Radiating top quarks

Gosselink, M.

Publication date
2010

6 Top quark pairs with additional jets

In this chapter a study on Monte Carlo simulated $t\bar{t}$ + n-jets events is presented. The aim of the study is (i) to understand the effect of $t\bar{t}$ analysis cuts on signal and background processes, (ii) to estimate the expected jet multiplicity for the ATLAS experiment, and (iii) to compare results between the different Monte Carlo generators MC@NLO, AcerMC, and Alpgen. The three generators use different techniques (explained in Chapter 2) and therefore differences in the jet distributions are expected. The comparison mayb be used to estimate the systematic uncertainty due to the Monte Carlo generator in the $t\bar{t}$ cross section determination as explained in Chapter 5.

6.1 Event selection

The same event selection criteria are used as in Section 5.3 of Chapter 5. First the effect of these criteria on QCD multi-jet background is estimated, and then the effect of adding b-tagging to the requirements is investigated.

6.1.1 QCD multi-jet background

To study QCD multi-jet background, samples are generated with Alpgen and Herwig/Jimmy. Compared to $t\bar{t}$, the cross section for this QCD multi-jet sample is very large: 18.9 μb. The total sample contains 1.67 million unscaled events[1]. The production of such large samples with full detector response simulation is problematic due to the required computing resources. Therefore the detector simulation is based on ATLFAST. Since the trigger information is absent in the simulation, the trigger efficiencies are assumed to be 100%. The relatively low number of events compared to the cross section means that also the integrated luminosity of the sample is low: $\int \mathcal{L}dt = N_{\text{evt}}/\sigma$ ranges from 0.5 to 9.1 pb^{-1} for the subsamples. Because all numbers are eventually scaled to 100 pb^{-1} in the analysis, the statistical uncertainties are large.

In Figure 6.1 the transverse momentum distributions of electrons, muons, and jets and the $\not{E}_T$ distribution are shown before event selection. In all plots the vertical dashed

[1]The total sample consists of four subsamples with different parton multiplicities weighted by the individual luminosities of the samples according to the MLM matching procedure, see Appendix A.

lines with the arrows on top indicate the cuts applied during event selection in the analysis. The selection of the transverse momentum of the electrons and muons remove many events. Most electrons and muons in QCD multi-jet events are relatively soft as can be seen from the rapidly decreasing number of leptons with increasing transverse momentum. The $\not\!\!E_T$ distribution shows a similar behaviour. It falls off steeply.

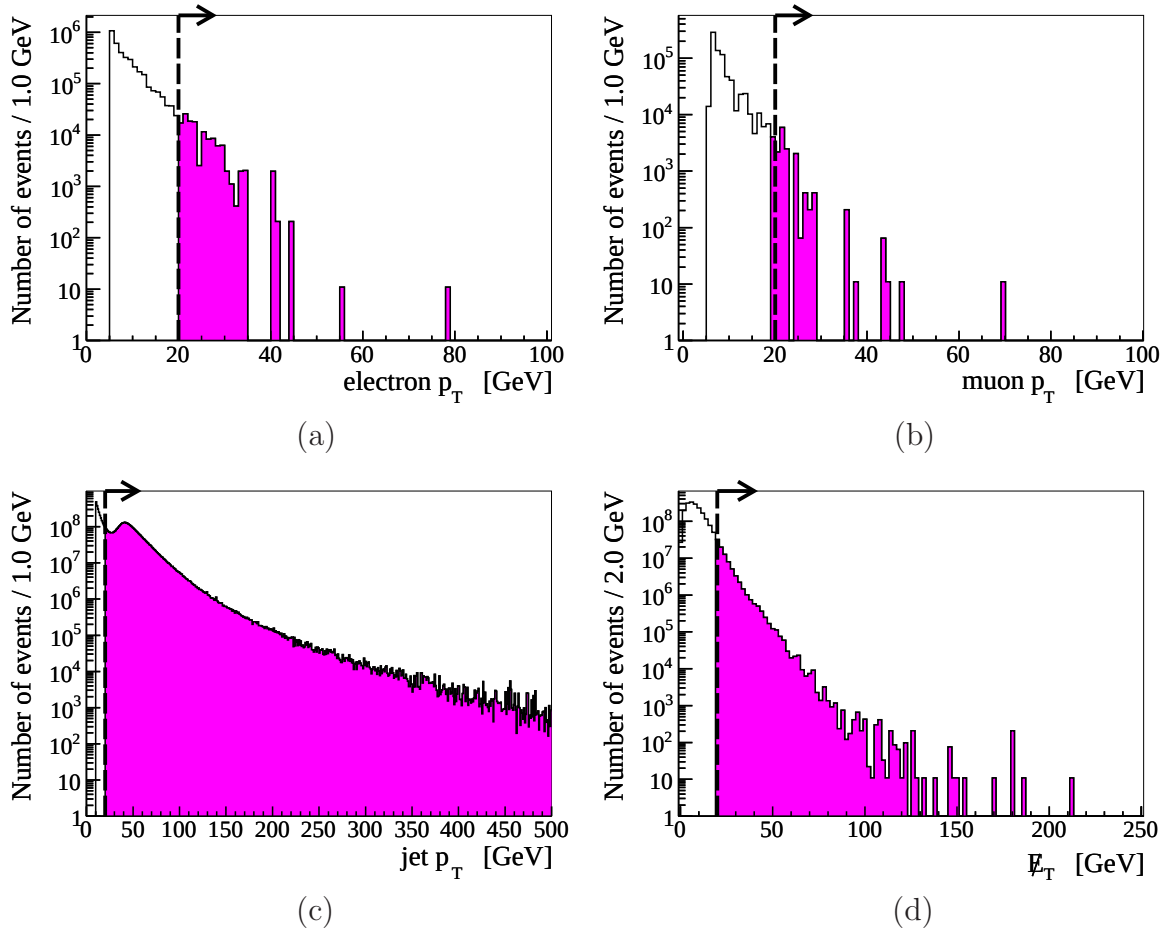


Figure 6.1: *The p_T distribution of (a) electrons, (b) muons, and (c) jets in QCD multi-jet events, and (d) the $\not\!\!E_T$ distribution before event selection. Normalised to 100 pb^{-1}.*

6

The jet transverse momentum distribution exhibits a remarkable dip around 27 GeV and a peak at 40 GeV. These features are due to Alpgen's MLM matching procedure. The minimum transverse momentum of partons generated with Alpgen in the hard subprocess is 40 GeV, while softer partons are generated only via the parton shower of Herwig. Effectively, this is a generator cut and results in a QCD multi-jet sample which gives only valid predictions for events with at least two jets, both with a minimum transverse momentum of approximately 40 GeV.

The number of events that pass the selection criteria are given in Table 6.1. For the $t\bar{t}(e)$ and $t\bar{t}(\mu)$ these numbers are in agreement with the efficiencies (without trigger

requirement) given in Table 5.1 in Chapter 5. The numbers differ slightly from the ones mentioned in Table 5.2 because in this chapter a branching ratio of $1/9^{\text{th}}$ was used for $W \to \ell\nu_\ell$ decay instead of the 10.8% in Chapter 5.

sample	initial	ℓ_{iso}	$\not\!\!E_T$	4j20	3j40	selected
$t\bar{t}(e)$	12,334	6,411	11,227	8,723	7,639	2,722
QCD	1.89×10^9	128×10^3	71×10^6	72×10^6	46×10^6	2×10^3
$t\bar{t}(\mu)$	12,341	8,477	11,301	8,079	7,066	3,677
QCD	1.89×10^9	14×10^3	71×10^6	72×10^6	46×10^6	44

Table 6.1: *Number of events that pass the separate cuts in the electron channel (top) and muon channel (bottom). Normalised to 100 pb^{-1}.*

Only a small fraction of the QCD multi-jet background passes all selection criteria. Most of the events are rejected by the isolated lepton requirement. Remember that the statistical uncertainties on these numbers are considerable due the limited amount of Monte Carlo events in combination with the large event weights. A discrepancy in the amount of QCD background in the electron channel with respect to the muon channel is however expected because the electron reconstruction efficiency is not included in the fast detector simulation [163]. Note that the number of events that pass the two different jet selection criteria are biased towards higher values due to the generator level cut.

6.1.2 Including b-tagging

Only a fraction of the background from W + jets and QCD multi-jet contains jets originating from b quarks[2]. On the other hand, each $t\bar{t}$ event typically has two such b-jets. The identification of these b-jets, called b-tagging, should therefore improve the $t\bar{t}$ purity. In this section the effect of b-tagging on the event selection is investigated.

Two distinct b-tagging requirements are considered here: at least one jet with a b-tag and at least two jets with a b-tag. A jet is b-tagged if it is a 'good' jet with a b-tag weight above 7.05. The b-tag weight is determined from a combination of the impact parameter and the secondary vertex tagger (`IP3D+SV1`). The former gives a weight according to the longitudinal and transverse distance of tracks inside a jet to the primary vertex, while the latter gives a weight according to the invariant mass, energy fraction, and number of two-track vertices of all tracks which could indicate a secondary vertex. The value of the weight (7.05) is the one agreed on in the ATLAS top quark physics working group [164] and corresponds to an average b-tagging efficiency of 60% for b-jets with a transverse momentum larger than 30 GeV, and a rejection factor for light jets of 154.

[2]From Table A.4 of Appendix A: the cross section for (filtered) W + jets events is ∼900 pb, while for $Wb\bar{b}$ + jets this is ∼20 pb. Furthermore, the (light) jet cross section is $\mathcal{O}$(mb) while the $b\bar{b}$ cross section is $\mathcal{O}$(μb).

Note that the QCD multi-jet samples only describe the production of 'light' jets (from u, d, s, and c quarks and gluons) in the hard scattering, not the production of $b\bar{b}$ + n-jets. In addition, the b-tagging implemented in fast detector simulation differs from full detector simulation: the b-tagging efficiency is a fixed parameter (at 60%) and is independent of the jet kinematics.

In Figure 6.2 the number of jets with a b-tag in the events that passed the event selection is shown. It demonstrates that, although requiring at least one or two jets with a b-tag will result in less signal events, the purity will indeed improve.

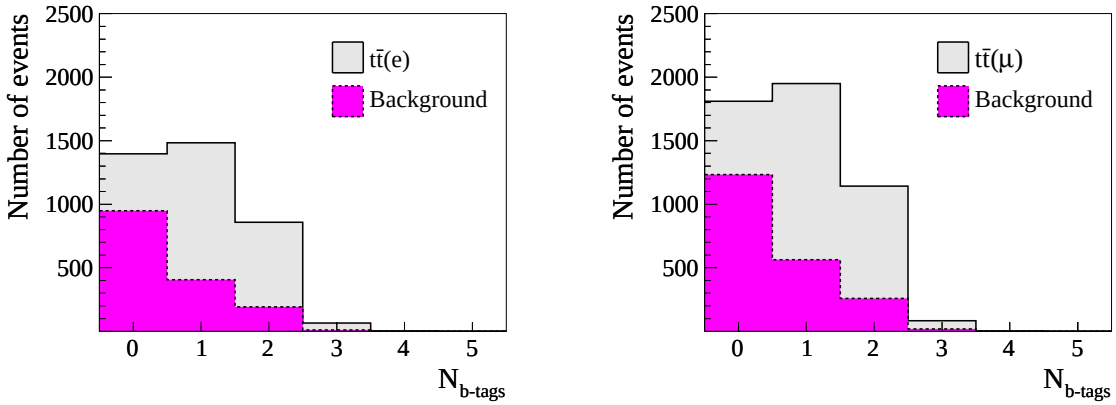


Figure 6.2: *Number of jets with a b-tag after the event selection in the electron channel (left) and muon channel (right). Normalised to 100 pb^{-1}.*

Selection efficiency

The number of events that pass the different b-tagging requirements are given in Table 6.2. As expected, the additional b-tagging reduces contribution from background to the selection significantly, especially from non-$t\bar{t}$. The largest contribution left is from dileptonic decaying $t\bar{t}$. Reason is the presence of large missing transverse energy, a lepton, and two b-jets. Since exactly one isolated lepton is required and di-leptonic $t\bar{t}$ has typically two leptons, this means that often one of the two leptons is not isolated or is outside the detector acceptance. Furthermore, although in this channel almost always two b-jets from the decay of top quarks are present, the fraction of events passing the b-tagging cut is lower than for fully hadronic and semi-leptonic $t\bar{t}$. A part of the explanation for this lies in the fact that dileptonic $t\bar{t}$ events contain less jets per event which can be mistagged.

The selection efficiencies for the two different b-tagging requirements individually (ϵ_{1b} and ϵ_{2b}), and the effect of the two b-tagging requirements in combination with the other analysis cuts on the overall selection efficiency ϵ_{sel}, are given in Table 6.3. Note that the individual selection efficiencies for the other analysis cuts and the trigger requirements remain the same as in Table 5.1.

sample	initial $\geq 0\ b_{\mathbf{tag}}$	selected e	selected μ	initial $\geq 1\ b_{\mathbf{tag}}$	selected e	selected μ	initial $\geq 2\ b_{\mathbf{tag}}$	selected e	selected μ
$t\bar{t}(e)$	12,334	2,248	7	8,896	1,800	5	2,968	722	1
$t\bar{t}(\mu)$	12,341	3	2,915	8,942	2	2,338	2,985	0	950
$t\bar{t}(\tau)$	12,411	172	249	9,016	135	200	3,046	54	79
$t\bar{t}(\ell^+\ell^-)$	9,190	276	354	6,487	219	287	2,018	85	113
$t\bar{t}$(jets)	37,022	11	42	27,302	6	32	9,803	0	13
single top	11,130	183	230	6,190	129	164	1,045	45	54
W + jets	91,157	742	1,041	4,773	70	100	146	5	4
Z + jets	64,131	113	79	1,243	10	7	17	0	0
$Wb\bar{b}$ + jets	5,119	46	59	2,517	36	44	550	14	14
WW	3,847	7	9	138	1	1	3	0	0
WZ	1,474	4	5	151	1	1	25	0	0
ZZ	271	1	0	52	0	0	11	0	0
QCD	$1.89 \cdot 10^9$	$2 \cdot 10^3$	44	$79 \cdot 10^6$	$2 \cdot 10^3$	22	$2.6 \cdot 10^6$	0	11
Signal	–	2,248	2,915	–	1,800	2,338	–	722	950
Background	–	1,557	2,075	–	609	842	–	203	279
S/B	–	1.4	1.4	–	3.0	2.8	–	3.6	3.4
$S/B_{\text{incl. QCD}}$	–	0.6	1.4	–	0.7	2.7	–	3.6	3.3

Table 6.2: *Number of events passing the analysis cuts, including the trigger requirement, with and without additional b-tagging in the electron and muon channel. Normalised to 100 pb^{-1}.*

The efficiencies to select events with b-tagged jets are nearly equal for $t\bar{t}(e)$ and $t\bar{t}(\mu)$. When requiring at least one jet that is b-tagged, the overall efficiency decreases with $\sim 20\%$ (compare Table 5.1). Though it drops almost with a factor three when demanding at least two jets that are b-tagged instead of one. From the selection efficiencies ϵ_{1b} and ϵ_{2b} a rough estimate of the average b-tagging efficiency (ϵ_{tag}) can be obtained. Since for $t\bar{t}$ events with two taggable b-jets the one or more b-tag selection efficiency ϵ_{1b} is approximately $1 - \epsilon^2_{\text{no-tag}} = 1 - (1 - \epsilon_{\text{tag}})^2$ and the two or more b-tag selection efficiency ϵ_{2b} is approximately ϵ^2_{tag}, it follows that the b-tagging efficiency is almost 50%. This is lower than the 60% b-tagging efficiency quoted earlier because a minimum transverse momentum for b-jets of 20 GeV was required instead of 30 GeV.

sample	$\geq$**1** $b_{\mathbf{tag}}$		$\geq$ **2** $b_{\mathbf{tag}}$	
	$\boldsymbol{\epsilon}_{1b}$	$\boldsymbol{\epsilon}_{\mathbf{sel}}$	$\boldsymbol{\epsilon}_{2b}$	$\boldsymbol{\epsilon}_{\mathbf{sel}}$
$t\bar{t}(e)$	72.1	14.6	24.1	5.9
$t\bar{t}(\mu)$	72.5	18.9	24.2	7.7

Table 6.3: *The selection efficiencies (in %) of the two b-tag requirements and their corresponding overall selection efficiencies for the $t\bar{t}(e)$ and $t\bar{t}(\mu)$ channel.*

Selection purity

The purity of the sample with $t\bar{t}$ candidates satisfying the selection and trigger requirements is calculated for the electron and muon channel in Table 6.4. The purity $\mathcal{P}$ is defined as:

$$\mathcal{P} \equiv \frac{N_{\text{sig}}}{N_{\text{sig}} + N_{\text{bkg}}} \tag{6.1}$$

where N_{sig} is the number of signal $t\bar{t}$ events and N_{bkg} the number of background events. Background from QCD multi-jet is disregarded because the uncertainty on the number of events remaining after event selection is too large to give sensible predictions. The numbers in each column indicate the purity of the sample after event selection for the three different scenarios: without b-tagging, with at least one jet that is b-tagged, and with at least two jets that are b-tagged. Note that it is only meaningful to quote the purities after all selection cuts since the various background samples contain generator level cuts.

sample	$\boldsymbol{\mathcal{P}}$		
	$\geq 0\ b_{\text{tag}}$	$\geq 1\ b_{\text{tag}}$	$\geq 2\ b_{\text{tag}}$
$t\bar{t}(e)$	59.6	75.4	78.8
$t\bar{t}(\mu)$	58.9	74.2	77.9

Table 6.4: *Purity (in %) of the $t\bar{t}(e)$ and $t\bar{t}(\mu)$ signal in the electron channel and muon channel respectively after event selection.*

6.2 Jet multiplicity

The characteristic jet multiplicity of semi-leptonic $t\bar{t}$ events and background after event selection, with and without using b-tagging, is determined. Then, a comparison is made between predictions with full and fast detector response simulation. Finally, results from three generators (MC@NLO, AcerMC, and Alpgen) are compared.

6.2.1 Jet spectrum

The predicted number of events in the electron and muon channel per jet multiplicity is given in Table 6.5. The negative numbers in the $N_{\text{jet}} = 11$ and 12 bins are a consequence of the usage of positive and negative event weights in MC@NLO in combination with statistical fluctuations due to the limited available amount of simulated events. It indicates that the expected number of events in these bins are close to zero. In Figure 6.3 the expected jet multiplicity spectra are shown for the electron and muon channel after the $t\bar{t}$ analysis cuts without b-tagging, at least one jet with b-tag, and at least two jets with a b-tag.

electron channel

N_{jet}	≥ 0 b_{tag}		≥ 1 b_{tag}		≥ 2 b_{tag}	
	$t\bar{t}(e)$	bkg	$t\bar{t}(e)$	bkg	$t\bar{t}(e)$	bkg
4	1,002.1	932.4	796.9	340.2	305.2	104.0
5	768.6	433.7	616.7	174.8	250.7	62.8
6	340.1	139.0	274.6	61.9	115.4	23.6
7	107.3	40.7	86.5	24.2	38.9	10.2
8	25.0	8.4	21.1	4.7	8.9	1.4
9	4.6	1.7	4.3	1.6	2.5	0.6
10	1.1	0.7	0.6	0.5	0.2	0.0
11	-0.6	0.0	-0.3	0.1	0.0	0.1
12	-0.2	0.1	-0.2	0.1	0.0	0.0
sum	2,248.0	1,556.8	1,800.2	608.6	721.8	202.7

muon channel

N_{jet}	≥ 0 b_{tag}		≥ 1 b_{tag}		≥ 2 b_{tag}	
	$t\bar{t}(\mu)$	bkg	$t\bar{t}(\mu)$	bkg	$t\bar{t}(\mu)$	bkg
4	1,267.4	1,176.2	995.2	421.7	386.3	134.9
5	998.7	599.3	812.6	267.1	336.2	84.5
6	457.3	217.6	369.8	105.9	158.0	42.1
7	146.0	62.4	121.6	36.4	53.5	13.7
8	34.9	14.7	29.0	8.1	11.3	2.8
9	7.8	3.0	6.6	1.6	3.4	0.6
10	2.3	1.6	2.7	0.8	1.3	0.2
11	0.5	0.1	0.1	0.1	0.0	0.0
12	0.1	0.0	0.1	0.0	0.1	0.0
sum	2,914.7	2,075.0	2,337.7	841.8	950.0	278.9

Table 6.5: *Expected number of events in the electron channel (top) and muon channel (bottom) for each jet multiplicity after event selection. Normalised to 100 pb^{-1} of data.*

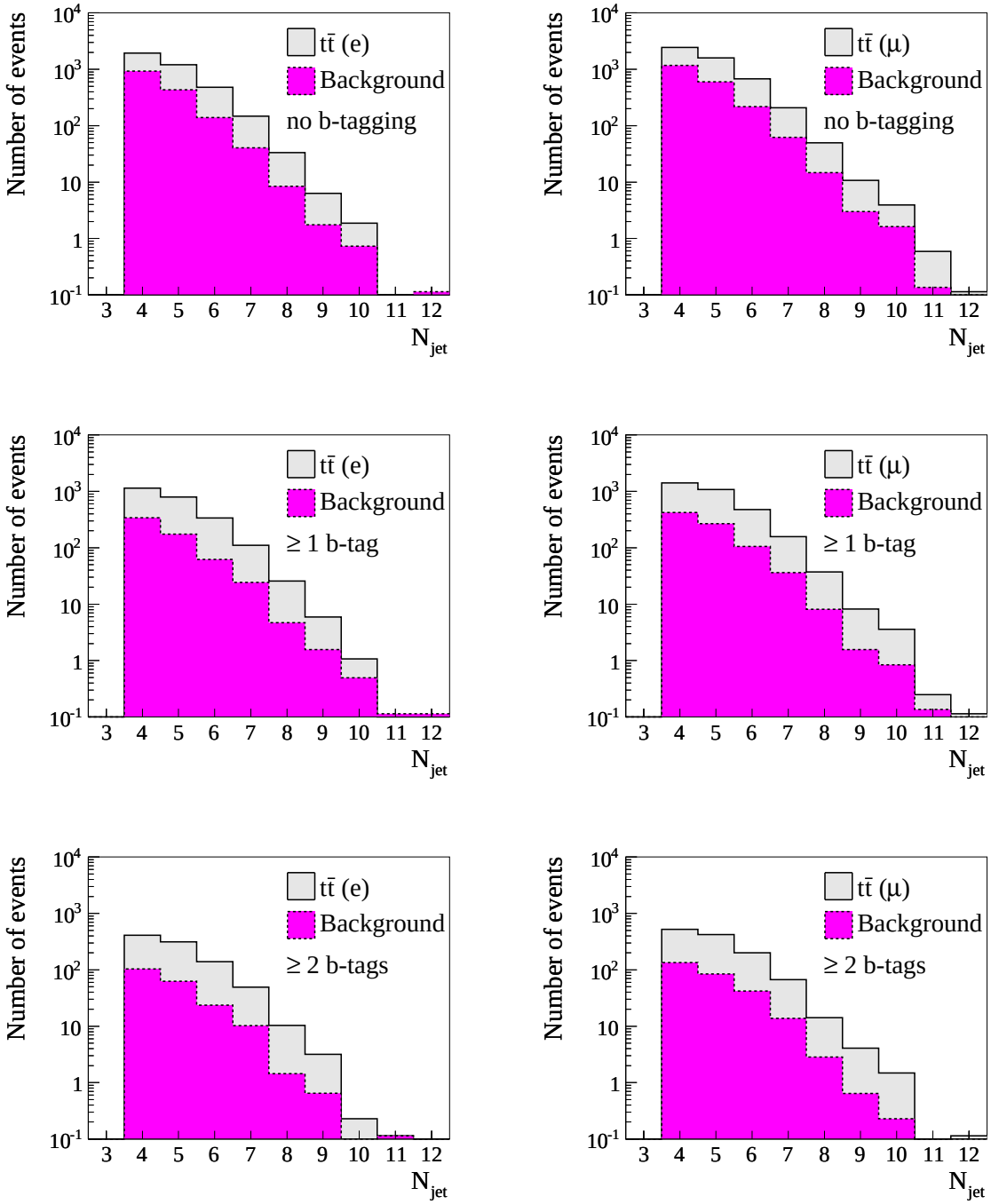
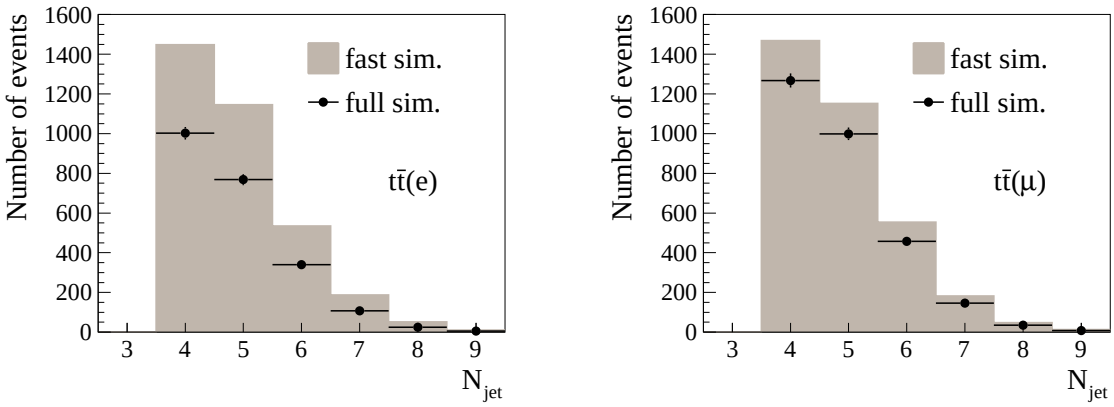


Figure 6.3: *Expected jet multiplicity distribution after event selection without b-tagging (top), at least on b-tagged jet (middle), and at least two b-tagged jets (bottom), for the the electron channel (left) and the muon channel (right) with 100 pb^{-1} of data.*

6.2.2 Fast versus full simulation

The results in the previous section are obtained with full detector simulation. The comparison between the three event generators in the next section is however performed with fast detector response simulation. To gain insight in the differences between fast and full simulation, the predictions of MC@NLO for the jet multiplicity using both types of detector simulation are compared first. The results are shown in Figure 6.4.



(a) (b)

Figure 6.4: *Predicted number of (a) $t\bar{t}(e)$ events and (b) $t\bar{t}(\mu)$ events per jet multiplicity compared between fast and full simulation. Normalised to 100 pb^{-1} data.*

The normalisations of the jet multiplicity distributions differ considerably between the fast and full simulation predictions, especially in the electron channel. The shapes of the distributions are identical though. This can be seen from Table 6.6, where the fraction of events per jet multiplicity is given. In both types of simulation the predicted fractions for the electron and muon channels are consistent with each other. For either lepton channel, the event fractions of fast simulation tend towards a higher jet multiplicity than those of full simulation, although the trend is similar.

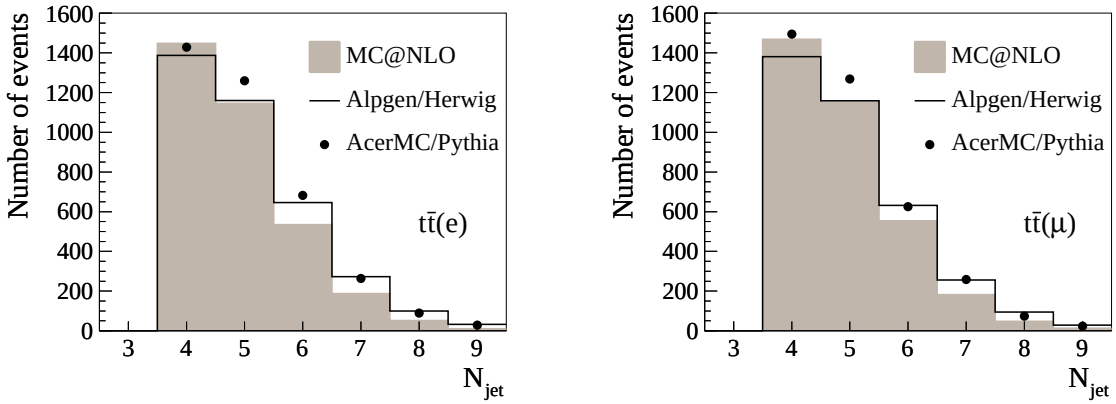
N_{jet}	$t\bar{t}(e)$		$t\bar{t}(\mu)$	
	full	**fast**	**full**	**fast**
4	44.6 ± 0.4	42.9 ± 0.3	43.5 ± 0.3	42.9 ± 0.3
5	34.2 ± 0.3	33.9 ± 0.3	34.3 ± 0.3	33.7 ± 0.3
6	15.1 ± 0.3	15.8 ± 0.2	15.7 ± 0.2	16.2 ± 0.2
7	4.8 ± 0.2	5.6 ± 0.1	5.0 ± 0.1	5.3 ± 0.1
8	1.1 ± 0.1	1.5 ± 0.1	1.2 ± 0.1	1.4 ± 0.1
≥ 9	< 0.3	< 0.4	< 0.4	< 0.5

Table 6.6: *Predicted fractions (in %) of $t\bar{t}$ signal events per jet multiplicity in the electron and muon channel using fast and full simulation.*

The discrepancies in the predictions arise from the following differences between the two types of simulation. In fast simulation the trigger is absent and the detector geometry is not described with great precision. Particles are reconstructed from the Monte Carlo truth list and not from simulated detector signals as in full simulation. Besides, the interactions of particles with detector material are not fully taken into account. This leads to an overestimation of the reconstruction efficiencies and, eventually, the selection efficiencies. In Section 6.3 the selection efficiencies will be further. Finally, the cone algorithm used for jet reconstruction in fast simulation is also different from the one in full simulation: the split and merge step is not included in the former case (Section 3.2.2).

6.2.3 Event generator comparison

In Figure 6.5 the predictions for the jet multiplicity in $t\bar{t}$ events after event selection are compared between MC@NLO, Alpgen, and AcerMC. All three predictions are based on fast detector simulation.



(a) (b)

Figure 6.5: *Comparison of the jet multiplicity in (a) $t\bar{t}(e)$ events and (b) $t\bar{t}(\mu)$ events after event selection as predicted by MC@NLO, Alpgen, and AcerMC. Normalised to 100 pb^{-1}.*

6

The spectra in the two lepton channels show a similar behaviour: AcerMC predicts more events to pass the event selection (overall larger number of events) and the spectra from MC@NLO and Alpgen cross over at $N_{\text{jet}} = 5$. MC@NLO predicts more 4-jet events while Alpgen expects more events with higher jet multiplicity. This behaviour can also be seen in Table 6.7 where the fraction of events per jet multiplicity is given.

The absolute differences between the fractions predicted by each generator are not very large. Especially the numbers of AcerMC and Alpgen are close to each other. Both expect less 4-jet events than MC@NLO and more higher jet multiplicity events. The agreement is remarkable, since the two generators are quite different. Differences with MC@NLO are larger than the statistical errors and might be significant.

	MC@NLO	Alpgen	AcerMC	stat.	MC@NLO	Alpgen	AcerMC	stat.
N_{jet}	$t\bar{t}(e)$				$t\bar{t}(\mu)$			
4	42.9	38.4	38.0	$\pm$ 0.4	42.9	38.7	39.9	$\pm$ 0.4
5	33.9	32.1	33.5	$\pm$ 0.4	33.7	32.5	33.8	$\pm$ 0.4
6	15.8	17.9	18.1	$\pm$ 0.3	16.2	17.7	16.7	$\pm$ 0.3
7	5.6	7.5	7.0	$\pm$ 0.2	5.3	7.2	6.9	$\pm$ 0.2
8	1.5	2.7	2.4	$\pm$ 0.1	1.4	2.7	2.0	$\pm$ 0.1
≥ 9	< 0.4	< 1.4	< 1.1	95%	< 0.5	< 1.3	< 0.9	95%

Table 6.7: *Predicted fractions of $t\bar{t}$ events (in %) per jet multiplicity for the electron channel (left) and muon channel (right) using three different event generators.*

6.3 $t\bar{t}$ cross section measurement

In the previous section it was pointed out that there are small differences in the jet spectra predicted by the three event generators. In this section the effect on the $t\bar{t}$ cross section measurement, presented in Chapter 5, will be studied.

6.3.1 Event selection

In Table 6.8 the selection efficiencies for the $t\bar{t}$ analysis cuts are given for the electron and muon channel. The numbers for the electron channel are very similar to those for the muon channel. (The numbers for full simulation are slightly different, especially for the single isolated lepton requirement, compared Table 5.1). Between the three event generators there are small differences. The values of the selection efficiencies for the two separate jet requirements (4j20 and 3j40) are higher for both ALPGEN and ACERMC compared to MC@NLO. This is purely due to the difference in jet spectrum. As was seen in the previous section: ALPGEN and ACERMC predict a larger fraction of high jet multiplicity events. These events are exactly the events that pass the two distinct jet requirements more easily, leading to the higher overall selection efficiency. These differences are larger than the statistical uncertainties, and the same for both channels. The efficiencies of the single isolated lepton and $\not\!\!E_T$ cuts are similar, as expected.

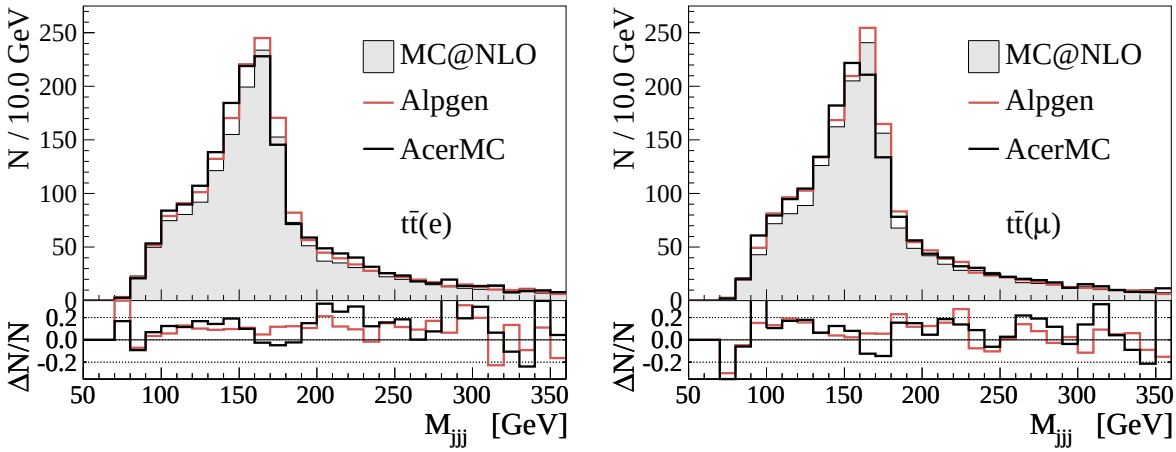
6.3.2 Mass reconstruction

In the $t\bar{t}$ analysis, the hadronic top mass is reconstructed by selecting from all jets in an event the three-jet combination which has the highest vector summed p_T. In addition, the invariant mass M_{jj} of at least one di-jet combination from these three jets is required to be within 10 GeV of the W boson mass M_W. Figure 6.6 shows the invariant three-jet

	Generator	ℓ_{iso}	$\not\!\!E_T$	4j20	3j40	sel	stat
$t\bar{t}(e)$	MC@NLO	66.6	90.8	64.2	56.3	27.4	± 0.1
	ALPGEN	67.4	91.0	67.2	58.5	29.3	± 0.2
	ACERMC	67.4	90.8	68.1	59.5	30.5	± 0.2
$t\bar{t}(\mu)$	MC@NLO	66.8	90.9	65.7	57.6	27.7	± 0.1
	ALPGEN	66.2	91.3	68.2	59.3	28.9	± 0.2
	ACERMC	67.4	91.7	68.8	60.8	30.4	± 0.2

Table 6.8: *Selection efficiencies (in %) for $t\bar{t}$ events in the electron channel (top) and muon channel (bottom) predicted by the three event generators.*

mass distribution, M_{jjj}, of all events that satisfy this M_W constraint.



(a) (b)

Figure 6.6: *Invariant mass distribution of the three-jet combinations which have at least one di-jet M_{jj} close to M_W in $t\bar{t}$ events that passed the selection cuts in the (a) electron channel and (b) muon channel. Normalised to 100 pb^{-1}.*

The distributions have similar features. The correctly reconstructed $t\bar{t}$ events are distributed like a Gaussian around a mass of 165 GeV. The peak is on top of the combinatorial background. However, the combinatorial background in the ACERMC distribution is overall larger than for MC@NLO, while the peak is smaller. This can be observed best in the ratio plots below the M_{jjj}-distributions, which indicate the relative difference for ACERMC and ALPGEN with respect to MC@NLO. The ALPGEN distributions are in both the electron and muon channel overall higher than MC@NLO predictions.

The subtle differences are also reflected quantitatively in the selection, reconstruction, and combined efficiency in Table 6.9. The reconstruction efficiency is measured by determining from Monte Carlo truth information how often a three-jet combination is within $\Delta R < 0.2$ of the four-vector of the top quark. The combined efficiency is the

product of the selection and reconstruction efficiencies.

	Generator	ϵ_{sel}	ϵ_{reco}	ϵ_{comb}
$t\bar{t}(e)$	MC@NLO			
	– full sim.	18.2 ± 0.1	17.8 ± 0.3	3.3 ± 0.1
	– fast sim.	27.4 ± 0.1	19.1 ± 0.2	5.2 ± 0.1
	ALPGEN	29.3 ± 0.2	19.8 ± 0.3	5.8 ± 0.1
	ACERMC	30.5 ± 0.2	18.2 ± 0.3	5.6 ± 0.1
$t\bar{t}(\mu)$	MC@NLO			
	– full sim.	23.6 ± 0.1	18.0 ± 0.2	4.3 ± 0.1
	– fast sim.	27.7 ± 0.1	19.3 ± 0.2	5.4 ± 0.1
	ALPGEN	28.9 ± 0.2	19.4 ± 0.3	5.6 ± 0.1
	ACERMC	30.4 ± 0.2	17.5 ± 0.3	5.3 ± 0.1

Table 6.9: *The selection, reconstruction, and combined efficiencies (in %) for $t\bar{t}$ events in the electron channel (top) and muon channel (bottom) as predicted by the three generators using fast simulation. For comparison the values from MC@NLO with full simulation are also given. Normalised to 100 pb^{-1}.*

The selection and combined efficiencies for ALPGEN are higher than for MC@NLO. The reconstruction efficiency is more or less the same within errors. This explains the overall larger number of events in the ALPGEN distributions in the two M_{jjj} plots. Remarkable is that the combined efficiencies of ACERMC and MC@NLO are equal, while the selection and reconstruction efficiencies differ significantly. Hence ACERMC predicts a larger combinatorial background and a relatively lower number of correctly reconstructed events, though the absolute number of reconstructed events is equal. Note that this compensation between the selection and reconstruction efficiency was also observed for jet energy scale variations discussed in Section 5.6.6: with a larger jet multiplicity it is more difficult to trace back the three (or more) jets associated with the hadronic top quark.

ACERMC relies on the parton showering of PYTHIA, thus the predicted amount of jet activity depends strongly on the initial state and final state radiation parameters of PYTHIA. The effect of varying these parameters, as is done to estimate the systematic uncertainty for the top cross section measurement in Chapter 5, is studied in more detail in Appendix B.2.

In Table 6.10 the selection efficiencies, reconstruction efficiencies, and combined efficiencies are specified per jet multiplicity bin as predicted by the three generators for semi-leptonic $t\bar{t}(e)$ and $t\bar{t}(\mu)$ events together. The electron channel and muon channel have been combined to reduce the statistical uncertainties on the calculated efficiencies. Note that the *total* selection, reconstruction, and combined efficiencies (Table 6.9) follow from the sum of these individual efficiencies over all jet multiplicity bins, weighted with

N_{jet}	Generator	ϵ_{sel}	ϵ_{reco}	ϵ_{comb}
4	MC@NLO	38.2 ± 0.2	23.4 ± 0.3	8.9 ± 0.4
	Alpgen	37.7 ± 0.2	25.3 ± 0.3	9.5 ± 0.4
	AcerMC	39.4 ± 0.3	22.5 ± 0.4	8.9 ± 0.5
5	MC@NLO	45.3 ± 0.2	17.4 ± 0.3	7.9 ± 0.5
	Alpgen	45.3 ± 0.3	18.5 ± 0.3	8.4 ± 0.5
	AcerMC	47.0 ± 0.3	17.3 ± 0.4	8.1 ± 0.6
6	MC@NLO	48.3 ± 0.4	14.6 ± 0.4	7.1 ± 0.7
	Alpgen	49.3 ± 0.4	14.9 ± 0.4	7.3 ± 0.7
	AcerMC	49.9 ± 0.5	13.1 ± 0.4	6.5 ± 0.9

Table 6.10: *Comparison of the predicted selection efficiencies, reconstruction efficiencies, and combined efficiencies (in %) per jet multiplicity bin for semi-leptonic $t\bar{t}(e)$ and $t\bar{t}(\mu)$ events with four, five, or six jets.*

the fraction of events in each jet multiplicity bin. Hence, differences between the generator predictions in Table 6.9 arise from both differences in the expected jet multiplicities and, if any, differences in the expected efficiencies per jet multiplicity bin.

As can be seen from Table 6.10, also for the individual jet multiplicities AcerMC tends to predict larger selection efficiencies and lower reconstruction efficiencies than MC@NLO and Alpgen. This is remarkable, since it could indicate, for example, that in four-jet events, the extra jet activity in AcerMC compensates for the loss of a jet from the decay of the hadronic top quark due to the acceptance cuts. Although the correct three-jet combination can not be found in that case for the hadronic top mass reconstruction, the event still passes the event selection criteria.

For all three event generators, the decrease in reconstruction efficiency is less with increasing jet multiplicity than one would naively expect when three jets where picked randomly in the event. Because in that case the chance of picking the correct three jets from the hadronic top in a four, five, or six jet event is 25%, 10%, and 6% respectively.

6.4 Conclusions and discussion

The selection of semi-leptonic $t\bar{t}$ candidates in the $t\bar{t}(e)$ and $t\bar{t}(\mu)$ channel has been investigated. Although only a rough estimate of the QCD multi-jet background could be included, it was shown that only a small part of this background remains after event selection. Furthermore, it was demonstrated that including b-tagging requirements to the selection criteria, results in a significant purification of the $t\bar{t}$ signal.

In addition, the jet multiplicity spectra for $t\bar{t}$ candidates after event selection have

been determined. Two variations using b-tagging requirements are thereby also considered. Comparison between three different event generators (MC@NLO, ALPGEN, and ACERMC) shows that ALPGEN and ACERMC predict slightly higher jet multiplicities per event than MC@NLO, resulting in higher estimates for the selection efficiencies.

How this higher jet multiplicity affects the expected results for the $t\bar{t}$ cross section measurement, depends on the event generator: predictions for the reconstruction efficiencies by ALPGEN and MC@NLO give equal values within statistical errors. This results in a 10% (5%) higher combined efficiency (selection and reconstruction) in the electron (muon) channel than MC@NLO according to ALPGEN. On the other hand, the reconstruction efficiencies predicted by ACERMC are lower than for MC@NLO. These lower reconstruction efficiencies compensate the larger selection efficiencies expected by ACERMC. In the end, the combined efficiencies for ACERMC are close (within statistical errors) to the predictions by MC@NLO for both the electron and muon channels.

The differences in the predicted selection, reconstruction, and combined efficiencies give an indication of the systematic uncertainty on the $t\bar{t}$ cross section measurement associated with the Monte Carlo generators.