



# ATLAS NOTE

## ATLAS-CONF-2013-109

December 26, 2013



### Calibration of the $b$ -tagging efficiency for $c$ jets with the ATLAS detector using events with a $W$ boson produced in association with a single $c$ quark

The ATLAS Collaboration

#### Abstract

A novel method is used to measure the  $b$ -jet tagging efficiency of  $c$  jets of the standard MV1 tagging algorithm in proton-proton collisions at  $\sqrt{s} = 7$  TeV using data collected by the ATLAS experiment at the LHC in the year 2011, corresponding to an integrated luminosity of about  $4.6 \text{ fb}^{-1}$ . The measurement is based on a sample of  $c$  jets produced in association with a  $W$  boson. The  $W$  boson is reconstructed via its decay into an electron and a neutrino, and the  $c$  jet is identified via a soft muon stemming from a semileptonic  $c$ -hadron decay. The charge correlation of the two leptons is exploited to extract a  $c$ -jet sample with high purity. The results are presented in the form of data-to-simulation efficiency scale factors. The total uncertainties range from 5% to 13%, increasing as the background rejection of the  $b$ -tagging algorithm increases.

# 1 Introduction

The identification of jets originating from the fragmentation of  $b$  quarks,  $b$ -tagging, constitutes an important instrument for the physics programme at the LHC. The background suppression by means of  $b$ -tagging is essential in precision measurements regarding top physics, in Higgs boson searches and measurements as well as in the search for new phenomena. In order to identify  $b$  jets and separate them from jets stemming from  $c$  quarks, light quarks or gluons,  $b$ -tagging algorithms exploit several distinct properties of the production and the decay of  $b$  hadrons e.g. their mass of about 5 GeV and their lifetime of about 1.5 ps. Their separation from  $c$  jets is particularly difficult since  $c$  hadrons have similar properties, such as the mass and lifetime, as  $b$  hadrons. Since simulations are not expected to describe all effects that impact the performance of  $b$ -tagging algorithms accurately, the efficiency with which a jet originating from the fragmentation of a  $b$  quark is identified as a  $b$  jet needs to be evaluated on data. It is equally important to determine the probability of mistakenly tagging a jet as a  $b$  jet that originates from a  $c$  quark, the  $c$ -jet tagging efficiency, or from a light quark or gluon, the mistag rate.

This note presents a novel method to determine the  $c$ -jet tagging efficiency, based on data collected by the ATLAS detector during the year 2011 at  $\sqrt{s} = 7$  TeV. The measurement of the  $c$ -jet tagging efficiency and its calibration by means of data-to-simulation scale factors are presented for the MV1  $b$ -tagging algorithm. The most commonly used tagging algorithm in ATLAS analyses is based on a neural network using the output information from the IP3D, SV1 and JetFitterCombNN  $b$ -tagging algorithms, described in Ref. [1]. Certain cuts on the output weights of the  $b$ -tagging algorithms define operating points that correspond to certain  $b$ -tagging efficiencies determined on a sample of simulated  $t\bar{t}$  events. The operating points of the MV1 algorithm calibrated in this note correspond to  $b$ -tagging efficiencies of 85%, 75%, 70% and 60%.

Precisely measuring the tagging efficiency of a  $b$ -tagging algorithm for a certain jet flavour requires a sample of jets extracted from data that is dominated by that jet flavour. The first measurement of the  $c$ -jet tagging efficiency by the ATLAS collaboration [2] is based on a sample of  $D^{*+}$  mesons reconstructed in the exclusive decay  $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ . In the present analysis an alternative approach to select a sample of  $c$  jets is chosen. A single  $c$  jet produced in association with a  $W$  boson is identified by a soft muon stemming from the semileptonic decay of a  $c$  hadron. The  $W$  boson is reconstructed via its decay into an electron and a neutrino. In proton-proton collisions at a centre-of-mass energy of  $\sqrt{s} = 7$  TeV, the dominant production mechanism is  $gs \rightarrow W^-c$  and  $g\bar{s} \rightarrow W^+\bar{c}$ , where the  $W$  boson is always accompanied by a  $c$  quark with an oppositely signed charge. Given that the soft muon carries a charge with the same sign as the  $c$  quark, requiring that the charge of the soft muon and the charge of the  $W$  boson (or its decay electron) to be of opposite sign selects  $W+c$  events with very high efficiency. Most of the background processes are evenly populated with events where the charges of the decay leptons are of opposite sign (OS) or of same sign (SS). Therefore, a very pure sample of  $c$  jets is obtained by extracting the number of  $W+c$  signal events as the difference between the number of events with opposite and with same charge (OS-SS). This fundamental strategy was already exploited in several  $W+c$  production cross section measurements [3–7]. In the remainder of this note jets that are soft-muon tagged (SMT) are referred to as SMT jets and a sample composed of such jets extracted as the number of OS-SS events is referred to as the SMT jet sample.

As a first step the  $c$ -jet tagging efficiency is measured using the SMT  $c$ -jet sample in data and simulation. The results are presented as data-to-simulation scale factors. Following that, an extrapolation procedure is performed to derive data-to-simulation scale factors that are applicable to an unbiased sample of inclusive  $c$  jets.

The present note is organized as follows. Section 2 summarizes the data and Monte Carlo (MC) samples used in the analysis. In Sec. 3 the object and event selection is detailed. The extraction of the  $W+c$  yield is discussed in Sec. 4. In Sec. 5 the measurement of the  $c$ -jet tagging efficiencies and data-

to-simulation scale factors for SMT  $c$  jets are presented. Section 6 describes the extrapolation procedure used to derive  $c$ -jet tagging efficiency scale factors for inclusive  $c$  jets. The systematic uncertainties are discussed in Sec. 7, while the final results are presented in Sec. 8. Conclusions are drawn in Sec. 9.

## 2 Data and simulated samples

This analysis is based on data collected with the ATLAS detector in the year 2011 during periods with stable  $pp$  collisions at  $\sqrt{s} = 7$  TeV and all relevant detector components operational. The resulting data sample corresponds to an integrated luminosity of  $4.6 \text{ fb}^{-1}$  with an uncertainty of 1.8% [8].

Simulated samples are used to compute efficiencies and model kinematic distributions of signal and background processes. The signal is defined to be the production of a  $W$  boson in association with a single charm quark. Background processes include the production of  $W$ +jets containing charm or bottom quark pairs or no heavy flavour quarks, as well as the production of  $Z$ +jets, top-quark pairs, single-top quarks, dibosons ( $WW$ ,  $WZ$  and  $ZZ$ ) and multijet events. The contribution from the production of a  $W$  boson in association with a single bottom quark is negligible.

$W$  bosons produced in association with  $c$  jets,  $b$  jets and light jets are generated separately using **ALPGEN** 2.13 [9]. The showering of the  $W+c$  signal process is done with **PYTHIA** 6.423 [10], while for all other  $W$ +jets processes **HERWIG** 6.520 [11] is used for the parton shower and **JIMMY** 4.31 [12] for the underlying event. An additional signal sample produced with **ALPGEN** and **HERWIG+Jimmy** is used to study systematic uncertainties. Exclusive samples with zero to four additional partons and an inclusive sample with five or more additional partons are used. Overlaps between different **ALPGEN** samples with heavy-flavour quarks originating from the matrix element and from the parton shower are removed. In addition, the **MLM** [13] matching scheme is applied to remove overlaps between events with a given parton multiplicity generated by the matrix element and the parton shower. The **CTEQ6L1** parton distribution function (PDF) [14] is used for all leading order (LO) MC generators.

To improve on known shortcomings in the **ALPGEN+PYTHIA** signal sample and to minimize systematic uncertainties, several  $c$ -quark fragmentation and  $c$ -hadron decay properties are corrected as explained in Sec. 6. While in the following this corrected sample is referred to as the **PYTHIA**-corrected sample, the sample without any of the fragmentation and decay corrections applied is called the **PYTHIA**-default sample. Additionally, signal samples generated with **ALPGEN** and **PYTHIA**, where the **EvtGen** [15] program is used to model hadron decays, are produced to study the  $c$ -hadron decay properties.

Background from  $Z$ +jets events is generated with **ALPGEN** interfaced to **HERWIG** and **JIMMY** using the same configuration as for  $W$ +jets events. For the diboson backgrounds, MC samples generated with **HERWIG** are used. The  $t\bar{t}$  background is obtained from **MC@NLO** 4.01 [16] with **HERWIG** used for the parton shower. The **CT10** [17] PDF is used for the next-to-leading order (NLO) matrix element calculations, while showering is performed with the **CTEQ6L1** PDF. Single-top production is based on the **ACERMC** 3.7 [18] MC generator (interfaced to **PYTHIA**) in the  $t$ -channel, and on the **MC@NLO** generator in the  $s$ -channel and for the associated production with a  $W$  boson. **TAUOLA** [19] and **Photos** [20] are employed to model the decay of  $\tau$  leptons and the QED radiation of photons, respectively.

The background processes are normalized to NNLO predictions in case of the inclusive  $W$ ,  $Z$  and  $t\bar{t}$  productions [21, 22] and to NLO predictions for the other processes [23, 24]. The properties of the multijet background events are determined using data-driven techniques.

Multiple  $pp$  collisions per bunch crossing (pileup) are modelled by overlaying minimum-bias events generated using **PYTHIA** with the hard process.

The simulated events are passed through a detailed simulation of the ATLAS detector response [25] based on **GEANT4** [26].

### 3 Event selection

The data sample used for this analysis is triggered by a single-electron trigger with pseudorapidity coverage of  $|\eta^e| < 2.47$  and a minimum threshold on the transverse momentum  $p_T^e$  of 20 GeV to 22 GeV, depending on the data-taking period.

Events are required to have at least one primary-vertex candidate. The vertex with the highest sum of the squared transverse momenta of associated tracks is selected as the primary vertex.

$W$  bosons are reconstructed in their leptonic decay channel to an electron and a neutrino. Electrons with  $p_T^e > 25$  GeV and  $|\eta^e| < 2.47$ , excluding the calorimeter transition region  $1.37 < |\eta| < 1.52$ , are selected. Electrons are required to pass the “tight” identification criteria described in Ref. [27] and re-optimized for the 2011 data-taking conditions. Calorimeter-based isolation requirements are applied to electrons: the sum of transverse energies in the calorimeter cells within a cone of radius  $\Delta R < 0.3$  around the electron direction,  $\sum_{\Delta R < 0.3} E_T^{\text{cells}}$ , is required to be less than 3 GeV. Exactly one lepton fulfilling isolation requirements is allowed in each event. Events with additional electrons or isolated muons are rejected to suppress background from  $Z$  and  $t\bar{t}$  events. The selection applied to veto leptons is looser than that used for signal leptons to ensure higher background rejection. Scale factors are applied to the MC simulation to bring the trigger and reconstruction efficiencies in accordance with the performance measured in data. The presence of a neutrino is inferred from the presence of missing transverse momentum ( $E_T^{\text{miss}}$ ). The magnitude and azimuthal direction of  $E_T^{\text{miss}}$  are measured from the vector sum of the transverse momenta of calibrated physics objects [28]. Low- $p_T$  tracks are used to recover soft particles which are missed in the calorimeters. A minimum  $E_T^{\text{miss}}$  of 25 GeV is required. In addition the  $W$ -boson transverse mass<sup>1</sup>,  $m_T^W$ , is required to be larger than 40 GeV.

Jets are reconstructed with the FastJet package [29] which uses the infrared- and collinear-safe anti- $k_t$  algorithm [30] with radius parameter  $R = 0.4$ . The detector input is based on topological clusters of calorimeter cells [31]. The jet energy is calibrated to account for the different response of the calorimeters to electrons and hadrons, for energy losses in un-instrumented regions, and for the energy offset introduced by pileup, by applying jet calibration factors dependent on  $p_T$ ,  $\eta$  and pileup conditions [32]. Jets with  $p_T > 25$  GeV and  $|\eta| < 2.5$  are selected. Furthermore, it is required that the ratio of the  $p_T$  sum of tracks inside the jet associated with the primary vertex divided by the  $p_T$  sum of all tracks inside the jet is larger than 0.75, in order to remove jets reconstructed from energy deposits stemming from particles produced in pileup interactions. Events are required to have exactly one jet fulfilling these requirements.

The selected jet is moreover required to contain a muon with  $p_T > 4$  GeV and  $|\eta| < 2.5$ . Muon candidates are formed from pairs of standalone tracks in the inner detector (ID) and the muon spectrometer that are combined using a  $\chi^2$ -matching procedure [33,34]. The reduced  $\chi^2$  value is required to be smaller than 3.2. In addition, two impact parameter requirements need to be fulfilled: a transverse impact parameter requirement,  $|d_0| < 3$  mm, and a longitudinal impact parameter requirement,  $|z_0 \cdot \sin \theta| < 3$  mm. The impact parameters  $d_0$  and  $z_0$  are defined at the point of closest approach to the primary vertex in the transverse plane and  $\theta$  is the polar angle of the muon with respect to the beam axis. A set of ID hit requirements [35] is applied to select high quality tracks also demanding at least one hit in the first pixel layer. Exactly one muon is required to be associated with the jet; the small fraction of events with jets containing more than one muon is discarded. The soft-muon tagging efficiency and mistag rate are measured in data [36]. The overall soft-muon  $c$ -jet tagging efficiency is about 4 %, mainly due to the low semimuonic branching ratio of  $c$  hadrons (approximately 10 %). The light-jet mistag rate is around 0.2 % depending on the jet kinematics. Scale factors are applied to correct the efficiencies in MC simulation to those measured in data.

---

<sup>1</sup> $m_T^W$  is defined as  $m_T^W = \sqrt{2p_T^l E_T^{\text{miss}}(1 - \cos\Delta\phi)}$  where  $\Delta\phi$  is the azimuthal separation between the directions of the lepton and the missing momentum in the transverse plane.

## 4 Determination of the $W+c$ yield

In order to enhance the signal contribution the charge correlation of the  $W+c$  signal process is exploited by measuring the number of OS-SS events,  $N_{\text{bkg}}^{\text{OS-SS}} = N_{\text{bkg}}^{\text{OS}} - N_{\text{bkg}}^{\text{SS}}$ . The remaining background after the OS-SS subtraction consists predominantly of  $W+\text{light}$  events in which the OS/SS asymmetry is due to the correlation of the charge of the  $W$  boson and the fragmentation products of the associated light quark. Multijet production is also an important background afflicted with large cross section uncertainties. The background from  $W+\text{light}$  and multijet events is estimated using data-driven methods. Smaller backgrounds from  $Z/\gamma^*+\text{jets}$ , top and diboson production are estimated from MC simulations. Backgrounds from  $W+b\bar{b}$  and  $W+c\bar{c}$  events are negligible since they are symmetric regarding the number of OS and SS events.

Since the SS sample is mainly composed of  $W+\text{light}$  and multijet events, their sum is constrained by the number of data events in the SS sample. The number of OS-SS events of the  $W+\text{light}$  and multijet backgrounds are therefore obtained from a constrained  $\chi^2$  fit to the number of data events in the SS sample and a subsequent propagation using the following equation

$$N_{\text{bkg}}^{\text{OS-SS}} = A_{\text{bkg}} \cdot N_{\text{bkg}}^{\text{OS+SS}} = \frac{2 \cdot A_{\text{bkg}}}{1 - A_{\text{bkg}}} N_{\text{bkg}}^{\text{SS}}. \quad (1)$$

Here  $N_{\text{bkg}}^{\text{OS+SS}}$  is the sum of the number of background events in the OS ( $N_{\text{bkg}}^{\text{OS}}$ ) and SS ( $N_{\text{bkg}}^{\text{SS}}$ ) samples and  $A_{\text{bkg}}$  the OS/SS asymmetry defined as

$$A_{\text{bkg}} = \frac{N_{\text{bkg}}^{\text{OS-SS}}}{N_{\text{bkg}}^{\text{OS+SS}}}. \quad (2)$$

In the fit, the sum of the  $W+\text{light}$  and multijet backgrounds plus the remaining backgrounds and a small  $W+c$  signal contribution is required to be equal to the total data count in the SS sample. The relative contributions of  $W+\text{light}$  and multijet events are allowed to vary in the fit, while all other backgrounds and the signal contribution are fixed to their values expected from simulations.

The OS/SS asymmetry of the multijet background,  $A_{\text{multijet}}$ , is determined by performing a binned maximum likelihood fit of templates to the  $E_{\text{T}}^{\text{miss}}$  distribution in data separately for the OS and SS samples. The fit is carried out in a control sample where the  $E_{\text{T}}^{\text{miss}}$  selection requirement is released. Two templates are used: one representing the multijet background and the other representing the contributions from all other sources. The multijet template is extracted from a data control region selected by inverting the electron isolation and some of the electron identification requirements. A residual contamination from  $W/Z$  and top quark events is estimated from simulations and subtracted. The non-multijet template representing all other processes, including the  $W+c$  signal,  $W/Z$ , diboson and top quark production, is obtained from simulations where the relative contributions of the individual processes are fixed according to their cross sections predictions.  $A_{\text{multijet}}$  derived according to Eq. 2 using the fit results in the signal region ( $E_{\text{T}}^{\text{miss}} > 25 \text{ GeV}$ ) is consistent with zero within the assigned total uncertainties. The uncertainties are dominated by the statistical component. The systematic uncertainties estimated by varying the fit range and using alternative multijet and non-multijet templates are found to be small.

The OS/SS asymmetry of the  $W+\text{light}$  background,  $A_{W+\text{light}}$ , is obtained from MC simulation and corrected by the asymmetry measured in a data control region using the relation:

$$A_{W+\text{light}} = A_{W+\text{light}}^{\text{MC}} \frac{A_{W+\text{light}}^{\text{data,tracks}}}{A_{W+\text{light}}^{\text{MC,tracks}}}. \quad (3)$$

$A_{W+\text{light}}^{\text{MC}}$  is the OS/SS asymmetry in the MC simulation for the signal region and  $A_{W+\text{light}}^{\text{MC,tracks}}$  ( $A_{W+\text{light}}^{\text{data,tracks}}$ ) is the OS/SS asymmetry in MC (data) estimated using the charges of the  $W$  boson and generic tracks

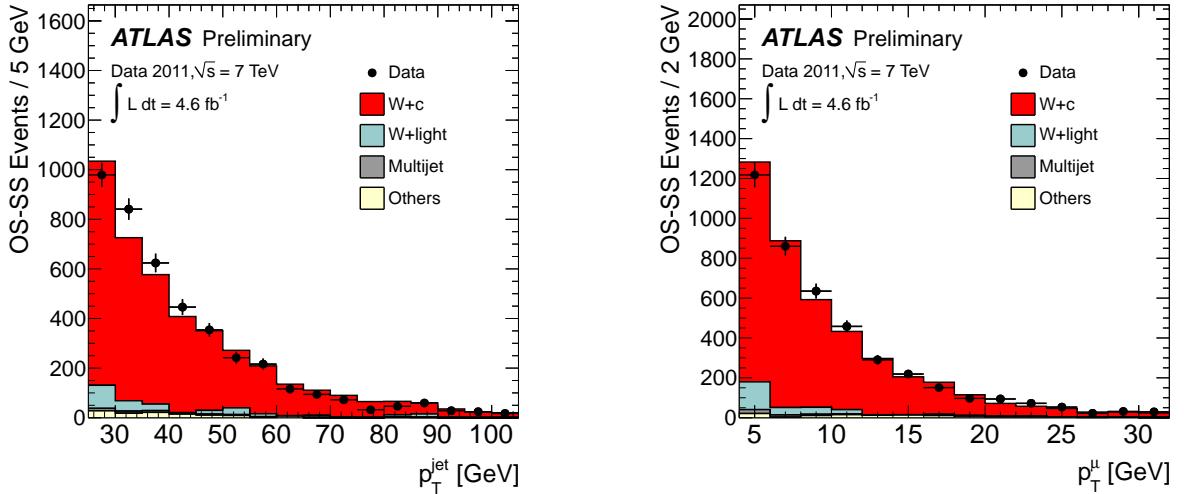


Figure 1: The  $p_T$  distributions of the SMT jet (left) and the soft muon (right) in OS-SS candidate events. The normalization of the  $W$ +light background and the shape and normalization of the multijet background are obtained using data-driven methods. All other backgrounds are estimated with MC simulations and normalized to their theoretical cross sections. The  $W+c$  signal contributions are normalized to the measured yield and the shapes are extracted from the ALPGEN+PYTHIA-corrected sample.

that are associated with the selected jet and pass the soft muon kinematic requirements.  $A_{W+\text{light}}^{\text{MC,tracks}}$  and  $A_{W+\text{light}}^{\text{data,tracks}}$  are derived from a control sample that is selected by relaxing the soft muon requirements.  $A_{W+\text{light}}$  is found to be approximately 10%. The uncertainty on  $A_{W+\text{light}}$  is dominated by the statistical uncertainty on  $A_{W+\text{light}}^{\text{MC}}$ . The sub-leading systematic uncertainty contains contributions from uncertainties on the background contamination in the control sample and the modelling of track properties.

The selected numbers of OS and SS events in the data SMT jet sample are 7445 and 3125. This results in a number of OS-SS data events of  $4320 \pm 100$  and an extracted number of SMT  $W+c$  events of  $3910 \pm 100$  (stat.)  $\pm 160$  (syst.). The estimated number of background events, the number of data events and the measured  $W+c$  yield are summarized in Table 1.

Figure 1 shows the  $p_T$  distributions of the selected SMT jet (left) and the associated soft muon (right) in OS-SS events. The signal contribution is normalized to the measured yield, the background contributions to the values listed in Table 1. The  $W+c$  signal shapes are derived from the ALPGEN+PYTHIA-corrected simulated signal sample. The multijet shapes are extracted from the data control region used to determine  $A_{\text{multijet}}$ . The predictions are in reasonable agreement with data.

## 5 Measurement of the $c$ -jet tagging efficiency of SMT $c$ jets

The selected OS-SS events have a high  $c$ -jet purity (approximately 90 %) and are therefore well suited to perform a measurement of the  $c$ -jet tagging efficiency of SMT  $c$  jets. The output weight of the MV1 tagging algorithm for which the  $c$ -jet tagging efficiency is measured is shown in Fig. 2. The  $W+c$  signal contribution is normalized to the measured yield while its shape is extracted from the ALPGEN+PYTHIA-corrected sample. The MV1 output weight is designed as such that real  $b$  jets tend to have high values close to one and light jets low values close to zero. The operating points calibrated in this note are defined by cuts on the MV1 output weight, namely 0.07, 0.40, 0.60 and 0.91, corresponding to  $b$ -tagging efficiencies of 85 %, 75 %, 70 % and 60 %, respectively, in simulated  $t\bar{t}$  events. The distribution of the

Number of events $N^{\text{OS-SS}}$	
W+light	$240 \pm 110$
Multijet	$50 \pm 130$
$t\bar{t}$	$13 \pm 5$
Single top	$62 \pm 10$
Diboson	$35 \pm 5$
Z+jets	$6 \pm 14$
Total background	$410 \pm 160$
$W+c$ (meas.)	$3910 \pm 190$
Data	$4320 \pm 100$

Table 1: Number of OS-SS events for different backgrounds and for the selected data sample. The measured  $W+c$  yields are also shown. The uncertainties on the yields include statistical and systematic uncertainties. Correlations between the uncertainties due to the constraint in the SS sample are taken into account when computing the total background uncertainties.

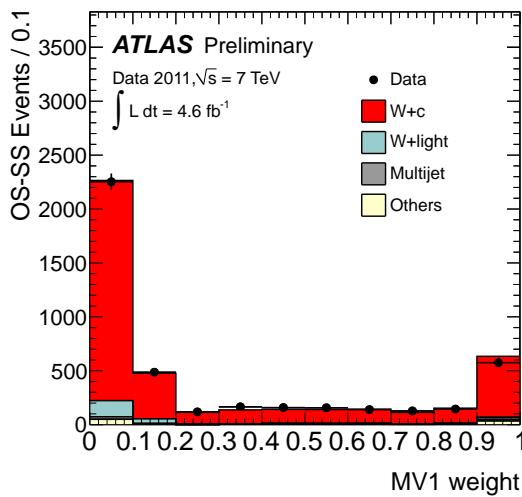


Figure 2: Output weight of the MV1 tagging algorithm for the SMT  $c$ -jet sample selected in data and simulation. The  $W+c$  signal contribution is normalized to the measured yield and the shape is extracted from the ALPGEN+PYTHIA-corrected sample.

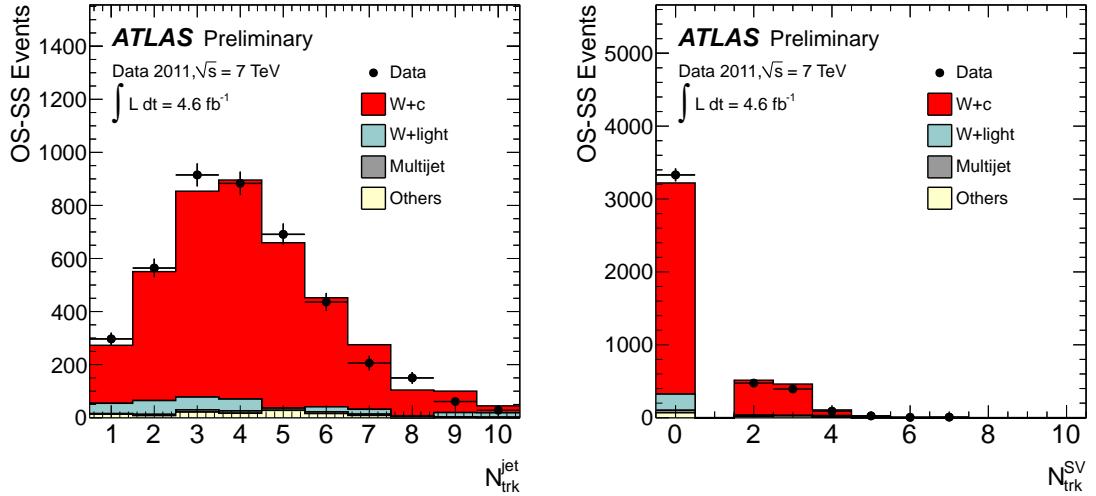


Figure 3: Track and secondary vertex (SV) properties of the selected SMT  $c$  jets: the number of tracks associated with the selected jet,  $N_{\text{trk}}^{\text{jet}}$ , used by the impact parameter (IP) based tagging algorithms (left); the number of tracks associated with a reconstructed SV,  $N_{\text{trk}}^{\text{SV}}$ , exploited by the SV1 tagging algorithm (right). The events with zero tracks correspond to the case of no SV being reconstructed. The  $W+c$  signal contributions are normalized to the measured yield and the shapes are extracted from the ALPGEN+PYTHIA-corrected sample.

number of tracks inside the selected SMT jets as well as of the number of tracks that are furthermore associated with a secondary decay vertex (SV) that serve as important discriminants when identifying  $b$  jets are shown in Fig. 3. It should be noted that in what follows the number of events refers to the number of OS-SS subtracted events unless indicated otherwise.

The  $c$ -jet tagging efficiency of SMT  $c$  jets,  $\varepsilon_{c(\mu)}^{\text{data}}$ , is derived as the fraction of  $W+c$  events selected in data that pass a certain  $b$ -tagging requirement

$$\varepsilon_{c(\mu)}^{\text{data}} = \frac{N_{Wc}^{\text{b-tag}}}{N_{Wc}}, \quad (4)$$

where  $N_{Wc}$  is the number of  $W+c$  events before applying the  $b$ -tagging requirement (hereinafter referred to as *pretag* level or sample) and  $N_{Wc}^{\text{b-tag}}$  is the number of  $W+c$  events passing the  $b$ -tagging requirement.<sup>2</sup>  $N_{Wc}$  is derived as described in Sec. 4. The number of  $b$ -tagged signal events  $N_{Wc}^{\text{b-tag}}$  is determined from

$$N_{Wc}^{\text{b-tag}} = N_{\text{data}}^{\text{b-tag}} - \sum_{\text{bkg}} N_{\text{bkg}}^{\text{b-tag}}, \quad (5)$$

where the number of events in data that pass the  $b$ -tagging requirement is denoted  $N_{\text{data}}^{\text{b-tag}}$  and the sum runs over all considered backgrounds. The number of  $b$ -tagged events of a given background,  $N_{\text{bkg}}^{\text{b-tag}}$ , is computed using the following expression

$$N_{\text{bkg}}^{\text{b-tag}} = \varepsilon_{\text{bkg}} \cdot N_{\text{bkg}}, \quad (6)$$

<sup>2</sup>The selected  $W+c$  sample has a small contamination of about 1 % of non- $c$  jets at pretag level and less than 0.2 % after applying  $b$ -tagging requirements, which is estimated using MC simulation and subtracted.

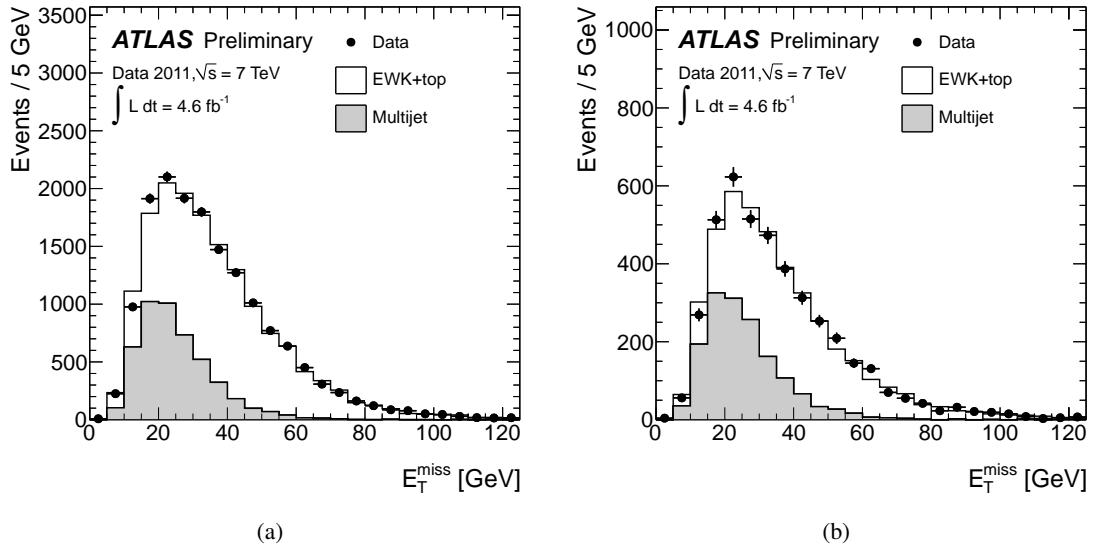


Figure 4: Fits to the  $E_T^{\text{miss}}$  distribution in the sum of the OS and SS samples used to determine the number of multijet events (a) before and (b) after applying a cut on the MV1 weight corresponding to the  $\varepsilon_b = 70\%$  operating point. “EWK+top” corresponds to the production of  $W$  and  $Z$  bosons, the production of single top and top quark pairs and the production of dibosons.

where  $N_{\text{bkg}}$  is the number of background events in the pretag sample listed in Table 1 and  $\varepsilon_{\text{bkg}}$  denotes the tagging rate of the considered background sample. The tagging rate, i.e. the fraction of events that pass the applied  $b$ -tagging requirement, depends both on the individual jet flavour composition of the considered background and on the tagging efficiencies or mistag rates of the different jet flavours.

The tagging rates of the  $W$ +light,  $t\bar{t}$ , single top, diboson and  $Z$ +jets backgrounds are extracted using MC simulation. The tagging efficiencies of the differently flavoured jets are corrected to match those in data by applying  $b$ -tagging scale factors [1, 37]. The corresponding systematic uncertainties are taken into account. The total uncertainties on the tagging rates are dominated by the statistical uncertainties due to the limited size of the simulated samples.

The tagging rate of the multijet background is estimated using a data-driven method. A binned maximum likelihood fit of templates to the  $E_T^{\text{miss}}$  distribution in data is performed both before and after applying the  $b$ -tagging requirement. The procedure follows closely the extraction of the OS/SS asymmetry at pretag level described in Sec. 4. Accordingly, the non-multijet templates are built using MC simulations and the multijet templates are extracted from a data control sample selected by inverting some of the electron identification criteria as well as the electron isolation requirement. The templates used to derive the number of  $b$ -tagged multijet events are obtained from control samples that are additionally required to pass the  $b$ -tagging requirement. Since the multijet-template shapes in the OS and SS control samples are found to be consistent, the multijet templates are derived from the sum of the number of OS and SS events to reduce the statistical uncertainties. The multijet tagging rates computed using the fit results for the OS and SS samples lead to compatible results. Therefore, the final multijet tagging rate is obtained from the fit results derived for the sum of the OS and SS samples. The fit results before and after applying the MV1  $b$ -tagging requirement corresponding to the  $\varepsilon_b=70\%$  operating point are shown in Fig. 4. The multijet tagging rates for the different operating points vary between 26 % and 55 % indicating that the multijet sample has a large heavy flavour component. The assigned total uncertainties range between 15 % and 23 % accounting for the dominating statistical uncertainties of the  $E_T^{\text{miss}}$  fits, the choice of the fit range and the shapes of the multijet and non-multijet templates.

Operating points ( $\varepsilon_b$ ) of the MV1 tagging algorithm			
85 %	75 %	70 %	60 %
$0.99 \pm 0.03 \pm 0.03$	$0.96 \pm 0.04 \pm 0.04$	$0.92 \pm 0.05 \pm 0.05$	$0.87 \pm 0.07 \pm 0.06$

Table 2: Data-to-simulation  $c$ -jet tagging efficiency scale factors for SMT  $c$  jets derived for the MV1 tagging algorithm with respect to a  $W+c$  sample simulated with ALPGEN+PYTHIA. The first uncertainty is the statistical uncertainty, the second is the systematic uncertainty due to the background determination, the event reconstruction and the simulated sample size.

The  $c$ -jet tagging efficiencies for the MV1 tagging algorithm derived for SMT  $c$  jets according to Eq. 4 are shown in Fig. 5(a). They vary between 13 % and 50 % with total uncertainties of 3-10 % that increase with the tightness of the operating point. The systematic uncertainties are dominated by the precision on the  $W$ +light and multijet background yields at pretag level, in particular on the data-driven OS/SS asymmetry estimates as discussed in Sec. 4, and on the  $W$ +light tagging rate. The statistical uncertainties are of the same order as the systematic uncertainties.

The expected  $c$ -jet tagging efficiency,  $\varepsilon_{c(\mu)}^{\text{sim}}$ , is defined as the fraction of SMT  $c$  jets selected in simulated ALPGEN+PYTHIA-default  $W+c$  events that pass the  $b$ -tagging requirement. In Figure 5(a)  $\varepsilon_{c(\mu)}^{\text{sim}}$  is compared to  $\varepsilon_{c(\mu)}^{\text{data}}$  for the different  $b$ -tagging operating points.

The data-to-simulation scale factor for SMT  $c$  jets  $\kappa_{c(\mu)} = \varepsilon_{c(\mu)}^{\text{data}} / \varepsilon_{c(\mu)}^{\text{sim}}$  is shown for different operating points in Fig. 5(b) and summarized in Table 2. The quoted systematic uncertainties arise from the previously discussed background determinations as well as from the  $W$  boson reconstruction and the SMT  $c$ -jets identification-efficiencies. A detailed discussion and a breakdown of the systematic uncertainties can be found in Sec. 7. For looser operating points  $\kappa_{c(\mu)}$  is found to be compatible with unity within the total uncertainty, but for tighter operating points  $\kappa_{c(\mu)}$  shows a trend towards lower values inconsistent with unity.

## 6 Calibration of the $c$ -jet tagging efficiency for inclusive $c$ -jet samples

Due to several differences between an inclusive sample of  $c$  jets and a sample of SMT  $c$  jets the derived  $c$ -jet tagging efficiency scale factors need to be extrapolated in order to be applicable to samples of inclusive  $c$  jets. Selecting a sample of  $c$  jets via the semimuonic decays of  $c$  hadrons results in a different composition of  $c$ -hadron types with regard to an inclusive sample due to the different semileptonic branching ratios of the  $c$  hadrons. Since the  $c$ -hadron types differ in several characteristics relevant for the performance of  $b$ -tagging algorithms, e.g. the lifetime or the charged decay multiplicity, their tagging efficiencies also differ. For the  $\varepsilon_b=70$  % operating point of the MV1 tagging algorithm the tagging efficiencies of the most prominent weakly decaying  $c$ -hadron types are for instance

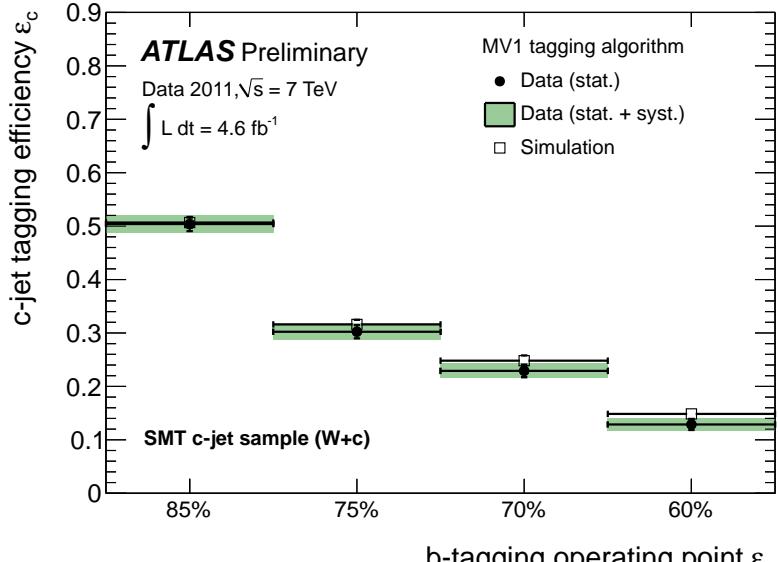
$$D^0 : 0.157 \pm 0.001, D^+ : 0.280 \pm 0.002, D_s : 0.152 \pm 0.003, \Lambda_c^+ : 0.041 \pm 0.002$$

as estimated from the ALPGEN+PYTHIA-corrected sample, where the quoted uncertainties are statistical only. Therefore, the overall  $c$ -jet tagging efficiency of a  $c$ -jet sample strongly depends on the sample composition.

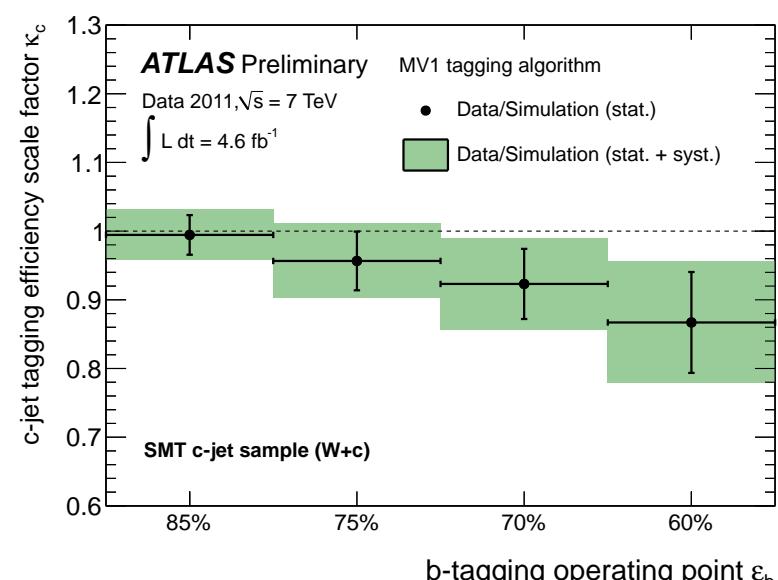
The  $c$ -jet tagging efficiency of an inclusive sample of  $c$  jets,  $\varepsilon_c$ , can be obtained from the  $c$ -jet tagging efficiency of SMT  $c$  jets,  $\varepsilon_{c(\mu)}$ , by applying an extrapolation factor  $\alpha$ :

$$\varepsilon_c = \alpha \cdot \varepsilon_{c(\mu)}. \quad (7)$$

A comparison of the expected  $\varepsilon_c^{\text{sim}}$  and  $\varepsilon_{c(\mu)}^{\text{sim}}$  for several operating points of the MV1 tagging algorithm derived using a  $W+c$  sample simulated with ALPGEN+PYTHIA-default is shown in Fig. 6.  $\varepsilon_c^{\text{sim}}$  is sys-



(a)



(b)

Figure 5: (a) Comparison of the  $c$ -jet tagging efficiencies of SMT  $c$  jets in data and the ALPGEN+PYTHIA-default simulation, derived for the MV1 tagging algorithm using  $W+c$  events. (b) Data-to-simulation  $c$ -jet tagging efficiency scale factors for SMT  $c$  jets derived for the MV1 tagging algorithm using  $W+c$  events.

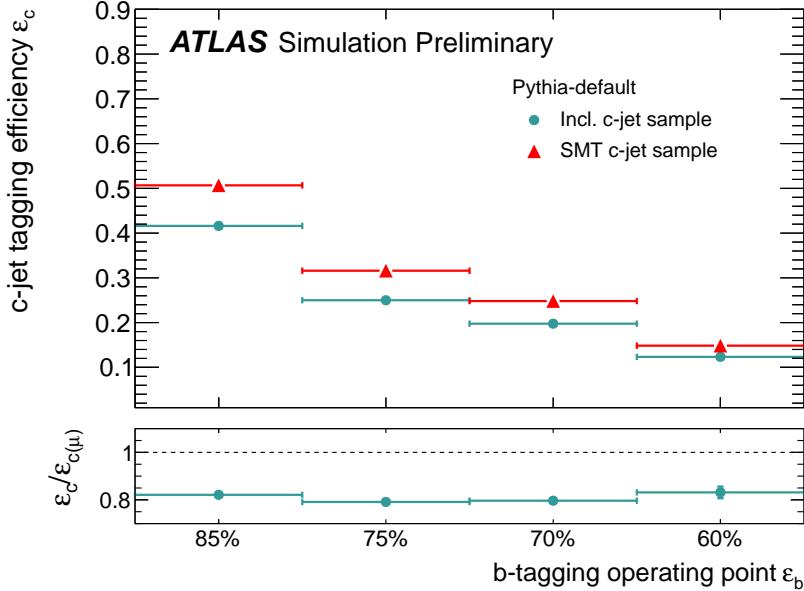


Figure 6: Comparison of the  $c$ -tagging efficiencies of SMT  $c$  jets and inclusive  $c$  jets predicted from the ALPGEN+PYTHIA-default signal sample. The ratio  $\epsilon_c^{\text{sim}} / \epsilon_{c(\mu)}^{\text{sim}}$  corresponds to the efficiency extrapolation factor  $\alpha_{\text{sim}}$ .

tematically lower than  $\epsilon_{c(\mu)}^{\text{sim}}$  resulting in a correction factor  $\alpha_{\text{sim}}$  of about 0.8 for the different operating points.

Similarly, the  $c$ -jet tagging efficiency scale factor  $\kappa_c$  for inclusive  $c$  jets can be computed from the measured  $c$ -jet tagging efficiency scale factor  $\kappa_{c(\mu)}$  by applying a correction factor  $\delta$

$$\kappa_c = \frac{\epsilon_c^{\text{data}}}{\epsilon_c^{\text{sim}}} \stackrel{\text{Eq. 7}}{=} \frac{\alpha_{\text{data}} \cdot \epsilon_{c(\mu)}^{\text{data}}}{\alpha_{\text{sim}} \cdot \epsilon_{c(\mu)}^{\text{sim}}} = \frac{\alpha_{\text{data}}}{\alpha_{\text{sim}}} \cdot \kappa_{c(\mu)} = \delta \cdot \kappa_{c(\mu)}, \quad (8)$$

where  $\delta$  is expressed as the ratio of the efficiency extrapolation factors in data,  $\alpha_{\text{data}}$ , and simulation,  $\alpha_{\text{sim}}$ . Mismodelling of the differences between SMT  $c$  jets and inclusive  $c$ -jet samples in MC simulation leads to different extrapolation factors  $\alpha$  and thus to a ratio  $\delta$  deviating from one.

The efficiency extrapolation factor  $\alpha_{\text{data}}$  is estimated using the simulated ALPGEN+PYTHIA-corrected sample that has several corrections applied to minimize  $c$ -quark fragmentation and  $c$ -hadron decay differences between data and simulation. The scale factor extrapolation factor  $\delta$  is thus estimated by

$$\delta \approx \frac{\alpha_{\text{sim}}^{\text{corr}}}{\alpha_{\text{sim}}}. \quad (9)$$

In order to correctly describe the  $c$ -hadron composition of the inclusive  $c$ -jet sample the fragmentation fractions of the relevant weakly decaying  $c$ -hadron types in the PYTHIA-default sample are re-weighted to those obtained by combining the results of measurements in  $e^+e^-$  and  $e^\pm p$  collisions [38]. The fragmentation fractions implemented in the PYTHIA-default simulation are compared to the data results in Fig. 7(a). By correcting the semileptonic branching ratios of  $c$  hadrons to match the world average values [39], also the modelling of the  $c$ -hadron composition of the simulated SMT  $c$ -jet sample is improved. The  $c$ -hadron fractions in the SMT  $c$ -jet sample predicted by the PYTHIA-default simulation are compared to those expected by combining the fragmentation fractions determined in  $e^+e^-$  and  $e^\pm p$  collisions with the measured semimuonic branching ratios of  $c$  hadrons in Fig. 7(b). While the dominant  $c$ -hadron type

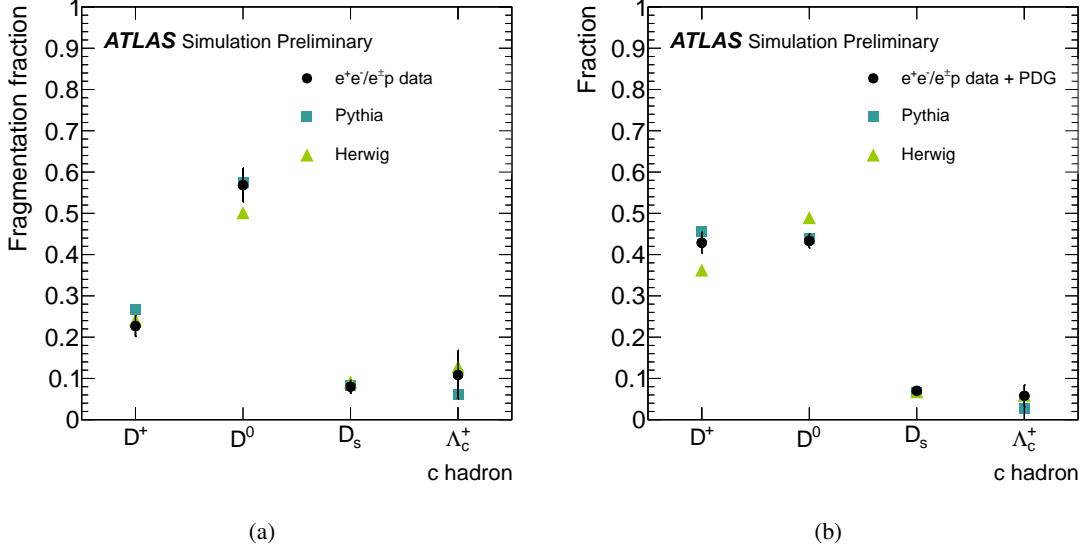


Figure 7: (a) Fragmentation fractions of the most relevant weakly decaying  $c$  hadrons. The fractions in the PYTHIA-default and HERWIG samples are compared to combined results of  $e^+e^-$  and  $e^\pm p$  measurements [38]. (b) Fractions of the different  $c$ -hadron types in a sample of  $c$  jets associated with semileptonically decaying  $c$  hadrons. The fractions in the PYTHIA-default and HERWIG samples are compared to the results expected from independent measurements of the  $c$ -hadron fractions [38] combined with the corresponding semimuonic branching ratios [39].

of the inclusive  $c$ -jet sample is the  $D^0$  meson ( $\sim 60\%$ ), the  $D^+$ -meson fraction is strongly enhanced in the SMT  $c$ -jet sample due to the relatively large semileptonic branching ratio of the  $D^+$  meson. Therefore, the SMT  $c$ -jet sample consists of a similar amount ( $\sim 43\%$ ) of  $D^0$  and  $D^+$  mesons.

Given that the main input variables to the  $b$ -tagging algorithms exploit track and vertex properties that can specifically be associated with  $b$ - and  $c$ -hadron decays, it is important that the charged decay multiplicity of  $c$ -hadron decays is well modelled by the MC simulation. In order to improve its description the relative branching fractions of the dominant semileptonic decay channels of the abundant  $D^+$  and  $D^0$  mesons are corrected in the simulation to match the world average values [39]. The less frequent decay channels, which are known to a lower precision, are adjusted to maintain the overall normalization. Since both of the dominant semileptonic  $D^0$  decays ( $D^0 \rightarrow K^-\mu^+\nu_\mu$ ,  $D^0 \rightarrow \bar{K}^0\pi^-\mu^+\nu_\mu$ ) have two charged decay products, referred to as 2-prong decays, the impact of the correction on the predicted  $c$ -jet tagging efficiency of SMT  $c$  jets is small. In case of the  $D^+$  meson the most dominant semileptonic decay ( $D^+ \rightarrow \bar{K}^0\mu^+\nu_\mu$ ) is a 1-prong decay, while the second dominant ( $D^+ \rightarrow K^-\pi^+\mu^+\nu_\mu$ ) and most of the less frequent decay channels are 3-prong decays. Therefore, a mismodelling of the relative fractions in case of the  $D^+$  meson has a noticeable impact on the  $c$ -jet tagging efficiency of SMT  $c$  jets. Furthermore, the hadronic  $n$ -prong branching ratios of  $c$  hadrons in the PYTHIA-default sample are corrected with a significant impact on the  $c$ -jet tagging efficiency of samples of inclusive  $c$  jets. The corrections in case of the  $D^0$  meson have been inferred from the measured inclusive  $n$ -prong branching ratios [39]. The hadronic  $n$ -prong branching ratios of the  $D^+$  and  $D_s$  mesons as well as the  $\Lambda_c^+$  baryon in the PYTHIA-default sample are re-weighted to the predictions of the EvtGen simulation. A comparison between the PYTHIA-default and EvtGen predictions, as well as the measured values in case of the  $D^0$  meson, reveals large differences in the hadronic  $n$ -prong distributions. Especially, the 2-to-0-prong ratio of the  $D^0$  meson and the 3-to-1-prong ratio of the  $D^+$  meson have a significant impact on the predicted inclusive  $c$ -jet

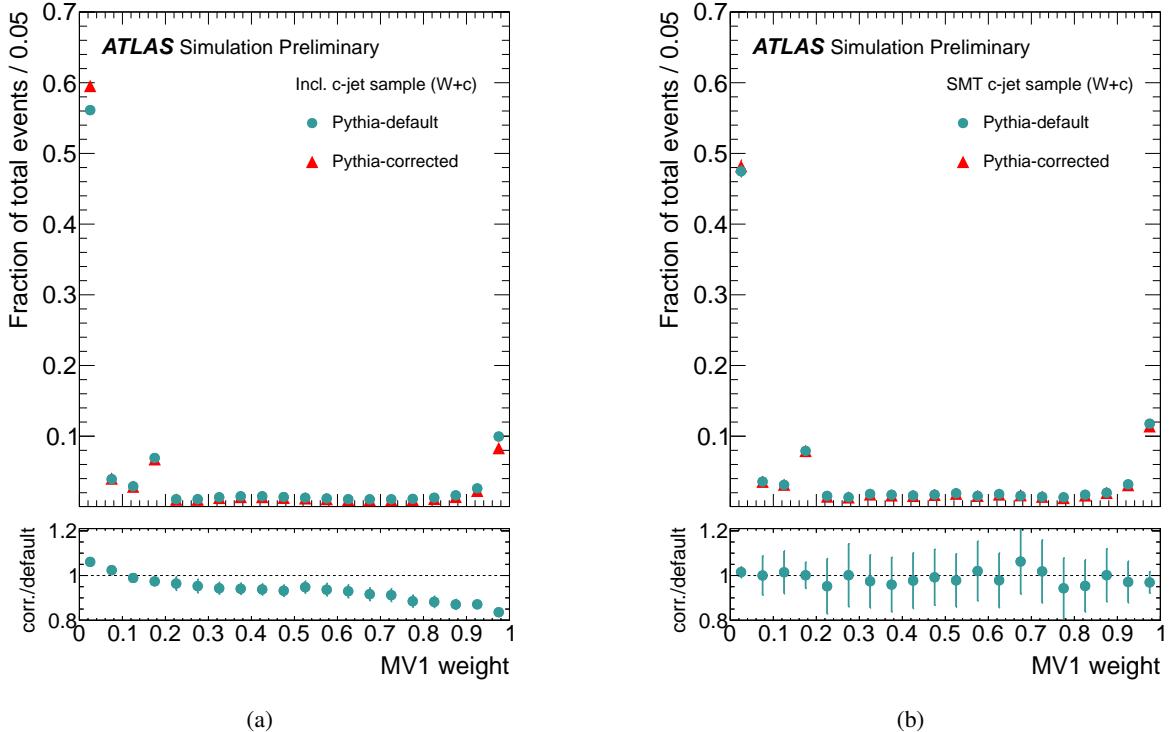


Figure 8: Comparison of the output weight of the MV1 tagging algorithm for (a) inclusive  $c$  jets and (b) SMT  $c$  jets derived with the ALPGEN+PYTHIA signal sample before (PYTHIA-default) and after applying all corrections regarding the  $c$ -quark fragmentation and  $c$ -hadron decay properties (PYTHIA-corrected).

tagging efficiency.<sup>3</sup>

The  $b$ -tagging performance is also dependent on the kinematic distributions of the  $c$ -hadron decay products: first the effect of any mismodelling of the momentum fraction of the  $c$  hadron ( $p_T^c \text{ hadron} / p_T^c \text{ jet}$ ) - which is sensitive to the  $c$ -quark fragmentation function - is evaluated by comparing different simulations; second the momentum of the decay muon in the rest frame of the  $c$  hadron ( $p^*$ ) is re-weighted to agree with the EvtGEN prediction.

The effect of these corrections on the output weight of the MV1 tagging algorithm for the ALPGEN+PYTHIA signal sample is shown in Fig. 8. The higher values in the tail of the output weight distribution of the inclusive  $c$ -jet sample (Fig. 8(a)) indicate that the ALPGEN+PYTHIA-default simulation predicts a higher  $b$ -tagging efficiency for  $c$  jets than the ALPGEN+PYTHIA-corrected simulation. This is largely explained by the significant correction of the hadronic  $n$ -prong decay branching ratios in the PYTHIA-default sample. The impact of the corrections on the output weight of the SMT  $c$ -jet sample (Fig. 8(b)) is less pronounced.

Computing the efficiency correction factor using the ALPGEN+PYTHIA-corrected sample leads to a results of  $\alpha_{\text{sim}}^{\text{corr}}$  (0.69-0.76) that is systematically lower than  $\alpha_{\text{sim}}$  (0.79-0.83). Moreover a dependence on the operating point is observed: the values decrease with an increase of the tightness of the operating point.

<sup>3</sup>For the  $D^0$  meson, PYTHIA-default predicts a fraction of hadronic 2-prong decays of  $(75.72 \pm 0.17)\%$  compared to  $(65 \pm 7)\%$  inferred from what is reported by the PDG; the respective 0-prong fractions are  $(5.03 \pm 0.04)\%$  and  $(17 \pm 7)\%$ . Thus the 2-to-0-prong ratio in PYTHIA-default is much larger. For the  $D^+$  meson PYTHIA-default predicts a fraction of hadronic 3-prong decays of  $(61.7 \pm 0.3)\%$ , while EvtGEN predicts  $(50.5 \pm 0.2)\%$ ; the respective 1-prong fractions are  $(36.1 \pm 0.2)\%$  and  $(50.5 \pm 0.2)\%$ . Hence the 3-to-1-prong ratio predicted by PYTHIA-default is higher. The uncertainties on the predicted fractions are statistical only.

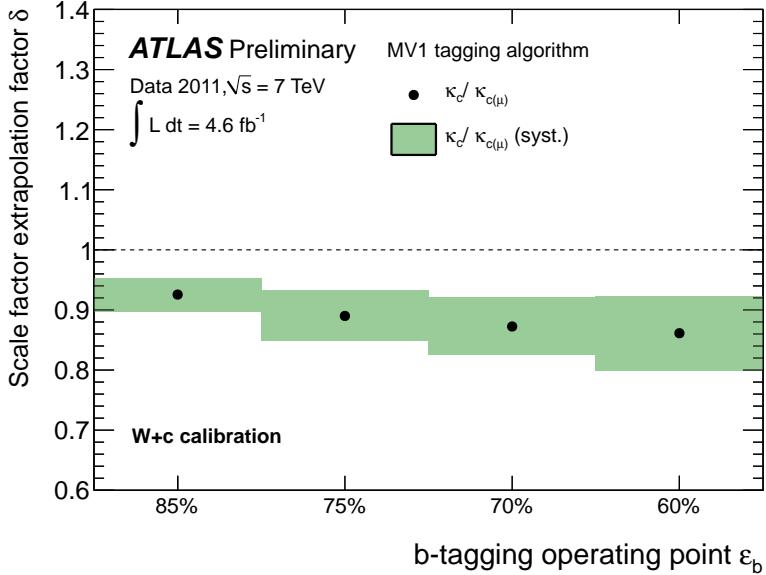


Figure 9: Scale factor extrapolation factor  $\delta$  between the  $c$ -jet tagging efficiency scale factors derived for SMT  $c$  jets and inclusive  $c$  jets for the MV1 tagging algorithm. The systematic uncertainties due to the correction procedure are indicated by the error bands. Statistical uncertainties are omitted since the numerator and the denominator of  $\delta$  are computed using the same events.

Therefore, the resulting scale factor extrapolation factors  $\delta$ , shown in Fig. 9, are systematically lower than unity with a decreasing trend towards tighter operating points. The total systematic uncertainties are due to the before-mentioned corrections and are discussed in detail in Sec. 7. Statistical uncertainties can be neglected since the numerator and the denominator of  $\delta$  are computed using approximately the same simulated events.

## 7 Systematic uncertainties

Systematic uncertainties on the  $c$ -jet tagging efficiency scale factors arise from the  $W$  boson reconstruction and SMT  $c$  jet identification, the pretag yield and tagging rate determination of the backgrounds as well as from the extrapolation procedure to correct the measured  $c$ -jet tagging efficiency scale factors for SMT  $c$  jets. The different contributions are summarized in Table 3 and discussed below.

### 7.1 Event reconstruction

The  $W$  boson reconstruction uncertainty arises from the electron trigger and reconstruction efficiencies, the electron energy scale and resolution as well as the  $E_T^{\text{miss}}$  reconstruction. There are two main sources for the uncertainty on the  $c$ -jet identification, namely the determination of the jet energy scale (JES) and resolution (JER) as well as the reconstruction and tagging efficiencies and energy resolution of the soft muon. The electron uncertainties are assessed by varying the electron efficiencies, energy scale and resolution in simulation independently within the range of their uncertainties as determined from data and re-calculating the resulting  $c$ -jet tagging efficiency. The soft muon uncertainties are determined accordingly: the reconstruction efficiency is varied within its measured precision, as are the soft muon tagging efficiency and mistag rate. The effect of the JES and JER uncertainties [40] is estimated by varying the jet four-momentum in the simulation and re-computing the  $c$ -jet tagging efficiency. The uncertainties

Source	Operating points ( $\varepsilon_b$ ) of the MV1 tagging algorithm			
	85 %	75 %	70 %	60 %
Event reconstruction	1.4	2.1	3.4	3.4
Background pretag yields	0.8	2.1	2.3	4.0
Background tagging rates	1.6	1.9	2.2	2.4
$c$ -quark fragmentation	0.7	0.7	0.9	1.0
Hadronic $c$ -hadron decays	2.1	3.7	4.8	6.3
Semileptonic $c$ -hadron decays	2.1	2.9	2.5	3.3
Simulated sample size	1.2	1.9	2.0	2.7
Total	4.0	6.2	7.4	9.6

Table 3: Summary of the systematic uncertainties on the  $c$ -jet tagging efficiency scale factor for inclusive  $c$  jets  $\kappa_c$ . The values are listed in percent.

on the lepton and jet energy scale and resolution are additionally propagated to the reconstruction of the missing transverse energy. Further systematic uncertainties due to the calibration of the soft jet energy scale ( $7 \text{ GeV} < p_T^{\text{jet}} < 20 \text{ GeV}$ ) and the calorimeter cells not associated with reconstructed objects that also affect the  $E_T^{\text{miss}}$  reconstruction are accounted for.

The systematic uncertainties due to the JES and JER calibrations are dominating the event reconstruction uncertainties, but are of the same order as the uncertainty due to the limited size of the simulated signal sample on which they are assessed.

## 7.2 Pretag yields and background tagging rates

The determination of the OS-SS background yields at pretag level and the assessment of the corresponding uncertainties is discussed in detail in Sec. 4. The main source of systematic uncertainties is the data-driven estimation of the OS/SS asymmetries of the  $W$ +light and multijet backgrounds. The uncertainty due to the background tagging rates is dominated by the uncertainty on the  $W$ +light tagging rate mainly because of the limited size of the simulated sample used to derive it, as discussed in Sec. 5. The systematic uncertainty due to the small contamination of non- $c$ -jets in the  $W+c$  sample is marginal.

## 7.3 Fragmentation and decay modelling

The  $c$ -quark fragmentation and  $c$ -hadron decay properties are corrected to improve the modelling of the ALPGEN+PYTHIA signal sample as described in Sec. 6. Whenever results from independent measurements are used to correct the MC description, the uncertainties assigned to those results are propagated to the extrapolated scale factors. This is done for the fragmentation fractions and the semileptonic decay branching ratios of the prominent weakly decaying  $c$ -hadrons as well as for the hadronic  $n$ -prong decay branching ratios of the  $D^0$  meson. Where corrections are derived from MC simulations because no measurements are available, the corresponding systematic uncertainties are assessed by comparing predictions from different MC generators. Therefore, the difference between the PYTHIA and HERWIG simulations is used to estimate the uncertainty due to the fragmentation function of  $c$  quarks. The systematic uncertainty due to a possible mismodelling of the  $p^*$  distribution is evaluated from the difference between the EvtGEN and PYTHIA simulations. The largest difference between the EvtGEN and either the PYTHIA or HERWIG simulations is used to estimate the uncertainties due to the hadronic  $n$ -prong decay branching ratios of the  $D^+$  and  $D_s$  mesons as well as the  $\Lambda_c^+$  baryon. The largest effect on the final scale factors computed for inclusive  $c$  jets arises from the correction of the  $n$ -prong decay branching ratios of

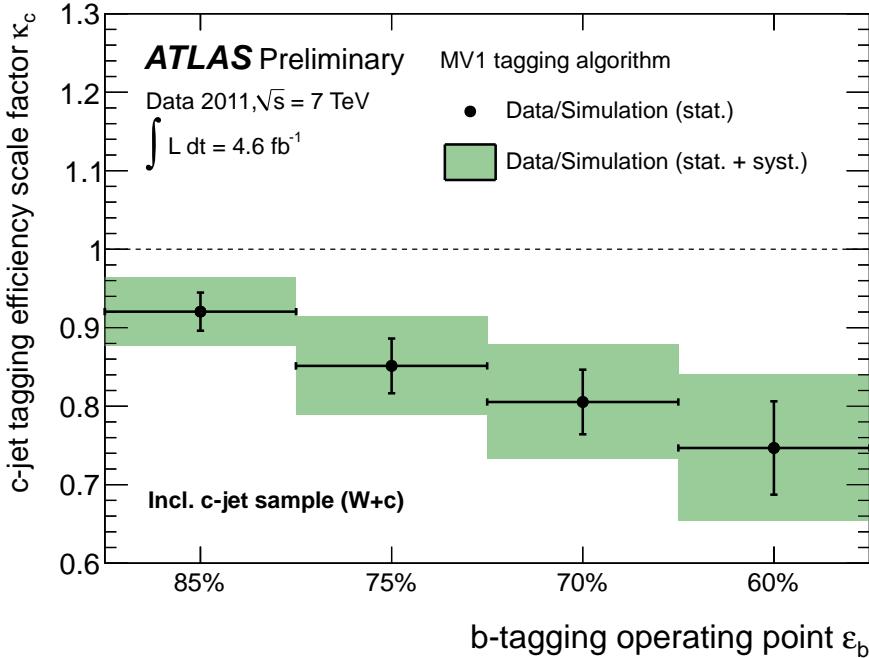


Figure 10: Data-to-simulation  $c$ -jet tagging efficiency scale factors for inclusive  $c$  jets derived for the MV1 tagging algorithm with respect to an ALPGEN+PYTHIA sample.

Operating points ( $\varepsilon_b$ ) of the MV1 tagging algorithm			
85 %	75 %	70 %	60 %
$0.92 \pm 0.02 \pm 0.02 \pm 0.03$	$0.85 \pm 0.03 \pm 0.03 \pm 0.04$	$0.81 \pm 0.04 \pm 0.04 \pm 0.04$	$0.75 \pm 0.06 \pm 0.05 \pm 0.05$

Table 4: Data-to-simulation  $c$ -jet tagging efficiency scale factors for inclusive  $c$  jets,  $\kappa_c$ , derived for the MV1 tagging algorithm with respect to an ALPGEN+PYTHIA sample. The given uncertainties are in order of appearance: statistical uncertainty due to the number of events in the data sample, the systematic uncertainty on the measured scale factors for the SMT  $c$ -jet sample and finally the systematic uncertainty due to the extrapolation procedure.

hadronically decaying  $c$  hadrons. Since only semileptonically decaying  $c$  hadrons to muons are used in the data measurement, a mismodelling of the properties of hadronically decaying  $c$  hadrons propagates fully to the scale factors for inclusive  $c$  jets.

## 8 Results

The data-to-simulation  $c$ -jet tagging efficiency scale factors for the MV1 tagging algorithm with respect to a  $W+c$  sample simulated with ALPGEN+PYTHIA-default are shown in Fig. 10 and listed in Table 4. Being applicable to inclusive samples of  $c$  jets, they are derived from the measured  $c$ -jet tagging efficiency scale factors for SMT  $c$  jets (see Sec. 5) by a simulation-based extrapolation procedure. Their values range between 0.75 and 0.92 with total relative uncertainties of 13 % to 5 %. Three sources of uncertainties are distinguished that are of the same order: the statistical uncertainty, the systematic uncertainty on the measured scale factors for the SMT  $c$ -jet sample and the systematic uncertainty due to the extrapolation procedure. The scale factors decrease and the corresponding uncertainties increase with increasing tightness of the operating point. The modelling of the charged particle multiplicity in  $c$ -hadron decays

is found to have a significant impact on the  $c$ -jet tagging efficiency. In particular the large differences observed between the reference **ALPGEN+PYTHIA** sample and the best knowledge from independent measurements [39] or an alternative signal sample simulated with the **EvtGEN** program lead to the low values of the scale factors for inclusive  $c$ -jet samples.

## 9 Conclusion

The  $b$ -tagging efficiency for  $c$  jets is measured with the ATLAS detector using data collected in proton-proton collisions at  $\sqrt{s} = 7$  TeV at the LHC, corresponding to an integrated luminosity of about  $4.6 \text{ fb}^{-1}$ .

The  $c$ -jet tagging efficiency calibration is performed on a sample where a single  $c$  jet is produced in association with a  $W$  boson. The  $W$  boson is reconstructed via its decay into an electron and a neutrino and the  $c$  jet is identified via a soft muon stemming from the semileptonic decay of the  $c$  hadron. Exploiting the charge correlation of the two decay leptons allows to extract a  $c$ -jet sample with very high purity of about 90 %.

The tagging efficiencies for  $c$  jets containing a muon are measured for several operating points of the **MV1** tagging algorithm and are compared to predictions from **ALPGEN+PYTHIA**. The corresponding data-to-simulation  $c$ -jet tagging efficiency scale factors vary between 0.87 and 0.99 with total relative uncertainties of 10 % to 4 % depending on the operating point. The statistical uncertainty is either of the same order as or larger than the systematic uncertainty.

A second set of data-to-simulation  $c$ -jet tagging efficiency scale factors that is applicable to an inclusive sample of  $c$  jets is presented. This second set is derived from the scale factors measured for  $c$  jets containing a muon by a simulation-based extrapolation procedure. The relative systematic uncertainties due to this extrapolation amount to 4-10 % depending on the considered operating point. The resulting inclusive  $c$ -jet tagging efficiency scale factors range between 0.75 and 0.92 with total uncertainties of 13 % to 5 %, where the scale factors decrease and the uncertainties increase with the tightness of the operating point. The three main contributions to the total uncertainties, the statistical uncertainties, the systematic uncertainties on the measured scale factors for the **SMT**  $c$ -jet sample and the systematic uncertainties due to the extrapolation procedure, are of the same order. The main source for the systematic uncertainties due to the performed extrapolation is the limited knowledge of the charged particle multiplicity of  $c$ -hadron decays which has a significant impact on the  $c$ -jet tagging efficiency.

## References

- [1] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102 (2011).  
<https://cds.cern.ch/record/1369219>.
- [2] ATLAS Collaboration, *b-jet tagging calibration on c-jets containing  $D^{*+}$  mesons*, ATLAS-CONF-2012-039 (2012). <http://cdsweb.cern.ch/record/1435193>.
- [3] CDF Collaboration, T. Aaltonen et al., *First Measurement of the Production of a W Boson in Association with a Single Charm Quark in  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.96$  TeV*, Phys. Rev. Lett. **100** (2008) 091803, [arXiv:0711.2901](https://arxiv.org/abs/0711.2901) [hep-ex].
- [4] CDF Collaboration, T. Aaltonen et al., *Observation of the Production of a W Boson in Association with a Single Charm Quark*, Phys. Rev. Lett. **110** (2013) 071801, [arXiv:1209.1921](https://arxiv.org/abs/1209.1921) [hep-ex].
- [5] D0 Collaboration, V. Abazov et al., *Measurement of the ratio of the  $p\bar{p} \rightarrow W+c$ -jet cross section to the inclusive  $p\bar{p} \rightarrow W+jets$  cross section*, Phys. Lett. **B666** (2008), [arXiv:0803.2259](https://arxiv.org/abs/0803.2259) [hep-ex].
- [6] CMS Collaboration, *Measurement of associated W + charm production in pp collisions at  $\sqrt{s} = 7$  TeV*, [arXiv:1310.1138](https://arxiv.org/abs/1310.1138) [hep-ex].
- [7] ATLAS Collaboration, *Measurement of the production of a W bosons in association with a charm hadron in pp collisions at  $\sqrt{s} = 7$  TeV*, ATLAS-CONF-2013-045 (2013).  
<https://cdsweb.cern.ch/record/1546800>.
- [8] ATLAS Collaboration, *Improved luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C73** (2013) 2518, [arXiv:1302.4393](https://arxiv.org/abs/1302.4393) [hep-ex].
- [9] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307** (2003) 001, [arXiv:hep-ph/0206293](https://arxiv.org/abs/hep-ph/0206293) [hep-ph].
- [10] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175) [hep-ph].
- [11] G. Corcella et al., *HERWIG 6.5 release note*, CERN-TH-2002-270 (2002), [arXiv:hep-ph/0210213](https://arxiv.org/abs/hep-ph/0210213) [hep-ph].
- [12] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. **C72** (1996) 637–646, [arXiv:hep-ph/9601371](https://arxiv.org/abs/hep-ph/9601371) [hep-ph].
- [13] M. L. Mangano, M. Moretti, and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions:  $Wb\bar{b} + n$  jets as a case study*, Nucl. Phys. **B632** (2002) 343–362, [arXiv:hep-ph/0108069](https://arxiv.org/abs/hep-ph/0108069) [hep-ph].
- [14] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **0207** (2002) 012, [arXiv:hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195) [hep-ph].
- [15] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152–155.

[16] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **0206** (2002) 029, arXiv:hep-ph/0204244 [hep-ph].

[17] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[18] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5*, arXiv:hep-ph/0405247 [hep-ph].

[19] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, *The tau decay library TAUOLA: Version 2.4*, Comput. Phys. Commun. **76** (1993) 361–380.

[20] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97–107, arXiv:hep-ph/0506026.

[21] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. **D69** (2004) 094008, arXiv:hep-ph/0312266 [hep-ph].

[22] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622, arXiv:1111.5869 [hep-ph].

[23] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, JHEP **1107** (2011) 018, arXiv:1105.0020 [hep-ph].

[24] J. M. Campbell, R. K. Ellis, and F. Tramontano, *Single top production and decay at next-to-leading order*, Phys. Rev. **D70** (2004) 094012, arXiv:hep-ph/0408158 [hep-ph].

[25] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

[26] GEANT4 Collaboration, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[27] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. **C72** (2012) 1909, arXiv:1110.3174 [hep-ex].

[28] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at  $\sqrt{s} = 7$  TeV*, ATLAS-CONF-2012-101 (2012). <https://cds.cern.ch/record/1463915>.

[29] M. Cacciari and G. P. Salam, *Dispelling the  $N^3$  myth for the  $k_t$  jet-finder*, Phys. Lett. **B641** (2006) 57–61, arXiv:hep-ph/0512210 [hep-ph].

[30] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- $k(t)$  jet clustering algorithm*, JHEP **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[31] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002 (2008). <https://cds.cern.ch/record/1099735>.

[32] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV*, Eur. Phys. J. **C73** (2013) 2304, arXiv:1112.6426 [hep-ex].

- [33] ATLAS Collaboration, *Muon Performance in Minimum Bias pp Collision Data at  $\sqrt{s} = 7 \text{ TeV}$  with ATLAS*, ATLAS-CONF-2010-036 (2010). <https://cds.cern.ch/record/1277675>.
- [34] ATLAS Collaboration, *Muon Reconstruction Performance*, ATLAS-CONF-2010-064 (2010). <https://cds.cern.ch/record/1281339>.
- [35] ATLAS Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, ATLAS-CONF-2011-063 (2011). <https://cdsweb.cern.ch/record/1345743>.
- [36] ATLAS Collaboration, *Measurement of the top quark pair production cross-section with ATLAS in pp collisions at  $\sqrt{s} = 7 \text{ TeV}$  in the single-lepton channel using semileptonic b decays*, ATLAS-CONF-2012-131 (2012). <https://cds.cern.ch/record/1478370>.
- [37] ATLAS Collaboration, *Calibrating the b-Tag Efficiency and Mistag Rate in  $35 \text{ pb}^{-1}$  of Data with the ATLAS Detector*, ATLAS-CONF-2011-089 (2011). <https://cds.cern.ch/record/1356198>.
- [38] E. Lohrmann, *A Summary of Charm Hadron Production Fractions*, arXiv:1112.3757 [hep-ex].
- [39] J. Beringer et al., *Review of Particle Physics*, Phys. Rev. **D86** (2012) 010001. <http://pdg.lbl.gov/>.
- [40] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at  $\sqrt(s) = 7 \text{ TeV}$  with ATLAS 2011 data*, ATLAS-CONF-2013-004 (2013). <https://cdsweb.cern.ch/record/1509552>.