

A SEARCH FOR EXOTIC PARTICLES OF CHARGE  
5/3 WITH THE CMS DETECTOR

MATTEO ABIS

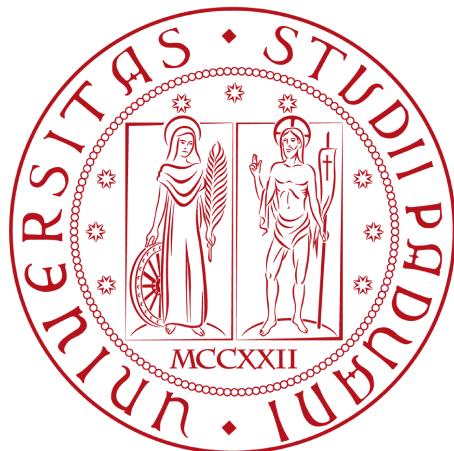
Dipartimento di Fisica e Astronomia  
Università degli Studi di Padova

Relatore: Patrizia Azzì      Relatore: Azzì  
A.A. 2011/2012



A SEARCH FOR EXOTIC PARTICLES OF CHARGE  $5/3$  WITH THE  
CMS DETECTOR

MATTEO ABIS



Dipartimento di Fisica e Astronomia  
Università degli Studi di Padova

Relatore: Patrizia Azzi  
A.A. 2011/2012



## ABSTRACT

---

The existence of heavy partners of the top quark with charge  $5/3$  is a common prediction of many solution to the Standard Model hierarchy problem. The search for the pair-production of top partners in pp collisions at  $\sqrt{s} = 7\text{ TeV}$  with the CMS detector at the LHC is reported. The data sample includes all of the statistics of 2011, corresponding to an integrated luminosity of  $5\text{ fb}^{-1}$ . New techniques and kinematical variables are developed in order to enhance the rejection of the backgrounds. The number of selected events is consistent with the expectations from Standard Model backgrounds and no excess is found. Exclusion limits are set at 95% confidence level.



## CONTENTS

---

<b>1 THE THEORY OF THE TOP PARTNERS</b>	<b>1</b>
1.1 The Standard Model	1
1.2 The unbroken SM	1
1.3 Spontaneous symmetry breaking: the Higgs mechanism	2
1.4 The hierarchy problem	3
1.5 The top partners	3
<b>2 THE LHC AND THE CMS DETECTOR</b>	<b>5</b>
2.1 The Large Hadron Collider	5
2.2 The Compact Muon Solenoid	5
2.2.1 Coordinate convention	6
2.2.2 Structure of the detector	6
2.2.3 The tracking system	7
2.2.4 The electromagnetic calorimeter	8
2.2.5 The hadron calorimeter	9
2.2.6 The muon detectors	9
2.2.7 The trigger and data acquisition system	9
<b>3 SIMULATION, RECONSTRUCTION AND DATASETS</b>	<b>11</b>
3.1 Simulation	11
3.1.1 The Full Simulation	11
3.1.2 The Fast Simulation	11
3.2 The reconstruction of physics objects	12
3.2.1 Track reconstruction	13
3.2.2 Vertex reconstruction	13
3.2.3 Muon reconstruction	13
3.2.4 Electron reconstruction	15
3.2.5 Jet reconstruction	17
3.3 Object selection	19
3.3.1 Muon selection	19
3.3.2 Electron selection	20
3.3.3 Jet selection	21
3.3.4 Event cleanup and vertex selection	21
3.3.5 Trigger requirements	22
3.4 Datasets	23
3.5 Signal Monte Carlo samples	23
3.6 Background Monte Carlo samples	24
3.7 Monte Carlo pile-up reweighting	25
<b>4 EVENT PRESELECTION</b>	<b>27</b>
<b>5 A DATA-DRIVEN BACKGROUND ESTIMATE</b>	<b>31</b>
5.1 Charge misidentification	31
5.2 Same-sign non-prompt background	32
5.2.1 Measuring the tight-to-loose ratio	33
5.2.2 Heavy flavour content of the samples	35

5.2.3	Prompt ratio approximation	36
5.2.4	Calculation of the expected event yields	36
5.2.5	Validation on test samples	38
6	THE RAZOR VARIABLES	43
6.1	The variable $M_R$	43
6.2	The transverse plane and the R variable	44
6.3	The razor variables for the top partners	44
6.4	Event reconstruction improvement and b-tagging	46
6.5	Final event selection	47
7	SYSTEMATIC UNCERTAINTIES	55
7.1	Monte Carlo systematics	55
7.2	Data-driven background systematics	55
8	RESULTS	59
8.1	Calculation of the observed limit	59
8.2	Calculation of the expected limit	62
8.3	Results	63
A	APPENDIX	67
	BIBLIOGRAPHY	79

## LIST OF FIGURES

---

Figure 1	The particles in the unbroken SM.	2
Figure 2	Radiative correction to the Higgs mass from the top quark.	3
Figure 3	Typical single and pair production of the top partner $T_{5/3}$ at the LHC.	4
Figure 4	A view of the LHC and the four experiments.	6
Figure 5	The CMS detector.	6
Figure 6	Transverse view of the CMS detector with some examples of tracks.	7
Figure 7	Energy resolution of the ECAL as measured in $Z \rightarrow ee$ events in the 2011 data.	8
Figure 8	The layout of the muon detectors.	9
Figure 9	Muon momentum resolution as a function of $p_T$ in regions with different pseudorapidity. The blue squares are for the muon system only, the green circles for the inner tracker only, and the red circles for the combined reconstruction	14
Figure 10	Illustration of the isolation observable.	15
Figure 11	An illustration of the <i>bremstrahlung</i> recovery algorithm.	16
Figure 12	The track distance for conversion rejection is the two dimensional distance in the xy-plane between B1 and B2. At these points, the two tracks from the photon conversion are parallel. The distance is defined to be negative if the two tracks overlap, and is positive other- wise.	17
Figure 13	Selection efficiency for the various working points, measured in $Z \rightarrow ee$ and $W \rightarrow ee$ events. In this analysis the HyperTightMC1 selection (light blue squares) is used.	21
Figure 14	Selection fake rate for the various working points, measured in QCD dijet events. In this analysis the HyperTightMC1 selection (light blue squares) is used.	22
Figure 15	Peak instantaneous luminosity delivered by the LHC in 2011.	25
Figure 16	Distribution of the number of vertices in un- weighted Monte Carlo and data. The Monte Carlo samples are weighted so that the two dis- tributions match.	26

- Figure 17 Pair-production of the  $T_{5/3}$ , decaying into same-sign leptons and at least four jets. 27
- Figure 18 Events with two same-sign leptons and two jets. Distribution of the number of jets in the backgrounds and signal for the 550 GeV mass point for the sum of the three decay channels, which are shown separately in figure 18. The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties. 28
- Figure 19 Events with two same-sign leptons and two jets. Distribution of the invariant mass of the leptons in the backgrounds and signal for the 550 GeV mass point for the sum of the three decay channels, which are shown separately in figure 46. The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties. 29
- Figure 20 The dilepton invariant mass spectrum for electrons in data. The opposite-sign distributions are shown in blue and the same-sign ones are in red. Both distributions are normalized to unit area. 32
- Figure 21 The corrected tight-to-loose ratio, loosely speaking the *fake rate*, as a function of  $p_T$  for electrons (left) and muons (right). The correction is accomplished by subtracting the MC-based  $W, Z$  and  $t\bar{t}$  contributions from the data. 39
- Figure 22 Expected and observed distributions of the  $p_T$  of the leading (top) and subleading (bottom) jet in events with two same-sign leptons and two jets. The sum of the three decay channels is shown. 40
- Figure 23 Expected and observed distributions of the  $p_T$  of the leading (top) and subleading (bottom) leptons in events with two same-sign leptons and two jets. The sum of the three decay channels is shown. 41

Figure 24	Expected and observed distributions of the invariant mass of jets (top) and leptons (bottom) in events with two same-sign leptons and two jets. The sum of the three decay channels is shown. The <i>optional b-tag</i> method for calculating the invariant mass of the jets is described in 6.4. 42
Figure 25	Simple example of a supersymmetric event, with two heavy pair-produced SUSY particles, two invisible lightest supersymmetric particles and two visible SM particles. 43
Figure 26	Comparison of a simple supersymmetric example with the top partner decay. 45
Figure 27	Invariant mass of the hadronic system for the sum of the three decay channels. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ , $\mu\mu$ channels are shown in figure 47 46
Figure 28	$M_R$ distributions for the sum of the three decay channels. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ , $\mu\mu$ channels are shown in figure 48 47
Figure 29	R distributions for the sum of the three decay channels. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ , $\mu\mu$ channels are shown in figure 49 48
Figure 30	$\Delta R$ between the b-jet and the lepton from the top decay. MC truth from the 550 GeV mass point. 49
Figure 31	Distribution of the $p_T$ of the b quark in the $T_{5/3}$ decay chain. If the quark has a low $p_T$ , it will most likely form a low- $p_T$ jet, which would not be collected by our selections. MC truth from the 550 GeV mass point. 49
Figure 32	Invariant mass of the hadronic system for the sum of the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ , $\mu\mu$ channels are shown in figure 50 50

- Figure 33  $M_R$  distributions for the sum of the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ ,  $\mu\mu$  channels are shown in figure 51 – 50 50
- Figure 34  $R$  distributions for the sum of the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ ,  $\mu\mu$  channels are shown in figure 52 – 51 51
- Figure 35 Optimization figure F for the invariant mass of the hadronic system. The selection on  $M_R > 200 \text{ GeV}$  and  $R > 0.2$  are applied. The point where F reaches a maximum is chosen as a cut on this variable. 51
- Figure 36 Optimization figure F for the  $M_R$  variable. The selection on  $M > 350 \text{ GeV}$  and  $R > 0.2$  are applied. The point where F reaches a maximum is chosen as a cut on this variable. 52
- Figure 37 Optimization figure F for the R variable. The selection on  $M > 350 \text{ GeV}$  and  $M_R > 200 \text{ GeV}$  are applied. The point where F reaches a maximum is chosen as a cut on this variable. 52
- Figure 38 Distribution of the R variable for events in the three decay channels, with the final selection applied without the selection on R. 53
- Figure 39 Distribution of the  $M_R$  variable for events in the three decay channels, with the final selection applied without the selection on  $M_R$ . 53
- Figure 40 Distribution of the invariant mass of the jets for events in the three decay channels, with the final selection applied without the selection on  $M(\text{jets})$ . 54
- Figure 41 Charge misidentification probability as a function of the pseudorapidity  $\eta$ . 56
- Figure 42 Distribution of the  $M_R$  variable for events in the three decay channels, with the final selection applied. 59
- Figure 43 Distribution of the number of jets for events in the three decay channels, with the final selection applied. 61
- Figure 44 Observed and expected limits for the production cross sections of the top partners. 63

- Figure 45 Events with two same-sign leptons and two jets. Distribution of the number of jets in the backgrounds and signal for the 550 GeV mass point for the three decay channels ee, e $\mu$  and  $\mu\mu$ . The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties. 71
- Figure 46 Events with two same-sign leptons and two jets. Distribution of the invariant mass of the leptons in the backgrounds and signal for the 550 GeV mass point for the three decay channels ee, e $\mu$  and  $\mu\mu$ . The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties. 72
- Figure 47 Invariant mass of the hadronic system for the three decay channels. Events with two same-sign leptons and at least two jets are shown. 73
- Figure 48  $M_R$  distributions for the three decay channels. Events with two same-sign leptons and at least two jets are shown. 74
- Figure 49  $R$  distributions for the three decay channels. Events with two same-sign leptons and at least two jets are shown. 75
- Figure 50 Invariant mass of the hadronic system for the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. 76
- Figure 51  $M_R$  distributions for the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. 77

Figure 52	R distributions for the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown.	78
-----------	---	----

## LIST OF TABLES

---

Table 1	List of triggers used in the analysis for the three decay channels: $\mu\mu$ , ee, e $\mu$ . CCTT is short for CaloIdL_CaloIsoVL_TrkIdVL_TrkIsoVL.	22
Table 2	Data samples used for the analysis.	23
Table 3	Signal Monte Carlo samples. The branching ratio is 0.21.	24
Table 4	Details of the background Monte Carlo samples used for the analysis.	24
Table 5	Expected event yields within the $Z$ invariant mass window ( $76 \text{ GeV} < M(\ell\ell) < 106 \text{ GeV}$ ) for the background contributions. The backgrounds not shown yield negligible contributions to this phase space region.	31
Table 6	Expected event yields within the $Z$ invariant mass window ( $76 \text{ GeV} < M(\ell\ell) < 106 \text{ GeV}$ ) for the eight $T_{5/3}$ signal points.	31
Table 7	Tight to loose ratio for electrons in the data and in the MC. The line Data-MC subtracts from the data the prompt lepton contribution due to $W + \text{jets}$ , $Z + \text{jets}$ and $t\bar{t}$ from the data.	34
Table 8	Tight to loose ratio for muons in the data and in the MC. The line Data-MC subtracts from the data the prompt lepton contribution due to $W + \text{jets}$ , $Z + \text{jets}$ and $t\bar{t}$ from the data.	34
Table 9	Tight to loose ratio for electrons in the data and in the MC with the $p_T$ range of the electron restricted to the 25-35 GeV range. The line Data-MC subtracts the prompt lepton contribution due to $W + \text{jets}$ , $Z + \text{jets}$ and $t\bar{t}$ from the data.	35
Table 10	Tight to loose ratio for muons in the data and in the MC with the $p_T$ range of the muon restricted to the 25-35 GeV range. The line Data-MC subtracts the prompt lepton contribution due to $W + \text{jets}$ , $Z + \text{jets}$ and $t\bar{t}$ from the data.	35

Table 11	Summary of the closure tests on Monte Carlo samples for events with two same-sign leptons. Statistical errors only.	38
Table 12	Systematic uncertainties.	55
Table 13	Systematic uncertainties for backgrounds that are taken from Monte Carlo.	56
Table 14	Systematic uncertainties for data-driven backgrounds	57
Table 15	Signal Monte Carlo yields for the various mass points in the three decay channels. Statistical and systematic uncertainties are shown.	60
Table 16	Expected backgrounds and observed yields for different cuts in the three decay channels. Statistical and systematic uncertainties are shown.	60
Table 17	Signal Monte Carlo yields for the various mass points in the ee channel. Statistical and systematic uncertainties are shown.	68
Table 18	Signal Monte Carlo yields for the various mass points in the e $\mu$ channel. Statistical and systematic uncertainties are shown.	68
Table 19	Signal Monte Carlo yields for the various mass points in the $\mu\mu$ channel. Statistical and systematic uncertainties are shown.	69
Table 20	Expected backgrounds and observed yields for different cuts in the ee channel. Statistical and systematic uncertainties are shown.	69
Table 21	Expected backgrounds and observed yields for different cuts in the e $\mu$ channel. Statistical and systematic uncertainties are shown.	69
Table 22	Expected backgrounds and observed yields for different cuts in the $\mu\mu$ channel. Statistical and systematic uncertainties are shown.	70



## THE THEORY OF THE TOP PARTNERS

---

### 1.1 THE STANDARD MODEL

The SM is a quantum field theory developed in the 1970s. It describes matter and the electromagnetic, weak and strong nuclear forces in terms of point-like particles [29, 40].

The most profound insight in the SM is that all of these interactions are determined by symmetry principles, called local gauge symmetries. This idea is connected with the fact that the conserved physical quantities, e.g. electric charge, are conserved at every point in space-time, and not just globally. This connection is given by Nöther's theorem in the lagrangian formalism.

The treatment of one particular property of the particles requires some ingenuity. Mass terms are not allowed in the SM lagrangian, because they break the fundamental symmetry principles. Yet it is obvious from observation that most particles have mass. The Higgs mechanism with spontaneous symmetry breaking was devised to solve this problem, and appears to be close to experimental confirmation at the time of writing.

We will first review the features of the unbroken SM, then introduce the Higgs field and finally present some shortcomings of this theoretical description.

A complete description can be found elsewhere, the following section will highlight the material that is relevant to the problems we are investigating in the rest of our work.

### 1.2 THE UNBROKEN SM

All of the matter particles in the SM are fermions, and are usually divided in two categories: particles that feel the strong force, called *quarks*, and particles that do not, called *leptons*. Quarks and leptons are grouped into three *families* or *generations*. Particles from different generations have exactly the same properties, except for mass. The origin of this family structure is still unknown.

A force is introduced by requiring the lagrangian for the matter fields to be invariant under a group of space-time dependent transformations, called *gauge* transformations.

The theory of QED was the first successful gauge theory. It describes the electromagnetic force, with a  $U(1)$  gauge invariance. As a consequence, a *gauge boson* has to be introduced to preserve the local symmetry, that is the photon.

The electromagnetic and weak forces are then unified by a gauge group  $SU(2)_L \times U(1)_Y$ , with the L subscript denoting the fact that the  $SU(2)$  gauge transformations only involve the left-handed fermions. The Y subscript is intended to indicate that this  $U(1)_Y$  group is not the same as the aforementioned group for QED. The number of gauge bosons must be the same as the dimension of the gauge group, so that we now get four vector bosons: the photon, the  $W^\pm$  and the Z.

The description of the strong force involves the group  $SU(3)$ . The charge of the strong force is called *colour*, it is carried by eight bosons called *gluons*. Figure 1 summarizes the particle contents of the unbroken SM.

Three Generations of Matter (Fermions)			
	I	II	III
mass→	2.4 MeV	1.27 GeV	171.2 GeV
charge→	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$
spin→	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
name→	u up	c charm	t top
	d down	s strange	b bottom
	v <sub>e</sub> electron neutrino	v <sub>μ</sub> muon neutrino	v <sub>τ</sub> tau neutrino
	e electron	μ muon	τ tau
Quarks			
Leptons			
Bosons (Forces)			
	γ photon	g gluon	Z <sup>0</sup> weak force
	W <sup>±</sup> weak force		

Figure 1: The particles in the unbroken SM.

### 1.3 SPONTANEOUS SIMMETRY BREAKING: THE HIGGS MECHANISM

We have only considered massless particles. This is not the case for all the known fermions, and for the W and Z bosons. A mechanism was proposed to preserve the gauge invariance of the lagrangian, while the vacuum state is no longer a singlet under the action of the gauge group.

Spontaneous symmetry breaking of the local gauge symmetry in the SM provides massive W and Z bosons and a new boson, called the Higgs boson. In addition, we can obtain massive fermions by coupling them to the Higgs field.

#### 1.4 THE HIERARCHY PROBLEM

Recently found evidence for a light Higgs boson, with a mass near 125 GeV [3, 14], are thought to be an indication of new symmetries and new particles beyond the SM.

The theory predicts that the mass  $m_H$  is subject to large radiative corrections from loop diagrams similar to figure 2, that should increase it by a large amount. In the SM, fine tuning is required to prevent this corrections from becoming too large [27]. Theoretical physicists generally dislike this fine tuning on grounds of *naturalness*.

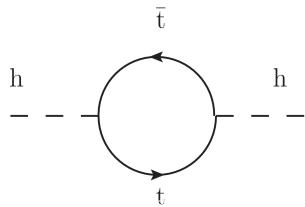


Figure 2: Radiative correction to the Higgs mass from the top quark.

The most notorious example of an extension to the SM that can eliminate this naturalness problem is supersymmetry, where the radiative corrections to the Higgs mass of the SM fermions are canceled by the contributions from their superpartners.

There exist many theories that do not invoke supersymmetry as the solution to the hierarchy problem. A robust and generic prediction of many of them is the existence of new particles, the top partners, again balancing the largest contribution to  $m_H$  coming from the top quark.

#### 1.5 THE TOP PARTNERS

Theoretical developments predicting the existence of top partners stem from the search for a way of introducing gravity in the SM, while solving the hierarchy problem at the same time.

Five-dimensional space-time theories including gravity can be formulated in terms of effective 4-d lagrangians where SM quarks mix with the top partners, which, again from arguments of naturalness, should have a mass below or slightly above the TeV scale. The lightest quarks have a small mass, indicating that their mixing with the new particles is negligible. The top quark, having a very large mass, should have a sizeable mixing with its partner, possibly making deviation from the SM top interactions detectable at the LHC.

We use the top partner model from [32] to describe our signal, with vertices with the vector bosons and the top quark. We study the possibility of observing the top partners in the very clean channel of two same-sign hard leptons. Typical production and decay diagrams

are shown in figure 3. The cross-section of the single production by weak interaction is smaller by a factor of at least two in all the mass range considered in the present work, and is therefore not taken into account.

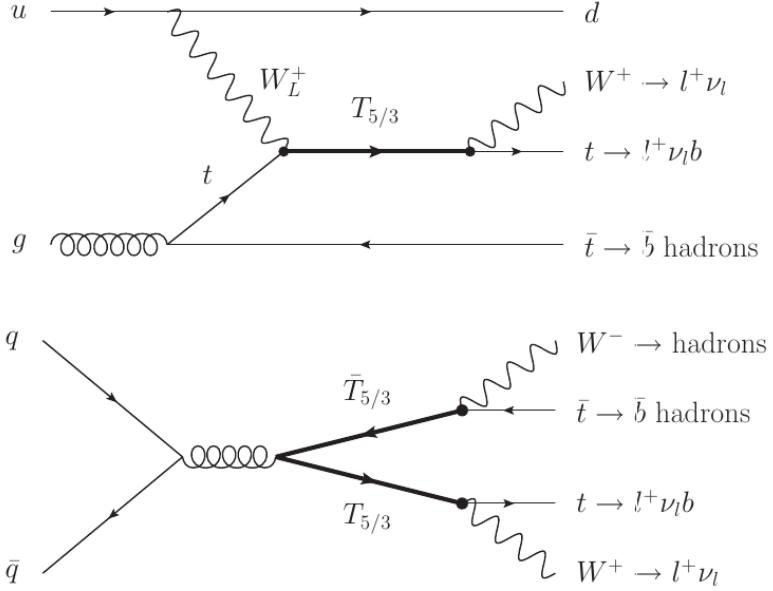


Figure 3: Typical single and pair production of the top partner  $T_{5/3}$  at the LHC.

# 2

## THE LHC AND THE CMS DETECTOR

---

### 2.1 THE LARGE HADRON COLLIDER

The LHC is a ring-shaped proton-proton accelerator with a circumference of 27 km. It is hosted in a tunnel 45 to 170 m below ground at the CERN laboratory in Geneva, Switzerland. In 2011, the LHC accelerated two proton beams up to an energy of 3.5 TeV each and an instantaneous luminosity of up to  $3.55 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ .

The two proton beams are pre-accelerated through four smaller systems before being injected in the LHC:

- the linear accelerator (up to 50 MeV)
- the proton synchrotron booster (up to 1.4 GeV)
- the proton synchrotron (up to 26 GeV)
- the super proton synchrotron (up to 450 GeV)

The protons travel in ultra-high vacuum around  $10^{-10}$  mbar, while different sets of magnets keep them in a circular orbit and provide acceleration and focusing.

Radio frequency cavities in the LHC accelerate the protons, providing an energy increase of 0.5 MeV/turn. Superconducting dipole magnets operate at currents of 11850 A to bend the trajectories of the protons with a magnetic field up to 8.3 T. Quadrupole magnets are employed to focus the beams, thus increasing the probability of an interaction when they collide. The magnets are cooled with superfluid helium at 1.9 K. The beams are made of “bunches” of protons with a time spacing of 50 ns.

At four different interaction points, the two proton beams are brought to collision where the experiments are located: CMS, ATLAS, LHCb and ALICE, as shown in figure 4. The first two are general purpose detectors. LHCb is designed for the study of CP violation in b physics, and ALICE for the analysis of the quark-gluon plasma produced in heavy ion collisions.

### 2.2 THE COMPACT MUON SOLENOID

CMS is a general purpose detector based on interlocking cylindrical subdetectors in the central *barrel* region, closed by two *endcaps*, all placed coaxially with the beam and centred on the beam interaction point.



Figure 4: A view of the LHC and the four experiments.

### 2.2.1 Coordinate convention

The conventional coordinate system has its origin at the nominal interaction point. The  $z$  axis points along the beam direction, the  $y$  axis points vertically upwards, and the  $x$  axis points towards the centre of the LHC circumference. The  $xy$  plane is thus called the *transverse* plane as it is orthogonal to the CMS cylinder.

We define two angles in the transverse plane: the azimuthal  $\varphi$  angle is measured from the  $x$  axis, and the polar angle  $\theta$ , measured from the  $z$  axis. *Pseudorapidity* is defined as  $\eta = -\log \tan \theta/2$ .

### 2.2.2 Structure of the detector

A global view and a transverse section of the detector are shown in figures 5 and 6.

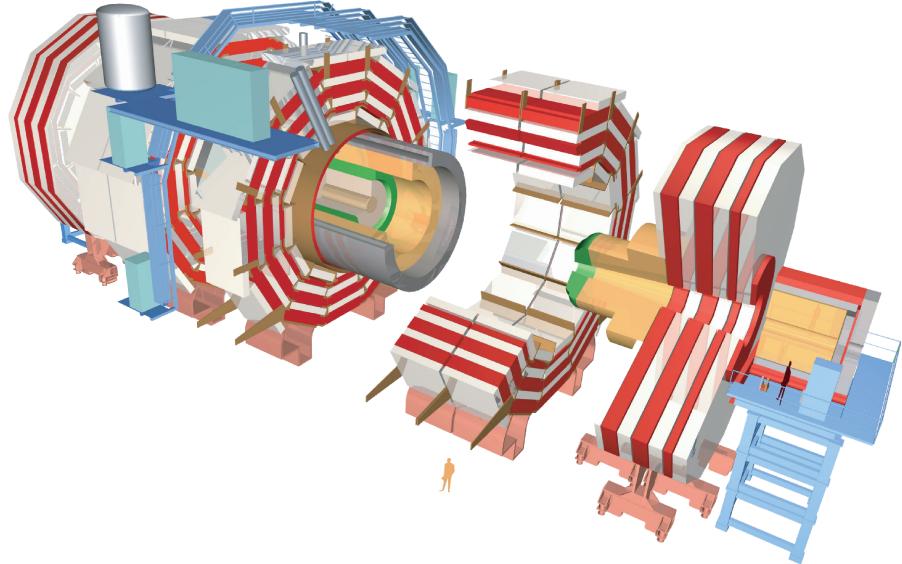


Figure 5: The CMS detector.

A superconducting solenoid 13.5 m and with a diameter of 5.9 m provides a uniform, axial magnetic field of 3.8 T. The return field sat-

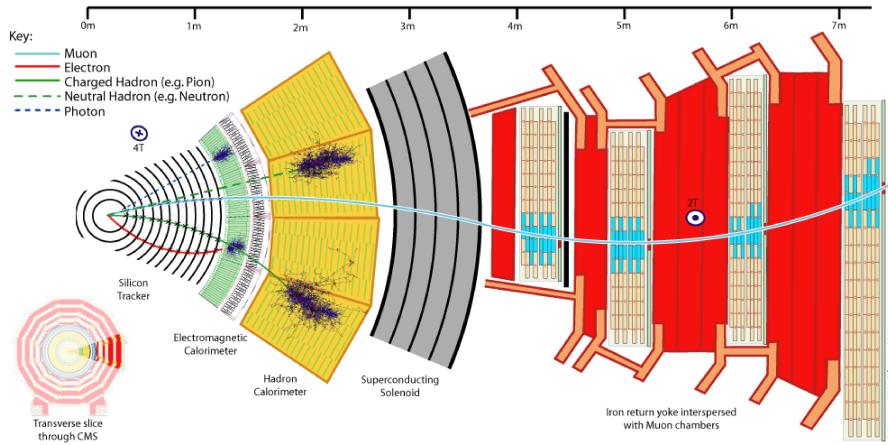


Figure 6: Transverse view of the CMS detector with some examples of tracks.

urates an iron yoke, which consists of five wheels in the barrel region and three disks in each of the two endcaps. Muon detectors are installed in the return yokes, with four stations of aluminium drift tubes and resistive plate chambers in the barrel and cathode strip chambers also with resistive plate chambers in the endcap.

The tracking system and the calorimeters are installed inside the solenoid. The tracker is a cylinder with a diameter of 2.6 m with three layers of pixel detectors, allowing exceptional measurements of the impact parameter of the particles and secondary vertex reconstruction, and ten layers of silicon microstrip detectors. The tracking system is surrounded by the calorimeters, which cover the region with  $|\eta| < 3$ . There is an electromagnetic calorimeter (ECAL), made of lead tungstate crystals and a brass/scintillator hadron calorimeter (HCAL). A forward calorimeter provides additional coverage up to  $|\eta| < 5$ .

### 2.2.3 The tracking system

Very high granularity and time resolutions are needed in the tough environment of the LHC experiments, with thousands of particles produced for every interaction. The high spatial resolution is particularly important in the region close to the primary interaction.

For these reasons, silicon detectors were chosen for the whole tracking system of the CMS experiment.

**THE PIXEL DETECTOR** is made of three cylinders in the barrel and two disks in each of the two endcaps. The spatial resolution on each hit is about  $10 \mu\text{m}$  in  $r-\varphi$  and about  $20 \mu\text{m}$  in  $z$ .

THE MICROSTRIP DETECTOR is made of concentrical layers in the barrel region, divided in two systems: the *tracker inner barrel* (TIB) and the *tracker outer barrel* (TOB).

The TIB has four layers in the region with  $20\text{ cm} < r < 55\text{ cm}$  and  $|z| < 65\text{ cm}$ . The sensors are  $10\text{ cm}$  long in  $z$ ,  $80$  to  $120\text{ }\mu\text{m}$  wide and  $320\text{ }\mu\text{m}$  thick. The single hit resolution is between  $23$  and  $34\text{ }\mu\text{m}$  in the  $r\text{-}\varphi$  plane, and about  $230\text{ }\mu\text{m}$  in  $z$ .

The TOB module is made of six layers, covering  $55\text{ cm} < r < 116\text{ cm}$  and  $|z| < 116\text{ cm}$ . The sensors have a length of  $25\text{ cm}$ , a width of  $180\text{ }\mu\text{m}$  and a thickness of  $500\text{ }\mu\text{m}$ . The single hit resolution is  $35$  to  $52\text{ }\mu\text{m}$  in the  $r\text{-}\varphi$  coordinates, and  $530\text{ }\mu\text{m}$  in  $z$ .

Two more modules are installed in the endcaps, arranged around the beam line: the *tracker inner disks* (TID) and *tracker outer endcaps* (TEC)

Each TEC module consists of nine disks, while the TID have three smaller disks. Both these modules have strips pointing radially toward the beam line with a thickness of  $320\text{ }\mu\text{m}$  to  $500\text{ }\mu\text{m}$ .

#### 2.2.4 The electromagnetic calorimeter

The ECAL is the inner calorimeter, made of  $61200$  lead tungstate crystals in the barrel and  $7324$  crystals in the endcaps. The base of the crystals is about  $2 \times 2\text{ cm}$  in size, with a depth of  $23\text{ cm}$ , corresponding to  $25$  radiation lengths.

The scintillation light is collected by avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps. They have a high gain and are insensitive to the magnetic field of the detector.

The energy resolution of the calorimeter has been measured in  $Z \rightarrow ee$  events from the 2011 data [24], as shown in figure 7.

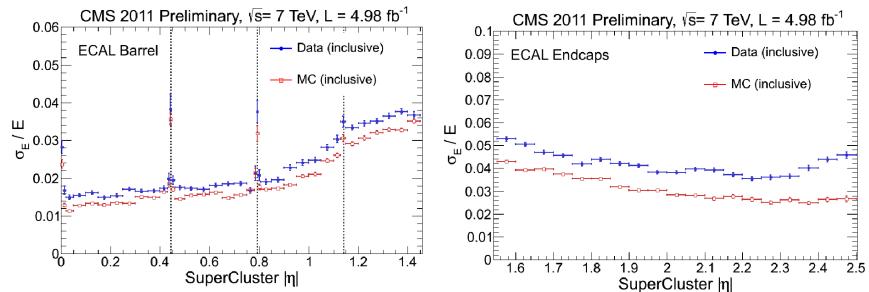


Figure 7: Energy resolution of the ECAL as measured in  $Z \rightarrow ee$  events in the 2011 data.

### 2.2.5 The hadron calorimeter

The HCAL surrounds the ECAL and is made of plastic scintillators interleaved with brass layers. It is composed of a *hadron barrel* up to  $|\eta| < 1.4$ , a *hadron endcap* for  $1.3 < |\eta| < 3.0$ . Two more systems provide very good hermeticity: the *hadron outer calorimeter* is placed outside the magnet in order to measure the energy of penetrating hadrons, while the *hadron forward calorimeter* extends the coverage up to  $|\eta| < 5$ .

### 2.2.6 The muon detectors

The muon system is installed in the return yokes of the magnet. Concentrical cylinders cover the barrel region up to  $|\eta| < 1.2$ , while four disks make up the endcap detectors ( $|\eta| < 2.4$ ). Three kind of gas detectors are employed: drift tubes in the barrel, cathode strip chambers in the endcaps, and resistive plate chambers in both. This last kind of detector has a fast response and is the most important for the trigger.

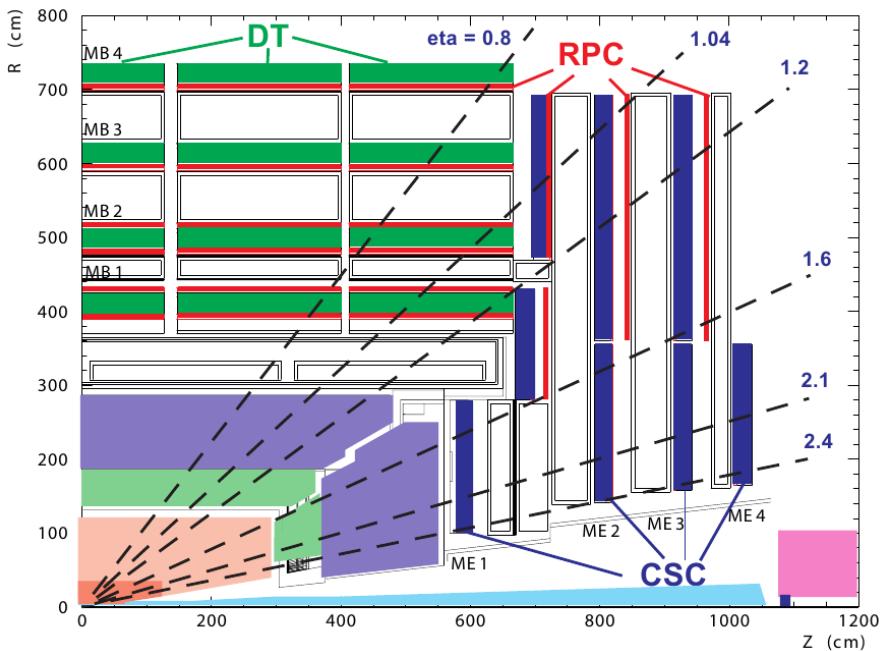


Figure 8: The layout of the muon detectors.

### 2.2.7 The trigger and data acquisition system

The design bunch spacing of the LHC is 25 ns, i.e. 40 MHz. About 100 interactions per second can be stored, so that the trigger has to provide a reduction factor of the order of  $10^5$ . The trigger system has two levels of selection: the *level 1* (L1) and the *high level trigger*.

**L1** is a hardware trigger, made of processors close to the detector components. It exploits coarse granularity information from calorimeters and muon chambers to identify electromagnetic clusters, muons, jets and missing  $E_t$ . Its processing time per event is on the order of a microsecond. It reduces the rate of events to less than 100 kHz.

**HLT** is a software trigger run on a farm of commercial CPUs, and has access to all event data, allowing for precise object reconstruction and energy/momentum evaluation. The HLT processing time is around 0.1 s per event.

# 3

## SIMULATION, RECONSTRUCTION AND DATASETS

---

### 3.1 SIMULATION

Accurate Monte Carlo simulations are needed in order to make sense of the data and model the behaviour of new physics signals in the detector. The physics of the collisions is simulated by the event generators MadGraph [7] and PYTHIA [39]. Then, the propagation of the particles in the detector has to be worked out. Two different types of simulation are used by the CMS collaboration: a GEANT4-based [28] simulation, known as the *Full Simulation*, and a detector model with simplified geometry, response evaluation and pattern recognition to decrease the processing time per event, the *Fast Simulation* [18]. Both the GEANT-based simulation and the Fast Simulation were employed in the generation of the background and signal samples for our studies.

#### 3.1.1 *The Full Simulation*

In the GEANT-based simulation, the energy deposits in the sensitive detector volumes are converted to electronic signals using algorithms based on the observed detector behavior, including the simulation of electronics noise and cross-talk. In many cases, the simulated electronic parameters are identical to those of the real devices; the constants specifying performance, calibrations, and noise behavior can be read from the same database used for the reconstruction of the real collision data. The output of this stage is simulated data in a format identical to that of the real raw data read from the detector. Further processing uses this data to simulate the formation of the L1 and HLT decisions using the same algorithms implemented online in the CMS trigger system. The simulated raw data that is produced is processed in a manner identical to that of the real data from LHC collisions.

#### 3.1.2 *The Fast Simulation*

This other approach makes a number of approximations:

**CMS GEOMETRY:** the Fast Simulation describes the detector with a simplified geometry of nested cylindrical layers. The particles are propagated from one layer to the next.

**MATERIAL EFFECTS:** five effects are taken into account. These are *bremssstrahlung*, photon conversion, multiple Coulomb scatter-

ing, energy loss through ionization and nuclear interactions. All of these are calculated analytically, except for nuclear interactions, since no analytical description is sufficient to describe the effect. Cross sections are taken from the PDG and the kinematics are derived from single particle collisions saved beforehand.

**TRACK RECONSTRUCTION:** the reconstruction is not usually part of the simulation. In particular, the track reconstruction is very time consuming. The hits of the simulated track are used to make a track candidate, but no fake tracks are simulated. This is possible because of the very low fake rate.

**MUONS:** muons are propagated through the tracker and calorimeters with average energy loss, then  $dE/dx$  and multiple scattering in the iron yokes of the muon chambers are computed. Muon simulated hits are produced in all sections of the muon systems.

**CALORIMETERS:** electron showers in the ECAL are simulated with the Grindhammer [37] parametrization. Photons undergo pair conversions based on the number of radiation lengths they have traversed. Detector effects such as energy leakage into the crystal gaps and into the HCAL are included, as well as electronic noise. Shower simulation in the HCAL is similar, with different types of particles parametrized from GEANT-based simulation results.

### 3.2 THE RECONSTRUCTION OF PHYSICS OBJECTS

The particle flow [19] (PF) event reconstruction aims at reconstructing and identifying all stable particles in the event. The essential idea is to analyse the event combining the information from all the available subdetectors in an optimal way.

The CMS detector, with its large silicon tracker and its extremely precise electromagnetic calorimeter, appears to be ideally suited for this purpose. The fundamental building bricks of the PF reconstruction are charged particle tracks, ECAL and HCAL clusters and muon tracks. These must be delivered with a high efficiency and low fake rate even in high-density environments like jets. A jet is a collection of particles resulting from the decay of a quark or gluon and emitted in the direction of the primary particle.

Reconsidering the design of the detector in figure 6, the logic used to interpret this diagram is akin to that employed by the PF. An ECAL cluster not linked to any track is a photon, an ECAL and HCAL cluster matched to a track gives a charged hadron, while an HCAL cluster without a track identifies a neutral hadron. Electrons are basically ECAL-only clusters linked to a track. Muons are recognized from the hits in the muon detectors.

From these basic elements, composite objects can be reconstructed, such as  $\tau$  leptons from their decay products, jets and missing transverse energy from all the particles in the event.

Quality requirements are needed in the reconstruction of the physics objects: thousands of particles are created in each event and their tracks can overlap. The quality selections are the result of detailed studies by the CMS collaboration, aiming for the best compromise between purity and efficiency. These recommendations are described in the following paragraphs.

### 3.2.1 *Track reconstruction*

The track reconstruction algorithm proceeds in an iterative fashion. The proton-proton interaction region, known as the beamspot, is used as a first estimate of the hard interaction point prior to the reconstruction of the primary vertex. The first step of the algorithm uses pixel hits only. Tracks are seeded from either triplets of hits or pairs of hits with an additional constraint from the beamspot or a pixel vertex. This first estimate of the trajectory is then propagated outwards in a search for compatible hits with the combinatorial track finder (CTF) [4]. An additional search for hits is performed from the outer boundaries of the detector and propagating inwards. Finally, the tracks are fitted and filtered to remove those that are likely fakes. Before starting the next iteration of the algorithm, hits that were unambiguously assigned to a track are removed from the collection to create a smaller, cleaner collection. The first iterations usually pick up prompt tracks, while the the following iterations find displaced tracks or tracks with missing hits in the pixel detector, allowing the reconstruction of conversions or nuclear interactions.

### 3.2.2 *Vertex reconstruction*

The reconstruction of the primary vertex starts from the track collection. Prompt tracks are selected based on the transverse impact parameter, number of hits, and  $\chi^2$ . They are then clustered, i. e. separated into groups forming vertex candidates. An adaptive vertex fit [26] is performed to find the point in space that is most compatible with the selected tracks.

The track and vertex reconstruction are intertwined as the additional information from the position of the vertex is then used to re-estimate track momenta and parameters.

### 3.2.3 *Muon reconstruction*

Muons are not stopped inside the CMS detector and leave only a tiny fraction of their energy in the calorimeters. The information from

the tracking system and the muon chambers is exploited for their reconstruction. The two systems are used independently in a first phase, where two algorithms are used:

**THE GLOBAL MUON RECONSTRUCTION**, or the *outside-in* approach, starts from a segment in the drift tubes or cathode strip chambers and extrapolates the seed layer by layer up to the tracker. If a matching track is found in the tracking system, the information from both tracks is combined to improve the resolution.

**THE TRACKER MUON RECONSTRUCTION**, or the *inside-out* method, extrapolates a track from the inner system to the muon chambers. The small energy loss due to interactions with the material of the magnets and calorimeters is taken into account, as well as an uncertainty arising from the possibility of multiple scattering.

The recommended selection requires these two algorithms to agree, as this improves the resolution for high- $p_T$  muons (figure 9).

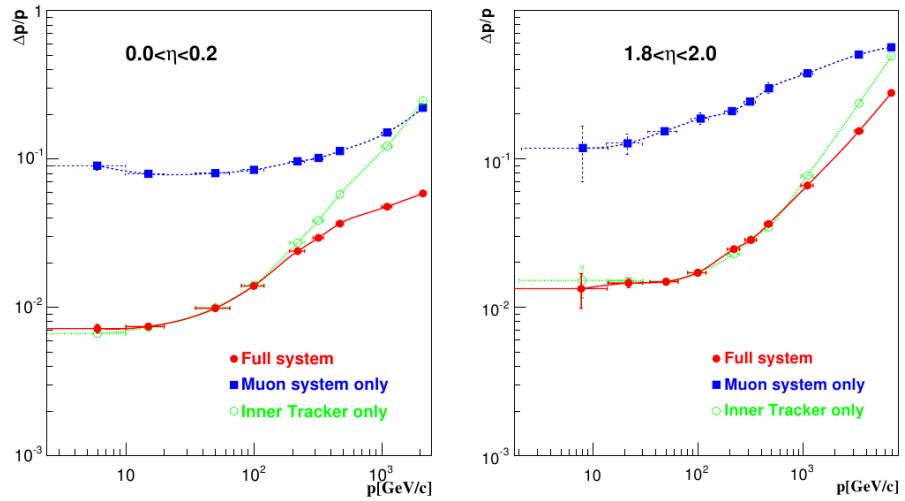


Figure 9: Muon momentum resolution as a function of  $p_T$  in regions with different pseudorapidity. The blue squares are for the muon system only, the green circles for the inner tracker only, and the red circles for the combined reconstruction

One of the most important observables for the muon candidates is the *relative isolation*. It indicates the amount of energy collected in the vicinity of the muon, by summing the contributions from the tracking system, the ECAL and the HCAL, divided by the  $p_T$  of the muon. The sum runs over all deposits within a cone of radius  $\Delta R = 0.3$  centred on the muon track, as illustrated in figure 10.

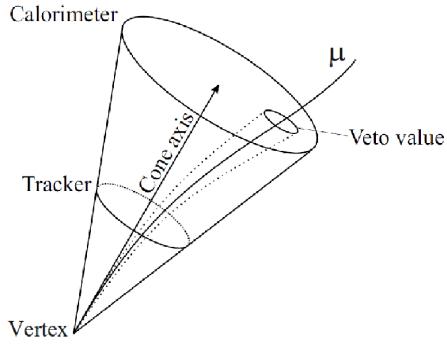


Figure 10: Illustration of the isolation observable.

### 3.2.4 Electron reconstruction

A customized track reconstruction was developed for electrons, where the hits are fitted using a *gaussian sum filter* (GSF) [16], i. e. the *bremsstrahlung* emission is modelled by a superposition of gaussian functions. However, this more sophisticated analysis demands more CPU power, at the order of a few hundreds milliseconds per track, and can be run only on a limited number of seeds.

The standard strategy to seed GSF tracks, hereafter called ECAL-driven seeding, heavily relies on the ability to gather into one single *supercluster* the entire energy deposit of the electron. The algorithm consists of three steps. First, cluster seeds are identified as local energy maxima above a given energy. Secondly, the seeds are grown by connecting cells with at least one side in common with a cell already in the cluster, and with an energy in excess of a given threshold. These thresholds are equivalent to two standard deviations of the electronic noise in the calorimeter, that is 80 MeV in the ECAL barrel, up to 800 MeV in the HCAL. Finally cluster energies and positions are determined from those clusters.

This implies collecting the electron and the *bremsstrahlung* photons energy deposits, which leads to clusters very extended along the azimuthal direction. This approach, which is efficient for isolated and high  $p_T$  electrons has unfortunately a limited efficiency in jets because the super-cluster collects additional contributions from other particles.

In addition to this ECAL-driven approach, the strategy developed for the Particle Flow [10] starts from the tracks and can be explained with two extreme examples. When the *bremsstrahlung* emission is negligible, the electron trajectory is determined with good precision by the standard tracking algorithm, and the track can be reconstructed up to the ECAL internal surface where it can be matched with the closest cluster. The momentum of the track can then be compared with the corresponding cluster energy, forming the  $E/p$  observable. If it is close to unity the track is selected to be then re-reconstructed with

the GSF algorithm. On the contrary, if the electron loses a substantial fraction of its energy by *bremstrahlung* emission, the characteristics of the track are exploited. Indeed, either the pattern recognition manages to accommodate for the changes of curvature, and the fitted track  $\chi^2$  is usually large, or it cannot follow the electron trajectory and the track is short. Electron tracks are then selected based on the attempt made by the standard algorithm to reconstruct it.

The tracker-driven seeding developed for the PF reaches a 80% efficiency in the barrel with a 10% probability to select wrongly a pion, with an acceptable CPU consumption; a more sophisticated treatment described below is further applied to reject fakes at the final identification level. In addition, this new algorithm improves the overall seeding efficiencies for isolated electron in CMS by 15% at 5 GeV with respect to the standard ECAL-driven seeding.

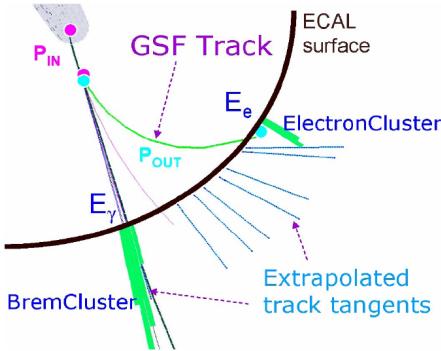


Figure 11: An illustration of the *bremstrahlung* recovery algorithm.

The procedure to collect all the calorimeter energy deposits, i.e., the *bremstrahlung* recovery, is driven by the GSF track (figure 11). For each tracker layer, where the material is mainly localized, a *bremstrahlung* photon emission is sought by computing a straight-line extrapolation, tangent to the track, up to the ECAL. If it matches a cluster, not already linked to a track, this cluster is selected as part of the electron. This procedure allows between 96% and 99% of the energy deposited in the ECAL to be collected.

The relative isolation is defined with the same method already described for the muons.

The charge of the electrons can be measured in three ways:

**THE GSF TRACK** whose curvature in the magnetic field of the detector defines the sign of the charge of the electron.

**THE CTF TRACK** algorithm for the reconstruction of the tracks. It accounts for multiple scattering with Kalman filter techniques [9]. The curvature of this track is another measurement of the charge.

**SUPERCLUSTER RELATIVE POSITION** the relative position of the track seed and the largest deposition in the ECAL provides an independent estimate of the electric charge.

Requiring these three methods to agree strongly decreases the probability of a charge mismeasurement. This is particularly important for our analysis, since the production of opposite-sign electrons is very abundant in Standard Model processes, and the accurate identification of the charge of the electrons almost eliminates these contributions.

Electrons can also be created from the conversion of an energetic photon. These are rejected if one of the following occurs:

**AT LEAST ONE MISSING HIT** in the tracker. Photons are neutral, so they usually miss hits in one or more layers.

**TRACK DISTANCE**  $< 0.02$  cm. An electron coming from a conversion has a partner positron track. If an oppositely charged track is found within this distance, the electron is rejected. The distance is calculated analytically by reconstructing the two helices as shown in figure 12

$\Delta \cot \theta < 0.02$ . A geometrical variable also related to the partner track:  $\theta$  is the angle between the two tracks in the  $yz$  plane.

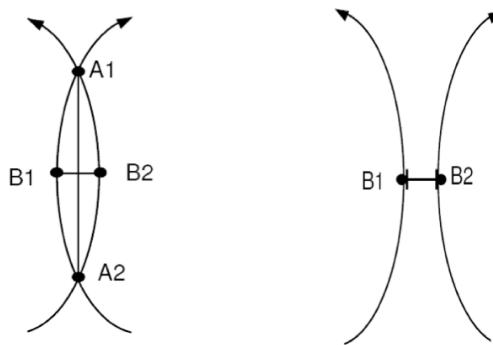


Figure 12: The track distance for conversion rejection is the two dimensional distance in the  $xy$ -plane between  $B_1$  and  $B_2$ . At these points, the two tracks from the photon conversion are parallel. The distance is defined to be negative if the two tracks overlap, and is positive otherwise.

### 3.2.5 Jet reconstruction

A quark or gluon produced in the hard scattering process undergoes hadronization due to colour confinement. Many hadrons are created, mostly in the same direction of the original parton, because of momentum conservation. All of these collimated particles form a jet.

However there can be difficulties in collecting all the particles in a jet:

- two or more jets may overlap, resulting in ambiguities in the assignment of the particles
- particles could be generated with a large momentum relative to the original parton direction, leading to the particle not being counted in the jet
- initial state radiation and particles from additional soft proton-proton interactions may contribute with additional tracks

The anti- $k_T$  algorithm [11] is used for the jet reconstruction. In this algorithm the distances between the particles  $i$  and  $j$   $d_{ij}$  are defined, as well as the distance between the  $i$ th-particle and the beam  $d_{iB}$ :

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta_{ij}}{R^2}$$

$$d_{iB} = p_{Ti}^{-2}$$

Where  $\Delta_{ij}^2 = (y_i - y_j)^2 + (\varphi_i - \varphi_j)^2$ , and  $p_{Ti}$ ,  $y_i$  and  $\varphi_i$  are the transverse momentum, rapidity and azimuthal angle of the  $i$ th particle;  $R = 0.5$  is a parameter representing a typical radius of the jet.

The algorithm then compares  $d_{ij}$  and  $d_{iB}$  iteratively:

- if  $d_{ij} < d_{iB}$  for some  $j$ ,  $i$  and  $j$  are merged into the same jet candidate
- if  $d_{ij} > d_{iB}$  for all  $j$ , then  $i$  is a jet.

After the jets have been reconstructed, the energy has to be scaled to account for possible mismeasurements [20]. The following corrections are used in this analysis:

**CHARGED HADRON SUBTRACTION:** charged hadrons originating from a secondary vertex are eliminated from the collection of particles in the jet reconstruction algorithm

**LEVEL 1 CORRECTION:** removes the energy coming from pile-up events. In principle this will remove any dataset dependence on luminosity so that the following corrections are applied upon a luminosity independent sample.

**LEVEL 2 CORRECTION:** makes the response independent from  $\eta$ , by using correction factors determined in dijet events.

**LEVEL 3 CORRECTION:** makes the response independent from the  $p_T$  of the jet. The correction factors are calculated from  $Z/\gamma +$  jets events.

**L2L3 RESIDUAL CORRECTIONS:** are only applied to the data, and not to Monte Carlo samples. This fixes the small differences between the data and Monte Carlo reconstructions, in a  $p_T$  and  $\eta$  dependent fashion.

The same corrections can be applied to the MC and data because of the very high quality of the simulation. The residual corrections account for the small difference. The recommendations from [38] are also applied, by selecting jets with:

- neutral hadron fraction  $< 0.99$
- charged hadron fraction  $< 0.99$
- number of constituents  $> 1$

### 3.2.5.1 Jet $b$ -tagging

The signal features two  $b$  quarks in the final state of each event. Many techniques have been developed to identify jets originated by a  $b$  quark. In this analysis, the *combined secondary vertex* algorithm [17] is used. This sophisticated and complex tag exploits all known variables, which can distinguish  $b$  from non- $b$  jets. Its goal is to provide optimal  $b$ -tag performance, by combining information about impact parameter significance, the secondary vertex and jet kinematics. The standard *loose* working point of this algorithm is employed, leading to a probability of a light-flavour (i. e. up, down or strange) jet passing this selection to 10%, as measured on jets with  $p_T$  of about 80 GeV in QCD Monte Carlo events [21].

## 3.3 OBJECT SELECTION

Signal events in our studies have a clean signature: two clean, isolated, same-sign leptons and many jets (see figure 17). To reduce the probability of pions and other objects of being wrongly identified as leptons further quality requirements are enforced.

### 3.3.1 Muon selection

The muons considered for our analysis have the following properties, mainly concerning the quality of the reconstructed track, besides high  $p_T$  and isolation:

- $p_T > 30 \text{ GeV}$
- relative isolation  $< 0.20$
- $|\eta| < 2.4$
- reconstructed as global and tracker muon
- at least eleven silicon hits
- transverse impact parameter  $< 0.02 \text{ cm}$
- normalized  $\chi^2 < 10$

- at least one hit in the muon system
- at least one hit in the pixel detectors
- at least two muon chambers whose track segments match

### 3.3.2 *Electron selection*

The selection of electron candidates is more complex due to the high material budget in the CMS detector and high magnetic field. More variables are needed to achieve good discriminating power against fakes. The CMS collaboration studied the following variables for electron identification:

**TRACKER-ECAL MATCHING**, with  $\Delta\phi$ ,  $\Delta\eta$  and  $E/p$  between the track reconstructed in the silicon detector and the ECAL energy deposits being compared

**HADRON FRACTION H/E**, where the energy collected in the HCAL directly behind the ECAL cluster is measured

**CLUSTER SHAPE** in  $\eta$

**IMPACT PARAMETER** with respect to the primary vertex

**CONVERSION REJECTION** with missing hits and partner track matching, as described in paragraph 3.2.4

**ISOLATION** of the electron in the tracker, ECAL and HCAL

The algorithm gives as output a selection for each electron candidate for nine defined severity levels. In this analysis, the HyperTightMC1 level is used. The efficiency and fake rate of this selection are shown in figures 13 and 14.

Our final electron selection includes:

- $p_T > 30 \text{ GeV}$
- relative isolation  $< 0.15$
- $|\eta| < 2.4$ , also excluding the transition region from the barrel to the endcap in the ECAL, with  $1.444 < |\eta| < 1.566$ .
- HyperTightMC1 category
- conversion rejection
- transverse impact parameter  $< 0.02 \text{ cm}$
- GSF-CTF-SC charge consistency (see paragraph 3.2.4)
- $\Delta R > 0.1$  from all muon tracks. It can happen that a muon track matches by chance with an ECAL energy deposit. It is particularly important for the counting of  $e\mu$  events that we remove such fakes.

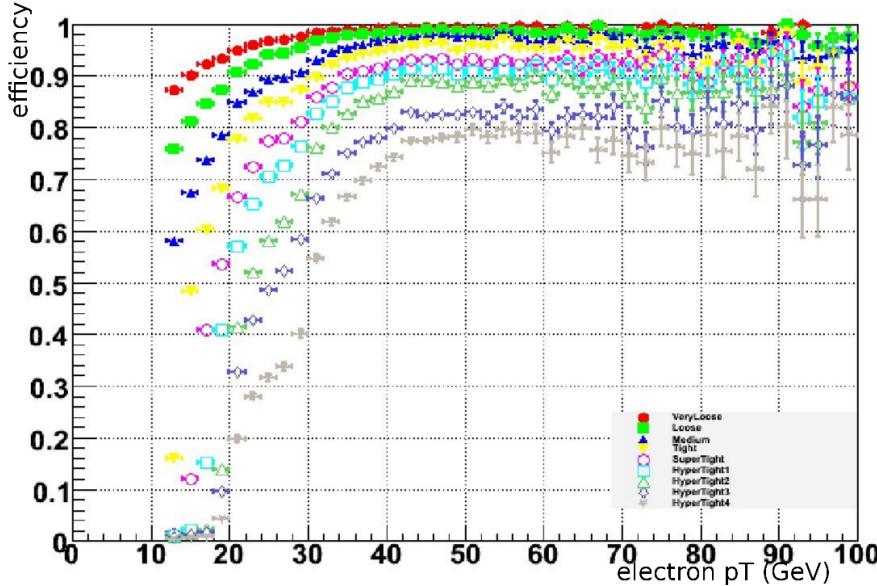


Figure 13: Selection efficiency for the various working points, measured in  $Z \rightarrow ee$  and  $W \rightarrow ee$  events. In this analysis the HyperTightMC1 selection (light blue squares) is used.

### 3.3.3 Jet selection

The jet selection has only few requirement in addition to the baseline reconstruction described in 3.2.5:

- $p_T > 30 \text{ GeV}$
- $|\eta| < 2.4$
- $\Delta R > 0.3$  between the jet and the selected leptons, to avoid double counting the leptons as jets.

### 3.3.4 Event cleanup and vertex selection

A pre-selection for the events includes some requirements on the quality of the overall reconstruction:

- the event is rejected if the fraction of high-purity tracks is less than  $1/4$  in events with at least ten tracks
- at least one primary vertex, with at least four degrees of freedom and  $\Delta z < 24 \text{ cm}$ ,  $\Delta r < 2 \text{ cm}$  with respect to the nominal interaction point
- events with high levels of noise in the HCAL are removed

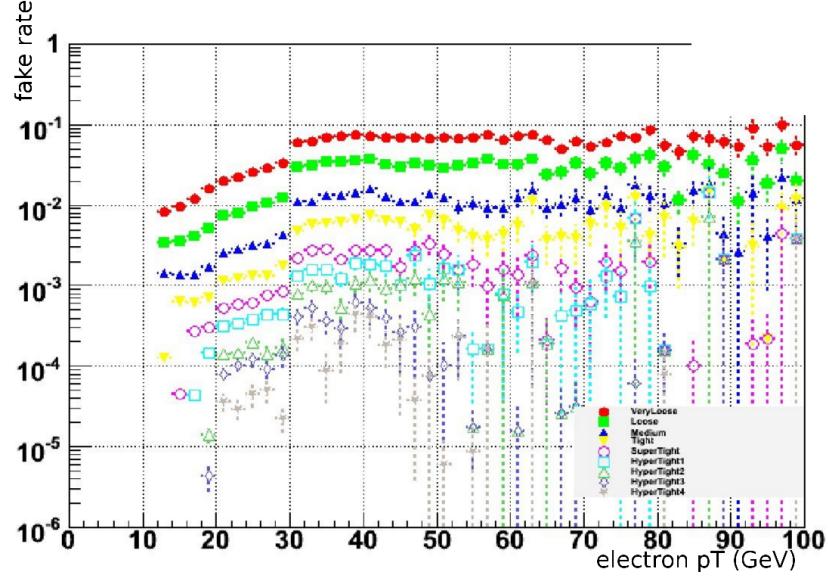


Figure 14: Selection fake rate for the various working points, measured in QCD dijet events. In this analysis the HyperTightMC1 selection (light blue squares) is used.

### 3.3.5 Trigger requirements

HLT selections are also used in our analysis. Double lepton triggers are employed with a relatively low  $p_T$  threshold. The triggers used for the three decay channels are listed in table 1.

---

HLT_DoubleMu7_v1,2 or
HLT_Mu13_Mu8_v2,3,4,6,7 or
HLT_Mu17_Mu8_v10,11
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v1,2,3,4,5,6 or
HLT_Ele17_CCTT_Ele8_CCTT_v6,7,8,9,10
HLT_Mu10_Ele10_CaloIdVL_v2,3,4,or
HLT_Mu17_Ele8_CaloIdVL_v1,2,3,4,5,6,8 or
HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_v4,7,8 or
HLT_Mu8_Ele17_CaloIdL_v1,2,3,4,5,6 or
HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_v3,4,7,8

---

Table 1: List of triggers used in the analysis for the three decay channels:  $\mu\mu$ ,  $ee$ ,  $e\mu$ . CCTT is short for CaloIdL\_CaloIsoVL\_TrkIdVL\_TrkIsoVL.

### 3.4 DATASETS

This analysis used the data recorded in 2011 at the CMS detector, corresponding to an integrated luminosity of  $5.0 \text{ fb}^{-1}$  of proton-proton collisions at a centre of mass energy of  $\sqrt{s} = 7 \text{ TeV}$ . The data is divided into *runs* in which the beam and detector conditions are stable. The runs are further divided into *lumisections*, corresponding to about 23 s.

The CMS collaboration certifies a list of good runs based on the conditions of the detector. This work is based on the official list of *golden runs* for physics analyses.

To enable the most effective access to CMS data, the data are first split into *primary datasets* (PDs). The division into PDs is done based on the trigger decision. The datasets used in this analysis are triggered with a pair of leptons and they are shown in table 2.

Dataset	Run range
/DoubleMuon/Run2011A-May10ReReco-v1/AOD	160329-163869
/DoubleMuon/Run2011A-PromptReco-v4/AOD	165071-168437
/DoubleMuon/Run2011A-05AugReReco-v1/AOD	170053-172619
/DoubleMuon/Run2011A-PromptReco-v6/AOD	172620-175770
/DoubleMuon/Run2011B-PromptReco-v1/AOD	175832-180296
/DoubleElectron/Run2011A-May10ReReco-v1/AOD	160329-163869
/DoubleElectron/Run2011A-PromptReco-v4/AOD	165071-168437
/DoubleElectron/Run2011A-05AugReReco-v1/AOD	170053-172619
/DoubleElectron/Run2011A-PromptReco-v6/AOD	172620-175770
/DoubleElectron/Run2011B-PromptReco-v1/AOD	175832-180296
/MuEG/Run2011A-May10ReReco-v1/AOD	160329-163869
/MuEG/Run2011A-PromptReco-v4/AOD	165071-168437
/MuEG/Run2011A-05AugReReco-v1/AOD	170053-172619
/MuEG/Run2011A-PromptReco-v6/AOD	172620-175770
/MuEG/Run2011B-PromptReco-v1/AOD	175832-180296

Table 2: Data samples used for the analysis.

### 3.5 SIGNAL MONTE CARLO SAMPLES

The model described in 1.5 has been implemented in the MadGraph event generators and eight samples corresponding to values of the  $T_{5/3}$  mass from 400 to 750 GeV were generated. In these samples, two of the same-sign W bosons arising from the decay of the top partners were forced to decay leptonically. The approximate next-to-next-to-

leading order cross sections were calculated using HATHOR [5]. The masses, cross sections and the number of generated events for the eight mass points are listed in table 3.

mass (GeV)	$\sigma \times \text{BR}$ (pb)	events
400	0.295	86205
450	0.139	86211
500	0.069	86684
550	0.036	86724
600	0.019	86965
650	0.011	87592
700	0.006	88145
750	0.004	88410

Table 3: Signal Monte Carlo samples. The branching ratio is 0.21.

### 3.6 BACKGROUND MONTE CARLO SAMPLES

Some rare Standard Model processes also produce isolated, same-sign leptons in their final state. These include the production of two or more vector bosons,  $WZ$ ,  $ZZ$ ,  $W^\pm W^\pm$  and  $WWW$ , and the production of a top-antitop pair with an associated vector boson,  $t\bar{t}W$  and  $t\bar{t}Z$ . Some of them were not available in the official productions from the CMS Collaboration and were privately produced with the Fast Simulation of the detector, described in paragraph 3.1.

The background samples used in this analysis are listed in table 4.

process	MC generator	$\sigma$ (pb)	events
$WZ + \text{Jets}$	MADGRAPH	0.879	1221134
$ZZ + \text{Jets}$	MADGRAPH	0.076	1185188
$W^+W^++\text{Jets}$	MADGRAPH	0.165	130000
$W^-W^-+\text{Jets}$	MADGRAPH	0.055	160000
$WWW + \text{Jets}$	MADGRAPH	0.038	1201777
$t\bar{t}W$	MADGRAPH	0.169	1029608
$t\bar{t}Z$	MADGRAPH	0.139	793155

Table 4: Details of the background Monte Carlo samples used for the analysis.

### 3.7 MONTE CARLO PILE-UP REWEIGHTING

The pile-up in events at the LHC is constantly evolving with the increase in instantaneous luminosity at which the machine operates (figure 15). Pile-up describes events coming from additional proton-

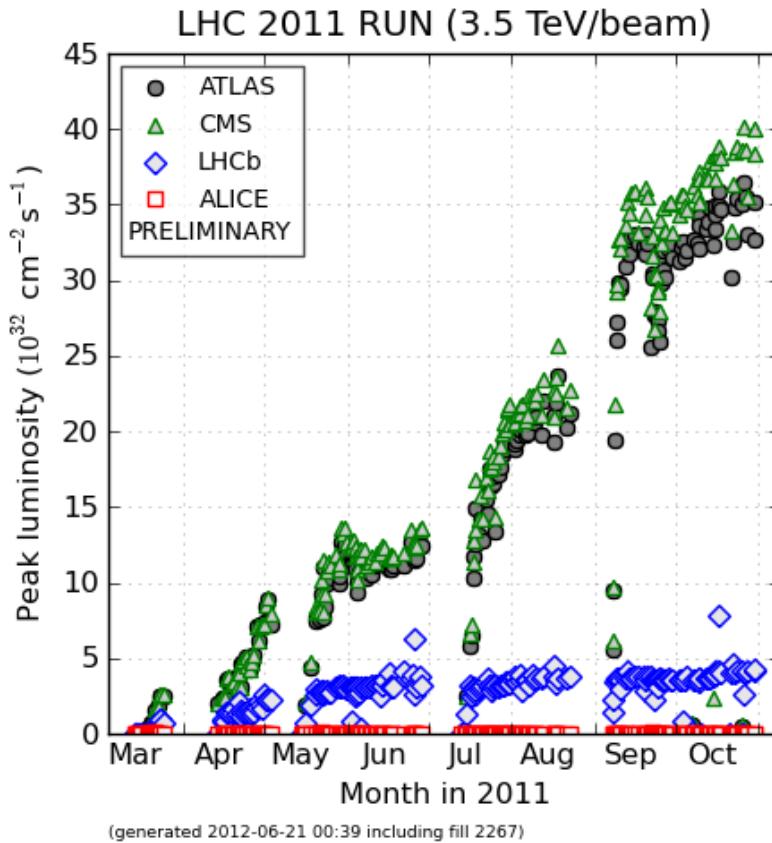


Figure 15: Peak instantaneous luminosity delivered by the LHC in 2011.

proton interactions. It is very unlikely to have two hard collisions occurring at the same time, but as the luminosity goes up, the density of the protons in the interaction region increases as well, thus giving rise to additional soft interactions between many proton pairs in the same bunch crossing.

In addition to this effect, known as *in-time* pile-up, we have the contributions from *out-of-time* pile-up: the design bunch spacing of the LHC is 25 ns, faster than the read-out response for some sub-detectors. It is therefore possible that the signals from particles generated in previous collisions are integrated. electronics from successive bunch crossings.

The simulated events are usually produced before the data are collected, with a choice of the pile-up distribution covering the expected condition, based on information from the accelerator. Therefore they

need taking into account the measured pile-up conditions in order to avoid a systematic bias, as shown in figure 16

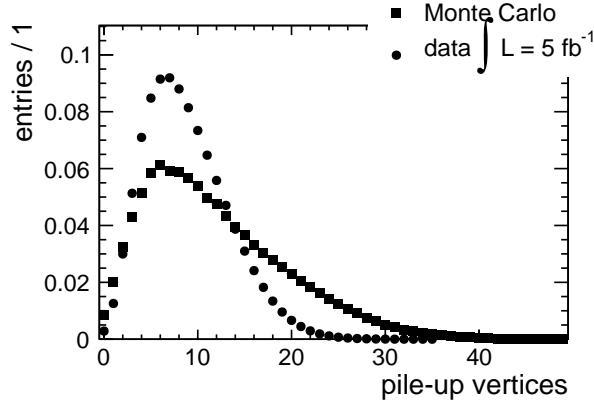


Figure 16: Distribution of the number of vertices in unweighted Monte Carlo and data. The Monte Carlo samples are weighted so that the two distributions match.

This is done by calculating a weight for every Monte Carlo event, depending on the number of pile-up vertices. The weights are calculated so that the pile-up distribution of the Monte Carlo matches the distribution measured in the data.

# 4

## EVENT PRESELECTION

The experimental signature of our signal is the presence of two same-sign prompt leptons and at least four jets (see figure 17).

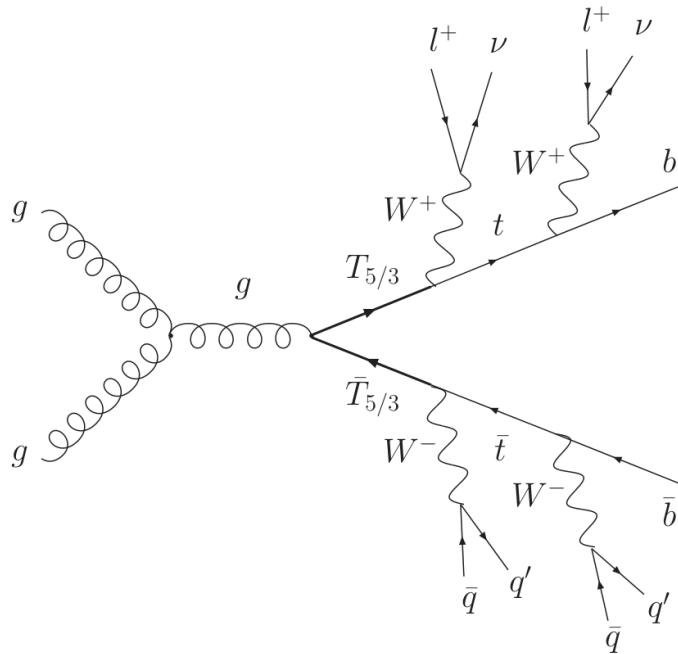


Figure 17: Pair-production of the  $T_{5/3}$ , decaying into same-sign leptons and at least four jets.

The backgrounds associated with this signal can be divided into three categories:

**TWO SAME-SIGN PROMPT LEPTONS** from rare SM processes.

**OPPOSITE-SIGN LEPTONS** whose charge is mismeasured, leading to a same-sign final state.

**ONE PROMPT AND ONE NON-PROMPT LEPTON WITH THE SAME CHARGE** (or two non-prompt same-sign leptons). This background arises mainly from  $t\bar{t}$  events where one of the leptons comes from the decay of a  $b$  quark, but is mistaken for a prompt lepton.

In order to reject all of these background sources, the following baseline selection is applied:

- event cleanup and trigger

- exactly two same-sign leptons, rejecting events with three or more leptons
- quarkonia veto, with the invariant mass of the leptons greater than 20 GeV
- Z veto, with the invariant mass of the ee and  $\mu\mu$  pairs not in the 76–106 GeV window. This selection is not needed for  $e\mu$  events for lepton number conservation.
- at least four jets

The event cleanup and trigger requirements are detailed in 3.3.4 and 3.3.5. The invariant mass selections are particularly well suited to reject events with charge misidentification, in addition to the charge consistency requirement on page 16, particularly from  $Z \rightarrow ee$  decays. Figure 19 clearly shows a peak at around the mass of the Z boson for the double electron channel. The effect is much smaller for the muons. The distribution of the number of jets in the signal and backgrounds (figure 18) also shows that this is a good discriminating variable against both the prompt and the  $t\bar{t}$  backgrounds.

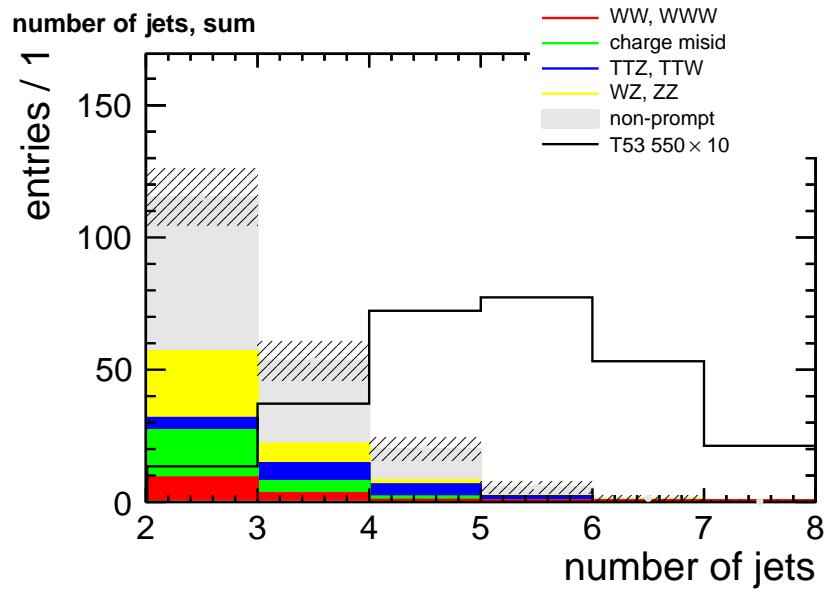


Figure 18: Events with two same-sign leptons and two jets. Distribution of the number of jets in the backgrounds and signal for the 550 GeV mass point for the sum of the three decay channels, which are shown separately in figure 18. The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties.

While the prompt same-sign background is well modelled by the Monte Carlo, we found that a data-driven method is needed for an ac-

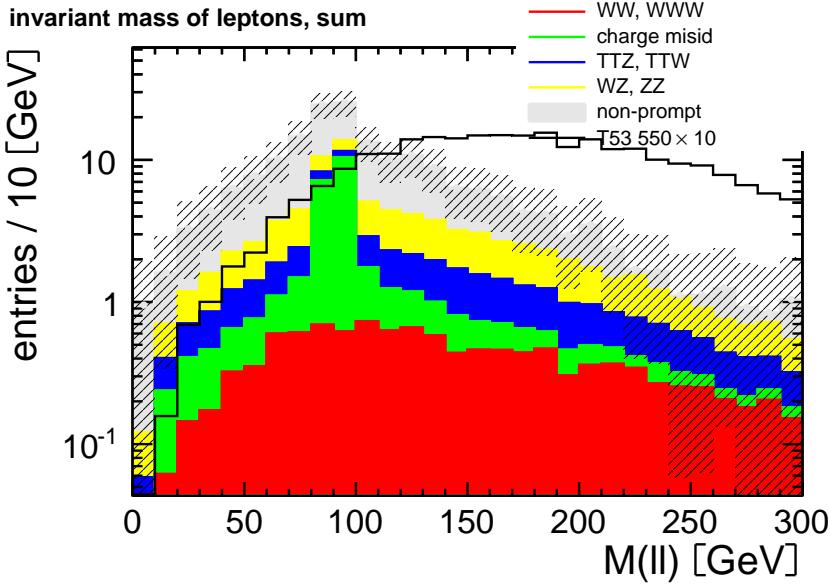


Figure 19: Events with two same-sign leptons and two jets. Distribution of the invariant mass of the leptons in the backgrounds and signal for the 550 GeV mass point for the sum of the three decay channels, which are shown separately in figure 46. The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties.

curate description of the contributions from charge misidentification and non-prompt leptons. That is the subject of the following chapter.



## A DATA-DRIVEN BACKGROUND ESTIMATE

---

In this chapter the data-driven method for estimating the backgrounds from charge misidentification and from non-prompt leptons is described.

### 5.1 CHARGE MISIDENTIFICATION

Standard Model processes with prompt leptons of opposite charge are very common at the LHC. They can contribute to the top partners background if the charge of one of the leptons is incorrectly measured. For muons in the  $p_T$  range considered in this analysis, the charge misidenification rate is extremely small ( $\approx 10^{-5}$ ) and their contribution to the background is negligible [15]. The magnitude of this contribution due to electrons can be derived from the data by using the Z boson resonance. The fraction of misidentified leptons is obtained by considering events with two electrons in which the invariant mass of the leptons falls inside the Z boson mass window:  $76\text{ GeV} < M(\ell\ell) < 106\text{ GeV}$ . These events must pass all cleaning and trigger requirements and the electrons are likewise required to pass all quality cuts, but the selection based on the number of jets. As can be seen in the tables 5 and 6, this sample is expected to be completely dominated by Z+jets.

	total	$t\bar{t}$	Z+jets	W+jets	WZ	ZZ	$t\bar{t} W$	$t\bar{t} Z$
in Z window	$9.3 \cdot 10^5$	644	$9.3 \cdot 10^5$	35	205	46.2	1.89	8.62
same-sign	1090	1.84	1050	9.97	8.9	1.02	0.62	0.20

Table 5: Expected event yields within the Z invariant mass window ( $76\text{ GeV} < M(\ell\ell) < 106\text{ GeV}$ ) for the background contributions. The backgrounds not shown yield negligible contributions to this phase space region.

$T_{5/3}$ mass [GeV]	400	450	500	550	600	650	700	750
in Z window	22.6	9.84	4.04	1.81	0.928	0.464	0.224	0.123
same-sign	12.1	4.74	1.79	0.716	0.359	0.145	0.075	0.0345

Table 6: Expected event yields within the Z invariant mass window ( $76\text{ GeV} < M(\ell\ell) < 106\text{ GeV}$ ) for the eight  $T_{5/3}$  signal points.

Applying the same method to data we obtain 858330 events in the  $Z$  mass window; 1014 events remain upon applying the same-sign requirement. The probability of misidentifying the charge of each lepton is computed by assuming that the entirety of the same-sign contribution is due to misidentified  $Z + \text{jets}$  events and is found to be  $5.89 \times 10^{-4}$ . To confirm that this upper limit is close to the true probability, we consider the dilepton invariant mass spectrum shown in Figure 20. Both the simulation and the mass spectrum suggest that the upper limit is close to the true probability (within approximately 1% according to the simulation). We note that this estimated charge misidentification probability is consistent with the measurement in [15].

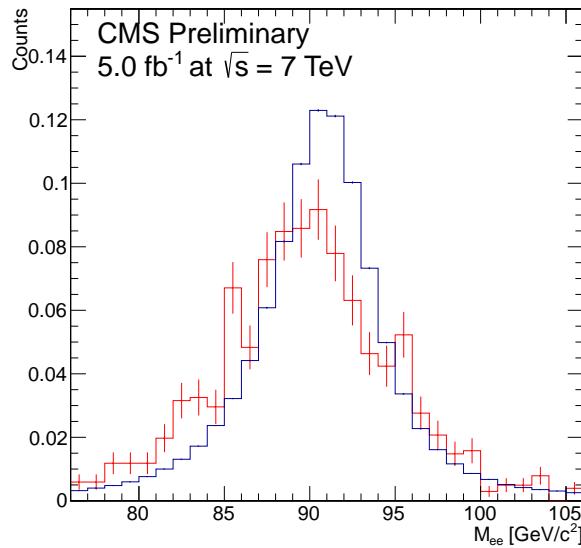


Figure 20: The dilepton invariant mass spectrum for electrons in data. The opposite-sign distributions are shown in blue and the same-sign ones are in red. Both distributions are normalized to unit area.

The number of expected same-sign events due to charge misidentification is estimated by considering the total number of events passing the full selection, but having oppositely charged leptons. These events are multiplied by the charge misidentification probability.

## 5.2 SAME-SIGN NON-PROMPT BACKGROUND

Non-prompt leptons are the result of the decays of heavy flavour quarks, decays in flight or conversions. The description of this background in the Monte Carlo samples is unsatisfactory, and a data-driven estimate is needed.

The so called *tight-loose* method is employed. The goal of this method is to measure the probability of a non-prompt lepton passing the anal-

ysis selection (5.2.1) and then to calculate the expected number of events contaminating our final sample by using this probability (5.2.4).

### 5.2.1 Measuring the tight-to-loose ratio

This method defines two types of leptons: a *tight* lepton, which is exactly the same as in 3.3.1 and 3.3.2, and a *loose* lepton, whose selection criteria are relaxed as follows:

#### LOOSE ELECTRON

- no transverse impact parameter requirements
- no conversion rejection
- relative isolation relaxed from 0.15 to 0.60

#### LOOSE MUON

- transverse impact parameter relaxed from 0.02 cm to 2 cm
- relative isolation relaxed from 0.20 to 0.40
- no requirements on the number of hits in the tracker, nor on the number of muon chambers with matching segments
- normalized  $\chi^2 < 50$  instead of 10

These selections are studied to include e. g. leptons from heavy flavour hadrons: they are usually close to a jet, hence not isolated, and they also have a larger impact parameter, as the hadron decays on a much larger timescale with respect to a  $T_{5/3}$  or a top quark. Conversion electrons also have to be taken into account, so that the requirement against conversion rejection has to be eliminated as well.

The relevant quantity is the tight-to-loose ratio, representing the probability that a non-prompt lepton is identified as tight:

$$f = \frac{\text{number of tight leptons}}{\text{number of loose leptons}}$$

However, this ratio is close to the probability for a non-prompt lepton to pass the tight selection only if it is calculated on a sample where the prompt component is negligible. Such a sample can be built from the `SingleMu` and `DoubleElectron` primary datasets by enforcing cuts aiming to reject leptons from W and Z bosons:

- exactly one loose lepton in each event
- at least one jet with  $p_T > 40\text{ GeV}$  and  $\Delta R > 1$  relative to the lepton
- $E_T^{\text{miss}} < 25\text{ GeV}$

- $m_T < 25 \text{ GeV}$
- extended Z veto: reject the event if the invariant mass of the lepton and any jet is in the  $76\text{-}106 \text{ GeV}$  window

The  $E_T^{\text{miss}}$  and  $m_T$  selections are meant to exclude the presence of W bosons, and the Z veto now includes all the jets, as a second lepton can be incorrectly identified as a jet.

Even with this selection, the contamination from prompt leptons is still significant, especially in the muon case. This can be estimated from the Monte Carlo W + jets and Z + jets samples. The results are shown in tables 7 and 8.

	loose	tight	ratio
W + jets	$31.6 \pm 0.21$	$29.3 \pm 0.20$	-
Z + jets	$21.9 \pm 0.08$	$20.4 \pm 0.08$	-
t̄t	$0.7 \pm 0.01$	$0.6 \pm 0.01$	-
Data	1974	274	$0.138 \pm 0.009$
Data-MC	$1919.8 \pm 44.43$	$223.7 \pm 16.55$	$0.117 \pm 0.009$

Table 7: Tight to loose ratio for electrons in the data and in the MC. The line Data-MC subtracts from the data the prompt lepton contribution due to W + jets, Z + jets and t̄t from the data.

	loose	tight	ratio
W + jets	$4505.0 \pm 26.36$	$4374.8 \pm 25.97$	-
Z + jets	$604.0 \pm 4.36$	$585.8 \pm 4.29$	-
t̄t	$85.5 \pm 1.20$	$80.4 \pm 1.16$	-
Data	33770	12170	$0.360 \pm 0.004$
Data-MC	$28575.5 \pm 185.70$	$7129.1 \pm 113.42$	$0.249 \pm 0.004$

Table 8: Tight to loose ratio for muons in the data and in the MC. The line Data-MC subtracts from the data the prompt lepton contribution due to W + jets, Z + jets and t̄t from the data.

In order to avoid being directly dependent on the MC, we note that the contamination from prompt leptons is negligible if we restrict the  $p_T$  range of the leptons to  $25 \text{ GeV} < p_T < 35 \text{ GeV}$ . The overlap of this  $p_T$  range with that of the signal selection in the analysis ( $p_T > 30 \text{ GeV}$ ) is small, but when the tight-to-loose ratio is corrected for the contribution of prompt leptons, it becomes nearly independent from the  $p_T$  (see figure 21).

	loose	tight	ratio
W + jets	$7.0 \pm 0.10$	$6.2 \pm 0.09$	-
Z + jets	$8.1 \pm 0.05$	$7.2 \pm 0.05$	-
t̄t	$0.1 \pm 0.00$	$0.1 \pm 0.00$	-
Data	1698	191	$0.112 \pm 0.009$
Data-MC	$1682.9 \pm 41.21$	$177.5 \pm 13.82$	$0.105 \pm 0.009$

Table 9: Tight to loose ratio for electrons in the data and in the MC with the  $p_T$  range of the electron restricted to the 25-35 GeV range. The line Data-MC subtracts the prompt lepton contribution due to W + jets, Z + jets and t̄t from the data.

	loose	tight	ratio
W + jets	$995.1 \pm 12.37$	$942.9 \pm 12.03$	-
Z + jets	$355.4 \pm 3.33$	$339.1 \pm 3.25$	-
t̄t	$13.9 \pm 0.48$	$11.8 \pm 0.45$	-
Data	40479	11536	$0.285 \pm 0.003$
Data-MC	$39114.6 \pm 201.60$	$10242.2 \pm 108.13$	$0.261 \pm 0.003$

Table 10: Tight to loose ratio for muons in the data and in the MC with the  $p_T$  range of the muon restricted to the 25-35 GeV range. The line Data-MC subtracts the prompt lepton contribution due to W + jets, Z + jets and t̄t from the data.

With this  $p_T$  restriction, the contributions from prompt leptons in the MC becomes negligible (tables 9 and 10), and in the following analysis we use only the values from the data:

$$f_e = 0.112 \pm 0.009 \text{ stat.}$$

$$f_\mu = 0.285 \pm 0.003 \text{ stat.}$$

### 5.2.2 Heavy flavour content of the samples

The samples in which these ratios are measured are selected without any requirements on the heavy flavour content. However, a large part of the non-prompt background is expected to arise from semileptonic t̄t. We therefore have to check that our prediction is stable with respect to variations in the flavour content.

The b-tagging algorithm described in 3.2.5.1 is used to enrich our samples of leptons coming from b quarks. This doesn't change the

values calculated in the previous section, within the statistical uncertainties:

$$f_e^{\text{btag}} = 0.120 \pm 0.009 \text{ (stat.)}$$

$$f_\mu^{\text{btag}} = 0.280 \pm 0.003 \text{ (stat.)}$$

### 5.2.3 Prompt ratio approximation

The *prompt ratio*  $p$  can also be defined as the probability for a prompt lepton to pass the tight selection. It can be studied with a tag and probe method in Drell-Yan events where the invariant mass of the leptons is within 10 GeV of the Z mass. This gives values close to one for both electrons ( $p_e = 0.92$ ) and muons ( $p_\mu = 0.98$ ). This efficiency is confirmed in Monte Carlo Z and W studies.

This justifies the approximation in all of the following formulae, where  $p_e = p_\mu = 1$  is assumed. This allows to simplify the calculation of the non-prompt background, as well as the treatment of the uncertainties.

### 5.2.4 Calculation of the expected event yields

The measure of the non-prompt rate  $f_\ell$  allows to make a prediction on the number of events with non-prompt leptons contaminating our final selection.

The idea of the method is best illustrated in the simpler case of single lepton events. The total number of leptons  $N_\ell$  passing the loose criteria is composed of  $N_p$  prompt leptons and  $N_f$  non-prompt leptons. These numbers are not directly measurable but they can be related to the number of leptons where no lepton ( $N_{t0}$ ) or one lepton  $N_{t1}$  passes the tight selection:

$$N_\ell = N_p + N_f = N_{t0} + N_{t1}$$

$$N_{t1} = N_p + fN_f$$

$$N_{t0} = (1 - f)N_f$$

These relations can be easily inverted to obtain  $N_p$  and  $N_f$ , and mostly the number of non-prompt leptons passing the tight selection, which is the expected number of background events in our final sample:

$$N_f^{\text{pass}} = fN_f = \varepsilon N_{t0}$$

Where we conveniently introduced the notation:

$$\varepsilon = \frac{f}{1-f}$$

This shows that the number of non-prompt leptons passing the tight cuts is expressed as a function of the ones failing the cuts, weighted by  $\varepsilon$ .

This approach can be generalized to our case of dilepton events, in the hypothesis that the probabilities for the two leptons are independent. We will first assume that the leptons have the same flavour, i. e. they are both electrons or muons.

Now the total number of events passing the loose cuts is the sum of three contributions:  $N_{pp}$  events with two prompt leptons,  $N_{fp}$  events with one prompt and one non-prompt lepton,  $N_{ff}$  with two non-prompt leptons. Again, these are related to the measurable numbers of events  $N_{tx}$  with  $x = 0, 1, 2$  leptons passing the tight cuts, the remaining ones failing these cuts.

$$\begin{aligned} N &= N_{ff} + N_{fp} + N_{pp} = N_{t0} + N_{t1} + N_{t2} \\ N_{t0} &= (1-f)^2 N_{ff} \\ N_{t1} &= (1-f)N_{fp} + 2f(1-f)N_{ff} \\ N_{t2} &= fN_{fp} + f^2N_{ff} \end{aligned}$$

The matrix of these linear equations can be inverted, and again with the same procedure illustrated for the single lepton case, we get the number of events with one non-prompt and two non-prompt leptons for our final selection:

$$\begin{aligned} N_{fp}^{\text{pass}} &= \varepsilon N_{t1} - 2\varepsilon^2 N_{t0} \\ N_{ff}^{\text{pass}} &= \varepsilon^2 N_{t0} \end{aligned}$$

As in the single lepton case, failing leptons get a weight  $\varepsilon$ , prompt passing leptons get a weight 1. For the mixed case of  $N_{fp}$ , in case both leptons fail, each of them is weighted alternatively by  $\varepsilon$ , the other by 1, hence the factor 2.

Finally these formulae can be extended to include the situation where an electron and a muon are part of the same configuration. The notation is now  $N_{pp}$ ,  $N_{fp}$ ,  $N_{pf}$  and  $N_{ff}$ , where the first subscript refers to the electron, the second one to the muon. The numbers of events passing and failing the selections is written as  $N_{txy}$ , with  $x$  giving the number of electrons passing the tight cuts and  $y$  giving the number for muons.

The basic equations are then:

$$\begin{aligned} N_{t00} &= (1-f_e)(1-f_\mu)N_{ff} \\ N_{t01} &= (1-f_e)N_{fp} + (1-f_e)f_\mu N_{ff} \\ N_{t10} &= (1-f_\mu)N_{pf} + (1-f_\mu)f_e N_{ff} \\ N_{t11} &= N_{pp} + f_\mu N_{pf} + f_e N_{fp} + f_e f_\mu N_{ff} \end{aligned}$$

From those, the number for the background expected events can be extracted with the now familiar technique:

$$\begin{aligned} N_{pf}^{\text{pass}} &= \varepsilon_\mu N_{t10} - \varepsilon_e \varepsilon_\mu N_{t00} \\ N_{fp}^{\text{pass}} &= \varepsilon_e N_{t01} - \varepsilon_e \varepsilon_\mu N_{t00} \\ N_{ff}^{\text{pass}} &= \varepsilon_e \varepsilon_\mu N_{t00} \end{aligned}$$

sample	event type	observed	expected
W+jets	ee	17	$25.11 \pm 6.60$
	$\mu\mu$	0	$0.68 \pm 0.36$
	e $\mu$	25	$30.61 \pm 8.15$
$t\bar{t}$	ee	43	$70.04 \pm 15.40$
	$\mu\mu$	33	$46.45 \pm 6.48$
	e $\mu$	83	$114.22 \pm 24.53$

Table 11: Summary of the closure tests on Monte Carlo samples for events with two same-sign leptons. Statistical errors only.

### 5.2.5 Validation on test samples

In order to make sure that the method applied for the non-prompt leptonic background estimate is sound we have tested it on independent Monte Carlo samples. The MC generator information provides the real non-prompt contribution to our selection. This number is what we call the *observed* number of events for the purposes of this paragraph.

The non-prompt ratio for Monte Carlo events was calculated in QCD samples, without the selections on the  $E_T^{\text{miss}}$  and  $m_T$ , as there are no vector bosons in the QCD Monte Carlo:

$$f_e = 0.091 \pm 0.017$$

$$f_\mu = 0.152 \pm 0.002$$

These numbers differ from the values in the data since the MC does not describe accurately the composition of the non-prompt events. The main bias is introduced by the filters at the generator level.

Monte Carlo events are selected with two same-sign leptons, discarding events where both leptons are actually prompt, i. e. the decay product of a W or Z boson. Therefore all of the dilepton events we observe after these selections have at least one non-prompt lepton. The method just described in 5.2.4 predicts a number of *expected* events.

The results are summarized in Table 11. The  $t\bar{t}$  studies show that the observed yields differ from the predicted by -39% for ee, -29% for  $\mu\mu$  and -27% for e $\mu$ . Based on these studies, we assign a 50% systematic uncertainty on the estimation of backgrounds due to fake leptons. This is in agreement with the studies done in reference [12].

A comparison between the observed and expected distributions for various kinematical variables (figures 22, 23 and 24) also shows good agreement.

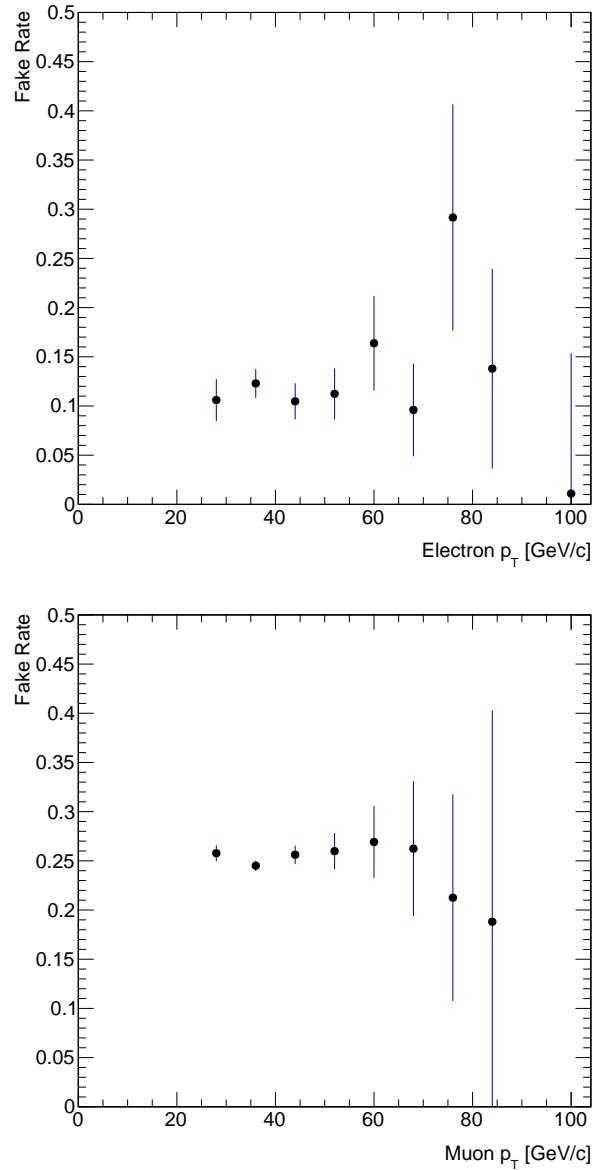


Figure 21: The corrected tight-to-loose ratio, loosely speaking the *fake rate*, as a function of  $p_T$  for electrons (left) and muons (right). The correction is accomplished by subtracting the MC-based  $W$ ,  $Z$  and  $t\bar{t}$  contributions from the data.

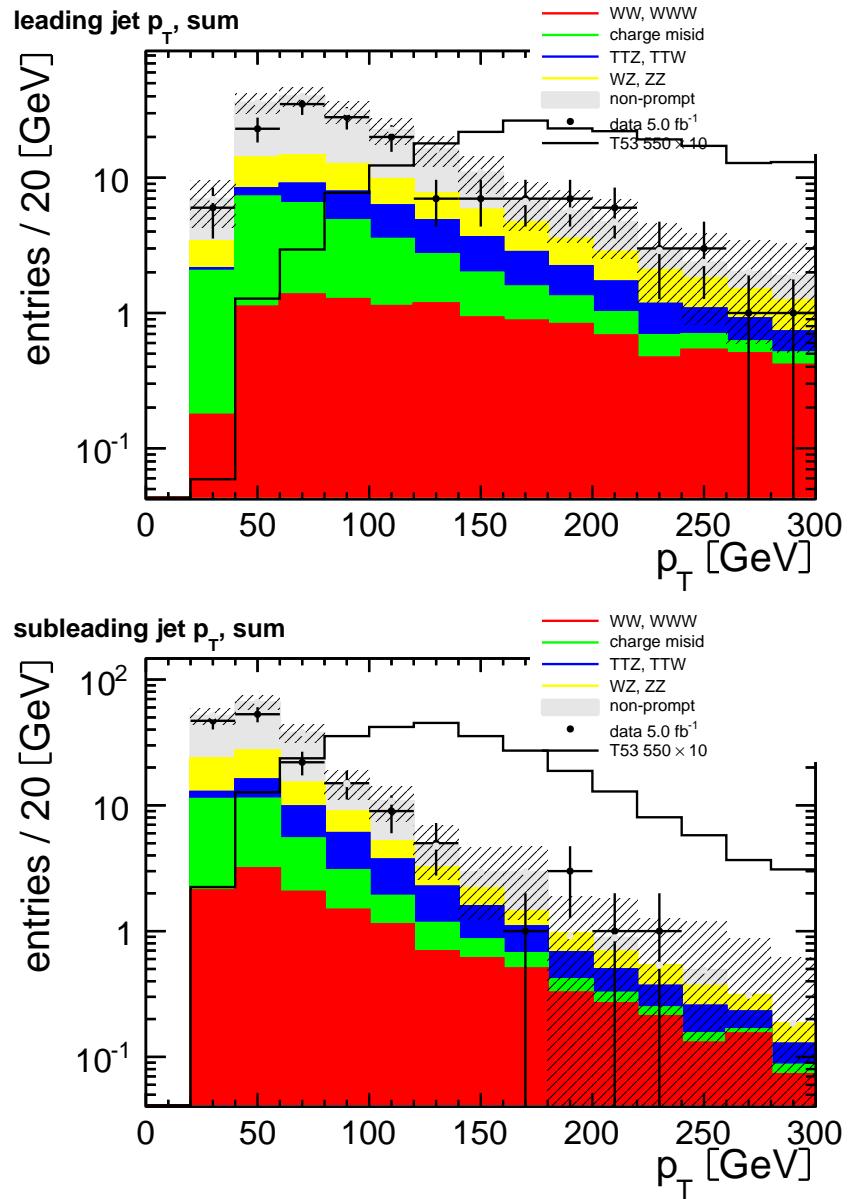


Figure 22: Expected and observed distributions of the  $p_T$  of the leading (top) and subleading (bottom) jet in events with two same-sign leptons and two jets. The sum of the three decay channels is shown.

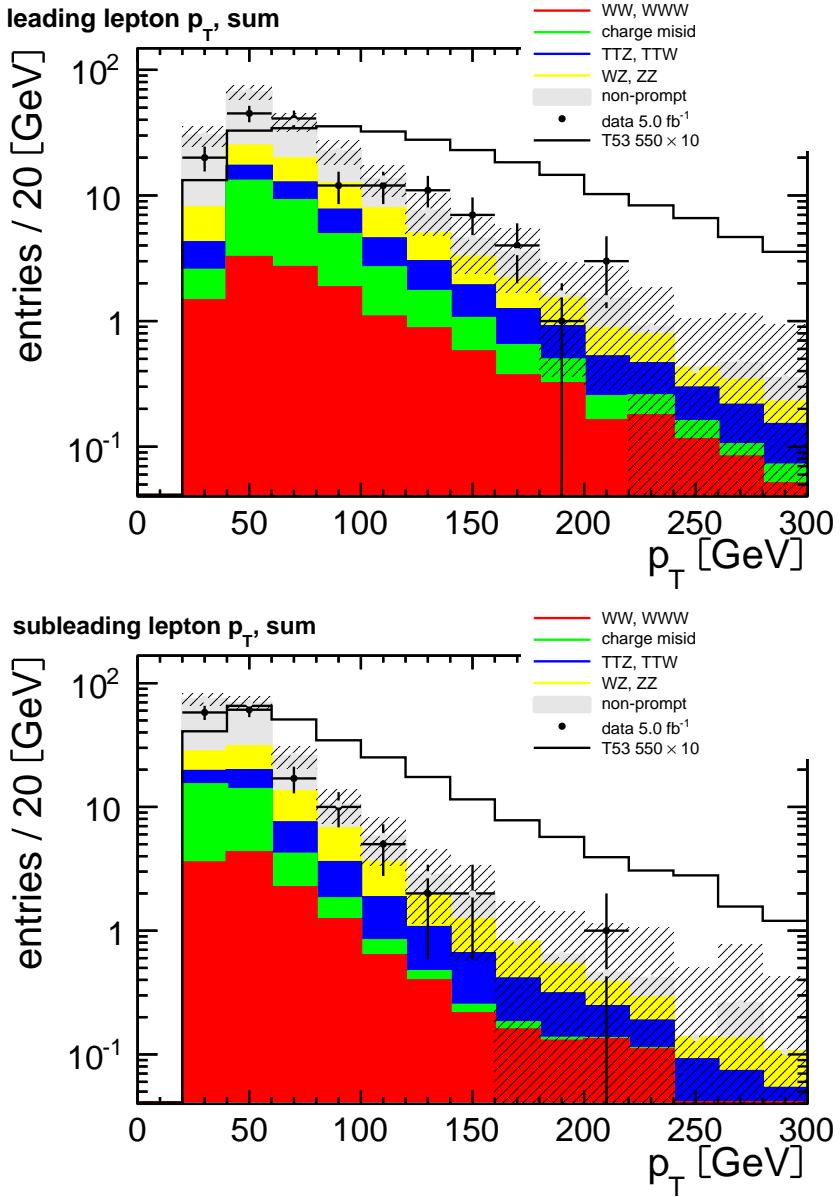


Figure 23: Expected and observed distributions of the  $p_T$  of the leading (top) and subleading (bottom) leptons in events with two same-sign leptons and two jets. The sum of the three decay channels is shown.

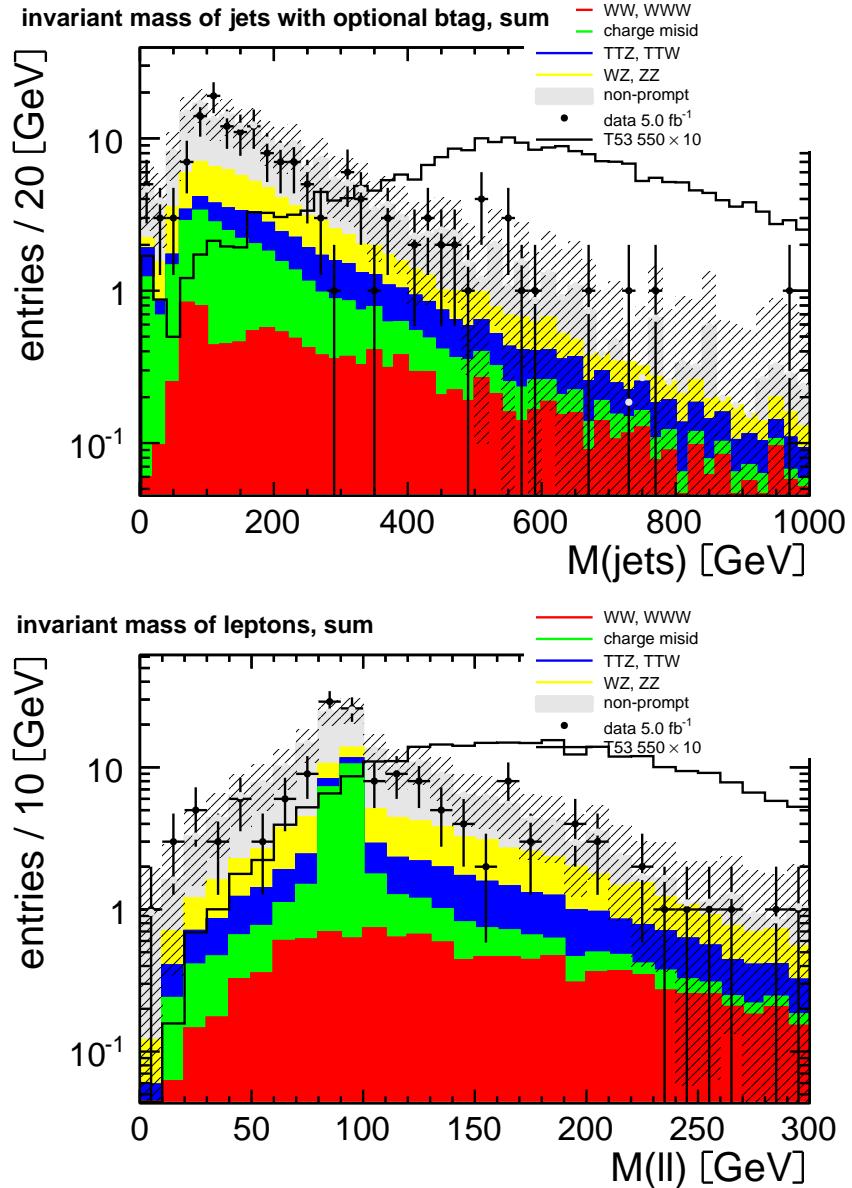


Figure 24: Expected and observed distributions of the invariant mass of jets (top) and leptons (bottom) in events with two same-sign leptons and two jets. The sum of the three decay channels is shown. The *optional b-tag* method for calculating the invariant mass of the jets is described in 6.4.

# 6

## THE RAZOR VARIABLES

The top partners are particles with a large mass. Therefore the energy scale of the events featuring a  $T_{5/3}$  is completely different with respect to Standard Model processes. Many studies have been devoted during the course of the last years to the development of kinematical variables that assist in the discovery of new physics [6, 30, 33, 35? ].

We now introduce the relevant variables, and discuss their application to our signal, which requires new techniques with respect to the standard treatment of these variables.

### 6.1 THE VARIABLE $M_R$

The razor variables were first introduced in the searches for supersymmetric particles [1]. The simplest events in these searches feature the pair production of two heavy supersymmetric particles. Each of these particles decay to a SM particle, which is detected and measured, and a non-interacting *lightest supersymmetric particle* which escapes detection and gives rise to  $E_T^{\text{miss}}$  (figure 25).

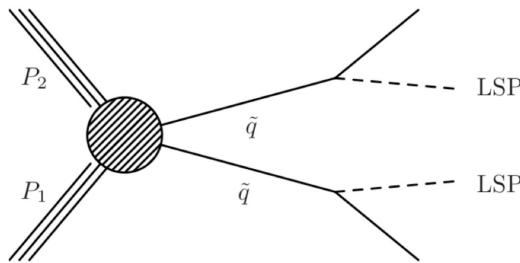


Figure 25: Simple example of a supersymmetric event, with two heavy pair-produced SUSY particles, two invisible lightest supersymmetric particles and two visible SM particles.

In this kind of events, there is an additional, well-motivated approximation that can be made. If the mass of the heavy particle is sufficiently large relative to the collider energy  $\sqrt{s}$ , the particles will be mostly produced near the threshold  $\sqrt{s} \approx 2M$ , where  $\sqrt{s}$  is the usual Mandelstam variable describing the hard partonic process. In this approximation, we can move from the laboratory reference frame to the centre-of-mass frame of the pair-produced particles by finding the frame where the magnitude of the momenta  $\vec{a}$  and  $\vec{b}$  of the visible particles are equal. This frame is called the *R frame*. We define the R-frame mass  $M_{R0}$  as:

$$M_{R0} := \sqrt{(a^0 - b^0)^2 - (a^z + b^z)^2}$$

Where  $a^\mu$  and  $b^\mu$  are the four momenta of the two visible particles, as measured in the laboratory frame. This quantity is invariant under longitudinal boosts, and has a distribution peaked at the value of the massive particle's mass.

It is not invariant under boosts in the transverse plane, but the effects of the  $p_T$  of the pair-produced particles can be taken into account. Introducing the notation  $\vec{v}_T = (v_x, v_y, 0)$  for the transverse components of any vector  $\vec{v}$ ,  $\vec{m}$  for the vector representing the  $E_T^{\text{miss}}$ , the corrections in [36] can be applied:

$$\begin{aligned}\vec{u} &:= \vec{a} + \vec{b} + \vec{m} \\ \vec{v} &:= \vec{a} + \vec{b} \\ M_R^2 &= \frac{1}{2} \left\{ M_{R0}^2 - \vec{u}_T \cdot \vec{v}_T + M_{R0} \sqrt{M_{R0}^2 + \vec{u}_T^2 - 2\vec{u}_T \cdot \vec{v}_T} \right\}\end{aligned}$$

## 6.2 THE TRANSVERSE PLANE AND THE R VARIABLE

Traditionally, the requirements used to improve the signal to background ratio in these processes is large magnitude of the missing transverse energy. However, by looking at the formula for  $M_R$  in the previous section, we see that there is additional kinematical information not yet used. That is, apart from the small correction for the  $p_T$  of the massive particle, the quantities in the transverse plane are not used in the calculation of the  $M_R$  variable.

We then assign half of the measured  $E_T^{\text{miss}}$  to each escaping particle, and define:

$$M_R^T = \sqrt{\frac{|\vec{m}|}{2} (|\vec{a}_T| + |\vec{b}_T|) - \frac{\vec{m}}{2} \cdot (\vec{a}_T + \vec{b}_T)}$$

This variable is an additional measurement of the scale of the process, that uses the information independent of the  $M_R$ . Finally we form the dimensionless R variable as the ratio:

$$R := \frac{M_R^T}{M_R}$$

For signal processes, the distribution of R peaks near  $1/2$  since it is the ratio of two measurements of the same scale, while for background events it is expected to fall after reaching a maximum for a lower value.

The effects of the  $p_T$  of the massive particle are taken into account by using the same formula for  $M_R^T$  but by boosting the vectors  $\vec{a}$ ,  $\vec{b}$  and  $\vec{m}$  by their sum  $\vec{a} + \vec{b} + \vec{m}$ .

## 6.3 THE RAZOR VARIABLES FOR THE TOP PARTNERS

The top partner event topology is different from the mentioned supersymmetric example, as the decay chains of the pair-produced par-

ticles are different: one of them decays into two leptons and one b-jet, while the other one decays into many jets (figure 26). The obvious

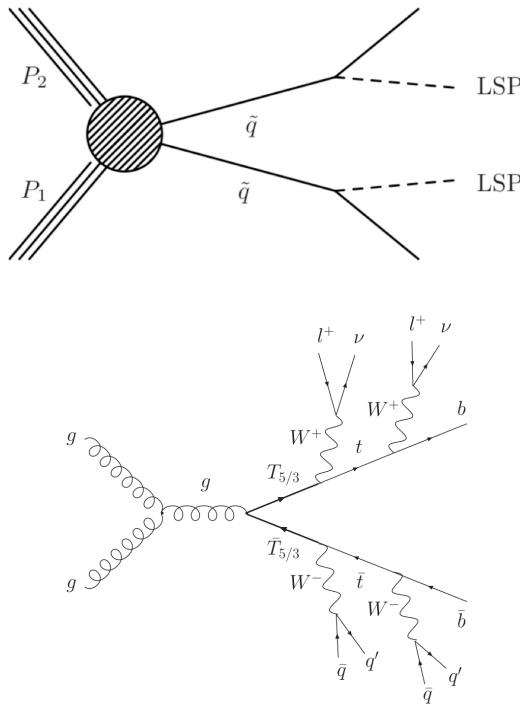


Figure 26: Comparison of a simple supersymmetric example with the top partner decay.

approach would be to try to build two *pseudoparticles* by summing the four momenta of the decay products of each  $T_{5/3}$  in the event. This was done in other SUSY searches for a variety of final states[1]. However, this does not give good results in  $T_{5/3}$  events for a number of reasons:

- there is a combinatorial background given by the fact that one of the jets belongs to one particle, and at least four jets to the other particle
- the  $E_T^{\text{miss}}$  is no longer equally shared by the two decay branches, but should be associated to the branch containing the two leptons
- there is high-quality information from the leptons, which are very well reconstructed at CMS. These kinematical features get smeared by the mixing with the jets, which are more complex and noisier objects

A new approach was sought for solving these problems, and the solution is at the same time simple, elegant and powerful. The idea is to build a *razor subsystem* containing only the two leptons and the two neutrinos. This follows exactly the basic topology outlined in 6.1

and 6.2 for SUSY events, and the  $M_R$  variable is expected to have a peak at half the mass of the  $T_{5/3}$ .

The jets can then be analyzed as a hadronic system. Since this system is, in principle, completely reconstructed, we measure its invariant mass.

The hadronic invariant mass and  $M_R$  in figures 27 and 28 show that the discriminating power of these variables is already very good. The prediction that the  $M_R$  variable peaks at half the mass of the  $T_{5/3}$  is also confirmed. We can also see a reconstruction of the mass of the Z boson in double electron events with charge misidentification, again showing the robustness of this method. On the contrary, the R variable has a very broad distribution.

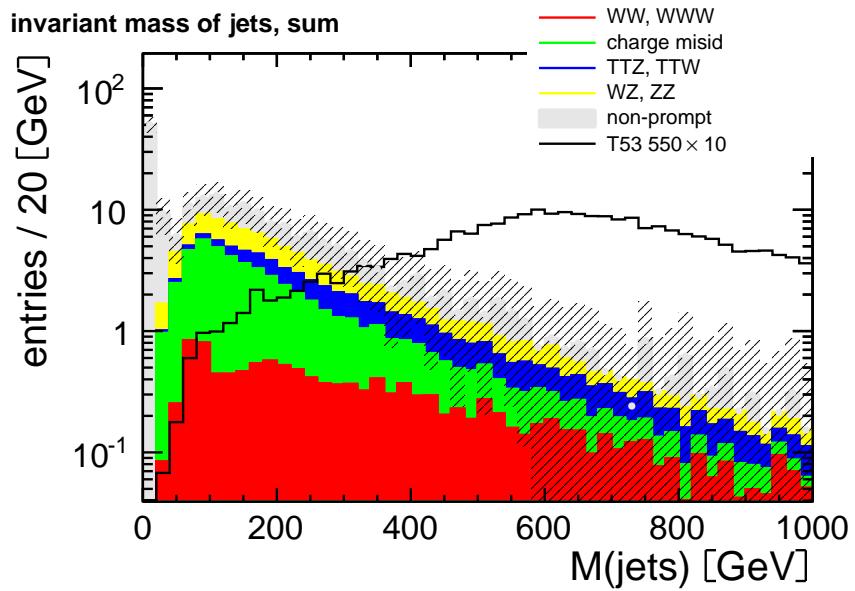


Figure 27: Invariant mass of the hadronic system for the sum of the three decay channels. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ ,  $\mu\mu$  channels are shown in figure 47

#### 6.4 EVENT RECONSTRUCTION IMPROVEMENT AND B-TAGGING

One more approximation remains, that is the association of the b-jet from the leptonic decay chain to the hadronic system. We can try to identify this jet and add it to the leptonic system, where it belongs. From the Monte Carlo information in the signal, we see that this jet is usually close to the lepton from the top decay (figure 30). This simple  $\Delta R$  matching fails for  $2/3$  of the signal events, mainly because there are two leptons in each event and many jets. Therefore, an unrelated jet can get close to a lepton just by chance. Moreover, the b-jet is sometimes not reconstructed because it fails to reach the  $p_T$  threshold

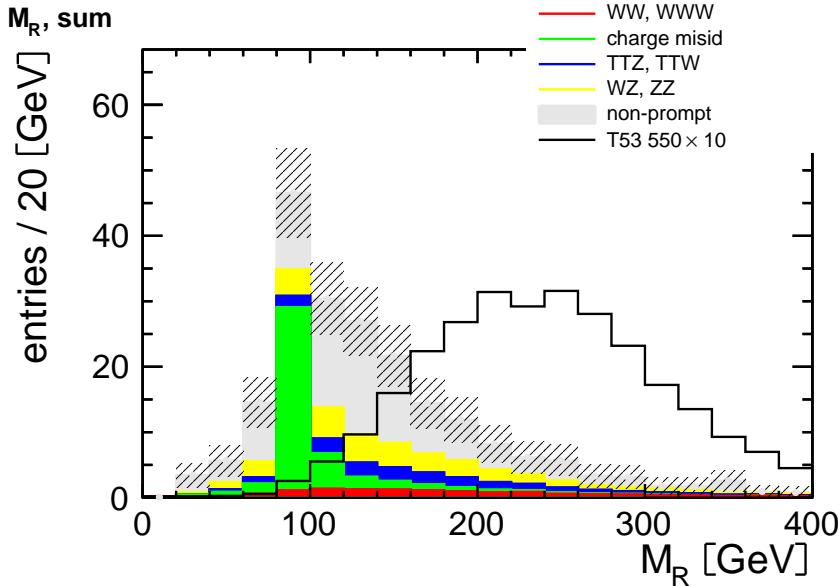


Figure 28:  $M_R$  distributions for the sum of the three decay channels. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ ,  $\mu\mu$  channels are shown in figure 48

(figure 31) or it is too close to the lepton itself, thus failing the  $\Delta R < 0.3$  selection of paragraph 3.3.3.

The event reconstruction can indeed be improved with the use of the jet b-tagging algorithm described in 3.2.5.1. If two b-tagged jets are found in the event, the one with the minimum  $\Delta R$  with respect to one of the leptons is moved from the hadronic system to the leptonic system. The hadronic invariant mass and the razor variables are recalculated accordingly by summing the four momentum of the b-jet to the leptons, and subtracting it from the hadronic system at the same time. If less than two b-tagged jets are found, no association is made. The studies on the signal show that this improves the event reconstruction for one in two events. The distributions of the invariant mass and the razor variables also show an improvement in the background rejection (figures 32, 33 and 34).

## 6.5 FINAL EVENT SELECTION

The razor variables of the last section have to be added to the preselection described in 4. There are three variables that can be used to select candidate  $T_{5/3}$  events: the invariant mass of the jets, the  $M_R$  and R razor variables. The following selections were chosen to optimize the ratio[34]:

$$F := \frac{\varepsilon(t)}{a/2 + \sqrt{B(t)}}$$

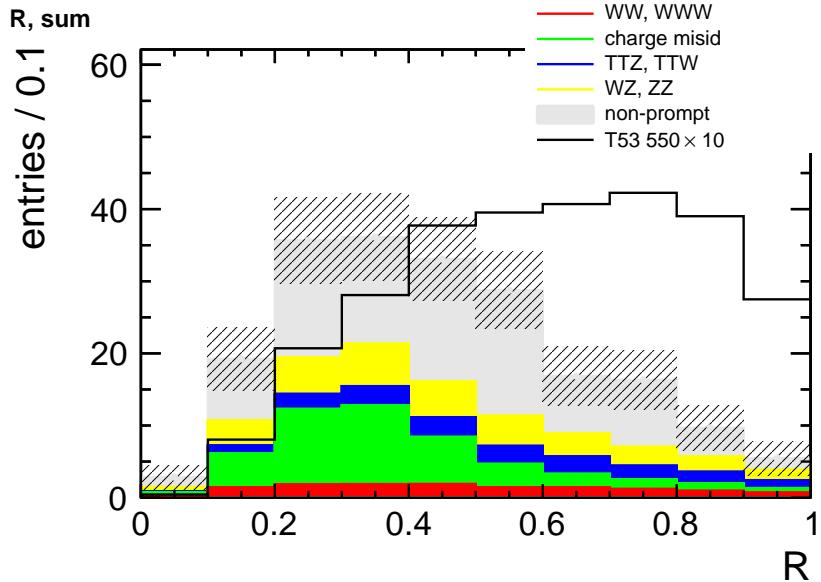


Figure 29:  $R$  distributions for the sum of the three decay channels. Events with two same-sign leptons and at least two jets are shown. The plots for the  $ee$ ,  $e\mu$ ,  $\mu\mu$  channels are shown in figure 49

Where  $\varepsilon$  is the efficiency of the selection  $t$  for the signal,  $B$  is the number of expected background events, and  $\alpha$  is the number of sigmas corresponding to a one-sided gaussian test at significance  $\alpha = 95\%$ . The commonly used  $S/\sqrt{B}$  figure is unsuitable for the present search, with a very low background, as the expression breaks down for  $B \rightarrow 0$ . The ratio  $F$  is plotted in figures 35, 36 and 37.

The final event selection includes:

- the event preselection from 4, with at least four jets and invariant mass of the leptons not in the quarkonia and  $Z$  boson windows.
- invariant mass of the hadronic system greater than  $350\text{ GeV}$
- razor  $M_R > 200\text{ GeV}$
- razor  $R > 0.2$

Figures 39, 40 and 38 show the distributions of the invariant mass of the jets,  $M_R$  and  $R$  with all the cuts applied, but the one on the plotted variable.

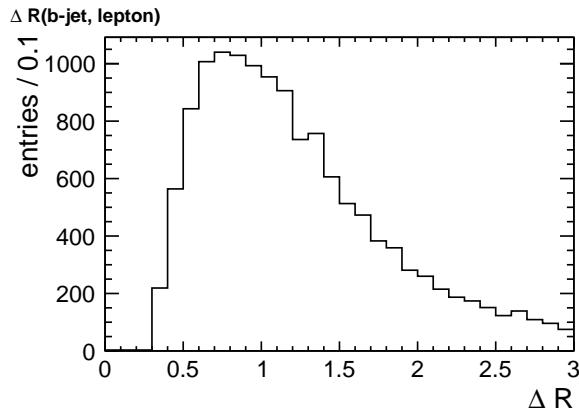


Figure 30:  $\Delta R$  between the b-jet and the lepton from the top decay. MC truth from the 550 GeV mass point.

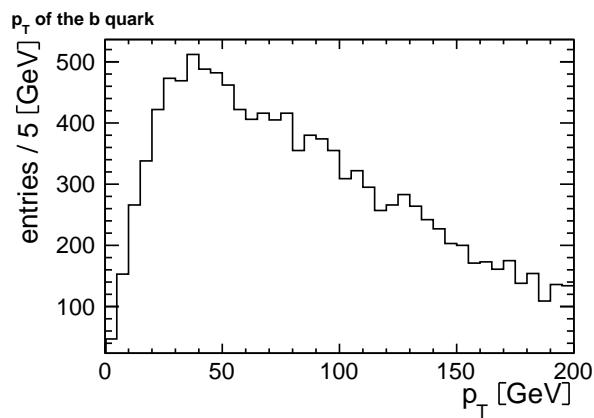


Figure 31: Distribution of the  $p_T$  of the b quark in the  $T_{5/3}$  decay chain. If the quark has a low  $p_T$ , it will most likely form a low- $p_T$  jet, which would not be collected by our selections. MC truth from the 550 GeV mass point.

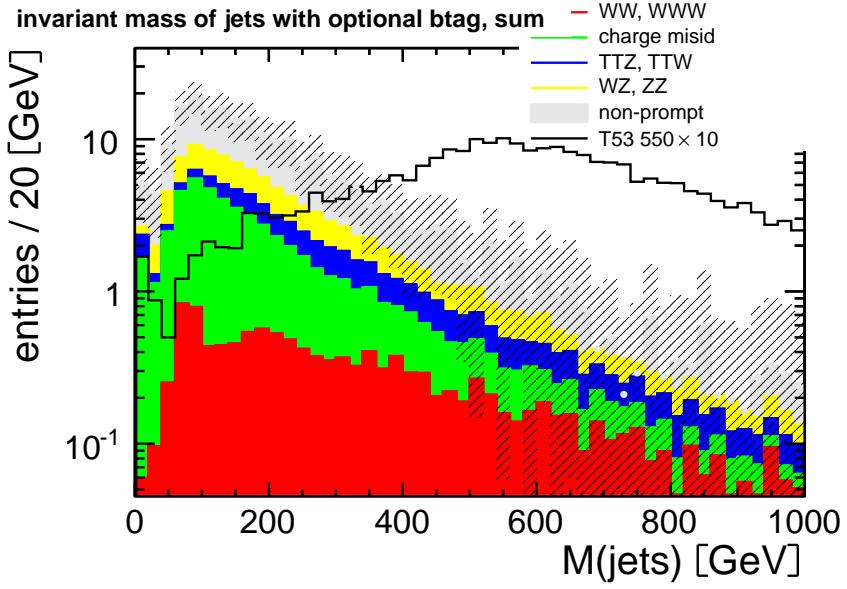


Figure 32: Invariant mass of the hadronic system for the sum of the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. The plots for the ee,  $e\mu$ ,  $\mu\mu$  channels are shown in figure 50

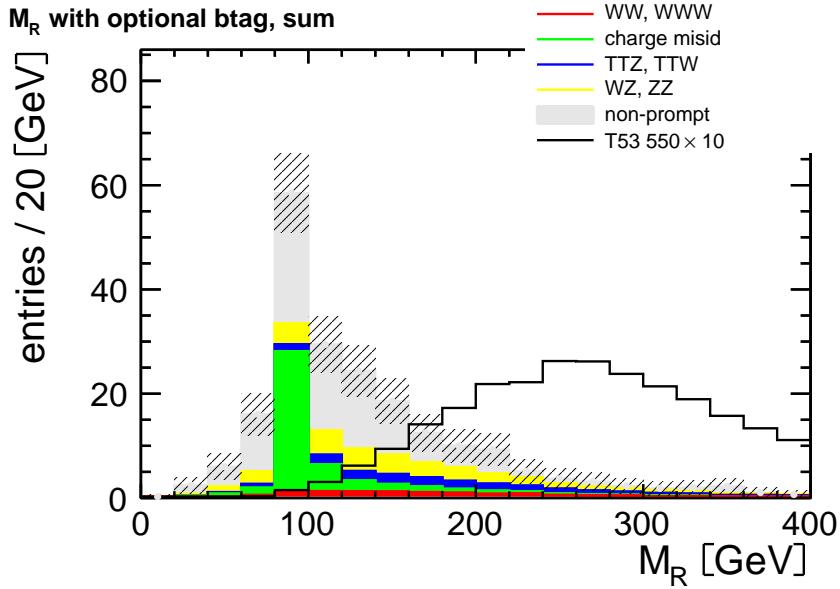


Figure 33:  $M_R$  distributions for the sum of the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. The plots for the ee,  $e\mu$ ,  $\mu\mu$  channels are shown in figure 51

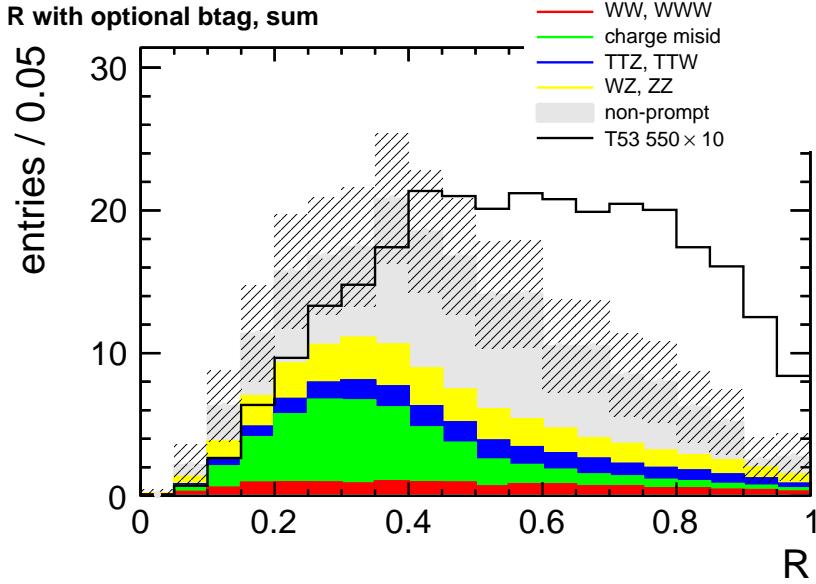


Figure 34:  $R$  distributions for the sum of the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown. The plots for the ee, e $\mu$ ,  $\mu\mu$  channels are shown in figure 52

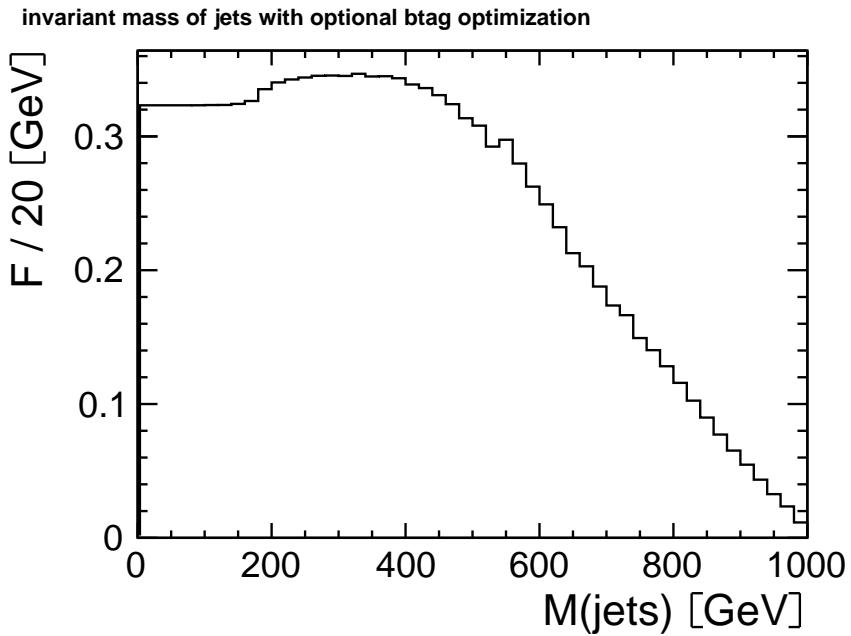


Figure 35: Optimization figure  $F$  for the invariant mass of the hadronic system. The selection on  $M_R > 200$  GeV and  $R > 0.2$  are applied. The point where  $F$  reaches a maximum is chosen as a cut on this variable.

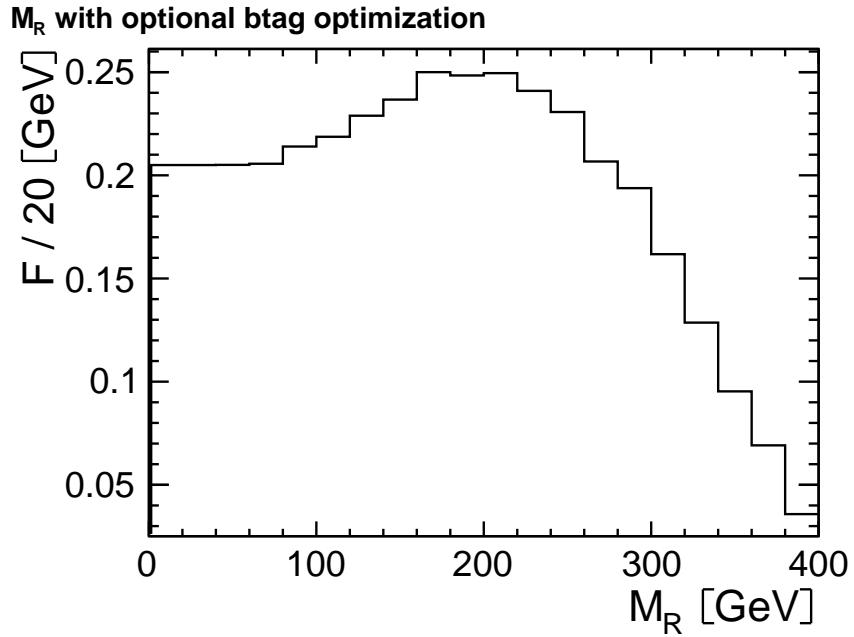


Figure 36: Optimization figure  $F$  for the  $M_R$  variable. The selection on  $M > 350$  GeV and  $R > 0.2$  are applied. The point where  $F$  reaches a maximum is chosen as a cut on this variable.

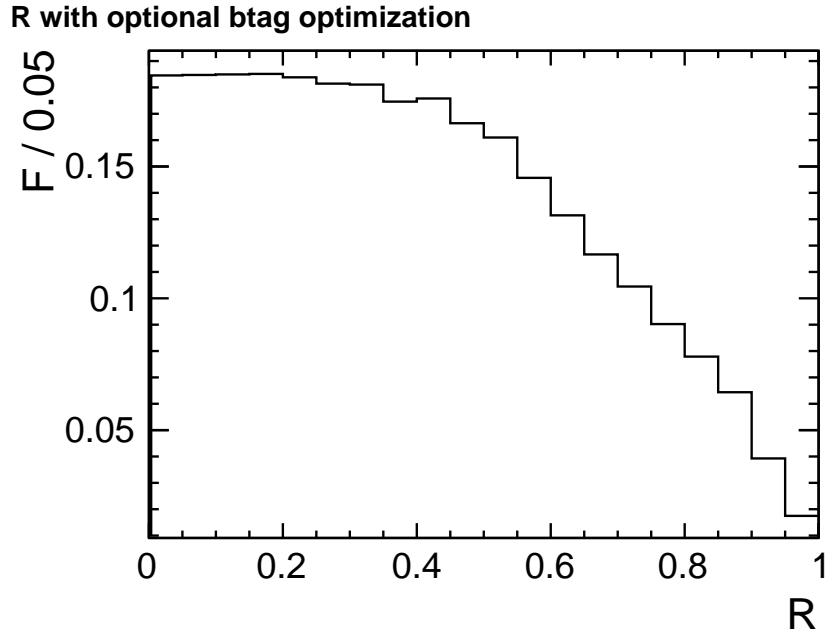


Figure 37: Optimization figure  $F$  for the  $R$  variable. The selection on  $M > 350$  GeV and  $M_R > 200$  GeV are applied. The point where  $F$  reaches a maximum is chosen as a cut on this variable.

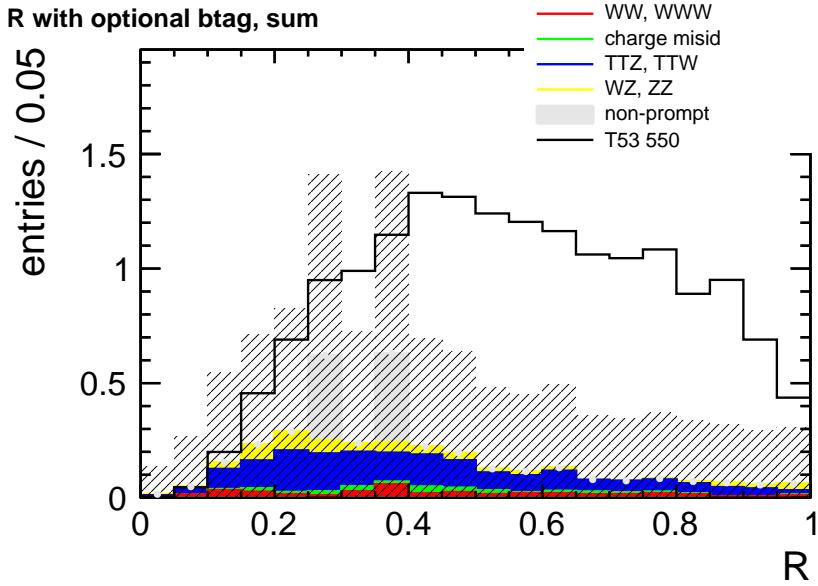


Figure 38: Distribution of the  $R$  variable for events in the three decay channels, with the final selection applied without the selection on  $R$ .

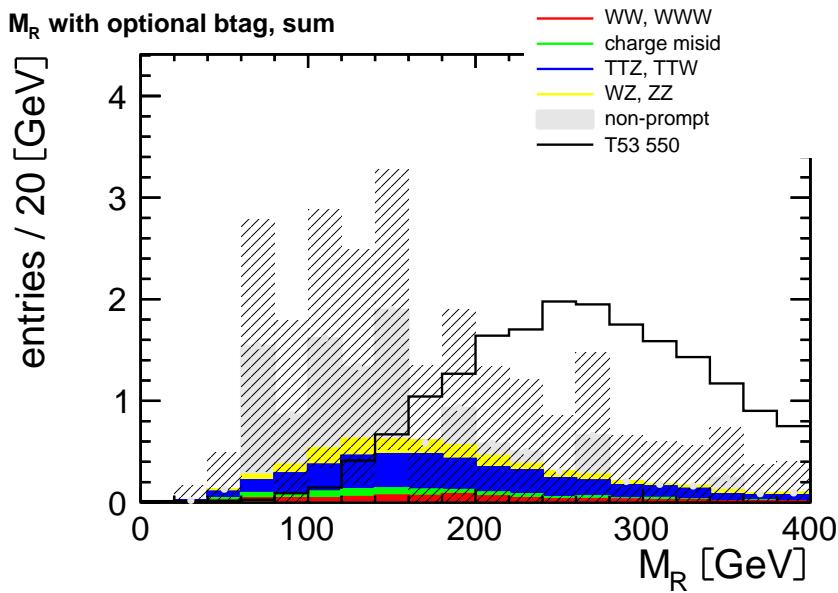


Figure 39: Distribution of the  $M_R$  variable for events in the three decay channels, with the final selection applied without the selection on  $M_R$ .

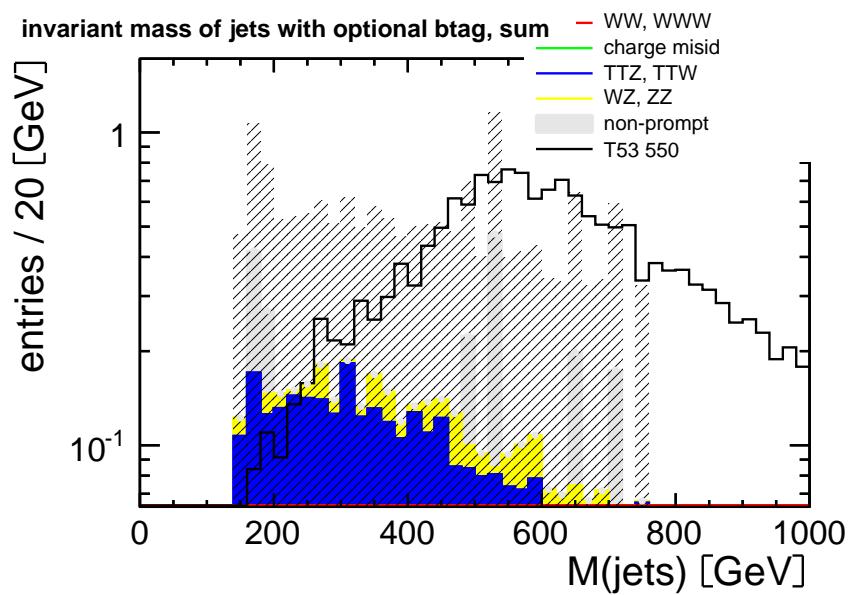


Figure 40: Distribution of the invariant mass of the jets for events in the three decay channels, with the final selection applied without the selection on  $M(\text{jets})$ .

## SYSTEMATIC UNCERTAINTIES

---

Systematic uncertainties need to be taken into account because of the imperfect knowledge of either detector effects or theory. Detector effects include uncertainties in the trigger efficiencies, the efficiency of the lepton reconstruction, the jet energy correction factors and pile-up conditions.

Lepton reconstruction efficiencies were studied in [22]. The uncertainty on the trigger efficiency is taken to be 5%. This is a conservative estimate based on the study of the dilepton triggers in [13]. We also include a 2.2% uncertainty on the luminosity [23]. These systematic uncertainties are summarized in Table 12.

### 7.1 MONTE CARLO SYSTEMATICS

The jet energy scale (JES) and pile-up uncertainties are calculated by varying the JES and pile-up according to the recommended recipes in the samples for which we take the result from Monte Carlo. The JES and pile-up uncertainties for the backgrounds are summarized in Table 13. For the signal points, we take the JES and pile-up uncertainties to be the maximum value of these for single point. This corresponds to 0.9% for the JES and 2% for the pile-up. Table 13 also contains the overall normalization uncertainty for each background sample.

The WZ and ZZ normalization uncertainties are taken from [2]. For the rare backgrounds, we assume a normalization uncertainty of 50%, as these processes have never been observed at the LHC or have not been measured well.

### 7.2 DATA-DRIVEN BACKGROUND SYSTEMATICS

We also include a 50% uncertainty associated with the estimation of the non-prompt lepton background 5.2.5 and a 20% uncertainty for

effect	uncertainty (%)
electron trigger	5.0
muon trigger	5.0
lepton efficiency	3.0
luminosity	2.2

Table 12: Systematic uncertainties.

sample	jet energy	pile-up	normalization
WZ	5.0%	1.8%	17%
ZZ	1.1%	0.6%	7.5%
$W^\pm W^\pm$	4.5%	2.4%	50%
WWW	3.7%	0.5%	50%
$t\bar{t}W$	3.4%	0.94%	50%
$t\bar{t}Z$	3.7%	0.25%	50%

Table 13: Systematic uncertainties for backgrounds that are taken from Monte Carlo.

the charge misidentification. This comes from the difference between the expected number of events derived with the method described in 5.1 and a more sophisticated  $\eta$  dependent estimate. We can see that the charge misidentification probability depends on  $\eta$  (figure 41), and is lower in the barrel with respect to the endcaps. However, given the

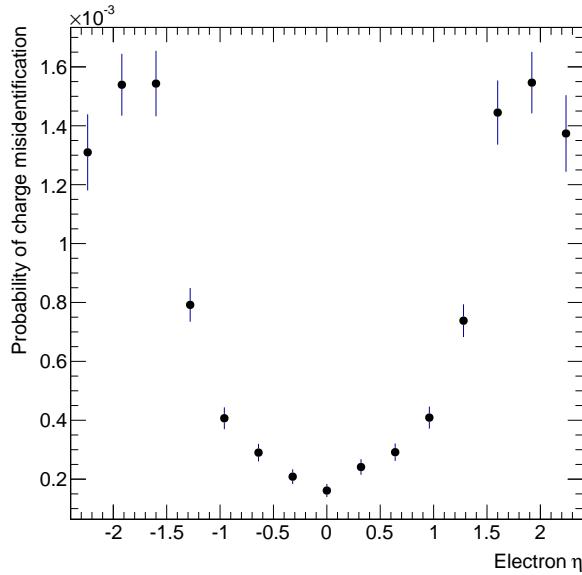


Figure 41: Charge misidentification probability as a function of the pseudo-rapidity  $\eta$ .

very small amount of events with charge misidentification passing our selections, this contribution to the total uncertainty is small as well. These uncertainties are summarized in table 14.

The inputs to the calculation of the limits on the  $T_{5/3}$  mass affected by the systematic uncertainties are the luminosity, the uncertainty on the signal efficiency and the uncertainty on the total background. The latter is the largest of the three and it is completely dominated by the 50% uncertainty associated with the estimation of the non-

data-driven method	uncertainty
charge misid	20%
non-prompt	50%

Table 14: Systematic uncertainties for data-driven backgrounds

prompt lepton background and the normalization uncertainty on the rare backgrounds.



## RESULTS

In this chapter we will extract the result from the comparisons of the event collected in the 2011 datasets with the background estimates described in the previous chapters. The events passing the kinematical selection are compared with expectations from Standard Model background sources. The number of expected events from the signal Monte Carlo is shown in table 15. The backgrounds are compared to the observed event yields in table 16.

No significant excess over the SM expectations is observed (also see figures 42 and 43), and we can set lower bounds on the masses of the top partners.

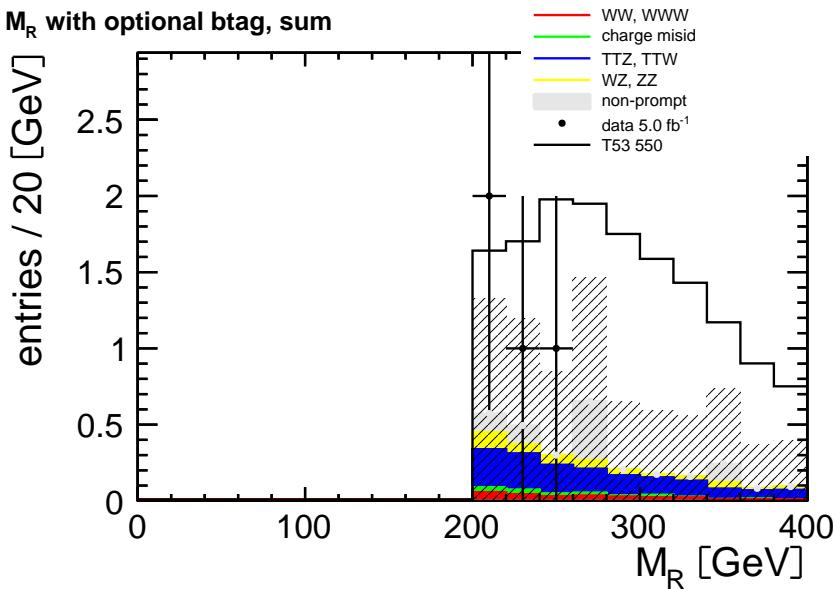


Figure 42: Distribution of the  $M_R$  variable for events in the three decay channels, with the final selection applied.

Exclusion limits are computed at the 95% confidence level for an event counting experiment [31].

### 8.1 CALCULATION OF THE OBSERVED LIMIT

A modified frequentist method is used at the LHC experiments [8], known as the  $CL_s$  method. The test statistic and the treatment of nuisance parameters are described in this section.

The expected  $T_{5/3}$  event yields are denoted by  $s$ , the backgrounds as  $b$ . We also introduce the usual *signal strength modifier*  $\mu$ . The system-

dataset	same-sign, 2 jets	preselection	razor
T <sub>53</sub> 400	<b>211.10</b> ±10.73	<b>151.16</b> ±7.73	<b>54.34</b> ±2.88
T <sub>53</sub> 450	<b>104.52</b> ±5.31	<b>77.93</b> ±3.98	<b>41.34</b> ±2.15
T <sub>53</sub> 500	<b>52.76</b> ±2.68	<b>40.09</b> ±2.04	<b>25.61</b> ±1.32
T <sub>53</sub> 550	<b>28.53</b> ±1.45	<b>22.47</b> ±1.14	<b>16.22</b> ±0.83
T <sub>53</sub> 600	<b>15.34</b> ±0.78	<b>12.07</b> ±0.61	<b>9.39</b> ±0.48
T <sub>53</sub> 650	<b>8.65</b> ±0.44	<b>6.87</b> ±0.35	<b>5.57</b> ±0.28
T <sub>53</sub> 700	<b>5.00</b> ±0.25	<b>3.98</b> ±0.20	<b>3.37</b> ±0.17
T <sub>53</sub> 750	<b>2.97</b> ±0.15	<b>2.35</b> ±0.12	<b>2.01</b> ±0.10

Table 15: Signal Monte Carlo yields for the various mass points in the three decay channels. Statistical and systematic uncertainties are shown.

dataset	same-sign, 2 jets	preselection	razor
WW, WWW	<b>13.28</b> ±6.67	<b>0.88</b> ±0.45	<b>0.28</b> ±0.14
TTZ, TTW	<b>18.57</b> ±9.31	<b>6.39</b> ±3.20	<b>1.41</b> ±0.71
WZ, ZZ	<b>34.62</b> ±6.17	<b>1.97</b> ±0.36	<b>0.42</b> ±0.08
total MC	<b>66.47</b> ±13.01	<b>9.25</b> ±3.25	<b>2.12</b> ±0.73
charge misid	<b>24.05</b> ±4.81	<b>0.83</b> ±0.17	<b>0.15</b> ±0.03
non-prompt	<b>105.24</b> ±52.80	<b>11.30</b> ±5.86	<b>0.76</b> ±0.56
observed	<b>157</b>	<b>15</b>	<b>4</b>

Table 16: Expected backgrounds and observed yields for different cuts in the three decay channels. Statistical and systematic uncertainties are shown.

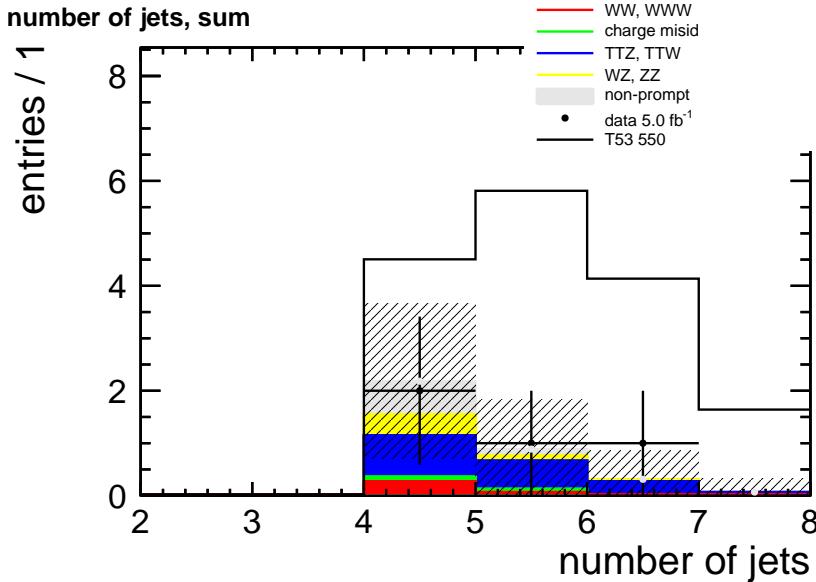


Figure 43: Distribution of the number of jets for events in the three decay channels, with the final selection applied.

atic uncertainties described in  $\gamma$  are introduced as nuisance parameters (denoted together as  $\theta$ ) for the number of expected signal and background events. They become then functions of these parameters  $s(\theta)$  and  $b(\theta)$ .

The systematic error probability density functions (*pdfs*)  $p(\theta|\tilde{\theta})$  are taken as gaussians. The mean of the gaussian  $\tilde{\theta}$  is the default value of the nuisance parameter. We then re-interpret these systematic errors *pdfs* as posteriors given by imaginary measurements  $\tilde{\theta}$ , as in Bayes' theorem (with flat  $\pi_\theta(\theta)$  priors):

$$p(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta)\pi_\theta(\theta)$$

This point of view allows to represent all statistical errors in a frequentist context: the  $p(\tilde{\theta}|\theta)$  can be used in a sampling fashion to construct the test statistics, and to constrain the likelihood of the main measurement.

The procedure for the calculation of the observed limit then includes the following steps:

1. We construct the likelihood function with a Poisson distribution:

$$\mathcal{L}(\text{data}|\mu, \theta) = \frac{(\mu s + b)^n}{n!} e^{-\mu s - b} p(\tilde{\theta}|\theta)$$

where *data* represents the actual observed event yields, or pseudo-experiments constructed by sampling distributions as discussed below.

2. In order to test the compatibility of the data with the background only and signal+background hypotheses we build a test statistic  $\tilde{q}_\mu$  based on the profile likelihood ratio:

$$\tilde{q}_\mu = -2 \log \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}$$

With the constraint that  $0 \leq \hat{\mu} \leq \mu$ . The  $0 \leq \hat{\mu}$  is obvious, as the signal rate must be positive, while the upper constraint means that fluctuations of the data such that  $\hat{\mu} \geq \mu$  are not considered as evidence against the signal hypothesis. The value of  $\hat{\theta}_\mu$  is the conditional maximum likelihood estimator of  $\theta$ , given  $\mu$ , while the values  $\hat{\mu}$  and  $\hat{\theta}$  give the global maximum of the likelihood.

3. The observed value  $\tilde{q}_\mu^{\text{obs}}$  for the  $\mu$  under test is calculated.
4. The values of the nuisance parameters best describing the observed data, i. e. maximizing the likelihood, are computed for the background only and signal+background hypotheses. We denote these values as  $\hat{\theta}_0^{\text{obs}}$  and  $\hat{\theta}_\mu^{\text{obs}}$  respectively.
5. Monte Carlo pseudo-experiments are made to construct *pdfs*  $f(\tilde{q}_\mu|\hat{\theta}_\mu^{\text{obs}})$  and  $f(\tilde{q}_\mu|\hat{\theta}_0^{\text{obs}})$ . Note that the nuisance parameters are fixed for the purposes of generating these pseudo-datasets, but are allowed to float in fits needed to evaluate the test statistic. This gives good coverage properties [25].
6. We calculate two p-values for the b-only and s + b hypotheses:

$$\begin{aligned} 1 - p_b &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | \text{b-only}) &= \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\hat{\theta}_0^{\text{obs}}) d\tilde{q}_\mu \\ p_\mu &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | s + b) &= \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\hat{\theta}_\mu^{\text{obs}}) d\tilde{q}_\mu \end{aligned}$$

and  $CL_s$  as the ratio of these two probabilities:

$$CL_s = \frac{p_\mu}{1 - p_b}$$

7. If for  $\mu = 1$ ,  $CL_s \leq 5\%$ , we state that the  $T_{5/3}$  is excluded with 95% confidence level. Limits calculated with this  $CL_s$  method are known to be conservative, i. e. the confidence level is actually higher than 95%.

## 8.2 CALCULATION OF THE EXPECTED LIMIT

In order to calculate the expected limit and the bands corresponding to 1 and  $2\sigma$  fluctuations, we generate many background-only pseudo-experiments and calculate  $CL_s$  as in the observed limit procedure.

Then, a cumulative probability distribution of results can be plotted, and the point at which the distribution crosses the quantile of 50% is the median expected value, to be taken as the expected limit. The  $\pm 1\sigma$  (68%) is defined by the crossings of the 16% and 84% quantiles. Crossings at the 2.5% and 97.5% define the  $\pm 2\sigma$  band.

### 8.3 RESULTS

The event yields from the three decay channels are combined when setting the limits. The expected and observed limits are shown in figure 44. The expected limit is 658 GeV, the observed limit is 633 GeV.

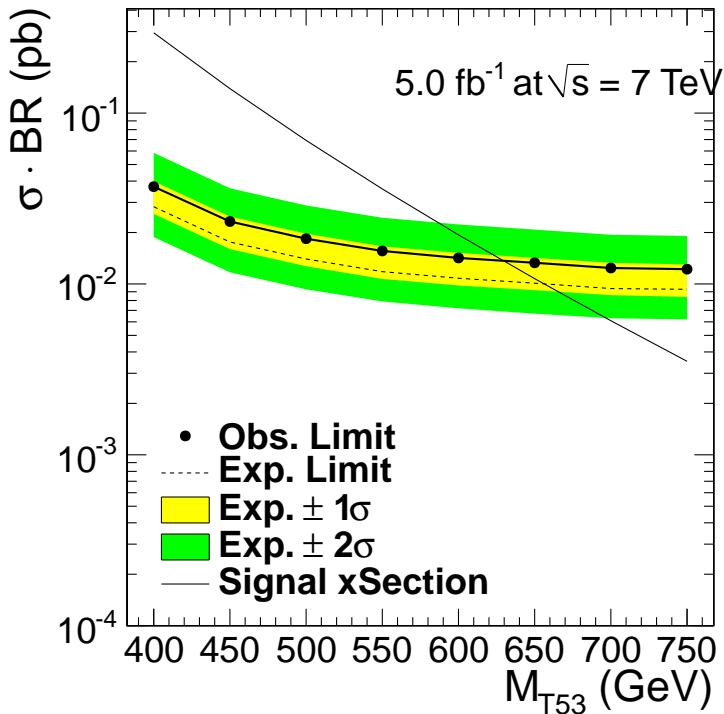


Figure 44: Observed and expected limits for the production cross sections of the top partners.



## CONCLUSIONS

---

The search for new physics at the LHC has entered a very challenging phase: the large amount and quality of the collected data allow to search for many signatures of physics beyond the Standard Model.

The techniques developed in this thesis for the analysis of signal events with the razor variables have been proven to be effective. The kinematical properties of these variables are very general and can be applied to a wide range of searches for new physics. It is the first time the razor variables are shown to be an accurate estimate for the energy scale of a process by considering only part of the decay chain, while previous searches only included them as global, event-by-event quantities.

This is particularly relevant for the analysis of the large amount of data the CMS experiment has recorded in 2012, amounting to  $15\text{ fb}^{-1}$ . A more detailed event description could possibly make the difference while looking for minute deviations from Standard Model expectations.

The analysis presented in this work is a search for heavy partners of the top quark in same-sign dilepton events. The data collected in 2011 by the CMS experiment at a centre-of-mass energy of 7 TeV are analysed. A counting experiment is established by selecting events in three decay channels: ee, e $\mu$  and  $\mu\mu$ . The Standard Model backgrounds are estimated with different Monte Carlo and data-driven techniques.

The observed yields agree with the Standard Model prediction. Hence, no evidence for new physics is found, and lower limits on the mass of the top partner are set.



# A

## APPENDIX

---

dataset	same-sign, 2 jets	preselection	razor
T53 400	$44.64 \pm 10.73$	$29.13 \pm 7.73$	$11.16 \pm 2.88$
T53 450	$22.63 \pm 5.31$	$15.89 \pm 3.98$	$8.69 \pm 2.15$
T53 500	$11.39 \pm 2.68$	$8.09 \pm 2.04$	$5.37 \pm 1.32$
T53 550	$6.16 \pm 1.45$	$4.67 \pm 1.14$	$3.43 \pm 0.83$
T53 600	$3.45 \pm 0.78$	$2.62 \pm 0.61$	$2.06 \pm 0.48$
T53 650	$1.82 \pm 0.44$	$1.39 \pm 0.35$	$1.14 \pm 0.28$
T53 700	$1.09 \pm 0.25$	$0.85 \pm 0.20$	$0.73 \pm 0.17$
T53 750	$0.67 \pm 0.15$	$0.53 \pm 0.12$	$0.45 \pm 0.10$

Table 17: Signal Monte Carlo yields for the various mass points in the ee channel. Statistical and systematic uncertainties are shown.

dataset	same-sign, 2 jets	preselection	razor
T53 400	$106.55 \pm 10.73$	$83.71 \pm 7.73$	$28.84 \pm 2.88$
T53 450	$52.72 \pm 5.31$	$42.26 \pm 3.98$	$21.87 \pm 2.15$
T53 500	$26.45 \pm 2.68$	$21.41 \pm 2.04$	$13.45 \pm 1.32$
T53 550	$14.40 \pm 1.45$	$11.95 \pm 1.14$	$8.52 \pm 0.83$
T53 600	$7.61 \pm 0.78$	$6.21 \pm 0.61$	$4.80 \pm 0.48$
T53 650	$4.31 \pm 0.44$	$3.52 \pm 0.35$	$2.82 \pm 0.28$
T53 700	$2.49 \pm 0.25$	$2.03 \pm 0.20$	$1.70 \pm 0.17$
T53 750	$1.47 \pm 0.15$	$1.18 \pm 0.12$	$1.00 \pm 0.10$

Table 18: Signal Monte Carlo yields for the various mass points in the e $\mu$  channel. Statistical and systematic uncertainties are shown.

dataset	same-sign, 2 jets	preselection	razor
T <sub>53</sub> 400	59.91±10.73	38.32±7.73	14.35±2.88
T <sub>53</sub> 450	29.17±5.31	19.77±3.98	10.78±2.15
T <sub>53</sub> 500	14.91±2.68	10.59±2.04	6.79±1.32
T <sub>53</sub> 550	7.98±1.45	5.85±1.14	4.26±0.83
T <sub>53</sub> 600	4.29±0.78	3.24±0.61	2.53±0.48
T <sub>53</sub> 650	2.52±0.44	1.95±0.35	1.61±0.28
T <sub>53</sub> 700	1.42±0.25	1.10±0.20	0.94±0.17
T <sub>53</sub> 750	0.83±0.15	0.65±0.12	0.56±0.10

Table 19: Signal Monte Carlo yields for the various mass points in the  $\mu\mu$  channel. Statistical and systematic uncertainties are shown.

dataset	same-sign, 2 jets	preselection	razor
WW, WWW	2.86±6.67	0.22±0.45	0.05±0.14
TTZ, TTW	3.97±9.31	1.33±3.20	0.31±0.71
WZ, ZZ	8.57±6.17	0.41±0.36	0.09±0.08
total MC	15.40±13.01	1.95±3.25	0.45±0.73
charge misid	18.38±4.81	0.22±0.17	0.04±0.03
non-prompt	34.99±1.91	2.27±0.48	0.12±0.11
observed	77	3	2

Table 20: Expected backgrounds and observed yields for different cuts in the  $ee$  channel. Statistical and systematic uncertainties are shown.

dataset	same-sign, 2 jets	preselection	razor
WW, WWW	6.63±6.67	0.48±0.45	0.16±0.14
TTZ, TTW	9.13±9.31	3.39±3.20	0.71±0.71
WZ, ZZ	17.45±6.17	1.12±0.36	0.19±0.08
total MC	33.21±13.01	4.99±3.25	1.07±0.73
charge misid	4.66±4.81	0.60±0.17	0.11±0.03
non-prompt	50.61±3.41	5.44±1.20	0.25±0.28
observed	48	8	2

Table 21: Expected backgrounds and observed yields for different cuts in the  $e\mu$  channel. Statistical and systematic uncertainties are shown.

dataset	same-sign, 2 jets	preselection	razor
WW, WWW	$3.78 \pm 6.67$	$0.18 \pm 0.45$	$0.07 \pm 0.14$
TTZ, TTW	$5.47 \pm 9.31$	$1.68 \pm 3.20$	$0.39 \pm 0.71$
WZ, ZZ	$8.60 \pm 6.17$	$0.45 \pm 0.36$	$0.14 \pm 0.08$
total MC	$17.86 \pm 13.01$	$2.31 \pm 3.25$	$0.60 \pm 0.73$
charge misid	$1.01 \pm 4.81$	$0.01 \pm 0.17$	$0.00 \pm 0.03$
non-prompt	$19.64 \pm 2.06$	$3.59 \pm 0.89$	$0.39 \pm 0.28$
observed	32	4	0

Table 22: Expected backgrounds and observed yields for different cuts in the  $\mu\mu$  channel. Statistical and systematic uncertainties are shown.

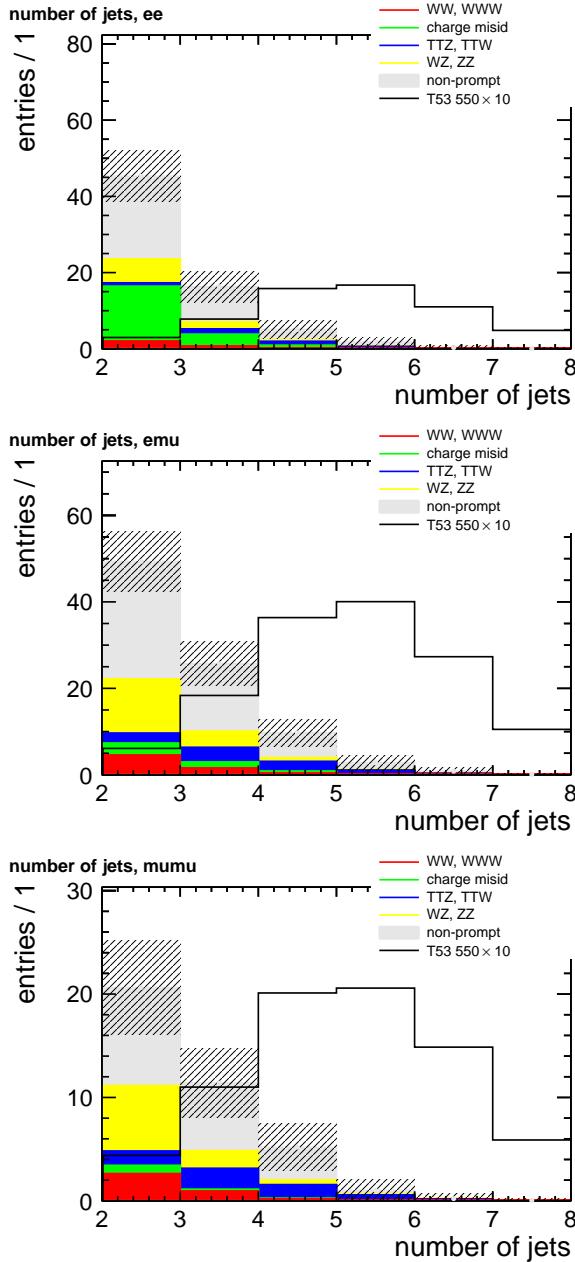


Figure 45: Events with two same-sign leptons and two jets. Distribution of the number of jets in the backgrounds and signal for the 550 GeV mass point for the three decay channels ee, e $\mu$  and  $\mu\mu$ . The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties.

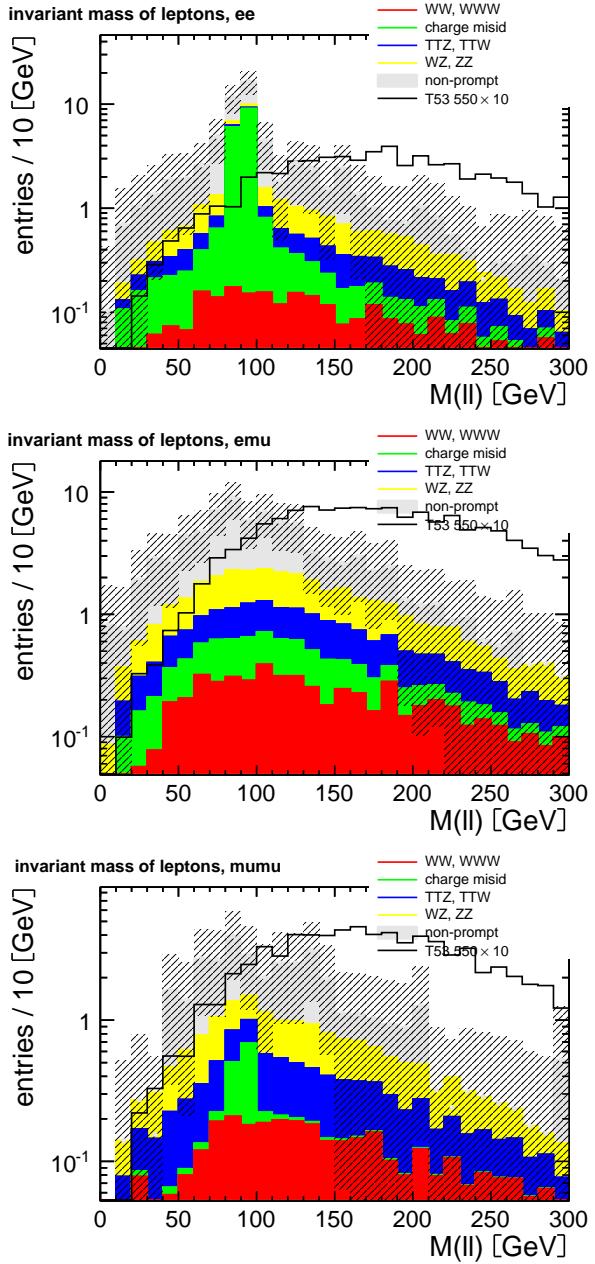


Figure 46: Events with two same-sign leptons and two jets. Distribution of the invariant mass of the leptons in the backgrounds and signal for the 550 GeV mass point for the three decay channels ee, e $\mu$  and  $\mu\mu$ . The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The shaded area includes statistical and systematic uncertainties.

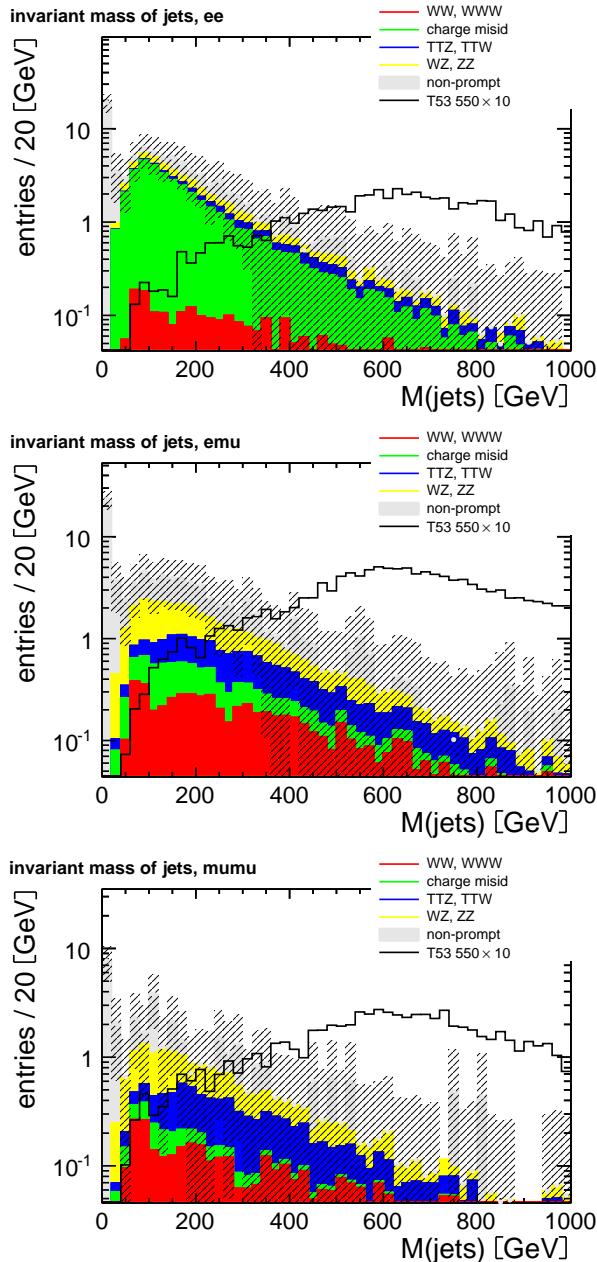


Figure 47: Invariant mass of the hadronic system for the three decay channels. Events with two same-sign leptons and at least two jets are shown.

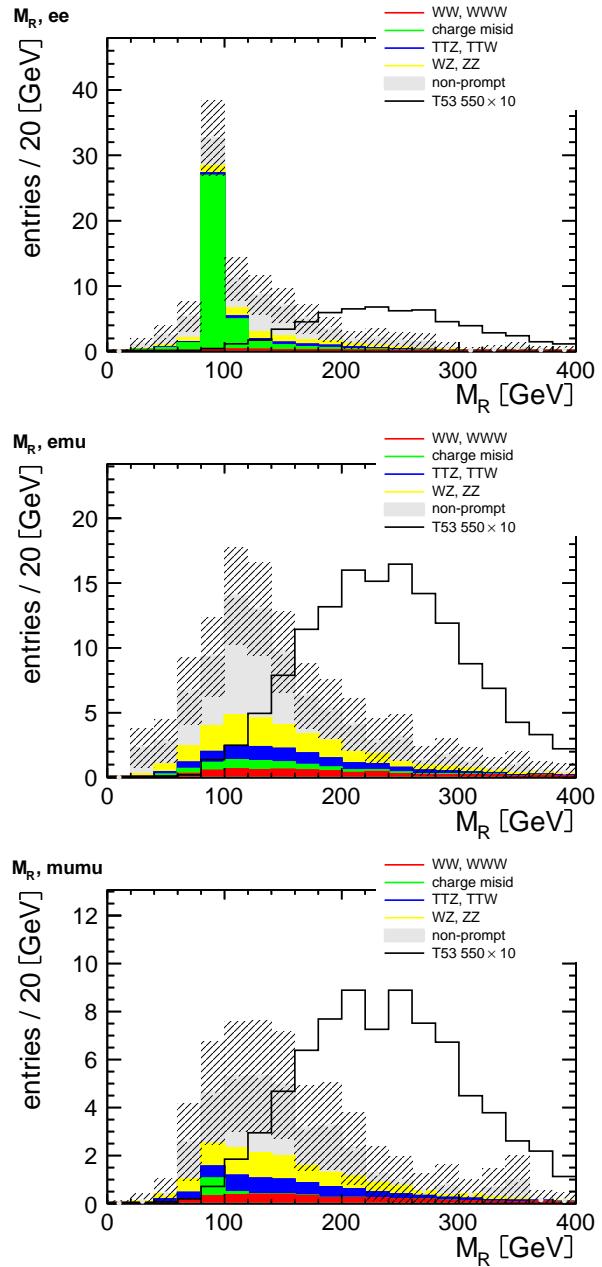


Figure 48:  $M_R$  distributions for the three decay channels. Events with two same-sign leptons and at least two jets are shown.

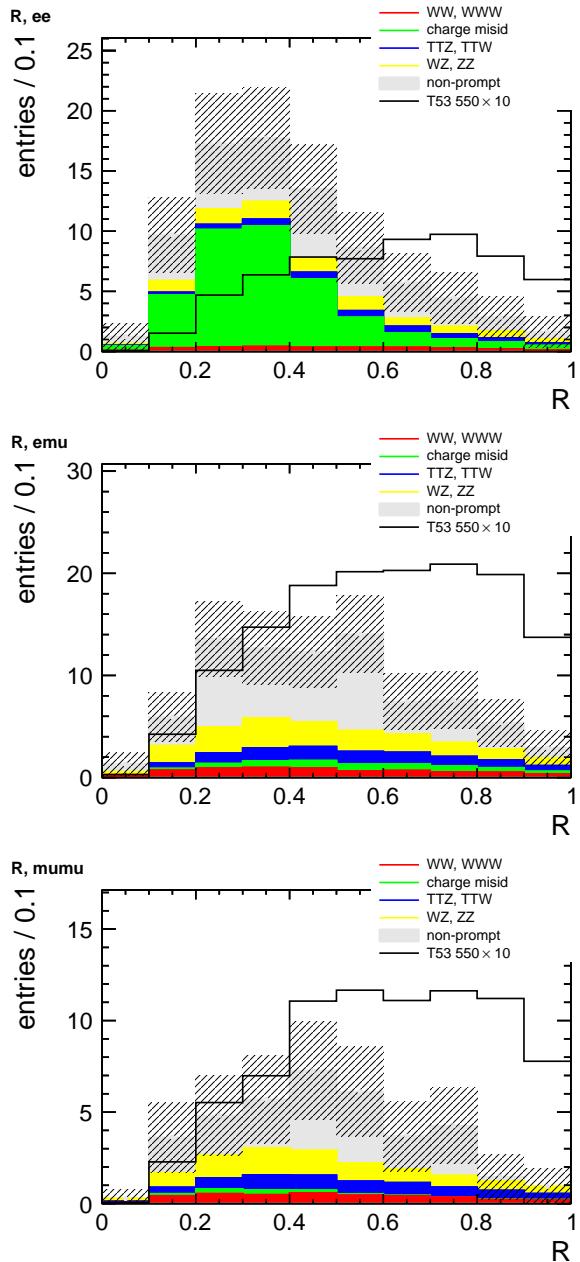


Figure 49:  $R$  distributions for the three decay channels. Events with two same-sign leptons and at least two jets are shown.

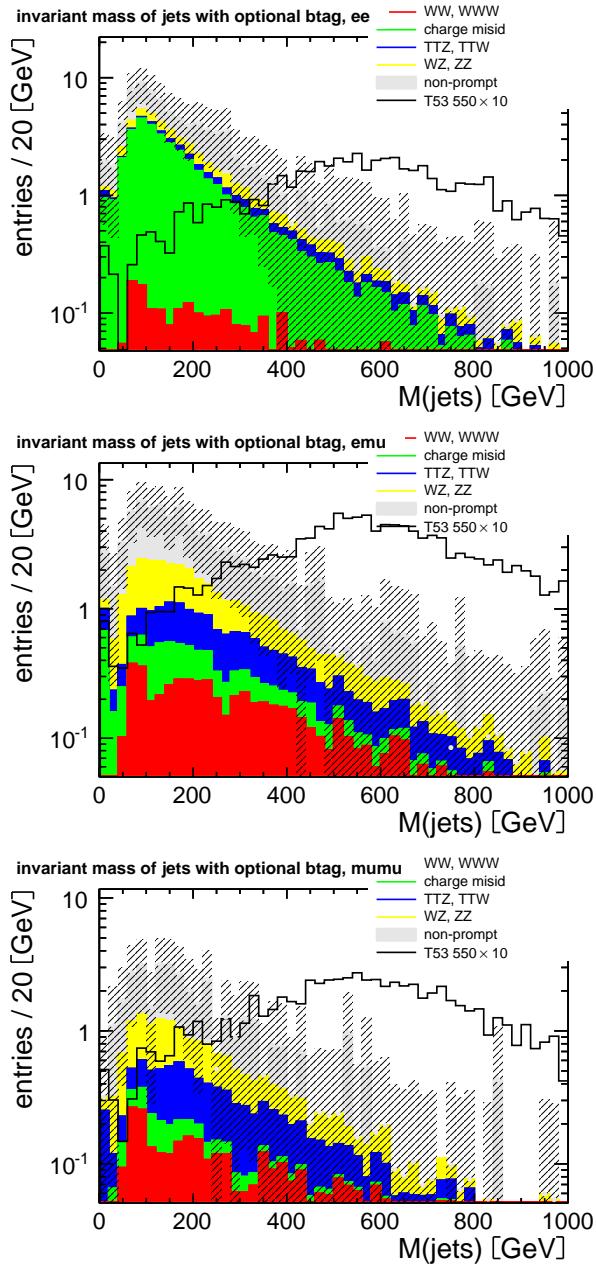


Figure 50: Invariant mass of the hadronic system for the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown.

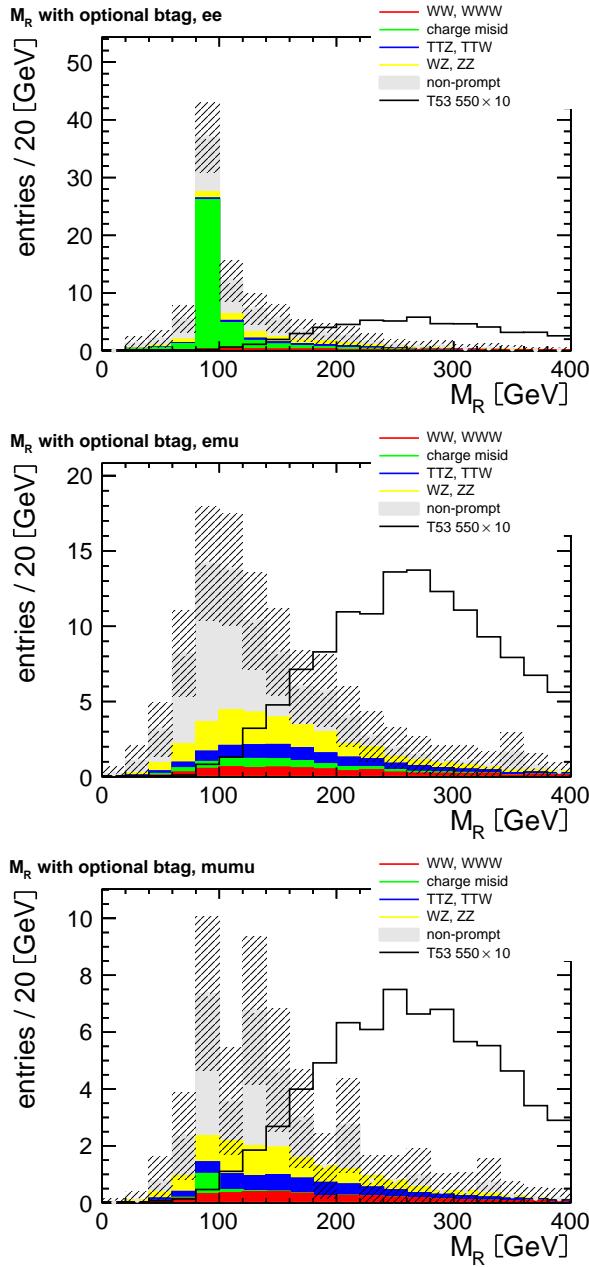


Figure 51:  $M_R$  distributions for the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown.

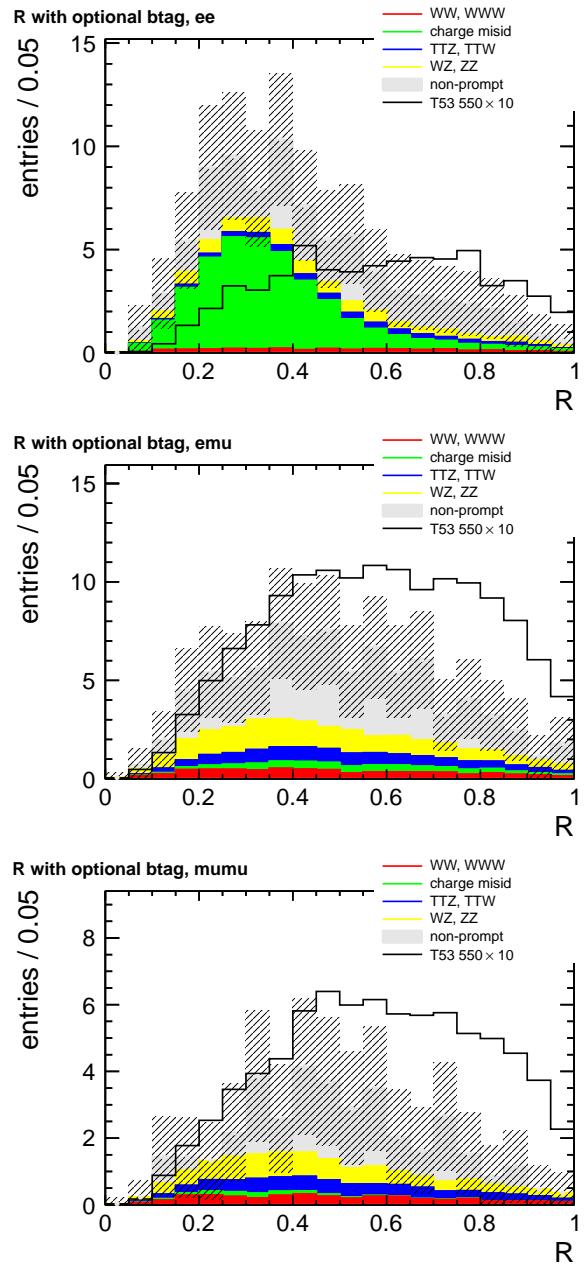


Figure 52: R distributions for the three decay channels, with the b-tag correction. Events with two same-sign leptons and at least two jets are shown.

## BIBLIOGRAPHY

---

- [1] Search for supersymmetry with the razor variables at CMS. 2012.
- [2] A. Adiguzel et al. Search for exotic wz resonances at cms. CMS Analysis Note CMS AN-2011/333, 2011.
- [3] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.
- [4] Wolfgang Adam, Boris Mangano, Thomas Speer, and Teddy Todorov. Track reconstruction in the cms tracker. Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec 2006.
- [5] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, et al. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput.Phys.Commun.*, 182:1034–1046, 2011. doi: 10.1016/j.cpc.2010.12.040.
- [6] B.C. Allanach, C.G. Lester, Michael Andrew Parker, and B.R. Webber. Measuring sparticle masses in nonuniversal string inspired models at the LHC. *JHEP*, 0009:004, 2000.
- [7] Johan Alwall et al. MadGraph/MadEvent v4: The New Web Generation. *JHEP*, 09:028, 2007.
- [8] The ATLAS. Procedure for the lhc higgs boson search combination in summer 2011. Technical Report CMS-NOTE-2011-005, CERN, Geneva, Aug 2011.
- [9] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron reconstruction in cms. *The European Physical Journal C - Particles and Fields*, 49:1099–1116, 2007. ISSN 1434-6044. URL <http://dx.doi.org/10.1140/epjc/s10052-006-0175-5>. 10.1140/epjc/s10052-006-0175-5.
- [10] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi. Electron reconstruction within the particle flow algorithm. *CMS Analysis Note*, (AN-2010/034), 2010.
- [11] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- $k_T$  jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008. URL <http://stacks.iop.org/1126-6708/2008/i=04/a=063>.

- [12] Serguei Chatrchyan et al. Search for new physics with same-sign isolated dilepton events with jets and missing transverse energy at the LHC. *JHEP*, 1106:077, 2011. doi: 10.1007/JHEP06(2011)077.
- [13] Serguei Chatrchyan et al. Search for heavy bottom-like quarks in 4.9 inverse femtobarns of pp collisions at  $\sqrt{s} = 7$  tev. Comments: Submitted to JHEP, Apr 2012.
- [14] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [15] Serguei Chatrchyan et al. Search for new physics in events with same-sign dileptons and b-tagged jets in pp collisions at  $\sqrt{s} = 7$  TeV. Submitted to JHEP, 2012.
- [16] CMS Collaboration. Reconstruction of electron tracks with the Gaussian-Sum Filter in the CMS tracker at LHC. *CMS Analysis Note*, (RN-2003-001), 2003.
- [17] CMS Collaboration. Algorithms for b jet identification in CMS. Jul 2009.
- [18] CMS Collaboration. Fast simulation of the CMS detector. *Journal of Physics, Conference Series*, (219), 2010.
- [19] CMS Collaboration. Particle flow event reconstruction in CMS and performance for jets, taus and  $E_T^{\text{miss}}$ . *CMS Physics Analysis Summary*, (PFT-9-001), 2010.
- [20] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6:11002, November 2011. doi: 10.1088/1748-0221/6/11/P11002.
- [21] CMS Collaboration. b-tagging performance (efficiency and mistag rate) measurements from CMS. July 2011.
- [22] CMS Collaboration. Search for heavy  $w'$  boson resonances decaying to a bottom quark and a top quark at  $\sqrt{s}=7$  TeV. Technical report, 2012. URL <http://cdsweb.cern.ch/record/1447758>.
- [23] CMS Collaboration. Absolute calibration of the luminosity measurement at cms: Winter 2012 update. Technical report, 2012. URL <http://cdsweb.cern.ch/record/1434360>.
- [24] The CMS collaboration. ECAL detector performance, 2011 data. May 2012.
- [25] K. Cranmer. Statistical Challenges for Searches for New Physics at the LHC. In L. Lyons and M. Karagöz Ünel, editors, *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, page 112, 2006. doi: 10.1142/9781860948985\_0026.

- [26] R Fruhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive vertex fitting. Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [27] G. 't Hooft. Recent Developments in Gauge Theories. In *NATO Science Series B: Physics*, page 438, 1969.
- [28] GEANT4 Collaboration. GEANT 4 — a simulation toolkit. *Nucl. Intrum. Meth.*, (056), 2003.
- [29] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961. ISSN 0029-5582. doi: 10.1016/0029-5582(61)90469-2. URL <http://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [30] Jay Hubisz, Joseph Lykken, Maurizio Pierini, and Maria Spiropulu. Missing energy look-alikes with  $100 \text{ pb}^{-1}$  at the LHC. *Phys.Rev.*, D78:075008, 2008. doi: 10.1103/PhysRevD.78.075008.
- [31] L. Moneta, K. Cranmer, G. Schott, and W. Verkerke. The RooStats project. In *Proceedings of the 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research. February 22-27, 2010, Jaipur, India*. p.57, 2010.
- [32] Jan Mrazek and Andrea Wulzer. A Strong Sector at the LHC: Top Partners in Same-Sign Dileptons. *Phys.Rev.*, D81:075006, 2010. doi: 10.1103/PhysRevD.81.075006.
- [33] Giacomo Polesello and Daniel R. Tovey. Supersymmetric particle mass measurement with the boost-corrected contransverse mass. *JHEP*, 1003:030, 2010. doi: 10.1007/JHEP03(2010)030.
- [34] G. Punzi. Sensitivity of Searches for New Signals and Its Optimization. In L. Lyons, R. Mount, and R. Reitmeyer, editors, *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, page 79, 2003.
- [35] Christopher Rogan. Kinematical variables towards new dynamics at the LHC. 2010.
- [36] Christopher Rogan.  $p_T$  corrections for razor variables. Apr 2012.
- [37] M. Rudowicz and S. Peters. Grindhammer G. *Nucl. Instrum. Meth. A*, (290):469, 1990.
- [38] N. Saoulidou. Particle flow jet identification criteria. *CMS analysis note*, (004), 2010.
- [39] T. Sjostrand, S.Mrenna, and P.Z.Skands. PYTHIA 6.4 physics and manual. *JHEP*, (05-026), 2006.

- [40] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. doi: 10.1103/PhysRevLett.19.1264. URL <http://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.