A Neural Network Approach to Muon Triggering in ATLAS

Ran Livneh

January 2007

CERN-THESIS-2007-029
19/03/2007

A Neural Network Approach to Muon Triggering in ATLAS

Presented By

Ran Livneh

Submitted in total fulfilment of the requirements of the degree of

Master of Science

Supervisor:
Prof. Erez Etzion

School of Physics and Astronomy
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel-Aviv University

December 2006

Abstract

The extremely high rate of events that will be produced in the future Large Hadron Collider requires the triggering mechanism to make precise decisions in a few nano-seconds. This poses a complicated inverse problem, arising from the inhomogeneous nature of the magnetic fields in ATLAS. This thesis presents a study of an application of Artificial Neural Networks to the muon triggering problem in the ATLAS end-cap. A comparison with realistic results from the ATLAS first level trigger simulation was in favour of the neural network, but this is mainly due to superior resolution available off-line. Other options for applying a neural network to this problem are discussed.

Table of Contents

Figures

Acknowledgments

I would like to thank Prof. Halina Abramowicz for giving me this unique opportunity to join the research program of the Tel Aviv University high-energy physics group.

I am deeply indebted to Prof. Erez Etzion for guiding me throughout the study, standing by my side, and putting up with all of my quirks.

I would like to thank Prof. David Horn and Dr. Gideon Dror for their part in the study, and for their notes and sources. I would like to thank Dr. Lorne Levinson for his helpful advice and ideas.

Finally, I would like to thank my friends and colleagues from the laboratory: Dr. Yan Benhammou, Jonathan Ginzburg and others for their professional and especially warm and personal support. I also very much appreciate the support given to me by Dr. Iuliana Cohen and Dr. Gideon Bella.

Publications

1. "Momentum reconstruction of particles in the forward muon trigger system of the ATLAS detector", April 1999, Gideon Dror, Erez Etzion, David Horn, Presented at *AIHENP'99*.

2. "Momentum reconstruction and triggering in the ATLAS detector", November 2000, Gideon Dror, Erez Etzion. AIP Conference proceedings 2001, p. 67 (*TAUP-2647* [hep-ex/0011061]).

3. "Using a neural network approach for muon reconstruction and triggering", February 2004, E. Etzion, H. Abramowicz, Y. Benhammou, G. Dror, D. Horn, L. Levinson and R. Livneh. (*Nucl. Instrum. Meth. A534 (2004) 222-227*).

Introduction

The Large Hadron Collider (LHC) is a new proton-proton collider currently under construction at the European Organization for Nuclear Research (CERN). The LHC is to be installed in an existing tunnel, which previously held the Large Electron Positron (LEP) collider.

The Tel-Aviv University (TAU) High-Energy Physics (HEP) group is a collaborator in ATLAS, one of the 4 LHC experiments in development. An Israeli collaboration of the Weizmann Institute, the Technion and the TAU HEP groups manufacture and test the Thin Gap Chamber muon detectors, a part of the ATLAS muon spectrometer. This muon spectrometer will supply high-precision muon momentum measurements.

Due to the high rate of events arising from the high design luminosity, real-time event classification and triggering is a main challenge. The first trigger level must reduce the rate of events by a few orders of magnitude, and generate these decisions within a time-interval of a few nanoseconds. A muon-based dedicated hardware triggering mechanism is designed to perform this task.

In this thesis we study an application of Artificial Neural Networks (ANN) to this triggering challenge. ANNs have been applied extensively to other HEP problems, and have shown promising results. It is hypothesized that the use of ANNs will improve on the performance of the dedicated hardware algorithm.

The first and second chapters give an overview of the accelerator and the detector. The focus of these background chapters is on the muon trigger algorithm and its implementation. The physical motivation, detector technology, and technical implementation difficulties are covered here.

The third chapter gives basic background on the subject of Artificial Neural Networks. The biological motivation is discussed, and the basic mathematical derivation is presented.

The fourth chapter is a review of the study. Different factors affecting the ANN performance are analysed, and an optimal architecture is chosen. The ANN performance is tested and compared with a simulation of the dedicated hardware trigger, showing promising results. The effect of the superior measurement resolution available off-line is studied and shown to be accountable for the high performance achieved by the ANN. Another such comparison is made in a low-resolution area of the detector, but the ANN performance is equivalent to the dedicated hardware in this region. Application of the ANN as an extra trigger level after the first hardware trigger level is shown to be a practical alternative for improving background rejection. Finally, using the ANN to configure the dedicated hardware trigger is discussed.

1. The Large Hadron Collider

The main purpose of ongoing particle physics research is to probe deeper into the structure of matter, searching for the origins of the Standard Model. New accelerators being constructed around the world recreate conditions prevailing in the universe just 10^{-12} seconds after the Big Bang, when the temperature was 10^{16} degrees. These conditions and the high rate of collisions will reveal the behaviour and allow the study of new fundamental particles.

The Large Hadron Collider[1] (LHC), currently under construction in the 27km LEP tunnel at CERN, is a two-ring superconductor accelerator. Once completed, it will generate proton-proton collisions at center of mass (CM) energies of 14 TeV, and at a luminosity of $10^{33} cm^{-2} s^{-1}$ and in 2-3 years reaching the target luminosity of $10^{34} cm^{-2} s^{-1}$. Protons will be accelerated in bunches of 10^{11} protons, at a rate of 40M bunches / sec. Each bunch crossing will produce about three collisions at the beginning and about 23 collisions when reaching the target luminosity. This high rate of events (1 GHz) calls for extremely fast triggering and classification solutions.

This accelerator is diverse in its design, and can also generate e-p collisions at energies of 1.3 TeV, or ion-ion collisions at energies up to 1148 TeV.

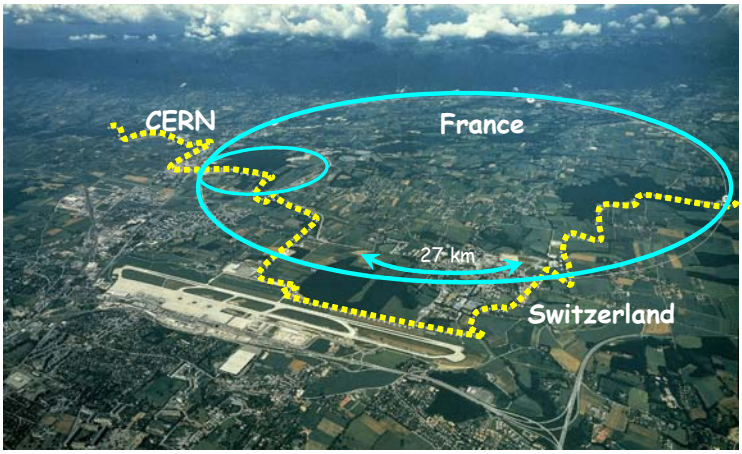

Figure 1: LHC - Aerial view

The LHC is a proton-proton collider, in contrast to LEP, which collided electrons with positrons. Though both LEP and LHC are synchrotrons, the LHC manages much higher CM collision energies, due to the reduced synchrotron radiation associated with accelerating protons. A novel twin-bore magnetic structure allows the two beams of equally-charged protons to circulate in opposite directions.

1.1 Acceleration stages

CERN's accelerator complex includes particle accelerators and colliders, and can handle beams of electrons, positrons, protons, antiprotons, and "heavy ions" (the nuclei of atoms, such as oxygen, sulphur, and lead)[2]. Each type of particle is produced in a different way, but then passes through a similar succession of acceleration stages, moving from one machine to another. The first steps are usually provided by linear accelerators, followed by larger circular machines. CERN has ten accelerators altogether, the biggest has been, until it was dismantled, the Large Electron Positron collider (LEP) followed by the Super Proton Synchrotron (SPS).

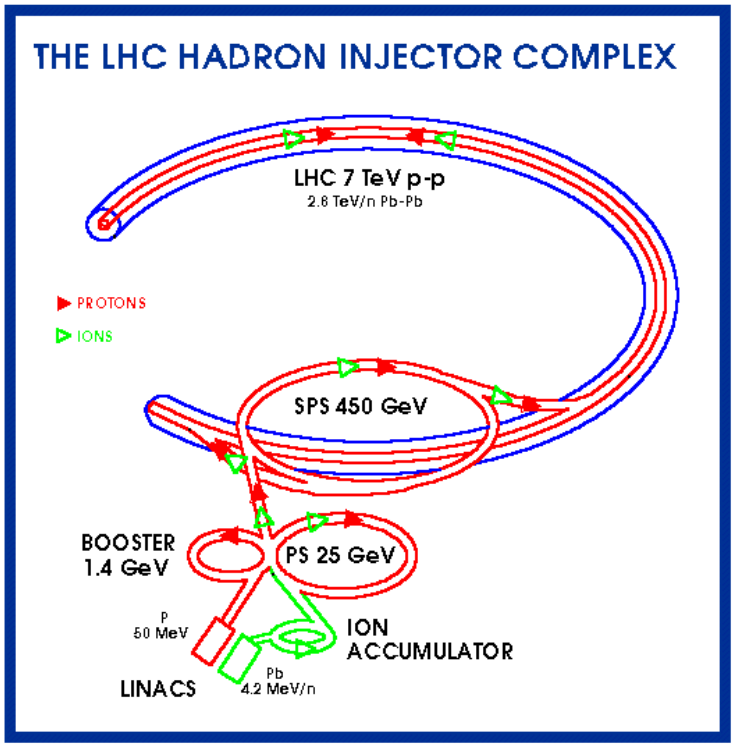

Figure 2: LHC acceleration stages

The first stage of acceleration happens in a linear accelerator (Linac), and each type of particle has its own line. This is because of their different masses - a lead ion is about 200 times heavier than a proton, and almost 400,000 times heavier than an electron. The final energies are 500 MeV for electrons, 50 MeV for protons, and 4.2 MeV/nucleon for lead (Pb) ions.

For the LHC experiment proton and very heavy ion beams then follows a 1.0 GeV "Booster" synchrotron to increase the energy prior to injection into the PS. Then all particle beams pass through the PS machine itself. The beams are then injected into the bigger rings for further acceleration: SPS, and then the LHC.

1.2 LHC experiments

The LHC has eight long straight sections symmetrically distributed around the ring. In these sections the two opposing proton beams can be steered to collide at mid-point. There are currently five experiments in development stages: ALICE, ATLAS, CMS, LHCb and TOTEM. In this thesis we will focus on the ATLAS, one of the two multipurpose experiments at at the LHC.

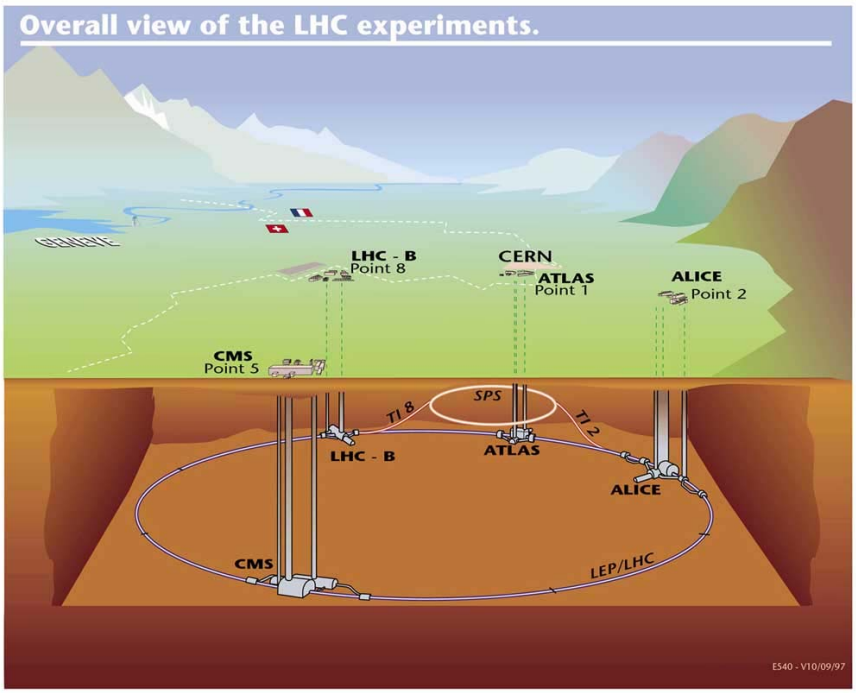

Figure 3[3]: LHC experiments

1.3 Timetable

Initial operation of the LHC is planned for summer 2007 at a reduced energy of 900 GeV (CM). Full commissioning up to 7 TeV will take place during the 2008 winter shutdown, and high-energy physics runs are planned for the rest of 2008.

The initial low-energy stage will be devoted to tuning the accelerator, the detectors and the triggers in order to achieve the required performance. The first physics runs will proceed at a low luminosity of $10^{33} cm^{-2} s^{-1}$, primarily allowing B-physics studies. High luminosity runs, expected after 2-3 years operation, are the main motivation for precise trigger algorithms and will be the driving force for discovering new physics.

2. The ATLAS detector

2.1 General description and nomenclature

ATLAS - A Large Toroidal LHC Apparatus - is the largest LHC experiment[4]. The purpose of the experiment is to explore the full range of new physics available in the LHC, while maintaining the capability to improve the accuracy of already measured phenomena.

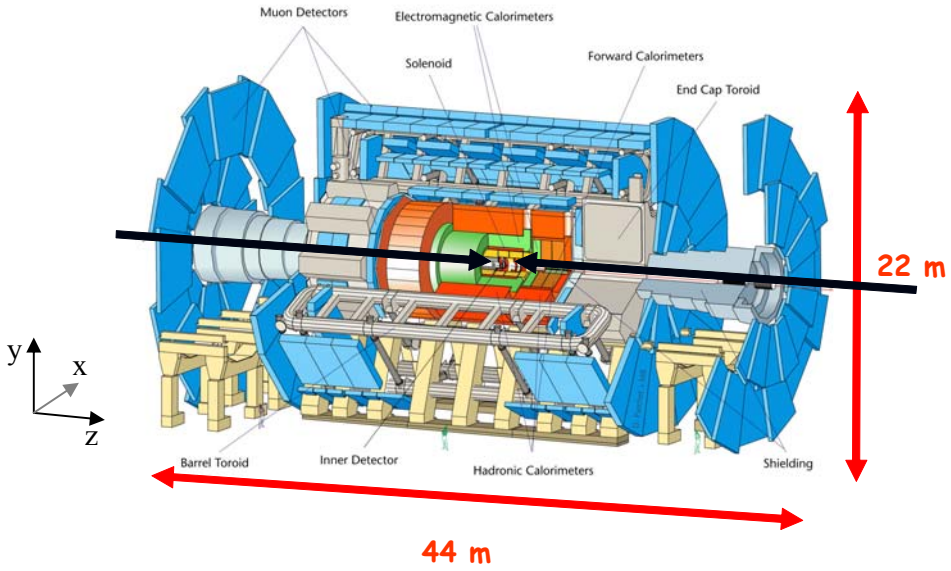

Figure 4: Overview of ATLAS detector

Measuring 44 meters long, 22 meters high and weighing 7000 tons, the ATLAS detector is the largest particle detector in the world. Proton beams enter the detector at both sides and collide inside the Inner Detector. Fragments from the collision spread in all directions, and are detected by the various elements in ATLAS.

Throughout this thesis, the following nomenclature will be adhered to. The beam direction defines the *z*-axis, and the *x*-*y* plane is the plane transverse to the beam direction. The positive *x*-axis is defined as pointing from the interaction point to the centre of the LHC ring, and the positive *y*-axis is pointing upwards (Figure 4). The azimuthal angle Φ ($\tan\phi = P_y / P_x$) is measured around the beam axis, and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$. The transverse momentum $P_T = \sqrt{P_x^2 + P_y^2}$ and the transverse energy E_T are defined in the *x*-*y* plane.

ATLAS is physically divided into two pseudorapidity regions – The barrel and the endcap. Each region contains the same basic building blocks, but the implementation varies.

2.2 Physics motivation

The LHC was designed to tackle some of the most fundamental questions regarding our understanding of the physical world. Among the physics goals of the LHC is the understanding of fundamental symmetry breaking, the search for the Higgs boson, the search for supersymmetrical particles and B-Physics research. The ATLAS detector, and in particular the muon triggering mechanism, are designed to suit these needs.

The desire to probe the origin of the electroweak scale lead ATLAS design to focus on the search for the predicted Higgs boson. The ATLAS detector is sensitive to the Higgs boson in the mass range 80 GeV – 1000 GeV. There are various decay channels for the Higgs boson, each with a different rate and different signal-to-background ratio in the various mass regions:

- $H \to \gamma\gamma$ direct production;
- $H \to \gamma\gamma$ from the associated production WH, ZH and $t\bar{t}H$, using a lepton (e, μ) tag from the vector boson or top quark decay;
- $H \to b\bar{b}$ from the associated production WH, ZH and $t\bar{t}H$, using a lepton (e, μ) tag and b-tagging;
- $H \to ZZ^* \to 4l$;
- $H \to ZZ \to 4l$ and $H \to ZZ \to ll\nu\nu$;
- $H \to WW \to l\nu jj$ and $H \to ZZ \to lljj$.

Of these channels, the $H \to ZZ^* \to 4l$ channel is dubbed the golden channel, as it offers a high branching ratio and a clean signature for the detection of a massive Higgs boson. This channel is indicated by four high-P_T leptons emerging from the interaction point (Figure 5). Salvaging this event from the minimum-bias background requires a triggering mechanism capable of identifying high-P_T leptons during the short 25ns interval between bunch crossings. This is one of the main motivation for the ATLAS muon trigger.

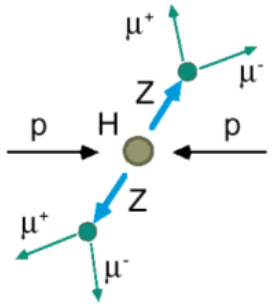

Figure 5: Higgs golden channel decay

The rate of B-hadron production at the LHC is enormous thanks to the large hadronic cross-section for b-quark production and the high luminosity of the machine. About one collision in every hundred will produce a b-quark pair, which is a considerably better signal-to-noise ratio than possible at lower-energy hadron machines such as the Tevatron.

In ATLAS, the inclusive-muon trigger with a P_T threshold of 6 GeV will make an initial selection of B-events. For example, one of the B-physics decay channels that will be triggered using the muon triggers is the J/Ψ decay described in Figure 6:

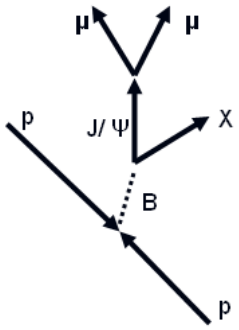

Figure 6: B-Physics J/Ψ decay

The study of Supersymmetry (SUSY) is another main goal of LHC and ATLAS. Theoretical models predict that most SUSY particles decay to the Lightest SUSY Particle (LSP) which is stable, neutral and weakly interacting. The LSP escapes the detector, giving the characteristic SUSY signature of missing transverse energy. Energetic muons are also good indicators of interesting physical processes involving these particles.

2.3 ATLAS design criteria

The basic design criteria[5] of the detector include the following:

- Very good electromagnetic calorimetry for electron and photon identification and measurements, complemented by full-coverage hadronic calorimetry for accurate jet and missing transverse energy measurements;
- High-precision muon momentum measurements, with the capability to guarantee accurate measurements at the highest luminosity using the external muon spectrometer alone;
- Efficient tracking at high luminosity for high-P_T lepton-momentum measurements, electron and photon identification, τ-lepton and heavy-flavour identification, and full event reconstruction capability at lower luminosity;
- Large acceptance in pseudorapidity (η) with almost full azimuthal angle (Φ) coverage everywhere;
- Triggering and measurements of particles at low-P_T thresholds, providing high efficiencies for most physics processes of interest at LHC.

ATLAS contains the four classical basic building blocks of a general particle detector: an Inner Detector, Electromagnetic and Hadronic Calorimeters, and a Muon Chamber. These are briefly described in the following sections.

2.4 The Inner Detector

The Inner Detector[6] (ID) combines high-resolution detectors at the inner radii with continuous tracking elements at the outer radii, all contained in the Central Solenoid (CS) which provides a nominal magnetic field of 2 Tesla.

The momentum and vertex resolution requirements from physics call for high-precision measurements to be made with fine-granularity detectors, given the very large track density expected at the LHC. Semiconductor tracking detectors, using silicon microstrip (SCT) and pixel technologies offer these features. The highest granularity is achieved around the vertex region using semi-conductor pixel detectors[7] (12μm in r-Φ and 66μm in z). The total number of precision layers must be limited because of the material they introduce, and because of their high cost. Typically, three pixel layers and eight strip layers (four space points) are crossed by each track. A large number of tracking points (typically 36 per track) is provided by the straw tube tracker (TRT), which provides continuous track-following with much less material per point and a lower cost. The combination of the two techniques gives very robust pattern recognition and high precision in both Φ and z coordinates. The straw hits at the outer radius contribute significantly to the momentum measurement, since the lower precision per point compared to the silicon is compensated by the large number of measurements and the higher average radius.

The outer radius of the ID cavity is 115 cm, fixed by the inner dimension of the cryostat containing the LAr EM calorimeter, and the total length is 7m, limited by the position of the end-cap calorimeters. Mechanically, the ID consists of three units: a barrel part extending over $\pm 80cm$, and two identical end-caps covering the rest of the cylindrical cavity. The precision tracking elements are contained within a radius of 56 cm, followed by the continuous tracking, and finally the general support and service region at the outermost radius.

The SCT system is an order of magnitude larger in surface area than previous generations of silicon microstrip detectors, and in addition must face radiation levels which will alter the fundamental characteristics of the silicon wafers themselves.

The TRT is based on the use of straw detectors, which can operate at the very high rates expected at the LHC by virtue of their small diameter and the isolation of the sense wires within individual gas volumes. Electron identification capability is added by employing xenon gas to detect transition-radiation photons created in a radiator between the straws. This technique is intrinsically radiation hard, and allows a large number of measurements, typically 36, to be made on every track at modest cost.

2.5 Calorimeters

The calorimetry consists of an electromagnetic (EM) calorimeter covering the pseudorapidity region $|\eta|<3.2$, a hadronic barrel calorimeter covering $|\eta|<1.7$, hadronic end-cap calorimeters covering $1.5<|\eta|<3.2$, and forward calorimeters covering $3.1<|\eta|<4.9$.

The EM calorimeter is a lead/liquid-argon (LAr) detector with accordion geometry[8]. Over the pseudorapidity range $|\eta|<1.8$, it is preceded by a presampler detector, installed immediately behind the cryostat cold wall, and used to correct for the energy lost in the material (ID, cryostats, coil) upstream of the calorimeter. Photons and electrons are stopped and absorbed in the EM calorimeter, and their energy is measured.

The hadronic barrel calorimeter is a cylinder divided into three sections: the central barrel and two identical extended barrels. It is based on a sampling technique with plastic scintillator plates (tiles) embedded in an iron absorber[9]. At larger pseudorapidities, where higher radiation resistance is needed, the intrinsically radiation-hard LAr technology is used for all the calorimeters: the hadronic end-cap calorimeter, a copper LAr detector with parallel-plate geometry, and the forward calorimeter, a dense LAr calorimeter with rod-shaped electrodes in a tungsten matrix. Hadrons are stopped and absorbed in this layer, while muons generally escape after some scattering within the iron material.

The barrel EM calorimeter is contained in a barrel cryostat, which surrounds the Inner Detector cavity. The solenoid which supplies the 2T magnetic field to the Inner Detector is integrated into the vacuum of the barrel cryostat and is placed in front of the EM calorimeter. Two end-cap cryostats house the end-cap EM and hadronic calorimeters, as well as the integrated forward calorimeter. The barrel and extended barrel tile calorimeters support the LAr cryostats and also act as the main solenoid flux return.

2.6 The Muon Spectrometer

Muons generally pass through the Inner Detector and calorimeter material with minimal energy loss. The muon spectrometer allows measurement of muon momenta – the muons pass through a magnetic field generated by the end-cap and barrel toroids, and their track is curved. The track curvature indicates the muon transverse momentum: low-P_T muons are deflected more than high-P_T ones.

The anticipated high level of particle fluxes has had a major impact on the choice and design of the spectrometer instrumentation, affecting required performance parameters such as rate capability, granularity, ageing properties and radiation hardness. Trigger and reconstruction algorithms have been optimised to cope with the difficult background conditions resulting from penetrating primary collision products and from radiation backgrounds, mostly neutrons and photons in the 1 MeV range, produced from secondary interactions in the calorimeters, shielding material, beam pipe and LHC machine elements.

In the barrel region, tracks are measured in chambers arranged in three cylindrical layers (‘stations’) around the beam axis; in the transition and end-cap regions, the chambers are installed vertically, also in three stations. Over most of the η-range, a precision measurement of the track coordinates in the principal bending direction of the magnetic field is provided by Monitored Drift Tubes (MDTs). At large pseudorapidities and close to the interaction point, Cathode Strip Chambers (CSCs) with higher

granularity are used in the innermost plane over $2<|\eta|<2.7$, to withstand the demanding rate and background conditions. Optical alignment systems have been designed to meet the stringent requirements on the mechanical accuracy and the survey of the precision chambers.

The precision measurement of the muon tracks is made in the *r*-*z* projection, in a direction parallel to the bending direction of the magnetic field; the axial coordinate (*z*) is measured in the barrel and the radial coordinate (*r*) in the transition and end-cap regions. The MDTs provide a single-wire resolution of ~80 mm when operated at high gas pressure (3 bars) together with robust and reliable operation thanks to the mechanical isolation of each sense wire from its neighbours.

The trigger system covers the pseudorapidity range $|\eta|\leq 2.4$. Resistive Plate Chambers (RPCs) are used in the barrel and Thin Gap Chambers (TGCs) in the end-cap regions. The trigger chambers for the ATLAS muon spectrometer serve a threefold purpose:

- Bunch crossing identification, requiring a time resolution better than the LHC bunch spacing of 25 ns;
- A trigger with well-defined P_T cut-offs in moderate magnetic fields, requiring a granularity of the order of 1 cm;
- Measurement of the second coordinate in a direction orthogonal to that measured by the precision chambers, with a typical resolution of 5–10 mm.

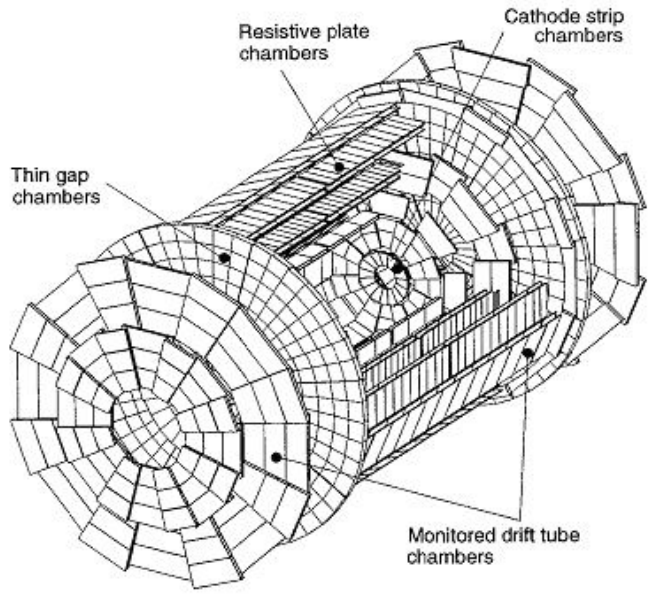

Figure 7: Muon chamber detectors

The overall layout of the muon chambers in the ATLAS detector is shown in Figure 7, indicating the different regions in which the four chamber technologies are employed. The chambers are arranged such that particles from the interaction point traverse three stations of chambers. The positions of these stations are optimized for good hermeticity

and optimum momentum resolution. In the barrel, particles are measured near the inner and outer field boundaries, and inside the field volume, in order to determine the momentum from the sagitta of the trajectory. In the end-cap regions, for $|\eta|>1.4$, the magnet cryostats do not allow the positioning of chambers inside the field volume. Instead, the chambers are arranged to determine the momentum with the best possible resolution from a point-angle measurement.

The following table gives an overview of the muon chamber instrumentation. 'Area covered' refers to chamber modules which normally contain several detector layers.

	Precision chambers		Trigger chambers	
	CSC	MDT	RPC	TGC
Number of chambers	32	1194	596	3588
Number of readout channels	67 000	370 000	355 000	440 000
Area covered (m^2)	27	5500	3650	2900

2.7 Magnet System

The magnet system[10] consists of three air-core superconducting toroids (Figure 8) designed to produce a large-volume magnetic field covering the rapidity range $0 \leq|\eta|\leq$ 2.7, with an open structure that minimizes the contribution of multiple scattering to the momentum resolution. The Barrel Toroid[11] (BT) extends over a length of 25 m, with an inner bore of 9.4 m and an outer diameter of 20.1 m. The two End-Cap Toroids[12] (ECTs) are inserted in the barrel at each end. They have a length of 5.0 m, an inner bore of 1.65 m and an outer diameter of 10.7 m. In Figure 8 the right-hand ECT is retracted.

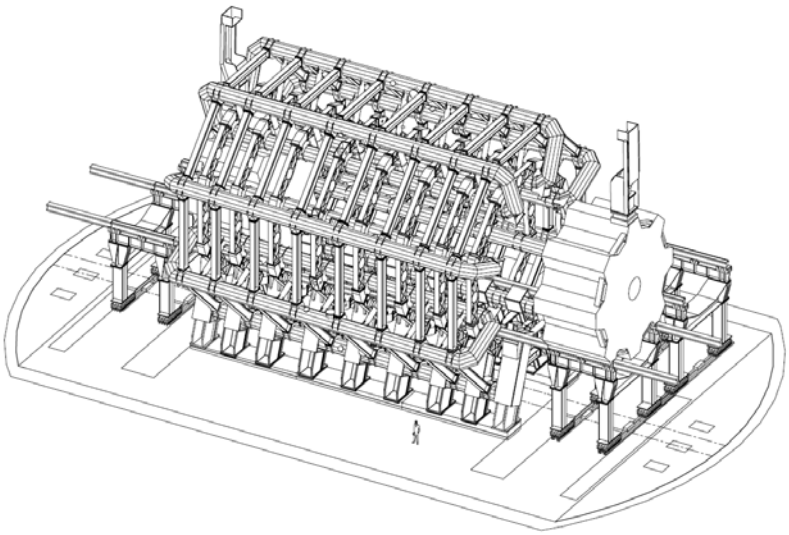

Figure 8: The superconducting air-core toroid magnet system

The Central Solenoid (CS), not shown here, provides a central field of 2 Tesla with a peak magnetic field of 2.6 T at the superconductor itself. The peak magnetic fields on the superconductors in the BT and ECT are 3.9 and 4.1 T respectively.

The performance in terms of bending power is characterised by the field integral $\int B \cdot dl$, where B is the azimuthal field component and the integral is taken on a straight line trajectory between the inner and outer radius of the toroids. The BT provides 2 to 6 Tm and the ECT contributes with 4 to 8 Tm in the 0.0-1.3 and 1.6-2.7 pseudorapidity ranges respectively. The bending power is lower in the transition regions where the two magnets overlap ($1.3<|\eta|<1.6$). Figure 9 shows the bending power for various pseudorapidity and Φ regions:

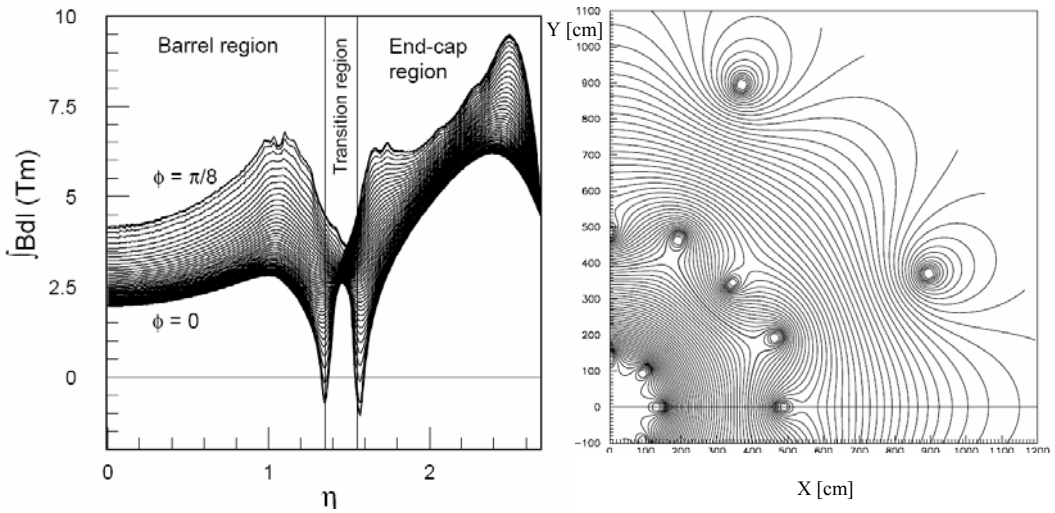

Figure 9: Magnetic bending power, Field map in transition region

The magnetic field exhibits a high degree of inhomogeneity, making the muon momentum reconstruction a difficult task. This is especially true in the transition region between the barrel and the end-cap, as shown in Figure 9 where on the left side seen the Magnetic bending power integrated between the first and the last muon chamber as a function of the psudorapidity, and on the right the field map in transition region. The field lines are shown in a plane perpendicular to the beam axis in the middle of the end-cap toroid. The scales are in centimeters.

.

2.8 Monitored Drift-Tube chambers

The basic detection elements of the MDT precision chambers[13] are aluminium tubes of 30 mm diameter and 0.4 mm wall thickness, with a 50 micron diameter central W–Re wire. The tubes are operated with a non-flammable Ar-CH_4-N_2 mixture at 3 bar absolute pressure. The envisaged working point provides for a highly linear space–time relation with a maximum drift time of ~ 500 ns, a small Lorentz angle, and good ageing

properties due to small gas amplification. The single-wire resolution is typically 80 micron, except very close to the anode wire.

The tubes are produced by extrusion from a hard aluminium alloy, and are available commercially. They are closed by endplugs which provide for accurate positioning of the anode wires, wire tension, gas tightness, and electrical and gas connections. The drift tubes can be manufactured to tight mechanical tolerances which are well matched to their intrinsic resolution properties, mostly using automated assembly procedures. The tube lengths vary from 70 cm to 630 cm.

To improve the resolution of a chamber beyond the single-wire limit and to achieve adequate redundancy for pattern recognition, the MDT chambers are constructed from 2x4 monolayers of drift tubes for the inner station and 2x3 monolayers for the middle and outer stations. The tubes are arranged in multilayer pairs of three or four monolayers, respectively, on opposite sides of a rigid support structure. The support structures ('spacer frames') provide for accurate positioning of the drift tubes with respect to each other, and for mechanical integrity under effects of temperature and gravity; for the barrel chambers which are not mounted in a vertical plane, they are designed to bend the drift tubes slightly in order to match them to the gravitational sag of the wires. The spacer frames also support most of the components of the alignment system.

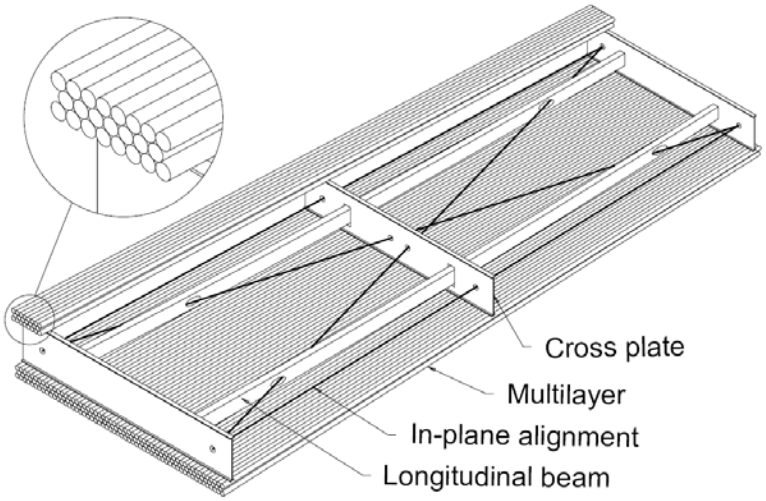

Figure 10: Schematic drawing of a rectangular MDT chamber

The structural components of the spacer frames are three 'cross-plates', to which the drift-tube multilayers are attached, and two 'long beams' connecting the cross-plates. The frames need to be constructed to a moderate mechanical accuracy of 0.5 mm only; accurate positioning of the drift tubes is provided by the assembly procedure. They will be attached to the rail structures of the spectrometer by three-point kinematic supports. Once a chamber is installed in its final location in the spectrometer, mechanical deformations are monitored by an in-plane optical system; hence the name 'monitored drift-tube chambers'.

Each drift tube is read out at one end by a low-impedance current sensitive preamplifier, with a threshold five times above the noise level. The preamplifier is followed by a differential amplifier, a shaping amplifier and a discriminator. The output of the shaping amplifier is also connected to a simple ADC, to correct the drift-time measurement for time-slewing using the charge integrated signal.

2.9 Thin Gap Chambers

Thin Gap Chambers[14, 15] (TGCs) provide two functions in the end-cap of the ATLAS muon spectrometer: the muon trigger capabilities and the azimuthal coordinate to complement the bending coordinate measured by the MDTs. The middle tracking plane of the MDT is complemented by seven layers of TGCs, which provide both functions. The inner tracking layer is complemented by two layers of TGCs, which provide only the azimuthal coordinate, while the azimuthal coordinate in the outer MDT layers is obtained by the extrapolation from the middle layer. The radial, or bending, coordinate is measured by reading which TGC wire-group is hit; the azimuthal coordinate is measured by reading which radial strip is hit.

As trigger chambers, the TGCs are required to have good time resolution to provide bunch-crossing identification, and fine granularity to provide a sharp cut-off in the momentum of the triggered muon. Good time resolution means assigning more than 99% of the triggered muons to the correct bunch-crossing. Fine granularity is needed since the trigger chambers must be located outside the magnetic field and can have only a relatively short lever arm of approximately 1 m. To match the geometric granularity to the needed momentum resolution, the number of wires in a wire-group varies, as a function of η, from 4 to 20 wires, i.e. from 7.2 to 36 mm. The alignment of the wire-groups in two (three) consecutive layers is staggered by half (third) the group width. This achieves good position resolution with a smaller number of electronics channels. The required azimuthal granularity of 2–3 mrad is obtained by staggering the radial strips. The design described here satisfies the time, momentum, and azimuthal coordinate resolution requirements.

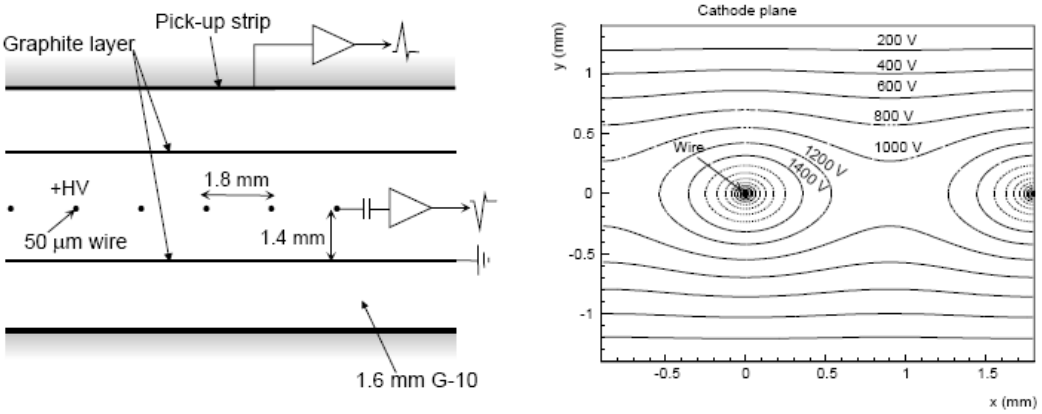

Figure 11: TGC schematic side view and the internal electric field.

Thin Gap Chambers operating in a saturated mode have a structure similar to Multi-Wire Proportional Chambers (MWPCs), except that the anode-to-anode, i.e. wire-to-wire, distance is larger than the cathode-to-anode distance. They were first used on a large scale in the OPAL[16, 17, 18] experiment. With the use of a highly quenching gas mixture of CO_2 and n-pentane (n-C_5H_{12}), 55%:45%, this type of cell geometry allows operation in saturated mode[19]. Among other advantages, saturated mode has the advantage of small sensitivity to mechanical deformations, which is important for large detectors.

The high electric field around the TGC wires (see Figure 11 on the left side) and the small distance between wires strongly reduce the drift component of ionization clusters, leading to very good time resolution. The best time resolution is achieved for the minimal anode-to-anode distance, 1.8 mm, and the lowest operational voltage is achieved for the smallest anode- to-cathode distance, 1.4 mm.

Taking into account mechanical tolerances, an operating voltage of 3.1 ± 0.1 kV and an anode-cathode gap of 1.4 mm ± 0.1 mm were chosen for the construction of large detectors. This achieves 99% efficiency for a minimum-ionizing particle within a time window of 25 ns (the LHC bunch-crossing period) without introducing stringent mechanical requirements.

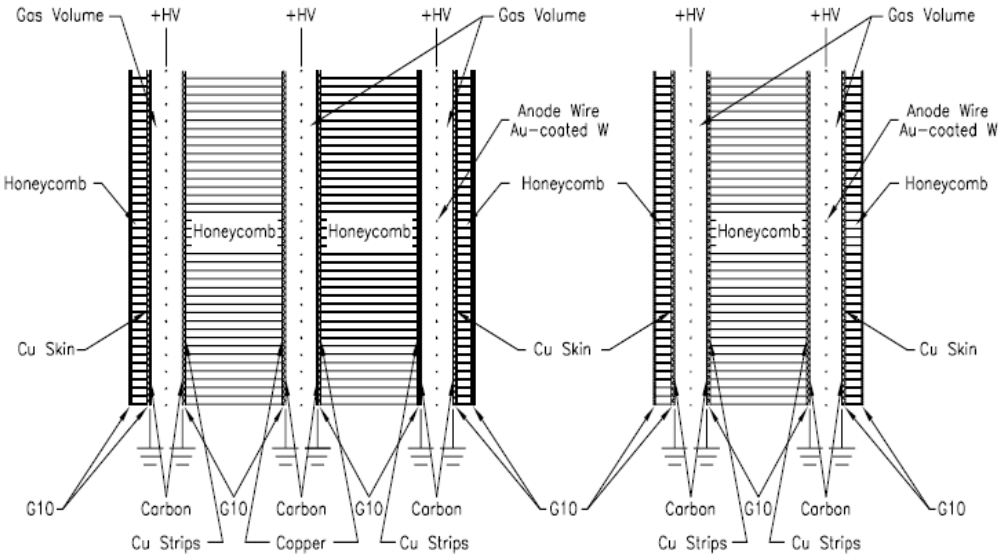

Figure 12: Cross-section of a triplet (left) and of a doublet of TGCs.

In the end-cap, the doublet furthest from the interaction point is referred to as the 'pivot' plane (see Figure 16). The layout geometry, chamber overlapping, and channel wiring have been arranged so that there are, to first order, no overlaps and no holes in this plane . Tracks passing through this plane can be given an unambiguous η-Φ coordinate depending only on their unique intercept in this plane. Because any track segment in the other planes must be in coincidence with this plane, double counting can be avoided in

the Level-1 trigger. The Level-1 trigger looks for tracks in a cone opening backwards from the pivot plane.

The very delicate structure of the TGC demands stringent validation tests for each TGC manufactured. Cosmic-ray telescope efficiency testbenches were carried out in order to test the individual TGC's efficiencies. Two cosmic ray telescopes were built in the Technion and Tel-Aviv University for this purpose. The testbench can test up to seven TGC units in parallel. Accumulating events for a period of one week provides a full mapping of the efficiency of each detector in the required resolution. Figure 13 shows a (rare) malfunctioning TGC chamber.

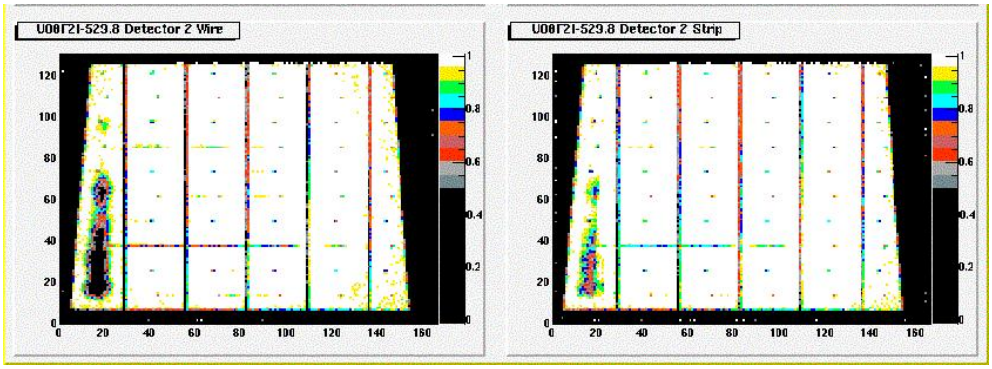

Figure 13: TGC testbench showing inefficiencies

2.10 Trigger and data-acquisition

The main challenges at the LHC that have an impact on the experiment's trigger system are an unprecedented rate of 10^9 interactions per second, the need to select rare predicted physics processes with high efficiency while rejecting much higher-rate background processes, and large and complex detectors with huge numbers of channels $O(10^7)$. Decisions must be taken every 25 ns; at high luminosity, each bunch crossing contains about 23 interactions. At the end of the decision chain, the event storage rate is limited to approximately 100 Hz, by practical limitations in the offline computing power and storage capacity. The average event size is 1 MB, restricting the final output event rate.

The ATLAS trigger and data-acquisition system[20] is based on three levels of online event selection. Each trigger level refines the decisions made at the previous level and, where necessary, applies additional selection criteria. Starting from an initial bunch-crossing rate of 40 MHz, the rate of selected events must be reduced to ~100 Hz for permanent storage. While this requires an overall rejection factor of 10^7 against 'minimum-bias' processes, excellent efficiency must be retained for the rare new physics, such as Higgs boson decays, that is sought in ATLAS.

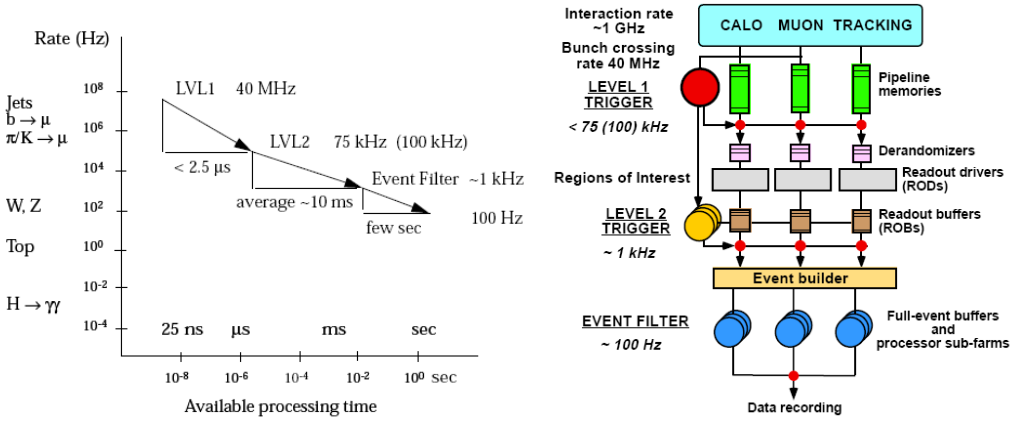

Figure 14: Event rates and decision stages, schematic trigger flow.

It is important to keep the latency (time taken to form and distribute the trigger decision) to a minimum. During this time information for all detector channels has to be retained in 'pipeline' memories. These memories are generally contained in custom integrated circuits, placed on or close to the detector, usually in inaccessible regions and in a high-radiation environment. The LVL1 latency, measured from the time of the proton–proton collision until the trigger decision is available to the front-end electronics, is required to be less than 2.5 μs. In order to achieve this, the LVL1 trigger is implemented as a system of purpose-built hardware processors.

Events selected by LVL1 are read out from the front-end electronics systems of the detectors into readout buffers (ROBs). A large number of front-end electronics channels are multiplexed into each ROB. Intermediate buffers average out the high instantaneous data rate at the output of the pipeline memories to match the available input bandwidth of the readout drivers (RODs). All of the data for the selected bunch crossing from all of the detectors are held in the ROBs either until the event is rejected by the LVL2 trigger or, in case the event is accepted by LVL2, until the data have been successfully transferred by the DAQ system to storage associated with the Event Filter (which makes the third level of event selection). The process of moving data from the ROBs to the Event Filter (EF) is called event building. Whereas before event building each event is composed of many fragments, with one fragment in each ROB, after event building the full event is stored in a single memory accessible by an EF processor.

The LVL2 trigger makes use of 'region-of-interest' (RoI) information provided by the LVL1 trigger. This includes information on the position (η and Φ) and P_T range of candidate objects, and energy sums. The RoI data are sent by LVL1 to LVL2, for all events selected by the LVL1 trigger, using a dedicated data path. Using the RoI information, the LVL2 trigger selectively accesses data from the ROBs, moving only the data that are required in order to make the LVL2 decision. The LVL2 trigger has access to all of the event data, if necessary with the full precision and granularity. However, typically only data from a small fraction of the detector, corresponding to limited regions

centred on the objects indicated by the LVL1 trigger, are needed by the LVL2 trigger. Hence, usually only a few per cent of the full event data are required thanks to the RoI mechanism.

After LVL2, a last stage of selection is performed in the EF. Here the algorithms will be based on offline code. The EF must reduce the rate to a level suitable for permanent storage, currently assumed to be ~100 Hz for full events of size ~1 Mbyte.

2.11 Muon trigger system

The muon and calorimeter LVL1 trigger systems[21] use simple algorithms to make fast decisions. Local pattern recognition and transverse-energy evaluation are performed on prompt, relatively coarse-grained information, which is provided by the fast muon trigger chambers and the tower summing electronics of the EM and hadronic calorimeters. In the following we will focus on the muon trigger which is the subject of this thesis.

The LVL1 algorithms are executed by custom electronics, programmed in terms of adjustable parameters. The decision time of ~2 μs includes the transmission of signals between the detector and the trigger electronics. During the LVL1 trigger processing, the data from all detector systems is held in pipeline memories. When LVL1 has accepted an event, the data is read out, formatted and initial preprocessing may be applied (e.g. calibration) before it is stored in readout buffers (ROBs) for use by the LVL2 trigger and the EF.

The signals generated by the TGCs in the end-cap are amplified, discriminated and shaped in the detector[22]. Detector-mounted electronics first identify the bunch-crossing and then find coincidences independently in r and Φ roads. The coincidences provide r, δr and Φ, $\delta\Phi$, where r and Φ are the track coordinates in the pivot plane and δr and $\delta\Phi$ are the track's deviation from the infinite momentum track.

Electronics situated outside the ATLAS cavern combine the measurements of r, δr and Φ, $\delta\Phi$ to make a trigger decision which is then passed through the Muon Interface to the Central Trigger Processor. The chamber hits and the intermediate r, δr and Φ, $\delta\Phi$ values are read out by the on and off-detector electronics.

A *Patch-Panel* receives TGC wire-group and strip signals from the ASD Boards, performs bunch-crossing identification and allows the correct treatment of physical overlaps in the TGCs by OR'ing signals from overlapping regions (see Figure 15)..

Slave Boards receive Bunch-crossing-ID assigned wire and strip signals from the Patch-Panel and perform coincidence operations on the input signals from TGC planes. There are four different types of Slave Board, two for the triplets and two for the pair of doublets, where wire and strip signals have different Slave Boards. Each Doublet Slave Board (DSB) receives signals from a pair of TGC doublets and performs 3-out-of-4 coincidence operations on them, to form low-P_T triggers. Each Triplet Slave Board (TSB) receives signals from a triplet plane of TGCs and performs coincidence operations on them. This is 2-out-of-3 coincidence for the wires and 1-out-of-2 for the strips. Outputs

from both the DSB and TSB are sent to the High-P_T Board where high-P_T trigger signals are formed.

The *High-P_T Board* receives the low-P_T signals from sets of four Doublet Slave Boards and hits from a projectively-matching set of Triplet Slave Boards, and performs a two-fold coincidence on these input signals to form high-P_T trigger signals. The board then merges the high-P_T output with the low-P_T signals, giving priority to the high-P_T signals. It has a similar structure to that of a Slave Board. Signals from the DSBs and TSBs are received by phase-adjust sections of the High-P_T Board. The coincidence operation is performed by a 256x288 two-fold coincidence matrix that combines the doublet and triplet results to produce high-P_T trigger signals. There are separate High-P_T Boards for r and Φ. The r, δr and Φ, $\delta\Phi$ information from the High-P_T Boards is passed to the Sector Logic.

The *Sector Logic* combines the δr information from the wires with the $\delta\Phi$ information from the strips, using the outputs of the High-P_T Boards. Each Sector Logic Board handles the trigger data from a single trigger sector. The Sector Logic Board contains a track selector which selects the two highest-P_T tracks in each sector. This Sector Logic consists of r-Φ coincidence matrices, a Track-Preselector and a Track Selector. Hit information for r and Φ is encoded so decoding is necessary immediately after the receivers. Combining the $\delta\Phi$ and δr information through the r-Φ coincidence matrix makes the track information more complete. Each level can make full use of both high and low-P_T input information.

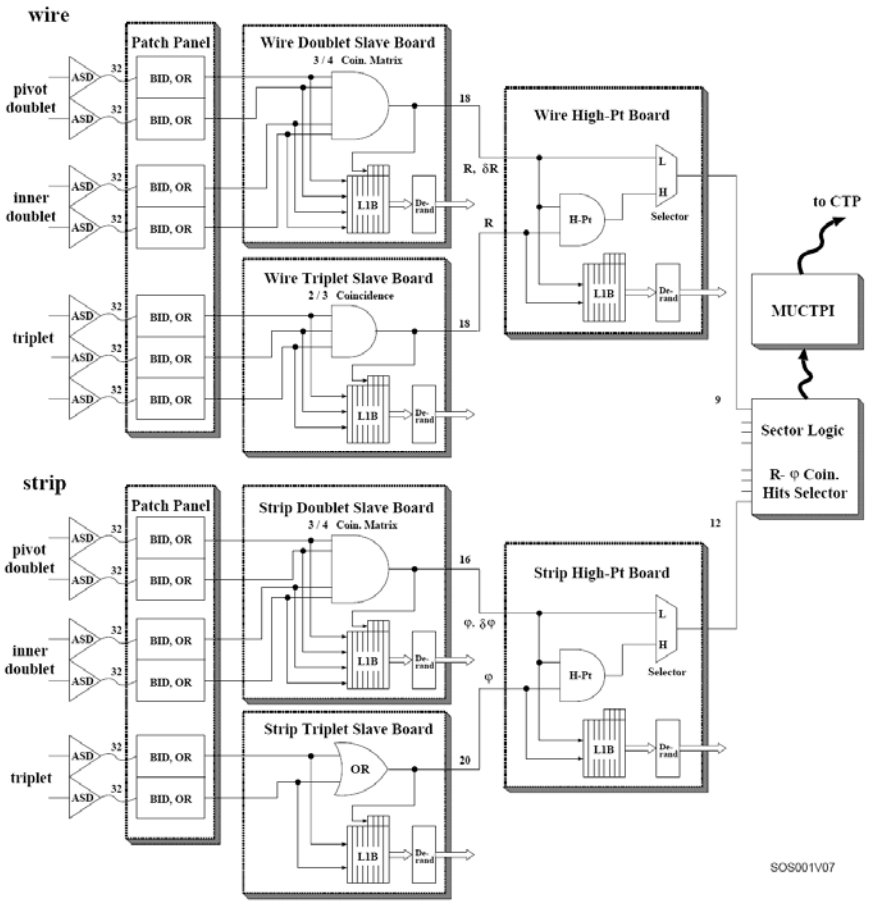

Figure 15: TGC LVL1 trigger scheme

2.12 Muon trigger algorithm

The level-1 muon-trigger[23] is based on the measurement of muon trajectories in two or three different planes (called stations). Muons are bent by the magnetic field generated by the toroids and their angle of deflection depends on their momentum, the field integral along their trajectory and the Coulomb scattering in the material lying in front of the trigger planes. The energy loss fluctuation is also important for low-P_T transverse-momentum triggers.

The differences from the infinite momentum track, 'residuals', are measured using three trigger stations. Figure 16 shows the three stations: M1, M2, and M3 (see also Appendix A). The trigger plane farthest from the interaction point, in the end-cap, and nearest to the interaction point in the barrel, is called the pivot plane. Two different lever arms from the pivot to the other two trigger planes provide two different measurements of the residuals. The two different lever arms allow trigger thresholds to cover a wide range of momenta: the shorter lever arm (the pivot and station 2) covers a lower-momentum region and the longer one (the pivot and station 1), a higher-momentum region.

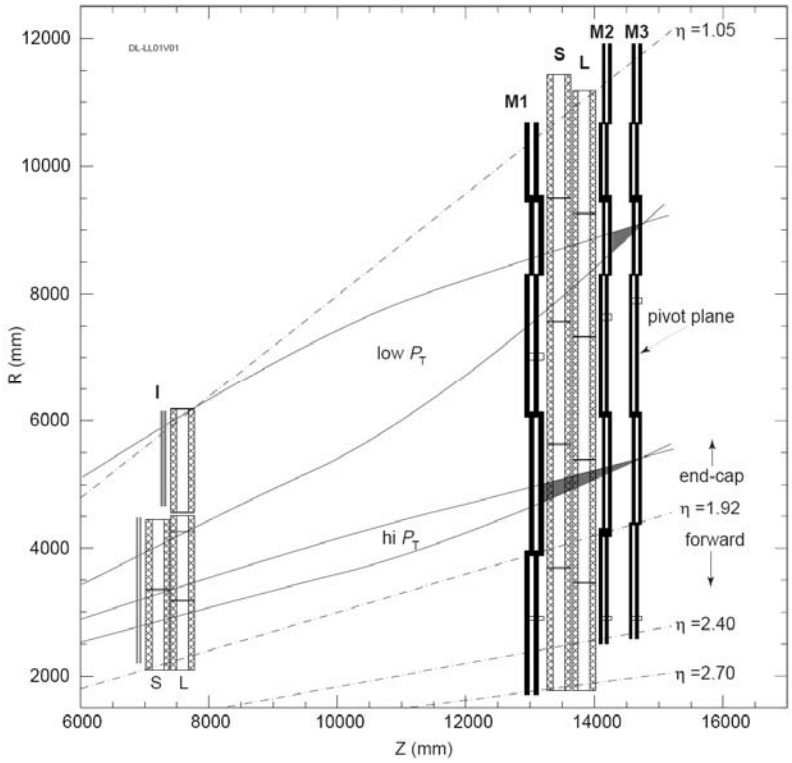

Figure 16[24]: End-cap TGC trigger planes

The residual distribution in a region (r, Φ) on the pivot plane reflects the momentum spectrum of muons passing through that region. The level-1 muon-trigger uses the residual distribution in order to discriminate muons with transverse momentum above some threshold from those below the threshold.

Each hit found in station TGC3 (the pivot) is extrapolated to station TGC2 along a straight line through the interaction point (this is the infinite-momentum track). A coincidence window is then defined around this point, where the window's size depends upon the required P_T threshold. The low-P_T trigger condition is then satisfied if, for both projections:

- There is at least one hit within the coincidence window;
- At least one of the two low-P_T stations has hits in both of the doublet trigger planes, to satisfy the three-out-of-four majority logic.

The size of the coincidence window defines the P_T threshold applied in the trigger: the broader the window the lower the threshold. Figure 17 shows coincidence windows for different trigger thresholds.

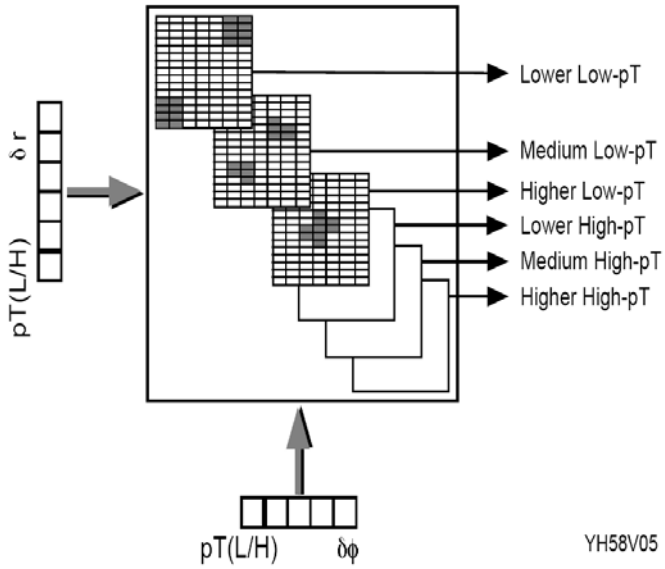

Figure 17[25]: Coincidence windows

A similar procedure is performed for the high- P_T trigger, where the planes of TGC1 together with the pivot plane are used. The high-P_T trigger is satisfied if:

- The track passes the low-P_T criteria, and
- At least two of the three planes of TGC1 in the η view, and one of the two planes of TGC1 in the r-Φ view, are within the appropriate coincidence window.

2.13 System segmentation

The muon-trigger is divided into regions[26] in η-Φ where independent thresholds can be set, i.e. where independent trigger windows can be used. The end-cap region of each octant is divided into six trigger sectors in Φ, where a trigger sector is a logical unit that is treated independently in the trigger. Thus in each end-cap of TGCs there are 48 end-cap trigger sectors. The smallest regions shown in Figure 18 are trigger subsectors which correspond to the smallest unit area of the trigger segmentation. A trigger subsector corresponds to eight channels of wire-groups and eight channels of read-out strips. An end-cap trigger sector contains 49 η rows by 4 Φ columns of trigger subsectors, a total of 196 trigger subsectors

Each η-Φ trigger subsector corresponds to one Region of Interest (RoI). Each subsector is treated independently in the trigger so that the δr and $\delta\Phi$ inputs that determine the P_T condition applied can be set separately for each subsector. In each trigger sector the two highest-P_T track candidates are selected and sent to the Muon Trigger / Central Trigger Processor Interface (MUCTPI).

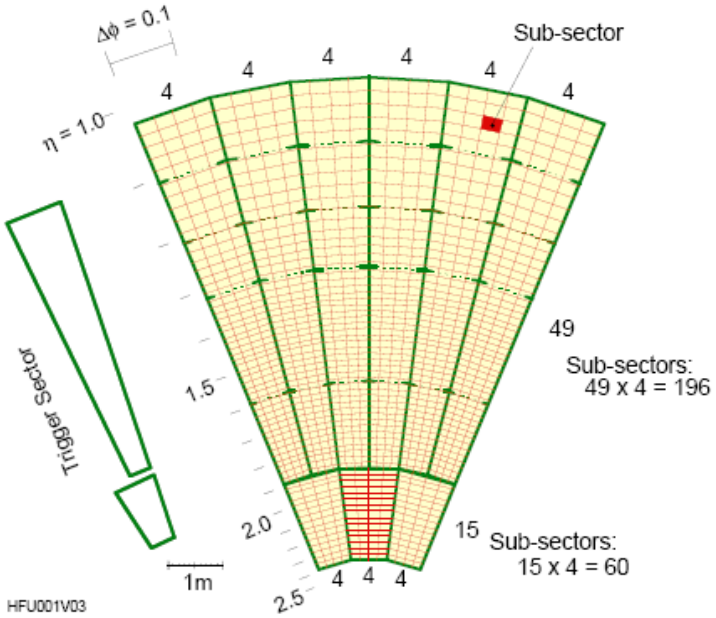

Figure 18: TGC LVL1 octant segmentation

2.14 Efficiency and resolution

The efficiency and resolution[27] of the ATLAS detector in its different regions is critical to the performance of the trigger and reconstruction algorithms.

The material through which muons pass prior to traversing the trigger counters plays an important part in determining the acceptance of the trigger coincidence windows through the contribution of multiple scattering to the measured residual in the trigger chambers. The material in front of the trigger chambers is that of the end-cap and barrel toroids and the inner detector and calorimeters, both hadronic and electromagnetic. In both the barrel and the endcap this material is dominated by the tile and liquid-argon calorimeters, as shown in Figure 19.

Total material between the interaction point and the muon system constitutes between 10 and 15 absorption lengths, with somewhat more material in the region of the barrel end-cap interface at $|\eta|$~1.15. For lower momentum particles this material serves as a barrier between the interaction point and the trigger chamber planes; in the barrel, muons with P_T below 3 GeV are absorbed.

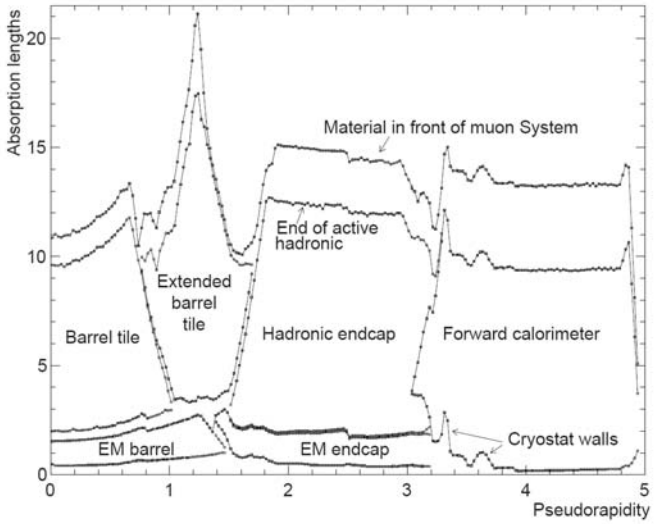

Figure 19[28]: Absorbtion length as a function Pseudorapidity as a result of material in front of muon system

The typical momentum-resolution figure is about 10% where this is mainly limited by the coulomb scattering in the calorimeter. However, there are Φ regions in the η interval $1.0 < |\eta| < 1.8$ where this resolution is significantly worse. This effect is due to the combination of the magnetic fields produced by the barrel and end-cap toroid coils (see section 2.7). Since in these regions the field integral is low, the resulting muon momentum resolution is worse than elsewhere. As shown in Figure 20 these poor resolution regions are located along the barrel and end-cap coils, with η values of ~1.3 (Φ coordinates of the end-cap coils) and ~1.6 (Φ coordinates of the barrel coils).

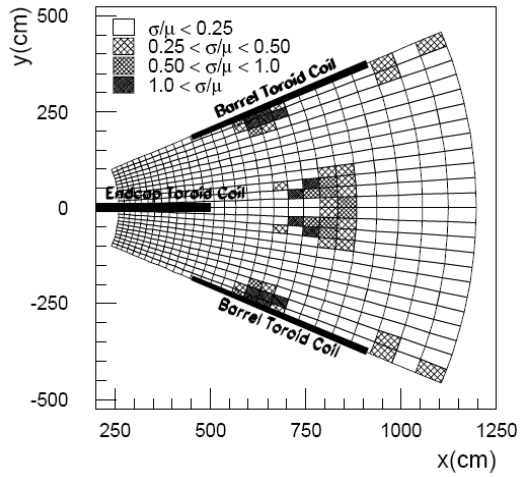

Figure 20[29]: Regions with poor momentum resolution

The momentum resolution is also affected by the finite width of the interaction region, which is around 5.5cm.

2.15 Simulation

The GEANT program (http://wwwasd.web.cern.ch/wwwasd/geant/) simulates the passage of elementary particles through the matter. Originally designed for the High Energy Physics experiments (1981) is currently used all over the world by hundreds of users to simulate HEP detectors. Continuous support and upgrades have kept the simulation program relevant through decades of progress. Along the time it has found applications also outside the HEP domain in areas such as medical and biological sciences, radio-protection and astronautics.

DICE – Detector Integration CodE[30] – was developed by the ATLAS Software Group in 1994 as an interface to the GEANT package. The package contains the geometry and material of the ATLAS detector, and is routinely updated according to the changes in geometry.

ATRIG[31] is a simulation package for the ATLAS Trigger. Information from the various detectors is processed, and a level-1 trigger output is produced.

In our study, we used DICE to generate muon events:

- The generated events were confined to the end-cap region ($|\eta|>1.05$) for focus and simplicity. The barrel region is also a relevant candidate for the same ANN solution, and the conclusions of this study should apply there.
- Only negatively-charged muons were generated. Positively-charged muons will be deflected in an opposite manner, though not exactly symmetrical[32]. Again we assume that the conclusions will be applicable for both signs.

Each simulated event generates 'hits' in the different TGC planes of the end-cap. The number of hits generated in each event can vary between 0-28 hits. The basic structure of the TGC detector (section 2.9) includes two doublets and one triplet – consequently, most of the events contain 7 hits. In some regions the TGC chambers overlap, giving rise to a higher number of hits, while in some regions the TGC chambers display inefficiencies, reducing the number of hits (see also Appendix B).

The hits generated by DICE are fed into the ATRIG simulation, and a simulated LVL1 trigger output is generated. ATRIG is configured to generate a low-P_T trigger tuned to 6 GeV, and thus it simulates the dedicated-hardware algorithm described in section 2.12.

In our study, events with less than four hits in the different TGC layers were removed. This selection is similar to the dedicated hardware coincidence logic described in 2.11. Also, we used hits from all TGC planes in order to generate a low-P_T trigger, while the dedicated hardware uses only the two doublets for this purpose. The effect of this difference is discussed later.

3. Artificial Neural Networks

The use of artificial neural networks seems to have earned its place among the analytical tools of the experimental high-energy physicists. Some classical examples are their use in jet tagging [33,34] and invariant mass reconstruction [35]. Further examples include a second-level trigger for the HERA experiment [36] and a possible use in the search for the Higgs boson at the Large Hadron Collider [37].

There are several advantages to using neural networks for our triggering application. The inhomogeneity of the magnetic field experienced by the muons does not allow a precise analytical solution. A neural network can sometimes generalize and solve a problem for which the logical structure is not entirely clear. This is achieved by training the network on sets of inputs, as will be described later. Also, the inherent parallel structure of a neural network makes it practical to implement in dedicated hardware, as shown in [38]. This dedicated hardware implementation allows computation in the limits of time imposed by the triggering demands.

This chapter[39] describes the theory behind the type of neural network I used in my study, the feedforward layered network, and the algorithm I used to train the networks (i.e. teaching them the relations between the parameters), the backpropagation algorithm.

3.1 Biology

The field of neural computation draws its inspiration from the biological central nervous systems of humans and animals. The recurring building block of the central nervous system is the neural cell, or the neuron. The neuron is a biological cell, with 3 main parts:

- The soma, or the body of the cell, is typically 10-80 μm in diameter.
- The dendrites function as the inputs of the cell – this can be a very large tree-like structure, on which axons form synapses (connections).
- The axon is the output of the cell – it can range from 100μm up to 1 meter in length, and can split up into several branches.

Transmission of signals between neurons (at the synapse) is chemical, and the transmission within a neuron (along the axon) is electrical. The electrical signal originates from a discharge that starts on the soma and travels along the axon. The interior of the neuron, the protoplasm, is negatively charged with respect to the surrounding neural liquid. The potential difference is about 70 mV and is supported by active pumping of the cell membrane and its impermeability to Na^+ ions, which causes a deficiency of positive ions in the protoplasm. Chemical substances called neurotransmitters, which are released from the pre-synapse, alter the permeability of the cell membrane, which causes a transient weakening or depolarization of the resting potential. When at some point the membrane potential drops below a certain threshold, usually about -60 mV, the membrane loses its impermeability to Na^+ and depolarises completely. The discharge starts out from the cell-body and propagates along the axon. Because the discharge in each previous section is complete, the pulse moves in one

direction and does not diminish in strength. When the pulse reaches the synapses at the end of the axon, it in turn can release neurotransmitters.

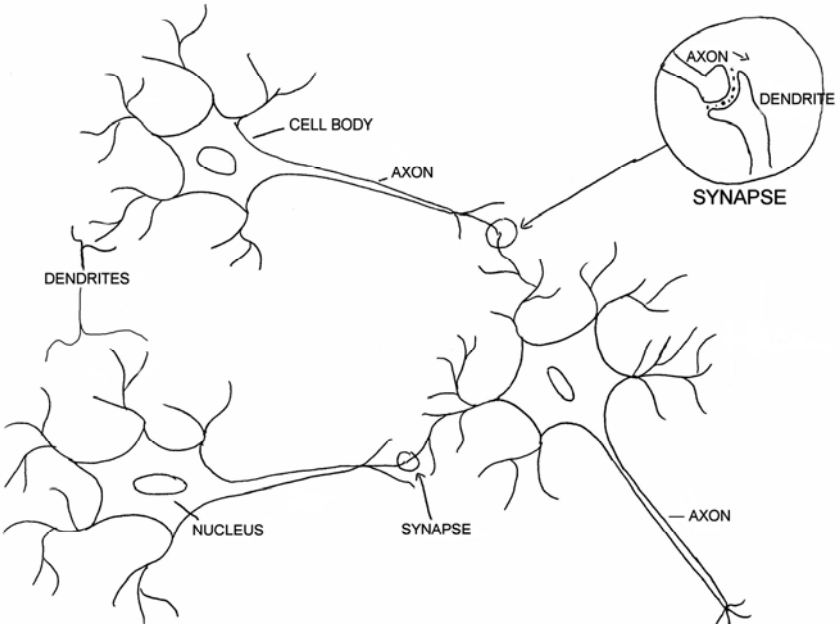

Figure 21: Schematic of a biological neuron

Signals can be encoded in the modulation of the pulsing frequency. Different types of neurotransmitters have different post-synaptic effects - they can even polarize the cell (inhibitors). Learning is basically the changing of synapses between neurons: their generation and destruction, and their change in strength, defined as their ability to polarize or depolarize the post-synaptic neuron. This and more leads to the complex signal processing of which the central nervous system is capable.

3.2 Mathematical theory

The mathematical abstraction of a neural network is defined (from [40]) as a directed graph with the following properties:

1. A state variable n_i is associated with each node i.
2. A real-valued weight ω_{ik} is associated with each link (ik) between nodes i and k.
3. A real-valued bias θ_i is associated with each node i.
4. A transfer function $f_i[n_k,\omega_{ik},\theta_i,(k\neq i)]$ is defined, for each node i, which determines the state of the node as a function of its bias θ_i, of the weights of its incoming links ω_{ik}, and of the states of the nodes connected to it by these links n_k.

3.2.1 Feed Forward Layered Networks

Feed forward layered networks are one subclass of neural networks. In this type of network, the information flows in only one direction. At one side of the network there is

the input layer – these nodes represent the inputs of the mapping represented by the network. The input layer feeds the next layer, which can be a hidden layer or the output layer. The last layer is the output layer. Any number of hidden layers can be used. In Figure 22 a feed-forward network with two hidden layers is shown. A feed-forward network with multiple layers is also referred to as a multi-layer perceptron. All possible connections are shown, but the connection weight matrix ω_{ik} will determine if a connection is present or not.

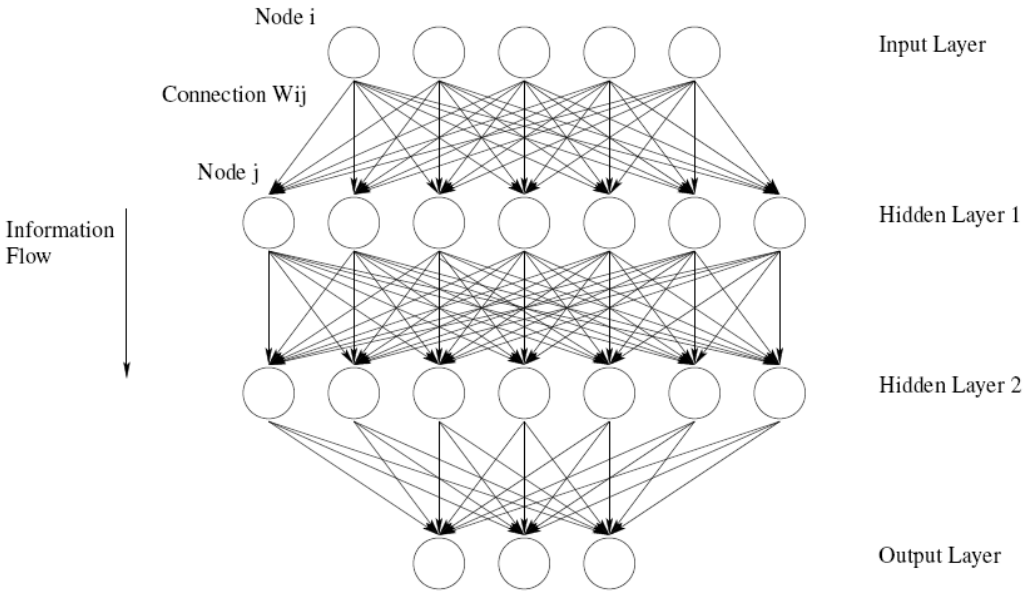

Figure 22: A feed-forward network with two hidden layers

How is the state or output of an individual node (neuron) calculated? It is almost equivalent to the biological case. First, the sum of the outputs of the nodes that feed into it multiplied by their respective connection weights is calculated. Note that in the case of a feed-forward layered network these nodes are the ones in the previous layer. After subtraction of the bias, the sum is fed into the transfer function of the node in question and the result of this gives the output of the node. If one takes as bias the weight of an extra node that always has unit output, the output of node *j* can be written as:

$$z_j = g_j(a_j) \tag{3-1}$$

with

$$a_j = \sum_{i=1}^{d} \omega_{ji} z_i + \omega_{j0} = \sum_{i=0}^{d} \omega_{ji} z_i \tag{3-2}$$

where d is the number of nodes in the previous layer, z_i the activation of node i, ω_{ji} ($i > 0$) the connection weight with node i, g_j the activation function belonging to node j and ω_{j0} the bias ($z_0 = 1$).

3.2.2 Network Topology and Activation Functions

The choice for a network topology, i.e. the number of layers and nodes, and the choice for the activation functions depends on the task that the network has to perform. The activation function commonly used for the hidden nodes in the network is the logistic sigmoidal function:

$$g(a) = \frac{1}{1+e^{-a}} \tag{3-3}$$

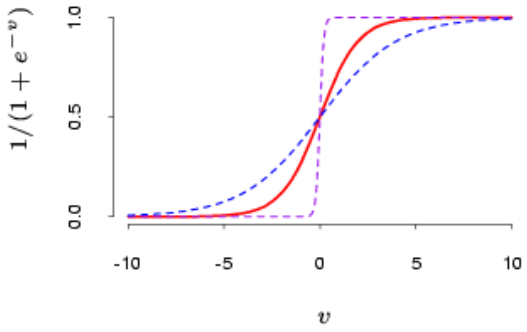

Figure 23: The logistic sigmoidal function

The logistic sigmoidal function maps the interval (-∞,∞) onto (0,1). Although this is not suitable for functional mapping, the function has properties that makes it a good choice for classification problems [41]. For functional mapping, the linear activation function is used:

$$g(a) = a \tag{3-4}$$

With one hidden layer of nodes and the choices stated above, any multi-variable function can be approximated to arbitrary accuracy (for proof see [41]). The error scales with $O(1/M)$, where M is the number of hidden nodes. Although one hidden layer is enough, the use of more tends to improve the efficiency of the network [42].

3.2.3 Training

Training of a neural network is basically minimizing an error-function by adjusting the free parameters in the model. The free parameters in a neural network are the connection strengths in the form of the matrix ω_{ji} and the biases θ_i. The biases can be replaced by the connection strengths ω_{j0} of a node that fires with unit strength.

For the training one uses a *training set* which contains n training patterns. A training pattern consists of the values of the input nodes x_n and the target output t_n. Each pattern is propagated through the network as described above. Generally this will yield an output y_n which differs from t_n. We suppose that the network error can be expressed as a differentiable function of the output variables so that:

$$E^n = E^n(y_1,\ldots,y_c) \tag{3-5}$$

for c output nodes. The total error function is the sum of the error functions of the

separate patterns:

$$E = \sum_n E^n \tag{3-6}$$

An error function typically used is the sum-of-squares error which, for a single input vector n, is:

$$E^n = \frac{1}{2}\sum_{k=1}^{c}(y_k - t_k)^2 \tag{3-7}$$

3.2.4 Backpropagation

There are numerous methods to minimize the error function. The method described in the following is the error backpropagation algorithm. In the following I will omit the subscript n for input and activation variables. The error function is minimized by using the derivatives of E^n with respect to the weights ω_{ji}. Note that E^n depends on the weight ω_{ji} only via the summed input a_j to node j. Using the chain rule for partial derivatives yields:

$$\frac{\partial E^n}{\partial \omega_{ji}} = \frac{\partial E^n}{\partial a_j}\frac{\partial a_j}{\partial \omega_{ji}} \tag{3-8}$$

We define:

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} \tag{3-9}$$

Using (3-2), we can write:

$$\frac{\partial a_j}{\partial \omega_{ji}} = z_i \tag{3-10}$$

Substituting (3-9) and (3-10) into (3-8) we obtain:

$$\frac{\partial E^n}{\partial \omega_{ji}} = \delta_j z_i \tag{3-11}$$

The values of z_i are known by the propagation of the pattern through the network. In order to get the derivative it is only necessary to calculate δ_j for each hidden node and each output node in the network. For the output nodes this is straightforward:

$$\delta_k \equiv \frac{\partial E^n}{\partial a_k} = g'(a_k)\frac{\partial E^n}{\partial y_k} \tag{3-12}$$

where z_k from equation (3-1) is denoted by the output activation y_k. Expressions for $g'(a)$ and $\frac{\partial E^n}{\partial y_k}$ can, for example, be found from equations (3-3), (3-4) and (3-7).

To obtain the δ's for hidden nodes we use the chain rule:

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k}\frac{\partial a_k}{\partial a_j} \tag{3-13}$$

The sum runs over all nodes k to which node j has output connections. Using definition (3-9) and formulae (3-1) and (3-2) we get the *back-propagation* formula:

$$\delta_j = g'(a_j)\sum_k \omega_{kj}\delta_k \tag{3-14}$$

By applying this formula recursively we can obtain all the δ's for the hidden nodes and input nodes by propagating the error backwards through the network. The weights are updated in the following way:

$$\Delta\omega_{ji} = -\mu\frac{\partial E}{\partial \omega_{ji}} \tag{3-15}$$

where μ is the step size and typically has a value between 0.1 and 1. The gradients are then recalculated with the new values of ω_{ji} over many iterations and the error converges to the smallest possible value.

This searching for the minimum in weight space has several problems. The local gradient does not always point to the minimum and the search could end in a local minimum. A method to overcome this problem is to add a momentum term to the gradient descent formula. This effectively adds inertia to the motion through weight space. It diminishes oscillations and helps to overcome local minima. The modified update formula is:

$$\Delta\omega_{ji}(t+1) = -\mu\frac{\partial E}{\partial \omega_{ji}} + \eta\Delta\omega_{ji}(t) \tag{3-16}$$

where η is called the momentum parameter.

The Levenberg-Marquardt algorithm[43] is a pseudo second-order method, estimating and taking into account the second derivatives of the functions to be optimized. The main benefit of this algorithm is the faster convergence time. Details of this method can be found in [44]. This was the primary method used for training the networks.

3.2.5 Data Samples

To train and test a neural network, several data samples are needed. A data sample consists of *patterns*. A pattern consists of data that are the network input and intended output variables.

- The *training set* is used for the actual training of the network. The error on this set is minimized during the training process.
- The *validation set* is needed to monitor the performance of the net on patterns that are not used during training. This is to verify the ability of the network to

generalize and to avoid overfitting on the patterns from the training set.

- A *test set* is used after training to measure the performance of the network.

It is important that the training set is different from the test and validation sets. The last two can be the same.

3.2.6 Training example

Training the networks adequately generally required around 1000 epochs. Figure 24 shows an example of a training run. The blue plot measures the mean square error (MSE) in each of 1000 epochs. The green dashed plot is the MSE of the validation set. It is clear to see that the training set outperforms the validation set, since the ANN learns to identify features specific to this set. Displaying the validation set performance helps to make sure that the net is not overtrained. Overtraining occurs when the network learns to choose specific events, and does not generalize the classification rule. When overtraining occurs, the performance of the validation group will decrease while the performance of the training group increases.

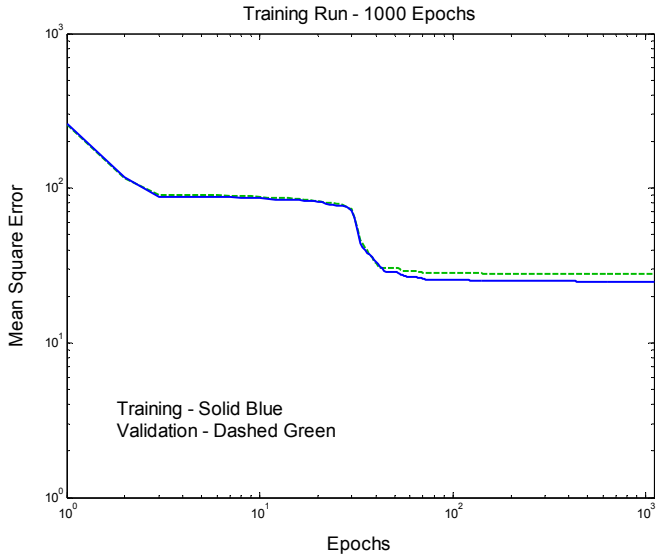

Figure 24: ANN training example

Typically:

- The training set outperforms the validation set.
- Overtraining never occurs.
- Most of the learning occurs during the first 100 epochs.

But, there were occasional exceptions:

- Sometimes the training did not converge at all – in these cases the training algorithm is probably stuck in a local minima.
- Rarely, the main part of the learning took place after 1000 epochs.

4. P_T reconstruction and triggering using an ANN

The task of extracting interesting physical events from the myriad of interactions taking place in the ATLAS detector is a complicated one. A seven orders-of-magnitude reduction is necessary (1GHz->100Hz). Much of the interference comes from background radiation and minimum-bias events. Luckily there are features that help to distinguish the few interesting events from the common uninteresting ones. High energy muons are one of the indications for interesting processes (see Figure 25 showing the P_T distribution of muon from different decay channels.Most of the muons originating from events of less interest such as shower muons or punch through, reside mostly in the low P_T region, below 10 GeV). For this reason the triggering mechanism is designed to tag high-P_T muons.

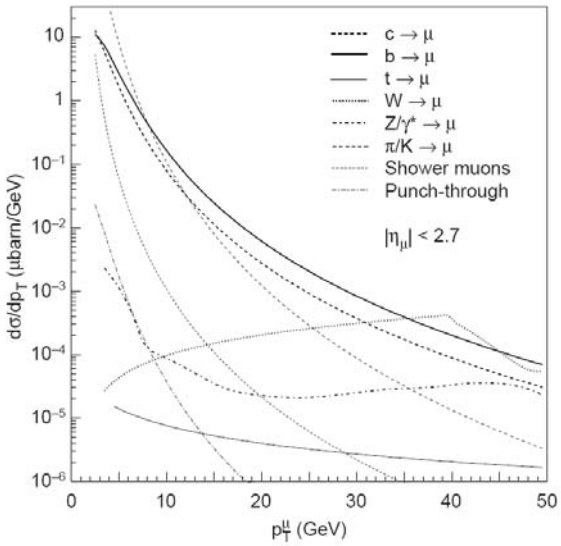

Figure 25[45]: Muon cross-sections for the different physics channels as a function of P_T at production

The triggering scheme for the muon spectrometer is a simple one: low-P_T muon tracks will bend more in a strong magnetic field. Thus, a strong magnetic field is applied, and the deviation from the infinite-momentum track originating from the interaction point is measured. The triggering scheme is described in detail in section 2.12. This is a simple and logical scheme, but in reality there are a few setbacks:

- The material through which muons pass on their way to the detectors induces multiple scattering processes. The material in front of the trigger chambers is that of the magnetic toroids, the inner detector and the hadronic and electromagnetic calorimeters. The total material between the IP and TGCs varies between 10 to 15 absorption lengths.

- The ATLAS magnet system consists of three air-core superconducting toroids. Due to the finite number of coils, the field configuration is not perfectly toroidal. This inhomogeneity, which is enhanced in the transition region between the barrel

and the endcap, significantly impacts upon the resolution of momentum measurements.

These considerations limit the resolution of the triggering mechanism. Limited resolution results in limited purity and efficiency of the trigger, which can have a direct impact on the effective luminosity possible at the LHC (see also section 2.14).

In this chapter, I will describe a study conducted on an application of Artificial Neural Networks (ANNs) to this problem. ANNs may present a feasible solution to the triggering mechanism, possibly improving on the resolution of the existing trigger algorithm.

The organization of this chapter is as follows: First, a general architecture is chosen and implemented for a net designed to estimate the P_T of generated muon events. This choice is based on the results of previous studies in this field. Different factors affecting the capabilities and performance of the net are analysed, and an optimal net is constructed. Next, the net is applied as a triggering network, and its performance is compared to a simulation of the dedicated hardware algorithm. Finally, other possible applications of an ANN to this problem are analysed and discussed.

4.1 ANN architecture

The chosen network topology is a feed-forward network with two hidden layers (see Figure 22). This enables the net to generalize complex problems, but leaves the simulation times reasonable. The net was created and trained using the Matlab Neural Networks Toolbox Version 4.0.

Our network architecture includes four input neurons. The reason for this choice and the appropriate input parameters and preprocessing choices will be discussed in the following sections. All neurons in the two hidden layers have sigmoidal transfer functions (see Figure 23). The choice of the number of neurons in each hidden layer will also be discussed. The output structure of the network includes one output neuron with a linear transfer-function. This output neuron will first be used to convey the estimated P_T of the selected muon, and later to output the triggering decision.

4.2 Transverse momentum estimator

In [46,47], it was shown that a 4x7x7x4 ANN is able to estimate the P_T of simulated muons. The study used DICE-generated (see section 2.15) positive and negative charged muons in the momentum range 1-50 GeV in the end-cap region $|\eta|>1.05$. The four input neurons were fed with the slopes and intercepts of the muon tracks in the *x-z* and *y-z* planes (this will be elaborated upon in section 4.4).

The output four-neuron vector included the charge and direction of the muon, but this is irrelevant to the current study – in our study the single output neuron represents only the estimated P_T (this is the only parameter relevant to triggering). The net was

trained using 1400 events, for several thousand epochs, and then tested on a set of 1829 events.

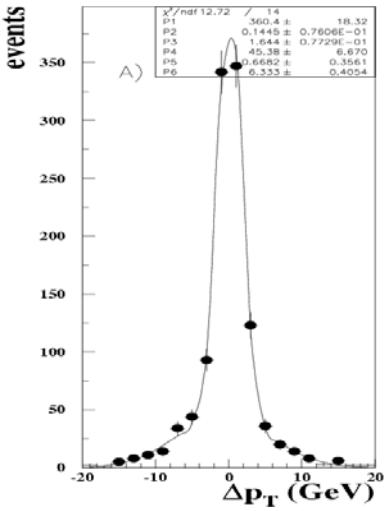

Figure 26: P_T reconstruction deviation histograms – previous study

The results of the study described in [47] concluded that the ANN is successful in reconstructing the momentum and charge of muons. The deviation of the estimated momentum from the real momentum is shown in Figure 26. The width of the gaussian shown in the figure is 1.6 GeV. The study also showed that a net trained specifically for the task of triggering performed better at this task than a net trained to reconstruct the P_T.

4.3 Hidden layers

In order to study the effect of the number of neurons in each hidden layer, we performed a set of runs with nets of varying size. All the nets received the same input format, and ran on similar data sets. The training sets consisted of 1333 samples, the validation sets contained 666 samples, and the test sets 2000. In our study, all events are generated by DICE in the momentum range 1-50 GeV in the end-cap region $|\eta|>1.05$.

In these runs, the samples were confined to a small *x*-*y* area in the M1 plane (around $|\eta|\sim1.8$), in order to avoid the geometrical effects of the detector (see section 2.14). Figure 27 shows the selected events within the relevant octant.

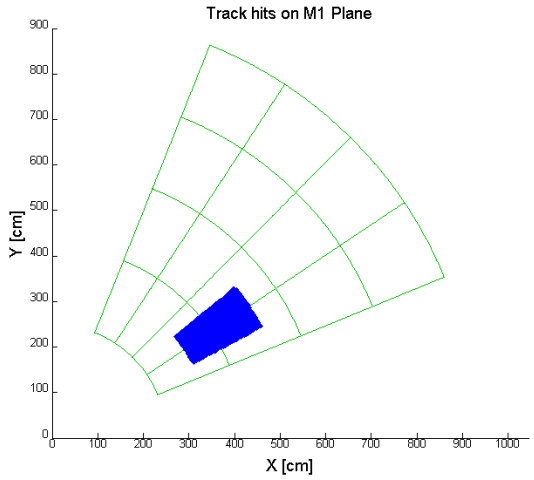

Figure 27: Distribution of hits in M1 plane

Nets containing 1 to 12 hidden neurons in each of two hidden layers were tested. The nets were trained, simulated, and the width of the gaussian fit to the momentum estimation error plot was calculated. For illustrative purposes, Figure 28 shows the performance of a "good" net with seven neurons in each hidden layer, achieving a width (σ) of 0.46GeV, and a "bad" net with seven neurons in each hidden layer, achieving a width of only 0.94GeV.

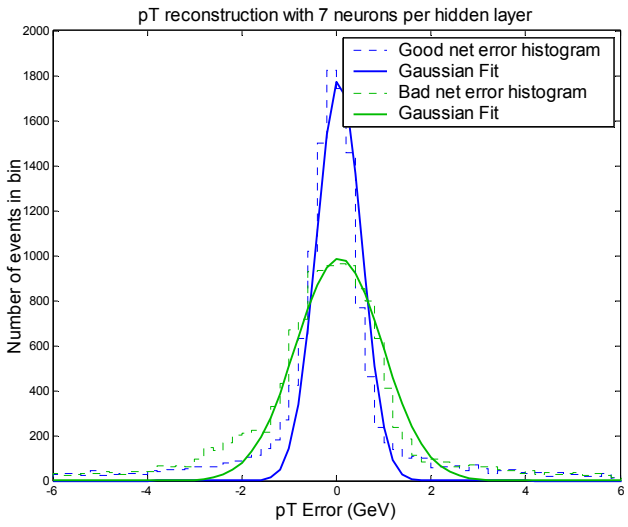

Figure 28: P_T error histogram for two nets

Fifty training and simulation runs were conducted for each net size (1-12). The results are given in Figure 29. The blue solid line indicates the performance (σ) of the best net achieved for each net-size, and the square (green) set indicates the average performance of the fifty nets tested in each category. The error bars indicate the standard deviation of the results in each set. The trained nets do not always converge – nets that did not converge were removed from the statistics. The red dashed plot indicates the percentage of nets trained that finally converged.

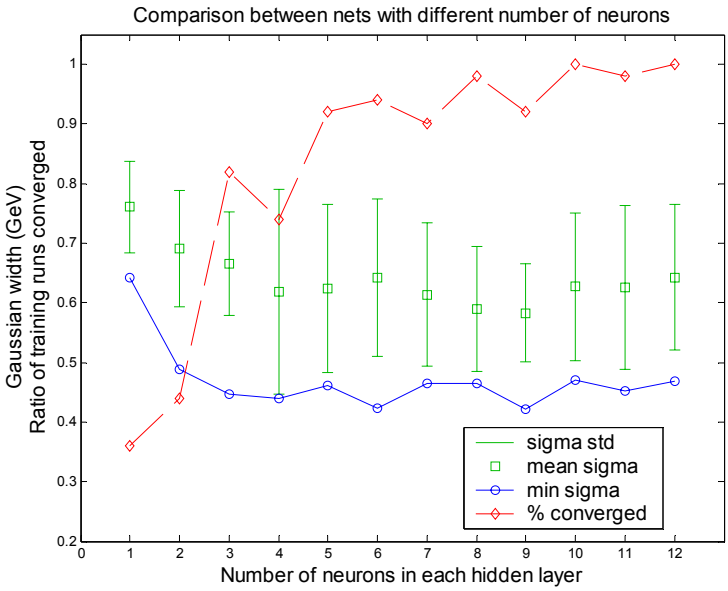

Figure 29: ANN performance with respect to the number of hidden neurons

It turns out that the minimum achievable gaussian width is less than 0.5GeV even for a small net containing two hidden layers with two neurons each. Although this is the case, small networks displayed erratic behaviour in terms of convergence. In contrast, large nets tend to converge, but are complex and consume long computation times. Thus for the remainder of the study we chose a medium net containing two hidden layers of seven neurons each. Notice that one training run on an ANN does not guarantee optimal results. This is a known drawback to using Neural Networks.

It should be mentioned that a large part of the error in the reconstruction of the momentum is due to the real stochasticity of the data. Perhaps this is the reason even the simplest nets can perform adequately.

4.4 Input vector and preprocessing

The raw data generated by the DICE simulation program is a set of events, each including 0-28 hits in the different layers of the detector. There are areas in which different elements in the same layer overlap. This, combined with inefficiencies in the TGC detector, causes each event to produce a different number of hits (see section 2.15). Feeding the hits directly into a net may be possible, but would require a much more

complicated net. In order to be able to work with a simple net, the need arises to canonize all tracks to a single representation.

In the end-cap region, the trigger chambers are located outside the toroidal field[48]. This, combined with the fact that we are dealing with energetic muons, allows us to assume that a muon track in this region follows a straight line. The input vector should represent a road – a straight line estimating the muon's flight path through the detector planes. Earlier studies [46] made use of the Hough transform in order to canonize the hits into a straight line, but it was established that a least-squares approximation algorithm is sufficient. We used the simpler least-squares approximation as the preprocessing stage in our study. Figure 30 shows the pre-processing
results on 2 events:

- A high-energy (41GeV) muon is approximated by a red (solid) line.
- A low-energy (4.4GeV) muon is approximated by a green (dashed) line.

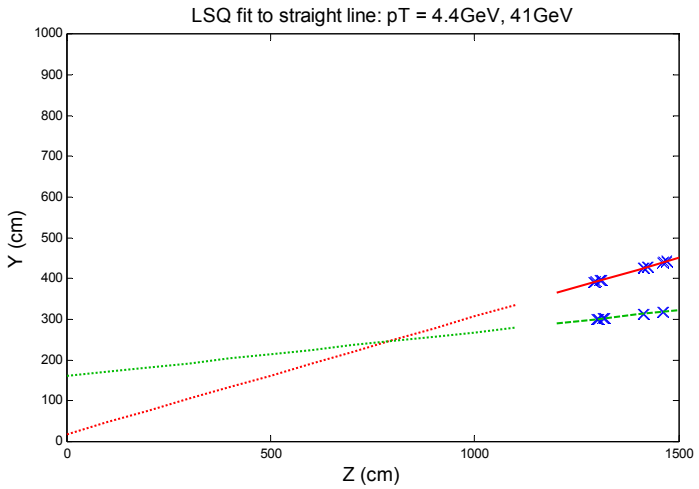

Figure 30: LSQ fit to a straight line

The interaction point is at (0,0,0) – thus it is evident that the "green" muon has been deflected to a greater degree than the red one. The dotted lines extrapolate the linear fit to z=0 in order to show this. The network receives four input parameters: The slope and intersection in the y-z plane (shown here), and the same parameters in the x-z plane.

Choosing this input structure defines the parameter hyperspace in which the network must make cuts. Figure 31 shows the parameter distribution in the slope plane and the intercept plane. Red (circled) dots are low-P_T (<6GeV) muons, and blue dots are higher P_T (6-50GeV) muons. The network is expected to perform a cut in the 4-D hyperspace – removing as many red events, while leaving maximum of the blue events.

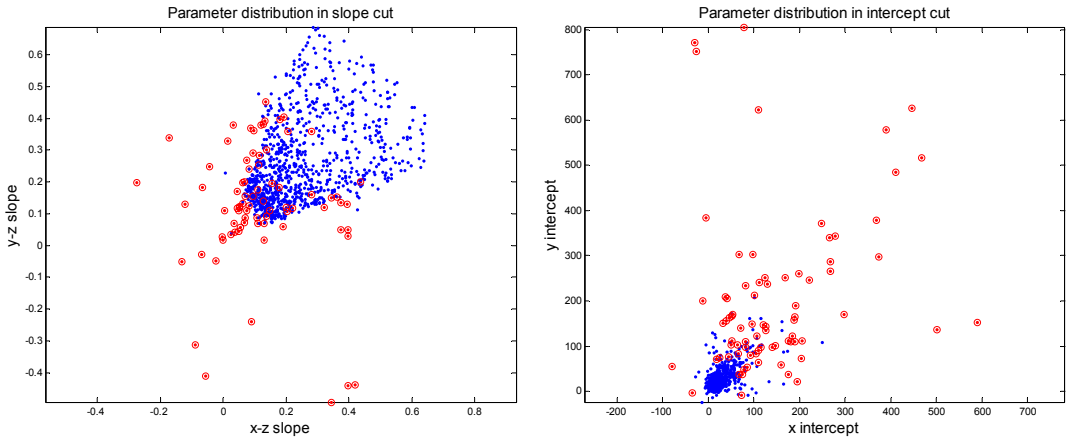

Figure 31: Slope and intercept parameter distribution

Since it is known that ANNs tend to be affected by the choice of inputs and by pre-processing choices, a few other input choices were studied:

- A six-neuron input layer, receiving second-order approximation parameters. This was attempted in order to account for residual magnetic field present in the space between the TGC planes. If this magnetic field was non-negligible, the muon tracks could not be approximated to a straight line. It was shown that this is not the case – the six-neuron input layer did not improve the performance of the net.
- In another test, the interaction point (0,0,0) was added to the second-order approximation, but this degraded the performance further.
- Transforming the inputs to polar coordinates, in accordance with the detector geometry. This produces slope and intercept values in r and Φ, which may be a more natural representation. This did not affect the performance of the net.
- Inputs similar to the ones used by the dedicated hardware algorithm: δr and $\delta\Phi$ residuals. This study is summarized in Appendix C.

Of all input structures attempted, the four-neuron slope and intersection structure described above proved to show the best results. The rest of this study is conducted with such a net.

4.5 Triggering network

The output structure of the net must also be chosen optimally. In [47], two types of output structures were studied: a four-neuron linear P_T and charge estimator, and a one-neuron tan-sig threshold trigger. The one-neuron threshold trigger structure was shown to perform better at the task of triggering. This is the structure that we chose for this study (although we chose to use a linear output neuron). This means that the net will not output the estimated P_T of the muon, but a value indicating if the event should be tagged. The value is compared to a threshold, producing a boolean value – trigger or no trigger.

The dedicated hardware trigger, described in section 2.12, can trigger on any of six different thresholds: three low-P_T thresholds, and three high-P_T ones. Throughout our study, we focused on the 6 GeV low-P_T threshold. The low-P_T thresholds are more important for removing background noise, since the background rate diminishes as P_T rises (see section 4.9). We assume (and verify) that results obtained for a 6 GeV threshold will be applicable for higher thresholds as well.

The triggering network is trained on 1333 events and validated on 666, exactly like the P_T estimating net, and it is tested on some 67000 events. In order to quantify the performance of the net, we plot the distribution of events that the net misclassified. That is, events generated with P_T>6GeV that did not cause the network to trigger, and events with P_T<6GeV that caused it to trigger.

The dedicated hardware trigger is divided into regions in η-Φ where independent thresholds can be set, as described in 2.13. It is logical to believe that the triggering network also performs better when it is trained and tested on events from a localized region in the detector. Figure 32 shows the results of two nets: The green (dashed) net was trained on events from a whole sector. The blue (solid) net was trained on the area shown in Figure 27.

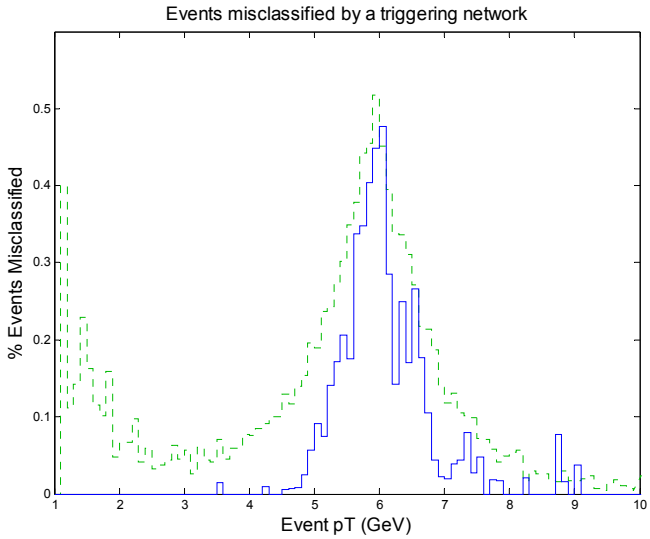

Figure 32: Events misclassified by a triggering network

It is obvious that the blue net performs better. This encourages us to work with local nets – each net specializing in a different region of the detector.

From Figure 32 we also deduce that we cannot naïvely assign a threshold of 6GeV to the net: in order for it to meet with the demands, the net must trigger on at least 90% of events[49] with an actual P_T of 6GeV (the physical acceptance should also be taken into account, but this is not relevant here). This condition is necessary in order to minimize the loss of interesting physics. There are two ways to control the threshold of the trigger

using an ANN. One way is to retrain the net, specifying a different threshold than the objective. The other way is to use the same trained net, but set a different threshold for the output neuron alone.

Figure 33 compares between these two methods. The solid blue plot shows the acceptance of the ANN (% of events accepted in the range 5.8GeV-6.2GeV). The dashed green plot shows the purity in this range (% of events with P_T>6GeV out of all events triggered). The left diagram was obtained by variation of the output neuron threshold. The diagram on the right was obtained by re-training each time with a different trigger threshold objective. This takes more time, and gives very similar results.

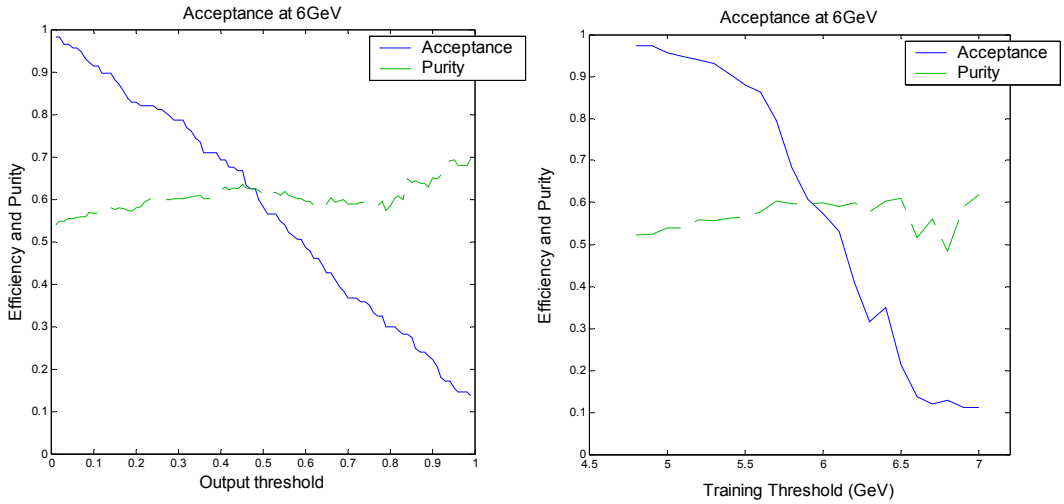

Figure 33: Acceptance and purity for different thresholds

We can see that in the left plot, the efficiency changes linearly, providing a convenient tuning method. This means that for each trigger the net must be trained only once, and slight modifications to the trigger threshold can be made by controlling a parameter. From Figure 33 we can see that to get 90% acceptance at 6GeV, we need to train the net to trigger at 6GeV, but set the output threshold to 0.11. Alternatively we can train the net to trigger at 5.33GeV, and leave the output threshold at 0.5. Changing the output threshold is a convenient method for fine-tuning, but the effect of this tuning is limited, and sometimes retraining will be necessary.

The purity plots in Figure 33 are not very informative. The test set is confined to the region 5.8GeV<P_T<6.2GeV, and as a result of this the purity is dominated by the real stochasticity of the data, and does not vary much. Actually, the selected threshold influences both the efficiency and the purity achieved by the net. A plot of the efficiency and purity for different thresholds in the region 4GeV<P_T<8GeV is shown in Figure 34. As expected, efficiency and purity are complementary.

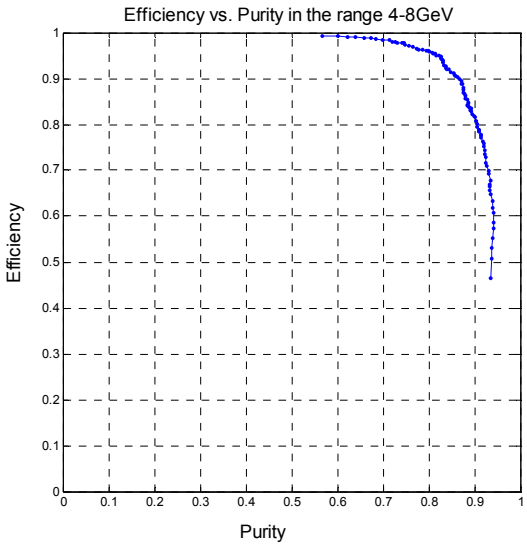

Figure 34: Efficiency vs. Purity

Figure 35 gives the misclassification histogram after setting the output threshold of the net to 0.11. Notice that now almost all misclassified events are events with P_T<6GeV that activated the trigger, i.e. the acceptance is high (90%).

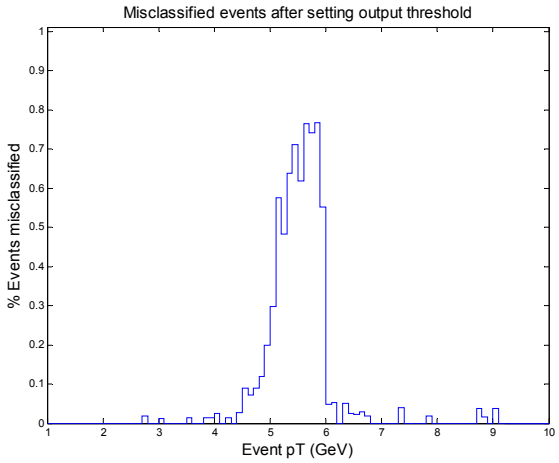

Figure 35: P_T distribution of misclassified events after setting threshold

4.6 Comparison with ATRIG

After choosing architecture and studying the various parameters that affect its performance, we proceed to compare its performance to that of the dedicated hardware trigger. The network we are using is the linear trigger output model, trained on a confined region of the detector (Figure 27), with an output threshold selected to give an acceptance of 90% at 6GeV. We used the ATRIG program (see 2.15) to simulate the behaviour of the hardware trigger. All events generated by DICE were passed to ATRIG, and the results were plotted.

Figure 36 shows the performance of the ANN compared with the ATRIG simulation. The plot on the left shows the histogram of misclassified events – both algorithms accept almost all of the events with P_T>6GeV, but the ANN performs better than ATRIG on events with P_T<6GeV. The plot on the right shows the percent of events triggered with respect to P_T. This is essentially the same plot, but it is illustrative for the next section. A test set of 14000 events was used to generate the plot.

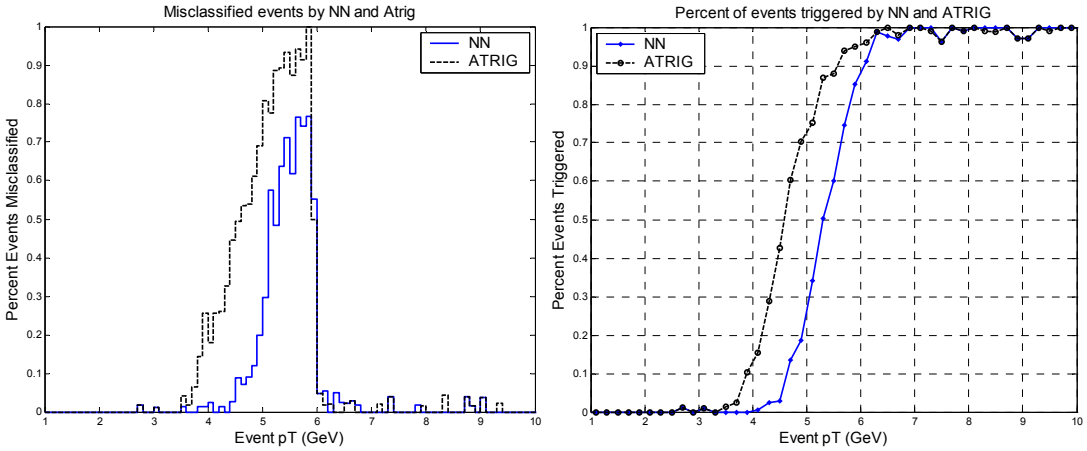

Figure 36: Comparison of ANN and ATRIG

Figure 37 plots the trigger functions obtained with the ANN and with ATRIG. The functions are fitted with a sigmoid of the form:

$$\frac{1}{1+e^{-a(x-b)}}$$

This sigmoid has two parameters: the slope (a) and the intersection point (b). The solid blue sigmoid is a fit of the ANN performance, and the dashed black one is a fit of the ATRIG performance. The dotted green sigmoid is a shifted ATRIG fit, plotted over the ANN fit (intersecting it at 6GeV) in order to emphasize the difference in the slope parameter.

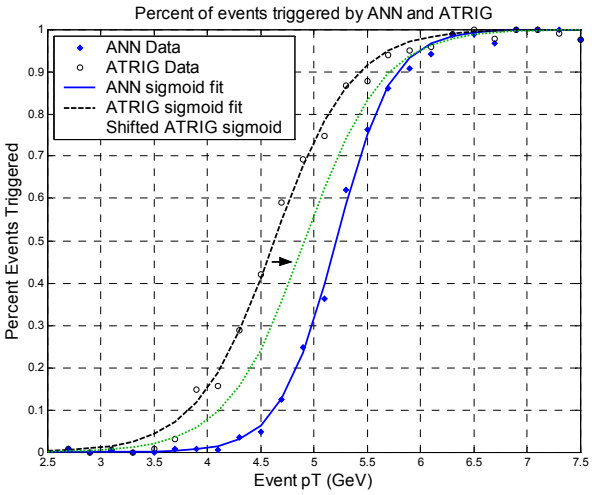

Figure 37: Comparison of ANN and ATRIG - sigmoid fit

Studying this figure, we note that the ANN outperforms ATRIG in two respects:

- The ANN is tuned to ~95% acceptance at 6GeV, while ATRIG gives much higher acceptance (~97.5%). This has a very pronounced negative impact on the purity achieved by ATRIG.
- The ANN shows a larger slope. If both of the algorithms were tuned to give the same efficiency at 6GeV, this parameter would determine the better of the two, as illustrated by the dotted green plot in Figure 37.

It should be stated that the plotted ANN trigger function is the best of 100 training runs. This specific net was fitted by a sigmoid with a slope parameter of 3.81. The mean slope parameter was 3.56, with a standard deviation of 0.16. The ATRIG trigger function was fitted by a sigmoid with a slope parameter of 2.75, ±0.15 with 95% confidence.

ATRIG	Slope = 2.75±0.15 GeV^{-1}
ANN	Slope = 3.56±0.16 GeV^{-1} Max = 3.81 GeV^{-1}

Tuning the ANN to a desired acceptance at 6GeV affects the slope of the trigger function, but this effect is small. This is shown in Figure 38 - the slope improves as the acceptance grows, but in this case the purity will diminish.

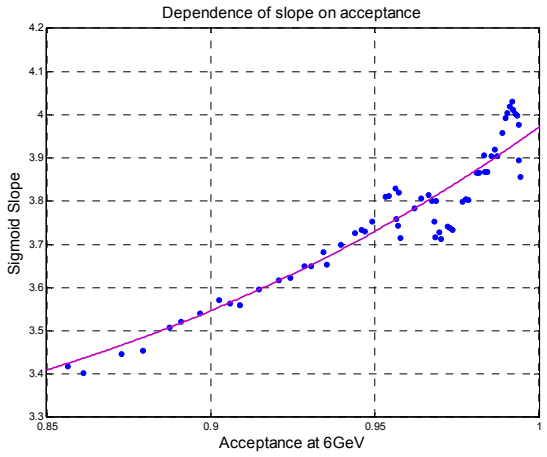

Figure 38: Dependence of slope on acceptance

We will assume that like the ANN, it is possible to tune ATRIG to any desired acceptance. Therefore in this respect, the improvement shown by the ANN is an artefact of different tuning considerations.

4.7 Effect of resolution

In this section we will show that the improved slope performance of the ANN arises due to the superior resolution of the ANN simulation. In the simulation, we have the privilege of preprocessing: generating a linear fit to all the hits in the different detector layers. This fit can be produced with infinite precision, whereas in the dedicated hardware trigger, the window resolution is that of the TGC – around 1cm (see section 2.9).

This limited resolution introduces "quantization noise" into the slope measurement, and can affect the trigger performance. We try to mimic this effect by introducing artificial quantization noise into the ANN inputs. To this effect all parameters were modified by a random, limited percentage:

$$\frac{dx}{x} = rand \cdot noiseLevel \qquad (-1<rand<1)$$

Introducing noise into the ANN inputs has the expected effect – the sigmoid broadens, and its slope lessens. Figure 39 shows the slope parameter obtained with respect to the noise level introduced. The ANN chosen for this run is one with average performance. The ANN and the data set were kept constant, except for the random variation of the input parameters. The dashed green line is the slope of the ATRIG trigger function (constant). The solid red line is a fit to the results obtained.

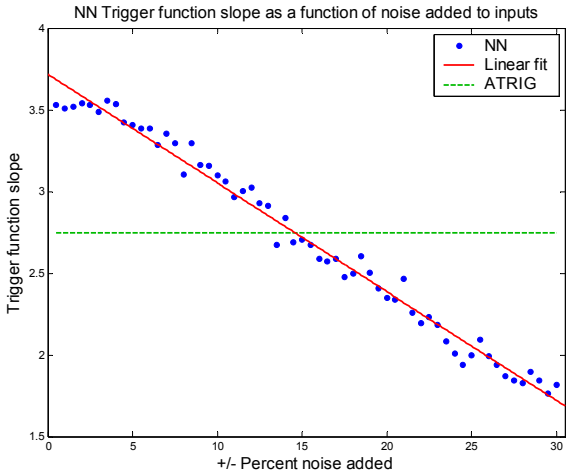

Figure 39: ANN trigger function slope vs added noise

From Figure 39 we see that the ANN obtains a slope similar to ATRIG after a ±15% noise level introduction. A histogram of the intercept parameters for events with a P_T of 4-6GeV is shown in Figure 40.

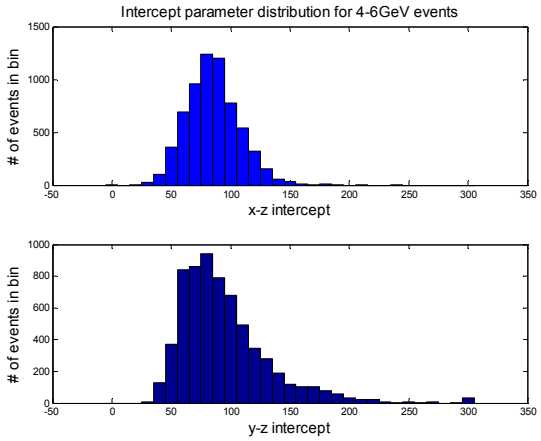

Figure 40: Intercept parameter distribution

From this histogram we can deduce that the scale of the deflection for critical events is around 100cm. This translates to a scale of around 10cm at the detector (due to the detector geometry – see section 2.12). A 15% error is equivalent to a ~1cm error at the detector, which is the level of the measurement resolution of the TGC.

4.8 Applying an ANN after ATRIG

Another relevant phase in which to implement the ANN is at Level 2 of the trigger, or as a Layer 1.5 after the dedicated hardware and before the more intensive computer processing. The ANN can offer a computation-cheap method of removing background events, thus saving computing power. This section describes a net trained for this purpose.

The same network architecture is used: four input neurons, two hidden layers of seven tan-sig neurons each and one linear output neuron. This time the network is trained only on events that have passed the ATRIG criterion. Thus, the ANN can specialize in rejecting the background that ATRIG failed to reject. In this case, in order to tune the net to 90% acceptance at 6GeV, it was necessary to train the net to a 5.3GeV P_T threshold, and fine-tune the results with the output neuron threshold. Tuning only the output neuron threshold did not suffice due to the new histogram created by the ATRIG filter.

Fifty ANNs were trained in this fashion, and each trigger function was fitted with a sigmoid. The resulting plots are similar to the one shown in Figure 37. The resulting slope parameter was also similar: a mean of 3.49 with a standard deviation of 0.23. The following table summarizes the results. The maximum slope achieved is also given, since it is possible to train many nets and choose the best performing one.

ATRIG	Slope = 2.75±0.15 GeV^{-1}
ANN	Slope = 3.56±0.16 GeV^{-1} Max = 3.81 GeV^{-1}
ANN after ATRIG	Slope = 3.49±0.23 GeV^{-1} Max = 3.70 GeV^{-1}

The conclusion is that an ANN trained and applied on events passing the first level dedicated hardware trigger cut is able to improve the purity of the cut. This is again a result of the improved resolution, but such a resolution may be available at a higher trigger level – due to the increased availability of data, or due to preprocessing algorithms.

4.9 Background distribution

The differences in the slopes of the different solutions seem at first miniscule. The importance of a slight improvement in the slope of the trigger function can be better understood if the distribution of background events is taken into account. Figure 41 shows the expected rates of minimum-bias events at LHC energies as predicted by different versions of several simulation programs (Pyhtia, Rhojet, Herwig, Isajet).

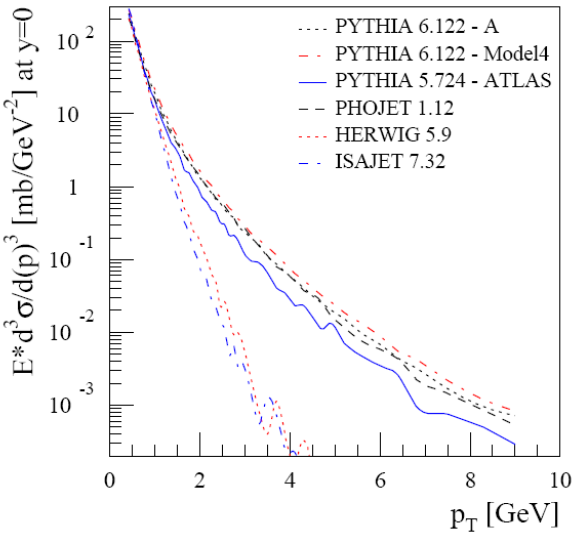

Figure 41[50]: P_T spectrum of charged particles in minimum-bias events

The rate of events falls quickly with rising P_T – in the vicinity of 6 GeV, the rate of events falls roughly like e^{-P_T}. This means that the background rejection ratio is strongly influenced by the slope of the trigger function near the threshold momentum. Figure 42 shows an estimate of the distribution of background events after passing selection by the different trigger functions shown in Figure 37.

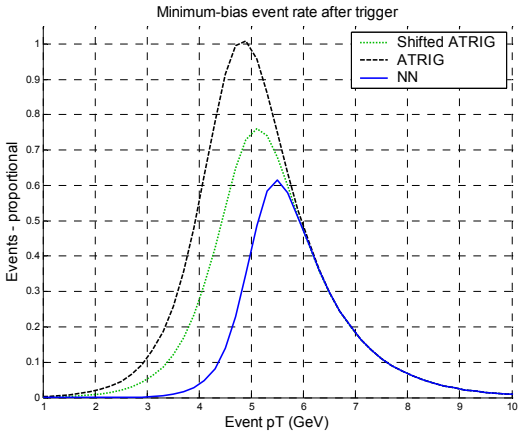

Figure 42: Background distribution after trigger

The dashed black plot depicts the minimum-bias event rate if ATRIG is used. The dotted green plot shows the improvement gained by shifting (tuning) ATRIG to accept 95% of real events at 6GeV. This reduces the rate of minimum-bias events by 25%. The solid blue plot is a maximal-slope ANN tuned to accept 95% of events at 6GeV. Using this mechanism reduces the rate by another 33%.

4.10 Geometrical effects

Up till now the set of events we have used has been localized to a small area of the detector in order to cancel out geometrical influences (Figure 27). In this section we will study how the comparison between ATRIG and the ANN varies with the detector geometry.

We choose events in a low-resolution region. The region $|\eta|\sim1.3$ with $|\Phi|<5.5°$ displays a large degree of magnetic field inhomogeneity, combined with increased coulomb scattering due to detector material (see section 2.14). We trained 100 nets in this region – a typical result can be seen in Figure 43. The scale is identical to Figure 37 for comparison.

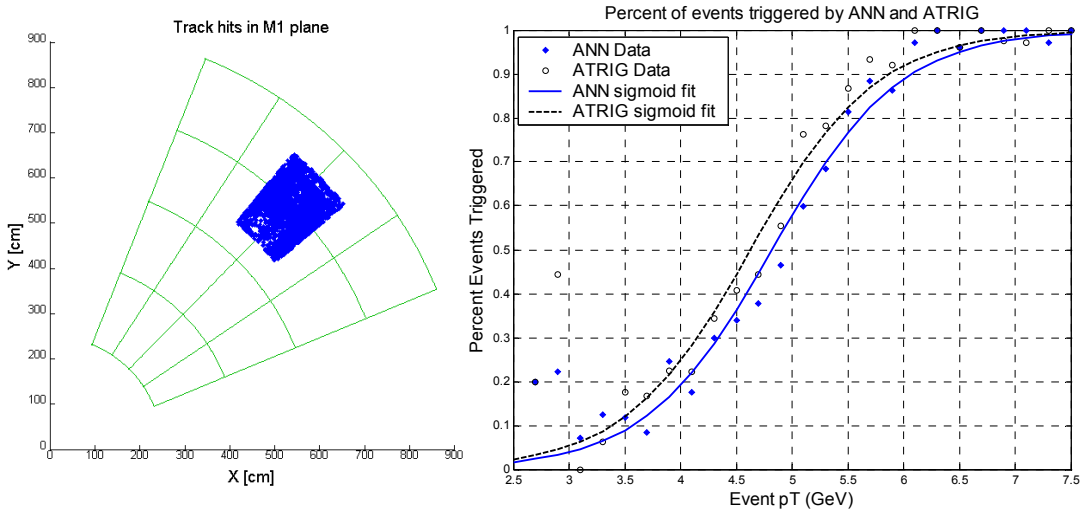

Figure 43: Performance in poor resolution area

The results indicate that the ANN does not generally perform better than ATRIG in this region. In this region both the algorithms perform poorly - the following table summarizes the results:

ATRIG	Slope = 1.77±0.14 GeV^{-1}
ANN	Slope = 1.6±0.19 GeV^{-1} Max = 2.01 GeV^{-1}

This indicates that the performance of the ANN is limited by the same factors affecting the performance of the dedicated hardware trigger.

4.11 Implementation options

In the previous sections we have shown that an ANN can perform at least as well as the dedicated electronics trigger, and in some cases better. Although the ANN relies on the same underlying classification principle, it does have certain inherent capabilities which may be put to use:

- Training – the net can be designed to adapt to changes in the detector geometry, changes in the magnetic field, etc.
- Simple low-resource hardware or calculation-cheap software realisation.
- Simple fine-tuning ability.

These benefits might give an ANN implementation an advantage over conventional methods in one of the following implementation alternatives.

4.11.1 Level 1

Implementation at Level 1 calls for a hardware realisation of the specific ANN. The simplicity of the ANN structure enables a fast hardware realization. This was directly demonstrated by the Pierre et Marie Curie University group which created and tested hardware dedicated machinery that successfully implemented the ANN structure used in this study [38].

In order to fully implement an ANN at Level 1, some form of fast hardware realisation would be necessary for the preprocessing stage. This stage transforms the numerous hits in the various layers into a linear track, using a least-squares approximation. Approximating the slope and intercept parameters from the two furthest hits simplifies this operation at the cost of reduced resolution – though we have shown that the improved performance of the ANN can be attributed to high resolution of the input parameters.

An ANN realisation that does not require a preprocessing step is discussed in Appendix C.

4.11.2 Level 1.5

Regrettably, implementing at Level 1 is probably not feasible at the current stage of progress at ATLAS. Another option is to implement the ANN in software between Level 1 and Level 2. Here the ANN can act as a computation-cheap algorithm for removing background events, reducing the amount of noise at the Level 2 input.

This implementation was discussed in section 4.8 and will require a higher resolution than available at Level 1. One of the benefits of realising a net at this triggering stage, even without improved resolution, is that it will allow simple fine-tuning of the Level 1 performance. Implementation at this stage does not require a hardware preprocessing engine, and coding in software automatically yields superior resolution.

4.11.3 Calibration aid

The basic underlying principal used by the ANN in order to select events is similar to the one used by the dedicated electronics (see section 2.12). That is, the ANN also makes the selection according to the deflection of the track from the infinite-momentum track. In order to show this equivalency, we will attempt to plot the virtual coincidence windows defined by the ANN.

An infinite momentum track (P_T>30GeV) is generated by DICE. This track has hits in M2 and M3, on a line crossing the IP. Now we generate new (virtual) events: The new tracks pass through the original hit in M3, but are offset in M2 (Figure 44: The blue line is the infinite-momentum track, and the green lines are virtual tracks). These virtual trajectories are fed into the trained ANN, and the resulting trigger decisions are mapped. In this fashion we can map a small area of M2 and generate a coincidence window around the infinite-momentum hit there.

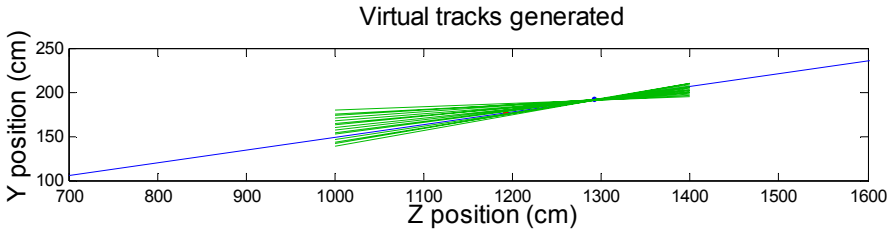

Figure 44: Virtual tracks

Figure 45 shows a coincidence window generated by a four-neuron per hidden layer ANN. The blue circle shows the infinite-momentum hit. The green squares are hits of events with P_T>6GeV. The red crosses are hits of events with P_T<6GeV. The blue line defines the $\hat{r}$ direction. The shaded areas are the ones passing the ANN selection (each pixel was generated as explained above).

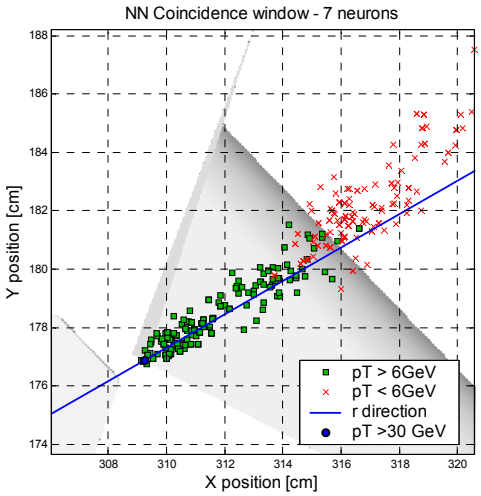

Figure 45: ANN coincidence window - no noise

We notice that we did not get exactly what we expected: we were expecting bound windows and we got extensive areas. The reason for this is the lack of background noise in the simulation – the ANN did not learn to reject sporadic background noise, especially in the $\hat{\phi}$ direction. In order to remedy this, we introduced synthetic noise – the results are shown in Figure 46.

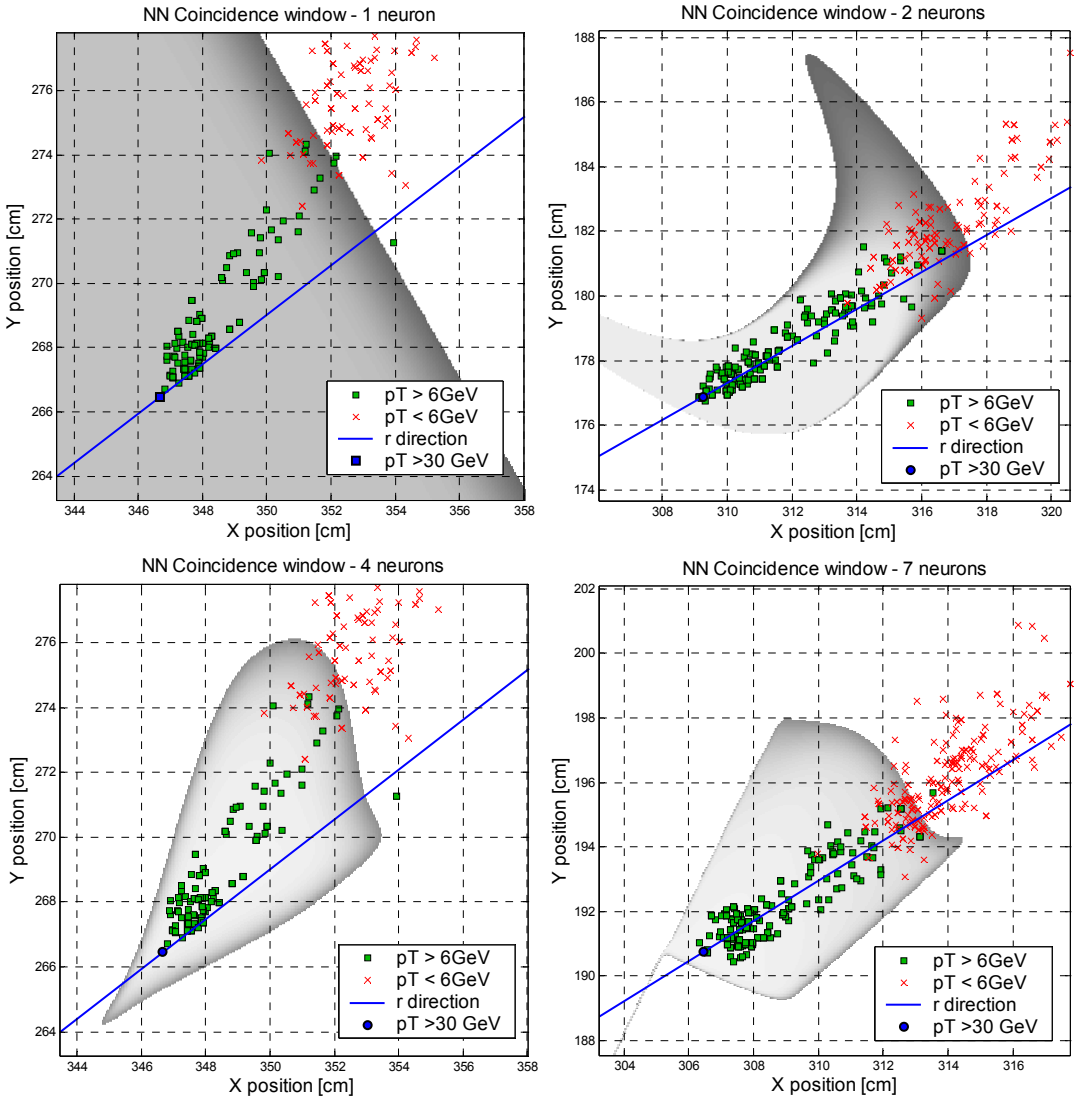

Figure 46: ANN coincidence windows for different neurons in the hidden layer.

In Figure 46 coincidence windows are plotted for four ANNs. A one-hidden-neuron per layer ANN can do no better than selection according to the deflection in $\hat{r}$. More intricate networks will create a closed window around the relevant events. Since the deflection is mainly in $\hat{r}$, we now understand why increasing the number of hidden neurons hardly has an impact on the performance. We can see that four neurons give a smooth contour, while seven neurons give an unnecessarily complex contour. Perhaps

more than four neurons will be necessary in order to allow positive and negative charge separation.

The above figures show that the basic underlying principle of the ANN and the dedicated hardware coincidence matrices is similar. Practically, these ANN-generated coincidence windows may be used to tune the dedicated hardware windows. It should be taken into account that the shape of these windows will vary with each ANN training run.

5. Discussion

Several artificial neural network architectures were studied for the purpose of identifying and selecting muons with a P_T above a defined threshold in the ATLAS end-cap. The current implementation of the trigger algorithm is based on coincidence windows implemented in dedicated hardware. Due to the complex nature of the inhomogeneous magnetic fields, and due to the coulomb scattering in the calorimeter, the performance of the trigger is limited. It was hypothesized that an ANN will outperform the existing algorithm.

An optimal ANN was chosen after studying the effect of different parameters on the ANN performance. Feed-forward nets containing two hidden layers with a variable number of hidden neurons were studied. It was shown that all nets with more than two neurons in each hidden layer managed to generalize and solve the problem adequately, but larger nets converged more often. These runs were conducted on P_T-estimating nets, while the rest of the study was conducted on triggering nets tuned to a threshold of 6GeV.

Different input structures were examined, and finally a four input-neuron structure was chosen. The network receives four input parameters: the slope and intersection in the *x*-*z* plane, and the same parameters in the *y*-*z* plane. These parameters are calculated by applying a linear LSQ fit to all available hits in all detector layers. This fact effectively increases the measurement resolution available to the ANN compared to the information available to the dedicated hardware.

The nets were applied to a confined area of the detector in order to avoid low-resolution areas. The neural net managed to generate a trigger-function superior to that of the dedicated hardware in two aspects: fine-tuning of the acceptance at 6GeV to 90%, and an increased slope. The importance of the slope parameter is critical due to the distribution of background events. A simple method for fine-tuning of the acceptance was introduced, and it was shown that tuning the acceptance has a negligible effect on the slope of the trigger-function.

In order to study the effect of the resolution on the ANN performance, we chose to mimic the effect of degraded resolution by adding noise to the ANN input parameters. It was shown that adding ~15% noise to the inputs causes the ANN performance to be equivalent to that of the dedicated hardware. This amount of noise is equivalent to the ~1cm resolution of the TGCs, implying that the improved ANN performance is due to the improved resolution available off-line. A δr-$\delta\Phi$ net described in Appendix C, using inputs similar to the dedicated hardware inputs, did not exhibit improved performance.

Implementing the ANN as another trigger level after the Level 1 dedicated hardware may improve on the Level 1 performance if superior resolution will be available – either due to new data or preprocessing. Implementing such a layer would also enable a simple method for fine-tuning the threshold.

In order to test the hypothesis that an ANN will outperform the dedicated hardware algorithm especially in low-resolution areas, we tested the net on such a region of the

detector. Both the ANN and the dedicated hardware simulation performed poorly, and thus the hypothesis was disproved.

In the final section we suggest using the ANN as a method of calculating coincidence windows for use in the dedicated hardware trigger. Through this experiment we gain a deeper understanding of the ANN structure. Since no noise was introduced, the ANN performed a cut mainly in the $\hat{r}$ direction. After remedying this effect, it was shown that a network with four hidden neurons per layer is sufficient for this purpose. It is also interesting to see that the bending occurs in both directions, and that high-P_T events are less affected by coulomb scattering.

This study focused only on the end-cap of the ATLAS detector and on the TGC chambers within it. Applying the same methods to the barrel region and its RPC chambers should yield similar results. This study simulated only one octant of the TGC detector, and did not take into account accurate noise models. Only one muon was generated at a time, and thus the capacity to register multiple muons was not studied. Only negative-charge muons were used, and the net was not trained to identify the charge of the muon. Also, the ATRIG simulation used was an old one. All of these reservations should not affect the validity of the results.

The results of this study suggest that the dedicated hardware algorithm is adequate as the triggering mechanism in ATLAS. The proposed ANN uses the same features used by the dedicated hardware algorithm in order to perform the classification, and therefore cannot outperform it. If increased trigger performance is desired, increased resolution is necessary. For example it is possible to use all hits in the M2 and M3 planes and implement an LSQ linear fit in hardware. Another example is the use of the triplet plane M1 for the low-P_T trigger. Applying an ANN can be used in conjunction with increasing the resolution and can introduce a new way of tuning the trigger threshold.

6. Appendices

Appendix A Trigger scheme

Figure 47 shows the triggering planes in the barrel and in the end-cap. In the end-cap, the furthest plane from the interaction point (TGC3) is used as the pivot, TGC2 gives the measurement for the low-P_T trigger, and TGC1 gives the measurement for the high-P_T trigger. In the barrel, the closest plane (RPC1) is used as the pivot plane, and the other RPCs give a low- and a high-P_T measurement.

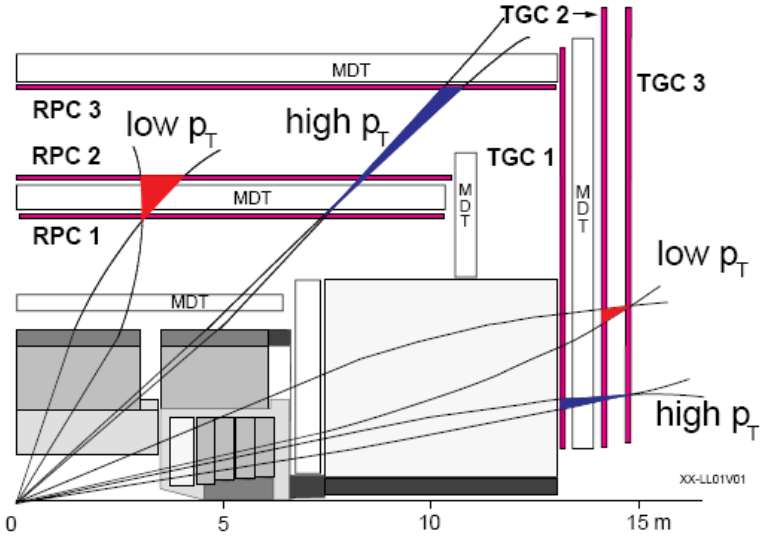

Figure 47: Muon LVL1 trigger – barrel and end-Cap

Appendix B Generated muon events

As explained in section 2.15, the simulated events are generated by DICE. The DICE output is a list of hits in the different TGC layers. Each event is of a muon with a different momentum, and contains a different number of hits.

An event structure is organized as follows:

- The tag <event> marks the start of an event structure
- The next entries represent the event serial number, and then Px, Py, Pz.

Each <hit> structure contains:

- The hit serial number
- X, Y and Z coordinates of the hit.

Examples:

```
<results>

<event> 1   13.667 11.999 63.964
<hit> 1 266.515 246.333 1306.811
<hit> 2 266.953 246.751 1309.191
<hit> 3 267.413 247.189 1311.682
<hit> 4 286.992 265.847 1417.812
<hit> 5 287.431 266.265 1420.191
<hit> 6 295.656 274.122 1464.812
<hit> 7 296.095 274.540 1467.191
</event>

<event> 3   0.810 1.016 3.336
</event>

<event> 5   35.891 20.099 67.426
<hit> 1 684.789 394.714 1306.811
<hit> 2 686.006 395.401 1309.191
<hit> 3 687.279 396.119 1311.682
<hit> 4 677.123 390.388 1291.811
<hit> 5 678.338 391.075 1294.191
<hit> 6 679.611 391.793 1296.682
<hit> 7 744.115 428.153 1422.812
<hit> 8 745.333 428.840 1425.191
<hit> 9 738.999 425.272 1412.812
<hit> 10 740.217 425.958 1415.191
<hit> 11 768.162 441.698 1469.812
<hit> 12 769.380 442.384 1472.191
<hit> 13 763.044 438.816 1459.812
<hit> 14 764.263 439.502 1462.191
</event>

</results>
```

Note the different number of hits in each event. Event #3 contains no hits: this is a soft muon with PT=1.3GeV, and was probably stopped by the calorimeter material.

The histogram of events used for training and testing the nets is shown in Figure 48. Larger number of events in the P_T range 4-6GeV were used, since this is the critical P_T region for a trigger tuned to 6GeV.

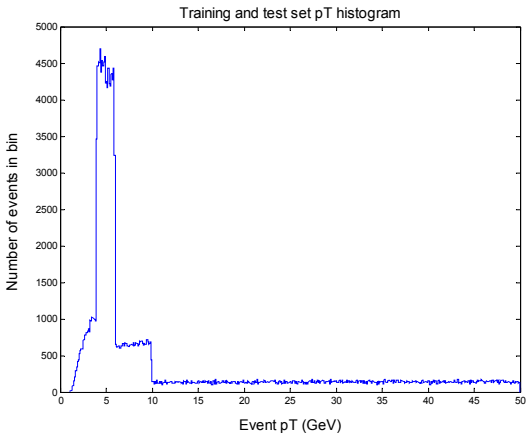

Figure 48: Event set P_T histogram

Appendix C δ*r*-δ*Φ* ANN

Another ANN input vector structure was studied. This network implements two input neurons. The corresponding input parameters are δr and δΦ – the deviations from the infinite momentum track ('residuals'), measured at M2. These parameters are identical to the parameters used by the dedicated hardware algorithm. An ANN based on these inputs does not require preprocessing, and is thus more adapted to implementation at Level 1.

Figure 49 plots the distribution of hits in the δr-δΦ coordinate system. Red hits mark events with PT>6GeV, and green hits mark events with PT>10GeV. It is evident that most of the bending occurs in the $\hat{r}$ direction. Coulomb scattering is more pronounced on soft muons than on hard ones.

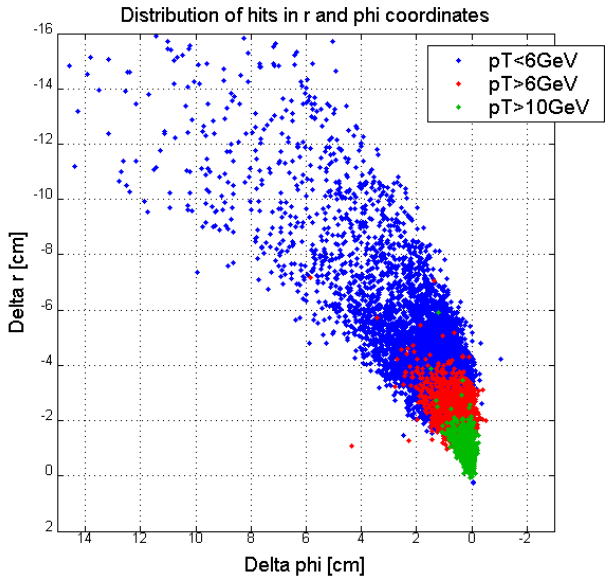

Figure 49: Distribution of hits in *r* and *Φ* coordinates

Fifty nets of the structure 2x7x7x1 were trained on these events, in the confined region of Figure 27. The ANN managed to learn the input structure, and classify the muons. The results did not show an improvement over the ATRIG algorithm. The results are summarized in the following table (relevant definitions in 4.6):

ATRIG	Slope = 2.75±0.15 GeV^{-1}
ANN	Slope = 2.55±0.10 GeV^{-1} Max = 2.74 GeV^{-1}

This result is in agreement with the conclusion of section 4.7 – that the superior performance achieved by the ANN is due to the improved resolution of the input parameters.

7. References

[1] ATLAS Collaboration, *The Large Hadron Collider Accelerator Project*: CERN/AC/93-03(LHC), 8 November 1993.

[2] LHC design report: <https://edms.cern.ch/file/445760/2/Vol_3_Chapter_1_v2.pdf>, and also <http://mcastro.home.cern.ch/mcastro/chapter1.pdf>, October 2006

[3] <http://livefromcern.web.cern.ch/livefromcern/antimatter/history/historypictures/LHC-drawing-half.jpg>, October 2006

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

[5] ATLAS Collaboration, *ATLAS Detector and Physics Performance Technical Design Report*, CERN/LHCC/99-14, May 1999, Volume I, Page 4.

[6] ATLAS Collaboration, *Inner Detector Technical Design Report*, Volume 1, CERN/LHCC/97-16, 30 April 1997.

[7] ATLAS Collaboration, *Pixel Detector Technical Design Report*, CERN/LHCC/98-13, 31 May 1998.

[8] ATLAS Collaboration, *Liquid Argon Calorimeter Technical Design Report*, CERN/LHCC/96-41, 15 December 1996.

[9] ATLAS Collaboration, *Tile Calorimeter Technical Design Report*, CERN/LHCC/96-42, 15 December 1996.

[10] ATLAS Collaboration, *Magnet System Technical Design Report*, CERN/LHCC/97-18, 1997

[11] ATLAS Collaboration, *ATLAS Barrel Toroid Magnet Technical Design Report*, CERN/LHCC/97–19, 1997.

[12] ATLAS Collaboration, *ATLAS End-cap Toroid Magnet Technical Design Report*, CERN/LHCC/97 – 20, 1997.

[13] ATLAS Collaboration, *Muon Spectrometer Technical Design Report*, CERN/LHCC/97- 22, 31 May 1997.

[14] S. Majawski et al., *A Thin Multiwire Chamber Operating in the High Multiplication Mode*, Nucl. Inst. Meth. 217, (1983) 265.

[15] ATLAS Collaboration, *Muon Spectrometer Technical Design Report*, CERN/LHCC/97- 22, 31 May 1997 – page 315.

[16] S. Dado et al., *A new high gain thin gap detector for the opal hadron calorimeter*, Nucl. Instrum. Methods A252 (1986) 511.

[17] G. Mikenberg, *Thin-gap gas chambers for hadronic calorimetry*, Nucl. Instrum. Methods A265 (1988) 223.

[18] C. Beard et al., *Thin, high gain wire chambers for electromagnetic presampling in OPAL*, Nucl. Instrum. Methods A286 (1990) 117.

[19] G. Bella et al., *Development of calorimeters using thin chambers operating in a high gain mode*, Nucl. Instrum. Methods A252 (1986) 503.

[20] ATLAS Collaboration, *ATLAS Detector and Physics Performance Technical Design Report*, CERN/LHCC/99-14, May 1999, Volume I, Page 23.

[21] ATLAS Collaboration, *ATLAS Detector and Physics Performance Technical Design Report*, CERN/LHCC/99-14, May 1999, Volume I, Page 349.

[22] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 297.

[23] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 373.

[24] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 293.

[25] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 312.

[26] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 300.

[27] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 367.

[28] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 368.

[29] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 378.

[30] DICE manual, ATLAS internal note SOFT-NO-10, 1994.

[31] ATRIG – ATLAS Trigger Simulation User Guide Revision 0.00, ATLAS note SOFT-94-017, 1994.

[32] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 380.

[33] L. Lonnblad, C. Peterson and T. Rognvaldsson, *Finding Gluon Jets with a Neural Trigger*, Phys. Rev. Lett. 65, 11 (1990).

[34] G. Marchesini, G. Nardulli and G. Pasquariello, *QCD coherence in tagging b jets by neural networks*, Nucl. Phys. B 394 (1993) 541.

[35] L. Lonnblad, C. Peterson and T. Rognvaldsson, *Mass reconstruction with a neural network*, Phys. Lett. B278 (1992) 181-186.

[36] J.H. Kohne, et al., *Realization of a second-level Neural Network Trigger for the H1 Experiment at HERA*, AIHENP'96 proceedings.

[37] T. Maggipinto, G. Nardulli, et al. *Role of Neural Networks in the Search of the Higgs Boson at LHC*, Phys. Lett. B409 (1997) 517-522 or [arXiv:hep-ex/9705020].

[38] E. Chorti, B. Granado, B. Denby and P. Garda, *An Electronic System for Simulation of Neural Network with a Micro-Second Real Time Constraint*, Batavia 2000 ,Advanced computing and analysis techniques in physics research, (2001) 76.

[39] R. Bruijn, *Muon Energy Reconstruction in ANTARES using Neural Networks*, Master's Thesis, August 2002.

[40] B.Muller, J. Reinhardt and M.T. Strickland, *Neural Networks, An Introduction*, Springer-Verlag Berlin Heidelberg 1990,1995.

[41] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press Inc.,1995.

[42] L. Lonnblad, C. Peterson and T. Rognvaldsson, *Mass reconstruction with a neural network*, Phys. Lett. B278 (1992) 181-186.

[43] Marquardt, D., *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, SIAM J. Appl. Math. 11, 431-441, 1963.

[44] Gill, P. R.; Murray, W.; and Wright, M. H. *The Levenberg-Marquardt Method* §4.7.3 in *Practical Optimization.* London: Academic Press, pp. 136-137, 1981.

[45] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 390.

[46] Gideon Dror et al., *Momentum reconstruction of particles in the forward muon trigger system of the ATLAS detector*, AIHENP99 proceedings, edited by G. Athanasiu, D. Perret-Gallix, Elsevier North-Holland Editors, (1999).

[47] Gideon Dror, Erez Etzion, *Momentum reconstruction and triggering in the ATLAS detector*, November 2000, (TAUP-2647 or [arXiv:hep-ex/0011061]).

[48] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 235.

[49] ATLAS Collaboration, *Level-1 Trigger Technical Design Report*, ATLAS TDR-12, June 1998, Page 231.

[50] ATLAS Collaboration, *ATLAS Detector and Physics Performance Technical Design Report*, CERN/LHCC/99-15, May 1999, Volume II, Page 481.