Journal of Physics: Conference Series **664** (2015) 072035 doi:10.1088/1742-6596/664/7/072035

Event Reconstruction Techniques in NOvA

M Baird[1], J Bian[2], M Messier[1], E Niner[1], D Rocco[2], K Sachdev[2]

[1]Indiana University Department of Physics, 727 E. Third St. Bloomington, IN 47405, USA
[2]University of Minnesota School of Physics & Astronomy, Tate Lab Room 148, 116 Church Street S.E., Minneapolis, MN 55455 USA

E-mail: mbaird42@fnal.gov, eniner@indiana.edu, ksachdev@physics.umn.edu

Abstract. The NOvA experiment is a long-baseline neutrino oscillation experiment utilizing the NuMI beam generated at Fermilab. The experiment will measure the oscillations within a muon neutrino beam in a 300 ton Near Detector located underground at Fermilab and a functionally-identical 14 kiloton Far Detector placed 810 km away. The detectors are liquid scintillator tracking calorimeters with a fine-grained cellular structure that provides a wealth of information for separating the different particle track and shower topologies. Each detector has its own challenges with the Near Detector seeing multiple overlapping neutrino interactions in each event and the Far Detector having a large background of cosmic rays due to being located on the surface. A series of pattern recognition techniques have been developed to go from event records, to spatially and temporally separating individual interactions, to vertexing and tracking, and particle identification. This combination of methods to achieve the full event reconstruction will be discussed.

1. Introduction

The NOvA experiment is an active long-baseline neutrino oscillation experiment using the recently upgraded NuMI beam at Fermilab to measure $\nu_\mu \rightarrow \nu_e$, $\overline{\nu}_\mu \rightarrow \overline{\nu}_e$, $\nu_\mu \rightarrow \nu_\mu$ and $\overline{\nu}_\mu \rightarrow \overline{\nu}_\mu$ oscillations [1]. The experiment uses two functionally identical detectors composed of alternating horizontal and vertical planes of PVC plastic cells. Each cell has a 4×6 cm cross section and is filled with liquid scintillator and a looped wavelength-shifting fiber attached to an avalanche photodiode (APD) for light collection. The detectors are separated by 809 kilometers and located 14 milliradians off-axis from the beam center in order to produce a narrow-band 2 GeV beam near the oscillation maximum for ν_e appearance. The Far Detector located in Ash River Minnesota is 15.6m $\times$ 15.6m $\times$ 60m, totaling 14 kilotons and 344,064 individual cells. The Near Detector is located one km from the target at Fermilab and is 4.2m $\times$ 4.2m $\times$ 15.8m for a total of 300 tons and 20,192 cells.

1.1. Event Topology

The goal of the reconstruction chain presented here is to identify ν_e charged-current interactions and reject ν_μ charged-current (CC) and neutral-current (NC) backgrounds. For comparison, a simulated event of each type is shown in Figure 1. Each event was simulated with identical four-vectors for the particles. The events contain a 0.78 GeV momentum proton and then a second 1.86 GeV momentum particle (e,μ,π^0) to represent a 2.15 GeV neutrino interacting with the detector. The figure shows one detector view for each event and the cell hits are colored by the charge deposited. The more difficult backgrounds are NC interactions with a single π^0, shown in

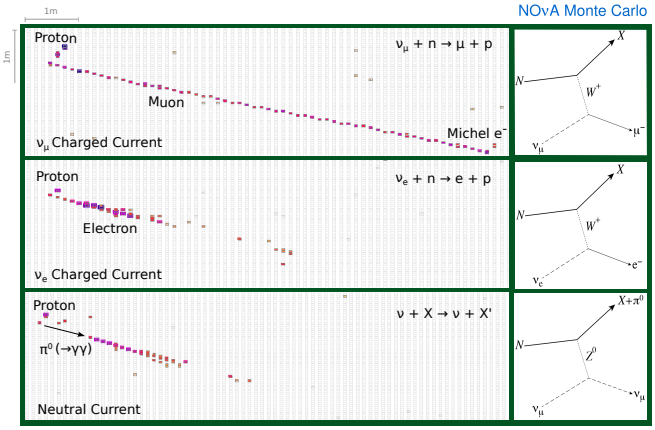

Figure 1. Simulated 2.15 GeV neutrino interactions showing a toy ν_μ CC interaction (top panel), ν_e CC interaction (middle panel), and NC interaction (bottom panel).

the lower panel. The π^0 has a 98.8% branching ratio to decay into two photons which produce electromagnetic showers that are difficult to distinguish from electrons. The photons travel some distance before converting into an e^-/e^+ pair and producing scintillation light, where the photon conversion distance in the NOvA detector is ~ 38 centimeters (6 plane widths). One handle on identifying π^0's is looking for a gap between the vertex and the start of a shower, which is why NOvA was designed with low-z materials to yield a longer conversion distance.

1.2. Reconstruction Philosophy

For the ν_e appearance analysis it was desired to have one reconstruction chain that was successful in picking out showery electrons while still performing reasonably well for the more track-like muons/protons/pions. It was decided to take a reconstruction approach of finding the global event vertex first before forming final particle reconstruction objects. This approach was chosen because short tracks of only a few cells (protons, neutrons, photons) become more significant in light of the vertex.

The full chain of reconstruction is outlined with a toy example in Figure 2. Reconstruction begins by separating individual event interactions from a larger readout window into reconstruction objects known in NOvA as "slices" (described in Section 2) which serve as the foundation for all later reconstruction stages. Next, a modified Hough transform is applied to identify prominent features or guidelines in a slice (Section 3). Then the guidelines are used as seeds to an algorithm to determine the global 3D vertex for the slice under the assumption that all activity in the slice has a common origin (Section 4). The vertex is then used as a seed to a fuzzy k-means algorithm that produces prongs (a collection of cell hits with a start point and direction) which contain the activity of particles in the event (Section 5). Finally, a variety of variables resulting from this reconstruction (dE/dx profiles, distance between vertex and prongs, length, direction, etc.) are fed into a neural net to classify the degree to which the slice was like a ν_e CC interaction (Section 6).

2. Interaction Separation with DBSCAN

For the NOvA experiment, a "hit" is the information from a single triggered channel and a "data event" is a collection of hits from an arbitrary readout window, currently 550 μsec. These hits can be divided into two categories, signal and noise where the signal hits can originate from multiple independent sources. In the Far Detector, located on the surface, the primary

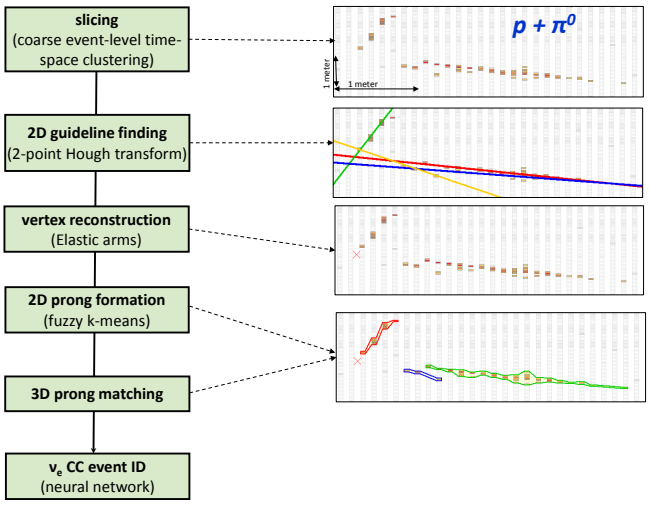

Figure 2. Flowchart of the reconstruction algorithms used to identify ν_e CC interactions.

concern is separating 50-70 cosmic rays in a readout window. For the Near Detector ∼4 neutrino interactions are expected in each 10 microsecond neutrino beam window. The physics interactions are separated using an expanding, density-based clustering algorithm (DBSCAN) initially outlined in [2] which makes use of spacial and temporal information.

A local density is computed for each hit by counting the number of neighbors that are within a certain "distance" of that hit. The clusters that this algorithm makes have two types of points, core points and border points. Core points have at least the minimum number of neighbors within the critical distance. Border points have less than the minimum number of neighbors, but are allowed to be included in the cluster if and only if they are the neighbor of a core point. The algorithm makes clusters by looping over all hits and expanding clusters around identified core points until all hits are either classified in a cluster or left as noise.

Most of the particles passing through the NOvA detector are all traveling at or very close to the speed of light. Therefore, hits that originated with the same individual physics event (one neutrino or one cosmic ray) are those that are separated by light-like intervals in 4D space. To take advantage of this, the neighbor score function for two hits includes a causality term, and two terms to penalize hits that are far separated in space. It is defined as follows:

$$\text{neighbor score} = \left(\frac{|\Delta T| - |\Delta\vec{r}|/c}{T_{res}}\right)^2 + \left(\frac{\Delta Z}{D_{pen}}\right)^2 + \left(\frac{\Delta XY}{D_{pen}}\right)^2 \quad (1)$$

where T_{res} is the timing resolution of the two hits added in quadrature and D_{pen} is a distance penalty.

For hits that occur in the same view, the neighbor score is calculated as shown above with $\Delta\vec{r}$ being a two-dimensional quantity. For hits that are in opposite views, ΔXY is zero and D_{pen} in the ΔZ term is replaced with a smaller term, $OppViewPlPen$ (opposite view plane penalty) with $\Delta\vec{r}$ being one-dimensional. Two hits are considered to be neighbors if their neighbor score is less than an overall sensitivity parameter ϵ which fulfills the role of the critical "distance" in the causality neighborhood space.

The parameters in the neighbor score function were all tuned with a data-driven method to achieve maximum performance and tested by examining two primary metrics, completeness

$$\text{Completeness} = \frac{\text{Energy from interaction deposited in slice}}{\text{Total energy from interaction deposited in detector}}, \quad (2)$$

and purity

$$\text{Purity} = \frac{\text{Energy from interaction deposited in slice}}{\text{Total energy in slice}}. \tag{3}$$

In later stages of the analysis, quality cuts are placed to remove small slices, so performance evaluations focused on slices with a minimum of ten hits in each view. In Far Detector cosmic simulations, slicing was found to have a completeness and purity of 99.3%. For the Near Detector neutrino simulations, slicing had a purity of 98.5% and completeness of 94.4%, making pileup concerns from multiple neutrino interactions negligible.

3. Finding Feature Guidelines with a Multi-Hough Transform

The Hough transform is a relatively standard algorithm used for pattern recognition problems designed to identify major lines or features in a two-dimensional image composed of points or pixels. To improve the robustness against noise, the classic algorithm is modified to work on pairs of points according to [3]. This algorithm is applied to each individual detector view separately by taking pairs of hits in the event, calculating the line that passes through this pair in polar coordinates of ρ, the perpendicular distance from the origin to the line, and θ, the angle between ρ and the x-axis, and filling a parameter space map (Hough map) with a Gaussian smeared vote. The Gaussian vote is calculated according to

$$vote = e^{-\frac{(\rho-\rho_0)^2}{2\sigma_\rho^2}} e^{-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}} \tag{4}$$

where $\sigma_\rho = 3/\sqrt{12}$ and $\sigma_\theta = 3/d\sqrt{6}$ (d is the distance between the two hits in the detector) which are determined from the physical size of the NOvA cells. In this manner, major event features will cause a build up of votes in certain regions of the Hough map resulting in peaks that can be taken as the Hough lines that characterize the event.

Once the Hough space map is created, it is then smoothed by averaging hits in the map using a Gaussian smoothing weight. Peaks in the Hough space map are required to be above a threshold value in order to be labeled as a valid line. Since each event can be composed of a different number of hits in the detector, this threshold value is determined separately for each Hough map using the average bin height.

If all peaks above the threshold value are taken as legitimate lines, then the modified Hough transform described above has a strong tendency to produce too many lines. To prevent this problem from happening, an alternative approach was employed. After the dominant line is identified from the tallest peak in the Hough map, the hits associated with that line are removed from consideration and then the Hough map is generated again to look for smaller lines. With the hits associated with the first line removed, the dominant features in the new Hough map will now be less representative of noise and much more representative of shorter, legitimate, physics tracks. This "Multi-Hough" procedure is repeated until either no more peaks are found above the threshold value, or a preset maximum number of lines is reached. Care is taken to ensure that the same or very similar lines are not repeated.

The main performance criteria for the algorithm is for the dominant Hough lines to pass close to and form intersections near the primary interaction point of the slice. It was found in simulated Far Detector interactions that the primary Hough line passed within an average of 6.9 (NC), 4.1 (ν_μ CC), and 2.7 (ν_e CC) cm of the vertex. For the secondary Hough line the average distance is 9.9 (NC), 8.2 (ν_μ CC), and 8.8 (ν_e CC) cm.

4. Vertex Identification with Elastic Arms

The Elastic Arms algorithm is designed to use the output from the Multi-Hough algorithm as a seed for finding the global event vertex. It is based on a method of the same name listed in [5]

which is sometimes referred to in the literature as the "method of deformable templates" [4–7]. The basic template for a NOvA event is a vertex with one or more particle tracks emanating outwards from that vertex. In the Elastic Arms method, once a vertex has been identified, each particle track is approximated by an "arm" (a vector pointing away from the vertex) whose direction can be adjusted to fit the event. For the application of this method to NOvA data, the number of arms is taken to be the largest number of good Hough lines found for the event in either the XZ or YZ views. To determine the location of the vertex, a list of vertex candidates must be generated and evaluated.

The optimum vertex candidate is chosen by minimizing an energy cost function given by

$$E = \sum_{i=1}^{N}\sum_{a=1}^{M} V_{ia}M_{ia} + \lambda\sum_{i=1}^{N}\left(\sum_{a=1}^{M} V_{ia} - 1\right)^2 + \frac{2}{\lambda_v}\sum_{a=1}^{M} D_a \tag{5}$$

where M_{ia} measures distance between cell hit i and arm a, V_{ia} is the strength of association between hit i and arm a, D_a is a measure of the distance between the vertex and the first hit on arm a, and λ and λ_v control the strength of the terms. The first term measures the goodness of fit between the hits and the arms while the second is a penalty term for hits not associated with any arm. The third term is not present in the literature and is a penalty for arms whose first hit is far from the vertex. This term is necessary in the NOvA application where the vertex is not known and is tuned to the distance scale of photon conversions (λ_v = 30cm for NOvA) since as described above, a common occurrence in NC backgrounds is a π^0 decaying into a pair of photons. The hit/arm association term V_{ia} is given by

$$V_{ia} = \frac{e^{-\beta M_{ia}}}{e^{-\beta\lambda} + \sum_{b=1}^{M} e^{-\beta M_{ib}}} \tag{6}$$

where $e^{-\beta\lambda}$ represents the likelihood that the hit is noise and β can be interpreted as a range over which hits are allowed to be associated with arm a.

From the Hough lines, vertex candidates are formed from the intersection points of the major lines in each view. The hits are also sorted by their Z coordinates and additional vertex candidates are formed from selected hits at fixed intervals (2%, 5%, ... 50%) in this list. Note that this introduces a bias to favor vertices at lower values of Z but that this is exactly what one expects for beam neutrino events in the NOvA detectors.

To determine the arm directions, a list of possible vectors is generated from the directions of the Hough lines (matched between views by their peak heights in the Hough map) plus vectors formed from the vertices of a dodecahedron. The arms are then set one-by-one by choosing the direction from this list that minimizes Equation 5 before moving on to the next arm. Care is taken to ensure that the same or very similar arms are not reused for each vertex. Once vertex and arm seeds have been set, a process of simulated annealing is applied using MINUIT to allow the vertex to settle into an optimal location. This is accomplished by cooling β from high to low temperatures, allowing the vertex to smoothly seek out the global minimum of Equation 5 while avoiding potential local minima within that function.

Together, the Multi-Hough and Elastic Arms algorithms achieve average event vertex resolutions of 11.6, 10.9, and 28.8 cm for ν_μ CC, ν_e CC, and NC events respectively. For both the ν_μ CC and ν_e CC events, 68% of the vertices are within 10 cm of the true vertex (38 cm for the NC events.) This puts the vertex for the CC events within approximately 2 cell widths of the true vertex the majority of the time.

5. Prong Formation with Fuzzy k-Means

Once the global event vertex has been identified by the Elastic Arms method, the next step is assign a prong membership to each cell hit within the event, with each prong representing the

hits from a single particle track or shower. This is accomplished with a possibilistic fuzzy-k means algorithm [8, 9]. The term "possibilistic" means that the sum of each hit's membership across all prongs is not forced to be one, which allows for isolated hits to be treated as noise. The "fuzziness" allows a hit to belong to more then one prong and the number of prongs is unknown at the start.

The philosophy behind this method is that when sitting at the global event vertex for either view, the cell hits within the slice should appear as peaks of deposited energy in a one-dimensional angular space around that vertex. The Fuzzy-K algorithm determines how many prong centers (peaks) are present and assigns a prong membership to the hits in the slice. Each hit is converted to an angle with respect to the vertex ranging from -π to π with 0 corresponding to the Z- axis of the detector, and assigned an angular uncertainty as a function of distance from the vertex. This uncertainty was determined empirically from a simulated sample of 1 and 2 GeV muons and electrons to incorporate the effects of multiple scattering.

Finding the prong centers and determining the membership for each hit is done with an iterative process. It begins by assuming there is only one prong centered on the region of highest density in the one-dimensional hit angular space. Then a degree of membership for each hit θ_j in each prong center θ_i is computed according to

$$U_{ij} = e^{\frac{m\sqrt{c}d_{ij}}{\beta}}, \tag{7}$$

where d_{ij} is the distance to prong centers given by

$$d_{ij} = \left(\frac{\theta_j - \theta_i}{\sigma_j}\right)^2. \tag{8}$$

Here σ_j is the angular uncertainty, m represents the "fuzziness" factor which allows hits to retain partial membership in multiple prongs, c is the number of prong centers, and β can be interpreted as a normalization term that represents how spread out the hits are expected to be around a "normal" prong center. The prong centers are then updated to θ_i' according to

$$\theta_i' = \theta_i + \frac{\sum_{j=1}^{n} \mathcal{A}_{ij}\,(\theta_j - \theta_i)}{\sum_{j=1}^{n} \mathcal{A}_{ij}} \tag{9}$$

where $\mathcal{A}_{ij} = \frac{U_{ij}^m}{\sigma_j^2}$ and the process is repeated until $\Delta\theta = |\theta_i' - \theta_i|$ is below tolerance for all centers.

Duplicate prong centers are removed and then a new one-dimensional distribution of cell hits in angular space is computed from all hits with less than 1% membership in any cluster. This distribution is used to identify the next most dense peak which is then added to the list of cluster center seeds. The next iteration of finding cluster centers is started with this new cluster center added to the list. This iterative process is repeated until either all cells belong to at least one cluster or until a maximum number of clusters is reached.

Since this prong formation process is done separately for each view, the last step is to attempt to match clusters between the views. To do this, a temporary track is formed out of every pair of two-dimensional clusters from each view. For each of these possible pairings, the cumulative cell hit energy distribution (normalized to the total energy for each cluster) is computed for each view as a function of distance along the track. Comparing this distribution between the views allows for the application of a Kupier metric which takes the sum of the largest absolute positive and negative vertical distances between the two distributions as a metric for each of these possible cluster combinations. The view matching proceeds by pairing together the clusters from each view best matched by this metric, and then continuing in this fashion until all clusters that can be matched have been matched.

The metric used to assess the performance is completeness for hits produced by the primary lepton in CC interactions. For ν_e CC events, the average completeness is 88%, 95% for QE events and 86% for non-QE events. For ν_μ CC events, these numbers are 93%, 98%, and 92% respectively.

6. ANN Event Classification

NOvA has a variety of ν_e event identifiers that use different methods of condensing the vast amount of information available about the neutrino interaction into a few manageable variables. One such method is LID, or Likelihood-base particle ID, that depends on the reconstruction chain described in the previous sections. Every particle deposits energy along it's trajectory in a characteristic manner, and LID uses these energy depositions per path length or dE/dx to identify the particles that produce tracks or showers in the NOvA detectors.

The inputs to LID are Fuzzy-K prongs. The most energetic prong in the neutrino interaction is picked and split into two views: the longitudinal view along the length of the shower, and a transverse view, perpendicular to the direction of the shower. In the longitudinal direction, the dE/dx is computed and recorded for each detector plane that the particle passes through. The transverse direction measures how concentrated the energy deposition is in the core of the shower. Therefore, the energy deposition is recorded radially, as a function of distance from the shower core, summed over the entire length of the shower.

Template histograms of the expected plane-by-plane dE/dx and transverse-view energy depositions are created for various particle hypotheses using simulated neutrinos interactions. These histograms are normalized to form probability distributions. The particle hypotheses that LID is currently capable of handling are electron, muon, charged pions, neutral pions, neutrons, protons and photons.

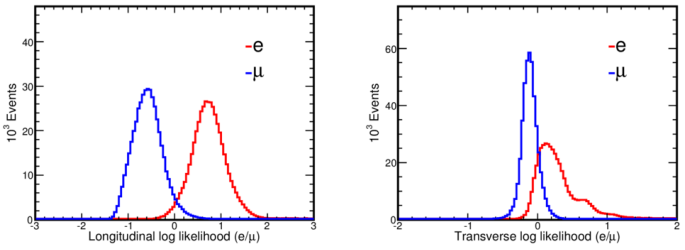

Figure 3. Log-likelihood difference $LL(e^-) - LL(\mu^-)$ for true electrons and true muons in longitudinal and transverse views. The Difference is a positive quantity for true electrons and negative for true muons.

The pattern of energy deposition of a prong of unknown identity is compared with these template histograms. For instance, if an unknown particle deposits energy E_i in the i^{th} plane from the start of the prong, the probability, P_i of various particle hypotheses for having in this manner is read from the template histograms for the i^{th} plane. The logarithms of these probabilities are then summed over all planes in the prong and used to construct a likelihood for that particle hypothesis:

$$LL(e^-) = \sum_i^N log\left(P_i(e^-)\right) \tag{10}$$

where $LL(e^-)$ is the likelihood of the unknown particle being an electron. The difference in the log-likelihoods indicates the identity of the particle as shown in Fig.3

These likelihood differences are used as an input variable to the ANN in ROOT's TMVA package. Note that these likelihoods are only computed and used for the most-energetic prong in the event. Additionally we use other event-level topological variables to separate ν_e CC signal interactions from the backgrounds. Some of the other variables include fraction of event energy contained in the most-energetic prong, angle of the most-energetic prong with respect to the neutrino beam, distance of the shower from the event vertex, number of minimum-ionizing planes at the start of the prong, reconstructed invariant mass with secondary prongs (for neutral pion rejection) and total energy close to the interaction vertex. The performance of LID in separating ν_e CC signal from other beam background can be seen in Fig.4

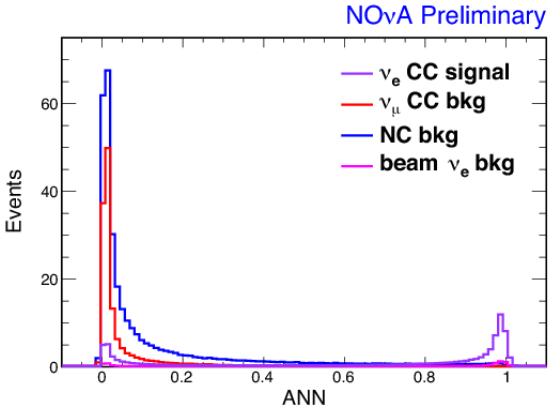

Figure 4. The performance of the likelihood-based PID on NOvA simulation.

7. Conclusions

The NOvA detectors are fine-grained detectors for their size. A unique reconstruction and particle-identification chain that takes advantage of this granularity has been developed to meet the goals of the ν_e appearance oscillation analysis.

Acknowledgements

The authors acknowledge that support for this research was carried out by the Fermilab scientific and technical staff. Fermilab is operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy.

References

[1] D.S. Ayers et al. The NOvA Technical Design Report. 2007.
[2] M. Ester and H. Kriegel and J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise.", Kdd **96**, 226-231 (1996).
[3] L. Fernandes and M. Oliveira, "Real-time line detection through an improved Hough transform voting scheme", Pattern Recognition **41**, 299 - 314 (2008).
[4] M. Gyulassy and M. Harlander, "Elastic Tracking and Neural Network Algorithms for Complex Pattern Recognition", Computer Physics Communications **66**, 31 - 46 (1991).
[5] M. Ohlsson, and C. Peterson, "Track Finding with Deformable Templates - The Elastic Arms Approach", Computer Physics Communications **71**, 77 - 98 (1992).
[6] M. Ohlsson, "Extensions and Explorations of the Elastic Arms Algorithm", Computer Physics Communications **77**, 19 - 32 (1993).
[7] R. Frühwirth and A. Strandlie, "Track Fitting with Ambiguities and Noise: A Study of Elastic Tracking and Non-linear Filters", Computer Physics Communications **120**, 197 - 214 (1999).
[8] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst. 1 (1993) 98110.
[9] M. S. Yang, K. L. Wu, Unsupervised possibilistic clustering, Pattern Recognition, 39 (2006), pp. 521.