Adaptive methods for the reconstruction of charged tracks in a noisy environment

Are Strandlie

Department of Physics
University of Oslo

Thesis presented for the degree of
Doctor Scientiarum

15th September 2000

Contents

Acknowledgements

The research work described in this thesis has been done during the three and a half-year period ranging from January 1997 to July 2000. The funding has been provided by the Norwegian Research Council through a doctoral fellowship (75 %) and the Department of Technology, Gjøvik College (25 %). In addition, the Norwegian Research Council has provided extra funding for a half-year stay at INRIA, Sophia Antipolis, France from October 1998 to April 1999.

The research has been carried out mainly at the Department of Physics, University of Oslo, at Gjøvik College and at INRIA. However, I have also had numerous short-time stays at the Institute for High Energy Physics, Austrian Academy of Sciences, Vienna and at CERN, Geneva.

First of all, I would like to express my gratitude to my main supervisor, Associate Professor Jørn Wroldsen, Gjøvik College. Jørn has been a continuous source of encouragement and sound guidance and was the main initiator of this project. It has been a pleasure to work together with him. My co-supervisor, Professor Steinar Stapnes at the Department of Physics, University of Oslo, also deserves thanks. He has always been more than willing to pay for my travels, and he has made it possible for me to access simulated data at CERN, data which have been used extensively in the thesis.

My sincere thanks go to Univ.-Doz. Rudolf Frühwirth, Institute for High Energy Physics, Austrian Academy of Sciences. Probably he knows more than anybody else about the details of what I have done in this work. His bright ideas and his warm encouragements have strongly influenced the quality of this thesis. It has been a great privilege both to be supervised and to be allowed to work together with one of the greatest experts in the world today in the field of track reconstruction.

Most of the figures in the introductory part of the thesis have been taken either from the ATLAS Inner Detector TDR [5] or from the ATLAS web pages. I hereby give thanks to the ATLAS community for allowing me to use this material herein.

It has also been a pleasure to work together with Research Director Bjørn Lillekjendlie, SINTEF, Oslo. His ideas have been vital for parts of this work. I will also give thanks to Research Director Josiane Zerubia, INRIA, Sophia Antipolis, France. She was kind enough to invite me to her group at INRIA for six months, a stay which turned out to be very fruitful for the research work.

The friendly and inspiring atmosphere in the Experimental Particle Physics group at the University of Oslo has been a source of inspiration during my Ph. D. project. In particular I would like to mention Professor Lars Bugge, who helped me come in contact with the Vienna group and thereby made a vital contribution to the successful completion of this work.

My sincere thanks go to my father and mother, Arne and Sissel Strandlie. The adolescence in my parents' home has been characterized by academic discussions and curious reasoning and has therefore strongly contributed to the desire of carrying out the piece of research described in this thesis. In this context I would also like to mention my parents-in-law, Ellen and Aasmund Haanes, for their care, support and interest in my work.

My good friends Øystein Elgarøy, Ivar Farup, Geir Otto Holmås and Bjørn Henrik Vangstein also deserve to be mentioned. The social life outside work hadn't been the same without them.

Last, but definitely not least, I will express my most heartfelt thanks to my precious wife, Hanne Kari, and my lovely daughter, Kristine. Hanne Kari has been an invaluable support in both the good and the bad times, and she has been very patient with a husband sometimes too much focused on his work. Kristine has, in a positive sense, turned our lives upside down and is a strong reminder of the fact that life also consist of things much more important than the daily work.

Gjøvik, July 2000
Are Strandlie

Summary

The amount and the complexity of the data that is to be analysed at the next generation of particle physics experiments at CERN will significantly exceed that of today's experiments in the same laboratory. This is mainly due to the fact that some of the sought events at LHC have very small cross-sections, something which necessitates many simultaneous collisions in the detectors as well as a very large collision rate. The *track reconstruction* algorithms applied at LHC will therefore have to reconstruct interesting tracks in a much larger background of low-momentum tracks than what is the case at LEP. It is safe to predict that a successful LHC algorithm will have to exhibit a larger degree of robustness towards noise than a LEP algorithm. Moreover, the large quantities of data that need to be analysed also require the algorithms to be fast, something which puts even stronger demands on the quality of the track reconstruction methods.

The main goal of the work described in this thesis is to try to meet some of the above-mentioned requirements. We present herein a number of new track reconstruction algorithms which we believe are well suited for analysing data of the LHC era. Some of them exhibit a larger degree of robustness than the currently most popular ones. Others will be shown to be very fast, but nevertheless more precise than existing algorithms of comparable computational complexity.

Another outcome of the present work is that we believe a bridge has been built between two classes of approaches, which earlier have been considered as distinct: algorithms based on filtering techniques and elastic tracking algorithms. One of the novelties of this work, the *Deterministic Annealing Filter (DAF)* of Paper 3, has properties resembling both. The basic filter structure of the DAF gives it the same appealing statistical properties as the Kalman filter. At the same time it exhibits the same pattern recognition abilities as the Elastic Arms algorithm. It is shown that this combination is a powerful one when applied to the reconstruction of particle tracks embedded in significant amounts of noise.

Since the LHC not will be operational until 2005 at the earliest, the experiments of this thesis have been based on data simulated by computers. All simulated data come from the tracking detectors of the ATLAS experiment, and there has been a particular focus on the Transition Radiation Tracker (TRT) of the ATLAS Inner Detector. The

TRT is the only part of the Inner Detector with ambiguous or drift time measurements. This, together with additional noise from nearby and crossing tracks, makes tracking in the TRT very challenging, and it has therefore served as an excellent testbed for our algorithms.

When evaluating the performance of track reconstruction algorithms based on simulated data, one should of course use data which as closely as possible resemble the real data of the experiment. The natural thing to do in a project like this would thus be to test out the algorithms by implementing them in one of the official track reconstruction programs in ATLAS. Regrettably, these software packages consist until today mostly of sets of closely interconnected algorithms, not being designed for simple implementation of and comparison with alternative algorithms. The algorithms developed herein have therefore been evaluated in stand-alone computer programs, mainly with respect to their basic statistical properties. However, there are now major efforts being put into transforming the existing code into object-oriented technology as well as modularizing it, making it much easier to test out alternative strategies. After the completion of this modularization phase, our algorithms should be implemented in the official framework and their performance be evaluated in such a realistic setting.

One of the major ideas behind this Ph. D. project was to create a possibility of new views and ideas concerning track reconstruction by establishing contacts with a variety of research groups, both physics and non-physics groups. This resulted in fruitful inspiration from people with a strong background in as different fields as physics, statistics and image processing, as can be seen on the author lists of the papers. Since track reconstruction indeed is an interdisciplinary subject, it is not surprising that cooperations between people with such different backgrounds turn out to be successful.

The thesis is organized as follows. First we give a general introduction and motivation for the work by briefly presenting some of the physics challenges at LHC. Then the LHC accelerator, the ATLAS detector and some of its subdetectors are described, followed by an introduction to the basic concepts of track finding and track fitting. A review over some of the most important track reconstruction methods is also given. The major part of the thesis, six research papers, follows thereafter. Four of these papers have been accepted for publication in international journals, and two have been published in proceedings from international, refereed conferences. The papers are succeeded by a summary and a discussion of the papers, and the thesis is concluded by a brief outlook to further research. Two additional papers, also published in proceedings from international, refereed conferences, are finally included in an appendix. The contents of these are quite similar to those of two of the journal papers in the main part of the thesis. However, they reflect a part of the work done during the project and are thus appended.

Introduction

1 General background

The new discoveries of physics during the 20th century have revolutionized our understanding of the basic structure of our world and the universe around us. At the turn of the century, some physicists believed that classical mechanics, thermodynamics and electromagnetism contained almost everything there was to say about nature. However, the quantum hypothesis of Max Planck in 1901 [1] was the beginning of a scientific revolution which at that time was beyond anyone's imagination. During the next three decades special and general relativity were established, as well as non-relativistic, one-particle quantum mechanics. Soon thereafter, the first attempts of creating relativistic many-particle theories were made.

In addition to the general theory of relativity, which describes the gravitational interaction, the most fundamental description of nature today is contained in a relativistic quantum field theory called the Standard Model (SM). One part of the SM is QCD (Quantum Chromodynamics), which describes the interaction between the quarks inside hadrons and mesons. This so-called *strong* interaction is mediated by massless particles called *gluons*. The other part of the SM is the unified theory of *weak* and *electromagnetic* interactions called the *electroweak* theory. The electroweak interaction is mediated by the massless *photon* and the massive W^+, W^- and Z^0 *bosons*. The existence of these massive bosons was predicted by the electroweak theory (developed by Glashow, Salam and Weinberg in the late 1960's [2]), and they were discovered by the UA1 experiment at the SPS collider at CERN in 1983. For this achievement, Carlo Rubbia and Simon van der Meer were awarded the Nobel prize in physics in 1984.

Since the time of UA1, a large number of experiments have been carried out in order to further verify the SM predictions. In addition, a lot of effort has been put into attempts of observing effects which cannot be explained within the framework of the model. Until today, no experimental results have shown evidence of incorrect SM predictions. On the contrary, the model has been verified to experimental errors significantly smaller than those of 15 years ago. For instance, one was in the 90's finally able to discover the top quark. The announcement of this discovery was made at Fermilab in the USA in 1995 [3].

Despite the remarkable successes of the SM, there are a number of important questions yet being unanswered. Some of these are:

- Why are there three generations of quarks and leptons?
- How do quarks and leptons acquire masses?
- Why are the force carrying W and Z bosons massive, while the photon is massless?
- Are the quarks and leptons really fundamental particles, or do they have a substructure?

The question of how quarks, leptons and bosons acquire mass is partially answered within the SM by postulating the existence of a Higgs field, leading to the prediction of a Higgs particle. This particle has not yet been discovered, and in present experiments, for instance at the LEP collider at CERN, significant efforts are being put into performing searches for the Higgs boson. If it is not discovered at LEP, a possible discovery of the Higgs will have to be made at experiments operating at higher energies than those accessible today.

Another shortcoming of the SM is the fact that the strong interaction is not described in the same theory as the electroweak interaction, something which is desirable for many theorists. Also, the SM is not a stable theory to higher energies, unless nature has provided us with very unlikely cancellations of radiative effects. This points towards general principles, yet unknown, that guarantee these cancellations. A central ingredient in the most promising attempts of making stable theories at higher energies is the concept of *supersymmetry*. If supersymmetry is a fundamental symmetry of nature, a large number of supersymmetric particles should exist. For instance, supersymmetry predicts that there is not only one but many Higgs bosons, and it is believed that supersymmetric particles are accessible at energies not far away from those of today's experiments. Experimental verification of supersymmetric particles would thus be a major leap forward in the understanding of the basic structure of the universe.

In order to try to answer some of the questions mentioned above, new accelerators are being designed. One of these is the Large Hadron Collider (LHC) at CERN, currently being constructed. Some of the major goals of the experiments at the LHC are to [4]:

- discover the Higgs particle or equivalent effects in a mass range from the lower limit given by today's experiments up to a mass of 1–2 TeV/c^2
- search for supersymmetric particles
- look for possible quark and lepton substructure
- investigate CP-violation in B-systems

2 The LHC accelerator and the ATLAS detector

The fulfilment of the above-mentioned goals requires an experimental setup which pushes the limits of today's technology. The LHC is a circular collider which will be built in the same tunnel as the existing LEP collider (consult Figure 1 for a schematic drawing). The LHC accelerator ring is approximately 27 km in circumference, and the accelerated particles will be protons. At certain *collision points* around the ring the protons will be forced into head-on collisions with one another, and here *particle detectors* will be installed. The largest of these detectors are ATLAS and CMS, and their positions are indicated on the figure.

The LHC will collide protons at a centre-of-mass energy of about 14 TeV. According to the current schedule, the LHC will start to take data during the year 2005. Since the most interesting cross-sections of LHC-processes are very small, the collision rate has to be very large. For this reason, about 40 million particle *bunches* will collide each second at the LHC. Each bunch will contain approximately 10^{11} protons, and the bunch spacing is 7.5 metres. The current in the LHC ring will be 0.56 A, and the stored

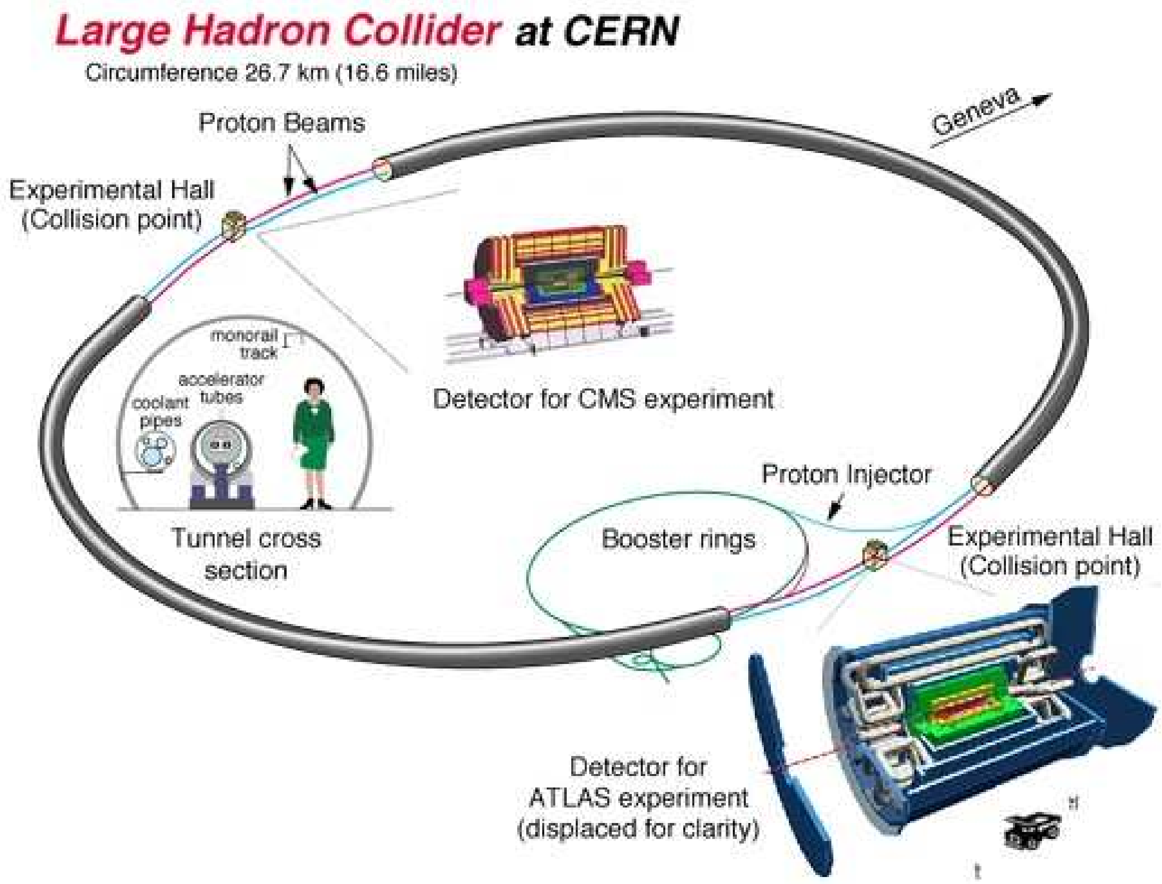


Figure 1: The LHC.

energy per beam will be 350 MJ.

One of the two major, general-purpose experiments at LHC is called *ATLAS* (see Figure 2). This detector will be approximately 25 m long and 20 m high, i.e. it will be as large as a five-storeyed building. Given the *pp* inelastic cross-section and the beam intensity at LHC, an average number of 23 collisions will be observed at the same time in the detector at full luminosity. Thus, the interaction rate in ATLAS will be about 1 GHz. Since the bunch-crossing rate is 40 MHz, and since the average event size in ATLAS is about 1 MB, this would imply a recording rate of about 40 TB/s. Such an amount of data is neither needed for the physics analyses nor technically possible to store. Therefore, a *trigger system* which selects the most interesting events has to be constructed. In ATLAS, the proposed trigger system consists of three levels, as shown in Figure 3. The level-1 trigger reduces the event rate to about 75 kHz, while the output rate of the level-2 trigger is about 1 kHz. The level-3 trigger or the *event filter* further reduces this rate by an order of magnitude, so that the final recording rate is about 100 Hz or 100 MB/s. This means that the storage requirements of data from the ATLAS detector will be about 1 PB or 10^{15} B per year. It is a big challenge to be able to store such an amount of data, as well as making the data available for physics analyses.

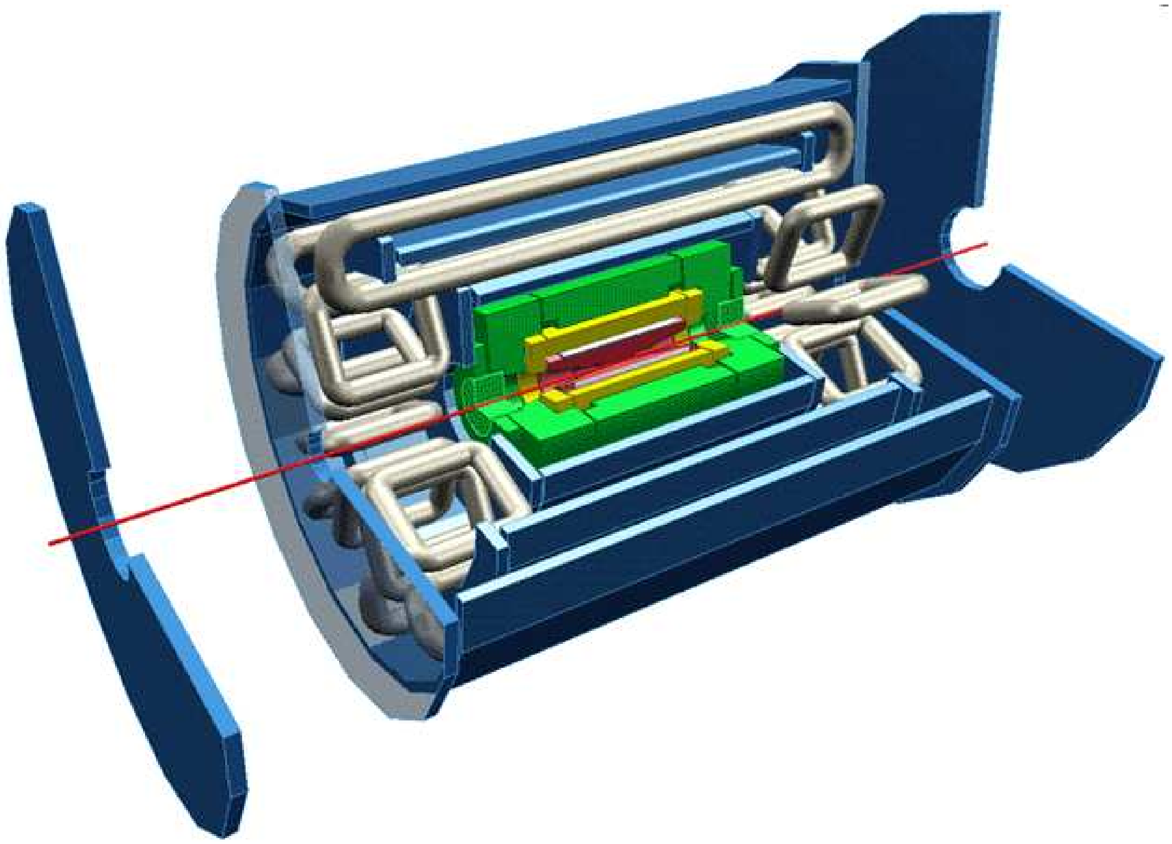


Figure 2: The ATLAS detector.

The ATLAS detector consists of five main parts:

- Inner Detector
- Electromagnetic Calorimeter
- Hadronic Calorimeter
- Muon Chambers
- Magnet Systems

The innermost part of the detector is the Inner Detector. The main tasks of the Inner Detector are to yield position measurements of the particles traversing the detector, making it possible to reconstruct the tracks and the vertices of the collision events with a high degree of accuracy, as well as to contribute to the identification of electrons, muons and photons. The Electromagnetic Calorimeter will measure the energy of electrons, positrons and photons, while the Hadronic Calorimeter will measure the energy of hadrons (e.g. protons, pions, kaons and neutrons). The Muon Chambers are specifically designed to determine the momenta and sign of charge of muons. There are two Magnet Systems: the central solenoid for the Inner Detector and the air-core toroid surrounding the Muon Chambers.

The Inner Detector [5] (see Figure 4) is the main tracking device of the ATLAS experiment. It provides a large number of high-resolution position measurements for each charged track, the positions of the measurements ranging from close to the beam to a distance of about 1 m radially out from the beam and about 3.5 m in the direction

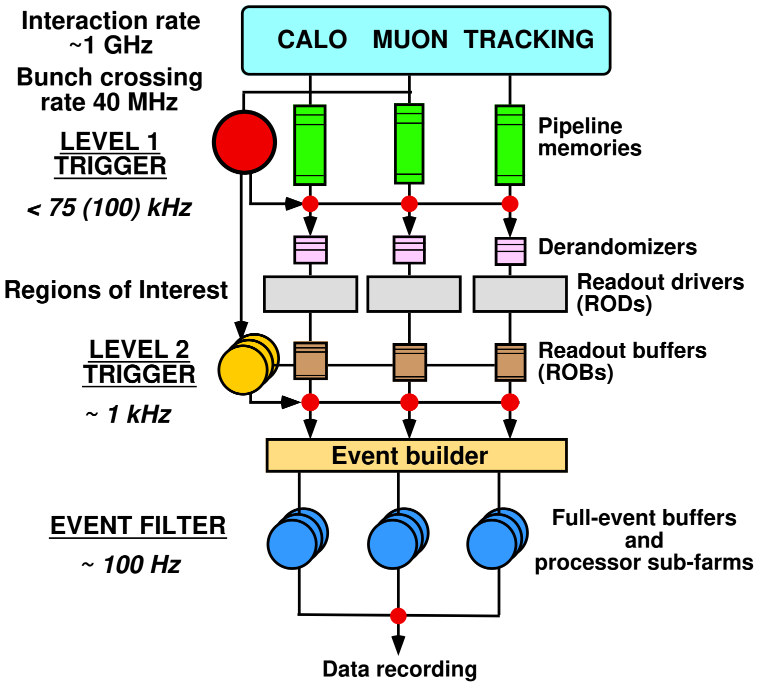


Figure 3: The trigger system in ATLAS.

parallel to the beam. The detector consists of three separate subsystems:

- The Pixel Detector
- The Semiconductor Tracker (SCT)
- The Transition Radiation Tracker (TRT)

Each of the subsystems consists of a *barrel* part and an *end-cap* part. The barrel detector elements are arranged in cylindrical layers with the symmetry axis along the beam line. The end-cap detector elements are arranged in wheels which are positioned perpendicularly to the direction of the beam. It is assumed that 3–4 % of the detector elements will be *inefficient*, i.e. with a probability of this percentage the detector elements will not give a signal above threshold, even though a charged particle has passed through. A transverse view of the barrel part of the Inner Detector is shown in Figure 5.

A plot of the *material budget* of the Inner Detector as a function of the absolute

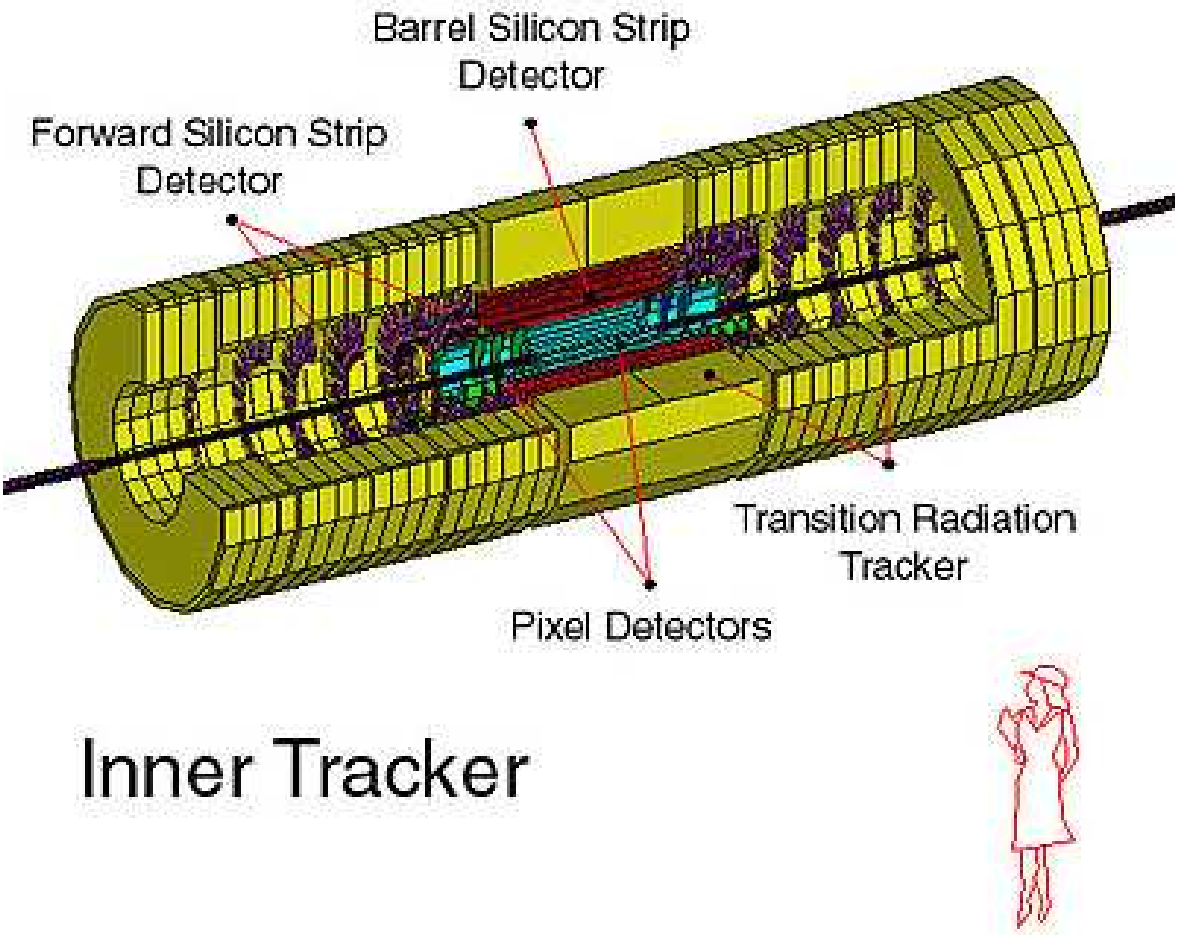


Figure 4: The ATLAS Inner Detector.

pseudorapidity is shown in Figure 6. Here the amount of material, in terms of radiation lengths, is plotted against the absolute pseudorapidity. The amount of material up to and including the different subdetectors is shown in histograms. The histogram labelled "Total" reflects also material being on the outside of the sensitive volume of the Inner Detector. It can be noted that the average amount of material up to and including the TRT is 43 % of a radiation length. Thus, multiple Coulomb scattering cannot be neglected for all but the highest momenta.

The Pixel detector elements are made of silicon. The size of a pixel is 50 μm times

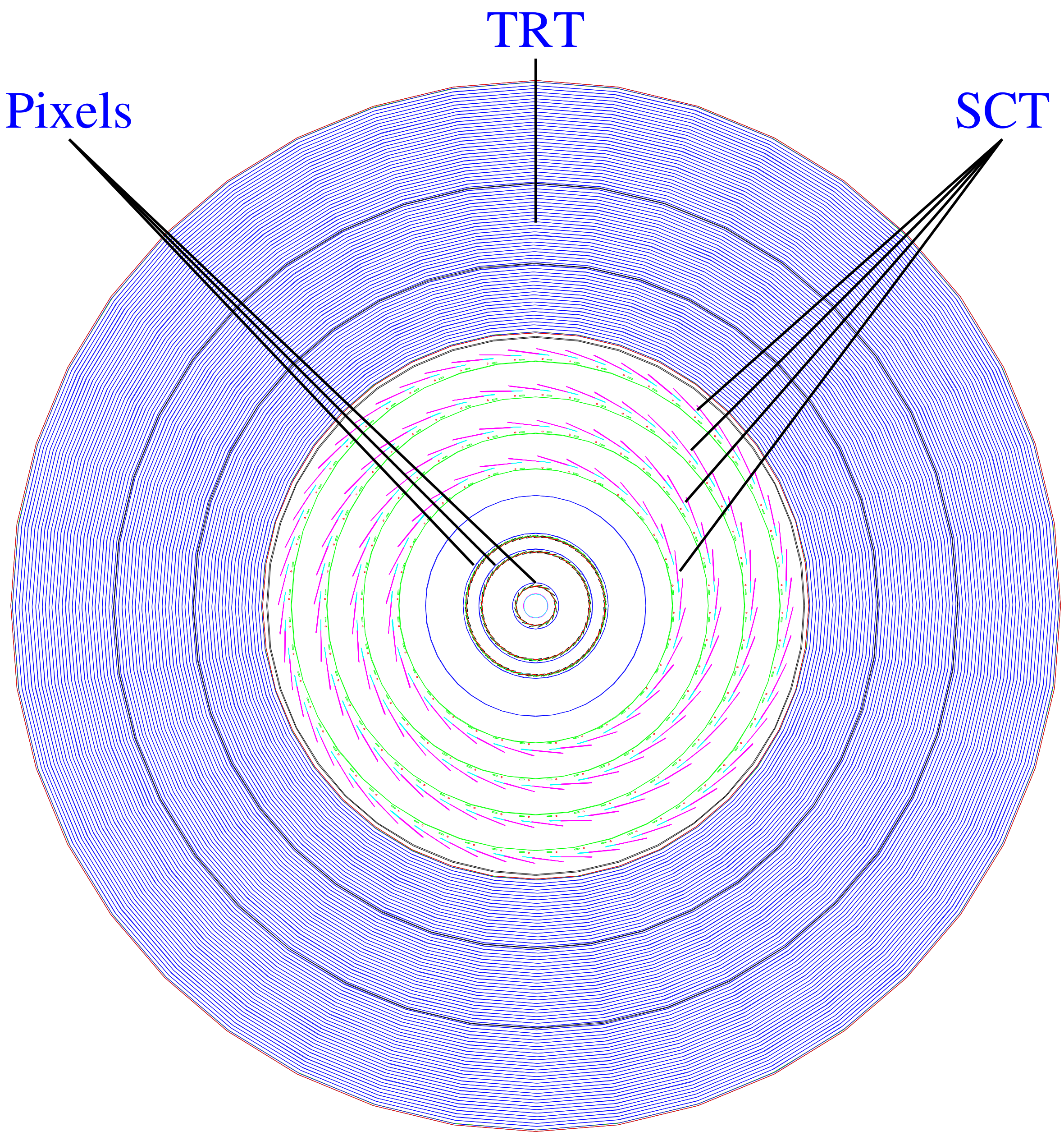


Figure 5: A transverse view of the barrel part of the ATLAS Inner Detector.

300 μm, which in the barrel gives resolutions of 12 μm in $R\phi$ and 66 μm in z.[1] There are in total about 140 million channels in the Pixel system.

The SCT detector elements are also made of silicon. However, the basic elements are *strips*, so that the read-out from one element has to be combined with the read-out of a nearby, slightly tilted *stereo layer* in order to yield a $3D$ measurement. Again for the barrel, the result is a very good resolution in $R\phi$ (16 μm), while the resolution in z is significantly worse (580 μm). The number of channels in the SCT is about 6 million.

1 The distance in $R\phi$ is defined as the distance along a circular curve defined as the intersection between a cylindrical detector layer and a plane perpendicular to the beam pipe, while the distance in z is the distance parallel to the beam pipe.

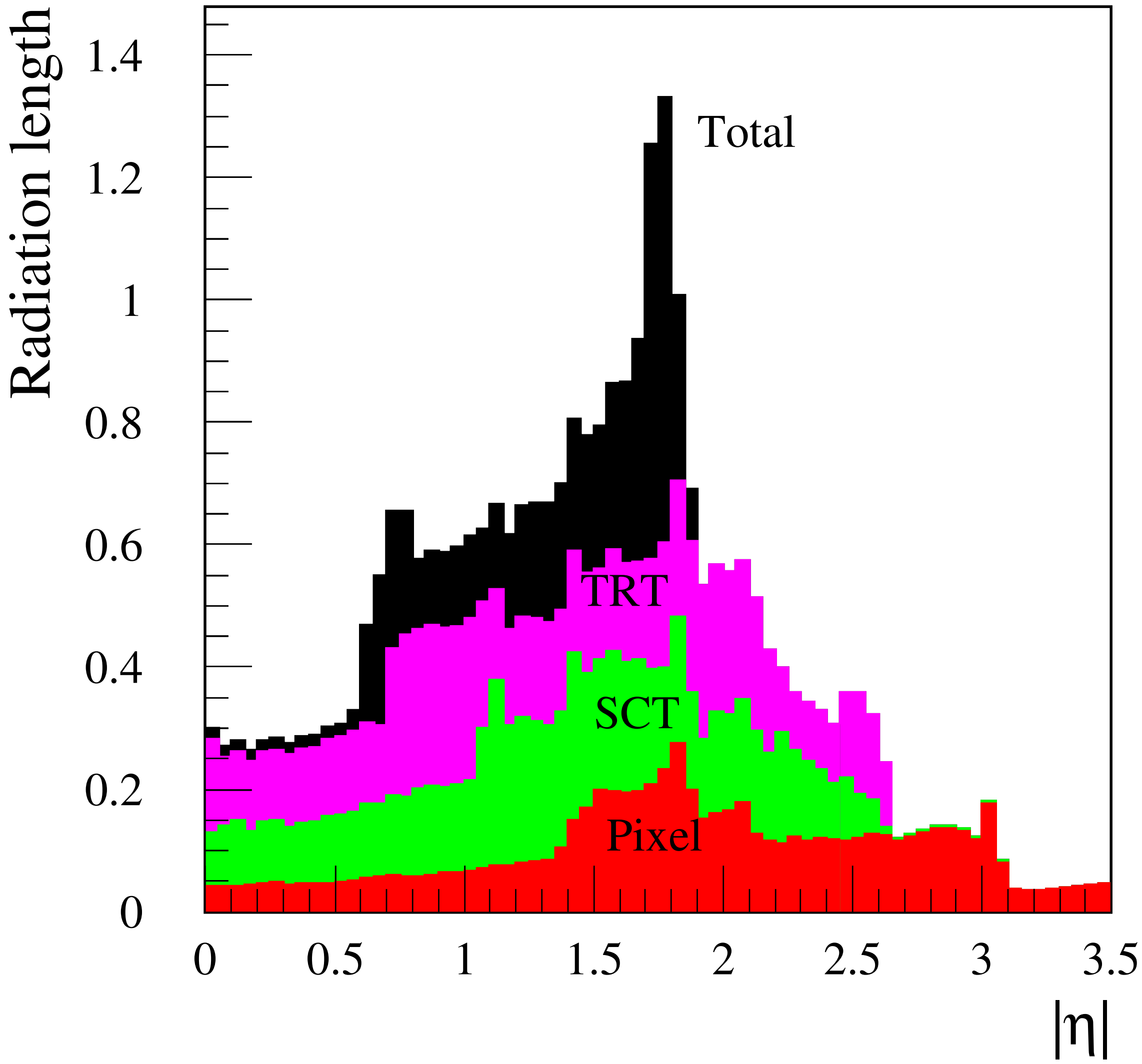


Figure 6: The material budget of the ATLAS Inner Detector.

The TRT is a *drift tube* detector, which yields two-dimensional measurements with a resolution of 170 μm. For the barrel part of the TRT, these tubes are positioned more or less along a cylindrical detector layer. The barrel ranges from a radial distance of about 0.5 m from the beam axis to about 1.0 m. There are about 75 layers of tubes, and there is a radial separation of 6.8 mm between the layers. Inside a layer, the separation in $R\phi$ between two drift tubes or *straws* is also approximately 6.8 mm. A straw has a diameter of 4 mm. In the barrel TRT, there are in total about 50000 straws. In the whole TRT, there are approximately 420000 read-out channels. See Figure 7 for a more detailed view. If a straw has fired during an event, the information about the *drift distance* is also supplied. The drift distance is the distance from the point of closest

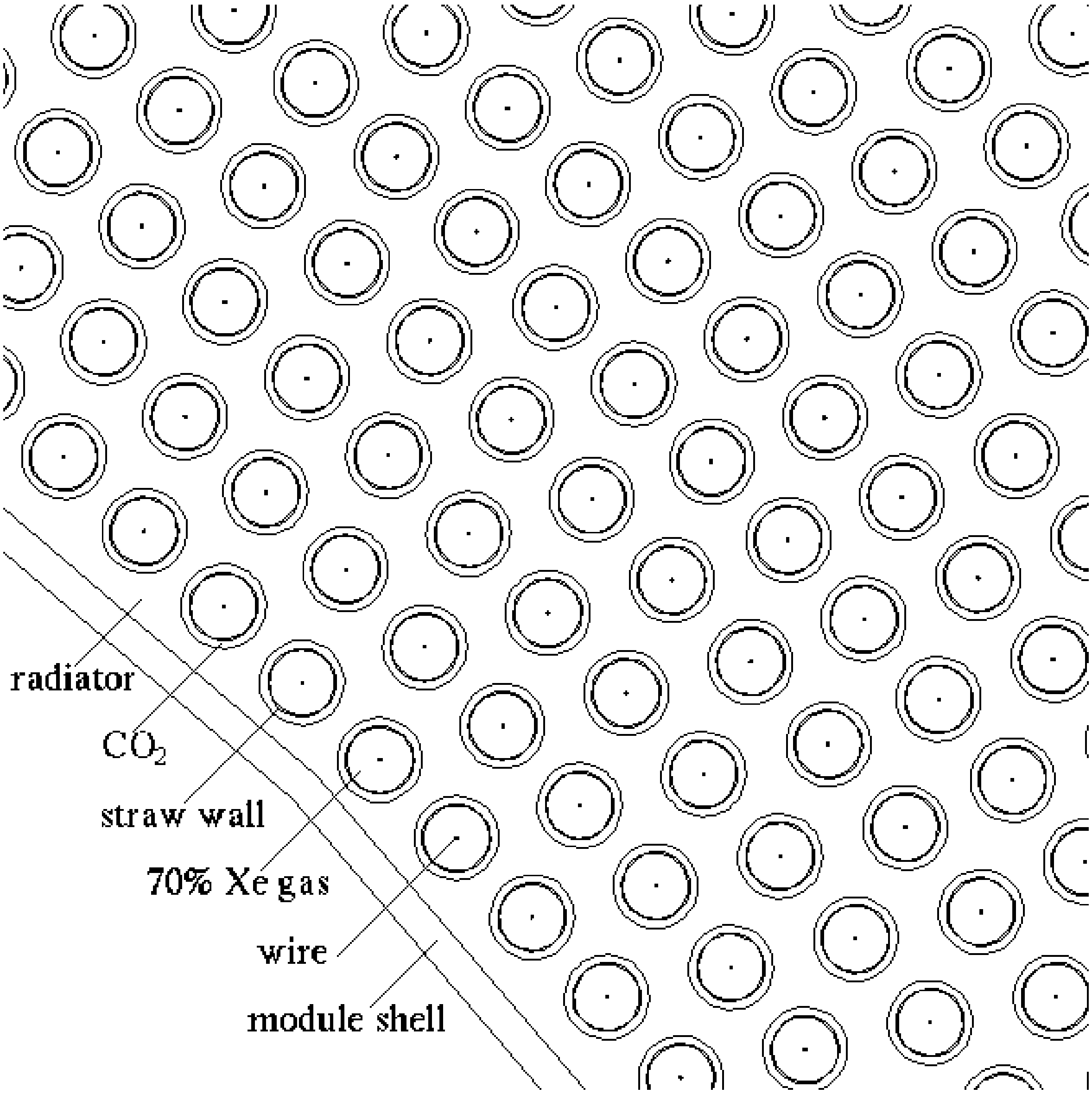


Figure 7: A detailed view over one part of the barrel TRT.

approach of the particle to the straw centre. However, it is not known at which side of the straw centre the particle has passed. This inherent ambiguity of drift tubes is often called the *left-right ambiguity.* It can be a non-trivial problem to efficiently resolve these ambiguities, and this is indeed the case with the data from the ATLAS TRT. Some of the work of this thesis is devoted to exactly this problem.

3 Track finding and track fitting

A simulated collision event in the barrel part of the ATLAS Inner Detector can be seen in Figure 8. This simulated event is a Higgs boson decaying through a b quark and its antiparticle $\bar{b}$. These quarks do again form *jets* of high-energetic particles, of which the ones with charge deposit a small amount of energy in the detector elements they pass through. This energy is registered and defines thus a *hit* in the detector. Due to the large number of simultaneous particle collisions at LHC, there are many *pile-up* events in addition to the interesting event in the detector. Moreover, at least in some detectors there is also pile-up from the preceding bunch-crossings. This makes the task of *track reconstruction* an extremely challenging one at the LHC experiments.

Track reconstruction can traditionally be divided into two main parts:

- **Track finding**: The input to this procedure is a list containing all hits in a detector or a region of a detector, as well as the associated position coordinates. The task is to divide these hits into a number of subsets, where the hits inside a subset are supposed to belong to the same particle. The number of subsets thus corresponds to the assumed number of particles. Track finding can be regarded as a specific case of a **pattern recognition** task.
- **Track fitting**: The input to the track fit consists of the position measurements of a subset as given by the track finder. The task of the track fitter is to, in an optimal manner, estimate the parameters of the curve describing the track, as well as to test the track hypothesis. For tracks originating from a detector embedded in a reasonably homogeneous magnetic field, the sought curves are helices. The output of the track fit is a list of the found particles, represented by estimates of the parameters of their tracks, the covariance matrices of the track parameters and the χ^2's of the fitted tracks.

Given the parameters of a track, the momentum vector of the particle at the collision point or the *vertex* can be calculated. This momentum vector forms the basis of an experimental determination of physical quantities such as differential cross-sections and invariant masses. It is therefore of the utmost importance to be able to make reliable and, in some statistical sense, optimal estimates of the track parameters.

At today's experiments, the strict division between track finding and track fitting is no longer valid. For instance, the complexity of the data at LEP experiments requires the application of methods which solve the problem of assigning hits to tracks

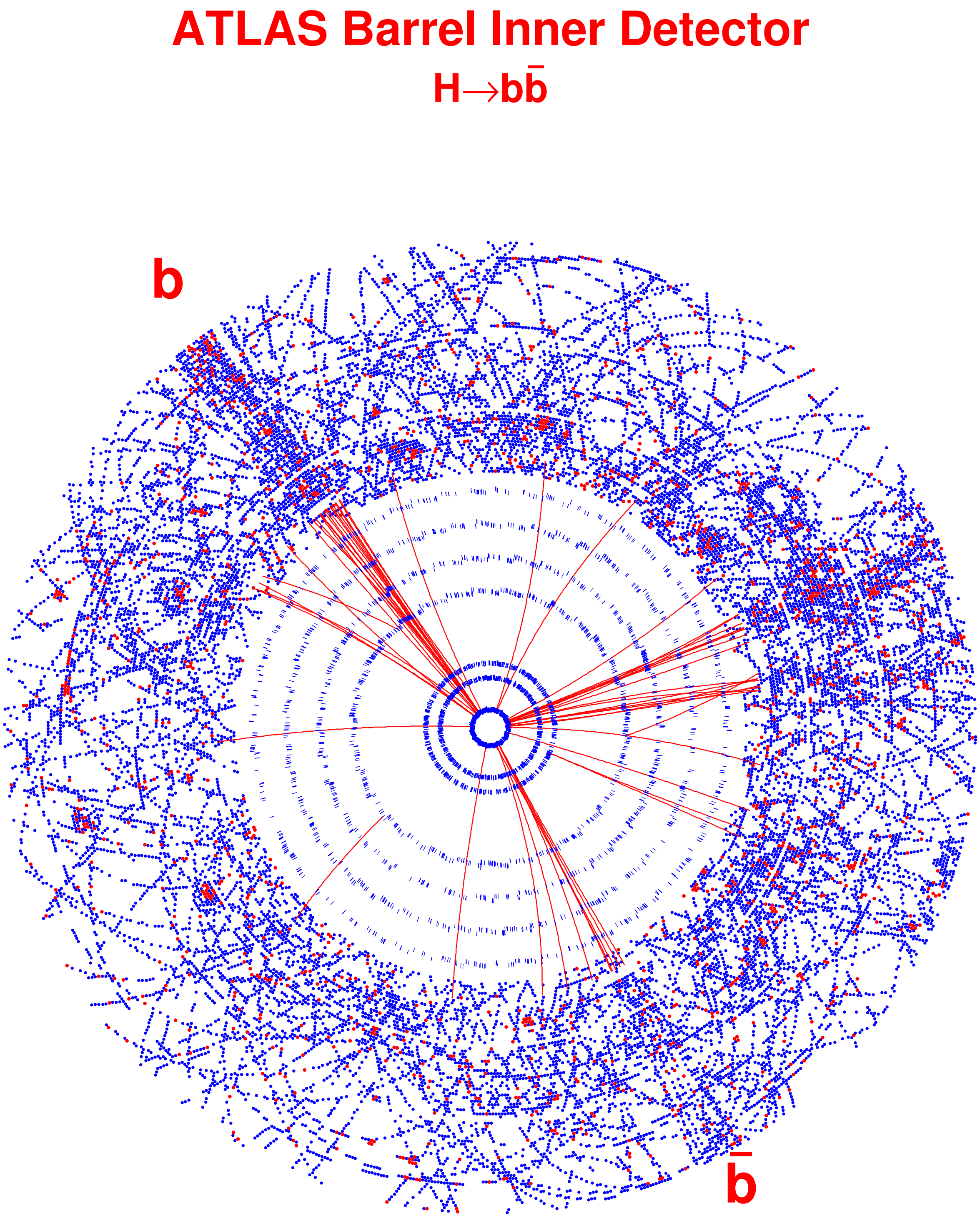


Figure 8: An $H \to b\bar{b}$ event.

concurrently with estimating the track parameters. An example of such a method is the *Kalman filter* [6], introduced at the DELPHI experiment at LEP and now widely used in experiments all over the world. In addition, the track reconstruction methods at LEP need to be able to identify and remove points belonging to other, nearby tracks or originating from noise in the electronics, i.e. they have to exhibit some kind of *robustness* towards *outliers*.

At the LHC experiments, the amount of data to be analysed during track reconstruction will be several orders of magnitude larger than at the LEP experiments. The background of low-energy tracks will also be larger than ever before. It is therefore strongly desirable to apply fast and robust track reconstruction methods at LHC. The work of this thesis is an attempt to supply the high-energy physics community with new track reconstruction algorithms exhibiting some of the properties needed for the LHC era.

4 Review of some pattern recognition and track fitting methods

We will in this section make a review over some popular methods for pattern recognition and track fitting. For an excellent and exhaustive review the reader is referred to the book by Frühwirth et al. [7].

4.1 The Hough transform

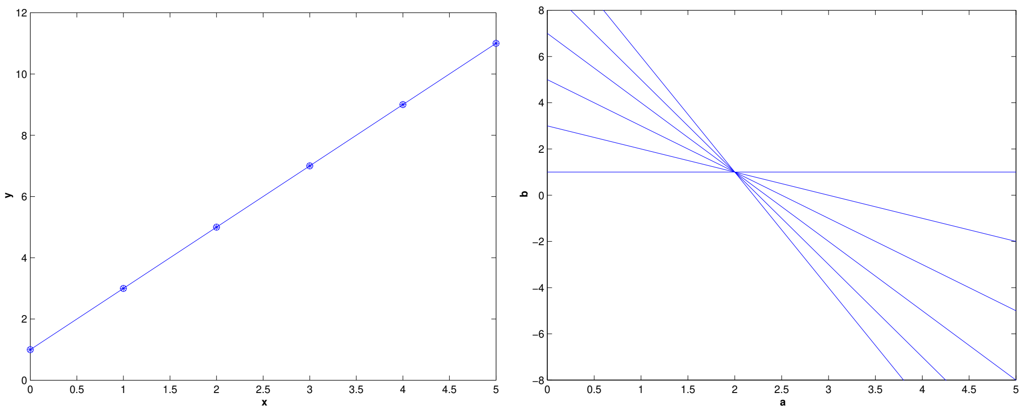
The Hough transform is a pattern recognition method which is extensively applied to line and curve detection in digital images. The method, however, was originally invented to detect particle tracks in bubble chamber pictures [8]. Some years later, the method was discovered by the image processing community and further developed there. During the later years, the method has again been given attention among physicists working in particle physics experiments, and it is now in widespread use. For instance, it has been successfully applied to heavy ion data from the PHENIX experiment at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven [9]. The Hough transform is also part of *XKalman* [10], one of the official software packages for track reconstruction in the ATLAS Inner Detector.

The basic idea is easily explained by looking at a simple example. Consider a straight line in the xy plane given by the equation $y = 2x + 1$. This is a special case of the general formula for a straight line, $y = ax+b$, with $(a, b) = (2, 1)$. The Hough transform arises naturally after a reformulation of the straight line equation: $b = -ax + y$. Thus, a point (x, y) is seen to define a straight line in a two-dimensional *parameter space*. This line in the space of the parameters a and b corresponds to all possible straight lines going through the point (x, y) in the xy plane. If a number of points (x_i, y_i) are lying along a straight line in the xy plane, the transformed lines in parameter space will intersect at the point which defines the actual parameters of the line. All this is illustrated in Figure 9.

In practice, the above-mentioned procedure is often done with a computer. One then makes parameter space *discrete* and defines one counter for each *cell* in parameter space. Instead of drawing a line for each point (x, y), one increments the relevant counters. For a two-dimensional parameter space, the counters can be visualized as pillars. The search for the correct line parameters is then performed by looking for *peaks* in the three-dimensional histogram defined by the counter pillars. An example of a histogram in parameter space for our specific case is shown in Figure 10. The position of the peak defines thus the estimated values of the parameters a and b.

The optimal number and size of the bins in parameter space depends on the actual problem at hand and must be evaluated on basis of aspects like the number and density of data points, the required resolution, the noise level and the available computer power. For instance, if the bin size is chosen too big, one might end up with two nearby lines inside one bin and therefore miss one of them. On the other hand, if the bin size is chosen too small, the measurement error of the data points could cause a line to be spread throughout several bins and not create a peak at all.

Due to the computational load, a straightforward application of the Hough transform with parameter space dimension higher than two is not feasible [11]. Thus, a three-dimensional problem such as circle detection is not entirely straightforward with



(a) The line $y = 2x + 1$.

(b) The lines in parameter space.

Figure 9: The line $y = 2x + 1$ and six points along this line are shown in (a). As shown in (b), each of these points corresponds to a line in parameter space. The correct values of the parameters, $(a, b) = (2, 1)$, can be deduced from the position of the intersection point.

the Hough transform. However, if one requires that the circles pass through a certain point, the problem is reduced to a two-dimensional one and can be solved along exactly the same lines as the line detection problem.

The *standard Hough transform* as described above is a *divergent* transform, i.e. a point is mapped onto a curve in parameter space. It is possible to reformulate the Hough transform so that it becomes a *convergent* transform. This can, in the line detection case, be done by picking two data points instead of one. Two points uniquely define a set of line parameters, and the corresponding cell in parameter space is incremented. This is done for all possible combinations of data points, and this version of the transform is therefore called the *combinatorial Hough transform.* A search for peaks in parameter space is done in the usual way. If there are many data points, this procedure can be

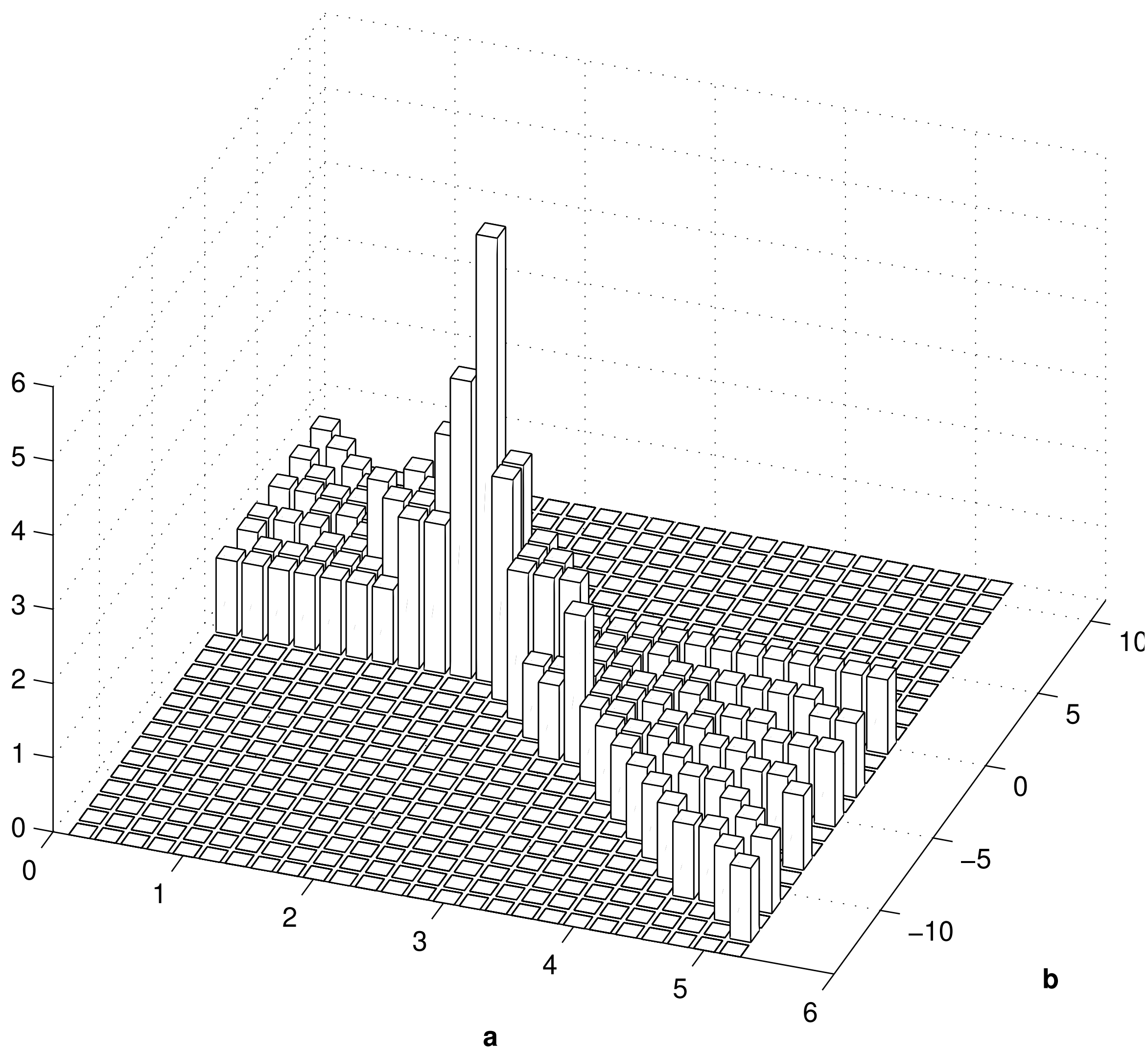


Figure 10: Histogram or accumulator array in parameter space for 6 points lying along the line $y = 2x + 1$.

slow. There exist several remedies in the literature, and a very popular approach is the so-called *randomized Hough transform* [12]. For the combinatorial case, the generic procedure is to randomly select only a subset of all the possible pairs of points and fill up the accumulator array with this subset.

Another way of lowering the computational load in the combinatorial case is to decrease the number of pairs of points by putting in some kind of prior information. When trying to detect circular particle tracks coming from the origin, a *local Hough transform* has been introduced [13]. It is based on picking only those pairs which have a distance between the points less than some maximum distance. In addition, since tracks from the origin are known to go radially outwards, the number of pairs can be further reduced.

In one of the papers of this thesis, the idea of using prior information to reduce the computational complexity has been pursued even further. This has led to a Hough transform procedure which avoids the accumulation of a histogram, and it is shown to be significantly faster than the standard Hough transform.

4.2 The Kalman filter

The Kalman filter [14] is a method for pattern recognition and track fitting which is very commonly used in modern high-energy physics experiments. Originally, however, the method was developed for radar tracking of spacecraft. It was introduced into high-energy physics by P. Billoir [15] and was called the "progressive" method of track fitting [16]. However, at that time the equivalence between the progressive method and the Kalman filter was not known. This was realized a few years later by R. Frühwirth [6]. It then became clear that the filter could be supplemented by a *smoothing* procedure, which enabled the optimal estimation of the track parameters *anywhere* along the track, not only at some reference surface.

The Kalman filter is a linear, recursive method of track fitting which is equivalent to a global, linear least-squares (GLS) method and is therefore the optimal, linear estimator of the track parameters. If the track model is truly linear and the errors involved are Gaussian, it is also *efficient*, i.e. no non-linear estimator can do better. The Kalman filter has some well-known, attractive features which makes it preferable over the GLS under certain circumstances:

- The recursive nature of the method makes it well-suited for combined track finding and track fitting.
- As mentioned above, the Kalman filter with its associated smoother gives optimal estimates anywhere along the track and not only at a reference surface (which is often chosen to be close to the vertex). This makes it easy to obtain optimal

predictions also in detectors lying on the outside or somewhere in between the tracking detectors.

- The ability to obtain smoothed predictions at any detector layer also enables the Kalman filter to efficiently resolve and remove *outliers*, i.e. measurements not belonging to the track to be fitted.
- In the presence of multiple scattering and many measurements, a large covariance matrix has to be inverted when applying the GLS. This is not the case for the Kalman filter.

In the framework of the Kalman filter, the change of the parameters of a track along its path is regarded as the dynamical evolution of a stochastic *state vector* $\boldsymbol{x}$. We will in the following assume that our detector is a collection of n shells or surfaces, and that both the measurements and the state vectors are defined at these surfaces only. A subscript k will denote quantities at layer number k.

The state vector at layer k is assumed to be a linear function of the state vector at layer $k-1$:

$$\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \boldsymbol{\omega}_k, \tag{1}$$

where the *process noise* $\boldsymbol{\omega}_k$ has mean value zero and a covariance matrix $\boldsymbol{Q}_k$. The *measurement vector* $\boldsymbol{m}_k$ is assumed to be a linear function of the state vector:

$$\boldsymbol{m}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{\epsilon}_k, \tag{2}$$

where the *measurement noise* $\boldsymbol{\epsilon}_k$ also has mean value zero, and its covariance matrix is $\boldsymbol{V}_k$. The process and measurement noise are assumed to be independent.

Given an estimate of the state vector in layer $k-1$, $\boldsymbol{x}_{k-1|k-1}$, the Kalman filter starts by extrapolating this state vector into the next layer:

$$\boldsymbol{x}_{k|k-1} = \boldsymbol{F}_k \boldsymbol{x}_{k-1|k-1}. \tag{3}$$

This is a *prediction* step of the Kalman filter. The covariance matrix of the state vector also needs to be propagated:

$$\boldsymbol{C}_{k|k-1} = \boldsymbol{F}_k \boldsymbol{C}_{k-1|k-1} \boldsymbol{F}_k^{\mathrm{T}} + \boldsymbol{Q}_k. \tag{4}$$

An updated estimate of the state vector is now obtained by including the information from the measurement $\boldsymbol{m}_k$, yielding the *filtered* estimate $\boldsymbol{x}_{k|k}$:

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k \left(\boldsymbol{m}_k - \boldsymbol{H}_k \boldsymbol{x}_{k|k-1} \right). \tag{5}$$

Here $\boldsymbol{K}_k$ is the *Kalman gain matrix* and is given by:

$$\boldsymbol{K}_k = \boldsymbol{C}_{k|k-1} \boldsymbol{H}_k^{\mathrm{T}} \left(\boldsymbol{V}_k + \boldsymbol{H}_k \boldsymbol{C}_{k|k-1} \boldsymbol{H}_k^{\mathrm{T}} \right)^{-1}. \tag{6}$$

The filtered covariance matrix is given by:

$$\boldsymbol{C}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\, \boldsymbol{C}_{k|k-1}, \tag{7}$$

where $\boldsymbol{I}$ denotes the identity matrix. It can be noted that the Kalman filter can be formulated in a totally equivalent manner with *weight matrices* or inverse covariance matrices instead of the covariance matrices themselves. When the dimension of the state vector is small, this approach is usually computationally less expensive.

The *smoothed state vector* in layer k given the information from all n layers is now

$$\boldsymbol{x}_{k|n} = \boldsymbol{x}_{k|k} + \boldsymbol{A}_k \left(\boldsymbol{x}_{k+1|n} - \boldsymbol{x}_{k+1|k} \right), \tag{8}$$

with the *smoother gain matrix*

$$\boldsymbol{A}_k = \boldsymbol{C}_{k|k} \boldsymbol{F}_k^{\mathrm{T}} \left(\boldsymbol{C}_{k+1|k} \right)^{-1}. \tag{9}$$

Finally, the covariance matrix of the smoothed state vector is

$$\boldsymbol{C}_{k|n} = \boldsymbol{C}_{k|k} + \boldsymbol{A}_k \left(\boldsymbol{C}_{k+1|n} - \boldsymbol{C}_{k+1|k} \right) \boldsymbol{A}_k^{\mathrm{T}}. \tag{10}$$

It can be noted that smoothing also can be achieved by combining the results from one filter running in the forward direction with those from another filter running in the backward direction.

It was soon realized that the Kalman filter, due to its recursive nature, could be used for combined track fitting and pattern recognition or track finding [17–19]. One possible procedure is what is called a *combinatorial Kalman filter* (CKF). The idea is to start up, usually from the outer layers of a tracking detector, with an initial estimate of the parameters of a track. The initial estimate can be obtained either by extrapolating estimates from other detectors or by combining a suitable number of measurements in the outer detector layers. This initial estimate is extrapolated inwards to the next layer, and measurements compatible with the prediction (i.e. those which survive a cut on the χ^2-statistic) are picked up. If there are several compatible measurements, all combinations of points which make up a valid track candidate must be stored. Each track candidate is then further extrapolated, and new compatible measurements are picked up. This procedure continues until one reaches the innermost layer. In the end, one has to select one or more of the surviving track candidates. If one is reasonably certain that there is only one real track in the traversed region, one can select the candidate with the least value of the χ^2 as the final candidate. If there can possibly be several tracks, one will of course have to keep several candidates. In this case it is possible to keep all candidates with a χ^2 below a certain cut, given that none of the candidates have too many measurements in common. A possible disadvantage with the CKF is the fact that the number of track candidates to be explored can become very large when the density of measurements is getting high. The computational cost can thus become very large. This problem will be even more pronounced at the LHC experiments than at the experiments of today.

As mentioned above, the Kalman filter can also be used as a method of detecting outliers [6]. For this purpose the filter is most effective when the number of outliers is small compared to the total number of measurements constituting the track. The strategy is to calculate the smoothed predictions in each layer with a measurement, using information from all the other layers with measurements. If the residual in a specific layer is too large, the measurement is identified as an outlier and removed from the candidate. However, if the number of outliers is too large, the smoothed predictions can be significantly biased during the first run of the filter and smoother. The quality of the outlier detection can thus decrease and the procedure might even break down completely.

Another approach to the problem of outlier detection is to define a *Gaussian mixture model* of the measurement error, yielding a *robustified Kalman filter*. The effect of this robustified Kalman filter is to give a smaller weight to the outlying observations in an automated manner. This was proposed already more than ten years ago [20,21], but the first journal paper on the subject appeared some years later [22]. The basic idea in the above-mentioned references is to assume that the measurement error is a sum of two Gaussian components. The first component expresses the belief that the measurement is a true measurement, and the covariance matrix of the first component is therefore equal to the covariance matrix of the true measurement. The second component expresses the belief that the measurement is an outlier, and because of this the covariance matrix of this component is taken to be much larger. If one assumes that the distribution of the predicted state vector is a single Gaussian, the *posterior* distribution of the state vector can by Bayes' theorem be shown to be a mixture of two components, the mean value of each component being given by a Kalman filter. In order to minimize the computational cost of the filter, the posterior is collapsed into a single Gaussian. The filtered state vector thus becomes the mean value of the posterior distribution, which turns out to be given as a linear combination of the state vectors of each of the components. The smoother is not affected by the robustification of the filter and is the same as with a standard Kalman filter.

The robust filter has more recently been generalized to the so-called *Gaussian-sum filter* (GSF) or the parallel filter [23]. With the GSF, there is no complete collapse of the posterior distribution of the state vector. Therefore, the filter becomes a collection of several Kalman filters running in parallel. It is now possible to have a mixture model description of both the measurement error and the process noise, something which means that long-tailed process noise as multiple Coulomb scattering can be more adequately described than with an ordinary Kalman filter. Of course, it can not be tolerated to have an exponentially increasing number of components as one traverses the detector. It is therefore necessary to invoke some kind of strategy in order to keep the number of components reasonably small. A very simple way is to keep the N largest components after each filter step. More sophisticated methods can be based for instance on clustering components with a small Kullback-Leibler distance.

Another difference between the GSF and the robustified Kalman filter is the development of a *Bayesian smoother*, which yields smoothed estimates using all the available information from the filter step. This has been shown to improve the quality of the estimates compared to a standard, weighted-mean approach [23].

4.3 Hopfield neural networks and elastic tracking methods

Hopfield networks

Neural networks of the type *Hopfield networks* [24] have been applied to track finding problems in high-energy physics. The basic idea is to set up an *energy function* whose minimum corresponds to the solution of the pattern recognition problem. This was proposed independently by C. Peterson [25] and B. Denby [26]. The standard, dynamical evolution of the network will always decrease the energy and rapidly converge to a solution [27]. However, the energy landscape usually has many local minima, so one will in general end up in the closest local minimum. In order to cure this behaviour, a so-called *mean-field annealing* technique has been developed [28]. This technique yields a set of deterministic equations whose solutions often are very close to the global optimum, with relatively small computational cost.

Hopfield networks have been tried out on data from the TPC detector of the ALEPH experiment at LEP [29] and found to give results comparable to those from the official pattern recognition program. More recently, one has also put efforts into exploring alternatives to mean-field annealing in searching the minimum of the energy function [30]. However, the conclusion is that mean-field annealing seems to be the winner when evaluating the methods with respect to a compromise between efficiency and computational speed. It also seems to be very difficult to beat a standard track finder such as the Kalman filter with respect to track finding efficiency. This can probably be ascribed to the fact that there is no explicit reference to the track model in a Hopfield net. (In a Hopfield net one favours situations where nearby links between measurements point more or less in the same direction.) In addition, the execution time of the Hopfield net scales nonlinearly with the event complexity, so it is supposed to be too slow for general use at the LHC.

As mentioned above, one of the major weaknesses of the Hopfield network is the absence of the track model in the formalism. In addition, the Hopfield net is a pure track finder, i.e. one has to supplement the algorithm with a separate track fitting procedure in the end. Both these shortcomings can be cured in an approach based on *deformable templates*, called *elastic tracking.* As with the Hopfield net, elastic tracking was proposed independently by two different groups. One flavour is called the *Elastic Tracking* algorithm (ET), developed by Gyulassy and Harlander [31]. Recently, a program package based on this algorithm has been developed for the STAR experiment at RHIC [32]. The other method of elastic tracking is called the *Elastic Arms* algorithm (EA), developed by Ohlsson, Peterson and Yuille [13]. It has become a very popular algorithm and has for instance been applied to vertex fitting in the CMS inner tracker [33].

Elastic Tracking

The ET is based on defining a charge distribution of a *template track*, and an *interaction energy* $R_{\mathrm{V}}(\boldsymbol{p}_{\mathrm{T}})$ between this template track and the charge distribution of the measurements in the event is then set up:

$$R_{\mathrm{V}}\left(\boldsymbol{p}_{\mathrm{T}}\right)=\int d\boldsymbol{x}d\boldsymbol{x}^{'}\rho\left(\boldsymbol{x}\right)V(\boldsymbol{x}-\boldsymbol{x}^{'})\rho_{\mathrm{T}}(\boldsymbol{x}^{'}). \tag{11}$$

Here $\boldsymbol{p}_{\mathrm{T}}$ defines the parameters of the template track, $\rho(\boldsymbol{x})$ is the charge density of the measurements in the event, $\rho_{\mathrm{T}}(\boldsymbol{x}^{'})$ is the charge density of the template track, and $V(\boldsymbol{x}-\boldsymbol{x}^{'})$ is the interaction *potential* between the template track and the measurements. Note that the charge densities of the measurements and the template track have opposite signs, so that there is an attractive force between them. In the original paper by Gyulassy and Harlander the potential is chosen to have a Lorentzian form:

$$V\left(\boldsymbol{x}\right)=\frac{w^{2}}{w^{2}+\boldsymbol{x}^{2}}. \tag{12}$$

The effect of a potential of this type is to attract the template track to points nearby it and make the template track insensitive to points being far away, i.e. it exhibits an implicit suppression of the outliers. Usually, the measurements and the template track are modelled as delta functions, and in this case the interaction energy simply becomes

$$R_{\mathrm{V}}\left(\boldsymbol{p}_{\mathrm{T}}\right)=-\sum_{i}\frac{w^{2}}{w^{2}+\boldsymbol{x}_{i}^{2}}, \tag{13}$$

where $\boldsymbol{x}_{i}^{2}$ is the squared distance from measurement i to the template track.

The minimum of the interaction energy with respect to the parameters of the template track defines the solution of the track reconstruction problem, but such a solution can in general not be found explicitly. For the ET one therefore applies a standard *gradient descent* procedure [34] in the space spanned by the parameters of the track. With gradient descent one starts out with an initial guess of the track parameters. The parameters are then changed with a value $\Delta\boldsymbol{p}_{\mathrm{T}}$ in a direction opposite to the gradient of the energy,

$$\Delta\boldsymbol{p}_{\mathrm{T}} = -\eta \frac{\partial R_{\mathrm{V}}\left(\boldsymbol{p}_{\mathrm{T}}\right)}{\partial \boldsymbol{p}_{\mathrm{T}}}, \tag{14}$$

where η, the "learning rate", is a constant determining how far in the direction of the negative gradient one goes. At the updated value of the parameter vector, the gradient is recalculated, and the parameter vector is again updated according to Equation (14). This is repeated until some convergence criterion is fulfilled. It can be noted that the constant w is supposed to be "time-dependent" or, in other words, a function of the current number of iterations. It is taken to be reasonably large at the beginning in order to not cut away good points early due to a bad initialization, i.e. one wants to avoid ending up in a *local minimum* of the interaction energy. During the iterations it should decrease, and, according to Gyulassy and Harlander, in the end be assigned a value comparable to the magnitude of the measurement error.

An event with many tracks is reconstructed by applying a sequential procedure. Firstly, the procedure described above is run until one track is found. Next, a second template track is initialized and run until convergence. The second track will during the iterations be attracted to the measurements and be repelled by the first template, so it should in principle not converge to the same solution. New template tracks are then added in the same manner as long as the algorithm converges properly.

Elastic Arms

The basic idea behind the EA algorithm of Ohlsson, Peterson and Yuille is to define an *energy function* of the following form:

$$E\left(\{S_{ia}\};\boldsymbol{p}\right) = \sum_{ia} S_{ia}M_{ia} + \lambda \cdot \sum_{i}\left(\sum_{a} S_{ia} - 1\right)^{2}. \tag{15}$$

Here $\boldsymbol{p} = \{\boldsymbol{p}_a\}$ is the set of parameter vectors $\boldsymbol{p}_a$ of the tracks, where $a = 1, \ldots, M$. Thus, it is assumed that there are M tracks in the event. The quantity M_{ia} denotes the squared distance from measurement i ($i = 1, \ldots, N$) to track or *arm* a, and $\{S_{ia}\}$ denotes the configuration of the *assignment variables* S_{ia}. The assignment variables S_{ia} are discrete variables which can take the values 0 and 1 only. A value 1 means that measurement i is assigned to track a, while it is 0 otherwise. The parameter λ governs the strength of a *penalty* or *regularizing term*. This term favours situations where a point i is assigned to only one track a.

As with the ET method, the solution of the track reconstruction problem is the minimum of the energy function with respect to the track parameter vectors and the assignment variables. However, it is computationally intractable to go through all possible combinations of the assignment variables, find and store the minimum value of the energy function for each of these combinations, and pick the solution with the least energy in the end. A way out of this problem is to introduce a temperature parameter T and require the configurations of the system to obey the Boltzmann distribution of statistical mechanics. The probabilities P of the different configurations will then depend on the energy as:

$$P\left(\{S_{ia}\}\,;\boldsymbol{p}\right) = \frac{1}{Z} e^{-\beta E(\{S_{ia}\};\boldsymbol{p})}, \tag{16}$$

where $\beta = 1/T$ is the inverse temperature and Z is a normalization constant or the *partition function* of the system. It is now possible to get rid of the assignment variables by focusing on the *marginal probability density function* $P_M(\boldsymbol{p})$. This is given as the sum of the above-defined probabilities over all different configurations of the assignment variables:

$$P_M(\boldsymbol{p}) = \sum_{\{S_{ia}\}} P\left(\{S_{ia}\}\,;\boldsymbol{p}\right). \tag{17}$$

It is possible to calculate this sum explicitly, and the result is

$$P_M(\boldsymbol{p}) = \frac{1}{Z} e^{-\beta E_{\text{eff}}(\boldsymbol{p})}, \tag{18}$$

where the *effective energy* E_{eff} is given by

$$E_{\text{eff}}(\boldsymbol{p}) = -\frac{1}{\beta} \sum_i \log\left(e^{-\beta\lambda} + \sum_a e^{-\beta M_{ia}}\right). \tag{19}$$

The strategy is now to start at a high temperature with an initial guess of the track parameters (obtained by a pattern recognition algorithm such as the Hough transform) and find the minimum of the effective energy with respect to the track parameters. The resulting track parameters are then used as initial parameters at a lower temperature, and the minimum of the effective energy is again sought. This procedure is called *annealing*, and it is repeated in principle until the zero temperature limit is reached. The track parameters found after the last iteration are taken to be the final estimate of the track parameters.

In the paper by Ohlsson et al. [13], gradient descent is used as the method of finding the minimum of the effective energy. This yields

$$\Delta\boldsymbol{p}_a = -\eta \frac{\partial E_{\text{eff}}}{\partial \boldsymbol{p}_a} = -\eta \sum_i V_{ia} \frac{\partial M_{ia}}{\partial \boldsymbol{p}_a}, \tag{20}$$

where V_{ia} is the *assignment probability* of measurement i to track a,

$$V_{ia} = \frac{e^{-\beta M_{ia}}}{e^{-\beta\lambda} + \sum_b e^{-\beta M_{ib}}}. \tag{21}$$

As with the ET, η is a free parameter deciding how far in the direction of the negative gradient one goes during the update of the parameter vector.

Gradient descent is but one of many ways of determining the search direction. Ohlsson [35] proposed to apply *Newton's method* [34], and this was tested out in a realistic setting by Lindström [36]. Newton's method is based on approximating the function to be minimized (in this case the effective energy) by a second order polynomial and searching for the value of the argument giving a zero gradient. This gives for the update of the parameter vectors

$$\Delta \boldsymbol{p}_a = -\eta \cdot \boldsymbol{H}^{-1} \frac{\partial E_{\text{eff}}}{\partial \boldsymbol{p}_a}, \tag{22}$$

where the parameter η usually is chosen to be 1, and $\boldsymbol{H}$ is the *Hessian matrix* or the matrix consisting of the second derivatives of the effective energy with respect to the track parameters.

The elastic arms algorithm has also been generalized to cope with ambiguous measurements. This was done by R. Blankenbecler [37]. In addition to the S_{ia}'s, one now introduces binary decision units s_{ia}^{+} and s_{ia}^{-} which can take values 0 and 1 and always sum up to one. The effect of these is to decide which one of the two ambiguous measurements that belong to the track. Introducing the new decision units leads to a slightly modified energy function:

$$E\left(\left\{S_{ia}, s_{ia}^{\pm}\right\}; \boldsymbol{p}\right) = \sum_{ia} S_{ia}\left(s_{ia}^{+} M_{ia}^{+} + s_{ia}^{-} M_{ia}^{-}\right) + \lambda \cdot \sum_i \left(\sum_a S_{ia} - 1\right)^2, \tag{23}$$

and the effective energy becomes in this case

$$E_{\text{eff}}(\boldsymbol{p}) = -\frac{1}{\beta} \sum_i \log\left(2e^{-\beta\lambda} + \sum_a \left(e^{-\beta M_{ia}^{+}} + e^{-\beta M_{ia}^{-}}\right)\right). \tag{24}$$

Otherwise, the algorithm is the same as in the standard case.

Introduction to the papers

We will here briefly introduce the six research papers that constitute the main part of this thesis:

Paper 1:

B. Lillekjendlie and A. Strandlie,

Particle Tracking in the ATLAS Inner Detector.

Proceedings of the International Conference on Computing in High Energy Physics (CHEP'98), Chicago, USA, September 1998.

Paper 2:

R. Frühwirth and A. Strandlie,

Track finding and fitting with the Gaussian-sum filter.

Proceedings of the International Conference on Computing in High Energy Physics (CHEP'98), Chicago, USA, September 1998.

Paper 3:

R. Frühwirth and A. Strandlie,

Track fitting with ambiguities and noise: a study of elastic tracking and non-linear filters.

Computer Physics Communications 120 (1999), 197-214.

Paper 4:

A. Strandlie, J. Wroldsen, R. Frühwirth and B. Lillekjendlie,

Particle tracks fitted on the Riemann sphere.

Computer Physics Communications 131 (2000), 95-108.

Paper 5:

A. Strandlie and J. Zerubia,

Particle tracking with iterated Kalman filters and smoothers: the PMHT algorithm.

Computer Physics Communications 123 (1999), 77-86.

Paper 6:

A. Strandlie and R. Frühwirth,

Adaptive Multitrack Fitting.

To appear in: Computer Physics Communications.

Summary and discussion of the papers

Paper 1: Particle Tracking in the ATLAS Inner Detector

This paper presents a pattern recognition algorithm specifically designed for finding high-energy tracks going entirely through the barrel part of the ATLAS Inner Detector TRT. The track candidates given by this algorithm have been used to initialize some of the adaptive track fitting methods studied in Papers 2 and 3. The algorithm is similar to a Hough transform procedure, but nevertheless different and constitutes therefore an alternative to a plain Hough transform. The main difference from a combinatorial Hough transform is the introduction of a track candidate *verification* step, instead of filling up an accumulator array in parameter space and looking for peaks in this space in the end.

We introduce a parameterization which handles circles and lines in exactly the same manner. This parameterization is computationally efficient and is based on mapping the measurements onto the *Riemann sphere*. (The mapping onto the Riemann sphere is thoroughly described in Paper 4.) To a larger extent than in the standard Hough transform, prior knowledge of the structure of the data and of the characteristics of the detector is put into the construction of the algorithm.

An analysis of the computational complexity of the algorithm shows that it is faster than a standard Hough transform, particularly at low occupancies. The track finding efficiency obtained is also better than what has been reported earlier by Lindström [36]. He used a combinatorial or convergent Hough transform and applied it to data very similar to ours. A more relevant competitor would have been the Hough transform part of the XKalman reconstruction program [10]. XKalman uses a Hough transform in the TRT for initializing a track search in the silicon detectors. However, the initial Hough transform procedure in XKalman is an integrated part of the whole track reconstruction program, and it is therefore non-trivial to extract the kind of results relevant for a sensible comparison.

A major weakness with our approach is the fact that it only handles tracks going entirely through the barrel part of the TRT. A Hough transform does not in principle exhibit this inherent limitation and is able to find also tracks leaving the barrel, i.e. entering the end-cap TRT, before reaching the outer radius of the TRT. One could then envisage a two-step procedure whose first step is to run our algorithm. Secondly, some Hough transform procedure could be applied. After the first step most of the high-energy tracks going entirely through the barrel would thus be removed from the event, leaving an event better suited for detection of tracks going only partially through. It would be interesting to implement such an approach on realistic data and compare it to the results from a plain Hough transform. A possible disadvantage of such a two-step procedure is of course the fact that the overall computational cost will increase.

Paper 2: Track finding and fitting with the Gaussian-sum filter

The main, new idea presented in this paper is the introduction of the Gaussian-sum filter (GSF) as a method of solving the left-right ambiguity of measurements in track candidates from a drift tube detector, concurrently with estimating the track parameters. In the earlier papers [22, 23], a mixture model of the measurement error was constructed mainly for two purposes:

- *Robustification* by hypothesizing that a measurement either is a valid track hit or an outlier. This is achieved by modelling the measurement as a two-component Gaussian mixture with equal means but different variances.
- If the measurement error is non-Gaussian simply due to the characteristics of the detector elements, a mixture model can yield a better representation of the actual shape of the error distribution than a purely Gaussian model. This is achieved again by modelling the measurement as a mixture consisting of components with equal means but different variances.

However, in order to construct a filter being able to resolve ambiguities, a mixture model of the measurement error has to consist of components with different means but equal variances. In other terms, one builds the hypothesis that the true measurement is one of the two possible options given by the read-out from the drift tube directly into the distribution of the measurement error. Due to strict limitations on the length of the manuscript, the derivation of the structure of the parallel filter resulting from this new principle is not shown in the paper. However, a derivation is included in the GSF section of Paper 3.

A classical method of solving a left-right ambiguity problem is an iterated Kalman filter. After convergence of the iterative procedure, the measurement in a layer being closest to the track is taken to be the correct one. The approach of the GSF is conceptually different from such a standard approach, because both ambiguous measurements are used at the same time in the update of the filter. Each of the two measurements are given weights according to their posterior probabilities. Thus, if one of them is close to the true track and the other is far away, the latter is strongly downweighted. However, if both are close to the track, they will both contribute significantly.

Another novelty of this paper is the application of a *robust prefilter* based on the Least Quantile of Squares (LQS) principle. It was shown that a standard least-squares initialization procedure could be biased due to outliers present in the track candidates. The robust prefilter was able to locate the correct track to a much higher degree of accuracy, something which turned out vital in this case.

Even though the GSF was shown to be an efficient tool for solving the left-right ambiguities of drift tube measurements together with estimating the track parameters, some deficiencies also became evident:

- A very good initialization of the track parameters was needed for the starting phase of the filter. This is not always easy to achieve, and a robust prefiltering procedure is also computationally quite heavy.
- Labeling those of the measurements furthest away from the initial track as noise candidates and thereby modelling them as a Gaussian mixture was necessary to make the filter robust enough. It is not obvious how to determine the mixture weights in such a case, and the noise hypothesis results in a further increase in the computational cost.

A desire to cure some of these shortcomings led to the development of a new and more robust approach: the Deterministic Annealing Filter (DAF). This algorithm was presented in Paper 3.

Paper 3: Track fitting with ambiguities and noise: a study of elastic tracking and non-linear filters

This paper contains a rather extensive study of two main classes of adaptive tracking: *non-linear filters* and *elastic tracking algorithms*. Four different algorithms have been studied and compared: the *Gaussian-sum filter (GSF)*, the *Elastic Arms algorithm (EA)*, the *Elastic Tracking algorithm (ET)* and a novel algorithm: the *Deterministic Annealing Filter (DAF)*.

Filtering methods and elastic tracking algorithms have traditionally been regarded as two distinct approaches. Filtering methods have since the paper by R. Frühwirth [6] become mainstream algorithms for track reconstruction in high-energy physics experiments. The elastic algorithms were partially developed as refinements of neural network tracking algorithms, which received significant attention about ten years ago. However, even though this class of methods gained appeal due to their theoretically elegant formulation, they have been regarded as too demanding with respect to computational power to be useful in practice.

It can be claimed that the invention of the DAF has builded a bridge between the "filter" and the "elastic" realms:

- The DAF can on one hand be regarded as a simplified version of the GSF. The main distinctions are:
  1. After each filter step, all components are collapsed into one.
  2. The posterior weights do not have to sum up to unity.
  3. It is iterative, with the possibility of having "inflated" variances during the first iterations.

  The DAF is thus a faster and more robust approach than the GSF. This has also been confirmed by the results of the simulation experiments in the paper.
- On the other hand, by applying the *EM algorithm* [38], a single-track version of the EA has in this paper been shown to be equivalent to the DAF. However, the DAF is able to include process noise such as multiple Coulomb scattering in a natural way, and it also supplies the covariance matrix of the estimated parameters. These properties makes the DAF better suited for concrete experimental use than the EA.

The simulation experiments have shown that the DAF has excellent abilities in discerning between correct track points and noise points (pattern recognition), concurrently with estimating the track parameters (track fitting). It should therefore be a viable alternative to conventional methods solving the same tasks, such as a combinatorial Kalman filter and an iterative Kalman filter. In fact, the soft measurement-to-track assignments of the DAF have been shown to yield even better results than the hard assignments of the combinatorial and the iterative Kalman filter.

The study of the EA algorithm has also yielded some interesting results:

- We have derived a modified energy function of the single-track version of the EA algorithm. The resulting algorithm introduces a competition between all hits in the same layer in a track candidate. This can be looked upon as a generalization of the idea of Blankenbecler [37], who, in the case of a detector with ambiguous measurements, modified the original EA algorithm to include a competition between a hit and its corresponding mirror hit.
- It has been shown that the minimization of the effective energy has to be done by advanced optimization methods in order to yield optimal results. We introduce the application of a quasi-Newton method for determining the search direction and an iterative line search along the chosen search direction. Only by following such a procedure until convergence can the performance of the DAF be achieved. Naive gradient descent performs significantly worse than the above-mentioned approach.
- It has been established that the $T \to 0$ limit of the EA algorithm is not the optimal one. Rather one should choose the "natural" temperature, which is related to the measurement error, as the final temperature.

For the ET algorithm, the highlights are:

- A new, Gaussian potential has been introduced as an alternative to the Lorentzian potential proposed by Gyulassy and Harlander [31]. The Gaussian potential has been shown to yield slightly better results than the Lorentzian.
- As for the EA algorithm, advanced and time-consuming minimization methods have to be applied in order to yield optimal results. This is particularly pertinent to the Lorentzian potential.
- Gyulassy and Harlander proposed an "annealing" on the width w of the potential [31], and they proposed that the final width of the potential should be approximately equal to the measurement error of the detector elements. We have shown that this choice of the final width of the potential is far too small to be optimal.

To our knowledge this paper also introduces the first thorough comparison between the EA and the ET algorithms. On our simulated data, the EA algorithm has turned out to be the best. This is due to the non-quadratic energy function of the ET, which makes it different from a least-squares estimator. Also, there is no inherent competition between several hits in the same detector layer. Moreover, the energy function of the ET has turned out to be numerically even more difficult to minimize than the energy function of the EA. However, the energy function of the ET is formulated as an interaction energy between two charge densities, something which enables the method to work also on non-clustered hits. If one of some reason wants to work on the raw hits instead of the clustered hits, one is obliged to choose the ET algorithm.

Paper 4: Particle tracks fitted on the Riemann sphere

The task of fitting circular arcs to a set of measurements is a very important one in many high-energy physics experiments. This is a *non-linear* problem, i.e. one will in general have to rely on an iterative and computationally complex approach. The need of developing fast circle fitting methods has led to a number of algorithms which solve the problem approximately [39–41]. The most precise of these seem to be the algorithms by Hansroul et al. [40] and Karimäki [41]. Hansroul et al. turn the circle fitting problem into a linear one by applying the *conformal mapping* method. By this transformation, circles going through the origin are mapped onto straight lines, while circles with a finite but small impact parameter are mapped onto parabolas. Karimäki's approach still leads to a non-linear problem, but he is able to find a unique solution to it. With respect to estimation accuracy, the two methods are comparable.

The fundamental idea in Paper 4 is to present a novel type of direct circle fit by mapping two-dimensional measurements onto the Riemann sphere and fitting a plane to the transformed, three-dimensional measurements. Fitting a plane to a set of measurements is basically a task of finding the eigenvectors and eigenvalues of a

3×3 matrix, which can be solved non-iteratively. It turns out that the accuracy of the resulting fit is very close to that which can be achieved by an optimal, non-linear approach, and it is very much faster. It is also comparable in speed to the above-mentioned, direct methods and significantly more precise.

One of the main results of Paper 3 is the realization that the Elastic Arms algorithm through an EM approach can be formulated as an iteratively reweighted least-squares procedure. The original EA solves the minimization step by invoking a non-linear minimization procedure, while the DAF does this by a Kalman filter and smoother. However, one can of course think of other ways of doing this, and in Paper 4 we propose to use the Riemann fit. The result is a very fast elastic tracking procedure (in fact, several tens of times faster than the original EA) : the *Elastic Planes (EP)* algorithm. We also introduce a global least-squares method (GLS) with annealing. In the absence of process noise this is equivalent to the DAF, but it is significantly faster.

A major shortcoming of the Riemann fit is the lack of a covariance matrix of the estimated track parameters. This needs to be derived in order to make the method a fully-fledged track fitting approach. Nevertheless, it is not obvious how to obtain such a covariance matrix, so this remains an open question.

With the invention of the DAF, the EP algorithm and the GLS with annealing, there are now a number of new elastic tracking algorithms on the scene. However, the original EA algorithm is formulated such that it can fit several tracks simultaneously to the measurements. A very natural question to ask is therefore: how do we formulate multi-track versions of the new elastic algorithms? This question has at least partially been answered in Papers 5 and 6.

Paper 5: Particle tracking with iterated Kalman filters and smoothers: the PMHT algorithm

Tracking is a subject which is not only limited to one part of the data analysis chain of a high-energy physics experiment. In fact, one encounters similar problems in a large range of applications. For instance, military applications as *sonar tracking* of missiles and submarines have been given significant attention.

We present in this paper a multi-track algorithm invented for sonar tracking, the *Probabilistic Multi-Hypothesis Tracking algorithm (PMHT)* by Streit and Luginbuhl [42], and make the necessary modifications in order to make it work on particle tracking in high-energy physics detectors. We also generalize the algorithm to include a deterministic annealing scheme. Since the PMHT is an example of an EM algorithm, it should come as no surprise that a deterministic annealing PMHT algorithm is derived by invoking a recently developed deterministic annealing EM algorithm [43]. In addition, a modification of the algorithm is made such that the assignment weights exhibit a competition between the different hits in a layer.

With these generalizations and modifications the PMHT has been shown in the single-track case and under the assumption of Gaussian noise to be equivalent to the DAF. The PMHT can thus be regarded as a multi-track generalization of the DAF. However, the formulation of the algorithm is not restricted to the Gaussian case only. In fact, any functional form of the probability densities of process noise and measurement noise can be handled. In this general case the maximization of the PMHT likelihood structure becomes much more complicated than in the Gaussian case. (With Gaussian noise the maximization is performed by running a set of Kalman filters and smoothers in parallel.) Other types of optimization methods then have to be considered, but they will in general require more computer power than the straightforward application of Kalman filters.

The simulation experiments of this paper have been performed on single tracks, and the main focus has been on showing the improvement of the performance due to the generalizations and modifications introduced. In particular, the annealing has turned out to be vital in order to obtain results with a satisfactory accuracy.

Even though the PMHT defines a very promising scheme for global track reconstruction, not much attention has been paid in this paper to the specific structure of the weights in the multi-track case. Moreover, no experiments considering the aspect of a measurement possibly belonging to several tracks have been performed. These issues have been discussed in Paper 6.

Paper 6: Adaptive Multitrack Fitting

Paper 6 contains a study of different multi-track algorithms in the context of reconstruction of a simulated sample of track pairs. The methods considered are:

1. A sequential application of the standard DAF.
2. The original Elastic Arms (EA) algorithm.
3. A novel method of this study introducing a global decision rule. We will denote it *Global Competition (GC).*
4. The EA algorithm generalized to handle ambiguous measurements.

We have tested out the algorithms with two different initialization methods. The first is to initialize the tracks with the true parameter values. Due to the highly fluctuating density of measurements in the track pair sample, it turned out non-trivial to apply standard track finding techniques such as a Hough transform or a robust initialization based on the LQS procedure described in Paper 2. Therefore, we have in addition developed a relatively simple procedure which works rather well: the "cog+/-" method described in the paper.

Obviously, finding the global minimum of the energy function or a very good approximation to the global minimum is a much harder problem in the multi-track case than in the single-track case. In the single-track case, there is little sensitivity to the starting temperature and to the annealing schedule. If the starting temperature is not too low and the annealing is not too slow, it works well. In the multi-track case, however, the choice of these has to be a compromise between conflicting requirements (see discussion in the paper). The convergence at the final temperature is also much more slow in the multi-track case: we had to spend four to five times more iterations at the final temperature in this paper than in the single-track simulations in the earlier papers.

The results of the experiments in the paper showed that all the different multi-track algorithms always work better than a sequential application of the standard DAF. The GC scheme also turned out to be the best, in particular at high noise levels. At very low noise levels the standard EA (without mirror hits) and the EA generalized to handle ambiguous measurements (with mirror hits) are the algorithms which seem to model the underlying structure of the data most realistically and thus give the best results in this limit.

However, it should be stressed that there are still a number of open issues. All multi-track methods applied in this study are to a certain extent depending on a good initialization of the track parameters. This is in general non-trivial to obtain in the case of close or overlapping tracks. Initializing the algorithms with the correct number of tracks is also a non-trivial task, but nevertheless vital in order to make the multi-track algorithms work properly. In addition, even though the GC algorithm has turned out to be the best in our studies, it is by no means sure that this contains the optimal choice of the assignment weights. The answers to these questions are very likely dependent on the specific problem to which a multi-track fit is applied. A detailed study in the context of a concrete experiment is therefore needed.

Outlook and conclusions

During the work of this thesis, the main choice has been to investigate the basic, statistical properties of the algorithms in an environment much simpler than the one one will encounter in the real world of an experiment. The next, natural step is then of course to implement the new algorithms and test them out in a realistic setting. For the DAF, some progress in this direction has been made already. It has been implemented in the official track reconstruction framework in CMS, *ORCA*, and is at the present time being evaluated. There is also a desire to implement the multi-track filter and the Riemann fit in the same framework [44].

Some of the new algorithms should also be well suited for implementation in the track reconstruction software in the ATLAS experiment. For instance, in XKalman [10], an iterative Kalman filter is used for fine-tuning tracks found by a combinatorial Kalman filter. It would be very interesting to implement the DAF for the same task and compare the two approaches. The DAF could also be an alternative to the current strategy in XKalman for the final track finding and fitting in the TRT. At the present stage, a straight line fit to TRT measurements after a conformal transformation is used to resolve the left-right ambiguity, and a straightforward Kalman filter is applied in the end. With the DAF, multiple scattering effects can be taken into account also in the assignment phase, and the annealed filter should thus in principle be more precise.

Track fitting with the Kalman filter is always based on the assumption that the different types of errors involved are distributed according to a Gaussian. With the Kalman filter, multiple scattering is taken into account by adding terms in the covariance matrix during the extrapolation phase. The same procedure is followed when including energy loss effects in the fitting of electron tracks [45]. However, neither multiple scattering nor energy loss follow a Gaussian distribution. Multiple scattering has a Gaussian core, but the tails are much longer than an ordinary Gaussian. The distribution of energy loss is highly non-Gaussian, but it has been shown that energy loss efficiently can be modelled by a Gaussian mixture and that the Gaussian-sum filter or the parallel filter gives an improved performance compared to a standard Kalman filter, at least under relatively simple conditions [46]. Recently, a very precise, two-component Gaussian mixture approximation of the multiple scattering distribution has been derived [47]. In the light of these innovations, there is an evident need of making a thorough investigation, under realistic conditions, of the application of the GSF for track fitting purposes as an alternative to the plain Kalman filter. It is even possible to propose a mixture model generalization of the DAF, which could be called the *deterministic annealing GSF*, so that a mixture model description of multiple scattering or even energy loss can be taken into account concurrently with assigning measurements to tracks.

A disadvantage of the GSF is the fact that the number of components increases exponentially as one traverses the detector. Therefore, some kind of selection procedure has to be invoked in order to limit this number. In the general case it might be non-trivial to decide how many components to keep to obtain results accurate enough, and this number might in addition be quite large. Under such circumstances an alternative would be to follow an approach similar to the general formulation of the PMHT algorithm presented in Paper 5. Recalling that the PMHT algorithm is well-defined for any potential shape of the distributions involved, one could instead of applying the GSF formulate the track reconstruction problem as the problem of finding the maximum of the PMHT likelihood structure. When dealing with non-Gaussian distributions, this has to be done by numerical optimization methods, but it can in principle be done. It would be interesting to compare the performance of this general PMHT with the GSF for some kind of specific application.

Vertex fitting is a data analysis problem in high-energy physics which since long has been solved by least-squares methods. It has been realized that the Kalman filter and a robust extension of it, the *M-estimator* [48], also can be applied to vertex fitting [20]. These methods have successfully been tested out on real data from the LEP experiment at CERN [49]. At the LHC, however, the track multiplicity will be significantly larger than at LEP. Due to the large number of "noise" vertices, existing methods as the Kalman filter and the M-estimator might be biased. Moreover, the standard methods do not handle the situation of fitting several vertices simultaneously to the tracks delivered by the track fitter. It would therefore be interesting to investigate the behaviour of the annealed filters as vertex fitters in an LHC environment, both the DAF for the fitting of single vertices as well as the multi-track filter for the fitting of multiple vertices.

As mentioned at the end of Paper 4, it would be very useful to have a similar, direct way of fitting points also in three spatial dimensions, i.e. to be able to fit a full helix to the measurements. A straightforward approach would be to first fit the coordinates in the bending plane as described in the paper. Then, armed with estimates of the parameters relevant for the movement in the bending plane, the projected path length as a function of the radius can be calculated. Since the movement in the direction parallel to the beam is a linear function of this projected path length, the coordinates of this direction can also be fitted directly. This also opens up the possibility of very fast elastic tracking in $3D$, which could be denoted an “elastic hyperplanes” algorithm. It would be interesting to compare both the execution speed and the accuracy of this approach to an approach based on a standard Kalman filter or a linearized, global least-squares fit.

A few words can also be said about possible improvements of the pattern recognition algorithm presented in Paper 1. The natural way to proceed is to include the drift time information and run the algorithm on both the hits and the corresponding mirror hits. The width of the band around the hypothesized track candidates could then be narrowed significantly, so one would at least expect a larger purity of the track candidates. Moreover, the width of the band could even be defined to depend on the expected amount of deflection due to multiple scattering, something which would possibly further improve the performance of the algorithm. The computational complexity will increase as a consequence of these modifications, which of course is a disadvantage. The actual strategy followed in a specific implementation will thus depend on the speed and accuracy requirements.

In conclusion, we have in this thesis presented a number of new, adaptive track reconstruction algorithms, as well as modifying and obtaining new knowledge of some of the existing ones. The main goal has been to supply the particle physics community with novel track reconstruction algorithms, capable of meeting the requirements of the experiments of the LHC era. One can hope that the outcome of this thesis is at least one step towards the fulfilment of this goal. However, one painfully realizes that there are at least as many open issues in the end of such a project as there were in the beginning. Some of the most important of these have been addressed above.

Appendix

The appendix consists of the two following research papers:

Paper 7:

A. Strandlie, J. Wroldsen, R. Frühwirth and B. Lillekjendlie,

Track fitting on the Riemann Sphere.

Proceedings of the International Conference on Computing in High Energy and Nuclear Physics (CHEP'00), Padova, Italy, February 2000.

Paper 8:

R. Frühwirth and A. Strandlie,

Adaptive Multi-track fitting.

Proceedings of the International Conference on Computing in High Energy and Nuclear Physics (CHEP'00), Padova, Italy, February 2000.

References

[1] M. Planck, *Ann. Physik* 4 (1901), 553.

[2] S. Weinberg, *Physical Review Letters* 19 (1967), 1264.

[3] *http://www.fnal.gov/pub/top95/top_news_release.html.*

[4] *ATLAS Detector and Physics Performance Technical Design Report.* CERN/LHCC 99-14, CERN, Geneva, 1999.

[5] *ATLAS Inner Detector Technical Design Report.* CERN/LHCC 97-16, CERN, Geneva, 1997.

[6] R. Frühwirth, Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research* A262 (1987), 444-450.

[7] R. Frühwirth, M. Regler, R. K. Bock, H. Grote and D. Notz, *Data analysis techniques for high-energy physics*, 2nd edition. Cambridge University Press, Cambridge, 2000.

[8] P. V. C. Hough, Machine analysis of bubble chamber pictures. *International Conference on High Energy Accelerators and Instrumentation*, CERN, Geneva, 1959.

[9] S. C. Johnson et al., Three-Dimensional Track Finding in the PHENIX Drift Chamber by a Combinatorial Hough Transform Method. *International Conference on Computing in High Energy Physics (CHEP'98)*, Chicago, USA, 1998.

[10] I. Gavrilenko, Description of Global Pattern Recognition Program (XKalman). *ATLAS Inner Detector Note 165*, CERN, Geneva, 1997.

[11] V. F. Leavers, Which Hough transform? *CVGIP: Image Understanding* 58 (1993), 250-264.

[12] L. Xu and E. Oja, Randomized Hough Transforms (RHT): Basic Mechanisms, Algorithms and Computational Complexities. *CVGIP: Image Understanding* 57 (1993), 131-154.

[13] M. Ohlsson, C. Peterson and A. Yuille, Track finding with deformable templates — the elastic arms approach. *Computer Physics Communications* 71 (1992), 77-98.

[14] A. Gelb (editor), *Applied Optimal Estimation.* The MIT Press, Cambridge, 1974.

[15] P. Billoir, Track fitting with multiple scattering: a new method. *Nuclear Instruments and Methods in Physics Research* 225 (1984), 352-356.

[16] P. Billoir, R. Frühwirth and M. Regler, Track element merging strategy and vertex fitting in complex modular detectors. *Nuclear Instruments and Methods in Physics Research* A241 (1985), 115-131.

[17] P. Billoir, Progressive track recognition with a Kalman-like fitting procedure. *Computer Physics Communications* 57 (1989), 390-394.

[18] P. Billoir and S. Qian, Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nuclear Instruments and Methods in Physics Research* A294 (1990), 219-228.

[19] P. Billoir and S. Qian, Further test for the simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nuclear Instruments and Methods in Physics Research* A295 (1990), 492-500.

[20] R. Frühwirth, *Application of Filter Methods to the Reconstruction of Tracks and Vertices in Events of Experimental High Energy Physics.* Ph. D. Thesis, Technical University of Vienna, 1988.

[21] M. Regler and R. Frühwirth, Reconstruction of charged tracks. In: *Techniques and Concepts of High Energy Physics V*, Plenum, New York, 1990, 407-499.

[22] R. Frühwirth, Track fitting with long-tailed noise: a Bayesian approach. *Computer Physics Communications* 85 (1995), 189-199.

[23] R. Frühwirth, Track fitting with non-Gaussian noise. *Computer Physics Communications* 100 (1997), 1-16.

[24] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79 (1982), 2554-2558.

[25] C. Peterson, Track finding with neural networks. *Nuclear Instruments and Methods in Physics Research* A279 (1989), 537-545.

[26] B. Denby, Neural networks and cellular automata in experimental high energy physics. *Computer Physics Communications* 49 (1988), 429-448.

[27] J. Hertz, A. Krogh and R. Palmer, *Introduction to the theory of neural computation.* Addison-Wesley, Redwood City, CA, USA, 1991.

[28] C. Peterson and J. R. Anderson, A Mean Field Theory Learning Algorithm for Neural Networks. *Complex Systems* 1 (1987), 995-1019.

[29] G. Stimpfl-Abele and L. Garrido, Fast track finding with neural networks. *Computer Physics Communications* 64 (1991), 46-56.

[30] M. Diehl, M. Jünger, R. Frühwirth and J. Scherzer, Global optimization for track finding. *Nuclear Instruments and Methods in Physics Research* A389 (1997), 180-183.

[31] M. Gyulassy and M. Harlander, Elastic tracking and neural network algorithms for complex pattern recognition. *Computer Physics Communications* 66 (1991), 31-46.

[32] B. Lasiuk et al., Development of an Elastic Tracking Package. *International Conference on Computing in High Energy Physics (CHEP'98)*, Chicago, USA, 1998.

[33] A. Caner et al., Track and vertex finding performance with the CMS inner tracker. *Nuclear Instruments and Methods in Physics Research* A435 (1999), 118-143.

[34] W. Press et al., *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.

[35] M. Ohlsson, Extensions and explorations of the elastic arms algorithm. *Computer Physics Communications* 77 (1993), 19-32.

[36] M. Lindström, Track reconstruction in the ATLAS detector using elastic arms. *Nuclear Instruments and Methods in Physics Research* A357 (1995), 129-149.

[37] R. Blankenbecler, Deformable templates — revisited and extended — with an OOP implementation. *Computer Physics Communications* 81 (1994), 318-334.

[38] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society B* 39 (1977), 1-22.

[39] J. F. Crawford, A non-iterative method for fitting circular arcs to measured points. *Nuclear Instruments and Methods in Physics Research* 211 (1983), 223-225.

[40] M. Hansroul, H. Jeremie and D. Savard, Fast circle fit with the conformal mapping method. *Nuclear Instruments and Methods in Physics Research* A270 (1988), 498-501.

[41] V. Karimäki, Effective circle fitting for particle trajectories. *Nuclear Instruments and Methods in Physics Research* A305 (1991), 187-191.

[42] R. L. Streit and T. E. Luginbuhl, Probabilistic Multi-Hypothesis Tracking. *NUWC-NPT Technical Report* 10428, Naval Undersea Warfare Center, Newport, Rhode Island, 1995.

[43] N. Ueda and R. Nakano, Deterministic Annealing EM algorithm. *Neural Networks* 11 (1998), 271-282.

[44] T. Todorov and R. Frühwirth, personal communication.

[45] D. Stampfer, M. Regler and R. Frühwirth, Track fitting with energy loss. *Computer Physics Communications* 79 (1994), 157-164.

[46] R. Frühwirth and S. Frühwirth-Schnatter, On the treatment of energy loss in track fitting. *Computer Physics Communications* 110 (1998), 80-86.

[47] R. Frühwirth and M. Regler, On the quantitative modelling of core and tails of multiple scattering by Gaussian mixtures. *To appear in: Nuclear Instruments and Methods in Physics Research A.*

[48] P. J. Huber, *Robust Statistics.* Wiley, New York, 1981.

[49] R. Frühwirth, P. Kubinec, W. Mitaroff and M. Regler, Vertex reconstruction and track bundling at the LEP collider using robust algorithms. *Computer Physics Communications* 96 (1996), 189-208.