This is to certify that the
thesis entitled

JET RECONSTRUCTION IN W + JETS EVENTS AT THE
LHC

presented by

ULRIKE SCHNOOR

has been accepted towards fulfillment
of the requirements for the

M. S. degree in Physics and Astronomy

Joey W Huston
Major Professor's Signature

8/27/2010
Date

JET RECONSTRUCTION IN W + JETS EVENTS AT THE LHC

By

Ulrike Schnoor

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Physics and Astronomy

2010

ABSTRACT

JET RECONSTRUCTION IN W + JETS EVENTS AT THE LHC

By

Ulrike Schnoor

Events with W bosons and jets in the final states play an important role in particle physics, both as a common background to interesting processes and as as an interesting process in its own right. They can be used in detector performance studies and to test perturbative QCD.

Studying multiple jet final states requires profound knowledge of the jet finding algorithms that are used to reconstruct the jets in an event. This thesis uses the software framework SpartyJet to conduct a comparison study of different jet reconstruction algorithms. A set of Monte Carlo simulations generated with ALPGEN at a center-of-mass energy of 7 TeV for the process of $W \rightarrow e\nu$ + jets has been investigated. For different jet multiplicities in the final state, the jet algorithms anti-k_T, k_T, and SISCone are tested. Finally, some early data pp collisions from ATLAS are studied as well.

ACKNOWLEDGMENT

It is a pleasure to thank the people who made this thesis possible. First of all, I would like to express my gratitude to my advisor Professor Joey Huston for letting me work on this project and for his guidance and support. Special thanks also goes to Brian Martin and Jessie Muir for the great cooperation and helpful comments and discussions. This thesis would not have been possible without the productive and friendly working environment in our office.

In addition, I am grateful to everyone who made my stay at MSU a successful and memorable year. This goes to people on both sides of the Atlantic Ocean: my friends in East Lansing as well as my family and friends in Germany for their moral support. I also want to gratefully acknowledge the Studienstiftung des deutschen Volkes whose cooperation with MSU and financial support made this stay possible.

TABLE OF CONTENTS

LIST OF FIGURES

Images in this thesis are presented in color

1. Introduction

At unprecedented high energy and luminosity, the Large Hadron Collider at the European Laboratory for Particle Physics (CERN) is colliding proton beams to investigate Standard Model and Beyond Standard Model processes. It is designed to make discoveries at the TeV scale such as finding the Higgs boson and exploring Beyond Standard Model signals like supersymmetry and extra dimensions.

However, in the early data taking at the LHC, new physics searches will not play a big role. At this stage, fundamental studies for the understanding of the detector's performance need to be carried out, such as calibration studies and underlying event production measurements as well as measurements of Standard Model processes [4]. To be able to make new discoveries with ATLAS, the detector first has to be well understood, and Standard Model processes have to be rediscovered and accurately investigated.

Also, all reconstruction methods have to be tested. Kinematic measurements and identification of electrons, muons, and missing E_T can be defined fairly precisely, whereas the measurement and reconstruction of jets is a more demanding task for which different jet algorithms for different final state topologies have been developed. SpartyJet, the software tool used in this study is a convenient way to simultaneously run different jet finding algorithms with different parameters.

A particularly important process for detector performance tests, as well as for new physics searches, is the production of W bosons in connection with jets. Due to their large cross section, W bosons are copiously produced at the LHC. This makes these processes well suitable for various studies.

In this thesis, different jet reconstruction algorithms are applied to a set of Monte Carlo simulations of W + n jets production to compare the transverse momentum distributions of the results of the algorithms. The Monte Carlo samples have been generated with ALPGEN [1] at a center-of-mass energy of 7 TeV. The data is from the L1Calo stream from 7 TeV proton-proton collisions in ATLAS.

The first chapter gives an introduction to the basic particle physics concepts that are useful for this study. A review of the ATLAS detector's structure and functionality is given in the next chapter. It is followed by a detailed description of the jet algorithms that are used in this thesis. The last two chapters contain transverse momentum distributions of their results on Monte Carlo samples and ATLAS data.

2. $W \rightarrow e\nu$ + Jets processes

2.1 Role of W production at LHC

Processes whose final state contains a W boson decaying to an electron or muon and a neutrino plus n jets ($n \geq 0$) play a special role at ATLAS. First of all, they are background to many interesting Standard Model and Beyond Standard Model processes, including the production of top quarks and Higgs bosons, as well as supersymmetry and processes involving extra dimensions [25].

Therefore, knowing the properties of $W \rightarrow e\nu + n$ jets production is necessary to be able to deal with the background in these physics searches and in order to make new discoveries. Deviations of the measured cross sections of high-p_T jets from QCD predictions could point to new physics [5].

In addition, $W \rightarrow e\nu$ + jets processes are particularly suitable to investigate Standard Model processes at ATLAS, as they allow us to "rediscover" known Standard Model properties and to test perturbative QCD (pQCD) predictions: At the interaction scale corresponding to the W boson's mass, perturbation theory can be applied. Also, abundant statistics are guaranteed by the high production rate of W bosons at the LHC due to their high cross sections. Thus, with precision measurements of relevant parameters, such as the mass of the W boson m_W, perturbative QCD can be tested.

Last but not least, the study of $W \rightarrow e\nu + n$ jets production is suitable to understand the detector in performance studies and to make precision tests of jet, lepton and missing E_T reconstruction. The latter is particularly important in the early data measurement. $W \rightarrow e\nu$ + jets processes will also be used to test and tune the Monte Carlo generators.

2.2 Standard Model and QED

The Standard Model is the current view of the interactions of elementary particles through the three fundamental forces: the electromagnetic, the weak, and the strong force. Gravitation is not included. The Standard Model proclaims a set of elementary particles containing the leptons, quarks and gauge bosons. It also describes the interactions of the elementary particles that are carried by the gauge bosons.

There are three generations of leptons: electrons, muons, and taus with their respective neutrinos. Quarks are grouped into three generations as well: up and down quark belong to the first generation, charm and strange to the second, and top and bottom to the third generation of quarks. There are four types of gauge bosons that mediate the forces: the photon belongs to the electromagnetic interaction, the W and Z bosons carry the weak force, and the gluons are the gauge bosons of the strong interaction.

Since electrons, muons, taus, W bosons and all quarks have an electric charge, they take part in the electromagnetic interaction. Furthermore, left-handed leptons and quarks of each generation interact through the weak force and are organized in weak isospin doublets.

Quantum Electrodynamics (QED), the relativistic quantum field theory of the electromagnetic interaction is unified with the theory of the weak interaction in the electro-weak theory. The gauge bosons W^+, W^-, Z^0, and the photon are the mediators of this force. In order for the fermions and gauge bosons to be massive, the $U(1) \times SU(2)$ symmetry of the electro-weak interaction has to be broken. This happens through the **Higgs mechanism**.

It introduces an additional field, called the Higgs field, that interacts with all other fields and itself. The Higgs field is a complex SU(2) doublet with four degrees of freedom. The Higgs potential in the Lagrangian goes to the fourth power of the field and has a non-zero vacuum expectation value. This spontaneously breaks the local symmetry and leads to Goldstone bosons that become the massive longitudinal modes of the gauge fields. W and Z bosons obtain massive longitudinal degrees of freedom, whereas the photon stays massless as it has only transverse components. This leaves the forth degree of freedom of the complex doublet to be a new particle, the Higgs boson, which has yet to be discovered. This is one of the goals of the LHC physics search.

2.3 Quantum Chromodynamics

In addition to the above-named leptons and gauge bosons, the Standard Model contains the color charged quarks and gluons. Quarks take part in all interactions: those of the electromagnetic, the weak, and the strong forces. Gluons are the mediators of the strong force which is described by Quantum Chromodynamics (QCD). This is the key interaction for the production of jets in hadron collisions.

Color is a charge similar to the electric charge. It lets quarks interact by exchanging gluons, which have color as well. Gluons can also interact with each other because the strong gauge group SU(3) is non-abelian, which leads to self-interaction terms in the QCD Lagrangian. There are three colors: red, blue, and green. Each quark has one of these colors, whereas gluons have one color and one anticolor.

The coupling constant of Quantum Electrodynamics, α_{QED}, is rather small, with a value of $\alpha_{QED} \approx 1/137$ at low energies. The coupling increases at higher energies.

In contrast to this behavior, the strong coupling constant α_s is small at high energies and increases when the energy scale decreases, reaching a value close to unity at low energies. This running of the coupling leads to two properties that are characteristic for QCD: **confinement**, which occurs at low energy scales, and **asymptotic freedom** at high energies.

Asymptotic freedom means that at high energy scales, according to the Heisenberg uncertainty principle corresponding to small distances, the strength of the strong interaction decreases significantly causing quarks and gluons to behave like free particles. Asymptotic freedom can be described perturbatively.

Quark or gluon confinement occurs at low energies and is the reason why colored particles are never found individually: If a quark-antiquark or gluon pair is pulled apart, at some point it becomes more efficient to use the energy spent on the separation for the production of new quark-antiquark or gluon pairs. As the particles are moved further apart, more and more of these pairs of colored objects are produced. Finally, these colored particles hadronize, i. e. they combine to form colorless particles. These can be mesons (containing a quark and an antiquark whose colors cancel) and baryons (containing three quarks with a different color each so the net color charge vanishes). In general, colorless states made of multiple quarks are called **hadrons**.

Besides these two properties, the running of the strong coupling constant also has an impact on the use of perturbation theory for QCD calculations. Usually, in a quantum field theory, a perturbative expansion can be made using the coupling constant as expansion parameter. As $\alpha_{QED} \ll 1$, perturbatition theory can be used to calculate QED processes at low energy scales, throughout the energy range reached at current colliders. However, at low energies, the strong interaction cannot be treated perturbatively: α_s becomes so large that the expansion does not converge.

Therefore, the factorization theorem has to be used in order to apply perturbation theory to QCD. It factorizes a QCD cross section into two parts: one long-distance piece that is not calculable, but universal, and one short-distance piece that is process dependent and calculable with perturbative QCD. The universal long-distance functions, including parton distribution functions, fragmentation functions, and form factors, are determined by globally fitting to experiments.

2.4 Jets in hadron collisions

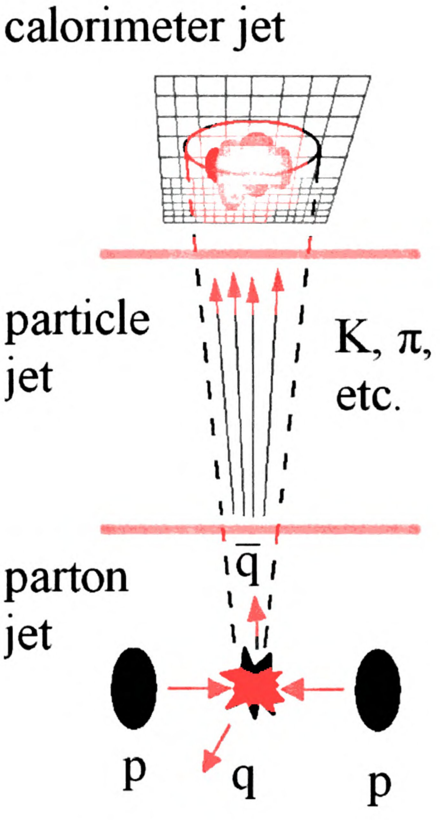


Figure 2.1: Jet production in pp collisions

Jets are sprays of elementary particles that are created in the hard scattering events of hadronic collisions. When two protons collide in the LHC, two of their partons take part in the hard scattering, producing quarks and gluons with high transverse momentum. They travel away from the interaction point into the detector.

On their way, they emit showers of gluons and quarks due to confinement. Then, these particles hadronize and the hadrons deposit their energy in the calorimeters of the detector. This energy is what is measured and used for jet reconstruction (see chapter 4).

The perturbative picture of jet production consists of a hard scattering process between two partons with high momentum transfer that can be calculated to fixed order perturbative QCD. It is the short-distance part of the interaction. Hadronization is the long-distance piece that can not be calculated with perturbation theory.

The three levels of jets visualized in figure (2.1) are used in Monte Carlo simulations of jet production. The parton level contains the original partons after the collision. With additional perturbative parton showering and non-perturbative hadronization, the particle or hadron level is generated. It also includes the non-perturbative underlying event.

To get to the calorimeter level (also called detector or cluster level), the particles are passed through a full ATLAS detector simulation, taking into account detector effects such as the cell resolution. It can also include the effects of pile-up: additional collision events occuring in the same bunch crossing as the event of interest. Pile-up adds soft radiation energy to the event. Real data is only available on detector level and can be corrected to particle level.

Jets have to be reconstructed from the clusters of hadronic particles that are measured in an event. In order to do this, different jet finding algorithms have been developed and are in use at the ATLAS experiment. Chapter 4 describes the jet reconstruction process and the most important algorithms in detail.

2.5 Other LHC physics goals

In addition to the search for a Higgs boson (see chapter 2.2), the LHC is looking to probe many other potential beyond standard model theories. Two of them will be briefly touched on below: **supersymmetry** and **extra dimensions**.

Supersymmetry is a theory that introduces new particles as supersymmetric partners to the Standard model particles: For each fermion, there is a SUSY boson, and for each boson, a SUSY fermion. Supersymmetry is a highly motivated theory. Besides providing an excellent dark matter candidate, it could solve the hierarchy problem and explain several other issues of beyond Standard Model searches.

Especially for the efforts of a grand unification of all forces, including gravitation, attempts are made to find extra spatial dimensions at the LHC. Dimensions additional to the three that we know of could be "curled-up", thus being invisible for us. However, extra dimensions would have an effect on gravitation since it occupies the entire universe. The effects would be small and difficult to measure, but could be seen on the TeV scale, leading to special signals like one high p_T jet and a lot of missing E_T when a virtual graviton disappears into the extra dimensions.

3. The ATLAS experiment at the Large Hadron Collider

3.1 The Large Hadron Collider at CERN

On March 30, 2010, the Large Hadron Collider (LHC) at the European Laboratory for Particle Physics (CERN) has started producing proton-proton collision events at a center-of-mass energy of 7 TeV. Colliding protons requires a complex design with two distinct rings for the proton beams. It is not possible to use one beam pipe for both beams like in a proton-antiproton collider such as the Tevatron. However, a proton-proton collider type is preferable because an antiproton beam cannot have sufficient intensity to reach the LHC's design luminosity of 10^{34} cm^{-2} s^{-1} [3].

The two collider rings that contain the counter-rotating beams are situated in 27 km long tunnel and are connected to the CERN accelerator complex via two transfer tunnels. From the accelerator complex, proton beams are injected into the LHC, where they are further accelerated by electric fields and guided by superconducting magnets.

There are four intersection points where the two proton beams collide. Six particle detector experiments, including the ATLAS detector, are performing a variety of studies, including precision tests of the interaction forces, especially QCD, top quark measurements and searches for a Standard Model or beyond Standard Model Higgs boson.

3.2 The ATLAS detector

The ATLAS experiment (A Toroidal LHC ApparatuS) is one of the general-purpose detectors at the LHC, constructed for measurements over a wide kinematic range and for a broad spectrum of physics searches. The ATLAS detector is designed to meet ambitious physics goals for the search of new physics signatures in a variety of processes, one of the benchmark processes being the Higgs boson search [5].

The layout of the detector is forward-backward symmetric and eight-fold rotational symmetric around the beam axis. The core part is the Inner Detector which covers the pseudorapidity region $|\eta| < 2.5$ and is responsible for charged particle tracking. It is surrounded by a thin 2 T superconducting solenoid [6]. With a combination of three sub-detectors, the Pixel Detector, SemiConductor Tracker (SCT), and the Transition Radiation Tracker (TRT), the Inner Detector performs pattern recognition, momentum and vertex measurements, and electron identification [5].

Beyond the solenoid are the electromagnetic and hadronic calorimeters. The electromagnetic calorimeter is a liquid-argon sampling calorimeter with high granulatity. It covers the pseudorapidity range $|\eta| < 3.2$ with an accordion-type calorimeter, and the range up to $|\eta| < 4.9$ with a compact Forward Calorimeter. It provides precision measurements of electrons and photons. Hadronic calorimetry is provided by a scintillating tile calorimeter (TileCal) in the rapidity region $|\eta| < 1.7$, extended up to $|\eta| < 4.9$ by the endcap and forward hadronic calorimeters that use liquid argon technology [7]. Hadronic calorimetry is responsible for jet reconstruction and E_T^{miss} measurements.

The muon spectrometer is situated beyond the calorimeters. A toroidal magnetic field bends the muons, with muon chambers to measure their tracks. There are three large superconducting toroid magnets, one beyond the barrel calorimeters and two in the end-cap regions of the detector. The muon chambers contain three layers each and are divided into two different types: cylindrically shaped Monitored Drift Tubes (MDT) in the barrel area, and Cathode Strip Chambers (CSC), arranged in disks, in the forward direction.

Finally, there are three smaller detector systems covering the large rapidity regions of the detector. Their tasks are to determine the luminosity of the beam and the centrality of heavy-ion collisions [5].

3.3 ATLAS triggering and data acquisition

Coherent data-taking at ATLAS is provided by the combination of the Trigger and Data Acquisition (TADQ) system, and the Detector Control System (DCS). Starting at an event rate of about 10^9 events per second, the trigger system's task is to gradually reduce this rate to about 200 Hz of signals.

The trigger system has three levels: level 1 (L1), level 2 (L2), and the event filter. The Data Acquisition System is responsible for data movement, but also controls hardware and software detector components involved in the data-taking process. The Detector Control System (DSC) coordinates the detector hardware operation and serves as an interface for the detector operator. Both systems are divided into sub-units corresponding to the sub-detectors of ATLAS [5].

At the planned luminosity, 10^9 interactions per second will occur in ATLAS, as proton bunches cross inside the detector at a rate of 40 MHz with an average of 23 events per bunch crossing [8]. Each signal is first processed by the sub-detectors' front-end electronics and then passed into the L1 trigger buffer, where it is stored for about 2.5 μs, the length of time the level 1 trigger takes for its decision. The L1 trigger is an online, hardware-based system that searches for high-p_T leptons, photons, and jets as well as large missing E_T. Processing reduced granularity information from the relevant sub-detectors for each selection, the L1 trigger decision reduces the event rate to $\sim$ 75 kHz. In addition, the first trigger level defines Regions of Interest (ROI's) according to interesting features found in a certain area of the event. To avoid large deadtimes in the case of temporally close L1 triggers, the accepted events are first buffered in the derandomising buffer and then transmitted into the Readout Drivers (ROD's).

The L2 trigger retrieves these events and refines the selection of the first level, biased by the first level's choice of ROI's. It uses software selection algorithms run on a farm of 500 processors to select the events, based on information from the entire detector at full granularity. That way, sharper thresholds can be applied on this level [9]. After L2 triggering, the rate goes down to $\sim$ 3.5 kHz.

The accepted data is collected and assembled by the event-builder system and then transfered to the event filter that uses the full event information to reduce the rate further to approximately 200 Hz. Finally, the CERN computer center stores all events that have passed the event filter [5].

3.4 ATLAS coordinate system [10]

The coordinate system used in the ATLAS detector is a right-handed system. Its z-axis follows the direction of the counter-clockwise rotating beam and the x-axis points to the center of the LHC ring. The y-axis points upwards but is slightly different to the vertical axis, as the entire collider ring is tilted about 1.23 % with respect to the horizontal plane. The transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ is the momentum perpendicular to the beam axis.

ϕ is the azimuthal angle circling around the beam, with $\phi = 0$ at the positive x-axis. The polar angle θ is the angle with respect to the positive z-axis. It is usually replaced by another coordinate, the pseudo-rapidity η which is a good approximation to the rapidity y. Using the rapidity is preferable to using the polar angle θ because the differential cross section $\frac{d\sigma}{dy}$ is invariant under Lorentz boosts. It depends on θ according to equation (3.1).

$$\eta = -\log\left(\tan\frac{\theta}{2}\right) \tag{3.1}$$

4. Jet reconstruction

4.1 ATLAS jet measurements

For many important ATLAS physics searches, jet reconstruction and jet energy measurements have to meet high efficiency and resolution requirements. An example is the measurement of the top quark mass in $t\bar{t}$ events with a semileptonic final state, for which the jet energy scale uncertainty should not exceed 1% [11].

First of all, a more extensive input to jet finding has to be built from the individual signals of the calorimeter cells. There are two approaches to do this: One way is to build **towers** by summing up the contents of cell bins of the dimensions $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$. This is an in-discriminant approach using all cells in the bin. Negative cell signals, which can occur as electronic noise after a signal has been registered, are recombined with positive ones until the net signal is positive. However, it does not provide actual noise suppression. The second way is the formation of three-dimensional **topological clusters** (short form: topo-clusters). Here, seed cells with an energy greater than a certain threshold are clustered together with their nearest neighbors. If these have energies above a certain smaller threshold, they are secondary seeds and get clustered with their own nearest neighbors. If no secondary seeds are found in the vicinity of a seed, all nearest neighbors are included into the cluster without regards to their energy deposit. This way provides noise suppression and a smaller number of clusters [5]. In the analysis in this thesis, the topo-cluster approach has been used.

In the next step of jet reconstruction, jet finding algorithms are run on the results of either one of the signal clustering methods. The following sections describe the various jet reconstruction algorithms that are in use at ATLAS. The calorimeter jets found by the respective algorithm undergo further calibration and corrections in order to correct to the particle level for the jets. These include hadronic jet calibration based on cell signal weighting and algorithm effects and finally, an in-situ calibration taking into account underlying event and other corrections. See figure (4.1) for a plan of the jet reconstruction process.

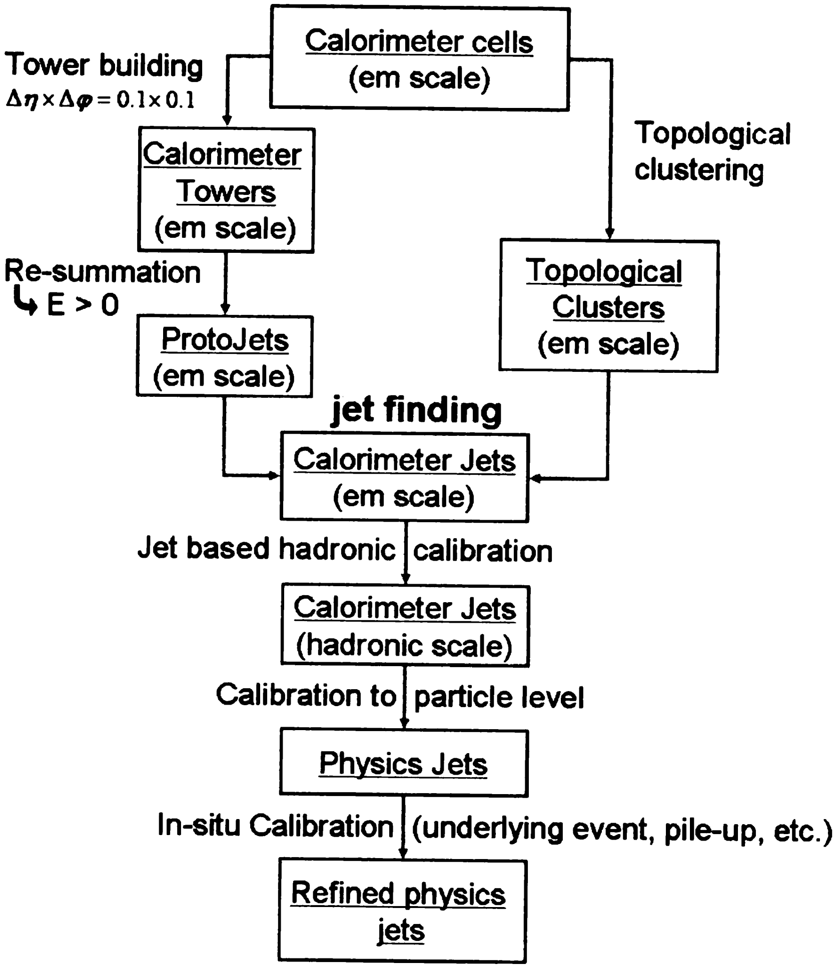


Figure 4.1: Calorimeter jet reconstruction, from [5]

4.2 Jet clustering algorithms

Jet reconstruction is the process of combining topological cluster particles or calorimeter towers into jets and assigning a four-momentum to these jets. An algorithm defines how the calorimeter signals are grouped into jets. Together with a distinct set of parameters and a certain recombination scheme that determines each new object's four-momentum, this is called a **jet definition**.

Jet algorithms can be run on different levels of input: parton level, hadron or truth level, and topo-cluster or detector level. Parton level and hadron level only exist in theory calculations or Monte Carlo simulations, whereas in experiments, the jet clustering will be carried out based on the detector's topo-clusters or towers.

Jet reconstruction requires the algorithm to be as similar as possible at all levels, to be detector independent, fast, and easy to calibrate.

Two major groups of jet finding algorithms are in use at ATLAS: algorithms that cluster particles to jets according to proximity in space are called **cone algorithms**, whereas **sequential recombination algorithms** cluster particles according to proximity of their momenta.

The next sections explain both types of jet reconstruction algorithms and their respective properties.

4.3 Cone algorithms

4.3.1 Iterative cone algorithms

There are different cone algorithm approaches, with most of them being iterative cone algorithms. In this case, an initial seed particle i is selected. Its momentum is added to the sum of the momenta of all particles j within a cone of radius R around the seed. The particles used are all those particles j for which the following relation is valid:

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 < R^2 \tag{4.1}$$

The resulting direction of the momentum sum is used as a seed particle for the next iteration. This iteration is repeated until a stable direction of each cone is determined. The different types of iterative cone algorithms can be distinguished by the way they deal with the following issues: first, how to find an appropriate initial seed particle, and second, how to handle the situation of overlapping stable cones in one event (i. e. when particles can be assigned to multiple cones).

The first important class of iterative cone algorithms are those that use the progressive removal method (IC-PR algorithms). In this approach, the particle with the highest transverse momentum is used as initial seed. After iteratively finding the stable cone position, all particles within this cone are removed from the event, and the iteration starts over with the highest p_T particle among the remaining clusters as the new seed. This is repeated until no particles are left in the event. Alternatively to this iterative approach, the same removal method can be used with fixed cones, i. e. a fixed cone is set up around the respective seed particle and all particles within the cone are removed. This is again repeated until no particles are left in the event (FC-PR algorithms, the nomenclature is adopted from [13]).

The second kind of iterative cone algorithms are the split-merge algorithms (IC-SM), using splitting and merging of cones to deal with the issue of overlapping cones. All particles, or optionally all particles above a certain p_T threshold, are used as seeds for the iteration. Once all stable cones are found, the splitting and merging procedure is performed. Two jet cones are merged if the particles they share contain at least a fraction f (typically $f = 0.75$ or 0.5) of the softer cone's transverse momentum. If this is not the case, the two cones are split by assigning the common particles to only one of the cones, usually the one whose axis is closer in the rapidity-azimuth-plane.

4.3.2 Infrared and collinear safety (IRC)

Infrared and collinear safety issues play an important role in the performance of cone algorithms in regard to comparisons to theoretical prediction. When applying an infrared and collinear safe jet finding algorithm to an event, adding a soft parton or collinearly splitting a particle/tower does not change the resulting jets reconstructed in the event. However, for the cone algorithms discussed above, collinear splitting of a particle's energy or a soft emission added to an event can lead to a different jet configuration. Collinear splitting of particles as well as the emission of soft partons occur randomly and with unpredictable properties in each event, so it is desirable that they do not affect the result of the jet algorithm. Also, in fixed-order QCD, the singularities from soft emissions and collinear splittings of partons usually cancel with divergent contributions from loops. With an IRC unsafe jet algorithm, both could lead to a different set of jets and thus to a possibly infinite cross section, because they might not cancel anymore. The two methods of iterative cone algorithms, IC-SM and IC-PR, have different issues with IRC safety.

IC-PR algorithms tend to be collinear unsafe, because they use the hardest particle of the event as starting seed. The collinear splitting of the hardest particle can result in a different initial seed, as another particle could become the hardest instead. This can lead to a different final jet configuration after the reconstruction, and therefore it is possible that singularities in the cross sections do not cancel with the loop corrections in the usual way, yielding infinite cross sections.

The issue for IC-SM algorithms is infrared unsafety. The emission of a soft particle can provide an additional seed. In cases where the corresponding cone overlaps with two neighboring harder cones, the split-merge procedure could merge these cones instead of having two separate cones which would be the case without the soft emission.

So again, the infinite cross sections from loop matrix elements and from the infrared singularity do not cancel, leading to an infinite jet cross section.

Extending cone algorithms to midpoint cone algorithms solves infrared unsafety partially, but still not completely. In a second run after applying the conventional IC-SM procedure, midpoint cone algorithms also iterate from seeds put in the middle of a pair of stable cones. That way, the final stable cones do not depend on the presence of seed cells between the jets. This is only a solution for simpler jet configurations, as it can still lead to infrared unsafety in other cases [13].

A complete solution of the cone-type algorithms IRC unsafety issues is provided by seedless cone algorithms.

4.3.3 Seedless cone algorithms - the SISCone algorithm

The primary idea for a seedless cone algorithm (SC algorithm) is to find all stable cones in an event with an exact procedure. The algorithms starts with finding all possible subsets of particles in the event. It then calculates the resulting momentum of each subset. A stable cone is found in those cases where the entire initial subset is included in the cone centered around the resulting axis. However, as all 2^n possible subsets of the n particles have to be processed and only very few will be stable cones at the end, this approach is very time consuming.

A more efficient seedless cone algorithm is the **SISCone** algorithm (Seedless infrared safe cone algorithm) [14]. It avoids long running times by using a computational geometry approach: among the 2^n possible subsets of particles only those that fit into a circle of radius R in the y-ϕ-plane are used, because all other subsets will never form a stable cone anyway. To find all relevant subsets of particles, i. e. the ones lying in a circle of radius R, all pairs of points within a distance of $2R$ have to be found and all possible circles through these pairs have to be drawn. Then, the resulting momentum is calculated and it is checked if the resulting cone is stable, i. e. the set of particles enclosed in the resulting cone corresponds to the initial subset. Finally, a split-merge procedure is run on the resulting stable cones.

By using the computational geometry approach, this algorithm is faster than the original seedless cone algorithm starting with all possible subsets of particles. It has shown to be infrared safe, in contrast to seeded cone algorithms. SISCone is one of the jet finding algorithms used in this thesis.

A last feature of cone algorithms that has to be mentioned are dark towers, particles that are not clustered into any jet. They occur for all cone-type algorithms using a split-merge procedure: Sometimes, clusters with lower p_T that are close to a high p_T jet, are not included in any stable cone. This can happen as a seed cone in the low p_T area is always drawn into the high p_T jet. In a split-merge procedure, only clusters inside a stable cone are considered. Therefore, this area becomes a dark tower.

In a progressive removal iteration, the area would be included in a stable cone once the high p_T jet is removed. SISCone does not produce dark towers as it runs the cone finding again on the energy remaining after the first run, and thereafter until no unclustered energy is left [13].

4.4 Sequential recombination algorithms

In general, sequential recombination algorithms calculate the distances between the initial particles with a certain algorithm specific measure of distance. Then they sequentially recombine the particles with the smallest distance. Contrary to cone algorithms, jets do not have to be split or merged, and no dark towers will appear, because every particle belongs to one and only one jet. Also, sequential recombination algorithms are infrared and collinear safe.

One has to distinguish between algorithms used at e^+e^- colliders and at hadronic colliders. At e^+e^- colliders, the total energy of the event is well-known and can be used in the distance measurement. This has been done for sequential recombination algorithms at lepton colliders. In a hadron collider event, the total energy of a hard scattering process is not known because each parton taking part in the hard scattering carries only a certain fraction of the proton's energy described by the parton distribution functions. Therefore, a different way to describe the distance has to be found.

The most commonly used representations of sequential recombination algorithms are the anti-k_T and the k_T algorithms that are described in the following sections. In the case of two incoming proton beams, both algorithms have in common the following distance measures with different values of the parameter p:

The distance between two particles i and j,

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p})\frac{\Delta R_{ij}^2}{R^2}, \tag{4.2}$$

and the distance between a particle and the beam

$$d_{iB} = p_{T,i}^{2p}, \tag{4.3}$$

with

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2. \tag{4.4}$$

Here, $p_{T,i}$ is the transverse momentum of particle i, while y_i and ϕ_i are its rapidity and angular coordinate. Different values can be chosen for the jet size R, usually between 0.4 and 1.2. As it is standard in proton colliders, d_{ij} and d_{iB} are invariant under longitudinal boosts. The different kinds of sequential recombination algorithms possess different values of p:

- $p = 1$: k_T algorithm

- $p = 0$: Cambridge-Aachen algorithm

- $p = -1$: anti-k_T algorithm

The algorithms k_T and anti-k_T algorithms will be used in this thesis and are described below. The Cambridge/Aachen algorithm is very similar but shows energy independent clustering.

4.4.1 The k_T algorithm

The (*inclusive*) k_T algorithm is used as proposed in [15]. For this algorithm, $p = 1$ in equations (4.2) and (4.3). After a preclustering procedure that reduces the number of initial particles [11], the algorithm proceeds as follows:

1. Determine d_{ij} and d_{iB} for each topo-cluster or tower.
2. Find d_{min}, the minimum of all d_{ij} and d_{iB}.
3. If $d_{min} = d_{ij}$, combine i and j to one new particle l and determine the four-momentum of l according to the recombination scheme that is used. Usually, this is just the four-momentum sum of i and j: $p_l^\mu = p_i^\mu + p_k^\mu$.
4. If $d_{min} = d_{iB}$, declare i to be a jet and remove it from the list of particles.
5. Stop iteration when no particle is left.

Through this procedure, a jet i is declared when the distance of i to each one of the other particles j, weighted by the ratio of their transverse momenta, is greater than the parameter R. Therefore, R is the crucial parameter in the k_T algorithm. It clusters soft particles with small relative momentum first. By favoring the clustering of soft particles, a jet can have arbitrarily small momentum. This is avoided by applying a transverse momentum cut to the final jets.

4.4.2 The anti-k_T algorithm

The anti-k_T algorithm uses equations (4.2) and (4.3) with $p = -1$ as distance measures as proposed in [16]. The clustering procedure is the same as the five steps of the k_T algorithm described above. In contrast to the k_T algorithm, anti-k_T starts clustering hard particles with particles that have small relative momenta to the hard particle. Therefore, anti-k_T's jets grow around a hard particle, leading to a circular shape of the resulting jets. Thus anti-k_T can be used as a substitute for cone-type algorithms, with similar properties but the advantage of being infrared and collinear safe like all sequential recombination algorithms.

4.5 SpartyJet

Kinematic reconstruction of jets at the LHC is more difficult than at previous colliders because the LHC covers a broader range of jet energies. Also, more pile-up events are taking place, adding energy from soft radiation to the event, which reduces the energy resolution significantly.

Therefore, at ATLAS, not only one, but several different algorithms and jet size parameters are used to analyze data sets. SpartyJet [18] is an analysis framework that serves as a jet finding tool providing all relevant jet algorithms for ATLAS simultaneously. This facilitates comparative jet finding studies like the one in chapter 5 of this thesis.

Structured in a modular way, SpartyJet is able to use any input in the form of four-momenta to perform any operation on the retrieved four-momentum sets. Most importantly, it carries out jet finding with the implemented algorithms: all ATLAS, CMS, CDF, and D0 algorithms, and PYTHIA's Celljet. Also, SpartyJet provides an interface to the FastJet library that includes k_T, Cambridge/Aachen, and anti-k_T algorithms, and a plugin for SISCone. In addition to jet reconstruction, various jet tools are implemented in SpartyJet as well, including jet area tools, input and output kinematic cut tools, a PDG ID[1] selection tool, geometric moment tools, and a p_T density tool.

First, the input four-momentum sets are converted into an initial jet list. The jet tools and the reconstruction algorithms are successively applied to this list, each modifying it further. Finally, the resulting jet collection is saved in ROOT ntuple format [19].

Another feature provided by SpartyJet is the addition of pile-up events. This is done by using a minimum bias event file and adding a certain number of these pile-up events to the original signal event. The number of pile-up interactions added can either be a fixed value, or can be drawn from a Poisson distribution with a fixed mean value. This feature allows for the direct comparison of signal events with signal plus pile-up events.

---

[1] Monte Carlo numbering scheme of the Particle Data Group, see http://pdg.lbl.gov/2002/montecarlorpp.pdf

5. Monte Carlo study

5.1 $W \rightarrow e\nu + \mathrm{n\,jets}$ ALPGEN samples

This study investigates $W \rightarrow e\nu + \mathrm{n\,jets}$ events and compares the analysis results of different jet reconstruction algorithms at the different jet levels (parton, hadron, and cluster levels). It is based on ALPGEN [1] samples for the process of W boson production associated with n jets at a center of mass energy of 7 TeV. There are several exclusive samples with the number of jets $n = 0, 1, 2, 3, 4$, and 5. Both W^+ and W^- production are considered. The W^- boson decays to an electron and an electron antineutrino, the W^+ to a positron and an electron neutrino. Parton showering and hadronization have been added to the tree-level ALPGEN Monte Carlo with HERWIG [20]. Underlying event effects have been added using the Jimmy generator [21] with double parton scattering.

The original samples are exclusive n jet samples. Exclusive distributions only include events with n and only n jets reconstructed with the respective jet definition on the respective level. For inclusive distributions, the resulting exclusive histograms have been merged by adding the weighted distributions. The distributions are weighted according to their respective ALPGEN cross sections that are taken from the ATLAS Database [22].

5.2 Cross section measurement

The relation between the number of events and the cross section of a certain process can be calculated according to

$$N = \mathcal{L} \cdot \sigma \cdot A \cdot \epsilon + B, \tag{5.1}$$

where $\mathcal{L}$ is the luminosity, N is the number of events passing all event selection cuts, σ is the cross section, A is the acceptance of the signal due to kinematic and angular cuts, ϵ is the reconstruction efficiency for this signal, and B the number of background events observed.

To measure the cross section of a certain process, use eq. 5.1 in the form

$$\sigma = \frac{N - B}{\mathcal{L} \cdot A \cdot \epsilon}. \tag{5.2}$$

As we are only making qualitative comparisons, the plots only show the number of events N divided by the respective luminosity as their cross section. We are assuming that the Monte Carlo simulation estimates the efficiency correctly. As the amount of data present is limited, background is not included.

5.3 Selection criteria and cuts

Event selection of the Monte Carlo sample is based on the Strawman A selection that the W/Z observation group at ATLAS has agreed on [25], with a few small deviations. All jets in the distributions have to pass the following kinematic cuts:

- $|\eta^{jet}| < 2.8$
- $p_T^{jet} > 30$ GeV

At parton level, a 30 GeV cut is applied to the parton transverse momentum. At cluster level, this cut is applied to the locally calibrated jets, which still have to be corrected for the jet energy scale. As the jet energy scale calibration has been performed incompletely in this study, cluster level jets are not calibrated to the full scale, so their p_T does not correspond to the parton p_T. Thus, the cut of 30 GeV is cutting more jets at cluster level than it would if the correct jet energy scale was used.

Kinematic cuts for leptons and W bosons:

- $E_T^{miss} > 25\,\mathrm{GeV}$
- electron $p_T > 20\,\mathrm{GeV}$
- electron $|\eta| < 2.47$, also excluding the crack region $1.37 < |\eta| < 1.52$
- electron IsEM cut is "robusttight"
- W transverse mass $m_T > 40\,\mathrm{GeV}$

The crack region at $1.37 < |\eta| < 1.52$ is excluded from photon and electron measurements because of the bad energy resolution in the transition area between barrel and endcap detectors. The trigger cut is EF_e20_loose, which triggers on a single isolated electron.

5.4 Transverse momentum distributions

5.4.1 Levels of reconstruction

The jets have been reconstructed at parton and cluster level. It was not possible to include hadron level results because with the truth particle information provided in the Monte Carlo D3PDs, an accurate truth level reconstruction is not available. The parton level corresponds to the Monte Carlo parton simulation. To get the hadron level, parton showering and hadronization are added to the parton level Monte Carlo events. Also, the non-perturbative contributions of the underlying event are added in. In addition to this, the cluster level takes into account detector effects via a full simulation of the ATLAS detector.

The cluster level reconstruction can also be corrected for underlying event: The p_T density ρ_{p_T} is calculated by running the k_T algorithm with R = 0.5 on all jets without applying a transverse momentum cut. The median transverse momentum density in bins of η is determined. We have used five bins with the divisions η = 0, 1.8, 2.4, 4.3, and 5.0.

In addition to the median p_T density, the jet area a is calculated. For each algorithm, a different jet area calculation method has been used: for k_T, this is the Voronoi area; for anti-k_T the active area; and for SISCone the optimized-passive area. They are determined by clustering so-called ghost particles with vanishing energy into the jets. For more details on jet area calculations see [24]. Then, the product of $\rho_{p_T} \cdot a$ is subtracted from the cluster level to correct for underlying event effects.

In the histograms, this corrected cluster level is labeled as "Corr". The jets on this level do not necessarily pass the jet p_T cut: Their uncorrected p_T does pass, but after subtracting the correction, the transverse momentum might be smaller than the original jet p_T cut. Still, a p_T cut of 30 GeV is applied on the corrected jets, so some jets are cut from the distributions (see chapter 5.5 for an investigation of the cluster correction).

5.4.2 Comparisons of all algorithms for exclusive samples

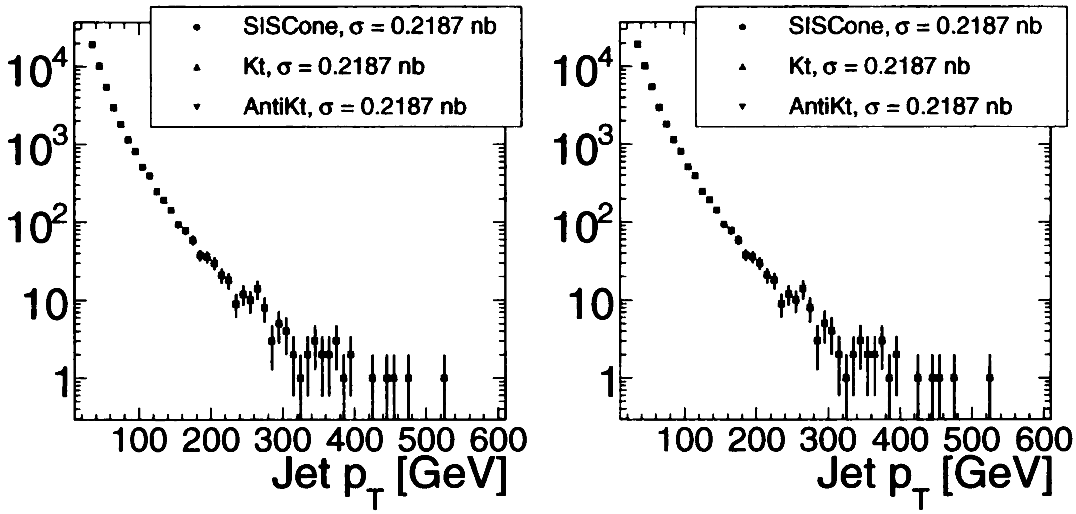
A jet definition contains the chosen jet algorithm and the parameter R in equations (4.1) and (4.2) to (4.4). In addition to this, the SISCone jet definition depends on the split-merge parameter, for which we have used the value f = 0.75. The results of the reconstruction also depend on the event level that has been used for the reconstruction. The following plots compare the three jet algorithms that have been used in this study. Each histogram contains the three algorithms' distributions with one given value of the jet size R. They also specify the total cross sections of each distribution σ. Figure 5.1 contains the parton level distributions and figure 5.2 the distributions at cluster level. All of these histograms are based on the W + 1 jet sample.

Anti-k_T and k_T are quite similar on one given level and with one given jet size, whereas SISCone shows bigger differences to both of them. This can be seen in the histograms in figure 5.1 and 5.2. They show the distributions of jet transverse momentum in events with a single jet found with the respective algorithm.

At parton level, since an exclusive W + 1 jet sample has been used for these plots, the p_T distributions of all three algorithms are exactly identical. At cluster level, anti-k_T and k_T give very similar shapes of the jet p_T distributions. Also, their cross sections only differ by less than 3 %, being smallest for $R = 0.7$. The deviation from anti-k_T to SISCone increases for larger jet sizes R and is of the order of 4 to 7 %.

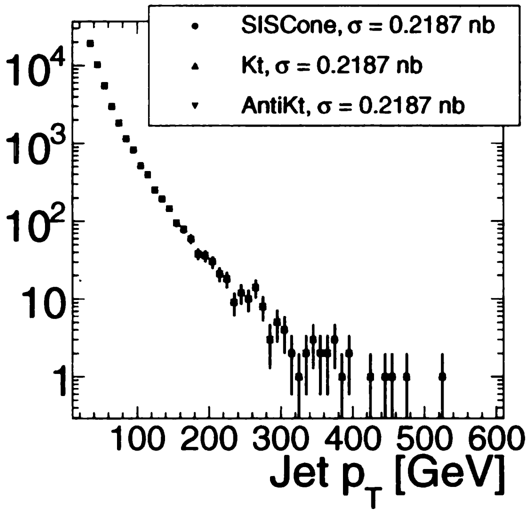
The reason for this behavior is that k_T and anti-k_T are both sequential recombination algorithms. Their main difference is that k_T starts the clustering with the particles that have smallest relative transverse momentum, sequentially matching these particles. Anti-k_T starts with the hardest particle in the event and clusters the particles with small relative p_T to it. SISCone is a cone algorithm that iterates over the particles in order to find a stable cone direction. The particles in this cone are included in the jet. This different approach leads to different results.

As the differences between the two sequential recombination algorithms are marginal, the following study only considers anti-k_T and SISCone algorithms. In the following section, the differences between SISCone and anti-k_T will be further investigated.



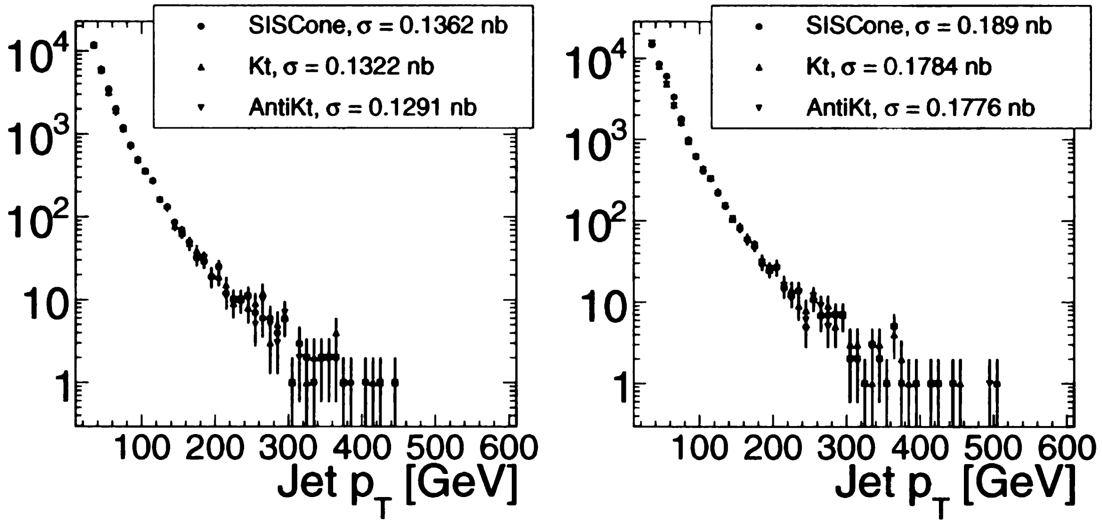
(a) Parton level, R = 0.4

(b) Parton level, R = 0.7



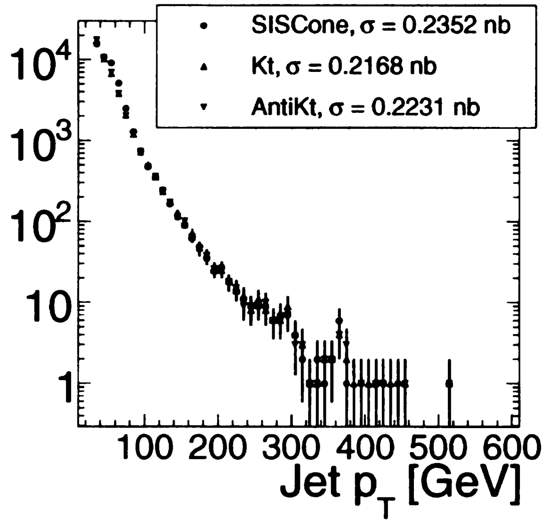
(c) Parton level, R = 1.0

Figure 5.1: Transverse momentum distribution of the jet in an exclusive W + 1 jet sample: Comparison of the cross sections of SISCone, k_T, and anti-k_T at parton level.



(a) Cluster level, R = 0.4

(b) Cluster level, R = 0.7



(c) Cluster level, R = 1.0

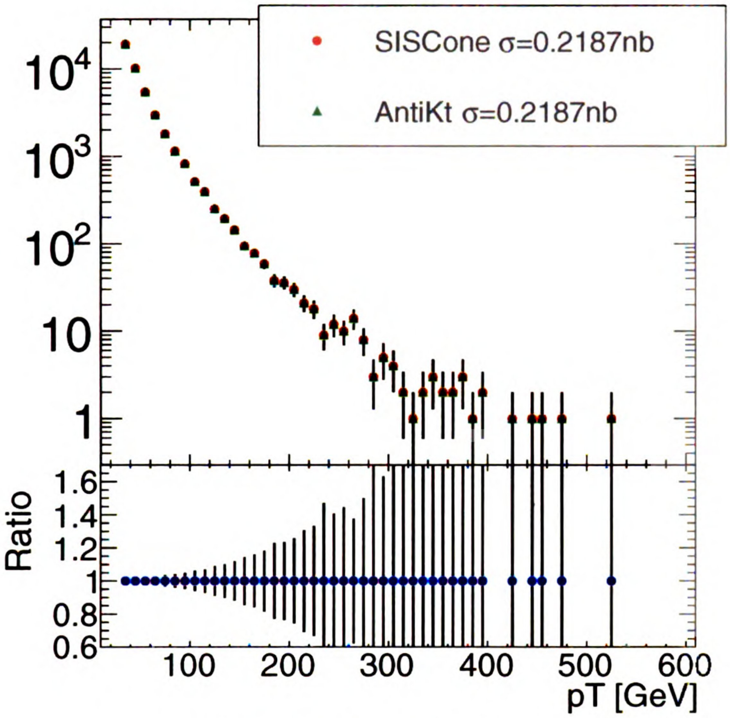
Figure 5.2: Transverse momentum distribution of the jet in an exclusive W + 1 jet sample: Comparison of the cross sections of SISCone, k_T, and anti-k_T at cluster level.

5.4.3 Comparison of SISCone and anti-k_T algorithms

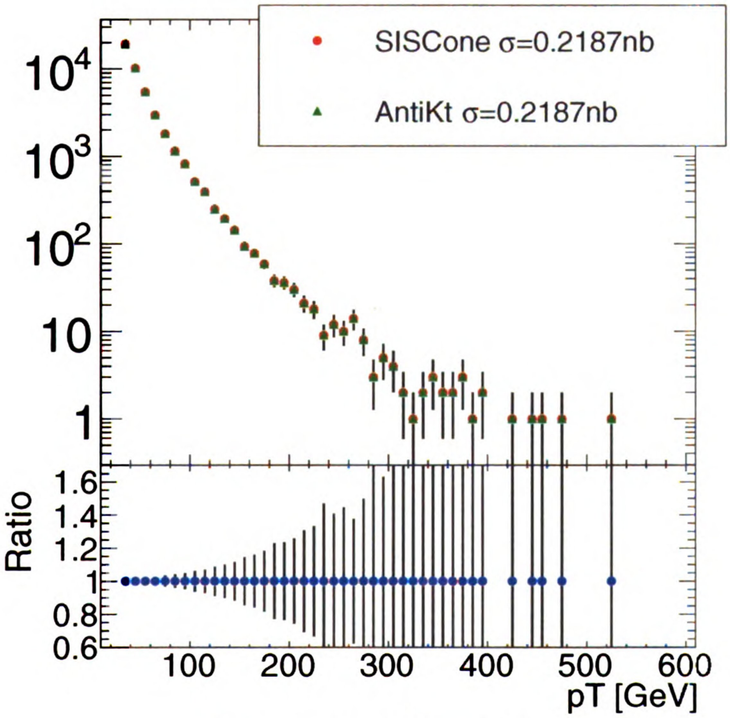
The following plots show the transverse momentum distributions of the W + 1, 2, and 3 jet samples. They are arranged according to the jet's order: The leading jet is the one with highest p_T in the event, the second leading jet has second highest p_T, etc.

Figures 5.3 and 5.4 show the W + 1 leading jet results. They are followed by the W + 2 results, with figures 5.5 and 5.6 for the leading jet and figures 5.7 and 5.8 for the second leading jet. The W + 3 jets distributions for the leading jet are in figures 5.9 and 5.10, and the ones for the third leading jet in figures 5.11 and 5.12.

Each plot shows the SISCone and the anti-k_T algorithm results of the transverse momentum distributions for one given jet parameter R at the respective level. The bottom part of each plot shows the ratio of $\frac{p_T(SISCone)}{p_T(anti-k_T)}$ in each p_T bin. The distributions are available at parton and cluster levels. The cluster level plots include an additional distribution that has been corrected for underlying event as described in section 5.4.1. These show two different ratios: One for the algorithms at the uncorrected level, and one at the corrected cluster level.

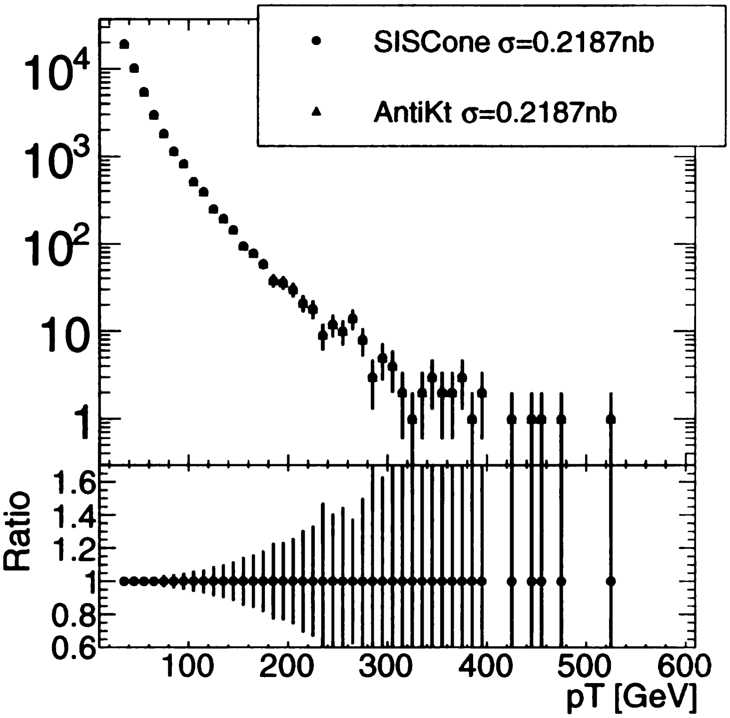


(a) Parton level, R = 0.4



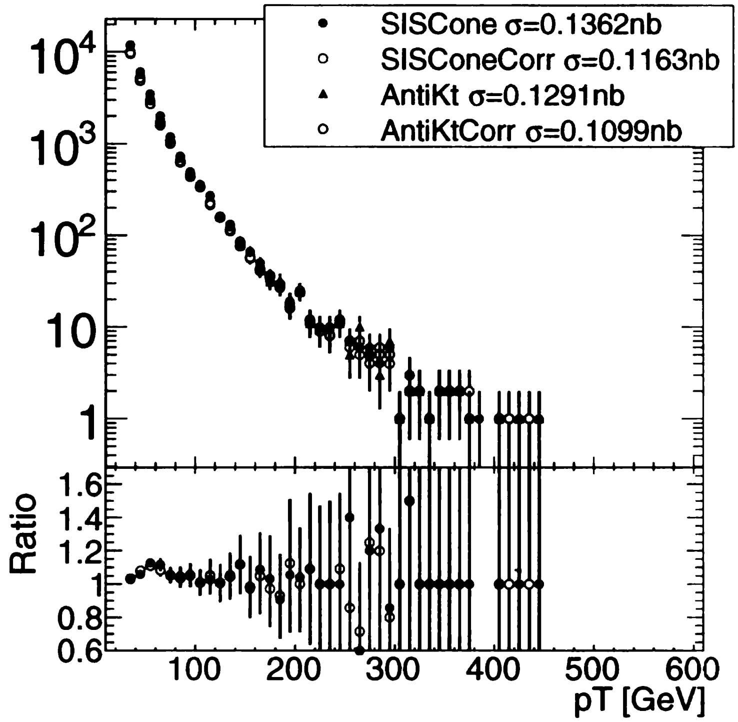
(b) Parton level, R = 0.7

Figure 5.3: Transverse momentum distribution of the jet in an exclusive W + 1 jet sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at parton level.

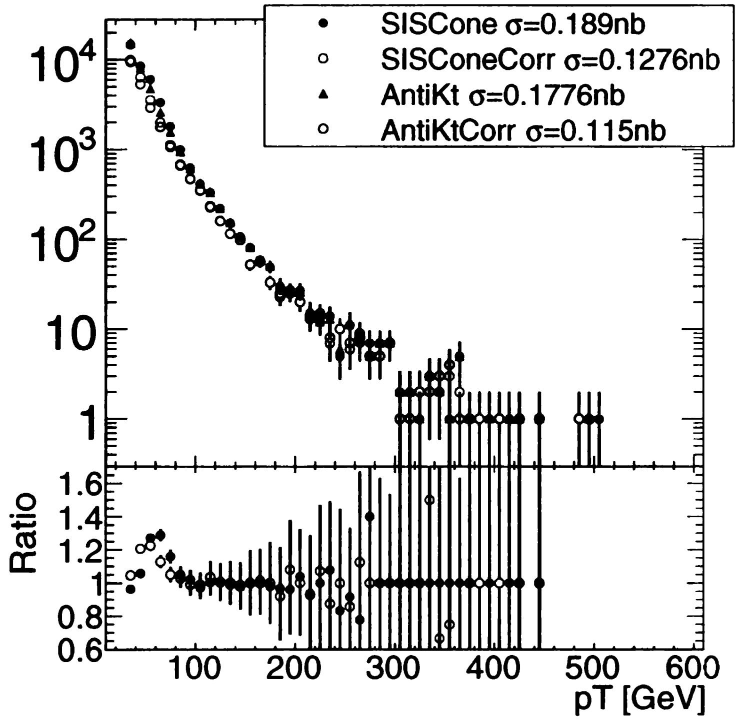


(c) Parton level, R = 1.0

Figure 5.3: [continued] Transverse momentum distribution of the jet in an exclusive W + 1 jet sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at parton level.

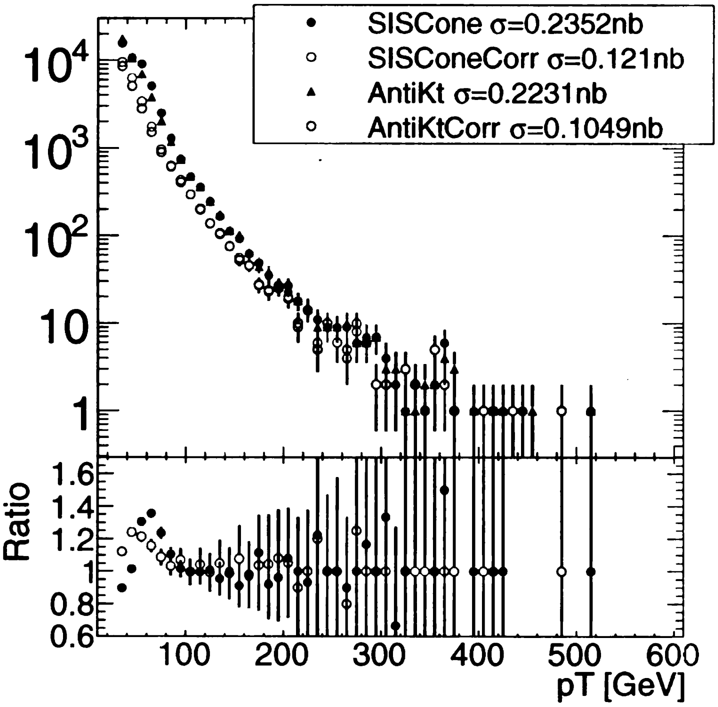


(a) Cluster level, R = 0.4



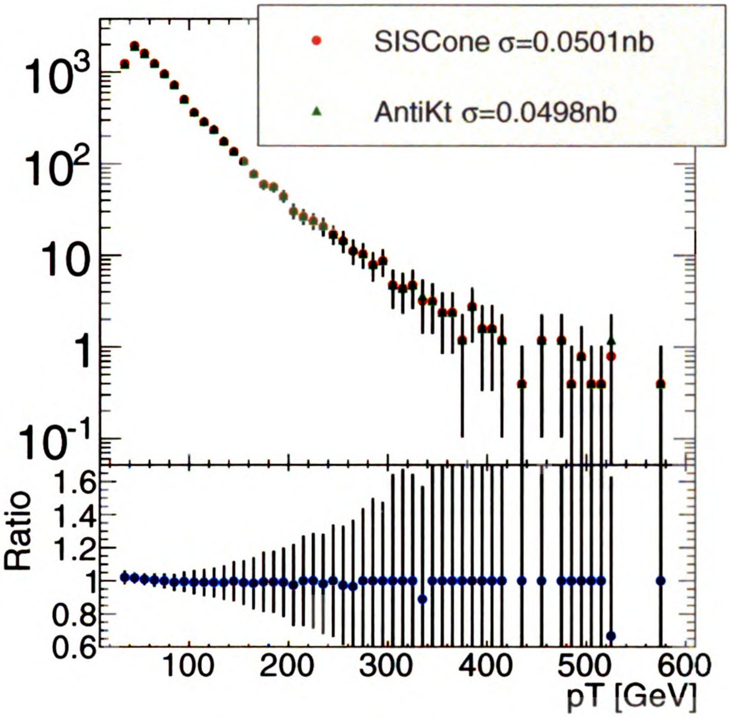
(b) Cluster level, R = 0.7

Figure 5.4: Transverse momentum distribution of the jet in an exclusive W + 1 jet sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at cluster level.

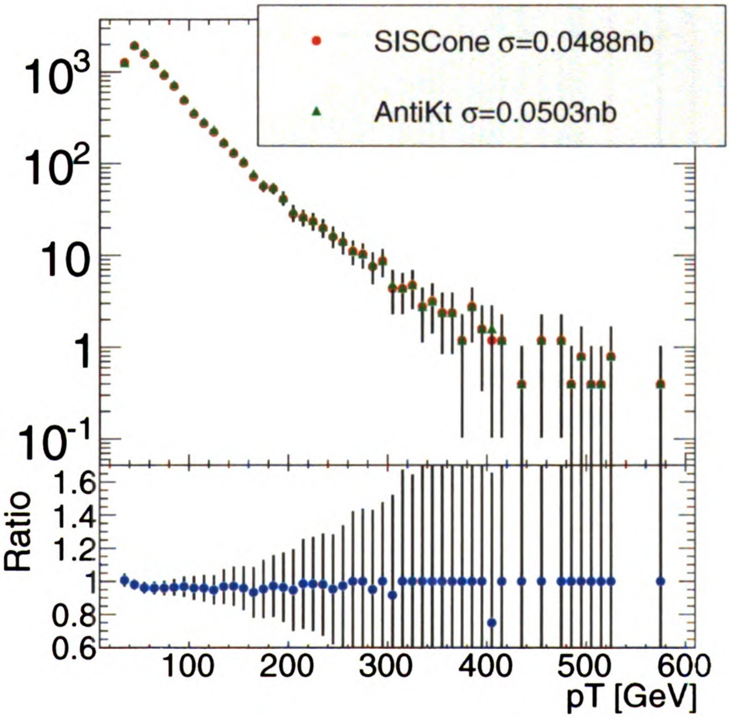


(c) Cluster level, R = 1.0

Figure 5.4: [continued] Transverse momentum distribution of the jet in an exclusive W + 1 jet sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at cluster level.

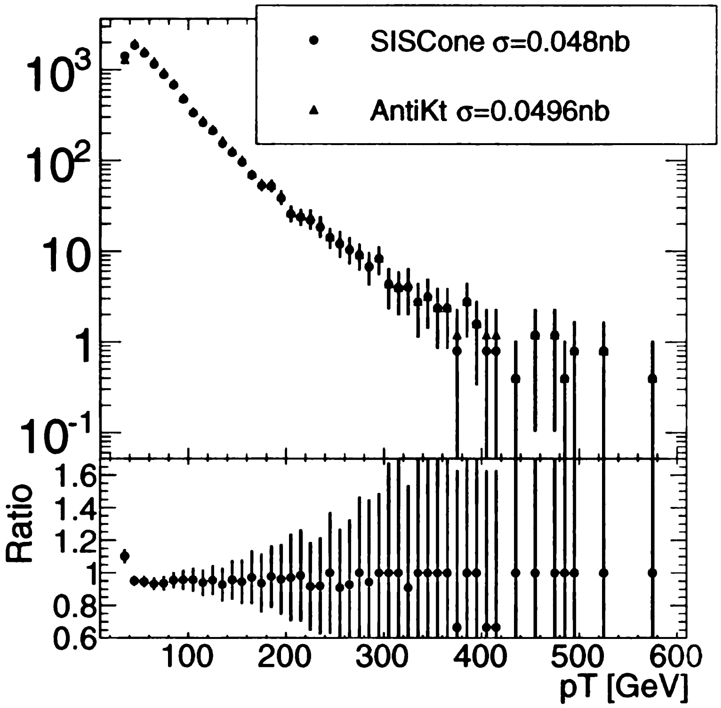


(a) Parton level, R = 0.4



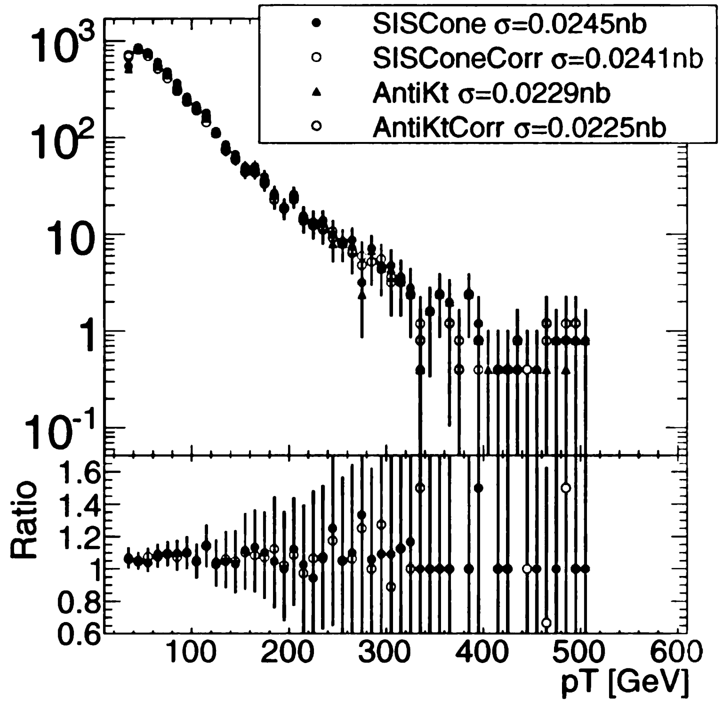
(b) Parton level, R = 0.7

Figure 5.5: Transverse momentum distribution of the leading jet in an exclusive W+ 2 jets sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at parton level.

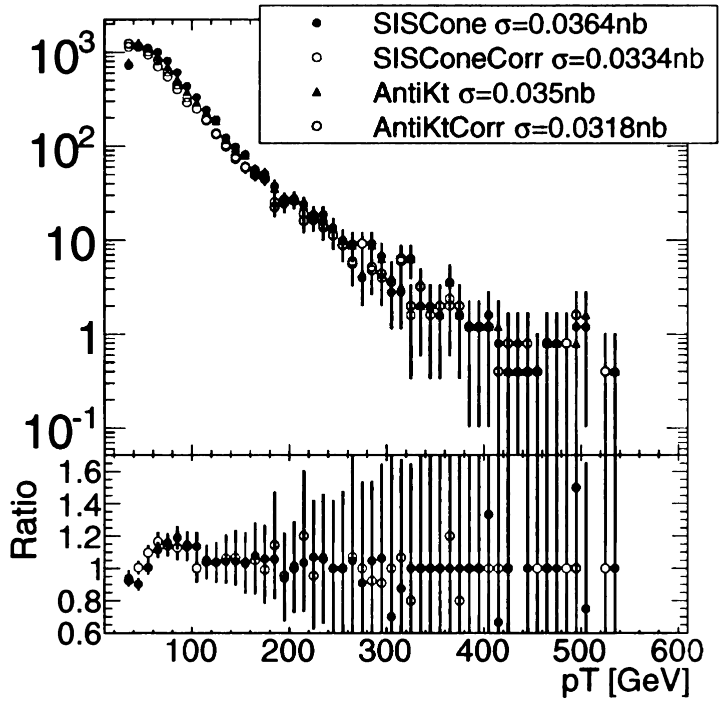


(c) Parton level, R = 1.0

Figure 5.5: [continued] Transverse momentum distribution of the leading jet in an exclusive W+ 2 jets sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at parton level.

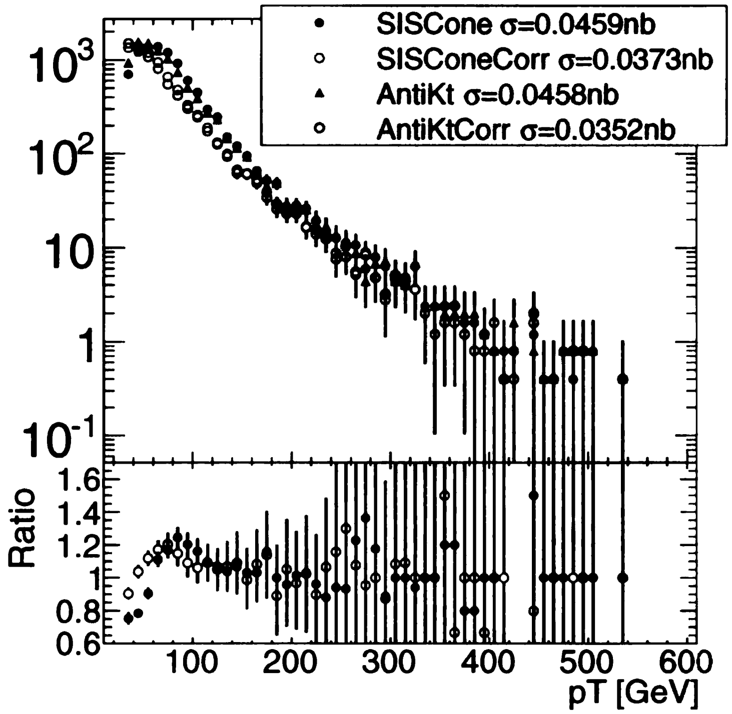


(a) Cluster level, R = 0.4



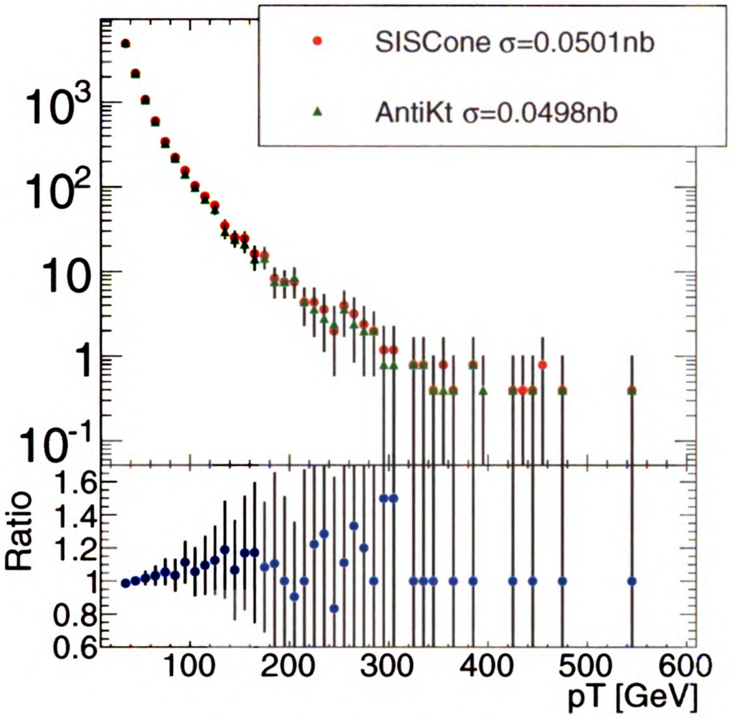
(b) Cluster level, R = 0.7

Figure 5.6: Transverse momentum distribution of the leading jet in an exclusive W+ 2 jets sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at cluster level.

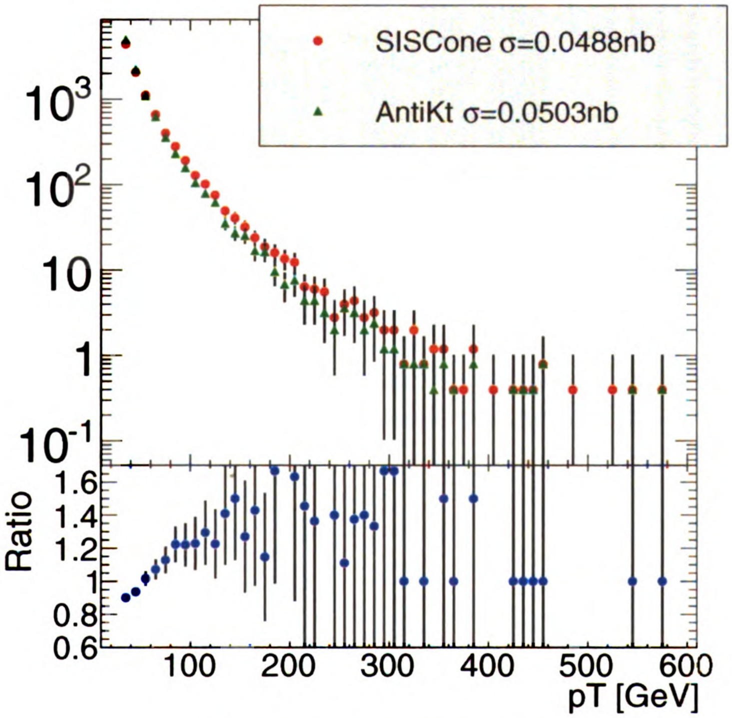


(c) Cluster level, R = 1.0

Figure 5.6: [continued] Transverse momentum distribution of the leading jet in an exclusive W+ 2 jets sample: Comparison and ratio of the cross sections from SISCone and anti-k_T at cluster level.

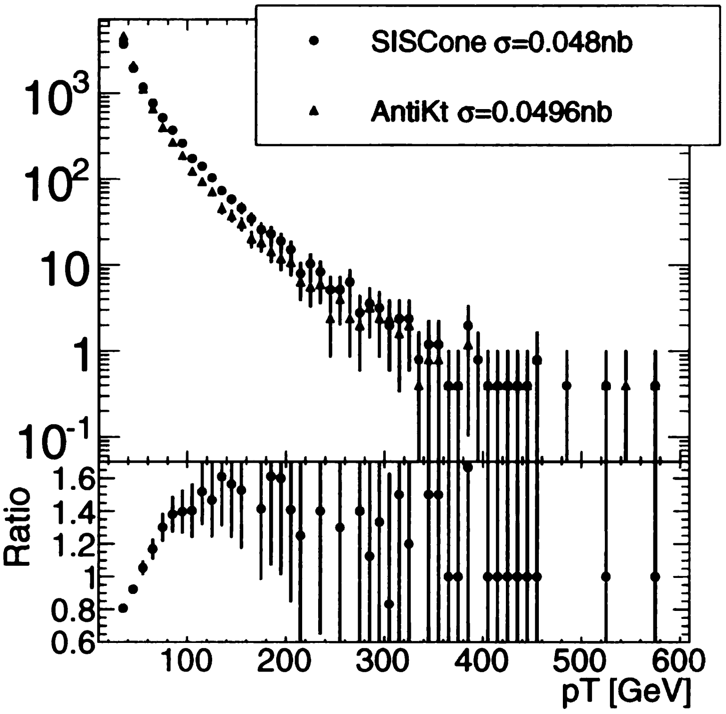


(a) Parton level, R = 0.4



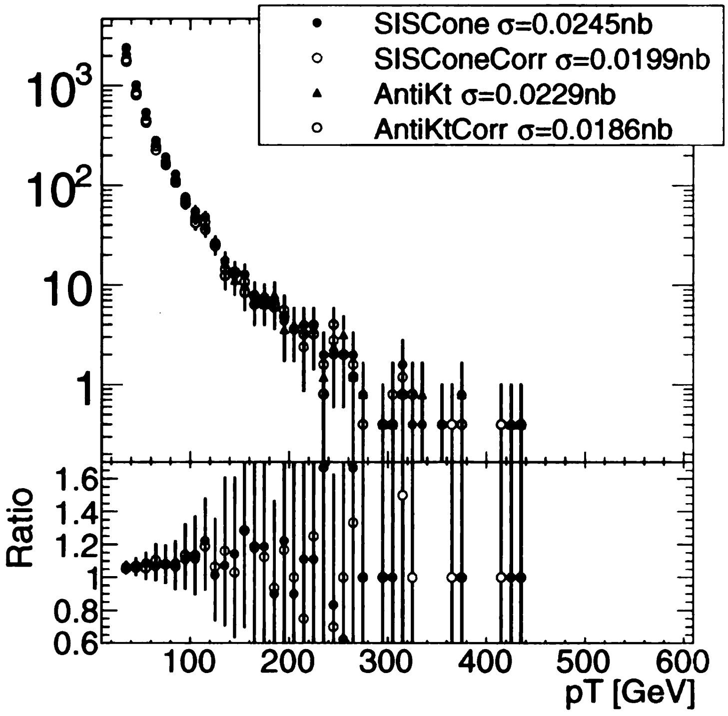
(b) Parton level, R = 0.7

Figure 5.7: Transverse momentum distribution of the second leading jet in an exclusive W+ 2 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at parton level.

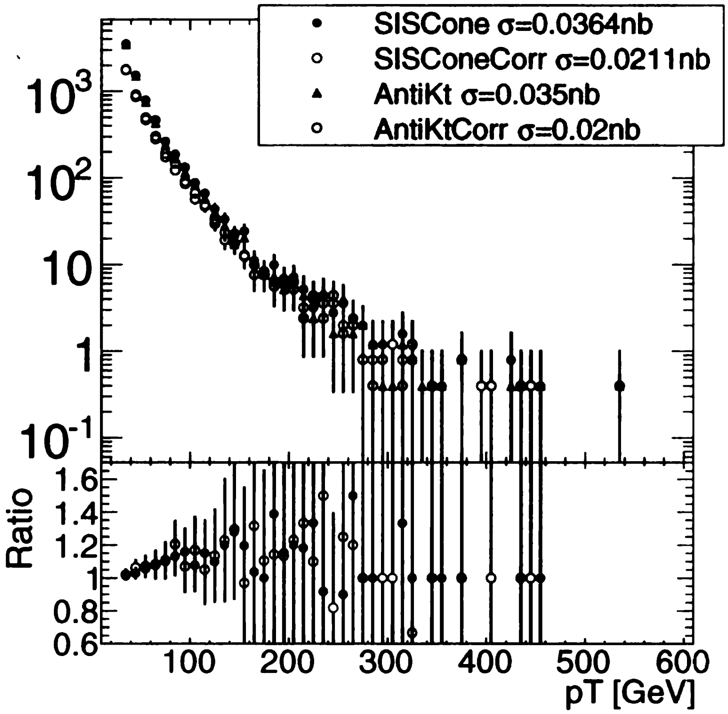


(c) Parton level, R = 1.0

Figure 5.7: [continued] Transverse momentum distribution of the second leading jet in an exclusive W+ 2 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at parton level.

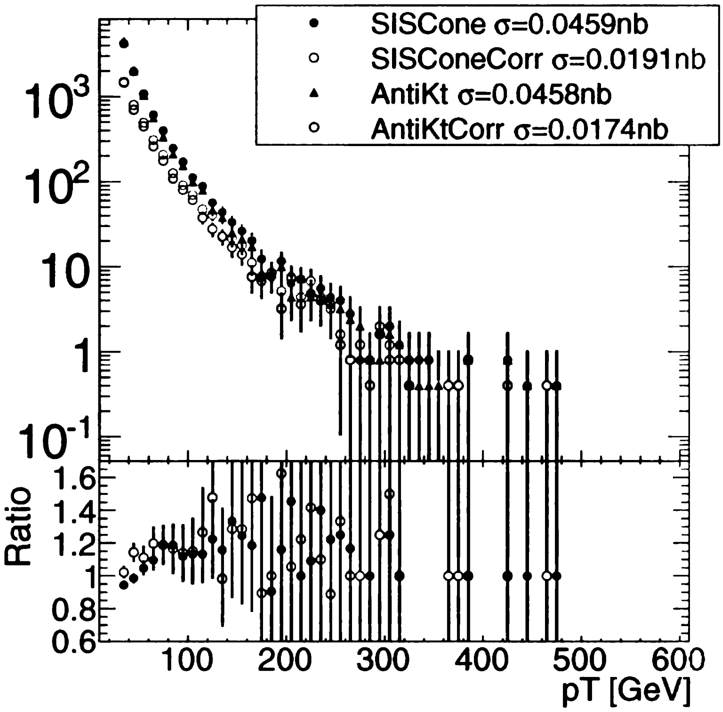


(a) Cluster level, R = 0.4



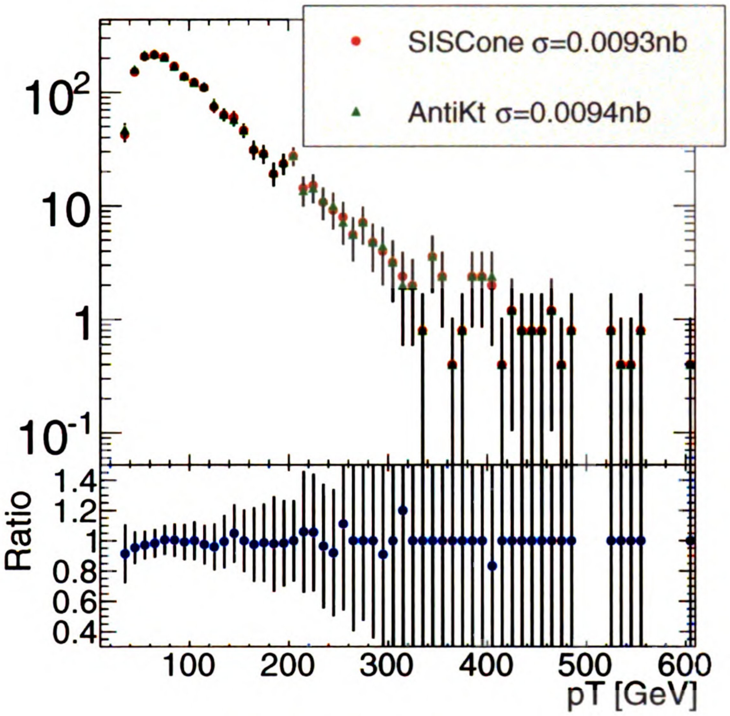
(b) Cluster level, R = 0.7

Figure 5.8: Transverse momentum distribution of the second leading jet in an exclusive W + 2 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at cluster level.

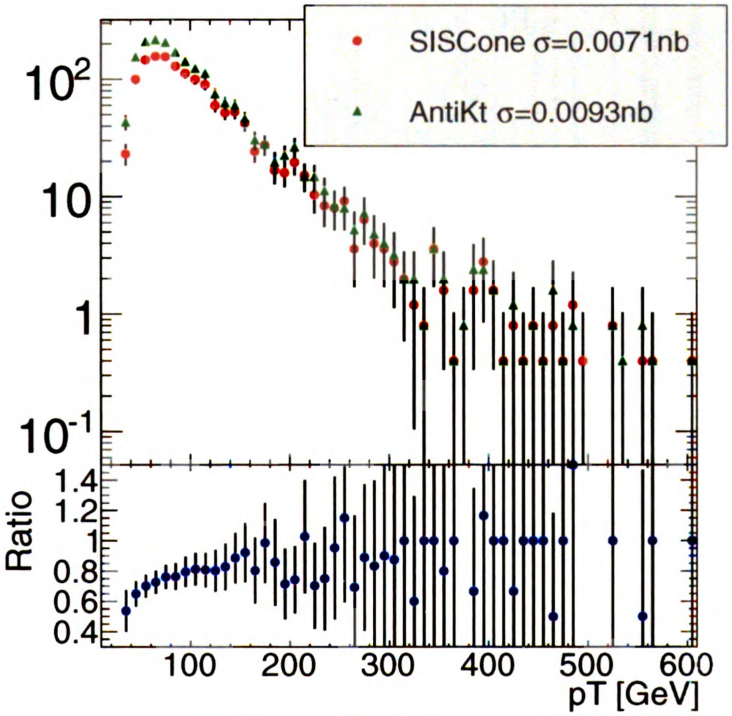


(c) Cluster level, R = 1.0

Figure 5.8: [continued] Transverse momentum distribution of the second leading jet in an exclusive W + 2 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at cluster level.

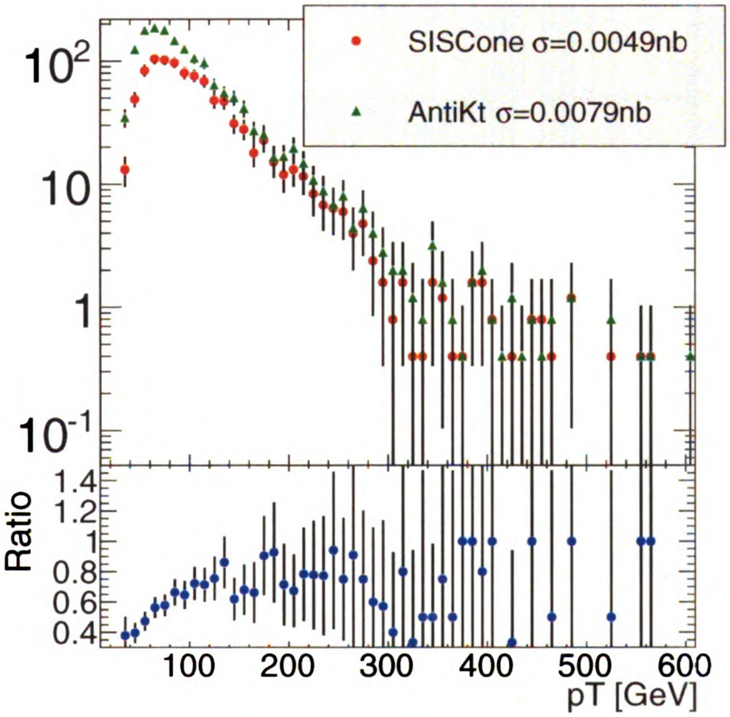


(a) Parton level, R = 0.4



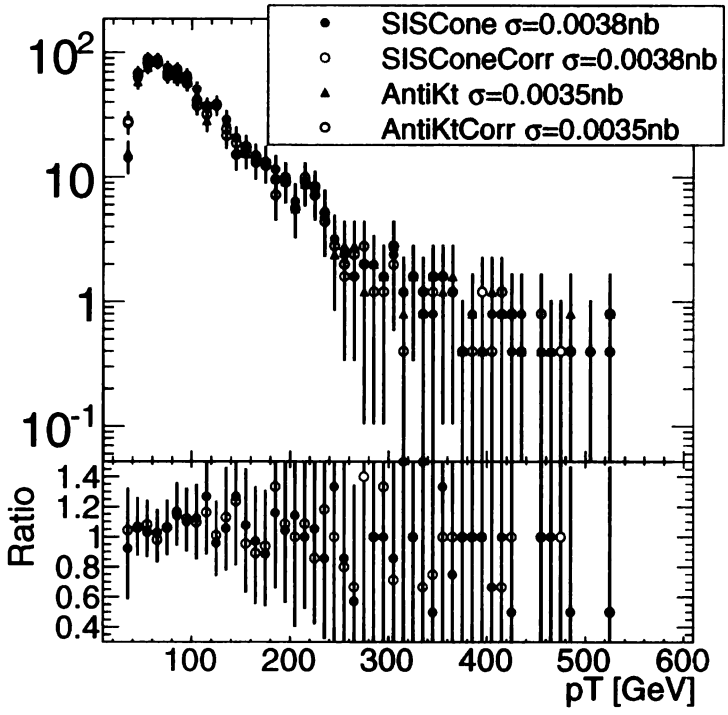
(b) Parton level, R = 0.7

Figure 5.9: Transverse momentum distribution of the leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at parton level.

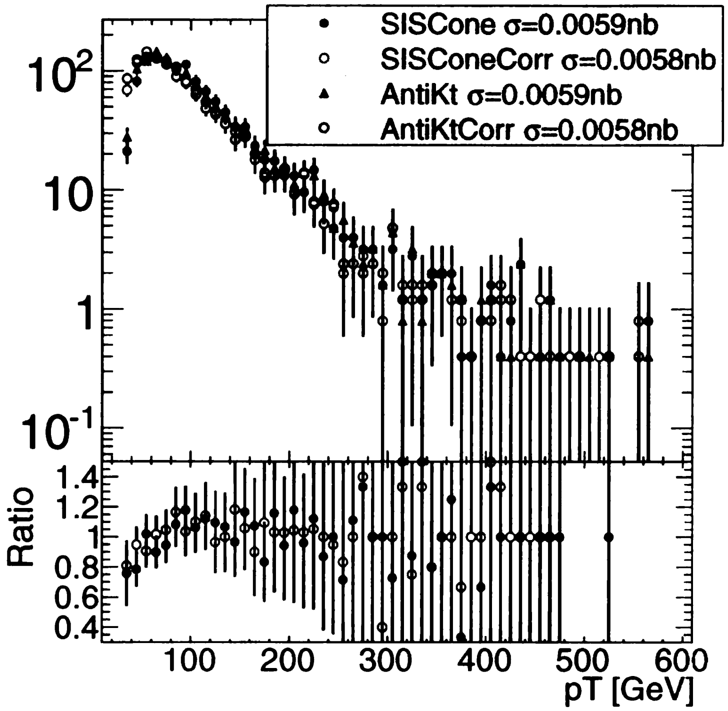


(c) Parton level, R = 1.0

Figure 5.9: [continued] Transverse momentum distribution of the leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at parton level.

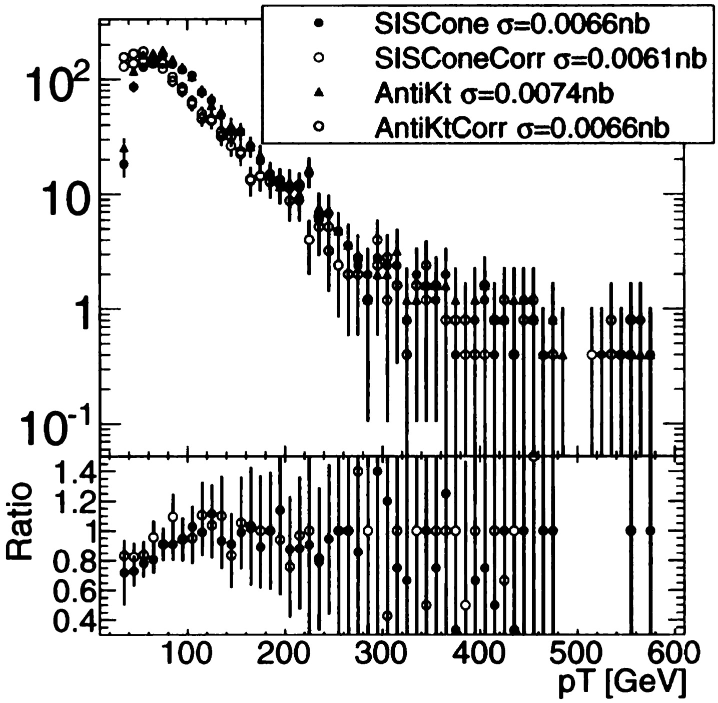


(a) Cluster level, R = 0.4



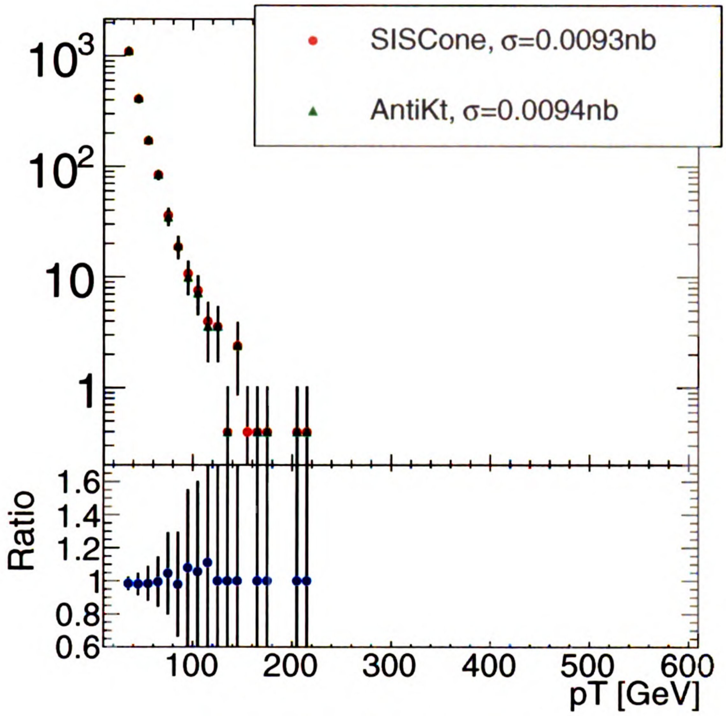
(b) Cluster level, R = 0.7

Figure 5.10: Transverse momentum distribution of the leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at cluster level.

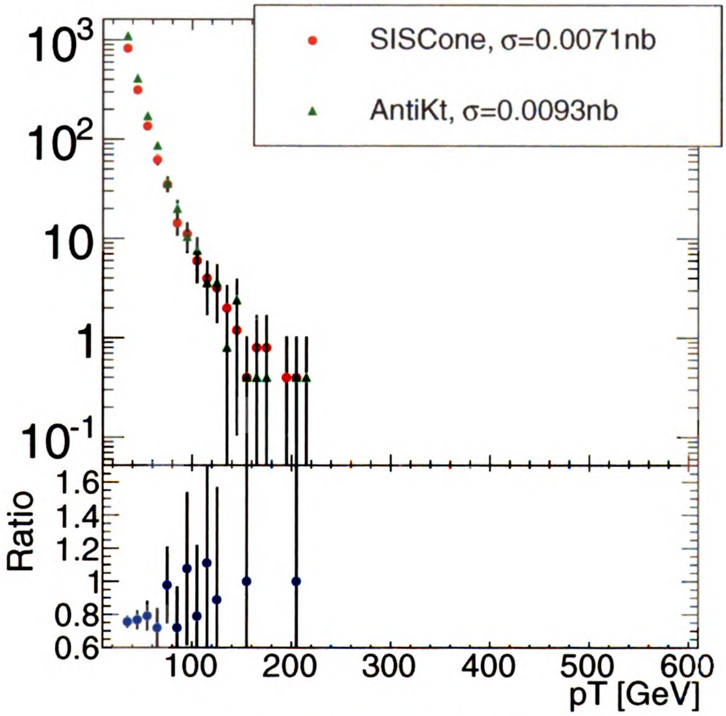


(c) Cluster level, R = 1.0

Figure 5.10: [continued] Transverse momentum distribution of the leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at cluster level.

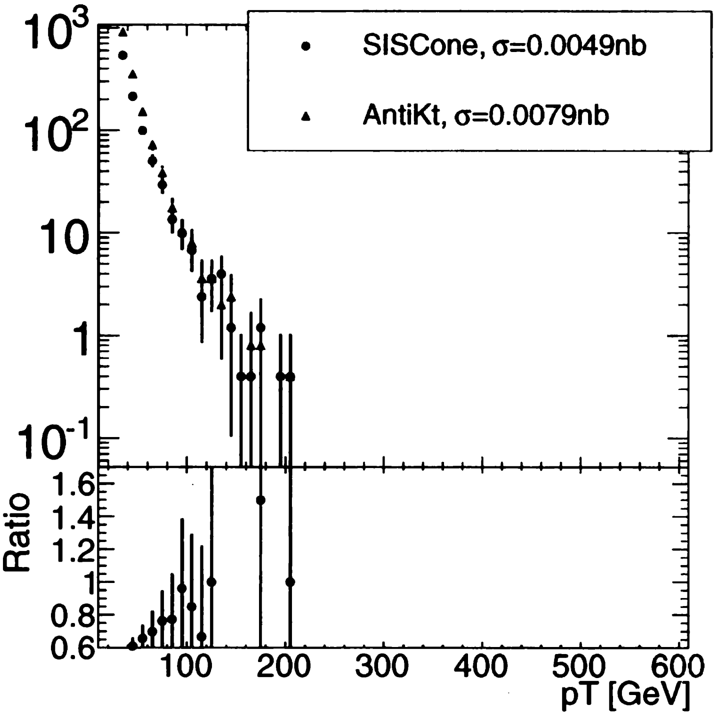


(a) Parton level, R = 0.4



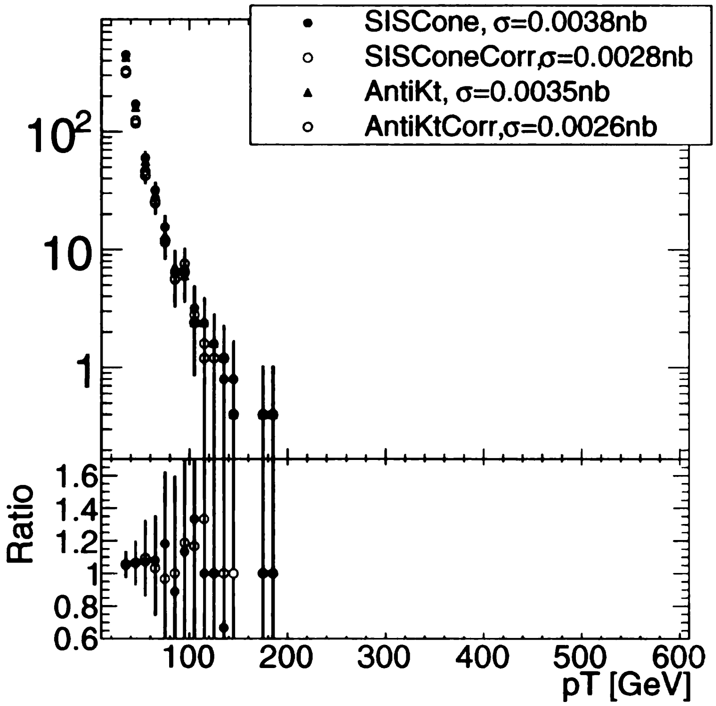
(b) Parton level, R = 0.7

Figure 5.11: Transverse momentum distribution of the third leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at parton level.

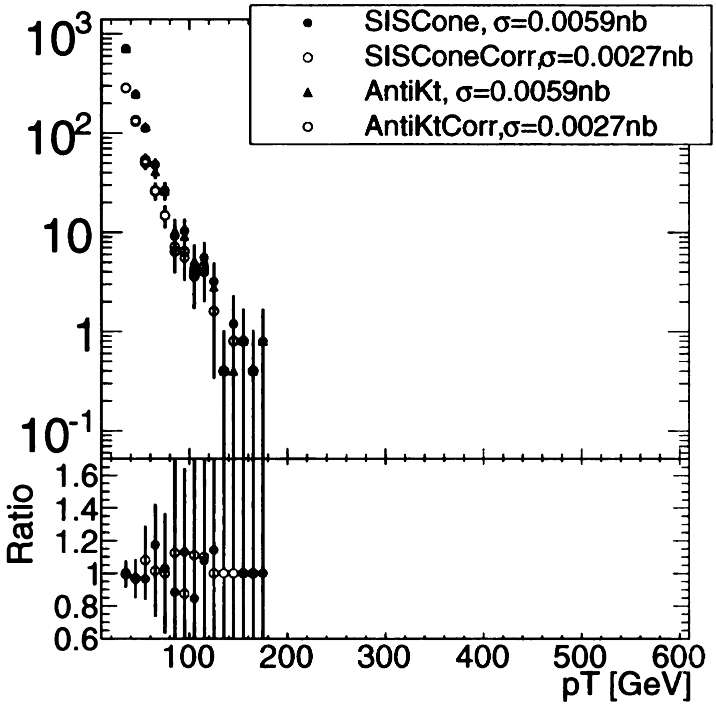


(c) Parton level, R = 1.0

Figure 5.11: [continued] Transverse momentum distribution of the third leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at parton level.

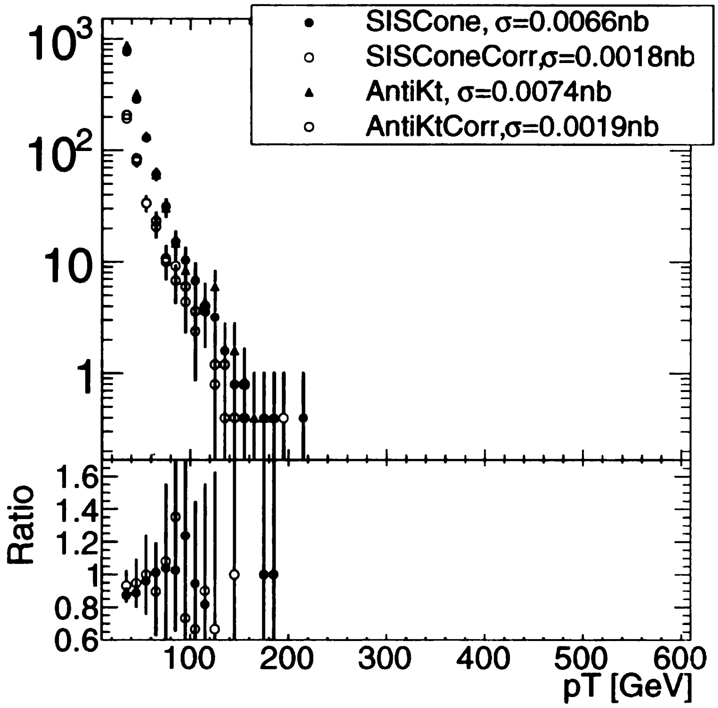


(a) Cluster level, R = 0.4



(b) Cluster level, R = 0.7

Figure 5.12: Transverse momentum distribution of the third leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at cluster level.



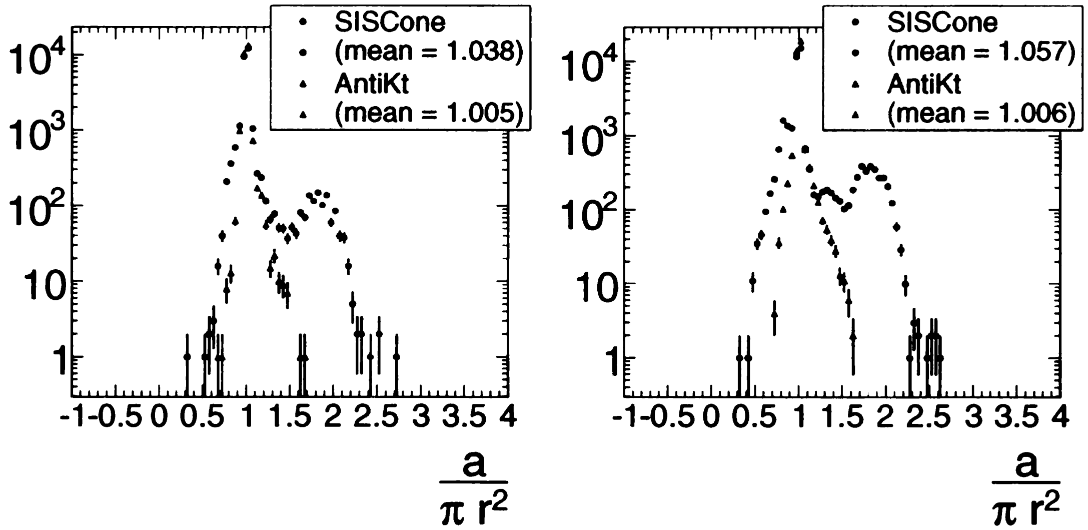
(c) Cluster level, R = 1.0

Figure 5.12: [continued] Transverse momentum distribution of the third leading jet in an exclusive W + 3 jets sample: Comparison and ratio of the cross section from SISCone and anti-k_T at cluster level.

At parton level for W + 1, all algorithms give the same results and their ratio equals unity. For $W \geq 2$ at parton level, and for all multiplicities at cluster level, there are differences between the two algorithms. The SISCone total cross sections at cluster level tend to be higher than the anti-k_T cross sections when considering W + 1 or 2 jets. In low p_T regions, SISCone cross sections are larger. In high p_T regions, the ratio of the two algorithms' cross sections approaches unity. This can be explained by the fact that areas of jets reconstructed with SISCone are usually larger than the ones found by recombination algorithms. Thus there are more objects whose momentum is recombined to the final jet's momentum, so the jets found with SISCone tend to have higher transverse momentum than those reconstructed with recombination algorithms.

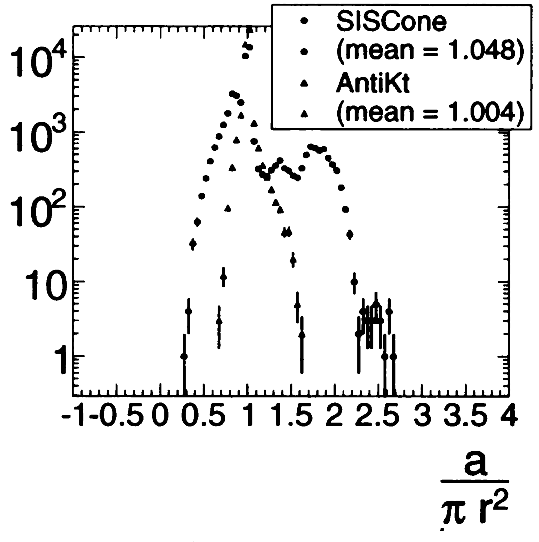
Figures 5.13 and 5.14 show that the jet areas of the SISCone algorithm are larger than the anti-k_T areas in the case of W + 1 and W + 2 jets. They display the ratio $\frac{a}{\pi R^2}$ of the leading jet area to the size of a circle in the $\eta - \phi$-plane with radius R corresponding to the jet size parameter. The distributions show, that the SISCone area distributions are wider, and especially show a second peak at larger areas that indicates that SISCone merges more soft particles in the jets than anti-k_T. The split-merge procedure is responsible for the discrepancy of the SISCone area to a regular cone.

However, when going to higher multiplicity samples such as W + 3 jets, another property of the SISCone algorithm comes into play: As the SISCone jet area is larger than the area of recombination algorithms, SISCone is more likely to merge two of the jets in the event that are not merged by the anti-k_T algorithm. If SISCone only finds two instead of three jets because it merges two of the jets, this event is removed from the distribution and is not counted in the cross section. This behavior does not have a large impact for smaller multiplicities as phase space is large enough so the jets do not get merged. For higher multiplicities, when the phase space for one jet decreases,



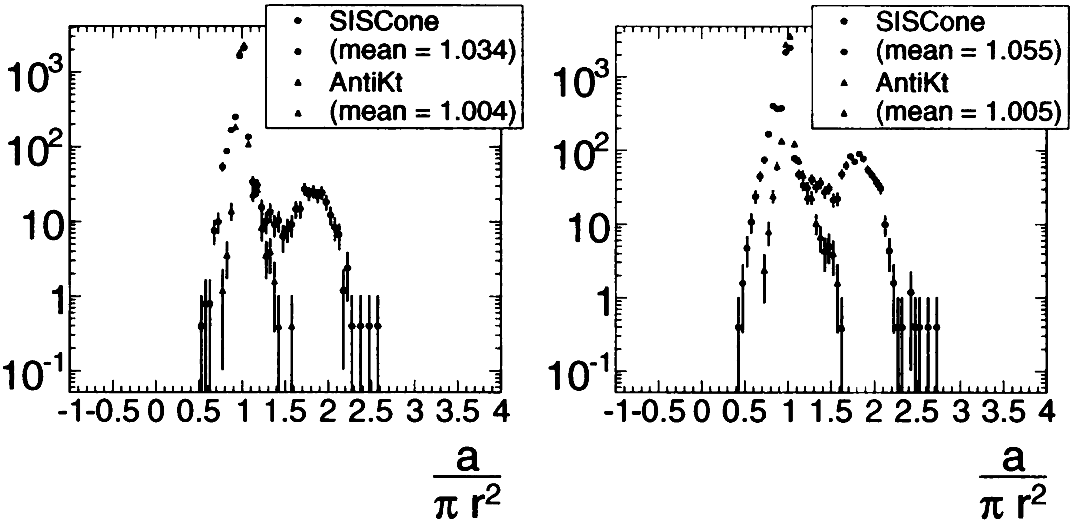
(a) R = 0.4

(b) R = 0.7

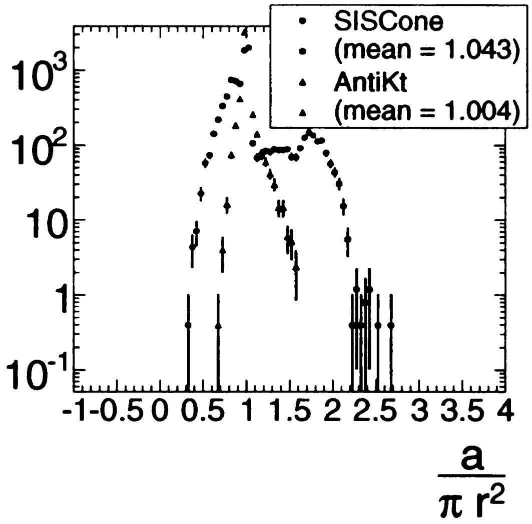


(c) R = 1.0

Figure 5.13: Size of the leading jet area $\frac{a}{\pi r^2}$ found with SISCone and anti-k_T for $W + 1$ jet events.



(a) R = 0.4 (b) R = 0.7



(c) R = 1.0

Figure 5.14: Size of the leading jet area $\frac{a}{\pi r^2}$ found with SISCone and anti-k_T for $W + 2$ jets events.

the SISCone cross section decreases with respect to recombination algorithm cross sections.

The fact that SISCone has a larger area also explains why the SISCone cross sections decrease when the jet size parameter increases: The probability of merging two of the jets produced in the event is increased when the jet size R is larger. If two jets are merged, the event is removed from the exclusive distribution, which decreases SISCone's cross section.

As the plots show, the differences between the algorithms do not depend on whether leading, second leading, or third leading jets are considered in the distribution.

At larger values of R, the differences between the algorithms increase, as the comparison of the distributions for $R = 0.4$, 0.7, and 1.0 show. At all levels, both algorithms are fairly similar for a jet size of $R = 0.4$. They are still similar for $R = 0.7$, but their distributions show bigger differences for a jet size of $R = 1.0$. This can be seen for example comparing figures 5.9a ($R = 0.4$), 5.9b ($R = 0.7$), and 5.9c ($R = 1.0$). This behavior could indicate that a jet size of $R = 0.5$ or 0.6 should be investigated as well. As both algorithms tend to agree more at a value of $R = 0.4$ or 0.7, the reconstruction results with a jet size parameter in this range could be more accurate.

5.4.4 Underlying event correction

The underlying event corrected cross section is a few percent smaller than the uncorrected one, since the jet transverse momentum is decreased by the amount of the underlying event p_T according to equation (5.3). The jets have to pass a p_T cut of 30 GeV after the correction, which not all of the corrected jets pass. This decreases the cross section of the corrected jets.

The histograms show that the subtraction from anti-k_T is slightly larger than what has been subtracted from SISCone. To explain this, consider equation (5.3). Two effects compete in this case: SISCone has the larger area of the two algorithms, which has been shown above in figures 5.13 and 5.14. This leads to a larger subtraction $a \cdot \rho_{p_T}$ for SISCone. However, $p_{T,uncorr}$ is larger for SISCone as well. All in all, the plots show that the SISCone correction is smaller than the anti-k_T correction.

$$p_{T,corr} = p_{T,uncorr} - \rho_{p_T} \cdot a \tag{5.3}$$

All of the following comparison plots show that the differences in cross sections of SISCone and anti-k_T depend on the jet parameter used for the reconstruction. In general, for larger values of R, the differences of the cross sections increase. This indicates that a smaller jet size, possibly in the area between 0.4 and 0.7, might be more accurate for all algorithms. This behavior is present at all levels and for all jet multiplicities except for the cluster level of the $W + 2$ sample.

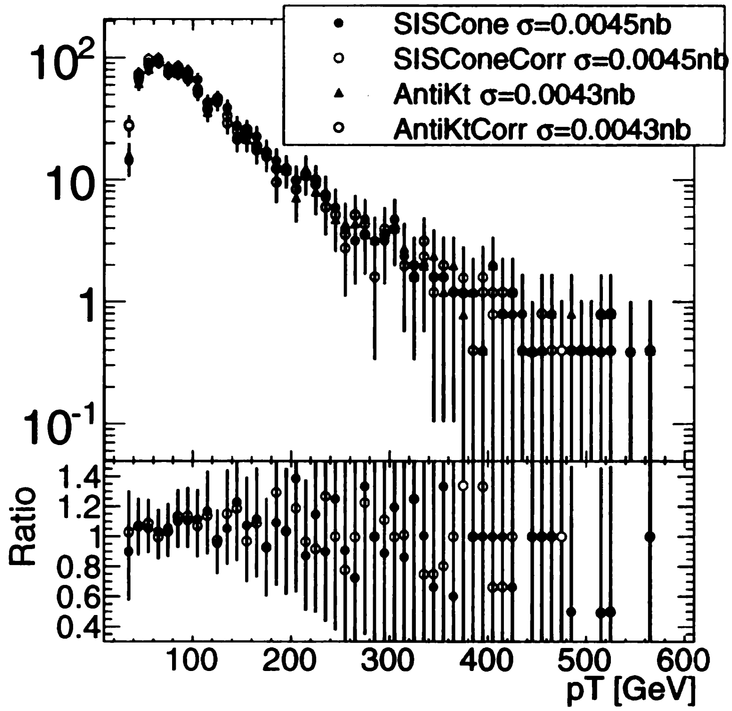
5.4.5 Inclusive $W+ \geq n$ jets algorithm comparison

Figures 5.15 and 5.16 show the inclusive p_T distributions at cluster level for inclusive $W+ \geq 3$ production that have been obtained by merging the weighted exclusive distributions of $W + 3$, 4, and 5 jets.

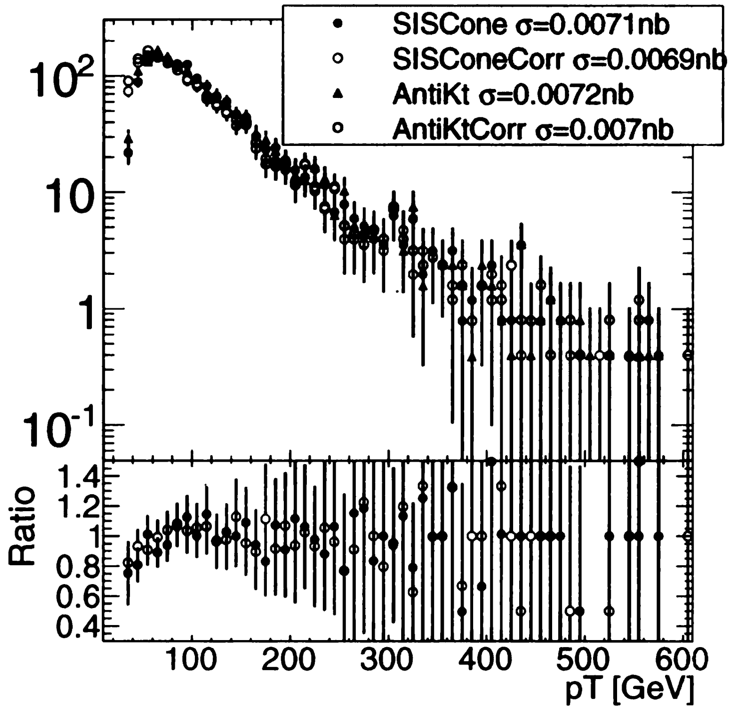
The inclusive distributions for a production of three or more jets do not show big differences from the exclusive production of 3 and only 3 jets: For the leading jet, compare the inclusive distributions in figure 5.15 to the exclusive distributions in figure 5.10; both are given at cluster level. For the third leading jet compare the inclusive distributions in figure 5.16 to the exclusive distributions in figure 5.12; again both at cluster level.

For both jets, the inclusive cross sections are slightly larger due to contributions from higher multiplicity events. However, the difference is not significant. Inclusive and exclusive W + 3 jets p_T distributions show very similar behavior. The SISCone cross section is slightly smaller than the anti-k_T cross sections like in the exclusive case. The cross sections of both algorithms are decreasing with increasing jet size R due to merging of jets.

The inclusive distributions for $W+ \geq 0$ jets production are given in chapter 6 to compare to the ATLAS data distributions (see figures 6.3 and 6.4).

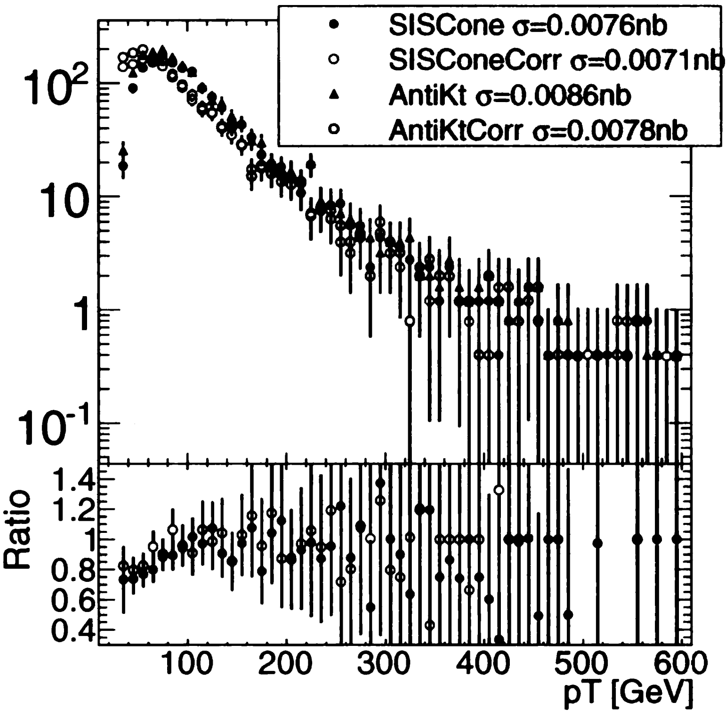


(a) $W+\geq 3$, R = 0.4



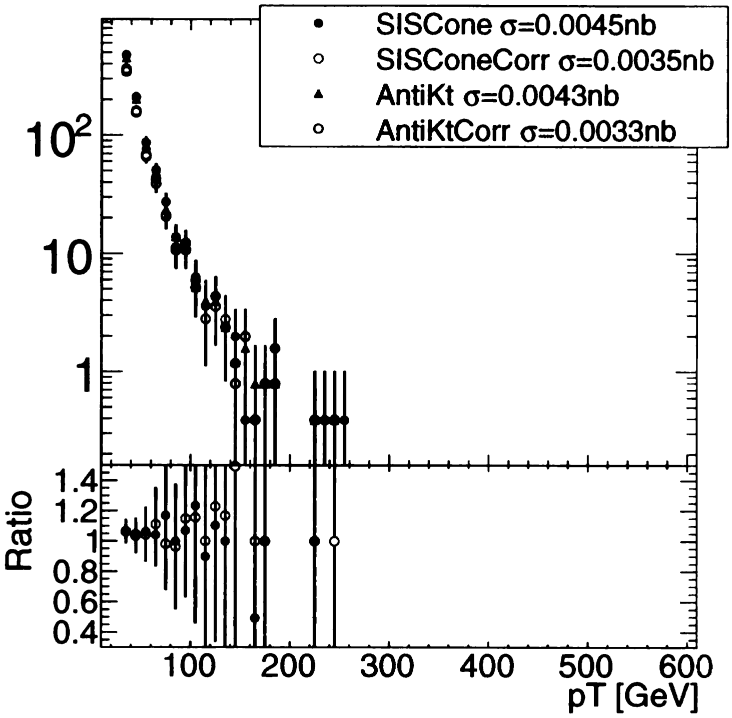
(b) $W+\geq 3$, R = 0.7

Figure 5.15: Leading jet p_T distributions for inclusive $W+\geq 3$ distributions at cluster level.

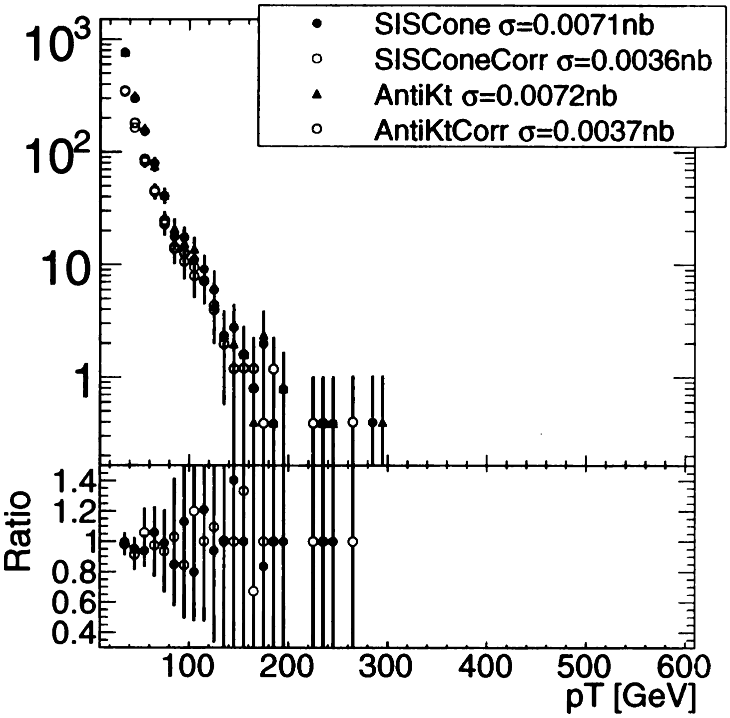


(c) $W+\geq 3$, R = 1.0

Figure 5.15: [continued] Leading jet p_T distributions for inclusive $W+\geq 3$ distributions at cluster level.

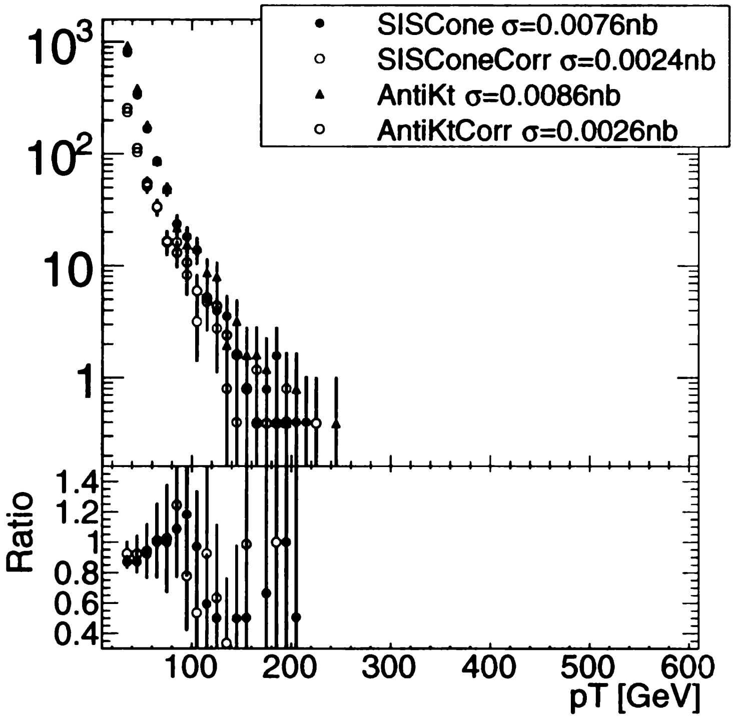


(a) $W+\geq 3$, R = 0.4



(b) $W+\geq 3$, R = 0.7

Figure 5.16: Third leading jet p_T distributions for inclusive $W+\geq 3$ distributions at cluster level.



(c) $W+\geq 3$, R = 1.0

Figure 5.16: [continued] Third leading jet p_T distributions for inclusive $W+\geq 3$ distributions at cluster level.

5.5 Comparison of levels of reconstruction

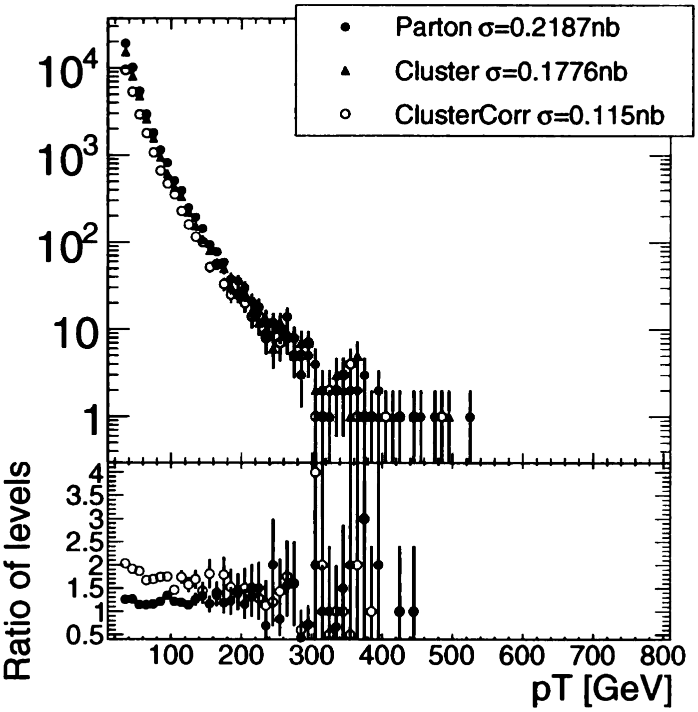
One of the requirements for jet algorithms is that they be equivalent at the parton, hadron, and cluster level. Of course, there are a number of differences between the levels that can change the results of the jet finding process: Besides running on different objects (i. e. partons and topo-clusters) on the different levels, underlying event and pileup are added at hadron and cluster level. However. results on the different levels should be as close as possible for a given jet definition.

In the analysis, other objects such as W bosons (at parton level) or electrons (at cluster and particle level) have to be removed. W bosons at parton level can be removed by SpartyJet directly by using the PDG ID selection tool. Electron or positron jets at hadron and cluster level have to be removed after the reconstruction. On these levels, electron and positron contributions are removed from the event by removing the nearest jet within $\Delta R = 0.2$ to a reconstructed electron or positron object.

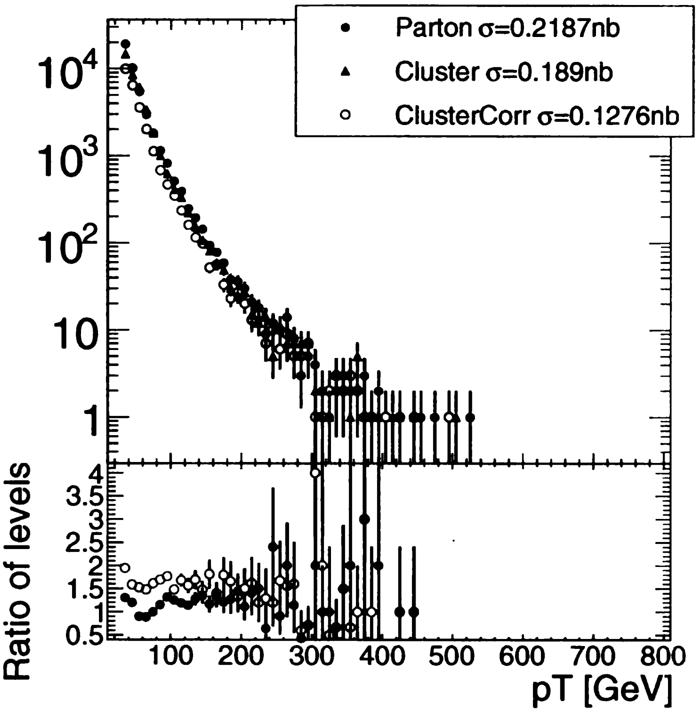
In this analysis, parton and hadron level jet reconstruction have been studied, as truth particle information was unavailable from the ATLAS Monte Carlos. However, the cluster level cross sections are not completely calibrated: Calibration of topo-clusters is based on cell signal weighting. The calorimeter is non-compensating, which means that its response to electromagnetic particles is higher than the response to hadronic particles. Therefore, the cluster signals of hadronic jets have to be corrected with an additional jet energy scale factor on top of the cell weighting in order to compare cluster level results to parton level calculations. In this analysis, the jet energy scale calibration is not complete, so the jet energies are between the electromagnetic and the full energy scale. For this reason, the cluster level transverse momenta are smaller than the parton level momenta. This leads to a uniform decrease of the cluster level cross sections with respect to the parton level cross sections.

Although the cluster level cross sections are not accurate, the shapes of cluster and parton level distributions are very similar. The histograms show that apart from the missing jet energy scale factor, both levels are fairly identical.

The following figures compare the distributions at cluster and parton levels. They contain the leading jet's transverse momentum distribution for one given jet algorithm with a radius of $R = 0.7$ at both levels. Figure 5.17 contains the level comparison for the W + 1 jet sample, figure 5.18 the one for the W + 2 jets sample, and figure 5.19 the one for the W + 3 jets sample.

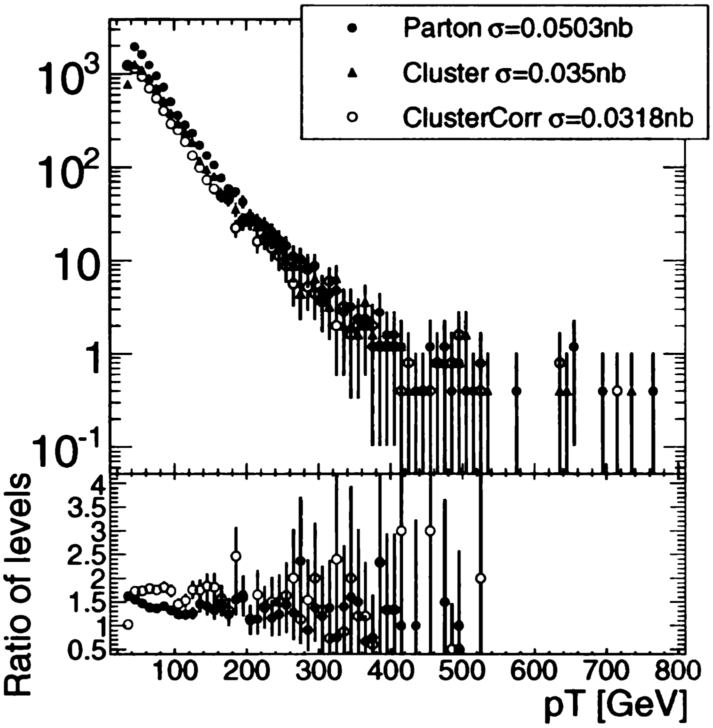


(a) anti-k_T

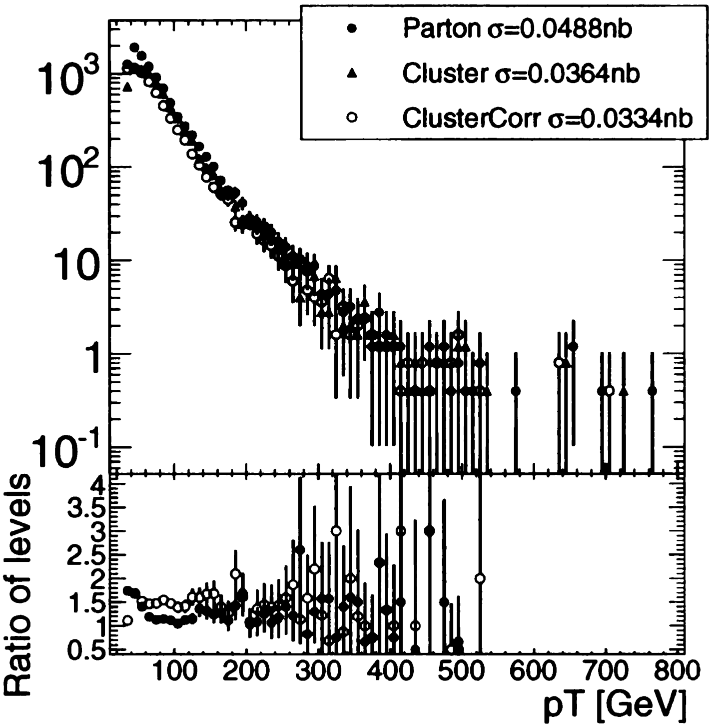


(b) SISCone

Figure 5.17: Level comparison for the W + 1 jet sample with a jet size R = 0.7 for both algorithms.

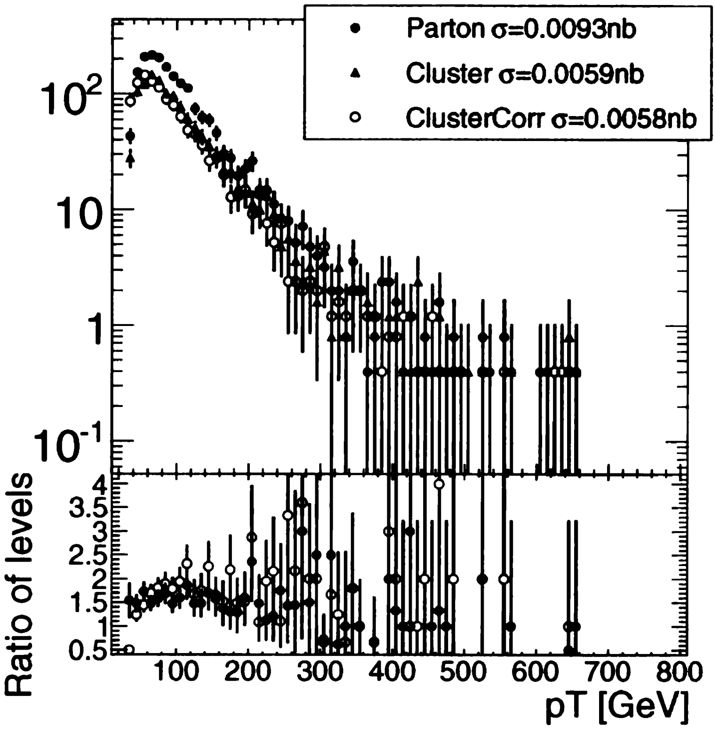


(a) anti-k_T

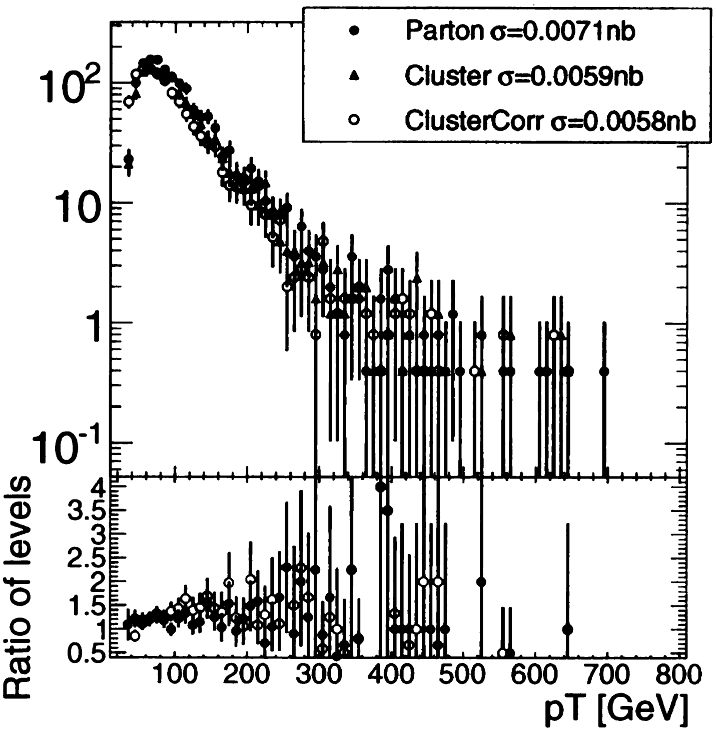


(b) SISCone

Figure 5.18: Level comparison for the W + 2 jets sample with a jet size R = 0.7 for both algorithms.



(a) anti-k_T



(b) SISCone

Figure 5.19: Level comparison for the W + 3 jets sample with a jet size R = 0.7 for both algorithms.

6. Study on ATLAS Data

The data available is from the L1Calo stream. Only those events that have at least one electron with a transverse momentum $p_T >$ 15 GeV are included.

6.1 Selection criteria and cuts

The cuts on the ATLAS Data are looser than the cuts on Monte Carlo samples due to the relatively small integrated luminosity currently available. The loosening of the cuts will allow more background into the sample, but will better allow the efficacy of SpartyJet in reconstructing ATLAS data to be demonstrated. These are the cuts in use for the data samples:

- $|\eta^{jet}| < 3.1$
- $p_T^{jet} > 20\,\mathrm{GeV}$

Kinematic cuts for leptons and reconstructed W bosons:

- $E_T^{miss} > 10\,\mathrm{GeV}$
- electron $p_T >$ 10 GeV
- electron $|\eta| < 2.47$, also excluding the crack region $1.37 < |\eta| < 1.52$
- electron IsEM cut is "loose"

- W transverse mass $m_T > 20\,\mathrm{GeV}$

The trigger cut is EF_e20_loose, which triggers on a single isolated electron.

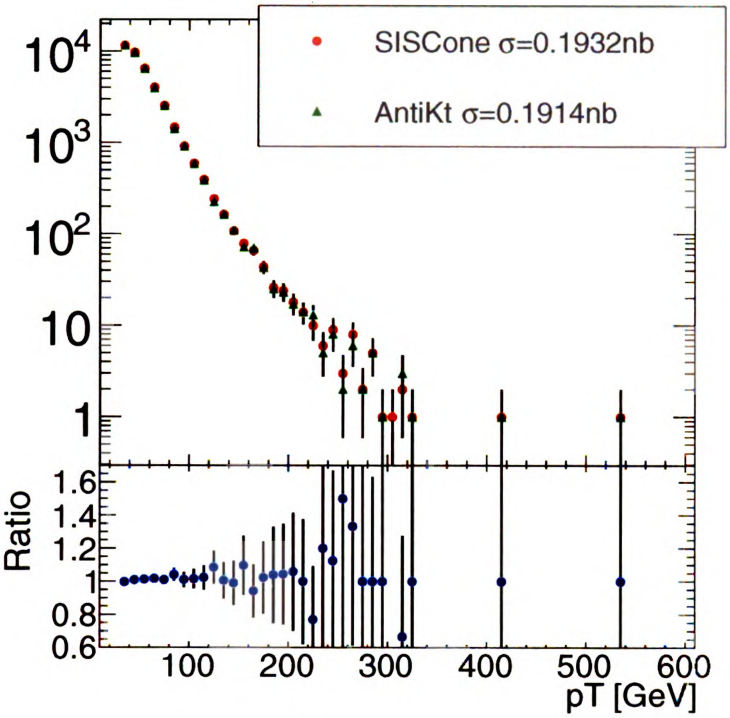
6.2 Transverse momentum distributions of ATLAS data

The data samples are only available on the cluster level, which corresponds to the detector output. In this case, pile-up effects are included. Similar to underlying event correction of the Monte Carlo cluster level, the data can be corrected for underlying event and pile-up effects by running a jet finding algorithm on all soft clusters with $p_T < 10$ GeV. These soft jets are subtracted from the event in the same manner as in the Monte Carlo analysis (see 5.4.1 for details on underlying event subtraction).

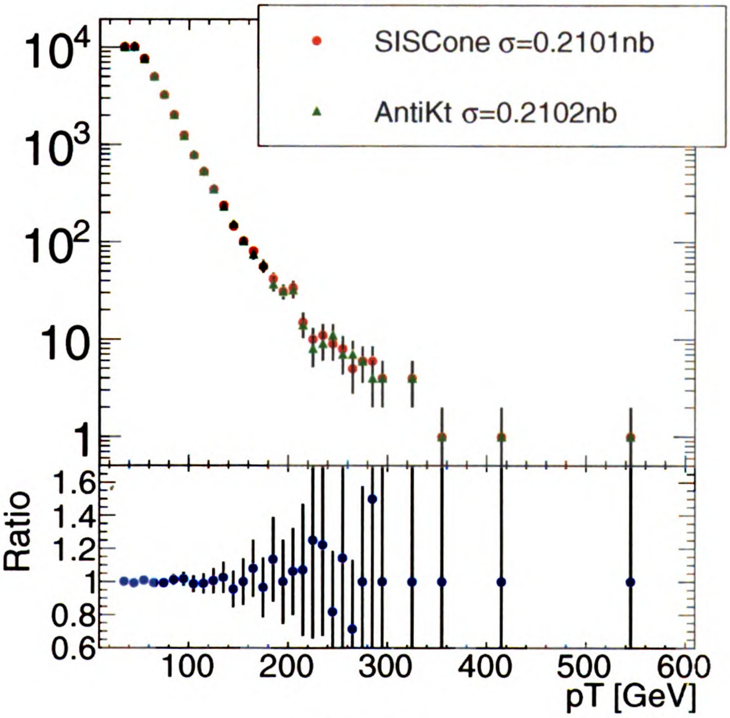
The following figures (6.1 to 6.2) show the transverse momentum distributions for the first and second leading jet, reconstructed from the data with anti-k_T or SISCone and different jet sizes R = 0.4, 0.7, 1.0.

In most cases, the SISCone cross section is slightly higher than the anti-k_T cross section over the entire transverse momentum range, especially for the higher p_T range. This is valid for both the the cluster and the corrected level. However, both algorithms are still quite similar.

As in the Monte Carlo study, the corrected cross section is few percent smaller than the uncorrected one.

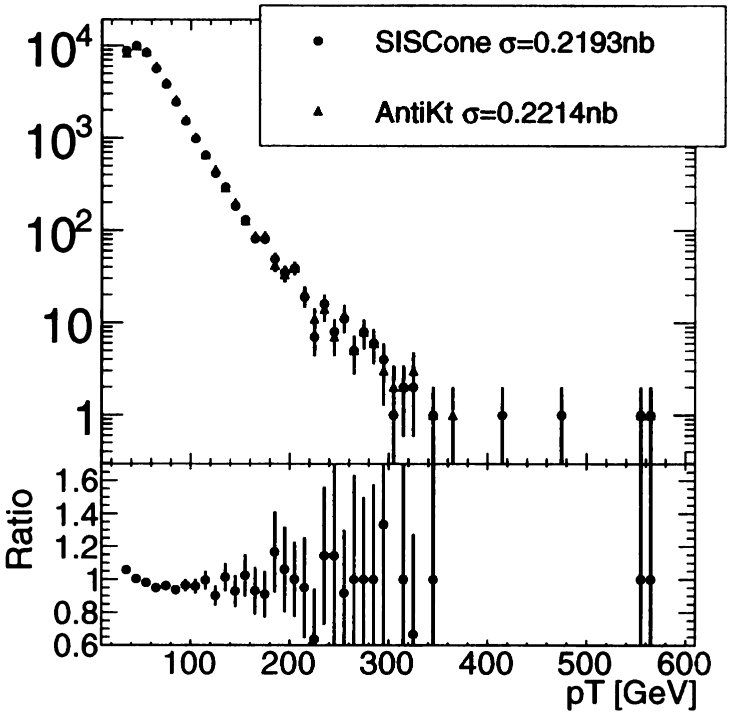


(a) R = 0.4



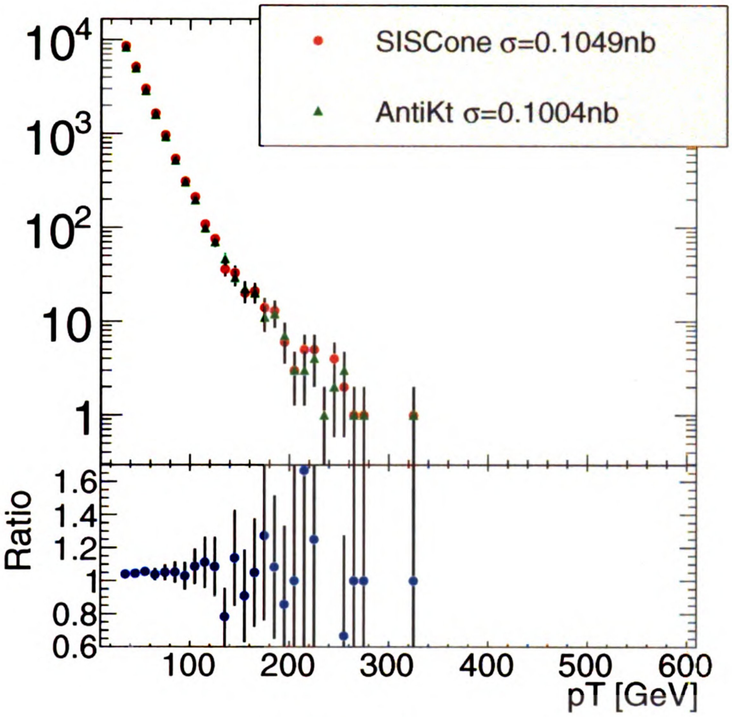
(b) R = 0.7

Figure 6.1: Transverse momentum distribution of the leading jet from the L1Calo data sample, comparing the SISCone and anti-k_T algorithms.

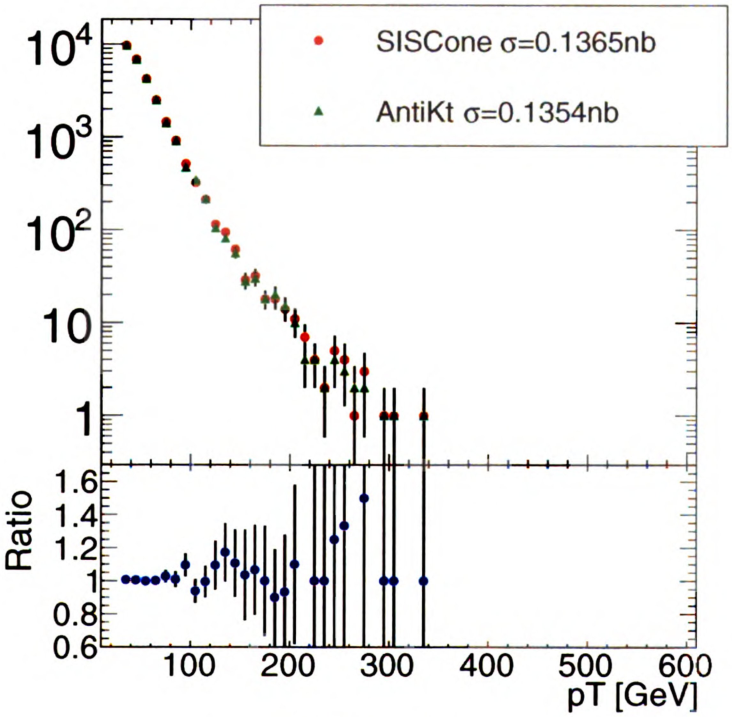


(c) R = 1.0

Figure 6.1: [continued] Transverse momentum distribution of the leading jet from the L1Calo data sample, comparing the SISCone and anti-k_T algorithms.

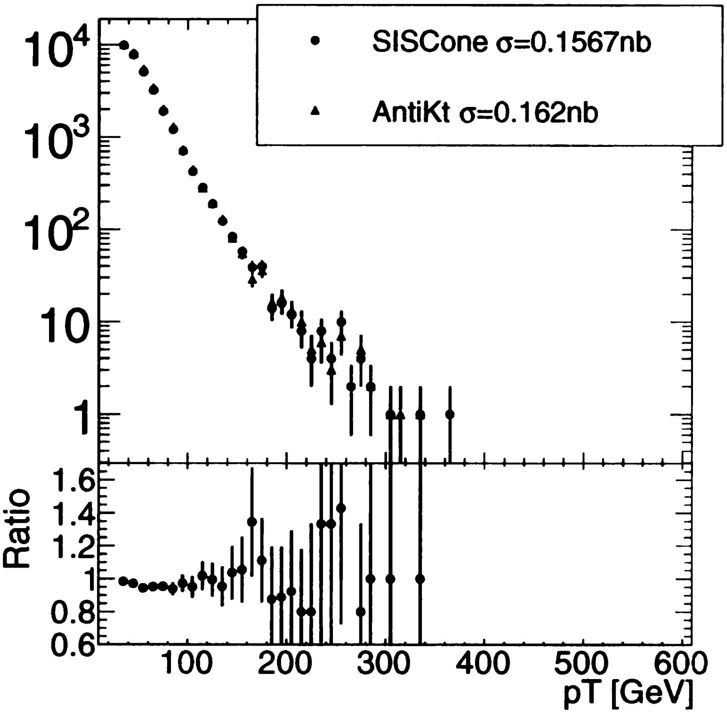


(a) R = 0.4



(b) R = 0.7

Figure 6.2: Transverse momentum distribution of the second leading jet from the L1Calo data sample, comparing the SISCone and anti-k_T algorithms.



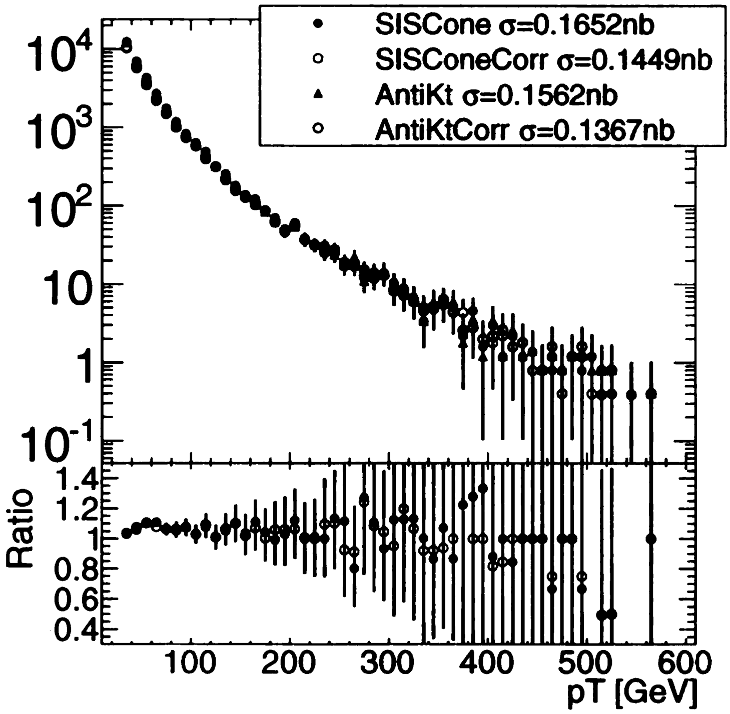
(c) R = 1.0

Figure 6.2: [continued] Transverse momentum distribution of the second leading jet from the L1Calo data sample, comparing the SISCone and anti-k_T algorithms.

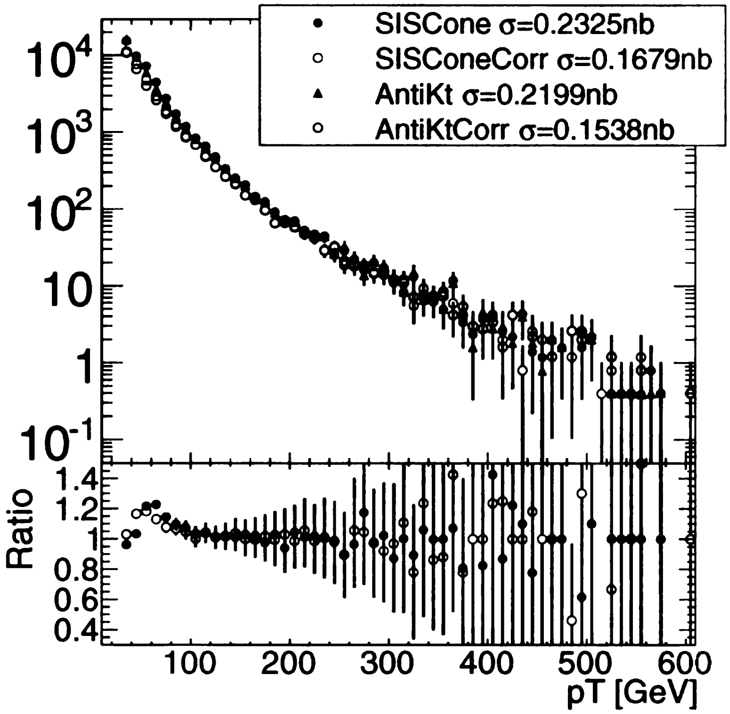
6.3 Qualitative comparison to Monte Carlo distributions

These distributions can be compared to cluster level Monte Carlo distributions. However, as the cluster level distributions used in this thesis are incompletely calibrated, only the shapes of the distributions can be compared since the Monte Carlo cross sections are not accurate. Therefore, different cuts for the data study are acceptable. For a direct comparison to Monte Carlo, the cuts would have to be identical.

Figures 6.3 and 6.4 show the inclusive Monte Carlo distributions for W + jets production at cluster level. They show similar shapes for the leading, second leading, and third leading jet p_T distributions as the ones reconstructed from data above. Also, the differences between the algorithms are similar in data as in Monte Carlo reconstruction. The data cross sections are higher due to looser cuts and incompletely calibrated cluster level Monte Carlo cross sections.

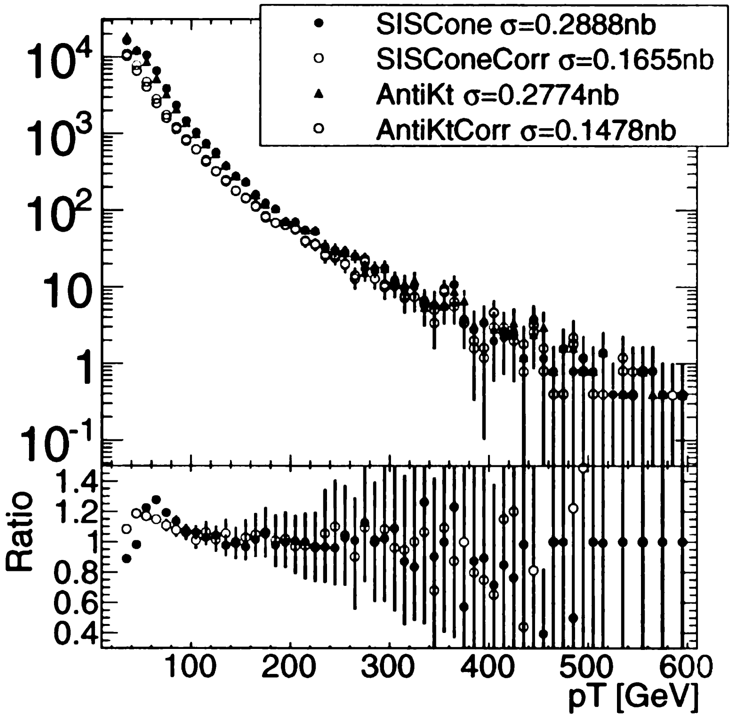


(a) $W+ \geq 0$, R = 0.4



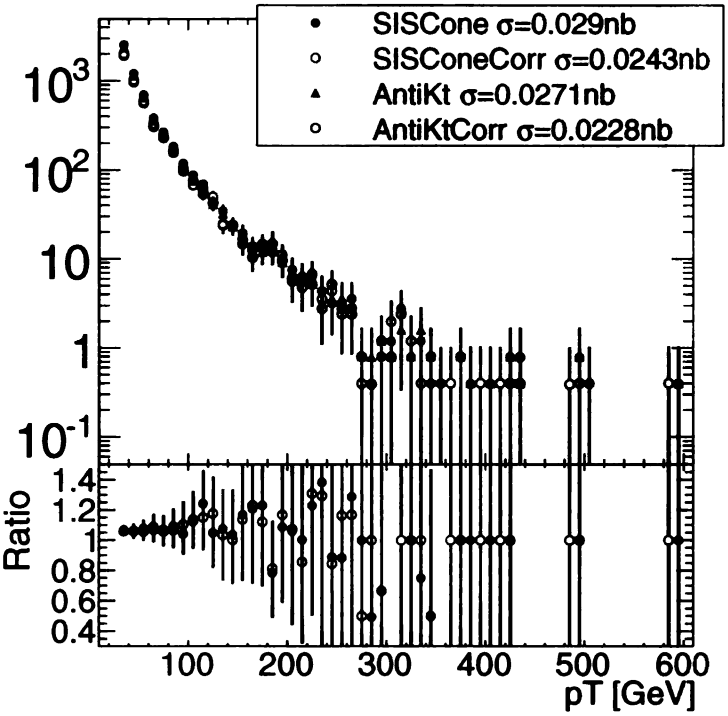
(b) $W+ \geq 0$, R = 0.7

Figure 6.3: Leading jet p_T distributions for inclusive $W+ \geq 0$ distributions at cluster level.

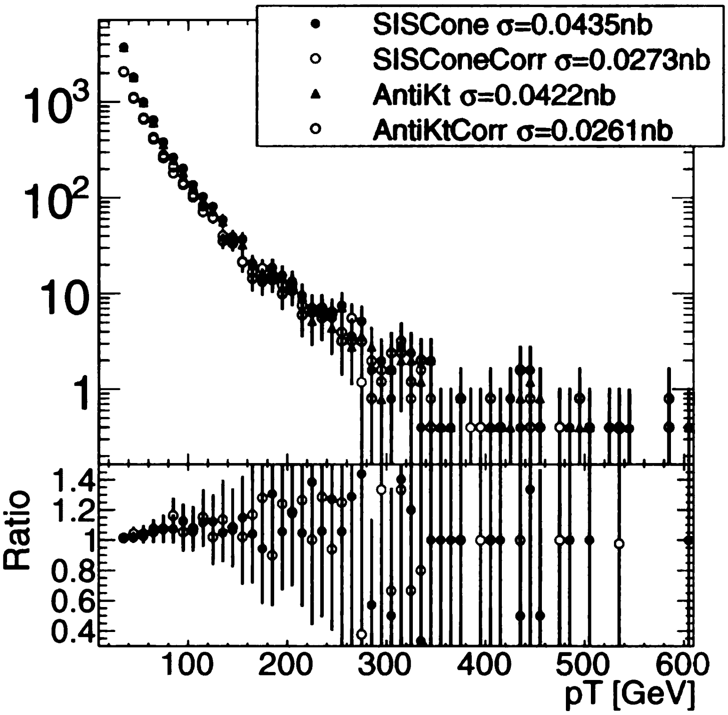


(c) $W+ \geq 0$, R = 1.0

Figure 6.3: [continued] Leading jet p_T distributions for inclusive $W+ \geq 0$ distributions at cluster level.

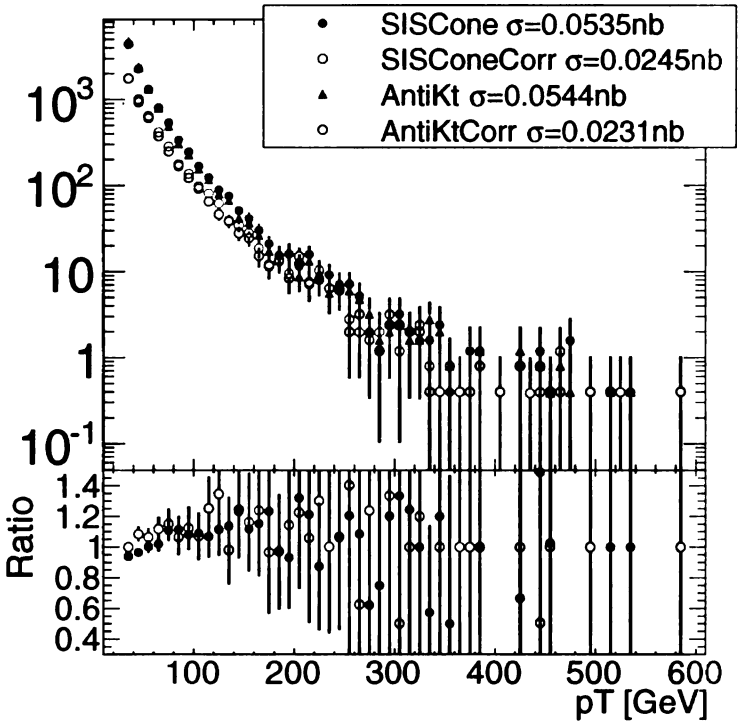


(a) $W+\geq 0$, R = 0.4



(b) $W+\geq 0$, R = 0.7

Figure 6.4: Second leading jet p_T distributions for inclusive $W+\geq 0$ distributions at cluster level.



(c) $W+ \geq 0$, R = 1.0

Figure 6.4: [continued] Second leading jet p_T distributions for inclusive $W+ \geq 0$ distributions at cluster level.

7. Conclusions

This study has investigated the transverse momentum distributions of exclusive and inclusive $W+$ n jets events at parton and cluster levels, using the algorithms k_T, anti-k_T and SISCone. The two sequential recombination algorithms k_T and anti-k_T have been shown to be very similar.

In the comparison between anti-k_T and SISCone, the transverse momentum distributions have displayed different cross sections. In $W + 1$ and $W + 2$ jets events, the SISCone cross section is usually higher, but for higher jet multiplicities, its cross sections are smaller than the anti-k_T's. The same behavior can be observed in inclusive $W+ \geq n$ production. This shows that SISCone is more likely to merge two jets.

In general, the SISCone, anti-k_T, and k_T jet clustering algorithms give similar results, but with differences that can illuminate different aspects of jet physics, and which have to be accounted for when comparing data to theory.

The jet transverse momentum distributions of a given algorithm at parton and cluster levels show similar shapes. The differences in the cross sections at the different levels depend on the jet size R but could not be further quantified in this study.

So far, only jet sizes of 0.4, 0.7, and 1.0 have been considered. A future study should also investigate jet algorithms with radii of 0.5 and 0.6, as it can be expected that the optimal jet size is between 0.4 and 0.7 rather than above 0.7.

BIBLIOGRAPHY

[1] Alpgen version 2.13, http://mlm.home.cern.ch/mlm/alpgen/

[2] http://www.kip.uni-heidelberg.de/atlas/projects/analysis/

[3] L. Evans, P. Bryant (editors), "The LHC Machine", J. Instrum. **3**, S08001 (2008)

[4] E. Etzion *et al.*, "B-Physics and Quarkonia studies with early ATLAS data", arXiv:1001.3806v1 [hep-ex]

[5] The ATLAS Collaboration (G. Aad *et al.*), "The ATLAS Experiment at the CERN Large Hadron Collider", J. Instrum. **3**, S08003 (2008)

[6] The ATLAS Collaboration (G. Aad *et al.*), "The ATLAS Inner Detector commissioning and calibration", ATLAS-IDET-2010-01-004 , arXiv:physics.ins-det/1004.5293v2 (2010)

[7] P.Puzo, Nuclear Instruments and Methods in Physics Research A, **494**, 340 (2002)

[8] S. Prasad, "The ATLAS trigger system", Harvard ATLAS meeting, (2007)

[9] P. Clarke, "The L2 Trigger for the ATLAS detector at the LHC", Nuclear Instruments and Methods in Physics Research A, **368**, 175 (1995)

[10] The ATLAS Computing Workbook, http://twiki.cern.ch/twiki/bin/view/Atlas/WorkBookAtlasExperiment

[11] I.Vivarelli,"Jet Measurements at ATLAS", Springer Proceedings in Physics, **108** Part 2, 63-66 (2006)

[12] D. Lelas *et al.* (ATLAS Collaboration), "Jet reconstruction with first data in ATLAS", ATL-COM-PHYS-2008-246 (2008)

[13] G. Salam, "Towards Jetography", Eur. Phys. J. C **67**, 637-686 (2010), arXiv:0802.118

[14] G. P. Salam, G. Soyez, "A practical Seedless Infrared-Safe Cone jet algorithm", JHEP **0705** 086 arXiv:0704.0292 [hep-ph] (2007)

[15] S. D. Ellis, D. E. Soper, "Successive combination Jet Algorithm For Hadron Collisions", Phys. Rev. D **48**, 3160 (1993)

[16] M. Cacciari, G. P. Salam, G. Soyez, "The anti-k_T jet clustering algorithm", JHEP **0804**, 063 (2008)

[17] C. Buttar *et al.*, "Standard Model Handles and Candles Working Group: Tools and Jets Summary Report", arXiv:0803.0678[hep-ph], (2008)

[18] http://www.pa.msu.edu/~huston/SpartyJet/SpartyJet.html

[19] SpartyJet Documentation, http://projects.hepforge.org/spartyjet/SJ_doc/index.html

[20] G. Corcella *et al.*, "Herwig 6.5", JHEP **0101** (2001) 010, arXiv:hep-ph/0011363 and hep-ph/0210213

[21] Jimmy generator, http://projects.hepforge.org/jimmy/

[22] ATLAS Database, http://twiki.cern.ch/twiki/bin/view/Atlas/DataBases

[23] M. Cacciari, G. P. Salam, "Pileup subtraction using jet areas", Phys. Lett. B **659** (2008) 119, arXiv:0707.1378[hep-ph]

[24] M. Cacciari, G. P. Salam, G. Soyez, "The catchment area of jets", JHEP **04** (2008) 005,

[25] J. F. Arguin *et al.*, "Prospects for W/Z+jets early data measurements with the ATLAS detector at the LHC", ATLAS Note (2010)