Monte Carlo Simulation of Top Quark pair production in dilepton channel

Summerstudent Report

Daniela Ramermann

Physics Department, Bielefeld University, Universitätsstrasse 25, D-33615 Bielefeld, Germany

Abstract: Monte Carlo simulations with three different event generators (Pythia 8, Cascade 2.3.14 and Powheg interfaced with Pythia) are made for recent CMS measurements of jet multiplicity of Top Quark events, evaluated in the dilepton channel. The Analysis is coded in Rivet and interfaced with each of the event generators.

1 Introduction

The Top quark is the heaviest of the quarks and the heaviest elementary particle so far ($m_t = 173 GeV$, almost as heavy as a gold atom. It is the only quark which decays before hadronization, so there are no decay products from bound states in the final state.

The Large Hadron Collider (LHC) with its current energy of 8 TeV is a "top factory", it produces a huge statistic of $t\bar{t}$ pairs, so also decay chanels with small branching ratio (BR) could be measured precisely. At this Energies the fraction of $t\bar{t}$ with additional hard jets is large ($\approx 50\%$) [1].

Also the $t\bar{t}$ process is here dominated by gluon-gluon fusion, so $t\bar{t}$ is a kind of benchmark for other standard model processes which share the production mechanism [3]. A better understanding of these processes is needed because in higher order QCD Calculations are also contributions from initial and final state radiation (ISR, FSR), so it is possible to test them. It is needed to get a good description of multijet processes, which are also background for new physics searches.

Region of interest is the dilepton channel, it is the decay channel with the smallest branching ratio $\approx 4.5\%$). Contributions are from ee, $\mu\mu$, $e\mu$ and from leptonic τ decays. It is the cleanest decay channel, since there are no additional hadronic decay products and low background.

There are 2 b-jets from top decay, the additional jets do not come from top decay but from ISR and FSR, and apart from this only leptons and missing transverse energy (mainly from neutrinos) in the final state.



Figure 1: Dileptonic decay of the $t\bar{t}$ -pair [5]

For parts of [1] (cross section, gap fraction) an analysis is written with Rivet (Robust Independent Validation of Experiment and Theory, http://rivet.hepforge.org/). Rivet is a framework for event generator testing and tuning and can be linked to various event generators. For comparison with data it is possible to get data from HepData records.

Trying to reproduce the results (also MC results!) from the paper is a cross check, for really being able to compare it is necessary to use one of the MC generators that were also used there. This will be Powheg, but unfortunately there were some problems, so correctly normalized Powheg results are only available for the cross section with "hard cuts".



Figure 2: The branching ratio for decay into dilepton channel is about 4.5%. [2]



Figure 3: Top quark pair production via gluon-gluon fusion and decay in dilepton channel

2 Generators

2.1 Pythia 8

Pythia 8 is a MC generator for high energy events, completely written in C++ and a standalone framework. Its initial- and final-state algorithms are based on p_{\perp} - ordered evolution. Initial-state evolution, multiple interactions and the final-state evolution are interleaved into one common decreasing p_{\perp} sequence [6].

This generator is leading-logarithm, its parton shower is based on the ordering of the shower outcome by transverse momentum $p_T[7]$. Next to leading order calculations are not taken into account.

2.2 Cascade

Cascade is a Hadron level MC generator for ep and pp scattering. It uses the CCFM evolution equation in a backward evolution approach for angular ordered parton showers, proton remnant and hadronization are done via a standard hadronization program like Pythia [8]. Also it agrees well with standard next to leading order calculations, where applicable [9].

2.3 Powheg

Powheg is a MC generator which implements next to leading order calculations in shower MC programs. It can be interfaced with all parton shower programs that support the Les Houches Interface. Here it is used interfaced with Pythia 6 for the parton Shower [10]. According to the POWHEG method, the hardest radiation is generated first, using exact next to leading order matrix elements. This output can be interfaced to any standard MC generator, which is either p_T ordered or allows implementation of a p_T -veto. Interfacing to angular ordered programs cannot guarantee the double-log accuracy of the whole result [11].

3 Event Selection

Region of interest is the dilepton Channel, where a produced $t\bar{t}$ -pair decays into $t \to b + W^+$ and $\bar{t} \to \bar{b} + W^-$, and the W bosons decay each into a charged lepton and a neutrino.

The reconstruction and selection of the data is described in more detail in [1]. Used were muons with transverse momentum $p_T > 20 \text{GeV}$ within the pseudorapidity region $|\eta| < 2.4$. They are required to have a relative isolation of $I_{rel} < 0.15$.

The Definition of I_{rel} is as follows: the sum of all transverse momenta of all neutral and charged reconstructed particle candidates, except the muon itself, inside an cone in $\eta - \phi$ space of $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} < 0.3$, divided by the muon transverse momentum [1]. Electrons are required to have a transverse energy $E_T > 20 \text{GeV}$, $|\eta| < 2.4$ and $I_{rel} < 0.15$. Jets

are selected if $|\eta| < 2.4$ and the transverse momentum is at least $p_T > 30$ GeV.

To avoid miscounting leptons as jets, "good" jets are the ones where no leading lepton is in a cone of 0.4 with respect to the jet. Only good jets are processed further, others rejected. A b-jet is defined to contain the decay products of a b-hadron [1].

For MC Simulation slightly different cuts than these "hard cuts" were used, the relative isolation of the leptons and the missing transverse energy was skipped, also the mass requirements for the lepton pair, only the kinematics in the final state were used to be able to compare with the experimental data and the corresponding MC simulations, since the same cuts were used there.

As accidentally first all the cuts were programmed in, it is possible to compare both simulations with data and with themselves.

	p_T	$ \eta $	additional Info
ee	$> 20 \mathrm{GeV}$	< 2.4	
$\mu\mu$	$> 20 \mathrm{GeV}$	< 2.4	
$e\mu$	$> 20 \mathrm{GeV}$	< 2.4	
b-jet	$> 30 \mathrm{GeV}$	< 2.4	ANTIKT algorithm $R = 0.5$,
			reject jet if leading leptons within a cone
			of 0.4 with respect to the jet

Table 1: selection criteria for the $t\bar{t}$ - dilepton channel \rightarrow "normal cuts"

	p_T	$ \eta $	dilepton mass	additional Info
ee	$> 20 \mathrm{GeV}$	< 2.4	$> 20 \mathrm{GeV}$	outside Z-boson mass window($\pm 15 GeV$)
$\mu\mu$	$> 20 \mathrm{GeV}$	< 2.4	$> 20 \mathrm{GeV}$	outside Z-boson mass window($\pm 15 GeV$)
$e\mu$	$> 20 \mathrm{GeV}$	< 2.4		
b-jet	$> 30 {\rm GeV}$	< 2.4		ANTIKT algorithm $R = 0.5$,
				reject jet if leading leptons within a cone
				of 0.4 with respect to the jet
E_{T}	$> 40 {\rm GeV}$			

Table 2: harder event selection criteria for the $t\bar{t}$ - dilepton channel (accidentally first also implemented) \rightarrow "hard cuts"

4 Results

4.1 Differential cross section as a function of jet multiplicity

The differential cross section as a function of jet multiplicity is defined as

$$\frac{d\sigma_{tt}}{dN} = \frac{N_{data}^i}{\Delta_x^i \mathcal{L}}$$

where N_{data}^i are the number of Events in data in bin i, Δ_x^i is the bin width and \mathcal{L} is the integrated luminosity.

Practically to do is counting how many jets (b-jets and additional jets) are there for each of the three p_T thresholds: 30GeV, 60GeV and 100GeV.

For comparison are here first the plots from the paper [1]:



4.1.1 Normal cuts



a) Differential cross section for $p_T > 30 GeV$ b) Differential cross section for $p_T > 60 GeV$



c) Differential cross section for $p_T > 100 GeV$

Results for the cross section: Cascade gives a quite reasonable description of the data, Pythia also but tends to overestimate the jet multiplicity. On the x-axis is the number of jets, the y-axis shows the differential cross section. In the last bin are also the higher jet multiplicities $\geq 6, \geq 4$, but plotting with the AIDA framework is somehow difficult...

4.1.2 Hard cuts



a) Differential cross section for $p_T > 30 GeV$ b) Differential cross section for $p_T > 60 GeV$ (hard cuts) (hard cuts)



c) Differential cross section for $p_T > 100 GeV$ (hard cuts)

Results for the cross section with the harder set of cuts: nearly the same result, here it was possible to include the Powheg MC, which gives a result very similar to Pythia 8. They share the p_T order in their mechanism, since Powheg was linked to Pythia 6. The description of the data is reasonably well. In the last bin are also the higher jet multiplicities ≥ 6 , ≥ 4 .



4.1.3 Comparison between normal and hard cuts

Left column: Comparison between normal	Right column: Comparison between nor-
and hard cuts for Pythia 8,	mal and hard cuts for Cascade,
$p_T > 30 GeV, p_T > 60 GeV$ and	$p_T > 30 GeV, p_T > 60 GeV$ and
$p_T > 100 GeV$. Pythia labeled are the hard	$p_T > 100 GeV$. Cascade labeled are the
cuts, Pythia0 labeled the normal cuts.	hard cuts, Cascade0 labeled the normal
The Differences are very small.	cuts. The Differences are also very small.

The differences between normal and hard cuts are very small, so there is good agreement between the whole event selection and the selection via kinematic cuts.

At last both Pythia 8 and both Cascade results are shown in comparison with data (Pythia, Cascade labeled are hard cuts, Pythia0, Cascade0 normal cuts):



a) Differential cross section for p_T 30 GeV, both cuts

> b) Differential cross section for $p_T > 60 GeV$, both cuts



c) Differential cross section for $p_T > 60 GeV$, both cuts

4.2 Gap Fraction

The gap fraction is defined as the fraction of events, which do not contain additional jets above a given threshold in p_T . So it is considering also initial- and final-state radiation and is therefore sensitive to jet activity arising from quark and gluon radiation. It is formally defined as

$$f(p_T) = \frac{N(p_T)}{N_{total}}$$

where $N(p_T)$: is the number of selected events that do not contain an additional jet above a chosen p_t value and N_{total} is the number of selected events.

Considered for gap fraction are the 1st additional jet (additional jet with highest p_T), the 2nd additional jet (additional jet with 2nd highest p_T) and the H_T . H_T is defined as the scalar sum of the transverse momentum of all additional jets in each selected event:

$$f(H_T) = \frac{N(H_T)}{N_{total}}$$

with $N(H_T)$ the number of events in which the sum of the scalar transverse momentum of the additional jets is less than a certain threshold [1].

For comparison are here first the plots from the paper [1]:



a) Gap fraction of the 1st additional jet [1] b) Gap fraction of the 2nd additional jet [1]



c) Gap fraction HT [1]

4.2.1 Normal cuts



 $400 \\ [p_T]$

c) Gap fraction HT

Cascade tends to overestimate the gap fraction, Pythia 8 underestimates it. Cascade describes the gap fraction of the 2nd additional jet quite good, the overall description is rather reasonable, "the truth seems to lie in between" the two different generation mechanisms (angular ordering of Cascade and p_T ordering of Pythia).





a) Gap fraction of the 1st additional jet b) Gap fraction of the 2nd additional jet (hard cuts) (hard cuts)



c) Gap fraction HT (hard cuts)

The result is again nearly the same as with the normal cuts.



4.2.3 Comparison between normal and hard cuts

Left column: Comparison between Right column: Comparison between normal and hard cuts for Pythia 8 for normal and hard cuts for Cascade for the the gap fraction of the 1st jet, 2nd jet gap fraction of the 1st jet, 2nd jet and and HT. Pythia labeled are the hard HT. Cascade labeled are the hard cuts, cuts, Pythia0 labeled the normal cuts. Cascade0 labeled the normal cuts. The The Differences are very small.

The differences between normal and hard cuts are again very small, so there is good agreement between the whole event selection and the selection via kinematic cuts shown.

At last both Pythia 8 and both Cascade results are shown in comparison with data (Pythia, Cascade labeled are hard cuts, Pythia0, Cascade0 normal cuts):



a) Gap fraction of the 1st additional jet, b) Gap fraction of the 2nd additional jet, both cuts both cuts



c) Gap fraction of HT, both cuts

4.3 Gap Fraction in different $|\eta|$ regions

It is also possible to obtain the gap fraction in different pseudorapidity (η) regions. Therefore only jets falling into the chosen region of $|\eta|$ are considered:

$$f(p_T) = \frac{\text{\#of events with no jet above the } p_T \text{ threshold in the considered } |\eta| \text{ region}}{\text{\#of selected events}}$$

In case of HT, only jets in the required $|\eta|$ region contribute to the scalar sum. All three cases were simulated for three different $|\eta|$ regions: $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.5$ and $1.5 < |\eta| < 2.1$, corresponding to the measured data and simulations in [1].

4.3.1 Normal cuts



c) Gap fraction of HT $1.5 < |\eta| < 2.1$



Left column: Gap faction of the first jet in $0<|\eta|<0.8,\ 0.8<|\eta|<1.5$ and $1.5<|\eta|<2.1$

Right column: Gap fraction of the 2nd jet in $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.5$ and $1.5 < |\eta| < 2.1$

The description gets better with rising $|\eta|$, Cascade does a little better job than Pythia, which might correspond to its angular ordering that let its result get better in non-extreme (not very low or high) $|\eta|$ regions - the description of the data is worst (and a little worse than Pythia) in the $|\eta| < 0.8$ region.



4.3.2 Hard cuts



Left column: Gap faction of the first jet Right column: Gap fraction of the 2nd in 0 < $|\eta|$ < 0.8, 0.8 < $|\eta|$ < 1.5 and jet in 0 < $|\eta|$ < 0.8, 0.8 < $|\eta|$ < 1.5 and $1.5 < |\eta| < 2.1$ (hard cuts) $1.5 < |\eta| < 2.1$ (hard cuts)



(hard cuts)

b) Gap fraction HT $0.8 < |\eta| < 1.5$ (hard cuts)



c) Gap fraction of HT $1.5 < |\eta| < 2.1$ (hard cuts)

The results for the hard cuts are again nearly the same as for the normal cuts.



4.3.3 Comparison between normal and hard cuts



Left column: Gap faction of the first jet Right column: Gap fraction of the 2nd in $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.5$ and jet in $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.5$ and $1.5 < |\eta| < 2.1$ $1.5 < |\eta| < 2.1$





c) Gap fraction of HT $1.5 < |\eta| < 2.1$ As can be seen in the histograms, there is very low difference between the two sets of cuts (in most bins they overlap).

4.3.4 Powheg

As there were some problems with the Powheg program, only many samples could be simulated and merged, which led to some problems with the correct normalization/scaling. Normalised samples were simply added, so comparison with data and the other MC generators is only possible by eye in the histogram. Also there was not enough time to run the simulation without the harder cuts, so only the hard cuts histograms are available. Although the ones for the gap fraction will be shown here:



c) Gap fraction of HT, Powheg



Left column: Gap faction of the first jet Right column: Gap fraction of the 2nd in $0 < |\eta| < 0.8, 0.8 < |\eta| < 1.5$ and jet in $0 < |\eta| < 0.8, 0.8 < |\eta| < 1.5$ and 1.5 < $|\eta| < 2.1$ (Powheg) 1.5 < $|\eta| < 2.1$ (Powheg)



a) Gap fraction HT 0 < $|\eta|$ < 0.8 (Powheg) b) Gap fraction HT 0.8 < $|\eta|$ < 1.5 (Powheg)



(I owneg)

The shape *seems* to fit quite good to the data, but it is not possible to say even if it is overor underestimating the gap fraction or other statements on the quality of the simulation. First thing for further research will be getting Powheg to run and compare this simulation to the ones from the paper in detail, since Powheg was used there also.

4.4 *tt* System transverse momentum

According to a recent paper [12] it seems to be interesting to take a closer look at the p_T of the $t\bar{t}$ system. Simulated was using Pythia 8 and Cascade, and two slightly different Rivet analyses were written: One (toppt.cc) directly accesses the transverse momentum of the $t\bar{t}$ system via questioning whether there are the decay products of a top quark pair in the event and then directly gets p_x and p_y and calculates $p_T = \sqrt{p_x^2 + p_y^2}$, the other analysis (top2.cc) performs first the event selection for the dilepton channel which was used for the cross section and gap fraction and then asks again directly for the p_T . This is only possible in simulations, as in experiment decayed particles have to be reconstructed and are not all known.

Also two slightly different approaches were used to get the p_T of the $t\bar{t}$ system: First the direct approach - if there is a top and an antitop in the event get its p_x and p_y and calculate the p_T . Second was questioned whether all decay products of the $t\bar{t}$ pair (b, \bar{b}, W^+, W^-) and itself were in the event and then again the momenta of the t-quarks were directly accessed and filled into the histogram.

Both ways were plotted in a histogram with binning of 10GeV and a fine binning of 1GeV width, to be able to take a closer look at the very low transverse momentum region of the $t\bar{t}$ system.



Figure 4: Direct approach to the p_T of the $t\bar{t}$ system

The toppt.cc analysis, which first performs the dilepton event selection has far less events in it, roughly 100 times, but the shape is roughly the same. The reason is, that about one percent of all generated top events match the dilepton channel criteria, this was also the case for the cross section and gap fraction analysis. Most $t\bar{t}$ pairs are produced at low p_T . Possibly one could normalize everything by the number of events and so be able to compare both analyses



Figure 5: Direct approach to the p_T of the $t\bar{t}$ system, fine binning



Figure 6: Look whether there are also b, \bar{b}, W^+, W^- in the event, then get the p_T of the $t\bar{t}$ system



Figure 7: Look whether there are also b, \bar{b}, W^+, W^- in the event, then get the p_T of the $t\bar{t}$ system, fine binning

better. Surprising is the difference between Cascade and Pythia in the histograms with fine binning. Cascade shows again a peak at low p_T , Pythia shows far less events in the first 10 GeV. The source of this behaviour might lie in the different mechanisms of the MC generators, with some data provided for the p_T of the $t\bar{t}$ system it can be decided which generator describes the data better and can be considered for further studies.

5 Acknowledgments

I want to thank my supervisors Hannes Jung and Carmen Diez Pardos for their help in this project, every time it was possible to ask a (even stupid) question which was kindly answered and gave me the possibility to do a next small step in this project.

Also I want to thank the organisers of the Summerstudent programme for choosing me to get this great experience here at DESY and for their help in the start, where I was not so happy, the other people of the CMS group, which I also could always bother with questions, the other summer students for discussion and fun

the other summer students for discussion and fun,

my professors at my home university, Jürgen Schnack and York Schroeder and PhD student Michael Czopnik for supporting my application for the Summerstudent programme and also my family for their overall support.

References

- [1] CMS Collaboration: Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at $\sqrt{s} = 8$ TeV, CMS Physics Analysis Summary, CMS-PAS-TOP-12-041.
- [2] Maria Aldaya, Top Quark Physics DESY Summer Student Lecture (21.8.2014) http:// summerstudents.desy.de/hamburg/e177730/e178988/top_physics_210814.pdf.
- [3] CMS Collaboration: Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV submitted to JHEP, CMS-TOP-11-005.
- [4] Analysis ATLAS_2012_1094568, ATLAS Collaboration 2012 https://rivet.hepforge. org/code/dev/a00444_source.html arXiv:1203.5015 [hep-ex].
- [5] Collection of diagrams of top signals and backgrounds, http://www-d0.fnal.gov/ Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html
- [6] A brief introduction to PYTHIA 8.1, Torbjörn Sjöstrand, Stephen Mrenna, Peter Skands arXiv:0710.3820
- [7] Pythia introduction notes, Stephen Mrenna, http://home.fnal.gov/~mrenna/HCPSS
- [8] H. Jung, G. P. Salam: Hadronic final state predictions from CCFM: the hadron-level Monte Carlo generator CASCADE arXiv:hep-ph/0012143v1
- [9] M. Deak, F. Hautmann, H. Jung, K. Kutak: The CASCADE MC event generator, Talk, Madrid 2009 https://indico.cern.ch/event/53294/session/3/contribution/ 278/material/slides/0.pdf
- [10] The POWHEG BOX, Project http://powhegbox.mib.infn.it
- [11] C. Oleari: The POWHEG BOX arXiv:1007.3893v1 [hep-ph]
- [12] Stefano Catani, Massimiliano Grazzini, Alessandro Torre: Transverse-momentum resummation for heavy-quark hadroproduction, arXiv:1408.4564v1 [hep-ph].