



# Anomalous Couplings Project

## Progress Report and documentation of Tools

Annik Olbrechts

Started 4 April 2014



# Contents

<b>1</b>	<b>Pending issues</b>	<b>1</b>
<b>2</b>	<b>Used tools and techniques</b>	<b>2</b>
2.1	MadGraph . . . . .	2
2.1.1	MadGraph_v155 version . . . . .	2
2.1.2	New MadGraph5_aMC@NLO version . . . . .	2
2.2	MadWeight . . . . .	3
2.2.1	MC_PERM MadWeight use . . . . .	4
2.2.2	Checking MadWeight on localgrid . . . . .	5
2.2.3	Influence of the used Transfer Function . . . . .	5
2.2.4	Adapting MadWeight to run continously on localgrid . . . . .	5
2.3	MadAnalysis . . . . .	5
2.3.1	How to run MadAnalysis in expert mode . . . . .	5
2.3.2	Content of analysis file in MadAnalysis . . . . .	6
2.3.3	Analyzing the MadGraph files . . . . .	7
2.3.4	Width of the decay . . . . .	7
2.4	FeynRules . . . . .	7
<b>3</b>	<b>Analyzing created FeynRules model</b>	<b>8</b>
3.1	Normalized coupling parameters . . . . .	8
3.2	Understanding parameters larger than 1 . . . . .	8
3.2.1	Performed checks . . . . .	9
<b>4</b>	<b>Event selection</b>	<b>14</b>
4.1	Choice of b-tag requirements . . . . .	14
4.1.1	Signal vs background comparison . . . . .	14
4.2	Use of $m_{lb}$ $\chi^2$ method for selecting the correct b-jets . . . . .	15
4.3	Histograms for event selection choice . . . . .	16
4.4	Considering 2 or 3 light jets . . . . .	17
4.4.1	Event selection numbers comparison . . . . .	19
4.4.2	Mlb-algorithm numbers comparison . . . . .	19
4.5	Studying optimal cut on $\chi^2$ value . . . . .	20
4.5.1	$\chi^2$ required to be smaller than 5 . . . . .	20
4.5.2	$\chi^2$ required to be smaller than 3 . . . . .	20
<b>5</b>	<b>Event corrections and reconstruction</b>	<b>22</b>
5.1	Which event corrections should be applied? . . . . .	22
5.1.1	Lumi- or PileUp Reweighting . . . . .	22
5.1.2	Lepton Scale Factors . . . . .	22

5.1.3	Jet Energy Correction factors . . . . .	22
5.1.4	Trigger choice . . . . .	22
5.1.5	Jet corrections (on the fly ...)	23
5.2	Choice of Monte Carlo samples . . . . .	23

# 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 ...?

# 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` (2.1)

`importmodelMassiveLeptons` (2.2)

`generatepp > tt , (t > bw+, w+ > mu + vm), (t > b w-, w- > jj)@1QED = 2` (2.3)

`outputdirName` (2.4)

`exit` (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 continuously 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 bzip command package has been installed on the m-machines, but not on mtop):

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

Whenever MadWeight is first used a personal directory should be initialized. 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 <sup>1</sup>:

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

---

<sup>1</sup>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 | grep aolbrech` (2.6)

`qdel 394402.cream02` (2.7)

### 2.2.3 Influence of the used Transfer Function

### 2.2.4 Adapting MadWeight to run continuously on localgrid

A single test with the following configuration (5000 events - 20 events/job - 10 000 init-Points) took almost 16h to finish so the full generator event sample ( $> 2\,470\,000$  events) would take more than 329 days!<sup>2</sup>

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 continuously 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 ...

---

<sup>2</sup>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:

$$./bin/ma5 \quad - -expert \quad (2.8)$$

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:

$$source \quad setup.sh \quad (2.9)$$

$$make \quad (2.10)$$

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:

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

### 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  $t t \rightarrow b jj b l \nu$  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-giving 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

Currently 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

# Chapter 3

## Analyzing created FeynRules model

### 3.1 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 7).

### 3.2 Understanding parameters larger than 1

Before starting to look at the  $t\bar{t}b\bar{b}$  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.

### 3.2.1 Performed checks

#### 3.2.1.1 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 3.1:** 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.

#### 3.2.1.2 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:

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

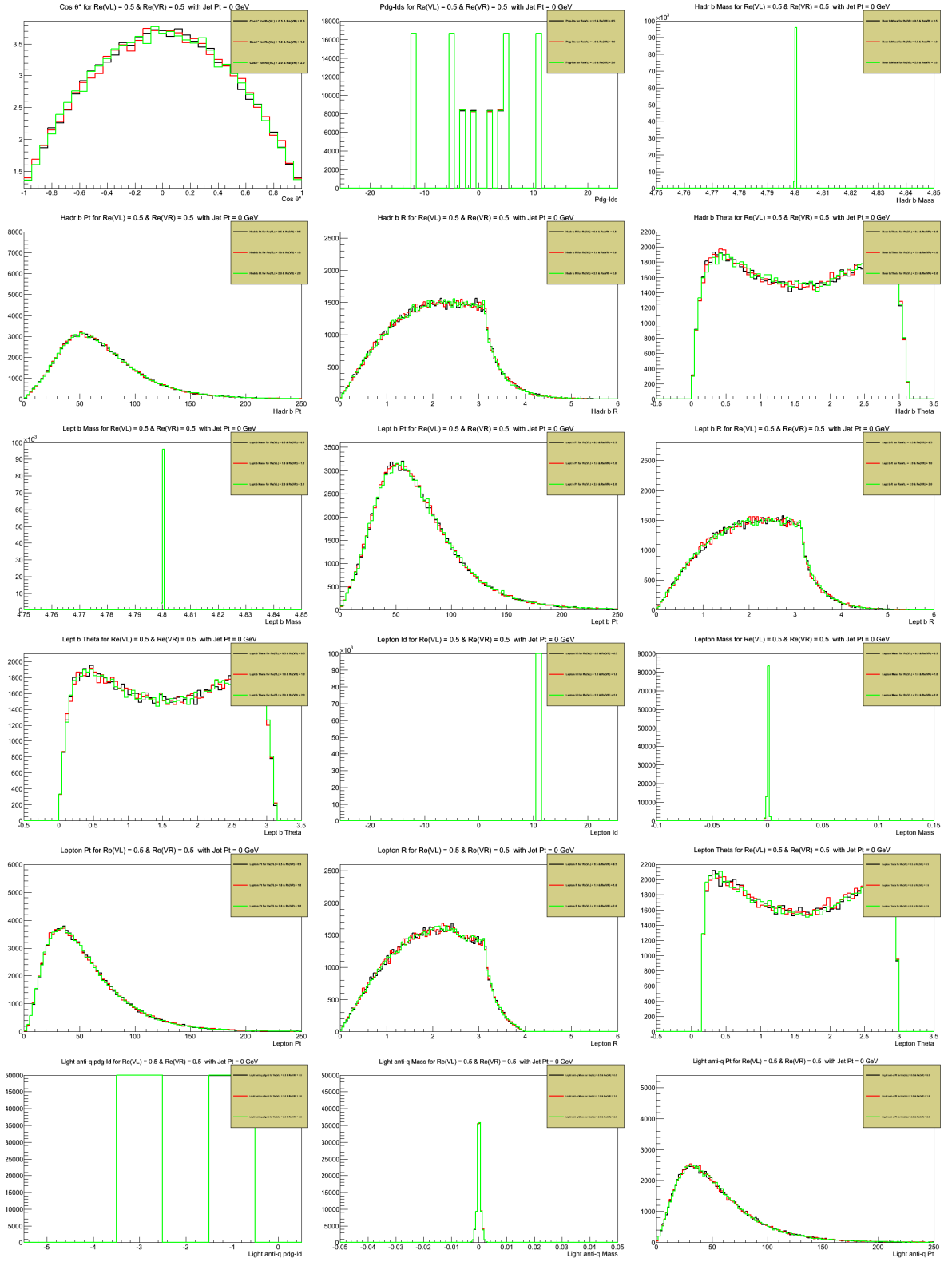
From the above numbers is clear that the Cross section becomes very large when the two coupling parameters increase. This can be understood quite easily since the second option allows much more decay options since the top quarks can decay both through the left-handed and the right-handed vector coupling 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.

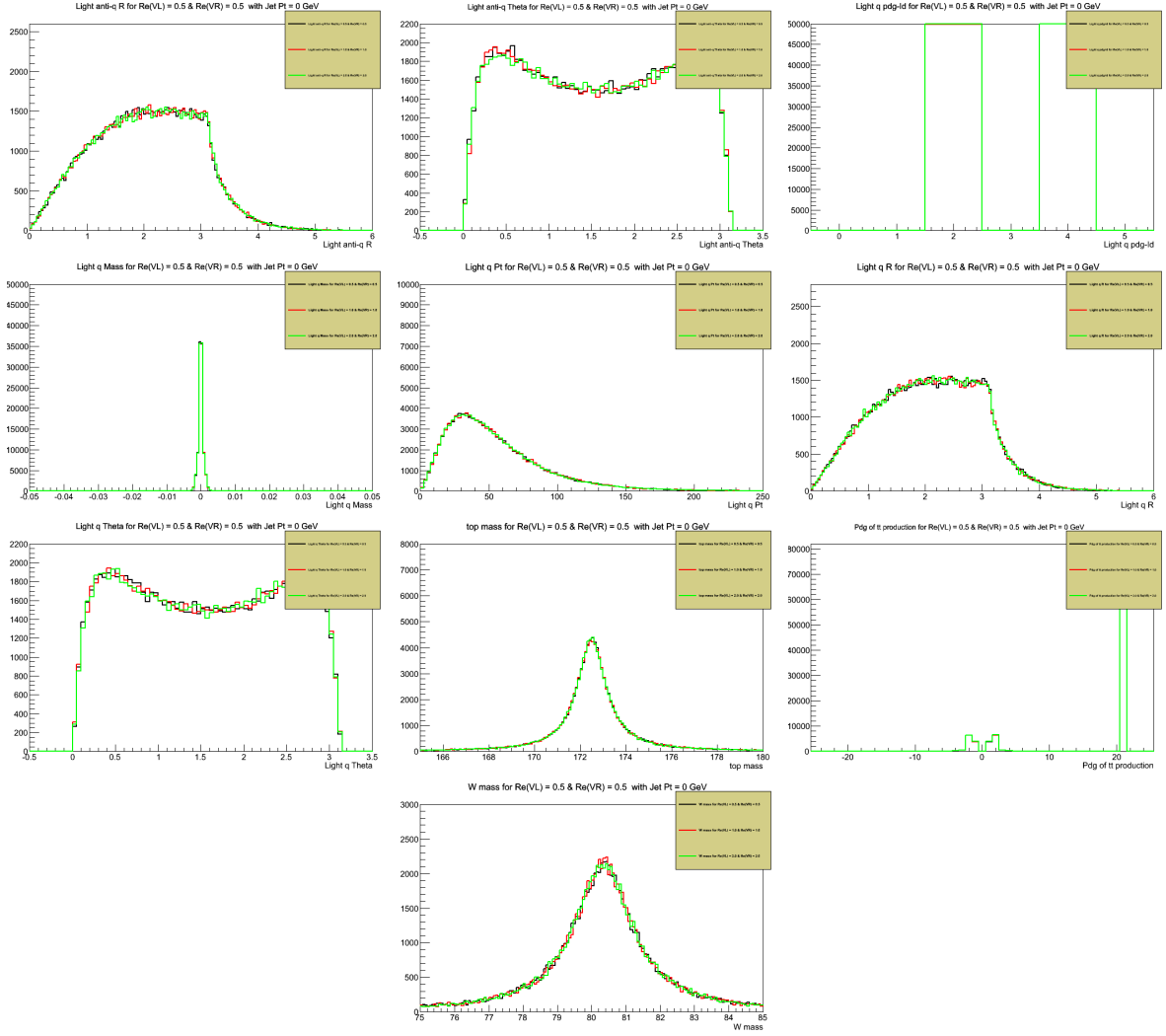
Looking at these plots clearly indicates that the kinematics doesn't change at all. Hence the created FeynRules model is able to deal in a correct way with these coupling parameters larger than 1.

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

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

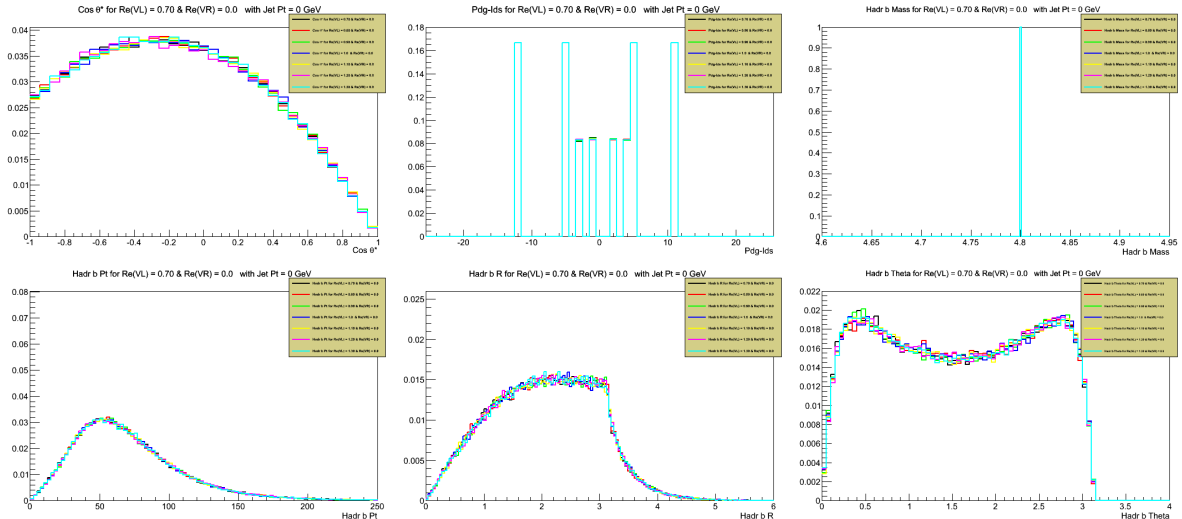
The distributions shown in this subsection are for fixed Jet Pt Cut value, set to 0. Also no Pt cut on the lepton was applied. Both coupling parameters have been changed proportionally.

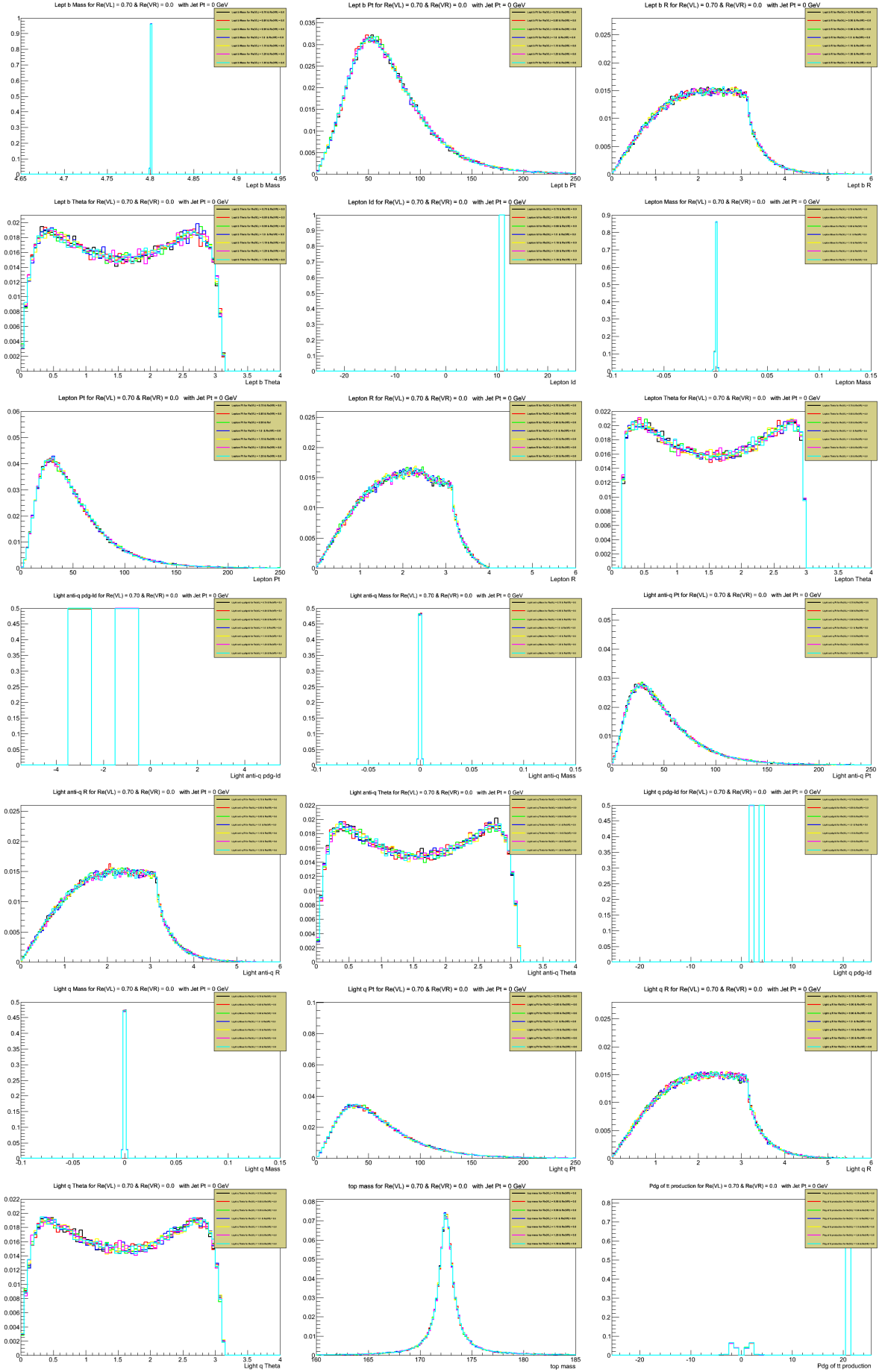


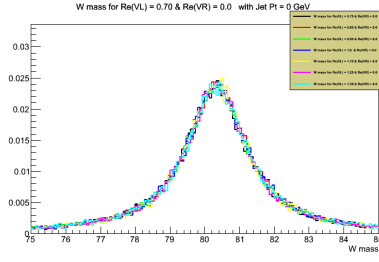


### 3.2.1.3 Model plots for fixed Pt Cut

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).







### 3.2.1.4 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 4

## Event selection

Not correctly considering tau's for the moment ...

### 4.1 Choice of b-tag requirements

#### 4.1.1 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 $\chi^2 m_{lb}$ )	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 4.1:** Overview of correct and wrong reconstructed events for the different b-tags without the use of a  $\chi^2 m_{lb} - m_{qqb}$  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 b-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

Option (no $\chi^2 m_{lb}$ )	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 4.2:** Overview of correct and wrong reconstructed b-jets for the different b-tags without the use of a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method

Option (no $\chi^2 m_{lb}$ )	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 4.3:** Overview of correct and wrong reconstructed light jets for the different b-tags without the use of a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method

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.

## 4.2 Use of $m_{lb} \chi^2$ method for selecting the correct b-jets

Option (with $\chi^2 m_{lb}$ )	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 4.4:** Overview of correct and wrong reconstructed events for the different b-tags when a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method is applied

Option (with $\chi^2 m_{lb}$ )	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 4.5:** Overview of the number of times the correct b-jet combination is chosen when using a  $\chi^2 m_{lb} - m_{qqb}$  method

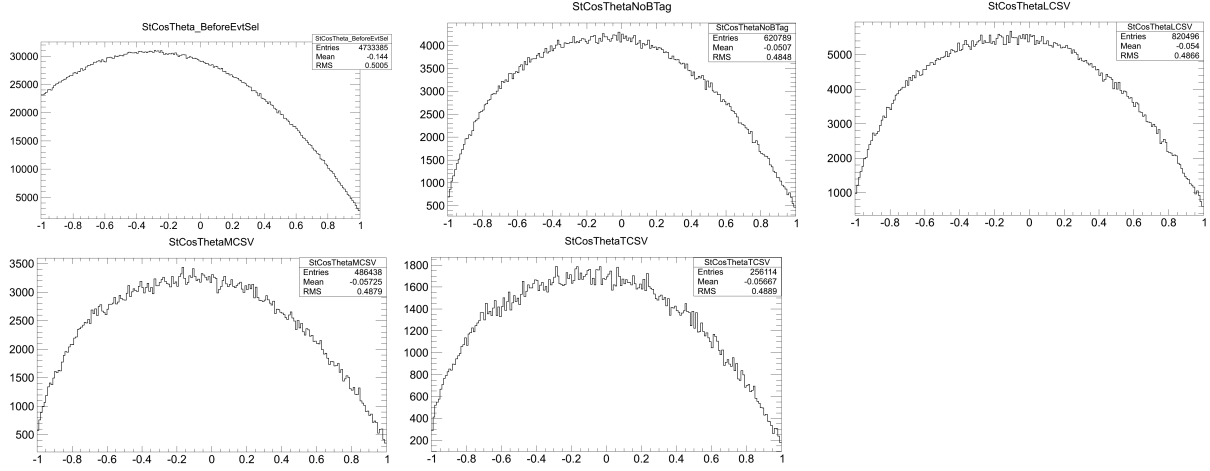
The two tables in this subsection require a different interpretation. The first one is actually a combined test of the  $\chi^2 m_{lb} - m_{qqb}$  method and the optimal b-tag choice while the second one is merely a performance check of the mlb method.

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  $\chi^2 m_{lb} - m_{qqb}$  method is applied.

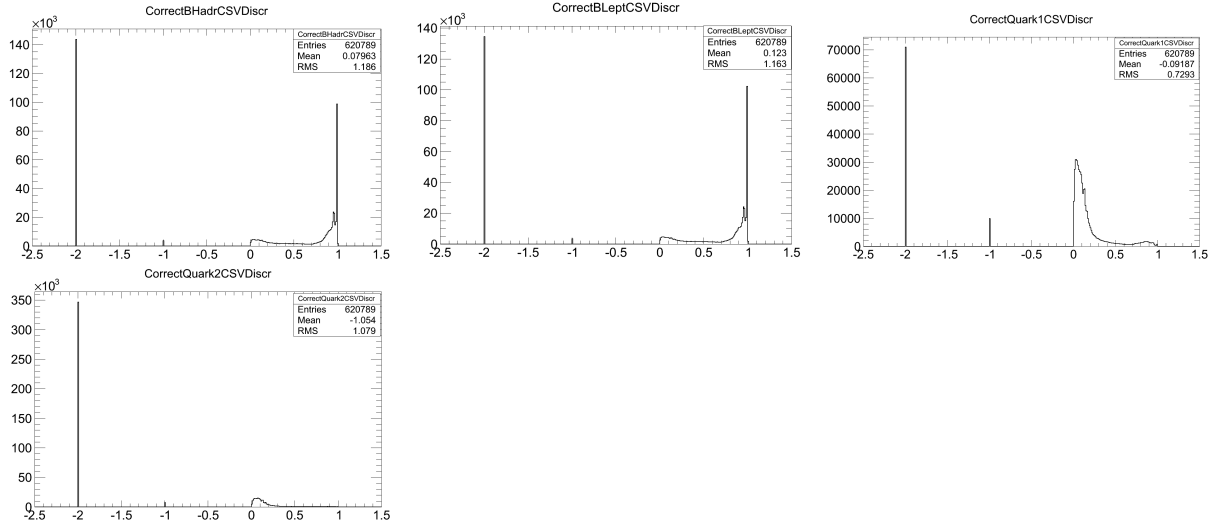
Therefore the second table is relevant in order to select whether some clear gain can be obtained when applying this method since it represents the number of times the correct b-jet combination. If this percentage is higher than 50%, which is the case, an improvement is obtained compared to an iteration 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  $\chi^2 m_{lb} - m_{qqb}$  method significantly enhances the correct choice of the b-jet combination. In about 84% of the cases the correct b-jet combination is chosen.

### 4.3 Histograms for event selection choice



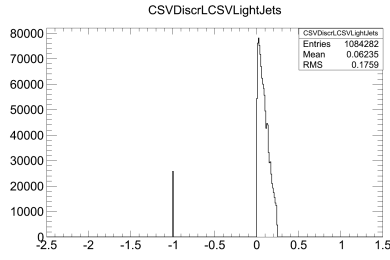
**Figure 4.1:** The  $\cos \theta^*$  distribution for the different b-tag options (all of them imply double b-tag), which are not really influenced by the application of a b-tag. The only relevant distortion is caused by the event selection which is applied.



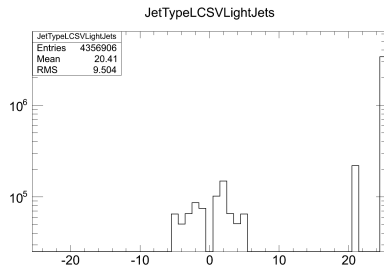
**Figure 4.2:** Distribution of the CSV discriminant for the different correctly matched quark-jet pairs. The value  $-2$  is used to represent a non-matched jet. As expected the b-jets have a large peak at 1, so the Tight b-tag of 0.898 will not take away too many correct b-jets. The problem in the matching is clearly represented in distribution of the second quark which is only reconstructed in less than half of the cases.

## 4.4 Considering 2 or 3 light jets

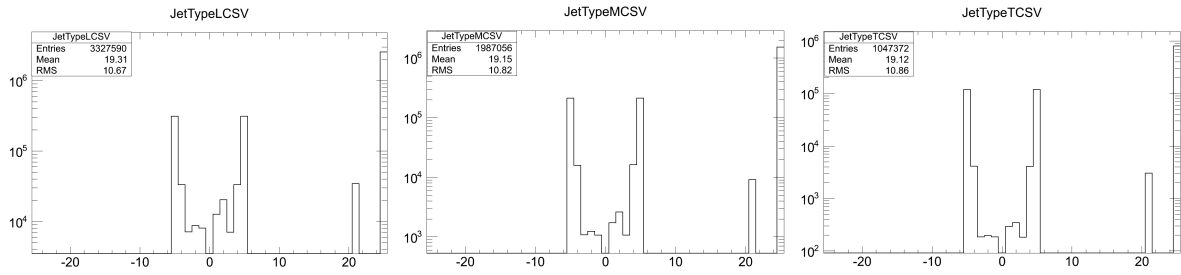
In order to try to improve the signal efficiency it can be considered to add a third light jet to the particles which have to be taken into account for the jet selection. Adding this third jet will however result in 4 additional combinations which have to be considered so this method will only benefit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



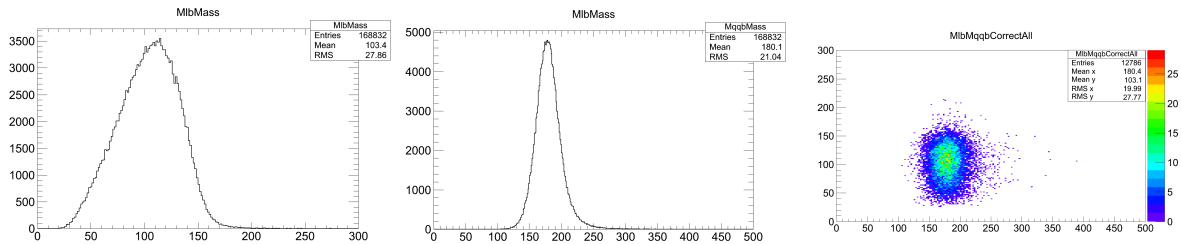
**Figure 4.3:** Distribution of the CSV discriminant of the selected light jets (all of them). Add same histograms for Medium and Tight option, this will show how many of the light jets actually have a large CSV discriminant (maybe only focus on the two/three leading light jets)



**Figure 4.4:** 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 with the JetPartonMatching method. Same for M and T ... ?



**Figure 4.5:** Jet type of the b-tagged jets (all of them) with the same convention for the non-matched jets.



**Figure 4.6:** 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.

Option (no $\chi^2 m_{lb}$ )	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 4.6:** Overview of correct and wrong reconstructed events for the different b-tags without the use of a  $\chi^2 m_{lb} - m_{qqb}$  method

Option (no $\chi^2 m_{lb}$ )	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 4.7:** Overview of correct and wrong reconstructed b-jets for the different b-tags without the use of a  $\chi^2 m_{lb} - m_{qqb}$  method

Option (no $\chi^2 m_{lb}$ )	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 4.8:** Overview of correct and wrong reconstructed light jets for the different b-tags without the use of a  $\chi^2 m_{lb} - m_{qqb}$  method

#### 4.4.1 Event selection numbers comparison

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 each other 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.

#### 4.4.2 Mlb-algorithm numbers comparison

In the  $m_{lb}$  method tables the 4 jets which are actually chosen by the  $\chi^2 m_{lb} - m_{qqb}$  method. 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

Option (with $\chi^2 m_{lb}$ )	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 4.9:** Overview of correct and wrong reconstructed events for the different b-tags when a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method is applied

Option (with $\chi^2 m_{lb}$ )	% 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 4.10:** Overview of the number of times the correct b-jet combination is chosen when using a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method

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.

## 4.5 Studying optimal cut on $\chi^2$ value

The tables shown here indicate that the application of a cut on the  $\chi^2$  value of the  $m_{lb} - m_{q\bar{q}b}$  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.

### 4.5.1 $\chi^2$ required to be smaller than 5

Option (with $\chi^2 m_{lb}$ )	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 4.11:** Overview of correct and wrong reconstructed events for the different b-tags when a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method is applied

### 4.5.2 $\chi^2$ required to be smaller than 3

<b>Option</b> (with $\chi^2 m_{lb}$ )	% 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 4.12:** Overview of the number of times the correct b-jet combination is chosen when using a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method

<b>Option</b> (with $\chi^2 m_{lb}$ )	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 4.13:** Overview of correct and wrong reconstructed events for the different b-tags when a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method is applied

<b>Option</b> (with $\chi^2 m_{lb}$ )	% 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 4.14:** Overview of the number of times the correct b-jet combination is chosen when using a  $\chi^2 m_{lb} - m_{q\bar{q}b}$  method



# Chapter 5

## Event corrections and reconstruction

### 5.1 Which event corrections should be applied?

#### 5.1.1 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](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\\_E1.cc](https://github.com/TopBrussels/FourTops/blob/master/FourTop_EventSelection_E1.cc) which contains the same information but about the electron case.

#### 5.1.2 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.

### 5.1.3 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 ...**

### 5.1.4 Jet Energy Correction factors

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

### 5.1.5 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 ...

## 5.2 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](https://docs.google.com/spreadsheet/ccc?key=0Apc0aJdnaVjSdFVaLVU2dlk4RDZHcjlaakE3NWIXTUE&usp=sharing_eil#gid=0)