Vrije Universiteit Brussel

FACULTEIT WETENSCHAPPEN EN BIO-INGENIEURSWETENSCHAPPEN

Anomalous Couplings Project

Progress Report and documentation of Tools

Annik Olbrechts

Started 4 April 2014

Contents

Chapter 1

Pending issues

- Does this created FeynRules model still contain Effective Field Theory?
- If the kinematics doesn't change for coupling parameters larger than 1, how is then possible to differentiate the different configurations which will be studied?
- Access to Mathematica for Width studies ... ?
- XS values which are used are the ones calculated using the MadGraph_v155 version. Should this be calculated again using the new MadGraphv5_aMC@NLO since normally this shouldn't change ...?
- Create branch for localgrid scripts! (Now already a branch exists for the documentation !)
- Understand *CorrectPhi* comment in MadWeight name
  $\rightarrow$ Maybe related with phi issues of neutrino's ...
- EventWeight calculated in analyzer should be integrated into MadWeight output (Hence multiply MadWeight weight with the EventWeight for each event!)
  Otherwise all the effort to include the JES and PU stuff has no influence at all and is not taken into account!
- Update EventNumberInformation file to only cout the information when the event has passed the eventSelection requirements. Otherwise just discard this event from the output file in order to avoid a long .txt file and long running time ...
- When creating ntuples for Data/MC comparison the EventNumberInformation output should be stored here as well. This will allow to calculate the likelihood and weight distributions for specific p_T cuts without having to run the entire base code again. So also the different python scripts which have been created should be moved to this .cc code in the future, so better not to spent too much time on improving these python scripts to avoid double work ...
  For storing the generator decay channel information the c++ enum should be used which can store multiple variables (**Check this out!**)
- Need to understand how MadWeight deals with the permutations ... Should this permutation of the light jets be done within MadWeight or should two .lhco files be created and sent separately to MadWeight. If the latter is the case, how should they be combined afterwards ?

- Be careful with used Beam Energy when using MadGraph. In the newest MadGraph directory this still seems to be set to 7TeV in stead of the required 4 GeV.
Could this have an influence on the obtained MadWeight results when using the new MadWeight version (connected to the newest MadGraph version) ...

- Specific MadWeight question for the latest version:
The 'refine' option allows you to realunch the computation of the weights which have a precision lower than X. ***But how can you find this precision?***

- Should a ScaleFactor correction be applied for the Branching ratio ?
This is not the case in code of James, but should find recommendations somewhere ...

- **To Check:** Does MadWeight give a different weight when the Neutrino Mass changes?
→ Currently mass is manually set to zero ...

- Running the analyzer code with the PileUp and JES/JER influences correct and active, the number of preselected events in the Event Selection Table has changed. Need to understand if it is expected that these corrections have an influence (lowering the number) on the number of preselected events ...

- Check whether the created root file can be trusted when the number of considered b-tags is greater than 1 ... !!

Chapter 2

Used Tools and Techniques

2.1 MadGraph

2.1.1 MadGraph_v155 version

Currently the MadGraph version which should be used is the MadGraph_v155 since this is the only one which was compatible with the used FeynRules version at the time of the creation of the model. According to a mail of Olivier Mattelaere on 6 February 2014 this issue should now be fixed.

In this project MadGraph is only used to create .lhe files in order to understand the created model. Especially for Cross Section studies it is extremely useful.
Hence two python scripts have been created in the following directory:

/user/aolbrech/AnomalousCouplings/MadGraph_v155/MassiveLeptons/
MadGraph5_v1_5_5/Wtb_ttbarSemiElMinus/RVL_RVR_XSGrid003.py
/user/aolbrech/AnomalousCouplings/MadGraph_v155/MassiveLeptons/
MadGraph5_v1_5_5/XSScript.py

The first one automatically creates all the desired configurations, for different RVL, RVR, Lepton Pt Cut and Jet Pt Cut values. The desired output directory can be specified in this file as well.
In order to save some disk space the events.lhe file is deleted using this script and only the unweighted_events.lhe are kept for analysis. These are the relevant events (*according to Alexis so understand what is the difference*).

The second file calculates the cross section of the desired configurations. It again loops over all the different variables and creates a .txt table and a .pdf table.
Maybe this can be created automatically when executing the first python script! This to avoid copying the considered variables from one file to another!

2.1.2 New MadGraph5_aMC@NLO version

This new MadWeight version was released in the spring of 2014 and is accompanied with a new MadWeight version as well.

Normally the cross section calculations shouldn't differ between the different MadGraph versions, but for at least one of the decay channels this should be checked!

In order to use MadWeight and calculate the cross section values for the desired processes and decay channels the following commands should be used:

$$./bin/mg5_aMC \tag{2.1}$$

$$importmodelMassiveLeptons \tag{2.2}$$

$$generatepp > tt~, (t > bw+, w+ > mu + vm), (t~ > b~w-, w- > jj)@1QED = 2 \tag{2.3}$$

$$outputdirName \tag{2.4}$$

$$exit \tag{2.5}$$

Once this directory is created the *index.html* file contains all the Feynmann Diagrams which correspond to the considered process. This file can be opened using $firefox$ on mtop.

Since the FeynRules diagram is created in such a way that it contains a **NP**, **QCD**, **QED** and **TEST** variable representing the different interaction vertices. Hence asking the variable **QED** to be equal to 2 results in the 16 independent diagrams which are required to represent the anomalies in the Wtb vertex. Currently the decay of the top quark is altered to contain all the different coupling constants while the other interaction vertices represent the Standard Model expectations.

When the cross section should be calculated for a specific process the following configuration files in the Cards directory should be updated.

- **run_card.dat** where the number of events which should be generated is defined together with the beam energy of the considered collision process. Also should be ensured that no kinematic cuts are active on the particles since these cuts will have a different effect then the cuts applied in the event selection of the analysis. These cuts are applied on the generator level instead of the jet-level which results in much more stringent and tighter effects when using the same p_T value.

- **param_card.dat** where all the parameters of the considered model should be defined.

- **proc_card.dat** contains the event which will be generated and the models which have been imported. This file shouldn't be changed but is particularly useful to ensure that the correct event is considered for generation.

All these files can still be changed once the *./bin/generate_events* command has been executed. The MadGraph software always asks whether one or more of the configuration files should be adapted before the command is submitted and allows a waiting time of 60s. Hence in order to run MadGraph continously using a script the **me5_configuration.txt** file should be adapted to enable this waiting time since otherwise when using nohup to run the script results in a crash and the script is terminated.

2.2 MadWeight

The running of the MadWeight event generator has to be done on localgrid since the events .lhco file has to be splitted in multiple jobs. The number of jobs which can be submitted is not limited, hence the events per job should be taken as low as possible. Running MadWeight with a large number of events per job implies that it will take a very long time since the submitted jobs have a long walltime which reduces their priority. Another important factor for the MadWeight event generator is defining the number of interaction points which have to be considered. This number has to be rather large (around 10 000 - 30 000 according to Lieselotte **TO CHECK MYSELF**) to ensure the uncertainty on the weight to be smaller than the weight itself.

Currently two different MadWeight event generator versions exist, and up to now (May 6 2014) only the oldest one has been tested and used extensively. The newer one was only installed beginning of May but should be less CPU intensive and allows the user to split the interaction points in two distinct steps. The versions are:

- madweight_mc_perm
- MadGraph5_aMC@NLO

The installation commands of these two versions are the following (the bzr command package has been installed on the m-machines, but not on mtop):

- bzr branch lp: maddevelopers/madgraph5/madweight
- bzr branch lp: maddevelopers/mg5amcnlo/madweight

Whenever MadWeight is first used a personal directory should be initalized. This can be done in the following way:

- Import the created model (called MassiveLeptons) in the model directory of MadGraph
- Activating the MadWeight event program and initializing a personal directory.
  1. ./bin/mg5_aMC (The mg5 executable should not be used!)
  2. import model MassiveLeptons
  3. generate p p > t t , (t > b w+ , w+ > mu+ vm) , (t > b w- , w- > j j) @1 QED = 2
  4. output madweight
  5. exit

2.2.1 MC_PERM MadWeight use

The most important steps which have to be executed to use this MadWeight version are the following ones [1]:

1. Initialize MadWeight to run on localgrid!
2. Update the MadWeight_card.dat
3. Set the correct transfer functions

[1] Full detailed explanation about the use of this madweight version can be found in the documentation of Bettina.

4. Run MadWeight

For this first step the following two files have to be changed in the /blablaMadWeight/bin/internal directory:

- madweight_interface.py
- cluster.py

2.2.2 Checking MadWeight on localgrid

Different command which should be used to check whether jobs are running are running on localgrid and how they should be killed when something went wrong:

$$qstat \;\; @cream02 \mid grep \;\; aolbrech \tag{2.6}$$

$$qdel \;\; 394402.cream02 \tag{2.7}$$

2.2.3 Influence of the used Transfer Function

2.2.4 Adapting MadWeight to run continously on localgrid

A single test with the following configuration (5000 events - 20 events/job - 10 000 initPoints) took almost 16h to finish so the full generator event sample (> 2 470 000 events) would take more than 329 days![2]
Even the full reconstructed event sample (> 210 000 events) requires more than 28 days of running this way.
TODO: Time of running should be compared against the new MadWeight version!!

A possible solution to reduce the CPU time to process these events is in stead of sending 2000 jobs, waiting until they are all finished and only then submitting the next bunch of 2000 jobs trying to adapt the MadWeight script to send continously 2000 jobs. This would imply that the script should check whether one of the jobs have been finished and immediatly submitting a new one.
For this the above mentioned scripts should be adapted!

According to Olivier (D) this can't be changed by the grid admins, but should be adapted by the MadGraph developers.

2.3 MadAnalysis

2.3.1 How to run MadAnalysis in expert mode

Version v112 should be used since the expert mode in this version works exactly as explained in the manual (arXiv: 1206.1599). In the more recent version v115 the expert mode doesn't work out-of-the-box and the python files need to be adapted ...

[2]From this can be concluded that the events on generator level should never be all run. Only a limited selection of these events should be considered.

Starting with a new analysis directory in MadAnalysis in expert mode can be done by typing the following command:

$$./bin/ma5 \quad --expert \tag{2.8}$$

In the questions asked by MadAnalysis after executing this command the name of the directory which needs to be created and the name of the analysis has to be given.
The name of this analysis shouldn't be made too complex since it has to be typed everytime when creating plots with MadAnalysis.

After the correct directory is created, the *Name*/SampleAnalyzer directory should be initialized by executing the following two commands:

$$source \;\; setup.sh \tag{2.9}$$
$$make \tag{2.10}$$

The actual analysis should be created in the Analysis directory, and a similar approach to user.cpp and user.h should be adopted.
Everytime a change has been made to these two files *make* should be executed in the SampleAnalyzer directory in order to process these changes.

The .lhe files (MadAnalysis cannot process .lhe.gz files so they should be unpacked using gunzip .lhe.gz) which should be considered should be wirtten down in a .txt files which is saved in the SampleAnalyzer directory.

The actual running of MadAnalysis is done with the following command:

$$./SampleAnalyzer \quad --analysis = "Name\ of\ analysis" \quad List.txt \tag{2.11}$$

2.3.2 Content of analysis file in MadAnalysis

The latest analysis file which has been used with all the necessary information can be found in the following directory on the m-machines:

$AnomalousCouplings/MadAnalysis\_v112/Wtb\_PtCutInfluence/SampleAnalyzer/Analysis$

The two analysis files with full detailed information are the LeptonPtCutInfluence.cpp and the JetPtCutInfluence.cpp files. They both consist of two different functions, namely the *Execute* and the *Finalize* function. The first one allows to access the information of each event while the second one is entered for each file. Therefore the particle content is reconstructed in the *Execute* function and the histograms for all the considered files are constructed in the *Finalize* function. Currently these analyzer files look at 28 different kinematic variables. All the kinematic information of each of the particles present in the expected semi-leptonic $t\bar{t}$ event is created.

In order to separate the two b-quarks in the event, the Particle Id of the leptonic top quark needs to be known. Therefore an integer *LeptonicTopPdgId* is used and the kinematic information of the b-quarks can only be stored when this integer is different from zero. This will normally not result in by-passing the b-quark information since the events in the .lhe files are read in in the same order as they are created by the MadGraph command. So first the top quarks are considered and only then the final state particles.

? Isn't it safer to use the daughter information of the top quark? Because in the case of t t -> b jj b l v the lepton comes after the b-quark ... Should investigate what happens in this case ... **TO CHECK:** I think some extra safety is incorporated and the actual loop of filling the event content is only done when the b's are reconstructed!

In order to automatically create the 28 histograms for all the different files and distinguish the different RVR and RVL values, an automatic name-givng loop is used. The name of the content of the histogram is each of the time combined with the correct name of the RVL/RVR content of the .lhe file and the decayChannel.

Currently the name of the considered .lhe files has to be adapted every time the files are changed. Maybe this should be made automatically read in from the .txt file to ease the processing of different configurations. If the .lhe files have a clear name where the RVL, RVR and PtCut values can be obtained from using a python script it shouldn't be too much work to save this in the .cpp file with this script. ***TO DO!***

2.3.3 Analyzing the MadGraph files

Currenlty the created model should be completely understood and the behavior of the model when the coupling coefficients are larger than 1 should be investigated. Therefore new MadGraph files have been processed for the following configuration:

$$\begin{aligned} Re(V_L) &\in [0.7,\ 1.3] \\ Re(V_R) &\in [-0.3,\ 0.3] \end{aligned}$$

These files can be found in the following directory on the m-machines and contain 100 000 events.

$$/user/aolbrech/AnomalousCouplings/MadGraph\_v155/MassiveLeptons/ \\ MadGraph5\_v1\_5\_5/Wtb\_ttbarSemiElMinus/ResultsXSGrid003$$

2.3.4 Width of the decay

2.4 FeynRules

Should be understood why the mass of the top quark is equal to 180 within the created FeynRules model!

Chapter 3

Generator level bottlenecks

The goal is to obtain as fast as possible results on generator level in order to ensure that no bottlenecks are found when running MadWeight. The advantage of only using generator level results is that one can be completely sure that the studied events are actual semi-leptonic ttbar events. Hence MadWeight should not have any problems calculating the weight for these kind of events and no CPU time will be spent on uncorrect events. This implies that any deviation from the expected results implies a bias, or even a problem, concerning the MadWeight output.

Once the results correspond to the expectations these preliminary results should be easily extended to reconstructed events. Finalizing the event selection then allows to fully trust the results obtained on reconstructed level and make sure that any deviation should be explained by the influence of the applied event selection.
These results can then be used to optimize the event selection with respect to the MadWeight output and CPU time needed.

3.1 Transfer Functions

In order to obtain results with MadWeight, the Transfer Functions which link the reconstructed energy distribution with the actual energy distributions should be calculated. In the case of generator level events this is not that important since no smearing of the energy is expected, but in order to avoid any bias from the used Transfer Functions it is adviced to use the real Transfer Functions with a smaller width.

The method to obtain the parameters describing the energy smearing is (partly) explained in the PhD Thesis of Arnaud Pin[1] which can be found in:

$$https://cp3.irmp.ucl.ac.be/upload/theses/phd/Pin\_Arnaud.pdf$$
$$/home/annik/Documents/Vub/PhD/ThesisSubjects/AnomalousCouplings/$$
$$PrepareGenLevelRunning\_Sep2014/TransferFunctions$$

The method to obtain the parameters of the Transfer Functions is based on the code

[1] It should be noted that in the PhD Thesis of Arnaud, and in the current double Gaussian transfer function syntax in MadWeight, only 5 parameters are used. The narrow gaussian distribution doesn't have a normalisation parameter in front, but is normalized afterwards in the MadWeight code.

received from Petra and Lieselotte, which was used in the Master thesis of Lieselotte[2]. This code is almost identical to the ROOT class FitSlicesY() but has some minor differences. Some of these have already been changed in order to match better with the ROOT class.
One of the most important differences between the two approaches was the treatment of the underflow and overflow bin. In the received code these two bins were respectively added to the first and last bin and hence included in the fitting range. This is not the desired behavior since the size of underflow/overflow bin can be relatively large compared to the first/last bin and significantly change the value of these bins. This would then imply that the position of the first and last bin is not located at the correct position and will, especially in the case of a limited number of bins, have a significant influence on the fit result. Now these underflow and overflow bins are just discarded from the fit range and will have no influence on the final result.

Another important, but useful, difference between the two methods is the number of histograms which are saved. The received code saves for each distribution which is considered the ProjectionY distribution together with the double Gaussian fit for this bin. This can not be changed in the ROOT class which only stores the distribution of the 6 parameters of the double Gaussian fit formula.

However, even after carefully ensuring that both methods are identical the obtained results are not. Up to now it is not clear what is the reason for the discrepancy between the two results and the only way to found out is comparing the distributions and results for a significant amount of statistics.
One possibility is the used fit ranges and number of bins. If the last bins are low on statistics their distribution might not agree with a double Gaussian distribution and hence result in a failed fit. Therefore the distribution for each ProjectionY bin is now closely studied for all of the considered histograms. This can be found is the following section 3.1.1.1.

<u>Remark:</u> Using transverse momentum in stead of energy
SHOULD MOVE TO OTHER SPOT !!! (gives too much MadWeight detail so move down in section) The Transfer Function configuration files in MadWeight are flexible enough to change the kinematic variables used for the Transfer Function calculations. Hence it seems to be more relevant to utilize the transverse momentum in stead of the energy of the considered partons and jets. Especially since the LHCO file used in the MadWeight calculations has the transverse momentum as input variable. Therefore it seems much more useful and realistic to use this parameter and not the energy. This implies that the Transfer Function configuration file should be adapted to use the *pt(p)* and *pt(pexp)* variables and not the currently used *p(0)* and *pexp(0)*.
After actually implementing the created Transfer Function files[3] and trying to use this Transfer Function one issue showed up which is currently not completel solved yet. MadWeight contains hard-coded files which only allow the use of the E, THETA and PHI variable for the Transfer Function dependencies. Therefore the so-called "block name"

[2] Additional feedback was also received from Arnaud, but he was using a completely different unbinned likelihood method. Since this only arrived after a couple of weeks waiting, it was decided to continue with the binned likelihood method used by Petra and Lieselotte.

[3] This was tested the first time on 30 November 2014.

PT is not accepted by the *change_tf.py* file. However it seems that there is no clear physical motivation for this limited choice of kinematic variables, but in order to be completely sure additional information is needed from Olivier and Pierre.
Once this issue is resolved the following file, which contains the limited amount of kinematic variables, should be adapted[4]:

$$bin/internal/madweight/change\_tf.py \tag{3.1}$$

The other relevant files used for the creation of the Transfer Functions are given in Equation **??**. The first one is the translation of the used $TF\_user.dat$ into a MadWeight readable file which can be implemented. This is build using the commands explained in Equation **??**. The second file is only relevant around line 292 where the function used for the Transfer Function creation is explained. This is the general MadWeight constructor file where all the different MadWeight functions are defined.

$$\begin{aligned} Source/MadWeight/transfer\_function/transfer\_function.f \\ bin/internal/madweight\_interface.py \end{aligned}$$

$$\begin{aligned} ./bin/mw\_options \\ define_transfer_fct \end{aligned}$$

3.1.1 Comparing ProjectionY distributions

Each of the considered histograms is a 2D histogram where the x-axis represents the energy of the generator level parton and the y-axis the difference between the generator level parton and the reconstructed matched particle. This is done for the energy distribution and the θ and ϕ angles.

Energy difference between parton and light jets
The first bin in the left-hand histogram is asymetric because of the applied event selection cuts on the reconstructed level. These event selection cuts will not have the same influence on the generator level partons explaining the slight asymetric behavior.

[4]Another solution is to keep the block name equal to E but just use the transverse momentum information. However this might result in confusion when in some files the hard-coded E parameter gets written down (for example on plots).

Figure 3.1: Energy of the generator parton versus the reconstructed jets for the two light quarks in the semi-muonic ttbar event topology (left) and their respective difference with respect to the generator level energy (right).

From these fit results can be concluded that the double Gaussian distribution is well recovered for lower energy values. As can be seen from the very first histogram, the largest bulk of events have an energy value lower than 100 GeV. The low statistics in the higher energy range can partly explain the failing fitting performance. The double Gaussian function form is used in order to properly reconstruct the narrow peak together with the wider Gaussian distribution representing the tail of the energy distribution. However in the high parton energy range, this first narrow distribution dissapeared due to low statistics in this region. Hence fitting these distributions will indeed result in an unsuccesful fit status.

A possible solution which can be considered is adding some of these histograms together and in this way enhancing the statistics in these combined bins. From the collection of distributions per bin could be concluded that the "ok" and "failed" fits are the ones with only about 2% of the total number of events. Hence it could be possible to adapt the code in such a way that in these cases the bin is combined with the following ones until a percentage higher than 3% is obtained.
Another solution consists of reducing the range of the fitted histograms and only fit the distributions with $E_{parton} < 150$GeV. However before continuing with this option it should be completely understood whether MadWeight is able to extrapolate in a correct way to higher energy values.

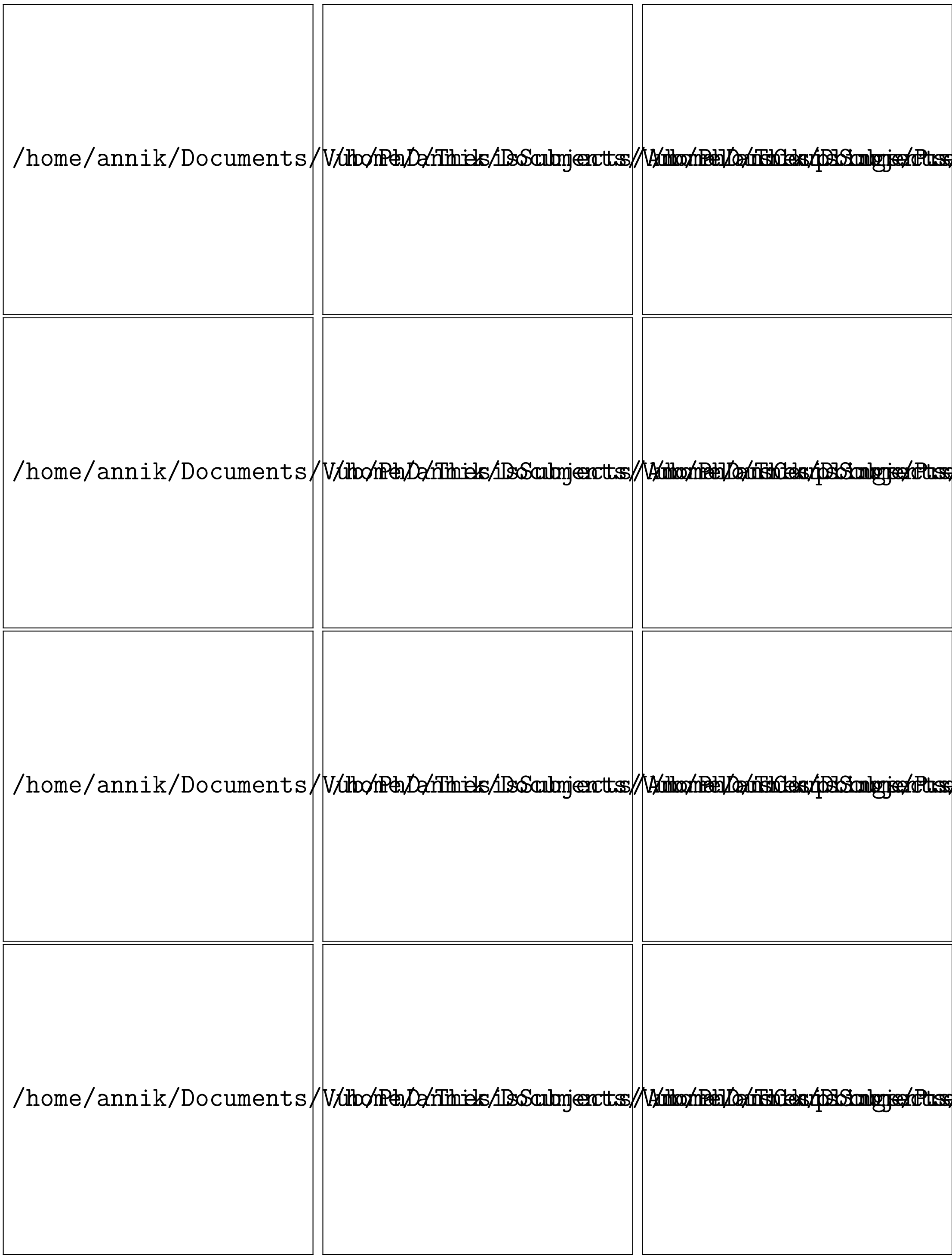


Figure 3.2: Distribution of the energy difference between the generator level parton and the corresponding light quark jet for each of the 10 considered bins and the overflow bin. All distributions were fitted with a double Gaussian function and besides bin 8, 9 and 10 returned with status “converged”. Bin 9 had status “failed” while the remaining two bins had status “ok”.
The final histogram (bottom right) shows the χ^2 distribution for this fit for each bin.

Figure 3.3: Energy dependency of the 6 parameters of the double Gaussian fit function. The two lowest rows show twice the same result, but with a different y-axis range. The double Gaussian fit parameters are combined in a histogram and then fitted with the Calorimeter energy function as explained in the PhD Thesis of Arnaud Pin.

3.2 Cross Section distribution for new grid

3.3 First results: Wrong log(likelihood) minimum

The first obtained MadWeight results for the enlarged grid ($V_L \in [0.8, 1.2]$ and $V_R \in [-1, 1]$) using only parton-level ttbar events did not result in the expected minimum of $(V_L, V_R) = (1, 0)$. This can be seen from Figure 3.4, which shows the distribution of the log(likelihood) for each point in the considered grid.

One of the possible influences on the displaced minimim of the log(likelihood) distribution could be the normalisation of the Cross Section influence. This XS normalisation ($\frac{XS}{XS^{SM}}$) should be multiplied with the likelihood value, not the log(likelihood). Hence in

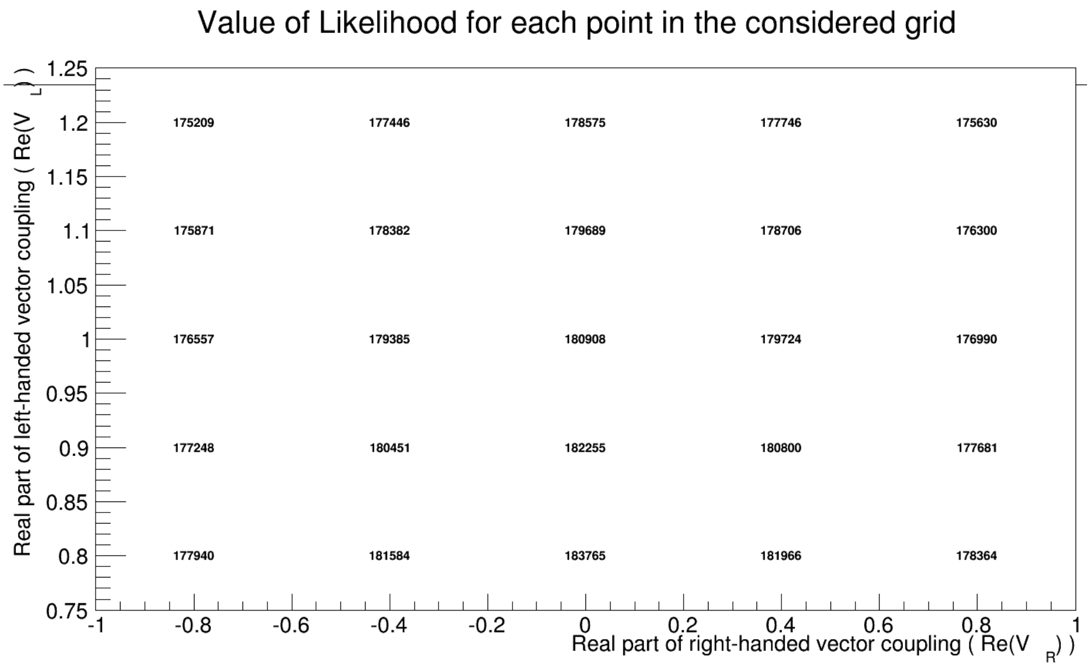


Figure 3.4: Distribution of the log(likelihood) for each point in the grid using 3200 parton-level positive semi-muonic ttbar events. The transfer function used to smear the parton-level kinematics is the single-gaussian function standard included in MadWeight.

order to correctly take this into account the obtained log(likelihood) value for each point in the grid should be corrected using the logarithm of this XS normalisation. The distribution of the normalisation on the Cross Section can be found in Figure 3.5 together with the log(likelihood) distribution after correctly taking into account this XS normalisation.

The formula which has been used is the following (Equation 3.1):

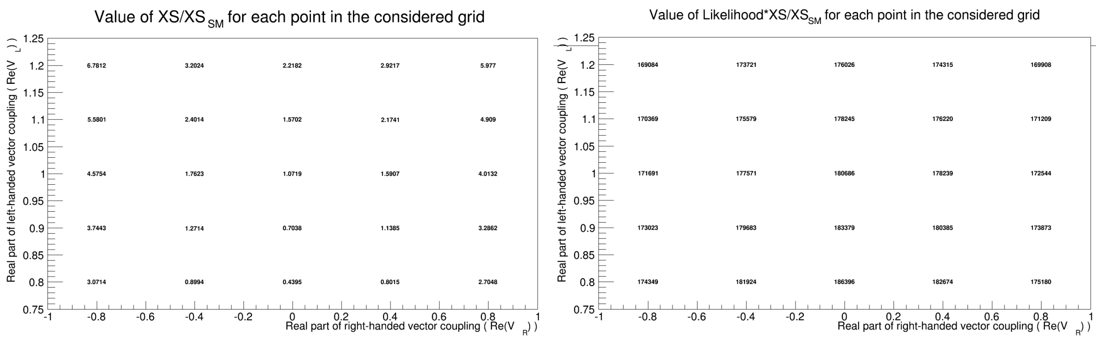


Figure 3.5: Distribution of the XS normalisation for positive semi-muonic ttbar events (left) and distribution of the log(likelhood) after taking into account this normalisation. As in the previous figure 3200 positive semi-muonic have been used to obtain this distribution and a single gaussian transfer function has been applied to smear the kinematics of these parton-level events.

$$P(y|a) = \frac{1}{\sigma(a) * Acc(a)} \int W(y|x,a) * Eff(x,a)|M(x,a)|^2 T(x,a) dx \tag{3.2}$$

$$\mathcal{L} = \prod P(y|a) \tag{3.3}$$

The normalisation which has been applied in Figure 3.5 is given in Equation 3.3:

$$\mathcal{L}_{Norm} = -ln(\sum P(y|a) * \frac{XS}{XS^{SM}}) = -ln(\mathcal{L}) - ln(\frac{XS}{XS^{SM}} * N) \qquad (3.4)$$

It should be checked whether this is the correct method to take into account the normalisation of the XS. Currently it has been assumed that this XS normalisation should be applied for each weight and hence is multiplied with the number of considered events. In the case that this normalisation should just be multiplied with the overall likelihood value ($\mathcal{L}$) the sum over the number of considered events drops out of the equation implying a very small influence of the XS on the log(likelihood) distribution.

As highlighted in the general Matrix Element Method formula (Equation 3.1), the probability to measure the observed quantities y already has a normalisation factor for the cross section. This factor is defined as the channel cross section and is calculated using Equation 3.4:

$$\sigma(a) = \int_{X_i} |M(x,a)|^2 T(x,a) dx \qquad (3.5)$$

Hence it should be investigated in detail whether this XS normalisation should still be applied. From the above equations could be concluded that the change in cross section is actually already incorporated in the weight obtained from MadWeight. This could make sense since MadWeight has all the necessary information to calculate the cross section for each point in the considered grid. The cross section values for each point have been calculated using MadGraph and the same model as used for the MadWeight calculations. Unfortunately it is not completely clear from the MadWeight documentation whether this is actually included in the weight or not.

3.4 Correct normalisation of Matrix Element probability

As could be seen in Equation 3.1 a term $\sigma(a)$ is included in the general Matrix Element Techniques formula. However it is not clear whether this cross section normalisation is actually performed within the MadWeight calculations or whether this normalisation should be done afterwards.
This question is closely related to the order of the current obtained weight values with MadWeight. Up to now no weight larger than 10^{-22} have been obtained, resulting in a very large log(likelihood) value.

This small value can be caused by many different reasons for which the most plausible ones are listed here:

- The normalisation of the MadWeight probability should still be done and is not performed within the Matrix Element Techniques formula.
  This can only be ruled out by contacting the Madweight experts and

asking explicitely what is done in the Madweight calculations. Also a possible hint could be found inside the MadWeight python files (but this should only be done if the received answer is not perfectly clear).

- The smallness of the weight could be caused by an error inside the created FeynRules model. *Should also look for the mail where one of the MadWeight experts (Olivier/Pierre or even Celine) answered about the possible explanation for the smallness of the weight and whether this implies some wrong assumptions).*
A possible way to exclude that the origin of this problem is the AnomalousCouplings FeynRules model is by comparing the results for the top mass fit when both the SM FeynRules model and the AnomalousCouplings model is used. If the weights are also this small when the SM model is used this smallness should be solved by an additional normalisation factor.

Update 31/10/2014: Probability function NOT normalized (according to mail Olivier)
As was expected from the smallness of the obtained MadWeight probabilities should the cross section normalisation be applied afterwards. Only in the older versions of MadWeight (based on MG) was this normalisation included automatically.

3.4.0.1 Measurement of top quark mass using Matrix Element Method

Comparing SM model with AnomalousCouplings model
As a first step the Feynmann diagrams belonging to the two different models should be compared. This information can be found in the *index.html* file in the following directories (and the files should be opened using firefox on mtop since this is the only m-machine with a working browser):

$$/AnomalousCouplings/MadGraph5\_aMC@NLO/madgraph5/SM\_ttbarSemiMuPlus$$
$$/AnomalousCouplings/MadGraph5\_aMC@NLO/madgraph5/ttbarSemiMuPlus\_QED2$$

Comparing SM cross section with MassiveLeptons cross section
In order to be sure that both models have the same Standard Model base, the cross sections for both models have been compared. This resulted in an unexpected outcome, namely that the obtained cross sections differ significantly depending on which MadGraph version is used to generate the considered events. A summary can be found in Table 3.1.

Top quark mass	MadGraph aMC@NLO		MadGraph v155	
	SM model	MassiveLeptons model	SM model	MassiveLeptons model
153	9.23 pb	9.645 pb	6.692 pb	6.984 pb
163	11.12 pb	11.63 pb	7.844 pb	8.199 pb
173	12.98 pb	13.54 pb	8.897 pb	9.281 pb
183	14.77 pb	15.4 pb	9.884 pb	10.3 pb
193	16.5 pb	17.22 pb	10.78 pb	11.25 pb

Table 3.1: Cross section values for semi-muonic (+) ttbar decay obtained using two different MadGraph versions.

From this table can be seen that there is, for both considered MadGraph versions, a small difference between the SM FeynRules model and the MassiveLeptons one. This could be caused by the different treatment of the leptons. In the SM model they are considered to be massless while in the MassiveLeptons one they are defined to have their actual mass.
A larger differrence occurs when both MadGraph versions are compared. From the answer received by Olivier it is not clear whether this difference is worrysome or could be explained by the LO theoretical uncertainties. Should also be investigated whether this difference is related to the NLO behavior of the newest MadGraph version. In case the MadGraph v155 version is not up to NLO a difference in cross section is definitely expected.

Chapter 4

Preliminary Results

4.1 Results on reconstructed level

4.2 Background influence

Chapter 5

Understanding Results Obtained With MadWeight

5.1 Comparing the two used MadWeight versions

Since it was found that the latest MadWeight version (aMC@NLO) resulted in many events with weight equal to 0 it was decided to compare this version with the previous one (mc_perm). It is expected that both versions result in similar weights when identical events are considered, otherwise the used version of MadWeight would have a too large influence on the analysis result.
Therefore two events which could succesfully run in the newest MadWeight version were also calculated using the old version. The obtained weights and their uncertainty can be compared in the following table and all the relevant information can be found in:

$$/home/annik/Documents/Vub/PhD/ThesisSubjects/AnomalousCouplings \\ /CompareMWVersions\_May2014$$

Event Number	V_L value	aMC@NLO version		mc_perm version	
		Weight	Uncertainty	Weight	Uncertainty
1	1.5	9.7610^{-28}	4.0710^{-30}	1.4410^{-26}	4.1910^{-29}
	1.0	1.9210^{-28}	8.0410^{-31}	2.8410^{-27}	8.2810^{-30}
	0.5	1.2010^{-29}	5.0310^{-32}	1.7810^{-28}	5.1710^{-31}
2	1.5	1.8610^{-23}	1.1510^{-25}	1.7710^{-24}	1.2310^{-26}
	1.0	3.6510^{-24}	2.2510^{-26}	3.5710^{-25}	2.7510^{-27}
	0.5	2.2710^{-25}	1.4710^{-27}	2.2310^{-26}	1.7210^{-28}

Table 5.1: Weight obtained from MadWeight for two specific ttbar semi-muonic (+) events. For these events the V_R was fixed to its Standard Model expectation value, which is 0, while the V_L value was varied.

Comparing these values clearly shows that there is a significant difference between the two MadWeight versions which were considered in this analysis. With some effort a general difference of a facto 10 can be identified between the two versions, with a higher weight value for the older mc_perm MadWeight version.
However when showing the relative differences between the different weights it can be seen that the behavior of these two MadWeight versions is actually very similar. Therefore

the histograms below give firstly the actual weight value and secondly the weight value normalised to the weight corresponding to the coupling parameter $V_L = 0$.

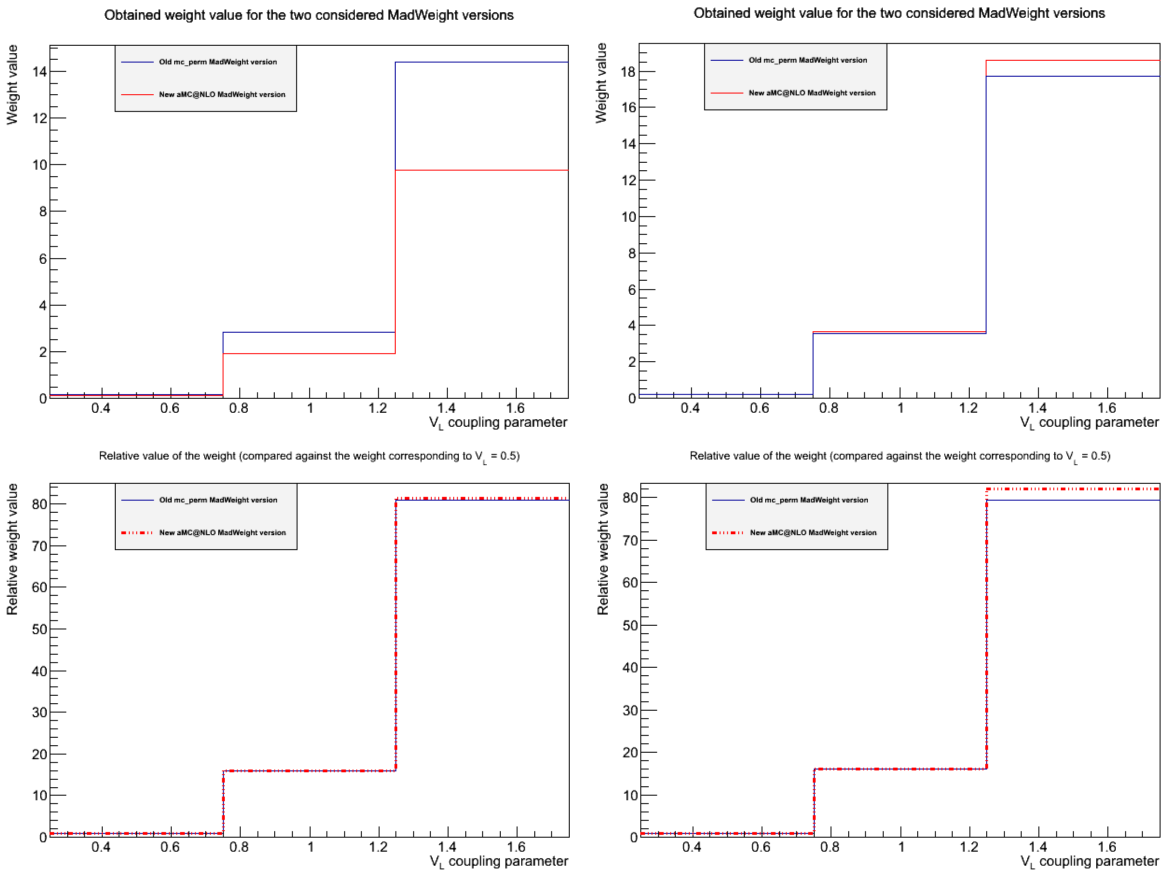


Figure 5.1: Distribution of the weights obtained from MadWeight for the two considered MadWeight versions (aMC@NLO and mc_perm) for two specific ttbar semi-muonic (+) events.

The above histograms clearly indicate that however the actual values of the weight significantly differ, the normalized results are almost identical.
This implies that the results obtained with MadWeight should always be normalized with respect to another MadWeight result (obtained using the same MadWeight version of course).

5.2 Individual weight distribution for condsidered grid-points

The first Likelihood distribution for the considered gridpoints gave rise to an unexpected result. In order to understand whether this strange behavior can be explained by a couple of events with a bad weight, the indiviudual weight distribution for a couple random events was studied. If these distributions indeed result in the correct behavior the events which influence the overal Likelihood distribution can be studied individually.
All the relevant information can be found in the following directory:

$$/home/annik/Documents/Vub/PhD/ThesisSubjects/AnomalousCouplings/UnderstandLikelihoodDistr\_July2014$$

And the creation of the MadWeight weights together with the python scripts making the corresponding histograms can be found in:

$$/localgrid/aolbrech/madweight/ttbarSemiMuPlus\_QED2/Events$$

The relevant python scripts are the following:

- **CalculateLikelihood.py** which simultaneously creates the histograms for the XS distribution, the raw and normalized Likelihood distribution, and if required the individual weight distributions for the selected events. All these histograms are saved in the Histos.root file.
- **RemoveZeroWeightEvents.py** which removes the events with weight equal to zero from the list and saves the non-zero events in a new .out file. On the other hand the events which failed the MadWeight computation are saved on a new .lhco file together with one succesful control event and are send again through MadWeight for a new weight calculation. This should be updated since the new computation of MadWeight also results in weights equal to zero for these failing events. So should be investigated what is different about these events ...

The first three histograms show the obtained Likelihood distribution for the different gridpoints which were considered. From these can be concluded that the behavior of the Likelihood normalized with the corresponding cross section divided by the Standard Model cross section is dominated by the distribution of the normalized cross section values. This means that any small deviations of the raw Likelihood values gets washed out by the multiplication with the normalized cross section values.

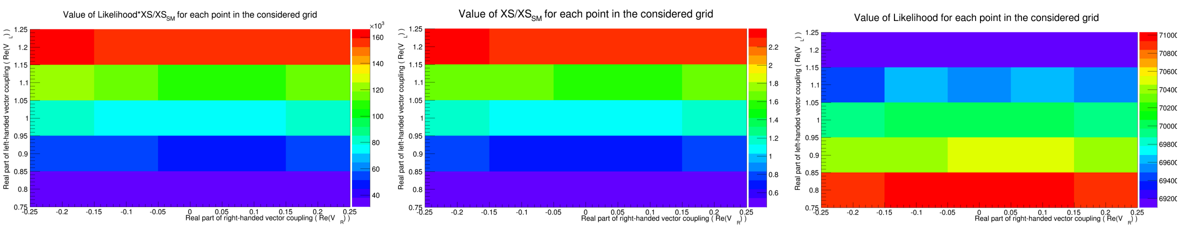


Figure 5.2: ...

The individual weight distribution for some random events can be found here.

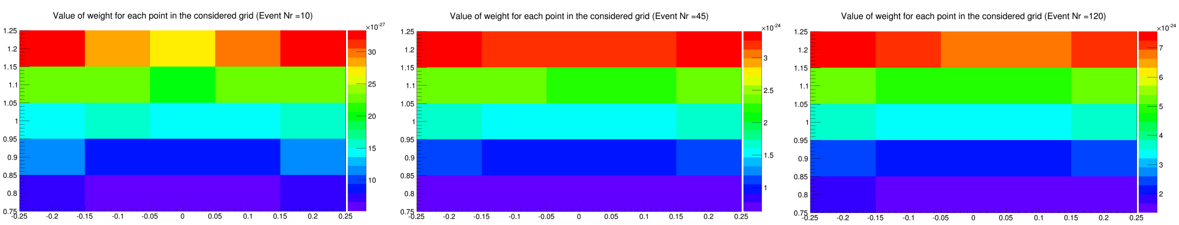


Figure 5.3: ...**Script should be ran again, and names for the axis should be added for these individual weight histograms!**

Chapter 6

Analyzing created FeynRules model

6.1 Normalized coupling parameters

In order to investigate the actual influence of the value of the coupling parameters on the kinematics of the event, the considered coupling parameters should be normalized to unitarity before any hard conclusions can be made. Therefore the configurations which should be investigated in large detail are the ones for which the width of the decay remains unchanged. This is explained in detail in one of the previous sections (2.3.4 on page 8).

6.2 Understanding parameters larger than 1

Before starting to look at the ttbar Monte Carlo and reweighting the events with MadWeight, the created model in FeynRules should be completely understood. Special care goes out to the behavior of the kinematic distributions for values of the coupling parameters larger than 1. Since the Standard Model expectation puts the real part of the left-handed vector coupling V_L almost equal to 1, simulation should be available around this Standard Model expectation value. Therefore the created model should be able to cope with coupling parameters larger than 1.

For this reason .lhco files were generated with MadGraph with the following configuration, as mentioned in 2.3.3:

$$\begin{aligned} Re(V_L) &\in [0.7,\ 1.3] \\ Re(V_R) &\in [-0.3,\ 0.3] \end{aligned}$$

For these generated events the main kinematic distributions have been investigated. No clear difference between the behavior below and above 1 has been found.

6.2.1 Performed checks

6.2.1.1 Cross section change

RVR RVL	-0.30	-0.20	-0.10	0.00	0.10	0.20	0.30
0.70	0.3775	0.3120	0.2724	0.3275	0.2632	0.2910	0.3436
0.80	0.5964	0.5097	0.4595	0.4385	0.4444	0.4816	0.5471
0.90	0.9011	0.7950	0.7308	0.7026	0.7103	0.3229	0.8356
1.00	1.3187	1.1874	1.1085	1.0711	1.0823	1.1317	1.2263
1.10	1.8700	1.7116	1.6154	1.5669	1.5763	1.6335	1.7522
1.20	2.5858	2.3996	2.2789	2.2200	2.2263	2.2983	2.4322
1.30	3.4896	3.2711	3.1278	3.0626	3.0655	3.1506	3.2983

Table 6.1: Cross sections for the different RVR-RVL couplings normalized to the SemiElMinus Standard Model Cross section (8.261 pb)

From this table can be seen that the cross section increases when the real component of V_L gets larger. The value of the right-handed vector coupling has only a minor influence on the cross section.

6.2.1.2 Relative increase visible in XS, but not in kinematic distributions

Since the observed model cannot represent physics at values larger than 1, one option is to look at specific fixed values of the real part of the left-handed and right-handed vector couplings. A proportional change in both of these coupling parameters should change the cross section values, but the kinematic should remain unchanged. Therefore the following configurations will be investigated:

$$
\begin{aligned}
Re(V_L) = 0.5 \quad &\& \quad Re(V_R) = 0.5 \rightarrow \; 2.07115 \; pb \\
Re(V_L) = 1.0 \quad &\& \quad Re(V_R) = 1.0 \rightarrow \; 33.1479 \; pb \\
Re(V_L) = 2.0 \quad &\& \quad Re(V_R) = 2.0 \rightarrow \; 530.027 \; pb
\end{aligned}
$$

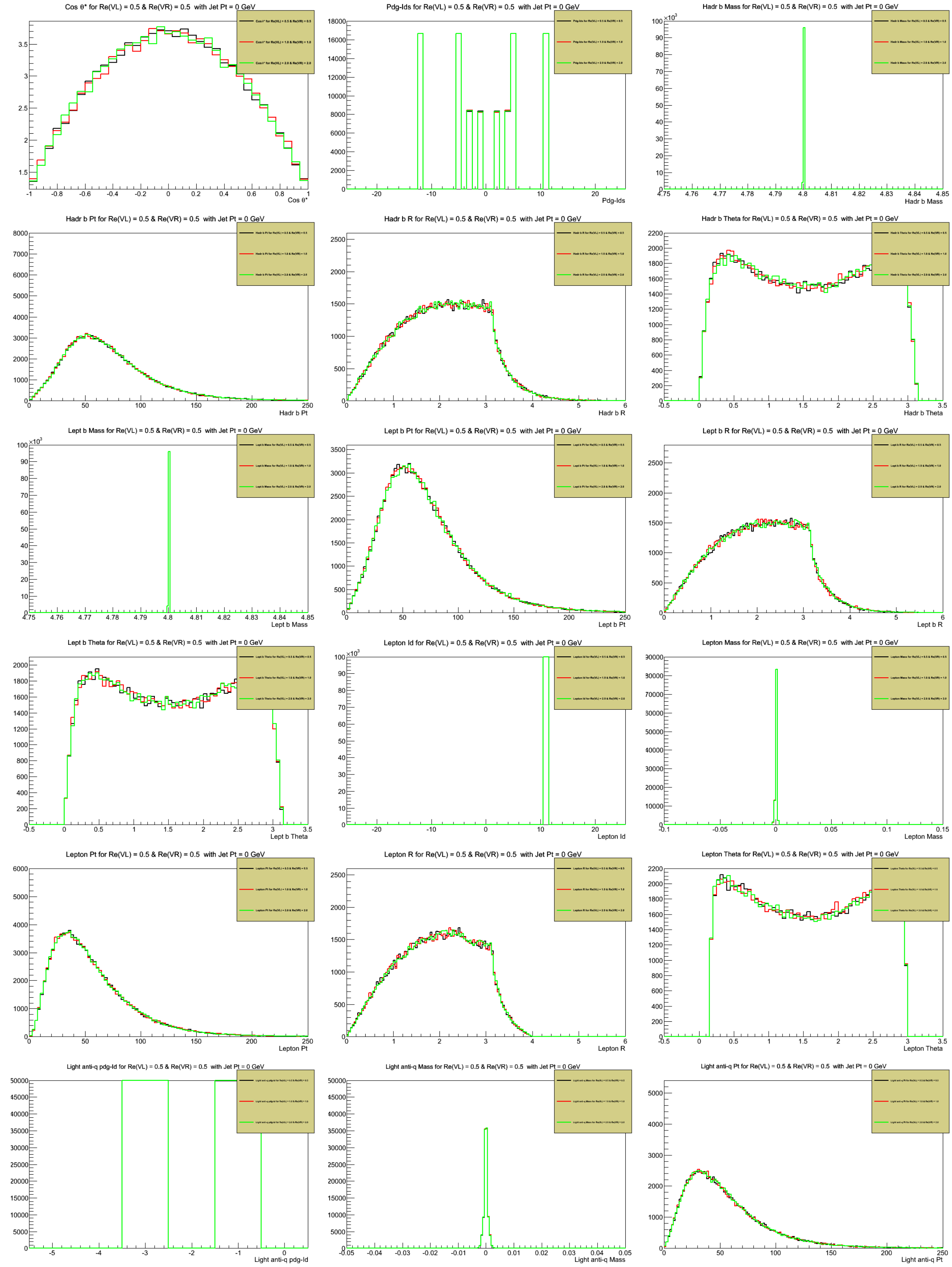
From the above numbers is clear that the Cross section becomes very large when the two coupling parameters increase. This can be understood quite easily since the second option allows much more decay options since the top quarks can decay both through the left-handed and the right-handed vector coupling side of the interaction vertex. The width of this configuration is not equal to the width of the Standard Model expectation and hence does not correspond to an actual physical solution. It is merely seen as a test of the model since the kinematics of the interaction should not differ.

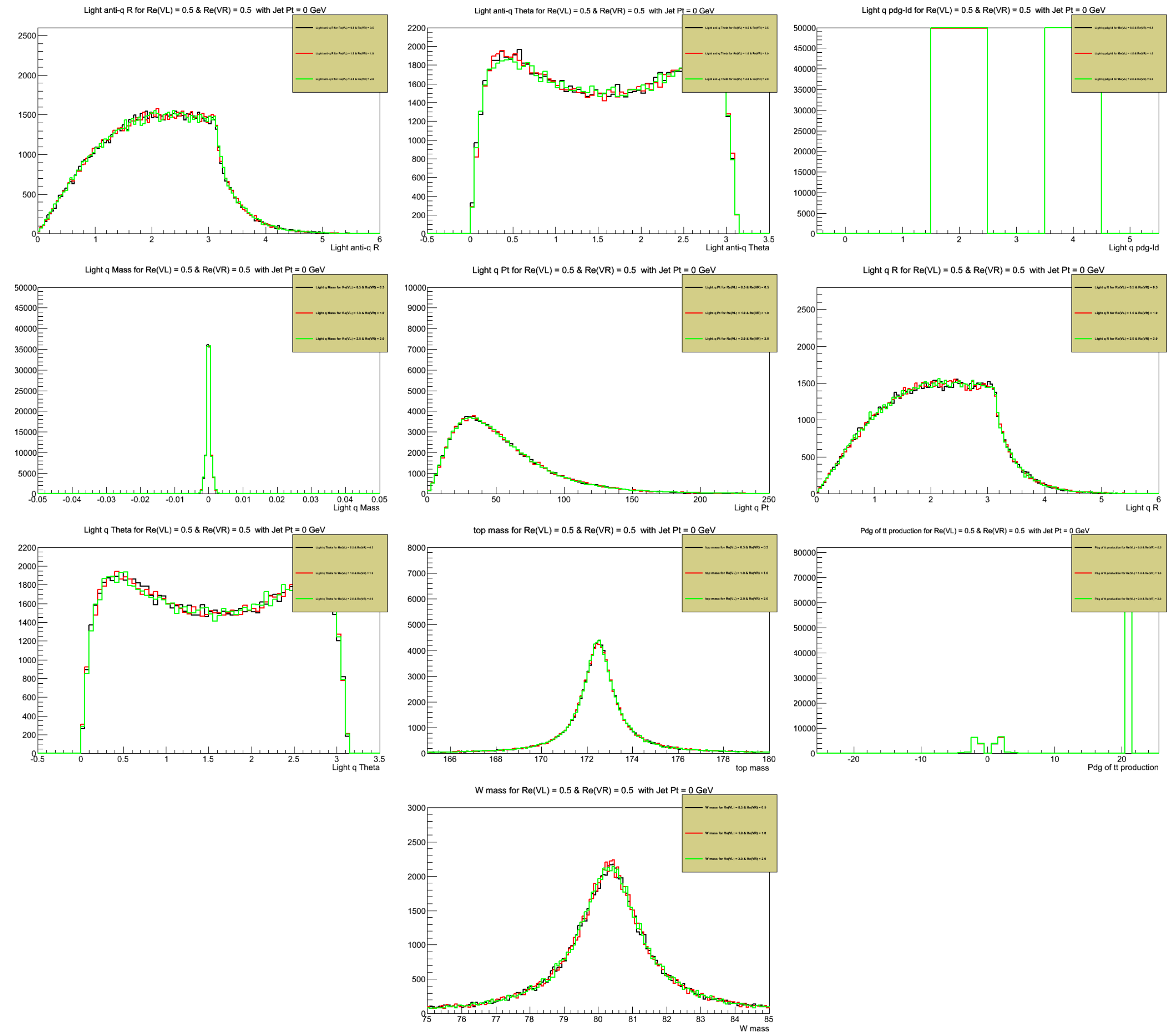
Looking at these plots clearly indicates that the kinematics doesn't change at all.
Hence the created FeynRules model is able to deal in a correct way with these coupling parameters larger than 1.

These MadGraph files have been created and can be found in:

$$
/user/aolbrech/AnomalousCouplings/MadGraph\_v155/MassiveLeptons/ \\
MadGraph5\_v1\_5\_5/Wtb\_ttbarSemiElMinus/RelativeChange
$$

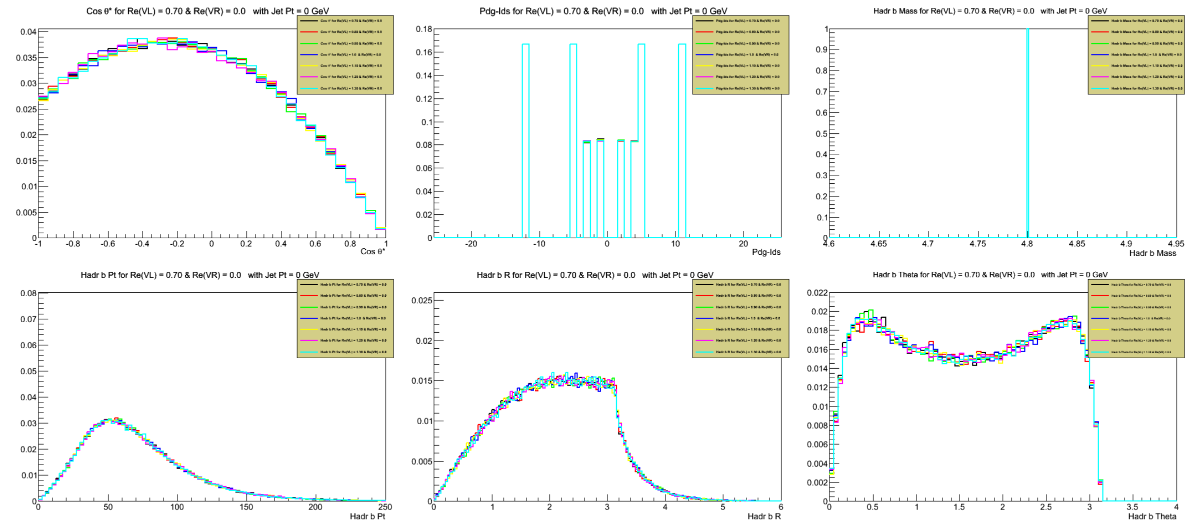
The distributions shown in this subsection are for fixed Jet Pt Cut value, set to 0. Also no Pt cut on the lepton was applied. Both coupling parameters have been changed proportionally.

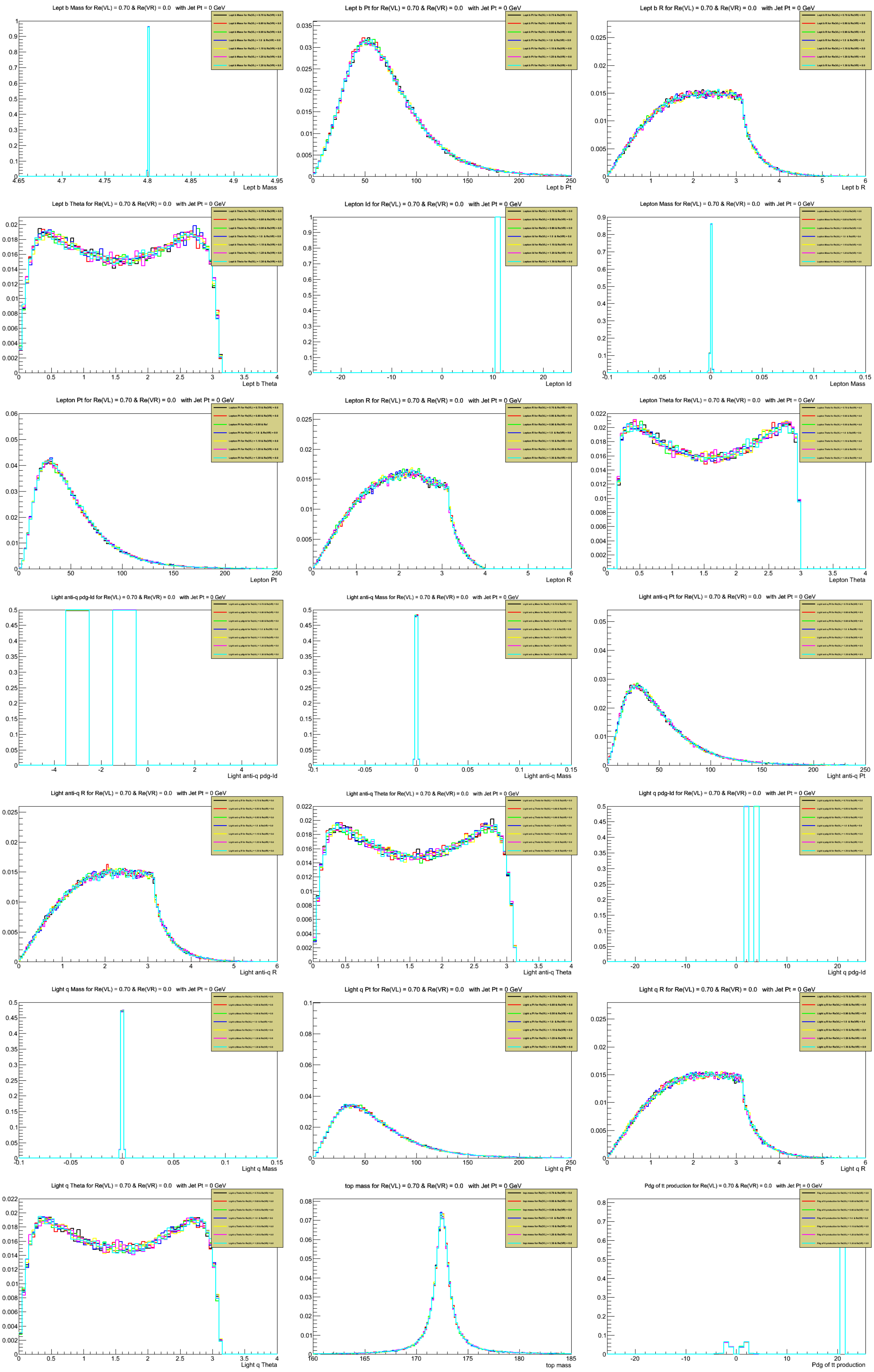


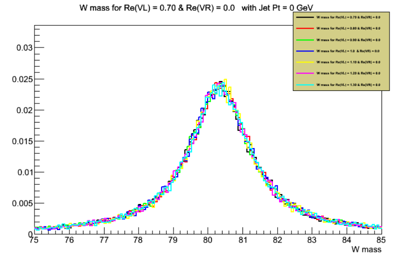


6.2.1.3 Model plots for fixed Pt Cut

The distributions shown in this subsection are for fixed Jet Pt Cut value, set to 0. Also no Pt cut on the lepton was applied. In this case the real part of the left-handed vector coupling has been varied between 0.7 and 1.3 in steps of 0.1 while the right-handed parameter has been fixed to its Standard Model value (0.0).




Lept b Mass for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lept b Mass
Lept b Pt for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lept b Pt
Lept b R for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lept b R
Lept b Theta for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lept b Theta
Lepton Id for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lepton Id
Lepton Mass for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lepton Mass
Lepton Pt for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lepton Pt
Lepton R for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lepton R
Lepton Theta for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Lepton Theta
Light anti-q pdg-Id for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light anti-q pdg-Id
Light anti-q Mass for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light anti-q Mass
Light anti-q Pt for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light anti-q Pt
Light anti-q R for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light anti-q R
Light anti-q Theta for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light anti-q Theta
Light q pdg-Id for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light q pdg-Id
Light q Mass for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light q Mass
Light q Pt for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light q Pt
Light q R for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light q R
Light q Theta for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Light q Theta
top mass for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
top mass
Pdg of tt production for Re(VL) = 0.70 & Re(VR) = 0.0 with Jet Pt = 0 GeV
Pdg of tt production



6.2.1.4 Model plots for varying Pt Cut

Script adapted, but error obtained when running the python script ...
Worked when everything was copied to the TestDir ...
Maybe the use of nohup gives the problem ...

Chapter 7

Event Selection

7.1 Choice of b-tag requirements

7.1.1 Signal vs background comparison

Since the complete event should be reconstructed as accurately as possible the 'signal' is represented by the case that all four particles are matched correctly while the 'background' events are the events for which at least one particle is matched wrongly. Events which are not matched using the JetPartonMatching algorithm (currently ptOrderedMinDist with dR of 0.3 is used) are not included in either of these two variables and are shown separately in the tables found below.
First the different b-tag options are compared for the four possible combinations which are still allowed, namely the interchange of both the two light quarks and the two b-jets.

Option (no χ^2 m_{lb})	all 4 correct	$\geq$ 1 wrong	$\frac{s}{\sqrt{b}}$	$\frac{s}{b}$	non-matched
2 L b-tags,	55967	50416	249.257	1.1101	225217
2 M b-tags,	49983	32012	279.361	1.56138	146633
2 M b-tags, light L-veto	39661	27751	238.081	1.42917	118389
2 T b-tags,	31444	16061	248.114	1.95779	78062
2 T b-tags, light M-veto	29160	15585	233.579	1.87103	73570
2 T b-tags, light L-veto	23159	14093	195.082	1.6433	59997

Table 7.1: Overview of correct and wrong reconstructed events for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

Option (no χ^2 m_{lb})	2 b's good	$\geq$ 1 b wrong	$\frac{s}{\sqrt{b}}$	$\frac{s}{b}$	non-matched
2 L b-tags,	78896	27487	475.873	2.8703	225217
2 M b-tags,	70073	11922	641.765	5.87762	146633
2 M b-tags, light L-veto	58778	8634	632.57	6.80774	118389
2 T b-tags,	43804	3701	720.036	11.8357	78062
2 T b-tags, light M-veto	41617	3128	744.11	13.3047	73570
2 T b-tags, light L-veto	34908	2344	721.018	14.8925	59997

Table 7.2: Overview of correct and wrong reconstructed b-jets for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

Option (no χ^2 m_{lb})	2 light good	$\geq$ 1 light wrong	$\frac{s}{\sqrt{b}}$	$\frac{s}{b}$	non-matched
2 L b-tags,	58707	47676	268.869	1.23137	225217
2 M b-tags,	50676	31319	286.351	1.61806	146633
2 M b-tags, light L-veto	40361	27051	245.398	1.49203	118389
2 T b-tags,	31690	15815	251.993	2.00379	78062
2 T b-tags, light M-veto	29466	15279	238.382	1.92853	73570
2 T b-tags, light L-veto	23456	13796	199.7	1.7002	59997

Table 7.3: Overview of correct and wrong reconstructed light jets for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

Above tables show a clear improvement for the 2 Tight b-tag case since suddenly the $\frac{s}{b}$ value goes up to almost 2. An additional benefit of the 2 Tight bt-tag case is that it as well will take care of a large part of the process backgrounds. Hence the motivation for selecting this b-tag option.
Comparing the normal 2 Tight b-tag case, where the light jets are defined as not being Tight b-tagged jets, against the two possibilities using a light-jet veto indicates no motivation to go for the light-veto option.

However the second table, showing only the reconstruction efficiency of the b-jets, shows an improvement when using a light-jet veto.
This means that, even if the number of selected events gets lower, the percentage of correct events does improve when asking for a light-jet veto since it ensures that mistagged b-jet events doesn't by mistake get identified as light jets. So events with a b-jet with a too low CSV discriminant now don't get selected anymore because these so-called light-jets don't survive the veto cut.

But, as confirmed by the last table, the efficiency of the light-jet reconstruction shows no distinct improvement. So for some reason the light-jets which are chosen with this veto method are not by definition the actual light-jets.

7.2 Use of m_{lb} χ^2 method for selecting the correct b-jets

Option (with χ^2 m_{lb})	all 4 correct	$\geq$ 1 wrong	$\frac{s}{\sqrt{b}}$	$\frac{s}{b}$	non-matched
2 L b-tags,	51055	55328	217.053	0.92277	514406
2 M b-tags,	45664	36331	239.572	1.25689	538794
2 M b-tags, light L-veto	36271	31141	205.539	1.16473	553377
2 T b-tags,	28580	18925	207.752	1.51017	573284
2 T b-tags, light M-veto	26513	18232	196.355	1.4542	576044
2 T b-tags, light L-veto	21073	16179	165.673	1.30249	583537

Table 7.4: Overview of correct and wrong reconstructed events for the different b-tags when a χ^2 m_{lb} - m_{qqb} method is applied

The two tables in this subsection require a different interpretation. The first one is actually a combined test of the χ^2 m_{lb} - m_{qqb} method and the optimal b-tag choice while

Option (with χ^2 m_{lb})	Correct b's	Wrong b's	$\frac{s}{\sqrt{b}}$	$\frac{s}{b}$	Correct option exists
2 L b-tags,	66882	12014	610.19	5.56701	78896
2 M b-tags,	59520	10553	579.395	5.6401	70073
2 M b-tags, light L-veto	49556	9222	516.04	5.37367	58778
2 T b-tags,	37013	6791	449.146	5.4503	43804
2 T b-tags, light M-veto	35015	6602	430.94	5.3037	41617
2 T b-tags, light L-veto	29139	5769	383.64	5.05096	34908

Table 7.5: Overview of the number of times the correct b-jet combination is chosen when using a χ^2 m_{lb} - m_{qqb} method

the second one is merely a performance check of the mlb method.
This second table had to be added since the first table can't be directly compared against the tables in the previous subsection since currently only one b-jet combination is left while in the previous case an iteration between the different b-jets was allowed. So the numbers will be lower by definition when the χ^2 m_{lb} - m_{qqb} method is applied.
Therefore the second table is relevant in order to select whether some clear gain can be obtained when applying this method since it represents the number of times the correct b-jet combination. If this percentage is higher than 50%, which is the case, an improvement is obtained compared to an iteration between the two possible combinations.

The first table indicates again that no real difference is found between the different b-tag options, but that as soon as 2 Tight b-tags are applied the $\frac{s}{b}$ improves slightly. Also the second table shows no difference in efficiency between the different b-tag options, but shows however that the use of this χ^2 m_{lb} - m_{qqb} method significantly enhances the correct choice of the b-jet combination. In about 84% of the cases the correct b-jet combination is chosen.

7.3 Histograms for event selection choice

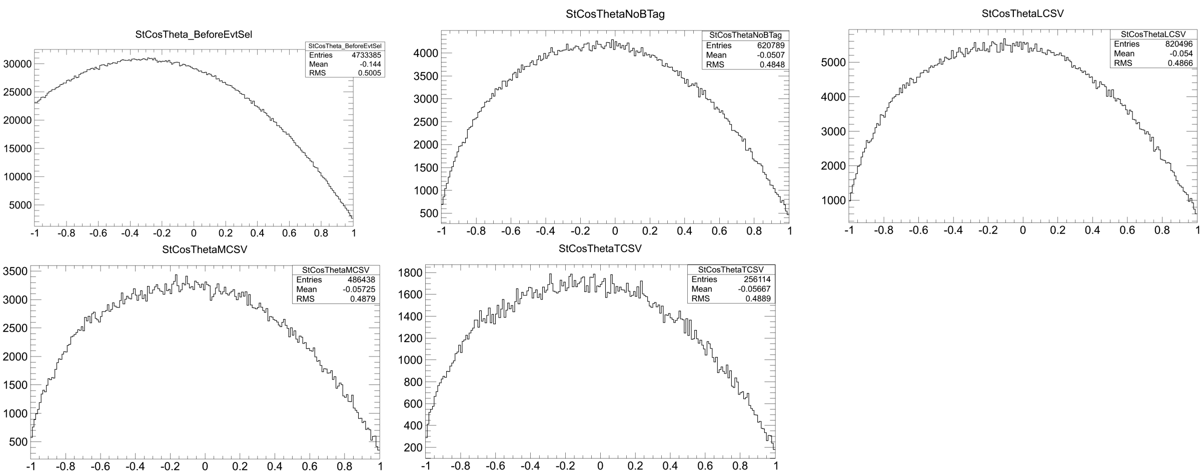


Figure 7.1: The $\cos\theta^*$ distribution for the different b-tag options (all of them imply double b-tag), which are not really influenced by the application of a b-tag. The only relevant distortion is caused by the event selection which is applied.

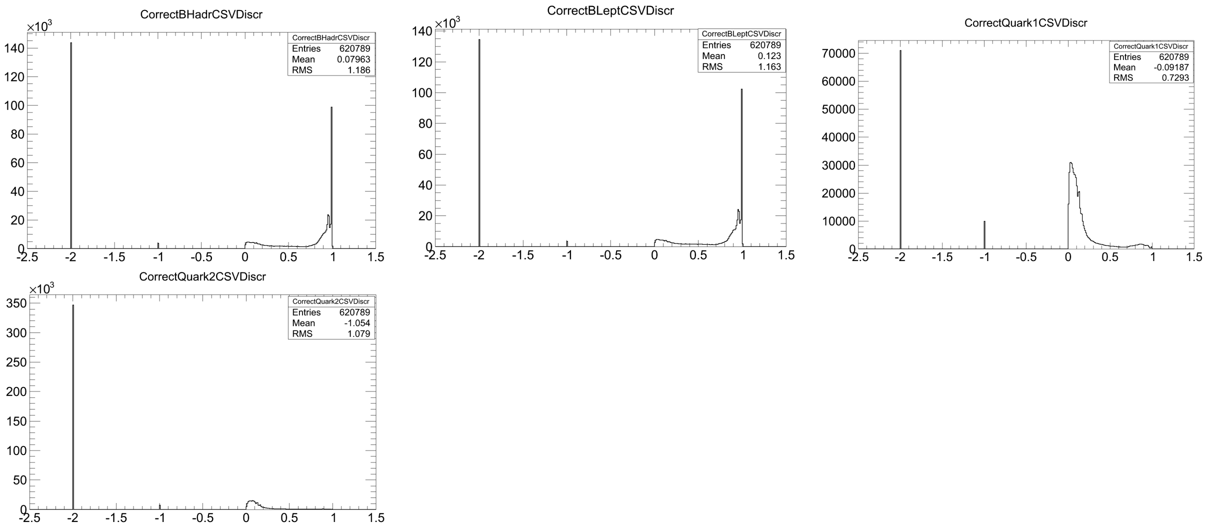


Figure 7.2: Distribution of the CSV discriminant for the different correctly matched quark-jet pairs. The value -2 is used to represent a non-matched jet. As expected the b-jets have a large peak at 1, so the Tight b-tag of 0.898 will not take away to many correct b-jets. The problem in the matching is clearly represented in distribution of the second quark which is only reconstructed in less than half of the cases.

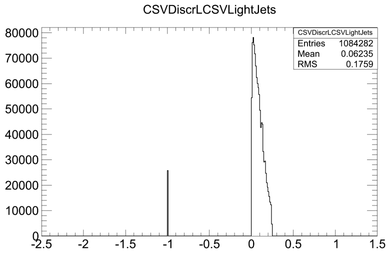


Figure 7.3: Distribution of the CSV discriminant of the selected light jets (all of them). **Add same histograms for Medium and Tight option, this will show how many of the light jets actually have a large CSV discriminant (maybe only focus on the two/three leading light jets)**

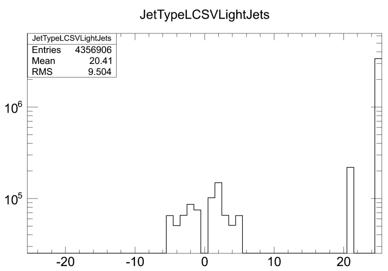


Figure 7.4: Jet type of the of the light jets (all of them). The value 25 means that the found light jet couldn't be matched to a Parton witht he JetPartonMatching method. **Same for M and T ... ?**

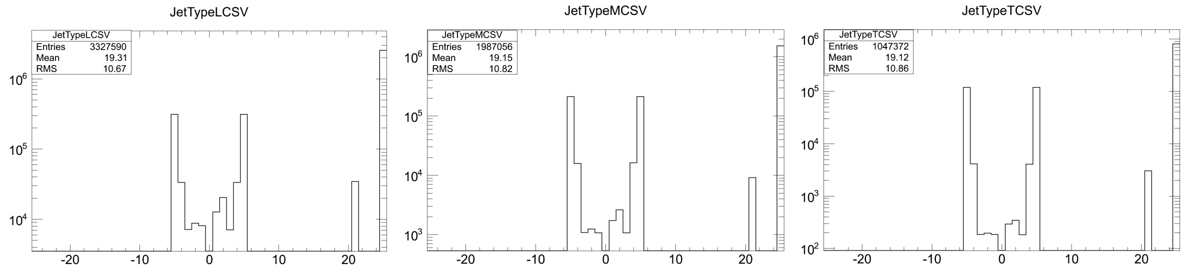


Figure 7.5: Jet type of the b-tagged jets (all of them) with the same convention for the non-matched jets.

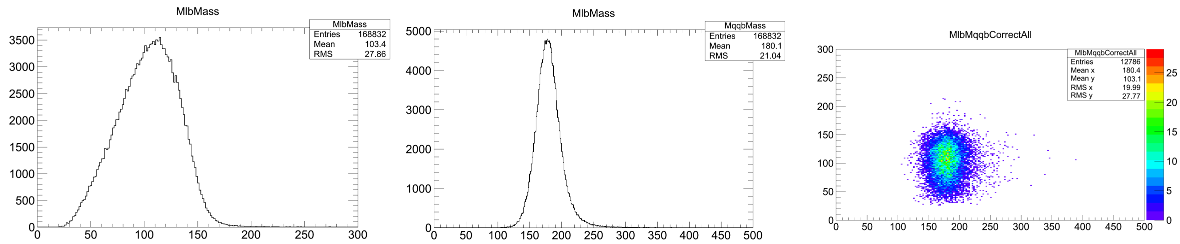


Figure 7.6: The first two histograms show the mass distribution for the correctly matched and reconstructed particles. A gaussian fit is applied in order to obtain both the m_{lb} and m_{qqb} mass and sigma. The last histogram shows the 2D behavior of these distributions.

7.4 Considering 2 or 3 light jets

In order to try to improve the signal efficiency it can be considered to add a third light jet to the particles which have to be taken into account for the jet selection. Adding this third jet will however result in 4 additional combinations which have to be considered so this method will only benefit when the χ^2 m_{lb} - m_{qqb} method is applied. Since MadWeight uses so much CPU sending the 6 possible combinations to MadWeight will not be beneficial.

7.4.1 Event selection numbers comparison

Option (no χ^2 m_{lb})	chosen jets are correct (%)	$\frac{s}{b}$	3rd jet is correct (%)
5 jet case, 2 T b-tags	76.2852	3.21677	70.244
4 jet case, 2 T b-tags	66.9091	2.02198	X

Table 7.6: Overview of correct and wrong reconstructed events for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

Option (no χ^2 m_{lb})	2 b's chosen correctly (%)	$\frac{s}{b}$	3rd jet is correct (%)
5 jet case, 2 T b-tags	90.5014	9.52784	64.5863
4 jet case, 2 T b-tags	92.3745	12.1138	X

Table 7.7: Overview of correct and wrong reconstructed b-jets for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

Option (no χ^2 m_{lb})	chosen light jets are correct (%)	$\frac{s}{b}$	3rd jet is correct (%)
5 jet case, 2 T b-tags	79.2892	3.8284	70.2241
4 jet case, 2 T b-tags	67.4722	2.0743	X

Table 7.8: Overview of correct and wrong reconstructed light jets for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

These three tables look at either a pure 5 jet case or a pure 4 jet case and comparing the numbers in each table with eachother is probably not extremely relevant. This because in the pure 5 jets case the matching requirement is loosened to two out of the three chosen light jets correctly matching with the partons. So in 1 out of 3 possibilities the so-called correct event will not be correct resulting in too high numbers for this case.

The first column gives the percentage how often the chosen jets are indeed the correct partons, hence in the 5-jet case the number of times the 5 possible jets match with the 4 correct partons. In the 4-jet case it implies that the four chosen jets are matched correctly with the 4 partons. The second column gives a similar value since the signal is defined as the number of times the matching was done correctly for the four partons while the background stands for the events where one of the matching is not succesful. The third column checks how often the third jet is one of the two correct light jets and compares it against the number in the first column. So it represents the number of times adding the third jet results in an improvement of the event reconstruction.

The numbers which are relevant in these tables are exactly these two last columns. These numbers show that in about 70% of the cases the third jet is actually one of the correct quarks. Therefore it can be decided from these numbers that in quite a lot of events, an improvement can be obtained when this third light jet is considered as well.

7.4.2 Mlb-algorithm numbers comparison

Option (with χ^2 m_{lb})	4 chosen jets are correct (%)	$\frac{s}{b}$	3rd jet is one of the 2 correct light jets (%)	3rd jet is chosen and correct (%)
5 jet case, 2 T b-tags	73.2413	2.73711	21.3059	89.2513
4 jet case, 2 T b-tags	76.9258	3.33384	0	-nan
Pure 5 jet case, 2 T b-tags	65.6683	1.91276	74.3984	89.2513

Table 7.9: Overview of correct and wrong reconstructed events for the different b-tags when a χ^2 m_{lb} - m_{qqb} method is applied

Option (with χ^2 m_{lb})	% b's correct	$\frac{s}{b}$		
5 jet case, 2 T b-tags	90.3229	9.33364	0	0
4 jet case, 2 T b-tags	90.7656	9.82911	0	0
Pure 5 jet case, 2 T b-tags	89.9057	8.90659	0	0

Table 7.10: Overview of the number of times the correct b-jet combination is chosen when using a χ^2 m_{lb} - m_{qqb} method

When the m_{lb} method is applied, the two first columns in the given tables represent similar quantities with the only difference that now the 4 jets which are actually chosen by the χ^2 m_{lb} - m_{qqb} method are considered. In this case the third column represents the number of times the third jet is chosen when the two light jets are matched correctly, implying that the third jet is one of the correct ones and that the second light jet is also correctly matched. The final column only looks at the third jet and puts no requirement on the matching of the second light jet. So it compares the number of times one of the chosen jets is the third jet and in how many cases this chosen third jet is one of the correct partons.

7.4.3 Compare efficiencies for 3^{rd} jet with 1^{st} and 2^{nd}

The obtained efficiency numbers for the 3^{rd} jet seemed to be rather high so to exclude any possible double-counting mistakes the percentages for the 1^{st} and 2^{nd} jet where also calculated and compared. Since the considered percentage represents the number of times the 3^{rd} jet corresponds to one of the actual light quarks should the sum of the three percentages not become any larger than 200%.
The percentages were calculated both before and after the application of the χ^2 algorithm and the results can be found in Tables 6.11 and 6.12. The first table shows the numbers before the application of the χ^2 algorithm implying that for correctly matched events 2 of the 3 light jets are correctly matched. The second table gives the results after the χ^2 minimization method.

	Number of events	Percentage (%)
Matched events	2455	
First jet	1596	65
Second jet	1740	70.9
Third jet	1574	64.1
Total	4910	200

Table 7.11: Number of times the first, second or third jet corresponds to one of the two correct light jets before the application of the χ^2 method.

	Number of events	Percentage (%)
Matched events	241	
First jet	162	67.2
Second jet	178	73.8
Third jet	142	58.9
Total	842	200

Table 7.12: Number of times the first, second or third jet corresponds to one of the two correct light jets after the application of the χ^2 method.

From these tables can be concluded that the obtained percentage of about 70% for the 3^{rd} jet is actually correct and that the result can be trusted. It also implies that in the 5-jet case (meaning that there actually is a third light jet) the three jets have a rather similar probability of being the correct jet.

7.4.4 Considering separate categories

In order to be certain whether the 3^{rd} light jet should be considered the considered events have been divided into two categories. The first only consists of events with exactly 2 light jets, hence 2 or more b-tagged jets[1] and 2 light jets, while the second category allows for more light jets. In case of multiple b-tagged jets only the two highest p_T jets are considered.
In the second category each event is treated in two separate ways. First the event is seen as a pure 4-jet event implying that only the two leading light jets are kept while for the second approach the third light jet is also included in the χ^2 algorithm resulting in 6 possible solutions.

For these three cases the matching and χ^2 minimization efficiency have been compared in order to ensure that the most efficient event selection will be used. The results can be found in Table 6.13 and indicates that including the third light jet doesn't result in a large gain of efficiency. On the contrary, including events with more than 4 jets but discarding the third light jet results in a significant decrease of efficiency.

	N(2 light jets)	N(2+ light jets)
# events	9328	6436
# matched events	4018	3319
# good combi chosen	3273	788 – 1604

Table 7.13: Number of events, number of matched events and number of events for which the correct jet combination is chosen using the χ^2 $m_{lb} - m_{qqb}$ algorithm for the two considered categories. The first number in the right-hand bottom corner respresents the number of good combinations chosen when the event is treated as a 4-jet event while the second number is for the treatment of a 5-jet event. An event is considered as matching if the 4 jets corresponding to the generator event are included in the collection of selected jets.

The obtained results seem rather suprising since it implies that only asking the 4 leading jets results in the worst efficiencies.
However it should be noted that for this configuration the 2 Tight b-tag constraint might influence this result. It could be possible that for looser b-tag requirements the mis-identification of the two b-jets results in a worse "combination choice" efficiency.

	Only 4-jet events	All, but treated as 4-jet	All, but treated as 5-jet
% matched events	43.07	46.5	46.5
% good combi	81.4	55.3	66.5

Table 7.14: Percentages for matching the reconstructed event with the generated and for selecting the good combination using the χ^2 $m_{lb} - m_{qqb}$ algorithm.

Is it possible to understand these results using the percentages for the first, second and third jet being correct .. ?

[1]The number of events with a third b-tagged jet is extremely small and will probably not really influence the efficiency as can be seen from the plots shown in 6.4.4.1.

When treating a 5-jet event as a 4-jet event the correct jet would have been the third, discarded, jet in $\frac{1}{3}$ of the cases which explains the significant reduction. However the difference between treating the event as a 4-jet or a 5-jet event doesn't result in a large difference. It seems that in much of the N(2+ light jet) cases the correct jet is actually still one of the following jets and is not included in the three leading light jets.

7.4.4.1 Histograms for number of selected, b-tagged and light jets

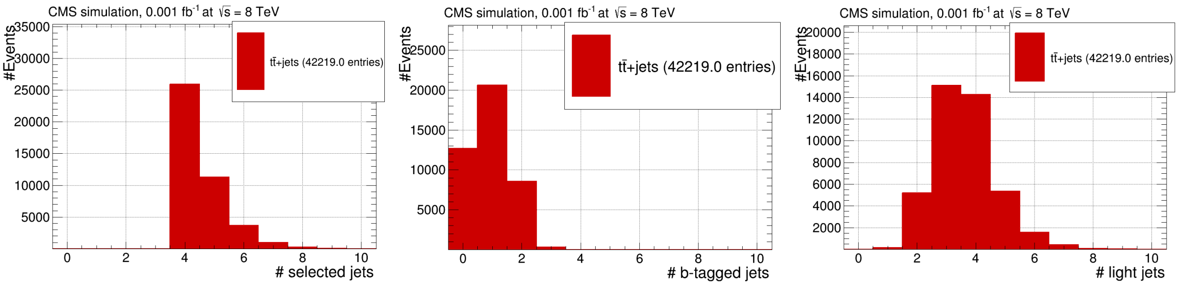


Figure 7.7: Number of selected, b-tagged and light jets, respectively, before requiring two Tight b-tags. These distributions are for the muon channel only.

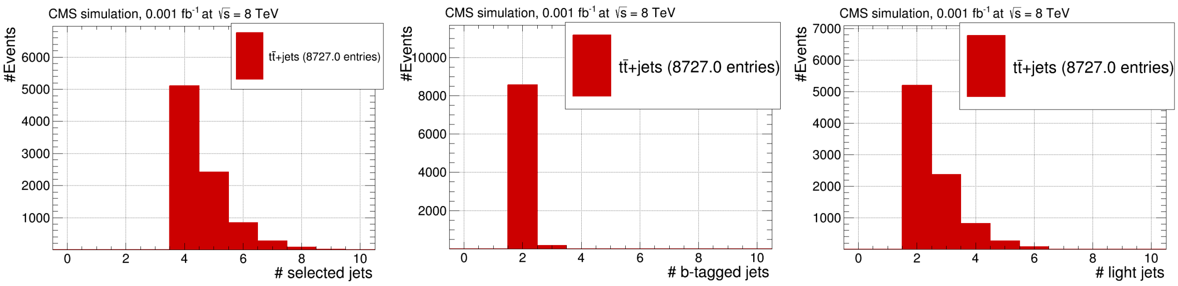


Figure 7.8: Number of selected, b-tagged and light jets, respectively, after requiring two Tight b-tags. These distributions are for the muon channel only.

7.5 Studying optimal cut on χ^2 value

The tables shown here indicate that the application of a cut on the χ^2 value of the m_{lb} - m_{qqb} method doesn't change the rates of good and wrong chosen events. Therefore it will only reduce the number of selected events, and hence reduce the needed CPU time, but still keep an as pure sample as obtained without any cut.
Since there is no real difference visible between setting the cut value to 3 or 5 it is advisable to use the cut value of 3 to reduce the number of selected events.

Option (with χ^2 m_{lb})	4 chosen jets are correct (%)	$\frac{s}{b}$	3rd jet is one of the 2 correct light jets (%)	3rd jet is chosen and correct (%)
5 jet case, 2 T b-tags	73.1485	2.72419	21.5302	89.0292
4 jet case, 2 T b-tags	76.89	3.32712	0	-nan
Pure 5 jet case, 2 T b-tags	65.5254	1.90068	74.5263	89.0292

Table 7.15: Overview of correct and wrong reconstructed events for the different b-tags when a χ^2 m_{lb} - m_{qqb} method is applied

Option (with χ^2 m_{lb})	% b's correct	$\frac{s}{b}$		
5 jet case, 2 T b-tags	90.3451	9.35743	0	0
4 jet case, 2 T b-tags	90.8573	9.93767	0	0
Pure 5 jet case, 2 T b-tags	89.8465	8.84884	0	0

Table 7.16: Overview of the number of times the correct b-jet combination is chosen when using a χ^2 m_{lb} - m_{qqb} method

Option (with χ^2 m_{lb})	4 chosen jets are correct (%)	$\frac{s}{b}$	3rd jet is one of the 2 correct light jets (%)	3rd jet is chosen and correct (%)
5 jet case, 2 T b-tags	72.8939	2.68921	21.7785	88.6241
4 jet case, 2 T b-tags	77.0226	3.3521	0	-nan
Pure 5 jet case, 2 T b-tags	64.7571	1.83745	74.192	88.6241

Table 7.17: Overview of correct and wrong reconstructed events for the different b-tags when a χ^2 m_{lb} - m_{qqb} method is applied

Option (with χ^2 m_{lb})	% b's correct	$\frac{s}{b}$		
5 jet case, 2 T b-tags	90.1586	9.16114	0	0
4 jet case, 2 T b-tags	90.8239	9.89789	0	0
Pure 5 jet case, 2 T b-tags	89.4472	8.47619	0	0

Table 7.18: Overview of the number of times the correct b-jet combination is chosen when using a χ^2 m_{lb} - m_{qqb} method

Option (with χ^2 m_{lb})	4 chosen jets are correct (%)	$\frac{s}{b}$	3rd jet is one of the 2 correct light jets (%)	3rd jet is chosen and correct (%)
5 jet case, 2 T b-tags	72.4364	2.62798	21.6013	87.5833
4 jet case, 2 T b-tags	77.3908	3.42298	0	-nan
Pure 5 jet case, 2 T b-tags	62.9538	1.69934	73.7615	87.5833

Table 7.19: Overview of correct and wrong reconstructed events for the different b-tags when a χ^2 m_{lb} - m_{qqb} method is applied

Option (with χ^2 m_{lb})	% b's correct	$\frac{s}{b}$		
5 jet case, 2 T b-tags	90.233	9.23861	0	0
4 jet case, 2 T b-tags	90.9792	10.0854	0	0
Pure 5 jet case, 2 T b-tags	89.5385	8.55882	0	0

Table 7.20: Overview of the number of times the correct b-jet combination is chosen when using a χ^2 m_{lb} - m_{qqb} method

7.5.1 χ^2 required to be smaller than 5

7.5.2 χ^2 required to be smaller than 3

7.5.3 χ^2 required to be smaller than 1

7.6 Influence of using p_T cuts suggested by the TOP reference selection Twiki

The TOP Reference Selection Twiki, and the different subgroup Twikis, suggest to use different event selection requirements than used before. The values suggested can be found in the table below, together with the values which were used for producing the tables and figures given above.
In this section the different results will be compared for these two event selections in order to ensure that the influence of this event selection can be ignored. If this is not the case, all the above tables have to be replaced and updated (which will however happen on a larger time-scale since these will be the correct values which will be used further in this analysis). The main goal is to quickly check whether the conclusion obtained from the above figures and tables still remain valid.

	Old values	TOP RefSel values
selected jets	40 GeV	30 GeV
selected muons	25 GeV	26 GeV
selected electrons	32 GeV	30 GeV
veto muons	10 GeV	10 GeV
veto electrons	10 GeV	20 GeV

In the following table the number of selected events after the different event selection cuts which are applied in this analysis can be found. The left-handed columns contain the information when the old p_T cuts are applied while the right-handed columns gives the number of selected events when the recommendations of the Top Reference Selection Twiki are followed.

From the comparison of the two columns can be seen that lowering the p_T cut on the selected jets significantly improves the percentage of selected events, especially because of the improved selection efficiency for the third and fourth jet. This can easily be understood from the p_T distribution histogram for the different leading jets which can be found below. These distributions show that in the case of the fourth jet, moving the p_T cut from 30 GeV to 40 GeV cuts away the largest percentage of this fourth jet since its distribution is peaked at a lower p_T value. The influence is much lower for the first and

Table 7.21: Event selection table before (left) and after (right) the pT cuts were updated to the ones suggested by the TOP reference twiki, for Semi-elec channel $t\bar{t} + jets$. (19600.8 pb^{-1} of int. lumi.)

	Old p_T cuts		New p_T cuts	
preselected	1.91516e+07		1.91516e+07	
trigged	3.52179e+06	18.4 %	3.52179e+06	18.4 %
Good PV	3.52179e+06	100 %	3.52179e+06	100 %
1 selected electron	2.76014e+06	78.4 %	2.85992e+06	81.2 %
Veto muon	2.75335e+06	99.8 %	2.8529e+06	99.8 %
Veto 2nd electron from Z-decay	2.7483e+06	99.8 %	2.84766e+06	99.8 %
Conversion veto	2.7483e+06	100 %	2.84766e+06	100 %
$\geq$ 1 jets	2.72998e+06	99.3 %	2.84406e+06	99.9 %
$\geq$ 2 jets	2.50625e+06	91.8 %	2.77442e+06	97.6 %
$\geq$ 3 jets	1.74245e+06	69.5 %	2.34978e+06	84.7 %
$\geq$ 4 jets	0.753281e+06	43.2 %	1.38732e+06	59.0 %

second jet because their distrubition is maximal around 100 GeV so changing this p_T cut only affects the left side of the tail of the distribution.

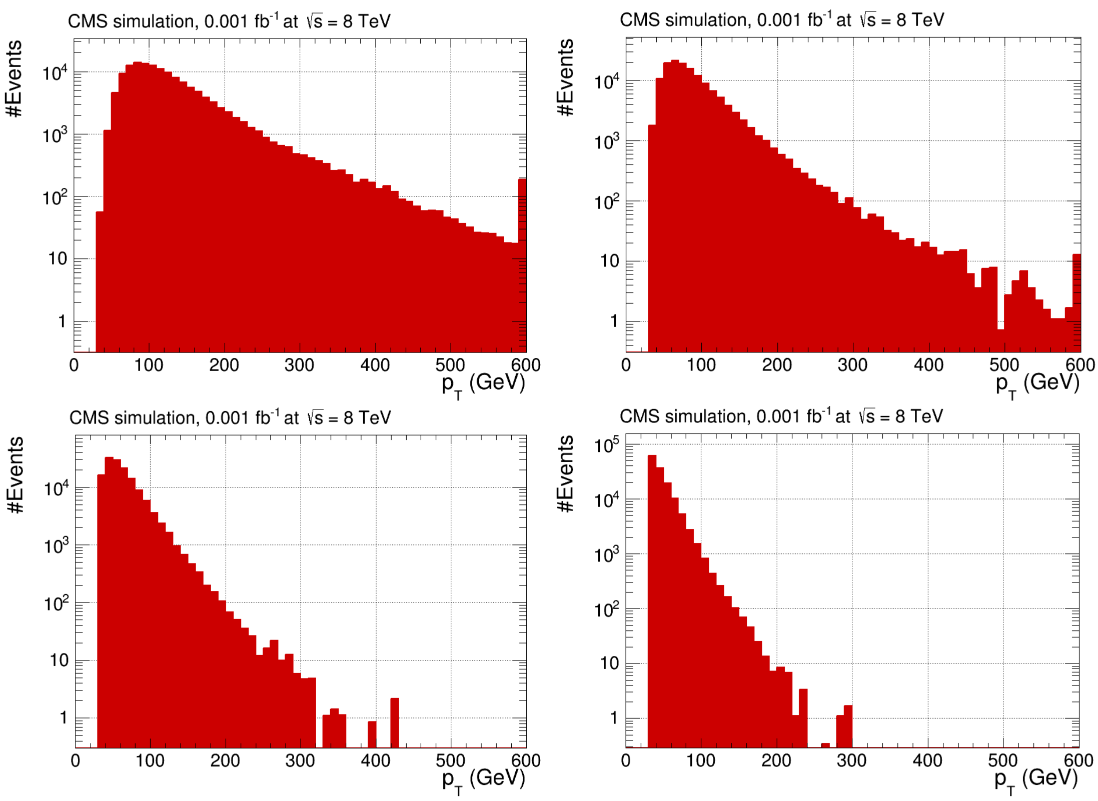


Figure 7.9: P_T distributions for the first, second, third and fourth jet respectively. All four histograms are for the semi-electronic decay channel. (Information about the Cross Section and number of events should be ignored for the moment. Correct Cross Section value is not yet used ...)

Table 7.22: Event selection table before (left) and after (right) the pT cuts were updated to the ones suggested by the TOP reference twiki, for Semi-muon channel $t\bar{t} + jets$. (19600.8 pb^{-1} of int. lumi.)

	Old p_T cuts		New p_T cuts	
preselected	1.91516e+07		1.91516e+07	
trigged	4.01265e+06	20.9 %	4.01265e+06	20.9 %
Good PV	4.01265e+06	100 %	4.01265e+06	100 %
1 selected muon	3.46903e+06	86.4 %	3.40822e+06	84.9 %
Veto 2nd muon	3.46369e+06	99.8 %	3.40291e+06	99.8 %
Veto electron	3.45244e+06	99.7 %	3.39184e+06	99.7%
$\geq$ 1 jets	3.4281e+06	99.3 %	3.38698e+06	99.9 %
$\geq$ 2 jets	3.14456e+06	91.7 %	3.30209e+06	97.5 %
$\geq$ 3 jets	2.19456e+06	69.8 %	2.80066e+06	84.8 %
$\geq$ 4 jets	944572	43.0 %	1.65345e+06	59.0 %

The same table can be created for the semi-muonic decay channel which shows a similar result.

7.6.1 Influence on choice of b-tag option and use of χ^2 m_{lb} - m_{qqb} method

With these new values considered for the event selection requirements, the obtained percentages and $\frac{s}{b}$ values should be compared again. The values wich are obtained using these new p_T cuts can be found in the following tables.
The first three tables show the results before the use of a χ^2 m_{lb} - m_{qqb} method while the last two tables give the obtained numbers when this χ^2 m_{lb} - m_{qqb} method is applied.

Analyzing the numbers in this tables clearly indicates that the obtained results for the old p_T cuts correspond to the values obtained earlier, which implies that the obtained values can easily be compared against each other and stil represent the same variables. The small differences which are however visible can be easily explained by statistical deviations.
Comparing the results obtained using the new p_T cut values with the ones using the old cut values clearly indicates that the old values resulted in slightly better selection efficiency and $\frac{s}{b}$ value. Both for the selection efficiency of the b-quark jets and the light jets a higher percentage is found when using the old p_T cut values.
A positive remark corresponding to the new p_T cut values is that the behavior of the different considered b-tag options is similar. The 2 Tight b-tag case without any veto on the light jets results in the highest selection efficiency and $\frac{b}{s}$ value. Hence the choice of the used b-tag option does not need to be updated.

Even the use of a χ^2 m_{lb} - m_{qqb} method doesn't improve the selection efficiency and $\frac{s}{b}$ values when changing to the newer p_T cut values. However compared to the previous case when no χ^2 m_{lb} - m_{qqb} method is applied the difference between the two p_T cut options becomes slightly less significant (from 18 % to 13 % difference). Still no improvement can be found impying that lowering the p_T cut on the jets only increases the number of selected events but doesn't insures selection more good events, on the contrary the selection efficiencies decreases even.

	Option (no χ^2 m_{lb})	all 4 correct	≥ 1 wrong	correct (%)	$\frac{s}{b}$	non-matched
New p_T cuts	L b-tags	31797	36132	46.8092	0.880023	103988
	2 M b-tags	28380	21485	56.9137	1.32092	63717
	2 M b-tags, light L-veto	21578	18821	53.4122	1.14649	50549
	2 T b-tags	17858	10684	62.5674	1.67147	33551
	2 T b-tags, light M-veto	16574	10496	61.2264	1.57908	31778
	2 T b-tags, light L-veto	12664	9552	57.004	1.3258	25440
Old p_T cuts	2 L b-tags	15285	13577	52.9589	1.1258	61307
	2 M b-tags	13480	8590	61.0784	1.56927	39800
	2 M b-tags, light L-veto	10718	7461	58.9581	1.43654	32048
	2 T b-tags	8555	4231	66.9091	2.02198	21234
	2 T b-tags, light M-veto	7915	4097	65.8924	1.9319	19946
	2 T b-tags, light L-veto	6327	3701	63.0933	1.70954	16156

Table 7.23: Overview of correct and wrong reconstructed events for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

	Option (no χ^2 m_{lb})	2 b's correct	$\geq$ 1 b wrong	b's correct (%)	$\frac{s}{b}$	non-matched
New p_T cuts	2 L b-tags	47926	20003	70.5531	2.39594	103988
	2 M b-tags	42045	7820	84.3177	5.3766	63717
	2 M b-tags, light L-veto	34622	5777	85.7001	5.99308	50549
	2 T b-tags	26117	2425	91.5037	10.7699	33551
	2 T b-tags, light M-veto	24974	2096	92.2571	11.9151	31778
	2 T b-tags, light L-veto	20585	1631	92.6584	12.6211	25440
Old p_T cuts	2 L b-tags	21456	7406	74.34	2.89711	61307
	2 M b-tags	18863	3207	85.469	5.88182	39800
	2 M b-tags, light L-veto	15859	2320	87.238	6.83578	32048
	2 T b-tags	11811	975	92.3745	12.1138	21234
	2 T b-tags, light M-veto	11203	809	93.2651	13.848	19946
	2 T b-tags, light L-veto	9424	604	93.9769	15.6026	16156

Table 7.24: Overview of correct and wrong reconstructed b-jets for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

	Option (no χ^2 m_{lb})	2 light good	$\geq$ 1 light wrong	light correct (%)	$\frac{s}{b}$	non-matched
New p_T cuts	2 L b-tags	33731	34198	49.6563	0.986344	103988
	2 M b-tags	28893	20972	57.9424	1.37769	63717
	2 M b-tags, light L-veto	22118	18281	54.7489	1.20989	50549
	2 T b-tags	18030	10512	63.1701	1.71518	33551
	2 T b-tags, light M-veto	16788	10282	62.017	1.63276	31778
	2 T b-tags, light L-veto	12868	9348	57.9222	1.37655	25440
Old p_T cuts	2 L b-tags	16015	12847	55.4882	1.24659	61307
	2 M b-tags	13688	8382	62.0208	1.63302	39800
	2 M b-tags, light L-veto	10938	7241	60.1683	1.51056	32048
	2 T b-tags	8627	4159	67.4722	2.0743	21234
	2 T b-tags, light M-veto	8005	4007	66.6417	1.99775	19946
	2 T b-tags, light L-veto	6409	3619	63.911	1.77093	16156

Table 7.25: Overview of correct and wrong reconstructed light jets for the different b-tags without the use of a χ^2 m_{lb} - m_{qqb} method

	Option (with χ^2 m_{lb})	all 4 correct	$\geq$ 1 wrong	4 chosen jets correct (%)	$\frac{s}{b}$	non-matched
New p_T cuts	2 L b-tags	24611	43318	77.4004	0.568147	103988
	2 M b-tags	21813	28052	76.8605	0.777592	63717
	2 M b-tags, light L-veto	16731	23668	77.5373	0.706904	50549
	2 T b-tags	13681	14861	76.6099	0.920598	33551
	2 T b-tags, light M-veto	12674	14396	76.4692	0.880383	31778
	2 T b-tags, light L-veto	9757	12459	77.0452	0.783129	25440
Old p_T cuts	2 L b-tags	11893	16969	77.8083	0.700866	61307
	2 M b-tags	10422	11648	77.3145	0.894746	39800
	2 M b-tags, light L-veto	8369	9810	78.0836	0.853109	32048
	2 T b-tags	6581	6205	76.9258	1.0606	21234
	2 T b-tags, light M-veto	6080	5932	76.8162	1.02495	19946
	2 T b-tags, light L-veto	4903	5125	77.4933	0.956683	16156

Table 7.26: Overview of correct and wrong reconstructed events for the different b-tags when a χ^2 m_{lb} - m_{qqb} method is applied

	Option (with χ^2 m_{lb})	Correct b's	Wrong b's	% b's correct	$\frac{s}{b}$	Correct option exists
New p_T cuts	2 L b-tags	28190	3607	88.6562	7.81536	31797
	2 M b-tags	25143	3237	88.5941	7.76738	28380
	2 M b-tags, light L-veto	19183	2395	88.9007	8.0096	21578
	2 T b-tags	15736	2122	88.1174	7.41565	17858
	2 T b-tags, light M-veto	14586	1988	88.0053	7.33702	16574
	2 T b-tags, light L-veto	11178	1486	88.266	7.52221	12664
Old p_T cuts	2 L b-tags	13935	1350	91.1678	10.3222	15285
	2 M b-tags	12306	1174	91.2908	10.4821	13480
	2 M b-tags, light L-veto	9808	910	91.5096	10.778	10718
	2 T b-tags	7765	790	90.7656	9.82911	8555
	2 T b-tags, light M-veto	7172	743	90.6128	9.65276	7915
	2 T b-tags, light L-veto	5743	584	90.7697	9.8339	6327

Table 7.27: Overview of the number of times the correct b-jet combination is chosen when using a χ^2 m_{lb} - m_{qqb} method

Chapter 8

Event corrections and reconstruction

8.1 Which event corrections should be applied?

8.1.1 Trigger choice

Two possible triggers exist, namely the so-called SingleLepton triggers or the CrossTriggers. The former kind only uses the information of the lepton while the latter one combines the lepton information together with the jets present in the event. These second kind of triggers were developed since it was expected that the p_T cut which had to be applied on this SingleLepton trigger would become too high to be physically relevant. However it has been found that the applied p_T cut doesn't differ that much between these two triggers, especially in the muon case the difference is almost negligible.

Since the scale factors which should be applied are much more complex in the case of the CrossTriggers preference is given to these SingleLepton triggers.
Minor disadvantage of these kind of triggers is that no information can be found on the TOP Twiki page (`https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopTrigger`) and currently no other Twiki page has been found with similar information ... Trigger contacts of TOP group prefer the use of the CrossTriggers since a lot of time and effort has been put into the development of these triggers. Probably the reason why no clear documentation is found on the Top Twiki page.
But James is using the same triggers in his analysis so all these triggers, and the different versions, can be found in the following two analyzer files:

- `https://github.com/TopBrussels/FourTops/blob/master/FourTop_EventSelection.cc` which contains information about the muon event selection and used triggers.
- `https://github.com/TopBrussels/FourTops/blob/master/FourTop_EventSelection_El.cc` which contains the same information but about the electron case.

8.1.2 Lumi- or PileUp Reweighting

The choice of the root file which should be used for the LumiReweighting is related to the Monte Carlo samples which will be used. If the Summer 12 MC samples, which describe the 8 TeV data, are used the *S10* file should be used.
LumiReweighting is actually taking into account the influence of PileUp but can't be

called this way when publishing results since a systematic influence of PileUp is no physical variable. However the luminosity of MinimumBias events is. This systematic influence can be calculated by comparing the influence of the up and down part.

8.1.3 Lepton Scale Factors

Lepton scale factors should be used and can be found in the code of James.
In the muon case the scale factors correspond to the entire dataset (ABCD) since different scale factors have been obtained for the different runs. Therefore caution should be applied when only running over a limited range of data in order to avoid to run only over the A part. However the influence shouldn't be too large ...

In the electron case the scale factors have been hard coded and do not vary for different parts of the data sample. **Therefore it should be checked whether the hard-coded numbers which are used in James code are still the correct ones which should be used. Hence a twiki with this information should be found ...**

8.1.4 Jet Energy Correction factors

Information available in the analyzer of James is up-to-date and can be copied.

8.1.5 Jet corrections (on the fly ...)

In James analyzer code Jet corrections are applied on the fly which is no problems, since even if the considered sample has these corrections already applied, the method of incorporating these jet corrections uses the gen information. So this avoids that these jet corrections would be applied twice ...

8.2 Choice of Monte Carlo samples

Same samples as James can be used, which are the latest Jan22 rereco samples. A full overview of these samples can be found on: `https://docs.google.com/spreadsheet/ccc?key=0Apc0aJdnaVjSdFVaLVU2dlk4RDZHcjlaakE3NWIxTUE&usp=sharing_eil#gid=0`