



# Studies on the Production of a $W$ Boson in Association with $b$ Jets in $pp$ Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector

Philipp Stolte

University of Göttingen, Germany

DESY Summer Student Program 2011

Supervisor: Takanori Kono

September 8, 2011

## Abstract

This report serves to present the results of my work during my participation in the DESY Summer Student Program in 2011. The aim of the project was to establish a measurement of the associated production of a  $W$  boson with one or two  $b$  jets, respectively, in proton-proton collisions at a center-of-mass energy of 7 TeV using the ATLAS detector at the LHC. All  $b$  jets are identified with the help of reconstructing the secondary vertices. The measurement is performed in a restricted  $W$  boson phase space using the electronic decay channels only. The measurement relies on  $1076 \text{ pb}^{-1}$  of data which was collected with the ATLAS detector in 2011.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Theory</b>	<b>4</b>
2.1	Production of Massive Gauge Bosons in Association With Jets . . . . .	4
2.2	Jets . . . . .	5
2.3	b-Tagging . . . . .	6
<b>3</b>	<b>The ATLAS Detector</b>	<b>7</b>
3.1	The ATLAS Experiment . . . . .	7
<b>4</b>	<b>Further Information about the Measurement</b>	<b>9</b>
4.1	Simulated Event Samples and Data . . . . .	9
4.2	Event Selection . . . . .	9
4.3	b-Tagging Algorithms . . . . .	10
<b>5</b>	<b>Results</b>	<b>11</b>
5.1	Distributions of Various Variables . . . . .	11
5.2	Reconstruction Efficiencies . . . . .	15
5.3	Cross Section Measurements . . . . .	19
<b>6</b>	<b>Conclusion and Outlook</b>	<b>23</b>

# 1 Introduction

The *Large Hadron Collider* (LHC) at CERN is currently the world's most powerful particle accelerator. It is designed to reach beam energies of more than 7 TeV which is sufficient to find the Higgs particle - if it exists - which has not been detected so far as well as to test the *Standard Model of Elementary Particle Physics* (SM) in more detail. Additionally, this new energy region may include something completely unexpected which can be commonly described as physics beyond the SM.

The LHC produces, among other particles, vast amounts of the massive gauge bosons  $W$  and  $Z$ . This report serves to present a measurement of the production cross section, to be more exact: the differential production cross section, of the gauge boson  $W$  in association with jets whereby one or two of these jets are  $b$  jets, in the following referred to as  $W + b$  jet production. The measurement is so important since it provides a test of quantum chromodynamics (QCD) and consequently serves to test the Standard Model further. Theoretical predictions for such a process reveal large uncertainties. A direct measurement of the corresponding cross section is crucial as  $W + b$  jet production constitutes also a large background in searches for the SM Higgs boson  $H$  via  $WH$  production associated with a decay of  $H \rightarrow b\bar{b}$  [1], to measurements of top quark properties via single [2] and via pair production [3] with  $t \rightarrow Wb$  and finally to searches for physics beyond the SM [4], as stated above.

The measurement of the cross section for  $W + b$  jet events in proton-proton ( $pp$ ) collisions at a center-of-mass energy of  $\sqrt{s} = 7$  TeV with the ATLAS detector at the LHC presented in this report is based on events in the electron decay mode. The integrated luminosity is about  $1076 \text{ pb}^{-1}$ . Selected events need to be consistent with the electronic decay of the  $W$  boson and are required to contain at least one jet. Jets originating from  $b$  quarks are identified by exploiting the comparatively long lifetime of  $B$  mesons of about 1.5 ps as well as the large mass of about 5.3 GeV of these hadrons. This identification of  $b$  jets is called  $b$ -tagging.

The cross section which is supposed to be calculated is defined in a restricted region of phase space for the kinematics and the multiplicity of all outgoing particles. This is done in order to simplify on the one hand comparisons with the theoretical predictions and to minimize on the other hand the extrapolation outside the experimentally accessible region. This phase space region is simplified as close as it is possible to the selection criteria. Exact cuts are given in one of the following paragraphs. Events need to contain a  $W$  boson decaying into an electron and the corresponding neutrino as well as at least one jet formed by clustering stable particles as described later on. For the later cross section measurements, at least one of the jets is required to be a  $b$  jet. The latter is defined by the existence of a  $b$  quark within the jet.

The following chapter can be regarded as an introductory review presenting some fundamental and theoretical aspects of the measurements mentioned above and described in detail later on. Special emphasis is placed on  $b$ -tagging. In the subsequent Chapter 3, the ATLAS detector built at the LHC is briefly presented. Some further information

about the measurement of the production cross section of  $W \rightarrow e\nu$  events in association with  $b$  jets is described in the fourth Chapter whereas Chapter 5 contains the results of this measurement, which was performed during the DESY Summer Student Program. Finally, a short summary can be found in Chapter 6 followed by a brief outlook.

## 2 Theory

In the first part of this chapter, theoretical aspects concerning  $W$  bosons and jets are explained in more detail. Then, fundamentals of  $b$ -tagging are outlined since various  $b$ -tagging methods are used for the measurement presented in this report.

### 2.1 Production of Massive Gauge Bosons in Association With Jets

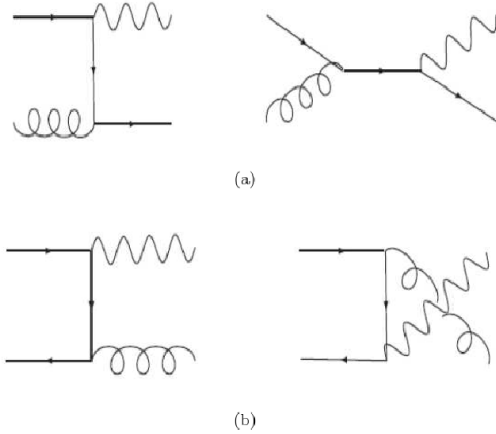


Figure 1: Feynman diagrams: Compton process (a) and  $q\bar{q}$  annihilation (b).

At the LHC, the massive gauge boson  $W$  is produced via Drell-Yan processes modified by higher order QCD corrections.

If these bosons are associated with one jet, two contributions are relevant: the Compton process and  $q\bar{q}$  annihilation. The Compton process is the dominant one at the LHC. In Fig. 1 the corresponding Feynman diagrams can be found.  $W$  bosons are presented by a wavy line, gluons by a spiral line and quarks by a straight one.

There also occur reactions with more than one jet in the final state. In the corresponding production processes of  $W$  bosons two or more quarks and gluons are radiated which then hadronize leading to jets.

The production of  $W$  bosons in association with  $b$  or  $c$  jets is of special interest for the later measurement described in this report.  $b$  jets are produced via gluon splitting, the corresponding Feynman diagram can be found in Fig. 2 on the left.  $c$  jet production is also possible with an  $s$  quark in the initial state, represented by the diagram in Fig. 2 on the right. The measurement of  $W + c$  jets is furthermore useful to constrain the parton distribution function (PDF) of  $s$  quarks. The production of a  $b$  jet according to such a flavor changing process is highly suppressed because a  $c$  quark with a comparatively small PDF is required in the initial state. The suppression can also be explained with the help of the CKM matrix and is, besides, another reason why the  $c$  jet production predominates the  $b$  jet production. The consequences will become obvious in the later description of the analysis.

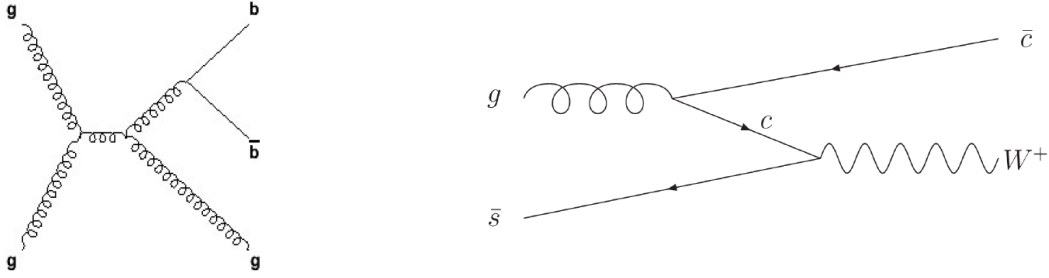


Figure 2: Feynman diagrams:  $b$  jet production via gluon splitting (left) and  $c$  jet production with an  $s$  quark in the initial state (right).

## 2.2 Jets

As shown in the introductory chapter, jets play an essential role for estimations of background processes. But jets also constitute an important signal. These two facts underline the importance of jets.

A jet can be described as a cone of hadrons and other particles produced by hadronization of partons (quarks or gluons). This process of producing jets out of partons is also called fragmentation. There are different processes or levels in the formation of a jet. This can be seen in Fig. 3.

At the LHC, as it is a  $pp$  collider, two protons with specific parton distributions collide and hard scattering processes occur. The partons participating in such hard processes can radiate further partons. The radiated quarks or gluons can radiate again, which leads to showers of partons. Since these partons cannot be free in the final states due to confinement, quarks and gluons are combined to colorless hadrons. This process is the so-called hadronization or fragmentation as described above. These hadrons then decay because of their limited lifetime.

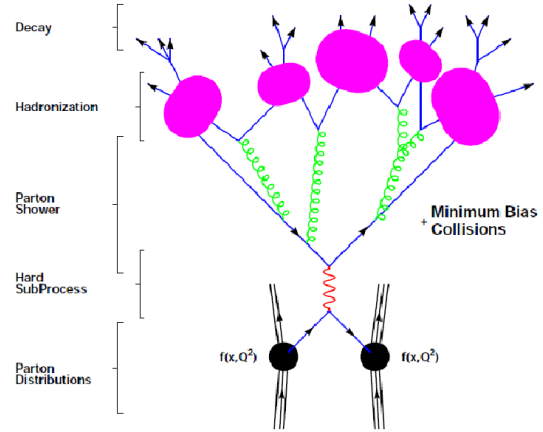


Figure 3: Schematic view of the basic structure of an event leading to a jet in a detector.

## 2.3 b-Tagging

Jets originating from bottom quarks possess unique properties that allow for a differentiation between these  $b$  jets and those coming from the hadronization of lighter quarks. Jets from  $u$ ,  $d$  and  $s$  quarks hadronize directly at the *primary vertex*, the original interaction point.  $b$  jets contain  $B$  mesons, which may have a decay vertex displaced from the primary one due to its long lifetime of about 1.5 ps, leading to a measurable flight length path of up to a few millimeters on average. That is why the decay of  $B$  hadrons can then take place at a *secondary vertex* [5].

The differentiation between  $b$  jets and light jets, the identification of  $b$  jets out of other jets, which is the so-called  $b$ -tagging, is typically based on the existence of this secondary vertex. A bottom jet can also be identified by measuring the *impact parameters* of tracks belonging to the decay. The impact parameter (IP) is defined as the distance from the closest approach of the track to the primary vertex. The IP can have different signs, which can be seen in Fig. 4. Because of the existence of a secondary vertex,  $b$  jets tend to have a positive impact parameter rather than a negative one which allows for an easier tagging of  $b$  jets [5]. Finding a lepton near a jet is also another indication for a  $b$  jet since the branching ratio of semileptonic  $B$  hadron decays is about 11% and hence considerably larger than the one of light hadrons.

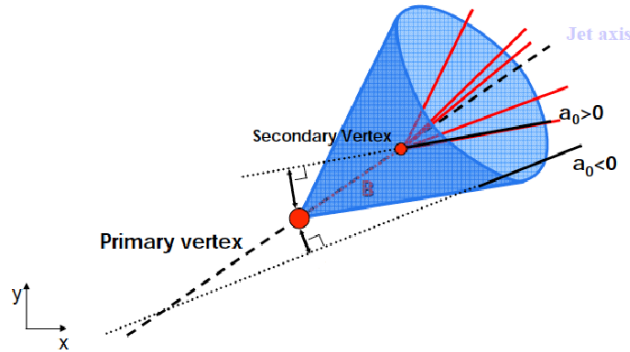


Figure 4: Illustration of the secondary vertex, displaced from the primary one, and of the sign of the impact parameter  $a_0$  in the transverse projection. “ $B$ ” represents a  $B$  meson.

The algorithms with which  $b$  jets are identified are called *taggers*. The ones used in the analysis presented here are described in Chapter 4.

### 3 The ATLAS Detector

ATLAS is next to CMS, ALICE and LHCb one of four detector experiments, excluding two small ones, which are located at the LHC at CERN in Geneva, which started operating at the end of 2009. This chapter deals with a short description of the ATLAS detector.

#### 3.1 The ATLAS Experiment

ATLAS (A Torodial LHC ApparatuS) is composed of several subdetectors. The four main parts are the inner detector, the calorimeter system, the muon spectrometer and the magnet system. An overview of the whole detector and its subsystems is given in Fig. 5. The detector is about 44 m long and 25 m in diameter at a weight of  $\approx 7000$  t [6].

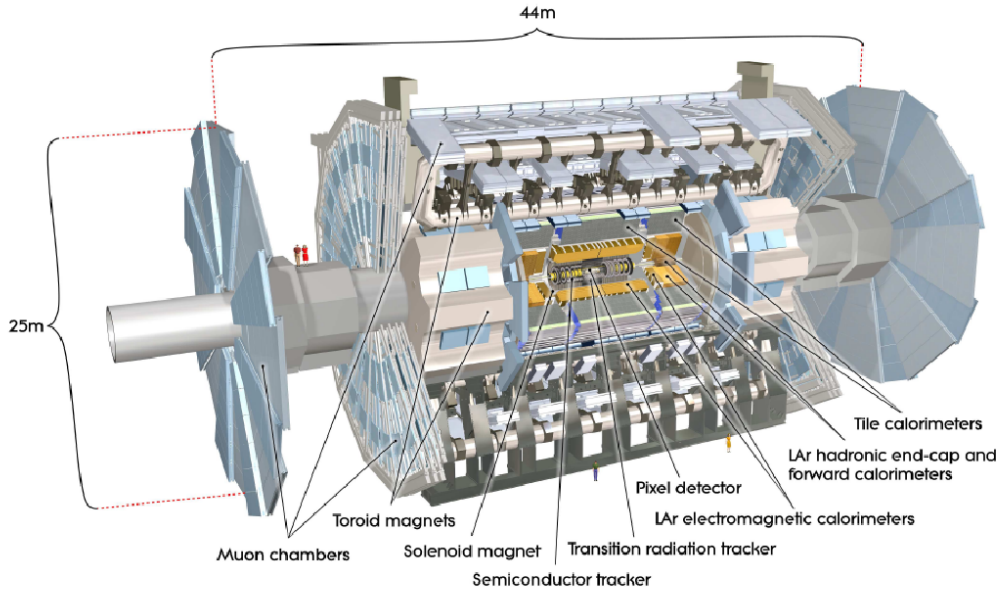


Figure 5: Cut-away view of the ATLAS detector and its subsystems [6].

Before the different parts of which the ATLAS detector consists are outlined in more detail, some frequently used detector variables are briefly introduced. *Spherical coordinates*  $(r, \theta, \phi)$  can be used to characterize positions inside the detector. However, the *pseudorapidity*  $\eta$  instead of  $\theta$  is commonly used as the third coordinate. It depends on  $\theta$  according to its definition:  $\eta = -\ln \tan \frac{\theta}{2}$ . Differences in  $\eta$  are invariant under Lorentz boosts. Usually, distances  $\Delta R$  are defined in the  $\eta$ - $\phi$ -plane according to:  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ .

The *inner detector* as the first ATLAS subdetector - beginning a few centimeters from the beam axis - is composed of the *Pixel Detector*, the *Semi-Conductor Tracker* (SCT) and the *Transition Radiation Tracker* (TRT). Because of over 80 million readout channels in the Pixel Detector, a spatial resolution up to the order of a  $\mu\text{m}$  is provided. Similar in concept and function is the SCT. Both the Pixel Detector and the SCT cover a range of  $|\eta| < 2.5$  and reach a high granularity. The TRT straw tubes are used to measure a particle's transition radiation. This allows for a differentiation between several particles, e.g. electrons and pions [6].

The ATLAS experiment uses several kinds of *calorimeters*, which surround the inner detector. There are two calorimeter systems covering a large  $|\eta|$ -range with  $|\eta| < 4.9$ : an inner *electromagnetic calorimeter system* and an outer *hadronic calorimeter system*. Both are *sampling calorimeters* which implies that passive absorber material and active material alternate. The calorimeter system absorbs the energy of particles in the detector so that the energy deposit of these particles can be measured. In most ATLAS calorimeters liquid argon serves as the active medium [6].

The energy resolution of a calorimeter is usually given as:  $\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b$  with a stochastic term  $a$  and a constant term  $b$ . In this case one obtains for the energy resolution of the electromagnetic calorimeter:  $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$  and for the one of the hadronic calorimeter:  $\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$  [7]. Furthermore, the electromagnetic calorimeter uses an accordion geometry which ensures a full coverage. This applies to electrons and photons. Most energy of the shower is collected in the second layer of the electromagnetic calorimeter (at high energy), which has a cell granularity of  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  implying an adequate spatial resolution. The majority of the EM shower energy is deposited in 16 radiation lengths  $X_0$  of this second layer which means that clusters with an energy below 50 GeV are fully contained. Also jet energies in the hadronic calorimeter are fully contained [6].

The *muon spectrometer* forms the outermost layer of the ATLAS detector and detects muons leaving the calorimeters as they are minimum ionizing particles, measures its momentum and triggers on them. The spectrometer is composed of four different sorts of muon chambers [6].

In order to bend charged particles and to measure their momenta, two large superconducting *magnet systems* are employed: The inner detector is surrounded by a *solenoid* which produces a magnetic field of about 2 T. Outside the calorimeters but within the muon system, an outer *toroidal* magnetic field is located [6].

Since each bunch crossing generates extremely large amounts of data and since most detected events are expected to be less interesting QCD scattering events, a *trigger system*, in this case a three level one, is used, trying to identify - in real time - the worthwhile signal events to retain for detailed studies and trying to reject background events [6].



## 4 Further Information about the Measurement

In this chapter, some further information concerning the measurement presented in this report are given. At first, the simulated event samples as well as the data samples are described in more detail. The following subchapter deals with the event selection while a further paragraph depicts the  $b$ -tagging algorithms used for this analysis. Only the most important information are given in this chapter - a detailed description would exceed the scope of this report by far.

### 4.1 Simulated Event Samples and Data

In order to model the  $W + b$  jet signal next to the various backgrounds Monte Carlo (MC) event samples with the full ATLAS detector simulation - which are furthermore corrected for all known detector effects - are used. On the basis of these MC samples cross sections can be obtained. The  $W + \text{jet}$  processes are simulated separately with  $W + bb + N$  jets,  $W + cc + N$  jets,  $W + c + N$  jets and  $W + N$  jets using the ALPGEN generator [8] which is interfaced to HERWIG [9] for parton shower and fragmentation and JIMMY [10] for the underlying event simulation.

The different backgrounds which are used in this analysis are the following:  $t\bar{t}$ , single top ( $t$ ),  $Z + \text{jets}$ ,  $W \rightarrow \tau + \text{jets}$ ,  $WW$ ,  $ZZ$  and  $WZ$ . It is worth mentioning that no further QCD multi-jet background is included in the MC samples. The backgrounds are normalized according to the integrated luminosity  $\int \mathcal{L} dt$ .

The analysis presented here is based on data which was taken in 2011 with a corresponding integrated luminosity of about  $\int \mathcal{L} dt = 1076 \text{ pb}^{-1}$ . All data on which the measurement is based are collected using ATLAS electron triggers. The currently used trigger is: EF\_e20\_medium.

### 4.2 Event Selection

In Tab. 1 several cuts are listed which are applied in order to define the phase space for the cross section calculations.

Requirement	Cut
Electron transverse momentum	$p_T^e > 25 \text{ GeV}$
Electron pseudorapidity	$ \eta^e  < 2.5$
Missing transverse energy	$\cancel{E}_T > 25 \text{ GeV}$
$W$ transverse mass	$M_T^W > 40 \text{ GeV}$
Jet transverse momentum	$p_T^j > 25 \text{ GeV}$
Jet pseudorapidity	$ \eta^j  < 4.9$
Jet isolation	$\Delta R(e, j) > 0.5$

Table 1: Chosen cuts for the cross section measurement.  $j$  refers to any jet.

Adequate final states, which are selected for the later cross section measurement, must contain exactly one electron with associated missing transverse energy consistent with originating from a  $W$  boson decay.

Electrons are required to pass  $E_T^e > 25$  GeV in the pseudorapidity region  $|\eta^e| < 2.5$ . Electrons in the region between the barrel and the end-cap calorimeters are rejected. A tight identification of electrons is required.

Jets are reconstructed by using the anti- $k_t$  algorithm [11]. For this reconstruction, a radius parameter of  $R = 0.4$  is chosen. Jets are required to have  $p_T^j > 25$  GeV and a pseudorapidity of  $|\eta^j| < 4.9$ . Additionally, all jets within  $\Delta R < 0.5$  of a selected electron are vetoed.

The missing transverse energy  $\cancel{E}_T$  is determined as the vector sum of all calorimeter energy deposits in the transverse plane. The transverse  $W$  mass is also an important parameter for the event selection. It is calculated from the measured electron momenta and the just defined  $\cancel{E}_T$  which is identified with the neutrino energy so that the transverse  $W$  mass can be determined according to the equation:  $M_T^W = \sqrt{2p_T^e p_T^\nu (1 - \cos(\phi^e - \phi^\nu))}$ , while  $\nu$  denotes the neutrino. The following cuts are chosen:  $\cancel{E}_T > 25$  GeV and  $M_T > 40$  GeV, as listed in Tab. 1.

### 4.3 b-Tagging Algorithms

In this analysis, two different  $b$ -tagging algorithms are used: the SV0 algorithm and the so-called JetProb algorithm.

The SV0-tagger is a comparatively simple one. It tags a jet as a bottom jet if the  $B$  hadron's flight distance significance, depending on the secondary vertex position, is larger than a certain cut value  $w_{\text{cut}}$ . The weight cut which is chosen is  $w_{\text{cut}} = 5.85$  so that a jet is tagged as a  $b$  jet if the weight  $w$  of the corresponding jet exceeds 5.85. A purity of 0.94 and a  $b$  jet tagging efficiency of 50% belong to this cut value [12].

The JetProb algorithm, which is also based on impact parameter information, identifies a jet as a bottom jet if  $-\log_{10}(w) > 3.25$  with the jet weight  $w$ . The purity is about 0.92 and the  $b$  jet tagging efficiency amounts to 50% as well.

Due to the fact that not all  $b$  jets are properly tagged as bottom jets, the  $b$ -tagging efficiency needs to be considered for the later cross section measurement. Furthermore, not all tagged jets in the selected data sample are in fact  $b$  jets. Charmed hadrons may have an appreciable lifetime sometimes also resulting in reconstructed displaced secondary vertices. Finally, also light-flavor jets can be misidentified as  $b$  jets because of hadronic interactions, photon conversions in the detector material and finally due to poorly measured track parameters.

The MC samples also contain the truth information. This allows for a check whether a tagged jet is really a  $b$  jet - as long as the MC event samples are used for the reconstruction. If a truth  $b$  quark can be found within  $\Delta R < 0.2$  of a selected jet, the jet is regarded as a truth  $b$  jet.

## 5 Results

This chapter deals with the presentation of the most important results of the analysis performed in the course of the DESY Summer Student Program and described in this report. In the first subchapter, different distributions of important variables of jets or other particles occurring in the analyzed decay channel are discussed. After the calculation of reconstruction efficiencies in the following subchapter the measurement of the  $W + b$  jet differential production cross section is described.

### 5.1 Distributions of Various Variables

Before the calculation of differential cross sections is feasible, certain preliminary steps are necessary. At first, it is useful to look at different distributions of various important variables of jets and, for example, electrons as these particles play an essential role in the analyzed decay channel. Interesting variables are  $\Delta R$ ,  $\Delta\phi$ , or  $\Delta\eta$  which can be measured between two jets participating in the event or between e.g. a jet and an electron. Furthermore, distributions of the rapidity  $y$  or of  $H_T$ , which is defined as the sum of the transverse momentum of all particles in one event, can be plotted for the later analysis. However, a detailed study is done on the distributions of the jet transverse momentum  $p_T$  - which represents the component of the jet momentum in the transverse plane - and the invariant mass  $M_{\text{inv}}$  of jets - which equals the mass in the rest frame:  $M_{\text{inv}} = E^2 - p^2$ . It is necessary to focus on these two variables in order to not exceed the scope of this report. During the Summer Student Program several hundred distributions for this measurement were plotted in total, which, of course, cannot be shown in such a short draft.

In Fig. 6 the jet multiplicity distribution of all events included in the data set can be found. Next to the measured data also the events originating from MC samples are plotted in this histogram. There are three different kinds of signal events. Events where a  $W$  boson is associated only by light jets (labeled as  $W + j$ ), events where a  $c$  jet is measured and finally events where a  $b$  jet is measured. The several visible backgrounds were already mentioned in the last chapter. Thus, taking a look at such distributions allows for a better understanding of backgrounds and the number of signal events, which is necessary for the later cross section calculations, can be estimated.

The plot illustrates that the number of events with a  $b$  jet is comparatively small. A lot of events with exactly one or two jets are included in the data sample which is used later on.

In Fig. 7 distributions of the transverse momentum of both the first and the second jet are presented. The name is chosen according to the transverse momentum value. That means that “first” refers to the jet which has the highest transverse momentum. Additionally, the distribution of the invariant mass can be seen in Fig. 7 These plots also reveal that the fraction of events with  $b$  or  $c$  jets is comparatively small while a large amount of background events are recorded.

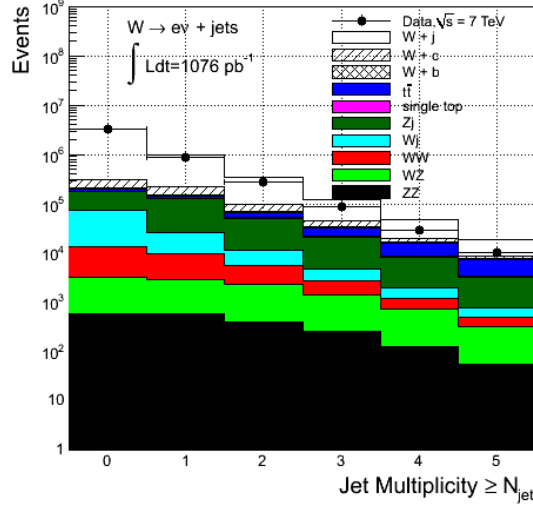


Figure 6: Jet multiplicity distribution including data and MC events.

***b*-Tagging** As mentioned above, the analysis presented here concentrates on events where at least one of the jets is a *b* jet. To separate *b* jets from all other jets *b*-tagging algorithms need to be used. In Chapter 4, two taggers were introduced: the JetProb tagger and the SV0 tagger. In this report, due to the limited scope, it is put emphasis on the Jetprob tagger. In Fig. 8 distributions of the transverse momentum of the leading jet can be found in the case that one or two *b* jets are tagged. The distributions are shown for both taggers.

These plots illustrate that the fraction of events in which *b* jets participate is still quite small whereas a lot of different background fractions are comparatively large. Since these distributions are used for the cross section measurement, it is useful to reduce the fraction of background events. Especially top antitop quark pair ( $t\bar{t}$ ) events and events with at least a single top quark constitute the largest backgrounds. That's why in particular these two ones need to be reduced.

**Lepton Veto and Varying Numbers of Jets in the Event** To remove background events in which top quarks and, most of all, top antitop quark pairs participate the signature of such decays has to be considered. The top decays almost exclusively into a *W* boson and a *b* quark:  $t \rightarrow Wb$ . The *W* boson then decays hadronically into two jets or leptonically into a lepton  $\ell$  and the corresponding antineutrino. Hence, the following decays need to be taken into consideration:  $t \rightarrow b + 2j$  and  $t \rightarrow b + \ell\nu$  while *j* denotes any jet except for *b* jets and  $\nu$  the corresponding neutrino. The final states of the  $t\bar{t}$  decay with at least one lepton are consequently:

$$t\bar{t} \rightarrow 2j + 2b + \ell\nu \quad \text{or} \quad t\bar{t} \rightarrow 2b + 2\ell + 2\nu. \quad (1)$$

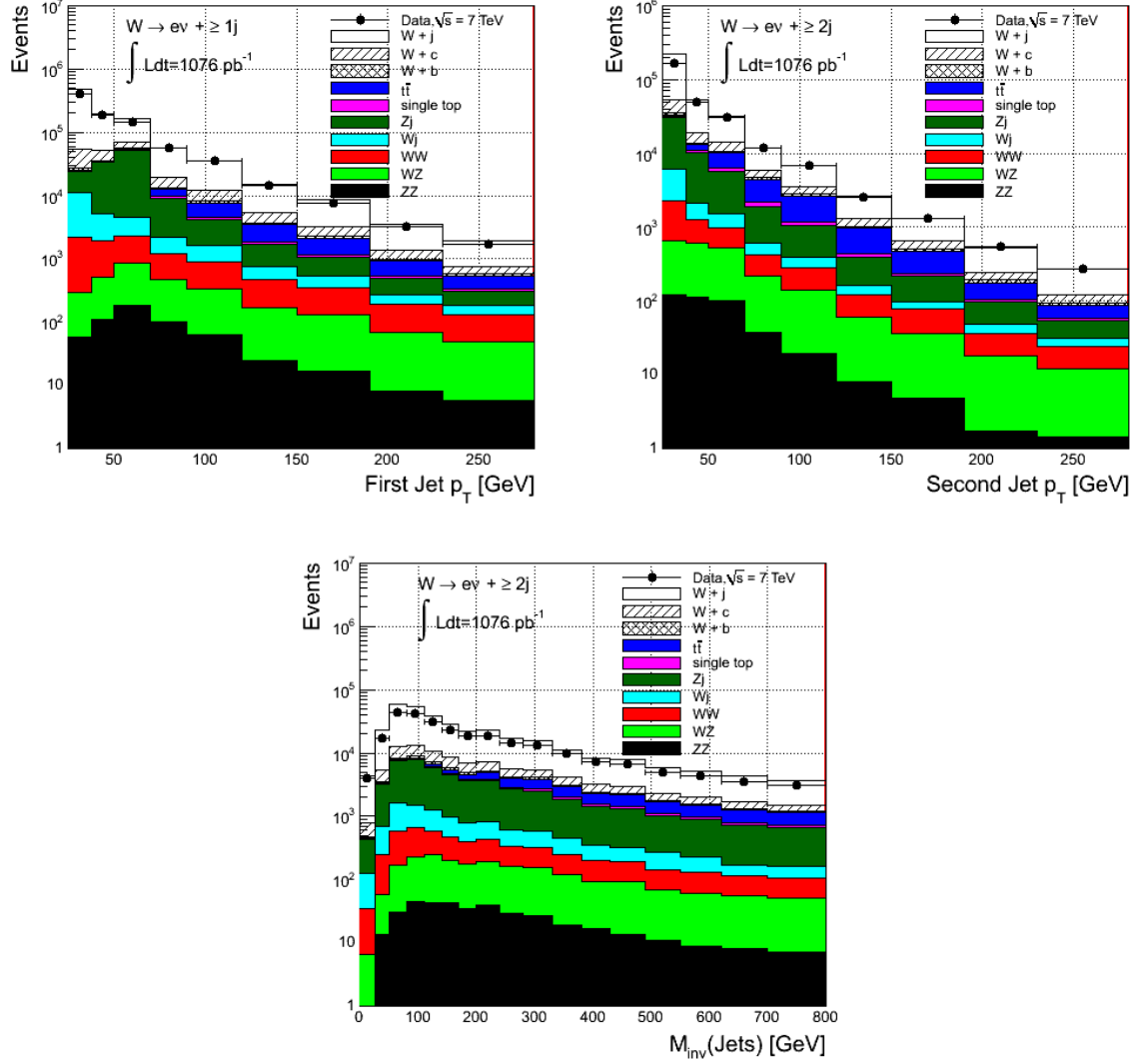


Figure 7: The first two plots above show the distribution in  $p_T$  of the first (left) and the second jet (right) in the case that events with  $\geq 1$  and  $\geq 2$  jets, respectively, are considered. The plot below shows the distribution in the invariant mass of jets in the case that events with  $\geq 2$  jets are considered.

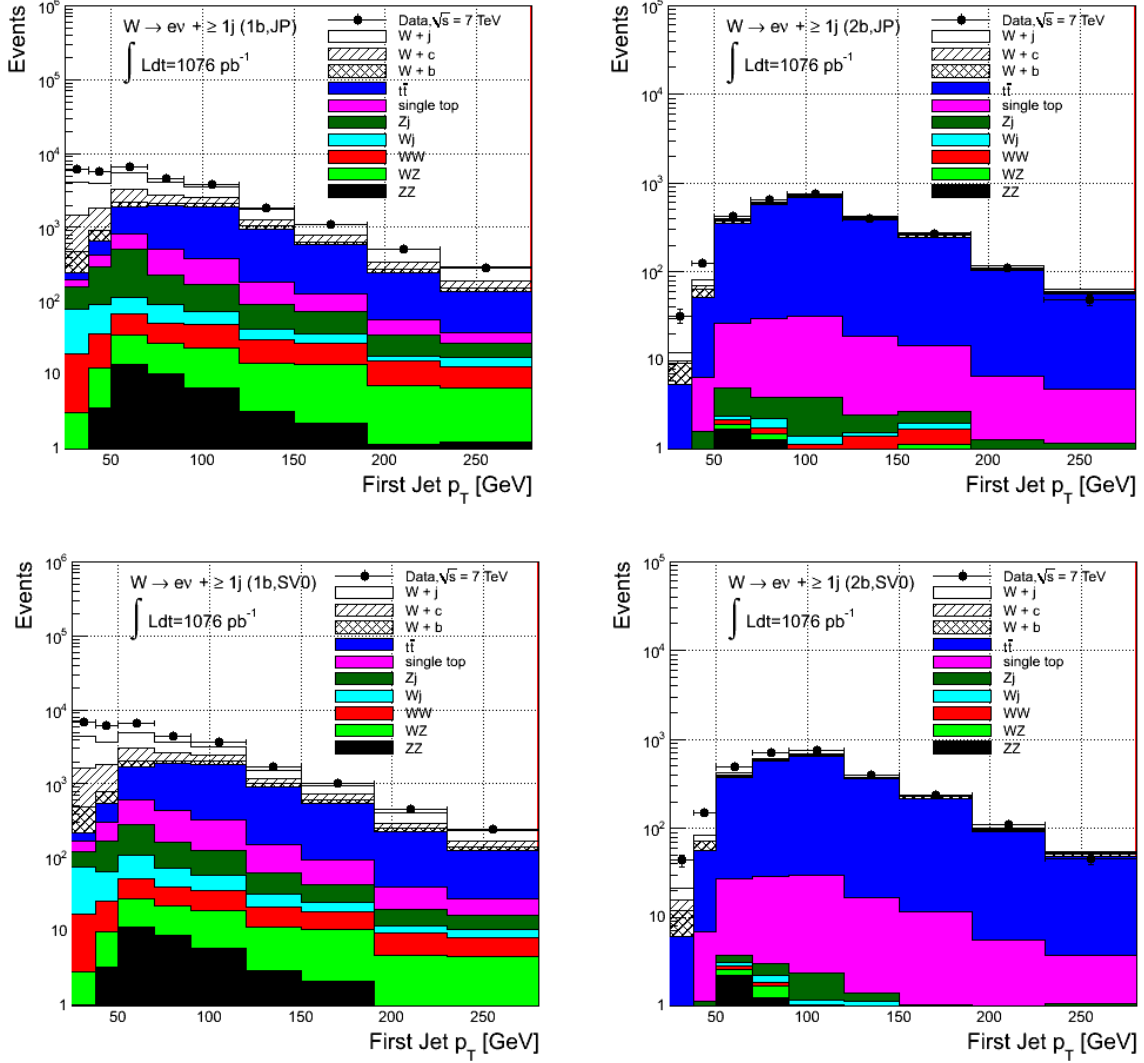


Figure 8: Distributions in  $p_T$  of the first jet in the case that  $b$ -tagging algorithms are used for events with  $\geq 1$  jets. The plots in the upper row belong to the JetProb algorithm, the ones in the lower row to the SV0 algorithm. In each row the distributions with one or two tagged  $b$  jets are presented.

The all-jets decay channel of  $t\bar{t}$  decays is not considered since only the electronic decay channels are used for the  $W + \text{jets}$  analysis presented here, as explained above, so that at least one lepton must participate in the  $t\bar{t}$  event. Eq. (1) reveals that the  $t\bar{t}$  final states can either be characterized by four jets or by two electrons, respectively. In order to increase the  $W + b$  jet fraction and in order to remove  $t\bar{t}$  background the following can thus be done: Only events with exactly one electron and no muon in the final state are considered, which is referred to as *lepton veto* and serves to remove dileptonic  $t\bar{t}$  decays. Furthermore, the number of jets  $N_{\text{jet}}$  in the event can be varied. So  $N_{\text{jet}} = 1$ ,  $N_{\text{jet}} = 2$  or  $N_{\text{jet}} = 3$  can be chosen to remove  $t\bar{t}$  events having a higher jet multiplicity with four or more jets in the final state.

In Fig. 9 different distributions which were plotted whilst taking the lepton veto into account can be seen. Additionally, the number of jets in the event was varied. To compare these new distributions with the older ones, plotted before adding the lepton veto and fixing the number of jets to a certain small value, some of these original, “old” distributions can also be found in Fig. 9. It becomes obvious that choosing a certain fixed number of jets - in the plots presented in the figure  $N_{\text{jet}} = 1$  and  $N_{\text{jet}} = 2$  is chosen, but the same applies to  $N_{\text{jet}} = 3$  - helps to reduce the fraction of  $t\bar{t}$  and some other backgrounds while the fraction of  $W + b$  jet events increases. Adding also the lepton veto improves the result further. However, a significant fraction of  $W + c$  jet events is visible in the distributions as well.

This fact gives rise to analyze in more detail to which extent a differentiation between  $b$  and  $c$  jets is possible. Therefore, a measurement of the secondary vertex mass of all jets which were tagged as  $b$  jets is performed. In Fig. 10 the results of this calculation can be seen in the case that events with exactly one jet (which is tagged as a  $b$  jet) and with exactly two jets (both jets tagged as  $b$  jets) are chosen. The JetProb tagger is used for this analysis. It is expected that  $b$  jets tend to have larger secondary vertex masses than  $c$  jets. This is also illustrated by the distributions in Fig. 10 but the effect is quite small so that  $W + c$  jet and  $W + b$  jet events can only be disentangled to a certain extent.

## 5.2 Reconstruction Efficiencies

This subchapter deals with the calculation of the reconstruction efficiency. This efficiency needs to be taken into consideration since not all jets are reconstructed properly by the anti- $k_t$  algorithm. Besides, not all electrons are identified. In the measurement presented in this report the reconstruction efficiency can be gained with the help of the MC samples as these samples contain also the truth information. In this way, it can be checked whether a reconstructed jet included in one of the MC samples is in fact a jet, called a truth jet. The reconstruction efficiency is defined as the number of all reconstructed jets divided by the number of all truth jets and is used in the next subchapter to perform the cross section measurement.

In Fig. 11 the efficiency distributions are plotted in the case that one jet is contained in the event. The transverse momentum of the first jet is the considered variable. The

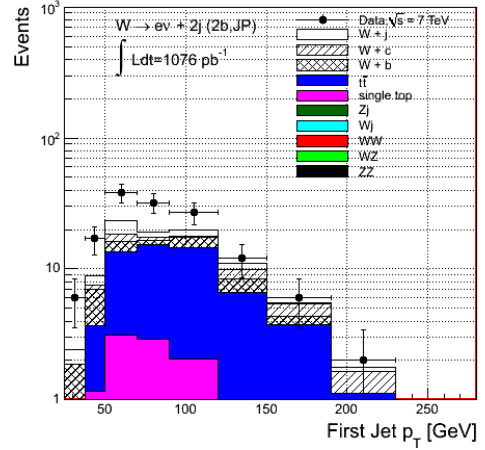
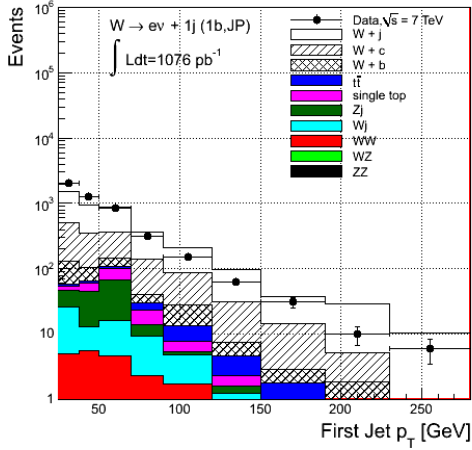
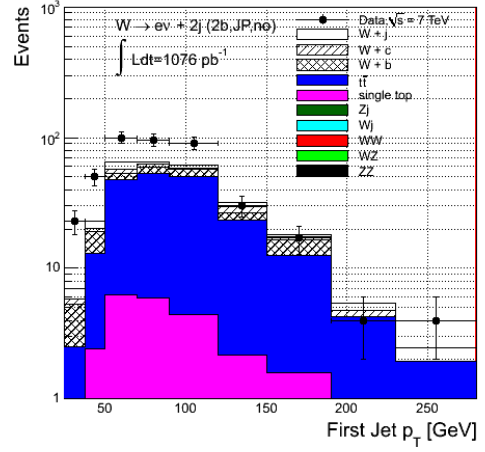
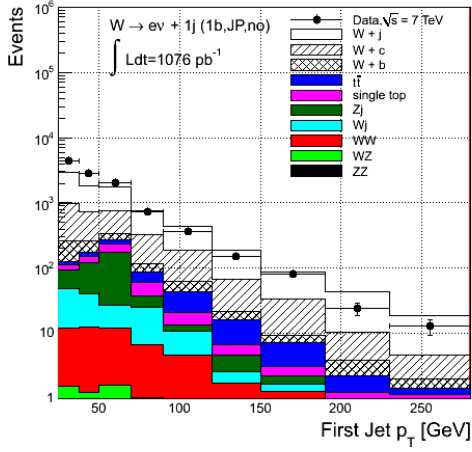
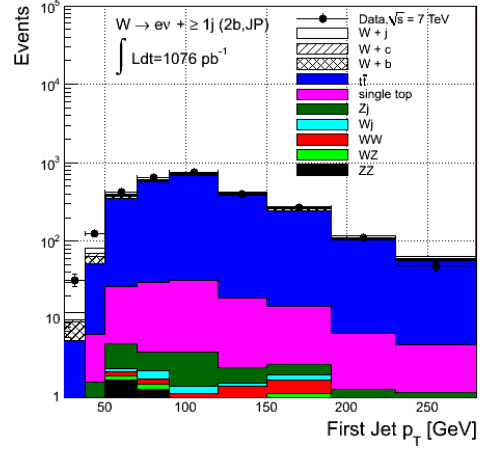
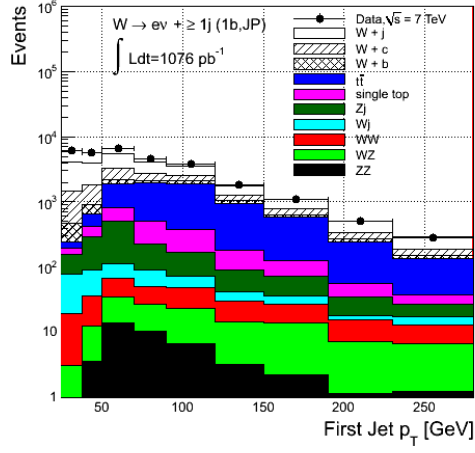


Figure 9: Distributions in  $p_T$  of the first jet. First row: Original plots with one or two  $b$ -tags (see Fig. 7) and events with  $\geq 1$  jet; second row:  $N_{\text{jet}} = 1$  (left) and  $N_{\text{jet}} = 2$  (right) chosen; last row: Lepton veto added in the case that  $N_{\text{jet}} = 1$  and  $N_{\text{jet}} = 2$ .



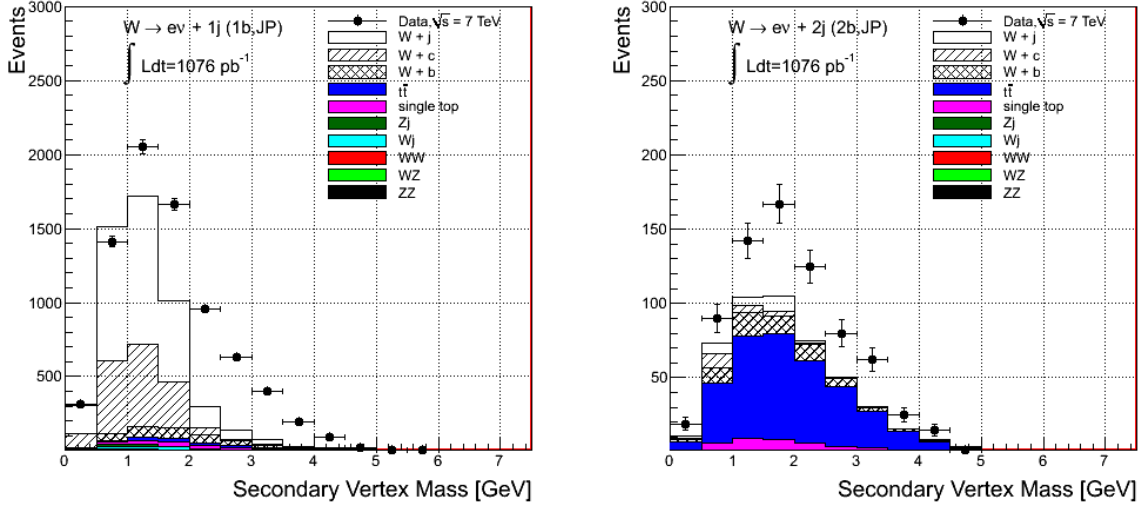


Figure 10: Secondary vertex mass distributions in the case that the JetProb tagger is used. Distributions of events with one jet (which is tagged as a  $b$  jet) and of events with two jets (which are both tagged as  $b$  jets) are plotted.

histograms at the top present the numbers of truth jets and numbers of reconstructed jets, the plots at the bottom show the resulting efficiency distributions. In the case that the jet in the event is tagged as a  $b$  jet - which is displayed in the plots on the right - the efficiency exceeds one although values smaller than one are expected. This is mainly due to the existence of fake jets; sometimes e.g. electrons are tagged as jets. At this point, the algorithms need to be improved to tag only or at least in most cases jets which are in fact jets so that the efficiency decreases to values smaller than one. This is discussed further in Chapter 6.

The efficiency as it is presented here can be calculated for all other distributions in the transverse momentum and the invariant mass and of course for all other ones mentioned at the beginning of this chapter. Due to lack of space these plots are not shown here, but the ones in Fig. 11 are some kind of representative.

As  $b$ -tagging plays a very important role in the measurement presented here, also the  $b$ -tagging efficiency can be discussed in more detail. The  $b$ -tagging efficiency is defined as the number of jets out of all reconstructed  $b$  jets which are tagged correctly as  $b$  jets. In Fig. 12 the  $b$ -tagging efficiency is visualized in distributions similar to the ones presented in Fig. 11. Again the transverse momentum of the first jet is used as variable. But in this case, no limit is set on the number of jets in the event. The calculations are done for the Jetprob - in Fig. 12 on the left - and the SV0 algorithm - on the right. In the upper row distributions of truth  $b$  jets, the distributions of reconstructed jets out of these truth jets as well as the distributions of jets out of these truth  $b$  jets which are even tagged as  $b$  jets are plotted together for each algorithm. In the bottom row the

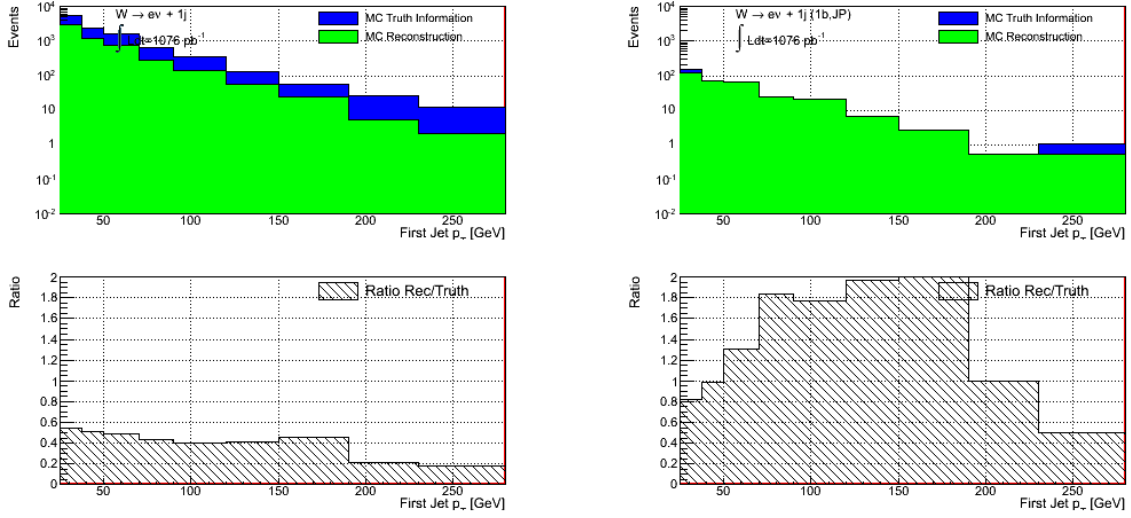


Figure 11: Distributions in  $p_T$  of the first jet in the case that one jet is included in the event - the number of truth jets and reconstructed jets are plotted separately - (upper row) and corresponding reconstruction efficiency distributions (bottom row). The plots on the right belong to events with one jet which is tagged as a  $b$  jet.

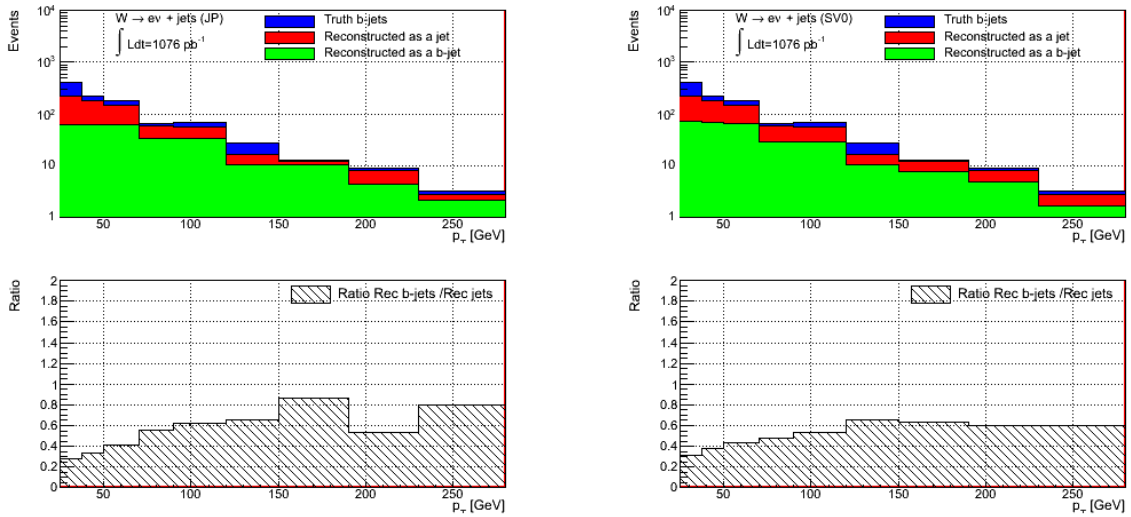


Figure 12: Distributions in  $p_T$  of the first jet - number of truth  $b$  jets, the number of reconstructed jets out of these truth jets and the number of jets out of these truth  $b$  jets which are tagged as  $b$  jets are plotted separately - (upper row) and corresponding  $b$ -tagging efficiency distributions (bottom row). The distributions are given for the JetProb (left) and the SV0 algorithm (right).

resulting reconstruction efficiencies can be found.

The  $b$ -tagging efficiencies reveal that the JetProb algorithm seems to work better. The efficiencies are slightly higher than the ones of the SV0 tagger. That is why it is put emphasis on the results obtained with the JetProb tagger in this report although the cross section measurements were performed for both taggers and although the difference in the efficiencies is not very remarkable. By the way, the  $b$ -tagging efficiency is in the expected range with values smaller than one.

### 5.3 Cross Section Measurements

After the various reconstruction efficiencies have been calculated, all information necessary for an adequate measurement of cross sections and differential cross section are available. In general, cross sections  $\sigma$  can be calculated according to:

$$\sigma = \frac{N - B}{\varepsilon \int \mathcal{L} dt} = \frac{S}{\varepsilon \int \mathcal{L} dt}.$$

In the measurement presented here, differential cross sections are of special interest. These can be calculated according to the following equation - similar to the one for  $\sigma$  - with respect to  $p_T$ :

$$\frac{d\sigma}{dp_T} = \frac{k(N - B)}{w\varepsilon \int \mathcal{L} dt} = \frac{kS}{w\varepsilon \int \mathcal{L} dt}. \quad (2)$$

$N$  denotes the number of measured events (in data),  $B$  the number of background events, which is - in this analysis - estimated from the available Monte Carlo samples.  $S = N - B$  is consequently the number of signal events. The efficiency  $\varepsilon$  and the integrated luminosity  $\int \mathcal{L} dt$  have already been introduced in the last chapters. The factor  $w$  is equivalent to the width of the histograms used in the plots while  $k$  denotes an additional correction factor, which is used in order to take account of higher order corrections if the differential cross section is calculated from MC samples, the so-called “truth” cross section (see below). If the cross section based on data is measured, this factor  $k$  is set to one.

The differential cross sections can also be calculated with respect to other variables mentioned above, so, for example, the differential cross section  $d\sigma/dM_{\text{inv}}$  can be determined and examined as well.

In Fig. 13 the results of several differential cross section measurements according to Eq. (2) can be found. The cross section is calculated separately for each histogram bin. Differential cross sections, determined as stated above, based on the number of measured events  $N$  are referred to as “Measured Xsec” in the legend of the histograms. The cross section can also be calculated from the MC samples. Since the MC samples contain the truth information, a “truth” number of signal events can be determined. This allows for

the measurement of differential cross sections based on this truth signal events. These cross sections are also plotted in Fig. 13 and referred to as “Truth Xsec”. To calculate such truth cross sections, the factor  $k = 1.7$  is used in this analysis.

Fig. 13 consists of six different plots, all of them corresponds to distributions in  $p_T$  of the first jet. The aim of the project was to measure differential cross sections of  $W + b$  jet events. However, the plots in the upper and middle row show cross section results in the case that  $W + b$  jet and  $W + c$  jet events are regarded as signal events, referred to as  $W + \text{heavyflavor jets}$ . Reasons are given later on.

Events with one or with two jets with one tagged  $b$  jet in each case are considered in Fig. 13. All plots on the left belong to the JetProb tagger, the ones on the right to the SV0 tagger. In the last row, events with two jets and one tagged  $b$  jets are used for the cross section measurement, in this case only the  $W + b$  jet fraction is used as signal, indicated by the label “ $Wb$ ” in the legend.

All calculations illustrate that there are notable discrepancies between the measured differential cross sections and the ones obtained from the MC samples. However, the shapes of both distributions are very similar to each other.

It is also worth mentioning that the differential cross section could not be calculated for each bin. This is due to the fact that the number of signal events  $S = N - B$  can be negative since the background is estimated from MC data which may lead to  $B > N$ . It can furthermore happen, in particular in the case that  $W + b$  constitutes the only signal, that certain histogram bins do not contain any events - due to limited statistics - so that no efficiency  $\epsilon$  can be calculated. Moreover, that is the reason why most of the plots presented here are based on measurements with  $W + \text{heavyflavor jets}$  as signal. Nevertheless, all plots illustrate - at least to a certain extent - that such a cross section measurement is in general possible.

The error of this cross section measurement - visualized by the error bars in the histograms - originates from the number of events, the uncertainty equals the square root of this number according to the Poisson distribution. No systematic uncertainties are considered in the measurements presented here so that e.g. no error of the integrated luminosity is taken into account. A detailed analysis of possible systematic uncertainties would have exceeded the scope of this project.

In Fig. 14 plots showing the results of some additional cross section measurements based on calculations with the JetProb tagger are presented. Events with exactly two jets are used. A further description can be found in the caption.

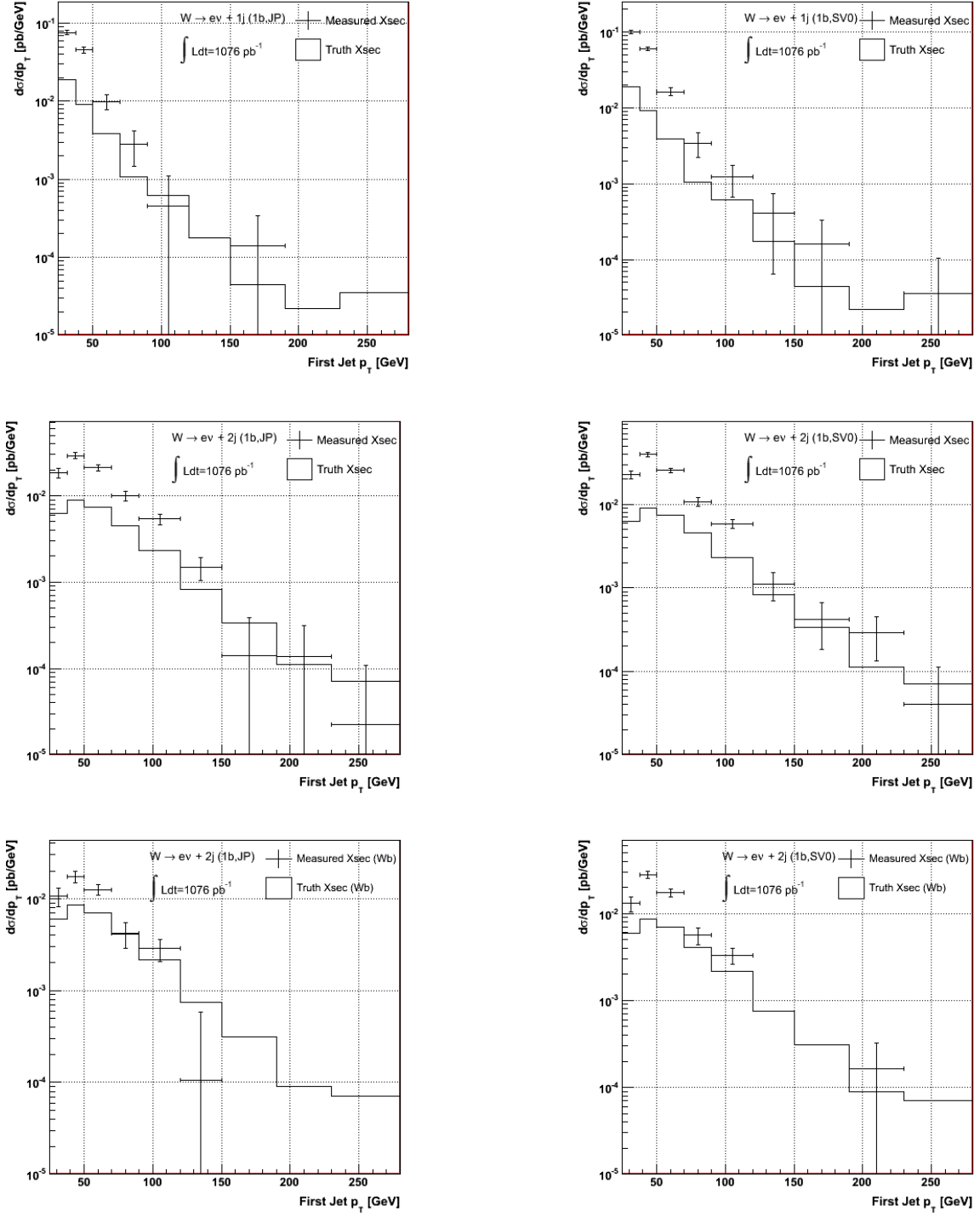


Figure 13: Calculations of  $d\sigma/dp_T$  ( $p_T$  of the first jet). Events with one or two jets and one tagged  $b$  jet are considered - JetProb tagger on the left, SV0 on the right. Upper and middle row:  $W$  + heavyflavor jets as signal; bottom row:  $W$  +  $b$  jets as signal.

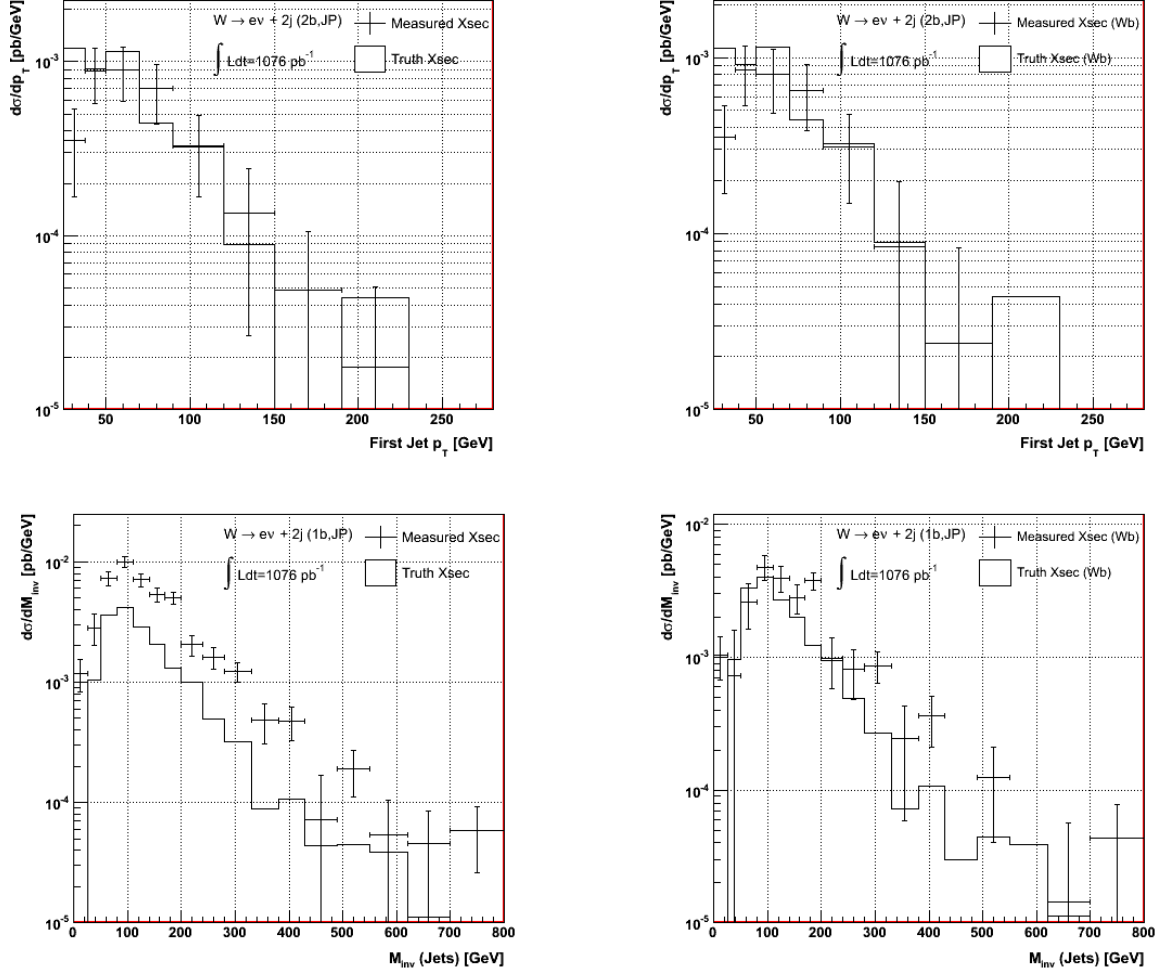


Figure 14: Differential cross section measurements based on calculations with the Jet-Prob tagger. Two-jet events are used. In the upper row, results for  $d\sigma/dp_T$  ( $p_T$  of the first jet) in the case that  $W + \text{heavyflavor}$  and  $W + b$  is chosen as signal are given, all events contain two tagged  $b$  jets. In the lower row, in contrast, results for  $d\sigma/dM_{\text{inv}}$  in the case that  $W + \text{heavyflavor}$  and  $W + b$  is chosen as signal are given, all events contain one tagged  $b$  jet.

## 6 Conclusion and Outlook

**Conclusion** As explained in more detail at the beginning of this report, the aim of the project presented here was to establish a measurement of the associated production of a  $W$  boson with one or two  $b$  jets, respectively, in proton-proton collisions at a center-of-mass energy of 7 TeV using the ATLAS detector at the LHC. Special emphasis was put on the calculation of differential cross sections. It could be shown that the differential cross section measurement of  $W + \text{heavyflavor jet}$  events (including  $b$  and  $c$  jets) as well as the desired differential cross section measurement of  $W + b$  jet events (with  $W + c$  jet events as an additional background) is possible - although in the latter case the number of signal events is comparatively small. The results seem to be adequate, however, deviations between the measured cross sections based on data and the truth cross sections become obvious.

The analysis also illustrates that a successful differential cross section measurement of events including  $b$  jets relies on special  $b$ -tagging algorithms like the SV0 and the JetProb tagger used in this analysis.

**Outlook** Although it could be shown that a differential cross section measured as described above is feasible, certain aspects need to be considered for a more precise measurement.

As stated in the last chapter, the calculation of the differential cross section error does not include any systematic uncertainties. Such uncertainties need to be taken into account in the future. For a more precise measurement of the described differential cross sections, it would also be useful to estimate the number of background events directly from data instead of using the given MC samples for this estimate. When the reconstruction efficiency is determined, it is furthermore necessary to understand the fake  $b$ -tagged jets better, which are probably the reason for the reconstruction efficiency exceeding one in several cases. Additionally, more sophisticated taggers to identify  $b$  jets out of all other jets would be useful. Then, for example, the separation of  $W + b$  jet events and  $W + c$  jet events using the vertex mass distribution could probably be more accurate.

## References

- [1] The ATLAS Collaboration, ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta, *ATL-PHYS-PUB-2009-088*, 2009.
- [2] ATLAS Collaboration, Searches for Single Top-Quark Production with the ATLAS Detector in pp collisions at  $\sqrt{s} = 7$  TeV, *ATLAS-CONF-2011-027*, 2011.
- [3] ATLAS Collaboration, Measurement of the top quark-pair production cross section with ATLAS in pp collisions at  $\sqrt{s} = 7$  TeV, *Eur. Phys. J.*, C71 (2011) 1577, 2011.
- [4] H.-S. Goh and S. Su, Phenomenology of the left-right twin Higgs model, *Phys. Rev.*, D 75 (2007) 075010, 2007.
- [5] M. Lehmacher, b-Tagging Algorithms and their Performance at ATLAS, *arXiv:0809.4896v3 [hep-ex]*, 2008.
- [6] The ATLAS Collaboration, G. Aad et al., The ATLAS Experiment at the CERN Large Hadron Collider, *JINST*, 3:S08003, 2008.
- [7] The ATLAS Collaboration, *ATLAS Detector and Physics Performance, Technical Design Report: Volume I*, CERN, 1999.
- [8] M. L. Langano et al., ALPGEN, a generator for hard multiparton processes in hadronic collisions, *JHEP*, 0307 (2003) 001, 2003.
- [9] G. Corcella et al., HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes), *JHEP*, 0101 (2001) 010, 2001.
- [10] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Multiparton interactions in photoproduction at HERA, *Z. Phys.*, C 72 (1996) 637, 1996.
- [11] M. Cacciari, G. P. Salam and G. Soyez, The anti- $k_t$  jet clustering algorithm, *JHEP*, 0804 (2008) 063.
- [12] ATLAS Collaboration, Calibrating the  $b$ -Tag Efficiency and Mistag Rate of the SV0  $b$ -Tagging Algorithm in  $3 \text{ pb}^{-1}$  of Data with the ATLAS Detector, *ATLAS-CONF-2010-099*, 2010.