**THE EUROPEAN
PHYSICAL JOURNAL C**

Special Article - Tools for Experiment and Theory

Implementation of a cellular automaton algorithm for neutrino interaction reconstruction in a liquid argon volume

J.J. Back, G.J. Barker, A.J. Bennieston, S.B. Boyd, B. Morgan, Y.A. Ramachers[a]

Dept. of Physics, University of Warwick, Coventry CV4 7AL, UK

Abstract A Cellular Automaton algorithm has been implemented in three dimensions for automated track reconstruction of neutrino interactions in a Liquid Argon Time Projection Chamber. We present details of the algorithm and characterise its performance on simulated data sets.

1 Introduction

Liquid Argon Time Projection Chambers (LAr-TPCs) are recognised as being a well-matched technology choice to meet the physics demands of a next-generation neutrino oscillation experiment. Their development is the subject of vibrant R&D programmes in Europe, Japan and the USA (see e.g. [1–3] and references therein).

The ability to achieve simultaneous tracking and calorimetry, with millimetric granularity, makes LAr-TPCs ideal to image neutrino interactions over a wide range of energies and allows us to identify the interaction components with robust particle identification as demonstrated by recent results from ICARUS [4]. However, a fully automated software event reconstruction in LAr has proven difficult to achieve. The mixture of ionisation tracks, electromagnetic and hadronic showers in the data set, in addition to not knowing a priori the neutrino interaction point, makes the algorithmic reconstruction of events a challenging task.

An active worldwide R&D programme has resulted in the construction of increasingly larger LAr-TPCs and has brought the issue of automated reconstruction to the fore. Recent developments in this area can be found in Refs. [5–7]. In this paper, we describe a first application of cellular automaton techniques to three-dimensional LAr-TPC data. In what follows we first describe the algorithm in Sect. 2.2 before going on to characterise its performance based on simulated neutrino interaction events in Sect. 3. We conclude in Sect. 4.

[a] e-mail: y.a.ramachers@warwick.ac.uk

2 Cellular automaton reconstruction in liquid argon

2.1 Motivation

The passage of ionising particles through LAr leaves a stream of released ionisation charge. The function of a LAr-TPC is to read out this charge, associating it with a 3D position in space and a corresponding measurement uncertainty. Therefore, The data can be considered as a spatial cell or 'voxel' linked with a charge deposit i.e. $(\Delta x, \Delta y, \Delta z; Q)$ and this vector of information we will call a 'hit'. The function of any reconstruction algorithm is split into three main tasks: (1) hit reconstruction; (2) the association of hits deriving from a common origin into hit-clusters; (3) the extraction of physics parameters from the clusters such as fitting to a model in order to extract the parameters of a particle trajectory. The aim of this study is to investigate an algorithm capable of making the association of hits into clusters with high efficiency. The first stage of the reconstruction chain, hit reconstruction, was therefore assumed to have already taken place and the input to the clustering stage was a set of hits in three spatial dimensions.

A primary motivation for investigating a cellular automaton (CA) approach to this application is the close match between the inherently cellular nature of the data and the way in which CA algorithms function. Cellular automata are dynamical systems which evolve in a space divided into a number of discrete cells, which in turn can be associated with as many dimensions as the application demands. For a reconstruction task, this definition slightly changes by incorporating the context of voxels or hits in a particle physics detector. The *cell definition* is changed from its original pixel-type nature to that of a connection between two hits (see Fig. 1), providing additional directional information. Specific features of CA algorithms relevant for our application include:

– The CA cells can take on several states which change by the application of a fixed rule relating the cell to its nearest

neighbours. This makes the approach explicitly *local* and independent of any prior assumption regarding the particle trajectories expected (prior knowledge of the expected event topology is an advantage that collider experiments have over a LAr-TPC exposed to a neutrino beam).

– The cell states change by iteratively stepping through the entire LAr volume of cells. This makes the algorithm relatively insensitive to the fact that the primary neutrino interaction point is initially unknown and its location cannot be assumed.

– CA algorithms are extremely simple, fast and naturally lend themselves to parallel processing.

Cellular automata have been used with some success before as part of track reconstruction strategies for HERA-B [8], NEMO-2 [9] and SciBar [10]. These previous applications have all involved relatively coarse-grained detectors and have all been inherently two-dimensional. We believe the present study is the first time a CA approach has been applied to a reconstruction task in 3D and in such a fine-grained environment as LAr.

2.2 The cellular automaton algorithm

Notions of 'forward' and 'backward' are central to the functionality of the algorithm and are defined to be with respect to a *run coordinate*. With a run coordinate along the x-axis, for example, a hit at location $(0, 1, 2)$ is backward of the hit at $(1, 1, 2)$ in the sense that its x-coordinate value is smaller. The algorithm scans the entire LAr volume along the run coordinate direction. A potential problem arises with particle trajectories that are perpendicular to the run coordinate, since in this case the definition of forward/backward are impossible to resolve. In our example, if the y- or z-axes were picked as the run coordinate, it would not be possible to say that one is backward of another since they share y- and z-coordinates. Therefore, it is important to the efficiency of the algorithm to pick the run coordinate that minimises hits piling-up into a single bin. This is achieved by choosing the run coordinate to be a measure of the overall direction of the event, and is calculated by running a principal component transformation (PCT) [11] on the hits and assigning the run coordinate to be the principal component axis.

The algorithm consists of three straightforward steps: (1) state initialisation, (2) finite state evolution and (3) post-processing.

The first stage of state initialisation involves the generation of an initial set of cells. This is done on a per-hit basis by finding all neighbouring hits within a given radius (equal to or larger than the smoothing radius) and filtering to select only those which are backward of the current central hit. If no hits exist in the backward direction within this *search radius*, the radius is expanded in 0.15 mm steps up to a maximum value, stopping as soon as a backward hit is encountered. Cells are then made by connecting the central hit to each backward hit found. If multiple cells are generated, those with the same slope are filtered out in favour of the cell with the shortest hit-to-hit distance. This prevents the CA from jumping over hits and producing multiple interleaved tracks.

Once a set of cells has been produced, the CA makes a forward step onto the next hit and the procedure repeats. This continues until no cell updates are made. At this stage, each cell found is assigned a weight, initially set to one, as illustrated in Fig. 1a. This weight corresponds to the initial state of each cell. The full cellular automaton has reached its initial state and the evolution process can be started.

Step (2) of the algorithm requires a prescription for automaton evolution where the state change rule consists of the following: each cell is tested in turn to find if there exists a *neighbour* cell in the backward direction. Neighbour cells share an end-point with the current cell and the angle made between the two cells (defined as the angle between straight-
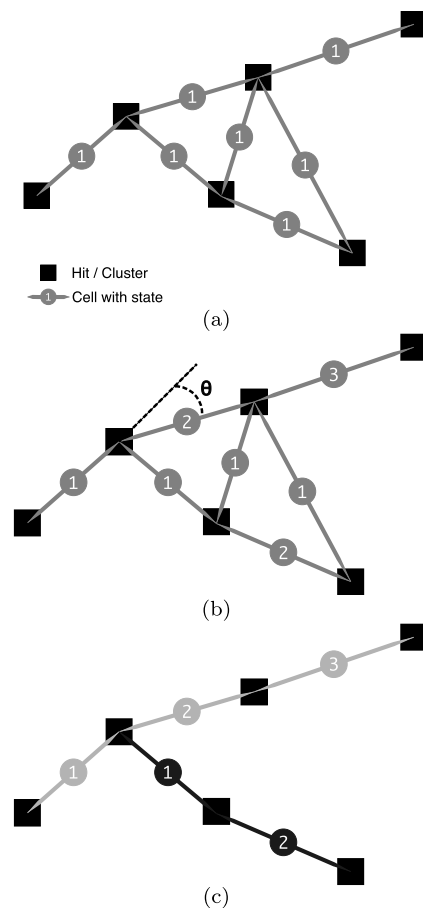

Fig. 1 The operation of the CA: (**a**) cells generated between space–points with an initial weight of one; (**b**) cell states after the forward run of the CA showing the definition of the CA angle; (**c**) the final result of the CA showing two clusters corresponding to linear tracks. Cells of weight one which are not the endpoint of a longer cluster are removed

line segments passing through the start and end hits of each cell) must be less than a specified angle (the *CA angle*).

The current cell is marked for a potential update if it has one or more neighbouring cells with the same weight. After forward-scanning through all cells, which could in principle be performed in parallel, cells marked for update have their weight increased by one. This procedure continues iteratively until no updates occur in a given forward scan, at which stage the state of the cell network is as illustrated in Fig. 1(b).

After this automaton evolution process, a reverse scan is performed in order to build cluster candidates. This stage is the post-processing, step (3), of the automaton and finishes the algorithm.

In a reverse scan, the highest-valued cell in the event is chosen as the starting point. If multiple cells have the same (high) weight, one is chosen at random. This starting cell is added to a new cluster candidate. The backward neighbours are then analysed to find a cell with a weight one less than the current cell. If multiple candidates exist, the one that makes the smallest angle with the current cell is chosen. This cell is added to the cluster candidate and set as the current cell for the next iteration. This process is repeated until a cell of weight one is encountered, at which point the reverse step is complete and a cluster has been built. The cells used to make this cluster are removed from the input set and the reverse scan algorithm is run again until the input cell set is empty. On completion of the reverse scan, the CA has isolated a set of clusters containing one or more cells as illustrated in Fig. 1c.

Clusters which contain only one cell are removed, since these isolated cells are dominated by noise. In addition, all clusters are examined for hits that are assigned to more than one cluster. Although cells have been uniquely assigned by the CA, hits can exist in multiple cells, which can in turn enter into multiple clusters. This situation is resolved by uniquely assigning a hit to the longest cluster in which it appears.

The output of the CA at this stage is comprised of a set of clusters, some of which may correspond to segments of a larger underlying physics object. This situation arises when the local propagation of the CA algorithm is interrupted by the presence of features such as delta electrons, bremsstrahlung or local kinks from multiple scattering in liquid argon. A stitching process is employed to recognise such localised breaks and recombine cluster segments that are likely to have originated from a common source. All pairs of segments are checked for consistency in space and in angle. To do this, the last five hits at either end of each segment are used to determine the mean position of the hits and a direction (based on a regression line using principal component analysis). Segments with end-points consistent with originating from the same point in space and which share

the same direction, to within some critical *stitching angle*, are merged. Clusters with less than five hits are discarded from further consideration.

3 Performance on simulated neutrino interactions

In order to test the algorithm on physics events, the Geant4 toolkit [13] was used to model a LAr-TPC as a stainless steel cylinder of 10 m radius and 10 m height filled with natural argon in the liquid state. Particles were tracked through the volume with interactions modelled by the QGSP_BIC_HP physics list with all electromagnetic and hadronic processes enabled. The LAr volume was divided into $(1 \times 1 \times 1)$ mm^3 voxels and all primary and secondary particles were tracked through voxels down to zero energy or until they left the TPC volume. Energy deposits by charged particles passing through the voxels were tallied into a map between voxel coordinates (x, y, z) and the total energy deposited E. No attempt was made to model the readout system or hit/voxel reconstruction from raw data as this is highly experiment-specific.

The GENIE [12] package was used to model muon neutrino interactions with a monoenergetic ν_μ spectrum at an energy of 0.77 GeV (corresponding to the JPARC neutrino beam mean energy). The neutrinos were directed in a beam along the x-axis through the centre of the TPC vessel.

LAr-TPC data was simulated by summing the ionisation charge released by particles traversing a LAr volume into 'voxels' of size $(1 \times 1 \times 1)$ mm^3. These voxels represent a hit-granularity finer than any planned large-scale LAr-TPC and so a charge-smoothing procedure was adopted in order to simulate the diffusion effects over any drift length. This was achieved by computing the charge-weighted average position of hits inside a sphere of a configurable radius. For this study, a *smoothing radius* of 6 mm was chosen to broadly match that expected for a large detector with drift distances in excess of 10 m.

3.1 Charged-current quasi-elastic interactions: $\nu_\mu + n \rightarrow \mu + p$

The performance of the CA on Charged-Current Quasi-Elastic (CCQE) events is of interest due to the importance of the CCQE channel to the long-baseline neutrino oscillation programme. This event class appears in the detector as a track-pair comprising one long track (the muon) and one short track (the proton), with variable opening angle. The proton can create a particle shower and the muon can decay to leave a Michel electron.

To determine the optimal choice of CA algorithm parameters for CCQE event reconstruction, we studied the effect of varying the CA parameter values on the efficiency for collecting the hits of final state particles in a single cluster with

Table 1 Summary of the optimal reconstruction parameter set as determined for CCQE events. See the text for a full description of parameters

Smoothing radius	CA angle	CA search radius/max. radius	Stitching angle	Stitching n end-points	Minimum cluster hit count
6 mm	20 degree	3 mm/12 mm	30 degree	5	5
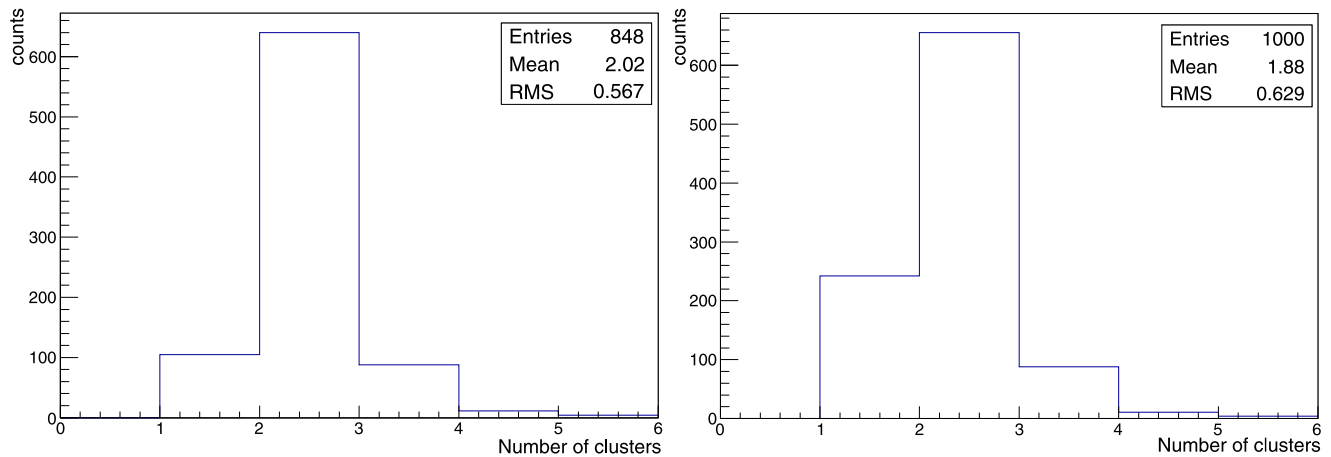
Fig. 2 The number of reconstructed clusters in a sample of 1000 CCQE events (*Left*) for events passing a range cut on the proton trajectory and (*Right*) with no range cut imposed

a high hit purity. For any generated particle, we define the hit efficiency to be the ratio of correct hits in the cluster compared to all hits produced by the particle, and hit purity to be the ratio of correct cluster hits compared to all hits in the cluster definition. The results showed that the performance was sensitive to the choice of the CA angle and, to a lesser extent, the stitching angle. Since reconstruction efficiency and hit purity are anti-correlated, the optimal choice of algorithm parameters is somewhat dependent on the needs of individual applications. In order to arrive at a set of 'baseline' parameters, the product of hit efficiency and purity was maximised, and it was found that this could be achieved over a rather wide range of reasonable parameter values. Based on 1000 $\nu_\mu + n \to \mu + p$ simulated events, the resulting set of optimised parameter values is listed in Table 1.

Figure 2 presents distributions of the number of clusters reconstructed in these events with (left histogram) and without (right histogram) a filter on the range of the proton track. The range cut required each proton to have deposited at least 25 hits corresponding to a maximum range of 2.5 cm, roughly equivalent to an energy threshold of 10 MeV for a minimum ionizing particle in liquid argon. With the range cut in place, the number of single clusters reconstructed (when at least two are expected) is significantly reduced, highlighting the fact that only trajectories long enough to be considered distinct track features can be efficiently reconstructed.

Of the events which passed the proton-range cut (84.8 %), the majority of reconstructed events (76 %) consisted of two distinct clusters, one associated with the muon and one

with the proton. Twelve percent were reconstructed as single clusters, while three or more clusters were reconstructed in 12 % of cases.

A close inspection of individual event displays revealed that the multi-cluster reconstructions occurred mostly in events where the proton induces a hadronic shower with fragments large enough to pass all cuts and qualify as clusters, together with instances of muon decay to Michel electrons, as illustrated in Fig. 3(a). Figure 3(b) shows an example of an event reconstructed with five distinct clusters resulting from the muon subsequently decaying to an electron, and the partner proton interacting to give two further protons.

All single cluster solutions in the sample were examined for possible failures of the algorithm, and the following three situations were seen to account for these cases: (a) Following the hit smearing procedure, the proton track was too short to be recognised (44 % of cases). This is possible for proton trajectories if the (more than 25) hits are tightly packed in the voxel grid resulting in the smearing producing a proton 'stub' that is short enough to be effectively invisible to the reconstruction. (b) Proton–muon trajectories appearing with a large opening angle (52 % of the time). This back-to-back topology will be reconstructed as a single track by any reconstruction algorithm unless the two clusters can be separated by additional external information—such as the location of the interaction vertex position or use of dE/dx information. (c) In a few cases (4 %), the proton produces a hadron shower immediately without leaving a track and the resulting individual shower clusters are
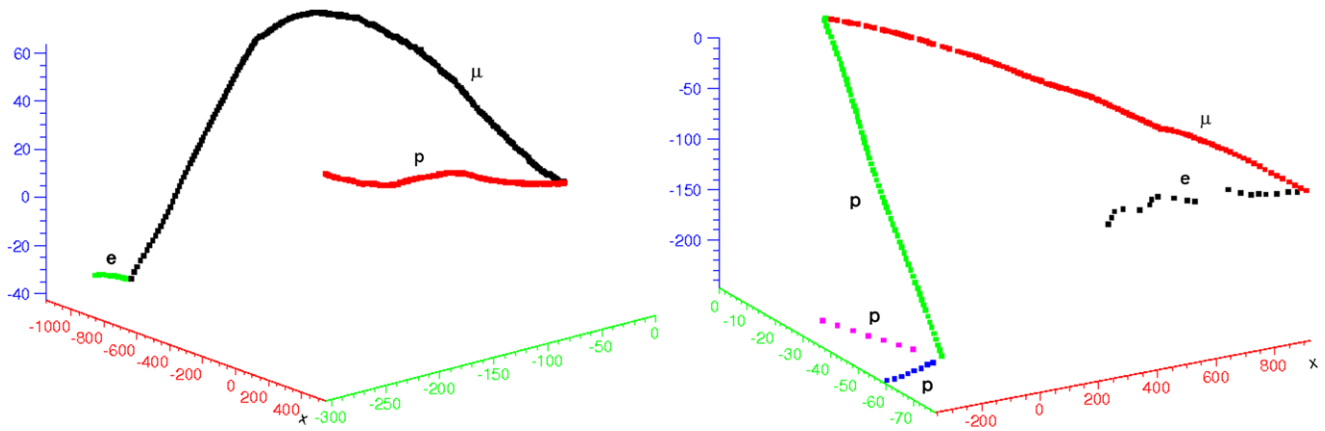
Fig. 3 CA reconstructed clusters of hits in events from the CCQE sample. The units on the axes are in mm. Hits associated to the same cluster have the same colour. (**a**) The CA has correctly reconstructed the muon decay electron. (**b**) The CA has identified all five-clusters resulting from the muon decay electron and a hadronic interaction of the final state proton

Table 2 Mean hit efficiencies and hit purities for muon and proton reconstruction under several conditions. Abbreviations: 'Ncls'—number of reconstructed clusters and 'proton cut'—proton range cut using Monte-Carlo truth information

Condition with proton cut	Muon efficiency [%]	Muon purity [%]	Proton efficiency [%]	Proton purity [%]
Ncls = 2	98.8	92.4	99.0	95.9
Ncls > 1	92.7	92.3	98.8	95.5
All Ncls	93.1	91.7	98.8	95.1

too small (number of hits below cutoff) to count as clusters themselves—so leaving the muon as the single cluster.

The conclusion of the cluster study is that the CA algorithm reliably reconstructs the multiplicity of track structures that exist in the underlying physics event. The reconstruction results were, in addition, analysed in terms of the cluster hit efficiency and hit purity. Events were selected where there were exactly two reconstructed clusters, and the simulation truth information was used to identify each cluster as either a muon or proton. The results are presented in Table 2. The high hit efficiencies mean that very few muon (or proton) hits are lost to other clusters in the reconstruction process and, likewise, the high hit purities reflect that there is little 'contamination' with hits from other objects in the event. Hit contamination, when it occurred, was found to be due to secondary particles—mostly delta electrons.

This clean data sample of two reconstructed clusters was inspected closely for failures of the algorithm. The muon cluster hit efficiency was seen to be less than 100 % in only 0.7 % of cases. The majority of these were found to be where the algorithm split the long muon track into two or more segments, usually corresponding to kinks in the muon trajectory. Occasionally, the algorithm can fail to associate hits at the end of a muon trajectory where excessive straggling of the ionisation deposits has occurred. This level of failure was acceptable to the current study, but could have been eliminated further with a tighter post-processing stitch-

ing together of clusters—albeit at the expense of producing more single-cluster solutions. Finally, in 2 % of cases it was found that a large part of the muon and proton hits were combined into the same reconstructed cluster. When this happens, the second reconstructed cluster was found to be either a fragment of the full muon track or made up of hits from a delta-electron. It is worth noting that combined proton-muon clusters could still, in principle, be disentangled by using the ionisation loss information, but this has not been implemented in the current study.

3.2 Charged-current single-pion production: $\nu_\mu + p \rightarrow \mu + p + \pi^+$

In order to test the algorithm on more complex event topologies, we ran on a sample of charged current events where the excitation of a baryon resonance and its subsequent decay ejects a pion to join the muon and proton in the final state. This reaction becomes comparable in rate to the CCQE process for neutrino energies above about 2 GeV.

The simplest event structure would feature exactly three ionisation tracks. However, pion decays in addition to all secondary particles which can emerge from muons and protons, mean that the final state often contains more than three physics objects to reconstruct. An example event reconstruction is shown in Fig. 4 (left), where the algorithm correctly

Table 3 Mean efficiencies and purities for muon, proton and pion reconstruction, respectively, under two selection conditions. Abbreviations: 'Ncls'—number of clusters and 'proton cut'—proton range cut using Monte-Carlo truth information

Condition with proton cut	Muon efficiency [%]	Muon purity [%]	Proton efficiency [%]	Proton purity [%]	Pion efficiency [%]	Pion purity [%]
Ncls > 2	96.9	86.9	98.3	90.1	92.6	85.5
All Ncls	97.3	85.6	98.5	89.0	93.1	84.1
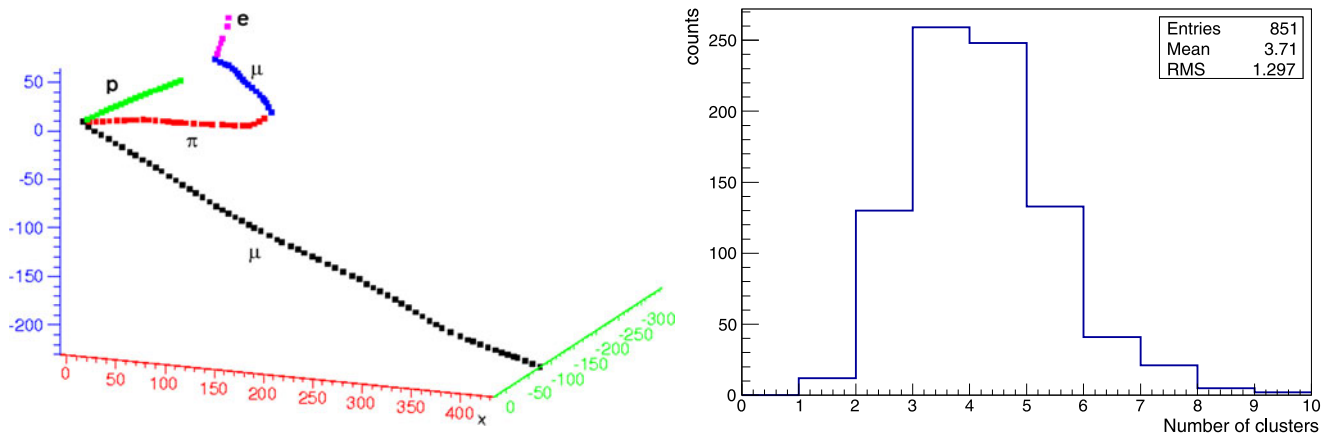
Fig. 4 (*Left*) CA reconstructed clusters of hits for an event from the $\nu_\mu + p \to \mu + p + \pi^+$ sample. The units on the axes are in mm. (*Right*) The number of reconstructed clusters for a total of 851 $\nu_\mu + p \to \mu + p + \pi^+$ events passing the proton range cut

found all five clusters resulting from the final state particles and the decay products of the pion.

Figure 4 (right) shows the distribution of reconstructed clusters resulting from running the algorithm on a sample of 1000 events. A total of 851 events passed the same proton range cut as was applied to the CCQE study. As expected, a substantial tail due to more than 3 reconstructed clusters is seen, and only 1.5 % of events are mis-reconstructed as a single cluster event.

The resulting hit efficiency and purity values for the muon, proton and charged pion objects in the final state are reported in Table 3. The algorithm is found to handle the extra complication in topology well, and high reconstruction efficiencies and purities are observed.

3.3 CA algorithm used as a track-shower discriminator

Correctly clustering hits from electromagnetic or hadronic shower objects is a challenging reconstruction task. Our development of the CA algorithm concentrated on track-like structures, but application of the algorithm to showers has been found to provide an efficient discriminator between track and shower objects. This could be a valuable addition to an overall reconstruction strategy for events in LAr where, for example, an initial decision on the event object type is made before applying further algorithms specialised in either track or shower reconstruction.
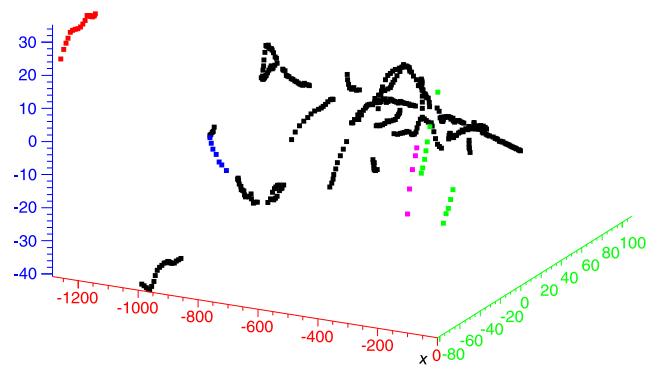

Fig. 5 Example of a CA processed electron shower from a single 1 GeV energy electron, started in liquid argon with event vertex at the origin of the coordinate system to the right of the picture. The units on the axes are in mm. Several clusters (colours) are reconstructed by the CA with the largest, core, cluster shown in *black*

Figure 5 illustrates the typical response of the CA applied to a 1 GeV isolated electron electromagnetic shower. The algorithm frequently reconstructs the core of the shower as a single cluster, surrounded by a number of smaller, satellite, clusters. The core contains enough hit information to provide an estimate of the shower direction and often includes the hits associated with the onset of the shower.

We have used this information to form three discriminating variables that are combined into a multivariate discriminator of tracks and showers in a procedure that follows our
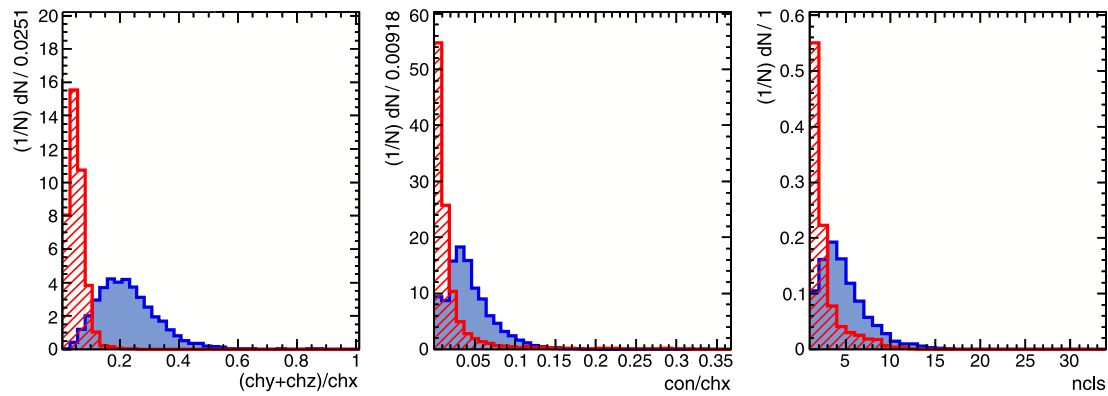
Fig. 6 The discriminating variables showing the individual separation achieved between the electron (*blue*) and muon (*hatched red*) samples. The variable '(chy+chz)/chx' labels the convex hull variables as dis-cussed in the text; 'con/chx' is the Coulomb energy based variable and 'ncls' labels the number of clusters variable (Color figure online)

earlier publication [14]. This technique is seen to be largely energy independent when validated with muon and electron events.

The first discrimination variable uses a convex hull [15] to define the extent of the core hit cluster from the CA processing, and forms the ratio of transverse to longitudinal extent of the convex hull. The hit cluster is first transformed to align with the main principal component axis, which then defines the longitudinal axis. The second variable is built as the ratio of the Coulomb energy to the longitudinal axis extent, where the Coulomb energy is defined as the sum of the amount of charge in each hit divided by its distance to the main principal component axis of the structure. The third variable used is simply the number of clusters produced by the CA. It is this third variable which renders the discrimination technique energy independent. While the first two variables are very effective at energies of 1 GeV and higher for muons and electrons, the low energy samples are only effectively separated when including the number of clusters. Electron showers are found to be far more likely to be broken up in separate clusters by the CA at low energies than muon tracks while the geometry variables yield similar results (extent and spread) for both.

The set of three variables are used to build a Boosted Decision Tree (BDT) based on the TMVA package [16]. To train and validate the BDT, showers from electrons at energies from 10 MeV to 3 GeV (signal) and muons in the same energy interval (background) were simulated (see Fig. 6). The optimal cut on the BDT is defined to be that which maximises the ratio of signal-to-total event rate. The BDT analysis gives 97.3 % as signal efficiency with 3.7 % contamination over all energies.

4 Conclusion

We have presented a cellular automaton approach to hit association (clustering) in three spatial dimensions applied to neutrino interaction events in liquid argon. These preliminary findings have confirmed that the algorithm is generally well-matched to the high-density cellular nature of LAr-TPC data sets.

For this study, operational parameters of the algorithm were optimised to maximise the product of the hit efficiency and hit purity of the resulting reconstructed clusters, but it is expected that users of the algorithm will tune its performance according to their specific application.

The focus of this study has been the reconstruction of track-type objects and the algorithm was shown to work efficiently on the two most important event classes at low energies: charged-current quasi elastic events and charged current single-pion production.

This work has spun out of studies whose ultimate goal is the fully automatic reconstruction of neutrino interaction event kinematics, together with a particle identification decision made on each object reconstructed. To this end, we have previously published results on finding key features in LAr interactions [17] and on particle identification of showers [14]. This study has shown that the CA approach can be used as the basis for an efficient track-shower discriminator, which can provide an initial classification of a reconstructed hit cluster as either a shower or track object so that specialised shower or track algorithms can be subsequently applied.

References

1. A. Rubbia, J. Phys. Conf. Ser. **308**, 012030 (2011)
2. B. Fleming, J. Phys. Conf. Ser. **308**, 012007 (2011)
3. M. Tanaka, J. Phys. Conf. Ser. **308**, 012022 (2011)
4. C. Rubbia et al., J. Instrum. **6**, P07011 (2011)
5. G.J. Barker, J. Phys. Conf. Ser. **308**, 012015 (2011)
6. C. Bromberg, J. Phys. Conf. Ser. **308**, 012014 (2011)
7. A. Badertscher et al., J. Instrum. **8**, P04012 (2013)
8. I. Abt, D. Emeliyanov, I. Gorbounov, I. Kisel, Nucl. Instrum. Methods A **490**, 546 (2002)
9. I. Kisel et al., Nucl. Instrum. Methods A **387**, 433 (1997)
10. H. Maesaka, Evidence for a muon neutrino oscillation in an accelerator based experiment, Kyoto University Thesis, 2005
11. I.T. Jolliffe, *Principal Component Analysis*, 2nd edn. (Springer, New York, 2002)
12. C. Andreopoulos, Nucl. Instrum. Methods A **614**, 87 (2010)
13. S. Agostinelli et al., Nucl. Instrum. Methods A **506**, 250 (2003)
14. J.J. Back et al., Eur. Phys. J. C **73**, 2369 (2013)
15. C.B. Barber, D.P. Dobkin, H.T. Huhdanpaa, ACM Trans. Math. Softw. **22**, 469 (1996)
16. P. Speckmayer, A. Hoecker, J. Stelzer, H. Voss, J. Phys. Conf. Ser. **219**, 032057 (2010)
17. B. Morgan, J. Instrum. **5**, P07006 (2010)