers of the calorimeter, electromagnetic cascades are created by electrons emitting photons in a process called bremsstrahlung, and photons converting to electron-positron pairs through pair production. The energy deposited by the cascade is then sampled by liquid-argon detectors between the absorption layers. The energy of the primary particle is reconstructed from the charged particles in the shower.

Hadrons interact with the nuclei in the iron absorbers of the hadronic calorimeter and create wide showers through inelastic interactions via the strong force. The mean-free path of a hadronic jet is about one order of magnitude greater than the mean-free path of a high energy electron, so while the electromagnetic calorimeter will stop electrons and photons, it will generally not stop hadrons. Relatively cheap scintillator tiles are used as sampling material. At small angles relative to the beam direction, there is a large amount of hadronic radiation that will damage the material it passes through. For this reason the end-caps of the calorimeter are made from the more radiation resistant liquid argon.

1.2.3 The Inner Detector

Each bunch collision can cause a large amount of interactions. Separating the different creation vertices of the particles flowing through the detector is vital for the analysis of the event. This is done by a high-granularity, high-resolution tracking detector close to the bunch crossings. In addition to determining the creation vertices, the tracking detector determines the charge and momentum of charged particles. With the large flux of particles passing through the subsystem, high radiation tolerance for sensors and electronics is needed.

The ATLAS inner detector is a tracking detector inside a 2 T magnetic field generated by a solenoid magnet. It is made up of three detector subsystems: the pixel system , the SemiConductor Tracker (SCT), and the Transition Radiation Tracker (TRT).

The outermost subsystem is the TRT. The TRT consists of gaseous drift tubes in both the barrel and end-cap regions. It is the system that takes up most of the space in the inner detector, and can provide over 30 position measurements to a charged particle. These measurements have large uncertainties compared to the other subsystems of the inner detector, and the measurements have left-right ambiguities. In addition to contributing measurements to the tracking detector, the TRT can identify electrons. When a highly relativistic, charged particle goes from a material to another with a different dielectric constant, it can emit low-energy photons, called transition radiation. The space between the tubes is filled with polypropylene foils and fibers, so that many such transitions occur. The low-energy photons are absorbed by the gas in the drift tubes, and deposit more energy than a charged particle flying through the detector without producing transition radiation. As the electron is the lightest charged particle, it becomes highly relativistic at lower momentum than the heavier particles, and will produce more transition radiation.

The two innermost subsystems consist of semiconductor detectors. Fast charged particles interacting with the electrons in semiconductor material create electron-hole pairs, pairs of free charge carriers. Creating an electron-hole pair only takes approximately one tenth of the energy needed to ionize a molecule in a gaseous detector, giving semiconductors excellent energy resolution. Semiconductors have high stopping power per unit volume compared to the noble gases used in detectors, and a workable signal can be obtained in smaller volumes. With modern processing techniques, a very large density of readout electrodes can be implanted in a semiconductor, producing excellent spatial resolution.

Inside the TRT lies the the SemiConductor Tracker (SCT), a silicon micro-strip detector, where electrodes are implanted as strips on opposite sides of the silicon wafer, producing two-dimensional measurements of the particle position. The SCT consists of four barrel layers, with nine end-caps on both sides.

Inside the SCT lies the pixel subsystem. The silicon pixel detectors making up the system have a matrix of readout electrodes that are implanted on the wafer, producing two dimensional measurements of the position of a charged particle. The pixel sensors have higher granularity than the SCT sensors to deal with the higher density of particles passing through it. The pixel system consists of a barrel region with three concentric cylindrical layers around the interaction point. At each end, there are three end-caps.

1.2.4 The insertable B-layer

B-mesons, which are mesons containing b-quarks, have lifetimes that are too short to make it into the detector systems, but large enough to travel a measurable distance away from the collision point, the primary vertex. When these particles decay, they create secondary creation vertices. Being able to reconstruct these vertices is an important part of particle identification. The innermost layer of the pixel system plays an important role in detecting these secondary vertices, and is therefore called the B-layer.

The pixel system was designed to withstand radiation up to a dose equivalent to a fluence of 10^{15} neutrons of 1 MeV per cm^2. This limit was expected to be reached within a few years of operating the LHC. When this limit is reached, the performance of the pixel system would start to deteriorate. For ATLAS to keep the ability of detecting secondary vertices, a new pixel layer has been inserted directly outside the beam pipe, only approximately 3 cm from the interaction point. This is called the insertable B-layer (IBL) [13].

The insertable B-layer consists of two types of detector technologies, planar pixels similar to the sensors in the original pixel system, and 3D silicon sensors with electrodes etched into the silicon substrate. The new pixel layer is designed to withstand a dose equivalent to a fluence of 6×10^{15} neutrons of 1 MeV per cm^2.

1.3 Test beam characterization of 3D sensors

Before new detector technology can be used in the ATLAS detector, it must go through thorough testing. An important part of this testing is done in a test beam experiment, where sensor response is studied in a particle beam produced by an accelerator. ATLAS 3D pixel sensors have been characterized in a collimated beam of high energy π^+ mesons produced by the CERN Super Proton Synchrotron accelerator.

A test beam experiment makes it possible to study the performance of detector technology with particles with much higher energies than the particles from a radioactive source, and with information about particle trajectories provided by the use of a tracking detector called a beam telescope. The study of 3D sensors in the test beam have focused on quantities such as the detection efficiency, the spatial resolution and the amount of signal recorded under various conditions. The performance has been compared to pixel sensors of the type used in the original pixel system.

To recreate the conditions in the insertable B-layer, test beam data have been taken with ATLAS pixel devices placed inside a magnetic field using the Bonn ATLAS beam telescope. The effects of a magnetic field on 3D sensors in the configuration of the insertable B-layer are expected to be small. Most of the test beam data have been taken without a magnetic field, using the EUDET beam telescope that at the time of data taking was not approved for use inside a magnet. The EUDET beam telescope can take data at a higher rate, and with improved spatial resolution of the reconstructed tracks, compared to the Bonn ATLAS telescope.

Reconstructing particle trajectories from a beam telescope consists of estimating hit positions in the telescope planes and extracting a mathematical description of the trajectories from the detector hits. Reconstructing trajectories from hits can be split into track finding, which is identifying a set of measurements that have been generated from the same particle, and track fitting, which is finding the optimal estimate of the particle trajectories.

Optimal performance from a beam telescope requires an optimal description of the tracking detector. This includes the distribution of material in the detector, the resolution and placement of all the sensor planes. A description with high accuracy can be obtained from numerical methods using information from fitted tracks.

When tracks are reconstructed, they are used to study the performance of the devices under test. Important quantities are the detection efficiency, the measurement uncertainties, the detector response, and how the deposited signal is distributed in the channels near the particle. The response of the sensor varies with the angle of the particle with respect to the sensor plane, and is affected by the presence of a magnetic field. A characterization of ATLAS pixel sensors have been performed both with and without a magnetic field, and with the sensor tilted at different angles with respect to the beam.

Chapter 2

Semiconductor detectors

Crystalline silicon is a semiconductor, a material with conduction properties between that of a conductor and an insulator. The conductivity of a semiconductor can be altered by introducing impurities to the crystal, called dopants. Applying a voltage over a junction of two types of doped silicon creates a volume that is depleted of free charge carriers, but where an energy deposit of a few eV is enough to create them.

When a fast charged particle passes through matter it takes part in a series of interactions with the material it passes through. Inelastic collisions with the electrons in a silicon crystal create free charge carriers, called electrons and holes. Interactions with the nuclei in the crystal can deflect the particle, or cause it to decelerate by emitting photons.

A charged particle passing through depleted silicon will leave a trail of electrons and holes. The free charge carriers move as a cloud through the electrical field, inducing a pulse of current on electrodes connected to the detector. Modern processing methods make it possible to create semiconductor pixel detectors with a large density of readout channels, giving high granularity, high resolution measurements of the positions of particles.

The original pixel layers in the ATLAS detector consists of sensors where electrodes are implanted on opposite sides of the silicon wafer. 3D sensors rely on new methods for implanting electrodes extending into the silicon substrate. This technology allows for smaller distances between electrodes, causing shorter collection distance for the charge carriers, and a lower depletion voltage. The result is a more radiation tolerant detector.

In this chapter a brief introduction to semiconductors and the p-n junction is given. Then the most important interactions between relativistic charged particles and matter are discussed. The focus is on the interactions that are of importance in a test beam experiment, meaning the interactions of fast charged particles and silicon sensors. Finally, the pixel detectors used in the ATLAS insertable B-layer are discussed.

For a more complete discussion of semiconductors, see [40]. For a more general and more complete discussion of semiconductor detectors see [25]. The interactions of a charged particle passing through matter are summarized in [5]. Energy loss in thin silicon sensors are described in [12].

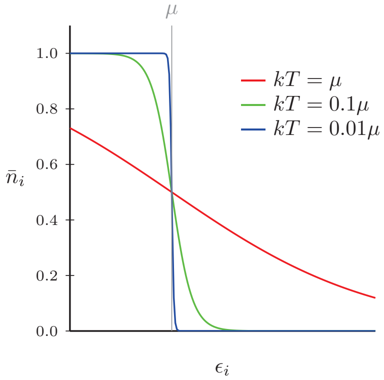


Figure 2.1: The Fermi-Dirac distribution, showing the occupancy of electrons ($\bar{n}_i$) as a function of increasing energy of the single electron state (ϵ_i) at different temperatures. Charge transport can occur in states that are partially filled, the states close to the Fermi level, μ.

2.1 Semiconductors

Silicon has four valence electrons and can form crystal, making a lattice of atoms that are bound together with four others through covalent bonds.

Electrons in an atom occupy orbitals, which form a discrete set of energy levels for the electrons. In the periodic potential in a crystalline solid, the energy levels available for electrons to occupy fall into near continuous bands. The bands are separated by gaps of energy with no available states, called band gaps. Electrons occupying low energy states are bound to an atom in the lattice. High energy states extend through the lattice, and are not localized to a single atom.

Electrons are fermions, and obey the Pauli exclusion principle, i.e., no two electrons can occupy the same state at the same time. The distribution of electrons in the energy levels in a solid is described by Fermi-Dirac statistics. The occupancy of a single electron state is the probability of the state being occupied at any given time. The occupancy of a single electron state i, with energy ϵ_i, and at a given absolute temperature, T, is

$$\bar{n}_i = \frac{1}{e^{(\epsilon_i - \mu)/kT} + 1}, \tag{2.1}$$

where k is Boltzmann's constant and μ is the Fermi level. The Fermi level is the energy needed to add one electron to the solid. Figure 2.1 shows the distribution of the occupancy as a function of increasing energy of the electron states.

At zero absolute temperature, every energy level below the Fermi level will have an occupancy of one, and every level above will have an occupancy of zero. At a temperature above zero, the energy states near the Fermi level become partially filled, meaning electrons occupy the states only a part of the time. Charge transport can only occur through states that are partially filled.

In a semiconductor like silicon, or a crystalline insulator like diamond, the Fermi level lies in a band gap, the difference between the two being the size of the band gap. The band below the Fermi level is called the valence band, and the band above is called the conduction band. The electron states in these bands extend through the lattice.

In a perfect silicon crystal at absolute zero temperature, every state in the valence band is occupied, and every state in the conduction band is empty. At above zero temperatures, electrons can be thermally excited to energy states in the conduction band. This leads to partially filled states in both bands, making charge transport possible. The charge transport occurs through the motion of electrons in the conduction band, and holes in the valence band. A hole is an available energy state, and can be seen as a positively charged particle.

The number of free charge carriers reaches an equilibrium when the rate of electron-hole pairs created by thermal excitation is the same as the rate of electron-hole recombination. Recombination is the process where an electron is deexcited to an available state in the valence band. Lattice defects and impurities in the crystal create energy states in the band gap that can increase the rate of electron-hole pair creation and recombination. Such impurities can also temporarily trap electrons or holes, impeding the motion of the charge carriers.

The number of charge carriers depend on the size of the band gap, the number of defects, and the temperature. At room temperature, there will be approximately 10^{10} electron-hole pairs per cm^3 of silicon [40]. The number of silicon atoms is approximately 5×10^{22} per cm^3 [40]. For comparison, the number of free electrons in copper is approximately 8.5×10^{22} per cm^3.

Electrons and holes moving in an electric field will be deflected by the lattice, limiting the average velocity of the carriers, called the drift velocity. The conductivity of a material depends both on the number of free charge carriers, and the drift velocity in the solid. In silicon, the drift velocity increases approximately linearly with the strength of the electrical field up to approximately 1 kV/cm. The drift velocities for electrons (v_e) and for holes (v_h) in this range are

$$v_e = \mu_e E \tag{2.2}$$

$$v_h = \mu_h E, \tag{2.3}$$

where μ_e and μ_e are the electron and hole mobilities respectively and E is the magnitude of the electrical field. The conductivity is

$$\sigma = q(n\mu_e + p\mu_h), \tag{2.4}$$

where n is the density of electrons in the conduction band, p is the density of holes in the valence band, and q is the elementary charge. The inverse of the conductivity is called resistivity, $\rho = 1/\sigma$.

With an increasing electrical field in a silicon lattice, the drift velocity asymptotically approaches a maximum, the saturation velocity.

For silicon at room temperature, the electron mobility is approximately 1400 cm^{2}V^{-1}s^{-1} and the hole mobility approximately 450 cm^{2}V^{-1}s^{-1} [3]. The saturation velocity for holes and electrons is in the order of 1×10^7 cm/s.

2.1.1 The p-n junction

The electrical properties of a semiconductor material can be modified by intentionally introducing impurities to the crystal in a process called doping.

Atoms from group IV in the periodic system (P, As, Sb) have one more valence electron than silicon. If such an atom is part of a silicon lattice, four of the electrons are part of covalent bonds. The extra electron is only weakly bound to the atom, and will act as a free charge carrier. This is an n-type, or donor, impurity.

Atoms from group III (B, Al, Ga, In) have one electron to few to complete the covalent bonds in the lattice structure. The missing covalent bond is also only weakly bound to the atom, and can move as a hole in the lattice structure. These are called p-type, or acceptor, impurities.

In an intrinsic semiconductor, the number of free electrons and holes are the same, as they are created and recombined in pairs. In a p-type or n-type doped semiconductor, most of the free charge carriers will come from impurities. The conductivity of the material increases with doping concentration.

Charge carriers in doped material are weakly bound to the impurities, as drifting away leaves an ionized impurity. Donor impurities introduce energy levels in the band gap slightly below the conduction band, acceptors introduce levels slightly above the valence band. At absolute zero temperature, the donor levels are filled and the acceptor levels are empty, but very little temperature is needed for the states to become partially filled.

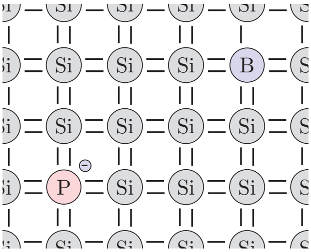


Figure 2.2: 2D representation of a silicon lattice with one p-type (boron, B) and one n-type (phosphor, P) impurity. The p-type impurity lacks one electron to complete 4 covalent bonds. The n-type impurity has one electron to spare after forming 4 bonds. Both these impurities are mobile.

A semiconductor detector is a diode consisting of a junction between a n-type and p-type material. In such a junction, there will be a diffusion current of holes drifting out of the p-type side, and electrons drifting out of the n-type side. This diffusion will leave fixed negative acceptor ions on the p-type side, and positive donor ions on the n-side, creating opposite space charges on the two sides of the junction. An equilibrium is reached when the drift current in the field due to the potential difference between the sides is equal to the diffusion current. The equilibrium potential difference is called the built in voltage, V_{bi}. The built in voltage leaves a region around the junction where free charge carriers are swept away, and that is practically free of charge carriers. This is called the depletion zone. With the charge carriers removed, the charge density in the depletion zone is equal to the net doping level. Figure 2.3 shows an illustration of a p-n junction.

Under reverse bias, an external voltage applied over the junction where the n-side is connected to the positive terminal and the p-side is connected to the negative, the effect is that the holes in the p-side and the electrons on the n-side are pulled further away from the junction. This causes the depletion region around the junction to grow. Under a forward bias the depletion region shrinks until the junction becomes conducting.

2.1.2 Full depletion approximation

Assuming the doping concentration is constant on both sides of the junction, the depletion zone is completely depleted of charge carriers, and that the zone outside is completely neutral, makes

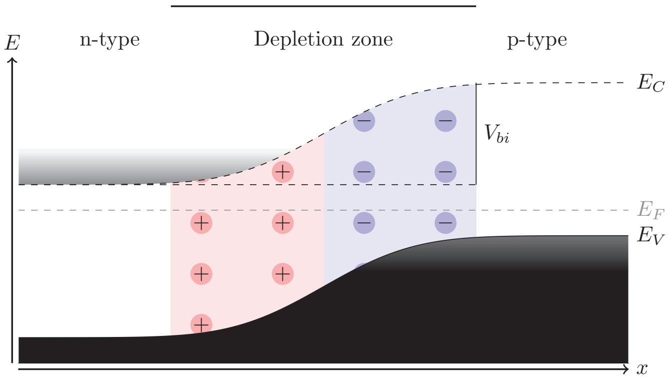


Figure 2.3: Band diagram of a p-n junction. E_C is the lowest energy level in the conduction band, E_V is the highest in the valence band. The built-in potential difference, V_{bi}, acts as a barrier preventing diffusion of electrons from the n-type to the p-type side, and diffusion of holes from the p-type to the n-type side. The color in the bands qualitatively represent the occupancy of electrons in the states, where black means an occupancy of one, white means an occupancy of zero. The potential difference causes free charge carriers to drift away from the junction.

it possible to calculate the depletion depth from Poisson's equation.

$$\nabla^2 \varphi = -\rho/\epsilon \tag{2.5}$$

Here φ is the electrical potential, ρ is the charge density, and ϵ is the permittivity. The charge density depends on the net charge due to free carriers in the volume, as well as the net charge due to the ionized impurities. Under the assumptions of full depletion, the density only relies on doping concentration.

$$\rho = q(N_d - N_a), \tag{2.6}$$

were q is the magnitude of the electron charge,and N_a and N_d are the densities of acceptor and donor impurities respectively, all of which are ionized.

There is no net charge in the material, so the space charge on the two sides must balance out. This means that

$$d_p N_a = d_n N_d, \tag{2.7}$$

where d_n is the depletion depth on the n-side, and d_p is the depletion depth on the p-side. The total width of the depleted region is

$$d = d_n + d_p = \sqrt{\frac{2\epsilon}{q} \frac{N_a + N_d}{N_a N_d} (V_{bi} - V)}, \tag{2.8}$$

where V_{bi} is the built in voltage, and V is the external voltage. The depletion zone grows proportional to the square root of the applied reverse bias voltage.

The junction can be seen as a parallel plate capacitance, where the capacitance decreases with increasing depletion depth.

$$C = \epsilon A/d \tag{2.9}$$

Here A is the area of the cross-section of the junction.

2.1.3 Leakage current

While the depleted zone is nearly non-conducting, it is not completely devoid of charge carriers, and there is a small fluctuating current passing through it, called leakage current. The reasons for the current are minority charge carriers, i.e., holes on the n-side or electrons on the p-side, creation of electron-hole pairs in the junction through excitation, and charge flowing through surface channels.

The amount of current passing through an ideal diode can be calculated from the Shockley diode equation,

$$I = I_s(e^{V/(nV_T)} - 1), \tag{2.10}$$

where I_s is the reverse saturation current, V is the voltage applied over the current, n is an ideality factor and V_T is the thermal voltage. The thermal voltage is

$$V_T = \frac{kT}{q}, \tag{2.11}$$

where T is the temperature, k is Boltzmann's constant, and q is the magnitude of the electron charge. The current as a function of voltage is shown in Figure 2.4. In an ideal diode, the current reaches a plateau where it would be constant with the bias voltage, and equal to the reverse saturation current. In a real sensor the reverse current can increase slightly with the voltage up to a point, where the device becomes drastically more conducting. This is called the breakdown voltage. Operating a device at or above this voltage can damage the sensor.

The reverse saturation current is not constant for a material, it increases exponentially with temperature, doubling approximately every 7°. The fluctuation of the leakage current is a source of noise. The amount of fluctuation increases with the mean.

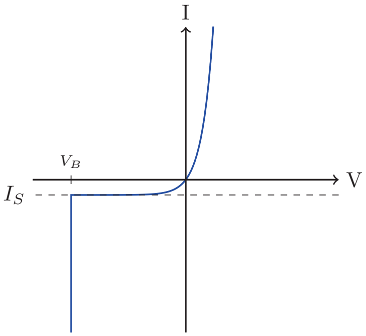


Figure 2.4: Current in an ideal diode as a function of applied voltage. In the reverse bias direction, the amount of current quickly reaches a plateau, I_s and stays constant with the voltage until a breakdown voltage, V_B, is reached. The breakdown is not included in the Shockley diode equation, but is included for illustration.

2.1.4 Electrodes in a semiconductor

In a junction of metal and semiconductor material, the energy levels of the electrons in the metal will extend into the semiconductor, and the energy levels in the semiconductor will extend into the metal. This affects the band structure of the semiconductor near the junction, and can cause a Schottky barrier, a potential energy barrier with a depletion zone extending into the semiconductor. This is illustrated in Figure 2.5.

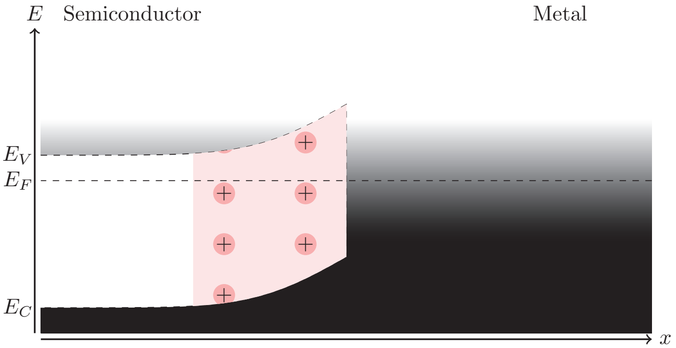


Figure 2.5: A junction of n-type semiconductor with metal.

Highly doped material, called n$^+$- or p$^+$-type material, causes the depletion zone in the metal-semiconductor junction to be small enough that charges can tunnel through. Using layers of highly doped material between the semiconductor and the metal creates Ohmic contact, where the current is proportional to the voltage over the junction.

A semiconductor sensor is often made with an asymmetric junction, for example between n$^+$ and p$^+$ electrodes in a lightly doped n-type bulk material, as illustrated in Figure 2.6. At a reverse bias that depletes the entire bulk material, very little of the highly doped volumes will be depleted. Typical doping concentrations of the bulk material is in the order of 10^{12} cm^{-3}, the concentration in heavily doped material can exceed 10^{18} cm^{-3}. In this case, the depletion zone into the bulk would extend 10^6 times farther than the depletion zone into the electrode.

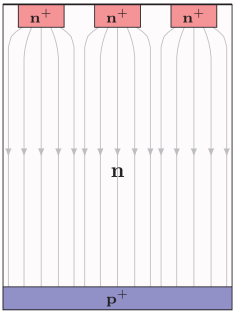


Figure 2.6: Cross section of a pixel device with n$^+$-type readout electrodes in n-type material. The gray lines are approximations of the electric field lines. The depletion zone will grow from the p$^+$ electrodes, and the n$^+$ electrodes will only be insulated from each other when the entire bulk material is depleted.

2.2 Signal formation

The signal on an electrode does not arrive when the electrons or holes are collected on it, it is induced on the electrode by the motion of charge through the electric field surrounding it, and in fact stops when all the charges have been collected. The Shockley-Ramo theorem states that the instantaneous current received by an electrode by a moving charge is:

$$i = E_v q v, \tag{2.12}$$

where q is the charge of the particle, v is the instantaneous velocity of the particle, and E_v is the component of the electric field in the direction of v at the instantaneous position of the particle under the following conditions: the electron is removed; the electrode is raised to unit potential; all other electrodes are grounded [34, 38]. If a cloud of electrons moves toward an electrode, it causes movement of the electrons in the electrode, and the induced current creates a positive charge build-up on the electrode. The charge build-up vanishes when all the electrons in the cloud are collected.

In a semiconductor, the electrons have higher mobility than holes. The electrons will cause a fast pulse with a high amplitude compared to the signal induced by the slower holes.

There is no electric field inside the undepleted material, and motion of electrons and holes outside the depleted zone does not generate any signal. Depleting the entire bulk material maximizes the amount of signal a particle will generate. Another benefit of complete depletion is that the capacitance of the detector decreases as the depletion zone grows. Capacitance is a source of noise.

2.3 Particles passing through matter

Relativistic, charged particles passing through a layer of semiconductor will go through a series of interactions with the material. Electronic interactions, inelastic collisions with the electrons in the material, is the main cause of energy loss for moderately relativistic particles. Elastic interactions with the heavy nuclei deflect the particle in a process called multiple Coulomb scattering. In the field of the charged particles in the material, a particle can decelerate by emitting bremsstrahlung, photons within a continuous energy spectrum that extends to large fraction of the initial particle energy. For high energy electrons, this is the dominating mode of energy loss.

All these interactions are of a stochastic nature, and are described by probability distributions.

2.3.1 Electronic energy loss

Electronic energy loss is due to inelastic collisions with the bound electrons in the material. The energy loss in each collision is generally relatively small, 90% of the collisions lead to an energy loss less than 100 eV. Less frequent collisions with large energy loss make the distribution of energy loss per collision have a large variance, and a long tail towards higher energy.

Energy loss from these collisions is the cause of signal in detectors, as it leads to ionization and excitation. In crystalline silicon, the energy deposited by electronic interactions is absorbed by the creation of electron-hole pairs and by vibrations in the lattice structure. A large number of these collisions occur, and the particle leaves a trail of electron-hole pairs behind it.

The amount of energy lost per distance traveled depends on the type and energy of the charged particle, and the type and density of the material it passes through. The shape of the energy loss as a function of $\beta\gamma$ of the incoming particle is similar for different particles in different materials. The mean energy loss of muons at varying energies in copper is shown in Figure 2.7. Particles with energies in the plateau around the minimum are called minimum ionizing particles.

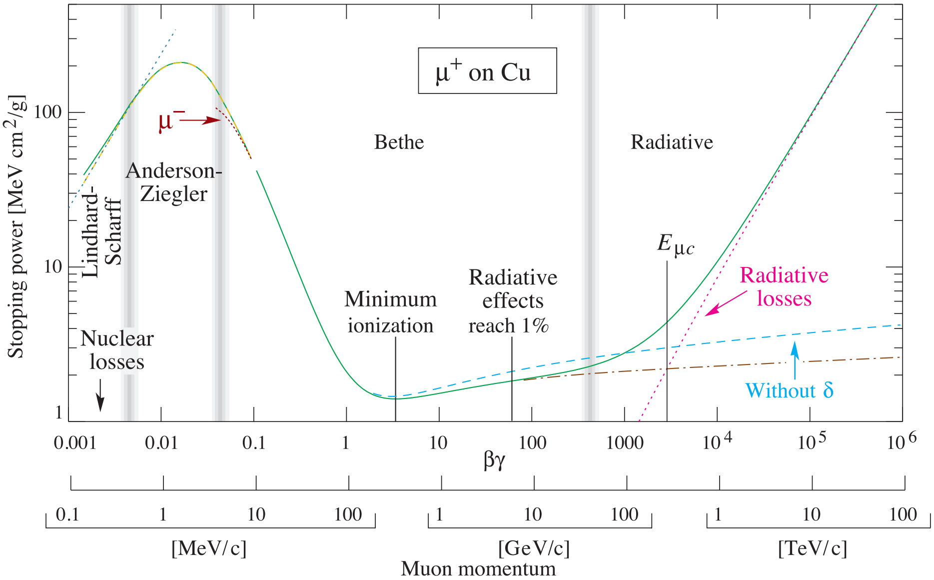


Figure 2.7: Mean energy loss of muons in copper as a function of $\beta\gamma = p/Mc$, where p and M are the momentum and mass of the fast charged particle. The plot is taken from [5].

The mean energy loss of a moderately relativistic charged particle can be calculated from the Bethe formula [5]. The mean energy loss is affected by rare events with large single collision energy losses, meaning that the probability distribution for energy loss in relatively thin sensors has a long tail towards high energies, and the most probable value for energy loss is considerably lower than the mean. The most probable value and the full-width-at-half-maximum of the energy loss is more descriptive for an average event than the mean.

The skewness of the probability distribution for energy loss varies with absorber thickness, but does not become truly Gaussian even for thick absorbers. Energy loss in moderately thick sensors, such as a 3 mm thick plastic scintillator, can be adequately described by the highly skewed Landau distribution, illustrated in Figure 2.8 [24, 43, 5]. For thin absorbers, such as silicon sensors of a few hundred μm, the distribution is significantly wider than the Landau distribution, and has a longer tail [10].

The most probable energy loss for a 100 GeV pion in a 400 μm thick silicon detector is

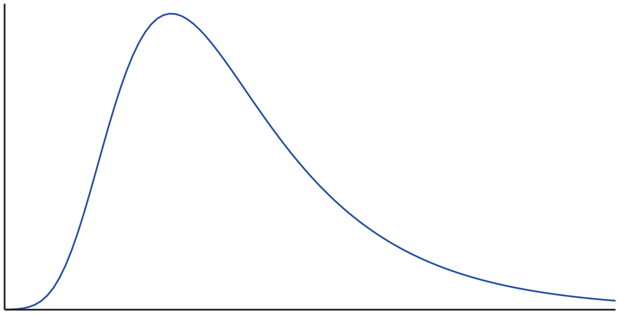


Figure 2.8: The Landau distribution, used to model energy loss in moderately thick absorbers. In silicon sensors with a width of a few hundred µm, the distribution is wider and has a longer tail.

approximately 100 keV, with a full-width-at-half-maximum of 43 keV and a tail that extends up to approximately 1 MeV [12]. Both the amount of energy loss and the variance is small compared to the energy of a minimum ionizing particle.

The most probable value for energy loss per distance in silicon increases with the thickness of the sensor. For a 200 to 300 µm thick sensor it is close to 280 eV/µm [9].

Not all the electronic energy loss goes into creating electron-hole pairs, some of it is absorbed as vibrations in the lattice. On average, one electron-hole pair is created for approximately every 3.62 eV of energy a particle loses in silicon at 300 K [25], and approximately 80 pairs will be created per µm. The band gap in silicon is approximately 1.1 eV.

The electromagnetic interactions with the charged particle create electron-hole pairs close to the trajectory of the particle. Many charge carriers freed from the lattice will absorb sufficient energy to become secondary ionizing particles, creating additional electron-hole pairs. This will generally occur close to the trajectory of the primary particle, but if a large amount of energy is transferred to a single electron free charge carriers will be created some distance away from it. These Energetic secondary electrons are called δ rays. If a δ ray is not stopped in the sensitive area of the detector, not all the energy loss of the primary particle will be detected.

2.3.2 Radiative energy loss

Radiative energy loss is the dominant mode of energy loss for highly relativistic particles. This occurs through transition radiation, Cherenkov radiation and bremsstrahlung. Cherencov radiation is low-energy photons emitted when a charged particle passes through a material at a speed higher than the phase velocity of light in the material. Transition radiation is low energy photons emitted when a highly relativistic charged particle passes through a border between two materials with different dielectric constants.

The electron is the lightest charged particle, and becomes highly relativistic at lower energies than the other particles. The main mode of energy loss by high energy electrons is bremsstrahlung [6]. Bremsstrahlung is emitted when the primary particle is deflected by a charged particle in the material, decelerating the primary particle. The energy of the emitted photons form a continuous spectrum, and a large fraction of the electron energy can be lost in a single emission.

A radiation length is a material characteristic, and is the average length where a high energy electron will lose all but 1/e of its energy by bremsstrahlung. Silicon has a radiation length of approximately 9.36 cm. Energy loss by bremsstrahlung is a highly non-Gaussian process, and in thin sensors the variance is large.

2.3.3 Multiple Coulomb scattering

A charged particle will also take part in elastic collisions with the nuclei in the material. This causes a series of small deflections to the particle, called multiple Coulomb scattering [7, 5]. The sum of these deflections lead to a change in the angle of the particle, as illustrated in Figure 2.9. The change in the angle nearly follows a Gaussian distribution with a mean change of zero. The standard deviation of the distribution of scattering in a plane parallel to the direction, σ_{plane}, can be calculated from the Highland formula [21].

$$\sigma_{\text{plane}} = \frac{0.0136}{E}\sqrt{\frac{X}{X_0}}(1 + 0.038ln(\frac{X}{X_0})) \tag{2.13}$$

Here E is the energy of the particle in units of GeV, and $\frac{X}{X_0}$ is the thickness of the plane in units of radiation lengths. Scattering in two orthogonal planes parallel to the trajectory can be modeled as two independent stochastic processes, each with a standard deviation that can be calculated from Eq. 2.13.

The Gaussian approximation describes 98% of the core scattering distribution, but the tails of the scattering distribution extend further than the tails of the Gaussian distribution.

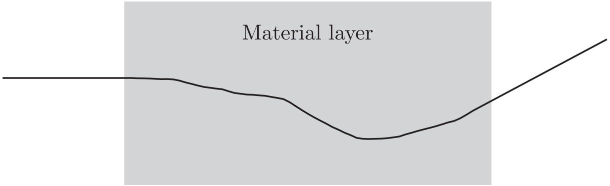


Figure 2.9: A particle deflected by multiple Coulomb scattering.

2.3.4 Radiation damage in silicon

Radiation can alter the lattice structure in silicon permanently through creating point defects in the crystal through non-ionizing energy loss [13, 23]. Nuclear interactions can knock atoms out of their position in the crystal, or transmute the nuclei.

These lattice defects alter the electrical properties of the material. Transmutation can lead to the removal or creation of donor or acceptor levels, altering the effective doping concentration in the material. Point defects lead to additional energy levels in the band gap, which act as trapping and recombination centers. Recombination centers increases the leakage current, and trapping reduces the mobility in the material.

In a detector with n-type bulk material, prolonged radiation has the effect of reducing the concentration of n-type impurities and increasing the p-type concentration. Eventually, a type inversion will occur, where the bulk acts as p-type material. The concentration of p-type impurities increase with the amount of radiation damage, resulting in higher voltages needed for complete depletion of the sensor bulk.

The amount of heat generated in the sensor increases linearly with the voltage applied over the junction and the amount of leakage current. As the amount of leakage current increases exponentially with the temperature in the device, a thermal runaway can occur. With an increasing doping concentration from transmutations, there comes a point when applying sufficient voltage for full depletion is not possible without risking thermal runaway.

The point defects are mobile at room temperatures, but much less so than the electrons or holes. The mobility of the defects rises with temperature, and the motion of the defects lead to annealing effects. Heating the sensors for short amounts of time lead to beneficial annealing. Storing sensors at room temperature is not problematic [14], but sensors are operated at cool temperatures to reduce the leakage current and power dissipation in the sensor.

2.4 ATLAS pixel detectors

ATLAS planar pixel sensors are made from high resistivity n-type silicon wafers with a width of approximately 300 µm in the original pixel detectors, and 200 µm in the IBL [1, 41]. The n^+- and p^+-type electrodes are implanted on opposite sides of the wafers, the readout electrodes are the n^+ electrodes. In a non-irradiated sensor, the depletion zone starts to grow from the p^+-side, and the readout electrodes are not isolated from each other until the bulk material is completely depleted. Full depletion occurs at 50-150 V. Irradiated sensors will eventually see a type inversion of the bulk material. The junction will then grow from the n^+-side, and the readout electrodes are isolated before the bulk is completely depleted. In the insertable B-layer, the maximum operating voltage is 1000 V.

Each n^+ electrode is connected to the input channel of a charge sensitive preamplifier in a front-end chip through a bump-bond, as illustrated in Figure 2.10. The current induced on the electrode is integrated on a capacitor that is discharged by a constant current. This leads to a triangular pulse shape where the charge builds up quickly on the capacitor, and then slowly discharges. The width of this signal is proportional to the deposited charge. The triangular signal goes into a discriminator, and is digitized as a time-over-threshold that is counted in units of bunch crossings. Figure 2.11 shows a simplified schematic representation of a preamplifier. Figure 2.12 illustrates the signals in the readout channels.

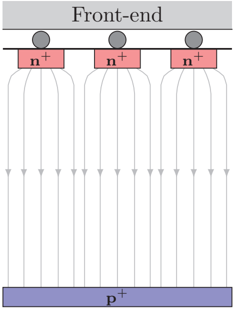


Figure 2.10: ATLAS planar pixel device bump-bonded to a front-end chip. The readout electrodes are connected to the input of the preamplifiers through the bump-bonds.

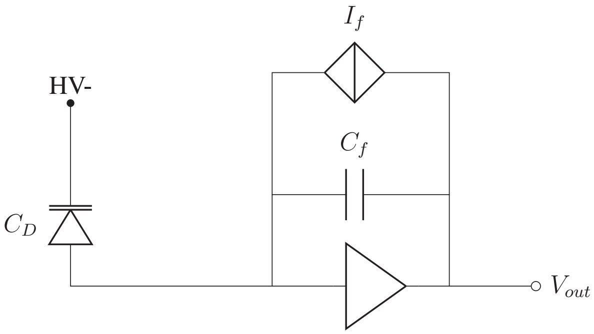


Figure 2.11: Simplified schematic of a charge sensitive peramplifier, where C_D is the charge generating sensor. The charge is integrated on C_f, which is discharged with a constant current over I_f.

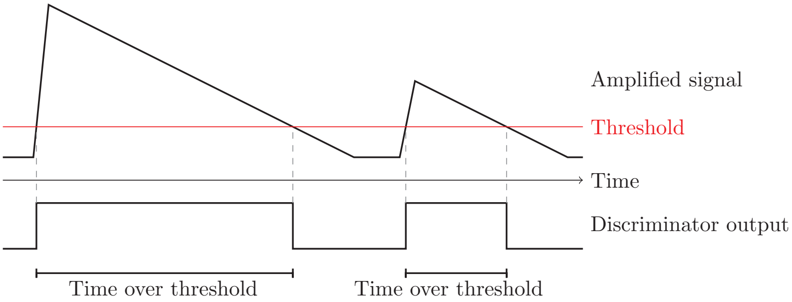


Figure 2.12: Amplifier and discriminator signals from a front-end channel. Two signals are shown, one stronger and one weaker. Both signals have approximately the same rise time, but the angle of inclination of the rising edge is higher for the stronger signal. The signals are discharged at the same rate. The length of the discriminator signal is counted by the 25 ns clock that gives the bunch crossing ID to the signal, and digitized as the time-over-threshold.

The feedback current that discharges the capacitor and the threshold for the discriminator are controlled by two digital-to-analog converters in each pixel, so the signal response can be tuned for each pixel individually to obtain a uniform response across the channels. For the test beam results presented in Appendixes C and D, the front-end chips have been tuned to a threshold corresponding to 3200 deposited electrons, and for a time-over-threshold of 60 to correspond to 20 000 electrons.

Two types of front-end chips are used by the pixel systems in ATLAS. The original pixel system uses the FE-I3 [32]. It consists of 18 × 160 readout channels, and is used with sensors with a pitch of 400 µm × 50 µm. Every 18th column is 600 µm wide to allow for continuous sensitive material between front-end chips. In every column, the eight first and eight last pixels are ganged in pairs to a common readout channel. This causes an ambiguity in these channels that must be resolved by tracking.

In the insertable B-layer, sensors are connected to modules consisting of one or two FE-I4 front-end chips [20]. The FE-I4 has higher granularity, with pixel size 250 µm × 50 µm and no ganged pixels. Each chip consists of 80 × 336 channels. The FE-I4 offers less dead time, at the cost of less resolution in signal, as the time over threshold value is stored in a 4 bit field, instead of 8 as in the FE-I3.

The original ATLAS pixel subsystem is made from pixel modules, where a sensor tile is connected to 16 front-end chips of the type FE-I3, that process and digitize the signal from the sensor. The front-end chips, that are arranged in two rows of eight chips, are connected to a module control chip (MCC), that receives and transmits digital data from the module. The front-end chips and the MCC are situated on a flexible printed circuit, called the flex-hybrid, that routs signals and power in the module. If a trigger is issued, information about the time, time-over-threshold, row and column number of the pixel going above threshold is transmitted.

The insertable B-layer consists of 14 staves, arranged cylindrically around the interaction point as illustrated in Figure 2.13 [13, 41]. Each stave consists of 32 single chip modules with 3D sensors, or 16 double chip modules with planar sensors. The sensors are connected to FE-I4 front-end chips. The detector are mounted so that the center of the sensors are tilted at 14° with respect to the radial direction. Across the sensors, the angle will vary between 0 and 27°.

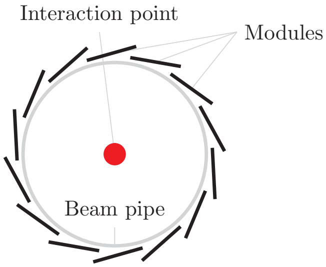


Figure 2.13: Illustration of a cross section of the IBL. 14 pixel modules are arranged around the beam pipe, tilted at 14° with respect to the radial axis. The IBL is installed within a radius from the interaction point between 31 and 40 mm.

In the cylindrical coordinate system around the beam pipe, charged particles will bend in the plane defined by the radial direction to the particle and the azimuthal angle. The detectors are mounted so that a measurement in the direction with 50 μm pitch give information about the azimuthal angle.

2.4.1 3D pixels

3D pixel sensors [31] aim to be more radiation hard than planar sensors by reducing the distance between the electrodes, while keeping the the amount of electron-hole pairs created by a particle large. To achieve this, vertical columns are etched into the silicon wafers, in a process called deep reactive-ion etching (DRIE). The n^+- and p^+-electrodes are then implanted in these columns. The electrode structure is illustrated in Figure 2.14. The distance a charge needs to travel depends on the distance between the etched columns, and not the thickness of the wafer, as is the case for planar devices. The amount of electron-hole pairs produced by a particle still depends on the thickness.

Due to the smaller inter-electrode distance in 3D devices, less voltage is needed to completely deplete the sensors. Reducing the distance the charges need to travel reduces charge trapping.

ATLAS devices have been produced with a varying amount of electrodes per readout channel. In one direction, the readout electrodes are always separated by 50 μm. The spacing in the other directions varies with different designs. The design for the insertable B-layer is a design with 2 readout electrodes per 250 µm pixel cell. This is roughly the same inter-electrode distance as using 3 electrodes per 400 µm pixel cell in sensors for the FE-I3.

3D sensors have the possibility of active edges. At the end of the sensor the silicon wafer needs to be cut. For planar sensors, this cut goes from the n^+-side to the p^+-side, and the sawing of the sensors can cause large amounts of leakage current between electrodes through channels in the cut surface, making it impossible to obtain sufficient depletion. To prevent surface leakage from becoming a problem, the cut must be performed at some distance to the sensitive area. This area is filled with guard rings for terminating the field, rings of electrodes around the sensitive area, where the voltage is reduced for each ring [45].

For a 3D device, a trench can be etched around the sensor and doped. This will terminate the field lines all the way around the sensor [22]. The sensor can be depleted all the way to this trench, maximizing the sensitive area. Etching through the edge instead of sawing removes the chances of any cracks forming.

Two types of 3D detectors have been considered for use in the insertable B-layer. One is of the type proposed in [31], where the electrodes are etched from one side and extend all the way through the detector. These sensors are called full 3D. The other type is called Double Side, Double Type Columns (DDTC). In these devices, one type of electrodes is etched from one side, the other type from the other [8]. 3D-DDTC devices can have a full or partial electrode overlap. In devices with partial electrode overlap the etched columns do not extend all the way through the silicon wafer. The electrical field of partially overlapping DDTC sensors share properties with both full 3D and planar sensors.

After the columns are etched and the electrodes are implanted, the holes can be filled with material that remain some detection efficiency, but the probability of detecting signal generated inside the electrodes is much lower than in the depleted silicon. If particles are traveling in paths parallel to the etched holes, the detection efficiency of the detector will be reduced, as most of the particles that do not travel through depleted bulk will go undetected. As the sensors in the insertable B-layer will be tilted, this effect is not expected to cause an overall drop in detection efficiency.

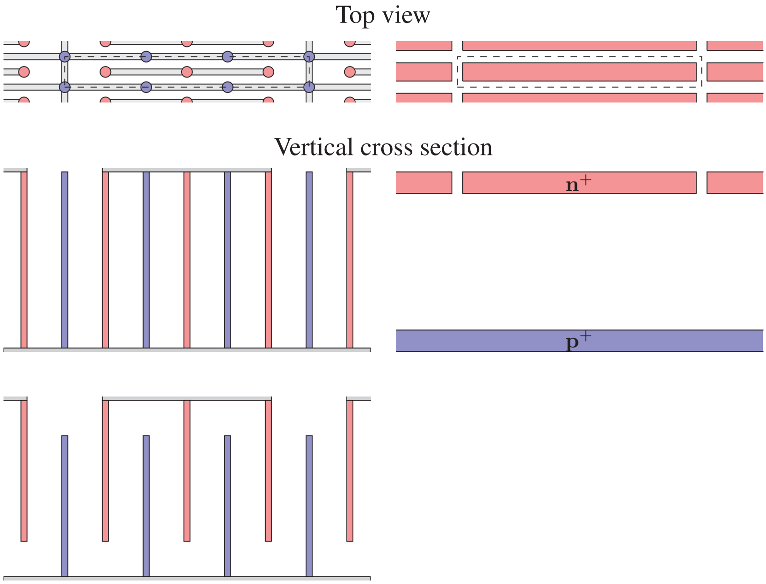
2.4.2 The shape of the charge cloud

A charged particle passing through silicon releases a cloud of charge carriers. The speed and direction of the drift of these charge carriers in the electric field determines the signal that is induced on the electrodes, which in turn affects important detector characteristics like detection efficiency and spatial resolution of the position estimates.

The initial shape and size of the charge cloud is stochastic, but largely determined by the position and direction of the charged particle. After this, the charges are affected by drift and diffusion. The Lorentz force will act upon all the charged particles drifting in the cloud.

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.14}$$

As the particles reach an average drift speed, the average drift direction will also reach an equilibrium. This direction is called the Lorentz angle.


Top view

Vertical cross section

Figure 2.14: Comparison of 3D and planar pixel layout. The left side shows 3D devices, the right side a planar pixel sensor. The two vertical cross section shows two different 3D devices, one with full electrode overlap and the other with partially overlapping electrodes. Both devices have 3 readout electrodes per pixel. The silicon bulk material is treated as transparent to show both electrode types in the 3D illustrations. The dashed lines outline the border for the pixel.

For planar devices, the presence of a magnetic field will have the effect of focusing the charge cloud created by a particle going through the detector in the same direction as the Lorentz angle, as can be seen in Figure 2.15. The charge cloud of particles that pass through the device parallel to the electrical field, or particles that are tilted in the opposite direction of the Lorentz angle, will be defocused. For full-3D devices, a slight change in drift angle compared to the electrical field will not affect the signal induced on the electrodes, and no focusing or defocusing effects are expected to occur from the presence of a magnetic field.

Defocusing of the charge cloud increase the probability of signal above threshold in more than one readout channel. This allows for interpolation techniques to be used to estimate the particle position, improving the resolution of the device. Spreading the signal over more than one channel also means less signal per channel, leading to a worse signal-to-noise ratio. As the signal-to-noise ratio also decreases with radiation damage, more signal sharing between channels can decrease the life-time of a sensor.

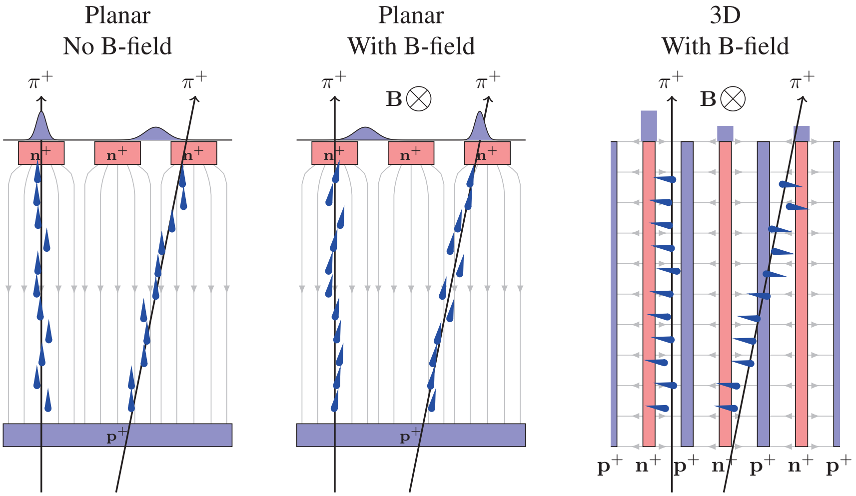


Figure 2.15: Cross section of pixel sensors with charged particles passing through. The direction of the electrical field is indicated by the gray lines, mainly vertical in the planar devices, and horizontal in the 3D device. 💧 indicates that the mean electron drift direction is upwards. The blue Gaussian shapes indicate the shape of the charge cloud as it approaches the readout electrodes. The blue bars on the top op the 3D illustration indicates the amount of signal induces in the channel. The directions of the B-fields are into the page.

Chapter 3

Track fitting with the Kalman filter

The experimental setup of the test beam, which will be discussed in more detail in the following section, consists of a beam telescope and devices under test that are placed in a collimated beam of high energy pions. The detector planes of the beam telescope and the devices under test are thin. Without a magnetic field, the pions in the beam go through the detector system as broken lines, due to scattering in the detector planes.

The Kalman filter is a recursive implementation of the least-squares estimator, and is used to obtain a parametric description of particle trajectories [16]. The Kalman filter alternates between updating the parameter estimate at a plane by including a measurement, and making predictions of the parameters at the next measurement plane. Under certain conditions which will be discussed in the following, the Kalman filter is the unbiased estimator with minimum variance. The Kalman filter can be used as a building block for pattern finding algorithms, and adaptive methods dealing with noisy input [26, 17].

In this chapter, track fitting with the Kalman filter in a test beam experiment consisting of thin planes without a magnetic field will be discussed. This includes a discussion of the track model, some background on least-squares estimation, and an introduction to the Kalman filter. A new technique for visualizing the Kalman filter estimates is introduced.

3.1 Particles moving through the detector system

The particles have relatively little interaction with air, and move as near straight lines between material layers in the beam. In the detector planes, or other dense material layers in the path, the particles interact with the material and change direction through multiple Coulomb scattering. Figure 3.1 shows an example of a track through a test beam experiment. For simplicity, the material planes in the experimental setup are assumed to have material distributed in planes with zero spatial thickness.

The test beam experiment is described in a right-handed Cartesian coordinate system, where x points in the horizontal direction, y in the vertical direction, and z is in the nominal direction of the particle beam.

The state of a particle is described with a parameter vector at the intersection with a set of planes it passes through. Without a magnetic field, no information about the momentum of a particle can be extracted from the trajectory, so the parameter vector only describes the position

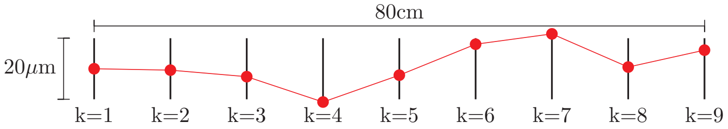


Figure 3.1: Track with multiple Coloumb scattering in a test beam experiment. The detector planes are the vertical black lines, the track is the broken red line.

and direction of the particle. The parameters that have been chosen to describe the track at a plane k in the test beam system are

$$\mathbf{x}_k = \begin{pmatrix} x \\ y \\ dx/dz \\ dy/dz \end{pmatrix}, \tag{3.1}$$

where x and y describe the position of the intersection in the global xy-plane, and dx/dz and dy/dz are direction tangents of the two angles with respect to the z-axis. This vector will be referred to as the state vector of the track at a plane. Uncertainties of the parameters are described in the 4×4 covariance matrix, $\mathbf{C}_k$.

3.1.1 Propagation of parameters and uncertainties

Assuming a particle does not interact with the air between the planes in the detector system, and that the true state vector of a particle is known directly in front of plane k, the state of the particle in front of the next plane, $\mathbf{x}_{k+1}^{true}$, is

$$\mathbf{x}_{k+1}^{true} = \mathbf{f}_{k+1}^{k}(\mathbf{x}_k^{true} + \boldsymbol{\omega}_k). \tag{3.2}$$

The function $\mathbf{f}_{k+1}^{k}$ is deterministic and describes the change in the state vector for the particle moving as a straight line between plane k and $k+1$. The changes to the parameters due to the interactions with the matter in plane k are described by the stochastic variable $\boldsymbol{\omega}_k$. If the material layers are thin the position of the particle in the xy-plane will be nearly the same when it exits the layer as it was when it entered it, but multiple scattering will change the direction tangents. All the particles passing through a sensor of uniform thickness traverse nearly the same amount of material if all the particles traverse it at nearly the same angle. Thin sensors in a collimated beam can be modeled as having zero spatial thickness with little loss of accuracy.

With multiple scattering being the only material interaction that affects the motion of the particle, the expectation value and covariance matrix of $\boldsymbol{\omega}_k$ are

$$E(\boldsymbol{\omega}_k) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } cov(\boldsymbol{\omega}_k) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{xz}^2 & 0 \\ 0 & 0 & 0 & \sigma_{yz}^2 \end{pmatrix}. \tag{3.3}$$

The standard deviations, σ_{xz}^2 and σ_{yz}^2, are equal and calculated from the Highland formula, Eq. 2.13.

Energy loss does not alter the trajectory of a particle directly, but the particle energy determines the amount of deflection from the Multiple Coulomb scattering in a plane. Energy loss in thin sensors is a highly non-Gaussian process. As the amount of energy loss and its variance is small compared to the total energy of the particle, the stochastic nature of energy loss does not greatly affect the amount of scattering that will occur. With no way of determining the amount of energy loss from the trajectories, all particles are assumed to have the same initial energy, and to lose the same amount of energy in each plane.

A prediction of the state vector in plane $k+1$ can be made from the true state in plane k.

$$\mathbf{x}_{k+1} = \mathbf{f}^{k}_{k+1}(\mathbf{x}^{true}_{k} + E(\boldsymbol{\omega}_k)) = \mathbf{f}^{k}_{k+1}(\mathbf{x}^{true}_{k}) \tag{3.4}$$

The uncertainties of the prediction are known from the linear error propagation of the uncertainties of the material interactions in plane k

$$\mathbf{C}_{k+1} = \mathbf{J}^{k}_{k+1} cov(\boldsymbol{\omega}_k) \mathbf{J}^{k\,T}_{k+1}, \tag{3.5}$$

where $\mathbf{J}^{k}_{k+1}$ is the 4×4 Jacobian matrix of the function $\mathbf{f}^{k}_{k+1}$.

$$\mathbf{f}^{k}_{k+1} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} \tag{3.6}$$

$$\mathbf{J}^{k}_{k+1} = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial (dx/dz)} & \frac{\partial f_1}{\partial (dy/dz)} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial (dx/dz)} & \frac{\partial f_2}{\partial (dy/dz)} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial (dx/dz)} & \frac{\partial f_3}{\partial (dy/dz)} \\ \frac{\partial f_4}{\partial x} & \frac{\partial f_4}{\partial y} & \frac{\partial f_4}{\partial (dx/dz)} & \frac{\partial f_4}{\partial (dy/dz)} \end{pmatrix} \tag{3.7}$$

The prediction of the state in plane $k+2$ becomes

$$\mathbf{x}_{k+2} = \mathbf{f}^{k}_{k+2}(\mathbf{x}^{true}_{k}) = \mathbf{f}^{k+1}_{k+2}\left(\mathbf{f}^{k}_{k+1}(\mathbf{x}^{true}_{k})\right). \tag{3.8}$$

To obtain the uncertainties of $\mathbf{x}_{k+2}$, the Jacobian of both $\mathbf{f}^{k}_{k+1}$ and $\mathbf{f}^{k+1}_{k+2}$, as well as the covariance matrix of $\boldsymbol{\omega}_k$ and $\boldsymbol{\omega}_{k+1}$, must be known.

$$\mathbf{C}_{k+2} = \mathbf{J}^{k+1}_{k+2}\left[\mathbf{J}^{k}_{k+1} cov(\boldsymbol{\omega}_k)\mathbf{J}^{k\,T}_{k+1} + cov(\boldsymbol{\omega}_{k+1})\right]\mathbf{J}^{k+1\,T}_{k+2}. \tag{3.9}$$

This can be implemented in a recursive fashion, as all the information needed to make a prediction in a plane $k+1$ is available in $\mathbf{x}_k$ and $\mathbf{C}_k$.

The modeling of multiple Coulomb scattering as a Gaussian distribution breaks down for electrons and positrons, which lose a large fraction of the particle energy through bremsstrahlung. In this case, the expected amount of scattering becomes non-Gaussian, with much longer tails than that of a Gaussian distribution.

3.1.2 Measurements and residuals

The measurement, $\mathbf{m}_k$, can be seen as the sum of the true position of the particle parameters projected into measurement space, $\mathbf{Hx}_k^{true}$, and a stochastic term, $\boldsymbol{\epsilon}_k$.

$$\mathbf{m}_k = \mathbf{Hx}_k^{true} + \boldsymbol{\epsilon}_k \tag{3.10}$$

The matrix $\mathbf{H}$ describes the projection from four-dimensional parameter space into two-dimensional measurement space. If the detector plane is orthogonal to the z-axis, the local coordinates of the measurement can be aligned with the global Cartesian coordinate system, and the projection simply becomes

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{3.11}$$

Assuming the measurements are unbiased, the expectation value of $\boldsymbol{\epsilon}_k$ is zero. The 2×2 covariance matrix of the measurement is $\mathbf{V}_k = cov(\boldsymbol{\epsilon}_k)$.

The residual vector describes the difference between the track prediction and a measurement in the plane. The residual vector in a plane can be calculated from an estimate of the track parameters that either includes the measurement in the plane, or does not. The residuals with respect to an estimate that includes the measurement in plane k is

$$\mathbf{r}_k = \mathbf{m}_k - \mathbf{Hx}_k, \tag{3.12}$$

with the corresponding covariance matrix

$$\mathbf{R}_k = \mathbf{V}_k - \mathbf{HC}_k\mathbf{H}^T. \tag{3.13}$$

The residual vector with respect to a track parameter estimate that does not include the measurement in plane k is

$$\mathbf{r}_k^* = \mathbf{m}_k - \mathbf{Hx}_k^*, \tag{3.14}$$

with the covariance matrix

$$\mathbf{R}_k^* = \mathbf{V}_k + \mathbf{HC}_k^*\mathbf{H}^T. \tag{3.15}$$

Vectors and matrices marked with the asterisk represent estimates obtained without including the measurement in plane k.

In real devices, the measurement errors can in some cases be systematic, and to a large degree be determined by the position of the particle in the measurement plane. This will be described in more detail in Section 5.10.

The measurements are defined to be on the front side of the planes, meaning they occur before the particle is scattered.

3.2 Track fitting by least-squares estimation

Ordinary least-squares estimation is done by finding the parameters that minimize the sum of squared residuals. For this method to be optimal, the residuals of all the measurements must have the same uncertainties. Due to the stochastic nature of the particle propagation this is generally not the case for a detector system, even if all measurement planes have the same resolution.

Estimating the track parameters in a reference plane can be done by finding the set of track parameters that minimize the sum of squared normalized residuals.

$$\chi^2 = \sum_{k=1}^{n} \mathbf{r}_k^T \mathbf{R}_k^{-1} \mathbf{r}_k \tag{3.16}$$

The residuals are here a function of the parameters in a reference plane, $\mathbf{x}_{ref}$.

$$\mathbf{r}_k = \mathbf{m}_k - \mathbf{H}\mathbf{f}_k^{ref}(\mathbf{x}_{ref}) \tag{3.17}$$

Due to scattering, the uncertainties in the parameter predictions will increase with the distance from the reference plane, leading to a down-weighing of measurements far from the reference plane where the parameters are estimated.

3.2.1 The Gauss-Markov theorem

The Gauss-Markov theorem states that the least-squares estimator is the linear, unbiased estimator with the lowest variance, if the following assumptions hold [37].

1. The measurements are unbiased.
2. All normalized residuals have the same variance.
3. The measurement errors are uncorrelated between planes.

The first assumption requires that the geometry of the sensor planes are correctly described. The resolution of modern semiconductor trackers can be in the order of a few µm, obtaining a good enough geometry description simply from measuring the positions of the devices during mounting is not enough, and numerical methods for geometry estimation are needed.

The second assumption requires that the covariance matrix of the residuals, $\mathbf{R}_k$, correctly describes the residual uncertainties. For this to be the case, the uncertainty of the measurements, as well as the uncertainties arising from particles interacting with matter, must be correctly described.

The third assumption requires that any systematic measurement errors do not cause correlations between planes.

In a complex experiment, such as the ATLAS detector, the track model is based on numerical methods and approximations, and is not truly linear. In the test beam without a magnetic field, however, the model is indeed linear.

If the Gauss-Markov assumptions hold, and the stochastic processes encountered are Gaussian, the parameters that minimize Eq. 3.16 cause each term in the sum to be a random number following the standard normal distribution. With uncorrelated errors, the expression becomes a sum of squared independent standard normal random numbers. Such a sum follows a χ^2 distribution characterized by its number of degrees of freedom, N_{dof}, which is the number of standard normal random numbers in the sum.

The number of degrees of freedom is $N_{\text{dof}} = N_m - N_p$, where N_m is the number of measurements and N_p is the number of estimated parameters, i.e. the dimension of the track parameter vector. A two-dimensional measurement in a detector plane increments N_m by two. If a fit is made to a number of measurements where N_{dof} is smaller than, or equal to, zero, the χ^2 of the fit will have a χ^2 of zero, meaning the fit passes through every measurement.

3.3 The Kalman filter

The Kalman filter is a recursive formulation of the least-squares estimator. It is based on Bayesian statistics, and works by reading in one measurement at the time, improving the estimate of the parameters as it moves through the detector. It can be used as a building block for track finding, outlier rejection, or dealing with non-Gaussian noise in measurements or material interaction [17, 26].

In order for the Kalman filter to work, it needs a track candidate, a list of all the measurements that are assumed to have been produced by the same particle.

The Kalman filter starts in the first or last detector plane with an initial guess of the track state. The measurement in the plane is then combined with the guess to produce an updated estimate of the track parameters at the plane. This updated state is used to predict the track parameters in the next measurement plane. The filter alternates between reading in a new measurement and predicting the state at the next measurement plane until all measurements are read in. The Kalman filter either runs in the forward direction, reading in measurements in the same order as the measurements were created by the particle, or in the backward direction.

The optimal estimate of the track parameters is obtained when all measurements have been included in the fit, meaning the last plane in a forward-running Kalman filter, or the first plane in a backward-running filter. Obtaining optimal estimates in the middle of a tracking detector requires a smoother that combines information from a forward running and a backward running Kalman filter.

3.3.1 Information filter formulation

The implementation that was used for reconstruction of the EUDET telescope data uses the information filter formulation of the Kalman filter. This is not the most common way of describing a Kalman filter in high-energy physics. The main differences from the standard formulation, are the way measurements are included in the track state, and the way the track state is described internally in the fitter. The details on how the measurements are included in the estimate will be discussed in the next section.

Instead of describing the track estimate as a parameter vector, $\mathbf{x}$, and its covariance matrix, $\mathbf{C}$, the information filter stores the track state in a weight matrix, $\mathbf{W} = \mathbf{C}^{-1}$, and an information vector, $\mathbf{i} = \mathbf{W}\mathbf{x}$.

In the very simple geometry of the test beam, the state of the track can be propagated in this implicit form. This offers some advantages in numerical stability, compared to the standard formulation. In more complex experiments, the explicit state vector is needed to look up the magnetic field, to obtain the amount of material at different positions along the path, and to calculate the trajectory length between measurement planes.

The information filter makes it possible to describe a complete lack of knowledge about the track state. The important part of the initial guess for the track estimates is not the initial guess of the track parameters, but the description of the uncertainties. The initial guess contains no information, and to reflect this it should have the covariance matrix

$$\mathbf{C}_0 = \begin{pmatrix} \infty & 0 & 0 & 0 \\ 0 & \infty & 0 & 0 \\ 0 & 0 & \infty & 0 \\ 0 & 0 & 0 & \infty \end{pmatrix}, \text{ or } \mathbf{W}_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.18}$$

Describing a zero matrix is a simple task for a computer, describing infinite uncertainties is not. Instead of infinite matrix elements, very large uncertainties are commonly used in the initial covariance matrix, but this will still give a small bias towards the initial guess of the track parameters. Numerical instabilities in the Kalman filter impose limits on how large the covariance matrix elements can be. These problems are avoided in the information formulation.

3.3.2 Including a measurement in the track estimate

Including a measurement in the track estimate is done by combining the information from a prediction of the track state in a plane with the information from the measurement. There are two equivalent ways of doing this.

The Kalman gain formalism, which is commonly used in high-energy physics, uses the gain matrix, $\mathbf{K}_k$, to weigh the measurement.

$$\mathbf{C}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{C}_{k|k-1} \tag{3.19}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k(\mathbf{m}_k - \mathbf{H}\mathbf{x}_{k|k-1}) \tag{3.20}$$

$$\mathbf{K}_k = \mathbf{C}_{k|k-1}\mathbf{H}^T(\mathbf{V}_k + \mathbf{H}\mathbf{C}_{k|k-1}\mathbf{H}^T)^{-1} \tag{3.21}$$

The updated state in plane k is described by $\mathbf{x}_{k|k}$ and $\mathbf{C}_{k|k}$, using all measurements up to and including the measurement in plane k. The predicted state in plane k, described by $\mathbf{x}_{k|k-1}$ and $\mathbf{C}_{k|k-1}$, uses all measurements up to and including the measurement in plane $k-1$. The measurement in plane k, $\mathbf{m}_k$, has the covariance matrix $\mathbf{V}_k$.

The weighted means formalism of the Kalman update is

$$\mathbf{C}_{k|k} = \left[\mathbf{C}_{k|k-1}^{-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{H}\right]^{-1} \tag{3.22}$$

$$\mathbf{x}_{k|k} = \mathbf{C}_{k|k}(\mathbf{C}_{k|k-1}^{-1}\mathbf{x}_{k|k-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{m}_k). \tag{3.23}$$

If the explicit state vector is needed, the gain matrix formalism is faster, but the weighted means formalism makes it easier to get an intuitive understanding of the process of including a measurement.

Assuming $\mathbf{H}$ is as in Eq. 3.11, we can project the measurement into the track parameter space.

$$\mathbf{x_m} = \mathbf{H}^T\mathbf{m}_k = \begin{pmatrix} m_x & m_y & 0 & 0 \end{pmatrix} \tag{3.24}$$

As the measurement does not contain any information about the direction of the particle, the covariance matrix would contain infinity elements. The inverse covariance matrix of the measurement in track parameter space is possible to obtain, though.

$$\mathbf{C}_\mathbf{m}^{-1} = \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{H} = \begin{pmatrix} \mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{3.25}$$

This is a block matrix, where $\mathbf{0}$ is the 2×2 null matrix. We also have that

$$\mathbf{C}_\mathbf{m}^{-1}\mathbf{x_m} = \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{H}\mathbf{H}^T\mathbf{m}_k = \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{m}_k, \tag{3.26}$$

so that the parenthesis in Eq. 3.23,

$$\mathbf{C}_{k|k-1}^{-1}\mathbf{x}_{k|k-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{m}_k = \mathbf{C}_{k|k-1}^{-1}\mathbf{x}_{k|k-1} + \mathbf{C}_\mathbf{m}^{-1}\mathbf{x_m}, \tag{3.27}$$

can be seen as the sum of two track states weighted by the inverse covariance matrix of the states. $\mathbf{x}_{k|k}$ is then the weighted mean of two track states, and $\mathbf{C}_{k|k}^{-1}$ is the sum of the weights.

For the information filter formulation, the update simply becomes

$$\mathbf{W}_{k|k} = \mathbf{W}_{k|k-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{H} \tag{3.28}$$

$$\mathbf{i}_{k|k} = \mathbf{i}_{k|k-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{m}_k. \tag{3.29}$$

With a diagonal $\mathbf{V}_k$, the update of the estimate consists of incrementing two elements in the weight matrix, and two elements in the information vector.

3.3.3 Material interactions

Propagating the track estimate to the next plane is done in two steps. The first step is taking the effects of the interactions with the matter in the current plane into account.

Including multiple scattering can be seen as propagating the estimate to the back side of a plane. If the state directly in front of the plane is described by $\mathbf{x}_k$ and $\mathbf{C}_k$, the state on the back side will be

$$\mathbf{x}_k^B = \mathbf{x}_k \tag{3.30}$$

$$\mathbf{C}_k^B = \mathbf{C}_k + \mathbf{U}^T\mathbf{Q}_k\mathbf{U}, \tag{3.31}$$

where $\mathbf{Q}$ is the 2×2 diagonal matrix containing the estimated uncertainties from scattering according to the Eq. 2.13, and U is the projection from state vector space to direction tangent space.

$$\mathbf{Q} = \begin{pmatrix} \sigma_{xz}^2 & 0 \\ 0 & \sigma_{yz}^2 \end{pmatrix}, \mathbf{U} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.32}$$

In the information filter, scattering can be added without inverting the weight matrix by using the Woodbury matrix identity [46].

$$\mathbf{W}_{k|k}^{B} = \left(\mathbf{C}_{k|k} + \mathbf{U}^T\mathbf{Q}\mathbf{U}\right)^{-1} \tag{3.33}$$

$$= \mathbf{W}_{k|k} - \mathbf{A}\mathbf{W}_{k|k}, \tag{3.34}$$

where

$$\mathbf{A} = \mathbf{W}_{k|k}\mathbf{U}^T(\mathbf{Q}^{-1} + \mathbf{U}\mathbf{W}_{k|k}\mathbf{U}^T)^{-1}\mathbf{U}. \tag{3.35}$$

As $\mathbf{i}_{k|k}$ depends on $\mathbf{W}_{k|k}$, it must also be updated.

$$\mathbf{i}_{k|k}^{B} = \mathbf{i}_{k|k} - \mathbf{A}\mathbf{i}_{k|k} \tag{3.36}$$

The definition of measurements being on the front side of a plane means that when a new measurement is included in a forward running filter at plane k, scattering for plane k has not yet been included in the estimate. For a backward running filter, it has. In reality, the measurement could be seen to be in the middle of the sensor, but as the detector plane consists of front ends and supporting structures, a much higher level of description of the geometry would be needed. It is doubtful that much would be gained from this in terms of tracking precision.

3.3.4 Predicting the track state in the next plane

The second step in propagating the estimate to the next plane, is the deterministic straight line propagation through air. This transformation of the track parameters is linear, and the transformation function can be expressed as a matrix operation.

$$\mathbf{x}_{k+1|k} = \mathbf{f}_k(\mathbf{x}_{k|k}^{B}) = \mathbf{F}_k\mathbf{x}_{k|k}^{B}, \tag{3.37}$$

The transport matrix, $\mathbf{F}_k$, for the straight line tracks is the Jacobian matrix of the transport function $f_k(\mathbf{x}_k^B)$.

$$\mathbf{F}_k = \mathbf{J}_{k+1}^{k} = \begin{pmatrix} 1 & 0 & \Delta z_k & 0 \\ 0 & 1 & 0 & \Delta z_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.38}$$

The element Δz_k is the propagation distance along the z-axis.

The propagation of the state covariance matrix is done by linear error propagation.

$$\mathbf{C}_{k+1|k} = \mathbf{F}_k \mathbf{C}^B_{k|k} \mathbf{F}^T_k \tag{3.39}$$

If Δz_k is known a priori, the propagation for the information filter state can be done without having to know the explicit track state vector.

$$\mathbf{W}_{k+1|k} = (\mathbf{F}_k \mathbf{C}^B_{k|k} \mathbf{F}^T_k)^{-1} = (\mathbf{F}^{-1}_k)^T \mathbf{W}^B_{k|k} \mathbf{F}^{-1}_k \tag{3.40}$$

$$\mathbf{i}_{k+1|k} = \mathbf{W}_{k+1|k} \mathbf{F}_k \mathbf{x}^B_{k|k} = (\mathbf{F}^{-1}_k)^T \mathbf{i}^B_{k|k} \tag{3.41}$$

The inverse transport matrix is the matrix performing a transformation in the opposite direction along the z-axis.

$$\mathbf{F}^{-1}_k = \begin{pmatrix} 1 & 0 & -\Delta z_k & 0 \\ 0 & 1 & 0 & -\Delta z_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.42}$$

The off-diagonal elements in $\mathbf{F}_k$ imply that the covariance matrix elements describing covariance between x and dx/dz as well as y and dy/dz become non-zero after any propagation step. The fact that the covariance or weight matrix contains the information about these correlations is the reason why the direction tangent estimates are updated when a new measurement is included in the fit, as will be shown in Section 3.3.7.

3.3.5 Test statistics

Studying the distribution of the fitted track χ^2 can give valuable information about the validity of the track model. If the distribution differs from the true χ^2 distribution with the same number of degrees of freedom, it is an indication that the Gauss-Markov assumptions are not fulfilled, or that the stochastic processes that are encountered are not Gaussian.

The χ^2 of the parameters in the final measurement plane can be calculated incrementally when fitting the track.

$$\chi^2 = \sum_{k=1}^{n} \mathbf{r}^T_{k|k-1} \mathbf{R}^{-1}_{k|k-1} \mathbf{r}_{k|k-1} \tag{3.43}$$

As the minimum residuals for a line passing through one or two measurements are zero, the χ^2 increment in the first and second measurement plane are also zero.

In addition to validating the track model, the fitted χ^2 value can be used to reject tracks that do not fit the hypothesis of the track model, tracks with a very large fitted χ^2. This can include tracks that are fitted from a track candidate that includes noise hits or hits from other tracks.

3.3.6 Smoother

To find the optimal estimates of the track state at every plane, the information from two information filters running in opposite directions are combined. This combination is analogous to the measurement update, Eq. 3.22 and 3.23, as it is a weighted mean of the two states where the weight is the inverse covariance matrix.

The state of the forward running filter is here denoted as $\mathbf{W}^{FW}_{k|k-1}$ and $\mathbf{i}^{FW}_{k|k-1}$, the backward running filter $\mathbf{W}^{BW}_{k|k+1}$ and $\mathbf{i}^{BW}_{k|k+1}$. Both these states describe the prediction in the plane without including the measurement.

$$\mathbf{C}^*_k = \left[\mathbf{W}^{FW}_{k|k-1} + \mathbf{W}^{BW}_{k|k+1}\right]^{-1} \tag{3.44}$$

$$\mathbf{x}^*_k = \mathbf{C}^*_k \left(\mathbf{i}^{FW}_{k|k-1} + \mathbf{i}^{BW}_{k|k+1}\right) \tag{3.45}$$

As the track state will be used for analysis, the explicit smoothed predictions is extracted, $\mathbf{C}^*_k$ and $\mathbf{x}^*_k$, where every measurement in the track except any measurement in plane k is included. A smoothed state that includes the measurement in plane k can be obtained by including the measurement in either the forward or the backward running filter before smoothing the states.

There also exists a gain matrix formalism for the smoother, which is faster if one uses the Kalman gain formalism for the fitter [16].

3.3.7 Visualizing the Kalman Filter

The evolution of the Kalman filter estimates can be described by visualizing the estimated track states as the Kalman filter moves through the detector system.

Only two parameters with non-zero covariance are visualized, x and dx/dz. The probability density function of the Gaussian track parameter estimate with the parameter vector $\mathbf{x}$ and covariance matrix $\mathbf{C}$ is

$$f(\mathbf{u}) = Ae^{-\frac{1}{2}(\mathbf{x}-\mathbf{u})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{u})}, \tag{3.46}$$

where

$$A = \frac{1}{2\pi\sqrt{|\mathbf{C}|}}, \tag{3.47}$$

and $|\mathbf{C}|$ is the determinant of the covariance matrix. The factor A is a normalization factor, the exponential factor gives the shape of the probability distribution.

Using the information filter formalism, the weight matrix, $\mathbf{W} = \mathbf{C}^{-1}$, is available at all times. However, the weight matrices from the information filter are non-invertible until two measurements or more have been included in the fit, as no information about the angle is available. With a non-invertible weight matrix, the determinant of the covariance matrix cannot be calculated, but the shape of the distribution can still be visualized.

Without an invertible weight matrix, the explicit parameter vector cannot be extracted from the information filter. The parameter estimate in the two first planes can be found by assuming the updated state in the first plane has an angle of zero and the position of the measurement in the plane. It can then be propagated to the next plane using Eq. 3.37.

As the measurement and its weigh metrix can be projected into parameter space, as in Eq. 3.24 and Eq. 3.25, the measurement can also be visualized in parameter space.

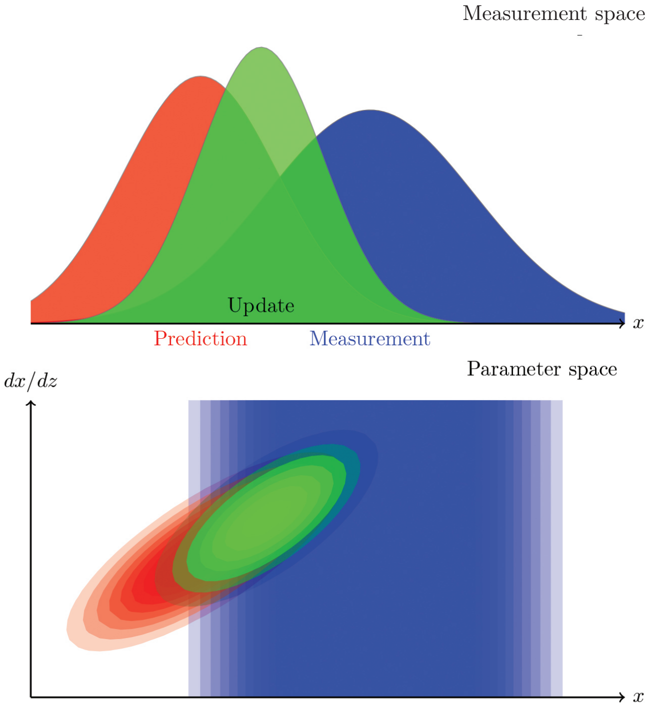
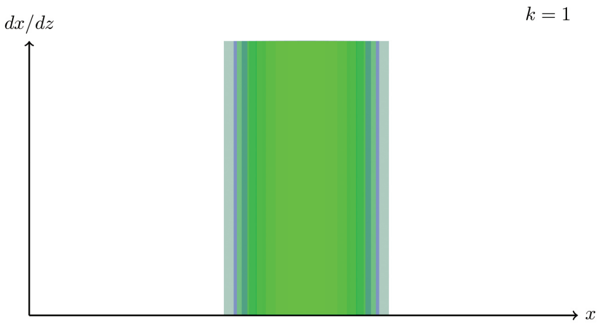
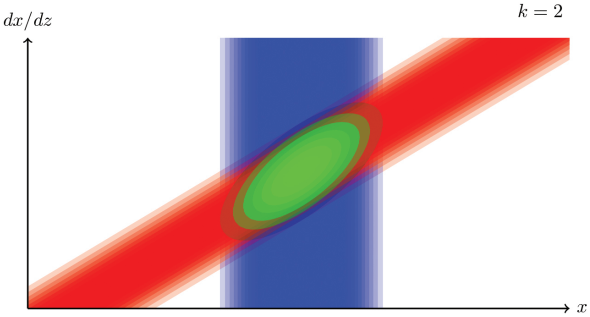


Figure 3.2: The probability densities of the measurement, prediction and updated state in both measurement space (top) and track parameter space (bottom). The measurement has a non-invertible weight matrix in parameter space, and infinite variance in dx/dz.

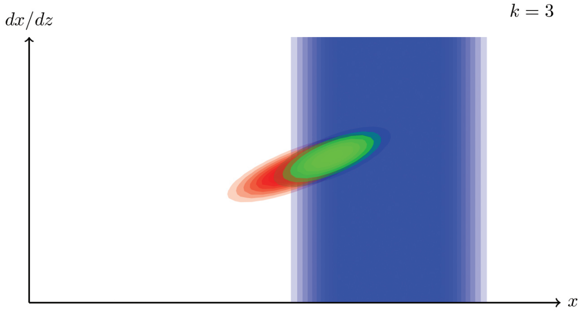
The probability density functions of the predicted state, the measurement and the updated state in a plane where more than two measurements have been included in the prediction is shown in Figure 3.2. The evolution of a Kalman filter in the first three planes is visualized in Figure 3.3. The propagation from one plane to the next is shown in Figure 3.4.



(a) A measurement projected into parameter space have well-defined uncertainties in the x direction, but infinite uncertainty in the angular direction. In the first plane, the updated state is exactly the same as the measurement.



(b) The prediction in the second plane has infinite uncertainties in both angle and position, but the correlation between the two is known due to off-diagonal elements in the transport matrix. When the measurement is included, the updated state gets an invertible weight matrix.



(c) The third plane, as well as every following plane, combines a well defined prediction with a measurement with infinite uncertainties in the angular parameters.

Figure 3.3: The evolution of the estimated track parameters for a two-dimensional Kalman filter through the three first planes.

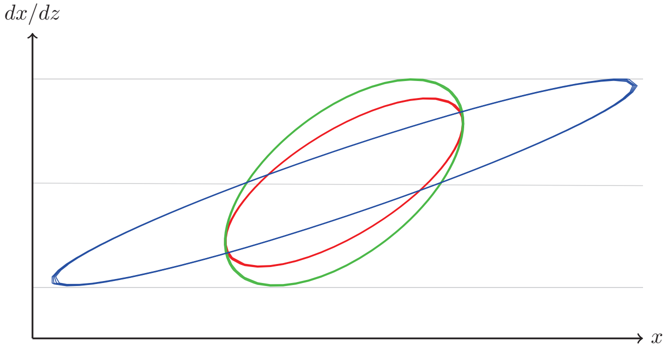


Figure 3.4: The evolution of the estimate as it is propagated from one plane to the next. The estimated track is parallel to the z-axis. The red estimate is at the front of plane $k > 2$. The green estimate is at the back of plane k. Including scattering increases the uncertainties in the angle, but not the uncertainties in the position. The blue estimate is at the front of plane $k + 1$. The linear error propagation increases the uncertainties in position, not the uncertainties in the angle.

Chapter 4

Test beam instrumentation

A test beam experiment characterizes devices under test by placing them in a beam of fast charged particles provided by a particle accelerator. Information about the trajectories of the particles are provided by a tracking detector called a beam telescope. A trigger system is needed to trigger readout to the telescope and the device under test.

The test beam area in the North Area of CERN is used for test beam characterization as a part of the research and development of new particle detector technology. A research and development group is awarded a time slot in a position in a beam line. In this slot, the experiment must be mounted, a sufficient amount of data must be taken for all the configurations of the experiment that are to be studied, and the experiment must be unmounted to make room for the next user.

Optimal usage of the time slot requires planning. The devices that are going into the beam, as well as the software to read out and monitor the data, must be tested and operational. Knowing which configuration to study and how many tracks that are needed for each configuration ensures that proper results are obtained. A workforce large enough to quickly mount the experiment, as well as continuously monitor the data taking, is needed.

4.1 The particle beam

The test beam experiments that took plance in the test beam hall in the North Area of CERN, used a beam produced by the CERN accelerator complex, illustrated in Figure 4.1. The SPS is injected with bunches of protons from the Proton Synchrotron (PS). The SPS accelerates the protons to 450 GeV, and then they are crashed into a target, creating secondary particles. The particles produced in the collisions go through a magnetic field where charged particles within a small momentum range are bent into a beam line. The particle beam goes through bending magnets, collimators and focusing magnets creating a narrow parallel beam. There are several beam lines in the test beam hall, and several experiments can use the same beam at the same time.

The beam in the test beam line is not constant in intensity in time, but comes in so called spills. When the SPS delivers a beam to the test beam area, or when it injects the bunches to the LHC, the accelerator is drained, and must again be injected with new protons.

The beam used in the testing of the ATLAS pixel devices at CERN consisted of π^+ mesons

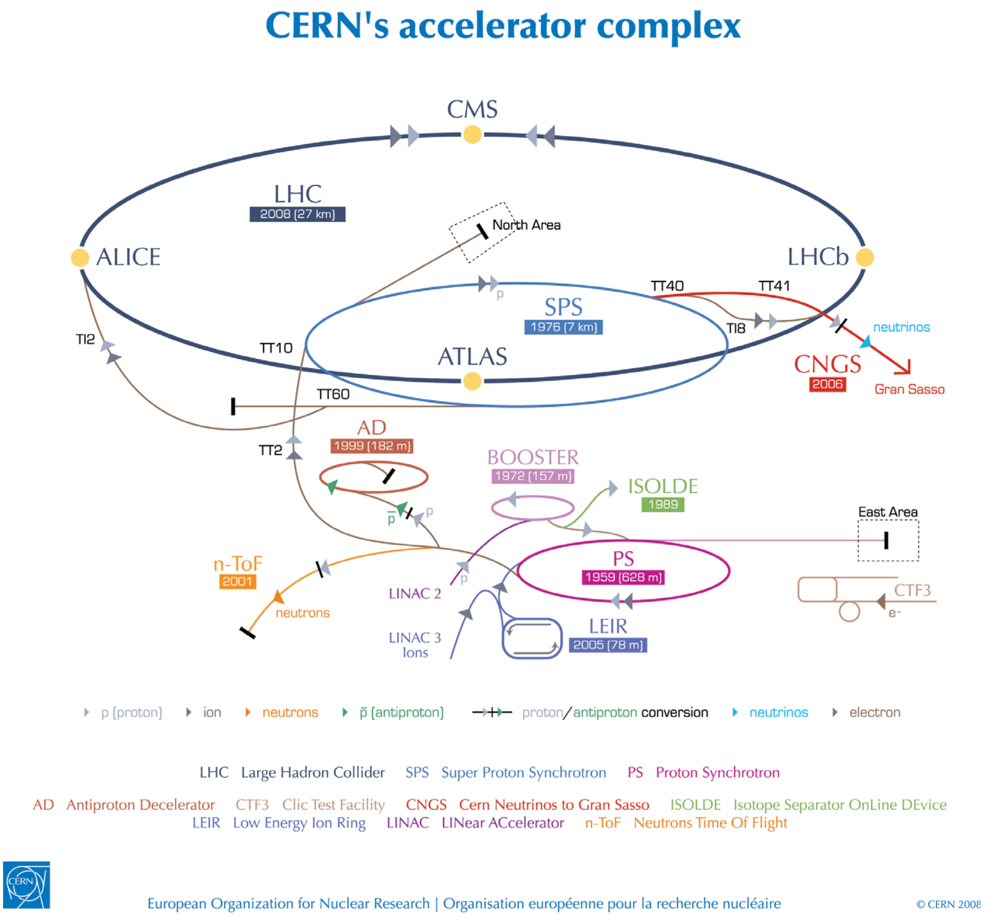


Figure 4.1: The CERN accelerator complex. The SPS provides accelerated protons to the North Area, where the test beam experiment was performed. CERN copyright.

with 120 or 180 GeV. The variance of the particle energies in the beam is small. After the beam shaping by magnets and collimators along the beam line, the angular variance is also small.

The width and intensity of the particle beam can be altered by adjusting collimators and magnets along the beam line. A high beam intensity can cause a too high trigger rate, which can cause problems for the data acquisition system. A low beam intensity leads to a low trigger rate, and inefficient use of the time slot.

Adjusting the magnets along the beam line can also move the beam spot, and can be used to focus it on the devices under test.

4.2 The Bonn ATLAS Telescope

The Bonn ATLAS telescope (BAT) [42], is a beam telescope made from three detector planes of two sided silicon micro strip sensors. Orthogonal n^+ and p^+ micro-strips are implanted on opposite sides, and both sides are read out and combined to give a two-dimensional estimate of the particle position. The detectors have a sensitive area of 32 mm × 32 mm, a thickness of approximately 300 µm, and a pitch of 50 µm on both sides. If a strip goes above threshold, the digitized signal is stored for the strip, as well as the two neighboring strips. This is done to improve the spatial resolution of the sensors by always allowing for charge interpolation. The standard deviations of the track parameters have been estimated to be approximately 6 µm at the position of the devices under test [39].

The devices under test were placed with two telescope planes on one side, and one on the other, as illustrated in Figure 4.2.

The system to trigger readout was based on two scintillators operated in coincidence. A third scintillator with a 15 mm diameter hole in the middle was placed downstream and operated in anti-coincidence, meaning that a signal from this scintillator would veto the trigger being issued. This was done to suppress any events with hadronic showering, and to narrow the region where triggers could be accepted to the region with the devices under test. The trigger logic was implemented with nuclear instrument modules (NIM).

The BAT can be operated within a magnetic field, and was used for tests inside the Morpurgo magnet situated in one of the beam lines [28].

4.3 The EUDET telescope

The EUDET telescope is a high resolution pixel telescope, developed by the EUDET and AIDA projects [35]. The telescope consists of two arms, each containing three pixel planes. The devices under test were placed between the arms, on a high precision XYϕ table, that can be controlled from outside the beam line.

The six telescope planes are sensors of the type Mimosa26. These are monolithic active pixel sensors (MAPS) with a 18.4 µm pitch. MAPS are silicon pixel devices with amplifiers integrated directly in each pixel. The resolution of the sensors depends on threshold settings in the electronics, but it is approximately 4-5 µm. The sensors consist of 576 × 1152 pixels, covering an area of 224 mm². The readout is binary, meaning no information about the amount

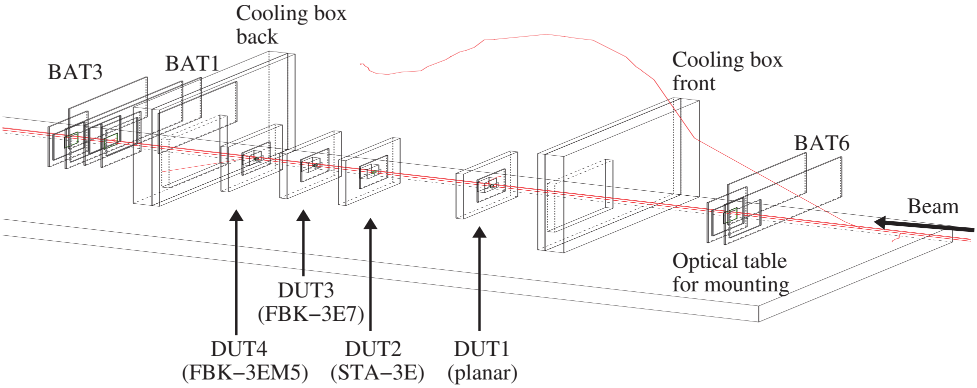


Figure 4.2: Illustration of the BAT telescope setup. The illustration is taken from [39], and shows the position of the telescope planes (BAT) and the devices under test (DUT)

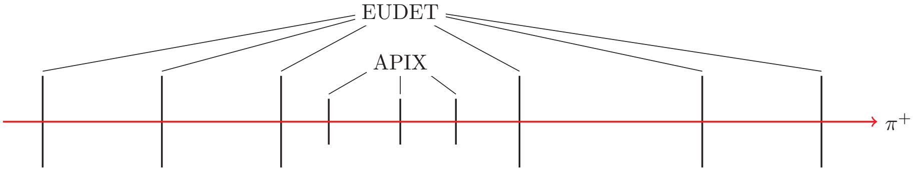


Figure 4.3: Schematic overview of the EUDET telescope with ATLAS pixel devices (APIX).

of deposited charge is stored, except that is went above threshold. Approximately 10 000 frames can be read out per second.

The telescope comes with a trigger system made from four scintillators operated in coincidence.

Compared to the BAT telescope, the EUDET telescope offers improved precision and allows for a much higher trigger rate. At the time of data taking, the EUDET telescope was not certified for use inside a magnetic field.

4.4 The devices under test

Devices of the full-3D type, as well as double side, double type column, have been tested. The devices under test in data that will be discussed in the following, are single chip assemblies. Single chip assemblies consist of a sensor bump bonded to a single front-end chip that is connected to the module control chip. Both sensors connected to FE-I4 and FE-I3 front-end chips have been tested. In addition to the 3D sensors, a planar pixel sensor of the type originally installed in the ATLAS inner detector has been used as a reference.

The devices under test were placed inside cooling boxes in the beam line. To prevent the formation of dew on the devices, nitrogen was supplied to the cooling box to avoid too high humidity. As dew can irreparably damage the electronics on the devices, the climate in the cooling box was monitored.

Two different data acquisition systems have been used to read out data from the devices under test in the test beam experiment. The VME based TurboDAQ, and the newer USB based USBPix. TurboDAQ works with FE-I3 devices, USBPix can work with both front-ends. The data acquisition systems have been integrated into the data acquisition systems of the telescopes, so synchronized data from both the telescope and devises under test is saved to the same file.

4.5 Test beam data taking

During data taking, the experiment is supervised in shifts to make sure the devices in the beam work, that triggers are accepted at a good rate, and that the data files produced by the data acquisition system are good.

The data acquisition systems come with online monitors, making it possible to study the signal response of the devices in the beam as the data are coming in, and to monitor the total number of triggers and the trigger rate. Correlation plots can be made between measurements in the different planes. This can be used to ensure that the data are synchronized, and that the devices are overlapping in the beam.

Shift work involves looking at data quality plots from the online monitor to make sure the detector planes produce the expected signal distributions, that all the channels in the device work, and that everything is well positioned in the beam. Shift workers also log any changes in the experimental setup, and mark any data files taken with bad beam conditions or faulty devices. Much of the data is taken as a part of different scans, for example scanning over different angles of inclination or different voltages applied over the junctions, or even different sensors. Shift workers keep track of the number of triggers that have been accepted in a given configuration, and make changes to the setup when enough data is taken.

The online monitors used in the test beams did not reconstruct tracks. Starting the reconstruction of tracks and full analysis of the data while taking data can further ensure that the system is able to produce the quality and rate of tracks that is expected.

Chapter 5

The test beam reconstruction chain

Before the data produced by a beam telescope like the BAT or EUDET can be used in the analysis of a device, the data must go through many processing steps, taking the data from a binary stream created by the data acquisition system to a parametric description of particle tracks. The data from the devices under test also go through these steps, as information about the devices are needed to optimize the description of the detector system.

The reconstruction chain consists of steps to create position estimates from the detector signals, finding the optimal description of the detector system, and extracting fitted tracks for analysis.

To get position estimates, the data stream from the experiment is decoded, noise hits in the data are removed, clusters of neighboring hits are identified and then used for the final position estimates. Obtaining the optimal description of the detector system consists of applying numerical methods to estimate the position and rotation of the detector planes, as well as the thickness of material layers in the setup and the resolution of the detector planes.

Track reconstruction consists of track finding, identifying sets of measurements that are created by the same particle, and track fitting, obtaining the optimal parametrized description of the particle trajectory.

5.1 Software

Several software frameworks and packages have been used or created to deal with the reconstruction of the ATLAS 3D silicon pixel test beam data.

- The EUDET telescope comes with an official reconstruction framework called EUTelescope. The framework processes the raw data all the way to analysis of the devices under test, and is written on top of the analysis framework Marlin. To make the framework live up to the high standards required in 3D pixel analysis, the alignment, pre-alignment, material and resolution estimation, as well as the final track reconstruction methods described in the following have been implemented and used in publications.

- Track reconstruction for the BAT telescope data was done with tbreco, a package written by Ole Røhne (University of Oslo). The published results from the data taken with the BAT have been reconstructed with this package.

- The analysis of the ATLAS 3D pixel test beam data was done within an analysis framework written for the experiment called tbmon. The framework reads files containing track information and decoded ATLAS pixel data created by either EUTelescope or tbreco.

The methods presented here are a mixture of the methods implemeted in the EUTelescope and tbmon frameworks. The pre-alignment, alignment, material estimation and track fitting are described as implemented in the EUTelescope package. The noise filtering, cluster finding and hit position estimation are descriptions of the implementations in tbmon.

5.2 Decoding the data

The data collected by the data acquisition systems are normally stored in a packed binary format in order to reduce bandwidth and storage requirements. The first step in the reconstruction of the particle tracks is therefore to decode this data and to translate it into a more accessible format. The devices that have been used in the experiments produce zero-suppressed data, meaning only information about channels going above threshold is retained. Both tbreco and EUTelescope store the decoded, zero-suppressed pixel and strip hits in instances of pixel or strip classes. In the tbreco package, this information is stored in memory for further processing, in EUTelescope the information is stored in a new file of the object oriented lcio format.

5.3 Noise filtering

A large amount of noise hits per readout trigger can multiply the amount of data the detector generates by a large factor. This can have an adverse effect on both the time needed to perform the reconstruction and the quality of the reconstructed tracks. If some channels in a device frequently produces noise hits, removing it from the reconstruction chain as early as possible is beneficial.

Removing these channels can be done by studying the occupancy of the channels. The occupancy of a channel is

$$\omega = N_{\text{hit}}/N_{\text{trig}}, \tag{5.1}$$

where N_{hit} is the number of times the channel has gone above threshold, and N_{trig} is the number of triggers the system has received. All channels with an occupancy above some cut-off value can be labeled as noisy, and removed from the data.

The occupancy of a channel depends both on rate of noise hits, and on the rate of hits created by a particle. If beam is not uniform in intensity across the sensors, the occupancy due to particle hits vary with position. An example of this can be found in the study of the active edges of the ATLAS 3D pixel devices presented in [36], where a narrow Gaussian beam aimed at the edge of the device under test was used.

The ATLAS pixel devices provide information about the time of each hit from the 25 ns bunch crossing clock. The distribution of the time of the hits with respect to the trigger is shown in Figures 5.1a and 5.1b. With timing information, one can define a window of time when the signal should arrive with respect to the trigger, and only consider the out-of-time occupancy,

$$\omega_{\text{ot}} = N_{\text{ot}}/N_{\text{trig}}. \tag{5.2}$$

Here N_{ot} is the number of hits occurring outside the time window where triggering particles are expected, divided by the number of triggers, N_{trig}. This method is less sensitive to spatial variations in the beam intensity. Figures 5.1c and 5.1d show the distribution of the out-of-time occupancy for the channels in two different ATLAS 3D pixel devices, one of them with several noisy channels.

The out-of-time occupancy estimated in Eq. 5.2 is not the real noise occupancy, the average number of noise hits per trigger, as noise hits occurring in the time window where triggering particles are expected to arrive are excluded. The noise occupancy can easily be estimated by scaling the out-of-time occupancy, by a factor $t/(t-t_{it})$, where t is the time window of the read out and t_{it} is the time window for particles that are considered in-time with the trigger.

Figures 5.1c shows that requiring an out-of-time occupancy of zero would exclude most of the channels in the device. Even a device with no clearly noisy channels will produce some noise hits.

5.4 Cluster finding

The position of the particle passing through the detector is estimated from clusters, not from the individual hits. A cluster is a group of one or more neighboring hits that are assumed to have been caused by the same particle.

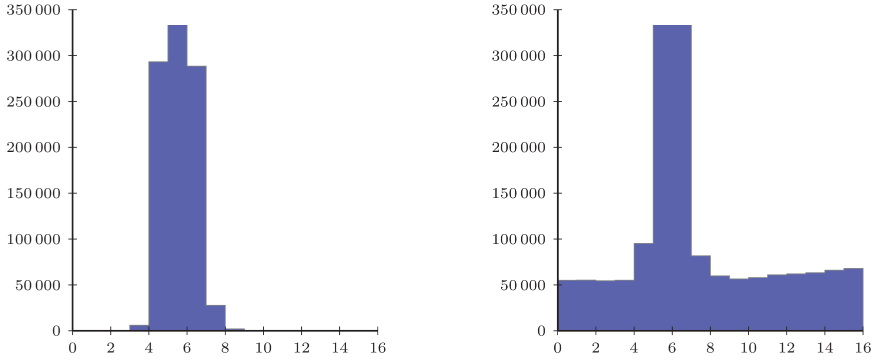
The method for cluster finding splits all hits after noise filtering into clusters. Initially all hits are stored in a list of available hits. Clusters are then created as follows.

- The first hit in the list of available hits is the start of a new cluster, and is removed from the list of available hits.
- For every hit that is added to the cluster:
  - All remaining available hits are compared to the new cluster hit.
  - If the available hit is connected to the new cluster hit, it is added to the cluster and removed from the list of available hits.
- When all hits in the cluster have been compared to the available hits, the cluster can no longer grow, and is saved.

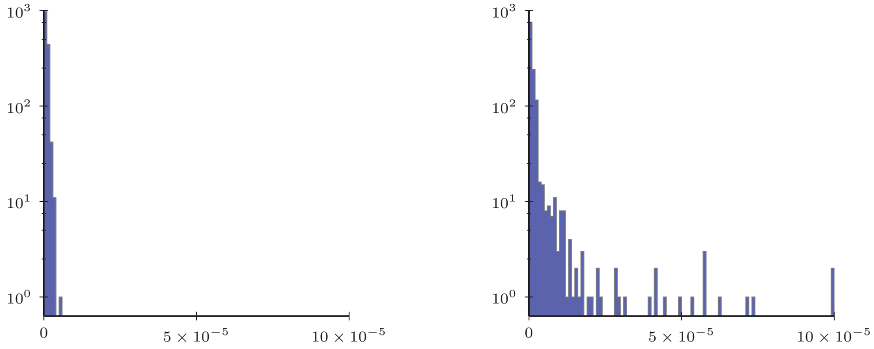
This process is repeated until there are no more available hits.

A hit is connected to a cluster hit if the distance between the two hits is smaller than a cutoff value. For the test beam results presented in Appendices C and D, hits that are direct neighbors horizontally, vertically or diagonally are connected.

Another possible test for inclusion of a hit is to compare the time of the hit to the average time of the hits in the cluster. This was not done for the results presented in Appendices C and D.



(a) Signal time (lv1) for hits in a non-noisy device. (b) Signal time (lv1) for hits in a noisy device.



(c) Out-of-time occupancy of a non-noisy device (d) Out-of-time occupancy of a noisy device

Figure 5.1: The left plots are of a 3D device with little noise, the right ones are of a noisy device. The two top plots show the distribution of the lv1, the time of the measurement in relation to the time of the trigger coincidence, counted by the internal 25 ns clock. Measurements caused by the triggering particle are assumed to come in the time window from three to ten clock ticks. The bottom plots show the distribution of the out-of-time occupancy, the occupancy of hits arriving outside of the time window. Note that the vertical axes have a logarithmic scale in the occupancy plots.

This is an agglomerative, connectivity based, single-linkage clustering algorithm. All available hits initially can be seen as individual clusters, that are then merged based on the connectivity. The size of the clusters are partially determined by the maximum allowed distance between the hits in the cluster.

In this case, all hits that are connected can form a cluster, and the clusters can grow to the final size without needing to recalculate and sort the distances between clusters for merging.

With the high density of particles in the ATLAS pixel detector, more advanced clustering algorithms can improve the performance of the tracker [33].

5.5 Hit position estimation

In the analysis of 3D pixel data, three methods for hit position estimation from clusters have been used and compared: the geometric mean, the charge weighted mean, and the corrected charge weighted mean. Which of these methods gives the best performance depends on the amount of charge sharing between neighboring channels, which in turn depends on the configuration of the experimental setup.

In test beam configurations where the devices are near orthogonal to the beam direction, particles that trigger hits in two channels tend to pass through the detector near the border between two channels, and particles that cause a single hit cluster must be assumed to pass through the center of the channel. At large tilt angles, this assumption is not valid.

5.5.1 The geometric mean

The simplest method for position estimation is the geometric mean, which is simply the average position of the hits in the cluster,

$$x = \frac{\sum\limits_{i=0}^{n} x_i}{n}, \tag{5.3}$$

where x is the estimated position, x_i is the center of the ith hit in the cluster, and n is the number of hits in the cluster. Pixel clusters contain information in two dimensions, but the two dimensions can be estimated independently.

If most of the clusters contain a single hit, with only a small fraction of two hit clusters, charge sharing happens in a narrow band around the border. The geometric mean of the two neighboring channels is right on the border, which is a good position estimate in this case.

5.5.2 The charge weighted mean

The charge weighted mean uses information about the pulse heights in the cluster channels to interpolate the hit position estimate.

$$x = \frac{\sum\limits_{i=0}^{n} w_i x_i}{\sum\limits_{i=0}^{n} w_i} \tag{5.4}$$

Here w_i is the signal strength of the ith hit.

If there is a large fraction of two hit clusters, the band around the border where charge sharing happens is wider. In this case, the charge weighted mean can give better position estimates than the geometric mean.

5.5.3 The corrected charge weighted mean

The position estimates from the charge weighted mean are often times not uniformly distributed within the pixel cells. If there is only a small fraction of two hit clusters, with a large amount of single hit clusters, the probability of estimating a hit near the border between two cells is low, and the probability of estimating hits near the center is high.

If the data contains mainly one or two hit clusters, the method that performs the best is the method of corrected charge weighted means [4]. This method is based on the assumption that the true position of the particles should be distributed uniformly within the pixel cells. The goal of the method is to find the transformation that when applied to the charge weighted means create a uniform distribution of position estimates.

The first step in achieving this, is to study the position of the charge weighted mean within the channels. This position is calculated as

$$\eta = x - \lfloor x \rfloor, \tag{5.5}$$

where x is the position estimate from the charge weighted mean method in units of pitch, and $\lfloor \rfloor$ is the floor function, the function that rounds downwards to the nearest integer. If the channels go from left to right, η is the distance between the charge weighted mean and the center of the channel to the left of the charge weighted mean. This is a number between 0 and 1 in units of pitch.

Figure 5.2a shows the distribution of η in the short direction for an ATLAS pixel device. From this distribution, the empirical cumulative distribution function ($eCDF$) can be obtained, as shown in Figure 5.2b. The cumulative distribution function is the function that if applied to all η will produce a uniform distribution.

The corrected charge weighted mean method simply applies this transformation to η.

$$x' = \lfloor x \rfloor + eCDF(\eta) \tag{5.6}$$

The empirical cumulative distribution function must be obtained in a calibration step before the corrections can be applied. This is done by obtaining the empirical probability distribution function of η in the discrete function $ePDF(x)$, which is simply the normalized histogram of the η distribution. The resulting empirical cumulative distribution is also a discrete function, and can be represented in a histogram.

$$eCDF(x) = \int_{-\infty}^{x} ePDF(x)dx \tag{5.7}$$

If parameters that affect the charge sharing of a device change, like the angle of inclination of the beam, the strength of the magnetic field, or the tuning of the front end electronics, this calibration must be redone.

All single hit clusters will have zero charge corrections, so there should be an asymmetry in Figure 5.2a. The probability of a particle passing exactly through the center of a pixel is essentially zero, and the large fraction of single hit clusters is created by the thresholds in the neighbor channels. One could imagine that 50% of the single hit clusters truly passed to the right of the channel, and should have a correction larger than 0. The other 50% passed to the left of the center, and should have a correction below 0. In the latter case, the wrong $\lfloor x \rfloor$ has been obtained, and the true correction should be close to 1. Simply moving half of the single channel cluster corrections from 0 to 1 restores the symmetry, and moves the region where charge sharing is assumed to happen to the border between the channels. The number of single hit clusters then determines the width around the center of the channel where no charge sharing is assumed to happen.

In real data there will be a smaller fraction of larger than 2 hit clusters, due to δ rays. If a δ ray is emitted in a direction that is close to parallel with the detector plane, it can trigger multiple channels. Such clusters will have large measurement uncertainties.

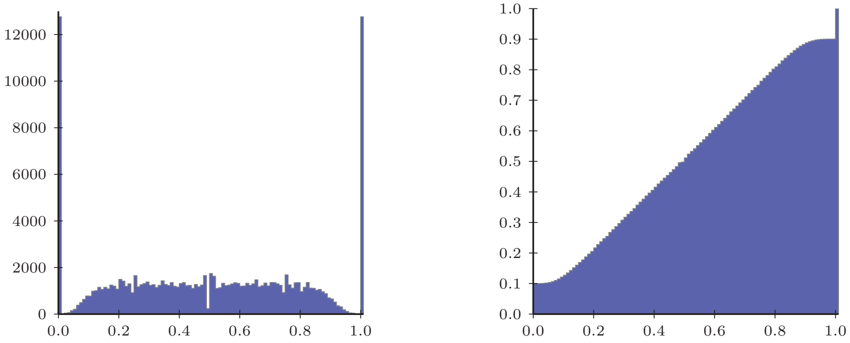
At high tilt angles, a large number of clusters can contain more than two hits. In this case, the basic assumption that charge sharing happens at the border between two hits no longer hold, and the charge weighted mean method can outperform the corrected charge weighted mean.

Figure 5.3 shows a comparison of the different methods using residuals from smoothed predictions and the estimated hit position of the cluster.

5.6 Global coordinate system, correlations and pre-alignment

When the experiment is being mounted, the position of all the detector planes, as well as any other material layer in the beam line between the first and last plane, are measured. This includes describing the rotations of the measurement coordinate system with respect to the global coordinate system. This information can be used to project the measurements in the detector planes into a global coordinate system. If for some reason the rotation of the detector planes are not known, correlation plots of the measurements in different planes can be useful in finding the correct rotations. Figures 5.4a and 5.4c show measurement correlations between two planes where the rotations are described correctly. In a collimated beam, the correlation of measurements of the x in the different planes should have a positive correlation, creating a correlation band with inclination near one.

For a beam that is collimated and traveling in a direction parallel to the z-axis of the detector



(a) Charge corrections

(b) eCDF

Figure 5.2: The distribution of charge corrections and the empirical CDF for a pixel device taking data with a beam at a 15° angle of inclination. The peaks at 0 and 1 in the charge corrections are due to single hit clusters.

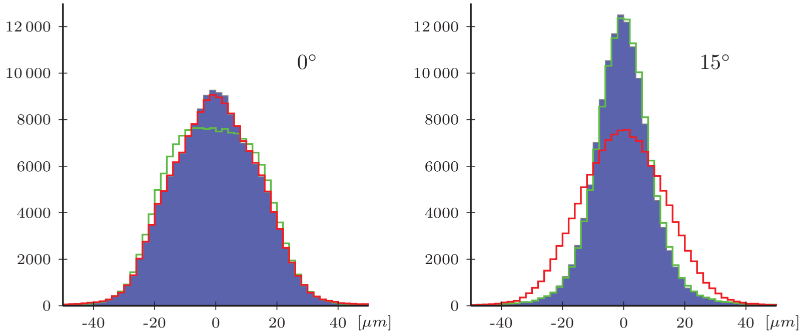
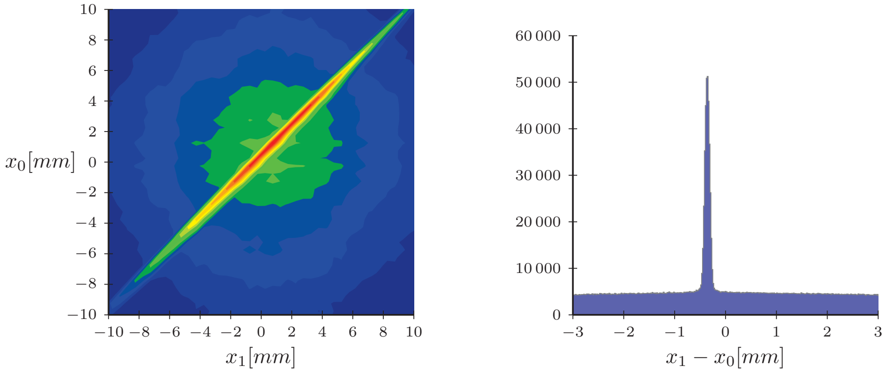
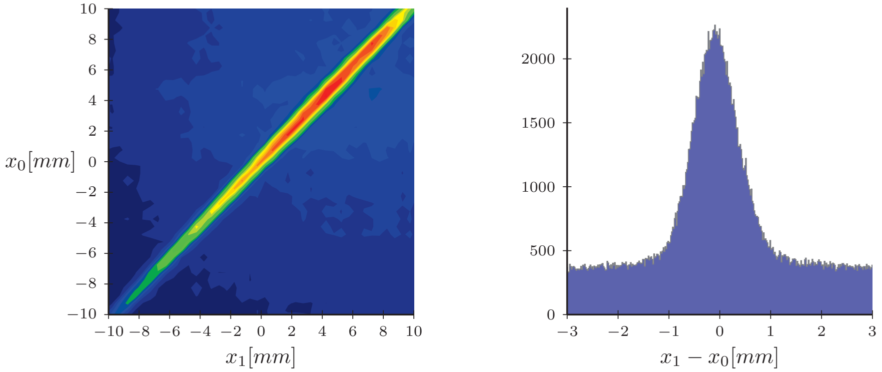


Figure 5.3: Residual distributions for charge corrected weighted means (🟦) , charge weighted mean (—) and geometric mean(—). for tracks with 0° and 15° angles of inclination without a magnetic field. In the 0° case, charge sharing happens in a narrow band around the pixel borders. In this case the geometric mean performs better than the charge weighted mean. In the 15° case the band is wider, and the charge weighted mean performs better than the geometric mean. The corrected charge weighted mean performs best in both cases by a small margin.



(a) x-position in plane 1 vs plane 0

(b) Shift in x-positions between the planes



(c) x-position in plane 1 vs plane 0

(d) Shift in x-positions between the planes

Figure 5.4: Plots showing the correlation of the measurements in the x-direction in EUDET plane 0 and 1. The two top plots are for 120GeV π^+, the two bottom plots are for 2GeV e^+ from a test beam performed at DESY. For the high energy beam, the shift in the x-positions can be used to improve the geometry description of the detector. For a low energy beam, where there is a lot of scattering, the estimated shifts are less likely to improve the geometry description.

system, the correlations between the planes can be used to improve the geometry description by translating the planes, if the position of the detector planes are not known to a high precision. The distribution of the difference between the position estimate in a reference plane and the correlated estimate in the plane to be aligned, should have a peak around the distance the plane is offset from the x-axis, as can be seen in Figure 5.4b and 5.4d. This pre-alignment step can make the remaining steps in the reconstruction chain simpler. If the detectors are noisy, or if there are many tracks per event, there will be noise in the distribution, and a robust peak finder is needed.

5.7 Track finding and fitting

The Kalman filter relies on a track finder for producing a list of hits assumed to have been created by the same particle. An optimal fit requires every measurement created by the track to be included in the list, and measurements not created by the track to be rejected.

The data read out after a trigger in the EUDET telescope rarely contains only a single measurement per plane. Several tracks can pass through the detector in the time window read out by a trigger, and the data can contain noise hits. The sensor planes have a less than 100% detection efficiency, and only considering tracks with measurements in all the planes would mean throwing away a large amount of data. To deal with track finding in this environment, several tools have been implemented.

The combinatorial Kalman filter is a track finder that uses Kalman filter predictions and updates to decide on including or rejecting measurements [26]. The decisions is based on the χ^2 of the difference between the prediction by the Kalman filter and the measurement in a plane, the same as the increments in Eq. 3.43.

The deterministic annealing filter (DAF) is an extension of the Kalman filter that improves the list of measurements that are fitted [17]. It is an iterative procedure that uses smoothed predictions and a weight function to make the decision to accept or reject measurements in the fit.

The cluster track finder is a new and very simple track finder that takes advantage of the parallel beam in the test beam experiment, and the ability of the DAF to deal with imperfections in the track finder. The method projects all measurements into the same measurement plane based on the beam angle, and groups all measurements that are close to each other based on an algorithm similar to the algorithm described in Section 5.4. All the clusters that are larger than a set number of hits are then fitted with the DAF. The details of these methods are presented in Appendix A.

One of the studies performed was a simulation experiment with perfect sensor efficiencies, with only one track per trigger, and a perfect description of the detector system. It was performed to test the implementation of the information filter. The results are repeated here to show perfect track quality plots.

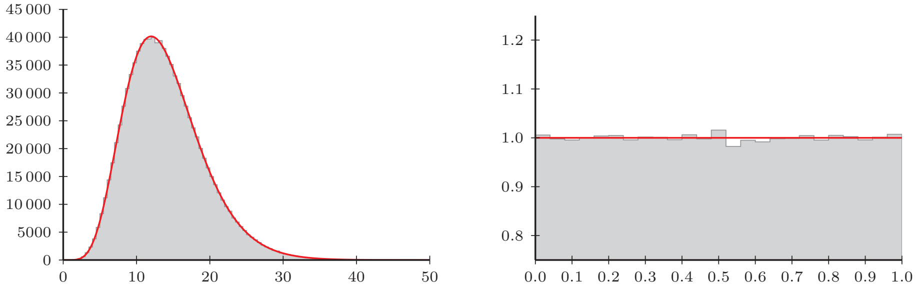
A p-value is the probability for observing a χ^2 value that is equal to, or greater than, the fitted χ^2 value given that the data truly is χ^2 distributed. If the data is χ^2 distributed, the p-values should form a uniform distribution from 0 to 1.

Pull distributions are one-dimensional residuals divided by their estimated standard deviations.

$$p_x = \frac{r_x}{\sigma_x}, p_y = \frac{r_y}{\sigma_y} \tag{5.8}$$

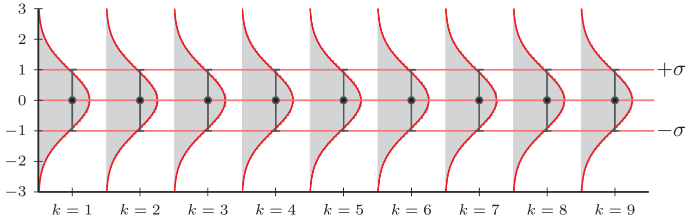
If the residuals follow a Gaussian distribution, and the estimated standard deviations are correct, the pull distributions should follow the standard normal distribution. Even if the residuals are not Gaussian, they should have a mean of zero and a standard deviation of one for the first and second Gauss-Markov assumption to be fulfilled.

The fitted χ^2, the p-values of the fitted χ^2, and pull distributions obtained in the simulation experiment are presented in Figure 5.5. These distributions are a good test that the Gauss-



(a) The distribution of the track χ^2 (▭), and the theoretical $\chi^2(N_{dof} = 14)$ (−).

(b) The p-values of the track χ^2 (▭), and the uniform distribution with the same area (−).



(c) Pull distributions in the x-direction in all the detector planes (▭). The black, vertical bars (I), show the mean and standard deviation of the pull distributions. The horizontal, red lines (−), show the mean and standard deviation of the standard normal distribution.

Figure 5.5: Track quality plots from 1 million simulated tracks, with a perfect geometry description. The empirical distributions are in gray (▭), the theoretical distributions are indicated in red (−).

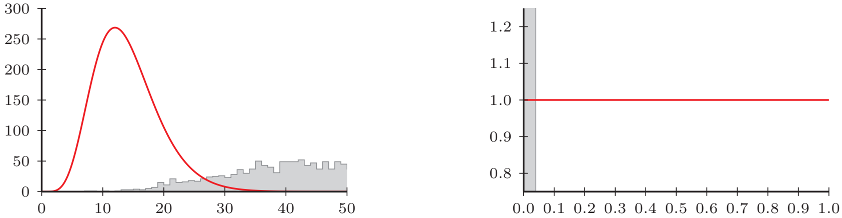
Markov assumptions hold. Large deviations from the theoretical distributions can be an indication that the tracker system is not properly described.

5.8 Alignment

The first Gauss-Markov assumption is that the measurements are unbiased. If the position or rotation of a detector plane in the tracking detector is not known exactly, it will lead to systematic measurement errors creating a measurement bias. Figure 5.6 shows the effects of such systematic errors on the pulls, χ^2 and p-values for real data.

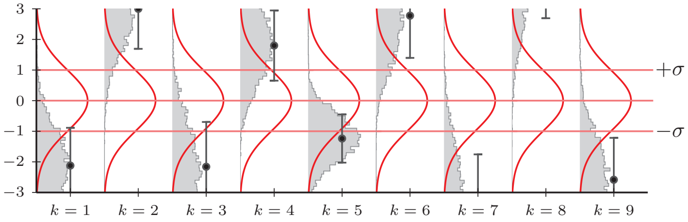
The geometry description starts with measuring the position of all the planes during the mounting of the experiment. Measurements obtained this way can be off by several mm. The resolution of the EUDET planes are in the order of a few µm, and the bias of the measurements due to errors in the geometry description must be small compared to this. To obtain high enough accuracy of the description, numerical methods for obtaining the optimal description of the geometry are needed.

Numerical alignment in the EUTelescope framework is performed with the program Millepede II.



(a) The distribution of the track χ^2 (▭) and the theoretical $\chi^2(N_{dof} = 14)$ (−).

(b) P-values for the fitted χ^2 (▭), and the uniform distribution with the same area (−).



(c) Pull distributions in the x direction for all the planes (▭), where the mean and standard deviations of the distributions are indicated with Ī. The horizontal lines (−) show the mean and standard deviations of the standard normal distribution.

Figure 5.6: Track quality plots for real 120 GeV π^+ data with a less than perfect geometry description. Only tracks with hits in all detector planes have been used to create the plots. The empirical distributions are in gray (▭), the theoretical distributions are indicated in red (−). This data has been misaligned by altering the geometry description of a well aligned system. In real data, without any numerical alignment, the description could be off by several hundred standard deviations in the pull distributions, making a visual representation hard.

5.8.1 The geometry description

The initial description of the geometry of each plane includes three parameters describing the x, y and z position of a reference point in the plane, and parameters describing the rotations around the x-, y- and z-axes. If this information is combined with information about the pitches of the device, a local measurement in units of pitch can be projected into a global coordinate system in units of µm.

The corrections to the positions of the measurement in the global coordinate system are described with two translations, two scale factors for the x- and y-axes, and a rotation around the z-axis.

$$x = x_0 x_{scale} + y_0 z_{rot} + x_{trans} \tag{5.9}$$

$$y = y_0 y_{scale} - x_0 z_{rot} + y_{trans} \tag{5.10}$$

$$z = z_0 \tag{5.11}$$

Here x_0, y_0 and z_0 are the measurement after they have been projected into the global coordinate system, x_{scale} and y_{scale} are the scale factors, x_{trans} and y_{trans} are the translations, and z_{rot} described the rotations around the z-axis.

Scale factors are used to avoid problems with using real rotations. Millepede II relies on information about the fitted residuals in x and y in the global coordinate system, and the partial derivatives of the residuals with respect to the alignment parameters. Corrections to the geometry description could have been described as rotations out of the global xy plane around axes parallel to the global axes through a reference point in the detector plane. If in this case the initial guess for the plane geometry was that it was parallel to the global xy plane, the partial derivatives of the residuals with respect to the rotations around the axes parallel to the global x- and y-axes would be $x_0 \sin(\theta_y)$ and $y_0 \sin(\theta_x)$ respectively, where θ_x is the rotation around an axis parallel to the x-axis passing through the reference point, and θ_x around the y-axis. With the initial guesses for the angles being zero, the residuals would be in a local minimum or maximum, and the partial derivatives would be zero. With a partial derivative of zero, the improvement by changing the parameter must be assumed to also be zero.

The geometries of the first and last planes have been fixed in the numerical alignment in order to define the global coordinate system. If the descriptions of the first and last planes assume that they are parallel to the global xy plane, but in fact they are not, then a plane in the middle of the detector that is truly parallel to the global xy plane could obtain scale factors larger than one. Such a correction is impossible to describe with rotations. Scale factors larger than one is an indication that global coordinate system is poorly defined, and if they are large the problem must be addressed. This can be done by fixing other planes in the system. The problem is illustrated in Figure 5.7.

Using scale factors for describing the rotations means that the z-positions of the measurements are not affected by corrections to the rotations around the x- and y-axes from numerical alignment. This leads to errors in the propagation distances in the Kalman filter. If the scale factors become large, meaning the initial guess for the detector rotation is poor, the initial projection of the measurements into the global coordinate system should be updated to avoid loss of precision.

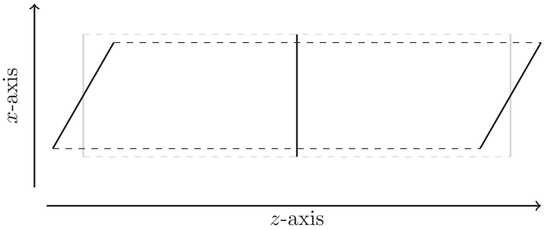


Figure 5.7: The coordinate system is defined by the outermost planes in the tracker system. If the planes are assumed to be parallel to the global x-axis as indicated by the gray vertical lines, but are in fact rotated around the y-axis, as illustrated by the black lines, the global x-axes will be compressed. The length and pitch of the plane in the middle of the system will appear larger than what they actually are if measured in the compressed coordinate system. In this case the numerical alignment will suggest a scale factor larger than one for the middle plane.

Due to the beam being collimated, estimating translation in the z-direction of the measurement planes numerically has proven to not be very robust, and measurements taken during the mounting have been used.

The implementation of the Kalman filter update uses the x and y of the numerically aligned measurement in the global coordinate system to describe the measurement. The z position of the measurement is used to calculate the initial Δz for the transport matrix, which can later be improved in the DAF iterations. This allows for simple transport and projection matrices for the fitter. If the measurement covariance matrix is defined in the local coordinate system, it must go through the same transformations to be projected into the global xy plane.

5.8.2 Numerical alignment

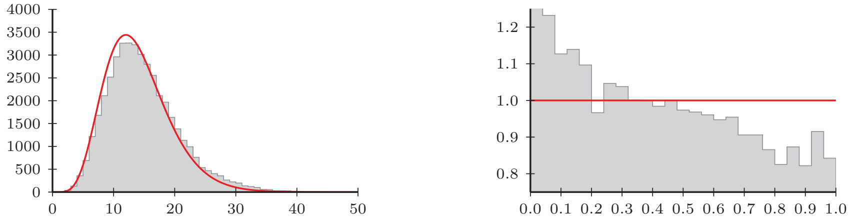
As the fitted χ^2 of each track is independent from the others, the sum of the χ^2 of the tracks in a sample is also χ^2 distributed. The number of degrees of freedom for the χ^2 of the track sample is the sum of the degrees of freedom for the individual track fits.

With a parametrized geometry, numerical alignment can be done by finding the geometry parameters that minimize the χ^2 of the track sample. For the test beam data, it is possible to do this by using general algorithms for multidimensional minimization, like the simplex algorithm of Nelder and Mead [29]. The simplex algorithm is an iterative method, that requires the entire track sample to be refitted for each iteration. This has been tested using the method as implemented by the GNU scientific library. The results presented in Appendix B used data that was aligned this way.

Even with the fast track fitter and simple geometry of the test beam data, iterative multidimensional optimization methods are too slow to deal with large amounts of data. The EUTelescope reconstruction chain uses the much faster program MILLEPEDE II[1] for alignment. Millepede II uses estimated residuals, and the partial derivatives of the residuals with respect to the geometry parameters and with respect to the track parameters to create a large system of linear equations that can be solved without iteration.

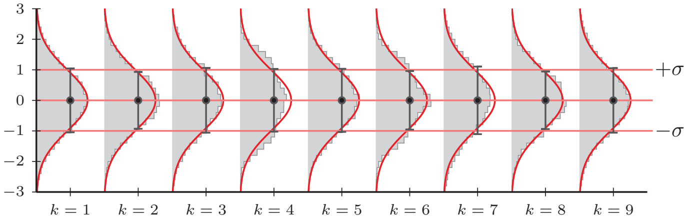
---

[1] see `http://http://www.desy.de/~blobel/mptalks.html`



(a) The distribution of the track χ^2 (▨) and the theoretical $\chi^2(N_{dof} = 14)$ (−).

(b) P-values for the fitted χ^2 (▨), and the uniform distribution with the same area (−).



(c) Pull distributions in the x direction for all the planes (▨), where the mean and standard deviations of the distributions are indicated with I. The horizontal red lines (−) show the mean and standard deviations of the standard normal distribution.

Figure 5.8: Track quality plots for real 120 GeV π^+ data after numerical alignment. Only tracks with hits in all detector planes have been used to create the plots. The empirical distributions are in gray (▨), the theoretical distributions are indicated in red (−). The fitted χ^2 values are slightly overestimated, leading to non-uniform p-values.

Misaligned measurements reduce the efficiency of a track finder for correctly accepting and rejecting measurements. Without pre-alignment, the cuts in the χ^2 increments used by the combinatorial Kalman filter might not be sufficient to achieve good track finding, and additional cuts in the residual distributions for the different planes might be needed. In the case of highly biased measurements, only tracks with measurements in all planes can be considered. The DAF does not work well with poor alignment.

A dilemma when aligning the system is whether to chose large or small data sets for performing alignment. Performing alignment on a large data set means the uncertainties in the estimated alignment parameters become smaller, but the data set can contain measurements taken over several hours. The actual position of the detector planes can change with time due to mechanical instabilities and these changes with time are not described by the chosen alignment parameters.

The best results were obtained by studying the evolution of the mean residuals within a time window as a function of time in a large data set aligned in a single step. This was used to determine the time window where the system was mechanically stable, and data from this time window was used in Appendix D. The mean of the residual distribution as a function of increasing run number is shown in Figure 5.9. The run number is a number identifying a data set stored to a file.

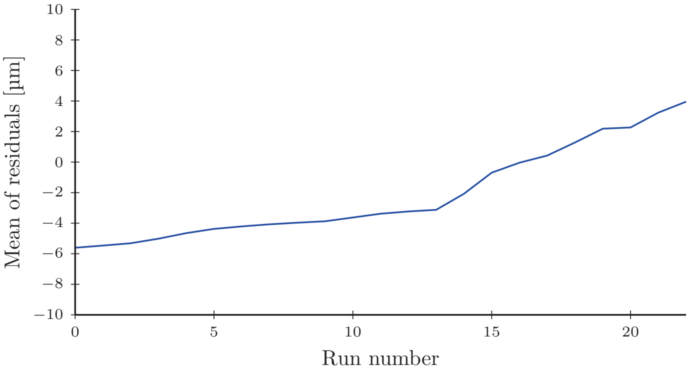


Figure 5.9: The mean of the residual distribution in the short pixel direction of an ATLAS pixel device as a function of increasing run number. The estimated geometry description is constant, but mechanical instabilities cause the planes to move as a function of time. Plots like this were used to determine when the experimental setup was most stable, and only stable runs were used for results that require high precision.

5.9 Material distribution and sensor resolutions

Figure 5.8 shows pull and χ^2 distributions for real data after numerical alignment. The distributions is closer to the theoretical distributions than in Figure 5.6, but not truly χ^2 distributed. This is a result of the second Gauss-Markov assumption not being fulfilled, the assumption that all the normalized residuals have the same variance.

The variances of the residuals depend on the amount of material the particle has passed through, the resolutions of all the measurements that have been included in the fit, and the propagation distance between the planes. If these parameters are not correctly described in the reconstruction, the normalization of the residuals will not be correct, and the Kalman filter will not perform optimally.

This leads to sub-optimal estimates of both the parameter vector and covariance matrix of the track, both of which are important in the analysis of the data.

Both the combinatorial Kalman filter and the DAF rely on the χ^2 of the difference between measurements and track predictions for the selection of hits in the detector system. If the observed χ^2 distributions vary from plane to plane, setting optimal cuts become hard.

Estimating the amount of material and the resolution of the planes in a beam telescope cannot be treated in the same way as alignment. The alignment methods rely on minimization of the χ^2 of the track sample. Increasing the estimated uncertainties of the measurements or the estimated amount of scattering in material planes would increase the estimated uncertainties of the residuals and minimize the χ^2 of the track sample. This is shown in Figure 5.10. Clearly the minimum χ^2 does not come from the correct estimates of measurement uncertainty and material distribution.

Novel numerical methods relying on a fast Kalman filter implementation and multidimensional minimization methods have been developed and tested on real and simulated test beam data for the estimation of these parameters. The methods and results are presented in Appendix B.

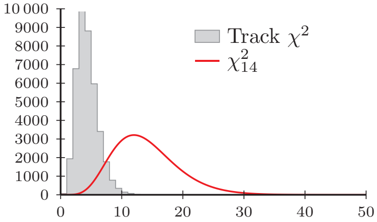


Figure 5.10: The distribution of the track χ^2 from simulated data (▭), and the theoretical $\chi^2(N_{dof} = 14)$ (−). The data was refitted with the standard deviation of the measurement errors and the amount of material in the planes overestimated by a factor two.

Figure 5.11 shows the pull and χ^2 distributions for real data after material and resolution estimation. The fitted χ^2 here follows the theoretical distributions very closely. The pull distributions are all unbiased and have unit variance.

In the geometry description, the thickness of measurement planes are described with a single number. If we assume that in reality the thicknesses are the same across the sensitive areas of the devices, but not outside it, the correct amount of material can only be estimated in the area where all the sensors overlap in the beam. A correct description of the areas that are not overlapping would require a more complex description of the material distribution in the planes.

All the measurements in a plane are in this model assumed to have the same uncertainties. The uncertainties of the measurements vary with the size of the cluster they are estimated from. It is likely that using this information could improve both the tracking resolution and the accuracy of the estimated covariance matrices.

Increasing the granularity of the description of the material distribution or the measurement resolutions would require more parameters to be estimated, requiring a larger track sample to obtain a sufficient estimate resolution.

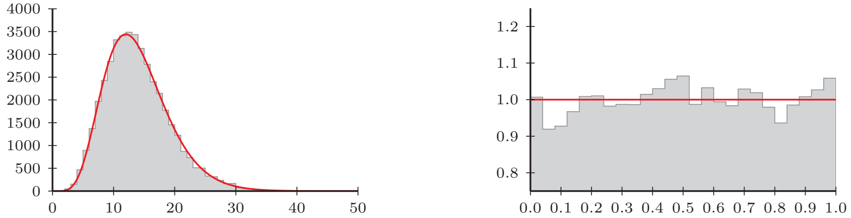
5.9.1 Material and resolution estimation in more complex experiments

The results presented in Appendix B come from estimating the material and resolution in the very simple geometry of the test beam. A complex modern high energy physics experiment would offer many challenges to the method that the test beam setup does not.

The main challenge is the speed of the multidimensional minimization procedure. A more complex detector would require more parameters to describe the system, which creates a requirement for larger track samples, and a larger number of iterations, in order to obtain estimates of sufficient resolution. The Kalman filter propagator for such an experiment becomes much more complex than for the test beam, as it needs to propagate through a magnetic field that varies in strength, and deal with more complex material distributions and interactions. The added complexity means that fitting a track require more floating point operations, and describing the system requires more memory.

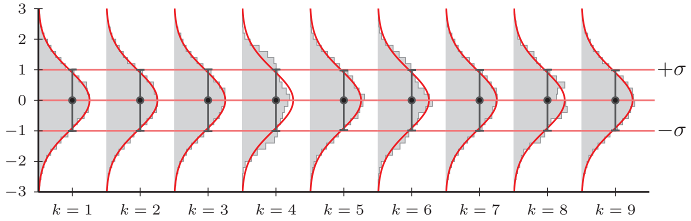
When track finding is solved, refitting a track sample falls into the category of embarrassingly parallel problems as refitting a track does not depend on information from any other tracks. Using modern multicore architectures, like a GPU, can potentially greatly accelerate the process. This requires that a track fitter can run effectively on a GPU, which might not be the case if random access to large databases describing magnetic fields or geometry are needed.

An implementation of the SDR2 method, presented in Appendix B, using openCL ran approximately 60 times faster on an NVIDIA GeForce GTX 460 compared to the original imple-



(a) The distribution of the track χ^2 (▨) and the theoretical $\chi^2(N_{dof} = 14)$ (−).

(b) P-values for the fitted χ^2 (▨), and the uniform distribution with the same area (−).



(c) Pull distributions in the x direction for all the planes (▨), where the mean and standard deviations of the distributions are indicated with I. The vertical red lines (−) show the mean and standard deviations of the standard normal distribution.

Figure 5.11: Track quality plots for real 120 GeV π^+ data after numerical alignment and estimation of resolution and plane thicknesses. Only tracks with hits in all detector planes have been used to create the plots. The empirical distributions are in gray (▨), the theoretical distributions are indicated in red (−). The fitted data is in good agreement with the theoretical distributions.

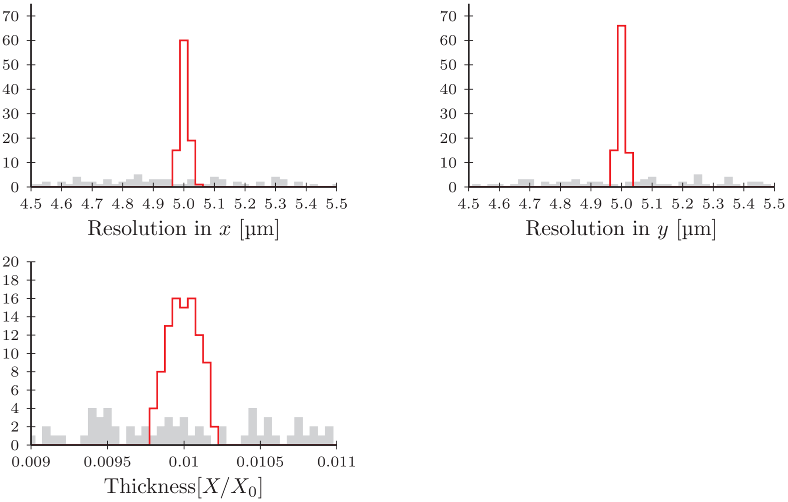


Figure 5.12: Estimated resolutions and thicknesses from a simulation experiment consisting of 500 000 pions of 40 GeV passing through 100 detector planes. All the planes have a resolution of 5 µm, and a thickness of 0.01 radiation lengths. All the thicknesses and resolutions were estimated simultaneously, except the thicknesses of the two first and two last planes, which were set to the true simulated parameters. The filled gray distribution shows the initial guesses, following a Gaussian distribution with a mean at the true simulated value, and a standard deviation equal to 10% of the mean. The red distribution shows the final estimates.

mentation on a 2.67GHz quadcode intel CPU. This GPU is a mid-range card from 2010. The openCL implementation was made to refit track candidates with hits in all planes, removing any branching. The entire system description, consisting of the z-positions, thicknesses and resolutions of the planes, is copied into the local memory of each work group. These optimizations would not be possible for a real physics experiment.

Another challenge is finding the correct minimum with a large number of parameters to estimate. Figure 5.12 shows that finding the correct minimum for a large number of parameters in a simulation experiment with a configuration like the test beam is possible. This is no guarantee it will work in other configurations or more complex geometries, with more parameters per detector element.

An advantage in a real physics experiment is the larger diversity in the track sample, with a range of energies and angles in particles. The uncertainties in the residuals that are studied in the methods depend on both the resolutions and the material effects in the beam. High energy tracks will be less scattered than low energy tracks, and the residuals will be more sensitive to errors in the resolution estimates. Tracks at large angles might traverse more material than perpendicular tracks, and the residuals will be more sensitive to errors in the estimated uncertainties from scattering. The presence of a magnetic field also would make information about the energy loss of the particles available.

Given a fast Kalman filter implementation, and a parametrized and mutable description of the detector, implementing the methods using an existing multidimensional minimizer is straight forward.

5.10 Error correlation

The third Gauss Markov assumption is that there is no correlation between measurement errors in different planes. This assumption does not necessarily hold for real data.

Eq. 3.10 describes the measurement errors in a plane with the stochastic variable ε. In a real detector, the measurement errors have contributions from processes that are not correlated between planes, like electrical noise and the creation and motion of electrons and holes. However, a particle passing through a pixel in a region where no charge sharing occurs, the measurement will be in the center of the pixel, and the errors are determined by the position of the particle in the pixel cell.

The ATLAS pixel cells are 50 µm×400 µm. Charge sharing in the 400 µm direction happens in a band that is very narrow compared to the pitch, and most measurements will be in the center of the 400 µm wide column.

The beam of particles in the test beam is collimated, with most particles traveling in a direction almost parallel to the z-axis. If the columns of two APIX detectors are aligned so they have a large overlap with each other in the beam, a situation as in Figure 5.13 can occur. In this case the errors are strongly correlated, and including the measurement in the track fit can reduce the resolution of the smoothed predictions.

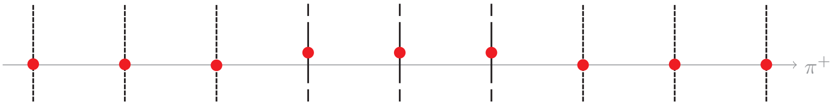
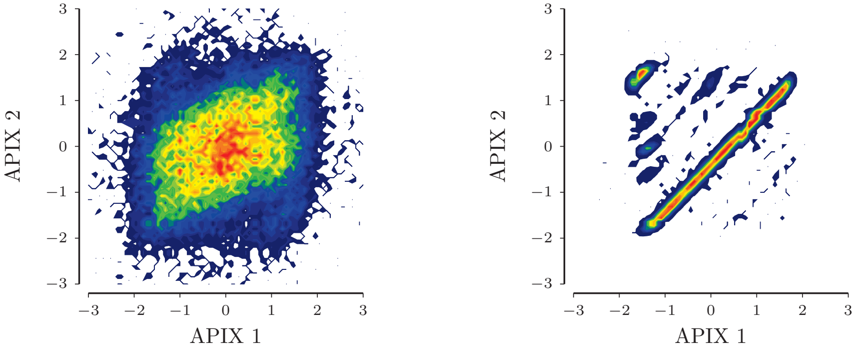


Figure 5.13: Illustration of error correlation in a detector system with three planes with little charge sharing surrounded by planes with better resolution. The vertical black liner represent readout channels, the red dots are the measurements and the horizontal line is the particle trajectory. The errors in the center planes are systematic and correlated in events with no charge sharing.

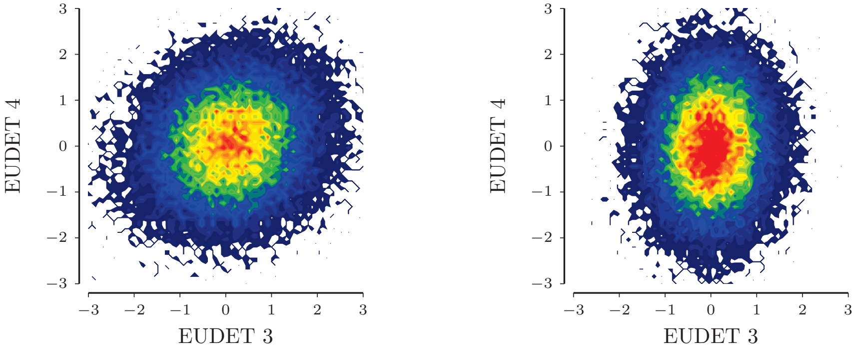
Figure 5.14 shows the error correlations between planes in the test beam experiment. In the 50 µm pixel direction there is a larger fraction of clusters that extend to more than one row, and the correlations between planes is not as pronounced.

5.11 The final track fit

When an acceptable description of the detector system has been obtained, the data is fitted one last time. The smoothed predictions are projected from the global xy plane into the two-dimensional coordinate systems of the detector planes. The track parameters, as well as information about all the hits in the devices under test are stored in a file, and the files are distributed to groups working on the analysis of the performance of the devices under test.



(a) Residual in the 50 µm ATLAS pixel direction. (b) Residuals in the 400 µm ATLAS pixel direction.



(c) y direction. (d) x direction.

Figure 5.14: The two top plots show the normalized residuals in the first APIX plane versus the second APIX plane. The two bottom plots show the residuals in the third EUDET plane versus the fourth. The residuals are calculated from smoothed predictions where both measurements are excluded from the fit. The plots are made from real data.

Chapter 6

Analysis of 3D pixel detectors

The test beam characterization of ATLAS 3D pixel sensors studied the performance of the devices in conditions similar to the ATLAS insertable B-layer (IBL), and demonstrated that the response of the devices are acceptable in these conditions.

With the insertion of the new B-layer, the ATLAS pixel system consists of two types of sensor technology, and understanding the difference in performance between the two is important. The main differences are due to the difference in drift of the charge cloud, the drop in detection efficiency in 3D sensors due to particles passing through the etched holes where the electrodes are implanted, and the active edges.

The analysis framework tbmon was developed for the test beam analysis of ATLAS 3D pixels, and has since been adopted by the IBL collaboration. The framework was designed to allow comparison of results from different groups working on analysis, and to share progress. The framework can process data from different experimental setups, using different beam telescopes and track reconstruction frameworks. The test beam results presented in Appendices C and D have been processed by this framework.

6.1 Detector characterization

Analyzing the ATLAS 3D silicon sensors require the same initial steps as preparing data for reconstruction of tracks. Noisy channels must be identified, clusters must be built and hit positions estimated. Information from tracking is used to match detector hits to particle tracks, to study the performance of the devices as a function of position in the sensor, and to estimate the spatial resolution of the measurements.

The main quantities that were studied in the experiments summarized in Appendices C and D are listed below.

- Detection efficiency of the device. The overall detection efficiency is calculated from tracks passing through the device, excluding tracks passing near the long or ganged pixels near the edge, or pixels that are flagged as noisy. An important study for 3D devices is the tracking efficiency as a function of the position within a pixel cell, as this shows the drop in efficiency in the etched out electrodes. The tracking efficiency as a function of position near the active edges is also an important measurement.

- The detector response. The sum of the charge buildup on the capacitors of the preamplifiers in the channels of a cluster should be proportional to the energy loss of the particle. The distribution of the sum of the time-over-threshold values should roughly show the distribution of the energy losses in the sensitive area of the detector. In 3D devices, this distribution differs from the distribution in a planar device due to tracks passing through the electrodes, which are not fully sensitive.

- Charge sharing behavior. The amount of charge sharing between channels in a cluster partially determines the resolution of the measurement. The readout channels have a threshold corresponding to a charge buildup of 3200 electrons that affects all the channels in a cluster. If a small signal is distributed over several channels, the probability rises that none of the channels go above threshold. Charge sharing is studied by studying the distribution of cluster sizes and the probability of having cluster sizes larger than one as a function of position within the pixel cell. When studying charge sharing as a function of position, tracks passing near the long pixels, the ganged pixels, or pixels that are flagged as noisy are excluded.

- The spatial resolution of measurements in the devices. Position estimates can be improved by charge interpolation techniques for clusters larger than one, so this is closely linked to charge sharing.

The performance depends on the behavior of the charge cloud, and vary with the direction of the particles with respect to the sensor plane, and the presence of a magnetic field. The devices in the IBL will detect particles coming in at angles from approximately 0 to -27° in a magnetic field of 2 T parallel to the long pixel direction. The test beam experiment studied the behavior of pixel devices both with and without a magnetic field of 1.6 T, tilting the devices so that particles came in at angles from -30 to 30°.

The results presented in Appendddix C are from the first test beam experiment with 3D detectors inside a magnetic field. The focus of this paper was to study the effect of the magnetic field on a 3D device with fully penetrating electrodes.

The results presented in Appendddix D are a summary of the test beam experiments that were performed during 2009. As the effect of the magnetic field on 3D devices is small, further tests were performed with the EUDET telescope without a magnetic field.

No sizable effect of a magnetic field with the same orientation as in the IBL on 3D sensors has been observed. With the configuration of the sensors in the IBL, the average cluster size will vary with the angle of inclination of the particles, which in turn vary with the position of the particle within the sensor. The charge cloud in a planar device will be de-focused for particles with small tilt angles, and focused for particles with higher tilt angles. Compared to planar sensors, 3D sensors have a relatively small amount of charge sharing for small tilt angles, and a relatively large amount of charge sharing for high tilt angles. The resolution of a device generally improves with cluster size, unless the clusters become very large.

For tracks that are orthogonal to the sensor plane, the overall detection efficiency of 3D devices is a few percent lower than in planar devices, due to low detection efficiency in the region of the etched electrodes. For tilted tracks, the efficiency is 99.9%, and comparable to the detection efficiency in planar sensors.

6.2 Efficiency counting with the EUDET telescope

The detection efficiency of a sensor is estimated by dividing the estimated number of times a particle creates a cluster in the sensor by the number of particles that has passed through the sensor in time with the readout. When calculating the overall efficiency, only tracks passing through the sensor away from the long or ganged pixels near the sensor edge and pixels flagged as noisy were considered.

When a trigger is issued, data from a time window of 16 × 25 ns is read out from the ATLAS pixel devices. The EUDET telescope reads out approximately 10 000 frames per second, which means a trigger contains data from a time window of approximately 100 µs. A track reconstructed from measurements in the telescope planes could have passed through the ATLAS device outside the time-window of the ATLAS readout. If the detection efficiency of the devices under test were to be estimated from all the tracks, the estimate would be artificially low.

The reconstruction software only considers tracks that are near a cluster in one or more of the devices under test to be in time with the readout, and only these tracks are stored for analysis.

This logic in the reconstruction software can cause another bias in the efficiency calculation. If only one ATLAS device had been in the beam, tracks that have not caused a cluster would be assumed to be out of time with the readout, and the estimated efficiency would always be 100%. With more than one device in the beam, the estimated efficiency would be 100% in the areas of the detectors that do not overlap in the beam.

The analysis framework only considers a track to be in time with a device under test if the track matches a cluster in one of the other devices under test in the beam. This means that the detection efficiency of the ATLAS pixel devices can only be estimated in the area where the devices overlap in the beam, as it is estimated from particles that pass through at least two devices under test.

The Bonn ATLAS telescope has a fast readout, so every track that is recorded in the telescope is in time with the ATLAS pixel readout and all reconstructed tracks are used in the efficiency estimation.

6.3 Covariance matrix of the track prediction

At the time of the study, accurate covariance matrices for the smoothed predictions were not available. The studies of the resolution in Appendix C do show the qualitative effect of the magnetic field on the sensor resolution, but do not estimate the measurement uncertainties.

The resolution of a device under test can be extracted from the width of the residual distribution if the covariance matrix of the track prediction at the plane correctly describes the tracking uncertainties.

The width of residual distributions are due to both measurement uncertainties and the uncertainties of smoothed prediction of the tracks. The efficiency as a function of position near the active edges or etched holes as observed in test beam data is also a convolution of the detector response and tracking uncertainty. To perform these studies, a high accuracy in the covariance matrix from the smoothed predictions are needed, relying on accurate estimates of the thicknesses and resolution of all the planes in the tracker system. The resolution of the track estimate varies as a function of z, and will be different for the different devices under test. Comparing convoluted resolutions or edge responses between devices directly gives limited information.

Numerical estimators, like the ones presented in Appendix B can be of help for obtaining estimated for the resolution of the devices under test.

6.4 The tbmon framework

Identifying noisy channels, building clusters, and estimating cluster hit positions have been redone in the tbmon framework instead of extracting the information from the files produced by the telescope reconstruction software. This was done to make sure the ATLAS pixel data is processed in exactly the same way in data taken with different telescopes. In addition to these tasks, the framework is used to produce histograms and calculate numbers of merit describing the detector performance.

The tbmon framework splits the task of analyzing the data task into building event objects, and the analysis of these objects. This is done to centralize certain decisions, and to simplify comparison of different data preparation methods.

An event object is created for each trigger, and contains information about the detector response of all planes in the system, as well as tracking information. The event object is created by a list of event builders. The first event builder reads data about the event from the file produced by the track reconstruction into the event object. Following event builders perform tasks such as noise filtering and cluster finding, as well as determine whether the track quality is good, and whether the track passed through sensitive material in the device on time with the readout.

After the event builders have prepared the data, analysis jobs that fill histograms and calculate figures of merit are ran. Analysis jobs can also generate calibration files, for example the empirical cumulative distribution function used to obtain the corrected charge weighted mean discussed in Section 5.5, or the channel occupancies used for noise filtering.

The modular event builders make it possible to run the same analysis jobs on data from different beam telescopes, and be sure that the same logic is used when processing the data, as this would be harder to guarantee if the performance analysis was implemented in the different reconstruction frameworks. It also makes it easy to compare different methods for the processing steps. If two event builders perform the same task, the results can be compared by running the same analysis job on the event objects produced by the different builders.

Decision making that can affect several analyses is pushed to the event builders or made in utilities provided by the framework. Examples of this are whether or not a track matches a cluster or what the best estimate position of a cluster is. Centralizing these decisions makes comparing results from different groups working on analysis simpler.

Chapter 7

Outlook and conclusions

Due to high radiation tolerance, ATLAS 3D silicon sensors are used in the ATLAS insertable B-layer, a new fourth layer of the pixel subsystem of the ATLAS detector at CERN. Test beam experiments have been performed to demonstrate that the novel sensor technology performs acceptably in conditions similar to the ATLAS insertable B-layer.

In this thesis the methods used for track reconstruction and analysis of data taken in the test beam experiments have been described. Results of the characterization of 3D devices have been presented.

The performance of the pixel devices has been tested at different angles both with and without the presence of a magnetic field. For 3D devices with electrodes penetrating the entire sensor wafer, the effect of a magnetic field in the same direction as the field in the IBL has been shown to be small.

The etched electrodes in the 3D design means that particles passing through the electrodes will generate less signal than tracks passing through the depleted silicon. For sensors that are orthogonal to the beam direction, a drop in detection efficiency of a few percent is observed in full 3D sensors. The efficiency drop is smaller in devices where the electrode do not penetrate the entire substrate. For devices tilted to angles similar to what will be observed in the insertable B-layer, the efficiency is very close to 100%.

Track reconstruction methods have been implemented for processing of data taken with the EUDET telescope without a magnetic field. Different methods have been tested, and compared in simulation experiments. An implementation of the combinatorial Kalman filter that takes advantage of the constraints of the collimated beam followed by the Deterministic Annealing Filter for track fitting gives excellent tracking efficiency and precision. A new track finder that is simple to implement has been introduced, called the cluster track finder. The cluster track finder produces crude track candidates, and it relies on the adaptive Deterministic Annealing Filter to refine the candidates. The tracking efficiency of the cluster track finder followed by the Deterministic Annealing Filter is lower than the combinatorial Kalman filter, but the performance is acceptable at low noise densities.

New methods for simultaneous estimation of detector resolution and the amount of material in the detector planes have been introduced. The methods are able to give accurate estimates in a simulation experiment with conditions similar to the test beam experiment. The methods have been able to improve tracking precision in real data, and give excellent track quality. A weakness of these methods is that they rely on numerical minimization methods requiring the track sample to be refitted several thousand times. When track finding is solved, refitting a track sample is an embarrassingly parallel problem, and implementing the fitter on a multicore system, like a GPU, can potentially greatly speed up the process. With a large number of parameters to estimate, finding the true minimum could prove to be a problem.

Following the studies presented in Appendixes C and D, several more test beam experiments have been performed on 3D and planar devices for the IBL both at CERN and in a positron beam at DESY. This includes tests of irradiated devices [27] and devices connected to FEI4 front-end chips [41]. The ATLAS 3D devices are used in the regions of the IBL furthest away from the interaction point, and tests where devices are tilted in the orthogonal direction to the azimuthal angle have been performed [41]. Many of the results presented in these publications have relied on the analysis framework and track reconstruction methods presented here. The software has also been used for data analysis in several other publications, presentations and theses, including [13, 19, 44, 30, 11, 39].

Bibliography

[1] G. Aad et al. The ATLAS experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[2] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012.

[3] N. Arora, J. R. Hauser, and D. Roulston. Electron and hole mobilities in silicon as a function of concentration and temperature. *Electron Devices, IEEE Transactions on*, 29(2):292–295, Feb 1982.

[4] E. Belau et al. Charge collection in silicon strip detectors. *Nucl. Instrum. Meth.*, 214:253–260, Sep 1983.

[5] J. Beringer et al. Review of Particle Physics (RPP). *Phys.Rev.*, D86:010001, 2012.

[6] H. Bethe and W. Heitler. On the Stopping of fast particles and on the creation of positive electrons. *Proc.Roy.Soc.Lond.*, A146:83–112, 1934.

[7] H. A. Bethe. Molière's theory of multiple scattering. *Phys. Rev.*, 89:1256–1266, Mar 1953.

[8] G. F. D. Betta, M. Boscardin, G. Darbo, C. Gemme, A. La Rosa, H. Pernegger, C. Piemonte, M. Povoli, S. Ronchin, A. Zoboli, and N. Zorzi. Development of 3D-DDTC pixel detectors for the ATLAS upgrade. *Nucl. Instrum. Methods Phys. Res., A*, 636(arXiv:0910.3629):S15–S23, Oct 2009.

[9] H. Bichsel. Straggling in Thin Silicon Detectors. *Rev.Mod.Phys.*, 60:663–699, 1988.

[10] H. Bichsel. A method to improve tracking and particle identification in TPCs and silicon detectors. *Nucl.Instrum.Meth.*, A562:154–197, 2006.

[11] M. Borri. *Characterization of 3D Silicon Assemblies for ATLAS Pixel Upgrade*. PhD thesis, The University of Manchester, 2013.

[12] M. Brigida, C. Favuzzi, P. Fusco, F. Gargano, N. Giglietto, F. Giordano, F. Loparco, B. Marangelli, M. Mazziotta, N. Mirizzi, S. Rainò, and P. Spinelli. A full monte carlo simulation code for silicon strip detectors. *Nuclear Physics B - Proceedings Supplements*, 2006.

[13] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva, Sep 2010.

[14] G. Casse, P. Allport, and A. Watson. Effects of accelerated annealing on p-type silicon micro-strip detectors after very high doses of proton irradiation. *Nucl.Instrum.Meth.*, A568:46–50, 2006.

[15] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012.

[16] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Reasearch A*, 262:444–450, 1987.

[17] R. Frühwirth and A. Strandlie. Track fitting with ambiguities and noise: A study of elastic tracking and nonlinear filters. *Computer Physics Communications*, 120:197–214, 1999.

[18] H. Gjersdal, A. Strandlie, and O. Røhne. Straight line track reconstruction for the ATLAS IBL testbeam with the EUDET telescope. Technical Report ATL-INDET-PUB-2014-003, CERN, Geneva, Jun 2014.

[19] P. Hansson, J. Balbuena, C. Barrera, E. Bolle, M. Borri, M. Boscardin, M. Chmeissan, G.-F. D. Betta, G. Darbo, C. D. Via, E. Devetak, B. DeWilde, D. Su, O. Dorholt, S. Fazio, C. Fleta, C. Gemme, M. Giordani, H. Gjersdal, P. Grenier, S. Grinstein, J. Hasi, K. Helle, F. Huegging, P. Jackson, C. Kenney, M. Kocian, I. Korolkov, A. L. Rosa, A. Mastroberardino, A. Micelli, C. Nellist, P. Nordahl, F. Rivero, O. Rohne, H. Sandaker, D. Silverstein, K. Sjoebaek, T. Slaviec, J. Stupak, I. Troyano, J. Tsung, D. Tsybychev, N. Wermes, and C. Young. 3d silicon pixel sensors: Recent test beam results. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 628(1):216 – 220, 2011. {VCI} 2010 Proceedings of the 12th International Vienna Conference on Instrumentation.

[20] T. Hemperek, D. Arutinov, M. Barbero, R. Beccherle, G. Darbo, S. Dube, D. Elledge, D. Fougeron, M. Garcia-Sciveres, D. Gnani, V. Gromov, M. Karagounis, R. Kluit, A. Kruth, A. Mekkaoui, M. Menouni, J. Schipper, and N. Wermes. Digital architecture of the new ATLAS pixel chip FE-I4. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, pages 791–796, Oct 2009.

[21] V. Highland. Some practical remarks on multiple scattering. *Nuclear Instruments and Methods in Physics Reasearch*, 129:497–499, 1975.

[22] C. Kenney, S. Parker, J. Segal, and C. Storment. Silicon detectors with 3-D electrode arrays: fabrication and initial test results. *Nuclear Science, IEEE Transactions on*, 46(4):1224–1236, Aug 1999.

[23] O. Krasel. *Charge Collection in Irradiated Silicon Detectors: A Study of the Operation Conditions of Silicon Sensors in the ATLAS Pixel Detector*. PhD thesis, University of Dortmund, 2004.

[24] L. Landau. On the energy loss of fast particles by ionization. *J.Phys.(USSR)*, 8:201–205, 1944.

[25] W. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-To Approach*. U.S. Government Printing Office, 1994.

[26] R. Mankel. A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 395(2):169 – 184, 1997.

[27] A. Micelli, K. Helle, H. Sandaker, B. Stugu, M. Barbero, F. Hügging, M. Karagounis, V. Kostyukhin, H. Krüger, J.-W. Tsung, N. Wermes, M. Capua, S. Fazio, A. Mastroberardino, G. Susinno, C. Gallrapp, B. D. Girolamo, D. Dobos, A. L. Rosa, H. Pernegger, S. Roe, T. Slavicek, S. Pospisil, K. Jakobs, M. Köhler, U. Parzefall, G. Darbo, G. Gariano, C. Gemme, A. Rovani, E. Ruscino, C. Butter, R. Bates, V. Oshea, S. Parker, M. Cavalli-Sforza, S. Grinstein, I. Korokolov, C. Pradilla, K. Einsweiler, M. Garcia-Sciveres, M. Borri, C. D. Vià, J. Freestone, S. Kolya, C. Lai, C. Nellist, J. Pater, R. Thompson, S. Watts, M. Hoeferkamp, S. Seidel, E. Bolle, H. Gjersdal, K.-N. Sjoebaek, S. Stapnes, O. Rohne, D. Su, C. Young, P. Hansson, P. Grenier, J. Hasi, C. Kenney, M. Kocian, P. Jackson, D. Silverstein, H. Davetak, B. DeWilde, D. Tsybychev, G.-F. D. Betta, P. Gabos, M. Povoli, M. Cobal, M.-P. Giordani, L. Selmi, A. Cristofoli, D. Esseni, P. Palestri, C. Fleta, M. Lozano, G. Pellegrini, M. Boscardin, A. Bagolini, C. Piemonte, S. Ronchin, N. Zorzi, T.-E. Hansen, T. Hansen, A. Kok, N. Lietaer, J. Kalliopuska, and A. Oja. 3D-FBK pixel sensors: Recent beam tests results with irradiated devices. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 650(1):150 – 157, 2011. International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010.

[28] M.Morpurgo. A large superconducting dipole cooled by forced circulation of two phase helium. *cryogenics*, pages 411–414, Jul 1979.

[29] J. Nelder and R. Mead. A Simplex Method for Function Minimization. *Computer Journal*, 7:308–313, 1965.

[30] C. Nellist. *Characterisation and Beam Test Data Analysis of 3D Silicon Pixel Detectors for the ATLAS Upgrade*. PhD thesis, The University of Manchester, 2013.

[31] S. I. Parker, C. J. Kenney, and J. Segal. 3-D: A New architecture for solid state radiation detectors. *Nucl. Instrum. Meth.*, A395:328–343, 1997.

[32] I. Perić, L. Blanquart, G. Comes, P. Denes, K. Einsweiler, P. Fischer, E. Mandelli, and G. Meddeler. The FEI3 readout chip for the ATLAS pixel detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 565(1):178 – 187, 2006. Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging PIXEL 2005 International Workshop on Semiconductor Pixel Detectors for Particles and Imaging.

[33] K. Prokofiev and K. E. Selbach. Neural network based cluster reconstruction in the atlas pixel detector. *Journal of Physics: Conference Series*, 396(2):022040, 2012.

[34] S. Ramo. Currents induced by electron motion. *Proceedings of the IRE*, 27(9):584–585, Sept 1939.

[35] I. Rubinsky. An EUDET/AIDA pixel beam telescope for detector development. *Physics Procedia*, 37, 2012.

[36] O. Røhne. Edge characterization of 3D silicon sensors after bump-bonding with the ATLAS pixel readout chip. *Nuclear Science Symposium Conference Record, NSS'08, IEEE*, pages 1929–1934, Oct 2008.

[37] A. Sen and S. Srivastava. *Regression Analysis: Theory, Methods, and Applications*. Lecture Notes in Statistics. Springer, 1990.

[38] W. Shockley. Currents to conductors induced by a moving point charge. *J. Appl. Phys.*, 9:635, 1938.

[39] K. Sjøbæk. Full simulation of a testbeam experiment including modeling of the Bonn ATLAS telescope and ATLAS 3D pixel silicon sensors. Master's thesis, University of Oslo, 2010.

[40] B. Streetman and S. Banerjee. *Solid State Electronic Devices*. Prentice Hall Series in Solid State Physical Electronics. Pearson Prentice Hall, 2006.

[41] The ATLAS IBL collaboration. Prototype ATLAS IBL modules using the FE-I4A front-end readout chip. *Journal of Instrumentation*, 7(11):P11010, 2012.

[42] J. Treis. *Development and operation of a novel PC-based high speed beam telescope for particle tracking using double sided silicon microstrip detectors*. PhD thesis, University of Bonn, 2002.

[43] P. Vavilov. Ionization losses of high-energy heavy particles. *Sov.Phys.JETP*, 5:749–751, 1957.

[44] J. Weingarten, S. Altenheiner, M. Beimforde, M. Benoit, M. Bomben, G. Calderini, C. Gallrapp, M. George, S. Gibson, S. Grinstein, Z. Janoska, J. Jentzsch, O. Jinnouchi, T. Kishida, A. L. Rosa, V. Libov, A. Macchiolo, G. Marchiori, D. Muenstermann, R. Nagai, G. Piacquadio, B. Ristic, I. Rubinskiy, A. Rummler, Y. Takubo, G. Troska, S. Tsiskaridtze, I. Tsurin, Y. Unno, P. Weigell, and T. Wittig. Planar pixel sensors for the ATLAS upgrade: beam tests results. *Journal of Instrumentation*, 7(10):P10028, 2012.

[45] N. Wermes and G. Hallewel. *ATLAS pixel detector: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998.

[46] M. A. Woodbury. *Inverting Modified Matrices*. Number 42 in Statistical Research Group Memorandum Reports. Princeton University, Princeton, NJ, 1950.

Appendix A

Straight line track reconstruction for the ATLAS IBL testbeam with the EUDET telescope.

Introduction

A readout event in the EUDET telescope can contain measurements from several tracks, as well as noise hits. The detector planes does not have 100% detection efficiency. A track fitter, like the Kalman filter, requires a set of measurements assumed to originate from the same particle. If this list contains hits that are not created by the particle, or misses measurements that are, tracking performance will suffer.

Several algorithms have been implemented for track reconstruction of EUDET data: the Combinatorial Kalman filter, a heuristic search through a directed graph; the Deterministic Annealing Filter, an iterative, adaptive fitting algorithm that uses all available information to determine if a hit should be rejected or accepted; and the cluster track finder, a new method that is very simple to implement based on a cluster finder. The implementation details of these methods, as well as the information filter used for track fitting are described in the following.

The implemented methods were validated in a simulation experiment, which found the combinatorial Kalman filter followed by the Deterministic Annealing Filter to give the best performance. At low noise densities, the cluster track finder performs acceptably. Results from real test beam data are presented, showing excellent track quality.

The implementations of the methods and the results from the simulated and real data were presented in an ATLAS note, an internally reviewed report-series from the ATLAS experiment.

ATLAS NOTE

June 11, 2014

Straight line track reconstruction for the ATLAS IBL testbeam with the EUDET telescope.

H. Gjersdal[a], O. Røhne[a], A. Strandlie[a,b]

[a]*Department of Physics, University of Oslo, Oslo, Norway*
[b]*Gjøvik University College, P. O. Box 191, N-2802 Gjøvik, Norway*

Abstract

Track reconstruction based on the Kalman filter has been implemented for straight line tracks in data taken with the EUDET beam telescope. The information filter formulation of the Kalman filter has been used to implement the combinatorial Kalman filter and the Deterministic Annealing Filter to deal with noisy input data. In addition to the Kalman filter based track finding, a new and simple to develop and implement track finding method based on cluster finding has been studied. The methods have been tested and compared using a simple simulation. The simulation studies show that both the combinatorial Kalman filter and the cluster track finder have high track finding efficiency, and that the Deterministic Annealing Filter is able to reduce the noise contamination and improve the precision of the track estimates. Real test beam data has been reconstructed with excellent goodness of fit.

1 Introduction

The EUDET telescope is a high precision beam telescope used in test beam characterization of particle detectors. The goal of the beam telescope is to provide information about the particles passing through it for analysis of devices under test.

Track reconstruction is the process of obtaining the optimal parametrized description of particle tracks at the point of intersection with the devices under test. This consists of track finding and track fitting. Track finding determines which measurements have been created by the same particle from data that contain noise hits and measurements created by other tracks. Track fitting is finding the optimal parameters for describing the track, as well as the correct description of the parameter uncertainties.

The Kalman filter [1] has been implemented for track fitting of EUDET data. The implementation has also been used as a core for methods for track finding. The simple geometry of the experiment, as well as the lack of a magnetic field, makes it possible to use the information filter formulation of the Kalman filter. This offers some advantages in numerical stability, as well as the ability to describe a complete lack of information for the initial guess of the track parameters.

Three different approaches to track reconstruction of test beam data have been implemented and compared with simulation experiments.

- The combinatorial Kalman filter [2], which is a track finder based on the Kalman filter.
- The combinatorial Kalman filter followed by the Deterministic Annealing Filter (DAF). The DAF is an iterative method that considers all the hits in the detector, using optimal track estimates in deciding on including or rejecting a hit [3]. The information filter can deal with several weighted hits per plane with only minor adjustments, making it ideal for implementing the DAF.
- The cluster track finder followed by the DAF. The cluster track finder is a new track finding method based on cluster finding. It takes advantage of the beam constraints in the test beam experiment, and the ability of the DAF to deal with imperfections in the track finder.

In the following, the experimental setup and the track model are described. The methods that were implemented for track fitting and track finding for the ATLAS IBL test beam with the EUDET telescope in 2010 and 2011 are then described. A validation of the implementations using a simple simulation experiment is presented, as well as results from real test beam data.

1.1 The experimental setup

The EUDET telescope was developed by the EUDET[1] and AIDA[2] projects [4]. The telescope consists of six MIMOSA26 sensors, each having 576×1152 pixels with a pitch of 18.4 μm in both directions, covering an area of approximately 10.6×21.2 mm². The sensor planes are distributed in two arms, where devices under test are placed in the middle, as illustrated in Figure 1. The reconstruction of the EUDET telescope data is done within the EUTelescope software framework[3].

The telescope has been used to reconstruct a monochromatic beam of π^+ generated by the Super Proton Synchrotron at CERN. The particle beam is collimated, and near orthogonal to the EUDET sensor planes.

---

[1] Detector R&D towards the International Linear Collider

[2] Advanced European Infrastructures for Detectors and Accelerators

[3] see `http://eutelescope.web.cern.ch/`

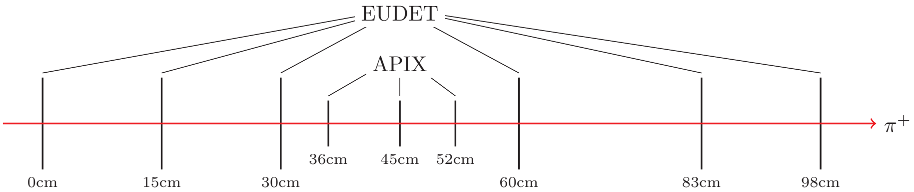


Figure 1: Schematic overview of the EUDET telescope, with ATLAS pixel devices (APIX). The position along the z-axis are noted under the detector planes. The exact z-positions of the planes vary between experiments, but these values are from the real and simulated data presented in the following.

1.2 The track model

The material in the test beam instrumentation is assumed to be concentrated in planes with no spatial thickness, and no interactions between the particles and air are assumed to occur. In these planes, the particles will interact with the material through multiple Coulomb scattering, changing the angle of the trajectory.

Without a magnetic field the energy or momentum of a particle can not be determined from the trajectory. The only effect of the beam energy on the trajectory is to determine the amount of scattering that can occur in a plane. As the energy loss can not be determined, all particles are assumed to have the same energy and momentum throughout the experiment.

The beam telescope is described in a three-dimensional, right-handed Cartesian coordinate system, where the y-axis is in the vertical direction, and the z-axis is parallel to the nominal beam direction. At a plane k, the particle track is described with the parameters

$$\mathbf{x}_k = \begin{pmatrix} x \\ y \\ dx/dz \\ dy/dz \end{pmatrix}. \tag{1}$$

The position where the particle intersects with the plane is described by x and y, and dx/dz and dy/dz are direction tangents of the angles in the xz and yz planes. The uncertainties of an estimate of $\mathbf{x}_k$ are described in the covariance matrix $\mathbf{C}_k$.

If there is no material between the planes, the parameters on the front side of plane k can be calculated from the parameters on the back side of plane $k-1$. This process is deterministic.

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1}^B \tag{2}$$

$$\tag{3}$$

The superscript ^{B} denotes parameters on the back side of a plane, after the particle has interacted with the material in the plane. The transport matrix, $\mathbf{F}_k$, depends on the propagation length along the z-axis, Δz_k.

$$\mathbf{F}_k = \begin{pmatrix} 1 & 0 & \Delta z_k & 0 \\ 0 & 1 & 0 & \Delta z_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4}$$

The change in angle of the particle due to multiple Coulomb scattering has an expectation value of zero. The uncertainties of scattering in orthogonal planes, parallel to the z-axis, are Gaussian and can be calculated from the Highland formula [5].

$$\sigma_{\text{plane}} = \frac{0.0136}{E}\sqrt{\frac{X}{X_0}}(1 + 0.038ln(\frac{X}{X_0})) \tag{5}$$

To estimate the uncertainties, the beam energy, E, as well as the amount of material, $\frac{X}{X_0}$, must be known. The beam energy is in units of GeV, the thickness in units of radiation lengths.

The changes in the track parameters due to material interactions in a plane are

$$\mathbf{x}_k^B = \mathbf{x}_k + \boldsymbol{\omega}, \tag{6}$$

where $\mathbf{x}_k$ are the track parameters on the front side of the plane, $\mathbf{x}_k^B$ are the track parameters on the back side of the plane, and $\boldsymbol{\omega}$ is a Gaussian stochastic term. The stochastic term has an expectation value

$$E\{\boldsymbol{\omega}\} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{7}$$

and the covariance matrix

$$cov\{\boldsymbol{\omega}\} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{xz}^2 & 0 \\ 0 & 0 & 0 & \sigma_{yz}^2 \end{pmatrix}. \tag{8}$$

The standard deviations in the two planes, σ_{xz} and σ_{yz}, are equal and calculated from Eq. 5.

The detectors in the test beam give two-dimensional measurements of the position of a particle in the detector planes. The measurements are modeled as a sum of the true position of the particle in the plane and a stochastic term due to measurement uncertainties.

$$\mathbf{m}_k = \mathbf{H}\mathbf{x}_k + \boldsymbol{\epsilon} \tag{9}$$

The measurement errors have an expectation value

$$E\{\boldsymbol{\epsilon}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{10}$$

and a covariance matrix $cov\{\boldsymbol{\epsilon}\} = \mathbf{V}$.

The matrix $\mathbf{H}$ describes the projection from the four-dimensional track parameter space to the two-dimensional measurement space. The projection is the same in all planes.

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{11}$$

The measurements are defined to be on the front side of the measurement planes.

2 Track fitting with the information filter

The Kalman filter is a recursive formulation of the least squares estimator. It starts at the first or last measurement plane in the detector system, and works by alternating between updating and predicting the track estimate. Updating is combining a prediction and a measurement in a plane, predicting is propagating the updated estimate to the next measurement plane.

The Kalman filter can run either in the forward direction, reading in the measurements in the same order they were created, or oppositely in the backward direction. The optimal estimate of the track parameters is obtained only when all measurements have been included in the fit. To obtain the optimal estimates in all the planes, estimates from a forward running and a backward running Kalman filter are combined in a process called smoothing.

The information filter is a reformulation of the Kalman filter that does not keep track of the explicit parameters and covariance matrix, but the weight matrix, $\mathbf{W} = \mathbf{C}^{-1}$, and the information vector, $\mathbf{i} = \mathbf{W}\mathbf{x}$.

In a complex high energy physics experiment, the explicit track parameters are needed for propagating the particle state. In the simple track model in the test beam experiment, however, it possible to perform the propagation with implicit parameters. In this case, the information filter formulation offers some advantages in numerical stability compared to the standard formulation.

The information filter is able to describe a complete lack of knowledge about the track parameters in both position and direction with a weight matrix that is a null matrix. This lack of knowledge can only be approximated with large diagonal elements in an explicit covariance matrix.

2.1 Update

A new measurement is included in the implicit parameter estimate by combining the information from the prediction and the measurement in a plane.

The measurement projected into parameter space can be described as

$$\mathbf{x}_m = \mathbf{H}^T \mathbf{m}_k, \tag{12}$$

with the weight matrix

$$\mathbf{W}_m = \mathbf{H}^T \mathbf{V}_k^{-1} \mathbf{H} = \begin{pmatrix} \mathbf{V}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{13}$$

where $\mathbf{0}$ is the 2×2 null matrix. The weigh matrix reflects that no information about the direction of the particle exists in the measurement.

The updated information vector is the weighted sum of the prediction and the measurement, where the weights are the weight matrices. The updated weight matrix is the sum of the weights.

$$\mathbf{W}_{k|k}\mathbf{x}_{k|k} = \mathbf{W}_{k|k-1}\mathbf{x}_{k|k-1} + \mathbf{W}_m\mathbf{x}_m \tag{14}$$

$$\mathbf{W}_{k|k} = \mathbf{W}_{k|k-1} + \mathbf{W}_m \tag{15}$$

The explicit parameters are the weighted average of the measurement and the prediction.

$$\mathbf{x}_{k|k} = (\mathbf{W}_{k|k-1} + \mathbf{W}_m)^{-1}(\mathbf{W}_{k|k-1}\mathbf{x}_{k|k-1} + \mathbf{W}_m\mathbf{x}_m) \tag{16}$$

The information filter does not need the explicit parameters, and the update simply becomes

$$\mathbf{i}_{k|k} = \mathbf{i}_{k|k-1} + \mathbf{H}^T \mathbf{V}_k^{-1} \mathbf{m}_k \tag{17}$$

$$\mathbf{W}_{k|k} = \mathbf{W}_{k|k-1} + \mathbf{H}^T \mathbf{V}_k^{-1} \mathbf{H}. \tag{18}$$

2.2 Prediction

The propagation of the parameter vector from the back side of plane $k-1$ to the front side of plane k is described in Eq. 2. The propagation of the uncertainties is done with linear error propagation.

$$\mathbf{C}_{k|k-1} = \mathbf{F}_k \mathbf{C}^B_{k-1|k-1} \mathbf{F}^T_k \tag{19}$$

In the simple track model, a track can be propagated without needing the explicit states as long as the z-position of the points where the track intersects with the material planes are known. The deterministic parameter transformations from the back side of plane $k-1$ to the front side of plane k are described by the inverse transport matrix, $\mathbf{F}^{-1}$.

$$\mathbf{W}_{k|k-1} = (\mathbf{F}_k \mathbf{C}^B_{k-1|k-1} \mathbf{F}^T_k)^{-1} = (\mathbf{F}^{-1}_k)^T \mathbf{W}^B_{k-1|k-1} \mathbf{F}^{-1}_k \tag{20}$$

$$\mathbf{i}_{k|k-1} = \mathbf{W}_{k|k-1} \mathbf{F}_k \mathbf{x}^B_{k-1|k-1} = (\mathbf{F}^{-1}_k)^T \mathbf{i}^B_{k-1|k-1} \tag{21}$$

The transport matrix describes the linear transformation of the parameters when they are propagated a distance Δz_k along the z-axis, where Δz_k is the difference between the z position of the points where the particle intersects with planes k and $k-1$. The case where the z-positions are not known exactly a priori will be discussed after the Deterministic Annealing Filter is introduced.

2.3 Multiple Coulomb scattering

Multiple scattering is taken into account by adding the covariance matrix of the material interactions to the covariance matrix of the parameters.

$$\mathbf{C}^B_k = \mathbf{C}_k + cov\{\omega\} \tag{22}$$

Inverting $\mathbf{W}_k$ to obtain $\mathbf{C}_k$ and $\mathbf{C}^B_k$ to obtain $\mathbf{W}^B_k$ would introduce numerical instabilities and require an initial weight matrix that is not a null matrix.

Only the covariance matrix elements that describe the uncertainties of the direction tangents are altered, and this uncertainty can be described with a 2×2 matrix

$$cov\{\omega\} = \mathbf{U}^T \mathbf{Q} \mathbf{U}, \tag{23}$$

where

$$\mathbf{Q} = \begin{pmatrix} \sigma^2_{xz} & 0 \\ 0 & \sigma^2_{xy} \end{pmatrix} \text{ and } \mathbf{U} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{24}$$

Using the the Woodbury matrix identity [6], this problem can be solved by inverting the 2×2 matrix $\mathbf{Q}$ instead of the 4×4 weight matrices.

$$\mathbf{W}^B_k = (\mathbf{C}_k + \mathbf{U}^T \mathbf{Q} \mathbf{U})^{-1} \tag{25}$$

$$= \mathbf{W}_k - \mathbf{A}_k \mathbf{W}_k, \tag{26}$$

where

$$\mathbf{A}_k = \mathbf{W}_k \mathbf{U}^T (\mathbf{Q}^{-1} + \mathbf{U} \mathbf{W}_k \mathbf{U}^T)^{-1} \mathbf{U}. \tag{27}$$

Since $\mathbf{i}_k$ depends on $\mathbf{W}_k$, this must also be updated.

$$\mathbf{i}^B_k = \mathbf{i}_k - \mathbf{A}_k \mathbf{i}_k \tag{28}$$

This method is faster than relying on inversion of the 4×4 matrices.

2.4 Test statistics

The difference between the predicted position and the measurement is described by the residual vector, $\mathbf{r}_{k|k-1}$, and the corresponding covariance matrix $\mathbf{R}_{k|k-1}$.

$$\mathbf{r}_{k|k-1} = \mathbf{H}^T\mathbf{x}_{k|k-1} - \mathbf{m}_k \tag{29}$$

$$\mathbf{R}_{k|k-1} = \mathbf{H}^T\mathbf{C}_{k|k-1}\mathbf{H} + \mathbf{V}_k \tag{30}$$

To test that the model assumptions hold, the fitted χ^2 of the tracks are studied.

$$\chi^2 = \sum_k^n \chi_k^2 \tag{31}$$

$$\chi_k^2 = \mathbf{r}_{k|k-1}^T\mathbf{R}_{k|k-1}^{-1}\mathbf{r}_{k|k-1} \tag{32}$$

If the measurements are unbiased, the stochastic processes that are encountered are Gaussian and correctly described, and the residuals in the different planes are uncorrelated, the fitted χ^2 should follow a χ^2 distribution with $N_{\text{dof}} = N_m - N_p$ degrees of freedom. N_m is the number of measurements, and $N_p = 4$ is the number of fitted track parameters. Each two-dimensional pixel measurement increments N_{dof} by 2. As a straight line can pass through one or two points with zero residuals, the χ^2 increments for the first two measurements that are included are zero.

The Kalman filter is a least squares estimator, finding the parameters that minimize χ^2 for each track. According to the Gauss-Markov theorem [7], the least squares estimator is the linear unbiased estimator with the smallest variance given that:

1. the measurements are unbiased,
2. the normalized residuals have the same variance,
3. and the residuals in the different planes are uncorrelated.

If the fitted χ^2 does not follow the theoretical χ^2 distribution with the same number of degrees of freedom, it is an indication that these assumptions do not hold, or that the residuals are not Gaussian.

If the geometry description of the experiment is not correct, the measurements will be biased. The normalized residuals will only have the same variance if the measurement uncertainties and the amount of scattering in the planes are correctly described. If the measurement errors are in part due to systematic effects, measurement errors can be correlated between planes.

2.5 Smoother

The track estimate can be optimal only when information from all the measurements have been included. For a plane in the middle of the telescope, information from Kalman filters running in opposite directions are combined. Analogously to the Kalman filter update, the information is combined by taking the weighted average of the two predictions, where the weights are the inverse covariance matrices.

$$\mathbf{C}_k^* = \left[\mathbf{W}_{k|k-1}^{FW} + \mathbf{W}_{k|k+1}^{BW}\right]^{-1} \tag{33}$$

$$\mathbf{x}_k^* = \mathbf{C}_k^*\left(\mathbf{i}_{k|k-1}^{FW} + \mathbf{i}_{k|k+1}^{BW}\right) \tag{34}$$

The implicit parameter estimate at the front side of plane k is described by $\mathbf{W}_{k|k-1}^{FW}$ and $\mathbf{i}_{k|k-1}^{FW}$ for a forward running filter, and $\mathbf{W}_{k|k+1}^{BW}$ and $\mathbf{i}_{k|k+1}^{BW}$ for a backward running filter. The explicit, smoothed estimate of the parameters in plane k are $\mathbf{x}_k^*$ and $\mathbf{C}_k^*$. The asterisk indicates that the estimates do not contain information from the measurement in plane k.

The sum of squared normalized residuals with respect to the smoothed predictions in all planes does not follow a χ^2 distribution, as the residuals are correlated.

3 Track finding

For the Kalman filter to work, it needs a list of measurements to be fitted. The process of going from all hits in the detector to a set of hits assumed to belong to the same track is called track finding. Because of detector noise, less than perfect detection efficiency, and the chance that more than one particle passes through the detector at each trigger, an event in the EUDET telescope rarely contains only a single hit per detector plane.

A good track finder must have a high probability for including a measurement that truly belong to the track, and a low probability of including a hit that does not. It must have a high probability of finding a real particle track, and a low probability of reporting a combination of hits not created by the same particle as a track.

Two pure track finders have been implemented. The combinatorial Kalman filter is an efficient search for hit combinations in a directed graph structure that is based on Kalman filter updates and predictions. A cruder and simpler to implement method, called the cluster track finder, has been developed, taking advantage of the small angular spread in the beam. This is based on clustering hits that are close in the global xy plane.

In addition to this, the Deterministic Annealing Filter (DAF) has been implemented. The DAF blurs the lines between track fitting and track finding. It is an iterative procedure that uses the information filter smoother to assign weights to all the hits in the planes. An initial weight must be assigned from a track finder.

3.1 The combinatorial Kalman filter

Assuming a particle will create either one or zero measurements per detector plane, the information from an event in the test beam setup can be seen as a directed acyclic graph, where every hit is a vertex. A vertex can be connected to another through a directed edge. In this case, every vertex emits a directed edge to every vertex corresponding to a hit further downstream in the experiment. Such a graph is drawn in Figure 2.

A path is a sequence of edges that connect a set of vertices. The length of the path is the number of edges it contains, a path containing a single vertex is of length zero. The measurements created by a track would create such a path. This path can be of any length, and is not required to start in the first measurement plane, or to end up in the last.

The adjacency matrix, $\mathbf{A}$, of a graph with n vertices is an $n \times n$ matrix, where the matrix element a_{ij} is the number of edges connecting vertex i to vertex j. The hits in the beam telescope form a graph where the adjacency matrix element a_{ij} is 1 in the case that vertex j is further downstream than vertex i, and 0 otherwise. The number of possible paths of length l can be calculated by summing all the matrix elements in the matrix $\mathbf{A}^l$. The total number of paths is $\prod_k (N_k + 1) - 1$, where N_k is the number of hits in plane k. There exists 728 distinct paths in the graph in Figure 2. Studying every possible combination of hits is impractical even at this hit density, and more efficient searches are needed.

The high detection efficiency of the EUDET telescope means that the probability of a real track in the detector creating a very short path is small. The probability of accepting a path not corresponding to a real particle track decreases with the length of the path. Only paths above some minimum length are considered acceptable track candidates, but this requirement alone does not sufficiently reduce the complexity. The graph in Figure 2 has 496 distinct paths containing four or more vertices.

The combinatorial Kalman filter is a heuristic search through the graph, where only edges likely to lead to an acceptable track are followed. The search starts at some vertex in the graph, and creates paths by following the edges to new vertices in depth-first. The recursive nature of the Kalman filter makes it possible to assess the likelihood of finding a good track by following an edge in the graph. This is done by calculating the χ^2 increment of including the connected vertex in a Kalman filter fit of

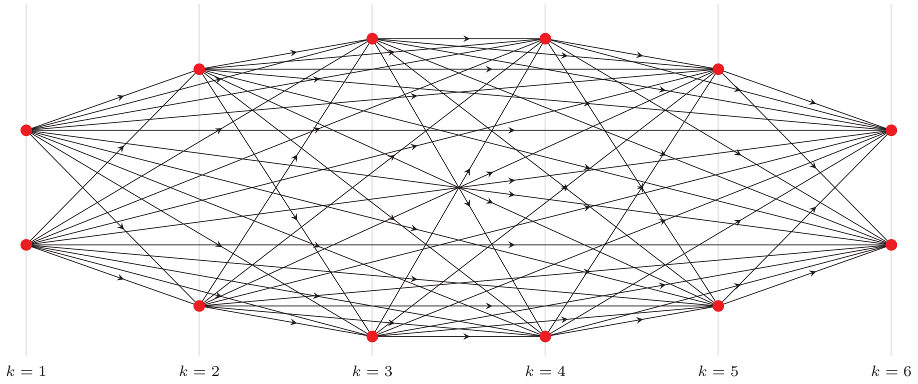


Figure 2: The directed acyclic graph for an event with two hits per plane and six measurement planes. A hit is a vertex, and every vertex is connected with an edge to all vertices representing hits further downstream. Vertices are drawn as red dots, the edges are black arrows. The position of the vertices in the horizontal direction corresponds to the measurement plane of the hit, the position of the vertex in the vertical direction has no significance. The measurements created by a particle would form a path in this graph, containing anywhere from zero to six vertices.

the path, Eq. 32. If the χ^2 increment is above a cut-off value, χ^2_{CKF}, the edge is not followed, and the entire subgraph starting with the connected vertex is excluded. With a tight cut in the χ^2 increment, the complexity of the problem is greatly reduced.

The χ^2 increment can only be calculated when the path contains two or more vertices. To speed up the search further, the collimated nature of the particle beam is taken into account. Cuts in the angles of the straight line between the two first vertices makes it possible to exclude edges emitting from the first vertex in the search.

A path of length larger than zero can be split into subset paths, paths made up of a subset of the vertices in the path. To avoid reporting subset paths of an accepted track as solutions, the search starts by considering the edges emitting from a vertex v that skip the fewest measurement planes. If an edge passes the cuts, the search follows this edge until a track is accepted or every path in the subgraph is excluded. After returning to the vertex v, further edges that skip planes are only considered if following edges that skip fewer planes has not lead to an accepted track.

The combinatorial Kalman filter has been implemented as a recursive function, that attempts to extend a path to a vertex in a measurement plane k. The function arguments are a path containing one or more vertices corresponding to hits upstream of the plane, as well as the Kalman filter prediction based on the vertices in the path. The last vertex in the path is $v_{i<k}$. The function performs the following tasks:

1. If every edge emitting from $v_{i<k}$ has been excluded, and the path is of sufficient length, the function will be called in plane $k = N_p + 1$, where N_p is the number of planes in the experiment. The track quality of the fitted path is then checked with cuts in χ^2/N_{dof}. If the candidate passes the cuts, the track is accepted. The function then returns.

2. A Kalman filter prediction is made from the estimate in the previous plane, containing all measurements in the path leading to plane k.

3. Every edge connecting $v_{i<k}$ to a vertex corresponding to a hit in plane k is considered. All the edges that lead to a path that passes cuts in the angle or χ^2 increment are followed. Following an

edge means the function calls itself in plane $k+1$ after updating the Kalman filter estimate, and extending the path with the new vertex.

4. If no vertex in the plane lead to the path being extended, the plane is skipped. This is only done if it is possible to obtain a path of sufficient length in the remaining planes. The function calls itself in plane $k+1$ without updating the estimate with a hit or extending the path.
5. The function returns.

The search is started in every vertex corresponding to a hit in the first plane. In following planes, searches are started in every vertex corresponding to a hit that has not been included in an accepted track originating from a search started further upstream. This is done for all planes where it is possible to achieve a path of sufficient length.

Both the speed and accuracy of the method depend on whether the χ^2 increments follow a real χ^2 distribution. A poor description of the detector geometry, the amount of scattering encountered, or measurement uncertainties, makes optimal edge rejection harder.

3.2 Deterministic Annealing Filter

The DAF is an extension of the Kalman filter, that deals with imperfections in track finding. While the combinatorial Kalman filter uses the χ^2 increments of a Kalman filter to accept or reject hits, the DAF uses smoothed predictions containing information from all the hits in the planes.

The DAF is an iterative procedure that assigns weights to all the hits. It alternates between using the information filter smoother for fitting the set of weighted hits, and improving the weights based on the previous fit. The fitter needs an initial set of weights for the hits, which is achieved by setting the weight of all the hits that the track finder has identified as belonging to the track to one, and all the other weights to zero. If more than one hit per plane is assigned a weight in the track finder, the weights are normalized so that the total weight in the plane is one.

3.2.1 Weight assignment

The ith hit in plane k, $\mathbf{m}_{k,i}$, is assigned the weight

$$w_{k,i} = \frac{\exp(-\chi^2_{k,i}/2T)}{\exp(-\chi^2_{DAF}/2T) + \sum_j \exp(-\chi^2_{k,j}/2T)}, \tag{35}$$

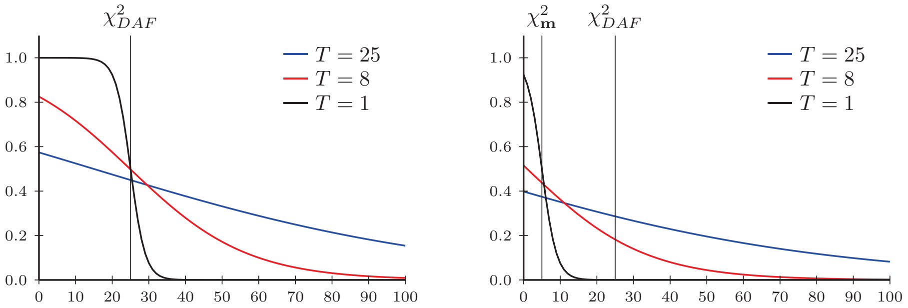
where $\chi^2_{k,i}$ is calculated from the smoothed predictions at the plane.

$$\chi^2_{k,i} = (\mathbf{m}_{k,i} - \mathbf{H}\mathbf{x}^*_k)^T (\mathbf{H}\mathbf{C}^*_{k,i}\mathbf{H}^T + \mathbf{V}_{k,i})^{-1} (\mathbf{m}_{k,i} - \mathbf{H}\mathbf{x}^*_k) \tag{36}$$

If the smoothed prediction is Gaussian and correctly estimated, the measurement uncertainties are known and Gaussian, and the measurement truly belongs to the track, $\chi^2_{k,i}$ should follow the χ^2 distribution with two degrees of freedom.

The weight depends on a cut-off value, χ^2_{DAF}, which is the χ^2 value that will obtain a weight of 0.5, and a "temperature", T, that determines how fast the weight will drop off around the cut-off. The weight function at different T is shown in Figure 3a. The term $\sum \exp(-\chi^2_{k,j}/2T)$ in the denominator introduces competition between hits, as illustrated in Figure 3b.

The DAF iterates through a predetermined set of temperatures, generally starting at a high temperature and ending up at a low temperature, with a harder cut-off, after several iterations.



(a) Weight function for a single hit.

(b) Weight function with a competing hit.

Figure 3: Weight function as a function of measurement χ^2 at different temperatures with and without a competing hit. $\chi^2_{\mathbf{m}}$ is the χ^2 of the competing hit, χ^2_{DAF} is the cut-off value.

3.2.2 DAF update and prediction

The information filter update becomes a sum over all the weighted hits in the plane

$$\mathbf{W}_{k|k} = \mathbf{W}_{k|k-1} + \sum_i w_{k,i}\mathbf{H}^T\mathbf{V}_{k,i}^{-1}\mathbf{H} \tag{37}$$

$$\mathbf{i}_{k|k} = \mathbf{i}_{k|k-1} + \sum_i w_{k,i}\mathbf{H}^T\mathbf{V}_{k,i}^{-1}\mathbf{m}_{k,i}. \tag{38}$$

In some cases it is possible to improve the transport matrices, $\mathbf{F}_k$, used in the information filter predictions. If one of the detector planes in the experiment is tilted such that it is not orthogonal to the z-axis, the propagation distance, Δz, depends on the position and direction of the track. Since the explicit state of the track is made available to estimate the hit weights, it is possible to also improve the Δz estimated for all the propagation matrices between the iterations in the DAF.

A good first estimate of the z-position of a plane is the z-position of the hit with the largest weight from the track finder.

3.2.3 DAF test statistics

Test statistics can be calculated for the DAF similar to the standard Kalman filter.

$$\chi^2 = \sum_{k=1}^{n}\sum_{i=1}^{m_k} w_{i,k}\mathbf{r}_{i,k|k-1}^T\mathbf{R}_{i,k|k-1}^{-1}\mathbf{r}_{i,k|k-1} \tag{39}$$

$$N_{\text{dof}} = -4 + 2\sum_{k=1}^{n}\sum_{i=1}^{m_k} w_{i,k}, \tag{40}$$

where n is the number of measurement planes, and m_k is the number of hits in plane k.

Since the weights, and the number of degrees of freedom, in this case are not integers, this is not a true χ^2 distribution. But if the temperature and cut-off for the weight function are tuned such that hits tend to get a weight very close to one, or very close to zero, the fitted χ^2 will be close to the theoretical χ^2 distribution with the number of degrees of freedom rounded to the nearest integer.

3.3 Cluster track finder

The DAF can reject outliers, pick up new hits that are close to the track, and deal with more than one hit per plane. This means that the requirements for accuracy in the track finder can be reduced.

A crude track finder that is very simple to implement has been developed for the test beam. The track finder works by projecting all hits to the first measurement plane using the nominal beam angle. Then a cluster finder groups all hits that are closer to any other hit in the cluster than a given radius. All hits in the cluster are then given a weight of 1, all hits outside the cluster get an initial weight of 0. If more than one hit in the same plane gets a weight of 1, the weights in the plane are normalized such that the sum is one. After assigning weights to the hits, the track is fitted with the DAF.

4 Simulation experiments

The implemented track reconstruction methods were tested using a simple simulation model, with a geometry and detector response similar to that of the test beam experiments with the EUDET telescope in the SPS π^+ beam, with three ATLAS pixel planes as devices under test.

The objects of the simulation experiments are to validate the information filter and the combinatorial Kalman filter, to see if the DAF is able to improve the tracks from the combinatorial Kalman filter, and to see if the cluster track finder followed by the DAF can compete with the combinatorial Kalman filter followed by the DAF.

This simulation is not intended as a complete and realistic model for the beam telescope or the ATLAS pixel devices.

4.1 Simulation model

The simulation model consists of nine planes orthogonal to the z-axis, and placed at the z positions in Figure 1. Each plane has a sensitive area of 5 mm × 5 mm.

Each simulated particle is initialized with a position that is uniformly distributed within a 3 mm × 3 mm square in the center of the sensors. The initial direction tangents are independently drawn from Gaussian distributions with a mean of zero and a standard deviation of 1×10^{-4}, similar to conditions observed in the SPS test beam experiments.

The beam energy is 100 GeV throughout the experiment, and all the planes have a thickness of $\frac{X}{X_0} = 0.01$. The amount of scattering a particle undergoes in a plane is drawn from a Gaussian distribution with a mean of zero and a standard deviation calculated from the Highland formula, Eq. 5.

The measurements errors are Gaussian with uncorrelated errors in x and y. The EUDET measurements have a standard deviation of 4.3 µm in both directions. The ATLAS pixel planes have a standard deviation of $400\ \mu\text{m}/\sqrt{12}$ in the x-direction, and $50\ \mu\text{m}/\sqrt{12}$ in the y-direction.

Noise hits are simulated by randomly placing hits uniformly within the 5 mm × 5 mm sensitive area of the sensors. This is a larger area than the area where the initial track states are generated, to prevent tracks from scattering out of the area where noise hits are simulated, and complicating the analysis. The performance of the methods were tested at different noise densities.

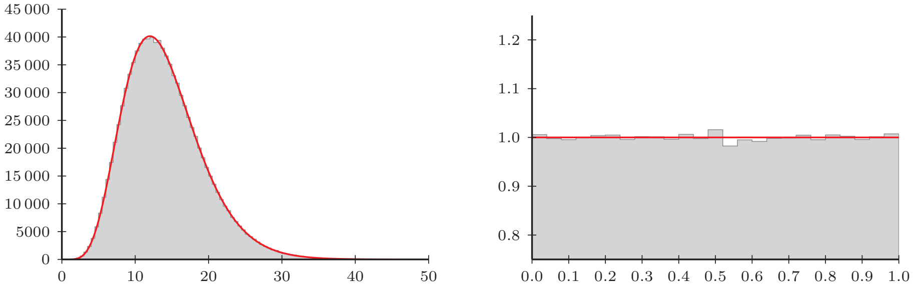
The simulated noise is distributed in a smaller area than that of the real EUDET sensors. The noise density rises much quicker with the number of noise hits per plane in simulated data than what it would do in real test beam data. This was done to reduce the time needed to study very high noise densities.

4.2 Validating the information filter implementation

To validate the information filter implementation, events with single tracks, 100% detection efficiency, and no noise hits were simulated. In this case, track finding is trivial. The quality of the track fits are then tested by studying the χ^2 distribution.

The geometry, amount of material, and measurement uncertainties are the same in the simulation and the reconstruction. In this case, a perfect χ^2 distribution is an indication that the optimal parameters are found, and that the stochastic processes are treated correctly in the information filter.

The probability of observing a χ^2 value that is equal to, or greater than the fitted χ^2 value, given that the observations truly are χ^2 distributed, is called the p-value. To better visualize how well the fitted χ^2 match the real χ^2 distribution, the p-values are also plotted. If the fitted χ^2 really does follow the χ^2 distribution, the p-values should follow a uniform distribution. The distributions are shown in Figure 4.



(a) The distribution of the track χ^2 (▭), and the theoretical $\chi^2(N_{\mathrm{dof}} = 14)$ (−).

(b) The p-values of the track χ^2 (▭), and the uniform distribution with the same width and area (−).

Figure 4: Track quality plots from 1 million simulated tracks, with a perfect geometry description. The empirical distributions are in gray (▭), the theoretical distributions are indicated in red (−). Both plots are a representation of the same data. The p-value plots are more sensitive to deviance from the model, and will be used to present the χ^2 distributions in the following.

4.3 The combinatorial Kalman filter

To challenge the combinatorial Kalman filter, noise hits as well as detector inefficiencies were introduced to the simulation. Each measurement plane was simulated to have a 95% detection efficiency, and the track finder was tested on data with an increasing number of noise hits per plane. Only the simulated hits in the EUDET planes were used in the track reconstruction. The three ATLAS pixel planes were treated as dead material planes.

The cumulative binomial probability of a particle generating four or more hits in the six detector planes with 95% detection efficiency per plane is approximately 99.78%. As the probability of finding ghost tracks is greatly increased if allowing tracks with only three hits, the track finder is at most allowed to skip two measurement planes per track.

The cuts in the direction tangents between the two first hits were set to $|dx/dz| < 0.0005 > |dy/dz|$, which are relatively conservative.

Only tracks with $\chi^2/N_{\mathrm{dof}} < 6.0$ were considered. If more than one track was found in an event, the track with the lowest χ^2/N_{dof} was used in the analysis.

To study the quality of the track finder, the following quantities calculated using simulation truth were studied:

- **Ghost tracks:** The percentage of events where a track was accepted where more than half the included hits are noise hits.
- **Tracking efficiency:** The percentage of events where a track that is not a ghost track is accepted.

- **Missed hits:** The percentage of real measurements that are not included in the fitted tracks. Measurements from events where a ghost track, or no track at all, was accepted are not included in the calculation.

- **Contamination:** The percentage of noise hits among all the hits included in accepted tracks, excluding ghost tracks.

- **Precision:** To measure the precision of the track sample, the generalized variance [8] of the estimated track parameters at the first ATLAS pixel plane has been used. The generalized variance is the determinant of the empirical covariance matrix, calculated from the difference between the estimated and the true parameters at the plane.

These figures of merit are shown in Figure 5 as functions of increasing noise density, at varying cut-off values in the χ^2 increment for the combinatorial Kalman filter. The inverse generalized variance of the combinatorial Kalman filter with $\chi^2_{CKF} = 36$ and 0 noise hits was used as a normalization factor for the precision.

At very low χ^2_{CKF}, the probability of missing hits is high. This reduces the tracking efficiency. With a lower average number of hits in the tracks, the capability of separating real measurements from fake hits is reduced.

As χ^2_{CKF} increases, the performance reaches a plateau. A hit with a very large χ^2 increment will cause the final track to fail the χ^2/N_{dof} cut, leading to the hit being rejected later on in the process. This is a much slower way of rejecting hits than having an appropriate χ^2_{CKF} cut.

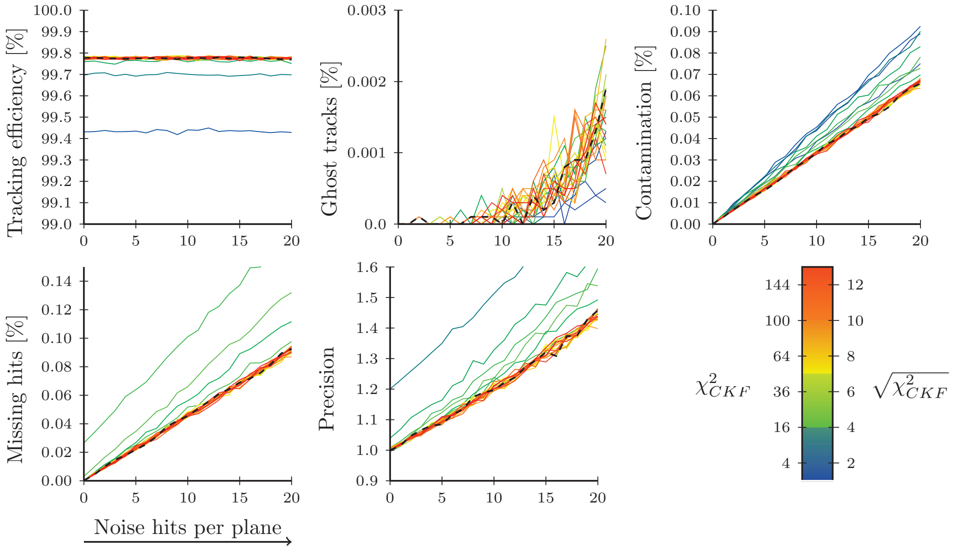


Figure 5: Figures of merit for the Combinatorial Kalman filter as a function of an increasing number of simulated noise hits per plane. The dashed line that lies in the region where the performance reaches a plateau, represents a cut in the χ^2 increment, χ^2_{CKF}, of 36. This will be used for comparison with the other reconstruction methods. The figures of merit are calculated from samples of 1 million tracks.

4.4 Combinatorial Kalman filter followed by the DAF

To compare the performance of the combinatorial Kalman filter followed by a DAF fit to the combinatorial Kalman filter alone, the same quantities as in the previous section were used, but adapted to deal with the weighted hits. For the DAF, the contamination is the percentage of the total weight that is assigned to noise hits. The total weight is the sum of all the weights assigned to hits. A missing hit is a real measurement assigned a weight less than 0.5. The requirement of having three or more hits, is changed to requiring $N_{\text{dof}} > 1.5$. A ghost track is a track where more than half of the total weight is assigned to noise hits.

The results are presented in Figure 6, where the performance is shown for different χ^2_{DAF} for the weighting function. The more accurate χ^2 calculated from smoothed predictions lead to decreased contamination, reduced percentage of missing hits, and improved precision for the DAF fitted tracks.

For reasonable χ^2_{DAF}, the tracking efficiency and the number of ghost tracks are approximately the same as for the combinatorial Kalman filter without the DAF. If χ^2_{DAF} is too high, the probability of including noise hits that bring the χ^2/N_{dof} of the track fit above the cut value is increased, and the tracking efficiency is reduced.

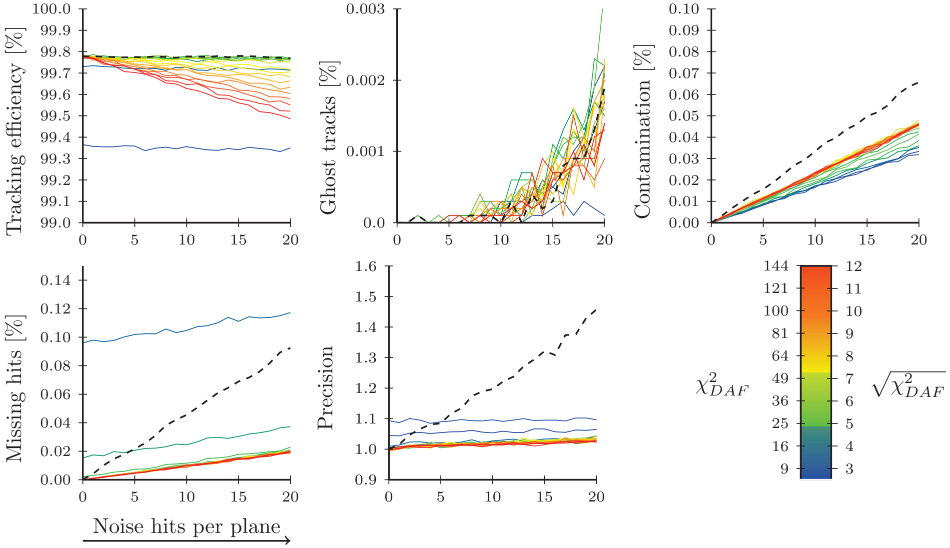
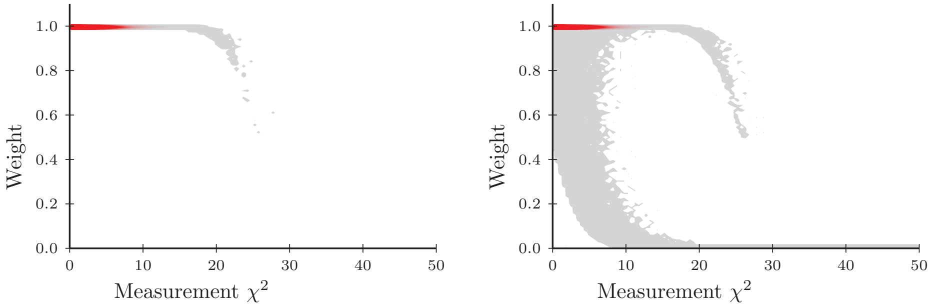


Figure 6: Figures of merit for the combinatorial Kalman filter followed by a DAF fit as a function of an increasing number of simulated noise hits per plane for varying χ^2 cut-off values, χ^2_{DAF}. The dashed line indicates the performance of the combinatorial Kalman filter without a DAF fit with $\chi^2_{CKF} = 36$. This is the same as the dashed line in Figure 5. The figures of merit are calculated from samples of 1 million tracks.

The DAF used 6 iterations, with decreasing temperatures of 25, 20, 14, 8, 4 and 1. These values were selected by studying simulated data with 20 noise hits per plane, and reconstructed with $\chi^2_{CKF} = \chi^2_{DAF} = 36$. The values are a compromise between high tracking efficiency and a low probability of missing hits. The precision was not improved by including lower temperatures, or iterating several times over the lowest temperature.

The weights assigned to each simulated hit is plotted in Figure 7. With no noise hits, the weights all follow the weight function from Eq. 35. With added noise, the hits start competing for weight, and the picture becomes more complex. The minimum weight that can be assigned to a hit is limited by the weight function with a competing hit with $\chi^2 = 0$. Note that the weights are calculated before the final track fit, and the χ^2 is calculated after the final fit.



(a) No noise hits per plane

(b) 20 noise hits per plane

Figure 7: The χ^2 versus the assigned weight after the final DAF fit. The weights are calculated with $\chi^2_{DAF} = 25$ and $T = 1$. The vast majority of hits in the visible χ^2 range are assigned a weight of one, and lie in the red regions of the plots.

4.5 The cluster track finder followed by the DAF

The performance of the cluster track finder followed by a DAF fit is shown in Figure 8. Every cluster with 3 or more hits was fitted with the DAF. The cluster radius needs to be large enough to connect the hits generated by the particle, but if the clusters become too large the DAF is not able to recover. When the clusters become too large, the performance of the method breaks down as the noise density increases.

With a larger spread in initial angular direction, or with more multiple scattering, the method would break down at lower noise densities.

The tracking efficiency is lower than that of the combinatorial Kalman filter, but above 99% even at very high noise densities. The other quantities behave as the combinatorial Kalman filter followed by the DAF. As all the tracks are fitted with the same χ^2_{DAF}, the contamination and amount of missing hits do not vary much with the cluster radius.

At high noise densities, a large number of clusters can be found. As each cluster must be fitted with the DAF, the method is somewhat slower than the combinatorial Kalman filter. Even at the unprecedented noise densities tested here, the speed of the track finding would not be an issue at the data rate produces by the EUDET telescope.

Finding the radius that maximizes the number of accepted tracks appears to be a good way of calibrating the method.

4.6 Comparison of the methods

To test whether the track states are estimated correctly in noisy conditions, the smoothed predictions in the first plane were compared to the true states of the simulated particles, $\mathbf{x}^{true}$. The estimated covariance matrix of the smoothed prediction should describe the uncertainties of this difference. The χ^2 of the difference is

$$\chi^2 = (\mathbf{x}^{true} - \mathbf{x})^T \mathbf{C}^{-1} (\mathbf{x}^{true} - \mathbf{x}), \tag{41}$$

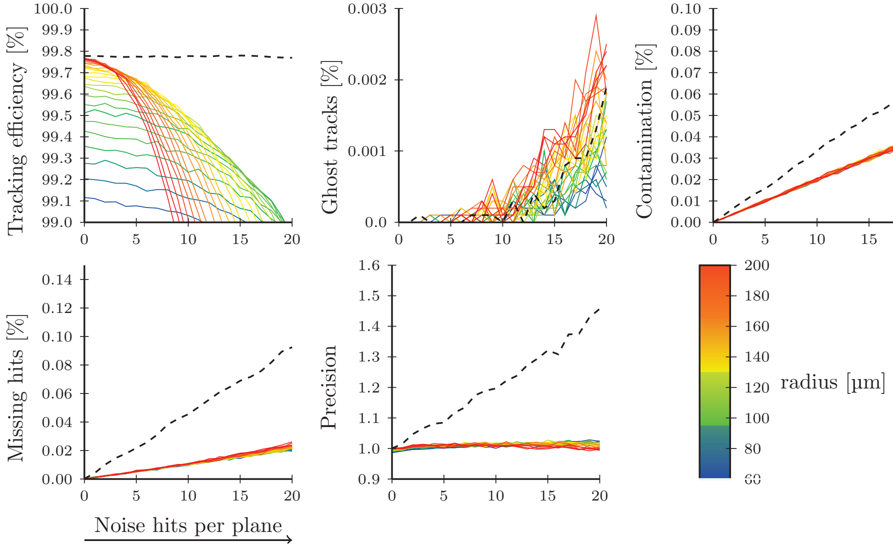


Figure 8: Figures of merit for the cluster track finder followed by the DAF with $\chi 2_{DAF} = 36$ as a function of an increasing number of simulated noise hits per plane. The radius is the maximum distance between hits in a cluster. The dashed line represents the combinatorial Kalman filter with χ^2_{CKF} cut of 36.This is the same as the dashed line in Figure 5. The figures of merit are calculated from samples of 1 million tracks.

with four degrees of freedom. Figure 9 shows the p-values of the χ^2 distribution for the combinatorial Kalman filter, the combinatorial Kalman filter followed by the DAF, and the cluster track finder. The combinatorial Kalman filter ran with a χ^2_{CKF} of 36, the DAF with χ^2_{DAF} of 36, and the cluster finder with a radius of 100 µm. The tracks fitted with the DAF have p-values that follow a uniform distribution. The p-values for the tracks fitted without the DAF has a small excess close to 0, meaning very high fitted χ^2.

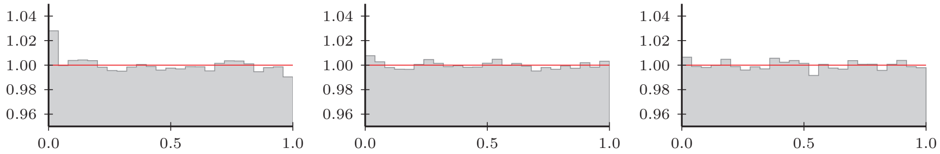
The combinatorial Kalman filter has a tracking efficiency very close to the theoretical 99.78% for tracks with four or more hits. The DAF offers improved noise rejection, and is able to produce tracks with less contamination and less probability of missing real hits, compared to the combinatorial Kalman filter alone.

The cluster track finder has a lower tracking efficiency than the combinatorial Kalman filter implementation, but still has a high tracking efficiency up to noise densities much higher than what has been observed in test beam data. The main advantage of this method is that it is very simple to implement compared to the combinatorial Kalman filter.

Real measurements become harder to separate from noise hits with more scattering and more noise hits. An event with several real tracks will behave differently from a single track with random noise hits. The optimal cut values for the combinatorial Kalman filter and the DAF will vary with these conditions, and a much more realistic simulation would be needed to tune the parameters for real data.

5 Track quality in real data

To test the implementation on real data, a sample consisting of 40000 tracks from a 120 GeV π^+ beam at CERN SPS was used. The data were recorded in November of 2010.



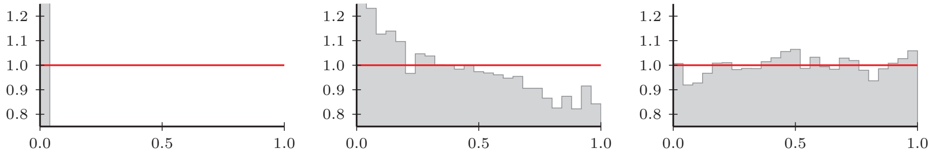
(a) CKF. (b) CKF + DAF. (c) CLU.

Figure 9: P-values of the differences between the true, simulated parameters and the estimated track parameters with 20 noise hits per plane. The distributions are normalized so that the area is 1. The p-values are for the combinatorial Kalman filter alone (CKF), the combinatorial Kalman filter followed by the DAF (CKF + DAF), and the cluster track finder (CLU). The small excess near 0 in the combinatorial Kalman filter without the DAF is likely due to the higher contamination.

5.1 Test statistics

In the simulation, the Gauss-Markov assumptions hold perfectly. For real data, this is not necessarily the case. For the measurements to be unbiased, the geometrical description must match reality. For the normalized residuals to have the same variance, the amount of material the particle passes through must be correctly accounted for, and the uncertainties of all the measurements must be known.

To obtain a correct description of geometry of the measurement planes, the Eutelescope package uses Millepede II[4]. To estimate the amount of material in the experimental setup, as well as the resolution for all the measurement planes, the method described in [9] was used. After this, the fitted χ^2 of the track sample closely follows the theoretical distribution, as shown in Figure 10.



(a) Before alignment. (b) After alignment. (c) After material and resolution estimation.

Figure 10: P-value plots for real 120 GeV π^+ data before numerical alignment, after numerical alignment, and after estimation of material and resolution. Only tracks with hits in all detector planes have been used to create the plots. The distributions are normalized so that the area is 1.

5.2 Error correlation

The third Gauss-Markov assumption is that the measurement errors are uncorrelated. The Kalman filter treats measurements as if they are a function of the true position of the particle, and smeared by a stochastic variable, $\boldsymbol{\varepsilon}$.

$$\mathbf{m}_k = \mathbf{H}x_{true} + \boldsymbol{\varepsilon} \tag{42}$$

The covariance matrix of $\boldsymbol{\varepsilon}$ is the covariance matrix of the measurement.

The ATLAS pixel devices included in the real data have cells that are 50 µm×400 µm. Most clusters in the test beam do not extend beyond one 400 µm wide column, so the position estimates tends to be in

---

[4] see `http://http://www.desy.de/~blobel/mptalks.html`

the center of the column. The measurement errors then depend more on the position of the particle within the pixel column, than the stochastic nature of the signal formation in the sensor. These measurements are also highly non-Gaussian, as the position of particles within the cell is close to a uniform distribution in the experiment.

If the columns of two ATLAS pixel sensors are aligned so they have a large overlap in the collimated beam, the errors will be correlated, as can be seen in Figure 11b. Including correlated measurements in the track fit can make the true resolution of the smoothed predictions worse for all the planes, even when the fitted χ^2 are still very close to the theoretical distribution.

In the short pixel direction, the problem is smaller, since the stochastic nature of the measurement errors are larger compared to the pitch.

6 Summary and conclusions

The Kalman filter has been implemented in the information filter formulation for straight line track fitting for data taken with the EUDET beam telescope. The implementation has been used in a combinatorial Kalman filter for track finding, as well as in an implementation of the DAF. The methods take advantage of the collimated beam, and a very simple track model.

In simulated data, with a sensor detection efficiency of 95%, and requiring 4 or more hits per track, the combinatorial Kalman filter is able to find close to the theoretical limit of 99.78% of all the tracks at much higher noise densities than what is common in real data.

The DAF is able to improve track finding by increasing the probability of rejecting noise hits as well as reducing the probability of rejecting a real measurement. The precision is improved compared to the combinatorial Kalman filter alone.

A very simple to implement track finder based on a cluster finding algorithm and the DAF has been implemented. The track fitter has a lower tracking efficiency than the combinatorial Kalman filter, but performs well at low noise densities. The fitted tracks are of the same quality as tracks fitted with the combinatorial Kalman filter followed by the DAF.

Due to the highly collimated beam in the SPS test beam, error correlations in the long pixel direction can occur between ATLAS pixel planes. Including correlated measurements in the fit can make the precision of the fit worse.

After numerical alignment and numerically obtained estimates of material amounts and sensor resolutions, the fitted χ^2 of tracks from real data is very close to following the χ^2, even with non-Gaussian, correlated measurement errors in the ATLAS pixel devices.

A Appendix: Speed optimization of the fitter

The information filter described here was used to implement the material and resolution estimation method described in [9]. The method requires the track sample to be refitted several thousand times, and a fast track fitter is needed.

Running two Kalman filters with two dimensions, one for x and dx/dz and one for y and dy/dz, is faster than running one with all four parameters. Covariance between x and y is introduced if the measurement covariance matrices have non-zero off-diagonal elements. The pixel devices give uncorrelated measurements in x and y in the local frame. But if the device is rotated around the z-axis, there is correlation between x and y in the global frame of the track fitter.

Since the misalignment in the rotation around the z-axis should be small, the corrections to the measurement covariance matrices from the alignment constant have been ignored to gain speed and simplicity.

Using 32-bit floating point numbers instead of 64-bit double precision floating point numbers, and using the single instruction, multiple data (SIMD) instruction sets available in modern computer architectures can improve the speed of track fitting [10]. The implementation described here has been created using the linear algebra package Eigen[5] that uses explicit vectorization, and using 32-bit floating point numbers.

References

[1] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Reasearch A*, 262:444–450, 1987.

[2] R. Mankel. A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. *Nuclear Instruments and Methods in Physics Reasearch A*, 395:169–184,1997.

[3] R. Frühwirth and A. Strandlie. Track fitting with ambiguities and noise: A study of elastic tracking and nonlinear filters. *Computer Physics Communications*, 120:197–214, 1999.

[4] I. Rubinsky. An EUDET/AIDA pixel beam telescope for detector development. *Physics Procedia*, 37, 2012.

[5] V. Highland. Some practical remarks on multiple scattering. *Nuclear Instruments and Methods in Physics Reasearch*, 129:497–499, 1975.

[6] W.W. Hager. Updating the Inverse of a Matrix, *SIAM Review*, 31:221–239, 1989.

[7] A. Sen and M. Srivastava. Regression Analysis, *Springer*, 1990.

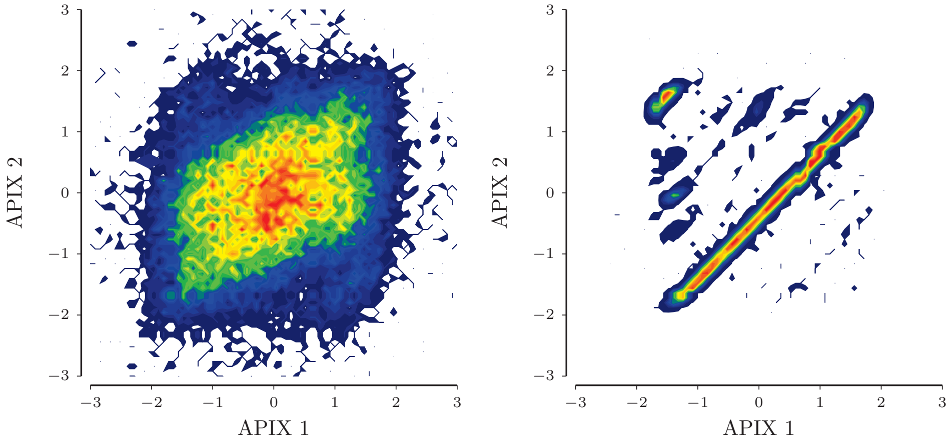
[8] S. S. Wilks. Certain generalizations in the analysis of variance Biometrika 24 (3-4): 471-494, 1932

[9] H. Gjersdal, Frühwirth, M. Nadler and A. Strandlie. Optimizing track reconstruction by simultaneous estimation of material and resolutions. *Journal of Instrumentation*, 8, 2013.

[10] S. Gorbunov, U. Kebschull, I. Kisel, V. Lindenstruth, Volker and W.F.J. Mueller. Fast SIMDized Kalman filter based track fit. *Computer Physics Communications*, 178:374–383, 2008.

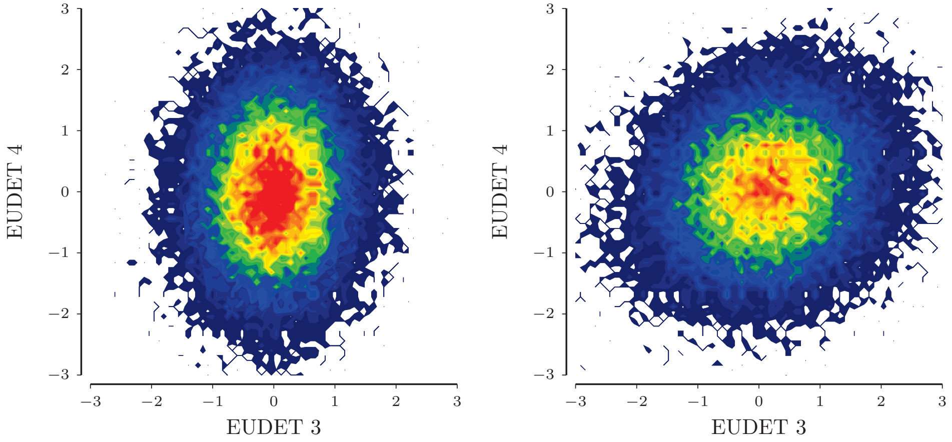
---

[5] see `http://eigen.tuxfamily.org/`



(a) x direction, 50 µm pixel pitch.

(b) y direction, 400 µm pixel pitch.



(c) x direction, 18.4 µm pixel pitch.

(d) y direction, 18.4 µm pixel pitch.

Figure 11: The two top plots show the normalized residuals in the first ATLAS pixel plane (APIX 1) versus the second ATLAS pixel plane(APIX 2). The ATLAS pixels are 50µm × 400µm, and correlations are clearly visible between the measurements in the long direction between the planes. The two bottom plots show the residuals in the third EUDET plane versus the fourth. The residuals are calculated from smoothed predictions where both hits are excluded from the fit. The plots are made from real data.

Appendix B

Optimizing track reconstruction by simultaneous estimation of material and resolutions.

Introduction

A particle going through the detector system of the test beam experiment will pass through sensors, support structures for the mounting of the devices and material that is in place for the cooling of the devices under test. The total amount of material is kept as low as possible, but the material distribution is not known exactly.

The resolution of the sensors vary depending on tuning of front-end electronics, the angle of the particle beam and the Lorentz angle. With new sensor technology, the resolution is not known exactly in any condition, and estimating it is one of the goals of the experiment.

Optimal performance of the track fitter, including the correct description of the track uncertainties, can only be achieved with the correct descriptions of the resolutions and thicknesses of the material planes in the beam.

The residual distributions are a convolution of measurement uncertainties and track prediction uncertainties. The track prediction uncertainties depend on the resolutions of all the measurements included in the track estimate, and the amount of material the particle has passed through. In order to extract information about the plane thickness or measurement uncertainties from the residuals, the correct descriptions of the resolutions and thicknesses of all other planes the track passes through are required.

Several new methods for obtaining thickness and resolution estimates were introduced in the following paper. The methods rely on multidimensional minimization methods, and the resolution and thickness parameters of a tracker system are estimated simultaneously.

A simulation study performed showed that the methods are able to correctly and robustly estimate the thicknesses and resolutions of a simulation experiment with a geometry similar to the test beam experiment.

The precision of the track fitter was improved in real data by applying the method to a test beam data set. Excellent track quality was obtained by using the estimated parameters when fitting the data.

Appendix C

Tracking Efficiency and Charge Sharing of 3D Silicon Sensors at Different Angles in a 1.4 Tesla Magnetic Field.

Introduction

The presence of a magnetic field affects the drift of the charge carriers in a sensor, and can affect the signal response of a semiconductor detector. ATLAS 3D pixels are expected to see smaller effects from the presence of a magnetic field in the configuration of the insertable B-layer than the planar sensors used in the original ATLAS pixel system.

The first results from a test beam experiment with 3D devices in a magnetic field are presented in the following. The data was taken with a full-3D sensor produced at Stanford, with electrodes penetrating the entire sensor wafer. The beam consisted of 180 GeV pions in a 1.4 T vertical magnetic field. The Bonn ATLAS telescope was used for tracking of the beam particles.

The studies presented are: the detection efficiency in the sensor in a magnetic field with tracks orthogonal to the sensor, and with the device tilted similarly to the configuration in the insertable B-layer; the detector response with and without a magnetic field, both tilted and non-tilted; how charge sharing behavior changes with tilt angle and the presence of a magnetic field; the residual distributions at the different tilt angles and with and without a magnetic field.

The observed effects of the magnetic field on the device was small compared to the effects observed in a reference sensor of the type used in the original pixel system.

This paper is published in the proceedings of the 7th International “Hiroshima” Symposium on the Development and Application of Semiconductor Tracking Detectors.

Appendix D

Test Beam Results of 3D Silicon Pixel Sensors for the ATLAS upgrade.

Introduction

The following paper is a summary of the test beam characterization that was performed in the high energy pion beam in the CERN North Area during 2009. The Bonn ATLAS telescope was used for tracking for data taken with the devices under test inside a magnetic field. Due to improved tracking resolution and the possibility of a higher trigger rate, the EUDET telescope was used for tracking with data taken without the magnetic field. The EUDET telescope was not certified for use within a magnetic field at the time of data taking.

A large part of the data was taken as part of an angle scan, with the devices tilted at different angles to the beam axis. Sensors from several manufacturers were tested, both full-3D, and double-sided, double-type column devices. The results were compared to a planar sensor of the type installed in the original pixel system.

The main differences between 3D sensors and planar sensors are the drift of the carriers in the sensor, the way a magnetic field affects the drift of the carriers, active edges in 3D sensors, and the loss of detection efficiency in 3D sensors due to particles passing through the etched out electrodes. Results on the detection efficiency, active edges, charge sharing probabilities, and signal response is presented.