First Year Report

WW Scattering at the LHC

Adam Davison

Supervisor: Prof J. Butterworth

June 11, 2007

1 Introduction

Since starting my postgraduate studies in September 2006, I have had the opportunity to explore many interesting new concepts and ideas within the field of high energy particle physics. This report marks the end of the first year of my degree and summarises my activities and achievements to date.

My work so far within the UCL HEP group has been as part of the ATLAS experiment. This work has been broadly divided into two categories, one Physics-analysis focused, the other more technical in nature. My analysis work has been searching for new Physics in the WW scattering channel, while my technical work has been on the Atlantis event display package.

This document begins with a short section of background information on the ATLAS experiment required to support the main body of the report, which will describe my work in more detail.

2 ATLAS and the LHC

The Large Hadron Collider (LHC) at CERN is a particle accelerator experiment located on the French-Swiss border near Geneva. It aims to collide protons with a center of mass energy of 14TeV at several interaction points around a 27km ring. By performing these collisions at an energy higher than ever before achieved, it is hoped that new physics will be observed.

ATLAS is a "general purpose" detector designed to detect the particles produced in these high energy collisions. It will be used to deduce information about the underlying processes which took place during the collision.

The ATLAS detector is composed of a number of different detection systems, with different, more specific purposes. The inner detector operates inside a magnetic field and provides position information about charged particles produced in the collisions from which "tracks" can be reconstructed which give information about the particles' momentum. The electromagnetic and hadronic calorimeters sit outside the inner detector and measure the energy of particles. The outermost functional part of the detector is the muon system. As the name would suggest, these chambers provide detection of muons passing through the detector[1].

Often in collider experiments it is convenient to express quantities such as momentum and energy in the plane transverse to the beam direction, usually denoted by a subscript t.

One important derived variable measurable using the ATLAS detector is "Missing E_t". Since particles like neutrinos do not interact with the detector, it is impossible to detect them directly. However, since we know transverse

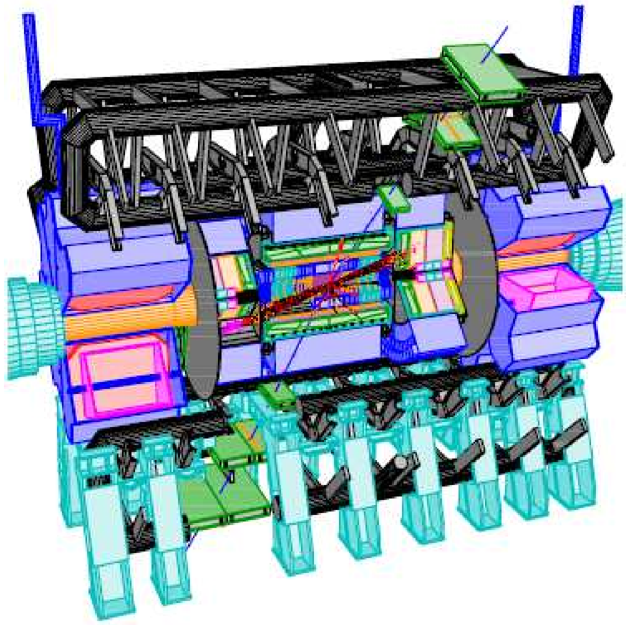

Figure 1: The ATLAS Detector[1]

momentum must be conserved in an event, we can sum over all the particles we can detect directly to form the quantity Missing E_t. Large values of Missing E_t imply that a non-interacting particle, such as a neutrino, was produced in the interaction.

3 The Physics of WW Scattering

Obviously there are a large number of potential processes which can take place when two protons interact with a center-of-mass energy of 14TeV. Generally in order to search for new physics, one chooses a specific process and final state (a channel) in order to make the analysis tractable. The focus of my work has been the WW scattering process, as shown in Figure 2.

In this process, two vector bosons are emitted from the interacting protons and then scatter off each other. Any potential new particle can be placed in the center of the scattering diagram. Assuming the cross-section for the resulting diagram is sufficiently high the new particle can be observed both in the cross-section and invariant mass spectrum of the process.

Although the process is named WW scattering, it is important to note that generally events which replace the hadronically decaying W with a Z boson are included in the signal for this channel. Although this document

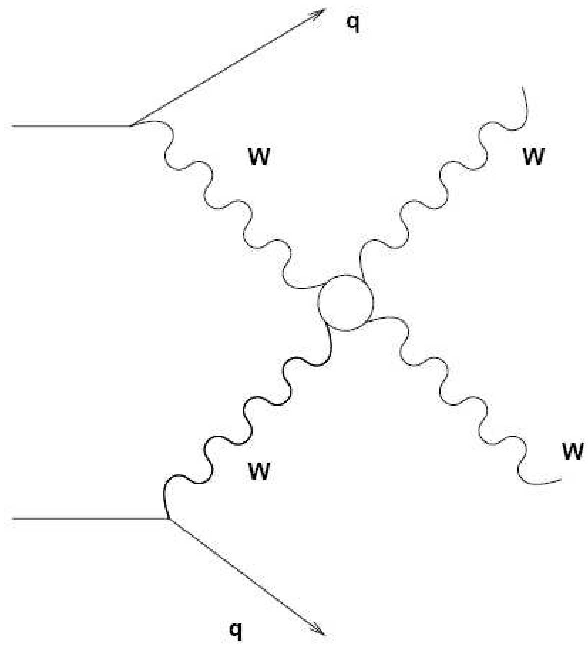

Figure 2: Feynman Diagram for Signal Process

will mainly discuss WW events, the results are generally unchanged for these WZ events.

In the absence of new particles beyond those already observed, standard model predictions of the cross-section for WW scattering become invalid at the TeV scale and violate unitarity. Generally theoretical models solve this by introducing new particles which create extra WW scattering diagrams which cancel out the divergence and restore unitarity. The Higgs is an example of such a particle but many other scenarios exist.

The theory which underpins the introduction of new resonances into the WW scattering invariant mass spectrum is the Electroweak Chiral Lagrangian (EWChL). The higher dimensional terms in this Lagrangian describe the form of any new Physics. In order to simplify the problem mathematically, only the first two unmeasured couplings, α_4 and α_5 are considered. This is equivalent to limiting the validity of the theory to being well below some energy scale Λ where new physics manifests itself. In order to extend the validity of this type of prediction up to at least $O(\Lambda)$, a "unitarisation" protocol is used. These protocols apply various conservation criteria in order to predict how the couplings might evolve with energy scale around $O(\Lambda)$.

As long as these results appear trustworthy, the values of α_4 and α_5 can then be varied to produce Lagrangians for various scenarios which introduce new scalar or vector resonances[2].

The experimental signature for a WW scattering event within ATLAS is jet(s) produced by the hadronic decay of one of the vector bosons, a high momentum lepton and Missing E_t. In order to reduce backgrounds to a manageable level the current WW analysis focuses on events where the hadronically decaying W boson has a very high (greater than 300GeV) P_t. In these events, the W is highly boosted and the decay products are usually resolved as a single, high p_t jet in the central region of the detector. Events where the leptonic W decays to a τ are also ignored currently.

The most basic cut applied as part of our analysis is to have at least one jet with a $p_t > 300GeV$. In this kinematic region, the main backgrounds for this process are $W + njets$ with the W decaying leptonically and $t\bar{t}$, with $W + njets$ being the main concern. Further cuts are then applied to this W candidate jet on the single jet mass and y scale, described in more detail in Section 4. The highest lepton and missing E_t must reconstruct to a W boson with a similar p_t cut of around 300GeV. There is an additional top-quark veto and cuts on the angular distribution of jets in the event designed to exploit the different signal and background topologies[2]

4 Jets

The most difficult part of the WW scattering analysis is constructing a set of cuts for the hadronically decaying W system which effectively reduce the background. Although applying a high p_t cut on the highest p_t jet in the event reduces background significantly, this is not sufficient. In the remaining background events, a jet from a QCD process cannot be easily distinguished from a jet produced from a highly boosted W decay using the jet variables alone. Further discriminating power can be obtained by looking inside the jet at the structure of the constituents. However, making use of this information requires a good understanding of the jet finding algorithm being used.

The jet finder is an algorithm for grouping 4-vectors according to some criteria. Usually the input 4-vectors are calorimeter deposits. The algorithm used in this report is $K_\perp$, which iterates over the input and makes binary decisions about combining pairs of inputs based on the "distance" between them in phase space as defined in Equation 1. The algorithm has a parameter R, which determines how near two 4-vectors have to be in order to be combined.

As long as the jet finder merges inputs using 4-vector addition, then the resulting 4-vectors have a mass, often referred to as the "single jet mass". For jets containing the decay products of a boosted vector boson, analogous to a dijet mass, the single jet mass should be the invariant mass of the parent vector boson, smeared by fragmentation, hadronization and detector resolution. For jets not produced from the decay of a heavy particle, the mass is generally lower, as the QCD splitting functions favour strongly ordered emissions which contribute very little to the jet mass. Therefore, the single jet mass is a variable with some power to discriminate between vector boson jets and light QCD jets.

Since the $K_{\perp}$ algorithm merges constituents into jets a pair at a time, it is also possible to explore the merging process in more detail. y values are the "distance" between two constituents which have been merged to make a jet, divided by the jet p_t to form a dimensionless quantity. Distance is defined as by the $K_{\perp}$ algorithm, the formula for which is given below:

$$d_{kl} \simeq min(E_k^2, E_l^2)\theta_{kl}^2 \tag{1}$$

The values are numbered such that y_2 represents the distance of the final merging while y_3 represents the penultimate etc... In this report only y_2 is discussed and all usage of y values implicitly refers to y_2.

y values can also be converted into a y energy scale by multiplying the jet p_t back in. Jets from the decay of a heavy particle X should have a $y(_2)$ scale of $O(M_X/2)$. As with the single jet mass, the QCD splitting functions which govern the structure of light QCD jets favour much softer splittings than seen in heavy particle decays. This tends to minimise one of E_k or E_l in equation 1 leading to relatively small y_2 and corresponding scale.

5 Software

ATLAS analysis is performed within the ATLAS offline software package. At the moment, as the LHC is not running, analysis is being performed on generated Monte Carlo events. As with many experiments, a typical scenario is to run a Monte Carlo event generator, followed by a detector simulation and finally some analysis code.

ATLAS has two main detector simulation packages. The first is a full detector simulation based on GEANT4[4]. This is understandably rather slow for a detector as complex as ATLAS. As a result, there is also a second, faster simulation called Atlfast. Atlfast makes a number of simplifications, the most important being that it uses vastly simplified geometry and parameterised smearing to approximate many of the effects observed in the full simulation.

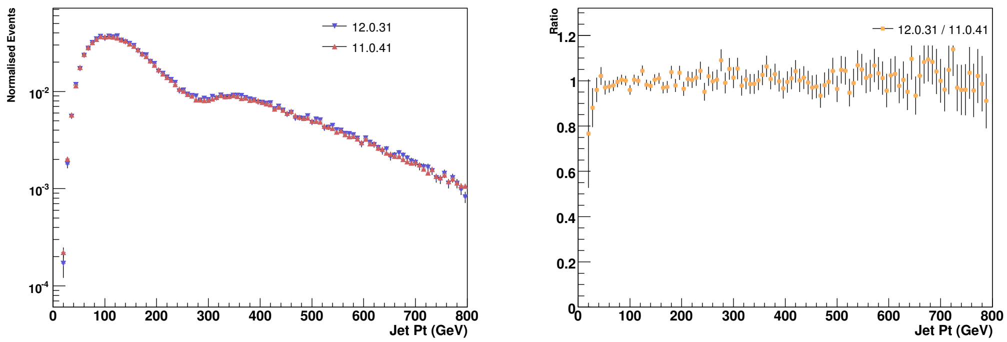

Figure 3: Leading jet p_t distributions comparison and ratio

6 Physics Contribution

Broadly speaking, my work follows on from the thesis of S. Stefanidis[5], which was submitted shortly after my joining the group. Stefanidis' work concentrated on the more theoretical aspects of a WW scattering analysis and explored the feasibility of such an analysis using the ATLAS experiment. Since we expect the first data from LHC experiments within the next year, the logical next step for the WW group is to take this understanding and use it to produce a robust and functional analysis.

6.1 Validation

The first task was to ensure that we were in a position to make use of our understanding of the signal and background. It was known that several important changes to both the Monte Carlo generators being used and also advances in the Atlfast detector simulation software could have had an impact on various variables used by the analysis.

In order to test both changes as independently as possible, the different versions of Atlfast were first compared using the same generated events. The versions of Atlfast compared were those included in version 11.0.41 and 12.0.31 of the ATLAS offline software package. The data sample used was a WW scattering signal sample in the case where no new resonances exist.

This comparison demonstrated that for all the key analysis variables the two versions of Atlfast agreed extremely well. Figure 3 shows the agreement for leading jet p_t. The good agreement was not surprising given that Atlfast is reasonably widely used throughout the ATLAS collaboration.

Once it has been shown that the different versions of Atlfast produced comparable output, it was possible to move on to using the different generator

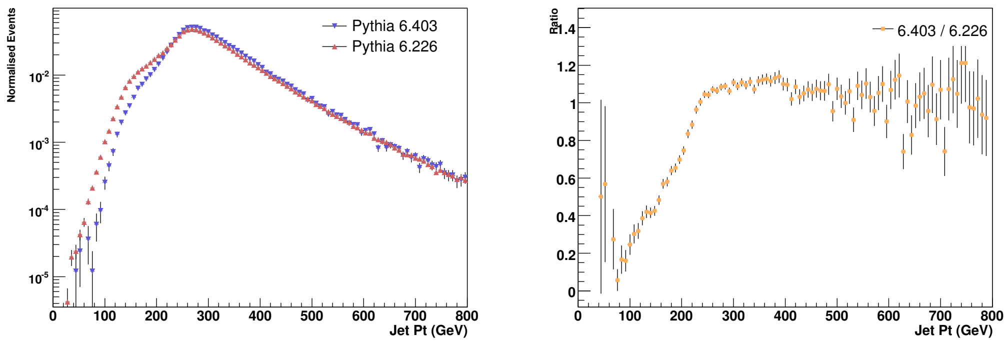

Figure 4: Leading jet p_t distributions comparison and ratio

versions as well as the different Atlfast versions. The generator versions being used were Pythia 6.403 and 6.226. The sample generated was a background, W+jets sample generated with a minimum hard scatter of 250GeV.

Agreement between most variables was again found to be very good. The only significant difference was in the leading jet p_t, shown in Figure 4. The main disagreement lies in the region below 250GeV, where the true leading jet has probably been missed by the detector. This implies that when the leading jet has a $p_t < 250GeV$ it has been produced from final state radiation. Therefore the shift in this distribution can be attributed to known changes in the parton shower algorithm between versions 6.403 and 6.226 of Pythia.

Since this shift is both understood and not significant to the WW scattering analysis, it is reasonable to assume that conclusions drawn using the older software version are still valid under more modern versions.

6.2 ATLAS Jet Reconstruction Software

As mentioned previously, the WW scattering analysis places greater demands on the ATLAS jet reconstruction (JetRec) software than most. It has therefore been a priority to be as involved as possible in the development of this software.

The single jet mass has been available by default as part of reconstructed jets for some time. However, the y values for $K_\perp$ jets have been less widely used and are currently not available. So far it has been reasonable (if technically complex) to generate them privately, however in the long-term it was obviously desirable to have them available as part of the default reconstruction.

One of my main contributions to the JetRec software has been to integrate the existing code for extracting y values from jets. This package, named

YSplitter (originally by P. Sherwood, J. Butterworth) was wrapped up as a tool of a suitable type to be included as part of the JetRec software. After consultation with the full-time JetRec developers, I was granted write access to this part of the ATLAS software and committed the code. It is anticipated that as part of an upcoming version of the JetRec software this tool will then be turned on by default, allowing for much wider (and simpler) usage of y values in analyses.

6.3 Jet Structure

As previously mentioned, the WW scattering analysis depends heavily on being able to make cuts on the y scale and mass of a jet. The paper WW Scattering at the LHC[2] evaluates the discriminating power of these cuts at the truth level. The work of S. Stefanidis explored the effects of detector resolution on these cuts using the Atlfast fast simulation package. My contribution has been to continue this work, exploring in more detail the detector resolution for these variables under both the fast and full detector simulations.

Two types of analysis were performed for the y scale and mass, the first being purely statistical comparisons between the truth, fast and full simulations and the second being event by event comparisons between reconstructed values and true values. The statistical comparisons demonstrate overall trends, while the event by event comparisons provide more direct access to the simulated detector resolutions. The statistical comparisons are normalised to unit area.

Figure 5 shows a statistical comparison between the truth and the two different detector simulations for single jet mass. The jets shown in this image are from a WW signal sample. This particular sample deals with a scenario where a new scalar particle exists with a mass of 1.15TeV. The jets shown here are those that pass a minimum p_t cut of 300GeV. The two graphs represent those jets that are or are not within dR ($\sqrt{d\eta^2 + d\phi^2}$) of 0.2 to a true hadronically decaying W boson. The agreement between the two detector simulations is very good except in the region below 10GeV, which spoils the normalisation somewhat for the anti-matched jets. This effect is discussed in more detail below.

Figure 6 shows an event by event comparison between the two detector simulations. Here, the leading jet from each event has been matched, by event number to the leading jet in the corresponding truth event. The value plotted is the reconstructed mass minus the true mass, divided by the true mass which gives a measure of the fractional change from the true value. Plotting this value makes it possible to explore the effects of the detector

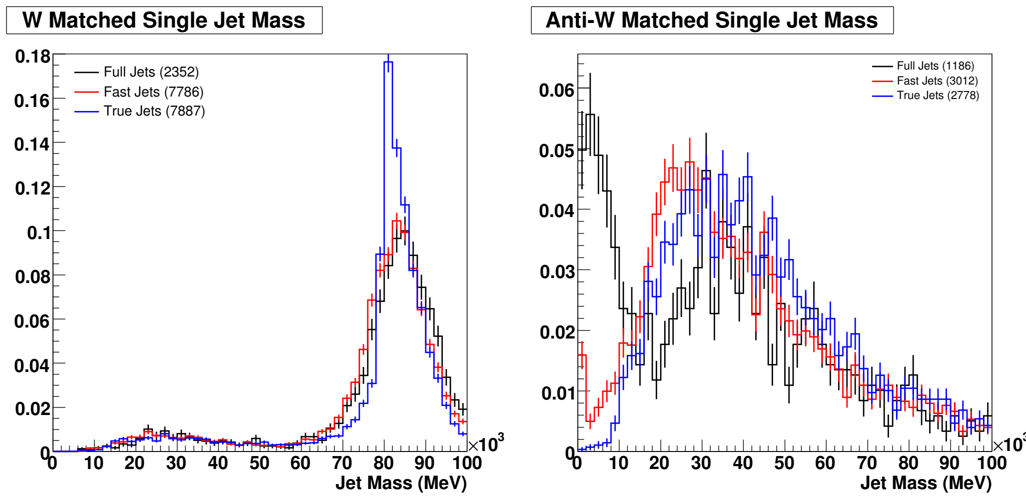

Figure 5: Single jet masses from different detector simulations

	σ
Full Simulation	$9.2\% \pm 0.2\%$
Atlfast	$9.1\% \pm 0.2\%$

Table 1: Gaussian fits for single jet mass resolution

on the measured value. The plot shows good agreement between the two simulations. Both appear to have similar resolutions for single jet mass measurements.

To try and quantify the resolutions somewhat, a Gaussian fit was performed to the central peaks for each simulation. The results are shown in Table 1. The two detector simulations have very similar resolutions, implying that the single jet mass is not as sensitive to changes in the detector geometry as might be expected.

An equivalent study was performed on the y scale variable, using the same criteria as for the single jet mass plots. The statistical comparison is shown in Figure 7 and is equivalent to Figure 5 for the single jet mass. This shows agreement to within statistical errors over almost the entire range. As seen for the single jet mass, at very low scales, the two simulations begin to diverge. This is an effect which needs to be further understood although it is worth noting that this is well separated from the region which this analysis considers to be signal ($>\sim 40GeV$). Presumably in this region the coarser calorimeter granularity employed by Atlfast affects ability to resolve y_2.

An equivalent event by event study was also performed, the results are given in Figure 8. From this plot it appears that the full simulation has a

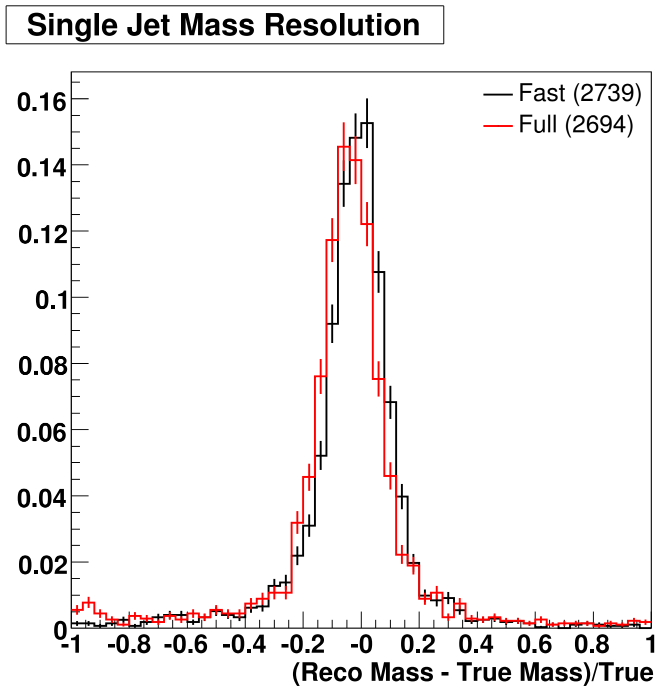

Figure 6: Single jet mass resolution from different simulations

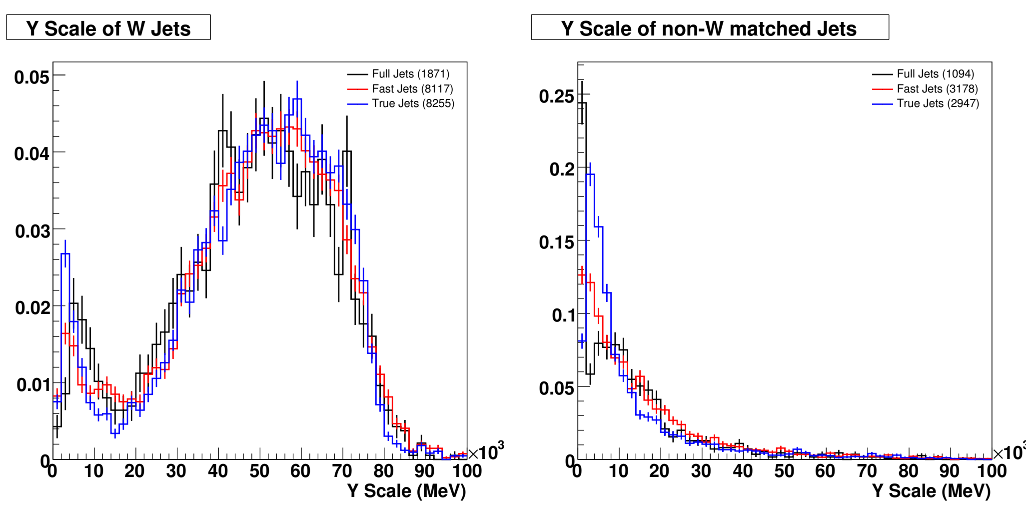

Figure 7: Y scales from different detector simulations

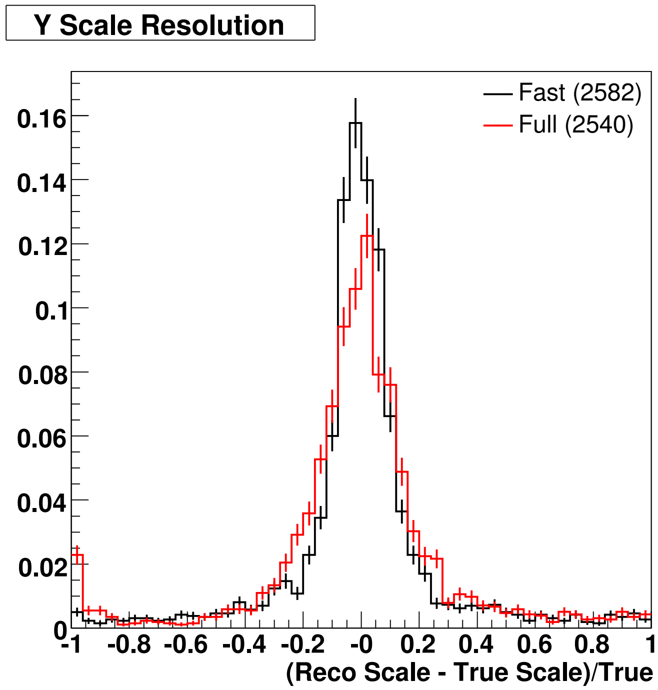

Figure 8: Y scale resolutions from different detector simulations

	σ
Full Simulation	$12.3\% \pm 0.3\%$
Atlfast	$8.8\% \pm 0.2\%$

Table 2: Gaussian fits for y scale resolution

slightly worse resolution while the approximations made by Atlfast underestimate the smearing somewhat.

As with the single jet mass above, Gaussian fits to the peaks in Figure 8 were performed, the results of which are shown in Table 2. As was expected based on the plot, the full simulation suggests worse resolution, which presumably more accurately models the final performance of the ATLAS detector.

The similarity between the two detector simulations for these two variables is an indication that they are reasonably well modeled and understood by both simulations.

Some efficiencies were also evaluated for signal and background samples, using Atlfast only for technical reasons. A set of sample cuts were chosen, $p_t > 300GeV$, jet mass $> 60GeV$ and y scale $> 40GeV$. These were then applied to the signal sample used above, a $W+3jets$ background sample and

All units fb	WW scalar 1.15TeV	W+4jets	W+3jets
Input	19.6	1530.0	164761.0
After p_t and jet mass cut	7.0	124.5	5004.4
After y scale cut	5.5	42.0	1855.8

Table 3: Cross-sections for signal and W+jets samples

	WW scalar 1.15TeV	W+4jets	W+3jets
Input	100.0%	100.0%	100.0%
After p_t and jet mass cut	36.0%	8.1%	3.0%
After y scale cut	78.7%	33.7%	37.1%

Table 4: Efficiencies per stage given as percentages

a $W + 4jets$ background sample. The calculated rates were then multiplied by the relevant cross-section and branching ratio to fill Table 3. These values can be used to calculate the efficiency of each individual cut when applied in this order as shown in Table 4. These values demonstrate that even after applying the single jet mass cut, the y scale cut still improves the signal to background ratio by a factor of 2-3 for these samples.

6.4 Collaboration and ExoticPhysView

The UCL WW scattering group are unsurprisingly not the only people working on a vector boson scattering analysis within the ATLAS collaboration. The UCL group co-operates with other groups working on similar channels such as $WZ \rightarrow qqll$ as part of the "CSC DiBosons" working group. The aim of this group is to produce an ATLAS note describing how a vector boson scattering analysis will be performed and the feasibility of such an analysis. The group is currently in the process of producing a common analysis framework, ExoticPhysView, allowing as much sharing of code as possible and providing a common n-tuple structure.

My contribution in this area has been to participate in the development process, especially providing testing and constructive criticism which has uncovered several important issues with the code.

6.5 Distributed Analysis

One of the big technical challenges faced by LHC experiments is making use of distributed analysis ("Grid Computing") to handle the large volumes of data which will be produced. In order to prepare for this, all the analysis groups are encouraged to start doing as much distributed analysis as possible, as early as possible. There are several packages which are designed to assist the Physicist in this regard, GANGA[3] being one of the most widely used. Theoretically, GANGA reduces the task of running a distributed analysis on the Grid to that of producing a configuration file describing the task. In practice, neither the tool-chain nor the underlying mechanisms are very mature. My contribution has been to be the first person from the DiBosons group to successfully run the ExoticPhysView software on the Grid. It is considered important for each analysis group to attempt this kind of task. Sharing my work and experiences has significantly helped others in this area. The best concrete example of this is the UCL wiki page, AtlasGanga[6] written by C. Bernius and myself which is a good resource for anyone starting to use GANGA.

7 Technical Contribution

My technical contribution to date has been related to the Atlantis event display package. Atlantis is written in the Java programming language and is capable of displaying a huge range of truth and reconstructed information about an event within the context of the ATLAS detector. It reads data in an XML (eXtensible Markup Language) based format produced by the JiveXML package. In this way it is capable of displaying data from a number of different sources within the ATLAS software including Monte Carlo generator output, low level detector output and reconstructed data. This wide range of features makes it extremely useful for both detector commissioning and analysis tasks.

My contributions to the Atlantis codebase so far have been mainly small introductory tasks, such as fixing bugs reported by users. Since I have prior experience profiling and optimising Java code, I have also been assigned several reports of slow behaviour to examine and eradicate where possible. For example, there were several reports that drawing of tracks was slow when highly zoomed into the center of the detector. Using profiling tools I was able to ascertain where the problem lay within the helix rendering algorithm and then rework that portion of the code to improve performance. I have also been an active contributor to the regular development meetings associated with the project.

My main contribution to the JiveXML package has been to explore the available options for using distributed analysis techniques to run JiveXML on the grid. Here, I was able to apply the knowledge gained as part of the equivalent Physics task. This will allow easier access to many of the currently available simulated data samples and will be an important ability when real data becomes available.

Finally, I also took part in an Atlantis tutorial session at CERN (3rd May 2007). I gave a 15 minute presentation on the basic features of Atlantis, followed by a 45 minute hands-on session which guided people through Atlantis in an interactive manner.

8 Personal

In December 2006, my work of Summer 2004 with Prof. M. Slater (Department of Computer Science, University College London) was published in the open access journal PLoS ONE[7]. The work was also featured on nature.com[8], the website associated with the Nature journal.

I have had the opportunity to give a number of talks on my work, both in phone meetings and in person. The highlight of these has been presenting my work on single jet mass and y scale resolutions at CERN to a large audience during the jets session of the latest ATLAS Trigger and Physics week.

9 Future Plans

By the time real data becomes available from ATLAS, the WW scattering group aims to be in a position to make full use of it. Therefore a significant part of my future Physics contribution, at least in the short to mid term, will be to perform whatever tasks are necessary to improve the analysis tools to the required level. I will also need to further understand the nature of our signal and background events, for example further studies of the y scales and related quantities will have to be performed in order to maximise the usefulness of these cuts.

In the longer term, the nature of my work becomes somewhat dependent on both the LHC startup schedule and what new Physics exists at the $> 100GeV$ scale. Presumably though it will involve contributing to the publication either of a WW scattering cross-section measurement or a related study, of the $W + njets$ background for example.

As mentioned previously, so far my technical work on Atlantis has been somewhat limited for logistical reasons. As things progress, I plan to expand my work in this area and where possible apply my technical knowledge to my Physics analysis.

10 Conclusions

My first year working on the ATLAS experiment has been an opportunity to work with many interesting people on one of the most interesting experiments in modern high-energy Physics. I have learnt many important concepts in relation to the WW scattering analysis and applied them to an ongoing study of the effectiveness of two important discriminating variables, the single jet mass and y scales. So far the outcome of this has been that these variables are reasonably well understood and simulated by the ATLAS software. My work so far has made a relevant contribution to the ATLAS collaboration and as my studies progress, I hope to expand this further.

11 Thanks

I would like to thank J. Butterworth for excellent supervision and guidance, E. Ozcan and S. Dean for all their help, Z. Maxa and N. Konstantinidis for help with Atlantis and everyone else who has welcomed me into the group.

References

[1] ATLAS Technical Design Report, Volume 1 (1999) - http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/TDR.html

[2] J. M. Butterworth, B. E. Cox and J. R. Forshaw, Phys. Rev. D **65** (2002) 096014 [arXiv:hep-ph/0201098].

[3] K. Harrison *et al.*, *In the Proceedings of 2003 Conference for Computing in High-Energy and Nuclear Physics (CHEP 03), La Jolla, California, 24-28 Mar 2003, pp TUCT002* [arXiv:cs.se/0306085].

[4] GEANT - Detector Description and Simulation Tool - http://wwwasd.web.cern.ch/wwwasd/geant/

[5] Stefanidis, E. (2007) *Study of WW scattering in the absence of a light Higgs boson using the ATLAS detector at the LHC* PhD Thesis, University College London

[6] AtlasGanga - Running a Grid Job using GANGA - https://www.hep.ucl.ac.uk/twiki/bin/view/Main/AtlasGanga

[7] Slater M, Antley A, Davison A, Swapp D, Guger C, et al. PLoS ONE 1(1): e39. (2006) doi:10.1371/journal.pone.0000039

[8] news @ nature.com, Virtual reality shocker, 22 Dec 2006 - http://www.nature.com/news/2006/061218/full/061218-17.html