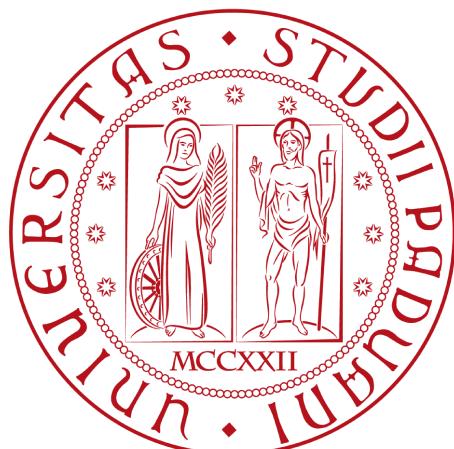


MATTEO ABIS

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

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

A.A. 2011/2012

Matteo Abis: *A search for exotic particles of charge $5/3$ with the CMS detector*, © A.A. 2011/2012

Ohana means family.
Family means nobody gets left behind, or forgotten.
— Lilo & Stitch

Dedicated to the loving memory of Rudolf Miede.

1939 – 2005

ABSTRACT

Short summary of the contents in English...

ZUSAMMENFASSUNG

Kurze Zusammenfassung des Inhaltes in deutscher Sprache...

ACKNOWLEDGMENTS

Put your acknowledgments here.

Many thanks to everybody who already sent me a postcard!

Regarding the typography and other help, many thanks go to Marco Kuhlmann, Philipp Lehman, Lothar Schlesier, Jim Young, Lorenzo Pantieri and Enrico Gregorio¹, Jörg Sommer, Joachim Köstler, Daniel Gottschlag, Denis Aydin, Paride Legovini, Steffen Prochnow, Nicolas Repp, Hinrich Harms, Roland Winkler, and the whole L^AT_EX-community for support, ideas and some great software.

Regarding LyX: The LyX port was initially done by Nicholas Mariette in March 2009 and continued by Ivo Pletikosić in 2011. Thank you very much for your work and the contributions to the original style.

¹ Members of GuIT (Gruppo Italiano Utilizzatori di T_EX e L^AT_EX)

CONTENTS

1 THE THEORY OF THE TOP PARTNERS	1
1.1 The Standard Model	1
1.1.1 The unbroken SM	1
1.1.2 Spontaneous symmetry breaking: the Higgs mechanism	2
1.1.3 The hierarchy problem	3
1.1.4 The top partners	3
2 THE LHC AND THE CMS DETECTOR	5
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
3 SIMULATION, RECONSTRUCTION AND DATASETS	11
3.1 Simulation	11
3.2 The reconstruction of physics objects	12
3.2.1 Muon reconstruction	13
3.2.2 Electron reconstruction	14
3.2.3 Jet reconstruction	16
3.3 Object selection	18
3.3.1 Muon selection	18
3.3.2 Electron selection	18
3.3.3 Jet selection	20
3.3.4 Event cleanup and vertex selection	20
3.3.5 Trigger requirements	21
3.4 Datasets	21
3.5 Signal Monte Carlo samples	22
3.6 Background Monte Carlo samples	22
3.7 Monte Carlo pile-up weighting	23
4 EVENT PRESELECTION	25
5 A DATA-DRIVEN BACKGROUND ESTIMATE	29
5.1 Charge misidentification	29
5.2 Same-sign non-prompt background	30
5.2.1 Heavy flavour content of the samples	33
5.2.2 Prompt ratio approximation	34
5.2.3 Calculation of the expected event yields	34
5.2.4 Closure tests	36
6 THE RAZOR VARIABLES	39

6.1	The variable M_R	39
6.2	The transverse plane and the R variable	40
6.3	The razor variables for the top partners	40
BIBLIOGRAPHY		43

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 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	13
Figure 10	Illustration of the isolation observable.	14
Figure 11	An illustration of the <i>bremstrahlung</i> recovery algorithm.	15
Figure 12	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.	19
Figure 13	Selection fake rate for the various working points, measured in QCD dijet events. In this analysis the HyperTightMC1 selection (light blue squares) is used.	20
Figure 14	Pair-production of the $T_{5/3}$, decaying into same-sign leptons and at least four jets.	25
Figure 15	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 ee , $e\mu$ and $\mu\mu$, and their sum. 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.	26

- Figure 16 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 ee, e μ and $\mu\mu$, and their sum. 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. 27
- Figure 17 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. 30
- Figure 18 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. 33
- Figure 19 Simple example of a supersymmetric event, with two heavy pair-produced SUSY particles, two invisible lightest supersymmetric particles and two visible SM particles. 39
- Figure 20 Comparison of a simple supersymmetric example with the top partner decay. 41

LIST OF TABLES

- Table 1 List of triggers used in the analysis for the three decay channels: $\mu\mu$, ee, e μ . CCTT is short for CaloIdL_CaloIsoVL_TrkIdVL_TrkIsoVL. 21
- Table 2 Data samples used for the analysis. 22
- Table 3 Signal Monte Carlo samples. The branching ratio is 0.21. 23
- Table 4 Details of the background Monte Carlo samples used for the analysis. 23
- 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. 29

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. 29
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. 32
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. 32
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 $25\text{-}35\text{ GeV}$ 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. 33
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 $25\text{-}35\text{ GeV}$ 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. 34
Table 11	Summary of the closure tests on Monte Carlo samples for events with two same-sign leptons. Statistical errors only. 37

LISTINGS

ACRONYMS

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.

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.

1.1.1 *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 _e electron neutrino	v _μ muon neutrino	v _τ tau neutrino
	e electron	μ muon	τ tau
Bosons (Forces)			
	0	0	0
	0	0	1
	1		
	γ photon	g gluon	Z weak force
	W ⁺ weak force	W ⁻ weak force	

Figure 1: The particles in the unbroken SM.

1.1.2 Spontaneous symmetry 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 with Yukawa terms.

1.1.3 The hierarchy problem

Recently found evidence for a light Higgs boson, with a mass near 125 GeV, 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. Theoretical physicists generally dislike this fine tuning on grounds of *naturalness*.

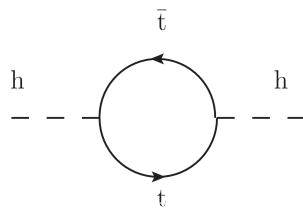


Figure 2: Radiative correction to the Higgs mass from the top quark.

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.

1.1.4 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 [15] 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.

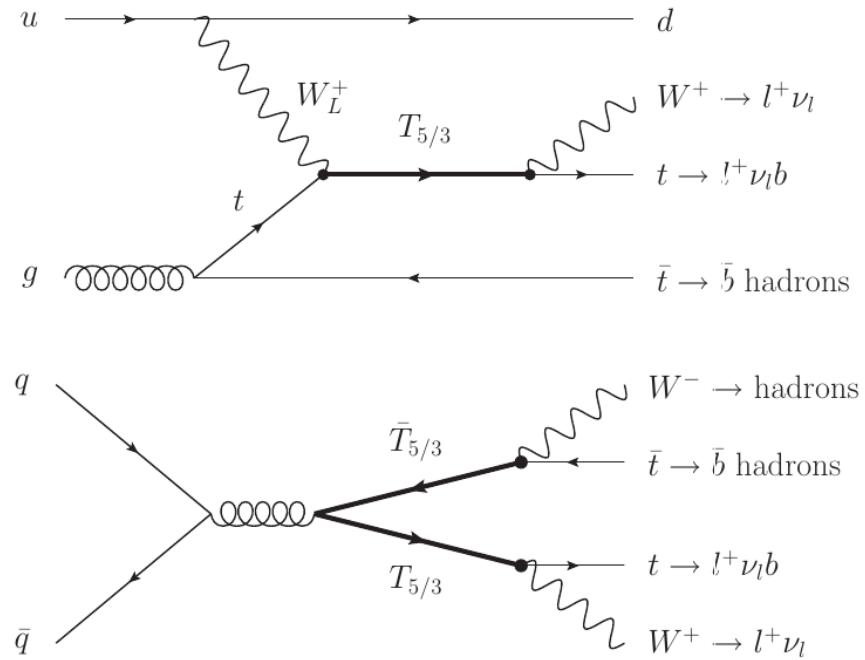


Figure 3: Typical single and pair production the top partner $T_{5/3}$ at the LHC.

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 75 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 mainly 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 φ angle is measured from the x axis, and the polar angle θ , 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.

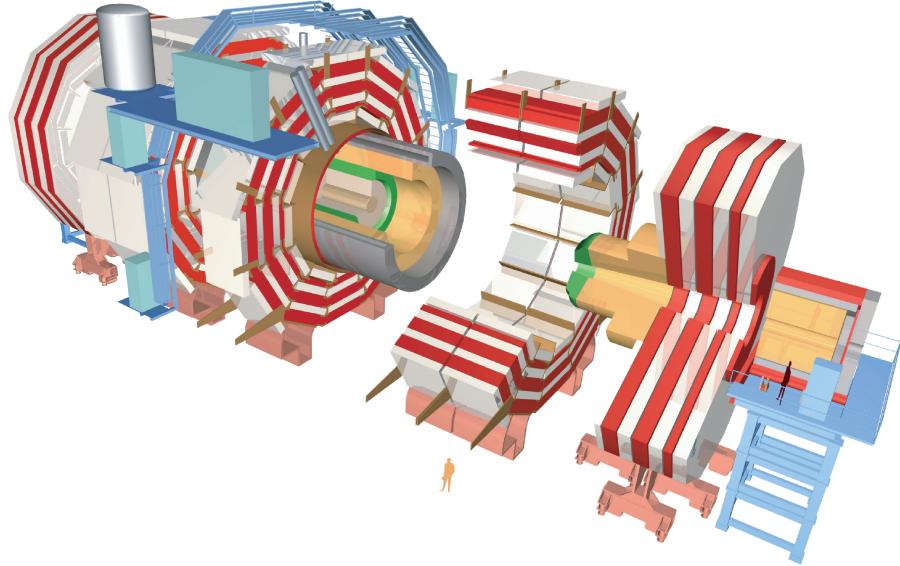


Figure 5: The CMS detector.

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-

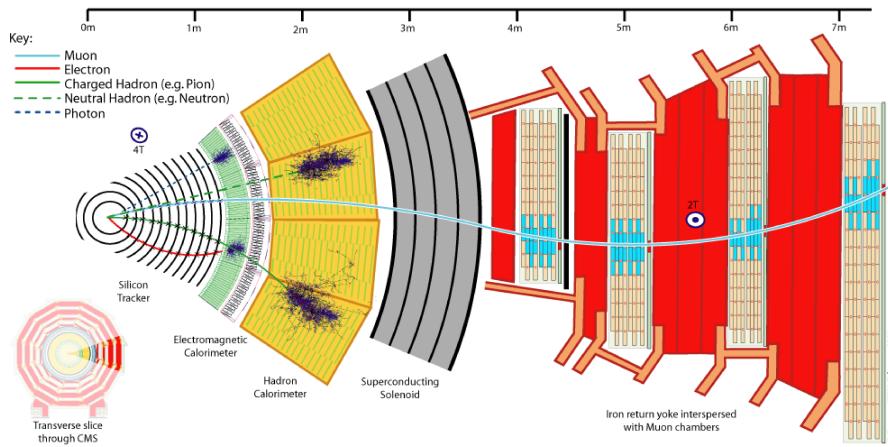


Figure 6: Transverse view of the CMS detector with some examples of tracks.

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 in the barrel and cathode strip chambers 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 $r\phi$ 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 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 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 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 [13], as shown in figure 7.

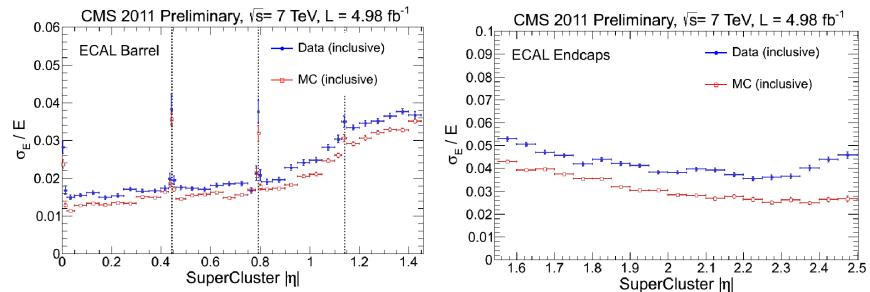


Figure 7: Energy resolution of the ECAL as measured in $Z \rightarrow ee$ events in the 2011 data.

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 id 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.

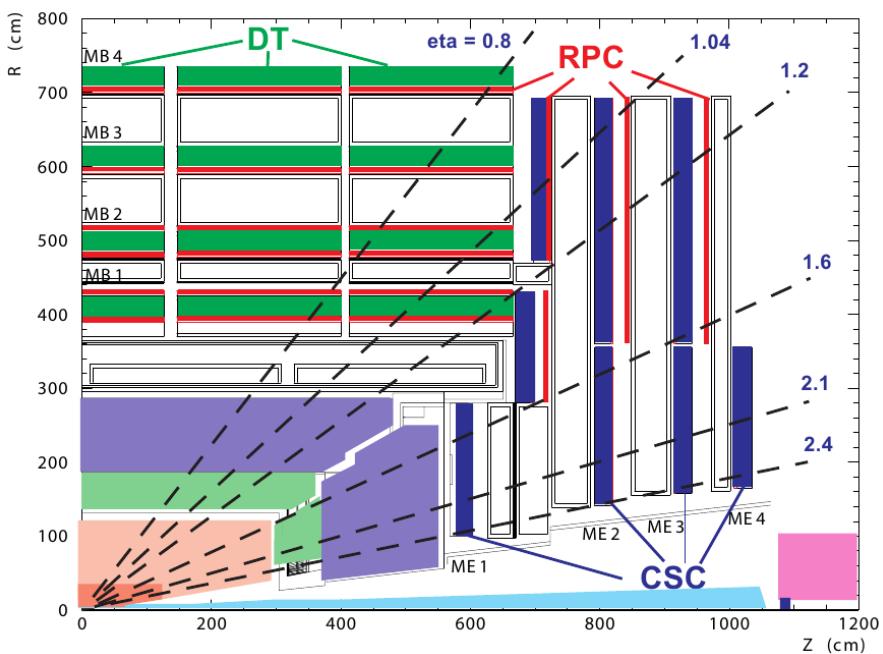


Figure 8: The layout of the muon detectors.

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 search for new physics phenomena. The physics of the collisions is simulated by the event generators MadGraph [?] and PYTHIA [18]. 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 [14] 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* [9]. Both the FullSim and the FastSim were employed in the generation of the background samples for our studies.

In the FullSim, the energy depositions 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 electronics; 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.

The FastSim makes a number of approximations:

CMS GEOMETRY: the FastSim 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 *bremsstrahlung*, photon conversion, multiple Coulomb scattering, 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. The FastSim includes it because, given the low

fake rate in the reconstruction, it is possible to emulate it at a lower computational cost. Only the hits of the simulated track are used to make a track candidate.

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 [16] 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 FullSim results.

3.2 THE RECONSTRUCTION OF PHYSICS OBJECTS

The particle flow [10] (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.

Reconsider the design of the detector in figure 6. The logic you use 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 by far the easiest particles to recognize.

From these basic elements, composite objects can be reconstructed, such as τ 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 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).

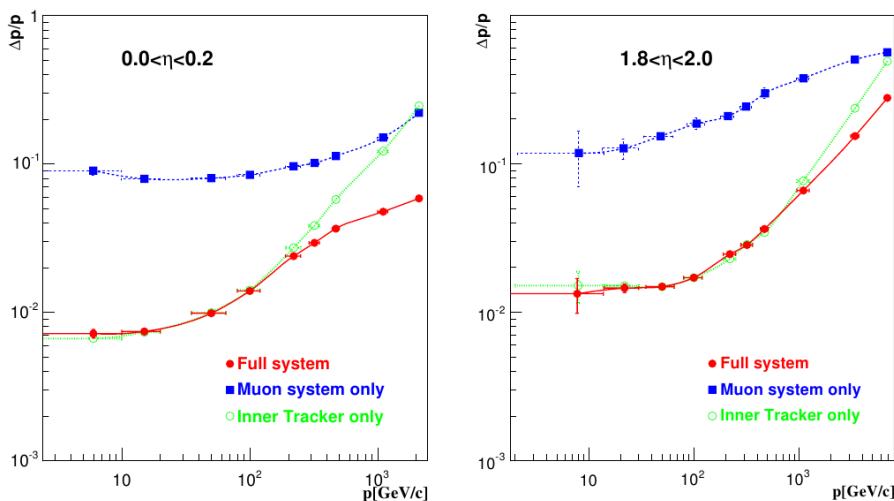


Figure 9: Muon momentum resolution as a function of p_T in regions with different pseudorapidity. The blue squares are for the muon system only, the green circles for the inner tracker only, and the red circles for the combined reconstruction

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.

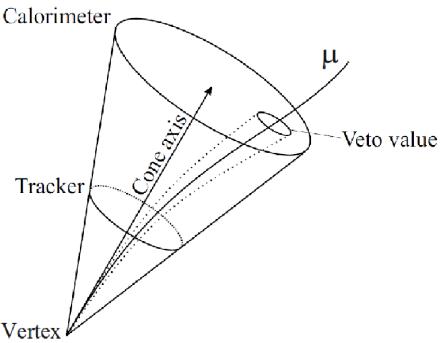


Figure 10: Illustration of the isolation observable.

3.2.2 Electron reconstruction

The tracks in CMS are reconstructed assuming that the particle is a muon. Coulomb scattering is the dominant effect on muons crossing material and its impact is modelled by gaussian fluctuations. This approach fails with electrons because of the highly non-gaussian *bremstrahlung* emission. A customized track reconstruction was developed for electrons, where the hits are fitted using a *gaussian sum filter* (GSF) [7], i.e. the *bremstrahlung* 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 *bremstrahlung* 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 contrast with this ECAL-driven approach, the strategy developed for the Particle Flow [3] starts from the tracks and can be explained with two extreme examples. When the *bremstrahlung* emission is negligible, the electron trajectory is determined with good precision by the standard tracking algorithm, and the track can be recon-

structed 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 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 χ^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.

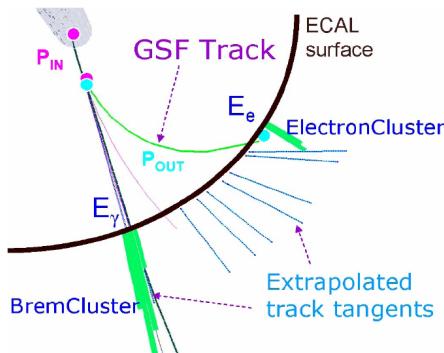


Figure 11: An illustration of the *bremstrahlung* recovery algorithm.

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 [2]. 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.

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

3.2.3 *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
- pile-up and initial state radiation may contribute with additional tracks

The anti- k_T algorithm [4] 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 φ_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 [11]. 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 removes 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 η , 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 η 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 [17] are also applied, by selecting jets with:

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

3.2.3.1 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 [8] 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 jet passing this selection to 10%, as measured on jets with p_T of about 80 GeV in QCD Monte Carlo events [12].

3.3 OBJECT SELECTION

Signal events in our studies have a clean signature: two clean, isolated, same-sign leptons and many jets. 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 η

IMPACT PARAMETER with respect to the primary vertex

CONVERSION REJECTION with missing hits and partner track matching, as described in paragraph 3.2.2

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 12 and 13.

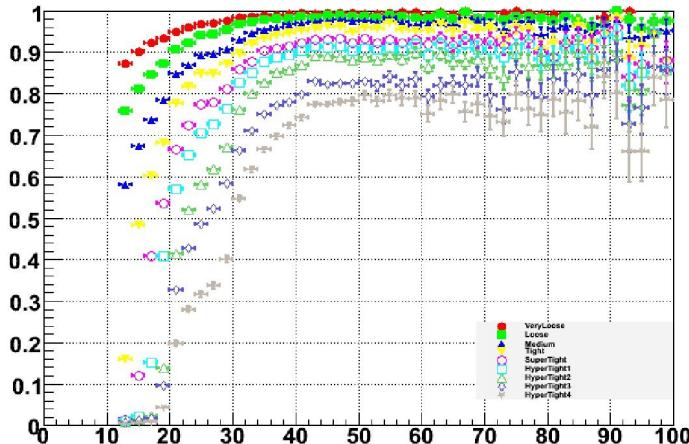


Figure 12: Selection efficiency for the various working points, measured in $Z \rightarrow ee$ and $W \rightarrow ev$ events. In this analysis the HyperTightMC1 selection (light blue squares) is used.

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

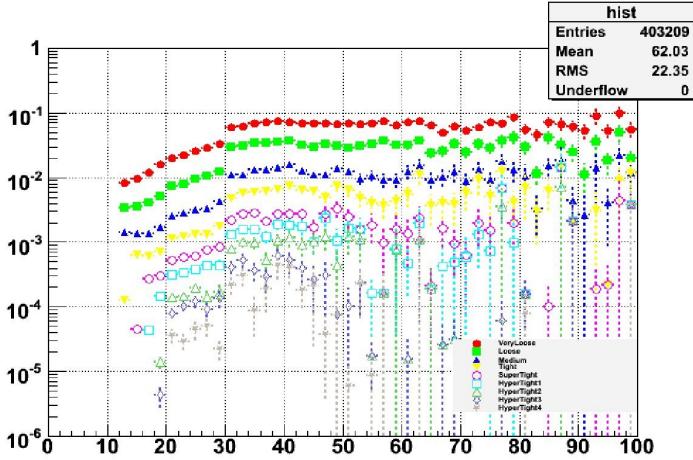


Figure 13: Selection fake rate for the various working points, measured in QCD dijet events. In this analysis the HyperTightMC1 selection (light blue squares) is used.

- conversion rejection
- transverse impact parameter $< 0.02 \text{ cm}$
- GSF-CTF-SC charge consistency (see paragraph 3.2.2)
- $\Delta R > 0.1$ 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.

3.3.3 Jet selection

The jet selection has only few requirement in addition to the baseline reconstruction described in 3.2.3:

- $p_T > 30 \text{ GeV}$
- $|\eta| < 2.4$
- $\Delta R > 0.3$ 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

3.3.5 Trigger requirements

HLT selections are also used in our analysis. Double lepton triggers are employed with a relatively high 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 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 ?? 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 [1]. The masses, cross sections and the number of generated events for the eight mass points are listed in table 3.

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, ttW and ttZ. 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.

mass (GeV)	$\sigma \times \text{BR} (\text{pb})$	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.

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

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

3.7 MONTE CARLO PILE-UP WEIGHTING

The pile-up in events at the LHC is constantly evolving with the increase in instantaneous luminosity at which the machine operates. The simulated events are usually produced before the data are collected, therefore they need taking into account the measured pile-up conditions in order to avoid a systematic bias, as shown in figure ??

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 14).

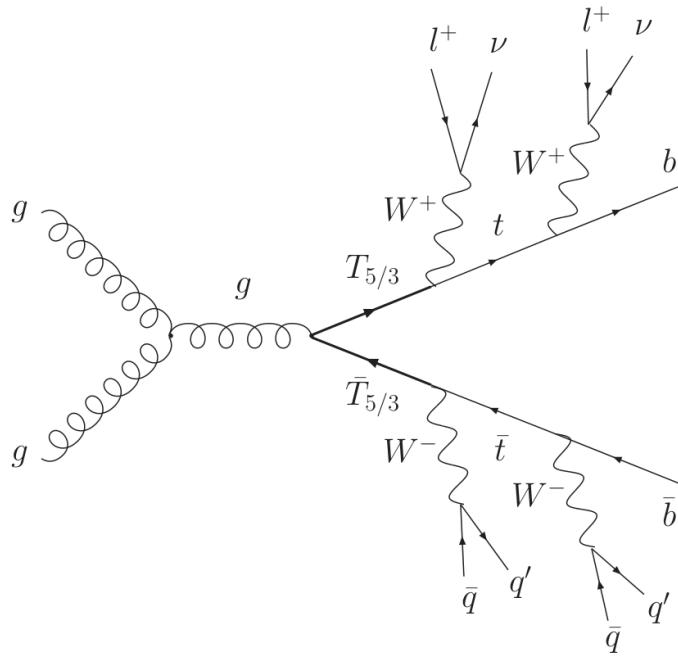


Figure 14: Pair-production of the $T_{5/3}$, decaying into same-sign leptons and at least four jets.

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 window. This selection is not needed for ee 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 15, particularly from $Z \rightarrow ee$ decays. Figure 16 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 15) also shows that this is a good discriminating variable against both the prompt and the $t\bar{t}$ backgrounds.

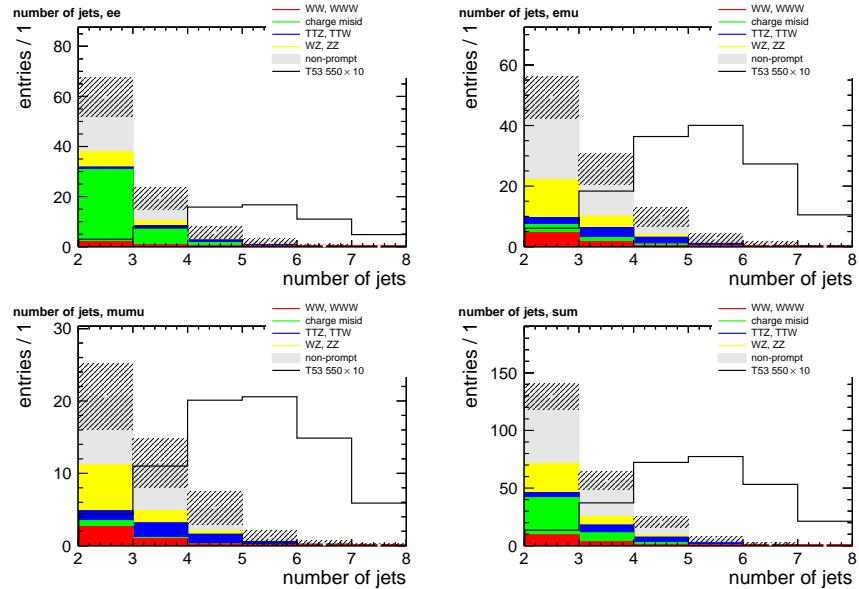


Figure 15: Events with two same-sign leptons and two jets. Distribution of the number of jets in the backgrounds and signal for the 550 GeV mass point for the sum of the three decay channels ee, $e\mu$ and $\mu\mu$, and their sum. The signal is amplified by factor of ten, the data-driven backgrounds 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 accurate description of the contributions from charge misidentification and non-prompt leptons. That is the subject of the following chapter.

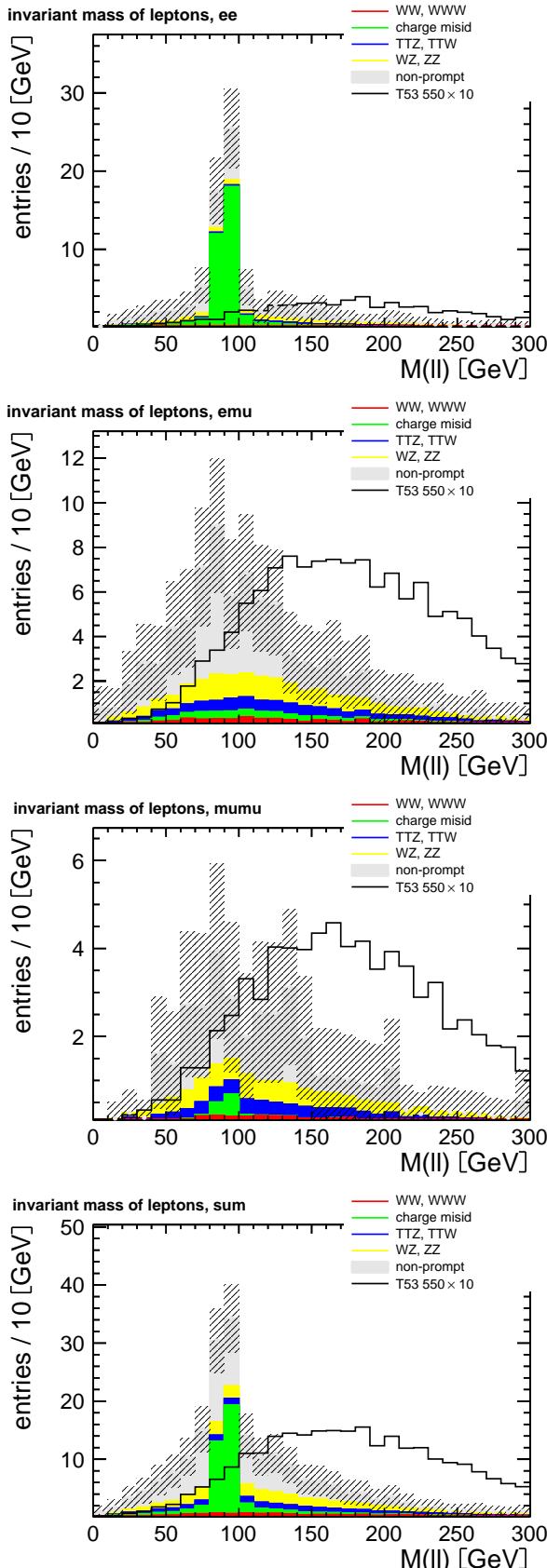


Figure 16: Events with two same-sign leptons and two jets. Distribution of the invariant mass of the leptons in the backgrounds and signal for the 550 GeV mass point for the sum of the three decay channels ee , $e\mu$ and $\mu\mu$, and their sum. The signal is amplified by factor of ten, the data-driven backgrounds for the non-prompt and charge misidentification contributions are detailed in chapter 5. The [18th August 2012 at 17:56 – classic thesis]

A DATA-DRIVEN BACKGROUND ESTIMATE

We identified three sources of backgrounds (page 25). The backgrounds from charge misidentification and from non-prompt leptons are analysed with data driven techniques.

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 misidentification rate is extremely small ($\approx 10^{-5}$) and their contribution to the background is negligible [6]. 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+jets events and is found to be 5.89×10^{-4} . To confirm that this upper limit is close to the true probability, we consider the dilepton invariant mass spectrum shown in Figure 17. 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 [6].

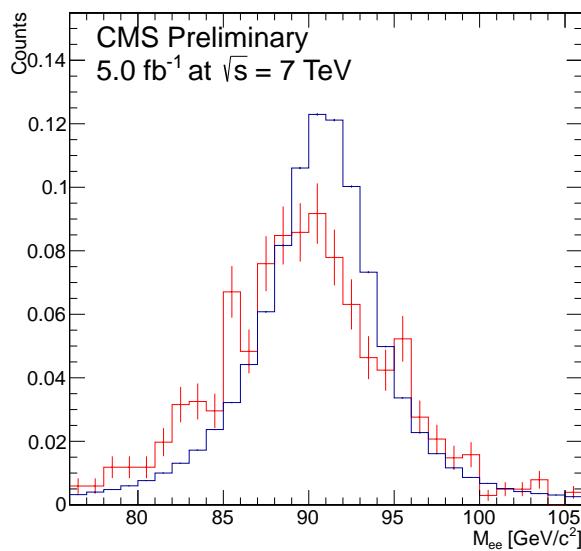


Figure 17: The dilepton invariant mass spectrum for electrons in data. The opposite-sign distributions are shown in blue and the same-sign ones are in red. Both distributions are normalized to unit area.

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. 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 76-106 GeV window

The E_T^{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 fairly large. 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 ± 0.21	29.3 ± 0.20	-
Z + jets	21.9 ± 0.08	20.4 ± 0.08	-
t̄t	0.7 ± 0.01	0.6 ± 0.01	-
Data	1974	274	0.138 ± 0.009
Data-MC	1919.8 ± 44.43	223.7 ± 16.55	0.117 ± 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 ± 26.36	4374.8 ± 25.97	-
Z + jets	604.0 ± 4.36	585.8 ± 4.29	-
t̄t	85.5 ± 1.20	80.4 ± 1.16	-
Data	33770	12170	0.360 ± 0.004
Data-MC	28575.5 ± 185.70	7129.1 ± 113.42	0.249 ± 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 18).

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:

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

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

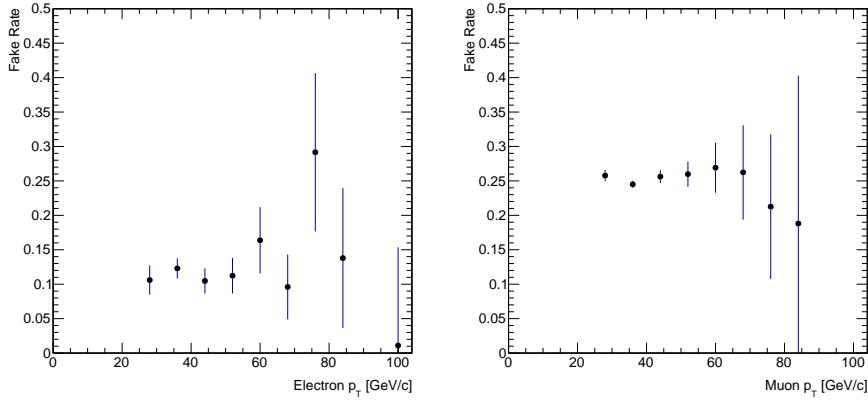


Figure 18: The corrected tight to loose ratio, loosely speaking the *fake rate*, as a function of p_T for electrons (left) and muons (right). The correction is accomplished by subtracting the MC-based W, Z and $t\bar{t}$ contributions from the data.

	loose	tight	ratio
W + jets	7.0 ± 0.10	6.2 ± 0.09	-
Z + jets	8.1 ± 0.05	7.2 ± 0.05	-
$t\bar{t}$	0.1 ± 0.00	0.1 ± 0.00	-
Data	1698	191	0.112 ± 0.009
Data-MC	1682.9 ± 41.21	177.5 ± 13.82	0.105 ± 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 $t\bar{t}$ from the data.

5.2.1 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 $t\bar{t}$. We therefore have to check that our prediction is stable with respect to variations in the flavour content.

The b-tagging algorithm described in 3.2.3.1 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:

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

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

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

5.2.2 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.

5.2.3 Calculation of the expected event yields

The measure of the non-prompt rate f_ℓ 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_ℓ 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:

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

$$N_{t1} = N_p + f N_f$$

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

These relations can be easily inverted to obtain N_p and N_f , and mostly the number of non-prompt leptons passing the tight selec-

tion, 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 ε .

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.

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

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 ε , 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 ε , the other by 1, hence the factor 2.

Finally these formulae can be extended to the 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:

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

5.2.4 Closure tests

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 fake rate was calculated in QCD samples (`QCD_Pt-15to3000` for electrons and `QCD_Pt-20_MuEnrichedPt-15` for muons), without the cuts on the E_T^{miss} and m_T , as there are no vector bosons in the QCD MC:

$$\begin{aligned} f_e &= 0.091 \pm 0.017 \\ f_\mu &= 0.152 \pm 0.002 \end{aligned}$$

These numbers differ from the values in the data since the MC does not describe accurately the composition of the non-prompt events.

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 5.2.3 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 [5].

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

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

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 the discovery of new physics [?].

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 [?]. 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^{miss} (figure 19).

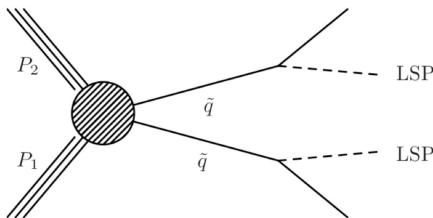


Figure 19: Simple example of a supersymmetric event, with two heavy pair-produced SUSY particles, two invisible lightest supersymmetric particles and two visible SM particles.

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:

$$M_{R0} := \sqrt{(a^0 - b^0)^2 - (a^3 + b^3)^2}$$

Where a^μ and b^μ are the four momenta of the 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^{miss} , the corrections in [?] 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^{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 20). The obvious

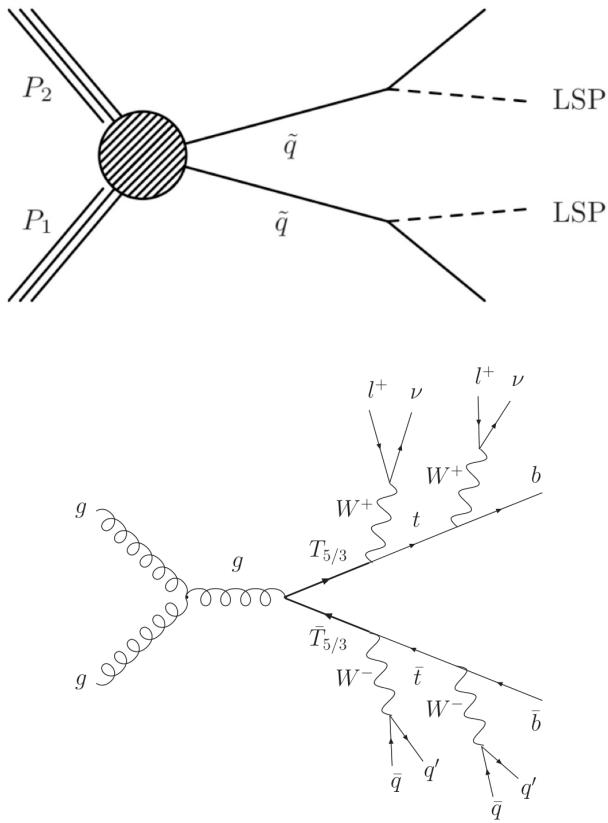


Figure 20: Comparison of a simple supersymmetric example with the top partner decay.

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

One more approximation remains, by associating 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.

BIBLIOGRAPHY

- [1] 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.
- [2] 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.
- [3] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi. Electron reconstruction within the particle flow algorithm. *CMS Analysis Note*, (AN-2010/034), 2010.
- [4] 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>.
- [5] 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.
- [6] 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.
- [7] CMS Collaboration. Reconstruction of electron tracks with the Gaussian-Sum Filter in the CMS tracker at LHC. *CMS Analysis Note*, (RN-2003-001), 2003.
- [8] CMS Collaboration. Algorithms for b jet identification in CMS. Jul 2009.
- [9] CMS Collaboration. Fast simulation of the CMS detector. *Journal of Physics, Conference Series*, (219), 2010.
- [10] CMS Collaboration. Particle flow event reconstruction in CMS and performance for jets, taus and E_T^{miss} . *CMS Physics Analysis Summary*, (PFT-9-001), 2010.
- [11] 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.

- [12] CMS Collaboration. b-tagging performance (efficiency and mistag rate) measurements from CMS. July 2011.
- [13] The CMS collaboration. ECAL detector performance, 2011 data. May 2012.
- [14] GEANT4 Collaboration. GEANT 4 — a simulation toolkit. *Nucl. Instrum. Meth.*, (056), 2003.
- [15] 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](https://doi.org/10.1103/PhysRevD.81.075006).
- [16] M. Rudowicz and S. Peters. Grindhammer G. *Nucl. Instrum. Meth. A*, (290):469, 1990.
- [17] N. Saoulidou. Particle flow jet identification criteria. *CMS analysis note*, (004), 2010.
- [18] T. Sjostrand, S.Mrenna, and P.Z.Skands. PYTHIA 6.4 physics and manual. *JHEP*, (05-026), 2006.