ATLAS Internal Note
INDET- DRAFT
7 January 1996

XKALMAN algorithm description

I.Gavrilenko
P.N.Lebedev Physical Institute, Moscow, Russia

Abstract

This document describes a pattern recognition algorithm for the ATLAS inner detector. The algorithm starts the pattern recognition in the TRT using a histogramming method. Results from the TRT are used as pointers for track search in the 'precision' layers using the Kalman filter-smoother formalism. Each track reconstructed in the 'precision' layers is then extrapolated back to the TRT for final reconstruction and fitting.

Contents

1 Introduction

XKALMAN is a package for global pattern recognition and track fitting in the Atlas inner detector for charged particle tracks with transverse momentum above 0.5 GeV. The essential ingredient in the algorithm is the treatment of multiple scattering noise contributions and bremsstrahlung energy losses during the track search procedure. Pattern recognition is initiated by finding track segments in the TRT and can be performed over a limited region or over the complete inner detector. For the reconstruction, XKALMAN uses two different techniques: a histogramming method and the Kalman filter-smoother formalism. The former is suitable for the track finding in the TRT and the latter for high-granularity detectors such as pixels or silicon strips. XKALMAN uses a set of standard programs from the ATLAS reconstruction package (ATRECON) as interfaces to the different subdetector digits and geometrical parameters and also to a set of standard ZEBRA output banks. XKALMAN also contains a library of routines for track fitting in the framework of Kalman filter-smoother formalism, which can be used outside the XKALMAN package itself. This note describes the overall structure of the algorithm and the main steps of the procedure. Finally some results from the implementation of the algorithm for B-physics studies are presented as conclusions.

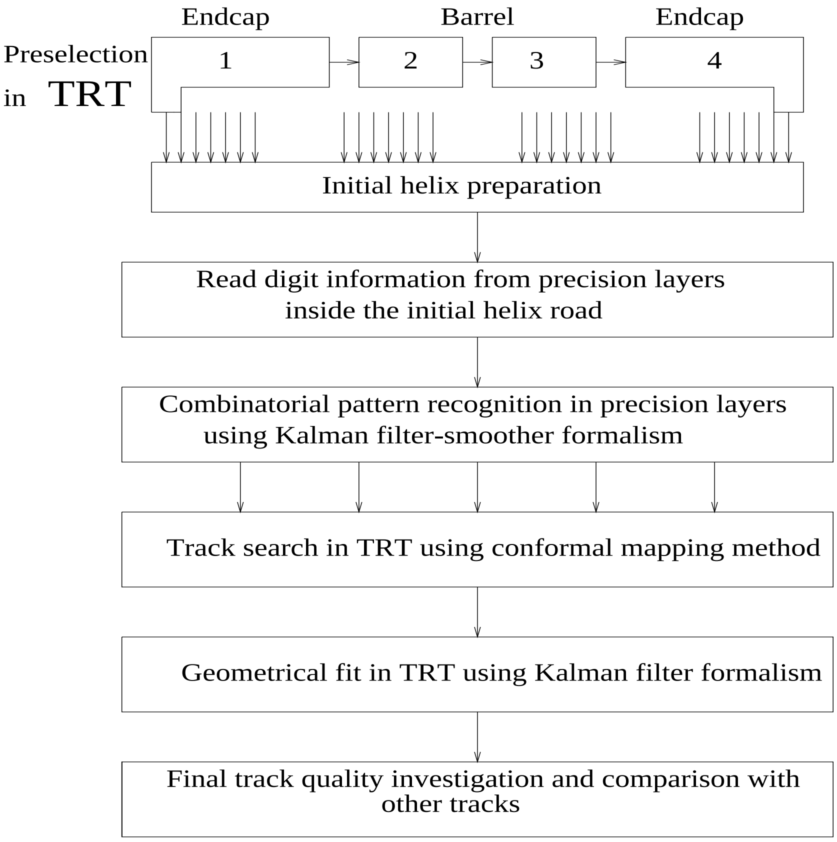

Figure 1: Framework of XKALMAN

2 The framework of XKALMAN

Any track reconstruction algorithm, which attempts to save computer time, contains global and local pattern recognition steps. The global pattern recognition should produce a map of all possible track trajectories and initiate a detailed track-finding process along each of them. The presence of this first step becomes essential in the case of a large number of extra hits arising from noise and low momentum particles, which will never be properly assigned to reconstructed tracks. All these hits will produce a sea of false combinations, which should be rejected during the reconstruction. Thus the global pattern recognition step should separate more or less promising track directions from obviously wrong ones, by using a simple and fast procedure. For this, one has to choose the sub-detector, which can initiate such a reconstruction in the simplest possible way.

The framework of XKALMAN is shown in Fig.1; it reflects the structure of the ATLAS Inner Detector, where high-granularity detectors (silicon strips and pixels) with a small number of 'precision' hits on the track direction are surrounded by a straw tube tracker (TRT) with a large number of track points and electron identification capability. There are several natural reasons to initiate the global pattern recognition process in the TRT despite its high occupancy, in order to avoid a possible combinatorial explosion in the global track search at very high luminosity:

- at the outer radius of the Inner Detector volume, the track density is significantly smaller than at lower radii;
- the single-point measurement accuracy in the TRT ($\sim$1.2 mm without the drift-time measurement) allows a straight-line track representation in most cases, thus considerably simplifying the pattern recognition;
- the large number of crossed straws per track ($\sim$40) allows the use of a simple and fast histogramming method, which nevertheless provides very low rates of fake tracks in a natural way (no complicated χ^2-cuts);
- a special treatment can be used for electron identification in the very first step of the pattern recognition, as for example in the case of a second-level trigger using $J/\psi \longrightarrow e^+e^-$ decays;
- the best efficiency for reconstructing tracks from secondary decays ($K_s^0 \longrightarrow \pi^+\pi^-$) or from photon conversions at large radii has been obtained by starting the pattern recognition from the outermost radius.

Obvious potential drawbacks of such a procedure may arise in the case of very high track densities, where the limited straw granularity will lead to the impossibility of separating multiple tracks crossing the same straws. This however occurs very rarely, even in the case of b-tagging for very high-p_T jets, due to the soft jet fragmentation functions. Effects of this type have however been observed, e.g. in the case of high-p_T $\tau \longrightarrow$3-prong decays [1], which, even if of very limited impact on physics in ATLAS, are of interest to understand the limitations of the method. Because of this intrinsic limitation, provision has been made in XKALMAN for allowing several reconstructed tracks, in the local pattern recognition step through the 'precision' layers, for each possible track trajectory found in the TRT in the first global pattern recognition step.

The first step of the track-finding process is therefore a global pattern recognition step in the TRT, which delivers as end-product a set of possible track-candidate trajectories over the full η-coverage of the Inner Detector and for all transverse momenta larger than 0.5 GeV (tracks with $p_T < 0.4$ GeV loop in the ATLAS Inner Detector). Each of these trajectories is defined as an initial helix, with its set of parameters and their covariance matrix. This helix is then used to define a track road through the 'precision' layers, along which are collected all the measured 'hits' in each 'precision' layer.

The second step of the XKALMAN track-finding process is then initiated, and the program attempts to find all possible helix trajectories within the initial road and with a sufficient number of 'precision hits'. During this step, the tracking and the fitting are performed simultaneously using the Kalman filter-smoother formalism. Each candidate helix trajectory is assigned a quality, calculated on the basis of the total number of 'hits' in the 'precision' layers and on overall χ^2. A cut is applied at the end of this second step to reject the ambiguous trajectories of insufficient quality.

In a third step, the accepted helix trajectories are then extrapolated back into the TRT, where a narrow road can be defined around the result of the extrapolation. The width of this road is obviously determined by the errors on the helix trajectory. All TRT drift-time hits found within this road are then included for the final track-finding and track-fitting steps.

All final accepted helix trajectories are then stored in an output bank for reconstructed tracks, using a well-specified and agreed-upon format [2]. This bank contains all relevant helix parameters at the innermost and outermost radius of the helix trajectory, as well as a list of all 'hits' used from the 'precision' layers.

In summary, the XKALMAN framework contains two independent steps: global and local pattern recognition. The filter is an independent module and the flexible structure of XKALMAN allows it to be used in various situations, such as:

- reconstruction and fitting of all tracks in an event (or part of an event);
- supplementary track-finding and fitting after another global pattern recognition procedure has been applied (e.g. after pixel pattern recognition);
- fitting of tracks reconstructed by a totally independent pattern recognition procedure.

3 Global pattern recognition in the TRT

3.1 Track representations in the TRT

The TRT is a straw drift-tube detector with 4 mm diameter Kapton straws with an average spacing of 8 mm. The free space between straws is filled with polypropylene foils or fibres which produce transition radiation photons. The TRT is positioned above a radius of 56 cm and over $\pm$ 340 cm along z and consists of two distinct geometrical parts:

- two barrel-halves consisting of axial straws with radii between 56 and 107 cm;
- two endcap parts consisting of layers of radial straws grouped into 18 wheels per endcap part. The active radii of the straws are between 64 and 103 cm, except for

the four wheels at largest-z, where the straw active radius extends down to 48 cm, thus maximising the coverage at large η.

The global pattern recognition algorithm is applied in the 2-dimensional projections most natural to the geometry of the TRT, i.e. in the R-ϕ plane for the barrel TRT and in the z-ϕ plane for the endcap TRT. The drift-time information from the straws is not used in this first global step. In a uniform magnetic field, parallel to z, the trajectory of a particle of charge q (in units of electron charge), transverse momentum p_T (in GeV), impact parameter ϵ (in cm), initial azimuthal angle ϕ_0 (in radians) and initial longitudinal position Z_0, can be described in the R-ϕ, z-ϕ and R-z projections as follows :

$$sin(\phi - \phi_o) = \frac{Rc_T}{2}\frac{(1-\epsilon^2/R^2)}{(1-c_t\epsilon)} + \frac{\epsilon}{r} \simeq \frac{Rc_T}{2} + \frac{\epsilon}{R} \tag{1}$$

$$\phi \simeq \phi_0 + \frac{c_T(z-z_0)}{2T} + \frac{T\epsilon}{(z-z_0)} \tag{2}$$

$$r \simeq 2 \mid \frac{sin(c_T(z-z_0)/2T)}{c_T} \mid \simeq \mid \frac{z-z_0}{T} \mid \tag{3}$$

where $c_T = .006q/p_T$ is the transverse curvature in a 2T field and $T = p_Z/p_T = 1/tan\theta$ with θ the polar angle of the trajectory.

The global pattern recognition is performed assuming that the impact parameter ϵ is zero, thus simplifying the above formulae to:

$$\phi = \phi_0 + Arcsin(\frac{Rc_T}{2}) \simeq \phi_0 + \frac{Rc_T}{2} \tag{4}$$

$$\phi = \phi_0^{'} + \frac{zc_z}{2} \tag{5}$$

where $\phi_0^{'} = \phi_0 - \frac{Z_0 c_T}{2T}$ and $c_Z = \frac{c_T}{T}$ is the longitudinal curvature. These simplified equations show that tracks originating from the primary vertex, with $p_T \geq 0.5$ GeV appear as almost straight lines in the R-ϕ plane for the barrel TRT (equation (4)), and as rigorous straight lines in the z-ϕ plane for the endcap TRT (equation (5)). The measurement of the slope of the track trajectory in these planes with respect to the horizontal axis (R or z) provides a direct measurement of p_T for the barrel and of p_Z for the endcap, as shown schematically in fig.2. As can be seen from fig.2, the main difference between the barrel and endcap TRT geometries, arises from the uncertainty in the knowledge of Z_{start} and Z_{stop} for tracks in the endcap part. This uncertainty contributes to the measurement uncertainty on the track p_T, as determined using $p_T = p_z tan\theta = p_z \frac{R_{stop}-R_{start}}{Z_{stop}-Z_{start}}$. If the reconstructed track trajectory originated from a secondary vertex with significant transverse impact parameter ϵ, the approximate equations used will result in a systematic error on the corresponding curvature measurement, which can be estimated to be:

$$\frac{\delta(c)}{c} \simeq \frac{\epsilon}{c < R >^2} \tag{6}$$

where $< R >$ is the average radius of the measured track trajectory, i.e $< R > \simeq 83$ cm. For an impact parameter of about 1 cm, the relative curvature error will then be about $2\% \times p_T$. Such a systematic error is not crucial for the result of the XKALMAN procedure, since the first global pattern recognition step is meant to provide only a coarse map of all possible trajectories.

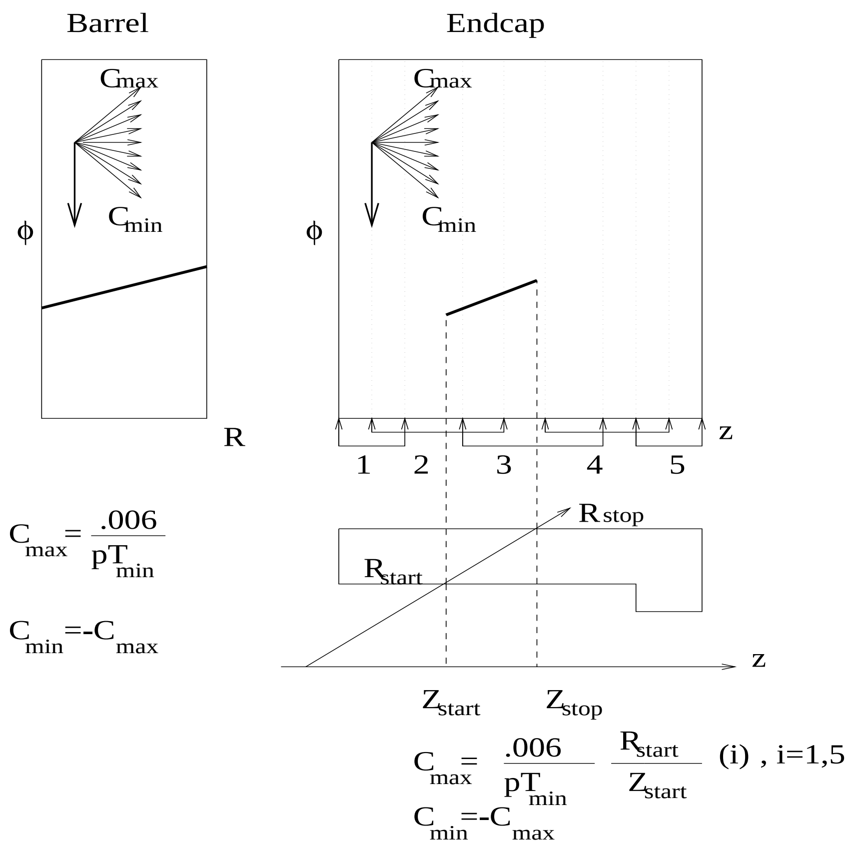

Figure 2: Track representations in the TRT

3.2 Histogramming method

The straight-line track representation of equations (4) and (5) allow the use of the same algorithm for pattern recognition in the different parts of the TRT. Since straight-line segments are among the simplest possible representations of the structure of an image, one can obviously find many well-known and widely used methods to extract straight lines from the image. One such is the Hough transformation [3], based on the very simple idea, that in the (R-ϕ) and (Z-ϕ) image planes with a fixed granularity, there are the same number of straight lines, which can be parameterised using two variables: (ϕ_0, c_T) for the barrel and (ϕ_0, c_Z) for the endcap. The possible values of these two variables can be plotted as points in the two-dimensional parameter plane. Thus, a straw hit in the image plane is transformed into many points in this parameter plane, but conversely straw hits lying on the same line in the image plane will all increase the content value of a single cell of the cumulative array in the (ϕ_0, c) parameter space. Once all cells have been filled appropriately, straight lines can then be detected by simply scanning the contents of the cumulative array.

As shown in fig.2, the track-candidate search is thus carried out by considering a limited set of successive track roads, determined by equations (4) and (5), for all values of ϕ_0, using different curvatures (c_T and c_z) and the two charge hypotheses (q $= \pm 1$). For each road or bin in the (ϕ_0, c) space, the number of straw hits crossed by the road is added to the content of a two-dimensional histogram representing the (ϕ_0, c) space. The roads are defined between R_{start} (resp. Z_{start}) and R_{stop} (resp. Z_{stop}) for the barrel (resp. endcap) parts. Whereas the definition of R_{start} and R_{stop} is uniquely defined R_{start}=R_{inner} and R_{stop}=R_{outer} for the barrel TRT, the situation in the endcap TRT is more complicated, as shown in fig.2, since Z_{start} and Z_{stop} are a function of T=$p_Z/p_T = 1/tan\theta$:

$$Z_{start} = TR_{inner} \quad and \quad Z_{stop} = TR_{outer} \tag{7}$$

This problem is solved for the endcap TRT, by dividing it into five overlapping parts along the Z-direction, and by carrying out the search for track candidates separately for each of these regions.

The algorithm initiates the search from a minimal curvature, c_{min}, to a maximum curvature, c_{max}, with a fixed step, dc of typically 0.0001 cm^{-1}. The values of c_{min} and c_{max} are:

$$c_{min} = -\mid \frac{.006}{p_T^{min}} \mid \quad and \quad c_{max} = -c_{min}, \tag{8}$$

with c = c_T for the barrel TRT, and

$$c_{min} = -\mid \frac{.006 R_{inner}(i)}{p_T^{min} Z_{start}(i)} \mid \quad and \quad c_{max} = -c_{min}, \tag{9}$$

with c = c_z for the endcap TRT, and where p_T^{min} is the minimum p_T of charged particles to be considered, and $R_{inner}(i)$ is the innermost radius for part i in the endcap TRT, as shown in fig.2. For each value of c considered, a straw hit can be represented as two boundaries ϕ_1 and ϕ_2 on the ϕ_0-axis, as shown schematically in fig.3. For the barrel TRT:

$$\phi_1 = \phi_{straw} - \frac{R_{straw}}{R} - Arcsin(\frac{Rc_T}{2}), \tag{10}$$

$$\phi_2 = \phi_{straw} + \frac{R_{straw}}{R} - Arcsin(\frac{Rc_T}{2}) \tag{11}$$

and for the endcap TRT :

$$\phi_1 = \phi_{straw} - \frac{R_{straw}}{< R >} - \frac{Zc_Z}{2}, \tag{12}$$

$$\phi_2 = \phi_{straw} + \frac{R_{straw}}{< R >} - \frac{Zc_Z}{2} \tag{13}$$

where ϕ_{straw} corresponds to the centre of the straw, R_{straw}=0.2 cm is the straw radius, and $< R > = (R_{inner}(i) + R_{outer}(i))/2$ in the endcap TRT. Thus, a track trajectory of given curvature crossing a given straw has as initial azimuthal angle a value ϕ_0 in the interval [ϕ_1,ϕ_2]. As one scans across increasing values of ϕ_0, ϕ_1 represents the given straw boundary as the trajectory enters the straw and is therefore assigned a positive weight of +1, whereas ϕ_2 represents the given straw boundary as the trajectory exits the straw and is therefore assigned a negative weight of -1. The total number of straw boundaries is thus twice the total number of straw hits. As illustrated by fig.3, the cumulative weight

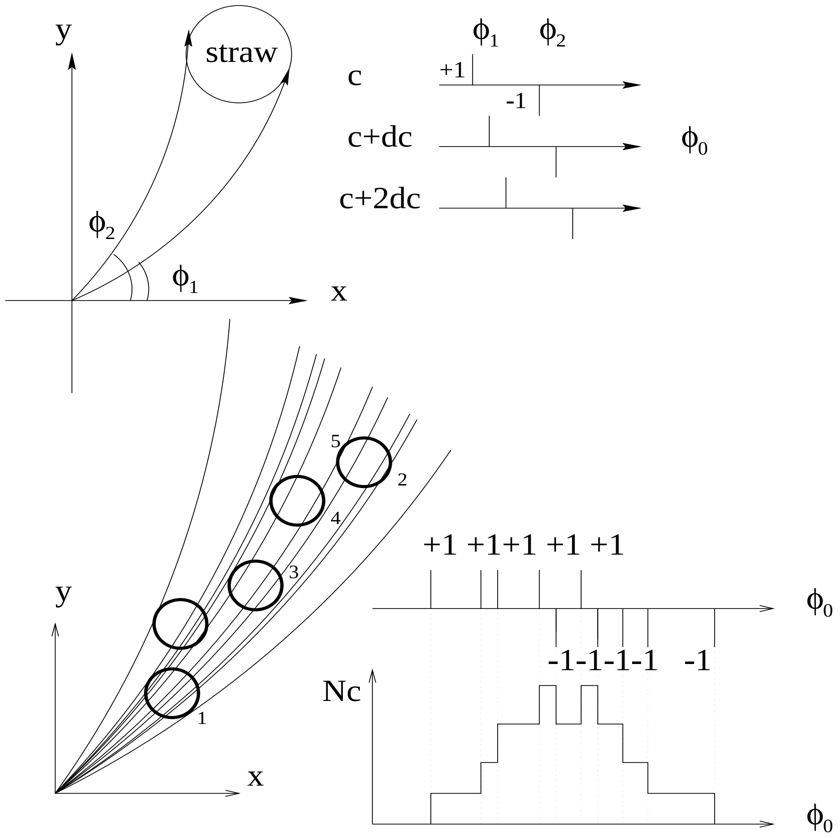

Figure 3: Representation of straw hits in (ϕ_0,c) space

of the straw boundaries scanned in increasing values of ϕ_0 for a given value of c will peak at a value N_c corresponding to the maximum total number of straw hits crossed by the particular road under investigation. All straw boundaries correspond to a bin of the histogram. For each bin in ϕ_0, the maximum number of crossed straw hits N_c^{max} can be extracted from the various values of N_c as c varies from c_{min} to c_{max}. The iterative procedure therefore consists of three types of elementary operations :

- calculation of azimuthal angles of all hit straw boundaries for a given value of the curvature c;
- ordering of hit straw boundaries in ascending values of ϕ_0;
- calculation of the number of hit straws crossed and comparison to maximum value found in previous steps.

The time needed to run this algorithm is proportional to Nlog(N), where N is the total number of straw hits.

3.3 Preparation of initial helix

From the result of the procedure described above, one has extracted set of boundary hits, corresponding to a cells in (ϕ_0 - c) space with a maximum number of straw hits N_c^{max}. This information is used to prepare the set of initial helix trajectories, for propagation into the 'precision' layers. Because the total number of boundary hits is much larger than the total number of real tracks to be reconstructed, some selection has to be applied at this stage. This selection will be based on the value of N_c^{max}, i.e the set of boundary hits are sorted in decreasing order of N_c^{max} and only those with $N_c^{max} > N_{cut}$ are retained. The procedure starts from the boundary hits with the largest N_c^{max}, for which four sequential operations are performed:

- for the given values of ϕ_0 and c_T or c_z corresponding to this boundary hit, determine the complete set of straws crossed by the track road, as defined by equations (4) and (5);
- determine the maximum value of $N_{hit} - N_{empty}$ in the road, where N_{hit} is the number of valid hit straws and N_{empty} is the number of crossed straws which were not hit. From this maximum value of $N_{hit} - N_{empty}$, the values of R_{start} and R_{stop} or Z_{start} and Z_{stop} can be determined;
- flag the boundary hits crossed by the track road under consideration as invalid for further use;
- prepare the initial helix parameters, if $N_{hit} > N_{cut}$.

The value of N_{cut} used for this iterative procedure is typically 8 to 10, for an average number of about 40 crossed straws.

In the ATLAS Inner Detector, the magnetic field in the active volume is sufficiently homogeneous and mostly parallel to the z-axis, so that the helix track model can be applied for propagation to the 'precision' layers in the second step of the pattern recognition procedure. The 'precision' layers can be best represented as a set of planes parallel to z in the barrel and orthogonal to z in the endcap. The most suitable helix representation in this case is the (x,y,z,ϕ,T,c,ϕ_x,I_{fix}) representation, together with the lower triangle of the covariance matrix, as shown in fig.4.

In this representation x and z are the parameters chosen for track propagation to barrel layers ($I_{fix} = 1$) and endcap layers ($I_{fix} = 3$). The local x-axis is chosen as the normal to the barrel layer surface for the layer considered, and ϕ_x is thus the azimuthal angle of this local x-axis with respect to the general coordinate system of the ATLAS Inner Detector. Initial helix parameters defined from boundary hits information. Fig.4 shows the values of the covariance matrix stored in the initial helix representation in the barrel and endcap cases. The initial covariance matrix terms include the uncertainties of the track-finding procedure described above, as well as those due to the primary vertex spread along Z.

4 Pattern recognition in 'precision' layers

At this stage, for each track-candidate found in the TRT, a progressive track-following procedure is applied to find all possible prolongations of the initial track-candidate in the high-granularity 'precision' layers, which consist of pixel detectors, silicon or GaAs strip

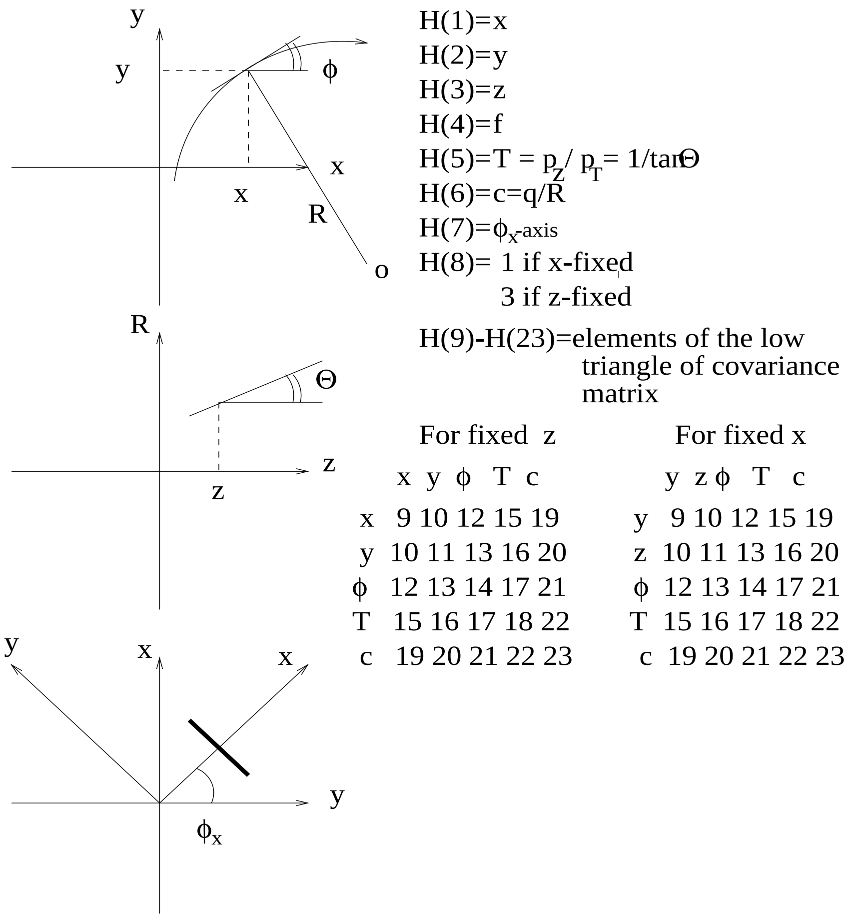

Figure 4: Track parameters for propagation through 'precision' layers

detectors and MSGC detectors (in older versions of the simulation). Thus each track candidate from the TRT is a pointer or an external seed indicating a region of interest in the 'precision' layers. The procedure consists of three main steps:

- preparation of 'precision' layer information inside track-candidate road;
- combinatorial pattern recognition;
- fine-tuning of tracks found and preparation of output helix information.

The Kalman filter-smoother formalism is applied to add the information from the 'precision' layers. This algorithm is now well-known and widely used [4], since it has been adopted by several high-energy physics experiments (for example DELPHI and ALEPH at LEP and ZEUS at HERA). The formalism is elegant and efficient, since it performs a geometrical fit to a charged track trajectory in a way, that allows the inclusion of new measurements along the trajectory one after the other. Unlike traditional pattern recognition algorithms, it always progressively updates the track-candidate information (helix parameters and covariance matrix) during the track-finding process, and thus predicts as precisely as possible of the next hit to be found along the trajectory. This feature can be used to merge into an unique recursive algorithm the track-finding procedure (pattern recognition) and the fitting of the track parameters (coordinates, angles and curvature). The reader is referred to [4] for the general principles of Kalman filtering as applied to track fitting.

More details, pertaining in particular to the context of the DELPHI experiment, can be found in [5]. In the procedure used for the ATLAS Inner Detector, the expressions published in the above references have been however reformulated to maximise efficiency, stability and speed.

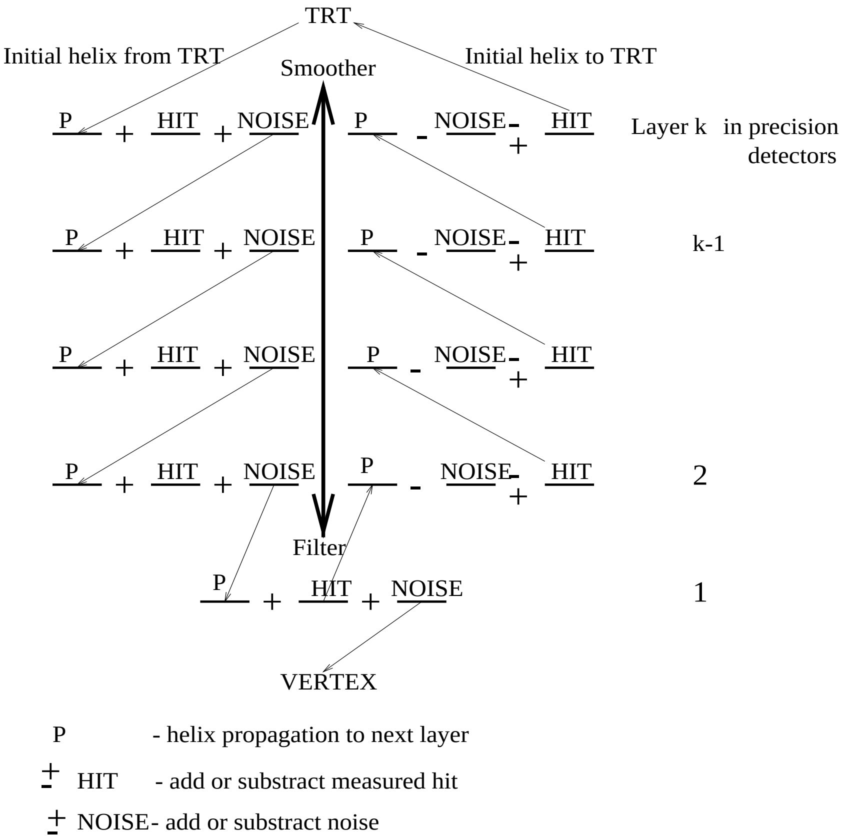

Figure 5: Procedure followed by the Kalman filter-smoother algorithm.

4.1 Description of the Kalman filter-smoother formalism

The Kalman filter requires a starting helix, to be provided by the general pattern recognition procedure. In XKALMAN, this helix is the result of the first step described above, i.e. global pattern recognition using the TRT. As illustrated in fig.5, the XKALMAN filter algorithm consists of a sequence of four types of elementary operations, iterated from the outermost to the innermost 'precision' layers:

- propagation of helix parameters and covariance matrix from layer k to layer k-1;
- comparison of helix prediction in layer k-1 to measured hits;
- addition of measured hit to current helix;
- addition of noise (multiple scattering, bremsstrahlung) for propagation between layers.

The XKALMAN smoother algorithm operates in a very similar way, but in the opposite direction, from the innermost to the outermost layers:

- propagation of helix parameters and covariance matrix from layer k to layer k+1;
- subtraction of noise after propagation between two layers;
- subtraction of old hit from the current helix;
- comparison of helix prediction in layer k+1 to measured hits;
- addition of best measured hit to current helix.

4.1.1 Filtering procedure

For the extrapolation of the helix parameters from layer to layer, the equation for the trajectory of a charge particle in a uniform magnetic field parallel to the z-direction is used:

$$H_{k-1}^{k} = f(H_k^k), \tag{14}$$

where H_k^k represents the filtered helix in layer k and H_{k-1}^k represents the projection of its parameters to layer k-1. For the error propagation, the Kalman filter uses a linear approximation:

$$C_{k-1}^{k} = F_{k-1}(C_k^k + Q_k)F_{k-1}^T, \tag{15}$$

where C_k^k is the covariance matrix of the filtered helix parameters in layer k, Q_k is the additional covariance to be added due to interaction (multiple scattering and/or bremsstrahlung) with the material of layer k, and F_{k-1} is the Jacobian matrix of the helix transformation, $F_{k-1} = \partial H_{k-1}/\partial H_k$. Because the 'precision' layers are planes parallel (barrel) or orthogonal (endcap) to the z-axis, the F_{k-1} matrices can be represented as:

$$F1_{k-1} = \frac{\partial(x, y, \phi, T, c)^{k-1}}{\partial(x, y, \phi, T, c)^k}, \tag{16}$$

$$F2_{k-1} = \frac{\partial(x, y, \phi, T, c)^{k-1}}{\partial(y, z, \phi, T, c)^k}, \tag{17}$$

$$F3_{k-1} = \frac{\partial(y, z, \phi, T, c)^{k-1}}{\partial(x, y, \phi, T, c)^k}, \tag{18}$$

$$F4_{k-1} = \frac{\partial(y, z, \phi, T, c)^{k-1}}{\partial(y, z, \phi, T, c)^k}, \tag{19}$$

where $F1_{k-1}$, $F2_{k-1}$, $F3_{k-1}$ and $F4_{k-1}$ are the respective Jacobian matrices for helix propagation from layer k to layer k-1 in the endcap-to-endcap, barrel-to-endcap, endcap-to-barrel and barrel-to-barrel cases.

After the helix propagation, XKALMAN starts a comparison between the predicted helix position in layer k-1 with the corresponding measured hits within a certain window. If no hits are found within this window, the layer is flagged as missing a hit, and the helix is extrapolated to the next layer. The comparison algorithm used exploits the 'binary' nature of the 'precision' layer measurements. A signal observed in one strip (or cluster of strips) means that the real track passed within a certain distance of the strip centre, but the probability that it passed through any given point within this distance is taken to be uniform. The incremental χ^2 of the helix track, which a new strip contributes, is taken to be 0 if the predicted helix coordinates are within the given strip. Otherwise it is taken to be the minimal $\partial\chi^2$ necessary for the helix to reach the closest boundary of the hit strip. If $\partial\chi^2$ is smaller than a given maximum (usually 10 to 20) the candidate hit is added to the helix. During this comparison procedure, XKALMAN uses only the helix coordinates and their covariance matrix elements as computed in the given layer. The actual filtering step or addition of a hit to the current helix is done by solving the following system of equations:

$$C_{k-1}^{k-1} = (1 + C_{k-1}^{k} U_{k-1})^{-1} C_{k-1}^{k}, \tag{20}$$

$$H_{k-1}^{k-1} = H_{k-1}^{k} + C_{k-1}^{k-1} U_{k-1} (M_{k-1} - H_{k-1}^{k}), \tag{21}$$

where M_{k-1} and U_{k-1} represent the measured hit parameters and their weight matrix, and H_{k-1}^{k-1}, C_{k-1}^{k-1} are a the updated helix parameters and covariance matrix. Solving equation (20) requires in principle the inversion of a 5x5 matrix, which can however be simplified, since the weight matrix U_{k-1} is usually 2x2 in terms of its non-trivial elements. Thus the matrix $1 + C_{k-1}^{k} U_{k-1}$ contains two columns of arbitrary elements, with the remaining diagonal elements equal to unity and off-diagonal elements equal to zero, and it can be easily inverted analytically.

4.1.2 Treatment of interactions with detector

Charged tracks traversing the Inner Detector inevitably interact with the material present in the layers and between layers. These interactions lead to disturbances in the trajectory, the most important of which are multiple scattering and energy loss, either through ionization or bremsstrahlung in the case of electrons. Each 'precision' layer amounts to a total material thickness of about 2 % of a radiation length, resulting in significant scattering angles for low-p_T tracks. Since the distances between successive layers are of order 10 cm, multiple scattering produces average lateral displacements of the trajectory of about $0.02/p_T$ in cm, i.e. about 200 μm for 1 GeV tracks. Such displacements are larger than the 'precision' layer measurement accuracy and have therefore to be taken into account. For charged particles, ionization losses do not influence in any significant way the track reconstruction: for particles of GeV energy, the ionization losses in the Inner Detector are typically 10 to 20 MeV and can be treated as a deterministic function of the material seen by the particle. In the case of electrons, however, the dominant energy-loss mechanism is bremsstrahlung. Electrons of a few GeV energy lose on average 20 to 60 % of their initial energy when passing through the ATLAS Inner Detector material. The treatment of these losses is complicated by the fact that their fluctuations are very large, even compared to the average loss itself.

In the context of Kalman filtering, multiple scattering and energy loss both contribute to the so-called 'noise' process, which has to be taken into account in the predictive equations of the filtering procedure. For all reconstructed trajectories, 'noise' terms have to be added to the covariance matrix of the track parameters at each measurement step. Multiple scattering affects both the track direction (ϕ,T) and transverse curvature (c_T), since the total track momentum is not affected by multiple scattering. Energy loss affects the total track momentum, and therefore only the transverse curvature. In order to calculate these contributions to the 'noise' process, it is sufficient to know the path length in units of radiation length of the track trajectory through the material considered at the current step in the filtering process. This path length through a given surface is in practice described by the track direction and the thickness of the layer. The covariance matrix contribution of the 'noise' process is thus calculated and added to the total covariance matrix of the track parameters before propagation into the next layer. The material surfaces crossed by the track trajectory are always considered in XKALMAN as "thin layers", and the actual material between layers (cables, supports,...) is only approximately taken into account in the procedure. Formulae for the computation of the covariance matrix contribution of multiple scattering exist in the literature [6,7] and are applied directly in XKALMAN. Ionization energy losses are accounted for in XKALMAN only as corrections to the curvature.

For energy loss due to bremsstrahlung in the case of electrons, XKALMAN uses the method proposed in [7], which considers bremsstrahlung 'noise' as a stochastic process, as in the case of multiple scattering: the helix curvature and its variance are corrected at each layer intersection, according to the average estimated bremsstrahlung energy loss in the layer. This correction decreases the absolute value of the curvature when the filtering is performed in the direction opposite to the particle trajectory, and increases it in the reverse case.

The multiple scattering 'noise' matrix contains four non-zero elements:

$$Q_k(3,3) = cov(\phi,\phi)_k = (\frac{.0136GeV}{p_T})^2 S(1 + .038ln(S))^2, \tag{22}$$

$$Q_k(4,4) = cov(T,T)_k = cov(\phi,\phi)_k(1+T^2), \tag{23}$$

$$Q_k(4,5) = cov(T,c)_k = cov(\phi,\phi)_k cT, \tag{24}$$

$$Q_k(5,5) = cov(c,c)_k = cov(\phi,\phi)_k \frac{(cT)^2}{1+T^2}, \tag{25}$$

where S is the path length in units of radiation length of the track trajectory through the layer under consideration.

The correction to the curvature due to ionization energy loss is :

$$\alpha_k = \frac{p + dE}{p}, \tag{26}$$

where p is the total track momentum and dE is the average energy loss in the layer under consideration.

The bremsstrahlung 'noise' matrix contains only one non-zero term:

$$Q_k(5,5) = cov(c,c)_k = S(.413 - .72S)c^2 \tag{27}$$

and the correction to the curvature due to bremsstrahlung is:

$$\alpha_k = 1 - S(1 - \frac{S}{2}). \tag{28}$$

For all candidate trajectories, XKALMAN calculates first a so-called 'μ-fit', which only accounts for multiple scattering and ionization energy losses. If the track candidate was identified as an electron in the TRT (or by the user), XKALMAN recalculates a so-called 'e-fit',which also accounts for bremsstrahlung energy losses. The result of the 'e-fit' is only used by XKALMAN for the final track output, if the 'e-fit' has increased the number of accepted measurement points from the 'precision' layers with respect to the 'μ-fit'.

4.1.3 Kalman smoother procedure

As described earlier, the smoother procedure is executed in the opposite direction to the filtering procedure, and propagates the complete information resulting from this filtering procedure outwards through all 'precision' layers. This smoother procedure thus permits the calculation of best estimates of the track parameters at any point along its trajectory. It also recovers many wrong or missing measurements, therefore considerably improving the individual layer efficiencies and also the overall track reconstruction efficiency and track quality. The propagation equations for this procedure in XKALMAN are :

$$H_k^{k-1} = f(H_{k-1}^n), \tag{29}$$

$$C_k^{k-1} = F_k C_{k-1}^n F_k^T, \tag{30}$$

where H_{k-1}^n and C_{k-1}^n are the smoothed helix and its covariance matrix representation at layer k-1, and F_k is the Jacobian matrix of the helix transformation described in section 4.1.1. The smoothed helix and its covariance matrix are calculated as :

$$H_k^n = H_k^{k-1} - BA(H_k^{k-1} - H_k^k), \tag{31}$$

$$C_k^n = B(C_k^{k-1} B^T + Q_k), \tag{32}$$

with

$$A = Q_k W_k^k \quad and \quad B = (1 + A)^{-1}, \tag{33}$$

where H_k^k, C_k^k, W_k^k are respectively the filtered helix and its covariance and weight matrices, and Q_k is the 'noise' matrix for filtering described in 4.1.2. In the absence of 'noise' processes, where Q_k=0, the smoothing procedure is equivalent to a pure outward-going extrapolation, with $H_k^n = H_k^{k-1}$ and $C_k^n = C_k^{k-1}$, as can be verified directly from the above equations. These equations contain only one direct inversion of a 5x5 matrix for the calculation of W_k^k. The matrix B contains only three columns of non-trivial elements and can therefore be inverted analytically.

The other steps of the smoother procedure, subtraction of old measurement point, addition of new best measurement point and selection of measured points in a given layer, are performed in the same way as in the filtering procedure. The subtraction of a measurement from the helix parameters is obtained by changing the sign of the measurement weight matrix (U_{k-1} is changed to -U_{k-1} in equations (20) and (21)).

4.1.4 Implementation in XKALMAN

The Kalman filter-smoother formalism is applied in XKALMAN through the use of five key subroutines:

- TDPROP : helix propagation;
- THELKS : comparison of helix with measurements in a given layer;
- TDADH : addition or subtraction of points from the helix;
- TDNOIS : calculation of 'noise' contribution to covariance matrix;
- THESMN : removal of 'noise' contribution during the smoothing procedure.

All operations with matrices are implemented explicitly, without subscripted variables or loops, in order to minimise the computation time used. XKALMAN allows the use of either a long or short Kalman filter-smoother procedure. The long procedure performs the complete set of operations described in this Section. The short procedure proceeds without any comparison step between helix and measurements, in the case where the complete list of hits on the track trajectory is already known.

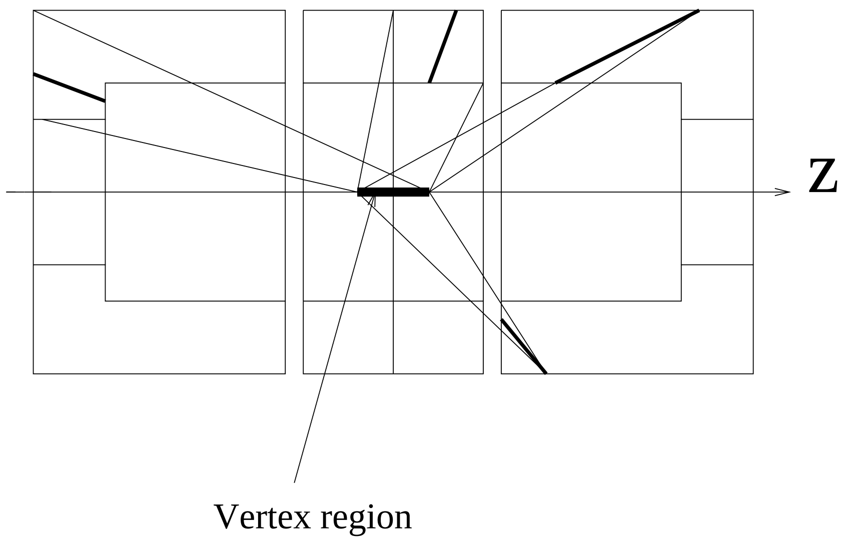

Figure 6: The width of the track road definition in the r-z plane.

4.2 Preparation of initial helix information

The initial helix, as given by the global pattern recognition in the TRT, consists of a set of helix parameters and their covariance matrix. In order to perform the Kalman filter-smoother procedure, the geometry of all sensitive layers and of all the material along the initial trajectory has to be given as input data to the algorithm. In the framework of ATRECON, XKALMAN uses a set of standard interface routines, which provide the measurement information from all 'precision' layers, as lists of clusters of hit pixels or strips. The routines are:

- PIXBCAND and PIXECAND for the barrel and endcap pixels;
- SCTTCAND and ZSCTCAND for the barrel and endcap silicon strips.

The input to these routines from XKALMAN is given as a track road, centred on the current helix and with the appropriate width in the R-ϕ and R-z planes. XKALMAN provides this information in a natural way by propagating the initial helix from the TRT pattern recognition to all the 'precision' layer planes. The width of the track road in the R-ϕ plane is typically taken to be 10-20 times as large as the helix coordinate errors after extrapolation. The width of the track road in the R-z plane, as illustrated in fig.6, has to take into account the primary vertex spread along z and the geometry of the TRT, and is much larger than the width in the R-ϕ plane, particularly in the barrel TRT region. The output is a set of clustered pixels and strips, which contains the complete information about each cluster found within the given road.

Before starting the Kalman filter-smoother procedure, XKALMAN sorts the information into an order appropriate for initiating the helix propagation, and organises it into two separate banks:

- a counter bank, containing the complete information about position, geometry and material thickness of each counter traversed by the initial helix, together with the list of measured clusters in each counter. The counter closest to the interaction vertex is labelled as counter 1, and the last counter in the list is the one closest to the initial helix position;
- a cluster bank, containing the coordinates and size of all measured clusters in the local counter coordinate system. The hit pixels and strips are presently clustered following a naive and somewhat conservative approach.

The Kalman filter-smoother procedure uses the counter bank for the helix propagation and the cluster bank for filtering and smoothing.

4.3 Combinatorial pattern recognition

XKALMAN applies a combinatorial pattern recognition through all 'precision' layers, in order to find all possible prolongations of the initial TRT helix within the track road. All found prolongations of sufficiently high quality can be defined as final track candidates, even if they originate from the same initial helix. The quality of any prolongation is determined from its length or number of 'precision' layer hits and its total χ^2, defined from the incremental χ^2-contributions of each layer (see Section 4.1).

In each layer, after comparing all measured cluster positions with the propagated helix information, the algorithm chooses the closest cluster and stores all other clusters with $\partial\chi^2 < 20$ in the cluster candidate bank. If no cluster is found with $\partial\chi^2 < 20$, the layer is counted as missing and the helix is propagated to the next layer. XKALMAN stops the propagation of the current helix prolongation in either of four cases:

- the algorithm has reached layer 1 closest to the interaction vertex;
- two consecutive layers are labelled as missing;
- the final total length of the given prolongation cannot be sufficient;

- the current prolongation is obviously worse than a previous one, which contains a large number of the identical 'precision'-layer hits.

The found prolongations with a sufficient number of 'precision'-layer hits (typically above 6-7) undergo a comparison with all other prolongations found previously in the same track road. If the number of identical 'precision'-layer hits used by two different prolongations is smaller than a given value, typically 2-3, both prolongations are kept, and otherwise the shorter one is removed. If the number of identical 'precision'-layer hits is above a given cut and the length of both prolongations is the same, the one with larger χ^2 is removed.

If a valid prolongation is retained, XKALMAN performs the short smoother procedure to extrapolate the helix back to the outermost 'precision' layer, and stores the helix parameters and covariance matrix together with a list of all clusters used.

After the current helix prolongation has been investigated, the algorithm investigates a new prolongation, starting from the neares layer containing 'precision'-layer hits added to the propagated helix. The algorithm then takes a new hit from the cluster candidate bank in the given layer. When all clusters in a layer from the cluster-candidate bank are exhausted, the algorithm also considers the possibility that the layer is missing. If all possible combinations in a given layer have already been investigated, the algorithm goes back to the previous layer with 'precision'-layer hits added to the propagated helix. The algorithm therefore stops the search, if for the last layer with 'precision'-layer hits (i.e the one closest to the initial helix), all combinations have been investigated.

The total number of elementary steps during this combinatorial pattern recognition can be very large, and it is important to introduce conditions, when the algorithm can skip some of them. At each step of the procedure, where the algorithm takes into account new 'precision'-layer hits or a missing point, XKALMAN performs the following analysis:

- as discussed above, the total number of elementary steps performed by XKALMAN is minimised by continuously comparing the current prolongation with all previously accepted prolongations. If the current prolongation contains a large number of identical 'precision'-layer hits as a previously accepted prolongation, but its final quality cannot be better, the algorithm stops the current investigation and takes into account a new 'precision'-layer hit;

- another possibility to decrease the number of steps is the introduction of so-called 'stable' hits, i.e. hits added to the helix, with a small disturbance of the trajectory. The estimator used to define 'stable' hit is the difference in trajectory $\partial\chi^2_H$ between the extrapolated helix and the updated one:

$$\partial\chi^2_H = (H^k_k - H^{k+1}_k)(C^k_k + Q_k)^{-1}(H^k_k - H^{k+1}_k)^T \tag{34}$$

  If this difference is small, the new helix obtained after addition of a 'stable' hit is not very different from the previous one and the quality of a prolongation which would miss this layer cannot be better;

- the total number of steps during the combinatorial pattern recognition can be decreased by the introduction of cuts on the coordinates of the first 'precision'-layer hits added to the helix. If the radius of these first 'precision'-layer hits (i.e. those closest to the initial helix) is smaller than a cut (typically 40-45 cm), XKALMAN stops the investigation of the current track road.

Thus, the combinatorial pattern recognition is a set of long filtering and short smoothing steps. At the end of this procedure, a track candidate bank corresponding to the initial TRT helix is updated, if at least one valid prolongations has been found. The algorithm then iterates over all other initial TRT helices.

4.4 Fine-tuning of found tracks

The fine-tuning procedure starts from the stored helix parameters after the combinatorial pattern recognition procedure. The track parameters are therefore stored at this stage at the outermost 'precision' layer. XKALMAN takes into account only the diagonal terms of the covariance matrix multiplied by a constant factor (∼10-100). The fine-tuning procedure is an iterative procedure performing a short filter and a long smoother for each track candidate, until the list of hits on the track are exactly identical after the short filter and long smoother procedures.

The input data for a given iteration step consists of the list of hits assigned to the track in the previous step. The short filter performs the calculation of the 'noise' covariance matrix contribution from the momentum of the input helix rather than the current one. Once the short filter is completed, the long smoother proceeds from the innermost layer (with assigned hits) outwards, and accepts in each layer only the closest cluster with $\partial\chi^2 < 20$. XKALMAN stops the fine-tuning procedure if the lists of hits assigned to the track before and after filter and smoother procedures are exactly the same, otherwise XKALMAN performs another iteration of the fine-tuning procedure beginning the 'short' filter with the first different hit in the outermost layer.

After this fine-tuning procedure, XKALMAN has calculated the best estimates of the track parameters at any point along its trajectory through the 'precision' layers. Only the helix parameters and covariance matrices at the innermost and outermost layers with the assigned hits are saved at this stage for storage in the final track bank, together with the list of 'precision'- layer hits.

5 Local pattern recognition in the TRT

The final step in the XKALMAN algorithm is the optimal assignment of hits in the TRT to all the tracks retained after the Kalman filter-smoother procedure through the 'precision' layers. Note that the straw drift-time information has not been used in the initial global pattern recognition step in the TRT.

The helix parameters at the outermost 'precision' layer, i.e that closest to the TRT, are used to define a narrow road through the TRT. The width of this road is defined typically as 10 times the helix coordinate errors after extrapolation. All straw hits inside this road, with their drift-time information, are used as input data for local pattern recognition in the TRT. The algorithm consists of two steps:

- search for the best track prolongation through the TRT;
- final track fit through all points.

The track search procedure uses the conformal mapping method proposed in [8] for fitting circles. This method transforms the equation of a circle of radius R:

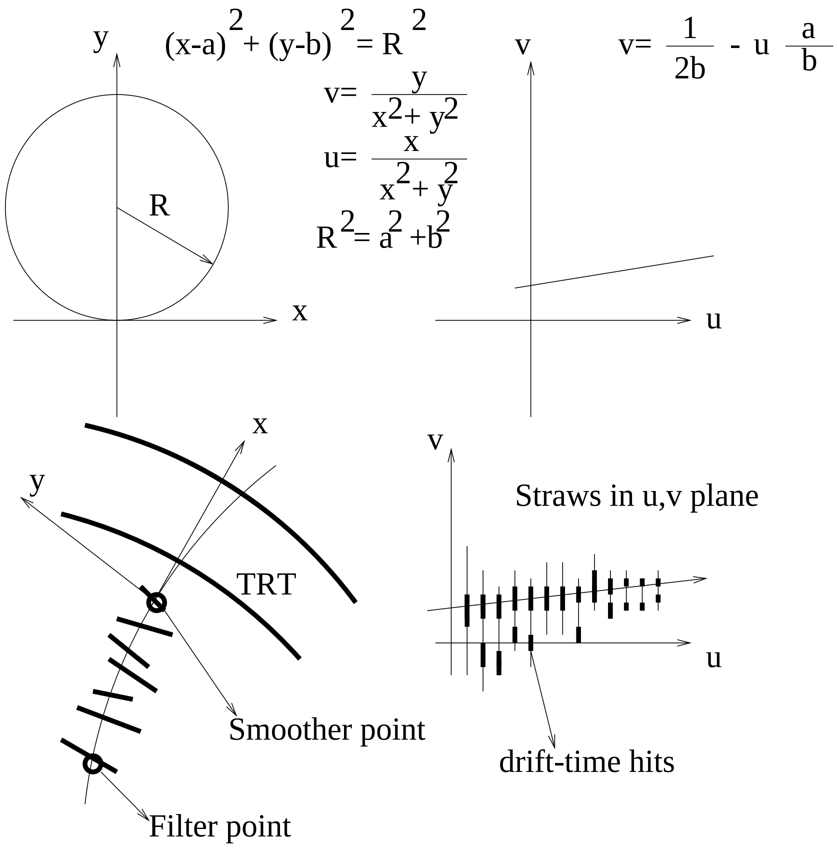

Figure 7: Conformal transformation for track search in the TRT.

$$(x-a)^2 + (y-b)^2 = R^2 \tag{35}$$

into a straight line with the prescription:

$$u = \frac{x}{x^2+y^2} \quad and \quad v = \frac{y}{x^2+y^2}. \tag{36}$$

Provided the circle passes through the origin of the coordinate system, i.e. $R^2 = a^2 + b^2$, then one obtains straight lines in the u-v plane of the form:

$$v = \frac{1}{2b} - u\frac{a}{b}. \tag{37}$$

A natural choice for the origin of the coordinate system for this step in the XKALMAN algorithm is the helix endpoint at the outermost 'precision' layer. The x-axis is then chosen to be parallel to the track direction, such that the result of the transformation is a track representation as a straight line in the u-v plane, almost parallel to the u-axis, as illustrated in fig.7. Each straw hit can be represented in the u-v plane as three segments normal to the u-axis. One segment represents the complete straw and the other two represent the possible track coordinates corresponding to the twofold ambiguity of the drift-time information in the straws. This drift-time accuracy is around 170μ and depends somewhat on the luminosity. The drift-time measurement efficiency at LHC design luminosity varies from 50% to 75% depending on the straw counting rate. The total width of the drift-time hits is taken to be 700μ or $\pm 2\sigma$.

The best track prolongation through the TRT can be defined as the straight line in the u-v plane (v=Au+B), which crosses the largest number of straw segments. The local pattern recognition through the TRT described here can therefore use the same unbinned histogramming method as for the global TRT pattern recognition procedure. Each straw segment is represented as two boundary hits on the v-axis, with weight -1 for the upper one and +1 for the lower one. As illustrated in fig.8, the straight line search in the u-v

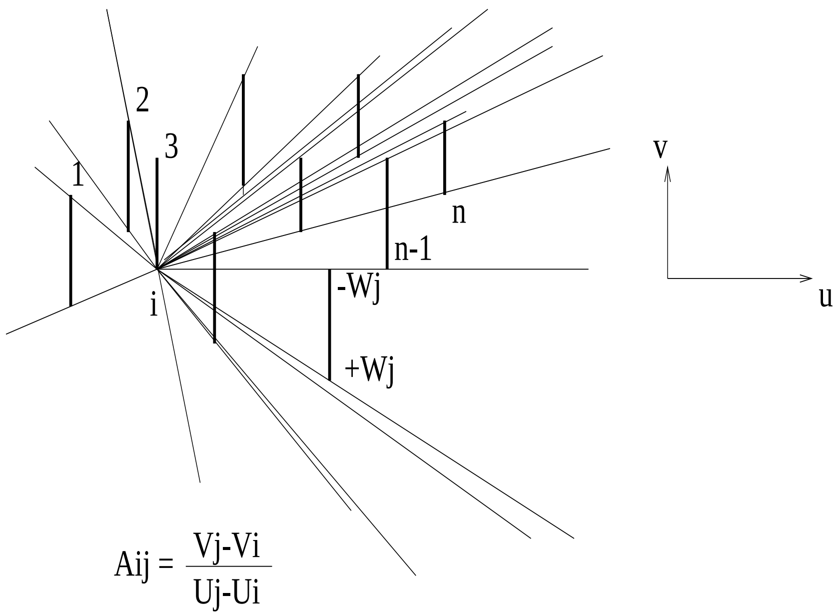

Figure 8: Track search in the TRT

plane is performed as a recursive algorithm over a set of track rotations around a given boundary hit. Each step of this algorithm, starting e.g. from some arbitrary boundary hit i, consists of three sequential operations :

- calculation of slope $A_{i,j}$ and weight $W_{i,j}$ for each i,j pair;
- sorting of all $A_{i,j}$,$W_{i,j}$ in ascending order of A;
- identification of the pair of adjacent boundary hits (j_1, j_2), which maximises the total weight $W_{i,j}$.

The new boundary hit j_1 or j_2 are used instead of i for the next step in the iteration, and the algorithm stops the search if both boundary hits j_1 and j_2 are already flagged as having been used in a previous step. The procedure outputs the parameters of the best straight line trajectory through the TRT, and all straws hits and drift-time hits intercepted by this line are used in the final track fit. This final track fit is performed using the short Kalman filter formalism, moving outwards from the straw closest to the vertex. As a initial helix, the filter uses the helix information at the outermost 'precision' layer from the track candidate bank.

6 Preparation of output tracks bank

After the final pattern recognition and track fitting steps of XKALMAN, the track candidate contains the following information:

- fit procedure used (μ-fit or e-fit);
- helix parameters and their covariance matrix at the end-points of the filter procedure in the 'precision' layers (point on trajectory closest to the vertex) and in the TRT (point on the trajectory closest to calorimeter);
- total χ^2 resulting from final fit procedure;
- list of all hits on track from all subdetectors;
- total number of 'precision' hits N_p;
- total number of straw hits N_s, of empty straws crossed N_e, and of drift-time hits N_t;

A track candidate is stored in the final output bank if it passes the following cuts:

- the number of 'precision' hits is larger than 5-7;
- the ratio $N_s/(N_s+N_e)$ is larger than 0.7-0.8;
- the ratio N_t/N_s is larger than 0.5-0.7;
- no previously accepted track has the same set of hits as the current one; this last cut removes full 'ghost' tracks.

The track information listed above is stored into the standard Inner Detector output banks defined for ATRECON. The helix information at the innermost radius can be used for vertexing algorithms, and that at the outermost radius can be used for propagation to the calorimeters and muon detectors.

XKALMAN contains a set of routines which transform given helix representations into others, according to the user needs. The list of hits on the track trajectory can be used to perform detailed comparisons of the reconstructed track with the simulated one.

7 Applications

Given algorithm was implemented for B-physics and b-tagging studies.

8 Conclusion

9 Acknowledgements

I wish to thank D.Froidevaux for continuos interest, encouragement and help with this work and note preparaton. I thank P.Nevski and U.Egede for interface between XKALMAN and ATRECON.

References

[1] P. Luthaus et al., ATLAS Internal Note PHYS-N0-071

[2] U. Egede, ATLAS Internal Note SOFT-NO-031 15 December 1994

[3] S.A. Dudani and A.L. Luk, Pattern Recognition 10,145-157 (1978); Chee-Woo Kang at al., Pattern Recognition 24,633-641 (1990).

[4] P. Billoir, Nucl. Inst. and Meth. 225 (1984) 352; P. Billoir, Comput. Phys. Comm. 57 (1989) 390; P. Billoir and S. Qian, Nucl. Inst. and Meth. A294 (1990) 219; R. Fruhwirth, Nucl. Inst. and Meth. A262 (1987) 444.

[5] R. Fruhwirth at al., Nucl. Inst. and Meth. A334 (1993) 528.

[6] E.J. Wolin and L.L. Ho, Nucl. Inst. and Meth. A239 (1993) 493.

[7] D. Stampfer at al., Comput. Phys. Comm. 79 (1994) 157.

[8] M. Hansroul at al., Nucl. Inst. and Meth. A270 (1988) 498.

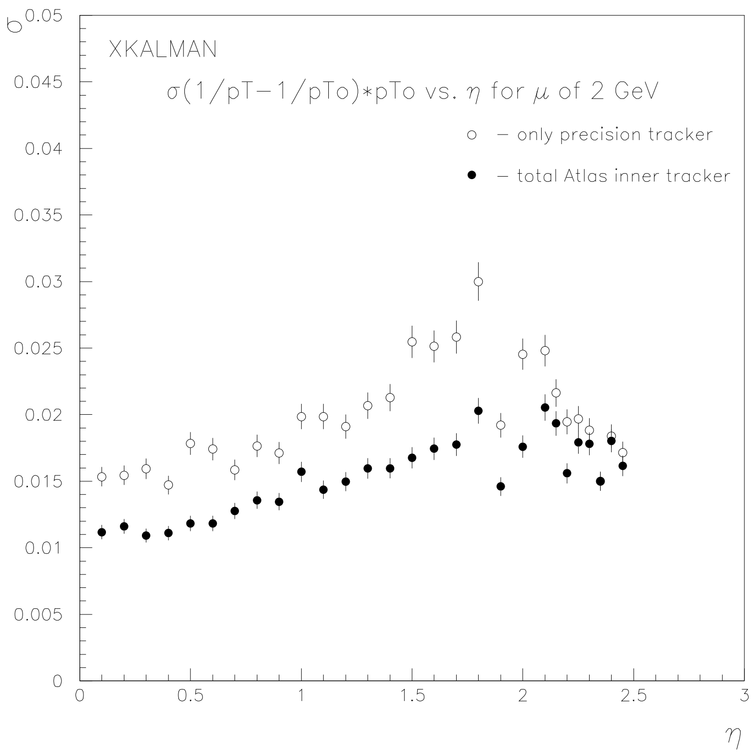

Figure 9: Momentum resolution vs. rapidity

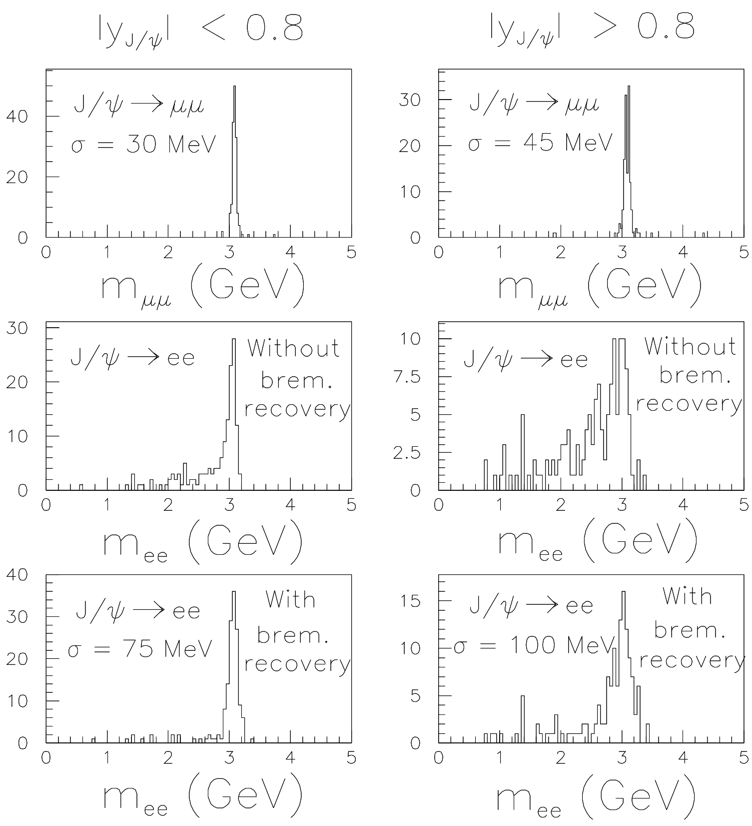

Figure 10: Distributions of reconstructed dilepton masses, for $J/\psi \to \mu^+ \mu^-$ (top), for $J/\psi \to e^+e^-$ without bremsstrahlung recovery (middle), and for $J/\psi \to e^+e^-$ with bremsstrahlung recovery (bottom). The results are shown separately for J/ψ rapidities smaller (left) and larger (right) than 0.8.